From randomness to prediction and probability

Seminar at University ot Tokyo
12 June 2013

Kenshi Miyabe
JSPS Research Fellow at The University of Tokyo

Overview

- Biography
- An introduction to algorithmic randomness
- History of randomness and probability
- No computability, no probability

Biography

Biography

- 宮部賢志（Kenshi Miyabe）
- Born in 岐阜（Gifu）in 1982
- Bachelor and Master degrees from Kyoto Univ.
- D.Sc. from Kyoto University, 2010
- Researcher for three years at Kyoto University
- JSPS fellow at Univ. of Tokyo from 2013

An introduction to algorithmic randomness

Math. in Science department

- Algebra（代数）
- Geometry（幾何）
- Analysis（解析）
- Foundation of math.（数学基礎論）

Foundation of math.

- Proof theory（証明論）
- Set theory（集合論）
- Model theory（モデル論）
- Computability theory（計算論）

Computability theory

- Degree theory
- Reverse mathematics
- Algorithmic randomness (Here!!)
- Computable model theory
- Generalized computability

Theoretical Computer science

- Theory of computation
  - automata, computability, complexity, cryptography, quantum
- Information and coding
- Algorithm
- Programming
- Formal methods

Applications

- Probability theory
- Universal induction
- Ergodic theory
- Quantum mechanics
- Computable measure theory
- Randomized algorithm

Problem 1

Given a class $\mathcal{C}$ of measure on 2^ω and a point $x \in 2^\omega$.

Find a measure μ in $\mathcal{C}$ such that x is random with respect to μ.

The function $x \mapsto \mu$ should be computable in a sense.

Exercise 1

Consider the fair-coin tossing game. Given $x \in 2^\omega$.

Find a Bernoulli measure μ such that x is Martin-Löf random with respect to μ.

The function $x \mapsto \mu$ is not computable in the sense of TTE theory but is layerwise computable.

Problem 2

Given a class $\mathcal{C}$ of measures.

Find a betting strategy f such that

1. f is computable in a sense,
2. for any $\mu \in \mathcal{C}$, if $x \in 2^\omega$ is random w.r.t. μ, then the conditional measure induced from f converges to the conditional measure induced from μ along x.

How about the convergence speed?

Problem 3

- It seems that we can do statistics without using the notion of probability.
- Then what is probability in this context?
- How does it relate to computability and prediction?
- Does this view contribute to statistics technically?

How to define randomness

- Statistical approach
- Gambling approach
- Computational approach

Martin-Löf's idea 1966

If a point is random, it is not contained in a small class.

This is false because each point $x \in 2^\omega$ is contained in a set $\{x\}$ with Lebesgue measure 0.

Then, we say that a point is random if it is not contained in a null set effectively represented.

Martin-Löf's randomness

Definition (Martin-Löf 1966)

A Martin-Löf test (ML-test) is a sequence of uniformly c.e. open sets $\{U_n\}$ such that $\mu(U_n) \leq 2^{-n}$ for all n.

We say that a set $A \in 2^\omega$ passes a ML-test $\{U_n\}$ if $A \notin \bigcap_n U_n$.

A set $A \in 2^\omega$ is called ML-random if it passes all ML-tests.

Basic observations

Proposition

The class of ML-random sets has measure 1.

There is a universal ML-test.

No computable sequence is ML-random. .

Gambling approach

Definition

A martingale is a function $d : 2^{<\omega} \to \mathbb{R}^+$ satisfying the average property:

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

for each $\sigma \in 2^{<\omega}$.

Theorem (Schnorr 1971)

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

Computational approach

Definition

$$K(\sigma) = \min\{|\tau| \ : \ U(\tau) = \sigma\}$$

Theorem (Levin 1973, Schnorr 1973)

A set A is ML-random iff $K(A \upharpoonright n) > n - O(1)$.

A generalization

Actually, we can define ML-randomness any computable topological space with a measure μ (Gács 2005, Hoyrup-Rojas 2009, M.).

Such a point is called μ-ML-random or ML-random with respect to μ.

Exercise 1

Consider the fair-coin tossing game. Given $x \in 2^\omega$.

Find a Bernoulli measure μ such that x is Martin-Löf random with respect to μ.

The function $x \mapsto \mu$ is not computable in the sense of TTE theory but is layerwise computable.

