Towards prediction via randomness

Kenshi Miyabe

Feb 22, 2011

AIT

Table of contents

Algorithmic randomness is highly related with probability theory, statistics, dynamical system and other many areas.

- Recent development of randomness enables us to apply.
- Some researchers come to think that we need to know what randomness means.

In this talk we discuss the following.

- What is algorithmic randomness?
- Some recent development and my contribution.
- My research plan to a prediction theory.

The topics in AIT

Algorithmic Information Theory is the result of putting Shannon's information theory and Turing's computability theory into a cocktail shaker and shaking vigorously. (G. J. Chaitin)

AC Algorithmic "Kolmogorov" Complexity

AP Algorithmic "Solomonoff" Probability

($\longrightarrow$ prediction or machine learning)

US Universal "Levin" Search

AR Algorithmic "Martin-Löf" Randomness

($\longrightarrow$ probability theory or statistical test)

From frequency to regularity

Casting a dice $\rightarrow$ Probability is just frequency.
The sunrise problem $\rightarrow$ Frequency is not appropriate.

- Even if you know the sun has risen for n days, we believe it more than $\frac{n}{n+1}$.
- If the sun did not rise today, the belief of the sunrise tomorrow would be much less than $\frac{n}{n+1}$.

Is there an appropriate initial distribution for Bayes rule?

Solomonoff's induction

Theorem 1 (Solomonoff '64)**.** *There is an optimal prediction $\mathcal{M}$.*

- The space is Σ^ω.
- Optimal in c.e. and up to a multiplicative constant
- Highly related with Kolmogorov complexity
- A good compression means a good prediction.

What does randomness mean?

Intuitively TFAE:

- Not random
- Has regularity
- Compressible
- Predictable

Randomness by martingale

Three approaches

Many notions of randomness has been proposed.
Martin-Löf Randomness has been regarded as the most natural one.
Randomness has (at least) three approaches.

- Typicality by a test concept (Martin-Löf 1966)
- Unpredictability by martingales (Schnorr 1971)
- Uncompressibility by Kolmogorov complexity (Levin, Schnorr 1973)

Most notions of randomness also have similar characterizations.

Computability

A set $A \subseteq \mathbb{N}$ is *computable* iff $1_A : \mathbb{N} \to \{0,1\}$ is computable.
A set A is *c.e.* iff $A = \mathrm{dom}(f)$ for a partial computable function.
A real $r \in [0,1]$ is *computable* iff its binary expantion is computable.
A real r is *c.e.* iff $\{q \in \mathbb{Q} \ : \ q < r\}$ is c.e.
There exist a c.e. set and a c.e. real that are not computable.
2^ω: Cantor space of infinite binary sequences
μ: generated from $\mu([\sigma]) = 2^{-|\sigma|}$
We often identify 2^ω with $[0,1]$.

Martingales in randomness

Definition 2. *A function* $d : 2^* \to \mathbb{R}^+ \cup \{0\}$ *is a* martingale *if*

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

The d *is a* supermartingale *if we replace* $=$ *by* $\geq$*.*

Proposition 3. *TFAE:*

- $A \in 2^\omega$ *is ML-random.*
- $\sup_n d(A \upharpoonright n) < \infty$ *for each c.e. martingale.*
- $\sup_n d(A \upharpoonright n) < \infty$ *for each c.e. supermartingale.*

The set of ML-random sequences has measure 1.

Universal martingale

Definition 4. *The* success *set of* d *is*

$$S[d] = \{A \ : \ d(A) = \sup_n d(A \upharpoonright n) = \infty\}.$$

A c.e. (super)martingale d *is* universal *if*

$$S[d] \supseteq S[f]$$

for each c.e. (super)martingale f.

Proposition 5. *There exists a universal c.e. (super)martingale.*

Optimal martingale

Definition 6. *A c.e. (super)martingale* d *is* optimal *if for each c.e. (super)martingale* f *there exists* c *s.t.*

$$f(\sigma) \leq c \cdot d(\sigma)$$

for all σ*.*

Proposition 7. *No c.e. martingale is optimal. There exists an optimal c.e. supermartingale.*

Remark 8. *This optimal c.e. supermartingale is nothing but an optimal prediction.*
Note that it is optimal up to a multicative constant.
There are (at least) two constructions: by a direct enumeration and by complexity.

