

Consistency check about mu0 and mu1 - 7월 29일 Version

1. Consistency check about mu0 and mu1

: covariate B의 분포는 이항분포에서 표준정규분포로 변경

: #of obs는 100000이고 exposure ratio는 0.76으로 고정

: 100번의 replication 동안 true_mu0($E[Y^0|E=1]$)와 true_mu1($E[Y^1|E=1]$) 계산

/ 각 true value와 ver01, ver04, ver05, ver08 추정량 통해 계산한 $\hat{\mu}_0, \hat{\mu}_1$ 과의 차이 계산

: 100번 반복시행 통해 얻은 결과를 이용해 추정량의 bias와 variance 계산

[추정량 Version]

Ver01 : $\hat{\mu}_0 = E[Y|E=0]$ / $\hat{\mu}_1 = E[Y|E=1]$

$$\begin{array}{l} \text{Ver 04} \left[\begin{array}{l} \hat{\mu}_0 = \log 1.5 \times P(B|E=1) + \log 2 \cdot E[C|E=1] \\ \hat{\mu}_1 = \hat{\mu}_0 + \log 2 \end{array} \right] \quad / \quad \text{Ver 05} \left[\begin{array}{l} \hat{\mu}_1 = \text{mean}(\text{result_model} \$ \text{fitted.values} \\ \text{[data\$E==1]}) \\ \hat{\mu}_0 = \text{mean}(\text{result_model} \$ \text{fitted.values} \\ \text{[data\$E==0]}) \end{array} \right] \\ \text{Ver 08} \left[\begin{array}{l} \hat{\mu}_1 = \frac{\sum \hat{w}_i A_i Y_i}{\sum \hat{w}_i A_i} \\ \hat{\mu}_0 = \frac{\sum \hat{w}_i (1-A_i) Y_i}{\sum \hat{w}_i (1-A_i)} \end{array} \right] \quad (\hat{w}_i \text{는 weight 값}) \end{array}$$

[Result]

① $\hat{\mu}_0$

	Bias	Variance
Ver01	-0.3760882052	8.071604e-05
Ver04	0.0009757375	1.036956e-05
Ver05	0.0018798240	7.747421e-05
Ver08	0.0018798240	7.747421e-05

: $E[Y^0|E=1]$ 의 경우, 추정량의 version이 Ver04일 때 최소의 bias를 가지는 것을 확인하였다.

: 모든 추정량에 대해 분산은 모두 적은 값을 가진다.

: Ver05, Ver08은 추정량의 bias와 variance값이 모두 같다는 점을 확인하였다.

② $\hat{\mu}_1$

	Bias	Variance
Ver01	0	0
Ver04	3.546110e-04	1.352759e-05
Ver05	-5.528911e-06	3.248597e-29
Ver08	4.440892e-18	6.454316e-33

: $E[Y^1|E=1]$ 의 경우, 추정량 version이 Ver01일 때 즉, 추정량이 $E[Y|E=1]$ 일 때 bias가 0이므로 consistency assumption이 만족하는 점을 확인하였다.

: 모든 추정량에 대해 편향, 추정량의 분산이 적은 값을 보인다.

Question) IPTW ATT Variance 함수 생성 때 $\hat{\mu}_0$ 은 Ver04 추정량, $\hat{\mu}_1$ 은 Ver01 추정량 이용 가능?

2. Consistency check about μ_0 and μ_1 - Version 2

- : $\text{true_mu0}(E[Y^0|E=1])$ 와 $\text{true_mu1}(E[Y^1|E=1])$ 값을 Potential outcome 이용해 계산하는 것이 아닌, 위에서 제시한 Version 4의 추정량이 true value라 간주하고 ver05, ver08 추정량의 편향, variance 계산
- : covariate B의 분포는 이항분포에서 표준정규분포로 변경
- : #of obs는 100000이고 exposure ratio는 0.76으로 고정
- : 100번의 replication 동안 $\text{true_mu0}(E[Y^0|E=1])$ 와 $\text{true_mu1}(E[Y^1|E=1])$ 계산
- / 각 true value와 ver01, ver05, ver08 추정량 통해 계산한 $\hat{\mu}_0, \hat{\mu}_1$ 과의 차이 계산
- : 100번 반복시행 통해 얻은 결과를 이용해 추정량의 bias와 variance 계산

[추정량 Version]

Ver01 : $\hat{\mu}_0 = E[Y|E=0]$ / $\hat{\mu}_1 = E[Y|E=1]$

$$\begin{aligned} \text{Ver 04} & \left[\begin{aligned} \hat{\mu}_0 &= \log 1.5 \times P(B|E=1) + \log 2 \cdot E[C|E=1] \\ \hat{\mu}_1 &= \hat{\mu}_0 + \log 2 \end{aligned} \right] \quad / \quad \text{Ver 05} & \left[\begin{aligned} \hat{\mu}_1 &= \text{mean}(\text{result_model} \$ \text{fitted.values} \\ & \quad [data\$E == 1]) \\ \hat{\mu}_0 &= \text{mean}(\text{result_model} \$ \text{fitted.values} \\ & \quad [data\$E == 0]) \end{aligned} \right] \\ \text{Ver 08} & \left[\begin{aligned} \hat{\mu}_1 &= \frac{\sum \hat{w}_i A_i Y_i}{\sum \hat{w}_i A_i} \\ \hat{\mu}_0 &= \frac{\sum \hat{w}_i (1-A_i) Y_i}{\sum \hat{w}_i (1-A_i)} \end{aligned} \right] \quad (\hat{w}_i \text{는 weight of } |D|) \end{aligned}$$

[Result]

① $\hat{\mu}_0$

	Bias	Variance
Ver01	-0.3760882052	6.673565e-05
Ver05	0.0009040866	6.964828e-05
Ver08	0.0009040866	6.964828e-05

: ver04 추정량이 $E[Y^0|E=1]$ true value라 간주한 경우, 추정량의 version이 Ver05, Ver08일 때 최소의 bias를 가지는 것을 확인하였고, 두 추정량의 편향과 분산값이 같은 것을 알 수 있음.

: 눈에 띄는 점은, true value를 potential outcome 통한 $E[Y^0|E=1]$ 값으로 간주했을 때와 Ver01의 편향이 같다는 점이다. (true value를 계산하는 공식이 Ver04이다...?!)

② $\hat{\mu}_1$

	Bias	Variance
Ver01	-0.000354611	1.352759e-05
Ver05	-0.000354611	1.352759e-05
Ver08	-0.000354611	1.352759e-05

: ver04 추정량이 $E[Y^1|E=1]$ true value라 간주한 경우, 모든 추정량 version이 동일한 편향과 분산을 가진다는 점을 확인하였다.