PLSC 30600

Week 2: Identification, ignorability, and propensity scores

Molly Offer-Westort

Department of Political Science,
University of Chicago

Winter 2026

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\text{Supp}[R_i] = \{0, 1\}$.

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\text{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Y_i is MCAR if:

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Y_i is MCAR if:
    - $Y_i \perp R_i$ (independence of outcome and response).

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\text{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Y_i is MCAR if:
  - $Y_i \perp R_i$ (independence of outcome and response).
  - $\Pr[R_i = 1] > 0$ (nonzero probability of response).

Recap MCAR: Missing Completely at Random

- Let Y_i and R_i be random variables with $\operatorname{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$.
- Y_i is MCAR if:
  - $Y_i \perp R_i$ (independence of outcome and response).
  - $\Pr[R_i = 1] > 0$ (nonzero probability of response).
- Note: implicitly, MCAR is with respect to *all* data, observed and unobserved. So, $R_i \perp (Y_i, X_i, Z_i, \ldots)$.

Missingness wrt the full data

Definition (Full data, observed data, and response indicator)

Let U_i denote the *full-data* random vector for unit i, and let $R_i \in \{0, 1\}$ indicate whether the outcome component Y_i of U_i is observed. Write $U_i = (Y_i, X_i)$ where X_i collects all covariates measured (or conceptually present) for unit i. We observe

$$Y_i^* = Y_i R_i + (-99)(1 - R_i).$$

MCAR: Missing Completely at Random

Definition (Missing Completely at Random (MCAR))

Let $U_i = (Y_i, \boldsymbol{X}_i)$ be the full-data vector and let $R_i \in \{0, 1\}$ be the response indicator. We say Y_i is *missing completely at random* if the following conditions hold:

1. $R_i \perp\!\!\!\perp U_i$ (missingness is independent of the full data),
2. $\Pr(R_i = 1) > 0$ (nonzero probability of response).

Equivalently, $R_i \perp\!\!\!\perp (Y_i, X_i)$ and $\Pr(R_i = 1) > 0$.

MAR: Missing at Random

Definition (Missing at Random (MAR))

Let $U_i = (Y_i, X_i)$ be the full-data vector and let $R_i \in \{0, 1\}$ be the response indicator. We observe X_i for all units (even when $R_i = 0$). We say Y_i is *missing at random conditional on* X_i if the following conditions hold:

1. $R_i \perp\!\!\!\perp Y_i \mid X_i$ (conditional independence of missingness and outcome),
2. $\exists\, \varepsilon > 0$ such that $\Pr(R_i = 1 \mid X_i) > \varepsilon$ a.s. (positivity).

Proposition (MCAR implies MAR)

If $R_i \perp\!\!\!\perp (Y_i, X_i)$, then $R_i \perp\!\!\!\perp Y_i \mid X_i$. Therefore, MCAR implies MAR (with the same positivity condition).

Recap: missing data → potential outcomes

- Potential outcomes: $Y_i(1)$ and $Y_i(0)$.

Recap: missing data → potential outcomes

- Potential outcomes: $Y_i(1)$ and $Y_i(0)$.
- Observed outcome: $Y_i = Y_i(d) : \quad D_i = d$.

Recap: missing data → potential outcomes

- Potential outcomes: $Y_i(1)$ and $Y_i(0)$.
- Observed outcome: $Y_i = Y_i(d) : \quad D_i = d$.
- Only one potential outcome is observed; the other is missing.

Recap: missing data → potential outcomes

- Potential outcomes: $Y_i(1)$ and $Y_i(0)$.
- Observed outcome: $Y_i = Y_i(d) : \quad D_i = d$.
- Only one potential outcome is observed; the other is missing.
- Identification requires assumptions about the assignment mechanism.

Recap: Manski and Nagin (1998) Utah juvenile court data

- How should judges sentence convicted juvenile offenders?

Recap: Manski and Nagin (1998) Utah juvenile court data

- How should judges sentence convicted juvenile offenders?
- Juvenile offenders in Utah may be assigned to residential or non-residential treatment programs.

  Treatment (sentencing): $D = 1$ residential, $D = 0$ non-residential

  Outcome (recidivism): $Y = 1$ recidivates, $Y = 0$ does not recidivate

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}[D_i] = \{0, 1\}$.

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\left[D_i\right] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **randomly assigned** if:

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **randomly assigned** if:
  - $(Y_i(0), Y_i(1)) \perp D_i$ (independence of potential outcomes and treatment).

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **randomly assigned** if:
  - $(Y_i(0), Y_i(1)) \perp D_i$ (independence of potential outcomes and treatment).

  - $0 < \Pr[D_i = 1] < 1$ (positivity).

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **randomly assigned** if:
  - $(Y_i(0), Y_i(1)) \perp D_i$ (independence of potential outcomes and treatment).

  - $0 < \Pr[D_i = 1] < 1$ (positivity).

- Under this assumption, $\mathrm{E}[\tau_i]$ is point identified.

