

# 機械学習による異質性分析

## 機械学習

川田恵介  
東京大学

keisukekawata@iss.u-tokyo.ac.jp

2025-12-22

## 1 一般化的な枠組み

### 1.1 一般化した部分線型モデル

$$Y = \underbrace{\beta(X) \times D}_{Interest} + \underbrace{f(X)}_{Nuisance} + \underbrace{u}_{E[u|D,X]}$$

- 異質性の推定 =  $\beta(X)$  の推定

### 1.2 前回の内容

- $\beta(X) \simeq \beta_0 + \beta_1 X_1 + ..$  を仮定
  - ▶ Best Linear Projections (BLP)
- 本日の内容: 機械学習等を用いることで、 $\beta(X)$  の定式化に関する強い仮定を用いずに推定できるか?
  - ▶ いくつか有力なアプローチが提案されている

### 1.3 T-learner

- $\beta(X) = E[Y | D = 1, X] - E[Y | D = 0, X]$  を利用

1.  $E[Y | D = 1, X]$  と  $E[Y | D = 0, X]$  を推定

2.  $\beta(X)$  を 推定値の差として推定

### 1.4 R-learner (Nie and Wager, 2021)

- 以下を利用

$$Y - E[Y | X] = \underbrace{\beta(X) \times (D - E[D | X])}_{Interest} + \underbrace{u}_{E[u|D,X]}$$

1.  $E[Y | X]$  と  $E[D | X]$  を推定

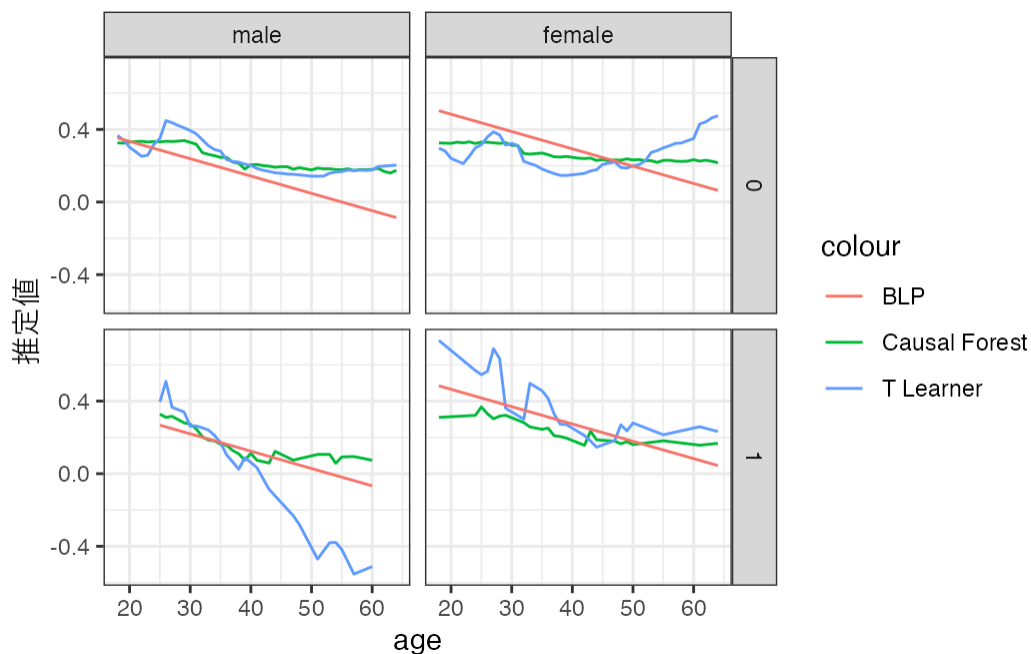
## 2. $\beta(X)$ を残差回帰として推定

### 1.5 Causal Forest

- Wager and Athey (2018) ; Athey, Tibshirani and Wager (2019) ; Athey and Wager (2019)
- R-learner の一種
  - ▶ Random Forest を用いて、 $\beta(X)$  を推定する

### 1.6 例

### 1.7 例



### 1.8 他の方法

- X-learner や DR-learner など、多くの方法が提案されている
  - ▶ CausalML の 14、15 章参照
- 課題: 一般に信頼区間の計算が難しい
  - ▶ Causal Forest は例外的に可能
- 経済分析にどう応用するか?

