Propensity score matching (PSM)

Jerzy Mycielski

Uniwersytet Warszawski

Maj 2010

Badania ewaluacyjne

- Ocena wpływu bodźca
- Przykłady:
  - ocena skuteczności programów aktywizacji zawodowej
  - ocena skuteczności dotacji mających na celu zachęcanie do innowacyjności
- Idealny eksperyment społeczny

1. jednostki należące do grupy eksperymentalnej i kontrolnej mają ten sam rozkład cech nieobserwowalnych
2. jednostki należące do grupy eksperymentalnej i kontrolnej mają ten sam rozkład cech obserwowalnych
3. ten sam kwestionariusz został zastosowany w przypadku grupy eksperymentalnej i kontrolnej - cechy i wyniki są mierzone w ten sam sposób
4. obie grupy znajdują się w tym samy otoczeniu ekonomicznym

Badania ewaluacyjne c.d.

- Idealną metodą ewaluacyjną jest eksperyment zrandomizowany
  - w przypadku takiego eksperymentu założenia (1)-(4) są automatycznie spełnione
- Tradycyjna analiza ekonomiczna dotycząca problemów związanych z ewaluacją w przypadku danych nieeksperymentalnych koncentrowały się na problemie z założeniem (1) - np. model Heckmana
- w praktyce jednak ważniejsze wydają się problemy z założeniami (2), (3), (4).
- Metoda PSM umożliwia znaczną redukcję obciążenia estymacji efektu bodźca jeśli nie jest spełnione założenie (2).

Notacja

- $i \in I_1$ grupa eksperymentalna
- $j \in I_0$ grupa kontrolna
- $D \in \{0, 1\}$
  - $D = 0$ jednostka nie została poddana działaniu bodźca (niebodźcowana - untreated)
  - $D = 1$ jednostka została poddana działaniu bodźca (bodźcowana - treated)
- $\mathbf{X}$ - wektor charakterystyk jednostki
- $\Pr(\mathbf{X})$ prawdopodobieństwo znalezienia się w grupie eksperymentalnej (propensity score)
- Efekt przyczynowy mierzony jest za pomocą zmiennej wynikowej Y
  - Y_0 wynik w przypadku jednostki niebodźcowanej
  - Y_1 wynik w przypadku jednostki bodźcowanej

Efekt przyczynowy

- Efekt przyczynowy (Treatment Effect) dla jednostki o charakterystykach $\mathbf{X}$

$$\mathrm{E}\left(Y_1 - Y_0 \middle| D = 1, \mathbf{X} \right)$$

- Średni efekt przyczynowy (Average Treatment Effect - ATE)

$$ATE = \mathrm{E}\left(Y_1 - Y_0 \right)$$

- Średni efekt przyczynowy dla grupy eksperymentalnej (Average Treatment Effect on the treated - ATE_1)

$$ATE_1 = \mathrm{E}\left(Y_1 - Y_0 \middle| D = 1 \right)$$

Efekt przyczynowy c.d.

- Obserwujemy

$$Y = DY_1 + (1 - D)\,Y_0$$

- Każdej jednostce przypisany jest efekt zaobserwowany i konfaktyczny

	I_0	I_1
$D = 0$	Y_0	Y_1
$D = 1$	Y_0	Y_1

- Nie jest możliwe bezpośrednie zaobserwowanie efektu przyczynowego dla grupy kontrolnej $\mathrm{E}\left(\,Y_0\right|\,D = 1, \mathbf{X})$ (efektu kontrfaktycznego)

Idea PSM

- Sprawdzamy jaka jest różnica między średnią wielkością zmiennej wynikowej dla jednostek bodźcowanych i podobnych niebodźcowanych.
- Im bardziej podobna jest jednostka niebodźcowana do jednostki bodźcowanej tym większą ma wagę w dokonywanym porównaniu
- Oszacowanie efekt przyczynowy dla jednostki i (Treatment Effect)

$$Y_{1i} - \sum_{j \in I_0} W_{N_0,N_1}(i,j)\, Y_{0j}$$

- Oszacowanie średniego efektu przyczynowego (ATE)

$$\sum_{i \in I_1} w_{N_0,N_1(i)} \left[Y_{1i} - \sum_{j \in I_0} W_{N_0,N_1}(i,j)\, Y_{0j} \right]$$

- Różne metody PSM są oparte na różnym doborze funkcji wag $w_{N_0,N_1(i)}$ i $W_{N_0,N_1}(i,j)$.

