STAT6061/STAT5008 – Causal Inference

Part 3-1. Stratification and Standardization

An-Shun Tai

[1]*Department of Statistics*
National Cheng Kung University

[2]*Institute of Statistics and Data Science*
National Tsing Hua University

Randomized experiment vs. observational study

- An observational study is characterized by an unknown functional form of the assignment mechanism, typically expressed as the propensity score $e(x) = \Pr(A = 1 | X = x, Y(0), Y(1))$, whereas in a randomized experiment, the assignment mechanism is known and explicitly specified by the study design.

- An observational study typically assumes that the assignment mechanism is regular—meaning it is individualistic, probabilistic, and unconfounded—whereas a randomized experiment ensures the assignment mechanism is regular by design and under experimental control.

Randomized experiment	Observational study
Assignment mechanism is **designed** to be individualistic.	Assignment mechanism is **assumed** to be individualistic (also known as the *no interference assumption*, as discussed in Part 1.3).
Assignment mechanism is **designed** to be probabilistic.	Assignment mechanism is **assumed** to be probabilistic (also known as the *positivity assumption* as discussed in Part 1.3).
Assignment mechanism is **designed** to be unconfounded.	Assignment mechanism is **assumed** to be unconfounded (also known as the *exchangeability assumption* as discussed in Part 1.3).
Propensity score is **known**	Propensity score is **unknown and needs to be estimated**

Randomization

- Randomization guarantees unconfounded treatment assignment, ensuring marginal exchangeability: $\{Y(1), Y(0)\} \perp A$

- Due to randomization, treated and control samples are exchangeable and considered "similar":
  – Similar in terms of **observed covariates** (e.g., age, gender, weight, baseline health status, socioeconomic factors)
  – More importantly, similar with respect to **unobserved covariates** and **potential outcomes**

- Conditional randomization (stratified randomization) ensures
  – Conditional exchangeability: $\{Y(1), Y(0)\} \perp A | X$.
  – That is, treated and control samples are considered "similar" within each stratum defined by covariates.

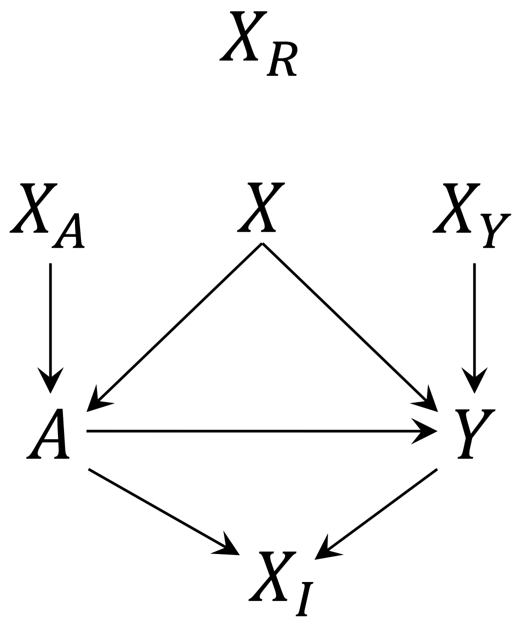
- The core rationale of causal inference in observational studies is to conceptualize them as conditionally randomized experiments, given observed covariates.

- Although observational studies lack randomization, we can assess "similarity" within each stratum—or aim to reduce "dissimilarity"—through appropriate **covariate adjustment methods**.

What covariates should we adjust for in observational studies?

(Ding, 2024)

The covariates have different features:



1. X $(= C)$ affects both the treatment and the outcome. Conditioning on X ensures ignorability, so we should control for X.
2. X_R is pure random noise not affecting either the treatment or the outcome. Including it in analysis does not bias the estimate but it introduces unnecessary variability in finite sample.
3. X_A is an instrumental variable that affects the outcome only through the treatment. Including X_A in analysis does not bias the estimate although it increases variability. However, with unmeasured confounding, including it in analysis amplifies the bias.
4. X_Y affects the outcome only but not the treatment. Without conditioning on it, the ignorability still holds. Since they are predictive to the outcome, including them in analysis often improves precision.
5. X_I is affected by the treatment and outcome. It is a post-treatment variable, not a pretreatment covariate. We should not include it if the goal is to infer the effect of the treatment on the outcome.

Assess causal effects from observational studies

- The covariates that require adjustment in observational studies are the confounders, denoted by C.

