CSDS 452: Causality and Machine Learning

Lecture 4: Identification & Estimation

Instructor: Jing Ma

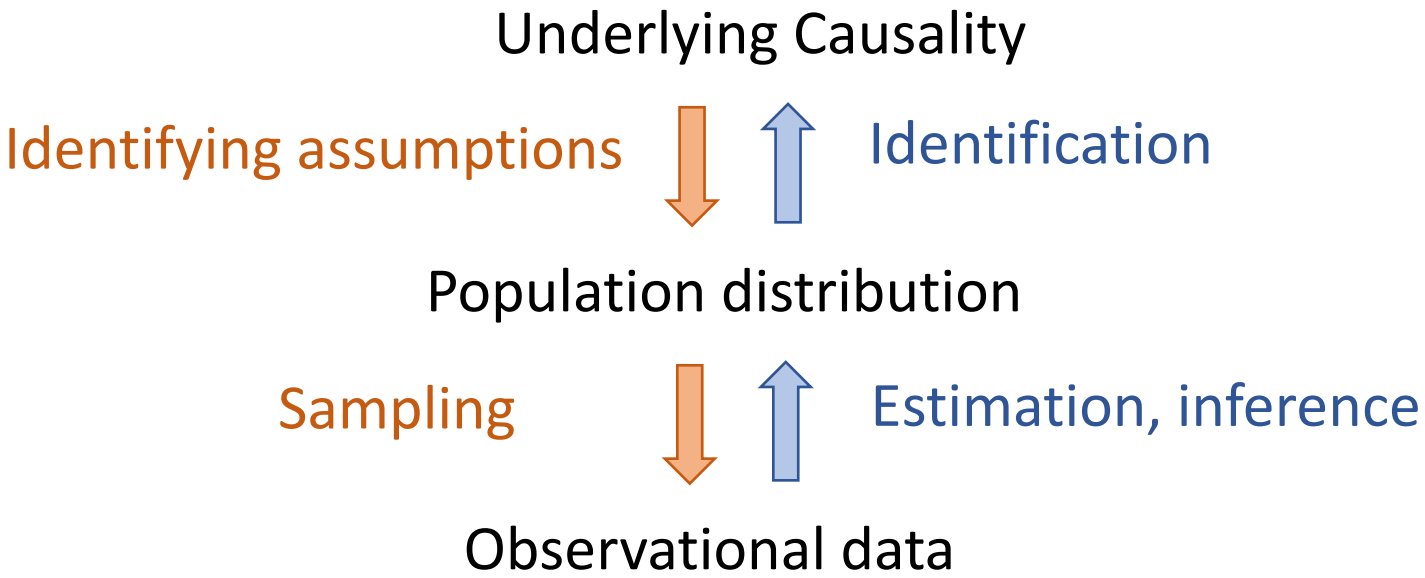
Fall 2024, CDS@CWRU

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

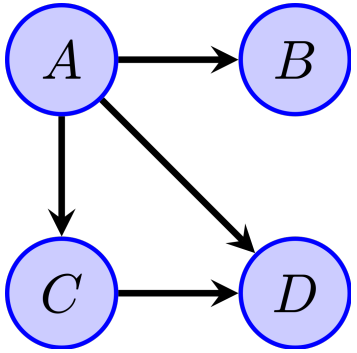
Recap: Identification and Estimation

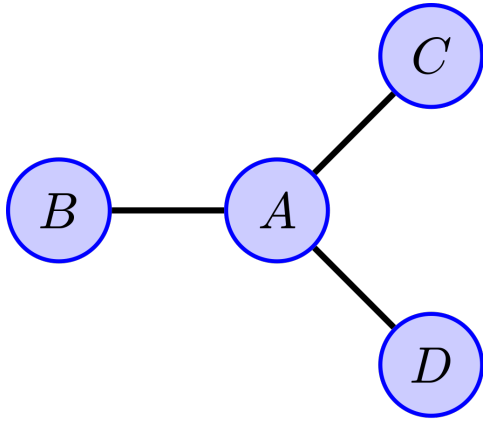
- Two components in learning causality
  - (1) Identification
  - (2) Estimation, inference



Recap: Graph Directionality

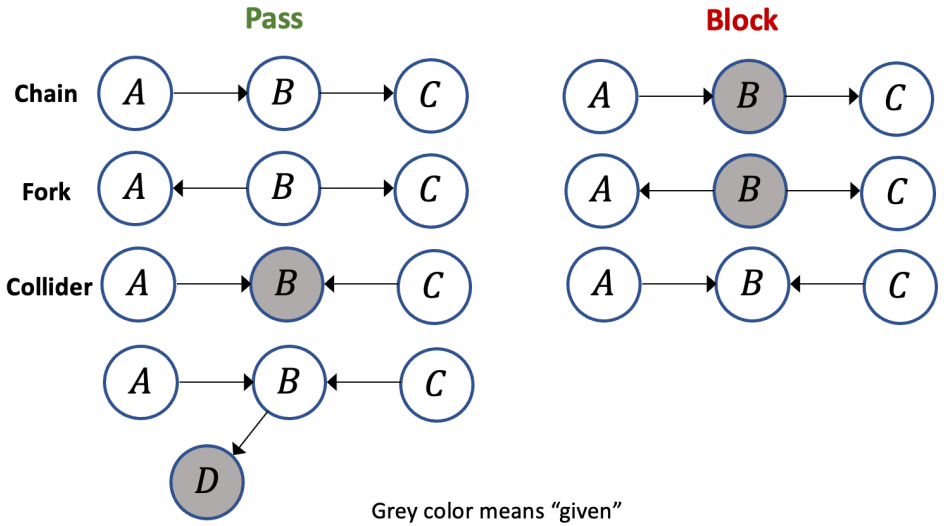
- Directed graphical models
  - Direction in edges
  - Bayesian networks
  - More popular in AI and statistics
- Undirected graphical models
  - Edges without direction
  - Markov random fields (MRFs)
    - Better suited to express soft constraints between variables
  - More popular in Vision and Physics



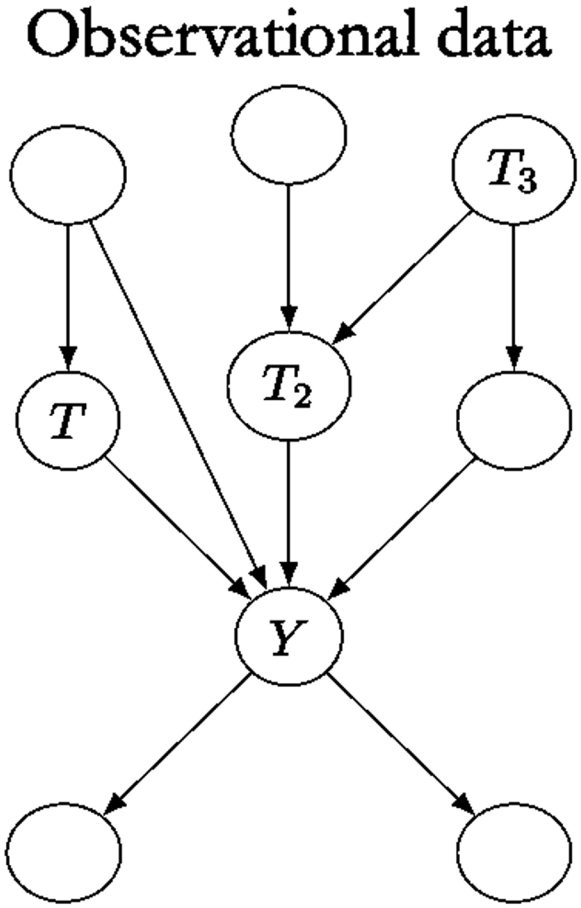


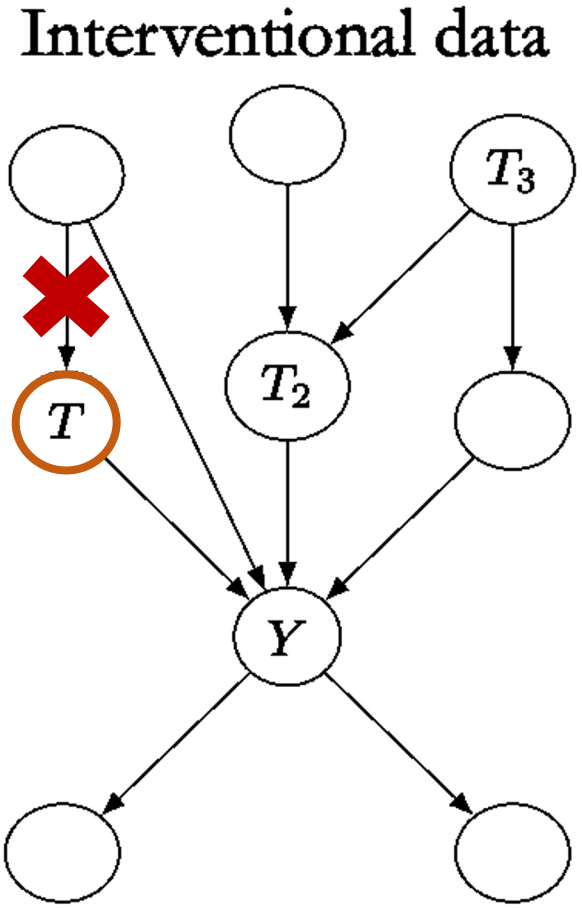
Recap: Blocked Paths

- A path between nodes X and Y is blocked by a (potentially empty) conditioning set Z if either of the following is true:
  - Along the path, there is a chain … $\rightarrow W \rightarrow \cdots$ or a fork … $\leftarrow W \rightarrow \cdots$ where W is conditioned on ($W \in Z$).
  - There is a collider W on the path that is not conditioned on ($W \notin Z$) and none of its descendants are conditioned on ($\mathrm{des}(W) \notin Z$).



