PAPER • OPEN ACCESS

On The Markov Chain Monte Carlo Convergence Diagnostic of Bayesian Finite Mixture Model for Income Distribution

To cite this article: I Susanto *et al* 2018 *J. Phys.: Conf. Ser.* **1090** 012014

View the article online for updates and enhancements.

On The Markov Chain Monte Carlo Convergence Diagnostic of Bayesian Finite Mixture Model for Income Distribution

I Susanto[1,2], **N Iriawan**[3*], **H Kuswanto**[3], **Suhartono**[3] , **K Fithriasari**[3], **B S S Ulama**[4], **W Suryaningtyas**[1] **and A A Pravitasari**[1]

[1] Doctoral Candidate at Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

[2] Statistics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Surakarta, Indonesia.

[3] Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya,Indonesia.

[4] Department of Business Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

E-mail: nur_i@statistika.its.ac.id

Abstract. The assessment and comparison of income inequality and poverty can be supported by estimating probability distribution of income. Income distributions which are typically heavy tailed and positively skewed have been estimated both parametric and nonparametric approach. In parametric approach, finite mixtures distributions have been usefully implemented in the modelling of income distributions which has multimodal characteristic. The Markov Chain Monte Carlo (MCMC) approach is one of estimation methods which has a good performance in estimating parameter of Bayesian finite mixture model. The convergence of the MCMC sampler to the posterior distribution is typically assessed using standard diagnostics methods,.i.e., Gelman-Rubin method, Geweke method, Raftery-Lewis method and Heidelberger-Welch method. Those methods can give different results to conclude MCMC convergence condition. In this paper, a real sample income data from the Indonesian Family Life Survey (IFLS) 2015 and BidikMisi 2015 are employed to demonstrate the performance of diagnostics tools that assess convergence of the MCMC algorithm in estimating parameter of Bayesian finite mixture models.

1. Introduction

Statistical modelling based on finite mixture distributions that also known as finite mixture modelling is an interesting research field with the considerable range of applications. Finite mixture model captures many specific properties of real data such as multimodality, skewness, kurtosis, and unobserved heterogeneity [1]. It means that mixture distribution can represent as data pattern in data-driven analysis perspective [2]. Presently, finite mixture models are implemented in such varied areas as biometrics, genetics, medicine, financ e and economics including income distribution modelling.

In parametric approach, there are many probability distributions alternatives which can be used to estimate the income distribution model in such population. This approach is implemented as part of the analyzing process of the income distribution in a region related to the economic and poverty imbalances which occur in that population. The mixture distribution is relatively flexible for

approaching the distribution of income with different sub-populations. Each sub-populations may reflect groups with economic homogeneity [3]. Finite mixture models provide a flexible extension of classical statistical models, though the inference for these models poses particular challenge on computational aspects [1].

Bayesian inference approach can overcome that computational problem, since it allows the complicated structure of a mixture model to be decomposed into a set of simpler structures [4]. However, their analytical parameter of posterior distribution which can contain large integral dimension equation is usually difficult to be found. Thus a simulation method called Markov Chain Monte Carlo (MCMC) is suggested to solve numerically this problem. MCMC approach involves the simulation process which is performed iteratively Markov chains using Monte Carlo method to obtain convergence condition on posterior distribution. Implementation of MCMC in Bayesian analysis requires a proper sampling algorithm to obtain a sample of a distribution. Gibbs sampler is one of algorithms that is frequently used as generator of random variables in MCMC [5].

Although MCMC algorithms allow an advantageous computation, it can also encounter a conceivably serious weakness that relates with convergence of parameter estimation process. Therefore, in Bayesian finite mixture modelling, it should use standard diagnostics tools that verify the convergence of estimated parameter in order to reach the target posterior inference for parameters. The assessment methods of convergence such as Gelman-Rubin method [6], Geweke method [7], Raftery-Lewis method [8] and Heidelberger-Welch method [9] can produce diverse results about the achievement of MCMC convergence condition.

