Introduction to Causal Inference

Lecture 5

Assignment Mechanism

邓婉璐

WANLUDENG@TSINGHUA.EDU.CN



清华大学统计与数据科学系

Outline

- Review and Inspiring Examples
- Definitions for Assignment Mechanism ★
- Assumptions for Assignment Mechanism ★
- Classification of Assignment Mechanisms ★
- Super Population (Notation, difference vs Finite Sample)
- Example: Lord’s Paradox



Review and Inspiring Examples



Assignment Mechanism (Introduction)

Treatment indicator representing the treatment assigned to unit i

Potential outcomes of unit i

The realized outcome

$$Y_i^{\text{obs}} = Y_i(W_i) = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1. \end{cases}$$

The missed outcome

$$Y_i^{\text{mis}} = Y_i(1 - W_i) = \begin{cases} Y_i(1) & \text{if } W_i = 0, \\ Y_i(0) & \text{if } W_i = 1. \end{cases}$$

The alternative treatment that is not assigned to unit i

$$Y_i(0) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 1, \\ Y_i^{\text{obs}} & \text{if } W_i = 0, \end{cases} \quad \text{and} \quad Y_i(1) = \begin{cases} Y_i^{\text{mis}} & \text{if } W_i = 0, \\ Y_i^{\text{obs}} & \text{if } W_i = 1. \end{cases}$$



An Example of Assignment

A case that W is related to Y

Table 1.4. ***Medical Example with Two Treatments, Four Units, and SUTVA: Surgery (S) and Drug Treatment (D)***

Unit	Potential Outcomes		Causal Effect
	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
Patient #1	1	7	6
Patient #2	6	5	−1
Patient #3	1	5	4
Patient #4	8	7	−1
Average	4	6	2

Table 1.5. ***Ideal Medical Practice: Patients Assigned to the Individually Optimal Treatment; Example from Table 1.4***

Unit i	Treatment W_i	Observed Outcome Y_i^{obs}
Patient #1	1	7
Patient #2	0	6
Patient #3	1	5
Patient #4	0	8

Simpson’s Paradox

A simple case:

Table 2. NAEP 1992 8th Grade Math Scores

	State	White	Black	Other Nonwhite
Nebraska	**277**	281	236	259
New Jersey	**271**	283	242	260





Simpson's Paradox

A case that W is also related to X.

A simple case:

Table 2. NAEP 1992 8th Grade Math Scores

	State	White	Black	Other Nonwhite	Standardized
Nebraska	**277**	281	236	259	**271**
New Jersey	**271**	283	242	260	**273**
		Proportion of population			
Nebraska		87%	5%	8%	
New Jersey		66%	15%	19%	
Nation		69%	16%	15%	

Which order should we use?

Depends on your questions.

$$281 \times 0.69 + 236 \times 0.16 + 259 \times 0.15 = 271.$$

$$283 \times 0.69 + 242 \times 0.16 + 260 \times 0.15 = 273.$$

Definitions for Assignment Mechanism



An Example

Table 1.4. ***Medical Example with Two Treatments, Four Units, and SUTVA: Surgery (S) and Drug Treatment (D)***

Unit	Potential Outcomes		Causal Effect	
	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$	$\vec{W}$
Patient #1	1	7	6	$W_1 = 0$
Patient #2	6	5	−1	$W_2 = 1$
Patient #3	1	5	4	$W_3 = 1$
Patient #4	8	7	−1	$W_4 = 0$
Average	4	6	2	

Randomness

Assignment Mechanism/Probability

Definition 3.1 (Assignment Mechanism)
Given a population of N units, the assignment mechanism is a row-exchangeable function $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$, *taking on values in* $[0, 1]$, *satisfying*

$$\sum_{\mathbf{W}\in\{0,1\}^N} \Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1, \quad \textit{for all } \mathbf{X}, \mathbf{Y}(0), \textit{ and } \mathbf{Y}(1).$$

- a function of all covariates and potential outcomes that describes the probability of any vector of assignments
- determines:
  - which units receive which treatments,
  - which potential outcomes are realized and can be observed,
  - which potential outcomes are missing;

