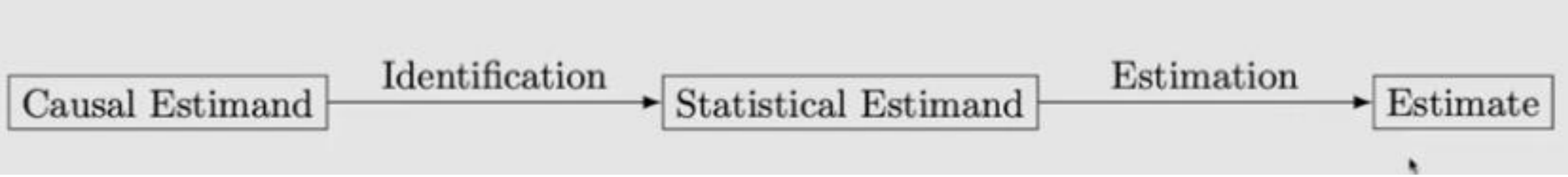
Causal Inference

Week6 Estimator

https://youtu.be/YzcOYU-s2t4

210623 전은주

Estimation portion of the flowchart



- Identification: Graph가지고 back-door, front-door path 막아서 flow 안 흐르게 하고, causal effect만 남기는
- Estimation: Estimate 숫자 값 추출하는 모델

Preliminaries

Conditional average treatment effects (CATEs):

$$\tau(x) \triangleq \mathbb{E}[Y(1) - Y(0) \mid X = x]$$

Tau(x), Y(1): Treatment group, Y(0): Controlled group, X는 변수

Always assuming unconfoundedness and positivity

Identification해서, W를 conditioning (W는 sufficient adjustment set)

$$\tau \triangleq \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_W\left[\mathbb{E}[Y \mid T = 1, W] - \mathbb{E}[Y \mid T = 0, W]\right]$$

Given W is a sufficient adjustment set

$$\tau(x) \triangleq \mathbb{E}[Y(1) - Y(0) \mid X = x] = \mathbb{E}_W[\mathbb{E}[Y \mid T = 1, X = x, W] - \mathbb{E}[Y \mid T = 0, X = x, W]]$$

Given $W \cup X$ is a sufficient adjustment set

X변수와 W 둘 다 conditioning
- X변수가 모든 측정 변수일 필요는 없지만, 그렇다면 Individual CATE

Contents

1. Conditional Outcome Modeling
2. Increasing Data Efficiency
3. Propensity Score and IPW (Inverse Probability Weighting)
4. Other Methods

1. Conditional outcome modeling (COM)

(좌) Causal estimate , (우) Statistical estimate

$$\tau = \mathbb{E}_W\left[\mathbb{E}[Y \mid T = 1, W] - \mathbb{E}[Y \mid T = 0, W]\right]$$

Model은 ML, DL 모두 가능

$$\tau = \mathbb{E}_W\left[\overbrace{\mu(1, W)}^{\text{model}} - \overbrace{\mu(0, W)}^{\text{model}}\right]$$

Model-assisted estimator:

W에 대한 Estimation 이니까,
Wi에 대한 mean 값으로 변경

$$\hat{\tau} = \frac{1}{n}\sum_i \left(\hat{\mu}(1, w_i) - \hat{\mu}(0, w_i)\right)$$

1. COM estimation of CATEs

ATE COM Estimator: $\hat{\tau} = \frac{1}{n}\sum_i (\hat{\mu}(1, w_i) - \hat{\mu}(0, w_i))$

CATE Estimand

$\tau(x) \triangleq \mathbb{E}[Y(1) - Y(0) \mid$ … $Y \mid T = 0, X = x, W]]$

COM estimation's many faces

- G-computation estimators
- Parametric G-formula
- Standardization
- S-learner where "S" is for "Single"

CATE COM Estimator: $\hat{\tau}(x) = \frac{1}{n_x}\sum_{i:x_i=x} (\hat{\mu}(1, w_i, x) - \hat{\mu}(0, w_i, x))$

