

# On Martin-Löf Convergence of Solomonoff's Mixture

Tor Lattimore and Marcus Hutter

Australian National University  
{tor.lattimore,marcus.hutter}@anu.edu.au

**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.

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

## 1 Introduction

Sequence prediction is the task of predicting symbol  $\alpha_n$  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.<sup>1</sup> 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  $\mu$  if it satisfies all the properties one would expect of an infinite sequence sampled from  $\mu$ . It 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  $\alpha$  (with respect to the Lebesgue measure  $\lambda$ ) for which  $M(\alpha_n|\alpha_{<n}) \not\rightarrow \lambda(\alpha_n|\alpha_{<n})$ . The  $\alpha$  used in their proof is computable from

---

<sup>1</sup> 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].

the halting problem, which presumably inspired the work in [7] where it is shown that if  $\alpha$  is 2-random, then every universal lower semicomputable semimeasure converges on  $\alpha$ . It is worth remarking that there are known semimeasures that do converge on all Martin-Löf random sequences, some of which are even lower semicomputable. Unfortunately, however, convergence rates for these semimeasures are unknown. For a detailed discussion see [3].

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.

## 2 Notation

Overviews of algorithmic information theory can be found in [5, 1].

**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  $\llbracket expr \rrbracket$ , 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  $A$ . The natural density of  $A \subseteq \mathbb{N}$  is  $d(A) := \lim_{n \rightarrow \infty} |\{a \in A : a \leq n\}| / n$ . and  $\bar{d}(A) := \limsup_{n \rightarrow \infty} |\{a \in A : a \leq n\}| / n$ . We use  $\vee$  and  $\wedge$  for logical or and and respectively.

**Strings.** A finite binary string  $x$  is a finite sequence  $x_1x_2x_3 \cdots x_n$  with  $x_i \in \mathcal{B} := \{0, 1\}$ . Its length is  $\ell(x)$ . An infinite binary string  $\omega$  is an infinite sequence  $\omega_1\omega_2\omega_3 \cdots$ . The empty string of length zero is denoted by  $\varepsilon$ . 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  $x_{s:t} := x_sx_{s+1} \cdots x_{t-1}x_t$  where  $s, t \in \mathbb{N}$  and  $s \leq t$ . If  $s > t$ , then  $x_{s:t} := \varepsilon$ . 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,  $\#010(1010) = 2$ . 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  $\forall x \in \mathcal{B}^*, \lim_{n \rightarrow \infty} \#x(\omega_{1:n})/n = 2^{-\ell(x)}$ .

**Measures and semimeasures.** A semimeasure is a function  $\mu : \mathcal{B}^* \rightarrow [0, 1]$  satisfying  $\mu(\varepsilon) \leq 1$  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  $\mu : \mathcal{B}^* \rightarrow \mathbb{R}$  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 recursively enumerable sequence  $\mu_1, \mu_2, \cdots$

of computable functions approximating  $\mu$  from below. For  $b \in \mathcal{B}$  and  $x \in \mathcal{B}^*$ ,  $\mu(b|x) := \mu(xb)/\mu(x)$  is the  $\mu$ -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  $T := \{(p^1, x^1), (p^2, x^2), \dots\}$  where  $p^k$  is the program for  $x^k$ . 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  $K_T : \mathcal{B}^* \rightarrow \mathbb{N}$  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  $\mu$  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  $\nu_1, \nu_2, \dots$  be such an enumeration and  $w : \mathbb{N} \rightarrow [0, 1]$  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 inconsistency 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].

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

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

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  $\mu$ -random strings.

**Lemma 1.** *The following hold:*

1. *If  $\omega \in \mathcal{B}^\infty$  is  $\lambda$ -random, then it is normal.*
2. *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  $n$ .*
3. *Let  $A, B \subseteq \mathbb{N}$  and  $\phi_n := \llbracket n \in A \rrbracket$ . If  $d(A) = 0$  and  $\bar{d}(B) > 0$ , then*
  - (a)  $\bar{d}(B - A) > 0$ .
  - (b)  $\lim_{n \rightarrow \infty} Km(\phi_{1:n})/n = 0$ .

