Estimate Partial Linear Model

機械学習の経済学への応用

川田恵介

問題設定

- $E[Y|D,X]$ の**特定の**特徴 (Estimand) を推定
- 母集団において、以下を**仮定**

$$E[Y|D,X] = \underbrace{\beta_D}_{Interest} D + \underbrace{f(X)}_{Unkwnon}$$

- Partial Linear Model (Robinson 1988)
- OLS: $f(X) = \beta_0 + \beta_1 X_1 + .. + \beta_L X_L$ の一般化
- 定式化に誤りがあっても解釈可能 (後日)

推定アルゴリズム: Partialling-out (Robinson 1988)

1. $E[Y|X], E[D|X]$ を推定 $\rightarrow f_Y(X), f_D(X)$

- 機械学習 + 交差推定 (Chernozhukov et al. 2018)

2. $Y - f_Y(X)$ を $D - f_D(X)$ で回帰（定数項は含めない）
3. $f_Y(X), f_D(X)$ の推定誤差を**無視**して、信頼区間を RobustStandardError を用いて推定

- 機械学習を Nuisance function を推定するツールとして使用

直感

$$Y = \beta_D D + f(X) + \underbrace{u}_{Y-E[Y|D,X]\ (Mean\ Zero)}$$

$$E[Y|X] = \beta_D E[D|X] + f(X)$$

- 両辺を引くと

$$Y - E[Y|X] = \beta_D \times (D - E[D|X]) + u$$

OLS との関係性: FWL 定理

- FWL 定理より、OLS 推定は以下のアルゴリズムで書き下せる

1. $E[Y|X], E[D|X]$ を推定 $\rightarrow f_Y(X), f_D(X)$

- **同じ** OLS + **非交差推定**

2. $Y - f_Y(X)$ を $D - f_D(X)$ で回帰（定数項は含めない）
3. $f_Y(X), f_D(X)$ の推定誤差を**無視**して、信頼区間を RobustStandardError を用いて推定

論点

- なぜ OLS ではダメなのか？
  - 過剰適合 あるいは 誤定式化が生じるため
- なぜ機械学習を応用した他のアルゴリズムではダメなのか？
  - 収束（漸近）性質が悪い

OLS の問題点

KeyConcept: Misspecificaition

- 推定モデルの定式化 = 推定されうる関数の**集合**
- Misspecification: $\beta \in R_L$ をどう選んでも、$f_Y(X) \neq E[Y|X]$
  - β を増やせば、Misspecification は避けられるが､､､､

ざっくり性質

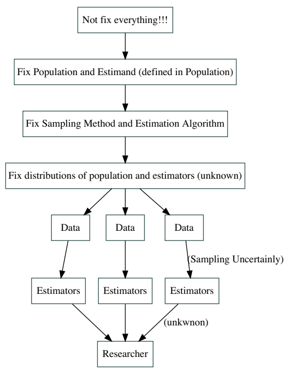
- "正しいモデル": Misspecification がない + β の数 $<<$ サンプルサイズ
  - 真のパラメータへの信頼区間を形成可能
- "間違ったモデル": Misspefication
  - 線形近似モデルへの信頼区間形成可能 $\neq$ 真のパラメータ

- 過剰なパラメータモデル
  - 過剰適合問題
  - 推定不可能、精度の大幅な悪化、信頼区間爆発

信頼区間

- 一般に推定値と真の値は、**一致し得ない**
  - 無限に大きいサンプルサイズが必要
  - 母集団への含意が不明瞭
- 代替的に信頼区間を用いて議論
  - **高い確率**で真の値を含んだ区間を得られるため

復習: Repeated Sampling Framework



KeyConcepts

- β_0 : 母集団におけるパラメータ (Fix but unknown)

- β_N：サンプルサイズ N のデータから計算された推定値 (Random with unknown distribution)

Key Property

- 大標本
  - Consistency：$\lim_{N\to\infty}(\beta_0 - \beta_N) \to 0$
    * 経済学では主張の根拠にしずらい
  - Asymptotic Normality：$\lim_{N\to\infty}\sqrt{N}(\beta_0 - \beta_N) \to N(0, \sigma^2)$
    * 信頼区間の近似計算の基盤
- 有限標本
  - Unbiasedness: $E[\beta_N] = \beta$

