Review Section

Observational Studies

Kentaro Nakamura

GOV 2002

December 8th, 2025

Today's Agenda

- Review of Mathematical Tools
  - Probability
    - Independence
    - Law of Total Probability
    - Bayes Rule
- Review of Class Materials

Review of Mathematical Tools: Probability (1)

- **Conditional Probability**:

$$\mathbb{P}(X_i = x \mid Y_i = y) = \frac{\mathbb{P}(X_i = x, Y_i = y)}{\mathbb{P}(Y_i = y)}$$

- **Law of Total Probability**:

$$\mathbb{P}(X_i = x) = \sum_y \mathbb{P}(X_i = x \mid Y_i = y)\, \mathbb{P}(Y_i = y)$$

  - This tells us that if you know joint probability $\mathbb{P}(X_i = x, Y_i = y)$, then you can calculate the marginal probability $\mathbb{P}(X_i = x)$ or $\mathbb{P}(Y_i = y)$ by

$$\mathbb{P}(X_i = x) = \sum_y \mathbb{P}(X_i = x, Y_i = y)$$

  - Also, if you know conditional probability $\mathbb{P}(X_i = x \mid Y_i = y)$ and one marginal probability $\mathbb{P}(Y_i = y)$, you can calculate the other marginal probability $\mathbb{P}(X_i = x)$ by

$$\mathbb{P}(X_i = x) = \sum_y \mathbb{P}(X_i = x \mid Y_i = y)\, \mathbb{P}(Y_i = y)$$

Review of Mathematical Tools: Probability (2)

- **Bayes Rule**:

$$\mathbb{P}(X_i = x \mid Y_i = y) = \frac{\mathbb{P}(Y_i = y \mid X_i = x)\mathbb{P}(X_i = x)}{\mathbb{P}(Y_i = y)}$$

- Be familiar with them!
  - Practice final question 2 requires them!

Review of Class Materials

- Quasi-Experimental Design for Observational Data
  - Selection on Observable
  - Regression Discontinuity Design (RDD)
  - Panel Data
    - Difference-in-Difference / Synthetic Control
    - Time-varying treatment / Mediation
  - (Instrumental Variable)

- Different Estimation Strategies
  - Outcome regression
  - Matching
  - Weighting
  - Doubly Robust Estimation

- Robustness Check
  - Sensitivity analysis
  - Partial Identification

Estimand, Identification, Estimation

- **Estimand**: Quantity of Interest / Target Parameter
  - In this class, we are interested in counterfactual
  - Example:
    - ATE: $\mathbb{E}[Y_i(1) - Y_i(0)]$
    - ATT: $\mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1]$
    - ATC: $\mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 0]$
    - CDE: $\mathbb{E}[Y_i(T_i = t, M_i = m) - Y_i(T_i = t', M_i = m)]$
- **Identification**: Write down your estimand with respect to observed data law
  - Example: Under conditional ignorability, ATE is identified as

$$\tau = \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 1, X_i]] - \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 0, X_i]]$$

- **Estimation**: Propose the estimator (the law you can calculate from the data)
  - Example: If you have outcome model $\hat{\mu}(T_i, X_i)$, then your estimator for ATE is

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^{N} \{\hat{\mu}(1, X_i) - \hat{\mu}(0, X_i)\}$$

Selection on Observable (1)

- Assumptions
  - **Conditional ignorability** given confounder X_i:

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

  - Positivity: $0 < \mathbb{P}(T_i = 1 \mid X_i = x) < 1$ for any $X_i = x$
  - Consistency: $Y_i(T_i) = Y_i$ for all i

- We can identify average treatment effect nonparametrically:

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

Selection on Observable (2)

- When estimand is ATT,

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

- We can do the same for ATC, too

Regression Discontinuity Design

- **Setup**:
  - $T_i \in \{0, 1\}$: Treatment
  - X_i: **Running variable** that perfectly determines the value of T_i with the cutpoint c

$$T_i = \mathbf{1}\{X_i \geq c\} = \begin{cases} 1 & \text{if } X_i \geq c \\ 0 & \text{if } X_i < c \end{cases}$$

- **Estimand**: Average treatment effect **on the threshold**

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c]$$

- **Assumption**: $\mathbb{E}[Y_i(t) \mid X_i = x]$ is continuous in x at $X_i = c$ for $t = 0, 1$
  - Continuity $\rightarrow$ Does not change abruptly
  - Formally, $\lim_{x \to c} \mathbb{E}[Y_i(t) \mid X_i = x] = \lim_{x \leftarrow c} \mathbb{E}[Y_i(t) \mid X_i = x]$
  - Example of violation (sorting): students strategically retaking an exam to just exceed a scholarship cutoff
  - Barely below and above the cutoff is no longer as-if random

