Algorithmic randomness

Kenshi Miyabe

June 16, 2010

ランダムを定義する

アルゴリズム的ランダムネスとは

アルゴリズム的ランダムネス (Algorithmic Randomness) とは、アルゴリズム情報理論 (Algorithmic Information Theory) とも呼ばれ、ランダムな二進無限列を計算の側面から具体的に定義し、その性質を調べる分野である。
G. Chaitin によれば、

> Shannon の情報理論と Turing の計算複雑性理論をシェイカーに入れて、力一杯シェイクしてできたもの。

- 具体的な列について議論できることが確率論と異なる。
- ランダムとは何かについて答える。
- ランダムは計算可能性に応じて定義される。
- 計算可能性との深い関連がある。

ランダムな列とは

ランダムな二進無限列を定義したい。

- $000000000000000000000000000000000000000000000\cdots$
- $010101010101010101010101010101010101010101010\cdots$
- $110111001110111101000000110101110000001111011\cdots$

少なくとも正規数 (0 と 1 が同じくらいの頻度で出てくる) だろう。
最後の列は円周率の十進展開の各桁が奇数か偶数かによって決まる列。
このような計算可能な列はランダムとは言わない。
計算可能な列は可算濃度、二進無限列は連続体濃度だから、すべてがランダムではない。
e や π の二進展開はランダムではないし、疑似乱数もランダムではない。

von Mises の collective

1919 年、von Mises は確率論の定式化を目指して、ランダムな列について考察を始めた。
ある選択規則による定まるすべての部分列に 1 が出てくる割合の極限が $1/2$ となる列を collective と呼び、ランダムな列だと思おうという提案。
von Mises は「選択規則」について厳密な定義を与えなかった。
その後計算可能性が定義され、Church により選択規則が計算可能な関数として厳密な定義が与えられることが指摘された。
Ville により重複対数の法則が満たされないことが示された。
ランダムな列を定義することは簡単ではなかった。

計算可能性

ランダムネスは計算可能性の理論 (Computability theory) を基礎とする。
二進無限列、実数、自然数の集合をそれぞれ同一視する。
自然数の集合が 計算可能 とはその特徴関数が計算可能であること。
集合が 計算可枚挙 または c.e. であるとは、含まれる元を値域にもつ計算可能な関数 Φ があることを言う。
ある実数 r が 計算可能 であるとは任意の精度で計算可能なこと。
つまり、$\Phi(m) = r \upharpoonright m$ となる Φ が存在する。ここで $r \upharpoonright m$ は r の最初の m 桁を表す。
ある実数 r が c.e. であるとは、単調増加の有理数近似列が存在することを言う。つまり、$\Phi(m) = a_m \in \mathbb{Q}$ となる単調増加有理数列 $\{a_m\}$ に対して $r = \lim_n a_n$。
万能チューリングマシンを使って、すべての c.e. な実数を列挙することができる。
しかし、計算可能な実数を列挙することはできない。

３つのアプローチ

ランダムの理論では以下の３方向からアプローチする。

典型性　わずかな実数しか持たないような特別な性質を持たない。
記述不可能性　記述するのが難しい。
予測不可能性　次の値を予想するのが難しい。

これらの適切な定式化が同値になることが知られている。
以下ではこれらの定式化について順番に説明する。

カントール空間

二進無限列の集合を カントール空間 と呼び、$2^\omega = \{0,1\}^\omega$ で表す。
有限列 σ に対して、$[\sigma] = \{Z \mid \sigma \preceq Z\}$ を σ の シリンダー と呼ぶ。
開基としてシリンダーの集合を取る。
二進有限列の集合 $S \subseteq \{0,1\}^*$ から作られる開集合を、

$$[S]^{\preceq} = \bigcup_{\sigma \in S} [\sigma]$$

で定義し、S が c.e. のとき、$[S]^{\preceq}$ を c.e. 開集合と呼ぶ。
測度 μ を $\mu([\sigma]) = 2^{-|\sigma|}$ から作られる測度とする。
カントール空間を $[0,1]$ の実数と同一視すれば、μ はルベーグ測度となる。

