Journal of Multivariate Analysis 143 (2016) 153–164

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Joint analysis of current count and current status data

Chi-Chung Wen [a], Yi-Hau Chen [b,*]

[a] Department of Mathematics, Tamkang University, Taiwan
[b] Institute of Statistical Science, Academia Sinica, Taiwan

ARTICLE INFO

Article history:
Received 24 September 2014
Available online 11 September 2015

AMS subject classifications:
62N01
62N02
62P10

Keywords:
Correlated data
Frailty model
Panel count
Self-consistency
Semiparametric maximum likelihood

ABSTRACT

We consider joint analysis of event times to a recurrent and a non-recurrent event, with the event time data subject to type I interval censoring. The motivation arises from a survey study, which collected current count data for time to occurrences of fracture (recurrent event), and current status (i.e., type I interval censored) data for time to osteoporosis (non-recurrent event). The aim of the study is to examine risk factors for, and levels of association between, the recurrent and non-recurrent events. We propose a joint analysis of current count and current status data based on a joint modeling for recurrent and non-recurrent events. In the proposed framework, a non-homogeneous Poisson process is assumed for the recurrent event, a proportional hazards model is assumed for failure time of the non-recurrent event, and the two event time processes share a common gamma frailty. A semiparametric maximum likelihood estimator, together with a stable computation algorithm, is developed for the joint model. The parametric (covariate effects and frailty) and nonparametric (baseline mean and cumulative hazard functions) components of the estimator are consistent at rates of square root and cubic root of the sample size, respectively. The asymptotic normality for the parametric component of the estimator is established. The application to the survey data mentioned above shows that, female is a common strong risk factor for both fracture and osteoporosis, and times to fracture and osteoporosis are highly associated.

1. Introduction

Current status data, also called type I interval censored data, commonly arise in survey studies. In such data, a survey (examination) time and an indicator of whether an event of interest has "failed" by the survey time are available for each subject. Inferences on the underlying failure time distribution with observed current status data have been widely investigated, which can in fact be considered under the more general context of interval-censored event time analysis, where there may be multiple examination times in each subject [7,3,12].

When the event of interest is recurrent by nature, such as repeated infections or injuries, but is only examined at an examination, the only information for each subject's event times includes the examination time and the number of events up to the examination time. We will call this type of data the "current count" data, which is a special case of the well known "panel count" data (e.g., [13,18,1,19,14]), where multiple examination times and the numbers of events up to the examination times are available for each subject.

This work focuses on joint analysis of current count and current status data, which has not been formally addressed in literature, though plenty of studies have been done for individual analysis of the two types of interval-censored data. Our

* Correspondence to: Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
E-mail address: yhchen@stat.sinica.edu.tw (Y.-H. Chen).

http://dx.doi.org/10.1016/j.jmva.2015.08.015
0047-259X/

motivation arises from the 2005 fracture-osteoporosis survey study in Taiwan, where one of the study aims is to understand the respective risk factors for fracture and osteoporosis, and the degrees of association between the underlying processes for the occurrences of fracture and osteoporosis. Note that fracture is recurrent while osteoporosis is non-recurrent by nature, and current count data were observed for the former while current status data were observed for the latter in this survey study, since there was only one examination time for each subject.

Our proposal to joint analysis of current count and current status data is a likelihood-based approach based on joint frailty models. Specifically, we assume the two underlying event time processes, one for times to the recurrent event and the other for failure time of the non-recurrent event, share a common gamma frailty whereby the association between the two processes is induced. Given the frailty, the recurrent event process is assumed to be a nonhomogeneous Poisson process, while the failure of the non-recurring event process is assumed to follow a proportional hazards model, and the two processes are independent. The modeling framework is semiparametric, in that the frailty distribution and the covariate effects in the two event time processes are parametrically modeled, while the baseline mean and cumulative hazard functions are fully unspecified (see Section 2). To make inferences on the proposed models, in Section 3 we propose semiparametric maximum likelihood estimation, where the two nonparametric functions (baseline mean and cumulative hazard) are treated as non-decreasing step functions with jumps only at the examination times. The resulting semiparametric maximum likelihood estimator (SPMLE) for the model parameter is shown to be consistent. The convergence rates are shown to be cubic root and square root of the sample size, respectively, for the nonparametric (the baseline mean and cumulative functions) and parametric components (regression and frailty parameters) of the parameter. We derive the asymptotic normality for the parametric component of the proposed SPMLE, and propose an estimator for the asymptotic variance–covariance matrix of the SPMLE using numerical second differentiation of the log-likelihood function. In Section 4, we develop a hybrid algorithm for computation of the SPMLE, which consists of a self-consistency algorithm for solving nonparametric baseline mean and cumulative hazard functions, and a conventional optimization algorithm for solving the regression and the gamma frailty parameters.