Sharp RDD: Identification

- Now, the estimand is $\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = c]$
- Then, for $T_i = 1$

$$\begin{aligned}\mathbb{E}[Y_i(1) \mid X_i = c] &= \lim_{x \leftarrow c} \mathbb{E}[Y_i(1) \mid X_i = x] \quad (\because \text{continuity}) \\ &= \lim_{x \leftarrow c} \mathbb{E}[Y_i \mid X_i = x] \quad (\because \text{consistency})\end{aligned}$$

- Similariy, for $T_i = 0$

$$\mathbb{E}[Y_i(0) \mid X_i = c] = \lim_{x \to c} \mathbb{E}[Y_i(0) \mid X_i = x] = \lim_{x \to c} \mathbb{E}[Y_i \mid X_i = x]$$

- Therefore,

$$\tau = \underbrace{\lim_{x \downarrow c} \mathbb{E}[Y_i \mid X_i = x]}_{= \mathbb{E}[Y_i(1) | X_i = c]} - \underbrace{\lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]}_{= \mathbb{E}[Y_i(0) | X_i = c]}$$

Panel Data Analysis: Difference-in-Difference (1)

- **Setup** (for the two time period):
  - G_i: treatment indicator ($G_i = 1$ for treatment group)
  - $D_{it} = tG_i$: treatment assignment indicator
  - Y_{it}: observed outcome for unit i at time t
  - $Y_{it}(d)$: potential outcome for unit i at time t
- **Estimand**: Average treatment effect for the treated (ATT)

$$\tau = \mathbb{E}[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1]$$

- Assumption: **Parallel trend**

$$\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1] = \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0]$$

  - We also assume **no anticipation** assumption

$$Y_{i0}(1) = Y_{i0}(0)$$

Panel Data Analysis: Difference-in-Difference (2)

$$
\begin{aligned}
&\{\mathbb{E}[Y_{i1} \mid G_i = 1] - \mathbb{E}[Y_{i1} \mid G_i = 0]\} - \{\mathbb{E}[Y_{i0} \mid G_i = 1] - \mathbb{E}[Y_{i0} \mid G_i = 0]\} \\
&= \{\mathbb{E}[Y_{i1}(1) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 0]\} \\
&\qquad - \{\mathbb{E}[Y_{i0}(0) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 0]\} \\
&= \underbrace{\mathbb{E}[Y_{i1}(1) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 1]}_{= \tau_{ATT}} + \mathbb{E}[Y_{i1}(0) \mid G_i = 1] \\
&\qquad - \mathbb{E}[Y_{i1}(0) \mid G_i = 0] - \mathbb{E}[Y_{i0}(0) \mid G_i = 1] + \mathbb{E}[Y_{i0}(0) \mid G_i = 0] \\
&= \tau_{ATT} + \underbrace{\Big(\mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0]\Big)}_{=0 \text{ under parallel trends}} \\
&= \tau_{ATT}.
\end{aligned}
$$

Causal Mediation Analysis / Time-varying treatment (1)

- In the case of DiD / SCM, we care about treatment at only one point
  - We might want to consider the treatment at time 1 and 2 (i.e., $Y_i(T_1 = t_1, T_2 = t_2)$)
  - This is connected to causal mediation analysis

- **Estimand**

$$\begin{aligned}
\text{Controlled Direct Effect}: \quad & \bar{\xi}(m) = \mathbb{E}[Y_i(1, m) - Y_i(0, m)] \\
\text{Natural Indirect Effect}: \quad & \bar{\delta}(m) = \mathbb{E}[Y_i(t, M_i(1)) - Y_i(t, M_i(0))] \\
\text{Natural Direct Effect}: \quad & \bar{\zeta}(m) = \mathbb{E}[Y_i(1, M_i(t)) - Y_i(0, M_i(t))]
\end{aligned}$$

Causal Mediation Analysis / Time-varying treatment (2)

- Assumptions for CDE: **Sequential Ignorability**

$$\{Y_i(t,m), M_i(t')\} \perp\!\!\!\perp T_i \mid X_i \quad \text{(Treatment Uncounfoundedness)}$$
$$Y_i(t,m) \perp\!\!\!\perp M_i \mid X_i = x, T_i, Z_i \quad \text{(Mediator Uncounfoundedness)}$$

