

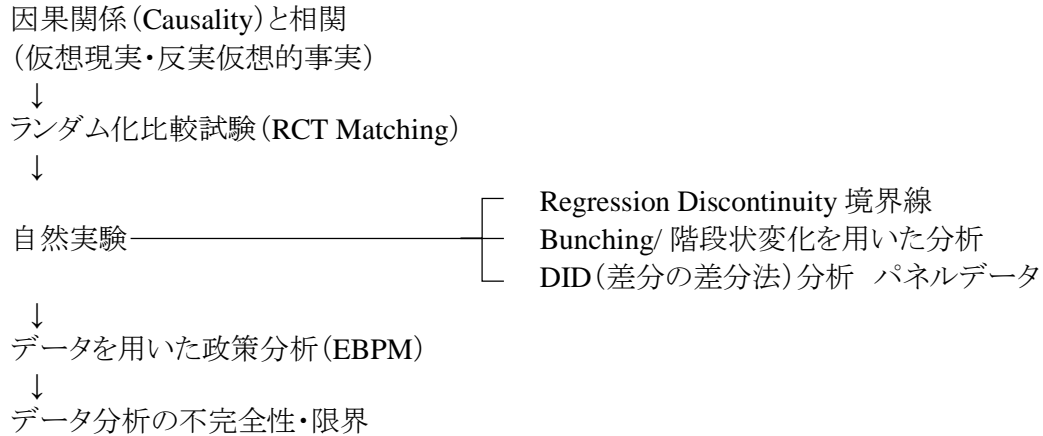
比較統計システム論 2018年度

政策評価の計量経済学(1)

第1講 因果推論概論	2
第2講 群間比較	6
第3講 統計的因果推論の枠組み.....	13
第4講 マッチング.....	17
第5講 傾向スコア法	21

第 1 講 因果推論概論

本講は岩崎 (2015, 第 1 章) を用いて、因果推論の基礎について論じる。



1.1 実験計画法 統計的因果推論 識別戦略 (Identification Strategy)

データ = 数値 + 背景情報

Ronald A. Fisher

①ランダム化(無作為化) ②反復(replication) ③局所管理(local control)

① ランダム化

「他の条件を一定にして」

② 反復

Repetition だけではなく、条件管理によって反復

③ 局所管理

(a) 単純無作為割付け (simple random allocation)

(b) 乱塊法計画 (randomized block design)

(c) 一対比較 (pair matching)

1.2 変数の種類と相互の関係

$$Z: \text{処理ダミー} \begin{cases} 1 & \text{処置} \\ 0 & \text{対照} \end{cases}$$

$$Y: \text{結果} \begin{cases} 1 & \text{有効} \\ 0 & \text{無効} \end{cases}$$

$$X: \text{観測される共変量 (Covariate)}$$

$$U: \text{観測されない共変量}$$

- (a) 共変量 X は、処置割付け Z にも、結果 Y にも影響を与えない

$$X$$

$$Z \rightarrow Y$$

- (b) 共変量 X は、処置割付け Z に影響を与えないが、結果 Y に影響を与える

$$X$$

$$\searrow$$

$$Z \rightarrow Y$$

- (c) 共変量 X は、処置割付け Z に影響を与えるが、結果 Y に影響を与えない

$$X$$

$$\swarrow$$

$$Z \rightarrow Y$$

- (d) 共変量 X は、処置割付け Z にも、結果 Y にも影響を与える

$$X$$

$$\swarrow$$

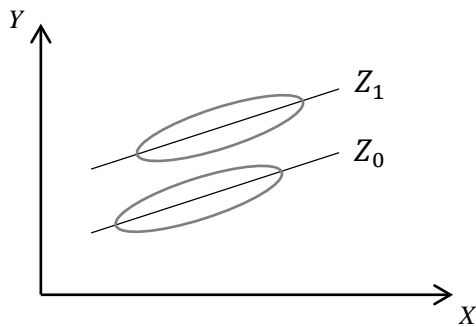
$$\searrow$$

$$Z \rightarrow Y$$

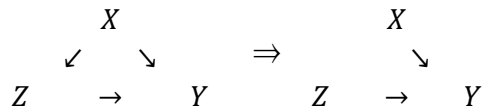
実験計画で処置の割付けがランダムであれば、 Z はいかなる共変量とも独立であるから、(a)あるいは(b)のみとなる。共変量が Y のみに影響を与える(b)の状況では、効果の修飾(modification)があるといい、 X を modifier と呼ぶ。

実験計画で処置への割付けがランダムでなく、処置の選択が被験者自身にゆだねられる自己選択(Self-selection)の場合、(c)か(d)である。(d)における X を交絡因子*(confounding factor)と呼ぶ。

* X は Z にも Y にも関係している



交絡の調整



共変量 X の値ごとに Z と Y との関係有条件付きで評価し、それを X の分布で統合することで調整。

共変量にする調整

- ① 実験計画段階で、処置群と対照群で共変量の分布が同じになるような抽出を行う。

層化 (Stratification) 小分類法 (Subclassification)

