Heterogeneity analysis

川田恵介

2025-07-08

1 $\beta(X)$ の記述

1.1 研究目標の実際

- 多くの研究目標: $\beta(X) = E[Y \mid D = 1, X] - E[Y \mid D = 0, X]$ の”特徴把握”を要求
  - 学歴 D 間での所得 Y 格差の実態を明らかにする
  - 東大経済学研究科 D への進学が、30 歳時点での所得 Y に与える効果を明らかにする
- 明らかにしたい具体的な”特徴”を、意識的に定めることが重要

1.2 ここまで

- 推定対象として、$\beta(X)$ の平均 (Parial Linear Model/AIPW)と Nonparametric な推定 (Causal Forest)を学ぶ

1.3 比較

- Partial Linear Model/AIPW: Valid な統計的推論が可能だが、X との関係性について”一切の”情報を持たない
  - $E[D = 1 \mid X] \simeq 0$ な事例が存在する場合、Average “treatment effect” on treated が有効 (Section 5)
- Nonparametric な推定: 無限大の事例/計算/認知資源があれば理想的だが、非現実的
  - 例えば、Valid な統計的推論が難しい
    - 例外的な状況: X の数が少ない場合、Causal Forest は標準誤差を提供 (CausalML 14 章参照)
- 適切な妥協が有効な場合が多い

2 比較研究のフロー: 私案

2.1 大前提

- 「$\beta(X)$ の特徴」が分析対象であるとしっかり認識 (同質性 ($\beta(X) = \beta$) を仮定しているわけではない)

- Overlap していないことが明白な事例は、事前に除外
- 異質性を確認したい変数 Z は、研究目標に照らして、事前に決める

2.2 分析フロー

1. Overlap の仮定を確認: 例えば推定された $f(d \mid X)$ の分布を確認
2. 最もシンプルな集計値 ($\beta(X)$ の平均値) である、PLR/AIPW を示す

- Overlap の仮定に注意: 場合によっては Average “Treatment” Effect on Treated/Control を使用

3. GATE を示し、$\beta(X)$ の分布を要約する
4. BLP を推定し、Z に関する要約された特徴を示す

2.3 実例

- 研究目標: 改装 (Reform)が、取引価格に与える効果を推定
- 重要な識別の仮定: Reform は、$X = [Size, Tenure, Distance, RoomNumber, Youseki]$ が同じであれば、Random に決まっている
- 推定目標: $\tau(X) = E[Y \mid D = 1, X] - E[Y \mid D = 0, X]$
- 懸案: ほとんど Reform していない物件の存在が予想される
  - “ググると”築 10 年以降から、Reform を検討し始めるのが多そうなので、メイン分析では築 10 年以上の物件に限定

2.4 Data

```r
library(tidyverse)

Data <- read_csv("Public/Data.csv") |>
  filter(TradeYear == 2024) |>
  filter(Tenure >= 10)

Y <- Data$Y
D <- Data$Reform
X <- model.matrix(~ Size + Tenure + Distance + RoomNumber + Youseki, Data)
X <- X[, -1]
```