Recap: Observation v.s. Intervention





$\boldsymbol{M}$: $T \coloneqq f_T(X, U_T)$

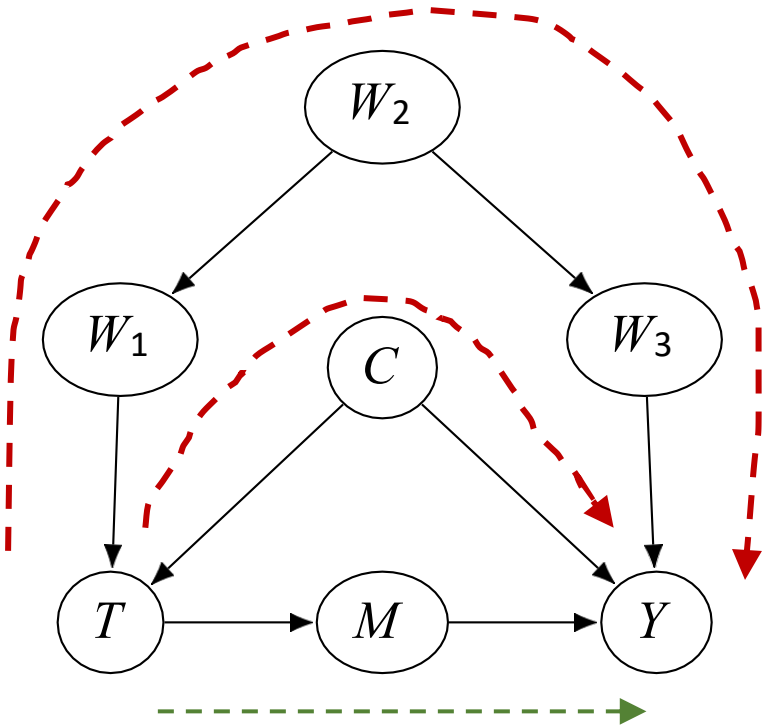
$Y \coloneqq f_Y(X, T, U_Y)$

$\boldsymbol{M_t}$: $T \coloneqq t$

$Y \coloneqq f_Y(X, T, U_Y)$

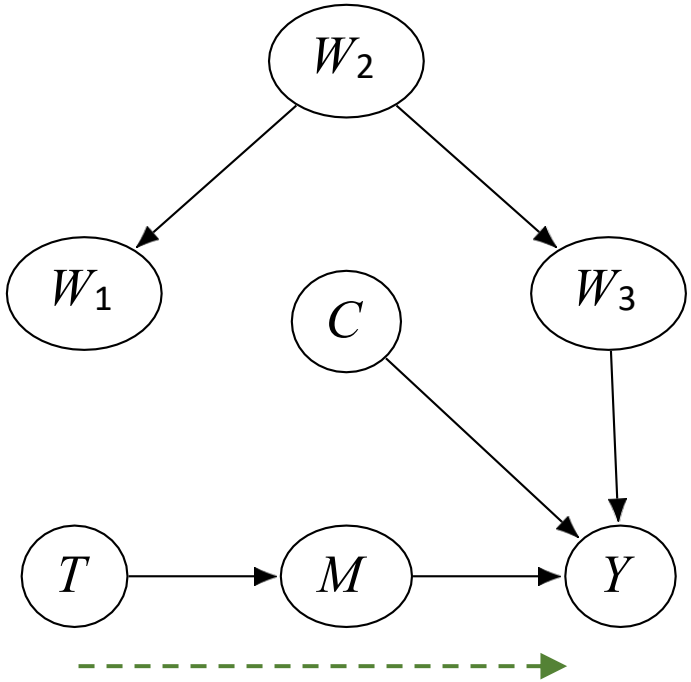
Recap: Backdoor Paths

$P(Y|t)$



Causal association

$P(Y|do(t))$



Causal association

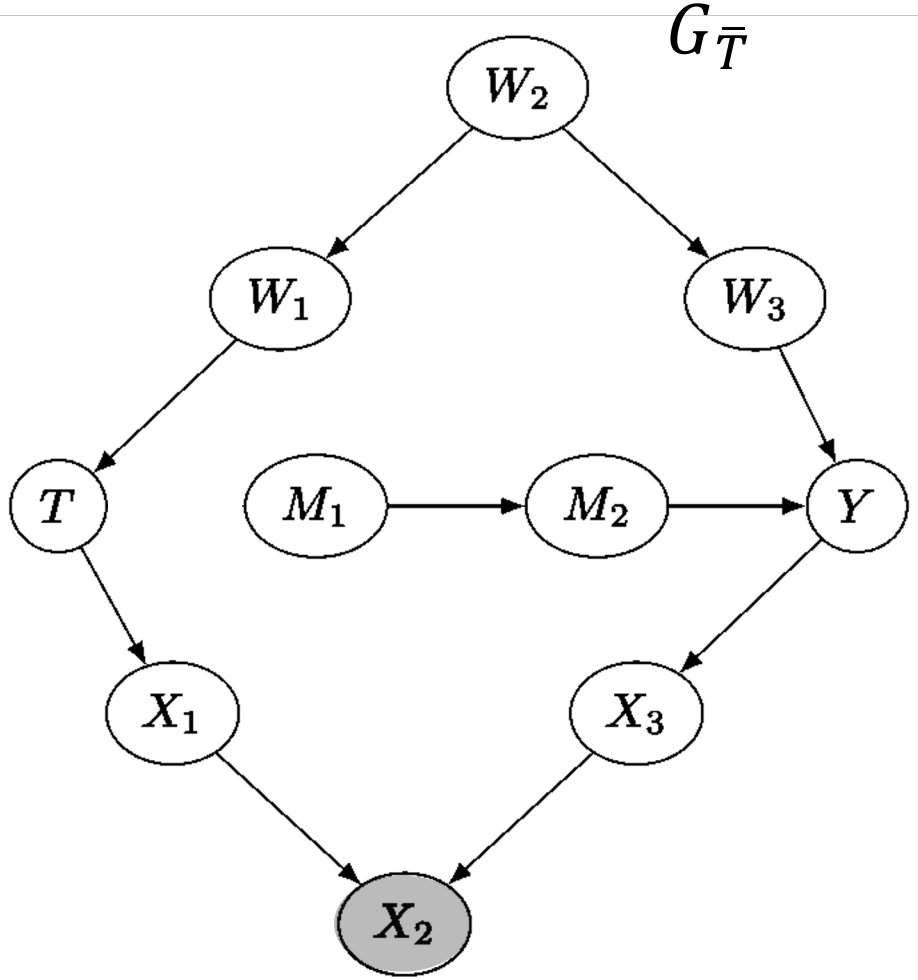
Recap: Backdoor criterion and backdoor adjustment

- A set of variables W satisfies the **backdoor criterion** relative to T and Y if the following are true:
  - W blocks all backdoor paths from T to Y
  - W does not contain any descendants of T

- Given the modularity assumption and that W satisfies the backdoor criterion, we can identify the causal effect of T on Y:

$$P(y|do(t)) = \sum_w P(w)P(y|t,w)$$

Recap: Backdoor criterion as d-separation



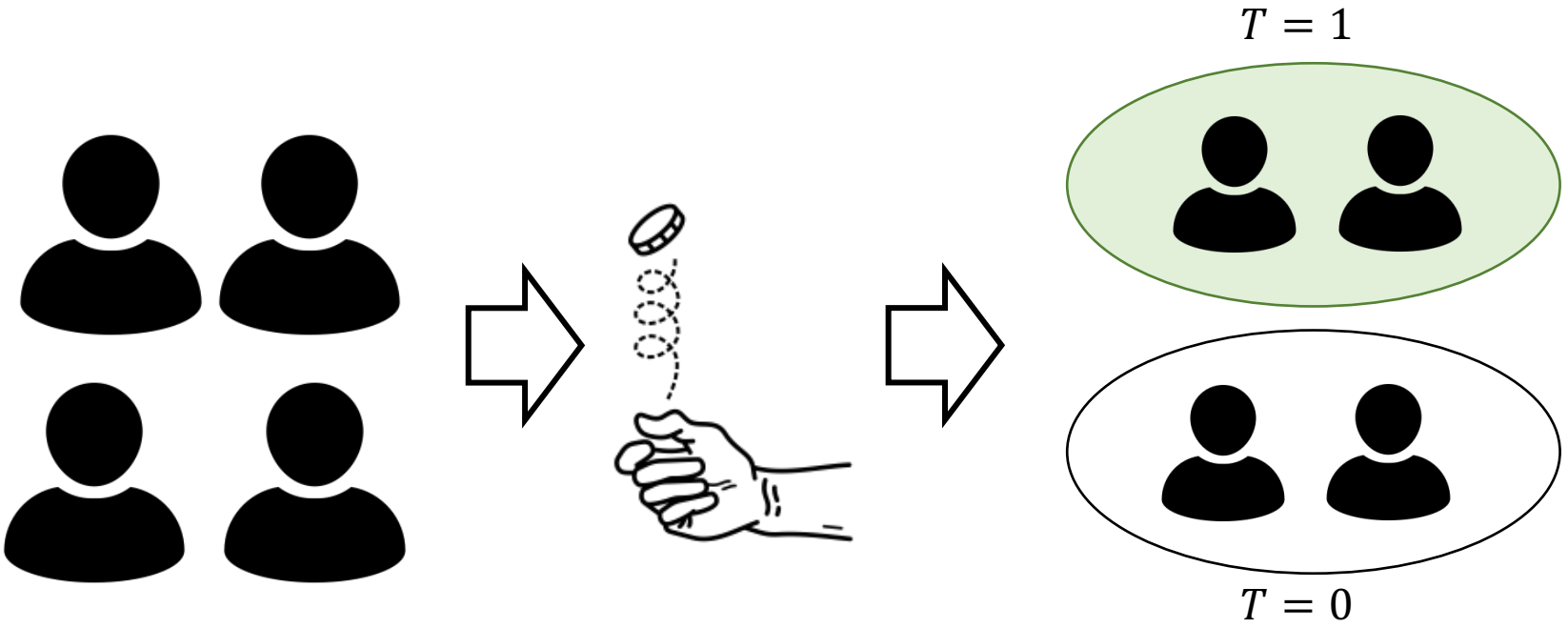
- W blocks all backdoor paths from T to Y
- W does not contain any descendants of T

$$Y \perp\!\!\!\perp_{G_{\bar{T}}} T \mid W$$

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Randomized Experiments



- Gold standard to assess the causal effect.
- The allocation of the treatment is under control. The distribution of the covariates for treated and control patients is balanced.