**Proof.** Part 1 is well known [5, §2.6]. For part 2 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  $\mu$  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

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

where we used the redundancy bound for the KT-estimator [12] and the definition of  $H(\theta)$ . Part 3a is immediate from the definition of the natural density. For 3b, let  $\theta_n := \#1(\phi_{1:n})/n$  and note that  $d(A) = 0$  implies that  $\lim_{n \rightarrow \infty} \theta_n = 0$  and so  $\lim_{n \rightarrow \infty} H(\theta_n) = 0$ . Finally apply part 2 to complete the proof. ■

### 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  $\alpha$  is sampled from computable measure  $\mu$ , then*

$$\lim_{n \rightarrow \infty} \sum_{b \in \mathcal{B}} (M(b|\alpha_{<n}) - \mu(b|\alpha_{<n}))^2 = 0, \quad w.\mu.p.1.$$

A subtle point is that convergence in Theorem 2 holds both off-sequence and on-sequence. A weaker (on-sequence only) statement would be that  $\lim_{n \rightarrow \infty} (M(\alpha_n|\alpha_{<n}) - \mu(\alpha_n|\alpha_{<n}))^2 = 0, w.\mu.p.1$ . Unfortunately, both results only hold with probability 1 while we are primarily interested in convergence on individual sequences.

## 4 Martin-Löf Convergence

We now ask whether there exists a universal mixture such that  $M(\alpha_n|\alpha_{<n}) \rightarrow \mu(\alpha_n|\alpha_{<n})$  for all  $\mu$ -random  $\alpha$ . 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 3.** *Let  $M$  be a universal mixture. Then there exists a  $\lambda$ -random  $\alpha$  such that  $\lim_{n \rightarrow \infty} \sum_{b \in \mathcal{B}} (M(b|\alpha_{<n}) - 1/2)^2 \neq 0$ .*

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

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

It is straightforward to check that  $\nu$  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  $M$ . By the definition of  $\nu$  we have that  $\nu(\alpha_{1:n}) = M(\alpha_{1:n})$  for all  $n$ . 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  $\alpha$ ,  $\nu$  and the conditional probability respectively. Therefore if  $\alpha_n = 0$ , then

$$\begin{aligned} M(0|\alpha_{<n}) + M(1|\alpha_{<n}) &\stackrel{(a)}{=} \sum_{i \in \mathbb{N}} w_i(\alpha_{<n}) (\nu_i(0|\alpha_{<n}) + \nu_i(1|\alpha_{<n})) \\ &\stackrel{(b)}{\leq} \sum_{i \in \mathbb{N}} w_i(\alpha_{<n}) - w_j(1 - M(0|\alpha_{<n})) \\ &\stackrel{(c)}{=} 1 - w_j(1 - M(0|\alpha_{<n})) \stackrel{(d)}{\leq} 1 - w_j M(1|\alpha_{<n}), \end{aligned} \quad (4)$$

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 fact that  $\sum_{i \in \mathbb{N}} w_i(x) = 1$ . For (d) we note that  $M$  is a semimeasure, which implies that  $1 - M(0|\alpha_{<n}) \geq M(1|\alpha_{<n})$ . Because  $\alpha$  is  $\lambda$ -random, it must contain infinitely many zeros by part 1 of Lemma 1 and the definition of a normal string. Let  $n_i$  be the position of the  $i$ th 0 in  $\alpha$  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  $\lambda$ . (c) follows by recalling that  $\lambda(1|\alpha_{<n_i}) = 1/2$  and the fact that  $\alpha$  is  $\lambda$ -random combined with Eq. 3. Then by Eq. 4,