2. Bayesian Finite Mixture Model

We explain finite mixture model in section 2.1 and Bayesian approach for finite mixture model is described in section 2.2.

2.1. Finite Mixture Model

A random variable vector $\mathbf{y}$ which has discrete or continuous type is said to be derived from a finite mixture distribution, if the probability density function $g(\mathbf{y})$ has a mixture density form that applies to all $\mathbf{y}$,

$$g(\mathbf{y}) = w_1 f_1(\mathbf{y}) + ... + w_K f_K(\mathbf{y}) \tag{1}$$

where $f_k(\mathbf{y})$ is probability density function for $k = 1, 2, \ldots, K$. $f_k(\mathbf{y})$ is distribution function of k-th mixture component and K is the number of mixture components. $w_1, \ldots, w_K$ are weighting parameters and vector $\mathbf{w} = (w_1, \ldots, w_K)'$ is weighting vector of mixture distribution. The value of $\mathbf{w}$ must fulfil $0 \leq w_k \leq 1$ and $w_1 + ... + w_K = 1$.

In some cases of application, if it is assumed that all components of the finite mixture distribution come from a probability distribution having parameters $\boldsymbol{\theta}$, then equation (1) can be redefined as

$$g(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\theta}) = w_1 f(\mathbf{y} \mid \boldsymbol{\theta_1}) + ... + w_K f(\mathbf{y} \mid \boldsymbol{\theta_K}) \tag{2}$$

with $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_K})'$ [1]. Statistical model that implements the concept of finite mixture distribution in the modeling is usually called the finite mixture model.

2.2. Bayesian Approach

Let $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ are uncategorized observations from the mixture distribution that are randomly selected. Let $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_d)' = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_K}, \mathbf{w})'$ denote all unknown parameters appearing in the mixture model. By using Bayes theorem, the posterior probability distribution $\pi(\boldsymbol{\Theta} \mid \mathbf{y})$ is proportional to the multiplication between the prior distribution of $\boldsymbol{\Theta}$, $p(\boldsymbol{\Theta})$, and the mixture likelihood, $f_L(\mathbf{y} \mid \boldsymbol{\Theta})$. It can be represented mathematically as follows,

$$\pi(\mathbf{\Theta}|\mathbf{y}) = \frac{f_L(\mathbf{y}|\mathbf{\Theta})p(\mathbf{\Theta})}{f(\mathbf{y})}$$

$$\propto f_L(\mathbf{y}|\mathbf{\Theta})p(\mathbf{\Theta}) \tag{3}$$

and the mixture likelihood function $f_L(\mathbf{y}|\mathbf{\Theta})$ takes the form,

$$f_L(\mathbf{y}|\mathbf{\Theta}) = \prod_{i=1}^{N} g(\mathbf{y}_i|\mathbf{\Theta}) = \prod_{i=1}^{N}\left(\sum_{k=1}^{K} w_k f(\mathbf{y}_i|\boldsymbol{\theta}_k)\right). \tag{4}$$

After determining prior distribution, the Bayesian approach will perform parameter estimation by integrating the posterior distribution. The integration process is approached by simulation procedure which is commonly known as MCMC method. When Bayesian approach is implemented on finite mixture modelling, parameter estimation result does not rely on asymptotic normality, and yields valid inference in cases where regularity conditions are violated, such as small data sets and mixtures with small component weights [10].

2.2.1. Gibbs Sampler. Advantage of the Gibbs sampler is that, in each step, random values only consider to be generated from univariate conditional distributions. Based on *[5]*, the algorithm can be summarized by the following steps:

1. Set initial values $\mathbf{\Theta}^{(0)}$.
2. For $t = 1, 2, ..., T$ repeat the following steps :

   (i) $\mathbf{\Theta} = \mathbf{\Theta}^{(t-1)}$

   (ii) for $j = 1, 2, ..., d$ update Θ_j from $\Theta_j \sim \pi(\Theta_j|\mathbf{\Theta}_{\backslash j}, \mathbf{y})$

      $\pi(\Theta_j|\mathbf{\Theta}_{\backslash j}, \mathbf{y})$ is full conditional posterior distribution with $\mathbf{\Theta}_{\backslash j} = (\Theta_1, ..., \Theta_{j-1}, \Theta_{j+1}, ..., \Theta_d)'$.

