Computational Statistics and Data Analysis 62 (2013) 171–180

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Estimation of the accelerated failure time frailty model under generalized gamma frailty

Pengcheng Chen [a], Jiajia Zhang [b], Riquan Zhang [a,c,*]

[a] School of Finance and Statistics, East China Normal University, Shanghai, 200241, PR China

[b] Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA

[c] Department of Mathematics, Shanxi Datong University, Datong, 037009, PR China

ARTICLE INFO

Article history:
Received 12 August 2011
Received in revised form 18 July 2012
Accepted 17 January 2013
Available online 23 January 2013

Keywords:
Accelerated failure time model
Frailty
Generalized gamma distribution
EM-like algorithm

ABSTRACT

The frailty model is one of the most popular models used to analyze clustered failure time data, where the frailty term is used to assess an association within each cluster. The frailty model based on the semiparametric accelerated failure time model attracts less attention than the one based on the proportional hazards model due to its computational difficulties. In this paper, we relax the frailty distribution to the generalized gamma distribution, which can accommodate most of the popular frailty assumptions. The estimation procedure is based on the EM-like algorithm by employing the MCMC algorithm in the E-step and the profile likelihood estimation method in the M-step. We conduct an extensive simulation study and find that there is a significant gain in the proposed method with respect to the estimation of the frailty variance with a slight loss of accuracy in the parameter estimates. For illustration, we apply the proposed model and method to a data set of sublingual nitroglycerin and oral isosorbide dinitrate on angina pectoris of coronary heart disease patients.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Clustered failure time data is very commonly seen in medical research and epidemiological cancer studies. For example, the dependence may be generated from the shared genetic structure or environment of subjects, such as family history of breast cancer patients. It is worth pointing out that applying the general proportional hazards (PH) model or the accelerated failure time (AFT) model directly to a clustered data set without considering the possible correlations in each cluster may lead to incorrect conclusions. Therefore, how to properly model the correlations among each cluster is a very important and interesting issue in analyzing survival data with clusters.

There exist many discussions about analyzing clustered failure times (details can be found in Hougaard, 2000). The frailty model (Vaupel et al., 1979) is one of the most important models, which model the correlation in each cluster by a random effect based on the PH model. Since then, several extensions have been discussed in the framework of the PH model. Hougaard (1986) discussed the frailty model based on the positive stable distribution. McGilchrist and Aisbett (1991) investigated the lognormal frailty model. Klein (1992) developed the EM algorithm for the gamma frailty model, and Yu (2006) improved the estimation method and applied it to handle large data sets. Therneau et al. (2003) proposed a penalized estimation method for frailty models by utilizing the partial likelihood function, and their method is available in Splus and R. Balakrishnan and Peng (2006) developed the estimation method for the generalized gamma frailty model based on the MCMC approach.

* Corresponding author at: School of Finance and Statistics, East China Normal University, Shanghai, 200241, PR China.
E-mail address: zhangriquan@163.com (R. Zhang).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.01.016

However, few studies have focused on the semiparametric AFT frailty model due to its computational difficulties, which arise directly from the semiparametric AFT model itself. The marginal approach (Jin et al., 2006a,b) is one of the most useful tools in analyzing clustered data in the AFT model. Pan (2001) proposed the AFT frailty model by assuming a frailty structure on the error term, which is called the AFT gamma frailty model. His estimation procedure was improved by Zhang and Peng (2007) through the M-estimator, and Xu and Zhang (2010) using rank estimation methods.

Whether the gamma distribution is flexible enough to capture associations in each cluster is questionable and may limit its usage in practice. The main reason for using the gamma distribution as the frailty distribution is its convenience in computation, since it has a closed form for the conditional distribution. In this paper, we propose to relax the gamma assumption by assuming the generalized gamma distribution. The advantage of the generalized gamma distribution is that it can include the gamma distribution, Weibull distribution and positive stable distribution as its special cases. The challenge of using the generalized gamma distribution comes directly from the complex expression of the complete likelihood function, unknown frailty term and unknown baseline distributions. We will address these issues by utilizing the MCMC approach in the E-step and the profile likelihood approach (Zeng and Lin, 2007) in the M-step.

