NIH Public Access
Author Manuscript
Biometrika. Author manuscript; available in PMC 2014 January 15.

Published in final edited form as:
Biometrika. 2013 ; 100(3): 741–755. doi:10.1093/biomet/ast012.

Kernel Smoothed Profile Likelihood Estimation in the Accelerated Failure Time Frailty Model for Clustered Survival Data

Bo Liu,
Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A

Wenbin Lu, and
Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695, U.S.A

Jiajia Zhang
Department of Epidemiology and Biostatistics, University of South Carolina, 800 Sumter Street, Columbia, South Carolina 29208, U.S.A

Bo Liu: bliu4@ncsu.edu; Wenbin Lu: lu@stat.ncsu.edu; Jiajia Zhang: jzhang@mailbox.sc.edu

Summary

Clustered survival data frequently arise in biomedical applications, where event times of interest are clustered into groups such as families. In this article we consider an accelerated failure time frailty model for clustered survival data and develop nonparametric maximum likelihood estimation for it via a kernel smoother aided EM algorithm. We show that the proposed estimator for the regression coefficients is consistent, asymptotically normal and semiparametric efficient when the kernel bandwidth is properly chosen. An EM-aided numerical differentiation method is derived for estimating its variance. Simulation studies evaluate the finite sample performance of the estimator, and it is applied to the Diabetic Retinopathy data set.

Keywords

Accelerated failure time model; Clustered survival data; EM algorithm; Kernel smoothing; Profile likelihood estimation

1. Introduction

Clustered survival data are a common type of multivariate survival data, often encountered in fields such as medicine, economics and epidemiology. Because multivariate survival models are important tools for analyzing clustered survival data, they have attracted considerable attention. There are two main approaches: marginal modelling and joint modelling via random effects. The first approach models the marginal distribution of correlated failure times without specifying the correlation structure. For example, Wei et al. (1989) proposed marginal regression analysis based on the proportional hazards model (Cox, 1972) for multivariate failure time data. A review of marginal approaches based on

© 2008 Biometrika Trust

Supplementary Material
Supplementary material available at *Biometrika* online includes additional simulation study results and technical derivations.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the proportional hazards model can be found in Lin (1994). In some applications, such as family studies, the within-cluster association is also important to investigators. Joint modelling uses random effects to describe the association among failure times within clusters. In addition, by appropriately taking into account the correlation structure, joint modelling can have better estimation efficiency than the marginal approach. Clayton & Cuzick (1985) introduced cluster-specific random effects or frailties to the proportional hazards model, which assumes that subjects within the same cluster can be considered independent conditional on the frailty. The multiplicative proportional hazards frailty model has been widely studied (Hougaard, 1987; Oakes, 1989; Nielsen et al., 1992) and various frailty distributions have been used to describe the within cluster correlation. The large sample properties of the associated nonparametric maximum likelihood estimators have been investigated by Murphy (1994, 1995) and Parner (1998).

The accelerated failure time model (Kalbfleisch & Prentice, 2002) is a useful alternative to the proportional hazards model. Many methods have been developed for parameter estimation in the accelerated failure time model for univariate survival data (Buckley & James, 1979; Tsiatis, 1990; Ying, 1993; Jin et al., 2003; Zeng & Lin, 2007). Recently, this model has been extended to clustered survival data. For example, Jin et al. (2006a,b) considered the marginal accelerated failure time model for clustered survival data: the former extended the Buckley-James estimation method; the latter extended the weighted log-rank estimation method. To improve the efficiency of the marginal approach, Li & Yin (2009) proposed a generalized moments estimation method, incorporating a posited correlation matrix into the rank-based estimating equations and minimizing a quadratic inference function. In addition, Johnson & Strawderman (2009) applied the induced smoothing technique to the weighted log-rank estimators for clustered survival data, which facilitates the resulting estimation and inference procedures. To characterize the correlation structure of failure times within clusters, Pan (2001) proposed to use frailties in the accelerated failure time model and developed an EM-like algorithm to estimate the coefficients in the accelerated failure time frailty model. Based on Pan's method, Zhang & Peng (2007) and Xu & Zhang (2010) developed more stable estimation procedures using M-estimation and rank-based estimation, respectively. More recently, Johnson & Strawderman (2012) introduced smoothing into the EM-like algorithm to facilitate parameter estimation. However, none of the above estimators are semiparametric efficient because the considered EM-like algorithms do not maximize the likelihood function. Moreover, the asymptotic properties of these estimators have not been studied. In this article, we develop a nonparametric maximum likelihood estimation method for the accelerated failure time frailty model.

2. The accelerated failure time frailty model

Let T_{ij} be the failure time, C_{ij} be the censoring time and X_{ij} be the p-dimensional vector of baseline covariates for the jth individual in the ith cluster, for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Here n is the total number of clusters and m_i is the size of the ith cluster. The observed data are $O = \{(\tilde{T}_{ij}, \delta_{ij}, X_{ij}) : i = 1, \ldots, n; j = 1, \ldots, m_i\}$, where $\tilde{T}_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I(T_{ij} \leq C_{ij})$.

The marginal accelerated failure time model is

$$\log T_{ij} = -\beta' X_{ij} + \varepsilon_{ij}, \quad (1)$$

where β is the p-dimensional vector of regression coefficients, and the error terms, $(\varepsilon_{i1}, \ldots, \varepsilon_{im_i})$, are independent across clusters and independent of $(X_{i1}, \ldots, X_{im_i})$. It is assumed that all ε_{ij} have a common unknown marginal distribution, and ε_{ij} and ε_{ik} may be correlated for $j \neq$

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

k. We assume that T_{ij} and C_{ij} are independent conditional on X_{ij}, and m_i is small compared to n and is noninformative, i.e., independent of T_{ij}, C_{ij} and X_{ij}.

To describe the dependence between clustered survival times, Pan (2001) proposed to consider the accelerated failure time frailty model. Specifically, given a positive latent variable α_i of mean 1 and variance σ^2, it is assumed that the hazard function of $e^{\varepsilon_{ij}}$ is

$$\lambda_{ij}(t)=\alpha_i\,\lambda(t)\quad(i=1,\ldots,n;\,j=1,\ldots,m_i),\qquad (2)$$

where $\lambda(\cdot)$ is an unspecified baseline hazard function. In addition, $\varepsilon_{i1}, \ldots, \varepsilon_{im_i}$ are assumed independent conditional on α_i and the magnitude of dependence among the ε_{ij} is characterized by the value of σ^2. There are many choices for the frailty distribution, e.g., the gamma distribution (Clayton, 1978), the positive stable distribution (Hougaard, 1986), the compound Poisson distribution (Aalen, 1992) and the log-normal distribution (McGilchrist & Aisbett, 1991).

3. Nonparametric maximum likelihood estimator

Let $f_\alpha(\cdot;\theta)$ denote the density of the latent variable α_i, where θ is an unknown finite dimensional vector of parameters. The log-likelihood function for the complete data, $\{(\tilde{T}_{ij}, \delta_{ij}, X_{ij}, \alpha_i) : i = 1, \ldots, n; j = 1, \ldots, m_i\}$, can be written as

$$l_n^c(\beta,\Lambda,\theta)=l_{n,1}^c(\theta)+l_{n,2}^c(\beta,\Lambda),$$

where $\Lambda(t)=\int_0^t\lambda(s)\,ds$, $R_{ij}(\beta) = \log(\tilde{T}_{ij}) + \beta' X_{ij}$,

$$l_{n,1}^c(\theta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}\log\alpha_i+\frac{1}{n}\sum_{i=1}^{n}\log f_\alpha(\alpha_i;\theta),\qquad (3)$$

$$l_{n,2}^c(\beta,\Lambda)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\delta_{ij}\left\{\beta' X_{ij}+\log\lambda\left(e^{R_{ij}(\beta)}\right)\right\}-\alpha_i\Lambda\left(e^{R_{ij}(\beta)}\right)\right].\qquad (4)$$

We use an EM algorithm to obtain the nonparametric maximum likelihood estimator. Let $\hat{\Omega}^{[k]} = (\hat{\beta}^{[k]}, \hat{\Lambda}^{[k]}, \hat{\theta}^{[k]})$ denote the parameter estimates at step k. In the expectation step, we obtain the conditional density of α_i given the observed data O and current parameter estimates $\hat{\Omega}^{[k]}$,

$$f_\alpha(\alpha_i|O,\hat{\Omega}^{[k]})=\frac{f_\alpha(\alpha_i;\hat{\theta}^{[k]})\alpha_i^{\sum_{j=1}^{m_i}\delta_{ij}}\exp\left\{-\alpha_i\sum_{j=1}^{m_i}\hat{\Lambda}^{[k]}\left(e^{R_{ij}(\hat{\beta}^{[k]})}\right)\right\}}{\int_0^{+\infty}f_\alpha(\alpha_i;\hat{\theta}^{[k]})\alpha_i^{\sum_{j=1}^{m_i}\delta_{ij}}\exp\left\{-\alpha_i\sum_{j=1}^{m_i}\hat{\Lambda}^{[k]}\left(e^{R_{ij}(\hat{\beta}^{[k]})}\right)\right\}d\alpha_i}.\qquad (5)$$