Example I

- **Setting**: 2 units $\{1, 2\}$, $2^2 = 4$ possible values for $\mathbf{W}$,

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

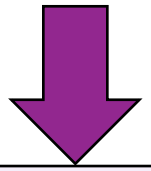
- **Assignment prob**: a randomized experiment with equal probability

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1/4, \quad \text{for } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

Assignment Mechanism/Probability

Definition 3.1 (Assignment Mechanism)
Given a population of N units, the assignment mechanism is a row-exchangeable function $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$, *taking on values in* $[0, 1]$, *satisfying*

$$\sum_{\mathbf{W}\in\{0,1\}^N} \Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1, \quad \textit{for all } \mathbf{X}, \mathbf{Y}(0), \textit{ and } \mathbf{Y}(1).$$



Definition 3.2 (Unit Assignment Probability)
The unit-level assignment probability for unit i is

$$E(W_i|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \sum_{\mathbf{W}:W_i=1} \Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)).$$

Example I

- **Setting**: 2 units $\{1, 2\}$, $2^2 = 4$ possible values for $\mathbf{W}$,

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Assignment prob**: a randomized experiment with equal probability

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1/4, \quad \text{for } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Unit assignment prob:**

  $p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = 1/2$ for both units $i = 1, 2$

Propensity Score

Definition 3.3 (Finite Population Propensity Score)

The propensity score at x is the average unit assignment probability for units with $X_i = x$,

$$e(x) = \frac{1}{N(x)} \sum_{i:X_i=x} p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$$

where $N(x) = \#\{i = 1, \ldots, N | X_i = x\}$ *is the number of units with* $X_i = x$. *For values x with* $N(x) = 0$, *the propensity score is defined to be zero.*

- the average of the unit-level assignment probabilities for subpopulations with a common value of the covariates $X_i = x$.



Example I

- **Setting**: 2 units {1, 2}, $2^2 = 4$ possible values for **W**,

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Assignment prob**: a randomized experiment with equal probability

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 1/4, \quad \text{for } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Unit assignment prob:**

  $p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = 1/2$ for both units $i = 1, 2$

- **Propensity score:**

  no covariates involved, $e(x)=1/2$ for any covariate value x



Example II

- **Setting**: 2 units $\{1, 2\}$, $2^2 = 4$ possible values for $\mathbf{W}$,

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Assignment prob**: a randomized experiment where only those assignments with exactly one treated and one control unit are allowed

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 1/2 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \\ 0 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \end{cases}$$

- **Unit assignment prob:**

  $p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = 1/2$ for both units $i = 1, 2$

- **Propensity score:**

  no covariates involved, $e(x)=1/2$ for any covariate value x

Example III

- **Setting**: 2 units {1, 2}, $2^2 = 4$ possible values for **W**,

$$\mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

- **Assignment prob**: The unit with more to gain from the active treatment is assigned to the treatment group, and the other to the control group

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 1 & \text{if } Y_2(1) - Y_2(0) > Y_1(1) - Y_1(0) \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ 1 & \text{if } Y_2(1) - Y_2(0) < Y_1(1) - Y_1(0) \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ 1/2 & \text{if } Y_2(1) - Y_2(0) = Y_1(1) - Y_1(0) \text{ and } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \\ 0 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}, \\ 0 & \text{if } Y_2(1) - Y_2(0) < Y_1(1) - Y_1(0) \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\ 0 & \text{if } Y_2(1) - Y_2(0) > Y_1(1) - Y_1(0) \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \end{cases}$$

- **Unit assignment prob:**

  $p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = 0,\ 1/2 \text{ or } 1$

- **Propensity score:**

  no covariates involved, $e(x)=1/2$ for any covariate value x

Example IV

- **Setting**: 3 units {1, 2, 3}, covariate $X_i = i$, $2^3 = 8$ possible values for $\mathbf{W}$,

$$\mathbf{W} \in \left\{ \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix} \right\}$$