X도 하나의 input으로 들어감. Sigma xi

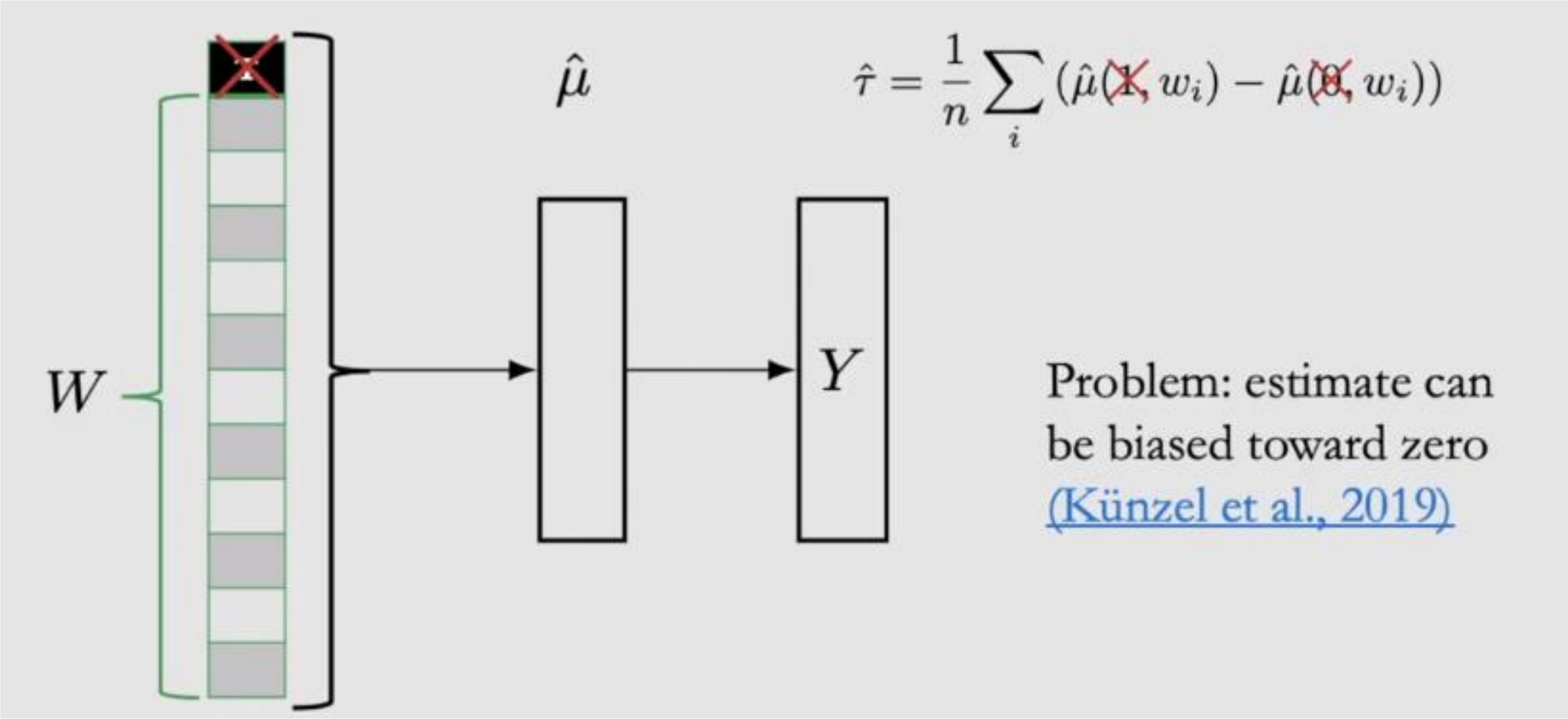
$\hat{\tau}_i = \hat{\tau}(x_i) = \hat{\mu}(1, w_i, x_i) - \hat{\mu}(0, w_i, x_i)$ Individual sample CATE

1. COM estimation of CATEs

COM estimation's many faces

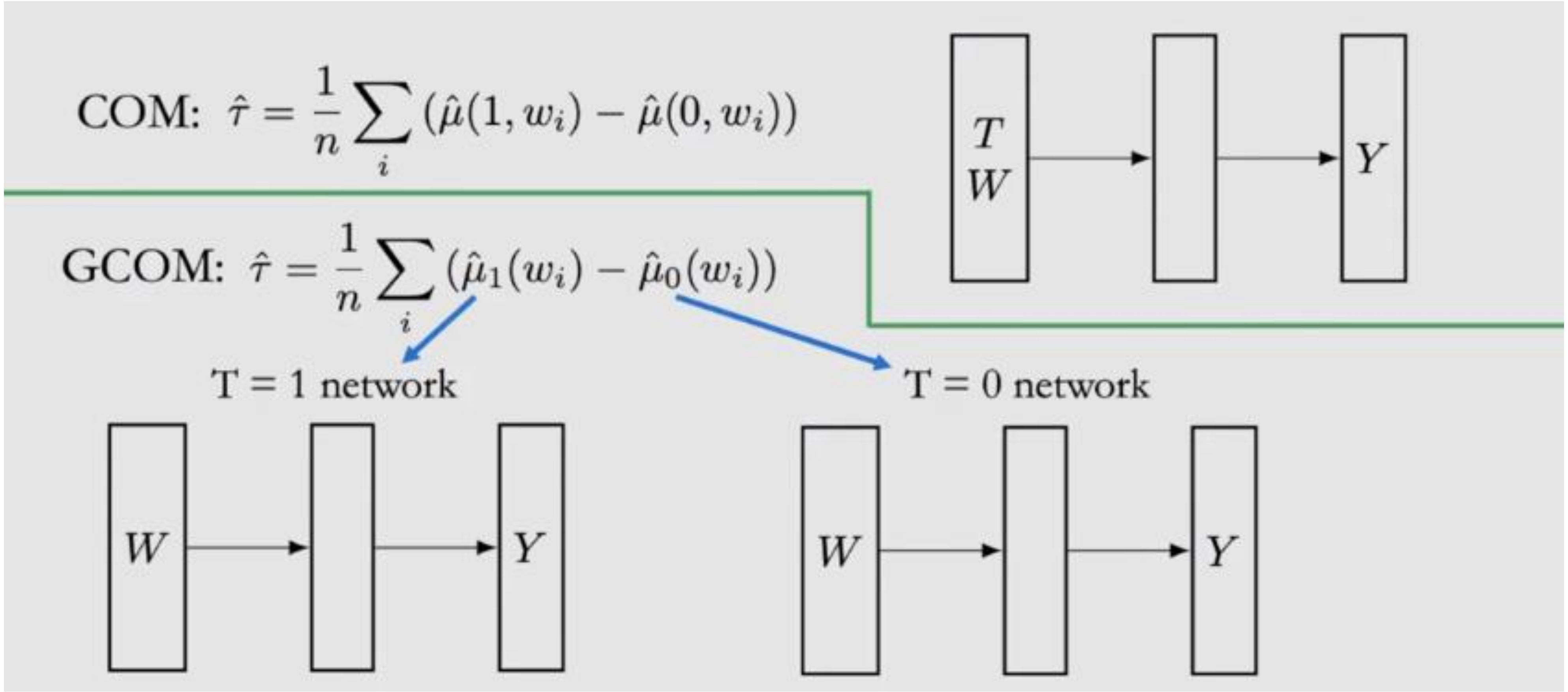
- G-computation estimators
- Parametric G-formula
- Standardization
- S-learner where "S" is for "Single"