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

Therefore  $\lim_{n \rightarrow \infty} M(0|\alpha_{<n}) + M(1|\alpha_{<n}) \neq 1$  and so  $\lim_{n \rightarrow \infty} \sum_{b \in \mathcal{B}} (M(b|\alpha_{<n}) - 1/2)^2 \neq 0$ , as required. ■

Coincidentally, the proof of Theorem 3 demonstrates the existence random sequences on which  $M$  fails to converge to a proper measure. This is interesting as it is a straightforward corollary of Theorem 2 that  $M$  converges to a measure with  $\mu$ -probability one with respect to any computable measure  $\mu$ .

We now present the on-sequence version of Theorem 3, which uses the same  $\alpha$  for a counter-example, but turns out to be significantly harder to prove.

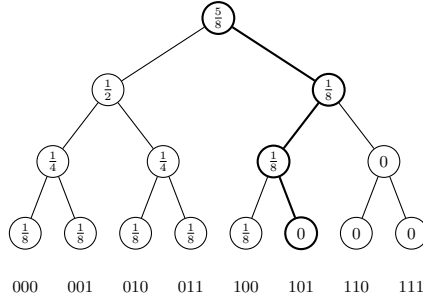
**Theorem 4.** *Let  $M$  be a universal mixture. Then there exists a  $\lambda$ -random  $\alpha$  such that  $\lim_{n \rightarrow \infty} M(\alpha_n|\alpha_{<n}) \neq 1/2$ .*

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

**Definition 5.** *Let  $M_t$  be a sequence of computable functions approximating  $M$  from below and define  $\alpha^t \in \mathcal{B}^\infty$  similarly to  $\alpha$  by  $\alpha_n^t := \llbracket M_t(\alpha_{<n}^t 0) > 2^{-n} \rrbracket$ . Now define  $\nu_t : \mathcal{B}^* \rightarrow [0, 1]$  by*

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

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



**Fig. 1.**  $\nu_3$  if  $\alpha_{1:3}^3 = 101$

It is shown in [3] that  $\lim_{t \rightarrow \infty} \alpha^t = \alpha$  and  $2^{-n} \equiv \lambda(\alpha_{1:n}) > M(\alpha_{1:n})$ . Additionally,  $\nu := \lim_{t \rightarrow \infty} \nu_t$  exists and is a lower semicomputable semimeasure with  $\nu(x) = \nu(x0) + \nu(x1)$  and  $\nu(\alpha_{1:n}) < 2^{-n}$ . Hutter and Muchnik then argued that if  $\alpha_{n:n+1} = 01$ , then  $\nu(\alpha_n|\alpha_{<n}) = 1$  and  $\nu(\alpha_{<n}) \geq M(\alpha_{<n})/2$ . They then set  $M' := \gamma M + (1 - \gamma)\nu$  for suitable  $\gamma$  and so poisoned convergence of either  $M$  or  $M'$ . Here we diverge from their work and consider  $\nu$  when predicting ones. For the remainder of this article  $\alpha$  and  $\nu$  refer to those defined above.

**Lemma 6.** *The following hold:*

1. *If  $\alpha_{n:n+1} = 10$  then  $\nu(1|\alpha_{<n}) \in (0, 1/3)$ .*

2. If  $\alpha_n = 1$  then  $\nu(1|\alpha_{<n}) \in (0, 1/2)$ .

**Proof.** For part one,

$$\begin{aligned} \nu(1|\alpha_{<n}) &\stackrel{(a)}{=} \frac{\nu(\alpha_{<n}1)}{\nu(\alpha_{<n})} \stackrel{(b)}{=} \frac{\nu(\alpha_{<n}10) + \nu(\alpha_{<n}11)}{\nu(\alpha_{<n}0) + \nu(\alpha_{<n}10) + \nu(\alpha_{<n}11)} \\ &\stackrel{(c)}{=} \frac{\nu(\alpha_{<n}10)}{\nu(\alpha_{<n}0) + \nu(\alpha_{<n}10)} \stackrel{(d)}{=} \frac{\nu(\alpha_{<n}10)}{2^{-n} + \nu(\alpha_{<n}10)} \stackrel{(e)}{<} \frac{2^{-n-1}}{2^{-n} + 2^{-n-1}} = \frac{1}{3}. \end{aligned}$$