The conditional expectations $E(\alpha_i|O, \hat{\Omega}^{[k]})$, $E(\log\alpha_i|O, \hat{\Omega}^{[k]})$ and $E\{\log f_\alpha(\alpha_i;\theta)|O, \hat{\Omega}^{[k]}\}$ can be calculated as the integrals of the corresponding terms with respect to the conditional density $f_\alpha(\alpha_i|O, \hat{\Omega}^{[k]})$. For example, when the frailty has a gamma density $f_\alpha(x;\theta) = x^{\theta-1} e^{-\theta x}\theta^\theta/\Gamma(\theta)$, where $x > 0$, $\theta > 0$ and $\Gamma(\theta)=\int_0^{+\infty}t^{\theta-1}e^{-t}dt$, we have

$$\hat{\alpha}_i^{[k]} \equiv E(\alpha_i|O,\hat{\Omega}^{[k]})=(D_i+\hat{\theta}^{[k]})/\{\hat{\theta}^{[k]}+\sum_{j=1}^{m_i}\hat{\Lambda}^{[k]}(e^{R_{ij}(\hat{\beta}^{[k]})})\},$$
$$E_{2,i}^{[k]} \equiv E(\log\alpha_i|O,\hat{\Omega}^{[k]})=\Psi(D_i+\hat{\theta}^{[k]})-\log\{\hat{\theta}^{[k]}+\sum_{j=1}^{m_i}\hat{\Lambda}^{[k]}(e^{R_{ij}(\hat{\beta}^{[k]})})\},$$
$$E_{3,i}^{[k]} \equiv E\{\log f_\alpha(\alpha_i;\theta)|O,\hat{\Omega}^{[k]}\}=(\theta-1)E_{2,i}^{[k]}-\theta\hat{\alpha}_i^{[k]}+\theta\log\theta-\log\Gamma(\theta),$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. For general frailty distributions, such as the log-normal distribution, these conditional expectations may not have closed analytical forms. In such cases we use gaussian quadrature. Therefore, the conditional expectations of (3) and (4) given O and $\hat{\Omega}^{[k]}$ are

$$E\{l_{n,1}^c(\theta)|O,\hat{\Omega}^{[k]}\}=\frac{1}{n}\sum_{i=1}^{n}E_{2,i}^{[k]}D_i+\frac{1}{n}\sum_{i=1}^{n}E_{3,i}^{[k]}, \quad (6)$$

$$E\{l_{n,2}^c(\beta,\Lambda)|O,\hat{\Omega}^{[k]}\}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\delta_{ij}\{\beta' X_{ij}+\log\lambda(e^{R_{ij}(\beta)})\}-\hat{\alpha}_i^{[k]}\Lambda(e^{R_{ij}(\beta)})\right], \quad (7)$$

where $D_i=\sum_{j=1}^{m_i}\delta_{ij}$ is the total number of observed events in cluster i.

In the maximization step, equation (6) can be easily maximized using standard gradient-based optimization algorithms. Let $\hat{\theta}^{[k+1]}$ denote the maximizer of (6). The conditional log-likelihood given in (7) cannot be directly maximized over β and Λ. We adopt a kernel smoothing technique similar to that used by Zeng & Lin (2007). Specifically, consider the piecewise constant hazard function $\tilde{\lambda}(t)=\sum_{l=1}^{J_n}c_l I(t_{l-1}\le t<t_l)$ on $[0, M]$, where $0 = t_0 < t_1 < \cdots < t_{J_n} = M$ are equally spaced, and M is an upper bound for all $e^{R_{ij}(\beta)}$. Accordingly, the cumulative hazard function is $\tilde{\Lambda}(t)=\sum_{l=1}^{J_n}c_l(t-t_l)I(t_{l-1}\le t<t_l)+(M/J_n)\sum_{l=1}^{J_n}c_l I(t\ge t_l)$. Using these expression for $\tilde{\lambda}(\cdot)$ and $\tilde{\Lambda}(\cdot)$ in (7) and maximizing (7) with respect to c_l ($l = 1, \ldots, J_n$), for fixed β, the following maximizers are obtained:

$$\hat{c}_l^{[k]}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)}{\sum_{i=1}^{n}\hat{\alpha}_i^{[k]}\sum_{j=1}^{m_i}\{(e^{R_{ij}(\beta)}-t_{l-1})I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)+(M/J_n)I(e^{R_{ij}(\beta)}\ge t_l)\}}.$$

The profile likelihood function for β constructed using the above expression for $\hat{c}_l^{[k]}$ is

$$\begin{aligned} l_{n,2}^{p,k}(\beta)= & \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}\beta' X_{ij}+\sum_{l=1}^{J_n}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)\right\}\right. \\ & \left.\times\log\left\{\frac{J_n}{nM}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)\right\}\right]-\sum_{l=1}^{J_n}\left(\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)\right\}\right. \\ & \left.\times\log\left[\frac{J_n}{nM}\sum_{i=1}^{n}\hat{\alpha}_i^{[k]}\sum_{j=1}^{m_i}\left\{(e^{R_{ij}(\beta)}-t_{l-1})I(t_{l-1}\le e^{R_{ij}(\beta)}<t_l)+\frac{M}{J_n}I(e^{R_{ij}(\beta)}\ge t_l)\right\}\right]\right). \end{aligned}$$

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

Following derivations similar to those given in Zeng & Lin (2007), $l_{n,2}^{p,k}(\beta)$ converges uniformly in β to a limiting function as $n \to \infty$, $J_n \to \infty$ and $J_n/n \to 0$, and the limiting function can be approximated by the smooth function

$$l_{n,2}^{s,k}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}\left[\beta' X_{ij}+\log\{\widehat{\lambda}^{[k+1]}(e^{R_{ij}(\beta)};\beta)\}\right], \quad (8)$$

where

$$\widehat{\lambda}^{[k+1]}(t;\beta)=\frac{1}{th_n}\frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}K[\{R_{ij}(\beta)-\log(t)\}/h_n]}{\sum_{i=1}^{n}\widehat{\alpha}_i^{[k]}\sum_{j=1}^{m_i}\int_{-\infty}^{\{R_{ij}(\beta)-\log(t)\}/h_n}K(u)du}, \quad t>0. \quad (9)$$

Let $\hat{\beta}^{[k+1]}$ denote the maximizer of $l_{n,2}^{s,k}(\beta)$. Given $\hat{\beta}^{[k+1]}$, a smooth estimator of $\lambda(t)$ can be obtained as $\hat{\lambda}^{[k+1]}(t;\hat{\beta}^{[k+1]})$ and $\widehat{\Lambda}^{[k+1]}(t) \equiv \int_0^t \widehat{\lambda}^{[k+1]}(s;\widehat{\beta}^{[k+1]})ds$.

From an initial estimator $\hat{\Omega}^{[0]}$, the E-step and M-step are repeated until convergence. The estimators of β, $\Lambda(\cdot)$ and θ at convergence are denoted by $\hat{\beta}_n$, $\hat{\Lambda}_n(\cdot)$ and $\hat{\theta}_n$, respectively. In our implementation, the starting value of β was taken to be the maximum smoothed profile likelihood estimator of Zeng & Lin (2007) assuming working independence among correlated failure times. An initial estimator for λ was obtained by setting $k = -1$ and $\widehat{\alpha}_i^{[-1]} \equiv 1$ in (9). Based on our numerical experience, the convergence of the proposed EM algorithm is not sensitive to the choice of the starting value for θ. For convenience, we chose $\hat{\theta}^{[0]} = 1$ for all scenarios considered in our numerical studies. Define $\widehat{\gamma}_n=(\widehat{\beta}_n', \widehat{\theta}_n')'$ and $\gamma_0=\left(\beta_0', \theta_0'\right)'$, the true value of γ. Let Λ_0 and λ_0 denote the true values of Λ and λ, respectively. Next, we establish the asymptotic properties for estimators $\hat{\gamma}_n$ and $\hat{\Lambda}_n$.

Theorem 1

Assume that the regularity conditions (C1)–(C8) given in the Appendix *hold. As* $n \to \infty$, $nh_n^2 \to \infty$ *and* $nh_n^4 \to 0$*: (i)* $\sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0$ *and* $\hat{\gamma}_n \to \gamma_0$ *almost surely; and (ii)* $n^{1/2}(\hat{\gamma}_n - \gamma_0)$ *converges in distribution to a mean-zero normal random vector with a covariance matrix that achieves the semiparametric efficiency bound* I^{-1}.

