Review Section

Causal Inference in Experiment (First Half)

Kentaro Nakamura

GOV 2002

October 8th, 2025

Today's Agenda

- Review of Mathematical Tools
  - Expectation
  - Variance
- Review of Class Materials

Review of Mathematical Tools: Expectation

- Rule 1: For constant c and random variable X, $\mathbb{E}[cX] = c\mathbb{E}[X]$
- Rule 2: For random variable X and Y, $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- Rule 3: If X and Y are independent, $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$
- Rule 4: **Law of Iterated Expectation**

$$\mathbb{E}[X] = \mathbb{E}\Big[\mathbb{E}[X \mid Y]\Big]$$

- **Recommendations**
  - Check which variable is random and which is not
  - Always think about how to use law of iterated expectation

Review of Mathematical Tools: Variance

- Rule 1: For constant c and random variable X,

$$\mathbb{V}[cX] = c^2\mathbb{V}[X]$$

- Rule 2: For random variable X and Y,
  $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathrm{Cov}(X, Y)$
  - If X and Y are independent, $\mathrm{Cov}(X, Y) = 0$

- Rule 3: Alternative representation

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- Rule 4: **Law of Total Variance**

$$\mathbb{V}[X] = \mathbb{V}[\mathbb{E}[X \mid Y]] + \mathbb{E}[\mathbb{V}[X \mid Y]]$$

- **Recommendations**
  - Check which variable is random and which is not
  - Check which variable is independent of other variable (Rule 2)
  - Always think about how to use rule 3 (typically easier to use)

Review of Class Materials

- Assumptions for Causal Inference
- Two frameworks
  - Finite-Population: Fisher's Permutation Test, SATE (Neyman)
  - Super-Population: PATE (Neyman)
- Better Experiments
  - Efficiency improvement with covariates
    - Block randomization, Matched-pair design
    - Regression (Post-Stratification)
  - Interference
    - Cluster Randomized Experiment
    - Conditional Randomization Test
  - Non-Compliance
    - Instrumental Variable

Assumptions for Causal Inference

- **Randomization** (a.k.a ignorability)[1]: Under complete randomization,

$$Y_i(t) \perp T_i$$

with $\mathbb{P}(T_i = t) > 0$ for all t.

- **Consistency**: $Y_i(t) = Y_i$ when $T_i = t$
  - It excludes interference (your potential outcome is function of other's treatment status)

[1] For block randomization, we assume complete randomization *within* block

Two inference frameworks: Overview

- **within sample inference** (finite-population framework)
  - Given data on n units ($i = 1, \cdots, n$), we are interested in the average treatment effect on that sample
  - Only source of randomness is *treatment assignment*
  - **Fisher's Permutation Test**: talk about individual (sharp null hypothesis)
  - **SATE (Neyman)**: Talk about sample averages

- **population inference** (super-population framework)
  - Generalizing the inference on the obtained sample to some population of interest
  - Source of randomness is (1) *treatment assignment* and (2) *sampling*
  - Example: PATE (Neyman)

Two inference frameworks: Overview

Population
(e.g., all students
in class)

Sampling

Finite
Population
(e.g., students in
section)

Y(1)	Y(0)
10	5
3	7
2	9

Treatment
Assignment

Observed
Data

T	Y
1	10
0	7
1	2

Efficiency Improvements with Covariates

- We often observe pre-treatment covariates $\rightarrow$ We want to use it to improve efficiency

- **Approach 1**: Better design with covariates
  - **Block randomization**: Randomize treatment *within* block
  - **Matched-pair design**: Randomize treatment *within* pair
    - Pair is the smallest block
  - You design experiments with covariates
    - Randomization is not complete randomization

- **Approach 2**: Better inference with regression
  - Use covariates to improve efficiency in inference
    - Randomization can be complete randomization
  - **Question**: Does linearity matter?
    - Regression: $Y_i = \alpha + \beta T + \delta \tilde{X}_i + \epsilon_i$
    - This model is pretty strong!
    - Everyone has same effect β (constant treatment effect)
    - Linearity in control variable

Regression and Causal Inference

- **Question**: Does linearity of regression matter?
  - Does it estimate ATE when violating linearity?
  - Does it improve efficiency compared to diff-in-means when violating linearity?

- **Case 1**: $Y_i = \alpha + \beta T_i + \epsilon_i$
  - Estimate ATE without assuming linearity (no constant treatment effect)
  - Same variance as difference-in-means
- **Case 2**: $Y_i = \alpha + \beta T_i + \boldsymbol{\gamma}^\top \tilde{\mathbf{X}}_i + \epsilon_i$
  - Consistent ATE without assuming linearity (no constant treatment effect)
  - No guarantee for efficiency improvement without correct modeling
- **Case 3**: $Y_i = \alpha + \beta T_i + \boldsymbol{\gamma}^\top \tilde{\mathbf{X}}_i + \boldsymbol{\delta}^\top T_i \tilde{\mathbf{X}}_i + \epsilon_i$
  - Consistent ATE without assuming linearity (no constant treatment effect)
  - Improve efficiency always

