

Article

Bayesian Inference for Finite Mixture Regression Model Based on Non-Iterative Algorithm

Ang Shan ¹ and Fengkai Yang ^{2,*} ¹ School of Mathematics, Shandong University, Jinan 250100, China; ang_shan@mail.sdu.edu.cn² School of Mathematics and Statistics, Shandong University, Weihai 264209, China

* Correspondence: yfk@sdu.edu.cn

Abstract: Finite mixtures normal regression (FMNR) models are widely used to investigate the relationship between a response variable and a set of explanatory variables from several unknown latent homogeneous groups. However, the classical EM algorithm and Gibbs sampling to deal with this model have several weak points. In this paper, a non-iterative sampling algorithm for fitting FMNR model is proposed from a Bayesian perspective. The procedure can generate independently and identically distributed samples from the posterior distributions of the parameters and produce more reliable estimations than the EM algorithm and Gibbs sampling. Simulation studies are conducted to illustrate the performance of the algorithm with supporting results. Finally, a real data is analyzed to show the usefulness of the methodology.

Keywords: finite mixture regression; non-iterative sampling; missing data; Gibbs sampling; EM algorithm



Citation: Shan, A.; Yang, F. Bayesian Inference for Finite Mixture Regression Model Based on Non-Iterative Algorithm. *Mathematics* **2021**, *9*, 590. <https://doi.org/10.3390/math9060590>

Academic Editor: Manuel Alberto M. Ferreira

Received: 4 February 2021

Accepted: 6 March 2021

Published: 10 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Finite mixtures regression (FMR) models are powerful statistical tools to explore the relationship between a response variable and a set of explanatory variables from several latent homogeneous groups. The aim of FMR is to discriminate the group an observation belongs to, and reveal the dependent relationship between the response and predictor variables in the same group after classification. Finite mixture regression model with normal error assumption (FMNR) is the earliest and most used mixture regression model in practice, see Quandt and Ramsey [1], De Veaux [2], Jones and McLachlan [3], Turner [4], McLachlan [5], Frühwirth-Schnatter [6] and Faria and Soromenho [7] et al. for early literatures. In recent years, many authors have extended the finite mixture normal regression models to other error distribution based mixture regression models, such as mixture Student's t regression (Peel [8]), mixture Laplace regression (Song et al. [9,10]), mixture regression based on scale mixture of skew-normal distribution (Zeller et al. [11]), and mixture quantile regression model (Tian et al. [12], Yang et al. [13]). The classical methods to deal with these mixture models are mainly based on Gibbs sampling for Bayesian analysis and EM algorithm (Dempster [14]) for finding the maximum likelihood estimator (MLE) from frequentist perspective, and the crucial technique in these methods is to employ a group of latent variables to indicate the group an observation belongs to, and formulate a missing data structure. Although EM algorithm and Markov Chain Monte Carlo (MCMC) based algorithm are widely used in dealing with mixture models, there are still some weak points in these algorithms, which should not be omitted. As to EM algorithm, the standard error of estimated parameter is always calculated as the square root of the diagonal element of the asymptotic covariance matrix motivated by the central limit theorem, but when the size of samples is small or even medium, this approximation may be unreasonable. In the case of Gibbs or other MCMC based sampling algorithms, the samples used for statistical inferences are iteratively generated, thus the accuracy of parameter estimation may decrease due to the dependency in samples. Besides, although

there are several tests and methods to study the convergence of the generated Markov Chain, no procedure can check convincingly whether the stage of convergence has been reached upon termination of iteration. So it would be a beneficial attempt to develop some algorithm with more effectiveness and computationally feasibility to deal with complex missing data problems.

Tan et al. [15,16] proposed a kind of non-iterative sampling algorithm (named IBF algorithm after Inverse Bayesian Formula) for missing data models, which can draw independently and identically distributed (i.i.d.) samples from posterior distribution. Without convergence diagnosis, the i.i.d. samples can be directly used to estimate model parameters and their standard errors, thus avoids the shortcomings in EM algorithm and Gibbs sampling. Recently, extensive applications of IBF algorithm have seen prosperous development in documents, to name a few, Yuan and Yang [17] developed a IBF algorithm for Student’s t regression with censoring, Yang and Yuan [18] applied the idea of IBF sampling to quantile regression models, also Yang and Yuan [19] designed a IBF algorithm for robust regression models with scale mixture of normal distribution, which includes normal, Student’s t, Slash, and contaminated normal as special cases. Tsionas [20] applied the idea to financial area and proposed a non-iterative method for posterior inference in stochastic volatility models.

Inspired by Tian et al. [15,16], in this paper, we propose an effective Bayesian statistical inference for FMNR model from a non-iterative perspective. We first introduce a group of latent multinomial distributed mixture component variables to formulate a missing data structure, and then combine the EM algorithm, inverse Bayes Formula, and sampling/importance resampling (SIR) ([21,22]) algorithm into a non-iterative sampling algorithm. Finally, we implement the IBF sampling to generate i.i.d. samples from posterior distributions and use these samples directly to estimate the parameters. We conduct simulation studies to assess the performance of the procedure by comparison with the EM algorithm and Gibbs sampling, and apply it to a classical data set with supporting results. The IBF sampling procedure can be directly extended to other mixture regression models, such as mixture regression model driven by Student’s t or Laplace error distributions.

2. Finite Mixture Normal Regression (FMNR) Model

In this subsection, we first introduce the usual FMNR model, and then display the complete likelihood function for observations and a set of latent variables, which are the mixture component variables, indicating the group an observation belongs to.

Let y_i represent the i th observation of a response variable and $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i,p-1})^T$ be the i th observation of a set of $p - 1$ explanatory variables. The FMNR model is described as follows

$$y_i \sim \sum_{j=1}^g \lambda_j N(\mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2), \quad i = 1, \dots, n,$$

therefore, the distribution of response variables is modeled as finite mixture normal distributions with g components. Here, λ_j is the mixture proportion with $0 < \lambda_j < 1$, and $\sum_{j=1}^g \lambda_j = 1$, $\boldsymbol{\beta}_j$ is the p -dimension regression coefficient vector in j th group with the first element denoting the intercept, σ_j^2 represents the variance of the error distribution in group j , and $N(\mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)$ represents the normal distribution with mean $\mathbf{x}_i^T \boldsymbol{\beta}_j$ and variance σ_j^2 .