1. Problem with COM estimation in high dimensions



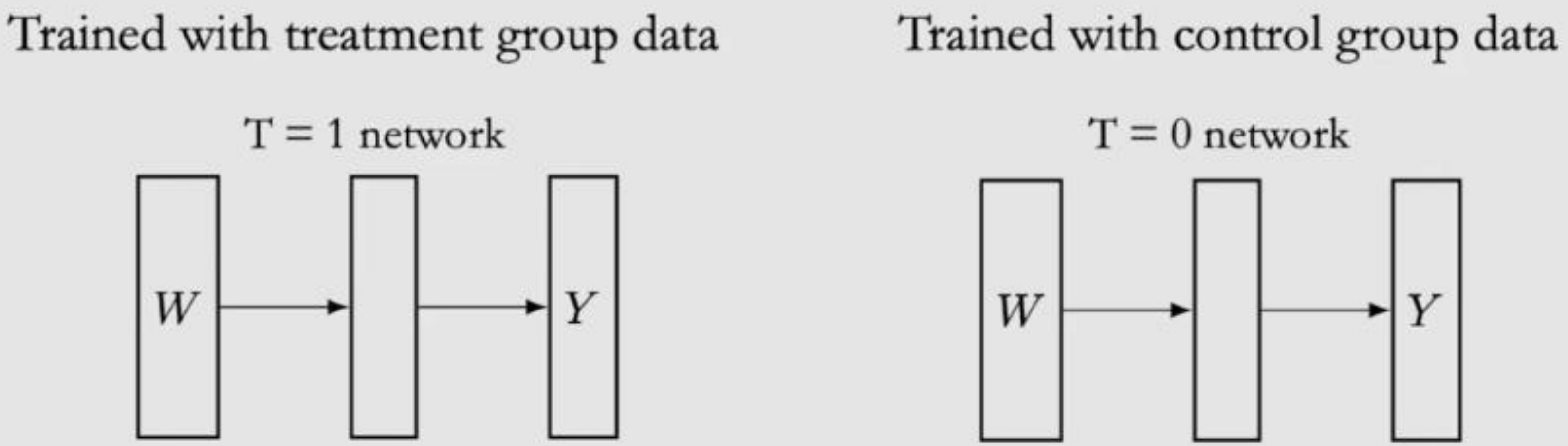
W 에 비해서 T (1,0)이 1dimension으로 너무 작으니까, 모델 (DL, ML)이 Treatment를 무시하게 된다.

1. Grouped COM (GCOM) estimation



따라서, T=1, T=0인 Group을 나눠서 CATE계산.
단점. 데이터가 반으로 잘라진다.

