Bayesian inference for the finite gamma mixture model of income distribution

I Susanto¹, N Iriawan², H Kuswanto², and Suhartono²

¹Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret, Surakarta, Indonesia.
²Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.
E-mail: irwansusanto@staff.uns.ac.id, nur_i@statistika.its.ac.id

Abstract. The income distribution model has provided an important aspect of economic inequality analysis. The determination of income inequality can be assisted by modeling a probability distribution of income which can be modeled by both parametric and nonparametric method. In the parametric perspective, the finite mixture distributions can perform a data-driven capability to model this income pattern of distributions which have particularly long-tailed, right-skewed and multimodal characteristics. The gamma distribution which has been widely used for estimating income distribution is used to develop the finite gamma mixture model which means the gamma distribution in each mixture component of the model. Bayesian approach pairs up with the Markov Chain Monte Carlo (MCMC) which has a valid inference without depending on normality asymptotic condition is used to estimate this finite mixture model. In this paper, the household income which was constructed based on the Indonesian Family Life Survey (IFLS) 2014-2015 data was utilized to show the work of the Bayesian inference performance through MCMC algorithm in estimating the parameter of the finite gamma mixture model. The goodness-of-fit comparisons of proposed finite gamma mixture models were made based on the widely applicable information criteria (WAIC).

1. Introduction

Finite mixture models provide a flexible extension of classical statistical models which have specific properties such as multimodality, asymmetric distribution, heavy-tailed, and undiscovered diversity [1]. In a data-driven analysis perspective, it gives the meaning that the mixture model can describe a pattern which is stimulated by a process that is triggered by data [2]. However, the inference of those models has a specific challenge on computational sides. Bayesian inference approach can handle that problem since it can build a simplified structure of mixture models [3]. Nevertheless, the analytical parameter of posterior distribution through Bayesian inference is generally problematic to be obtained. Therefore a simulation method named Markov Chain Monte Carlo (MCMC) is proposed to determine that difficulty numerically. The simulation process of MCMC is executed Markov chains iteratively using Monte Carlo method to reach a convergence state on posterior distribution. The fulfillment of MCMC in Bayesian analysis demands a convenient sampling algorithm. One of those algorithms is the Gibbs sampler which is commonly implemented as a generator of random variables in MCMC [4].

The simulation process of MCMC is an essential process in Bayesian analysis that needs an important examination in the achievement of convergence. A usual method in diagnosing the convergence of MCMC is using the Gelman-Rubin method which assesses the convergence by analyzing the difference between multiple Markov chains [5].
2. Finite gamma mixture model

Suppose a random variable vector \( y \) which has a positive and continuous type; then the finite gamma mixture model can be defined by

\[
g(y \mid w, \theta) = w_1 f(y \mid \theta_1) + \ldots + w_k f(y \mid \theta_k) = \sum_{k=1}^{K} w_k f(y \mid \theta_k),
\]

(1)

Where \( K \) is the number of mixture components, \( w_1, \ldots, w_k \) are weighting parameters and the vector \( w=(w_1, \ldots, w_k) \) is a weighting vector of mixture distribution. The value of \( w \) must fulfill \( 0 \leq w_k \leq 1 \) and \( w_1 + \ldots + w_k = 1 \). \( f(y \mid \theta_k) \) has a gamma distribution with two parameters \( \alpha_k \) and \( \beta_k \), so \( f(y \mid \theta_k) = f(y \mid \alpha_k, \beta_k) = \beta_k^{\alpha_k} (\Gamma(\alpha_k))^{-1} y^{\alpha_k-1} e^{-\beta_k y} \) with \( \theta_k=(\alpha_k, \beta_k) \) and \( \theta=(\theta_1, \ldots, \theta_k) \).

3. Bayesian inference

We explain Bayesian method in sub-section 3.1, Gibbs sampler algorithm is described in subsection 3.2, the convergence diagnostic of the Markov Chain Monte Carlo in sub-section 3.3 and the widely applicable information criteria (WAIC) in sub-section 3.4.

3.1 Bayesian method

Let \( y=\{y_1, y_2, \ldots, y_N\} \) are unclassified observations from the mixture distribution that are randomly chosen. Let \( \Theta=(\Theta_1, \Theta_2, \ldots, \Theta_d) = (\theta_1, \ldots, \theta_K, w) \) are known as all parameters which have to be estimated in the mixture model. Based on Bayes theorem, the posterior distribution \( \pi(\Theta \mid y) \), is proportional to the multiplication between the prior distribution of \( \Theta \), \( p(\Theta) \), and the mixture likelihood \( f_L(y \mid \Theta) \). It can be represented mathematically as follows,

\[
\pi(\Theta \mid y) = \frac{f_L(y \mid \Theta) p(\Theta)}{f(y)}
\]

\[\propto f_L(y \mid \Theta) p(\Theta)\]