Denote parameters

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_g^T)^T, \boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_g^2)^T,$$

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_g)^T, \boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\sigma}^{2T}, \boldsymbol{\lambda}^T)^T,$$

and observation vector $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, the likelihood of \mathbf{y} in the FMNR model is that

$$\begin{aligned}
 L(\boldsymbol{\theta}|\mathbf{y}) &= \prod_{i=1}^n \sum_{j=1}^g \lambda_j N(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2) \\
 &= \prod_{i=1}^n \sum_{j=1}^g \lambda_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{2\sigma_j^2}\right\},
 \end{aligned}
 \tag{1}$$

where $N(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)$ denotes the density function of distribution $N(\mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)$ evaluated at y_i . It is challenging to get the maximum likelihood estimations of the parameters by directly maximizing the above likelihood function, and it is also difficult to get the full conditional distributions used in Bayesian analysis. In documents, one effective approach to deal with mixture models is to introduce a group of latent variables which establishes a missing data structure, in this way, some data augmentation algorithms, such as EM algorithm and Gibbs sampling can be easily performed, besides, by taking advantage of the missing data structure, a new non-iterative sampling algorithm can be smoothly carried out.

Here, we introduce a group of latent membership-indicators $\mathbf{G}_i = (G_{i1}, \dots, G_{ig})^T$, such that

$$G_{ij} = \begin{cases} 1, & \text{if } y_i \text{ belongs to group } j, \\ 0, & \text{otherwise,} \end{cases}$$

with $P(G_{ij} = 1) = \lambda_j$ for $j = 1, 2, \dots, g$, and $\sum_{j=1}^g G_{ij} = 1$. Given the mixing probabilities $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_g)^T$, the latent variables $\mathbf{G}_1, \dots, \mathbf{G}_n$ are independent of each other, with multinomial densities,

$$f(\mathbf{G}_i|\boldsymbol{\lambda}) \sim \lambda_1^{G_{i1}} \dots \lambda_{g-1}^{G_{i,g-1}} (1 - \lambda_1 - \dots - \lambda_{g-1})^{G_{ig}}.$$

Note that given $G_{ij} = 1$, we have $y_i|G_{ij} = 1 \sim N(\mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)$, and thus get the following complete likelihood of observations \mathbf{y} and latent variables $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_n)$:

$$\begin{aligned}
 L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) &= \prod_{i=1}^n \prod_{j=1}^g (\lambda_j N(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2))^{G_{ij}} \\
 &= \prod_{i=1}^n \prod_{j=1}^g \left(\lambda_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{2\sigma_j^2}\right\} \right)^{G_{ij}}.
 \end{aligned}$$

The above complete likelihood is crucial to implement EM algorithm, Gibbs sampling and IBF algorithm.

3. Bayesian Inference Using IBF Algorithm

3.1. The Prior and Conditional Distributions

In Bayesian framework, suitable prior distributions should be adopted for the parameters. With the purpose to emphasize the importance of data and for simplicity, a set of independent non-informative prior distributions are used, which are

$$\pi(\boldsymbol{\beta}_j, \sigma_j^2) \propto \frac{1}{\sigma_j^2}, \quad \pi(\lambda_1, \dots, \lambda_g) \propto 1.$$

In fact, the prior of $(\lambda_1, \dots, \lambda_g)$ is Dirichlet distribution with concentration equals to 1, also known as a uniform prior over the space $\lambda_1 + \dots + \lambda_g = 1$. These priors are improper and noninformative, but the posterior distributions are proper and meaningful. Then given complete data (\mathbf{y}, \mathbf{G}) , the posterior density of $\boldsymbol{\theta}$ is shown as

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) &\propto L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) \prod_{j=1}^g \pi(\boldsymbol{\beta}_j, \sigma_j^2) \pi(\sigma_j^2) \pi(\boldsymbol{\lambda}) \\ &\propto L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G}) \prod_{j=1}^g \frac{1}{\sigma_j^2}. \end{aligned}$$

In order to implement the IBF sampling in FMNR model, we now present several required conditional posterior distributions, which are all proportional to the complete posterior density $\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{G})$.

1. Conditional distribution $\pi(\mathbf{G}_i|\boldsymbol{\theta}, \mathbf{y})$

For $i = 1, \dots, n$, the conditional posterior distribution of $\mathbf{G}_i = (G_{i1}, \dots, G_{ig})^T$ is

$$\pi(G_{i1}, \dots, G_{ig}|\boldsymbol{\theta}, \mathbf{y}) \propto \prod_{j=1}^g (\lambda_j \mathbf{N}(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2))^{G_{ij}},$$

the right side of (1) is the kernel of the probability density function of multimomial distribution, thus we obtain that

$$\mathbf{G}_i|\boldsymbol{\theta}, \mathbf{y} \sim \text{Mult}(1, r_{i1}, \dots, r_{ig}), \tag{2}$$

where

$$r_{ij} = \frac{\lambda_j \mathbf{N}(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma_j^2)}{\sum_{k=1}^g \lambda_k \mathbf{N}(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_k, \sigma_k^2)}, \tag{3}$$

with $\text{Mult}(1, r_{i1}, \dots, r_{ig})$ standing for the multinomial distribution with parameters (r_{i1}, \dots, r_{ig}) .

2. Conditional distribution $\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G})$

The conditional density of $\boldsymbol{\beta}_j$ is given by

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}) \propto \exp\left\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j)^2}{2\sigma_j^2}\right\}.$$

Let $\mathbf{G}_{(j)} = (G_{1j}, \dots, G_{nj})$, $\mathbf{W}_j = \text{diag}(\mathbf{G}_{(j)})$,

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{p-1})^T,$$

and

$$\mathbf{B}_j^{-1} = \mathbf{X}^T \mathbf{W}_j \mathbf{X}, \quad \mathbf{b}_j = \mathbf{B}_j \mathbf{X}^T \mathbf{W}_j \mathbf{y},$$

then, we have

$$\pi(\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)}) \propto \exp\left\{-\frac{(\boldsymbol{\beta}_j - \mathbf{b}_j)^T \mathbf{B}_j^{-1} (\boldsymbol{\beta}_j - \mathbf{b}_j)}{2\sigma_j^2}\right\},$$

consequently,

$$\boldsymbol{\beta}_j|\sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)} \sim \mathbf{N}(\mathbf{b}_j, \sigma_j^2 \mathbf{B}_j). \tag{4}$$

