Day 1, Practical 2

Helene Charlotte Wiese Rytgaard

May 14, 2023

In this practical we will continue to work with the simulated data from Practical 1.

1 Simulate data

We will work with the simulation function defined in the first practical.

Task 1: Use the simulation function from the first practicals from day 1 (Task 1) to draw a random dataset with sample size $n = 1000$.

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

2 Implement the estimating equation estimator

Task 2: Implement the estimating equation estimator, as outlined below:

1. Fit the models below for the outcome regression f and the propensity score π.
2. Use `fit.f` to predict the conditional expectations $\mathbb{E}_P[Y \mid A, X]$ and $\mathbb{E}_P[Y \mid A = a, X]$. Add these as columns to the dataset.
3. Use `fit.pi` to estimate the propensity score $\pi(a \mid X) = P(A = a \mid X)$. Add this as a column to the simulated dataset.
4. Implement $\hat{\psi}_n^{\text{ee}}$ based on Equation (1):

$$\begin{aligned}\hat{\psi}_n^{\text{ee}} &= \tilde{\Psi}_{\text{ee}}(\hat{f}_n, \hat{\pi}_n, \hat{P}_n) \\ &= \frac{1}{n}\left\{\left(\frac{A_i}{\hat{\pi}_n(1 \mid X_i)} - \frac{1 - A_i}{\hat{\pi}_n(0 \mid X_i)}\right)\left(Y_i - \hat{f}_n(A_i, X_i)\right) + \hat{f}_n(1, X_i) - \hat{f}_n(0, X_i)\right\}. \qquad (1)\end{aligned}$$

5. Implement the variance estimator based on Equation (2):

$$\hat{\sigma}_n^2 = \mathbb{P}_n\{\tilde{\phi}^*(\hat{f}_n, \hat{\pi}_n)\}^2/n = \frac{1}{n^2}\sum_{i=1}^{n}\{\tilde{\phi}^*(\hat{f}_n, \hat{\pi}_n)(O)\}^2 \qquad (2)$$

3 Compare with the TMLE estimator

Task 3. Load the `tmle` package and use the `tmle()` function to get the TMLE estimate and variance using the same models as in **Task 2** using the code below. Check that you get about the same, and comment.

```
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
```

4 Look at results of simulation studies

Task 4. You can access simulation results from **Task 10** (Practical 1) as follows by downloading the file:

data/sim-data-output/save-est-sim-setting-1.rds

from github. Load it to R as below (changing the path):

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

Task 5. The code below shows how vectors of estimates and estimated variances saved across the simulation repetitions are extracted from the object above. These results are for correctly specified models. Compute the bias, variance, mean squared error and coverage for the TMLE estimator and the estimating equation (EE) estimators. Comment on the results.

```
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)
```

Task 6. The code below shows how vectors of estimates and estimated variances saved across the simulation repetitions are extracted from the object above, for the situation when misspecified models were used. Compute the bias, variance, mean squared error and coverage for the TMLE estimator and the estimating equation (EE) estimators. Comment on the results.

```
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)
```

Task 7. You can access simulation results from a simulation study positivity issues as in Section 4 (Task 9) from Practical 1 by downloading the file:

data/sim-data-output/save-est-sim-setting-2.rds

from github. Load it to R as below (changing the path), and then repeat **Task 5** and **Task 6** above. Comment on the results.

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

NB: Note that for these results are produced for TMLE both with and without weight truncation. Those *with* weight truncation are accessed as follows:

```
#-- 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)
```