Computable Prediction

Kenshi Miyabe[✉]

Department of Mathematics, Meiji University, Kawasaki, Kanagawa, Japan
research@kenshi.miyabe.name
http://kenshi.miyabe.name/wordpress

Abstract. We try to predict the next bit from a given finite binary string when the sequence is sampled from a computable probability measure on the Cantor space. There exists the best betting strategy among a class of effective ones up to a multiplicative constant, the induced prediction from which is called algorithmic probability or universal induction by Solomonoff. The prediction converges to the true induced measure for sufficiently random sequences. However, the prediction is not computable.

We propose a framework to study the properties of computable predictions. We prove that all sufficiently general computable predictions also converge to the true induced measure. The class of sequences along which the prediction converges is related to computable randomness. We also discuss the speed of the convergence. We prove that, even when a computable prediction predicts a computable sequence, the speed of the convergence cannot be bounded by a computable function monotonically decreasing to 0.

Keywords: Algorithmic probability · Universal induction · Computable randomness

1 Introduction

Given data, one finds regularity and predicts the future. This is what all living things are doing and what we want to make machines do. Is there a universal way of doing this? If so, what properties should the prediction have?

Solomonoff's algorithmic probability or universal induction answers. For simplicity, consider the case of infinite binary sequences, that is, the Cantor space 2^ω. We try to predict the next bit from a given finite binary string. It is known that there is an optimal c.e. semi-measure M on the Cantor space. Here, a function is c.e. if it can be computably approximated from below and optimality means that it dominates all c.e. semi-measures up to a multiplicative constant. See the definition in Sect. 2.1. By the usual correspondence between measures and martingales (or semi-measures and supermartingales), this roughly means that the prediction by M behaves at least as well as one by any c.e. semi-measure. The prediction induced from an optimal c.e. semi-measure is called *algorithmic probability* or *universal induction*.

Algorithmic probability has some desirable properties. One of them is the convergence to the true induced measure (Theorem 2.3 below). This roughly means that algorithmic probability can find any computable regularity unknown in advance.

In this paper, we propose a framework showing that all sufficiently general prediction should have some properties. That a program has a property does not mean that the program is general enough, however, that a program does not have a property means that the program can be modified to a more general one. By evaluating computational complexity that a function with a property should have, we can also discuss how difficult to add the property although we do not discuss much this in this paper.

We focus on the speed of the convergence. One of our results (Theorem 4.4) says that, for all sufficiently general computable predictions, the speed of the convergence to the true measure cannot be bounded by a computable function. Thus, incomputability of the rate of the convergence is not by the incomputability of algorithmic probability. Rather than that, it is by the existence of computable measures that are "close" to each other.

The structure of the paper as follows. In Sect. 2 we review some notions and results on algorithmic randomness and algorithmic probability. In Sect. 3 we prove the convergence result for computable predictions along computably random sequences. In Sect. 4 we consider the case of Dirac measures and show the incomputability of the rate of the convergence.

2   Preliminaries

2.1   Algorithmic Randomness

We follow the notation in computability theory (see e.g. [10]) or the theory of algorithmic randomness (see e.g. [1,8]).

The Cantor space 2^ω is the class of all infinite binary sequences equipped with the topology generated by the base elements of the cylinders $[\sigma] = \{X \in 2^\omega : \sigma \prec X\}$ where $\prec$ denotes the prefix relation. A function $f : \omega \to \omega$ is *computable* if it is computable by a Turing machine. The computability on $\mathbb{Q}$ or $2^{<\omega}$ is naturally induced by their natural representation by ω. A real $x \in \mathbb{R}$ is called *computable* if there exists a computable sequence $(a_n)_{n\in\omega}$ of rationals such that $|x - a_n| \leq 2^{-n}$ for all n. A real $x \in \mathbb{R}$ is called *left-c.e.* if there exists an increasing computable sequence $(a_n)_{n\in\omega}$ of rationals.

