Causal Inference with Matching

Lê Việt Phú
Fulbright School of Public Policy and Management

Ngày 1 tháng 7 năm 2019

Thiết lập quan hệ nhân quả bằng matching

- Matching là một thiết kế nghiên cứu dựa trên giả định quan sát được đặc tính giải thích cho vấn đề lựa chọn mẫu (selection on observables).
- Matching khác với hồi quy là không dựa trên tham số (nonparametric methods).
- Có rất nhiều phương pháp matching, tuy nhiên cốt lõi của tất cả các phương pháp là đảm bảo điều kiện cân bằng giữa hai nhóm hưởng lợi và đối chứng.

Cơ chế của phương pháp matching

- Với cơ chế matching 1-1:

$$ATT = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)})$$

trong đó $Y_{j(i)}$ là biến kết quả của quan sát $j(i)$ có các đặc tính quan sát được $X_{j(i)}$ gần với X_i nhất.

- Mở rộng matching với M quan sát gần nhất:

$$ATT = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_i - (\frac{1}{M} \sum_{m=1}^{M} Y_{jm(i)}) \right\}$$

Ví dụ với matching chỉ với một biến quan sát

unit	Potential Outcome (D=1)	Potential Outcome (D=0)		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	?	1	3
2	1	?	1	1
3	0	?	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

$$ATT = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)}) = ?$$

unit	Potential Outcome (D=1)	Potential Outcome (D=0)		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	9	1	3
2	1	0	1	1
3	0	9	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

$$ATT = \frac{1}{3}\Big\{(6-9)+(1-0)+(0-9)\Big\} = -3.7$$

Matching khi có nhiều đặc tính quan sát được

Lời nguyền về dữ liệu đa chiều (Curse of dimensionality): Độ khó của việc ghép được dữ liệu tăng theo cấp số mũ mỗi khi thêm một chiều không gian dữ liệu.

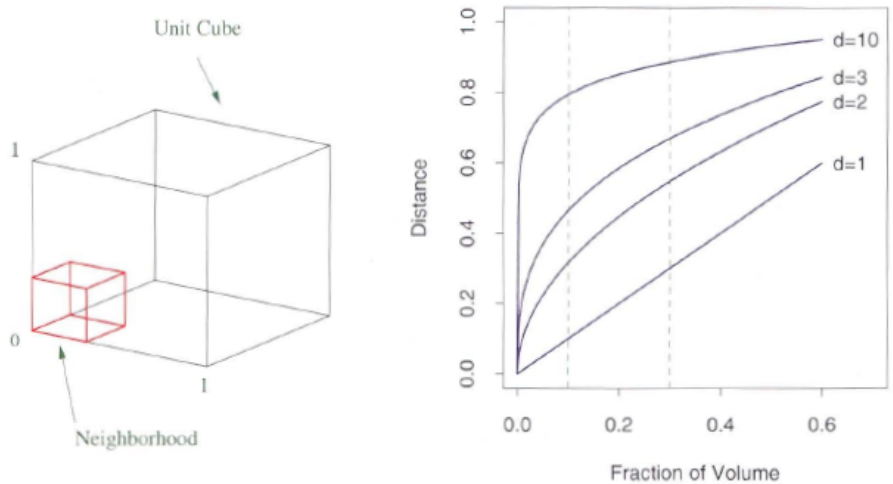


FIGURE 2.6. *The curse of dimensionality is well illustrated by a subcubical neighborhood for uniform data in a unit cube. The figure on the right shows the side-length of the subcube needed to capture a fraction r of the volume of the data, for different dimensions p. In ten dimensions we need to cover 80% of the range of each coordinate to capture 10% of the data.*

Matching với dữ liệu đa chiều

- Propensity score matching: xây dựng xác suất tham gia chương trình bằng hồi quy với các đặc tính quan sát được.
- Xây dựng hàm khoảng cách (distance metric) để xác định tính chất giống nhau giữa các nhóm tham gia và đối chứng dựa trên các đặc điểm quan sát được.

