アルゴリズム的ランダムネスと学習可能性(3)

宮部賢志(明治大学)

2024年7月9日(火) -- 12日(金)
RIMS共同研究（公開型）
数理論理学の最近の進展
Symposium on Advances in Mathematical Logic

目次

1. ランダムネスの理論の概要・歴史: (1)前半
2. Martin-Löfランダム性: (1)後半
3. マルチンゲール: (2)前半
4. Kolmogorov複雑性: (2)後半
5. Solovay還元: (3)前半
6. 学習可能性: (3)後半

定義(Martin-Löf 1966)

Martin-Löf検定(ML検定): 一様c.e.開集合の列$(U_n)_{n\in\omega}$で，すべての$n \in \omega$で$\mu(U_n) \leq 2^{-n}$となるもの．

$X \in 2^\omega$とML検定$(U_n)_n$に対して，$X \notin \bigcap_n U_n$となるとき，XはML検定$(U_n)_n$に**合格する**(pass)という．

$X \in 2^\omega$がすべてのML検定に合格するとき，Xは**MLランダム**であるという．

Solovay還元

left-c.e. MLランダム

実数xがleft-c.e.であるとは，単調増加な計算可能な有理数列$(a_n)_n$で$\lim_n a_n = x$となるものが存在すること．

テーマ：left-c.e. MLランダム，みんな同じ性質を持つ．

- Turing次数が$\mathbf{0}'$
- 停止確率として書ける
- Solovay完全
- 複雑性が定数しか違わない
- 収束速度はすべて同じ

停止確率

定義

万能非接頭機械Uに対し，その**停止確率**(halting probability)を，

$$\Omega = \sum_{\sigma \in \mathrm{dom}(U)} 2^{-|\sigma|}$$

として定義する．

ChaitinのΩと呼ばれることもある．

停止確率は機械Uに依存するため，Ω_Uと書くこともある．
省略した場合は，万能非接頭機械であれば成り立つ性質を議論する．

停止確率の性質

Kraftの不等式から$\Omega \leq 1$が分かる

また左c.e.実数(left-c.e. real)であることも分かる．

命題
停止確率は停止問題とTuring同値である，つまり，$\Omega \equiv_T \emptyset'$

証明
Ωが与えられたら，$\Omega_s \upharpoonright n = \Omega \upharpoonright n$となる$s$まで待つことで，長さ$n$以下の文字列で停止するか判定できる．

逆に各$r \in \mathbb{Q}$に対し$\exists s.\Omega_s > r$ならば止まる，そうでなければ止まらないプログラムを考えることで，Ωを計算できる．

停止確率の性質

定理 (Chaitin)
ΩはMLランダムである．

証明
Ωは停止問題とTuring同値なので有理数ではない．
よって，すべてのnに対してsが存在して$\Omega_s \upharpoonright n = \Omega \upharpoonright n$となる．

十分大きい$c \in \omega$を使って，以下のように機械Mを定義する．
入力$\tau \in 2^{<\omega}$に対して，

$$U(\tau)[s] = \Omega_s \upharpoonright n, \ |\tau| < n - c$$

(これは$K(\Omega_s \upharpoonright n)[s] < n - c$を意味している)ならば，

$\mu \notin \mathrm{rng} U[s]$となる$\mu \in 2^{<\omega}$を見つけ，$M(\tau) = \mu$とする．
c, Mの情報を持つ長さc以下の文字列をρとする．

$$U(\rho\tau) = \mu,\ U(\rho\tau)[s]\uparrow$$

が成り立っていると思って良い．

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

であり，$\Omega \restriction n \neq \Omega_s \restriction n.$
すなわち$|\tau| < n - c$ならば$U(\tau) \neq \Omega \restriction n$なので，$K(\Omega \restriction n) \geq n - c$

証明終

特別な機械Mを作ることで，Uがそれを自動で追いかけてくれるというのが面白い．

Ωは，停止問題を計算することができるような情報をMLランダムという圧縮した形で持っている．

後に逆も成り立つことを紹介する．

Solovay還元

left-c.e.実数がMLランダムあることは収束がとても遅いことを意味する．

定義 (Solovay)

