平均差の NonParametric 推定

異質性分析

川田恵介

Table of contents

1 平均差の機械学習による推定

- Chap 13 and 14 in CausalML 参照

1.1 条件付き平均差の近似モデル

- Estimand:
$$\theta_\tau(X) = E[Y|D=1,X] - E[Y|D=0,X]$$
の近似モデル $g_\tau(X)$
- 中心極限定理などにはこだわらず、近似精度 $E[(\tau(X) - g_\tau(X))^2]$ の最小化を目標とする

1.2 応用: 個人因果効果の予測

- 潜在結果の枠組み (Imbens 2022) を用いると、個人因果効果を定義できる
  - $D=0$ の世界線における結果 $Y_i(0)$ と $D=1$ の結果 $Y_i(1)$ の差 $\tau_i = Y_i(1) - Y_i(0)$
    - ∗ “合理的” 意思決定の基礎情報 (限界効果)
- 異なる世界線は原理的に観察不可能なので、τ_i を推定することは” 不可能”
  - 因果推論の根本問題

1.3 応用: 個人因果効果の予測

- 通常の枠組みに収められる
- $$E[(\tau_i - g_\tau(X_i))^2]$$

を最小にする関数 g_τ を推定する

- 理想の予測モデル:
$$g_\tau(X_i) = E_P[\tau_i|X_i]$$
- τ_i が観察できないので、直接的な推定はできない

1.4 分解

$$\tau_i - g_\tau(X_i) = \underbrace{\tau_i - E_P[\tau|X]}_{IrreducibleError}$$

$$+\underbrace{E_P[\tau|X] - g_{\tau,\infty}(X)}_{ApproximationError} + \underbrace{g_{\tau,\infty}(X) - g_\tau(X)}_{EstimationError}$$

1.5 分解: BLP

- $g_\tau(X) \sim \beta_0 + .. + \beta_L X_L$ の推定法は、前々回議論

$$\tau_i - g_\tau(X_i) = \underbrace{\tau_i - E_P[\tau|X]}_{IrreducibleError}$$

$$+\underbrace{E_P[\tau|X] - g_{\tau,\infty}(X)}_{ApproximationError \neq 0} + \underbrace{g_{\tau,\infty}(X) - g_\tau(X)}_{EstimationError \sim N(0,\sigma^2)}$$

- Approximation error が大きい可能性がある

1.6 分解: ここからの手法

$$\tau_i - g_\tau(X_i) = \underbrace{\tau_i - E_P[\tau|X]}_{IrreducibleError}$$

$$+\underbrace{E_P[\tau|X] - g_{\tau,\infty}(X) + g_{\tau,\infty}(X) - g_\tau(X)}_{\text{同時に最小化を目指す}}$$