- Assumptions for NIE / NDE

$$\{Y_i(t,m), M_i(t')\} \perp\!\!\!\perp T_i \mid X_i$$
$$Y_i(t',m) \perp\!\!\!\perp M_i(t) \mid X_i = x, T_i \quad \text{(Cross-world Counterfactual)}$$

  - Importantly, we cannot have Z_i for NIE / NDE

- Problem of Mediation / Time-varying Treatment: Post-treatment bias
  - Mediator is by definition post-treatment
  - For CDE, the confounder for mediator can be post-treatment

Identification of CDE

$$
\begin{aligned}
&\mathbb{E}[Y(t,m)] \\
&= \mathbb{E}[\mathbb{E}[Y(t,m) \mid X_i]] \\
&= \mathbb{E}[\mathbb{E}[Y(t,m) \mid T_i = t, X_i]] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(t,m) \mid T_i = t, X_i, Z_i] \mid T_i = t, X_i]] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y(t,m) \mid T_i = t, X_i, Z_i, M_i = m] \mid T_i = t, X_i]] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[Y \mid T_i = t, X_i, Z_i, M_i = m] \mid T_i = t, X_i]]
\end{aligned}
$$

- Make sure you understand each step!

Identification of NDE / NIE

$$
\begin{aligned}
&\mathbb{E}[Y(t, M(t'))|X] \\
&= \sum_m \mathbb{E}[Y(t,m) \mid X, M(t') = m]\mathbb{P}(M(t') = m \mid X) \quad (\because \text{L.I.E.}) \\
&= \sum_m \mathbb{E}[Y(t,m) \mid X, M(t') = m, T = t']\mathbb{P}(M(t') = m \mid X) \\
&= \sum_m \mathbb{E}[Y(t,m) \mid X, T = t']\mathbb{P}(M(t') = m \mid X) \\
&= \sum_m \mathbb{E}[Y(t,m) \mid X, T = t]\mathbb{P}(M(t') = m \mid X, T = t') \\
&= \sum_m \mathbb{E}[Y(t,m) \mid X, T = t, M(t) = m]\mathbb{P}(M(t') = m \mid X, T = t') \\
&= \sum_m \mathbb{E}[Y \mid X, T = t, M = m]\mathbb{P}(M = m \mid X, T = t')
\end{aligned}
$$

- Make sure you understand each step!

Estimation: Outcome Regression

- Based on the identification formula, we propose the estimation strategies
- Strategy 1: Outcome regression, such as

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

- Example 1 (ATE): The identification formula of ATE is given by

$$\tau_{\mathrm{ATE}} = \mathbb{E}\Big[\mathbb{E}[Y_i \mid T_i = 1, X_i] - \mathbb{E}[Y_i \mid T_i = 0, X_i]\Big]$$

  - We can estimate each $\mathbb{E}[Y_i \mid T_i = 1, X_i = x]$ using regression
- Example 2 (ATT): Based on identification formula,

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

- But outcome model depends on modeling assumption in the case

Matching

- 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

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

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

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

  - 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 is used in many places, including panel data (panel match)

Weighting

- 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)

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

- Weighting is also used in other settings, including mediation and DiD
  - Make sure that you can derive weighting estimator for each setting

Doubly Robust Estimation (1)

- We learn two approaches to estimate causal effect: outcome model and weighting

$$\mathbb{E}[Y_i(1) - Y_i(0)]$$
$$= \begin{cases} \mathbb{E}[\mathbb{E}[Y_i \mid T_i = 1, X_i] - \mathbb{E}[Y_i \mid T_i = 0, X_i]] \text{ (Outcome)} \\ \mathbb{E}\left[\frac{T_i Y_i}{\pi(X_i)} - \frac{(1-T_i)Y_i}{1-\pi(X_i)}\right] \quad \text{(weighting)} \end{cases}$$

- **Doubly Robust Estimator / Augmented IPW (AIPW)**: Combine weighting (IPW) with outcome model so that if either works, we can estimate causal effect

$$\hat{\tau}_{\mathrm{AIPW}} = \frac{1}{n}\sum_{i=1}^{n}\Big(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)\Big) + \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{T_i(Y_i - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \frac{(1 - T_i)(Y_i - \hat{\mu}_0(X_i))}{1 - \hat{\pi}(X_i)}\Big)$$

Doubly Robust Estimation (2)

