Causal Paper Review

SeongSik Choi

January 26, 2022

Seoul National University

Notation

Notation :

Y : Outcome

A : Treatment

X : Covariate

Z : IV(instrumental variable)

Basically, we consider single binary treatment A.

Z(Multi-dim in regression, Polytonomous in weighting)

Table of Contents

Motivation :

IPTW(inverse probability of treatment weighted) 방법 이용을 위해 PS(propensity score)에 대해 모델링을 하고 싶은 상황.

가지고 있는 covariate(confounding의 후보)이 매우 많다. 이들 중, ATE 추정에 도움이 되는 것들만 남기고 싶다.

이를 위해서는 'Y와 관계된' variable만 남겨야 한다.

1. Outcome-Adaptive Lasso

For $\mathbf{X} = (X_1, X_2, \ldots, X_d)$,

Indices of covariates는 다음과 같이 4가지로 나뉜다.

$\mathcal{C}$: related with both outcome and exposure.

$\mathcal{P}$: related with outcome, not with exposure.

$\mathcal{I}$: related with exposure, not with outcome.

$\mathcal{S}$: not related with both outcome and exposure.

$\mathcal{C}$는 포함해야 하며, $\mathcal{P}$는 efficiency 향상을 위해 포함해야 한다. $\mathcal{I}$는 올바른 PS 추정은 돕겠지만, ATE 추정에 도움을 주지는 않는다. 그리고 우리의 관심은 PS의 추정이 아닌 ATE의 추정이다. $\mathcal{S}$는 빼야한다.

Relatedness with outcome이 핵심.

1. Outcome-Adaptive Lasso

Relatedness with outcome이 핵심.

Consider the outcome model $Y_i = \eta A + \sum_{j=1}^{d} \beta_j X_j + \epsilon_i$.

If $\hat{\beta}_j \approx 0$, X_j should not be included for the PS estimation.

1. Outcome-Adaptive Lasso

Consider the PS model.

(Although it is not true, it is helpful to estimate ATE.)

$$\begin{aligned}\text{logit}\{\pi(\mathbf{X})\} &= \text{logit}\{P(A = 1 \mid \mathbf{X})\} \\ &= \sum_{j \in \mathcal{C}} \alpha_j X_j + \sum_{j \in \mathcal{P}} \alpha_j X_j\end{aligned}$$

If $j \in \mathcal{I} \cup \mathcal{S}$, $\alpha_j = 0$.

So we want to set $\widehat{\alpha}_j = 0$ when $\hat{\beta}_j \approx 0$.

1. Outcome-Adaptive Lasso

We want to set $\widehat{\alpha}_j = 0$ if $\hat{\beta}_j \approx 0$.

$$\widehat{\alpha}(OAL) = \underset{\alpha}{\operatorname{argmin}} \left[\sum_{i=1}^{n} \left\{ -a_i \left(\mathbf{x}_i^T \alpha \right) + \log \left(1 + e^{\mathrm{x}_i^T \alpha} \right) \right\} + \lambda_n \sum_{j=1}^{d} \widehat{\omega}_j \left| \alpha_j \right| \right]$$

where $\widehat{\omega}_j = \left| \widehat{\beta}_j \right|^{-\gamma}$ s.t. $\gamma > 1$

Unlike Adaptive Lasso, $\widehat{\omega}_j$ comes from the outcome model ■

Table of Contents

2. Regression and Weighting Methods Using IV

Motivation :

Why do we use IV(instrumental variable) method?

1. To estimate ATE in the existence of unmeasured confounders
2. **To estimate LATE(Local ATE)**

2. Regression and Weighting Methods Using IV

What is IV?

Instrumental Variable Z satisfies

1. Z is associated with A
2. Z does not affect Y except through its potential effect on A
3. Z and Y do not share causes

also additional condition is needed to be identifiable.

(Homogeneity or Monotonicity)

(Example. A : 학원 등록 여부, Y : 성적, Z : 학원 권유 여부)

What is LATE?