The proof of Theorem 1 is given in the Appendix. To estimate the variance of $\hat{\beta}_n$, we adopt the EM-aided numerical differentiation method proposed by Chen & Little (1999), which numerically computes the empirical Fisher information matrix of the profile likelihood. A similar method was used by Lu (2010) for variance estimation in the accelerated failure time model with a cure fraction. Specifically,

$$E\left\{l_{n,1}^{c}(\theta)+l_{n,2}^{c}(\beta,\Lambda)\,|O,\Omega\right\}=\frac{1}{n}\sum_{i=1}^{n}\widetilde{l}_i(\beta,\Lambda,\theta).$$

The jth component of $\hat{\beta}_n$ is perturbed by a small value, d. The pair of perturbed estimates is denoted by $\hat{\beta}_{n,j-} = (\hat{\beta}_{n,1}, \ldots, \hat{\beta}_{n,j} - d, \ldots, \hat{\beta}_{n,p})'$ and $\hat{\beta}_{n,j+} = (\hat{\beta}_{n,1}, \ldots, \hat{\beta}_{n,j} + d, \ldots, \hat{\beta}_{n,p})'$ for $j = 1, \ldots, p$. The β is fixed at $\hat{\beta}_{n,j-}$, and the above EM algorithm is implemented until convergence. The estimates of Λ and θ at convergence are denoted by $\hat{\Lambda}_{n,j-}$ and $\hat{\theta}_{n,j-}$, respectively. The estimates $\hat{\Lambda}_{n,j+}$ and $\hat{\theta}_{n,j+}$ can be similarly obtained. For $i = 1, \ldots, n$ and $j = 1, \ldots, p$, define

NIH-PA Author Manuscript

$$\tilde{S}_{ij}=\{\tilde{l}_i(\hat{\beta}_{n,j+},\hat{\Lambda}_{n,j+},\hat{\theta}_{n,j+})-\tilde{l}_i(\hat{\beta}_{n,j-},\hat{\Lambda}_{n,j-},\hat{\theta}_{n,j-})\}/(2d).$$

Let $\tilde{S}_i = (\tilde{S}_{i1}, \ldots, \tilde{S}_{ip})'$ and $\tilde{I}_n=\sum_{i=1}^n \tilde{S}_i \tilde{S}_i'$. Then, the covariance matrix of $\hat{\beta}_n$ can be estimated by $\tilde{I}_n^{-1}$.

4. Numerical Examples

4·1. Simulation studies

We generated clustered failure times from the following model

$$\log T_{ij}=X_{ij1}-X_{ij2}+\varepsilon_{ij}\ (i=1,\ldots,100; j=1,\ldots,5),$$

where X_{ij1} follows a Bernoulli distribution with a success probability of 0.5, X_{ij2} follows a uniform distribution on [-1,1] and ε_{ij} follows the frailty model (2). We considered two frailty distributions: gamma frailty with mean 1 and variance $\sigma^2 = 1/\theta$; and log-normal frailty with mean 1 and variance $\sigma^2 = e^{\theta} - 1$. Further, we considered three choices for $\lambda_0(t)$: Weibull-type, $\lambda_0(t) = at^b$; log-normal-type, $\lambda_0(t) = t^{-1} \varphi\{\log(t)\}/[1 - \Phi\{\log(t)\}]$, where $\varphi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard normal random variable; and reciprocal-type, $\lambda_0(t) = c/(1 + t)$. Here, a, b and c are positive constants. Censoring times were generated from a uniform distribution on $[0, \tau_c]$, where τ_c was chosen to yield censoring proportions of 15% and 40%. For each setting, we conducted 2000 simulation runs.

For the bandwidth parameter, h_n, of the kernel smoother, we adapted the optimal bandwidths proposed by Jones (1990) and Jones & Sheather (1991) for density estimation. Such bandwidths were also used by Zeng & Lin (2007) for smoothing the profile likelihood in the standard accelerated failure time model. Specifically, we set $h_n = \zeta \hat{\sigma}_e n^{-1/3}$, where ζ is a positive constant, and $\hat{\sigma}_e$ is the sample standard deviation of the fitted residuals, $\log \tilde{T}_{ij}+\hat{\beta}_{\rm LS}' X_{ij}$. Here $\hat{\beta}_{\rm LS}$ is the least squares estimate based on all of the data, including censored data. In our simulations, we tried a range of values for ζ and found that $0.8 \le \zeta \le 1.8$ works well in all of the scenarios. For comparison, we also included the Gehan rank estimator (Jin et al., 2006b), the induced smoothing estimator (Johnson & Strawderman, 2009) and the smoothed EM-like estimator (Johnson & Strawderman, 2012). The former two estimators are based on the marginal accelerated failure time model.

The results for the gamma and log-normal frailties are summarized in Tables 1 and 2, respectively. Because the results for the various hazard functions are similar, we present only the results for the Weibull-type and reciprocal-type. In addition, as reported in Johnson & Strawderman (2009), the Gehan rank estimator and induced smoothing estimator have very similar performances. Therefore, we exclude the results for the Gehan rank estimator. All three estimators for the regression parameters are essentially unbiased under all settings and the averages of the estimated standard errors obtained using the proposed EM-aided numerical differentiation method for the nonparametric maximum likelihood estimator are close to their standard deviations with the empirical coverage probabilities close to the nominal level. In most cases, the nonparametric maximum likelihood estimator is more efficient than the Gehan rank estimator and the induced smoothing estimator. The efficiency gain is more substantial when the variance of the frailty is large, but it decreases as the variance decreases. This result agrees with our expectation since when the variance of the frailty is large, the survival times within the same cluster are strongly correlated. Thus, the nonparametric maximum likelihood estimator is expected to be more efficient since it effectively accounts for the within-cluster correlation. The nonparametric maximum likelihood estimator is generally more efficient than the smoothed EM-like estimator in terms of the mean square error for the Weibull-type hazard function, especially when the correlation between clustered survival times is strong and the censoring proportion is low. However, for the reciprocal-type hazard function, the smoothed EM-like estimator may have better efficiency than the nonparametric maximum likelihood estimator when the correlation is weak or the censoring proportion is high. This result is attributed to the smaller biases of the smoothed EM-like estimators. Finally, the proposed nonparametric maximum likelihood estimator for the variance of the frailty is nearly unbiased. The mean estimated survival curves for $S_0(t) \equiv \exp\{-\Lambda_0(t)\}$ are given in Figures 1 – 6 of the Supplementary Material. For all the scenarios, the mean estimated survival curves are close to the true survival curves.

We also conducted a sensitivity analysis to study the performance of the nonparametric maximum likelihood estimator when the frailty distribution is misspecified. Specifically, clustered survival data were generated from the accelerated failure time frailty model with the log-normal frailty as considered previously. However, the nonparametric maximum likelihood estimator was computed based on the gamma frailty. The simulation results are given in Table 3. The nonparametric maximum likelihood estimator for the regression parameters shows very small biases that are comparable to those reported in Table 2 when the log-normal frailty distribution was correctly specified. The means of the estimated standard errors are close to the standard deviations with proper coverage probabilities. Based on the limited simulations we have conducted, the performance of the nonparametric maximum likelihood estimator for the regression parameters is relatively robust to the misspecification of the frailty distribution. However, the estimate for the variance of the frailty shows large biases when the frailty distribution is misspecified. The mean estimated survival curves for $S_0(t)$ are given in Figures 7–9 of the Supplementary Material. When the log-normal frailty is misspecified as the gamma frailty, the mean estimated survival curves slightly overestimate the true survival curves for the cases with large frailty variance, i.e., $\sigma^2 = 3.48$, while they are nearly unbiased for cases with smaller variances.

We conducted additional simulations with cluster size of 2 and $n = 200$. The simulation results are given in the Supplementary Material. The findings are similar to those reported here.