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **randomly assigned** if:
  - $(Y_i(0), Y_i(1)) \perp D_i$ (independence of potential outcomes and treatment).
  - $0 < \Pr[D_i = 1] < 1$ (positivity).

- Under this assumption, $\mathrm{E}[\tau_i]$ is point identified.

- Aside: Does this independence need to hold for the full data?

Random assignment

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$.
- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.
- D_i is **randomly assigned** if:
  - $(Y_i(0), Y_i(1)) \perp D_i$ (independence of potential outcomes and treatment).
  - $0 < \Pr[D_i = 1] < 1$ (positivity).
- Under this assumption, $\mathrm{E}[\tau_i]$ is point identified.
- Aside: Does this independence need to hold for the full data?
- Implicitly: $R_i \perp\!\!\!\perp U_i^{pre}$ (missingness is independent of the full pre-intervention data)

ATE under random assignment

- If D_i is randomly assigned, then:

ATE under random assignment

- If D_i is randomly assigned, then:

$$\mathrm{E}\big[Y_i(1)\big] = \mathrm{E}\big[Y_i \mid D_i = 1\big], \quad \mathrm{E}\big[Y_i(0)\big] = \mathrm{E}\big[Y_i \mid D_i = 0\big].$$

$$\mathrm{E}\left[\tau_i\right] = \mathrm{E}\big[Y_i \mid D_i = 1\big] - \mathrm{E}\big[Y_i \mid D_i = 0\big].$$

Difference-in-means estimator

- Plug-in estimator under random assignment:

Difference-in-means estimator

- Plug-in estimator under random assignment:

$$\widehat{\mathrm{E}}[\tau_i] = \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i(1-D_i)}{\sum_{i=1}^n (1-D_i)}.$$

Difference-in-means estimator (matrix form)

- Let

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix}, \quad \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Difference-in-means estimator (matrix form)

- Let

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix}, \quad \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

- Let $n_1 = \mathbb{1}^\top \boldsymbol{D}$ and $n_0 = \mathbb{1}^\top(\mathbb{1} - \boldsymbol{D})$.

Difference-in-means estimator (matrix form)

- Let

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix}, \quad \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

- Let $n_1 = \mathbb{1}^\top \boldsymbol{D}$ and $n_0 = \mathbb{1}^\top(\mathbb{1} - \boldsymbol{D})$. Mechanically,

$$n_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_n \end{bmatrix} = \sum_{i=1}^{n} 1 D_i,$$

$$n_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 - D_1 \\ 1 - D_2 \\ \vdots \\ 1 - D_n \end{bmatrix} = \sum_{i=1}^{n} 1(1 - D_i).$$

Difference-in-means estimator (matrix form)

$$\widehat{\mathrm{E}}\left[\tau_i\right] = \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right)^{\top} \boldsymbol{Y}$$

Difference-in-means estimator (matrix form)

$$\widehat{\mathrm{E}}[\tau_i] = \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right)^\top \boldsymbol{Y} \qquad \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right) = \begin{bmatrix} \frac{D_1}{n_1} - \frac{1-D_1}{n_0} \\ \frac{D_2}{n_1} - \frac{1-D_2}{n_0} \\ \vdots \\ \frac{D_n}{n_1} - \frac{1-D_n}{n_0} \end{bmatrix}$$

Difference-in-means estimator (matrix form)

$$\widehat{\mathrm{E}}[\tau_i] = \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right)^\top \boldsymbol{Y} \qquad \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right) = \begin{bmatrix} \frac{D_1}{n_1} - \frac{1-D_1}{n_0} \\ \frac{D_2}{n_1} - \frac{1-D_2}{n_0} \\ \vdots \\ \frac{D_n}{n_1} - \frac{1-D_n}{n_0} \end{bmatrix}$$

$$\widehat{\mathrm{E}}[\tau_i] = \left[\frac{D_1}{n_1} - \frac{1-D_1}{n_0}, \frac{D_2}{n_1} - \frac{1-D_2}{n_0}, \ldots, \frac{D_n}{n_1} - \frac{1-D_n}{n_0}\right] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

Difference-in-means estimator (matrix form)

$$\widehat{\mathrm{E}}[\tau_i] = \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right)^\top \boldsymbol{Y} \qquad \left(\frac{\boldsymbol{D}}{n_1} - \frac{\mathbb{1} - \boldsymbol{D}}{n_0}\right) = \begin{bmatrix} \frac{D_1}{n_1} - \frac{1-D_1}{n_0} \\ \frac{D_2}{n_1} - \frac{1-D_2}{n_0} \\ \vdots \\ \frac{D_n}{n_1} - \frac{1-D_n}{n_0} \end{bmatrix}$$

$$\widehat{\mathrm{E}}[\tau_i] = \left[\frac{D_1}{n_1} - \frac{1-D_1}{n_0}, \frac{D_2}{n_1} - \frac{1-D_2}{n_0}, \cdots, \frac{D_n}{n_1} - \frac{1-D_n}{n_0}\right] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} \left(\frac{D_i}{n_1} - \frac{1 - D_i}{n_0}\right) Y_i = \frac{1}{n_1} \sum_{i=1}^{n} D_i Y_i - \frac{1}{n_0} \sum_{i=1}^{n} (1 - D_i) Y_i$$

$$= \frac{\sum_{i=1}^{n} Y_i D_i}{\sum_{i=1}^{n} D_i} - \frac{\sum_{i=1}^{n} Y_i (1 - D_i)}{\sum_{i=1}^{n} (1 - D_i)}.$$

Utah example: random assignment intuition

- Treat D_i as randomly assigned to residential vs. non-residential.

