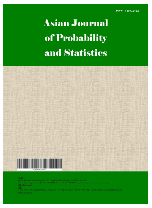

Asian Journal of Probability and Statistics

10(4): 60-71, 2020; Article no.AJPAS.64432
ISSN: 2582-0230

Do Prior Type and Sample Size have Effect on Mixtures of Normal? The Monte Carlo evidence

Ojo O. Oluwadare[1*], Owonipa R. Oluremi[2] and Enesi O. Lateifat[2]

[1]Department of Statistics, Federal University of Technology Akure, Nigeria.
[2]Department of Mathematical Sciences, Kogi State University, Anyigba, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author OOO conceived the presented idea, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors ORO and EOL managed the analyses of the study. Author OOO managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2020/v10i430255
Editor(s):
(1) Dr. Belkacem Chaouchi, Khemis Miliana University, Algeria.
Reviewers:
(1) Eric David Cohen, State University of Campinas, Brazil.
(2) Kamil Alakuş, Ondokuz Mayıs University, Turkey.
(3) Sin Jose Hejase, Al Maaref University (MU), Lebanon.
Complete Peer review History: http://www.sdiarticle4.com/review-history/64432

Received: 15 November 2020
Accepted: 22 January 2021
Published: 06 February 2021

Original Research Article

Abstract

Overtime finite mixtures of Normal in regression have gained popularity and also shown to be useful in modelling heterogeneous data. This study examines the effects of prior and sample size in regression mixtures of Normal models with Bayesian approach. Monte Carlo experiment was carried out on the Normal mixtures model in order to examine the strength of priors and also to know the suitable sample size to produce stable results. Results obtained from the experiment indicate that an informative prior gives a reliable estimate than non-informative prior while large sample sizes maybe needed to obtain stable results.

Keywords: Bayesian; montecarlo; normal models; prior.

1 Introduction

In regression models, assumptions about the functional forms and distributions can be made. However, real life situations especially with the use of economic theory do not tells us the functional forms and

**Corresponding author: E-mail: ojooo@futa.edu.ng;*

distributions of an economic model. There are many techniques that have been developed to overcome this problem; one of such techniques is by mixing together different distributions also termed as mixture of models. There are numerous literature on mixture of models [1-5].

A typical example of mixture models is called normal models. Normal model is a kind of model that allows whole distributions to have unknown forms. It also has a flexible distribution that can be obtained by mixing several distributions together while the resulting flexible distribution can in turn be used for approximation of unknown distribution of interest The major advantage of using mixtures of Normal distribution is the flexibility in modelling strategy and also simple to work with [6]. Mixture of Normal distribution can also accommodate skewness and multimodality in error distribution [7].

Apart from Normal mixtures model, other notable examples of non-normal mixtures have been largely demonstrated in the literature. Lee and McLachlan [8] used a multivariate skew t-distribution, Andrews and McNicholas [9] utilized a multivariate t-distribution, Weibull distribution by Sultan et al. [10], Lakshmi and Vaidyanathan [11] employed Gamma distribution and skew-normal distribution by Zellner et al. [12] among others. However, due to computational convenience and applicability to other methods such as Bayesian and Maximum Likelihood (ML) methods in the estimation of model parameters, Normal mixtures model has been found useful in many applied works than non-normals distribution [13].

Prior distribution plays a major role in Bayesian modelling. It reflects the information about an uncertain parameter that is combined with new data to yield a posterior distribution. However, wrong choice of prior can lead to incorrect inferences and decisions [14]. The major challenge in the estimation of mixture models is the incorporation of right prior in the model being considered for estimation. Just as the prior distribution plays a key role in Bayesian inference, sample size must also be taken into consideration. If the sample size is small or available data gives indirect information about the parameters of interest, the prior distribution will also become more relevant.

This present work examines the effects of sample size and prior type on regression mixtures model. It will help to determine the kind of prior for the estimation of Normal mixtures while suitable sample size will be determined with the aid of Monte Carlo study. The remainder of the paper is arranged as follows. In section 2, we briefly give the regression model and Bayesian estimation procedures involving a mixture of Normals. For comparative purposes, the performance of the priors across different sample sizes with the aid of numerical studies is provided in Section 3. Section 4 provides results of analyses from Monte Carlo experiment. Section 5 concludes.

