Journal of Mathematics and Statistics

Original Research Paper

An Algorithm for Determining the Number of Mixture Components on the Bayesian Mixture Model Averaging for Microarray Data

[1]Ani Budi Astuti, [2]Nur Iriawan, [3]Irhamah and [4]Heri Kuswanto

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Jl Veteran, Malang 65145, Indonesia and Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Jl. Arif Rahman Hakim, Surabaya 60111, Indonesia
[2, 3, 4]Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Jl. Arif Rahman Hakim, Surabaya 60111, Indonesia

Article history
Received: 27-09-2014
Revised: 21-08-2015
Accepted: 16-09-2015

Corresponding Author:
Ani Budi Astuti
Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Brawijaya, Jl Veteran, Malang 65145, Indonesia and Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Jl. Arif Rahman Hakim, Surabaya 60111, Indonesia
Email: ani_budi@ub.ac.id and ani_budi@staff.ub.ac.id

Abstract: The major challenges on the statistical analysis of microarray data are the limited availability of samples, large number of measured variables and the complexity of distribution of the data obtained (e.g., multimodal). These phenomena could be considered in Bayesian method, used Bayesian Mixture Model (BMM) methods and Bayesian Model Averaging (BMA) methods. Modeling of Bayesian Mixture Model Averaging (BMMA) for microarray data was developed based on these two studies. One of the most important stages in BMMA is determination of the number of mixture components in the data setting as the most appropriate BMMA models. This paper proposes an algorithm for determining the number of mixture components in BMMA for microarray data. The algorithm is developed based on the simulation data generated from a case study of Indonesian and it has been implemented on the outside Indonesian microarray data. The results have succed to demonstrate two step algorithms, called Preliminary Process and Smoothing Process Algorithms, to the Indonesian case microarray data with the accuracy rate of 99.3690% and 99.9094% for the outside Indonesian microarray data.

Keywords: Algorithm, Number of Mixture Components, Bayesian Mixture Averaging, Microarray

Introduction

Microarray is an analysis technique to monitor the activity of thousands of gene expressions simultaneously. Gene expression data is the data obtained from microarray experiments (Ibrahim *et al*., 2002; Knudsen, 2004; Chun *et al*., 2014; Harijati and Keane, 2012). The major challenges for the statistical analysis of microarray data are limited availability of the samples, large number characteristics of the variables would be measured and the distribution of the data would be very complex. It can be shown in various studies that have been conducted. Wholey (2012) was only able to replicate 3 observations to identify the effect of stress levels on E. Coli as a result of *hypochlorous* acid using cluster analysis. Li (2009) was also only able to replicate 3 observations to identify patterns and development of the small intestine of mouse through *epithelial* and *mesenchymal* signal using t-test couple with cluster analysis. In addition, the nature of the distribution of obtained microarray data in most cases is very complex (Muller *et al*., 2004; Eman *et al*., 2012).

Bayesian method had been proposed as a statistical analysis that did not consider the number of samples and hence this method could be used for analyzing small or large number of samples with any distribution (Congdon, 2006; Ghosh *et al*., 2006; Ahmed *et al*., 2010; Anggorowati *et al*., 2012; Diana *et al*., 2013; Amran *et al*., 2013; Astuti *et al*., 2014; 2015; Yuan-Ying *et al*., 2014). All parameters in the model are treated as random variable in Bayesian method (Gelman *et al*., 1995). According to Mengersen (2009), the advantage of Bayesian method is able to provide an inference of unknown variables based on the posterior distribution from the data directly.

