Journal of the Korean Data & Information Science Society
2015, **26**(6), 1513–1521

http://dx.doi.org/10.7465/jkdi.2015.26.6.1513
한국데이터정보과학회지

Variable selection in Poisson HGLMs using h-likelihoood†

Il Do Ha[1] · Geon-Ho Cho[2]

[1]Department of Statistics, Pukyong National University
[2]Faculty of Medical Industry Convergence, Daegu Haany University

Received 23 July 2015, revised 24 August 2015, accepted 17 September 2015

Abstract

Selecting relevant variables for a statistical model is very important in regression analysis. Recently, variable selection methods using a penalized likelihood have been widely studied in various regression models. The main advantage of these methods is that they select important variables and estimate the regression coefficients of the covariates, simultaneously. In this paper, we propose a simple procedure based on a penalized h-likelihood (HL) for variable selection in Poisson hierarchical generalized linear models (HGLMs) for correlated count data. For this we consider three penalty functions (LASSO, SCAD and HL), and derive the corresponding variable-selection procedures. The proposed method is illustrated using a practical example.

Keywords: LASSO, penalized h-likelihood, Poisson HGLMs, SCAD, variable selection.

1. Introduction

In regression analysis, determining relevant variables for a statistical model with a large number of covariates is very important. Recently, variable selection methods using a penalized likelihood have been widely studied in various regression models such as linear models and generalized linear models. The main advantage of these methods is that they select important variables and estimate the regression coefficients of the covariates, simultaneously; i.e. they delete insignificant variables by estimating their coefficients as zero. For example, such methods include the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), smoothly clipped absolute deviation (SCAD; Fan and Li, 2001), elastic net (Zou and Hastie, 2005), adaptive-LASSO (Zou, 2006), and h-likelihood penalty (HL; Lee and Oh, 2014), etc.

Very recently, Ha *et al.* (2014) developed a penalized h-likelihood procedure for variable selection of fixed effects in semi-parametric frailty models. Here they considered three penalty functions (LASSO, SCAD and HL). In this paper we propose a simple variable-selection procedure in Poisson HGLMs (Lee and Nelder, 1996) using Ha *et al.*'s (2014) method. Here,

† This work was supported by Research Grant of Pukyong National University (2014 year).
[1] Corresponding author: Professor, Department of Statistics, Pukyong National University, Busan 608-737, Korea. E-mail: idha1353@pknu.ac.kr
[2] Professor, Faculty of Medical Industry Convergence, Daegu Haany University, Gyeongsan 712-715, Korea.

the Poisson HGLMs with random effects are very useful for analyzing correlated count data. For the distribution of random effects we use a normal distribution, which is useful for modeling of multi-component or correlated random effects (Ha *et al.*, 2014). For the variable selection we use the hierarchical likelihood (h-likelihood; Lee and Nelder, 1996) as in Ha and Cho (2012) and Ha *et al.* (2014). The h-likelihood avoids the need for the marginalization over the random-effect distribution and provides a statistically efficient procedure in various random-effect models such as HGLMs and frailty models (Rondeau *et al.*, 2008; Ha *et al.*, 2011), but the marginal likelihood often requires the computation of intractable integrals when eliminating the random effects, particularly for normally distributed random effects.

The penalty functions are very important in conducting the variable selection. The SCAD penalty provides good properties such as oracle property, while the HL penalty is unbounded at the origin and gives a very good performance in various high dimensional problems (Lee and Oh, 2014). Note that the SCAD penalty method leads to an oracle maximum likelihood (ML) estimator, whereas the HL penalty approach gives an oracle shrinkage estimator (Kwon *et al.*, 2014). In other words, an oracle ML estimator is the ML estimator when all covariates with nonzero coefficients are known. Fan and Peng (2004) showed that a local solution of the SCAD penalty is asymptotically equivalent to an oracle ML estimator. Similarly, an oracle shrinkage estimator is the shrinkage estimator when all covariates with nonzero coefficients are known. Kwon *et al.* (2014) showed that a local solution for the HL penalty is an oracle shrinkage estimator. It is well known that shrinkage estimations would be preferred for prediction (Efron and Morris, 1975; Lee and Nelder, 2006). Ha *et al.* (2014) have showed via simulations that the HL has higher probability of choosing the true model than the LASSO and SCAD methods without losing prediction accuracy.

