Nonparametric and Semiparametric Regression with Missing Outcomes Using Weighted Kernel and Profile Estimating Equations

Xihong Lin
Harvard University

We consider nonparametric and semi-parametric regression when an outcome is missing at random (MAR). We first consider nonparametric regression of a scalar outcome on a covariate under MAR. We show that nonparametric kernel regression estimation based only on complete cases is generally inconsistent. We propose inverse probability weighted (IPW) kernel estimating equations and a class of augmented inverse probability weighted (AIPW) kernel estimating equations for nonparametric regression under MAR. Both approaches do not require specification of a parametric model for the error distribution and the estimators are consistent when the probability that a sampling unit is observed, i.e., the selection probability or the response probability, is known by design or is estimated using a correctly specified model. We show that the AIPW kernel estimator is double-robust in that it yields valid inference if either the model for the response probability is correctly specified or a model for the conditional mean of the outcome given covariates and auxiliary variables is correctly specified. In addition, we argue that adequate choices of the augmented term in the AIPW kernel estimating equation help increase the efficiency the estimator of the nonparametric regression function. We study the asymptotic properties of the proposed IPW and AIPW kernel estimators. We extend the results to semi-parametric regression under MAR where one covariate effect is modeled non-parametrically and some covariate effects are modeled parametrically. IPW and AIPW profile-kernel estimating equations are proposed to estimate the parametric component. Asymptotic semi-parametric efficiency is studied. We perform simulations to evaluate their finite sample performance, and apply the proposed methods to the analysis of the AIDS Costs and Services Utilization Survey data.