An Introduction to Causal Inference
Final Presentation

Christina Feng, Sherry Chang, Gengyang Chen

Dec. 4, 2023

Can a Heart Transplant Save Lives?

- Zeus is waiting for a heart transplant. He dies five days later after receiving a new heart.
- Would Zeus have been alive if he had not received the heart transplant?
- We want to know if receiving a heart transplant improves patients' survival.
- Our population of interest: Zeus's extended family.

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Definition

- Definition of **a causal effect for an individual**:
  The treatment A has a causal effect on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$ for the individual.
- The variables $Y^{a=1}$ and $Y^{a=0}$ are referred to as potential outcomes or counterfactual outcomes.
- Only one of the potential outcomes can be observed in reality.
- Definition of **average causal effect** in the population:
  An average causal effect of treatment A on outcome Y is $ACE = E[Y^{a=1}] - E[Y^{a=0}]$ in the population of interest.

Identifiability conditions in Observational Studies

- **Exchangeability:**

$$Y^a \perp A | L = l \quad \text{for all } a$$

- **Positivity:**

$$P[A = a | L = l] > 0 \quad \text{for all values of } l \text{ where } P[L = l] \neq 0$$

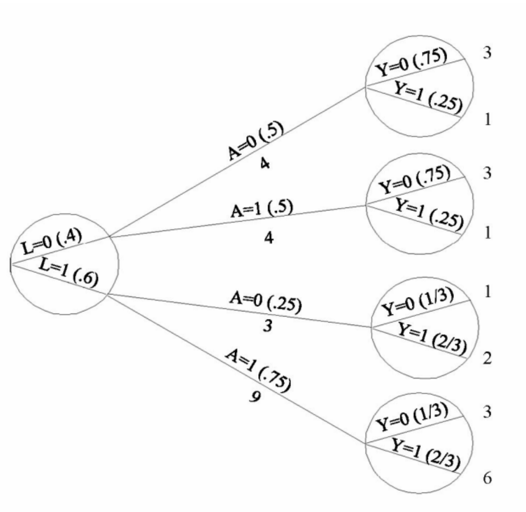
- **Consistency:**

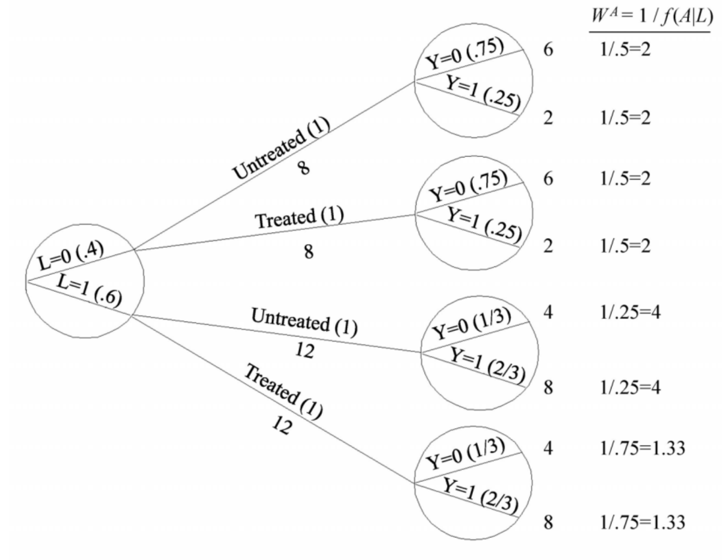
$$Y^a = Y \quad \text{for all individuals with } A = a$$

Methods to estimate ACE - Standardization

- By consistency and conditional exchangeability, $Pr[Y^a = 1|L = l] = Pr[Y = 1|L = l, A = a]$
- Causal risk ratio $= \frac{Pr[Y^{a=1}=1]}{Pr[Y^{a=0}=1]} = \frac{\sum_l Pr[Y=1|L=l,A=1]Pr[L=l]}{\sum_l Pr[Y=1|L=l,A=0]Pr[L=l]}$ by standardization

Methods to estimate ACE - Inverse Probability Weighting





- Weighting each individual by the inverse of the conditional probability of receiving the treatment level to create a pseudo-population that simulates what would have happened if all individuals in the population had been untreated or treated.
- Causal risk ratio $\frac{Pr[Y^{a=1}=1]}{Pr[Y^{a=0}=1]}$ can be calculated using pseudo-pupulation under conditional exchangeability.

