

計量経済学特論 B 不明点

前川 大空

2025 年 10 月 21 日

1 Introduction to Causal Inference

■p.7 Simpson's paradox あるデータを全体で見ると一方の傾向が現れるのに、グループに分けて詳しく見ると逆の傾向が出てしまう現象.

p.8 Ex. ワクチン 基礎免疫という交絡因子が邪魔

■p.10 DiD について DevEcon とかはここの修正に近年悩まされている? 平行トレンドの不適切性?

p.11 Ex. Card and Krueger (1994) 労働経済学では、最低賃金上げは企業退出により失業率を上昇させるとされる. 理論と NJ のデータあってなくて、原因は平行トレンドの不成立?

■p.13 Shigeoka (2014) 医療費負担率の設計によるモラルハザードの説明を RDD で. 被説明変数がいろいろ変わっていて、Overall outpatient visits のジャンプだけが大きいことがその証左になっていることを確認せよ. 『制度の記述』から始まることも注目しよう.

■p.18 線形の特定化

$$Y_i = \alpha_i + \beta_i D_i + \varepsilon_i \quad \text{where } D_i \in \{0, 1\}$$

β_i のインデックスからも分かるように、効果には異質性が許容される. $Y_i(1)$, $Y_i(0)$ は簡約表現に過ぎない.

■Exercise 1.1 セレクションバイアスとその原因を明らかにする練習問題だ.

- $\hat{\beta}$ の確率極限を求め、 $Y_i(1) - Y_i(0)$ と比較してこれを解釈せよ.
- $(Y_i(1), Y_i(0)) \not\perp D_i$ の下で、確率極限を処置効果とセレクションバイアスに分解せよ.

Proof. $D_i \in \{0, 1\}$ は処置割当てを表す. $Y_i = \alpha + \beta D_i + u_i$ で OLS 推定を行うと、 $\hat{\beta}$ の確率極限は:

$$\text{plim } \hat{\beta} = \frac{\text{Cov}(D_i, Y_i)}{\text{Var}(D_i)}.$$

ここで Y_i に潜在結果の表現 $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ を代入すると:

$$\text{Cov}(D_i, Y_i) = \text{Cov}(D_i, Y_i(0)) + \text{Cov}(D_i, (Y_i(1) - Y_i(0)) D_i).$$

次に $\tau_i := Y_i(1) - Y_i(0)$ と置くと:

$$\begin{aligned} \text{Cov}(D_i, \tau_i D_i) &= \mathbb{E}[D_i^2 \tau_i] - \mathbb{E}[D_i] \mathbb{E}[D_i \tau_i] = \mathbb{E}[D_i \tau_i] - \mathbb{E}[D_i] \mathbb{E}[D_i \tau_i] & \because D_i^2 = D_i \\ &= \mathbf{P}(D_i = 1)(1 - \mathbf{P}(D_i = 1)) \mathbb{E}[\tau_i | D_i = 1] \end{aligned}$$

最後の変形は $\mathbb{E}[D_i \tau_i] = \mathbb{E}[\tau_i | D_i = 1] \mathbf{P}(D_i = 1)$, $\mathbb{E}[D_i] = \mathbf{P}(D_i = 1)$ を利用した. 同様に:

$$\text{Cov}(D_i, Y_i(0)) = \mathbb{E}[D_i Y_i(0)] - \mathbb{E}[D_i] \mathbb{E}[Y_i(0)] = \mathbf{P}(D_i = 1) (\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)])$$

一方:

$$\text{Var}(D_i) = \mathbf{P}(D_i = 1)(1 - \mathbf{P}(D_i = 1))$$

であるから, これを用いると:

$$\frac{\text{Cov}(D_i, Y_i(0)) + \text{Cov}(D_i, (Y_i(1) - Y_i(0))D_i)}{\text{Var}(D_i)} = \mathbb{E}[\tau_i | D_i = 1] + \frac{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)]}{1 - \mathbf{P}(D_i = 1)}.$$

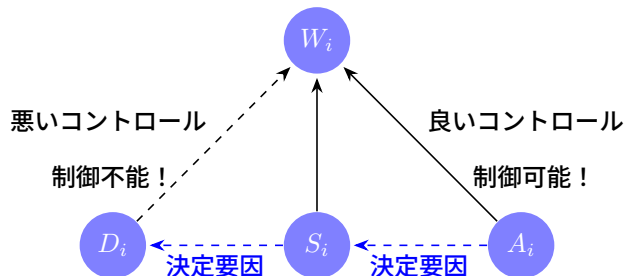
LIE による $\mathbb{E}[Y_i(0)] = \mathbf{P}(D_i = 1)\mathbb{E}[Y_i(0) | D_i = 1] + (1 - \mathbf{P}(D_i = 1))\mathbb{E}[Y_i(0) | D_i = 0]$ から:

$$\frac{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0)]}{1 - \mathbf{P}(D_i = 1)} = \mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]$$

$$\text{plim } \hat{\beta} = \underbrace{\mathbb{E}[\tau_i | D_i = 1]}_{\text{処置群の平均効果}} + \underbrace{(\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0])}_{\text{選択バイアス}} = \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]$$

従って確率極限は「処置群における平均処置効果 (ATT)」と「選択バイアス項」の和に分解できる.
 $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$ の下では第二項は消えず, $\hat{\beta}$ は β の一致推定量とはならない. \square

■p.22 bad control 語の定義としては, 一般的には **興味のある説明変数が決定要因になるようなコントロール変数** のことなのでズレてる気がする.



2 Randomized Control Trials

■p.2 RCT の意義 セレクションバイアスは $(Y_i(1), Y_i(0)) \not\perp\!\!\!\perp D_i$ の下で発生する

$$\begin{aligned} \text{plim } \hat{\beta} &= \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0] \\ &= \underbrace{\mathbb{E}[\tau_i | D_i = 1]}_{\text{処置群の平均効果}} + \underbrace{(\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0])}_{\text{選択バイアス}} \end{aligned}$$

これを消すための方法として第一に考えられることは**ランダムな割当**:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$$

$$\mathbf{P}(Y_i(t) \leq y | D_i = 0) = \mathbf{P}(Y_i(t) \leq y | D_i = 1) \quad \forall y \in \text{supp}(Y_i)$$

■Brook et al. (1983) より良い保険を提供することで、家計の健康状態は改善されるかどうか? これに答えを与えるため、米国の家族に異なる保険プランをランダムに割り当てる。

p.5 結果の解釈 不明。

■p.8 ネイマン推定量 RCT のうち、特に完全無作為化実験で、処置・統制群の平均の差を推定量とする:

$$\tau_{\text{dif}} \equiv \frac{1}{N_1} \sum_{i=1}^N Y_i \mathbb{1}\{D_i = 1\} - \frac{1}{N_0} \sum_{i=1}^N Y_i \mathbb{1}\{D_i = 0\} \quad \text{where } N_d = \sum_{i=1}^N \mathbb{1}\{D_i = d\}$$

このとき、Neyman 推定量は次の性質をもつ:

$$\mathbb{E}_D[\tau_{\text{dif}}] = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)).$$