   (iii) $\mathbf{\Theta}^{(t)} = \mathbf{\Theta}$ and save it as the generated set of values at t + 1 iterations of the algorithm.

Therefore, by giving the chain $\mathbf{\Theta}^{(t)}$, algorithm generates the new parameter values as follows

$\Theta_1^{(t)}$ from $\pi(\Theta_1|\Theta_2^{(t-1)}, \Theta_3^{(t-1)}, ..., \Theta_d^{(t-1)}, \mathbf{y})$

$\Theta_2^{(t)}$ from $\pi(\Theta_2|\Theta_1^{(t)}, \Theta_3^{(t-1)}, ..., \Theta_d^{(t-1)}, \mathbf{y})$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad\quad \vdots \quad \vdots$

$\Theta_j^{(t)}$ from $\pi(\Theta_j|\Theta_1^{(t)}, \Theta_2^{(t)}, ..., \Theta_{j-1}^{(t)}, \Theta_{j+1}^{(t-1)}, ..., \Theta_d^{(t-1)}, \mathbf{y})$

$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad\quad \vdots \quad \vdots$

$\Theta_d^{(t)}$ from $\pi(\Theta_d|\Theta_1^{(t)}, \Theta_2^{(t)}, ..., \Theta_{d-1}^{(t)}, \mathbf{y})$

The generating process, $\pi(\Theta_j|\mathbf{\Theta}_{\backslash j}, \mathbf{y}) = \pi(\Theta_j|\Theta_1^{(t)}, \Theta_2^{(t)}, ..., \Theta_{j-1}^{(t)}, \Theta_{j+1}^{(t-1)}, ..., \Theta_d^{(t-1)}, \mathbf{y})$, is quite simple and it has proportional relation as $\pi(\Theta_j|\mathbf{\Theta}_{\backslash j}, \mathbf{y}) \propto \pi(\mathbf{\Theta}|\mathbf{y})$, where all the variables except Θ_j have constant values.

3. Markov Chain Monte Carlo Convergence Diagnostics

3.1. Gelman-Rubin

The Gelman-Rubin method approach is done by defining several Markov Chains with different initial values and comparing the variance between some Markov Chains with the variance in each Markov Chain. If there is *m Markov Chains* which are mutually independent and if each markov chain has been taken a number of T iterations, $t = 1, 2, ..., T$, then MCMC convergence can be monitored through estimation of potential scale reduction factor (PSRF) formulated as follows

$$\hat{R} = \frac{\hat{V}(\Theta)}{avg} \tag{5}$$

where $\hat{V}(\Theta) = \frac{T-1}{T} avg + \frac{1}{T} B$ is the variance estimate of Θ, $avg = \frac{1}{m}\sum_{i=1}^{m} s_i^2$ with $s_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(\Theta_{it} - \overline{\Theta_i}\right)^2$

and $B = \frac{T}{m-1}\sum_{i=1}^{m}\left(\overline{\Theta}_i - \overline{\overline{\Theta}}\right)^2$ with $\overline{\overline{\Theta}} = \frac{1}{m}\sum_{i=1}^{m}\overline{\Theta}_i$. If the PSRF value is close to 1, then every m Markov Chains converge to the target distribution. Conversely, if the value of PSRF is large, it is necessary to take an extended stage of simulation that allows to reduce the value of $\hat{V}(\Theta)$ or to increase the value of *avg* such that m Markov Chains converge to target distribution [11]. The Gelman-Rubin method for one parameter is further developed by Brooks and Gelman for vector of Θ and also known as the Brooks-Gelman-Rubin (BGR) method [12].

