Statistical Modelling HW6

Korawat Tanwisuth

October 20, 2018

1.

$$
\begin{aligned}
&p(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\
&= p(X_1 = x_1) \prod_{t=2}^{n} p(X_t = x_t | X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}) \\
&= p(X_1 = x_1) \prod_{t=2}^{n} P(X_t = x_t | X_{t-1} = x_{t-1}) \quad \text{(By Markov Assumption)} \\
&= p(X_1 = x_1) \prod_{t=2}^{n} p_{x_{t-1} x_t}
\end{aligned} \tag{1}
$$

Let N_{ij} be the number of counts from state i to state j. Rewriting the likelihood in terms of transition probability p_{ij}.

$$
\begin{aligned}
L(p_{ij}) &= p(X_1 = x_1) \prod_{i=1}^{K} \prod_{j=1}^{K} p_{ij}^{N_{ij}} \\
l(p_{ij}) &= \log(p(X_1 = x_1) \prod_{i=1}^{K} \prod_{j=1}^{K} p_{ij}^{N_{ij}}) \\
&= \log p(X_1 = x_1) + \sum_{i,j} N_{ij} \log p_{ij}
\end{aligned} \tag{2}
$$

We want to maximize $l(p_{ij})$ but we have to add the constraint that $\sum_j p_{ij} = 1$ for all i.

Thus, our objective function becomes:

$$
l(p_{ij}) = \log p(X_1 = x_1) + \sum_{i,j} N_{ij} \log p_{ij} - \sum_{i=1}^{K} \lambda_i (\sum_j p_{ij} - 1)
$$

$$
\begin{aligned}
\frac{\partial l(p_{ij})}{\partial p_{ij}} &= \frac{N_{ij}}{p_{ij}} - \lambda_i \\
\frac{\partial l(p_{ij})}{\partial p_{ij}} &= 0 \\
\rightarrow p_{ij} &= \frac{N_{ij}}{\lambda_i}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\frac{\partial l(p_{ij})}{\partial \lambda_i} &= \sum_j p_{ij} - 1 \\
\frac{\partial l(p_{ij})}{\partial \lambda_i} &= 0 \\
&\to \sum_{j=1}^{K} p_{ij} = 1 \\
&\to \sum_{j=1}^{K} \frac{N_{ij}}{\lambda_i} = 1 \\
&\to \lambda_i = \sum_{j=1}^{K} N_{ij}
\end{aligned} \tag{4}
$$

Combining the two equations above, $\hat{p}_{ij} = \dfrac{N_{ij}}{\sum_{j=1}^{K} N_{ij}}$.

$$
\hat{P} = \begin{bmatrix} 0.09696093 & 0.503617945 & 0.399421129 \\ 0.97191011 & 0.019662921 & 0.008426966 \\ 0.99641577 & 0.003584229 & 0.000000000 \end{bmatrix} \tag{5}
$$

To find stationary distribution π, we solve for the equation $\pi \hat{P} = \pi$. Equivalently, we solve for the left eigenvector of $\hat{P}$.

$$
\pi = [0.5211161 \ 0.2684766 \ 0.2104072]
$$

2. We will work with the precision instead of the variance. Denote $\tau_k = \frac{1}{\sigma_k^2}$.

Given

$\pi_j \sim \text{Dir}(1/K, \ldots, 1/K)$

$\mu_k \sim \text{Normal}(\mu_0, \sigma_0^2)$

$\sigma_k^2 \sim \text{Inv-Gam}(a_0, b_0)$

$\tau_k \sim \text{Gamma}(a_0, b_0)$

We can write joint distribution for all parameters of a Normal HMM as:

$$
\begin{aligned}
p(\rho_k, \mathbf{A}, \mu_k, \tau_k) \propto \prod_{k=1}^{K} \rho_k^{I(z_1=k)} \prod_{k=1}^{K}\prod_{j=1}^{K} \pi_{jk}^{N_{jk}} \prod_{k=1}^{K}\prod_{t:z_t=k}^{T} \sqrt{\tau_k} \exp\{-\frac{\tau_k (y_t - \mu_k)^2}{2}\} \\
\times \prod_{k=1}^{K} \rho_k^{\frac{1}{K}-1} \prod_{k=1}^{K}\prod_{j=1}^{K} \pi_{jk}^{\frac{1}{K}-1} \times \prod_{k=1}^{K} \exp\{-\frac{(\mu_k - \mu_0)^2}{2\sigma_0^2}\} \times \prod_{k=1}^{K} \tau_k^{a_0 - 1} \exp\{-b_0 \tau_k\}
\end{aligned}
$$

where π_{jk} is the jth row and kth column of the matrix $\mathbf{A}$ (the transition matrix). Here, ρ_k is the initial distribution of z_1.