4·2. Analysis of diabetic retinopathy data

We applied our estimation methods to clustered survival data from the diabetic retinopathy study conducted by the National Eye Institute (Huster et al., 1989). The study enrolled 197 patients with proliferative diabetic retinopathy representing a 50% simple random sample of patients with high-risk. For each patient, the photocoagulation treatment was randomly assigned to one eye, while the other eye was an untreated control. The endpoint of interest is the time to severe visual loss after treatment. In addition to the treatment indicator, 1 for treated with photocoagulation and 0 for untreated, there are three prognostic factors: age at diagnosis of diabetes, type of diabetes, 1 for adult and 0 for juvenile, and risk group ranging from 6–12. A primary goal is to study the effects of treatment and risk factors on the time to severe visual loss. This data set has been previously studied. For example, Lu (2007) studied the data using a marginal bivariate accelerated failure time model based on the weighted log-rank estimation method of Jin et al. (2006b). Lu also developed a statistical test for the association between pairs of failure times after adjusting for covariates. It was found that the null hypothesis of independence was rejected with a small p-value, which implies that there

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

is a strong correlation between the pair of error terms in the bivariate accelerated failure time model. Here, we fit the accelerated failure time frailty model with all four covariates using the proposed nonparametric maximum likelihood estimation method and the Gehan rank estimation method. For our method, both the gamma and log-normal frailties were considered. The bandwidth parameters for the nonparametric maximum likelihood estimators were selected as in the simulation studies. The results are given in Table 4. The nonparametric maximum likelihood estimators have much smaller standard errors than the Gehan rank estimator, which indicates the efficiency gained by taking the correlation between error terms into account in the nonparametric maximum likelihood estimators. In addition, the nonparametric maximum likelihood estimators with the gamma and log-normal frailty distributions show very similar performances, which may imply that the analysis results are not sensitive to the choice of the frailty distribution. The estimated variance of the frailty is 0.88 under the gamma frailty and 1.16 under the log-normal frailty. Finally, all of the methods found that treatment and risk group are significantly associated with time to severe visual loss, whereas age at diagnosis of diabetes and type of diabetes are not.

5. Discussion

The proposed kernel-smoothing based nonparametric maximum likelihood estimation method can be extended to other types of multivariate survival data, such as recurrent event data. Specifically, let $N_i^*(t)$ denote the number of events observed on subject *i* by time *t*. We assume that $N_i^*(t)$ is a nonhomogeneous Poisson process and model its conditional intensity function given covariates Z_i and frailty a_i by $a_i e^{\beta' Z_i} \lambda(e^{\beta' Z_i} t)$. This model is an extension of the accelerated failure time model for counting processes considered by Lin et al. (1998) and was studied by Strawderman (2006) using an EM-like algorithm. The nonparametric maximum likelihood estimation and its associated inference for the above model require further investigation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the editor, an associate editor and two referees for very insightful comments. Lu and Zhang's research was supported by the National Cancer Institute.

Appendix

Throughout the proofs, we assume the following regularity conditions:

(C1) The hazard function $\lambda_0(\cdot)$ is positive and thrice-continuously differentiable with $\dot{\lambda}_0(0) > 0$, where $\dot{\lambda}_0(0)$ is the right derivative of $\lambda(t)$ at $t = 0$. In addition, $\Lambda_0(\tau) < \infty$.

(C2) There exists some positive constant c_0, such that $\mathrm{pr}(C_{ij} e^{\beta_0' X_{ij}} \geq \tau | X_{ij}) \geq c_0$.

(C3) The covariates X_{ij} are bounded. If there exists a constant vector a, such that $a' X_{ij} = 0$ almost surely, then $a = 0$.

(C4) The true regression parameters β_0 belong to the interior of a known compact set B, and $0 < \theta_0 < \infty$.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

(C5) The kernel function $K(\cdot)$ is thrice-continuously differentiable. In addition, $K^{(r)}(\cdot)$ (r = 0, 1, 2, 3), have bounded variations in R, where $K^{(r)}(\cdot)$ is the rth derivative of $K(\cdot)$.

(C6) The cluster size m_i is completely random. In addition, there exists a positive integer m_0, such that $1 \le m_i \le m_0$ and $\mathrm{pr}(m_i \ge 2) > 0$.

(C7) For any $1 \le k \le m_0$, $\sup_\theta \int_0^\infty u^k |f_\alpha^{(m)}(u;\theta)| du < \infty$ for m = 0, 1, 2, where $f_\alpha^{(m)}(u;\theta)$ is the mth derivative of $f_\alpha(u;\theta)$ with respect to θ. In addition, $\int_0^\infty u^k f_\alpha^{(1)}(u;\theta_0) du \ne 0$ for some k.

(C8) The information matrix I is finite and positive definite.

Conditions (C1)–(C5) are similar to those used in Zeng & Lin (2007). Condition (C7) is assumed to establish the consistency of the proposed estimators, which is satisfied by many commonly used frailty distributions, e.g., the gamma and log-normal distributions. Condition (C8) is assumed to establish the asymptotic normality of the estimators.

Proof of Theorem 1

To establish the consistency of the estimators, we introduce the following quantity

$$\tilde{\Lambda}_n(t) \equiv \int_0^t \frac{(nh_n s)^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} K[\{R_{ij}(\beta_0) - \log s\}/h_n]}{n^{-1}\sum_{i=1}^n \alpha_i^* \sum_{j=1}^{m_i} \int_{-\infty}^{\{R_{ij}(\beta_0)-\log s\}/h_n} K(u)du} ds,$$

where $\alpha_i^* = E(\alpha_i | \delta_{ij}, \tilde{T}_{ij}, X_{ij})$. Following Lemma 2.4 of Schuster (1969), we can show that as $n \to \infty$,

$$\sup_{s\in[0,\tau]} \left| \frac{1}{nh_n s} \sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} K\left\{\frac{R_{ij}(\beta_0)-\log s}{h_n}\right\} - \frac{dE\{\sum_{j=1}^{m_i} I(\delta_{ij}=1, e^{R_{ij}(\beta_0)} \le s)\}}{ds} \right| \to 0,$$
$$\sup_{s\in[0,\tau]} \left| \frac{1}{n} \sum_{i=1}^n \alpha_i^* \sum_{j=1}^{m_i} \int_{-\infty}^{\{R_{ij}(\beta_0)-\log s\}/h_n} K(u)du - E\left\{\alpha_i^* \sum_{j=1}^{m_i} I(e^{R_{ij}(\beta_0)} \ge s)\right\} \right| \to 0.$$

almost surely. Moreover,

$$\frac{dE\{\sum_{j=1}^{m_i} I(\delta_{ij}=1, e^{R_{ij}(\beta_0)} \le s)\}}{ds} = \lambda_0(s) E\left\{\alpha_i e^{-\alpha_i \Lambda_0(s)} \sum_{j=1}^{m_i} S_c(se^{-\beta_0' X_{ij}} | X_{ij})\right\},$$
$$E\left\{\alpha_i^* \sum_{j=1}^{m_i} I(e^{R_{ij}(\beta_0)} \ge s)\right\} = E\left\{\alpha_i \sum_{j=1}^{m_i} I(e^{R_{ij}(\beta_0)} \ge s)\right\} = E\left\{\alpha_i e^{-\alpha_i \Lambda_0(s)} \sum_{j=1}^{m_i} S_c(se^{-\beta_0' X_{ij}} | X_{ij})\right\},$$

where $S_c(t \mid x) = \mathrm{pr}(C_{ij} \ge t \mid X_{ij} = x)$. Therefore,

$$\sup_{s\in[0,\tau]} \left| \frac{(nh_n s)^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \delta_{ij} K[\{R_{ij}(\beta_0) - \log s\}/h_n]}{n^{-1}\sum_{i=1}^n \alpha_i^* \sum_{j=1}^{m_i} \int_{-\infty}^{\{R_{ij}(\beta_0)-\log s\}/h_n} K(u)du} - \lambda_0(s) \right| \to 0,$$

almost surely, which implies $\tilde{\Lambda}_n(t) \to \Lambda_0(t)$ almost surely for any $t \in [0, \tau]$. This pointwise consistency can be strengthened to uniform consistency on $[0, \tau]$ due to the monotonicity and boundedness of $\tilde{\Lambda}_n(t)$ and $\Lambda_0(t)$.

