Day 1, Practical 2, Hely's solution

Helene Charlotte Wiese Rytgaard

May 30, 2023

Task 1.

```
set.seed(15)
head(sim.data <- sim.fun(1000))
```

```
   id         X1           X2 X3 A Y
1:  1  0.4084562  0.38996075  0 0 0
2:  2 -1.2198243 -1.67449303  1 0 0
3:  3  1.8658349 -2.22881407  0 1 1
4:  4  0.6036221 -0.01388672  0 0 0
5:  5 -0.5317124  0.57686435  0 0 0
6:  6  1.9554368  0.15718650  0 0 1
```

1 Implement the estimating equation estimator and its variance

Task 2.

1.

```
fit.f <- glm(Y~A+X1+X2+X3, family=binomial, data=sim.data)
fit.pi <- glm(A~X1+X2+X3, family=binomial, data=sim.data)
sim.data[, pred.EY1:=predict(fit.f, type="response", newdata=copy(sim.data)[, A:=1])]
sim.data[, pred.EY0:=predict(fit.f, type="response", newdata=copy(sim.data)[, A:=0])]
sim.data[, pred.pi1:=predict(fit.pi, type="response", newdata=sim.data)]
```

2.

```
(est.ee <- sim.data[, mean(A/pred.pi1*(Y-pred.EY1) - (1-A)/(1-pred.pi1)*(Y-pred.EY0) +
    pred.EY1 - pred.EY0)])
```

```
[1] 0.06638761
```

3.

```
(var.eic <- sim.data[, mean((A/pred.pi1*(Y-pred.EY1) - (1-A)/(1-pred.pi1)*(Y-pred.EY0)
    + pred.EY1 - pred.EY0)^2)/nrow(sim.data)])
```

```
[1] 0.000861793
```

```
(ci.ee <- est.ee + c(-1,1)*1.96*sqrt(var.eic))
```

```
[1] 0.008849243 0.123925984
```

2 Compare with the TMLE estimator

Task 3.

```r
library(tmle)
tmle.fit <- tmle(Y=sim.data$Y, A=sim.data$A,
        cbind(X1=sim.data$X1,
              X2=sim.data$X2,X3=sim.data$X3),
        gform=A~X1+X2+X3, ## treatment model
        Qform=Y~A+X1+X2+X3, ## outcome model
        family="binomial",
        cvQinit=FALSE)
#-- get the ATE estimate:
tmle.fit$estimates$ATE$psi
#-- get the variance estimate:
tmle.fit$estimates$ATE$var
#-- get the confidence interval:
tmle.fit$estimates$ATE$CI
```

```
[1] 0.06626344
[1] 0.0008581074
[1] 0.008848237 0.123678637
```

3 Look at results of simulation studies

Task 4.

```r
library(here)
estimator.list <- readRDS(paste0(here(), "/data/sim-data-output/",
                "save-est-sim-setting-1",
                ".rds"))
```

Task 5.

```r
fit.tmle <- unlist(estimator.list$fit.tmle2)
fit.ee <- unlist(estimator.list$fit.ee2)
var.tmle <- unlist(estimator.list$fit.tmle2.var)
var.ee <- unlist(estimator.list$fit.ee2.var)
```

```r
message(paste0("mse tmle: ", mean((fit.tmle - mean(fit.tmle))^2)))
message(paste0("mse ee estimator: ", mean((fit.ee - mean(fit.ee))^2)))
message(paste0("variance tmle: ", var(fit.tmle)))
message(paste0("variance ee estimator: ", var(fit.ee)))
message(paste0("bias tmle: ", mean(fit.tmle-ATE)))
message(paste0("bias ee estimator: ", mean(fit.ee-ATE)))
message(paste0("coverage tmle: ", mean(fit.tmle-1.96*sqrt(var.tmle)<=ATE & fit.tmle
    +1.96*sqrt(var.tmle)>=ATE)))
message(paste0("coverage ee estimator: ", mean(fit.ee-1.96*sqrt(var.ee)<=ATE & fit.ee
    +1.96*sqrt(var.ee)>=ATE)))
```

```
mse tmle: 0.000556845093131028
mse ee estimator: 0.000556839212085858
variance tmle: 0.000557961015161351
```

```
variance ee estimator: 0.000557955122330519
bias tmle: -0.00160891683180579
bias ee estimator: -0.00159067911571245
coverage tmle: 0.96
coverage ee estimator: 0.962
```

We can also compare:

```r
message(paste0("sd tmle: ", sd(fit.tmle)))
message(paste0("mean se tmle: ", mean(sqrt(var.tmle))))
message(paste0("sd ee estimator: ", sd(fit.ee)))
message(paste0("mean se ee estimator: ", mean(sqrt(var.ee))))
```

```
sd tmle: 0.0236211984277121
mean se tmle: 0.0242444770105457
sd ee estimator: 0.0236210736913147
mean se ee estimator: 0.0243424187350596
```

