Statistical methods for bivariate survival data with interval sampling and application to biomedical studies

Hong Zhu, Ph.D. Candidate

Department of Biostatistics
Johns Hopkins School of Public Health

Abstract

In collection of registry or surveillance data of a disease, it is common to identify incidence of disease within a calendar time interval, then subsequently collect bivariate or multivariate survival data as markers for progression of the disease. We consider an interval sampling scheme, where the first failure event (i.e., cancer onset) is identified within a calender time interval, the occurrence of the initiating event (i.e., birth) can be retrospectively confirmed, and the observation of the second failure event (i.e., death) is subject to right censoring. To analyze this type of bivariate survival data, it is important to recognize the presence of bias arising due to interval sampling. Under stationary and semi-stationary assumptions, we develop nonparametric and semiparametric methods to estimate the joint survival function. Simulation studies demonstrate the proposed estimating methods perform well. We apply the methods of joint survival function estimation to NCI's SEER ovarian cancer registry data for illustration of the methods and theory. Moreover, it is important to incorporate sampling bias in analyzing this type of bivariate survival data based on both uncensored and censored observations. We extend the work by developing a copula models approach and studying the dependence structure of bivariate survival data . The copula models method is evaluated by simulations and illustrated by Rakai Human Immunodeficiency Virus (HIV) seroconversion data to study the disease progression of HIV infection for treatment-naive individuals.