

Causal Inference with Matching

Lê Việt Phú
Fulbright School of Public Policy and Management

Ngày 18 tháng 7 năm 2020

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.

Các bước thực hiện đánh giá tác động can thiệp bằng matching

1. Thiết lập mức độ giống nhau giữa các quan sát dựa trên các đặc tính quan sát được nhưng có tình trạng can thiệp khác nhau, từ đó có thể ghép được với nhau.
2. Thực hiện các phương pháp ghép dựa trên mức độ giống nhau ở bước 1.
3. Đánh giá chất lượng ghép cặp, lặp lại bước 1-2 cho đến khi đạt chất lượng ghép mong muốn.
4. Phân tích kết quả và ước lượng tác động can thiệp.

Ước lượng mức độ giống nhau (measure of closeness) để thiết lập nhóm đối chứng

Cần xây dựng khái niệm khoảng cách/mức độ cận kề/tương đồng giữa hai quan sát bất kỳ. Sử dụng ký hiệu D_{ij} để mô tả khái niệm khoảng cách/mức độ tương đồng giữa hai quan sát i và j . Ví dụ một số cách xây dựng chỉ số khoảng cách:

1. Giống nhau tuyệt đối (exact matching):

$$D_{ij} = \begin{cases} 0 & \text{if } X_i = X_j \\ \infty & \text{if } X_i \neq X_j \end{cases}$$

với X_i và X_j là các thuộc tính quan sát được.

2. Giống nhau dựa trên xác suất tham gia (propensity score matching):

$$D_{ij} = |P(T_i = 1|X_i) - P(T_j = 1|X_j)|$$

với $P(\cdot)$ là xác suất tham gia, ước lượng từ hàm hồi quy xác suất của các đặc tính quan sát được X .

3. Dùng hàm khoảng cách tổng quát (generalized distance) để so sánh sự khác biệt giữa các quan sát (Mahalanobis distance¹):

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

với X là các thuộc tính quan sát được, Σ^{-1} là nghịch đảo của ma trận phương sai/hiệp phương sai của X (inverse variance-covariance matrix).

¹Tương tự như khoảng cách Euclidean đo chiều dài dựa trên hình học mặt phẳng.

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

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

Matching khi có nhiều đặc tính quan sát được (dữ liệu đa chiều)

Lời nguyên 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.

- ▶ Exact matching để tìm đối chứng là tối ưu, tuy nhiên khi số đặc tính (chiều của dữ liệu) tăng lên thì khó tìm quan sát tương đồng, dẫn đến mất dữ liệu.
- ▶ 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.
- ▶ Ghép bằng propensity score matching có ưu điểm là chỉ dùng một chỉ số được xây dựng từ các đặc tính quan sát được.

Hàm khoảng cách Mahalanobis

- ▶ Giả sử $X = (X_1, X_2, \dots, X_k)$ 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

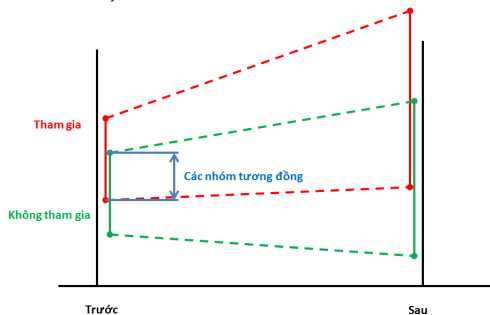
$$D_M(X_i, X_B) = \sqrt{(-4 \ 0) \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}^{-1} (-4 \ 0)^T}$$

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 $D_M(X_i, X_B) = ?$

Phương pháp so sánh bằng điểm xu hướng (propensity score matching-PSM)



- ▶ Bản chất của PSM là tìm nhóm hưởng lợi (treatment group) và nhóm đối chứng (control group) có tương đồng với nhau trong tổng thể mẫu dữ liệu bằng mô hình thống kê.
- ▶ Chúng ta xây dựng một chỉ số gọi là điểm xu hướng - propensity score. Điểm xu hướng là xác suất quan sát được một hộ có tham gia chính sách hay không.
- ▶ Các hộ gia đình có cùng điểm xu hướng gần nhau được ghép cặp và làm đối chứng cho nhau.

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):

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

- ▶ 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

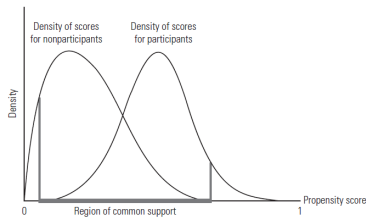
- ▶ Ước lượng mô hình xác suất $P(D_i = 1|X) = f(X_i)$ bằng hồi quy logit hay probit. Lưu ý phải lựa chọn các biến giải thích và cấu trúc hàm phù hợp.
- ▶ Ước lượng xác suất tham gia chương trình đối với mỗi quan sát i tại các giá trị X_i , gọi là điểm xu hướng (propensity score).
- ▶ Ghép các nhóm hưởng 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.