(2)

and the mixture likelihood function \( f_L(y \mid \Theta) \) takes the form,

\[
f_L(y \mid \Theta) = \prod_{i=1}^{N} g(y_i \mid \Theta) = \prod_{i=1}^{N} \left( \sum_{k=1}^{K} w_k f(y_i \mid \theta_k) \right).
\]

(3)

By giving prior distribution, the Bayesian approach will perform parameter estimation by integrating the posterior distribution which is approached by the MCMC method. In Bayesian approach, the result of parameter estimation does not depend on asymptotic normality and provides a valid inference in cases such as small data sets, unbalanced data, and mixtures with small component weights [6].

3.2 Algorithm

In each step of Gibbs sampler, a random value simply takes into account to be produced from the univariate conditional distributions. The algorithm can be developed by the following steps:

1. Set initial values \( \Theta^{(0)} \).
2. For \( t=1, 2, \ldots, T \) repeat the following steps:
   (i) Set \( \Theta = \Theta^{(t-1)} \)
   (ii) generate \( \Theta_j \) for \( j=1, 2, \ldots, d \) from \( \Theta_j \sim \pi(\Theta_j \mid \Theta_{-j}, y) \).

\( \pi(\Theta_j \mid \Theta_{-j}, y) \) Represents a full conditional posterior distribution with \( \Theta_{-j}=(\Theta_1, \ldots, \Theta_{j-1}, \Theta_{j+1}, \ldots, \Theta_d) \).
The full conditional posterior distribution \( \pi(\Theta_j | \Theta_{\setminus j}, y) \) has proportional relation as

\[
\pi(\Theta_j | \Theta_{\setminus j}, y) \propto \pi(\Theta_j | y),
\]

with all the variables except \( \Theta_j \) have constant values.

(iii) set \( \Theta^{(i)} = \Theta \)

(iv) keep it as the generated set of values at \( t + 1 \) iterations of the algorithm.

(v) by giving the chain \( \Theta^{(i)} \), the algorithm sets up the new parameter values as follows

\[
\Theta_1^{(i)} \text{ from } \pi(\Theta_1 | \Theta_2^{(i-1)}, \Theta_3^{(i-1)}, ..., \Theta_d^{(i-1)}, y) \\
\Theta_2^{(i)} \text{ from } \pi(\Theta_2 | \Theta_1^{(i)}, \Theta_3^{(i-1)}, ..., \Theta_d^{(i-1)}, y) \\
\vdots \\
\Theta_j^{(i)} \text{ from } \pi(\Theta_j | \Theta_1^{(i)}, \Theta_2^{(i)}, ..., \Theta_{j-1}^{(i)}, \Theta_{j+1}^{(i-1)}, ..., \Theta_d^{(i-1)}, y) \\
\vdots \\
\Theta_d^{(i)} \text{ from } \pi(\Theta_d | \Theta_1^{(i)}, \Theta_2^{(i)}, ..., \Theta_{d-1}^{(i)}, y)
\]

### 3.3 MCMC convergence diagnostics

The Gelman-Rubin method is constructed by specifying several Markov chains with different initial values and comparing the variance between some Markov chains with the variance in each Markov chain. Suppose \( m \) Markov chains are mutually independent with each Markov chain iterated for \( t = 1, 2, ..., T \). The convergence of MCMC is monitored through estimation of potential scale reduction factor (PSRF) formulated as follows.

\[
\hat{R} = \frac{\hat{V}(\Theta)}{\text{avg}}
\]

Where \( \hat{V}(\Theta) = \frac{T-1}{T} \text{avg} + \frac{1}{T} B \) is the variance estimate of \( \Theta \), \( \text{avg} = \frac{1}{m} \sum_{i=1}^{m} s_i^2 \) with \( s_i^2 = \frac{1}{T-1} \sum_{t=1}^{T} (\Theta_{ji} - \overline{\Theta})^2 \)

\[
B = \frac{T}{m-1} \sum_{i=1}^{m} (\overline{\Theta}_i - \overline{\Theta})^2 \text{ with } \overline{\Theta} = \frac{1}{m} \sum_{i=1}^{m} \overline{\Theta}_i . \] If the PSRF value approach to 1, then every \( m \) Markov Chains converge to the target distribution. Contrarily, if the PSRF has a large value, it is necessary to have an extended simulation that permits to decrease \( \hat{V}(\Theta) \) or to raise \( \text{avg} \) such that \( m \) Markov Chains converge to target distribution [7]. The Gelman-Rubin method is expanded for the vector of \( \Theta \) and also recognized as the Brooks-Gelman-Rubin (BGR) method [8].

