Chapter 49
On the Markov Chain Monte Carlo Convergence Diagnostic of Bayesian Bernoulli Mixture Regression Model for Bidikmisi Scholarship Classification

Nur Iriawan, Kartika Fithriasari, Brodjol Sutijo Suprih Ulama, Irwan Susanto, Wahyuni Suryaningtyas and Anindya Apriliyanti Pravitasari

Abstract The Bidikmisi scholarship program is an education assistance program by the government of Indonesia which aims to achieve equitable access and learning opportunities at University. Bidikmisi acceptance status having a binary type (i.e. 0 and 1) produces a structure of Bernoulli mixture model with two components. The characteristics of each component can be identified through the Bernoulli mixture regression modeling by involving the covariates of Bidikmisi scholarship grantees. The estimating parameter of Bernoulli mixture regression model was performed using Bayesian-Markov Chain Monte Carlo (MCMC) approach. One of the challenges in using Bayesian-MCMC algorithm is determining the convergence of the sampler to the posterior distribution which is typically assessed using diagnostics tools. In this paper, we present that the diagnostics tools such as Geweke method, Gelman-Rubin method, Raftery-Lewis method and Heidelberger-Welch method can give different results to conclude MCMC convergence. The improvement of convergence indicators occurs on Gelman-Rubin method and Heidelberger-Welch method when the number of iterations is increased.

Keywords Markov chain monte carlo · Convergence diagnostic · Bernoulli mixture regression · Bayesian computation · Bidikmisi

---

N. Iriawan (✉) · K. Fithriasari · I. Susanto · W. Suryaningtyas · A. A. Pravitasari
Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: nur_i@statistika.its.ac.id

I. Susanto
e-mail: irwansusanto@staff.uns.ac.id

B. S. S. Ulama
Department of Business Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
e-mail: brodjol_su@statistika.its.ac.id

L.-K. Kor et al. (eds.), *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*,
https://doi.org/10.1007/978-981-13-7279-7_49

1 Introduction

Bernoulli mixture model is developed based on mixture distribution which has an adaptive capability to represent data pattern in data-driven analysis perspective [1]. Bidikmisi acceptance status that has a binary type can be formed as Bernoulli mixture model with two components. The characteristics of each component can be identified through the Bernoulli mixture regression model by involving the covariates of Bidikmisi scholarship recipients.

The inference for Bernoulli mixture regression model with Bayesian-Markov Chain Monte Carlo (MCMC) can overcome a particular challenge on computational aspects. Nevertheless it encounters a weakness that relates with the convergence of estimation process. Therefore the parameter estimation process of Bayesian Bernoulli mixture regression model has to be assessed on it convergence achievement.

2 Methodology

2.1 Bayesian Bernoulli Mixture Regression Model

Nadif and Govaert [2] introduced Bernoulli mixture model which was further developed by Grun and Leisch [3] in the generalized linear model framework. Suppose $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ is a random sample of a binary vector which has a linear relationship with covariates $X_1, X_2, \ldots, X_p$ on each Y_i such that

$$\eta_i = g(\mu_i) = g(\mathrm{E}(Y_i|X)) = \mathrm{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \sum_{j=1}^{p} \beta_j X_{ij} \qquad (1)$$

where η is linear predictor, $g(.)$ is the link function which is defined as logit function for Bernoulli distribution, μ_i is expected value of random variable Y_i and β is regression parameter. Therefore Bernoulli mixture regression model can be defined as

$$f(\mathbf{Y}|L, \boldsymbol{\pi}, \mathbf{X}, \boldsymbol{\beta}) = \sum_{\ell=1}^{L} \pi_\ell p_\ell(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}_\ell) \qquad (2)$$

where L is the number of mixture components, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_L)$ is the mixture proportion which has property $\sum_{\ell=1}^{L} \pi_\ell = 1$ and $p_\ell(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}_\ell) \sim Be(\mathrm{logit}_\ell^{-1}(\boldsymbol{\mu}))$, i.e., $p_\ell(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}_\ell)$ has a Bernoulli distributed with parameter $\mathrm{logit}_\ell^{-1}(\boldsymbol{\mu})$ with $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$, $\mathbf{X}' = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ and $\boldsymbol{\beta}_\ell = (\beta_1, \beta_2, \ldots, \beta_p)'$. Let $\boldsymbol{\Theta} =$

