Journal of the Korean Data & Information Science Society
2016, **27**(5), 1389–1397
http://dx.doi.org/10.7465/jkdi.2016.27.5.1389
한국데이터정보과학회지

ML estimation using Poisson HGLM approach in semi-parametric frailty models†

Il Do Ha[1]

[1]Department of Statistics, Pukyong National University
Received 28 June 2016, revised 17 July 2016, accepted 18 July 2016

Abstract

Semi-parametric frailty model with nonparametric baseline hazards has been widely used for the analyses of clustered survival-time data. The frailty models can be fitted via an auxiliary Poisson hierarchical generalized linear model (HGLM). For the inferences of the frailty model marginal likelihood, which gives MLE, is often used. The marginal likelihood is usually obtained by integrating out random effects, but it often requires an intractable integration. In this paper, we propose to obtain the MLE via Laplace approximation using a Poisson HGLM approach for semi-parametric frailty model. The proposed HGLM approach uses hierarchical-likelihood (h-likelihood), which avoids integration itself. The proposed method is illustrated using a numerical study.

Keywords: H-likelihood, Laplace approximation, marginal likelihood, Poisson HGLMs, semi-parametric frailty models.

1. Introduction

For inference of semi-parametric frailty models with nonparametric baseline hazards, the ML estimation of parameters requires the computation of a marginal likelihood (i.e. observed-data likelihood). However, the marginal likelihood often involves an intractable integral in integrating out the frailty terms (i.e. random effects). For the log-normal frailty model with normally distributed random effect, the marginal likelihood does not give an explicit form, leading to an numerical integration. Recently, for the log-normal frailty model the ML estimator, which maximizes the marginal likelihood, is available via the `phmm()` function in the **phmm** R package (Donohue and Xu, 2013), but care must be taken to ensure that the Monte Carlo expectation maximization (MCEM) algorithm converges (Ha *et al.*, 2012). Ma *et al.* (2003) and Ha and Lee (2005) have shown that semi-parametric frailty models can be fitted via an auxiliary Poisson HGLM (Lee and Nelder, 1996). This is clearly an extension of a Poisson generalized linear model for fitting Cox's (1972) proportional hazards model (Whitehead, 1980).

Thus, in this paper we are interested in the ML estimation in semi-parametric frailty models using a Poisson HGLM approach. For this purpose, we propose a Laplace approximation (LA) method based on h-likelihood (Lee and Nelder, 1996). Here, the h-likelihood obviates

† 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-2015R1D1A3A01015663).
[1] Professor, Department of Statistics, Pukyong National University, Busan 608-737, Korea.
E-mail: idha1353@pknu.ac.kr

the integration itself and gives a statistically efficient procedure for various random-effect models such as HGLMs (Lee *et al.*, 2006; Ha and Noh, 2013; Ha and Cho, 2015). In Poisson HGLMs and frailty models, the use of gamma random-effect distribution gives an explicit marginal likelihood, but other distributions such as normal random-effect distribution do not. Thus, in this paper we consider the Poisson HGLM with normal random effects. In this case, the Gauss-Hermite Quadrature (GHQ) method based on SAS NLMIXED procedure provides the MLE. Accordingly, the performance of proposed LA method can be evaluated via the comparison with the marginal GHQ method.

The organization of this paper is as follows. In Section 2 we describe the structures of data and frailty models. Then we review the corresponding h-likelihood procedures and explain briefly an equivalence between semi-parametric frailty model and Poisson HGLM, which provides a motivation for the use of the proposed LA. In Section 3 we show how to find the LA estimators based on the h-likelihood using a Poisson HGLM approach for semi-parametric frailty model. In section 4 the proposed method is demonstrated using an numerical study based on a practical data set. Our LA results are also compared with those from the SAS NLMIXED procedure. Finally, some discussions are given in Section 5.

2. Fitting frailty model via a Poisson HGLM

2.1. Frailty model and h-likelihood

Data structures considered are as follows. Let T_{ij} $(i = 1, \dots, q,\ j = 1, \dots, n_i)$ be the survival time (i.e. event time or time-to-event) for the jth observation of the ith cluster (or subject) and let C_{ij} be the corresponding censoring time. Then observable random variables are given by

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

where $I(\cdot)$ is the indicator function. Here $n = \sum_i n_i$ is the total sample size, n_i is the cluster size and q is the number of clusters.

Given an unobserved frailty $U_i = u_i$ for the ith subject, suppose that the form of conditional hazard function of T_{ij} is given by

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

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, $x_{ij} = (x_{ij1}, \dots, x_{ijp})^{\mathrm{T}}$ is a vector of fixed covariates and $\beta = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of unknown regression parameters. We assume the frailties U_i are independent and identically distributed random variables with frailty parameter α. The popular distributions for U_i are gamma and lognormal. Traditionally it is also assumed that the frailty distributions have mean 1 and variance, say, α, i.e. $E(U_i) = 1$ and $\mathrm{var}(U_i) = \alpha$ for the gamma frailty model and $V_i = \log U_i \sim N(0, \alpha)$ for the log-normal frailty model.

In model (2.1) the functional form of the baseline hazard $\lambda_0(t)$ is unknown. Along the lines of Breslow (1972), consider the baseline cumulative hazard function $\Lambda_0(t)$ to be a step function with jumps at the observed event (e.g. death) times:

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

where $y_{(k)}$ is the kth $(k = 1, \dots, r)$ smallest distinct event time among y_{ij}'s and $\lambda_{0k} = \lambda_0(y_{(k)})$. In this paper we are interested in the estimation of β and α; here λ_{0k} are nuisance parameters. For simplicity of argument we consider models with one frailty term only, even though it can be extended to multi-component frailty terms as in Lee *et al.* (2006).

Let $w = (w_1, \ldots, w_r)^{\mathrm{T}}$, where $w_k = \log \lambda_{0k}$. Along the lines of Lee and Nelder (1996) and Ha *et al.* (2001), the h-(log)likelihood for semi-parametric frailty models (2.1), denoted by h, is defined by

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

where

$$\sum_{ij} \ell_{1ij} = \sum_{k} d_{(k)} w_k + \sum_{ij} \delta_{ij} \eta_{ij} - \sum_{k} \exp(w_k) \Big\{ \sum_{(i,j) \in R(y_{(k)})} \exp(\eta_{ij}) \Big\}.$$

Here

$$\ell_{1ij} = \ell_{1ij}(w, \beta; y_{ij}, \delta_{ij} | u_i)$$

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, $d_{(k)}$ is the number of events at $y_{(k)}$,

$$\eta_{ij} = x_{ij}^{\mathrm{T}} \beta + v_i \text{ with } v_i = \log u_i,$$

is the linear predictor on the log-hazard, and

$$R_{(k)} = R(y_{(k)}) = \{(i, j) : y_{ij} \geq y_{(k)}\}$$

is the risk set at $y_{(k)}$.

Lee and Nelder (1996) proposed the use of h for inferences about $\theta = (w^{\mathrm{T}}, \beta^{\mathrm{T}}, v^{\mathrm{T}})^{\mathrm{T}}$ with $v = (v_1, \ldots, v_q)^{\mathrm{T}}$ and hence that of an adjusted profile likelihood, $p_\theta(h)$, after eliminating θ, for inferences about α, given by

$$p_\theta(h) = [h - \frac{1}{2} \log \det\{D(h, \theta)/(2\pi)\}]|_{\theta=\hat{\theta}}, \tag{2.3}$$

where $D(h, \theta) = -\partial^2 h / \partial \theta^2$ and $\hat{\theta}$ solves $\partial h / \partial \theta = 0$. We note that $p_\theta(h)$ is the first-order Laplace approximation to restricted likelihood of α.

2.2. Equivalence of procedures of frailty model and Poisson HGLM

Following Ha and Lee (2005), below we outline the semi-parametric frailty model can be fitted via a Poissin HGLM. Let $y_{ij,k}$ be 1 if the (i, j)th individual experiences an event at $y_{(k)}$ and 0 otherwise. Ma *et al.* (2003) showed that frailty models could be fitted by using the following auxiliary Poisson random-effect model. Given $U_i = u_i$, assume that $y_{ij,k}$ are conditionally independent with

$$y_{ij,k} | u_i \ \sim \ \text{Poisson}(\mu_{ij,k} u_i), \ (i, j) \in R_{(k)}, \tag{2.4}$$

where

$$\mu_{ij,k} = \exp(w_k + x_{ij}^{\mathrm{T}} \beta) = \exp(x_{ij,k}^{\mathrm{T}} \gamma).$$

Here $x_{ij,k} = (e_k^{\mathrm{T}}, x_{ij}^{\mathrm{T}})^{\mathrm{T}}$, e_k is a vector of components 0 and 1 such that $e_k^{\mathrm{T}} w = w_k$, and $\gamma = (w^{\mathrm{T}}, \beta^{\mathrm{T}})^{\mathrm{T}}$. Note here that $e_k = (0, \cdots, 1, \cdots, 0)^T$ and $w = (w_1, \cdots, w_k, \cdots, w_r)^T$.

Let y and U denote the vectors of the $y_{ij,k}$'s and the U_i's, respectively. Then this auxiliary model provides the conditional Poisson log-likelihood for y given u:

$$\ell_{\mathrm{P1}}(\gamma; y|u) = \sum_k \sum_{(i,j)\in R_{(k)}} \{y_{ij,k} \log(\mu_{ij,k} u_i) - \mu_{ij,k} u_i\}. \tag{2.5}$$

From (2.2) and (2.5) we see that ℓ_{P1} is equivalent to the first term, $\sum_{ij} \ell_{1ij}$, in h of (2.2) because

$$\sum_k \sum_{(i,j)\in R_{(k)}} y_{ij,k} \log(\mu_{ij,k} u_i) = \sum_k d_{(k)} w_k + \sum_{ij} \delta_{ij} \eta_{ij}$$

and $\mu_{ij,k} u_i = \exp(w_k + \eta_{ij})$.

Accordingly, for fitting frailty models we can use either Lee and Nelder's (1996, 2001) procedure with the Poisson HGLM (2.4) or Ha *et al.*'s (2001) profile-likelihood procedure (i.e. h-likelihood procedure after eliminating nuisances w) for the frailty model (2.1). In fact, the two procedures are equivalent. In this paper we focus on the former, rather than the latter, because an objective in this paper is to present how to obtain the ML estimators of parameters in semi-parametric frailty models (2.1) using the Poisson HGLM procedure. Here, we consider the log-normal frailty model where the explicit form of marginal likelihood is not available. For this we propose the use of LA approach based on h-likelihood, as shown in next section.

3. Laplace approximated MLE in Poisson HGLM

Consider the Poisson HGLM (2.4) with normal random effect. Then, from (2.5) the corresponding h-likelihood is given by

$$h = h(\gamma, v, \alpha) = \ell_{\mathrm{P1}}(\gamma; y|u) + \ell_2(\alpha; v), \tag{3.1}$$

where ℓ_{P1} is given in (2.5), $\ell_2(\alpha; v) = \sum_i \ell_{2i}$ with

$$\ell_{2i} = \ell_{2i}(\alpha; v_i) = -\frac{1}{2}\log(2\pi\alpha) - \frac{1}{2\alpha} v_i^2,$$

which is the logarithm of the density function for v_i.

The corresponding marginal likelihood m is given by integrating out random effects v from the h-likelihood (3.1):

$$m = m(\beta, \alpha) = \log\left\{\int \exp(h)\ dv\right\}, \tag{3.2}$$

where h is given in (3.1) which includes the density function of log-frailty v. Maximizing the marginal likelihood m gives MLEs. However, the integration in (3.2) cannot be computed explicitly except for Poisson-gamma HGLM (Lee and Nelder, 1996). Thus, an approximation (e.g. Laplace approximation) of this integral is needed.

3.1. The first-order Laplace approximation

Firstly we consider the first-order LA (LA1), $p_v(h)$, to the marginal likelihood m. Following Barndorff-Nielsen and Cox (1989), we can show that as $N = \min_{1\le i\le q} n_i \to \infty$,

$$m = p_v(h) + O(N^{-1}). \tag{3.3}$$

Here, from the definition of (2.3), $p_v(h)$ is given by

$$p_v(h) = \left[h - \frac{1}{2}\log\det\left\{\frac{D(h,v)}{(2\pi)}\right\}\right]\Big|_{v=\hat{v}},$$

where $D(h,v) = -\partial^2 h/\partial v^2$ and $\hat{v}$ solves $\partial h/\partial v = 0$. Note that $p_v(h)$ in (3.3) leads to an adjusted profile h-likelihood (i.e. a restricted likelihood) for (γ,α) with v eliminated (Lee and Nelder, 2001). Accordingly, maximizing $p_v(h)$ gives the LA1 estimate; its detailed form is given by

$$p_v(h) = \hat{h} - \frac{1}{2}\log\det(\hat{D}) + \frac{q}{2}\log(2\pi), \tag{3.4}$$

where $\hat{h} = h|_{v=\hat{v}}$, $\hat{D} = D(h,\hat{v}) = Z^T\hat{W}Z + U$, and $\hat{W} = \mathrm{diag}(\hat{\mu})$ is a diagonal weight matrix with $\hat{\mu} = \exp(X\gamma + Z\hat{v})$ and $U = \alpha^{-1}I_q$ with $q\times q$ identity matrix I_q. Here, X and Z are model matrices of γ and v, respectively.

3.1.1. Estimation of fixed effects

The LA1 estimators for fixed effects $\gamma = (w^T, \beta^T)^T$, given α, are obtained by solving iteratively

$$\frac{\partial p_v(h)}{\partial \gamma_k} = 0 \quad (k = 1,\ldots,r,r+1,\ldots,r+p). \tag{3.5}$$

Then the score equation of (3.5) is computed as follows. From (3.4) we obtain

$$\frac{\partial p_v(h)}{\partial \gamma_k} = \frac{\partial \hat{h}}{\partial \gamma_k} - \frac{1}{2}\mathrm{tr}\left(\hat{D}^{-1}\frac{\partial \hat{D}}{\partial \gamma_k}\right).$$

Here

$$\frac{\partial \hat{h}}{\partial \gamma_k} = \frac{\partial h}{\partial \gamma_k}|_{v=\hat{v}}$$

since $\partial\hat{h}/\partial\gamma_k = \{(\partial h/\partial\gamma_k) + (\partial h/\partial v)(\partial\hat{v}/\partial\gamma_k)\}|_{v=\hat{v}}$ and $(\partial h/\partial v)|_{v=\hat{v}} = 0$ (Ha *et al.*, 2001), and $\partial h/\partial\gamma_k = (y-\mu)^T X_k$ with the kth column vector X_k of X, and

$$\frac{\partial \hat{D}}{\partial \gamma_k} = Z^T\hat{W}_k' Z,$$

where $\hat{W}_k' = \partial\hat{W}_k/\partial\gamma_k$. We also have the following equations (Lee and Nelder, 1996):

$$\begin{aligned}\frac{\partial \hat{v}}{\partial \gamma_k} &= -\left(\frac{-\partial^2 h}{\partial v^2}\right)^{-1}\left(\frac{-\partial^2 h}{\partial v\partial\gamma_k}\right)|_{v=\hat{v}} \\ &= -(Z^T\hat{W}Z + U)^{-1}(Z^T\hat{W}X_k).\end{aligned}$$

In order to solve the estimating equations (3.5), we use the Newton-Raphson method, with the following negative second derivatives

$$-\frac{\partial^2 p_v(h)}{\partial\gamma_k\partial\gamma_l} = -\frac{\partial^2\hat{h}}{\partial\gamma_k\partial\gamma_l} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\gamma_k}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\gamma_l} + \frac{\partial^2\hat{D}}{\partial\gamma_k\partial\gamma_l}\right). \tag{3.6}$$

3.1.2. Estimation of dispersion parameter

The LA1 dispersion estimator for α is also obtained by solving iteratively

$$\frac{\partial p_v(h)}{\partial\alpha} = 0 \tag{3.7}$$

with

$$\frac{\partial p_v(h)}{\partial\alpha} = \frac{\partial h}{\partial\alpha}|_{v=\hat{v}} - \frac{1}{2}\mathrm{tr}\left(\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha}\right).$$

Here $\partial h/\partial\alpha = \sum_i \partial\ell_{2i}/\partial\alpha = \sum_i\{-1/(2\alpha) + v_i^2/(2\alpha^2)\}$, and

$$\frac{\partial\hat{D}}{\partial\alpha} = Z^T\left(\frac{\partial\hat{W}}{\partial\alpha}\right)Z - \alpha^{-2}I_q.$$

In order to solve the estimating equation (3.7), we also use the Newton-Raphson method with the negative second derivative, given by

$$-\frac{\partial^2 p_v(h)}{\partial\alpha^2} = -\frac{\partial^2\hat{h}}{\partial\alpha^2} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha} + \frac{\partial^2\hat{D}}{\partial\alpha^2}\right). \tag{3.8}$$

3.2. The second-order Laplace approximation

The $p_v(h)$ in (3.3) may not provide sufficiently accurate approximation to m when the cluster size n_i are small as in $n_i = 1$ or 2 for all i, or the dispersion parameter α is larger (Lee and Nelder, 2001; Lee *et al.*, 2006). Thus, Lee and Nelder (2001) have recommended the use of the second-order Laplace approximation. As $N = \min_{1\le i\le q} n_i \to \infty$ we have

$$m = s_v(h) + O(N^{-2}),$$

where

$$s_v(h) = p_v(h) - F(h)/24.$$

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 $D(h,v)^{-1}$. In this paper, we call dispersion estimator of α maximizing $s_v(h)$, the second-order Laplace approximation (LA2) estimator; similarly, it is obtained by solving the estimating equation for α:

$$\frac{\partial s_v(h)}{\partial\alpha} = 0.$$

For the estimation of γ, LA2 method uses $p_v(h)$, not $s_v(h)$, because we have found that the use of $p_v(h)$ works well (Lee *et al.*, 2006).

3.3. Variance estimation for parameter estimators

We see that the asymptotic covariance matrix of $\hat{\gamma}$ and $\hat{\alpha}$ is obtained from the inverse of observed information matrix, given by $-\partial^2 m/\partial\psi^2$ with $\psi = (\gamma^T, \alpha)^T$. However, the integration in m is generally intractable. Thus we use the first-order approximation, $p_v(h)$. Therefore, the variances of $\hat{\psi}$ are estimated from the inverse of $H = -\partial^2 p_v(h)/\partial\psi^2$, defined by

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{pmatrix}, \tag{3.9}$$

where $H_{11} = -\partial^2 p_v(h)/\partial\gamma^2$ is given in (3.6), $H_{22} = -\partial^2 p_v(h)/\partial\alpha^2$ is given in (3.8) and $H_{12} = -\partial^2 p_v(h)/\partial\gamma\partial\alpha$ with its entries

$$-\frac{\partial^2 p_v(h)}{\partial\gamma_k\partial\alpha} = -\frac{\partial^2\hat{h}}{\partial\gamma_k\partial\alpha} + \frac{1}{2}\mathrm{tr}\left(-\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\gamma_k}\hat{D}^{-1}\frac{\partial\hat{D}}{\partial\alpha} + \frac{\partial^2\hat{D}}{\partial\gamma_k\partial\alpha}\right).$$

4. Numerical example

The proposed method is illustrated using a well-known real data set, which is the kidney infection data introduced by McGilchrist and Aisbett (1991). Here, we focus upon comparisons of LA1 and LA2 estimators. We also include the marginal GHQ method using SAS NLMIXED procedure. For the model fitting and computation, we used SAS/IML.

The data set consists of times to until the first and second recurrences ($n_i = 2$) of kidney infection in 38 patients ($q = 38$) using a portable dialysis machine. The catheter is later removed if infection occurs and can be removed for other reasons, which is treated as censoring; about 23.7% of the data were censored. Here, each survival time is time to infection since insertion of the catheter. The covariates of interest are Age and Sex (1=female, 0=male). The survival times from the same patient may be correlated due to a shared patient effect.

Table 4.1 Estimation results of parameters in the kidney infection data (single covariate)

Method	$\hat{\beta}_1$ (SE)	$\hat{\alpha}$ (SE)
LA1	-1.316 (0.447)	0.384 (0.288)
LA2	-1.301 (0.443))	0.362 (0.275)
GHQ	-1.304 (0.448)	0.364 (0.294)

Note: β_1, Sex effect; α, variance of random effect; SE, the estimated standard error.

For the data set we perform two numerical analyses. For the first analysis, we fitted the Poisson HGLM (2.4) with one single covariate, Sex. For the second analysis, we used the two covariates, Sex and Age. The results are summarized in Tables 4.1 and 4.2, respectively. The results in Table 4.1 show that for the Sex effect β_1 and dispersion parameter α, the LA2 estimates are about the same as the GHQ estimates using SAS NLMIXED procedure. However, the absolute magnitudes of LA1 estimates of β_1 and α are larger than those of the GHQ estimates. We observe that the estimated SE of $\hat{\alpha}$ by the LA2 in Table 4.1 is slightly smaller than that of the GHQ. A possible reason is that for the computation of the SE we use the LA1 method with $p_v(h)$, not $s_v(h)$. The trends of the results in Table 4.2 are overall similar to those evident in Table 4.1. It is well known that the GHQ with a single quadrature point (i.e. Q=1) results in the first-order Laplace approximation of the log likelihood (SAS Institute Inc., 2014). We have confirmed that our LA1 results in Tables 1 and 2 are the same as the NLMIXED results with Q=1. Note that we used Q=20 for the GHQ computation in Tables 4.1 and 4.2.

Table 4.2 Results on the estimation of parameters in the kidney infection data (two covariates)

Method	$\hat{\beta}_1$ (SE)	$\hat{\beta}_2$ (SE)	$\hat{\alpha}$ (SE)
LA1	-1.318 (0.449)	0.004 (0.012)	0.390 (0.289)
LA2	-1.304 (0.445)	0.004 (0.011)	0.368 (0.277)
GHQ	-1.303 (0.449)	0.004 (0.011)	0.366 (0.292)

Note: β_1, Sex effect; β_2, Age effect.

5. Discussion

For the ML estimation in semi-parametric frailty models, we have proposed a Poisson HGLM approach based on the LA method. In particular, we have found via an numerical study in Section 4 that the LA1 estimates give somewhat larger bias for the estimation of α, but that the LA2 estimates are about the same as the GHQ estimates by SAS NLMIXED procedure. Thus, we recommend the use of the LA2 method. In this paper, we used the LA1 method for the estimation of SE of $\hat{\alpha}$. In Tables 4.1 and 4.2 we have also found that such SEs were underestimated in the LA2 method. For the improvement the application of LA2 method would be necessary even if it requires more computations.

In this paper we considered only models with one random-effect term. The proposed method can be extended to the frailty models with more than one random-effect term. The kidney data set we use in Section 4 is somewhat small. However, in the Poisson HGLM (2.4), the number of nuisance parameters w_k's in fixed effects $\gamma = (w^T, \beta^T)^T$ is increased with sample size n, leading to a computation problem of a high-dimensional inverse matrix in using the Newton-Raphson method. Developing an estimation procedure to overcome this problem would be an interesting further work.

References

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.

Donohue, M. and Xu, R. (2013). Phmm: Proportional hazards mixed-effects model, `http://CRAN.R-project.org/package=phmm.Rpackageversion0.7-5`.

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**, 1513-1521.

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.

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. (2012). frailtyHL: A package for fitting frailty models with h-likelihood. *R Journal*, **4**, 307-320.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461-466.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society 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. (2006). *Generalised linear models with random effects: Unified analysis via h-likelihood*, Chapman and Hall, London.

Ma, R., Krewski, D. and Burnett, R. T. (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika*, **90**, 157-169.

SAS Institute Inc. (2014). *SAS/STAT 13.2 users guide: The NLMIXED procedure*, SAS Institute Inc., Cary, NC, USA.

Whitehead, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Applied Statistics*, **29**, 268-275.