Sains Malaysiana 48(1)(2019): 227–235
http://dx.doi.org/10.17576/jsm-2019-4801-26

Bayesian Inference for Linear Regression under Alpha-Skew-Normal Prior

(Pentaabiran Bayesian untuk Model Regresi Linear Prior Normal-Pencong-Alfa)

ALHAMIDE, A.A., KAMARULZAMAN IBRAHIM, ALODAT, M.T. & WAN ZAWIAH WAN ZIN*

ABSTRACT

A study on Bayesian inference for the linear regression model is carried out in the case when the prior distribution for the regression parameters is assumed to follow the alpha-skew-normal distribution. The posterior distribution and its associated full conditional distributions are derived. Then, the Bayesian point estimates and credible intervals for the regression parameters are determined based on a simulation study using the Markov chain Monte Carlo method. The parameter estimates and intervals obtained are compared with their counterparts when the prior distributions are assumed either normal or non-informative. In addition, the findings are applied to Scottish hills races data. It appears that when the data are skewed, the alpha-skew-normal prior contributes to a more precise estimate of the regression parameters as opposed to the other two priors.

Keywords: Alpha skew normal distribution; Bayesian linear regression model; simulation

ABSTRAK

Suatu kajian tentang pentaabiran Bayesian untuk model regresi dijalankan untuk kes taburan prior bagi parameter regresi yang diandaikan mengikuti taburan normal-pencong-alfa. Taburan posterior dan taburan bersyarat penuh yang berkaitan diterbitkan. Seterusnya, anggaran titik dan selang boleh percaya Bayesian ditentukan berdasarkan satu kajian simulasi menggunakan kaedah rantai Markov Monte Carlo. Anggaran titik dan selang yang diperoleh dibandingkan dengan keputusan apabila taburan diandaikan normal dan tak bermaklumat. Di samping itu, penemuan ini digunakan untuk data perlumbaan bukit Scottish. Kajian ini mendapati bahawa dalam kes data pencong, penganggaran parameter adalah lebih tepat apabila prior normal-pencong-alfa diandaikan berbanding prior normal dan tak bermaklumat.

Kata kunci: Model regresi linear Bayesian; simulasi; taburan normal-pencong-alfa

INTRODUCTION

There had been a growing interest on the study of non-Gaussian parametric distributions since the last twenty years. This was in response due to the increase on the demand of analyzing datasets which exhibit skewness or heavy tail distributions. Several statistical distribution and techniques have shed the light on how to treat data sets which show either skewness, heavy tail or both. A review of these techniques and distributions can be found in several articles such as Arellano-Valle and Azzalini (2006), Azzalini and Capitanio (2014) and Genton (2004). Furthermore, a large class of such distributions which is called the skew-symmetric (SS) distributions has been discussed by Alodat et al. (2014) and Wang et al. (2004). These new families of skew distributions have been used for re-studying a large number of statistical methodologies under the new setups. For example, the problems of inference and prediction for linear and non-linear regression models have been considered under the assumption that the error term has a skew distribution (Arellano-Valle et al. 2005). Also, some authors such as Alodat and Al-Momani (2014) have considered the regression problem when both error term and regression coefficients have skew distributions. Alhamide et al. (2016) has considered the linear regression modeling when the error term is assumed to follow the extended skew distribution. In all the cases, the analysis of the regression under the skew setups shows an improvement in terms of the accuracy measures such Akaike Information criterion (AIC), bias and mean-squared errors.

Although the regression model has been treated under several skew normal families, from a Bayesian point of view, there is still much work which could be added to the literature in this field. This is due to some new findings in the area of the skew normal distributions. One of the recent findings in the skew normal distributions is the so called alpha-skew normal distribution which was introduced by Elal-Olivero (2010). Let YY be a random variable which follows an alpha skew normal distribution given by,

$$f(y;\alpha)=\frac{1+(1-\alpha y)^2}{2+\alpha^2}\varphi(y),\quad y\in\mathbb{R},\alpha\in\mathbb{R} \tag{1}$$

where ϕ is the probability density function (pdf) of a standard normal random variable. This pdf can be denoted by $ASN(\alpha)$. The parameter α reflects the properties of asymmetricity and unibimodality of the distribution. Furthermore, the distribution reduces to the normal distribution when $\alpha = 0$. In order to give the above distribution a higher ability to capture the features of data, we define the following location- scale version of (1):

$$f(y;\alpha;\mu;\sigma) = \frac{1+(1-\frac{\alpha}{\sigma}(y-\mu))^2}{(2+\alpha^2)\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right),$$

$$y \in \mathbb{R}, \alpha \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0 \qquad (2)$$

Thus (2) can be denoted by $ASN(\alpha,\mu,\sigma)$. The aim beyond the above definition is to define a new family of skew distributions that have enough flexibility to fit both unimodal and bimodal shape.

