Quantitative Bio-Science 31(1), 39~45(2012)

Comparison of Estimation Methods for Semi-parametric Frailty Models

Il Do Ha*

Department of Asset Management, Daegu Haany University, Gyeongsan 712-715, Korea
(Received April 5, 2012; Revised May 10, 2012; Accepted May 15, 2012)

ABSTRACT

Frailty models have been widely used for the analysis of univariate or multivariate time-to-event data. We firstly review likelihood-based methods including marginal likelihood and hierarchical likelihood (h-likelihood). In this paper, we study and discuss various likelihood approaches based on marginal likelihood and h-likelihood under semiparametric frailty models with nonparametric baseline hazards. For the comparison, with semiparametric gamma frailty models the practical-data examples and simulation studies are presented. Furthermore, both h-likelihood and Bayesian approaches are also compared and their relationship is studied.

Key words : Bayesian approach, Frailty models, Hierarchical likelihood, Marginal likelihood, Profile likelihood

Introduction

Frailty models have been widely used for the analysis of univariate or correlated time-to-event data [1,2]. For the inference of semiparametric frailty models, extensions of Cox's proportional hazards models, many authors have proposed likelihood-based methods [3-6]. In particular, the gamma frailty model has been often used because it gives an explicit marginal likelihood. The usual marginal likelihood methods such as the EM algorithm use the discrete nonparametric Breslow estimates for unknown baseline hazards playing the role of nuisance parameters: see for example Nielsen et al. [3] and Andersen et al. [7].

Recently, in semiparametric gamma frailty models Rondeau et al. [8] and Barker & Henderson [9] have numerically showed that the use of Breslow estimates in the EM can lead to small-sample underestimation of parameters, particularly for frailty parameters. For the reduction of the bias they have proposed the use of the continuous nonparametric estimates, instead of the Breslow estimate, for the baseline hazards. However, the bias problem may also occur because the number of nuisance parameters in baseline hazards increases with sample size. Under this situation the uncertainties in the nuisance parameter estimation should be considered in estimating the frailty parameter. Thus, this problem can be solved by the use of appropriate profile likelihood methods for eliminating them.

In this paper we study and discuss various likelihood approaches based on marginal likelihood [3][7] and h-likelihood [10-12], and show how to profile the nuisance parameters. Note here that we still use the Breslow estimates. For the illustration, with semiparametric gamma frailty models the real-data examples and small-sample simulation studies are presented. Furthermore, both h-likelihood and Bayesian approaches are also compared and their relationship is studied.

Frailty Models

Let $T_{ij}(i=1, ..., q, j=1, ..., n_i, n=\sum_i n_i)$ be the time-to-event (survival time) for the jth observation of the ith individual (e.g. subject or cluster) and C_{ij} be the corresponding censoring time. Let the observable random variables be $y_{ij}=\min(T_{ij}, C_{ij})$ and $\delta_{ij}=I(T_{ij}\leq C_{ij})$, where $I(\cdot)$ is the indicator function. Denote by U_i the unobserved frailty random variable (or random effect) for

* Correspondence should be addressed to Dr. Il Do Ha, Department of Asset Management, Daegu Haany University, Gyeongsan 712-715, South Korea. Tel: +82-53-819-1457, Fax: +82-53-819-1276, E-mail: idha@dhu.ac.kr

the ith individual.

The frailty models are described as follows. Given $U_i=u_i$ the conditional hazard function of T_{ij} is of the form

$$\lambda_{ij}(t|u_i)=\lambda_0(t)\exp(x_{ij}^T\beta)u_i, \quad (1)$$

where $\lambda_0(\cdot)$ is a parametric or nonparametric baseline hazard function, $x_{ij}=(x_{ij1}, ..., x_{ijp})^T$ is a vector of fixed covariates and is a vector of the corresponding regression parameters. For identifiability purposes the term $x_{ij}^T\beta$ does not include an intercept term if the baseline hazard $\lambda_0(\cdot)$ is unspecified. The frailties U_i are assumed to be independent and identically distributed random variables with a density function having frailty parameter α.

For the gamma distribution the marginal likelihood is explicitly obtainable, whereas for the lognormal distribution it is not, but the penalized partial likelihood [13], an approximation of the marginal likelihood, is available. For both frailty distributions the h-likelihood provides a simple unified framework [11, 14]. It is worthy to note that the h-likelihood method can be easily applied to other frailty distributions such as the inverse Gaussian.