### 3.4 Widely applicable information criteria

Widely applicable information criteria which can be named as Watanabe-Akaike information criteria (WAIC) is suitable for comparing Bayesian model which has hierarchical and mixture structure [9,10]. WAIC which is constructed as the estimation of expected log pointwise predictive density (elppd) is defined by

\[
elppd_{\text{waiac}} = \text{lpdp} - \text{p}_{\text{waiac}} .
\]

where \( \text{lpdp} \) is the log pointwise predictive density,

\[
\text{lpdp} = \log \prod_{i=1}^{n} p_{\text{post}} (y_i) = \sum_{i=1}^{n} \log \int f(y_i | \theta) \pi(\theta) \, d\theta ,
\]

With \( p_{\text{post}} (y_i) \) is the predictive density function based on posterior distribution \( \pi(\theta) \) and the likelihood function \( f(y_i | \theta) \). The effective number of parameter \( p_{\text{waiac}} \), is formulated as

\[
p_{\text{waiac}} = \sum_{i=1}^{n} \text{var}_{\text{post}} (\log f(y_i | \theta)) ,
\]
With \( \text{var}_{\text{post}}(\log f(y_i | \theta)) \) is the posterior variance of log-likelihood function. The determination of \( \text{lppd} \) and \( p_{\text{waic}} \) values can be estimated through MCMC computation. Therefore the estimated \( \text{lppd} \) can be given by

\[
\text{lppd} = \sum_{i=1}^{n} \log \left( \frac{1}{S} \sum_{s=1}^{S} f(y_i | \theta^s) \right),
\]

Where \( \theta^s \) with \( s = 1, 2, ..., S \) is simulated from the posterior distribution \( \pi(\theta) \). Whereas the estimation of \( p_{\text{waic}} \) is defined by

\[
\hat{p}_{\text{waic}} = \sum_{i=1}^{n} V_i \left( \log(f(y_i | \theta^i)) \right)
= \sum_{i=1}^{n} \frac{1}{S-1} \sum_{s=1}^{S} \left( \log(f(y_i | \theta^s) - E[\log(f(y_i | \theta^s))] \right)^2.
\]

4. Results

4.1. Estimating the finite gamma mixture model
In model (1), based on [11], three parameters \( \alpha, \beta, w \) which have to be estimated involve the prior distributions as

\[
p(w) \propto \text{Dirichlet}(\beta)
\]
\[
p(\alpha_k) \propto \text{Exp}(\beta)
\]
\[
p(\beta_k) \propto \text{InverseGamma}(\alpha_k, \beta_k).
\]

4.2. Empirical study
In this subsection, household income data which was determined from the Indonesian Family Life Survey (IFLS) 2014-2015 was employed to demonstrate the Bayesian inference performance through MCMC algorithm in estimating a parameter of the finite gamma mixture model. MultiBUGS which is an OpenBUGS software with parallel computation capability [12], was used to estimate the parameter of the finite gamma mixture model. The computation of MCMC convergence and with those diagnostic methods has been integrated with MultiBUGS. Whereas computing through R software is done with Convergence Diagnostic and Output Analysis (CODA) package [13].

4.2.1. Data. Income distribution was developed by household income response data which were processed based on survey data from Indonesian Family Life Survey (IFLS) 2014-2015 [14]. In this paper, we determine the income of 690 households.

4.2.2. Estimating parameter and convergence diagnostics. In the first process, regarding effective sample size of MCMC chains for parameters of interest, we generated 10,000 iterations as an effective sample size [15] and 100,000 iterations as a further simulation in consideration of the determination of the effect of increased number iterations on MCMC convergence [16]. Table 1 shows that the model with four mixture components has a relatively small WAIC than the other number of mixture components. As stated on [9,10], WAIC gives an asymptotically unbiased estimate of the out-of-sample prediction error which can be used as a measure of model accuracy. It means that in the goodness of fit aspects the model with a smaller value of WAIC than other models can be stated as a much better model than the other models. Therefore a gamma mixture model with four mixture component is selected. Referring on (1), the model can be formulated by

\[
g(y | w, \theta) = \sum_{k=1}^{K} w_k f(y | \theta_k),
\]
where \( f(y|\theta_k) \sim \text{Gamma}(\alpha_k, \beta_k) \), \( k = 1,2,3,4 \).