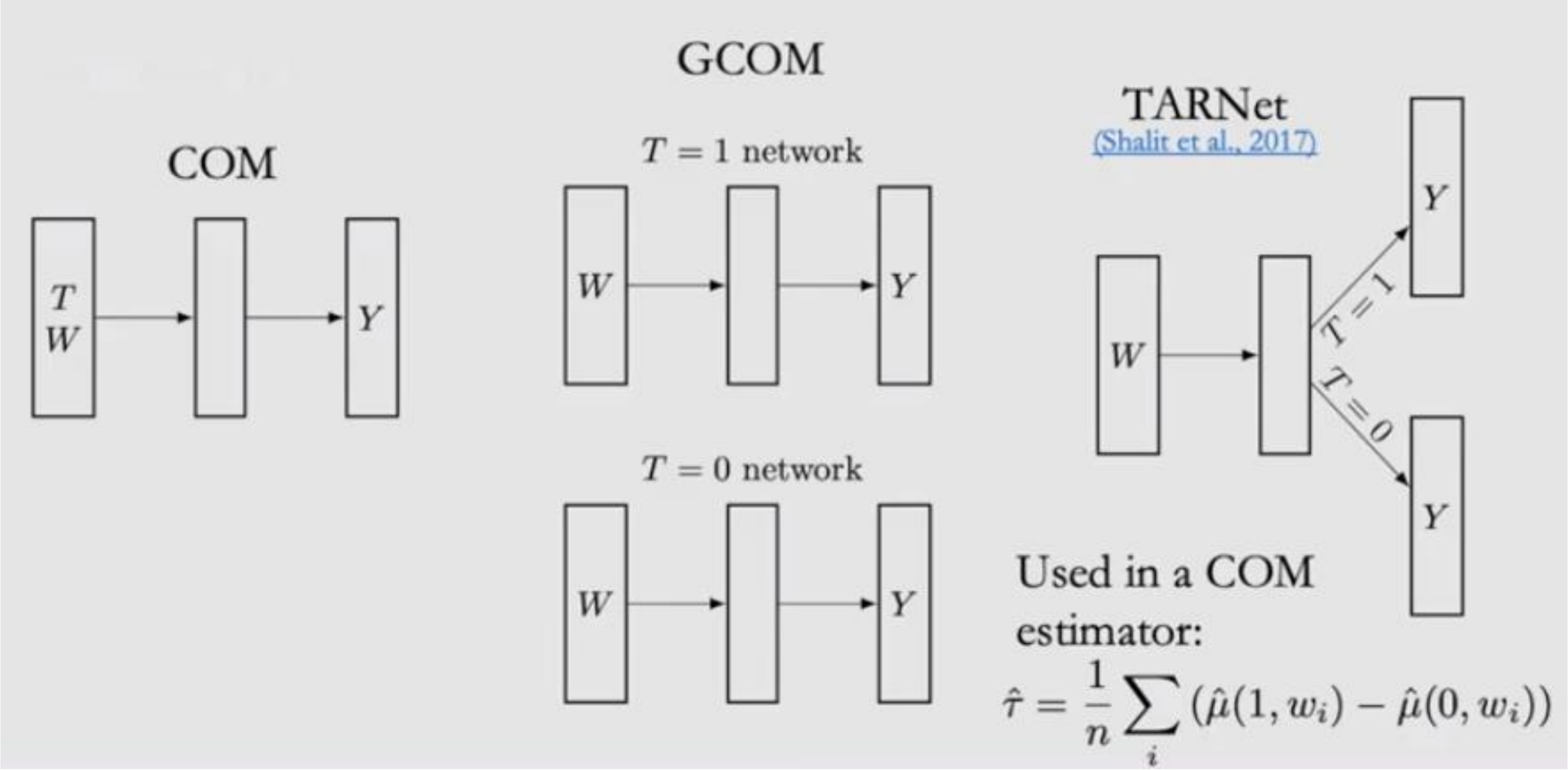
1. Grouped COM (GCOM) estimation



Problem: networks have higher variance than they would if they were trained with all the data (not efficient)

따라서, T=1, T=0인 Group을 나눠서 CATE계산.
단점. 데이터가 반으로 잘라진다.

2. Increasing Data Efficiency

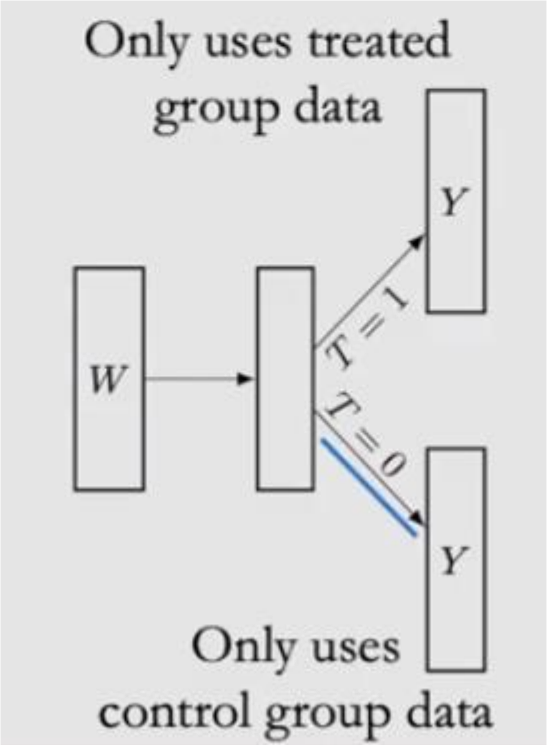


Too biased!

Too much variance

All data 로 training
T=1, T=0으로 나뉨 (하지만 사실 나뉠 때 데이터 ½로 나뉘는건 여전함)

TARNet inefficiency



All data 로 training
T=1, T=0으로 나뉨 (하지만 사실 나뉠
때 데이터 ½로 나뉘는건 여전함)

Metalearner

X-Learner (Kunzel, 2019)

Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the national academy of sciences* 116.10 (2019): 4156-4165.