3. Conditional distribution $\pi(\sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)})$

The conditional density of σ_j^2 is obtained as

$$\pi(\sigma_j^2|\mathbf{y}, \mathbf{G}_{(j)}) \propto \left(\frac{1}{\sigma_j^2}\right)^{\frac{\sum_{i=1}^n G_{ij} - p}{2} + 1} \exp\left\{-\frac{\sum_{i=1}^n G_{ij}(y_i - \mathbf{x}_i^T \mathbf{b}_j)^2}{2\sigma_j^2}\right\}, \tag{5}$$

noticing that the right side of Equation (4) is the kernel of inverse-Gamma density, we get

$$\sigma_j^2 | \mathbf{y}, \mathbf{G}_{(j)} \sim \text{IG} \left(\frac{\sum_{i=1}^n G_{ij} - p}{2}, \frac{\sum_{i=1}^n G_{ij} (y_i - \mathbf{x}_i^T \mathbf{b}_j)^2}{2} \right). \tag{6}$$

4. Conditional distribution $\pi(\lambda | \mathbf{G})$

Finally, we obtain the conditional density of $\lambda = (\lambda_1, \dots, \lambda_g)^T$, that is

$$\pi(\lambda_1, \dots, \lambda_g | \mathbf{G}) \propto \prod_{j=1}^g \lambda_j^{\sum_{i=1}^n G_{ij} + 1 - 1}, \tag{7}$$

the right side of (6) is the kernel of Dirichlet distribution, thus

$$\lambda | \mathbf{G} \sim \mathcal{D} \left(\sum_{i=1}^n G_{i1} + 1, \dots, \sum_{i=1}^n G_{ig} + 1 \right). \tag{8}$$

3.2. IBF Sampler

Based on the following expression

$$\pi(\theta | \mathbf{y}) = \sum_{\mathbf{G}} \pi(\theta | \mathbf{y}, \mathbf{G}) \pi(\mathbf{G} | \mathbf{y}),$$

and

$$\begin{aligned} \pi(\theta | \mathbf{y}, \mathbf{G}) &= \pi(\beta, \sigma^2, \lambda | \mathbf{y}, \mathbf{G}) \\ &= \prod_{j=1}^g \pi(\beta_j | \sigma_j^2, \mathbf{y}, \mathbf{G}_{(j)}) \pi(\sigma_j^2 | \mathbf{y}, \mathbf{G}_{(j)}) \pi(\lambda | \mathbf{y}, \mathbf{G}), \end{aligned}$$

in order to draw i.i.d. samples from $\pi(\theta | \mathbf{y})$, we should first get i.i.d. samples from $\pi(\mathbf{G} | \mathbf{y})$. According to Tan et al. [15,16], and based on the inverse Bayes formula, the posterior distribution of \mathbf{G} can be expressed as

$$\pi(\mathbf{G} | \mathbf{y}) \propto \frac{\pi(\mathbf{G} | \mathbf{y}, \theta_0)}{\pi(\theta_0 | \mathbf{y}, \mathbf{G})},$$

where θ_0 is the posterior mode of $\pi(\theta | \mathbf{y})$ obtained by EM algorithm. The idea of IBF sampler in the EM framework is to use EM algorithm to obtain $\pi(\mathbf{G} | \mathbf{y}, \theta_0)$, an optimal important sampling density which can best approximate $\pi(\mathbf{G} | \mathbf{y})$, thus by implementing SIR algorithm, i.i.d. samples can be generated approximately from $\pi(\mathbf{G} | \mathbf{y})$.

Specifically, the IBF algorithm for FMNR model includes the following four steps:

Step 1. Based on (2), draw i.i.d. samples

$$\mathbf{G}_i^{(l)} = (G_{i1}^{(l)}, \dots, G_{ig}^{(l)})^T \sim \text{Mult}(r_{i1}^0, \dots, r_{ig}^0), l = 1, \dots, L,$$

for $i = 1, 2, \dots, n$, where r_{ij}^0 is the same as r_{ij} in (3), with θ replaced by θ_0 . Denote $\mathbf{G}^{(l)} = (\mathbf{G}_1^{(l)}, \mathbf{G}_2^{(l)}, \dots, \mathbf{G}_n^{(l)})$.

Step 2. Calculate the weights

$$\omega_l = \frac{\frac{1}{\pi(\theta_0 | \mathbf{y}, \mathbf{G}^{(l)})}}{\sum_{l=1}^L \frac{1}{\pi(\theta_0 | \mathbf{y}, \mathbf{G}^{(l)})}}, \quad l = 1, 2, \dots, L,$$

where $\pi(\theta_0 | \mathbf{y}, \mathbf{G}^{(l)})$ is calculated as $\pi(\theta_0 | \mathbf{y}, \mathbf{G})$ with \mathbf{G} replaced by $\mathbf{G}^{(l)}$.

Step 3. Choose a subset from $\{\mathbf{G}^{(l)}\}_{l=1}^L$ via resampling without replacement from the discrete distribution on $\{\mathbf{G}^{(l)}\}_{l=1}^L$ with probabilities $\{\omega_l\}_{l=1}^L$ to obtain an i.i.d. sample of size $K (< L)$ approximately from $\pi(\mathbf{G}|\mathbf{y})$, denoted by $\{\mathbf{G}^{(k)}\}_{k=1}^K$.

Step 4. Based on (8), generate

$$\boldsymbol{\lambda}^{(k)}|\mathbf{G}^{(k)} \sim \mathcal{D}\left(\sum_{i=1}^n G_{i1}^{(k)} + 1, \dots, \sum_{i=1}^n G_{ig}^{(k)} + 1\right).$$

For $j = 1, \dots, g$, based on (6), generate

$$\sigma_j^{2(k)}|\mathbf{y}, \mathbf{G}^{(j)} \sim \text{IG}\left(\frac{\sum_{i=1}^n G_{ij}^{(k)} - p}{2}, \frac{\sum_{i=1}^n G_{ij}^{(k)} (y_i - \mathbf{x}_i^T \mathbf{b}_j^{(k)})^2}{2}\right),$$

and based on (4), generate

$$\boldsymbol{\beta}_j^{(k)}|\sigma_j^{2(k)}, \mathbf{y}, \mathbf{G}^{(j)} \sim \mathbb{N}(\mathbf{b}_j^{(k)}, \sigma_j^{2(k)} \mathbf{B}_j^{(k)}),$$

where

$$\mathbf{G}^{(j)} = (G_{1j}^{(k)}, \dots, G_{nj}^{(k)}), \quad \mathbf{W}_j^{(k)} = \text{diag}(\mathbf{G}^{(j)}),$$

and

$$\mathbf{b}_j^{(k)} = \mathbf{B}_j^{(k)} \mathbf{X}^T \mathbf{W}_j^{(k)} \mathbf{y}, \quad \mathbf{B}_j^{(k)} = (\mathbf{X}^T \mathbf{W}_j^{(k)} \mathbf{X})^{-1}.$$

Finally, $\{(\boldsymbol{\beta}^{(k)}, \sigma^{2(k)}, \boldsymbol{\lambda}^{(k)})\}_{k=1}^K$ are the i.i.d. samples approximately generated from $\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}|\mathbf{y})$.

