Section: Module 8

Matching / Weighting

Kentaro Nakamura

GOV 2002

November 14th, 2025

Logistics

- Problem Set 8 is due on Next Wednesday
    - We decided to make it due on Wednesday
    - Check Google Calendar about the schedule

- Today's agenda
    - Matching
    - Weighting

Overview: Nonparametric Identification

- This week is about the different estimation strategy.
- We assume the **conditional ignorability** given confounder X_i:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i$$

  - We also need positivity: $0 < \mathbb{P}(T_i = 1 \mid X_i = x) < 1$ for any $X_i = x$

- We can identify average treatment effect nonparametrically:

$$\begin{aligned}
\tau_{\text{ATE}} &= \mathbb{E}[Y_i(1) - Y_i(0)] \\
&= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\
&= \mathbb{E}[\mathbb{E}[Y_i(1) \mid X_i]] - \mathbb{E}[\mathbb{E}[Y_i(0) \mid X_i]] \quad (\because \text{L.I.E}) \\
&= \mathbb{E}[\mathbb{E}[Y_i(1) \mid T_i = 1, X_i]] - \mathbb{E}[\mathbb{E}[Y_i(0) \mid T_i = 0, X_i]] \ (\because \text{Ignorability}) \\
&= \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 1, X_i]] - \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 0, X_i]] \quad (\because \text{Consistency})
\end{aligned}$$

- Equivalently,

$$\tau_{\text{ATE}} = \int \{\mathbb{E}[Y_i \mid T_i = 1, X_i = x] - \mathbb{E}[Y_i \mid T_i = 0, X_i = x]\} dF(x)$$

Overview: Nonparametric Identification

- When estimand is ATT,

$$\begin{aligned}
\tau_{\text{ATT}} &= \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1] \\
&= \mathbb{E}[Y_i(1) \mid T_i = 1] - \mathbb{E}[Y_i(0) \mid T_i = 1] \\
&= \mathbb{E}[Y_i \mid T_i = 1] - \mathbb{E}[Y_i(0) \mid T_i = 1] \quad (\because \text{consistency}) \\
&= \mathbb{E}[Y_i \mid T_i = 1] - \mathbb{E}[\mathbb{E}[Y_i(0) \mid X_i, T_i = 1] \mid T_i = 1] \quad (\because \text{L.I.E}) \\
&= \mathbb{E}[Y_i \mid T_i = 1] - \mathbb{E}[\mathbb{E}[Y_i(0) \mid X_i, T_i = 0] \mid T_i = 1] \ (\because \text{Ignorability}) \\
&= \mathbb{E}[Y_i \mid T_i = 1] - \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 0, X_i] \mid T_i = 1] \quad (\because \text{Consistency})
\end{aligned}$$

- NOTE: These identifications are **nonparametric**
  - We do not assume the model assumption like linear models
  - The question is how to **estimate** the target parameter and what assumption we will make

Overview: Different Adjustment Strategies

- There are many methods to control confounder
- Outcome model
- Matching
    - Propensity Score Matching
    - Coarsened Exact Matching (CEM)
    - Cardinality Matching
- Weighting
    - Inverse Probability Weighting (IPW)
    - Covariate Balancing Propensity Score
    - Entropy Balancing / Stable Weights
- Doubly Robust Estimation (next week)
    - Augmented IPW
    - Double Machine Learning / Semiparametric Estimation

Outcome model / Regression Adjustment

- Let's start with outcome model. Recall that

$$\mathbb{E}[Y_i \mid T_i, X_i] = \alpha + \beta T_i + \gamma^\top X_i$$

- Recall that the identification formula of ATE is given by

$$\tau_{\text{ATE}} = \int \{\mathbb{E}[Y_i \mid T_i = 1, X_i = x] - \mathbb{E}[Y_i \mid T_i = 0, X_i = x]\} dF(x)$$

  - If we assume model is correct, then $\tau_{\text{ATE}} = \beta$

- If your estimand is ATT,

$$\begin{aligned}\hat{\tau}_{\text{ATT}} &= \mathbb{E}[Y_i \widehat{\mid T_i} = 1] - \mathbb{E}[\mathbb{E}[Y_i \mid T_i \widehat{= 0, X}_i] \mid T_i = 1] \\ &= \frac{1}{n_1}\sum_{i=1}^{n} T_i(Y_i - \underbrace{\{\hat{\alpha} + \hat{\gamma}^\top X_i\}}_{\mathbb{E}[Y_i \widehat{\mid T_i = 0}, X_i]})\end{aligned}$$

which is imputing the outcome under control.

