Doubly robust estimators

16.06.21 11:12

- a.k.a. augmented inverse probability of treatment weighting (IPTW)

Background: IPTW estimation of $E(Y^1)$:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{A_i Y_i}{\pi_i(X_i)} = E(Y^1)$$

indicates treatment

propensity score

unbiased estimator iff the propensity score is $\boxed{\text{correctly specified}}$
↳ meaning that $P(A=1 \mid X_i) \equiv \pi(X_i) \quad (\forall X_i)$

Background: Regression-based estimation of $E(Y^1)$:

outcome model $m_1(X) = E[Y \mid A=1, X]$

$$\frac{1}{n}\sum_{i=1}^{n}\{\underbrace{A_i \cdot Y_i}_{} + \underbrace{(1-A_i)\, m_1(X_i)}_{}\}$$

if outcome model $m_1(X)$ is correctly specified, then this estimator is unbiased

for subjects with $A=1$, use observed outcome Y

for subjects with $A \neq 1$, use predicted value of Y given their covariates X if their A would have been 1 instead

1. learn $m_1(X)$ from subjects with $A=1$.
2. predict Y with $m_1(X)$ for subjects with $A \neq 1$.

Doubly Robust Estimators:

- a DR estimator is an estimator that is unbiased if either the propensity score model OR the outcome regression model are correctly specified. (only one model needs to be correctly specified)

Example of a DR estimator:

- $m_1(X)$ is not correctly specified if $m_1(X) \neq E(Y \mid A=1, X)$

Scenario I: $m_1(X)$ incorrectly specified, $\pi(X)$ correctly specified:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i)} - \underbrace{\frac{A_i - \pi(X_i)}{\pi(X_i)}}_{}\, m_1(X_i)\right\}$$

expectation of A_i is equal to $\pi(X_i)$, the propensity score. So in expectation $A_i - \pi(X_i)$ will cancel out.

=> thus the RH term must have expectation of 0.

-> then the rest of the estimator left is a valid estimator b/c the prop. score model is correctly specified: $\boxed{\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i)}\right\}}$ the standard IPTW estimator of $E(Y^1)$.

Scenario II: $m_1(X)$ correctly specified, $\pi(X)$ incorrectly specified: (i.e. $E(Y \mid X) = m_1(X)$) (i.e. $E(A_i) \neq \pi(X_i)$)

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i)} - \frac{A_i - \pi(X_i)}{\pi(X_i)}\, m_1(X_i)\right\} =$$

$= +1$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{m_1(X_i) + \frac{A_i(Y_i - m_1(X_i))}{\pi(X_i)}\right\} =$$

this part goes to 0, b/c $m_1(X_i)$ is correctly specified (i.e. $E(Y \mid X) = m_1(X)$).

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\{m_1(X_i)\}}_{\text{goes to } E(Y^1).}$$

Average of $E(Y \mid A=1, X)$ over the dist. of X → thus X is being out-marginalized giving us $E(Y \mid A=1) = E(Y^1)$.

Doubly robust estimators: (allows us to specify and use 2 models w/ guaranteed unbiasedness if 1 model is incorrectly specified)

- DR estimator a.k.a. Augmented IPTW (AIPTW) estimator.
  ↳ combines vanilla IPTW w/ an outcome model $m_1(X)$

=> in general: AIPTW estimators should be more efficient than regular IPTW estimators

meaning the estimator has a smaller variance associated w/ its estimations.