Martin-Löf ランダムネス

ランダムな列を「わずかな列しか持たないような特別な性質を持たない。」で定義したい。
「わずかな列」を「測度 0 の開集合に含まれる列」と解釈したいが、そんな列は存在しない。
そこで計算可能性を使って以下のように制限する。

定義 1 (Martin-Löf 1966). 一様に *c.e.* の開集合の列 $\{V_n\}$ が $\mu(V_n) \leq 2^{-n}$ を満たす時、*Martin-Löf* テスト と呼ぶ。
列 X が $X \notin \bigcap_n V_n$ となる時、X はこのテストを パスする と言う。
すべての *Martin-Löf* テストをパスする列を *Martin-Löf* ランダム または *ML* ランダム であると言う。

万能 *Martin-Löf* テスト

c.e. の開集合の列は一つのマシンで数え上げることができるので、実は一つの Martin-Löf テストだけを考えれば良い。

定義 2. *Martin-Löf* テスト V_n が、任意の *Martin-Löf* テスト U_n に対して、$\bigcap_n U_n \subseteq \bigcap_n V_n$ であるとき、万能 であると言う。

命題 3. 万能 *Martin-Löf* テストが存在する。

証明. e 番目の Martin-Löf テスト U_n^e に対して、$V_n = \bigcup_e G^e_{n+e+1}$ とすれば良い。 □

命題 4. *Martin-Löf* ランダムな列の集合は測度 *1* を持つ。

証明. 万能 Martin-Löf テストをパスしない列の集合は測度 0 である。 □

ランダムを定義する（その２）

Kolmogorov複雑性

次に記述不可能性を定式化する。
Kolmogorov 複雑性は Kolmogorov, Solmonoff, Chaitin らにより独立に定義された。
万能チューリングマシン U を固定し、有限列 x の Kolmogorov 複雑性 $C(x)$ は以下で定義される。

$$C(x) = \min\{|p| \ : \ U(p) = x\}$$

自明な上限として、$C(x) \leq |x| + O(1)$ が分かる。
$C(x) \geq |x| - c$ となる x を c 圧縮不能という。
長さ n の列のうち c 圧縮不能な文字列は全体の $1 - 2^{-c}$ より多い。

prefix-free Kolmogorov 複雑性

Martin-Löf ランダムネスとの関係を調べる中で、prefi x-freeの概念が導入された。
文字列の集合が prefi x-freeとはどの二つの文字列も互いの接頭辞にならないことを言う。
チューリングマシンが prefi x-freeとは定義域が prefi x-freeであることを言う。
prefi x-freeチューリングマシン V に対し、prefi x-free Kolmogorov 複雑性 $K(x)$ が以下で定義される。

$$K(x) = \min\{|p| \ : \ V(p) = x\}$$

上限は、$K(x) \leq |x| + K(|x|) + O(1)$ である。

定理 5 (Schnorr 1971). 列 A が *Martin-Löf* ランダムであることと $K(A \restriction n) \geq n - O(1)$ となることは同値である。

丁か半か

次に予測不可能性の概念を定式化する。
「丁半」

- 2 つのサイコロをツボの中に入れて降り、合計が偶数か奇数かを当てるゲーム
- 偶数または奇数になる確率はそれぞれ 50%（のはず）
- 当たれば賭け金の 2 倍がもらえる
- 初期投資は有限で借金はできない

もし、予測可能であれば、何らかの規則があるということになる。
元手を何倍にもできるような規則があるか、ないか。
あればランダムではない、なければランダムであると言って良いだろう。