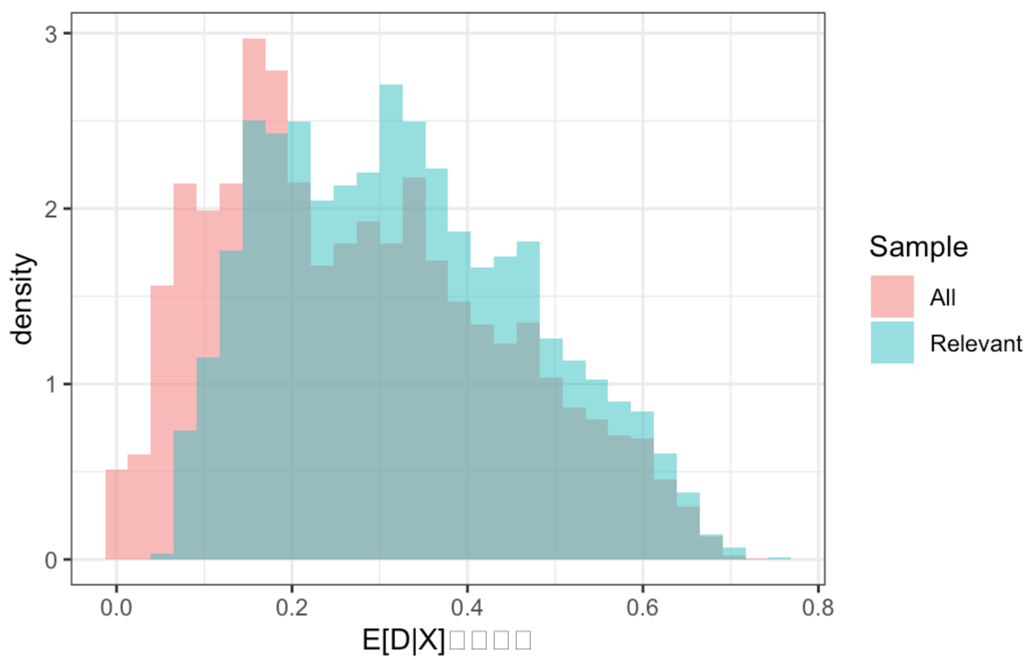
2.5 Nuisance[1]

```
library(SuperLearner)

HatY <- SuperLearner(
  Y = Y,
  X = X,
  newX = X,
  SL.library = c("SL.mean", "SL.lm", "SL.ranger"),
  cvControl = list(V = 2)
)

HatD <- SuperLearner(
  Y = D,
  X = X,
  newX = X,
  SL.library = c("SL.mean", "SL.lm", "SL.ranger"),
  family = binomial(),
  cvControl = list(V = 2)
)
```

2.6 Check



2.7 grf

[1]Section 6

```r
Tau <- grf::causal_forest(
  Y = Y,
  W = D,
  X = X,
  Y.hat = HatY$SL.predict[, 1],
  W.hat = HatD$SL.predict[, 1]
)
```

2.8 Average

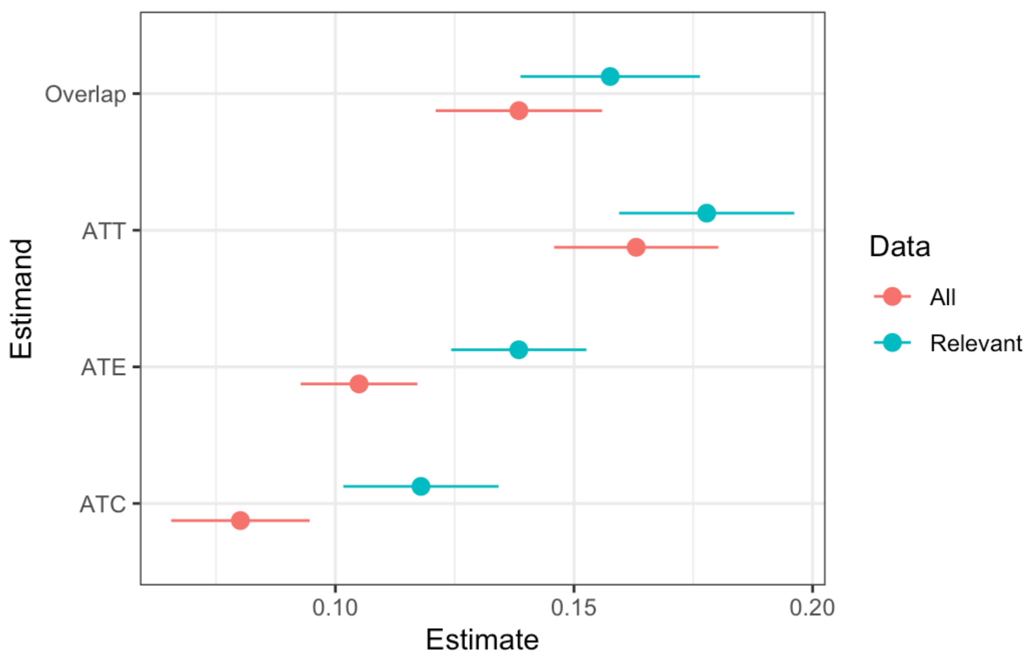
```r
ATE <- grf::average_treatment_effect(Tau)

ATT <- grf::average_treatment_effect(Tau, target.sample = "treated")

ATC <- grf::average_treatment_effect(Tau, target.sample = "control")

OR <- grf::average_treatment_effect(Tau, target.sample = "overlap")
```

