INTRODUCTION

In almost every field of study, studying the distribution of a random curve can give needed information. In the field of in vitro fertilization, particular interest is given to modeling the temperature curves of women across their menstrual cycle. Fitting the curve can lead to categorizing cycles as healthy or unhealthy and can help identify the most fertile days of a cycle.

THE HIERARCHICAL MODEL

First, we will define terms necessary to describing the model. A prior distribution of an unknown quantity X is the probability distribution that would express one’s certainty about X before the data is taken into account. Similarly, a posterior distribution of an unknown quantity X is the probability distribution of that quantity conditional on the observed data. A hyperparameter is parameter of a prior distribution that is not treated as random. For example, consider a random variable X that came from a \( N_2(\mu, \Sigma) \) distribution, where \( \Sigma \) is known and we can use a \( N_2(\mu_0, \Sigma_0) \) distribution to model \( \mu \). Then \( \mu \) is a parameter, \( N_2(\mu_0, \Sigma_0) \) is a prior distribution, and \( \mu_0 \) and \( \Sigma_0 \) are hyperparameters.
The General Model

Let \( i \in \{1, 2, \ldots, n\} \) refer to woman \( i, j \in \{1, 2, \ldots, n\} \) refer to cycle \( j \) of woman \( i \), and \( t \in \{1, 2, \ldots, n_t\} \) refer to day \( t \) in cycle \( j \) of woman \( i \). These indices will retain these definitions throughout the remainder of the paper. We can establish the following model for bbt curves \([3]\):

\[
y_{ij}(t) = \eta_{ij}(t) + \epsilon_{ij}(t),
\]

where \( \epsilon_{ij} \) is the measurement error and \( \eta_{ij} \) is a smooth function from \( \{1, 2, \ldots, n_t\} \) to \( \mathbb{R} \). Since cycles of the same woman are expected to be correlated, we can establish a prior distribution for the distribution of functions for woman \( i \), \( \eta_{ij} \), call it \( G_i \). However, healthy menstrual cycles will follow a similar pattern among women, so we can establish a prior distribution for the collection of distributions for the different women \( \eta_i \), call it \( P \). Thus if we assume our error is normally distributed, we have the following model \([3]\):

\[
y_{ij}(t) = \eta_{ij}(t) + \epsilon_{ij}(t), \quad \epsilon_{ij}(t) \sim N(0, \sigma^2), \quad \eta_{ij} \sim G_i, \quad \mathcal{G} = \{G_i\}_{i=1}^{n} \sim P.
\]

Basis representation

Now that we have our general model, there are a number of different representations that could be used for \( \eta_{ij} \). A common strategy is to consider a basis representation \([3]\):

\[
\eta_{ij}(t) = \sum_{k=1}^{d} \theta_{ij}(t) b_k(t),
\]

where \( \{b_k(t)\}_{k=1}^{d} \) are pre-specified basis functions and \( \theta_{ij} \sim \mathcal{G} = \{\theta_{ij}\}_{i=1}^{n} \) are cycle-specific basis coefficients. Since \( \{b_k(t)\}_{k=1}^{d} \) is pre-specified, our major objective is to model the \( \theta_{ij} \)s. The approach we use is to give the \( \theta_{ij} \)s a hierarchical normal model \([3]\):

\[
\theta_{ij} \sim N_k \left( \alpha_j, \Omega_j^{-1} \right), \quad \alpha_j \sim N_k \left( \alpha, \Omega_j^{-1} \right).
\]

Where \( N_k(\mu, \Sigma) \) is the multivariate normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \). The woman-specific basis coefficients are given by \( \alpha \), \( \alpha \) gives the global mean basis coefficients, and \( \Omega \) and \( \Omega_j \) are the precision matrices for the within and between woman variability respectively. So to complete the model, we need to specify the basis functions \( \{b_k(t)\}_{k=1}^{d} \) and establish the components of \( \gamma \). We’ll do this in the next section when we consider a parametric hierarchical structure.