In this paper, we derive the variable-selection procedure for Poisson HGLMs. The proposed method is illustrated using a well-known practical example with epilepsy seizure count data (Thall and Vail, 1990). The paper is organized as follows. In Section 2 we describe Poisson HGLMs and the corresponding h-likelihood. In Section 3 we review various penalty functions, and we derive a penalized h-likelihood procedure for variable selection using Ha *et al.*'s (2014) method. A practical example with the epilepsy count data is used to illustrate our method in Sections 4. Finally, further discussions are given in Section 5.

2. Poisson HGLMs and h-likelihood

We describe a formulation of Poisson HGLMs and then outline the construction of h-likelihood. Let y_{ij} $(i = 1, \ldots, q,\ j = 1, \ldots, n_i,\ n = \sum_i n_i)$ be the response count variable for the jth observation in the ith cluster (or subject). Denote by v_i an unobserved random effect associated with the ith cluster. We assume that given v_i, y_{ij} follows a Poisson distribution with

$$\log(\mu_{ij}) = \eta_{ij}, \tag{2.1}$$

where $\mu_{ij} = E(y_{ij}|v_i)$ is the conditional mean of y_{ij} given v_i,

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

is the linear predictor, and $x_{ij} = (1, x_{ij1}, \ldots, x_{ijp})^T$ is $(p + 1) \times 1$ covariate vectors corresponding to fixed effects $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$. We assume that the random effect v_i are independent and follow a distribution with dispersion parameter θ. Although the results of this paper can be extended to non-normal random effect (e.g. log-gamma random effect), for simplicity, we assume a normal distribution,

$$v_i \sim N(0, \theta),$$

which is useful for modelling multi-components or correlated random effects (Lee and Nelder, 1996; Ha *et al.*, 2007, 2011).

Along the lines of Lee and Nelder (1996), the hierarchical log-likelihood (h-likelihood) for the Poisson HGLMs (2.1) is defined by

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

where

$$\sum_{ij} \ell_{1ij} = \sum_{ij} \{y_{ij} \log(\mu_{ij}) - \mu_{ij} - \log \Gamma(y_{ij} + 1)\}$$

with $\mu_{ij} = \exp(x_{ij}^T \beta + v_i)$, $\ell_{1ij} = \ell_{1ij}(\beta; y_{ij}|v_i)$ is the logarithm of the conditional density function for y_{ij} given v_i, and $\ell_{2i} = \ell_{2i}(\theta; v_i)$ is the logarithm of the density function for normally distributed random effect v_i with parameter θ, given by

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

The h-likelihood method avoids intractable integrations necessary in computing the marginal likelihood eliminating the random effects, and provides a statistically efficient and unified inference procedure for various random-effect models (Lee *et al.*, 2006; Ha *et al.*, 2007, 2014). In particular, it gives statistical justifications for inferences of random effects and also useful materials for data analysis (Ha and Noh, 2013; Paik *et al.*, 2015).

3. Variable selection procedure

3.1. Penalty function for variable selection

Following Ha *et al.* (2014), we consider variable selection of fixed effects β in the Poisson HGLM (2.1) via maximization of a penalized h-likelihood, ℓ_p, using h in (2.2) and a penalty; it is defined by

$$\ell_p(\beta, v, \theta) = h - n \sum_{j=1}^{p^*} J_\gamma(|\beta_j|), \tag{3.1}$$

where $p^* = p + 1$ is the number of fixed effects (i.e. regression parameters), $J_\gamma(|\cdot|)$ is a penalty function that controls model complexity using the tuning parameter γ. Here, we don't impose any penalty on the dispersion parameter θ. Typically, setting $\gamma = 0$ results in the standard Poisson HGLM, whereas the regression coefficient estimates $\hat{\beta}$ tend to 0 as $\gamma \to \infty$. That is, a larger value of γ tends to choose a simple model, whereas a smaller value of γ inclines to a complex model (Fan and Lv, 2010). A method for choosing an optimal value of γ will be presented later.