Methods to estimate ACE - Stratification

- Goal: To identify effect modification
- Effect Modification: The average causal effect of A on Y varies across levels of V.
- **Stratification**: The causal effect of A on Y is computed in each stratum of V. For dichotomous V , the stratified causal risk differences are:
  $Pr[Y^{a=1} = 1|V = 1] - Pr[Y^{a=0} = 1|V = 1]$
  and
  $Pr[Y^{a=1} = 1|V = 0] - Pr[Y^{a=0} = 1|V = 0]$
- If the average causal effect differs between the two strata, we conclude that there is effect modification by V of the causal effect of A on Y.

Methods to estimate ACE - Matching

- Goal: To construct a subset of the population in which the variables L have the same distribution in both the treated and the untreated.
- **Matching**: For each untreated individual in non critical condition $(A = 0, L = 0)$ randomly select a treated individual in non critical condition $(A = 1, L = 0)$, and for each untreated individual in critical condition $(A = 0, L = 1)$ randomly select a treated individual in critical condition $(A = 1, L = 1)$.
- Often one chooses the group with fewer individuals and uses the other group to find their matches.
- Under the assumption of conditional exchangeability given L, the result of this procedure is unconditional exchangeability of the treated and the untreated in the matched population.
- Matching needs not be one-to one (matching pairs), but it can be one-to-many (matching sets).

Comparison Between Methods

- Standardization and Inverse Probability Weighting are used to compute either marginal or conditional effects.
- Stratification and Matching are used to compute conditional effects in certain subsets of the population.

About the paper

- Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. Statistical Science, 22(4), 523–539. http://www.jstor.org/stable/27645858
- **Purpose:** Investigating the practical behavior of the estimators under different model specifications in terms of estimating a population mean from an incomplete dataset.
- **Assumptions:** The outcome y_i's are missing at random. $P(X, T, Y) = \prod_i P(x_i)P(t_i|x_i)P(y_i|x_i)$

Problem of interest

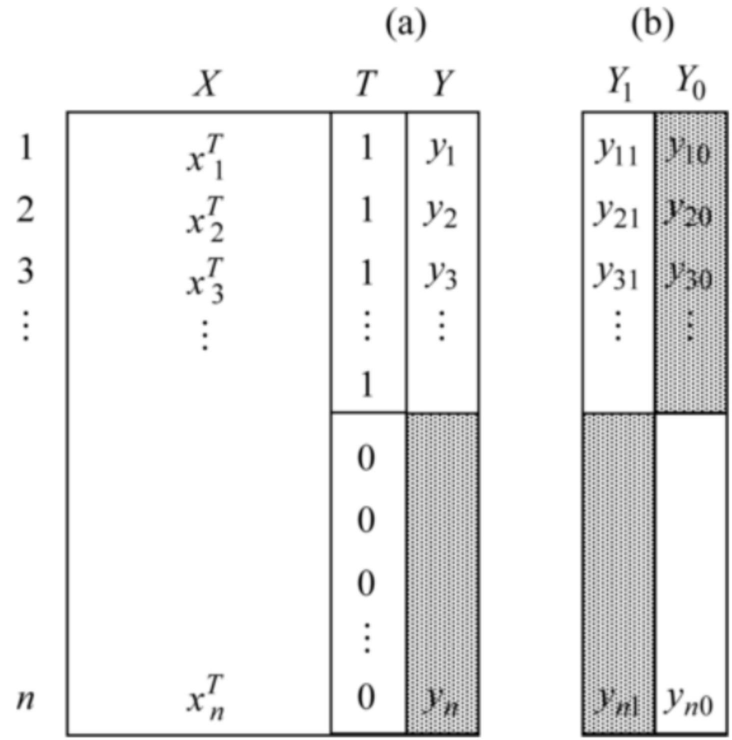


FIG. 1. *Schematic representation of sample data for estimating (a) a population mean and (b) an average causal effect, with missing values denoted by shading.*

