Debiased Machine Learning

川田恵介

1 一般的な命題

1.1 一般化

- 全スライドの議論は、一般的な推定対象に適用できる (Laan and Rose, 2011, , Chernozhukov et al. (2018) , Chernozhukov et al. (2022))
  - 「推定対象が低次元のパラメタであり、モーメント条件が母分布について微分可能」であれば、確立されている
  - Double/Debiased machine learning や Targeted machine learning (Laan and Rose, 2011) と呼ばれる

1.2 推定対象の一般的な表現

- 未知の nuisance 関数 g を含む、(既知の)モーメント条件として、推定対象 β を定義する。

$$E[m(W, \beta, g)] = 0$$

- W = 変数 (例: $W = [Y, D, X]$)
- 最小化問題の一階条件と解釈しても良い

$$\beta \in \arg\min L(W, \beta, g)$$

1.3 例: Partial Linear Model

- 以下の最小化問題として定義

$$\beta \in \arg\min \underbrace{E\left[(Y - g_Y(X) - \beta \times (D - g_D(X)))^2\right]}_{=L(\beta,g)}$$

- 一階条件より、以下は同値の定義

$$0 = E\left[\underbrace{(D - g_D(X)) \times (Y - g_Y(X) - \beta \times (D - g_D(X)))}_{=m(W,\beta,g)}\right]$$

1.4 一般的な命題

- 以下を仮定 (Chernozhukov et al., 2018)

1. Neyman's orthogonality を満たす: $\partial E[m(W, \theta, \tilde{g})]/\partial g \mid_{\tilde{g}=g} = 0$
2. 事例数について、推定された nuisance 関数は真の関数に収束し、かつ収束速度が、**事例数**$^{-1/4}$ よりも早い

- ある程度の事例数で、母平均を近似できる

3. nuisance 関数が交差推定されている

1.5 一般的な命題

- + regularity conditions を満たす場合、
- 推定値の分布は、漸近的に (**事例数**$^{-1/2}$ 以上の速度で) バイアスのない正規分布に収束する
  - 信頼区間を(解析的/Bootstrap)で近似計算できる

1.6 VS OLS

- nuisance 関数を OLS で推定するとすると
  - 誤定式化がない場合: 一致推定量かつ収束速度は**事例数**$^{-1/2}$
    - 多くの機械学習の推定手法よりも、早い速度で収束する
  - 誤定式化が存在する場合: 一致性を満たさない
    - より現実的な状況であり、Stacking が推奨される理由
- 相場観: 後者の可能性を重視

1.7 収束速度への要求

- 機械学習等で推定した場合の収束速度が、**事例数**$^{-1/4}$ よりも早いことを厳密に保証することは依然として困難
  - 相場観: 「誤定式化がない」ことを仮定し、OLS で推定するよりもマシ
    - ただし、OLS を含めた Stacking を用いることを推奨

1.8 Neyman's orthgonality の導出

- 母分布について微分可能な推定対象について、Neyman's orthogonality を満たす定式化を解析的に導出できる
  - わかりやすい解説は、Hines et al. (2022) を推奨
  - 微分できない推定対象への拡張も議論 (Hirano and Porter, 2012; Park, 2024)
  - "自動化的な導出"も議論されている (Chernozhukov et al., 2022; Chernozhukov, Newey and Singh, 2022; Luedtke, 2024; Laan et al., 2025)

2 応用例: AIPW

2.1 推定対象

- $D = 0/1$ を想定し、

$$\left(\underbrace{\tau(X)}_{E[Y|D=1,X]-E[Y|D=0,X]} \times f(X)\right)\text{の母集団における総和}$$

- 母集団全体での比率 $f(X)$ をウェイトに用いた、平均差の”平均”
- モーメント条件: $E[m(\theta, g)] = 0$

$$m(\theta, g) = \tau(X) - \theta$$

2.2 非推奨

- $m(\theta, g) = \underbrace{E[Y \mid 1, X] - E[Y \mid 0, X]}_{=\tau(X)}$ について

1. $E[Y \mid d, X]$ を推定
2. $m(\theta, g)$ のデータ上の平均値を導出

- Neyman’s orthgonality を満たさないので、step 1 の推定誤差の影響がθ の推定に無視できない影響を与える

2.3 推定対象の書き換え

- Neyman’s orthogonality を満たすように書き換える

$$m(\theta, g) = \tau(X)$$

$$+\frac{D(Y - E[Y \mid 1, X])}{f(D = 1 \mid X)} - \frac{(1 - D)(Y - E[Y \mid 0, X])}{f(D = 0 \mid X)}$$

