
A Bayesian Approach to Principal Stratification with Multiple Binary Post-treatment Covariates

Bo Liu

Department of Statistical Science
Duke University
Durham, NC 27705
b1226@duke.edu

Haoliang Zheng

Department of Statistical Science
Duke University
Durham, NC 27705
hz228@duke.edu

Abstract

Abstract

Key words: Key words

1 Introduction

Randomization, when possible, is desirable in studying the causal effect of one treatment against the other in clinical trials and other experiments. By randomization, the subpopulation receiving either treatment is homogeneous in both observed and unobserved covariates. Hence, the source of bias from the treatment assignment is eliminated, and therefore any difference in outcome between the two treatment groups can be interpreted as the causal effect between the treatments.

In real studies, there may frequently exist post-treatment covariates which are highly correlated with the randomized treatment and have non-ignorable effect on the outcome. The effect of these covariates is not homogeneous within each randomized treatment group, thus introducing bias and increasing difficulty in inference on the causal effect. A common example of such post-treatment covariates is non-compliance, where the actual treatment one received might be opposite from what they are randomized to, as in the studies of [add citations here](#). In these situations, the actual treatment is not randomized, and there is unobserved confounding between the actual treatment and the outcome.

One approach is the standard intention-to-treat (ITT) method, which ignores the actual treatment and compares the outcome between two randomized treatment groups. This preserves the randomization, but the estimand is the effectiveness of the treatment instead of the efficacy, which might be undesirable when the efficacy is of clinical interest. Another direction, pioneered by [citation of first IV paper](#) and introduced to the context of non-compliance in a landmark paper [add citation Angrist et al](#), is the instrumental variable (IV) approach. Here, the assigned treatment can be viewed as an instrumental variable, as it is highly correlated with the actual treatment, but might not have direct effect on the outcome. When the no-direct-effect assumption is questionable by domain knowledge, principal stratification [add citation](#) is also a feasible approach. The idea is to define principal strata by the potential values of post-treatment covariates under both treatment arms, which is not dependent on the treatment assignment and is determined prior to the treatment. The principal stratification approach identifies the underlying strata and estimates the causal effect within the strata of interest.

In this report, we generalize the idea of principal stratification to multiple binary post-treatment covariates. We also provide a user-friendly software to facilitate Bayesian inference on estimating the causal effect.

2 Notation