On Martin-Löf Convergence of Solomonoff’s Measure

Tor Lattimore and Marcus Hutter
Sequence Prediction

Can you guess the next number?
1, 2, 3, 4, 5, ...
3, 1, 4, 1, 5, ...
1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, ...
Sequence Prediction

- $x$ is a binary sequence of length $\ell(x)$
- $\mu(x)$ is the $\mu$-probability of observing $x$
- $\mu(b|x)$ is the $\mu$-probability of observing $b \in \{0, 1\}$ given $x$
- $\mathcal{M}$ is a countable set of semimeasures (sequence generators)
- $\lambda(x) := 2^{-\ell(x)}$ is the Lebesgue measure

Goal: Construct predictor $\rho$ such that for all $\mu \in \mathcal{M}$

$$\rho(x_t|x_{<t}) \xrightarrow{fast} \mu(x_t|x_{<t})$$

when $x$ is sampled from $\mu$
<table>
<thead>
<tr>
<th>Name</th>
<th>Principle</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Epicurus</strong></td>
<td>Never discard a hypothesis consistent with the data</td>
<td></td>
</tr>
<tr>
<td><strong>Bayes</strong></td>
<td>Update your beliefs as data is observed</td>
<td></td>
</tr>
<tr>
<td><strong>Church-Turing</strong></td>
<td>The world is (stochastically) computable</td>
<td></td>
</tr>
<tr>
<td><strong>Occam</strong></td>
<td>The world is more likely to be simple</td>
<td></td>
</tr>
<tr>
<td><strong>Solomonoff</strong></td>
<td>Combine the above for induction/prediction</td>
<td></td>
</tr>
</tbody>
</table>

\[ M := \text{set of all lower semi-computable semimeasures} \]

\[ K(\nu) := \text{prefix Kolmogorov complexity of } \nu \]

\[ M(x) := \sum_{\nu \in M} 2^{-K(\nu)} \nu(x) \]
Universal Bayesian Sequence Prediction

- \( f \) is lower semi-computable (l.s.c.) if there exists \( f_t \uparrow f \)
- \( \mathcal{M} := \) set of all l.s.c. semimeasures
- \( w : \mathcal{M} \to (0, 1) \) is a l.s.c. distribution on \( \mathcal{M} \)
- \( M(x) := \sum_{\nu \in \mathcal{M}} w_{\nu} \nu(x) \) is a universal l.s.c. mixture semimeasure
- \( M(x) \geq w_{\nu} \nu(x) \) for all \( \nu \in \mathcal{M} \) (Universality)
- \( M \) is not unique – depends on choice of \( w \)

**Theorem (Solomonoff)**

*If \( \mu \in \mathcal{M} \), then*

(a) \( \lim_{t \to \infty} |\mu(x_t|x_{<t}) - M(x_t|x_{<t})| = 0 \), \( w.\mu.p.1 \)

(b) \( E_\mu \sum_{t=1}^{\infty} |\mu(x_t|x_{<t}) - M(x_t|x_{<t})|^2 \leq \ln \frac{1}{w_\mu} \)
Characterising Convergence to Lebesgue

- $\lambda(b|x) = \frac{1}{2}$ is the Lebesgue measure
- $\lim_{t \to \infty} |M(x_t|x_{<t}) - \lambda(x_t|x_{<t})| = 0$, w.l.p.1
- $C_M := \{ x \in B^\infty : \lim_{t \to \infty} M(x_t|x_{<t}) = \frac{1}{2} \}$
- **Goal of paper:** What does $C_M$ look like?
- $\lambda(C_M) = 1$ is immediate, but imprecise

**Intuition**

Expect sequences in $C_M$ to appear *typical* of sequences sampled from $\lambda$

<table>
<thead>
<tr>
<th>not</th>
<th>111111111111111111111111111111111111111111111111111111111111...</th>
</tr>
</thead>
<tbody>
<tr>
<td>maybe</td>
<td>1010000010111101111010001010001100001001...</td>
</tr>
</tbody>
</table>
Characterising Randomness

Intuition

$x$ is (Martin-Löf) random $\equiv x$ typical of sequences sampled from $\lambda$

not

\[ 11111111111111111111111111111111 \cdots \]

maybe

\[ 1010000010111101111010001010001100001001 \cdots \]

Equivalent definitions:

- $x$ passes all effective statistical tests
  - $x$ satisfies law of large numbers
  - $x$ satisfies law of iterated logarithm
  - $x$ is normal
  - $\ldots$
  - $x$ does not belong to an effective measure $0$ set
- $x$ is incompressible $K(x_{<n}) \geq n - c$ for all $n$
- $M(x_{<n}) \propto 2^{-n}$

$\mathcal{R} := \{ x : x \text{ is Martin-Löf random} \}$
Conjecture

\[ R \equiv \text{set of Martin-Löf random infinite sequences} \]
\[ C_M \equiv \text{set of infinite sequences on which } M \text{ converges to } \lambda \]

Conjecture

For all universal l.s.c. mixtures \( M \), \( R \subseteq C_M \)

Theorem (Hutter & Muchnik)

There exists a universal l.s.c. mixture \( M \) and \( x \in R \) such that

\[ \lim_{n \to \infty} M(x_n|x_{<n}) \neq \frac{1}{2} \]

\[ \implies R \not\subseteq C_M \]

Conjecture is false, but maybe the quantifier can be improved.
New Results

**Theorem**

For each universal l.s.c. mixture $M$ there exists an $x \in \mathcal{R}$ such that

\[
\lim_{n \to \infty} M(x_n | x < n) \neq \frac{1}{2}
\]

\[\implies \mathcal{R} \not\subseteq \mathcal{C}_M\]

**Theorem**

There exists an $x \notin \mathcal{R}$ such that for each universal l.s.c. mixture $M$

\[
\lim_{n \to \infty} M(x_n | x < n) = \frac{1}{2}
\]

\[\implies \bigcap_M \mathcal{C}_M \not\subseteq \mathcal{R}\]
Summary & History

- For each universal l.s.c. mixture $M$ there exists a Martin-Löf random real $x$ such that $M$ does not converge to $\lambda$ on $x$

\[ \forall M, \quad \mathcal{R} \not\subseteq C_M \]

- There exists a non-Martin-Löf random real $x$ such that every universal l.s.c. mixture $M$ converges to $\lambda$ on $x$

\[ \bigcap_{M} C_M \not\subseteq \mathcal{R} \]

- Miyabe, 2011: The 2-random reals are a subset of $C_M$ for all $M$

- Hutter & Muchnik, 2007: There exists a (non-universal) l.s.c. semimeasure $W$ such that

\[ \mathcal{R} \subseteq C_W \]

- Open questions:
  - What about randomness classes between Martin-Löf and 2-randomness?
  - Does there exist a universal l.s.c. semimeasure $M$ that is not a mixture such that $\mathcal{R} \subseteq C_M$?