Communications for Statistical Applications and Methods
2017, Vol. 24, No. 5, 457–480

https://doi.org/10.5351/CSAM.2017.24.5.457
Print ISSN 2287-7843 / Online ISSN 2383-4757

Bayesian analysis of random partition models with Laplace distribution

Minjung Kyung[1,a]

[a]Department of Statistics, Duksung Women's University, Korea

Abstract

We develop a random partition procedure based on a Dirichlet process prior with Laplace distribution. Gibbs sampling of a Laplace mixture of linear mixed regressions with a Dirichlet process is implemented as a random partition model when the number of clusters is unknown. Our approach provides simultaneous partitioning and parameter estimation with the computation of classification probabilities, unlike its counterparts. A full Gibbs-sampling algorithm is developed for an efficient Markov chain Monte Carlo posterior computation. The proposed method is illustrated with simulated data and one real data of the energy efficiency of Tsanas and Xifara (*Energy and Buildings*, 49, 560–567, 2012).

Keywords: Laplace mixture, model-based cluster, random partition model, Dirichlet process prior

1. Introduction

Clustering algorithms attempt to understand a partition of a finite set of objects into a potentially predetermined number of nonempty subsets; in addition, the number of partitions is often unknown beforehand. We focus on probability models for partitions and avoid purely algorithmic methods. As a special case, product partition models (PPMs), introduced by Hartigan (1990) and Barry and Hartigan (1992), are based on modeling random partitions of the sample space. These assume that observations in different elements of a random partition of the data are actually independent. So if the probability distribution for the random partitions is in a product form prior to obtaining observations, it is also then in product form after obtaining the observations (Jordan *et al*, 2007). In inference can therefore be made by conditioning on and averaging over partitions, with a random partition:

$$P(\rho_n = \{S_1, \ldots, S_k\}) = K \prod_{j=1}^{k} c(S_j),$$

where ρ_n is a partition of the objects in a family of subsets $S_1, S_2, \ldots, S_k$ of $S_0 = \{1, 2, \ldots, n\}$ and $c(S)$ is a non-negative *cohesion* that is specified for each subset of S_0. Here, the normalizing constant $K = \sum_{\rho \in \mathcal{P}} \prod_{j=1}^{|\rho|} c(S_j)$, where $\mathcal{P}$ is the set of all possible partitions into nonempty sets. Together with independent sampling across clusters, a PPM can be described as

$$\mathcal{P}(\mathbf{y}|\rho_n = \{S_1, \ldots, S_k\}) \propto \prod_{j=1}^{k} c(S_j) P(\mathbf{y}_{S_j}), \tag{1.1}$$

[1] Department of Statistics, Duksung Women's University, 33 Samyangro 144-gil, Seoul 01369, Korea.
E-mail: mkyung@duksung.ac.kr

Published 30 September 2017 / journal homepage: http://csam.or.kr

where $P(\mathbf{y}_{S_k})$ is the density for subcluster S_k and the cohesion is $c(S_j)$

Cohesion is the measure of the strength of the functional relationship of the elements in each subsets that then controls the partition of subsets that can be roughly thought of as a probability. A popular choice is $c(S) = m(|S|-1)!$ where m is a precision parameter and $|S|$ is the number of elements in S. It follows that the resulting probability model for ρ_n is

$$P(\rho_n) = \frac{m^{k-1}\prod_{j=1}^{k}\left(n_j-1\right)!}{\prod_{i=1}^{n}(m+i-1)}, \tag{1.2}$$

where $n_j = |S_j|$ is the number of elements in cluster j that is known as the Dirichlet process (DP) random partition (Blackwell and MacQueen, 1973). Details are in McCullagh and Yang (2007), Müller *et al.*(2015), Pitman (1996), Quintana and Iglesias (2003), and references therein.

A similarly popular prior on random partitions is model-based clustering and its extended models, which fit a finite mixture of multivariate Gaussian distributions with various variance structures to the data (Banfield and Raftery, 1993; Fraley and Raftery, 2002, 2007; McLachlan and Peel, 2000; Wolfe, 1970). It implements an Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) to obtain a local optimum of the log-likelihood. To select the best number of clusters, model selection criterion such as Bayesian information criterion (BIC) was employed after fitting several mixture models with different numbers of clusters.

In the Bayesian literature, the nonparametric Bayesian clustering approach is usually based on a mixture of the DP (Antoniak, 1974; Ferguson, 1973) and an unknown number of clusters. Especially, a Dirichlet process mixture (DPM) of regression models has been widely used as a flexible semiparametric approach for clustering and density estimation (Escobar and West, 1995). The implementation of the DP mixture models has been made feasible by the modern method of Bayesian computation and efficient algorithms (MacEachern and Müller, 1998; Neal, 2000). Product partition type priors on a normal mixture of regression model also have been widely used for the tractable, probability-based, objective function to identify good partitions (Booth *et al.*, 2008; Crowley, 1997; Quintana and Iglesias, 2003).

Recently, a natural extension of the random partition model has been considered with incorporating covariate values in its definition. MacEachern (1999, 2000) proposed a collection of dependent random probability measures with marginal distributions given by the DP. This idea has been extended and applied to the construction of various types of random probability measures such as the density regression (Dunson *et al.*, 2007; Tokdar *et al.*, 2010). A covariate-dependent extension was proposed by Müller *et al.*, (2011) and some alternative extensions to build covariate-dependent random partition models can be found in Park and Dunson (2010), and Argiento *et al.*, (2014). Airoldi *et al.*, (2014) also provided a general family of nonexchangeable species sampling sequences dependent on the realizations of a set of latent variables. Murua and Quintana (2017) recently provided the construction of a covariate-dependent prior distribution based on the Potts clustering model by covariate proximity in both the formation of clusters, and the prior predictive distributions for the multivariate multiple linear regression of the multivariate normal error.

Park and Dunson (2010) argued that a semiparametric Bayesian approach with an infinite number of clusters can be considered by letting $y_i \sim f(\phi_i)$, with $\phi_i \sim G$ and G assigned a DP prior. In marginalizing out G, a prior on the partition of subjects into clusters is formed with cluster-specific parameters consisting of independent draws from G_0 as base distribution in the DP. This prior is a type of PPM (Quintana and Iglesias, 2003) and it is appealing to marginalize out G in order to increase efficiency in computation and simplify interpretation. They also argued that the DP induces a

particular prior on the partition and one can develop alternative classes of PPMs by replacing the DP prior on G with an alternative choice such as species sampling models (Ishwaran and James, 2003; Pitman, 1996) which are a very broad class of nonparametric priors that include the DP as a special case.

Most of the mixture of the regression model are considered with a normality assumption for the distribution of subcluster $P(\mathbf{y}_{S_k})$. With normality assumption on the distribution of error in each cluster, various form of mean might be able to be estimated easily based on feasible computation and efficient algorithms. Instead in this research, we propose a Laplace distribution for the distribution of subclusters. Least absolute deviation (LAD) regression has been widely used in practice by assuming that the error terms follow a Laplace distribution. Because it is known that the least absolute value (LAV) estimator is statistically more efficient than the least squares estimator (normal error regression model) when disturbances come from heavy-tailed distributions such as non-normal stable distributions, the Laplace distribution or contaminated normal distribution (Dielman, 1984). He also argued that the asymptotic distribution of the LAV estimator is known under a fairly general set of assumptions, allowing for statistical inference in large samples. Details on the theoretical properties of LAD can be found in Dielman (1984, 2005).

Song *et al.*, (2014) recently proposed a robust estimation procedure for mixture linear regression models with error terms that follow a Laplace distribution. They argued that LAD regression has been widely used in practice to consider the impact of outliers. Outliers are known to impact more heavily on mixture linear regression models than on the usual linear regression models since the outliers affect the estimation of the regression parameters as well as totally blur the mixture structure. The estimation procedure of the EM algorithm has been studied using the fact that the Laplace distribution can be written as a scale mixture of a normal and a latent distribution.

In this research, we develop a full Bayesian estimation procedure for the linear regression mixture model of the full conditional posterior distribution with Laplace distribution. For the prior on the clustering structure, we consider a random partition model of the DP process based on a truncated approximation of stick-breaking priors (Ishwaran and Zarepour, 2000) because the proposed model leads to a tractable, probability-based, objective function to identify good partitions. For the full posterior distribution of Laplace distribution, we consider that the Laplace distribution is a scale mixture of a normal distribution with an exponential mixing density (Andrews and Mallows, 1974). Details are discussed in the following section.

We also apply a post process to posterior samples for parameters of the proposed model to choose a single clustering estimate to compromise the "label-switching" problem (Richardson and Green, 1997; Stephens, 2000). We follow Fritsch and Ickstadt (2009), which finds a single clustering estimate by maximizing the posterior expected adjusted Rand index with the posterior probabilities of two observations being clustered together.

We use hierarchical models and Gibbs sampling to obtain estimators for Laplace distribution mixture models. In Section 2, we consider the hierarchical structure of models and the basic identity of a scale mixture of a normal distribution for Laplace distribution. Section 3 provides details on Markov chain Monte Carlo (MCMC) procedures based on the full conditional distribution of parameters and post process based on the posterior similarity matrix to choose a single cluster and important oscillating functions in each curve based on the posterior expected adjusted Rand index. We compare the proposed Laplace regression mixture and the normal mixture in Section 4, using simulations and data sets. There is a discussion in Section 5.

2. Random partition model of Laplace distribution

We begin with construction of the random partition model with DP prior based on a Laplace linear regression. We discuss the mixture structure and the basic identity of the Laplace distribution which is a scale mixture of a normal distribution with an exponential mixing density.

2.1. Random partition model

We discussed in Section 1 that a PPM with a cohesion function $c(S) = m(|S|-1)!$ where m is a precision parameter and $|S|$ is the number of elements in S, is the DP random partition and the resulting probability model for the random partition is in (1.2).