In this section, we employ a non-informative prior setting for the FMNR model and develop a IBF sampler. We will extend this procedure to the conjugate prior situation in the supplement.

4. Results

4.1. Simulation Studies

To investigate the performance of the proposed algorithm, we conduct some simulations under different situations by means of four criteria such as the mean, mean square error, mean absolute deviance, and the coverage probability to evaluate the method by comparison with the classical EM and Gibbs sampling.

The data $\{(x_i, y_i), i = 1, \dots, n\}$ are generated from the following mixture model

$$y_i = \begin{cases} 5 - 5x_i + \varepsilon_i, & \text{if } z_i = 1; \\ -5 + 5x_i + \varepsilon_i, & \text{if } z_i = 2, \end{cases}$$

where z_i is a group indicator of y_i with $P(z_i = 1) = 0.5$. Here, x_i is generated from the uniform distribution on interval $(-1, 1)$, and ε_i is the random error. We consider the following symmetric distributions to describe error:

1. Normal distribution: $\varepsilon \sim N(0, 1)$;
2. Student's t distribution: $\varepsilon \sim t(3)$;
3. Laplace distribution: $\varepsilon \sim L(0, 1)$;
4. Logistic distribution: $\varepsilon \sim \text{Logistic}(0, 1)$.

In these error distributions, cases 2–4 are heavy-tailed symmetric distributions and often used in the literature to mimic the outlier situations. We use these distributions to illustrate the robustness and effectiveness of the procedure, and implement a 2-components mixture normal regression model with IBF sampling algorithm, EM algorithm and Gibbs sampling to fit these simulated data. The prior setting is the same as in Section 3.1.

To evaluate the finite sample performance of the proposed methods, the experiment is replicated for 200 times with sample size $n = 100$ and 300, respectively, for each error

distribution. In the i -th replication, we set $L = 6000$ and $K = 3000$, then get 3000 i.i.d samples from the posterior distributions using the proposed non-iterative sampling algorithm. Similarly, when performing Gibbs sampling, we generate 6000 Gibbs samples and discard the first 3000 as burn-in samples, then use the later 3000 after burn-in samples as effective samples to estimate the parameters. We record the mean of these samples as $\hat{\theta}^{(i)}$, for $i = 1, 2, \dots, 200$, and calculate the means (Mean), mean squared errors (MSE) and mean absolute deviances (MAD) of $\{\hat{\theta}^{(i)}\}_{i=1}^{200}$, which are expressed as

$$\text{Mean} = \frac{1}{200} \sum_{i=1}^{200} \hat{\theta}^{(i)}, \text{MSE} = \frac{1}{200} \sum_{i=1}^{200} (\hat{\theta}^{(i)} - \theta)^2, \text{MAD} = \frac{1}{200} \sum_{i=1}^{200} |\hat{\theta}^{(i)} - \theta|.$$

Besides, we calculate the coverage probability (CP) of 200 intervals with 95% confidence level which cover the true values of parameters. The evaluation criteria for EM algorithm are the same as the above four statistics, with $\theta^{(i)}$ denoting posterior mode instead of posterior mean in the i -th replication. These four statistics are used to assess the performance of the three algorithms.

Simulation results are presented in Tables 1–3. From these tables, we are delighted to see that:

1. All the three algorithms recover the true parameters successfully, since most of the Means based on the 200 replications are very close to the corresponding true values of parameters, and the MSEs and MADs are rather small. Besides, almost all of the coverage probabilities are around 0.95, which is the true level of confidence interval. We also see that as sample size increases from 100 to 300, the MSEs and MADs decrease and fluctuate steadily in a small range;
2. In general, for fixed size of samples, the MSEs and MADs in IBF algorithm are smaller than the ones in EM algorithm and Gibbs sampling, which show that the IBF sampling algorithm can produce more accurate estimators.

Table 1. Monte Carlo simulation results for IBF sampler.

Distribution	Parameter	IBF ($n = 100$)				IBF ($n = 300$)			
		Mean	MSE	MAD	CP	Mean	MSE	MAD	CP
Normal	β_{10}	5.0105	0.0225	0.1160	0.93	4.9971	0.0069	0.0664	0.96
	β_{11}	−5.000	0.0748	0.2202	0.93	−4.9983	0.0250	0.1228	0.94
	β_{20}	−4.9940	0.0214	0.1171	0.965	−5.0061	0.0071	0.0662	0.93
	β_{21}	4.9949	0.0685	0.2053	0.940	5.0240	0.0230	0.1209	0.95
	λ_1	0.4937	0.0025	0.0411	0.965	0.4968	0.0008	0.0230	0.94
t (3)	β_{10}	5.0110	0.0574	0.1884	0.96	5.0206	0.0184	0.1100	0.96
	β_{11}	−5.0290	0.2251	0.3545	0.91	−5.0501	0.0648	0.1975	0.955
	β_{20}	−4.9817	0.0549	0.1785	0.96	−5.0056	0.0165	0.1048	0.985
	β_{21}	5.0243	0.2083	0.3496	0.95	5.0348	0.0852	0.2284	0.92
	λ_1	0.4973	0.0032	0.0401	0.95	0.5005	0.0010	0.0262	0.93
Logistic (0,1)	β_{10}	4.9907	0.0782	0.2215	0.945	5.0064	0.0327	0.1443	0.940
	β_{11}	−4.9812	0.2347	0.3890	0.905	−4.9892	0.1026	0.2491	0.905
	β_{20}	−5.0054	0.0752	0.2187	0.955	−5.0053	0.0314	0.1428	0.915
	β_{21}	4.9892	0.2388	0.3838	0.915	4.9810	0.0890	0.2308	0.920
	λ_1	0.4917	0.0035	0.0477	0.92	0.4972	0.0009	0.0248	0.955
Laplace (0,1)	β_{10}	5.0182	0.0420	0.1631	0.95	5.0119	0.0138	0.0950	0.97
	β_{11}	−4.9800	0.1385	0.2799	0.93	−5.0029	0.0540	0.1856	0.935
	β_{20}	−5.0161	0.0475	0.1644	0.915	−5.0133	0.0160	0.0986	0.925
	β_{21}	4.9743	0.1691	0.3163	0.925	5.0014	0.0428	0.1705	0.950
	λ_1	0.4996	0.0021	0.0429	0.95	0.4992	0.0010	0.0231	0.94

Table 2. Monte Carlo simulation results for Gibbs sampling.