π-model and y-model

- π-model
  - A π-model is a proposed functional form for $P(T_i = 1|\boldsymbol{x_i}) = \pi_i(\boldsymbol{x_i}) = \pi_i$.
  - π_i is the propensity score.
  - In the paper and our simulation, $\hat{\pi}_i = \frac{exp(\boldsymbol{x_i}^T \hat{\boldsymbol{\alpha}})}{1+exp(\boldsymbol{x_i}^T \hat{\boldsymbol{\alpha}})}$, where $\hat{\boldsymbol{\alpha}}$ is the maximum-likelihood estimate of the coefficients from the logistic regression of T on X.
- y-model
  - Let us define $E(y_i|\boldsymbol{x_i}) = m(\boldsymbol{x_i}) = m_i$, so that $y_i = m(\boldsymbol{x_i}) + \epsilon_i$ with $E(\epsilon_i) = 0$.
  - A y-model is a functional form for $m(\boldsymbol{x_i})$.
  - In the paper and our simulation, $\hat{m}_i = \boldsymbol{x_i}^T \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the ordinary least square estimates of the regression coefficients for Y.

Setting

- Sample size $n = 200,\ 1000$
- Simulation rounds $m = 1000$
- Let Y denote the outcome and T denote the treatment.
  Y_1 is the outcome under $T = 1$ and Y_0 is the outcome under $T = 0$.
- $Y = TY_1 + (1 - T)Y_0$
- Average causal effect (ACE): $E[Y_1] - E[Y_0]$
- Compare the performance of each method based on:
  - Bias $= \frac{1}{m}\sum_{i=1}^{m}(\hat{\mu}_{i1} - \hat{\mu}_{i0} - ACE)$
  - MSE $= \frac{1}{m}\sum_{i=1}^{m}(\hat{\mu}_{i1} - \hat{\mu}_{i0} - ACE)^2$
  - Relative Bias = Bias / ACE

Data Generation

- Covariates: $X_1 \sim N(0,1)$, $X_2 \sim Bin(1,0.7)$, $X_3 \sim Gamma(2,2)$
- $\boldsymbol{X} = (X_1, X_2, X_3)$
- Coefficients for Y_1: $\boldsymbol{\beta_1} = (99, 66.6, 18.8, -52.1)$
- Coefficients for Y_0: $\boldsymbol{\beta_0} = (19.9, 6.6, 2.33, 8.88)$
- Mechanism for Y_1: $Y_1 = \boldsymbol{X}^T\boldsymbol{\beta_1} + \epsilon_1,\ \epsilon_1 \sim N(0,1)$
- Mechanism for Y_0: $Y_0 = \boldsymbol{X}^T\boldsymbol{\beta_0} + \epsilon_0,\ \epsilon_0 \sim N(0,1)$
- Treatment: $T_i \sim Bernoulli(\pi_i)$, where $\pi_i = expit(\boldsymbol{X_i}^T\boldsymbol{\alpha})$ and $\boldsymbol{\alpha} = (0.1, 0.6, 0.07, -0.4)$

Model Misspecification

- Let S be the sample we obtained, S_1 be the data in the treatment group and S_0 be the data in the control group.
- So the data in the treatment group is $\{(Y_{1j}, \boldsymbol{X_j}, T_j = 1), j \in S_1\}$, the data in the control group is $\{(Y_{0j}, \boldsymbol{X_j}, T_j = 0), j \in S_0\}$.
- We only use two covariates X_1 and X_2 for the misspecification of π-model and y-model, respectively.