- Double Robust な定式化/Argumented Inverse Propensity Weight と呼ばれてきた

2.4 実装

```
library(tidyverse)

data("CPSSW9204", package = "AER")

Y <- CPSSW9204$earnings

D <- if_else(CPSSW9204$degree == "bachelor",1,0)

X <- model.matrix(~ gender + age + year, CPSSW9204)
```

```
X <- X[,-1]
```

2.5 実装

```
ddml::ddml_plm(
  y = Y, D = D, X = X,
  learners = list(list(fun = ddml::ols),list(fun = ddml::mdl_ranger)),
  silent = TRUE,
  shortstack = TRUE
) |> summary() # 推定対象 = Partial Linear Model
```

```
PLM estimation results:

, , nnls

             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.0155     0.0535  -0.291    0.771
D_r            5.6586     0.1143  49.487    0.000
```

```
ddml::ddml_ate(
  y = Y, D = D, X = X,
  learners = list(list(fun = ddml::ols),list(fun = ddml::mdl_ranger)),
  silent = TRUE,
  shortstack = TRUE
) |> summary() # 推定対象 = AIPW
```

```
ATE estimation results:

     Estimate Std. Error t value Pr(>|t|)
nnls     5.64      0.113    49.9        0
```

2.6 PLR VS AIPW

- $D = 0,1$ の場合、どちらも同じような研究対象に使われる
  - 例: D の平均効果の推定
- 厳密には、研究対象が異なる
  - 平均効果を算出する際の、Weight が違う

2.7 PLR VS AIPW

- AIPW:

$$\tau(X) \times f(X)\text{の総和}$$

- Partial Linear Model:

$$\tau(X) \times \underbrace{f(D=1 \mid X) \times f(D=0 \mid X)}_{Overlap\ Weight}$$

$\times f(X)$の総和

2.8 性質

- $\tau(X)$ または $f(D=1 \mid X)$ が均質な場合、Partial Linear Model と AIPW は同じ値を定義
  - ほとんどの経済現象で異なる
- Partial Linear Model は、D の分布に偏りがないサブグループ ($f(D=1 \mid X)$ が 0.5 に近い)を重点的に反映
  - 直感的な解釈が難しい (解釈の試みとしては、Zhou and Opacic (2022) など)

2.9 性質

- AIPW の解釈は容易: 平均差の”単純”平均
  - $f(1 \mid X)$ の偏りが激しい場合、“oracle”推定においても信頼区間が爆発的に増加
    - 練習問題: なぜ?
  - $f(1 \mid X) = 0$ ないし 1 が存在すれば、AIPW は定義不可能
    - 比較研究として、“無理筋”であることに注意
      - 例: 男性と女性の政治的指導者を、日本やアメリカ、フランス内で比較

2.10 Overlap/Positivity

- Estimand: $\tau(X)$ の平均値を”定義する”ためには、母集団において $f(1 \mid X) \in (0,1)$ である必要がある
  - $f(1 \mid X) \in [0,1]$ ではないことに注意
    - 因果推論の文脈では Overlap/Positivity の仮定と呼ばれる
- 満たされない場合、 比較研究として”無理筋”
  - 例: 男性と女性の政治的指導者を、日本やアメリカ、フランスで比較

2.11 対処

- 比較不可能なグループは、極力分析計画の時点で排除
- Partial Linear model を用いる (解釈の難しさを受け入れる)
- AIPW + overlap の弱さへの対処を併用
  - Dorn (2025) とその引用文献参照

3 2 段階機械学習への拡張

3.1 推定対象

- Partial linear model で定義される X の”関数” $\beta_D(X)$

$$E[Y \mid D, X] = \beta(X) \times D + f(X)$$

- 以下を最小化する”関数”として、定義する

$$E[(Y - \beta(X) \times D - f(X))^2]$$

  - 有限のパラメタではなく、関数であることに注意

3.2 研究対象

- $D = 0, 1$ の場合、$\beta(X) = \tau(X) = E[Y \mid 1, X] - E[Y \mid 0, X]$
- もし D が、X 内でランダムに決まっているのであれば、 $\beta(X) = X$ 内での平均効果
  - 個人因果効果 $\tau_i = Y_i(1) - Y_i(0)$ の最善の予測値!!!
  - 練習問題: 個人の結果 Y_i の最善の予測値は、 $E[Y \mid X]$ であることと同じ理由

3.3 実装例

- 代表的なアルゴリズムは、Causal Forest (Wager and Athey, 2018; Athey, Tibshirani and Wager, 2019)
  - 紹介論文 (Athey and Wager, 2019) , Causal ML 15 章

3.4 Get Started: Education premium

```
set.seed(111)
library(tidyverse)

data("CPSSW9204", package = "AER")

Y <- CPSSW9204$earnings

D <- if_else(CPSSW9204$degree == "bachelor",1,0)

X <- model.matrix(~ gender + age + year, CPSSW9204)

X <- X[,-1]
```