3.2. Geweke

Suppose that Θ_1 is a parameter of interest and $\Theta_1^{(1)}$, ..., $\Theta_1^{(T)}$ are related simulated values with subsamples A and B as the beginning and the end respectively. The diagnostic test computes

$$Z = \frac{\overline{\Theta}^B - \overline{\Theta}^A}{\sigma_{(\overline{\Theta}^B - \overline{\Theta}^A)}} \tag{6}$$

where $\overline{\Theta}^A$ and $\overline{\Theta}^B$ are means of subsamples and $\sigma_{(\overline{\Theta}^B - \overline{\Theta}^A)}$ is an estimated standard deviation of difference $\overline{\Theta}^B - \overline{\Theta}^A$. Considering Z asymptotically follows the standardized normal distribution, $Z \sim N(0,1)$, if $|Z| > 2$ then the chain is not convergent [5].

3.3. Raftery-Lewis

Suppose $N_{\min}$ as the minimum number of iterations that is needed to achieve the required estimation precision for some function of parameter Θ and the quantile of interest is q and s, then $N_{\min}$ is given by

$$N_{\min} = \left\{ F^{-1}\left(\frac{s+1}{2}\right) \right\}^2 \frac{q(1-q)}{r^2} \tag{7}$$

where $F(\cdot)$ is the standard normal cumulative distribution function. The value of dependence factors, $I = N / N_{\min}$, can be used to indicate convergence condition of the chain. If the value is greater than 5, then it implies a convergence failure.

3.4. Heidelberger-Welch

The Heidelberger and Welch method which is applied for the analysis of single chains from univariate observations is developed based on Brownian bridge theory [9]. Suppose $\Theta^{(j)}$ is the j-th iterate in the chain, $f(0)$ is the spectral density of the chain evaluated at zero and $[nt]$ is the greatest integer less than or equal to nt. Let

$$B_n(t) = \frac{S_{[nt]} - [nt]\overline{\Theta}}{\left(nf(0)\right)^{1/2}}, \quad 0 \le t \le 1 \tag{8}$$

where $S_0 = 0$, $S_k = \sum_{j=1}^{k} \Theta^{(j)}$ $k \geq 1$, $\bar{\Theta} = \left(\sum_{j=1}^{n} \Theta^{(j)} \right) n^{-1}$, then $B_n = \{ B_n(t),\ 0 \leq t \leq 1 \}$ converges in distribution to the Brownian bridge as $n \to \infty$.

4. Application

In this section, real sample income data which determines from the Indonesian Family Life Survey (IFLS) 2015 and BidikMisi 2015 are employed to demonstrate the performance of diagnostics tools that assess convergence of the MCMC algorithm in finite mixture models. Computation of MCMC convergence with those diagnostic methods has been integrated in BUGS software : WinBUGS [13] or OpenBUGS [14]. Whereas computing through R software is done with Convergence Diagnostic and Output Analysis (CODA) package [15].

4.1. Indonesian Family Life Survey (IFLS) 2015.
Source of household income response data is processed based on survey data from Indonesian Family Life Survey (IFLS) 2015 which examines the life of household in Indonesia [16]. In this paper, we use household income data in Province of Daerah Istimewa Yogyakarta (DIY) with a sample of 690 households in five districts / cities: Kulonprogo Regency, Bantul Regency, Gunungkidul Regency, Sleman Regency and Yogyakarta City. The histogram kernel plot of the distribution is shown in Figure 1 below
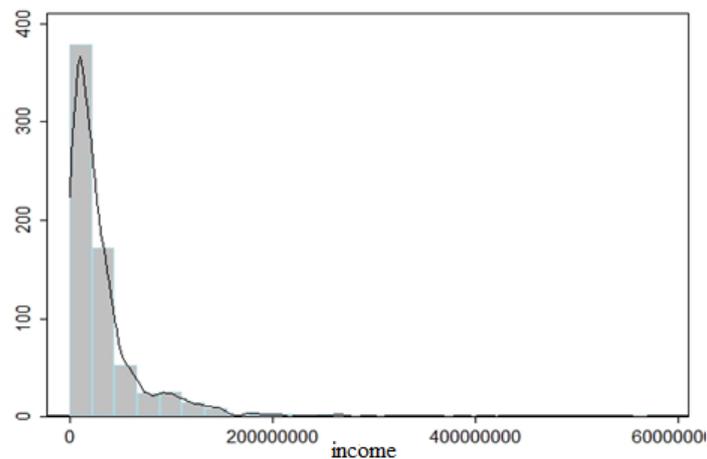

Figure 1 Kernel Histogram Plot Distribution of Household Income Data in Province of DIY.