Looking at the joint posterior, we can derive the full conditional as the following (See the appendix for the details):

$$p(\mu_k|-) = \text{Normal}(\frac{\sigma_0^2(\sum_{t:z_t=k} y_t) + \mu_0\sigma_k^2}{\sigma_0^2 N_k + \sigma_k^2}, \frac{\sigma_0^2\sigma_k^2}{N_k\sigma_0^2 + \sigma_k^2})$$

$$p(\sigma_k^2|-) = \text{InvGamma}(a_0 + 0.5N_k, b_0 + 0.5\sum_{t:z_t=k}(y_t - \mu_k)^2)$$

OR

$$p(\tau_k|-) = \text{Gamma}(a_0 + 0.5N_k, b_0 + 0.5\sum_{t:z_t=k}(y_t - \mu_k)^2)$$

$$p(\mathbf{A}_{j.}|-) \propto \prod_{k=1}^{K}\prod_{j=1}^{K} a_{jk}^{N_{jk}+\frac{1}{K}-1} = \text{Dir}(N_{j1} + \frac{1}{K}, N_{j2} + \frac{1}{K}, \ldots, N_{jk} + \frac{1}{K})$$

Here, $\mathbf{A}_{j.}$ denotes the j th row of the $\mathbf{A}$ matrix and N_{jk} denotes the number of transition from state j to state k.

$$p(\rho_k|-) \propto \prod_{k=1}^{K} \rho_k^{I(z_1=k)+\frac{1}{K}-1} = \text{Dir}(I(z_1 = k) + \frac{1}{K})$$

Below are the steps in the Gibb sampling algorithm: At $b = 0$, initialize $\mathbf{z}^{(0)}, \pi^{(0)}, \mathbf{A}^{(0)}, \boldsymbol{\mu}^{(0)}, \boldsymbol{\tau}^{(0)}$.

For $b = 1, 2, \ldots, B$:

1. Update μ_k

Sample $\mu_k^{(b+1)}$ from $\text{Normal}(\frac{\sigma_0^2(\sum_{t:z_t=k} y_t) + \mu_0\sigma_k^2}{\sigma_0^2 N_k + \sigma_k^2}, \frac{\sigma_0^2\sigma_k^2}{N_k\sigma_0^2 + \sigma_k^2})$

2. Update τ_k

Sample $\tau_k^{(b+1)}$ from $\text{Gamma}(a_0 + 0.5N_k^{(b)}, b_0 + 0.5\sum_{t:z_t=l}(y_t - \mu_k^{(b+1)})^2)$

3. Update $\mathbf{A}$

Compute $N_{jk}^{(b)}$ where N_{jk} denotes the number of transition from state j to state k.

Sample the j th row of $\mathbf{A}^{(b+1)}$ from $\text{Dir}(N_{j1}^{(b)} + \frac{1}{K}, N_{j2}^{(b)} + \frac{1}{K}, \ldots, N_{jk}^{(b)} + \frac{1}{K})$

4. Update $\mathbf{z}$

Here, we compute backward probabilities first $\beta_t(z_t) = p(y_{t+1}, \ldots, y_T|z_t = j)$.

First, we initialized $\beta_T(z_T) = 1$

Then for $t = T - 1, \ldots, 2$, we use the recursive relationship to update.

$\beta_t(z_t) = \sum_{j=1}^{K} p(y_{t+1}|z_{t+1} = j)p(z_{t+1} = j|z_t)\beta_{t+1}(z_{t+1})$

$(\rho_1, \rho_2, \ldots, \rho_K) \sim \text{Dir}(I(z_1 = 1) + 1/K, I(z_1 = 2) + 1/K, \ldots, I(z_1 = K) + 1/K)$

Sample z_1 with probabilities:

$p(z_1 = k|-)^{(b+1)} \propto \rho_k N(y_1|\mu_k^{(b+1)}, (\sigma_k^2)^{(b+1)})p(y_{2:T}|z_1 = k)^{(b)}$

Then, for $t = 1, 2, \ldots, T$, sample z_t with probabilities:

$p(z_t = k|z_{t-1} = j)^{(b+1)} \propto \pi_{jk}^{(b+1)} N(y_t|\mu_k^{(b+1)}, (\sigma_k^2)^{(b+1)})p(y_{t+1:T}|z_t = k)^{(b)}$

$\pi_{jk}^{(b+1)}$ is the jth row kth column of the transition matrix $\mathbf{A}^{(b+1)}$

5. Now, we compute forward probabilities to calculate the predictive density.
$\alpha_t(z_t) = p(y_1, y_2, \ldots, y_t, z_t)$

First, we initialized $\alpha_1(z_1) = p(y_1|z_1)p(z_1)$.

