Journal of the Korean Data & Information Science Society
2012, **23**(1), 199–207

한국데이터정보과학회지

H-likelihood approach for variable selection in gamma frailty models†

Il Do Ha[1] · Geon-Ho Cho[2]

[12]Daegu Haany University

Received 11 November 2011, revised 30 November 2011, accepted 14 December 2011

Abstract

Recently, variable selection methods using penalized likelihood with a shrink penalty function have been widely studied in various statistical models including generalized linear models and survival models. In particular, they select important variables and estimate coefficients of covariates simultaneously. In this paper, we develop a penalized h-likelihood method for variable selection in gamma frailty models. For this we use the smoothly clipped absolute deviation (SCAD) penalty function, which satisfies a good property in variable selection. The proposed method is illustrated using simulation study and a practical data set.

Keywords: Frailty models, h-likelihood, marginal likelihood, penalized likelihood, variable selection.

1. Introduction

Selecting relevant variables from a high-dimensional statistical modelling is very important in data analysis. In ordinary linear regression models, there are a variety of classical techniques for variable selection such as forward selection, backward elimination and stepwise selection. However, these classical methods can be computationally expensive for large data set and often suffer from high variability (Fan and Li, 2001; Lu and Zhang, 2007). Thus, Tibshirani (1996) proposed a new variable-selection technique in linear models, called "least absolute shrinkage and selection operator". It shrinks some coefficients and sets others to zero; that is, the true regression coefficients that are zero are automatically estimated as zero, and the remaining coefficients are estimated well as if the correct submodel were known in advance. However, the LASSO method can lead to biased estimates for the large coefficients and does not satisfy a good property such as oracle property for variable selection. Thus, the improved shrinkage methods including SCAD (Fan and Li, 2001, 2002) and adaptive-LASSO (Zou, 2006) have been developed using the penalized likelihood for various models such as generalized linear models (GLMs) and Cox's proportional hazards (PH) models. In particular, Fan and Li (2002) proposed the penalized marginal likelihood

† The 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).

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

[2] Professor, Department of Asset Management, Daegu Haany University, Gyeongsan 712-715, Korea.

method using the SCAD penalty function for gamma frailty models with correlated survival data. The marginal likelihood approach often requires an intractable integration in order to eliminate the frailties (i.e. random effects) (Ha *et al.*, 2001; Ha and Lee, 2003).

In this paper, we propose a penalized h-likelihood approach for variable selection in gamma frailty models. For this we use the SCAD penalty. The h-likelihood (or hierarchical likelihood; Lee and Nelder, 1996) avoids such intractable integration and provides a statistically efficient procedure in random-effect models such as hierarchical GLMs (HGLMs; Lee and Nelder, 1996; Ha, 2008; Kim *et al.*, 2011) and frailty models (Lee *et al.*, 2006; Ha *et al.*, 2010). The proposed method is illustrated using simulation studies and a well-known survival data, i.e. recurrent infection time data from chronic granulomatous disease (CGD) patients (Fleming and Harrington, 1991). We also demonstrate that our method gives similar results to the penalized marginal likelihood method (Fan and Li, 2002).

The paper is organized as follows. In Section 2 we briefly describe the gamma frailty models. In Section 3 we outline standard estimation procedures and then propose a penalized h-likelihood procedure. Simulation study and data analysis are provided in Section 4. Finally, discussion is given in Section 5.

2. Gamma frailty models

Let T_{ij} $(i = 1, \cdots, q,\ j = 1, \cdots, n_i,\ n = \sum_i n_i)$ be time-to-event (i.e. survival time) for the j th observation of the i th cluster (or subject) and C_{ij} be the corresponding censoring time. Then we have the observable random variables:

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

where $I(\cdot)$ is the indicator function. Denote by u_i the unobserved frailty (or random effect) for the i th cluster. Gamma frailty models are described as follows. Given u_i, the conditional hazard function of T_{ij} takes the form