2 Materials and Methods

2.1 Bayesian inference based on Normal mixture model

Consider a linear regression model given as:

$$y = x\theta + \varepsilon \tag{1}$$

Where,

$$y = (y_1, \ldots, y_N)', \quad \theta = (\theta_1, \ldots, \theta_k)'$$

$$x = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N2} & \cdots & x_{NK} \end{pmatrix}$$

and

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$

If the assumption where ε is iid N(0, h^{-1}) for $i = 1, \ldots, N$ is replaced by letting ε_i to have a mixture of different distributions, we have:

$$\varepsilon = \sum_{j=1}^{m} e_{ij}(\alpha_j + h_j^{1/2}\, \eta_{ij}) \tag{2}$$

where,

η_{ij} is iid N(0,1), for $i = 1, \ldots, N$ and $j = 1, \ldots, m$, it means $(\alpha_j + h_j^{-1/2}\, \eta_{ij})$ is also normal random variable with mean and precision of α_j and h_j, respectively.

α_j, h_j and e_{ij} are parameters to be estimated and $\sum_{j=1}^{m} e_{ij} = 1$.

But e_{ij} shows the components in mixture where ith error can be drawn. Hence, e_{ij} can be zero or 1, for $= 1, \ldots, m$, since it is difficult to know when the ith error is taken from, we let P_j for $j = 1, \ldots, m$ be the probabilities of error being drawn.

Therefore, equation (2) denotes that the regression error component is a weighted average of m with different density functions. Furthermore, we can stack the parameters α, h, e_i, and e as:

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}, \quad h = \begin{pmatrix} h_1 \\ \vdots \\ h_m \end{pmatrix}, \quad e_i = \begin{pmatrix} e_{i1} \\ \vdots \\ e_{im} \end{pmatrix}, \text{ and } e = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}$$

Thus, e_i is drawn from a multinomial density function and is defined as:

$e_i \sim M\ (1, P)$

Where,

$$P = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix}$$

Therefore, the likelihood function is given as:

$$\mathrm{P}(y|\ \theta, h, \alpha, P) = \mathrm{v}\ \frac{1}{(2\pi)^{N/2}} \prod_{i=1}^{N} \{\sum_{j=1}^{m} p_j \sqrt{h_j}\ \exp\ [\frac{-h_j}{2}\ (y_i - \alpha_j - \theta' x_i)^2]\} \tag{3}$$

The prior for this study is the one that will allow for computation conveniences. Hence, we will use a Dirichlet distribution and is given by:

$$P \sim D\ (P_o), \tag{4}$$

Where,

$0 \le p_j \le 1$ and $\sum_{j=1}^{m} P_j = 1$

Because of identification problem of the model as obtained in (literature), we place the following restrictions on the prior as:

$h_{j-1} < h_j$ (5)

$\alpha_{j-1} < \alpha_j$ (6)

and

$p_{j-1} < p_j$ (7)

Where, $j = 2, \dots, m$

A prior has to be assumed for restriction parameter in (6) in order for the α to be normal [6]. Therefore, we have:

$$\mathrm{P}(\alpha) \propto f_N(\alpha|\alpha_o, Q_o).\, I\,(\alpha_1, < \alpha_2 < \dots < \alpha_m) \quad (8)$$

Where, $I(.)$ is an indicator function and is given as:

$$I(A) = \begin{cases} 1, & if\ A\ holds \\ 0 & otherwise \end{cases}$$

We also assume an independent Normal-Gamma prior using the following:

$$\theta \sim N(\theta^o, Q^o) \quad (9)$$

$$h_j \sim G(s_j^{o\ \ -2}, v_j^o) \quad (10)$$

In order to obtain the posterior distribution we treat e as a latent data, the likelihood can then be written as:

$$\mathrm{P}(y|e, \theta, h, \alpha, P) = \frac{1}{(2\pi)^{N/2}} \prod_{i=1}^{N} \{\sum_{j=1}^{m} e_{ij} \sqrt{h_j}\ \exp\,[\frac{-h_j}{2}\ (y_i - \alpha_j - \theta' x_i)^2]\} \quad (11)$$