Blackwell and MacQueen (1973) proved that for $Y_1, \ldots, Y_n$ iid from $G \sim \mathcal{DP}$, the joint distribution of $\mathbf{Y}$ is a product of successive conditional distributions of following form:

$$y_i|y_1, \ldots, y_{i-1}, m, \mu, \tau^2 \sim \frac{1}{i-1+m}\sum_{l=1}^{i-1}\delta(y_l = y_i) + \frac{m}{i-1+m} f\left(y_i|\mu, \tau^2\right), \tag{2.1}$$

where $f(y_i|\,\mu, \tau^2)$ is a probability density function: the base measure, and $\delta(\cdot)$ denotes the Dirac delta function. Quintana and Iglesias (2003) also show that the joint marginal distribution of (2.1) can be expressed as the PPM as

$$\begin{aligned}
\mathcal{P}(\mathbf{y}) &= \prod_{i=1}^{n}\left(\frac{1}{i-1+m}\sum_{l=1}^{i-1}\delta(y_l = y_i) + \frac{m}{i-1+m} f\left(y_i|\mu, \tau^2\right)\right) \\
&= \sum_{k=1}^{n}\frac{1}{\prod_{i=1}^{n}(m+i-1)}\prod_{j=1}^{k} m(n_j-1)!\left(\prod_{l=1}^{n_j} f\left(y_l|\mu, \tau^2\right)\right)\prod_{l=2}^{n_j}\delta(y_l = y_j) \\
&= K^*\sum_{k=1}^{n}\prod_{j=1}^{k} c(S_j) f_j(y_l),
\end{aligned} \tag{2.2}$$

where $f_j(y_l)$ is the density function of y_l, n_j is the sample size in cluster j, and $l \in S_j$, S_j is the subset of $S_0 = \{1, 2, \ldots, n\}$ for cluster j and K^* is the normalizing constant. This expression is known as the Blackwell and MacQueen (1973)'s Pólya urn representation of the DP.

The algorithms of Bush and MacEachern (1996) are some of the most widely-used approaches for the posterior computation of Pólya urn DP. They argued that their approach first updates the configuration of subjects to clusters based on the Pólya urn scheme in (2.1), and then separately updated cluster specific parameters given the cluster configuration with conjugate priors. Extension to non-conjugate priors is discussed by MacEachern and Müller (1998) and Neal (2000) based on Metropolis-Hastings. Park and Dunson (2010) also considered a generalized Pólya urn scheme based on a distance metric through a flexible nonparametric model for the joint distribution of the predictors.

Here, for the DP process prior, we consider the stick-breaking representation of the DP for the infinite number of clusters. According to Sethuraman (1994), if G is assigned a DP prior with precision m and base measure G_0, the stick-breaking representation of G is

$$G = \sum_{h=1}^{\infty} p_h \delta_{\theta_h}, \quad p_h = V_h\prod_{l<h}(1-V_l),\ V_h \sim \mathrm{Be}(1, m),\ \theta_h \sim G_0,$$

where δ_θ is a probability measure concentrated at θ and all V_h's and θ's are independent. Gibbs sampling methods for stick-breaking priors are provided in many articles.

Ishwaran and James (2001) presented two Gibbs sampling methods for fitting Bayesian nonparametric hierarchical models based on stick-breaking priors. The first type of Gibbs sampler, referred to as a Pólya urn Gibbs sampler, applies to stick-breaking priors with a known Pólya urn characterization, that are priors with an explicit and simple prediction rule. The second method, the blocked Gibbs sampler, works by directly sampling values from the posterior of the random measure. They argue that the blocked Gibbs sampler avoids marginalizing over the prior and allows the prior to be directly involved in the Gibbs sampling scheme. This allows direct sampling of the nonparametric posterior and leads to several computational and inferential advantages. Thus, in this paper, we consider the blocked Gibbs sampler of Ishwaran and James (2001) based on the stick-breaking representation of the DP as a prior on the clustering structure.

For the index of cluster, we consider an indicator variable of mixture Z_i for observation i, $i = 1, \ldots, n$. Then we re-express the model structure with the stick-breaking prior as

$$\begin{aligned} Y_i|X_i, Z_i, \{\boldsymbol{\theta}_h\}_{h=1}^{\infty} &\sim f\left(y_i|\mathbf{X}_i, \boldsymbol{\theta}_{Z_i}\right), \\ Z_i &\sim \sum_{h=1}^{\infty} p_h \delta_h, \\ p_h &= V_h \prod_{l<h} (1 - V_l), \end{aligned} \tag{2.3}$$

where $V_h \sim \text{Be}(1, m)$.

2.2. Basic identity

For the density function of y_i in cluster k, we consider a Laplace distribution such that

$$f(y_i|\mathbf{X}_i, \boldsymbol{\theta}_k) = \frac{1}{2b_k} \exp\left(-\frac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_k\right|}{b_k}\right), \tag{2.4}$$

where $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, b_k)$, $\boldsymbol{\beta}_k$ is a regression parameter of the location parameter, and b_k is a scale parameter.

The Laplace (double-exponential) distribution is a scale mixture of a normal distribution with an exponential mixing density (Andrews and Mallows, 1974), that is

$$\frac{a}{2} \exp(-a|z|) = \int_0^{\infty} \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{z^2}{2\tau}\right) \frac{a^2}{2} \exp\left(-\frac{a^2}{2}\tau\right) \mathrm{d}\tau. \tag{2.5}$$

Details of the equation and the proof have been discussed by Kyung *et al.* (2010). The main idea is to introduce appropriate latent parameters. We develop the posterior distribution of Laplace distribution parameters based on the equation in (2.5).

3. Sampling scheme

Ishwaran and Zarepour (2000) proposed a truncated approximation with $N < \infty$ to the DPM model in (2.3) to improve the mixing of its Gibbs sampler. They argued that the key to work with random probability measures of a truncated approximation is that it allows us to perform blocked updates for the probability $p_1, \ldots, p_N$ and $Z_1, \ldots, Z_N$ in (2.3). This then will result in a rapid mixing Markov

chain that permits a direct inference for the posterior of the random probability measure G. All the detailed derivation of the posterior distributions of $\mathbf{Z}$ and $\mathbf{p}$ and how to determine the truncation level N can be found therein. Therefore, we consider a truncated approximation with N instead of ∞ mixing properties for the hierarchical structure of a Laplace random partition model in (2.3) with the probability density function of (2.4).

For the regression parameters in cluster h, we consider the following priors:

$$\boldsymbol{\beta}_h \sim \text{MVN}_p\left(\mathbf{0}, c_h \boldsymbol{I}\right) \quad \text{and} \quad b_h^2 \sim \pi\left(b_h^2\right) \propto \frac{1}{b_h^2}.$$

We begin with construction of a cluster structure and discuss how to estimate parameters in each cluster.

Step 1. Cluster structure

With an appropriate approximation level of N, the subject-specific latent variable Z_i in (2.3) follows a discrete distribution with $\mathbf{p} = (p_1, \ldots, p_N)$. Then, the conditional posterior distribution of Z_i updated with observed $\mathbf{y}_i$ is specified as

$$\text{P}\left(Z_i = h | y_i, \mathbf{p}, \{\boldsymbol{\theta}_l\}_{l=1}^N\right) = \frac{p_h f\left(y_i | \mathbf{X}_i, \boldsymbol{\theta}_h\right)}{\sum_{l=1}^N p_l f\left(y_i | \mathbf{X}_i, \boldsymbol{\theta}_l\right)}.$$

Upon sampling $\mathbf{Z}$, the index set $S_h = \{i; Z_i = h\}$ for $h = 1, \ldots, N$ is also updated, inducing the cluster structure among n genes. Letting $n_h = n\left(S_h\right)$ be the cardinality of S_h, $(N-1)$ beta random variables, $V_1, \ldots, V_{N-1}$ can be updated by sampling from

$$V_h | \alpha, \mathbf{Z} \sim \text{Beta}\left(1 + n_h,\ m + \sum_{j=h+1}^{N} n_j\right),$$

for $h = 1, \ldots, N-1$, and it is set that $V_N = 1$ to ensure $\sum_{h=1}^N p_h = 1$ with $p_h = V_h \prod_{l<h}(1 - V_l)$ for $h = 1, \ldots, N$.

Step 2. Model parameters

Given the cluster indices of each observation $\mathbf{Z}$, the joint posterior distribution of cluster h based on a hierarchical model with priors can be written as

$$\begin{aligned}
&\pi\left(\boldsymbol{\beta}_h, \boldsymbol{\tau}_h, b_h^2 \middle| \mathbf{X}, \mathbf{Z}, by\right) \\
&\propto \left[\prod_{Z_i = h} (2\pi\tau_i)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\tau_i}\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right)^2\right\} \frac{1}{2b_h^2} \exp\left(-\frac{\tau_i}{2b_h^2}\right)\right] \frac{1}{b_h^2} \exp\left(-\frac{1}{2c_h}\boldsymbol{\beta}_h'\boldsymbol{\beta}_h\right),
\end{aligned}$$

where $\boldsymbol{\tau}_h = \{\tau_i : Z_i = h\}$ is a vector variances of observations in cluster h. Thus the full conditional posterior distribution of model parameters can be obtained based on data augmentation methods. For cluster h, the full conditional posterior distributions of parameters are

$$t_i \equiv \left.\frac{b_h \left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{\tau_i}\right| Z_i = \begin{cases}
\boldsymbol{\beta}_h | \boldsymbol{\tau}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim \text{MVN}_p\left(\boldsymbol{\beta}_h^*,\ \Sigma_{\boldsymbol{\beta}_h^*}\right) \\
h, \boldsymbol{\beta}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim \text{inverseGaussian}\left(\mu_i = 1,\ \lambda_i = \dfrac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{b_h}\right) \\
b_h^2 | \boldsymbol{\beta}_h, \boldsymbol{\tau}_h, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim \text{IG}\left(n_h, \dfrac{1}{2}\sum_{Z_i=1} \tau_i\right),
\end{cases}$$

where

$$\boldsymbol{\beta}_h^* = \left(\sum_{Z_i=h} \frac{1}{\tau_i}\mathbf{X}_i'\mathbf{X}_i + \frac{1}{c_h}\boldsymbol{I}\right)^{-1}\left(\sum_{Z_i=h} \frac{1}{\tau_i}\mathbf{X}_i' y_i\right), \qquad \Sigma_{\boldsymbol{\beta}_h^*} = \left(\sum_{Z_i=h} \frac{1}{\tau_i}\mathbf{X}_i'\mathbf{X}_i + \frac{1}{c_h}\boldsymbol{I}\right)^{-1},$$

and $n_h = |S_h|$ is the number of observations in cluster h. Details of derivation are in Appendix A.

Step 3. Post process

Mixture models suffer from a well-known "label-switching" problem, which arises due to the identical likelihood for any permutation of component-specific parameters. Inheriting the properties of the adjusted Rand index, Fritsch and Ickstadt (2009) proposed the posterior expected adjusted Rand index and showed the outperformance of the posterior expected adjusted Rand index over competing methods such as maximum a posteriori (MAP) estimate. In addition, implementing the method is easy with R package *mcclust*. Details are in Fritsch and Ickstadt (2009) and references therein.

4. Application

We conduct simulation studies to evaluate our proposed random partition model with LAD regression. We implement the full conditional Gibbs sampler using DP prior on the cluster structure to analyze the energy efficiency data set. As a competing model, we consider the normal mixture model (NMM) of DP prior. Also, to compare the proposed method, we consider the model-based clustering of Fraley and Raftery (2002) based on the NMM, for which the R package *mclust* (Fraley *et al.*, 2012) is available. Mixture models of normal distributions with various covariance structures are fitted via the EM algorithm. The NMM was implemented with no specific prior distribution, which was the default setting provided by the R package *mclust*.

