# CS 477/677 Causal Inference: Homework 2
# Ignorable Causal Models

Ha Bui
hbui13@jhu.edu

## 1 Analytical (26 Points)

**1 (9 points)** Assume we have data on a randomized controlled trial (RCT) where the relevant variables are: an outcome $Y$, a set of covariates $\vec{X}$, and a treatment $A$. Since we are dealing with an RCT, $A \perp\!\!\!\perp Y(a), \vec{X}$ for all values $a$.

- Show that the ACE is identified by $\sum_{\vec{x}} \{\mathbb{E}[Y|A=1,\vec{x}] - \mathbb{E}[Y|A=0,\vec{x}]\}\, p(\vec{x})$.

  **Solution**:

  Due to RCT, $A \perp\!\!\!\perp Y(a), \vec{X}$ for all values $a$, then:

  $$A \perp\!\!\!\perp Y(a) \mid \vec{X} \text{ (weak union)} \tag{1}$$

  Then, we have:

  $$\begin{aligned} p(Y(a)) &= \sum_{\vec{x}} p(Y(a)|\vec{x})p(\vec{x}) \\ &= \sum_{\vec{x}} p(Y(a)|A,\vec{x})p(\vec{x}) \text{ (since } A \perp\!\!\!\perp Y(a) \mid \vec{X} \text{ in 1)} \\ &= \sum_{\vec{x}} p(Y|A=a,\vec{x})p(\vec{x}) \text{ (by consistency, } Y(A) = Y \text{ when } A = a) \end{aligned}$$

  So, we obtain ACE:

  $$\begin{aligned} E(Y(1)) - E(Y(0)) &= E(Y|A=1,\vec{X}) - E(Y|A=0,\vec{X}) \\ &= \sum_{\vec{x}} \{\mathbb{E}[Y|A=1,\vec{x}] - \mathbb{E}[Y|A=0,\vec{x}]\}\, p(\vec{x}) \end{aligned}$$

- Give one advantage of using the parametric g-formula estimator (which includes $\vec{x}$) over the Neyman estimator (which ignores $\vec{x}$).

  **Solution**:

  Assume Ignorability doesn't hold, $A \not\perp\!\!\!\perp Y(a)$, then $E(Y(a)) \neq E(Y|A=a)$. Therefore, with Neyman estimator, ACE is unidentifed.

  However, we have Conditional Ignorability from 1, $A \perp\!\!\!\perp Y(a) \mid \vec{X}$, then with g-formula estimator, ACE is identified by:

  $$\mathbb{E}(Y(1)) - \mathbb{E}(Y(0)) = \frac{1}{n}\left\{\sum_i \mathbb{E}(Y|A=1,\vec{x_i};\hat{\eta_1}) - \mathbb{E}(Y|A=0,\vec{x_i};\hat{\eta_1})\right\} \tag{2}$$

  , where $\hat{\eta_1}$ is estimator parameters from $\mathbb{E}(Y|A,\vec{X};\eta_1)$

So, g-formula estimator is more flexible if causation is not association.

- Give one advantage of using the Neyman estimator (which ignores $\vec{x}$) over the parametric g-formula estimator (which includes $\vec{x}$).

  **Solution**:

  Due to RCT, $A \perp\!\!\!\perp Y(a), \vec{X}$ for all values $a$, we got the Ignorability:

  $$A \perp\!\!\!\perp Y(a) \text{ (decomposition)}$$

  Combining with Consistency $Y(a) = Y$ when $A = a$, we have:

  $$\begin{cases} p(Y(1)) = p(Y(1)|A = 1) = p(Y|A = 1) \\ p(Y(0)) = p(Y(0)|A = 0) = p(Y|A = 0) \end{cases} \tag{3}$$

  Apply Neyman Estimator, we got ACE:

  $$\begin{aligned} \mathbb{E}(Y(1)) - \mathbb{E}(Y(0)) &= \mathbb{E}(Y|A = 1) - \mathbb{E}(Y|A = 0) \text{ (since 3)} \\ &= \left(\frac{1}{k}\sum_{i=1}^{n} Y_i \cdot A_i\right) - \left(\frac{1}{n-k}\sum_{i=1}^{n} Y_i \cdot (1 - A_i)\right) \end{aligned}$$

  So, compared to ACE with g-form estimation in 2 (which need to do MLE), the Neyman estimator is a less complex model than the g-form.