A function $f : 2^{<\omega} \to \mathbb{R}$ is called *computable* or *c.e.* if $f(\sigma)$ is computable or left-c.e. uniformly in $\sigma \in 2^{<\omega}$, respectively. A measure μ on 2^ω is *computable* if the function $\sigma \mapsto \mu([\sigma]) =: \mu(\sigma)$ is computable. A *semi-measure* is a function $\mu : 2^{<\omega} \to [0,1]$ such that $\mu(\lambda) \leq 1$ and $\mu(\sigma) \geq \mu(\sigma 0) + \mu(\sigma 1)$ for every $\sigma \in 2^{<\omega}$ where λ denotes the empty string. A measure μ with $\mu(\lambda) = \mu(2^\omega) = 1$ is called a *probability* measure. Notice that each semi-measure μ satisfying $\mu(\lambda) = 1$ and $\mu(\sigma) = \mu(\sigma 0) + \mu(\sigma 1)$ for every $\sigma \in 2^{<\omega}$ can be seen as a probability measure.

Let μ, ν be measures or semi-measures on 2^ω. We say μ *multiplicatively dominates* (or *m-dominates*) ν if, there exists $C \in \omega$ such that $\nu(\sigma) \leq C\mu(\sigma)$ for all

$\sigma \in 2^{<\omega}$. A c.e. semi-measure μ is called *optimal* if μ m-dominates every c.e. semi-measure. An optimal c.e. semi-measure exists.

Fix a computable probability measure μ. Martin-Löf randomness (or ML-randomness) is an important concept to talk about randomness of individual sequences. ML-randomness is usually defined by tests, but we give an equivalent characterization to compare it with the definition of computable randomness below. Let $X_{\leq n} = X_1 X_2 \cdots X_n$ be the initial segment of X with length n.

Theorem 2.1 ([6])**.** *A sequence $X \in 2^{\omega}$ is* ML-random w.r.t. μ *(or μ-ML-random) if and only if there exists a constant $C \in \omega$ such that $\xi(X_{\leq n}) \leq C\mu(X_{\leq n})$ for all n, where ξ is an optimal c.e. semi-measure.*

By the optimality, this is equivalent to the following statement: For every c.e. semi-measure ξ, there exists a constant $C \in \omega$ such that $\xi(X_{\leq n}) \leq C\mu(X_{\leq n})$ for all n.

The central notion in this paper is computable randomness. Computable randomness for the uniform measure is defined by martingales. However, Rute [9] has suggested the following definition for a general computable probability measure.

Definition 2.2. *A sequence $X \in 2^{\omega}$ is* computably random w.r.t. μ *(or μ-computably random) if $\mu(X_{\leq n}) > 0$ for all n and $\limsup_n \xi(X_{\leq n})/\mu(X_{\leq n}) < \infty$ for all computable measures ξ.*

It is not difficult to see that this is equivalent to the following statement: For every computable measure ξ, there exists a constant $C \in \omega$ such that $\xi(X_{\leq n}) \leq C\mu(X_{\leq n})$ for all n. ML-randomness implies computable randomness, but the converse does not hold in general.

2.2   Algorithmic Probability

We review some results from algorithmic probability. For details, see e.g. [2].

Let μ be an optimal c.e. semi-measure. Fix a sequence $X \in 2^{\omega}$. We are interested in the ratio

$$\mu(k|X_{<n}) = \frac{\mu(X_{<n}k)}{\mu(X_{<n})},$$

where $X_{<n} = X_1 \cdots X_{n-1}$ and $k \in \{0, 1\}$. Notice that X_0 denotes the empty string. The function $k \mapsto \mu(k|X_{<n})$ is a measure on $\{0, 1\}$ but the measure of the whole space $\{0, 1\}$ need not be 1. The ratio can be understood as the conditional probability of the n-th bit given the initial $(n-1)$ bits of X, and is called *algorithmic probability*.

One of the desirable properties of algorithmic probability is the following convergence to the true induced measure.