Utah example: random assignment intuition

- Treat D_i as randomly assigned to residential vs. non-residential.
- Then treated and control groups have the same potential outcome distributions.

Utah example: random assignment intuition

- Treat D_i as randomly assigned to residential vs. non-residential.
- Then treated and control groups have the same potential outcome distributions.
- Sample means within $D = 1$ and $D = 0$ identify $\mathrm{E}[Y_i(1)]$ and $\mathrm{E}[Y_i(0)]$.

Science table: random assignment

What would the difference in recidivism rates be if all juveniles were assigned to residential instead of non-residential treatment?

Target: $\mathrm{E}[Y_i(1)] - \mathrm{E}[Y_i(0)]$

i	$Y_i(0)$	$Y_i(1)$	D_i	Y_i
1	0	?	0	0
2	?	1	1	1
3	1	?	0	1
4	?	1	1	1
5	0	?	0	0
6	?	1	1	1
7	1	?	0	1
8	?	0	1	0
9	1	?	0	0
10	?	1	1	1

$$\widehat{\mathrm{E}}[Y_i(1)] = \frac{1+1+1+0+1}{5} = 0.8, \quad \widehat{\mathrm{E}}[Y_i(0)] = \frac{0+1+0+1+1}{5} = 0.6$$

$$\widehat{\mathrm{E}}[\tau_i] = 0.8 - 0.6 = 0.2$$

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\left[D_i\right] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **strongly ignorable conditional on** X_i if:

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\operatorname{Supp}[D_i] = \{0, 1\}$.
- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.
- D_i is **strongly ignorable conditional on** X_i if:
    - $(Y_i(0), Y_i(1)) \perp D_i \mid X_i$ (conditional independence).

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **strongly ignorable conditional on** X_i if:
    - $(Y_i(0), Y_i(1)) \perp D_i \mid X_i$ (conditional independence).
    - $\exists \varepsilon > 0$ such that $\varepsilon < \Pr[D_i = 1 \mid X_i] < 1 - \varepsilon$ (positivity).

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **strongly ignorable conditional on** X_i if:
  - $(Y_i(0), Y_i(1)) \perp D_i \mid X_i$ (conditional independence).
  - $\exists \varepsilon > 0$ such that $\varepsilon < \Pr[D_i = 1 \mid X_i] < 1 - \varepsilon$ (positivity).

- Random assignment implies strong ignorability (with pre-treatment X_i).

Strong Ignorability

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- D_i is **strongly ignorable conditional on** X_i if:
  - $(Y_i(0), Y_i(1)) \perp D_i \mid X_i$ (conditional independence).

  - $\exists \varepsilon > 0$ such that $\varepsilon < \Pr[D_i = 1 \mid X_i] < 1 - \varepsilon$ (positivity).

- Random assignment implies strong ignorability (with pre-treatment X_i). Strong ignorability does *not* imply random assignment.

ATE under Strong Ignorability

- If D_i is strongly ignorable conditional on X_i, then:

ATE under Strong Ignorability

- If D_i is strongly ignorable conditional on X_i, then:

$$\mathrm{E}\left[\tau_i\right] = \sum_x \mathrm{E}\big[Y_i \mid D_i = 1, X_i = x\big] \Pr[X_i = x] - \sum_x \mathrm{E}\big[Y_i \mid D_i = 0, X_i = x\big] \Pr[X_i = x].$$

Conditional ATE under Strong Ignorability

- For any $x \in \mathrm{Supp}[X_i]$:

Conditional ATE under Strong Ignorability

- For any $x \in \mathrm{Supp}[X_i]$:

$$\mathrm{E}[\tau_i \mid X_i] = \mathrm{E}\big[Y_i \mid D_i = 1, X_i\big] - \mathrm{E}\big[Y_i \mid D_i = 0, X_i\big].$$

Post-stratification plug-in estimator under MAR

- Under MAR, $Y \perp R \mid X$, so:
$$\mathrm{E}[Y] = \sum_{x \in \mathrm{Supp}[X]} \mathrm{E}[Y \mid R = 1, X = x] \Pr[X = x].$$

Post-stratification plug-in estimator under MAR

- Under MAR, $Y \perp R \mid X$, so:

$$\mathrm{E}[Y] = \sum_{x \in \mathrm{Supp}[X]} \mathrm{E}[Y \mid R = 1, X = x] \Pr[X = x].$$

