

計量経済学応用

14. マッチング

矢内 勇生

2018年5月31日

高知工科大学 経済・マネジメント学群

今日の目標

- マッチング法の使い方を理解する

マッチングの目的

- 調査・観察データで、ランダム化比較実験に準じた分析を行う
 - ▶ 処置を受けた個体と処置を受けていない個体の間でバランスのとれた部分標本を探す
 - バランス：共変量の値が等しい
 - バランスしていない個体を除外することもある
 - ▶ 2段階の手続き
 - デザイン：結果変数を除いて、マッチングする
 - 分析：マッチングで得られた部分標本を使い、因果効果を推定する

マッチングの概要

- (1) 処置を受けたそれぞれの個体に対し、処置を受けていない個体の中から、処置変数（原因）以外の観察された共変量の値が等しい個体を1つまたは複数見つける
 - (2) 処置を受けていない個体のうち、(1) でマッチング相手に選ばれなかった個体を取り除く
- マッチングされた部分標本では、処置群と統制群の間で、共変量の分布は同じか非常に似たものになるはず
 - 処置群と統制群で結果変数の平均値に差があれば、それは処置の違いによるものであると考えられる

マッチングで何が推定できるか？ (1)

- 平均処置効果 (average treatment effect: ATE)

$$ATE = E[Y(1)] - E[Y(0)]$$

- 処置群に対する平均処置効果 (average treatment effect on the treated: ATT)

$$ATT = E[Y(1) \mid D = 1] - E[Y(0) \mid D = 1]$$

- 統制群に対する平均処置効果 (average treatment effect on the controlled: ATC)

$$ATC = E[Y(1) \mid D = 0] - E[Y(0) \mid D = 0]$$

マッチングで何が推定できるか？ (2)

- ATE を推定したい
 - ▶ 処置された個体も統制された個体も棄てられない
- ATT を推定したい
 - ▶ 処置群の個体は棄てられない：比較的大きな統制群が必要？
- ATC を推定したい
 - ▶ 統制群の個体は棄てられない：比較的大きな処置群が必要？

マッチング法を使った分析の手順

- (1) マッチングに利用する共変量を選ぶ
- (2) 「近さ」を定義する
- (3) マッチングを実行する
- (4) マッチング結果を評価し、うまくマッチングできるまで
(1) から (3) を繰り返す
- (5) 「いい」マッチング結果が得られたら、それを利用して
平均処置効果（因果効果）を推定する

どの共変量を使う？

- マッチングの目的：処置の割当がランダムになるようにしたい
 - ▶ 処置の割当に影響を与える（と考えられる）変数を共変量として利用する
 - 一般論として、共変量が多ければ多いほど良い
 - 処置に影響する共変量を入れ損なうとバイアスが生じる
 - 傾向スコアを使う場合、関係ない共変量を入れても問題はない
 - ▶ 処置の影響を受ける変数は使わない
 - ▶ 結果変数は絶対に使わない

「近さ」を定義する

- 個体同士をマッチングする際に使う基準（どれくらい似ていたらマッチングするか）を決める
- 様々な距離が利用される（詳細は省略）
 - ▶ 重み付きユークリッド距離
 - ▶ マンハッタン距離（City-block 距離）
 - ▶ マハラノビス距離
 - ▶ 傾向スコア距離

傾向スコア (Propensity Score)

- 個体 i と j の傾向スコア距離 : $|e_i - e_j|$
 - ▶ 傾向スコア : $e_i = \Pr(D_i = 1 \mid X_i)$
 - 共変量の値が X_i のときに、処置を受ける確率

マッチングの仕方

- いくつの個体をマッチングするか
 - ▶ 1対1?
 - ▶ k 対1? k は定数?
- 復元抽出 or 非復元抽出?
- マッチングされなかった個体をどうするか
 - ▶ 捨てる?
 - ▶ なんらかの方法でデータ分析に取り込む?

正確マッチング (Exact Matching)

- 共変量の値が完全に同じ個体同士をマッチングする
 - ▶ 長所：（うまくいけば）共変量については、処置群と統制群が完全にバランスする
 - ▶ 短所：共変量が完全に一致する個体を見つけるのは困難
 - 結果として、多くの個体がマッチング相手を見つけられない
 - 処置群（統制群）のすべての個体にマッチング相手を見つけないと、ATT (ATC) が推定できない（ATEもちろん推定できない）

最近傍マッチング

(Nearest Neighbor Matching)

- k 対1 最近傍マッチング：処置群のそれぞれの個体に対し、統制群の中から共変量の値に近い k 個の個体をマッチングする
 - ▶ ATT が推定できる
 - ▶ 最も単純な方法： $k = 1$ にする（1対1マッチング）
- マッチング相手に選ばれなかった統制群の個体は捨てる
 - ▶ 標本サイズが縮小するから検出力が小さくなる？
 - 2群のバランスが良くなるので、検出力は大きくなる！
- 十分な数のマッチング相手を見つけられないときはどうする？
 - ▶ うまくないマッチングとマッチングなしのどちらがいいか考える