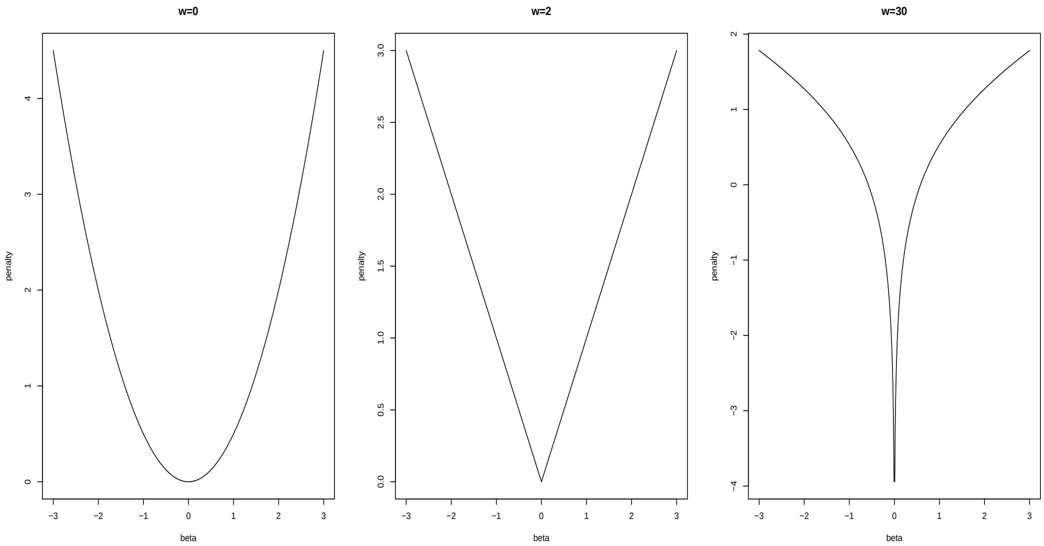

Figure 3.1 HL penalties with $w = 0, 2$ and 30

Recently, various penalty functions have been used in the literature on variable selection in statistical models including the GLMs and Cox proportional hazards models. As in Ha *et al.* (2014), we consider the following three penalty functions (LASSO, SCAD and HL). However, our results can be applied to other penalty functions which are not discussed here.

(i) LASSO (Tibshirani, 1996):

$$J_\gamma(|\beta|) = \gamma|\beta|, \tag{3.2}$$

(ii) SCAD (Fan and Li, 2001):

$$J'_\gamma(|\beta|) = \gamma I(|\beta| \le \gamma) + \frac{(a\gamma - |\beta|)_+}{a-1} I(|\beta| > \gamma), \tag{3.3}$$

where x_+ denotes the positive part of x; i.e., x_+ is x if $x > 0$, zero otherwise.

(iii) HL (Lee and Oh, 2014):

$$J_\gamma(|\beta|) \equiv J_{(a,w)}(|\beta|) = \log\Gamma(1/w) + \frac{\log w}{w} + \frac{\beta^2}{2au(|\beta|)} + \frac{(w-2)\log u(|\beta|)}{2w} + \frac{u(|\beta|)}{w}, \tag{3.4}$$

where $u(|\beta|) = [\{8w\beta^2/a + (2-w)^2\}^{1/2} + 2 - w]/4$.

It is well known that a good penalty function should produce estimates that satisfy the oracle properties (unbiasedness, sparsity and continuity): Fan and Li (2001). The LASSO in (3.2) is the most common penalty as L_1 penalty, but it does not simultaneously satisfy these three properties. Fan and Li (2001) showed that the SCAD penalty satisfies all the 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, simultaneously. Fan and Li (2001, 2002) have also shown that the choice of $a = 3.7$ in (3.3) performs well in a variety of situations; here we also follow their suggestion by taking $a = 3.7$.

Very recently, Lee and Oh (2014) proposed a new penalty unbounded at the origin within the framework of a random effect model. The new unbounded HL penalties in (3.4), $J_\gamma(|\beta|)$, at various values of $w = 0, 2$ and 30 and $a = 1$ are displayed in Figure 3.1. The form of the penalty changes from a quadratic shape ($w = 0$) for ridge regressions to a cusped form ($w = 2$) for LASSO and then to an unbounded form ($w > 2$) at the origin. In the case of $w > 2$, it allows an infinite gain at zero.

