Spring 1-1-2006

Bayesian Analysis of Discrete Longitudinal Data

Nicholas Bernini

Bayesian Analysis of Discrete Longitudinal Data

A Thesis

Presented to the Faculty

of the Department of Mathematics and Computer Science

McAnulty College and Graduate School of Liberal Arts

Duquesne University

in partial fulfillment of

the requirements for the degree of

Masters of Science in Computational Mathematics

by

Nicholas A. Bernini

April 3, 2006

Nicholas A. Bernini

Bayesian Analysis of Discrete Longitudinal Data

Master of Science in Computational Mathematics

Department of Mathematics & Computer Science
Duquesne University, Pittsburgh, PA, USA

April 3, 2006

APPROVED_____
John Kern, Ph.D., Assistant Professor
Department of Mathematics & Computer Science

APPROVED_____
Frank D'Amico, Ph.D., Department Chair
Department of Mathematics & Computer Science

APPROVED_____
Kathleen Taylor, Ph.D., Graduate Director
Department of Mathematics & Computer Science

APPROVED_____
Francesco C. Cesareo, Ph.D., Dean
McAnulty College and Graduate School of Liberal Arts

Nicholas A. Bernini

Bayesian Analysis of Discrete Longitudinal Data

Advisor: John C. Kern II

Abstract

This thesis explores a Bayesian hierarchical model to compare treatment effectiveness for menopausal symptom relief. Specifically, this model recognizes the discrete nature of the data, as well as its time dependency. Bayesian analisys is used to make inference on each individual profile, as well as on a group profile for each treatment group.

Acknowledgements

I would like to thank my advisor John C. Kern II for his ideas and support in this work. I would also like to thank Sujit Ghosh for his contribution in the theoretical model, and Susan Cohen for supplying the study data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Objective

The purpose of this research is to develop a hierarchical longitudinal regression model for frequency data collected on several subjects over a fixed period of time. This model will recognize time dependence that exists in an individual profile, and will use all profiles from a particular group to provide a corresponding group profile. Given actual data from a study conducted to investigate an alternative treatment for menopausal symptom relief, we apply this model to compare the effectiveness of different treatments.

We propose a piecewise linear function, dependent on time, to describe the frequency measurments of each individual from various treatment groups. One of the main benefits of this model is its flexibility to describe the profile of a single subject, and treat this one profile as a sample from some population. In doing this, each of the individuals assigned to a given treatment goup can be considered random samples from that treatment population distribution. Bayesian analysis will be applied to estimate the parameters for the profile of each individual; the hierarchical com-

ponent of the model will then allow for the estimation of the group profile parameters.

Previous models applied to these data included a Generalized Poisson (GP) Regression Model (Borgesi, et al 2003), and a Negative Binomial Regression Model (Kern and Cohen, et al 2005). These models allow for a description of a group mean vs. time, but do not recognize time dependence in an individual's profile. A Bayesian hierarchical model which explicitly recognizes time dependence in an individual's daily hot flush frequency was also applied to these data (Jordan, et al 2005); this model does not describe a group's mean vs. time, however. The model considered in this research incorporates the best of both of these models by allowing each patient's individual profile to represent a random sample from the group's population distribution. For other tailored hierarchical Bayesian piecewise regression models, see (Lopes, et al 2002).

1.2 The Data

The motivation for this model comes from a study conducted by Yale University, with funding from The Patrick and Catherine Weldon Donaghue Medical Foundation. The study aimed to ascertain the effectiveness of different treatments for the relief of menopausal symptoms among women in breast cancer remission. Traditional treatment for menopausal symptom relief included hormone treatment, which is highly discouraged for patients in breast cancer remission. The alternative method thought to be beneficial is acupuncture.

The study selected 39 women in breast cancer remmision suffering from menopausal symptoms and assigned them randomly to one of three treatment groups. One treatment group (16 individuals) recieved acupuncture in benefical locations. Another

group (17 individuals) received acupuncture in locations thought to be non-beneficial. A third treatment group (6 individuals) recieved educational sessions that emphasized techniques for healthy midlife living. Each of the patients were monitored for 13 weeks with the first week being the baseline; the treatment was administered for the remaining 12 weeks. Daily hot flush frequency was measured (self-recorded) over this time period. It is from this observed frequency data that we apply our piecewise linear model to determine the effectiveness of acupuncture for menopausal symptom relief.

1.3 The Model

Due to the discrete nature of this data, we let the frequency observed by patient i at time t, denoted by y_{it}, be Poisson

$$y_{it} \backsim Pois(\mu_{it}),$$

and define the Poisson mean μ_{it} as a piecewise linear funtion of time $t \in \{1, 2, \ldots, T\}$. Then, for a fixed number of partitions n, the $n-1$ points joining each line segment, combined with the two extreme endpoints, will suffice in describing this function. We introduce, for person i,

$$\boldsymbol{K_i} = \{K_{i1}, K_{i2}, \ldots, K_{i(n+1)}\} \quad \text{and} \quad \boldsymbol{\lambda_i} = \{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i(n+1)}\},$$

where (K_{ij}, λ_{ij}) represents the right endpoint of the $(j-1)^{st}$ line segment and the left endpoint of the j^{th} line segment of this piecewise linear function for person i. Then, for time t, there exists $j \in \{1, 2, \ldots, n\}$ such that $K_{ij} < t < K_{i(j+1)}$ which gives

$$\mu_{it} = \frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij}$$

as the mean frequency for individual i at time t. In doing this, we recognize that y_{it} is distributed as a Poison random variable, whose mean μ_{it} is dependent on time. In selecting a prior distribution for μ_{it}, we assign prior distributions for $\boldsymbol{K_i}$ and $\boldsymbol{\lambda_i}$. Due to the restriction of $\lambda_{ij} \geqslant 0$, we assign a gamma prior for λ_{ij} of

$$\lambda_{ij} \backsim gamma(\alpha_j, \beta_j), \text{ where } \alpha_j, \beta_j > 0,$$

and noniformative hyper-priors for α_j and β_j of

$$\alpha_j \backsim Unif(0, M) \text{ and } \beta_j \backsim Unif(0, M),$$

where $M = 1000$. In assigning a prior distribution for $\boldsymbol{K_i}$, we interpret the time partition locations $\boldsymbol{K_i}$ as a function of the distances $\boldsymbol{x_i} = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ between them. This means $x_{i1} = K_{i1} - 0,\ x_{i2} = K_{i2} - K_{i1},\ \ldots$, etc. Then we assign a prior for $\boldsymbol{x_i}$ of

$$\boldsymbol{x_i} \backsim multinomial(N; \theta_1, \theta_2, \ldots, \theta_n),$$

where $\theta_j \in [0, 1]$, $\sum_{j=1}^{n} \theta_j = 1$, $x_{ij} \in [0, N] \cap \mathbb{Z}^+$, and $\sum_{j=1}^{n} x_{ij} = N$. We also assign a hyper-prior distribution on $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$ of

$$\boldsymbol{\theta} \backsim Dirichlet(a_1, a_2, \ldots, a_n), \text{ where } a_j > 0,\ \forall j.$$