OLS + 正しいモデル

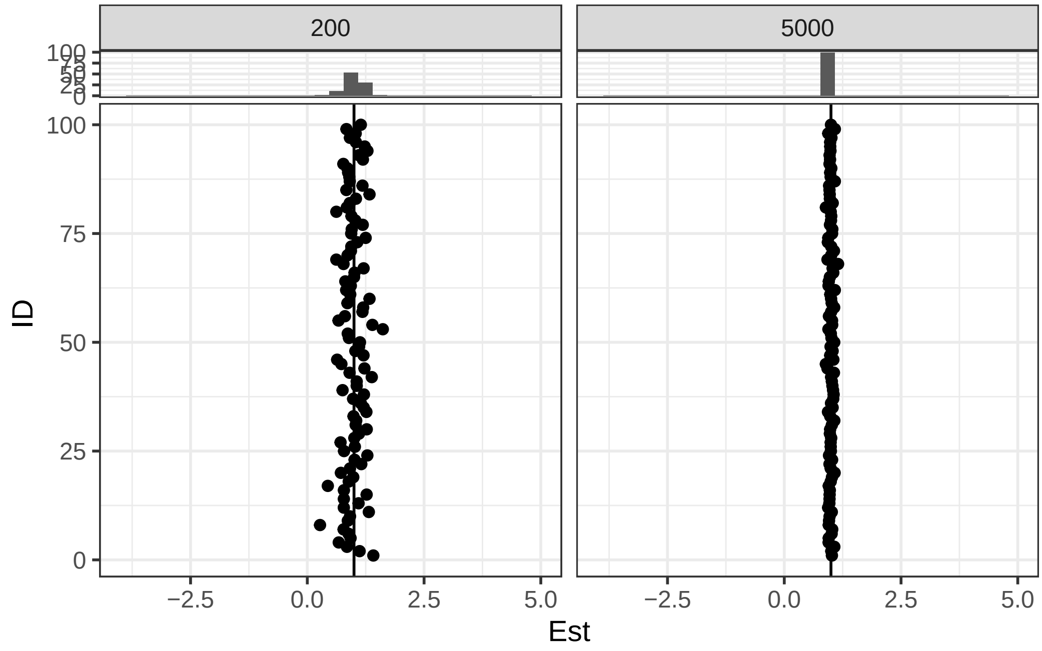
- ランダムサンプルデータであれば、Asymptotic Normality を満たす
  - Consistency, Unbiasedness も
- 独立して研究を行う大量の研究者をイメージ
  - 結論は全員異なる
  - サンプルが大きくなるにつれて、真の値に近い推定値を得る
  - **多くの**研究者が真の値を含む信頼区間を得る

Numerical Example

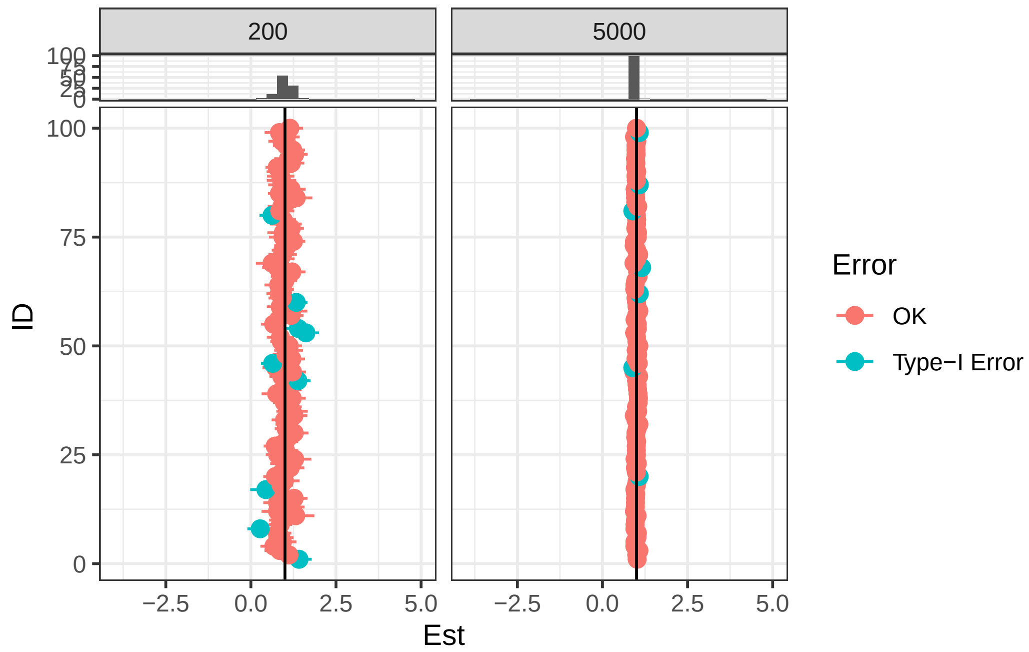
- $Y = D + X^2 + Normal(0, 5)$
- $E[D|X] = 0.5 - 0.4 \times I(X >= 1|X <= -1)$
- $Data = \{D, X, Z_1, .., Z_{190}\}$
  - $X, Z_1, .., Z_{190} \sim$ Uniform(-2,2)