1. Estimate $\hat{\mu}_1(x)$ and $\hat{\mu}_0(x)$ — Assume X is a sufficient adjustment set and is all observed covariates

ITE (Individual Treatment Effect) 를 Treatment & Control group 따로따로 구한다. (모든 데이터를 활용해서)

2a. Impute ITEs

Treatment group: $\hat{\tau}_{1,i} = Y_i(1) - \hat{\mu}_0(x_i)$

Control group: $\hat{\tau}_{0,i} = \hat{\mu}_1(x_i) - Y_i(0)$

Y(1)은 observed data, μ 는 모델의 estimate

2b. Fit a model $\hat{\tau}_1(x)$ to predict $\hat{\tau}_{1,i}$ from x_i in treatment group

Fit a model $\hat{\tau}_0(x)$ to predict $\hat{\tau}_{0,i}$ from x_i in control group

3. $$\hat{\tau}(x) = g(x)\,\hat{\tau}_0(x) + (1 - g(x))\,\hat{\tau}_1(x)$$

where $g(x)$ is some weighing function between 0 and 1. Example: propensity score

3. Propensity Score and IPW (Inversed Probability weighting)

Propensity score(PS)
- 관찰 대상자가 가지고 있는 여러가지 특성(변수)을 고려하여 실험군과 대조군으로 구분할 때 각 집단으로 배정될 조건부 확률
- 다시 말하면 관찰된 특성변수들의 집합을 가진 어떤 개체가 treated될 확률로 정의할 수 있다

$$e(W) \triangleq P(T = 1 \mid W)$$

Given positivity, unconfoundedness given W implies unconfoundedness given the propensity score $e(W)$.

Even if W is high-dimensional, $e(W)$ is only 1-dimensional!

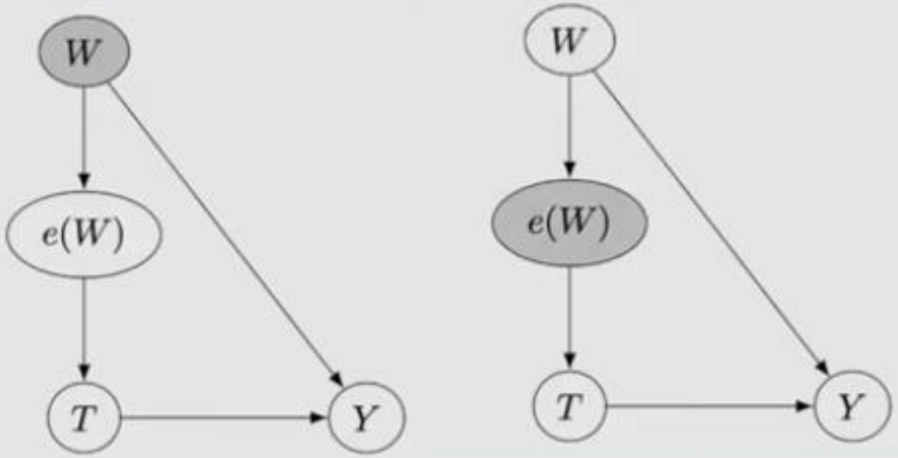
3. Propensity Score and IPW (Inversed Probability weighting)

Propensity Score Theorem

Given positivity, unconfoundedness given W implies unconfoundedness given the propensity score $e(W)$. Equivalently,

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid W \implies (Y(1), Y(0)) \perp\!\!\!\perp T \mid e(W)$$

Graphical Proof:



W to T는 P(T|W) -> T가 binary면, P(T=1|W)로 변경 가능 (P=0|W)는 1에서 빼면 구해지니까.

P(T=1|W) == PS, e(W)

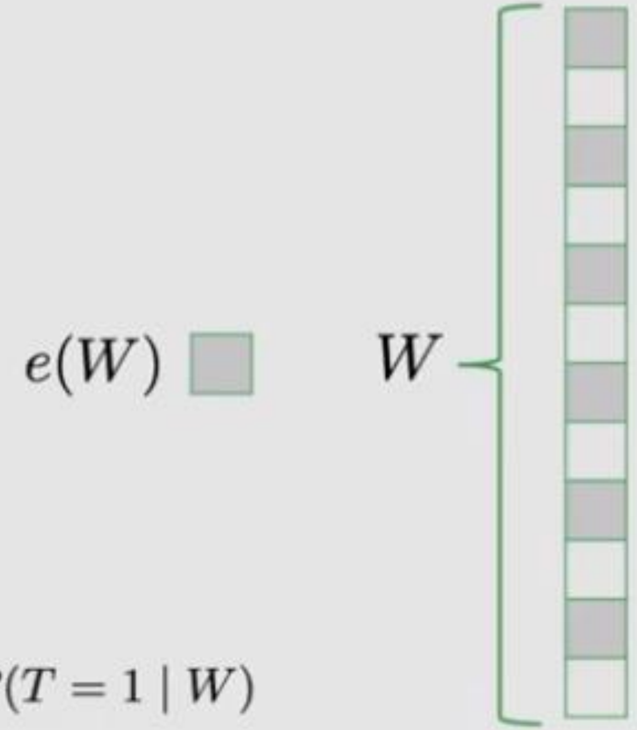
W를 막으면, e(W)를 막아도 같은 현상 (모든 backdoor path 막음)

Implication for the Positivity-Unconfoundedness Tradeoff

Recall that overlap decreases with the dimensionality of the adjustment set

The propensity score magically reduces the dimensionality of the adjustment set done to 1!