- **Assignment prob**: a sequential randomized experiment allows for dependence of the assignment mechanism on the potential outcomes

  - Unit 1 ($X_i = 1$): a fair coin toss ⟹ Y_1^{obs}
  - Unit 2 ($X_i = 2$): assign to alternative ⟹ Y_2^{obs}
  - Unit 3 ($X_i = 3$): assign to the active or control treatment that appears better, based on a comparison of Y_1^{obs} & Y_2^{obs}

Assumptions for Assignment Mechanism



Restrictions on Assignment Mechanism (I)

- Individualistic assignment: a unit's assignment probability does not depend on the values of covariates and potential outcomes for other units.

Definition 3.4 (Individualistic Assignment)

An assignment mechanism $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ *is individualistic if, for some function* $q(\cdot) \in [0, 1]$,

$$p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1)), \quad \text{for all } i = 1, \ldots, N,$$

and

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} q(X_i, Y_i(0), Y_i(1))^{W_i} \left(1 - q(X_i, Y_i(0), Y_i(1))\right)^{1-W_i},$$

for $(\mathbf{W}, \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \in \mathbb{A}$, *for some set* $\mathbb{A}$, *and zero elsewhere (c is the constant that ensures that the probabilities sum to unity).*



Restrictions on Assignment Mechanism (I)

- Individualistic assignment: a unit's assignment probability does not depend on the values of covariates and potential outcomes for other units.
- Individualistic assignment is violated in sequential experiments such as Example 4.

$$p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1)), \quad \text{for all } i = 1, \ldots, N$$

$$p_3(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 0 & \text{if } Y_1(0) > Y_2(1) \text{ and } Y_1(1) < Y_2(0), \\ 1 & \text{if } Y_1(1) \geq Y_2(0) \text{ and } Y_1(0) \leq Y_2(1), \\ 1/2 & \text{otherwise.} \end{cases}$$

- Given individualistic assignment, the propensity score simplifies to:

$$e(x) = \frac{1}{N(x)} \sum_{i:X_i=x} p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) \quad \Rightarrow \quad e(x) = \frac{1}{N_x} \sum_{i:X_i=x} q(X_i, Y_i(0), Y_i(1)).$$

Restrictions on Assignment Mechanism (II)

- Probabilistic assignment:
  - imply a non-zero probability for each treatment value, for every unit.
  - or, every unit has the possibility of being assigned to the active treatment and the possibility of being assigned to the control treatment.

Definition 3.5 (Probabilistic Assignment)
An assignment mechanism $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$ *is probabilistic if the probability of assignment to treatment for unit i is strictly between zero and one:*

$$0 < p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) < 1, \quad \text{for each possible } \mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1),$$

for all $i = 1, \ldots, N$.



Restrictions on Assignment Mechanism (III)

- Unconfounded assignment:
  - disallows dependence of the assignment mechanism on the potential outcomes.
  - If an assignment mechanism is unconfounded, we can drop the two potential outcomes and write the assignment mechanism as $\Pr(\mathbf{W}|\mathbf{X})$.

Definition 3.6 (Unconfounded Assignment)

An assignment mechanism is unconfounded if it does not depend on the potential outcomes:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}'(0), \mathbf{Y}'(1)),$$



$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \Pr(\mathbf{W}|\mathbf{X})$$

for all $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}(0)$, $\mathbf{Y}(1)$, $\mathbf{Y}'(0)$, *and* $\mathbf{Y}'(1)$.