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

where $\lambda_0(t)$ is an unspecified baseline hazard function, β is a $d \times 1$ vector of unknown regression parameters corresponding to fixed covariates $x_{ij} = (x_{ij1}, \cdots, x_{ijd})^{\mathrm{T}}$. The distribution of frailties u_i's is assumed to follow a gamma distribution with mean $E(u_i) = 1$ and var $(u_i) = \alpha$ (Hougaard, 2000; Ha *et al.*, 2001). Here, the variance α is called frailty or dispersion parameter. The gamma frailty gives a mathematical simplicity; for other frailty distributions see Duchateau and Janssen (2008) and Ha (2011). Notice that the model (2.1) becomes Cox's PH model if $\alpha = 0$ (i.e. $u_i \equiv 1$).

The functional form of $\lambda_0(t)$ in (2.1) is unknown. Following Breslow's (1972) idea, we consider the baseline cumulative hazard function $\Lambda_0(t)$ to be a step function with jumps at the r distinct observed event times (Ha *et al.*, 2001),

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

where $y_{(k)}$ is the k th $(k = 1, \cdots, r)$ smallest distinct event time among the y_{ij}'s, and $\lambda_{0k} = \lambda_0(y_{(k)})$.

3. Penalized h-likelihood approach

We firstly outline standard h-likelihood and marginal likelihood procedures for gamma frailty models (2.1). Then we propose a penalized h-likelihood procedure for the variable selection using the SCAD function.

3.1. Standard estimation procedures

Following Ha *et al.* (2001) and Ha and Lee (2003), the h-likelihood for the gamma frailty model (2.1) is defined by

$$h = h(\beta, \lambda_0, \alpha) = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \tag{3.1}$$

where

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

$\ell_{1ij} = \ell_{1ij}(\beta, \lambda_0; y_{ij}, \delta_{ij}|u_i)$ is the logarithm of the conditional density function for y_{ij} and δ_{ij} given u_i,

$$\ell_{2i} = \ell_{2i}(\alpha; u_i) = \alpha^{-1}(v_i - u_i) + c(\alpha)$$

is the logarithm of the density function for $v_i = \log(u_i)$ with parameter α. Here $\lambda_0 = (\lambda_{01}, \cdots, \lambda_{0r})^T$, $\eta_{ij} = x_{ij}^T\beta + v_i$, $d_{(k)}$ is the number of deaths at $y_{(k)}$, $R_{(k)} = R(y_{(k)}) = \left\{(i,j) : y_{ij} \geq y_{(k)}\right\}$ is the risk set at $y_{(k)}$, and $c(\alpha) = -\log\Gamma(\alpha^{-1}) - \alpha^{-1}\log(\alpha)$.

Since the dimension of λ_0 increases with sample size n, for the estimation of (β, v) with $v = (v_1, \cdots, v_q)^T$ Ha *et al.* (2001) proposed to the use of the profile h-likelihood h^* with λ_0 eliminated:

$$h^* = h|_{\lambda_0=\widehat{\lambda_0}} = \sum_{ij} \delta_{ij}\eta_{ij} - \sum_{k} d_k \log \left\{ \sum_{(i,j)\in R_{(k)}} \exp(\eta_{ij}) \right\} + \sum_{i} \ell_{2i}, \tag{3.2}$$

where

$$\widehat{\lambda_{0k}} = \frac{d_{(k)}}{\sum_{(i,j)\in R(y_{(k)})} \exp(x_{ij}^T\beta + v_i)}$$

are solutions of the estimating equations, $\partial h/\partial\lambda_{0k} = 0$, for $k = 1, \cdots, r$. With h^* in (3.2) we estimate fixed parameters (β, α) and random effects v as follows. Ha *et al.* (2001) further showed that given α the estimation of $\tau = (\beta, v)^T$ is obtained by solving

$$\partial h^*/\partial\tau = (\partial h/\partial\tau)|_{\lambda_0=\widehat{\lambda}_0} = 0. \tag{3.3}$$