- **Proof Strategy**:
  - Check the following two cases separately
  - (1) correct outcome model: replace $\hat{\mu}_t(X_i)$ with $\mathbb{E}[Y_i \mid T_i = t, X_i]$
  - (2) correct propensity score model: replace $\hat{\pi}(X_i)$ with $\mathbb{E}[T_i \mid X_i]$

- Try Problem Set 8 Question 2 for Stat286
  - Also, try practice final Question 3 (panel version of doubly robust estimator)
  - Try to show doubly robust estimator for mediation

Proof of Double Robustness

- We only prove that the AIPW of $\mathbb{E}[Y_i(1)]$ part is unbiased if either propensity score model or outcome model is correctly specified.

$$
\begin{aligned}
\text{Bias} := & \; \mathbb{E}\left[\hat{\mu}_1(X_i) + \frac{T_i(Y_i - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)}\right] - \mathbb{E}[Y_i(1)] \\
= & \; \mathbb{E}\left[\frac{T_i(Y_i - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \Big(Y_i(1) - \hat{\mu}_1(X_i)\Big)\right] \\
= & \; \mathbb{E}\left[\frac{\mathbb{E}[T_i Y_i \mid X_i] - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \Big(\mathbb{E}[Y_i(1) \mid X_i] - \hat{\mu}_1(X_i)\Big)\right] \quad \text{(L.I.E)} \\
= & \; \mathbb{E}\left[\frac{\mathbb{E}[T_i Y_i(1) \mid X_i] - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \Big(\mathbb{E}[Y_i(1) \mid X_i] - \hat{\mu}_1(X_i)\Big)\right] \\
= & \; \mathbb{E}\left[\frac{\mathbb{E}[T_i \mid X_i]\mathbb{E}[Y_i(1) \mid X_i] - \hat{\mu}_1(X_i))}{\hat{\pi}(X_i)} - \Big(\mathbb{E}[Y_i(1) \mid X_i] - \hat{\mu}_1(X_i)\Big)\right] \\
= & \; \mathbb{E}\left[\left(\frac{\mathbb{E}[T_i \mid X_i]}{\hat{\pi}(X_i)} - 1\right)\Big(\mathbb{E}[Y_i(1) \mid X_i] - \hat{\mu}_1(X_i)\Big)\right] \\
= & \; \mathbb{E}\left[\left(\frac{\mathbb{E}[T_i \mid X_i]}{\hat{\pi}(X_i)} - 1\right)\Big(\mathbb{E}[Y_i \mid T_i = 1, X_i] - \hat{\mu}_1(X_i)\Big)\right]
\end{aligned}
$$

Approaches for Robustness Check

- In observational studies, these assumptions can often be violated

- Approach 1: Sensitivity Analysis
  - The goal is still point identification
  - Ask how the point estimate changes if assumptions are violated to the certain extent
  - Regression-based approach (partial R^2)
  - Risk-based approach (cornfield condition)
  - Check section slide Module 6.5 for the derivation

- Approach 2: Partial Identification
  - How much can we know with the minimal amount of assumptions we are willing to make?
  - Try problem set 7 again to check your understanding

- Approach 3: Modeling selection bias
  - Heckman's selection model (see recording of Module 7)
  - Note that this is based on the strong model assumption

Partial Identification: Case of Binary Outcome (1)

- Let's think about the binary outcome and treatment. We have the following principal strata:

$$(Y_i(0), Y_i(1)) \in \{(0,0), (0,1), (1,0), (1,1)\}$$

- Suppose that we want to assign treatment to maximize the effect $Y_i(1) - Y_i(0)$
  - That is, assigning treatment to the strata $(Y_i(0), Y_i(1)) = (0,1)$ and not assigning to the strata $(Y_i(0), Y_i(1)) = (1,0)$
  - The only people whose outcome is 0 is those in strata $(Y_i(0), Y_i(1)) = (0,0)$
- **Question**: How can we maximize the outcome value by optimizing the treatment assignment?
- If we optimally assign the treatment effect, the observed outcome δ will be

$$\begin{aligned}\delta :=& 1 \times \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1) \ + 1 \times \mathbb{P}(Y_i(0) = 0, Y_i(1) = 1) \\ & + 1 \times \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) + 0 \times \mathbb{P}(Y_i(0) = 0, Y_i(1) = 0)\end{aligned}$$

Partial Identification: Case of Binary Outcome (2)

- Now,

$$\mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(0) = 0, Y_i(1) = 1) + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1)$$
$$\underbrace{\mathbb{P}(Y_i(0) = 1)}_{\text{Identifiable}} = \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1)$$

  but we do not observe the probability of each principal strata.

