Summer 2005

Bayesian Hierarchical Modeling for Longitudinal Frequency Data

Joseph Patrick Jordan

Bayesian Hierarchical Modeling for Longitudinal Frequency 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

Master of Science

by

Joseph Jordan

June 22, 2005

Joseph Jordan

Bayesian Hierarchical Modeling for Longitudinal Frequency Data

Master of Science in Computational Mathematics

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

June 22, 2005

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

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

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

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

Contents

List of Figures

Chapter 1

Introduction

1.1 Objective

The objective of this research is to develop a longitudinal frequency model for data collected regularly for several individuals over an extended time period. This model must recognize explicitly the discrete nature of the data, as well as any dependence that exists among an individual's time consecutive measurements. Motivated by a study investigating alternative treatments for relief of menopausal symptoms, we apply this model to actual study data in an effort to compare treatment effectiveness.

We propose a Bayesian hierarchical model to describe not only the frequency measurements (profile) of each individual, but also the parameters that govern individual profiles. One of the main benefits of such a hierarchical model is that by utilizing a population distribution to structure dependence into the parameters, the model can adequately fit the data without the problem of overfitting (Gelman, et al 1995). This hierarchical model will be built upon hyperparameters that take into account previous measures in estimating the parameters of the proposed

distribution.

1.2 The Model

Let y_{ij} represent the observed frequency at time i for person j. We recognize the discrete nature of the data by assuming

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}),$$

and define $\mu_{ij} = \ln(\lambda_{ij})$ to more easily model the prior distribution on λ_{ij}. Specifically, we let

$$\mu_{ij} \sim \text{N}(\beta_j + \rho_j(\mu_{(i-1)j} - \beta_j), \sigma_j^2).$$

In this way we recognize an individual's data-generating mean at time i to be normally distributed about a number that is a function of the data-generating mean at time $i-1$. Note that β_j can be thought of as the overall mean of the jth profile; ρ_j the strength of the time dependence; and σ_j^2 the variability of daily means. These hyperparameters β_j, ρ_j, and σ_j^2 are modeled as follows:

$$\beta_j \sim \text{N}(\beta_0, \sigma_\beta^2); \quad \eta_j = \frac{1+\rho_j}{2} \sim \text{Beta}(\kappa\omega, \kappa(1-\omega)); \quad \sigma_j^2 \sim \text{Inv-gamma}(a, b).$$

It is through the distributions of these hyperparameters that we can compare effectiveness among treatment groups. In particular, by applying this model to different treatment groups, inference on β_0 for each group can be compared. Finally, we assign non-informative or vague priors to the following:

$$\beta_0, \sigma_\beta^2 \overset{iid}{\sim} \mathrm{U}(0, 1000); \quad a, b \overset{iid}{\sim} \mathrm{U}(0, 50); \quad \omega \sim \mathrm{U}(0, 1); \quad \kappa \sim \mathrm{Gamma}(0.1, 0.1).$$

These vague prior distributions were chosen to reflect a lack of parameter knowledge.

1.3 The Data and Model Simulation

We point out that it is possible to select parameter values that yield profile realizations reminiscent of those observed in the motivating clinical trial. Shown in Figures 1 and 2 are simulated profiles with fixed hyperparameters β_j, ρ_j, and σ_j^2. Note that these profiles are representative of those that are observed in the actual clinical dataset (profiles pictured in Figures 3 and 4). That the proposed model is capable of simulating representative profiles is further support for its use. Any set of simulated test data is unique from that of the clinical data, as we know the parameter values from which the test data was generated. Unknown parameter values associated with the clinical data will be estimated via standard Markov Chain Monte Carlo (MCMC) techniques, including Metropolis-Hastings updating (Gilks, et al 1996).

The clinical data was obtained from a study conducted at Yale University, with funding from The Patrick and Catherine Weldon Donaghue Medical Foundation. This study observed menopausal women in breast cancer remission. The aim of this study was to examine the effects of acupuncture as a treatment to relieve symptoms associated with menopause. Since these women have a history of breast cancer, traditional hormone therapy for the relief of the symptoms of menopause

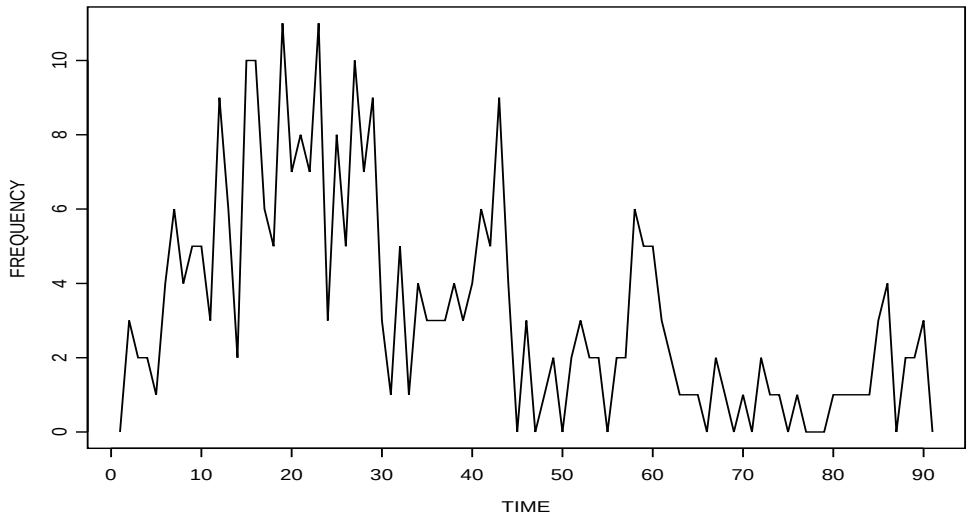

Figure 1.1: Simulated subject profile: $\beta_j = .5,\ \rho_j = .9,\ \sigma_j^2 = .5$.

