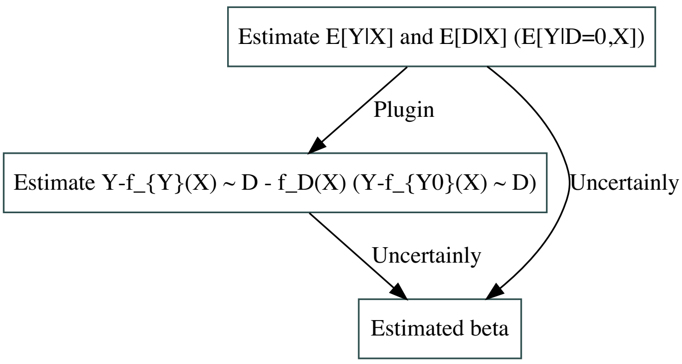
Adaptivity Property

機械学習の経済学への応用

川田恵介

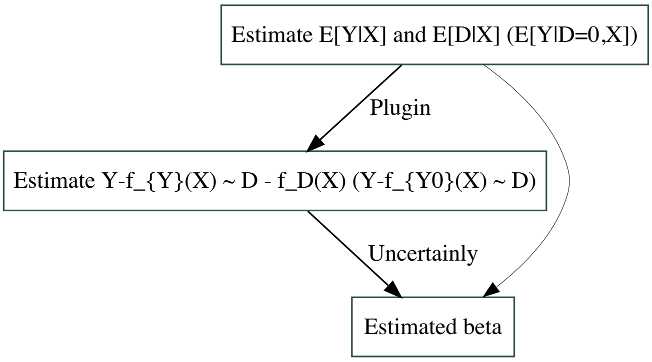
イメージ

- 機械学習を応用しても、“OLS”(FWL 定理) でも、



アプローチ

- Nuisance function の推定誤差の転嫁を減らす
  - 漸近分布を計算する際に” 無視できる” に!!!!



別のアプローチ

- $f_Y(X), f_D(X)$ の推定誤差をそもそも減らす｜収束速度を上げる
  - できれば苦労しない、、、
  - 正しいモデルを推定できるのであれば、OLS は理想的な収束速度
  - 誤定式であれば収束しない
- 機械学習による推定値は一般に収束が遅いので、それを**補正**する

Concept: 収束速度

- 収束速度:

$$\underbrace{n^a}_{\to\infty} \times \underbrace{(f_Y(X) - E[Y|X])}_{\to 0}$$

  - どの a に負けるか
- 正しいモデル
  - $a < 1/2 \to 0$
  - $a > 1/2 \to \infty$

- $a = 1/2 \rightarrow Normal$

Concept: Oracle estimator

- 重要な参照点

$$\beta_N^{Oracle} \in \arg\min \sum [Y - E[Y|X] - \beta \times [D - E[D|X]]]^2$$

- Oracle Estimator の呼称例 (Nie and Wager 2021)
  - $E[Y|X], E[D|X]$ はデータに非依存 (非確率変数)
  - $\lim_{N\rightarrow\infty} \sqrt{N}[\beta_N^{oracle} - \beta_0] \rightarrow Normal$

Adaptive Property

- Adaptive Property の呼称例 (Chernohzukov and Fernandez-Val)
- $\lim \sqrt{N}(\beta_N - \beta_0)$ の振る舞いに焦点を当てる
  - 経済学における" 結論"(信頼区間など) が" 依拠"
- $\lim \sqrt{N}(\beta_N - \beta_N^{Oracle}) = 0$ であれば、漸近的に無視できる

