計算可能な予測の収束速度

宮部賢志

明治大学数学科

2021 年 12 月 22 日 (水)

RIMS 共同研究（公開型）「証明と計算の理論と応用」

つぎのかずは

どんな列か？

- 円周率 π
- 自然対数の底 e
- 独立に確率 $\frac{1}{3}$ で 1 が出るランダムな列
- 奇数番目は必ず 1 で，偶数番目は独立に確率 $\frac{1}{4}$ で 1 が出る部分的にランダムな列
- などなど

答えが存在するか？

統計的学習理論では,

最適な予測は存在しない．No Free Lunch 定理より．

Solomonoff の帰納的推論では,

最適な予測が存在する．計算可能ではないけれど．

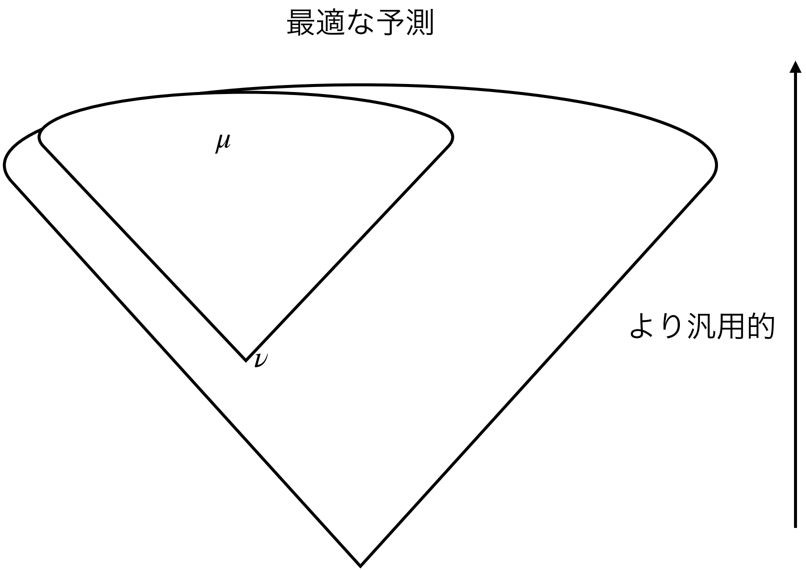
矛盾ではない．条件が異なる．また「最適」の言葉の意味も異なる．

ゴール

最適な (optimal) 予測とは，最も汎用的な (general) 予測である．

> より汎用的な予測であろうと追い求めていくと，あるところからはずっとこの性質 P を持つ．

ということがある．このとき，十分汎用的な予測は性質 P を持つ，という．
特に，収束速度に関する性質を調べたい．


最適な予測
μ
ν
より汎用的

目次

- 設定
- 汎用的予測
- 収束速度

問題設定

Cantor 空間: $\{0,1\}^{\mathbb{N}}$

μ: Cantor 空間上の未知の計算可能測度，モデル測度

$X = X_1X_2X_3\cdots \in \{0,1\}^{\mathbb{N}}$: μ に対してランダムな列

条件付き確率，$X_{<n}$ が与えられたときに X_n を予測したい：

$$\mu(k|X_{<n}) = \frac{\mu(X_{<n}k)}{\mu(X_{<n})}, \quad (k \in \{0,1\})$$

最適な予測

$\xi : \{0,1\}^* \to [0,1]$ が半測度 (semi-measure) であるとは,

$$\xi(\epsilon) \leq 1,\ \xi(\sigma) \geq \xi(\sigma 0) + \xi(\sigma 1)$$

ϵ は空文字.

$\xi : \{0,1\}^* \to [0,1]$ が c.e.(or lower semicomputable) であるとは, $\xi(\sigma)$ が σ に関して一様に下から計算可能に近似できること.

c.e. 半測度 ξ が最適 (optimal) とは, 任意の c.e. 半測度 μ に対して, ある $c \in \mathbb{N}$ が存在して, すべての $\sigma \in \{0,1\}^*$ で,

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

となることをいう. 最適な c.e. 半測度は存在する.

Solomonoff の結果

ξ: 予測測度

定理 1 (Solomonoff 1960s-70s)

ξ: 最適な *c.e.* 半測度，μ: 計算可能測度，X: μ ランダムな列，

$$|\xi(k|X_{<n}) - \mu(k|X_{<n})| \to 0,\ n \to \infty, k \in \{0,1\}$$

Martin-Löf ランダム性では不十分，2 ランダムなら十分．

(独り言：たぶん density randomness くらい．)

批判

計算可能でない予測の話であり，実際の機械学習の研究にはほとんど役に立たない.

目次

domination

ξ が ν を dominate するとは,

$$\exists c \in \mathbb{N} \forall \sigma . \nu(\sigma) \leq c \cdot \xi(\sigma)$$

Solomonoff の結果は，すべての計算可能測度を dominate する半測度は良い振る舞いをするということ．計算可能測度の集合上に，この関係から定まる順序を入れたとき，「良い振る舞いをする」とはどういうことか？

ゴール：より多くのモデル測度に対して速い収束をする

ξ は ν より汎用的な (general) 測度であると呼ぶことにする

KL-divergence

$\{0,1\}$ 上の測度 μ, ξ に対して，μ の ξ に対する KL-divergence を，

$$d(\mu||\xi) = \sum_{k\in\{0,1\}} \mu(k) \ln \frac{\mu(k)}{\xi(k)}$$

とする.

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

domination と収束

定理 2

ξ, ν に関して以下は同値.

- ξ が ν を *dominate* する
- すべての μ に対し，$D_\infty(\mu||\nu) < \infty \Rightarrow D_\infty(\mu||\xi) < \infty$
- ある $c \in \mathbb{N}$ が存在して，すべての μ で，$D_\infty(\mu||\xi) \leq D_\infty(\mu||\nu) + c$.

