PARAMETER ESTIMATION FOR LATENT MIXTURE MODELS WITH APPLICATIONS TO PSYCHIATRY

by

Lulu Ren

B.S. Statistics, Nanjing University, 2000

M.A. Statistics, University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

Faculty of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Lulu Ren

It was defended on

April 20, 2006

and approved by

Satish Iyengar

Leon J. Gleser

Wesley K. Thompson

Shirley Y. Hill (WPIC)

Dissertation Director: Satish Iyengar

PARAMETER ESTIMATION FOR LATENT MIXTURE MODELS WITH APPLICATIONS TO PSYCHIATRY

Lulu Ren, PhD

University of Pittsburgh, 2006

Longitudinal and repeated measurement data commonly arise in many scientific research areas. Traditional methods have focused on estimating single mean response as a function of a time related variable and other covariates in a homogeneous population. However, in many situations the homogeneity assumption may not be appropriate. Latent mixture models combine latent class modeling and conventional mixture modeling. They accommodate the population heterogeneity by modeling each subpopulation with a mixing component. In this paper, we developed a hybrid Markov Chain Monte Carlo algorithm to estimate the parameters of the latent mixture model. We show through simulation studies that MCMC algorithm is superior than the EM algorithm when missing value percentage is large.

As an extension of latent mixture models, we also propose the use of cubic splines as a curve fitting technique instead of classic polynomial fitting. We show that this method gives better fits to the data, and our MCMC algorithm estimates the model efficiently. We apply the cubic spline technique to a data set which was collected in a study of alcoholism. Our MCMC algorithm shows several different P300 amplitude trajectory patterns among children and adolescents.

Other topics that are covered in this thesis include the identifiability of the latent mixture model and the use of such model to predict a binary outcome. We propose a bivariate version of the latent mixture model, where two courses of longitudinal responses can be modeled at the same time. Computational aspects of such models remain to be completed in the future.

KEY WORDS: Markov Chain Monte Carlo algorithm; EM algorithm; Gibbs sampler; Metropolis-Hastings algorithm; cubic spline; missing data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

I would like to thank the members of my committee, Dr. Leon J. Gleser, Dr. Shirley Y. Hill and Dr. Wesley K. Thompson for their insightful advice and kind help throughout the writing of this thesis.

I want to especially thank my advisor, Dr. Satish Iyenger, for his guidance during my graduate studies. I am very grateful to have the opportunity to work with Dr. Iyengar. His encouragement and patience helped me went through many hard times in my research. This thesis could not have been finished without him.

I also want to thank my wonderful friends, Ana-Maria Iosif, Jim Scisciani, Ahmet Sezer, Zhuoxin Sun, Melissa Ziegler, Jeongeun Kim and Kimberly Thomas. They have made my life in Pittsburgh so colorful. I would miss them dearly.

Last but not least, I want to dedicate this thesis to my parents, Yizhi Ren and Yunzhi Li, for their unconditional love and support. Also to my dearest cat friend, mimisan.

1.0 INTRODUCTION

1.1 OVERVIEW

Repeated measurement and longitudinal data commonly arise in many psychological or social research areas. Traditional methods have focused on the relationships between response variables and covariates. Usually, the data population is considered to be homogeneous. A single mean response is estimated as a function of a time related variable and other covariates. However, in many situations the data come from a heterogeneous population. In these cases researchers are interested in not only the responses' change over time but also the differences between each subpopulation. Examples of the methods dealing with such problems include cluster analysis, finite mixture models, and latent class analysis. In these methods, "the goal is to group individuals into categories, each one of which contains individuals who are similar to each other and different from individuals in other categories" [24]. Heterogeneity among the longitudinal data can be represented as multiple developmental trajectories, which are commonly seen in alcohol, drug, and mental health research. See [27] for more examples.

In 1998, Muthén and Shedden proposed a generalized latent variable modeling framework [23, 26], which we will call a latent mixture model in this thesis. Recall the classical mixed model:

$$Y_i = X_i\alpha + Z_i\beta_i + \epsilon_i, \qquad i = 1, \cdots, n.$$

where α is the fixed effect, β_i represent the individual-specific random effect and ϵ_i are independent normally distributed error terms. Unlike the random effect model, latent mixture models assume β_i are sampled from a mixture distribution instead of one probability distribution. To determine the mixing proportions, a categorical latent variable C_i, a class

membership indicator, is introduced into the model. The heterogeneity of the population is therefore modeled by combining continuous latent variables and categorical latent variables together.

To further understand the generalized latent variable modeling, it is helpful to take a look at the conventional latent variable analysis and related modeling from which the latent mixture modeling framework was drawn. Early applications of latent class analysis in medicine were done by Rindskopf and Rindskopf [32], and Ubersax and Grove [43]. The goal is to associate observed categorical variables, for instance, diagnosis or symptoms, with unobserved latent classes; that is to find the smallest number of latent classes possible to group the categorical variables and at the same time preserve the diversity of the data. A continuous version of latent class analysis is the latent class growth analysis [27]. It mainly focuses on longitudinal data, where there are multiple measurements over time for each subject. The object of the analysis is to group subjects into classes where the mean growth curve shape for each class is different from others, and to estimate the posterior probability of being in each class.

Finite mixture models [42] have a very broad scope. The choice of mixing components is flexible: Gaussian mixtures are often used for heterogeneous continuous data whereas Poisson mixtures are commonly used for categorical data. Often, both the component distributions and mixing proportions depend on the covariates of the subject: see [18]. As an extension to the models above, latent mixture modeling estimates the mean growth curve for each class and individual variation within the class. Both goals are carried out by assuming individual-specific random effects. The details of the model will be illustrated in the section 2.1.2.

Parameter estimation for the mixture models is rather complicated. Different methodologies had been developed in the past three decades. Likelihood based approaches are central with the *Expectation Maximization* (EM) algorithm [6] being a milestone. The EM algorithm treats the unobserved class indicators as latent variables and augments the observed incomplete data. It is an especially appealing idea when the augmented likelihood has a simplified form, in which case maximum likelihood estimators can be easily obtained. Further description of the EM algorithm along with some variations such as ECM and Monte Carlo

EM will be given in the section 2.2.1.

In practice, one also needs to take account of the missing data problem. Missing data is a common occurrence when the measurements are taken repeatedly on the same subject over time. Subjects may be lost due to attrition, high expense of follow up or by study design constrains. Missing data makes standard analysis more difficult and less accurate. EM-type algorithms can be set up to deal with the missing data problem: the missing values are regarded as latent as well as other unknown hyperparameters. The EM algorithm augments the data as if missing values were observed and works with the augmented likelihood. As the complexity of the model and the missing proportion increase, EM-type algorithms become more difficult to handle because finding the expectation of the predictive likelihood can become very complicated and tedious.

An alternative way to analyze this complicated model is through Bayesian methods, e.g. the Markov Chain Monte Carlo (MCMC) method. MCMC algorithms are becoming increasingly popular because advanced software and high speed computers considerably reduce the computational complexity. MCMC algorithms are stochastic sampling procedures and they are set up to approximate the full posterior distribution instead of obtaining a point estimator as in the EM algorithm. In this paper, we implemented a hybrid MCMC algorithm, a combination of Gibbs sampler and Metroplis-Hastings algorithm, to estimate the parameters in the latent mixture model. The details will be given in chapter 2.

1.2 MOTIVATING DATA SET

The data to be analyzed were first collected by Hill et al. [16] for the study of the association between event-related potential (ERP) component P300 and familial risk for alcohol dependence. An ERP is a series of positive and negative voltage deflections in the ongoing electroencephalography (EEG) in response to certain stimuli, often visual or auditory. The P300 component is a positive ERP component peaking at approximately 300 millisecond (ms) after stimulus onset [50]. In the neurobiology literature, many studies have shown that both P300 amplitude and latency are related to gender, age and psychopathology. The

P300 component has been studied as a potential biological marker for development of many psychiatric disorders, including schizophrenia [28] and alcoholism [29, 35].

1.2.1 Subjects and Data Collection

Children between ages 8 to 18 whose parents enrolled in a large family study were followed annually. The children are either at high risk or low risk for developing alcohol dependence based on familial loading for alcohol dependence [17]. Each child was also administered the Schedule for Affective Disorders and Schizophrenia for School-Aged Children (K-SADS; [2]). The presence or absence of any childhood diagnosis was used to further classify subtypes of high- and low-risk children.

Considerable evidence exists suggesting that P300 amplitude abnormality in childhood is a risk marker for later development of alcohol dependence [17]. There is also evidence that P300 amplitude is heritable [16]. The main goal of the analysis is to determine if risk status confers a different developmental pattern across childhood and adolescence. The second goal is to determine if the pattern types are associated with the individual diagnoses of the children during the course of the study as well. We will also provide methods to study the relationship between the P300 patterns and development of alcoholism in young adulthood.

There are 137 subjects in total, 68 boys and 69 girls. Among all subjects, 78 children are from high risk families and 59 children are from low risk families. The response variable is the annual measurement of P300 amplitude in microvolts (μV). We consider the association between P300 and the presence of the child's psychiatric disorders, which was measured by the K-SADS. To avoid the high dimensionality of the data, we collapsed the K-SADS diagnoses into two groups: internalizing and externalizing disorder diagnoses. Internalizing disorders include depression, mania, panic attacks, separation anxiety, phobias, generalized anxiety, dysthymia, cyclothymia, and overanxious disorder. Externalizing disorders include conduct disorder, attention deficit disorder, oppositional disorder, alcohol abuse and drug use disorder. Thus, the covariates for each subject included in the study were gender, risk status, internalizing and externalizing psychiatric disorder diagnoses. All 137 subjects had complete covariates measured and were followed into adulthood (age 21 and older), when

Table 1: P300 data set

	High-Risk Children		Low-Risk Children		
	Int./Ext.	None	Int./Ext.	None	
Male	24	13	11	20	68
Female	22	19	8	20	69
Total	46	32	19	40	137

the alcohol dependence and other outcomes were recorded.

Table 1 shows a break down of our data set.

1.2.2 Missing Data Problem

It would have been best if all subjects were observed each year from age 8 to 18. However, certain subjects did drop out the study or skip one or more yearly interviews. Some subjects only have a few observations whereas others attended almost all the interviews. In fact, for each child we only observe a part of the P300 amplitude trajectory. The total missing value percentage in our data set is 50.50% (defined as $\frac{\text{total No. of missing observations}}{\text{total No. of possible observations}}$). Missing value frequency varies for different ages; see figure 1 for details. In this thesis, we propose the use of a hybrid MCMC algorithm to fit the latent mixture model. This algorithm treats missing observations as latent variables and augments the observed data with values sampled from the conditional posterior distributions of the latent variables given observed data and parameters. The sampling method and assessment for convergence will be discussed in chapter 2.

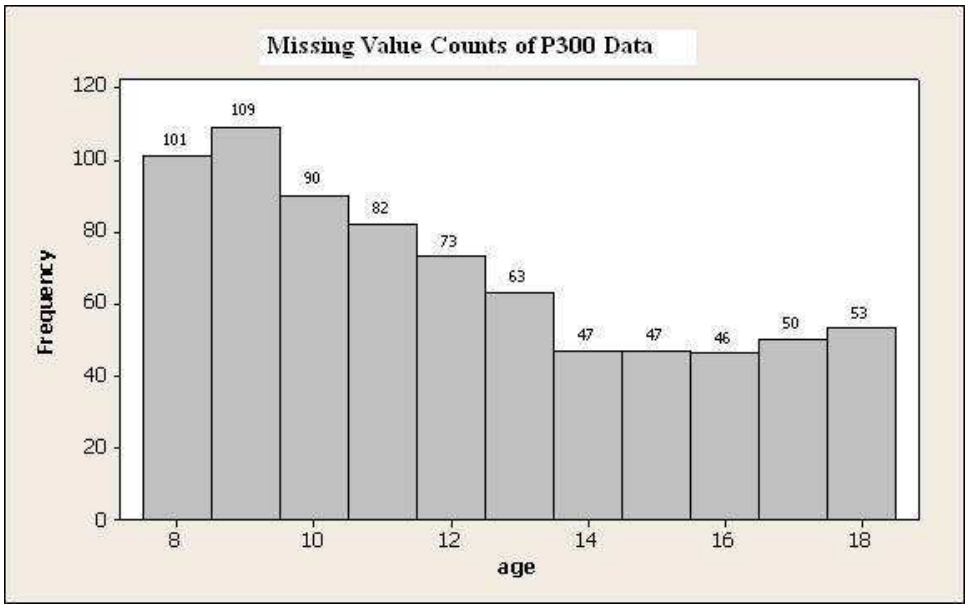

Figure 1: Missing data count vs. age

1.3 CURVING FITTING TECHNIQUES

In this study, our goal is to model the heterogeneity amongst the longitudinal courses of P300 amplitude. To do so, an individual developmental curve must be fitted to each subject. Múthen and Shedden proposed a polynomial fit in their latent mixture model framework. We instead propose using cubic splines for curve fitting because spline methods have more flexibility in describing the shape of the trajectory, especially when the curve has local fluctuations. In our work, the knots of the spline functions are fixed first and then set as free parameters. In section 4.5.2, we give a detailed introduction to this technique.

2.0 BACKGROUND

2.1 LATENT MIXTURE MODELING FRAMEWORK

In the longitudinal data analysis setting, a latent mixture modeling framework [23] offers a flexible way to study the outcome's change over time. In many studies, it is believed that there are different subtypes of subjects in a population. To capture the heterogeneity, the latent mixture modeling proposes to combine the classical random effect mixture model and the latent class model. We now will give brief introductions to both methodologies.

2.1.1 Latent Class Modeling

Latent class models are used to identify subtypes of related subjects from multivariate categorical data. They can be applied in cluster analysis, factor analysis and regression. For example, the observed categorical variables may be a set of diagnostic variables and the latent class variables may correspond to the presence or absence of certain risk factor. In short, by using latent class analysis, one hopes to find a small number of underlying subtypes that can describe the associations among the observed categorical variables. Latent class models refer to the unobserved subtypes as latent classes. Let $Y_1, Y_2, \ldots, Y_J$ denote the set of observed categorical variables, and C denote the latent variable. Suppose C has K categories, and the probability associated with class k is $\pi_C(k), \quad k = 1, \ldots, K$. Since $\sum_{k=1}^{K} \pi_C(k) = 1$, there are $(K-1)$ free parameters in the model.

There are two standard assumptions in latent class analysis. First is called internal homogeneity: the subjects in a certain latent class have the same distribution. Second is called local independence: for each subject, $Y_1, \ldots, Y_J$ are independent within the same

latent class [5]. Using these two assumptions, we easily obtain the joint distribution of Y and C:

$$P_{Y,C}(y,k) = \pi_C(k)\, P_{Y|C(k)}(y) = \pi_C(k) \prod_{j=1}^{J} P_{Y_j|C(k)}(y_j). \tag{2.1}$$

The latent class model can therefore be written as:

$$P_Y(y) = \sum_{k=1}^{K} P_{Y,C}(y,k) = \sum_{k=1}^{K} \pi_C(k)\, P_{Y|C(k)}(y). \tag{2.2}$$

One common application of the latent class model is the prediction of latent class membership given parameter estimates, i.e. the prediction of C given the value of Y:

$$\pi_{C|Y}(k) = P_{Y,C}(y,k) / P_Y(y)$$

For some examples of applications, see the work of Rindskopf [32], and Young et al. [48].

2.1.2 Random Effects Mixture Model

In classical longitudinal data analysis, random effects are often assumed to be normally distributed, i.e.

$$y_i = X_i\alpha + Z_i\beta_i + \epsilon_i, \quad i = 1, \ldots, N \tag{2.3}$$

where y_i is the n_i dimensional vector of response for the ith subject, with $\beta_i \sim N(0, D)$ and $\epsilon_i \sim N(0, \sigma^2 I_{n_i})$. X_i and Z_i are the covariate matrices; α represents the fixed effects and β_i is the subject-specific random effect vector. It is often assumed that β_i is independent of the error term ϵ_i. Since the random effects are sampled from a *single* multivariate normal distribution, this model can be called the *homogeneity* model.

Verbeke and Lesaffre showed that when the random effects in fact are sampled from a finite mixture of normal distributions, the random effects may be badly estimated if homogeneity is assumed [44]. They proposed a *heterogeneity* model to accommodate clustered $\beta_i's$. This is of importance in longitudinal models where there is heterogeneity in the subject population. For instance, in the context of the alcohol study, the subjects come from a family with low risk or high risk for alcohol dependence.

The heterogeneity model assumes that random effect β_i is sampled from a mixture of K multivariate normal distributions with means μ_k and common covariance matrix D. Each component of the mixture represents a subtype of the population.

Let p_{ik} be the mixing proportion of the kth component for subject i, $\sum_{k=1}^{K} p_{ik} = 1$. The marginal distribution of y_i can be written as:

$$y_i \sim \sum_{k=1}^{K} p_k N\left(X_i\alpha + Z_i\mu_k, Z_i D Z_i^{'} + \sigma^2 I_{n_i}\right) \tag{2.4}$$

Redner and Walker discussed the details of fitting the model using EM algorithm [30].

To check whether the correct number of mixture components was used, Verbeke and Lesaffre suggested a goodness-of-fit test for heterogeneity using the Kolmogorov - Smirnov test. A linear combination of the repeated measures of each subject $a_i^{'}y_i$ can be used to avoid the evaluation of the multivariate normal distribution. The Kolmogorov-Smirnov test then is performed on the cumulative density function of $a_i^{'}y_i$, which are uniformly distributed under correct model. A drawback for this test is the choice of a, the ideal choice is the one with maximal variability in $a_i^{'}y_i$ due to random effects compared to the variability due to the error terms. For more on such tests, see [44].

2.1.3 Latent Mixture Model

The latent mixture model framework was proposed by Muthén and Shedden in 1999 [23]. It incorporates the ideas of random effect mixture models and latent class models by using both continuous and categorical latent variables.

Consider an n_i-dimensional vector y_i of continuous variables and an r-dimensional vector u_i of binary outcomes, which are related to each other in the model via latent variables. x_i is the vector of covariates, η is the vector of continuous latent variables and c is a vector of latent categorical variables. Also, let $c_i = (c_{i1}, \ldots, c_{iK})^{'}$ have a multinomial distribution, where $c_{ik} = 1$ when subject i belongs to the kth latent class and is zero otherwise. The random effect mixture part of the model can be written as:

$$y_i = \Lambda_y \eta_i + \epsilon_i \tag{2.5}$$

$$\eta_i = Ac_i + \Gamma x_i + \xi_i \tag{2.6}$$

where Λ_y is a $n_i \times p$ matrix of constants whose values depend on the context. ϵ_i is the error term that is independent with other variables and is distributed as $N(0, \Sigma_\epsilon)$, with Σ_ϵ diagonal. A and Γ are parameter matrices relating the classes and covariates to the outcome. ξ_i is another residual vector that is distributed as $N(0, \Psi)$.

Next, suppose that the values of binary variables u are independent given the class membership of subject i. That is: $p(u_{i1}, \ldots, u_{ir}|c_i) = p(u_{i1}|c_i) \ldots p(u_{ir}|c_i)$. Let $\tau_i = (\tau_{i1}, \ldots, \tau_{ir})'$ where $\tau_{ij} = p(u_{ij} = 1|c_i)$. An ordinary logit model gives:

$$\text{logit}(\tau_i) = \Lambda_u c_i. \tag{2.7}$$

Finally, define $p_i = (p_{i1}, \ldots, p_{iK-1})'$ where $p_{ik} = p(c_{ik} = 1|L_i)$. The $(K-1)$-dimensional vector $\text{logit}(p_i) = (\log[p_{i1}/p_{iK}], \ldots, \log[p_{iK-1}/p_{iK}])$ can be modeled using logistic regression:

$$\text{logit}(p_i) = \gamma_k' \tilde{x}_i. \tag{2.8}$$

where $\gamma_k \quad k = 1, \ldots, K-1$ are parameter vectors associated with the covariates and $\tilde{x}_i$ represents the covariate vector of subject i with an intercept term.

Equations 2.5 - 2.8 represent the latent mixture model in a hierarchical structure. Individual response is related to the covariates via continuous latent variables, i.e. random effects. The random effects also depend on the class membership. This model allows us to estimate the mean response vector for each class, and to predict the binary outcomes of each individual. Also the posterior probability that individual i falls in the kth class can be estimated. It is common in mixture models that a subject is classified to the class for which he/she has the highest posterior probability.

2.2 COMPUTATIONAL APPROACHES

2.2.1 EM Algorithm

The *Expectation Maximization* (EM) algorithm was first named by Dempster et al. in 1977 [6]. Earlier applications of the EM algorithm were widely scattered in the literature. To obtain the maximum likelihood estimators (MLE) for the parameters, traditional methods

usually involve differentiating the log-likelihood function and solving the resulting normal equations to locate the mode of the log-likelihood. When the model is complicated, these steps can turn into massive computations because they often do not have analytic solutions. Rather than performing a difficult maximization directly, the EM algorithm augments the data with "latent data" that simplifies the calculation and subsequently performs a series of simple maximizations [38]. The following is a detailed description of the algorithm.

Let y and z be the observed data and the augmented latent data respectively. Assume that the likelihood of y can be written as $g(y|\theta)$. The goal is to get the MLE for θ by maximizing the function $l(\theta) = \log g(y|\theta)$. In many cases, $l(\theta)$ has no easy closed form. On the other hand, the complete data likelihood $f(y, z|\theta)$ usually is easier to deal with. The EM algorithm therefore calculates the MLE by using $f(y, z|\theta)$ instead of the observed likelihood $l(\theta)$. The EM algorithm is an iterative procedure, which consists of two steps: the E-step and the M-step. At the ith iteration:

1. E-step:

   Calculate $Q\left(\theta, \theta^{(i)}\right) = E_z\left(\log p(\theta|y, z)\,|y, \theta^{(i)}\right)$

2. M-step:

   Find $\theta^{(i+1)}$ which maximizes the Q function in E-step

The algorithm is iterated until $||\theta^{(i+1)} - \theta^{(i)}||$ is sufficiently small.