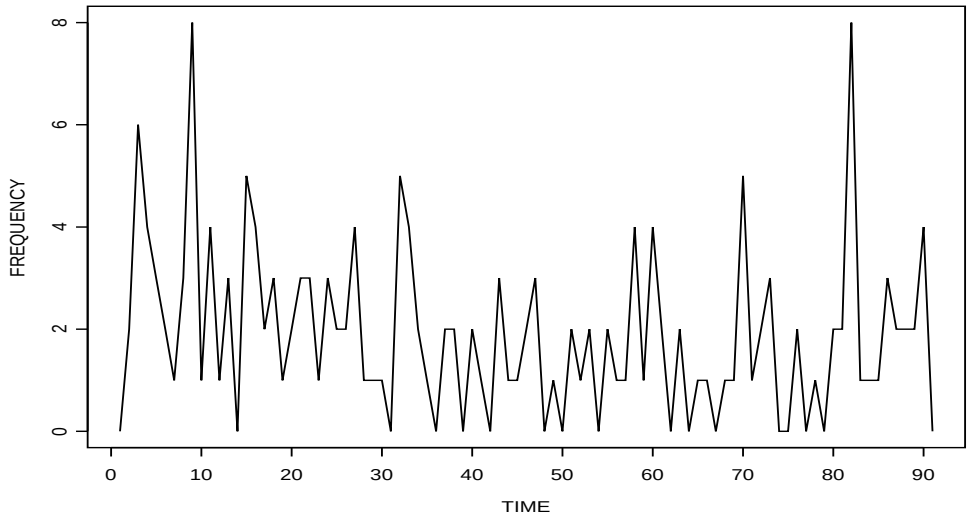

Figure 1.2: Simulated subject profile: $\beta_j = .5,\ \rho_j = .5,\ \sigma_j^2 = .5$.

was not a recommended option. Women enrolled in this study were randomly placed into one of three different groups: a control group (6 individuals who were given weekly educational sessions to help the subject's understanding of midlife healthy living), a treatment group (16 individuals who were given weekly acupuncture on effective bodily areas), or a placebo group (17 individuals who

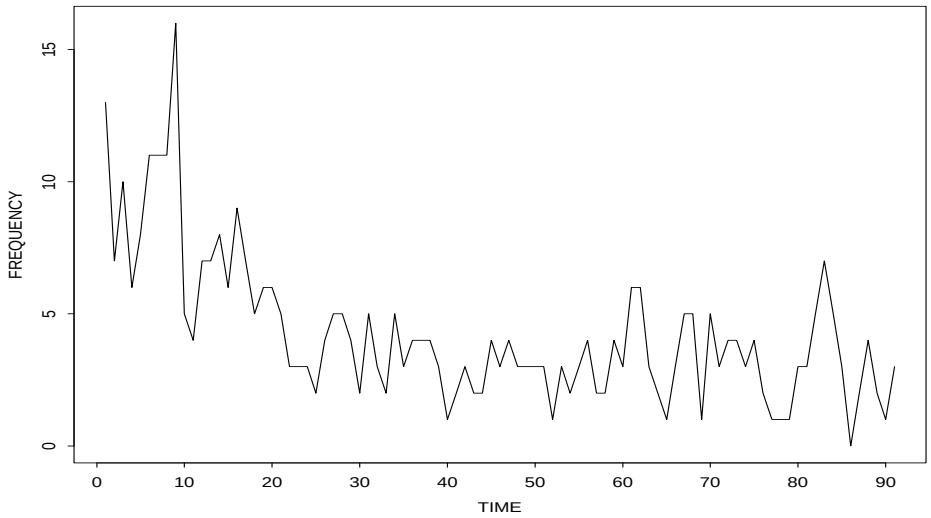

Figure 1.3: Actual subject profile.

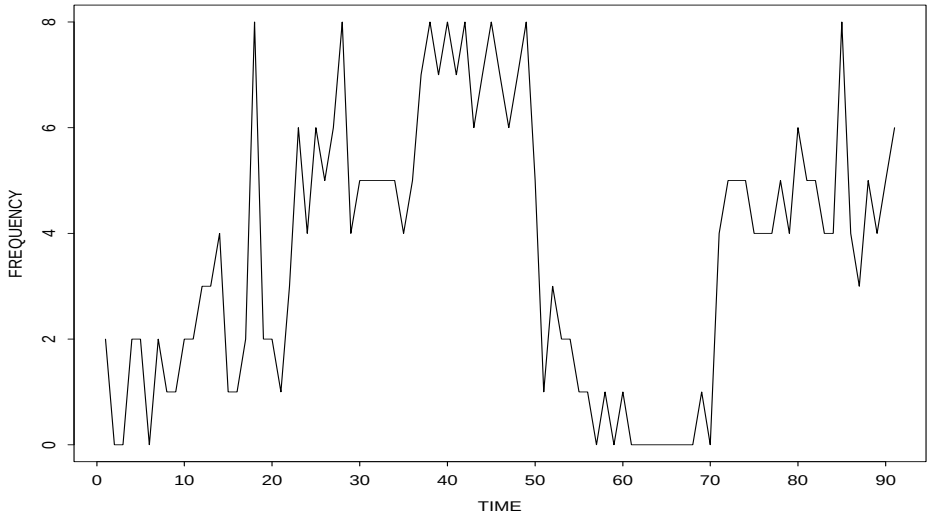

Figure 1.4: Actual subject profile.

were given weekly acupuncture on non-effective bodily areas). The length of this study was thirteen weeks, with the first week being the baseline week; treatment was administered for the following 12 weeks. Although several variables were recorded, the measurement of interest for this investigation is the number of hot flushes afflicting the women for each of the 91 days of observation. Unlike previous models fit to this data (Kern and Cohen 2005 and Borgesi 2004), the model we

implement explicitly recognizes time dependence through the prior distribution on the μ_{ij}'s. The resulting inference will allow for the measurement of treatment effectiveness over time. We will apply the model separately to competing groups and then compare posterior distributions of appropriate parameters.

Chapter 2

Model Implementation

2.1 Sampling Techniques

Using the model described in Section 1.2, we utilize MCMC techniques for parameter estimation. Specifically, we use Metropolis, Metropolis-Hastings, and Gibbs Sampling to draw from the joint posterior distribution of all parameters. To demonstrate these techniques, let $\mathbf{y}$ represent the data and $\theta_1, \theta_2, \ldots, \theta_m$ a vector of all model parameters. The joint posterior distribution $\pi(\theta_1, \ldots, \theta_m \mid \mathbf{y})$ is the product of the likelihood and prior distributions for $\theta_1, \ldots, \theta_m$, and is used in these sampling methods as follows:

Metropolis Sampling: Given a current value for our parameter, θ_1^c, propose a new value, θ_1^*, from a proposal density. The proposal density we choose is uniform with length $2k$, $k \in \mathbb{R}^+$, and centered on θ_1^c. Thus, θ_1^* is randomly chosen from the interval, $(\theta_1^c - k, \theta_1^c + k)$. We then accept θ_1^* with probability α:

$$\alpha = \frac{\pi(\theta_1^*, \theta_2^c, \ldots \mid \mathbf{y})}{\pi(\theta_1^c, \theta_2^c, \ldots \mid \mathbf{y})}.$$

We maintain our current value, θ_1^c with probability $1 - \alpha$.