Likelihood-based Methods

Following Lee and Nelder [10] and Ha et al. [11], the h-likelihood for frailty models (1) is defined by

$$h=h(\beta, \lambda_0, \alpha)=\sum_{ij} l_{1ij}+\sum_i l_{2i}, \quad (2)$$

where

$$l_{1ij}=l_{1ij}(\beta, \lambda_0; y_{ij}^*|u_i)=\delta_{ij}\{\log\lambda_0(y_{ij})+\eta_{ij}\}-\{\Lambda_0(y_{ij})\exp(\eta_{ij})\}$$

is the logarithm of the conditional density function for $y_{ij}^*=(y_{ij}, \delta_{ij})$ given $U_i=u_i$, $l_{2i}=l_{2i}(\alpha; u_i)$ is the logarithm of the density function for $V_i=\log(U_i)$ with parameter α, $\Lambda_0(t)=\int_{-\infty}^{t}\lambda_0(k)$, is the baseline cumulative hazard function, and $\eta_{ij}=x_{ij}^T\beta+u_i$ with $u_i=\log(u_i)$.

Marginal likelihood, denoted by m, has been often used for inference; it can be obtained by integrating out the frailties from the h-likelihood of (2):

$$m=m(\beta, \lambda_0, \alpha)=\sum_i \log\left\{\int\exp(h_i)\,dv_i\right\}, \quad (3)$$

where $h_i=\sum_j l_{1ij}+l_{2i}$ is the contribution of the ith individual to h in (2). The marginal likelihood in (3) often requires intractable integration (e.g. lognormal frailty) except for gamma frailty [11].

Assume that the functional form of $\lambda_0(t)$ is known; for example, $\lambda_0(t)=\phi_1\phi_2 t^{\phi 2-1}$ is Weibull baseline hazard and $\lambda_0=(\phi_1,\phi_2)^T$. Let $l=l(\beta^*, \theta)$ with $\beta^*=(\beta, \lambda_0)$ be a likelihood, either an h-likelihood, h, or a marginal likelihood m, with nuisance parameters θ. Lee and Nelder [15] considered a function $p_\theta(l)$, defined by

$$p_\theta(l)=\left[l-\frac{1}{2}\log\det\left\{D(l, \theta)/(2\pi)\right\}\right]\Big|_{\theta=\hat\theta},$$

where $D(l, \theta)=-\partial^2 l/\partial\theta^2$ and $\hat\theta$ solves $\partial l/\partial\theta=0$. The function $p_\theta(\cdot)$ produces an adjusted profile likelihood, eliminating nuisance effects θ, which can be fixed effects or β^* random effects u or both. In general, $p_u(h)$ is the first-order Laplace approximation to m(i.e. $p_u(h)\simeq m$) and $p_{\beta^*,u}(h)\simeq p_{\beta^*}(m)$ [14,15]. In principle, we argue, that one should use the h-likelihood, h, for inferences about u; the marginal-likelihood, m, for β^*; and the restricted likelihood, $p_{\beta^*}(m)$, for the dispersion parameter α. When m is numerically difficult to obtain, we can use $p_v(h)$ and $p_{\beta^*,u}(m)$ as approximations to m and $p_{\beta^*}(m)$, respectively. Furthermore, the second-order Laplace approximation [15,16], denoted by $p_v^s(h)$, to m is defined by

$$p_v^s(h)=p_u(h)-F(h)/24, \quad (4)$$

where $F(h)=\text{trace}(S)_{u=\hat u}$ with

$$S=-\{3(\partial^4 h/\partial u^4)+5(\partial^3 h/\partial u^3)D(h, u)^{-1}(\partial^3 h/\partial u^3)\}D(h, u)^{-2}.$$

Now, consider the gamma frailty models with $E(U_i)=1$ and $\text{var}(U_i)=\alpha$. From (3) we have an explicit marginal likelihood:

$$\begin{aligned} m=m(\beta, \lambda_0, \alpha)=&\sum_{ij}[\delta_{ij}\{x_{ij}^T\beta+\log\lambda_0(y_{ij})\}] \\ &+\sum_i\{-(\alpha^{-1}+\delta_{i+})\log(\alpha^{-1}+\mu_{i+}) \\ &+\log\Gamma(\alpha^{-1}+\delta_{i+})-c(\alpha)\}, \end{aligned} \quad (5)$$

where $\delta_{i+}=\sum_j\delta_{ij}$, $\mu_{i+}=\sum_j\mu_{ij}=\sum_j\Lambda_0(y_{ij})\exp(x_{ij}^T\beta)$ and $c(\alpha)=\log\Gamma(\alpha^{-1})+\alpha^{-1}\log\alpha$. The corresponding h-likelihood is given by

$$\begin{aligned} h=h(\beta, \lambda_0, \alpha)=&\sum_{ij}[\delta_{ij}\{x_{ij}^T\beta+\log\lambda_0(y_{ij})\}] \\ &+\sum_i\{(\alpha^{-1}+\delta_{i+})u_i-(\alpha^{-1}+\mu_{i+})u_i-c(\alpha)\}. \end{aligned}$$

