On Martin-Löf Convergence of Solomonoff's Measure

Tor Lattimore and **Marcus Hutter**

Australian National University

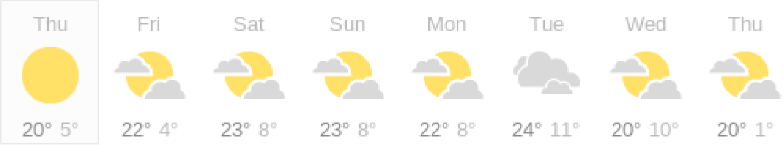
Sequence Prediction

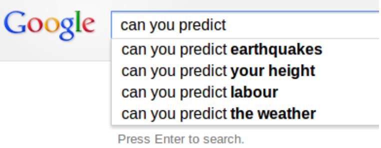
Can you guess the next number?

$1, 2, 3, 4, 5, \ldots$

$3, 1, 4, 1, 5, \ldots$

$1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, \ldots$





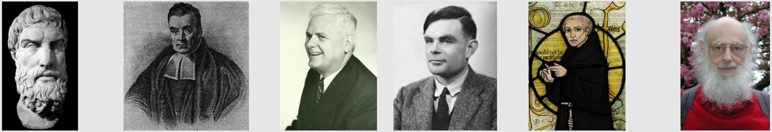
Sequence Prediction

- x is a binary sequence of length $\ell(x)$
- $\mu(x)$ is the μ-probability of observing x
- $\mu(b|x)$ is the μ-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 ρ 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 μ



Epicurus Bayes Church-Turing Occam Solomonoff

Epicurus	Never discard a hypothesis consistent with the data
Bayes	Update your beliefs as data is observed
Church-Turing	The world is (stochastically) computable
Occam	The world is more likely to be simple
Solomonoff	Combine the above for induction/prediction

$\mathcal{M}$:= set of all lower semi-computable semimeasures

$K(\nu)$:= prefix Kolmogorov complexity of ν

$$M(x) := \sum_{\nu \in \mathcal{M}} 2^{-K(\nu)} \nu(x)$$

Universal Bayesian Sequence Prediction

- f is lower semi-computable (l.s.c.) if there exists $f_t \nearrow f$
- $\mathcal{M} :=$ 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) $\lim_{t \to \infty} |\mu(x_t|x_{<t}) - M(x_t|x_{<t})| = 0, \quad w.\mu.p.1$

(b) $\mathbf{E}_\mu \sum_{t=1}^{\infty} |\mu(x_t|x_{<t}) - M(x_t|x_{<t})|^2 \leq \ln \frac{1}{w_\mu}$

Characterising Convergence to Lebesgue

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

Intuition

Expect sequences in $\mathcal{C}_M$ to appear *typical* of sequences sampled from λ

not $1111111111111111111111111111111111111111\cdots$

maybe $1010000010111101111010001010001100001001\cdots$

Characterising Randomness

Intuition

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

not $1111111111111111111111111111111111111111\cdots$

maybe $10100000101111011111010001010001100001001\cdots$

Equivalent definitions:

- x passes all effective statistical tests
  - x satisfies law of large numbers
  - x satisfies law of iterated logarithm
  - x is normal
  - $\cdots$
  - 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

$$\mathcal{R} \equiv \text{ set of Martin-Löf random infinite sequences}$$
$$\mathcal{C}_M \equiv \text{ set of infinite sequences on which } M \text{ converges to } \lambda$$

Conjecture

For all universal l.s.c. mixtures M, $\mathcal{R} \subseteq \mathcal{C}_M$

Theorem (Hutter & Muchnik)

There exists a universal l.s.c. mixture M and $x \in \mathcal{R}$ such that

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

$\implies \mathcal{R} \not\subseteq \mathcal{C}_M$

Conjecture is false, but maybe the quanitifer 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 λ on x

$$\forall M, \quad \mathcal{R} \nsubseteq \mathcal{C}_M$$

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

$$\bigcap_M \mathcal{C}_M \nsubseteq \mathcal{R}$$

- Miyabe, 2011: The 2-random reals are a subset of $\mathcal{C}_M$ for all M
- Hutter & Muchnik, 2007: There exists a (non-universal) l.s.c. semimeasure W such that

$$\mathcal{R} \subseteq \mathcal{C}_W$$

- 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 $\mathcal{R} \subseteq \mathcal{C}_M$?