Metropolis-Hastings Sampling: Follow the same algorithm as Metropolis sampling, with one exception: a Hastings ratio is used to adjust the acceptance probability α. Hastings ratios are necessary when updating bounded parameters and are discussed in Section 2.3.

Gibbs Sampling: From the joint posterior, obtain the full conditional distribution for the parameter of interest. We then sample the parameter of interest from its full conditional distribution and always accept the sampled value of our parameter. In cases where the full conditional is not recognizable, we resort to Metropolis or Metropolis-Hastings Sampling.

We iterate these sampling techniques appropriately through all θ_i's. For our research we implemented the above algorithm in a C program with 25 million iterations.

2.2 MCMC Calculations

In order to utilize MCMC one must know the joint posterior distribution (i.e. both the likelihood and prior) for all parameters of interest. Since the y_{ij}'s come from a Poisson distribution, the likelihood function for μ_{ij} is the product of Poisson mass functions. Each μ_{ij} is modeled to be the mean hot flush frequency for days $2i$ and $2i - 1$ for $i = 1, \ldots, 44$, with μ_{45j} representing the mean hot flush frequency for days 89, 90, and 91. This gives a total of 45 μ_{ij}'s, whose likelihood $L(\mu_{1j}, \mu_{2j}, \ldots, \mu_{45j})$ is the product of the following Poisson mass functions:

$$L(\mu_{1j}, \mu_{2j}, \ldots, \mu_{45j}) = \prod_{i=1}^{44} \left(\left(\frac{e^{-e^{\mu_{ij}}} e^{\mu_{ij} y_{(2i)j}}}{y_{(2i)j}!} \right) \left(\frac{e^{-e^{\mu_{ij}}} e^{\mu_{ij} y_{(2i-1)j}}}{y_{(2i-1)j}!} \right) \right)$$

$$
\begin{aligned}
&\times \left(\frac{e^{-e^{\mu_{45j}}} e^{\mu_{45j} y_{89j}}}{y_{89j}!}\right)\left(\frac{e^{-e^{\mu_{45j}}} e^{\mu_{45j} y_{90j}}}{y_{90j}!}\right)\left(\frac{e^{-e^{\mu_{45j}}} e^{\mu_{45j} y_{91j}}}{y_{91j}!}\right) \\
\propto\ & \prod_{i=1}^{44}\left[\left(e^{-e^{\mu_{ij}}}\right)^2\left(e^{\mu_{ij}(y_{(2i)j}+y_{(2i-1)j})}\right)\right]\left(e^{-e^{\mu_{45j}}}\right)^3\left(e^{\mu_{45j}(y_{89j}+y_{90j}+y_{91j})}\right). \\
=\ & \left(\prod_{i=1}^{44} e^{-2e^{\mu_{ij}}} e^{\mu_{ij}(y_{(2i-1)j}+y_{(2i)j})}\right)\left(e^{-3e^{\mu_{45j}}} e^{\mu_{45j}(y_{89j}+y_{90j}+y_{91j})}\right). \quad (2.1)
\end{aligned}
$$

It should be noted here that the above likelihood can evaluate to values that exceed our available computational precision. To alleviate this problem, we will instead evaluate the log-likelihood function for each parameter. Taking the natural log of (2.1) gives:

$$
\ln(L(\boldsymbol{\mu}_j)) = \left(\sum_{i=1}^{44} -2e^{\mu_{ij}} + (y_{(2i-1)j} + y_{(2i)j})\mu_{ij}\right) + (-3e^{\mu_{45j}} + (y_{89j} + y_{90j} + y_{91j})\mu_{45j}),
$$

where $\boldsymbol{\mu}_j = \{\mu_{1j}, \mu_{2j}, \ldots, \mu_{45j}\}$. Since we assumed that our μ_{ij}'s come from a normal distribution, their prior is the product of 45 normal densities:

$$
\begin{aligned}
\pi(\mu_{1j}, \mu_{2j}, \ldots, \mu_{45j}) &= \prod_{i=1}^{45}\left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\right) e^{\frac{-(\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{2\sigma_j^2}} \\
&\propto \prod_{i=1}^{45}\left(e^{\frac{-(\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{2\sigma_j^2}}\right).
\end{aligned}
$$

Thus,

$$
\pi(\boldsymbol{\mu}_j) \propto e^{\sum_{i=1}^{45} \frac{-(\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{2\sigma_j^2}}.
$$

Again, we will utilize the natural logarithm for computer implementation.

$$\ln(\pi(\boldsymbol{\mu}_j)) \;=\; \sum_{i=1}^{45} \frac{-(\mu_{ij} - (\beta_j + \rho_j(\mu_{(i-1)j} - \beta_j)))^2}{2\sigma_j^2}. \tag{2.2}$$

When implementing the Metropolis sampling algorithm to update μ_{ij}, we need to compute the difference in log posterior densities for μ_{ij}, as evaluated at the proposed μ_{ij}^* and the current μ_{ij}^c, respectively. Thus, the natural log of the acceptance probability α is given by

$$\begin{aligned}
\ln(\alpha) \;=\; & \left(\sum_{i=1}^{44} -2e^{\mu_{ij}^*} + (y_{(2i-1)j} + y_{(2i)j})\mu_{ij}^*\right) + \left(-3e^{\mu_{45j}^*} + (y_{89j} + y_{90j} + y_{91j})\mu_{45j}^*\right) \\
& + \sum_{i=1}^{45} \frac{-(\mu_{ij}^* - (\beta_j + \rho_j(\mu_{(i-1)j}^c - \beta_j)))^2}{2\sigma_j^2} \\
& - \left(\sum_{i=1}^{44} -2e^{\mu_{ij}^c} + (y_{(2i-1)j} + y_{(2i)j})\mu_{ij}^c\right) + \left(-3e^{\mu_{45j}^c} + (y_{89j} + y_{90j} + y_{91j})\mu_{45j}^c\right) \\
& - \sum_{i=1}^{45} \frac{-(\mu_{ij}^c - (\beta_j + \rho_j(\mu_{(i-1)j}^* - \beta_j)))^2}{2\sigma_j^2}.
\end{aligned}$$