OLS の性質

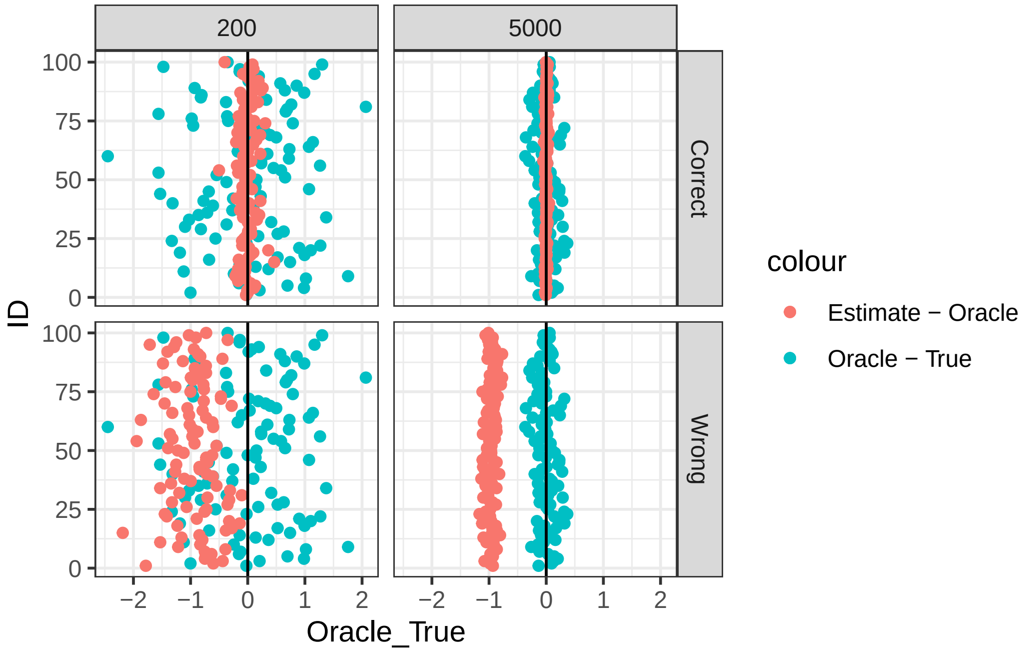
- $f_Y(X), f_D(X)$ が正しいならば: $\sqrt{N}[\beta_N - \beta_N^{Oracle}] \rightarrow 0$
  - $\sqrt{N}[\beta_N - \beta_0] = \sqrt{N}[\beta_N^{Oracle} - \beta_0]$
  - **推定値の漸近的な振る舞い**は、Oracle Estimator と一致
- $f_Y(X), f_D(X)$ が正しくなければ: $\sqrt{N}[\beta_N - \beta_N^{Oracle}] \nrightarrow 0$
  - 不一致

Numerical Example

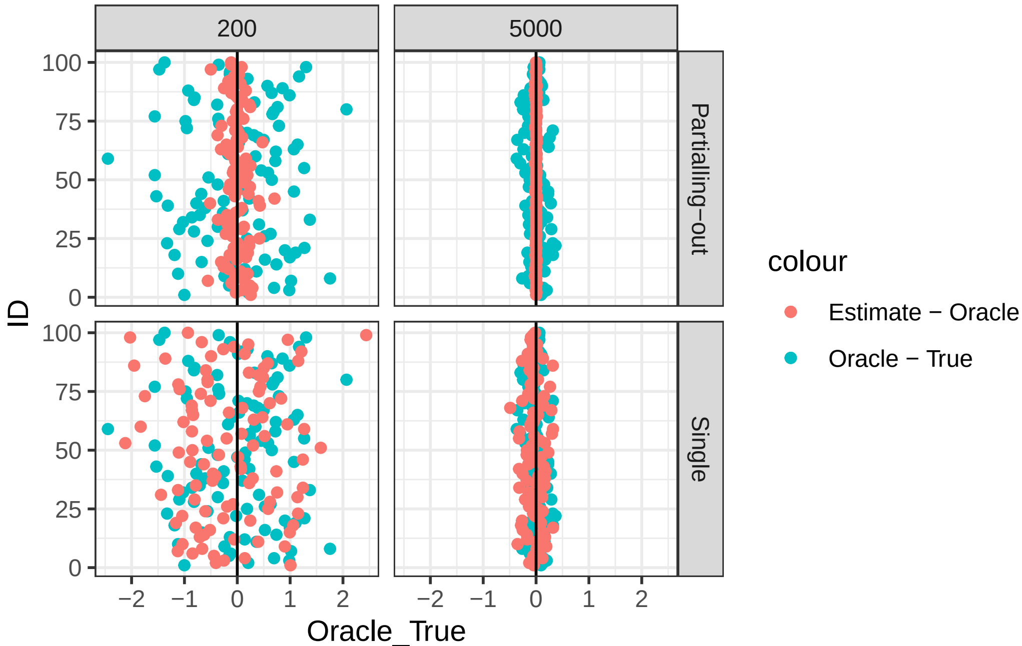
- $Y = D + X^2 + Normal(0, 5)$
- $E[D|X] = 0.5 - 0.4 \times I(X >= 1|X <= -1)$
- $Data = \{D, X, Z_1, .., Z_{190}\}$
  - $X, Z_1, .., Z_{190} \sim$ Uniform(-2,2)

OLS

- $E[Y|D,X] = \beta_D D + \beta_{X^2} X^2 + \beta_1 I(X \geq 1) + \beta_2 I(X \geq -1)$
- $E[Y|D,X] = \beta_D D + \beta_X X$



機械学習の応用



Partialling-out 推定

$$\tilde{\beta} \in \arg\min_{\beta} (Y_i - f_Y(X_i) - \beta \times [D_i - f_D(X_i)])^2$$

- $f_Y, f_D =$ 推定された予測関数

一階条件

$$0 = \sum_i \phi(Data_i, f_Y, f_D, \tilde{\beta})$$

$$\phi(Data_i, f_Y, f_D, \tilde{\beta}) = (D_i - f_D(X_i))[Y_i - f_Y(X_i) - \tilde{\beta}(D_i - f_D(X_i))]$$