Notice that the SCAD method provides oracle ML estimators (least squares estimators), while the HL approach gives oracle shrinkage estimators (Kwon *et al.*, 2014). When there is a multi-collinearity, shrinkage estimation is better than the ML estimation. Lee *et al.* (2010; 2011a, 2011b) have shown its advantages of HL method over LASSO and SCAD methods, especially when the number of covariates is larger than the sample size (i.e. $p^* > n$); it actually has a better property for variable selection without losing prediction power. Since a in (3.4) has a greater sensitivity to change of penalty than w, we consider only a few of values for w, e.g., $w = 2.1, 3, 10, 30, 50$ representing small, medium and large values of w. Thus, we conduct a simultaneous choice of w and a in this sense.

3.2. Penalized h-likelihood procedure

We derive a penalized h-likelihood procedure for variable selection in Poisson HGLM (2.1), using Ha *et al.*'s (2014) method for frailty models. The variable selection procedure based on the penalized h-likelihood ℓ_p in (3.1) is as follows:

1) **Estimation of** (β, v)**:** Given the dispersion parameter θ, the maximum penalized h-likelihood (MPHL) estimates of (β, v) are obtained by solving the joint estimating equations of β and v, $\partial\ell_p/\partial(\beta, v) = 0$. Following Lee and Nelder (1996) and Ha *et al.* (2014), we can show that the joint equations can be explicitly expressed as a simple matrix form:

$$\begin{pmatrix} X^TWX + n\Sigma_\gamma & X^TWZ \\ Z^TWX & Z^TWZ + U \end{pmatrix} \begin{pmatrix} \widehat{\beta} \\ \widehat{v} \end{pmatrix} = \begin{pmatrix} X^TWw \\ Z^TWw \end{pmatrix}, \tag{3.5}$$

where X and Z are model matrices of fixed effects β and random effects $v = (v_1, \dots, v_q)^T$, respectively, and $w = \eta + (y - \mu)/\mu$ is the Poisson GLM adjusted dependent variable with $\mu = \exp(\eta)$ and $\eta = X\beta + Zv$, $W = \mathrm{diag}(\mu_{ij})$ with $\mu_{ij} = \exp(x_{ij}^T\beta + v_i)$ is a diagonal weight matrix, and $\Sigma_\gamma = \mathrm{diag}\{J'_\gamma(|\beta_j|)/|\beta_j|\}$.

The score equations (3.5) are extensions of existing estimation procedures. For example, under no-penalty (i.e., $\Sigma_\gamma = 0$) they become the standard score equations of Poisson HGLMs of Lee and Nelder (1996). For the GLM without random effect, they also reduce to

$$(X^TWX + n\Sigma_\gamma)\widehat{\beta} = X^TWw. \tag{3.6}$$

We thus see that the new equations (3.5) provide a general penalized equation (3.6) for the GLM, as in frailty models (Ha *et al.*, 2014).

2) **Estimation of** θ**:** For estimation of θ we use an adjusted profile h-likelihood $p_\tau(\ell_p)$ (Ha and Lee, 2003; Lee *et al.*, 2006) which eliminates (β, v) from ℓ_p in (3.1), defined by

$$p_\tau(\ell_p) = \left[\ell_p - \frac{1}{2}\log\det\left\{\frac{H(\ell_p;\tau)}{(2\pi)}\right\}\right]\Bigg|_{\tau=\hat{\tau}}, \tag{3.7}$$

where $\tau = (\beta^T, v^T)^T$, $\hat{\tau} = \hat{\tau}(\theta) = (\hat{\beta}^T(\theta), \hat{v}^T(\theta))^T$ and $H(\ell_p; \tau) = -\partial^2 \ell_p / \partial\tau\partial\tau^T$. The estimate of θ is obtained by solving the score equation $\partial p_\tau(\ell_p)/\partial\theta = 0$.

3) **Standard-error formula:** Following Fan and Li (2001) and Ha *et al.* (2014), an approximated standard error (SE) of $\hat{\beta}$ is obtained from a sandwich formula based on ℓ_p:

$$\text{cov}(\hat{\beta}) = (H_{\beta\beta} + n\Sigma_\gamma)^{-1} H_{\beta\beta} (H_{\beta\beta} + n\Sigma_\gamma)^{-1}, \tag{3.8}$$

where $H_{\beta\beta} = \{(X^T WX) - (X^T WZ)(Z^T WZ + U)^{-1}(Z^T WX)\}|_{v=\hat{v}}$.