Ha and Lee (2003) showed that given α, the equations (3.3) lead to the maximum h-likelihood estimator (MHLE) or HGLM-type score equations:

$$\begin{pmatrix} X^TW^*X & X^TW^*Z \\ Z^TW^*X & Z^TW^*Z+U \end{pmatrix}\begin{pmatrix} \widehat{\beta} \\ \widehat{v} \end{pmatrix} = \begin{pmatrix} X^Tw^* \\ Z^Tw^*+R \end{pmatrix}, \tag{3.4}$$

where X is the $n \times d$ matrix whose i th row vector is x_{ij}^T, Z is the $n \times q$ group indicator matrix whose i th row vector is z_{ij}^T, $W^* = W^*(\beta, v)$ is the $n \times n$ symmetric matrix given in Appendix 2 of Ha and Lee (2003), $U = \text{diag}\left\{-\partial^2\ell_{2i}/\partial v_i^2\right\}$ is the $q \times q$ diagonal matrix, $w^* = W^*\eta + (\delta - \mu)$, $\mu = \Lambda_0 \exp(\eta)$, $R = Uv + (\partial\ell_2/\partial v)$. The square matrix of left-hand side of (3.4) is the negative Hessian matrix, $H^* = -\partial^2h^*/\partial\tau^2$, leading to an asymptotic variance of $\widehat{\beta}$ (Lee and Nelder, 1996; Ha and Lee, 2003):

$$\text{var}(\widehat{\beta}) = (X^TV^{-1}X)^{-1} \text{ with } V = W^{*-1} + ZR^{-1}Z^T. \tag{3.5}$$

Next, for the estimation of the gamma-frailty parameter α we use the second-order Laplace approximation (Lee and Nelder, 2001) to marginal likelihood (3.8), defined by

$$s_v(h) = p_v(h) - F(h)/24, \tag{3.6}$$

where $p_v(h) = [h - (\log\det H(h,v)/(2\pi))/2]|_{v=\widehat{v}}$, $H(h,v) = -\partial^2h/\partial v^2$, $\widehat{v}$ solves $\partial h/\partial v = 0$ and $F(h) = \sum_i -2(\alpha^{-1} + \delta_{i+})^{-1}$. The maximum adjusted h-likelihood estimator for α is obtained by solving

$$\partial s_v(h)/\partial\alpha = 0. \tag{3.7}$$

In summary, the estimates $(\widehat{\tau}, \widehat{\alpha})$ are obtained by the alternation between the two estimating equations (3.4) and (3.7) until convergence is achieved. On the other hand, 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 h-likelihood in (3.1):

$$m = m(\lambda_0, \beta, \alpha) = \sum_i \log\left\{\int \exp(h_i)du_i\right\}, \tag{3.8}$$

where $h_i = \sum_j \ell_{1ij} + \ell_{2i}$. For the gamma frailty models (2.1), from (3.1) we have an explicit marginal likelihood (Hougaard, 2000):

$$\begin{aligned} m(\lambda_0, \beta, \alpha) &= \sum_{ij} \delta_{ij}\left\{x_{ij}^T\beta + \log\lambda_0(y_{ij})\right\} \\ &\quad + \sum_i \left\{-(\alpha^{-1} + \delta_{i+})\log(\alpha^{-1} + \mu_{i+}) + \log\Gamma(\alpha^{-1} + \delta_{i+}) + c(\alpha)\right\} \end{aligned} \tag{3.9}$$

where $\delta_{i+} = \sum_j \delta_{ij}$, $\mu_{i+} = \sum_j \Lambda_0(y_{ij})\exp(x_{ij}^T\beta)$, $c(\alpha) = -\log\Gamma(\alpha^{-1}) - \alpha^{-1}\log\alpha$. The maximum likelihood estimators are obtained by maximizing the marginal likelihood (3.9) either directly or by using the expectation maximization (EM) algorithm (Ha *et al.*, 2010). For details see Nielsen *et al.* (1992) and Andersen *et al.* (1997).