Restrictions on Assignment Mechanism (III)

- Unconfounded assignment:
  - disallows dependence of the assignment mechanism on the potential outcomes.
  - If an assignment mechanism is unconfounded, we can drop the two potential outcomes and write the assignment mechanism as $\Pr(\mathbf{W}|\mathbf{X})$. $\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \Pr(\mathbf{W}|\mathbf{X})$

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \begin{cases} 1/2 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}, \\ 0 & \text{if } \mathbf{W} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \end{cases}$$

Example 2

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \begin{cases} 1/2 & \text{if } Y_1(0) > Y_2(1), \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(1) \geq Y_2(0), \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(0) \leq Y_2(1), \text{ and } \mathbf{W} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \\ 1/2 & \text{if } Y_1(1) < Y_2(0), \text{ and } \mathbf{W} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \end{cases}$$

Example 4

Combination of Restrictions

- Individualistic assignment

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} q(X_i, Y_i(0), Y_i(1))^{W_i} \left(1 - q(X_i, Y_i(0), Y_i(1))\right)^{1-W_i},$$

$$p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1)), \quad \text{for all } i = 1,\ldots,N,$$

- Unconfounded assignment

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \Pr(\mathbf{W}|\mathbf{X})$$

Assignment mechanism:

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = c \cdot \prod_{i=1}^{n} q(X_i)^{W_i}\left(1 - q(X_i)\right)^{1-W_i}$$

Propensity score: $e(x) = q(x),$

- The assignment mechanism is the product of the propensity scores
- the propensity score can also be interpreted as the unit-level assignment probability for units with covariate $X_i = x$ (not just the average)

Combination of Restrictions

- Individualistic assignment

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} q(X_i, Y_i(0), Y_i(1))^{W_i} \, (1 - q(X_i, Y_i(0), Y_i(1)))^{1-W_i},$$

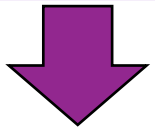
$$p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = q(X_i, Y_i(0), Y_i(1)), \quad \text{for all } i = 1,\ldots,N,$$

- Probabilistic assignment:

$$0 < p_i(\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) < 1,$$

- Unconfounded assignment

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \Pr(\mathbf{W}|\mathbf{X})$$



- Strongly ignorable treatment assignment (Rosenbaum and Rubin, 1983a).
- More generally, ignorable treatment assignment refers to the weaker restriction: the assignment mechanism can be written in terms of $\mathbf{W}, \mathbf{X}$, and $\mathbf{Y}^{obs}$ only, without dependence on $\mathbf{Y}^{mis}$ (Rubin, 1978).

Classification of Assignment Mechanisms



Randomized Experiments



- **Randomized Experiment**: *A randomized experiment is an assignment mechanism that*
  1) *is probabilistic, and* ----▸ random experiment
  2) *has a known functional form that is controlled by the researcher.*
- **Classical Randomized Experiment**: a special case of randomized experiment whose *assignment mechanism is*
  1) *individualistic, and*
  2) *unconfounded.*

- Randomized experimental designs are the most credible basis for causal inference
- U.S. FDA typically relies on such experiments in its approval for pharmaceutical treatments.

Observational Studies



- **Observational Study**: An assignment mechanism corresponds to an observational study if the functional form of the assignment mechanism is **unknown**.

The key difference from experiments

- **Regular Assignment Mechanism**: An assignment mechanism is **regular** if it is
  1) individualistic,
  2) probabilistic,
  3) unconfounded.

- **Irregular Assignment Mechanism**:
  - ✧ Assignment to treatment may differ for some units from the receipt of treatment, i.e., noncompliance
  - ✧ We often assume that assignment to treatment itself is unconfounded, but allow receipt of treatment to be confounded

Classic Randomized Experiments

- Four requirements for assignment mechanism:
  1) is individualistic
  2) is probabilistic
  3) is unconfounded
  4) has a known functional form that is controlled by the researcher

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} e(X_i)^{W_i} \cdot (1-e(X_i))^{1-W_i}$$

$$c = \left(\sum_{\mathbf{W}\in\mathbb{W}^+} \prod_{i=1}^{N} e(X_i)^{W_i} \cdot (1-e(X_i))^{1-W_i} \right)^{-1}.$$

- Four important special cases:
  1) Bernoulli trials
  2) completely randomized experiments
  3) stratified randomized experiments (randomized blocks)
  4) paired randomized experiments

The key difference is $\mathbb{W}^+$.

$\mathbb{W}$: the set of all possible values

$\mathbb{W}^+$: the set of possible values for W with positive probability



Bernoulli Trails

Drawback?

