

SEMIPARAMETRIC INFERENCE FOR LONGITUDINAL DATA WITH INFORMATIVE OBSERVATION TIMES AND TERMINAL EVENT

Shirong Deng, Kin-yat Liu, Wen Su and Xingqiu Zhao*

*Wuhan University, The Chinese University of Hong Kong,
City University of Hong Kong and The Hong Kong Polytechnic University*

Abstract: In many longitudinal studies, irregularly repeated measures are often correlated with observation times. Also, there may exist a dependent terminal event such as death that stops the follow-up and is subject to right censoring. To deal with such complex data, we propose a class of flexible semiparametric marginal conditional mean models for longitudinal response processes. The new models include the interaction between the observation history and some covariates, and an unknown functional form of the length from the observation time to the terminal event time, while leaving the within-subject dependence structure of the response process and patterns of the observation process to be arbitrary. For estimation of both scalar and functional parameters in the proposed models, we develop a two-stage spline-based least squares estimation approach and establish the asymptotic properties of the proposed estimators. The performance of the proposed estimation procedure is examined by simulation studies, and a longitudinal data example is provided for illustration.

Key words and phrases: Conditional modeling, empirical process, informative observation times, longitudinal data, terminal event time.

1. Introduction

Longitudinal data occur frequently in a wide variety of settings, including epidemiological studies, clinical trials, economic applications and others. The response variables and covariates are observed repeatedly at irregular time points for different subjects under study, and the observations are independent among different subjects and may be correlated within each subject. For the analysis of such longitudinal data, various parametric and semiparametric methods have been studied by Laird and Ware (1982), Liang and Zeger (1986) and Zeger and Diggle (1994) among others, and excellent reviews have been provided by Lin and Ying (2001) and Diggle et al. (2002).

A basic assumption behind the methods above is that the observation times are independent of response variable, completely or given covariates. However, such an assumption can be violated in many applications, that is, the observation

*Corresponding author. E-mail: xingqiu.zhao@polyu.edu.hk