Considering the $ASN(\alpha,\mu,\sigma)$ as a choice of the prior distribution is reasonable as opposed to the other form of skew normal distributions as given by Azzalini and Capitanio (2014) due to its flexibility in allowing for the skewness property in the data and also, according to Elal-Olivero (2010), the later distribution has less theoretical tractability because it has the pdf which includes the CDF of the normal distribution that has no closed form. This paper is concerned with the study of the statistical inference by using the Bayesian approach for the linear regression models under alpha skew normal prior distribution (LR-ASN). This paper also includes the simulation results on the estimation of the parameters of the LR-ASN as compared to the linear regression models under normal prior distribution (LR-ND) and non-informative prior (LR-NI). Based on the comparison, we demonstrate the value of two informative priors, which are the normal prior and the alpha skew normal prior, and a non-informative prior for depicting our lack of knowledge on the parameters, in the Bayesian linear regression modeling of skew data. The remainder of this paper is organized as follows: first, we present the linear regression model under alpha skew normal prior. Next, we introduce the linear regression model under normal prior distribution and non-informative prior. This is followed by a simulation study for computing the Bayes estimates and credible intervals for the parameters under ASN prior and then comparison of results found based on normal and non-informative priors. Finally, we apply the findings to the Scottish hills races data and summarized the results found.

THE MODEL

In this section, we consider the linear regression model:

$$y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, n, \qquad (3)$$

where y_i are responses, $\boldsymbol{\beta}^T = (\beta_0, \ldots, \beta_k)$ and $\boldsymbol{X}_i^T = (1, X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ik})$ is a vector for the values of explanatory variables and the random errors $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. In order to analyze the above regression model using Bayesian methods involving $ASN(\alpha, \mu, \sigma^2)$ prior, we adopt a prior distribution with the following hierarchical representation:

$$\beta_0, \beta_1, \ldots, \beta_k|\alpha, \mu, \sigma^2 \sim ASN(\alpha, \mu, \sigma^2)$$

$$\alpha|a, \sigma^2 \sim N(a, \sigma^2)$$

$$\mu|b, \sigma^2 \sim N(b, \sigma^2)$$

$$\sigma^2|c, d \sim IG(c, d),$$

where a, b, c and d are the hyperparameters. It is known that the joint pdf of $\boldsymbol{y}$ is:

$$f(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\beta},\mu;\alpha;\sigma) = (2\pi)^{\frac{-n}{2}}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right).$$

Let $\boldsymbol{\Theta} = (\boldsymbol{\beta}^T, \mu, \alpha, \sigma^2)$ and $\pi(\boldsymbol{\Theta})$ denotes the joint prior distribution. Then

$$\pi(\boldsymbol{\Theta}) = \pi(\boldsymbol{\beta}|\alpha,\mu,\sigma^2) \times \pi(\alpha|a, \sigma^2) \times \pi(\mu|b, \sigma^2) \times \pi(\sigma^2|c,d).$$

Therefore, the posterior distribution of $\boldsymbol{\Theta}$ given $\boldsymbol{y}$ and $\boldsymbol{X}$ is

$$\pi(\boldsymbol{\Theta}|\boldsymbol{y}, \boldsymbol{X}) \propto \frac{1}{\sigma^n}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right)$$

$$\times\left[\prod_{i=0}^{k}\frac{\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu)\right)^2}{(2+\alpha^2)\sigma^3}\exp\left(-\frac{1}{2\sigma^2}(\beta_i-\mu)^2\right)\right]$$

$$\times\frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(\alpha-a)^2\right)\frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(\mu-b)^2\right)$$

$$\frac{1}{(\sigma^2)^{c+1}}\exp\left(-\frac{1}{d\sigma^2}\right).$$

In order to make the inferences about $\boldsymbol{\Theta}$, we draw a random sample from $\pi(\boldsymbol{\Theta}|\boldsymbol{y}, \boldsymbol{X})$, and implement the Metropolis-Hastings algorithm by using R (Robert & Casella 2010). When implementing the algorithm, the normal distribution is considered as the proposal distribution. This requires finding the full conditional distributions corresponding to $\pi(\boldsymbol{\Theta}|\boldsymbol{y}, \boldsymbol{X})$ as the following:

(i) The distribution of $\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \alpha, \mu, \sigma^2$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \alpha, \mu, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right)$$

$$\times\prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\mu})^T(\boldsymbol{\beta}-\boldsymbol{\mu})\right)$$

$$\propto\prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}(-2\boldsymbol{y}^T\boldsymbol{X\beta}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta}+(\boldsymbol{\beta}-\boldsymbol{\mu})^T(\boldsymbol{\beta}-\boldsymbol{\mu}))\right),$$

where $\boldsymbol{\mu} = \mu\mathbf{1}_{k+1}$. Hence,

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \alpha, \mu, \sigma^2) \propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta}+\boldsymbol{\beta}^T\boldsymbol{\beta}-2\boldsymbol{y}^T\boldsymbol{X\beta}-2\boldsymbol{\mu}^T\boldsymbol{\beta})\right)$$

$$\propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{I}_{k+1})\boldsymbol{\beta}-2(\boldsymbol{X}^T\boldsymbol{y}+\boldsymbol{\mu})^T\boldsymbol{\beta})\right)$$

$$\propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}^T\boldsymbol{A\beta}-2\boldsymbol{w}^T\boldsymbol{\beta})\right),$$

where

$$\boldsymbol{A} = \boldsymbol{X}^T X + \boldsymbol{I}_{k+1} \quad \text{and} \quad \boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{y} + \boldsymbol{\mu}.$$

By completing the squares for the term $\boldsymbol{\beta}^T\boldsymbol{A\beta} - 2\boldsymbol{w}^T\boldsymbol{\beta}$ and eliminating the constant terms, we have