Theorem 2.3. *Let μ be a computable probability measure on 2^ω and ξ be an optimal c.e. semi-measure. Suppose $X \in 2^\omega$ is sampled from μ. Then,*

$$\xi(k|X_{<n}) - \mu(k|X_{<n}) \to 0 \ as \ n \to \infty \tag{1}$$

for both $k \in \{0,1\}$ and

$$\frac{\xi(X_n|X_{<n})}{\mu(X_n|X_{<n})} \to 1 \ as \ n \to \infty \tag{2}$$

with μ-probability 1.

The convergence (1) is called the *convergence in difference* by Solomonoff [11]. The convergence (2) is called the *convergence in ratio* in Li-Vitányi's book [7, Theorem 5.2.2, p. 433]. Remark the difference between on-sequence and off-sequence. The speed of the convergence is one of our interest, which has been discussed briefly in [3] but has not been established.

2.3   Distance Measures Between Probability Measures

The following notions are important in the proof of the convergence. For probability measures μ, ξ on $\{0,1\}$, we define the *squared Hellinger distance $H^2(\nu, \xi)$* by

$$H^2(\mu, \xi) = \frac{1}{2} \sum_{k \in \{0,1\}} (\sqrt{\mu(k)} - \sqrt{\xi(k)})^2 = 1 - \sum_{k \in \{0,1\}} \sqrt{\mu(k)\xi(k)}.$$

From the equalities above, $0 \le H^2(\mu, \xi) \le 1$. We also use the Kullback-Leibler divergence (or KL-divergence) of μ with respect to ξ defined by

$$D(\mu||\xi) = \sum_{k \in \{0,1\}} \mu(k) \ln \frac{\mu(k)}{\xi(k)}$$

where ln is the natural logarithm and $0 \cdot \ln \frac{0}{z} = 0$ for $z \ge 0$ and $y \ln \frac{y}{0} = \infty$ for $y > 0$. The two notions are related by the following inequality:

$$2H^2(\mu, \xi) \le D(\mu||\xi). \tag{3}$$

One can check this by direct calculation or see Hutter [2, Lemma 3.11h].

3   Convergence Along Computable Random Sequences

3.1   Convergence Results

Algorithmic probability is computably approximable or Δ_2^0 but not computable while all the functions we can implement are computable. The correct or ideal prediction may have some properties that algorithmic probability has, however,

implementing a program with one of such properties may be impossible. Thus, this does not say anything about whether our implementable prediction should have the properties. That a program does not have one of such properties does not say that the program is not general enough.

The goal of this paper is to give a framework to study the properties of computable measures or predictions. Algorithmic probability uses an optimal c.e. semi-measure while no computable measure m-dominates all computable measures. We abandon to pursue the unique correct prediction. Instead, we ask the following:

Which properties do all sufficiently general predictions have?

Notice that this statement is about a prediction that can be implemented in reality.

As a definition of the "generality" above, we use m-domination inspired by the definition of optimality. More concretely, we construct a computable measure ν such that a property P holds for all computable measures m-dominating ν. This means that P holds for all sufficiently general predictions. There are many quantifiers, and we will see that their order is important.

Suppose that, a property P is witnessed by a computable measure ν_P, that is, all computable measures ξ m-dominating ν_P have the properties P. Similarly, suppose that a property Q is witnessed by a computable measure ν_Q. Then, the property $P \wedge Q$ is witnessed by the computable measure $(\nu_P + \nu_Q)/2$. The composition of properties can be extended into computable countable sum.

Some property P may be witnessed by a measure μ executable in feasible time. If some good prediction induced from ν does not have the property P, then the prediction by $\epsilon\mu + (1-\epsilon)\nu$ for a positive rational ϵ is more general than ν in the sense above, and the computation cost may be still reasonable.

Now, we give computable versions of Theorem 2.3 as follows.

Theorem 3.1. *Let μ be a computable probability measure on 2^ω. For all computable probability measures ξ m-dominating μ and for all μ-computably random sequence $X \in 2^\omega$, we have*

$$\sum_{n=1}^{\infty} D(\mu(\cdot|X_{<n}) \,||\xi(\cdot|X_{<n})) < \infty.$$

In particular,

$$\xi(k|X_{<n}) - \mu(k|X_{<n}) \to 0 \text{ as } n \to \infty$$

for both $k \in \{0,1\}$.