4) **Tuning-parameter selection:** The performance of penalized likelihood methods depends on the choice of the tuning parameter in the penalty functions. For the choice of tuning parameters γ, a generalized cross-validation (GCV) statistic has been extensively used (Tibshirani, 1996, 1997; Fan and Li, 2001, 2002; Androulakis *et al.*, 2012). However, Wang *et al.* (2007) showed that the GCV approach can not select the tuning parameters satisfactorily, with a non-ignorable overfitting effect in the resulting model (Fan and Lv, 2010; Zhang *et al.*, 2010). Thus, they proposed to use a BIC-based selection criterion. Following Wang *et al.* (2007) and Ha *et al.* (2014), we use a BIC-type criterion based on the h-likelihood for selecting tuning parameters γ, defined by

$$\text{BIC}(\gamma) = -2p_v(h) + e(\gamma)\log(n), \tag{3.9}$$

where $e(\gamma) = \text{tr}[\{H_{\beta\beta} + n\Sigma_\gamma\}^{-1} H_{\beta\beta}]$ is the effective number of parameters (Lee and Nelder, 1996; Ha *et al.*, 2007). Note that $\hat{\gamma} = \text{argmin}_\gamma\{BIC(\gamma)\}$ is calculated using a simple grid search method as in Fan and Li (2002).

In summary, the variable-selection procedure above is easily implemented via a slight modification to the existing h-likelihood procedures (Lee *et al.*, 2006; Ha *et al.*, 2014). An outline of variable-selection algorithm can be described as follows.

1. In the inner loop, we maximize ℓ_p for $\tau = (\beta^T, v^T)$ (i.e., we solve (3.5)) and the adjusted profile h-likelihood $p_\tau(\ell_p)$ in (3.7) for θ, respectively.
2. In the outer loop, we find γ that minimizes BIC(γ) in (3.9).
3. At convergence, we compute the estimates of the SEs for $\hat{\beta}$ using (3.8).

Remark 3.1: Notice that to avoid a numerical difficulty in solving (3.5), we employ $\Sigma_{\gamma,\epsilon} = \text{diag}\{J'_\gamma(|\beta_j|)/(|\beta_j| + \epsilon)\}$ for a small positive value of ϵ, say, $\epsilon = 10^{-8}$, instead of Σ_γ. Then $\Sigma_{\gamma,\epsilon}$ is always defined (Lee and Oh, 2014). As long as ϵ is small, the diagonal elements of $\Sigma_{\gamma,\epsilon}$ are very close to those of Σ_γ. In fact, this algorithm is identical to that of Hunter and Li (2005) for improvement of a local quadratic approximation (Fan and Li, 2001): Johnson *et al.* (2008). Here, we report $\hat{\beta} = 0$ if all five printed decimals are zero. In case of the SCAD and HL penalties, there may exist several local maximizers. Thus, a good initial value is needed for getting a proper estimate $\hat{\beta}$. As in Ha *et al.* (2014), we use a LASSO solution as the initial value for the SCAD and HL penalties.

4. Illustration

We illustrate the proposed method using the epilepsy seizure count data from a clinical trial presented by Thall and Vail (1990). The data come from the randomized clinical trial conducted among patients suffering from simple or complex partial seizures to receive either the antiepileptic drug progabide or a placebo, as an adjuvant to standard chemotherapy. Primary outcome of interest (y) is the number of seizures occurring over the previous 2 weeks measured at each of four successive postrandomization clinic visits. The data consist of four repeated measures ($n_i = 4$) of $q = 59$ epileptic patients, with covariates Constant, Base (x_1), Trt (x_2, placebo=0, progabide=1), Base*Trt (x_3), Age (x_4) and Visit ($x_5 = 1$ for the fourth visit, $x_5 = 0$ otherwise). Here, Base is the logarithm of a quarter of the number of epileptic seizures recored in the 8-week period preceding the trial, and Age is the logarithm of age. The response variables (y) may be correlated due to four repeated measures of the same patient. We thus assume that $y_{ij}|u_i$ ($i = 1, \ldots, 59; j = 1, 2, 3, 4$) follow Poisson distribution with mean $\mu_{ij} = \exp(\eta_{ij})$ and

$$\eta_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + v_i \tag{4.1}$$

is the linear predictor and random effect $v_i \sim N(0, \theta)$.