Taking the product of these prior distributions gives the prior distribution for $\boldsymbol{\mu_i}$, denoted $\pi(\boldsymbol{\mu_i})$, where $\boldsymbol{\mu_i} = \{\mu_{i1}, \mu_{i2}, \ldots, \mu_{iT}\}$. Given $\boldsymbol{\mu_i}$ and the data of the i^{th} person $\boldsymbol{y_i}$, we can obtain the likelihood function for $\boldsymbol{\mu_i}$, denoted $L(\boldsymbol{\mu_i}|\boldsymbol{y_i})$. As given by Bayes' Theorem, we define the posterior distribution $\pi(\boldsymbol{\mu_i}|\boldsymbol{y_i})$ as being proportional to the product of the likelihood function and the prior distribution

$$\pi(\boldsymbol{\mu_i}|\boldsymbol{y_i}) \propto L(\boldsymbol{\mu_i}|\boldsymbol{y_i}) \cdot \pi(\boldsymbol{\mu_i}).$$

Given this posterior distribution for $\boldsymbol{\mu_i}$, we use Markov Chain Monte Carlo (MCMC) sampling techniques to draw parameter values from this distribution (Gilks, et al 1996). Inference is then made based on these parameter values.

Chapter 2

Model Implementation

2.1 Markov Chain Monte Carlo Sampling Techniques

Using the model described in Section 1.3, we use MCMC sampling techniques to estimate the parameters $\boldsymbol{\lambda_i}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{K_i}$, $\boldsymbol{x_i}$, and $\boldsymbol{\theta}$. These sampling techniques consist of Metropolis, Metropolis-Hastings and Gibbs sampling. We use these techniques to sample draws for each parameter from its posterior distribution. To demonstrate the process, let $\mathbf{y}$ represent the data of one individual and $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$ represent the m parameters in need of estimation. We define the posterior distribution of $\boldsymbol{\theta}$, denoted $\pi(\boldsymbol{\theta}|\mathbf{y})$, as being proportional to the product of the liklihood function $L(\boldsymbol{\theta}|\mathbf{y})$ and the prior distribution $\pi(\boldsymbol{\theta})$.

Gibbs sampling: From the joint posterior of $\boldsymbol{\theta}$, obtain the full conditional distribution for the parameter to be updated. From this full conditional distribution, sample a random draw and accept this as the updated value for our parameter. In the case where the full conditional distribution is unobtainable for a parameter, Metropolis or Metropolis-Hastings sampling will be used.

Metropolis sampling: Metropoils sampling is used to sample an unbounded parameter from its posterior distribution. In this case, given a current value of our parameter θ_1^c, propose a new value θ_1^* from a proposal density. We chose a uniform proposal density centered at θ_1^c with length $2k$, $k \in \mathbb{R}^+$. Thus, θ_1^* is randomly chosen from the interval $(\theta_1^c - k, \theta_1^c + k)$. We note that the size of k varies with respect to the parameter. We then accept θ_1^* with probability $\boldsymbol{\gamma^*}$,

$$\boldsymbol{\gamma^*} \quad = \quad min\{1, \quad \frac{\pi(\theta_1^*, \theta_2^c, \ldots, \theta_m^c|\mathbf{y})}{\pi(\theta_1^c, \theta_2^c, \ldots, \theta_m^c|\mathbf{y})}\}. \tag{2.1}$$

We note that the above fraction can exceed values of 1, and thus violate the rules of probability. We correct this by using the *min* function as in Equation (2.1). This process is used in calculating all acceptance probabilities. We maintain our current value θ_1^c with probability $1 - \boldsymbol{\gamma^*}$.

Metropolis-Hastings sampling: Metropolis-Hastings sampling is used when the full conditional distribution of the parameter is unobtainable and the parameter is bounded. The same technique is used here as in Metropolis sampling, except for the inclusion of a correction factor to adjust the acceptance probability $\boldsymbol{\gamma^*}$. More discussion on Hasting ratios is given in Section 2.3.

We iterate each of these sampling techniques for all θ_i's, using the most current value of θ_i at each step. In estimation of the parameters for our model, we implemented a Java program with 9.5 million iterations.

2.2 MCMC Calculations

In order to update the parameters using MCMC sampling techniques, the posterior distribution for all parameters must be obtained. To update the mean μ_{it} for person i at a time $t \in \{1, 2, \ldots 91\}$, we obtain both the likelihood function $L(\boldsymbol{\mu_i}|\boldsymbol{y_i})$ and

the prior distribution $\pi(\boldsymbol{\mu_i})$ where $\boldsymbol{\mu_i} = \{\mu_{i1}, \mu_{i2}, \ldots, \mu_{i91}\}$ are the estimated mean hotflushes to be updated and $\boldsymbol{y_i} = \{y_{i1}, y_{i2}, \ldots, y_{i91}\}$ is the i^{th} person's data. Letting μ_{it} be the mean of a Poisson random variable, the likelihood function for $\boldsymbol{\mu_i}$ is the product of 91 Poisson density functions

$$\begin{aligned} L(\boldsymbol{\mu_i}|\boldsymbol{y_i}) &= \prod_{t=1}^{91} \frac{{\mu_{it}}^{y_{it}}}{y_{it}!} e^{-\mu_{it}} \\ &\propto \prod_{t=1}^{91} {\mu_{it}}^{y_{it}} e^{-\mu_{it}}. \end{aligned} \tag{2.2}$$

Defining μ_{it} as a piecewise linear function of time, we partition the time support into 4 intervals. Then, for person i, we define this function in terms of its five knot locations $\boldsymbol{K_i}$ and node heights $\boldsymbol{\lambda_i}$

$$\boldsymbol{K_i} = \{K_{i1}, K_{i2}, \ldots, K_{i5}\} \quad \text{and} \quad \boldsymbol{\lambda_i} = \{\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i5}\},$$

where (K_{ij}, λ_{ij}) is the left end point of the j^{th} line segment and the right end point to the $(j-1)^{st}$ line segment. To avoid confusion about where to evaluate our piecwise linear function at a time which shares an endpoint to two different line segments, we assign each K_{ij} a half integer value ranging from 0.5 to 91.5. To effectively model the data, we must also constrain some of the ${K_{ij}}'s$ in value. Since our data was collected over 13 weeks, we fix $K_{i1} = 0.5$ and $K_{i5} = 91.5$ as endpoints. This ensures that we model the data over the correct time period. Also, to account for the baseline week, we fix $K_{i2} = 7.5$. Now, we have

$$\boldsymbol{K_i} = \{0.5, 7.5, K_{i3}, K_{i4}, 91.5\}.$$

Then for $t \in \{1, 2, \ldots, 91\}$ there exists $j \in \{1, 2, \ldots, 4\}$ such that $K_{ij} < t < K_{i(j+1)}$. This yields

$$\mu_{it} = \frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij}. \tag{2.3}$$

Substituting (2.3) into (2.2) gives

$$\begin{aligned} L(\boldsymbol{\mu_i}|\boldsymbol{y_i}) = \prod_{t=1}^{91} \left[\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right)^{y_{it}} \right] \times \\ \prod_{t=1}^{91} e^{-\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right)}. \end{aligned} \tag{2.4}$$