Multiplying (11) and (4) to obtain marginal distribution of P, we have

$$P \sim D\ (P^*) \quad (12)$$

Where,

$$P^* = P_o + \sum_{i=1}^{N} e_i \quad (13)$$

Again, multiplying (11) and (8), we have:

$$\mathrm{P}(\alpha|y, e, \theta, h) \propto f_N(\alpha|\alpha^*, Q_\alpha^*).\, I\,(\alpha_1, < \alpha_2 < \dots < \alpha_m) \quad (14)$$

Where,

$$Q_\alpha^* = \frac{1}{\left[Q_{(o)\ \alpha}^{-1} + \sum_{i=1}^{N} (\sum_{j=1}^{m} e_{ij} h_i) e_i e_i'\right]}$$

$$\alpha^* = Q_\alpha^*\ \ [Q_{(o)\alpha}^{-1}\ \alpha_o + \ \sum_{i=1}^{N} (\sum_{j=1}^{m} e_{ij} h_i) e_i\, (y_i - \theta' x_i)] \quad (15)$$

Lastly, to obtain the conditional posterior for both θ and h, we combine (11) with (9) and (10), we have:

$$\theta \mid y, e, h, \alpha \sim N(\theta^{*}, Q^{*}) \quad (16)$$

$$h_j \mid y, e, \theta, \alpha \sim G(s *_j^{-2}, v_j^{*}) \quad (17)$$

where

$$Q^{*}= \frac{1}{[\frac{1}{Q^{o}} + \sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij} h_j x_i x_i^{'}]}$$

$$\theta^{*}= Q^{*}\ [\frac{1}{Q^{o}}\ \theta^{o} + \sum_{i=1}^{n}\sum_{j=1}^{m} e_{ij} h_j x_i (y_i - \ \alpha_j)]$$

$$v_j^{*} = \sum_{i=1}^{n} e_{ij} + \ v_j^{o}$$

$$S_j^{*\ 2} = \frac{\sum_{i=1}^{n} e_{ij}\, (y_i - \alpha_j - x_i^{'}\ \theta)^{'}\ (y_i - \alpha_j - x_i^{'}\ \theta) + v_j^{o}\, S_j^{o\ 2}}{v_j^{*}}$$

Thus, (13), (15), (17) are then used to carry out posterior inference for desired different estimates.

2.2 Model comparison

In this study, three popular information criteria will be used to select best component among the competing components in the mixtures. These information criteria have been found to be easy in calculations and also do not rely on prior information [15]. These information criteria are; Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn (HQ).

Akaike Information Criterion (AIC): This was proposed by Akaike [16]. It is defined as:

$$\text{AIC} = 2\ln\,[P(y|)] - 2p \quad (18)$$

A model with a smallest AIC is chosen to be the best. In simulation study, it has been observed that method of AIC is upwardly biased especially with respect to class enumeration mixture models (Van Horn et al., 2009). This makes most researchers to use BIC and HQ.

Bayesian Information Criterion (BIC): It is the most popular information criterion and does not rely on specific sampling distributions but the observed data. It takes into account of sample size, likelihood function, and number of parameters. BIC was proposed by Schwarz [17] and is simply defined as:

$$\text{BIC} = 2\ln[P(y|)] - P\ln(N) \quad (19)$$

A model with a smallest BIC is chosen to be the best.

Hannan-Quinn (HQ): It is an alternative to both AIC and BIC. This was introduced by Hannan and Quinn [18]. It is given as:

$$\text{HC} = L_{max} + 2k\ \ln\,(\ln(N)) \quad (20)$$

where, L_{max} is the log-likelihood, k is the number of parameters, and N is the number of sample size.

3 Monte Carlo Experiment

In order to examine the effects of sample size and prior kind in Normal mixture model, we present a Monte Carlo experiment in this section. Thus, the model for the experiment is given as:

$y_i = \varepsilon_i$ (21)

Where, ε_i is the mixture of Normal distributions

y_i is the dependent variables for different models

The sample sizes for the Normal mixtures model are set as: N= 10, 25,100, and 10000

We will use Gibbs sampler, a typical example of Markov Chain Monte Carlo (MCMC) for Posterior simulation [19, 6] for more details). This Gibbs sampler has advantages due to its flexibility in evaluation of conditional distributions. An initial burn-in-period of R_0 = 1000 will be discarded and include R_1 = 10,000 replications in the experiment.