Then for $t = 2, \ldots, T$, we use the recursive relationship to update the message.

$\alpha_{t+1}(z_{t+1}) = p(y_{t+1}|z_{t+1}) \sum_{j=1}^{K} p(z_{t+1}|z_t = j)\alpha_t(z_t)$

(The derivation for predictive densities are done in the appendix.)

Let ϕ_T be the forward message of $z_T^{(b+1)}$.

$p(y_{t+1} = y|y_1, y_2, \ldots, y_T) = \sum_{k=1}^{K} (\phi_T \mathbf{A}^{(b+1)})_k N(y|\mu_k^{(b+1)}, (\sigma_k^2)^{(b+1)})$

where $(\phi_T \mathbf{A})_k$ denotes the k th entry of the product between the forward probabilities and the transition matrix.

Please note that we are using the most updated values of the parameters in the updating equations.

The results are shown in the next page. (Numerical results are in the appendix)

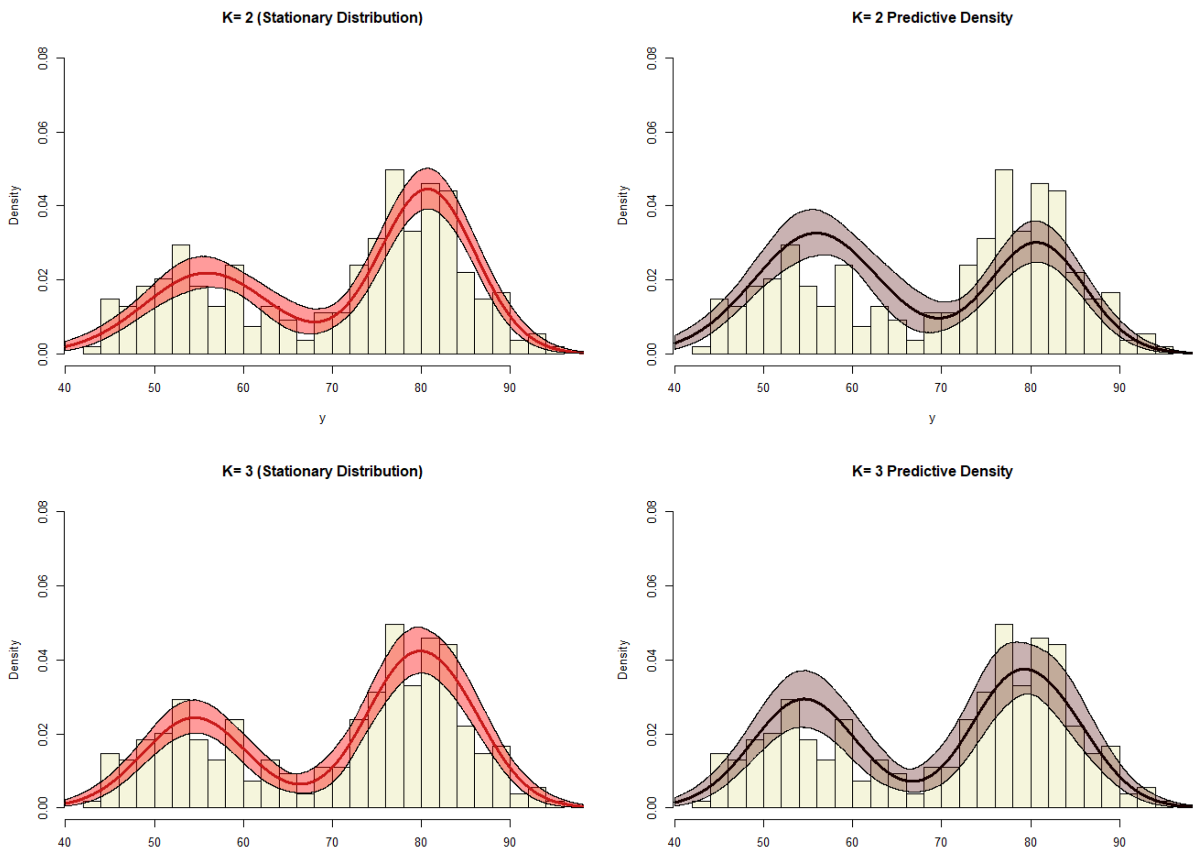

Figure 1: The left hand plot shows 90% confidence interval and the posterior mean. The right hand side shows the predictive density and 90% confidence interval.