3.2. Penalized h-likelihood procedure

Consider a penalized h-likelihood using h^* in (3.2):

$$h_p(\beta, v, \alpha) = h^* - n\sum_{j=1}^{d} p_\lambda(|\beta_j|). \tag{3.10}$$

Here $p_\lambda(|\beta|)$ is a penalty function with tuning parameter λ. For example, $p_\lambda(|\beta|) = \lambda|\beta|^p$ is called L_p penalty; it becomes L_1 penalty (i.e. LASSO penalty; Tibshirani, 1996) if $p = 1$, and L_2 penalty if $p = 2$ which leads to the ridge regression. Fan and Li (2001) proposed the use of SCAD penalty, defined by

$$p'_\lambda(|\beta|) = \lambda I(|\beta| \le \lambda) + \frac{(a\lambda - |\beta|)_+}{a-1} I(|\beta| > \lambda), \tag{3.11}$$

where x_+ denotes the positive part of x ; x_+ is x if $x > 0$, zero otherwise. A good penalty function should produce estimates that satisfy unbiasedness, sparsity and continuity (Fan and Li, 2001). However, all of L_p penalty do not simultaneously satisfy these three properties. Fan and Li (2001) showed that the SCAD penalty satisfies all three properties and that it can perform well as the oracle procedure in terms of selecting the correct subset model and estimating the true non-zero coefficients. Thus, in this paper we use the SCAD penalty (3.11) with a reasonable choice of $a = 3.7$ (Fan and Li, 2001) in oder to maximize (3.10) with respect to (β, v) given α. In particular, the estimating equations for β_j $(j = 1, \cdots, d)$ are given by

$$\partial h_p/\partial\beta_j = (\partial h/\partial\beta_j)|_{\lambda_0=\widehat{\lambda}_0} - np'_\lambda(|\beta_j|)\mathrm{sgn}(|\beta_j|) = 0$$

for $\beta_j \ne 0$. However, the SCAD functions become non-differentiable at the origin and they do not have continuous second-order derivatives. These make it difficult to maximize the penalized h-likelihood (3.10). Thus, we use a local quadratic approximation (Fan and Li, 2001) to the SCAD penalty (3.11). Then we can show that given α, the maximum penalized h-likelihood estimators for (β, v) are obtained from the following score equations:

$$\begin{pmatrix} X^TW^*X + n\Sigma_\lambda & X^TW^*Z \\ Z^TW^*X & Z^TW^*Z + U \end{pmatrix} \begin{pmatrix} \widehat{\beta} \\ \widehat{v} \end{pmatrix} = \begin{pmatrix} X^Tw^* \\ Z^Tw^* + R \end{pmatrix} \tag{3.12}$$

with $\Sigma_\lambda = \mathrm{diag}{p_\lambda}'(|\beta_j|)/|\beta_j|$, which is an extension of the standard score equations (3.4). The standard errors of $\widehat{\beta}$ are obtained from a sandwich formula (Fan and Li, 2001):

$$\mathrm{cov}(\widehat{\beta}) = \{H_{\beta\beta} + n\Sigma_\lambda\}^{-1}\,\mathrm{cov}(\partial\widehat{h}/\partial\beta)\,\{H_{\beta\beta} + n\Sigma_\lambda\}^{-1}, \tag{3.13}$$

where $H_{\beta\beta} = -\partial^2\widehat{h}/\partial\beta^2$ with $\widehat{h} = h^*|_{v=\widehat{v}}$. Here, we use $H_{\beta\beta}$ for $\mathrm{cov}(\partial\widehat{h}/\partial\beta)$. For the estimation of α, we use $s_v(h_p) = s_v(h) - n\sum_{j=1}^d p_\lambda(|\beta_j|)$ as in the penalized marginal likelihood (PML), $m_p(\beta, \alpha) = m - n\sum_{j=1}^d p_\lambda(|\beta_j|)$, by Fan and Li (2002). That is, the corresponding maximized PML estimators are obtained by maximizing m_p.

Furthermore, for the estimation of tuning parameter λ we use a generalized cross-validation (GCV) statistic (Craven and Wahba, 1979; Cho *et al.*, 2010), given by

$$\mathrm{GCV}(\lambda) = -s_v(h)/n\,\{1 - e(\lambda)/n\},$$

where $e(\lambda) = \mathrm{tr}[\{H_{\beta\beta} - \Sigma_\lambda\}^{-1} H_{\beta\beta}]$ is the effective number of parameters. Note that $\widehat{\lambda} = \mathrm{argmin}_\lambda\{\mathrm{GCV}(\lambda)\}$ is selected using a simple grid method as in Fan and Li (2002).

In summary, in the inner loop, we maximize h_p for (β, v) and $s_v(h_p)$ for α, respectively. In the outer loop, we find λ that minimizes $\mathrm{GCV}(\lambda)$.

4. Illustration

4.1. Simulation study

Following Fan and Li's (2002) scheme, simulated studies, based upon 100 replications of simulated data, are presented to evaluate the performance of the proposed method. We generate data from the gamma frailty model (2.1) assuming the exponential baseline hazard $\lambda_0(t) = 1$, regression parameters $\beta = (0.8, 0, 0, 1, 0, 0, 0.6, 0)$ and the variance $\alpha \equiv \sigma^2 = 0.5$. Here, covariates $x = (x_1, x_2, \cdots, x_8)^T$ are also generated from AR(1) structure with a correlation $\rho = 0.5$. Note that x_1, x_4 and x_7 are important covariates. We also set the sample size $n = \sum_{i=1}^q n_i = 50, 100, 200$ with $n_i = 2$ for all i. The corresponding censoring times C_{ij} are generated from an exponential distribution with parameter empirically determined to achieve approximately the right censoring rate, around 30%.

For the 100 replications we computed the average number of zero coefficients of $\widehat{\beta}$. We also computed the mean, standard deviation (SD), and standard error (SE) which is obtained from the sandwich formula (3.13). Here, the mean and SD are the average and standard deviation of estimates of $\widehat{\beta}$ in 100 simulations, respectively. The SE is also the average of 100 estimated standard errors for $\widehat{\beta}$. For more details of the computation see Ha and Lee (2003) and Ha *et al.* (2011). For the frailty parameter σ^2 the corresponding mean and SD are also given. For comparison, we include Fan and Li's (2002) penalized marginal likelihood method. For the model fitting and computation we used SAS/IML.

The simulation results for zero coefficients and parameter estimates are summarized in Table 4.1. Here, the column labeled "C" indicates the average number of coefficients, of the five true zeros, correctly set to zero, and "IC" indicates the average number of the three non-zeros incorrectly set to zero. The proposed h-likelihood method works well with sample size n and it gives very similar results to the marginal likelihood method (Fan and Li, 2002). Furthermore, in Table 4.1 SD is the estimates of the true $\left\{\mathrm{var}(\widehat{\beta})\right\}^{1/2}$ and SE is the average of 100 estimated standard-errors for nonzero-coefficient estimates $\widehat{\beta} = (\widehat{\beta_1}, \widehat{\beta_4}, \widehat{\beta_7})^T$ (Ha and Lee, 2003). Our SE is underestimated as compared to SD under a smaller sample $n = 50$, but it is largely improved when n increases as in $n = 100$ and $n = 200$ in terms of a good agreement between SE and SD.

4.2. A practical example

The chronic granulomatous disease (CGD) data (Fleming and Harrington, 1991) relate to a placebo-controlled randomized trial of gamma interferon in chronic granulomatous disease. The trial aimed to investigate the effectiveness of gamma interferon (gamma-IFN) in reducing the rate of serious infections in CGD patients. In total, 128 patients were followed for about 1 year. Of the 63 patients in the treatment group, 14 patients experienced at least

