

「ランダム性と計算可能性」

宮部賢志

明治大学理工学部数学科

random

Algorithmic Randomness

- computability theory
- ~~information theory~~
- measure theory (or probability theory)
- information theory

の融合分野

関連分野: ergodic theory, statistical prediction

前半 ランダムの概念の重要性

後半 randomness and computability

ランダム性の概念の重要性

2 June, '16

(1)

@筑波大学数学域談話会

科学の概念の
見直しが必要

正しい

② ランダムの概念の重要性

②

ランダムな引はどれか?

• 01010 10101 01010 10101 01010 10101 01010 10101
0と1が交互

• 00100 10000 11111 10110 10101 00010 00100 00101
出現率の少数を以下に2進展開

• 00011 00100 11000 10001 10111 01110 11001 11001
奇数番目は2番目の引、偶数番目は1番目の引

• 01010 00101 01110 10101 00101 01011 10101 01001

確率としてはどれも 2^{-40} で区別できない

ランダム $\xrightarrow[\text{規則が}\bracket{\text{分かる}}]{} \text{ランダムでない}$ 主観的?

ランダム性

(3)

典型性、予測不可能性、圧縮不可能性

(Def) (1930s)

$f: \Sigma^{\omega} \rightarrow \Sigma^{\omega}$ が 計算可能 とは、

プログラミング

その函数を計算する Turing machine が存在すること。言語

(Prop)

カズキ

プログラムを並べ上げることができる。

すべての Turing machine を模倣する universal machine U が存在する。

$\forall f: \text{計算可能}, \exists \delta \in \Sigma^{\omega} \cup (\delta \tau) = f(\tau)$

(Def)

$S \subseteq \Sigma^{\omega}$ が prefix-free とは、

$\forall \sigma, \tau \in S, \sigma \neq \tau \Rightarrow \sigma \not\preceq \tau \text{ かつ } \tau \not\preceq \sigma$.

電話番号

Turing machine が prefix-free $\stackrel{\text{def}}{\iff} \text{dom}(\Phi)$ が prefix-free

(Prop) $\exists U: \text{prefix-free universal prefix-free machine. s.t.}$

$\forall f: \text{prefix-free machine } \exists \delta \in \Sigma^{\omega}, U(\delta \tau) = f(\tau) \text{ for all } \tau \in \Sigma^{\omega}$

(Def) Kolmogorov complexity (19⁶0s-70s)

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

Counting Theorem

$$|\{\sigma : |\sigma| = n \wedge K(\sigma) \leq n + K(n) - r\}| \leq 2^{n-r+O(1)}$$

$$\text{特に } \max\{K(\sigma) : \sigma \in 2^n\} = n + K(n) \pm O(1)$$

後半の $\leq$ だけ示す. (後半の $\geq$ は前半から出る)

以下のような機械 M を考える

$$M(\sigma\tau) = \begin{cases} \sigma\tau & \text{if } M(\sigma) = \tau \\ \text{undefined} & \text{otherwise} \end{cases}$$

τ は prefix-free な σ のみを見つけておける.

よって M は prefix-free

また $K_M(\sigma\tau) \leq n + K(n)$ より.

τ の最小性から $K(\tau) \leq n + K(n) + O(1)$.

前半の主張は、

ほぼ 2^n だけ

r だけ圧縮できる文字列は、全体の 2^{-r} しか存在しない.

逆に、ほとんどの列は、ほとんど圧縮できない.
ことを意味している.

このことは、長さ n の文字列が 2^n であることから.

長さ $\leq n$ の文字列のほとんどが長さ n に近いことから
予想される.

ただし、証明はもう少しややこしい.

(例)

$$K(n) \leq 2 \log n + 2 \quad \text{底は 2.}$$

(証明)

n という文字列は、2進数値で表せば $\log_2$ 桁で表せる。

0 は 00 , 1 は 01 と重複を許して表現し。

最後に 01 とする事で、prefix-free な表現が得られる。

~~同じ理由で~~、 $K(n) \leq (1+\varepsilon) \log n + O(1)$ が示せる。
前定理を使えば、

(例)

$$K(n) \leq \log n + O(1) \quad \text{は成り立たない。}$$

(定理)

(Kraft の不等式)

M は prefix-free machine とすると、

$$\sum_{\sigma \in \text{dom}(M)} 2^{-|\sigma|} \leq 1$$

(証明)