Various studies that have been previously conducted for microarray data, showed that its distribution tends to follow multimodal distribution and a suitable arbitrary distribution. Do *et al*. (2005) used a Bayesian Mixture Model (BMM) methods to identify gene expression differences from microarray data, whilst Sebastiani *et al*. (2006) used Bayesian Model Averaging (BMA) methods to identify those differences. This paper would introduce Bayesian Mixture Model Averaging (BMMA) for microarray data based on these two studies, which is proposed to answer the challenges of modeling microarray data. There are several important stages that must be carried out in the BMMA in which, the most important step is to determine the number of mixture components which suitables to the data. If this step reports the wrong number of component, then BMMA model would fail to represent the real condition of the data (Iriawan, 2012). To do those purposes, this paper firstly proposes an algorithm for determining the number of mixture components in BMMA through the simulation data generated from a case study of Indonesian microarray data (Harijati, 2007) and demonstrates the work of this algorithm to be implemented on the data used in Do *et al*. (2005).

Materials and Methods

Data Sets

To develop the algorithm, the simulation data generated based on the pattern of some genes expression differences of Chickpea (*Cicer Arietinum*) plant tissues which was researched by Harijati (2007) in Indonesia. Her data was collected only for 3 observations on each gene ID to all of 4,000 genes ID. There were only 62 genes ID of those 4,000, however, were expressed completely. Among these 62 genes ID, there were found fifteen genes ID with the very important function (as the defence function). These fifteen genes ID, which are normally distributed, were taken for the simulated data in this research. The characteristics of these fifteen data are then used to simulate the number of 10,000 simulated genes ID and repeated 10 times to guarantee the accuracy of these algorithms. Implementation of the algorithms would be done on Do *et al*. (2005) data by taking a sample size of 3-combination.

Bayesian Analysis

The concept of Bayesian analysis using Bayes theorem was invented by Thomas Bayes in 1702-1761, where in this analysis, model parameter, $\theta \in \Omega$, is treated as a random variable. This concept explains that the Bayesian analysis is a statistical analysis method based on the posterior probability distribution model. This distribution is a blend of two information i.e., the information of past data as prior information and the information of observations (samples) data as a constituent likelihood function to update the prior information (Box and Tiao, 1973; Zellner, 1971; Gelman *et al*., 1995; Aitkin, 2001; Iriawan, 2003).

The process of Bayesian analysis can be described as follows: Suppose if there are observational data x with a likelihood function $f(x|\theta)$ then the information known about the parameters θ before observations were made is referred to as the prior θ, namely $p(\theta)$. Furthermore, to determine the posterior probability distribution of θ, namely $p(\theta|x)$ based on the rules of probability in Bayes' theorem as in Equations 1 (Ghosh *et al*., 2006):

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)} \quad (1)$$

Where:

$$f(x) = \begin{cases} \int_{\theta \in R} f(x|\theta) f(\theta) d\theta, & \theta \text{ is continuous} \\ \sum_{\theta \in B} f(x|\theta)\ p(\theta), & \theta \text{ is discrete} \end{cases}$$

$f(x)$ is a normalized constant (Gelman *et al*., 1995) so Equations 1 can be rewritten as:

$$p(\theta|x) \propto f(x|\theta)p(\theta) \quad (2)$$

Based on the Equations 2 we obtain that the posterior probability is proportional to the product of the likelihood function and the prior probability of the model parameters. This means, it updates the prior information using the information of samples in the data likelihood to obtain the posterior information that will be used for decision making (Iriawan, 2003).

Mixture Models

Mixture model is a special model for the data having the characteristics of multimodal i.e., data consisted sub-sub-population or groups, where each proportion of sub-population or group is a constituent component of the mixture models. Mixture model called the particular model because this model is able to combine some different data but still retains the characteristics of the original data (McLachlan and Basford, 1988; Gelman *et al*., 1995; Astuti, 2006). This model will able to provide flexible parametric framework in modeling and statistical analysis (Marin *et al*., 2005).