Randomness deficiency

1. $t(x) = \sup\{n \ : \ x \notin U_n\}$.
2. $u : 2^\omega \to \overline{\mathbb{R}}^+$ is an integral test.
3. $\mathrm{CPS}(x) = \sup_n 2^{n-K(x\upharpoonright n)}$.
4. $\mathrm{JM}(x) = \sum_n 2^{n-K(x\upharpoonright n)}$.
5. $d(x) = \sup_n d(x \upharpoonright n)$.

Then, x is ML-random iff one of these is finite. However, they vary more than a constant.

Similar to "significance level" in statistics.

Open question

Which is the natural one? This problem is closely related to the following open question:

Question (Levin?)
How should we define the mutual information of two infinite sequences?

Layerwise computability

(Hoyrup and Rojas 2009)

A function $f :\subseteq X \to Y$ is layerwise computable if $f(x)$ is computable from x and its randomness deficiency.

Learnability

Bienvenu and Monin (2012) gave a weak condition of that the function is layerwise computable.

Thus, they gave an answer to a formulation of the problem 1 .

Connected to

- Parametrized complexity
- Learnability with discrete advice
- Randomized algorithm

History of randomness and probability

How about Problem 2?

- Before giving an answer to Problem 2, we see the history of randomness and probability.

Hilbert (1862-1943)

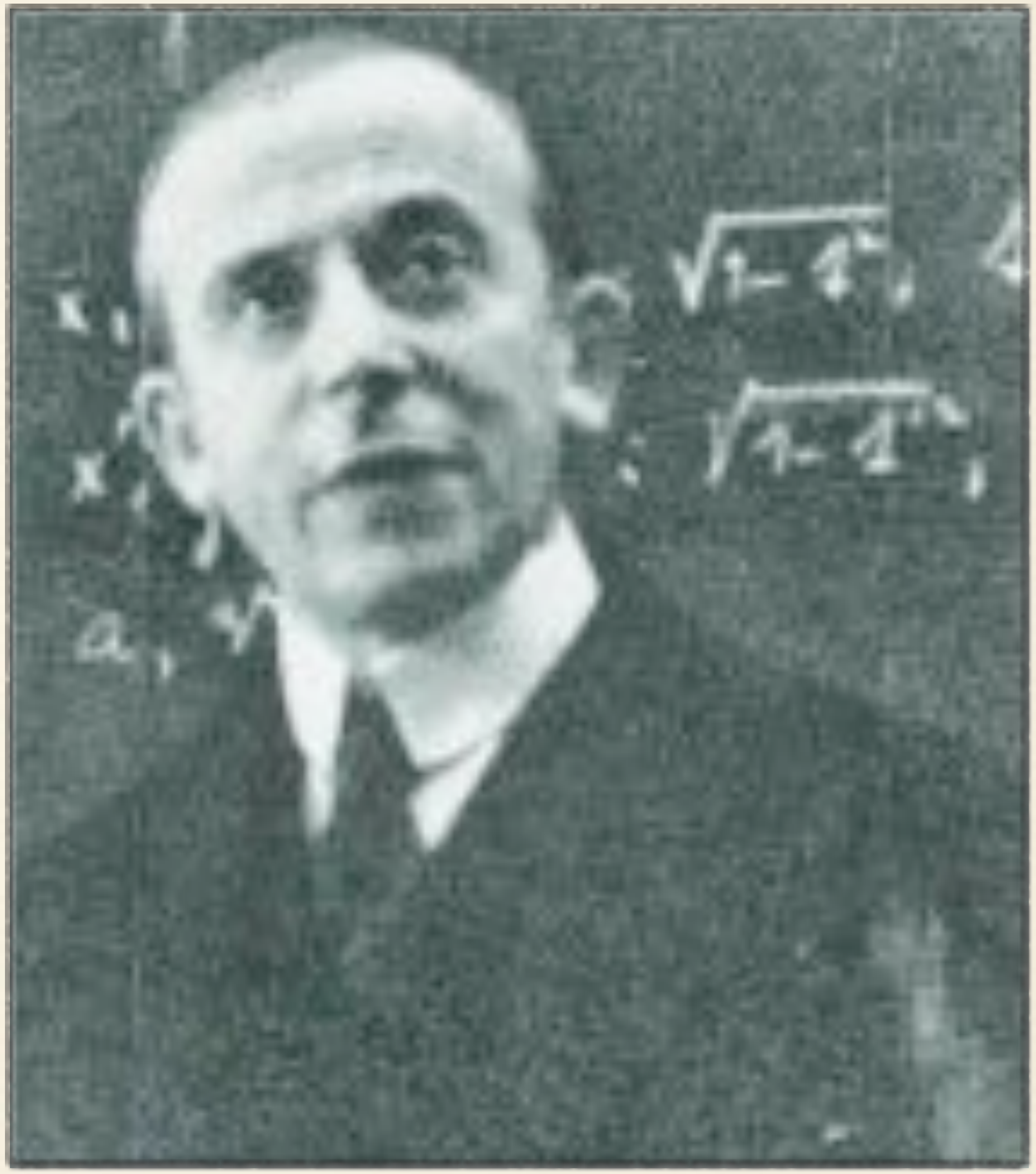
- Hilbert's 23 problmes on ICM
- 6th "Mathematical treatment of the axioms of physics"