PointEstimation

- $E[Y|D, X] = \beta_D D + \beta_{X^2} X^2$

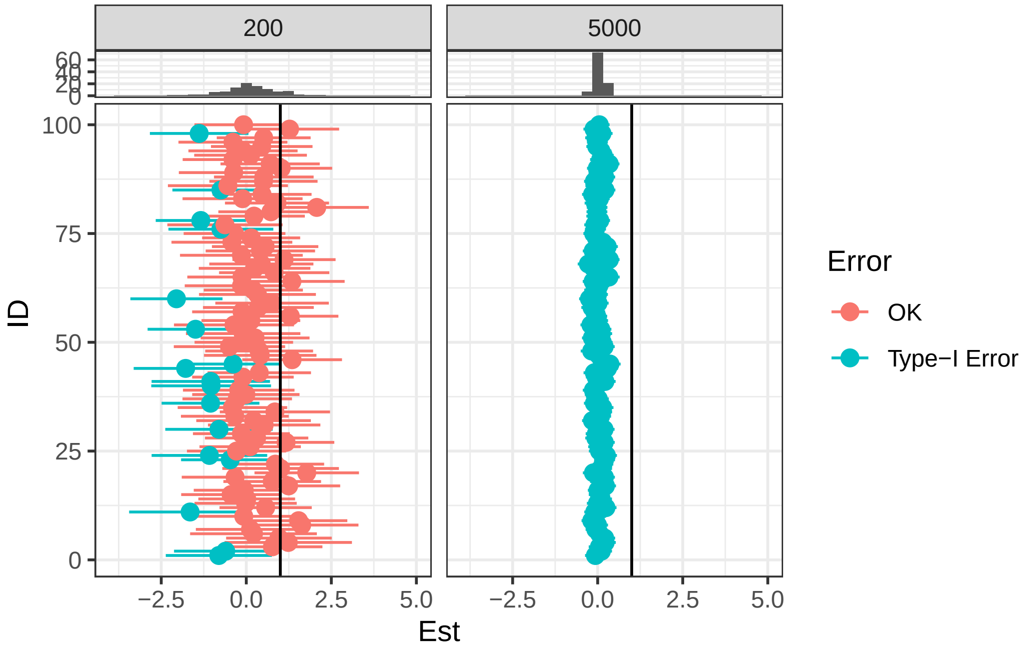


ConfidenceInterval



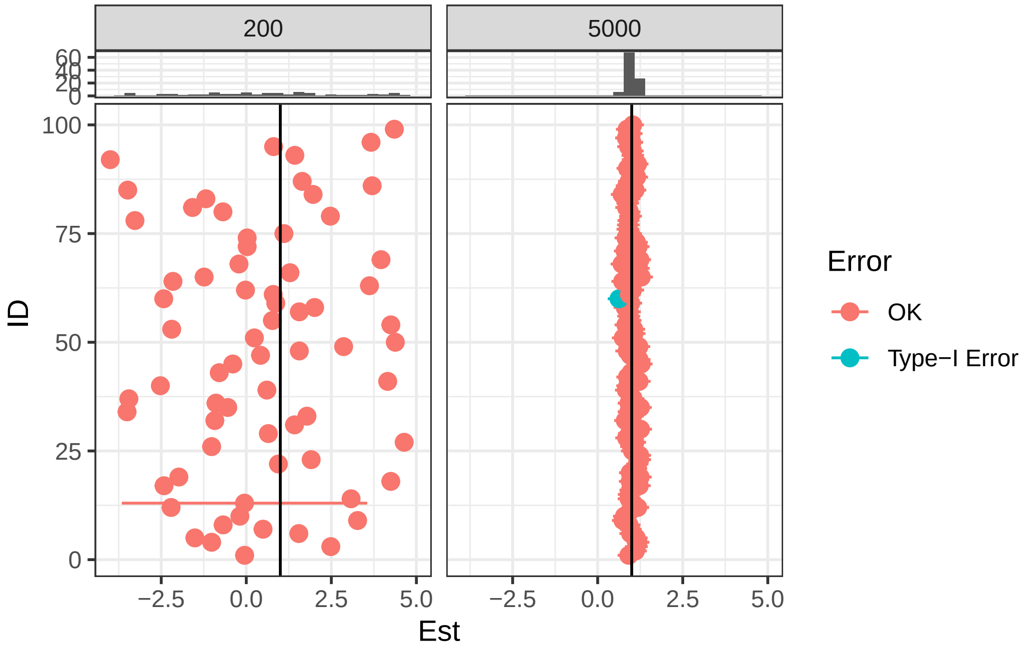
Wrong OLS

- $E[Y|D,X] = \beta_0 + \beta_D D + \beta_1 X$

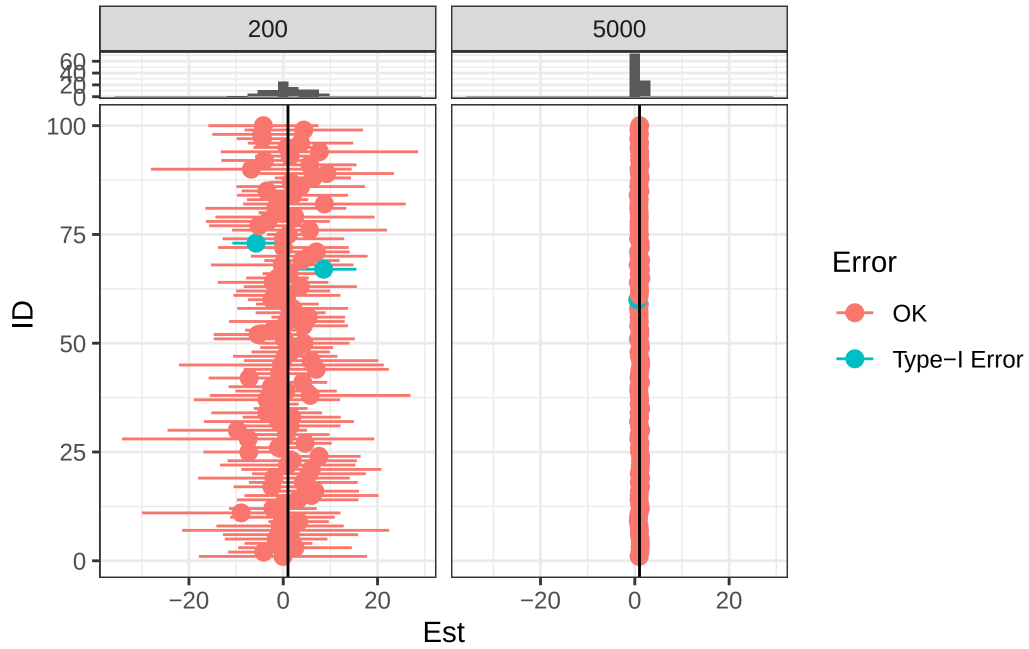


OverParametrization

- $E[Y|D, X, Z_1, .., Z_{190}] = \beta_D D + \beta_X X + \beta_{X^2} X^2 + \beta_{Z1} Z_1 + .. + \beta_{Z_{190}} Z_{190}$



OverParametrization



OLS まとめ

- OLS + 少数のパラメタ without miss-speficaition: 素晴らしいパフォーマンス
  - 非現実的？
- OLS + Miss-speficaition
  - Invalid Confidence Interval, Non-Consistent
- OLS + 大量のパラメータ
  - 過剰適合、標準誤差の爆発、信頼区間が広すぎる|計算できない

付録: RCT

- シンプルな線形モデルを用いた因果推論は、“TopJournal” でも散見される
- 重要な例外: D がランダムに決定
  - X を導入しても漸近性質は悪化しない (Lin 2013)
  - “Top Journal” における因果推論の多くは、RCT | 綺麗な自然実験を用いたものが多い

- 機械学習の応用にも有益

付録: Saturated Model

- $f(X) =$ ありうる全ての X の組み合わせについてのダミー変数を導入し、OLS 推定 (Angrist and Pischke 2009)
  - Saturated Model
- Partial Linear Model の範囲内では誤定式化は起きない
  - より一般的なケースでも、解釈可能 (後日)
- 過剰適合は大丈夫???

