Journal of the Korean Data & Information Science Society
2018, **29**(6), 1679–1685

http://dx.doi.org/10.7465/jkdi.2018.29.6.1679
한국데이터정보과학회지

Laplace approximation approach for frailty survival models†

Il Do Ha[1] · Geon Ho Cho[2]

[1]Department of Statistics, Pukyong National University
[2]Division of Cosmetic Science and Technology, Daegu Haany University

Received 17 October 2018, revised 17 November 2018, accepted 20 November 2018

Abstract

For multivariate or correlated survival data semi-parametric frailty models with nonparametric baseline hazards, extensions of Cox's (1972) proportional hazards models, has been often used. The marginal likelihood has been usually used for the inferences, but it often requires the computation of difficult integration in integrating out the frailty terms. In this paper we propose a Laplace approximation approach based on hierarchical likelihood for the frailty models. The proposed method is demonstrated via simulation study and two well-known real data sets. In particular, the simulation results show that our method is better than standard h-likelihood method in terms of bias.

Keywords: Frailty models, hierarchical likelihood, Laplace approximation, multivariate survival data.

1. Introduction

Various estimation methods have been developed for the inference of semi-parametric frailty models with nonparametric baseline hazards, i.e. Cox's proportional hazards models (Lee and Lee, 2017) allowing for random effects (Ha *et al.*, 2017). It is well known that the likelihood-based estimation methods are useful. For example, the maximum likelihood (ML) method is popular, but it requires the computation of a marginal likelihood (i.e. observed likelihood); it often involves an intractable integral in integrating out the frailty terms (i.e. random effects). The marginal likelihood generally does not give an explicit form except for gamma frailty model, leading to an numerical integration. For example, log-normal frailty model with normally distributed random effect does not provide a closed form for the marginal likelihood. Thus, Ha *et al.* (2001) proposed the use of the hierarchical likelihood

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2017R1E1A1A03070747).

[1] Corresponding author: Professor, Department of Statistics, Pukyong National University, Busan 608-737, Korea. E-mail: idha1353@pknu.ac.kr

[2] Professor, Division of Cosmetic Science and Technology, Industrial Quality Engineering, Daegu Haany University, Gyeongsan 712-715, Korea

(h-likelihood; Lee and Nelder, 1996) which avoids the integration itself. The h-likelihood provides a statistically efficient procedure for various random-effect models including generalized linear mixed models (GLMMs), hierarchical generalized linear models (HGLMs) and survival random-effect models such as frailty models and mixed-effects survival models (Lee *et al.*, 2017; Ha and Noh, 2013; Ha and Cho, 2015; Kim *et al.*, 2016; Ha *et al.*, 2017).

The h-likelihood performs well the semi-parametric frailty models under moderate censoring, but we need to study the performance for the high-censoring. For this purpose, we propose a Laplace approximation (LA) method based on the h-likelihood under the semi-parametric frailty models. It is also well known that the LA method works well as the cluster size increases, but that it may lead to a serious bias when cluster size is small (Lee *et al.*, 2001). The higher-order approximation generally reduces such bias (Lee *et al.*, 2017).

The paper is organized as follows. In Section 2 we describe the data structures and the h-likelihood for the frailty models. In Section 3 we propose the LA estimator based on the h-likelihood. In Section 4 we present a simulation study to evaluate the performance the proposed method. The proposed method is illustrated using two well-known data examples in Section 5. Finally, some discussions are given in Section 6.

2. H-likelihood approach for frailty models

Suppose that data consist of censored time-to-event (survival time) observations from q subjects (or clusters), with cluster size n_i, $i = 1, \dots, q$. Let T_{ij} ($i = 1, \dots, q$; $j = 1, \dots, n_i$; $n = \sum_i n_i$) be the survival time for the jth observation of the ith subject. Here, q is the number of subjects and n is the total sample size. Let U_i be an unobserved frailty (random effect) of the $i$the subject. Given the unobserved frailty for the ith subject $U_i = u_i$, the conditional hazard function of T_{ij} is of the form

$$\lambda_{ij}(t|u_i) = \lambda_0(t)\exp(x_{ij}^T\beta)u_i, \tag{2.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ is a vector of fixed covariates, and $\beta = (\beta_1, \dots, \beta_p)^T$ is a vector of the corresponding regression parameters. For identifiability purposes, the term $x_{ij}^T\beta$ in the model (2.1) does not include an intercept term. Suppose that the frailties U_i are independent and identically distributed random variables with frailty parameter α; the gamma and lognormal frailty models assume gamma and lognormal distributions for U_i, respectively. In this paper we are interested in lognormal frailty model with log-frailty $V_i = \log U_i \sim N(0, \alpha)$; its marginal likelihood dose not give a closed form. Let C_{ij} be censoring time corresponding to T_{ij}. Then the observable random variables be

$$Y_{ij} = \min(T_{ij}, C_{ij}) \text{ and } \delta_{ij} = I(T_{ij} \le C_{ij}),$$

where $I(\cdot)$ is the indicator function. In this paper we assume the non-informative censoring given frailty u_i (Ha *et al.*, 2001, 2017).

The functional form of $\lambda_0(t)$ in the model (2.1) is unknown. For the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(k)dk$, Breslow (1972) proposed the use of a step function with jumps at the observed event times, i.e.

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

where $y_{(k)}$ is the kth ($k = 1, \ldots, r$) smallest distinct event (e.g. death) time among Y_{ij}, and $\lambda_{0k} = \lambda_0(y_{(k)})$. Along the lines of Lee and Nelder (1996), the h-likelihood (Ha *et al.*, 2001, 2017) for the model (2.1) is defined as

$$h = h(\beta, \lambda_0, v, \alpha) = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}, \tag{2.2}$$

where

$$\begin{aligned}\sum_{ij} \ell_{1ij} &= \sum_{ij} \delta_{ij}\{\log \lambda_0(y_{ij}) + \eta_{ij}\} - \sum_{ij}\{\Lambda_0(y_{ij})\exp(\eta_{ij})\} \\ &= \sum_k d_{(k)} \log \lambda_{0k} + \sum_{ij} \delta_{ij}\eta_{ij} - \sum_k \lambda_{0k}\Big\{\sum_{(i,j)\in R_{(k)}} \exp(\eta_{ij})\Big\},\end{aligned}$$

$\ell_{1ij} = \ell_{1ij}(\beta, \lambda_0; y_{ij}, \delta_{ij}|u_i)$ is the logarithm of the conditional density function for Y_{ij} and δ_{ij} given $U_i = u_i$, $\ell_{2i} = \ell_{2i}(\alpha; v_i)$ is the logarithm of the density function for $V_i = \log(U_i)$ with parameter α, and the linear predictor on the log-hazard

$$\eta_{ij} = x_{ij}^T\beta + v_i$$

with $v_i = \log(u_i)$. Here $v = (v_1, \ldots, v_q)^T$, $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0r})^T$, $d_{(k)}$ is the number of events at $y_{(k)}$ and $R_{(k)} = R(y_{(k)}) = \{(i,j) : y_{ij} \geq y_{(k)}\}$ is the risk set at $y_{(k)}$. In (2.2), the loglikelihood of the ith subject is the logarithm of the joint density of (Y_i, δ_i, V_i), where $Y_i = (Y_{i1}, \ldots, Y_{in_i})^T$ and $\delta_i = (\delta_{i1}, \ldots, \delta_{in_i})^T$.

3. Proposed Laplace approximation

For estimating fixed effects $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0r})^T$ and $\beta = (\beta_1, \ldots, \beta_p)^T$, the standard h-likelihood methods (Ha *et al.*, 2001; Ha and Lee, 2003) have, respectively, used h and the profile h-likelihood

$$h^* = h|_{\lambda_0 = \widetilde{\lambda}_0}, \tag{3.1}$$

where $\widetilde{\lambda}_{0k}$ slove $\partial h/\partial\lambda_{0k} = 0$. Note here that the dimension of λ_0 increases with the sample size n and that r is the number of distinct event times among the Y_{ij}'s. We have shown that the standard h-likelihood method based on h^* in (3.1) provides a reasonable estimator by eliminating many nuisance parameters λ_{0k} caused by no censoring or a lower censoring (Ha and Lee, 2005; Ha *et al.*, 2010). In such cases, the marginal ML estimators under gamma frailty models could suffer from substantial biases, particularly for frailty parameter α (Barker and Henderson, 2005; Ha et al., 2010). However, we have found that the standard h-likelihood methods can give a bias estimator for lognormal frailty with high censoring: see the simulation results in Section 4. Thus, below we propose the use of a LA method based on h-likelihood.

Let $\ell = \ell(\beta, \theta)$ be a likelihood, either an h-likelihood, h, or a marginal likelihood, m, with nuisance parameters θ. Lee and Nelder (2001) considered a function $p_\theta(\ell)$, defined by

$$p_\theta(\ell) = [\ell - \frac{1}{2}\log\det\{D(\ell,\theta)/(2\pi)\}]|_{\theta=\widehat{\theta}}\ , \tag{3.2}$$

where $D(\ell,\theta) = -\partial^2\ell/\partial\theta^2$ and $\widehat{\theta}$ solves $\partial\ell/\partial\theta = 0$. The function $p_\theta(\cdot)$ produces an adjusted profile likelihood, eliminating nuisance effects θ, which can be fixed effects (β,λ_0) or random effects v or both. Following the notation of (3.2), we define an adjusted profile h-likelihood $p_v(h)$ as

$$p_v(h) = [h - \frac{1}{2}\log\det\{D(h,v)/(2\pi)\}]|_{v=\widehat{v}}\ , \tag{3.3}$$

where $D(h,v) = -\partial^2 h/\partial v^2$ and $\widehat{v}$ solves $\partial h/\partial v = 0$. Then $p_v(h)$ is the function of parameters (λ_0,β,α) and it becomes the first-order LA to marginal likelihood m, given by

$$m = \log\{\int \exp(h)dv\}.$$

For the estimation of fixed effects (β,λ_0) we propose the use of $p_v(h)$. Following Barndorff-Nielsen and Cox (1989), we see that as $n^* = \min_{1\le i\le q} n_i \to \infty$, we have the relationship between m and $p_v(h)$ (Lee and Nelder, 2001; Ha *et al.*, 2017):

$$m = p_v(h) + O(n^{*-1}).$$

Note here that $p_v(h)$ produces an adjusted profile h-likelihood for (β,λ_0,α) with v eliminated (Lee and Nelder, 2001); that is, it can be expressed as

$$p_v(h) = \hat{h} - \frac{1}{2}\log\det(\hat{D}) + \frac{q}{2}\log(2\pi),$$

where $\hat{h} = h|_{v=\hat{v}}$ is a profile likelihood for (β,λ_0,α) after eliminating v and $\hat{D} = D(h,\hat{v}) = (-\partial^2 h/\partial v^2)|_{v=\hat{v}}$. In the gamma frailty model, given α, for the estimation of (β,λ_0) the uses of $p_v(h)$ and h are equivalent because $\hat{D} = \alpha^{-1} + \sum_j \delta_{ij}$ which is free of (β,λ_0). However, in the log-normal frailty model, given α, the use of $p_v(h)$ for the estimation of (β,λ_0) may be crucial when censoring rate is high because $p_v(h)$ and h give different estimators for (β,λ_0). We see that in gamma frailty model, given α, the marginal MLE for fixed effects (β,λ_0) can be obtained by maximization of h (or $p_v(h)$) as in Poisson-gamma HGLM. Thus, for the gamma frailty model the LA estimators leads to the marginal ML estimators for (β,λ_0), given α.

Next, for the estimation of frailty parameter α, the standard h-likelihood method (Ha and Lee, 2003; Lee *et al.*, 2017) based on the profile h-likelihood h^* uses the second-order LA, $s_{\beta,v}(h^*)$ as in the HGLM procedure. Here, $s_{\beta,v}(h^*)$ is given by

$$s_{\beta,v}(h^*) = p_{\beta,v}(h^*) - F(h^*)/24. \tag{3.4}$$

Here

$$F(h^*) = \sum_{i=1}^{q}\left\{-3\left(\frac{\partial^4 h^*}{\partial v_i^4}\right)b_{ii}^2 - 5\left(\frac{\partial^3 h^*}{\partial v_i^3}\right)^2 b_{ii}^3\right\}|_{v=\widehat{v}},$$

where b_{ii} is the ith diagonal element of $D(h^*,v)^{-1}$. For efficiency of computation, we use $F(h)$ instead of $F(h^*)$ (Ha *et al.*, 2010, 2017).

In summary, we propose the use of $p_v(h)$ for the estimation of (β,λ_0) and that of $s_{\beta,v}(h)$ for the estimation of α, respectively. In other words, (i) given α, the LA estimates of (β,λ_0) are obtained by maximizing $p_v(h)$ and (ii) given the estimates of (β,λ_0), the LA estimate of α is obtained by maximizing $s_{\beta,v}(h^*)$. Their performances are assessed via simulation studies below.

4. Simulation study

Numerical studies, based on 200 replications of simulated data, are presented to evaluate the performance of the proposed method. The simulated data are generated from lognormal frailty model (2.1) having exponential baseline hazard $\lambda_0(t) = 1$, one single binary covariate with β and the variance of Normal frailty $\alpha \equiv \sigma^2$. Here, the true values are $\sigma^2 = \beta = 1$. The censoring distribution is an exponential and the censoring rate is considered as a high censoring with about 70% censoring. We consider a sample size $n = 200$ with $(q, n_i) = (100, 2)$ for all i:

We computed the mean and mean squared error (MSE) of estimates of parameters $\theta = (\beta, \sigma^2)^T$. The MSE is defined by $\sum_i (\widehat{\theta}^{(i)} - \theta)^2/M$, where $\widehat{\theta}^{(i)}$ is the estimate of θ in the ith replication and $\bar{\theta} = \sum_i \widehat{\theta}^{(i)}/M$ is the mean of $\widehat{\theta}^{(i)}$'s, and $\theta = \sigma^2$ or β.

From Table 4.1 we find that the h-likelihood (HL) method shows a downward bias for σ^2, leading to also a downward bias for β. However, we observe that the LA method improves the h-likelihood (HL) method in terms of biases for (σ^2, β). In particular, the LA method for σ^2 gives smaller MSEs as compared to the HL method. These results indicate that the proposed LA method is reasonable under a high censoring.

Table 4.1 Simulation results based on 200 replications for the estimation of parameters in log-normal frailty model (70% censoring).

	$\hat{\sigma^2}$		$\hat{\beta}$	
Method	Mean	MSE	Mean	MSE
HL	0.813	0.479	0.909	0.133
LA	0.975	0.436	0.982	0.153

Note: HL, h-likelihood method; LA, Laplace approximation method; β, regression parameter; σ^2, variance of log-frailty.

5. Real data examples

In this section we illustrate the proposed LA method using two practical examples with well-known real data with high censoring. For the model fitting and computation, we used SAS/IML.

Example 1: Litter matched rat data. Mantel *et al.* (1977) presented a survival data set from a tumorigenesis study of 50 litters of female rats. For each litter, one rat was selected to receive the study drug and the other two rats were treated with placebo. Here, each litter can be treated as a cluster. Thus, the data consists of $n = 150$ having $q = 50$ litters with each three litter size ($n_i = 3$ for all i). Here, survival time (time-to-event) is time to development of tumor, measured in weeks. Forty rats among 150 rats developed a tumor, leading to about 73.3% censoring. Survival times for rats within a litter may be correlated due to the shared genetic or environmental effects. We thus fitted the log-normal frailty model with a single covariate, treatment (1 for drug and 0 for placebo). The fitted results are summarized in Table 5.1.

Table 5.1 Results on the estimation of parameters in the litter-matched rat data (73.3% censoring).

Method	$\hat{\beta}_1$ (SE)	$\hat{\sigma^2}$
HL	0.903 (0.321)	0.315
LA	1.056 (0.323)	0.472

Note: HL, h-likelihood method; LA, Laplace approximation method; β, regression parameter; σ^2, variance of log-frailty; SE, the estimated standard error.

Table 5.1 shows that the LA estimates of β and α are larger than the corresponding HL estimates. The trends of the results in Table 5.1 are overall similar to those evident in Table 4.1.

Example 2: CGD data. The CGD (chronic granulomatous disease) data set is from a placebo-controlled randomized trial of gamma interferon in CGD (Fleming and Harrington, 1991). This trial aimed to investigate the effectiveness of gamma interferon (γ-IFN) in reducing the rate of serious infections in the CGD patients. In total, 128 patients from 13 hospitals were followed for about 1 year, and they had recurrent infections with different cluster size n_i, ranging from 1 to 8. The number of patients accrued per hospital ranged from 4 to 26. Among 63 patients in the treatment group, 14 patients experienced at least one infection and a total of 20 infections were recorded. In the placebo group, 30 out of 65 patients experienced at least one infection, with a total of 56 infections being recorded.

Here, time to event (i.e. survival time) is the gap time (i.e. inter-arrival time) between recurrent infection times. Censoring occurred at the last follow-up for all patients, except one, who experienced a serious infection on the date he left the study. In this study, about 63.3% of the individuals were censored. We fitted the log-normal frailty model for the gap times, with a single covariate, treatment (0 = placebo, 1 = γ-IFN).

Table 5.2 again shows that the LA estimates of the absolute of β and α are larger than the HL estimates. These LA results in Table 5.2 also confirm the simulation results in Table 4.1.

Table 5.2 Results on the estimation of parameters in the CGD data (63.3% censoring).

Method	$\hat{\beta}_1$ (SE)	$\hat{\sigma^2}$
HL	-1.062 (0.325)	0.836
LA	-1.151 (0.347)	1.065

Note: HL, h-likelihood method; LA, Laplace approximation method; β, regression parameter; σ^2, variance of log-frailty; SE, the estimated standard error.

6. Discussion

We have found via simulation studies that the proposed LA method reduces the bias caused by the standard h-likelihood method when the censoring rate is high. Note that the estimates of nuisance parameters λ_0 or $\Lambda_0(t)$ are plugged in the h-likelihood procedure. The standard h-likelihood method of nuisances parameters using h may give biased estimates in such extreme cases such as a high censoring because plugging these poor estimates in the likelihood may result in a bias in parameter estimation. However, with the proposed LA method the λ_0 are obtained by maximizing $p_v(h)$, not h. In Table 4.1 we presented the simulation results of a high censoring rate only such as 70%. Thus, simulation studies with various censoring rates seem to be necessary as a future work.

Furthermore, for the gamma frailty model the h-likelihood method for λ_0 given (β, σ^2) provides the same estimators as the ML method. Thus, the standard h-likelihood method may work well even in high censoring even if it require a further study including simulation study. However, theoretical justifications related to the estimation of λ_0 would be an interesting future work.

References

Barker, P. and Henderson, R. (2005). Small sample bias in the gamma frailty model for univariate survival. *Lifetime Data Analysis*, **11**, 265-284

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic techniques for use in statistics*, Chapman and Hall, New York.

Breslow, N. E. (1972). Discussion of professor Cox's paper. *Journal of the Royal Statistical Society B*, **34**, 216-217.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society B*, **74**, 187-220.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley, New York

Ha, I. D. and Cho, G.-H. (2015). Variable selection in Poisson HGLMs using h-likelihood. *Journal of the Korean Data & Information Science Society*, **26**, 1513-521.

Ha, I. D., Jeong, J. H. and Lee, Y. J. (2017). *Statistical modelling of survival data with random effects: H-likelihood approach*, Springer, Singapore.

Ha, I. D. and Lee, Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational Graphical Statistics*, **12**, 663-681.

Ha, I. D., Lee, Y. and Song, J. K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**, 233-243.

Ha, I. D. and Noh, M. (2013). A visualizing method for investigating individual frailties using frailtyHL R-package. *Journal of the Korean Data & Information Science Society*, **24**, 931-940.

Ha, I. D., Noh, M. and Lee, Y. (2010). Bias reduction of likelihood estimators in semi-parametric frailty models. *Scandinavian Journal of Statistics*, **37**, 307-320

Kim, B., Ha, I. D. and Lee, D. (2016). Analysis of multi-center bladder cancer survival data using variable selection method of multi-level frailty models. *Journal of the Korean Data & Information Science Society*, **27**, 499-510.

Lee, T. and Lee, M. (2017). Analysis of stage III proximal colon cancer using the Cox proportional hazards model. *Journal of the Korean Data & Information Science Society*, **28**, 349-359.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619-678.

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.

Lee, Y., Nelder, J. A. and Pawitan, Y. (2017). *Generalised linear models with random effects: Unified analysis via h-likelihood*, 2nd edn., London: Chapman and Hall.

Mantel, N., Bohidar, N. R. and Ciminera, J. L. (1977). Mantel-Haenszel analysis of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Research*, **37**, 3863-3868.