An Open Problem Regarding the Convergence of Universal A Priori Probability

Marcus Hutter

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
IDSIA, Galleria 2, CH-6928 Manno-Lugano, Switzerland
marcus@idsia.ch, http://www.idsia.ch/~marcus

2003

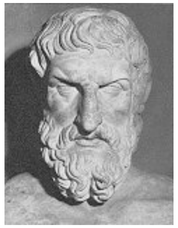
Abstract

Solomonoff unified Occam's razor and Epicurus' principle of multiple explanations to one elegant, formal, universal theory of inductive inference, which initiated the field of algorithmic information theory. His central result is that the posterior of his universal semimeasure M converges rapidly to the true sequence generating posterior μ, if the latter is computable. Hence, M is eligible as a universal predictor in case of unknown μ. Convergence holds in difference and in ratio, with probability 1 and in mean sum. Contrary to what was believed before, we show that the problem of individual convergence for all Martin-Löf random sequences is still open. The latter is particularly interesting and natural, since Martin-Löf randomness can be defined in terms of M itself.

Table of Contents

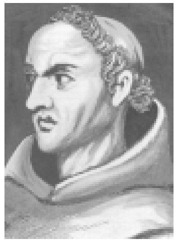
- Induction = Predicting the Future
- Solomonoff's Universal Prior M
- (Semi)measures & Universality
- Martin-Löf Randomness
- Convergence of Random Sequences
- Posterior Convergence
- Failed Attempts to Proof $M \xrightarrow{\text{M.L.}} \mu$
- Generalization & Conclusions & Prize

Induction = Predicting the Future



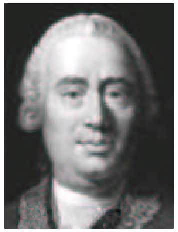
Epicurus' principle of multiple explanations

If more than one theory is consistent with the observations, keep all theories.



Ockhams' razor (simplicity) principle

Entities should not be multiplied beyond necessity.

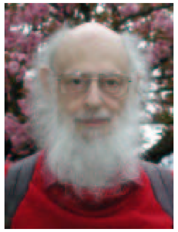


Hume's negation of Induction

The only form of induction possible is deduction as the conclusion is already logically contained in the start configuration.

Bayes' rule for conditional probabilities

Given the prior believe/probability one can predict all future probabilities.



Solomonoff's universal prior

Solves the question of how to choose the prior if nothing is known.

Solomonoff's Universal Prior M

Strings: $x = x_1 x_2 ... x_n$ with $x_t \in \{0,1\}$ and $x_{1:n} := x_1 x_2 ... x_{n-1} x_n$ and $x_{<n} := x_1 ... x_{n-1}$.

Probabilities: $\rho(x_1 ... x_n)$ is the probability that an (infinite) sequence starts with $x_1 ... x_n$.

Conditional probability: $\rho(x_t | x_{<t}) = \rho(x_{1:t}) / \rho(x_{<t})$ is the ρ-probability that a given string $x_1 ... x_{t-1}$ is followed by (continued with) x_t.

The universal prior $M(x)$ is defined as the probability that the output of a universal Turing machine starts with x when provided with fair coin flips on the input tape. Formally, M can be defined as

$$M(x) \;:=\; \sum_{p\,:\,U(p)=x*} 2^{-l(p)}$$

Semimeasures & Universality

Continuous (Semi)measures: $\mu(x) \overset{(\geq)}{=} \mu(x0) + \mu(x1)$ and $\mu(\varepsilon) \overset{(\leq)}{=} 1$.
$\mu(x) =$ probability that a sequence starts with string x.

Universality of M (Solomonoff:78): M is an enumerable semimeasure.
$M(x) \geq w_\rho \cdot \rho(x)$ with $w_\rho = 2^{-K(\rho)-O(1)}$ for all an enum. semimeas. ρ.

Explanation: Up to a multiplicative constant, M assigns higher probability to all x than any other computable probability distribution.

Martin-Löf Randomness

- Martin-Löf randomness is a very important concept of randomness of individual sequences.

- Characterization by Levin:73: Sequence $x_{1:\infty}$ is μ-Martin-Löf random (μ.M.L.) $\Leftrightarrow \exists c : M(x_{1:n}) \leq c \cdot \mu(x_{1:n}) \forall n$.

- A μ.M.L. random sequence $x_{1:\infty}$ passes all thinkable effective randomness tests, e.g. the law of large numbers, the law of the iterated logarithm, etc. Especially, the set of all μ.M.L. random sequences has μ-measure 1.

