Semiparametric regression analysis of panel count data and interval-censored failure time data

Bin Yao

This dissertation discusses three important research topics on semiparametric regression analysis of panel count data and interval-censored data. Both types of data arise commonly in real-life studies in many fields such as epidemiology, social science, and medical research. In these studies, subjects are usually examined multiple times at periodical or irregular follow-up examinations. For panel count data, the response variable is the counts of some recurrent events, whose exact occurrence times are usually unknown. For interval-censored data, the response variable is the time to some events of interest, often called survival time or failure time, and the exact response time is never observed but is known to fall within some interval formed by two examination times. The primary goal for both types of data is to study effects of covariates on the response variable and can be completed by regression analysis. The three topics are: (i) we propose a gamma frailty non-homogeneous Poisson process model for the regression analysis of panel count data to account for the within-subject correlation. We propose an efficient estimation approach based on an EM algorithm. An R package \textit{PCDSpline} has been developed and available on CRAN to disseminate our approach. (ii) we study regression analysis of case 1 interval-censored data, also referred to as current status data, using the generalized odds-rate hazards (GORH) models. We propose an efficient estimation approach with fixed $\rho$ in the GORH models based on a novel EM algorithm. A working model approach is proposed when true value of $\rho$ is known but does not require to fit the GORH models with different $\rho$ values. (iii) We study the joint modeling of panel count data and interval-censored failure time data. The motivating example is a real-life data set about sexually transmitted infections (STI). The failure time of interest is the onset of a new STI, which has an interval-censored data structure. Another response is the number of condom non-use, which has a panel count data structure. The research interests include to estimate the covariate effects on onset time of STI and on the number of condom use as well as the statistical association between the two responses. A frailty model is proposed for this joint analysis and an EM algorithm is derived for the estimation approach.