## 2 Group average treatment effect

### 2.1 推定対象

- $\beta(X)$  の推定結果を、異質性分析の指標として用いる
- $\beta(X)$  に基づく Group average treatment effects (GATE)
  - Chernozhukov et al. (2025) ; Athey and Wager (2019)

### 2.2 手順

1. データをランダムに分割し、一方のデータで  $\beta(X)$  を推定する
2. もう一方のデータで、 $\beta(X)$  の代表でサブグループ分けを行う
3. サブグループごとに、 $\beta(X)$  の平均値を推定する

### 2.3 例

```
CPS1985 <- CPS1985

Y <- CPS1985$wage |> log()

D <- if_else(CPS1985$occupation == "technical",
             1,
             0)

X <- CPS1985 |>
  select(-wage, -occupation) |>
  fastDummies::dummy_columns(remove_selected_columns = TRUE)

Group <- sample(1:2,
               nrow(CPS1985),
               replace = TRUE)
```

### 2.4 例

```
model_train <- grf::causal_forest(
  X = X[Group == 1,],
  W = D[Group == 1],
  Y = Y[Group == 1]) |>
  predict(X[Group == 2,])

model_estimation <- grf::causal_forest(
  X = X[Group == 2,],
  W = D[Group == 2],
  Y = Y[Group == 2])
```

## 2.5 例: GATE

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

```
estimate   std.err  
0.2489658  0.1046948
```

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

```
estimate   std.err  
0.08616792 0.10224537
```

## 2.6 CLAN

- サブグループの特徴は？
- Chernozhukov et al. (2025) では、シンプルにサブグループごとに記述統計量を計算

## 2.7 例: CLAN

```
Z <- CPS1985 |>  
  filter(Group == 2) |>  
  select(-wage, -occupation) |>  
  mutate(  
    High = model_train$predictions >  
    median(model_train$predictions)  
  )  
  
gtsummary::tbl_summary(  
  Z,  
  by = High)
```

## 2.8 例: CLAN

	FALSE	TRUE
Characteristic	N = 144 <sup>†</sup>	N = 143 <sup>†</sup>
education	12.00 (12.00, 14.00)	13.00 (12.00, 16.00)

<sup>†</sup> Median (Q1, Q3); n (%)

<b>Characteristic</b>	<b>FALSE</b> N = 144 <sup>†</sup>	<b>TRUE</b> N = 143 <sup>†</sup>
experience	26 (20, 35)	8 (4, 12)
age	44 (39, 54)	28 (24, 32)
ethnicity		
cauc	120 (83%)	126 (88%)
hispanic	3 (2.1%)	6 (4.2%)
other	21 (15%)	11 (7.7%)
region		
south	33 (23%)	48 (34%)
other	111 (77%)	95 (66%)
gender		
male	77 (53%)	70 (49%)
female	67 (47%)	73 (51%)
sector		
manufacturing	26 (18%)	27 (19%)
construction	5 (3.5%)	3 (2.1%)
other	113 (78%)	113 (79%)
union	34 (24%)	18 (13%)
married	112 (78%)	83 (58%)
college	31 (22%)	41 (29%)

<sup>†</sup> Median (Q1, Q3); n (%)

## 2.9 発展的議論

- Athey and Wager (2019) の手順がわかりやすいが、問題点も指摘されている
  - その後の議論 (Kallus, 2023; Wager, 2024; Fava, 2025)

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

Chernozhukov, V. et al. (2025) “Fisher–schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india,” *Econometrica*, 93(4), pp. 1121–1164.

Fava, B. (2025) “Training and Testing with Multiple Splits: A Central Limit Theorem for Split-Sample Estimators,” *arXiv preprint arXiv:2511.04957* [Preprint].

Kallus, N. (2023) “Treatment effect risk: Bounds and inference,” *Management Science*, 69(8), pp. 4579–4590.

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

Wager, S. (2024) “Sequential validation of treatment heterogeneity,” *arXiv preprint arXiv:2405.05534* [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.