Since $\boldsymbol{\mu_i}$ is defined in terms of $\boldsymbol{K_i}$ and $\boldsymbol{\lambda_i}$, we update each K_{ij} and λ_{ij} to obtain an updated value of μ_{it}. We define the likelihood function for $\boldsymbol{K_i}$ and $\boldsymbol{\lambda_i}$, denoted $L(\boldsymbol{K_i}, \boldsymbol{\lambda_i}|\boldsymbol{y_i})$, as in (2.4). That is

$$\begin{aligned} L(\boldsymbol{K_i}, \boldsymbol{\lambda_i}|\boldsymbol{y_i}) = \prod_{t=1}^{91} \left[\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right)^{y_{it}} \right] \times \\ \prod_{t=1}^{91} e^{-\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right)}. \end{aligned}$$

It should be noted that equation (2.5) may evaluate to values that exceed our computational precision. To alleviate this problem, we compute the log-likelihood function for each parameter. Taking the natural log of (2.5) yields

$$\begin{aligned} \ln(L(\boldsymbol{K_i}, \boldsymbol{\lambda_i}|\boldsymbol{y_i})) = \sum_{t=1}^{91} \left[y_{it} \ln \left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right) \right] + \\ - \sum_{t=1}^{91} \left[\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}}(t - K_{ij}) + \lambda_{ij} \right) \right]. \end{aligned} \tag{2.5}$$

We define the posterior distributions for $\boldsymbol{K_i}$ and $\boldsymbol{\lambda_i}$, denoted $\pi(\boldsymbol{K_i}|\boldsymbol{\lambda_i}, \boldsymbol{y_i})$ and $\pi(\boldsymbol{\lambda_i}|\boldsymbol{K_i}, \boldsymbol{y_i})$ respectively, as the product of their likelihood function $L(\boldsymbol{K_i}, \boldsymbol{\lambda_i}|\boldsymbol{y_i})$ and prior distri-

butions, denoted $\pi(\boldsymbol{K_i})$ and $\pi(\boldsymbol{\lambda_i})$ respectively.

2.2.1 Updating λ_{ij}

Keeping with the Poisson parameter inherent restriction $\lambda_{ij} \geqslant 0$, for $j \in \{1, 2, \ldots, 5\}$, we choose a gamma prior distribution for each λ_{ij}. Consequently, the prior distribution for $\boldsymbol{\lambda_i}$ is the product of the gamma density functions for each λ_{ij}. Specifically

$$\begin{aligned}\pi(\boldsymbol{\lambda_i}) &= \prod_{j=1}^{5} \left[\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)} {\lambda_{ij}}^{\alpha_j - 1} e^{-\beta_j \lambda_{ij}} \right] \\ &\propto \prod_{j=1}^{5} \left[{\lambda_{ij}}^{\alpha_j - 1} e^{-\beta_j \lambda_{ij}} \right].\end{aligned}$$

Again, utilizing the natural log for computer implementation gives:

$$\ln(\pi(\boldsymbol{\lambda_i})) = \sum_{j=1}^{5} \left[(\alpha_j - 1) \ln(\lambda_{ij}) - \beta_j \lambda_{ij} \right].$$

When implementing the Metropolis sampling algorithm for a proposed value of λ_{ij}, we need to compute the difference in log posterior densities for μ_{it}, as evaluated at the proposed value $\mu^*_{it\lambda}$ and the current value $\mu^c_{it\lambda}$. Here, we define $\mu^*_{it\lambda}$ and $\mu^c_{it\lambda}$ as in (2.3) using the proposed and current value of λ_{ij}, λ^*_{ij} and λ^c_{ij} respectively. Hence, the natural log of the acceptance probability $\boldsymbol{\gamma^*}$ is given by

$$\begin{aligned}\ln(\boldsymbol{\gamma^*}) = &\sum_{t=1}^{91} \left[y_{it} \ln(\mu^*_{it\lambda}) - \mu^*_{it\lambda} \right] + \sum_{j-1}^{5} \left[(\alpha_j - 1) \ln(\lambda^*_{ij}) - \beta_j \lambda^*_{ij} \right] \\ &- \sum_{t=1}^{91} \left[y_{it} \ln(\mu^c_{it\lambda}) - \mu^c_{it\lambda} \right] - \sum_{j=1}^{5} \left[(\alpha_j - 1) \ln(\lambda^c_{ij}) - \beta_j \lambda^c_{ij} \right]\end{aligned}$$

We accept λ^*_{ij} with probability $\boldsymbol{\gamma^*}$ and keep our current value λ^c_{ij} with probability $1 - \boldsymbol{\gamma^*}$. We update each λ_{i1} for all people in the group $i \in \{1, 2, \ldots, N\}$ using the

current values of all other parameters when evaluating from the posterior distribution of λ_{i1}.

2.2.2 Updating α_j

Inorder to update α_j, we must assign a likelihood function and prior distribution for $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_5\}$. Since α_j is the shape parameter of a gamma distribution, for N people in the group, its likelihood function is the product of N gamma densities

$$\begin{aligned} L(\alpha_j|\boldsymbol{\lambda_j}, \beta_j) &= \prod_{i=1}^{N} \left[\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)} {\lambda_{ij}}^{\alpha_j - 1} e^{-\beta_j \lambda_{ij}} \right] \\ &= \left(\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)} \right)^N \prod_{i=1}^{N} \left[{\lambda_{ij}}^{\alpha_j - 1} e^{-\beta_j \lambda_{ij}} \right] \\ &\propto \left(\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)} \right)^N \prod_{i=1}^{N} {\lambda_{ij}}^{\alpha_j - 1}, \end{aligned}$$

where $\boldsymbol{\lambda_j} = \{\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{Nj}\}$ represents every person in the group's λ_j. Taking the natural log of (2.6) gives

$$\ln(L(\boldsymbol{\alpha}|\boldsymbol{\lambda_j}, \boldsymbol{\beta})) = N(\alpha_j \ln(\beta_j) - \ln(\Gamma(\alpha_j))) + \sum_{i=1}^{N} (\alpha_j - 1) \ln(\lambda_{ij}).$$

We assign a noninformative uniform hyper-prior on $\boldsymbol{\alpha}$. For $j \in \{1, 2, \ldots, 5\}$, we get

$$\pi(\alpha_j) = \frac{1}{1000} \propto 1.$$

Again, utilizing the natural log yields

$$\ln(\pi(\alpha_j)) = \ln(1) = 0.$$

When implementing the Metropolis sampling algorithm, the log acceptance probability $\ln(\boldsymbol{\gamma^*})$ for α_j^* is

$$\begin{aligned}\ln(\boldsymbol{\gamma^*}) = N(\alpha_j^* \ln(\beta_j) - \ln(\Gamma(\alpha_j^*))) + \sum_{i=1}^{N}(\alpha_j^* - 1)\ln(\lambda_{ij}) \\ - N(\alpha_j^c \ln(\beta_j) - \ln(\Gamma(\alpha_j^c))) - \sum_{i=1}^{N}(\alpha_j^c - 1)\ln(\lambda_{ij}).\end{aligned}$$