For the simulation studies, we considered the maximum truncation level is 30 ($N = 30$) to perform blocked updates for the probability $p_1, \ldots, p_N$ and $Z_1, \ldots, Z_N$ in (2.3). An MCMC algorithm also ran for 50,000 iterations with a burn-in period of 20,000. We collected every 10^{th} sampler among 30,000 iterations to prevent the correlation of Gibbs. With 3,000 Gibbs sampler, we conducted the post process. With the unknown number of clusters, for the Gibbs sampling, we consider the posterior expected adjusted Rand index of Fritsch and Ickstadt (2009) in R package *mcclust* to choose the optimal number of clusters and the indices of clusters. For the model-based clustering, the BIC is used to identify the optimal number of clusters and covariance structure for a given data set, and a MAP estimate is obtained.

4.1. Simulation study I

We first evaluated the performance of our method with simulated data, where the classes are known. We simulated data according to the following regression models with $n = 300$ and $k = 3$

$$Y_{ih} = \mathbf{X}_i\boldsymbol{\beta}_h + \epsilon_{ih},$$

where $i = 1, \ldots, n$ and $h = 1, \ldots, k$. We considered three clusters ($k = 3$) and the cluster indicator Z_i follows

$$Z_i \sim \text{Multinomial}\,(1, \mathbf{p} = (0.3,\ 0.3,\ 0.4))\,.$$

For regression, we generated two exploratory variables, X_1 from $N(-3, 0.01)$ and X_2 from $N(2, 0.01)$. We set a design matrix as $\mathbf{X} = (1, X_1, X_2)$. We fix the regression parameters in each cluster as:

$$\text{Cluster } 1 : \boldsymbol{\beta}_1 = (0, 0, 2), \quad \text{Cluster } 2 : \boldsymbol{\beta}_2 = (-1, 0, -2), \quad \text{Cluster } 3 : \boldsymbol{\beta}_3 = (1, 1, 0).$$

For various situation of data structure, we considered three different sets of errors for each clusters:

Set 1. $\epsilon_{i1} \sim N(0, 0.5)$, $\epsilon_{i2} \sim N(0, 0.2)$ and $\epsilon_{i3} \sim N(0,\ 0.1)$

Set 2. $\epsilon_{i1} \sim \text{Laplace}(0,\ \sqrt{0.5})$, $\epsilon_{i2} \sim \text{Laplace}(0,\ \sqrt{0.2})$ and $\epsilon_{i3} \sim \text{Laplace}(0,\ \sqrt{0.1})$

Set 3. $\epsilon_{ih} \sim t\,(\text{df} = 5)$ for $i = 1, \ldots, n$ and $h = 1, \ldots, k$.

For normally distributed error data (Set 1), means of cluster $\mathbf{X}\boldsymbol{\beta}_h$ are set to be well separable such that $\mu_1 = \mathbf{X}\boldsymbol{\beta}_1 \approx 4$, $\mu_2 = \mathbf{X}\boldsymbol{\beta}_2 \approx -5$, and $\mu_3 = \mathbf{X}\boldsymbol{\beta}_3 \approx -2$, and the true variances are small numerically as $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.2$, and $\sigma_3^2 = 0.1$.

For the model-based normal mixture, to identify the optimal number of clusters and covariance structure for a given data set, the BIC is considered and the BIC plots of each data sets for the number of clusters are in Figure C.1 in Appendix C. By the R package *mclust*, the mixture models of normal distributions with various covariance structures are considered via the EM algorithm and the BIC are computed. In our simulation studies, for all data sets, the BIC has chosen only one cluster with most of multivariate covariance structures except "spherical, equal volume (EII)" and "spherical, varying volume (VII)" structures. It might be reason for large scale parameter values of each data sets for each cluster, or the limitation of the BIC computation based on the EM algorithm. We already know the number of clusters as 3 for each sets of data. Therefore, for the comparison, we consider the EM based MAP estimation of 3 clusters with spherical and varying volume (VII) covariance.

Based on the posterior mean and 95% credible interval, the proposed model correctly estimates the mean functions, but it fails to capture the linear trends correctly. However, the estimated μ_h's are numerically close to the true values. Based on the post process, we compute the adjusted Rand index between the Laplace regression partition model and the true cluster index of the same objects, and Z_i's from our proposed model are perfectly matched to the true indices of clusters. NMM also shows similar results to the proposed model. The estimated curves with true curves are in Figure 1. Both the Laplace regression partition model and the normal regression mixture model estimated the true mean curve adequately. The EM based Gaussian mixture models also estimate the mean of each clusters close to the true values and the computed adjusted Rand index shows that the measured classification index of the EM perfectly matched to true indices of clusters.

Table 1 shows the estimated scale parameters and 95% credible intervals of both Laplace partition models and NMMs. Estimated scale parameters of the proposed Laplace regression partition model are $\hat{b}_1 = 0.49(0.34.0.98)$, $\hat{b}_2 = 0.37(0.27, 0.95)$, and $\hat{b}_3 = 0.24(0.17, 0.93)$, and these are numerically similar to the standard deviation of true models for each clusters. The estimated standard deviation of the NMM are $\hat{\sigma}_1 = 0.58(0.47, 0.84)$, $\hat{\sigma}_2 = 0.44(0.36, 0.58)$, and $\hat{\sigma}_3 = 0.30(0.25, 0.54)$ of cluster 1, 2, and 3. Based on the 95% credible intervals of the proposed Laplace regression partition model, we observe that the posterior distributions of scale parameters are skewed to right and have a wider credible interval than NMM. The posterior distributions of the standard deviations are symmetric and have shorter credible intervals. The estimated standard deviation of the EM NMM are $\hat{\sigma}_1 = 0.28$, $\hat{\sigma}_2 = 0.22$, and $\hat{\sigma}_3 = 0.15$, which are almost half values of the estimated mean standard deviation of the Gibbs normal mixtures.

For Laplace random partition data (Set 2), means of cluster $\mathbf{X}\boldsymbol{\beta}_h$ are also set to be well separable such that $\mu_1 = \mathbf{X}\boldsymbol{\beta}_1 \approx 4$, $\mu_2 = \mathbf{X}\boldsymbol{\beta}_2 \approx -5$, and $\mu_3 = \mathbf{X}\boldsymbol{\beta}_3 \approx -2$, but the true scale parameters are not

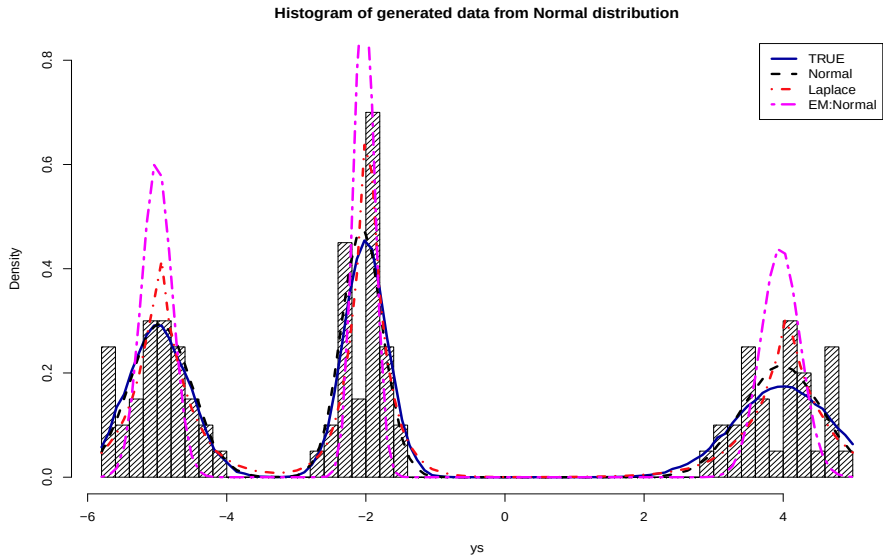


Figure 1: *Histogram of generated data from normal mixture model (Set 1). The cluster-specific mean curves (black solid lines) with two estimated curves by the Dirichlet process normal mixture model (dotted blue lines), the model-based normal mixture EM model (dot-dashed magenta lines), and the Laplace regression random partition model (red dot lines) are on histogram.*

Table 1: Posterior median and 95% CI for cluster-specific residual scale parameter σ_h or b_h of NMM and LPM, and MAP estimate from the model-based clustering of EM

Model	Cluster	Truth	NMM		LPM		MAP
			Mean	95% CI	Mean	95% CI	
Normal	1	0.71	0.58	(0.47, 0.84)	0.49	(0.34, 0.98)	0.28
	2	0.45	0.44	(0.36, 0.58)	0.37	(0.27, 0.95)	0.22
	3	0.32	0.30	(0.25, 0.54)	0.24	(0.17, 0.93)	0.15
Laplace	1	0.71	0.73	(0.45, 3.27)	0.49	(0.26, 1.00)	0.43
	2	0.45	0.41	(0.19, 2.46)	0.47	(0.18, 1.67)	0.36
	3	0.32	0.60	(0.20, 5.38)	0.85	(0.11, 2.71)	0.17
	4	-	0.36	(0.19, 8.08)	0.63	(0.20, 2.35)	
	5	-	-	-	0.21	(0.06, 1.41)	
t (df = 5)	1	1.67	0.84	(0.57, 11.80)	0.40	(0.25, 3.12)	0.73
	2	1.67	2.75	(2.12, 25.80)	1.03	(0.61, 5.20)	0.66
	3	1.67	-	-	1.35	(0.57, 4.24)	0.30

CI = credible interval; NMM = normal mixture model; LPM = Laplace partition model; MAP = maximum a posteriori.

small numerically to be separable clusters clearly, $b_1 = 0.71$, $b_2 = 0.45$, and $b_3 = 0.32$. Therefore, there might be the grey zone (which is a subregion that is not separable clearly as different clusters).