- **Examples:** Independent tosses of a fair coin

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 0.5^N$$

Independent tosses of a biased coin

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q^{N_t} \cdot (1-q)^{N_c}$$

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i}\right]$$ ----→ Independent tosses of coins

- **Bernoulli Trial (a formal definition):**

  a classical randomized experiment with an assignment mechanism such that the assignments for all units are independent.

- **Theorem 4.1 (Assignment Mechanism for a Bernoulli Trial)**

  If the assignment mechanism is a Bernoulli trial, then

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i}\right],$$

  where $e(x)$ is the propensity score, which must be strictly between zero and one for all i, implying $\mathbb{W}^+ = \{0, 1\}^N$

Completely Randomized Experiment

- A leading case of a classical randomized experiment Drawback?
- The number of treated units N_t is fixed
- N_t units are randomly selected, from a population of N units, to receive the active treatment, with the remaining N_c assigned to the control group.
- Each unit has unit assignment probability $q = N_t / N$
- The assignment mechanism:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 1 \Big/ \binom{N}{N_t} & \text{if } \sum_{i=1}^{N} W_i = N_t, \\ 0 & \text{otherwise,} \end{cases}$$

$$\binom{N}{N_t} = \frac{N!}{N_t!(N - N_t)!}$$

- subset of the set of possible values for $\mathbf{W}$ with positive probability

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i=1}^{N} W_i = N_t \right\}$$

Stratified Randomized Experiment

- Divide N units into J blocks based on covariates, block of unit i: $B_i=B(X_i)$
- The size of unit block j is N_j
- Randomly assign $N_t(j)$ units in block j to receive active treatment, the other units in the block to receive control treatment
- **Stratified Randomized Experiment (a formal definition):**

A stratified randomized experiment with J blocks is a classical randomized experiment with an assignment mechanism satisfying

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i:B_i=j}^{N} W_i = N_t(j), \text{for } j = 1, 2, \ldots, J \right\}$$

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} \prod_{j=1}^{J} \binom{N(j)}{N_t(j)}^{-1} & \text{if } \mathbf{W} \in \mathbb{W}^+, \\ 0 & \text{otherwise,} \end{cases} \Rightarrow e(X_i) = \frac{N_t(j)}{N(j)}, \quad B_i = j$$

for some preset $N_t(j)$ such that $N_j > N_t(j) > 0$, for $j = 1, \ldots, J$

Paired Randomized Design

- Example: an educational experiment with a pre-test score as covariate
  - Rank the students from high to low on their scores on the pre-test
  - The top 2 form the first pair, the next 2 form the next pair, and so on
  - In each pair, one unit is randomly assigned to the treatment with probability 1/2.
- An extreme version of the stratified randomized experiment
- Two units within each block, and a fair coin is tossed to decide which member of the pair gets the active/control treatment

- **Paired Randomized Experiment (a formal definition):**
  A paired randomized experiment is a stratified randomized experiment with $N(j) = 2$ *and* $N_t(j) = 1$ *for* $j = 1,\ldots,N/2$, *so that*

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i:B_i=j}^{N} W_i = 1, \text{for } j = 1, 2, \ldots, N/2 \right\}$$

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \begin{cases} 2^{-N/2} & \text{if } \mathbf{W} \in \mathbb{W}^+, \\ 0 & \text{otherwise.} \end{cases}$$

A Comparison with a toy example

2 units for $X_i = x, x \in \left\{1, \ldots, \frac{N}{2}\right\}; \quad e(x) = \frac{1}{2}$

Table 4.1. ***Number of Possible Values for the Assignment Vector by Design and Sample Size***

Type of Experiment and Design	Number of Possible Assignments Cardinality of $\mathbb{W}^+$	Number of Units (N) in Sample			
		4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely randomized experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified randomized experiment	$\binom{N/2}{N/4}^2$	4	36	4,900	0.2×10^9
Paired randomized experiment	$2^{N/2}$	4	16	256	65,536

Eliminate "unhelpful" assignment vectors that are *a priori* unlikely to lead to precise causal inferences.



