Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Juniper Publishers
key to the Researchers

Review Article
Volume 10 Issue 2 - August 2020
DOI: 10.19080/BBOAJ.2020.10.555782

Biostat Biom Open Access J

Bayesian Mixed Effects Model with Variable Selection

Mingan Yang*

Division of Biostats and Epidemiology, San Diego State University, USA

Submission: June 22, 2020; **Published:** August 21, 2020

***Corresponding author:** Mingan Yang, Division of Biostats and Epidemiology, School of Public Health, San Diego State University, USA.

Abstract

Recently, many approaches have been proposed to address the problem of selecting both fixed and random effects in mixed effects models. In this article, we review several approaches by comparing their procedures and performances, discussing their similarities and differences, and explaining their advantages and disadvantages.

Keywords: Variable selection; Random effects; Mixed effects model; Bayesian model selection; Parameter expansion; Stochastic search

Abbreviations: LME: Linear Mixed Effects; MCMC: Markov Chain Monte Carlo; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; GIC: Generalized Information Criterion; SSVS: Stochastic Search Variable Selection

Introduction

Linear mixed effects (LME) models [1] are widely used in longitudinal studies to analyze correlated or clustered data. Generally, the random effects are incorporated to account for heterogeneity among the subjects. In analysis of an LME model, a primary objective is to select potential significant fixed effects and random effects of the outcome variables. Practically, one might be able to use a selection approach (e.g., back-ward elimination or forward selection) and apply a standard criterion, such as the Akaike information criterion (AIC), generalized information criterion (GIC), Bayesian information criterion (BIC), and Bayesian factor, to choose a preferred model by fitting all the possible models repeatedly. However, the number of competing models increases exponentially with the number of predictors. Thus, there are many challenging issues associated with the problem of joint selection of both fixed and random effects such as: intense computation, increased prediction error with increased number of covariates [2], bias associated in estimated variance of the fixed effects and near singular random effect covariance matrix for underfitting and overfitting of the random effects, respectively [3]. In this paper, we review several exemplary approaches by comparing their procedures and performances, investigating their similarities and differences, and explaining their advantages and disadvantages.

The Method

Suppose we have n subjects in a study and each subject has n_i repeated observations. For i=1; … , n. Let y_{ij} denote the response variable for subject i at observation j; X_{ij} the corresponding l x 1 predictor; and Z_{ij} a predictor vector of dimension q x1. Then, we de ne the LME model as follow:

$$y_{ij} = X_{ij}'\beta + Z_{ij}'\varsigma_i + \in_{ij}, \qquad 1$$

Where $\beta = (\beta_1,...,\beta_l)'$ is the fixed effect coefficient vector, $\varsigma_i = (\varsigma_{i1},...,\varsigma_{iq})'$ $N(0,\Omega)$ is the ith random effect, and the elements of the residual errors $\in_{ij}$ $N(0,\sigma^2)$. Generally, we choose Z_{ij} as a subvector of X_{ij} that include all the candidate predictors. We are interested in selecting a subset of important predictors in the model.

Chen & Dunson [4] addressed the problem by using the reparameterization approach of a modified Cholesky decomposition of the random effects covariance matrix. The covariance Ω can be decomposed as

$$\Omega = \Delta\Gamma\Gamma'\Delta',$$

where $\Delta = Diag(\lambda_1,...,\lambda_q)$, and λ_k is a non-negative element, proportional to the k^{th} random effect standard deviation, for $k = 1,...,q$. When $\lambda_k = 0$ it indicates that the k^{th} random effect is excluded from the model. Γ is a lower triangular matrix,

$$\Gamma = \begin{pmatrix} 1 & 0 & ... & 0 \\ \gamma_{21} & 1 & ... & 0 \\ ... & ... & ... & 0 \\ \gamma_{q1} & \gamma_{q2} & ... & 1 \end{pmatrix}$$

with all diagonal elements being 1 and the other free elements characterizing correlations between the random effects. With this

random effects covariance matrix decomposition, model (1) takes the form:

$$y_{ij} = X_{ij}'\beta + Z_{ij}'\Delta\Gamma\xi_i + \in_{ij}, \quad \in_{ij} \sim N(0,\sigma^2). \tag{2}$$

Chen & Dunson [4] showed that by rearranging terms, the covariance matrix of the random effects can be expressed as a function of $\lambda = (\lambda_1,...,\lambda_q)'$ and the free elements of Γ, $r = (\gamma_{fm} : f = 2,...,q; m = 1,...,f-1)$. Hence, the variance parameters λ and r keep desirable conditional conjugacy properties and we can construct a Gibbs sampling algorithm for sampling the posterior distribution using the SSVS approach.

