Statistical Learning in Mediation Analysis

Lab for Chapter 5: Estimating interventional randomized direct and indirect effects in R

David Benkeser
Emory University

Iván Díaz
Weill Cornell Medicine

MODULE 17

Summer Institute in Statistics for Clinical and Epidemiological Research

August 2022

Contents of this lab

1. Illustration of estimation of the interventional direct and indirect effects using the `medoutcon` R package
2. Do-it-yourself analysis of a real dataset

Illustrative dataset

- Recall the dataset `weight_behavior` from the mma R package used in Lab 3.
- We set up the dataset the same way as for Lab 3, removing missing data:

```
library(mma)
library(tidyverse)
# load and examine data
data(weight_behavior)
dim(weight_behavior)
```

```
## [1] 691  15
```

```
# drop missing values
weight_behavior <- weight_behavior %>%
  drop_na() %>%
  as_tibble()
```

Setting up the problem

As before, we focus on the causal effects of participating in a sports team (`sports`) on the BMI of children (`bmi`), taking into consideration mediators given by (`exercises`, `overweigh`). All other measured covariates are taken to be potential baseline confounders.

For this lab we will use `snack` as an intermediate confounder.

The function `medoutcon()`

Recall the main function of the package: `medoutcon()`. In Lab 3 we discussed the arguments to this function, but left one pending: Z. This is the intermediate confounder of interest. In this function, if Z is null, the function estimates the NDE and NIE, if Z is a vector, the function estimates the interventional direct and indirect effects

Estimating the interventional direct effect

First, we set up the super learner libraries as in Lab 3:

```
library(sl3)
# instantiate learners
fglm_lrnr <- Lrnr_glm_fast$new()
lasso_lrnr <- Lrnr_glmnet$new(alpha = 1, nfolds = 3)
rf_lrnr <- Lrnr_ranger$new(num.trees = 200)
# create learner library and instantiate super learner ensemble
lrnr_lib <- Stack$new(fglm_lrnr, lasso_lrnr, rf_lrnr)
sl_lrnr <- Lrnr_sl$new(learners = lrnr_lib, metalearner = Lrnr_nnls$new())
```

Estimating the interventional direct effect

We now compute the effects:

```
library(medoutcon)
# compute one-step estimate of the natural direct effect
ide_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = (as.numeric(weight_behavior$snack) - 1),
  M = weight_behavior[, c("exercises", "overweigh")],
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "direct",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
```

Estimating the interventional indirect effect

```r
# compute one-step estimate of the natural indirect effect
iie_onestep <- medoutcon(
  W = weight_behavior[, c("age", "sex", "race", "tvhours")],
  A = (as.numeric(weight_behavior$sports) - 1),
  Z = (as.numeric(weight_behavior$snack) - 1),
  M = weight_behavior[, c("exercises", "overweigh")],
  Y = weight_behavior$bmi,
  g_learners = lasso_lrnr,
  h_learners = lasso_lrnr,
  b_learners = lasso_lrnr,
  effect = "indirect",
  estimator = "onestep",
  estimator_args = list(cv_folds = 5)
)
```

Results

```
summary(ide_onestep)
```

```
## # A tibble: 1 x 7
##   lwr_ci param_est upr_ci var_est eif_mean estimator param
##    <dbl>     <dbl>  <dbl>   <dbl>    <dbl> <chr>     <chr>
## 1 -0.321     0.209  0.739  0.0732 1.98e-15 onestep   direct~
```

```
summary(iie_onestep)
```

```
## # A tibble: 1 x 7
##   lwr_ci param_est upr_ci var_est eif_mean estimator param
##    <dbl>     <dbl>  <dbl>   <dbl>    <dbl> <chr>     <chr>
## 1  0.431     0.989   1.55  0.0810 1.38e-15 onestep   indire~
```

Note that, in contrast with the results of Lab 3 where the effect of snacking was part of the indirect effect, here the effect of snacking is part of the direct effect.

Do-it-yourself analysis of a real dataset

Please revisit the exercise at the end of Lab 4, and compute the estimate of the interventional direct and indirect effects for that problem/dataset.