Table 4.1 Simulations results using the SCAD under gamma frailty model

n	Method	Zeros		$\widehat{\beta}_1$		$\widehat{\beta}_4$		$\widehat{\beta}_7$		$\widehat{\sigma^2}$	
		C	IC	Mean	SD(SE)	Mean	SD(SE)	Mean	SD(SE)	Mean	SD
True		5	0	0.8		1		0.6		0.5	
50	PHL	4.08	0.04	0.79	0.246(0.196)	1.08	0.289(0.244)	0.61	0.285(0.212)	0.25	0.285
	PML	4.06	0.05	0.79	0.246(0.196)	1.06	0.285(0.242)	0.60	0.259(0.210)	0.24	0.272
100	PHL	4.27	0.02	0.78	0.151(0.134)	0.98	0.191(0.161)	0.59	0.187(0.144)	0.37	0.259
	PML	4.27	0.02	0.78	0.153(0.135)	0.98	0.194(0.161)	0.59	0.188(0.145)	0.37	0.258
200	PHL	4.54	0	0.81	0.121(0.098)	1.00	0.124(0.117)	0.60	0.115(0.105)	0.49	0.210
	PML	4.54	0	0.81	0.121(0.098)	1.00	0.124(0.117)	0.60	0.118(0.106)	0.49	0.209

Note: The simulation is conducted with 100 replications for gamma frailty model assuming the true regression parameter $\beta = (0.8, 0, 0, 1, 0, 0, 0.6, 0)$ and frailty variance $\sigma^2 = 0.5$, with sample sizes $n = 50$, 100 , 200 with $n_i = 2$. PHL, penalized h-likelihood method; PML, penalized marginal likelihood method; C, average number of coefficients, of the five true zeros, correctly set to zero; IC, average number of the three non-zeros incorrectly set to zero; SD, standard deviation of estimates in 100 simulations; SE, average of 100 estimated standard errors.

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.

The survival times are the gap times for recurrent infection time data. Censoring occurred at the last observation for all patients, except one, who experienced a serious infection on the date he left the study. In the CGD study about 63% of the data were censored. The survival times for a given patient are likely to be correlated due to recurrent times of the same patients. Thus we model the recurrent infection times using the gamma frailty (2.1) and the nine fixed covariates x_{ij}'s: treatment (0=placebo, 1=gamma-IFN), pattern of inheritance (0=autosomal recessive, 1=X-linked); age (in years); height (in cm); weight (in kg); use of corticosteroids at time of study entry (0=no, 1=yes); use of prophylactic antibiotics at time of study entry (0=no, 1=yes); sex (0=male, 1=female), and hospital region (0=U.S., 1=Europe).

Table 4.2 Variable selection for CGD data using gamma frailty model

Variable	SHL		PHL		PML	
	Est	(SE)	Est	(SE)	Est	(SE)
Gamma-IFN	-1.140	(0.345)	-1.069	(0.335)	-1.071	(0.328)
Inheritance	-0.792	(0.384)	-0.817	(0.378)	-0.810	(0.370)
Age	-0.099	(0.048)	-0.030	(0.013)	-0.030	(0.013)
Height	0.011	(0.014)	0	(0)	0	(0)
Weight	0.010	(0.021)	0	(0)	0	(0)
Corticosteroids	2.478	(1.020)	2.046	(0.948)	2.005	(0.931)
Prophylactic	-0.759	(0.481)	-0.840	(0.455)	-0.829	(0.445)
Sex	-0.950	(0.558)	-1.048	(0.540)	-1.029	(0.532)
Hospital region	-0.827	(0.412)	-0.880	(0.408)	-0.872	(0.401)
Frailty (σ^2)	0.884	(0.571)	0.902	(0.604)	0.792	(0.522)

Note: Est, estimate; SHL, standard h-likelihood method;
PHL: penalized h-likelihood method; PML: penalized marginal likelihood method.