**2 (3 points)** Show that $\sum_{a',m} \mathbb{E}[Y|m, a']p(a')p(m|a)$ is equal to $\mathbb{E}[Y\frac{p(M|a)}{p(M|A)}]$.
**Solution**:

$$\begin{aligned} \sum_{a',m} \mathbb{E}[Y|m, a']p(a')p(m|a) &= \sum_{a',m}\sum_{y} yp(y|m, a')p(a')p(m|a) \\ &= \sum_{a',m,y} y\frac{p(y, m|a')}{p(m|a')}p(a')p(m|a) \\ &= \sum_{a',m,y} y\frac{p(y, m, a')}{p(m|a')}p(m|a) \\ &= \mathbb{E}[Y\frac{p(M|a)}{p(M|A)}] \end{aligned}$$

**3 (5 points)** Assume we have a dataset with variable $X_1$, which is always observed, and variables $X_2$ and $X_3$ which are potentially missing. Let $R_2$ and $R_3$ be variables indicating whether $X_2$ and $X_3$ are observed. Assume the following:

- $X_2(R_2 = 1) \perp\!\!\!\perp R_2 \mid X_1$
- $X_3(R_3 = 1) \perp\!\!\!\perp R_3 \mid X_1, X_2(R_2 = 1)$
- $R_2 = 0 \implies R_3 = 0$, that is, if $R_2$ is 0, $R_3$ must be 0

With the above given, show that $p(X_1, X_2(R_2 = 1), X_3(R_3 = 1))$ is identified.

**Solution**:

From $X_2(R_2 = 1) \perp\!\!\!\perp R_2 \mid X_1$, (with $X_1$ fully observed), then:

$$\begin{aligned} p(X_2(R_2 = 1) \mid X_1) &= p(X_2(R_2 = 1) \mid R_2, X_1) \\ &= p(X_2(R_2 = 1) \mid R_2 = 1, X_1) \\ &= p(X_2 \mid R_2 = 1, X_1) \text{ (since } X_2 = X_2 \text{ if } R_2 = 1, X_1) \end{aligned} \tag{4}$$

Due to Missing data at Random in $X_2$ and $X_3$ while $X_1$ fully observed, then:

$$\begin{cases} X_2 = X_2 \text{ if } R_2 = 1, X_1 \\ X_2 =? \text{ if } R_2 = 0, X_1 \end{cases}, \quad \begin{cases} X_3 = X_3 \text{ if } R_3 = 1, X_1 \\ X_3 =? \text{ if } R_2 = R_3 = 0, X_1 \text{ (by assume } R_2 = 0 \implies R_3 = 0) \end{cases} \tag{5}$$

From $X_3(R_3 = 1) \perp\!\!\!\perp R_3 \mid X_1, X_2(R_2 = 1)$ (with $X_1$ fully observed), then:

$$\begin{aligned} p(X_3(R_3 = 1) \mid X_1, X_2(R_2 = 1)) &= p(X_3(R_3 = 1) \mid R_3, X_1, X_2(R_2 = 1)) \\ &= p(X_3(R_3 = 1) \mid R_3 = 1, X_1, X_2(R_2 = 1)) \\ &= p(X_3 \mid R_3 = 1, X_1, X_2(R_2 = 1)) \\ &= p(X_3 \mid R_3 = 1, X_1, X_2, R_2 = 1) \text{ (since 4 and 5)} \end{aligned} \tag{6}$$

Therefore, we obtain:

$$\begin{aligned} & p(X_1, X_2(R_2 = 1), X_3(R_3 = 1)) \\ & = p(X_1)p(X_2(R_2 = 1), X_3(R_3 = 1)|X_1) \\ & = p(X_1)p(X_3(R_3 = 1)|X_2(R_2 = 1), X_1)p(X_2(R_2 = 1)|X_1) \\ & = p(X_1)p(X_3|R_3 = 1, X_1, X_2, R_2 = 1)p(X_2|R_2 = 1, X_1) \text{ (since 4 and 6)} \end{aligned}$$

So, $p(X_1, X_2(R_2 = 1), X_3(R_3 = 1))$ is identified (because depends only on observed variable $X_1, X_2, X_3$ given $R_2 = 1, R_3 = 1$).