Statystyczna niezależność

- Definicja

$$\Pr\left(D = 1 | Y_0, Y_1, \mathbf{X}\right) = \Pr\left(D = 1 | \mathbf{X}\right) \tag{1}$$

- Selekcja do próby następuje w oparciu jedynie o czynniki <u>obserwowalne</u>
- Zakładamy, że niewystępuje samoselekcja.
- Wykluczona selekcja w oparciu o czynniki nieobserwowalne wpływające na wynik bodźcowania - tak jak w modelu Heckmana
- Inna notacja

$$\left(Y_0, Y_1\right) \perp D | \mathbf{X}$$

- Stable Unit Treatment Assumption SUTVA - wynik dla poszczególnych jednostek nie zależy od wyników pozostałych jednostek

Silna statystyczna niezależność

- Jeśli dodatkowo prawdą założenie o wspólnej dziedzinie (common support):
$$0 < \Pr(\mathbf{X}) < 1 \tag{2}$$
- Jeśli warunki (1) (1), to mówimy o silnej statystycznej niezależności.
- Warunek (2) jest warunkiem stosowania PSM.
- Jeśli $\Pr(\mathbf{X}) = 0$ lub $\Pr(\mathbf{X}) = 1$, to jednostka jest nigdy bądź zawsze bodźcowana i dobranie dla niej odpowiedników w próbie kontrolnej jest niemożliwe.
- W przypadku, gdy niespełniony jest warunek (2) musimy ograniczyć się do podpopulacji, w której jest on spełniony
- Zdefiniujmy
$$S_0 = Supp(\mathbf{X}|\, D = 0)$$
$$S_1 = Supp(\mathbf{X}|\, D = 1)$$
- Wspólna dziedzina
$$S = S_0 \cap S_1$$

Dekompozycja obciążenia oszacowania w badaniach ewaluacyjnych

- Różnica średnich wyników w grupie eksperymentalnej i kontrolnej

$$\mathrm{E}\left(Y_0 \middle| D = 1, X \in S_1\right) - \mathrm{E}\left(Y_0 \middle| D = 0, X \in S_0\right) = B_1 + B_2 + B_3$$

- Obciążenie wynikające z niepokrywających się dziedzin

$$\begin{aligned} B_1 = & \mathrm{E}\left[\mathrm{E}\left(Y_0 \middle| D = 1, X \in S_1 \setminus (S_0 \cap S_1)\right) \middle| D = 1\right] \\ & - \mathrm{E}\left[\mathrm{E}\left(Y_0 \middle| D = 0, X \in S_0 \setminus (S_0 \cap S_1)\right) \middle| D = 0\right] \end{aligned}$$

- Obciążenie wynikające z odmienności rozkładów w grupie eksperymentalnej i kontrolnej

$$\begin{aligned} B_2 = & \mathrm{E}\left[\mathrm{E}\left(Y_0 \middle| D = 0, X \in S_0 \cap S_1\right) \middle| D = 1\right] \\ & - \mathrm{E}\left[\mathrm{E}\left(Y_0 \middle| D = 0, X \in S_0 \cap S_1\right) \middle| D = 0\right] \end{aligned}$$

Dekompozycja obciążenia oszacowania w badaniach ewaluacyjnych c.d.

- Obciążenie wynikające z wynikające z różnic rozkładów charakterystyk nieobserwowalnych w grupie eksperymentalnej i kontrolnej

$$B_3 = \mathrm{E}\left[\mathrm{E}\left(Y_0 | D = 1\right) - \mathrm{E}\left(Y_0 | D = 0\right) | D = 1, X \in S_0 \cap S_1\right]$$

- Zastosowanie PSM umożliwia skorygowanie obciążenia B_2 oraz określenie wspólnej dziedziny, dla której nie występuje obciążenie B_1

Implikacje silnej statystycznej niezależności c.d.

- Z silnej niezależności wynika, że średni efekt kontrfaktyczny można policzyć w następujący sposób

$$\begin{aligned} \mathrm{E}\left(Y_0 \mid D=1\right) &= \mathrm{E}\left[\mathrm{E}\left(Y_0 \mid D=1,\mathbf{X}\right) \mid D=1\right] \\ &= \mathrm{E}\left[\mathrm{E}\left(Y_0 \mid D=0,\mathbf{X}\right) \mid D=1\right] \end{aligned}$$

- Załóżmy, że udało nam się uzyskać parametryczne bądź nieparametryczne oszacowania $\widehat{r}_0\left(\mathbf{X}_i\right) = \mathrm{E}\left(Y_0 \mid D=0,\mathbf{X}_i\right)$ i $\widehat{r}_1\left(\mathbf{X}_i\right) = \mathrm{E}\left(Y_1 \mid D=1,\mathbf{X}_i\right)$
- Oszacowanie ATE_1 możemy uzyskać jako średnią

$$\frac{1}{N_1}\sum_{i\in I_1}\left(\widehat{r}_1\left(\mathbf{X}_i\right) - \widehat{r}_0\left(\mathbf{X}_i\right)\right)$$

- Do przeprowadzenia tego przekształcenia wystarczy nam założenie, że

$$Y_0 \perp D \mid \mathbf{X}$$

- Do oszacowania ATE_1 wystarczy jeszcze słabsze założenie niezależności średnich

