

# 後半戦

## 労働経済学 1

川田恵介

## 1 Stacking and Debiased Machine Learning

### 1.1 ここまでの内容

- 線型モデルをベースに議論
  - ▶ LASSO による予測
  - ▶ Double-selection による、パラメタの推論
- ここからは、より柔軟な枠組みを紹介

### 1.2 これからの内容

- より多様な機械学習/推定目標に対応可能な Debiased (Double/Target) Machine Learning を紹介
  - ▶ 講義中では、異質性分析への応用を紹介
  - ▶ 他の応用例: 操作変数法、パネルデータ(固定効果モデル) (Chernozhukov et al., 2024)、媒介分析 (Farbmacher et al., 2022)、組織評価 (付加価値推定) (Keele et al., 2023)
- 線型予測モデルを補完する機械学習の手法
  - ▶ Random Forest/Boosting/Stacking

### 1.3 Debiased Machine Learning

1. 研究課題に回答可能な母分布の関数を定義 (識別)
2. Neyman の直行条件を満たすように変換し、推定対象を導出 (one-step adjustment)
3. Nuisance 関数を機械学習等を用いて推定 (下準備となる推定)
4. Nuisance の推定結果を代入し、推定対象を推定 (推定本番)

### 1.4 Generalized Partial Linear Model

- Debiased Machine Learning の応用
- Robinson (1988) を一般化したモデル上で推定対象/nuisance を定義

$$E[Y \mid d, x] = \underbrace{\beta(x)}_{\text{Interest}} \times d + \underbrace{f(x)}_{\text{Nuisance (局外/厄介)}}$$

- 例:  $D$  が  $X$  内でランダムに決まっているのであれば、 $\beta(X) = E[y_i(1) - y_i(0) \mid X]$ 
  - ▶ 因果効果の異質性を解明する推定目標

## 1.5 One-step adjustment

- $\beta(x) = \beta$  ならば、以下の Population OLS の結果として推定可能

$$Y - E[Y \mid X] \sim D - E[D \mid X]$$

- $E[Y \mid X]/E[D \mid X]$  の推定に機械学習を活用可能
- $\beta(X)$  は、以下の母平均最小化問題として推定できる

$$(Y - E[Y \mid X] - D + E[D \mid X])^2$$

## 1.6 応用例: Shiba et al. (2021)

- 自然災害による被害は、高齢者の認知能力にどのような効果を持つのか?
  - ▶ Target population: 岩沼市(東日本大震災で大きな被害を受けた)
  - ▶  $D$  = 震災による住宅損壊,  $Y$  = 認知能力
    - $X$  = 震災前の人口学的属性/SES/健康状態/心理社会行動的要因

## 1.7 Shiba et al. (2021)

- 識別:  $X$  が同じであれば、Mean independence が成り立つ
  - ▶ 例えば、「住宅の損壊はランダムに発生」
- 推定: Debiased Machine Learning
- 結果:
  - ▶ 災害について特に脆弱なグループ ( $\beta_D(X)$  が大きくマイナス) は、以下の特徴を持つ
    - 高齢、未婚、独居、教育水準の低さ、無職である可能性が高く、ベースラインの健康問題

## 1.8 異質性分析の重要性

- データの事例数が増える中で、詳細な異質性分析が可能になっている
  - ▶ 重要な政策/学術的含意を持つ
  - ▶ コミュニケーションを改善できる可能性
    - 「去年と今年で”平均的な収入”があまり変化していない」のみを発信すると、“全員”の収入が変化していないと勘違いされる可能性がある

## 1.9 データ主導の重要性

- 伝統的な異質性分析では、研究者が”重要そうなサブグループ”を設定してきた
  - ▶ 分析に必要な時間の増大
  - ▶ 不適切な分析: 「“こっそり”膨大なサブグループ分析を行い、“好ましい結果”のみを報告する」 (Cherry pick) を誘発する恐れ
- 伝統的には、最低限のサブグループ分析 (例: 男女別分析) に止める傾向もあった

## 1.10 まとめ

- Debiased Machine Learning は、Double-selection の一般化
  - ▶ 利用できる機械学習: OLS や LASSO 以外の、有力な予測モデルの推定方法 (Random Forest, Boosting, Deep Learning 等) も活用可能
  - ▶ 活用できる推定対象: Double-selection は現状、OLS や Two stage least square (Chernozhukov et al. (2015)) 等にはしか応用できない

## 2 Reference

### Bibliography

- Chernozhukov, V., Hansen, C., & Spindler, M. (2015). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review*, 105(5), 486–490.
- Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024, ). Applied Causal Inference Powered by ML and AI. <https://arxiv.org/abs/2403.02467>
- Farbmacher, H., Huber, M., Laff ers, L., Langen, H., & Spindler, M. (2022). Causal mediation analysis with double machine learning. *The Econometrics Journal*, 25(2), 277–300.
- Keele, L. J., Ben-Michael, E., Feller, A., Kelz, R., & Miratrix, L. (2023). Hospital quality risk standardization via approximate balancing weights. *The Annals of Applied Statistics*, 17(2), 901–928.
- Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56, 931–954.
- Shiba, K., Daoud, A., Hikichi, H., Yazawa, A., Aida, J., Kondo, K., & Kawachi, I. (2021). Heterogeneity in cognitive disability after a major disaster: A natural experiment study. *Science Advances*, 7(40), eabj2610.