Semiparametric analysis of recurrent events: artificial censoring, truncation, pairwise estimation and inference

Debashis Ghosh

Abstract The analysis of recurrent failure time data from longitudinal studies can be complicated by the presence of dependent censoring. There has been a substantive literature that has developed based on an artificial censoring device. We explore in this article the connection between this class of methods with truncated data structures. In addition, a new procedure is developed for estimation and inference in a joint model for recurrent events and dependent censoring. Estimation proceeds using a mixed U-statistic based estimating function approach. New resampling-based methods for variance estimation and model checking are also described. The methods are illustrated by application to data from an HIV clinical trial as with a limited simulation study.

Keywords Accelerated failure time model · Cause-specific hazard · Comparability · Competing risks · Empirical process · Semi-competing risks data

1 Introduction

In many medical and scientific settings, analysts must deal with recurrent events data. Recurrent failure time data represents an event that can potentially repeat itself during the course of a study. Examples of recurrent events are tumor recurrences from oncology studies (Byar 1980), repeated cardiovascular events in patients (Cui et al. 2008) and births by a mother in demography studies. A problem arises when subjects who have repeated events tend to withdraw from the study and become censored observations. Such a dropout mechanism violates the usual independent censoring mecha-
nism necessary for the validity of nonparametric and semiparametric inferences from recurrent failure time data (e.g., Andersen and Gill 1982; Pepe and Cai 1993; Lawless and Nadeau 1995).

To accommodate dependent censoring, a very popular approach has been to use inverse probability censoring weighted estimation techniques (Robins and Rotnitzky 1992). Such an approach requires that there exist a sufficiently rich set of covariates that fully explain the hazard of the dependent censoring mechanism. One then constructs a set of weights based on estimating this model that is then used to reweight censored observations in a manner that leads to asymptotically unbiased estimation. A generalization of this procedure is to construct so-called “doubly robust” estimating equations that yield consistent estimators if either the model for dependent censoring or for the event of interest is specified correctly. A summary of this approach can be found in van der Laan and Robins (2002), and an application to the recurrent events problem has been given recently by Miloslavsky et al. (2004).

An alternative approach to this problem was initially proposed in the single-event case by Lin et al. (1996). They developed semiparametric estimation and inferential procedures for a bivariate accelerated failure time model for the time to event and for dependent censoring. While the model for the dependent censoring time can be estimated using standard techniques, that for the time to event cannot be handled in a straightforward way. This is because the dependent censoring induces a covariate-dependent dropout mechanism with respect to the time to event. Lin et al. (1996) introduce an artificial censoring device in order to construct an unbiased estimating function for the time to event. Recently, Peng and Fine (2007) developed a related estimation procedure for the same model. We will describe both of these approaches in Sect. 2.1.

For the situation of recurrent events, this approach has been extended by many authors. Ghosh and Lin (2003) developed an estimation procedure in which the target estimand was the expected number of events as a function of time. Such a model was extended to account for noncompliance in clinical trials by Matsui (2004). One could alternatively look at the times between the events, which are called interevent or gap times. For this approach, Chang (2000) proposed a scale-change model for analyzing the gap time between multiple events in the presence of dependent censoring. She also used the artificial censoring technique to construct an estimation procedure that would yield consistent estimators.

In this article, we revisit the model of Ghosh and Lin (2003) and use the ideas from Peng and Fine (2007) to construct an unbiased U-statistic-based estimating function. A secondary goal is provide a new perspective from which to interpret it. This is done using ideas from the analysis of truncated survival data (Bhattacharya et al. 1983; Wang et al. 1986). In particular, we shall see that an effect of the artificial censoring is to achieve comparability between pairs of observations, which will be shown in Sect. 2.2. A different interpretation of the artificial censoring was provided in a nice article by Joffe (2001), who used ideas from causal inference to explain the justification. The structure of this paper is as follows. In Sect. 2, we describe the observed data structure and review some previous methods for a univariate version of the problem (Lin et al. 1996; Peng and Fine 2007). It also provides an equivalence between the observed data structure with event times that are subject to right-truncation. We then