Interference

- **Interference**: Your potential outcome is influenced by other's treatment status
  - Under consistency, $Y_i(T_i = t) = Y_i$ when $T_i = t$
  - If there is interference, $Y_i(T_1 = t_1, \cdots, T_n = t_n) \neq Y_i(T_i = t)$

- How to deal with it?
  - **Cluster Randomization**: Design experiments that allow interference
  - **Conditional Randomization Test**: Detect spillover effects

Cluster Randomization

- **Cluster randomized experiment**: assign treatment at the cluster level
  - We allow spillover *within* each cluster
  - We assume no spillover *across* clusters

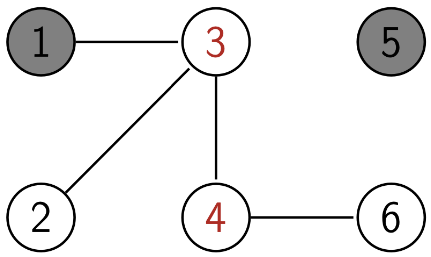
- Because everyone in each cluster is in the same treatment status,

$$\underbrace{Y_{ij}(T_{1j}, \cdots, T_{m_j j})}_{\text{Allowing interference within cluster } j} \qquad = Y_{ij}(T_j)$$

- Inference under cluster randomization
  - Regard each cluster as a unit of analysis
  - Apply Neyman's analysis

Conditional Randomization Test (Example)

- Let's understand the example from the class again
  - Gray units: treated / White units: untreated
  - Focal units: unit 3 and 4



- Question: Are units 3 and 4 affected by the treatment of their friends?

Conditional Randomization Test (Example)

- Null Hypothesis: $Y_3, Y_4 \perp\!\!\!\perp T_1, T_2, T_5, T_6 \mid T_3, T_4$
  - Under the null, given the fixed value of T_3 and T_4, the permutation of T_1, T_2, T_5, T_6 should not affect Y_3, Y_4

- Remark: connection from permutation test
  - Permutation Test: Assume sharp null. Randomization is guaranteed by design
  - CRT: Randomization is justifiable under the null

- **Procedure**
  - STEP 1: Calculate test statistic $\mathrm{Corr}(Y_{\mathrm{focal}}, \bar{T}_{\mathrm{friend}})$ on the observed data
  - STEP 2: Permute the treatment assignment T_1, T_2, T_5, T_6 given the original value of T_3, T_4
  - STEP 3: For each permutation, calculate the test statistic and create the reference distribution
  - STEP 4: Compare the observed test statistic with the reference distribution

Non-Compliance

- Sometimes we cannot directly force treatments
  - We instead give **encouragement** to take the treatment
  - But some units might refuse to take the treatment

- **Method 1**: Intention-to-Treat (ITT) analysis
  - Estimate the effect of encouragement, not treatment

- **Method 2**: Instrumental Variable
  - Use encouragement to identify **Local ATE (LATE)**

Non-Compliance: Complier

- Four principal strata (or compliance types):
- Compliers: $T_i(Z_i = 1) = 1$ and $T_i(Z_i = 0) = 0$
- Always-takers: $T_i(Z_i = 1) = T_i(Z_i = 0) = 1$
- Never-takers: $T_i(Z_i = 1) = T_i(Z_i = 0) = 0$
- Defiers: $T_i(Z_i = 1) = 0$ and $T_i(Z_i = 0) = 1$

	$Z_i = 1$	$Z_i = 0$
$T_i = 1$	Complier / Always-taker	Defier / Always-taker
$T_i = 0$	Defier / Never-taker	Complier / Never-taker

Instrumental Variable

- **Assumption 1 (Randomization)**: Instruments are randomized so that

$$\{Y_i(T_i = t, Z_i = z), T_i(Z_i = z)\} \perp Z_i$$

- **Assumption 2 (Exclusion Restriction)**: Instruments affect outcome only through treatment so that

$$Y_i(T_i = t, Z_i = z) = Y_i(T_i = t)$$

- **Assumption 3 (Monotonicity)**: No defiers
  - **Defiers**: Those who would take treatment when not encouraged, but would not when encouraged

$$T_i(Z_i = 1) \geq T_i(Z_i = 0)$$

Instrumental Variable: Identification (1)

- ITT effect is defined as

$$\mathrm{ITT} := \mathbb{E}[Y_i(Z_i = 1) - Y_i(Z_i = 0)]$$

- By randomization and consistency, we have

$$\mathbb{E}[Y_i(Z_i = z)] = \mathbb{E}[Y_i(Z_i = z) \mid Z_i = z] = \mathbb{E}[Y_i \mid Z_i = z]$$

- Thus, ITT effect is identified as

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

Instrumental Variable: Identification (2)

- Now, notice that

$$Y_i(Z_i = z) = Y_i(T_i = T_i(z), Z_i = z)$$

- This is because $T_i(z)$ is a value of treatment under $Z_i = z$
  - So, $Y_i(T_i = T_i(z), Z_i = z)$ is function of only $Z_i = z$