Hàm khoảng cách Mahalanobis

- Giả sử $X = (X_1, X_2, ..., X_k)$ là vector các đặc tính quan sát được. Hàm khoảng cách giữa hai quan sát i và j theo phương pháp Mahalanobis được tính như sau:

$$D_M(X_i, X_j) = \sqrt{(X_i - X_j)^T \Sigma_X^{-1} (X_i - X_j)}$$

  trong đó Σ_X là ma trận phương sai và hiệp phương sai của X.
- Khoảng cách càng nhỏ thì quan sát X_i và X_j càng gần nhau, hay i và j có thể ghép cặp được với nhau. $D_M = 0$ thì chúng ta có cặp ghép hoàn hảo.
- Khái niệm hàm khoảng cách tương tự như cách đo chiều dài (Euclidean distance), tuy nhiên áp dụng trong không gian đa chiều.

Ví dụ tính khoảng cách Mahalanobis

unit	X_1	X_2
Treated	0	0
Control A	5	5
Control B	4	0

với

$$\Sigma_X = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

Quan sát A hay B làm đối chứng tốt hơn?

$$D_M(X_i, X_j) = \sqrt{(X_i - X_j)^T \Sigma_X^{-1} (X_i - X_j)}$$

$$D_M(X_i, X_A) = \sqrt{(-5 \ \ -5) \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}^{-1} (-5 \ \ -5)^T}$$

và

$$D_M(X_i, X_B) = \sqrt{(-4 \ 0) \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}^{-1} (-4 \ 0)^T}$$

Cho biết:

$$\begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 5.2 & -4.7 \\ -4.7 & 5.2 \end{bmatrix}$$

Từ đó tính được $D_M(X_i, X_A) =?$ và $D_M(X_i, X_B) =?$

Điều kiện độc lập với dữ liệu thử nghiệm ngẫu nhiên và dữ liệu quan sát được

- Đối với thử nghiệm ngẫu nhiên đảm bảo việc phân bổ vào nhóm tham gia hay đối chứng hoàn toàn độc lập với kết quả chương trình:

$$Y_i^1, Y_i^0 \perp D_i$$

- Đối với dữ liệu quan sát được, và giả định việc lựa chọn mẫu dựa trên các đặc tính quan sát được (selection on observables):

$$Y_i^1, Y_i^0 \perp D_i | X$$

Ghép cặp dựa vào Propensity score

- Có thể chuyển đổi điều kiện trên thành lựa chọn mẫu dựa trên propensity score:

$$Y_i^1, Y_i^0 \perp D_i | p(X)$$

với $p(X) = P(D = 1|X)$. Điều kiện này được gọi là điều kiện "unconfoundedness", có nghĩa là nếu thay vì dùng các đặc tính quan sát được X_i để lựa chọn nhóm đối chứng và hưởng lợi, chúng ta có thể sử dụng điểm xu hướng.

Các bước thực hiện propensity score matching

- Ước lượng mô hình xác xuất $P(D_i = 1|X) = f(X_i)$ bằng hồi quy logit hay probit. Lưu ý phải lựa chọn các biến giải thích và cấu trúc hàm phù hợp.
- Ước lượng xác suất tham gia chương trình đối với mỗi quan sát i tại các giá trị X_i, gọi là điểm xu hướng (propensity score).
- Ghép các nhóm hưởng lợi và đối chứng dựa trên giá trị $p(D_i = 1|X)$ tương đồng. Có nhiều phương pháp ghép cặp khác nhau.
  - 1-1, 1-M, NN, caliper, kernel, entropy, genetic
- Kiểm tra các điều kiện cân bằng. Nếu không đảm bảo thực hiện lại từ đầu. Lặp lại cho đến khi điều kiện cân bằng được đảm bảo.
- Ước tính ATT từ các nhóm đối tượng được có thể ghép cặp.

