On Martin-Löf (Non-)Convergence of Solomonoff's Universal Mixture

Tor Lattimore[a,*], Marcus Hutter[a]

[a] *RSCS, Australian National University*

April 14, 2014

Abstract

We study the convergence of Solomonoff's universal mixture on individual Martin-Löf random sequences. A new result is presented extending the work of Hutter and Muchnik (2004) by showing that there does not exist a universal mixture that converges on all Martin-Löf random sequences. We show that this is not an artifact of the fact that the universal mixture is not a proper measure and that the normalised universal mixture also fails to converge on all Martin-Löf random sequences.

Keywords: Solomonoff induction; Kolmogorov complexity; theory of computation.

1. Introduction

Sequence prediction is the task of predicting symbol α_n having seen $\alpha_{1:n-1} = \alpha_1 \cdots \alpha_{n-1}$. Solomonoff approached this problem by taking a Bayesian mixture over all lower semicomputable semimeasures where complex semimeasures were assigned lower prior probability than simple ones.[1] He then showed that, with probability one, the predictive mixture converges (fast) to the truth for any computable measure [9]. Solomonoff induction arguably solves the sequence prediction problem and has numerous attractive properties, both technical [9, 2, 5] and philosophical [8]. There is, however, some hidden unpleasantness, which we explore in this paper.

Martin-Löf randomness is the usual characterisation of the randomness of individual sequences [6]. A sequence is Martin-Löf random if it passes all effective tests, such as the laws of large numbers and the iterated logarithm. Intuitively, a sequence is Martin-Löf random with respect to measure μ if it satisfies all the properties one would expect of an infinite sequence sampled from μ. It

*Principal corresponding author

Email addresses: `tor.lattimore@anu.edu.au` (Tor Lattimore), `marcus.hutter@anu.edu.au` (Marcus Hutter)

[1] Actually, Solomonoff mixed over proper measures. The use of semimeasures was introduced later by Levin to ensure that the mixture itself was lower semicomputable [14].

has previously been conjectured that the set of Martin-Löf random sequences is precisely, or contained within, the set on which the Bayesian mixture converges.

This question has seen a number of attempts with a partial negative solution and a more detailed history of the problem by Hutter and Muchnik [3]. They showed that there exists a universal lower semicomputable semimeasure M and Martin-Löf random sequence α (with respect to the Lebesgue measure λ) for which $M(\alpha_n|\alpha_{<n}) \not\to \lambda(\alpha_n|\alpha_{<n})$. The α used in their proof is computable from the halting problem, which presumably inspired the work in [7] where it is shown that if α is 2-random, then every universal lower semicomputable semimeasure converges on α. It is worth remarking that there are exist semimeasures that do converge on all Martin-Löf random sequences, some of which are even lower semicomputable. Unfortunately, however, they are not universal and may not enjoy the same fast convergence rates in expectation as universal measures do. For the construction and a detailed discussion, see [4, §5].

While Hutter and Muchnik showed that there exists a universal lower semicomputable semimeasure and Martin-Löf random sequence on which it fails to converge, the question of whether or not this failure occurs for all such semimeasures has remained open. We prove that for every universal lower semicomputable Bayesian mixture there exists a Martin-Löf random sequence on which it fails to converge. This result is interesting for a few reasons. The choice of universal mixture is akin to choosing an optimal universal Turing machine when computing Kolmogorov complexity. In both cases, asymptotic results are rarely dependent on this choice and so it is useful to confirm this trend here. On the other hand, if the result had been positive then the existence of a universal mixture that did converge on all Martin-Löf random strings would be a nice property that might justify the choice of one universal mixture over another.

The universal mixture is not a proper measure in the sense that the sum of conditional probabilities $M(0|x) + M(1|x) < 1$ for all x. For this reason it is common to use a normalised version M_{norm} where normalisation is chosen to preserve the ratio $M_{\text{norm}}(x0)/M_{\text{norm}}(x1) = M(x0)/M(x1)$. We show that the situation is not improved by normalisation and that M_{norm} also fails to converge to the Lebesgue measure on some Martin-Löf random sequences.

Our paper is structured as follows. We present the required notation and some basic results in algorithmic information theory (Section 2). We then present Solomonoff's original theorem showing that the universal mixture converges to the truth with probability one (Section 3). The main theorems are then presented of which Theorem 6 is the most important stating for any universal mixture M that there exists a Martin-Löf random sequence α such that the predictive distribution $M(\alpha_n|\alpha_{<n})$ does not converge to $\frac{1}{2}$ and actually is bounded away from $\frac{1}{2}$ for a non-zero fraction of the time (Section 4). We then show that this is also true of the normalised version of the universal mixture (Section 5) and that there exists an infinite sequence that is not Martin-Löf random, but on which all universal mixtures converge to $\frac{1}{2}$ (Section 6). We conclude in Section 7.

2. Notation

Overviews of algorithmic information theory can be found in [5, 1]. A table of notation may be found in Appendix B.

General. The natural, rational and real numbers are denoted by $\mathbb{N}$, $\mathbb{Q}$ and $\mathbb{R}$. Logarithms are taken with base 2. A real $\theta \in (0,1)$ has entropy $H(\theta) := -\theta \log\theta - (1-\theta)\log(1-\theta)$. The indicator function is $[\![expr]\!]$, which takes value 1 if *expr* is true and 0 otherwise. For sets A and B we write $A - B$ for their difference and $|A|$ for the size of A and $A^c = \mathbb{N} - A$ for the complement of A. The empty set is denoted by $\emptyset$. If $A \subseteq \mathbb{N}$ and $n \in \mathbb{N}$, then $A[n] := \{a \in A : a \leq n\}$. We use $\vee$ and $\wedge$ for logical or/and respectively.

Natural density. Let $A \subseteq B \subseteq \mathbb{N}$. Then the (upper) natural density of $A \subseteq B$ are

$$d(A, B) := \lim_{n\to\infty} \frac{|A[n]|}{|B[n]|} \qquad\qquad \bar{d}(A, B) := \limsup_{n\to\infty} \frac{|A[n]|}{|B[n]|}$$

where the latter quantity is useful in the case when the former does not exist. If $B = \mathbb{N}$, then we abbreviate $d(A) \equiv d(A, \mathbb{N})$ and $\bar{d}(A) \equiv \bar{d}(A, \mathbb{N})$.

Strings. A finite binary string x is a finite sequence $x_1 x_2 x_3 \cdots x_n$ with $x_i \in \mathcal{B} := \{0, 1\}$. Its length is $\ell(x)$. An infinite binary string ω is an infinite sequence $\omega_1\omega_2\omega_3\cdots$. The empty string of length zero is denoted by ϵ (distinct from $\varepsilon > 0 \in \mathbb{R}$). The sets $\mathcal{B}^n$, $\mathcal{B}^*$ and $\mathcal{B}^\infty$ are the sets of all strings of length n, all finite strings and all infinite strings respectively. Substrings of $x \in \mathcal{B}^* \cup \mathcal{B}^\infty$ are denoted by $x_{s:t} := x_s x_{s+1} \cdots x_{t-1} x_t$ where $s, t \in \mathbb{N}$ and $s \leq t$. If $s > t$, then $x_{s:t} := \epsilon$. A useful shorthand is $x_{<t} := x_{1:t-1}$. Let $x, y \in \mathcal{B}^*$, then $\#x(y)$ is the number of (possibly overlapping and wrapping around) occurrences of x in y and xy is their concatenation. For example, $\#010(1010) = 2$ (because we count the wrap around match when starting at the last bit). If $\ell(y) \geq \ell(x)$ and $x_{1:\ell(x)} = y_{1:\ell(x)}$, then we write $x \sqsubseteq y$ and say x is a prefix of y. Otherwise we write $x \not\sqsubseteq y$. A string $\omega \in \mathcal{B}^\infty$ is normal if $\forall x \in \mathcal{B}^*, \lim_{n\to\infty} \#x(\omega_{1:n})/n = 2^{-\ell(x)}$.

Measures and semimeasures. A semimeasure is a function $\mu : \mathcal{B}^* \to [0, 1]$ satisfying $\mu(\epsilon) \leq 1$ and $\mu(x) \geq \mu(x0) + \mu(x1)$ for all $x \in \mathcal{B}^*$. It is a measure if both inequalities are replaced by equalities. A function $\mu : \mathcal{B}^* \to \mathbb{R}$ is lower semicomputable if the set $\{(x, r) : r < \mu(x), r \in \mathbb{Q}, x \in \mathcal{B}^*\}$ is recursively enumerable. In this case there exists a recursive sequence $\mu_1, \mu_2, \cdots$ of computable functions approximating μ from below. For $b \in \mathcal{B}$ and $x \in \mathcal{B}^*$, $\mu(b|x) := \mu(xb)/\mu(x)$ is the μ-probability that x is followed by b. The Lebesgue measure is $\lambda(x) := 2^{-\ell(x)}$.