Distribution	Parameter	Gibbs ($n = 100$)				Gibbs ($n = 300$)			
		Mean	MSE	MAD	CP	Mean	MSE	MAD	CP
Normal	β_{10}	5.0060	0.0228	0.1213	0.955	5.0031	0.0076	0.0707	0.96
	β_{11}	-5.0047	0.0871	0.2381	0.940	-5.0024	0.0313	0.1413	0.93
	β_{20}	-5.000	0.0268	0.1311	0.945	-4.9943	0.0073	0.0683	0.95
	β_{21}	5.0356	0.0810	0.2272	0.950	5.0079	0.0265	0.1281	0.94
	λ_1	0.4989	0.0024	0.0392	0.965	0.5008	0.0008	0.0238	0.955
t(3)	β_{10}	4.9670	0.0732	0.1919	0.955	4.9724	0.0616	0.1909	0.975
	β_{11}	-5.0424	0.2138	0.3486	0.955	-5.0850	0.1967	0.3215	0.960
	β_{20}	-4.9766	0.0789	0.2105	0.965	-4.9761	0.0664	0.1928	0.95
	β_{21}	5.1234	0.2851	0.4129	0.955	5.0654	0.2010	0.3415	0.935
	λ_1	0.5001	0.0032	0.0449	0.92	0.5006	0.0027	0.0408	0.955
Logistic (0,1)	β_{10}	4.9644	0.0861	0.2355	0.960	4.9937	0.0246	0.1264	0.960
	β_{11}	-5.0682	0.3350	0.4700	0.935	-5.0430	0.0872	0.2344	0.925
	β_{20}	-4.9775	0.0894	0.2444	0.96	-4.9958	0.0367	0.1483	0.915
	β_{21}	5.0036	0.2960	0.4289	0.955	4.9960	0.0979	0.2489	0.945
	λ_1	0.4983	0.0031	0.0437	0.93	0.50497	0.0009	0.0244	0.97
Laplace (0,1)	β_{10}	5.0046	0.0447	0.1645	0.950	5.0094	0.0145	0.0953	0.965
	β_{11}	-5.0263	0.1370	0.2893	0.975	-5.0065	0.0598	0.1921	0.940
	β_{20}	-4.9849	0.0395	0.1561	0.975	-5.0019	0.0137	0.0940	0.975
	β_{21}	5.0549	0.1705	0.3376	0.950	4.9823	0.0489	0.1717	0.950
	λ_1	0.5050	0.0028	0.0438	0.96	0.4983	0.0009	0.0237	0.955

Table 3. Monte Carlo simulation results for EM algorithm.

Distribution	Parameter	EM ($n = 100$)				EM ($n = 300$)			
		Mean	MSE	MAD	CP	Mean	MSE	MAD	CP
Normal	β_{10}	4.9921	0.0226	0.1199	0.930	5.0195	0.0278	0.1346	0.955
	β_{11}	-4.9789	0.0753	0.2203	0.935	-4.9964	0.0227	0.1214	0.960
	β_{20}	-5.0097	0.0243	0.1244	0.93	-4.9959	0.0078	0.0684	0.935
	β_{21}	4.9887	0.0690	0.2045	0.95	5.0009	0.0236	0.1246	0.960
	λ_1	0.5020	0.0024	0.0386	0.945	0.4984	0.0008	0.0241	0.955
t(3)	β_{10}	4.9919	0.0569	0.1889	0.92	5.0245	0.0206	0.1184	0.945
	β_{11}	-4.9531	0.2392	0.3826	0.89	-0.5033	0.0837	0.2219	0.905
	β_{20}	-4.9900	0.0694	0.2001	0.915	-4.9987	0.0203	0.1086	0.935
	β_{21}	5.0178	0.2087	0.3505	0.920	5.0320	0.0876	0.2336	0.915
	λ_1	0.4970	0.0032	0.0449	0.925	0.4972	0.0010	0.0261	0.96
Logistic (0,1)	β_{10}	5.0291	0.0786	0.2229	0.925	5.0195	0.0278	0.1346	0.955
	β_{11}	-4.9373	0.2377	0.3913	0.945	-5.0232	0.0992	0.2514	0.915
	β_{20}	-5.0034	0.0791	0.2233	0.925	-5.0061	0.0276	0.1310	0.950
	β_{21}	4.9162	0.2440	0.3903	0.940	4.9769	0.0901	0.2316	0.935
	λ_1	0.4962	0.0035	0.0487	0.95	0.5026	0.0009	0.0247	0.965
Laplace (0,1)	β_{10}	4.9928	0.0425	0.1638	0.940	4.9982	0.0177	0.1076	0.935
	β_{11}	-4.9637	0.1714	0.3282	0.895	-4.9834	0.0459	0.1749	0.970
	β_{20}	-5.0185	0.0486	0.1697	0.895	-5.0125	0.0133	0.0926	0.965
	β_{21}	5.0064	0.1831	0.3334	0.925	4.9880	0.0525	0.1795	0.940
	λ_1	0.4987	0.0031	0.0445	0.95	0.4957	0.0009	0.0255	0.955

Thus, from these simulation studies, we can draw the conclusion that the IBF sampling algorithm is comparable to, and in most cases more effective than the EM algorithm and Gibbs sampling when modelling mixture regression data. Besides that, the IBF algorithm is much faster than the iterative Gibbs sampling, for example, when $n = 100$, the mean running time is 6 s for IBF and 10.39 s for Gibbs, and when $n = 300$, it is 19.58 for IBF and 31.16 s for Gibbs.

4.2. Real Data Analysis

The tone perception data stem from an experiment of Cohen [23] on the perception of musical tones by musicians. In this study, a pure fundamental tone was played to a trained musician with electronically generated overtones added, which is determined by a stretch ratio. The predictor variable is stretch ratio (X) and the response variable is the tuning ratio (Y), which is defined as the ratio of the adjusted tone to the fundamental tone. 150 pairs of observations with the same musician were recorded. The data is typically suitable for mixture regression modelling for that two separate trends clearly emerge in the scatter plot Figure 1. The dataset “tonedata” can be found in the R package `mixtools` authored by Benaglia et al. [24]. Many articles have analyzed this dataset using a mixture of mean regressions framework, see Cohen [23], Viele and Tong [25], Hunter and Young [26], and Young [27]. Recently, Song et al. [20], Wu and Yao [28] analyzed the data using a 2-component mixtures median regression models from parametric and semi-parametric perspective, respectively.

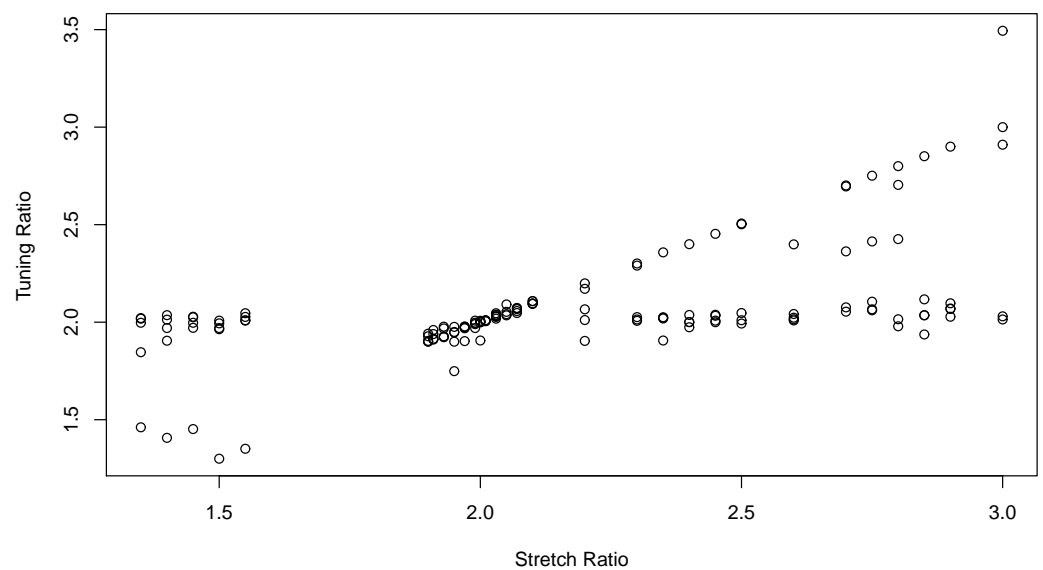


Figure 1. Scatter plot of tone perception data.

Now we re-analyze this data with the following 2-components FMNR model:

$$Y \sim \lambda_1 N(\beta_{10} + \beta_{11}X, \sigma_1^2) + \lambda_2 N(\beta_{20} + \beta_{21}X, \sigma_2^2),$$

and implement IBF algorithm by comparison with the EM and Gibbs sampling. For EM algorithm, we let $\beta_{10} = 1$, $\beta_{11} = -1$, $\beta_{20} = -1$, $\beta_{21} = 1$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$, $\lambda_1 = 0.6$, $\lambda_2 = 0.4$ as initial values, and use the R function `mixreg` and `covmix` in R package `mixreg` (authored by Turner [29]) to estimate the parameters and calculate the standard deviance. As to IBF algorithm, we set $K = 10,000$ and $J = 60,000$, thus get 6000 posterior samples. Similarly, in Gibbs sampling, we generate 10,000 Gibbs samples and discard the first 4000 as burn-in samples, then use the later 6000 after burn-in samples as effective samples to estimate the parameters.

Table 4. Estimations with EM, Gibbs, and IBF in Tone data analysis.

Algorithm	Estimates	β_{10}	β_{11}	β_{20}	β_{21}	σ_1^2	σ_2^2	λ_1	λ_2
EM	Mean	1.9163	0.0425	−0.0192	0.9922	0.0021	0.0176	0.6977	0.3022
	Sd	0.0226	0.0102	0.1021	0.0441	0.0003	0.0041	0.0484	0.0484
IBF	Mean	1.9162	0.0426	−0.0211	0.9930	0.0196	0.0022	0.6951	0.3048
	Sd	0.0228	0.0103	0.1072	0.0462	0.0048	0.0003	0.0444	0.0444
Gibbs	Mean	1.9162	0.0427	−0.0198	0.9921	0.0202	0.0022	0.6983	0.3017
	Sd	0.0233	0.0105	0.1124	0.0481	0.0055	0.0003	0.0475	0.0475

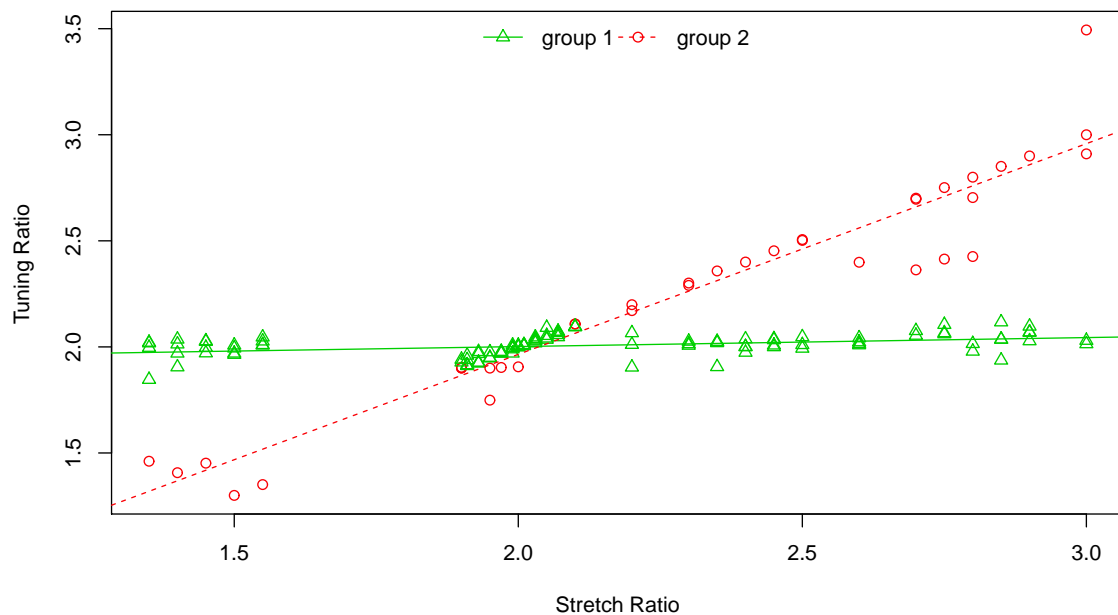


Figure 2. Scatter plot of tone perception data with estimated mixture regression lines by IBF algorithm.

Table 4 reports the estimated means and standard deviations of the parameters based on the above three algorithms. Clearly, the IBF algorithm are comparable to the EM and Gibbs sampling with very slight differences in means and standard deviations. We should notice that, although EM algorithm is effective for parameter estimation in mixture models, the standard deviations in EM algorithm are calculated as the square roots of the diagonal elements for the observed inverse information matrix, which are based on the asymptotic theory of MLE, however when the sample size is small or moderate, the estimated standard deviations may be unreliable. Therefore, compared with EM algorithm, we prefer to IBF sampling, for that the standard deviations can be directly estimated based on the i.i.d. posterior samples without extra effort. Figure 2 is the scatter plot with fitted mixture regression lines based on IBF, which clearly describes the linear relationship between turning ratio and stretch ratio in two different groups. In this figure, the observations are shown in different symbols, and classified using the mean of $\{\mathbf{G}^{(l_k)}\}_{k=1}^{3000}$ by comparison with 0.5, when the i -element of the mean vector is greater than 0.5, the i observation is classified to group one. In the supplement, we present the plots of ergodic means of posterior samples for all parameters based on IBF sampler and Gibbs sampler. We can see from these plots that for each parameter, the ergodic means based on IBF algorithm converge much faster than the ones based on Gibbs sampling, which is not surprising for that the IBF samples are i.i.d generated from the posterior distributions. Considering this point, the IBF sampling outperforms the Gibbs sampler.

For illustrating purpose, in the supplement, the histograms with kernel density estimation curves based on 6000 IBF samples and 6000 Gibbs samples are plotted in Figures S5–S8, with the red lines denoting the posterior means. In general, the posterior distributions of

each parameter based on IBF algorithm and Gibbs sampler are very similar. These plots vividly depict the posterior distributions of the parameters, for example, the distributions of regression coefficients are nearly symmetric, resemble normal distributions, but the distributions of variances σ_1^2 and σ_2^2 are both right skewed with a long tail at the right side of the distributions, which cannot be spotted by the asymptotically normality theory for MLE (or posterior mode).

Therefore, the IBF sampling algorithm provides an access to estimate the posterior distributions of parameters with more accuracy than the EM algorithm and Gibbs sampling by utilizing the i.i.d. posterior samples.

4.3. Algorithm Selection

There are several Bayesian model selection criteria to compare different models in fitting the same data set, here we take advantage of the idea of model selection to help determine which sampling algorithm is the “best”, the IBF algorithm or the Gibbs sampling. The deviance information criterion (DIC), proposed by Spiegelhalter et al. [30], was used as a measure of fit and complexity. The *deviance* is defined as

$$D(\theta) = -2\log L(\mathbf{y}|\theta) = -2l(\mathbf{y}|\theta),$$

where $L(\theta|\mathbf{y})$ is the likelihood function for FMNR model given by (2), and $l(\theta|\mathbf{y})$ is the log-likelihood. The DIC statistic is defined as

$$\text{DIC} = 2\bar{D} - \hat{D},$$

with $\bar{D} = \sum_{k=1}^K D(\theta^{(k)})/K$ and $\hat{D} = D(\hat{\theta})$. Here, $\{\theta^{(k)}\}_{k=1}^K$ are the posterior samples obtained by IBF sampler or Gibbs sampling, and $\hat{\theta}$ are the estimated posterior means.

In Bayesian framework, the Akaike information criterion (AIC) proposed by Brooks [31] and the Bayesian information criterion (BIC) proposed by Carlin and Louis [32], can be estimated by

$$\text{AIC} = \bar{D} + 2s, \quad \text{BIC} = \bar{D} + s\log(n),$$

where s is the number of the model parameters. We choose the algorithm with smaller DIC, AIC, and BIC as a better algorithm to fit the model.

In the tone data study, the estimated measures based on FMNR model are given in Table 5. This table tells us that for all the three criteria, the values based on IBF are slightly smaller than the Gibbs counterparts, which justifies the superiority of the IBF sampler over the Gibbs sampler, although this superiority is not significant.

Table 5. Model selection criteria for FHNR model with IBF sampler and Gibbs algorithm in the tone data analysis.

Algorithm	DIC	AIC	BIC
IBF	−268.9490	−259.5399	−235.4548
Gibbs	−268.4472	−259.2297	−235.1446

5. Discussion

This paper propose an effective Bayesian inference procedures for fitting FMNR models. With the in-complete data structure introduced by a group of multinomial distributed mixture component variables, a non-iterative sampling algorithms named IBF under a non-informative prior settings is developed. The algorithm combines the EM algorithm, inverse Bayesian formula and SIR algorithm to obtain i.i.d samples approximately from the observed posterior distributions of parameters. Simulation studies show that the IBF algorithm can estimate the parameters with more accuracy than the EM algorithm and Gibbs sampling, and runs much faster than Gibbs sampling. Real data analysis shows the usefulness of the methodology, also illustrates the advantages of the IBF sampler,

for example, the ergodic means of samples from IBF sampling converge much faster than the ones from Gibbs sampling, and density estimation based on i.i.d. IBF samples are more reliable than the asymptotically normal distribution from EM algorithm. The IBF sampling algorithm can be viewed as an important supplement of the classical EM, Gibbs sampling and other missing data algorithms. The procedure can be extended to other error distribution based finite mixture regression models, such as mixture Student's *t* regression. In this case, the Student's *t* distribution should be firstly represented as a mixture of normal distribution with the latent mixture variable following a gamma distribution, and the IBF algorithm should be carried out twice. In this paper, the number of component is known, when *g* is unknown, the Reversible-jump MCMC method (Green [33], Richardson and Green [34]) may be used to estimate *g*, and this point is under consideration. Future research directions may also be highlighted.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2227-7390/9/6/590/s1>.