2.2.3 Updating β_j

We define β_j as the scale parameter of a gamma distibution. Therefore, β_j has a likelihood equivalent to the product of N gamma density functions

$$\begin{aligned}L(\beta_j|\boldsymbol{\lambda_j}, \alpha_j) &= \prod_{i=1}^{N}\left[\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)}{\lambda_{ij}}^{\alpha_j - 1}e^{-\beta_j\lambda_{ij}}\right] \\ &= \left(\frac{{\beta_j}^{\alpha_j}}{\Gamma(\alpha_j)}\right)^N \prod_{i=1}^{N}\left[{\lambda_{ij}}^{\alpha_j - 1}e^{-\beta_j\lambda_{ij}}\right] \\ &\propto {\beta_j}^{N\alpha_j}\prod_{i=1}^{N}e^{-\beta_j\lambda_{ij}}. \qquad (2.6)\end{aligned}$$

Taking the natural log of (2.6) gives

$$\ln(L(\beta_j|\boldsymbol{\lambda_j}, \alpha_j)) = N\alpha_j \ln(\beta_j) + \sum_{i=1}^{N} -\beta_j\lambda_{ij}.$$

We again assign a noninformative uniform hyper-prior for $\boldsymbol{\beta}$. Then, in updating β_j, we calculate the difference in log postierior densities for β_j^* and β_j^c which gives:

$$\begin{aligned}\ln(\boldsymbol{\gamma^*}) = N\alpha_j \ln(\beta_j^*) + \sum_{i=1}^{N} -\beta_j^*\lambda_{ij} \\ - N\alpha_j \ln(\beta_j^c) - \sum_{i=1}^{N} -\beta_j^c\lambda_{ij}.\end{aligned}$$

We accept β_j^* as a new hierarchical parameter with probability $\boldsymbol{\gamma^*}$. After updating

β_1, we repeat this process for the remaining $\lambda_{ij}{}'s$, $\alpha_j{}'s$, and $\beta_j{}'s$. We then implement the Metropolis algorithm for sampling values of K_{i3} and K_{i4} for all people in the group.

2.2.4 Updating $\boldsymbol{K_i}$

In updating $\boldsymbol{K_i}$, we must first obtain both the likelihood function and the prior distribution. We define the log-likelihood function for $\boldsymbol{K_i}$ as in (2.5):

$$\ln(L(\boldsymbol{K_i}, \boldsymbol{\lambda_i}|\boldsymbol{y_i})) = \sum_{t=1}^{91} \left[y_{it} \ln \left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}} (t - K_{ij}) + \lambda_{ij} \right) \right] - \sum_{t=1}^{91} \left[\left(\frac{\lambda_{i(j+1)} - \lambda_{ij}}{K_{i(j+1)} - K_{ij}} (t - K_{ij}) + \lambda_{ij} \right) \right].$$

Assigning a prior distribution for $\boldsymbol{K_i}$ can be accomplished by interpreting these time partitions in terms of the distance between them. Since K_{i1},K_{i2}, and K_{i5} are fixed, we need only to consider the distance between K_{i2} and K_{i3}, K_{i3} and K_{i4}, and also K_{i4} and K_{i5}. To maintain 4 distinct time partitions, we restrict the distance between any to $K_{ij}{}'s$ to be greater than or equal to one. We introduce for person i

$$\boldsymbol{x_i} = \{x_{i1}, x_{i2}, x_{i3}\}, \quad \text{where } \; x_{ij} \in \{0, 1, 2, \ldots, 80\} \text{ and } \sum_{j=1}^{3} x_{ij} = 80.$$

We can now define K_{i3} and K_{i4} as

$$K_{i3} = K_{i2} + 1 + x_{i1} \;\; = \;\; 8.5 + x_{i1} \tag{2.7}$$

$$K_{i4} = K_{i3} + 1 + x_{i2}. \tag{2.8}$$

We let $\boldsymbol{x_i}$ be from a multinomial distribution

$$\boldsymbol{x_i} \backsim multinomial(80, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3\}$ are probabilities satisfying $\sum_{j=1}^{3} \theta_j = 1$. Thus, the prior distribution for $\boldsymbol{K_i}$ is represented through a prior distribution for $\boldsymbol{x_i}$

$$\pi(\boldsymbol{x_i}) \;=\; 80! \prod_{j=1}^{3} \frac{{\theta_j}^{x_{ij}}}{x_{ij}!} \;\propto\; \prod_{j=1}^{3} \frac{{\theta_j}^{x_{ij}}}{x_{ij}!}. \tag{2.9}$$

The natural log of (2.9) is then

$$\ln(\pi(\boldsymbol{x_i})) = \sum_{j=1}^{3} \left[x_{ij} \ln(\theta_j) - \ln(x_{ij}!) \right].$$

We propose a value for ${\boldsymbol{x_i}}^* = \{x_{i1}^*, x_{i2}^*, x_{i3}^*\}$ based on ${\boldsymbol{x_i}}^c = \{x_{i1}^c, x_{i2}^c, x_{i3}^c\}$, and define ${\boldsymbol{K_i}}^* = \{0.5, 7.5, K_{i3}^*, K_{i4}^*, 91.5\}$ and ${\boldsymbol{K_i}}^c = \{0.5, 7.5, K_{i3}^c, K_{i4}^c, 91.5\}$ as in (2.7) and (2.8) using ${\boldsymbol{x_i}}^*$ and ${\boldsymbol{x_i}}^c$ respectively. We also define μ_{itK}^* and μ_{itK}^c as in (2.3) using ${\boldsymbol{K_i}}^*$ and ${\boldsymbol{K_i}}^c$. We now calculate the log-acceptance probability as

$$\begin{aligned} \ln(\boldsymbol{\gamma^*}) = &\sum_{t=1}^{91} \left[\ln(\mu_{itK}^*) - \mu_{itK}^* \right] + \sum_{j=1}^{3} \left[x_{ij}^* \ln(\theta_j) - \ln(x_{ij}^*!) \right] \\ &- \sum_{t=1}^{91} \left[\ln(\mu_{itK}^c) - \mu_{itK}^c \right] - \sum_{j=1}^{3} \left[x_{ij}^c \ln(\theta_j) - \ln(x_{ij}^c!) \right]. \end{aligned}$$

We accept our new value of ${\boldsymbol{x_i}}^*$ and ${\boldsymbol{K_i}}^*$ with probability $\boldsymbol{\gamma^*}$ and retain our current values with probability $1 - \boldsymbol{\gamma^*}$.