The estimated number of cluster is 5 based on our proposed model and 4 based on the NMM. Thus, the computed adjusted Rand index between the estimated Z_i and the true index is 0.83 and 0.89 for our model and NMM, separately. The Laplace regression random partition model seems to detect partition sensitively for the grey zone between partition 2 and 3, compared to the normal regression mixture model. Figure 2 shows the estimated curves with true curves. Regardless of the estimated number of clusters, the estimated curves are quite close to the true curve for both Laplace and normal

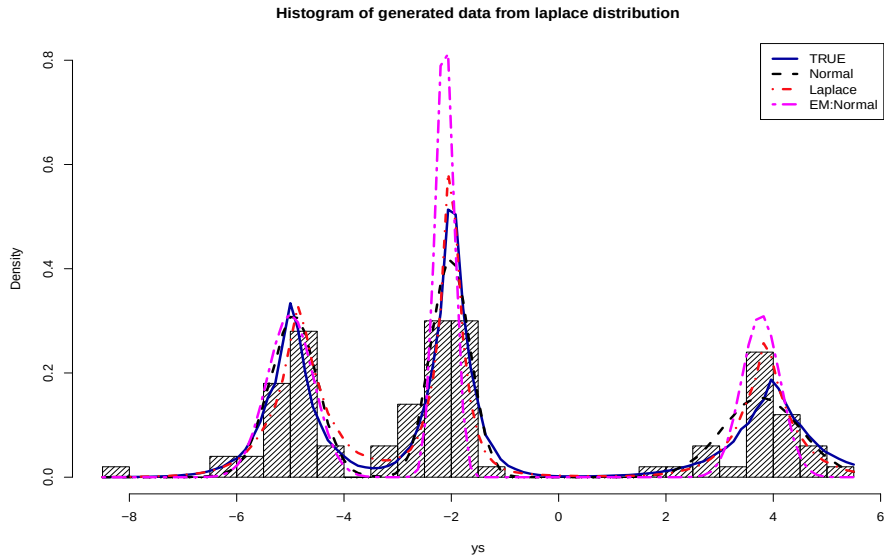


Figure 2: *Histogram of generated data from Laplace random partition model (Set 2). The cluster-specific mean curves (black solid lines) with two estimated curves by the normal mixture model (dotted blue lines) and the Laplace regression random partition model (red dot lines) are on histogram.*

regression models.

The proposed model correctly estimates the mean functions based on the posterior mean and 95% credible interval; however, it fails to capture the linear trends correctly and similar results of the NMM. Table 1 shows that the estimated scale parameters and 95% credible intervals of the proposed Laplace regression partition model are $\hat{b}_1 = 0.49(0.26, 1.00)$, $\hat{b}_2 = 0.47(0.18, 1.67)$, $\hat{b}_3 = 0.85(0.11, 2.71)$, $\hat{b}_4 = 0.63(0.20, 2.35)$, and $\hat{b}_5 = 0.21(0.06, 1.41)$, and the estimated median standard deviation of the NMM are $\hat{\sigma}_1 = 0.73(0.45, 3.27)$, $\hat{\sigma}_2 = 0.41(0.19, 2.46)$, $\hat{\sigma}_3 = 0.60(0.20, 5.38)$, and $\hat{\sigma}_4 = 0.36(0.19, 8.08)$. For the scale parameters, credible intervals of standard deviations of normal regression mixture models are quite wider compared to the credible intervals of scale parameters of Laplace regression partition models. The posterior distributions of the standard deviations are also highly skewed right. Data set is generated from the Laplace random partition models; however, it might be a reason for an unstable estimation of the standard deviation in NMMs.

With EM algorithm of fixed 3 clusters, the estimated scale parameters are $\hat{\sigma}_1 = 0.43$, $\hat{\sigma}_2 = 0.36$, and $\hat{\sigma}_3 = 0.17$. There might exist underestimation problem even with known number of clusters for the EM. The computed adjusted Rand index between the estimated Z_i and the true index is 0.91. As discussed above, it might be the reason for data generation setting and the forced separation to 3 clusters. We also observe that the estimated curves are quite close to the true curve based on the estimated curve in Figure 2.

For t (df = 5) (Set 3), means of cluster $\mathbf{X}\boldsymbol{\beta}_h$ are also set to be well separable such that $\mu_1 = \mathbf{X}\boldsymbol{\beta}_1 \approx 4$, $\mu_2 = \mathbf{X}\boldsymbol{\beta}_2 \approx -5$, and $\mu_3 = \mathbf{X}\boldsymbol{\beta}_3 \approx -2$, but with df = 5, t-distribution has heavy tails. Therefore, the generated data set might not be well separable.

The estimated number of cluster is 3 based on our proposed model and 2 based on the NMM. Therefore, the computed adjusted Rand index between the estimated Z_i and the true index is 0.50 and

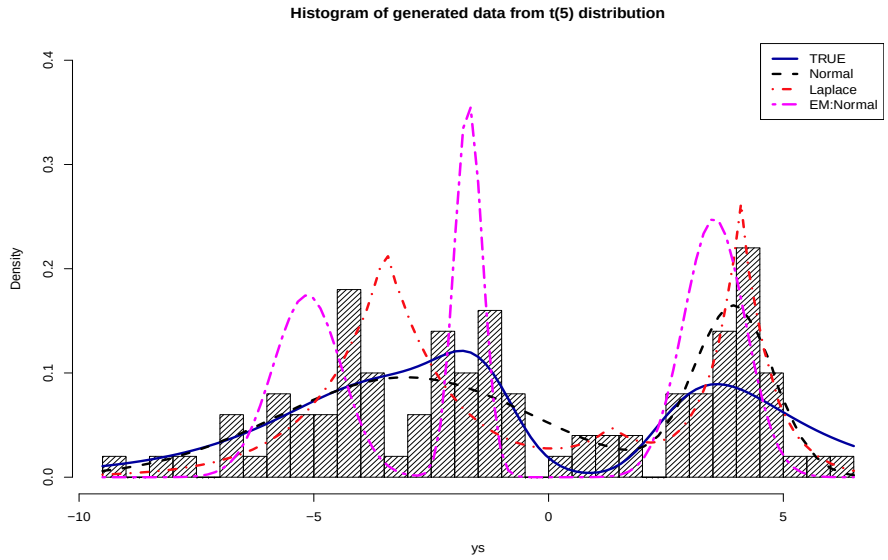


Figure 3: *Histogram of generated data from t mixture model of df = 5 (Set 3). The cluster-specific mean curves (black solid lines) with two estimated curves by the normal mixture model (dotted blue lines) and the Laplace regression random partition model (red dot lines) are on histogram.*

0.47 for our model and NMM, separately. The Laplace regression random partition model seems to detect partition sensitively compared to the normal regression mixture model. However, the computed adjusted Rando index is 0.70 for the EM Gaussian mixture model. For the estimated mean of each cluster μ_h, we observe that the NMM combine cluster 2 and 3, then estimates the mean of cluster 1 as around 4 $\hat{\mu}_1 \approx 4$ and of cluster 2 as around -4 $\hat{\mu}_2 \approx -4$, the mean of the true means of cluster 2 and 3. However, the Laplace random partition model estimate the mean of cluster 1 as around 4 $\hat{\mu}_1 \approx 4$; however, cluster 2 is around -3.5 $\hat{\mu}_2 \approx -3.5$ and around 1.3 for cluster 3 $\hat{\mu}_3 \approx 1.3$. From the histogram of generated data in Figure 3, we observe that there is a group of data between 0 and 3, and the distribution of negative valued data are skewed left. This distribution might be because the true distribution seems to have two clusters with a normal regression mixture model that estimates the number of parameters as 2. The proposed Laplace random partition model also seems to partition a subset of data between 0 and 3 as a different cluster due to the distribution of generated data. However, with fixed 3 clusters, the estimated curve of the EM on the histogram seems not to consider the distribution of data at all even though the computed adjusted Rando index is 0.70. Thus, if our goal is density estimation, we should better use the mixture models with Gibbs sampling, but the EM will provide more hidden information if our goal is the detection of the cluster indices.

The proposed model and the NMM correctly estimates the mean functions based on the posterior mean and 95% credible interval; however, it fails to capture the linear trends correctly. Figure 3 includes the estimated curves with true curves. With heavy tailed mixture, we observe that any method might be unable to capture the true clustering structure in data. Estimated scale parameters and 95% credible intervals in Table 1 of the proposed Laplace regression partition model are $\hat{b}_1 = 0.40(0.25, 3.12)$, $\hat{b}_2 = 1.03(0.61, 5.20)$, and $\hat{b}_3 = 1.35(0.58, 4.24)$. However, the estimated standard deviations of the NMM $\hat{\sigma}_1 = 0.84(0.57, 11.80)$ and $\hat{\sigma}_2 = 2.75(2.12, 25.80)$ are unstable and

Table 2: Computed the adjusted Rand index between partitions and the true indices of clusters of the same objects for NMM and LPM, and MAP estimate from the model-based clustering model of EM

	Set 4			Set 5		
	NMM	LPM	MAP	NMM	LPM	MAP
Rand index	0.467	0.434	0.00	0.427	0.401	0.102

NMM = normal mixture model; LPM = Laplace partition model; MAP = maximum a posteriori.

highly skewed. True values are included in the credible intervals of scale parameters of Laplace partition models; however, the true values are not included in the credible intervals of NMM because it estimated the number of cluster as 2 with large value of standard deviation. The estimated standard deviations of the EM are $\hat{\sigma}_1 = 0.73$, $\hat{\sigma}_2 = 0.66$, and $\hat{\sigma}_3 = 30$, and these are underestimated.

4.2. Simulation study II

For more complicated structure of the data generation process with clusters, we generated two more data sets to evaluate our proposed random partition model with LAD regression. We simulated data according to the following regression models with $n = 400$ and $k = 2$

$$Y_{ih} = \mathbf{X}_i \boldsymbol{\beta}_h + \epsilon_{ih}$$

where $i = 1, \ldots, n$ and $h = 1, \ldots, k$. We considered two clusters ($k = 2$) and the cluster indicator Z_i follows

$$Z_i \sim \text{Multinomial}\,(1, \mathbf{p} = (0.6, 0.4))\,.$$

For regression, we generated two exploratory variables, X_1 from $N(0,\ 1)$ and X_2 from $N(0,\ 1)$ and set a design matrix as $\mathbf{X} = (1, X_1, X_2)$. The fixed regression parameters in each clusters are:

$$\text{Cluster 1}: \ \boldsymbol{\beta}_1 = (0, 1, 1), \qquad \text{Cluster 2}: \ \boldsymbol{\beta}_2 = (0, -1, -1).$$

We mimic the 5^{th} and 6^{th} settings of the simulation studies in Song *et al.* (2014) and these are;

Set 4. $\epsilon_{ih} \sim 0.95N(0,\ 1) + 0.05N(0,\ 25)$ for $h = 1, 2$

Set 5. $\epsilon_{ih} \sim N(0,\ 1)$ with 5

The error in Set 4 is a mixture of two normal distributions and this complexity causes the generated data to appear to have at least four clusters and not easy to partition. This would produce 5% data likely to be low leverage outliers and unsmooth curved data. Based on the posterior mean and 95% credible interval of parameters in each clusters, we observe that the estimate fails to capture the linear trends correctly.