- Note that outcome model can be model-dependent

Matching: Overview

- For any outcome regression model $\mathbb{E}[Y_i \mid T_i = 0, X_i] = \hat{\mu}_0(X_i)$, the regression-based estimator for ATT is written as

$$\hat{\tau}_{\text{ATT}} = \frac{1}{n_1} \sum_{i=1}^{n} T_i (Y_i - \hat{\mu}_0(X_i))$$

- Matching is the way to find the observation under control which is closer to treated observation; formally,

$$\hat{\tau}_{\text{Matching}} = \frac{1}{n_1} \sum_{i=1}^{n} T_i \left(Y_i - \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} Y_{i'} \right)$$

  - Notice that in the case of exact matching, $\mathcal{M}_i$ is the set of observations with $X_{i'} = X_i$ for all $i' \in \mathcal{M}_i$ and $T_{i'} = 0$
  - This is why matching is the nonparametric imputation (i.e., reducing model dependence)

Matching: Overview

- **Matching**: Impute missing potential outcomes using the observed outcomes of "closest" units
  - **Goal of Matching**: Remove all imbalances in observed covariates
  - While there are many matching methods, you should choose the one that gives you the best covariate balances!
- When the number of covariates is small and all are discrete, we can do **exact** matching
  - However, often the exact matching is not feasible
- How should we do when we have continuous variables / many controls?
  - Option 1: Matching based on Distance Measures
    - Measure the distance and pick up the nearest neighbor
  - Option 2: Coarsened Exact Matching
    - Coarsening the variable into discrete variables / bins
  - Option 3: Propensity Score Matching
    - Create one dimensional summary of covariates (propensity score)
- However, non-exact matching allows covariate imbalances, which leads to the bias.

Bias Decomposition (Heckman et al. 1998)

- Matching can deal with
  - Bias due to lack of common support
  - Bias due to imbalances in covariates within their common supports
- Matching cannot deal with bias due to unobserved confounder.

$$
\begin{aligned}
&\mathbb{E}[Y_i(0) \mid T_i = 1] - \mathbb{E}[Y_i \mid T_i = 1] \\
&= \underbrace{\int_{S_1 \setminus S} \mathbb{E}[Y_i(0) \mid T_i = 1, X_i] dF_{X_i \mid T_i = 1}(X_i) - \int_{S_0 \setminus S} \mathbb{E}[Y_i(0) \mid T_i = 0, X_i] dF_{X_i \mid T_i = 0}(X_i)}_{\text{Bias due to lack of common support}} \\
&+ \underbrace{\int_{S} \mathbb{E}[Y_i(0) \mid T_i = 0, X_i] \{ dF_{X_i \mid T_i = 1}(X_i) - dF_{X_i \mid T_i = 0}(X_i) \}}_{\text{Bias due to imbalances in covariates within their common support}} \\
&+ \underbrace{\int_{S} \Big(\mathbb{E}[Y_i(0) \mid T_i = 1, X_i] - \mathbb{E}[Y_i(0) \mid T_i = 0, X_i] \Big) dF_{X_i \mid T_i = 1}(X_i)}_{\text{Bias due to unobservable variables within their common support}}
\end{aligned}
$$