From

$$\partial h/\partial u_i=(\delta_{i+}+\alpha^{-1})-(\mu_{i+}+\alpha^{-1})u_i=0,$$

we have

$$\hat u_i=\frac{\alpha^{-1}+\delta_{i+}}{\alpha^{-1}+\mu_{i+}}$$

Note here that the adjustment term for $p_u(h)$,

$$D(h, u)|_{u_i=\hat{u}_i} = -\partial^2 h/\partial u^2|_{u_i=\hat{u}_i} = (\alpha^{-1}+\mu_{i+})\hat{u}_i = \alpha^{-1}+\delta_{i+},$$

is free of (β, λ_0) but depends upon α. We have that

$$\begin{aligned} p_u(h) &= \left[h - \frac{1}{2}\log\det\left\{D(h, u)/(2\pi)\right\}\right]\Big|_{u=\hat{u}} \\ &= \sum_{ij}[\delta_{ij}\{x_{ij}^T\beta + \log\lambda_0(y_{ij})\}] \\ &+ \sum_i\{-(\alpha^{-1}+\delta_{i+})\log(\alpha^{-1}+\mu_{i+}) \\ &+ (\alpha^{-1}+\delta_{i+})\log(\alpha^{-1}+\delta_{i+}) - (\alpha^{-1}+\delta_{i+}) \\ &- \log(\alpha^{-1}+\delta_{i+})/2 + \log(2\pi)/2 - c(\alpha)\}, \end{aligned} \tag{6}$$

which is equivalent to approximating m of (5) by the first-order Stirling approximation

$$\log\Gamma(x) \doteq (x-1/2)\log(x) + \log(2\pi)/2 - x$$

for $\Gamma(\alpha^{-1}+\delta_{i+})$. Thus, the marginal maximum likelihood (ML) estimator for β (maximizing $p_u(h)$) can be obtained by maximization of h. Furthermore, a good approximation to the ML estimator for α can be obtained by using $p_u(h)$ if the first-order Stirling approximation works well. It can be further shown that the second-order Laplace approximation $p_u^s(h)$ is equivalent to approximating m by the second-order Stirling approximation

$$\log\Gamma(x) \doteq (x-1/2)\log(x) + \log(2\pi)/2 - x + 1/(12x).$$

Here the term, S, in (4) is given by $S = \mathrm{diag}\{-2(\alpha^{-1}+\delta_{i+})^{-1}\}$.

Nonparametric Baseline Hazard Models

The model (1) can be directly fitted using likelihoods based on h of (2) or m of (3) if the parametric form of $\lambda_0(t)$ in (1) is specified. When the functional form of $\lambda_0(t)$ is unknown, following Breslow [17], we consider the baseline cumulative hazard function $\Lambda_0(t)$ to be a step function with jumps at the distinct observed death times,

$$\Lambda_0(t) = \sum_{k:y_{(k)}\le t}\lambda_{0k} \tag{7}$$

where $y_{(k)}$ is the kth ($k=1, \ldots, s$) earliest distinct death time among the y_{ij}'s, and $\lambda_{0k} = \lambda_0(y_{(k)})$.

Let $w = (w_1, \ldots, w_r)^T$, where $w_k = \log\lambda_{0k}$. The first term in h-likelihood of (2) can be rewritten as follows:

$$\begin{aligned} l_1(w, \beta) = \sum_{ij} l_{1ij} &= \sum_{ij}\delta_{ij}\{\log\lambda_0(y_{ij}) + \eta_{ij}\} - \sum_{ij}\{\Lambda_0(y_{ij})\exp(\eta_{ij})\} \\ &= \sum_k d_{(k)}w_k + \sum_{ij}\delta_{ij}\eta_{ij} - \sum_k \exp(w_k)\left\{\sum_{(i,j)\in R_{(k)}}\exp(\eta_{ij})\right\}, \end{aligned} \tag{8}$$

where $d_{(k)}$ is the number of deaths at $y_{(k)}$ and $R_{(k)} = R(y_{(k)}) = \{(i, j) : y_{ij} \ge y_{(k)}\}$ is the risk set at $y_{(k)}$.