Few different perspectives on RCT

- Comparability and covariate balance
- Exchangeability
- No backdoor paths

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T = 1) = P(X|T = 0)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

$$P(y|do(t))$$

$$P(y|t)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

$$P(y|do(t)) = \sum_x P(y|t,x)P(x)$$

$P(y|t)$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

$$P(y|do(t)) = \sum_x P(y|t,x)P(x) = \sum_x \frac{P(y|t,x)P(t|x)P(x)}{P(t|x)}$$

$$P(y|t)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

$$P(y|do(t)) = \sum_x P(y|t,x)P(x) = \sum_x \frac{P(y|t,x)P(t|x)P(x)}{P(t|x)}$$

$$= \sum_x \frac{P(y,t,x)}{P(t|x)}$$

$$P(y|t)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

Covariate balance => association is causation

$$P(y|do(t)) = \sum_x P(y|t,x)P(x) = \sum_x \frac{P(y|t,x)P(t|x)P(x)}{P(t|x)}$$

$$= \sum_x \frac{P(y,t,x)}{P(t|x)} = \sum_x \frac{P(y,t,x)}{P(t)}$$

$$P(y|t)$$

Comparability and covariate balance

- Treatment and control groups are the same in all aspects except treatment
- Covariate balance: the distribution of covariates X is the same across treatment groups.

$$P(X|T=1) = P(X|T=0)$$

Randomization => covariate balance

$$\mathrm{P}(X) = P(X|T=1) = P(X|T=0)$$

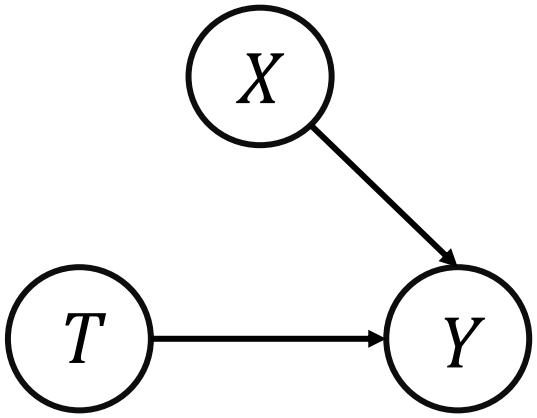
Covariate balance => association is causation

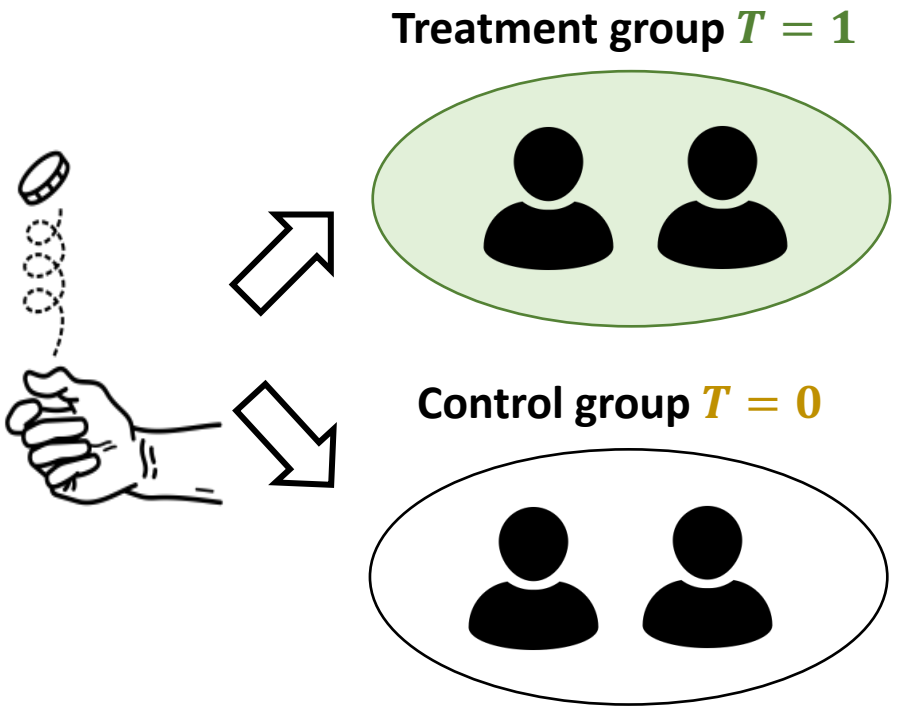
$$P(y|do(t)) = \sum_x P(y|t,x)P(x) = \sum_x \frac{P(y|t,x)P(t|x)P(x)}{P(t|x)}$$

$$= \sum_x \frac{P(y,t,x)}{P(t|x)} = \sum_x \frac{P(y,t,x)}{P(t)} = \sum_x P(y,x|t) = P(y|t)$$

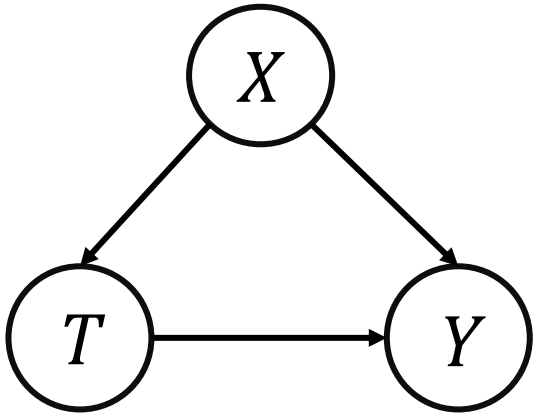
Exchangeability

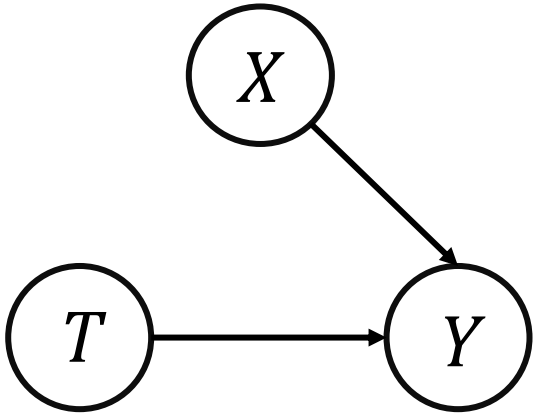
- $\big(Y(1), Y(0)\big) \perp\!\!\!\perp T$
- T is randomly assigned, will not change $Y(1), Y(0)$





No backdoor paths

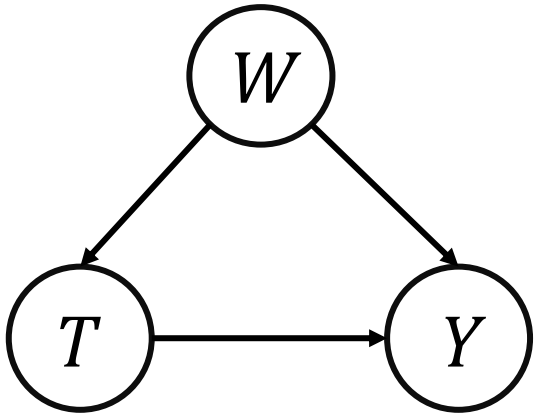




Recall Backdoor Adjustment

- Backdoor adjustment:

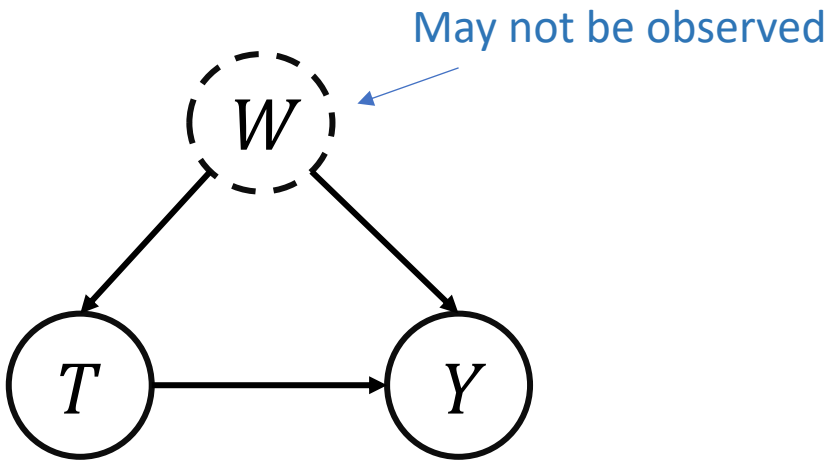
$$P(y|do(t)) = \sum_w P(w)P(y|t,w)$$



Recall Backdoor Adjustment

- Backdoor adjustment:

$$P(y|do(t)) = \sum_w P(w)P(y|t,w)$$



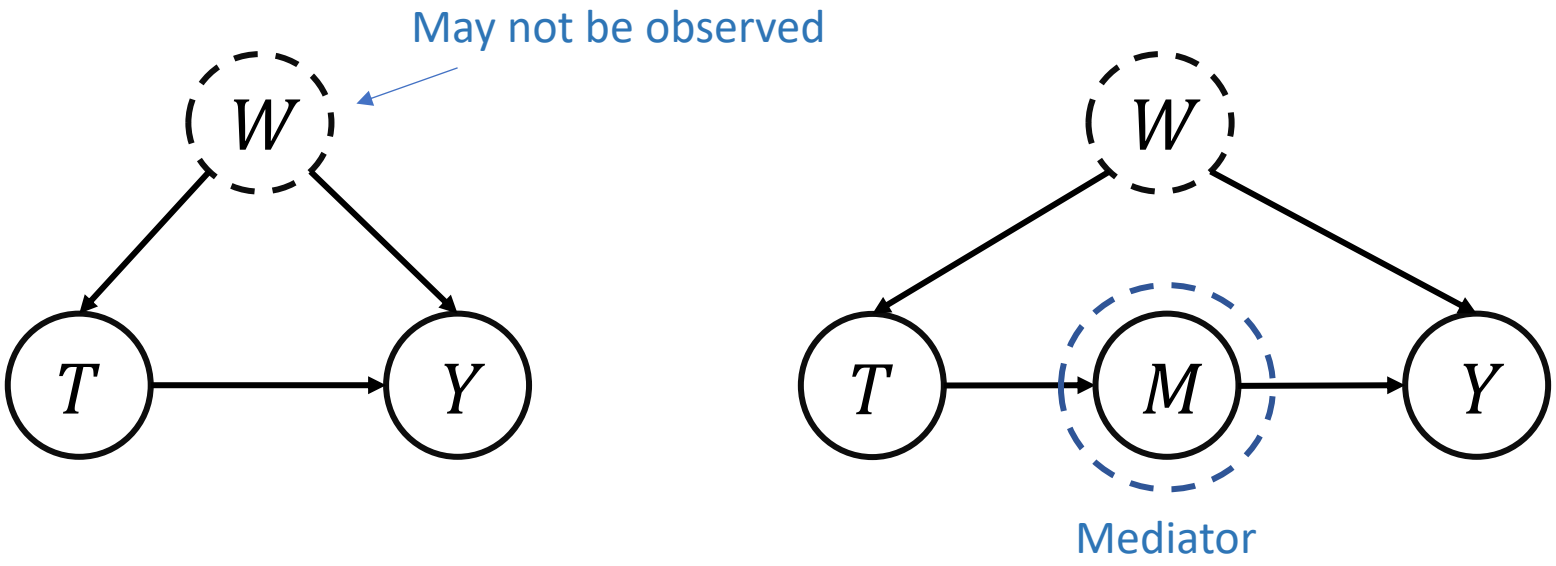
Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Frontdoor Adjustment