The organization of this paper is as follows. In Section 2, we give the description of the AFT generalized gamma frailty model. Section 3 outlines the estimation procedure; we list some derivation details in the Appendix. The simulation study is conducted in Section 4. Then we apply the proposed model and method to a data set of coronary heart disease patients in Section 5. Some discussions and conclusions are given in Section 6.

2. Accelerated failure time model with generalized gamma frailty

Let us define T^*_{ij} to be the true failure time for the jth subject in the ith group, and X_{ij} to be a $p \times 1$ vector of covariates associated with this subject, $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. The observed time T_{ij} is the smaller value of the true failure time T^*_{ij} and censor time C_{ij}, i.e., $T_{ij} = \min(T^*_{ij}, C_{ij})$. The censoring indicator $\delta_{ij} = 1$ if $T_{ij} = T^*_{ij}$ and 0 otherwise. We assume that the censoring is noninformative. For the sake of simplicity, O is used to denote the observed data set $(T_{ij}, \delta_{ij}, X_{ij})$.

AFT frailty model

The AFT frailty model can be specified as

$$T_{ij} = \exp(\beta' X_{ij}) V_{ij} \tag{1}$$

$$h(V_{ij}|W_i = w_i) = w_i h_0(V_{ij}) \tag{2}$$

where β is a p vector of unknown parameters of interest, $h(\cdot)$ is the hazard function, and $h_0(\cdot)$ is the baseline hazard function. Basically, model (1) is the AFT model, where $\log T$ is the linear regression of the covariate; and model (2) models the association of V_{ij} among clusters by the PH frailty model. Specifically, the frailty, W_i, is the random effect shared by subjects in the ith cluster. Given W_i, the conditional hazard function, $h(V_{ij}|W_i = w_i)$ follows the PH frailty model. Furthermore, we assume that $W_i, i = 1 \ldots, n$ are independent identically-distributed ($i.i.d.$) and independent of X_{ij}. Given the frailty random variable, the hazard function and the survival function can be expressed as

$$h(t_{ij}|X_{ij}, w_i) = w_i \exp(-\beta' X_{ij}) h_0(t_{ij} \exp(-\beta' X_{ij}))$$
$$S(t_{ij}|X_{ij}, w_i) = S_0(t_{ij} \exp(-\beta' X_{ij}))^{w_i}.$$

Note that we have the standard AFT model when $W_i \equiv 1$. Let $f(w)$ denote the distribution of the frailty random effects. The complete likelihood function can be written as

$$\prod_{i=1}^{n} \prod_{j=1}^{n_i} h(t_{ij}|X_{ij}, w_i)^{\delta_{ij}} S(t_{ij}|X_{ij}, w_i) f(w_i).$$

The estimates of parameters can be obtained by maximum likelihood if the baseline hazard function and frailty distribution are known. However, the baseline hazard function is usually unknown and the estimates should be obtained through the EM algorithm.

Generalized gamma frailty distribution

Usually, we assume the frailty distribution follows the gamma distribution, Weibull distribution, positive stable distribution or lognormal distribution. For the purpose of a model's identifiability (Peng and Zhang, 2008), the mean of the frailty distribution is assumed to be one. In order to enable the model's flexibility, in this paper, we assume the generalized gamma distribution as the frailty distribution, which can be specified as

$$g(w; q, \sigma, \lambda) = \begin{cases} |q|(q^{-2})^{q^{-2}} (\lambda w)^{q^{-2}(q/\sigma)} \exp[-q^{-2}(\lambda w)^{q/\sigma}]/[\Gamma(q^{-2})\sigma w] & q \neq 0 \\ \left(\sqrt{2\pi}\, \sigma w\right)^{-1} \exp\{-[\log(\lambda w)]^2/(2\sigma^2)\} & q = 0 \end{cases} \tag{3}$$

where $\sigma (> 0)$ is the shape parameter, $\lambda (> 0)$ is the scale parameter, and $-\infty < q < \infty$.

Download English Version:

https://daneshyari.com/en/article/415702

Download Persian Version:

https://daneshyari.com/article/415702

Daneshyari.com