Theorem 3.2. *Let μ be a computable probability measure. For all computable probability measure ξ m-dominating μ and for all μ-computably random sequence $X \in 2^\omega$, we have*

$$\frac{\xi(X_n|X_{<n})}{\mu(X_n|X_{<n})} \to 1 \text{ as } n \to \infty.$$

Notice that both Theorems 3.1 and 3.2 claim the existence of a computable measure $\nu(=\mu)$ such that for all computable measures ξ m-dominating ν have some properties, that is, all sufficiently general computable measures have some properties.

Hutter and Muchnik [3] has shown that "μ-probability 1" in Theorem 2.3 cannot be replaced by "for all μ-ML-random sequences X." The above theorem says that, for a computable probability measure, we only need μ-computable randomness.

3.2   Martingale Characterization and Convergence

We use a martingale characterization of computable randomness and a convergence theorem of martingales. The following characterization is due to [9].

Definition 3.3. *Let μ be a computable probability measure. A martingale M with respect to μ is a partial function $M :\subseteq 2^{<\omega} \to \mathbb{R}^+$ such that the following two conditions hold:*

(i) (Impossibility condition) If $M(\sigma)$ is undefined, then $\mu(\sigma) = 0$.
(ii) (Fairness condition) For all $\sigma \in 2^{<\omega}$, we have

$$M(\sigma 0)\mu(\sigma 0) + M(\sigma 1)\mu(\sigma 1) = M(\sigma)\mu(\sigma)$$

where undefined $\cdot\, 0 = 0$ and $\mathbb{R}^+$ is the set of all non-negative reals.

We say M is an almost-everywhere computable martingale *(or a.e. computable martingale) if M is a partial computable function. We say M succeeds on $X \in 2^\omega$ if $\limsup_{n\to\infty} M(X_{\leq n}) = \infty$.*

Proposition 3.4. *Let μ be a computable probability measure on 2^ω. Then, $X \in 2^\omega$ is μ-computably random if and only if $\mu(X_{\leq n}) > 0$ for all n and there is no a.e. computable martingale M which succeeds on X.*

Computable randomness with respect to the uniform measure can be characterized as the existence of the limit along the sequence for all computable martingales (see, e.g. [1, Theorem 7.1.3]) by using Doob's upcrossing argument. The same method can be applied for any computable measure.

Proposition 3.5. *Let μ be a computable probability measure on 2^ω. Then, $X \in 2^\omega$ is μ-computably random if and only if $\lim_{n\to\infty} M(X_{\leq n})$ exists for all a.e. computable martingales M.*

3.3   Proof of Theorem 3.1

Proof. Since ξ m-dominates μ, there exists a constant $C \in \omega$ such that

$$\mu(\sigma) \leq C\xi(\sigma) \tag{4}$$

for all $\sigma \in 2^{<\omega}$. Let $D(\sigma)$ be the KL-divergence of μ w.r.t. ξ at σ, that is,

$$D(\sigma) = D(\mu(\cdot|\sigma) \ || \ \xi(\cdot|\sigma)).$$

We define a function $M : \subseteq 2^{<\omega} \to \mathbb{R}^+$ by

$$M(\sigma) = \ln C - \ln \frac{\mu(\sigma)}{\xi(\sigma)} + \sum_{t=1}^{|\sigma|} D(\sigma_{<t})$$

for every $\sigma \in 2^{<\omega}$.