Bayesian inference will be done using m =3 and the equations of the model are given as:

Equation 1: First Normal

α_1= -2, h_1 =16, p_1= 0.65

Equation 2: Second Normal

α_2= 2, $h_2 = 16$, p_2= 0.25

Equation 3: Third Normal

α_3= 0, $h_3 = 4$, p_3=0.10

Informative prior hyper parameters:

$\alpha^o = 1$, Q^o= $(10000)^2 \; I_m$, $S_j^{o\;-2}$ = 1, v_j= 0.1, and $p^o = l_m$

Where, l_m is an m-vector of ones

We create a non-informative prior by setting the hyper-parameters as follows:

$\alpha^o = 0_m$, Q^o= I_m, $S_j^{o\;-2}$ = 0, v_j= 0, and $p^o = l_m$

It is necessary to assess the convergence of MCMC simulation. This is done by calculating the Convergence Diagnostic (CD) test statistics [20, 21, 22, 6] for more details). The primary aim of CD is to test the equality of the means of the first and latter part of a Markov chain. It has a property of asymptotically standard normal distribution. Hence, the CD statistics based on Geweke [19] is given by:

$$\mathrm{CD} = \frac{\hat{g}_{R_A} - \hat{g}_{R_C}}{\sqrt{\frac{\hat{\sigma}_A}{R_A} + \frac{\hat{\sigma}_C}{R_C}}} \quad (22)$$

Where, $\hat{\sigma}_A$ and $\hat{\sigma}_C$ are variance of first and last set draws

$\hat{g}_{R_A}$ and $\hat{g}_{R_C}$ are estimates of mean of first and last set of the draws

R_A and R_C are number of first and last draws

Decision rule: If CD is less than |1.96| for all the parameters considered, it indicates that convergence of the MCMC algorithms has been achieved.

Another MCMC diagnostic is the Numerical Standard Error (NSE); this measures the approximation due to error. It is computed as:

$$\text{NSE} = \frac{\hat{\sigma}}{\sqrt{R}} \tag{23}$$

Where R is the number of replication

$\hat{\sigma}$ is standard error

4 Results and Discussion

The results of Monte Carlo are reported in this section. Tables 1, 3, 5, and 7 contain information criteria for the simulated data while summary of the parameter estimates for the selected model for each simulated data were presented in Tables 2, 4, 6, and 8. The summary of the parameter estimates entails posterior means & standard deviation and MCMC results; NSE and Geweke's convergence diagnostic test. The information criteria for model comparison are AIC, BIC, and HQ. Figs. 1-4 show the histogram for mixtures of Normal models for different samples with both informative and non-informative priors in appendix. The primary aim of the figures is to show that all mixtures of Normals are flexible.

In Tables 1 and 3, model m =2 is preferred for all the information criteria presented using informative prior but the values for non-informative prior were not available. All the information criteria used select the correct value of m =3 as shown in Tables 5 and 7. These show that information criteria are useful for selecting the number of components in Normal mixtures. Non-informative prior does not yields any value for AIC, BIC, and HQ for small sample sizes (10 and 25), this is due to small data sets and lack of vital information used in the analysis.

It can be seen that the estimation results (mean and standard deviation) for the parameters appear quite reasonable. For example, the true model is estimated with reasonably accurate results for both informative and non-informative priors (see Tables 2, 4, 6, and 8) except for small sample sizes (10 and 25) in the case of non-informative. Also, the Posterior means obtained for both informative and non-informative priors are closer to their true values.

Results from Tables 2, 4, 6, and 8 show that convergence of our algorithms has been achieved with the use of Geweke's CD for both informative and non-informative priors across the sample sizes except for small sample sizes (10 and 25) in the case of non-informative prior. The NSE revealed in Tables 2 and 4 for informative prior indicate that we are achieving reasonably precise estimates. However, Tables 6 and 8 show that all the estimates are more accurate for both informative and non-informative priors when sample sizes are 100 and 500.