Extra: Why matching works

- $\mathcal{M}_i$ is generated given $\{T_i, X_i\}_{i=1}^n$ w/o looking outcome so that

$$\{Y_i\}_{i=1}^n \perp\!\!\!\perp \mathcal{M}_i \mid \{T_i, X_i\}_{i=1}^n$$

Then, with this and i.i.d. sampling (both used in third equality),

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} Y_{i'}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} Y_{i'} \mid X_1,\dots,X_n,\ T_1,\dots,T_n,\ \mathcal{M}_i\right]\right] \\
&= \mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} \mathbb{E}\left[Y_{i'} \,\middle|\, X_1,\dots,X_n,\ T_1,\dots,T_n,\ \mathcal{M}_i\right]\right] \\
&= \mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} \mathbb{E}[Y_{i'} \mid X_{i'}, T_{i'} = 0]\right] \\
&= \mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} \mu_0(X_{i'})\right] \\
&= \mathbb{E}\left[\frac{1}{|\mathcal{M}_i|}\sum_{i'\in\mathcal{M}_i} \mu_0(X_i)\right] \quad (\because X_{i'} = X_i \text{ for all } i') \\
&= \mathbb{E}[\mu_0(X_i)]
\end{aligned}$$

Bias Correction (Abadie and Imbens 2011)

- Let's first derive the bias of matching estimator. Now, recall that

$$\hat{\tau}_{\mathrm{Matching}} = \frac{1}{n_1} \sum_{i=1}^{n} T_i \Big(Y_i - \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} Y_{i'} \Big)$$

Then, for treated observation (i.e., $T_i = 1$),

$$\begin{aligned}
&\mathbb{E}\Big[Y_i - \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} Y_{i'} \mid \{T_j, X_j\}_{j=1}^{N}, \mathcal{M}_i \Big] \\
&= \mathbb{E}[Y_i \mid T_i = 1, X_i] - \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} \mathbb{E}[Y_{i'} \mid T_{i'} = 0, X_{i'}] \\
&= \mathbb{E}[Y_i \mid T_i = 1, X_i] - \frac{1}{|\mathcal{M}_i|} \sum_{i' \in \mathcal{M}_i} \mu_0(X_{i'})
\end{aligned}$$

However, because of non-exact matching, $X_i \neq X_{i'}$ for $i' \in \mathcal{M}_i$ (notice the difference from the slide "Why Matching works").

Bias Correction (Abadie and Imbens 2011)

- Hence, the bias is given by

$$
\begin{aligned}
&\mathbb{E}[\tau_{\text{match}} - \tau] \\
&= \mathbb{E}\Big[\mathbb{E}[Y_i \mid T_i = 1, X_i] - \frac{1}{|\mathcal{M}_i|}\sum_{i' \in \mathcal{M}_i} \mu_0(X_{i'}) \\
&\qquad - \mathbb{E}[Y_i \mid T_i = 1, X_i] + \mathbb{E}[Y_i \mid T_i = 0, X_i]\Big] \\
&= \mathbb{E}\Big[\frac{1}{|\mathcal{M}_i|}\sum_{i' \in \mathcal{M}_i}\Big(\mu_0(X_i) - \mu_0(X_{i'})\Big)\Big]
\end{aligned}
$$

- **Idea of Bias Correction**: Model bias with regression $\mu_0(X_i) = \beta^\top X_i$
    - Then, the bias term is $\beta^\top (X_i - X_{i'})$

Propsneity Score (1)

- **Propensity Score**: $\pi(X_i) = \mathbb{P}(T_i = 1 \mid X_i)$
  - In both matching / weighting, propensity score is used
- Propensity score has the balancing properties:

$$T_i \perp\!\!\!\perp X_i \mid \pi(X_i)$$

- Under ignorability $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i$, balancing property implies ignorability given propensity score:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid \pi(X_i)$$

- **Takeaway**: Instead of conditioning on X_i, we only need to control $\pi(X_i)$
  - Since $\pi(X_i)$ is one-dimensional, it can be used for matching easily
  - But we need to estimate propensity score

Propsneity Score (2): Proof of Balancing Property

- To prove the independence between two random variables A and B, all you need is to show that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

$$\mathbb{P}(T_i = 1 \mid \pi(X_i), X_i) = \mathbb{P}(T_i = 1 \mid X_i) = \pi(X_i)$$