つまり，より多くの μ に対して，KL-divergence の意味で誤差和が有限であることと同値.

注意

- $D_\infty(\mu||\xi)$ は，μ の ξ に対する KL-divergence に一致する．
- dominate するなら絶対連続．絶対連続性については角谷の同値定理という古い結果がある．
- KL-divergence 以外の測度の距離でうまくいくかどうかはよく分からない．
- 計算可能性は仮定していない．純粋な測度論の結果．

目次

十分良い計算

Kolmogorov 複雑性は計算不可能なので，

$$f(n) \leq K(n) + O(1),\ \sum_n 2^{-f(n)} < 1$$

を満たす計算可能な関数 f は存在しない．しかし，

$$f(n) \leq K(n) + O(1)\ i.o.,\ \sum_n 2^{-f(n)} < 1$$

となる計算可能関数は存在し，Solovay 関数と呼ばれている．

十分汎用的な予測

十分大きな自然数 $n \in \mathbb{N}$ について性質 P が成り立つとは，ある自然数 $N \in \mathbb{N}$ が存在して，すべての $n > N$ で性質 P が成り立つこと．

十分汎用的な計算可能測度が性質 P を満たすとは，ある計算可能測度 ν が存在して，ν を dominate するすべての計算可能測度 ξ が性質 P を満たすことをいう．

- 計算可能な測度についての話ができる．
- Solomonoff の結果に相当する議論ができる．

収束速度

今回は性質として収束速度について議論したい．なぜか．

- 最適な c.e. 半測度の収束速度についてはきれいな解析ができない
- 計算可能測度 μ, ξ に対しては，$D_\infty(\mu||\xi)$ は左 c.e. 実数になる．その収束速度は Solovay 還元という立派な理論がある．

ML ランダム性

定理 3

μ: 計算可能測度，ξ: 十分汎用的な測度

$$D_{\infty}(\mu||\xi)$$

有限な値で，*Martin-Löf* ランダムな左 *c.e.* 実数.

注意：ルベーグ測度に対する ML ランダム性.

収束する実数列の中で，収束が最も遅いということ.

具体例

- $\ell_p(\mu,\xi) = \sum_{k\in\{0,1\}} |\mu(k) - \xi(k)|^p$
- $L_p(\mu,\xi) = \sum E_{X\sim\mu}[\ell_{p,X_{<k}}(\mu,\xi)]$

とおくと，$L_p(\mu,\xi)$ が ML ランダムになる p は存在すればただ 1 つ．それを $R(\mu,\xi)$ と書く．

定理 4

Dirac 測度 $\mu = \mathbf{1}_A$ の場合は，十分汎用的な ξ に対し $R(\mu,\xi) = 1$.

μ が *separated*($\inf_{k,\sigma} \mu(k|\sigma) > 0$) の場合は，十分汎用的な ξ に対し $R(\mu,\xi) = 2$.

分かること

- 十分汎用的な予測は適切な測度に収束する
- その収束速度は μ の性質によりほぼ一意に定まり，遅すぎたり速すぎたりはしない

証明のスケッチ

$D_\infty(\mu||\xi)$ の ML ランダム性の証明の流れ.

1. μ と似ているが少しだけ異なる測度 ν を作る
2. $D_\infty(\mu||\nu)$ が ML ランダム性を示す
3. ξ が ν を dominate するなら，$D_\infty(\mu||\xi) - D_\infty(\mu||\nu)$ が左 c.e. になることを示す
4. $D_\infty(\mu||\xi)$ の ML ランダム性を示す.

3 の性質も ν の構成に依存するように見えるが，反例があるわけではない.

μ の構成

z_n: 計算可能な正の有理数列で $s = \sum_n z_n < 1$ が ML ランダムなもの

$Z^\sigma \in \{0,1\}^{\mathbb{N}}$: $\sigma \in Z^\sigma$, $\mu(Z^\sigma) = 0$

$$\mu_n(\sigma) = \begin{cases} \mu(\sigma) & \text{if } |\sigma| \leq n \\ \mu(\tau)\mathbf{1}_{Z^\tau} & \text{if } |\sigma| > n,\ \tau = \sigma_{\leq n} \end{cases}$$

$$\nu = \sum_n z_n \mu_n + (1-s)\mu$$

割合 z_n 分は，n 桁目までは μ と同じで，それ以降は μ と特異になる 1 点に重みを置く

ML ランダム性

Radon-Nikodym 微分が,

$$\frac{d\mu}{d\nu} = \frac{1}{1-s}$$

で，s のランダム性から $D(\mu||\nu)$ のランダム性が得られる.

また，$\frac{d\mu}{d\nu}$ が定数であることから,

$$D(\mu||\xi) = D(\mu||\nu) + \frac{1}{1-s} \cdot D(\nu||\xi)$$

で，$D(\mu||\xi)$ のランダム性が得られる.

誤差

Dirac 測度 $\mu = \mathbf{1}_A$ の場合,

$$-\log(1 - \xi(A_n | A_{<n})) \approx K^h(n)$$

ここで, $K^h(n)$ は time-bounded Kolmogorov complexity.
誤差がある意味で完全に記述できる.

まとめ

- 「十分汎用的」の概念を導入することで，良い計算可能な予測の性質を議論する枠組みを提案した.
- 十分汎用的な予測の収束速度を議論した.

今後

- Bernoulli 測度に限定した場合どうなるか？ ⇒3 月の数学会にて.
- separated 測度で n 番目での誤差は評価できるか？
- 多項式時間でできるか？
- 収束が計算可能に抑えられるのはどのような場合か？
- 一般空間上で有界閉集合にラベルがある場合はどうか？

終わり