Since the dimension of w increases with sample size, for estimation of (β, u) Ha et al. [11] proposed the use of profile likelihood

$$\begin{aligned} h^* &= h|_{w=\hat{w}} \\ &= \left\{\sum_k d_{(k)}\hat{w}_k + \sum_{ij}\delta_{ij}\eta_{ij} - \sum_k d_{(k)}\right\} + \sum_i l_{2i} \end{aligned} \tag{9}$$

with w eliminated, where

$$\exp(\hat{w}_k) = \frac{d_{(k)}}{\sum_{(i,j)\in R_{(k)}}\exp(x_{ij}^T\beta)u_i}$$

are solutions of the estimating equations, $\partial h/\partial w_k = 0$, for $k=1, \ldots, r$. For gamma or lognormal frailty models, h^* becomes the kernel of the penalized partial likelihood [13]. Ha et al. [11] further showed that given α the joint estimating equations for $\tau = (\beta^{\mathrm{T}}, u^{\mathrm{T}})^{\mathrm{T}}$ are obtained from

$$\partial h^*/\partial\tau = (\partial h/\partial\tau)|_{w=\hat{w}}$$

For inference of the frailty parameter α, Lee and Nelder [15] and Ha and Lee [14] have proposed the use of the adjusted profile h-likelihood, $p_\tau(h^*)$, i.e. after eliminating τ, defined by

$$p_\tau(h^*) = \left[h^* - \frac{1}{2}\log\det\left\{D/(2\pi)\right\}\right]\Big|_{\tau=\hat{\tau}}, \tag{10}$$

where $\hat{\tau} = \hat{\tau}(\theta) = (\hat{\beta}^T(\theta), \hat{u}^T(\theta))^T$, $\hat{u}$ solve $\partial h^*/\partial u = (\partial h/\partial u)|_{w=\hat{w}}$, and $D = D(h^*; \tau) = -\partial^2 h^*/\partial\tau^2$ is an information matrix for τ. Note that the first-order approximation in (10) performs well for lognormal frailty. However, for a non-lognormal frailty such as gamma frailty, Ha and Lee [6,14] have demonstrated that the second-order Laplace approximation (i.e. $p_\tau^s(h^*)$) works better.

The usual inference methods using marginal likelihood also use the assumption (7) for the baseline hazard. Since the nuisance parameters λ_{0k} are unknown, their estimates are substituted into m. That is, the estimates $\tilde{\lambda}_{0k}$ from $\partial m/\partial\lambda_{0k} = 0$ are substituted into the marginal likelihood after frailties are integrated out. For the inference the usual profile marginal likelihood based on of (3) has been often used, defined by

$$m^* = m|_{w=\hat{w}}, \tag{11}$$

where $\tilde{m}_k = \log\tilde{\lambda}_{0k}$: for example, for the gamma frailty models see Nielsen et al. [3] and Andersen et al. [7].

In this paper we are interested in the inference of the frailty parameter α in semiparametric frailty models. Since the number of nuisance parameters w increases with sample size n, the

use of m^* of (11) (hence $p^*_u(h)$ or $p^{*s}_u(h)$) can lead to the underestimation for α: see the simulation results of Table 3. Here,

$$p^*_u(h)=p_u(h)\big|_{w=\hat{w}} \text{ and } p^{*s}_u(h)=p^s_u(h)\big|_{w=\hat{w}}.$$

It is recommended to use an adjusted profile likelihood, $p_w(m)$, after eliminating w from m if m is available. We have found via simulation studies that the use of $p_w(m)$ with $w_k=\log\lambda_{0k}$ works better than in that of $p_{\lambda_0}(m)$. Theses profile likelihoods have the following relationships.

Result 1. Following Ha et al. [12], we can show that in the semiparametric gamma frailty models (1), we have

(i) $p_w(m)\simeq p_{w,u}(h)=p_u(h^*)+c$,
(ii) $p_w(m)\simeq p^s_{w,u}(h)=p^s_u(h^*)+c$,
(iii) $p_{w,\beta}(m)\simeq p_{w,\beta,u}(h)=p_{\beta,u}(h^*)+c$,
(iv) $p_{w,\beta}(m)\simeq p^s_{w,\beta,u}(h)=p^s_{\beta,u}(h^*)+c$

where $c=\frac{1}{2}\sum_k \log\{d_{(k)}/(2\pi)\}$ does not depend on the frailty parameter α.