- So, we can write ITT Effect as

$$\begin{aligned}&\mathbb{E}[Y_i(Z_i = 1, T_i(Z_i = 1)) - Y_i(Z_i = 0, T_i(Z_i = 1))]\\&= \underbrace{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}_{\text{Estimatable from Data (Identified)}}\end{aligned}$$

Instrumental Variable: Identification (3)

- By law of iterated Expectation, we can write

$$\mathbb{E}[Y_i(1, T_i(1)) - Y_i(0, T_i(0))] = \mathbb{E}\Big[\mathbb{E}[\Delta \mid \underbrace{T_i(1), T_i(0)}_{\text{Principal Strata}}]\Big]$$

where $\Delta = Y_i(1, T_i(1)) - Y_i(0, T_i(0))$

- This is equivalent to

$$
\begin{aligned}
&= \mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 0]\,\underbrace{\mathbb{P}(T_i(1) = 1, T_i(0) = 0)}_{\text{Complier}} \\
&\quad + \mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 1]\,\underbrace{\mathbb{P}(T_i(1) = 1, T_i(0) = 1)}_{\text{Always-Taker}} \\
&\quad + \mathbb{E}[\Delta \mid T_i(1) = 0, T_i(0) = 0]\,\underbrace{\mathbb{P}(T_i(1) = 0, T_i(0) = 0)}_{\text{Never-Taker}} \\
&\quad + \mathbb{E}[\Delta \mid T_i(1) = 0, T_i(0) = 1]\,\underbrace{\mathbb{P}(T_i(1) = 0, T_i(0) = 1)}_{\text{Defier}}
\end{aligned}
$$

Instrumental Variable: Identification (4)

- Notice that $\mathbb{P}(T_i(1) = 0, T_i(0) = 1) = 0$ (i.e., no defier)
- Also notice that for always taker,

$$\begin{aligned}
&\mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 1] \\
&= \mathbb{E}[Y_i(Z_i = 1, T_i(1)) - Y_i(Z_i = 0, T_i(0)) \mid T_i(1) = 1, T_i(0) = 1] \\
&= \mathbb{E}[Y_i(Z_i = 1, T_i = 1) - Y_i(Z_i = 0, T_i = 1) \mid T_i(1) = 1, T_i(0) = 1]
\end{aligned}$$

- By exclusion, we have $Y_i(T_i = t, Z_i = z) = Y_i(T_i = t)$. So,

$$= \mathbb{E}[\underbrace{Y_i(T_i = 1) - Y_i(T_i = 1)}_{=0} \mid T_i(1) = 1, T_i(0) = 1] = 0$$

- Same holds for never-taker

Instrumental Variable: Identification (5)

- Hence,

$$
\begin{aligned}
&\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0] \\
&= \mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 0] \underbrace{\mathbb{P}(T_i(1) = 1, T_i(0) = 0)}_{\text{Complier}} \\
&\quad + \underbrace{\mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 1]}_{=0} \mathbb{P}(T_i(1) = 1, T_i(0) = 1) \\
&\quad + \underbrace{\mathbb{E}[\Delta \mid T_i(1) = 0, T_i(0) = 0]}_{=0} \mathbb{P}(T_i(1) = 0, T_i(0) = 0) \\
&\quad + \mathbb{E}[\Delta \mid T_i(1) = 0, T_i(0) = 1] \underbrace{\mathbb{P}(T_i(1) = 0, T_i(0) = 1)}_{=0} \\
&= \mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 0] \underbrace{\mathbb{P}(T_i(1) = 1, T_i(0) = 0)}_{\text{Complier}}
\end{aligned}
$$

Instrumental Variable: Identification (6)

- Finally, notice that

$$
\begin{aligned}
&\mathbb{E}[T_i(1) - T_i(0)] \\
&= \mathbb{E}[T_i(1) \mid Z_i = 1] - \mathbb{E}[T_i(0) \mid Z_i = 0] \quad (\because \text{Randomization}) \\
&= \mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0] \quad (\because \text{Consistency}) \\
&= \mathbb{P}(T_i = 1 \mid Z_i = 1) - \mathbb{P}(T_i = 1 \mid Z_i = 0) \quad (\because T_i \text{ is binary}) \\
&= \mathbb{P}(\text{Always-taker or Complier}) - \mathbb{P}(\text{Always-Taker}) \quad (\because \text{no defier}) \\
&= \mathbb{P}(\text{Complier}) = \mathbb{P}(T_i(1) = 1, T_i(0) = 0)
\end{aligned}
$$

- Therefore,

$$
\underbrace{\mathbb{E}[\Delta \mid T_i(1) = 1, T_i(0) = 0]}_{\text{Late (ATE among complier)}} = \underbrace{\frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[T_i \mid Z_i = 1] - \mathbb{E}[T_i \mid Z_i = 0]}}_{\text{Estimatable from Data}}
$$