Super Population



The Concept of Super Population

- A finite population of units: $\{1, 2, \cdots, N\}$
- The finite units of size N can be viewed as a random sample from an infinite super-population, i.e.,
  - $(Y_i(0), Y_i(1), W_i, X_i)$ follow a distribution indexed by a global parameter (ϕ, θ, ψ) of the form below:

$$f_{W|Y(0),Y(1),X}(W_i|Y_i(0), Y_i(1), X_i, \phi) \cdot f_{Y(0),Y(1)|X}(Y_i(0), Y_i(1)|X_i, \theta) \cdot f_X(X_i|\psi),$$

  - $(Y_i(0), Y_i(1), X_i)$ are i.i.d. draws

The causal estimand for a super population

$$\tau_{sp} = E(Y_i(1) - Y_i(0))$$



Super-Population Propensity Score

Definition 3.7 (Super-Population Propensity Score)

The propensity score at x is the population average unit assignment probability for units with $X_i = x$,

$$e(x) = E_Y[E_W(W_i|X_i = x, Y(0), Y(1)] = E_{SP}[W_i|X_i = x]$$
$$= \Pr(W_i = 1|X_i = x, \phi, \theta)$$

for all x in the support of X_i; $e(x)$ *is here a function of* ϕ, *a dependence that we usually suppress notationally.*

Restrictions on Assignment Mechanism for a Super-Population

Definition 3.8 (Super-Population Probabilistic Assignment)
An assignment mechanism is super-population probabilistic if the probability of assignment to treatment for unit i is strictly between zero and one:

$$0 < f_{W|Y(0),Y(1),X}(1|Y_i(0), Y_i(1), X_i, \phi) < 1, \quad \text{for each possible } X_i, Y_i(0), Y_i(1).$$

Definition 3.9 (Super-Population Unconfounded Assignment)
An assignment mechanism is super-population unconfounded if it does not depend on the potential outcomes:

$$f_{W|Y(0),Y(1),X}(w|y_0, y_1, x, \phi) = f_{W|Y(0),Y(1),X}(w|y'_0, y'_1, x, \phi),$$

for all y_0, y_1, x, y'_0, y'_1, ϕ, *and for* $w = 0, 1$.



