Accepted Manuscript

Semiparametric regression analysis of clustered survival data with semi-competing risks

Mengjiao Peng, Liming Xiang, Shanshan Wang

PII: S0167-9473(18)30040-9
DOI: https://doi.org/10.1016/j.csda.2018.02.003
Reference: COMSTA 6566

To appear in: *Computational Statistics and Data Analysis*

Received date : 2 August 2017
Revised date : 11 February 2018
Accepted date : 15 February 2018

Please cite this article as: Peng M., Xiang L., Wang S., Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics and Data Analysis* (2018), https://doi.org/10.1016/j.csda.2018.02.003

Semiparametric Regression Analysis of Clustered Survival Data with Semi-competing Risks

Mengjiao Peng, Liming Xiang*, Shanshan Wang

School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

Abstract

Analysis of semi-competing risks data is becoming increasingly important in medical research in which a subject may experience both nonterminal and terminal events, and the time to the intermediate nonterminal event (e.g. onset of a disease) is subject to dependent censoring by the terminal event (e.g. death) but not vice versa. Typically, both two types of events are dependent. In many applications, subjects may also be nested within clusters, such as patients in a multi-center study, leading to possible association among event times due to unobserved shared factors across subjects. To incorporate dependency within clusters and association between two types of event times, we propose a new flexible semiparametric modeling framework where a copula model is employed for the joint distribution of the nonterminal and terminal events, and their marginal distributions are modeled by Cox proportional hazards models with random effects. A nonparametric maximum likelihood estimation procedure is developed and implemented through a Monte Carlo EM algorithm. The proposed estimator is also shown to enjoy desirable asymptotic properties. Results from extensive simulation studies indicate that the proposed method performs very well in finite samples and is especially robust against misspecification of the random effects distribution. We further illustrate the practical utility of the method by analyzing data from a multi-institutional study of breast cancer.

Key Words: Copula, Clustered data, Monte Carlo EM algorithm, Proportional hazards model, Random effects, Semi-competing risks.

* Corresponding author. Tel.: +65 65137451 ; Fax: +65 65158213.
E-mail address: lmxiang@ntu.edu.sg .