CAUSAL INFERENCE WITH TIME-VARYING TREATMENT IN MOBILE HEALTH

Tiantian Ye and Brook Luers
Department of Statistics, College of Literature, Science and Art, University of Michigan

Summary

In randomized trials of mobile health interventions, treatments, which often consist of messages on a smartphone, can be delivered to each participant on many occasions (e.g. twice a day or once a week). Scientific interest lies in the causal effect of treatment on a specified near-term health behavior, such as the number of steps walked in a day. Because treatment assignment, behavioral outcomes, and potential covariates (e.g. current location) vary within each participant over time, common regression methods can lead to bias when estimating the average causal effect of a mobile health treatment. More specifically, GEE can produce biased estimates when modeling the conditional mean of a longitudinal outcome using time-varying covariates, unless a diagonal variance-covariance matrix is used in estimation (Sullivan Pepe and Anderson 1994). And mixed effects models can suffer from the same bias.

We wish to extend this estimation method to permit mixed effects modeling and hence estimation of individual causal effects. Here we present simulations which show that generalized estimating equations (GEE) and linear mixed models can produce biased estimates of average causal effects when participants are repeatedly randomized to receive treatment. We propose a solution based on a regression model for individual causal effects.

Introduction

As a motivating example, HeartSteps is a mobile application meant to encourage physical activity. In a micro-randomized trial (Klass et al. 2015) involving HeartSteps, participants were asked each evening, with probability 0.5, to create a plan for physical activity during the next day. One scientific goal is to understand the causal effect of creating this activity plan on the following day’s step count. In this case the health outcome of interest for participant i at time t, denoted Y_{it}, is the number of steps taken on day t. Letting $A_{it}$ indicate treatment assignment, where $A_{it} \equiv 1$ if participant i created a physical activity plan, the randomization probabilities are $P(A_{it} = 1) = 0.5$. Using potential outcomes notation, $Y_{it}$ is the number of steps taken by participant i when she receives treatment on day t and $Y_{it}^-\mathbb{I}(A_{it} = 0)$ is the number of steps taken by participant i when she does not receive treatment on day t. The observed number of steps is $Y_{it}^\mathbb{I}(A_{it})$. So only one of $Y_{it}^\mathbb{I}(A_{it})$ or $Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0))$ is observed for each participant.

Goals

This paper: Estimate the average causal effect: $E(Y_{it}^\mathbb{I}(A_{it}) - Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0)))$.

(Future work) Estimate the individual causal effects: $\Delta_{it} = Y_{it}^\mathbb{I}(A_{it}) - Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0))$.

Notations

$A_{it}$ = binary treatment indicator for participant i at occasion t sequentially randomized with $P(A_{it} = 1) = 0.5$

$Y_{it}^\mathbb{I}(A_{it})$ = potential outcome for participant i at occasion t when receiving treatment $A_{it} = 1$

$Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0))$ = observed outcome measured for participant i at occasion t when receiving treatment $A_{it} = 0$

$\Delta_{it} = Y_{it}^\mathbb{I}(A_{it}) - Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0))$ = individual causal effect for participant i at occasion t.

Proposed Solution

Model the individual causal effect, $\Delta_{it}$, using $\Delta_{it} = \beta Y_{it}^\mathbb{I}(A_{it})$ as the response variable.

It can be shown that $E(\Delta_{it}) = E(\Delta_{it})$.

Simulations

These simulations illustrate the bias described above and explore the feasibility of regression models for the individual causal effect using $\Delta_{it}$ as the response variable.

Two generative models

$$Y_{it} = \theta + (\beta \cdot Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0))) + \epsilon_{it}$$

$$Y_{it} = \theta + \eta_{it} + \beta Y_{it}^\mathbb{I}(A_{it}^-\mathbb{I}(A_{it} = 0)) + \epsilon_{it}$$

where $\epsilon_{it}$ is the unobserved error term, $\eta_{it}$ is the (random error) $\sim N(0, \sigma^2)$, and $u_{it}$ is independent of $\epsilon_{it}$ and all random variables are independent across participants $i = 1, \ldots, N$.

These generative models lead to the same individual and average causal effects:

$\Delta_{it} = \beta + u_{it} + \epsilon_{it}$, $\Delta = \beta + N(0, \sigma^2)$

A weighted and centered estimation method, similar to GEE, consistently estimates causal effects of mobile health treatments moderated by time-varying covariates (Boruvka et al. 2016). We wish to extend this estimation method to permit mixed effects modeling and hence estimation of individual causal effects.

Conclusion

The conditions given in Sullivan Pepe and Anderson 1994 for unbiased estimation are satisfied by the generating model (1) but not by model (2). Both estimation methods produced unbiased estimates for generating model (1). Mixed models and GEE with exchangeable working covariance produced biased estimates of $\beta$ in generating model (2) when fitting regression model (3). This bias is alleviated when we model the individual causal effect with response variable $\Delta_{it}$.

These simulations illustrate the bias that can arise when estimating the average causal effect of a time-varying treatment on a time-varying response using GEE or mixed models. Regression models for the individual causal effect, $\Delta_{it}$, using $\Delta_{it}$ as the response variable show promise in overcoming this bias. Future simulations using more realistic generative models will include time-varying covariates which moderate the effect of treatment as well as treatment assignment probabilities that depend on prior treatment and prior outcome measurements.

Reference