2.2.5 Updating $\boldsymbol{\theta}$

Next, we must update the hierarchical parameters $\boldsymbol{\theta}$, based on the values of all $\boldsymbol{x_i}'s$, denoted $\boldsymbol{X}$, and defined as the $N \times 3$ matrix of each person's $\boldsymbol{x_i}$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & x_{N3} \end{bmatrix}.$$

Defining $\boldsymbol{\theta}$ as the parameters for a multinomial distribution, the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{X}) &= \prod_{i=1}^{N} \left[\frac{{\theta_1}^{x_{i1}}}{x_{i1}!} \frac{{\theta_2}^{x_{i2}}}{x_{i2}!} \frac{{\theta_3}^{x_{i3}}}{x_{i3}!} \right] \\ &\propto \prod_{i=1}^{N} \left[{\theta_1}^{x_{i1}} {\theta_2}^{x_{i2}} {\theta_3}^{x_{i3}} \right]. \end{aligned} \tag{2.10}$$

We assign a Dirichelet hyper-prior for $\boldsymbol{\theta}$

$$\boldsymbol{\theta} \backsim Dirichlet(1, 1, 1),$$

which gives a prior density of

$$\pi(\boldsymbol{\theta}) = \Gamma(3) \prod_{j=1}^{3} \frac{{\theta_j}^{(1-1)}}{\Gamma(1)} \propto 1. \tag{2.11}$$

Taking the product of (2.10) and (2.11) gives us the posterior distribution for $\boldsymbol{\theta}$

$$\begin{aligned} \pi(\boldsymbol{\theta}|\boldsymbol{x_i}) &= \prod_{i=1}^{N} \left[{\theta_1}^{x_{i1}} {\theta_2}^{x_{i2}} {\theta_3}^{x_{i3}} \right] \\ &= {\theta_1}^{\sum_{i=1}^{N} x_{i1}} {\theta_2}^{\sum_{i=1}^{N} x_{i2}} {\theta_3}^{\sum_{i=1}^{N} x_{i3}}. \end{aligned} \tag{2.12}$$

We recognize the full conditional distribution of $\boldsymbol{\theta}$ from (2.12) as

$$\boldsymbol{\theta} \backsim Dirichlet\left(1 + \sum_{i=1}^{N} x_{i1},\ 1 + \sum_{i=1}^{N} x_{i2},\ 1 + \sum_{i=1}^{N} x_{i3} \right).$$

Lastly, we implement Gibbs sampling technique to update $\boldsymbol{\theta}$ from its full conditional distribution.

2.3 Calculations of Metropolis-Hastings Correction Factors

Whenever there is a restriction placed on a the support for the parameter being updated through MCMC (i.e. $\theta > 0$), a correction factor, also known as a Hastings ratio, must be utilized to adjust the acceptance probability $\boldsymbol{\gamma^*}$. It should be noted here that these restrictions are natural due to the corresponding probability distributions for each parameter (except for α_j and β_j which where capped at 100 and 10 respectively to keep from sampling unrealistic variances for the gamma distributions). We then multiply the acceptance probability $\boldsymbol{\gamma^*}$ by these Hasting ratios to adjust the acceptance probability by preventing the sampling technique from being biased against a parameter value near a boundary. Table (2.1) represents the log-Hasting ratios for λ_{ij}, α_j, and β_j.

Parameter	**Propsal Radius**(k)	**Restriction**	$\ln($**Hastings ratio**$)$
λ_{ij}	$k = .5$	$\lambda_{ij} > 0$	$\ln(min(2k, k + \lambda_{ij}^c))$ $- \ln(min(2k, k + \lambda_{ij}^*))$
α_j	$k = 1$	$0 < \alpha_j < 100$	$\ln(min(2k, k + \alpha_j^c, k + (100 - \alpha_j^c))$ $- \ln(min(2k, k + \alpha_j^*, k + (100 - \alpha_j^*))$
β_j	$k = 1$	$0 < \beta_j < 10$	$\ln(min(2k, k + \beta_j^c, k + 10 - \beta_j^c))$ $- \ln(min(2k, k + \beta_j^*, k + 10 - \beta_j^*))$

Table 2.1: Restricted Parameters λ_{ij}, α_j, and β_j and Corresponding Natural Log Hasting ratios.

It should be noted that the proposal density for λ_{ij}, α_j, and β_j is uniform with radius $k \in \mathbb{R}^+$. In proposing a value for $\boldsymbol{x_i}$, we propose $\boldsymbol{x_{i1}^*}$ and x_{i2}^* from a discrete

uniform distribution of radius 1 centered at the current value of x_{ij}. That is

$$x_{i1}^* \in \{x_{i1}^c - 1, x_{i1}^c, x_{i1}^c + 1\}, \quad \text{and}$$
$$x_{i2}^* \in \{x_{i2}^c - 1, x_{i2}^c, x_{i2}^c + 1\}$$

This prevents the Hastings ratio for $\boldsymbol{x_i}$ from being presented in closed form. Table (2.2) describes the sequential logical conditions for the determination of the log-Hastings ratio.

Step	**Case**	**One part of** $\ln$**(Hastings ratio)**
1	$x_{i2} = 0 \quad \&$ $x_{i1} = 0$	$\ln(\frac{1}{2})$
2	$x_{i2} = 0 \quad \&$ $x_{i1} = 80$	$\ln(\frac{1}{2})$
3	$x_{i2} = 0 \quad \&$ $0 < x_{i1} < 80$	$\ln(\frac{1}{6})$
4	$x_{i2} = 80$	$\ln(\frac{1}{2})$
5	$0 < x_{i2} < 80 \quad \&$ $x_{i1} = 0$	$\ln(\frac{1}{6})$
6	$0 < x_{i2} < 80 \quad \&$ $0 < x_{i1} < 80$	$\ln(\frac{1}{9})$

Table 2.2: Log-Hastings Ratio for $\boldsymbol{x_i}$

For clarity on how to use Table (2.2) to obtain a Hastings ratio for $\boldsymbol{x_i}$, we offer two examples. The first is when neither $\boldsymbol{x_i}^c$ nor $\boldsymbol{x_i}^*$ are close to a boundary. Let

$$\boldsymbol{x_i}^c = \{25, 25, 30\} \text{ and } \boldsymbol{x_i}^* = \{24, 26, 30\}.$$

We define the Hastings ratio HR as

$$HR = \frac{q(\boldsymbol{x_i}^c|\boldsymbol{x_i}^*)}{q(\boldsymbol{x_i}^*|\boldsymbol{x_i}^c)}. \tag{2.13}$$

Here, $q(\boldsymbol{x_i}^c|\boldsymbol{x_i}^*)$ and $q(\boldsymbol{x_i}^*|\boldsymbol{x_i}^c)$ represent the probability of proposing $\boldsymbol{x_i}^c$ given $\boldsymbol{x_i}^*$,

and the probability of proposing $\boldsymbol{x_i}^c$ given x^*. Taking the natural log of (2.13) gives

$$\ln(HR) = \ln(q(\boldsymbol{x_i}^c|\boldsymbol{x_i}^*)) - \ln(q(\boldsymbol{x_i}^*|\boldsymbol{x_i}^c)). \tag{2.14}$$

Using (2.14) to find the log-HR for updating $\boldsymbol{x_i}$, we see that the log-HR for our first example is

$$\ln(HR) = \ln\left(\frac{1}{9}\right) - \ln\left(\frac{1}{9}\right) = 0.$$

For a second example, let $\boldsymbol{x_i}^c = \{1, 79, 0\}$ and $\boldsymbol{x_i}^* = \{0, 80, 0\}$. Then, using (2.14) to find the log-HR, we get

$$\ln(HR) = \ln\left(\frac{1}{9}\right) - \ln\left(\frac{1}{2}\right).$$

It should be noted that x_{i3} is strictly dependent on x_{i1} and x_{i2}. Therefore, in proposing $\boldsymbol{x_i}^*$, we propose only x^*_{i1} and x^*_{i2} and calculate $x^*_{i3} = 80 - x^*_{i1} - x^*_{i2}$. Thus, in calculating the HR for $\boldsymbol{x_i}$, we need not consider x_{i3}.