(a) is the definition of the conditional measure. (b) follows because  $\nu(x) = \nu(x0) + \nu(x1)$ . (c) and (d) are true, since  $\alpha_{n:n+1} = 10$  and so  $\alpha_{<n}11 > \alpha_{1:n+1}$  and  $\alpha_{<n}0 < \alpha_{1:n}$ , which imply that  $\nu(\alpha_{<n}11) = 0$  and  $\nu(\alpha_{<n}0) = 2^{-n}$ . (e) follows from algebra and because  $\nu(\alpha_{1:n}) < 2^{-n}$  for all  $n$ . For the second part we use the same reasoning to obtain

$$\nu(1|\alpha_{<n}) = \frac{\nu(\alpha_{<n}1)}{\nu(\alpha_{<n}0) + \nu(\alpha_{<n}1)} = \frac{\nu(\alpha_{<n}1)}{2^{-n} + \nu(\alpha_{<n}1)} < \frac{1}{2},$$

as required. ■

**Lemma 7.** Let  $n_i$  be the position of the  $i$ th 1 in  $\alpha$  and  $j \in \mathbb{N}$  be such that  $\nu = \nu_j$  in the enumeration of all lower semicomputable semimeasures used by  $M$ . If  $\lim_{n \rightarrow \infty} M(\alpha_n|\alpha_{<n}) = \frac{1}{2}$ , then the function  $\bar{M}(x) := M(x) - w_j\nu(x)/2$  satisfies the following properties:

1.  $\bar{M}$  is a universal mixture.
2.  $\liminf_{i \rightarrow \infty} \bar{M}(1|\alpha_{<n_i}) \geq \frac{1}{2}$ .
3. There exists  $\gamma \in (0, 1)$  such that for all sufficiently large  $n$  with  $\alpha_{n:n+1} = 10$ ,  $\bar{M}(1|\alpha_{<n}) > \frac{1}{2\gamma^2}$ .

**Proof.** The first part is trivial. For the third part, let  $w_k$  be the prior weight that  $M$  assigns to itself,  $n$  be such that  $\alpha_{n:n+1} = 10$  and  $\epsilon_n := \frac{1}{2} - \frac{M(\alpha_{<n}1)}{M(\alpha_{<n})}$ . Then

$$\begin{aligned} \bar{M}(1|\alpha_{<n}) &\stackrel{(a)}{=} \frac{M(\alpha_{<n}1) - w_j\nu(\alpha_{<n}1)/2}{M(\alpha_{<n}) - w_j\nu(\alpha_{<n})/2} \stackrel{(b)}{>} \frac{M(\alpha_{<n}1) - w_j\nu(\alpha_{<n})/6}{M(\alpha_{<n}) - w_j\nu(\alpha_{<n})/2} \\ &\stackrel{(c)}{=} \frac{1}{2} + \frac{w_j\nu(\alpha_{<n}) - 12M(\alpha_{<n})\epsilon_n}{12\bar{M}(\alpha_{<n})} \stackrel{(d)}{>} \frac{1}{2} + \frac{w_j}{24} - w_k\epsilon_n, \end{aligned}$$