Once we sample μ_{ij} for $i = 1, \ldots, 45$, we then sample β_j for $j = 1, \ldots, n$. This implementation of Metropolis sampling will utilize the current values of the μ_{ij}'s. We present the likelihood function for the β_j's, and derive the log-likelihood for implementation purposes. The likelihood for the β_j's is the product of 45 normal densities; the same normals that were used for the prior on the $\boldsymbol{\mu}_j$'s. The difference between the two is in which parameter is treated as unknown. In this case we treat the μ_{ij}'s as known and the β_j's as random:

$$L(\beta_j) \;=\; \prod_{i=1}^{45} \left(\frac{1}{\sqrt{2\pi\sigma_j^2}}\right) e^{\frac{\mu_{ij} - (\beta_j + \rho_j(\mu_{(i-1)j} - \beta_j)))^2}{-2\sigma_j^2}}$$

$$
\begin{aligned}
&\propto \prod_{i=1}^{45} \left(e^{\frac{\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{-2\sigma_j^2}} \right) \\
&= \prod_{i=1}^{45} e^{\frac{\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{-2\sigma_j^2}}. \\
\ln(L(\beta_j)) &= \sum_{i=1}^{45} \left(\frac{\mu_{ij}-(\beta_j+\rho_j(\mu_{(i-1)j}-\beta_j)))^2}{-2\sigma_j^2} \right). \qquad (2.3)
\end{aligned}
$$

We specify the prior for β_j to be a normal density:

$$
\begin{aligned}
\pi(\beta_j) &= \frac{1}{\sqrt{2\pi\sigma_\beta^2}} e^{-\frac{1}{2\sigma_\beta^2}(\beta_0-\beta_j)^2} \\
&\propto e^{-\frac{1}{2\sigma_\beta^2}(\beta_0-\beta_j)^2}.
\end{aligned}
$$

The natural log of this prior is:

$$
\ln(\pi(\beta_j)) = -\frac{1}{2\sigma_\beta^2}\left(\beta_0-\beta_j\right)^2.
$$

When implementing the Metropolis sampling algorithm, the log of the acceptance probability α for a proposed β_j^* is

$$
\begin{aligned}
\ln(\alpha) &= -\frac{1}{2\sigma_j^2}\left(\sum_{i=1}^{45} (\mu_{ij}-(\beta_j^*+\rho_j(\mu_{(i-1)j}-\beta_j^*)))^2) - (\mu_{ij}-(\beta_j^c+\rho_j(\mu_{(i-1)j}-\beta_j^c)))^2) \right) \\
&\quad -\frac{1}{2\sigma_\beta^2}\left((\beta_0-\beta_j^*)^2 - (\beta_0-\beta_j^c)^2 \right).
\end{aligned}
$$

After updating all β_j^*'s, we then sample new values of ρ_j for $j = 1, \ldots, n$. The likelihood function for ρ_j is the same as in (2.3), except we treat the ρ_j's as random. A beta prior for ρ_j gives:

$$
\pi(\rho_j) = \frac{\Gamma(\kappa\omega+\kappa(1-\omega))}{\Gamma(\kappa\omega)\Gamma(\kappa(1-\omega))} \left(\frac{1+\rho_j}{2}\right)^{\kappa\omega-1} \left(1-\frac{1+\rho_j}{2}\right)^{\kappa(1-\omega)-1}
$$

$$\propto \left(\frac{1+\rho_j}{2}\right)^{\kappa\omega-1}\left(\frac{1-\rho_j}{2}\right)^{\kappa(1-\omega)-1}.$$

The natural log of this prior is given by

$$\begin{aligned}\ln(\pi(\rho_j)) &= (\kappa\omega-1)(\ln(1+\rho_j)-\ln(2)) \\ &\quad +(\kappa(1-\omega)-1)(\ln(1-\rho)-\ln(2)). \qquad (2.4)\end{aligned}$$

The natural log of the acceptance probability α for a proposed ρ_j^* is then

$$\begin{aligned}= & -\frac{1}{2\sigma_j^2}\left(\sum_{i=1}^{45}(\mu_{ij}-(\beta_j+\rho_j^*(\mu_{(i-1)j}-\beta_j)))^2\right) \\ & +\frac{1}{2\sigma_j^2}\left(\sum_{i=1}^{45}(\mu_{ij}-(\beta_j+\rho_j^c(\mu_{(i-1)j}-\beta_j)))^2\right) \\ & +(\kappa\omega-1)(\ln(1+\rho_j^*)-\ln(2))+(\kappa(1-\omega)-1)(\ln(1-\rho_j^*)-\ln(2)) \\ & -(\kappa\omega-1)(\ln(1+\rho_j^c)-\ln(2))+(\kappa(1-\omega)-1)(\ln(1-\rho_j^c)-\ln(2)).\end{aligned}$$

We next turn our attention to updating σ_j^2, where we again have the same likelihood function as in (2.3). We choose an inverse-gamma prior

$$\begin{aligned}\pi(\sigma_j^2) &= \frac{b^a}{\Gamma(a)}(\sigma_j^2)^{-(a+1)}e^{\frac{-b}{\sigma_j^2}} \\ &\propto (\sigma_j^2)^{-(a+1)}e^{\frac{-b}{\sigma_j^2}}\end{aligned}$$

whose natural log is given by

$$\ln(\pi(\sigma_j^2)) = -(a+1)\ln(\sigma_j^2)-\frac{b}{\sigma_j^2}.$$

Given a current value σ_j^{2c}, the natural log of the acceptance probability α for a

proposed σ_j^{2*} is

$$\begin{aligned}\ln(\alpha) &= -\left(\frac{45}{2}\right)\ln(2\pi\sigma_j^{2*}) - \frac{1}{2\sigma_j^{2*}}\left(\sum_{i=1}^{45}(\mu_{ij} - (\beta_j + \rho_j(\mu_{(i-1)j} - \beta_j)))^2 - (a+1)\ln(\sigma_j^{2*}) - \frac{b}{\sigma_j^{2*}}\right) \\ &\quad + \left(\frac{45}{2}\right)\ln(2\pi\sigma_j^{2c}) - \frac{1}{2\sigma_j^{2c}}\left(\sum_{i=1}^{45}(\mu_{ij} - (\beta_j + \rho_j(\mu_{(i-1)j} - \beta_j)))^2 - (a+1)\ln(\sigma_j^{2c}) - \frac{b}{\sigma_j^{2c}}\right)\end{aligned}$$

Above we implemented the Metropolis-Hastings algorithm for $\mu_{ij}, i = 1, \ldots, 45$. We include the parameter μ_{0j} whose posterior is given by

$$-\frac{1}{2\sigma_j^2}\left((\mu_{1j} - (\beta_j + \rho_j(\mu_{0j}^* - \beta_j)))^2\right) + \frac{1}{2\sigma_j^2}\left((\mu_{1j} - (\beta_j + \rho_j(\mu_{0j}^c - \beta_j)))^2\right)$$

and is calculated without any data.