The true number of cluster is 2, and the NMM of Gibbs choose 2 clusters and the Laplace partition model of Gibbs have chosen 5 clusters based on the posterior expected adjusted Rand index. The BIC of the EM model based-cluster consider 2 to 3 clusters with "spherical, varying volume (VII)" variance structure, but we choose to have 2 clusters. Based on the chosen number of clusters of models, the computed adjusted Rand index between clusterings/partitions and the true indices of the clusters are in Table 2. We observe that the computed value of adjusted Rand index with true indices of the Gibbs NMM is slightly larger than that of the Gibbs Laplace partition model numerically. However, the EM model-based model seems not to detect true clustering indices correctly compared to other Gibbs models.

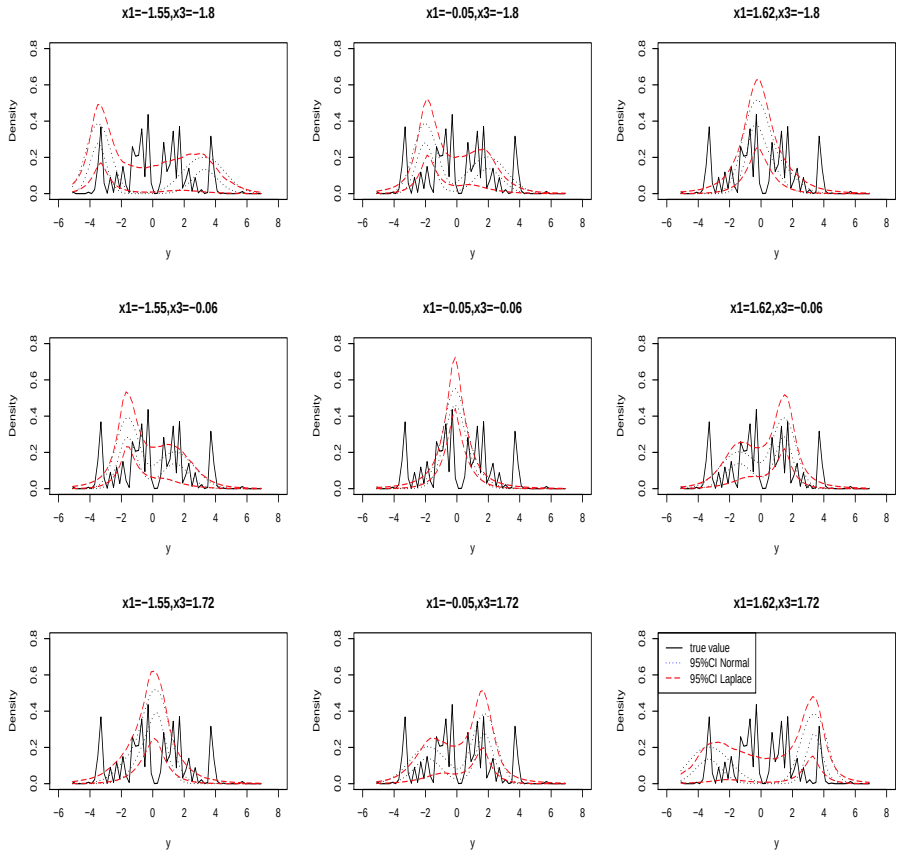


Figure 4: *Estimated curves based on normal mixture model (bold black line) and on Laplace random partition model (bold blue line) with 95% credible intervals of normal mixture (dotted black lines) and of Laplace random partition model (dotted blue lines) for Set 4 data at selected data points.*

Figure 4 shows the estimated curves of NMM and LPM with 95% credible intervals on selected data. Based on the 95% credible intervals, we observe that the 95% credible interval of LPM is wider than NMM as discussed in the previous section. The estimated curves do not seem to adequately estimate the true curve at each data point due to the complexity of data generation setting. However, the true number of clusters is 2, and the estimated curves of NMM and LPM seem to capture the true number of clusters around the data points that can be easily partitioned.

In the generating setting of the 5^{th} data (Set 5), 5% of the observations are replicated serving as high leverage outliers, used to check the robustness of estimation procedures against the high leverage outliers. The BIC of the EM model based-cluster consider 2 to 3 clusters "spherical, equal volume (EII)" variance structure; however, we choose to have 3 clusters because of the 5% high leverage outliers. The NMM and the Laplace partition model of Gibbs choose 4 clusters based on the posterior expected adjusted Rand index. From the computed adjusted Rand index in Table 2, we observe that the computed value of adjusted Rand index of the Gibbs normal is a little larger than it of the Gibbs Laplace partition model. It is unexpected that our proposed method performs no better than the Gibbs

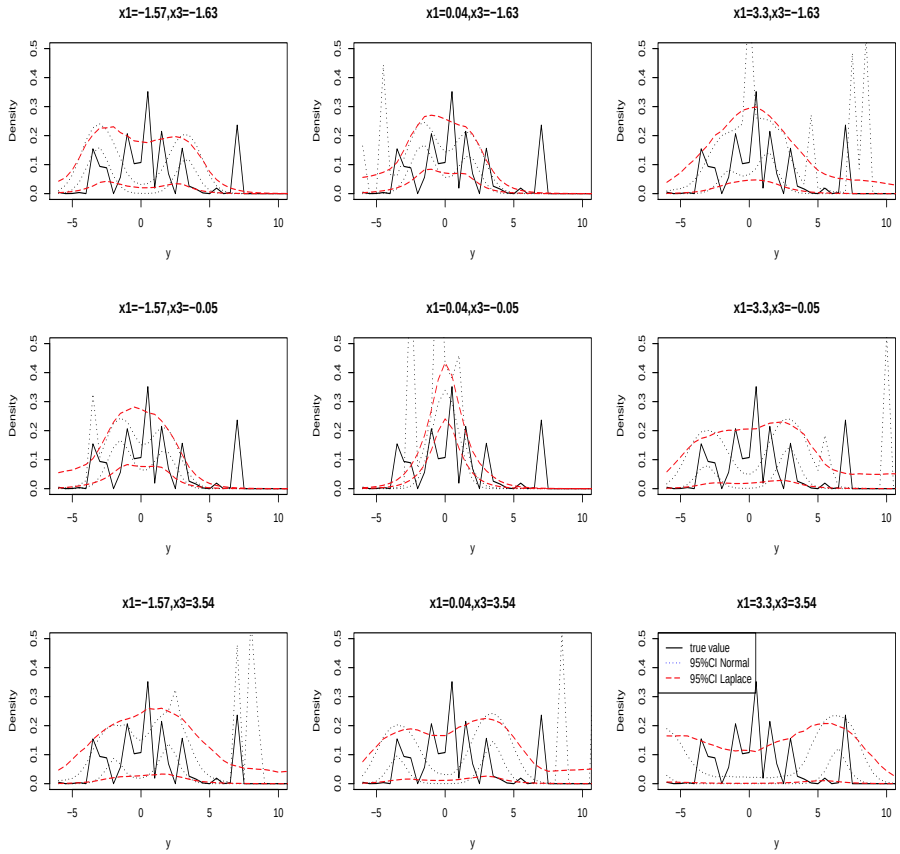


Figure 5: *Estimated curves based on normal mixture model (bold black line) and on Laplace random partition model (bold blue line) with* 95% *credible intervals of normal mixture (dotted black lines) and of Laplace random partition model (dotted blue lines) for Set 5 data at selected data points.*

NMM. We observe that the estimated curve of the Gibbs normal mixture does not capture the true curve correctly from the estimated curves in Figure 5. However, the 95% credible interval of the proposed Gibbs Laplace partition model seems to adequately estimate for the hidden structure of the data with high leverage outliers, even though the 95% credible interval is wider.

4.3. Energy efficient data analysis

The energy efficient dataset was created and processed by Tsanas and Xifara (2012) using 12 different building shapes simulated in Ecotect. The buildings differ with respect to glazing area, glazing area distribution, and orientation, amongst other parameters. They originally simulate various settings as functions of the afore-mentioned characteristics to obtain 768 building shapes and the dataset comprises 768 samples and 8 features, aiming to predict two real valued responses. Two responses are "Heating Load" (Y_1) and "Cooling Load" (Y_2), and eight attributes are relative compactness (X_1), surface area (X_2), wall area (X_3), roof area (X_4), overall height (X_5), orientation (X_6), glazing area (X_7), and glazing area distribution (X_8). Correlations between explanatory variables are very strong among

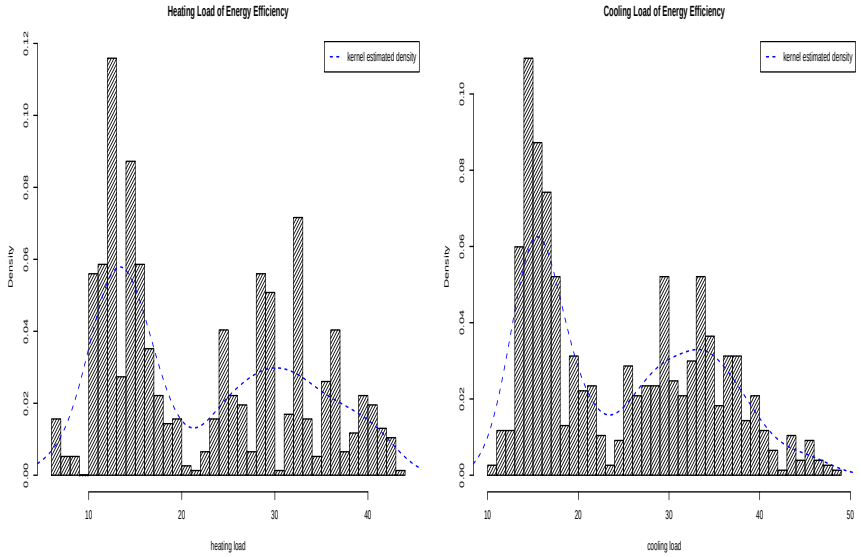


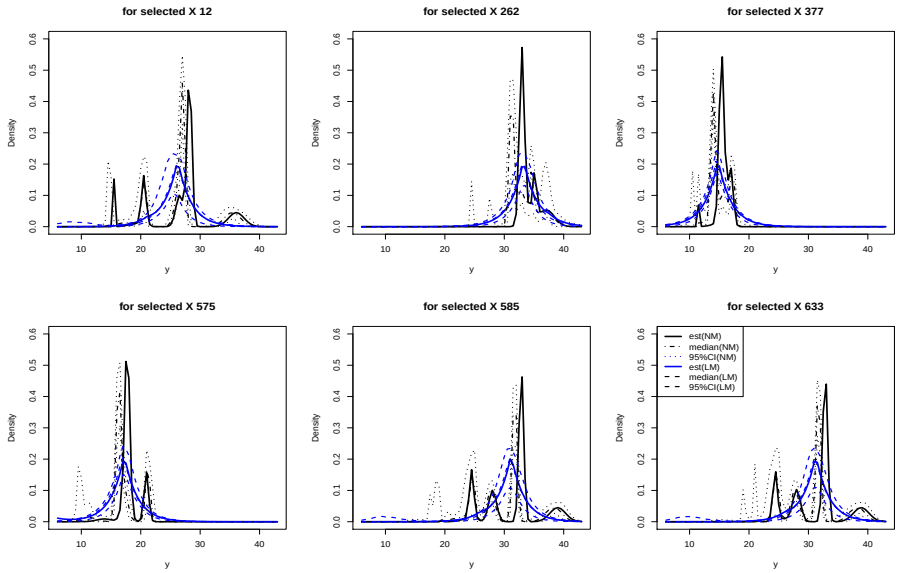
Figure 6: *Histograms of heating load and cooling load with estimated curve of responses based on kernel method.*

X_1, X_2, X_4, and X_5; however, there is no relationship with X_6, X_7, and X_8. In addition, there exist mild correlation between X_3 and (X_1, X_2, X_4, X_5), and between X_7 and X_8.