Based on figure 1, it can be shown that income distribution indicates a multimodal characteristic that can be approached by finite mixture distribution. Using *Mathematica* 11 software, we preliminary presume that income distribution can be modelled by mixture model with three components of the two parameter Gamma distribution, $Gamma(\alpha, \beta)$,which can be called as three components Gamma mixture,

$$g(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\theta}) = \sum_{k=1}^{3} w_k f(\mathbf{y} \mid \boldsymbol{\theta}_k)$$

where $f(\mathbf{y} \mid \boldsymbol{\theta}_k) = f(\mathbf{y} \mid \alpha_k, \beta_k) = \beta_k^{\alpha_k} \left(\Gamma(\alpha_k) \right)^{-1} y^{\alpha_k - 1} e^{-\beta_k y}$ is two parameter Gamma distribution.

In that model, there are three parameters α, β, w which have to be estimated. Based on [17], the prior distributions which are implemented for each of parameters are

$$p(\mathbf{w}) \propto Dirichlet(\vartheta)$$
$$p(\alpha_k) \propto Exp(\theta)$$
$$p(\beta_k) \propto InverseGamma(\alpha_k, \beta_k)$$

In our research, we focus on estimation convergence of weighting parameters w_1, w_2, w_3. In the first process, we generated 8,000 iterations which produced indicator values of CODA diagnostic that indicated some convergent problems. It was necessary to take an extended stage of iterations as stated on [11]. Therefore we performed further simulations,i.e.,100,000 iterations so that the effect of raised number iterations could be verified on MCMC convergence. The results of CODA diagnostic are described on table 1 with case 1 for 8,000 iterations and case 2 for 100,000 iterations.
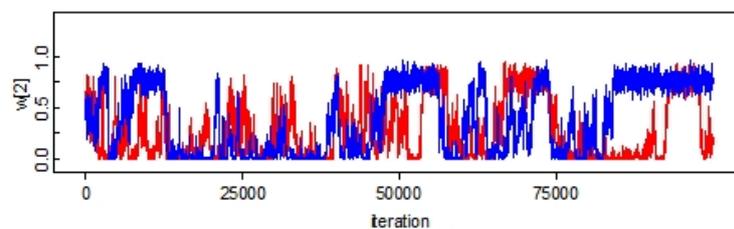
Table 1. Indicator values of CODA diagnostics for weighting parameters w_1, w_2, w_3 .

Diagnostics	w_1		w_2		w_3	
	case 1	case 2	case 1	case 2	case 1	case 2
BGR	5.74	1.01	3.05	1.01	1.43	1.04
Geweke	-4.555	-1.260	5.653	1.56	0.116	-0.249
Raftery-Lewis	41.20	60.3	9.37	17.8	22.20	50.8
Heidelberger-Welch	failed	failed	failed	failed	passed	passed

Referring to table 1, for BGR and Geweke methods, the increasing number of iterations can improve the convergence of MCMC. However, it does not occur for Raftery-Lewis and Heidelberger-Welch methods. In all cases and all parameters, Raftery-Lewis method gives unconvergent indicator values. If we regard from three diagnostics methods: BGR, Geweke and Heidelberger-Welch, it seems that only estimation parameter of w_3 which tends to converge. This problem is caused by a computation process that is trapped on one of the mixture components and the process cannot get out from that mixture component. Historical simulations with two different chains on figure 1 present that estimation process of w_3 tends a convergent condition which does not be achieved by estimation processes of w_1, w_2.
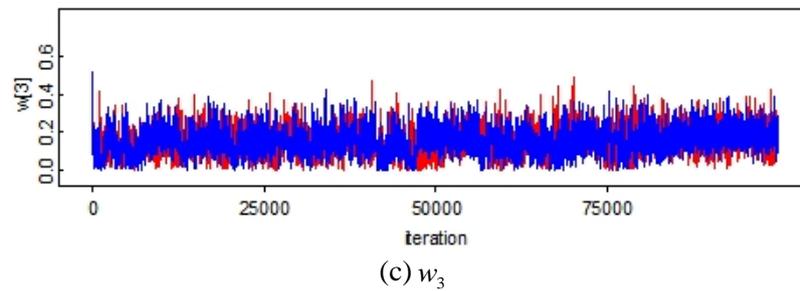