- ② 解析段階での調整には共分散分析 (analysis of covariance: ANCOVA) 及び、ロジスティック回帰がある。

$$Y = \alpha + \tau Z + \beta X + \varepsilon$$

$$\log\{P/(1-P)\} = \alpha + \tau Z + \beta X \quad (Y = \{0, 1\})$$

逆確率重み付け (Inverse Probability Weighting: IPW) 個体 i が抽出される確率を P_i として、結果 Y_i の観測値に抽出確立の逆数 $1/P_i$ の重みを加えて分析

シンプソン (1901) は分割表を無条件で併合してはならない、因果の方向はどちらかといった条件に依存することを示している。

Rubin Causal Model (RCM)

1.3 因果関係の統計的評価法

- (a) 潜在的な結果に基づく反事実モデル(Counterfactual model)
- (b) 同時方程式モデル(Simultaneous equation model)
- (c) 構造方程式モデル(Structural equation model)
- (d) 有向非巡回グラフ(Directed Acyclic Graph: DAG)に基づくグラフィカルモデリング

参考文献

岩崎学(2015)『統計的因果推論』朝倉書店.

第2講 群間比較

本講は岩崎(2015, 第2章)を用いて、群間比較の基礎について論じる。

2.1 有効率比較

処置効果を有効(1)、無効(0)の2項評価とする場合、それぞれの有効率は

$$P_1 = P(Y_1 = 1) = P(Y = 1|Z = 1)$$

ここで処理群($Z = 1$)、対照群($Z = 0$)とする。

$$P_0 = P(Y_0 = 1) = P(Y = 1|Z = 0)$$

これらの期待値

$$P_1 = E(Y_1) = E(Y|Z = 1), \quad P_0 = E(Y_0) = E(Y|Z = 0)$$

処置効果は有効率 P_1 と P_0 の比較で評価

差: $\delta = P_1 - P_0$	$-1 \leq \delta \leq 1$	$\delta = 0 (P_1 = P_0)$
比: $\gamma = P_1/P_0$	$0 \leq \gamma < \infty$	$\gamma = 1 (P_1 = P_0)$
オッズ比: $\omega = \frac{P_1/(1-P_1)}{P_0/(1-P_0)}$	$0 < \omega < \infty$	$\omega = 1 (P_1 = P_0)$

個体を各部にランダムに振り分けて観測する場合、観測結果は独立。

2.2 平均値の比較

結果変数 $Y_1 \sim N(\mu_1, \sigma_1^2)$ $Y_2 \sim N(\mu_0, \sigma_0^2)$ の正規分布に従っており、それぞれ独立しているとする。

平均値の差 $\delta = \mu_1 - \mu_0$ を評価する。

通常、 $\sigma_1^2 = \sigma_0^2 = (\sigma^2)$ と仮定: 等分散性の仮定

効果の大きさ (Effect Size: ES) $ES = (\mu_1 - \mu_0)/\sigma$

仮説検定

$$H_0 = \mu_1 - \mu_0 \text{ (帰無仮説)} \quad H_1 = \mu_1 \neq \mu_0 \text{ (対立仮説)}$$

2標本で検定(自由度に注意)を行う

	誤りではない H_0 が(正しい)	H_0 が誤り
H_0 を棄却しない	①	②
H_0 を棄却する	③	④

①④は正しいが、②③を選択すると問題

③ 第一種の誤り: H_0 が正しい時に、これを棄却する誤り

② 第二種の誤り: H_0 が誤っている時に、これを採択する誤り

品質管理の分野では第一種の誤りは生産者のリスク、第二種の誤りは消費者のリスクとされる。一般には消費者のリスクを避けることが重視されるので、第二種の誤りに対してより慎重になる。刑事訴訟で推定無罪とするのも、第二種の誤りを回避する態度の表われ。

科学では「正しい」かどうかは分からないが「誤っている」ことは判断できる。統計的に棄却できないということは「正しい」とか「採択」されたと考えるより、とりあえずこの仮説は誤っていると判定されていないので、次の仮説が出てくるまで保留しておこうというぐらいのこと。

2.3 回帰分析

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_u X_u + \varepsilon$$

$$Y = \beta_0 + \beta_1 Z + \varepsilon \quad Z = \begin{cases} 1 \text{ (処置)} \\ 0 \text{ (対照)} \end{cases}$$

であれば、等分散を仮定した 2 標本 t 検定と同一

結果変数 Y に対して、説明変数 X_i および U とし、 X_i は観測されるが、 U は観測されない要因すべてを表すとす。

$$E \begin{pmatrix} Y \\ X \\ U \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_x \\ \mu_u \end{pmatrix} \quad V \begin{pmatrix} Y \\ X \\ U \end{pmatrix} = \begin{pmatrix} \sigma_y^2 & \sigma_{yx} & \sigma_{yu} \\ \sigma_{yx} & \sigma_x^2 & \sigma_{xu} \\ \sigma_{yu} & \sigma_{xu} & \sigma_u^2 \end{pmatrix}$$

$$(i) \quad \begin{matrix} U \\ \searrow \\ X \end{matrix} \rightarrow Y \quad \text{であれば、} \sigma_{xu} = 0$$