$\alpha \leq_S \beta$とは，部分計算可能な関数$f :\subseteq \mathbb{Q} \to \mathbb{Q}$と$c \in \omega$が存在して，

$$q \in \mathbb{Q}\text{かつ}q < \beta \Rightarrow f(q) \downarrow < \alpha\text{かつ}\alpha - f(q) < c(\beta - q)$$

βの下からの近似有理数qが与えられたら，αのよりよい近似有理数$f(q)$を与えることができる．
つまり，βの方が下から近似しにくい．

ランダム性の尺度

定理

$$\alpha \leq_S \beta \Rightarrow \alpha \leq_T \beta$$

定理

$$\alpha \leq_S \beta \Rightarrow \alpha \leq_K \beta$$

ここで，

$$\alpha \leq_K \beta \iff K(\alpha \upharpoonright n) < K(\beta \upharpoonright n) + O(1)$$

Solovay還元の特徴づけ

定理 (Downey, Hirschfeldt, and Nies 2002)
α, βをleft-c.e.実数とする．
$\alpha \leq_S \beta$であることと以下は同値：ある$c \in \omega$とleft-c.e.実数γが存在して，

$$c\beta = \alpha + \gamma$$

気持ち：αがある一定量変化したら，βもその後でそれと同じぐらい変化する．

left-c.e.実数βが**Solovay完全**(Solovay complete)とは，任意のleft-c.e.実数αに対して，$\alpha \leq_S \beta$となること．

left-c.e.なMLランダム実数が存在するから，Solovay完全なleft-c.e.実数はMLランダム．

Kučeraｰ-Slamanの定理

定理 (Kučera-Slaman 2001)
αをMLランダムなleft-c.e.実数とすると，αはSolovay完全．

証明
βをleft-c.e.実数として，$\beta \leq_S \alpha$を示す．

近似列$(\alpha_s), (\beta_s)$を固定する．
ML検定$(U_n)_n$を以下のように構成する．
stage sにおいて，$\alpha_s \in U_n[s]$であれば何もしない．
そうでなければ，最後にU_nが変化したstage(なければ0)をtとし，
$(\alpha_s,\ \alpha_s + 2^{-n}(\beta_s - \beta_t))$を$U_n$に入れる．

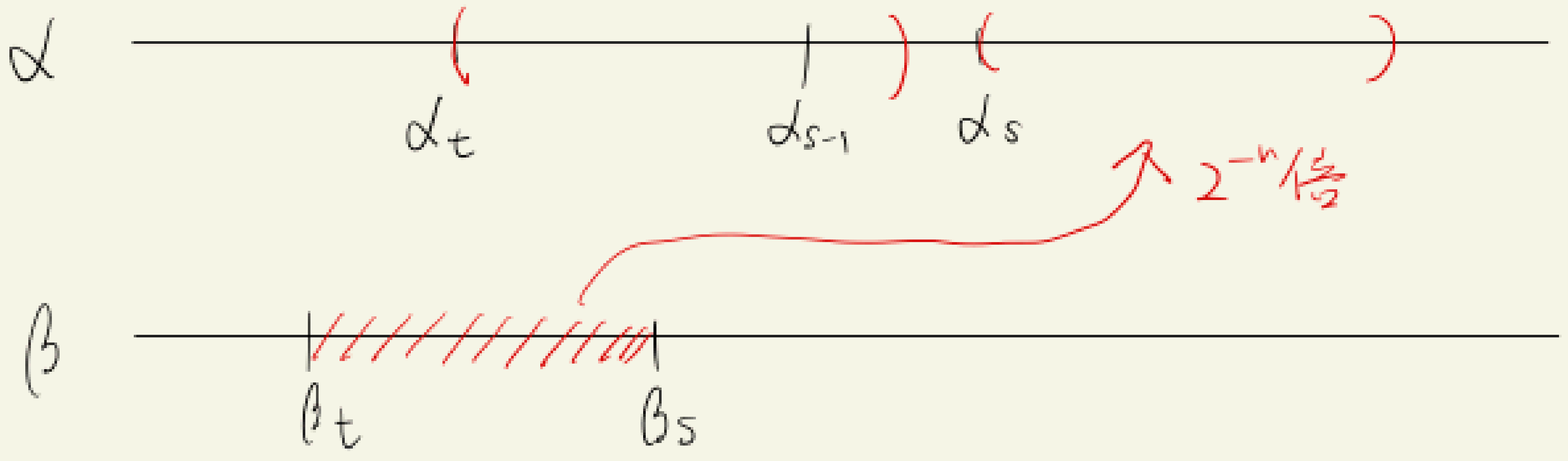
Kučera-Slamanの定理(続き)

明らかに$\mu(U_n) \leq 2^{-n}\beta$で，$(U_n)_n$はML検定.
αはMLランダムなので，ある$n \in \omega$で$\alpha \notin U_n$.