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \alpha, \mu, \sigma^2) \propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{w})^T\boldsymbol{A}(\boldsymbol{\beta}-\boldsymbol{A}^{-1}\boldsymbol{w})\right). \quad (4)$$

(ii) The conditional distribution of $\alpha|\boldsymbol{\beta}, \boldsymbol{X}, \sigma^2, \mu, \boldsymbol{y}$ is given by

$$\pi(\alpha|\boldsymbol{\beta}, \boldsymbol{X}, \mu, \sigma^2, \boldsymbol{y}) \propto \prod_{i=0}^{k}\frac{\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)}{(2+\alpha^2)\sigma^3}$$

$$\exp\left(-\frac{1}{2\sigma^2}(\alpha-a)^2\right)$$

$$\propto \prod_{i=0}^{k}\frac{\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)}{2+\alpha^2}\exp\left(-\frac{1}{2\sigma^2}(\alpha-a)^2\right). \quad (5)$$

(iii) The conditional distribution of $\mu|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \sigma^2, \boldsymbol{y}$ is given by

$$\pi(\mu|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \sigma^2, \boldsymbol{y}) \propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=0}^{k}(\beta_i-\mu)^2+(\mu-b)^2\right)\right).$$

Then, $\pi(\mu|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \sigma^2, \boldsymbol{y})$ simplifies to

$$\pi(\mu|\boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{y}) \propto \prod_{i=0}^{k}\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)$$

$$\exp\left(-\frac{k+2}{2\sigma^2}\left(\mu-\frac{\mu^*}{k+2}\right)^2\right), \quad (6)$$

where $\mu^* = b + \sum_{i=0}^{k}\beta_i$.

(iv) The conditional distribution of $\sigma^2|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \mu, \boldsymbol{y}$ is given by

$$\pi(\sigma^2|\boldsymbol{\beta}, \alpha, \mu, \boldsymbol{y}) \propto \frac{1}{\sigma^n}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}+\boldsymbol{X\beta})\right)$$

$$\times\left(\prod_{i=0}^{k}\frac{\left(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2\right)}{\sigma^3}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=0}^{k}(\beta_i-\mu)^2\right)\right.$$

$$\left.\times\frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(\alpha-a)^2\right)\right)\times\left(\frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(\mu-b)^2\right)\right.$$

$$\left.\times\frac{1}{(\sigma^2)^{c+1}}\exp\left(-\frac{1}{d\sigma^2}\right)\right).$$

By some simplification, we find that

$$\pi(\sigma^2|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \mu, \boldsymbol{y}) \propto \frac{1}{(\sigma^2)^{\frac{n+3k+2c+7}{2}}}\prod_{i=0}^{k}\left(\sigma^2+\left(\sigma-\alpha\left(\beta_i-\mu\right)\right)^2\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})+\right.\right.$$

$$\left.\left.\sum_{i=0}^{k}(\beta_i-\mu)^2+(\alpha-a)^2+(\mu-b)^2+\frac{2}{d}\right]\right).$$

Hence,

$$\pi(\sigma^2|\boldsymbol{\beta}, \boldsymbol{X}, \alpha, \mu, \boldsymbol{y}) \propto \frac{1}{(\sigma^2)^{a^*+1}}\prod_{i=0}^{k}$$

$$(\sigma^2+(\sigma-\alpha(\beta_i-\mu))^2)\exp\left(-\frac{d^*}{\sigma^2}\right). \quad (7)$$

where

$$a^* = \frac{n+3k+2c+5}{2},$$

and

$$d^* = \frac{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})+\Sigma_{i=0}^{k}(\beta_i-u)^2+(\alpha-a)^2+(\mu-b)^2+\frac{2}{d}}{2}.$$

BAYESIAN INFERENCE FOR LINEAR REGRESSION UNDER NORMAL PRIOR

We consider the linear regression model in (3) where the errors are assumed to follow the normal distribution $N_n(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$. Then the joint density of $\boldsymbol{y}$ is:

$$f(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = (2\pi)^{\frac{-n}{2}}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right).$$

To completely specify the Bayesian model, we need to specify the prior distributions for the parameters $\boldsymbol{\Theta} = (\boldsymbol{\beta}^T, \sigma^2)$. We assume the following prior structure for $\boldsymbol{\Theta}$:

$$\boldsymbol{\beta}|\sigma^2 \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) \text{ and } \sigma^2|c, d \sim IG(c, d).$$

In order to implement the MCMC algorithm, we need to obtain the full conditional distributions for the posterior distribution. Following Hoff (2009), the full conditional distributions corresponding to $\pi(\boldsymbol{\Theta}|\boldsymbol{y},\boldsymbol{X})$ can be easily shown as the followings.