(a) is the definition of  $\bar{M}$  and conditional probability. (b) follows from part 1 of Lemma 6. In (c) we substituted  $\epsilon_n$ . (d) by substituting inequalities  $\nu(\alpha_{<n}) \geq \nu(\alpha_{<n}0) = 2^{-n} > M(\alpha_{<n})/2 > \bar{M}(\alpha_{<n})/2$  and  $\bar{M}(\alpha_{<n}) > w_k M(\alpha_{<n})$ . Since  $\epsilon_n \rightarrow 0$  for sufficiently large  $n$  with  $\alpha_{n:n+1} = 10$ , we have  $\bar{M}(1|\alpha_{<n}) > 1/2 + w_j/48 = \frac{1}{2\gamma^2}$  where  $\gamma^2 := \frac{1}{1+w_j/24} \in (0, 1)$ . For the second part

$$\begin{aligned} \bar{M}(1|\alpha_{<n_i}) &= \frac{M(\alpha_{<n_i}1) - w_j\nu(\alpha_{<n_i}1)/2}{M(\alpha_{<n_i}) - w_j\nu(\alpha_{<n_i})/2} \stackrel{(a)}{>} \frac{M(\alpha_{<n_i}1) - w_j\nu(\alpha_{<n_i})/4}{M(\alpha_{<n_i}) - w_j\nu(\alpha_{<n_i})/2} \\ &= \frac{1}{2} - \frac{M(\alpha_{<n_i})\epsilon_{n_i}}{M(\alpha_{<n_i})} \geq \frac{1}{2} - w_k\epsilon_{n_i}, \end{aligned}$$

where (a) follows from part 2 of Lemma 6. Taking the limit as  $i \rightarrow \infty$  completes the result.  $\blacksquare$

To prove the main theorem we construct a pair of infinite binary sequences  $\chi$  and  $\psi$  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  $\lambda$  on  $\alpha$  and show that this implies  $\liminf_{n \rightarrow \infty} Km(\chi_{1:n})/n < 1$  and  $\lim_{n \rightarrow \infty} K(\psi_{1:n})/n = 0$ . But  $\alpha$  is  $\lambda$ -random, so  $\lim_{n \rightarrow \infty} Km(\alpha_{1:n})/n = 1$ , which leads to a contradiction.

**Proof of Theorem 4.** Let  $\alpha$  be as in the proof of Theorem 3. Define  $\{m_i\}$  and  $\{n_i\}$  inductively by

$$\begin{aligned} m_1 &:= \min \{m : \alpha_m = 1\} \\ n_i &:= \min \{n \geq m_i : \alpha_{n+1} = 0\} \\ m_i &:= \min \{m > n_{i-1} : \alpha_m = 1\}, \end{aligned}$$

which are chosen so that  $\alpha_{m_i-1:n_i+1} = 01^{n_i-m_i+1}0$ . Since  $\alpha$  is  $\lambda$ -random, by part 1 of Lemma 1,  $d(\{n_i : i \in \mathbb{N}\}) > 0$ . Furthermore,  $\bar{M}$  is universal so by Eq. 3 there exists an  $\epsilon > 0$  such that  $1 \geq 2^{n_i} \bar{M}(\alpha_{1:n_i}) > \epsilon$  for all  $i$ . Let  $\gamma$  be as in the proof of Lemma 7. Therefore we can choose a  $c \in \mathbb{Q}$  such that:

1.  $\bar{d}(A := \{i : c < 2^{n_i} \bar{M}(\alpha_{1:n_i}) \leq c/\gamma\}) > 0$ .
2.  $\bar{d}(B := \{i : 2^{n_i} \bar{M}(\alpha_{1:n_i}) > c/\gamma\}) = 0$ .

Define  $F \subset \mathbb{N}$  by

$$F := \{i : \exists j \in \{m_i, \dots, n_i - 1\} \text{ such that } 2^j \bar{M}(\alpha_{1:j}) > c\} - B.$$

Now define indicators  $\chi$  and  $\psi$  by

$$\begin{aligned} \chi_n &:= \llbracket \alpha_n = 1 \vee \exists i : (n = n_i + 1 \wedge i \in A - F) \rrbracket \\ \psi_n &:= \llbracket \exists i : n = m_i \wedge i \in F \cup B \rrbracket. \end{aligned}$$

Let  $M_t$  and  $\bar{M}_t$  be computable approximations of  $M$  and  $\bar{M}$  from below respectively and  $m(x) := \max \{m \leq \ell(x) : x_{m-1} = 0 \vee m = 1\}$ . Then

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

The equation above is computable given  $\chi_{1:n}$ ,  $\psi_{1:n}$  and  $\alpha_{<n}$  by the following argument.