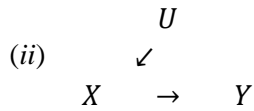
$$U \text{ が観測されなければ } Y = \alpha + \beta X + \varepsilon$$

$$U \text{ が観測されれば } Y = \alpha + \beta X + \gamma U + \varepsilon$$

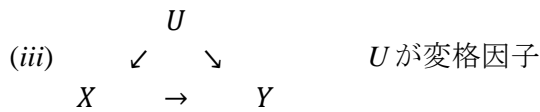
ε は X と U とともに独立な誤差項

$$\begin{aligned} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} &= \begin{pmatrix} \sigma_x^2 & \sigma_{xu} \\ \sigma_{xu} & \sigma_u^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{yx} \\ \sigma_{yu} \end{pmatrix} = \frac{1}{\sigma_x^2 \sigma_u^2 - \sigma_{xu}^2} \begin{pmatrix} \sigma_u^2 & -\sigma_{xu} \\ -\sigma_{xu} & \sigma_x^2 \end{pmatrix} \begin{pmatrix} \sigma_{yx} \\ \sigma_{yu} \end{pmatrix} \\ &= \frac{1}{\sigma_x^2 \sigma_u^2 - \sigma_{xu}^2} \begin{pmatrix} \sigma_u^2 \sigma_{yx} - \sigma_{xu} \sigma_{yu} \\ -\sigma_{xu} \sigma_{yx} + \sigma_x^2 \sigma_{yu} \end{pmatrix} \end{aligned}$$

$\sigma \times U \Rightarrow \beta = \frac{\sigma_{yx}}{\sigma_x^2}$ となり、 U とは独立して推定できる。母集団パラメータ β の不偏推定量となる。また $\gamma = \frac{\sigma_{yu}}{\sigma_u^2}$ U の値によって Y への効果が異なる。



X の値で条件付けされた期待値 Y を求めれば良いので、この場合も推定可能。



$$\beta = \frac{\sigma_u^2 \sigma_{yx} - \sigma_{xu} \sigma_{yu}}{\sigma_x^2 \sigma_u^2 - \sigma_{xu}^2} = \frac{\sigma_{yx} - \sigma_{xu} \sigma_{yu} / \sigma_u^2}{\sigma_x^2 - \sigma_{xu}^2 / \sigma_u^2}$$

右辺のパラメータに標本から求めた分散共分散を代入し、推定量

$\beta = (S_u^2 S_{yx} - S_{xu} S_{yu}) / (S_x^2 S_u^2 - S_{xu}^2)$ が計算される。 X と ε 、 U と ε は独立しているので、

$$\sigma_{yx} = \text{cov}[X, \beta X_t \gamma U_t \varepsilon] = \beta \sigma_x^2 + \gamma \sigma_{xu}$$

$$\sigma_{yu} = \text{cov}[U, \beta X_t \gamma U_t \varepsilon] = \beta \sigma_{xu} + \gamma \sigma_u^2$$

これを用いて β の不偏推定量が得られる。

X の係数は U を与えたとの条件の下で、 U の影響を取り除いた残りの残差に関する X と Y の間の単回帰係数となっている。

2.4 共分散分析(ANCOVA)

Cochran (1957)は共分散分析の役割として、

- (i) 実験研究における推測の精度向上
- (ii) 観察研究における共変量の偏りの調整
- (iii) 処置効果の意味付けへの寄与
- (iv) 分類データによる回帰の当てはめ
- (v) 欠測データへの対処

共分散分析は重回帰モデルにおいて、1つの変量を処置ダミー変数としたものに相当している。

結果変数 Y に対し、処置の割付けを表すダミー変数 Z ($1=$ 処理、 $0=$ 対照)。観測される共変量を X 、観測されない共変量を U とする。

(Y, Z, X) の期待値と分散共分散行列を次のように表わす。

$$E \begin{pmatrix} Y \\ Z \\ X \end{pmatrix} = \begin{pmatrix} \mu_Y \\ \mu_Z \\ \mu_X \end{pmatrix} \quad V \begin{pmatrix} Y \\ Z \\ X \end{pmatrix} = \begin{pmatrix} \sigma_Y^2 & \sigma_{YZ} & \sigma_{YX} \\ \sigma_{YZ} & \sigma_Z^2 & \sigma_{ZX} \\ \sigma_{YX} & \sigma_{ZX} & \sigma_X^2 \end{pmatrix}$$

$$p^{(1)} = p(Z = 1) = E(Z)$$

$$p^{(0)} = p(Z = 0) = E(1 - Z) = 1 - p^{(1)} \quad \text{とすると}$$

$$\sigma_Z^2 = E(Z^2) - E([Z])^2 = p^{(1)} - (p^{(1)})^2 = p^{(1)}p^{(0)}$$

$$E(Y) = p^{(1)}E(Y|Z = 1) + p^{(0)}E(Y|Z = 0) = p^{(1)}\mu_y^{(1)} + p^{(0)}\mu_y^{(0)}$$