According to McLachlan and Basford (1988); Gelman *et al*. (1995) and Iriawan (2001), mixture probability function of an observation $x=(x_1,x_2,\ldots,x_n)$ taken from a number of *k*-sub-population can be expressed as in Equations 3:

$$f(x|\theta,w) = \sum_{j=1}^{k} w_j g_j(x|\theta_j) \quad (3)$$

where, $f(x|\theta,w)$ is a function of the probability mixture, $g_j(x|\theta_j$ is a j^{th} probability function of k number of sub-population that make up a model and θ is a mixture model parameters containing of (θ_1, θ_2, …θ_k). Parameter θ_j, j = 1,2,…,k represents the characteristic distribution of g_j on each component in mixture models. While w is the parameter vector of proportions (weighted) mixture model containing of (w_1, w_2,…, w_k), where $0 < w_j < 1$, $\forall j$ and $\sum_{j=1}^{k} w_j = 1$ for each model parameter θ_j. Mixture model described in Equations 3 would be applied to model a finite mixture of k-particular number of components (Astuti, 2006).

Bayesian Mixture Model (BMM) Analysis

According to Richardson and Green (1997), to model such data into a mixture model, each observation x_i would be classified on each unknown number of sub-population. If the allocation of each observation on each sub-population in Equations 3 is denoted by z, then the allocation of each observation z_i, i= 1, 2,…,n could be determined based on Equations 4:

$$p(z_i = j) = w_j, j = 1,2,\cdots,k \tag{4}$$

Given the value of z_i then the observation data x_i can be derived from the sub-populations as in Equations 5:

$$x_i | z_i \sim f(x|\theta_{z_i}), i = 1,2,\cdots,n \tag{5}$$

Thus the resulting joint posterior distribution of all parameter in the mixture model can be expressed as in Equations 6:

$$\begin{aligned} &p(k, w, z, \theta, x) = \\ &p(k)p(w|k)p(z|w,k)p(\theta|z,w,k)p(x|\theta,z,w,k) \end{aligned} \tag{6}$$

The next process is to estimate each parameter in Equations 6 by employing the full conditional distribution of each parameter (Richardson and Green, 1997).

Full Conditional Distributions

Suppose there is a parameter θ which has a stationary distribution of $p(\theta)$. Full conditional distribution of parameter θ is constructed by making a partition of θ as shown in Equations 7 (Wati, 2006):

$$\theta = (\theta_S, \theta_{-S}) \tag{7}$$

where, θ_s denotes the-s^{th} parameter to be estimated and θ_{-S} denote the complement of θ_s that is parameter θ without including the-s^{th} component.

Gilks (1995) and Wati (2006) explains that the full conditional distribution can be established based on the joint posterior distribution as in Equations 8:

$$p(\theta_S | \theta_{-S}) = \frac{p(\theta_S, \theta_{-S})}{\int p(\theta_S, \theta_{-S}) d\theta_S} \tag{8}$$

Bayesian Model Averaging (BMA) Analysis

The basic principle of BMA is to form the best single model by considering all possible models. BMA is a Bayesian solution for uncertainty model, where the completion of the model is done by averaging the posterior distribution of all best models (Madigan and Raftery, 1994; Montgomery and Nyhan, 2010; Kuswanto and Sari, 2013; Astuti *et al*., 2014; 2015).

According to Hoeting *et al*. (1999), parameters estimation using BMA approach is done by combination of all best possible hierarchical models to data. If {M_1, M_2,…, M_l} are the set of models from M and Δ is paramters that would be estimated, then BMA estimation will begin with determining model M_k, k =1,2,…,l and each prior probability distribution of all parameters in each model. Implicitly, probabilityy of each model, M_k, k =1,2,…,l to fit the data (x) would influence to the expected value of the parameter Δ. Therefore, this parameters can be obtained by averaging all of model M_k, k =1,2,…,l, that is as in Equations 9:

$$E(\Delta | x) = \sum_{k=1}^{l} P(M_k | x) E(\Delta | M_k, x) \tag{9}$$

Equations 9 shows weighted expected value of Δ in every possible combination models, whilst the variance of ($\Delta|x$) is shown in Equations 10:

$$Var(\Delta | x) = \sum_{k=1}^{l} \begin{pmatrix} var(\Delta | x, M_k) + \\ [E(\Delta | M_k, x]^2) P(M_k | x) - E(\Delta | x)^2) \end{pmatrix} \tag{10}$$

