Public Health 243D — Fall 2018 Syllabus
Adaptive Designs

Class meets TuTh 9:30–11A in Cory 289
Class Number: 32680
Course Website: https://bcourses.berkeley.edu/courses/1473876

Instructor: Mark van der Laan
Email: laan@berkeley.edu
Office Hours: Th 11A–12P in Berkeley Way West 5311

Course Description

This course examines theory and statistical methods for analyzing data generated by adaptive group sequential designs, or experimental designs that permit dynamic modifications of their course through data-driven adjustments of randomization probabilities. It also considers the construction of targeted adaptive designs that are optimized with respect to a criterion that translates what the investigator is most concerned about (i.e., minimizing the number of patients in a clinical trial that are assigned to their corresponding inferior treatment). For example, we will develop sequential adaptive designs that are tailored to learn the optimal individualized treatment rule, while simultaneously randomly assigning new subjects according to the learned rule: this estimation problem is often referred to as the contextual multi-arm bandit problem, a type of reinforcement learning.

A variety of potential design adaptations of great practical importance for which this theory applies will be investigated. Topics to be covered include: group sequential adaptive designs, targeted adaptive designs, targeted maximum likelihood estimation of causal contrasts for adaptive designs, analysis of such estimators based on the corresponding martingale central limit theorem and maximal inequalities for martingale processes, nonparametric statistical models for stationary time-series data, and adaptive designs within a single time-series for learning optimal individualized treatment rules.

Credit Hours: 3

Learning Outcomes

At the completion of this course, students should be able to:

1. Formulate the statistical estimation problem for a group sequential adaptive design and a single time-series; which involves defining the data, model, and target parameter.
2. Simulate a group sequential adaptive design for a causal question. Justify simulation design.
3. Construct a targeted group sequential adaptive design that converges to an optimal fixed design, such as one that optimizes information for answering a particular causal question.
4. Learn and interpret the optimal individualized treatment rule within a group sequential adaptive design, with statistical inference regarding its effect relative to a control.
5. Implement a targeted maximum likelihood estimator for estimation and inference based on data generated by a group sequential adaptive design. Explain the formal theoretical basis in terms of the corresponding martingale CLT and maximal inequality for martingale process.
6. Evaluate optimal dynamic treatments within a single time-series that regularly measure covariates, outcomes and treatments.

Prerequisites

The following prerequisites are recommended but not required. For students who are concerned about their background affecting their success in the course, please consult Prof. van der Laan.

- Public Health C240A / Statistics C245A.
- Statistics 201A-B (or older version Statistics 200A-B).
- Statistics C239A / Political Science C236A or Public Health 252D.

Requirements and Materials

There will be optional readings from the following textbooks that serve as excellent references, encompassing many of the topics covered in this course and more. These textbooks are freely available in PDF for students to download through SpringerLink:

- Van der Laan, Mark J., & Rose, Sherri (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data.* {abbr. "vdL&R (2011) book" in schedule}
- Van der Laan, Mark J., & Rose, Sherri (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies.* {abbr. "vdL&R (2018) book" in schedule}

Mark van der Laan's blog (`https://vanderlaan-lab.org/post/`) is encouraged for web-based discussion. Students are invited to submit a question to the blog at any time by sending an email to `vanderlaan.blog@berkeley.edu` at any time. Course assignments will require submission of questions to the blog.

Course capture recordings, project guidelines, the syllabus, and attendance will be posted on the private bCourses page, `https://bcourses.berkeley.edu/courses/1473876`.

Grading

- *Project (**60%**):* Consists of two brief presentations, one comprehensive presentation, and a final report due at the end of the term. Typical projects involve setting up a simulation representing targeted adaptive designs, implementation of estimators of causal questions of interest, formal statistical inference, and evaluation of performance of procedures. Group sizes may range from 1 to 3 students. Project guidelines with details on the required deliverables are posted on bCourses.
- *Participation (**30%**):* Includes attendance, engagement during class, and completion of short, open-note worksheets during class. Students must email Prof. van der Laan if they cannot attend class. Attendance records will be made available on bCourses.
- *Blog Question (**10%**):* In a group of 1 to 3, students are required to ask a question to the blog by sending an email to `vanderlaan.blog@berkeley.edu`. The question can be related to (but is not limited to) final projects, research, or course content. The criteria is broad, just focus on a statistical question. Students will receive no credit if their question is easily "Google-able".

Important Due Dates

Project Presentations I – Proposal	Tuesday, Sept. 25
Project Presentations II – Progress	Tuesday, Oct. 30
Blog Question Assignment	Tuesday, Nov. 20
Project Presentations III – Comprehensive	Tuesday, Dec. 04
Project Report	Sunday, Dec. 09

Course Policies

Accommodations: Please speak with Prof. van der Laan as soon as possible if you require any particular accommodations and the necessary arrangements will be worked out.

Scheduling Conflicts: Notify Prof. van der Laan by the second week of the term about any known or potential conflicts (e.g., religious observances, interviews, team activities, conferences).

Honor Code

"*As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.*"

The purpose of the Honor Code is to enhance awareness of the need for the highest possible levels of integrity and respect on campus, both within and outside the academic context. We hope and believe that the code will catalyze a series of ongoing conversations about our principles and practices. Together, through engagement, we can create a consistent message and ethos in our classrooms, labs, departments, and throughout the academic enterprise, to ensure that the core values of academic integrity and honesty are being embraced by both students and faculty. Please carefully read the Honor Code (`http://asuc.org/honorcode/index.php`).

