ABSTRACT
Causal inference concerns not only the average effect of the treatment on the outcome but also the underlying mechanism through an intermediate variable of interest. Principal stratification characterizes such mechanism by targeting subgroup causal effects within principal strata, which are defined by the joint potential values of an intermediate variable. Due to the fundamental problem of causal inference, principal strata are inherently latent, rendering it challenging to identify and estimate subgroup effects within them. A line of research leverages the principal ignorability assumption that the latent principal strata are mean independent of the potential outcomes conditioning on the observed covariates. Under principal ignorability, we derive various nonparametric identification formulas for causal effects within principal strata in observational studies, which motivate estimators relying on the correct specifications of different parts of the observed-data distribution. Appropriately combining these estimators further yields new triply robust estimators for the causal effects within principal strata. These new estimators are consistent if two of the treatment, intermediate variable, and outcome models are correctly specified, and they are locally efficient if all three models are correctly specified. We show that these estimators arise naturally from either the efficient influence functions in the semiparametric theory or the model-assisted estimators in the survey sampling theory. We evaluate different estimators based on their finite-sample performance through simulation, apply them to two observational studies, and implement them in an open-source software package.
Authors for paper- Zhichao Jiang, Shu Yang, Peng Ding