Complexity. A Turing machine T is a recursively enumerable set of pairs of binary strings $T := \{(p^1, x^1), (p^2, x^2), \cdots\}$ where the program p^k outputs x^k. It is a prefix machine if the set of programs is prefix free, $p^k \not\sqsubseteq p^j$ for all $j \neq k$. T is a monotone machine if $p^k \sqsubseteq p^j \implies x^k \sqsubseteq x^j \vee x^j \sqsubseteq x^k$. For prefix machine

T the prefix complexity with respect to T is a function $K_T : \mathcal{B}^* \to \mathbb{N}$ defined by

$$K_T(x) := \min_p \{\ell(p) : (p, x) \in T\}$$

If T is a monotone machine, then the monotone complexity with respect to T is defined by

$$Km_T(x) := \min_p \{\ell(p) : (p, y) \in T \wedge x \sqsubseteq y\}$$

There exists an additively optimal prefix machine U such that for all prefix machines T there exists a constant c_T with $K_U(x) < K_T(x) + c_T$. In identical fashion there exists an additively optimal monotone machine. As is usual in algorithmic information theory, we fix a pair of additively optimal prefix and monotone machines and write $K(x) := K_U(x)$ and $Km(x) := Km_U(x)$. The choice of reference machine is irrelevant for this work.

A lower semicomputable semimeasure M is universal if for every lower semicomputable semimeasure μ there exists a constant $c_\mu > 0$ such that $\forall x, M(x) > c_\mu \mu(x)$. Zvonkin and Levin [14] showed that the set of all lower semicomputable semimeasures is recursively enumerable (possibly with repetition). Let $\nu_1, \nu_2, \cdots$ be such an enumeration and $w : \mathbb{N} \to [0, 1]$ be a lower semicomputable sequence satisfying $\sum_{i \in \mathbb{N}} w_i \leq 1$, which we view as a prior on the lower semicomputable semimeasures. Then the universal mixture is defined by

$$M(x) := \sum_{i \in \mathbb{N}} w_i \nu_i(x). \tag{1}$$

There are, of course, many possible enumerations and priors, and hence there are many universal mixtures. This paper aims to prove certain non-convergence results about all universal mixtures, regardless of the choice of prior. Defining $w_i(x) := w_i \nu_i(x)/M(x)$ and substituting into Eq. 1 leads to

$$M(b|x) = \sum_{i \in \mathbb{N}} w_i(x) \nu_i(b|x). \tag{2}$$

There exist universal lower semicomputable semimeasures that are not representable as universal mixtures, but we do not consider these here [13].

Normalised mixture. It is well known that no universal mixture is a proper measure. In fact

$$(\forall x \in \mathcal{B}^*) \qquad M(0|x) + M(1|x) < 1 \qquad [5, \text{Ex. } 4.5.1].$$

For this reason the universal mixture is often normalised by defining

$$(\forall b \in \mathcal{B}, x \in \mathcal{B}^*) \qquad M_{\text{norm}}(b|x) := \frac{M(b|x)}{M(0|x) + M(1|x)}.$$

Then the normalised measure can be defined $M_{\text{norm}} : \mathcal{B}^* \to [0,1]$ by

$$M_{\text{norm}}(\epsilon) := 1 \qquad M_{\text{norm}}(x) := M(x) \prod_{n=1}^{\ell(x)} \frac{1}{M(0|x_{<n}) + M(1|x_{<n})},$$

which satisfies $M_{\text{norm}}(0|x) + M_{\text{norm}}(1|x) = 1$ and $M_{\text{norm}}(\epsilon) = 1$. This is not the only possible normalisation, but is standard in the literature [5]. Of particular importance for this paper are the facts that $M_{\text{norm}}(x) > M(x)$ and $M_{\text{norm}}(b|x) > M(b|x)$, which follow immediately from the definition of M_{norm}.

Martin-Löf randomness. Let μ be a computable measure and M a universal lower semicomputable semimeasure. An infinite binary string ω is μ-Martin-Löf random (μ-random) if and only if there exists a $c > 0$ such that

$$(\forall n \in \mathbb{N}) \qquad \mu(\omega_{<n})/M(\omega_{<n}) > c. \tag{3}$$

Observe that the definition does not depend on the choice of universal lower semicomputable semimeasure since for any two universal lower semicomputable semimeasures M and M' there exists a constant $c > 0$ such that $cM'(x) > M(x) > M'(x)/c$, $\forall x$ [5]. We write $\mathcal{R}_\mu \subset \mathcal{B}^\infty$ for the set of μ-random strings.

Lemma 1. *The following hold:*

(i) *If $\omega \in \mathcal{B}^\infty$ is λ-random, then it is normal.*
(ii) *If $x \in \mathcal{B}^*$ with $\ell(x) = n$ and $\theta := \#1(x)/n$, then $Km(x) < nH(\theta) + \frac{1}{2}\log n + c$ for some $c > 0$ independent of x and n.*
(iii) *Let $A, B \subseteq \mathbb{N}$ and $\phi_n := [\![n \in A]\!]$. If $d(A) = 0$ and $\bar{d}(B) > 0$, then*
   (a) $\bar{d}(B - A) > 0$.
   (b) $\lim_{n\to\infty} Km(\phi_{1:n})/n = 0$.

Proof. Part (i) is well known [5, §2.6]. For part (ii) we use the KT-estimator, which is defined by

$$\mu(x) := \int_0^1 \frac{1}{\pi\sqrt{(1-\theta)\theta}} \theta^{\#1(x)}(1-\theta)^{\#0(x)} d\theta.$$

Because μ is a measure and is finitely computable using a recursive formula [12], we can apply Theorem 4.5.4 in [5] to show that there exists a constant $c_\mu > 0$ such that

$$\begin{aligned} Km(x) &< -\log\mu(x) + c_\mu \le \frac{1}{2}\log n + 1 + \log\theta^{\#1(x)}(1-\theta)^{\#0(x)} + c_\mu \\ &= \frac{1}{2}\log n + 1 + nH(\theta) + c_\mu, \end{aligned}$$

where we used the redundancy bound for the KT-estimator [12] and the definition of $H(\theta)$. Part $(iii)(a)$ is immediate from the definition of the natural density. For (b), let $\theta_n := \#1(\phi_{1:n})/n$ and note that $d(A) = 0$ implies that $\lim_{n\to\infty}\theta_n = 0$ and so $\lim_{n\to\infty} H(\theta_n) = 0$. Finally apply part (ii) to complete the proof. $\square$

3. Almost Sure Convergence

Before Martin-Löf convergence is considered we present a version of the celebrated theorem of Solomonoff with which we will contrast our results [10].

Theorem 2 (Solomonoff, 1978)**.** *If M is a universal lower semicomputable semimeasure and α is sampled from computable measure μ, then with μ-probability 1 (w.μ.p.1.)*

$$\lim_{n\to\infty} \sum_{b\in\mathcal{B}} \left(M(b|\alpha_{<n}) - \mu(b|\alpha_{<n})\right)^2 = 0.$$

We say that M converges on-sequence to λ on α if

$$\lim_{n\to\infty} M(\alpha_n|\alpha_{<n}) = \frac{1}{2}. \tag{4}$$

It converges off-sequence if

$$\lim_{n\to\infty} \sum_{b\in\mathcal{B}} M(b|\alpha_{<n}) = \frac{1}{2}. \tag{5}$$

It is trivial that (5) implies (4). Unfortunately, Theorem 2 only ensures convergence to α with λ-probability 1 while we are primarily interested in convergence on individual sequences. An easy consequence of Theorem 2 is that all universal lower semicomputable semimeasures converge to a proper measure with μ.p.1. for all computable measures μ.

Corollary 3. *If M is a universal lower semicomputable semimeasure and α is sampled from computable measure μ, then*

$$\lim_{n\to\infty} M(1|\alpha_{<n}) + M(0|\alpha_{<n}) = 1 \quad \textit{with } \mu\textit{-probability } 1.$$

4. Martin-Löf Convergence

Recall that λ is the Lebesgue measure defined by $\lambda(x) := 2^{-\ell(x)}$. We now ask whether there exists a universal mixture such that $\lim_{n\to\infty} M(\alpha_n|\alpha_{<n}) = \frac{1}{2}$ for all λ-random α. Two new theorems are presented, the first is subsumed by the second, but admits an easy proof and serves as a nice warm-up.