We claim that M is an a.e. computable martingale w.r.t. μ. The function M is non-negative because of (4) and the non-negativeness of D. For the impossibility condition of Definition 3.3, notice that, if $\mu(\sigma) > 0$, then $\xi(\sigma) > 0$ because ξ m-dominates μ, thus $M(\sigma)$ is defined. Then, the a.e. computability of M follows from the computability of μ, ξ, and D. For the fairness condition,

$$\sum_{k \in \{0,1\}} \mu(\sigma k) M(\sigma k) - \mu(\sigma) M(\sigma) = - \sum_{k \in \{0,1\}} \mu(\sigma k) \ln \frac{\mu(k|\sigma)}{\xi(k|\sigma)} + \mu(\sigma) D(\sigma) = 0.$$

Since X is μ-computably random, we have $\limsup_n M(X_{\leq n}) < \infty$. Since both $\ln C - \ln \frac{\mu(\sigma)}{\xi(\sigma)}$ and $D(\sigma)$ are always non-negative, $\sum_{n=1}^{\infty} D(X_{\leq n})$ also converges. Finally, the last claim of the theorem follows by (3). □

3.4   Proof of Theorem 3.2

Proof. Suppose that ξ is a computable measure m-dominating the measure μ. We define a function $M : 2^{<\omega} \to \mathbb{R}^+$ by

$$M(\sigma) = \frac{\xi(\sigma)}{\mu(\sigma)}.$$

Then, M is a a.e. computable martingale w.r.t. μ. Hence, $\lim_n M(X_{\leq n}) = \alpha$ exists for all μ-computably random sequences $X \in 2^\omega$.

Since ξ m-dominates μ, there exists $C \in \omega$ such that $\mu(\sigma) \leq C\xi(\sigma)$ for every $\sigma \in 2^{<\omega}$. Then,

$$M(X_{\leq n}) = \frac{\xi(X_{\leq n})}{\mu(X_{\leq n})} \geq \frac{1}{C}$$

for every n. Thus, $\alpha \geq \frac{1}{C}$.

Fix $\epsilon > 0$. Then, there exists $N \in \omega$ such that

$$\left| \frac{\xi(X_{\leq n})}{\mu(X_{\leq n})} - \alpha \right| = |M(X_{\leq n}) - \alpha| \leq \frac{\epsilon}{3C}$$

for all $n \geq N$. Thus,

$$\frac{\xi(X_n|X_{\leq n})}{\mu(X_n|X_{\leq n})} = \frac{\xi(X_{\leq n})}{\mu(X_{\leq n})} \cdot \frac{\mu(X_{<n})}{\xi(X_{<n})} \leq \frac{\alpha + \epsilon/(3C)}{\alpha - \epsilon/(3C)} = 1 + \epsilon \cdot \frac{2}{3\alpha C - \epsilon} < 1 + \epsilon$$

for all $n \geq N + 1$ if ϵ is sufficiently small. Similarly, $\frac{\xi(X_n|X_{\leq n})}{\mu(X_n|X_{\leq n})} > 1 - \epsilon$ for all $n \geq N + 1$. Since ϵ is arbitrary, the claim follows. □

4   Non-computability of the Convergence

From now on, we only consider the case that μ is the Dirac measure on a point $A \in 2^\omega$. If μ is computable, then A should be computable. Theorem 3.1 in this case can be written as follows.

Corollary 4.1. *Let $A \in 2^\omega$ be a computable sequence. There exists a computable measure ν such that*

$$\sum_n (1 - \xi(A_n|A_{<n})) < \infty \tag{5}$$

for all computable measures ξ m-dominating ν. In particular, $\xi(A_n|A_{<n}) \to 1$ as $n \to \infty$.

Proof. Let μ be the Dirac measure on the point $A \in 2^\omega$. Then, A is μ-computably random. By Theorem 3.1, we have $\sum_{n=1}^\infty \ln \frac{1}{\xi(A_n|A_{<n})} < \infty$. Finally, notice that

$$\sum_n \ln(1 - (1 - \xi(A_n|A_{<n}))) > -\infty \iff \sum_n (1 - \xi(A_n|A_{<n})) < \infty.$$

$\square$

All sufficiently general computable measures can detect the pattern of a computable sequence A while no computable measure can detect the pattern of all computable sequences. One needs to pay attention to the order or quantifiers.

Remark 4.2. *For each computable measure ξ, there exists a computable sequence A such that $\xi(A_n|A_{<n})$ does not converge to 1.*

This claim is essentially the same as a famous fact in algorithmic learning theory that the class of all computable sequences is not BC-learnable. See e.g. [12] for a survey on algorithmic learning theory. A stronger result in the context of universal induction is in [4, Lemma 5.2.4]. For the sake of self-containedness, we give a short proof here in our terminology.