Bayesian Mixture Model Averaging (BMMA) Analysis

BMMA is proposed method which is build by combining the two methods, BMM and BMA, wherein the model parameter estimators are obtained by averaging all of the possible mixture models. Therefore, the form of BMMA can be determined by applying Equations 9 and 10 to the some mixture models. This finding BMMA, therefore, have the value of $E(\Delta|x)$ as the weighted value of the parameters in any combination of mixture models and $P(M_k|x)$ is the posterior distribution of the k^{th} mixture model.

Results and Discussion

Differences of Gene Expression Data in Harijati (2007)

Gene expressions differences data in Harijati (2007) are used as the basis pattern for simulation data. The Pattern of data for 15 genes ID of Chickpea (*Cicer Arietinum*) plant tissues from Harijati (2007) are shown in Fig. 1.

Based on the Fig. 1, it can be seen that the data looks as unimodal or multimodal distribution. As an example, gene ID LS0024-Dfc represents a bimodal distribution. Therefore the mixture distribution approach must be applied. The detection to number of components in the mixture, the two step algorithms is proposed, called Preliminary Process Algorithm and Smoothing Process Algorithm.

The Proposed Preliminary Process Algorithm

Preliminary Process Algorithm is proposed as an algorithm for determining the boundary value as the separator of different mixture components. The algorithm has 5 steps of the Proposed Preliminary Process Algorithm.

Algorithm 1: Preliminary Process

Step (1). Sort each observed gene ID, from smallest to largest, namely $X_{(1)p}$, $X_{(2)p}$, $X_{(3)p}$ with p= 1,2,3,…, q and q is the number of gene ID.

Step (2). Determine the mean of the data of each gene ID based on the results of step (1) using

$$\bar{X}_p = \frac{\sum_{i=1}^{3} X_{(i)p}}{3}$$

Step (3). Determine the grand mean of all gene ID from step (2) using $(\bar{X})_{total} = \frac{\sum_{j=1}^{q} \bar{X}_p}{q}$.

Step (4). Determine the first boundary value using

$$c_{boundary1} = \left[\frac{(\bar{X})_{total}}{2} + 0.017485114\right]$$

Step (5). Determine the second boundary value using $d_{boundary2} = \left[\frac{(\bar{X})_{total}}{3} - 0.021175691\right]$

Based on the Algorithm 1, it can be seen that the boundary value would be as the separator of different mixture components in the microarray data. Derived from Algorithm 1, we propose an algorithm for determining the number of mixture components in the microarray data, called Smoothing Process Algorithm.

The Proposed Smoothing Process Algorithm

This algorithm is used to determine the optimization of the number of mixture components. The steps of the Proposed Smoothing Process Algorithm are as in Algorithm 2.