(i) The conditional distribution of $\boldsymbol{\beta}$ given $\boldsymbol{y}$, $\boldsymbol{X}$ and σ^2 is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\left((\boldsymbol{\beta}-\boldsymbol{V}^{-1}\boldsymbol{W})^T\boldsymbol{V}(\boldsymbol{\beta}-\boldsymbol{V}^{-1}\boldsymbol{W})\right)\right),$$

i.e.,

$$\boldsymbol{\beta} \sim N_n(\boldsymbol{V}^{-1}\boldsymbol{W}, \sigma^2\boldsymbol{V}^{-1}).$$

(ii) The conditional distribution for $\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}$ is given by

$$\pi(\sigma^2|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\beta}) \propto \frac{1}{(\sigma^2)^{c^*+1}}\exp\left(-\frac{d^*}{\sigma^2}\right),$$

where

$$c^* = \frac{n}{2} + c$$

and

$$d^* = \frac{(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})}{2} + \frac{1}{d}.$$

BAYESIAN INFERENCE FOR LINEAR REGRESSION UNDER NON-INFORMATIVE PRIOR

Consider the linear regression models

$$y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, n,$$

where $\epsilon_i \sim \boldsymbol{N}_n(\mathbf{0}, \sigma^2 I_n)$, $= (y_1, \ldots, y_n)^T$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ and $\boldsymbol{X}_i^T = (1, X_{i1}, X_{i2}, \ldots, X_{ip})$.

Then, the joint density function of $\boldsymbol{Y}$ is given by

$$f(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}, \sigma^2) = (2\pi)^{\frac{-n}{2}}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\boldsymbol{y}-\boldsymbol{X\beta})\right).$$

Following Berger et al. (2001), we assume the standard non-informative prior as given by $\pi(\beta, \sigma^2) \propto \frac{1}{\sigma^2}$. So, the joint posterior density function is

The conditional distributions for each of the parameters are as the followings:

(i) The conditional distributions for $\boldsymbol{\beta}$ given $\boldsymbol{y}$, $\boldsymbol{X}$ and σ^2 is

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}, \sigma^2 \sim N_n(\boldsymbol{V}^{*-1}\boldsymbol{W}^*, \sigma^2\boldsymbol{V}^{*-1}).$$

(ii) It can be easily seen that the conditional distribution for σ^2 conditional on $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ is given by an inverse gamma pdf, denoted as $IG\left(\frac{n}{2}, \frac{d^{**}}{2}\right)$ where $d^{**} = \frac{(y-X\beta)^T(y-X\beta)}{2}$

SIMULATION STUDY

In this section, we conduct a simulation study using the Metropolis-Hastings algorithm by considering samples of sizes $n = 15$, 20 and 25, with 10 000 iterations inclusive of the burn-in iterations of 2000, where X_1 and X_2 represent the predictor variables for the regression model that are generated using the normal distribution with known mean and variance. In the simulation, the true values of the regression parameters β_0, β_1, β_2, variance σ^2 and the hyper-parameters which describe the ASN distribution through a hierarchical structure are assumed known. The true regression parameters are given in Tables 1, 2, and 3 while the values of hyper-parameters are given by $a = 2$, $b = 3$, $c = 10$ and $d = 1$. The point estimates and credible intervals found for the parameters of ASN distribution are compared to those found for LR-ND and LR-NI. The estimated posterior mean and standard error (S.E) found for each parameter under the different priors are shown in Tables 1, 2, and 3. In addition to these point estimates, the computed 95% credible intervals for the mean posterior of the parameters of the three models are determined. For the purpose of illustration, as given in Table 4, we provide the 95% credible intervals for the case when $n = 25$.

For assessing the convergence of the iterative simulation of the posterior mean of the parameters, plots of the simulated values as given in Figure 1 are studied.

Based on the Tables 1-4 and Figure 1, we may report the following concluding remarks:

We note that the standard errors associated with the posterior means of the linear regression estimates under alpha-skew-normal prior are less than the standard errors of the corresponding posterior means for the linear regression estimates under normal and non-informative prior distributions. In general, the standard error for all the parameters decrease when the values of n increase. We conclude that the posterior mean for the parameters of LR-ASN are found to be more precise than the posterior

TABLE 1. Posterior mean and standard error for the parameters under the three prior distributions with n = 15

Parameters	LR-ND		LR-NI		LR-ASN	
	Mean	S.E	Mean	S.E	Mean	S.E
β_0 (2.0617)	2.0805	0.8553	2.0154	0.0785	2.0288	0.0646
β_1 (3.2027)	3.2056	0.0836	3.1177	0.0923	3.2129	0.0089
β_2 (2.3862)	2.3799	0.0847	2.3155	0.0951	2.3828	0.0051
μ (2.6701)	-	-	-	-	2.6879	0.1139
α (2.1169)	-	-	-	-	2.0268	0.2470
σ^2 (0.0945)	0.1052	0.0832	0.0882	0.0426	0.0668	0.0060

TABLE 2. Posterior mean and standard error for the parameters under the three prior distributions with n = 20

Parameters	LR-ND		LR-NI		LR-ASN	
	Mean	S.E	Mean	S.E	Mean	S.E
β_0 (2.0617)	2.0992	0.7629	2.0312	0.0616	2.1152	0.0532
β_1 (3.2027)	3.2006	0.0731	3.3054	0.0620	3.2025	0.0057
β_2 (2.3862)	2.3754	0.0663	2.3429	0.0629	2.3819	0.0051
μ (2.6701)	-	-	-	-	2.7039	0.1122
α (2.1169)	-	-	-	-	2.0489	0.2335
σ^2 (0.0945)	0.1031	0.0801	0.0669	0.0262	0.0602	0.0034