■p.9 標本平均となる理由 標本平均である理由は、抽出に関する不確実性を考慮していないため。

- 標本 ATE $\frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$: 固定された標本 での ATE
- 母集団 ATE $\mathbb{E}[Y_i(1) - Y_i(0)]$: 母集団全体 での ATE

したがって、Neyman 推定量の不偏性は、所与の標本での 割当 D_i のランダム性のもとで 不偏。母集団で不偏とは限らない。i.i.d. 抽出の場合は次が成り立つ:

$$\mathbb{E}[\tau_{\text{dif}}] = \mathbb{E}[Y_i(1) - Y_i(0)],$$

ここで $\mathbb{E}[\cdot]$ は 抽出・割当両方 で取られ、推定量は母集団 ATE でも不偏。

■p.11 線形回帰での代用 共変量の表現が容易で、かつネイマン推定量を構成することが可能であるため、一般的には線形回帰を利用する。現状は標本数は固定されているため、大標本理論 (一致性) は考慮せず、推定量の 不偏性 のみを考える。

■p.12 回帰による RCT の推定 ランダム化実験において、単純な回帰モデル

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

を考える。ここで $D_i \in \{0, 1\}$ は処置割当てであり、潜在結果表記で

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

が成り立つ。OLS 推定量 $\hat{\tau}$ は標本共分散/分散の比で表される:

$$\hat{\tau} = \frac{\widehat{Cov}(D_i, Y_i)}{\widehat{Var}(D_i)} = \frac{\frac{1}{n} \sum_{i=1}^n D_i Y_i - \left(\frac{1}{n} \sum_{i=1}^n D_i\right) \left(\frac{1}{n} \sum_{i=1}^n Y_i\right)}{\frac{1}{n} \sum_{i=1}^n D_i^2 - \left(\frac{1}{n} \sum_{i=1}^n D_i\right)^2}.$$

不偏性を示そう。

Proof. まず次の母集団における期待値を考える:

$$\mathbb{E}[D_i Y_i] = \mathbb{E}[D_i (D_i Y_i(1) + (1 - D_i) Y_i(0))] = \mathbb{E}[D_i Y_i(1)] \quad \because D_i(1 - D_i) = 0.$$

同様に全体の期待値は

$$\mathbb{E}[Y_i] = \mathbb{E}[D_i Y_i(1)] + \mathbb{E}[(1 - D_i) Y_i(0)] = \mathbb{E}[D_i] \mathbb{E}[Y_i(1)] + \mathbb{E}[1 - D_i] \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$$

これらを用いて分子 (の母集団における期待値) を計算する:

$$\begin{aligned} \mathbb{E}[D_i Y_i] - \mathbb{E}[D_i] \mathbb{E}[Y_i] &= \mathbb{E}[D_i Y_i(1)] - \mathbb{E}[D_i] (\mathbb{E}[D_i Y_i(1)] + \mathbb{E}[(1 - D_i) Y_i(0)]) \\ &= \mathbb{E}[D_i] \mathbb{E}[Y_i(1)] - \mathbb{E}[D_i] (\mathbb{E}[D_i] \mathbb{E}[Y_i(1)] + (1 - \mathbb{E}[D_i]) \mathbb{E}[Y_i(0)]) \\ &= \mathbb{E}[D_i] (1 - \mathbb{E}[D_i]) (\mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]). \end{aligned}$$

分母は

$$\begin{aligned} \text{Var}[D_i] &= \mathbb{E}[D_i] (1 - \mathbb{E}[D_i]). \\ \mathbb{E}[\hat{\tau}] &= \frac{\mathbb{E}[D_i Y_i] - \mathbb{E}[D_i] \mathbb{E}[Y_i]}{\text{Var}[D_i]} = \mathbb{E}[Y_i(1) - Y_i(0)]. \end{aligned}$$

が成り立つ. $\hat{\tau}$ は不偏推定量である. □

p.13 別解: 条件期待値を使った直観

Proof. 条件付き期待値を用いるとより直観的に説明できる.

$$\varepsilon_i \equiv Y_i - \mathbb{E}[Y_i | D_i]$$

とおくと,

$$\begin{aligned} \varepsilon_i &= Y_i - \mathbb{E}[Y_i(1) | D_i = 1] D_i - \mathbb{E}[Y_i(0) | D_i = 0] (1 - D_i) \\ &= Y_i - \mathbb{E}[Y_i(1) - Y_i(0)] D_i - \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i. \end{aligned}$$

つまり条件付き期待値関数は線形の形を取り,

$$\mathbb{E}[Y_i | D_i] = \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(1) - Y_i(0)] D_i,$$

これは回帰モデルの $\alpha + \tau D_i$ に対応する. したがって OLS による推定は不偏性を満たす. □

■p.14 Q. ネイマン推定量と OLS は同等か? どちらも不偏推定量, 不均一分散下で同等.

p.14 Ex. 完全無作為化実験: ネイマン分散推定量は頑健 OLS 分散推定量と同等.

■p.15 ネイマン推定量との違い 共変量が入られること. 以下の (真の) 構造モデルを考えよう:

$$Y_i = \alpha + \tau D_i + X_i \beta + \varepsilon_i$$

X_i は共変量で $(Y_i(d), X_i) \perp\!\!\!\perp D_i$ とすると, 過小定式化により得られる $\hat{\tau}$ の OLSE は $\mathbb{E}[Y_i(1) - Y_i(0)]$ に対し不偏性を持たないが, 一貫性は保たれる. これは, 欠落変数バイアス (確率収束先のバイアス) が $\text{Cov}(X_i, D_i)$ の形で表現されるためである. 次の練習問題で確認しよう:

■p.16 Exercise 2.1 ネイマン推定量と同等な $\hat{\tau}$ は, 共変量を落とした過小定式化に基づく推定量として得られて, $(Y_i(d), X_i) \perp\!\!\!\perp D_i$ によって, 真の ATE との間に一貫性は保持される. 真のデータ生成過程が:

$$Y_i = \alpha + \tau D_i + X_i \beta + \varepsilon_i, \quad (\varepsilon_i, X_i) \perp\!\!\!\perp D_i$$

を満たすとする。ここで D_i は二値の確率変数。誤った定式化 (過小定式化)

$$Y_i = \tilde{\alpha} + \tilde{\tau} D_i + \tilde{\varepsilon}_i$$

で得られる OLS 推定量 $\hat{\tau}$ は τ に一致性を持つことを示せ。

Proof. 真のモデル $Y = \mathbf{1}\alpha + D\tau + X\beta + \varepsilon$ として、過小定式化に基づく OLSE $\hat{\tau}$ は:

$$\begin{aligned} (\hat{\alpha}, \hat{\tau})^\top &= (Z^\top Z)^{-1} Z^\top Y = (Z^\top Z)^{-1} Z^\top (\mathbf{1}\alpha + D\tau + X\beta + \varepsilon) \\ &= (Z^\top Z)^{-1} Z^\top (\mathbf{1}\alpha + Z^\top D + X\beta) + (Z^\top Z)^{-1} Z^\top \varepsilon \\ &= (Z^\top Z)^{-1} (\mathbf{1}\alpha + Z^\top D + X\beta) + (Z^\top Z)^{-1} Z^\top \varepsilon \\ &= \left(\frac{Z^\top Z}{n}\right)^{-1} \left(\frac{Z^\top \mathbf{1}\alpha + Z^\top D\tau + Z^\top X\beta}{n}\right) + \left(\frac{Z^\top Z}{n}\right)^{-1} \frac{Z^\top \varepsilon}{n} \end{aligned}$$