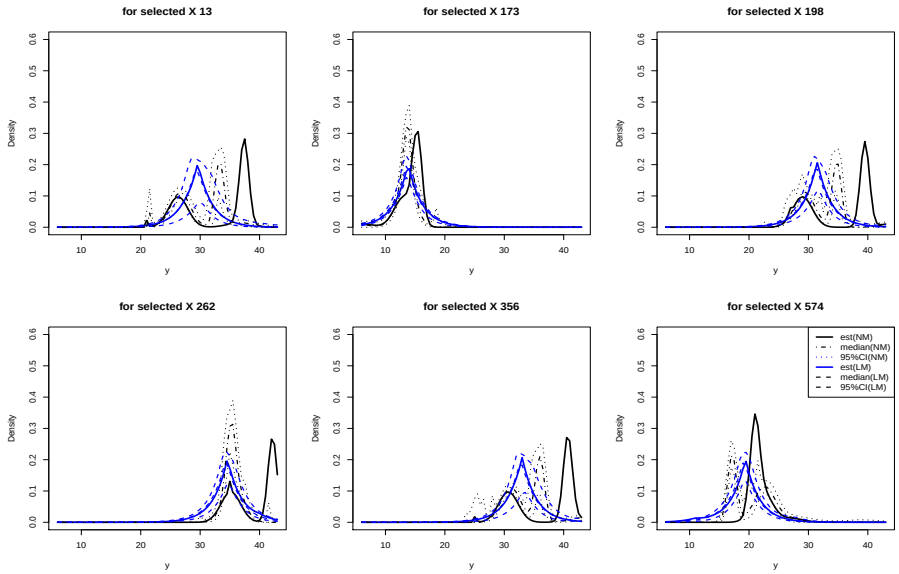
Tsanas and Xifara (2012) investigated the association strength of each input variable with each of the output variables using a variety of classical and non-parametrical statistical analysis tools to identify the most strongly related input variables. They compared a linear regression approach and random forests to estimate heating load (HL) and cooling load (CL). Tsanas and Xifara did not considered standardization and intercept in the model for the linear regression and random forest models. Tsanas and Xifara concluded that based on the random forest, X_7 (glazing area) is the most important predictor for both HL and CL and similar interpretation for a regression model. However, X_7 variable varies from 0 to 0.4 and the observed X_2 variable is in $(514.5,\ 808.5)$. It might be a reason why the estimated impact of X_7 is larger than other variables.

We instead consider a normal regression mixture model and a Laplace regression random partition model for HL and CL because a simple linear regression is inadequate to explain the relationship of input variables to output variable. Figure 6 includes histograms of HL and CL with estimated density based on Gaussian kernels. We observe from histograms that the HL might be able to explained with mixture of few normal distributions, and the CL can be explained with one normal distribution with small variance and one normal distribution with large variance.

Estimated cluster-specific parameters of normal linear regression mixture model and Laplace linear regression random partition model are in Table B.2 in Appendix B. The estimated number of clusters of NMM is 6 and of Laplace partition model is 2. For NMM, cluster 1 is specified with positive parameters of X_5 and X_8, cluster 2 is with no input variables, cluster 3 is with negative parameter of X_1 and positive parameters of X_5 and X_7, cluster 4 is very similar with cluster 3, cluster 5 is positive parameters of X_5 and X_7, and cluster 6 is with X_7. Tsanas and Xifara (2012) explained that X_7 (glazing area) is the most important predictor for HL, and it is in NMM of cluster 3 and 4. However, the impact of X_7 is the same as X_5 (orientation) in cluster 5. There are also clusters which do not detect X_7 as an



(a) Estimated curves of Heating load on selected data points



(b) Estimated curves of Cooling load on selected data points

Figure 7: *Estimated curves based on normal mixture model (bold black line) and on Laplace random partition model (bold blue line) with 95% credible intervals of normal mixture (dotted black lines) and of Laplace random partition model (dotted blue lines) for heating and cooling load at selected data points.*

impact input, and 95% credible intervals are significantly skewed left. With the number observations in each clusters, we observe that most of HL can be explained with X_5 and X_7.

The estimated number of clusters of Laplace random partition model is 2, and cluster 1 is specified

with positive parameters of X_5 and X_7 and cluster 2 is with positive parameter of X_7 only. We also observed that the X_7 input variable is the most important variable to explain HL. Even though the observed values of X_7 is smaller than other variables, in cluster 1, we observe that the 95% credible interval is not wide to suspect the impact of the small observed values. We select few of data points and plot curves of HL and CL in Figure 5 that compare the estimated curves of both the NMM and Laplace random partition model. The estimated curves of selected NMM data points show many bumps compared to the estimated curves of Laplace random partition model due to the estimated number of clusters and cluster-specific parameter estimation. The estimated curves of HL is in the upper part of Figure 7.

For the Cooling Load, estimated cluster-specific parameters of the normal linear regression mixture model and Laplace linear regression random partition model are in Table B.1 in Appendix B. The estimated number of clusters of NMM is 4 and of Laplace partition model is 2. For Laplace random partition model, cluster structure on CL is similar and like clusters on HL. For NMM, cluster 1 is specified with X_5, cluster 2 and 4 are with X_5 and X_7, but cluster 3 is with no significant input variables. The estimated curves of the normal mixture and Laplace partition models on selected data points are in Figure 7.

Unlike the conclusion of Tsanas and Xifara (2012), we observe that X_5 (orientation) is also an important variable to explain HL and CL with X_7 (glazing area). Tsanas and Xifara (2012) argued that the most important variable (glazing area) is not the most correlated with either output variable and other input variables. It can also be intuitively understood that the glazing area is of paramount significant to determine the energy performance of buildings. However, we argue that there are various cluster structures to explain HL and CL with significant input variables in each cluster. Therefore, various linear combinations of orientation and glazing area are important elements to determine the energy performance of buildings because the amount of glazing and the orientation of buildings determine that the heat absorbed in a building due to the sun as well as a similar orientation and glazing is a source of heat leakage from the building to the environment.

5. Discussion

We have developed a random partition procedure based on a DP prior with Laplace distribution. A full Gibbs-sampling algorithm for the linear regression mixture model of the full conditional posterior distribution with Laplace distribution is developed for an efficient MCMC posterior computation. For the prior on the clustering structure, we consider a random partition model of the DP, because the proposed model leads to a tractable, probability-based, objective function to identify good partitions. For the full posterior distribution of the Laplace distribution, we consider the fact the Laplace distribution is a scale mixture of a normal distribution with an exponential mixing density (Andrews and Mallows, 1974). We also have applied a post process to posterior samples for parameters of the proposed model to choose a single clustering estimate to compromise the "label-switching" problem based on maximizing the posterior expected adjusted Rand index of Fritsch and Ickstadt (2009).

For the comparison of the proposed methods, we considered the model-based clustering, Gaussian mixture model, based on the EM methods in our simulation studies. To choose the optimal number of clusters, we considered the BIC values on each sets. However, in our simulation studies, strangely for all data sets, the BIC has chosen only one cluster with most of multivariate covariance structures except "spherical, equal volume (EII)" and "spherical, varying volume (VII)" structures. It might be the reason of large scale parameter values of each sets of data for each clusters or the limitation of the BIC computation based on the EM algorithm. In the simulation, we already know the number

of clusters and we fixed the number of clusters as 3 that is the true number of clusters for the data generation.

For the first set of simulations, we considered three different sets of error distributions, normal, Laplace, and t with df = 5. Based on the posterior mean and 95% credible interval, the proposed model and the NMM correctly estimates the mean functions, but it fails to capture the linear trends correctly for all sets of generated data. With light tailed error such as Laplace distribution, for the scale parameters, credible intervals of standard deviations of normal regression mixture models are quite wider compared to the credible intervals of scale parameters of Laplace regression partition models. The posterior distributions of the standard deviations are also highly skewed right. However, with heavy tailed errors such as t distribution with df = 5, we observed that in the credible intervals of scale parameters of Laplace partition models, true values are included, but in the credible intervals of NMM, the true values are not included because it estimated the number of clusters smaller with a large value of standard deviation than the true number of clusters.

The two data sets in the second simulation section were with the 5% low leverage outliers and with the 5% high leverage outliers, respectively. The NMM and even the EM model based clustering algorithm failed to capture the linear trends correctly in the proposed model; in addition, the estimated curves were not on the generated data points correctly. However, for the data with 5% high leverage outliers, the 95% credible interval of the proposed Gibbs Laplace partition model seems to adequately estimate for the hidden structure of data with high leverage outliers, even though the 95% credible interval is wider.

The EM NMM seems to underestimate the scale parameters of each clusters on each set of data compared to other Gibbs methods. Also, with fixed number of clusters as the true number of clusters, the estimated curve of the EM on the histogram seem not to consider the distribution of data with heavy tailed error, but the indices of clusters based on the EM seem close to the true indices of clusters. It is best use the mixture models with Gibbs sampling if our goal is density estimation; however, the EM will provide more hidden information if our goal is the detection of the cluster indices.

We observe that X_5 (orientation) is also an important variable to explain HL and CL with X_7 (glazing area) in the energy performance data of buildings. The estimated numbers of clusters for HL are 6 of NMM and 2 of Laplace random partition model; in addition, the estimated number of clusters for CL are 4 of NMM and 2 of Laplace random partition model. We conclude that there are various cluster structures to explain HL and CL with significant input variables in each cluster. Thus various linear combinations of orientation and glazing area are important elements to determine the energy performance of buildings, because the amount of glazing and the orientation of buildings determine the heat absorbed in a building due to the sun; in addition, similarly orientation and glazing is a source of heat leakage from the building to the environment.