Unfortunately, we don't have access to it. The best we can do is model it, shifting the high-dimensionality problem to the modeling of $e(W) \triangleq P(T=1 \mid W)$



Overlap decreases with the dimensionality of the adjustment set
= more likely to have positivity violation get worse

W dimension이 크면 클수록, positivity violation 심해 짐 -> e(W)쓰면 1dimension임으로 굳
e(W) estimation하는 model 활용해서 e(W)쓰자

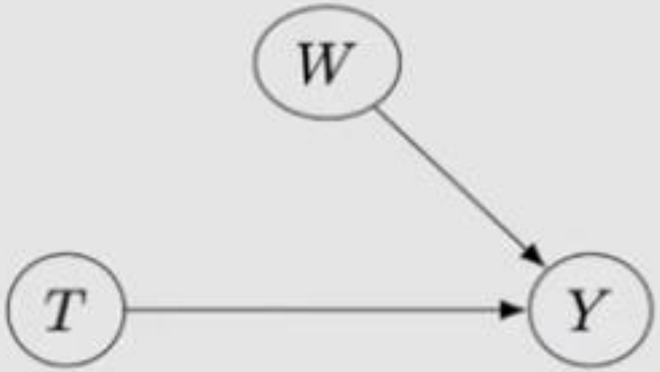
Inverse probability weighting (IPW)

Y = Observation outcome, Reweighting 을 한다 (역수)

$$\mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\mathbb{1}(T=t)Y}{P(t \mid W)}\right]$$

$$\frac{Y}{P(t \mid W)}$$

See proof in Appendix A.3 of the course book



$$\tau \triangleq \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}\left[\frac{\mathbb{1}(T=1)Y}{e(W)}\right] - \mathbb{E}\left[\frac{\mathbb{1}(T=0)Y}{1-e(W)}\right]$$

ATE구하는데 적용 (e(W), 1-e(W))

$$\hat{\tau} = \frac{1}{n_1}\sum_{i:t_i=1}\frac{y_i}{\hat{e}(w_i)} - \frac{1}{n_0}\sum_{i:t_i=0}\frac{y_i}{1-\hat{e}(w_i)}$$

IPW CATE estimation

ATE까지는 구했는데, CATE는? (Conditional-ATE)

Not quite as natural with IPW as with COM, so beyond scope of course

Simple extension:

$$\hat{\tau}(x) = \frac{1}{n_x} \sum_{i:x_i=x} \left(\frac{\mathbb{1}(t_i = 1)y_i}{\hat{e}(w_i)} - \frac{\mathbb{1}(t_i = 0)y_i}{1 - \hat{e}(w_i)} \right)$$

See, e.g., Abrevaya et al. (2015) and references therein

Specific X에 대하여 계산,
근데 잘 안됨. Data points가 많으면,

4. Other Methods

COM과 PS 동시 활용

1) Using Both conditional outcome models and propensity score models

Model both $\mu(t, w)$ and $e(w)$

Example:

W를 e(w)로 변경

$$\hat{\tau} = \frac{1}{n}\sum_i [\hat{\mu}(1, w_i) - \hat{\mu}(0, w_i)]$$

$$\hat{\tau} = \frac{1}{n}\sum_i [\hat{\mu}(1, \hat{e}(w_i)) - \hat{\mu}(0, 1 - \hat{e}(w_i))]$$

4. Other Methods

COM과 PS 동시 활용

1) Using Both conditional outcome models and propensity score models

Model both $\mu(t, w)$ and $e(w)$

Example:

W를 e(w)로 변경

$$\hat{\tau} = \frac{1}{n} \sum_i [\hat{\mu}(1, w_i) - \hat{\mu}(0, w_i)]$$

$$\hat{\tau} = \frac{1}{n} \sum_i [\hat{\mu}(1, \hat{e}(w_i)) - \hat{\mu}(0, 1 - \hat{e}(w_i))]$$

4. Other Methods

COM과 PS 동시 활용

1) Using Both conditional outcome models and propensity score models

Doubly robust methods

- Model both $\mu(t, w)$ and $e(w)$
- Consistent if either $\hat{\mu}(t, w)$ or $\hat{e}(w)$ is consistent
- Theoretically converge to the estimand at a faster rate than COM/IPW
- See Section 7.7 of the course book for references to relevant papers