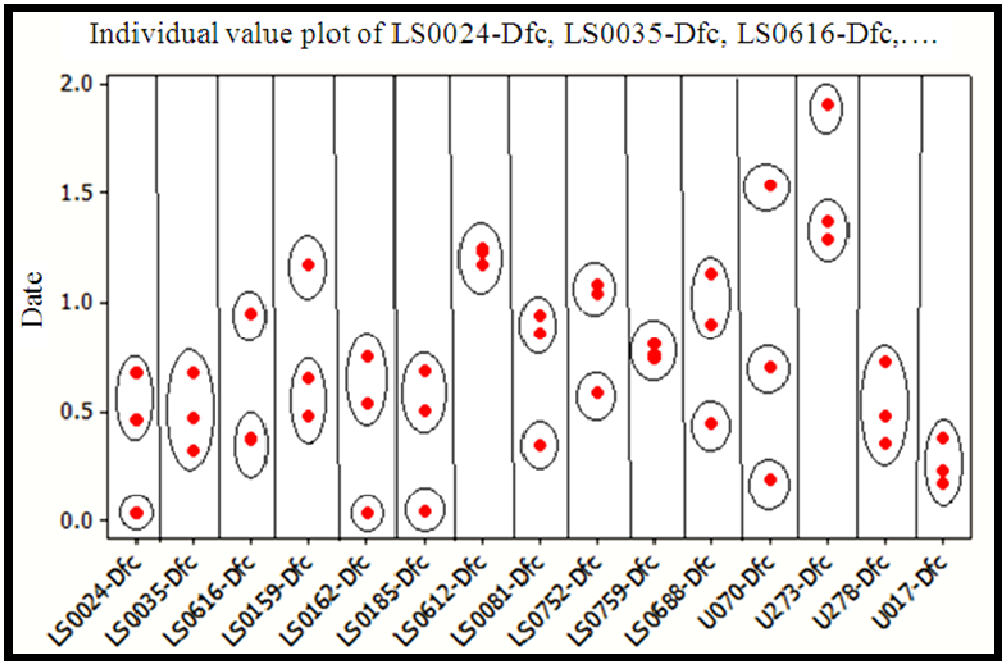

Fig. 1. The mixture pattern of Harijati (2007) data

Algorithm 2. Smoothing process

Step (1). Do step (1) to step (5) on the Preliminary Process Algorithm in Algorithm 1.

Step (2). Calculate the standard deviation value for each p^{th} gene ID from $X_{(1)p}$ and $X_{(2)p}$ namely stdev_{12p} calculated as $\text{stdev}_{12p} = \frac{\sum_{i=1}^{2}(X_{(i)p} - \bar{X}_{(1)(2)p})^2}{(2-1)}$ and standard deviation for each p^{th} gene ID from $X_{(2)p}$ and $X_{(3)p}$ namely stdev_{23p} calculated as $\text{stdev}_{23p} = \frac{\sum_{i=2}^{3}(X_{(i)p} - \bar{X}_{(2)(3)p})^2}{(2-1)}$.

Step (3). Determine the minimum value of standard deviation for each p^{th} gene ID based on the result of step (2) that is $\text{stdev}_{\text{min}\_p} = \min(\text{stdev}_{12p}, \text{stdev}_{23p})$.

Step (4). Calculate the value of the lower limit and upper limit of the data $X_{(i)}$ for each p^{th} gene ID using the formula $X_{(i)p} \pm(1.96\times\text{stdev}_{\text{min}\_p}/\text{sqrt}(3))$. This lower and upper limit for $X_{(i)p}$ called $lwX_{(i)p}$ and $upX_{(i)p}$.

Step (5). Determine the value of $a_{1p} = abs(X_{(1)p}\text{-}X_{(2)p})$, $a_{2p} = abs(X_{(2)p}\text{-}X_{(3)p})$, $c_p = abs(X_{(1)p}\text{-}X_{(3)p})$ and $d_p = abs(a_{1p}\text{-}X_{(3)p})$ $d_p = abs(a_{1p}\text{-}a_{2p})$ of each p^{th} gene ID.

Step (6). Check if the value of $c_p < c$ boundary1then the data is unimodal.

Step (7). Check if the value of $c_p \geq c_{boundary1}$ and $d_p \geq d_{boundary2}$ and $upX_{(1)p} < lwX_{(1)p}$ and $lwX_{(3)p} \leq upX_{(2)p}$ then the data is left skewed and forms a mixture distribution with 2 components.

Step (8). Check if the value of $c_p \geq c_{boundary1}$ and $d_p \geq d_{boundary2}$ and $lwX_{(2)p} \leq upX_{(1)p}$ and $upX_{(2)p} < lwX_{(3)p}$ then the data is right skewed and forms a mixture distribution with 2 components.

