DS-GA 3001.009 Applied Statistics: Homework #7 Solutions

Due on Thursday, November 30, 2023

Please hand in your homework via Gradescope (entry code: RKXJN2) before 11:59 PM.

1. In the estimation of ATE, in class we modeled the mean outcomes for each group:

$$\mu_0(X) = \mathbb{E}[Y \mid X, W = 0], \quad \mu_1(X) = \mathbb{E}[Y \mid X, W = 1].$$

Another modeling is to model a single mean outcome $m(X) = \mathbb{E}[Y \mid X]$ and consider the following estimating function

$$f_{(m,e,\tau)}(W, X, Y) = (Y - m(X) - (W - e(X))\tau)(W - e(X)),$$

where $e(X) = \mathbb{P}(W = 1 \mid X)$ is the propensity score.

(a) Find the expression of $m(X)$ in terms of $(\mu_0(X), \mu_1(X), e(X))$.

(b) Assuming that $\mu_1(x) = \mu_0(x) + \tau$ for all x, show that $f_{(m,e,\tau)}(W, X, Y)$ is a valid estimating function, i.e.

$$\mathbb{E}[f_{(m,e,\tau)}(W, X, Y)] = 0.$$

(c) Show that $f_{(m,e,\tau)}(W, X, Y)$ is Neyman orthogonal with respect to (m, e), i.e.

$$\begin{aligned} \mathbb{E}[\nabla_m f_{(m,e,\tau)}(W, X, Y)] &= 0, \\ \mathbb{E}[\nabla_e f_{(m,e,\tau)}(W, X, Y)] &= 0. \end{aligned}$$

(d) Show that $f_{(m,e,\tau)}(W, X, Y)$ is *not* doubly robust, by arguing that in general

$$\mathbb{E}[f_{(m,\widehat{e},\tau)}(W, X, Y)] \neq 0.$$

Solution:

(a) By the law of total probability,

$$\begin{aligned} m(X) &= \mathbb{E}[Y \mid X] \\ &= \mathbb{P}(W = 1 \mid X) \cdot \mathbb{E}[Y \mid X, W = 1] + \mathbb{P}(W = 0 \mid X) \cdot \mathbb{E}[Y \mid X, W = 0] \\ &= e(X)\mu_1(X) + (1 - e(X))\mu_0(X). \end{aligned}$$

(b) It holds that

$$\begin{aligned} \mathbb{E}[f_{(m,e,\tau)}(W, X, Y)] &= \mathbb{E}\left\{\mathbb{E}\left[(Y - m(X) - (W - e(X))\tau)(W - e(X)) \mid X\right]\right\} \\ &= \mathbb{E}\left\{e(X) \cdot \mathbb{E}\left[(Y(1) - m(X) - (1 - e(X))\tau)(1 - e(X)) \mid X, W = 1\right]\right\} \\ &\quad - \mathbb{E}\left\{(1 - e(X)) \cdot \mathbb{E}\left[(Y(0) - m(X) + e(X)\tau)e(X) \mid X, W = 0\right]\right\} \\ &= \mathbb{E}\left\{e(X)(1 - e(X)) \cdot (\mu_1(X) - m(X) - (1 - e(X))\tau)\right\} \\ &\quad - \mathbb{E}\left\{e(X)(1 - e(X)) \cdot (\mu_0(X) - m(X) + e(X)\tau)\right\} \\ &= \mathbb{E}\left\{e(X)(1 - e(X)) \cdot (\mu_1(X) - \mu_0(X) - \tau)\right\} = 0. \end{aligned}$$

(c) The first identity follows from

$$\mathbb{E}[\nabla_m f_{(m,e,\tau)}(W,X,Y)] = -\mathbb{E}[W - e(X) \mid X] = 0.$$

The second identity follows from

$$\begin{aligned}
\mathbb{E}[\nabla_e f_{(m,e,\tau)}(W,X,Y)] &= -\mathbb{E}[Y - m(X) \mid X] + 2\mathbb{E}[(W - e(X))\tau \mid X] \\
&= -(\mathbb{E}[Y \mid X] - m(X)) + 2\tau(\mathbb{E}[W \mid X] - e(X)) \\
&= 0.
\end{aligned}$$