Table 1. Results of Information criteria for model comparison when the sample size is 10

	Informative prior			**Non-informative prior**		
Model	**AIC**	**BIC**	**HQ**	**AIC**	**BIC**	**HQ**
m =1	-28.4896	-29.3973	-29.9959	NaN	NaN	NaN
m =2	-2.1991	-4.0146	-5.2116	NaN	NaN	NaN
m =3	-9.8260	-12.5493	-14.3449	NaN	NaN	NaN

Table 2. Parameter estimates for Mixtures of Normals when the sample size is 10

	Informative prior				Non-informative prior			
	Mean	Standard deviation	Geweke CD	NSE	Mean	Standard deviation	Geweke CD	NSE
α_1	1.9	0.1	-0.8	0.0025	NaN	NaN	NaN	NaN
α_2	2.1	0.3	-1	30.9197	NaN	NaN	NaN	NaN
α_3	787.11	5998	-0.3	59.7089	NaN	NaN	NaN	NaN
h_1	15.0	8.7	0.3	0.00023	NaN	NaN	NaN	NaN
h_2	15.7	15.2	-0.2	0.0433	NaN	NaN	NaN	NaN
h_3	1	4.8	-0.5	0.0455	NaN	NaN	NaN	NaN
p_1	0.6	0.1	-1.5	0.0007	0.3343	0.2364	-0.2096	0.0021
p_2	0.3	0.1	1.1	0.0005	0.3327	0.2346	-0.0860	0.0015
p_3	0.1	0.1	1	0.0006	0.3330	0.2344	0.3390	0.0025

Table 3. Results of information criteria for model comparison when the sample size is 25

	Informative prior			Non-informative prior		
Model	AIC	BIC	HQ	AIC	BIC	HQ
m =1	-48.9656	-52.6222	-53.4868	NaN	NaN	NaN
m =2	-19.3209	-26.6342	-28.3635	NaN	NaN	NaN
m =3	-27.1439	-38.1137	-40.7077	NaN	NaN	NaN

Table 4. Parameter estimates for mixtures of normals when the sample size is 25

	Informative prior				Non-informative prior			
	Mean	Standard deviation	Geweke CD	NSE	Mean	Standard deviation	Geweke CD	NSE
α_1	-2	0.1	-0.8	0.0008	NaN	NaN	NaN	NaN
α_2	0.6	0.7	0.6	0.0072	NaN	NaN	NaN	NaN
α_3	7890	5962.2	-0.4	50.3148	NaN	NaN	NaN	NaN
h_1	13.6	5	-0.7	0.0525	NaN	NaN	NaN	NaN
h_2	0.6	0.4	-0.3	0.0029	NaN	NaN	NaN	NaN
h_3	1	4.5	0.8	0.0334	NaN	NaN	NaN	NaN
p_1	0.7	0.1	-0.1	0.0006	0.3360	0.2376	-0.6283	0.0022
p_2	0.3	0.1	-0.3	0.0006	0.3344	0.2386	0.8530	0.0015
p_3	0	0	1.2	0.0003	0.3296	0.2352	-0.1148	0.0119

Table 5. Results of information criteria for model comparison when the sample size is 100

	Informative prior			Non-informative prior		
Model	AIC	BIC	HQ	AIC	BIC	HQ
m =1	-219.6922	-227.5077	-227.4368	-216.5908	-224.4063	-224.3354
m =2	-12.5749	-28.2059	-28.0641	-66.0433	-81.6744	-81.5326
m =3	-23.3656	-0.0809	0.1318	-22.2968	-45.7433	-45.5306

Table 6. Parameter estimates for Mixtures of Normals when the sample size is 100

	Informative prior				Non-informative prior			
	Mean	Standard deviation	Geweke CD	NSE	Mean	Standard deviation	Geweke CD	NSE
α_1	-2.0344	0.0242	-0.5044	0.0001	-1.9732	0.0282	0.3717	0.0001
α_2	-0.0093	0.2066	-1.8023	0.0029	0.0296	0.1375	-0.6136	0.0019
α_3	1.9819	0.0547	-0.3113	0.0004	1.9842	0.0474	-0.2296	0.0005
h_1	26.8084	4.7438	-0.8602	0.0499	22.5401	4.4384	0.5951	0.0487
h_2	4.4379	2.8282	0.5376	0.0301	4.7663	2.1953	1.3604	0.0378
h_3	14.8194	4.4074	0.4357	0.0529	18.9650	5.6422	-0.5589	0.0793
p_1	0.6492	0.0472	-0.5574	0.0004	0.5712	0.0491	0.0175	0.0003
p_2	0.0910	0.0300	0.6091	0.0004	0.1591	0.0367	0.3505	0.0003
p_3	0.2597	0.0434	0.1779	0.0005	0.2698	0.0438	-0.3364	0.0003