LATE over a subpopulation X = x is

$$\text{LATE}\left(\mathbf{x}, \mathbf{z}, \mathbf{z}'\right) = E\left(Y_1 - Y_0 \mid \mathbf{X} = \mathbf{x}, A_{\mathrm{z}} < A_{\mathrm{z}'}\right)$$

The population LATE is

$$\text{LATE}\ \left(\mathbf{z}, \mathbf{z}'\right) = E\left(Y_1 - Y_0 \mid A_{\mathrm{z}} < A_{\mathrm{z}'}\right)$$

(Example. A : 학원 등록 여부, Y : 성적, Z : 학원 권유 여부)

2. Regression and Weighting Methods Using IV

Additional condition is needed to be identifiable.

(Homogeneity or Monotonicity)

- Homogeneity

For any a, $E\left(Y^{a=1} - Y^{a=0} \mid Z = z, A = a\right)$ is constant with respect to z.

- Monotonicity

For any $\mathbf{z}, \mathbf{z}'$, and $\mathbf{x}$, either $A_{\mathrm{z}}(\omega) \leq A_{\mathrm{z}'}(\omega)$ for all ω in the set $\{\mathbf{X} = \mathbf{x}\}$ or $A_{\mathrm{z}}(\omega) \geq A_{\mathrm{z}'}(\omega)$ for all ω in the set $\{\mathbf{X} = \mathbf{x}\}$.

(Example. A : 학원 등록 여부, Y : 성적, Z : 학원 권유 여부)

2. Regression and Weighting Methods Using IV

기존 논문에서는 IV 세팅하에 (Y, X, Z, A)의 분포를 알 때, LATE 추정량이 Identifiable함을 보였다.

본 논문은 이에 대한 Estimation 방법을 제시한 논문이다.

기존 논문들보다 general한 세팅 에서의 estimation을 다루고 있다.

사용하는 Modeling assumption에 따라 Regression method와 Weighting method로 나뉜다. (Causal assumption은 서로 동일하다.)

사용하는 Modeling assumption에 따라 Regression method와 Weighting method로 나뉜다. (Causal assumption은 앞서 제시된 조건들로, 서로 동일하다.)

Regression method

$$P(A = 1 \mid \mathbf{X}, \mathbf{Z}) = \pi(\mathbf{X}, \mathbf{Z})$$
$$E(Y \mid A, \mathbf{X}, \mathbf{Z}) = \eta(A, X, \pi(\mathbf{X}, \mathbf{Z}))$$

(Modeling assumptions for treatment propensity score and the outcome regression function)

Weighting method

$$P(\mathbf{Z} = \mathbf{z} \mid \mathbf{X} = \mathbf{x}) = p(\mathbf{z} \mid \mathbf{x})$$

(Modeling assumptions for instrument propensity score)

2. Regression and Weighting Methods Using IV

Regression method

For $\pi < \pi', \pi = \pi(\mathbf{x}, \mathbf{z})$, and $\pi' = \pi(\mathbf{x}, \mathbf{z}')$,

Theorem 1. (b) If the IV assumptions hold, then

$$E(Y_1 \mid A_z < A_{z'}, \mathbf{X} = \mathbf{x}) = \frac{\pi' \eta(1, \mathbf{x}, \pi') - \pi \eta(1, \mathbf{x}, \pi)}{\pi' - \pi}$$

and

$$E(Y_0 \mid A_z < A_{z'}, \mathbf{X} = \mathbf{x}) = \frac{(1-\pi)\eta(0, \mathbf{x}, \pi) - (1-\pi')\eta(0, \mathbf{x}, \pi')}{\pi' - \pi},$$

Regression method

Theorem 1. (c) If, further, the population monotonicity assumption holds, then

$$E\left(Y_1 \mid A_{\mathrm{z}} < A_{\mathrm{z}'}\right) = \frac{E\left[\pi' \eta\left(1, \mathbf{X}, \pi'\right)\right] - E[\pi \eta(1, \mathbf{X}, \pi)]}{E\left(\pi'\right) - E(\pi)}$$

and

$$E\left(Y_0 \mid A_{\mathrm{z}} < A_{\mathrm{z}'}\right) = \frac{E[(1-\pi)\eta(0, \mathbf{X}, \pi)] - E\left[\left(1-\pi'\right)\eta\left(0, \mathbf{X}, \pi'\right)\right]}{E\left(\pi'\right) - E(\pi)},$$

where $\pi < \pi', \pi = \pi(\mathbf{X}, \mathbf{z})$, and $\pi' = \pi\left(\mathbf{X}, \mathbf{z}'\right)$.