We pause from utilizing true Metropolis-Hastings sampling in our program when we update β_0 and σ_β^2. In order to implement Metropolis-Hastings, we do not need to know the full conditional distribution for the parameter we are updating. This is especially helpful since full conditionals are not always easily found. In the case where the full conditional distribution is easily obtained, Gibbs sampling provides a more efficient algorithm for updating parameters. We implemented Gibbs sampling for these two parameters since their Normal Inverse-Gamma joint posterior easily yields the following full conditional distributions (Gelman, et al 1995):

$$\beta_0 \mid \sigma_\beta^2 \sim \mathrm{N}\left(\bar{\beta}, \frac{\sigma_\beta^2}{n}\right); \qquad \sigma_\beta^2 \mid \beta_0 \sim \mathrm{IG}\left(\frac{n}{2}, \frac{1}{2}\left(\sum_{j=1}^{n}(\beta_j - \bar{\beta})^2 + n(\bar{\beta} - \beta_0)^2\right)\right),$$

where $\bar{\beta}$ represents the mean of the β_j's and n is the sample size.

After we implement Gibbs sampling to update β_0 and σ_β^2, we return to Metropolis-Hastings sampling for updating κ. The beta likelihood function comes from the

same beta distribution used as the prior for ρ_j in (2.4):

$$\begin{aligned} L(\kappa) &= \prod_{j=1}^{n} \frac{\Gamma(\kappa\omega + \kappa(1-\omega)}{\Gamma(\kappa\omega)\Gamma(\kappa(1-\omega))} \left(\frac{1-\rho_j}{2}\right)^{\kappa\omega-1} \left(1 - \frac{1-\rho_j}{2}\right)^{\kappa(1-\omega)-1} \\ &= \prod_{j=1}^{n} \frac{\Gamma(\kappa)}{\Gamma(\kappa\omega)\Gamma(\kappa(1-\omega))} \left(\frac{1-\rho_j}{2}\right)^{\kappa\omega-1} \left(\frac{1+\rho_j}{2}\right)^{\kappa(1-\omega)-1} \end{aligned} \tag{2.5}$$

whose natural log is:

$$\begin{aligned} \ln(L(\kappa)) &= \sum_{j=1}^{n} \left(\frac{\Gamma(\kappa)}{\Gamma(\kappa\omega)\Gamma(\kappa(1-\omega))} + (\kappa\omega - 1) \ln\left(\frac{1-\rho_j}{2}\right)\right) \\ &\quad + \sum_{j=1}^{n} \left((\kappa(1-\omega) - 1) \ln\left(\frac{1+\rho_j}{2}\right)\right). \end{aligned} \tag{2.6}$$

We specify a vague gamma prior for κ:

$$\begin{aligned} \pi(\kappa) &= \left(\frac{.1^{.1}}{\Gamma(.1)}\right) \kappa^{.1-1} e^{-.1\kappa} \\ &\propto \kappa^{.1-1} e^{-.1\kappa}. \end{aligned}$$

By taking the natural log, we get:

$$\ln(\pi(\kappa)) = -.9\ln(\kappa) - .1\kappa.$$

We then calculate an acceptance probability α for a new value κ^*

$$\ln(\alpha) = \sum_{j=1}^{n} \left(\frac{\Gamma(\kappa^*)}{\Gamma(\kappa^*\omega)\Gamma(\kappa^*(1-\omega))} + (\kappa^*\omega - 1) \ln\left(\frac{1-\rho_j}{2}\right) + (\kappa^*(1-\omega) - 1) \ln\left(\frac{1+\rho_j}{2}\right)\right)$$

$$-\sum_{j=1}^{n}\left(\frac{\Gamma(\kappa^c)}{\Gamma(\kappa^c\omega)\Gamma(\kappa^c(1-\omega))}+(\kappa\omega-1)\ln\left(\frac{1-\rho_j}{2}\right)+(\kappa(1-\omega)-1)\ln\left(\frac{1+\rho_j}{2}\right)\right)$$
$$-.9\kappa^*-.1\kappa^*+.9\kappa^c+.1\kappa^c.$$

The posterior for ω uses the same likelihood as κ (2.6), however, the prior for ω is chosen to be uniform (noninformative). Since this prior is a constant, it is proportional to 1. Thus, we do not need to concern ourselves with it in the calculation of the acceptance probability. Thus, we only need to calculate the difference of $\ln(L(\omega^*))$ and $\ln(L(\omega^c))$:

$$\begin{aligned}\ln(\alpha) &= \sum_{j=1}^{n}\left(\frac{1}{\Gamma(\kappa\omega^*)\Gamma(\kappa(1-\omega^*))}+(\kappa\omega^*-1)\ln\left(\frac{1-\rho_j}{2}\right)+(\kappa(1-\omega^*)-1)\ln\left(\frac{1+\rho_j}{2}\right)\right)\\ &\quad-\sum_{j=1}^{n}\left(\frac{1}{\Gamma(\kappa\omega^c)\Gamma(\kappa(1-\omega^c))}+(\kappa\omega^c-1)\ln\left(\frac{1-\rho_j}{2}\right)+(\kappa(1-\omega^c)-1)\ln\left(\frac{1+\rho_j}{2}\right)\right),\end{aligned}$$

where n represents the number of subjects.