By Helly's theorem, there exists a convergent subsequence $(\hat{\beta}_{n_k}, \hat{\theta}_{n_k}, \hat{\Lambda}_{n_k})$ such that $(\hat{\beta}_{n_k}, \hat{\theta}_{n_k}, \hat{\Lambda}_{n_k}) \to (\beta^*, \theta^*, \Lambda^*)$ almost surely, where Λ^* is a monotonically increasing function. Define the observed log-likelihood function

$$l_n^o(\beta,\theta,\Lambda)=\sum_{i=1}^{n}\log\int_0^\infty \left[\prod_{j=1}^{m_i}\{\alpha_i\,\lambda(e^{R_{ij}\,(\beta)})e^{\beta'\,X_{ij}}\}^{\delta_{ij}}\ e^{-\alpha_i\,\Lambda\,(e^{R_{ij}\,(\beta)})}\right] f_\alpha(\alpha_i;\theta)d\alpha_i.$$

We have

$$0 \le n_k^{-1}\, l_{n_k}^o\,(\hat{\beta}_{n_k}, \hat{\theta}_{n_k}, \hat{\Lambda}_{n_k}) - n_k^{-1}\, l_{n_k}^o(\beta_0, \theta_0, \tilde{\Lambda}_{n_k}).$$

Letting $k \to \infty$ leads to

$$0 \le E\left(\log\frac{\int_0^\infty\left[\prod_{j=1}^{m_i}\{\alpha_i\,\lambda^*\,(e^{R_{ij}\,(\beta^*)})e^{\beta^{*\prime}\,X_{ij}}\}^{\delta_{ij}}\ e^{-\alpha_i\,\Lambda^*\,(e^{R_{ij}\,(\beta^*)})}\right]\ f_\alpha(\alpha_i;\theta^*)\,d\alpha_i}{\int_0^\infty\left[\prod_{j=1}^{m_i}\{\alpha_i\,\lambda_0\,(e^{R_{ij}\,(\beta_0)})e^{\beta_0'\,X_{ij}}\}^{\delta_{ij}} e^{-\alpha_i\,\Lambda_0(e^{R_{ij}\,(\beta_0)})}\right]\ f_\alpha(\alpha_i;\theta_0)\,d\alpha_i}\right),$$

where λ^* is the derivative of Λ^*. Due to the nonnegativity of the Kullback–Leibler information,

$$\begin{aligned}&\int_0^\infty\left[\prod_{j=1}^{m_i}\{\alpha_i\,\lambda^*(e^{R_{ij}\,(\beta^*)})e^{\beta^{*\prime}\,X_{ij}}\}^{\delta_{ij}} e^{-\alpha_i\,\Lambda^*(e^{R_{ij}\,(\beta^*)})}\right]\ f_\alpha(\alpha_i;\theta^*)d\alpha_i\\ &=\int_0^\infty\left[\prod_{j=1}^{m_i}\{\alpha_i\,\lambda_0\,(e^{R_{ij}\,(\beta_0)})e^{\beta_0'\,X_{ij}}\}^{\delta_{ij}} e^{-\alpha_i\,\Lambda_0(e^{R_{ij}\,(\beta_0)})}\right]\ f_\alpha(\alpha_i;\theta_0)d\alpha_i.\end{aligned}$$

Set $\delta_{i1} = 1$ and $\tilde{T}_{i1} = 0$. For $j = 2, \ldots, m_i$, if $\delta_{ij} = 0$, set $\tilde{T}_{ij} = \tau$; if $\delta_{ij} = 1$, integrate $\tilde{T}_{ij}$ from 0 to τ. We have

$$\begin{aligned}&\int_0^\infty\alpha_i\,\lambda^*(0)e^{\beta^{*\prime}\,X_{ij}}\prod_{j=2}^{m_i}\{1-e^{-\alpha_i\,\Lambda^*(\tau e^{\beta^{*\prime}\,X_{ij}})}\}^{\delta_{ij}}\{e^{-\alpha_i\,\Lambda^*(\tau e^{\beta^{*\prime}\,X_{ij}})}\}^{1-\delta_{ij}}\ f_\alpha(\alpha_i;\theta^*)d\alpha_i\\ &=\int_0^\infty\alpha_i\,\lambda_0(0)e^{\beta_0'\,X_{ij}}\prod_{j=2}^{m_i}\{1-e^{-\alpha_i\,\Lambda_0(\tau e^{\beta_0'\,X_{ij}})}\}^{\delta_{ij}}\{e^{-\alpha_i\,\Lambda_0(\tau e^{\beta_0'\,X_{ij}})}\}^{1-\delta_{ij}}\ f_\alpha(\alpha_i;\theta_0)d\alpha_i.\end{aligned}$$

The two sides of the above equation are summed over all possible combinations of δ_{ij} ($j = 2, \ldots, m_i$) to obtain

$$\lambda^*\,(0)\,e^{\beta^{*\prime}\,X_{ij}}=\lambda_0\,(0)e^{\beta_0'X_{ij}}$$

since $E(\alpha_i) = 1$. Therefore, $(\beta^* - \beta_0)'\,X_{ij} = \log\{\lambda_0(0)/\lambda^*(0)\}$. By (C3), $\beta^* = \beta_0$. It follows that $\lambda_0(0) = \lambda^*(0)$. In addition, following similar steps, we can obtain

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\int_0^\infty \alpha_i^k\, f_\alpha\,(\alpha_i;\theta_0)\, d\alpha_i = \int_0^\infty \alpha_i^k\, f_\alpha\,(\alpha_i;\theta^*) d\alpha_i,$$

for any $1 \le k \le m_i$, which leads to $\theta_0 = \theta^*$. Finally, set $\delta_{i1} = 1$ and integrate $\tilde{T}_{i1}$ from 0 to t. For $j = 2, \ldots, m_i$, if $\delta_{ij} = 0$, set $\tilde{T}_{ij} = \tau$; if $\delta_{ij} = 1$, integrate $\tilde{T}_{ij}$ from 0 to τ. The two sides of the equation are summed over all possible combinations of δ_{ij} ($j = 2, \ldots, m_i$) to obtain

$$\int_0^\infty \{1 - e^{-\alpha_i\, \Lambda^*\,(te^{\beta^{*'}\, X_{i1}})}\}\, f_\alpha\,(\alpha_i;\theta^*)\, d\alpha_i = \int_0^\infty \{1 - e^{-\alpha_i\, \Lambda_0\,(te^{\beta_0^{'}\, X_{i1}})}\} f_\alpha\,(\alpha_i;\theta_0)\, d\alpha_i.$$

It follows that $\Lambda^*(t) = \Lambda_0(t)$ for $t \in [0, \tau]$. Therefore, $(\hat{\beta}_n, \hat{\theta}_n, \hat{\Lambda}_n(t)) \to (\beta_0, \theta_0, \Lambda_0\,(t))$ almost surely by Helly's theorem, which can be strengthened to uniform convergence on $[0, \tau]$.

Next, we show that $\hat{\beta}_n$ is asymptotically normal and its variance achieves the semiparametric efficiency bound. Let BV$[0, \tau]$ denote the space of bounded variation functions on $[0, \tau]$ and define class $H = \{h = (h_{11}, h_{12}, h_2) : h_{11} \in \mathbb{R}^p$ with $\|h_{11}\|_1 < \infty$, $|h_{12}| < \infty$, $h_2 \in$ BV$[0, \tau]\}$. For $h \in H$, define the norm $\|h\| = \|h_{11}\|_1 + |h_{12}| + \|h_2\|_\upsilon$, where $\|h_{11}\|_1$ is the L_1 norm of h_{11}, and $\|h_2\|_\upsilon$ is the absolute value of $h_2(0)$ plus the total variation of h_2 on the interval $[0, \tau]$. Consider submodels $\beta_d = \beta + dh_{11}$, $\theta_d = \theta + dh_{12}$ and $\Lambda_d\,(t) = \int_0^t \{1 + dh_2(u)\} d\Lambda(u)$. Further, define $U_n\,(\beta, \theta, \Lambda)\,(h_{11}, h_{12}, h_2) = n^{-1}\, \{\partial l_n^o(\beta_d, \theta_d, \Lambda_d)/\partial d\}|_{d=0}$, where $(h_{11}, h_{12}, h_2) \in H$. For simplicity of notation, we denote

$$A_i = A_i\,(\beta, \Lambda) = \prod_{j=1}^{m_i} \{\alpha_i \lambda\,(e^{R_{ij}\,(\beta)}) e^{\beta'\, X_{ij}}\}^{\delta_{ij}}\, e^{-\alpha_i \Lambda\,(e^{R_{ij}\,(\beta)})}.$$