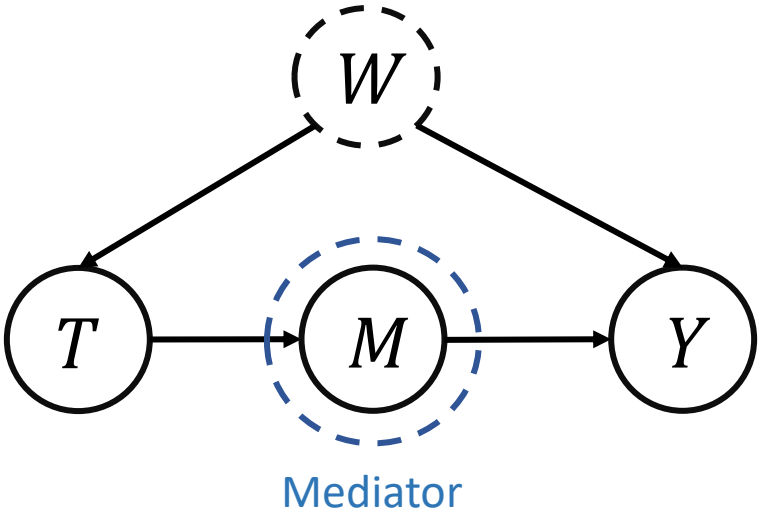
- Backdoor adjustment:

$$P(y|do(t)) = \sum_w P(w)P(y|t,w)$$



Frontdoor Adjustment

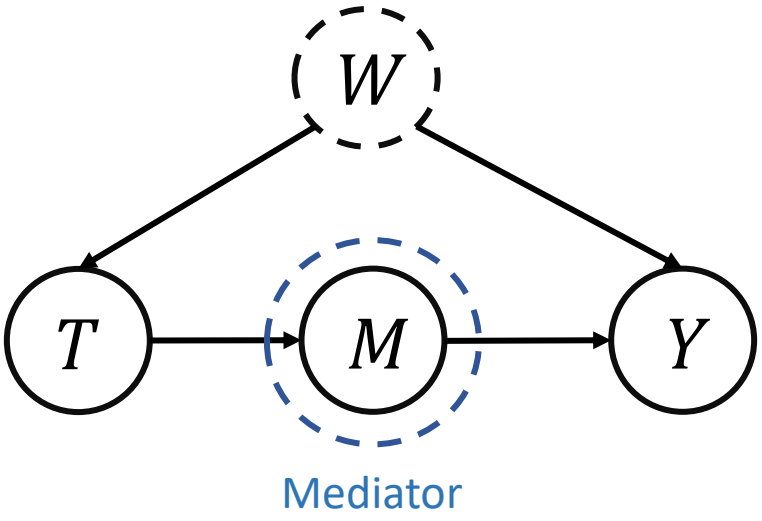
- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

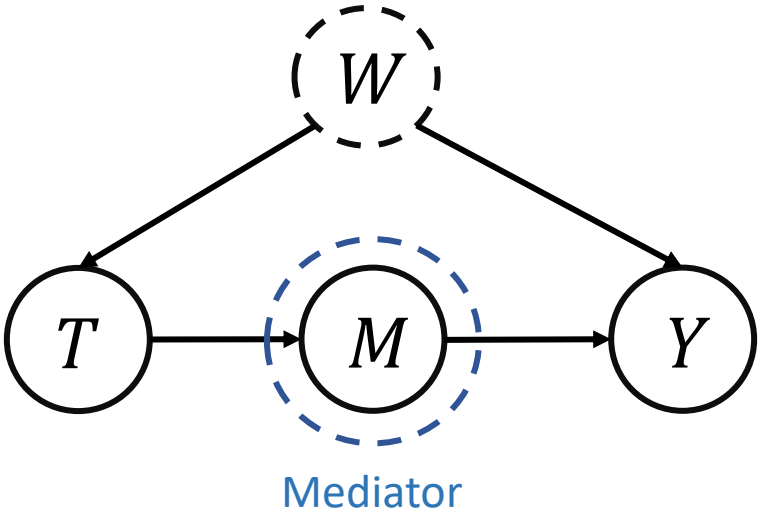
$P(m|do(t))$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

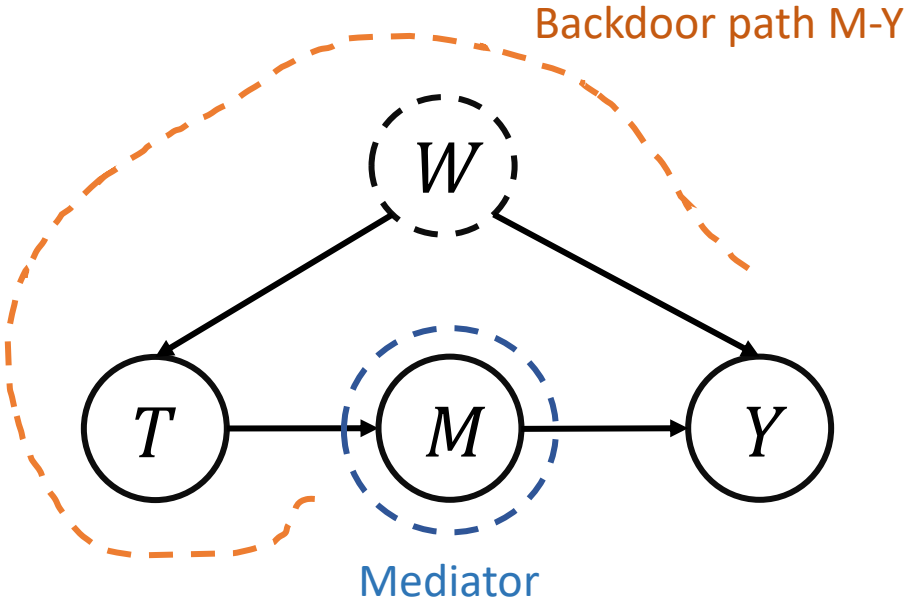
$$P(m|do(t)) = P(m|t)$$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

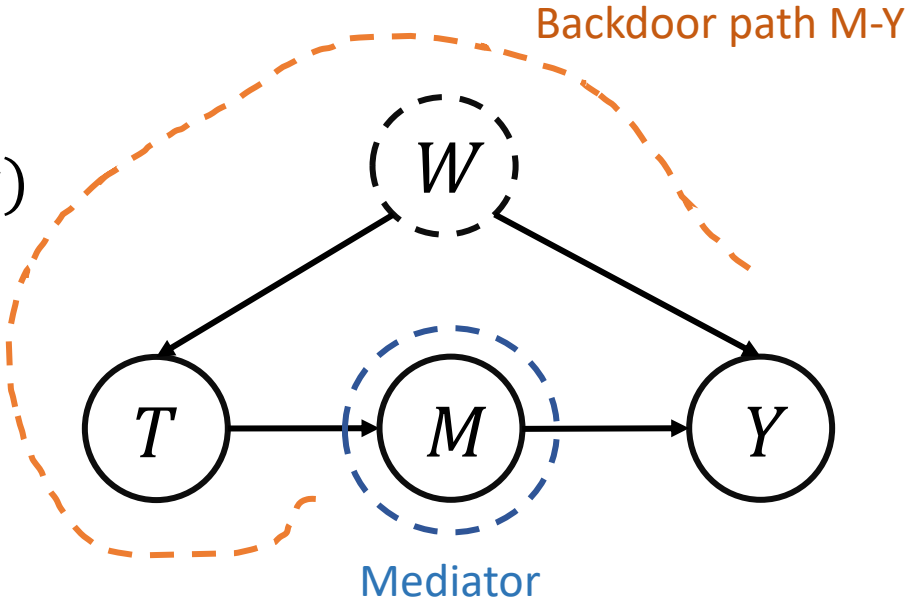
$P(y|do(m))$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

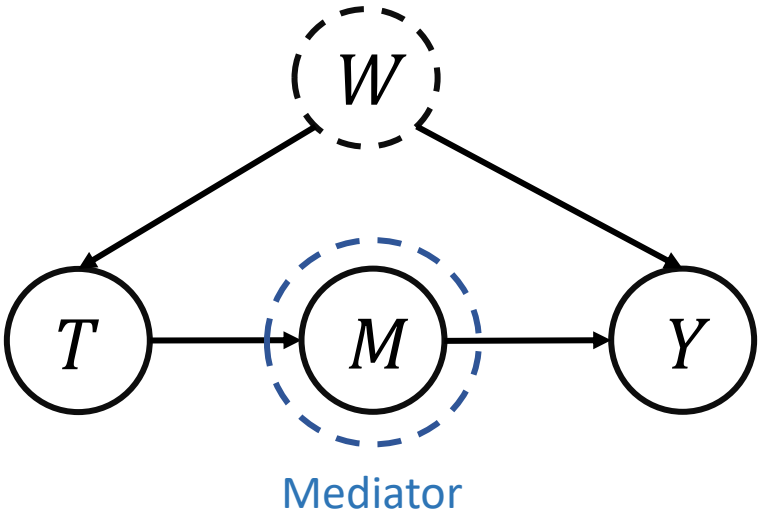
$$P(y|do(m)) = \sum_t P(y|m,t)P(t)$$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