$(\Theta_1, \Theta_2, \ldots, \Theta_d)' = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L, \boldsymbol{\pi})'$ denote all unknown parameters appearing in the Bernoulli mixture regression model. The posterior probability distribution $f(\boldsymbol{\Theta}|\mathbf{Y}, L, \mathbf{X})$ can be represented as $f(\boldsymbol{\Theta}|\mathbf{Y}, L, \mathbf{X}) \propto f(\mathbf{Y}|\boldsymbol{\Theta}, L, \mathbf{X})p(\boldsymbol{\Theta})$ where $p(\boldsymbol{\Theta})$ is the prior distribution of $\boldsymbol{\Theta}$ and $f(\mathbf{Y}|\boldsymbol{\Theta}, L, \mathbf{X})$ is the mixture likelihood of Eq. (2).

2.2 *Markov Chain Monte Carlo Convergence Diagnostics*

As referred in [4], in Bayesian inference perspective, if Markov chain is convergent imply that the chain reaches the true posterior distribution. Thus the convergence of estimated parameters should be checked in order to get the true posterior inference for parameters.

Based on Gelman and Rubin [5], if there are *m Markov Chains* which are mutually independent and it has been taken a number of T iterations, $t = 1, 2, \ldots, T$, MCMC convergence can be monitored through estimation of potential scale reduction factor (PSRF). If the PSRF value is close to 1, then every m Markov Chains converge to the true posterior distribution.

As stated in [6], the diagnostic test of Geweke compute indicator Z that has means of subsamples A, $\bar{\Theta}^A$, and means of subsamples B, $\bar{\Theta}^B$ as the beginning and the end of samples respectively. Considering $\sigma_{(\bar{\Theta}^B - \bar{\Theta}^A)}$ is an estimated standard deviation of difference $\bar{\Theta}^B - \bar{\Theta}^A$ and Z asymptotically follows the standardized normal distribution, $Z \sim N(0, 1)$, so if $|Z| > 2$ then the chain is not convergent.

Raftery and Lewis [7] defined $N_{\min}$ as the minimum number of iterations that would be needed to achieve the required estimation precision for some function of parameter. If the value of dependence factors, $I = N/N_{\min}$, is greater than 5, then it implies convergence failure of Markov chain.

Heidelberger and Welch [8] proposed the method consists of two tests for assessing convergence. Firstly, a stationary test which verifies whether the Markov chain occurs from a stationary stochastic process. Secondly, the half-width test which determines if there is sufficient sample size for a chain to estimate the mean values of the process with appropriate accuracy. Markov chain has not convergent when it fails to meet those two tests.

2.3 *Data and Model*

The Data in this research are Bidikmisi 2015 data of all districts in East Java Province Indonesia which are composed of 33,603 Bidikmisi registrants from 35 regencies. Research variables used in this study consisted of the response variable (Y) and the predictor variable (X) which is constructed by dummy variables as follows

Y: the acceptance status of Bidikmisi scholarship (1 = accepted, 0 = not accepted)
X_1: father's job is formed by dummy variables d_{11}, d_{12}, d_{13}, and d_{14}.

d_{11}: farmer, fisherman or others job which relate with agriculture.
d_{12}: civil servants, police, and army.
d_{13}: entrepreneur.
d_{14}: private employees

X_2: mother's Job is formed by dummy variables d_{21}, d_{22}, d_{23}, and d_{24}.

d_{21}: farmer, fisher and others job which relate with agriculture.
d_{22}: civil servants, police, and army.
d_{23}: entrepreneur.
d_{24}: private employees

X_3: father's education is formed by dummy variables d_{31}, d_{32}, and d_{33}.

d_{31}: not continue to school.
d_{32}: elementary, junior high, or senior high school graduate level.
d_{33}: higher education level

X_4: mother's education is formed by dummy variables d_{41}, d_{42}, and d_{43}.

d_{41}: not continue to school.
d_{42}: elementary, junior high, or senior high school graduate level.
d_{43}: higher education level