(a) w_1

(b) w_2

(c) w_3
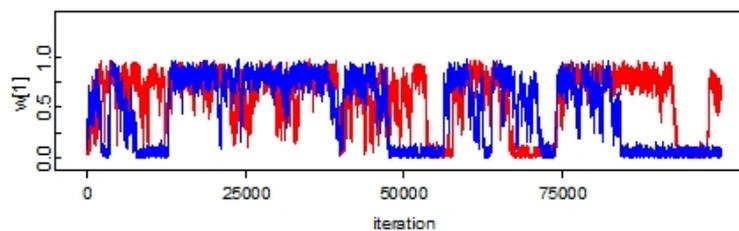
Figure 1. Historical simulation of w_1, w_2, w_3 on 100,000 iterations.

Those convergence problems can happen, because there is not restriction condition on mean of each component Gamma mixture, i.e., $\mu_1 < \mu_2 < \mu_3$, which can assure identifiability of component mixture. So re-parameterization of Gamma mixture model that implements restriction condition is done through BUGS program which is given on Appendix. The results of CODA diagnostics after re-parameterization are shown on table 2.

Table 2. Indicator values of CODA diagnostics for weighting parameters w_1, w_2, w_3

Diagnostics	w_1		w_2		w_3	
	case 1	case 2	case 1	case 2	case 1	case 2
BGR	1.02	1.00	1.00	1.01	1.02	1.00
Geweke	-0.9699	-0.7212	-0.5645	-0.5383	0.7292	0.2442
Raftery-Lewis	98.4	63	7.78	17	18.7	35.7
Heidelberger-Welch	passed	passed	passed	passed	passed	passed

Table 2 shows that indicator values for BGR, Geweke and Heidelberger-Welch methods indicate the Markov chain reaches convergence condition for both cases. However, the convergence conditions are not fulfilled by the Raftery-Lewis method.

Other possibly solution to avoid a trapped estimation process on one of the mixture components is to reduce the number of mixture components [10]. So, we simplified the mixture model from three component mixtures to two component mixtures. The CODA results for two component Gamma mixture are exposed on table 3.

Table 3. Indicator values of CODA for two components Gamma mixtures.

Diagnostics	w_1		w_2	
	case 1	case 2	case 1	case 2
BGR	1.03	1	1.03	1
Geweke	0.764	1.115	-0.764	-1.115
Raftery-Lewis	16.8	26.5	9.78	18.5
Heidelberger-Welch	passed	passed	passed	passed

In regarding to table 3, it can be seen that BGR, Geweke and Heidelberger-Welch methods give similar conclusion about convergence parameters achieved. Nevertheless, Raftery-Lewis method still gives an opposite convergence result. For this reason, it has to be required further exploration which concerns about implementation of Raftery-Lewis on finite mixture models.

4.2. BidikMisi 2015
Source of income response data is processed based on survey data from BidikMisi in 2015. In this paper, we used income data in Province of Daerah Istimewa Yogyakarta (DIY) with a sample of 1,149 households. In this paper the income distribution is modelled by mixture model with three components of Normal distribution $N(\mu, \sigma^2)$,

$$g(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\theta}) = \sum_{k=1}^{3} w_k f(\mathbf{y} \mid \boldsymbol{\theta}_k)$$

where $f(\mathbf{y} \mid \boldsymbol{\theta}_k) = f(\mathbf{y} \mid \mu_k, \sigma_k^2) = \left(2\pi\sigma^2\right)^{-1/2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is Normal distribution. The prior distribution for weighting parameter is $p(\mathbf{w}) \propto Dirichlet(\vartheta)$. Similar simulations as IFLS data are conducted for assessment process of MCMC convergence. The results of CODA diagnostic are presented on table 4.

