M - 13

BAYESIAN WITH FULL CONDITIONAL POSTERIOR DISTRIBUTION APPROACH FOR SOLUTION OF COMPLEX MODELS

Pudji Ismartini

BPS-Statistics Indonesia, Sekolah Tinggi Ilmu Statistik (STIS)
Jakarta, Indonesia
Email : ismartini@bps.go.id, ismartini09@yahoo.com

Abstract

Complex models are often arise in social field applications. Model complexity might be caused by the number of variables in the model, or due to the complex data structures. Problems are often encountered in complex models is a difficulty to obtain a solution in the model parameter estimation process. Bayesian methods can overcome these problems with the modern approach of Bayesian analysis using special simulation procedure based on the posterior distribution of parameters, i.e. the Markov Chain Monte Carlo (MCMC). Implementation of MCMC methods for Bayesian analysis requires proper sampling algorithm in order to obtain a sample from a distribution. The algorithm which is efficient and often used by MCMC is Gibbs Sampling. One of the advantages of Gibbs sampling is the generation of random variables is done using the concept of one-dimensional distribution which are structured as a form of full conditionals, i.e. the full conditional posterior distribution of parameter. This paper propose to describe the process of parameter estimation for complex model using Bayesian through full conditional posterior distribution of parameters.

Keywords: Bayesian, Complex models, MCMC, Gibbs Sampling, Full Conditional Posterior Distribution

1 INTRODUCTION

In a real world phenomena, it is often faced complex models due to some reasons. First, it is caused by big number predictors used in the model. It is arisen in many areas of science and technology, especially in social science since social phenomena are generally influenced by huge dimensions of life. Thus modelling in social science might deal with huge predictors. Second, complex model might occur because of the uncommon particular distribution of data. In general, data modeling using the normal distribution assumption. However, the data are often encountered cases that are not normally distributed. For example, data of per capita household expenditure which are log normal distributed with three parameters. Third, complex models may be caused by the complex structure of the data. In the social field, in general, the data has a hierarchical structure where data is nested within its territory.

Estimation methods commonly used in statistical modeling is the classical approach using the data likelihood function. In the classical approach, the model parameters to be estimated is assumed to be single-valued. Estimation process with the classical approach, commonly done to obtain parameter values that maximizes the Likelihood function which is considered as a function of these parameters. In some cases, the estimation process with the classical methods generally use numerical optimization techniques to obtain the solution. However, in the case of complex models is often difficult to obtain the solution of classical methods.

Bayesian inference approach is different from the classical approach even though both methods use the data Likelihood function. The Bayesian approach, all unknown parameters are viewed as random variables characterized by prior distributions of the parameters (Ntzoufras, 2009; Gelman et al., 2004; Congdon, 2006). Prior distribution of the parameters stated variations of these parameters.

In contrast to classical approaches, Bayesian methods do not involve the optimization process in inference. Bayesian approach is applying Bayes' theorem based on the joint posterior distribution of all parameters (King, Morgan, Gimenez and Brooks, 2010). Furthermore, the Bayesian method will perform parameter estimation using marginal posterior distributions of the parameters. The marginal posterior distributions obtained by integrate joint posterior distribution. At this stage, especially for the case of fairly complex models generally arise the problems of the integration process becomes very complicated and difficult to obtain the solution. However, Bayesian methods can overcome these problems. In this case, the method used in the modern approach of Bayesian analysis is not using the integration of joint posterior distribution analytically, but by using the data simulation procedure that follows a joint posterior distribution through the full conditional form in order to obtain the marginal posterior distribution of each parameter which will be estimated. Thus, the optimization process is performed in a classical analysis is replaced with the process of integration in the Bayesian analysis approach (King et al., 2010). Integration in Bayesian analysis is not done analytically to the joint posterior distribution of the parameters, but with a special approach to the simulation procedure that produces a set of sample from the posterior distribution. This process became known as the Markov Chain Monte Carlo (MCMC) (Koop 2003; King et al., 2010). This paper propose to describe the process of parameter estimation for complex model using Bayesian through full conditional posterior distribution of parameters.