Các phương pháp ước lượng khác sử dụng propensity score

- Sử dụng PS để điều chỉnh hàm hồi quy (regression adjustments with propensity score).
- Có thể dùng PS để làm quyền số để ước lượng ATT.
- Kết hợp cả hai phương pháp trên.

Hồi quy điều chỉnh sử dụng PS

- Với giả định “unconfoundedness", chúng ta có thể sử dụng hàm hồi quy:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 p(D_i = 1|X_i) + \varepsilon_i$$

và $ATT = \beta_2$

- Nếu cho phép tác động khác biệt, (heterogeneous effects), chúng ta sẽ ước lượng hàm hồi quy sau:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 p(D_i = 1|X_i) + \beta_3 D_i * p(D_i = 1|X_i) + \varepsilon_i$$

và $ATT = \beta_2 + \beta_3 * p(D_i = 1|X_i)$

Dùng PS làm quyền số (Weighting by propensity score)

- $$ATT_{naive} = \bar{Y}_T - \bar{Y}_C = \frac{\sum D_i Y_i}{\sum D_i} - \frac{\sum (1 - D_i) Y_i}{\sum (1 - D_i)}$$

  Ước lượng ATT_{naive} bị chệch do vấn đề lựa chọn mẫu với dữ liệu phi thử nghiệm.

- $$ATT_{p(X_i)} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{D_i Y_i}{p(X_i)} - \frac{(1 - D_i) Y_i}{1 - p(X_i)} \right\}$$

  Ước lượng $ATT_{p(X_i)}$ được gọi là inverse propensity score weighting (IPW).

- Hàm ý của sử dụng PS làm quyền số khi ước lượng ATT là gì?

Chuẩn hóa ước lượng IPW bằng công thức sau:

$$ATT_{p(X_i)} = \left(\sum_{i=1}^{N} \frac{D_i Y_i}{p(X_i)} \Big/ \sum_{i=1}^{N} \frac{D_i}{p(X_i)}\right) - \left(\sum_{i=1}^{N} \frac{(1-D_i) Y_i}{1-p(X_i)} \Big/ \sum_{i=1}^{N} \frac{1-D_i}{1-p(X_i)}\right)$$

- Ước lượng bằng IPW nhất quán nhưng bị chệch với cỡ mẫu nhỏ.
- Mức độ chệch có thể khá nghiêm trọng khi quyền số quá lớn hoặc quá nhỏ $\rightarrow$ Liên hệ với điều kiện overlapping/common support?

Kết hợp cả hai phương pháp hồi quy và điều chỉnh quyền số với PS

Chúng ta sẽ ước lượng hàm hồi quy sau

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \varepsilon_i$$

và sử dụng quyền số

$$\omega_i = \sqrt{\frac{D_i}{p(X_i)} + \frac{1 - D_i}{1 - p(X_i)}}$$

- Ước lượng bằng phương pháp kết hợp vững hơn các ước lượng khác khi cấu trúc hàm ước lượng tác động hoặc hàm ước lượng propensity bị sai (“double robustness").

Matching hay Hồi quy?

- Parametric or non-parametric?
  - Hồi quy cần giả định mạnh về cấu trúc hàm, trong khi non-parametric matching chỉ cần đảm bảo điều kiện cân bằng $\rightarrow$ minh bạch hơn và kết quả ít phụ thuộc vào kỹ thuật xây dựng mô hình.
  - Tuy nhiên, nếu dùng PS để giảm chiều của dữ liệu $\rightarrow$ sẽ gặp phải các vấn đề của hồi quy khi ước lượng mô hình PS.
- Phương pháp nào dễ thuyết phục hơn?
  - Matching chỉ sử dụng kết quả đầu ra để so sánh $\rightarrow$ Có thể thiết kế các nghiên cứu đảm bảo kết quả không bị chi phối bởi ý muốn chủ quan (p-hacking).