$P(y|do(t))$

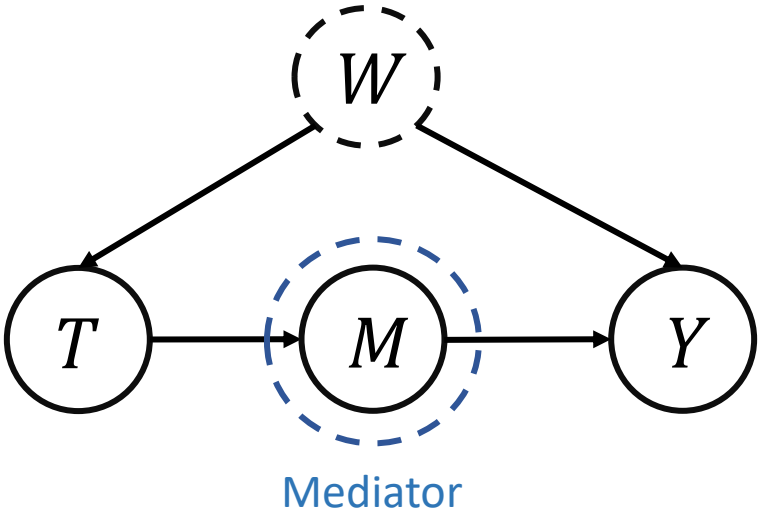


Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

$$P(y|do(t))$$

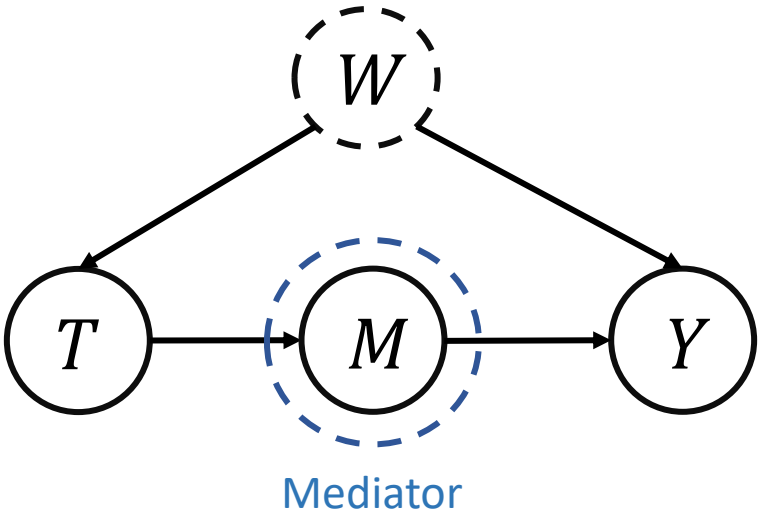
$$P(m|do(t))P(y|do(m))$$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

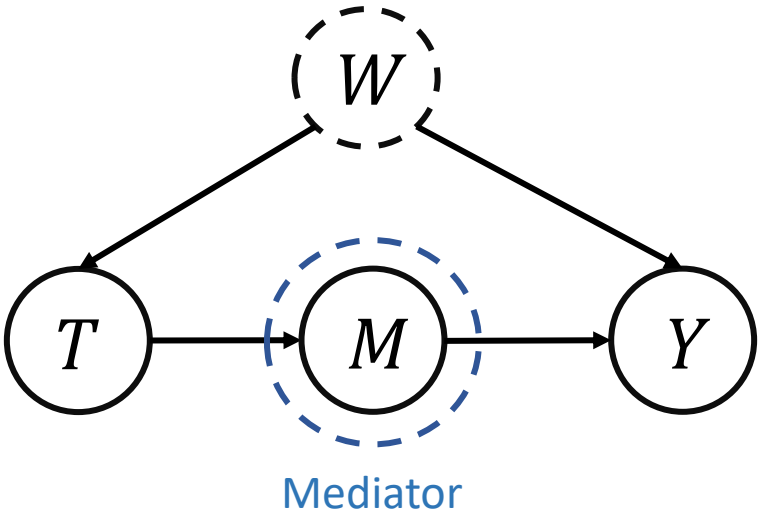
$$P(y|do(t)) = \sum_{m} P(m|do(t))P(y|do(m))$$



Frontdoor Adjustment

- Step 1. Identify the causal effect of T on M
- Step 2. Identify the causal effect of M on Y
- Step 3. Based on the above two steps, identify the causal effect of T on Y

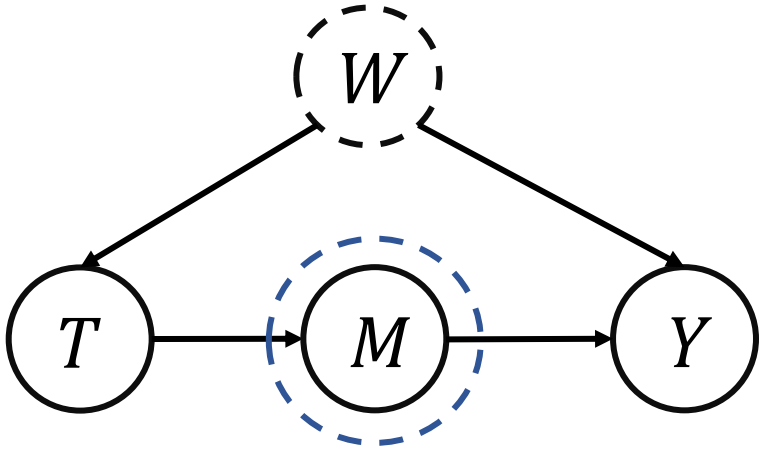
$$P(y|do(t))$$
$$= \sum_m P(m|do(t))P(y|do(m))$$
$$= \sum_m P(m|t) \sum_{t'} P(y|m,t')P(t')$$



Frontdoor Criterion

- A set of variables M satisfies the frontdoor criterion relative to T and Y if the following are true:
  - 1. M completely mediates the effect of T on Y (i.e., all causal paths from T to Y go through M).
  - There is no unblocked backdoor path from T to M.
  - All backdoor paths from M to Y are blocked by T.

$$P(y|do(t))$$
$$= \sum_m P(m|do(t))P(y|do(m))$$
$$= \sum_m P(m|t) \sum_{t'} P(y|m,t')P(t')$$



Can the causal effect be identified if neither the backdoor criterion nor the frontdoor criterion is satisfied?

Can the causal effect be identified if neither the backdoor criterion nor the frontdoor criterion is satisfied?

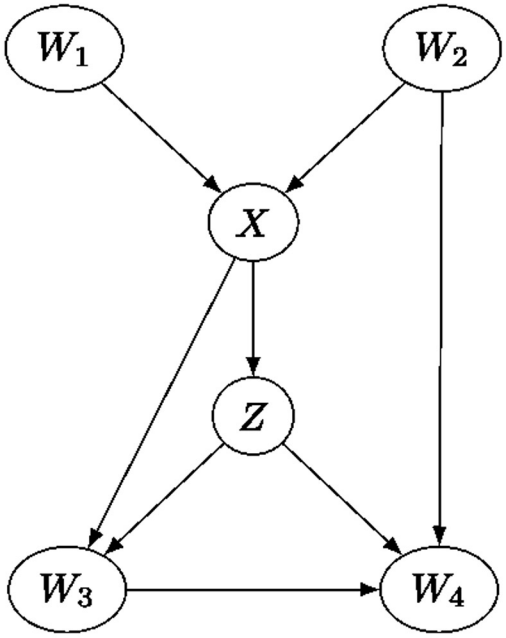
Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

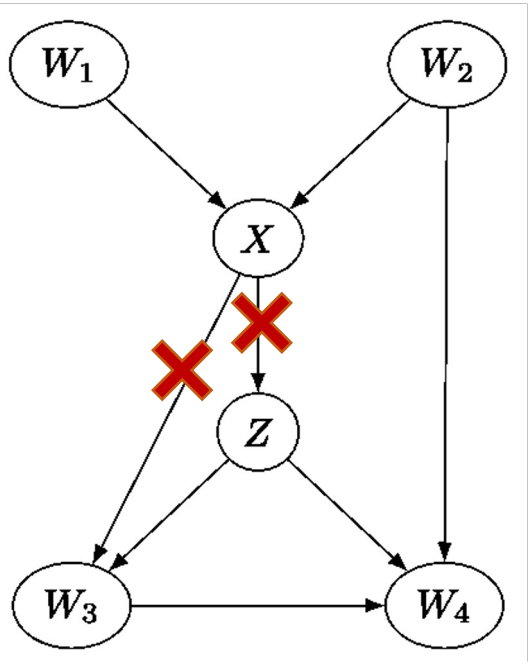
Do-calculus

- Question: how to identify causal quantities (with do operator)?
- The do-calculus is an axiomatic system for replacing probability formulas containing the do operator with ordinary conditional probabilities.
- 3 rules
  - Not limited to backdoor or frontdoor criterion, but the rules of do-calculus can be used to prove backdoor/frontdoor.

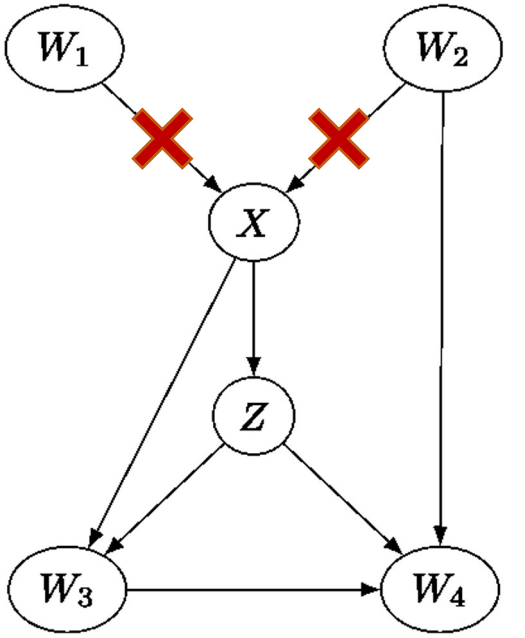
G



$G_{\underline{X}}$



$G_{\overline{X}}$



Do-calculus: Rule 1

- If $Y \perp\!\!\!\perp_{G_{\overline{T}}} Z | T, W$:

$$P(y|do(t), z, w) = P(y|do(t), w)$$

Do-calculus: Rule 1

- If $Y \perp\!\!\!\perp_{G_{\overline{T}}} Z \mid T, W$:

$$P(y|do(t), z, w) = P(y|do(t), w)$$

Interventional distribution w.r.t. $do(t)$

Do-calculus: Rule 1

- If $Y \perp\!\!\!\perp_{G_{\overline{T}}} Z | T, W$:

$$P(y|do(t), z, w) = P(y|do(t), w)$$

Interventional distribution w.r.t. $do(t)$

When we remove $do(t)$:

- If $Y \perp\!\!\!\perp_{G} Z | W$:

D-separation

$$P(y|do(t), z, w) = P(y|do(t), w)$$

- Rule 1 is a generalization of d-separation to interventional distribution

Do-calculus: Rule 2

- If $Y \perp\!\!\!\perp_{G_{\overline{T},\underline{Z}}} Z | T, W$:

$$P(y|do(t), do(z), w) = P(y|do(t), z, w)$$

Do-calculus: Rule 2

- If $Y \perp\!\!\!\perp_{G_{\overline{T},\underline{Z}}} Z | T, W$:

$$P(y|do(t), do(z), w) = P(y|do(t), z, w)$$

When we remove $do(t)$:

- If $Y \perp\!\!\!\perp_{G_{\underline{Z}}} Z | W$:

$$P(y|do(t), do(z), w) = P(y|do(t), z, w)$$

- Rule 2 is a generalization of backdoor adjustment/criterion

Do-calculus: Rule 3

- If $Y \perp\!\!\!\perp_{G_{\bar{T},\overline{Z(W)}}} Z \mid T, W$

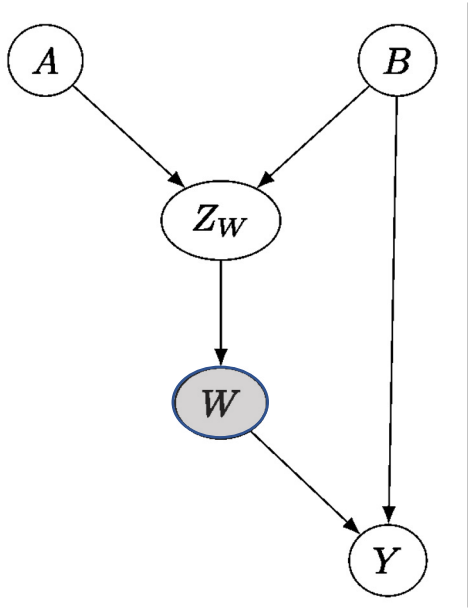
$$P(y|do(t), do(z), w) = P(y|do(t), w)$$

$Z(W)$: the set of nodes of Z that aren't ancestors of any node of W in $G_{\bar{T}}$

When we remove $do(t)$:

- If $Y \perp\!\!\!\perp_{G_{\overline{Z(W)}}} Z \mid T, W$

$$P(y|do(t), do(z), w) = P(y|do(t), w)$$



3 Rules in do-calculus

- **Rule 1:** If $Y \perp\!\!\!\perp_{G_{\overline{T}}} Z \mid T, W$:

$$P(y|do(t), z, w) = P(y|do(t), w)$$

- **Rule 2**: If $Y \perp\!\!\!\perp_{G_{\overline{T},\underline{Z}}} Z \mid T, W$:

$$P(y|do(t), do(z), w) = P(y|do(t), z, w)$$

- **Rule 3**: If $Y \perp\!\!\!\perp_{G_{\overline{T},\overline{Z(W)}}} Z \mid T, W$

$$P(y|do(t), do(z), w) = P(y|do(t), w)$$

Identification in SCM: Use the 3 rules, until there is no do-operator.

Completeness of do-calculus

- Are there identifiable causal quantities that cannot be identified with do-calculus?

Completeness of do-calculus

- Are there identifiable causal quantities that cannot be identified with do-calculus?

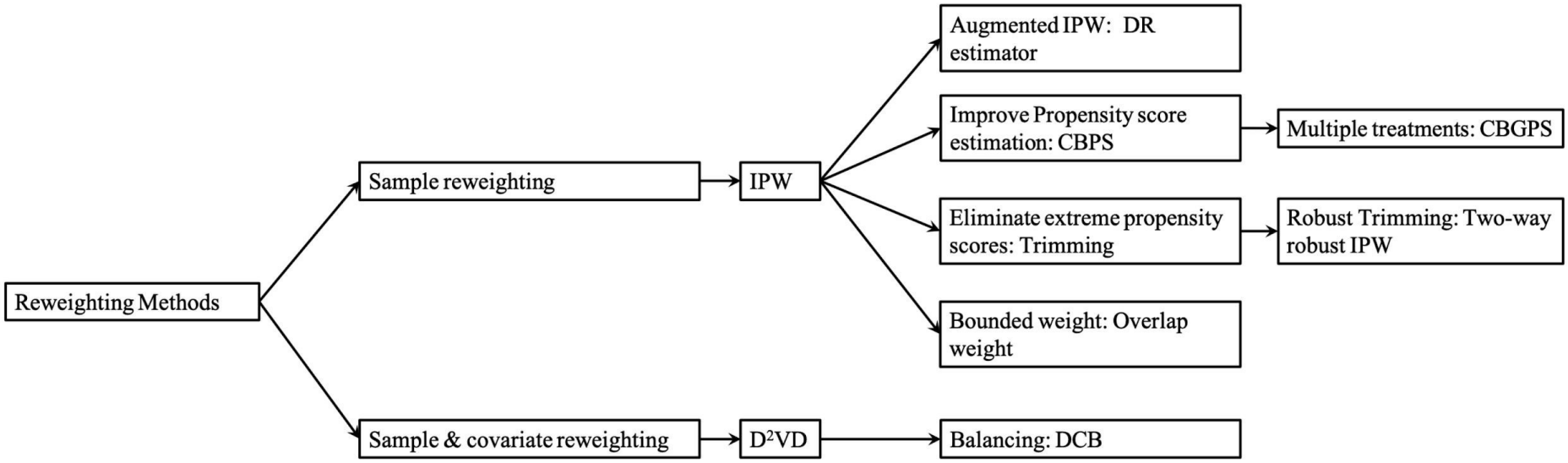
Do-calculus is complete [1]

[1] Huang Y, Valtorta M. Pearl's calculus of intervention is complete[J]. arXiv preprint arXiv:1206.6831, 2012.

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
  - Identification from graph
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Summary of Re-weighting Methods



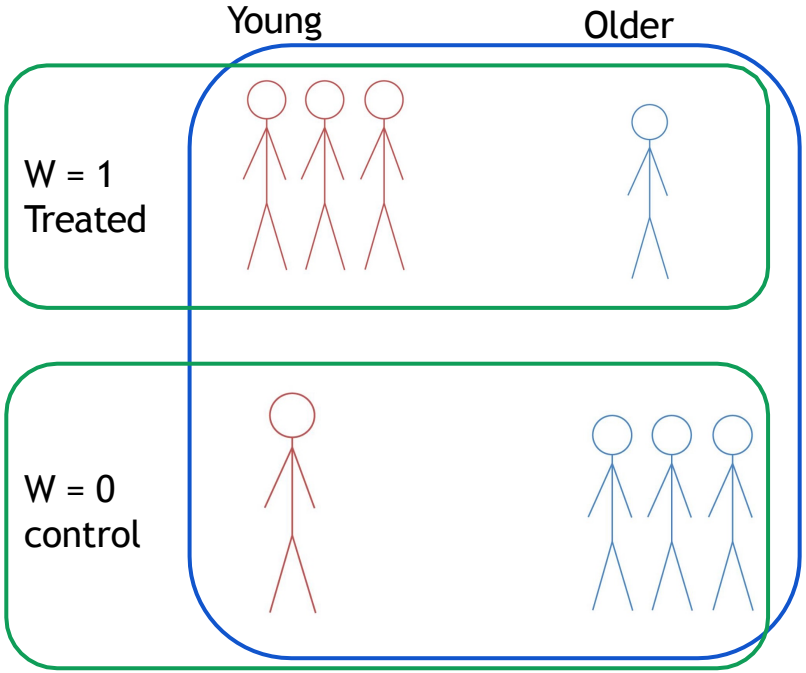
Yao L, Chu Z, Li S, et al. A survey on causal inference[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021, 15(5): 1-46.

Re-weighting Methods

- **Challenge of causal effect estimation**: different distributions of treated and control groups bring bias
- Sample re-weighting is a common way to overcome the selection bias problem
- **Idea**: By *assigning appropriate weight to each sample* in the observation dataset, a pseudo-population is created on which the distributions of the treated group and control group are similar

Sample Re-weighting Methods

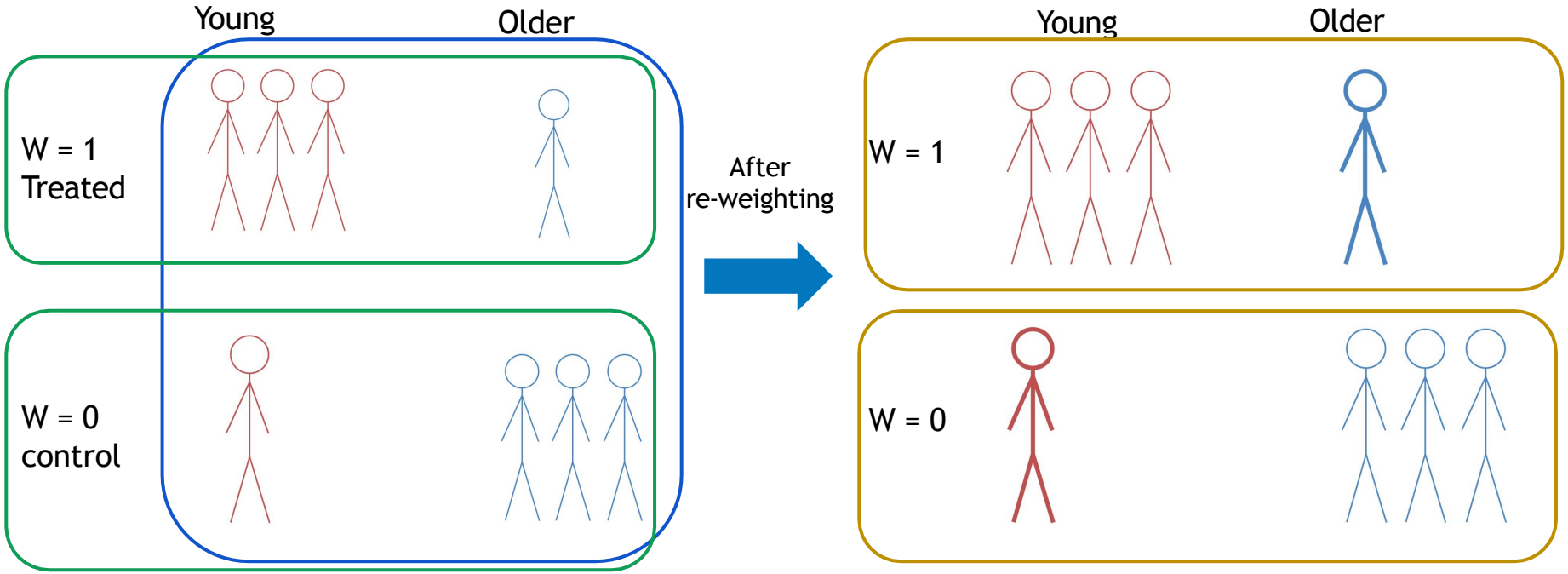
- Intuition example: **Age** (Young/older) as the confounder



http://www.rebeccabarter.com/blog/2017-07-05-ip-weighting/

Sample Re-weighting Methods

- Intuition example: **Age** (Young/older) as the confounder



http://www.rebeccabarter.com/blog/2017-07-05-ip-weighting/

Sample Re-weighting Methods

- **Balancing Score**: Balancing score $b(\boldsymbol{X})$ is a general weighting score, which is the function of covariates $\boldsymbol{X}$ satisfying: $W \perp\!\!\!\perp \boldsymbol{x} | \, b(\boldsymbol{x})$.
  - $b(\boldsymbol{x})$ contains all information about treatment assignment
  - Able to approximate the whole population using balancing score
- One representative method: **Propensity Score based Re-weighting**
- **Propensity Score**: Conditional probability of assignment to a particular treatment given a vector of observed covariates

$$e(x) = P(W = 1 | X = x)$$

Sample Re-weighting Methods

- **Inverse propensity weighting (IPW)**
- The weight assigned for each unit is:

$$r = \frac{W}{e(x)} + \frac{1-W}{1-e(x)}$$

$W = 1 \Rightarrow r = \frac{1}{e(x)}$

$W = 0 \Rightarrow r = \frac{1}{1-e(x)}$

where W is the treatment and $e(x)$ is the propensity score

- After re-weighting, the IPW estimator of ATE is defined as:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

  - Theoretical results show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates.
  - However, IPW highly relies on the correctness of propensity scores

J. Robins, A. Rotnitzky, and L. Zhao. "Estimation of regression coefficients when some regressors are not always observed." *Journal of the American statistical Association* 89.427 (1994): 846-866.