It can be shown that the EM algorithm increases the posterior $p(\theta|y, z)$ at each iteration, i.e. $p\left(\theta^{(i+1)}|y, z\right) \geq p\left(\theta^{(i)}|y, z\right)$ Also, if the iterates $\theta^{(i)}$ converge, they converge to a stationary point of $p(\theta|y, z)$. This implies that to reach the global maximum, multiple starting values may be needed to avoid the situation where EM algorithm is stuck at some local maximum or saddle points. For more details, see [19] and [45]. Dempster et al. also showed that the EM algorithm converges at a linear rate, with the rate depending on the proportion of information about θ that is observed. This means that the computation time can be quite long when a large portion of the data is missing.

"Despite of the slow convergence, the EM algorithm has become a very popular computational method in statistics"[45]. One of the advantages of the EM algorithm is its ability of handling missing data. This feature makes the EM algorithm a useful tool for parameter

estimation of latent mixture models. Therefore the EM algorithm is used as the default estimation method in many commercial packages such as Mplus [25] and Amos (Distributed by SPSS), which are designed to analyze models with latent variables.

We use notation that is in line with latent class models: y_i is the observed longitudinal response, each observation belongs to one of the K underlying classes. The class indicators c_i are unobservable. Augment c_i to obtain the complete data likelihood:

$$L(\theta|y,c) = \prod_{i=1}^{n}\prod_{k=1}^{K} p_{ik} f_k(\theta|y_i) \tag{2.9}$$

where p_{ik} denotes the probability that ith subject belongs to the kth class, and $c_i = (c_{i1},\ldots,c_{iK})$ has multinomial distribution with parameters 1 and $(p_{i1},\ldots,p_{iK})$. The E-step and M-step in each iteration can be written as the following:

1. E-step:
   Calculate $Q\left(\theta,\theta^{(i)}\right) = E_c\left(\log p(\theta|y,c)\,|y,\theta^{(i)}\right) = \sum_{k=1}^{K}\left[\log p(\theta|y,c_k)\right] p\left(c_k = 1|\theta^{(i)},y\right)$
2. M-step:
   Maximize $Q\left(\theta,\theta^{(i)}\right)$ with respect to θ to get the new iterate $\theta^{(i+1)}$.

2.2.2 Other EM-type Algorithms

There are many variants of the EM algorithm created to facilitate either the E-step or the M-step. The ECM and Monte Carlo EM algorithms are two examples.

2.2.2.1 ECM algorithm ECM stands for *Expectation Conditional Maximization* algorithm. EM algorithm becomes less appealing when the complete likelihood is complicated. The computation required for the M-step may be difficult. Fortunately, complete likelihood maximization can be simplified when conditional on some function of the parameters [46]. In ECM, the M-step is replaced by a CM step where for $c = 1,\cdots,C$, we find $\theta^{i+c/C}$ that maximizes the Q function conditional on $g_c(\theta) = g_c\left(\theta^{(t+(c-1)/C)}\right)$. Simply speaking, ECM algorithm would maximize the Q function over a subvector of the parameter vector holding the rest fixed and then maximize the Q function over those fixed parameters given the new

value for previous ones. To ensure the resulting mode is an unconstrained maximum in the parameter space, g_s has to be a *space filling* function. For more details, see [46].

2.2.2.2 Monte Carlo EM algorithm

Monte Carlo Expectation Maximization refers to a Monte Carlo E-step in the EM algorithm. The Monte Carlo method can be used to facilitate the calculation of the Q function. Recall that $Q\left(\theta,\, \theta^{(i)}\right) = E_z\left(\log\, p\left(\theta|y,z\right)|y,\theta^{(i)}\right)$; based on the current value $\theta^{(i)}$, Q function can be approximated by the following steps:

1. Draw $z_1, \cdots, z_m \sim p(Z|Y,\theta^{(i)})$
2. Let $\hat{Q}_{i+1} = \frac{1}{m}\log p(\theta|z_j, Y), \;\; j = 1, \cdots, m.$

The choice of which algorithm to use in general depends on the data and computational power that is available. In this paper, The conventional EM algorithm was implemented by using a statistical package Mplus. We first fit the latent mixture model to a simulated data set that contains three latent classes. No missing values are assumed at this point. Besides the EM algorithm, we also developed a MCMC algorithm for parameter estimation, which will be introduced in the following section. When we compared the results from the two algorithms, both methods performed well. Then we simulated a data set with a large proportion of missing values and the missing value pattern is the same as the P300 data. It turned out that EM algorithm was not able to achieve convergence before it hit the maximum number of iterations. In these difficult cases, the MCMC method still gave good estimates and confidence intervals for the parameters. For more details, see section 3.2.

2.2.3 MCMC Algorithms

Even with the help of data augmentation methods, the algebra required for the EM algorithm can be overwhelming so that obtaining the MLE is very difficult. To overcome this problem, many statisticians resort to Bayesian approaches, namely *Markov Chain Monte Carlo* (MCMC) algorithms. For a wide class of problems, the method of Markov chain simulation appears to be the easiest way to get reliable results [12]. The idea is to simulate a random walk in the parameter space which converges to a stationary distribution that is the full conditional posterior distribution of the parameter $p\left(\theta|y,z\right)$. MCMC algorithms, stochastic

in nature, are set up to approximate the full distribution, which is a more ambitious task than the point estimation needed for the EM algorithm [37].

2.2.3.1 Gibbs Sampler The Gibbs sampler is one of the most widely used MCMC algorithms. It is especially appealing for its ability to reduce the complex multidimensional problems to a sequence of much lower dimensional ones. Suppose the parameter vector to be estimated is $\theta = (\theta_1, \ldots, \theta_d)$. It samples one component or one subvector of the parameter vector at a time using the full conditional posterior distribution $p\left(\theta_s|\theta_1, \ldots, \theta_{s-1}, \theta_{s+1}, \ldots, \theta_d, y, z\right)$. Given the starting point $\theta^{(0)} = \left(\theta_1^{(0)}, \ldots, \theta_d^{(0)}\right)$, the sampling steps are as following:

1. Sample $\theta_1^{(i+1)}$ from $p\left(\theta_1|\theta_2^{(i)}, \ldots, \theta_d^{(i)}, y, z\right)$

2. Sample $\theta_2^{(i+1)}$ from $p\left(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \ldots, \theta_d^{(i)}, y, z\right)$

$\vdots \qquad \vdots$

d. Sample $\theta_d^{(i+1)}$ from $p\left(\theta_d|\theta_1^{(i+1)}, \ldots, \theta_{d-1}^{(i+1)}, y, z\right)$.

The vectors $\theta^{(0)}$, $\theta^{(1)}, \ldots, \theta^{(i)}, \ldots$ are a realization of a Markov chain. In practice, the Gibbs sampler is usually run N iterations with a burn-in period at the beginning. The iterates obtained after burn-in can be regarded as a random sample from the conditional posterior distribution of θ. Inferences of the parameters can be done based on the sample.

The following convergence results had been shown by Chan [3] and Wu, Wong and Kong [20, 21] and Tierney [41]:

Result 1. The joint distribution of $\left(\theta_1^{(i)}, \ldots, \theta_d^{(i)}\right)$ converges geometrically to $p\left(\theta_1, \ldots, \theta_d|y, z\right)$ as $i \to \infty$.

Result 2. $\frac{1}{N}\sum_{i=1}^{N} p\left(\theta^{(i)}\right) \overset{a.s.}{\to} \int p\left(\theta\right) p\left(\theta|y, z\right) d\theta$, as $N \to \infty$.

From the sampling steps 2.2.3.1 – 2.2.3.1, we can see that the Gibbs sampler is easy to implement when the conditional posterior distribution of the parameter has a simple form. Unfortunately, it is not always the case in practice. One remedy is sampling from

approximate distributions (also called a candidate density) and correcting those draws to approximate the target posterior distribution. Many clever methods have been developed for this purpose. The *Metropolis-Hastings* (M-H) algorithm is a general term for a family of Markov chain simulation methods that are useful for drawing samples from Bayesian posterior distributions [12]. The Metropolis algorithm (Metropolis et al. 1953) was first developed to study the equilibrium properties of large systems of particles such as electrons in atoms. Hastings (1970) suggested a generalization of the Metropolis algorithm [38]. In this section, we present the general Metropolis-Hastings algorithm.

2.2.3.2 Metropolis-Hastings algorithm Assume that the target distribution $p(\theta|y)$ can be computed up to a normalizing constant. The Metropolis algorithm simulates a sequence of random points $\left(\theta^{(1)}, \theta^{(2)}, \ldots\right)$ whose distributions converge to the target distribution. The algorithm proceeds as following:

1. Specify a starting value $\theta^{(0)}$.

2. For $i = 1, 2, \ldots$

a. Draw a candidate point $\theta^{(*)}$ from the candidate density $q\left(\theta^{(*)}, \theta^{(i)}\right)$, where $\theta^{(i)}$ is the current value of θ.

b. Update the chain to $\theta^{(*)}$ with probability $\alpha\left(\theta^{(*)}, \theta^{(i)}\right)$, where

$$\alpha\left(\theta^{(*)}, \theta^{(i)}\right) = \min\left\{\frac{p(\theta^{(*)}|y)q(\theta^{(i)}, \theta^{(*)})}{p(\theta^{(i)}|y)q(\theta^{(*)}, \theta^{(i)})}, 1\right\} \tag{2.10}$$

Note: if the candidate draw is not accepted, let $\theta^{(i+1)} = \theta^{(i)}$ and the algorithm moves to the next iteration.

Chib and Greenberg suggested using multivariate normal or multivariate t distribution as the candidate density [4]. To better approximate the target distribution, the density mode and inverse of the Hessian evaluated at the mode will be used as the mean and the covariance matrix of the candidate density respectively.

2.2.3.3 Prior Distributions The choices for the prior distributions are essential to the convergence of the MCMC algorithm. Generally, three different kind of priors are used in estimating finite mixtures. Take a simple mixture of normal distributions as an example.

- Independent Priors [10]: Parameters are assumed a priori independent, e.g.

$$p\left(\mu, \sigma^2\right) = p\left(\mu\right) p\left(\sigma^2\right) \tag{2.11}$$

- Conjugate Priors [8]: Diebolt and Robert suggested a conjugate prior of the following form:

$$p\left(\mu, \sigma^2\right) = p\left(\mu|\sigma^2\right) p\left(\sigma^2\right) \tag{2.12}$$

- Hierarchical Priors [31]: Green and Richardson proposed the hierarchical prior in 1997, where a hyperparameter β is introduced into the prior distribution.

$$p\left(\mu, \sigma^2, \beta\right) = p\left(\beta\right) p\left(\mu\right) p\left(\sigma^2|\beta\right) \tag{2.13}$$

2.2.3.4 Label Switching Problem The identifiability problem for mixture model can be very complicated. One common phenomenon in this context is called the label switching problem. The likelihood of mixture model is invariant with respect to any permutation of the mixing components. If there is no prior information to distinguish them, the likelihood will have $K!$ symmetric modes. There are many ways to solve such problems. One approach is to put artificial constraints on the parameters. In the context of finite mixture model, one can put the constrains on either the means or the variances, e.g. $\mu_1, < \mu_2 < \cdots < \mu_K$ or $\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_K^2$. Under such constraints, the mixing components are always sorted in certain order so that the mixture model is identifiable.

In this paper, we implemented a hybrid MCMC method, i.e. embedding a Metropolis random-walk subchain in the Gibbs sampler in the situation where it is difficult to sample directly from a conditional posterior distribution. This strategy was suggested by Müller in 1993 [22]. We chose independent priors as the prior distributions for the parameters. And we overcame the label switching problem by imposing constraints on mixing components.

3.0 PARAMETER ESTIMATION

To illustrate that the MCMC method is an appropriate approach when there is a considerable amount of missing information in the data set, we simulated data sets each containing the same number of subjects and same covariates as the P300 data. We also generated data sets which have the same missing data pattern as the P300 data.

We first started with a well-separated 3-class mixture data set. We analyzed the complete data using both the EM algorithm, a built-in procedure in M-plus package and MCMC algorithm that we developed. We will see that the two methods agree in general. Moreover, the MCMC algorithm yielded narrower confidence intervals for parameter estimates. Then we deleted some points so that the simulated data has the same missing pattern as P300 data. Both methods were applied to the data again. This time, the EM algorithm had difficulty with convergence while the MCMC algorithm still produced reasonable estimates and confidence intervals.

We further generated a 3-class mixture data set where the class means were poorly separated. MCMC algorithm was applied to the data. Again, the estimates and confidence intervals for model parameters are quite good. After demonstrating MCMC algorithm is an appropriate approach to problems of this sort, we applied it to the P300 data.

3.1 THE LATENT MIXTURE MODEL

Recall the latent mixture model framework which was introduced in section 2.1.3. To best serve the purpose of our P300 study, we chose a matrix which contains the time points when the measurements were taken as Λ_y. A second degree polynomial curve was employed to fit

the P300 trajectory. Moreover, polynomial coefficients were treated as random effects and latent class membership was introduced using a categorical latent variable c.

We first take the mixture part of the model to address the multiple trajectory patterns. Assume that there are K underlying classes; n subjects in the study; n_i is the number of observations that are taken for ith subject; and each subject has Q covariates. The modified latent mixture model then can be written as:

$$y_i = T\eta_i + \epsilon_i \quad \text{or} \quad y_{ij} = T_j^{'}\eta_i + \epsilon_{ij} \tag{3.1}$$

where y_{ij} is the the jth observation of the ith subject with $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$. T is a matrix contains rows of $\left(1,\, t_j,\, t_j^2\right)$, where t_j is the subject's age when the jth observation was taken. ϵ_i is a residual term which is uncorrelated with other variables and normally distributed with mean zero and diagonal covariance matrix $\Sigma_\epsilon = \texttt{diag}\left(\sigma_1^2, \ldots, \sigma_{n_i}^2\right)$.

The random effect η_i can be modeled as:

$$\eta_i = Ac_i + \Gamma x_i + \xi_i, \tag{3.2}$$

where A is a parameter matrix containing columns of intercepts of each class. Γ is another parameter matrix containing coefficients which relate to the covariates, x_i, for the ith subject. ξ_i is an error term that is normally distributed with mean zero and covariance matrix Ψ. As in (2.6), c_i is a K-dimensional latent categorical vector, $c_{ik} = 1$ if subject i belongs to the kth class, and zero otherwise. c_i has multinomial distribution with parameters 1 and $(p_{i1}, \ldots, p_{iK})^{'}$ and is modeled using multinomial logit regression:

$$\log(\frac{p_{ik}}{p_{iK}}) = \gamma_k^{'}\tilde{x}_i, \tag{3.3}$$

for $k = 1, \ldots, K-1$. $\tilde{x}_i$ is a vector contains a constant 1, and the covariates x_i for subject i.

3.1.1 Augmented Likelihood

The MCMC algorithm augments the data with latent variables so that the likelihood can be expressed using a series of conditional densities. In this case, all missing data, the random coefficients and class indicators are treated as latent variables.

The parameters are: σ_j^2 $(j = 1, \ldots, n_i)$, A, Γ, Ψ γ_k $(k = 1, \ldots, K)$.

The latent variables are: c_i, η_i $(i = 1, \ldots, n)$, $y_{i^*j^*}$, i^*, j^* are the indices for missing values.

Thus the complete data likelihood is:

$$
\begin{aligned}
L &= \prod_{i=1}^{n} f\left(y_i|\eta_i\right) f\left(\eta_i|c_i, x_i\right) f\left(c_i|x_i\right) \\
&= \prod_{i=1}^{n} \left[\phi\left(T\eta_i, \Sigma_\epsilon\right) \phi\left(Ac_i + \Gamma x_i, \Psi\right) \prod_{k=1}^{K} p_{ik}^{c_{ik}} \right] \\
&\propto \prod_{i=1}^{n} \left\{ |\Psi|^{-1/2} \exp\left[-\frac{1}{2} \left(\eta_i - Ac_i - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - Ac_i - \Gamma x_i\right) \right] \right. \\
&\quad \left. \prod_{j=1}^{n_i} \left(\sigma_j^{-1}\right) \exp\left(-\frac{\left(y_{ij} - T_j'\eta_i\right)^2}{2\sigma_j^2} \right) \prod_{k=1}^{K} \left[\frac{\exp\left(\gamma_k'\tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left(\gamma_l'\tilde{x}_i\right)} \right]^{c_{ik}} \right\} \\
&\propto |\Psi|^{-\frac{n}{2}} \prod_{i=1}^{n} \left\{ \prod_{k=1}^{K} \left\{ \exp\left[-\frac{1}{2} \left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right] \frac{\exp\left(\gamma_k'\tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left(\gamma_l'\tilde{x}_i\right)} \right\}^{c_{ik}} \right. \\
&\quad \left. \prod_{j=1}^{n_i} \left(\sigma_j^{-1}\right) \exp\left(-\frac{\left(y_{ij} - T_j'\eta_i\right)^2}{2\sigma_j^2} \right) \right\}.
\end{aligned}
$$

where $\phi(\cdot, \cdot)$ is the density function of the multivariate normal distribution. A_k is the kth column of parameter matrix A, and $T_j = \left(1, t_i, t_j^2\right)'$.

3.1.2 Prior and Posterior Distribution

When choosing priors, we split the parameter matrices into columns, and sample one column at a time. Here, we gave the same prior to those columns coming from the same parameter matrix:

$p\left(A_k\right) = N\left(0, \sigma_A^2 I\right), \quad k = 1, \ldots, K;$

$p\left(\Gamma_q\right) = N\left(0, \sigma_\Gamma^2 I\right), \quad q = 1, \ldots, Q;$

$p\left(\sigma_j^2\right) = IG\left(a_0, b_0\right), \quad j = 1, \ldots, n_i.$

It is more convenient to sample from Ψ^{-1} instead of sampling from Ψ directly. Thus, let $p\left(\Psi^{-1}\right) = WI\left(v, S\right)$, a Wishart prior with v a small constant and S a positive definite matrix.

Finally, let $p\left(\gamma_k\right) = N\left(0, \sigma_\gamma^2 I\right), \;\; k = 1, \ldots, K-1$.

It follows that:

a.

$$\begin{aligned} p\left(A_k | \sigma_A^2, \Gamma, \Psi^{-1}, \left\{\eta_i\right\}, \left\{c_i\right\}, \left\{y_{ij}\right\}\right) &\propto \exp\left\{-\frac{1}{2}\left[A_k^{'}\Sigma_A^{-1}A_k - 2\sum_{i=1}^{n} c_{ik}A_k^{'}\Psi^{-1}\left(\eta_i - \Gamma x_i\right)\right]\right\} \\ &\sim N\left(\Sigma_A \Psi^{-1}\sum_{i=1}^{n} c_{ik}\left(\eta_i - \Gamma x_i\right),\; \Sigma_A\right) \end{aligned} \tag{3.4}$$

where $\Sigma_A = \left(\sum_{i=1}^{n} c_{ik}\Psi^{-1} + \frac{1}{\sigma_A^2} I\right)^{-1}$, $k = 1, \ldots, K$.

b.

$$\begin{aligned} p\left(\Gamma_q | \sigma_\Gamma^2, A, \Psi^{-1}, \left\{c_i\right\}\right) &\propto \exp\left\{-\frac{1}{2}\left[\Gamma_q^{'}\Sigma_\Gamma^{-1}\Gamma_q - 2\sum_{i=1}^{n}\Gamma_q^{'}X_{iq}\Psi^{-1}\left(\eta_i - Ac_i - \sum_{r\neq q} x_{ir}\Gamma_r\right)\right]\right\} \\ &\sim N\left(\Sigma_\Gamma \sum_{i=1}^{n} x_{iq}\Psi^{-1}\left(\eta_i - Ac_i - \sum_{r\neq q} x_{ir}\Gamma_r\right),\; \Sigma_\Gamma\right) \end{aligned} \tag{3.5}$$

where $\Sigma_\Gamma = \left(\sum_{i=1}^{n}\sum_{k=1}^{K} c_{ik}x_{iq}^2\Psi^{-1} + \frac{1}{\sigma_\Gamma^2} I\right)^{-1}$, $q = 1, \ldots, Q$.

c.

$$\begin{aligned} p\left(\sigma_j^2 | a_0, b_0, \left\{\eta_i\right\}, \left\{y_{ij}\right\}\right) &\propto \left(\sigma_j^2\right)^{-a_0 - \frac{n}{2} - 1} \exp\left\{-\frac{1}{\sigma_j^2}\left[\frac{1}{2}\sum_{i=1}^{n}\left(y_{ij} - T_j^{'}\eta_i\right)^2 + b_0\right]\right\} \\ &\sim IG\left(a_0 + \frac{n}{2},\; \frac{1}{2}\sum_{i=1}^{n}\left(y_{ij} - T_j^{'}\eta_i\right)^2 + b_0\right) \end{aligned} \tag{3.6}$$

d.

$$\begin{aligned} p\left(\Psi^{-1} | v, S, A, \Gamma, \left\{\eta_i\right\}, \left\{c_i\right\}\right) &\propto \left|\Psi^{-1}\right|^{\frac{n+v-3-1}{2}} \exp\left\{-\frac{1}{2}\mathtt{tr}\left[\sum_{i=1}^{n}\alpha_i^{'}\Psi^{-1}\alpha_i + S^{-1}\Psi^{-1}\right]\right\} \\ &\sim WI\left(n + v,\; \left(\sum_{i=1}^{n}\alpha_i\alpha_i^{'} + S^{-1}\right)^{-1}\right) \end{aligned} \tag{3.7}$$

where $\alpha_i = \eta_i - Ac_i - \Gamma x_i$.

e.

$$p\left(\gamma_k|\sigma_\gamma,\{c_i\}\right) \propto \exp\left\{-\frac{\gamma_k^{'}\gamma_k}{2\sigma_\gamma^2}\right\}\prod_{i=1}^{n}\frac{\exp\left(\gamma_k^{'}\tilde{x}_i\right)^{c_{ik}}}{\sum_{l=1}^{K}\exp\left(\gamma_l^{'}\tilde{x}_i\right)} \tag{3.8}$$

Note that for the multinomial part of the model to be identifiable, γ_K is fixed as 0. Moreover, the posterior distribution of γ_k does not have a standard form.