TABLE 3. Posterior mean and standard error for the parameters under the three prior distributions with n = 25

Parameters	LR-ND		LR-NI		LR-ASN	
	Mean	S.E	Mean	S.E	Mean	S.E
β_0 (2.0617)	2.0821	0.6558	2.1787	0.0539	2.0982	0.0467
β_1 (3.2027)	3.1856	0.0657	3.1506	0.0589	3.1935	0.0048
β_2 (2.3862)	2.3811	0.0652	2.3715	0.0585	2.3866	0.0045
μ (2.6701)	-	-	-	-	2.6957	0.0830
α (2.1169)	-	-	-	-	2.0243	0.2188
σ^2 (0.0945)	0.1054	0.0792	0.0692	0.0234	0.0518	0.0015

TABLE 4. The 95% credible intervals for the parameters under different priors with n = 25

Parameters	LR-ND		LR-NI		LR-ASN	
	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
β_0 (2.0617)	0.8169	3.3738	2.07194	2.28813	1.9944	2.1906
β_1 (3.2027)	3.0574	3.3147	3.0358	3.26774	3.1837	3.2029
β_2 (2.3862)	2.2512	2.5087	2.25612	2.48437	2.3789	2.3971
μ (2.6701)	-	-	-	-	2.5356	2.8604
α (2.1169)	-	-	-	-	1.5822	2.4458
σ^2 (0.0945)	0.0154	0.3098	0.0377	0.1264	0.0508	0.0951

mean for linear regression models under normal and non-informative prior distributions. Given certain sample size, say n = 25, it appears that the 95% credible intervals for the parameters of LR-ASN are shorter than those credible intervals for LR-ND and LR-NI, which also indicates that a more precise estimation is obtained when the prior is assumed to follow the alpha-skew-normal distribution. The chains of the simulated posterior mean of the parameters of LR-ASN converge and become stable about certain particular mean value, indicating that sufficient information is available for accurate inference.

AN APPLICATION: THE SCOTTISH HILLS RACES DATA

The data set considered in this application is the Scottish hills races data which consists of n = 33 observations where the response variable y = time (in seconds) taken by a particular horse and two other predictor variables, denoted as x_1 = distance (in miles) and x_2 = climb (in feet). This data set has also been applied by Chatterjee and Hadi (2012). It is appropriate to study this data by plotting the histogram and Q-Q normal plot of the residuals found by fitting a multiple linear regression model under normal error which can be easily found as $\hat{y} = -10.361 + 6.6921x_1$

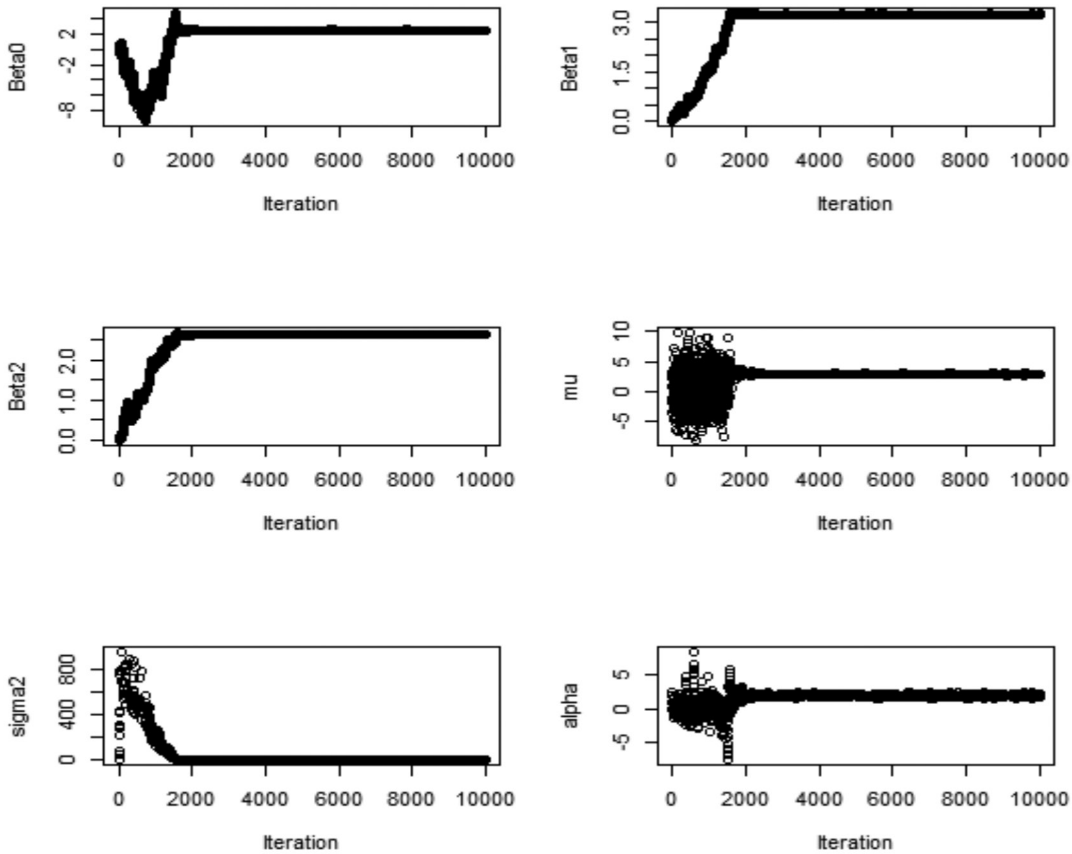

FIGURE 1. The plots of iterative simulation of the estimated posterior mean for the parameters of LR-ASN