Chapter 3

Discussion

3.1 Results

Having applied our model seprately to all three groups, we compare the posterior distributions of $\boldsymbol{\mu_i}$ for each application to compare treatness effectiveness between the three groups. Figures (3.1), (3.2), and (3.3) represent these posterior distributions. It should be noted that the high variance in μ_{it} near the endpoints of the time support is natural for this model. For a given time t, μ_{it} is dependent on values both before and after itself (i.e $\mu_{i(t-1)}$ and $\mu_{i(t+1)}$). When t is closer to an endpoint, we lose information for μ_{it} on the bounded side of t. This in turn presents a higher variance in μ_{it} when t is near 0 or 91, as compared to when t is near the midpoint of the time support. When comparing Figures (3.1), (3.2), and (3.3) for treatment effectiveness, it is important to consider the trend of the mean for each distribution over time, rather than the differences in the function values at a specific time. Since the placebo frequencies are, on day 1, significantly higher on average than the treatment frequencies, we expect there to be a lower group mean profile for the treatment group. "Parallel" profiles, as depicted in Figures (3.1) and (3.2), suggest no difference between treatment and placebo.

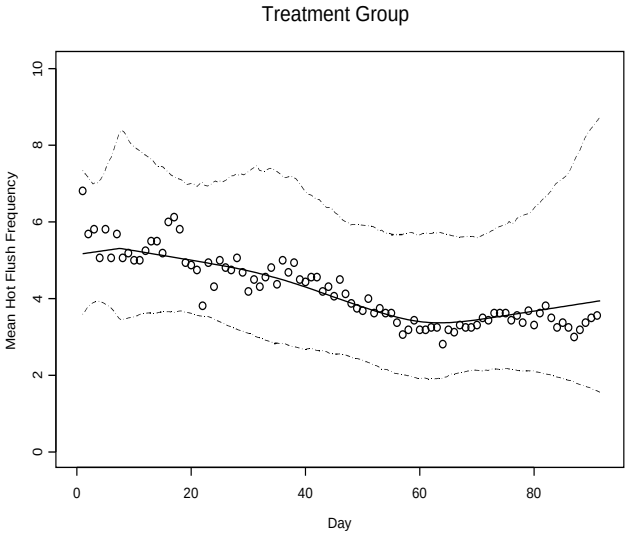

Figure 3.1: Posterior Distributions for Treatment Group.

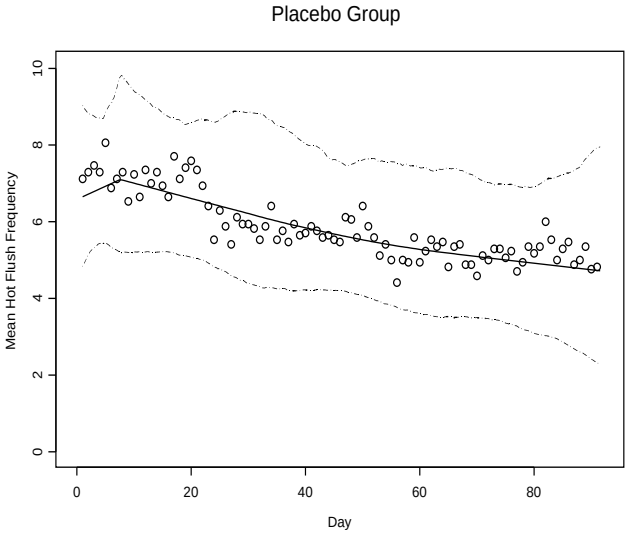

Figure 3.2: Posterior Distributions for Placebo Group.

It is evident from these plots that the group mean for the treatment group is less than both the education group mean and the placebo group mean over the entire time support. It is also apparent that the means of the posterior distributions for the treatment and placebo group decrease over time, as where the educational group

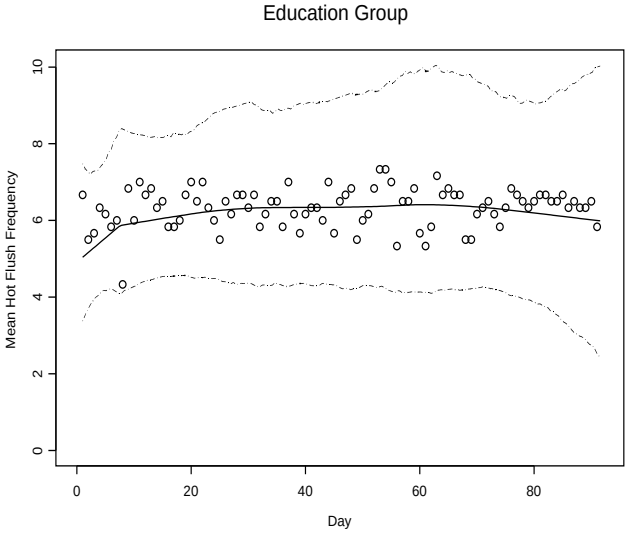

Figure 3.3: Posterior Distributions for Education Group.

mean tends to stay constant. Figure (3.4) (left) represents the probability that the treatment mean is less than the placebo mean at a given time. Since both posterior distribution's means where decreasing at the same rate, we see that the probability in Figure (3.4) (left) is constant for the first 12 weeks. This implies that both effective and uneffective acupuncture relieve menopausal symptoms. Figures (3.5) (left) and (3.6) (left) also imply the same result. As time increases, the probability that the treatment or placebo group mean is less than the educaton group mean increases over time. It should be noted that the dip in probability towards the end of the time interval is due to the high variance in μ_{it} as t approaches 91.

Figures (3.4) (right), (3.5) (right), and (3.6) (right) further support these results. These graphs are a time series of boxplots. At each time is a boxplot representing a random sample of size 1000; each event in the sample is the difference in two randomly sampled μ_{it}'s from their respected posterior distributions. The increasing patterns in Figures (3.5) (right) and (3.6) (right) suffice in showing that as time increases, the difference in group means is increasing; implying a positive acupuncture effect in

relieving menopausal symptoms.