2 METHODS

The Bayesian method is adopted from the name of the inventor of the method, namely Thomas Bayes (1702-1761). However, the new Bayesian method became known in 1764 after Thomas Bayes died. Although Bayesian methods have existed since the 18th century, but until the early 20th century, the Bayesian method is less popular than classical methods (frequentist). The use of Bayesian methods begins to increase as the development of information technology so the difficulty in data analysis in order to get the solution analytically, can be obtained using computer simulations (Gelman et al., 2004).

2.1 Bayesian Method

Bayesian method is based on the Bayes' Theorem (Box and Tiao, 1992; Gelman and Hill, 2007). According to the rules of probability in the Bayes theorem, the posterior distribution of the parameters can be expressed as followed:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(\mathbf{y})}, \tag{1}$$

where, $\boldsymbol{\theta}$ and $\mathbf{y}$ are both random, $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_d \end{bmatrix}^T$ is parameter vector dan $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$ denotes vector of observations from the sample. $f(\mathbf{y})$ is defined as normalized constant with respect to $\boldsymbol{\theta}$. $f(\boldsymbol{\theta})$ is prior distribution of parameter $\boldsymbol{\theta}$ and $f(\mathbf{y} \mid \boldsymbol{\theta})$ is a likelihood function of data. Then, the posterior in Equation (1) can be represented as a proportional form as follows:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta}). \tag{2}$$

It is shown in Equation (2), the posterior is proportional to the combination of prior information and current information of data. All information about the unknown parameter of interest is included in their joint posterior distribution.

2.2 Markov Chain Monte Carlo (MCMC)

MCMC is a method that is widely implemented in the Bayesian modeling approach to obtain solutions for existing posterior distribution which requires complicated calculation since it involves integration process with high dimensions (Carlin and Chib, 1995). The idea of MCMC is to generate a set of parameter data based on Markov Chain process by using Monte Carlo simulation iteratively to obtain the posterior distribution which is stationary (steady state) (Ntzoufras, 2009). Markov Chain adalah suatu proses stokastik dari $\left\{\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},...,\boldsymbol{\theta}^{(T)}\right\}$ sedemikian sehingga

$$f(\boldsymbol{\theta}^{(t+1)} \mid \boldsymbol{\theta}^{(t)},...,\boldsymbol{\theta}^{(1)}) = f(\boldsymbol{\theta}^{(t+1)} \mid \boldsymbol{\theta}^{(t)}).$$

The MCMC algoritm is as follow (Ntzoufras, 2009):

1. Define initialize value of parameters $\boldsymbol{\theta}^{(0)}$
2. Generate sample of parameters through iteratively suppose M times of iteration
3. *burn in* process for the first B iteration in order to reach equilibrium state
5. Use $\left\{\boldsymbol{\theta}^{(B+1)},\boldsymbol{\theta}^{(B+2)},...,\boldsymbol{\theta}^{(M)}\right\}$ as a sample in posterior analysis

2.3 Gibbs Sampling

Implementation of MCMC methods for Bayesian analysis requires proper sampling algorithm to obtain a sample from a distribution. The algorithm is often used as a generator of random variables in the MCMC is the Gibbs sampling (Casella and George, 1992; Gelman et al., 2004). Gibbs Sampling is a technique for generating random variables from a marginal distribution directly without having to calculate the density function of the distribution (Casella and George, 1992). Gibbs sampling process is done by taking a sample of Gibbs Sequence based on the basic properties of the Markov Chain (Casella and George, 1992). Gibbs Sampling algorithm is illustrated in Figure 1.

