An optimal superfarthingale and its convergence

Kenshi Miyabe

1 Dec 2011

Algorithmic Probability

Universal Prediction

"Gold standard" for prediction of a sequence from finite alphabet:

$$\xi(x) = \sum_{\nu \in \mathcal{M}} w_\nu \nu(x)$$

or

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

Many favorable properties!!
See Rathmanner & Hutter (2011) etc.

Some Problems

(i) Dependence on the universal machine

(ii) Inconsistency to the axioms of probability

(iii) Restriction of the space

(iv) Convergence in the sense of randomness

(v) ...

What is probability?

Question 1 (4a in Hutter (2009)). *Not perform well for short sequences because of the constant dependent on the universal machine.*

Is it really a problem?

Probability

> Probability is one way to express this confidence.
>
> By Solomonoff (2009)

For him, probability is subjective and nothing but

$$M(x_n|x_{<n}) = \frac{M(x_{1:n})}{M(x_{<n})}.$$

Subjectivity

Subjectivity

For quite some time I felt that the dependence of ALP (Algorithmic Probability) on the reference machine was a serious flaw in the concept, and I tried to find some "objective" universal device, free from the arbitrariness of choosing a particular universal machine. When I though I finally found a device of this sort, I realized that I really didn't want it - that I had no use for it at all!

By Solomonoff (2009)

I agree with him and this answers the question.

Solved?

(i)Dependence $\Rightarrow$ It is needed!

(ii)Inconsistency $\Rightarrow$

Subjective probability need not be additive!

(See Glenn Shafer 1978 and follow an argument by Beroulli.)

Other problems

Question 2. *Can one construct probability theory based on his interpretation and formulation?*

- Generalization of the underlying space
- Convergence in the sense of stronger randomness

Theorem 3 (Solomonoff 1978)**.**

$$\frac{M(x_n|x_{<n})}{\mu(x_n|x_{<n})} \to 1$$

as $n \to \infty$ with μ-probability 1.

An interpretation

"Probability" converges to the true "probability"?

Surprisingly we can not replace it with ML-randomness!

Theorem 4 (Hutter & Muchnik 2007)**.**
$\exists M$: an optimal semimeasure,
$\exists \mu$: a computable measure,
$\exists x$: a μ-ML-random sequence,

$$M(x_n|x_{<n}) - \mu(x_n|x_{<n}) \not\to 0$$

as $n \to \infty$.

We should ask WHY (related with 4i in Hutter (2009).)

Algorithmic Randomness

Cantor space

Convergence follows from optimality of the semimeasure.
On Cantor space, a seq. is Martin-Löf random iff no c.e. supermartingale d succeeds on it:

$$d(\sigma) \geq \frac{d(\sigma 0) + d(\sigma 1)}{2}, \quad \sup_{n \to \infty} d(x_{1:n}) = \infty.$$

Exists an optimal supermartingale d_0:

$$d(\sigma) \leq c \cdot d_0(\sigma).$$

We also have a characterization by integral tests.

Generalization

Randomness was generalized to a computable metric space.
Exists an optimal integral test by Hoyrup & Rojas.
Generalize it to any topological space by Weihrauch approach.

Theorem 5 (M.). *There exists an optimal μ-integral test on an SCT_3 space.*

T_3 is also called "regular".
The theorem does not hold if one drops the condition of SCT_3.
How about replacing with SCT_2 (Hausdorff)?

Game-theoretic probability

Superfarthingale

X: a topological space,
$\mathcal{M}$: a space of probability measures on X,
$\Pi = (\mathcal{M} \times X)^\infty$, $\Pi^\diamond = (\mathcal{M} \times X)^*$.

Definition 6 (Dawid & Vovk 1999). *A* superfartingale *is a function* $V : \Pi^\diamond \to [-\infty, \infty]$ *satisfying*

$$V(p_1, y_2, \ldots, p_{n-1}, y_{n-1}) \geq \int_X V(p_1, y_1, \ldots, p_n, x) dp_n.$$

Compare to

$$d(\sigma) \geq \frac{d(\sigma 0) + d(\sigma 1)}{2}.$$

Game-randomness

Definition 7 (essentially Vovk & Shen 2010). *A sequence* $\pi \in \Pi$ *is* game-random *if* $\sup_n V(\pi^n) < \infty$ *for all "c.e." superfarthingales* V*.*

They defined the "c.e." only on Cantor space but it can be generalized to a general space. Game-randomness is a generalization of Martin-Löf randomness.

Theorem 8 (M.). *If* $\mathbf{X}$ *is an* SCT_3 *space, then there is an optimal superfarthingale.*

Convergence

From the proof

Consider Cantor space.
A simpler proof of the convergence by Hutter & Muchnik 2007.
Hellinger distance

$$h_n(\nu, \mu|\omega_{<n}) = \sum_{a=0,1} \left(\sqrt{\nu(a|\omega_{<n})} - \sqrt{\mu(a|\omega_{<n})}\right)^2$$

We simply write $h_n = \sum_a (\sqrt{\nu_n} - \sqrt{\mu_n})^2$.
Let $N_n = \sum_a \sqrt{\nu_n \mu_n}$ then

$$N_n \leq \exp(-h_n/2).$$

μ-supermartingale

Let

$$d(\sigma) = \sqrt{\frac{\nu(\sigma)}{\mu(\sigma)}} \exp\left(\frac{1}{2}\sum_{i=1}^{n} h_i\right).$$

Then

$$\frac{\sum_a \mu(\sigma a) d(\sigma a)}{\mu(\sigma) d(\sigma)} = N_n \exp\left(\frac{h_n}{2}\right) \leq 1.$$

So d is a μ-supermartingale.
Then d is not c.e.!! because

$$d < \infty \Rightarrow h_n \to 0.$$

Main result

V: optimal c.e. superfarthingale

$\overline{V}(\pi^n) = \frac{V(\pi^n)}{V(\pi^{n-1})}$

μ_n: $\mu_n(A) = \int_A \overline{V}(\pi^{n-1}, p_n, x) dp_n$

W: non-negative c.e. farthingale

ν_n: $\nu_n(A) = \int_A \overline{W}(\pi^{n-1}, p_n, x) dp_n$

Theorem 9 (M.). *The Hellinger distance between μ_n and ν_n converges to 0 a.s. w.r.t. $\prod \mu_n$.*

Stronger randomness

Definition 10. *μ-2-ML-random iff $d(\sigma) < \infty$ for all c.e. μ-supermartingales relative to the halting problem.*

Proposition 11 (M.)**.** *For a μ-2-ML-random sequence α,*

$$\frac{\mu_0(\alpha_n|\alpha_{<n})}{\mu(\alpha_n|\alpha_{<n})} \to 1$$

as $n \to \infty$.

An interpretation

An interpretation

"Probability" converges to a measure for which the sequence is random.

Another probability is not needed anymore!

Remark 12. *Such a measure may not exist.*

Summary

What we did

- Generalized to sequences of points in a SCT_3.
- Proved convergence for a stronger random sequence.
  $\Rightarrow$ The fact of non-c.e. helps to solve (4i) in Hutter (2009)?
- Gave an interpretation of probability without another probability.

Future work

- How about difference randomness?
- Can we further generalize to a non-sequence space?
- Can we directly define an optimal semimeasure?

Thank you!