Parametric Hierarchical Model

As we discussed in Section 1, bbt levels of a healthy cycle follow a biphasic pattern where the bbt levels start at the low plateau in the follicular phase then quickly rise during ovulation to the high plateau in the luteal phase. Thus we can model \( \eta_{ij}(t) \) as follows \([3]\):
where $\theta_{ij} = \begin{pmatrix} \theta_{ij1} \\ \theta_{ij2} \end{pmatrix}$, where $\theta_{ij1}$ and $\theta_{ij2}$ indicate the temperature during the follicular phase and the increase in bbt during ovulation respectively for cycle $j$ of woman $i$. We can further define $k_{ij}$ to be the last day of the follicular phase and $r_{ij}$ to be the number of days the bbt rises during ovulation (Figure 2). Thus, if we consider $Y_{ij}$ as $Y_{ij}$ a vector of $n_{ij}$ bbt values, we can represent our hierarchical model in matrix form [3]:

$$Y_{ij} = X_{ij} \theta_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, I_{n_{ij}}, \sigma^2)$$

$$X_{ij} = \begin{bmatrix} 1 \cdots 1 \\ 0 \cdots 0 \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \\ 1/2 \cdots 1/2 \\ \vdots \cdots \vdots \\ 1/2 \cdots 1/2 \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \end{bmatrix}^T$$

Thus $X_{ij}(t)$, the $t^{th}$ row of $X$, corresponds to the $t^{th}$ day of cycle $j$ of woman $i$. So we have defined our pre-specified basis function $(b_1(t), b_2(t)) = X_{ij}(t)$ and our cycle-specific basis coefficients $\theta_{ij} = \begin{pmatrix} \theta_{ij1} \\ \theta_{ij2} \end{pmatrix}$.

**Bayesian Specification**

Now we have within ($\Omega$) and between ($\Omega'$) woman variability, the last day of the follicular phase ($k_{ij}$), the number of days during ovulation ($r_{ij}$), the global mean ($\alpha$), and measurement error variance ($\sigma^2$) left to model. We’ll use the following equations to model the within and between woman variability:

$$\Omega = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \quad \Omega' = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$$

We complete the Bayesian specification with the following priors [3]:

$$k_{ij} \sim U(\bar{m}_j, m_j + 20)$$

$$r_{ij} \sim U(1, 15)$$

$$\alpha \sim N(\alpha_0, \Sigma_\alpha)$$

$$\sigma^2 \sim G(\epsilon, d)$$

$$\omega_h \sim G(a_h, b_h) \text{ for } h = 1, 2$$

$$\lambda_h \sim G(a_{h1}, b_{h1}) \text{ for } h = 1, 2$$

where $G(\epsilon, d)$ is the gamma distribution with shape parameter $\epsilon$ and inverse scale parameter $d$, $U(a, b)$ is the discrete uniform distribution between $a$ and $b$, and $\bar{m}_j$ is the first day after menstruation. Thus, our pre-specified hyperparameters are $\alpha_0, \Sigma_\alpha, \epsilon, d, a_h, b_h, a_{h1},$ and $b_{h1}$ (for $h = 1, 2$). A summary of the parameters, hyperparameters, their descriptions, and distributions is given in the appendix.

**GIBBS SAMPLING**

Our goal is to find the parameters of our model given a set of observations. Since finding the joint distribution is a complicated process, we need to use a different method. A common method to do this is to use a Monte Carlo Markov Chain as an iterative procedure. The specific type of Monte Carlo Markov Chain that we’ll use is called a Gibbs sampler.
Definition

Suppose we have random variables $X_1, X_2, \ldots, X_n$.

1. Initialization: Let $x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}$ be given some initial value.

2. Iteration: For $1 \leq i \leq K$, sample $x_j^{(i)}$ from the conditional distribution of $X_j$ given $X_1^{(i)}, \ldots, X_{j-1}^{(i)}$ and $X_{j+1}^{(i-1)}, \ldots, X_n^{(i-1)}$ for $1 \leq j \leq n$.

