Kernel density estimation with imputed group membership information

Søren Feodor Nielsen[1]
[1] Copenhagen Business School, Denmark

The problem I will consider in this talk is the estimation of an unknown density of a subgroup of continuous measurements, when the group membership is missing for some of the observations. So, the data consists of a continuous outcome variable and a partially observed grouping variable, and we are trying to estimate the conditional density of the outcome given the group. If the probability of not having information on the group membership depends on the continuous measurement, then an estimator based on complete cases, i.e. restricting attention to observations with known group membership, will be biased. Wang, He & Gunzler (2012) have suggested an inverse probability weighted estimator, which is consistent and asymptotically normal. A more traditional solution to this problem is to use imputation: "Guess" which group the observations belong to and base the estimator on the observations that are known or guessed to belong in the relevant group. Over the last two decades inverse probability weighting estimators have become increasingly popular in non- and semi-parametric missing data problems such as this one. Whereas imputation may seem somewhat ad hoc, weighting seems more principled. Inverse probability weighting is however inefficient, and by imputing the missing data we are able to utilize more of the information in the data. In this talk I will compare weighting and different imputation strategies. I will also discuss optimal estimation and outline some results that improve on the usual expression for the asymptotic variance.

Keywords: Kernel density estimation, imputation, optimal estimation, variance estimation

References:

Tang, W., H. He, D. Gunzler (2012) Kernel smoothing density estimation when group membership is subject to missing *Journal of Statistical Planning and Inference* **142**, 685–694.