$$E(YZ) = p^{(1)}E(Y|Z = 1) = p^{(1)}\mu_y^{(1)}$$

$$\Rightarrow \sigma_{yz} = E(YZ) - E(Y)E(Z) = p^{(1)}p^{(0)}(\mu_y^{(1)} - \mu_y^{(0)})$$

$$\sigma_{yx} = E(ZX) - E(Z)E(X) = p^{(1)}p^{(0)}(\mu_x^{(1)} - \mu_x^{(0)})$$

ただし、 $Z = 0, 1$ に対して $\mu_y^{(Z)} = E(Y|Z = z)$ が、 $Z = z$ とした時の Y の条件付き期待値を表すのに対し、 $\mu_x^{(Z)} = E(X|Z = z)$ は結果として、 $Z = z$ であった時の X の条件付き期待値という意味である。 Z が定められた後に Y が観測されるのに対して、 X は Z の後に観測されるのではなく、 X が先にあってそれに基づいて Z が定められるためである。

共分散分析のモデル

$$Y = \alpha + tZ + \beta X + U$$

$$t = \frac{\sigma_x^2 \sigma_{yz} - \sigma_{zx} \sigma_{yx}}{\sigma_z^2 \sigma_x^2 - \sigma_{zx}^2} = \frac{\sigma_{yz} - \sigma_{zx} \sigma_{yx} / \sigma_x^2}{\sigma_z^2 - \sigma_{zx}^2 / \sigma_x^2} = \frac{(\mu_y^{(1)} - \mu_y^{(0)}) - (\mu_x^{(1)} - \mu_x^{(0)}) \sigma_{yx} / \sigma_x^2}{1 - p^{(1)} p^{(0)} (\mu_x^{(1)} - \mu_x^{(0)})^2 / \sigma_x^2}$$

もし $\mu_x^{(1)} - \mu_x^{(0)} = 0$ であれば、 $\mu_y^{(1)} - \mu_y^{(0)} = \tau$ となる。

$$(i) \quad \begin{array}{ccc} X & \leftarrow & U \\ & \searrow & \downarrow \\ Z & \rightarrow & Y \end{array}$$

処置の割付け Z がランダムに行われる場合、 Z から Y への因果効果は不偏推定できる。

$$(ii) \quad \begin{array}{ccc} X & \leftarrow & U \\ \downarrow & \searrow & \downarrow \\ Z & \rightarrow & Y \end{array}$$

処理の割付け Z には X のみが影響を与え、 U は影響を与えない場合、 X は変格因子であり、 Z から Y への因果効果の推定では X を調整する。共分散分析モデルでは、不偏推定が可能。

U が Z 及び X に影響を与えている場合には、不偏推定は不可能である。できるだけ U を減らして偏りを最小限に止める。

$$(iii) \quad \begin{array}{ccccc} & & U & & \\ & & \downarrow & \searrow & \\ X & \rightarrow & Z & \rightarrow & Y \end{array}$$

U は Z と Y には影響を及ぼすが、 X には影響を及ぼしていない場合、 X は操作変数 (instrumental variable) と呼ばれ、観測されない共変量 U の影響を排除して、処置効果の偏りのない推定が可能となる。

2.5 ロジスティック回帰モデル

結果変数 Y が二項分布 (1: 有効、0: 無効) になっている場合の分析方法。確率 P を対数オッズにロジット変換し、

$$\omega = \text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \alpha + \beta x$$

というモデルを考える $-\infty < \omega < \infty$

$$P = F(\omega) = \frac{1}{1 + \exp(-(\alpha + \beta x))}$$

ここで $\exp(x) = e^x$ で与えられる確率分布を、累積分布関数が $F(\omega)$ ロジスティック分布という。ロジスティック分布の確率密度関数は

$$f(\omega) = F'(\omega) = \exp(-\omega)/(1 + \exp(-\omega))^2$$

となる。

処置の割付け Z (1: 処置、0: 対照) とし、有効率を $P_1 = P(Y = 1|Z = 1)$ 、 $p_0 = P(Y = 1|Z = 0)$ とした時のロジスティック回帰モデルは

$$\log(P/(1-P)) = \alpha + \tau Z$$

$Z = 0$ で $\log(P_0/(1-P_0)) = \alpha$ 、 $Z = 1$ で $\log(P_1/(1-P_1)) = \alpha + \tau$ であるので、 α が対照群での有効率の対数オッズ、 t が処置群に対する有効率の対数オッズ比 $\tau = \log\left[\frac{P_1/(1-P_1)}{P_0/(1-P_0)}\right]$ である。