We fitted the above Poisson HGLM (4.1) using the penalized h-likelihood procedure in Section 3. Here, the Base and Age covariates were standardized as other covariates are binary. For computations, we used SAS/IML. The selected values of the tuning parameters γ by the BIC (3.9) were 0.048, 0.082 and $(w, a) = (50, 0.017)$ for the LASSO, SCAD and HL, respectively. The estimates of the dispersion parameter θ for no-penalty, LASSO, SCAD and HL are 0.281, 0.324, 0.308 and 0.294, respectively. The estimated coefficients and their standard errors are shown in Table 4.1. The covariates, Base and Visit including Constant, are very significant in all the four methods. The three variable-selection methods select slight different variables. That is, the LASSO and HL choose four covariates (Constant, Base, Age, Visit) and (Constant, Base, Trt, Visit), respectively, whereas the SCAD chooses three covariates (Constant, Base, Visit). Here, the LASSO selects a non-significant covariate (i.e. Age) under no-penalty, as compared to other two methods (i.e., SCAD and HL). We find that the HL selects a significant covariate, Trt, but that other two methods (i.e., SCAD and HL) do not. This is expected from the fact that the HL has higher probability of choosing the true model than the LASSO and SCAD methods without losing prediction accuracy (Ha *et al.*, 2014).

Table 4.1 Case 1: estimated coefficients and standard errors (in parentheses) for Poisson HGLM with epilepsy seizure count data

Variable	No-penalty	LASSO	SCAD	HL
Constant	1.852 (0.110)	1.609 (0.078)	1.677 (0.080)	1.749 (0.086)
x_1: Base	0.651 (0.102)	0.668 (0.073)	0.735 (0.079)	0.726 (0.076)
x_2: Trt	-0.930 (0.416)	0 (0)	0 (0)	-0.145 (0.072)
x_3: Bxt	0.335 (0.212)	0 (0)	0 (0)	0 (0)
x_4: Age	0.103 (0.080)	0.001 (0.001)	0 (0)	0 (0)
x_5: Visit	-0.160 (0.055)	-0.119 (0.042)	-0.110 (0.037)	-0.137 (0.047)

In addition, we also fitted the above model including the interactions among Base, Trt, Age and Visit. The estimated results are summarized in Table 4.2. As expected, the HL selects the significant covariate (Trt), but the LASSO and SCAD still do not. Furthermore, the HL does not select the non-significant 4 interactions (x_6, x_7, x_8 and x_9), whereas the LASSO and SCAD select the two non-significant interactions, x_7 and x_8. That is, we again find that the HL selects well important variables, as compared to the two methods (LASSO and SCAD); this confirms such findings by Ha *et al.* (2014).

Table 4.2 Case 2: estimated coefficients and standard errors (in parentheses) for Poisson HGLM with epilepsy seizure count data

Variable	No-penalty	LASSO	SCAD	HL
Constant	1.831 (0.113)	1.605 (0.077)	1.671 (0.080)	1.749 (0.086)
x_1: Base	0.658 (0.104)	0.671 (0.072)	0.739 (0.079)	0.726 (0.076)
x_2: Trt	-0.907 (0.428)	0 (0)	0 (0)	-0.145 (0.072)
x_3: Bxt	0.341 (0.217)	0 (0)	0 (0)	0 (0)
x_4: Age	0.103 (0.122)	0 (0)	0 (0)	0 (0)
x_5: Visit	-0.063 (0.088)	-0.066 (0.033)	-0.071 (0.027)	-0.137 (0.047)
x_6: Trt*Age	0.027 (0.163)	0 (0)	0 (0)	0 (0)
x_7:Trt*Visit	-0.155 (0.118)	-0.077 (0.037)	-0.028 (0.012)	0 (0)
x_8: Base*Visit	-0.028 (0.053)	-0.014 (0.016)	-0.016 (0.011)	0 (0)
x_9: Age*Visit	-0.063 (0.061)	0 (0)	0 (0)	0 (0)

5. Discussion

In this paper, we have shown how to select important variables in Poisson HGLM through a penalized h-likelihood procedure. In Section 4 we have demonstrated via data analysis that the proposed procedure with the HL penalty performs well. An advantage of our method can be easily implemented by a slight modification to the existing HGLM estimation procedure (Lee *et al.*, 2006). Thus our method can be straightforwardly applied to variable selection in practical random-effect models such as general generalized linear mixed models or HGLMs with various response-variable distributions and random-effect structures (Lee *et al.*, 2006; Shin and Kim, 2014).