Treatment vs. Placebo

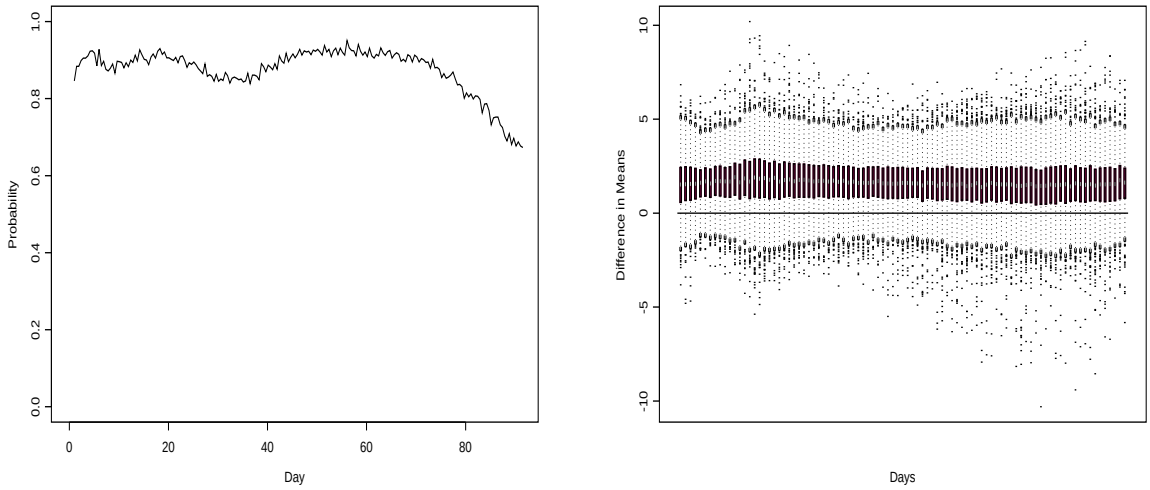

Figure 3.4: **Left:** P($\mu_{it}{}^{Tr} < \mu_{it}{}^{Pl}$) **Right:** Boxplot of ($\mu_{it}{}^{Pl} - \mu_{it}{}^{Tr}$)

Treatment vs. Education

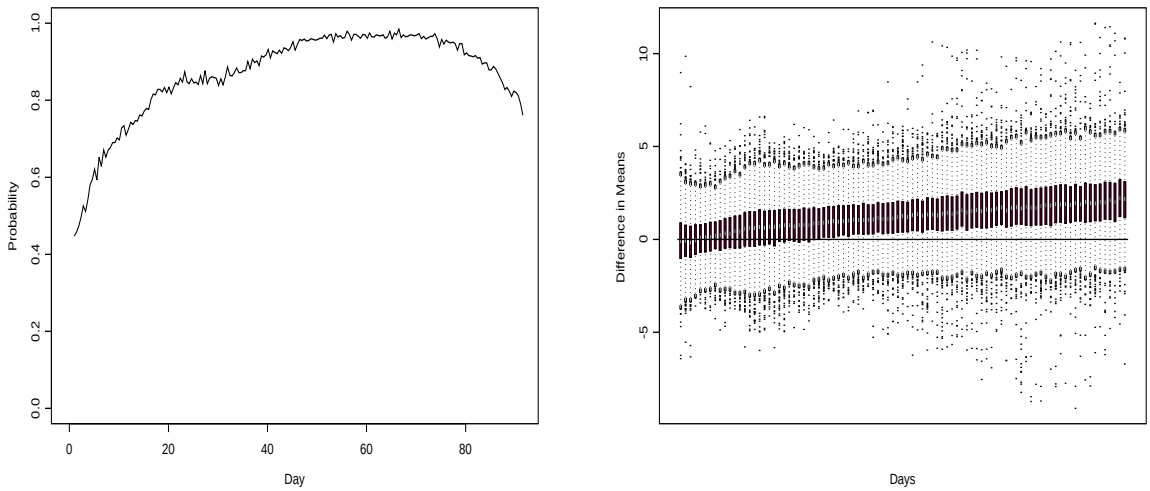

Figure 3.5: **Left:** P($\mu_{it}{}^{Tr} < \mu_{it}{}^{Ed}$) **Right:** Boxplot of ($\mu_{it}{}^{Ed} - \mu_{it}{}^{Tr}$)

One benefit of this model is its ability to describe a single profile within a group, and treat this profile as a random sample from the group's distribution. Figures (3.7),

Placebo vs. Education

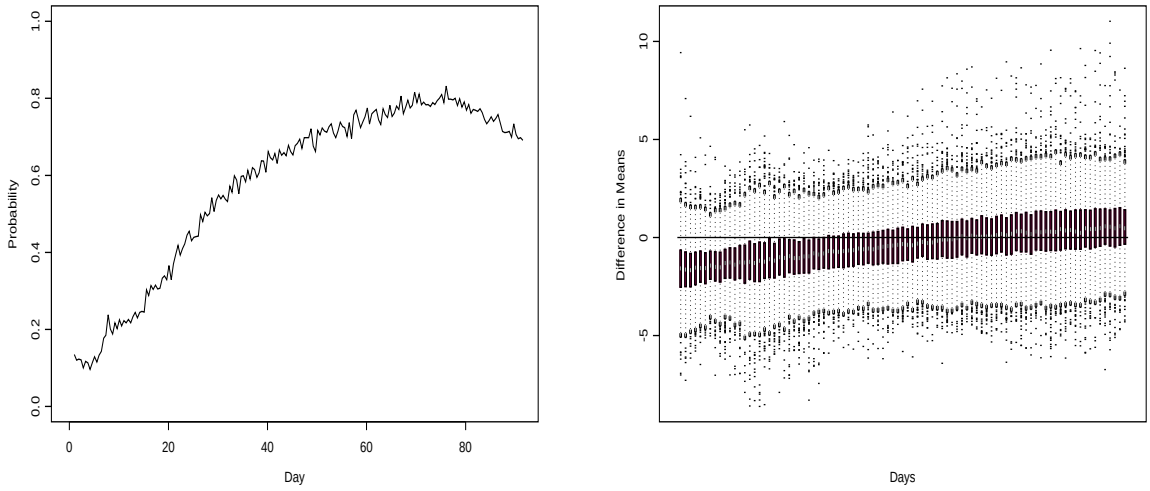

Figure 3.6: **Left:** P($\mu_{it}{}^{Pl} < \mu_{it}{}^{Ed}$) **Right:** Boxplot of ($\mu_{it}{}^{Pl} - \mu_{it}{}^{Ed}$)

(3.8), and (3.9) show two estimated profiles for two individuals from each group. Actual study data is included in the plot, along with a 95% credible interval for μ_{it}. We point out that infomation from each profile of the group was used to make inferences about the corresponding group mean in Figures (3.1), (3.2), or (3.3).

3.2 Future Work

Further research includes using Bayes Factors (Berger, et al 1996)to compare the fit of this model with that of previously applied models mentioned in Section 1.1. In addition, coding a general version of this model (which accepts any study duration, number of fixed and random time knots, location of random and fixed knots, etc.) would greatly improve its appeal.