マルチンゲール

次の値の予測戦略に当たる概念を以下で定義する。

定義 6. 次を満たす関数 $d : \{0,1\}^* \to \mathbb{R}^+$ を <u>マルチンゲール</u> と呼ぶことにする。

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

マルチンゲール d が、実数 A で成功するとは、$\limsup_n d(A \upharpoonright n) = \infty$ となることを言う。

定理 7 (Schnorr 1971). ある列が *Martin-Löf* ランダムであることとすべての *c.e.* マルチンゲールに対して成功しないことは同値である。

命題 8. 万能な *c.e.* マルチンゲール d が存在する。

万能な c.e. マルチンゲール d が大きくならないように列を選ぶことで Martin-Löf ランダムな列を得ることができる。

Chaitin の Ω

Martin-Löf ランダムな列として重要なものに Chaitin の Ω がある。

定義 9.

$$\Omega = \sum_{U(\sigma)\downarrow} 2^{-|\sigma|}$$

停止確率とも呼ばれる。
停止集合とチューリング同値である。
どんな prefi x-free万能チューリングマシンに対しても収束して 1 未満である。

定理 10 (Calude, Hertling, Khoussainov and Wang (1998), Kučera and Slaman (2001))**.** 実数が *c.e.* で *Martin-Löf* ランダムであることと、ある *prefix-free* 万能チューリングマシンの停止確率になることは同値である。

自然な性質

Martin-Löf ランダムは、ランダムな列として自然な性質を多く持つ。

- 正規数である。（大数の法則の成立）
- 最初の有限個の入れ替えで閉じている。
- 計算可能な並び替えで閉じている。
- 重複対数の法則が成り立つ。

独立性定理

ある列が別の列から見てランダムであるということを定式化する。

定義 11. ランダムネスの定義で A をオラクルとしても利用することを許したものを *A-Martin-Löf* ランダムと言う。

ランダムであれば、どの部分も他の部分の情報を持たないのが自然である。
列 A と B をそれぞれ奇数番目と偶数番目に持つ列を $A \oplus B$ で表す。

定理 12 (独立性定理、van Lambalgen 1987). 列 $A \oplus B$ が *ML* ランダムであることと、A が *ML* ランダムかつ B が *A-ML* ランダムであることは同値。

ランダムネスが自然であるかどうかの基準の一つとなっている。

様々なランダムネス

Schnorr の批判

Schnorr はランダムの概念を定義するテストとして計算可能なものだけを許すべきだと主張した。

定義 13 (Schnorr 1971). 一様に *c.e.* な開集合の列 $\{V_n\}$ が $\mu(V_n) = 2^{-n}$ を満たすとき、$\{V_n\}$ を *Schnorr* テスト と呼ぶ。すべての *Schnorr* テストをパスする列を *Schnorr* ランダム と呼ぶ。

$\mu(V_n)$ が n に関して一様に計算可能で 0 に収束するとしても良い。測度は計算可能であるが、計算可能な開集合である必要はない。

定義 14 (Schnorr 1971). 計算可能なマルチンゲールで成功しない列を 計算可能ランダム と呼ぶ。

弱いランダムネス

様々なランダムネスが定義され、その性質が調べられた。
計算の強さにより以下のような包含関係がある。
ML ランダム ⇒ 計算可能ランダム ⇒Schnorr ランダム ⇒Kurtz ランダム
計算資源の制限によるランダムが Lutz らによって研究されている。
多項式時間ランダムネス、指数時間ランダムネス、多項式領域ランダムネスなど。
予想する順番を入れ替えを許すことによって得られる、Kolmogorov-Loveland ランダムネスもランダムネスとして自然な性質を持っている。

強いランダムネス

ジャンプは $A' = \{e : \Phi_e^A(e)\downarrow\}$ で定義される。

$$\phi' \equiv_T K \equiv_T \Omega$$

定義 15. 列が $\phi^{(n-1)}$ ランダムのとき、<u>n ランダムネス</u> と呼ぶ。

定理 16 (Nies, Stephan and terwijn 2005). 列 A が 2 ランダムと無限に多くの n で $C(A\upharpoonright n) \geq n - O(1)$ であることは同値。