The proposed h-likelihood framework is based on the LASSO, SCAD or HL penalty. However, the LASSO or SCAD method may not be directly applicable to the high dimensional case with $p^* > n$ (Zou and Hastie, 2005; Lee *et al.*, 2010; Fan and Lv, 2010; Lee, 2015). With the HL penalty method, extension to such high dimensional case in HGLMs would be an interesting topic.

References

Androulakis, E., Koukouvinos, C. and Vonta, F. (2012). Estimation and variable selection via frailty models with penalized likelihood. *Statistics in Medicine*, **31**, 2223-2239.

Efron, B. and Morris, C. (1975). Data analysis using Steins estimator and its generalizations. *Journal of the American Statistical Association*, **70**, 311-319.

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.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**, 101-148.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, **32**, 928-961.

Ha, I. D. and Cho, G.-H. (2012). H-likelihood approach for variable selection in gamma frailty models. *Journal of the Korean Data & Information Science Society*, **23**, 199-207.

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.

Ha, I. D., Lee, Y. and MacKenzie, G. (2007). Model selection for multi-component frailty models. *Statistics in Medicine*, **26**, 4790-4807.

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. Pan, J., Oh, S. and Lee, Y. (2014). Variable selection in general frailty Models using penalized h-Likelihood. *Journal of Computational and Graphical Statistics*, **23**, 1044-1060.

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.

Hunter, D. and Li, R. (2005). Variable selection using MM algorithms. *The Annals of Statistics*, **33**, 1617-1642.

Johnson, B. A., Lin, D. Y. and Zeng, D. (2008). Penalized estimating functions and variable selection in semiparametric regression models. *Journal of the American Statistical Association*, **103**, 672-680.

Kwon, S., Oh, S. and Lee Y. (2014). The use of random-effect models for high-dimensional variable selection problems. revision sent to *Scandinavian Journal of Statistics.*

Lee, D., Lee, W., Lee, Y. and Pawitan, Y. (2010). Super sparse principal component analysis for high-throughput genomic data. *BMC Bioinformatics*, **11**, 296.

Lee, D., Lee, W., Lee, Y. and Pawitan, Y. (2011a). Sparse partial least-squares regression and its applications to high-throughput data analysis. *Chemo-metrics and Intelligent Laboratory Systems*, **109**, 1-8.

Lee, S. (2015). A note on standardization in penalized regressions. *Journal of the Korean Data & Information Science Society*, **26**, 505-516.

Lee, W., Lee, D., Lee, Y. and Pawitan, Y. (2011b). Sparse canonical covariance analysis for high-throughput data. *Statistical Applications in Genetics and Molecular Biology*, **10**, 1-24.

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. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series C*, **55**, 139-185.

Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalised Linear Models with Random Effects: Unified Analysis via h-Likelihood*, London, Chapman and Hall.

Lee, Y. and Oh, H. S. (2014). A new sparse variable selection via random-effect model. *Journal of Multivariate Analysis*, **125**, 89-9.

Paik, M. C., Lee, Y. and Ha, I. D. (2015). Frequentist inference on random effects based on summarizability. *Statistica Sinica*, **25**, 11071132.

Rondeau, V., Michiels, S., Liquet, B. and Pignon, J. P. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*, **27**, 1894-1910.

Shin, S. B. and Kim, Y. J. (2014). Statistical analysis of recurrent gap time events with incomplete observation gaps. *Journal of the Korean Data & Information Science Society*, **25**, 327-336.

Thall and Vail (1990) Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657-671.

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

Tibshirani, R. (1997). The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, **16**, 385-395.

Wang, H., Li, R. and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, **94**, 553-568.

Zhang, Y., Li, R. and Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, **105**, 312-323.

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

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society B*, **67**, 301-320.