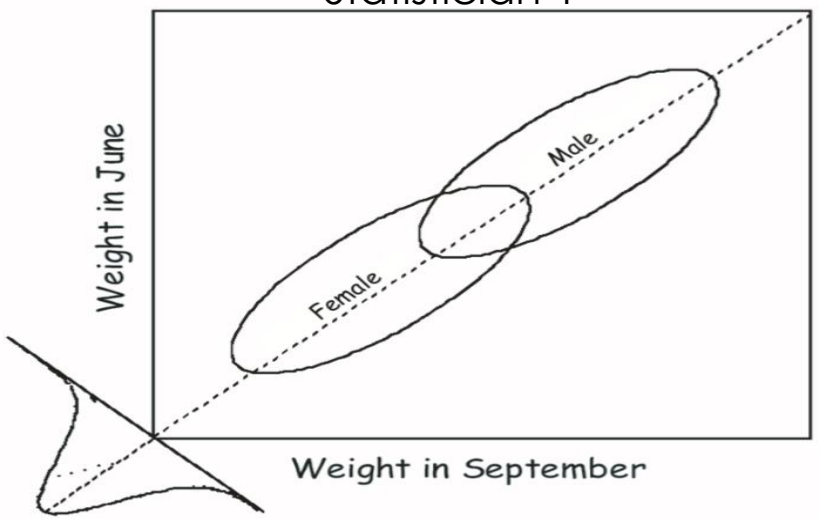
Example: Lord’s Paradox Revisit



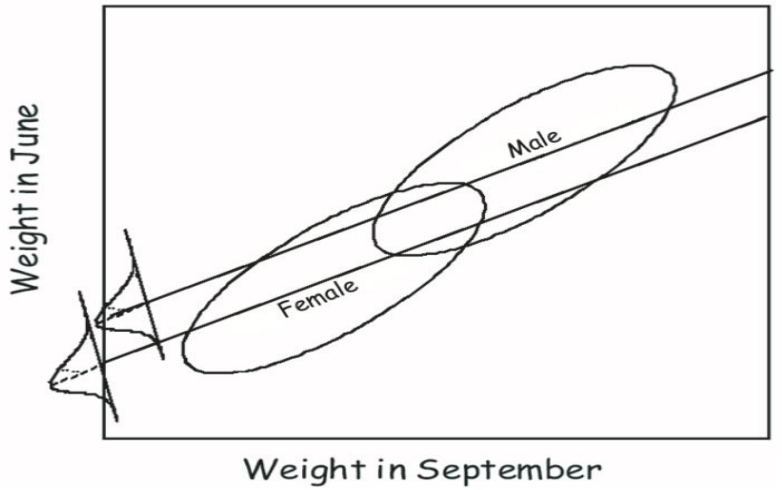
Lord's Paradox revisit

Two hypothetical statisticians got opposite conclusions for the same data

Statistician 1



Statistician 2



The Neyman-Rubin framework

Unit: the students

Treatment: the university diet (Sep. - June) ⟹ Potential outcome Y_t

Control: not defined at all ⟹ Potential outcome Y_c ⇢ ?

The observed weights in June

Assignment mechanism: assigned everyone to the treatment

Covariates: gender, weight in Sep.

Lord's Paradox revisit

Statistician 1:

The effects of medical college on males and females are:

$$D_i = E(Y_T - Y_C | G = i), \qquad i = 1, 2.$$

Thus, the difference of causal effects between males and females:

$$\begin{aligned} D &= D_1 - D_2. \\ &= [E(Y_T|G=1) - E(Y_C|G=1)] - [E(Y_T|G=2) - E(Y_C|G=2)]. \\ &= [E(Y_T|G=1) - E(Y_T|G=2)] - [E(Y_C|G=1) - E(Y_C|G=2)]. \end{aligned}$$

Assumption: $Y_C = X,$

$$\Rightarrow \quad E(Y_C|G=i) = E(X|G=i).$$

Estimation:

$$\hat{D} = [\bar{Y}_{T,G=1} - \bar{Y}_{T,G=2}] - [\bar{X}_{G=1} - \bar{X}_{G=2}] = 0$$



Lord's Paradox revisit

Statistician 2:

Adjust the effects of the university diet based on covariate X (weight in Sep.):

$$E(Y_T|X, G = i), \qquad i = 1, 2,$$

Improvement on each level of X: $U_i(X) = E(Y_T - X|X, G = i), \qquad i = 1, 2.$

Thus, the difference: $U(X) = U_1(X) - U_2(X).$

Assumption 1: $E(Y_T|X, G = i) = a_i + bX, \qquad i = 1, 2.$

$\Rightarrow U_i(X) = a_i + (b-1)X, \qquad i = 1, 2. \Rightarrow U(X) = a_1 - a_2.$

Assumption 2 (build connection between U and D):

$$E(Y_C \mid X, G = i) = a + b'X, \qquad i = 1, 2$$

The difference of conditional causal effects between males and females:

$$D(X) = E(Y_T - Y_C \mid X, G = 1) - E(Y_T - Y_C \mid X, G = 2)$$
$$= [E(Y_T \mid X, G = 1) - E(Y_C \mid X, G = 1)] - [E(Y_T \mid X, G = 2) - E(Y_C \mid X, G = 2)]$$
$$= (a_1 - a) - (a_2 - a) = a_1 - a_2 = U(X)$$



Summary

- Definitions for Assignment Mechanism:
  - ✓ Assignment Mechanism
  - ✓ Unit Assignment Probability
  - ✓ Propensity Score
- Assumptions for Assignment Mechanism:
  - ✓ Individualistic assignment
  - ✓ Probabilistic assignment
  - ✓ Unconfounded assignment
- Classification of Assignment Mechanisms:
  - ✓ Randomized Experiment: Bernoulli trials, CRE, SRE, PRE, etc.
  - ✓ Observational Study: RAM.
- Super Population: Notation; difference vs Finite Sample
- Example - Lord’s Paradox: The power of clear causal framework.



