Journal of the Korean Statistical Society 45 (2016) 156–165

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Frailty model approach for the clustered interval-censored data with informative censoring☆

Jinheum Kim [a], Youn Nam Kim [b], Seong W. Kim [c,*]

[a] Department of Applied Statistics, University of Suwon, Gyeonggi-Do 445-743, Republic of Korea

[b] Clinical Trials Center Severance Hospital, Yonsei University Health System, Seoul 120-749, Republic of Korea

[c] Department of Applied Mathematics, Hanyang University, Gyeonggi-Do 426-791, Republic of Korea

ARTICLE INFO

Article history:
Received 25 February 2015
Accepted 9 September 2015
Available online 9 October 2015

AMS 2000 subject classifications:
primary 62N01
secondary 62F10

Keywords:
Clustered interval-censored data
EM algorithm
Frailty effect
Gauss–Hermite approximation
Informative censoring

ABSTRACT

Interval censoring is frequently encountered in many clinical trials with periodic follow-up as the time of a specific event, such as death, is determined within an interval. Most existing methodologies with regression analysis were extended and developed under the assumption of non-informative censoring mechanism. However, this assumption sometimes does not hold. Subsequently, it is impossible to test the dependence or independence assumption of the censoring mechanism. One remedy to circumvent these difficulties is to impose extra assumptions or modeling. In this article, we employ the Cox proportional hazards models with a shared frailty effect incorporated with clustered interval-censored data for which there exists a dependency between the failure and visiting times. The parameters are estimated via the EM algorithm. Simulations are performed to investigate the finite-sample properties of the proposed method. Finally, two real datasets are analyzed to demonstrate our methodologies.

1. Introduction

Interval-censored outcomes often arise as the time of a specific event is known to have occurred within a period. This type of censoring mechanism can be regarded as more general compared to right censoring. In many clinical trials in conjunction with periodic follow-up, each subject is observed through several examinations. However, a subject may skip one or more pre-appointed visits and return with the failure already occurred. In these situations, the event time lies on an interval of the form $(L, R]$, where L is the time last seen without disease, and R is the first time the subject appeared with disease. Thus, subjects without any disease should have $R = \infty$, hence right-censored. On the contrary, a subject with $L = 0$ corresponds to left-censored. Extensive work has been conducted to analyze interval-censored data. Sun (2006) provides comprehensive literature on interval censoring.

Most existing methodologies with regression analysis were developed under the assumption of non-informative censoring mechanism (Zhang, Sun, & Sun, 2005). That is, the failure time and the visiting times of subjects are assumed to be independent. However, in some situations, this assumption does not hold. For instance, when failure occurs, a patient could

☆ This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2011-0010889).

* Corresponding author.
E-mail address: seong125@gmail.com (S.W. Kim).

http://dx.doi.org/10.1016/j.jkss.2015.09.002
1226-3192/

experience some symptoms prior to or together with failure. This makes the patients tend to visit the doctor earlier than scheduled (Wang, Sun, & Tong, 2010; Zhang, Sun, Sun, & Finkelstein, 2007). Under these circumstances, it is more plausible to expect that the exact failure time would be closer to the right endpoint of the observed interval. However, it is virtually impossible to observe both the failure time and the visiting times simultaneously. Subsequently, testing the dependence or independence assumption of the censoring mechanism is impossible. To circumvent these difficulties, one solution could be to impose extra assumptions or modeling. In this article, we employ the Cox proportional hazards models with a shared frailty effect incorporated with interval-censored data for which there exists a dependency between the failure time and the visiting times.

Huang and Wolfe (2002) have dealt with the clustered right-censored data assuming the dependence between the failure time and the censoring time. Zhang et al. (2005, 2007) and Wang et al. (2010) have utilized frailty models to explain a dependence structure between the failure time and the censoring times for the interval-censored data with informative censoring. Kim and Kim (2012) recently proposed a procedure of the regression parameter estimation using the Cox proportional hazards model with a shared frailty for the clustered interval-censored data under the assumption that both the failure time and the censoring times are independent. In this article, we extend the arguments of Huang and Wolfe (2002) and Kim and Kim (2012) to the clustered interval-censored data in the presence of informative censoring.

The rest of this article is organized as follow. Section 2 introduces the proposed models and their parameter estimation procedures. Section 3 provides simulation results. The proposed method is presented with two real examples in Section 4. Finally, brief concluding remarks are given in Section 5. All the theoretical derivations are given in the Appendix.

2. Models and parameter estimation

Let T_{ij} denote the failure time of interest for the jth subject within the ith cluster ($i = 1, \dots, n; j = 1, \dots, n_i$). Let U_{ij} and V_{ij} be two observation times with $U_{ij} \le V_{ij}$. Although we cannot observe the exact failure time, T_{ij}, it is only less than or equal to U_{ij}, between U_{ij} and V_{ij}, or greater than V_{ij}. Define the gap time W_{ij} as $W_{ij} = V_{ij} - U_{ij}$ if V_{ij} is available; otherwise $W_{ij} = \infty$. Let $\boldsymbol{x}_{ij}$ be a $p \times 1$ vector of covariates. Thus, the observed data for the jth subject within the ith cluster have the form of

$$\boldsymbol{o}_{ij} = (U_{ij}, V_{ij}, \delta_{1ij}, \delta_{2ij}, \boldsymbol{x}_{ij}')',$$

where $\delta_{1ij} = I(T_{ij} \le U_{ij})$ and $\delta_{2ij} = I(U_{ij} < T_{ij} \le V_{ij})$. Subsequently, the entire observations can be expressed as $\boldsymbol{o} = (\boldsymbol{o}_1', \dots, \boldsymbol{o}_n')'$, where $\boldsymbol{o}_i = (\boldsymbol{o}_{i1}', \dots, \boldsymbol{o}_{in_i}')'$.