Kolmogorov (1903-1987)

- Probability axioms (1933)
- Frequentist

von Mises (1883-1953)

- von Mises (1883-1953)
- Frequency approach (1919-)
- "First the Collective - then the Probability."



Kolmogorov 1963

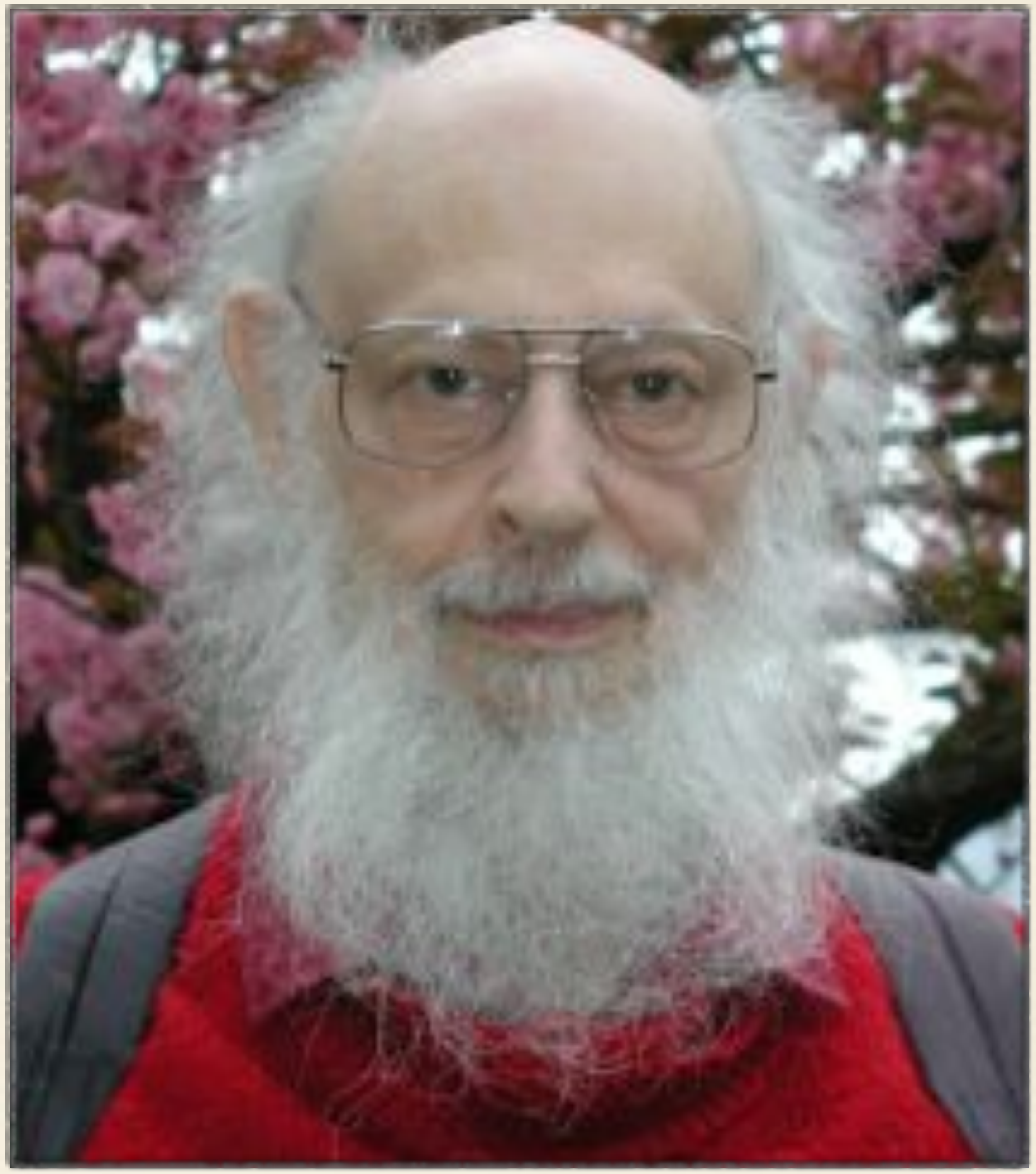
I have already expressed the point of view that the basis of the applicability of the mathematical theory of probability to random events of the real world is the frequency approach to probability in one form or another, which was so strongly advocated by von Mises.