**Table 1.** Indicator values of WAIC for the number of mixture components which is simulated 10,000 and 100,000.

| Number of mixture components | WAIC 10,000 | WAIC 100,000 |
|-----------------------------|-------------|--------------|
| 2                           | 25,300      | 25,300       |
| 3                           | 25,290      | 25,290       |
| 4                           | 25,270      | 25,270       |
| 5                           | 25,280      | 25,280       |

The result of BGR diagnostic for the MCMC convergence of estimated parameter \( w_k \) is shown by figure 1. It can be shown that the graphics which have a green color line describe the dynamical movement of the pooled posterior variance, whereas the blue color line represents an evolution of the average within-sample variance, and the PSRF is shown by a red color line. The PSRF which tends to one indicates that the MCMC convergence of estimated parameter \( w_k \) is achieved for \( k = 1,2,3,4 \).
Figure 1. Dynamical movement of PSRF indicator for the MCMC of the estimated parameter $w_1$, $w_2$, $w_3$, and $w_4$ on 100,000 iterations.

BGR indicators for each $w_k$, for $k = 1, 2, 3, 4$, which describe the convergent of MCMC chains are supported by historical simulations of $w_1$, $w_2$, $w_3$, and $w_4$ on figure 2. Trace plots of historical simulations which look like “fat hairy caterpillar” indicate that the Markov chains are convergent as stated on [6].
**Figure 2.** Historical simulation of the MCMC of an estimated parameter $w_1, w_2, w_3,$ and $w_4$ on 100,000 iterations.

The convergent conditions of Markov chains imply that the chains reach the posterior distribution. So it produces the valid posterior inference for parameters and estimated parameters represent on table 2 can give valid results.

**Table 2.** Estimated parameter of $w_k, \alpha_k,$ and $\beta_k$

| The $k$-th mixture component | $w_k$ | $\alpha_k$ | $\beta_k$ |
|------------------------------|------|-----------|----------|
| 1                            | 0.4915 | 0.9077 | 4.070-8 |
| 2                            | 0.1044 | 0.9077 | 2.971-8 |
| 3                            | 0.0859 | 0.9077 | 2.315-8 |
| 4                            | 0.3181 | 0.9077 | 1.915-8 |

The final model can be arranged as

$$g(y \mid w, \theta) = \sum_{k=1}^{4} w_k f(y \mid \theta_k)$$

$$= 0.4915 \text{Gamma}(0.9077, 4.070-8) + 0.1044 \text{Gamma}(0.9077, 2.971-8)$$

$$+ 0.0859 \text{Gamma}(0.9077, 2.315-8) + 0.3181 \text{Gamma}(0.9077, 1.915-8).$$

**5. Conclusion**

Bayesian inference through Markov chain Monte Carlo (MCMC) method and widely applicable information criteria (WAIC) can give an assurance result on the parameter estimation process of the finite gamma mixture model. Based on the household income which is constructed through the
Indonesian Family Life Survey (IFLS) 2014-2015 data, it can be modeled by a finite mixture model with four mixture components of the gamma distribution.

Acknowledgments
We are grateful to the Ministry of Research, Technology and Higher Education Indonesia which has funded this research under PDD research grant no. 474/UN27.21/PP/2018.

References
[1] Frühwirth-Schnatter S 2006 *Finite mixture and Markov switching models* (Springer Science & Business Media)
[2] Iriawan N 2012 *Pemodelan dan Analisis Data-Driven* (Surabaya: ITS Press)
[3] Marin J-M, Mengersen K and Robert C P 2005 Bayesian modeling and inference on mixtures of distributions *Handbook of statistics* 25 459–507
[4] Ntzoufras I 2011 *Bayesian modeling using WinBUGS* vol 698 (John Wiley & Sons)
[5] Gelman A and Rubin D B 1992 *Statistical science* 7 457–72
[6] Tatarinova T and Schumitzky A 2015 *Nonlinear Mixture Models: A Bayesian Approach* (World Scientific)
[7] Brooks S P and Roberts G O 1998 *Statistics and Computing* 8 319–35
[8] Brooks S P and Gelman A 1998 *Journal of computational and graphical statistics* 7 434–55
[9] Watanabe S 2010 *Journal of Machine Learning Research* 11 3571–3594
[10] Gelman A, Hwang J and Vehtari A 2014 *Stat Comput* 24 997–1016
[11] Wiper M, Insua D R and Ruggeri F 2001 *Journal of Computational and Graphical Statistics* 10 440–54
[12] Goudie R J B, Turner R M, De Angelis D, and Thomas A 2017 MultiBUGS: Massively parallel MCMC for Bayesian hierarchical models arXiv:1704.03216 [stat]
[13] Plummer M, Best N, Cowles K, and Vines K 2006 *R news* 6 7–11
[14] Strauss J, Witoelar F and Sikoki B 2016 *The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report* (Santa Monica: RAND Corporation)
[15] Kruschke J 2014 *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Academic Press)
[16] Susanto I, Iriawan N, Kuswanto H, Suhartono, Fithriasari K, Ulama B S S, Suryaningtyas W, and Pravitasari A A 2018 *O J. Phys.: Conf. Ser.* 1090 012014