This Gibbs sampling procedure is a Markov chain, where the stationary distribution is joint distribution of $X_1, X_2, \ldots, X_n$. Thus, for large $k$, the sample $x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}$ approximates the joint distribution of $X_1, X_2, \ldots, X_n$. A blocked Gibbs sampler uses the same steps, except it groups some of the random variables together so that the variables in a block are sampled from their joint distribution conditioned on all other variables.

Gibbs Sampling Algorithm

Here is the blocked Gibbs sampling algorithm for our model after we’ve given the parameters some initial value [1]:

1. Cycle-specific coefficients: $\theta_j^{(i)} | y_{ij} \ldots \sim N_2(h_1^{(i)}, V_1^{(i)})$

$$
V_1^{(i)} = (\Omega + \sigma^2 X_j X_j) \quad h_1^{(i)} = V_1^{(i)} (\Omega + \sigma^2 X_j y_{ij})
$$

2. Women-specific means: $\alpha_i | y_{ij} \ldots \sim N_2(h_2^{(i)}, V_2^{(i)})$

$$
V_2^{(i)} = (\Omega + n_i \Omega) \quad h_2^{(i)} = V_2^{(i)} (\Omega + \sum h_{ij} \theta_j y_{ij})
$$

3. Global mean: $\alpha | y_{ij} \ldots \sim N_2(h_3^{(i)}, V_3^{(i)})$

$$
V_3^{(i)} = (\Sigma + n_i \Sigma) \quad h_3^{(i)} = V_3^{(i)} (\sum \alpha_i + \Omega \sum h_{ij} \theta_j y_{ij})
$$

4. Components of $\Omega$ ( $h = 1, 2$): $\omega_{ij} | y_{ij} \ldots \sim G(p_1, q_1)$

$$
p_1 = a_{h} + \frac{1}{2} b_{h} = b_{h} + \frac{1}{2} \sum_{i=1}^{n} \frac{q_{ij}}{\theta_{ij} - \alpha_{ih}}
$$

5. Components of $\Omega$ ( $h = 1, 2$): $\omega_{ij} | y_{ij} \ldots \sim G(p_2, q_2)$

$$
p_2 = a_{ih} + \frac{n}{2} \quad q_2 = h_{ih} + \frac{1}{2} \sum_{i=1}^{n} \left( \alpha_{ih} - \alpha_{h} \right)^2
$$

6. Error variance: $\sigma^2 | y_{ij} \ldots \sim G(p_3, q_3)$

$$
p_3 = c + \frac{n}{2} \sum_{i=1}^{n} \frac{q_{ij}}{\theta_{ij} y_{ij}} \quad q_3 = d + \sum_{i=1}^{n} \sum_{j=1}^{n} \left( y_{ij} (t) - y_{ij} (r) \theta_{ij} \right)^2
$$

7. Last day of follicular phase: Use the change-point stopping rule for $k_{ij}$ (see Section 4).

8. Ovulation period: Use the change-point stopping rule for $r_{ij}$ (see Section 4).

So if we use the blocked Gibbs sampling steps to update the unknown parameters from their conditional distributions, then after a large number of iterations, we’ll be sampling from the joint distribution of all the unknown parameters.

CHANGE-POINT PROBLEM

Now we need to consider $k_{ij}$ and $r_{ij}$ since these are specific points (called change-points) in the bbt curve where the shape of the curve changes. The process for finding a change-point $P$ is intuitive. Given a set of observations $y_m = \{y_{1}, y_{2}, \ldots, y_{m}\}$ we’ll find the probability that $P \leq m$ given $y_m$. We want this probability to be higher than some high threshold $Q$. Thus we have a "stopping rule" [4]:

Terminate at $m$ if and only if $Pr(P \leq m | y_m) > Q$.