Acknowledgements

Minjung Kyung was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science, ICT & Future Planning (Grant No. NRF-2015R1C1A1A01051837).

Appendix A: Posterior distribution of model parameters

The joint posterior distribution of parameters for cluster h in Section 3 is

$$\pi\left(\boldsymbol{\beta}_h, \boldsymbol{\tau}_h, b_h^2 | \mathbf{X}, \mathbf{Z}, by\right) \propto \left[\prod_{Z_i=h} (2\pi\tau_i)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\tau_i}\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right)^2\right\} \frac{1}{2b_h^2} \exp\left(-\frac{\tau_i}{2b_h^2}\right)\right] \frac{1}{b_h^2} \exp\left(-\frac{1}{2c_h}\boldsymbol{\beta}_h'\boldsymbol{\beta}_h\right).$$

For the posterior distribution of $\boldsymbol{\beta}_h$,

$$\begin{aligned}\pi\left(\boldsymbol{\beta}_h|\boldsymbol{\tau}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y}\right) &\propto \exp\left\{-\frac{1}{2}\sum_{Z_i=h}\frac{1}{\tau_i}\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right)'\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right) - \frac{1}{2c_h}\boldsymbol{\beta}_h'\boldsymbol{\beta}_h\right\} \\ &\propto \exp\left[-\frac{1}{2}\left\{\boldsymbol{\beta}_h'\left(\sum_{Z_i=h}\frac{1}{\tau_i}\mathbf{X}_i'\mathbf{X}_i + \frac{1}{c_h}\boldsymbol{I}\right)\boldsymbol{\beta}_h - 2\boldsymbol{\beta}_h'\left(\sum_{Z_i=h}\frac{1}{\tau_i}\mathbf{X}_i'y_i\right)\right\}\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}_h - \boldsymbol{\beta}_h^*\right)'\Sigma_{\boldsymbol{\beta}_h^*}^{-1}\left(\boldsymbol{\beta}_h - \boldsymbol{\beta}_h^*\right)\right],\end{aligned}$$

where

$$\boldsymbol{\beta}_h^* = \left(\sum_{Z_i=h}\frac{1}{\tau_i}\mathbf{X}_i'\mathbf{X}_i + \frac{1}{c_h}\boldsymbol{I}\right)^{-1}\left(\sum_{Z_i=h}\frac{1}{\tau_i}\mathbf{X}_i'y_i\right), \qquad \Sigma_{\boldsymbol{\beta}_h^*} = \left(\sum_{Z_i=h}\frac{1}{\tau_i}\mathbf{X}_i'\mathbf{X}_i + \frac{1}{c_h}\boldsymbol{I}\right)^{-1}.$$

Therefore,

$$\boldsymbol{\beta}_h|\boldsymbol{\tau}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim MVN_p\left(\boldsymbol{\beta}_h^*,\ \Sigma_{\boldsymbol{\beta}_h^*}\right).$$

For the posterior distribution of τ_i of $\boldsymbol{\tau}_h = \{\tau_i|Z_i = h\}$,

$$\begin{aligned}\pi\left(\tau_i|Z_i = h, \boldsymbol{\beta}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y}\right) &\propto \tau_i^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\tau_i}\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right)^2 - \frac{\tau_i}{2b_h^2}\right\} \\ &\propto \tau_i^{-\frac{1}{2}}\exp\left[-\frac{1}{2\tau_i b_h^2}\left\{b_h^2\left(y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right)^2 + \tau_i^2\right\}\right] \\ &\propto \tau_i^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\tau_i}\left(\frac{\tau_i}{b_h} - \left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|\right)^2\right\}.\end{aligned}$$

Let

$$\frac{\tau_i}{b_h\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|} = \frac{1}{t_i}$$

then

$$\begin{aligned}\pi\left(t_i|Z_i = h, \boldsymbol{\beta}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y}\right) &\propto \left(\frac{b_h\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{t_i}\right)^{-\frac{1}{2}} b_h\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right| t_i^{-2}\exp\left\{-\frac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|^2}{2\frac{b_h\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{t_i}}\left(\frac{1}{t_i} - 1\right)^2\right\} \\ &\propto \left(\frac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|/b_h}{2\pi t_i^3}\right)^{\frac{1}{2}}\exp\left\{-\frac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|/b_h}{2t_i}\left(t_i - 1\right)^2\right\}.\end{aligned}$$

Therefore,

$$t_i \equiv \left.\frac{b_h\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{\tau_i}\right| Z_i = h, \boldsymbol{\beta}_h, b_h^2, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim \text{inverse Gaussian}\left(\mu_i = 1,\ \lambda_i = \frac{\left|y_i - \mathbf{X}_i\boldsymbol{\beta}_h\right|}{b_h}\right),$$

and

$$\tau_i = \frac{b_h \left| y_i - \mathbf{X}_i \boldsymbol{\beta}_h \right|}{t_i}.$$

For the posterior distribution of b_h^2,

$$\pi\left(b_h^2 | \boldsymbol{\beta}_h, \boldsymbol{\tau}_h, \mathbf{X}, \mathbf{Z}, \mathbf{y}\right) \propto \left(\frac{1}{b_h^2}\right)^{n_h+1} \exp\left(-\frac{\sum_{Z_i=h} \tau_i}{2b_h^2}\right).$$

Therefore,

$$b_h^2 | \boldsymbol{\beta}_h, \boldsymbol{\tau}_h, \mathbf{X}, \mathbf{Z}, \mathbf{y} \sim \text{IG}\left(n_h, \frac{1}{2}\sum_{Z_i=1} \tau_i\right).$$

Appendix B: Parameter estimation

Table B.1: Posterior median and 95% credible interval of cluster-specific model parameters of NMM and LPM for Heating load

Model	Parameter	Cluster		
		1	2	3
NMM	β_1	0.29 (−1.89, 2.28)	0.23 (−1.81, 2.33)	−2.20 (−4.01, −0.71)
	β_2	−0.01 (−0.83, 0.77)	−0.02 (−0.83, 0.75)	0.00 (−0.78, 0.78)
	β_3	0.04 (−0.76, 0.86)	0.06 (−0.70, 0.87)	0.09 (−0.69, 0.88)
	β_4	−0.01 (−1.58, 1.63)	−0.02 (−1.56, 1.59)	−0.10 (−1.65, 1.49)
	β_5	1.87 (1.46, 2.77)	2.01 (−0.41, 2.72)	1.21 (0.82, 2.47)
	β_6	−0.01 (−0.44, 0.24)	−0.11 (−0.60, 0.41)	−0.02 (−0.11, 0.06)
	β_7	0.22 (−1.45, 2.63)	0.67 (−1.41, 2.79)	10.33 (2.27, 11.27)
	β_8	2.01 (1.05, 2.98)	2.07 (−0.38, 3.44)	−0.09 (−0.15, 2.16)
	σ	0.07 (0.04, 2.14)	1.30 (0.35, 2.20)	0.38 (0.32, 1.05)
	n_h	20	22	116
Model	Parameter	Cluster		
		4	5	6
NMM	β_1	−1.64 (−2.72, −0.33)	−1.40 (−2.77, 0.07)	−0.15 (−2.20, 1.71)
	β_2	0.01 (−0.84, 0.79)	0.02 (−0.79, 0.80)	0.07 (−0.71, 0.85)
	β_3	0.01 (−0.77, 0.86)	0.03 (−0.76, 0.83)	−0.03 (−0.81, 0.76)
	β_4	−0.05 (−1.63, 1.64)	−0.07 (−1.63, 1.53)	0.04 (−1.51, 1.60)
	β_5	4.11 (2.97, 4.25)	3.10 (2.89, 3.27)	−0.62(−1.46, 0.32)
	β_6	−0.01 (−0.10, 0.05)	−0.03 (−0.16, 0.10)	0.00 (−0.23, 0.24)
	β_7	11.85 (2.88, 12.32)	3.47 (2.39, 4.40)	7.82 (5.09, 10.68)
	β_8	−0.05 (−0.09, 0.06)	0.00 (−0.10, 0.11)	−0.11 (−0.29, 0.08)
	σ	0.39 (0.35, 0.71)	0.65 (0.54, 0.79)	1.32 (1.05, 1.65)
	n_h	371	125	114
Model	Parameter	Cluster		
		1	2	
LPM	β_1	−1.77 (−6.69, 1.42)	−1.38 (−3.30, 1.29)	
	β_2	0.00 (−0.78, 0.76)	−0.04 (−0.87, 0.75)	
	β_3	0.04 (−0.74, 0.82)	0.13 (−0.88, 1.41)	
	β_4	−0.04 (−1.58, 1.52)	−0.10 (−1.74, 1.52)	
	β_5	3.68 (2.98, 4.28)	3.62 (−1.43, 3.93)	
	β_6	−0.03 (−0.19, 0.13)	−0.03 (−0.84, 0.91)	
	β_7	11.63 (10.51, 14.38)	11.25 (0.88, 12.48)	
	β_8	0.09 (−0.03, 0.27)	0.09 (−0.73, 0.95)	
	b	2.23 (1.76, 2.43)	2.21 (1.00, 2.47)	
	n_h	729	39	

NMM = normal mixture model; LPM = Laplace partition model.

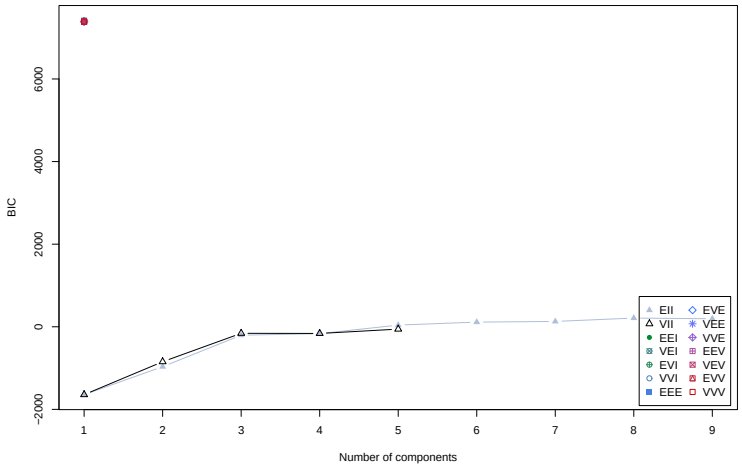
Table B.2: Posterior median and 95% credible interval of cluster-specific model parameters of NMM and LPM for cooling load