Let r_i be a frailty for the ith cluster. Assume that T_{ij}'s within the ith cluster share an unobservable frailty and they are conditionally independent given $\boldsymbol{x}_{ij}$ and r_i. Suppose T_{ij} depends on U_{ij} and V_{ij} (or W_{ij}). In order to incorporate this dependency, conditional on $\boldsymbol{x}_{ij}$ and r_i, we consider the Cox proportional hazards models with a shared frailty for T_{ij}, U_{ij}, and W_{ij}, respectively, as follows:

$$\lambda_t(t|\boldsymbol{x}_{ij}, r_i) = \lambda_{0t}(t) \exp\{\boldsymbol{\beta}_t'\boldsymbol{x}_{ij} + r_i\}, \tag{1}$$

$$\lambda_u(t|\boldsymbol{x}_{ij}, r_i) = \lambda_{0u}(t) \exp\{\boldsymbol{\beta}_u'\boldsymbol{x}_{ij} + \alpha_u r_i\}, \tag{2}$$

$$\lambda_w(t|\boldsymbol{x}_{ij}, r_i) = \lambda_{0w}(t) \exp\{\boldsymbol{\beta}_w'\boldsymbol{x}_{ij} + \alpha_w r_i\}, \tag{3}$$

where $\boldsymbol{\beta}_t$, $\boldsymbol{\beta}_u$, and $\boldsymbol{\beta}_w$ are the regression coefficients; $\lambda_{0t}(\cdot)$, $\lambda_{0u}(\cdot)$, and $\lambda_{0w}(\cdot)$ are the baseline hazard functions for T_{ij}, U_{ij}, and W_{ij}, respectively; and α_u and α_w are unknown parameters representing the degree of dependency between T_{ij} and U_{ij} and between T_{ij} and W_{ij}, respectively. Finally, r_i is assumed to be a normal random variable with a mean of 0 and variance θ. That is, $r_i \sim N(0, \theta)$. Further, assume that T_{ij}, U_{ij}, and W_{ij} are conditionally independent given $\boldsymbol{x}_{ij}$ and r_i.

Given $\boldsymbol{x}_{ij}$ and r_i, the likelihood function L_{ij} for the jth subject within the ith cluster can be expressed as follows: when $\delta_{1ij} = 1$, that is, T_{ij} is left-censored at u_{ij} and therefore W_{ij} is right-censored at 0, noting that $P(W_{ij} > 0|\boldsymbol{x}_{ij}, r_i) = 1$,

$$\begin{aligned} L_{ij} &= P(U_{ij} = u_{ij}, W_{ij} > 0, T_{ij} \in (0, u_{ij}]|\boldsymbol{x}_{ij}, r_i) \\ &= P(U_{ij} = u_{ij}, T_{ij} \in (0, u_{ij}]|\boldsymbol{x}_{ij}, r_i) \\ &= P(U_{ij} = u_{ij}|\boldsymbol{x}_{ij}, r_i)P(T_{ij} \in (0, u_{ij}]|\boldsymbol{x}_{ij}, r_i); \end{aligned} \tag{4}$$

when $\delta_{2ij} = 1$, that is, T_{ij} is interval-censored in $(u_{ij}, v_{ij}]$, but W_{ij} is exactly observed as $(v_{ij} - u_{ij})$,

$$\begin{aligned} L_{ij} &= P(U_{ij} = u_{ij}, T_{ij} \in (u_{ij}, v_{ij}], W_{ij} = v_{ij} - u_{ij}|\boldsymbol{x}_{ij}, r_i) \\ &= P(U_{ij} = u_{ij}|\boldsymbol{x}_{ij}, r_i)P(T_{ij} \in (u_{ij}, v_{ij}]|\boldsymbol{x}_{ij}, r_i)P(W_{ij} = v_{ij} - u_{ij}|\boldsymbol{x}_{ij}, r_i); \end{aligned} \tag{5}$$

when $\delta_{3ij} = 1$ and $\psi_{ij} = 0$, that is, both T_{ij} and W_{ij} are right-censored at u_{ij} and 0, respectively,

$$\begin{aligned} L_{ij} &= P(U_{ij} = u_{ij}, W_{ij} > 0, T_{ij} \in (u_{ij}, \infty)|\boldsymbol{x}_{ij}, r_i) \\ &= P(U_{ij} = u_{ij}, T_{ij} \in (u_{ij}, \infty)|\boldsymbol{x}_{ij}, r_i) \\ &= P(U_{ij} = u_{ij}|\boldsymbol{x}_{ij}, r_i)P(T_{ij} \in (u_{ij}, \infty)|\boldsymbol{x}_{ij}, r_i); \end{aligned} \tag{6}$$