3.2 HYBRID MARKOV CHAIN MONTE CARLO ALGORITHM

3.2.1 Sampling Scheme

As stated in the last section, some parameters' posterior densities have a very simple form. Therefore, they can be sampled directly from their posterior distributions. However, some posterior distributions are quite complicated. In order to obtain random sample from these posterior distributions, we implemented a combination of two sampling methods: Gibbs sampler and Metropolis-Hastings algorithm.

Given the initial values: $\left(c_i^{(0)},\eta_i^0,y_{i^*j^*}^{(0)},\sigma_j^{(0)},A^{(0)},\Gamma^{(0)},\Psi^{-1(0)},\gamma_k^{(0)}\right)$, where $i=1,\ldots,n,\ j=1,\ldots,n_i,\ k=1,\ldots,K$, the sampling scheme iterates the following steps:

1. Sample $c_i=(c_{i1},\ldots,c_{iK})^{'}$ from multinomial $(1,\ (p_{i1},\ldots,p_{iK})^{'})$. Where

$$p_{ik}=\frac{\exp\left\{-\frac{1}{2}\left(\eta_i-A_k-\Gamma x_i\right)^{'}\Psi^{-1}\left(\eta_i-A_k-\Gamma x_i\right)\right\}\exp\left(\gamma_k^{'}\tilde{x}_i\right)}{\sum_{k=1}^{K}\exp\left\{-\frac{1}{2}\left(\eta_i-A_k-\Gamma x_i\right)^{'}\Psi^{-1}\left(\eta_i-A_k-\Gamma x_i\right)\right\}\exp\left(\gamma_k^{'}\tilde{x}_i\right)}$$

   $i=1,\ldots,n,\ k=1,\ldots,K$. Again, $\gamma_K^{(0)}$ and γ_K are fixed at 0.
2. Sample $y_{i^*j^*}$ from $N(T_{j^*}^{'}\eta_{i^*},\ \sigma_{j^*}^2)$.
3. Sample η_i from $N\left(\Sigma_\eta\left[\Psi^{-1}\left(Ac_i+\Gamma x_i\right)+T^{'}D^{-1}y_i\right],\ \Sigma_\eta\right)$,
   where $D=\mathtt{diag}\left(\sigma_1^2,\ldots,\sigma_{n_i}^2\right)$, $\Sigma_\eta=\left(\Psi^{-1}+T^{'}D^{-1}T\right)^{-1}$.
4. Sample σ_j^2 from $IG\left(a_0+\frac{n}{2},\ \frac{1}{2}\sum_{i=1}^{n}\left(y_{ij}-T_j^{'}\eta_i\right)^2+b_0\right)$, $j=1,\ldots,n_i$.
5. Sample A_k from $N\left(\Sigma_A\Psi^{-1}\sum_{i=1}^{n}c_{ik}\left(\eta_i-\Gamma x_i\right),\ \Sigma_A\right)$, $k=1,\ldots,K$.
6. Sample Γ_q from $N\left(\Sigma_\Gamma\sum_{i=1}^{n}x_{iq}\Psi^{-1}\left(\eta_i-Ac_i-\sum_{r\neq q}x_{ir}\Gamma_r\right),\ \Sigma_\Gamma\right)$, $q=1,\ldots,Q$.
7. Sample Ψ^{-1} from $WI\left(n+v,\ \left(\sum_{i=1}^{n}\alpha_i\alpha_i^{'}+S^{-1}\right)^{-1}\right)$.

8. Sample from the posterior distribution of γ_k using the Metropolis-Hastings algorithm. Let $\log p(\gamma_k)$ denote the log conditional posterior density of γ_k as in (3.8).

i. Obtain $m_0 = \arg\max_{\gamma_k} \log p(\gamma_k)$ via Newton-Raphson algorithm by using quantities:

$$\frac{\partial \log p(\gamma_k)}{\partial \gamma_k} = -\frac{\gamma_k}{\sigma_\gamma^2} + \sum_{i=1}^{n} \left(c_{ik} - \frac{\exp \gamma_k' \tilde{x}_i}{\sum_{l=1}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)} \right) \tilde{x}_i \tag{3.9}$$

$$\frac{\partial^2 \log p(\gamma_k)}{\partial^2 \gamma_k} = -\frac{1}{\sigma_\gamma^2} I - \sum_{i=1}^{n} \frac{\exp\left(\gamma_k' \tilde{x}_i\right) \sum_{l \neq k}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)}{\left(\sum_{l=1}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)\right)^2} \tilde{x}_i \tilde{x}_i' \tag{3.10}$$

ii. Calculate $V_0 = \tau \left(-\frac{\partial^2 \log p(\gamma_k)}{\partial^2 \gamma_k} |_{m_0} \right)^{-1}$, where τ is a tunning parameter. It can be adjusted to control the dispersion of the candidate density and the acceptance rate of the random draws. We chose the candidate density:

$$q(\gamma_k | m_0, V_0, \tau) \propto |V_0|^{-1/2} \exp\left\{ -\frac{1}{2} (\gamma_k - m_0)' V_0^{-1} (\gamma_k - m_0) \right\}, \tag{3.11}$$

which is a multivariate normal distribution with mean m_0 and covariance matrix V_0.

iii. Draw a random deviate $\gamma_k^{(*)}$ from the candidate density, and update the Markov chain of γ_k from the current value to $\gamma_k^{(*)}$ with probability

$$\alpha\left(\gamma_k^{(*)}, \gamma_k\right) = \min\left\{ \frac{p(\gamma_k^{(*)}) q(\gamma_k | m_0, V_0, \tau)}{p(\gamma_k) q(\gamma_k^{(*)} | m_0, V_0, \tau)},\ 1 \right\}. \tag{3.12}$$

This completes the last step of the Gibbs sampler.

3.3 SIMULATION RESULTS

As already mentioned at the beginning of this chapter, we generated three data sets, each has three latent classes. In each simulation, we estimated the model parameters using EM algorithm (Mplus package) and MCMC method. The Markov chain in MCMC algorithm ran for 8,000 iterations after a 2000-iteration burn-in procedure. Multiple starting values were tried on each data set, and we chose the one giving highest observed likelihood among the ones led to convergence.

The first data set contain 137 subjects with no missing values. We tested both the EM algorithm and MCMC method. The results are shown in the following tables[1](numbers in bold are the results from MCMC method). we can see that the estimates obtained from the EM algorithm and the MCMC method are relatively close. Almost all the 90% CI's contain the true parameter values. Moreover, the MCMC method gave slightly smaller standard errors and narrower confidence intervals.

Table 2: Mixing proportions – simulation 1

Mixing Proportion	Class 1	Class 2	Class 3
True Value	10.08%	13.85%	76.07%
ML (M-plus Program)	16.90%	7.04%	76.06%
MCMC Method	**11.23%**	**12.68%**	**76.06%**

[1]For other parameters, see appendix.

Table 3: Estimate of A – simulation 1

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.E./S.D.
A_{11}	20.0	17.7132	19.837	21.9608	1.295
		17.4521	**18.951**	**20.4625**	**0.905**
A_{21}	-2.3	-3.0076	-1.794	-0.5804	0.740
		-2.9506	**-2.125**	**-1.2901**	**0.504**
A_{31}	-0.8	-1.8489	-0.819	0.2109	0.628
		-1.4770	**-0.694**	**0.07970**	**0.473**
A_{12}	35.0	24.8902	35.478	46.0658	6.456
		33.3011	**34.828**	**36.3443**	**0.929**
A_{22}	-0.9	-2.1124	-0.556	1.0004	0.949
		-1.8164	**-0.977**	**-0.1354**	**0.511**
A_{32}	1.2	-1.5532	1.138	3.8292	1.641
		0.7645	**1.562**	**2.3430**	**0.482**
A_{13}	42	39.7116	42.27	44.8284	1.560
		40.2360	**41.574**	**42.9120**	**0.813**
A_{23}	-5.0	-5.9176	-4.886	-3.8544	0.629
		-5.4835	**-4.752**	**-3.9977**	**0.456**
A_{33}	0.04	-0.9140	0.088	1.0900	0.611
		-0.5043	**0.200**	**0.8913**	**0.424**

Next, we deleted a portion of points so that the resulting data set has the same missing value pattern as the P300 data. In this case, the EM algorithm treats every missing value as a latent variable. Finding the expectation of the predictive conditional distribution became much more difficult. We implemented the EM algorithm on this data set using Mplus package. The algorithm did not converge before it hit the maximum number of iterations. Thus, only the results from the MCMC method are presented[2].

Table 4: Mixing proportions – simulation 2

Mixing Proportion	Class 1	Class 2	Class 3
True Value	10.08%	13.85%	76.07%
MCMC Method	14.08%	17.35%	68.57%

[2]For other tables see appendix

Table 5: Estimate for A – simulation 2

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
A_{11}	20.0	16.86056	19.5285	22.09346	1.59773
A_{21}	-2.3	-3.78259	-2.4204	-0.87782	0.90047
A_{31}	-0.8	-1.72956	-0.7660	0.16025	0.58249
A_{12}	35.0	33.14202	35.8333	38.50763	1.61999
A_{22}	-0.9	-2.33911	-1.19234	0.11313	0.74396
A_{32}	1.2	0.51944	1.47692	2.41357	0.58771
A_{13}	42	40.26159	42.37972	44.71559	1.33021
A_{23}	-5.0	-6.45723	-5.44123	-4.34767	0.64115
A_{33}	0.04	-0.76917	0.05202	0.81102	0.48051

Finally, we generated another 3-class mixture data where the 3 classes are poorly separated. Again, MCMC algorithm was able to provide resonable estimates and confidence intervals for the parameters.

Table 6: Mixing proportions – simulation 3

Mixing Proportion	Class 1	Class 2	Class 3
True Value	10.67%	67.60%	21.73%
MCMC Method	5.09%	73.78%	21.13%

Table 7: Estimate for A – simulation 3

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
A_{11}	8.0	3.1983	6.2295	9.5857	1.9790
A_{21}	1.2	-1.6532	0.2628	2.1611	1.1479
A_{31}	-2.8	-4.0392	-2.2104	-0.3677	1.1204
A_{12}	16.0	12.7120	14.540	16.7814	1.2442
A_{22}	1.8	0.2157	1.4014	2.6655	0.7384
A_{32}	-3.0	-3.4910	-2.2547	-0.9661	0.7658
A_{13}	30.0	27.2673	29.2864	31.6624	1.3841
A_{23}	1.0	-1.1527	0.1551	1.5147	0.7989
A_{33}	-2.8	-3.8310	-2.5288	-1.1848	0.8045

It is known that random samples from Gibbs sampler sampling procedure are usually correlated. We used "thinning" strategy, i.e. take one sample for every three or five iterations. The resulting estimates were not very different from averaging the whole random sample. We did encounter label switching problem in the simulation study. The following plot shows a path plot of mixing proportions for a 2-class mixture model. We can see the mixing proportions switched over after the Markov chain had already stabilized for many iterations.

To prevent label switching, we impose ordering constraints on the mixing components, i.e:

$A_{11} < A_{12} <, \ldots, < A_{1K}$,

Or $A_{21} < A_{22} <, \ldots, < A_{2K}$ when the first components are equal.

...

The simulation study showed that MCMC algorithm can efficiently estimate the latent mixture model even when there is a large amount of missing data. In the next chapter, we fit the latent mixture model to P300 data using the MCMC algorithm.

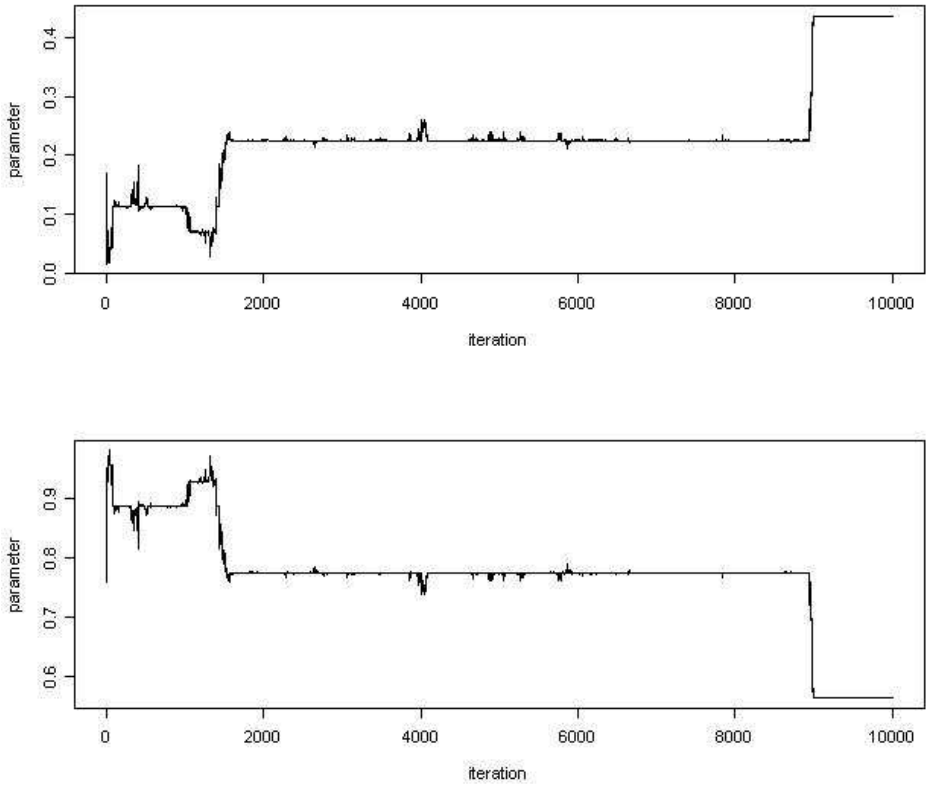

Figure 2: Label switching in Gibbs sampler

4.0 CUBIC SPLINE TECHNIQUE

4.1 MOTIVATION

Second and third degree polynomials have been widely used in growth curve models. They have also been adapted by latent mixture model users. Polynomial fitting is appropriate when the underlying curve is somewhat smooth and the function is only evaluated at a relatively small number of points. In many cases, though, the function which we intend to estimate can have many fluctuations. Such problem arose when we used a second degree polynomial to fit the P300 trajectory. A few initial runs indicated that the behavior of the P300 amplitude over time was roughly linear. This finding was inconsistent with what had been found in neurobiology literature. Further inspection of the raw data showed that P300 amplitude increases during early childhood to a high point and then levels off as the subject gets older. A second degree polynomial cannot model such behavior well over a broad age range. To overcome such shortcomings, we propose the use of splines in latent mixture modeling.

4.2 CUBIC SPLINES

4.2.1 Piecewise Polynomial Splines

Suppose $f(X)$ is the underlying continuous function we want to estimate. We can divide the domain of X into several contiguous subintervals and represent $f(X)$ by a separate polynomial in each interval. When the piecewise polynomials are continuous and have continuous

first and second order derivatives at the knots, which are the break points of the subintervals, it is known as a spline. A common choice, the cubic spline, consists of polynomials of degree three or less. Its claimed that cubic splines are the lowest-order spline for which the knot-discontinuity is not visible to the human eye [14], and that there is rarely any good reason to go beyond cubic-splines. One intuitive way to represent cubic spline is to use a truncated polynomial basis. Generally a spline of degree M-1 with K knots will have $M+K$ basis functions:

$$\begin{aligned} h_j(x) &= x^{j-1}, \quad j = 1, \ldots, M. \\ h_{M+1}(x) &= (x - \xi_l)_+^{M-1}, \quad l = 1, \ldots, K. \end{aligned}$$

where $\{\xi_l\}$ are the interior knots of the spline.

4.2.2 Natural Cubic Splines

A natural cubic spline adds boundary constraints to the ordinary spline, e.g. $f(X)$ is linear beyond the boundary knots. A natural cubic spline with K knots can be represented by $K+2$ basis functions. One can start from an ordinary cubic spline basis and then reduce the basis by imposing boundary constrains.

Truncated power basis is not the only way to represent the spline. The B-spline basis is another popular choice. B-spline bases have computational advantage over truncated polynomial bases. They are defined recursively as the following:

$$\begin{aligned} B_{i,1}(x) &= \left\{ \begin{array}{ll} 1 & \text{if } \xi_i \leq x < \xi_{i+1} \\ 0 & \text{otherwise} \end{array} \right) \\ B_{i,M}(x) &= \frac{x - \xi_i}{\xi_{i+M} - \xi_i} B_{i,M-1}(x) + \frac{\xi_{i+M+1} - x}{\xi_{i+M+1} - \xi_{i+1}} B_{i+1,M-1}(x) \end{aligned}$$

where $i = 1, \ldots, K+M$.

Fortunately, most current statistical software packages can generate spline basis functions for any order and sequence of knots, including the B-spline basis matrix. The B-spline basis has important computational advantages because it has minimal local support, so there are many off diagonal zeros when the observation points are sorted. In our paper, we propose the use of B-spline basis to generate piecewise polynomials.

4.3 SIMULATION STUDY

To use cubic spline in latent mixture model, we substitute the second degree polynomial basis by B-Spline basis:

$$T \quad \rightarrow \quad B$$

$$\begin{pmatrix} 1 & t_1 & t_1^2 \\ & \cdots & \\ 1 & t_{ni} & t_{ni}^2 \end{pmatrix} \rightarrow \begin{pmatrix} B_{1,4}(t_1) & B_{2,4}(t_1) & \cdots & B_{K+M,4}(t_1) \\ & & \cdots & \\ B_{1,4}(t_{ni}) & B_{2,4}(t_{ni}) & \cdots & B_{K+M,4}(t_{ni}) \end{pmatrix},$$

where $B_{i,4}, \quad (i = 1, \ldots, 4 + K)$ is the ith B-Spline basis function for a cubic spline with K knots.

Returning to the P300 data, as we said earlier, the P300 amplitude shows a high point at early age. Since we only evaluate the spline at 11 (ages 8-18) points, it is reasonable to use one knot in this context. We fixed the knot to be at age 12 based on the fact that the P300 amplitude presents different pattern before and after puberty.

Using the B-spline basis matrix increases the number of parameters while the sampling procedure still follows through as before. To show that our algorithm performs well using cubic spline fitting technique, we simulated a 3-class data set where the mean of each class is a cubic spline with a knot placed at age 12.

The following tables contain the parameter estimates, standard deviation of the samples from the posterior distribution and the 90% confidence intervals.

Table 8: Estimate for A – fixed knot spline

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
A_{11}	-8	-8.665147	-7.171415	-5.6265476	0.9240607
A_{21}	2	1.394544	2.405728	3.5379609	0.6380144
A_{31}	-5	-7.206609	-5.761632	-4.5539684	0.7760513
A_{41}	-5	-5.297267	-4.238954	-3.1396257	0.6815811
A_{51}	4	2.501114	3.436957	4.3711204	0.5639592
A_{12}	2	1.943493	2.6581024	3.3414943	0.4325731
A_{22}	6	5.166365	5.7693049	6.4070957	0.3832608
A_{32}	8	7.760301	8.6684092	9.5120036	0.5172460
A_{42}	-2	-2.841543	-2.1875127	-1.5100157	0.4077285
A_{52}	-1	-1.294250	-0.7661024	-0.2273716	0.3221489
A_{13}	9	8.258286	9.428654	10.6400421	0.7184915
A_{23}	-2	-3.417012	-2.219586	-1.0507394	0.7282726
A_{33}	7	5.759844	7.325635	8.9646741	0.9935571
A_{43}	-4	-4.970921	-3.757044	-2.5465597	0.7546951
A_{53}	6	4.745251	5.640232	6.5316172	0.5539389

Table 9: Mixing proportions – fixed knot spline

Mixing Proportion	Class 1	Class 2	Class 3
True Value	31.25%	31.50%	37.25%
MCMC Algorithm	32.12%	33.58%	34.31%

We chose multiple initial values to start the MCMC algorithm. Among the ones yielded convergent results, we present the one leading to the highest likelihood.