$\sigma \in \text{dom}(M)$ に対して、 $[\sigma]$ は互いに素

$$\text{よって} \quad \sum_{\sigma \in \text{dom}(M)} 2^{-|\sigma|} = \mu\left(\bigcup_{\sigma \in \text{dom}(M)} [\sigma]\right) \leq 1$$

↑

(例の証明)

Cantor 空間上の一様測度

もし、 $K(n) \leq \log n + c$ とすれば、

$$\sum_n 2^{-K(n)} \geq \sum_n \frac{1}{2^c n} = \infty$$

だが、これは、

$$\sum_n 2^{-K(n)} \leq \sum_{\sigma \in \text{dom}(M)} 2^{-|\sigma|} \leq 1 \quad \text{に反する。}$$

(1970s) Chaitin-Solomonoff-Kolmogorov

Def $X \in 2^\omega$ が 1-random とは、
 $K(X^n) > n - o(1)$ という。

Rem 1ML-random と同値

Thm 1-random な列の集合は測度 1

Proof $\forall c \in \omega$

$V_{c,n} = \bigcup \{[\sigma] : K(\sigma) \leq n - c, |\sigma| = n\} \leftarrow \text{長 } n \text{ で } c \text{ だけ圧縮できる文字列の集合}$

$$\mu(V_{c,n}) \leq 2^{-n} \cdot 2^{n - K(n) - c + d} = 2^{-K(n) - c + d}.$$

$$\mu\left(\bigcup_n V_{c,n}\right) \leq \sum_n 2^{-K(n) - c + d} \leq 2^{-c + d}$$

↑ c だけ圧縮できる列の集合。
 ランダムでない列の集合は、

$\bigcap_c \bigcup_n V_{c,n}$ に含まれるので、その測度は 0

↑
 任意の長さ c だけ圧縮できる列の集合。

//

Thm 計算可能な列はランダムではない

Proof $n \mapsto A \upharpoonright n$ が計算可能なので、 $n \in \text{prefix-tree}$ に表し、
これを計算する machine を M とすれば、

$$K_M(A \upharpoonright n) \leq 2 \log n + 2$$

よって

$$K(A \upharpoonright n) \leq 2 \log n + 2 + c.$$

よって A はランダムではない //

すなわち、ランダムな列とランダムでない列が存在する。

ランダムな列の例は後に見る。

Thm A: 1-random $\Rightarrow \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n A(k)}{n} = \frac{1}{2}$
 大数の法則.

証明

$$S_n = k = \frac{n}{2} + f, \quad f \geq 0 \text{ かつ } 3 \leq n.$$

$$K(A|n) \leq K(n) + K(f) + \log_2 \binom{n}{k}$$

Stirling の公式より

$$\ln n! = (n + \frac{1}{2}) \ln n - n + o(1).$$

$$0 \leq x \leq \frac{1}{2} \text{ とき}$$

$$\ln(1+x) \geq x - \frac{1}{2}x^2$$

$$\ln(1-x) \geq -x - x^2$$

$$\begin{aligned} \ln k &= \ln\left(\frac{n}{2} + f\right) \\ &= \ln\left(\frac{n}{2}\right) + \ln\left(1 + 2 \cdot \frac{f}{n}\right) \\ &\geq \ln n - \ln 2 + 2 \cdot \frac{f}{n} - 2 \cdot \frac{f^2}{n^2} \end{aligned}$$

$$\ln(n-k) \geq \ln n - \ln 2 - 2 \cdot \frac{f}{n} - 4 \cdot \frac{f^2}{n^2}$$

$$\begin{aligned} \text{よって} \quad \ln k! &\geq \left(\frac{n}{2} + f + \frac{1}{2}\right) (\ln n - \ln 2 + 2 \cdot \frac{f}{n} - 2 \cdot \frac{f^2}{n^2}) - \left(\frac{n}{2} + f\right) \\ \ln(n-k)! &\geq \left(\frac{n}{2} - f + \frac{1}{2}\right) (\ln n - \ln 2 - 2 \cdot \frac{f}{n} - 4 \cdot \frac{f^2}{n^2}) - \left(\frac{n}{2} - f\right) \\ \ln(k!(n-k)!) &\geq (n+1)(\ln n - \ln 2) + \frac{(n-3)f^2}{n^2} + 2 \cdot \frac{f^3}{n^2} - n \end{aligned}$$

$$\ln \binom{n}{k} \leq (\ln 2)(n+1) - \frac{1}{2} \ln n - \frac{(n-3)f^2}{n^2} - 2 \cdot \frac{f^3}{n^2}$$

$$\log \binom{n}{k} \leq n+1 - \frac{1}{2 \ln 2} \ln n - \frac{(n-3)f^2}{(\ln 2)n^2} - \frac{2}{\ln 2} \frac{f^3}{n^2}$$

$f \geq \varepsilon n$ ならば、右辺は n より $o(\ln n)$ 以上小さくなる。

・大数の法則 と 確率概念 は関係ない?