Step (9). Check if the value of $c_p \geq c_{boundary1}$ and $d_p < d_{boundary2}$ then the data forms a mixture distribution with 3 components.

Step (10). Calculate the number of gene ID that cannot be identified.

Derived from the proposed two algorithms above, it can be determined the number of mixture components for microarray data. If the exact number of mixture components for microarray data is known then the modeling of BMMA for microarray data could be done.

These algorithms are only proposed and applied to detect the mixture distribution of 3 observations in each gene ID for maximum of 3 components in the mixture. When the observation of each gene ID more than 3, then the algorithm can be applied by randomly choosing of the data with sample size of 3-combination.

Accuracy of Smoothing Process and Preliminary Process Algorithms

Two algorithms above have been built based on the pattern of data Harijati (2007). Table 1 shown the accuracy of those two algorithms applied to 100,000 generated observations.

From Table 1, it can be seen that those two consecutive algorithms have a fairly high degree of accuracy, that is 99.3690%. This means that if there are 100,000 observations of genes ID then 99,369 genes ID of them can be determined the number of mixture components, whilest 631 genes ID cannot be determined.

Table 1. The accuracy of algorithms

Random	C3[a]	C4[b]	Accuracy (%)	Inaccuracy (%)
1	9,933.0	67.0	99.330	0.670
2	9,928.0	72.0	99.280	0.720
3	9,935.0	65.0	99.350	0.650
4	9,936.0	64.0	99.360	0.640
5	9,930.0	70.0	99.300	0.700
6	9,930.0	70.0	99.300	0.700
7	9,931.0	69.0	99.310	0.690
8	9,937.0	63.0	99.370	0.630
9	9,953.0	47.0	99.530	0.470
10	9,956.0	44.0	99.560	0.440
Total	99,369.0	631.0	993.690	6.310
Mean	9,936.9	63.1	99.369	0.631

[a] The number of data where mixture component can be detected from 10,000 observations generated

[b] The number of data where mixture component cannot be detected from 10,000 observations generated

Table 2. Implementation of algorithms on data in Do *et al.* (2005)

Description	Total
The number of observations tested (with sample size 3-combination of patient	9,930
Undetected mixture components	9 (0.0906%)
Detected mixture components	9,921 (99.9094%)

Implementation of Algorithms on Data in Do et al. (2005)

In this section, the two algorithms above are tested by employing data in Do *et al*. (2005), which were taken at randomly from two groups with the size of n=10 and n=12 respectively. The sample size of 3 patients as an observation is increasingly selected from 10 patients for each of 30 genes ID in the first group and from 12 patients for each of 30 genes ID in the second group. This scenario would give 10,200 observations with a sample size of 3-combination of patients. Among those observations, the sample of 9,930 observations were selected randomly. The work of the proposed couple algorithms have been demonstrated to this sample and give the result that only 9 observations of them could not be identified their number of mixture components. The accuracy of the proposed algorithms, therefore, is 99.9094%. More results can be seen in the Table 2.

Conclusion

The proposed Preliminary Process and Smoothing Process Algorithms developed based on the pattern of the Indonesian case microarray data as in Harijati (2007), has been proven with accuracy of 99.3690% when it applied to the 100,000 genes ID randomly generated from Harijati (2007) data itself. This performance is increase to 99.9094% when it was implemented to the 9,930 genes ID randomly generated from Do *et al*. (2005) data. The Preliminary Process and Smoothing Process Algorithms proposed and applied to detect the mixture distribution of 3 observations in each gene ID for maximum of 3 components in the mixture.

Acknowledgements

This paper is part of the Doctoral Research at Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS) and Surabaya, Indonesia. We would like firstly thank to Directorate General of Higher Education that has funded this research and secondly thank to Harijati and reseach group of Do, Muller and Tang to use his data and to anonymous refewer to this paper.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Ahmed, A.O.M., H.S. Al-Kutubi and N.A. Ibrahim, 2010. Comparison of the Bayesian and maximum likelihood estimation for weibull distribution. J. Math. Statistics, 6: 100-104. DOI: 10.3844/jmssp.2010.100.104.