+ $0.008x_2$. Based on the plots in Figure 2, the histogram of the residuals is found to be skewed to the left and the Q-Q normal plot of the residuals indicates departures of several points from the straight line, possibly be due to the presence of outliers in the data, in which we believe contributing to the skewness of the response values. It will be further shown in the analysis that the linear regression under alpha skew normal prior can nicely capture the skewness property inherent in the data. All the three models, which are LR-ASN, LR-ND and LR-NI, are fitted to the Scottish hills races data and the estimated posterior mean and standard deviation found for the parameters are given in Table 5. Also, note that the credible intervals for the parameters which are shown in Table 6 indicate that LR-ASN outperform LR-ND and LR-NI since the widths are found to be shorter under alpha skew normal prior as opposed to the other two priors.

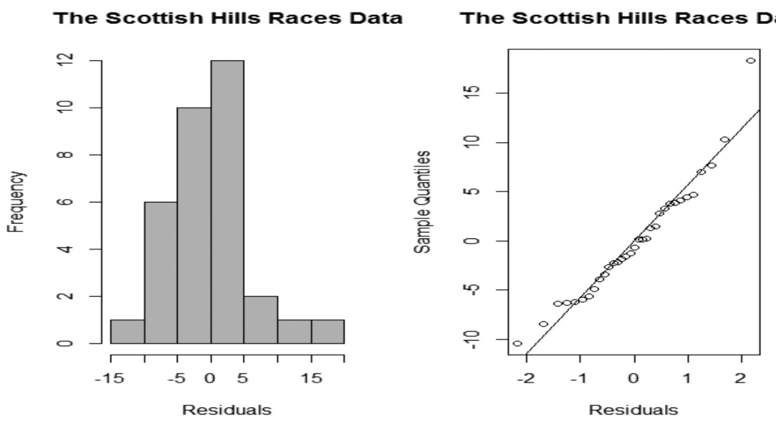

FIGURE 2. Histogram and Q-Q normal plot of the residuals found for the Scottish hills races data based on the multiple linear regression model

Thus, we have the following conclusions:

The posterior standard errors for the parameters of the linear regression model are found to be smaller for the model under alpha-skew-normal prior as compared to those found based on normal and non-informative prior distributions. This indicates that the model with alpha skew normal prior is more precise. The 95% credible intervals for LR-ASN are shorter than those credible intervals found for LR-ND and LR-NI, indicating a more precise estimation. Based on the computed values of deviance information criterion (DIC), the Bayesian linear regression model under alpha-skew-normal prior appears to fit the data better than the other models due to the smallest DIC value (Spiegelhalter et al. 2002).

The histograms of the estimated parameters found based on conditional posterior distributions for LR-ASN, LR-ND and LR-non-informative that are shown in Figures 3, 4 and 5 respectively. These figures describe the distribution of the estimated parameters. Tables 5 and 6 provide a summary of the information displayed by the histograms in terms of the estimated posterior means and the 95% credible intervals.

CONCLUSION

In this paper, we study the Bayesian statistical inference for the linear regression models. The linear regression models have been treated under three types of prior distributions, namely, the alpha-skew-normal, normal, and non-informative. Under the alpha-skew-normal prior, the conditional posterior distributions for each parameter of interest are derived. For point estimation of the parameters, Metropolis-Hastings algorithm is applied since the conditional posterior distributions are not available in a closed form. Based on the simulated data, it is found that the posterior means of the parameters under the alpha-skew-normal prior are more precise than the corresponding posterior means under normal and non-informative priors. These findings are further supported by using the Scottish hills races data. Under both the simulation study and application of Scottish hills races data, the interval estimates are found to be narrower as the sample size increase. Based on this study, we have introduced the alpha-skew-normal distribution as an alternative prior distribution for describing the uncertainty on the parameters in the linear regression model. This