Weekness of von Mises's theory

- The axiomatic approach is much simpler
- The notion of collectives is not the proper notion of randomness (Ville 1939)
- How to know the limit?

Solomonoff

- Solomonoff (1926-2009)
- algorithmic probability, algorithmic information theory
- artificial intelligence based on machine learning



- Typicalness - statistics
- Unpredictability - game-theoretic probability
- Incompressibility
  - algorithmic probability,
  - Minimal Descriptive Length,
  - Information Criterion ?

Problem 2

Given a class $\mathcal{C}$ of measures.

Find a betting strategy f such that

1. f is computable in a sense,
2. for any $\mu \in \mathcal{C}$, if $x \in 2^\omega$ is random w.r.t. μ, then the conditional measure induced from f converges to the conditional measure induced from μ along x.

How about the convergence speed?

Solomonoff's Universal induction

Theorem (Solomonoff 1978)

Let μ be a computable measure and M be an optimal semimeasure. Then,

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

with probability 1.

Convergence fails for a random set

Theorem(Hutter and Muchnik 2007)

There are a computable measure μ, an optimal semimeasure M and a μ-ML-random sequence such that

$$M(x_n|x_{<n}) - \mu(x_n|x_{<n}) \not\to 1.$$

- 2-randomness
- weak 2-randomness
- weak Demuth randomness
- Difference randomness
- ML-randomness
- Schnorr randomness
- Kurtz randomness

Theorem (Many)

The following are equivalent for a ML-random set z:

1. z is a density-one point for each Π^0_1-class containing z.
2. z is a convergence point for each c.e. martingale.
3. $f'(z)$ exists for each interval-c.e. function f.
4. z is Madison random.

Question

Is Madison randomness sufficient for the convergence?

Theorem (Pathak, Rojas and Simpson)

The following are equivalent for z:

1. z is Schnorr random.
2. z is a Lebesgue point for each L^1-computable function.

Question

Is Schnorr randomness sufficient for the convergence for the measure with L^1-computable density?

Solomonoff called the ratio $M(x_n|x_{<n})$ algorithmic probability because of universality, and he tried to convince that this is the objective probability.

Just before he passed away in 2009, he changed his mind and wrote the following.

Subjective probability

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!

Solomonoff (2009)

Subjectivity in science has usually been regarded as Evil. - that it is something that does not occur in “true science” - that if it does occur, the results are not “science” at all. The great statistician, R. A. Fisher, was of this opinion. He wanted to make statistics “a true science” free of the subjectivity that had been so much a part of its history.
I feel that Fisher was seriously wrong in this matter, and that his work in this area has profoundly damaged the understanding of statistics in the scientific community - damage from which it is recovering all too slowly.

Solomonoff (2011)

No computability, no probability

Problem 3

- It seems that we can do statistics without using the notion of probability.
- Then what is probability in this context?
- How does it relate to computability and prediction?
- Does this view contribute to statistics technically?

- Downey & Hirshfeldt "Algorithmic Randomness and Complexity" (2010)
- Nies "Computability and Randomness" (2009)
- Li & Vitányi "An Introduction to Kolmogorov Complexity and Its Applications" 3rd (2009)
- Hutter "Universal Artificial Intelligence" (2004)
- Shoenfield "Recursion theory" (1993)
- Weihrauch "Computable Analysis" (2001)