- Sample plug-in estimator:

$$\widehat{\mathrm{E}}[Y] = \sum_{x \in \mathrm{Supp}[X]} \left(\frac{\sum_{i=1}^{n} Y_i^* \, \mathbb{1}(R_i = 1)\mathbb{1}(X_i = x)}{\sum_{i=1}^{n} \mathbb{1}(R_i = 1)\mathbb{1}(X_i = x)} \right) \left(\frac{\sum_{i=1}^{n} \mathbb{1}(X_i = x)}{n} \right).$$

Science table: ignorability (conditioning on X)

i	X_i	$Y_i(0)$	$Y_i(1)$	D_i	Y_i
1	A	0	?	0	0
2	A	?	1	1	1
3	B	1	?	0	1
4	B	?	1	1	1
5	A	0	?	0	0
6	A	?	1	1	1
7	B	1	?	0	1
8	B	?	0	1	0
9	A	1	?	0	0
10	B	?	1	1	1

$$\widehat{\mathrm{E}}[Y_i \mid D_i = 1, X_i = \mathsf{A}] = 1, \quad \widehat{\mathrm{E}}[Y_i \mid D_i = 0, X_i = \mathsf{A}] = 1/3$$

$$\widehat{\mathrm{E}}[Y_i \mid D_i = 1, X_i = \mathsf{B}] = 2/3, \quad \widehat{\mathrm{E}}[Y_i \mid D_i = 0, X_i = \mathsf{B}] = 1$$

$$\widehat{\mathrm{E}}[Y_i(1)] = \frac{5}{10} \cdot 1 + \frac{5}{10} \cdot \frac{2}{3} = 0.833, \quad \widehat{\mathrm{E}}[Y_i(0)] = \frac{5}{10} \cdot \frac{1}{3} + \frac{5}{10} \cdot 1 = 0.667$$

$$\widehat{\mathrm{E}}[\tau_i] = 0.833 - 0.667 = 0.167$$

Response Propensity Function

- Suppose we observe (R_i, X_i) with $\mathrm{Supp}\,[R_i] = \{0, 1\}$.

Response Propensity Function

- Suppose we observe (R_i, X_i) with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- The response propensity function is:

Response Propensity Function

- Suppose we observe (R_i, X_i) with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- The response propensity function is:
$$p_R(x) = \Pr[R_i = 1 \mid X_i = x], \quad \forall x \in \mathrm{Supp}[X_i].$$
- The random variable $p_R(X_i)$ is the (response) propensity score for unit i.

Response Propensity Function

- Suppose we observe (R_i, X_i) with $\mathrm{Supp}[R_i] = \{0, 1\}$.

- The response propensity function is:

$$p_R(x) = \Pr[R_i = 1 \mid X_i = x], \quad \forall x \in \mathrm{Supp}[X_i].$$

- The random variable $p_R(X_i)$ is the (response) propensity score for unit i.

- It is the conditional probability of response given covariates.

MAR and the propensity score

- Let Y_i and R_i be random variables with $\text{Supp}[R_i] = \{0, 1\}$.

MAR and the propensity score

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$ and let X_i be a random vector.

MAR and the propensity score

- Let Y_i and R_i be random variables with $\text{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$ and let X_i be a random vector.
- If Y_i is MAR conditional on X_i, then:

MAR and the propensity score

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$ and let X_i be a random vector.
- If Y_i is MAR conditional on X_i, then:
  - $R_i \perp X_i \mid p_R(X_i)$ (balance conditional on $p_R(X_i)$).

MAR and the propensity score

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$ and let X_i be a random vector.
- If Y_i is MAR conditional on X_i, then:
  - $R_i \perp X_i \mid p_R(X_i)$ (balance conditional on $p_R(X_i)$).
  - $Y_i \perp R_i \mid p_R(X_i)$ (independence of outcome and response conditional on $p_R(X_i)$).

MAR and the propensity score

- Let Y_i and R_i be random variables with $\mathrm{Supp}[R_i] = \{0, 1\}$.
- Let $Y_i^* = Y_i R_i + (-99)(1 - R_i)$ and let X_i be a random vector.
- If Y_i is MAR conditional on X_i, then:
  - $R_i \perp X_i \mid p_R(X_i)$ (balance conditional on $p_R(X_i)$).
  - $Y_i \perp R_i \mid p_R(X_i)$ (independence of outcome and response conditional on $p_R(X_i)$).
  - $\exists \varepsilon > 0$ such that $\varepsilon < \Pr[R_i = 1 \mid p_R(X_i)] < 1 - \varepsilon$.