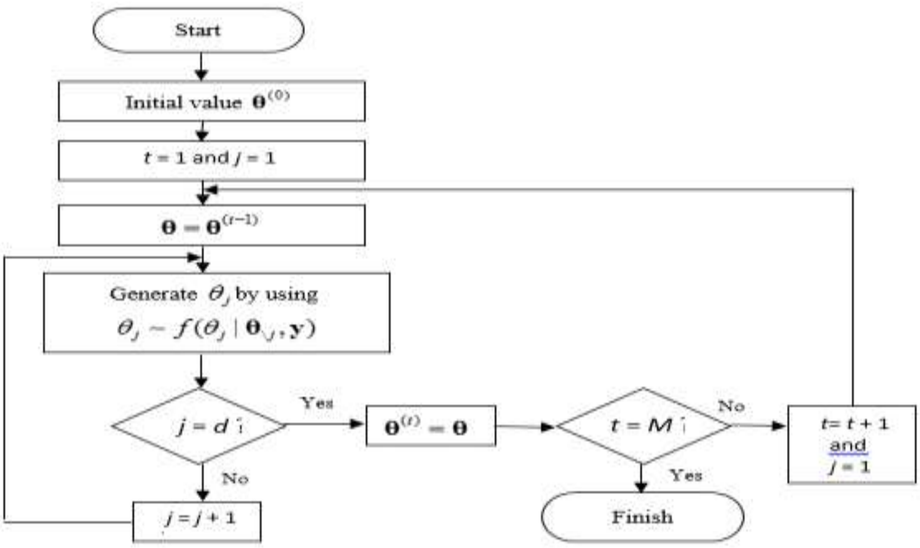

Figure 1. Gibbs Sampling Algorithm

3 Discussions

Gibbs Sampling using the concept of phasing set of unknown parameters and estimate the parameter or set of parameters one at a time with a given value of the other parameters and the data sample used (Lunn, Thomas, Best, and Spiegelhalter, 2000). This is one of the advantages of Gibbs Sampling because it generates the random variables by using the concept of unidimensional distributions which are structured as a form of full conditionals.

Assume $\boldsymbol{\theta}$ is a vector of a set of parameters to be estimated, the full conditional posterior for each parameter can be formulated. The full conditional posterior of parameter θ_j is specified from the joint posterior distribution of all parameters in the vector $\boldsymbol{\theta}$ by setting constant value for other parameter than θ_j, i.e. $\boldsymbol{\theta}_{\setminus j} = \begin{bmatrix} \theta_1 & \cdots & \theta_{j-1} & \theta_{j+1} & \cdots & \theta_d \end{bmatrix}^T$, where $\boldsymbol{\theta}_{\setminus j}$ is vector of parameter $\boldsymbol{\theta}$ without θ_j and *d* is number of parameter to be estimated. Thus, the full conditional posterior distribution of θ_j, i.e. $f(\theta_j \mid \boldsymbol{\theta}_{\setminus j}, \mathbf{y})$, is proportional form of joint posterior distribution of all parameters which only contains the component parameters θ_j, while the other components are eliminated. The full conditional posterior distribution for each parameter was then used in the iteration process of parameter estimation using MCMC with Gibbs sampling.

Suppose Y is random variable which follow Log Normal Three Parameters distribution (LN3) or $Y \sim LN3(\mu_{[y]}, \sigma^2_{[y]}, \lambda)$. If $\tau_{[y]} = 1/\sigma^2_{[y]}$ is precision parameter, the probability density function of LN3 is as follow (Norstad, 2011):

$$f(y \mid \mu_{[y]}, \tau_{[y]}, \lambda) = \frac{1}{(y-\lambda)} \sqrt{\frac{\tau_{[y]}}{2\pi}} \exp\left[-\frac{\tau_{[y]}}{2} (\ln(y-\lambda) - \mu_{[y]})^2 \right].$$