U_nが変化したstageを$s_0, s_1, \cdots$とすると，

$$\alpha_{s_{i+1}} - \alpha_{s_i} \geq 2^{-n}(\beta_{s_i} - \beta_{s_{i-1}})$$

これは，$\beta \leq_S \alpha$を示している.


α
α_t
α_{s-1}
α_s
2^{-n}倍
β
β_t
β_s

同値性

定理 (Calude, Hertling, Khoussainov, and Wang 1998)
αがSolovay完全なleft-c.e.実数ならば，ある万能非接頭機械Uが存在して，$\Omega_U = \alpha$となる．

つまり，left-c.e.実数に対しては，MLランダム性，Solovay完全性，停止確率となることはすべて同値．
Solovay還元はK還元を導くことから，すべてのMLランダムleft-c.e.実数はその非接頭Kolmogorov複雑性は定数を除いて一致する．

収束の速さ

定理 (Barmpalias and Lewis-Pye 2017)

ΩをMLランダムなleft-c.e.実数とし，その近似列$(\Omega_s)_s$を固定する．
αをleft-c.e.実数とし，$(\alpha_s)_s$をその近似列とする．

- αがMLランダムならば，$\lim_s \dfrac{\alpha - \alpha_s}{\Omega - \Omega_s}$が正の値として存在する．
- αがMLランダムでなければ，$\lim_s \dfrac{\alpha - \alpha_s}{\Omega - \Omega_s} = 0$

MLランダムな列はどのような計算可能な近似列をとったとしても，同じように収束が遅い．

c.a.実数

left-c.e.以外の実数に対するSolovay還元を考えたい．

$x \in \mathbb{R}$が**弱計算可能**(weakly computable)とは，計算可能な有理数列$(a_n)_n$で$\sum_n |a_{n+1} - a_n| < \infty$となるものが存在して，$\lim_n a_n = x$となること．

2つのleft-c.e.実数の差になることと同値．
left-c.e.を含む最小の実閉体．

$x \in \mathbb{R}$が**計算近似可能**(computably approximable, c.a.)とは，計算可能な有理数列$(a_n)_n$が存在して$\lim_n a_n = x$となること．

実数を2進展開と同一視したとき，$x \leq_T \emptyset'$と同値．

拡張Solovay還元

定義 (Zheng and Rettinger 2004)
α, βをc.a.実数とする．
$\alpha \leq_S \beta$とは，ある計算可能な近似列$(a_n)_n, (b_n)_n$と$c \in \omega$が存在して，
すべての$n \in \omega$で，

$$|\alpha - a_n| < c(|\beta - b_n| + 2^{-n})$$

となること．

注意：left-c.e.実数に対しては，もとのSolovay還元と同値になる．

Solovay還元の性質

定理 (Rettinger and Zheng 2005)
αを弱計算可能実数とする．

もしαがMLランダムならば，αはleft-c.e.かright-c.e.

証明
αの計算可能近似列$(a_n)_n$で有界変動$\sum_n |a_{n+1} - a_n| < \infty$を固定

$\alpha < a_n$となるnが有限個であれば，αはleft-c.e.である．

$\alpha > a_n$となるnが有限個であれば，αはright-c.e.である．

$(a_n)_n$がαを無限回跨ぐとする．有界変動であることから，m回以上跨ぐ部分は全体の長さの$1/m$以下である．

このことから，αを覆うML検定を作ることができる．

Solovay還元の性質

定理 (Rettinger and Zheng 2005)
αをc.a.実数とする．

$$S(\alpha) = \{\beta \in \mathrm{CA} \ : \ \beta \leq_S \alpha\}$$

は体(field)をなす．

観察 (隈部・宮部・鈴木 to appear)
上記の集合は実閉体にもなる．

Solovay還元の特徴づけ

定理 (隈部・宮部・鈴木 to appear)
α, βをc.a.実数とする．
$\alpha \leq_S \beta$と以下は同値:ある計算可能なα, βの近似列$(a_n)_n, (b_n)_n$と$c \in \omega$が存在して，