Then we can write $U_n(\beta, \theta, \Lambda)\,(h_{11}, h_{12}, h_2) = U_{n1}(h_{11}) + U_{n2}(h_2) + U_{n3}(h_{12})$, where

$$\begin{aligned} U_{n1}\,(h_{11}) &= \frac{1}{n}\frac{\partial}{\partial d}|_{d=0}\, l_n^o\,(\beta_d, \Lambda, \theta) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\int_0^\infty A_i\, \left[\sum_{j=1}^{m_i} h_{11}^{'} X_{ij} \{\delta_{ij}\, \dot{\lambda}\,(e^{R_{ij}\,(\beta)})\, e^{R_{ij}\,(\beta)}/\lambda\,(e^{R_{ij}\,(\beta)})/ + \delta_{ij} - \alpha_i\, \lambda(e^{R_{ij}\,(\beta)}) e^{R_{ij}\,(\beta)}\}\right]\, f_\alpha\,(\alpha_i\,;\theta) d\alpha_i}{\int_0^\infty A_i\, f_\alpha\,(\alpha_i\,;\theta) d\alpha_i}, \end{aligned} \quad \text{(A1)}$$

$$U_{n2}\,(h_2) = \frac{1}{n}\frac{\partial}{\partial d}|_{d=0}\, l_n^o\,(\beta, \Lambda_d, \theta) = \frac{1}{n}\sum_{i=1}^n \frac{\int_0^\infty A_i\, \left\{\sum_{j=1}^{m_i} \delta_{ij}\, h_2\,(e^{R_{ij}\,(\beta)}) - \alpha_i \sum_{j=1}^{m_i} \int_0^{e^{R_{ij}\,(\beta)}} h_2\,(u)\, d\Lambda(u)\right\}\, f_\alpha\,(\alpha_i;\theta) d\alpha_i}{\int_0^\infty A_i\, f_\alpha(\alpha_i;\theta) d\alpha_i}, \quad \text{(A2)}$$

$$U_{n3}\,(h_{12}) = \frac{1}{n}\frac{\partial}{\partial d}|_{d=0}\, l_n^o(\beta, \Lambda, \theta_d) = \frac{1}{n}\sum_{i=1}^n \frac{\int_0^\infty h_{12}\, A_i \{\partial f_\alpha\,(\alpha_i;\theta)/\partial\theta\} d\alpha_i}{\int_0^\infty A_i\, f_\alpha\,(\alpha_i;\theta) d\alpha_i} \quad \text{(A3)}$$

Define $u(\beta, \theta, \Lambda)\,(h_{11}, h_{12}, h_2) = \lim_{n\to\infty} U_n(\beta, \theta, \Lambda)\,(h_{11}, h_{12}, h_2) \equiv u_1(h_{11}) + u_2(h_2) + u_3(h_{12})$. It can be easily shown that $u(\beta_0, \theta_0, \Lambda_0)\,(h_{11}, h_{12}, h_2) = 0$. In addition, it is easy to show that $u(\beta, \theta, \Lambda)$ is Fréchet differentiable since $u(\beta, \theta, \Lambda)$ is a smooth function of β, θ and Λ. Let $u(\beta_0, \theta_0, \Lambda_0)\,(\beta - \beta_0, \theta - \theta_0, \Lambda - \Lambda_0)\,(h)$ denote the corresponding Fréchet derivative of $u(\beta, \theta, \Lambda)$ at $(\beta_0, \theta_0, \Lambda_0)$. After some algebra, we have

$$\dot{u}(\beta_0,\theta_0,\Lambda_0)(\beta-\beta_0,\theta-\theta_0,\Lambda-\Lambda_0)(h)=(\beta-\beta_0)'Q_1(h)+\int_0^\infty Q_2(t,h)\,d\{\Lambda(t)-\Lambda_0(t)\}+(\theta-\theta_0)Q_3(h),$$

where $h = (h_{11}, h_{12}, h_2)$,

$$\begin{aligned} Q_1(h) &= B_1 \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} + \int_0^\infty D_1(t)\, h_2(t)\, dt, \\ Q_2(t,h) &= B_2(t)' \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} + c_2(t)\, h_2(t) + \int_0^\infty D_2(t,u)\, h_2(u)\, du, \\ Q_3(h) &= B_3' \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} + \int_0^\infty D_3(t)\, h_2(t)\, dt, \end{aligned}$$

where B_1 is a $p \times (p + 1)$ matrix, $B_2(t)$ and B_3 are $(p + 1)$-dimensional vectors, $D_1(t)$ is a p-dimensional function, and $c_2(t)$, $D_2(t, u)$ and $D_3(t)$ are scalar functions. Therefore, $Q(h) \equiv (Q_1(h), Q_2(h), Q_3(h))$ is a continuous linear operator from the linear span of H to itself.

Consider two classes of functions:

$$\begin{gathered} A_1(\beta_0,\theta_0,\Lambda_0)=\left\{h_{11}' U_1^*(\beta_0,\theta_0,\Lambda_0)+h_{12}U_3^*(\beta_0,\theta_0,\Lambda_0):\|h_{11}\|_1<\infty, |h_{12}|<\infty\right\}, \\ A_2(\beta_0,\theta_0,\Lambda_0)=\{U_2^*(\beta_0,\theta_0,\Lambda_0)(h_2):h_2 \in BV[0,\tau]\}, \end{gathered}$$

where

$$\begin{gathered} U_1^*(\beta_0,\theta_0,\Lambda_0)=\frac{\int_0^\infty A_{i0}\left[\sum_{j=1}^{m_i} X_{ij}\left\{\delta_{ij}\,\dot{\lambda}_0(e^{R_{ij}(\beta_0)})e^{R_{ij}(\beta_0)}/\lambda_0(e^{R_{ij}(\beta_0)})+\delta_{ij}-\alpha_i\,\lambda(e^{R_{ij}(\beta_0)})e^{R_{ij}(\beta_0)}\right\}\right] f_\alpha(\alpha_i;\theta_0)d\alpha_i}{\int_0^\infty A_{i0}\, f_\alpha(\alpha_i;\theta_0)d\alpha_i}, \\ U_2^*(\beta_0,\theta_0,\Lambda_0)(h_2)=\frac{\int_0^\infty A_{i0}\left\{\sum_{j=1}^{m_i}\delta_{ij}\,h_2(e^{R_{ij}(\beta_0)})-\alpha_i\sum_{j=1}^{m_i}\int_0^{e^{R_{ij}(\beta_0)}} h_2(u)d\Lambda_0(u)\right\} f_\alpha(\alpha_i;\theta_0)d\alpha_i}{\int_0^\infty A_{i0}\, f_\alpha(\alpha_i;\theta_0)d\alpha_i}, \\ U_3^*(\beta_0,\theta_0,\Lambda_0)=\frac{\int_0^\infty A_{i0}\{\partial f_\alpha(\alpha_i;\theta_0)/\partial\theta\}d\alpha_i}{\int_0^\infty A_{i0}\, f_\alpha(\alpha_i;\theta_0)d\alpha_i} \end{gathered}$$

and $A_{i0} = A_i(\beta_0, \Lambda_0)$. Since $U_1^*(\beta_0,\theta_0,\Lambda_0)$ and $U_3^*(\beta_0,\theta_0,\Lambda_0)$ are bounded functions based on assumptions (C1)–(C5), A_1 is a Donsker class. In addition, since $h_2 \in \mathrm{BV}[0, \tau]$, A_2 can be written as the summation of bounded Donsker classes, which is also a Donsker class. Therefore, we have $n^{1/2}\{U_n(\beta_0, \theta_0, \Lambda_0)(h) - u(\beta_0, \theta_0, \Lambda_0)(h)\}$ converges weakly to a Gaussian process G^* on $l^\infty(H)$.

In addition, since $\|\beta - \beta_0\|_1 + |\theta - \theta_0| = o_p(1)$ and $\sup_{t\in[0, \tau]} |\Lambda(t) - \Lambda_0(t)| = o_p(1)$, we can show that $A_1(\beta, \theta, \Lambda)$ and $A_2(\beta, \theta, \Lambda)$ are Donsker classes. This implies that

$$\begin{aligned} &\sup_{h\in H}\left|(U_n-u)(\hat{\beta}_n,\hat{\theta}_n,\hat{\Lambda}_n)(h)-(U_n-u)(\beta_0,\theta_0,\Lambda_0)(h)\right| \\ &=o_p\left\{\max\left(n^{-1/2},\|\hat{\beta}_n-\beta_0\|_1+|\hat{\theta}_n-\theta_0|+\sup_{t\in[0,\tau]}|\hat{\Lambda}_n(t)-\Lambda_0(t)|\right)\right\}. \end{aligned} \tag{A4}$$

Finally, we show that $\dot{u}(\beta_0, \theta_0, \Lambda_0)$ is continuously invertible. It is equivalent to show that $Q(h)$ is a one to one map, i.e., $Q(h) = 0$ implies $h = 0$. If $Q(h) = 0$, $\dot{u}(\beta_0, \theta_0, \Lambda_0) = 0$ for $(\beta, \theta,$

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Λ) in a neighbourhood of (β_0, θ_0, Λ_0). We choose $\beta = \beta_0 + dh_{11}$, $\theta = \theta_0 + dh_{12}$ and $\Lambda(t) = \Lambda_0(t) + d\int_0^t h_2(u)\, d\Lambda_0(u)$ for a small constant d. By the definition of $u(\beta_0, \Lambda_0, \theta_0)$, we have

$\dot{u}(\beta_0, \Lambda_0, \theta_0) = dE[\{h_{11}' U_1^*(\beta_0, \theta_0, \Lambda_0) + U_2^*(\beta_0, \theta_0, \Lambda_0)(h_2) + h_{12} U_3^*(\beta_0, \theta_0, \Lambda_0)\}^2] = 0$. This implies that $h_{11}' U_1^*(\beta_0, \theta_0, \Lambda_0) + U_2^*(\beta_0, \theta_0, \Lambda_0)(h_2) + h_{12} U_3^*(\beta_0, \theta_0, \Lambda_0) = 0$ almost surely. Following the techniques used to derive the consistency of the estimators, we can show that $h = 0$. The details are given in the Supplementary Material.