Task 6.

```r
fit.miss.tmle <- unlist(estimator.list$fit.tmle)
fit.miss.ee <- unlist(estimator.list$fit.ee)
var.miss.tmle <- unlist(estimator.list$fit.tmle.var)
var.miss.ee <- unlist(estimator.list$fit.ee.var)
```

```r
message(paste0("mse tmle: ", mean((fit.miss.tmle - mean(fit.miss.tmle))^2)))
message(paste0("mse ee estimator: ", mean((fit.miss.ee - mean(fit.miss.ee))^2)))
message(paste0("variance tmle: ", var(fit.miss.tmle)))
message(paste0("variance ee estimator: ", var(fit.miss.ee)))
message(paste0("bias tmle: ", mean(fit.miss.tmle-ATE)))
message(paste0("bias ee estimator: ", mean(fit.miss.ee-ATE)))
message(paste0("coverage tmle: ", mean(fit.miss.tmle-1.96*sqrt(var.miss.tmle)<=ATE &
    fit.miss.tmle+1.96*sqrt(var.miss.tmle)>=ATE)))
message(paste0("coverage ee estimator: ", mean(fit.miss.ee-1.96*sqrt(var.miss.ee)<=ATE
    & fit.miss.ee+1.96*sqrt(var.miss.ee)>=ATE)))
```

```
mse tmle: 0.00089473471916182
mse ee estimator: 0.000893969530476735
variance tmle: 0.000896527774711243
variance ee estimator: 0.000895761052581899
bias tmle: -0.000783695637668334
bias ee estimator: -0.000777411300778136
coverage tmle: 0.936
coverage ee estimator: 0.938
```

We can also compare:

```r
message(paste0("sd tmle: ", sd(fit.miss.tmle)))
message(paste0("mean se tmle: ", mean(sqrt(var.miss.tmle))))
message(paste0("sd ee estimator: ", sd(fit.miss.ee)))
message(paste0("mean se ee estimator: ", mean(sqrt(var.miss.ee))))
```

```
sd tmle: 0.0299420736541617
mean se tmle: 0.0300289535191218
sd ee estimator: 0.0299292674915692
mean se ee estimator: 0.0301175322459571
```

Task 7.

```r
library(here)
estimator.list <- readRDS(paste0(here(), "/data/sim-data-output/",
                    "save-est-sim-setting-2",
                    ".rds"))
```

```r
fit.tmle <- unlist(estimator.list$fit.tmle2)
fit.ee <- unlist(estimator.list$fit.ee2)
var.tmle <- unlist(estimator.list$fit.tmle2.var)
var.ee <- unlist(estimator.list$fit.ee2.var)
message(paste0("mse tmle: ", mean((fit.tmle - mean(fit.tmle))^2)))
message(paste0("mse ee estimator: ", mean((fit.ee - mean(fit.ee))^2)))
message(paste0("variance tmle: ", var(fit.tmle)))
message(paste0("variance ee estimator: ", var(fit.ee)))
message(paste0("bias tmle: ", mean(fit.tmle-ATE)))
message(paste0("bias ee estimator: ", mean(fit.ee-ATE)))
message(paste0("coverage tmle: ", mean(fit.tmle-1.96*sqrt(var.tmle)<=ATE & fit.tmle
    +1.96*sqrt(var.tmle)>=ATE)))
message(paste0("coverage ee estimator: ", mean(fit.ee-1.96*sqrt(var.ee)<=ATE & fit.ee
    +1.96*sqrt(var.ee)>=ATE)))
```

```
mse tmle: 0.00181014222691649
mse ee estimator: 0.00177221890373885
variance tmle: 0.00181376976644939
variance ee estimator: 0.00177577044462811
bias tmle: 0.000883767504956664
bias ee estimator: 0.00101642732487951
coverage tmle: 0.946
coverage ee estimator: 0.956
```

```r
message(paste0("sd tmle: ", sd(fit.tmle)))
message(paste0("mean se tmle: ", mean(sqrt(var.tmle))))
message(paste0("sd ee estimator: ", sd(fit.ee)))
message(paste0("mean se ee estimator: ", mean(sqrt(var.ee))))
```

```
sd tmle: 0.0425883759545888
mean se tmle: 0.0397125330316192
sd ee estimator: 0.0421398913694389
mean se ee estimator: 0.0403883886578484
```

```r
fit.miss.tmle <- unlist(estimator.list$fit.tmle)
fit.miss.ee <- unlist(estimator.list$fit.ee)
var.miss.tmle <- unlist(estimator.list$fit.tmle.var)
var.miss.ee <- unlist(estimator.list$fit.ee.var)
message(paste0("mse tmle: ", mean((fit.miss.tmle - mean(fit.miss.tmle))^2)))
```

```
message(paste0("mse ee estimator: ", mean((fit.miss.ee - mean(fit.miss.ee))^2)))
message(paste0("variance tmle: ", var(fit.miss.tmle)))
message(paste0("variance ee estimator: ", var(fit.miss.ee)))
message(paste0("bias tmle: ", mean(fit.miss.tmle-ATE)))
message(paste0("bias ee estimator: ", mean(fit.miss.ee-ATE)))
message(paste0("coverage tmle: ", mean(fit.miss.tmle-1.96*sqrt(var.miss.tmle)<=ATE &
    fit.miss.tmle+1.96*sqrt(var.miss.tmle)>=ATE)))
message(paste0("coverage ee estimator: ", mean(fit.miss.ee-1.96*sqrt(var.miss.ee)<=ATE
    & fit.miss.ee+1.96*sqrt(var.miss.ee)>=ATE)))
```