1.7 T Learner (Künzel et al. 2019)

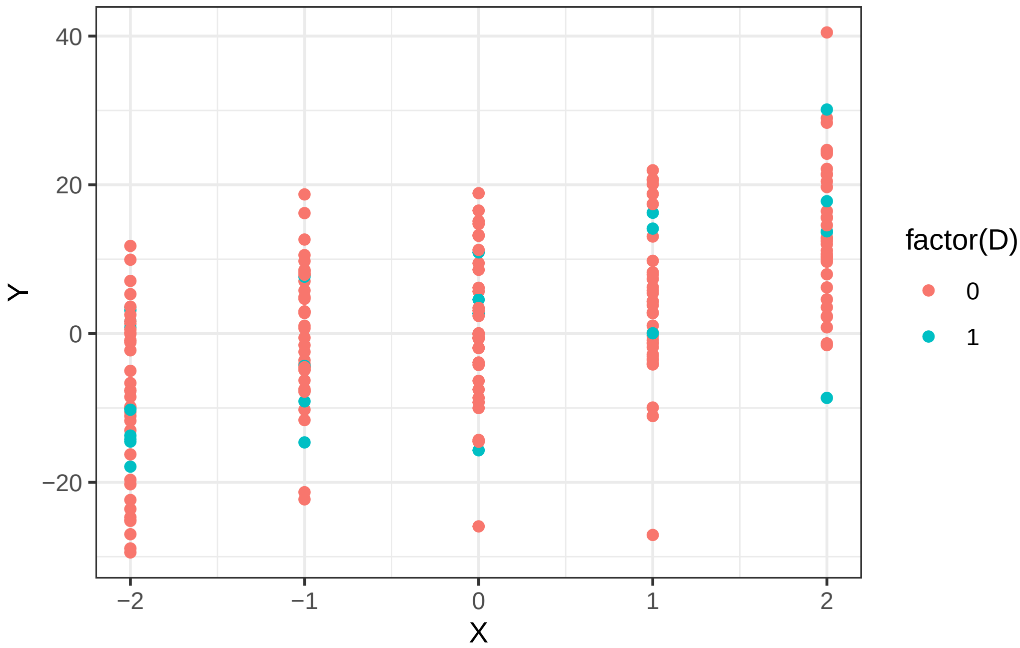
- Varian (2014)

1. $E[Y|1,X] \sim g_Y(1,X), E[Y|0,X] \sim g_Y(0,X)$ を教師付き学習で推定
2. $g_\tau(X) = g_Y(1,X) - g_Y(0,X)$

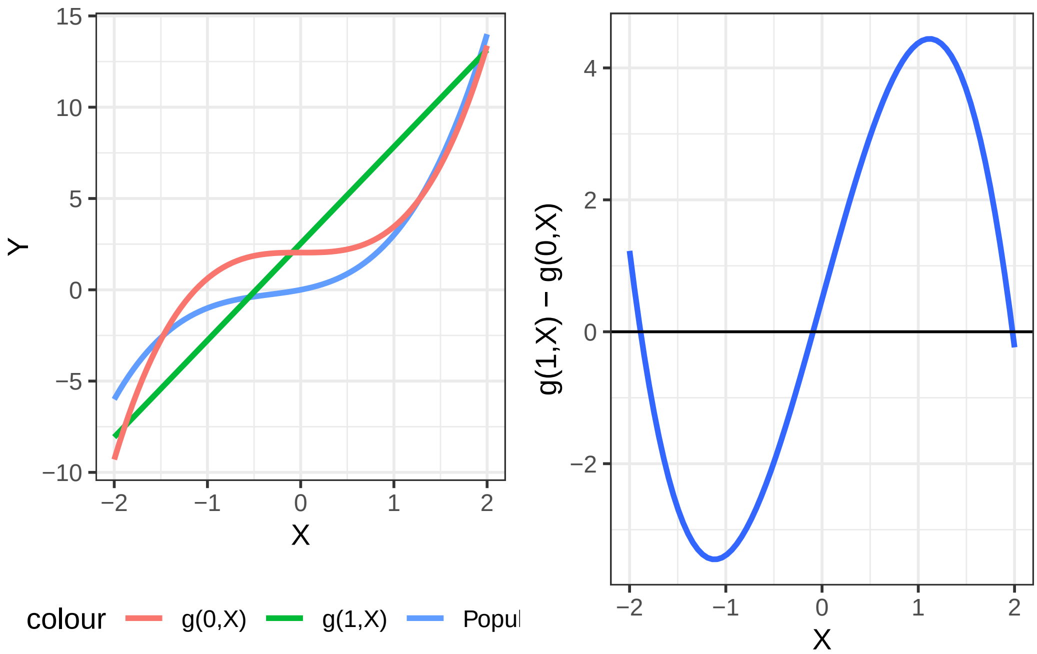
1.8 数値例

- $E[Y|D,X] = X + X^2 + X^3$
  - $D = 1/0$ のサブグループごとに LASSO を用いて、$Y \sim X + X^2 + X^3$ を推定
  - $\Pr[D == 1] = 0.1$
  - $\Pr[D == 0] = 0.9$

