An optimal semimeasure and its convergence

Kenshi Miyabe

Aug 23-24, 2011

What is prediction?

Prediction?

- Given an unknown object,
- assume it behaves probabilistically,
- assume the probability is contained in a family,
- given a loss function, a utility function or a risk function,
- select a parameter in the family,
- discuss the quality of the prediction probabilistically.

Statistical decision problem

According to A. Wald $\cdots$

- *parameter space* $\{P_\theta\}$ of probabilistic measures
- *loss function* or *risk function*
- *decision function*

An appropriate assumption $\Rightarrow$ non-parametric?
Probability is used to predict and to evaluate.
Another way?
Not best prediction $\Rightarrow$ relativism?

Solomonoff's idea

Solomonoff's idea

Predictable = Compressible

- relation to Occam's razor (1285 - 1349) and Epicurus principle (342? B.C. - 270 B.C.)
- existence of an optimal prediction
- convergence to the true probability

Critics

- dependence on an optimal prediction for short sequences
- extra information
- non-computability
- restriction of the setting (*)
- interpretation of probability (**)
- more

(*) Partially solved in my paper
(**) The main theme of this talk

AIT

Semimeasure

semimeasure $\mu : 2^* \to \mathbb{R}$ s.t.

$$\mu(\lambda) \leq 1 \text{ and } \mu(\sigma) \geq \mu(\sigma 0) + \mu(\sigma 1)$$

probability measure when $=$
computable and c.e. if so is $\mu(\sigma)$ respectively
optimal if $\mu_0(\sigma) \geq c \cdot \mu(\sigma)$

$$\mu(\sigma_n | \sigma_{<n}) = \frac{\mu(\sigma_{1:n})}{\mu(\sigma_{<n})}$$

Convergence

μ_0: optimal c.e. semimeasure
μ: computable measure

Theorem 1 (Solomonoff 1978)**.**

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

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

μ_0 can completely find the regularity!

Probability

An interpretation

"Probability" converges to the true "probability".

What does it mean?
Evaluation needs probability?
How about Martin-Löf random sequences?

ML-randomness

μ-supermartingale $d : 2^* \to \mathbb{R}^+$ s.t.

$$\mu(\sigma)d(\sigma) \geq \mu(\sigma 0)d(\sigma 0) + \mu(\sigma 1)d(\sigma 1)$$

Definition 2. *μ-Martin-Löf random iff $d(\sigma_{1:n}) < \infty$ for all c.e. μ-supermartingales.*

Proposition 3 (Levin 73)**.** *M: optimal semimeasure μ-Martin-Löf random iff $M(\sigma_{1:n}) \leq c \cdot \mu(\sigma_{1:n})$*

Non-convergence

Surprisingly we can not replace it with ML-randomness!

Theorem 4 (Hutter & Muchnik 2007)**.** *There exists an optimal semimeasure M, a computable measure μ and a μ-ML-random sequence α s.t.*

$$M(\alpha_n|\alpha_{<n}) - \mu(\alpha_n|\alpha_{<n}) \not\to 0$$

as $n \to \infty$.

Question 5. *Does there exist another optimal semimeasure that satisfies the property?*

More importantly we should ask WHY.

Convergence

From the proof

A simpler proof of the convergence by Hutter and Muchnik 2007.
Hellinger distance

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

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.
Note that

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

Compare

In probability theory,

$$d < \infty \text{ with probability } 1.$$

For Martin-Löf randomness,

$$d \text{ is c.e. } \Rightarrow d < \infty.$$

So d is not c.e.!

Stronger randomness

Definition 6. *μ-2-ML-random iff $d(\sigma) < \infty$ for all c.e. μ-supermartingales relative to the halting problem.*

Proposition 7. *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 8. *Such a measure may not exist.*

Summary

What we did

- Proved convergence for a stronger random sequence.
- Generalized to sequences of points in a SCT_3.
- Gave an interpretation without probability.

Future work

- When does the sequence converge?
- Can we define a semimeasure on a general space?
- Can we characterize optimality by a loss function?
- Do we have an optimal decision making or hypothesis testing?
- Does it have a practical application?

Thank you!

If time is left

What is probability?

Two main interpretations

Objective probability (Aleatory probability)

- frequency
- propensity

Subjective probability (Epistemic probability)

- the degree of belief

Arnauld (1612 - 1694) first identified them!

Algorithmic probability

Objective probability because determined?
Subjective probability for a short sequence? (by Solomonoff 2009)
Epistemic!
Also adapted to the prequential approach.

Kolmogorov's axioms

Why insufficient?

- optimality?
- buy and sell
- semimeasure?
- game-theoretic?