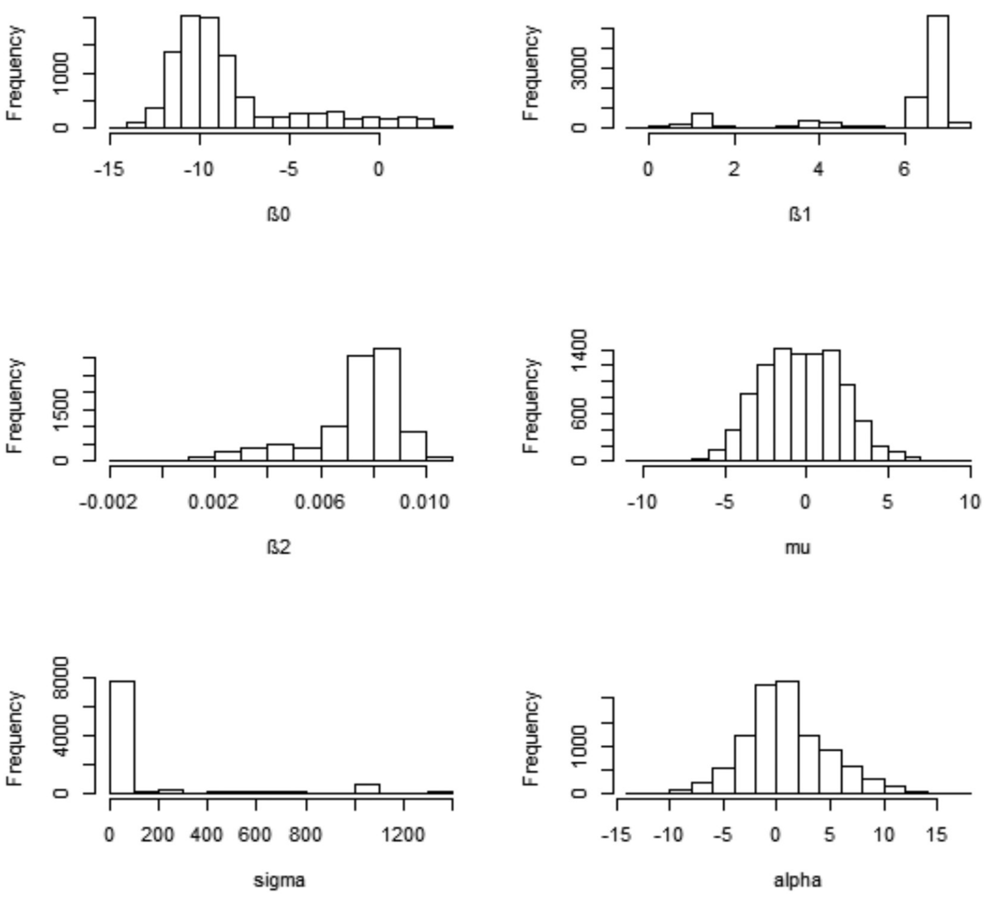

FIGURE 3. The histogram for all the estimated parameters of the LR-ASN for the Scottish hills races data

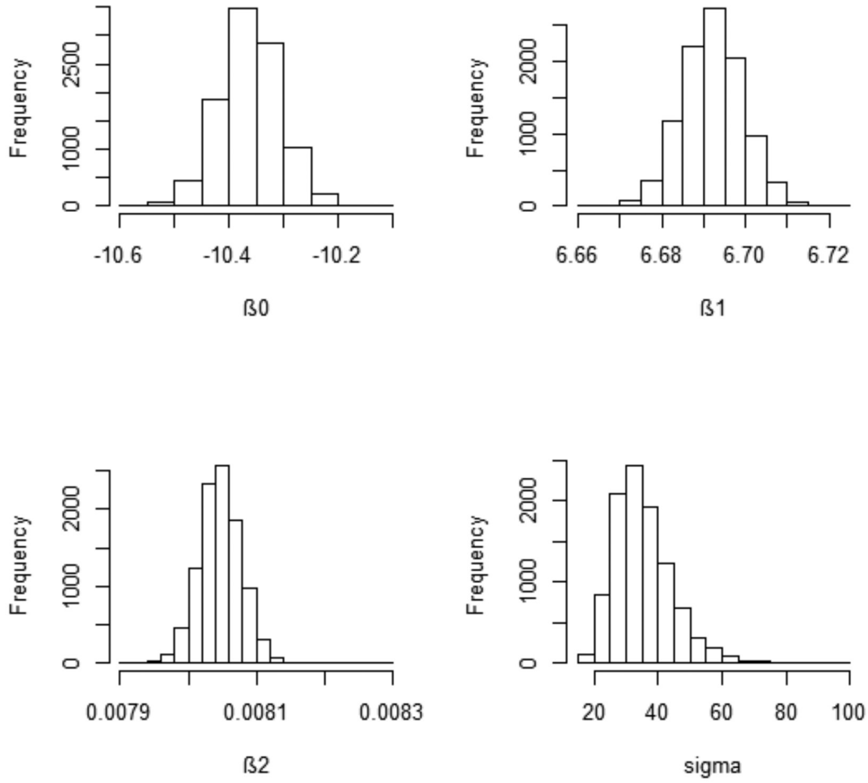

FIGURE 4. The histogram for all the estimated parameters of the LR-ND for the Scottish hills races data