Theorem 4. *Let M be a universal mixture. Then there exists a λ-random α such that*

$$\lim_{n\to\infty} \sum_{b\in\mathcal{B}} \left(M(b|\alpha_{<n}) - \frac{1}{2}\right)^2 \neq 0.$$

Proof. We use the same λ-random string α as Hutter and Muchnik [3], which is defined inductively by $\alpha_n := [\![M(\alpha_{<n}0) > 2^{-n}]\!]$. Define $\nu : \mathcal{B}^* \to [0, 1]$ by

$$\nu(x) := M(x)[\![\forall n \leq \ell(x) : x_n = 0 \vee M(x_{<n}0) > 2^{-n}]\!].$$

It is straightforward to check that ν is both lower semicomputable and a semimeasure. Therefore there exists a $j \in \mathbb{N}$ such that $\nu = \nu_j$ in the enumeration of all lower semicomputable semimeasures used by M. Now if $\alpha_n = 1$, then $M(\alpha_{<n}0) > 2^{-n}$ by the definition of α. Therefore $\alpha_n = 0 \vee M(\alpha_{<n}0) > 2^{-n}$ is true for all n and so by the definition of ν we have that $\nu(\alpha_{1:n}) = M(\alpha_{1:n})$ for all n. Therefore $w_j(\alpha_{<n}) := w_j\nu(\alpha_{<n})/M(\alpha_{<n}) = w_j$. Furthermore,

$$\alpha_n = 0 \implies M(\alpha_{<n}0) \leq 2^{-n} \implies \nu(\alpha_{<n}1) = 0 \implies \nu(1|\alpha_{<n}) = 0,$$

where we used the definitions of α, ν and the conditional probability respectively. Therefore if $\alpha_n = 0$, then

$$\begin{aligned} M(0|\alpha_{<n}) + M(1|\alpha_{<n}) &\stackrel{(a)}{=} \sum\nolimits_{i\in\mathbb{N}} w_i(\alpha_{<n})\left(\nu_i(0|\alpha_{<n}) + \nu_i(1|\alpha_{<n})\right) \\ &\stackrel{(b)}{\leq} \sum\nolimits_{i\in\mathbb{N}} w_i(\alpha_{<n}) - w_j(\alpha_{<n})(1 - M(0|\alpha_{<n})) \\ &\stackrel{(c)}{=} 1 - w_j(1 - M(0|\alpha_{<n})) \\ &\stackrel{(d)}{\leq} 1 - w_j M(1|\alpha_{<n}) \end{aligned} \tag{6}$$

where (a) follows directly from Eq. 2. (b) follows by extracting $w_j(\alpha_{<n})$ from the sum and using the facts that $\nu_j(0|\alpha_{<n}) + \nu_j(1|\alpha_{<n}) = M(0|\alpha_{<n})$ and $\nu_i(0|\alpha_{<n}) + \nu_i(1|\alpha_{<n}) \leq 1$ for all i. (c) follows from the facts that $\sum_{i\in\mathbb{N}} w_i(x) = 1$ and $w_j(\alpha_{<n}) = w_j$. For (d) we note that M is a semimeasure, which implies that $1 - M(0|\alpha_{<n}) \geq M(1|\alpha_{<n})$. Because α is λ-random, it must contain infinitely many zeros by Lemma 1(i) and the definition of a normal string. Let n_i be the position of the ith 0 in α and $k \in \mathbb{N}$ be such that $\nu_k = \lambda$. Therefore there exists a $c > 0$ such that

$$M(1|\alpha_{<n_i}) \stackrel{(a)}{=} \sum_{i\in\mathbb{N}} w_i(\alpha_{<n})\nu(1|\alpha_{<n_i}) \stackrel{(b)}{\geq} w_k(\alpha_{<n})\lambda(1|\alpha_{<n_i}) \stackrel{(c)}{>} c,$$

where (a) is the same as Eq. 2 and (b) follows by extracting the contribution of the Lebesgue measure λ. (c) follows by recalling that $\lambda(1|\alpha_{<n_i}) = 1/2$ and the fact that α is λ-random combined with Eq. 3. Then by Eq. 6,

$$\liminf_{i\to\infty} M(0|\alpha_{<n_i}) + M(1|\alpha_{<n_i}) \leq 1 - w_j c < 1.$$

Therefore $\lim_{n\to\infty} M(0|\alpha_{<n}) + M(1|\alpha_{<n}) \neq 1$ and so

$$\lim_{n\to\infty} \sum_{b\in\mathcal{B}} (M(b|\alpha_{<n}) - 1/2)^2 \neq 0$$

as required. □

The proof of Theorem 4 demonstrates the existence of random sequences on which M fails to converge to a proper measure. This is interesting when compared to Corollary 3, which shows that M converges to a measure with μ-probability one with respect to any computable measure μ.

Proposition 5. *There exists an α that is λ-random where*

$$\lim_{n\to\infty} M(0|\alpha_{<n}) + M(1|\alpha_{<n}) \neq 1.$$

We now present the on-sequence version of Theorem 4, which uses the same α for a counter-example, but turns out to be significantly harder to prove.

Theorem 6. *Let M be a universal mixture. Then there exists a λ-random α and $\varepsilon > 0$ such that*

$$\bar{d}\left(\left\{n : \left|M(\alpha_n|\alpha_{<n}) - \frac{1}{2}\right| > \varepsilon\right\}\right) > 0.$$

Corollary 7. *Let M be a universal mixture. Then there exists a λ-random α such that $\lim_{n\to\infty} M(\alpha_n|\alpha_{<n}) \neq \frac{1}{2}$.*

Initially we follow the proof in [3] by constructing a lower semicomputable semimeasure ν that dominates M on α infinitely often, but where $\nu(0|\alpha_{<n}) = 1$ if $\alpha_n = 0$. The semimeasure defined below is identical to that given in [3].

Definition 8. Let $M_t : \mathcal{B}^* \to \mathbb{Q}$ be a sequence of computable functions approximating M from below and define $\alpha^t \in \mathcal{B}^\infty$ similarly to α by $\alpha^t_n := [\![M_t(\alpha^t_{<n}0) > 2^{-n}]\!]$. Now define $\nu_t : \mathcal{B}^* \to [0,1]$ by

$$\nu_t(x) := \begin{cases} 2^{-t} & \text{if } \ell(x) = t \wedge x < \alpha^t_{1:t} \\ \nu_t(x0) + \nu_t(x1) & \text{if } \ell(x) < t \\ 0 & \text{otherwise,} \end{cases}$$

where $x < \alpha^t_{1:t}$ is decided by lexicographical order.

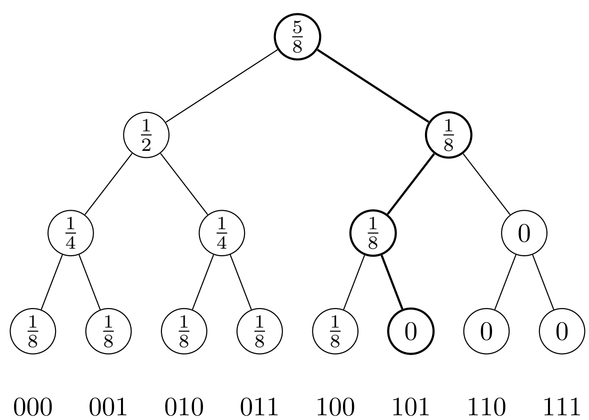

Figure 1: ν_3 if $\alpha^3_{1:3} = 101$

The following lemma is due to Hutter and Muchnik [3] and we make extensive use of it here also.

Lemma 9 (Hutter & Muchnik)**.**

(i) $\lim_{t\to\infty} \alpha^t = \alpha$.
(ii) $\nu := \lim_{t\to\infty} \nu_t$ *exists and is a lower semicomputable semimeasure.*
(iii) $\nu(x) = \nu(x0) + \nu(x1)$ *for all* x.
(iv) $\nu(\alpha_{<n}) < 2^{-n+1}$ *for all* $n \in \mathbb{N}$.
(v) *If* $\alpha_n = 1$*, then* $\nu(\alpha_{<n}0) = 2^{-n}$.
(vi) *If* $\alpha_n = 0$*, then* $\nu(\alpha_{<n}1) = 0$.

