‹New Simulation Scenario Result›

- 2월 20일 ~ 21일 Version

‹Preview›

: 새롭게 생성한 Simulation Scenario에 대하여, p=0.125일 때 Simulation 결과 살펴봄.

[Version 0]

① P(A=1 | B=b, C=c, D=d) = expit(-5.5 + 0.7B + 0.7C + 0.5D) -- p = 0.125

② μ = 1 + 2.5B + 2.5C + A + U

For Scenario Version 4, ATT Performance Version 0

Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00823075561	0.080399423	0.949	0.999067603		
IPW	-0.00543264361	0.084441398	0.918	0.754513407	0.953	1.019217829
DR	0.00914683432	0.082134824	0.272	0.0296606278	0.955	1.076876684

Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.397393736	2.088859905	0.036	1.584882157		
IPW	-0.00543264361	0.084441398	0.918	0.754513407	0.953	1.019217829
DR	-0.00606167914	0.084731758	0.347	0.051913523	0.953	1.104810623

Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00823075561	0.080399423	0.949	0.999067603		
IPW	1.395569397	2.083829891	0.006	0.507176493	0.038	0.994576588
DR	0.00811929842	0.080512604	0.263	0.0290941522	0.949	1.075486746

Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.397393736	2.088859905	0.036	1.013840914		
IPW	1.395569397	2.083829891	0.006	0.507176493	0.038	0.994576588
DR	1.395569392	2.083829876	0	0.0572982065	0.046	1.280184245

: 분산 추정량의 Coverage와 SD Ratio는 GOOD

/ Scenario 별 추정량 간 bias가 동일한 현상이 나타나지 않음.

/ DR Estimator의 Bias가 예상과 다름. (Scenario가 mis-specified 되었을 때 bias가 더 낮음)

[Version 1]

① P(A=1 | B=b, C=c, D=d) = expit(-5.4 + 0.7B + 0.5C + 0.5D) -- p = 0.125

② μ = 1 + 2.5B + 1.5C + A + U

For Scenario Version 4, ATT Performance Version 1

Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00972862041	0.076766223	0.958	0.945588195		
IPW	-0.00151390088	0.080166145	0.915	0.681860175	0.957	0.964514753
DR	0.0108805514	0.078633623	0.292	0.0277543407	0.962	1.020682199

Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.401653093	2.095919056	0.03	1.499921762		
IPW	-0.00151390088	0.080166145	0.915	0.681860175	0.957	0.964514753
DR	-0.00280168742	0.080516273	0.355	0.0484502522	0.965	1.04480528

Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00972862041	0.076766223	0.958	0.945588195		
IPW	1.400742154	2.093498101	0.006	0.451576005	0.029	0.936335905
DR	0.0096870198	0.076929826	0.284	0.027197232	0.965	1.01875336

Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.401653093	2.095919056	0.03	0.95415514		
IPW	1.400742154	2.093498101	0.006	0.451576005	0.029	0.936335905
DR	1.400742153	2.093498096	0	0.0536592546	0.04	1.208102977

: Input Control 하는 Confounder "B"의 effect를 줄였음에도 Version 0과 동일한 결과가 나오고 있음.

∴ Random error라 간주하고, p 값 다양하게 바꾸어가며 결과 살펴보자.

[p = 0.5]

① P(A=1 | B=b, C=c, D=d) = expit(-2.8 + 0.7B + 0.7C + 0.5D)

② μ = 1 + 2.5B + 2.5C + A + U

For Scenario Version 4, ATT Performance Version 0 (p = 0.5)

Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.0028841906	0.039327509	0.95	1.003818749		
IPW	0.0032872935	0.049593818	0.976	1.339187649	0.954	1.001122625
DR	0.0016566883	0.043236989	0.699	0.270220437	0.957	1.02089948

Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.157974732	1.405153099	0.008	1.554649834		
IPW	0.0032872935	0.049593818	0.976	1.339187649	0.954	1.001122625
DR	0.0028615676	0.049059448	0.834	0.567247894	0.956	1.160060347

Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.0028841906	0.039327509	0.95	1.003818749		
IPW	1.15694812	1.404374816	0.011	1.055516929	0.008	0.996112142
DR	0.0024027246	0.040014804	0.691	0.257539824	0.953	1.025047382

Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.157974732	1.405153099	0.008	0.998478048		
IPW	1.15694812	1.404374816	0.011	1.055516929	0.008	0.996112142
DR	1.156948119	1.404374814	0	0.294336547	0.011	1.068459982

[p = 0.25]

① P(A=1 | B=b, C=c, D=d) = expit(-4.5 + 0.7B + 0.7C + 0.5D)

② μ = 1 + 2.5B + 2.5C + A + U

For Simulation Scenario Version 4, ATT Performance Version 0 - p = 0.25

Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.0110746815	0.054812802	0.943	1.000424934		
IPW	-0.00881398144	0.058594677	0.965	1.140475003	0.95	1.012846768
DR	0.0091945884	0.056328517	0.391	0.0766409487	0.954	1.05513604

Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.264840368	1.685360455	0.004	1.562936267		
IPW	-0.00881398144	0.058594677	0.965	1.140475003	0.95	1.012846768
DR	-0.00904215482	0.058946739	0.521	0.148781451	0.955	1.111871296

Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.0110746815	0.054812802	0.943	1.000424934		
IPW	1.262010796	1.678071469	0.003	0.795601046	0.005	0.989013239
DR	0.0106003596	0.054964231	0.377	0.0741227337	0.954	1.057919318

Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.264840368	1.685360455	0.004	1.000067248		
IPW	1.262010796	1.678071469	0.003	0.795601046	0.005	0.989013239
DR	1.262010796	1.678071469	0	0.108280065	0.009	1.177263871

[p = 0.125]

① P(A=1 | B=b, C=c, D=d) = expit(-5.5 + 0.7B + 0.7C + 0.5D)

② μ = 1 + 2.5B + 2.5C + A + U

For Scenario Version 4, ATT Performance Version 0						
Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00823075561	0.080399423	0.949	0.999067603		
IPW	-0.00543264361	0.084441398	0.918	0.754513407	0.953	1.019217829
DR	0.00914683432	0.082134824	0.272	0.0296606278	0.955	1.076876684
Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.397393736	2.088859905	0.036	1.584882157		
IPW	-0.00543264361	0.084441398	0.918	0.754513407	0.953	1.019217829
DR	-0.00606167914	0.084731758	0.347	0.051913523	0.953	1.104810623
Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	0.00823075561	0.080399423	0.949	0.999067603		
IPW	1.395569397	2.083829891	0.006	0.507176493	0.038	0.994576588
DR	0.00811929842	0.080512604	0.263	0.0290941522	0.949	1.075486746
Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.397393736	2.088859905	0.036	1.013840914		
IPW	1.395569397	2.083829891	0.006	0.507176493	0.038	0.994576588
DR	1.395569392	2.083829876	0	0.0572982065	0.046	1.280184245

[p = 0.065]

① P(A=1 | B=b, C=c, D=d) = expit(-6.5 + 0.7B + 0.7C + 0.5D)

② μ = 1 + 2.5B + 2.5C + A + U

For Simulation Scenario Version 4, ATT Performance Version 0 - p = 0.0625						
Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	-0.000184011744	0.16953389	0.941	1.012635211		
IPW	-0.00552950516	0.17631739	0.764	0.405170642	0.94	1.034540371
DR	0.00208804442	0.17205414	0.14	0.0109685337	0.953	1.104574273
Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.528801181	2.61095665	0.163	1.638639646		
IPW	-0.00552950516	0.17631739	0.764	0.405170642	0.94	1.034540371
DR	-0.00660569391	0.17676508	0.171	0.0162163134	0.95	1.118138639
Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	-0.000184011744	0.16953389	0.941	1.012635211		
IPW	1.527810354	2.6075284	0.031	0.261209274	0.152	0.996817
DR	-0.000430763826	0.16939005	0.139	0.0109149366	0.951	1.103729957
Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.528801181	2.61095665	0.163	1.026210978		
IPW	1.527810354	2.6075284	0.031	0.261209274	0.152	0.996817
DR	1.527810347	2.60752838	0	0.0287014236	0.211	1.381897502

[p = 0.015]

① P(A=1 | B=b, C=c, D=d) = expit(-8 + 0.7B + 0.7C + 0.5D)

② μ = 1 + 2.5B + 2.5C + A + U

For Scenario Version4, ATT Performance Version 0 (p = 0.015)						
Scenario (i)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	-0.00150434493	0.54284485	0.958	1.025477602		
IPW	-0.000950405343	0.55042221	0.495	0.11204699	0.9246988	1.020269255
DR	0.00268710934	0.54922209	0.06	0.00291074059	0.964	1.12882388
Scenario (ii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.670856819	3.72097981	0.587	1.695013823		
IPW	-0.000950405343	0.55042221	0.495	0.11204699	0.9246988	1.020269255
DR	-0.00255174337	0.55134379	0.071	0.00328732437	0.962	1.133777749
Scenario (iii)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	-0.00150434493	0.54284485	0.958	1.025477602		
IPW	1.670665767	3.72060519	0.099	0.0697029399	0.566	0.973815179
DR	-0.00164595374	0.54323854	0.066	0.00290227941	0.962	1.125868586
Scenario (iv)	Bias	rMSE	Naive_var_coverage	Naive_var_SD_Ratio	Sandwich_robust_var_coverage	Sandwich_var_SD_Ratio
Outcome_reg	1.670856819	3.72097981	0.587	1.039498509		
IPW	1.670665767	3.72060519	0.099	0.0697029399	0.566	0.973815179
DR	1.67066576	3.72060516	0.017	0.00872275964	0.78	1.499172522

(나의 생각)

1) model이 올바르게 적합 되었느냐의 여부에 상관없이 DR Estimator는 로버스트하다. Bias, rMSE, Sandwich robust variance의 Coverage probability에 큰 변화가 없음. - 단, Scenario (iv) 제외

2) model correctly specification 여부에 Outcome regression estimator, IPW estimator는 크게 영향 받는다.
   - Outcome regression estimator의 경우 bias가 1.66 정도 증가, 분산은 과대추정됨
   - IPW Estimator의 경우 bias가 1.669 정도 증가, 분산은 과소추정됨.

3) Naive variance estimator의 성능은 Exposure prevalence가 낮아질수록 나빠지는 반면(Coverage probability 확인), Sandwich robust variance estimator는 그런 경향이 보이지 않음. / 따라서, Sandwich variance estimator 사용하는 것을 추천한다.

4) Scenario (iv)에서 Exposure prevalence가 낮아질수록 Sandwich variance estimator의 Coverage probability가 높아진다.