- Identification assumptions:
  – Conditional exchangeability: $\{Y(1), Y(0)\} \perp A|C$.
  – SUTVA: no interference + consistency
  – Positivity

- Under the three identification assumptions, the causal parameter $\mathbb{E}(Y(a))$ can be identified as

$$\int \mathbb{E}(Y|A = a, C = c)\Pr(C = c)dc$$

The ATE, $\mathbb{E}(Y(1) - Y(0))$, is identified as

$$\int \{\mathbb{E}(Y|A = 1, C = c) - \mathbb{E}(Y|A = 0, C = c)\}\Pr(C = c)dc$$

- This formulation aligns with the core principle of estimating causal effects in conditionally (stratified) randomized experiments.

Estimation of ATE

- How can we estimate ATE

$$\tau = \int \{\mathbb{E}(Y|A = 1, C = c) - \mathbb{E}(Y|A = 0, C = c)\}\Pr(C = c)dc$$

- Following the concept of stratified randomized experiments, we use stratification to estimate causal effects within each stratum, and standardization to derive the marginal effect by weighting each stratum-specific estimate by its population proportion.

- Discrete covariate/confounder: $\{Y(1), Y(0)\} \perp A|C = c$ for $c = 1,2,3 \dots, K$

$$\hat{\tau} = \sum_{i=1}^{N} \pi_c \left\{\frac{I(A_i = 1, C_i = c)Y_i}{N_{c,1}} - \frac{I(A_i = 0, C_i = c)Y_i}{N_{c,0}}\right\}$$

$\pi_c = \#\{i: C_i = c\}$
$N_{c,a} = \#\{i: C_i = c, A_i = a\}$

Estimation of ATE: Outcome regression

- For continuous C, stratification can be implemented by fitting an outcome regression model with covariate adjustment.

$$\mu_a(C) = \mathbb{E}(Y|A = a, C)$$

- For example, we can assume a linear regression model: $\mu_a(C) = \mathbb{E}(Y|A, C) = \beta_0 + \beta_a A + \beta_c C$
- These parameters can be estimated using ordinary least squares (OLS) or estimating equations.
- The predictor is denoted as $\hat{\mu}_a(C)$

- Standardization can be achieved nonparametrically by using sample mean:

$$\hat{\tau}_o = \frac{1}{N}\sum_{i=1}^{N}\{\hat{\mu}_1(C_i) - \hat{\mu}_0(C_i)\}$$

- The standard error of the outcome regression estimator is typically estimated using a **nonparametric bootstrap**.

Further discussion on outcome regression estimators

- In contrast to completely randomized experiments (where covariates are not confounders), the estimator becomes inconsistent if the model is misspecified. Why?

- Flexible modeling: can incorporate various machine learning techniques (e.g., linear/logistic regression, random forests, SVMs, deep learning) to estimate

- Limitation: Does not inherently ensure the positivity (overlap) assumption; relies on extrapolation in regions with limited data, potentially leading to unstable estimates.

  - Positivity remains essential for standardization because if $\Pr(A = a|C = c) = 0$ while $\Pr(C = c) \neq 0$, then the conditional mean outcome $\mathbb{E}(Y|A = a, C = c)$ is undefined.

Overview of causal inference methods in observational studies

- **Standardization**
  – Outcome regression
  – G-computation
- **Weighting Methods**
  – Inverse Probability of Treatment Weighting (IPTW)
  – Stabilized weights
- **Matching Methods**
  – Exact matching
  – Mahalanobis distance matching
- **Propensity Score Methods**
  – Propensity score stratification
  – Propensity score weighting
  – Propensity score in regressions
  – Propensity score matching
- **Double Robust Methods**
  – Augmented Inverse Probability Weighting (AIPW)
  – Targeted Maximum Likelihood Estimation (TMLE)

G-computation

Step 1: Construct a regression model for outcome Y: $\mathbb{E}(Y|A,C) = g(A,C;\theta)$

Step 2: Fit models with real data to obtain MLE for all parameters: $\widehat{\mathbb{E}}\,(Y|A,C) = g(A,C;\hat{\theta})$

Step 3: Conduct g-computation algorithm using MLE and bootstrap.

(3a) For each individual in your sample (with covariate values C_i), predict the potential outcomes under treatment level $A = 1$: $\hat{Y}_i(1) = g(1, C_i; \hat{\theta})$.

(3b) For each individual in your sample (with covariate values C_i), predict the potential outcomes under treatment level $A = 0$: $\hat{Y}_i(0) = g(0, C_i; \hat{\theta})$.

(3c) Compute the means $Y(a)$ for $a = 1,2$ which is the g-computation algorithm approximation estimation of $\mathbb{E}(Y(a))$: $\sum_i \hat{Y}_i(a)\,/N$

(3d) Bootstrap to obtain the standard errors and corresponding 95% confidence intervals.

References

Ding, P. (2024). *A First Course in Causal Inference*.

Hernán M.A., Robins J.M. (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC