Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 3, pp. 999 ~ 1007

On Profile Likelihood for Gamma Frailty Models

Il Do Ha[1)]

Abstract

The semiparametric gamma frailty models have been often used for multivariate survival analysis because they give an explicit marginal likelihood. The commonly used estimation procedure is the profile likelihood method based on marginal likelihood, which provides the same parameter estimates as the EM algorithm. In this paper we show in finite samples the standard profile-likelihood method can lead to an underestimation of parameters, particularly for the frailty parameter. To overcome this problem, we propose an adjusted profile-likelihood method. For the illustration a numerical example and a small-sample simulation study are presented.

Keywords : Adjusted profile likelihood, Gamma frailty models, Marginal likelihood, Profile likelihood

1. Introduction

Frailty models, extensions of Cox's proportional hazards models, have been widely used for the analysis of various correlated and/or heterogeneous event-time data, for example, from the study of biomedicine (Hougaard, 2000) or econometrics (Horowitz, 1999; Clapp et al., 2006). For the inferences many authors have proposed several likelihood-based methods. The marginal likelihood (i.e., observed data likelihood), which is obtained by integrating out the frailties, has been often used. In particular, gamma frailty models give an explicit marginal likelihood (Nielsen et al., 1992; Andersen et al., 1997).

In this paper we focus on the estimation of frailty parameter in the semiparametric gamma frailty models allowing unspecified baseline hazard. Note here that the handling of the nuisance parameters in the baseline hazard is a main

1) Associate Professor, Department of Asset Management, Daegu Haany University, Gyeongsan, 712-715, Korea.
E-mail : idha@dhu.ac.kr

issue. The usual solution is to profile out the nuisance parameters (Murphy and van der Vaart, 2000). The profile likelihood method based on marginal likelihood has been usually used, and it leads to the same parameter estimates as the EM algorithm (Nielsen et al., 1992; Andersen et al., 1997; Murphy and van der Vaart, 2000). For the elimination of nuisance parameters, these methods all use the types of the discrete nonparametric Breslow (1972) estimates.

Recently, in semiparametric gamma frailty models Rondeau et al. (2003) and Baker and Henderson (2005) have numerically showed that the use of such Breslow estimates in the EM method can lead to finite sample underestimation of parameters, particularly for frailty parameters. For the reduction of such bias they have proposed the use of the continuous nonparametric estimate, instead of the Breslow estimate, for the baseline hazard.

The bias problem may also occur because the number of nuisance parameters in the baseline hazard increases with sample size. That is, under this situation the uncertainties in the nuisance parameter estimation should be considered in estimating the frailty parameter. However, there is no such consideration in the standard profile likelihood, leading to a downward bias, especially for the frailty parameter (Ha and Lee, 2005). This problem can be solved by an appropriate modification (e.g. an adjusted profile likelihood) for the standard profile likelihood (Severini, 1998; Lee and Nelder, 1996, 2001). Thus, in this paper we propose a new method to improve the standard profile-likelihood method. Notice here that we still use the Breslow estimate.

The paper is organized as follows. In Section 2 we review the standard profile marginal-likelihood method for the gamma frailty models, and then propose a new profile-likelihood method. In Section 3 the proposed method is illustrated by using a numerical example based on a well-known real data set and a small-sample simulation study. Finally, some remarks are given in Section 4.

2. Profile Likelihoods for Frailty Models

Let T_{ij} be the survival time for the jth observation on the ith subject ($i=1,\cdots,q$; $j=1,\cdots,n_i$; $n=\sum_i n_i$) and C_{ij} be the corresponding censoring time. Let the observable random variables be $y_{ij}=\min(T_{ij},C_{ij})$ and $\delta_{ij}=I(T_{ij}\le C_{ij})$, where $I(\cdot)$ is the indicator function. Denoted by U_i the unobserved frailty(random effect) for the ith subject.