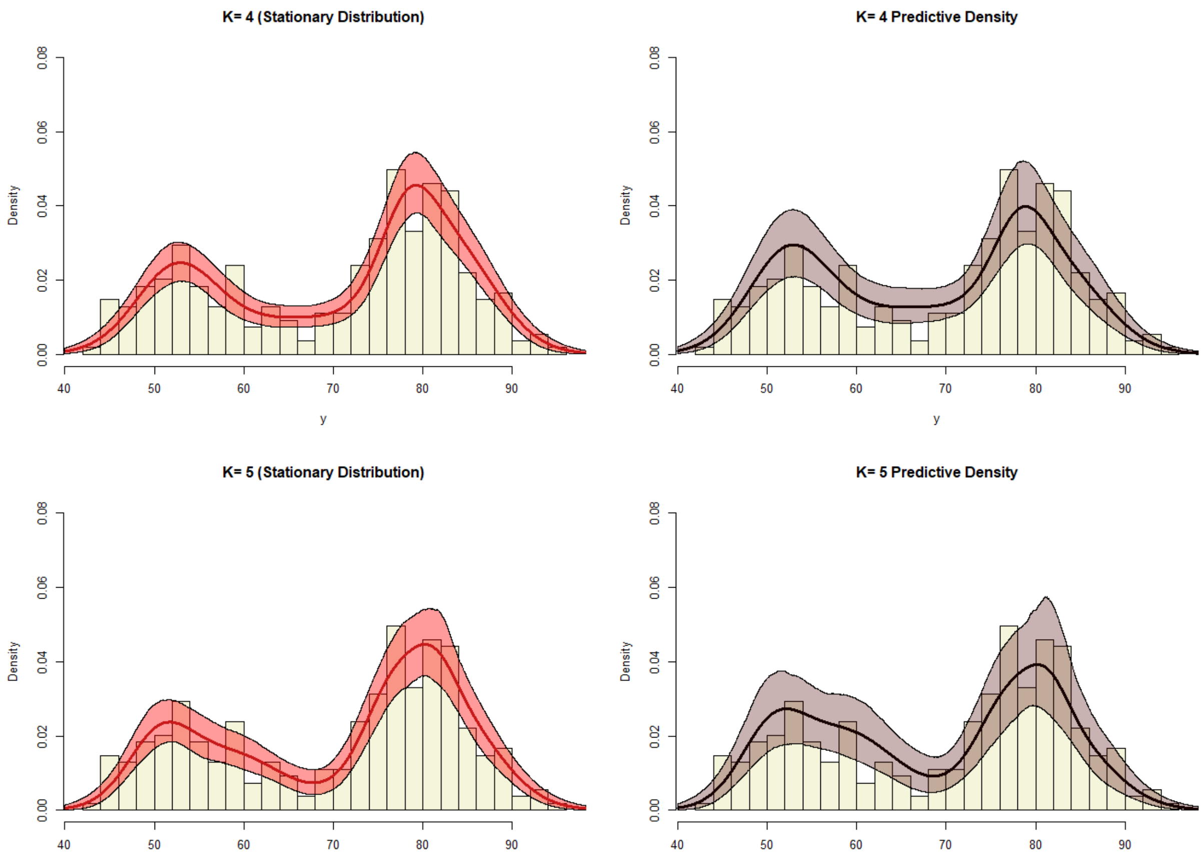

Figure 2: The left hand plot shows 90% confidence interval and the posterior mean. The right hand side shows the predictive density and 90% confidence interval.