General randomness

Computable topological space

Definition 9 (Hertling and Weihrauch 2009)**.** *A* computable topological space *or* CTS *is a 4-tuple* $\mathbf{X} = (X, \tau, \beta, \nu)$ *such that*

- (X, τ) *is a topological* T_0*-space,*
- $\nu :\subseteq \Sigma^* \to \beta$ *is a notation of a base* β *of* τ*,*
- *dom*(ν) *is recursive and*
- $\nu(u) \cap \nu(v) = \bigcup\{\nu(w) \ : \ (u, v, w) \in S\}$ *for all* $u, v \in$ *dom*(ν) *for some r.e. set* $S \subseteq ($*dom*$(\nu))^3$*.*

Computability

Define a representation $\delta :\subseteq \Sigma^\omega \to X$ of the points as

$$x = \delta(p) \iff (\forall w \in \Sigma^*)(w \ll p \iff x \in \nu(w)).$$

A point is δ-computable if it has a computable δ-representation.
Computability of sets and functions are similarly defined.

Computability of measures

$M(X)$: the space of bounded non-negative Borel measures

Proposition 10. *G: the finite union of base sets*
A countable subbase of the A-topology τ_A:

$$\{\mu \ : \ \mu(G) > q\}, \ \{\mu \ : \ \mu(X) < q\}$$

Theorem 11. $(M(X), \tau_A, \beta_A, \nu_A)$ *is a CTS.*

Definition 12. *A measure is* computable *if it is a computable point in the CTS.*

Measure randomness

Let $(X, \mathcal{A}, \mu)$ be a measure space.
A *filtration* is a sequence of sub-σ-algebra $(\mathcal{A}_n)$ such that $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ for each n.
A sequence of $\mathcal{A}$-measurable functions $(f_n, \mathcal{A}_n)$ is called a *supermartingale* if $\int f_n d\mu < \infty$ and $\int_A f_n d\mu \geq \int_A f_{n+1} d\mu$ for all $A \in \mathcal{A}_n$.

Theorem 13. *A point x is measure μ-random iff $\sup_n f_n(x) < \infty$ for each $(\nu_{\mathbb{N}}, \delta, \overline{\rho}_<)$-computable supermartingale $(f_n, \mathcal{A}_n)$*

Computably Regular

Definition 14 (regular). *A non-negative measure μ is* regular *if $\forall A \in \mathcal{A}$ and $\forall \epsilon > 0$, there exists a closed set F_ϵ s.t. $F_\epsilon \subset A$, $A \backslash F_\epsilon \in \mathcal{A}$ and $\mu(A \backslash F_\epsilon) < \epsilon$.*

Definition 15 (computably regular). *A non-negative measure μ is* computably regular *if for all u there exists v_n and p_n uniformly in u and n s.t.*

- $\mu(\bigcup_n \nu(v_n)) \to \mu(\nu(u))$,
- $\nu(v_n) \subseteq \psi^-(p_n) \subseteq \nu(u)$,
- $\mu(\psi^-(p_n) \backslash \nu(v_n)) \leq 2^{-n}$.

Proposition 16. *If the measure is computably regular, there exists a universal $(\nu_\mathbb{N}, \delta, \overline{\rho}_<)$-comp. martingale.*

Optimality?