3.5 Get Started: Education premium

```
library(SuperLearner)

Model_Y <- SuperLearner(
```

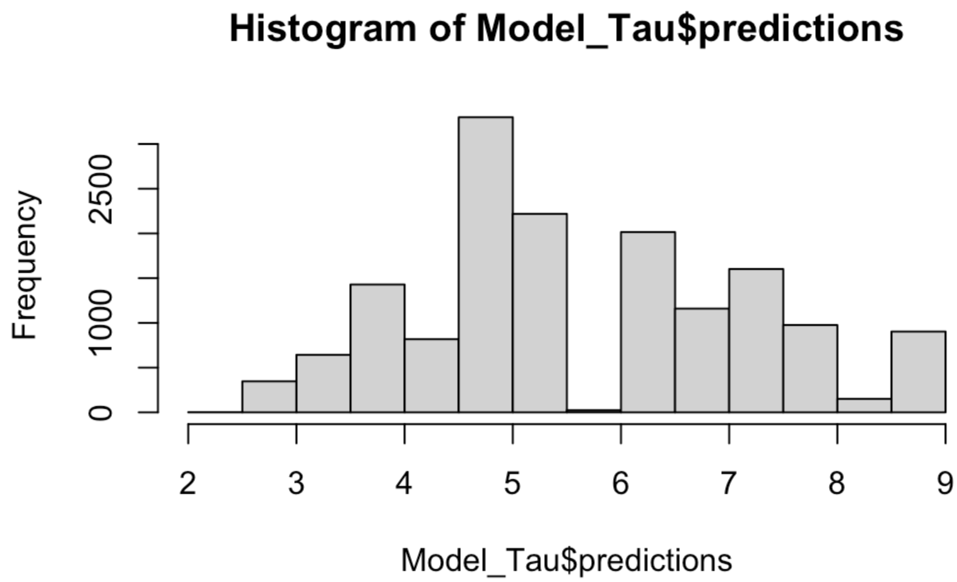
```r
  Y = Y,
  X = X,
  newX = X,
  SL.library = list(
    "SL.lm",
    "SL.ranger"
  ))

Model_D <- SuperLearner(
  Y = D,
  X = X,
  newX = X,
  SL.library = list(
    "SL.lm",
    "SL.ranger"
  ))
```

3.6 Get Started: Education premium

```r
Model_Tau <- grf::causal_forest(
  X = X,
  W = D,
  Y = Y,
  Y.hat = Model_Y$SL.predict,
  W.hat = Model_D$SL.predict)

hist(Model_Tau$predictions)
```



3.7 Get Started: Education premium

```
# A tibble: 5 × 4
  genderfemale   age year2004 Prediction
         <dbl> <dbl>    <dbl>      <dbl>
1            0    33        1       8.88
2            0    33        1       8.87
3            0    33        1       8.85
4            0    33        1       8.85
5            0    33        1       8.85
```

```
# A tibble: 5 × 4
  genderfemale   age year2004 Prediction
         <dbl> <dbl>    <dbl>      <dbl>
1            0    25        0       2.48
2            0    25        0       2.50
3            0    25        0       2.51
4            0    25        0       2.51
5            0    25        0       2.54
```

3.8 推定対象

- Nuisance について Local Robust に再定義
- 以下を最小化する”関数”として、定義する

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

- R-learner と呼ばれる枠組み (Nie and Wager, 2021) の一例

3.9 推定方法

1. nuisance 関数を機械学習を用いて推定する
2. 母集団における誤差を最小化するように、機械学習等を用いて、複雑な関数を推定する

- Neyman's ortogonality は、2 段階目に機械学習を活用する場合でも、理論的性質を改善する (Foster and Syrgkanis, 2023)
- Causal Forest は 2 段階目を Random Forest で行う

3.10 他の選択肢

- 様々な手法が提案されている
  - Kennedy (2020) : AIPW による定義をベースとする手法
- $E[Y \mid D=1,X]$ と $E[Y \mid D=0,X]$ を個別に推定し、差をとる (T-learner)
  - 特に$E[Y \mid D,X]$ を単純なモデルで近似できる場合、有力な手法
- $\beta(X)$ の推定結果を Stacking の手法も議論される (CausalML 15 章)

4 補論: 優先順位付への応用

4.1 Treatment assignment

- どのような介入をどのような優先順位で行うべきか?
  - 個別因果効果の予測値は判断基準の一つ
  - 効果の"大きな"事例から優先的に介入を行う
  - 評価指標も開発される (Yadlowsky et al., 2025)
- 実証実験: Ida et al. (2022)