K=2

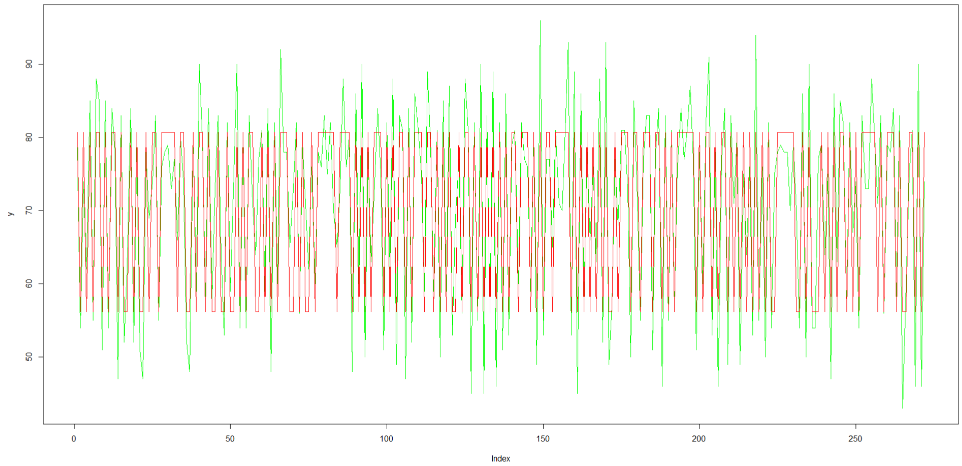

Figure 3: The plot shows the posterior mean superimposed on a plot of data points for K=2

K=3

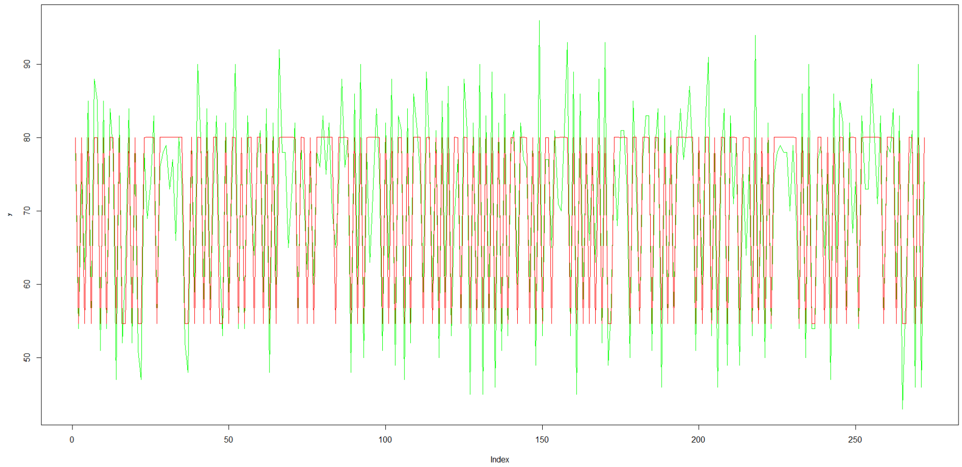

Figure 4: The plot shows the posterior mean superimposed on a plot of data points for K=3

K=4

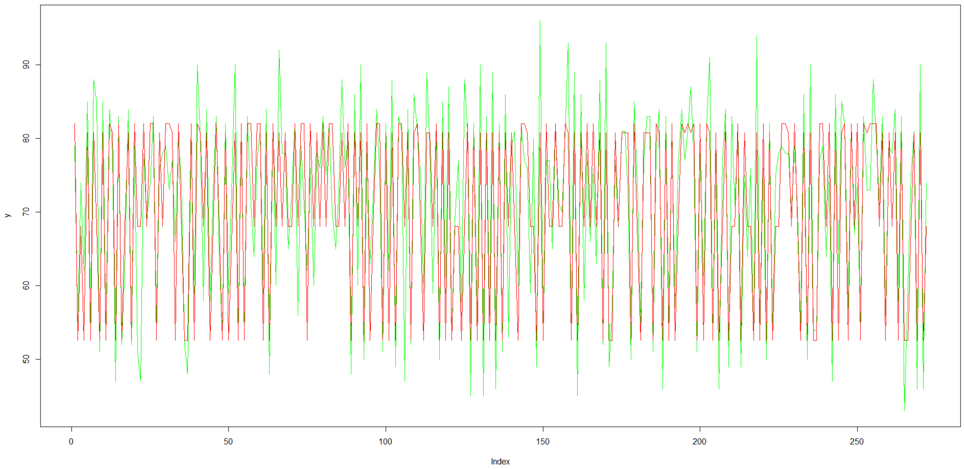

Figure 5: The plot shows the posterior mean superimposed on a plot of data points for K=4