1.9 数値例



1.10 数値例



1.11 Regulization bias

- 教師付き学習は、“適切に単純化する” はずだが、過剰に複雑化させている
  - D の分布が偏っているケースにおいて、非常に深刻
- 問題点: 最適化問題の設定ミス
  - $\min E[(\tau_i - g_\tau)^2]$ ではなく、$\min E[(Y_i - g_\tau)^2]$ を目指して単純化が行われるため

1.12 DR-learner

- 母集団にモデルを適合できれば、$E[Y|1,X] - E[Y|0,X]$ を識別できる最適化問題を解く
- AIPW を応用: Psude-outcome を定義

$$\phi(X) = g_Y(1,X) - g_Y(0,X)$$
$$+\frac{D(Y - g_Y(1,X))}{g_D(X)} - \frac{(1-D)(Y - g_Y(0,X))}{1 - g_D(X)}$$

- $E[\phi(X)|X] = \tau(X)$ なので、$\phi(X)$ の近似モデルは平均差の良い近似モデルであることが期待できる

1.13 DR-learner (Kennedy 2020)

1. $g_Y(d, X) \sim E[Y|d, X], g_D(X) \sim E[D|X]$ を交差推定
2. Psude outcome $\phi(X)$ を計算
3. $\phi(X) \sim X$ をなんらかの Algorithm で推定
   - 機械学習も使用可能

1.14 R-learner (Nie and Wager 2021)

- Partialling Out の一般化

$$\min E[(Y - E[Y|X] - \tau(X) \times [D - E[D|X]])^2]$$

- PartialingOut した Y と D について、　母集団における誤差を最小化するように $\tau_P(X)$ を定義
  - 最も母集団に適合する $\tau(X)$ 関数を推定

1.15 R-learner (Nie and Wager 2021)

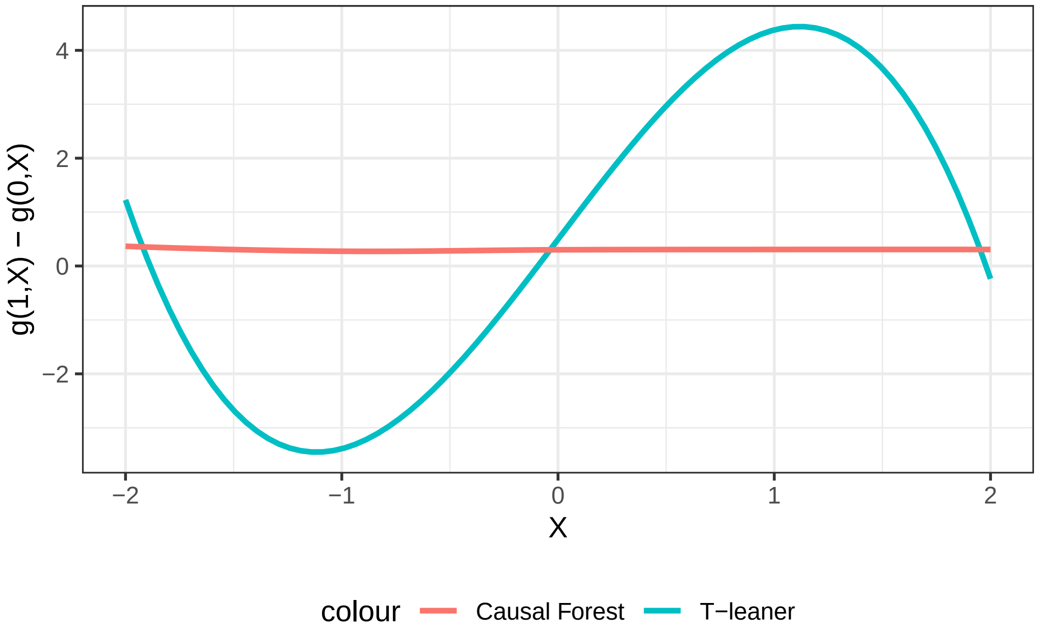
1. $g_Y(X) \sim E[Y|X], g_D(X) \sim E[D|X]$ を交差推定
2. $E_P[(Y - g_Y(X) - \tau(X) \times [D - f_D(X)])^2]$ を近似的に最小化するよう τ を推定