4.2 Effect VS Level

- 「"効果の大きさ"のみで優先順位をつける」ことを正当化する Social Welfare function は限定的
  - 「効果の総和最大化」が社会的に望ましいことを要求
- 他の指標としては、介入を受けられなかった場合の Y の予測値
  - 「放置すると状況が悪い事例」から優先する
    - 一般に$\tau(X)$ よりも Yの方が予測しやすい点も利点
- 比較研究も行われている (Haushofer et al., 2025; Athey, Keleher and Spiess, 2025)

4.3 他の基準

- Li et al. (2023) : no-harm criteria
- Kitagawa and Tetenov (2018) : Resource constraints
  - Kitagawa, Lee and Qiu (2025) : Regret aversion

4.4 Reference

Bibliography

Athey, S. and Wager, S. (2019) “Estimating treatment effects with causal forests: An application,” Observational studies, 5(2), pp. 37–51.

Athey, S., Keleher, N. and Spiess, J. (2025) “Machine learning who to nudge: causal vs predictive targeting in a field experiment on student financial aid renewal,” Journal of Econometrics, p. 105945.

Athey, S., Tibshirani, J. and Wager, S. (2019) “Generalized Random Forest,” The Annals of Statistics, 47(2), pp. 1148–1178.

Chernozhukov, V. et al. (2018) “Double/debiased machine learning for treatment and structural parameters.” Oxford University Press Oxford, UK.

Chernozhukov, V. et al. (2022) “Locally robust semiparametric estimation,” Econometrica, 90(4), pp. 1501–1535.

Chernozhukov, V., Newey, W. and Singh, R. (2022) “Automatic debiased machine learning of causal and structural effects,” Econometrica, 90(3), pp. 967–1027.

Chernozhukov, V. et al. (2022) “Riesznet and forestriesz: Automatic debiased machine learning with neural nets and random forests,” in International Conference on Machine Learning, pp. 3901–3914.

Dorn, J. (2025) “How Much Weak Overlap Can Doubly Robust T-Statistics Handle?.”

Foster, D.J. and Syrgkanis, V. (2023) “Orthogonal statistical learning,” The Annals of Statistics, 51(3), pp. 879–908.

Haushofer, J. et al. (2025) “Targeting impact versus deprivation,” American Economic Review, 115(6), pp. 1936–1974.

Hines, O. et al. (2022) “Demystifying statistical learning based on efficient influence functions,” The American Statistician, 76(3), pp. 292–304.

Hirano, K. and Porter, J.R. (2012) “Impossibility results for nondifferentiable functionals,” Econometrica, 80(4), pp. 1769–1790.

Ida, T. et al. (2022) Choosing who chooses: Selection-driven targeting in energy rebate programs.

Kennedy, E.H. (2020) “Towards optimal doubly robust estimation of heterogeneous causal effects,” arXiv preprint arXiv:2004.14497 [Preprint].

Kitagawa, T. and Tetenov, A. (2018) “Who should be treated? empirical welfare maximization methods for treatment choice,” Econometrica, 86(2), pp. 591–616.

Kitagawa, T., Lee, S. and Qiu, C. (2025) “Leave No One Undermined: Policy Targeting with Regret Aversion,” arXiv preprint arXiv:2506.16430 [Preprint].

Laan, L. van der et al. (2025) “Automatic Debiased Machine Learning for Smooth Functionals of Nonparametric M-Estimands,” arXiv preprint arXiv:2501.11868 [Preprint].

Laan, M.J. Van der and Rose, S. (2011) Targeted learning. Springer.

Li, H. et al. (2023) “Trustworthy policy learning under the counterfactual no-harm criterion,” in International Conference on Machine Learning, pp. 20575–20598.

Luedtke, A. (2024) “Simplifying debiased inference via automatic differentiation and probabilistic programming,” arXiv preprint arXiv:2405.08675 [Preprint].

Nie, X. and Wager, S. (2021) “Quasi-oracle estimation of heterogeneous treatment effects,” Biometrika, 108(2), pp. 299–319.

Park, G. (2024) “Debiased Machine Learning when Nuisance Parameters Appear in Indicator Functions,” arXiv preprint arXiv:2403.15934 [Preprint].

Wager, S. and Athey, S. (2018) “Estimation and inference of heterogeneous treatment effects using random forests,” Journal of the American Statistical Association, 113(523), pp. 1228–1242.

Yadlowsky, S. et al. (2025) “Evaluating treatment prioritization rules via rank-weighted average treatment effects,” Journal of the American Statistical Association, 120(549), pp. 38–51.

Zhou, X. and Opacic, A. (2022) “Marginal interventional effects,” arXiv preprint arXiv:2206.10717 [Preprint].