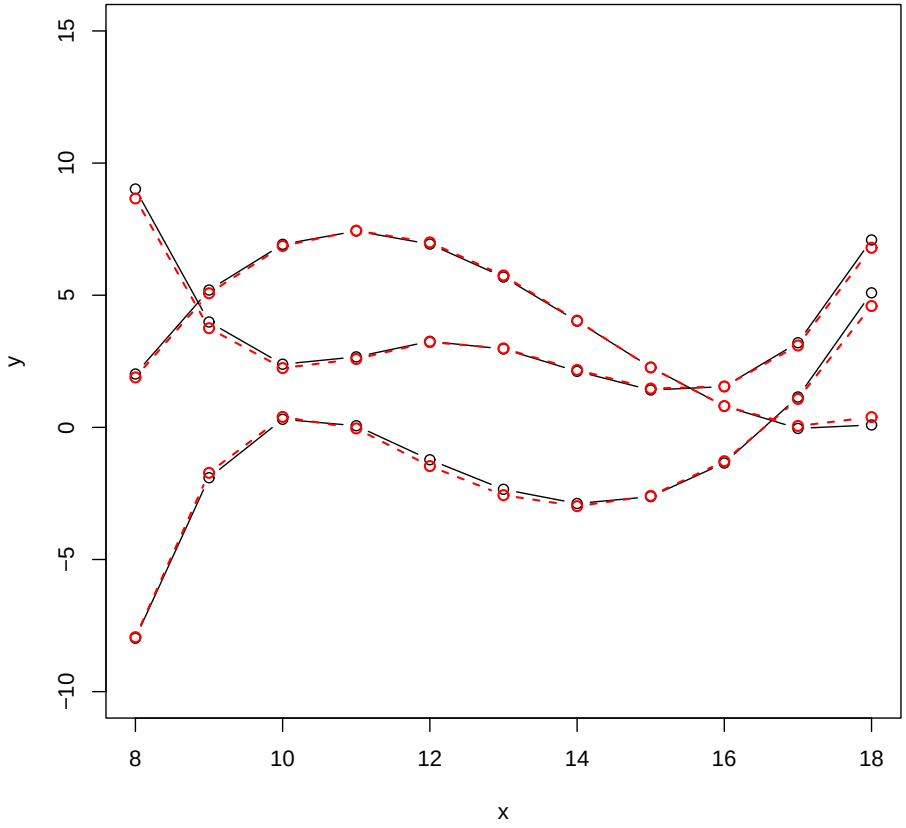

Figure 3: solid lines: generated true curves; dashed lines: fitted cubic splines (fixed knot at 12)

4.4 APPLICATION

4.4.1 Modification

In this section, we show the results from using the MCMC algorithm to analyze the P300 data. We started with a second degree polynomial basis in the model. The results showed linear trend of P300 amplitude trajectory. As we mentioned before, this finding is inconsistent with existing knowledge about P300 amplitude of children and adolescents: the P300

amplitude trajectories are expected to increase during childhood, reach a peak and leveling off during adolescence period. We then applied the spline fitting technique to the data set. Before do so, we made some modifications to the latent mixture model. These modifications are necessary adjustments to deal with actual data.

- The data was transferred to a logarithm scale. P300 amplitude is always positive by its definition. The logarithm transform avoids the error of generating negative values for the P300 amplitude.
- The covariance matrix Ψ was set to 0. We knew from the initial runs that the values in Ψ are very small compare to the those in matrix Σ_ϵ. In another words, we changed the model to a fixed effect mixture model.
- The matrix Γ is set to be 0. We reasoned that in the logit model for mixing proportions, the contribution of x_i has already been taken account for: the posterior probability of being in either class is a function of the covariates. We use the weaker assumption so that the class means do not necessarily depend on the covariates.

Based on the above modifications, the latent mixture model can be written as:

$$y_i = B\left(Ac_i\right) + \epsilon_i \tag{4.1}$$

$$\text{logit}\left(\frac{p_{ik}}{p_{iK}}\right) = \gamma_k^{'}\tilde{x}_i \tag{4.2}$$

where $i = 1, \cdots, n,\ k = 1, \cdots, K-1$.

To determine the appropriate number of classes for the P300 data set, We ran both 2-class and 3-class mixture models. The MCMC algorithm typically ran for 8,000 iterations with a 1500 burn-in period. The whole sampling procedure takes about 6-7 hours on a Pentium 4 PC with 2G RAM. The convergence of the Markov Chain was monitored by examining the iteration path plots of the parameters.

4.4.2 Three Classes Mixture

We can see that the majority class contains about 57% of the subjects. Other two minority classes having similar sizes contain about 20% and 22% of the subjects respectively. In figure 4.4.2 we can see three different P300 trajectory patterns. They are consistent with the theory that P300 increases first and then levels off. To show that our algorithm converges, we present an iteration path plot for the observed log likelihood function as well, see figure 5.

Table 10: Three-class mixture – P300 data

Mixing Proportions	Class 1	Class 2	Class 3
MCMC Algorithm	57.67%	20.44%	21.9%

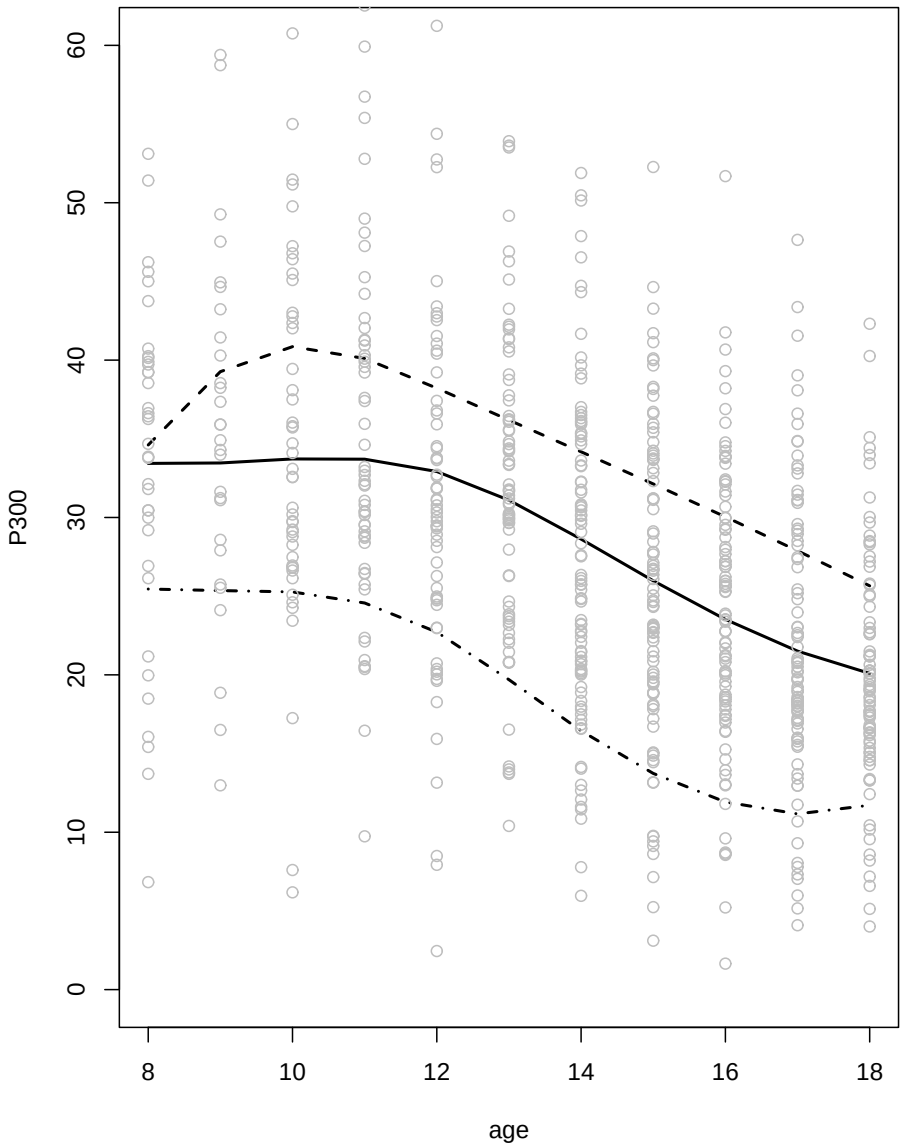

Figure 4: Three-class mixture – P300 data: solid line class1; dotted/dashed line class 2; dashed line: class 3

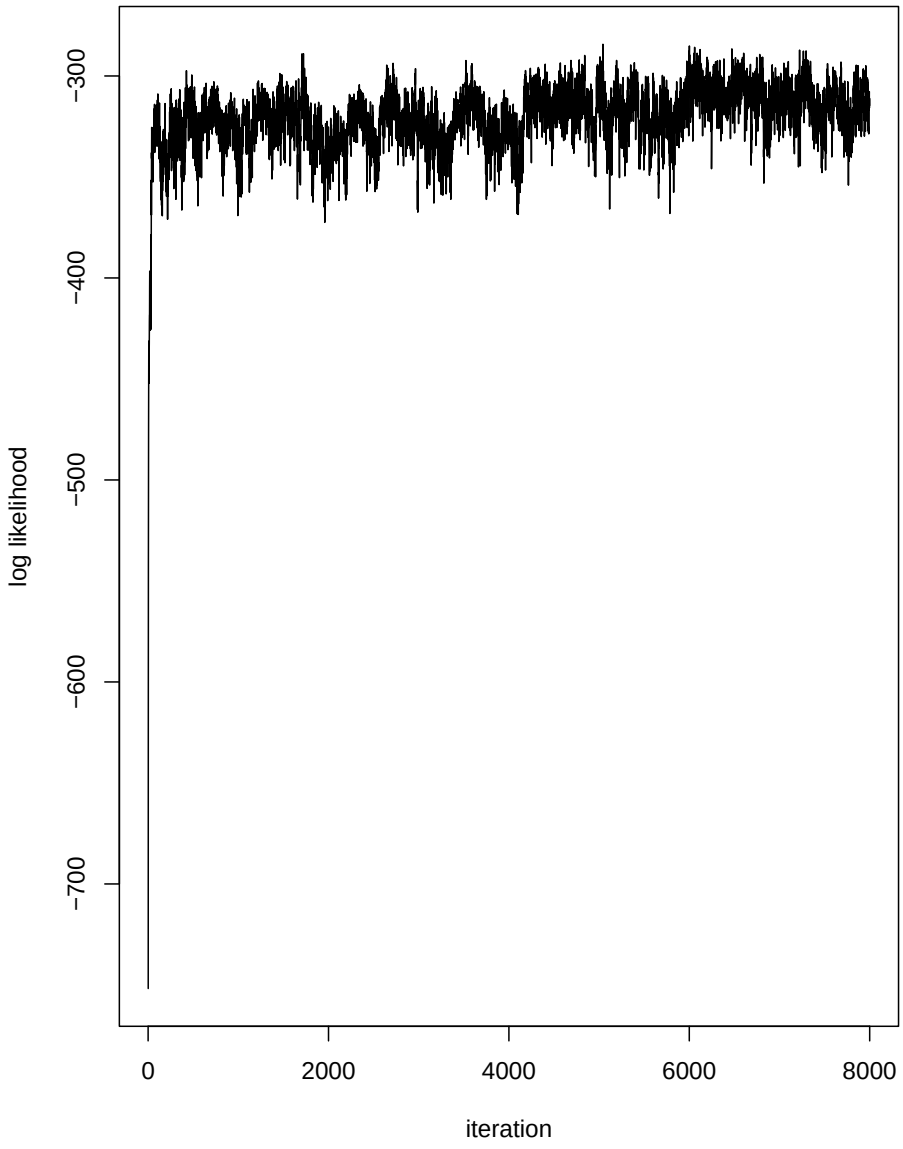

Figure 5: Observed log likelihood – P300 data (3-class)

4.4.3 Two Classes Mixture

Now a natural question is how well a 2-class mixture model fits the data; can the two upper classes with similar shape be combined? The following plot showed the result from the 2-class fit. Again an iteration plot of observed log likelihood is presented to show the convergence, see figure 7.

Table 11: Two-class mixture – P300 data

Mixing Proportions	Class 1	Class 2
MCMC algorithm	12.94%	87.06%

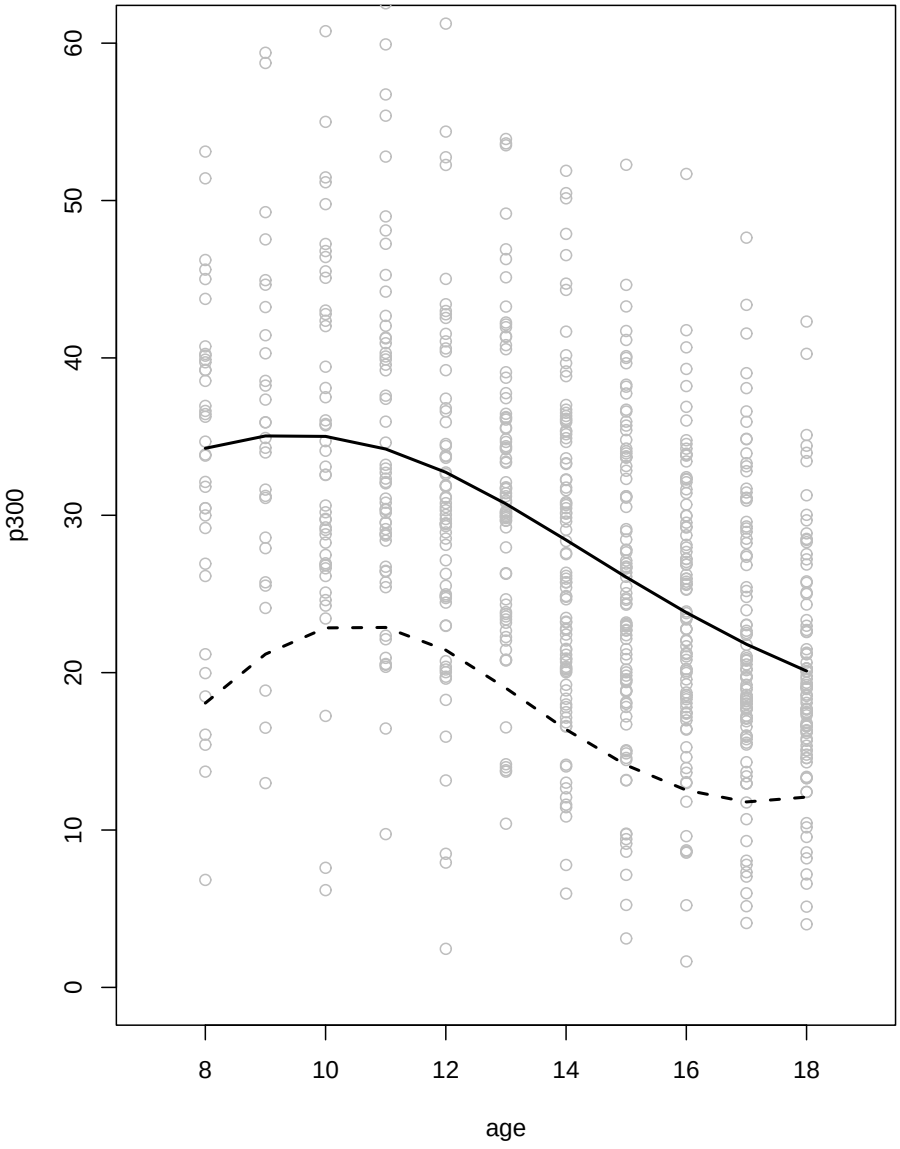

Figure 6: Two-class mixture – P300 data: dashed line class 1; solid line class 2

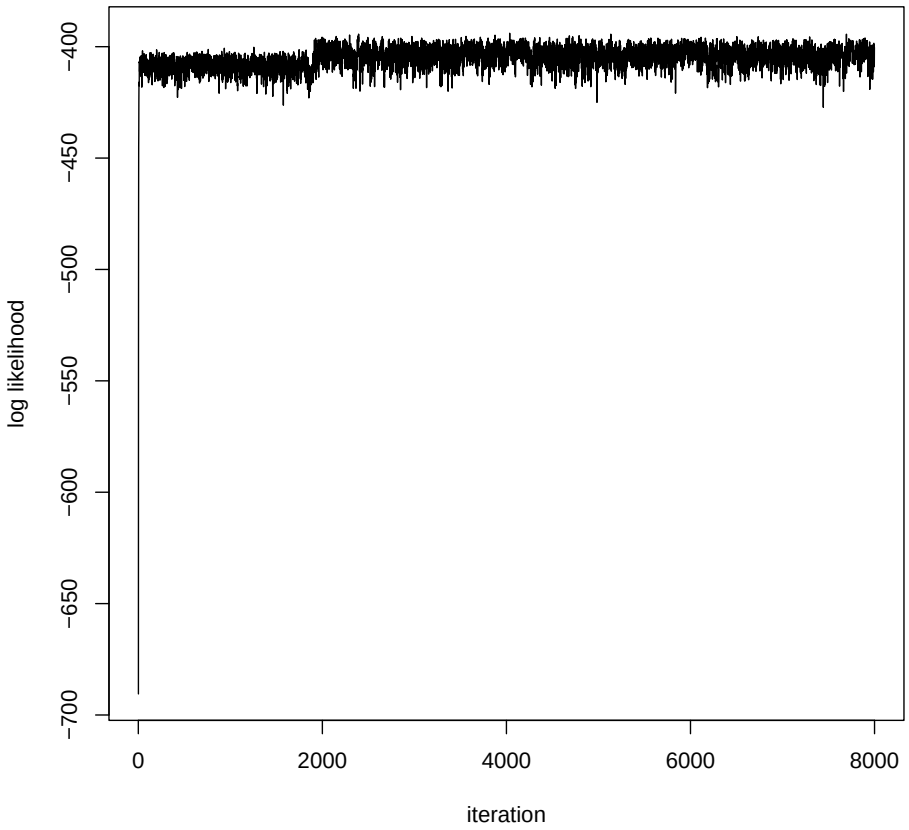

Figure 7: Observed log likelihood – P300 data (2-class)

4.4.4 Conclusion

The majority class seem to preserve the shapes of the 2 upper classes in the 3-class mixture. We calculate the *Bayesian Information Criterion* (BIC) to determine which model fits the data better, where BIC$=-2(logL)+log(n)d$. In the 3-class mixture, $BIC = 1057.7$, while in 2-class mixture $BIC = 1112.2$. BIC takes the number of parameters into consideration, so it tends to penalize more when the number of mixture increases. Since the BIC score for 3-class mixture is slightly smaller, we conclude that the 3-class mixture fits the data better than the 2-class mixture. This finding is also consistent with the results from earlier paper by Hill. et al. [17].

4.5 FREE-KNOT CUBIC SPLINE

In the previous section, the location of the knot was chosen based on prior knowledge about the curve. It is not hard to imagine situations where the location of the knot is unknown. In such cases, the location of the knot is also a free parameter.

Dimatteo et al. described a fully Bayesian method for curve-fitting with free knot splines for data drawn from an exponential family distribution [9]. We tailored that approach to the latent mixture model.

Recall that:

$$\begin{aligned} y_i &= B_t\eta_i + \epsilon_i \\ \eta_i &= Ac_i + \Gamma x_i + \xi_i \\ \log(\frac{p_{ik}}{p_{iK}}) &= \gamma_k^{'}\tilde{x}_i \end{aligned}$$

Note that the B-Spline basis matrix B_t replaced the second degree polynomial basis matrix T. Moreover, the value of the basis matrix depends on the location of the knot t.

4.5.1 Sampling Scheme

The augmented likelihood is the following:

$$
\begin{aligned}
L &= \prod_{i=1}^{n} f\left(y_i|\eta_i\right) f\left(\eta_i|c_i,x_i\right) f\left(c_i|x_i\right) \\
&= \prod_{i=1}^{n}\left[\phi\left(B_t\eta_i,\Sigma_\epsilon\right)\phi\left(Ac_i+\Gamma x_i,\Psi\right)\prod_{k=1}^{K} p_{ik}^{c_{ik}}\right] \\
&\propto \prod_{i=1}^{n}\left\{|\Psi|^{-1/2}\exp\left[-\frac{1}{2}\left(\eta_i-Ac_i-\Gamma x_i\right)'\Psi^{-1}\left(\eta_i-Ac_i-\Gamma x_i\right)\right]\right. \\
&\quad \left.\prod_{j=1}^{n_i}\left(\sigma_j^{-1}\right)\exp\left(-\frac{\left(y_{ij}-\left(B_t\right)_j'\eta_i\right)^2}{2\sigma_j^2}\right)\prod_{k=1}^{K}\left[\frac{\exp\left(\gamma_k'\tilde{x}_i\right)}{\sum_{l=1}^{K}\exp\left(\gamma_l'\tilde{x}_i\right)}\right]^{c_{ik}}\right\} \\
&\propto |\Psi|^{-\frac{n}{2}}\prod_{i=1}^{n}\left\{\prod_{k=1}^{K}\left\{\exp\left[-\frac{1}{2}\left(\eta_i-A_k-\Gamma x_i\right)'\Psi^{-1}\left(\eta_i-A_k-\Gamma x_i\right)\right]\frac{\exp\left(\gamma_k'\tilde{x}_i\right)}{\sum_{l=1}^{K}\exp\left(\gamma_l'\tilde{x}_i\right)}\right\}^{c_{ik}}\right. \\
&\quad \left.\prod_{j=1}^{n_i}\left(\sigma_j^{-1}\right)\exp\left(-\frac{\left(y_{ij}-\left(B_t\right)_j'\eta_i\right)^2}{2\sigma_j^2}\right)\right\}
\end{aligned}
$$

The parameter vector for this model is: $\left\{\left\{\sigma_j^2\right\}_{j=1}^{n_i}, A, \Gamma, \Psi, \left\{\gamma_k\right\}_{k=1}^{K-1}, t\right\}$. Adding an extra parameter t only affects the posterior distribution of $\left(\sigma_j^2\right)_{j=1}^{n_i}$, the missing value $y_{i^*j^*}$ and η_i. We now discuss the Gibbs sampler and Metropolis-Hastings algorithm sampling scheme.

