Variable Selection in Semiparametric Regression Modeling

Speaker: Runze Li, Pennsylvania State University

Time: 3:00-5:00 Monday, May 14, 2007

Venue: C406

Abstract:

In this talk, I will introduce how to select significant variables in semiparametric modeling. Variable selection for semiparametric regression models consists of two components: model selection for nonparametric components and selection of significant variables for the parametric portion. Thus, semiparametric variable selection is much more challenging than parametric variable selection (e.g., linear and generalized linear models) because traditional variable selection procedures including stepwise regression and the best subset selection now require separate model selection for the nonparametric components for each submodel. This leads to very heavy computational burden. In this paper, we propose a class of variable selection procedures for semiparametric regression models using nonconcave penalized likelihood. We establish the rate of convergence of the resulting estimate. With proper choices of penalty functions and regularization parameters, we show the asymptotic normality of the resulting estimate, and further demonstrate that the proposed procedures perform as well as an oracle procedure. A semiparametric generalized likelihood ratio test is proposed to select significant variables in the nonparametric component. We investigate the asymptotic behavior of the proposed test and demonstrate that its limiting null distribution follows a chi-squared distribution, which is independent of the nuisance parameters. Extensive Monte Carlo simulation studies are conducted to examine the finite sample performance of the proposed variable selection procedures.