Regression method

The fitted values $\hat{\pi}$ and $\hat{\eta}$ can be substituted in **Theorem 1. (b), (c)** to estimate causal effects.

The LATE at $\mathbf{X} = \mathbf{x}$ can be estimated by

$$\hat{E}\left(Y_1 - Y_0 \mid A_{\mathbf{z}} < A_{\mathbf{z}'}, \mathbf{X} = \mathbf{x}\right) = \frac{\hat{E}\left(Y \mid \mathbf{x}, \mathbf{z}'\right) - \hat{E}(Y \mid \mathbf{x}, \mathbf{z})}{\hat{E}\left(A \mid \mathbf{x}, \mathbf{z}'\right) - \hat{E}(A \mid \mathbf{x}, \mathbf{z})}$$

The population LATE can be estimated by

$$\hat{E}\left(Y_1 - Y_0 \mid A_{\mathbf{z}} < A_{\mathbf{z}'}\right) = \frac{\tilde{E}\left[\hat{E}\left(Y \mid \mathbf{X}, \mathbf{z}'\right)\right] - \tilde{E}[\hat{E}(Y \mid \mathbf{X}, \mathbf{z})]}{\tilde{E}\left[\hat{E}\left(A \mid \mathbf{X}, \mathbf{z}'\right)\right] - \tilde{E}[\hat{E}(A \mid \mathbf{X}, \mathbf{z})]}$$

where $\hat{E}(Y \mid \mathbf{x}, \mathbf{z}) = \hat{\pi}\hat{\eta}(1, \mathbf{x}, \hat{\pi}) + (1 - \hat{\pi})\hat{\eta}(0, \mathbf{x}, \hat{\pi})$ and $\hat{E}(A \mid \mathbf{x}, \mathbf{z}) = \hat{\pi}(\mathbf{x}, \mathbf{z})$, and $\tilde{E}$ denotes sample average.

Weighting method

For $P(\mathbf{Z} = \mathbf{z} \mid \mathbf{X} = \mathbf{x}) = p(\mathbf{z} \mid \mathbf{x})$,

Theorem 3. (a) For $F(\mathbf{z}) = P\left(A_{\mathrm{z}} = 1\right), G_1(\mathbf{z}) = E\left(Y_1 \mid A_{\mathrm{z}} = 1\right)$ and $G_0(\mathbf{z}) = E\left(Y_0 \mid A_{\mathrm{z}} = 0\right)$,

$$F(z) = E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}A\right] / E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}\right]$$

$$G_1(z) = E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}AY\right] / E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}A\right]$$

$$G_0(z) = E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}(1 - A)Y\right] / E\left[\frac{Z}{\hat{p}(z \mid \mathbf{X})}(1 - A)\right]$$

Weighting method

For $F(\mathbf{z}) = P(A_{\mathrm{z}} = 1)\,, G_1(\mathbf{z}) = E(Y_1 \mid A_{\mathrm{z}} = 1)$ and $G_0(\mathbf{z}) = E(Y_0 \mid A_{\mathrm{z}} = 0)$,

Theorem 3. (b) If the population monotonicity assumption holds, then

$$E(Y_1 \mid A_{\mathrm{z}} < A_{\mathrm{z}'}) = \frac{F(\mathbf{z}')\,G_1(\mathbf{z}') - F(\mathbf{z})G_1(\mathbf{z})}{F(\mathbf{z}') - F(\mathbf{z})}$$

and

$$E(Y_0 \mid A_{\mathrm{z}} < A_{\mathrm{z}'}) = \frac{(1 - F(\mathbf{z}))G_0(\mathbf{z}) - (1 - F(\mathbf{z}'))\,G_0(\mathbf{z}')}{F(\mathbf{z}') - F(\mathbf{z})}$$

where $F(\mathbf{z}) < F(\mathbf{z}')$ and hence $A_{\mathrm{z}}(\omega) \leq A_{\mathrm{z}'}(\omega)$ for all ω.

2. Regression and Weighting Methods Using IV

Regression method

The population LATE can be estimated similarly to the regression method.

