Invited sessions for ISCB42 (Lyon, 2021)
Session 1

Causal inference in continuous time for dense longitudinal data from wearable devices

Organizers : Linda Valeri : Columbia University Mailman School of Public Health, US

Mark van der Laan : University of California, Berkeley, US
TMLE for Causal Effects based on continuous time longitudinal data structures

Suchi Saria, John C. Malone Associate Professor of Computer Science, Statistics and Health Policy and Director of the Machine Learning and Healthcare Lab, Johns Hopkins University

Roy Adams, Johns Hopkins University, USA
The Impact of Time Series Length and Discretization on Longitudinal Causal Estimation Methods

Tianchen Qian : Department of Statistics, University of California, Irvine, United States
Modeling the Time-varying Effect of Mobile Health Intervention Using Nested Longitudinal Data with Zero-Inflation

Title: TMLE for Causal Effects based on continuous time longitudinal data structures

Mark van der Laan[1] and Helene Rytgaard[2]

[1] UC Berkeley, USA - presenting author : laan@berkeley.edu

[2] University of Copenhagen, Denmark

Abstract:

In many applications one is concerned with estimation of the causal impact of a multiple time point intervention on a final outcome based on observing a sample of longitudinal data structures. We consider the case that subjects are monitored at a finite set of time-points on a continuous time-scale, and at these monitoring times treatment actions and or time-dependent covariates and outcomes are collected. Current methods based on sequential regression break down under this setting. We develop a new targeted maximum likelihood estimator that still avoids estimation of the conditional densities for outcome and covariates of likelihood, but instead estimates a conditional mean function. We also consider a TMLE that involves estimation of the conditional densities. We develop highly adaptive lasso estimators of the nuisance functions and establish asymptotic efficiency of the TMLE under minimal conditions. In particular, we demonstrate these new TMLEs for estimation of treatment specific survival functions for single time-point interventions on competing survival times. Advantages relative to first discretizing the time scale and using currently avaiable corresponding TMLE are discussed.

The Impact of Time Series Length and Discretization on Longitudinal Causal Estimation Methods

Roy Adams[1], Suchi Saria[1,2,3], Michael Rosenblum[4]

[1] Department of Computer Science, Johns Hopkins University, USA
[2] Department of Applied Mathematics and Statistics, Johns Hopkins University, USA
[3] Bayesian Health, USA
[4] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, USA
(presenting author)

The use of observational time series data to assess the impact of multi-time point interventions is becoming increasingly common as more health and activity data are collected and digitized via wearables, social media, and electronic health records. Such time series may involve hundreds or thousands of irregularly sampled observations. One common analysis approach is to simplify such time series by first discretizing them into sequences before applying a discrete-time estimation method that adjusts for time-dependent confounding. In certain settings, this discretization results in sequences with many time points; however, the empirical properties of longitudinal causal estimators have not been systematically compared on long sequences. In this talk, we compare three representative longitudinal causal estimation methods on simulated and real clinical data and analyze the impact of sequence length and discretization bin width on estimator performance. Our simulations and analyses assume a Markov structure and that longitudinal treatments/exposures are binary-valued and have at most a single jump point. We identify sources of bias that arise from temporally discretizing the data and provide practical guidance for discretizing data and choosing between methods when working with long sequences. Additionally, we compare these estimators on electronic health record data, evaluating the impact of early treatment for patients with a life-threatening complication of infection called sepsis.

Modeling the Time-varying Effect of Mobile Health Intervention Using Nested Longitudinal Data with Zero-Inflation

Tianchen Qian[1], Justin Zhu[2], Ashley Walton[3], Predrag Klasnja[4], Susan Murphy[2,5]

[1] Department of Statistics, University of California, Irvine, United States (presenting author)
[2] Department of Statistics, Harvard University, United States
[3] Department of Philosophy, Dartmouth College, United States
[4] School of Information, University of Michigan, Ann Arbor, United States
[5] Department of Computer Science, Harvard University, United States

Mobile health intervention studies, where interventions such as push notifications are delivered to individuals repeatedly during the course of the study, provide intensive longitudinal data with time-varying treatments. This provides unprecedented opportunity to understand how the causal effect of such interventions changes over time and is modified by contextual information, and it in turn informs the development and optimization of these interventions. Motivated by a physical activity study (HeartSteps micro-randomized trial), we develop methods to model the time-varying effect of push notifications that suggest exercise, where the outcome of interest at each decision point is the minute-level step count over the subsequent hour. We develop marginal and conditional models for the causal effect on the nested longitudinal data (decision points within each participant and minutes within each decision point), and we incorporate techniques such as penalized splines, Gaussian Processes, and generalized additive models to allow for flexible causal effect curves over time. In addition, we address the challenge that the outcome (step count at each minute) is zero-inflated.