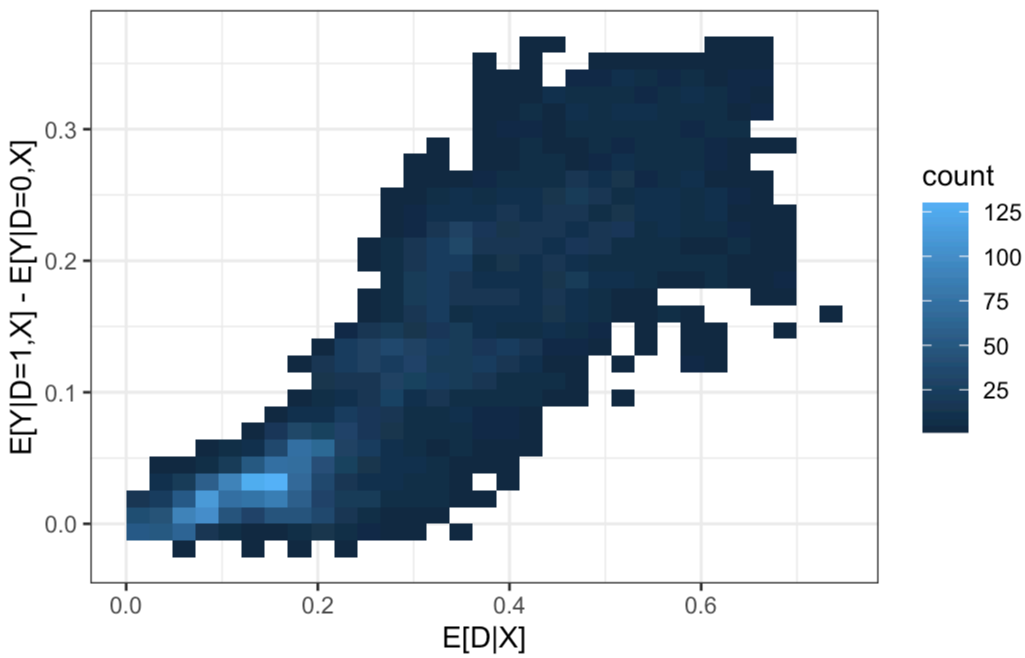
2.9 Result



2.10 直感

```r
tibble(
  Tau = Tau_All$predictions,
  HatD = HatD_All$SL.predict[, 1]
) |>
```

```
ggplot(
  aes(
    x = HatD,
    y = Tau
  )
) +
geom_bin2d() +
theme_bw() +
ylab("E[Y|D=1,X] - E[Y|D=0,X]") +
xlab("E[D|X]")
```



3 Best Linear Projection

3.1 線型モデルの把握への活用

- 線型モデルの役割は、予測だけではなく、変数間の関係性の"把握"
  - 人間が把握できるモデルは通常、βの数が少なく、現実的な事例数のもとで適切に推定できる
- 完全に Nonparametric な推定は、信頼区間の推定が難しい
  - "仮にできたとしても"、X の組み合わせ分ある推定値を人間が認知できない
  - "過剰な推定目標"である可能性

3.2 Best Linear Projection

- Semenova and Chernozhukov (2021)
- $\beta(X) \simeq \beta_0 + \beta_1 Z_1 + ..$ で近似
  - X を全てZとして使っても、一部を使っても OK
- 注: 把握が目的なので、人間が理解できないほど複雑化できない
  - 典型的には関心となる変数 Z を人間が指定し、線型モデルを推定する
    - Z を標準化すれば、β_0 は"平均差"として解釈できる