定理 17 (Miller 2009). 列 A が 2 ランダムと無限に多くの n で $K(A\upharpoonright n) \geq n + K(n) - O(1)$ であることは同値。

複雑性とランダムのきれいな対応が見られる。

ランダムと計算可能性

2 ランダムネスが自然であるという主張もある。

定理 18 (Kučera (1984), Gács (1986))**.** すべての列はある *Martin-Löf* ランダムな列に wtt 還元可能。

定義 19. A が Low for Ω とは、Ω が A ランダムであることを言う。

定理 20 (Nies, Stephan and Terwijn 2005)**.** 列 A が *ML* ランダムかつ *Low for* Ω であることは、A が 2 ランダムであることと同値。

Proof. A が 2 ランダム $\iff$ A が Ω ランダム
$\iff$ $A \oplus \Omega$ が Martin-Löf ランダム $\iff$ Ω が A ランダム □

記述集合論との関係

計算可能性を超算術的まで広げてランダムネスを定義したのは Martin-Löf で 1968 年のことである。
現在 Δ^1_1 ランダムネスと呼ばれる。
現在もっとも自然なランダムネスと考えられているのは Martin-Löf ランダムネスであるが、Martin-Löf 自身は Δ^1_1 ランダムネスこそが自然なランダムネスであると主張した。
本格的な研究が始まったのは 2007 年で Nies らによる。
超算術的な関数でのランダムネスを考えることは、記述集合論での解析的階層の精密化に他ならない。
よって、ランダムネスと記述集合論を区別する必要はないという研究者もいる。

全くランダムではない列と計算可能性

ランダムネスの研究における最近の最も大きな成果の一つが以下の４つの概念の同値性である。

K-trivial $(\forall n)K(A \upharpoonright n) \leq K(n) + O(1)$ (Chaitin 1975)
low for ML $MLR^A \subseteq MLR$ (Zambella 1990)
low for K $(\forall n)K(n) \leq K^A(n) + O(1)$ (Andrej A. Mucknik 1999)
basis for ML ある A-ML ランダムな Z に対して $A \leq_T Z$(Kučera 1993)

定理 21 (Hirschfeldt, Nies, Stephan 2004). 上の４つは同値である。

フラクタル次元とランダムネス

基礎とする測度を計算可能な別の測度にした場合のランダムネスも考察されている。
特に Hausdorff 測度 H^s を使って、Hausdorff 次元の計算可能版が定義される。

$$\dim_H A = \inf\{s \in \mathbb{Q}^+ \ : \ A \text{ は } H^s \text{ランダムではない}\}$$

Kolmogorov 複雑性による特徴付けも存在し、

$$\dim_H A = \liminf_{n\to\infty} \frac{K(A \restriction n)}{n}$$

さらに Lutz(2003) により修正されたマルチンゲールによる特徴付けも与えられている。

適切な相対化

Schnorr ランダムネスの不自然さ

今まではランダムネスの概説をした。
以下では Schnorr ランダムネスの適切な相対化を与えたという結果について話をする。
最近、Schnorr ランダムネスは独立性定理が成り立たないことが示された。

定理 22 (Merkle, Miller, Nies, Reimann, Stephan 2006)**.** $A \oplus B$ は *Schnorr* ランダムだが、A は B-*Schnorr* ランダムではないような A, B が存在する。

しかし、この逆は成り立つ。

定理 23 (Yu 2007)**.** A が *Schnorr* ランダムで B が A*Schnorr* ランダムならば、$A \oplus B$ はシュノアランダムである。

独立性定理は適切なランダムの概念かどうかの指標の一つであるので、自然なランダムネスである Schnorr ランダムネスで成立しないのは不自然。

Schnorr ランダムネスの不自然さ（その２）

４つの概念の同値性の Schnorr ランダムネス版が中途半端にしか成り立たない。

定理 24 (Kjos-Hanssen, Nies and Stephan 2005, Downey, Greenberg Mikhailovich and Nies 2008)**.** 以下は同値。