Question 17. *Does there exist an optimal* $(\nu_{\mathbb{N}}, \delta, \overline{\rho}_<)$*-computable supermartingale?*

Application

SLLN for randomness

Theorem 18. *Let μ_1 be a continuous computable measure on R such that $m = \int |z| d\mu_1 < \infty$ is computable.Let $\mu = \prod \mu_1$ be the products of infinite same measures.Then*

$$\{y_i\} \text{ is } \mu\text{-random} \Rightarrow \frac{\sum_{i=1}^n y_i}{n} \to \int x d\mu.$$

Theorem 19. *Let μ_1 be a continuous computable measure on R such that $m = \int |z| d\mu_1 = \infty$.Let $\mu = \prod \mu_1$ be the products of infinite same measures.Then*

$$\{y_i\} \text{ is } \mu\text{-random} \Rightarrow \limsup_n \frac{|\sum_{i=1}^n y_i|}{n} = \infty.$$

Hartman-Wintner LIL for randomness

Theorem 20 (Hartman-Wintner LIL). *Let $\mu = \prod \mu_1$ be a computable measure on $\mathbf{R}^{\mathbb{N}}$ such that μ_1 is continous, $\int z\mu(dz) = 0$ and $\int z^2\mu(dz) = \sigma^2$ is computable. Then*

$$\{x_i\} \text{ is } \mu\text{-random} \Rightarrow \limsup_n \frac{\sum x_i}{\sqrt{2n \ln \ln n}} = \sigma.$$

Convergence speed

Let $K^c = \{A \ : \ (\forall n) K(A \restriction n) > n - c\}$.

Theorem 21 (Davie '01). *There exists a computable function* $n(c,\epsilon)$ *s.t. if* $A \in K^c$ *then for all* $n > n(c,\epsilon)$*,*

$$\left| \frac{S_n(A)}{n} - \frac{1}{2} \right| < \epsilon.$$

For c *and* $\lambda > 1$*, there exists a computable function* $n(c,\lambda)$ *s.t. if* $A \in K^c$ *then for all* $n > n(c,\lambda)$*,*

$$S_n(A) \le \frac{n}{2} + \lambda \sqrt{\frac{n}{2} \log \log n}.$$

Effective Birkhoff's ergodic theorem

Definition 22. *T is $\ln^2$-ergodic if $\forall f, g$*

$$\gamma_n(f,g) = \left|\frac{1}{n}\sum_{i<n}\int f\circ T^i \cdot g d\mu - \int f d\mu \int g d\mu\right| \leq \frac{c_{f,g}}{(\ln(n))^2}$$

Theorem 23 (Birkhoff's ergodic theorem for random points; V'yugin '97, Nandakumar '08, Hoyrup and Rojas '09)**.**
(X,μ): a CPS
$T: X \to X$ an effectively measurable measure-preserving map
$f \in L^1(X.\mu)$ an effectively measurable function
T: $\ln^2$-ergodic
Then $\exists n(c,\epsilon)$ comp. function s.t.
if $x \in K_c$ then for all $n > n(c,\epsilon)$,
$\left|\frac{1}{n}\sum_{i<n} f\circ T^i \cdot g d\mu - \int f d\mu \int g d\mu\right| < \epsilon$

Probability and prediction

What is probability?

Three main interpretations of probability:

- Frequency probability $\longrightarrow$ von Mises
- Subjective probability $\longrightarrow$ some researchers
- Propensity probability $\longrightarrow$ Kolmogorov

Optimal prediction

X: a computable topological space with computable regularity

Question 24. *Does there exist an optimal $(\nu_{\mathbb{N}}, \delta, \overline{\rho}_{<})$-computable supermartingale on $X^{\mathbb{N}}$?*

There are (at least) two approaches:

- Define monotone complexity Km on a general space.
- Enumerate all such supermartingales.

The latter approach is more simple.

Radon-Nikodym operator

However there are some difficulties.

Theorem 25 (Weihrauch)**.** *Radon-Nikodym operator on* $[0, 1]$ *is not computable.*

Then how about using the setting of game-theoretic probability?

Game-theoretic probability

Parameters: Measurable space $(X, \mathcal{F}),\ K_0 \in \mathbb{R}$
Players: Forecaster, Skeptic, Reality
Protocol:
FOR $n = 1, 2, \ldots$:
Forecaster announces p_n on $(\Omega, \mathcal{F})$.
Skeptic announces f_n on Ω
Reality announces $x_n \in \Omega$.
$K_n := K_{n-1} + f_n(x_n) - \int_\Omega f_n dp_n$.
Effectivezed game-theoretic probability is essentially the same as generalized randomness.

End

Thank you!