Hazard function estimation with cause-of-death data missing at random

Qihua Wang · Gregg E. Dinse · Chunling Liu

Abstract Hazard function estimation is an important part of survival analysis. Interest often centers on estimating the hazard function associated with a particular cause of death. We propose three nonparametric kernel estimators for the hazard function, all of which are appropriate when death times are subject to random censorship and censoring indicators can be missing at random. Specifically, we present a regression surrogate estimator, an imputation estimator, and an inverse probability weighted estimator. All three estimators are uniformly strongly consistent and asymptotically normal. We derive asymptotic representations of the mean squared error and the mean integrated squared error for these estimators and we discuss a data-driven bandwidth selection method. A simulation study, conducted to assess finite sample behavior,

demonstrates that the proposed hazard estimators perform relatively well. We illustrate our methods with an analysis of some vascular disease data.