2.1 Translating data generating process into simulations.

# 1. setting the seed

set.seed(252)

#2. setting the number of observations

n<- 5000

#3. simulating the Us

U.W1<- runif(n, min=0, max=1)

U.W2<- runif(n, min=0, max=1)

U.A<- runif(n, min=0, max=1)

U.Y<- rnorm(n, mean=0, sd=0.3)

#4. Given the random input, deterministically evaluate the F equations.

W1<- as.numeric(U.W1<0.2)

W2<- as.numeric(U.W2<plogis(0.5\*W1))

A<- as.numeric(U.A<plogis(W1\*W2))

Y<- 4\*A + 0.7\*W1 - 2\*A\*W2 + U.Y

# 5. intervene to set A=a and generate the counterfactual outcomes Y.a

Y.1<- 4\*1 + 0.7\*W1 - 2\*1\*W2 + U.Y

Y.0<- 4\*0 + 0.7\*W1 - 2\*0\*W2 + U.Y

# 6. Create a data frame with endogenous factors and counterfactual outcomes

X<- data.frame(W1, W2, A, Y, Y.1, Y.0)

head(X)

W1 W2 A Y Y.1 Y.0

1 0 0 0 -0.39069139 3.609309 -0.39069139

2 0 1 0 0.27579209 2.275792 0.27579209

3 0 1 0 0.13800411 2.138004 0.13800411

4 0 0 0 -0.03862696 3.961373 -0.03862696

5 0 1 1 2.08010486 2.080105 0.08010486

6 0 0 0 -0.02693322 3.973067 -0.02693322

summary(X)

W1 W2 A Y Y.1

Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :-0.91451 Min. :1.025

1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.: 0.08653 1st Qu.:2.090

Median :0.0000 Median :1.0000 Median :1.0000 Median : 1.66352 Median :3.032

Mean :0.1854 Mean :0.5184 Mean :0.5258 Mean : 1.66258 Mean :3.098

3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.: 3.05952 3rd Qu.:4.044

Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. : 5.32635 Max. :5.361

Y.0

Min. :-0.9749

1st Qu.:-0.1505

Median : 0.0816

Mean : 0.1346

3rd Qu.: 0.3800

Max. : 1.6252

# 7. Evaluate the causal parameter

Psi.F<- mean(Y.1 - Y.0)

Psi.F

[1] 2.9632

3 Defining target causal parameter with working MSM

#1 Generate exogenous factors, covariates, and counterfactuals

U.V.msm <- runif(n, min=0, max=3)

U.W1.msm <- U.W1

U.W2.msm <- U.W2

U.A.msm <- U.A

U.Y.msm <- rnorm(n, mean=0, sd=0.1)

V.msm <- 2 + U.V.msm

W1.msm <- W1

W2.msm <- W2

A.msm <- as.numeric(U.A.msm<plogis(W1.msm\*W2.msm + V.msm\*0.2))

Y.msm <- 2\*A.msm + 0.3\*W1.msm + 2\*A.msm\*W2.msm + 0.5\*A.msm\*V.msm + U.Y.msm

Y.1.msm <- 2\*1 + 0.3\*W1.msm + 2\*1\*W2.msm + 0.5\*1\*V.msm + U.Y.msm

Y.0.msm <- 2\*0 + 0.3\*W1.msm + 2\*0\*W2.msm + 0.5\*0\*V.msm + U.Y.msm

Y.a.msm <- c(Y.1.msm, Y.0.msm)

a.msm<- c( rep(1,n), rep(0, n) )

#2 Create data frame X.msm

X.msm<- data.frame(V.msm, a.msm, Y.a.msm)

#3 Evaluate target causal parameter

workMSM <- glm(formula=Y.a.msm ~ a.msm\*V.msm, data=X.msm)

workMSM

Call: glm(formula = Y.a.msm ~ a.msm \* V.msm, data = X.msm)

Coefficients:

(Intercept) a.msm V.msm a.msm:V.msm

0.0552329 3.0900971 -0.0002508 0.4848198

Degrees of Freedom: 9999 Total (i.e. Null); 9996 Residual

Null Deviance: 63640

Residual Deviance: 5348 AIC: 22130

4. Interpret results: Treatment A increases outcome Y by approximately 3 units (or 3 pounds at the end of the study). This is somewhat close to the estimands we calculated earlier. V (age) has a relatively negligible effect on counterfactual outcome. The effect term (age\*treatment) increases the child’s weight by approximately ½ pound by the end of the study.