Targeted Minimum Loss-based Estimation (TMLE) for Causal Inference in Biostatistics

Helene Charlotte Wiese Rytgaard (hely@sund.ku.dk)

September 9, 2021

1 Welcome to the course

Welcome to the course: *Targeted Minimum Loss-based Estimation (TMLE) for Causal Inference in Biostatistics*. The course starts on Monday the 27th of September, 2021.

Please read the following instructions carefully.

1. The practicals will mainly consist of computer exercises. To participate, you should bring your own laptop.
2. A list of relevant `R` packages for the practicals can be found in Section 5.
3. To prepare for the lectures, see Section 4. Section 3 further gives an overview of the course.

2 Course location

The course will take place in the following rooms of CSS:

Monday	September 27th	8am–13pm	CSS 7.0.18
Monday	September 27th	13–15pm	CSS 2.1.42
Tuesday	September 28th	8am–15pm	CSS 35.3.13
Wednesday	September 29th	8am–15pm	CSS 7.0.18

3 Overview of course

Targeted minimum loss-based estimation (TMLE) is a general framework for estimation of causal effects that combines semiparametric efficiency theory and machine learning in a two-step procedure. The main focus of the course is to understand overall concepts, the theory, and the application of TMLE. A sufficient background in mathematics and statistics is needed, although we emphasize that making the theory *practical* is really the point here (thus, many mathematical details will skipped). For the larger part of the course, we focus on the simple example of estimating an average treatment effect, with the general principles being similar for other parameters.

The course runs over three full days (8am–3pm, lunch from 12–1pm), planned largely as follows:

Day 1. On day 1 we go through the roadmap of targeted learning (both from a theoretical and a practical angle) and give a brief introduction to basic concepts of causal inference. The afternoon will be about the background theory of nonparametric efficiency theory.

Day 2. On day 2 we introduce the TMLE, relating its foundation to the theory from the afternoon of day 1. We go through the targeting step, its purpose and how it is carried out for an average treatment effect. The afternoon will cover super learning.

Day 3. On day 3 we move on to time-varying settings. We discuss time-varying treatments and time-dependent confounding that hinders the use of "classical" statistical methods.

4 Reading plan

For **Day 1** and **Day 2** we recommend that you read Kennedy (2016). This text provides a valuable introduction to the concepts of nonparametric efficiency theory that we need to understand TMLE. You may want to focus on pages 1–13, although Section 4.1 is also quite useful.

For **Day 2**, you may read Chapter 5 of van der Laan and Rose (2011) as an introduction to TMLE. You may further read Chapter 3 of the same book as an introduction to super learning.

For **Day 3**, you should read Kreif et al. (2017); we will use this paper in one of the practicals. You may further read Schwab et al. (2014) about the `ltmle` (longitudinal TMLE) software (skip Section 4 on marginal structural models).

5 Relevant `R` packages

```r
install.packages("tmle")
install.packages("ggplot2")
install.packages("data.table")
install.packages("randomForestSRC")
install.packages("SuperLearner")
install.packages("ltmle")
```

References

Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research*, pp. 141–167. Springer.

Kreif, N., L. Tran, R. Grieve, B. De Stavola, R. C. Tasker, and M. Petersen (2017). Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: a demonstration of longitudinal targeted maximum likelihood estimation. *American journal of epidemiology 186*(12), 1370–1379.

Schwab, J., S. Lendle, M. Petersen, and M. van der Laan (2014). ltmle: Longitudinal targeted maximum likelihood estimation. *R package version 0.9 3.*

van der Laan, M. J. and S. Rose (2011). *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media.