- *computably traceable*
- *low for Schnorr randomness*
- *low for computable measure machine*

定理 25 (Franklin and Stephan 2008)**.** 以下は同値。

- *Schnorr trivial*
- *truth-table low for Schnorr randomness*
- ある A-*tt Schnorr* ランダムな B に対して $A \leq_{snr} B$

truth-table low の概念による適切な相対化を行うことでこの不自然さを解消できる。

別の相対化とは

それぞれのランダムの概念はオラクルの列に対して定義され、ランダムな列の族と同一視できる。
よって列から列の族への関数 S と思うことができる。
独立性定理が成立するとは、以下が成り立つことである。

$$A \oplus B \in S(\phi) \iff A \in S(\phi) \& B \in S(A)$$

S' が S の別の相対化であるとは、$S(\phi) = S'(\phi)$ が成り立つことである。
よって独立性定理が成立する相対化は本質的に一つしかなく、それが自然な相対化であることを意味する。

独立性定理の証明の再検討

独立性定理の一方向が、ML ランダムで成り立ち、Schnorr ランダムで成り立たない理由を探る。

定理 26. $A \oplus B$ が *ML* ランダムならば、B は A-*ML* ランダムである。

証明. B が A-ML ランダムでないとしよう。
ML テスト V_n^A があって、$B \in \bigcap_n V_n^A$ である。
新たな ML テスト W_n を以下で定義する。

$$W_n = \{X \oplus Y \mid X \in 2^\omega, Y \in \widetilde{V}_n^X\}.$$

ここで $\widetilde{V}_n^X$ は V_n^X として測度が 2^{-n} を超えないようにしたものとする。
$\widetilde{V}_n^A = V_n^A$ であるから、$A \oplus B \in \bigcap_n W_n$。 □

独立性定理を成り立たせるには

上の証明が Schnorr ランダムで成り立たない理由は、

$$W_n = \{X \oplus Y \mid X \in 2^\omega, Y \in \widetilde{V}_n^X\}$$

において、この V_n^X が Schnorr テストで模倣できないからである。
模倣できるようにするために V_n に対して以下の制限を考える。

- すべての X に対して、V_n^X が X で計算可能なもの。
- V_n^X は X を使って良いが計算時間は計算可能な関数で抑えられるもの。

実はこの 2 つの制限は同値である。

相対化の解釈

普通、相対化は計算能力で考える。
これはランダムネスの理論が計算可能性の理論を基盤に作られているからであり、伝統的なものである。
Schnorr ランダムはテストとして計算可能なものに制限したという見方がされてきた。
独立性定理の観点からすれば、テストに対応するチューリングマシンを制限しているのだと解釈できる。
これは同時に、チューリングマシンとして適切な制限ならばそれぞれにランダムの概念が定義できる可能性があることを示唆している。

真理値表 Schnorr ランダム

真理値表還元

Φ_e を e 番目のチューリングマシンとする。
列 A, B に対して、ある e があって $A = \Phi_e^B$ となるとき、A は B にチューリング還元可能であると言い、$A \leq_T B$ と書く。
特に、すべての Z に対して Φ_e^Z が定義されるとき、真理値表還元可能であると言い、$A \leq_{tt} B$ と書く。
この tt は truth-table(真理値表) を表している。

命題 27. $A \leq_{tt} B$ であることと $A = \Phi^B$ かつ $\Phi^B(n)$ を計算するために必要な時間が n に関して計算可能な関数で抑えられることは同値。

真理値表 *Schnorr* ランダム

定義 28. *c.e.* 開集合の列 $\{V_n\}$ が一様に A に真理値表還元可能な *ML* テストであるとき、$\{V_n\}$ は <u>A 真理値表 *Schnorr* テスト</u> であると言う。
すべての A 真理値表 *Schnorr* テストをパスする列を <u>A 真理値表 *Schnorr* ランダム</u> と言う。