Lastly, we update the hyperparameters a and b which both have inverse-gamma likelihoods and uniform priors. Again, we only need to write the likelihoods for the Metropolis-Hastings algorithm, as their priors are uniform. The likelihood for a is

$$\begin{aligned}L(a) &= \prod_{j=1}^{n}\left(\frac{b^a}{\Gamma(a)}\right)(\sigma_j^2)^{-(a+1)}e^{\frac{-b}{\sigma_j^2}}\\ &\propto \prod_{j=1}^{n}\left(\frac{b^a}{\Gamma(a)}\right)(\sigma_j^2)^{-(a+1)}.\end{aligned}$$

Taking the natural log yields:

$$\ln(L(a)) = \sum_{j=1}^{n}\left(a\ln(b)-\ln(\Gamma(a))+(-(a+1)\ln(\sigma_j^2)\right).$$

Thus, the log acceptance probability for a^* is

$$\begin{aligned}\ln(\alpha) &= \sum_{j=1}^{n}\left(a^*\ln(b)-\ln(\Gamma(a^*))+(-(a^*+1)\ln(\sigma_j^2)\right)\\ &\quad -\sum_{j=1}^{n}\left(a^c\ln(b)-\ln(\Gamma(a^c))+(-(a^c+1)\ln(\sigma_j^2)\right).\end{aligned}$$

The likelihood and log-likelihood for b is

$$\begin{aligned}L(b) &= \prod_{j=1}^{n}\left(\frac{b^a}{\Gamma(a)}\right)(\sigma_j^2)^{-(a+1)}e^{\frac{-b}{\sigma_j^2}}\\ &\propto \prod_{j=1}^{n} b^a e^{\frac{-b}{\sigma_j^2}}\\ \ln(L(b)) &= \sum_{j=1}^{n}\left(a\ln(b)-\frac{b}{\sigma_j^2}\right).\end{aligned}$$

Lastly, the log acceptance probability for b^*

$$\ln(\alpha) = \sum_{j=1}^{n}\left(a\ln(b^*)-\frac{b^*}{\sigma_j^2}\right)-\sum_{j=1}^{n}\left(a\ln(b^c)-\frac{b^c}{\sigma_j^2}\right).$$

2.3 Calculation of Metropolis-Hastings Correction Factors

Whenever there is a restriction placed on a 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 α. The parameters that have such restrictions in our model are identified in Table 2.1. It should be noted here that these restrictions are natural due to the corresponding probability distributions for each parameter (with the exception of μ_{0j}, which was capped at 40 to prevent

sampling unrealistic values). The Hastings ratios are then utilized through multiplication with the Metropolis acceptance probability α. This product represents the adjusted acceptance probability and prevents the sampling method from being biased against parameter values near a boundary. Table 2.1 also shows the natural log of the Hastings ratios for all necessary parameters:

Parameter	**Restriction**	$\ln($**Hastings Ratio**$)$
μ_{0j}	$\mu_{0j} < \ln(40)$	$\ln(min(2k, k + \ln(40) - \mu_{0j}^{c}))$ $- \ln(min(2 * k, k + \ln(40) - \mu_{0j}^{*}))$
ρ_j	$-1 \leq \rho_j \leq 1$	$\ln(min(2k, k + (\rho_j^c + 1), k + (1 - \rho_j^c))$ $- \ln(min(2k, k + (\rho_j^c + 1), k + (1 - \rho_j^*))$
σ_j^2	$\sigma_j^2 \geq 0$	$\ln(min(2k, k + \sigma_j^{2c}))$ $- \ln(min(2k, k + \sigma_j^{2*}))$
κ	$\kappa > 0$	$\ln(min(2k, k + \kappa^c))$ $- \ln(min(2k, k + \kappa^*))$
ω	$0 < \omega < 1$	$\ln(min(2k, k + \omega^c, k + (1 - \omega^c))$ $- \ln(min(2k, k + \omega^*, k + (1 - \omega^*))$
a	$0 < a < 50$	$\ln(min(2k, k + a^c, k + 50 - a^c))$ $- \ln(min(2k, k + a^*, k + 50 - a^*))$
b	$0 < b < 30$	$\ln(min(2k, k + b^c, k + 30 - b^c))$ $- \ln(min(2k, k + b^*, k + 30 - b^*))$

Table 2.1: Restricted Parameters and Corresponding Natural Log Hastings Ratios.

Chapter 3

Discussion

Having applied our model separately to the three experimental groups, we compare the posterior distributions of β_0 for each application to gauge overall treatment effectiveness. Shown in Figure 3.1 are boxplots of posterior β_0 samples for each application. It is evident from these boxplots that the education group's overall mean is slightly higher than that of the placebo groups. It is especially interesting to note that from the boxplots, we can see that the median β_0 value for the treatment group is significantly lower than that of the other two groups. Through simulation, we found that $\mathrm{P}(\beta_0^{Tr} < \beta_0^{Pl}) \approx .95$. Also, when an analysis of variance was conducted to test for equality of the means of the three β_0 distributions, a p-value of 0.00 was found. We therefore conclude that the treatment is significant in lowering the overall mean hot flush frequency as compared to the placebo and education groups.

A nice feature of this model is its ability to describe in detail the mean hot flush frequency on an individual profile-by-profile basis. Figures 3.2 and 3.3 show two individual profiles from the treatment group accompanied by boxplots of posterior λ_{ij} samples. The left plot in these two figures is the hot flush frequency profile

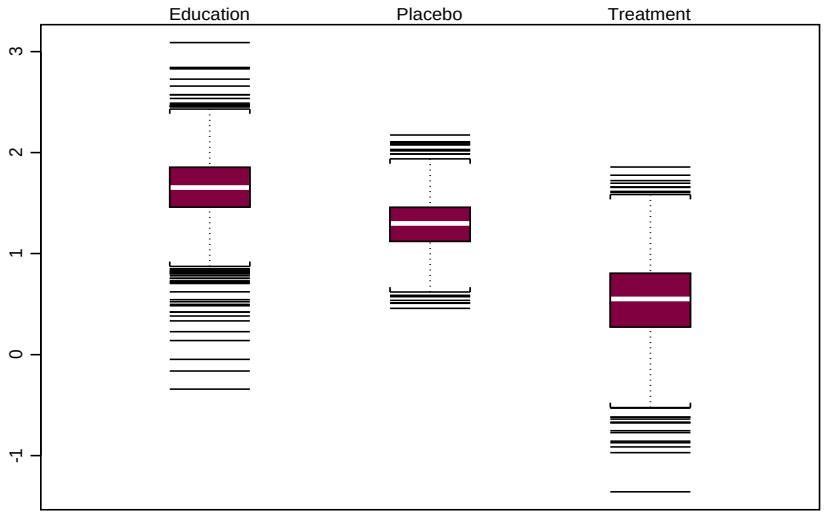

Figure 3.1: Boxplots of β_0.

for each of the two individuals; on the right are boxplots of the exponentiated μ_{ij}'s from the Metropolis-Hastings sampling (there are 1250 μ_{ij} samples used for each individual boxplot). As you can see, the MCMC did well to match the hot flush patterns for these individuals. Figures 3.4 and 3.5 show two individual profiles from the placebo group (left) accompanied by corresponding boxplots of posterior λ_{ij} samples (right). In the same way as for the treatment group, the MCMC sampling was again accurate in matching the hot flush patterns. In similar fashion figures 3.6 and 3.7 show two individual profiles from the education group (left) accompanied by corresponding boxplots of posterior λ_{ij} samples (right). We remind the reader that each μ_{ij} represents the mean hot flush frequency for 2 days. This helps explain why in situations like those displayed in Figure 3.4 the μ_{ij} boxplots respond gently to sudden, acute spikes in hot flush frequency.

