Kernel density estimation with missing data

Suzanne R. Dubnicka
Kansas State University

Abstract: In most parametric statistical analyses, knowledge of the distribution of the response variable, or of the errors, is important. As this distribution is not typically known with certainty, one might initially construct a histogram or estimate the density of the variable of interest to gain insight regarding the distribution and its characteristics. However, when the response variable is incomplete, a histogram will only provide a representation of the distribution of the observed data. In the AIDS Clinical Trial Study (ACT) protocol 175, for example, interest lies in CD4 counts at final follow-up, but CD4 counts collected at final follow-up are missing for more than one third of the patients. We propose methods for estimating the density of an incomplete response variable when auxiliary data are available. The proposed estimator is based on the Horvitz-Thompson estimator, and the propensity scores are estimated nonparametrically. The density estimator will be evaluated and applied to the ACT data. Extensions will also be discussed.