The priors are specified as follows. Let $\beta^J = \{\beta_k : J_{k=1}\}$ be the vector of coefficients for the selected fixed effects in the current model, X^J be the corresponding covariates matrix. An i.i.d. Bernoulli prior is assumed for $J_k \overset{iid}{\sim} Bern(po)$. β^J is assumed to be a Zellner g-prior [5],

$$\beta^J \sim N(0,\sigma^2(X^{J'}X^J)^{-1}/g), \quad g \sim G(1/2, N/2),$$

Where G(a,b) denotes a Gamma distribution with mean a/b and variance a/b^2, $\sigma^2 \propto 1/\sigma^2$ (or inverse Gamma distribution) and $N = \sum n_i$.

For $\Delta = Diag(\lambda_1,...,\lambda_q)$, the elements of Δ are specified i.i.d. as

$$\lambda_k \sim_{pk} \delta_0 + (1-_{pk})N_+(0,1) \tag{3}$$

where IG(.) is an inverse Gamma distribution, $\delta_0(.)$ denotes a point mass at zero, $N_+(.)$ is a truncated positive normal distribution. The lower triangular free elements of Γ is put a normal distribution prior. For easy notation, we denote above zero-inflated truncated positive normal prior as $\lambda_k \sim ZIN_+(_{pk,}0,1)$. g is put a Gamma prior G(1/2; 1/2) and σ^2 a Jeffrey's prior $1 = \sigma^2$ or an inverse Gamma prior.

Chen & Dunson [4] specified the prior $\xi_i \sim N(0,I)$. Kinney & Dunson [6] used the approach of Gelman [7] and specified the covariance matrix $V(\xi_i) - Diag(d1,...,d_q)$. With this specification, the parameters Δ,Γ and $V(\xi_i)$ are not identifiable. In fact, Kinney & Dunson [6] took the parameter-expansion approach [8,9] that improved computational efficiency and reduces dependence among the parameters. We should note the Chen & Dunson [4] only considered variable selection of random effects. Kinney & Dunson [6] extended it to joint selection of both fixed and random effects for linear and logist models; in addition, the approach of Kinney & Dunson [6] also overcame the computational inefficiency due to slow mixing of the Gibbs sampler.

Although it is simple and convenient to assume that the random effects are normally distributed. However, there are several limitations with such specifications: the assumption is often not reasonable; thus misspecification of random effects might result in misleading interpretation and even incorrect results. In addition, it is challenging to specify nonparametric distribution for the random effects since there is bias associated with fixed effects estimates when the expected values of random effects are not zero. To resolve the bias of random effects, Yang [10] and Yang (2013) [11] used the approaches of the Probit stick-breaking (PSB) and location-scale symmetrized PSB (sPSB) [12] for linear and logist models with joint variable selection for both fixed and random effects. They define

$$G_U = \sum_{h=1}^{\infty} l_h(U)\delta_{\theta h}, \; l_h(U) = \phi(_{kh}(U))\prod_{l<h}(1-\phi(_{kh}(U))) \tag{4}$$

Where $_{kh}(U) = ch + f_h(U)$ with $_{ch} \sim N(\mu,1)$ and $f_h(U)$ generally takes a Gaussian kernel. The heteroscedastic scale PSB mixture and heteroscedastic sPSB location-scale mixtures are defined as follows:

$$PSB: \xi_i \sim \sum_{h=1}^{H} {}_{lh}(U)N(0,\sum{}_h) \tag{5}$$

$$_sPSB: \xi_i \sim \sum_{h=1}^{H} {}_{lh}(U)/2\{N(-\mu_h,\sum{}_h) + N(\mu_h,\Sigma_h)\} \tag{6}$$

for the random effects in truncated form. Obviously, the nonidentifiability issue is resolved since both PSB and sPSB are centered at vector zero.