機械学習の応用

- 基本戦略「過剰に複雑なモデルを適度に単純化する」を踏襲
- Partialling-out を推奨
  - Nuisance 関数の推定に機械学習を使用

非推奨: MachineLearning as Modelling

- $E[Y|D,X]$ を直接推定
- 例: 以下を推定

$$\min\sum_i[Y-\beta_D\times D-\underbrace{\tilde{f}(X)}_{\text{十分に複雑な線形モデル}}]^2+\lambda\sum_l|\beta_l|$$

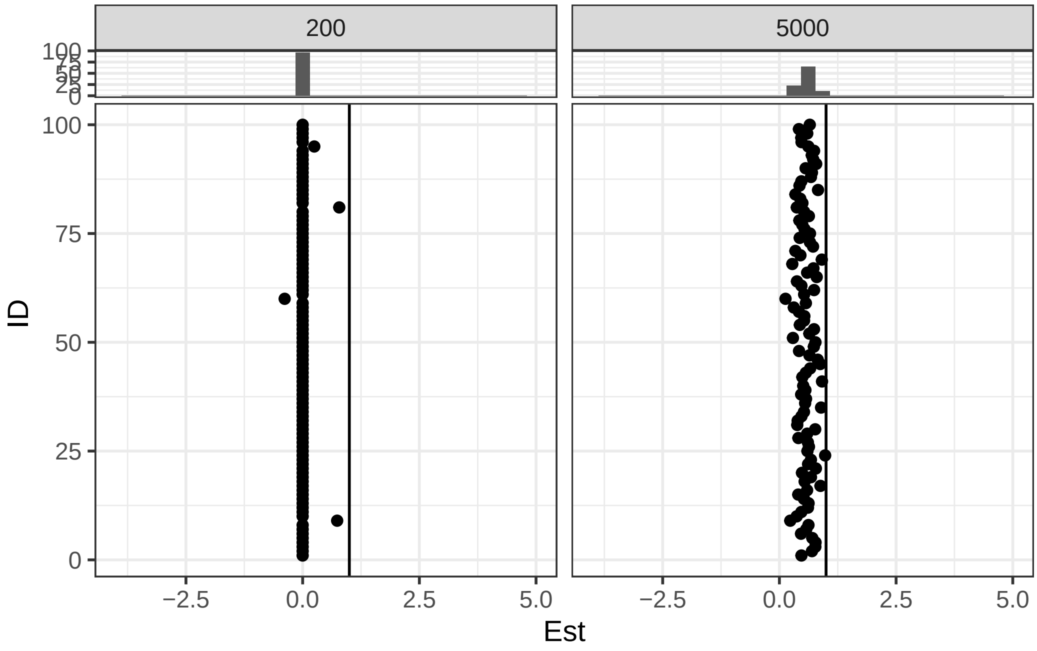
- LASSO

非推奨: Single MachineLearning

1. $D\in\{0,1\}$ を想定
2. $E[Y|D=0,X]$ を推定 $\rightarrow f_{Y0}(X)$
3. $Y-f_{Y0}(X)$ を D で回帰