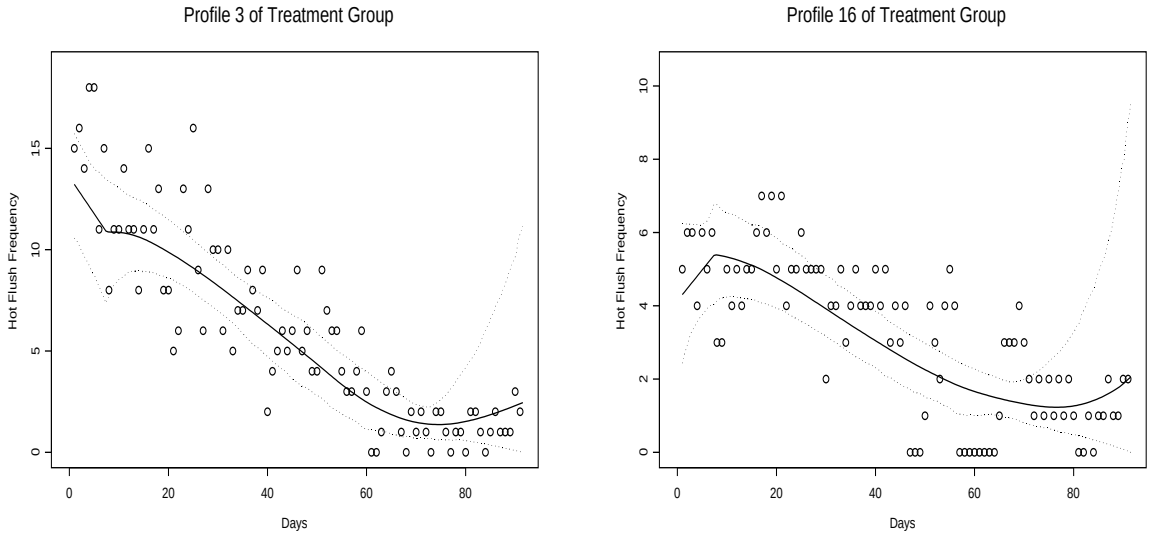

Figure 3.7: Profile plots for person 3 and 16 in Treatment Group.

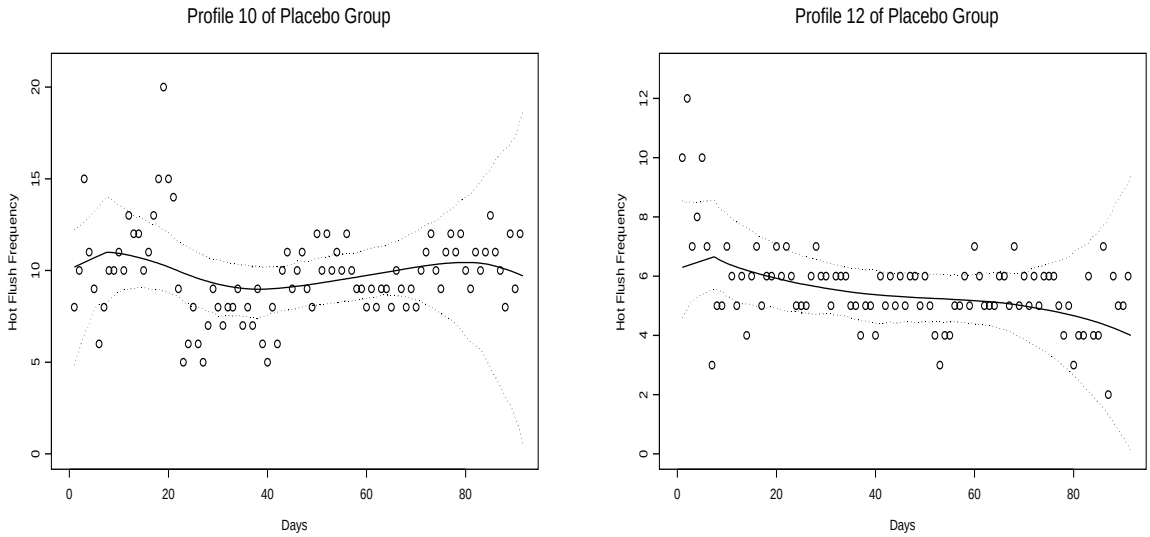

Figure 3.8: Profile plots for person 10 and 12 in Placebo Group.

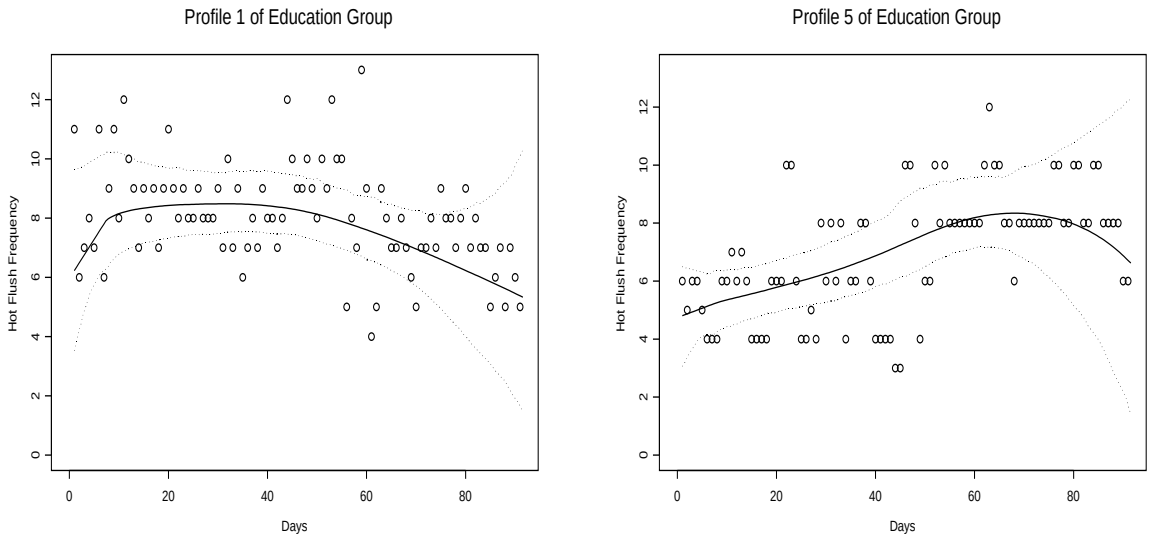

Figure 3.9: Profile plots for person 1 and 5 in Education Group.

References

Berger, J.O., Pericchi, R. (1996). "The Intricnsic Bayes Factor for Model Selection and Prediction." *Journal of American Statistical Association*, Vol 91, No. 433, pp 109–121.

Borgesi, Jennifer. (2003). A Piecwise Linear Generalized Poisson Regression Approach to Modeling Longitudinal Frequency Data. *Master's Thesis, Duquesne University.*

Gelman, A., Carlin, J.B.,Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis.* London: Chapman and Hall.

Jordan, Joeseph. 2005. Bayesian Hierarchical Modeling for Longitudinal Frequency Data. *Master's Thesis, Duquesne University.*

Kern II, J.C. and Cohen, S. (2005). "Menopausal Symptom Relief with Acupuncuture: Bayesian Analysis using Piecwise Regression." *Communications in Statistics: Simulation and Computation*, Vol 34, 3, pp 738–798.

Lopes, H.F., Muller, P. and Rosner, G.L. (2003). "Bayesian Meta-Analysis for Longitudinal Data Modes Using Multivariate Mixture Priors." *Biometrics*, 59, 66–75.