注意 29. 上の表現に合わせて *Schnorr* テストを表現すると以下のようになる。
c.e. 開集合の列 $\{V_n\}$ が一様に A にチューリング還元可能な *ML* テストであるとき、$\{V_n\}$ は A-*Schnorr* テストであると言う。

命題 30. 任意の A に対して、A-*Schnorr* ランダムならば A 真理値表 *Schnorr* ランダム。

真理値表 *Schnorr* ランダムと *Schnorr* ランダムは似ている

命題 31. A が計算可能な列の時、A-*Schnorr* ランダムと A 真理値表 *Schnorr* ランダムは一致する。

よって真理値表 Schnorr ランダムは Schnorr ランダムの別の相対化になっていることが分かる。

命題 32. $A \le_T B$ のとき、A は B-*Schnorr* ランダムではない。
$A \le_{tt} B$ のとき、A は B 真理値表 *Schnorr* ランダムではない。

真理値表 *Schnorr* ランダムの同値な定義

ML ランダムネスや Schnorr ランダムネスには３つの同値な定義があった。
同様に真理値表 Schnorr ランダムネスにも３つの同値な定義を与えた。
マルチンゲールによる定義は Franklin と Stephan に与えられた定義と同値になることも示した。

真理値表 *Schnorr* ランダムの独立性定理

定理 33. 真理値表 *Schnorr* ランダムでは独立性定理が成り立つ。

よって A-Schnorr ランダムと A 真理値表 Schnorr ランダムが異なるような A が存在する。
Schnorr ランダムで独立性定理が一方向だけ成立する理由は Schnorr ランダムが不適切に小さく定義されたからだと解釈できる。

命題 34 (Kjos-Hanssen). $A \equiv_T B$ かつ $A \oplus B$ が *Schnorr* ランダムとなるような A, B が存在する。

命題 35. $A \equiv_T B$ かつ A は B 真理値表 *Schnorr* ランダムとなるような A, B が存在する。

A-Schnorr ランダムと A 真理値表 Schnorr ランダムが異なる A の特徴付けはまだできていない。

Schnorr ランダムの列が独立性定理を満たさないのは

Schnorr ランダムが独立性定理を満たさないことが分かってから、どのような Schnorr ランダムの列に対して成り立たないのかについて調べられてきた。
Franklin と Stephan はどのような Schnorr ランダムの列が独立性定理を満たさない Schnorr ランダムの半分になるかを問うた。
真理値表 Schnorr ランダムに関する次の定理は、これに対して明快な解答を与える。

定理 36. 任意の A に対して、チューリング同値の A 真理値表 *Schnorr* ランダムが存在する。

空間の一般化

空間の一般化

ランダムの概念は確率の基礎となりうるだけでなく、統計や機械学習、人工知能への応用が期待できる。
しかし現在までそのような応用が乏しい理由の一つが、ランダムネスを考える空間が限定されていることである。
Martin-Löf による定義は自然に位相空間に拡張できることを Zvonkin と Levin(1970) が指摘した。
それを受けて Hertling と Weihrauch(1998) により考察されている。
その後複雑性による定義とその同値性についても考察が進んできた。

一般化の歩み

	measure	complexity	martingales
$2^\omega, \mathbb{R}$	Martin-Löf 1966	Schnorr 1971	Schnorr 1971
compact CMS	-	Gács 2005	?
CMS	-	Hoyrup&Rojas 2009	?
CTS	Hertling&Weihrauch 1998	?	?

ここで「CMS」は computable metric space を、「CTS」は computable topological space を表している。

最近、computable topological spaces 上での complexity と martingale による特徴付けを与えることができた。

ランダムネスの応用に向けて一歩一歩進んでいきたい。

終り

ありがとうございました。