4.5.1.1 Prior distribution

In addition to the prior distributions in 3.1.2, we assign a uniform distribution on the parameter t:

$p\left(t\right) = U\left(t_1, t_{n_i}\right)$ t_1 and t_{n_i} are end points of the time interval

4.5.1.2 Gibbs sampler and M-H algorithm

In Gibbs sampler, the posterior for σ_j^2 changes to:

$$
\begin{aligned}
p\left(\sigma_j^2|a_0,b_0,\left\{\eta_i\right\},\left\{y_{ij}\right\},B_t\right) &\propto \left(\sigma_j^2\right)^{-a_0-\frac{n}{2}-1}\exp\left\{-\frac{1}{\sigma_j^2}\left[\frac{1}{2}\sum_{i=1}^{n}\left(y_{ij}-\left(B_t\right)_j'\eta_i\right)^2+b_0\right]\right\} \\
&\sim IG\left(a_0+\frac{n}{2},\ \frac{1}{2}\sum_{i=1}^{n}\left(y_{ij}-\left(B_t\right)_j'\eta_i\right)^2+b_0\right)
\end{aligned}
$$

for $j = 1, \cdots, n_i$

The posterior distribution of t is:

$$p\left(t|\left\{\eta_i\right\},\left\{y_{ij}\right\},\left\{\sigma_j^2\right\}\right) \quad \propto \quad \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(y_i - B_t\eta_i\right)'\Sigma_\epsilon^{-1}\left(y_i - B_t\eta_i\right)\right\}$$

which obviously is not an easy distribution function of t. Therefore, we use the Metropolis-Hastings algorithm to obtain random sample for t. Since t can only be in the interval (t_0, t_{n_i}), we propose two candidate densities from exponential family to approximate the posterior distribution.

- Rescaled Beta Distribution: the candidate density is a beta distribution centered at the current value of t with certain variance.
  - Draw a random sample $t^{(*)}$ from beta $(t\nu, (1-t)\nu)$, which has mean t and variance $\frac{t(1-t)}{\nu+1}$. ν is a tuning parameter which can be adjusted to control the variance. In our study, ν takes on values around 40.
  - Calculate the acceptance probability

$$\alpha\left(t^{(*)}, t\right) = \min\left\{\frac{p(t^{(*)})q(t|t,\nu)}{p(t)q(t^{(*)}|t,\nu)},\ 1\right\}$$

    where $p(t)$ is the posterior distribution of t as in , $q(t|t,\nu) \propto t^{t\nu-1}(1-t)^{(1-t)\nu-1}$
  - Update the Markov chain to $t^{(*)}$ with probability α.
- Truncated Normal Distribution: the candidate density is a normal distribution centered at the current value of t and truncated at end points of the time domain.
  - Draw a random sample $t^{(*)}$ from $N(t, V)$. Variance V is a tunning parameter. In our algorithm, it takes on value around 2.
  - Calculate the acceptance probability

$$\alpha\left(t^{(*)}, t\right) = \min\left\{\frac{p(t^{(*)})q(t|t,V)}{p(t)q(t^{(*)}|t,V)},\ 1\right\}$$

    where $q(t^*|t,V) \propto \exp\left\{-\frac{1}{2}(t^*-t)^2/V^2\right\}$
  - Update the Markov Chain of t to $t^{(*)}$ with probability α.

Using the M-H algorithm described above, the complete Gibbs sampling procedure for latent mixture model using free-knot cubic spline is the following:
Given the initial values: $\left(c_i^{(0)}, \eta_i^{(0)}, y_{i^*j^*}^{(0)}, \sigma_j^{(0)}, A^{(0)}, \Gamma^{(0)}, \Psi^{-1(0)}, \gamma_k^{(0)}, t^{(0)}\right)$. where $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $k = 1, \ldots, K$.

The sampling scheme iterates the following steps:

1. Sample $c_i = (c_{i1}, \ldots, c_{iK})^{'}$ from multinomial $(1,\ (p_{i1}, \ldots, p_{iK})^{'})$, where

$$p_{ik} = \frac{\exp\left\{-\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)^{'} \Psi^{-1}\left(\eta_i - A_k - \Gamma x_i\right)\right\} \exp\left(\gamma_k^{'} \tilde{x}_i\right)}{\sum_{k=1}^{K} \exp\left\{-\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)^{'} \Psi^{-1}\left(\eta_i - A_k - \Gamma x_i\right)\right\} \exp\left(\gamma_k^{'} \tilde{x}_i\right)}$$

   for $i = 1, \ldots, n,\ k = 1, \ldots, K$. Again, $\gamma_K^{(0)}$ and γ_K are fixed at 0.
2. Sample $y_{i^*j^*}$ from $N\left((B_t)_{j^*}^{'} \eta_{i^*},\ \sigma_{j^*}^2\right)$
3. Sample η_i from $N\left(\Sigma_\eta\left[\Psi^{-1}\left(Ac_i + \Gamma x_i\right) + B_t^{'} D^{-1} y_i\right],\ \Sigma_\eta\right)$
   $D = \texttt{diag}\left(\sigma_1^2, \ldots, \sigma_{n_i}^2\right)$, $\Sigma_\eta = \left(\Psi^{-1} + B_t^{'} D^{-1} B_t\right)^{-1}$.
4. Sample σ_j^2 from $IG\left(a_0 + \frac{n}{2},\ \frac{1}{2}\sum_{i=1}^{n}\left(y_{ij} - (B_t)_j^{'}\eta_i\right)^2 + b_0\right)$, $j = 1, \ldots, n_i$.
5. Sample A_k from $N\left(\Sigma_A \Psi^{-1} \sum_{i=1}^{n} c_{ik}\left(\eta_i - \Gamma x_i\right),\ \Sigma_A\right)$, $k = 1, \ldots, K$.
6. Sample Γ_q from $N\left(\Sigma_\Gamma \sum_{i=1}^{n} x_{iq} \Psi^{-1}\left(\eta_i - Ac_i - \sum_{r \neq q} x_{ir}\Gamma_r\right),\ \Sigma_\Gamma\right)$, $q = 1, \ldots, Q$.
7. Sample Ψ^{-1} from $WI\left(n + v,\ \left(\sum_{i=1}^{n} \alpha_i \alpha_i^{'} + S^{-1}\right)^{-1}\right)$
8. Sample γ_k via the Metropolis-Hastings algorithm described in section 3.2.1.
9. Sample t via the Metropolis-Hastings algorithm described above.

4.5.2 Simulation Study

To test the hybrid MCMC algorithm, we ran simulation studies using both the beta and truncated normal distributions as the candidate densities. We present a case where we generated three cubic splines on the interval of $(8, 18)$. A single knot was chosen to be 13.5. Again, the simulated data set has same amount of missing data percentage as the P300 data. The following tables show the estimates for selected parameters, for a complete list of estimates, see appendix. Figures showing the fitted spline curves are also shown below.

Using truncated normal distribution as the candidate density:

Table 12: Mixing proportions – free knot spline (truncated normal)

Mixing Proportion	Class 1	Class 2	Class 3
True Value	22.92%	64.81%	12.27%
MCMC Method	21.17%	65.30%	13.52%

Table 13: Knot location – free knot spline (truncated normal)

knot	True	5%	Estimate	95%	S.D.
t	13.5	13.4213	13.5703	13.7154	0.0934

Table 14: $\hat{A}$ – free knot spline (truncated normal)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
A_{11}	12	11.1708319	13.7283714	16.8506389	1.696141
A_{21}	-6	-4.5270101	-0.8462035	3.1240802	2.318974
A_{31}	8	0.3337492	4.2388006	8.2927925	2.469080
A_{41}	-4	-5.7134314	-3.1476607	-0.2202592	1.658824
A_{51}	6	3.5843491	5.7759939	8.0558584	1.296931
A_{12}	15	13.9624362	14.872854	15.8622766	0.5699230
A_{22}	20	18.3260861	19.774648	21.1369606	0.8550956
A_{32}	-2	-4.1360351	-3.002130	-2.0405981	0.6374539
A_{42}	10	9.8723769	10.683996	11.4376022	0.4706184
A_{52}	11	10.3721068	10.937952	11.4412316	0.3221205
A_{13}	19	17.2628774	19.377193	21.6487260	1.322247
A_{23}	4	1.7125507	4.789802	8.2920609	2.007608
A_{33}	2	-1.4682981	2.020994	4.7842285	1.897184
A_{43}	10	8.1465558	9.864022	11.7365675	1.090248
A_{53}	2	0.1643264	1.469370	2.7316290	0.760857

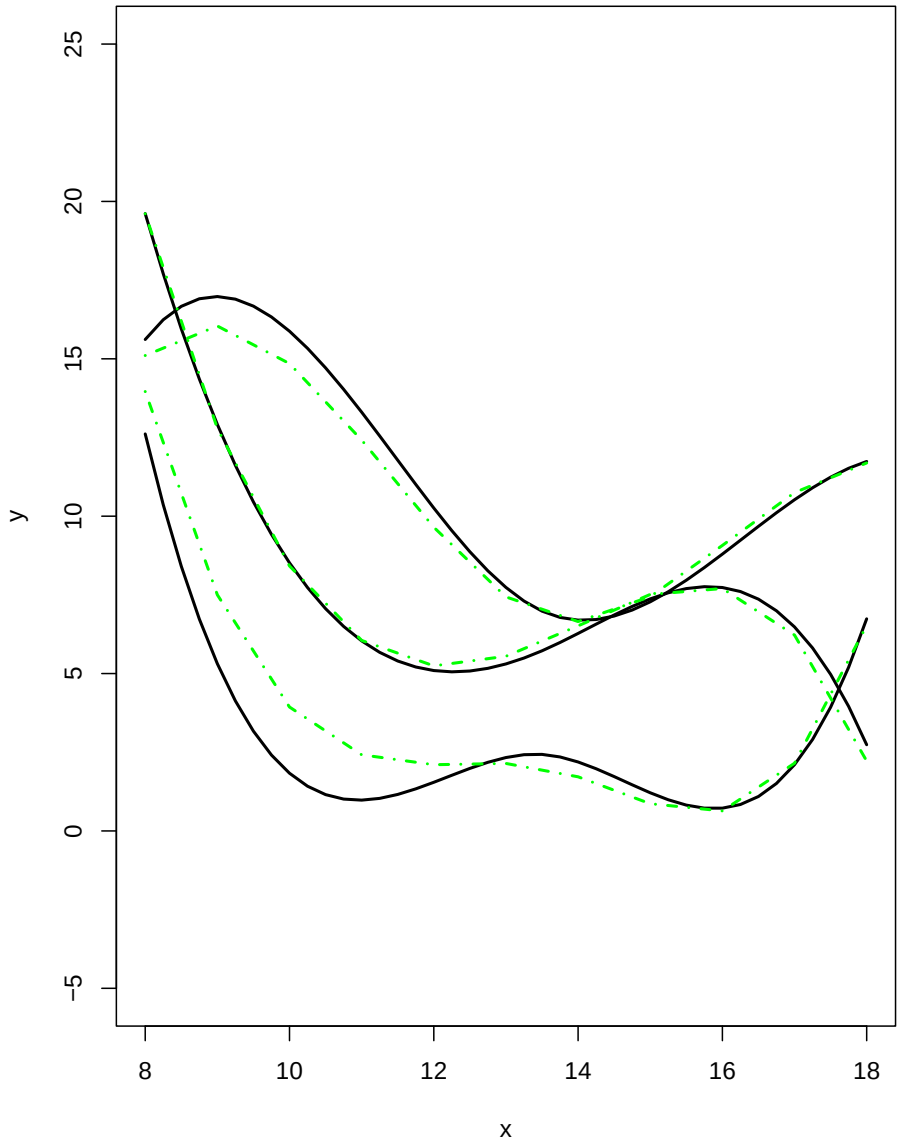

Figure 8: solid lines: generated curves; dashed lines: free knot spline (truncated normal)

Table 15: Mixing Proportions – free knot spline (beta)

Mixing Proportion	Class 1	Class 2	Class 3
True Value	22.92%	64.81%	12.27%
MCMC Method	21.90%	64.29%	13.81%

Table 16: Knot location – free knot spline (beta)

knot	True	5%	Estimate	95%	S.D.
t	13.5	13.4968	13.6952	13.8874	0.0928

Table 17: $\hat{A}$ – free knot spline (beta)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
A_{11}	12	8.4460017	11.2191761	14.0128535	1.490795
A_{21}	-6	-4.9502522	-1.9921842	1.7322297	1.878599
A_{31}	8	1.6699078	6.0913502	10.5989925	2.281091
A_{41}	-4	-4.7127975	-0.6722037	4.8174379	2.268255
A_{51}	6	-0.9532512	6.9259549	10.8814017	2.582900
A_{12}	15	14.1756547	14.894371	15.5963046	0.3757461
A_{22}	20	19.1698777	20.499688	22.2654261	0.8282394
A_{32}	-2	-3.5506547	-1.925634	-0.4151692	0.8264122
A_{42}	10	8.1842763	9.519120	10.7530173	0.6542392
A_{52}	11	10.3539654	11.027033	11.7478272	0.3491679
A_{13}	19	16.7017987	18.812027	20.6085828	0.9920208
A_{23}	4	5.0687347	7.544126	10.5683583	1.4897026
A_{33}	2	-3.9384831	-1.213159	1.5731351	1.4523661
A_{43}	10	8.9945348	11.162569	13.3823585	1.1314062
A_{53}	2	1.4657984	3.312512	4.9325699	0.8687165

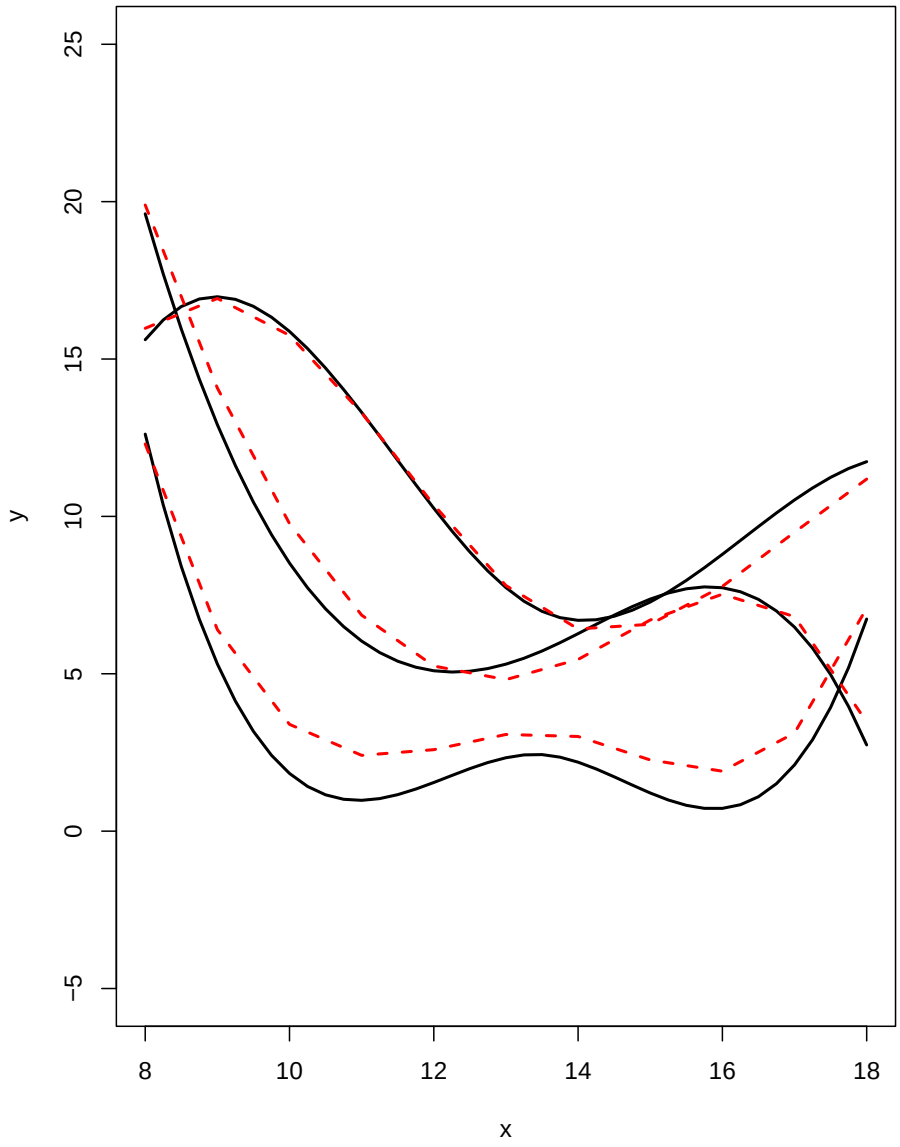

Figure 9: solid lines: generated curves; dashed lines: free knot spline fitting (beta)

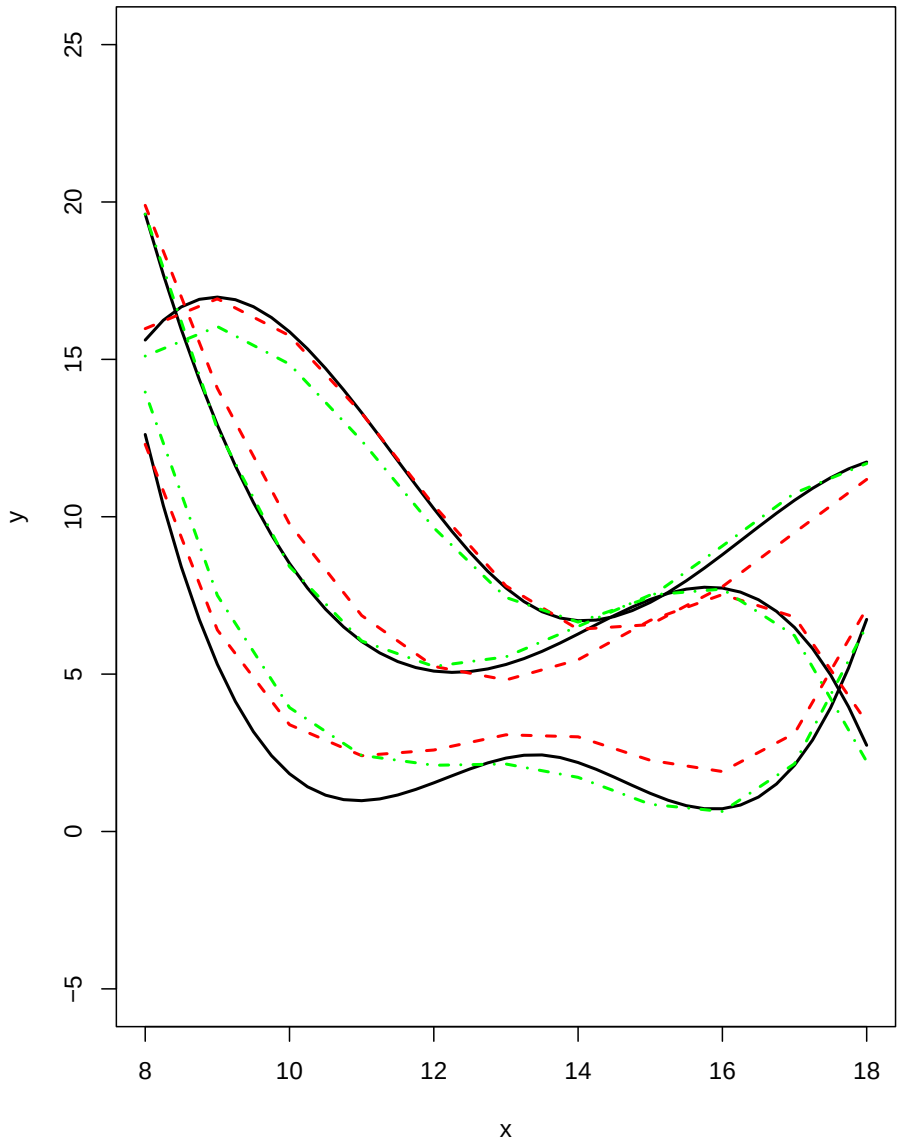

Figure 10: solid lines: generated curves; dashed lines: fitted cubic splines

The last figure shows the curve fitting using both the truncated normal and rescaled beta distribution as the candidate density. In our data set, they both perform very well.

5.0 OTHER APPLICATION AND BIVARIATE EXTENSION

This chapter consists of two separate parts. First, we discuss the use of latent mixture model to predict a binary outcome; second, we extend the latent mixture model to the bivariate case. For both problems, we developed the model and a way to fit the model to the data. The computational applications remain to be completed in the future.

5.1 PREDICT A BINARY OUTCOME

Prediction for a binary outcome is useful in our application, with the subject developing alcoholism in adulthood represented by a binary variable u_i. Ordinary logistic regression can relate this variable to the latent class membership c_i.