共変量 X がある場合、

$$\text{logit}(P) = \log\{P/(1-P)\} = \alpha + \tau Z + \beta X + \xi ZX$$

簡単化のために($X = 0, 1$)の二項変数であるとする。

$$P_{ZX} = P(Y = 1|Z = z, X = x) \text{ とすると}$$

$$Z = 1, X = 1: \log\{P_{11}/(1 - P_{11})\} = \alpha + \tau + \beta + \xi$$

$$Z = 1, X = 0: \log\{P_{10}/(1 - P_{10})\} = \alpha + \tau$$

$$Z = 0, X = 1: \log\{P_{01}/(1 - P_{01})\} = \alpha + \beta$$

$$Z = 0, X = 0: \log\{P_{00}/(1 - P_{00})\} = \alpha$$

これより α, τ, β, ξ について解くと

$$\alpha = \log \frac{P_{00}}{1 - P_{00}}, \tau = \log \frac{P_{10}}{1 - P_{10}} - \log \frac{P_{00}}{1 - P_{00}} = \log \frac{P_{10}/(1 - P_{10})}{P_{00}/(1 - P_{00})}$$

$$\beta = \log \frac{P_{01}}{1 - P_{01}} - \log \frac{P_{00}}{1 - P_{00}} = \log \frac{P_{01}/(1 - P_{01})}{P_{00}/(1 - P_{00})}$$

$$\xi = \left(\log \frac{P_{11}}{1 - P_{11}} - \log \frac{P_{01}}{1 - P_{01}} \right) - \left(\log \frac{P_{10}}{1 - P_{10}} - \log \frac{P_{00}}{1 - P_{00}} \right) = \log \frac{P_{11}/(1 - P_{11})}{P_{01}/(1 - P_{01})} - \log \frac{P_{10}/(1 - P_{10})}{P_{00}/(1 - P_{00})}$$

定数項 α は $X = 0$ での対照群での有効対数オッズ、 ξ は共変量ごとに処置効果が異なるかどうかという交互作用を表すパラメータであり、 $\xi \neq 0$ の時には共変量ごとの処置効果の大きさが異なることから結果の解釈は難しくなる。

τ は共変量にかかわらず、処置効果を処置群と対照群との間の対数オッズ比としたものである。

参考文献

岩崎学(2015)『統計的因果推論』朝倉書店。

第3講 統計的因果推論の枠組み

本講は岩崎 (2015, 第3章) を用いて、統計的因果推論の基礎について論じる。

3.1 処理効果の定義

- 個体処理効果: 個体 i の潜在的な結果を $\{Y_i(1), Y_i(0)\}$ とする時、その差 $\tau_i = Y_i(1) - Y_i(0)$ を個体処理効果 (Individual Treatment Effect: ITE) という。
- 平均処理効果: 母集団全体での個体処理効果の期待値 $\tau = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)] = \tau_1 - \tau_0$ を平均処理効果 (Average Treatment Effect: ATE) という。

$$\tau = PATE = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \frac{1}{N} \sum_{i=1}^N Y_i(1) - \frac{1}{N} \sum_{i=1}^N Y_i(0)$$

母集団平均処理効果

- 部分集合での平均処理効果

$$\tau_A = E(Y_i(1) - Y_i(0)|A) = \frac{1}{N_A} \sum_{i \in A} (Y_i(1) - Y_i(0))$$

N_τ を処置を受けた個体群として処置群での平均処理効果 (ATE) は

$$\tau_t = E(Y(1) - Y(0)|Z = 1) = \frac{1}{N_\tau} \sum_{Zi=1} (Y_i(1) - Y_i(0))$$

対照群での平均処理効果 (ATC) を同様に定義

$$\tau_c = E(Y(1) - Y(0)|Z = 0) = \frac{1}{N_c} \sum_{Zi=0} (Y_i(1) - Y_i(0))$$

$E(Y(0)|Z = 1)$ 及び $E(Y(1)|Z = 0)$ はともに反事実であり、 τ_t 、 τ_c も推定不可能である。

3.2 SUTVA 条件

母集団から得られた n 個の個体からなる標本での平均処理効果を標本平均処理効果 (Sample average treatment effect: SATE) という

$$\tau_s = SATE = \frac{1}{N} \sum_{i=1}^n (Y_i(1) - Y_i(0))$$

これらの個体処置効果が意味を持つためには、次の SUTVA 条件を満たす必要がある。

SUTVA 条件: 個体 i を特徴づける潜在的な結果 $\{Y_i(1), Y_i(0)\}$ が次の 2 条件を満足する時、SUTVA 条件 (Stable unit treatment value assumption) を満足するという。(a) 個体 i の潜在的な結果 $\{Y_i(1), Y_i(0)\}$ は他の個体の受ける処置に依存しない。(b) 個体 i に対する処置は 1 通りに定まる

- (a) は相互干渉がない (no interference) こと
- (b) 処置の内容を明確にし、曖昧にしないこと

3.3 処置の割付けと識別性

個体 i への処置の割付けを表す変数を Z_i (1: 処置、0: 対照) とすると、実際に観測される結果 (Observable outcome) は

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

と表せる。定義正值性 (Positivity) 処置を受ける確率が $0 < P(Z = 1) < 1$ となる。

個体が処置を受ける確率 $P(Z = 1)$ も対照となる確率 $P(Z = 0)$ も 0 でないと仮定するもの。

定義独立性 (independence) 割付け Z が潜在的な結果 $\{Y(1), Y(0)\}$ に依存しないことを $\{Y(1), Y(0)\}$ と Z は独立であるといい、 $\{Y(1), Y(0)\} \perp Z$ と書く。