ここで、 $Z = [\mathbf{1} \ D]$, $\gamma = [\alpha, \tau]^\top$ は過小定式化で用いた説明変数行列で:

$$Z^\top Z = \begin{pmatrix} \mathbf{1}^\top \mathbf{1} & \mathbf{1}^\top D \\ D^\top \mathbf{1} & D^\top D \end{pmatrix}, Z^\top D = \begin{pmatrix} \mathbf{1}^\top D \\ D^\top D \end{pmatrix}, Z^\top \mathbf{1} = \begin{pmatrix} \mathbf{1}^\top \mathbf{1} \\ D^\top \mathbf{1} \end{pmatrix}, Z^\top X = \begin{pmatrix} \mathbf{1}^\top X \\ D^\top X \end{pmatrix}, Z^\top \varepsilon = \begin{pmatrix} \mathbf{1}^\top \varepsilon \\ D^\top \varepsilon \end{pmatrix}.$$

仮定 $(\varepsilon_i, X_i) \perp\!\!\!\perp D_i$ から $\mathbb{E}[X_i D_i] = 0$, $\mathbb{E}[D_i \varepsilon_i] = 0$ であって,

$$\mathbb{E}[Z^\top Z] = \begin{pmatrix} n & np \\ np & np^2 \end{pmatrix}, \mathbb{E}[Z^\top D] = \begin{pmatrix} np \\ np^2 \end{pmatrix}, \mathbb{E}[Z^\top \mathbf{1}] = \begin{pmatrix} n \\ np \end{pmatrix}, \mathbb{E}[Z^\top X] = \begin{pmatrix} n\mathbb{E}[X_i] \\ \mathbf{0} \end{pmatrix}, \mathbb{E}[Z^\top \varepsilon] = \mathbf{0}$$

$$\mathbb{E}[Z^\top Z]^{-1} = \frac{1}{n^2 p - n^2 p^2} \begin{pmatrix} np & -np \\ -np & n \end{pmatrix} = \frac{1}{n(1-p)} \begin{pmatrix} p & -p \\ -p & 1 \end{pmatrix} \quad \text{if } p \in (0, 1),$$

ここで、 $p = \mathbb{E}[D_i]$. LLN より:

$$\begin{aligned} \text{plim} \begin{pmatrix} \hat{\alpha} \\ \hat{\tau} \end{pmatrix} &= (\mathbb{E}[Z^\top Z])^{-1} (\mathbb{E}[Z^\top \mathbf{1}]\alpha + \mathbb{E}[Z^\top D]\tau + \mathbb{E}[Z^\top X]\beta) + (\mathbb{E}[Z^\top Z])^{-1} \mathbb{E}[Z^\top \varepsilon] \\ &= (\mathbb{E}[Z^\top Z])^{-1} (\mathbb{E}[Z^\top \mathbf{1}]\alpha + \mathbb{E}[Z^\top D]\tau) \\ &= \frac{1}{n(1-p)} \begin{pmatrix} p & -p \\ -p & 1 \end{pmatrix} \begin{pmatrix} n\alpha + np\tau \\ np\alpha + np\tau \end{pmatrix} \\ &= \frac{1}{n(1-p)} \begin{pmatrix} \cdot \\ -p(n\alpha + np\tau) + (np\alpha + np\tau) \end{pmatrix} = \begin{pmatrix} \cdot \\ \tau \end{pmatrix} \end{aligned}$$

題意は示された。FWL 定理を用いるならば最初の $\tilde{\alpha}$ に関して利用すればよかったわけだ。 \square

■補足: FWL 定理 真の構造モデルにおける D_i の係数 τ は、次の手順で得られる係数と一致する:

1. Y_i を $1, X_i$ に回帰して残差 $\tilde{Y} = M_{1,X}Y = Y - P_{1,X}Y$ を得る ($P_{1,X}$: 射影行列)
2. D_i を $1, X_i$ に回帰して残差 $\tilde{D}_i = M_{1,X}D_i$ を得る
3. \tilde{Y}_i を \tilde{D}_i に回帰したときの係数が、 τ と等しい

即ち以下が成り立つ:

$$\hat{\tau} = (D^\top M_{1,X} D)^{-1} D^\top M_{1,X} Y.$$

Proof. FWL 定理を用いて示そう (期待値が分からない):

$$\begin{aligned}
 \tilde{Y}_i &= Y_i - \bar{Y}, \quad \tilde{\epsilon}_i^* \equiv Y_i - X_i(X_i^\top X_i)^{-1} X_i^\top Y_i = M_X Y_i \\
 \hat{\tau}_i^* &\xrightarrow{p} \frac{\mathbb{E}[(M_X \tilde{D}_i)^\top (M_X Y_i)]}{\mathbb{E}[(M_X \tilde{D}_i)^\top (M_X \tilde{D}_i)]} = \frac{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{Y}_i]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} \\
 &= \frac{\mathbb{E}[\tilde{D}_i^\top M_X (\tau \tilde{D}_i + \tilde{X}_i \beta + \varepsilon_i)]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} = \tau + \frac{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{X}_i]}{\mathbb{E}[\tilde{D}_i^\top M_X \tilde{D}_i]} = \tau \\
 &\because \mathbb{E}[\tilde{D}_i M_X \tilde{X}_i] = \mathbb{E}[\tilde{D}_i (I - \tilde{X}_i (\tilde{X}_i^\top \tilde{X}_i)^{-1} \tilde{X}_i^\top) \tilde{X}_i] = 0 \\
 \hat{\tau} &\xrightarrow{p} \frac{\mathbb{E}[\tilde{D}_i \tilde{Y}_i]}{\mathbb{E}[\tilde{D}_i^2]} = \frac{\mathbb{E}[\tilde{D}_i (\tau \tilde{D}_i + \tilde{X}_i^\top \beta + \varepsilon_i)]}{\mathbb{E}[\tilde{D}_i^2]} = \tau + \beta \frac{\mathbb{E}[\tilde{D}_i \tilde{X}_i^\top]}{\mathbb{E}[\tilde{D}_i^2]} = \tau (\because X \perp\!\!\!\perp D)
 \end{aligned}$$

□

■p.17 Estimations with RCT: Regression with Covariates 真のモデルが

$$Y_i = \alpha + \tau D_i + X_i \beta + D_i (X_i - \bar{X}) \gamma + \varepsilon_i$$

で与えられるとする. $(Y_i(d), X_i) \perp\!\!\!\perp D_i$ の下で, OLSE $\hat{\tau}$ は ATE の一致推定量. ここで $\mathbb{E}[Y_i(d) \mid X_i]$ の関数形は一致性には関係なく, $\hat{\gamma}$ はこの線形射影と見なすことが出来る.

p.18 条件付期待関数 (CEF) と線形射影 処置効果は線形なら:

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \tau d_i + d_i (X_i - \bar{X}) \gamma$$

では? OLS 推定は完全に正しい定式化である. しかし一般には,

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x] = \alpha(x) + \tau(x) D_i$$

のように非線形である. このとき OLS は $\tau(x)$ の線形近似として

$$\tau(x) \approx \tau + x' \gamma$$

を推定しているとみなせる.