Method 1: Inverse-Propensity Weighting

- $\hat{\pi}_i = expit(\boldsymbol{x_i}^T \hat{\boldsymbol{\alpha}}) = \frac{exp(\boldsymbol{x_i}^T \hat{\boldsymbol{\alpha}})}{1+exp(\boldsymbol{x_i}^T \hat{\boldsymbol{\alpha}})}$
- $\hat{\mu}_1 = \frac{\sum_{i=1}^n t_i \hat{\pi}_i^{-1} y_i}{\sum_{i=1}^n t_i \hat{\pi}_i^{-1}}$
- $\hat{\mu}_0 = \frac{\sum_{i=1}^n (1-t_i)(1-\hat{\pi}_i)^{-1} y_i}{\sum_{i=1}^n (1-t_i)(1-\hat{\pi}_i)^{-1}}$
- IPW estimator: $\hat{\mu}_1 - \hat{\mu}_0$

Method 2: Regression Estimation

- $\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}^T \hat{\boldsymbol{\beta}}_{\mathbf{1}}$
- $\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}^T \hat{\boldsymbol{\beta}}_{\mathbf{0}}$
- Ordinary least-squares regression estimator: $\hat{\mu}_1 - \hat{\mu}_0$

Method 3: Regression Estimation with Residual Bias Correction (Doubly Robust Estimator)

- Combine model-based predictions for y_i with inverse-probability weights.
- IPW estimate can in turn be used to correct the OLS estimate for bias arising from y-model failure.
- $\hat{\mu}_1 = \hat{\mu}_{1,OLS} + \frac{1}{n}\sum_{i=1}^{n} t_i \hat{\pi}_i^{-1}(y_i - \boldsymbol{x_i}^T \hat{\boldsymbol{\beta}}_{\mathbf{1}})$
- $\hat{\mu}_0 = \hat{\mu}_{0,OLS} + \frac{1}{n}\sum_{i=1}^{n}(1 - t_i)(1 - \hat{\pi}_i)^{-1}(y_i - \boldsymbol{x_i}^T \hat{\boldsymbol{\beta}}_{\mathbf{0}})$
- Bias-corrected regression estimator: $\hat{\mu}_1 - \hat{\mu}_0$

Simulation result for Method 1

Sample Size	π-model	Bias	Relative Bias	MSE
200	Correct	0.474	0.016	52.547
200	Incorrect	4.812	0.162	67.432
1000	Correct	-0.004	0.000	9.097
1000	Incorrect	4.644	0.157	29.072

Table 1: Performance of IPW estimators of ACE over 1000 samples

Simulation result for Method 2

Sample Size	y-model	Bias	Relative Bias	MSE
200	Correct	0.008	0.000	27.744
200	Incorrect	4.531	0.153	54.149
1000	Correct	-0.045	-0.002	4.998
1000	Incorrect	4.672	0.158	27.411

Table 2: Performance of ordinary least-squares regression estimators of ACE over 1000 samples

Simulation result for Method 3

Sample Size	π-model	y-model	Bias	Relative Bias	MSE
200	Correct	Correct	0.007	0.000	27.743
200	Correct	Incorrect	0.238	0.008	33.365
200	Incorrect	Correct	0.007	0.000	27.742
200	Incorrect	Incorrect	4.538	0.153	54.551
1000	Correct	Correct	-0.044	-0.001	4.999
1000	Correct	Incorrect	-0.057	-0.002	6.199
1000	Incorrect	Correct	-0.045	-0.002	4.999
1000	Incorrect	Incorrect	4.665	0.157	27.416

Table 3: Performance of bias-corrected regression estimators of ACE over 1000 samples

.

Performance Comparison

- If both π- model and y-model are correct, doubly robust estimator outperforms IPW and OLS estimator when $n = 200$; all three estimators perform equally well when $n = 1000$.
- IPW estimators are sensitive to misspecification of the π- model.
- Regression estimators are sensitive to misspecification of the y- model.
- Doubly robust estimator provides consistent estimate of the ACE even if one of the models is misspecified.

Thank you!