Implikacje silnej statystycznej niezależności

- Silna statystyczna niezależność implikuje, że (Resenbaum, Rubin 1983)

$$(Y_0, Y_1) \perp D | \Pr(\mathbf{X})$$

- Dowód

$$\begin{aligned}
E[D | Y_0, Y_1, \Pr(\mathbf{X})] &= E\{E[D | Y_0, Y_1, \mathbf{X}] | \Pr(\mathbf{X})\} \\
&= E[E[D | Y_0, Y_1, \mathbf{X}] | \Pr(\mathbf{X})] \\
&= E[\Pr(D = 1 | Y_0, Y_1, \mathbf{X}) | \Pr(\mathbf{X})] \\
&= E\{\Pr(D = 1 | \mathbf{X}) | \Pr(\mathbf{X})\} \\
&= E\{\Pr(\mathbf{X}) | \Pr(\mathbf{X})\} = \Pr(\mathbf{X}) \\
&= E[D | \Pr(\mathbf{X})]
\end{aligned}$$

Implikacje silnej statystycznej niezależności c.d.

- Wniosek Resenbauma i Rubina oznacza, że niezależność wyniku od przynależności do grupy kontrolnej zachodzi nie tylko warunkowo względem cech ale także warunkowo względem $\Pr(\mathbf{X})$
- Zauważmy, że jeśli mamy obserwacje z grupy eksperymentalnej i kontrolnej o identycznym $\Pr(\mathbf{X})$, to ATE_1 można oszacować jako:

$$\mathrm{E}\left[\left. Y_1 \right| D = 1, \Pr(\mathbf{X})\right] - \mathrm{E}\left[\left. Y_0 \right| D = 0, \Pr(\mathbf{X})\right] = \mathrm{E}\left[\left. Y_1 - Y_0 \right| \Pr(\mathbf{X})\right]$$

- Dobierając podobne jednostki nie musimy uwzględniać podobieństwa wszystkich cech a jedynie zbliżoną wartość $\Pr(\mathbf{X})$
- Procedura PSM korzysta z tej fundamentalnej własności:

W PSM grupa kontrolna dobierana jest z puli kontrolnej (zbioru obserwacji nie objętych działaniem bodźca) na podstawie podobieństwa oszacowanych wartości propensity score $\Pr(\mathbf{X})$

PSM - szacowanie propensity score

- Metody parametryczne (probit, logit)
- Metody nieparametryczne
- Kluczowym problemem przy szacowaniu $\Pr(\mathbf{X})$ jest zwrócenie uwagi na to, czy spełnione jest założenie o wspólnej dziedzinie
    - zakres wartości dopasowanych prawdopodobieństw powinien być zbliżony
    - nie powinny występować predyktory doskonale przewidujące wynik
- Z tego powodu w literaturze pojawia się pogląd, że model dla $\Pr(\mathbf{X})$ nie powinien być "zbyt dobry"

PSM - dobieranie obserwacji do grupy kontrolnej

- Dobieranie bez zwracania lub ze zwracaniem
- Liczba jednostek z grupy kontrolnej przypadająca na jednostkę bodźcowaną: 1 do 1, 1 do n
- Algorytm dobierania:
  - najbliższy sąsiad (nearest neighbor)
  - najbliższy sąsiad z limitem (nearest neighbor with caliper)
  - metoda z promieniem (radius matching)
  - metoda jądrowa

$$w_{ij} = \frac{K\left(\frac{Pr(X_i)-Pr(X_j)}{h}\right)}{\sum_{j \in I_0}\left(\frac{Pr(X_i)-Pr(X_j)}{h}\right)}$$

przy czym h jest szerokością okna

- Dobór algorytmu dobierania odpowiedników zależy głównie od wielkości próby

PSM - ocena jakości łączenia

- W idealnym przypadku grupa kontrolna uzyskana na podstawie PSM powinna być duża i mieć identyczny rozkład charakterystyk obserwowalnych jak grupa eksperymentalna
- Przy doborze algorytmu doboru obserwacji powinniśmy brać pod uwagę dwa elementy:
    - obciążenie oszacowania efektu przyczynowego: tym mniejsze im podobniejsze obserwacje w grupie kontrolnej
    - wariancję oszacowania efektu przyczynowego: tym mniejsza im większa grupa kontrolna
- Do zbadania jakości łączenia można porównać średnie wartości charakterystyk w grupie eksperymentalnej i grupie kontrolnej
- Można procedurę tę sformalizować przeprowadzając testy równości średnich

PSM - zalety i ograniczenia

- Zalety
  - W przypadku stosowania PSM nie musimy przyjmować parametrycznych postaci zależności między efektem przyczynowym a charakterystykami jednostek - zależność $\mathrm{E}\left(Y_1 \middle| D = 1, \mathbf{X} \right)$ i $\mathrm{E}\left(Y_0 \middle| D = 1, \mathbf{X} \right)$ może mieć całkiem ogólną formę
  - eliminuje najważniejsze źródła obciążenia
- Ograniczenia
  - muszą być spełnione założenia silnej statystycznej niezależność
    - brak samoselekcji
    - brak selekcji w oparci o cechy nieobserwowalne
  - duże wymagania w odniesieniu do wielkości zbioru danych - aby dobrać podobne obserwacje do grupy kontrolnej musimy mieć dużą pulę kontrolną