Table 7. Results of Information criteria for model comparison when the sample size is 500

	Informative prior			Non-informative prior		
Model	AIC	BIC	HQ	AIC	BIC	HQ
m =1	-1036	-1048.6	-1046.46	-1036	-1048.6	-1046.4
m =2	-172.3097	-197.5974	-193.1940	-172.3120	-197.5997	-193.1963
m =3	6.5322	-31.3992	-24.7941	6.5681	-31.3633	-24.7582

Table 8. Parameter estimates for Mixtures of Normals when the sample size is 500

	Informative prior				Non-informative prior			
	Mean	Standard deviation	Geweke CD	NSE	Mean	Standard deviation	Geweke CD	NSE
α_1	-2.0099	0.0130	1.2286	0.0001	-2.0096	0.0129	1.2619	0.0001
α_2	0.0067	0.0705	0.3555	0.0004	0.0068	0.0701	0.2912	0.0004
α_3	1.9613	0.0232	-0.3974	0.0002	1.9602	0.0230	-0.4129	0.0002
h_1	18.3615	1.4447	1.3921	0.0155	18.4600	1.4539	0.3272	0.0153
h_2	3.9135	0.9566	0.0848	0.0128	3.9410	0.9663	0.1282	0.0141
h_3	16.5943	2.2060	0.3556	0.0169	16.8236	2.2405	0.3255	0.0171
p_1	0.6594	0.0209	0.5661	0.0003	0.6594	0.0209	0.5801	0.0003
p_2	0.1068	0.0140	0.8358	0.0002	0.1068	0.0140	0.7070	0.0002
p_3	0.2338	0.0186	-1.3586	0.0002	0.2338	0.0186	-1.2717	0.0002

5 Conclusion

The importance of mixtures of Normal in modelling of fat-tailed and multi-modal distributions has necessitated its investigation. Inappropriate choices of prior and sample sizes can leads into wrong inference in applied research works especially in regression mixture modelling. In this paper, Bayesian inference was carried out on regression with Normal mixtures using finite mixture modelling to know the kind of prior and sample sizes that are suitable for posterior inference. It was deduced that all the information criteria were consistent with one another and gave the same conclusive results for model comparison of Normal models. The parameter estimates of the posterior were also reasonable especially for informative prior while a great variability was demonstrated with large sample sizes (100 and 500). The numerical standard errors are accurate for large sample sizes. However, with the use of small samples gave imprecise results and these can lead to wrong inference.