Proof. Let $(\epsilon_n)_n$ be a computable sequence of positive rationals such that $\epsilon_n < \frac{1}{2}$ for all n and $\prod_n (1 + \epsilon_n) < \infty$. For each σ, at least one of $i \in \{0,1\}$ satisfies $\xi(\sigma i) < \frac{1+\epsilon_{|\sigma|}}{2}\xi(\sigma)$. This is a c.e. relation and one can compute such i from σ uniformly. By iterating this, one can compute a sequence A such that $\xi(A_{\leq n}) < \frac{1+\epsilon_n}{2}\xi(A_{<n})$ for all n. Since $\epsilon_n \to 0$, we have

$$\limsup_n \xi(A_n|A_{<n}) \leq \limsup_n \frac{1 + \epsilon_n}{2} \leq \frac{1}{2}.$$

$\square$

The following theorem says that all sufficiently general predictions $\xi(A_n|A_{<n})$ are not too close to 1; they have almost the same convergence speed (5) up to a multiplicative constant. The convergence speed $\xi(\overline{A_n}|A_{<n})$ to 0 is the slowest one among the sequences whose sum converge. Let $\bar{k} = 1 - k$ for $k \in \{0,1\}$.

Theorem 4.3. *Let $A \in 2^\omega$ be a computable sequence and $(a_n)_n$ be a computable sequence of positive rationals such that $\sum_n a_n < \infty$. Then, there exists a computable measure ν with the following property: For each computable measure ξ m-dominating ν, there exists a natural number $C \in \omega$ such that*

$$\xi(\overline{A_n}|A_{<n}) \geq \frac{a_n}{C}$$

for all n.

Notice that $(a_n)_n$ need not be monotone.

Proof. Without loss of generality, we can assume that $s = \sum_n a_n < 1$. Define a measure ν by

$$\nu = \sum_n a_n \mathbf{1}_{A_{<n}\overline{A_n}0^\omega} + (1-s)\mathbf{1}_A$$

where $\mathbf{1}_X$ is the point-mass measure on $X \in 2^\omega$.

We claim that this measure ν is computable. It suffices to show that $\nu(\sigma)$ is computable uniformly in $\sigma \in 2^{<\omega}$. If $\sigma \prec A$, then

$$\nu(\sigma) = \sum_{n \geq |\sigma|} a_n + 1 - s = 1 - \sum_{n < |\sigma|} a_n.$$

If $\sigma = A_{<k}\overline{A_k}0^i$ for some $k, i \in \omega$, then

$$\nu(\sigma) = a_k.$$

If $\sigma = A_{<k}\overline{A_k}0^i 1\tau$ for some $k, i \in \omega$ and $\tau \in 2^{<\omega}$, then

$$\nu(\sigma) = 0.$$

In any case, $\nu(\sigma)$ is computable from n.

Suppose that a computable measure ξ m-dominates ν. Then, there exists $C \in \omega$ such that $\nu(\sigma) \leq C\xi(\sigma)$ for all $\sigma \in 2^{<\omega}$. Then,

$$\xi(\overline{A_n}|A_{<n}) = 1 - \frac{\xi(A_{\leq n})}{\xi(A_{<n})} = \frac{\xi(A_{<n}\overline{A_n})}{\xi(A_{<n})} \geq \frac{\nu(A_{<n}\overline{A_n})}{C} = \frac{a_n}{C}.$$

$\square$

The rate of convergence of $\xi(\overline{A_n}|A_{<n})$ to 0 is not monotone. In fact, it cannot be bounded by a decreasing computable function converging to 0.