Aitkin, M., 2001. Likelihood and Bayesian analysis of mixtures. Statistical Modelling, 1: 287-304. DOI: 10.1177/1471082X0100100404.

Amran, N. Iriawan, Subiono and Irhamah, 2013. Identification and characterization of extreme rainfalls distribution in malang residence. J. Math. Statistics, 9: 357-366. DOI: 10.3844/jmssp.2013.357.366.

Anggorowati, M.A., N. Iriawan, Suhartono and H. Gautama, 2012. Restructuring and expanding technology acceptance model: Structural equation model and bayesian approach. Am. J. Applied Sci., 9: 496-504. DOI: 10.3844/ajassp.2012.496.504

Astuti, E.Y., 2006. Implementation of the Bayesian Markov chain Monte Carlo in optimal portfolio modeling with mixture model approach of multiple mixture. MSc Thesis, ITS, Surabaya, Indonesia.

Astuti, A.B., N. Iriawan, Irhamah and H. Kuswanto, 2014. Kolmogorov-Smirnov and continuous ranked probability score validation on the Bayesian model averaging for microarray data. Applied Math. Sci., 146: 7277-7287. DOI: 10.12988/ams.2014.49760.

Astuti, A.B., N. Iriawan, Irhamah and H. Kuswanto, 2015. Occam's window selection in Bayesian model averaging modeling for gene expression data from chickpea plant. Int. J. Applied Math. Statistics, 53: 160-165.

Box, G.E.P. and G.C. Tiao, 1973. Bayesian Inference in Statistical Analysis. MA: Addison-Wesley, Massachusetts.

Chun, Y., S. Xiu-Juan, L. Xin-Min and S. Jia, 2014. The problem of power spectrum and the signal-to-noise ratio and its fast algorithm implementation in gene identification. Int. J. Applied Math. Statistics, 52: 66-72.

Congdon, P., 2006. Bayesian Statistical Modelling. 2nd Edn., John Wiley and Sons, USA, ISBN-10: 9780470035931, pp: 596.

Diana, R., I.N. Budiantara, Purhadi and S. Darmesto, 2013. Smoothing spline in semiparametric additive regression model with bayesian approach. J. Math. Statistics, 9: 161-168. DOI: 10.3844/jmssp.2013.161.168

Do, K.A., P. Muller and F. Tang, 2005. A Bayesian mixture model for differential gene expression. J. Royal Statistical Society: Series C (Applied Statistics), 54: 627-644. DOI: 10.1111/j.1467-9876.2005.05593.x.

Eman, A., E.G. Neamat and A.E.A. Iman, 2012. Support vector machine ensembles using feature-subset selection for enhancing microarray data classification. Int. J. Applied Math. Statistics, 28: 1-11.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin 1995. Bayesian Data Analysis. Chapman and Hall, London, ISBN-10: 0412039915, pp: 552.

Gilks, W.R. 1995. Full Conditional Distribution. In: Markov Chain Monte Carlo in Practice, Gilks, W.R., S. Richardson and D.J. Spiegelhalter (Eds.), CRC Press, London, ISBN-10: 0412055511, pp: 75-88.

Ghosh, J.K., M. Delampady and T. Samanta, 2006. An Introduction to Bayesian Analysis: Theory and Methods. 1st Edn., Springer-Verlag, New York, ISBN-10: 978-0-387-35433-0, pp: 354.

Harijati, N., 2007. A study of the resistance of chickpea (*Cicer Arietinum*) to *Ascochyta Rabiei* and the effect of age of plant tissue on disease development. PhD. Thesis, La Trobe University, Australia.