The proof of all results are found in [3, Proof of Thm.6]. Before proving the main result we discuss some useful properties of ν and briefly summarise the proof in [3]. First, if $\alpha_n = 0$, then by part (vi) we have $\nu(\alpha_{<n}1) = 0$ and so by part (iii) we obtain $\nu(\alpha_{<n}) = \nu(\alpha_{<n}0)$ and so $\nu(0|\alpha_{<n}) = 1$. Already something is feeling fishy. We have constructed a lower semicomputable semimeasure that correctly predicts all zeros of a random sequence. Additionally, if $\alpha_{n+1} = 1$, then $\nu(\alpha_{<n}) \geq \nu(\alpha_{1:n}0) = 2^{-n-1}$. Now define a new universal mixture by $M' = (1-\gamma)M + \gamma\nu$ with γ chosen sufficiently close to 1. If $\alpha_{n:n+1} = 01$, then convergence of M' is poisoned by the non-converging ν. Since α is normal there are infinitely many n for which $\alpha_{n:n+1} = 01$, which completes the result that M' does not converge to λ on α. But M' is a very specific universal mixture while Theorem 6 demands that we prove non-convergence for all universal mixtures. We start by partitioning α and proving additional properties of ν.

Define a sequences of sets $\{N_k\}_{k=0}^\infty$ by $N_k := \left\{n : \alpha_{n:n+k} = 1^k 0\right\}$ as illustrated in Figure 2. Observe that $\{N_k\}$ are disjoint and cover $\mathbb{N}$, $\mathbb{N} = \bigcup_{k=0}^\infty N_k$. Later it will be important to know the density of N_k, which is easily computed by exploiting the fact that α is λ-random and therefore (by Lemma 1) normal. Since each $n \in N_k$ is the start of a sub-sequence $1^k 0$ and by normality the total proportion of the sequence $1^k 0$ in α must be 2^{-k-1} as is demonstrated formally in the following lemma.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
α	1	0	0	1	1	0	1	0	1	1	1	0	1	1	0
$N_0 =$		2	3			6		8				12			15
$N_1 =$	1				5		7				11			14	
$N_2 =$				4						10			13		
$N_3 =$									9						

$N_0 = \{2, 3, 6, 8, 12, 15\}$, $N_1 = \{1, 5, 7, 11, 14\}$
$N_2 = \{4, 10, 13\}$ and $N_3 = \{9\}$.

Figure 2: The sets N_k

Lemma 10. *If* $k \in \mathbb{N}$*, then* $d(N_k) = 2^{-k-1}$ *and* $d(\bigcup_{k=\kappa+1}^\infty N_k) = 2^{-\kappa-1}$

Proof. α is λ-random, and therefore normal. Then

$$2^{-k-1} = \lim_{n\to\infty} \frac{\#1^k 0(\alpha_{1:n})}{n} = \lim_{n\to\infty} \frac{N_k[n-k]}{n} = \lim_{n\to\infty} \frac{N_k[n]}{n} = d(N_k).$$

For the second part we make use of Lemma 17 in the Appendix.

$$d\left(\bigcup_{k=\kappa+1}^{\infty} N_k\right) \underset{\substack{\uparrow \\ \text{Lemma 17(ii)}}}{=} 1 - d\left(\bigcup_{k=1}^{\kappa} N_k\right) \underset{\substack{\uparrow \\ \text{Lemma 17(ii)}}}{=} 1 - \sum_{k=0}^{\kappa} d(N_k) \underset{\substack{\uparrow \\ d(N_k) = 2^{-k-1}}}{=} 2^{-\kappa-1}$$

as required. ☐

Lemma 11. *The following hold:*

i *If* $\alpha_n = 1$, *then* $\nu(1|\alpha_{<n}) < \frac{1}{2}$.
ii *If* $k \geq 1$ *and* $n \in N_k$, *then*

$$\nu(1|\alpha_{<n}) < \frac{2^k - 1}{2^{k+1} - 1}.$$

Proof. We use the properties of ν and the fact that $f(x) = x/(a+x)$ is monotone increasing for positive constants a (Lemma 15).

$$\nu(1|\alpha_{<n}) \overset{\substack{\text{Definition of cond. prob.} \\ \downarrow}}{=} \frac{\nu(\alpha_{<n}1)}{\nu(\alpha_{<n})} \underset{\substack{\uparrow \\ \text{Lemma 9(iii)}}}{=} \frac{\nu(\alpha_{<n}1)}{\nu(\alpha_{<n}0) + \nu(\alpha_{<n}1)} \underset{\substack{\uparrow \\ \text{Lemma 9(v)}}}{=} \frac{\nu(\alpha_{<n}1)}{2^{-n} + \nu(\alpha_{<n}1)}. \qquad (\star)$$

Part (i) now follows from Lemma 15 and Lemma 9(iv). For part (ii) we start by bounding $\nu(\alpha_{<n}1)$.

$$\nu(\alpha_{<n}1) \overset{\substack{\text{Iterate } \nu(x1) = \nu(x10) + \nu(x11) \\ \downarrow}}{=} \sum_{i=1}^{k} \nu(\alpha_{<n}1^i0) + \nu(\alpha_{<n}1^{k+1}) \overset{(a)}{<} \sum_{i=1}^{k} 2^{-n-i} \overset{\substack{\text{Geometric series} \\ \downarrow}}{=} (1 - 2^{-k})2^{-n}.$$

(a) follows from the facts that $\nu(\alpha_{<n}1^{k+1}) = 0$ and $\nu(\alpha_{<n}1^i0) \leq 2^{-n-i}$ for $i < k$ ($n \in N_k$ and Lemma 9(v)) and strict inequality for $i = k$ (Lemma 9(iv)). Substituting into ($\star$) and applying Lemma 15 in the Appendix leads to

$$\nu(1|\alpha_{<n}) < \frac{(1 - 2^{-k})2^{-n}}{2^{-n} + (1 - 2^{-k})2^{-n}} = \frac{2^k - 1}{2^{k+1} - 1}$$

as required. ☐

Auxiliary mixture. We now exploit the fact that M is a universal mixture. Since ν is a lower semicomputable semimeasure, there exists an i such that $\nu_i = \nu$ where ν_i is the ith element in the enumeration of lower semicomputable semimeasures used to define M in Eq. 1. Then define $\tilde{w} : \mathbb{N} \to [0, 1]$ and auxiliary mixture $\tilde{M} : \mathcal{B}^* \to [0, 1]$ by

$$\tilde{w}_j := w_j - [\![j = i]\!]w_j/2 \qquad \tilde{M}(x) := \sum_{i \in \mathbb{N}} \tilde{w}_i \nu_i(x) = M(x) - \frac{w_i}{2}\nu(x).$$

Now $\tilde{M}$ is lower semicomputable To see this we only need to check that $\tilde{w}$ is lower semicomputable, which is clear since i is a natural number and so computable.

We now show that if M converges on α, then $\tilde{M}$ overestimates the probability of sequences of 1's.

Lemma 12. *Let $D \subseteq \mathbb{N}$. If $d(D) = 1$ and $\lim_{n\in D} M(\alpha_n|\alpha_{<n}) = \frac{1}{2}$, then there exists a $\gamma \in (0,1)$ and sequence of sets $C_k \subseteq N_k$ such that $d(N_k - C_k) = 0$ and*

$$(\forall n \in C_k) \qquad \tilde{M}(1^k|\alpha_{<n}) \geq \frac{1}{\gamma}\left(\frac{1}{2}\right)^k.$$