From (i) and (ii) of Result 1 we find the following facts. The $p_w(m)$ is a proper profile likelihood, but it requires the computation of m. When m is hard to obtain the use of $p_{w,u}(h)$ or $p_u(h^*)$ is possible. Note here that the difference between $p_{w,u}(h)$ and $p_u(h^*)$ is constant [6]. However, the computation of $p_{w,u}(h)$ is also difficult because the dimension of nuisance parameters w increases with sample size. Thus, the use of $p_u(h^*)$ or $p^s_u(h^*)$ is recommended; Next, from the Result (iii) and (iv) we see that $p_{\beta,u}(h^*)$ or $p^s_{\beta,u}(h^*)$ is recommended when the number of fixed covariates is several. We investigate the performance of these profile likelihood methods by simulation below.

Real-data Examples

Numerical examples using well-known two real data set are presented to compare the various marginal- and h-likelihood methods. We consider the semiparametric gamma frailty model which gives an explicit marginal likelihood. Here, given the frailty parameter α, the marginal- and h-likelihood methods provide the same estimates for β. However, the methods give different estimators for α. The estimates of α were obtained by maximized the various profile likelihoods for α.

Example 1: *Kidney infection data* [18]. The data consist of times to the first and second recurrences of infection in 38 kidney patients using a portable dialysis machine. Infections can occur at the location of insertion of the catheter. The catheter is later removed if infection occurs and can be removed for other reasons, which we regard as censoring. Here, each survival time is time to infection since insertion of the catheter. The survival times from the same patient are likely to be related because of frailty describing the patient's effect.

Table 1. The estimated results of various likelihood methods under semiparametric gamma frailty model for the kidney infection data (m, marginal likelihood; h, h-likelihood; α, gamma frailty variance; β, regression parameter)

Method	Maximized likelihood	$\hat{\alpha}$	$\hat{\beta}$
m^*	−240.164	0.388	−1.535
$p^*_u(h)$	−240.906	0.317	−1.461
$p^{*s}_u(h)$	−240.162	0.389	−1.536
$p_u(h^*)$	−240.136	0.404	−1.549
$p^s_u(h^*)$	−239.266	0.487	−1.618
$p_w(m)$	−195.748	0.486	−1.617
$p_{\beta,u}(h^*)$	−240.082	0.436	−1.577
$p^s_{\beta,u}(h^*)$	−239.107	0.522	−1.643
$p_{\beta,w}(m)$	−195.589	0.520	−1.642

Note: $p_w(m)\simeq p^s_{w,n}(h)=p^s_u(h^*)+c$,
$p_{w,\beta}(m)\simeq p^s_{w,\beta,u}(h)=p^s_{\beta,u}(h^*)+c$ and $c=43.521$

We use a single covariate, the sex of the patients, coded as 1 for male and 2 for female. The results are given in Table 1. As expected, m^* & $p^{*s}_u(h)$, $p_w(m)$ & $p^s_u(h^*)$ and $p_{\beta,w}(m)$ & $p^s_{\beta,u}(h^*)$ give, respectively, about the same estimation results. However, we find the estimates from m^* and $p^{*s}_u(h)$ lead to smaller values than those from $p^s_u(h^*)$ and $p^s_{\beta,u}(h^*)$. These results indicates that the maximization of m^* and $p^{*s}_u(h)$ give an underestimation for both frailty and regression parameters (α, β). We claim the nuisance parameters have to be eliminated properly in order to concentrate the inference on the parameter of interest: see also simulation results of Table 3.

Example 2: *Mammary tumour data*. Gail et al. [19] presented data on multiple occurrences of mammary tumours for 48 female rats. The observations are the times to the development of a mammary tumour for 23 female rats in the treatment group and 25 female rats in the control group. Initially, 76 rats were injected with a carcinogen, and each rat was treated with retinyl acetate for the next 60 days. Some 48 rats were tumour-free after 60 days. These rats were randomly assigned to continued retinoid prophylaxis or to the control group, where they received no treatment. Rats were palpated for tumours twice weekly and observed for 122 days. The main objective of the study was to evaluate treatment. The time origin is the day of the initial carcinogen injection. The survival time T_{ij}(j=1, ..., n_i) is then calculated as $t_{i,j}-t_{i,j-1}$, where $t_{i,j}$ with $t_{i,0}$=0 is the jth tumour occurrence time of the ith rat, which is the inter-arrival (gap) time between the tumour recurrences. Here the cluster sizes n_i range from 1 to 14. Censoring (approximately 17%)