- 2 段階目にも、教師付き学習も活用可能
  - 前回は OLS

1.16 Causal Forest

- R learner の特殊ケース
- 2 段階目を RandomForest で実装
- 様々な工夫 (Chap 14.4 in CausalML 参照)

1.17 数値例: Causal Forest VS T-Learner



1.18 推定誤差

- 2 段階目も教師付き学習で推定した場合、一般に推定誤差 (母平均との乖離リスク) を推定することは困難
- Causal Forest は例外的に可能
  - ただし X の数は少ない必要がある
  - (川田の経験上)、信頼区間がかなり大きくなる

1.19 Example. Britto, Pinotti, and Sampaio (2022)

- 大規模 (整理) 解雇 $= D$ は、犯罪 $= Y$ を増加するか?
  - 司法データ (裁判日誌) と雇用データ (Employee-Employer matched data) を名寄せ !!! したパネルデータ (Difference-in-Difference)
    - ブラジルの男性について、平均効果 23%

1.20 Example. Britto, Pinotti, and Sampaio (2022)

- BLP 分析より、若年・短い勤続年数の労働者において大きい
- 98 % の労働者について、“有意” な効果が検出
  - 幅広い犯罪の拡大効果をもつ

1.21 補論: 予測性能

- 色々な提案
  - Rank Average Treatment Effect (Yadlowsky et al. 2021) は grf に実装ずみ

1.22 補論: Stacking

- $g_{\tau(X)}$ 自体も集計できる

1. サンプルを Prediction/Confirm データに分割
2. Prediction データのみを用いて、$g_{\tau,1}(X),..,g_{\tau,L}(X)$ を推定する
3. Confirm データを用いて、psud-outcome $\phi(X)$ を推定し、$\phi(X) \sim g_{\tau,1}(X) + .. + g_{\tau,L}(X)$ を OLS 回帰する

2 異質性の発見

- 機械学習により推定された関数について、安定的な統計的推論は難しい
  - Causal forest でも、X の数が多い、ないし Random Forest が適さない母集団である可能性がある
- “顕著な効果” をもつグループの存在を示すのみであれば、より選択肢が広がる

2.1 Motivaing example

X	Tau(X)	Pr[X]
1	10	0.1
2	1	0.8
3	-2	0.1

- 平均効果: $10 \times 0.1 + 1 \times 0.8 - 2 \times 0.1 = 0.7$

- 上位 10 % の平均効果 = 10
- 下位 10 % の平均効果 = -2

2.2 Group average treatment effect (Chernozhukov et al. 2018)

- Estimand: $E[Y|1,X]-E[Y|0,X]$ の予測値 $g_{\tau}(X)$ を" 前提" として、

$$E[\tau(X)|g_{\tau}(X)\in\{\tau_{-},\tau_{+}\}]$$

- 予測モデルは、因果効果の異質性を探索するための、シグナルとしてのみ用いる
  - 予測モデルについての統計的推論は" 不要"

2.3 Algorithm

1. サンプルを Prediction/Confirm データに分割
2. Prediction データのみを用いて、$g_{\tau}(X)$ を推定する
3. Confirm データを用いて、Nuisance 関数および GATE を推定する

- 交差推定を行う場合は、sequential test が必要 (Wager 2024)

2.4 Example. Fukai, Ichimura, and Kawata (2021)

- 2019/2020 年 $= D$ の就業状態 $= Y$ を比較
  - COVID の" 効果"
- 過去の就業状態や年齢など $= X$ について GATE を推定
  - COVID の" 効果" は一部の層に集中している

