Evaluación de impacto

Gabriel V. Montes-Rojas

Causalidad vs. casualidad

- Supongamos que nuestro interés está en el análisis de relaciones causales, “si hago esto-entonces pasa esto”.
- El impacto causal de interés se puede dar por $D_i = \{0, 1\}$. Ejemplos: si la familia recibe AUH, si el trabajador recibe capacitación, etc.
- El resultado (outcome) de interés es

$$Y_i = \begin{cases} Y_{1i} & si \;\; D_i = 1 \\ Y_{0i} & si \;\; D_i = 0 \end{cases}$$

- Y_{1i} es el valor del resultado tomando en cuenta el **efecto** del tratamiento.
- Y_{0i} es el valor del resultado sin el **efecto** del tratamiento.
- Solo observamos $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$

Average treatment effect

- Nos interesa estimar
  - $\delta_{ATE} = E[Y_{1i} - Y_{0i}]$, average treatment effect;
  - $\delta_{ATT} = E[Y_{1i} - Y_{0i} | D_i = 1]$, average treatment effect on the treated
- El principal problema es que solo vemos a los individuos en uno solo de los estados del tratamiento. Entonces si comparamos aquellos con $D = 1$ con aquellos con $D = 0$ corremos el riesgo de no solo evaluar el tratamiento, sino diferencias subyacentes entre los dos grupos. Tenemos $E(Y_i|D_i = 1) = E(Y_{1i}|D_i = 1)$ y $E(Y_i|D_i = 0) = E(Y_{0i}|D_i = 1)$.
- Supongamos un estimador naïve

$$\underbrace{E(Y_i|D_i = 1) - E(Y_i|D_i = 0)}_{Total} = \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{ATT} + \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{Sesgo\ de\ selecc.}$$

Aleatoriedad e independencia

- Supongamos que D, quienes reciben el tratamiento, es aleatorio y es independiente de los potenciales resultados (random assignment). Esto hace que $E(Y_{1i}|D_i = 1) = E(Y_{1i}|D_i = 0)$ y $E(Y_{0i}|D_i = 1) = E(Y_{0i}|D_i = 0)$.
- La aleatoriedad genera independencia: $\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i$: La asignación a un estado de tratamiento es independiente de los resultados potenciales.
- Entonces,

$$\begin{aligned} E(Y_i|D_i = 1) - E(Y_i|D_i = 0) &= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) \\ &= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1) \\ &= E(Y_{1i} - Y_{0i}|D_i = 1) \end{aligned}$$

Conditional independence assumption (CIA)

- La CIA (conditional independence assumption, no la Central Intelligence Agency) [también se define como *unconfoundness*] establece que se puede lograr independencia después de condicionar en X:

$$\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i | X_i$$

- Entonces,

$$\begin{aligned} E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i) &= E(Y_{1i}|D_i = 1, X_i) - E(Y_{0i}|D_i = 0, X_i) \\ &= E(Y_{1i}|D_i = 1, X_i) - E(Y_{0i}|D_i = 1, X_i) \end{aligned}$$

Modelos de regresión

- Tomemos ahora un modelo de efectos causales constantes (constant-effect causal model)

$$Y_i = \alpha + \rho D_i + \eta_i$$

donde $\alpha = E(Y_{0i})$, $\rho = E(Y_{1i} - Y_{0i})$, $\eta_i = Y_{0i} - E(Y_{0i})$. Así, $E[Y_i|D_i = 1] = \alpha + \rho + E[\eta_i|D_i = 1]$ y $E[Y_i|D_i = 0] = \alpha + E[\eta_i|D_i = 0]$.

- Por otra parte, descompongamos η_i en dos partes tal que $\eta_i = X_i'\gamma + \nu_i$. Por contrucción, $E(\eta_i|X_i) = X_i'\gamma$ y $E(\nu_i|X_i) = 0$.
- Por otro lado, por la CIA,

$$E(Y_i|X_i, D_i) = E(Y_i|X_i) = \alpha + \rho D_i + E(\eta_i|X_i) = \alpha + \rho D_i + X_i'\gamma + \nu_i,$$

con ν_i independiente de X_i y D_i.