処置が条個体に対して無作為に決められるランダム割付けであれば割付け変数 Z は個体のすべての属性とは無関係であり、潜在的な結果とも無関係となる。

事象 A, B に対し $A \perp B$ であれば、 $P(A|B) = P(A)$ である。

$$E(Y|Z = 1) = E(Y(1)|Z = 1) = E(Y(1))$$

$$E(Y|Z = 0) = E(Y(0)|Z = 0) = E(Y(0))$$

$$\Rightarrow \tau = E[Y|Z = 1] - E[Y|Z = 0] = E(Y(1)) - E(Y(0))$$

正規性および独立性の仮定によって、平均処置効果 τ は推定可能となる。

観測可能量によって母集団パラメータが一意的に推定可能であるとき、パラメータは識別 (identify) される。正規性及び独立性は平均処置効果 τ が識別されるための識別可能条件である。

3.4 共変量と条件付き独立

共変量 X が与えられたとき、潜在的な結果 $\{Y(1), Y(0)\}$ に割付け Z が依存しないことを、 X が与えられた条件の下で $\{Y(1), Y(0)\}$ と Z は条件付き独立であるといい、

$$\{Y(1), Y(0)\} \perp Z | X \quad (\text{A})$$

と書く。観測される共変量 X のみが処置の割付けに影響を与える変格因子であり、 X の他に割付け Z に影響を与えるような観測されない交絡因子はないということを意味している。

潜在的な結果 $\{Y(1), Y(0)\}$ が実際に比較可能であるためには、条件付き正値性

$$0 < P(Z = 1|X) < 1 \quad (\text{B})$$

の条件が必要となる。

(A)、(B)2つの条件は、因果効果の推定では基本的な条件であり、Rosenbaum and Rubin (1983a) は、これは強い意味での無視可能な割付け (Strongly ignorable treatment assignment) と命名している。

一般に事象 A, B, C に対して、 $A \perp B | C$ であれば $P(A|B, C) = P(A|C)$ であるので、条件付き独立性及び正値性が成り立つとき

$$E(Y|Z = 1, X) = E(Y(1)|Z = 1, X) = E(Y(1)|X)$$

$$E(Y|Z = 0, X) = E(Y(0)|Z = 0, X) = E(Y(0)|X)$$

が成り立つ。これらはいずれも一般にすべてが観測されるとは限らない潜在的な結果 $Y(Z)$ の条件付き期待値が、観測可能量 Y の条件付き期待値と等しくなるという条件であり、これらを X の分布で期待値をとることにより

$$Ex[E[Y|Z = 1, X]] = Ex[E[Y(1)|X]] = E[Y(1)] = \tau_1$$

$$Ex[E[Y|Z = 0, X]] = Ex[E[Y(0)|X]] = E[Y(0)] = \tau_0$$

これより平均処置効果は

$$\tau = \tau_1 - \tau_0 = Ex[E[Y|Z = 1, X]] - Ex[E[Y|Z = 0, X]]$$

と観測可能量によって推定可能となる。

これらの条件を満たすとき、共変量 X が与えられれば、観察研究は実験結果と同じような枠組みでの統計的推測が可能となる。

参考文献

岩崎学(2015)『統計的因果推論』朝倉書店.

第4講 マッチング

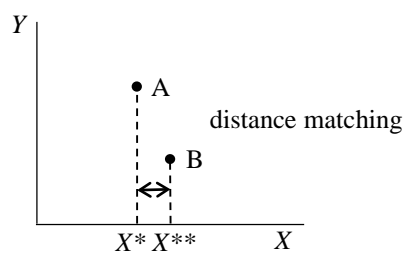
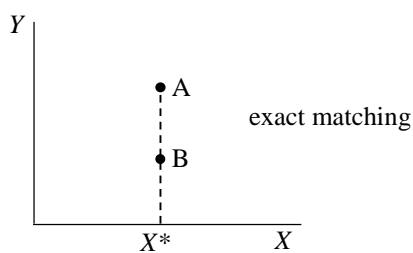
本講は岩崎(2015, 第5章)を用いて、マッチングの基礎について論じる。

4.1 マッチングの効用

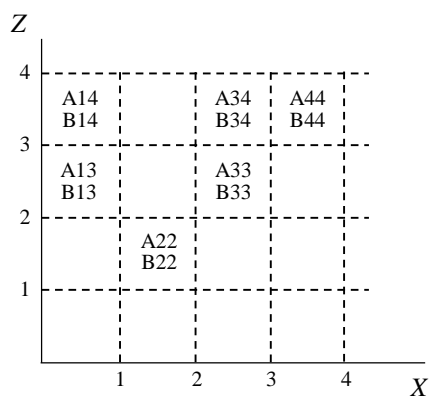
妥当性(validity)
処置効果を偏りなく推定できるか
不偏性(unbiasedness)