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$

$E[Y(1)] =$

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

$$E[Y(1)] = E_X[E[Y(1)|X]]$$ law of total expectation

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

$$E[Y(1)] = E_X[E[Y(1)|X]] = E_X\left[\frac{E[Y(1)|X]E[W|X]}{e(X)}\right]$$

Definition of propensity score

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

$$E[Y(1)] = E_X[E[Y(1)|X]] = E_X\left[\frac{E[Y(1)|X]E[W|X]}{e(X)}\right]$$

$$= E_X\left[E[\frac{Y(1)W}{e(X)}|X]\right]$$

$W \perp\!\!\!\perp Y(1) \mid X$

Assumption: no unmeasured confounders

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

$$E[Y(1)] = E_X[E[Y(1)|X]] = E_X\left[\frac{E[Y(1)|X]E[W|X]}{e(X)}\right]$$

$$= E_X\left[E[\frac{Y(1)W}{e(X)}|X]\right] = E_X\left[E\left[\frac{YW}{e(X)}\middle|X\right]\right]$$

$W = 0 \Rightarrow\ Y(1)W = YW = 0;$
$W = 1 \Rightarrow Y(1)W = YW$

IPW is Unbiased

- Estimator of IPW:

$$\hat{\mathrm{ATE}}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\frac{W_i Y_i^F}{\hat{e}(x)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)Y_i^F}{1-\hat{e}(x)}$$

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

$$E[Y(1)] = E_X[E[Y(1)|X]] = E_X\left[\frac{E[Y(1)|X]E[W|X]}{e(X)}\right]$$

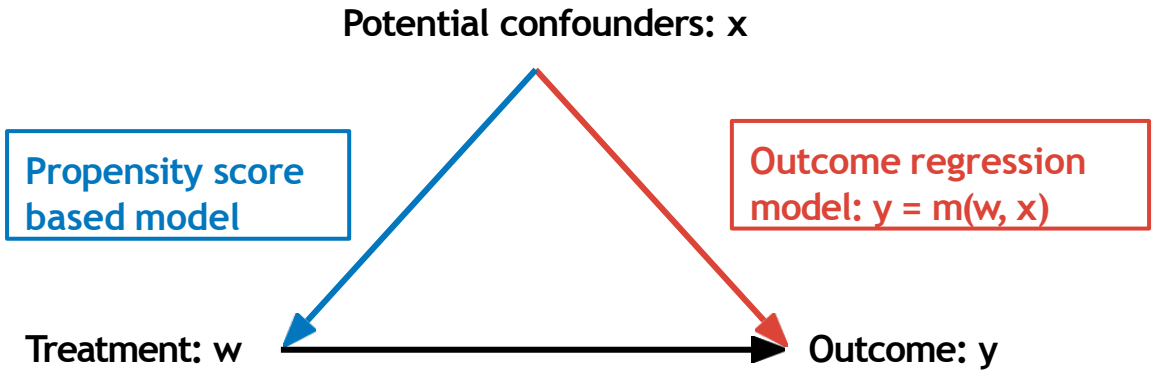
$$= E_X\left[E[\frac{Y(1)W}{e(X)}|X]\right] = E_X\left[E\left[\frac{YW}{e(X)}\middle|X\right]\right] = E\left[\frac{YW}{e(X)}\right]$$

IPW

- Theoretical results show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates.
- However, IPW highly relies on the correctness of propensity scores

Sample Re-weighting Methods

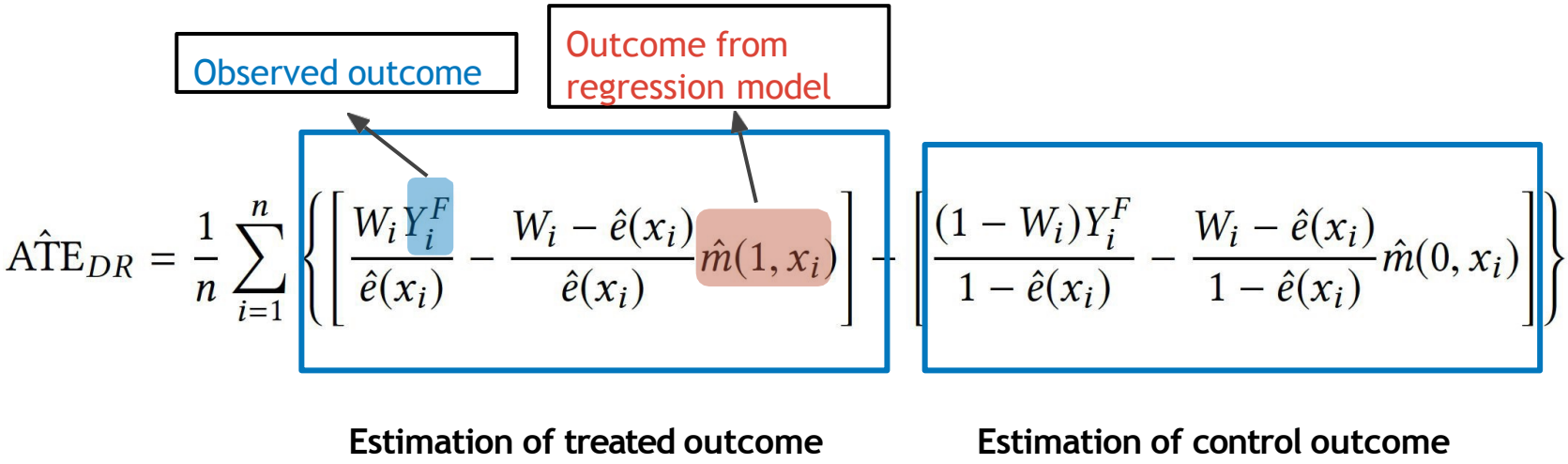
- **Doubly Robust Estimator (DR)** or **Augmented IPW**
- It combines the propensity score weighting with the outcome regression



J. Robins, A. Rotnitzky, and L. Zhao. "Estimation of regression coefficients when some regressors are not always observed." *Journal of the American statistical Association* 89.427 (1994): 846-866.

Sample Re-weighting Methods

- **Doubly Robust Estimator (DR)** or **Augmented IPW**
  - Unbiased when one of the propensity score or outcome regression is correct



Sample Re-weighting Methods

- **Doubly Robust Estimator (DR)** or **Augmented IPW**
  - Unbiased when one of the propensity score or outcome regression is correct

$$\hat{\mathrm{ATE}}_{DR} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{W_i Y_i^F}{\hat{e}(x_i)} - \frac{W_i - \hat{e}(x_i)}{\hat{e}(x_i)}\hat{m}(1,x_i)\right] - \left[\frac{(1-W_i)Y_i^F}{1-\hat{e}(x_i)} - \frac{W_i - \hat{e}(x_i)}{1-\hat{e}(x_i)}\hat{m}(0,x_i)\right]\right\}$$

IPW augmented IPW augmented

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
  - Identification from graph
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Stratification

- **Stratification** adjusts the selection bias by splitting the entire group into subgroups, where within each subgroup, the treated group and the control group are similar under some measurements
- Stratification is also named as ***subclassification*** or ***blocking***
- ATE for stratification is estimated as

$$\hat{\tau}^{\text{strat}} = \sum_{j=1}^{J} q(j) \left[\bar{Y}_t(j) - \bar{Y}_c(j)\right]$$

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
  - Identification from graph
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Matching

- Matching methods estimate the counterfactuals and meanwhile reduce the estimation bias brought by the confounders
- Potential outcomes of the i-th unit estimated by matching are:

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0, \\ \frac{1}{\#\mathcal{J}(i)} \sum_{l \in \mathcal{J}(i)} Y_l & \text{if } W_i = 1; \end{cases}$$

$$\hat{Y}_i(1) = \begin{cases} \frac{1}{\#\mathcal{J}(i)} \sum_{l \in \mathcal{J}(i)} Y_l & \text{if } W_i = 0, \\ Y_i & \text{if } W_i = 1; \end{cases}$$

Where $J(i)$ is the matched neighbors of unit i in the opposite treatment group

Matching

- **Distance Metrics for Matching**
  - *Original Data Space*
    - Euclidean distance
    - Mahalanobis distance
  - *Transformed Feature Space*
    - Propensity score based transformation
    - Other transformations (e.g., prognosis score)

Matching

- **Propensity Score Matching (PSM)**
  - Propensity scores denote conditional probability of assignment to a particular treatment given a vector of observed covariates.