Proof. Let $D_a := \{n : n + a \in D\}$ and $D_{\leq k} := \bigcap_{a=0}^k D_a$. Now $d(D) = 1$ and D_k is a translation of D. Therefore $d(D_k) = 1$ and by Lemma 17(iii,vi) $d(D_{\leq k}) = 1$ and $d(D_{\leq k} \cap N_k, N_k) = 1$. Define $\varepsilon_n := 2M(1|\alpha_{<n})$, which by the assumption that M converges to λ on α satisfies $\lim_{n\in D} \varepsilon_n = 1$. Suppose $\alpha_n = 1$, then

$$\begin{aligned}
\tilde{M}(1|\alpha_{<n}) &\overset{(a)}{=} \frac{M(\alpha_{<n}1) - w_j\nu(\alpha_{<n}1)/2}{\tilde{M}(\alpha_{<n})} \\
&\overset{(b)}{=} \frac{\varepsilon_n M(\alpha_{<n})/2 - w_j\nu(\alpha_{<n}1)/2}{\tilde{M}(\alpha_{<n})} \\
&\overset{(c)}{=} \frac{\varepsilon_n M(\alpha_{<n})/2 - w_j\nu(\alpha_{<n})/4}{\tilde{M}(\alpha_{<n})} + w_j\frac{\nu(\alpha_{<n})/4 - \nu(\alpha_{<n}1)/2}{\tilde{M}(\alpha_{<n})} \\
&\overset{(d)}{=} \frac{1}{2} + \frac{(\varepsilon_n - 1)M(\alpha_{<n})}{2\tilde{M}(\alpha_{<n})} + w_j\frac{\nu(\alpha_{<n})/4 - \nu(\alpha_{<n}1)/2}{\tilde{M}(\alpha_{<n})} \\
&\overset{(e)}{\geq} \frac{1}{2} + \frac{(\varepsilon_n - 1)}{2} + \frac{w_j}{8} - \frac{w_j\nu(1|\alpha_{<n})}{4}
\end{aligned}$$

where (a) follows by substituting the definition of $\tilde{M}$. (b) is the definition of ε_n. (c) by expanding. (d) by the definition of $\tilde{M}$. (e) by bounding $M \geq \tilde{M}$ and $\nu(\alpha_{<n}) \geq \nu(\alpha_{<n}0) = 2^{-n}$ and $\tilde{M}(\alpha_{<n}) \leq M(\alpha_{<n}) \leq 2^{1-n}$. Therefore

$$(\forall 1 \leq a \leq k) \qquad \liminf_{n\in D_{\leq k}\cap N_k} \left(\tilde{M}(1|\alpha_{<n}1^{k-a}) + \frac{w_j\nu(1|\alpha_{<n}1^{k-a})}{4}\right) \geq \frac{1}{2} + \frac{w_j}{8}.$$

By Lemma 11(ii), if $n \in N_k$ we have that $\nu(1|\alpha_{<n}1^{k-1}) < \frac{1}{3}$ and by Lemma 11(i), $\nu(1|\alpha_{<n}1^{k-a}) < \frac{1}{2}$ for $2 \leq a \leq k$. Therefore for any $2 \leq k \leq a$ and $\gamma := \frac{12}{12+w_j} \in (0,1)$

$$\liminf_{n\in D_{\leq k}\cap N_k} \tilde{M}(1|\alpha_{<n}1^{k-1}) \geq \frac{1}{2} + w_j\left(\frac{1}{8} - \frac{1}{12}\right) = \frac{1}{\gamma}\cdot\frac{1}{2} \qquad (\star)$$

$$\liminf_{n\in D_{\leq k}\cap N_k} \tilde{M}(1|\alpha_{<n}1^{k-a}) > \frac{1}{2}. \qquad (\star\star)$$

Combining $(\star)$ and $(\star\star)$ and using the fact that for positive sequences a_i and b_i

it holds that $\liminf a_i b_i \geq (\liminf a_i)(\liminf b_i)$ we obtain

$$\begin{aligned}
\liminf_{n \in D_{\leq k} \cap N_k} \tilde{M}(1^k | \alpha_{<n}) &= \liminf_{n \in D_{\leq k} \cap N_k} \prod_{a=1}^{k} \tilde{M}(1 | \alpha_{<n} 1^{k-a}) \\
&\geq \left(\liminf_{n \in D_{\leq k} \cap N_k} \tilde{M}(1 | \alpha_{<n} 1^{k-1}) \right) \left(\prod_{a=2}^{k} \liminf_{n \in D_{\leq k} \cap N_k} \tilde{M}(1 | \alpha_{<n} 1^{k-a}) \right) \\
&> \frac{1}{\gamma} \left(\frac{1}{2} \right)^k .
\end{aligned}$$

Therefore there exists a c_k such that for all $n > c_k$ and $n \in D_{\leq k} \cap N_k$

$$\tilde{M}(1^k | \alpha_{<n}) > \frac{1}{\gamma} \left(\frac{1}{2} \right)^k .$$

Let $C_k := \{n : n > c_k \wedge n \in D_{\leq k} \cap N_k\}$. Therefore $d(C_k, D_{\leq k} \cap N_k) = 1$ and so by Lemma 17(i,v) and the previously shown fact that $d(D_{\leq k} \cap N_k, N_k) = 1$ we obtain $d(N_k - C_k) = 0$ as required. $\square$

To prove the main theorem we construct a pair of infinite binary sequences χ and ψ such that $\alpha_{1:n}$ is computable from $\chi_{1:n}$ and $\psi_{1:n}$. This implies that $Km(\alpha_{1:n}) < Km(\chi_{1:n}) + K(\psi_{1:n}) + O(1)$, which holds because you can construct a program for $\alpha_{1:n}$ using the concatenation of a prefix program for $\psi_{1:n}$ and a monotone program for $\chi_{1:n}$. Finally we assume that M converges on-sequence to λ on α and show that this implies $\liminf_{n \to \infty} Km(\chi_{1:n})/n < 1$ and $\lim_{n \to \infty} K(\psi_{1:n})/n = 0$. But α is λ-random, so $\lim_{n \to \infty} Km(\alpha_{1:n})/n = 1$, which leads to a contradiction.

Proof of Theorem 6. Let α be as in the proof of Theorem 4 and γ be as in the proof of Lemma 12. Define

$$A := \left\{ n \in N_1 : c < 2^n \tilde{M}(\alpha_{1:n}) \leq c/\gamma \right\} \quad B := \left\{ n \in N_1 : 2^n \tilde{M}(\alpha_{1:n}) > c/\gamma \right\}$$

where $c \in \mathbb{Q}$ is chosen such that $\bar{d}(A) > 0$ and $d(B) = 0$. This is necessarily possible since by the universality of $\tilde{M}$ and the fact that α is λ-random there exists constants $\mathbb{Q} \ni c', c'' > 0$ such that $2^n \tilde{M}(\alpha_{1:n}) \in [c', c'']$. Define $F \subset \mathbb{N}$ by

$$F := \left\{ n : \alpha_n = 1 \wedge 2^{n-1} \tilde{M}(\alpha_{<n}) > c \right\}$$

Now define sequences χ and ψ by

$$\begin{aligned}
\chi_n &:= [\![\alpha_n = 1 \vee 2^{n-1} \tilde{M}(\alpha_{<n}) > c]\!] \\
\psi_n &:= [\![n \in F]\!]
\end{aligned}$$

Let M_t and $\tilde{M}_t$ be computable approximations of M and $\tilde{M}$ from below respectively. Define $\beta \in \mathcal{B}^\infty$ by

$$\beta_n := \begin{cases} 0 & \text{if } \chi_n = 0 \\ 1 & \text{if } \chi_n = 1 \wedge \psi_n = 1 \\ 1 & \text{if } \chi_n = 1 \wedge \psi_n = 0 \wedge \exists t : M_t(\alpha_{<n}0) > 2^{-n} \\ 0 & \text{if } \chi_n = 1 \wedge \psi_n = 0 \wedge \exists t : 2^{n-1}\tilde{M}_t(\alpha_{<n}) > c. \end{cases}$$

We claim that $\alpha = \beta$ and prove this using four cases.

Case 1 ($\chi_n = 0$). By the definition χ and the fact that $\chi_n = 0$ we have $\alpha_n = 0$ and by the definition of β we have $\beta_n = 0 = \alpha_n$.

Case 2 ($\chi_n = 1 \wedge \psi_n = 1$). By the definition of ψ and the fact that $\psi_n = 1$ we have that $\beta_n = 1$ and $n \in F$, which implies that $\alpha_n = 1 = \beta_n$.

Case 3 ($\chi_n = 0 \wedge \psi_n = 0 \wedge \exists t : M_t(\alpha_{<n}) > 2^{-n}$). If $\exists t : M_t(\alpha_{<n}) > 2^{-n}$, then $\alpha_n = 1$ by the definition of α and the fact that M_t is a monotone increasing approximation of M.

Case 4 ($\chi_n = 0 \wedge \psi_n = 0 \wedge \exists t : 2^{n-1}\tilde{M}_t(\alpha_{<n}) > c$). Since $\psi_n = 0$ we have that $n \notin F$. Therefore either $\alpha_n = 0$ or $2^{n-1}\tilde{M}_t(\alpha_{<n}) \leq c$, but the latter is not true. Therefore $\alpha_n = 0 = \beta_n$, which completes the proof that $\alpha = \beta$.