Matching

- Tomemos,

$$\delta_{ATT} = E(Y_{1i} - Y_{0i}|D_i = 1) = E_X \left\{ E(Y_{1i}|X_i, D_i = 1) - E(Y_{0i}|X_i, D_i = 1)|D_i = 1 \right\},$$

donde estamos aplicando la ley de esperanzas iteradas.

- $E(Y_{0i}|X_i, D_i = 1)$ es no observado y contrafáctico. Sin embargo, bajo la propiedad de la CIA, $E(Y_{0i}|X_i, D_i = 1) = E(Y_{0i}|X_i, D_i = 0)$.
- Entonces,

$$\delta_{ATT} = E_X \left\{ E(Y_{1i}|X_i, D_i = 1) - E(Y_{0i}|X_i, D_i = 0)|D_i = 1 \right\} = E_X \left\{ \delta_X|D_i = 1 \right\},$$

donde $\delta_X = E(Y_{1i}|X_i, D_i = 1) - E(Y_{0i}|X_i, D_i = 0)$.

- Supongamos que el dominio de X es discreto ($\mathcal{X}$), entonces

$$\delta_{ATT} = \sum_{X \in \mathcal{X}} \delta_X P[X_i = x|D_i = 1]$$

$$\delta_{ATE} = \sum_{X \in \mathcal{X}} \delta_X P[X_i = x]$$

Nearest neighbor matching (NNM)

- The NNM method of treatment-effect estimation imputes the missing potential outcome for each individual by using an average of the outcomes of similar subjects that receive the other treatment level. Similarity between subjects is based on a weighted function of the covariates for each observation.
- It determines the nearest by using a weighted function of the covariates for each observation. By default, the Mahalanobis distance is used, in which the weights are based on the inverse of the covariates variance–covariance matrix. You may also request exact matching for categorical covariates. For example, you may want to force all matches to be of the same gender or race.

Nearest neighbor matching (NNM)
STATA

Main reference: `http://www.stata.com/manuals13/te.pdf`

- In `STATA` use
  `teffects nnmatch`
- Example of use: `teffects nnmatch (`Y X`) (`D`)`

  where
  - Y: outcome variable of interest.
  - X: set of covariates for the match.
  - D: treatment indicator.
- Optional to specify number of neighbors to match, default is 1: `teffects nnmatch (`Y X`) (`D`), nneighbor(#)`
- Optional to request ATE or ATT, default is ATE: `teffects nnmatch (`Y X`) (`D`), ate OR atet`
- Example of use: `teffects nnmatch (`Y X_1`) (`D`), ematch(`X_2`)`

  where
  - X_1: set of covariates for the match, no exact matching.
  - X_2: set of covariates for the match, exact matching (should be discrete valued covariates).

Nearest neighbor matching (NNM)
STATA

- We will illustrate the use of `teffects nnmatch` by using data from a study of the effect of a mother's smoking status during pregnancy (`mbsmoke`) on infant birthweight (`bweight`) as reported by Cattaneo (2010, Journal of Econometrics 155: 138-154).
- dataset also contains information about each mother's age (`mage`), education level (`medu`), marital status (`mmarried`), whether the first prenatal exam occurred in the first trimester (`prenatal1`), whether this baby was the mother's first birth (`fbaby`), and the father's age (`fage`).

Nearest neighbor matching (NNM)
STATA

```
use http://www.stata-press.com/data/r13/cattaneo2
teffects nnmatch (bweight mage prenatal1 mmarried fbaby) (mbsmoke)
* Requesting 4 matches
teffects nnmatch (bweight mage prenatal1 mmarried fbaby) (mbsmoke), nn(4)

* Exact match on prenatal1 mmarried fbaby
teffects nnmatch (bweight mage) (mbsmoke), ematch(prenatal1 mmarried
fbaby) metric(euclidean)
```

Matching

- Los modelos de matching crean el contrafactual para *cada* valor de X. (covariate matching)
- El problema es cuando tenemos muchos valores de X, que a su vez depende de la cardinalidad de $\mathcal{X}$.
- Una solución práctica es implementar la condicionalidad en el propensity score (Rosenbaum y Rubin, 1983, propensity score matching):

Propensity score: $p(X_i) \equiv P[D_i = 1|X_i] = E[D_i|X_i]$.