と
有効性(efficiency)
推定の精度が良いか

4.2 個別マッチング



Xの1次元



X, Zの2次元

マッチング推定効率の評価

- (a) マッチングによって選び出された個体の結果は、これから観測する
- (b) 対照となるべき個体の集合については、すべての結果は観測されている。

処置群と対照群とで背景因子の似通った個体を選び出し、それら同士を比較することで、背景情報のばらつきを抑えようという考え方

Mckinlay (1977) による Matching の利点

- (1) わかりやすく、統計の知識がなくても使える
- (2) 独立なサンプリングよりも、推定効率が良い
- (3) データの解析が容易

Matching の欠点

- (1) マッチングにはコストがかかる
- (2) 推定効率が下がることもある。個体の選択によりサンプルサイズが小さくなることで、標準誤差は大きくなる。
- (3) 事後的な共分散分析や事後層別に比べて、マッチングによる方法は必ずしも効率が良いわけではない。

4.3 距離の定義

各個体は一般に m 個の変量からなるベクトル $X = (x_1, \dots, x_m)^T$ によって特徴づけられ、第 i 個体の値を $X_i = (x_{i1}, \dots, x_{im})^T$ 、第 j 個体の値を $X_j = (x_{j1}, \dots, x_{jm})^T$ とする。

- (a) 重み付きユークリッド距離 (Weighted Euclidian Distance)

$$D_{ij} = \sum_{k=1}^m \omega_k (x_{ik} - x_{jk})^2$$

- (b) 重み付き市街地距離 (Weighted City-block Distance)

$$D_{ij} = \sum_{k=1}^m \omega_k |x_{ik} - x_{jk}|$$

(c) マハラノビス距離 (Mahalanobis Distance)

$$D_{ij} = (X_i - X_j)^T \Sigma^{-1} (X_i - X_j)$$

ここで Σ は共変量の分散共分散行列、ATT の推定では Σ は対照群全体での分散共分散行列、ATE の推定では Σ は処置群と対照群をプールした分散共分散行列とする。

(d) 傾向スコア (Propensity Score)

$$D_{ij} = |e(X_i) - e(X_j)|$$

(e) 線形傾向スコア (Linear Propensity Score)

$$D_{ij} = |\text{logit}(e(X_i)) - \text{logit}(e(X_j))|$$

$$\text{ただし、} \text{logit}(p) = \log(p/(c(-p)))$$

(a) (b) ですべての k に対し、 $\omega_k = 1$ とすれば、通常マージクド距離、及び市街地距離となる。ある k に対して、 $\omega_k = \infty$ とすれば、その変数に関しては、必ず一致しなければならない。すべての ω_k を ∞ とすれば、全変数が一致する必要があり、Exact Matching という。

傾向スコアマッチングの場合、重要と思われる共変量に関しては、マハラノビス距離などで直接的な近さを保証した方が望ましい。

あまりにも遠い個体同士をマッチングさせない工夫も必要。

$$D_{ij} = \begin{cases} (X_i - X_j)^T \Sigma^{-1} (X_i - X_j) (|\text{logit}(e(X_i)) - \text{logit}(e(X_j))| \leq c) \\ (|\text{logit}(e(X_i)) - \text{logit}(e(X_j))| > c) \end{cases}$$

とすれば、線形傾向スコアでの距離が c を超えたものはマッチングの対象とならない。この時 c をカリパー (Caliper) という。カリパーの目安は、線形傾向スコアの標準偏差の 0.2-0.25。

4.4 マッチング法

処置群での平均処置効果 ATT の推定では、処置群の個体 X_i に対し、対照群でのマッチング相手をいくつ選ぶかの選択の余地がある。

1:1 マッチング (one-to-one matching)

1:k マッチング (one-to-k matching) (k は処置群ごとに変えることも可能)

マッチング相手の選択を非復元 (without replacement) か、復元 (with replacement) かも選べる。

マッチングでは、処置群の個体と距離が近いものを対照群から選ぶ最近傍マッチング (Nearest Neighbor Matching) が用いられる。

処置群の個体を 1 つずつ順に選び、それらに対し距離が最も近い対照群での個体を非復元で選ぶというのが最も単純なマッチング法 (貪欲マッチング: Greedy Matching) であるが、「早い者勝ち」の方法であり、処置群の個体の選択順序に依存する。全体としての近さの尺度を定義し、その尺度が小さくなるように反復法によって全体のバランスをよくする方法 (最適マッチング: Optimal Matching) もある。

マッチング相手の選択では、対照群でのマッチング候補のプールが大きい場合、全データの一部しか選ばれないことになり、処置効果の推定の効率が落ちる。そこで、個体すべてに重み付けをして、全て利用する方法も考えられている。具体的には平均処置効果 ATE の推定では、 Z_i を処置を表すダミー変数 (1: 処置, 0: 対照) とした時、個体 i への重みを $\omega_i = Z_i/e(X_i) + (1 - Z_i)1 - e(X_i)$ とする。すなわち処置群の個体には傾向スコアの逆数 $1/e(X_i)$ を、対照群の個体には、 $1/(1 - e(X_i))$ を重みとして与える方法が用いられる。