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

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

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function and β is a $p\times 1$ vector of fixed effects associated with fixed covariate vector $x_{ij}=(x_{ij1},\cdots,x_{ijp})^T$. Here, the frailties U_i's are assumed to be independent and identically distributed from a gamma distribution with mean $E(U_i)=0$ and $\mathrm{var}(U_i)=\alpha$ (Nielsen et al., 1992). For the U_i other frailty distributions such as lognormal can be assumed (McGilchrist and Aisbett, 1991; Hougaard, 2000; Ha et al., 2001).

2.1 Standard Profile Likelihood

For the inference the marginal likelihood, denoted by m, has been often used (e.g., Nielsen et al., 1992; Andersen et al., 1997); it can be obtained by integrating out the frailties from the hierarchical likelihood (h-likelihood, Lee and Nelder, 1996; Ha et al., 2001):

$$m=m(\lambda_0,\beta,\alpha)=\sum_i \log\left\{\int \exp(h_i)\,du_i\right\}, \tag{2}$$

where $h_i=\sum_j \ell_{1ij}+\ell_{2i}$, $\ell_{1ij}=\ell_{1ij}(\lambda_0,\beta;y_{ij}|u_i)$ is the logarithm of the conditional density function for (y_{ij},δ_{ij}) given $U_i=u_i$, and $\ell_{2i}=\ell_{2i}(\alpha;u_i)$ is the logarithm of the density function for U_i with parameter α.

For the gamma frailty models (1), we have from (2) an explicit marginal likelihood:

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

(3)

where $\delta_{i+}=\sum_j\delta_{ij}$, $\mu_{i+}=\sum_j\mu_{ij}=\sum_j\Lambda_0(y_{ij})\exp(x_{ij}^T\beta)$, $\Lambda_0(t)=\int_{-\infty}^{t}\lambda_0(k)\,dk$ is the baseline cumulative hazard function and $f(\alpha)=\log\Gamma(\alpha^{-1})+\alpha^{-1}\log\alpha$.

The model (1) can be directly fitted using m of (3) if the parametric form(e.g. Weibull) for the baseline hazard $\lambda_0(t)$ in (1) is specified (Ha and Lee, 2003). However, in this paper the functional form of the $\lambda_0(t)$ is unknown.

For the model (1) the usual profile-likelihood method based on marginal likelihood is as follows. Following Breslow (1972), many authors (e.g. Andersen et al., 1997; Ha et al., 2001) considered the baseline cumulative hazard function $\Lambda_0(t)$ to be a step function with jumps at the s distinct observed death times,

$$\Lambda_0(t) = \sum_{k: y_{(k)} \le t} \lambda_{0k}, \tag{4}$$

where $y_{(k)}$ is the kth ($k = 1, \cdots, s$) smallest distinct death time among the y_{ij}'s, and $\lambda_{0k} = \lambda_0(y_{(k)})$. Under the assumption (4), the estimates $\widehat{\lambda_{0k}}$ from $\partial m / \partial \lambda_{0k} = 0$ are substituted into the marginal likelihood (3). That is, the standard profile likelihood eliminating the nuisance parameters λ_0, denoted by PL1, based on m of (3) is defined by

$$\text{PL1}(\beta, \alpha) = m|_{\lambda_0 = \widehat{\lambda_0}}, \tag{5}$$

where $\lambda_0 = (\lambda_{01}, \cdots, \lambda_{0s})^T$ and $\widehat{\lambda_0} = (\widehat{\lambda_{01}}, \cdots, \widehat{\lambda_{0s}})^T$ with

$$\widehat{\lambda_{0k}} = \frac{d_{(k)}}{\sum_{(i,j) \in R(y_{(k)})} \exp(x_{ij}^T \beta)\, \widehat{u_i}} \quad \text{and} \quad \widehat{u_i} = \frac{\alpha^{-1} + \delta_{i+}}{\alpha^{-1} + \mu_{i+}}.$$

Here, $R(y_{(k)}) = \{(i,j): y_{ij} \ge y_{(k)}\}$ is the risk set at $y_{(k)}$. In particular, the maximization of profile likelihood (PL1) gives the same parameter estimates for (β, α) as the EM algorithm (Nielsen et al., 1992; Murphy and van der Vaart, 2000).