K=5

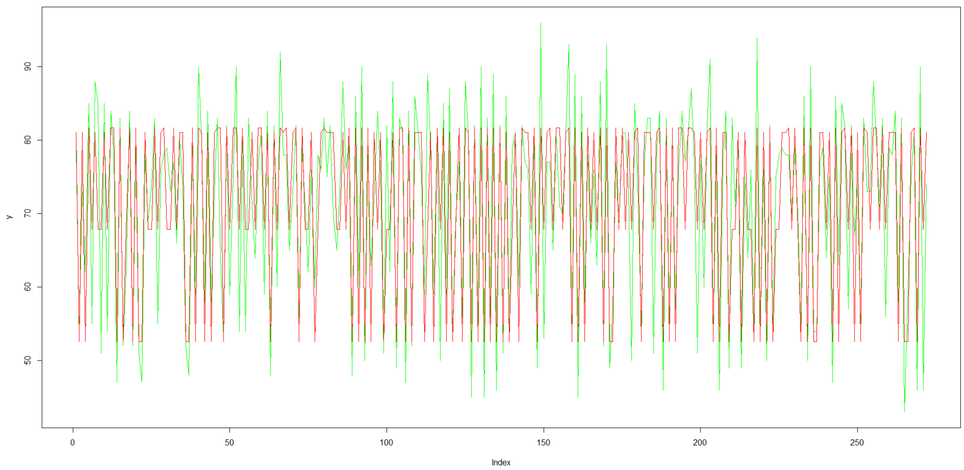

Figure 6: The plot shows the posterior mean superimposed on a plot of data points for K=5

3. We will work with the precision instead of the variance. Denote $\tau = \frac{1}{\sigma^2}$.

Given

$\pi_j \sim \text{Dir}(1/K, \dots, 1/K)$

$\mu_1 \sim \text{Normal}(\mu_{01}, \sigma_{01}^2) = \text{Normal}(-0.5, 1/6)$

$\mu_2 \sim \text{Normal}(\mu_{02}, \sigma_{02}^2) = \text{Normal}(0, 10^{-6})$

$\mu_3 \sim \text{Normal}(\mu_{03}, \sigma_{03}^2) = \text{Normal}(0.5, 1/6)$

$\sigma^2 \sim \text{Inv-Gam}(a_0, b_0) = \text{Inv-Gam}(1, 1)$

$\tau \sim \text{Gamma}(a_0, b_0) = \text{Gamma}(1, 1)$

We can write joint distribution for all parameters of a Normal HMM as:

$$
\begin{aligned}
p(\rho_k, \mathbf{A}, \mu_{\mathbf{k}}, \tau) \propto & \prod_{k=1}^{K} \rho_k^{I(z_1=k)} \prod_{k=1}^{K}\prod_{j=1}^{K} \pi_{jk}^{N_{jk}} \prod_{k=1}^{K}\prod_{t:z_t=k}^{T} \sqrt{\tau}\exp\{-\frac{\tau(y_t-\mu_k)^2}{2}\} \\
& \times \prod_{k=1}^{K} \rho_k^{\frac{1}{K}-1} \prod_{k=1}^{K}\prod_{j=1}^{K} \pi_{jk}^{\frac{1}{K}-1} \times \prod_{k=1}^{K} \exp\{-\frac{(\mu_k-\mu_{0k})^2}{2\sigma_{0k}^2}\} \times \tau^{a_0-1}\exp\{-b_0\tau\}
\end{aligned}
$$

where π_{jk} is the jth row and kth column of the matrix $\mathbf{A}$ (the transition matrix). Looking at the joint posterior, we can derive the full conditional as the following (See the appendix for the details):

$$p(\mu_k|-) = \text{Normal}(\frac{\sigma_{0k}^2(\sum_{t:z_t=k} y_t) + \mu_{0k}\sigma^2}{\sigma_{0k}^2 N_k + \sigma^2}, \frac{\sigma_{0k}^2\sigma^2}{N_k\sigma_{0k}^2 + \sigma^2})$$

$$p(\sigma^2|-) = \text{InvGamma}(a_0 + 0.5N_k, b_0 + 0.5\sum_{k=1}^{K}\sum_{t:z_t=k}(y_t-\mu_k)^2)$$