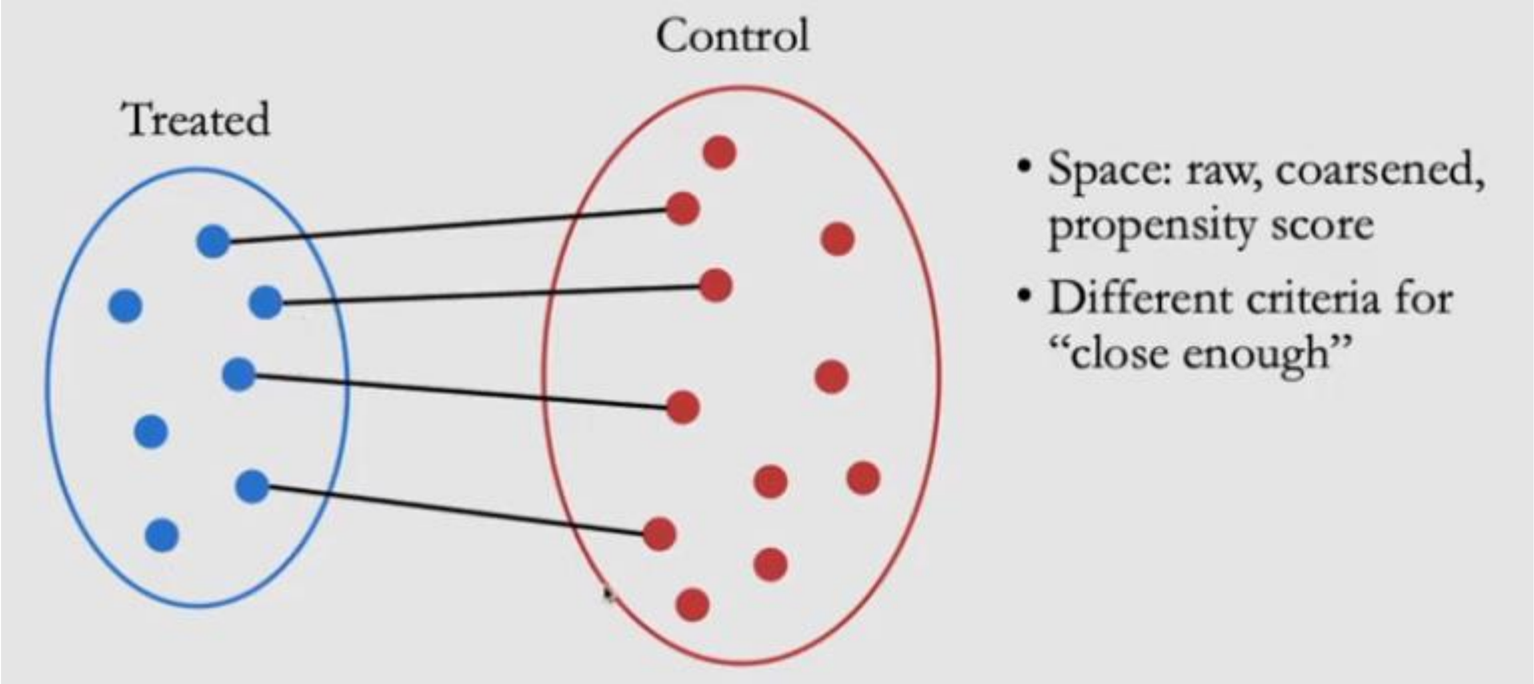
Consistent = given infinite data our model is correct
데이터가 무한대로 있으면, μ estimation이 μ 실제와 같다.

W 를 e(W)로 대체한 Doubly robust method가 더 쉽게 converge한다.

4. Other Methods

2) Matching

Treatment group과 control group subject의 covariance비슷한 것 끼리 matching



Matching안된 거는 분석 안함. Adjustment properly 하기 위해서
Matching 위해서 raw data, coarsened, propensity score dimension에서 close 계산

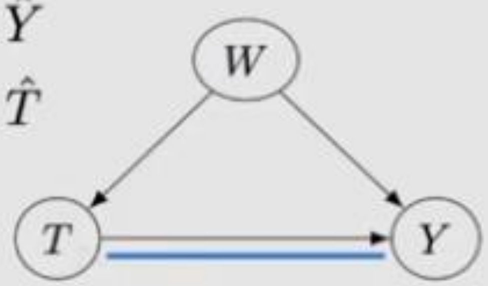
4. Other Methods

3) Double machine learning

Step1. W->Y를 fit / W->T를 fit
Step2. T->Y ($Y - \hat{Y}$ 예측, $T - \hat{T}$ 활용해서)

Stage 1:

- Fit a model to predict Y from W to get the predicted $\hat{Y}$
- Fit a model to predict T from W to get the predicted $\hat{T}$