4.5 マッチング効果の評価

共変量分布の重なり: Common Support (共通部分)

傾向スコア分布の重なりが全くないか、ほとんどない場合、それは処置群及び対照群が共変量によって判別されていることを意味する。

ほとんど共通部分がなければ、カリパーを設定したマッチングでは、マッチングできないことになり、処置効果の推定はできないという判断になる。

参考文献

岩崎学 (2015) 『統計的因果推論』朝倉書店。

Elizabeth A. Stuart (2010) “Matching methods for causal inference: A review and a look forward”, *Statistical Science* 25(1), pp.1-21. doi: 10.1214/09-STS313

第 5 講 傾向スコア法

本講は岩崎 (2015, 第 4 章)、Rosenbaum and Rubin (1983)、Guo and Fraser (2015) を参考にしながら、傾向スコアマッチング法について考える。

処置の割付けに関して最も広範に用いられているのが傾向スコア (propensity score) 法である。

定義 5.1 傾向スコア

処置の割付けを表すダミー変数 Z 及び観測された共変量の集合 X に対し、 X が与えられたときに個体が処置に割付けられる確率

$$e(X) = P(Z = 1|X)$$

を傾向スコアという。

傾向スコア $e(X)$ は共変量 X のもつ処置の割付けに関する情報をすべて集約したものである。傾向スコア法は Rosenbaum and Rubin (1983) によって導入され、その応用が多分野に広がっている。応用例に関しては、Guo and Fraser (2015) 等を参照。

傾向スコアの利用では、共変量 X が平均処置効果の識別可能条件を満たすかどうかことが重要となる。

条件 5.1 平均処置効果の識別可能条件

- (a) 条件付き独立性: 観測された共変量 X 以外に割付け Z に影響を与える変数はなく、潜在的な結果 $\{Y(1), Y(0)\}$ と割付け Z は条件付き独立となる。すなわち

$$\{Y(1), Y(0)\} \perp Z | X$$

が成り立つ。

- (b) 条件付き正值性: 与えられた共変量 X の下で、処置に割付けられる確率は 0 より大きく 1 未満である。

$$0 < e(X) < 1$$

条件 5.1 を満たせば、処置の割付けに関しては観測された共変量 X のみを考えればよいことになり、処置群と対照群間の共変量のインバランスの調整が可能となる。

定理 5.1 条件付き独立性

傾向スコア $e(X)$ が与えられたとき、潜在的な結果 $\{Y(1), Y(0)\}$ と割付け変数 Z は条件付き独立となる。すなわち

$$\{Y(1), Y(0)\} \perp Z | e(X)$$

が成り立つ。

[証明] 岩崎 (2015) p.101 を参照。

定理 5.2 バランシング

同じ傾向スコア $e(X)$ の値に対応した個体の共変量の分布は処置群と対象群で等しい(バランシングする)。すなわち、 $e(X)$ の条件付きで X と Z は条件付き独立

$$X \perp Z | e(X)$$

である。

[証明] 岩崎 (2015) p.101 を参照。

一般に、 X の関数 $b(X)$ で $X \perp Z | b(X)$ を満たすものをバランシングスコア (balancing score) と呼ぶ (Rosenbaum and Rubin (1983))。

傾向スコア $e(X)$ の値が与えられれば、その条件付きで処置の割付けはランダムとみなすことが出来る。 $e(X)$ の条件付きでの平均処置効果の推定が

$$\begin{aligned} E[Y(1) - Y(0) | e(X)] &= E[Y(1) | e(X)] - E[Y(0) | e(X)] \\ &= E[Y(1) | Z = 1, e(X)] - E[Y(0) | Z = 0, e(X)] \\ &= E[Y | Z = 1, e(X)] - E[Y | Z = 0, e(X)] \end{aligned}$$

ここで最終行は観測される値の期待値である。

母集団全体が平均処置効果 τ の推定は

$$\tau = E[Y(1) - Y(0)] = E_{e(X)}[E[Y|Z = 1, e(X)] - E[Y|Z = 0, e(X)]]$$

により示しうる。

定理 5.2 より、同じ $e(X)$ を持つ個体においては、それらの観測された共変量 X の分布は理論上両群で等しくなる。

統計的推論において、確率分布のパラメータに関する情報を十分統計量がすべて担っているとき、十分統計量で条件をつけた残りの変数の条件付き分布はパラメータに依存しないことと対応 (岩崎 (2015) p.99)

参考文献

岩崎学 (2015) 『統計的因果推論』朝倉書店.

Rosenbaum, Paul R. and Rubin, Donald B. (1983) “The Central Role of the Propensity Score in Observational Studies for Causal Effects” *Biometrika*, 70(1), pp.41–55. DOI:10.2307/2335942

Guo, Shenyang Y. and Fraser, Mark W. (2015). *Propensity Score Analysis -Statistical Methods and Applications*, 2nd edition, Sage Publishing. 参照先: <https://uk.sagepub.com/en-gb/asi/propensity-score-analysis/book238151>