最近傍マッチング：貪欲なマッチング

- 処置群の個体を何らかの方法で（例えば、ランダムに）並べ替える
- 最初の個体に対し、統制群の中で最も近い k 個体をマッチングさせる
- 第2の個体に対し、統制群の中で最も近い k 個体をマッチング
- ...
 - ▶ マッチング相手は、非復元抽出で行う： $i < j$ なら、 j 番目の処置群個体は、 i 番目の処置群個体のマッチング相手を選べない
- 問題：マッチングの順番によって答えが変わってしまう

最近傍マッチング：復元抽出する？

- 復元抽出を用いたマッチングの長所
 - ▶ バイアスが小さくなる
 - ▶ マッチングの順番を気にしなくてよい
 - ▶ 処置群のすべての個体にマッチング相手を見つけるために必要な統制群のサイズが、比較的小さい
- 復元抽出を用いたマッチングの短所
 - ▶ マッチングされた個体が独立ではなくなる
 - ▶ 重みを考える必要がある
 - ▶ 統制群の特定の個体は何回処置群にマッチングされた数える必要がある

最近傍マッチングの弱点

- すべての観測値を使うわけではない：情報が失われる
- 最近傍マッチング：マッチングされた各個体に重み w を与える ($0 < w < 1$)
- 代替案：すべての個体に重み w を与える ($0 \leq w \leq 1$)
 - ▶ 層化 (subclassification; 区間マッチング [interval matching])
 - ▶ 全員マッチング (full matching)
 - ▶ 重み付け (Weighting)

Subclassification (区間マッチング)

- 似た個体を集めた階級 (subclass) を複数作る
 - ▶ 例) 傾向スコアのパーセンタイルを使って、10%分ずつの階級を10作る
- ATT または ATE が推定できる
- 階級の数、5から10にすることが多い
- 標本サイズが大きいときは、10から20の階級を作れる

全員マッチング

- 階級の数自動的に決める
- 各階級は、処置群の個体と統制群の個体を少なくとも1つずつ含む
- ATT または ATE が推定できる
- 各階級内で、処置群の個体と統制群の個体の平均距離が最小化される

重み付けによる調整

- 傾向スコアの逆数を重みとして利用する
- IPTW (inverse probability of treatment weighting):

- ▶ ATE が推定できる

$$w_i = \frac{D_i}{\hat{e}_i} + \frac{1 - D_i}{1 - \hat{e}_i}$$

\hat{e}_i は個体 i の傾向スコアの推定値

- ▶ 処置群と統制群のすべての個体に重みを与える
- この他に、オッズによる重み付けも考えられる（詳細は割愛）

共有サポート (Common Support)

- 傾向スコアの分布が処置群と統制群で重なっているか確認する必要がある
- 重なっていないと、
 - ▶ 統制群における多くの個体の傾向スコアが、処置群の個体の傾向スコアからかけ離れている：ATT を推定することは適切ではない
 - ▶ 分析対象の制限が必要
- 共有サポート
 - ▶ 傾向スコアの分布に重なっている部分があるとき、
 - ▶ 重なっていない部分の観測値は捨てて、残った部分进行分析する

バランスを評価する

- マッチングがうまくいったか確認するため、処置群と統制群の間に観察された共変量がバランスしているか調べる
- 調べること
 - ▶ 理想：共変量の同時分布を2群間で比較する
 - ▶ 実践：各共変量の周辺分布を2群間で比較する
 - 平均値、分散、相関係数などが同じかどうか調べる

数値によるバラン診断

- 標準化バイアス (standardized bias or standardized difference in means):

$$\frac{\bar{x}_t - \bar{x}_c}{\sigma_t}$$

- ▶ 1つひとつの共変量、交差項、二乗項について調べる
- すべての標準化バイアスが0.1より小さくなるようにする

図によるバランス診断

- 2群間で、傾向スコアの分布を比較する
- Q-Q プロット（マッチング後の統制群 vs マッチング後の処置群）を作る
 - ▶ マッチング前の処置群 vs マッチング前の統制群のQ-Q プロットと比較し、バランスの改善具合も確認する
- 標準化バイアスをプロットする：すべてのデータ vs マッチングされたデータ

バランス検定？

- バランスを確かめるために、統計的検定（例えば、 t 検定）をしてはいけない！
- バランスするかしないかは標本（サンプル）の特性；検定は母集団の特徴を知るために実施すること
- この文脈で統計的検定をするのは混乱の元
 - ▶ 私たちは、群間に差がないことを示したい
 - ▶ 検定の帰無仮説は、「差がない」こと
 - ▶ したがって、帰無仮説を棄却して「差がある」ことは示せる
 - ▶ しかし、「差がない」ことは決して示せない：バランスしている証拠は出せない
 - ▶ 標本サイズが小さいと帰無仮説を棄却しにくくなる（検出力が小さくなる）ので、標本サイズが小さい方が有利というおかしい事態になってしまう

バランス後の分析

- デザインの段階（マッチング）が終了後、データを分析する
- マッチングがうまくいけば、処置が無作為に割り当てられたかのように（つまり、ランダム化比較実験の場合と同じように）データ分析できる
 - ▶ ATE、ATT、ATC のうちどれを推定するかに注意
- マッチングに使った共変量を統制変数として重回帰を行った方が良い：残差のバランスを改善するため
- 利用したマッチング法によっては、推定の段階で重みをつける必要がある