3.3 推定方法

1. 機械学習などにより、$E[Y \mid d, X], E[D \mid X]$ を推定
2. AIPW のスコア ($\theta = E[\phi]$) を算出

$$\phi = \tau(X)$$

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

3. ϕ を Z で OLS 推定し、通常の信頼区間を算出

3.4 注意点

- 多重共線性による、ミスリードする結果に注意
  - 相関が強い変数は、一つの変数にまとめる
    - 例: 部屋の広さと部屋の数は相関が強いので、部屋の広さで代表する
- Z について標準化 ((Z-Z の平均)/Z の Standard Error) を推奨
  - 特に β_0 が$\beta(X)$ の代表的な値として解釈できる
    - 標準化しないと $\beta_0 \simeq \tau(Z = 0)$ であり、解釈困難

3.5 BLP

```r
Z <- Data |> select(Size, Tenure, Distance)

grf::best_linear_projection(Tau, scale(Z))
```

```

Best linear projection of the conditional average treatment effect.
Confidence intervals are cluster- and heteroskedasticity-robust (HC3):

             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.1384357  0.0071136 19.4606  < 2e-16 ***
```

```
Size         0.0180019  0.0081167  2.2179  0.02661 *
Tenure       0.0835991  0.0075295 11.1028  < 2e-16 ***
Distance     0.0169091  0.0073172  2.3109  0.02089 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4 Group Average Difference

4.1 グループ化

- X の相関が強い場合/非線形性が強い場合、線型モデルは$\tau(X)$異質性の特徴をうまく捉えきれない
- X の情報をもとに、事例をグループ化して、平均差の違いを推定する方法が有力
  - $\tau(X)$ によって、“sort”する (Chernozhukov et al., 2018; Kallus, 2022)

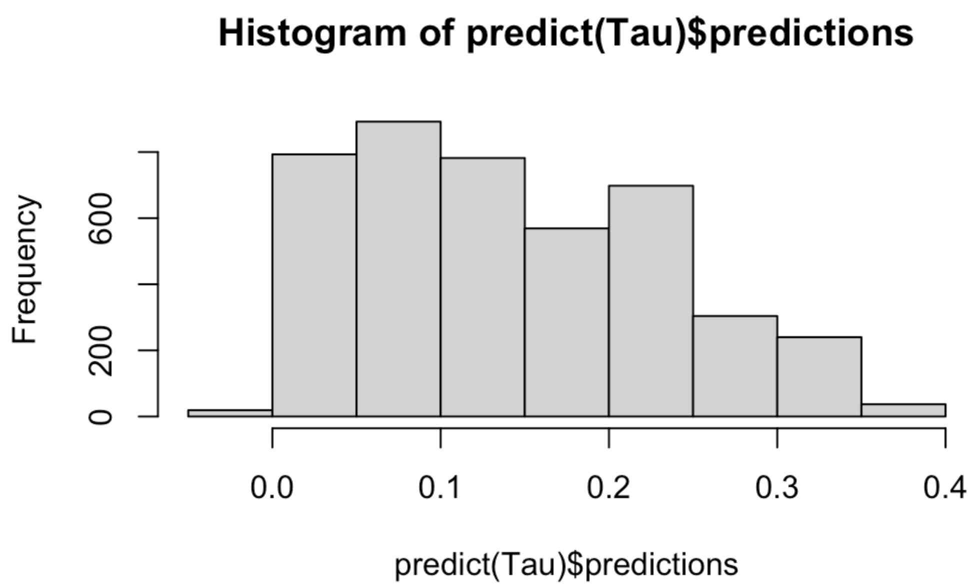
4.2 Athey and Wager (2019)

1. Causal Forest などを用いて、$\tau(X)$ を交差推定[2]
2. $\tau(X)$ でソートし、事前に指定したルールに基づいて、サブグループ分け (例: median よりも高い/低い)
3. サブグループごとに平均差を AIPW などで推定 (因果推論では、Group Average Treatment Effect と呼ばれる)
4. サブグループ間でのXの違いを記述 (Classification Analysis)

4.3 Hetero

```
hist(predict(Tau)$predictions)
```

[2]Section 7



4.4 Group Average Difference