Let $w_i = P(u_i = 1)$, so the latent mixture model with prediction for the binary outcome is:

$$\begin{aligned} y_i &= B_t \eta_i + \epsilon_i \\ \eta_i &= A c_i + \Gamma x_i + \xi_i \\ \log(\frac{p_{ik}}{p_{iK}}) &= \gamma_k^{'} \tilde{x}_i \\ \log(\frac{w_i}{1-w_i}) &= \lambda^{'} c_i \end{aligned}$$

The augmented likelihood for this model is:

$$\begin{aligned} L &= \prod_{i=1}^{n} f\left(y_i|\eta_i\right) f\left(\eta_i|c_i,x_i\right) f\left(u_i|c_i\right) f\left(c_i|x_i\right) \\ &\propto |\Psi|^{-\frac{n}{2}} \prod_{i=1}^{n} \left\{ \prod_{k=1}^{K} \left\{ \exp\left[-\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right] \frac{\exp\left(\gamma_k' \tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)} \right\}^{c_{ik}} \right. \\ &\quad \left. \frac{\exp\left(\lambda' c_i\right)^{u_i}}{1+\exp\left(\lambda' c_i\right)} \prod_{j=1}^{n_i} \left(\sigma_j^{-1}\right) \exp\left(-\frac{\left(y_{ij} - \left(B_t\right)_j' \eta_i\right)^2}{2\sigma_j^2} \right) \right\}. \end{aligned}$$

5.2 SAMPLING DISTRIBUTION

It is clear that u_i contains additional information about latent class membership. The sampling distribution for c_i is changed accordingly:

$$\begin{aligned} & p\left(c_i|u_i,x_i\right) \\ &\quad \propto \prod_{k=1}^{K} \left\{ \exp\left\{ -\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right\} \frac{\exp\left(\gamma_k' \tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)} \right\}^{c_{ik}} \frac{\exp\left(\lambda' c_i\right)^{u_i}}{1+\exp\left(\lambda' c_i\right)} \\ &\quad \propto \prod_{k=1}^{K} \left\{ \exp\left\{ -\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right\} \frac{\exp\left(\gamma_k' \tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left(\gamma_l' \tilde{x}_i\right)} \frac{\exp\left(u_i \lambda_k\right)}{1+\exp\left(\lambda\right)} \right\}^{c_{ik}} \end{aligned}$$

Thus, c_i can be sampled from a multinomial distribution with parameter 1 and $p_i = (p_{i1}, \cdots, p_{iK})$, where

$$p_{ik} = \frac{\exp\left\{ -\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right\} \exp\left(\gamma_k' \tilde{x}_i\right) \frac{\exp(u_i \lambda_k)}{1+\exp(\lambda_k)}}{\sum_{k=1}^{K} \exp\left\{ -\frac{1}{2}\left(\eta_i - A_k - \Gamma x_i\right)' \Psi^{-1} \left(\eta_i - A_k - \Gamma x_i\right) \right\} \exp\left(\gamma_k' \tilde{x}_i\right) \frac{\exp(u_i \lambda_k)}{1+\exp(\lambda_k)}},$$

for $i = 1, \cdots, n$, and $k = 1, \cdots, K$. γ_K is fixed at 0.

To sample λ, we gave independent prior $N\left(0, \frac{1}{\sigma_\lambda^2} I\right)$. Thus the posterior distribution of λ is:

$$p\left(\lambda|\sigma_\lambda^2, \{c_i\}, \{u_i\}\right) \propto \exp\left\{ -\frac{\lambda'\lambda}{2\sigma_\lambda^2} \right\} \prod_{i=1}^{n} \left(\frac{\exp\left(\lambda' c_i\right)}{1+\exp\left(\lambda' c_i\right)} \right)^{u_i} \left(\frac{1}{1+\exp\left(\lambda' c_i\right)} \right)^{1-u_i}$$

We propose the Metropolis-Hastings algorithm similar to section 3.2.1 to sample λ:

i. Obtain $m_0 = \arg\max_{\lambda} \log p(\lambda)$ via Newton-Raphson algorithm by using quantities:

$$\frac{\partial \log p(\lambda)}{\partial \lambda} = -\frac{\lambda}{\sigma_\lambda^2} + \sum_{i=1}^{n} \left(u_i - \frac{\exp(\lambda' c_i)}{1 + \exp(\lambda' c_i)} \right) c_i \tag{5.1}$$

$$\frac{\partial^2 \log p(\lambda)}{\partial \lambda \partial \lambda'} = -\frac{1}{\sigma_\lambda^2} I - \sum_{i=1}^{n} \frac{\exp(\lambda' c_i)}{(1 + \exp(\lambda' c_i))^2} c_i c_i' \tag{5.2}$$

ii. Calculate $V_0 = \tau \left(-\frac{\partial^2 \log p(\lambda)}{\partial \lambda \partial \lambda'} |_{m_0} \right)^{-1}$, where τ is a tuning parameter. It can be adjusted to control the dispersion of the candidate density and the acceptance rate of the random draws. The candidate density is then specified as:

$$q(\lambda | m_0, V_0, \tau) \propto |V_0|^{-1/2} \exp\left\{ -\frac{1}{2} (\lambda - m_0)' V_0^{-1} (\lambda - m_0) \right\}, \tag{5.3}$$

which is a multivariate normal distribution with mean m_0 and covariance matrix V_0.

iii. Draw a random deviate $\lambda^{(*)}$ from the candidate density, update the Markov chain of λ from current value to $\lambda^{(*)}$ with probability

$$\alpha(\lambda^{(*)}, \lambda) = \min \left\{ \frac{p(\lambda^{(*)}) q(\lambda | m_0, V_0, \tau)}{p(\lambda) q(\lambda^{(*)} | m_0, V_0, \tau)},\ 1 \right\}$$

Therefore the Gibbs sampler for the latent mixture model with prediction for a binary outcome iterates between the following steps. Given the initial values:
$\left(c_i^{(0)}, \eta_i^{(0)}, y_{i^* j^*}^{(0)}, \sigma_j^{(0)}, A^{(0)}, \Gamma^{(0)}, \Psi^{-1(0)}, \gamma_k^{(0)}, t^{(0)} \right)$, where $i = 1, \ldots, n, \quad j = 1, \ldots, n_i$, $k = 1, \ldots, K$. The sampling scheme iterates the following steps:

1. Sample $c_i = (c_{i1}, \ldots, c_{iK})'$ from multinomial $(1,\ (p_{i1}, \ldots, p_{iK})')$ as in 5.1.
2. Sample $y_{i^* j^*}$ from $N\left((B_t)'_{j^*} \eta_{i^*},\ \sigma_{j^*}^2 \right)$
3. Sample η_i from $N\left(\Sigma_\eta \left[\Psi^{-1} (A c_i + \Gamma x_i) + B_t' D^{-1} y_i \right],\ \Sigma_\eta \right)$,
   $D = \mathtt{diag}\left(\sigma_1^2, \ldots, \sigma_{n_i}^2 \right)$, $\Sigma_\eta = \left(\Psi^{-1} + B_t' D^{-1} B_t \right)^{-1}$.
4. Sample σ_j^2 from $IG\left(a_0 + \frac{n}{2},\ \frac{1}{2} \sum_{i=1}^{n} \left(y_{ij} - (B_t)'_j \eta_i \right)^2 + b_0 \right)$ $j = 1, \ldots, n_i$.
5. Sample A_k from $N\left(\Sigma_A \Psi^{-1} \sum_{i=1}^{n} c_{ik} (\eta_i - \Gamma x_i),\ \Sigma_A \right)$ $k = 1, \ldots, K$.
6. Sample Γ_q from $N\left(\Sigma_\Gamma \sum_{i=1}^{n} x_{iq} \Psi^{-1} \left(\eta_i - A c_i - \sum_{r \neq q} x_{ir} \Gamma_r \right),\ \Sigma_\Gamma \right)$ $q = 1, \ldots, Q$.
7. Sample Ψ^{-1} from $WI\left(n + v,\ \left(\sum_{i=1}^{n} \alpha_i \alpha_i' + S^{-1} \right)^{-1} \right)$
8. Sample γ_k via Metropolis-Hastings algorithm described in section 3.2.1.
9. Sample λ vis Metropolis-Hastings algorithm described above.

5.3 BIVARIATE LATENT MIXTURE MODEL

There are many times when researchers measure more than one response variable at each time point. The resulting developmental pathways often are closely related, so that two responses together may provide more information than one response variable alone. In the ERP study, there are other ERP components being recorded along with P300, and those components are the candidates for the secondary response variable.

In order to reduce the complexity of the model, only the intercepts of estimated curves are considered random. That is, the intercept term is individual specific. We continue to use the notations in the previous sections, and let y_{ij1}, y_{ij2} be the two responses at jth time point for subject i respectively. Then the bivariate response variable can be written as: $y_{ij} = \begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix}$.

We model the developmental trajectory of y_{ij} as the following:

$$y_{ij} = \alpha_i + B^{'} T_j + \epsilon_{ij}$$

Parameter specification for this model is the following:

- $B = (\beta_1, \beta_2)$ is a parameter matrix contains columns of polynomial coefficients (slope and quadratic term) for each ERP component.
- $\alpha_i = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \end{pmatrix}$ is a 2-dimensional random coefficient vector, which plays the role of a random intercept for each ERP component trajectory. Again, its believed that there are K underlying trajectory classes. To depict the population heterogeneity, we assume α_i's are sampled from a mixture of bivariate normal distributions with mixing probability p_{ik}.

$$\alpha_i \sim \sum_{k=1}^{K} p_{ik} N_2 \left(\mu_k, \Sigma_\alpha\right).$$

where $\mu_k = \begin{pmatrix} \mu_{k1} \\ \mu_{k2} \end{pmatrix}$, $\Sigma_\alpha = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix}$. At this point we put no further constraints on the covariance matrix structure.

- Let $I_i = (I_{i1}, \ldots, I_{iK})^{'}$ be the class indicator for subject i. Hence it has a multinomial distribution with parameter 1 and $(p_{i1}, \ldots, p_{iK})^{'}$, with mixing proportions modeled as:

$$\text{logit}\,(p_{ik}) = \log\left(\frac{p_{ik}}{p_{iK}}\right) = \eta_k^{'}\tilde{x}_i \quad k = 1, \ldots, K-1 \tag{5.4}$$

  It follows that $p_{ik} = \frac{\exp(\eta_k^{'}\tilde{x}_i)}{\sum_{l=1}^{K}\exp(\eta_l^{'}\tilde{x}_i)}$, with constraints: $\eta_K = 0$, and $\sum_{k=1}^{K} p_{ik} = 1$.
- ϵ_{ij} is a residual term that is independent from other variables,

$$\epsilon_{ij} \sim N_2\left(0,\, \sigma_\epsilon^2 I\right)$$

5.3.1 Conditional Distribution and Augmented Likelihood

We propose a MCMC algorithm similar to section 2 to fit the model. First we will give the conditional distribution of each parameter and later we will present the sampling scheme to be used in MCMC algorithm.

- $f\,(I_i|\tilde{x}_i) \sim \texttt{Multinomial}\left(1,\, (p_{i1}, \ldots, p_{iK})^{'}\right), \quad p_{ik} = \frac{\exp(\eta_k^{'}\tilde{x}_i)}{\sum_{l=1}^{K}\exp(\eta_l^{'}\tilde{x}_i)}$
- $f\,(\alpha_i|I_i) \sim N_2\,(\mu_i,\, \Sigma_\alpha)\,, \quad \mu_i = \mu I_i \quad \mu = (\mu_1, \cdots, \mu_K)$
- $f\,(y_{ij}|\alpha_i) \sim N_2\,(v_{ij},\, \sigma_\epsilon^2 I)\,, \quad v_{ij} = \alpha_i + B^{'}T_j$

The augmented likelihood therefore is:

$$\begin{aligned}
L &= \prod_{i=1}^{n}\prod_{j=1}^{n_i} f\,(y_{ij}|\alpha_i)\, f(\alpha_i|I_i) f(I_i|\tilde{x}_i) \\
&\propto \prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{1}{\sigma_\epsilon^2}\exp\left\{-\frac{(y_{ij1} - \beta_1^{'}T_j - \alpha_{i1})^2}{2\sigma_\epsilon^2}\right\}\exp\left\{-\frac{(y_{ij2} - \beta_2^{'}T_j - \alpha_{i2})^2}{2\sigma_\epsilon^2}\right\} \\
&\quad |\Sigma_\alpha|^{-1/2}\exp\left\{-\frac{1}{2}\left(\alpha_i - \mu I_i\right)^{'}\Sigma_\alpha^{-1}\left(\alpha_i - \mu I_i\right)\right\}\prod_{k=1}^{K}\left(\frac{\exp(\eta_k^{'}\tilde{x}_i)}{\sum_{l=1}^{K}\exp(\eta_l^{'}\tilde{x}_i)}\right)^{I_{ik}}
\end{aligned}$$

5.3.2 Prior and Posterior Distribution

The parameters to be estimated are $\{ B, \sigma_\epsilon^2, \{\mu_k\}, \Sigma_\alpha, \{\eta_k\} \}$, $k = 1, \ldots, K$. Again, we assign the following independent priors:

$p(\beta_1) = p(\beta_2) \sim N\left(0,\, \sigma_\beta^2 I\right)$

$p(\sigma_\epsilon^2) \sim IG(a_1,\, b_1)$

$p(\Sigma_\alpha^{-1}) \sim WI(v_1,\, S_1)$

$p(\mu_k) \sim N\left(0,\, \sigma_\mu^2 I\right), \quad k = 1, \ldots, K$

$p(\eta_k) \sim N\left(0,\, \sigma_\eta^2 I\right), \;\; k = 1, \ldots, K-1$

To obtain the posterior distributions, we collect the corresponding terms in the augmented likelihood and multiply it by the priors:

a.

$$
\begin{aligned}
& p\left(\beta_1 | \sigma_\beta^2, \sigma_\epsilon^2, \{y_{ij1}\}, \{\alpha_{i1}\}\right) \\
& \quad \propto \quad \exp\left\{ -\frac{1}{2}\left[\beta_1' \left(\frac{n}{\sigma_\epsilon^2} \sum_{j=1}^{n_i} T_j T_j' + \frac{1}{\sigma_\beta^2} I \right) \beta_1 - \frac{2\beta_1'}{\sigma_\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} T_j \left(y_{ij1} - \alpha_{i1} \right) \right] \right\} \\
& \quad \sim \quad N\left(\frac{\Sigma_\beta}{\sigma_\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} T_j \left(y_{ij1} - \alpha_{i1} \right),\; \Sigma_\beta \right) \qquad (5.5)
\end{aligned}
$$

$$
p\left(\beta_2 | \sigma_\beta^2, \sigma_\epsilon^2, \{y_{ij2}\}, \{\alpha_{i2}\}\right) \sim N\left(\frac{\Sigma_\beta}{\sigma_\epsilon^2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} T_j \left(y_{ij2} - \alpha_{i2} \right),\; \Sigma_\beta \right) \qquad (5.6)
$$

Where $\Sigma_\beta = \left(\frac{n}{\sigma_\epsilon^2} \sum_{j=1}^{n_i} T_j T_j' + \frac{1}{\sigma_\beta^2} I \right)^{-1}$.

b.

$$
\begin{aligned}
& p\left(\sigma_\epsilon^2 | a_1, b_1, B, \{y_{ij}\}, \{\alpha_i\}\right) \\
& \quad \propto \quad (\sigma_\epsilon^2)^{-(a_1+1)} \exp\left\{ -\frac{b_1}{\sigma_\epsilon^2} \right\} \prod_{i=1}^{n} \prod_{j=1}^{n_i} (\sigma_\epsilon^2)^{-1} \exp\left\{ -\frac{\left(y_{ij} - B' T_j - \alpha_i \right)' \left(y_{ij} - B' T_j - \alpha_i \right)}{2\sigma_\epsilon^2} \right\} \\
& \quad = \quad (\sigma_\epsilon^2)^{-(a + n \cdot n_i + 1)} \exp\left\{ -\frac{1}{\sigma_\epsilon^2} \left[b + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - B' T_j - \alpha_i \right)' \left(y_{ij} - B' T_j - \alpha_i \right)}{2} \right] \right\} \\
& \quad \sim \quad IG\left(a + n \cdot n_i,\; b + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(y_{ij} - B' T_j - \alpha_i \right)' \left(y_{ij} - B' T_j - \alpha_i \right)}{2} \right) \qquad (5.7)
\end{aligned}
$$

c.

$$
\begin{aligned}
& p\left(\mu_k|\sigma_\mu^2, \Sigma_\alpha, \{\alpha_{i1}\}, \{\alpha_{i2}\}, \{\eta_k\}\right) \\
& \propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\left(\alpha_i-\mu_k\right)^{'}\Sigma_\alpha^{-1}\left(\alpha_i-\mu_k\right)\right\}^{I_{ik}} \exp\left\{-\frac{\gamma_{k1}^{'}\gamma_{k1}}{2\sigma_\gamma^2}\right\} \\
& = \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} I_{ik}\left(\alpha_i-\mu_k\right)^{'}\Sigma_\alpha^{-1}\left(\alpha_i-\mu_k\right) - \frac{\mu_k^{'}\mu_k}{2\sigma_\mu^2}\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\mu_k^{'}\left(\sum_{i=1}^{n} I_{ik}\Sigma_\alpha^{-1}+\frac{I}{\sigma_\mu^2}\right)\mu_k^{'} - 2\sum_{i=1}^{n}\alpha_i^{'}\Sigma_\alpha^{-1}\mu_k\right]\right\} \qquad (5.8) \\
& \sim N\left(\mu_{\alpha k},\, \Sigma_k\right) \qquad (5.9)
\end{aligned}
$$

$\mu_{\alpha k} = \Sigma_k \sum_{i=1}^{n} I_{ik}\alpha_i^{'}\Sigma_\alpha^{-1}$ $\Sigma_k = \left(\frac{1}{\sigma_\mu^2}I + \sum_{i=1}^{n} I_{ik}\Sigma_\alpha^{-1}\right)^{-1}$

Where $k = 1, \ldots, K$

d.

$$
\begin{aligned}
& p\left\{\Sigma_\alpha^{-1}|v_1, S_1, \{\alpha_i\}, \{\gamma_k\}, \{I_i\}\right\} \\
& \propto |\Sigma_\alpha^{-1}|^{\frac{v_1-3-1}{2}} \exp\left\{-\frac{1}{2}\mathtt{tr}\left(S_1^{-1}\Sigma_\alpha^{-1}\right)|\Sigma_\alpha^{-1}|^{\frac{n}{2}} + \left\{-\frac{1}{2}\left(\alpha_i-\mu I_i\right)^{'}\Sigma_\alpha^{-1}\left(\alpha_i-\mu I_i\right)\right\}\right\} \\
& \propto |\Sigma_\alpha^{-1}|^{\frac{v_1-3-1}{2}} \exp\left\{-\frac{1}{2}\mathtt{tr}\left[\left(S^{-1}+\left(\alpha_i-\mu I_i\right)\left(\alpha_i-\mu I_i\right)^{'}\right)\Sigma_\alpha^{-1}\right]\right\} \\
& \sim WI\left\{v_1+n,\ \left(S^{-1}+\left(\alpha_i-\mu I_i\right)\left(\alpha_i-\mu I_i\right)^{'}\right)^{-1}\right\} \qquad (5.10)
\end{aligned}
$$

e. The mixing proportion part of the bivariate latent mixture model is similar to the univariate case. Each subject's class membership is related to the covariates of that subject. The relationship is modeled by multinomial logistic regression:

$$
p\left(\eta_k|\sigma_\eta^2, \{I_i\}\right) \propto \exp\left\{-\frac{\eta_k^{'}\eta_k}{2\sigma_\eta^2}\right\}\prod_{i=1}^{n}\frac{\exp\left(\eta_k^{'}\tilde{x}_i\right)^{I_{ik}}}{\sum_{l=1}^{K}\exp\left(\eta_l^{'}\tilde{x}_i\right)}
$$

As in section 3.1.2, the vector η_K will be fixed at 0 in order for the multinomial model to be identifiable. The posterior distributions of η_k $k = 1, \ldots, K-1$ do not have simple form therefore sampling from their posteriors is very complicated. Metropolis-Hastings algorithm introduced in section 2 will be used to obtain a approximate random sample of η_k.

5.3.3 Sampling Scheme

In this section, we present Gibbs sampler and M-H algorithm sampling scheme for the parameters.

1. Sample $I_i = (I_{i1}, \ldots, I_{iK})'$ from Multinomial $\left\{1,\ (p_{i1}, \ldots, p_{iK})'\right\}$. where

$$p_{ik} = \frac{\exp\left\{-\frac{1}{2}(\alpha_i - \mu_k)' \Sigma_\alpha^{-1}(\alpha_i - \mu_k)\right\} \exp\left(\eta_k' \tilde{x}_i\right)}{\sum_{l=1}^{K} \exp\left\{-\frac{1}{2}(\alpha_i - \mu_l)' \Sigma_\alpha^{-1}(\alpha_i - \mu_l)\right\} \exp\left(\eta_l' \tilde{x}_i\right)}$$

   $i = 1, \ldots, n.$ and $k = 1, \ldots, K.$

2. Sample $y_{i^*j^*} = \begin{pmatrix} y_{i^*j^*1} \\ y_{i^*j^*2} \end{pmatrix}$ from $N_2\left(B'T_{j^*} + \alpha_{i^*},\ \sigma_\epsilon^2 I\right)$.

   It is noted that in the ERP study, the bivariate ERP components always are missing in pairs. In general, when there is only one component missing, one could sample the missing component from the conditional distribution given the other observed component.

3. Sample α_i from $N_2\left(\mu_{\alpha i},\ \Sigma_i\right)$.

   where $\mu_{\alpha i} = \Sigma_i \left(\frac{1}{\sigma_\epsilon^2} \sum_{j=1}^{n_i} \left(y_{ij} - B'T_j\right) + \Sigma_\alpha^{-1} \mu I_i\right)$,

   $\Sigma_i = \left(\frac{n_i}{\sigma_\epsilon^2} + \Sigma_\alpha^{-1}\right)^{-1},\ \ i = 1, \ldots, n.$

4. Sample β_1 and β_2 from bivariate normal distributions as in 5.5 and 5.6 respectively.

5. Sample σ_ϵ^2 from the Inverse-Gamma distribution as in 5.7.

6. Sample μ_k from $N\left(\mu_{\alpha k},\ \Sigma_k\right)$ as in 5.3.2, $\ k = 1, \ldots, K.$

7. Sample Σ_α^{-1} from the Inverse-Wishart distribution as in 5.10.

8. Sample η_k via Metropolis-Hastings algorithm. The sampling steps are almost the same as section 3.2.1 except the tuning parameter needs to be adjusted accordingly. We do not repeat the details here.