$$(\forall n, k)[k < n \Rightarrow |a_n - a_k| < c(|b_n - b_k| + 2^{-k})]$$

考える差が数列同士なので使いやすい．

定理 (Rettinger and Zheng 2005)
αをc.a.実数とする.
$\alpha \leq_S \Omega$とαが弱計算可能であることは同値.

証明
$S(\Omega)$はすべてのleft-c.e.を含む体である.
弱計算可能実数の集合はleft-c.e.を含む最小の体なので，$S(\Omega)$に含まれる.

逆は上記の定理を使うと楽に示せる.

Solovay還元の性質

$\alpha \leq_S \Omega$として，αが弱計算可能であることを示す．
仮定よりα, Ωの近似列$(a_n)_n, (b_n)_n$と$c \in \omega$が存在して，

$$(\forall n, k)[k < n \Rightarrow |a_n - a_k| < c(|b_n - b_k| + 2^{-k})]$$

ここで，$(a_n)_n, (b_n)_n$は有界変動ではないかもしれない．
$b_n \to \Omega$で，Ωには単調増加の近似列$(\Omega_s)_s$が存在するので，

$$|b_{n_i} - \Omega_{n_i}| < 2^{-i}$$

となる計算可能な部分列$(n_i)_i$が取れる．

Solovay還元の性質

ここで，$(a_{n_i})_i$はαに収束する計算可能な列である．また，

$$
\begin{aligned}
|a_{n_{i+1}} - a_{n_i}| &< c(|b_{n_{i+1}} - b_{n_i}| + 2^{-n_i}) \\
&< c(|\Omega_{n_{i+1}} - \Omega_{n_i}| + 2^{-i-1} + 2^{-i} + 2^{-n_i})
\end{aligned}
$$

右辺の和は有限であるから，$(a_{n_i})_i$は有界変動．

left-c.e.実数についてはSolovay次数の構造もよく知られている．

命題

left-c.e.実数α, βに対し，$\alpha + \beta$はα, βのSolovay次数における最小上界を与える．

証明

$\alpha, \beta \leq_S \alpha + \beta$はDowneyらの特徴づけから明らか．

$\alpha, \beta \leq_S \gamma$とすれば，$\gamma$の近似から$\alpha, \beta$の近似を作ることができるので，$\alpha + \beta$の近似も作ることができる．

Solovay次数の構造

定理 (Downey, Hirschfeldt, and Nies 2002)
MLランダムでないleft-c.e.実数αに対し，あるleft-c.e.実数β, γが存在して，$\beta, \gamma <_S \alpha$かつ$\alpha = \beta + \gamma$となる．

MLランダムでないleft-c.e.実数αに対し，あるleft-c.e.実数βが存在して，$\alpha <_S \beta <_S \Omega$となる．

Solovay次数の構造

問題

弱計算可能実数に対するSolovay次数の構造はどうなっているか？
最小上界は常に存在するか？稠密か？
c.a.次数ではどうか？

学習可能性

学習とは

機械学習とはデータから規則を抽出する技術.
理論的な解析のためには設定を定める必要がある.
ここではSolomonoffの枠組みを用いる.

2^ω上の未知の確率測度μからランダムな列$X \in 2^\omega$が最初から順に与えられるとする.
$\sigma = X \upharpoonright n$を見て次のビット$X(n)$の確率を推定する.
真の確率は

$$k \mapsto \mu(k|\sigma) = \frac{\mu(\sigma k)}{\mu(\sigma)}$$

である.