Vùng hỗ trợ chung - Common support

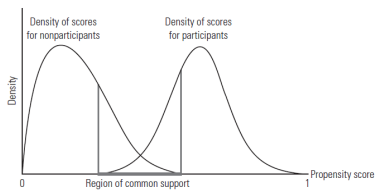
Vùng hỗ trợ chung (còn gọi là điều kiện trùng lặp - overlapping condition) là vùng có ước lượng điểm xu hướng (hoặc xác suất tham gia) của cả nhóm tham gia và nhóm kiểm soát. Hai nhóm phải có một số lượng quan sát có cùng điểm xu hướng thì mới ghép cặp được với nhau.

$$0 < P(T_i = 1|X_i) < 1$$

Vùng hỗ trợ chung tốt



Vùng hỗ trợ chung kém



Điều kiện có vùng hỗ trợ chung

- ▶ Điều kiện này đảm bảo có thể tìm được các quan sát đối chứng đối với một số đối tượng tham gia. Nếu các quan sát có điểm xu hướng khác biệt nhau quá thì không thể so sánh được với nhau.
- ▶ Có càng nhiều đối tượng tham gia và kiểm soát trong vùng hỗ trợ chung càng tốt.
- ▶ Quan sát không nằm trong vùng hỗ trợ chung sẽ bị loại do không tìm được nhóm đối chứng.

Tính tác động can thiệp trung bình lên đối tượng tham gia trong mô hình PSM

Nếu các điều kiện trên được thỏa mãn, thì tác động can thiệp trung bình đối với những người tham gia (ATT hay ATOT) được tính như sau:

$$ATT_{PSM} = \frac{1}{N_T} \left[\sum_{i \in T} Y_i^T - \sum_{j \in C} \omega(i, j) Y_j^C \right]$$

- ▶ ATT_{PSM} là khác biệt trung bình (có quyền số) giữa nhóm tham gia và nhóm đối chứng có cùng điểm xu hướng.
- ▶ $\omega(i, j)$ là quyền số, tùy thuộc vào phương pháp ghép cặp.
- ▶ Do tồn tại vấn đề lựa chọn mẫu nên $ATE \neq ATT$ trong phương pháp PSM.

Các hình thức xác định nhóm đối chứng và quyền số $\omega(i, j)$ dựa vào điểm xu hướng

1. Phương pháp lựa chọn n láng giềng gần nhất (n nearest neighbor matching): Với mỗi đối tượng tham gia, lựa chọn n quan sát không tham gia có chỉ số xu hướng gần nhất làm nhóm đối chứng.
2. Chọn theo bán kính (caliper or radius matching): Ghép quan sát có chỉ số xu hướng nằm trong khoảng $\pm r$ cho trước.
3. Chọn phân tầng và so sánh trong từng khoảng giá trị cho trước (stratification or interval matching).
4. Chọn bằng hồi quy nội tại sử dụng phương pháp phi tham số (kernel and local linear matching).

Ví dụ ghép cặp 1-n

- ▶ 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:

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

Các phương pháp ước lượng khác sử dụng propensity score

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

Hồi quy điều chỉnh sử dụng propensity score

- ▶ Với giả định "unconfoundedness", chúng ta có thể ước lượng tác động can thiệp bằng hàm hồi quy điều chỉnh cho xác suất can thiệp:

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

$$\text{và } ATT = \beta_1$$

- ▶ Nếu nghi ngờ can thiệp có tính chất khác biệt (heterogeneous effects), chúng ta sẽ ước lượng hàm hồi quy sau:

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

$$\text{và } ATT = \beta_1 + \beta_3 * p(D_i = 1|X_i)$$

Dùng propensity score làm quyền số (Weighting by propensity score)

- ▶ Tác động can thiệp trung bình:

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

Dùng ATE_{naive} để ước lượng ATT bị chệch do vấn đề lựa chọn mẫu với dữ liệu phi thử nghiệm.

- ▶ Dùng propensity score để điều chỉnh plugged-in estimator:

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

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

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

Chuẩn hóa (normalize) ước lượng IPW bằng công thức sau:

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

- ▶ 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 → Liên hệ với điều kiện overlapping/common support?

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

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").

Thực hành ước lượng tác động can thiệp và so sánh kết quả giữa các phương pháp ghép cặp

- ▶ Cài đặt package psmatch2 (findit psmatch2).
- ▶ Thực hành với dữ liệu mô phỏng và các hình thức ước lượng bằng psmatch2.
- ▶ So sánh kết quả giữa ghép bằng Mahalanobis distance với propensity score matching/weighting.
- ▶ Một giải pháp khác là coarsened exact matching nhằm khắc phục tính đa chiều của dữ liệu. CEM làm nhám/thô bộ dữ liệu (coarsen) sau đó match trên dữ liệu đã làm thô. Cài đặt package cem (findit cem).

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 → minh bạch hơn và kết quả ít phụ thuộc vào kỹ thuật xây dựng mô hình.
 - Tuy nhiên, nếu dùng propensity score để giảm chiều của dữ liệu → sẽ gặp phải các vấn đề của hồi quy khi ước lượng mô hình propensity score.
- ▶ Phương pháp nào dễ thuyết phục hơn?
 - Matching chỉ sử dụng kết quả đầu ra để so sánh → 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:

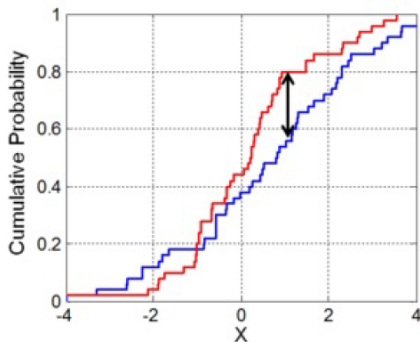
- ▶ Kiểm định T-test về giá trị trung bình. Giả thuyết H_0 là giá trị trung bình tương đồng giữa các nhóm. Cần sử dụng mức ý nghĩa α thấp (hay chấp nhận xác suất xảy ra sai lầm loại 1 thấp).
- ▶ Kiểm định tương đồng (equivalence tests). Giả thuyết H_0 là các nhóm khác biệt nhau. Dùng để xác định xác 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.
- ▶ Sử 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_0 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.
- ▶ **Stata implementation: "findit escftest"**