It is apparent from the results that the choice of prior plays a major role in regression mixture of Normal models with Bayesian approach. Informative prior performed well in both large and small sample sizes for model comparison of Normal models and parameter estimates. Thus, the researcher is advised to choose an informative prior and large sample when dealing with Normal mixtures of models to obtain stable results. There are so many directions for further research. We assume standard linear model for regression mixtures of Normal model characterized by correlated predictors and a situation when sample sizes are both balanced and unbalanced, We plan to further the research in that regards in our future paper.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Cohen A, Amin R. The effects of normal mixtures and autocorrelation on the fraction non-conforming. Communications in Statistics: Simulation and computation. 2016;46:10, 8105-8117.
[2] Lamont AE, Vermunt JK, van Horn ML. Regression Mixture models: Does modelling the covariance between independent variables and latent classes improve the results? Multivariate Behavioural Research. 2016;51(1):35-52.
[3] McLachlan G, Peel D. Finite mixture models. New York, NY: Wiley; 2000.
[4] Van Horn ML, Smith J, Fagan AA, Jaki T, Feaster DJ, Masyn K, Howe G. Not quite normal: Consequences of violating the assumption of normality in regression mixture models. Structural Equation Modelling: A Multidisciplinary Journal. 2012;19u:227-249.
[5] Kim M, Lamont AE, Jaki T, Feaster D, Howe G, Van Horn ML. Impact of an equality constraint on the class-specific residual variances in regression mixtures: AA Monte Carlo simulation study. Behavioural Research Methods. 2016;48(2):813-826.
[6] Koop G. Bayesian Econometrics. Wiley; 2003.
[7] Koop G, Poirier DJ, Tobias JL. Bayesian Econometric Methods. Cambridge; 2005.
[8] Lee S, McLachlan GJ. Finite mixtures of multivariate skew t-distributions: Some recent and new results. Stat. Comput. 2014;24(2):181-202.
[9] Andrews JL, McNicholas PD. Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions. Stat. Comput. 2012;22(5):1021-1029.
[10] Sultan KS, Ismail MA, Al-Moisheer AS. Mixture of two inverse Weibull distributions: Properties and estimation. Comput. Stat. Data. Anal. 2007;51(11):5377-5387.
[11] Lakshmi RV, Vaidyanathan VS. Parameter estimation in Gamma mixture model using Normal –based approximation. Journal of Statistical Theory and Applications. 2015;15(1):25-35.
[12] Zellner CB, Cabral CR, Lachos VH. Robust mixture regression modelling based on scale mixtures of skew-normal distributions. Mathematics????; 2016.
[13] Zhang H, Huang Y. Finite mixture models and their applications: A review. Austin Biom and Biostat. 2015;2(1).
[14] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis, Chapman & Hall, London; 1995.
[15] Lo Y, Mendel NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika. 2001;88:767-778.
[16] Akaike H. Infromation theory and an extension of the maximum likelihood principle. In Petror BN, Csaki F. International ymposium on Inference theory, Akademia Kiado, Budapest. 1973;267-281
[17] Schwarz GE. Estimating the dimension of a model. Annals of Statistics. 1978;6(2):461-464.
[18] Hannan EJ, Quinn BG. The determination of the order of an Autoregression. JRSS. B41. 1979;190-195.
[19] Geweke J. Using Simulation methods for Bayesian Econometric Models: Inference, development, and Communication. Econometric Reviews. 1999;18:1-126.
[20] Gelman A. Inference and monitoring convergence, in Gilks, Richardson and Speigelhalter. Missing; 1996.
[21] Raftery AE, Lewis SM. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. Statistical Science. 1992;7:493–497.
[22] Geweke J. Evaluating the accuracy of sampling-based approaches o calculating posterior moments. In: Bayesian Statistics. Oxfors Clarendon press. 1992;4:169-193.

Appendix

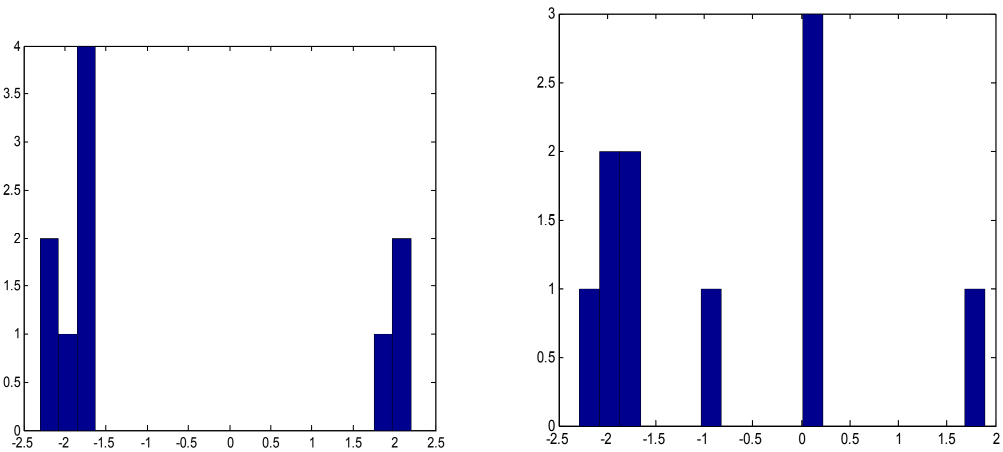

Fig. 1. Histogram for mixtures of Normals when sample size 10 with informative prior