■p.20 共変量を追加したときの分散への影響 共変量を追加することによって推定量の分散は弱く減少する:

$$\text{Var}(Y_i - \mathbb{E}[Y_i \mid D_i = d]) \geq \text{Var}(Y_i - \mathbb{E}[Y_i \mid D_i = d, X_i = x]).$$

ただし, サンプルにおいてはトレードオフが存在する:

- 共変量を追加しても母集団分散は決して増加しない.
- 無関連な変数を追加すると標本分散はむしろ増加しうる.

証明が **Exercise 2.2** で行われる.

Proof. 仮定 $\mathbb{E}[Y_i^2] < \infty$ の下で, 次の 2 つの記法を用いる:

$$m(D_i) = \mathbb{E}[Y_i \mid D_i], \quad m(D_i, X_i) = \mathbb{E}[Y_i \mid D_i, X_i].$$

このとき示すべき不等式は

$$\text{Var}(Y_i - m(D_i)) \geq \text{Var}(Y_i - m(D_i, X_i)).$$

まず LIE より

$$m(D_i) = \mathbb{E}[m(D_i, X_i) \mid D_i]$$

さらに全分散の法則から:

$$\text{Var}(Y_i \mid D_i) = \mathbb{E}[\text{Var}(Y_i \mid D_i, X_i) \mid D_i] + \text{Var}(m(D_i, X_i) \mid D_i).$$

この式はすべての D_i に対して成り立つ. 凸関数 $\kappa(x) = x^2$ に Jensen の不等式を適用すると:

$$\mathbb{E}[m(D_i, X_i)^2 \mid D_i] \geq (\mathbb{E}[m(D_i, X_i) \mid D_i])^2 = m(D_i)^2.$$

$$\text{Var}(m(D_i, X_i) \mid D_i) = \mathbb{E}[m(D_i, X_i)^2 \mid D_i] - m(D_i)^2 \geq 0.$$

D_i について期待値を取ると:

$$\begin{aligned} \mathbb{E}[\text{Var}(Y_i \mid D_i)] &= \mathbb{E}[\mathbb{E}[\text{Var}(Y_i \mid D_i, X_i) \mid D_i]] + \mathbb{E}[\text{Var}(m(D_i, X_i) \mid D_i)] \\ &\geq \mathbb{E}[\mathbb{E}[\text{Var}(Y_i \mid D_i, X_i) \mid D_i]] = \mathbb{E}[\text{Var}(Y_i \mid D_i, X_i)]. \end{aligned}$$

一方,

$$\text{Var}(Y_i - m(D_i)) = \mathbb{E}[(Y_i - m(D_i))^2] = \mathbb{E}[\text{Var}(Y_i \mid D_i)],$$

$$\text{Var}(Y_i - m(D_i, X_i)) = \mathbb{E}[(Y_i - m(D_i, X_i))^2] = \mathbb{E}[\text{Var}(Y_i \mid D_i, X_i)].$$

よって,

$$\text{Var}(Y_i - m(D_i)) \geq \text{Var}(Y_i - m(D_i, X_i)).$$

$$\text{Var}(Y_i - \mathbb{E}[Y_i \mid D_i]) \geq \text{Var}(Y_i - \mathbb{E}[Y_i \mid D_i, X_i])$$

題意は示された.

□

■p.24 層化無作為化実験 等確率割当 等確率の割り当ての場合には, $(Y_i(d), G_i) \perp\!\!\!\perp D_i$ だが $G_i \not\perp\!\!\!\perp Y_i(d)$ で, コントロール変数と同じ扱いが出来ることを注意せよ. 一致性は保たれるため, 入れても入れなくともよい.

■p.25 等確率割当てでない場合 等確率割当て無ければ, $G_i \not\perp\!\!\!\perp D_i$ であって, **ダミーとして入れる必要がある**.

1. ダミーとして追加する場合: $\beta^W = \sum_{g \in G} \mathbf{P}(G_i = g)e(g)(1 - e(g))\beta_g$ への一致推定量
2. 交差項も追加する場合: $\sum_{g \in G} \beta_g \mathbf{P}(G_i = g)$ への一致推定量

後者の方が分散が大きくなるため, トレードオフがある.

■p.28 no spillover effect 適切な効果量の **範囲設定** が必要.

■p.29 Ex. Angrist and Lavy (2009) 成績優秀賞の効果分析. 治療群と対照群の割り当ては **学校単位**. 個人に割り当てた場合, 足の引っ張り合いが起きうることが理由.

■p.32 SUTVA 以下のように定式化される:

Assumption: SUTVA

潜在的結果が以下のように表されることを指す.

$$Y_i = Y_i^*(Z_i) = \begin{cases} Y_i^*(1) & \text{if } D_i = 1 \\ Y_i^*(0) & \text{if } D_i = 0 \end{cases} \quad \text{where } D_i \in \{0, 1\}$$

つまり, 以下の 2 条件がこの式で表現されている:

1. no spillover: 潜在的結果は他者の処置の影響を受けない (引数が D_i のみ)
2. consistency: 処置が施されたとき, 観測されるのは潜在結果をその処置で評価した値
3. binary treatment: 処置が均一 (デジタルな処置)

■p.33 Ex. Angrist and Lavy (2009) 以下のクラス内平均ならば他クラスの処置 C' に依存しない:

$$\bar{Y}_C(d) := \frac{1}{n_C} \sum_{i=1}^{n_C} Y_{i,C}(d)$$

■p.36 フィッシャーの正確確率検定 以下のように, すべての潜在結果が特定化できる帰無仮説を考える:

$$H_0^{\text{Fisher}} : Y_i(0) = Y_i(1) \quad \forall i \in \{1, \dots, N\}$$

この下では $Y_i = Y_i(d)$ は固定されるため, 以下は D_i に依存しない:

$$T(D, Y) = \left| \frac{\sum_i Y_i \mathbf{1}\{D_i = 1\}}{\sum_i \mathbf{1}\{D_i = 1\}} - \frac{\sum_i Y_i \mathbf{1}\{D_i = 0\}}{\sum_i \mathbf{1}\{D_i = 0\}} \right|$$

この D_i に関する全組合せが帰無分布を構成する. これは厳密分布であり, 漸近理論に依存しない. これを用いて行われるのが **ランダム化検定** である. 一方で Neyman の帰無仮説は平均処置効果についての仮説で:

$$H_0^{\text{Neyman}} : \mathbb{E}[Y_i(1) - Y_i(0)] = 0.$$

Fisher の検定はより強い仮定に基づく検定である.