Since $u(\beta_0, \theta_0, \Lambda_0)$ is continuously invertible on its range, based on Theorem 3.3.1. of van der Vaart & Wellner (1996), we have that $n^{1/2}[\{\hat{\gamma}_n, \hat{\Lambda}_n(t)\} - \{\gamma_0, \Lambda_0(t)\}]$ converges weakly to a tight Gaussian process $G = \{u(\beta_0, \theta_0, \Lambda_0)\}^{-1} G^*$. In addition, the variance of G is

$$\mathrm{var}\{G(h)\} = \int_0^\infty h_2(t) Q_2^{-1}(t, h)\, d\Lambda_0(t) + (h_{11}', h_{12}) \begin{pmatrix} Q_1^{-1}(h) \\ Q_3^{-1}(h) \end{pmatrix},$$

where $Q^{-1}(h) \equiv \left\{Q_1^{-1}(h), Q_2^{-1}(h), Q_3^{-1}(h)\right\}$ is the inverse of $Q(h)$. We derive the semiparametric efficiency bound I^{-1} (Bickel et al., 1993) and show that the asymptotic variance of $n^{1/2}(\hat{\gamma}_n - \gamma_0)$ achieves the semiparametric efficiency bound. The details are given in the Supplementary Material.

References

Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. The Annals of Applied Probability. 1992; 2:951–972.

Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. Baltimore: Johns Hopkins University Press; 1993.

Buckley J, James I. Linear regression with censored data. Biometrika. 1979; 66:429–436.

Chen HY, Little RJA. Proportional hazards regression with missing covariates. Journal of the American Statistical Association. 1999; 94:896–908.

Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. 1978; 65:141.

Clayton DG, Cuzick J. Multivariate generalizations of the proportional hazards model (with discussion). Journal of the Royal Statistical Society A. 1985; 48:82–117.

Cox DR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society Series B. 1972; 34:187–220.

Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Biometrika. 1986; 73:387–396.

Hougaard P. Modelling multivariate survival. Scandinavian Journal of Statistics. 1987; 14:291–30.

Huster WJ, Brookmeyer R, Self SG. Modelling paired survival data with covariates. Biometrics. 1989; 45:145–156. [PubMed: 2655727]

Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. Biometrika. 2003; 90:341–353.

Jin Z, Lin DY, Ying Z. On least-squares regression with censored data. Biometrika. 2006a; 93:147–161.

Jin Z, Lin DY, Ying Z. Rank regression analysis of multivariate failure time data based on marginal linear models. Scandinavian Journal of Statistics. 2006b; 33:1–23.

Johnson L, Strawderman R. Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. Biometrika. 2009; 96:577–590. [PubMed: 23049117]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Johnson L, Strawderman R. A smoothing expectation and substitution algorithm for the semiparametric accelerated failure time frailty model. Statistics in Medicine. 2012; 31:2335–2358. [PubMed: 22437629]

Jones MC. The performance of kernel density functions in kernel distribution function estimation. Statistics and Probability Letter. 1990; 9:129–132.

Jones MC, Sheather SJ. Using non-stochastic terms to adavantage in kernelbased estimation of integrated squared density derivatives. Statistics and Probability Letter. 1991; 11:511–514.

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2. New York; Wiley: 2002.

Li H, Yin GS. Generalized method of moments estimation for linear regression with clustered failure time data. Biometrika. 2009; 96:293–306.

Lin DY. Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine. 1994; 13:2233–2247. [PubMed: 7846422]

Lin DY, Wei LJ, Ying Z. Accelerated failure time models for counting processes. Biometrika. 1998; 85:605–618.

Lu W. Tests of independence for censored bivariate failure time data. Lifetime Data Analysis. 2007; 13:75–90. [PubMed: 17221306]

Lu W. Efficient estimation for an accelerated failure time model with a cure fraction. Statistica Sinica. 2010; 20:661–674. [PubMed: 20414460]

McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics. 1991; 47:461–466. [PubMed: 1912255]

Murphy SA. Consistency in a proportional hazard model incorporating a random effect. Annals of Statistics. 1994; 22:712–731.

Murphy SA. Asymptotic theory for the frailty model. Annals of Statistics. 1995; 23:182–198.

Nielsen GG, Gill RD, Andersen PK, Sorensen TIA. A counting process approach to maximum likelihood estimation in frailty models. Scandinavian Journal of Statistics. 1992; 19:25–44.

Oakes D. Bivariate survival models induced by frailties. Journal of the American Statistical Association. 1989; 84:487–493.

Pan W. Using frailties in the accelerated failure time model. Lifetime Data Analysis. 2001; 7:55–64. [PubMed: 11280848]

Parner E. Asymptotic theory for the correlated gamma-frailty model. Annals of Statistics. 1998; 26:183–214.

Schuster EF. Estimation of a probability density function and its derivatives. The Annals ofMathematical Statistics. 1969; 40:1187–1195.

Strawderman R. A regression model for dependent gap times. International Journal of Biostatistics. 2006; 2:1–33.

Tsiatis AA. Estimating regression parameteters using linear rank tests for censored data. The Annals of Statistics. 1990; 18:354–372.

van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. New York: Springer-Verlag; 1996.

Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. Journal of the American Statistical Association. 1989; 84:1065–1073.

Xu L, Zhang J. An EM-like algorithm for the semiparametric accelerated failure time gamma frailty model. Computational Statistics and Data Analysis. 2010; 54:1467–1474.

Ying Z. A large sample study of rank estimation for censored regression data. The Annals of Statistics. 1993; 21:76–99.

Zeng D, Lin DY. Efficient estimation for the accelerated failure time model. Journal of the American Statistical Association. 2007; 102:1387–1396.

Zhang J, Peng Y. An alternative estimation method for the accelerated failure time frailty model. Computational Statistics and Data Analysis. 2007; 51:4413–4423.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Simulation results for gamma frailty.

					NPMLE				J-S[1]		J-S[2]	
CR	σ^2	$\hat{\sigma^2}$	SD	β	Bias(%)	SD	SE	CP(%)	Bias(%)	SD	Bias(%)	SD
					$\lambda_0(t)$: Weibull type							
15%	2	1.98	.30	$\hat{\beta_1}$	−0.5	.068	.08	97	0.3	.078	0.0	.104
				$\hat{\beta_2}$	0.5	.061	.07	97	−0.4	.069	0.1	.090
40%	2	1.96	.31	$\hat{\beta_1}$	−1.6	.085	.09	95	1.2	.090	−1.0	.109
				$\hat{\beta_2}$	1.5	.076	.08	94	−0.9	.081	1.3	.096
15%	1	0.98	.17	$\hat{\beta_1}$	−0.5	.069	.07	95	0.3	.075	−0.4	.082
				$\hat{\beta_2}$	0.4	.060	.06	94	−0.3	.064	0.3	.071
40%	1	0.97	.19	$\hat{\beta_1}$	−2.2	.082	.10	97	0.4	.089	−1.1	.093
				$\hat{\beta_2}$	2.3	.071	.08	96	−0.3	.076	1.1	.081
15%	0.5	0.48	.12	$\hat{\beta_1}$	−0.9	.062	.08	97	−0.1	.069	−0.5	.069
				$\hat{\beta_2}$	1.1	.055	.07	97	0.1	.061	0.7	.062
40%	0.5	0.47	.14	$\hat{\beta_1}$	−2.4	.077	.09	96	−0.2	.083	−1.7	.082
				$\hat{\beta_2}$	2.9	.070	.08	96	0.2	.075	1.9	.074
					$\lambda_0(t)$: reciprocal type							
15%	2	2.00	.30	$\hat{\beta_1}$	−0.3	.154	.17	96	0.6	.161	0.3	.234
				$\hat{\beta_2}$	1.4	.138	.14	95	0.5	.143	0.8	.208
40%	2	2.00	.32	$\hat{\beta_1}$	−2.5	.173	.18	95	1.0	.179	0.3	.219
				$\hat{\beta_2}$	4.0	.154	.16	94	0.8	.157	1.0	.201
15%	1	0.99	.17	$\hat{\beta_1}$	−1.5	.143	.15	96	−0.5	.144	−0.9	.166
				$\hat{\beta_2}$	1.4	.125	.13	96	0.8	.128	0.5	.148
40%	1	0.98	.22	$\hat{\beta_1}$	−4.2	.163	.18	96	−0.8	.160	−1.6	.172
				$\hat{\beta_2}$	5.2	.146	.16	96	1.0	.145	1.0	.152
15%	0.5	0.48	.11	$\hat{\beta_1}$	−1.0	.136	.13	94	0.6	.136	0.2	.139