mse tmle: 0.00504335124157197
mse ee estimator: 0.0055914632575187
variance tmle: 0.00505345815788774
variance ee estimator: 0.00560266859470812
bias tmle: 0.00498946280696468
bias ee estimator: 0.00403961703692303
coverage tmle: 0.902
coverage ee estimator: 0.922

```
message(paste0("sd tmle: ", sd(fit.miss.tmle)))
message(paste0("mean se tmle: ", mean(sqrt(var.miss.tmle))))
message(paste0("sd ee estimator: ", sd(fit.miss.ee)))
message(paste0("mean se ee estimator: ", mean(sqrt(var.miss.ee))))
```

sd tmle: 0.0710876793677198
mean se tmle: 0.0666878315039104
sd ee estimator: 0.0748509759101918
mean se ee estimator: 0.0730752786729624

```
#-- correctly specified:
fit.wt.tmle <- unlist(estimator.list$fit.wt.tmle2)
var.wt.tmle <- unlist(estimator.list$fit.wt.tmle2.var)
#-- misspecified:
fit.wt.miss.tmle <- unlist(estimator.list$fit.wt.tmle)
var.wt.miss.tmle <- unlist(estimator.list$fit.wt.tmle.var)
```

```
#-- correctly specified:
message(paste0("mse tmle: ", mean((fit.wt.tmle - mean(fit.wt.tmle))^2)))
message(paste0("variance tmle: ", var(fit.wt.tmle)))
message(paste0("bias tmle: ", mean(fit.wt.tmle-ATE)))
message(paste0("coverage tmle: ", mean(fit.wt.tmle-1.96*sqrt(var.wt.tmle)<=ATE & fit.
    wt.tmle+1.96*sqrt(var.wt.tmle)>=ATE)))
```

mse tmle: 0.00144353715294481
variance tmle: 0.00144643001297075
bias tmle: 0.000336486603559796
coverage tmle: 0.946
```

```
#-- correctly specified (without weight truncation):
message(paste0("mse tmle: ", mean((fit.tmle - mean(fit.tmle))^2)))
message(paste0("variance tmle: ", var(fit.tmle)))
message(paste0("bias tmle: ", mean(fit.tmle-ATE)))
message(paste0("coverage tmle: ", mean(fit.tmle-1.96*sqrt(var.tmle)<=ATE & fit.tmle
    +1.96*sqrt(var.tmle)>=ATE)))
```

mse tmle: 0.00181014222691649
variance tmle: 0.00181376976644939
bias tmle: 0.000883767504956664
coverage tmle: 0.946

```
#-- misspecified:
message(paste0("mse tmle: ", mean((fit.wt.miss.tmle - mean(fit.wt.miss.tmle))^2)))
message(paste0("variance tmle: ", var(fit.wt.miss.tmle)))
message(paste0("bias tmle: ", mean(fit.wt.miss.tmle-ATE)))
message(paste0("coverage tmle: ", mean(fit.wt.miss.tmle-1.96*sqrt(var.wt.miss.tmle)<=
    ATE & fit.wt.miss.tmle+1.96*sqrt(var.wt.miss.tmle)>=ATE)))
```

mse tmle: 0.00329099813961518
variance tmle: 0.00329759332626771
bias tmle: 0.0139252005009493
coverage tmle: 0.878

```
#-- misspecified (without weight truncation):
message(paste0("mse tmle: ", mean((fit.miss.tmle - mean(fit.miss.tmle))^2)))
message(paste0("variance tmle: ", var(fit.miss.tmle)))
message(paste0("bias tmle: ", mean(fit.miss.tmle-ATE)))
message(paste0("coverage tmle: ", mean(fit.miss.tmle-1.96*sqrt(var.miss.tmle)<=ATE &
    fit.miss.tmle+1.96*sqrt(var.miss.tmle)>=ATE)))
```

mse tmle: 0.00504335124157197
variance tmle: 0.00505345815788774
bias tmle: 0.00498946280696468
coverage tmle: 0.902