Future research includes implementing an expanded version of this model that

can account for covariates in addition to hot flush frequency. Other discrete daily measurements, such as loss of concentration or mood swings, in addition to continuous pre- and post-study cortisol level measurements can be incorporated to help better determine the effectiveness of acupuncture as a treatment for menopausal symptoms.

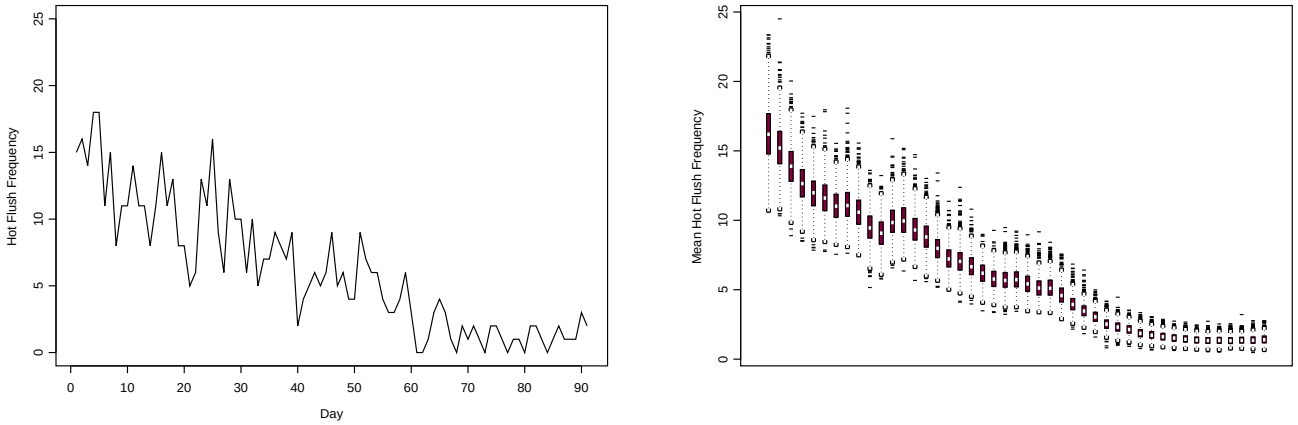

Figure 3.2: Profile 3 in Treatment Group (left), and posterior distributions of λ_{ij}'s (right).

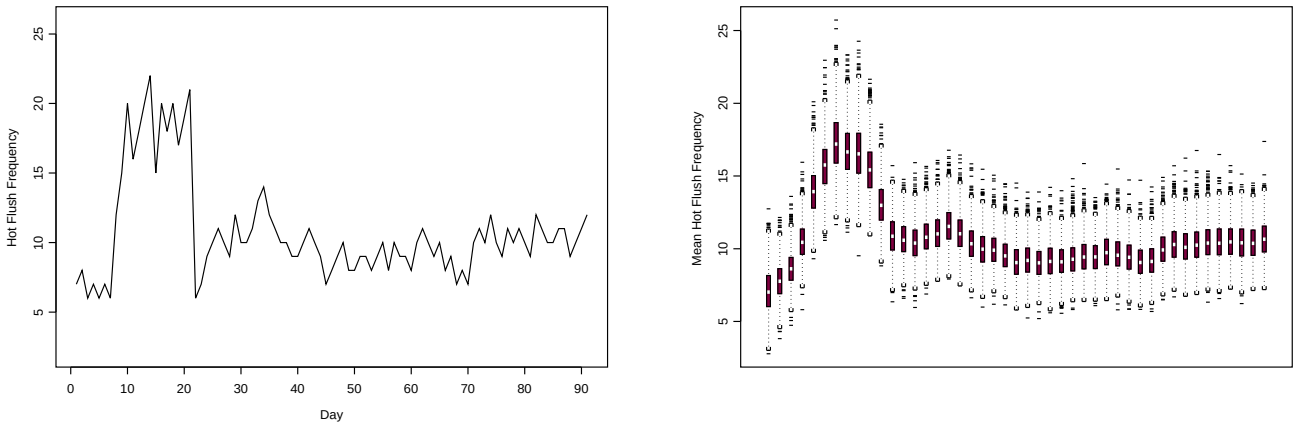

Figure 3.3: Profile 14 in Treatment Group (left), and posterior distributions of λ_{ij}'s (right).

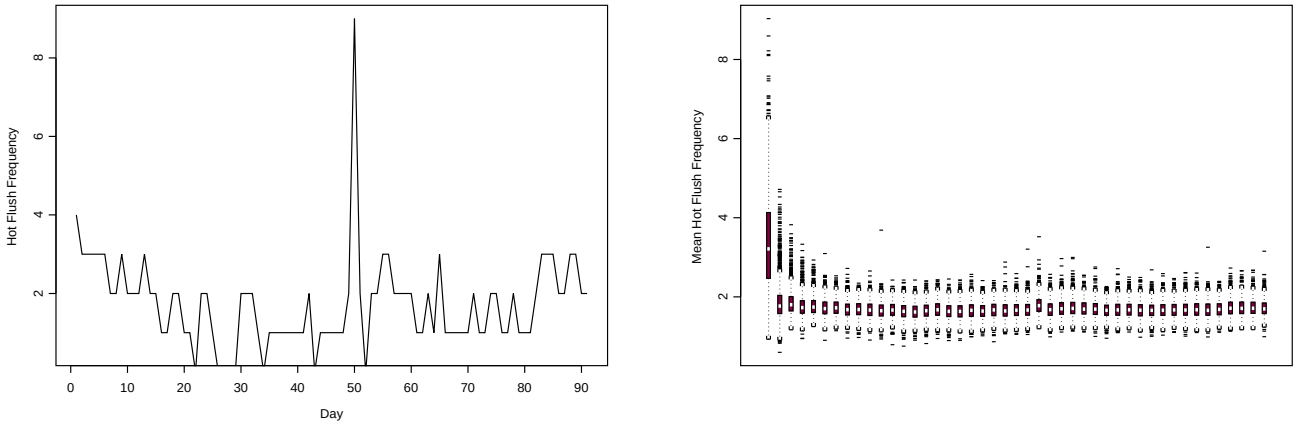

Figure 3.4: Profile 6 in Placebo Group (left), and posterior distributions of λ_{ij}'s (right).