All of dummy variables defined above are valued by 1, and otherwise are 0.
The Bernoulli mixture regression model which has to be estimated is defined by

$$f(\mathbf{y}|\boldsymbol{\pi}, \mathbf{x}, \boldsymbol{\beta}) = \pi_1\, Be\left(\frac{e^{g_1(\mathbf{x})}}{1 + e^{g_1(\mathbf{x})}}\right) + \pi_2\, Be\left(\frac{e^{g_2(\mathbf{x})}}{1 + e^{g_2(\mathbf{x})}}\right) \tag{3}$$

with π_1 and π_2 are mixture proportions which have properties $0 \le \pi_1 \le 1$, $0 \le \pi_2 \le 1$ and $\pi_1 + \pi_2 = 1$. $f(\mathbf{y}|\boldsymbol{\pi}, \mathbf{x}, \boldsymbol{\beta})$ represents two mixture components namely a component of wrong acceptance condition and a component of right acceptance condition. While $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ are formed by

$$g_1(\mathbf{x}) = \beta_{(1)0} + \beta_{(1)11}d_{11} + \beta_{(1)12}d_{12} + \beta_{(1)13}d_{13} + \beta_{(1)14}d_{14} + \beta_{(1)21}d_{21} + \beta_{(1)23}d_{23} + \beta_{(1)24}d_{24} + \beta_{(1)31}d_{31} + \beta_{(1)32}d_{32} + \beta_{(1)33}d_{33} + \beta_{(1)41}d_{41} + \beta_{(1)42}d_{42} + \beta_{(1)43}d_{43}$$

and

$$g_2(\mathbf{x}) = \beta_{(2)0} + \beta_{(2)11}d_{11} + \beta_{(2)12}d_{12} + \beta_{(2)13}d_{13} + \beta_{(2)14}d_{14} + \beta_{(2)21}d_{21} + \beta_{(2)23}d_{23} + \beta_{(2)24}d_{24} + \beta_{(2)31}d_{31} + \beta_{(2)32}d_{32} + \beta_{(2)33}d_{33} + \beta_{(2)41}d_{41} + \beta_{(2)42}d_{42} + \beta_{(2)43}d_{43}.$$

In that model, there are two parameters $\pi_\ell, \beta_{(\ell)kj}$ which have to be estimated. The prior distributions which are implemented for the each of the parameters are $p(\pi_\ell) \sim dirichlet(\vartheta)$ and $p(\beta_{(\ell)kj}) \sim N(\mu, \sigma)$. Coefficient of $\beta_{(\ell)kj}$ indicates the number of units (as coded) of change in $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ between the category for which dummy variable $d_{kj} = 0$ and the category for which dummy variable $d_{kj} = 1$.

3 Results

Computation of estimated parameters was performed on OpenBUGS [9]. Whereas diagnostic processes of MCMC convergence were done through R software with convergence diagnosis and output analysis (CODA) package [10]. In the first process, we generated 10,000 iterations which produced $\pi_1 = 0.604$ and $\pi_2 = 0.396$ as a significant estimated mixture proportions. CODA diagnostic for some parameters which have convergent problems is described on Table 1.

Referring to Table 1, Gelman-Rubin method shows that MCMC is convergent for all estimated parameter. The Geweke method indicates MCMC for the estimated parameter $\hat{\beta}_{41}$ on $g_1(x)$ is not convergent. Whereas Raftery-Lewis method gives unconvergent MCMC on estimated parameters $\hat{\beta}_0$, $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$ in $g_1(x)$ and $g_2(x)$. Based on Heidelberger-Welch method [8], MCMC for estimated parameter $\hat{\beta}_{24}$ is failed to converge. It means that by discarding of 10% increment until 50% of the iterations, the stationary tests are still failed. Therefore, when the simulation is run 10,000 iterations, there is only Gelman-Rubin which has a convergent indicator for all parameters. If Markov chain does not converge to the posterior distribution of parameters, then valid inferences of parameters on the Bernoulli mixture regression model cannot be accomplished. Based on [9], we conducted further simulations, i.e.,100,000 iterations in order to know the effect of increased number iterations on MCMC convergence. The significant estimated mixture proportions are $\pi_1 = 0.6041$ and $\pi_2 = 0.3959$. The result of CODA diagnostic is presented on Table 2.

In regard to Table 2, it can be seen that the Gelman-Rubin method and the Heidelberger-Welch method present an improvement of indicator values. Those

Table 1 Indicator values of CODA diagnostics with 10,000 iterations

Param	Sig. est. value		Gelman-Rubin		Geweke		Raftery-Lewis		Heidel-Welch	
	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$
$\hat{\beta}_0$	1.201	−1.79	1.01	1.03	0.223	−0.11	18.9	23.7	passed	passed
$\hat{\beta}_{11}$	−1.288	0.87	1.00	1.01	0.306	0.15	9.03	12.8	passed	passed
$\hat{\beta}_{21}$	−2	1.14	1.02	1.01	−0.77	0.115	20.7	20.3	passed	passed
$\hat{\beta}_{24}$	−1.3	0.07	1.01	1.00	−0.69	−0.71	4.54	3.63	failed	failed
$\hat{\beta}_{41}$	−0.257	−0.07	1.00	1.00	2.084	0.185	3.71	4.01	passed	passed