OR

$$p(\tau|-) = \text{Gamma}(a_0 + 0.5N_k, b_0 + 0.5\sum_{k=1}^{K}\sum_{t:z_t=k}(y_t-\mu_k)^2)$$

$$p(\mathbf{A}_{j.}|-) \propto \prod_{k=1}^{K}\prod_{j=1}^{K} a_{jk}^{N_{jk}+\frac{1}{K}-1} = \text{Dir}(N_{j1} + \frac{1}{K}, N_{j2} + \frac{1}{K}, \dots, N_{jk} + \frac{1}{K})$$

Here, $\mathbf{A}_{j.}$ denotes the j th row of the $\mathbf{A}$ matrix and N_{jk} denotes the number of transition from state j to state k.

$$p(\rho_k|-) \propto \prod_{k=1}^{K} \rho_k^{I(z_1=k)+\frac{1}{K}-1} = \text{Dir}(I(z_1 = k) + \frac{1}{K})$$

Below are the steps in the Gibb sampling algorithm: At $b = 0$, initialize $\mathbf{z}^{(0)}, \pi^{(0)}, \mathbf{A}^{(0)}, \boldsymbol{\mu}^{(0)}, \tau^{(0)}$.

For $b = 1, 2, \dots, B$:

1. Update μ_k

Sample $\mu_k^{(b+1)}$ from Normal$(\frac{\sigma_{0k}^2(\sum_{t:z_t=k} y_t)+\mu_{0k}\sigma^2}{\sigma_{0k}^2 N_k+\sigma^2}, \frac{\sigma_{0k}^2\sigma^2}{N_k\sigma_{0k}^2+\sigma^2})$

2. Update τ

Sample $\tau^{(b+1)}$ from Gamma$(a_0+0.5N_k, b_0+0.5\sum_{k=1}^K\sum_{t:z_t=k}(y_t-\mu_k)^2)$

3. Update $\mathbf{A}$

Compute $N_{jk}^{(b)}$ where N_{jk} denotes the number of transition from state j to state k.

Sample the j th row of $\mathbf{A}^{(b+1)}$ from Dir$(N_{j1}^{(b)}+\frac{1}{K}, N_{j2}^{(b)}+\frac{1}{K}, \ldots, N_{jk}^{(b)}+\frac{1}{K})$

4. Update $\mathbf{z}$

Here, we compute backward probabilities first $\beta_t(z_t)=p(y_{t+1},\ldots,y_T|z_t=j)$.

First, we initialized $\beta_T(z_T)=1$

Then for $t=T-1,\ldots,2$, we use the recursive relationship to update the message.

$\beta_t(z_t)=\sum_{j=1}^K p(y_{t+1}|z_{t+1}=j)p(z_{t+1}=j|z_t)\beta_{t+1}(z_{t+1})$

$(\rho_1,\rho_2,\ldots,\rho_K)\sim$ Dir$(I(z_1=1)+1/K, I(z_1=2)+1/K,\ldots,I(z_1=K)+1/K)$

Sample z_1 with probabilities:

$p(z_1=k|-)^{(b+1)}\propto\rho_k N(y_1|\mu_k^{(b+1)},(\sigma^2)^{(b+1)})p(y_{2:T}|z_1=k)^{(b)}$

Then, for $t=1,2,\ldots,T$, sample z_t with probabilities:

$p(z_t=k|z_{t-1}=j)^{(b+1)}\propto\pi_{jk}^{(b+1)}N(y_t|\mu_k^{(b+1)},(\sigma^2)^{(b+1)})p(y_{t+1:T}|z_t=k)^{(b)}$

$\pi_{jk}^{(b+1)}$ is the jth row kth column of the transition matrix $\mathbf{A}^{(b+1)}$

5. Now, we compute forward probabilities to calculate the predictive density.
$\alpha_t(z_t)=p(y_1,y_2,\ldots,y_t,z_t)$

First, we initialized $\alpha_1(z_1)=p(y_1|z_1)p(z_1)$.

Then for $t=2,\ldots,T$, we use the recursive relationship to update the message.

$\alpha_{t+1}(z_{t+1})=p(y_{t+1}|z_{t+1})\sum_{j=1}^K p(z_{t+1}|z_t=j)\alpha_t(z_t)$

(The derivation for predictive densities are done in the appendix.)

Let ϕ_T be the forward message of $z_T^{(b+1)}$.

$p(y_{t+1}=y|y_1,y_2,\ldots,y_T)=\sum_{k=1}^K(\phi_T\mathbf{A}^{(b+1)})_k N(y|\mu_k^{(b+1)},(\sigma^2)^{(b+1)})$

where $(\phi_T\mathbf{A})_k$ denotes the k th entry of the product between the forward probabilities and the transition matrix.