- But we know that

$$\delta = \underbrace{\mathbb{P}(Y_i(1) = 1)}_{\text{Identifiable}} + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$$

  so we need to think about how to maximize $\mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$

Partial Identification: Case of Binary Outcome (3)

- Let's write down all the constraints:
  - Firstly, each probability is bounded between 0 and 1
  - Then, we can identify $\mathbb{P}(Y_i(1) = 1)$ and $\mathbb{P}(Y_i(0) = 1)$

- In this case, each strata probability can be written as observed quantity and $\mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$. I.e.,

$$0 \leq \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) \leq 1$$
$$0 \leq \underbrace{\mathbb{P}(Y_i = 1 \mid T_i = 0) - \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)}_{=\mathbb{P}(Y_i(0)=1, Y_i(1)=1)} \leq 1$$

- You can also derive $\mathbb{P}(Y_i(0) = 0, Y_i(1) = 1)$ and $\mathbb{P}(Y_i(0) = 0, Y_i(1) = 0)$

- Under these constraints, think about how much you can maximize the quantity of interest.

Linear programming

- **Optimization problem** contains two components:
  - Objective function: the function to minimize / maximize
  - Constraints that solution need to satisfy

- Standard approach: transform the optimization problem to the specific form so that solver can solve automatically

- **Linear programming**: One form of optimization problem that can be easily solved by solver
  - Both constraint and objective function are linear

$$\max_x \quad c^\top x$$
$$\text{such that} \quad x \geq 0, Ax \leq b$$

Lee's bounds (1)

- If outcome is not binary, the previous approach does not work
  - This is exactly the setting of the practice final question 1
  - Let's review how we can deal with the continuous case together

- **Setup**
  - X_i: self-reported income
  - $T_i = \mathbf{1}\{X_i \geq c\}$: treatment indicator
    - whether the household is eligible for the program (c is threshold)
  - Y_i: outcome of interest (continuous)
  - M_i: misreporting status

- **Assumption**:
  - $\mathbb{E}[Y_i(t) \mid X_i = x, M_i = 0]$ is continuous
    - I.e., among those who do not misreport, continuity holds
  - If $X_i < c$, then $M_i = 0$
    - No units with $X_i < c$ are manipulators

Lee's bounds (2)

- Let $f_{X,M}(x, m)$ be the joint density of $X = x, M = m$.
- Above the cutoff, we have mixture of manipulators and non-manipulators

$$f_+(c) := \lim_{x \uparrow c} f_X(x) = f_{X,M}(c, 0) + f_{X,M}(c, 1)$$

while below the cutoff we only have the nonmanpulators

$$f_-(c) := \lim_{x \downarrow c} f_X(x) = f_{X,M}(c, 0)$$

- Therefore,

$$\mathbb{P}(M_i = 1 \mid X_i = c) = \frac{f_{X,C}(c, 1)}{f_+(c)} = \frac{f_+(c) - f_-(c)}{f_+(c)}$$

Lee's bounds (3)

- Now, let's think about the bounds of LATE

$$\mathbb{E}\big[Y_i(1) - Y_i(0) \mid X_i = c, M_i = 0\big]$$

- As everyone below the cutoff is non-manipulator, by continuity

$$\mathbb{E}\big[Y_i(0) \mid X_i = c, M_i = 0\big] = \lim_{x \uparrow c} \mathbb{E}[Y_i \mid X_i = x]$$

  which is point identified.

- We thus need to bound

$$\mathbb{E}\big[Y_i(1) \mid X_i = c, M_i = 0\big]$$

  since just above the cutoff, we have mixture of manipulators and nonmanipulators

Lee's bounds (4)

- However, now we know how many people are manipulators at the cutoff; i.e.,

$$p = \mathbb{P}(M_i = 1 \mid X_i = c) = \frac{f_+(c) - f_-(c)}{f_+(c)}$$

which means that $1 - p$ people are non-manipulators

- We also observe the distribution of outcomes
- **Idea**: Think about how to allocate these $1 - p$ people
    - If everyone is at the bottom of outcome distribution, we then obtain the lower bounds
    - If everyone is at the top of outcome distribtuion, we then obtain the upper bounds
    - Formally, with quantile function $Q^+(u) = \inf\{y : F^+(y) \geq u\}$

$$\underline{\mu}_1 = \frac{1}{1-p} \int_0^{1-p} Q^+(u)\, du, \qquad \overline{\mu}_1 = \frac{1}{1-p} \int_p^1 Q^+(u)\, du$$