Since $E\left(\ln\left(y_i - \lambda\right)\right) = \mathbf{x}_i^T \boldsymbol{\beta},$ the likelihood function of LN3 random variable is :

$$\begin{aligned} f_L(\mathbf{y} \mid \boldsymbol{\beta}, \tau_{[y]}, \lambda) &= \prod_{i=1}^{n} f(y_i \mid \boldsymbol{\beta}, \tau_{[y]}, \lambda) \\ &= \prod_{i=1}^{n} \frac{1}{(y_i - \lambda)} \sqrt{\frac{\tau_{[y]}}{2\pi}} \exp\left[-\frac{\tau_{[y]}}{2} (\ln(y_i - \lambda) - \mathbf{x}_i^T \boldsymbol{\beta})^2 \right] \end{aligned}$$

(3)

The prior of parameter $\beta, \tau_{[y]}$, and λ are defined as follow:

$$\beta_r \sim N(\mu_{[\beta]r}, \sigma^2_{[\beta]r}), \tag{4}$$

$$\tau_{[y]} \sim \text{Gamma}(\text{a}_{[\tau_{[y]}]}, \text{b}_{[\tau_{[y]}]}), \tag{5}$$

$$\lambda \sim N(\mu_{[\lambda]}, \sigma^2_{[\lambda]}). \tag{6}$$

The set of parameter to be estimated is then $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau_{[y]}, \lambda)$ where $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_k \end{bmatrix}^T$. According to Equation (2), joint posterior distribution of parameters is formulated by using Equation (3), (4), (5) and (6) :

$$f(\boldsymbol{\beta},\tau_{[y]},\lambda\,|\,\mathbf{y}) \propto \prod_{i=1}^{n} f(y_i\,|\,\boldsymbol{\beta},\lambda,\tau_{[y]})\prod_{r=0}^{k} p(\beta_r)\; p(\tau_{[y]})\,p(\lambda)$$

$$\propto \frac{\tau_{[y]}^{n/2}}{\prod_{i=1}^{n}(y_i-\lambda)}\exp\left[-\frac{1}{2}\tau_{[y]}\left\{\sum_{i=1}^{n}(\ln(y_i-\lambda)-\mathbf{x}_i^T\boldsymbol{\beta})^2\right\}\right]\;\tau_{[\beta]r}^{n/2}\exp\left[-\frac{1}{2}\sum_{r=0}^{k}\tau_{[\beta]r}(\beta_r-\mu_{[\beta]r})^2\right]$$

$$\tau_{[y]}^{\mathbf{a}_{[\tau_{[y]}]}-1}\exp\left[-\frac{\tau_{[y]}}{\mathrm{b}_{[\tau_{[y]}]}}\right]\;\tau_{[\lambda]}^{1/2}\exp\left[-\frac{\tau_{[\lambda]}}{2}(\lambda-\mu_{[\lambda]})^2\right].$$

(7)

The full conditional posterior distribution of each parameter is a proportional form of joint posterior distribution in Equation (7) which is only contain a parameter that will be estimated. Thus, other component of parameters are eliminated since it is treated as a constant value. The full conditional posterior distribution of each parameter is then defined as follow:

(i) $$f(\beta_r\,|\,\mathbf{y},\boldsymbol{\beta}_{\backslash r},\lambda,\tau_{[y]}) \propto \exp\left[-\frac{1}{2}\tau_{[y]}\left\{\sum_{i=1}^{n}(\ln(y_i-\lambda)-\mathbf{x}_i^T\boldsymbol{\beta})^2\right\}\right]\;\exp\left[-\frac{1}{2}\sum_{r=0}^{k}\tau_{[\beta]r}(\beta_r-\mu_{[\beta]r})^2\right], \quad (8)$$

(ii) $$f(\lambda\,|\,\mathbf{y},\boldsymbol{\beta},\tau_{[y]}) \propto \frac{1}{\prod_{i=1}^{n}(y_i-\lambda)}\exp\left[-\frac{1}{2}\tau_{[y]}\left\{\sum_{i=1}^{n}(\ln(y_i-\lambda)-\mathbf{x}_i^T\boldsymbol{\beta})^2\right\}\right]\exp\left[-\frac{\tau_{[\lambda]}}{2}(\lambda-\mu_{[\lambda]})^2\right], \quad (9)$$