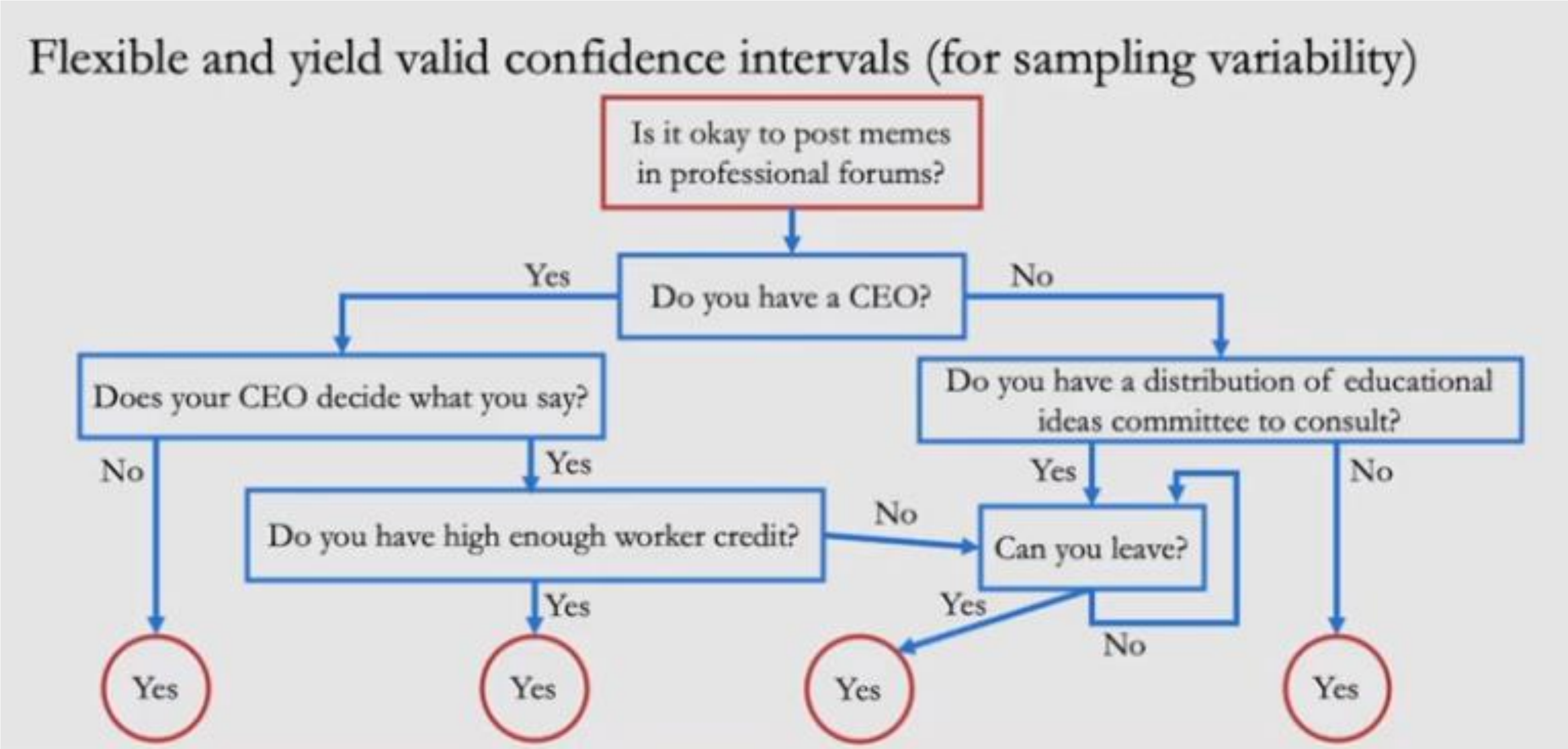
Stage 2:

Partial out W by fitting a model to predict $Y - \hat{Y}$ from $T - \hat{T}$

4. Other Methods

4) Causal trees and forests

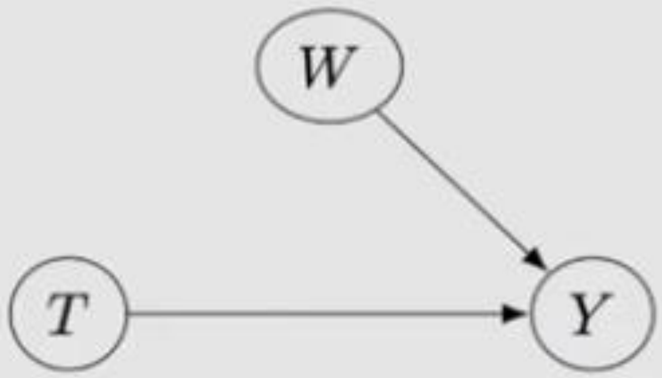
Flexible하지만, sampling에 따라 valid confidence interval 달라질 수도 (unobserved confounder 처리는?? 다음시간 주제)



Inverse probability weighting (IPW)

Y = Observation outcome, Reweighting 을 한다 (역수)

$$\mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\mathbb{1}(T=t)Y}{P(t \mid W)}\right] \qquad \frac{Y}{P(t \mid W)}$$



See proof in Appendix A.3 of the course book

$$\tau \triangleq \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}\left[\frac{\mathbb{1}(T=1)Y}{e(W)}\right] - \mathbb{E}\left[\frac{\mathbb{1}(T=0)Y}{1-e(W)}\right]$$

ATE구하는데 적용 (e(W), 1-e(W))

$$\hat{\tau} = \frac{1}{n_1}\sum_{i:t_i=1}\frac{y_i}{\hat{e}(w_i)} - \frac{1}{n_0}\sum_{i:t_i=0}\frac{y_i}{1-\hat{e}(w_i)}$$

1. Conditional outcome modeling (COM)

(좌) Causal estimate , (우) Statistical estimate

$$\tau = \mathbb{E}_W\left[\mathbb{E}[Y \mid T=1, W] - \mathbb{E}[Y \mid T=0, W]\right]$$

Model은 ML, DL 모두 가능

$$\tau = \mathbb{E}_W\left[\overbrace{\mu(1, W)}^{\text{model}} - \overbrace{\mu(0, W)}^{\text{model}}\right]$$

Model-assisted estimator:

W에 대한 Estimation 이니까,
Wi에 대한 mean 값으로 변경

$$\hat{\tau} = \frac{1}{n}\sum_i \left(\hat{\mu}(1, w_i) - \hat{\mu}(0, w_i)\right)$$