■p.43 Ex. Young (2019) AEA 掲載論文の RCT を Fisher 検定で再評価し, およそ 13 ~ 22% の結果が有意でなくなることを示した. これは Fisher 検定が有限標本でより厳格であること, 特に **unbalanced な割当** の影響が大きい場合に有意差が消えることを意味する (p.47).

■p.51 検定法の選択 $N < 250$ のとき, 下方バイアスに対応するための頑健標準推定量としては **HC3** が適切だが, Stata のデフォルトは HC1. 十分大きな標本数では逆に **HC1** を利用することが適切. ランダム化検定は, 保守的になりすぎるところがあるため, あまりにも標本数が小さいときのみ利用すべき.

■教科書 p.79 多重検定問題 様々な種類の潜在結果を考える際に, ただ仮説検定を繰り返すだけでは, 本来制御の対象である検定のサイズを制御することが出来ないという問題.

■教科書 p.83 ボンフェローニ検定 多重検定問題の対策として, 和集合不等式に基づいて, それぞれの過誤率を $5/M\%$ とすることで族過誤率 (FWER) を 5% に抑える検定.

■教科書 p.85 ボンフェローニ・ホルム検定 明らかに帰無仮説から生成されていない検定統計量を逐次的に消去し棄却水準を下げて, 検出力をボンフェローニ検定よりも高める一方で, 族過誤率 (FWER) は依然 5% に抑える検定. 少なくとも一つの帰無仮説を棄却する確率自体はボンフェローニ検定と変わらない ($5/M\%$).

■教科書 p.87 適切な制御対象と分析結果の利用目的 ある処置が企業における複数の業績指標に及ぼす効果を評価するような場合、どの項目に差があるかによって異なる対応をとるような場合は族過誤率を気にする必要がある。だがある処置が企業の業績に何らかの影響があるかを見るだけならば気にする必要がなく、**偽検出比率 (FDP)** の期待値である **偽検出率 (FDR)** を考えれば十分である。

$$\text{FDP} := \frac{\#\{i \in \{1, \dots, n\} : H_i \text{ が真かつ } H_i \text{ が棄却される}\}}{\max\{\#\{i \in \{1, \dots, n\} : H_i \text{ を棄却する}\}, 1\}} \quad (3.5)$$

$$\text{FDR} = \mathbb{E}[\text{FDP}] \leq \alpha$$

■教科書 p.89 ベンジャミン・ホッシュバーク検定 10% が水準として慣例となっている。手順としては、今までとは逆に、**棄却しにくい** ものから順に p 値の検証を行い、棄却された場合にはそれ以上に棄却しやすいものを全て棄却するような検定となっている。例えば j 番目に小さい $p_{(j)}$ 、すなわち j 番目に棄却されやすい帰無仮説については、棄却の基準を少し厳しくして、 $p_{(j)} \leq j\alpha/M$ ならば $H_{(1)} \sim H_{(j)}$ まで全て棄却する。検定統計量が独立である際には、論理的には怪しいところがあるが、相関構造に対してはある程度頑健なことを考慮して、よく用いられる検定といえる。

■p.52 多重検定問題 ステップアップ法 (Romano and Shaikh, 2006), (See List et al., 2023) あたりを使うのが検出力を保ったうえで、族過誤率の制御をするには良いと思われる。FDR に関しては Benjamini-Hochberg 検定を用いればいいだろう。

3 Instrumental Variables

■p.3 片側遵守者 処置受取 D_i の、操作変数たる処置割当 Z_i との関係性から判断して、今回の設定は片側遵守者のケースである。無作為化実験は $Y_i(d) \perp\!\!\!\perp Z_i$ を保証する。まず **処置割当効果 (ITT 効果)** について考察する: ITT 効果は、割当変数による平均の差

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]$$

として定義される。独立性 $(Y_i(1), Y_i(0)) \perp\!\!\!\perp Z_i$ の下で、

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbb{E}[D_i | Z_i = 1].$$

右辺第 2 項 $\mathbb{E}[D_i | Z_i = 1] = \mathbf{P}(D_i = 1 | Z_i = 1)$ は、割当群のうち実際に処置を受けた割合を意味する。したがって、ITT 効果は処置効果に、実際の処置受取率を掛け合わせた形で表される。

Proof. 以下のように、片側遵守者の仮定と、独立性によって確認ができる:

$$\begin{aligned} & \mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] \\ &= \mathbb{E}[Y_i(1)D_i + Y_i(0)(1 - D_i) | Z_i = 1] - \mathbb{E}[Y_i(1)D_i + Y_i(0)(1 - D_i) | Z_i = 0] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0))D_i | Z_i = 1] + \mathbb{E}[Y_i(0) | Z_i = 1] - \mathbb{E}[Y_i(0) | Z_i = 0] \quad \because D_i |_{Z_i=0} \equiv 0 \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbf{P}[D_i = 1 | Z_i = 1] \\ &+ \mathbb{E}[0 | D_i = 0, Z_i = 1] \mathbf{P}[D_i = 0 | Z_i = 1] - \mathbb{E}[Y_i(0)] + \mathbb{E}[Y_i(0)] \quad \because (Y_i(1), Y_i(0)) \perp\!\!\!\perp Z_i \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1] \mathbb{E}[D_i | Z_i = 1] \end{aligned}$$

これで題意は満たされた。

□

■p.7 Bloom (1984) ITT 効果を用いると, ATT を次のように推定できる:

$$ATT = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbf{P}(D_i = 1 | Z_i = 1)} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1, Z_i = 1].$$

これは **Bloom (1984)** により提案された推定式であり, 一般形として

$$\frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]}$$

なる **Wald 推定量** として知られる.

■p.8 Wald 推定量と IV 推定量の関係 Wald 推定量は, 二値の Z_i を操作変数とした IV 推定量と一致する:

$$\frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)}$$

背後にある構造方程式系は

$$\begin{cases} Y_i = \alpha + \beta D_i + \varepsilon_i, \\ D_i = \pi_0 + \pi_1 Z_i + \xi_i, \end{cases}$$

であり, Z_i が D_i に関連し ($\pi_1 > 0$), かつ Z_i が Y_i に直接影響しない ($Y_i(D_i, Z_i) = Y_i(D_i)$, $(\varepsilon_i, \xi_i) \perp\!\!\!\perp Z_i$, 排除制約) ならば, 一致性を持つ因果効果推定量となる.

■p.9 Exercise 3.1 Wald 推定量を導出しよう. $0 < Cov(D, Z)$ のとき,

$$\frac{Cov(Y, Z)}{Cov(D, Z)} = \frac{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}$$

が成り立つことを示せ.