- Also, we can also show that

$$\begin{aligned} \mathbb{P}(T_i = 1 \mid \pi(X_i)) &= \mathbb{E}[T_i \mid \pi(X_i)] \quad (\because T \text{ is binary}) \\ &= \mathbb{E}[\mathbb{E}[T_i \mid X_i, \pi(X_i)] \mid \pi(X_i)] \quad (\because \text{L.I.E}) \\ &= \mathbb{E}[\mathbb{E}[T_i \mid X_i] \mid \pi(X_i)] \\ &= \mathbb{E}[\pi(X_i) \mid \pi(X_i)] = \pi(X_i) \end{aligned}$$

- Therefore,

$$\mathbb{P}(T_i = 1 \mid \pi(X_i), X_i) = \mathbb{P}(T_i = 1 \mid \pi(X_i))$$

which implies $T_i \perp\!\!\!\perp X_i \mid \pi(X_i)$.

Propsneity Score (3): Proof of Ignorability

$$\begin{aligned}
&\mathbb{P}(T_i = 1 \mid Y_i(1), Y_i(0), \pi(X_i)) \\
&= \mathbb{E}[T_i \mid Y_i(1), Y_i(0), \pi(X_i)] \\
&= \mathbb{E}[\mathbb{E}[T_i \mid Y_i(1), Y_i(0), X_i] \mid Y_i(1), Y_i(0), \pi(X_i)] \\
&= \mathbb{E}[\mathbb{E}[T_i \mid X_i] \mid Y_i(1), Y_i(0), \pi(X_i)] \\
&= \mathbb{E}[\pi(X_i) \mid Y_i(1), Y_i(0), \pi(X_i)] \\
&= \pi(X_i)
\end{aligned}$$

- Notice that we already proved $\mathbb{P}(T_i = 1 \mid \pi(X_i)) = \pi(X_i)$, which means

$$\mathbb{P}(T_i = 1 \mid Y_i(1), Y_i(0), \pi(X_i)) = \mathbb{P}(T_i = 1 \mid \pi(X_i))$$

which means

$$T_i \perp\!\!\!\perp \{Y_i(1), Y_i(0)\} \mid \pi(X_i)$$

Extra: Balancing Score and Ignorability

- $B(X_i)$ is a balancing score (i.e., $X_i \perp\!\!\!\perp T_i \mid B(X_i)$) if and only if there exists a function f such that $\pi(X_i) = f(B(X_i))$.
  - Propensity score is a coarsest balancing score.
  - For proof, see Theorem 2 of Rosenbaum and Rubin (1983)

- **Implication**: Under strong ignorability, any balancing score leads to ignorability given balancing score

- Proof: Similar to the proof in the previous page,

$$
\begin{aligned}
\mathbb{P}(T_i = 1 \mid B(X_i)) &= \mathbb{E}[T_i \mid B(X_i)] \quad (\because T \text{ is binary}) \\
&= \mathbb{E}[\mathbb{E}[T_i \mid X_i, B(X_i)] \mid B(X_i)] \quad (\because \text{L.I.E}) \\
&= \mathbb{E}[\mathbb{E}[T_i \mid X_i] \mid B(X_i)] \\
&= \mathbb{E}[\pi(X_i) \mid B(X_i)] \\
&= \mathbb{E}[\pi(X_i) \mid f(\pi(X_i))] = \pi(X_i)
\end{aligned}
$$

Extra: Balancing Score and Ignorability: Cont.