Table 2 Indicator values of CODA diagnostics with 100,000 iterations

Param	Sig. est. value				Gelman-Rubin		Geweke			
	Raftery-Lewis								Heidel-Welch	
$g_1(x)$	$g_2(x)$	$g_1(x)$		$g_2(x)$	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$	$g_1(x)$	$g_2(x)$
$\hat{\beta}_0$	1.19	−1.72	1.00	1.00	1.594	−1.77	53	25.2	passed	d passed
$\hat{\beta}_{11}$	−1.287	0.87	1.00	1.00	−0.50	1.87	13.8	13	passed	passed
$\hat{\beta}_{21}$	−1.99	1.13	1.00	1.00	−1.39	1.28	38.7	19.2	passed	passed
$\hat{\beta}_{24}$	−1.29	0.067	1.00	1.00	−2.09	1.09	4.91	3.92	passed	passed
$\hat{\beta}_{41}$	−0.256	−0.069	1.00	1.00	−0.74	−1.00	3.74	3.68	passed	passed

results can confirm significant outcome of estimated parameters. Four predictor variables,i.e., father's job, mother's job, father's education and mother's education significantly influence $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ which can affect on the parameter of Bernoulli distribution according to the Eq. (3). Nevertheless, the Geweke method shows inconsistent diagnostic results, whereas, the Raftery-Lewis method has not significant change on convergent indicator results and still gets unconvergent MCMC indicator values on $\hat{\beta}_0$, $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$. These results affirm that application of MCMC diagnostic tests to Bayesian Bernoulli mixture regression model can assure classification of two acceptance conditions in Bidikmisi, i.e., wrong acceptance condition and right acceptance condition.

4 Conclusion

The MCMC diagnostic methods give different results to conclude MCMC convergence. On the Gelman-Rubin method and the Heidelberger-Welch method, the increasing number of iterations improve convergence indicators for estimated parameters of Bernoulli mixture regression model.

Acknowledgements The Authors are grateful to DRPM-Kemenristekdikti Indonesia which supported this research under PUPT research grant no. 608/PKS/ITS/2017.

References

1. Iriawan, N.: Pemodelan dan Analisis Data-Driven. ITS Press, Surabaya (2012)
2. Nadif, M., Govaert, G.: Clustering for binary data and mixture models-choice of the model. Appl. Stoch. Mod. Bus. Ind. **13**, 269–278 (1997). https://doi.org/10.1002/(SICI)1099-0747(199709/12)13:3/4%3c269:AID-ASM321%3e3.0.CO;2-7

3. Grun, B., Leisch, F.: Finite mixtures of generalized linear regression models. In: Shalabh, C. H. (ed.) Recent Advances in Linear Models and Related Areas, pp. 205–230. Springer, Heidelberg (2008)
4. Alkan, N.: Assessing convergence diagnostic tests for bayesian cox regression. Comm. Stat. Sim. Comp. **46**(4), 3201–3212 (2017). https://doi.org/10.1080/03610918.2015.1080835
5. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**(4), 457–511 (1992). https://doi.org/10.1214/ss/1177011136
6. Geweke, J.: Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Smith, A.F.M., Dawid, A.P., Berger, J.O. (eds.) Bayesian Statistics 4, pp. 169–193. Oxford University Press, New York (1992)
7. Raftery, A.E., Lewis, S.M.: How many iterations in the gibbs sampler? In: Bernardo, J.M., Smith, A.F.M., Dawid, A.P., Berger, J.O. (eds.) Bayesian Statistics 4, pp. 762–773. Oxford University Press, New York (1992)
8. Heidelberger, P., Welch, P.D.: Simulation run length control in the presence of an initial transient. Oper. Res. **31**(6), 1109–1144 (1983). https://doi.org/10.1287/opre.31.6.1109
9. Lunn, D., Spiegelhalter, D., Thomas, A., Best, N.: The BUGS project: evolution, critique and future directions (with discussion). Stat. Med. **28**(25), 3049–3082 (2009). https://doi.org/10.1002/sim.3680
10. Plummer, M., Best, N., Cowles, K., Vines, K.: CODA: convergence diagnosis and output analysis for MCMC. R News **6**(1), 7–11 (2006)