Table 2. The estimated results of various likelihood methods under semiparametric gamma frailty models for the mammary tumour data (m, marginal likelihood; h, h-likelihood; α, gamma frailty variance; β, regression parameter)

Method	Maximized likelihood	$\hat{\alpha}$	$\hat{\beta}$
m^*	−1180.620	0.247	−0.822
$p_u^*(h)$	−1181.152	0.222	−0.818
$p_u^{*s}(h)$	−1180.619	0.247	−0.822
$p_u(h^*)$	−1180.539	0.254	−0.823
$p_u^s(h^*)$	−1179.958	0.282	−0.827
$p_w(m)$	−1153.494	0.283	−0.827
$p_{\beta,u}(h^*)$	−1181.147	0.273	−0.825
$p_{\beta,u}^s(h^*)$	−1180.541	0.301	−0.829
$p_{\beta,w}(m)$	−1154.077	0.302	−0.829

Note: $p_w(m) \simeq p_{w,n}^s(h) = p_u^s(h^*) + c$,
$p_{w,\beta}(m) \simeq p_{w,\beta,u}^s(h) = p_{\beta,u}^s(h^*) + c$ and c=26.463

occurred when no new tumour was found. We also use a single fixed covariate x_{ij} (=1 for treatment and for control). We find that the results in Table 2 are very similar to those evident in Example 1.

Simulation Study

Numerical studies, using 200 replications of simulated data, are presented to investigate the performances of the likelihood methods based on marginal- and h-likelihood. We again consider the semiparmetric gamma frailty model.

Under no censoring we generate data assuming the exponential baseline hazard $\lambda_0(t)=1$, one standard normal covariate with $\beta=1$, and $\alpha=0.5$, 1.0. We consider $n=\sum_{i=1}^{q} n_i$ with $n=$ 100, 200 and $(q, n_i)=(100, 1)$, (100, 2), (200, 1). Notice here that we chose fairly extreme cases, the no censoring and small sample size, because these situations yielded the most biased estimates of $\hat{\alpha}$ in the simulation studies by Nielsen et al. [3] and Barker and Hendersion [9]. Moreover, if satisfactory results could be obtained for these, good results would follow more generally. For the fitting we used the four comparable likelihoods, $p_u^{*s}(h)$, $p_u^s(h^*)$, and $p_{\beta,u}^s(h^*)$. From 200 replications of simulated data we compute the mean, standard deviation and mean squared error for $\hat{\beta}$ and $\hat{\alpha}$. For the computation we used SAS/IML.

The results are summarized in Table 3. Overall, these results confirm those from numerical examples in Section 5. As expected, the bias increases with frailty and decreasing sample size. The estimates from m^* and $p_u^{*s}(h)$ show about the same results, but they gives severely downward biases in all cases considered, especially in or frailty parameters α. Moreover, the underestimation of α leads to that of β. Table 3 also demonstrates that the two profile likelihood methods, $p_u^s(h^*)$ and $p_{\beta,u}^s(h^*)$, eliminating the nuisance parameters w reduce effectively such biases. Here the $p_u^s(h^*)$ method slightly performs better than in the $p_{\beta,u}^s(h^*)$ method. This may be a result due to the consideration of a single covariate or small frailty variance.

Comparison of h-likelihood and Bayesian Approach

In Bayesian framework fixed parameters (β, α) are treated as random variables, so that they require a prior (distribution) $\pi(\beta, \alpha)$. Here we assume its independence, $\pi(\beta, \alpha)=\pi(\beta) \cdot \pi(\alpha)$.

Applying the Bayes theorem, the joint posterior density for semiparametric frailty models (1) with nonparametric baseline hazards is proportional to [20]:

$$\pi(\beta, u, \alpha | y^*) \propto \pi(y^* | u, \beta) \cdot \pi(u | \alpha) \cdot \pi(\beta) \cdot \pi(\alpha).$$

Note that $f_\beta(y^*|u)=\pi(y^*|u, \beta)$ and $f_\alpha(u)=\pi(u|\alpha)$. From (2) and (9) we see that $f_\beta(y^*|u)$ is a conditional partial likelihood eliminating λ_0 given u, leading to $f_\beta(y^*|u) \cdot f_\alpha(u)=\exp(h^*)$. Thus we have

$$\log\{\pi(\beta, u, \alpha | y^*)\} \propto h^* + \log\pi(\beta) + \log\pi(\alpha).$$

In particular, for the Bayesian inference Ducrocq & Casella [21] and Legrand et al. [20] assumed uniform priors (i.e. flat priors), $\pi(\beta)=1$ and $\pi(\alpha)=1$. Under the assumptions, we also see that the log-joint posterior $\log\{\pi(\beta, u, \alpha|y^*)\}$ is equivalent to the profile h-likelihood h^* in (9): see also Rigby & Stasinopoulos [22] and Abrahantes et al. [23]. For the estimation of β and u, the log-joint posterior is maximized over (β, u) given α [20].