Proof. まず, 共分散の定義と $Z \in \{0, 1\}$ を利用して:

$$Cov(Y, Z) = \mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z].$$

$$\mathbb{E}[YZ] = \mathbb{E}[Y | Z = 1]\mathbf{P}(Z = 1) = \mathbb{E}[Y | Z = 1]\mathbb{E}[Z].$$

$$\mathbb{E}[Y] = \mathbb{E}[Y | Z = 1]\mathbb{E}[Z] + \mathbb{E}[Y | Z = 0]\mathbb{E}[1 - Z].$$

したがって,

$$\begin{aligned} Cov(Y, Z) &= \mathbb{E}[Z](\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y]) \\ &= \mathbb{E}[Z](\mathbb{E}[Y | Z = 1] - (\mathbb{E}[Y | Z = 1]\mathbb{E}[Z] + \mathbb{E}[Y | Z = 0]\mathbb{E}[1 - Z])) \\ &= \mathbb{E}[Z]\mathbb{E}[1 - Z]\{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]\} \\ &= Var(Z)\{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]\}. \end{aligned}$$

同様に,

$$Cov(D, Z) = Var(Z)\{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]\}.$$

これらの代入によって以下を得る:

$$\frac{Cov(Y, Z)}{Cov(D, Z)} = \frac{\mathbb{E}[Y | Z = 1] - \mathbb{E}[Y | Z = 0]}{\mathbb{E}[D | Z = 1] - \mathbb{E}[D | Z = 0]}.$$

よって二値 Z の場合における Wald 推定量を得た。 □

■p.16 Exercise 3.2 両側遵守の場合、は追加の仮定を置かない限り ITT の解釈が出来ない。独立性:

$$(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$$

の下で、以下が成り立つことを示せ:

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] = \mathbb{E}[Y_i(1) - Y_i(0) | \text{cp}] \mathbf{P}(\text{cp}) - \mathbb{E}[Y_i(1) - Y_i(0) | \text{df}] \mathbf{P}(\text{df}).$$

Proof. 独立性 $(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$ を用いる。タイプ集合を:

at (always-taker), nt (never-taker), cp (complier), df (defier)

とする。まず条件付き期待の差をタイプごとに分解する:

$$\begin{aligned} & \mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] \\ &= \mathbb{E}[D_i(1)Y_i(1) + (1 - D_i(1))Y_i(0)] - \mathbb{E}[D_i(0)Y_i(1) + (1 - D_i(0))Y_i(0)] \\ &= \mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \end{aligned}$$

ここで $D_i(1) - D_i(0)$ は各タイプで

$$D_i(1) - D_i(0) = \begin{cases} 1 & (\text{complier}), \\ 0 & (\text{always-taker または never-taker}), \\ -1 & (\text{defier}) \end{cases}$$

を取る。従って期待値はタイプの差として書ける:

$$\begin{aligned} \mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] &= \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{\text{cp}\}] - \mathbb{E}[(Y_i(1) - Y_i(0))\mathbb{1}\{\text{df}\}] \\ &= \mathbb{E}[Y_i(1) - Y_i(0) | \text{cp}] \mathbf{P}(\text{cp}) - \mathbb{E}[Y_i(1) - Y_i(0) | \text{df}] \mathbf{P}(\text{df}). \end{aligned}$$

これがスライド記載の式に一致する。 □

■p.21 LATE と単調性 単調性により defier の不存在を仮定することによって, ITT は **LATE と complier の割合の積** に分解できる。これを变形することにより, Wald 推定量と LATE の一致が確認できる。

■p.22 LATE と共変量 Wald 推定量は共変量を考慮していない。存在する場合, 仮定を再構成すべき:

1. 条件付き独立性: $(Y_i(0), Y_i(1), D_i(0), D_i(1)) \perp\!\!\!\perp Z_i | X_i$
2. 排除制約: $Y_i(d, z) = Y_i(d)$
3. 関連性: $\mathbb{E}[D_i | Z_i = 1, X_i] \neq \mathbb{E}[D_i | Z_i = 0, X_i]$
4. 単調性: $D_i(1) \geq D_i(0)$ a.s.

X_i が離散変数であり, Z_i, D_i が二値変数のとき, 2SLS は次の **完全飽和モデル** で構成される:

$$\begin{aligned} \text{第 1 段階: } D_i &= \sum_{k=1}^K \mathbb{1}\{X_i = k\}(\pi_{0k} + \pi_{1k}Z_i) + \xi_i, \\ \text{第 2 段階: } Y_i &= \sum_{k=1}^K \mathbb{1}\{X_i = k\}(\alpha_{0k}) + \tau D_i + \varepsilon_i, \end{aligned}$$

この設定下では、各 $X_i = k$ ごとに LATE を考えられる:

$$\beta(X_i = k) = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = k, D_i(1) > D_i(0)].$$

このとき、2SLS の推定量 $\hat{\tau}_{2SLS}$ は、LATE の加重平均に一致する:

■p.25 Exercise 3.3 2SLS 推定量 $\hat{\tau}_{2SLS}$ が以下のような加重平均で表せることを示す:

$$\hat{\tau}_{2SLS} \xrightarrow{p} \mathbb{E}[\omega(X)\tau(X)] \quad \text{where } \omega(X_i) = \frac{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)}{\mathbb{E}[\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)]}.$$

この重み $\omega(X_i)$ は、 $X_i = x$ での Z_i による D_i の変動の大きさに比例する。したがって、**操作変数がより強く効いている集団 (compliance のばらつきが大きい集団)** ほど、推定量において重く評価される。

Proof. 完全飽和モデルのもとでは、各 $X_i = k$ ごとに独立した第 1 段階 $D_i = \pi_{0k} + \pi_{1k}Z_i + \xi_i$ と、第 2 段階 $Y_i = \alpha_{0k} + \tau D_i + \varepsilon_i$ を構築できる。このとき、各 k ごとに Wald 推定量:

$$\tau(k) = \frac{\mathbb{E}[Y_i \mid Z_i = 1, X_i = k] - \mathbb{E}[Y_i \mid Z_i = 0, X_i = k]}{\mathbb{E}[D_i \mid Z_i = 1, X_i = k] - \mathbb{E}[D_i \mid Z_i = 0, X_i = k]}$$

を計算できる。完全飽和モデルでは、FWL 定理で、 X_i ごとに偏差を取った 2SLS 推定量が

$$\hat{\tau}_{2SLS} = \frac{\text{Cov}(P_{Z,X}D_i, P_{Z,X}Y_i)}{\text{Var}(P_{Z,X}D_i)} \xrightarrow{p} \frac{\text{Cov}(\mathbb{E}[Y_i \mid X_i, Z_i], \mathbb{E}[D_i \mid X_i, Z_i])}{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i])}$$

と表される。ここで、確率収束先を整理すると:

$$(\text{分子}) = \text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z]) = \mathbb{E}[\text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z] \mid X)]$$

条件付き共分散の中身を代入すると:

$$\begin{aligned} \mathbb{E}[\text{Cov}(\mathbb{E}[Y \mid X, Z], \mathbb{E}[D \mid X, Z] \mid X)] &= \mathbb{E}[\text{Cov}(a(X) + \tau(X)m(X, Z), m(X, Z) \mid X)] \\ &= \mathbb{E}[\tau(X) \text{Var}(m(X, Z) \mid X)], \end{aligned}$$