(d) Repeating the computation in (b) gives that

$$\begin{aligned}
\mathbb{E}[f_{(m,\widehat{e},\tau)}(W,X,Y)] &= \mathbb{E}\left\{e(X)(1-\widehat{e}(X)) \cdot (\mu_1(X) - m(X) - (1-\widehat{e}(X))\tau)\right\} \\
&\quad - \mathbb{E}\left\{\widehat{e}(X)(1-e(X)) \cdot (\mu_0(X) - m(X) + \widehat{e}(X)\tau)\right\} \\
&= \mathbb{E}\left\{(e(X) - \widehat{e}(X)) \cdot (\mu_0(X) - m(X) + \widehat{e}(X)\tau)\right\} \\
&= -\tau\mathbb{E}\left\{(e(X) - \widehat{e}(X))^2\right\} \neq 0,
\end{aligned}$$

where the last step follows from

$$m(X) = e(X)(\mu_0(X) + \tau) + (1 - e(X))\mu_0(X) = \mu_0(X) + e(X)\tau,$$

thanks to (a) and the assumption in (b).

2. Consider the same setting for the AIPW estimator in class, but now we aim to estimate the average treatment effect on the treated (ATTE): $\tau^{\text{ATTE}} = \mathbb{E}[\mu_1(X) - \mu_0(X) \mid W = 1]$. Consider the following estimating function:

$$f_{(\mu_0,e,\tau^{\text{ATTE}})}(W,X,Y) = \frac{W(Y - \mu_0(X) - \tau^{\text{ATTE}})}{m} - \frac{e(X)(1-W)(Y - \mu_0(X))}{m(1-e(X))},$$

where $e(x) = \mathbb{P}(W = 1 \mid X = x)$ is the propensity score, and $m = \mathbb{P}(W = 1)$ is the marginal probability of treatment. For simplicity we assume that m is known.

(a) Let $p(x)$ be the pmf of $X = x$. Using the Bayes rule, show that

$$\mathbb{P}(X = x \mid W = 1) = \frac{p(x)e(x)}{m}.$$

(b) Use (a) to prove the following identity:

$$\tau^{\text{ATTE}} = \mathbb{E}\left[\frac{e(X)}{m}(\mu_1(X) - \mu_0(X))\right].$$

(c) Show that $f_{(\mu_0,e,\tau^{\text{ATTE}})}(W,X,Y)$ is a valid estimating function, i.e.

$$\mathbb{E}[f_{(\mu_0,e,\tau^{\text{ATTE}})}(W,X,Y)] = 0.$$

(d) (*Bonus 5 points*) Show that $f_{(\mu_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)$ is doubly robust, i.e. for any $(\widehat{\mu}_0(x),\widehat{e}(x))$,

$$\mathbb{E}[f_{(\widehat{\mu}_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)] = 0,$$
$$\mathbb{E}[f_{(\mu_0,\widehat{e},\tau^{\mathrm{ATTE}})}(W,X,Y)] = 0.$$

Solution:

(a) The Bayes rule gives

$$\mathbb{P}(X=x\mid W=1) = \frac{p(x)\cdot\mathbb{P}(W=1\mid X=x)}{\mathbb{P}(W=1)} = \frac{p(x)e(x)}{m}.$$

(b) By (a), we have

$$\begin{aligned}
\tau^{\mathrm{ATTE}} &= \mathbb{E}[\mu_1(X)-\mu_0(X)\mid W=1] \\
&= \sum_x (\mu_1(x)-\mu_0(x))\mathbb{P}(X=x\mid W=1) \\
&= \sum_x \frac{e(x)(\mu_1(x)-\mu_0(x))}{m}p(x) \\
&= \mathbb{E}\left[\frac{e(X)}{m}(\mu_1(X)-\mu_0(X))\right].
\end{aligned}$$