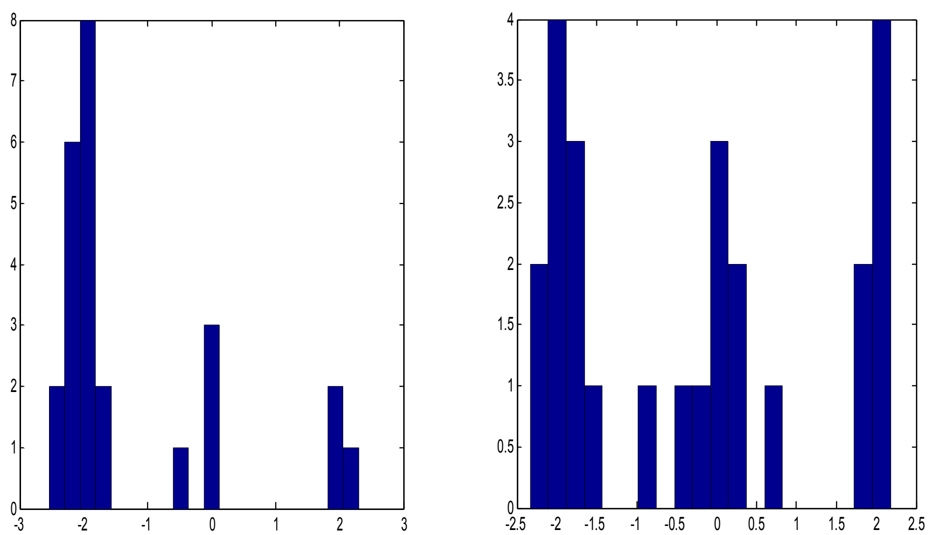

Fig. 2. Histogram for mixtures of Normals when sample size 25 with informative prior

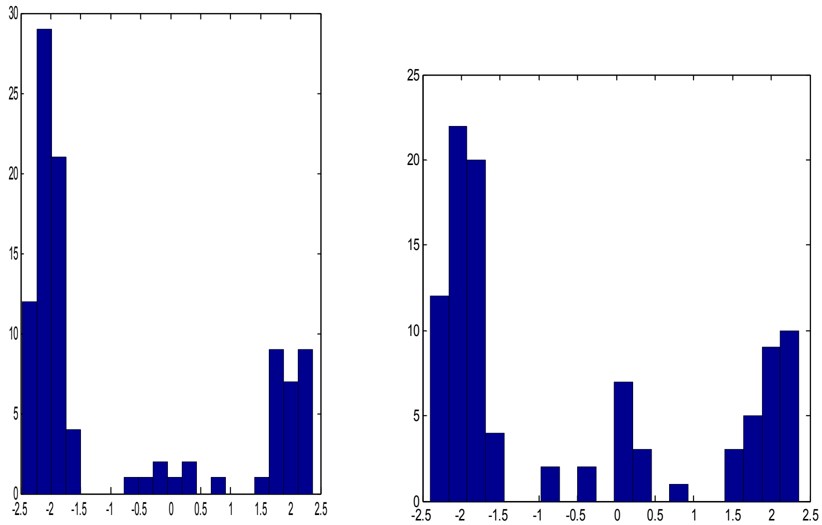

Fig. 3. Histogram for mixtures of Normals when sample size 100 with informative prior

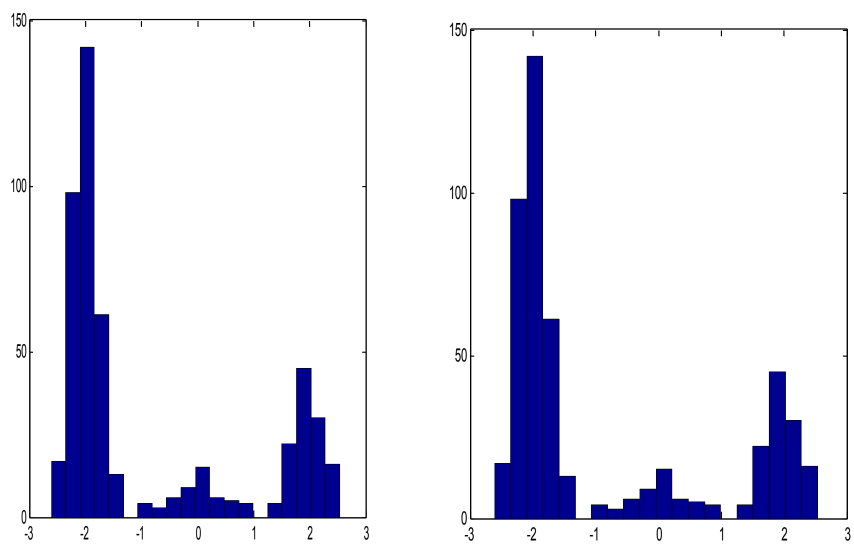

Fig. 4. Histogram for mixtures of Normals when sample size 500 with informative prior

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/64432