Reference

- Reference: [T1] Chapter 3 & 4
- Further thinking about Simpson’s Paradox:
  - Is it always more accurate with finer stratification?
  - Can we always do stratification?

- Rubin, D. B. (1974), “Estim
Randomized and Non-random
Educational Psychology, Vol.
- Lord, F. M. (1967), “A Parado
Comparisons”, Psychological

Appendix: Another Example for Simpson's Paradox



Simpson’s Paradox: Example 2

- **Two important exams to be a doctor in USA**:

  MCAT (Medical College Admission Test),

  USMLE (United States Medical Licensing Exam) Step 1, usually taken after two years in the medical college.

- **Data**: students (black and white) attending for these two exams during 1993 – 1995



Simpson's Paradox: Example 2

Table 1. Selected Medical School Application and Licensing Statistics

	Frequencies											
	Last MCAT score all MCAT takers 1993–1995			*Applied to medical school 1994–1996*			*Accepted at medical school 1994–1996*			*USMLE Step 1 test volumes 1996–1998*		
MCAT-BS score	*Black*	*White*	*Total*	*Black*	*White*	*Total*	*Black*	*White*	*Total*	*Black*	*White*	*Total*
3 or less	1,308	1,168	2,476	404	238	642	8	1	9	6	1	7
4	1,215	2,094	3,309	482	557	1,039	52	10	62	39	10	49
5	1,219	3,547	4,766	582	1,114	1,696	202	45	247	116	36	152
6	1,269	5,289	6,558	752	1,983	2,735	417	163	580	256	140	396
7	1,091	6,969	8,060	748	3,316	4,064	518	636	1,154	338	589	927
8	1,234	11,949	13,183	868	6,698	7,566	705	2,284	2,989	537	2,167	2,704
9	702	13,445	14,147	544	8,628	9,172	476	4,253	4,729	340	4,003	4,343
10 or more	660	29,752	30,412	511	20,485	20,996	475	14,244	14,719	334	12,786	13,120
Totals	**8,698**	**74,213**	**82,911**	**4,891**	**43,019**	**47,910**	**2,853**	**21,636**	**24,489**	**1,966**	**19,732**	**21,698**

Simpson's Paradox: Example 2

>

	Selected conditional probabilities											
	Probability of MCAT taker applying to medical school			Probability of MCAT taker being accepted to medical school			Probability of MCAT taker taking USMLE Step 1			Probability of medical school acceptee taking USMLE Step 1		
MCAT-BS score	Black	White	Total	Black	White	Total	Black	White	Total	Black	White	Total
3 or less	0.31	0.20	0.26	0.01	0.00	0.00	0.00	0.00	0.00	0.75		0.78
4	0.40	0.27	0.31	0.04	0.00	0.02	0.03	0.00	0.01	0.75	1.00	0.79
5	0.48	0.31	0.36	0.17	0.01	0.05	0.10	0.01	0.03	0.57	0.80	0.62
6	0.59	0.37	0.42	0.33	0.03	0.09	0.20	0.03	0.06	0.61	0.86	0.68
7	0.69	0.48	0.50	0.47	0.09	0.14	0.31	0.08	0.12	0.65	0.93	0.80
8	0.70	0.56	0.57	0.57	0.19	0.23	0.44	0.18	0.21	0.76	0.95	0.90
9	0.77	0.64	0.65	0.68	0.32	0.33	0.48	0.30	0.31	0.71	0.94	0.92
10 or more	0.77	0.69	0.69	0.72	0.48	0.48	0.51	0.43	0.43	0.70	0.90	0.89
Totals	**0.56**	**0.58**	**0.58**	**0.33**	**0.29**	**0.30**	**0.23**	**0.27**	**0.26**	**0.69**	**0.91**	**0.89**

<