Furthermore, for the estimation of frailty parameters α Legrand et al. [20] used the marginal posterior of α, given by

$$\pi(\alpha | y^*) = \int\int \pi(\beta, u, \alpha | y^*)\, d\beta du \propto \int\int \exp(h^*)\, d\beta du,$$

whose Laplace approximation becomes the adjusted profile h-likelihood, $\exp\{p_{\beta,u}(h^*)\}$) in (10): see also Rigby & Stasinopoulos [22] and Rue et al. [24].

In summary, under the uniform prior we have that

$$\log\{\pi(\beta, u, \alpha | y^*)\} \propto h^* \text{ and } \log\{\pi(\alpha | y^*)\} \simeq p_{\beta,u}(h^*). \quad (12)$$

Hence, from (12) we confirm that in frailty models, as in HGLMs [25], the h-likelihood method also gives an approxi-

Table 3. Simulation results about the estimators $\hat{\alpha}$ and $\hat{\beta}$ using marginal likelihood (m) and h-likelihood (h) methods under semiparametric gamma frailty models. The simulation is conducted with 200 replications at each gamma frailty variance with true regression parameter $\beta=1$ (No censoring)

α	q	n_i	Method	$\hat{\alpha}$			$\hat{\beta}$		
				Mean	SD	MSE	Mean	SD	MSE
0.5	100	1	m^*	0.246	0.241	0.1221	0.891	0.203	0.0531
			$p^{*s}_u(h)$	0.248	0.243	0.1224	0.892	0.204	0.0533
			$p^{s}_u(h^*)$	0.510	0.392	0.1532	1.021	0.261	0.0682
			$p^{s}_{\beta,u}(h^*)$	0.584	0.418	0.1812	1.054	0.264	0.0724
	100	2	m^*	0.432	0.182	0.0376	0.977	0.235	0.0553
			$p^{*s}_u(h)$	0.431	0.182	0.0377	0.977	0.235	0.0553
			$p^{s}_u(h^*)$	0.484	0.202	0.0410	0.994	0.240	0.0575
			$p^{s}_{\beta,u}(h^*)$	0.509	0.194	0.0375	1.005	0.239	0.0570
	200	1	m^*	0.320	0.227	0.0838	0.921	0.166	0.0337
			$p^{*s}_u(h)$	0.322	0.229	0.0841	0.922	0.167	0.0338
			$p^{s}_u(h^*)$	0.487	0.297	0.0879	1.001	0.179	0.0320
			$p^{s}_{\beta,u}(h^*)$	0.531	0.315	0.0994	1.023	0.187	0.0352
1.0	100	1	m^*	0.419	0.363	0.4694	0.795	0.228	0.0935
			$p^{*s}_u(h)$	0.423	0.368	0.4679	0.797	0.229	0.0935
			$p^{s}_u(h^*)$	0.885	0.617	0.3925	0.958	0.296	0.0891
			$p^{s}_{\beta,u}(h^*)$	1.090	0.685	0.4753	0.962	0.327	0.1077
	100	2	m^*	0.902	0.240	0.0667	0.978	0.281	0.0790
			$p^{*s}_u(h)$	0.902	0.240	0.0669	0.978	0.281	0.0790
			$p^{s}_u(h^*)$	0.994	0.258	0.0662	0.997	0.286	0.0813
			$p^{s}_{\beta,u}(h^*)$	1.015	0.260	0.0673	1.002	0.287	0.0817
	200	1	m^*	0.625	0.301	0.2306	0.865	0.188	0.0535
			$p^{*s}_u(h)$	0.632	0.306	0.2280	0.868	0.189	0.0531
			$p^{s}_u(h^*)$	0.992	0.439	0.1922	0.996	0.231	0.0531
			$p^{s}_{\beta,u}(h^*)$	1.068	0.424	0.1838	1.024	0.208	0.0435

Note: q, No. of clusters; n_i, cluster size; SD, standard deviation; MSE, mean squared error.

mation of the Bayesian inference under uniform prior. Notice, however, that the Bayesian eliminates parameters by integration, while h-likelihood them by conditioning (or profiling) [25].