Convergence of Random Sequences

Let $z_1(\omega), z_2(\omega), ...$ be a sequence of real-valued random variables.
z_t is said to converge for $t \to \infty$ to random variable $z_*(\omega)$

$i)$ with probability 1 $(\mathbf{w.p.1})$ $:\Leftrightarrow \mathbf{P}[\{\omega : z_t \to z_*\}] = 1$,

$ii)$ in mean sum $(\mathbf{i.m.s.})$ $:\Leftrightarrow \sum_{t=1}^{\infty} \mathbf{E}[(z_t - z_*)^2] < \infty$,

$iii)$ for every μ-Martin-Löf random sequence $(\mu\mathbf{.M.L.})$ $:\Leftrightarrow$
$\forall\omega : [\exists c \forall n : M(\omega_{1:n}) \leq c \cdot \mu(\omega_{1:n})]$ implies $z_t(\omega) \stackrel{t\to\infty}{\longrightarrow} z_*(\omega)$,

where $\mathbf{E}[..]$ denotes the expectation and $\mathbf{P}[..]$ denotes the probability of $[..]$.

Remarks

(i) In statistics, convergence **w.p.1** is the "**default**" characterization of convergence of random sequences.

(ii) Convergence **i.m.s.** is **very strong**: it provides a rate of convergence in the sense that the expected number of times t in which z_t deviates more than ε from z_* is finite and bounded by $\sum_{t=1}^{\infty}\mathbf{E}[(z_t - z_*)^2]/\varepsilon^2$. Nothing can be said for **which** t these deviations occur.

(iii) **Martin-Löf**'s notion of randomness of **individual** sequences.

Convergence i.m.s. implies convergence w.p.1.

Convergence M.L. implies convergence w.p.1.

Posterior Convergence

Universality $M(x) \geq w_\mu \mu(x)$ implies the following posterior convergence results:

$$i) \quad \sum_{t=1}^n \mathbf{E} \sum_{x'_t} \left(\mu(x'_t|x_{<t}) - M(x'_t|x_{<t})\right)^2 \leq \ln w_\mu^{-1} < \infty$$

$M(x'_t|x_{<t}) \to \mu(x'_t|x_{<t})$ for any x'_t i.m.s. for $t \to \infty$.

$$ii) \quad \sum_{t=1}^n \mathbf{E} \left[\left(\sqrt{\frac{M(x_t|x_{<t})}{\mu(x_t|x_{<t})}} - 1 \right)^2 \right] \leq \ln w_\mu^{-1} < \infty$$

$\sqrt{\frac{M(x_t|x_{<t})}{\mu(x_t|x_{<t})}} \to 1$ i.m.s. for $t \to \infty$.

An interesting open question is whether M converges to μ (in difference or ratio) individually for all Martin-Löf random sequences.

Clearly, convergence μ.M.L. may at most fail for a set of sequences with μ-measure zero.

Failed Attempts to Proof $M \xrightarrow{\text{M.L.}} \mu$

- Conversion of bounds (i) or (ii) to effective μ.M.L. randomness tests fails, since they are not enumerable.

- The proof given in Vitanyi&Li:00 is incomplete. The implication "$M(x_{1:n}) \leq c \cdot \mu(x_{1:n}) \forall n \Rightarrow \lim_{n\to\infty} M(x_{1:n})/\mu(x_{1:n})$ exists" has been used, but not proven, and may indeed be wrong.

- Vovk:87 shows that for two finitely computable (semi)measures μ and ρ and $x_{1:\infty}$ being μ.M.L. random that

$$\sum_{t=1}^{\infty}\left(\sqrt{\mu(x_t|x_{<t})} - \sqrt{\rho(x_t|x_{<t})}\right)^2 < \infty \;\Leftrightarrow\; x_{1:\infty} \text{ is } \rho\text{.M.L. random.}$$

  If M were recursive, then this would imply $M \to \mu$ for every μ.M.L. random sequence $x_{1:\infty}$, since every sequence is M.M.L. random.

Difficulty of Proving $M \xrightarrow{\text{M.L.}} \mu$

$M \xrightarrow{\text{M.L.}} \mu$ cannot be decided from M being a mixture distribution or from dominance or enumerability alone. Further structural properties of M have to be employed.

Conclusions

Contrary to what was believed before, the question of posterior convergence $M/\mu \to 1$ (also $M \to \mu$) for all μ-random sequences is still open.

Prize

A prize (of probably grossly inappropriate magnitude) of 128 Euro for a solution of this problem is offered. [http://www.idsia.ch/~marcus]