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

CR	σ^2	$\hat{\sigma^2}$	SD	β	NPMLE Bias(%)	NPMLE SD	NPMLE SE	NPMLE CP(%)	J-S[1] Bias(%)	J-S[1] SD	J-S[2] Bias(%)	J-S[2] SD
				$\hat{\beta_2}$	1.1	.124	.12	93	−1.0	.122	0.2	.122
40%	0.5	0.47	.12	$\hat{\beta_1}$	−5.1	.156	.17	95	0.8	.157	−0.3	.156
				$\hat{\beta_2}$	5.2	.140	.15	94	−0.9	.139	1.0	.138

†CR, censoring rate; σ^2, true variance of frailty; $\hat{\sigma^2}$, estimate for σ^2; SD, sample standard deviation; SE, mean of estimated standard errors; CP, empirical coverage probability of 95% Wald-type confidence interval; NPMLE, proposed nonparametric maximum likelihood estimator; J-S[1], smoothed EM-like estimator; J-S[2], induced smoothing estimator.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Simulation results for log-normal frailty.

					NPMLE				J-S[1]		J-S[2]	
CR	σ^2	$\hat{\sigma^2}$	SD	β	Bias(%)	SD	SE	CP(%)	Bias(%)	SD	Bias(%)	SD
					$\lambda_0(t)$: Weibull type							
15%	3.48	3.43	1.26	$\hat{\beta_1}$	−1.0	.066	.08	97	0.3	.077	−0.6	.083
				$\hat{\beta_2}$	1.0	.059	.07	97	−0.2	.066	0.6	.073
40%	3.48	3.10	1.31	$\hat{\beta_1}$	−2.3	.085	.10	96	0.5	.093	−1.5	.099
				$\hat{\beta_2}$	2.5	.074	.09	97	−0.4	.080	1.7	.085
15%	1.72	1.70	0.55	$\hat{\beta_1}$	−0.8	.066	.07	96	−0.2	.073	−0.5	.076
				$\hat{\beta_2}$	0.9	.058	.06	96	0.1	.063	0.5	.066
40%	1.72	1.59	0.59	$\hat{\beta_1}$	−2.8	.083	.09	95	0.5	.089	−1.6	.092
				$\hat{\beta_2}$	3.0	.073	.08	96	−0.3	.077	1.7	.078
15%	0.65	0.64	0.21	$\hat{\beta_1}$	−0.7	.063	.07	96	−0.1	.070	−0.5	.067
				$\hat{\beta_2}$	0.8	.056	.06	96	0.2	.060	0.5	.059
40%	0.65	0.63	0.23	$\hat{\beta_1}$	−3.1	.080	.09	95	−0.2	.086	−1.6	.083
				$\hat{\beta_2}$	3.0	.069	.08	95	0.5	.075	1.6	.072
					$\lambda_0(t)$: reciprocal type							
15%	3.48	3.41	1.24	$\hat{\beta_1}$	−0.8	.143	.17	98	0.8	.151	0.4	.171
				$\hat{\beta_2}$	1.1	.128	.15	97	−0.2	.133	−0.3	.157
40%	3.48	2.94	1.05	$\hat{\beta_1}$	−3.8	.171	.19	96	1.0	.173	−0.2	.187
				$\hat{\beta_2}$	4.7	.151	.17	96	−0.5	.152	0.5	.169
15%	1.72	1.69	0.53	$\hat{\beta_1}$	−1.0	.140	.16	97	0.7	.146	0.2	.153
				$\hat{\beta_2}$	1.5	.126	.14	96	−0.3	.132	−0.1	.140
40%	1.72	1.53	0.51	$\hat{\beta_1}$	−4.5	.168	.18	96	1.2	.167	−0.3	.172
				$\hat{\beta_2}$	5.1	.149	.16	95	0.5	.148	0.7	.156
15%	0.65	0.65	0.19	$\hat{\beta_1}$	−1.0	.134	.14	95	0.8	.136	0.1	.135

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

					NPMLE				J-S[1]		J-S[2]	
CR	σ^2	$\hat{\sigma^2}$	SD	β	Bias(%)	SD	SE	CP(%)	Bias(%)	SD	Bias(%)	SD
				$\hat{\beta_2}$	1.5	.123	.13	94	−0.3	.124	0.1	.124
40%	0.65	0.62	0.21	$\hat{\beta_1}$	−5.1	.160	.18	95	0.5	.159	−0.4	.155
				$\hat{\beta_2}$	6.1	.142	.16	95	0.3	.141	0.6	.142

†CR, censoring rate; σ^2, true variance of frailty; $\hat{\sigma}^2$, estimate for σ^2; SD, sample standard deviation; SE, mean of estimated standard errors; CP, empirical coverage probability of 95% Wald-type confidence interval; NPMLE, proposed nonparametric maximum likelihood estimator; J-S[1], smoothed EM-like estimator; J-S[2], induced smoothing estimator.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Sensitivity analysis results for misspecified frailty distribution

CR	σ^2	$\hat{\sigma^2}$	SD	β	Bias(%)	SD	SE	CP(%)
				$\lambda_0(t)$: Weibull type				
15%	3.48	1.01	0.17	$\hat{\beta_1}$	−1.1	.067	.08	97
				$\hat{\beta_2}$	1.0	.060	.07	97
40%	3.48	1.04	0.20	$\hat{\beta_1}$	−3.4	.086	.10	96
				$\hat{\beta_2}$	3.3	.075	.09	97
15%	1.72	0.70	0.12	$\hat{\beta_1}$	−0.9	.067	.07	96
				$\hat{\beta_2}$	1.0	.059	.07	97
40%	1.72	0.72	0.15	$\hat{\beta_1}$	−3.3	.084	.09	95
				$\hat{\beta_2}$	3.2	.072	.08	96
15%	0.65	0.39	0.10	$\hat{\beta_1}$	−0.8	.063	.07	96
				$\hat{\beta_2}$	0.8	.056	.06	97
40%	0.65	0.39	0.12	$\hat{\beta_1}$	−3.2	.080	.09	96
				$\hat{\beta_2}$	3.0	.069	.08	96
				λ_0 (t): reciprocal type				
15%	3.48	1.04	0.16	$\hat{\beta_1}$	−1.2	.145	.17	98
				$\hat{\beta_2}$	1.4	.130	.15	97
40%	3.48	1.14	0.22	$\hat{\beta_1}$	−5.3	.174	.19	96
				$\hat{\beta_2}$	5.4	.153	.17	96
15%	1.72	0.74	0.13	$\hat{\beta_1}$	−1.6	.141	.16	97
				$\hat{\beta_2}$	1.7	.127	.14	96
40%	1.72	0.79	0.16	$\hat{\beta_1}$	−5.4	.169	.18	95
				$\hat{\beta_2}$	5.6	.151	.16	95
15%	0.65	0.38	0.09	$\hat{\beta_1}$	−1.6	.134	.14	95
				$\hat{\beta_2}$	1.8	.123	.12	94

CR	σ^2	$\hat{\sigma^2}$	SD	β	Bias(%)	SD	SE	CP(%)
40%	0.65	0.41	0.11	$\hat{\beta_1}$	−6.6	.160	.18	95
				$\hat{\beta_2}$	6.8	.146	.16	95

†CR, censoring rate; σ^2, true variance of frailty; $\hat{\sigma}^2$, estimate for σ^2; SD, sample standard deviation; SE, mean of estimated standard errors; CP, empirical coverage probability of 95% Wald-type confidence interval.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Analysis results for diabetes data.

	NPMLE$_g$			NPMLE$_l$			GehanR		
	Est.	**SE**	**pv**	**Est.**	**SE**	**pv**	**Est.**	**SE**	**pv**
treatment	−0.929	.104	.000	−0.949	.105	.000	−0.981	.201	.000
age	0.011	.006	.078	0.008	.006	.162	0.010	.016	.551
type	−0.029	.102	.779	0.070	.096	.467	−0.301	.453	.506
risk	1.660	.353	.000	1.491	.346	.000	2.568	1.036	.013

†NPMLE$_g$, nonparametric maximum likelihood estimator with the gamma frailty; NPMLE$_l$, nonparametric maximum likelihood estimator with the log-normal frailty; GehanR, Gehan rank estimator; Est., estimated coefficients; pv, *p*-value; age, age at diagnosis of diabetes; type, type of diabetes; risk, risk group.