Harijati, N. and P.J. Keane, 2012. Disease development caused by *Ascochyta rabiei* on chickpea detached-leaves in petri dishes. Am. J. Plant Sci., 10: 1369-1375. DOI: 10.4236/ajps.2012.310165.

Hoeting, J.A., D. Madigan, A.E. Raftery and C.T. Volinsky, 1999. Bayesian model averaging: A tutorial. Statistical Sci., 14: 382-417.

Ibrahim, J.G., M.H. Chen and R.J. Gray, 2002. Bayesian models for gene expression with DNA microarray data. J. Am. Statistical Associat., 97: 88-99. DOI: 10.1198/016214502753479257.

Iriawan, N., 2001. Univariable normal mixture model estimation: A Bayesian methods approach with MCMC. Proceedings of National Seminar and Konferda VII on Mathematics in DIY and Central Java Region, Yogyakarta, Indonesian, pp: 105-110.

Iriawan, N., 2003. Simulation Technique. Teaching Modules, ITS, Surabaya, Indonesian.

Iriawan, N., 2012. Modeling and Analysis of Data-Driven. ITS Press, Surabaya, Indonesian.

Knudsen, S., 2004. Guide to Analysis of DNA Microarray Data. 2nd Edn., John Wiley and Sons, Inc., New Jersey, ISBN-10: 0471656046, pp: 160.

Kuswanto, H. and M.R. Sari, 2013. Bayesian model averaging with Markov chain Monte Carlo for calibrating temperature forecast from combination of time series models. J. Math. Statistics, 9: 349-356. DOI: 10.3844/jmssp.2013.349.356.

Li, X., 2009. Bioinformatic analysis of epithelial: Mesenchymal crosstalk during mouse gut development and patterning. PhD Thesis, University of Michigan, USA.

Madigan, D. and A.E. Raftery, 1994. Model selection and accounting for model uncertainty in graphical models using Occam's window. J. Am. Statistical Associat., 89: 1535-1546. DOI: 10.1080/01621459.1994.10476894.

Marin, J.M., K. Mengersen and C.P. Robert, 2005. Bayesian modelling and inference on mixtures of distributions. Handbook Statistics, 25: 459-507. DOI: 10.1016/SO169-7161(05)25016-2

McLachlan, G.J. and K.E. Basford, 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York, ISBN-10: 0824776917, pp: 253.

Mengersen, K., 2009. Module 1 Bayesian analysis, short course on Bayesian modeling. Statistics Department, ITS, Surabaya.

Montgomery, J. and B. Nyhan, 2010. Bayesian model averaging: Theoretical developments and practical applications. Political Analysis, 18: 245-270. DOI: 10.1093/pan/mpq001.

Muller, P., G. Parmigiani, C. Robert and J. Rousseau, 2004. Optimal sample size for multiple testing: The case of gene expression microarrays. J. Am. Statistical Associat., 99: 990-1001. DOI: 10.1198/016214504000001646.

Richardson, S. and P.J. Green, 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). J. Royal Statistical Society, 59: 731-792. DOI: 10.1111/1467-9868.00095

Sebastiani, P., H. Xie and M.F. Ramoni 2006. Bayesian analysis of comparative microarray experiments by model averaging. Int. Society Bayesian Analysis, 1: 707-732.

Wati, D.M., 2006. Analysis of Bayesian Markov chain Monte Carlo in normal mixture modeling with unknown number of components (Case Study: Rate of Inflation in Indonesia). MSc Thesis, ITS, Surabaya, Indonesian.

Wholey, W.Y.C., 2012. Hypochlorous acid stress responses in bacteria. PhD Thesis, Michigan University, USA.

Yuan-Ying, Z., X. Deng-Ke, D. Xing-De and D. Liang, 2014. Bayesian estimation of simplex distribution nonlinear mixed models for longitudinal data. Int. J. Applied Math. Statistics, 52: 1-10.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. 1st Edn., John Wiley, New York, ISBN-10: 0471981656, pp: 431.