Science table: MAR and the propensity score

i	$X_{[1]i}$	$X_{[2]i}$	$Y_i(0)$	R_i	$Y_i^*(0)$
1	A	0	0	1	0
2	A	0	?	0	-99
3	B	0	1	1	1
4	B	0	?	0	-99
5	A	1	0	1	0
6	A	1	?	0	-99
7	B	1	1	1	1
8	B	1	?	0	-99
9	A	0	1	1	1
10	B	1	?	0	-99

Science table: add the (empirical) propensity score

i	$X_{[1]i}$	$X_{[2]i}$	$\widehat{p}_R(X_i)$	$Y_i(0)$	R_i	$Y_i^*(0)$
1	A	0	0.67	0	1	0
2	A	0	0.67	?	0	-99
3	B	0	0.50	1	1	1
4	B	0	0.50	?	0	-99
5	A	1	0.50	0	1	0
6	A	1	0.50	?	0	-99
7	B	1	0.33	1	1	1
8	B	1	0.33	?	0	-99
9	A	0	0.67	1	1	1
10	B	1	0.33	?	0	-99

$$\widehat{p}_R(X_i) = \frac{\sum_{i=1}^{n} R_i \mathbb{1}(X_{[1]i}, X_{[2]i})}{\sum_{i=1}^{n} \mathbb{1}(X_{[1]i}, X_{[2]i})}$$

$X_{[1]}$	$X_{[2]}$	n	$\widehat{p}_R$
A	0	3	0.67
A	1	2	0.50
B	0	2	0.50
B	1	3	0.33

Estimation with the propensity score

$$\widehat{\mathrm{E}}[Y_i(0)] = \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.67] + \frac{4}{10}\widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.50] + \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.33]$$

$$\widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.67] = \frac{1}{2}, \quad \widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.50] = \frac{1}{2},$$

$$\widehat{\mathrm{E}}[Y_i \mid R_i = 1, \widehat{p}_R(X_i) = 0.33] = 1$$

$$= \frac{3}{10} \cdot \frac{1}{2} + \frac{4}{10} \cdot \frac{1}{2} + \frac{3}{10} \cdot 1 = 0.65$$

Treatment Propensity Function

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}[D_i] = \{0, 1\}$.

Treatment Propensity Function

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

Treatment Propensity Function

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- The treatment propensity function is:

Treatment Propensity Function

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

- The treatment propensity function is:
$$p_D(x) = \Pr[D_i = 1 \mid X_i = x], \quad \forall x \in \mathrm{Supp}\,[X_i].$$

- The random variable $p_D(X_i)$ is the (treatment) propensity score for unit i.

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\left[D_i\right] = \{0, 1\}$.

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\left[D_i\right] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\text{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.

- If D_i is strongly ignorable conditional on X_i, then:

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.
- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.
- If D_i is strongly ignorable conditional on X_i, then:
  - $D_i \perp X_i \mid p_D(X_i)$ (balance conditional on $p_D(X_i)$).

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}\,[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.

- If D_i is strongly ignorable conditional on X_i, then:
  - $D_i \perp X_i \mid p_D(X_i)$ (balance conditional on $p_D(X_i)$).
  - $(Y_i(0), Y_i(1)) \perp D_i \mid p_D(X_i)$ (conditional independence).

Strong Ignorability and the propensity score

- Let $Y_i(0)$, $Y_i(1)$, and D_i be random variables with $\mathrm{Supp}[D_i] = \{0, 1\}$.

- Let $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$ and $\tau_i = Y_i(1) - Y_i(0)$.

- If D_i is strongly ignorable conditional on X_i, then:
  - $D_i \perp X_i \mid p_D(X_i)$ (balance conditional on $p_D(X_i)$).
  - $(Y_i(0), Y_i(1)) \perp D_i \mid p_D(X_i)$ (conditional independence).
  - $\exists \varepsilon > 0$ such that $\varepsilon < \Pr[D_i = 1 \mid p_D(X_i)] < 1 - \varepsilon$.

Science table: strong ignorability and the propensity score

i	$X_{[1]i}$	$X_{[2]i}$	$Y_i(0)$	$Y_i(1)$	D_i	Y_i
1	A	0	0	?	0	0
2	A	0	?	1	1	1
3	B	0	1	?	0	1
4	B	0	?	1	1	1
5	A	1	0	?	0	0
6	A	1	?	1	1	1
7	B	1	1	?	0	1
8	B	1	?	0	1	0
9	A	0	1	?	0	0
10	B	1	?	1	1	1

Science table: add the (empirical) propensity score

i	$X_{[1]i}$	$X_{[2]i}$	$p_D(X_i)$	$Y_i(0)$	$Y_i(1)$	D_i	Y_i
1	A	0	0.33	0	?	0	0
2	A	0	0.33	?	1	1	1
3	B	0	0.50	1	?	0	1
4	B	0	0.50	?	1	1	1
5	A	1	0.50	0	?	0	0
6	A	1	0.50	?	1	1	1
7	B	1	0.67	1	?	0	1
8	B	1	0.67	?	0	1	0
9	A	0	0.33	1	?	0	0
10	B	1	0.67	?	1	1	1

$X_{[1]}$	$X_{[2]}$	n	$p_D(X)$
A	0	3	0.33
A	1	2	0.50
B	0	2	0.50
B	1	3	0.67

Estimation with the treatment propensity score

$$\widehat{\mathrm{E}}[Y_i(1)] = \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.33] + \frac{4}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.50] + \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.67]$$

$$= \frac{3}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot \frac{1}{2} = 0.85$$