Discussion

We have studied various inference methods for semiparametric frailty models. We have showed that the h-likelihood approach is useful in profiling properly the nuisance parameters, leading to reduce the bias of maximum likelihood estimator from standard marginal likelihood. For complex frailty models such as multi-component [5,26] or correlated frailties [27], the h-likelihood can be easily extended because it avoids the integration itself, whereas the marginal likelihood requires intractable integrations. In particular, the h-likelihood approach is very useful to inference (e.g. prediction of heterogeneity) of random effect (frailty) [25,27,28], but the marginal likelihood can not be used directly for such inference because it eliminates them by integration.

Furthermore, we have showed that under uniform prior (i.e. flat prior) h-likelihood and Bayesian approaches lead to very similar estimation results. However, the results of both approaches may depend on the choice of priors for fixed parameters (β, α). It would be very interested to study the sensitivity analysis to prior specification, particularly for frailty parameter α [29].

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0021165).

References

1. Hougaard P. Analysis of Multivariate Survival Data. New York:

Springer; 2000.
2. Duchateau L, Janssen P. The frailty model. New York: Springer; 2008.
3. Nielsen GG, Gill RD, Andersen PK, Sørensen TI. A counting process approach to maximum likelihood estimation in frailty models. Scand J Stat 1992;19;25-44.
4. Vaida F, Xu R. Proportional hazards model with random effects. Stat Med 2000;19:3309-3324.
5. Yau KKW. Multilevel models for survival analysis with random effects. Biometrics 2001;57:96-102.
6. Ha ID, Lee Y. Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. Biometrika 2005;92:717-723.
7. Andersen PK, Klein JP, Knudsen K, Palacios RT. Estimation of variance in Cox's regression model with shared gamma frailties. Biometrics 1997;53:1475-84.
8. Rondeau V, Commenges D, Joly P. Maximum penalized likelihood estimation in a gamma frailty model. Lifetime Data Anal 2003;9:139-153.
9. Barker P, Henderson R. Small sample bias in the gamma frailty model for univariate survival. Lifetime Data Anal 2005;11:265-284.
10. Lee Y, Nelder JA. Hierarchical generalized linear models (with discussion). J R Stat Soc B 1996;58:619-678.
11. Ha ID, Lee Y, Song JK. Hierarchical likelihood approach for frailty models. Biometrika 2001;88:233-243.
12. Ha ID, Noh M, Lee Y. Bias reduction of likelihood estimators in semiparametric frailty models. Scand J Stat 2010;37:307-320.
13. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics 2000;56:1016-1022.
14. Ha ID, Lee Y. Estimating frailty models via Poisson hierarchical generalized linear models. J Comput Graph Stat 2003;12:663-681.
15. Lee Y, Nelder JA. Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika 2001;88:987-1006.
16. Lee Y, Nelder JA, Pawitan Y. Generalised linear models with random effects: unified analysis via h-likelihood. London: Chapman and Hall; 2006.
17. Breslow NE. Discussion of Professor Cox's paper. J R Stat Soc B 1972;34:216-217.
18. McGilchrist, CA, Aisbett CW. Regression with frailty in survival analysis. Biometrics 1991;47:461-466.
19. Gail MH, Santner TJ, Brown CC. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. Biometrics 1980;36:255-266.
20. Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. Stat Med 2005;24:3789-3804.
21. Ducrocq V, Casella G. A Bayesian analysis of mixed survival models. Genet Sel Evol 1996;28:505-529.
22. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc C 2005;54:507-554.
23. Abrahantes JC, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. Comparison of different estimation procedures for proportional hazards model with random effects. Comput Stat Data An 2007;51:3913-3930.
24. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). J R Stat Soc B 2009;71:319-392.
25. Lee Y, Nelder JA. Likelihood inference for models with unobservables: another view (with discussion). Stat Sci 2009;24:255-293.
26. Ha ID, Lee Y, MacKenzie G. Model selection for multi-component frailty models. Stat Med 2007;26:4790-4807.
27. Ha ID, Sylvester R, Legrand C, MacKenzie G. Frailty modelling for survival data from multi-centre clinical trials. Stat Med 2011; 30:2144-2159.
28. Lee Y, Ha ID. Orthodox BLUP versus h-likelihood methods for inferences about random effects in Tweedie mixed models. Stat and Comput 2010;20:295-303.
29. Jang MJ, Lee Y, Lawson AB, Browne WJ. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling. Environmetrics 2007;18:809-821.