6.0 FUTURE RESEARCH

In this last chapter of the dissertation, we will give a summary of our proposed work and results that we have obtained. Moreover, we will discuss some interesting problems and unsolved issues related to the topic.

Our research in this dissertation was motivated by a study of alcoholism. In the study, the P300 amplitude trajectories of the subjects were examined through childhood and adolescent period. We used a latent mixture model frame work to analyze the heterogeneity in the trajectory population. The main focuses of our research are: first, to develop efficient computational algorithm to estimate such complexed model when confronted with large amount of missing data; second, to propose the use of cubic spline technique in curve fitting. The computational method for this improved model also was developed. It is our expectation that above algorithm and modeling can be widely applied in many other scientific areas.

6.1 PREDICTION FOR BINARY OUTCOME

In chapter 5.3.3, we described the use of latent mixture model for predicting a binary outcome. So far we have not found a significant relationship between the latent class membership and the onset of alcoholism in adulthood. Non-significant results also showed in post hoc analysis, where we used simple logistic regression with alcoholism presence as response and latent class membership as explanatory variable.

The phenomena may be caused by low occurrence of the alcoholism amongst the subjects. One possible way to overcome this problem is increasing the frequency of the outcome by grouping different psychiatric disorders as a whole. The candidate outcome variables are:

drug dependence and abuse; depression; and eating disorder etc. However, the detailed study of various outcomes is beyond the scope of this thesis.

6.2 LABEL SWITCHING PROBLEM

In section 2.2.3.4, we mentioned that the label switching problem was solved by imposing artificial constrains on the mixing components, for instance,

$A_{11} < A_{12} <, \ldots, < A_{1K}$
Or $A_{21} < A_{22} <, \ldots, < A_{2K}$ when the first components are equal
...

Unfortunately, above method can be problematic [1]. Consider two mixing components that are close to each other, imposing ordering constrains can cause bias in parameter sampling. For example, the samples for the component with smaller mean value are always bounded from above by another component. In the future, we plan to improve our MCMC algorithm by using other methods to avoid label switching problem. Celeux, Hurn and Robert gave summary of possible methods; for details, see [1].

6.3 UNKNOWN NUMBER OF KNOTS

In this dissertation, we proposed the use of a single knot cubic spline in latent mixture model. The location of the knot is a free parameter. We developed a MCMC algorithm to accommodate the free knot cubic spline. Sometimes, there is no prior information about the number of knots of the spline. In that case, the number of knots can be represented by a free parameter. A reversible jump MCMC algorithm proposed by Green at al.[13] can be used to analyze such model. Note that, during the sampling, the dimensionality of the parameter space may change at each iteration. An application of such method can be found in [9].

APPENDIX A

IDENTIFIABILITY

We now will take a close look at the latent mixture model to see if there exists a unique characterization. It is important to consider identifiability in practice because otherwise the estimation is not well defined.

The concept of identifiability of mixture models goes back to Teicher (1961). He gave a sufficient condition that a class of finite mixtures be identifiable and from that, established the identifiability of all finite mixtures of one-dimensional Gaussian distributions [40]. Yakowitz and Spragins modified the definition given by Teicher to include multidimensional cdf's [47]. Let $\mathcal{F}$= $\left\{F\left(x,\theta\right),\, \theta \in \Theta,\, x \in \Re^d\right\}$ be the d-dimensional distribution functions from which mixtures are to be formed. Then the class of finite mixtures of $\mathcal{F}$ with the appropriate class of distribution functions, $\mathcal{H}$.

Identifiability. $\mathcal{H}$= $\Big\{H(x) : H(x) = \sum_{k=1}^{K} \pi_k F\left(x,\theta_k\right),\, \pi_k > 0,\, \sum_{k=1}^{K} \pi_k = 1,$ $F\left(x,\theta_k\right) \in \mathcal{F},\, K = 1,2,\cdots,x \in \Re^d\Big\}$ *is the convex hull of* $\mathcal{F}$. *The definition of "identifiability" implies* $\mathcal{F}$ *generates identifiable finite mixtures if and only if* $\mathcal{H}$ *has the uniqueness of representation property:*

$$\sum_{k=1}^{K} \pi_k F_k = \sum_{k=1}^{K'} \pi_k^{'} F_k^{'} \tag{A.1}$$

implies $K = K^{'}$ *and for each* k, $1 \leq k \leq K$ *there is some* j, $1 \leq j \leq K^{'}$, *such that* $\pi_k = \pi_j^{'}$ *and* $F_k = F_j^{'}$.

Theorem. *(Yakowitz and Spragins, 1968)*

A necessary and sufficient condition that $\mathcal{H}$ be identifiable is that $\mathcal{F}$ be a linearly independent set over the field of real numbers, $\Re$.

The proof of the theorem can be found in the Titterington, Smith and Makov [42].

In the latent mixture model (3.1) and (5.1), the joint distribution of the longitudinal responses for subject i is the following:

$$
\begin{aligned}
f_{\Theta}\left(y_i\right) &= f_{\Theta}\left(y_{i1}, \cdots, y_{in_i}\right) && \text{(A.2)}\\
&= \int_{c,\eta} f\left(y_i|\eta_i\right) f\left(\eta_i|c_i, x_i\right) f\left(c_i|x_i\right)\, dc_i d\eta_i && \text{(A.3)}\\
&= \int_{\eta} f\left(y_i|\eta_i\right) \int_{c} \phi\left(\eta_i; Ac_i + \Gamma x_i, \Psi\right) \prod_{k=1}^{K} p_{ik}^{c_{ik}} dc_i\, d\eta_i && \text{(A.4)}\\
&= \int_{\eta} f\left(y_i|\eta_i\right) \sum_{k=1}^{K} p_{ik} \phi\left(\eta_i; A_k + \Gamma x_i, \Psi\right) d\eta_i && \text{(A.5)}\\
&= \int_{\eta} \sum_{k=1}^{K} p_{ik} \phi\left(\eta_i; A_k + \Gamma x_i, \Psi\right)\, \phi\left(y_i; T\eta_i, \Sigma_{\epsilon}\right) d\eta_i && \text{(A.6)}\\
&= \sum_{k=1}^{K} p_{ik} \phi\left(y_i; T\left(A_k + \Gamma x_i\right), T\Psi T^{'} + \Sigma_{\epsilon}\right) && \text{(A.7)}
\end{aligned}
$$

Where $p_{ik} = \frac{\exp\ \gamma_k^{'} \tilde{x}_i}{\sum_{l=1}^{K} \exp\left(\gamma_l^{'} \tilde{x}_i\right)}$.

The joint distribution shows that the model has K components, each represents a multivariate normal distribution. Yakowitz and Spragins showed that a finite mixture of multivariate Gaussian family is identifiable [47]. The proof is based on coming to a contradiction to Teicher's identifiability of one-dimensional Gaussian family result by assuming the multivariate Gaussian family is not identifiable. For the details, see [47] and [40].

Since the mean and the covariance matrix of a finite mixture of multivariate Gaussian are identifiable, parameter matrix A and Γ of the latent mixture model are identifiable as long as for each different x_i, the resulting y means are different and there are more distinct x combinations than corresponding parameters [23]. The covariance matrix is class invariant, the variance components can be uniquely defined with some constrains on matrix Ψ. In our Bayesian parameter estimation approach, the identifiability is governed by the priors and posteriors of the variance components.

APPENDIX B

SIMULATION STUDY 1

Table 18: Diagonal elements of Σ_ϵ – simulation 1

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.E./**S.D.**
σ_1^2	1.20	-0.1012	0.914	1.9292	0.619
		0.9612	**1.384**	**1.9116**	**0.293**
σ_2^2	1.06	0.5285	1.032	1.5355	0.307
		0.9373	**1.293**	**1.7375**	**0.249**
σ_3^2	0.86	0.3347	0.725	1.1153	0.238
		0.6924	**0.955**	**1.2782**	**0.182**
σ_4^2	0.58	0.16420	0.525	0.8858	0.220
		0.2328	**0.350**	**0.4957**	**0.082**
σ_5^2	0.65	0.17952	0.519	0.8585	0.207
		0.3978	**0.568**	**0.7836**	**0.119**
σ_6^2	1.02	0.38160	0.931	1.4804	0.335
		0.7637	**1.046**	**1.3978**	**0.196**
σ_7^2	0.75	0.08628	0.657	1.2277	0.348
		0.6558	**0.899**	**1.2076**	**0.170**
σ_8^2	0.32	0.0610	0.348	0.6350	0.175
		0.2989	**0.420**	**0.5737**	**0.085**
σ_9^2	0.21	0.0670	0.208	0.3490	0.086
		0.1378	**0.208**	**0.2968**	**0.049**
σ_{10}^2	0.31	-0.0334	0.188	0.4094	0.135
		0.2505	**0.384**	**0.5521**	**0.094**
σ_{11}^2	0.54	0.2318	0.855	1.4782	0.380
		0.4329	**0.715**	**1.0644**	**0.195**

Table 19: Parameter matrix Γ – simulation 1

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.E./S.D.
Γ_{11}	0.80	0.0129	0.828	1.6431	0.497
		0.5087	**1.035**	**1.5601**	**0.315**
Γ_{21}	0.05	-0.5707	-0.059	0.4527	0.312
		-0.3380	**-0.044**	**0.2324**	**0.174**
Γ_{31}	-0.04	-0.4042	0.014	0.4322	0.255
		-0.3630	**-0.093**	**0.1744**	**0.164**
Γ_{12}	-0.80	-2.2382	-1.00	0.2382	0.755
		-1.2156	**-0.660**	**-0.1081**	**0.340**
Γ_{22}	0.02	-0.5765	0.050	0.6765	0.382
		-0.3846	**-0.067**	**0.2379**	**0.189**
Γ_{32}	0.04	-0.6662	-0.084	0.4982	0.355
		-0.2262	**0.071**	**0.3692**	**0.181**
Γ_{13}	0.65	-0.3093	0.637	1.5833	0.577
		-0.1019	**0.432**	**0.9740**	**0.327**
Γ_{23}	-0.03	-0.7091	-0.140	0.4291	0.347
		-0.3969	**-0.098**	**0.1961**	**0.181**
Γ_{33}	0.06	-0.2682	0.214	0.6962	0.294
		-0.4781	**-0.198**	**0.0844**	**0.171**
Γ_{14}	0.08	-0.9099	0.056	1.0220	0.589
		-0.3447	**0.210**	**0.7705**	**0.343**
Γ_{24}	0.02	-0.4114	-0.003	0.4054	0.249
		-0.1911	**0.112**	**0.4128**	**0.185**
Γ_{34}	-0.01	-0.4170	-0.048	0.3210	0.225
		-0.3977	**-0.110**	**0.1822**	**0.177**

Table 20: $\{\gamma_k\}$ – simulation 1

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.E./S.D.
γ_{11}	-2.510	-13.9125	-2.427	9.0612	7.005
		-4.7159	**-0.936**	**2.8798**	**2.310**
γ_{12}	1.293	-2.3310	2.181	6.6730	2.739
		-0.4222	**1.011**	**2.4556**	**0.862**
γ_{13}	-1.420	-5.4969	-1.984	1.5289	2.142
		-4.7107	**-2.422**	**-0.5478**	**1.278**
γ_{14}	0.015	-3.6947	-0.418	2.8587	1.998
		-1.6611	**-0.263**	**1.0863**	**0.842**
γ_{15}	0.76	-2.6361	0.493	3.6221	1.908
		-0.5400	**0.870**	**2.3847**	**0.899**
γ_{21}	-3.20	-27.7948	-2.373	23.0488	15.5011
		-6.1208	**-2.784**	**0.4653**	**2.0440**
γ_{22}	1.40	-7.0850	1.041	9.1670	4.9549
		-0.2025	**1.062**	**2.4137**	**0.8190**
γ_{23}	-0.55	-16.9544	-1.482	13.9904	9.4344
		-2.2753	**-0.849**	**0.5138**	**0.8593**
γ_{24}	-0.48	-6.4716	0.288	7.0476	4.1217
		-0.8192	**0.508**	**1.8386**	**0.8227**
γ_{25}	0.65	-9.7577	0.646	11.0497	6.3437
		-1.2829	**0.080**	**1.4172**	**0.8169**

Table 21: Covariance matrix Ψ – simulatioin 1

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.E./S.D.
Ψ_{11}	0.8124	0.2174	0.875	1.53264	0.401
		0.6676	**1.088**	**1.6336**	**0.302**
Ψ_{21}	0.2028	-0.1235	0.193	0.50952	0.193
		0.1528	**0.333**	**0.5301**	**0.117**
Ψ_{31}	0.3000	0.0117	0.302	0.59228	0.177
		0.2690	**0.462**	**0.7021**	**0.134**
Ψ_{22}	0.2919	0.0277	0.254	0.48032	0.138
		0.2660	**0.397**	**0.5643**	**0.093**
Ψ_{32}	0.1260	-0.0435	0.091	0.22548	0.082
		0.0730	**0.175**	**0.2947**	**0.068**
Ψ_{33}	0.2976	0.1664	0.291	0.41564	0.076
		0.3477	**0.473**	**0.6339**	**0.089**

APPENDIX C

SIMULATION STUDY 2

Table 22: Diagonal elements of Σ_ϵ – simulation 2

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
σ_1^2	1.20	0.24805	1.26025	2.94303	0.88174
σ_2^2	1.06	0.17383	0.72246	1.82923	0.61386
σ_3^2	0.86	0.42937	1.10092	2.15264	0.57190
σ_4^2	0.58	0.14433	0.35096	0.66410	0.16705
σ_5^2	0.65	0.45513	0.82945	1.34371	0.27753
σ_6^2	1.02	0.48157	0.97990	1.65542	0.37852
σ_7^2	0.75	0.47724	0.71540	1.02851	0.17055
σ_8^2	0.32	0.20193	0.32654	0.48799	0.08815
σ_9^2	0.21	0.11656	0.18885	0.28284	0.05196
σ_{10}^2	0.31	0.10388	0.21171	0.36107	0.07977
σ_{11}^2	0.54	0.30044	0.59130	0.96111	0.20377

Table 23: Parameter matrix Γ – simulation 2

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Γ_{11}	0.80	-0.32530	0.59925	1.69591	0.60759
Γ_{21}	0.05	-0.19802	0.29719	0.69823	0.27570
Γ_{31}	-0.04	-0.32413	0.03487	0.38948	0.22171
Γ_{12}	-0.80	-1.70199	-0.78831	0.15393	0.56843
Γ_{22}	0.02	-0.45669	-0.03181	0.39051	0.25517
Γ_{32}	0.04	-0.36745	-0.09583	0.20151	0.17273
Γ_{13}	0.65	-0.13979	0.70109	1.53325	0.51054
Γ_{23}	-0.03	-0.32365	0.03506	0.38584	0.21813
Γ_{33}	0.06	-0.18431	0.05538	0.29537	0.14606
Γ_{14}	0.08	-0.70188	0.28200	1.21620	0.57942
Γ_{24}	0.02	-0.49875	-0.04174	0.39194	0.27293
Γ_{34}	-0.01	-0.22450	0.06850	0.38268	0.18576

Table 24: $\{\gamma_k\}$ – simulation 2

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
γ_{11}	-2.510	-6.69242	-0.56118	9.36762	4.95987
γ_{12}	1.293	-0.03808	1.84863	3.74193	1.15575
γ_{13}	-1.420	-13.74307	-4.6108	-0.40166	4.36722
γ_{14}	0.015	-1.61856	-0.14802	1.24490	0.87907
γ_{15}	0.76	-0.41868	1.64326	4.44887	1.49160
γ_{21}	-3.20	-8.19680	-3.97211	-0.33766	2.37936
γ_{22}	1.40	0.015073	1.77787	3.85231	1.20613
γ_{23}	-0.55	-1.95052	-0.63141	0.65688	0.79225
γ_{24}	-0.48	-1.55743	-0.25498	1.00355	0.77428
γ_{25}	0.65	-0.11487	1.17299	2.57253	0.81474

Table 25: Covariance matrix Ψ

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Ψ_{11}	0.8124	0.3981	1.2656	0.5247	0.6856
Ψ_{21}	0.2028	-0.3636	-0.0214	0.2385	0.1865
Ψ_{31}	0.3000	-0.0033	0.2094	0.4565	0.1422
Ψ_{22}	0.2919	0.2256	0.3954	0.6222	0.1241
Ψ_{32}	0.1260	0.0443	0.1666	0.3042	0.0800
Ψ_{33}	0.2976	0.2284	0.3343	0.4739	0.0759

APPENDIX D

SIMULATION STUDY 3

Table 26: Diagonal elements of Σ_ϵ – simulation 3

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
σ_1^2	1.5	0.2877	0.7689	1.5356	0.4076
σ_2^2	1.8	0.6347	1.9177	4.1037	1.2354
σ_3^2	1.0	0.4596	0.9169	1.5903	0.3658
σ_4^2	0.5	0.2616	0.6178	1.1947	0.3053
σ_5^2	1.2	0.8705	1.4054	2.1066	0.3870
σ_6^2	0.8	0.4760	0.7986	1.2264	0.2338
σ_7^2	1.4	0.7422	1.1211	1.5984	0.2680
σ_8^2	0.6	0.4221	0.6773	1.0103	0.1800
σ_9^2	0.9	0.8784	1.2825	1.7973	0.2847
σ_{10}^2	1.1	0.6743	1.1775	1.8048	0.3472
σ_{11}^2	2.0	0.4885	1.3040	2.4303	0.6040

Table 27: Parameter matrix Γ – simulation 3

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Γ_{11}	0.80	-0.0892	0.7947	1.6433	0.5314
Γ_{21}	0.10	-0.5867	-0.0663	0.4210	0.3060
Γ_{31}	-0.25	-1.1783	-0.6692	-0.1778	0.3023
Γ_{12}	-0.20	-0.4048	0.6170	1.5152	0.5854
Γ_{22}	0.50	0.2759	0.8107	1.3215	0.3215
Γ_{32}	-0.15	-0.6572	-0.1231	0.4170	0.3281
Γ_{13}	0.65	0.2641	1.1731	2.0051	0.5278
Γ_{23}	-0.05	-0.2718	0.2045	0.6628	0.2842
Γ_{33}	0.12	-0.2262	0.2544	0.7446	0.2935
Γ_{14}	-0.15	-1.2713	-0.2869	0.8108	0.6331
Γ_{24}	-0.08	-0.5306	-0.01395	0.5037	0.3147
Γ_{34}	-0.06	-0.8913	-0.3659	0.1550	0.3212

Table 28: $\{\gamma_k\}$ – simulation 3

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
γ_{11}	-3.0	-14.2580	-6.9935	-1.1571	3.8058
γ_{12}	1.2	0.3062	1.8627	3.9980	1.2555
γ_{13}	-0.5	-1.5783	0.9363	3.8481	1.5733
γ_{14}	2.0	-7.9389	-0.5751	2.4732	3.2837
γ_{15}	1.0	-1.9670	0.4764	3.1430	1.5622
γ_{21}	2.0	-1.1682	1.6900	4.4781	1.7709
γ_{22}	0.8	1.0200	2.1880	3.4651	0.7707
γ_{23}	-1.50	-3.8596	-2.5195	-1.1829	0.8512
γ_{24}	0.5	-0.6255	0.5595	1.7762	0.7429
γ_{25}	0.6	1.3612	2.7861	4.4526	0.9506

Table 29: Covariance matrix Ψ – simulation 3

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Ψ_{11}	1.20	0.3535	1.1346	2.4951	0.6895
Ψ_{21}	0.38	-0.5385	0.01789	0.5143	0.3193
Ψ_{31}	0.25	-0.7300	-0.2410	0.2452	0.2983
Ψ_{22}	0.80	0.5850	0.9325	1.3742	0.2457
Ψ_{32}	0.50	0.2750	0.5558	0.8889	0.1881
Ψ_{33}	1.40	1.0167	1.3773	1.8399	0.2561

APPENDIX E

APPLICATION TO P300 DATA

Table 30: Diagonal elements of Σ_ϵ – two-mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
σ_1^2	0.08986	0.13767	0.20281	0.03617
σ_2^2	0.0603	0.09586	0.14852	0.02785
σ_3^2	0.1192	0.17024	0.23765	0.03676
σ_4^2	0.08268	0.11472	0.15678	0.02299
σ_5^2	0.14738	0.20119	0.27363	0.0388
σ_6^2	0.06825	0.0916	0.1226	0.01707
σ_7^2	0.11493	0.15038	0.19476	0.02506
σ_8^2	0.13182	0.17301	0.23125	0.0314
σ_9^2	0.15013	0.19456	0.25245	0.03156
σ_{10}^2	0.12955	0.17121	0.22632	0.03066
σ_{11}^2	0.11539	0.15314	0.20177	0.02667

Table 31: Estimate for A – two-mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
A_{11}	2.57718	2.89406	3.16567	0.18164
A_{21}	2.77913	3.16475	3.55941	0.23713
A_{31}	2.7623	3.23338	3.70915	0.28689
A_{41}	1.95497	2.33458	2.81164	0.25971
A_{51}	2.30593	2.49253	2.72612	0.12745
A_{12}	3.43384	3.53385	3.63841	0.06302
A_{22}	3.43604	3.58032	3.72243	0.08724
A_{32}	3.3363	3.5313	3.71793	0.1164
A_{42}	3.00729	3.15245	3.29589	0.08799
A_{52}	2.92423	3.00091	3.07612	0.04621