- f_D, f_Y の $E[D|X], E[Y|X]$ からの" 局所的な乖離" について、$\tilde{\beta}(\phi)$ は"Robust"

Parametric SubModel

- 局所的な変動を定義するために以下の仮想的な予測モデルを導入 (Fisher and Kennedy 2021)

$$f_Y^t(X_i) = t \times f_Y(X_i) + (1-t) \times E[Y|X_i]$$

$$f_D^t(X_i) = t \times f_D(X_i) + (1-t) \times E[D|X_i]$$

- $t \in [0,1]$：ポイント ＝ スカラー
- 推定値 $(t=1)$ から真の値 (理想の予測モデル) $(t=0)$ の間を行きき

Taylor 近似

$$\sum_i \phi(W_i, f_Y, f_D, \tilde{\beta})$$

$$\simeq \sum_i \phi(W_i, E[Y|X], E[D|X], \tilde{\beta})$$

$$+ \underbrace{\sum_i \frac{\partial \phi(W_i, f_Y^t, f_D^t, \tilde{\beta})}{\partial t}\Big|_{t=0} \times [1-0]}_{\to 0?(faster\ than \sqrt{N}?)}$$

- $W_i = \{Y_i, D_i, X_i\}$

Taylor 近似

$$\sum_i \frac{\partial \phi(Data_i, f_Y^t, f_D^t, \tilde{\beta})}{\partial t}\Big|_{t=0}$$

$$= -\sum_i \underbrace{[Y_i - E[Y|X_i] - 2\tilde{\beta}(D_i - E[D|X_i])]}_{:=IF_D} \times \underbrace{(f_D(X_i) - E[Y|X_i])}_{ReducibleError}$$

$$-\sum_i \underbrace{(D_i - E[D|X_i])}_{:=IF_Y} \times \underbrace{(f_Y(X_i) - E[Y|X_i])}_{ReducibleError}$$

Neyman's ohthogonality

- $E[IF_Y] = 0$ (Neyman's orthogonality)
- f_Y が $E[Y|X]$ の一致推定量ならば、

$$\sum_i IF_Y \times (f_Y(X_i) - E[Y|X_i])$$

$$= \underbrace{(E[f_Y(X_i)] - E[Y|X_i])}_{Bias:\text{ゆっくり収束}} \times \underbrace{\sum_i IF_Y}_{\to 0},\ (Debiased)$$

$$+ \underbrace{\sum_i IF_Y \times (f_Y(X_i) - E[f_Y(X_i)])}_{?}$$

Cross fitting

- 一般に IF_Y と $(f_Y(X_i) - E[f_Y(X_i)])$ は、Sampling Uncertainly の影響を受ける
- 同じデータで双方を計算すると、相関する $\to$ 収束速度が遅くなる
- 例: Y と D が正に相関している場合、データ上である x について、偶然 D_i が上振れ

$$IF_Y = (D_i - E[D|X_i]) \uparrow \& f_Y(X_i) - E[f_Y(X_i)] \uparrow$$

- f_D を交差推定することで相関を断ち切り、収束速度改善

まとめ

- $IF_D \times (f_D(X_i) - E[f_D(X_i)])$ にも適用可能
- f_Y, f_D の $E[Y|X], E[D|X]$ への収束速度が $n^{1/4}$ 以上であり、Neyman's ohthogonality が成り立てば、$\lim n^{1/2}(\tilde{\beta} - \beta^{Oracle}) = 0$
  - $\lim_{n\to\infty} n^{1/2}(f_Y(X) - E[Y|X]) \to \infty$ であったとしても、
  - $\lim_{n\to\infty} n^{1/4}(f_Y(X) - E[Y|X]) \to 0$ であれば OK

まとめ: 非推奨アルゴリズム

$$\tilde{\beta} \in \arg\min_{\beta} (Y_i - f_{Y0}(X_i) - \beta \times D_i)^2$$

- 一階条件: $0 = \sum_i \phi(Data_i, f_Y, f_D, \tilde{\beta})$

$$\phi(W_i, f_Y, f_D, \tilde{\beta}) = D_i \times [Y_i - f_{Y0}(X_i) - \tilde{\beta} \times D]$$

- Neyman's ohthogonality を満たさない: $E[D] \neq 0$
  - Reducible error $E[Y|D = 0, X] - f_{Y0}(X_i)$ の影響をまともに受ける

Reference

Fisher, Aaron, and Edward H Kennedy. 2021. "Visually Communicating and Teaching Intuition for Influence Functions." *The American Statistician* 75 (2): 162–72.

Nie, Xinkun, and Stefan Wager. 2021. "Quasi-Oracle Estimation of Heterogeneous Treatment Effects." *Biometrika* 108 (2): 299–319.