Theorem 4.4. *Let $A \in 2^\omega$ be a computable sequence. Then, there exists a computable measure ν such that no decreasing computable sequence $(b_n)_n$ converging to 0 m-dominates $\xi(\overline{A_n}|A_{<n})$ for all computable measures ξ m-dominating ν.*

Proof. Let $U :\subseteq 2^{<\omega} \to 2^{<\omega}$ be a universal prefix-free machine. By the usual convention, $U(\sigma)[s] \uparrow$ for each $s < |\sigma|$. For all n, let

$$a_n = \sum_\sigma \{2^{-|\sigma|} : U(\sigma) \downarrow \text{ at stage } n\}.$$

Note that $(a_n)_n$ is a computable sequence because the possible σ should satisfy $|\sigma| \leq n$ by the convention. Furthermore,

$$\sum_n a_n = \sum_{\sigma \in \text{dom}(U)} 2^{-|\sigma|} < 1.$$

Then, by Theorem 4.3, there exists a computable measure ν such that $\nu(\overline{A_n}|A_{<n})$ m-dominates $(a_n)_n$.

Suppose that there exists a decreasing computable $(b_n)_n$ such that $b_n \to 0$ as $n \to \infty$ and $(b_n)_n$ m-dominates $\nu(\overline{A_n}|A_{<n})$. Then, $(b_n)_n$ also m-dominates $(a_n)_n$, and let $C \in \omega$ such that $a_n \leq 2^C b_n$ for all n.

For each σ, search the least $n \in \omega$ such that $b_n < 2^{-|\sigma|-C}$. If $U(\sigma)[n] \uparrow$ and $U(\sigma)[s] \downarrow$ for some $s > n$, then $a_s \geq 2^{-|\sigma|}$ by the definition of $(a_n)_n$, and

$$a_s \leq 2^C b_s \leq 2^C b_n < 2^{-|\sigma|},$$

which is a contradiction. Thus, $U(\sigma) \downarrow$ if and only if $U(\sigma)[n] \downarrow$. Since n is computable from σ uniformly, this means that the halting problem is computable, which is a contradiction. Hence, such $(b_n)_n$ does not exist. $\square$

It may be interesting to compare the above result to Laplace's answer to the sunrise problem. The answer $\frac{n}{n+1}$ is slightly slower than (5). Theorem 4.4 means that all sufficiently general prediction violates Nicod's criterion as the usual Solomonoff induction does [5].

References

1. Downey, R.G., Hirschfeldt, D.R.: Algorithmic randomness and complexity. Theory and Applications of Computability. Springer, New York (2010). https://doi.org/10.1007/978-0-387-68441-3
2. Hutter, M.: Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability. Springer, Heidelberg (2005). https://doi.org/10.1007/b138233
3. Hutter, M., Muchnik, A.: On semimeasures predicting Martin-Löf random sequences. Theor. Comput. Sci. **382**, 247–261 (2007)
4. Legg, S.: Machine super intelligence. Ph.D. thesis, Università della Svizzera italiana (2008)
5. Leike, J., Hutter, M.: Solomonoff induction violates Nicod's criterion. In: Chaudhuri, K., Gentile, C., Zilles, S. (eds.) ALT 2015. LNCS (LNAI), vol. 9355, pp. 349–363. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24486-0_23
6. Levin, L.A.: On the notion of a random sequence. Sov. Math. Dokl. **14**, 1413–1416 (1973)

7. Li, M., Vitányi, P.: An introduction to Kolmogorov complexity and its applications. Graduate Texts in Computer Science, 3rd edn. Springer, New York (2009). https://doi.org/10.1007/978-0-387-49820-1

8. Nies, A.: Computability and Randomness. Oxford Logic Guides, vol. 51. Oxford University Press, Oxford (2009). https://doi.org/10.1093/acprof:oso/9780199230761.001.0001

9. Rute, J.: Computable randomness and betting for computable probability spaces. Math. Log. Q. **62**(4–5), 335–366 (2016)

10. Soare, R.I.: Turing Computability. Theory and Applications of Computability. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-642-31933-4

11. Solomonoff, R.J.: Complexity-based induction systems: comparisons and convergence theorems. IEEE Trans. Inf. Theor. **IT-24**, 422–432 (1978)

12. Zeugmann, T., Zilles, S.: Learning recursive functions: a survey. Theor. Comput. Sci. **397**, 4–56 (2008)