Academic Integrity

One of the most important values of an academic community is the balance between the free flow of ideas and the respect for the intellectual property of others. Researchers don't use one another's research without permission; scholars and students always use proper citations in papers; professors may not circulate or publish student papers without the writer's permission; and students may not circulate or post materials (handouts, exams, syllabi — any class materials) from their classes without the written permission of the instructor.

Any test, paper or report submitted by you and that bears your name is presumed to be your own original work that has not previously been submitted for credit in another course unless you obtain prior written approval to do so from your instructor. In all of your assignments, you may use words or ideas written by other individuals in publications, web sites, or other sources, but only with proper attribution. If you are not clear about the expectations for completing an assignment or taking a test or examination, be sure to seek clarification from your instructor or GSI beforehand. Finally, you should keep in mind that as a member of the campus community, you are expected to demonstrate integrity in all of your academic endeavors and will be evaluated on your own merits. The consequences of cheating and academic dishonesty including a formal discipline file, possible loss of future internship, scholarship, or employment opportunities, and denial of admission to graduate school are simply not worth it.

Tentative Course Schedule

Week	Content
1: Aug. 23	• *Topics:* Course introduction.
2: Aug. 28 – 30	• *Topics:* Recap of fundamentals for single experiment: data, model, target parameter, efficient influence curve, exact second order remainder, least favorable sub-model, loss function, TMLE, analysis of TMLE. • *Reading:* Chapters 4 – 5 vdL&R (2011) book.
3: Sept. 04 – 06	• *Topics:* Sequential adaptive design, likelihood of sequential adaptive design, statistical model, statistical estimation problem, oracle fixed design, adaptive design converging to oracle fixed design, examples of oracle fixed design. • *Reading:* Section 1 vdL (2008) tech report.
4: Sept. 11 – 13	• *Topics:* Sequential adaptive design learning optimal individualized treatment rule, multiple bandit problem, Sequential adaptive design maximizing information for answering question of interest. • *Reading:* Chapter 29 vdL&R (2011) book, Chapter 24 vdL&R (2018) book, Section 2 vdL (2008) tech report.
5: Sept. 18 – 20	• *Topics:* MLE for adaptive design likelihood assuming parametric model, analysis of MLE, martingale CLT, asymptotic equicontinuity of martingale process, MLE treating parametric model as working model, analysis of this MLE, parametric versus realistic models for adaptive designs. • *Reading:* Chapter 29 vdL&R (2011) book, Sections 4–5 vdL (2008) tech report.
6: Sept. 25 – 27	• *Topics:* TMLE of target parameter for sequential adaptive design, general. • *Reading:* vdL (2008) tech report, Chapter 24 vdL&R (2018) book. • **Project Presentations I – Proposal on Tues Sept. 25**.
7: Oct. 02 – 04	• *Topics:* TMLE of ATE for sequential adaptive design, analysis of TMLE. • *Reading:* Chambaz & vdL (2014) SJoS article, vdL (2008) tech report, Chapter 29 vdL&R (2011) book.
8: Oct. 09 – 11	• *Topics:* TMLE of mean outcome under optimal rule for sequential adaptive design, analysis of TMLE. • *Reading:* Chapter 24 vdL&R (2018), Chapter 29 vdL&R (2011), Chambaz, Zheng & vdL (2017) AoS article.
9: Oct. 16 – 18	• *Topics:* Highly Adaptive Lasso (HAL) for sequential adaptive designs and more general time-series (i.e. ordered sequence of experiments with common parameters), analysis of HAL. • *Reading:* TBA
10: Oct. 23 – 25	• *Topics:* Online super-learning for sequential adaptive designs and more general time-series, oracle inequality for online cross-validation selector. Ivana Malenica presents online super-learner `R` package. • *Reading:* Chapter 18 vdL&R (2018) book, Benkeser et al. (2017) SiM article.

Week	Content
11: Oct. 30 – Nov. 1	• *Topics:* Examples of sequential adaptive sampling design, with application to controlling an epidemic. Aurelien Bibaut presents research, in collaboration with UCSF, Hugh Sturrock, and Ben Arnold. • **Project Presentations II – Progress on Tues Oct. 30**.
12: Nov. 6 – 8	• *Topics:* General time-series model and estimation problem, estimation of marginal causal effects based on observing a single time series with alternating treatment and outcomes, Online targeted learning. • *Reading:* Chapter 19 vdL&R (2018), vdL & Lendle (2014) tech report.
13: Nov. 13 – 15	• *Topics:* Estimation of context specific short-term causal effects based on time series with alternating treatment and outcomes. • *Reading:* TBA
14: Nov. 20	• *Topics:* Adaptive design within a single time series learning the optimal rule for the unit. Comparing optimal dynamic treatments for i.i.d. data for population versus optimal dynamic treatment for a single unit/individual. • *Reading:* TBA • **Blog question due Tues Nov. 20, email** `vanderlaan.blog@berkeley.edu`.
15: Nov. 27 – 29	• *Topics:* Generalizations to observing multiple time series, allowing for network dependence, asymptotics in number of time points and/or number of time series, future research topics.
16: Dec. 4 – 6	• **Project Presentations III – Comprehensive on Tues Dec. 4** • **Project Report due Sun Dec. 09, email pdf to** `laan@berkeley.edu`.