2.2 Proposed Profile Likelihood

In this paper we focus on the estimation of the frailty parameter α. Note that the number of nuisance parameters λ_0 in (1) increases with sample size n. Under this situation, the uncertainties in the nuisance parameter estimation should be considered in estimating the α. However, there is no such consideration in PL1 of (5), leading to a severly downward bias for α (Ha and Lee, 2005). To overcome this problem, an appropriate modification for PL1 would be useful.

Following Lee and Nelder (2001), it is recommended to use an adjusted profile likelihood. Let ℓ be a likelihood, for example, a marginal likelihood m of (2). Lee and Nelder (2001) considered a function $p_\Theta(\ell)$, defined by

$$p_\Theta(\ell) = [\ell - \frac{1}{2} \log \det\{D(\ell, \Theta)/(2\pi)\}]|_{\Theta = \widehat{\Theta}}, \tag{6}$$

where $D(\ell, \Theta) = -\partial^2 \ell / \partial \Theta^2$ and $\widehat{\Theta}$ solves $\partial \ell / \partial \Theta = 0$. The function $p_\Theta(\cdot)$ in (6) produces an adjusted profile likelihood, eliminating nuisance parameters Θ, which can be fixed effects β or λ_0, or random effects $u = (u_1, \cdots, u_q)^T$. The second term on the right-hand side of (6) can be interpreted as a penalty term, which subtracts from the ordinary profile likelihood the undeserved information on the nuisance parameters θ; for the further details see Lee and Nelder (2001) and Pawitan (2001).

From (6) we have found via simulation studies that the use of $p_w(m)$, not $p_{\lambda_0}(m)$, performs well. Here $\omega=(\omega_1,\cdots,\omega_s)^T$ with $\omega_k=\log\lambda_{0k}$. In particular, we have experienced that the estimate of α cannot be obtained by maximizing $p_{\lambda_0}(m)$ because the $p_{\lambda_0}(m)$ is indeed increasing with increasing values of α. Thus, the proposed profile marginal likelihood, denoted by PL2, is defined by

$$\mathrm{PL2}(\beta,\alpha)=p_w(m)=[m-\frac{1}{2}\log\det\{D(m,\omega)/(2\pi)\}]|_{\omega=\hat{\omega}}, \quad (7)$$

where $D(m,\omega)=-\partial^2 m/\partial\omega^2$ and $\hat{\omega}$ solves $\partial m/\partial\omega=0$. Note here that the solutions of w_k's are given by $\widehat{\omega_k}=\log\widehat{\lambda_{0k}}$ with $\widehat{\lambda_{0k}}$ in (5).

3. Illustration

For the illustration, we compare the performance of the proposed PL2 method in (7) with that of the standard PL1 method in (5). Here we present a numerical example and a simulation study. We have found that the PL1 performs well for the estimation of β given α. Thus, for the estimation of β given α we suggest the PL2 method also uses the PL1. Note that, given α, both methods (PL1 and PL2) provide the same estimates for β, but that they give different estimates for α. As a result, both methods give different estimates for (β,α).