Since $\alpha_{1:n}$ is λ-random and can be computed from $\psi_{1:n}$ and $\chi_{1:n}$ using the equation above, there exist constants $c_1, c_2 > 0$ such that

$$Km(\chi_{1:n}) + K(\psi_{1:n}) + c_2 > Km(\alpha_{1:n}) > n - c_1$$

where the second inequality follows from [5, Example 4.5.3]. We now work by contradiction Assume contrary to the theorem statement that there does not exist an ε such that

$$d\left(\left\{n : \left|M(\alpha_n|\alpha_{<n}) - \frac{1}{2}\right| > \varepsilon\right\}\right) > 0.$$

Therefore by Lemma 16 there exists a set $D \subseteq \mathbb{N}$ such that $d(D) = 1$ and $\lim_{n\in D} M(\alpha_n|\alpha_{<n}) = \frac{1}{2}$. Then by Lemma 12 there exists a sequence of sets C_k such that $d(N_k - C_k) = 0$ and for all $n \in C_k$

$$\tilde{M}(1^k|\alpha_{<n}) > \frac{1}{\gamma}\left(\frac{1}{2}\right)^k.$$

We start by showing that $d(F) = 0$. Let

$$A_k := \{n \in N_k : n + k \notin B\} \qquad B_k := \{n \in N_k : n + k \in B\},$$

which are translated subsets of A and B respectively and so $d(B_k) = 0$ for all

k. Now fix $\kappa \in \mathbb{N}$ and decompose F as a subset of four sets as follows

$$
\begin{aligned}
F &\overset{(a)}{\subseteq} \left(\bigcup_{k=1}^{\kappa} F \cap N_k\right) \cup \left(\bigcup_{k=\kappa+1}^{\infty} N_k\right) \\
&\subseteq \left(\bigcup_{k=1}^{\kappa} F \cap N_k \cap C_k\right) \cup \left(\bigcup_{k=1}^{\kappa} N_k - C_k\right) \cup \left(\bigcup_{k=\kappa+1}^{\infty} N_k\right) \\
&\subseteq \left(\bigcup_{k=1}^{\kappa} F \cap N_k \cap C_k \cap A_k\right) \cup \left(\bigcup_{k=1}^{\kappa} B_k\right) \cup \left(\bigcup_{k=1}^{\kappa} N_k - C_k\right) \cup \left(\bigcup_{k=\kappa+1}^{\infty} N_k\right)
\end{aligned}
$$

where in (a) we used the facts that $\mathbb{N} = \bigcup_{\kappa=0}^{\infty} N_k$ and $\alpha_n = 1$ for all $n \in F$. We now bound the natural density of each of the four sets in order. Suppose $n \in A_k \cap C_k$. Then $n \in A_k$ and so $2^{n+k-1}\tilde{M}(\alpha_{<n+k}) \leq c/\gamma$. Additionally, $n \in C_k$ so

$$
2^{n-1}\tilde{M}(\alpha_{<n}) \underset{\substack{\uparrow \\ \text{Def. of cond. prob}}}{=} 2^{n-1}\frac{\tilde{M}(\alpha_{<n+k})}{\tilde{M}(1^k|\alpha_{<n})} \underset{\substack{\uparrow \\ n \in C_k}}{\leq} 2^{n-1}\frac{2^{1-n-k}c/\gamma}{\frac{1}{\gamma}\left(\frac{1}{2}\right)^k} = c,
$$

which implies $n \notin F$. Therefore $F \cap N_k \cap A_k \cap C_k = \emptyset$ and

$$
d\left(\bigcup_{k=1}^{\kappa} F \cap N_k \cap C_k \cap A_k\right) = 0.
$$

Also

$$
d\left(\bigcup_{k=1}^{\kappa} B_k\right) = d\left(\bigcup_{k=1}^{\kappa} N_k - C_k\right) = 0
$$

Finally apply Lemma 10 to bound

$$
d\left(\bigcup_{k=\kappa+1}^{\infty} N_k\right) = 2^{-\kappa-1}.
$$

Therefore $\bar{d}(F) < 2^{-\kappa-1}$ for all $\kappa \in \mathbb{N}$. Therefore $d(F) = 0$. Therefore by Lemma 1(iii) we obtain $\lim_{n\to\infty} Km(\psi_{1:n})/n = 0$. Since $|Km(x) - K(x)| < O(\log \ell(x))$ for all x [5, §4.5.5], $\lim_{n\to\infty} K(\psi_{1:n})/n = 0$ as well.

Let $\theta_n := \#1(\chi_{1:n})/n$. By Lemma 1(iii) we have that $\bar{d}(A - F) > 0$ and by Lemma 10 we have that $d(N_0) = \frac{1}{2}$. Since $A \subseteq N_1$ and $N_1 \cap N_0 = \emptyset$, by Lemma 17 we have $\bar{d}(N_0 \cup A - F) > \frac{1}{2}$. Since $\chi_n = 1$ if $n \in N_0 \cup A - F$ there exists a $0 < c_3 \in \mathbb{Q}$ such that $\limsup_{n\to\infty} \theta_n > \frac{1}{2} + c_3$. If $\theta_n > \frac{1}{2} + c_3$ then by Lemma 1(ii) there exists a $c_4 > 0$ such that $Km(\chi_{1:n}) < nH\left(\frac{1}{2} + c_3\right) + \frac{1}{2}\log n + c_4$. Therefore for all $\varepsilon > 0$ there exists an arbitrarily large n such that

$$
\begin{aligned}
n - c_1 &< Km(\alpha_{1:n}) < Km(\chi_{1:n}) + K(\psi_{1:n}) + c_2 \\
&< \varepsilon n + nH\left(\frac{1}{2} + c_3\right) + \frac{1}{2}\log n + c_2 + c_4.
\end{aligned}
$$

This is a contradiction since $H\left(\frac{1}{2}+c_3\right)<1$. Therefore there exists an $\varepsilon>0$ such that

$$d\left(\left\{n : \left|M(\alpha_n|\alpha_{<n}) - \frac{1}{2}\right|\right\}\right) > 0$$

as required. $\square$

5. Normalised non-convergence

We now show that the non-convergence is not only an artifact of the fact that M is not a proper measure and that convergence still fails if M is replaced by the normalised mixture M_{norm}.

Theorem 13. *Let M be a universal mixture and M_{norm} be its normalisation. Then there exists a λ-random α and $\varepsilon>0$ such that*

$$\bar{d}\left(\left\{n : \left|M_{\text{norm}}(\alpha_n|\alpha_{<n}) - \frac{1}{2}\right| > \varepsilon\right\}\right) > 0.$$

The proof uses the same α as in Theorem 6 and runs roughly as follows. It is known that $M_{\text{norm}}(\alpha_n|\alpha_{<n}) > M(\alpha_n|\alpha_{<n})$. Therefore if $M(\alpha_n|\alpha_{<n}) > \frac{1}{2}+\varepsilon$ on some non-zero density set, then $M_{\text{norm}}(\alpha_n|\alpha_{<n})$ does not converge to $\frac{1}{2}$ on that set. But by Theorem 6 we know that there exists a set of non-zero density set on which $M(\alpha_n|\alpha_{<n})$ is bounded away from $\frac{1}{2}$. Therefore if $M_{\text{norm}}(\alpha_n|\alpha_{<n})$ is to converge on some dense set to λ, then $M(\alpha_n|\alpha_{<n})$ must fail by being smaller than $\frac{1}{2}$ for a non-zero proportion of the time. But this implies that the ratio $\lambda(\alpha_{1:n})/M(\alpha_{1:n})$ is increasing with n, which is a contradiction by the dominance $M(\alpha_{1:n}) > c\lambda(\alpha_{1:n})$ for all n and some constant $c>0$.

Proof. Let α be as in the proof of Theorem 6, For $0<\gamma<1<\zeta<2$ define

$$\begin{aligned}
L_{<\gamma} &:= \{n : M(\alpha_n|\alpha_{<n}) < \gamma/2\} \\
L_{\gamma,\zeta} &:= \{n : \gamma/2 \le M(\alpha_n|\alpha_{<n}) \le \zeta/2\} \\
L_{>\zeta} &:= \{n : M(\alpha_n|\alpha_{<n}) > \zeta/2\},
\end{aligned}$$

which are chosen to be disjoint and satisfy $L_{<\gamma}\cup L_{\gamma,\zeta}\cup L_{>\zeta}=\mathbb{N}$. We proceed by contradiction. Assume that $d(L_{>\zeta})=0$ for all $\zeta>1$, which implies that the set for which $M(\alpha_n|\alpha_{<n})>\zeta/2$ has zero density. By Theorem 6 there exists an ε such that

$$d\left(F := \left\{n : \left|M(\alpha_n|\alpha_{<n}) - \frac{1}{2}\right| > \varepsilon\right\}\right) > 0.$$

By definition, $F = L_{<1-2\varepsilon}\cup L_{>1+2\varepsilon}$. By assumption $d(L_{>1+2\varepsilon})=0$. Therefore setting $\gamma=1-2\varepsilon$ leads by Lemma 17(ii) to $\bar{d}(L_{<1-2\varepsilon})=\bar{d}(F)>0$.