Matching

Teorema del propensity score: Si la propiedad CIA es válida (ej. $\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i|X_i$) entonces $\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i|p(X_i)$.

Prueba: La prueba está basada en Rosenbaum y Rubin (1983). Se necesita probar que $P[D_i = 1|Y_{ji}, p(X_i)]$ no depende de Y_{ji} para $j = 0, 1$:

$$
\begin{aligned}
P[D_i = 1|Y_{ji}, p(X_i)] &= E[D_i|Y_{ji}, p(X_i)] = E_X\{E[D_i|Y_{ji}, p(X_i), X_i]|Y_{ji}, p(X_i)\} \\
&= E_X\{E[D_i|Y_{ji}, X_i]|Y_{ji}, p(X_i)\} = E_X\{E[D_i|, X_i]|Y_{ji}, p(X_i)\} \text{ (CIA)} \\
&= E_X\{p(X_i)|Y_{ji}, p(X_i)\} = p(X_i)
\end{aligned}
$$

Teorema del balanceo: $X_i \perp\!\!\!\perp D_i|p(X_i)$.

Prueba: La prueba está en Rosenbaum y Rubin (1983).

Propensity score matching (NNM)
STATA

- En `STATA` usar `teffects psmatch`
- Ejemplo: `teffects psmatch` (Y) (D X), `nneighbor(#) ate 0 att`

Propensity score matching (NNM)
STATA

```
use http://www.stata-press.com/data/r13/cattaneo2
teffects psmatch (bweight) (mbsmoke mmarried mage fbaby medu)
```

Common support

(Common support) For all $x \in domain(X)$ *we have that*

$$0 < \underline{p} \leq p(x) \leq \bar{p} < 1$$

- This condition ensures that treatment observations have comparison observations “nearby” in the propensity score distribution to match with.
- In `STATA` type `teffects overlap` after `teffects psmatch` (you need to specify the option `, gen(ps)`

Main identification result using the propensity score (inverse probability weighting)

Hirano, K., Imbens, G.W. and Ridder, G. (1996) "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161-1189

- Under the CIA and common support assumptions, $ATE = E[Y_1 - Y_0]$ can be identified by $ATE = E\left[\frac{Y.D}{p(X)} - \frac{Y.(1-D)}{1-p(X)}\right]$.

  Proof: (p.1169)

  $E\left[\frac{Y.D}{p(X)}\right] = E_X\left\{E\left[Y.D|X\right]\frac{1}{p(X)}\right\} = E_X\left\{E[Y_1|X]E[D|X]\frac{1}{p(X)}\right\} = E_X\left\{E[Y_1|X]p(X)\frac{1}{p(X)}\right\} = E[Y_1]$

  $E\left[\frac{Y.(1-D)}{1-p(X)}\right] = E_X\left\{E\left[Y.(1-D)|X\right]\frac{1}{1-p(X)}\right\} = E_X\left\{E[Y_0|X]E[(1-D)|X]\frac{1}{1-p(X)}\right\} =$
  $E_X\left\{E[Y_0|X](1-p(X))\frac{1}{p(X)}\right\} = E[Y_0]$.

- Under the CIA and common support assumptions, $ATT = E[Y_1 - Y_0|D = 1]$ can be identified by $E[p(X)].ATT = E\left[p(X)\left(\frac{Y.D}{p(X)} - \frac{Y.(1-D)}{1-p(X)}\right)\right]$. (Note: $E[p(X)] = Pr[D = 1]$.)