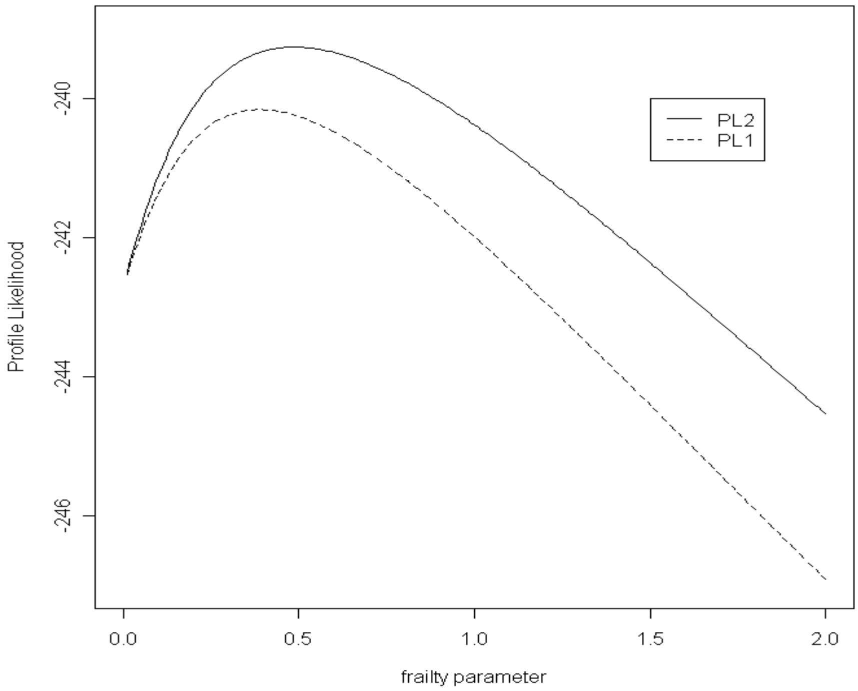
3.1 Numerical Example

The kidney infection data set from McGilchrist and Aisbett (1991) consists 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 correlated because of frailty describing the patient's effect. We use a single covariate, the sex of the patients, coded as 1 for male and 2 for female. The estimation results on the model (1) are given in Table 1.

<Table 1> The estimation results using the two profile likelihoods in the kidney infection data

Method	$\hat{\alpha}$	$\hat{\beta}$
PL1(Standard)	0.39	-1.54
PL2(Proposed)	0.49	-1.62

Note: $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of frailty parameter and regression parameter, respectively.



<Figure 1> Profile likelihoods (PL1 and PL2) for frailty parameter α

As expected, the absolute magnitude of estimates from the PL1 is smaller than that from the PL2. These results indicate that the maximization of PL1 may give an underestimation for frailty and regression parameters (α, β). In addition, the two profile likelihoods (PL1 and PL2) are computed at each α and the results are plotted in Figure 1. We again see that the two profile likelihoods for α are maximized at the different values of α, leading to $\hat{\alpha}=0.39$ for PL1 and $\hat{\alpha}=0.49$ for PL2. Thus, we claim the nuisance parameters have to be eliminated properly in order to concentrate the inference on the parameter of interest.

3.2 Simulation Study

Simulation studies, using 200 replications of simulated data, are presented to investigate the performances of the proposed PL2 method over the standard PL1. Under the model (1) we generate data assuming the exponential baseline hazard $\lambda_0(t)=1$, one binary covariate x_{ij} with $\beta=1$, and $\alpha=0.5, 1.0$. In particular, we set x_{ij} to 0 for the first $q/2$ subjects, to form the control group, and x_{ij} to 1 for the remaining $q/2$, to form the treatment group: see also Ha et al. (2001). We consider the two small samples as in 50 pairs and 100 pairs, i.e., $n=\sum_{i=1}^{q} n_i=100, 200$ with $(q, n_i)=(50,2), (100,2)$. Here, we chose the no-censoring case because such situation yielded larger biased estimates for α than in censoring case: see for example the simulation results by Nielsen et al. (1992). Moreover, if satisfactory results could be obtained for these, good results would follow more generally. From 200 replications of simulated data we compute the mean, standard deviation (SD) and mean squared error (MSE) for $\hat{\alpha}$ and $\hat{\beta}$ For the computation we used SAS/IML.