Regression :
(1) Estimate $\pi(\hat{x}, z)$, $\eta(a, x, \hat{\pi}(x, z))$
(2) Estimate the LATE at $(X = x)$ and population LATE

Weighting :
(1) Estimate $p(z|x)$
(2) Estimate $F(z), G_1(z), G_0(z)$
(3) Estimate the population LATE ■

Table of Contents

Motivation :

TMLE는 일반적인 G-computation과 PS method와 달리, **Targeting step**을 통해 ATE에 대한 bias-variance tradoff를 optimize하는 것을 목표로 한다.

Doubly robust하다는 장점을 가진다.

3. Targeted Maximum Likelihood Estimation

Three steps.

1. Estimate $E(Y|A,X)$, $E(Y|A=a,X)$ **(Outcome method)**
2. Estimate $P(A|X)$ **(PS method)**
3. Regress the residuals from outcome estimation in Step 1 onto some function of the PS.
   $\text{logit}(E(Y|A,X)) = \text{logit}(\hat{E}(Y|A,X)) + \epsilon H(A,X)$
   (Targeting step)

$\text{logit}(\hat{E}^*(Y|A,X)) = \text{logit}(\hat{E}(Y|A,X)) + \hat{\epsilon} H(A,X)$

Finding $H(A,X)$ might be the crux.

3. Targeted Maximum Likelihood Estimation

1. Estimate $E(Y|A, X)$, $E(Y|A = a, X)$

Y	X1	X2	X3	X4	A
1	1	0	6	1	1
1	0	1	6	3	0
0	0	0	3	2	0
1	0	1	5	1	1
1	0	0	5	2	0

1. Estimate $E(Y|A, X)$, $E(Y|A = a, X)$

For $Q(A, X) = E(Y|A, X)$, we can estimate $Q \to Q_A, Q_a$.

Y	A	Q_A	Q_0	Q_1
1	1	0.83	0.65	0.83
1	0	0.72	0.72	0.87
0	0	0.48	0.48	0.70
1	1	0.83	0.65	0.83
1	0	0.62	0.62	0.81

2. Estimate $P(A|X)$

For $\pi(X) = P(A|X)$, we can estimate π (PS method)

3. Targeted Maximum Likelihood Estimation

3. Regress the residuals from outcome estimation in Step 1 onto some function of the PS.

Calculate the function of the PS,

$$H(A, \mathbf{X}) = \frac{\mathrm{I}(A = 1)}{\Pr(A = 1 \mid \mathbf{X})} - \frac{\mathrm{I}(A = 0)}{\Pr(A = 0 \mid \mathbf{X})}$$

for each individual.

Y	A	Q_A	Q_0	Q_1	H_1	H_0	H_A
1	1	0.83	0.65	0.83	2.18	−1.85	2.18
1	0	0.72	0.72	0.87	1.61	−2.65	−2.65
0	0	0.48	0.48	0.70	3.40	−1.42	−1.42
1	1	0.83	0.65	0.83	2.39	−1.72	2.39
1	0	0.62	0.62	0.81	2.30	−1.77	−1.77

3. Targeted Maximum Likelihood Estimation

3. Regress the residuals from outcome estimation in Step 1 onto some function of the PS.
$\text{logit}(E(Y|A,X)) = \text{logit}(\hat{E}(Y|A,X)) + \epsilon H(A,X)$

Y	A	Q_A	Q_0	Q_1	H_1	H_0	H_A
1	1	0.83	0.65	0.83	2.18	-1.85	2.18
1	0	0.72	0.72	0.87	1.61	-2.65	-2.65

```
glm _fit <- glm(Y ~  - 1 + offset(qlogis(Q _ A)) + H _ A,
                data=dat, family=binomial)
```

$$\text{logit}(\hat{E}^*(Y|A,X)) = \text{logit}(\hat{E}(Y|A,X)) + \hat{\epsilon}H(A,X)$$

$$\text{logit}(\hat{E}^*(Y|A=a,X)) = \text{logit}(\hat{E}(Y|A=a,X)) + \hat{\epsilon}H(A,X)$$

Reference

1. Shortreed, S. M., Ertefaie, A. (2017). Outcome-adaptive lasso: variable selection for causal inference. Biometrics, 73(4), 1111-1122.

2. Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. Journal of the American Statistical Association, 101(476), 1607-1618.

3. Schuler, M. S., Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. American journal of epidemiology, 185(1), 65-73.