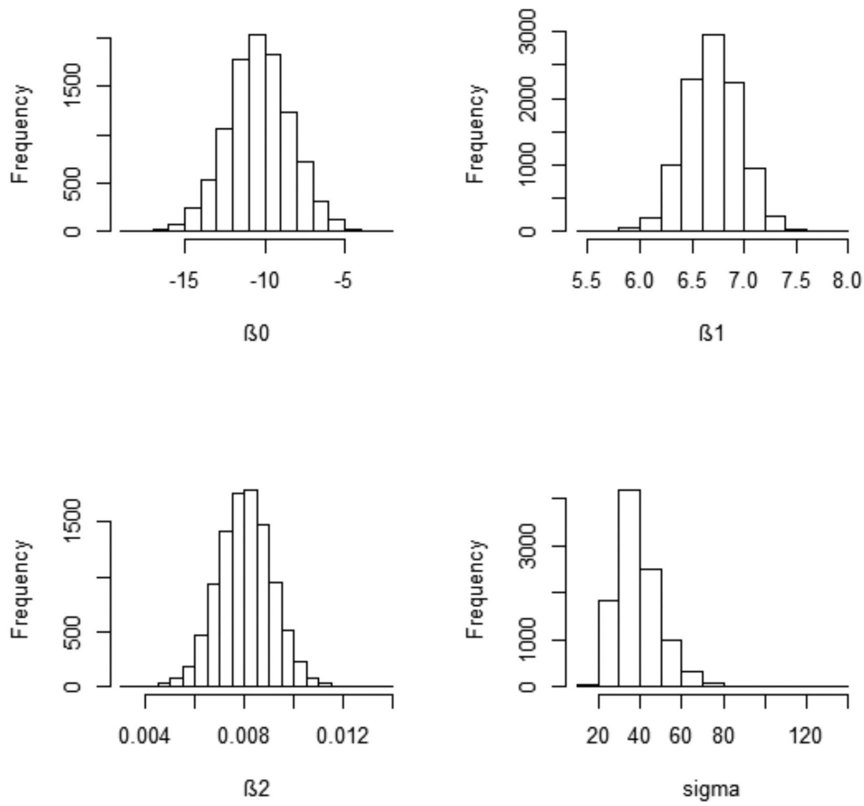

FIGURE 5. The histogram for all the estimated parameters of the LR-NI for the Scottish hills races data

TABLE 5. Posterior means and standard error for the parameters under the three fitted models for the Scottish hills races data

Parameters	LR-ND		LR-NI		LR-ASN	
	Mean	S.E	Mean	S.E	Mean	S.E
β_0	-9.9606	1.9207	-10.3697	1.9647	-9.6113	1.3425
β_1	6.6679	0.2671	6.6889	0.2649	6.6415	0.1814
β_2	0.0079	0.0011	0.0081	0.00109	0.0078	0.00068
μ	-	-	-	-	0.3227	0.2439
α	-	-	-	-	2.0109	0.1738
σ^2	39.1579	10.4515	39.3519	11.0158	21.3771	4.0284
DIC	222.4021		217.4239		209.1234	

TABLE 6. The 95% credible intervals for the parameters under the three models for the Scottish hills races data

Parameters	LR-ND		LR-NI		LR-ASN	
	2.5 %	97.5%	2.5 %	97.5%	2.5 %	97.5%
β_0	-13.708	-6.1498	-14.216	-6.4938	-12.437	-6.5543
β_1	6.1499	7.1948	6.16542	7.2308	6.2444	7.0523
β_2	0.0058	0.0101	0.00592	0.01024	0.0062	0.0093
μ	-	-	-	-	-2.8982	3.1927
α	-	-	-	-	1.7140	2.4067
σ^2	23.3717	64.9754	23.3743	65.1886	15.1076	31.9544

prior distribution is found to be particularly suitable in the case when the data exhibit the skewness property. A further study can be carried out for comparing the choices of different skew normal distributions when the Bayesian linear regression modeling is applied for analyzing the skew data.

REFERENCES

Alhamide, A., Ibrahim, K. & Alodat, M. 2016. Inference for multiple linear regression model with extended skew normal errors. *Pakistan Journal for Statistics* 32(2): 81-96.

Alodat, M. & Al-Momani, E. 2014. Skew Gaussian process for nonlinear regression. *Communications in Statistics-Theory and Methods* 43(23): 4936-4961.

Alodat, M., Al-Salman, S. & Al-Rawwash, M. 2014. A unified weighted family of distributions which contains the skew-elliptical family. *Probability and Mathematical Statistics-Poland* 34(2): 215-235.

Arellano-Valle, R. & Azzalini, A. 2006. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 33(3): 561-574.

Azzalini, A. & Capitanio, A. 2014. *The Skew-Normal and Related Families*. 3rd ed. New York: Cambridge University Press.

Arellano-Valle, R., Bolfarine, H. & Lachos, V. 2005. Skew-normal linear mixed models. *Journal of Data Science* 3(4): 415-438.

Berger, J.O., Pericchi, L.R., Ghosh, J., Samanta, T., De Santis, F., Berger, J. & Pericchi, L. 2001. Objective Bayesian methods for model selection: Introduction and comparison. *Lecture Notes-Monograph Series* 38: 135-207.

Chatterjee, S. & Hadi, A.S. 2012. *Regression Analysis by Example*. 5th ed. Hoboken, New Jersey: John Wiley & Sons.

Elal-Olivero, D. 2010. Alpha-skew-normal distribution. *Proyecciones (Antofagasta)* 29(3): 224-240.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2014. *Bayesian Data Analysis*. 3rd ed. Boca Raton: Chapman & Hall/CRC.

Genton, M.G. 2004. *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton: CRC Press.

Hoff, P.D. 2009. *A First Course in Bayesian Statistical Methods*: New York: Springer Science & Business Media.

Robert, C. & Casella, G. 2010. *Introducing Monte Carlo Methods with R*. New York: Springer Science & Business Media.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4): 583-639.

Alhamide, A.A. & Alodat, M.T.
Department of Mathematics, Statistics and Physics
Qatar University
Qatar

Kamarulzaman Ibrahim & Wan Zawiah Wan Zin*
Pusat Pengajian Sains Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: w_zawiah@ukm.edu.my

Received: 5 September 2017
Accepted: 3 August 2018