- Proof (cont.): Also,

$$\begin{aligned}
&\mathbb{P}(T_i = 1 \mid Y_i(1), Y_i(0), B(X_i)) \\
&= \mathbb{E}[T_i \mid Y_i(1), Y_i(0), B(X_i)] \\
&= \mathbb{E}[\mathbb{E}[T_i \mid Y_i(1), Y_i(0), X_i] \mid Y_i(1), Y_i(0), B(X_i)] \\
&= \mathbb{E}[\mathbb{E}[T_i \mid X_i] \mid Y_i(1), Y_i(0), B(X_i)] \\
&= \mathbb{E}[\pi(X_i) \mid Y_i(1), Y_i(0), B(X_i)] \\
&= \mathbb{E}[\pi(X_i) \mid Y_i(1), Y_i(0), f(\pi(X_i))] \\
&= \pi(X_i)
\end{aligned}$$

- Therefore,

$$\mathbb{P}(T_i = 1 \mid Y_i(1), Y_i(0), B(X_i)) = \mathbb{P}(T_i = 1 \mid B(X_i))$$

  which implies $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid B(X_i)$.

- This explains why achieving balance is important
  - We only need to balance the observed covariates
  - The problem is how to achieve the best balance

Cardinality Matching

- **Idea**: Match pairs so that we can have minimal imbalances

$$
\begin{aligned}
\max \quad & \sum_{i:T_i=1} \sum_{j:T_j=0} M_{ij} \\
\text{s.t.} \quad & \underbrace{\sum_{j:T_j=0} M_{ij} \leq 1 \text{ for each } i \text{ with } T_i = 1}_{\text{For treatment } i\text{, at most one match from control}} \\
& \underbrace{\sum_{i:T_i=1} M_{ij} \leq 1 \text{ for each } j \text{ with } T_j = 0}_{\text{For control } j\text{, at most one match from treatment}} \\
& \underbrace{\frac{\left|\sum_{i:T_i=1}\sum_{j:T_j=0} M_{ij}\{f_k(X_{ik}) - f_k(X_{jk})\}\right|}{\sum_{i:T_i=1}\sum_{j:T_j=0} M_{ij}} \leq \epsilon_k, \text{ for each } k}_{\text{Acceptable imbalances between matched pair}}
\end{aligned}
$$

Balance Test / Equivalence Test

- As balance of covariate is important, we often want the low-dimensional summary on it
  - Balance Test: measure the standardized difference in covariates and report if we reject the null
  - **Problem**: Failure to reject the null $\neq$ Acceptance of null hypothesis
    - Null hypothesis: Covariates are balanced

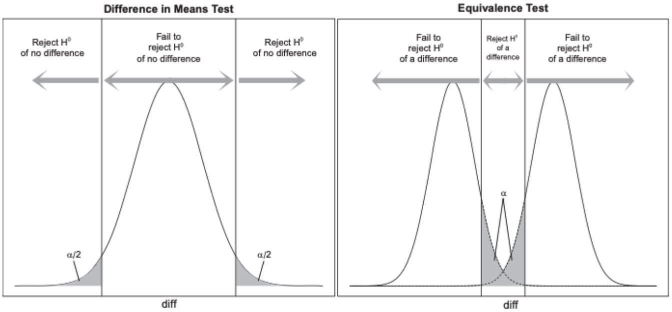
- **Equivalence Test**: Null hypothesis is "covariates are not balanced"
  - By rejecting the null, we can statistically say that covariates are balanced.
  - Formally, with pre-specified equivalence margin $\Delta > 0$,

$$H_0 : |\tau| \geq \Delta \quad \text{(Difference is too large)}$$
$$H_1 : |\tau| \leq \Delta \quad \text{(Difference is small enough)}$$

Two One-Sided Tests (TOST)

- **Two One-Sided Tests**: Most classical procedure for equivalence test
  - Specifically, test two times:

$$1.H_0^{(1)} : \tau \leq -\Delta \quad \text{v.s.} \quad H_1^{(1)} : \tau > -\Delta$$
$$2.H_0^{(2)} : \tau \geq +\Delta \quad \text{v.s.} \quad H_1^{(1)} : \tau < +\Delta$$



Note: The left panel depicts the logic of tests of difference under the null hypothesis of no difference. The right panel depicts the logic of one type of equivalence test—the two one-sided t-test (TOST)—under the null hypothesis of difference.