```
grf::average_treatment_effect(
  Tau,
  subset = Tau$predictions >= median(Tau$predictions)
)
```

```
  estimate    std.err
0.22638253 0.01089877
```

```
grf::average_treatment_effect(
  Tau,
  subset = Tau$predictions < median(Tau$predictions)
)
```

```
   estimate     std.err
0.050488890 0.009131508
```

4.5 CLAN

```
Data |>
  mutate(
    Group = if_else(
      Tau$predictions >= median(Tau$predictions),
```

```
      "Hige",
      "Low"
    )
  ) |>
  select(Group, Size:Youseki) |>
  gtsummary::tbl_summary(
    by = Group
  )
```

Characteristic	**Hige** N = 2,167[1]	**Low** N = 2,167[1]
Size	55 (40, 65)	30 (20, 60)
Tenure	36 (26, 44)	17 (13, 21)
Distance	7.0 (4.0, 10.0)	6.0 (4.0, 8.0)
RoomNumber		
1	667 (31%)	1,357 (63%)
2	715 (33%)	390 (18%)
3	726 (34%)	401 (19%)
4	57 (2.6%)	19 (0.9%)
5	2 (<0.1%)	0 (0%)
Youseki	300 (200, 400)	400 (300, 500)

[1] Median (Q1, Q3); n (%)

5 Appnedix: Propensity score weight

5.1 様々な Weight

- $ATE = \int \tau(X) f(X) dX$
- $PLR = \int \tau(X) \frac{f(D=1|X) f(D=0|X)}{\int f(D=1|X) f(D=0|X) dX} f(X) dX$
- $ATT = \int \tau(X) \times f(X \mid D=1) dX$

$$= \int \tau(X) \times \frac{f(D=1 \mid X)}{f(D=1)} \times f(X) dX$$

- $ATC = \int \tau(X) \times f(X \mid D=0) dX$

$$= \int \tau(X) \times \frac{f(D=0 \mid X)}{f(D=0)} \times f(X) dX$$

5.2 ATT/ATC の特徴

- 解釈が用意: 平均差 $\tau(X)$ を $D = 1/D = 0$ におけるXの分布を用いて集計
  - D がX ないでランダムに決まっていれば、$D = 1/0$ グループ内での "平均効果"と解釈できる
- Neyman's orthogonality を満たすように書き換えられる
  - ATT: $E[D \mid X] \simeq 0$ が存在しても推定できる
    - $E[D \mid X] \simeq 1$ が存在すれば問題

6 Appendix: Stacking

6.1 Stacking

- nuisance 関数の推定については、stacking を使用を推奨
  - 使用するアルゴリズムの数が、非常に多い場合、２重の交差推定が必要
    - Ahrens et al. (2025)

7 Appendix: Causal Forest の特徴

7.1 Out of sample prediction

- Bagging 系統のアルゴリズムでは、ブートストラップ法を用いて、大量の予測モデルを生成
  - 特定の事例 i が含まれないデータのみから生成された予測モデルも存在
    - そのようなモデルのみを集計すれば、事例 i について交差推定できる
- 追加的なサンプル分割なしで、交差推定できる

7.2 AIPW

- grf パッケージにおける AIPW 推定は、以下の定式化を使用

$$E[\theta - \phi] = 0$$

where $\phi = \tau(X) +$

$$\frac{D - E[D \mid X]}{E[D \mid X](1 - E[D \mid X])}$$

$$\times (Y - E[Y \mid X] - (D - E[D \mid X]) \times \tau(X))$$

7.3 Reference

Bibliography

Ahrens, A. et al. (2025) “Model averaging and double machine learning,” Journal of Applied Econometrics [Preprint].

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

Chernozhukov, V. et al. (2018) Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India.

Kallus, N. (2022) “Treatment Effect Risk: Bounds and Inference,” 2022 ACM Conference on Fairness, Accountability, and Transparency (Minor revision in Management Science) [Preprint].

Semenova, V. and Chernozhukov, V. (2021) “Debiased machine learning of conditional average treatment effects and other causal functions,” The Econometrics Journal, 24(2), pp. 264–289.