Table 32: Estimate of γ_k – two-mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
γ_{11}	-37.73	-27.33402	-26.78509	2.38905
γ_{12}	19.03443	19.30655	24.46	1.18429
γ_{13}	-2.04095	-1.93743	0.023	0.45052
γ_{14}	-0.49	5.42361	5.73586	1.35898
γ_{15}	15.17726	15.71505	25.9	2.34055

Table 33: Diagonal elements of Σ_ϵ – three mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
σ_1^2	0.11186	0.17214	0.25749	0.04712
σ_2^2	0.06002	0.09911	0.15497	0.02943
σ_3^2	0.10172	0.14747	0.2072	0.03262
σ_4^2	0.05778	0.08142	0.1129	0.01725
σ_5^2	0.1401	0.19343	0.2626	0.03798
σ_6^2	0.0443	0.05941	0.07912	0.01087
σ_7^2	0.08348	0.10983	0.14334	0.01823
σ_8^2	0.08228	0.11011	0.14635	0.01959
σ_9^2	0.08661	0.11467	0.14991	0.0197
σ_{10}^2	0.07826	0.10585	0.14226	0.01949
σ_{11}^2	0.0702	0.09468	0.12566	0.01715

Table 34: Estimate for A – three mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
A_{11}	3.36497	3.50963	3.66027	0.09067
A_{21}	3.32143	3.49955	3.67096	0.10748
A_{31}	3.40813	3.61913	3.82799	0.12839
A_{41}	2.95539	3.10699	3.25897	0.09265
A_{51}	2.9218	3.00033	3.082	0.04864
A_{12}	3.03465	3.23683	3.44358	0.12589
A_{22}	2.94653	3.22021	3.49987	0.16576
A_{32}	3.01995	3.3496	3.66452	0.19747
A_{42}	1.9394	2.22551	2.51373	0.17283
A_{52}	2.33623	2.46081	2.58794	0.07661
A_{13}	3.34269	3.54407	3.74554	0.12345
A_{23}	3.54999	3.78281	4.0242	0.14605
A_{33}	3.31137	3.61277	3.91611	0.1851
A_{43}	3.17257	3.42096	3.66519	0.14938
A_{53}	3.09481	3.24478	3.38794	0.08944

Table 35: Estimate of γ_k – three mixture

PARAMETER	5%	ESTIMATE	95%	S.D.
γ_{11}	2.84036	4.06048	5.6442	0.85065
γ_{12}	-0.01904	1.3117	2.83205	0.7758
γ_{13}	-13.4399	-5.17196	-2.75648	3.26461
γ_{14}	-0.29865	1.04653	4.23765	1.31103
γ_{15}	-3.30135	-1.27727	0.86646	1.65808
γ_{21}	-3.72696	-3.72696	-3.72696	0
γ_{22}	2.56945	2.56945	2.56945	0
γ_{23}	1.73152	1.73152	1.73152	0
γ_{24}	0.30461	0.30461	0.30461	0
γ_{25}	1.73225	1.73225	1.73225	0

APPENDIX F

FIXED KNOT SPLINE SIMULATIONS

Table 36: Diagonal elements of Σ_ϵ – fixed knot spline

PARAMETER	TRUE VALUE	5%	ESTIMATE	95%	S.D.
σ_1^2	1.20	0.35918387	1.8163	3.4822170	0.9552
σ_2^2	0.50	0.04079826	0.1417	0.3096920	0.0870
σ_3^2	0.80	0.48681225	0.7576	1.0963182	0.1887
σ_4^2	1.00	0.40879705	0.6252	0.9029152	0.1524
σ_5^2	0.90	0.66587359	0.9609	1.3353276	0.2064
σ_6^2	1.20	0.89193902	1.2114	1.6022659	0.2207
σ_7^2	0.40	0.25128991	0.3743	0.5167648	0.0824
σ_8^2	1.10	0.65385495	0.8821	1.1692545	0.1584
σ_9^2	0.50	0.34535708	0.4943	0.6789783	0.1039
σ_{10}^2	0.65	0.38323461	0.5790	0.8154604	0.1328
σ_{11}^2	1.40	0.53447683	1.0262	1.6117949	0.3346

Table 37: Parameter matrix Γ – fixed knot spline

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Γ_{11}	0.40	-0.96379800	-0.1343131	0.6254843	0.4781829
Γ_{21}	1.00	0.21164321	0.8805075	1.5850198	0.4153896
Γ_{31}	1.20	-0.16601821	0.7820478	1.6891867	0.5504022
Γ_{41}	0.80	0.15322117	0.8577256	1.5652377	0.4345983
Γ_{51}	-0.30	-0.86938522	-0.2520641	0.3598583	0.3751993
Γ_{12}	-0.40	-2.94282989	-1.5570944	-0.2249653	0.8400174
Γ_{22}	0.50	-0.35690548	0.7400830	1.8095950	0.6377515
Γ_{32}	1.10	0.20833818	1.3870110	2.6715480	0.7571014
Γ_{42}	0.45	-0.88662590	0.2045928	1.1706297	0.6404632
Γ_{52}	1.30	0.66954932	1.4615164	2.3089926	0.5047502
Γ_{13}	-0.70	-1.88587036	-1.0971038	-0.3666841	0.4592052
Γ_{23}	0.65	-0.09028497	0.6289569	1.3542915	0.4345952
Γ_{33}	0.80	-0.48441946	0.4712866	1.3123085	0.5594229
Γ_{43}	-1.30	-1.73361358	-1.1029953	-0.4076910	0.4030641
Γ_{53}	0.90	-0.04055456	0.4626317	0.9621692	0.3049029
Γ_{14}	1.00	0.92787207	1.9868170	3.3010095	0.6943474
Γ_{24}	1.30	0.30593133	1.2407297	2.0897410	0.5452900
Γ_{34}	-0.20	-2.04324303	-0.8534608	0.2862777	0.7005595
Γ_{44}	1.40	0.52144472	1.3398864	2.1966744	0.5127865
Γ_{54}	0.85	0.54303529	1.1635087	1.8092946	0.3854619

Table 38: $\{\gamma_k\}$ – fixed knot spline

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
γ_{11}	-2.00	-9.4846366	-4.546989	-1.058199	2.5958424
γ_{12}	-5.00	-5.3266520	-3.896036	-2.670768	0.7847861
γ_{13}	6.00	4.3426730	7.843844	12.912302	2.6397498
γ_{14}	2.00	0.1630886	1.256321	2.248609	0.6506906
γ_{15}	-4.00	-4.2890295	-2.844878	-1.550955	0.8463214
γ_{21}	3.00	2.2476022	3.7778836	5.540983	1.0093364
γ_{22}	-4.00	-5.6639550	-4.0657961	-2.581002	1.0122094
γ_{23}	-5.00	-7.4945965	-5.4858906	-3.621376	1.1817404
γ_{24}	1.00	-0.5691191	0.4957370	1.565129	0.6817153
γ_{25}	2.00	-1.0276077	0.4545812	1.927834	0.9056638

Table 39: Covariance matrix Ψ – fixed knot spline

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Ψ_{11}	0.60	0.0809	0.7641	2.0140	0.6168
Ψ_{22}	0.40	0.0569	0.2707	0.6290	0.1857
Ψ_{33}	0.25	0.0601	0.3207	0.7729	0.2351
Ψ_{44}	0.60	0.4203	0.8724	1.3844	0.2971
Ψ_{55}	0.80	0.5917	1.1309	1.7615	0.3550

APPENDIX G

FREE KNOT SPLINE SIMULATIONS

Table 40: Diagonal elements of Σ_ϵ – free knot (truncated normal)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
σ_1^2	1.00	0.06734012	0.8116	2.418597	0.7776
σ_2^2	1.20	1.49412453	3.2950	6.059737	1.4574
σ_3^2	1.50	2.75697865	4.7201	7.160512	1.3647
σ_4^2	2.00	0.73333549	1.3219	2.071580	0.4118
σ_5^2	1.80	0.82016789	1.2343	1.765789	0.2984
σ_6^2	0.90	0.46095885	0.7081	1.012589	0.1691
σ_7^2	1.00	0.87062310	1.1873	1.579599	0.2178
σ_8^2	2.20	1.34652253	1.7910	2.322503	0.2979
σ_9^2	1.50	1.22098753	1.6417	2.150969	0.2837
σ_{10}^2	0.85	0.72850252	1.0142	1.370584	0.1977
σ_{11}^2	1.00	0.47131993	1.2988	2.054765	0.4750

Table 41: Parameter matrix Γ – free knot (truncated normal)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Γ_{11}	0.20	-1.11445436	-0.2094	0.6502545	0.5374775
Γ_{21}	0.50	-1.15616691	0.2359	1.8465877	0.9095538
Γ_{31}	0.30	-0.88130033	0.3417	1.3812423	0.6894895
Γ_{41}	0.40	0.04561498	0.7353	1.4920484	0.4331153
Γ_{51}	0.50	-0.28465026	0.2681	0.7830298	0.3286610
Γ_{12}	0.75	-0.04328120	0.9617	1.9631243	0.6026198
Γ_{22}	0.25	-2.58123745	-0.6582	1.0172517	1.1207814
Γ_{32}	1.00	0.41975911	1.8638	3.4146192	0.9166958
Γ_{42}	0.40	-0.99684440	0.0024	0.9642876	0.5927909
Γ_{52}	0.20	0.03099701	0.5545	1.1432309	0.3432278
Γ_{13}	0.50	-0.14859702	0.7755	1.7560901	0.5820782
Γ_{23}	1.00	0.38517244	2.0536	3.6951880	1.0001010
Γ_{33}	0.60	-0.65873191	0.5772	1.8663959	0.7748862
Γ_{43}	0.50	-0.57881472	0.2057	0.9611890	0.4724274
Γ_{53}	0.80	-0.07419930	0.4496	0.9858645	0.3250405
Γ_{14}	-0.25	-4.54370776	-1.710	0.7627850	1.597272
Γ_{24}	1.10	-6.47339820	-2.9183	0.2836884	1.995871
Γ_{34}	0.40	-1.43023479	2.0452	5.6908706	2.296158
Γ_{44}	0.30	-2.05707087	0.4500	2.9099030	1.540209
Γ_{54}	0.50	-1.47753327	0.6335	2.8618018	1.278740

Table 42: $\{\gamma_k\}$ – free knot (truncated normal)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
γ_{11}	-10	-8.373225	-5.282897	-1.344543	2.348120
γ_{12}	15	5.168424	8.023552	11.869139	2.295260
γ_{13}	-12	-11.908442	-8.100832	-5.334269	1.996313
γ_{14}	13	9.731781	12.067107	15.066111	1.554092
γ_{15}	20	7.290714	10.494941	15.797119	2.319339
γ_{21}	5	4.757292	7.808832	11.375684	1.849136
γ_{22}	4	1.639368	4.112789	6.502078	1.567111
γ_{23}	-5	-11.176965	-8.011558	-4.973060	1.756590
γ_{24}	10	2.642880	7.321311	10.854664	2.427488
γ_{25}	-4	-15.046937	-7.986390	-2.855800	3.765636

Table 43: Covariance matrix Ψ – free knot (truncated normal)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Ψ_{11}	0.45	0.1615	1.9412	3.7001	1.0845
Ψ_{22}	0.45	0.6064	3.8952	7.8327	2.2911
Ψ_{33}	0.45	0.0742	0.4897	1.2780	0.4110
Ψ_{44}	0.45	0.0962	0.4663	1.0190	0.2929
Ψ_{55}	0.45	0.0783	0.6442	1.4860	0.4383

Table 44: Diagonal elements of Σ_ϵ – free knot (beta)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
σ_1^2	1.00	0.11490577	0.9118	2.136885	0.5738
σ_2^2	1.20	0.33245201	0.9428	2.022528	0.4374
σ_3^2	1.50	1.58004348	2.5529	4.018304	0.6227
σ_4^2	2.00	2.64779607	4.1787	6.261933	0.9220
σ_5^2	1.80	1.88869282	2.8065	4.077812	0.5571
σ_6^2	0.90	0.50436582	0.9166	1.448328	0.2466
σ_7^2	1.00	0.74113164	1.1360	1.655222	0.2338
σ_8^2	2.20	1.18377276	1.7599	2.526270	0.3468
σ_9^2	1.50	0.98770833	1.5628	2.298878	0.3347
σ_{10}^2	0.85	0.52826365	1.0107	1.628164	0.2818
σ_{11}^2	1.00	0.03854691	0.4317	1.322698	0.3575

Table 45: Parameter matrix Γ – free knot (beta)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Γ_{11}	0.20	-0.4346073	0.28092495	0.9726666	0.3666822
Γ_{21}	0.50	-1.1813306	0.13954767	1.3558389	0.6695729
Γ_{31}	0.30	-0.4946720	0.85263772	2.2456146	0.7066673
Γ_{41}	0.40	-1.3206724	-0.15157410	1.0402946	0.6077953
Γ_{51}	0.50	-0.7200610	-0.01474894	0.6811170	0.3539668
Γ_{12}	0.75	-0.2064355	0.6253033	1.5613692	0.4639864
Γ_{22}	0.25	-1.5297735	0.1879357	1.6805446	0.7771341
Γ_{32}	1.00	0.1036115	1.6561443	3.2295323	0.8114257
Γ_{42}	0.40	-0.5366710	0.7759965	2.0419937	0.6498700
Γ_{52}	0.20	-0.4090822	0.4090063	1.1825971	0.4079976
Γ_{13}	0.50	-0.3401601	0.4057803	1.2552120	0.3993963
Γ_{23}	1.00	-0.9440144	0.5435716	2.0826422	0.7575399
Γ_{33}	0.60	-0.7215947	0.7708687	2.2680413	0.7574871
Γ_{43}	0.50	-1.7375071	-0.4672812	0.8701831	0.6604271
Γ_{53}	0.80	-0.4015500	0.3152499	1.0388198	0.3702179
Γ_{14}	-0.25	-0.6173260	1.7565596	4.3253389	1.365589
Γ_{24}	1.10	-6.0045916	-2.0184605	0.7582995	1.668890
Γ_{34}	0.40	-3.9313324	1.0492038	5.5834895	2.246305
Γ_{44}	0.30	-8.1627780	-2.6663879	1.0715512	2.193610
Γ_{54}	0.50	-4.5493651	-0.6224611	7.0391040	2.549119

Table 46: $\{\gamma_k\}$ – free knot (beta)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
γ_{11}	-10	-11.363018	-4.922416	0.565344	3.211789
γ_{12}	15	4.159225	7.503412	11.819270	1.901914
γ_{13}	-12	-11.643400	-6.817899	-3.176648	2.131709
γ_{14}	13	7.653873	10.921227	14.925553	2.018018
γ_{15}	20	5.792401	10.870971	16.379508	2.864669
γ_{21}	5	3.018138	7.154200	12.771643	2.497693
γ_{22}	4	3.813814	8.607736	14.267106	2.868526
γ_{23}	-5	-13.537123	-7.907969	-3.889450	2.479663
γ_{24}	10	3.241150	8.659103	15.722992	3.096204
γ_{25}	-4	-16.017718	-8.416017	-2.106191	3.933162

Table 47: Covariance matrix Ψ – free knot (beta)

PARAMETER	TRUE VALUE	5%	MEAN	95%	S.D.
Ψ_{11}	0.45	0.0582	0.5044	1.4401	0.3884
Ψ_{22}	0.45	0.0397	0.5391	2.5228	0.6445
Ψ_{33}	0.45	0.3988	2.1872	4.3988	1.0433
Ψ_{44}	0.45	2.7281	4.8708	7.6422	1.2767
Ψ_{55}	0.45	0.8271	2.0346	3.1971	0.5796

BIBLIOGRAPHY

[1] Celeux G., Hurn M. and Robert C.P. (2000). Computational and Inferential Difficulties With Mixture Posterior Distributions. *Journal of the American Statistical Association* **95**, 957-970.

[2] Chambers W.J., Puig-Antich J, Hirsch M, Paez P, Ambrosini PJ, Tabrizi MA, and Davies M. (1985). The assessment of affective disorders in children and adolescents by semistructured interview: Testretest reliability of the schedule for affective disorders and schizophrenia for school-age children. Present episode version. *Arch Gen Psychiatry* **42**, 696702.

[3] Chan, K.S. (1993). Asymptotic behavior of the Gibbs sampler. *Journal of the American Statistical Association* **8**, 320-326.

[4] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician* **49**, 327-335.

[5] Clogg, C.C. (1995). Latent class models. In *The Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Arminger, G., Clogg, C.C. and Sobel, E. New York: Plenum. 331-359.

[6] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B* **39**, 1-22.

[7] Dias JG, Wedel M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing* **14**, 323-332.

[8] Diebolt, J. and Robert, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B* **56**, 363-375.

[9] Dimatteo, I., Genovese, C.R. and Kass, R.E. (2001). Bayesian curve-fitting with free knot splines. *Biometrika* **88**, 1055-1071.

[10] Escobar M.D. and West M. (1995).Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577-588.

[11] Garrett, E.S. and Zeger, S.L. (2000). Latent class model diagnosis. *Biometrics* **56**, 1055-1067.

[12] Gelman, A., Carlin, JB., Stern, HS., and Rubin, DB. (1995). *Bayesian data analysis*. Chapman&Hall/CRC.

[13] Green P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**; 711-32.

[14] Hastie, T., Tibshirani, R. and Friedman, J. *The Elements of Statistical Learning*. Springer.

[15] Hill, SY., Shen, S., Locke, J., Steinhauer SR., Konicky, C., and Lowers. L., et al. (1999). Developmental delay in P300 production in children at high risk for developing alcohol-related disorders. *Biol Psychiatry* **46**, 970-981.

[16] Hill, SY., Yuan, H. and Locke, J. (1999). Path analysis of P300 amplitude of individual from families at high and low risk for developing alcoholism. *Biol Psychiatry* **45**, 346-359.

[17] Hill, SY. and Shen, S. (2002). Neurodevelopmental patterns of visual P3b in association with familial risk for alcohol dependence and childhood diagnosis. *Society of Biological Psychiatry* **51**, 621-631.

[18] Jacobs, R.A., Jordan, M.I., Nowlan, S.J., and Hinton, G.E. (1991). Adaptive mixtures of local experts. *Neural Computation* **3**, 79-87.

[19] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. J. Wiley & Sons, New York.

[20] Liu, JS., Wong, W.H., Kong,A. (1994). Covariance structure of the Gibbs sampler with the applications to the comparisons of estimates and augmentation schemes. *Biometrika* **81**, 27-40.

[21] Liu, JS., Wong, WH., Kong,A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society* **B57**, 157-169.

[22] Müller, P. (1993). A generic approach to posterior integration in Gibbs sampling. Technical Report. Purdue University.

[23] Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55**, 463-469.

[24] Muthén, B. and Muthén, LK. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcohol Clin Exp Res* **24**,882-891.

[25] Muthén, LK. and Muthén, B.(1998). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

[26] Muthén, B. (1998). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. *New Methods for the Analysis of Change*. APA, Washington, DC. in press.

[27] Nagin, DS. and Tremblay, RE. (1993). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Dev* **70**, 1181-1196.

[28] Pfefferbaum, A., Ford, J.M., White, P.M., and Roth, W.T. (1989). P3 in schizophrenia is affected by stimulus modality, response requirements, medication status, and negative symptoms. *Arch Gen Psychiatry* **46**, 1035-1044.

[29] Pfefferbaum, A. Ford, J.M., White, P.M., and Mathalon, D. (1991). Event-related potentials in alcoholic men: P3 amplitude reflects family history but not alcoholic consumption. *Alcohol Clin Exp Res* **15**, 839-850.

[30] Redner, RA. and Walker, HF. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**, 195-239.

[31] Richardson, S. and Green P.J. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B* **59**; 731-792.

[32] Rindskopf, D. and Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Stat Med* **5**, 21-27.

[33] Rodriguez-Yam, G., Davis, R.A., and Scharf, LL. (2004). Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression.

[34] Schervish, MJ. and Carlin, BP. (1992). On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics* **1**, 111-127.

[35] Steinhauer, S.R., Hill, S.Y., and Zubin, J. (1987). Event-related potentials in alcoholics and their first degree relatives. *Alcohol* **4**, 307-314.

[36] Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of Royal Statistical Society B* **62**, 795-809.

[37] Sahu, S.K. and Roberts G.O. (1999) On convergence of the EM algorithm and the Gibbs sampler. *Statistics and Computing* **9**, 55-64.

[38] Tanner, MA. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions,* 3rd ed. New York: Springer.

[39] Tanner, M.A., and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528-540.

[40] Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics* **34**, 1265-1269.

[41] Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701-1762.

[42] Titterington, D.M. Smith, A.F.M. and Makov, UE. (1985) *Statistical Analysis of Finite Mixture Distributions.* Chichester, U.K.: Joen Wiley and Sons.

[43] Ubersax, JS. and Grove, WM. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine* **9**, 559-572.

[44] Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91**, 217-221.

[45] Wu, CFJ. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics* **11**, 95-103.

[46] Meng,X.L. and Rubin D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267-278.

[47] Yakowitz, J.S. and Spragins, J.D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics* **39**, 209-214.

[48] Young, M.A., Tanner, M.A., and Meltzer, H.Y. (1982). Operational definitions of schizophrenia: What do they identify? *Journal of Nervous and Mental Disease* **170**, 443-447.

[49] Zubin, J., and Steinhauer, S.R. (1981). How to break the logjam in schizophrenia: A look beyond genetics. *The Journal of Nervous and Mental Disease* **169**, 477-492.

[50] `http://www.bnl.gov/neuropsychology/monetary_rewards_erp.asp`.