2.5 Example. SetUp

```r
set.seed(111)

library(tidyverse)
library(mlr3verse)
library(DoubleML)

Data = read_csv("Public/Data.csv") |>
```

```r
  filter(TradeYear == 2022)

Group = sample(
  1:2,
  nrow(Data),
  replace = TRUE
)

Y = Data$Price |> log()

D = Data$Reform

Z = Data |>
  model.matrix(
    ~ Size + Distance + Tenure ,
    data = _
  )

Z = Z[,-1]

X = Data |>
  model.matrix(
    ~ poly(Size,2) + poly(Distance,2) + poly(Tenure,2),
    data = _
  )

X = X[,-1]
```

2.6 Nuisance

```r
TaskPrediction = double_ml_data_from_matrix(
  y = Y[Group == 1],
  d = D[Group == 1],
  X = X[Group == 1,]
)

TaskConfirm = double_ml_data_from_matrix(
  y = Y[Group == 2],
  d = D[Group == 2],
```

```
  X = X[Group == 2,]
)

ModelPrediction = DoubleMLPLR$new(
  TaskPrediction,
  lrn("regr.cv_glmnet", s = "lambda.min"),
  lrn("classif.cv_glmnet", s = "lambda.min"),
  n_folds = 2
)

ModelConfirm = DoubleMLIRM$new(
  TaskConfirm,
  lrn("regr.cv_glmnet", s = "lambda.min"),
  lrn("classif.cv_glmnet", s = "lambda.min"),
  n_folds = 2
)

lgr::get_logger("mlr3")$set_threshold("warn")

ModelPrediction$fit(store_predictions = TRUE)

ModelConfirm$fit(store_predictions = TRUE)
```

2.7 Example. Prediction

```
FitTau = grf::causal_forest(
  X = Z[Group == 1,],
  W = D[Group == 1],
  Y = Y[Group == 1],
  Y.hat = ModelPrediction$predictions$ml_l,
  W.hat = ModelPrediction$predictions$ml_m
  ) |>
  predict(Z)
```

2.8 Example. GATE

```
Cutoff = FitTau$predictions[Group == 2] |> quantile(probs = c(1/3,2/3))
```

```
Cutoff
```

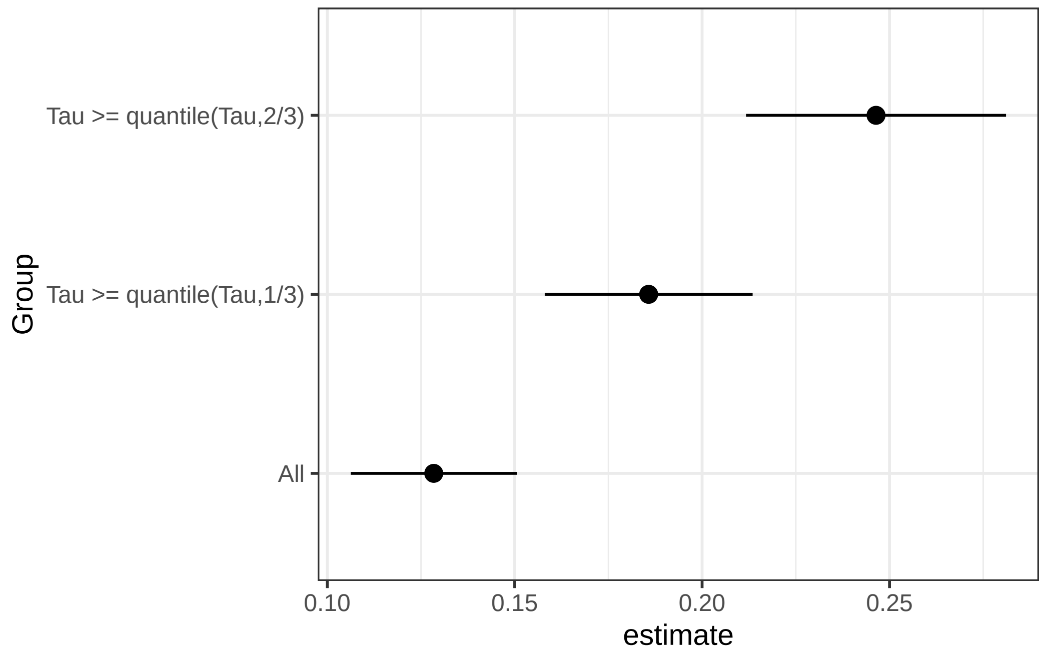
```
 33.33333%  66.66667%
0.03954255 0.13648334
```

```
Score = ModelConfirm$psi_b[,1,1]

estimatr::lm_robust(
  Score ~ 1,
    subset = FitTau$predictions[Group == 2] > Cutoff[2])
```

```
             Estimate Std. Error  t value     Pr(>|t|)  CI Lower CI Upper   DF
(Intercept) 0.2464015 0.01768526 13.93259 4.393406e-42 0.2117169 0.281086 1899
```

