補足：マッチング推定量

講師：遠山祐太

最終更新：2024-11-16

はじめに

補足資料の概要

1. マッチング推定量の詳細
    - 近傍マッチング、傾向スコアマッチング、など。
    - 技術的な詳細については、例えば 因果推論入門〜ミックステープ：基礎から現代的アプローチまで-Scott-Cunninghamを参照

補足：マッチング推定量

推定手法

- $E[Y_i|D_i = 1, X_i = x]$ と $E[Y_i|D_i = 0, X_i = x]$ を推定したい。
- 実装方法
  1. 回帰：パラメトリック・ノンパラメトリック
  2. 最近傍マッチング
  3. 傾向スコアマッチング

方法その１：回帰・対応アプローチ

- $k \in \{0,1\}$ について、 $\hat{\mu}_k(x)$ を $\mu_k(x) = E[Y_i|D_i = k, X_i = x]$ の推定量とする。
- 対応推定量（analog estimator）は

$$\hat{ATE} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)$$

$$\hat{ATT} = \frac{N^{-1}\sum_{i=1}^{N} D_i(Y_i - \hat{\mu}_0(X_i))}{N^{-1}\sum_{i=1}^{N} D_i}$$

- $\mu_k(x) = E[Y_i|D_i = k, X_i = x]$ はどのように推定すればよいか？

ノンパラメトリック推定

- 小さな K について $X_i \in \{x_1, \cdots, x_K\}$ が離散であるとする。
  - 例：人口統計上の属性（男女・白人/非白人）ならば $K = 4$
- このとき、ノンパラメトリックな離散化（binning）推定量は

$$\hat{\mu}_k(x) = \frac{\sum_{i=1}^{N} \mathbf{1}\{D_i = k, X_i = x\} Y_i}{\sum_{i=1}^{N} \mathbf{1}\{D_i = k, X_i = x\}}$$

- ここでは、 $\mu_k(x) = E[Y_i | D_i = k, X_i = x]$ の推定について、いかなる仮定もおいていない。

次元の呪い

- 問題：ノンパラメトリック推定は、 K が大きくなると共変量との関係でパフォーマンスが下がる。
  - 潜在的に考えられる群がたくさんある一方、それぞれの群について観測がほとんどない。
  - K 個の変数があり L 種類の値を取る場合、 L^K の可能な群がある。
- この事実は**次元の呪い（curse of dimensionality）**として知られている。
- X が連続の場合は、ノンパラメトリック手法としてカーネル回帰を用いることができる。

パラメトリック推定

- 次のように、パラメトリックな仮定をおくことを考える。

$$E[Y_i|D_i = 0, X_i = x] = \beta' x_i$$
$$E[Y_i|D_i = 1, X_i = x] = \beta' x_i + \tau_0$$

このとき、

$$y_i = \beta' x_i + \tau D_i + \epsilon_i$$

なるモデルを得る。

- マッチング推定量は、共変量（統制変数） x_i を追加することによって、欠落変数バイアスを統制するものだと捉えることができる。

方法その2： M -近傍マッチング

- アイデア：別の群で自分に近い対応を探索する。
- $\hat{y}_i(0)$ および $\hat{y}_i(1)$ を、処置が施される場合と施されない場合の（仮説的な）アウトカムの推定量とする。

$$\hat{y}_i(0) = \begin{cases} y_i & if\ D_i = 0 \\ \frac{1}{M}\sum_{j \in L_M(i)} y_j & if\ D_i = 1 \end{cases}$$

- 個体 i に「近い」別の群の M 個体の集合を $L_M(i)$ とする。
- X_i と X_j の距離を定義する方法はいくつもあり、例えば

$$dist(X_i, X_j) = ||X_i - X_j||^2$$

- (1) M と (2) 距離の種類を選ぶ必要がある。

方法その３：傾向スコアマッチング

- 傾向スコア $P(D_i = 1 | X_i = x)$ を最近傍を定義する距離として採用する。
- Step 1: X_i に関する柔軟な関数を用いて、ロジスティック回帰またはプロビット回帰で傾向スコアを推定する。
- Step 2 : それぞれの観測について傾向スコアを計算し、ペアを決める。