・ランダム (→ 大数の法則)

↑
確率的振る舞い

① モデル選択



説明モデル ⇒ 計算可能であるはず!
(予測)

計算可能な予測 (見つけ戦略) の中で、

"最善のもの" が存在する。

それは、記述が最少のプログラムを重視する予測りとなっている。

Solomonoff の universal induction
(1960s -)

⑨エルゴード理論

エルゴード仮説

確率1で「長時間平均 = 位相平均」が成立。
or 測度1で

軌道は決定的なのに確率概念が?

軌道のランダム性 = 初期値のランダム性

↓

長時間平均 = 位相平均

↓

確率1

確率論では $\Leftarrow$ だけ、

ランダムネスの理論では $\Rightarrow$ も示せる。

Hoyrup - Rojas の研究

(2010s -)

② ランダム性と計算可能性

ランダムは使えない? — ランダム=むだづめ、規則性がない
人為的でない、秩序がない
作為的でない

ランダムは使える? — ランダム化比較実験、モンテカルロ
確率モデルのシミュレーション

Φ_e : e 番目の Turing machine.

(定理) $H = \{ \langle e, k \rangle : \Phi_e(k) \downarrow \}$ は計算可能ではない

(Proof) $K = \{ e : \Phi_e(e) \downarrow \}$ が計算可能であるとすると.

$$h(n) = \begin{cases} \Phi_n(n) + 1 & (n \in K) \\ 0 & (n \notin K) \end{cases}$$

も計算可能、その指数 e, e' とすると.

$$e' \in K \Rightarrow \Phi_{e'}(e') = h(e') = \Phi_e(e') + 1$$

$$e' \notin K \Rightarrow \Phi_{e'}(e') = \dagger(e') = 0$$

矛盾

//

H は構成が難しい (計算可能でない)

H は有用 (H を使えば様々な集合が計算できる)

ランダム性は?

問い

・ランダムな列は必ず有用である?

・ランダムな列には有用な列とそうでない列がある?

・ランダムな列はすべて有用ではない?

定義 Ω が ^{universal machine} 万能機械 とは、

任意の computable function f に対し、 $\exists \delta \forall \tau$

$$\Omega(\delta\tau) = f(\tau)$$

模倣できる。

Def $\Omega_U = \sum_{\sigma \in \text{dom}(U)} 2^{-|\sigma|}$ 停止確率 (halting probability)

Thm $\Omega \equiv_T \underbrace{H \equiv_T K}_{\text{使う}}$

Proof

Ω は T -から 近似可能 な τ 。

$\Omega > \frac{1}{2}$ ならば 止まる machine と考えることで、
これは H を使、て 判定可能。

$$\text{よって } \Omega \leq_T H$$

逆を示す。

$$M(0^e 1) \downarrow = \lambda \stackrel{\text{def}}{\iff} \Phi_e(e) \downarrow \quad \text{この } \Phi_e$$

U は universal な τ である。

$$U(e0^e 1) \downarrow \iff e \in K$$

Ω を使、て $e \in K$ と判定する

$$\Omega - \Omega_s < 2^{-(|e|+e+1)} \quad \text{まで待つ}$$

もし、 $U(e0^e 1)[s] \downarrow$ であれば、 $M(0^e 1) \downarrow$ して $e \in K$
 $U(e0^e 1)[s] \uparrow$ であれば、