2.9 Example. GATE



2.10 Classification analysis

- Chernozhukov et al. (2018)
- どのようなサブグループについて、効果が大きい/小さいか?
- Estimand: ある背景変数 X_l について

$$E[X_l|g_\tau(X) > Q(q)] - E[X_l|g_\tau(X) \leq Q(q)]$$

–
$$Q(q) =$$

qth quantile of $q_{\tau}(X)$

2.11 Example. Classification analysis

```
Data |>
  filter(
    Group == 2
  ) |>
  select(
    Size,
    Distance,
    Tenure
  ) |>
  mutate(
    Group = FitTau$predictions[Group == 2] > Cutoff[2]
  ) |>
  gtsummary::tbl_summary(by = Group)
```

Characteristic	**FALSE**, N = 3,802	**TRUE**, N = 1,900
Size	35 (20, 60)	50 (30, 60)
Distance	6.0 (4.0, 9.0)	6.0 (4.0, 10.0)
Tenure	12 (7, 17)	36 (28, 43)

2.12 付論: Treatment effect risk

- Kallus (2023) により提案
- $$E_P[\tau(X)|\tau(X) \leq Q(q)]$$
  –
  $$Q(q) =$$

  qth quantile of $\tau(X)$

  – Estimand が完全に母集団上で定義できており、解釈がより明確

  – **Neyman の直行条件**を満たすので、g の推定誤差は漸近分布を計算する際に無視できる

Reference

Britto, Diogo GC, Paolo Pinotti, and Breno Sampaio. 2022. "The Effect of Job Loss and Unemployment Insurance on Crime in Brazil." *Econometrica* 90 (4): 1393–423.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2018. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India." National Bureau of Economic Research.

Fukai, Taiyo, Hidehiko Ichimura, and Keisuke Kawata. 2021. "Describing the Impacts of COVID-19 on the Labor Market in Japan Until June 2020." *The Japanese Economic Review* 72 (3): 439–70.

Imbens, Guido W. 2022. "Causality in Econometrics: Choice Vs Chance." *Econometrica* 90 (6): 2541–66.

Kallus, Nathan. 2023. "Treatment Effect Risk: Bounds and Inference." *Management Science* 69 (8): 4579–90.

Kennedy, Edward H. 2020. "Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects." *arXiv Preprint arXiv:2004.14497*.

Künzel, Sören R, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. "Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning." *Proceedings of the National Academy of Sciences* 116 (10): 4156–65.

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

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.

Wager, Stefan. 2024. "Sequential Validation of Treatment Heterogeneity." *arXiv Preprint arXiv:2405.05534*.

Yadlowsky, Steve, Scott D. Fleming, Nigam Haresh Shah, Emma Brunskill, and Stefan Wager. 2021. "Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects." In.