となる ($a(X)$ は X のみに依存する関数)。一方、分母は全分散の法則により:

$$\text{Var}(\mathbb{E}[D \mid X, Z]) = \mathbb{E}[\text{Var}(m(X, Z) \mid X)] + \text{Var}(\mathbb{E}[m(X, Z) \mid X]),$$

であるが、完全飽和モデルでの 2SLS は X による「間の変動」を除いた (すなわち M_X による偏差をとった) 成分に基づくため、分母に使うのは内部変動 $\mathbb{E}[\text{Var}(m \mid X)]$ に対応する部分である (標準的な導出では M_X を掛けた後の経験分散の期待が $\mathbb{E}[\text{Var}(m \mid X)]$ に収束する)。したがって、射影後の 2SLS の母集団極限は次の形に帰着する:

$$\hat{\tau}_{2SLS} \xrightarrow{p} \mathbb{E}\left[\frac{\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)}{\mathbb{E}[\text{Var}(\mathbb{E}[D_i \mid X_i, Z_i] \mid X_i)]} \cdot \tau(X_i)\right].$$

したがって、 $\hat{\tau}_{2SLS}$ は各 X_i ごとの LATE $\tau(X_i)$ の加重平均に一致する。 \square

Proof. 第 1 段階を D を (X, Z) に回帰した射影とし、

$$\hat{D} = P_{Z,X}D, \quad \hat{D} \xrightarrow{p} \mathbb{E}[D \mid X, Z].$$

第2段階は Y の (\hat{D}, X) への回帰。FWL 定理より、 X による偏差を取った回帰式は、

$$Y - \widehat{\mathbb{E}}[Y | X] \text{ を } \hat{D} - \widehat{\mathbb{E}}[\hat{D} | X] \text{ に回帰する OLS の係数}$$

で表される。したがって、母集団形は次のようになる:

$$\hat{\tau}^{\text{2SLS}} \xrightarrow{p} \tau^* = \frac{\text{Cov}(Y - \mathbb{E}[Y | X], \mathbb{E}[D | X, Z] - \mathbb{E}[\mathbb{E}[D | X, Z] | X])}{\text{Var}(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X])}. \quad (*)$$

分子を条件付き期待を用いて展開すると、

$$\text{Num} = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y | X])(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X]) | X]].$$

ここで、 X で条件付けた内側の期待値を、 Z の二値性を利用して書き換える:

$$\text{Num} = \mathbb{E}[\mathbb{E}[(Y - \mathbb{E}[Y | X])(Z - \mathbb{E}[Z | X])\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\} | X]].$$

Z が X から独立にランダム化されているため、 $\mathbb{E}[(Z - \mathbb{E}[Z | X])^2 | X] = \text{Var}(Z | X)$ であり、また、

$$\mathbb{E}[Y | X, Z = 1] - \mathbb{E}[Y | X, Z = 0] = \tau(X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}.$$

したがって、

$$\text{Num} = \mathbb{E}[\tau(X) \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2].$$

同様に、分母は

$$\begin{aligned} \text{Den} &= \text{Var}(\mathbb{E}[D | X, Z] - \mathbb{E}[D | X]) = \mathbb{E}[\text{Var}(\mathbb{E}[D | X, Z] | X)] \\ &= \mathbb{E}[\text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]. \end{aligned}$$

分子・分母を対応させると、

$$\tau^* = \frac{\mathbb{E}[\tau(X) \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]}{\mathbb{E}[\text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2]}.$$

したがって、

$$\tau^* = \mathbb{E}[\omega(X)\tau(X)], \quad \omega(X) \propto \text{Var}(Z | X)\{\mathbb{E}[D | X, Z = 1] - \mathbb{E}[D | X, Z = 0]\}^2.$$

題意は示された。 □

■p.27 Abadie の κ 上記の推定量は、 X_i が離散変数である場合にのみ「完全飽和モデル」として構成できる。現実には X_i が連続あるいは高次元であるため、この方法を直接適用することは困難である。そこで **Abadie (2003)** により提案されたのが、共変量を含む状況でも **complier だけの平均**を識別できる加重法 (κ 重み) である。Abadie は、任意の可積分関数 $g(Y_i, D_i, X_i)$ に対して次を示した:

$$\mathbb{E}[g(Y_i, D_i, X_i) | D_i(1) > D_i(0)] = \frac{\mathbb{E}[\kappa_i g(Y_i, D_i, X_i)]}{\mathbb{E}[\kappa_i]},$$

ただし、重み κ_i は

$$\kappa_i = 1 - \frac{D_i(1 - Z_i)}{1 - P(Z_i = 1 | X_i)} - \frac{(1 - D_i)Z_i}{P(Z_i = 1 | X_i)}.$$

$D_i = Z_i$ の個体 (すなわち complier) のみが非ゼロの重み κ_i を持つ. κ_i を用いることで, $\mathbb{E}[Y \mid X, D, D(1) > D(0)]$ に対する関数 $h(D_i, X_i; \theta)$ の近似のため, 次の最小二乗問題を考えればよい:

$$\theta_0 = \arg \min_{\theta} \mathbb{E}[\kappa_i \{Y_i - h(D_i, X_i; \theta)\}^2],$$

この方法により, X_i が連続変数であっても, complier 平均効果を整合的に推定できる.

■p.30 線形確率モデル (LPM) の利用 IV では, 第 1 段階で $\mathbb{E}[D_i \mid Z_i, X_i]$ を推定する必要があるが, D_i が二値変数の場合, その条件付き期待値は確率として $[0, 1]$ に制約される. しかし, 多くの実証研究では, 第 1 段階を 線形確率モデル (Linear Probability Model, LPM) で近似しており, 確率としての性質を満たさない予測を返す. 本来線形射影によって推定され機械的にその残差は第 2 段階の誤差と無相関となるが, 仮に第 1 段階を ロジットやプロビット などの非線形モデルで推定した場合はこれが成り立たない. 誤った定式化の懸念もある以上, 線形モデルでの第一段階推定が無難だと結論できる.

4 Conditional Independence Approaches

■p.2 LaLonde (1986) 実験データで得られた真の因果効果を, 観察データと回帰による補正でどこまで再現できるかについての研究.

■p.6 バイアス 従来のバイアス概念は無条件の差

$$\mathbb{E}[Y(0) \mid D = 1] - \mathbb{E}[Y(0) \mid D = 0]$$

で表されるが, これには少なくとも三種類の要因が含まれると説明される:

1. $X \mid D = d$ のサポートの違い
2. $X \mid D = d$ の分布の違い
3. 共通サポートの選択バイアス $\mathbb{E}[Y(0) \mid D = 1, X = x] - \mathbb{E}[Y(0) \mid D = 0, X = x]$

■p.7 Heckman et al. Heckman らは LaLonde の問題点を整理し, バイアスの源泉を示すための表を提示している. 概念的には, 単純な平均差では原因を特定できないことを強調している.

■p.8 共変量分布のバランス 共変量間の差は主に二つに分かれる:

- **unbalanced:** 同一サポート上で分布が異なる
- **not overlapped:** サポート自体が異なり共通領域が小さい (または無い)

共変量バランスやオーバーラップを評価するため, 共変量ごとの平均や標準偏差の比較が確認できる.