もし $t > s$ で $U(e0^e 1)[t] \downarrow$ となる。

$$\Omega_t \geq \Omega_s + 2^{-(|e|+e+1)} > \Omega \quad \text{より矛盾。}$$

すなわち $U(e0^e 1) \uparrow$ なので、 $e \notin K$ //

Thm Ω is 1-random

Proof

再帰定理より M の指数 c を使, て良い

$$V(0^c | \sigma) = M(\sigma) \text{ である.}$$

M は、入力 $\tau \in 2^{<\omega}$ に対し、

$$V(\tau)[s] = \Omega_s \upharpoonright n \text{ かつ } |\tau| < n - c$$

となる s, n を探す。

つまり $K(\Omega_s \upharpoonright n) < n - c$ である。

そのような n, s を見つけたら、

$$\mu \neq \text{rng } V[s] \text{ であり、 } M(\tau) = \mu \text{ である.}$$

$$\text{このとき、 } V(0^c | \tau) = \mu \text{ かつ } V(0^c | \tau)[s] \uparrow \neq \mu$$

$$\Omega - \Omega_s \geq 2^{-(c+1+n)} > 2^{-n}$$

$$\text{よって } \Omega_s \upharpoonright n \neq \Omega_s \upharpoonright n$$

すなわち、 $|\tau| < n - c$ ならば、 $V(\tau) \neq \Omega \upharpoonright n$ 。

$$\text{よって、 } K(\Omega \upharpoonright n) \geq n - c. //$$

有用なランダム列の存在!!

(van Lambalgen 1998)

Thm $A \oplus B$ が 1 -random $\Leftrightarrow A$ が 1 -randomかつ B が A に対して 1 -randomProof $\Rightarrow$ B が A に対して 1 -randomでないとする. $\forall d \in \omega \quad |K^A(B \upharpoonright n)| < n - d$ $\{ \langle \sigma, \tau \rangle : K^\sigma(\tau) < |\tau| - d, |\sigma| \geq |\tau| \}$

で、prefix-free となるように取る.

それ以外のプログラム p に対して、入力: p, d, σ, x の各場合 $\bar{d}$ は d を表現する $2 \log_2 d + 2$ の文字列
長さ x は $|x| = |\sigma| - |\tau|$ となる任意の文字列出力: $\sigma \oplus (\tau x)$

prefix-free となる.

こゝに示す、

入力の長さは $|p| + 2 \log_2 d + 2 + \overbrace{|x|}^{|\sigma|} + |x|$ $< |\tau| - d + 2 \log_2 d + 2 + |\sigma| + |x|$ より $(d - 2 \log_2 d - 2)$ の余裕がでる.

Proof 終止

($\Leftarrow$) A が 1-random かつ B が A に対して 1-random ならば

$$K(A \oplus B | 2n) > 2n - O(1)$$

を示せば良い

n が十分大 $\leq C$ だけ固定する。

$$K(A|n) = n + c_n \text{ である。}$$

$$d_n = K(B|n | (A|n)^*) > n - c_n - O(1) \text{ を示せば良い。}$$

$$(\circledast) K(G, \tau) = K(\sigma) + K(\tau | \sigma^*) + O(1) \text{ より}$$

$$K(A \oplus B | 2n) = K(A|n, B|n)$$

$$\begin{aligned} &= K(A|n) + K(B|n | (A|n)^*) \\ &= n + c_n + K(B|n | (A|n)^*) \end{aligned}$$

$$\text{したがって } K^x(A|n) \leq d_n + K(|d_n - n|) + K(c_n)$$

より得られる。

$$\Omega = A \oplus B \text{ とおく.}$$

A, B は共に 1-random

$$\forall K \not\leq_T A, K \not\leq_T B$$

もし $K \leq_T A$ なら, ~~それは~~

$$B \leq_T \Omega \leq_T K \leq_T A \text{ あり}$$

B は A に関して 1-random であることに反する.

For expert,

Prop some Δ_2^a set is A -1-random
Nies 3.4, 10 $\Rightarrow A$ is GL,
2005

Cor $\Omega = A \oplus B, \Rightarrow A, B$ are low

ほんで

~~これは~~ 有用でないランダムな系列

Thm (Kučera-Saman Theorem) 2001.

1-random and left-c.e. $\Rightarrow K \leq_r A$

Thm Thm

Def $K(X|n) \geq n + K(n) - o(1) \iff X$ is 1-random relative to K
 $\iff$ def
 X is 2-random.

Thm

2-random $\Rightarrow GL_1$

ほとんど有用ではない。

おおざ、18に言えば、

よりランダムになると、有用ではなくなる。

様々なランダム性があり、様々な有用性がある。

どのランダム性の概念がどのように有用なのか。

特色が知られており、まだまだ分からないことも多い。