

<Debugging Simulation Result>

- 2월 10일 ~ 2월 16일 Version

[2월 10일 ~ 14일]

<What TO DO>

: Sandwich variance estimator of IPW ATE의 Coverage probability 확인
: Sandwich variance estimator of IPW ATE 생성 때 참고한 선행논문 확인 — Coverage prob이 0.95 인지

<Check Result>

1) PS model & Outcome model all correctly specified된 상황에서, Sandwich variance estimator of IPW ATE의 Coverage probability가 0.951정도 나옴을 확인 — 이때, Exposure ratio는 0.5

2) Sandwich variance estimator of IPW ATE 생성 때 참고한 선행논문의 Simulation에서는 해당 분산 추정량의 Coverage probability가 0.95 정도 됨을 확인함

∴ Debugging 한 후 다시 얻은 결과 file 명은 “Ver2.12_ATT_Performance.xlsx”

[2월 16일]

<What TO DO>

: 시나리오와 Replication이 변함에 따라 IPW ATT Estimator와 DR ATT Estimator 추정치가 동일한지 확인

: $\hat{\mu}_0 = E[Y^0 | A=1]$ 추정할 때 Regression 통해 얻은 fitted.values 이용하거나 formula 통해 얻는 방법 2가지 있음. / 두 방법을 통해 얻은 추정치 차이의 절댓값이 1e-06보다 작은지 확인

<Check Result>

1) IPW ATT Estimator와 DR Estimator가 Replication에 따라 동일한 추정치 제시하는 것을 확인

1)-①. PS model만 correctly specified 되면, IPW ATT, DR ATT 추정치가 모두 동일함을 확인

1)-②. Outcome regression model만 correctly specified 되면, Outcome regression estimator, DR ATT 추정치가 모두 똑같음.

2) $\hat{\mu}_0 = E[Y^0 | A=1]$ 추정하는 두 가지 방법 이용해 얻은 추정치들 차이의 절댓값이 모든 Replication에 대해 1e-06보다 적음을 확인.

⇒ Estimator Debugging

3)-①. IPW ATT Estimator / DR ATT Estimator에서 μ_1 추정치가 같은 부분은 이해할 수 있음
 -- 추정량 공식이 동일함.

3)-②. IPW ATT Estimator / DR ATT Estimator에서 μ_0 추정치가 같으면 안되나, 같음.

3)-②-1. IPW ATT 추정 때, DR ATT 추정 때 적합한 PS model이 일치함을 확인.

3)-②-2. “sum(A)”와 “sum(weight*(1-A))” 값이 동일함. → B가 이항변수이면 충분히 일어날 수 있는가?

3)-②-2. 각 추정량 분자 값도 모두 같음. 즉, “sum((1-A)*weight*(Y-mu0_hat) + A*mu0_hat)” 와
 sum((1-A)*weight*Y) 값 또한 같음.

∴ Confounder “B”와 공변량 “C”, “U”의 분포를 모두 표준정규분포로 바꿈. -- Scenario Version 3

4) 공변량들의 분포를 바꾼 뒤, 각 추정량들의 bias, 추정치 다시 확인

[각 방법론 별 Bias]

: $P(A=1 \mid B=b, C=c) = \text{expit}(-1+B+0.3C) / \mu = 1 + 2.5B + A + 0.3C$

Scenario	Outcome regression	IPW Estimator	DR Estimator
(i)	0.002340689	0.002331311	0.001981060
(ii)	1.570398041	0.002331311	0.00304989
(iii)	0.002340689	1.570398041	0.002340689
(iv)	1.570398	1.570398	1.570398

[각 방법론 별 Bias 측면, 시뮬레이션 결과 해석 — 나의 생각]

Scenario (i) : DR Estimator의 성능이 Best

Scenario (ii) : Outcome model이 mis-specified 된 시나리오이기 때문에 Outcome regression estimator의 편향이 급격히 증가 — make sense

/ DR Estimator도 Outcome regression estimator보다는 아니지만, 편향이 어느 정도 증가

/ IPW Estimator에는 변화가 없는 것이 맞다고 생각됨.

Scenario (iii) : PS model이 mis-specified 된 시나리오이므로 IPW Estimator의 편향이 100배 이상 증가한 것은 make sense

/ DR Estimator 또한 Scenario (i) 결과에 비해 편향이 증가 — 그러나, Scenario (ii)보다는 아님

Q) PS model의 misspecification이 더 영향이 큰가?

/ Outcome regression estimator와 DR Estimator 추정 결과가 동일함.

Q) formula 측에서 살펴볼 수 있는 방법이 있을까?

/ Scenario (ii)에서의 Outcome regression estimator 편향 결과와 IPW Estimator 편향 결과가 동일

Q) formula 측에서 살펴볼 수 있는 방법이 있을까?

Scenario (iv) : 모든 추정치들이 동일함.

/ Scenario (iii)에서의 질문을 해결하면 답을 알 수 있을 듯 하다.

Outcome regression estimator & IPW Estimator : model의 적합성에 크게 영향을 받는다.

DR Estimator : 다른 두 추정량보다는 Robust함이 두드러지게 보임.