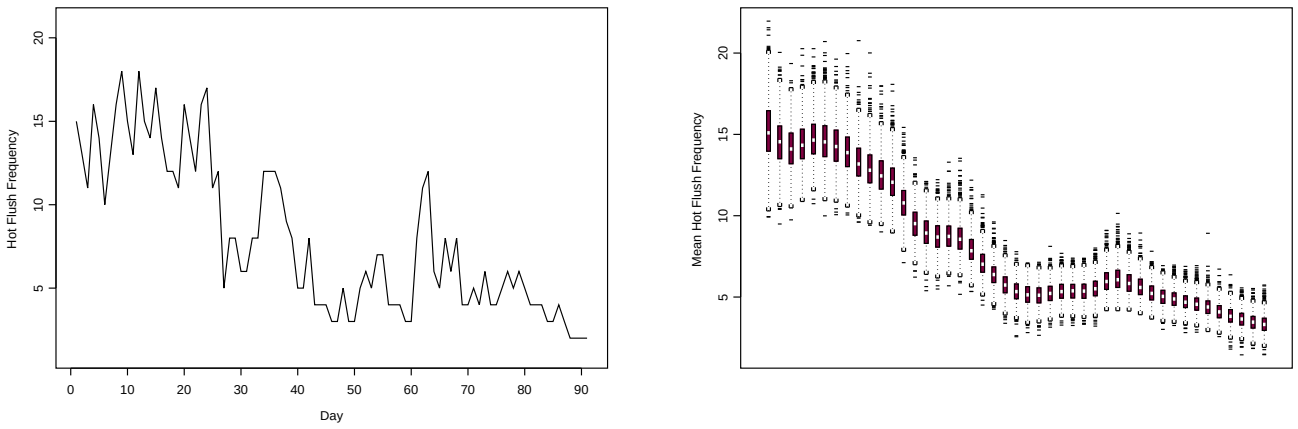

Figure 3.5: Profile 15 in Placebo Group (left), and posterior distributions of λ_{ij}'s (right).

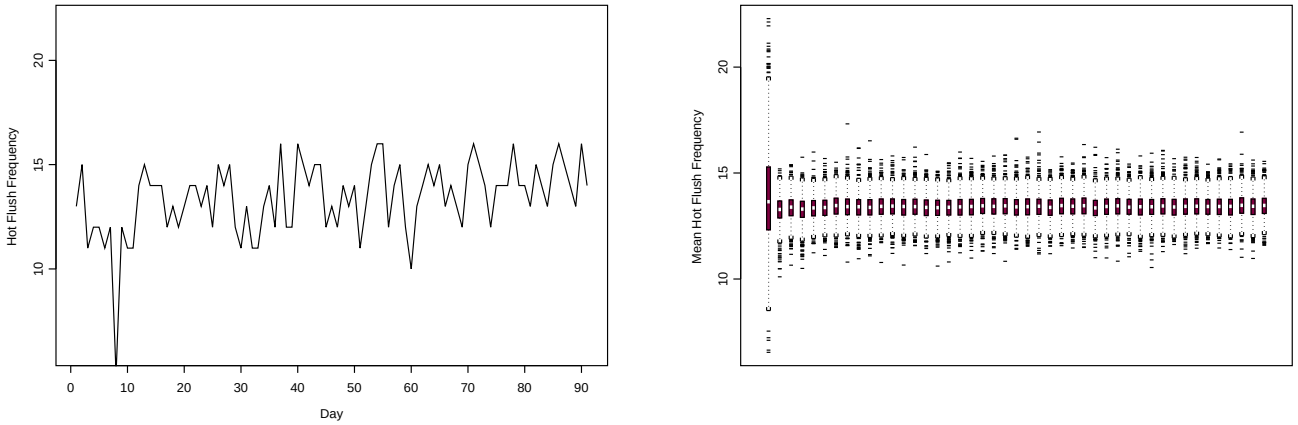

Figure 3.6: Profile 2 in Education Group (left), and posterior distributions of λ_{ij}'s (right).

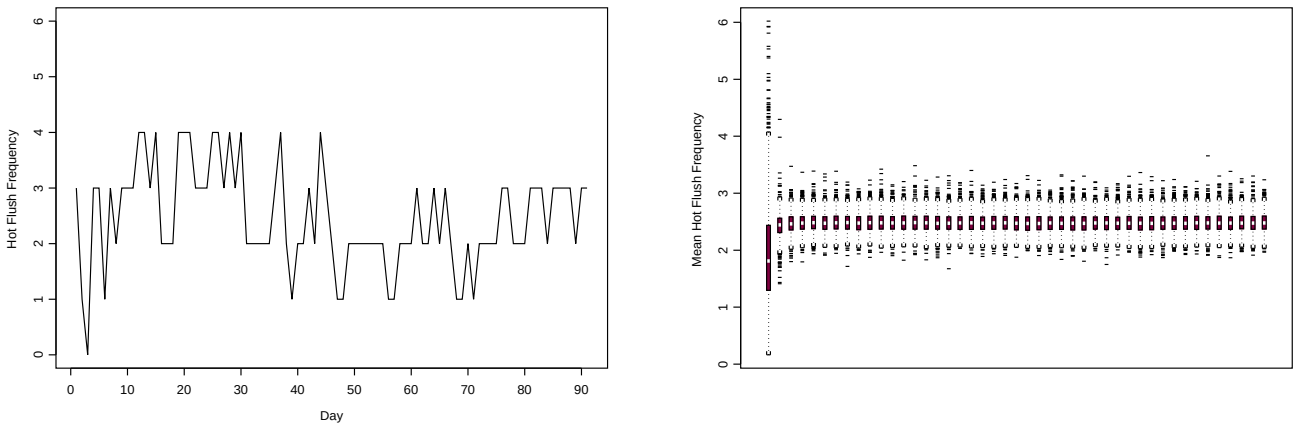

Figure 3.7: Profile 4 in Education Group (left), and posterior distributions of λ_{ij}'s (right).

References

Borgesi, J. 2004. *A Piecewise Linear Generalized Poisson Regression Approach to Modeling Longitudinal Frequency Data.* Unpublished masters thesis, Duquesne University, Pittsburgh, PA, USA.

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

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

Kern, J. and S.M. Cohen. 2005. Menopausal symptom relief with acupuncture: modeling longitudinal frequency data. Vol 34, 3: *Communications in Statistics: Simulation and Computation.*