Semiparametric regression analysis for length-biased and interval-censored data with a cure fraction

Chyong-Mei Chen

National Yang-Ming University
E-mail: cmchen2@ym.edu.tw

Abstract: Left-truncated (LT) data are often encountered in epidemiological cohort studies, where individuals are recruited according to a certain cross-sectional sampling criterion. Length-biased data, a special case of LT data, assume that the incidence of the initial event follows a homogeneous Poisson process. In this article, we consider analysis of length-biased and interval-censored (LBIC) data with a nonsusceptible fraction. We first point out the importance of a well-defined target population, which depends on the priori knowledge for the support of the failure times of susceptible individuals. Given the appropriate target population, we can proceed a length-biased sampling and draw valid inferences from sample. When there is no covariate, we show that to maximize the full likelihood function, it suffices to consider discrete version of the survival function for the susceptible individuals with jump points at left-end points of the censoring intervals. Based on this result, we propose an EM algorithm to obtain the nonparametric maximum likelihood estimates (NPMLEs) of nonsusceptible rate and the survival function of the susceptible individuals. We also develop a novel graphical method for assessing the stationarity assumption. When covariate is present, we consider analyzing LBIC data under the Cox proportional hazards model with a nonsusceptible fraction, where the probability of being susceptible is determined by the logistic regression model. We construct the full likelihood function and obtain the NPMLEs of the regression parameters by employing the EM algorithm. The large sample properties of the NPMLEs are established. The performance of the NPMLE is assessed by simulations. The metabolic syndrome data are analyzed to illustrate our method.