Starting with $m = 1$, the first time that the probability is greater than $Q$ is our change-point $P$.

Finding $k_{ij}$
First we wish to find $k_{ij}$ assuming that the other parameters are known. Let $X_k$ be the first $m$ rows of the matrix $X_{ij}$ established above but with $k_{ij} = k$. Thus

$$X_k = \begin{bmatrix} \omega_{ij,k} \end{bmatrix},$$

$\omega_{ij,k}$ be the prior distribution for $k_{ij} = k$, $\pi(.)$ be the density function of $\theta_{ij}$, and $f(.)$ be the density function of the data $y_{ij,m}$. We now need to find the posterior distribution of $k_{ij}$ in order to find $Pr(k \leq m | y_m)$ for our stopping rule.

Let $L_k = f(i_j, m | k_{ij} = k)$ be the change-point likelihood. Then implementing Bayes’ theorem, we get \(^(4)\)

$$L_k = \frac{f(i_j, m | \theta_{ij} = k) \pi(\theta_{ij} = k) | y_{ij,m} = k}}{\pi(\theta_{ij}) | y_{ij,m} = k}.$$ 

Since $L_k$ is not a function of $\theta_{ij}$ we may replace $\theta_{ij}$ with 0. Also $f(y_{ij,m} | 0, k)$ does not depend on $k$, so $L_k$ is proportional in $k$ to $W_k = (\pi(0 | y_{ij,m}, k))^{-1}$.

The posterior distribution of $\theta_{ij}$ is $N_2(\mu_k, V_k)$ where $V_k = (\sigma^2 X_k X_k' + \Omega)^{-1}$.

$$\mu_k = V_k (\sigma^2 X_k y_{ij,m} + \Omega \alpha).$$

Note that this is the same posterior distribution found for $\theta_{ij}$ found in Section 3, but with the altered matrix $X_k$. Using this posterior distribution for $\theta_{ij}$, we find that

$$W_k = (det_k)^{-2} \exp\left(\frac{1}{2} \mu_k' V_k \mu_k\right).$$

Since we need to find $Pr(k \leq m | y_{ij,m})$, we’ll first find $Pr(k = k | y_{ij,m})$. From applying Bayes’ theorem to $Pr(k = k | y_{ij,m})$, we get

$$Pr(k = k | y_{ij,m}) = L_k \pi_k k \sum_{s=1}^{n_y} L_s \pi_s.$$ 

Since $L_k$ is proportional in $k$ to $W_k$, then we can substitute $W_k$ in for $L_k$, resulting in

$$Pr(k = k | y_{ij,m}) = W_k \pi_k k \sum_{s=1}^{n_y} W_s \pi_s^{-1}.$$ 

These probabilities are easy to compute as the data becomes available. When $k \geq m$,

$$X_k = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

So $X_k$ does not depend on $k$ (and hence $W_k = W_m$). Thus, our stopping rule is to terminate at $m$ if and only if

$$(1/Q) \sum_{k=1}^{n_y} W_k \pi_k > QW_m \sum_{k=m+1}^{n_y} \pi_k.$$ 

Once we hit our stopping rule \(^{(4)}\), let $k = m$.

Finding $r_{ij}$

We can apply the same approach to finding $r_{ij}$ by using a transformation to examine the data “backwards”, i.e. consider the last day first, and proceed through the cycle in reverse order. So our new reversed data is $y_y^{*} = (y_1(n_y), y_1(n_y - 1), \ldots, y_1(1))$.