$$e(x) = Pr(W = 1|X = x)$$

  - Based on propensity scores, the distance between two units is

$$D(\mathbf{x}_i, \mathbf{x}_j) = |e_i - e_j|$$

  - Alternatively, linear propensity score based distance metric

$$D(\mathbf{x}_i, \mathbf{x}_j) = |\text{logit}(e_i) - \text{logit}(e_j)|$$

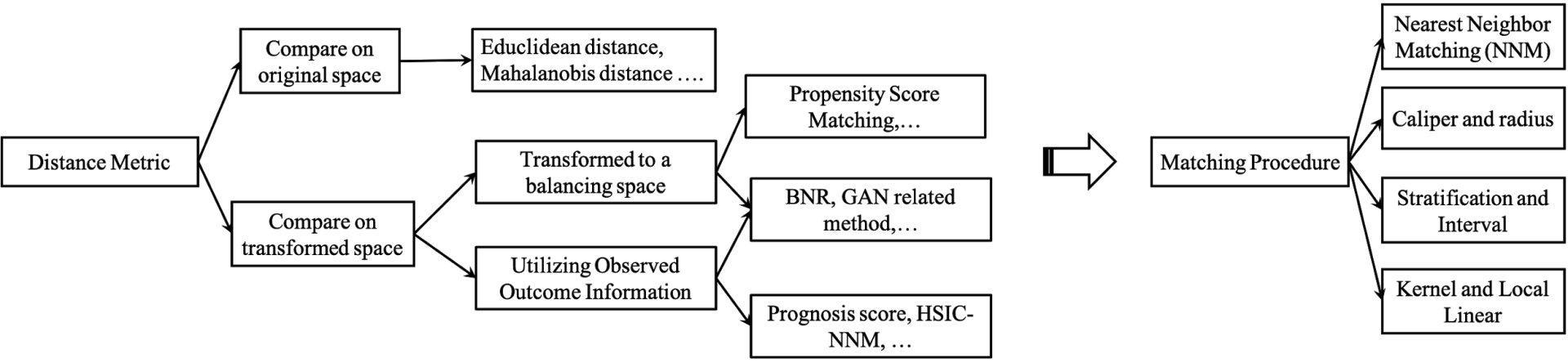
P. Rosenbaum, and D. Rubin. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70.1 (1983): 41-55.

Matching

- **Choosing Matching Algorithm**
  - Nearest Neighbors
  - Caliper
  - Stratification
  - Kernels
- **Variable Selection**
  - The more, the better?
  - Post-treatment variables

Matching

❑ Summary of Matching Methods



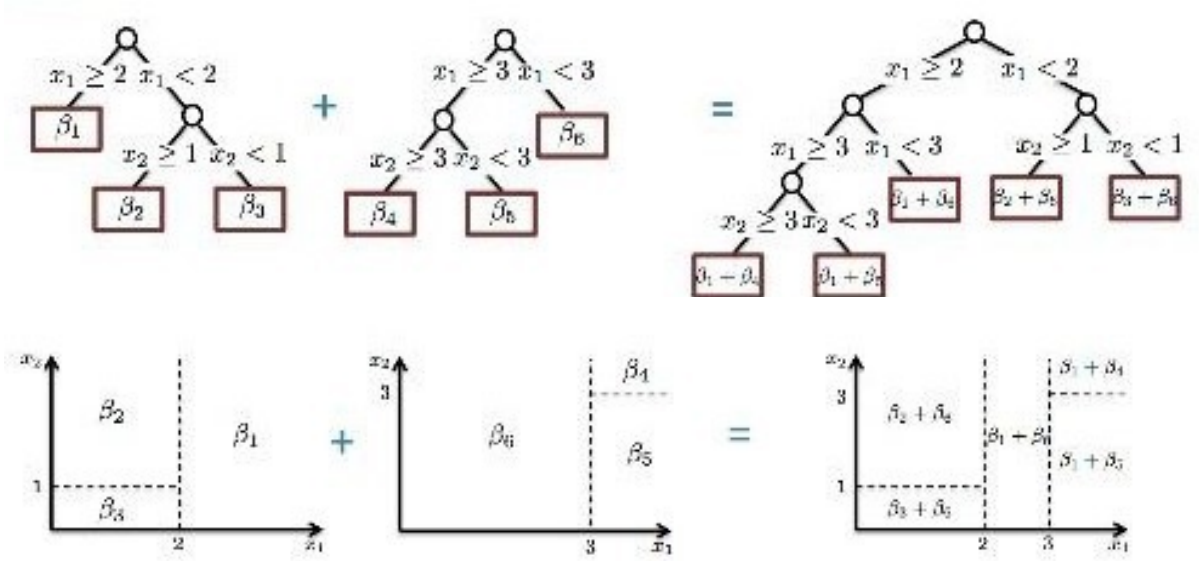
Yao L, Chu Z, Li S, et al. A survey on causal inference[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021, 15(5): 1-46.

Outline

- Identifiability
  - Randomized experiments
  - Front door
  - Do-calculus
  - Identification from graph
- Classical causal effect estimation methods
  - Re-weighting methods
  - Stratification methods
  - Matching methods
  - Tree-based methods

Tree-based Methods

- **Bayesian Additive Regression Trees (BART)**
  - A Bayesian "sum-of-trees" model
  - Nonparametric Bayesian regression model



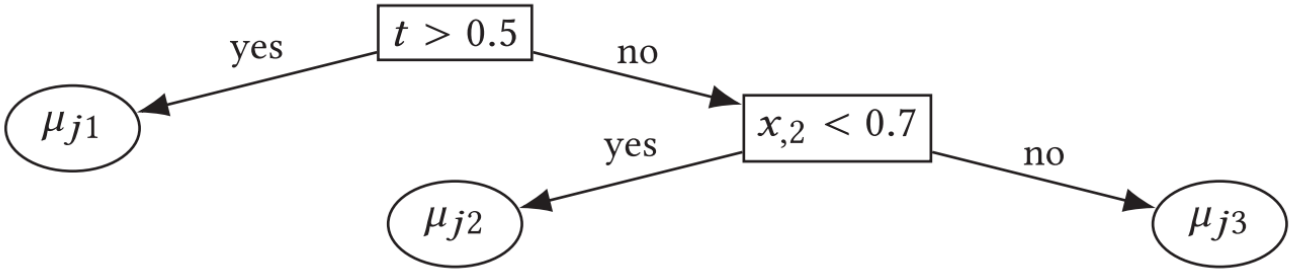
Hill, Jennifer L. "Bayesian nonparametric modeling for causal inference." *Journal of Computational and Graphical Statistics* 20.1 (2011): 217-240.
https://www.slideshare.net/SAMSI_Info/mums-bayesian-fiducial-and-frequentist-conference-multiscale-analysis-of-bart-veronika-rockova-april-29-2019

Tree-based Methods

- **Bayesian Additive Regression Trees (BART) Formulation:**
  - BART takes the features and the treatment as input and output the distribution of potential outcomes as

$$f(x,t) = \sum_j g_j(x,t) \qquad j\text{: index of Bayesian regression tree}$$

  - CATE for x can be estimated by $f(x,1) - f(x,0)$



Example of a subtree $g(x,t)$ in BART

Tree-based Methods

- Advantages of BART:
  - Easy to implement. Less requirement for parameter tuning
  - Posterior can provide uncertainty of the estimation
  - BART can deal with a mass of predictors and handle continuous treatment variables and missing data

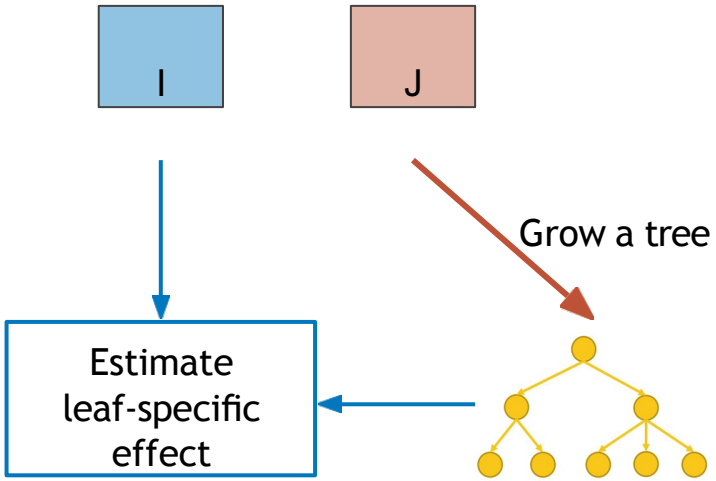
Tree-based Methods

- **Causal Forests**
  - Single tree is noisy -> using forest
  - It is based on Breiman's random forest algorithm
  - Trees and forests help find neighbors adaptively
  - Extended to multiple treatments

S. Wager, and S. Athey. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113.523 (2018): 1228-1242.

Tree-based Methods

- **Causal Forest:**
- Single tree as the Double sample tree
  - Split the sample into I samples and J samples
  - **Goal**: estimate the outcome

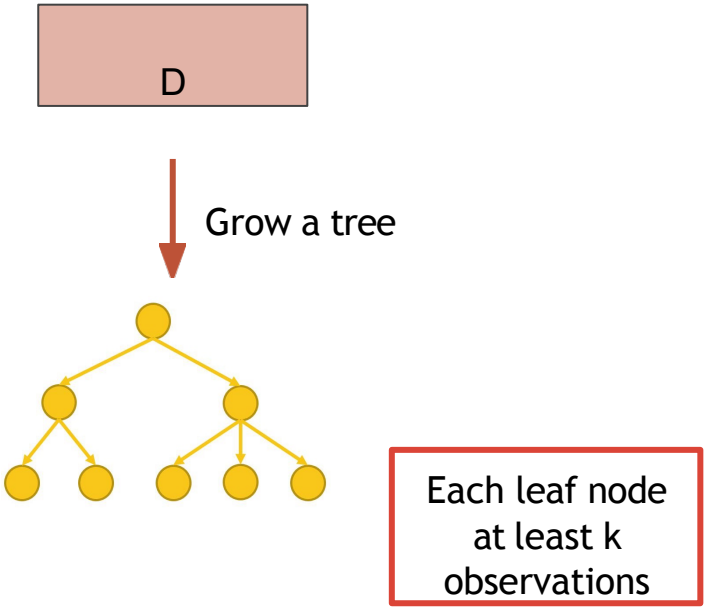


Advantage:

- Can estimate CATE
- Consistent to true CATE
- Nice Asymptotic properties

Tree-based Methods

- **Causal Forest:**
- Single tree as the propensity tree
  - **Goal**: estimate the treatment assignment W
- Use full samples



Summary

- **Classical Causal Inference Methods**
  - Simple methods with theoretical guarantee
  - May not be sufficient to handle high dimension data

Reading Materials

- Guo R, Cheng L, Li J, et al. A survey of learning causality with data: Problems and methods[J]. ACM Computing Surveys (CSUR), 2020, 53(4): 1-37.
- Yao L, Chu Z, Li S, et al. A survey on causal inference[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021, 15(5): 1-46.

Thank you!
Q&A

jing.ma5@case.edu