  Proof:

  $E\left[p(X)\frac{Y.D}{p(X)}\right] = E_X\left\{E\left[Y.D|X\right]\right\} = E_X\left\{E[Y_1|D = 1, X]p(X)\right\}$, using
  $E[Y.D|X] = E[Y_1.D|X] = E[Y_1|D = 1, X]p(X)$

  $E\left[p(X)\frac{Y.(1-D)}{1-p(X)}\right] = E_X\left\{p(X)E\left[Y_0|D = 0, X\right]\frac{1-p(X)}{1-p(X)}\right\} = E_X\left\{E\left[Y_0|D = 1, X\right]p(X)\right\}$

  Then, $E\left[p(X)\frac{Y.D}{p(X)} - p(X)\frac{Y.(1-D)}{1-p(X)}\right] = E_X\left\{p(X)E[Y_1 - Y_0|D = 1, X]\right\}$. Finally note that
  $E[Y_1 - Y_0|D = 1] = E_X\{E[(Y_1 - Y_0)|D = 1, X]|D = 1\} = E_X\{E[(Y_1 - Y_0)|X]|D = 1\} =$
  $E_X[ATE(X)|D = 1] = \int ATE(x)dF(x|D = 1) = \frac{\int ATE(x)p(x)dF(x)}{\int p(x)dF(x)}$ (p.1174).

QTE

Firpo, S. (2007) "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259-276.

- The QTE is written as $\Delta_\tau = q_{1\tau} - q_{0\tau}$, where $q_{j\tau} = inf_q Pr[Y(j) \leq q] \geq \tau$, $j = 0, 1$.
- The QTT is written as $\Delta_\tau = q_{1\tau|D=1} - q_{0\tau|D=1}$, where $q_{j\tau|D=1} = inf_q Pr[Y(j) \leq q|D = 1] \geq \tau$, $j = 0, 1$.
- Identification: Under CIA, common support, and uniqueness of quantiles: $\tau = E[\frac{D}{p(X)} 1\{Y \leq q_{1\tau}\}]$, $\tau = E[\frac{1-D}{1-p(X)} 1\{Y \leq q_{0\tau}\}]$, $\tau = E[\frac{D}{E[p(X)]} 1\{Y \leq q_{1\tau|D=1}\}]$, $\tau = E[\frac{(1-D)p(X)}{E[p(X)](1-p(X))} 1\{Y \leq q_{0\tau|D=1}\}]$.
- Estimation:

$$\hat{q}_{1\tau} = argmin_q \sum_{i=1}^{N} D_i / \hat{p}(X_i) \rho_\tau(Y_i - q)$$

$$\hat{q}_{0\tau} = argmin_q \sum_{i=1}^{N} (1 - D_i)/(1 - \hat{p}(X)) \rho_\tau(Y_i - q)$$

$$\hat{q}_{1\tau|D=1} = argmin_q \sum_{i=1}^{N} \frac{D_i}{\sum_{i=1}^{N} D_i} \rho_\tau(Y_i - q)$$

$$\hat{q}_{0\tau|D=1} = argmin_q \sum_{i=1}^{N} \frac{(1 - D_i)\hat{p}(X_i)}{(1 - \hat{p}(X_i)) \sum_{i=1}^{N} D_i} \rho_\tau(Y_i - q)$$

Matching y regresión

- Hay una comparación interesante que sale del libro de Angrist y Pischke.
- $\delta_{ATT}^{match} = \frac{\sum_{X \in \mathcal{X}} \delta_X P[D_i=1|X_i=x] P[X_i=x]}{\sum_{X \in \mathcal{X}} P[D_i=1|X_i=x] P[X_i=x]}$ $\Rightarrow$promedio ponderado por la probabilidad de aparición.
- $\delta_{ATT}^{reg} = \frac{\sum_{X \in \mathcal{X}} \delta_X P[D_i=1|X_i=x](1-P[D_i=1|X_i=x]) P[X_i=x]}{\sum_{X \in \mathcal{X}} P[D_i=1|X_i=x](1-P[D_i=1|X_i=x]) P[X_i=x]}$ $\Rightarrow$promedio ponderado por la varianza de $E[D_i|X_i]$.