Simulation studies present in Section 5 reveal that the proposed SPMLE, together with its asymptotic theory, performs well in scenarios with moderate sample sizes. We apply the proposed method to joint analysis of the current count data for fracture and the current status data for osteoporosis observed in the 2005 fracture-osteoporosis survey study in Taiwan. The analysis provides results regarding risk factors for, as well as the levels of association between, the occurrences of fracture and osteoporosis (see Section 6).

2. Model description

Let $\mathsf{N}(t)$ be the number of occurrences of a recurrent event over $(0, t]$, and T the failure time of another related, non-recurrent event for one subject. Denote by C the examination time, and Z_1 and Z_2 the vectors of covariates that may affect $\mathsf{N}(t)$ and T respectively. The covariates Z_1 and Z_2 are allowed to have elements in common. When current count data for $\mathsf{N}(t)$ and current status data for T are observed, the observation for each subject then consists of $Y = (C, N, \Delta, Z_1, Z_2)$, where $N \equiv \mathsf{N}(C)$ indicates the number of occurrences of the recurrent event up to the examination time C, and $\Delta \equiv I(T \leq C)$ indicates whether the failure time T precedes C or not.

To assess the effect of Z_1 on $\mathsf{N}(t)$ and the effect of Z_2 on T, while accounting for association between two counting processes $N_1(t)$ and $N_2(t)$, with $N_1(t) = \mathsf{N}(t)$ for a recurrent event and $N_2(t) = I[T \leq t]$ for a non-recurrent event induced by the event time T, a shared and unobserved frailty η is introduced. Suppose η is a gamma random variable with mean 1 and variance $\gamma > 0$. We assume that, conditional on (η, Z_1, Z_2), N and T are independent. The recurrent process N is a nonhomogeneous Poisson process with the mean function, which is also the cumulative incidence function, given by

$$E(\mathsf{N}(t)|\eta, Z_1, Z_2) = \eta \exp(\beta_1' Z_1) \Lambda_1(t). \tag{1}$$

The cumulative hazard function of the failure time T is given by

$$\Lambda(t|\eta, Z_1, Z_2) = \eta \exp(\beta_2' Z_2) \Lambda_2(t). \tag{2}$$

In this joint modeling, β_1 and β_2 denote the vectors of unknown regression parameters, Λ_1 the increasing baseline mean function, and Λ_2 the baseline cumulative hazard function. In the model, large γ represents strong association between the recurrent process and the failure time. It can be shown that the marginal mean function of $\mathsf{N}(t)$ is $E(\mathsf{N}(t)|Z_1) = \exp(\beta_1' Z_1)\Lambda_1(t)$, and the marginal survival function of T is $S(t|Z_2) = (1 + \gamma e^{\beta_2' Z_2} \Lambda_2(t))^{-1/\gamma}$.

Assume C and $(\eta, \mathsf{N}(t), T)$ are independent conditioned on (Z_1, Z_2) and the conditional distribution of C given (Z_1, Z_2) does not depend on the parameters of interest. Then the likelihood for a single observation $Y = (C, N, \Delta, Z_1, Z_2)$ is

$$\begin{aligned} L(\theta, \Lambda_1, \Lambda_2)(Y) &\propto E_\eta \left\{ [\eta e^{\beta_1' Z_1} \Lambda_1(C)]^N \exp(-\eta e^{\beta_1' Z_1} \Lambda_1(C)) [1 - \exp(-\eta e^{\beta_2' Z_2} \Lambda_2(C))]^{\Delta} [\exp(-\eta e^{\beta_2' Z_2} \Lambda_2(C))]^{1-\Delta} \right\} \\ &= \frac{\Gamma(N + \gamma^{-1})}{\Gamma(\gamma^{-1})} [\gamma e^{\beta_1' Z_1} \Lambda_1(C)]^N [(1 + \gamma e^{\beta_1' Z_1} \Lambda_1(C))^{-N-\gamma^{-1}} \\ &\quad - (1 + \gamma e^{\beta_1' Z_1} \Lambda_1(C) + \gamma e^{\beta_2' Z_2} \Lambda_2(C))^{-N-\gamma^{-1}}]^{\Delta} [(1 + \gamma e^{\beta_1' Z_1} \Lambda_1(C) + \gamma e^{\beta_2' Z_2} \Lambda_2(C))^{-N-\gamma^{-1}}]^{1-\Delta}, \end{aligned}$$

where $\theta = (\beta_1', \beta_2', \gamma)'$ and E_η is the expectation with respect to η.