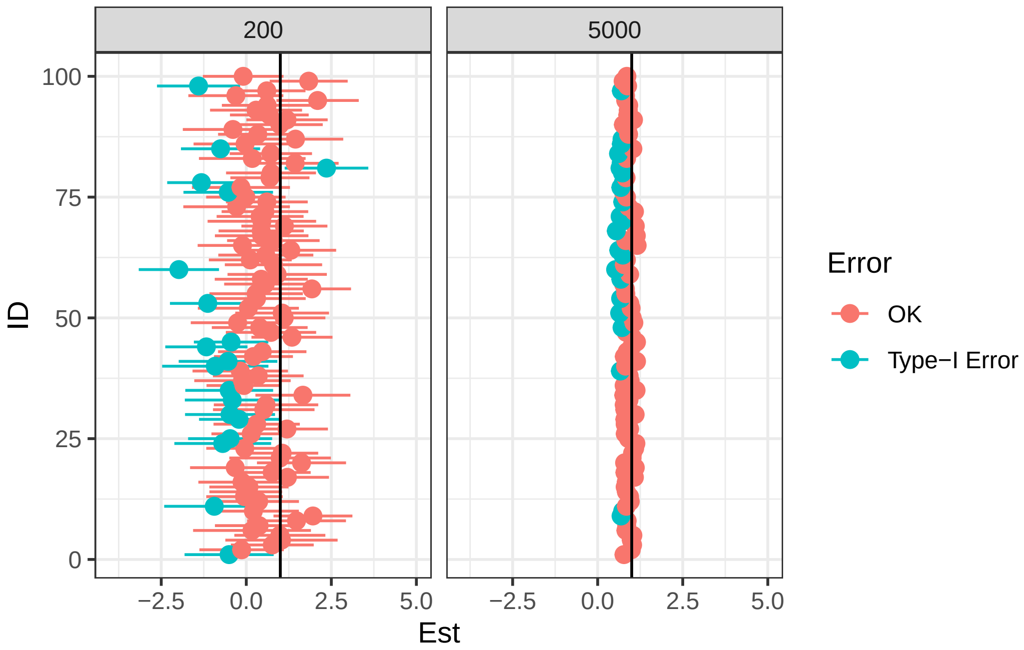
LASSO

- $f(X)=\beta_0+\beta_X X+\beta_{X^2}X^2+\beta_{Z_1}Z_1+..+\beta_{190}Z_{190}$



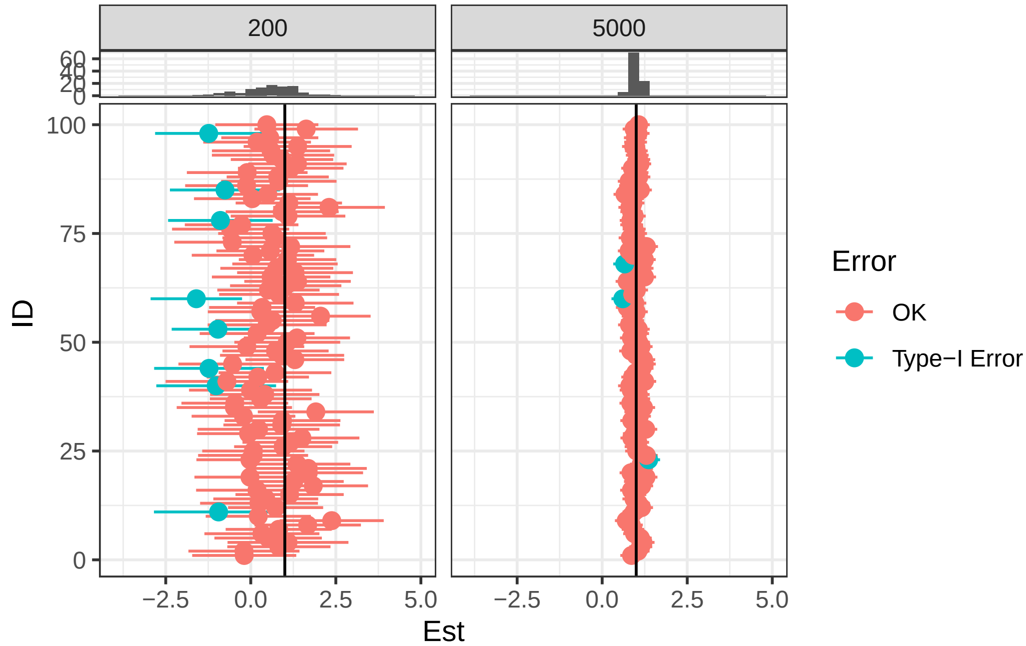
Single LASSO

- $f_{Y0}(X) = \beta_0 + \beta_X X + \beta_{X^2} X^2 + \beta_{Z_1} Z_1 + .. + \beta_{190} Z_{190}$