Weighting: Overview

- Limitation of Matching
  - It can throw away many observations
  - It may not be able to balance covariates
- **Idea**: Weight each observation so that the covariate is balanced
- **Horvitz-Thompson estimator** (a.k.a inverse probability weighting)

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{\pi}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{\pi}(X_i)}\right\}$$

$$\widehat{\text{ATT}} = \frac{1}{n_1}\sum_{i=1}^{n}\left\{T_i Y_i - \frac{\hat{\pi}(X_i)(1-T_i)Y_i}{1-\hat{\pi}(X_i)}\right\}$$

$$\widehat{\text{ATC}} = \frac{1}{n_0}\sum_{i=1}^{n}\left\{\frac{(1-\hat{\pi}(X_i))T_i Y_i}{\hat{\pi}(X_i)} - (1-T_i)Y_i\right\}$$

Unbiasedness of HT-estimator (ATE)

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{T_iY_i}{\pi(X_i)}-\frac{(1-T_i)Y_i}{1-\pi(X_i)}\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{T_iY_i(1)}{\pi(X_i)}-\frac{(1-T_i)Y_i(0)}{1-\pi(X_i)}\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\{\mathbb{E}\left[\frac{T_iY_i(1)}{\pi(X_i)}-\frac{(1-T_i)Y_i(0)}{1-\pi(X_i)}\mid X_i\right]\right\}\quad (\because \text{L.I.E})$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\mathbb{E}[T_iY_i(1)\mid X_i]}{\pi(X_i)}-\frac{\mathbb{E}[(1-T_i)Y_i(0)\mid X_i]}{1-\pi(X_i)}\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{\overbrace{\mathbb{E}[T_i\mid X_i]}^{=\pi(X_i)}\mathbb{E}[Y_i(1)\mid X_i]}{\pi(X_i)}-\frac{\overbrace{\mathbb{E}[1-T_i\mid X_i]}^{=1-\pi(X_i)}\mathbb{E}[Y_i(0)\mid X_i]}{1-\pi(X_i)}\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[\mathbb{E}[Y_i(1)\mid X_i]-\mathbb{E}[Y_i(0)\mid X_i]\Big]$$

$$=\mathbb{E}[Y_i(1)-Y_i(0)]$$

Hajek estimator / Stabilized weights

- **Problem**: Weights can be extreme
  - Thus, we need to stabilize the weights
- **Hajek Estimator**: Use the normalized weights; i.e., for $\widehat{\mathbb{E}[Y_i(1)]}$, use

$$\frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{T_i/\hat{\pi}(X_i)}{\sum_{j=1}^{n} T_j/\hat{\pi}(X_j)}}_{\text{weight of Hajek}} Y_i \quad \left(\leftrightarrow \text{HT is } \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{T_i}{\hat{\pi}(X_i)}}_{\text{HT Weight}} Y_i\right)$$

- Hajek estimator is asymptotically consistent (close to truth for large N)
  - The denominator is $\frac{1}{n}\mathbb{E}[\sum_{j=1}^{n} \frac{T_j}{\pi(X_j)}] = \frac{1}{n}\mathbb{E}[\sum_{j=1}^{n} \mathbb{E}[\frac{T_j}{\pi(X_j)} \mid X_j]] = 1$
  - The numerator is same as HT estimator, which is in expectation equal to $\mathbb{E}[Y_i(1)]$

Toward Better Estimation of Propensity Score

- Both HT and Hajek estimators require estimation of propensity score
  - i.e., if propensity score is misspecified, we have the bias

- Three different approaches (next week)
  1. Covariate Balancing Propensity Score (CBPS)
     - Estimate propensity score s.t. we achieve balance
     - But still assume parametric assumption on propensity score function
  2. Calibration: Entropy Balancing / Stable Weights
     - Estimate weight so that we achieve balance
     - We no longer estimate propensity score
  3. Causal Machine Learning / Semiparametric Estimation
     - Flexibly estimate propensity score / outcome models
     - Relax parametric assumption as much as possible