Let $A = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$ be our transformation matrix. Then

$$\theta_y = A \theta_y = \begin{bmatrix} \theta_{y1} + \theta_{y2} \\ -\theta_{y2} \end{bmatrix}$$

$$y_y^{*}(t) = \begin{cases} \theta_{y1} + \theta_{y2} & 1 \leq t \leq n_y - r \\ \theta_{y1} + \theta_{y2} - \theta_{y3} \left( \frac{t-n_y+r}{n_y-r} \right) & n_y - r \leq t \leq n_y - k_y \\ \theta_{y1} & n_y - k_y \leq t \leq n_y \end{cases}$$

Thus if we let $X^*_y$ correspond to the first $m$ rows of the above system, then
\[ y_{ij,m}^* = X_{ij}^* \theta_j^* + e_{ij,m}^* \]

This transformation lets us use the same process that we used to find \( k_j \). Let \( \{ \pi_j \} \) be the prior distribution for \( r \) (the second change-point). The new posterior distribution for \( \theta_j^* \) is \( N_r(\mu_r, V_r) \), where

\[ V_r = \left( \sigma^{-2} \left( X_r' X_r + A \Omega A' \right) \right)^{-1} \]

\[ \mu_r = V_r \left( \sigma^{-2} \left( X_r' y_{r,a} + A \Omega A' \alpha_i \right) \right). \]

Note that since \( \theta_j^* = A \theta_j \Omega_j \sim N_r(A \alpha_i, A \Omega A') \) thus giving a similar posterior distribution found in Section 3 for \( \theta_j \). Using this posterior distribution for \( \theta_j^* \) we find that

\[ W_r = (\det V_r)^{-1} \exp \left( \frac{1}{2} \mu_r' V_r \mu_r \right). \]

Thus, our stopping rule is to terminate at \( m \) if and only if

\[ (1 = Q) \sum_{r=1}^{m} W_r \pi_r > QW_m \sum_{r=m+1}^{n} \pi_r. \]

Once we hit our stopping rule, \( r = n_j - m \) (since we transformed our data). Since \( r_j \) is the number of days between \( k_j \) and \( r \), then \( r_j = r - k_j \).

### SIMULATION

#### Data simulation

Now that the method has been described, we’ll introduce a simulation in order to illustrate the method. The simulated data consisted of 20 women, each with 10 cycles, so we had a total of 200 cycles. Each cycle had a length of 30 days [5]. We arbitrarily assigned the global mean and covariance matrices to be

\[ \alpha = (36, 5, 0.4) \quad \Omega = \Omega = \begin{bmatrix} 0.25 & 0.01 \\ 0 & 0.01 \end{bmatrix}. \]

Each woman specific mean was generated as per the model discussed in Section 2, so \( \alpha_i \sim N_i(\alpha, \Omega) \). We then generated \( k_j \) and \( r_j \) for each cycle according to \( k_j \sim U(1, 14) \) and \( r_j \sim U(1, 15) \). Finally, we generated the cycle specific coefficients according to \( \theta_j \sim N_2(\alpha_i, \Omega) \) and the measurement error terms according to \( \varepsilon_{ij} \sim N(0, \sigma^2 I) \) where \( \sigma^2 = 0.01 \). Now that we generated all the parameters needed for our model, we generated the bbt values according to

\[ y_i = X_i \theta_i + \varepsilon_i. \]

#### Initialization

In order to run the algorithm, we need to specify the hyperparameters (our initial guesses). In order to account for error, we intentionally perturb the hyperparameters away from the actual parameters established in Section Data simulation So we set our hyperparameters to be

\[ \alpha_s = \begin{bmatrix} 36 \\ 1 \end{bmatrix} \quad \Sigma_s = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.01 \end{bmatrix} \]

\[ a_1 = a_2 = 1 \quad b_1 = b_2 = 1 \]

\[ a_1 = a_2 = 1 \quad b_1 = b_2 = 1 \]

\[ c = 1 \quad d = 1 \]

These hyperparameters are used in the Gibbs sampling steps outlined in Section 4. But before we can implement the Gibbs sampling algorithm, we need to initialize our unknowns from their prior distributions using the hyperparameters, as seen below:

\[ \alpha \sim N \left( \begin{bmatrix} 36 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.25 & 0 \\ 0 & 0.01 \end{bmatrix} \right) \]

\[ \alpha_1, \alpha_2 \sim G(1, 1) \]

\[ \alpha_1, \alpha_2 \sim G(1, 1) \]

\[ \sigma^2 \sim G(1, 1) \]

We also need to initialize the \( \alpha_i \) from their prior distributions since the first Gibbs sampling step is finding the posterior distribution of \( \theta_j \) conditional on \( \alpha_i \). So we’ll generate \( \alpha_i \) values from the prior distribution.
\[ \alpha_i = N_2(\alpha, \Omega) \]

The final step for initialization was to set the \( k_{ij} \) and \( r_{ij} \) values all to a fixed value of 7.