Using the penalized h-likelihood and marginal likelihood procedures in Section 3.2, the tuning parameter λ, selected by the corresponding GCV, is all 0.13. The estimates (PHL and PML) corresponding to the selected λ are given in Table 4.2. From Table 4.2, the t-statistic (=estimate/SE) of standard h-likelihood method (SHL) indicates that the effects of 4 covariates (i.e. Height, Weight, Prophylactic and Sex) are not significant; particularly, the effects of Height and Weight are very non-significant. As expected, we find that the proposed method (PHL) chooses 7 out of 9 covariates and that the non-zero estimates of PHL are overall similar to those of SHL. The PHL for $\widehat{\beta}$ also gives similar results to the PML, which confirms the simulation results of Section 4.1. However, we observe that the estimates of frailty parameters (σ^2) of PHL method and their SEs are slightly different from those of PML; a possible reason is that the estimates of β are somewhat robust to those of σ^2 (Fan and Li, 2002). Here, the SEs of both PHL and PML for σ^2 are obtained from the inversion of Hessian matrices $-\partial^2 p_v(h)/\partial(\beta,\lambda_0,\sigma^2)^2$ and $-\partial^2 m/\partial(\beta,\lambda_0,\sigma^2)^2$, respectively (Ha *et al.*, 2010).

4.3. Discussion

We have proposed a variable selection method using a penalized h-likelihood with SCAD penalty under gamma frailty models. We have demonstrated via numerical studies that our method performs well and found out that it gives about the same results as the penalized marginal likelihood approach (Fan and Li, 2002). These facts are not surprising because the adjusted profile h-likelihood $s_v(h)$ in (3.6) is the second-order Laplace approximation to marginal likelihood m in (3.8): see also Ha and Lee (2003). The proposed method can be extended to various frailty models such as frailty models with log-normal distribution (Ha *et al.*, 2001; Ha, 2009) and frailty models with correlated distribution (Duchateau and Janssen, 2008; Hong, 2010; Ha *et al.*, 2011) because the h-likelihood avoids integration itself. However, the corresponding marginal likelihood approach is computationally intensive because it requires very intractable integrations. We used only SCAD penalty in the h-likelihood framework. However, our method can be applied to other penalty functions such as LASSO, ridge or adaptive LASSO. The comparison with our method would be an interested further work.

References

Andersen, P. K., Klein, J. P. and Zhang M.-J. (1999). Testing for centre effects in multi-centre survival studies: A monte carlo comparison of fixed and random effects tests. *Statistics in Medicine*, **18**, 1489-1500.

Breslow, N. E.(1972). Discussion of professor Cox's paper. *Journal of the Royal Statistical Society B*, **34**, 216-7.

Cho, D., Shim, J and Seok, K. H. (2010). Doubly penalized kernel method for heteroscedastic autoregressive data. *Journal of Korean Data & Information Science Society*, **21**, 155-162.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, **31**, 377-403.

Duchateau, L. and Janssen, P. (2008). *The frailty model*, Springer-Verlag, New York.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348-1360.

Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics*, **30**, 74-99.

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

Ha, I. D. (2008). A HGLM framework for meta-analysis of clinical trials with binary outcomes. *Journal of Korean Data & Information Science Society*, **19**, 1429-1440.

Ha, I. D. (2009). A correction of SE from penalized partial likelihood in frailty models. *Journal of Korean Data & Information Science Society*, **20**, 895-903.

Ha, I. D. (2011). Inference for heterogeneity of treatment effect in multi-center clinical trial. *Journal of Korean Data & Information Science Society*, **22**, 605-612.

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

Ha, I.D., Sylvester, R., Legrand, C. and MacKenzie, G. (2011). Frailty modelling for survival data from multi-centre clinical trials. *Statistics in Medicine*, **30**, 2144-2159.

Hong, Y. W. (2010). Failure rate of a bivariate exponential distribution. *Journal of Korean Data & Information Science Society*, **21**, 173-177.

Hougaard, P. (2000). *Analysis of multivariate survival data*, Springer-Verlag, New York.

Kim, H. K., Noh, M. and Ha, I. D. (2011). A study using HGLM on regional difference of the dead due to injuries. *Journal of Korean Data & Information Science Society*, **22**, 137-148.

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. and Nelder, J. A. and Pawitan, Y. (2006). *Generalised linear models with random effects: Unified analysis via h-likelihood,* Chapman and Hall, London.

Lu, W. and Zhang, H. H. (2007). Variable selection for proportional odds model. *Statistics in Medicine*, **26**, 3771-3781.

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.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B*, **58**, 267-288.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418-1429.