Author Contributions: Conceptualization, S.A. and F.Y.; methodology, S.A. and F.Y.; investigation, S.A. and F.Y.; software, S.A.; writing-original draft preparation, S.A. and F.Y.; writing-review and editing, S.A. and F.Y.; Both authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation of Shandong province of China under grant ZR2019MA026.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The “tonedata” used in this paper are available from the free R package *mixtools* authored by Benaglia et al. [24].

Acknowledgments: The authors would like to express their gratitude to Benaglia et al. [24] for providing the “tonedata” in the free R package *mixtools*.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Quandt, R.; Ramsey, J. Estimating mixtures of normal distributions and switching regression. *J. Am. Stat. Assoc.* **1978**, *73*, 730–738. [[CrossRef](#)]
2. De Veaux, R.D. Mixtures of linear regressions. *Comput. Stat. Data Anal.* **1989**, *8*, 227–245. [[CrossRef](#)]
3. Jones, P.N.; McLachlan, G.J. Fitting finite mixture models in a regression context. *Aust. J. Stat.* **1992**, *34*, 233–240. [[CrossRef](#)]
4. Turner, T.R. Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *J. R. Stat. Soc. C-Appl.* **2000**, *49*, 371–384. [[CrossRef](#)]
5. McLachlan, J.; Peel, D. *Finite Mixture Models*, 1st ed.; Wiley: New York, NY, USA, 2000.
6. Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*, 1st ed.; Springer: New York, NY, USA, 2006.
7. Faria, S.; Soromenho, G. Fitting mixtures of linear regressions. *J. Stat. Comput. Simul.* **2010**, *80*, 201–225. [[CrossRef](#)]
8. Peel, D.; McLachlan, G.J. Robust mixture modelling using the *t* distribution. *Stat. Comput.* **2000**, *10*, 339–348. [[CrossRef](#)]
9. Song, W.X.; Yao, W.X.; Xing, Y.R. Robust mixture regression model fitting by Laplace distribution. *Comput. Stat. Data Anal.* **2014**, *71*, 128–137. [[CrossRef](#)]
10. Song, W.X.; Yao, W.X.; Xing, Y.R. Robust mixture multivariate linear regression by multivariate Laplace distribution. *Stat. Probab. Lett.* **2017**, *43*, 2162–2172.
11. Zeller, C.B.; Cabral, C.R.B.; Lachos, V.H. Robust mixture regression modelling based on scale mixtures of skew-normal distributions. *Test* **2016**, *25*, 375–396. [[CrossRef](#)]
12. Tian, Y.Z.; Tang, M.L.; Tian, M.Z. A class of finite mixture of quantile regressions with its applications. *J. Appl. Stat.* **2016**, *43*, 1240–1252. [[CrossRef](#)]
13. Yang, F.K.; Shan, A.; Yuan, H.J. Gibbs sampling for mixture quantile regression based on asymmetric Laplace distribution. *Commun. Stat. Simul. Comput.* **2019**, *48*, 1560–1573. [[CrossRef](#)]
14. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **1977**, *39*, 1–38.

15. Tan, M.; Tian, G.L.; Ng, K. A non-iterative sampling algorithm for computing posteriors in the structure of em-type algorithms. *Stat. Sin.* **2003**, *43*, 2162–2172.
16. Tan, M.; Tian, G.L.; Ng, K. *Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation*, 1st ed.; Chapman & Hall/CRC: New York, NY, USA, 2010.
17. Yuan, H.J.; Yang, F.K. A non-iterative Bayesian sampling algorithm for censored Student-t linear regression models. *J. Stat. Comput. Simul.* **2016**, *86*, 3337–3355.
18. Yang, F.K.; Yuan, H.J. A non-iterative posterior sampling algorithm for linear quantile regression model. *Commun. Stat. Simul. Comput.* **2017**, *46*, 5861–5878. [[CrossRef](#)]
19. Yang, F.K.; Yuan, H.J. A Non-iterative Bayesian Sampling Algorithm for Linear Regression Models with Scale Mixtures of Normal Distributions. *Comput. Econ.* **2017**, *45*, 579–597. [[CrossRef](#)]
20. Tsionas, M.G. A non-iterative (trivial) method for posterior inference in stochastic volatility models. *Stat. Probab. Lett.* **2017**, *126*, 83–87. [[CrossRef](#)]
21. Rubin, D.B. Comments on “The calculation of posterior distributions by data augmentation,” by M.A. Tanner and W.H. Wong. *J. Am. Statist. Assoc.* **1987**, *82*, 543–546.
22. Rubin, D.B. Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics*; Bernardo, J.M., DeGroot, M.H., Lindley, D.V.A., Simith, A.F.M., Eds.; Oxford University Press: Oxford, UK, 1988; Volume 3, pp. 395–402.
23. Cohen, E. The Influence of Nonharmonic Partial on Tone Perception. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1980.
24. Benaglia, T.; Chauveau, D.; Hunter, D.R.; Yong, D. mixtools: An R package for analyzing finite mixture models. *J. Stat. Softw.* **2009**, *32*, 1–19. [[CrossRef](#)]
25. Viele, K.; Tong, B. Modeling with mixtures of linear regressions. *Stat. Comput.* **2002**, *12*, 315–330. [[CrossRef](#)]
26. Hunter, D.R.; Young, D.S. Semiparametric Mixtures of regressions. *J. Nonparametr. Stat.* **2012**, *24*, 19–38. [[CrossRef](#)]
27. Young, D.S. Mixtures of regressions with change points. *Stat. Comput.* **2014**, *24*, 265–281. [[CrossRef](#)]
28. Wu, Q.; Yao, W.X. Mixtures of quantile regressions. *Comput. Stat. Data Anal.* **2016**, *93*, 162–176. [[CrossRef](#)]
29. Turner, R. Mixreg: Functions to Fit Mixtures of Regressions. R Package Version 0.0-6. 2018. Available online: <http://CRAN.R-project.org/package=mixreg> (accessed on 30 March 2018).
30. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B* **2002**, *64*, 583–639. [[CrossRef](#)]
31. Brooks, S.P. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *J. R. Stat. Soc. B* **2002**, *64*, 616–618.
32. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2000.
33. Green, P.J. Reversible jump markov chain monte carlo computation and Bayesian model determination. *Biometrika* **1995**, *82*, 711–732. [[CrossRef](#)]
34. Richardson, S.; Green, P.J. On bayesian analysis of mixtures with an unknown number of components. *J. R. Stat. Soc. B* **1997**, *59*, 731–784. [[CrossRef](#)]