(iii) $$f(\tau_{[y]}\,|\,\mathbf{y},\boldsymbol{\beta},\lambda) \propto \tau_{[y]}^{n/2}\exp\left[-\frac{1}{2}\tau_{[y]}\left\{\sum_{i=1}^{n}(\ln(y_i-\lambda)-\mathbf{x}_i^T\boldsymbol{\beta})^2\right\}\right]\tau_{[y]}^{\mathbf{a}_{[\tau_{[y]}]}-1}\exp\left[-\frac{\tau_{[y]}}{\mathrm{b}_{[\tau_{[y]}]}}\right]$$

(10)

The process of parameter estimation is conducted using MCMC and Gibbs Sampling. This process is iterative process that follows the Markov Chain process iteratively to estimate the parameters. This iterative estimation process use a given value of the parameter information from the previous iteration step. This process will take place in full conditionals form of each parameter and these are arranged alternately as the next phase of iterative stochastic simulations (Gelman, et al, 2004). Gibbs sampling iteration procedure is done with the following steps (Ismartini, Iriawan, Setiawan dan Ulama, 2012):

Step 1 : Define initial value for each parameter which will be estimated ($\boldsymbol{\beta}^{(0)},\boldsymbol{\lambda}^{(0)},\boldsymbol{\tau}_{[y]}^{(0)}$).

Step 2 : Generate sample of each parameter by conducted *M* iteration process through the full conditional posterior distribution, i.e.:

i) Generate β_r using Equation (8),
ii) Generate λ using Equation (9),
iii) Generate $\tau_{[y]}$ using Equation (10).

This step is conducted iteratively.

4 CONCLUSIONS

Parameter estimation technique using Bayesian approach with full conditional posterior distributions of the parameters is an alternative solution that can be applied to complex modeling cases. This is possible since the full conditional posterior distribution is using the concept of unidimesional distribution. Thus, the solution of a complex estimation process can be obtained by using this iterative method.

REFERENCES

Box, G.E.P dan Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley.

Carlin, B.P. dan Chib, S. (1995), Bayesian model choice via Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society*, Ser. B, **57**(3): 473-484.

Casella, G. dan George, E.I. (1992), Explaining the Gibbs Sampler. *Journal of the American Statistical Association* **46**(3): 167-174.

Congdon, P.D. (2006), *Bayesian Statistical Modelling*, 2nd edition, John Wiley & Sons, England.

Gelman, A., Carlin, J B., Stern, H. S. dan Rubin, D. B. (2004), *Bayesian Data Analysis*, 2nd edition, Chapman & Hall, Florida.

Gelman, A., dan Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, UK.

King, R., Morgan, B.J.T, Gimenez, O., dan Brooks, S.P. (2010), *Bayesian Analysis for Population Ecology*, Chapman & Hall/CRC, USA.

Ismartini, P., Iriawan, N., Setiawan, dan Ulama, B.S.S. (2012c), Toward a Hieararchical Bayesian Framework for Modelling the Effect of Regional Diversity on Household Expenditure, *Journal of Mathematics and Statistics, Science Publication*, **8**(2): 283-291, 25 June 2012. ISSN: 1549-3644, DOI: 10.3844/jmssp.2012.283.291.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons, Chichester, England.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), WinBUGS – A Bayesian Modelling Framework: Concepts, Structure and Extensibility, *Statistics and Computing*, **10**, pp. 325-337.

Norstad, J. (2011), The Normal and Lognormal Distributions, *unpublished paper*, www.norstad.org/finance/normdist.pdf (31 July 2012).

Ntzoufras, I. (2009), *Bayesian Modeling Using WinBUGS*. Wiley, New Jersey, USA.