Yang (2012) [10] & Yang (2013) [11] provided nonparametric approaches for linear and logist models for joint variable selection of both fixed and random effects. Their approaches are much more flexible than those of Chen & Dunson [4] and Kinney & Dunson [6]. However, the computation is more intense. Later Yang et. al. [13] used the shrinkage priors for mixed effects models with variable selection. The approach is efficient in shrinking small coefficients to zero while minimally shrinking large coefficients due to the heavy tails. They use several popular shrinkage priors: generalized double pareto prior [14], the horse shoe prior [15], and normal-exponential-gamma prior [16], respectively, as follows for the fixed effects:

$$\beta_{k:J_k=1} \sim N(0,\sigma^2/gk_k), \quad k_k \sim \exp(\varphi_k^2/2), \quad \varphi_k \sim Gamma(\nu_1,\vartheta_1), \tag{7}$$

$$\beta_{k:J_k=1} \sim N(0,\sigma^2/gk_k^2), \quad k_k \sim Ca^+(0,\nu_2), \quad \nu_2 \sim Ca^+(0,\vartheta_2), \tag{8}$$

$$\beta_{k:J_k=1} \sim N(0,\sigma^2/gk_k), \quad k_k \sim \exp(\varsigma^2/2), \quad \varsigma^2 \sim Gamma(\nu_3,\vartheta_3), \tag{9}$$

where $Ca^+(0; c)$ denotes a standard half-Cauchy distribution on the positive reals with scale parameter c. The performances of the shrinkage approaches are very good while the computations are not that intense.

Conclusion

In this article, we reviewed several approaches of linear and logistic models for joint selections of both xed and random effects. The approaches of Chen & Dunson [4] and Kinney & Dunson [6] are simple in implementation and provide reasonably good results. The approaches of Yang (2012) [10] & Yang (2013) [11] are much more flexible and provide much better results though the computations are intense. The approach of Yang et.al. [13] maintains a good balance of benefits of the above-mentioned parametric and nonparametric approaches. In summary, the performance and computation intensity by descending orders are Yang (2012) [10] & Yang (2013) [11], Yang et al. [13], Chen & Dunson [4] and Kinney & Dunson [6].

How to cite this article: Mingan Y. Bayesian Mixed Effects Model with Variable Selection. Biostat Biom Open Access J. 2020; 10(2): 555782.
DOI: 10.19080/BBOAJ.2020.10.5557782

References

1. Laird N and Ware J (1982) Random-effects models for longitudinal data. Biometrics 38(4): 963-974.
2. Miller A (2002) Subset selection in regression. In: 2nd (edn.), Boca Raton, Chapman and Hall/CRC, USA, pp. 1-256.
3. Lange N, Laird, NM (1989) The effect of covariance structures on variance estimation in balance growth-curve models with random parameters. Journal of the American Statistical Association 84(405): 241-247.
4. Chen Z, Dunson DB (2003) Random effects selection in linear mixed models. Biometrics 59(4): 762-769.
5. Zellner A, Siow A (1980) Posterior odds ratios for selected regression hypotheses. Hypothesis Testing 31: 585-603.
6. Kinney SK, Dunson DB (2007) Fixed and Random E effects Selection in Linear and Logistic Models. Biometrics 63(3): 690-698.
7. Gelman A (2005) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1(3): 515-534.
8. Liu C, Rubin DB, Wu YN (1998) Parameter expansion to accelerate EM: the PX-EM algorithm. Biometrika 85(4): 755-770.
9. Liu JS, Wu YN (1999) Parameter expansion for data augmentation. Journal of the American Statistical Association 94(448): 1264-1274.
10. Yang M (2012) Bayesian variable selection for logistic mixed model with nonparametric random effects. Computational Statistics and Data Analysis 56(9): 2663-2674.
11. Yang M (2013) Bayesian nonparametric centered random effects models with variable selec-tion. Biometrical Journal 55(2): 217-230.
12. Pati D, Dunson DB (2010) Bayesian Nonparametric regression with varying residual density. Annals of the Institute of Statistical Mathematics 66: 1-31.
13. Yang M, Wang M, Dong G (2020) Bayesian variable selection for mixed effects model with shrinkage prior. Computational Statistics and Data Analysis, 35: 227-243.
14. Armagan A, Dunson DB, Lee J (2013) Generalized double pareto shrinkage. Statistica Sinica 23(2013): 119-143.
15. Carvalho C, Polson N, Scott J (2010) The horseshoe estimator for sparse signals. Biometrika 97(2): 465-480.
16. Griffin J, Brown P (2007) Bayesian adaptive lassos with non-convex penalization. Australian & New Zealand journal of Statistics 53(4): 423-442.

DOI: 10.19080/BBOAJ.2020.10.555782

How to cite this article: Mingan Y. Bayesian Mixed Effects Model with Variable Selection. Biostat Biom Open Access J. 2020; 10(2): 555782.
DOI: 10.19080/BBOAJ.2020.10.5557782