Table 4. Indicator values of CODA diagnostics for weighting parameters w_1, w_2, w_3 of Normal mixture.

Diagnostics	w_1		w_2		w_3	
	case 1	case 2	case 1	case 2	case 1	case 2
BGR	1.04	1	1.01	1	1.07	1
Geweke	-0.8569	-0.6666	-0.8146	0.4141	1.0277	0.0883
Raftery-Lewis	12.00	23.3	7.93	12.0	55.5	99.0
Heidelberger-Welch	failed	passed	failed	passed	passed	passed

Indicator values which are shown on table 4 give an affirmation that the increasing number of iterations can provide better indicator values of MCMC convergence. Even though, that condition only happen on BGR, Geweke and Heidelberger-Welch methods, while indicator value of convergence base on Raftery-Lewis method has not changed.

Historical simulations with two different chains on figure 2 describe that estimation process of w_1, w_2, w_3 tends a convergent condition.
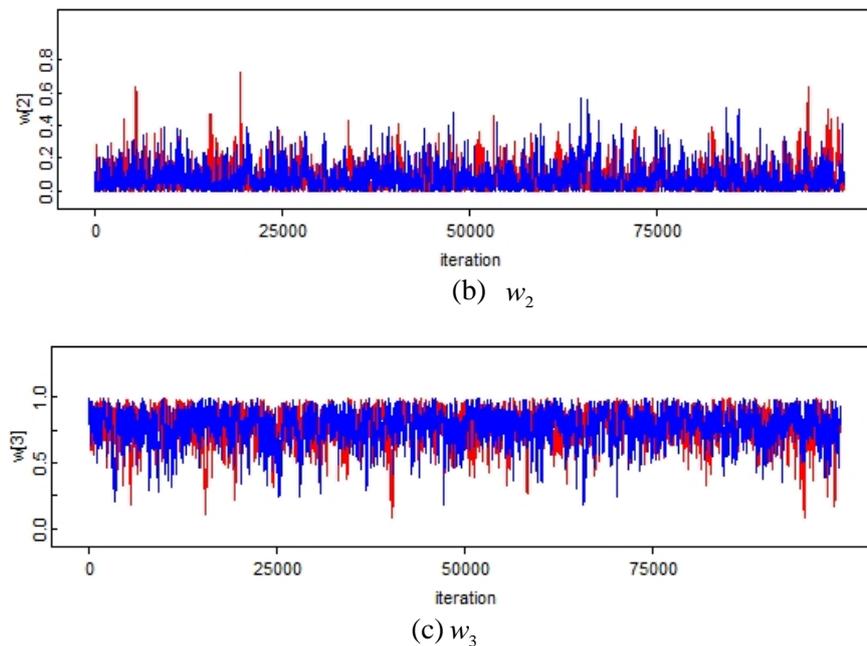

(a) w_1

(b) w_2

(c) w_3
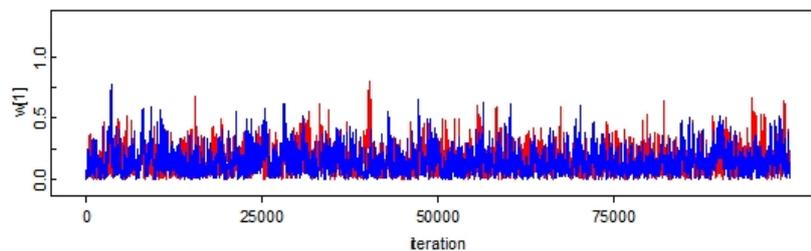
Figure 2. Historical simulation of w_1, w_2, w_3 on 100,000 iterations for Normal mixture model.