定義

$\xi : 2^{<\omega} \to [0,1]$が**半測度**(semimeasure)とは，$\lambda$を空文字とし，任意の$\sigma \in 2^{<\omega}$に対し，

$$\xi(\lambda) \leq 1,\ \xi(\sigma 0) + \xi(\sigma 1) \leq \xi(\sigma)$$

を満たすこと．

半測度がc.e.であるとは，関数として下側半計算可能(lower semicomputable)であること．

c.e.半測度が**万能**(universal)とは，任意のc.e.半測度νに対し，ある$c \in \omega$が存在して，すべての$\sigma \in 2^{<\omega}$に対し，以下を満たすこと．

$$\nu(\sigma) \leq c\xi(\sigma)$$

定理 (Solomonoff 1964)
μ: 計算可能な確率測度

ξ: 万能c.e.半測度

$k \in \{0,1\}$に対し，μ-確率1で，以下が成立する．

$$|\mu(k|X_{<n}) - \xi(k|X_{<n})| \to 0 \quad (n \to \infty)$$

ξはμの事前知識はないが，Xの情報からμを推定可能．

Solomonoffはこの$\xi(\cdot|\cdot)$を**アルゴリズム的確率**と呼んだ．

これで人工知能の研究終わり．．．というわけにはいかない．

従来の見解

Marcus Hutter’s beautiful (and in some respects beautifully ugly) theory of Universal AI , building on the earlier theory of Solomonoff induction, is the paradigm case of the theory-first approach.(BEN GOERTZELのblogより 2021)

These answers are useless if they cannot be approximated in practice, i.e., by a regular Turing machine. (Leike and Hutter 2018)

One might worry that for this reason, Solomonoff prediction is completely useless as a practical guide for assigning prior probabilities. (A Dilemma for Solomonoff Prediction, Neth 2023)

その他，impracticalとも言われる.

問い

- 計算不可能なpriorを如何に近似するか？
  - 近似は難しい
- 収束に必要なランダム性は？
  - MLランダム性では不十分

学習理論の違い

- 古典的統計的学習理論
  - 過学習，損失関数は凸，1対1対応
- 深層学習，KAN
  - 二重降下，損失関数は非凸，1対1対応しない
- 万能推論
  - すべての計算可能な確率分布を考える，自然数でパラメタ化される
  - 非連続，損失関数は非凸，1対1対応しない

ニューラルネットワークとTuring機械というモデルに差はあるが，似たことを考えている．

汎用的な予測とは

計算可能な確率分布に万能なものは存在しない．
dominateにより半順序を入れれば「より汎用的な予測」の概念は定義できる．

定義

十分汎用的な計算可能な予測確率に対して性質Pが成り立つとは，ある計算可能な予測確率νが存在して，νをdominateするすべての計算可能な予測確率ξに対し，ξが性質Pを持つことを意味する．

KL-divergence

離散分布において，μのξに対するKL-divergenceは，

$$D(\mu\|\xi) = \sum_{\sigma\in\{0,1\}} \mu(\sigma)\log\frac{\mu(\sigma)}{\xi(\sigma)}$$

- $d_\sigma(\mu||\xi) = d(\mu(\cdot|\sigma)||\xi(\cdot|\sigma))$
- $D_n(\mu||\xi) = \sum_{k=1}^{n} E_\mu[d_{X_{<k}}(\mu||\xi)]$
- $D_\infty(\mu||\xi) = \lim_{n\to\infty} D_n(\mu||\xi)$

μとξが計算可能ならば，$D_\infty(\mu\|\xi)$は左c.e.実数か無限大

誤差の収束

定理 (M.)
μ: 計算可能な確率測度
十分汎用的な予測測度ξに対し，$D_\infty(\mu\|\xi)$は有限で，MLランダムな左c.e.実数となる．

直感的な説明
ξは十分汎用的なので，μとμに似た分布両方をdominateする必要がある．
その違いがいくらでも遅く起こるので，その誤差の収束は最も遅く，MLランダムである必要がある．
また十分汎用的であれば，どんなξでもほぼ同じ速さで収束する．

まとめ

- ランダムネスの理論は確率の公理化，データの計算論的な理解などを目指して始まった
- ランダム性は典型性(検定)，予測不可能性(マルチンゲール)，圧縮不可能性(複雑性)などにより特徴づけられる．(積分検定などもある．)
- MLランダムな左c.e.実数は，停止確率，Solovay完全性など多くの性質を持つ面白い実数のクラス．近似の収束が遅い実数．
- MLランダムな左c.e.実数は，学習の文脈で誤差の和として現れる．おそらくそのような例は他にもあると思われる．