

〈Simulation Scenario 구현 결과〉

- 12월 14일 Version

〈수정사항〉

1) “Simulation_Monte_Carlo_SD_Ratio_ver1.1.R” &
“Simulation_Performance_record_p_0.12_ver1.1.R” file 삭제

2) Hardcoding functions들 중 Sandwich robust IPW ATT Variance function 부분에서

$$\hat{\mu}_0 = \frac{\sum_{i=1}^n \hat{w}_i (1 - A_i) Y_i}{\sum_{i=1}^n \hat{w}_i (1 - A_i)} \text{로 변환 (기준에 true 값으로 설정해 두었었음} \rightarrow \text{추정량으로 변환)}$$

〈What TO DO〉

1) Outcome regression estimator 비교

: (검색해서 찾은) 강의노트 기반 Outcome regression estimator과 회귀모델 적합 후 얻은 회귀계수 추정치가 동일한 형태인지 비교

1)-1. Outcome regression estimator의 신뢰구간 계산 시, SD에 또 한번 root를 씌운 부분 수정

2) 각 방법론별 ATT, ATE 차이 계산 때 차이의 절댓값이 1e-06보다 작은지 확인

3) 새로운 성능평가지표인 $\frac{\widehat{var}(\hat{\beta})}{var(\hat{\beta})}$ 의 분모 계산할 때 N수는 1000으로 고정, REPL만 기존 1000번보다 크게 해야 함. (기본적으로 추정량은 sample size에 의존하기 때문)

〈Result〉

① 각 방법론 별 ATE, ATT 추정치들 간 차이 확인

```
> table((ATE_1_result$Outcome_reg_est-ATT_1_result$Outcome_reg_est)<=1e-06)
TRUE
1000
> table((ATE_1_result$IPW_est-ATT_1_result$IPW_est)<=1e-06)
FALSE TRUE
  480   520
> table((ATE_1_result$DR_est-ATT_1_result$DR_est)<=1e-06)
FALSE TRUE
  490   510
```

: Outcome regression 기반 estimator들의 값은 동일한 것이 당연 / IPW, DR Estimator들의 경우 같은 경우가 나오는데 무조건 다 다르게 나와야 하는 부분인지

② Outcome regression estimator 공식 비교

Check)

: Outcome regression model 적합 결과를 " $\hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha} A_i$ " 라고 하자.

: 강의노트 기반 (Duke) Outcome regression model based estimator $\hat{\tau}_{ATE} = \frac{1}{N} \sum_{i=1}^N A_i (Y_i - \hat{\mu}_0(X_i))$

$$+ (1 - A_i) (\hat{\mu}_1(X_i) - Y_i) / N$$

$$\rightarrow = A_i (Y_i - (\hat{\beta}_0 + \hat{\beta}_1^T X_i)) + (1 - A_i) (\hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha} - Y_i)$$

$$= A_i Y_i - A_i (\hat{\beta}_0 + \hat{\beta}_1^T X_i) + \hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha} - Y_i - A_i (\hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha}) + A_i Y_i$$

① $A_i = 0$ 일 때 : $\frac{1}{N} \sum_{i=1}^N \hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha} - Y_i / N \neq \hat{\alpha}$

② $A_i = 1$ 일 때 : $\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1^T X_i + \hat{\beta}_0 + \hat{\beta}_1^T X_i + \hat{\alpha} - Y_i - \hat{\beta}_0 - \hat{\beta}_1^T X_i - \hat{\alpha} + Y_i) / N$

$$= \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1^T X_i) / N \neq \hat{\alpha}$$

: 같지 않음을 확인하여서, 회귀모델 적합한 후 얻은 회귀계수 값을 이용하는 것으로 code 변환.

③ 새로운 성능평가지표인 SD Ratio의 분모 N수는 동일하게, REPL만 바꾸어서 SD Ratio 재계산 / 성능 전체적으로 재확인

```
print(ATE_1_performance)
#               Bias               rMSE Naive_var_coverage Naive_var_SD_Ratio Sandwich_robust_var_coverage Sandwich_var_SD_Ratio
# Outcome_reg -0.002379192 0.009928812           0.941          10.1284737              NA              NA
# IPW         -0.002150330 0.010115352           0.867           7.9216994              0.982          1.489921
# DR          -0.002304139 0.009961594           0.938           0.9841385              0.980          1.223564
```

```
print(ATT_1_performance)
#               Bias               rMSE Naive_var_coverage Naive_var_SD_Ratio Sandwich_robust_var_coverage Sandwich_var_SD_Ratio
# Outcome_reg -0.002379192 0.009928812           0.941          10.1284737              NA              NA
# IPW         -0.002438935 0.009934539           0.870           8.0668766              0.995          2.001384
# DR          -0.002399891 0.009940280           0.377           0.0638348              1.000          10.818076
```

-- Outcome regression estimator 기반 Naive variance의 SD Ratio 값이 큰데, Coverage는 나쁘지 않음.

-- DR ATT의 Sandwich variance estimator의 SD Ratio 값이 10 이상으로 매우 크며, Coverage는 1이다. (분산이 과대추정되고 있는 상황)

-- 분산 추정량을 다시 살펴볼 필요가 있는지?