<Table 2> Simulation results on $\hat{\alpha}$ and $\hat{\beta}$ using the two profile likelihoods

α	n	Method	$\hat{\alpha}$ Mean	$\hat{\alpha}$ SD	$\hat{\alpha}$ MSE	$\hat{\beta}$ Mean	$\hat{\beta}$ SD	$\hat{\beta}$ MSE
0.5	100	PL1	0.34	0.238	0.084	0.94	0.293	0.089
		PL2	0.44	0.270	0.077	0.98	0.306	0.093
	200	PL1	0.43	0.182	0.038	0.98	0.235	0.055
		PL2	0.49	0.194	0.038	1.00	0.239	0.057
1.0	100	PL1	0.80	0.284	0.119	0.95	0.381	0.147
		PL2	0.96	0.327	0.108	0.98	0.395	0.156
	200	PL1	0.90	0.240	0.067	0.98	0.281	0.079
		PL2	0.99	0.258	0.066	1.00	0.286	0.081

Note: The simulation is conducted with 200 replications at each gamma frailty variance α and true regression parameter $\beta=1$.

The results are summarized in Table 2. Overall, these results confirm those of Table 1 and Figure 1. As expected, the bias increases with frailty and decreasing sample size. The PL1 method shows a slight bias for regression parameter β, but leads to severely downward biases, in all cases considered, for frailty parameter α. The proposed PL2 method removes noticeably such biases. We observe that the PL2 method has larger SD than the PL1 method, caused by the underestimation of the PL1. The results of Table 2 demonstrate that the proposed PL2 method reduces effectively such biases from the PL1.

4. Remarks

For the finite samples on the model (1) we have numerically showed that the proposed method (PL2) works well and improves largely the standard method (PL1). Thus, care is necessary in using the profile likelihood when the number of nuisance parameters is large: see also Ha and Lee (2005). The further theoretical work is required on the PL2, for example, about the formal proof of parameter orthogonality of ω and β in the sense of Cox and Reid (1987) and the asymptotic justification as in Parner (1998).

The PL2 method needs a marginal likelihood, which usually has no explicit form; the marginal likelihood often requires an intractable integration, for example, for lognormal frailty mode1s. Thus, the use of PL2 can be restricted. For such cases, an adjusted profile likelihood based on h-likelihood (Ha et al., 2001; Ha and Lee, 2005) would be very useful.

References

1. Andersen, P. K., Klein, J. P., Knudsen, K. and Palacious, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties, *Biometrics*, 53, 1475-1484.
2. Barker, P. and Henderson, R. (2005). Small sample bias in the gamma frailty model for univariate survival, *Lifetime Data Analysis*, 11, 265-284.
3. Breslow, N. E. (1972). Discussion of Professor Cox's paper. *Journal of the Royal Statistical Society B*, 34, 216-217.
4. Clapp, J. M., Dend, Y. and An, X. (2006). Unobserved heterogeneity in models of competing mortgage termination risks, *Real Estate Economics*, 34, 243-273.
5. Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society B*, 49, 1-39.
6. Ha, I. D. and Lee, Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models, *Journal of Computational and Graphical Statistics*, 12, 663-681.
7. Ha, I. D. and Lee, Y. (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models, *Biometrika*, 92, 717-723.
8. Ha, I. D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models, *Biometrika,* 88, 233-243.
9. Horowitz, J. L. (1999). Semiparametric estimation of a proportional

hazard model with unobserved heterogeneity, *Econometrica*, 67, 1001-1028.

10. Hougaard, P. (2000). *Analysis of multivariate survival data,* New York: Springer-Verlag.
11. Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society B,* 58, 619-678.
12. Lee, Y and Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions, *Biometrika,* 88, 987-1006.
13. McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis, *Biometrics*, 47, 461-466.
14. Murphy and van der Vaart, A. W. (2000). On profile likelihood, *Journal of the American Statistical Association*, 95, 449-465.
15. Nielsen, G.G., Gill, R.D., Andersen, P.K. and Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models, *Scandinavian Journal of Statistics,* 19, 25-44.
16. Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model, *The Annals of Statistics*, 26, 183-214.
17. Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood,* Oxford: Clarendon Press.
18. Rondeau, V., Commenges, D. and Joly, P. (2003). Maximum penalized likelihood estimation in a gamma frailty model, *Lifetime Data Analysis*, 9, 139-153.
19. Severini, T. A. (1998). An approximation of the modified profile likelihood function, *Biometrika*, 85, 403-411.

[received date : Jun. 2006, accepted date : Aug. 2006]