1. The first two cases are straightforward since  $m(\alpha_{<n}1)$  is computable.



2. If neither the first nor second case match, then by the definitions of  $\chi$  and  $\psi$  exactly one of the 3rd or 4th cases must hold. Therefore the conditions can be computed in parallel for increasing  $t$  until one completes.

Since  $\alpha_{1:n}$  is  $\lambda$ -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 and show that if  $\lim_{n \rightarrow \infty} M(\alpha_n | \alpha_{<n}) = \frac{1}{2}$  then  $Km(\chi_{1:n}) + K(\psi_{1:n})$  is smaller than  $n - c_1 - c_2$  for sufficiently large  $n$ .

We start by showing that  $d(F) = 0$ . By Lemma 7, for each  $k \in \mathbb{N}$  there exists an  $N_k$  such that if  $i > N_k$ , then  $\bar{M}(\alpha_{n_i} | \alpha_{<n_i}) > 1/(2\gamma^2)$  and  $\bar{M}(1 | \alpha_{<n}) > \gamma^{1/k}/2$  whenever  $\alpha_n = 1$  and  $n \geq m_i$ . Suppose  $N_k < i \notin B$  and  $j \in \{m_i, \dots, n_i - 1\}$  with  $\ell_i := n_i - m_i + 1 \leq k$ , then

$$\begin{aligned} 2^j \bar{M}(\alpha_{1:j}) &\stackrel{(a)}{=} 2^j \frac{\bar{M}(\alpha_{1:n_i})}{\bar{M}(1 | \alpha_{<n_i})} \prod_{n=j+1}^{n_i-1} \frac{1}{\bar{M}(1 | \alpha_{<n})} \\ &\stackrel{(b)}{<} \gamma^2 2^{n_i} \bar{M}(\alpha_{1:n_i}) \gamma^{-(n_i-j-1)/k} \stackrel{(c)}{\leq} c \gamma^{1-(n_i-j-1)/k} \stackrel{(d)}{\leq} c, \end{aligned}$$

where (a) follows from the definition of the conditional measure. (b) follows from the inequalities  $\bar{M}(1 | \alpha_{<n_i}) > 1/(2\gamma^2)$  and  $\bar{M}(1 | \alpha_{<n}) > \gamma^{1/k}/2$ . (c) is true by the assumption that  $i \notin B$ , which implies that  $2^{n_i} \bar{M}(\alpha_{1:n_i}) \leq c/\gamma$ . Finally (d) follows because  $n_i - j - 1 \leq n_i - m_i + 1 \leq k$ . Therefore  $i \notin F$  and

$$\begin{aligned} \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \llbracket i \in F \rrbracket &\stackrel{(a)}{\leq} \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} (\llbracket \ell_i \leq k \rrbracket \llbracket i \in F \rrbracket + \llbracket \ell_i > k \rrbracket) \\ &\stackrel{(b)}{\leq} \frac{N_k}{\mathcal{I}} + \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \llbracket \ell_i > k \rrbracket \stackrel{(c)}{=} \frac{N_k}{\mathcal{I}} + 1 - \sum_{\kappa=1}^k \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \llbracket \ell_i = \kappa \rrbracket. \end{aligned}$$

(a) and (c) follow by algebra. (b) because if  $i > N_k$  and  $\ell_i \leq k$ , then  $i \notin F$ . Now  $\ell_i$  is the length of a contiguous block of 1's surrounded by zeros. Since  $\alpha$  is  $\lambda$ -random, by Lemma 1 the asymptotic proportion of such contiguous blocks of length  $\kappa$  is  $2^{-\kappa}$  by the following argument.