Since M is universal there exists a constant $c_\lambda>0$ such that $M(x)>c_\lambda\lambda(x)$ for all $x\in\mathcal{B}^*$ and so

$$(\forall n\in\mathbb{N}) \qquad \frac{1}{n}\log\frac{\lambda(\alpha_{1:n})}{M(\alpha_{1:n})} < \frac{1}{n}\log\frac{1}{c_\lambda} \overset{n\to\infty}{\longrightarrow} 0. \tag{$\star$}$$

Now choose ζ so that $1 < \zeta < 2^{-d(L_{<\gamma})\log\gamma}$. Then we obtain a contradiction by

$$\begin{aligned}
0 &\overset{(a)}{=} \limsup_{n\to\infty} \frac{1}{n} \log \frac{\lambda(\alpha_{1:n})}{M(\alpha_{1:n})} \\
&\overset{(b)}{=} \limsup_{n\to\infty} \frac{1}{n} \left(\sum_{t=1}^{n} \log \frac{1}{M(\alpha_t|\alpha_{<t})} - n \right) \\
&\overset{(c)}{\geq} \limsup_{n\to\infty} \frac{1}{n} \left(\sum_{t\in L_{<\gamma}[n]} \log \frac{1}{\gamma} + \sum_{t\in L_{\gamma,\zeta}} \log \frac{1}{\zeta} + \sum_{t\in L_{>\zeta}[n]} \log \frac{1}{2} \right) \\
&\overset{(d)}{=} \bar{d}(L_{<\gamma}) \log \frac{1}{\gamma} + \bar{d}(L_{\gamma,\zeta}) \log \frac{1}{\zeta} + \bar{d}(L_{>\zeta}) \overset{(e)}{\geq} d(L_{<\gamma}) \log \frac{1}{\gamma} + \log \frac{1}{\zeta} \overset{(f)}{>} 0
\end{aligned}$$

where (a) follows from ($\star$) (b) by the definition of the Lebesgue measure λ. (c) by bounding $M(\alpha_n|\alpha_n)$ for $n \in L_{<\gamma}$, $L_{\gamma,\zeta}$ and $L_{>\zeta}$. (d) by the definition of the natural density. (e) since $\bar{d}(A) \leq 1$ for all A. (f) by choosing $1 < \zeta < 2^{-d(L_{<\gamma})\log\gamma}$. Therefore there exists a $\zeta > 1$ such that $\bar{d}(L_{>\zeta}) > 0$. Finally, for $n \in L_{>\zeta}$ we have $M_{\text{norm}}(\alpha_n|\alpha_{<n}) > M(\alpha_n|\alpha_{<n}) > \frac{\zeta}{2}$, which implies that

$$(\forall n \in L_{>\zeta}) \qquad \left| M_{\text{norm}}(\alpha_n|\alpha_{<n}) - \frac{1}{2} \right| > \frac{\zeta - 1}{2} > 0.$$

as required. $\square$

6. Convergence on non-random sequences

Here we give something of a converse to Theorem 6 via a small extension to Theorem 7 in [3]. Not only can the predictive distribution of M fail to converge to λ on some λ-random sequences, but it can also succeed in converging to λ on sequences that are not λ-random. The proof relies on defining a measure μ for which the predictive distribution $\mu(\cdot|\omega_{<n})$ converges to $\lambda(\cdot|\omega_{<n}) = \frac{1}{2}$ at just the right rate to ensure that if ω is μ-random, then it is not λ-random.

Proposition 14. *There exists an $\omega \in \mathcal{B}^\infty$ such that*

1. *ω is not λ-random.*
2. *For all universal lower semi-computable semimeasures M*

$$\lim_{t\to\infty} \sum_{b\in\mathcal{B}} \left(M(b|\omega_{<t}) - \frac{1}{2} \right)^2 = 0.$$

Proof. Define computable measure ν inductively by

$$\mu(1|x) := \frac{1}{2} + \frac{1}{2\sqrt{1 + \ell(x)}}$$

For universal lower semi-computable semimeasure M define the set of μ-random sequences on which M converges to μ by

$$A_M := \left\{ \omega : \lim_{t\to\infty} \sum_{b\in\mathcal{B}} (M(b|\omega_{<t}) - \mu(1|\omega_{<t}))^2 = 0 \wedge \omega \text{ is } \mu\text{-random} \right\}.$$

Now $\mu(A_M) = 1$ by Theorem 2 and the well-known fact that $\mu(\mathcal{R}_\mu) = 1$ for all computable measures μ. Therefore since there are only countably many universal lower semi-computable semimeasures, we have $\mu\left(A := \bigcap_M A_M\right) = 1$. Let $\omega \in A$, which is μ-random. Then

$$\sum_{t=1}^{\infty} \sum_{b\in\mathcal{B}} \left(\sqrt{\mu(b|\omega_{<t})} - \sqrt{\lambda(b|\omega_{<t})}\right)^2 \geq \sum_{t=1}^{\infty} \left(\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{t}}} - \sqrt{\frac{1}{2}}\right)^2 = \infty.$$

Therefore ω is not λ-random by Theorem 3 of [11]. Finally by the definition of $\omega \in A$ and μ we have that for all universal lower semi-computable semimeasures M

$$\lim_{t\to\infty} \sum_{b\in\mathcal{B}} \left(M(b|\omega_{<t}) - \frac{1}{2}\right)^2 = \lim_{t\to\infty} \sum_{b\in\mathcal{B}} \left(M(b|\omega_{<t}) - \mu(b|\omega_{<t})\right)^2 = 0$$

as required. $\square$

7. Summary

We have shown that for every universal mixture M there exists an infinite λ-random sequence α and $\varepsilon > 0$ such that

$$\bar{d}\left(\left\{ n : \left| M(\alpha_n|\alpha_{<n}) - \frac{1}{2} \right| > \varepsilon \right\}\right) > 0, \qquad (\star)$$

which means that the predictive distribution of M is wrong by at least ε for a non-zero fraction of the time. This extends the previously known results that there existed a universal mixture for which this kind of failure occurred [3]. We also showed that $(\star)$ holds even if M is replaced by the normalised version of the universal mixture M_{norm}. It is known that the totally incomputable mixture over all computable measures does converges on all Martin-Löf sequences [4], so the failure of M_{norm} is somewhat surprising and shows that the distortions caused by normalisation are substantial.

We also showed that there exists a single infinite sequence α that is not λ-random, but on which the predictive distributions of all M converge to λ. This result is unsurprising, since α was constructed to be Martin-Löf random with respect to some measure for which the predictive distribution converges to the that of the Lebesgue measure, but does so sufficiently slowly that α is not itself λ-random.

Open Problems. There are a number of natural questions remaining. Suppose M is a universal lower semi-computable semimeasure and define $\mathcal{C}_M$ and $\mathcal{C}$ by

$$\mathcal{C}_M := \left\{\omega : \lim_{t\to\infty} M(\omega_n|\omega_{<n}) = \frac{1}{2}\right\} \quad \text{and} \quad \mathcal{C} := \bigcap_M \mathcal{C}_M$$

where the intersection is taken over all universal lower semi-computable semimeasures. What is the nature of $\mathcal{C}_M$ and $\mathcal{C}$? It follows from [3] that there exists an M such that $\mathcal{R}_\lambda \not\subseteq \mathcal{C}_M$, which implies that $\mathcal{R}_\lambda \not\subseteq \mathcal{C}$. In [7] it is shown that the 2-random reals are a subset of $\mathcal{C}$. In this work we showed that for all universal mixtures $\mathcal{R}_\lambda \not\subseteq \mathcal{C}_M$. Obvious open questions are:

1. Does there exists a universal lower semi-computable semimeasure (not a mixture) such that $\mathcal{R}_\lambda \subseteq \mathcal{C}_M$? An example of a non-trivial universal enumerable semimeasure that is not (essentially) a mixture may also be of interest.
2. As above, but where $\mathcal{R}_\lambda$ is replaced with a different class of random reals somewhere on the hierachy between Martin-Löf random and 2-random reals, such as the weak 2-random reals.