■p.11 オーバーラップの評価基準 オーバーラップ評価の一つの方法として, ロジット (またはロジットの線形予測) に基づく近傍判定が提示されている. 十分に小さい ℓ^U について:

$$\xi_i = \begin{cases} 1 & \text{if } \sum_{i': D_{i'} \neq D_i} \mathbb{1}\{\hat{\ell}(X_{i'}) - \hat{\ell}(X_i) \leq \ell_u\} \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \ell(x) = \log \frac{p(x)}{1 - p(x)}$$

これは「反対群に十分近い傾向スコアを持つ観測が存在するか」を示す. オーバーラップの要約指標は:

$$q_c = \frac{1}{N_c} \sum_{i: D_i=0} \xi_i, \quad q_t = \frac{1}{N_t} \sum_{i: D_i=1} \xi_i$$

処置群・対照群がどれだけ重なっているかを数値化する．LaLonde では $q_c = 0.21$, $q_t = 0.97$ が報告された．ATE の推定は困難であり ATTE しか見られず, overlapping のない処置グループに関しては, 傾向スコアの低い群を除外することで観測を除外することが検討される．

■p.13 マッチング法 共変量がバランスしていない場合やオーバーラップが不十分な場合の解決策としてマッチング法が紹介される．理想は「同一の X を持つ対」を作ることであり, これにより疑似ランダム化の近似を行う．単純なマッチング差分の推定量は, ある観測 i に対して同じ X を持つ m_i^c を見つけ,

$$\hat{\tau}_i^{\text{match}} = Y_i - Y_{m_i^c},$$

処置群に対する ATTE 推定量は

$$\hat{\tau}^{\text{match}} = \frac{1}{N_t} \sum_{i:D_i=1} (Y_i(1) - Y_{m_i^c}(0))$$

と表される．マッチング推定量の推論は非標準的で, ブートストラップが無効になりうる．マッチング推定の分散推定やブートストラップ改良の研究が続いている (Ex. Abadie & Imbens (2006): 分散推定量, Otsu & Rai (2017): 実現可能なブートストラップ法・同対照群を流用すると失敗しうる)．

■p.17 非完全一致の場合 完全一致が困難な場合, X が最も近い観測を採る最近傍マッチングを用いる:

$$m_i^c = \arg \min_{i': D_{i'} \neq D_i} \|X_i - X_{i'}\|$$

距離はユークリッドや傾向スコア差 $|p(X_i) - p(X_j)|$ を使う．傾向スコアは **balancing** の性質を持つ:

$$X \perp\!\!\!\perp D \mid b(X)$$

を満たす $b(X)$ を balancing score と呼ぶ．つまり $b(X)$ を固定すれば, 処置・対照群の X 分布は同一．

■p.20 Rosenbaum & Rubin (1983)

1. 傾向スコアはバランススコア:

$$X \perp\!\!\!\perp D \mid p(X) \quad p(X) = \Pr(D = 1 \mid X)$$

2. 傾向スコアは最小次元のバランススコア:

$b(X)$ がバランススコアなのは $b(X)$ が $p(X)$ より細かい ($p(x_1) \neq p(x_2) \Rightarrow b(x_1) \neq b(x_2)$) とき

■p.22 傾向スコアの因果的含意 もし潜在結果が条件付き独立

$$(Y(1), Y(0)) \perp\!\!\!\perp D \mid X$$

を満たすなら, 任意のバランススコア $b(X)$ について

$$(Y(1), Y(0)) \perp\!\!\!\perp D \mid b(X)$$

も成立する． $p(X)$ は (1) 共変量の balancing に有効で (2) 条件付き独立の利用において最も効率的．

■p.23 傾向スコアの利用: Heckman et al. (1998) 条件付き独立の下で

$$E[Y(0) \mid D = 1, X \in S] = E[E[Y \mid D = 0, p(X) = p(X_i)] \mid D = 1, X \in S]$$

が成り立つ．ここで S は共通サポート $S = \text{Supp}(X \mid D = 1) \cap \text{Supp}(X \mid D = 0)$ である．

■p.24 傾向スコアによる推定式 推定量はカーネル法を用いて

$$\hat{\tau} = \frac{1}{N_t} \sum_{i:D_i=1} [Y_i - \hat{E}(Y \mid D=0, p(X) = p(X_i))] I\{X_i \in \hat{S}\}$$

で与えられる。傾向スコア推定は一次段階として扱われ、標準的な二段階推定の漸近理論を適用できる。

■p.25 傾向スコアマッチング (Abadie & Imbens, 2016) 非完全一致 (M 個の最近傍) マッチングを定式化:

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (2D_i - 1) \left(Y_i - \frac{1}{M} \sum_{j \in J_M(i)} Y_j \right),$$

ここで $J_M(i)$ は $D_j = 1 - D_i$ かつ $|p(X_i) - p(X_j)|$ が最も近い M 個の集合。Wang & Zubizarreta (2023) は、バランス条件を保ちつつ最大の M を選ぶマッチング法を提案し、その推定量が \sqrt{n} -一致することを示した。

■p.26 Balancing through Propensity Score (演習 4.1) 仮定:

- 潜在結果の条件付き独立 $(Y(1), Y(0)) \perp D \mid X$
- すべての x で $0 < p(x) < 1$

このとき以下が成立する:

$$\mathbb{E} \left[\left(\frac{D}{p(X)} - \frac{1-D}{1-p(X)} \right) Y \right] = E[Y(1) - Y(0)].$$

つまり、傾向スコアを用いた逆確率重み付けにより母平均の因果効果を識別できる。

■p.27 Inverse Probability Weighting (IPW) IPW 推定量には二形式がある:

1. Horvitz–Thompson (1952) 型:

$$\hat{\tau}_{HT} = \frac{1}{n} \sum_i \left(\frac{D_i Y_i}{p(X_i)} - \frac{(1-D_i) Y_i}{1-p(X_i)} \right)$$

2. Hájek (1971) 型:

$$\hat{\tau}_H = \frac{1}{n} \sum_i h(D_i, X_i) Y_i,$$

ここで $h(D_i, X_i)$ は処置・統制群の重みをそれぞれの標本全体で標準化したもの。

■p.28 傾向スコア推定の誤った定式化とバランス推定 傾向スコア推定の誤った定式化は深刻な影響を与える。Hirano et al. (2003) は非パラメトリックな第一段階推定を提案。Imai & Ratkovic (2014) は、尤度推定の代わりにバランス条件

$$E \left[\left(\frac{D}{p(X; \beta_0)} - \frac{1-D}{1-p(X; \beta_0)} \right) f(X) \right] = 0$$

を満たすように β_0 を推定する「Covariate Balancing Propensity Score (CBPS)」法を提案した。

■p.29 二重頑健推定 Robins et al. (1994) の Augmented IPW (AIPW), 二重頑健推定 は, $p(X; \gamma)$ または $E[Y \mid X] = X'\beta$ の, いずれか一方の定式化が正しければ一致推定量が得られる。したがって AIPW は, モデルの不確実性に対して頑健な推定法として広く用いられる。