(c) It holds that

$$\begin{aligned}
&\mathbb{E}[f_{(\mu_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)] \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{W(Y-\mu_0(X)-\tau^{\mathrm{ATTE}})}{m}\middle|X\right] - \mathbb{E}\left[\frac{e(X)(1-W)(Y-\mu_0(X))}{m(1-e(X))}\middle|X\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{e(X)(Y(1)-\mu_0(X)-\tau^{\mathrm{ATTE}})}{m}\middle|X,W=1\right] - \mathbb{E}\left[\frac{e(X)(Y(0)-\mu_0(X))}{m}\middle|X,W=0\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\frac{e(X)(\mu_1(X)-\mu_0(X)-\tau^{\mathrm{ATTE}})}{m}\middle|X,W=1\right]\right\} \\
&= \mathbb{E}\left\{\frac{e(X)(\mu_1(X)-\mu_0(X)-\tau^{\mathrm{ATTE}})}{m}\right\} \\
&= \mathbb{E}\left\{\frac{e(X)(\mu_1(X)-\mu_0(X)}{m}\right\} - \frac{\mathbb{E}[e(X)]}{m}\cdot\tau^{\mathrm{ATTE}} \\
&= \tau^{\mathrm{ATTE}} - \tau^{\mathrm{ATTE}} = 0,
\end{aligned}$$

where the last step is due to (b) and $\mathbb{E}[e(X)] = \mathbb{E}\{\mathbb{P}[W=1\mid X]\} = \mathbb{P}(W=1) = m$.

(d) The expected difference between $f_{(\widehat{\mu}_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)$ and $f_{(\mu_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)$ is

$$\mathbb{E}[f_{(\widehat{\mu}_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y) - f_{(\mu_0,e,\tau^{\mathrm{ATTE}})}(W,X,Y)]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\frac{W(\mu_0(X)-\widehat{\mu}_0(X))}{m} - \frac{e(X)(1-W)(\mu_0(X)-\widehat{\mu}_0(X))}{m(1-e(X))}\right] \\
&= \mathbb{E}\left\{(\mu_0(X)-\widehat{\mu}_0(X)) \cdot \mathbb{E}\left[\frac{W}{m} - \frac{e(X)(1-W)}{m(1-e(X))}\middle| X\right]\right\} \\
&= \mathbb{E}\left\{(\mu_0(X)-\widehat{\mu}_0(X)) \cdot \left(\frac{e(X)}{m} - \frac{e(X)}{m}\right)\right\} = 0,
\end{aligned}
$$

so the first identity follows from (c). Similarly, for the second identity we have

$$
\begin{aligned}
&\mathbb{E}[f_{(\mu_0,\widehat{e},\tau^{\text{ATTE}})}(W,X,Y) - f_{(\mu_0,e,\tau^{\text{ATTE}})}(W,X,Y)] \\
&= \mathbb{E}\left[\left(\frac{\widehat{e}(X)}{1-\widehat{e}(X)} - \frac{e(X)}{1-e(X)}\right)\frac{(1-W)(Y-\mu_0(X))}{m}\right] \\
&= \mathbb{E}\left\{\left(\frac{\widehat{e}(X)}{1-\widehat{e}(X)} - \frac{e(X)}{1-e(X)}\right) \cdot \mathbb{E}\left[\frac{(1-W)(Y-\mu_0(X))}{m}\middle| X\right]\right\} \\
&= \mathbb{E}\left\{\left(\frac{\widehat{e}(X)}{1-\widehat{e}(X)} - \frac{e(X)}{1-e(X)}\right) \cdot \mathbb{E}\left[\frac{(1-e(X))(Y(0)-\mu_0(X))}{m}\middle| X, W=0\right]\right\} \\
&= 0,
\end{aligned}
$$

where the last step is because $\mathbb{E}[Y(0)-\mu_0(X) \mid X, W] = 0$.

3. Coding: we will compare the IPW and AIPW estimators on a synthetic dataset. Based on inline instructions, fill in the missing codes in `https://tinyurl.com/y22fams3`. Be sure to submit a pdf with your codes, outputs, and colab link.

   Solution: see `https://tinyurl.com/2ec3tajd`.