Estimation with the treatment propensity score

$$\widehat{\mathrm{E}}\left[Y_i(1)\right] = \frac{3}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.33\right] + \frac{4}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.50\right] + \frac{3}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.67\right]$$

$$= \frac{3}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot \frac{1}{2} = 0.85$$

$$\widehat{\mathrm{E}}\left[Y_i(0)\right] = \frac{3}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.33\right] + \frac{4}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.50\right] + \frac{3}{10}\widehat{\mathrm{E}}\left[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.67\right]$$

$$= \frac{3}{10} \cdot 0 + \frac{2}{10} \cdot 0 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot 1 = 0.5$$

Estimation with the treatment propensity score

$$\widehat{\mathrm{E}}[Y_i(1)] = \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.33] + \frac{4}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.50] + \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 1, \widehat{p}_D(X_i) = 0.67]$$

$$= \frac{3}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot \frac{1}{2} = 0.85$$

$$\widehat{\mathrm{E}}[Y_i(0)] = \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.33] + \frac{4}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.50] + \frac{3}{10}\widehat{\mathrm{E}}[Y_i \mid D_i = 0, \widehat{p}_D(X_i) = 0.67]$$

$$= \frac{3}{10} \cdot 0 + \frac{2}{10} \cdot 0 + \frac{2}{10} \cdot 1 + \frac{3}{10} \cdot 1 = 0.5$$

$$\widehat{\mathrm{E}}[\tau_i] = 0.85 - 0.5 = 0.35$$

Post-treatment variables

- Conditioning on post-treatment variables can break ignorability.

Post-treatment variables

- Conditioning on post-treatment variables can break ignorability.
- A variable affected by treatment is not a valid adjustment set.

Post-treatment variables

- Conditioning on post-treatment variables can break ignorability.
- A variable affected by treatment is not a valid adjustment set.
- DAG intuition: conditioning on descendants can open biasing paths.

Post-treatment variables

- Conditioning on post-treatment variables can break ignorability.
- A variable affected by treatment is not a valid adjustment set.
- DAG intuition: conditioning on descendants can open biasing paths.
- Preview: we will formalize this with DAGs in Weeks 3–4.

Post-treatment variables: why independence fails

- Example: $D \rightarrow M \rightarrow Y$ with M a post-treatment mediator.

Post-treatment variables: why independence fails

- Example: $D \to M \to Y$ with M a post-treatment mediator.

- Even if $D \perp (Y(0), Y(1))$ (random assignment), $D \not\perp M$ because M is caused by D.

Post-treatment variables: why independence fails

- Example: $D \to M \to Y$ with M a post-treatment mediator.
- Even if $D \perp (Y(0), Y(1))$ (random assignment), $D \not\perp M$ because M is caused by D.
- Conditioning on M forces treated/control units to have the same mediator value, which mixes causal pathways and changes the estimand.

Post-treatment variables: why independence fails

- Example: $D \to M \to Y$ with M a post-treatment mediator.
- Even if $D \perp (Y(0), Y(1))$ (random assignment), $D \not\perp M$ because M is caused by D.
- Conditioning on M forces treated/control units to have the same mediator value, which mixes causal pathways and changes the estimand.

$$\Pr[M = 1 \mid D = 1] \neq \Pr[M = 1 \mid D = 0] \quad \Rightarrow \quad D \not\perp M.$$

Generalized Potential Outcomes Model

- Allow multi-valued treatments: $D_i \in \mathrm{Supp}\left[D_i\right]$.

Generalized Potential Outcomes Model

- Allow multi-valued treatments: $D_i \in \mathrm{Supp}[D_i]$.
- Generalized potential outcomes: $Y_i(d)$ for each $d \in \mathrm{Supp}[D_i]$.

Generalized Potential Outcomes Model

- Allow multi-valued treatments: $D_i \in \mathrm{Supp}\left[D_i\right]$.
- Generalized potential outcomes: $Y_i(d)$ for each $d \in \mathrm{Supp}\left[D_i\right]$.
- Switching equation: $Y_i = \sum_{d \in \mathrm{Supp}\left[D_i\right]} Y_i(d)\, \mathbb{1}(D_i = d)$.

Generalized Potential Outcomes Model

- Allow multi-valued treatments: $D_i \in \mathrm{Supp}\left[D_i\right]$.
- Generalized potential outcomes: $Y_i(d)$ for each $d \in \mathrm{Supp}\left[D_i\right]$.
- Switching equation: $Y_i = \sum_{d \in \mathrm{Supp}\left[D_i\right]} Y_i(d)\, \mathbb{1}(D_i = d)$.
- Preview: generalized propensity scores summarize $\Pr[D_i = d \mid X_i]$.