$$\begin{aligned} \lim_{\mathcal{I} \rightarrow \infty} \frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \llbracket \ell_i = \kappa \rrbracket &\stackrel{(a)}{=} \lim_{\mathcal{I} \rightarrow \infty} \frac{1}{\mathcal{I}} \#01^\kappa 0(\alpha_{1:n_{\mathcal{I}+1}}) \\ &\stackrel{(b)}{=} \lim_{\mathcal{I} \rightarrow \infty} \frac{(n_{\mathcal{I}} + 1)}{\#10(\alpha_{1:n_{\mathcal{I}+1}})} \cdot \frac{\#01^\kappa 0(\alpha_{1:n_{\mathcal{I}+1}})}{(n_{\mathcal{I}} + 1)} \stackrel{(c)}{=} 2^{-\kappa}, \end{aligned}$$

where (a) and (b) follow from the definitions of the intervals and (c) follows the definition of normal numbers and from part 1 of Lemma 1. Therefore  $\frac{1}{\mathcal{I}} \sum_{i=1}^{\mathcal{I}} \llbracket \ell_i > k \rrbracket < 2^{1-k}$  for sufficiently large  $\mathcal{I}$ . Sending  $k \rightarrow \infty$  gives  $d(F) := \lim_{\mathcal{I} \rightarrow \infty} \sum_{i=1}^{\mathcal{I}} \llbracket i \in F \rrbracket / \mathcal{I} = 0$ . It follows from  $d(B) = d(F) = 0$  and Lemma 1 that  $d(B \cup F) = 0$  and  $\lim_{n \rightarrow \infty} Km(\psi_{1:n})/n = 0$ . Since  $|Km(x) - K(x)| < O(\log \ell(x))$  for all  $x$  [5, §4.5.5],  $\lim_{n \rightarrow \infty} K(\psi_{1:n})/n = 0$  as

well. Let  $\theta_n := \#1(\chi_{1:n})/n$ . By Lemma 1 we have that  $\bar{d}(A - F) > 0$ . Therefore there exists a  $0 < c_3 \in \mathbb{Q}$  such that  $\limsup_{n \rightarrow \infty} \theta_n > \frac{1}{2} + c_3$ , where we also used the fact that  $\alpha_n = 1 \implies \chi_n = 1$  and  $\bar{d}(\{n_i : i \in \mathbb{N}\}) > 0$ . If  $\theta_n > \frac{1}{2} + c_3$  then by Lemma 1 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  $\epsilon > 0$  there exists an arbitrarily large  $n$  such that

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

This is a contradiction since  $H\left(\frac{1}{2} + c_3\right) < 1$ . Therefore  $\lim_{n \rightarrow \infty} M(\alpha_n | \alpha_{<n}) \neq \frac{1}{2}$  as required.  $\blacksquare$

## 5 Summary

We have shown that for every universal mixture there exists an infinite  $\lambda$ -random sequence on which it fails to converge.

**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

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

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 exist 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 hierarchy 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  $\alpha \in \mathcal{C}$  that is not  $\lambda$ -random. See Proposition 8 in the appendix, which is adapted from Theorem 7 in [3]. Note that it is known that there exists a lower semicomputable semimeasure  $W$  that converges on all  $\lambda$ -random sequences, but  $W$  is not universal [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.

## A Convergence on non-random sequences

**Proposition 8.** *There exists an  $\alpha \in \mathcal{B}^\infty$  such that*

1.  $\alpha$  is not  $\lambda$ -random.
2. For all universal lower semi-computable semimeasures  $M$

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

**Proof.** Define computable measure  $\nu$  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  $\mu$ -random sequences on which  $M$  converges to  $\mu$  by

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

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

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

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

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

as required. ■