Double ML for Sample Selection. Suppose we are interested in the Average Treatment Effect parameter

$$\theta = \mathrm{E}[Y(1) - Y(0)],$$

where $Y(1)$ and $Y(0)$ are potential outcomes with and without treatment, respectively. The treatment D is randomly assigned given X:

$$(Y(1), Y(0)) \perp D \mid X, \quad (1)$$

and the propensity score is

$$\pi(X) = \Pr(D = 1 \mid X).$$

Furthermore, the outcome Y is partially missing, and the selection indicator $S = 1$ if and only if Y is observed. To sum up, the data vector

$$W = (D, X, S, S \cdot Y),$$

consists of the pre-treatment covariates X, the treatment indicator $D \in \{1, 0\}$, the selection indicator S, and the outcome Y if and only if $S = 1$. The selection is exogenous, that is,

$$S \perp Y \mid D = d, X \quad \forall d \in \{1, 0\}. \quad (2)$$

which implies

$$S(d) \perp Y(d) \mid D = d, X \quad \forall d \in \{1, 0\}.$$

(i) Write down any valid, unconditional moment equation for θ_0. **Note!** A moment function for θ must depend only on data vector W, the parameter θ, and, possibly, other identified functions (e.g., the propensity score $\pi(X)$).

The actual/realized outcome and selection variable are related to its potential counterparts as

$$Y = DY(1) + (1 - D)Y(0), \quad S = DS(1) + (1 - D)S(0).$$

Equation (1) implies a moment equation

$$\theta = \mathrm{E}[Y(1) - Y(0)] = \mathrm{E}[\frac{DY}{\pi(X)} - \frac{(1 - D)Y}{1 - \pi(X)}]$$

if there is no sample selection problem. However, the terms DY or $(1-D)Y$ are not observed if $S=0$, and this moment is infeasible. Our identifying assumptions imply that

$$\mathrm{E}[Y(d) \mid X] = \mathrm{E}[Y(d) \mid D=d, X] = \mathrm{E}[Y(d) \mid D=d, X, S(d)=1].$$

where the first equality follows from (1) and the second equality from (2). Therefore, the mean potential outcome identified by

$$\mathrm{E}[Y(d) \mid X] = \mathrm{E}[Y \mid D=d, X, S=1] = \mathrm{E}[\frac{1\{D=d\}SY}{\Pr(S=1 \mid D=d, X)\Pr(D=d \mid X)} \mid X].$$

Law of Iterated Expectations $\mathrm{E}_X[\mathrm{E}[Y(d) \mid X]] = \mathrm{E}[Y(d)]$ gives

$$\mathrm{E}[Y(d)] = \mathrm{E}[\frac{1\{D=d\}SY}{\Pr(S=1 \mid D=d, X)\Pr(D=d \mid X)}] =: \mathrm{E}[\psi(d)].$$

Thus,

$$\theta = \mathrm{E}[Y(1)-Y(0)] = \mathrm{E}[\psi(1)-\psi(0)].$$

(iii) Write down an orthogonal moment equation for θ_0. Demonstrate the orthogonality property. Define regression function

$$\mu(d, X) = \mathrm{E}[Y \mid S=1, D=d, X=x].$$

Robins and Rotnitzky moment equation is orthogonal with respect to the first-stage nuisance functions $\eta(x) := \{\mu(d,x), \Pr(S=1 \mid D=d, X=x), \pi(x)\}, d \in \{1,0\}$. It takes the form

$$\mathrm{E}[\phi(d,\eta)] = \frac{1\{D=d\}S(Y-\mu(d,X))}{\Pr(S=1 \mid D=d, X)\Pr(D=d \mid X)} + \mu(d,X)$$

and the moment equation for θ is

$$\theta = \mathrm{E}[Y(1)-Y(0)] = \mathrm{E}[\phi(1,\eta_0)-\phi(0,\eta_0)].$$

(iv) Propose a Double ML approach to estimate θ_0. Clearly state your assumptions, describe the method, and provide a formal central limit theorem statement for it.

Step 1. Partition date in two halves J_1 and J_2. On each half, estimate the first-stage nuisance functions using some nonparametric/ML methods. Let the estimate be $\widehat{\eta}_1$ on J_1 and $\widehat{\eta}_2$ on J_2.

Step 2. Compute sample average

$$\widehat{\theta} := N^{-1}(\sum_{i \in J_2}[\phi(1, \widehat{\eta}_1) - \phi(0, \widehat{\eta}_1)] + \sum_{i \in J_1}[\phi(1, \widehat{\eta}_2) - \phi(0, \widehat{\eta}_2)]).$$

invoking the estimate $\widehat{\eta}$ from the partition that does not contain i.

Grading criteria for this question:

- Good or Pass. An attempt to write down moment equation, or general grasp of double ML as expressed by generic ML algorithm, depending on the details/progress.
- Excellent or Good. Some specific progress involving calculations above, depending on the details/progress.