Model	Parameter	Cluster			
		1	2	3	4
NMM	β_1	0.10 (−1.54, 2.12)	1.13 (−0.58, 2.77)	−0.21 (−2.24, 1.80)	−1.36 (−3.42, 1.99)
	β_2	−0.01 (−0.80, 0.80)	−0.01 (−0.83, 0.79)	−0.04 (−0.84, 0.77)	−0.01 (−0.79, 0.76)
	β_3	0.08 (−0.72, 0.88)	0.08 (−0.73, 0.89)	0.26 (−0.55, 1.06)	0.40 (−0.76, 0.85)
	β_4	−0.03 (−1.63, 1.56)	−0.03 (−1.64, 1.60)	−0.19 (−1.79, 1.42)	−0.02 (−1.58, 1.54)
	β_5	0.76 (0.20, 1.15)	1.90 (0.77, 2.17)	0.61 (−1.40, 1.43)	5.31 (0.35, 5.62)
	β_6	−0.04 (−0.15, 0.08)	0.14 (−0.07, 0.32)	−0.06 (−0.57, 0.48)	0.04 (−0.12, 0.25)
	β_7	0.69 (−1.55, 2.52)	9.52 (1.52, 10.94)	1.81 (−0.36, 4.15)	9.95 (0.19, 10.86)
	β_8	1.10 (−0.00, 2.42)	0.03 (−0.10, 2.35)	0.35 (−0.16, 0.80)	0.01 (−0.07, 1.42)
	σ	0.15 (0.08, 0.38)	1.65 (0.14, 1.86)	2.06 (1.42, 2.80)	0.76 (0.13, 1.98)
	n_h	10	309	49	400

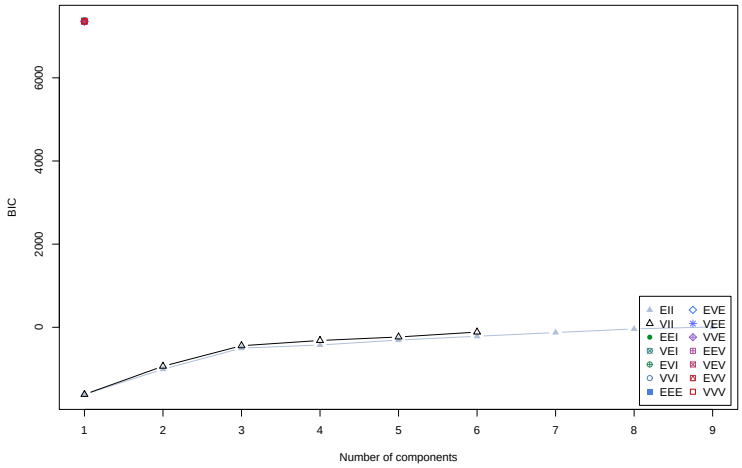
Model	Parameter	Cluster	
		1	2
LPM	β_1	−1.96 (−8.86, 0.82)	−1.02 (−3.39, 1.60)
	β_2	−0.00 (−0.81, 0.80)	−0.03 (−0.91, 0.84)
	β_3	0.03 (−0.77, 0.83)	0.11 (−0.93, 1.41)
	β_4	−0.02 (−1.62, 1.58)	−0.10 (−1.84, 1.59)
	β_5	3.80 (3.15, 5.05)	3.57 (−1.71, 4.16)
	β_6	0.12 (−0.13, 0.29)	0.13 (−0.83, 1.06)
	β_7	8.70 (7.37, 10.68)	7.74 (1.20, 9.51)
	β_8	0.06 (−0.11, 0.20)	0.05 (−0.86, 1.04)
	b	2.37 (1.98, 2.60)	2.35 (1.00, 3.70)
	n_h	739	29

NMM = normal mixture model; LPM = Laplace partition model.

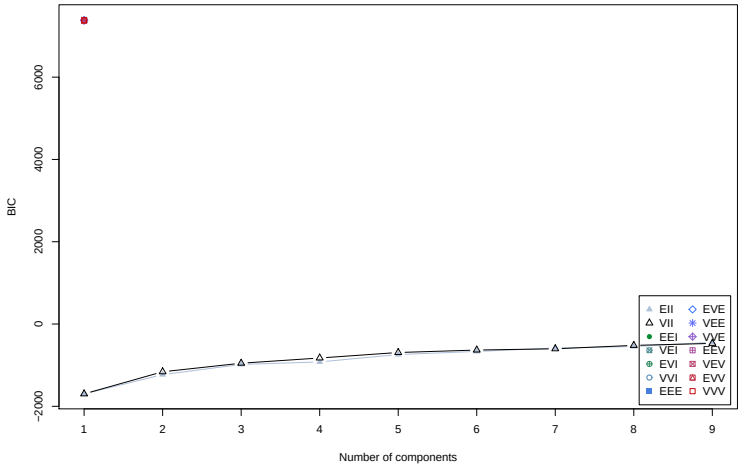
Appendix C: BIC plots of EM model based clustering



(a) Normal mixture



(b) Laplace random partition



(c) t(df = 5) Mixture

Figure C.1: *BIC plots of model based clustering based on the generated data from normal mixture model (Set 1), Laplace random partition model (Set 2), and t mixture model of df = 5 (Set 3). BIC = Bayesian information criterion.*

References

Airoldi EM, Costa T, Bassetti F, Leisen F, and Guindani M (2014). Generalized species sampling priors with latent Beta reinforcements, *Journal of the American Statistical Association*, **109**, 1466–1480.

Andrews DF and Mallows CL (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society Series B (Methodological)*, **36**, 99–102.

Antoniak CE (1974). Mixture of Dirichlet processes with applications to Bayesian nonparametric problems, *The Annals of Statistics*, **2**, 1152–1174.

Argiento R, Cremaschi A, and Guglielmi A (2014). A "density-based" algorithm for cluster analysis using species sampling Gaussian mixture models, *Journal of Computational and Graphical Statistics*, **23**, 1126–1142.

Banfield JD and Raftery AE (1993). Model-based Gaussian and non-Gaussian clustering, *Biometrics*, **49**, 803–821.

Barry D and Hartigan JA (1992). Product partition models for change point problems, *The Annals of Statistics*, **20**, 260–279.

Blackwell D and MacQueen JB (1973). Ferguson distributions via Pólya urn schemes, *The Annals of Statistics*, **1**, 353–355.

Booth JG, Casella G, and Hobert JP (2008). Clustering using objective functions and stochastic search, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **70**, 119–139.

Bush CA and MacEachern SN (1996). A semiparametric Bayesian model for randomized block design, *Biometrika*, **83**, 275–285.

Crowley EM (1997). Product partition models for normal means, *Journal of the American Statistical Association*, **92**, 192–198.

Dempster AP, Laird NM, and Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B (Methodological)*, **39**, 1–38.

Dielman TE (1984). Least absolute value estimation in regression models: an annotated bibliography, *Communications in Statistics Theory and Methods*, **4**, 513–541.

Dielman TE (2005). Least absolute value regression: recent contributions, *Journal of Statistical Computation and Simulation*, **75**, 263–286.

Dunson DB, Pillai N, and Park JH (2007). Bayesian density regression, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **69**, 163–183.

Escobar MD and West M (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, **90**, 577–588.

Ferguson TS (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, **1**, 209–230.

Fritsch A and Ickstadt K (2009). Improved criteria for clustering based on the posterior similarity matrix, *Bayesian Analysis*, **4**, 367–392.

Fraley C and Raftery AE (2002). Model-based clustering, discriminant analysis, and density estimation, *Journal of the American Statistical Association*, **97**, 611–631.

Fraley C and Raftery AE (2007). Bayesian regularization for normal mixture estimation and model-based clustering, *Journal of Classification*, **24**, 155–181.

Fraley C, Raftery AE, Murphy TB, and Srucca L (2012). mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation, University of Washington, Department of Statistics.

Hartigan JA (1990). Partition models, *Communications in Statistics Theory and Methods*, **19**, 2745–2756.

Ishwaran H and James LF (2001) Gibbs sampling methods for stick-breaking priors, *Journal of the American Statistical Association*, **96**, 161–173.

Ishwaran H and James LF (2003) Some further developments for stick-breaking priors: finite and infinite clustering and classification, *Sankhyā: The Indian Journal of Statistics*, **65**, 577–592.

Ishwaran H and Zarepour M (2000) Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models, *Biometrika*, **87**, 371–390.

Jordan C, Livingstone V, and Barry D (2007). Statistical modelling using product partition models, *Statistical Modelling*, **7**, 275–295.

Kyung M, Gill J, Ghosh M, and Casella G (2010). Penalized regression, standard errors, and Bayesian lassos, *Bayesian Analysis*, **5**, 369–412.

MacEachern SN (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*, Alexandria, VA.

MacEachern SN (2000). *Dependent Dirichlet processes*. Department of Statistics, The Ohio State University, Columbus, OH.

MacEachern SN and Müller P (1998). Estimating mixture of Dirichlet process models, *Journal of Computational and Graphical Statistics*, **7**, 223–238.

McCullagh P and Yang J (2007). Stochastic classification models. In *Proceedings of the International Congress of Mathematicians (Madrid, 2006)*, Madrid, 669–686.

McLachlan GJ and Peel D (2000). *Finite Mixture Models*, John Wiley & Sons, New York.

Müller P, Quintana F, Jara A, and Hanson T (2015). *Bayesian Nonparametric Data Analysis*, Springer, Cham.

Müller P, Quintana F, and Rosner GL (2011). A product partition model with regression on covariates, *Journal of Computational and Graphical Statistics*, **20**, 260–278.

Murua A and Quintana FA (2017). Semiparametric Bayesian regression via Potts model, *Journal of Computational and Graphical Statistics*, **26**, 265–274.

Neal RM (2000). Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Park JH and Dunson DB (2010). Bayesian generalized product partition model, *Statistica Sinica*, **20**, 1203–1226.

Pitman J (1996). Some developments of the Blackwell-MacQueen urn scheme, *Statistics, Probability and Game Theory*, 245-267, IMS Lecture Notes Monograph Series, **30**, Institute of Mathematical Statistics, Hayward, CA.

Richardson S and Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society, Series B*, **59**, 731–792.

Sethuraman J (1994). A constructive definition of Dirichlet priors, *Statistica Sinica*, **4**, 639–650.

Song W, Yao W, and Xing Y (2014). Robust mixture regression model fitting by Laplace distribution, *Computational Statistics and Data Analysis*, **71**, 128–137.

Stephens M (2000) Dealing with label switching in mixture models, *Journal of the Royal Statistical Society, Series B*, **62**, 795–809.

Tokdar ST, Zhu YM, and Ghosh JK (2010). Bayesian density regression with logistic Gaussian process and subspace projection, *Bayesian Analysis*, **5**, 319–344.

Tsanas A and Xifara A (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, *Energy and Buildings*, **49**, 560-567.

Quintana FA and Iglesias PL (2003). Bayesian clustering and product partition models, *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **65**, 557–574.

Wolfe JH (1970). Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Research*, **5**, 329–350.

Received April 25, 2017; Revised July 24, 2017; Accepted August 21, 2017