**4 (4 points)** Assume $\{Y(1), Y(0)\} \perp\!\!\!\perp A \mid \mathbf{X}$. Is $E[Y(a) \mid \mathbf{W}]$, for $\mathbf{W} \subseteq \mathbf{X}$, identifiable? If not, give an intuitive argument for why. If so, derive the identifying formula.

**Solution**: Yes. Because:

Due to Conditional Ignorability: $\{Y(1), Y(0)\} \perp\!\!\!\perp A \mid \mathbf{X}$ and Consistency $Y(a) = Y$ when $A = a$, then:

$$\begin{aligned} \mathbb{E}[Y(a)|\mathbf{X}] &= \mathbb{E}[Y(a)|A, \mathbf{X}] \\ &= \mathbb{E}[Y|A = a, \mathbf{X}] \\ &= \sum_y y p(Y = y|A = a, \mathbf{X}) \end{aligned} \tag{7}$$

Due to $\mathbf{W} \subseteq \mathbf{X}$, then:

$$\begin{aligned} \mathbb{E}(Y(a) \mid \mathbf{W}) &= \mathbb{E}\left[\mathbb{E}[Y(a)|\mathbf{X}] \mid \mathbf{W}\right] \\ &= \mathbb{E}\left[\sum_y y p(Y = y|A = a, \mathbf{X}) \mid \mathbf{W}\right] \text{ (since 7)} \\ &= \sum_g \sum_y g y p(Y = y|A = a, \mathbf{X} = g), \text{ where } g \in \mathbf{W} \subseteq \mathbf{X} \end{aligned}$$

**5 (5 points)** The average effect of treatment on the treated (AETT) is defined as

$$E[Y(1)|A=1] - E[Y(0)|A=1]$$

Assume $A$ is binary, and that consistency is true. Prove that the AETT is identifiable if $p(Y(0))$ is identifiable.

**Solution**:

Due to $A \in \{0,1\}$, then:

$$\begin{aligned} & p(Y(0)) = p(Y(0), A=1) + p(Y(0), A=0) \\ & \Leftrightarrow p(Y(0)) = p(Y(0)|A=1)p(A=1) + p(Y(0)|A=0)p(A=0) \\ & \Leftrightarrow \mathbb{E}[Y(0)] = p(A=1)\mathbb{E}[Y(0)|A=1] + p(A=0)\mathbb{E}[Y(0)|A=0] \end{aligned} \tag{8}$$

If $p(Y(0))$ is identifiable, then $\mathbb{E}(Y(0))$ is identifiable. We have:

$$\begin{aligned} \mathbb{E}(Y(0)|A=1) &= \frac{\mathbb{E}[Y(0)] - p(A=0)\mathbb{E}[Y(0)|A=0]}{p(A=1)} \text{ (since 8)} \\ &= \frac{\mathbb{E}[Y(0)] - p(A=0)\mathbb{E}[Y|A=0]}{p(A=1)} \\ & \text{(by consistency, } Y(a) = Y \text{ when } A = a) \end{aligned} \tag{9}$$

Therefore, AETT is:

$$\begin{aligned} & \mathbb{E}([Y(1)|A=1] - \mathbb{E}([Y(0)|A=1] \\ & = \mathbb{E}([Y|A=1] - \mathbb{E}([Y(0)|A=1] \text{ (by consistency, } Y(a) = Y \text{ when } A = a) \\ & = \mathbb{E}([Y|A=1] - \left[\frac{\mathbb{E}[Y(0)] - p(A=0)\mathbb{E}[Y|A=0]}{p(A=1)}\right] \text{ (since 9)} \end{aligned}$$

So, AETT is identifiable if $p(Y(0))$ is identifiable.