Điều kiện tiên quyết với matching

Đảm bảo các đặc tính quan sát được cân bằng:

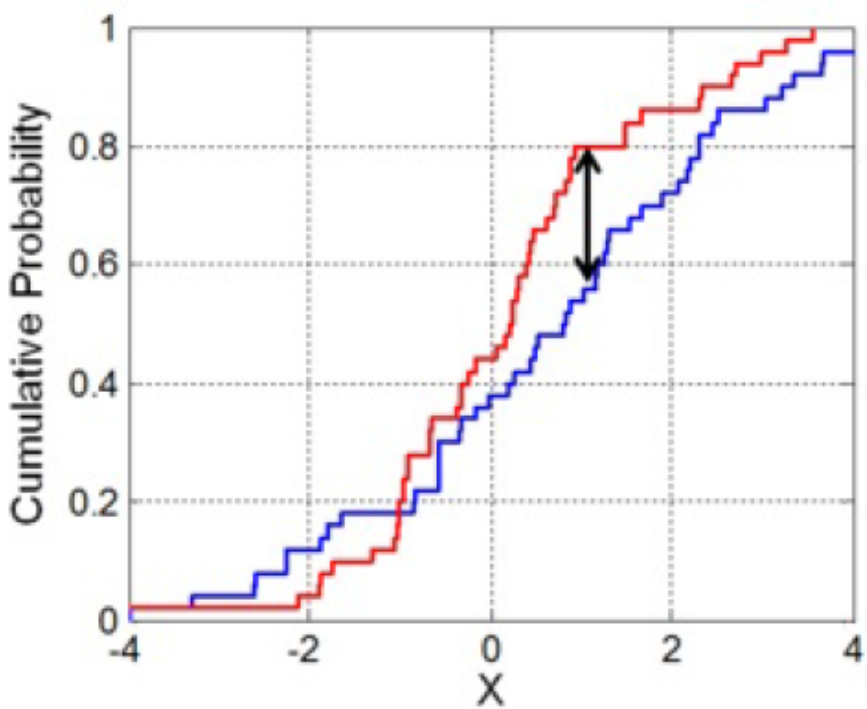
- Kiểm định T-test về giá trị trung bình. Giả thuyết H_o là giá trị trung bình tương đồng giữa các nhóm. Cần sử dụng mức ý nghĩa α thấp (hay chấp nhận xác suất xảy ra sai lầm loại 1 thấp).
- Kiểm định tương đồng (equivalence tests). Giả thuyết H_o là các nhóm khác biệt nhau. Dùng để xác định xác suất xảy ra sai lầm loại 2. Để giảm sai lầm loại 2 hay tăng độ mạnh thống kê thì cần tăng cỡ mẫu.
- Kiểm định phân phối Kolmogorov-Smirnov:
  - Xác định liệu hai nhóm dữ liệu quan sát được thu thập từ cùng một phân phối.
  - Sử dụng để nhận định sự khác biệt về phân phối.
- Sửu dụng thống kê mô tả, đồ thị phân phối, QQ plot...

Kiểm định Kolmogorov-Smirnov

Trị kiểm định là khoảng cách cực đại giữa hai phân phối thực nghiệm (empirical CDF) của hai nhóm hưởng lợi và kiểm soát.

$$D = sup_x |\hat{F}_1(x) - \hat{F}_0(x)|$$

với $\hat{F}_0(x)$ và $\hat{F}_1(x)$ là hai phân phối thực nghiệm của X_0 và X_1.



Kiểm định Kolmogorov-Smirnov

- Giả thuyết H_o là không có sự khác biệt về hàm phân phối thực của X_0 và X_1.
- D có phân phối Kolmogorov, và giá trị cực trị tại mức ý nghĩa α được tính như sau:

$$D_{critical} = c_\alpha \sqrt{(n_1 + n_0)/n_1 n_0}$$

và

α	.1	.05	.01
c_α	1.22	1.36	1.63

- Khi D lớn thì kết luận hai mẫu lấy từ hai phân phối khác nhau.