Double LASSO

- $f_Y(X) = \beta_0 + \beta_X X + \beta_{X^2} X^2 + \beta_1 I(X \geq -1) + \beta_2 I(X \geq -1) + \beta_{Z_1} Z_1 + .. + \beta_{190} Z_{190}$
- $f_D(X) = \beta_0 + \beta_X X + \beta_{X^2} X^2 + \beta_1 I(X \geq -1) + \beta_2 I(X \geq -1) + \beta_{Z_1} Z_1 + .. + \beta_{190} Z_{190}$



まとめ: 機械学習

- Population Risk (MSE) を最小化するように設計
  - パラメータの推論が主目的ではない
- 非常に緩やかな仮定の元で、一致性が成り立つアルゴリズムは複数存在
- 一般に（有限標本）バイアスが発生
- 収束が遅い、漸近正規性が成り立たなず、信頼区間計算が難しい
  - Bootstrap 法でも同じ

まとめ: 理想的な OLS

$$\beta_{interest} - \beta_N = \underbrace{\beta_{interest} - \beta_0}_{Identification}$$

$$+\underbrace{\beta_0-\beta_{N\to\infty}}_{=0}+\underbrace{\beta_{N\to\infty}-E[\beta_N]}_{=0}+\underbrace{E[\beta_N]-\beta_N}_{正規分布で近似}$$

- 一般に $\lim_{N\to\infty}\sqrt{N}(\beta_N-\beta_0)\to Normal$

まとめ: 誤定式化

$$\beta_{interest}-\beta_N=\underbrace{\beta_{interest}-\beta_0}_{Identification}$$

$$+\underbrace{\beta_0-\beta_{N\to\infty}}_{\neq 0}+\underbrace{\beta_{N\to\infty}-E[\beta_N]}_{=0}+\underbrace{E[\beta_N]-\beta_N}_{正規分布で近似}$$

- 一般に $\lim_{N\to\infty}\sqrt{N}(\beta_N-\beta_0)\to\infty$

まとめ: 複雑すぎるモデル

$$\beta_{interest}-\beta_N=\underbrace{\beta_{interest}-\beta_0}_{Identification}$$

$$+\underbrace{\beta_0-\beta_{N\to\infty}}_{=0}+\underbrace{\beta_{N\to\infty}-E[\beta_N]}_{=0}+\underbrace{E[\beta_N]-\beta_N}_{爆発の恐れ}$$

- 一般に $\lim_{N\to\infty}\sqrt{N}(\beta_N-\beta_0)\to Normal$ だが、、、

まとめ: LASSO

$$\beta_{interest}-\beta_N=\underbrace{\beta_{interest}-\beta_0}_{Identification}$$

$$+\underbrace{\beta_0-\beta_{N\to\infty}}_{"=0"}+\underbrace{\beta_{N\to\infty}-E[\beta_N]}_{\neq 0}+\underbrace{E[\beta_N]-\beta_N}_{SamplingUncetainly}$$

- 一般に $\lim_{N\to\infty}\sqrt{N}(\beta_N-\beta_0)\to\infty$

まとめ: Single MachineLearning

$$\beta_{interest}-\beta_N=\underbrace{\beta_{interest}-\beta_0}_{Identification}$$

$$+\underbrace{\beta_0-\beta_{N\to\infty}}_{"=0"}+\underbrace{\beta_{N\to\infty}-E[\beta_N]}_{\neq 0}+\underbrace{E[\beta_N]-\beta_N}_{SamplingUncetainly}$$

- 一般に $\lim_{N\to\infty} \sqrt{N}(\beta_N - \beta_0) \to \infty$

まとめ: PartiallingOut

$$\beta_{interest} - \beta_N = \underbrace{\beta_{interest} - \beta_0}_{Identification}$$

$$+ \underbrace{\beta_0 - \beta_{N\to\infty}}_{"=0"} + \underbrace{\beta_{N\to\infty} - E[\beta_N]}_{"\to 0"} + \underbrace{E[\beta_N] - \beta_N}_{\text{正規分布で近似}}$$

- 一般に $\lim_{N\to\infty} \sqrt{N}(\beta_N - \beta_0) \to Normal$

Reference

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1): C1C68. https://doi.org/10.1111/ectj.12097.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7 (1): 295–318.

Robinson, P. M. 1988. "Root-n-Consistent Semiparametric Regression." *Econometrica* 56 (4): 931954. https://doi.org/10.2307/1912705.