5. Conclusion

The increasing number of iterations can provide a better indicator value of MCMC convergence on diagnostics tools such as Gelman-Rubin method, Geweke method and Heidelberger-Welch method. Nonetheless indicator value of Raftery-Lewis method is not affected by the increment of iteration. Therefore, further researches that concern on Raftery-Lewis method for assessment convergence of Bayesian finite mixture modelling are required.

Appendix

```
model{
        for( i in 1 : N ) {
                y[i] ~ dgamma(alpha, beta[i])
                beta[i] <- alpha/mu[T[i]]
                T[i] ~ dcat(w[])
        }
        w[1:3] ~ ddirich(alp[])
        theta ~ dunif(0.0, 10000000)
        theta2 ~ dunif(0.0,10000000)
        mu[3] <- mu[2] + theta2
        mu[2] <- mu[1] + theta
        mu[1] <- 1/lamb
```

```
    lamb ~ dgamma(1.5, 1)
    alpha ~ dexp(1)
}
```

References

[1] Frühwirth-Schnatter S. Finite mixture and Markov switching models New York: Springer Science+Business Media; 2006.

[2] Iriawan N. Modelling and Data-Driven Analysis Surabaya: ITS Press; 2012.

[3] Pittau MG, Zelli R. Empirical evidence of income dynamics across EU regions. Journal of Applied Econometrics. 2006; 21(5): p. 605–628. DOI : 10.1002/jae.855.

[4] Marin JM, Mengersen K, Robert CP. Bayesian Modelling and Inference on Mixtures of Distributions. In Dey DK, Rao CR, editors. Handbook of Statistics Vol.25 : Bayesian Thinking Modeling and Computation. Amsterdam: Elsevier; 2005. p. 459-507.

[5] Ntzoufras I. Bayesian modeling using WinBUGS New york. DOI: 10.1002/9780470434567: John Wiley & Sons; 2009.

[6] Gelman A, Rubin D. Inference from Iterative Simulation using Multiple Sequences. Statistical Science. 1992; 7(4): p. 457 - 511. DOI : 10.1214/ss/1177011136.

[7] Geweke J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bernardo JM, Smith AFM, Dawid AP, Berger JO, editors. Bayesian Statistics 4. New York: Oxford University Press; 1992. p. 169 - 193. DOI : 10.1.1.27.2952.

[8] Raftery AE, Lewis SM. How Many Iterations in the Gibbs Sampler ? In Bernardo JM, Smith AFM, Dawid AP, Berger JO, editors. Bayesian Statistics 4. New York: Oxford University Press; 1992. p. 762 - 773.DOI : 10.1.1.41.6474.

[9] Heidelberger P, Welch PD. Simulation Run Length Control in the Presence of an Initial Transient. Operations Research. 1983; 31(6): p. 1109 - 1144. DOI : 10.1287/opre.31.6.1109.

[10] Tatarinova T, Schumitzky A. NonLinear Mixture Models : A Bayesian Approach London: Imperial College Press; 2015.

[11] Brooks SP, Roberts GO. Convergence assessment techniques for Markov Chain Monte Carlo. Statistics and Computing. 1998; 8(4): p. 319-335. DOI : 10.1023/A:1008820505350.

[12] Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics. 1998; 7(4): p. 434 - 455. DOI : 10.1080/10618600.1998.10474787.

[13] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing. 2000; 10: p. 325-337. DOI : 10.1023/A:1008929526011.

[14] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions (with discussion). Statistics in Medicine. 2009; 28: p. 3049 - 3082. DOI : 10.1002/sim.3680.

[15] Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. R News. 2006; 6(1): p. 7-11.

[16] Strauss J, Witoelar F, Sikoki B. The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report. Working Paper. Santa Monica: RAND Corporation; 2016. Report No.: WR-1143/1-NIA/NICHD.

[17] Wiper M, Insua DR, Ruggeri F. Mixtures of Gamma Distributions with Applications. Journal of Computational and Graphical Statistics. 2001; 10: p. 440 - 454.