**Running the algorithm**

Now that initialization is completed, we can run the algorithm. We ran the algorithm for a total of 6,000 iterations with a 1,000 iteration burn-in. To make sure that the parameters were converging, we performed traceplots of a number of the variables. Figures 3 and 4 show the traceplots for \( \alpha_{11} \) and \( \alpha_{12} \) respectively.

**Tables 1 and 2** summarize the results from estimating the parameters in our model. The cycle specific parameters (\( k_{ij}, r_{ij}, \theta_{ij1}, \) and \( \theta_{ij2} \)) are summaries over all cycles, while the women specific parameters (\( \alpha_{i1} \) and \( \alpha_{i2} \)) are summaries over all women. As evidenced by the tables, the distribution of these parameters was well estimated. Figures 5 and 6 show graphs of the true and estimated distributions for two example cycles.

**Table 1.** Posterior summaries of the true cycle and woman specific parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>First Quartile</th>
<th>Median</th>
<th>Third Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_{ij} )</td>
<td>7.07</td>
<td>4.00</td>
<td>4</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>( r_{ij} )</td>
<td>8.60</td>
<td>4.60</td>
<td>4</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>( \theta_{ij1} )</td>
<td>36.49</td>
<td>0.13</td>
<td>36.40</td>
<td>36.50</td>
<td>36.58</td>
</tr>
<tr>
<td>( \theta_{ij2} )</td>
<td>0.40</td>
<td>0.13</td>
<td>0.31</td>
<td>0.41</td>
<td>0.47</td>
</tr>
<tr>
<td>( \alpha_{i1} )</td>
<td>36.49</td>
<td>0.08</td>
<td>36.43</td>
<td>36.51</td>
<td>36.55</td>
</tr>
<tr>
<td>( \alpha_{i2} )</td>
<td>0.41</td>
<td>0.09</td>
<td>0.35</td>
<td>0.38</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Table 2. Posterior summaries of the estimated cycle and woman specific parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>First Quartile</th>
<th>Median</th>
<th>Third Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>7.91</td>
<td>4.04</td>
<td>4.5</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>$r_j$</td>
<td>7.83</td>
<td>4.87</td>
<td>3</td>
<td>8</td>
<td>12.5</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>36.49</td>
<td>0.13</td>
<td>36.39</td>
<td>36.50</td>
<td>36.57</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>0.40</td>
<td>0.13</td>
<td>0.30</td>
<td>0.40</td>
<td>0.47</td>
</tr>
<tr>
<td>$\alpha_j$</td>
<td>36.50</td>
<td>0.08</td>
<td>36.44</td>
<td>36.50</td>
<td>36.54</td>
</tr>
<tr>
<td>$\beta_j$</td>
<td>0.40</td>
<td>0.10</td>
<td>0.34</td>
<td>0.39</td>
<td>0.47</td>
</tr>
</tbody>
</table>

CONCLUSION

So, we can see that the blocked Gibbs sampling algorithm was an accurate method for estimating the various parameters of our hierarchical model. Most importantly, the change-point analysis gave accurate estimates for the last day of the follicular stage and the number of days that the temperature rises during ovulation (Appendix). These two figures are critical for determining the most fertile period of the menstrual cycle. A natural extension of this topic would be to model unhealthy cycles using a nonparametric approach. Combining the parametric and nonparametric models would allow for broader analysis and more unusual data sets.

REFERENCES