Average Marginal Causal Effect

- If $Y_i(d)$ is differentiable with respect to d, the **average marginal causal effect** is:

Average Marginal Causal Effect

- If $Y_i(d)$ is differentiable with respect to d, the **average marginal causal effect** is:

$$\mathrm{AMCE} = \mathrm{E}\left[\frac{\partial \mathrm{E}\left[Y_i(D_i) \mid D_i, X_i\right]}{\partial D_i}\right].$$

- Under conditional independence, the CEF is causal, so:

Average Marginal Causal Effect

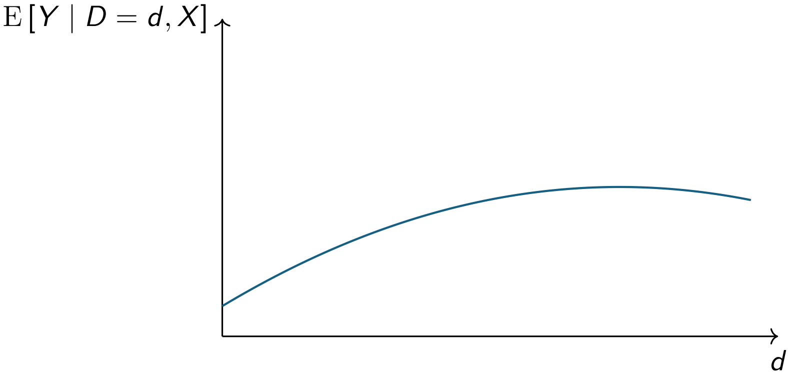
- If $Y_i(d)$ is differentiable with respect to d, the **average marginal causal effect** is:

$$\text{AMCE} = \mathrm{E}\left[\frac{\partial \mathrm{E}\left[Y_i(D_i) \mid D_i, X_i\right]}{\partial D_i}\right].$$

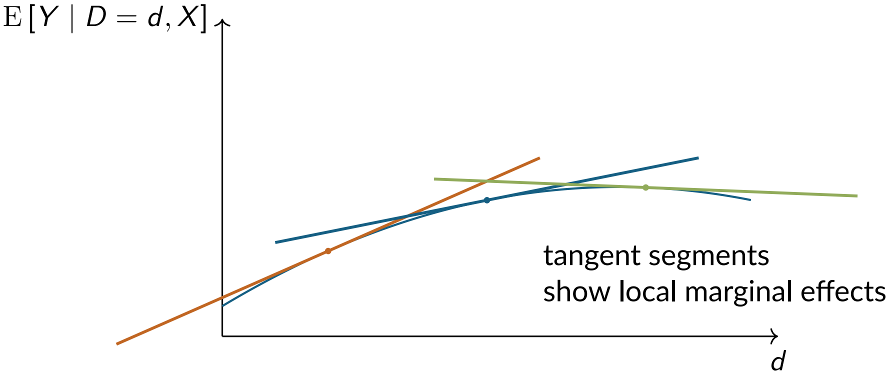
- Under conditional independence, the CEF is causal, so:

$$\text{AMCE} = \mathrm{E}\left[\frac{\partial \mathrm{E}\left[Y_i \mid D_i, X_i\right]}{\partial D_i}\right].$$

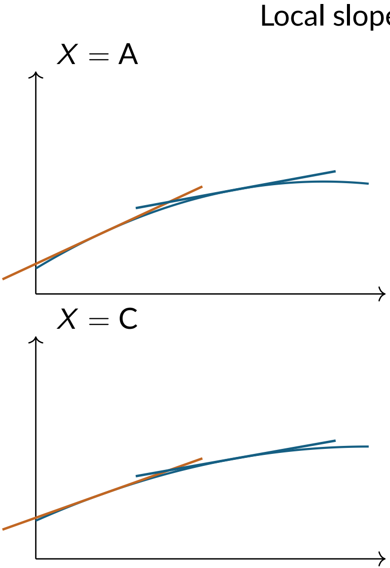
AMCE: local marginal effects



AMCE: local marginal effects



AMCE heterogeneity across covariates



Practice: AMCE with a non-uniform treatment

- Let $D \sim \mathrm{Beta}(2, 2)$ on $(0, 1)$ and $X \sim \mathrm{Bernoulli}(p)$, independent.

Practice: AMCE with a non-uniform treatment

- Let $D \sim \mathrm{Beta}(2, 2)$ on $(0, 1)$ and $X \sim \mathrm{Bernoulli}(p)$, independent.

- Suppose the conditional mean function is

$$m(d, x) \;=\; \alpha \;+\; \beta d^{1/2} \;+\; \gamma d^2 \;+\; \eta x d.$$

Practice: AMCE with a non-uniform treatment

- Let $D \sim \mathrm{Beta}(2,2)$ on $(0,1)$ and $X \sim \mathrm{Bernoulli}(p)$, independent.

- Suppose the conditional mean function is

$$m(d,x) \;=\; \alpha \;+\; \beta d^{1/2} \;+\; \gamma d^2 \;+\; \eta x d.$$

- Compute the average marginal causal effect:

$$\mathrm{AMCE} \;\equiv\; \mathrm{E}\left[\frac{\partial}{\partial d} m(D,X)\right].$$

Practice: AMCE with a non-uniform treatment

- Let $D \sim \mathrm{Beta}(2,2)$ on $(0,1)$ and $X \sim \mathrm{Bernoulli}(p)$, independent.

