Analysis of semiparametric regression models for the cumulative incidence functions under the two-phase sampling designs

Unkyung Lee

Preprint no. 2016-05

Abstract

Competing risks often arise where a subject may be exposed to two or more mutually exclusive causes of failure. In the competing risks setting, the effects of covariates on the semiparametric model for cumulative incidence function can be assessed by using the direct binomial regression approach. In epidemiologic cohort studies, case-cohort study designs have been widely used to evaluate the effects of covariates on failure times when the occurrence of the failure event is rare. In this dissertation, we study estimating procedures for the cumulative incidence function based on competing risks data under case-cohort/two-phase sampling designs. First, we introduce the semiparametric model for the cumulative incidence function with missing data. The estimation procedure is based on the direct binomial regression model, which enables us to evaluate the effects of the covariates directly when there exists competing risks. We develop an estimating equation for the missing model by using the inverse probability weighting of the complete case method. However, the IPW method loses the efficiency because it uses only complete data of subjects. To overcome this inefficiency, we also propose an estimating equation for the semiparametric model by using augmented inverse probability of complete case method. The AIPW method is doubly robust and it can improve efficiency. The asymptotic properties of the proposed IPW and AIPW estimators are established. The finite-sample properties of those estimators are investigated by the simulation studies. The proposed estimating methods are applied to analyze data from the RV144 vaccine efficacy trial.

Department of Mathematics, UNC Charlotte, Charlotte, NC 28223