Unfortunately, an elegant characterisation of $\mathcal{C}_M$ and $\mathcal{C}$ seems unlikely because there exists an $\omega \in \mathcal{C}$ that is not λ-random (Proposition 14). Note that it is known that there exists a lower semicomputable semimeasure W that converges on all λ-random sequences, but W is not universal. A mixture over all computable measures also converges on all λ-random sequences, but is not lower semicomputable [4].

References

[1] Cristian Calude. *Information and Randomness: An Algorithmic Perspective.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2nd edition, 2002.

[2] M. Hutter. On universal prediction and Bayesian confirmation. *Theoretical Computer Science*, 384(1):33–48, 2007.

[3] M. Hutter and A. Muchnik. Universal convergence of semimeasures on individual random sequences. In *Proc. 15th International Conf. on Algorithmic Learning Theory (ALT'04)*, volume 3244 of *LNAI*, pages 234–248, Padova, 2004. Springer, Berlin.

[4] M. Hutter and A. Muchnik. On semimeasures predicting Martin-Löf random sequences. *Theoretical Computer Science*, 382(3):247–261, 2007.

[5] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, Verlag, 3rd edition, 2008.

[6] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9(6):602 – 619, 1966.

[7] K. Miyabe. An optimal superfarthingale and its convergence over a computable topological space. In *Solomonoff Memorial*, Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2011.

[8] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, 13(6):1076–1136, 2011.

[9] R. Solomonoff. A formal theory of inductive inference, Part I. *Information and Control*, 7(1):1–22, 1964.

[10] R. Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *Information Theory, IEEE Transactions on*, 24(4):422–432, 1978.

[11] V. Vovk. On a randomness criterion. *Soviet Mathematics Doklady*, 35:656–660, 1987.

[12] F. Willems, Y. Shtarkov, and T. Tjalkens. The context tree weighting method: Basic properties. *IEEE Transactions on Information Theory*, 41:653–664, 1995.

[13] I. Wood, P. Sunehag, and M. Hutter. (Non-)equivalence of universal priors. In *Solomonoff Memorial*, Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2011.

[14] A. Zvonkin and L. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83, 1970.

Appendix A. Technical Results

Lemma 15. *If $a > 0$, then the function $f(x) = x/(a + x)$ is monotone increasing.*

Proof. $\frac{\partial}{\partial x} f(x) = \frac{1}{a+x} - \frac{x}{(a+x)^2} = \frac{a}{(a+x)^2} > 0$. □

Lemma 16. *Let $\{\delta_k\}_{k=1}^{\infty}$ be a sequence of non-negative reals. For $\varepsilon > 0$ define $D_\varepsilon := \{n : \delta_n > \varepsilon\}$. If $d(D_\varepsilon) = 0$ for all k, then there exists a $C \subseteq \mathbb{N}$ such that $d(C) = 1$ and $\lim_{n \in C} \delta_n = 0$.*

Proof. Define $\varepsilon_k := 2^{-k}$ and $C_k := \mathbb{N} - D_{\varepsilon_k}$. Then $d(C_{\varepsilon_k}) = 1$ for all k. Furthermore, let $\{n_k\}_{k=1}^{\infty}$ be a monotone increasing sequence such that $C_k[n]/n > 1 - 2^{-k}$ for all $n > n_k$, which must exist since $d(C_k) = 1$. Then let $C := \bigcup_{k=0}^{\infty} C_k(n_{k+1})$. If $n_k \leq n \leq n_{k+1}$, then $C[n]/n \geq C_k[n]/n \geq 1 - 2^{-k}$. Since n_k is finite for all k it follows that $d(C) = \lim_{n \to \infty} C[n]/n = 1$. □

Lemma 17. *Let $A, B, C \subseteq \mathbb{N}$ such that $A, B \subseteq C$. Then (provided all quantities exist)*

$$\begin{array}{rrcl}
(i) & A \subseteq B & \Longrightarrow & d(A,C) = d(A,B)d(B,C) \\
(ii) & A \cap B = \emptyset & \Longrightarrow & d(A \cup B) = d(A) + d(B) \\
(iii) & & & d(A) + d(B) - d(A \cap B) = d(A \cup B) \\
(iv) & A \subseteq B & \Longrightarrow & d(A) \leq d(A,B) \\
(v) & d(B,C) = 1 & \Longrightarrow & d(C - B) = 0 \\
(vi) & d(A) = 1 \wedge d(B) > 0 & \Longrightarrow & d(A \cap B, B) = 1 \\
(vii) & A \subseteq B & \Longrightarrow & \bar{d}(A) \leq \bar{d}(B)
\end{array}$$

Proof.

$$(i) \quad \lim_{n\to\infty} \frac{|A[n]|}{|C[n]|} \overset{\substack{\text{given limits exist} \\ \downarrow}}{=} \left(\lim_{n\to\infty} \frac{|A[n]|}{|B[n]|}\right)\left(\lim_{n\to\infty} \frac{|B[n]|}{|C[n]|}\right) \overset{\substack{\text{definition of density} \\ \downarrow}}{=} d(A,B)d(B,C)$$

$$(ii) \quad d(A \cup B) \overset{\substack{\text{def. of density} \\ \downarrow}}{=} \lim_{n\to\infty} \frac{|(A \cup B)[n]|}{n} \overset{\substack{\text{since } A \cap B = \emptyset \\ \downarrow}}{=} \lim_{n\to\infty} \frac{|A[n]|}{n} + \lim_{n\to\infty} \frac{|B[n]|}{n} \overset{\substack{\text{def. of density} \\ \downarrow}}{=} d(A) + d(B)$$

$$(iii) \quad d(A \cup B) \overset{\substack{\text{part (ii)} \\ \downarrow}}{=} d(A) + d(B - A \cap B) \overset{\substack{\text{part (ii)} \\ \downarrow}}{=} d(A) + d(B) - d(A \cap B)$$

$$(iv) \quad d(A) = \lim_n \frac{A[n]}{n} \overset{\substack{|B[n]| \leq n \\ \downarrow}}{\leq} \lim_n \frac{|A[n]|}{|B[n]|} = d(A,B)$$

$$(v) \quad d(B - A) \overset{\substack{\text{part (iv)} \\ \downarrow}}{\leq} d(B - A, B) = d(B,B) - d(A,B) \overset{\substack{d(A,B) = 1 \\ \downarrow}}{=} 0$$

$$(vi) \quad d(A \cap B, B) \overset{\substack{\text{part (i)} \\ \downarrow}}{=} d(A \cap B)/d(B) \overset{\substack{\text{part (vii)} \\ \downarrow}}{=} 1$$

$$(vii) \quad \bar{d}(A) = \limsup_n \frac{|A[n]|}{n} \overset{\substack{|A[n]| \leq |B[n]| \\ \downarrow}}{\leq} \limsup_n \frac{|B[n]|}{n} = \bar{d}(B)$$

□

Appendix B. Table of Notation

$\wedge$, $\vee$	logical and/or respectively
log	logarithm with base 2
ε, δ	small things

$[\![expr]\!]$	indicator function
$\mathbb{N}$	natural numbers
$\mathbb{R}$	real numbers
$\mathbb{Q}$	rational numbers
$\mathcal{B}$	binary alphabet $\mathcal{B} = \{0, 1\}$
$\mathcal{B}^*$	set of finite binary strings
$\mathcal{B}^\infty$	set of infinite binary strings
x, y	finite binary strings
$x \sqsubseteq y$	x is a prefix of y
$\#x(y)$	number of (possibly overlapping) occurrences of x in y
$\ell(x)$	length of string x
ϵ	empty string of length zero
$A, B, C, \cdots$	subsets of the integers
$A[n]$	set of elements in A smaller or equal to n
A^c	complement of A
$d(A)$	natural density of A
$\bar{d}(A)$	upper natural density of A
μ, ν	lower semicomputable semimeasures
λ	Lebesgue measure
$\nu_1, \nu_2, \nu_3, \cdots$	enumeration of lower semicomputable semimeasures
M	universal mixture
M_{norm}	normalised universal mixture
w_k	prior belief in environment ν_k
$w_k(x)$	posterior belief in environment ν_k having observed x
$K(x)$	prefix Kolmogorov complexity of x
$Km(x)$	monotone Kolmogorov complexity of x
$\omega,\ \chi,\ \psi$	infinite binary strings
α	infinite binary Martin-Löf random string
γ, ζ	numbers in $(0, 2)$
θ	number in $(0, 1)$
$H(\theta)$	entropy of θ