- Suppose the conditional mean function is

$$m(d,x) \;=\; \alpha \;+\; \beta d^{1/2} \;+\; \gamma d^2 \;+\; \eta x d.$$

- Compute the average marginal causal effect:

$$\mathrm{AMCE} \;\equiv\; \mathrm{E}\left[\frac{\partial}{\partial d} m(D,X)\right].$$

Note: for $D \sim \mathrm{Beta}(2,2)$, the density is $f_D(d) = 6d(1-d)$ on $(0,1)$.

AMCE

Step 1: Differentiate $m(d,x)$ with respect to d.

$$m(d,x) = \alpha + \beta d^{1/2} + \gamma d^2 + \eta x d.$$

Differentiate term-by-term:

$$\frac{\partial}{\partial d} m(d,x) = 0 + \beta \frac{1}{2} d^{-1/2} + 2\gamma d + \eta x = \frac{\beta}{2\sqrt{d}} + 2\gamma d + \eta x.$$

AMCE

Step 1: Differentiate $m(d, x)$ with respect to d.

$$m(d, x) = \alpha + \beta d^{1/2} + \gamma d^2 + \eta x d.$$

Differentiate term-by-term:

$$\frac{\partial}{\partial d} m(d, x) = 0 + \beta \frac{1}{2} d^{-1/2} + 2\gamma d + \eta x = \frac{\beta}{2\sqrt{d}} + 2\gamma d + \eta x.$$

Step 2: Use linearity of expectation.

$$\mathrm{AMCE} = \mathrm{E}\left[\frac{\beta}{2\sqrt{D}} + 2\gamma D + \eta X\right] = \frac{\beta}{2}\mathrm{E}\left[D^{-1/2}\right] + 2\gamma \mathrm{E}\left[D\right] + \eta \mathrm{E}\left[X\right].$$

AMCE

Step 1: Differentiate $m(d, x)$ with respect to d.

$$m(d, x) = \alpha + \beta d^{1/2} + \gamma d^2 + \eta x d.$$

Differentiate term-by-term:

$$\frac{\partial}{\partial d} m(d, x) = 0 + \beta \frac{1}{2} d^{-1/2} + 2\gamma d + \eta x = \frac{\beta}{2\sqrt{d}} + 2\gamma d + \eta x.$$

Step 2: Use linearity of expectation.

$$\mathrm{AMCE} = \mathrm{E}\left[\frac{\beta}{2\sqrt{D}} + 2\gamma D + \eta X\right] = \frac{\beta}{2}\mathrm{E}\left[D^{-1/2}\right] + 2\gamma \mathrm{E}\left[D\right] + \eta \mathrm{E}\left[X\right].$$

Since $X \sim \mathrm{Bernoulli}(p)$,

$$\mathrm{E}\left[X\right] = p.$$

AMCE

Step 3: Compute the needed moments for $D \sim \mathrm{Beta}(2, 2)$.

Compute $\mathrm{E}[D]$. For the Beta distribution,

$$\mathrm{E}[D] = \frac{\alpha}{\alpha + \beta} = \frac{2}{2+2} = \frac{1}{2}.$$

AMCE

Step 3: Compute the needed moments for $D \sim \mathrm{Beta}(2,2)$.

Compute $\mathrm{E}[D]$. For the Beta distribution,

$$\mathrm{E}[D] = \frac{\alpha}{\alpha+\beta} = \frac{2}{2+2} = \frac{1}{2}.$$

Compute $\mathrm{E}[D^{-1/2}]$.

$$\mathrm{E}[D^{-1/2}] = \int_0^1 u^{-1/2} f_D(u)\,du = \int_0^1 u^{-1/2} 6u(1-u)\,du = 6\int_0^1 u^{1/2}(1-u)\,du$$

$$= 6\int_0^1 \left(u^{1/2} - u^{3/2}\right)\,du = 6\left(\int_0^1 u^{1/2}\,du - \int_0^1 u^{3/2}\,du\right).$$

Using $\int_0^1 u^a\,du = \frac{1}{a+1}$ for $a > -1$:

$$\int_0^1 u^{1/2}\,du = \frac{2}{3}, \qquad \int_0^1 u^{3/2}\,du = \frac{2}{5}.$$

Thus,

$$\mathrm{E}[D^{-1/2}] = 6\left(\frac{2}{3} - \frac{2}{5}\right) = \frac{8}{5}.$$

AMCE

Step 4: Substitute back into AMCE.

$$\text{AMCE} = \frac{\beta}{2}\frac{8}{5} + 2\gamma\frac{1}{2} + \eta p = \frac{4}{5}\beta + \gamma + \eta p.$$

$$\boxed{\text{AMCE} = \frac{4}{5}\beta + \gamma + \eta p.}$$

References I

Manski, C. F. and Nagin, D. S. (1998). Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological methodology*, 28(1):99–137.