Please note that we are using the most updated values of the parameters in the updating equations.

The results are shown in the next page. (Numerical results are in the appendix)

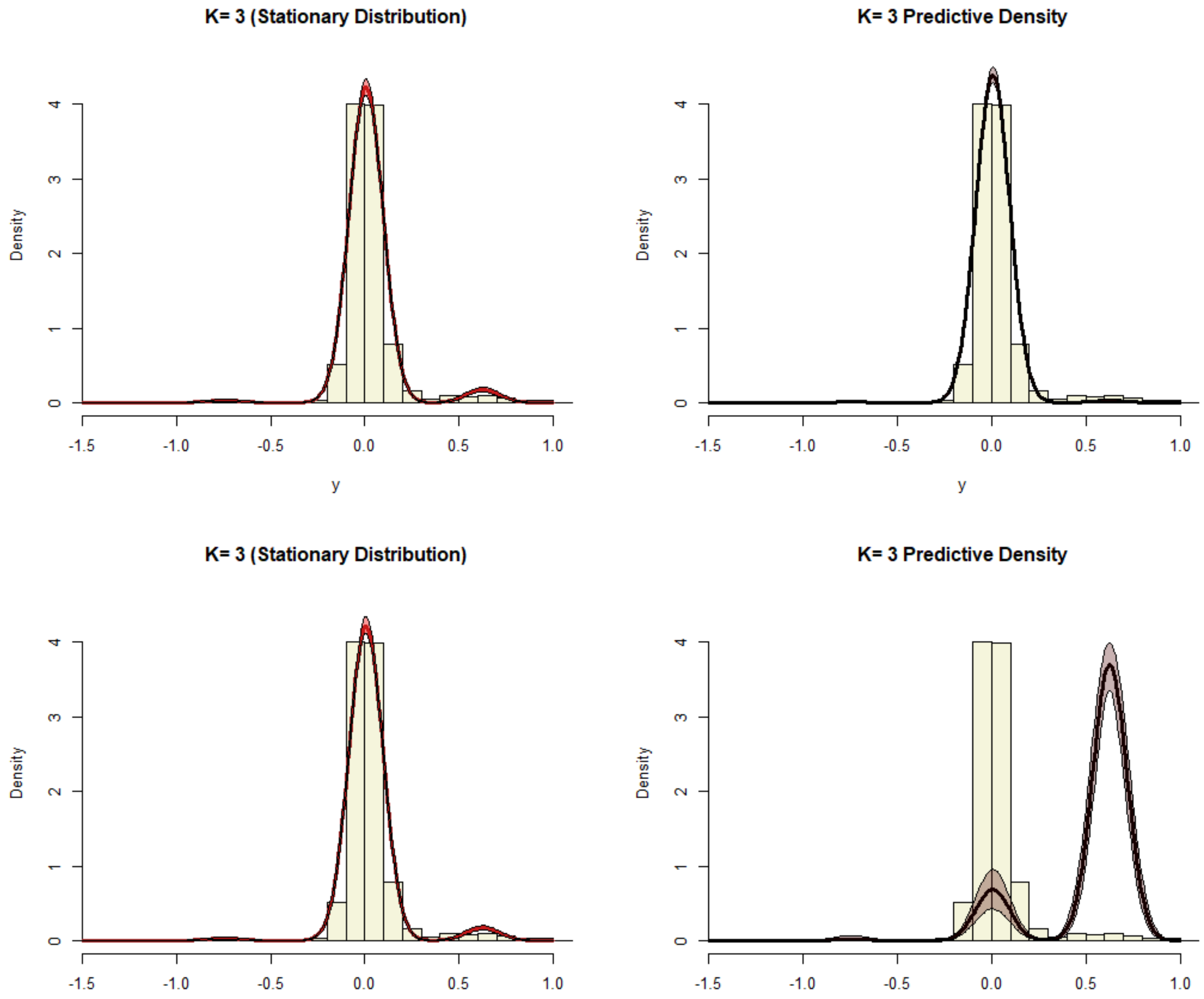

Figure 7: The left plot in the first row shows 90% confidence interval and the posterior mean. The right plot shows the predictive density and 90% confidence interval for 2271 data points. The left plot in the second row shows 90% confidence interval and the posterior mean. The right plot shows the predictive density and 90% confidence interval for 2270 data points.

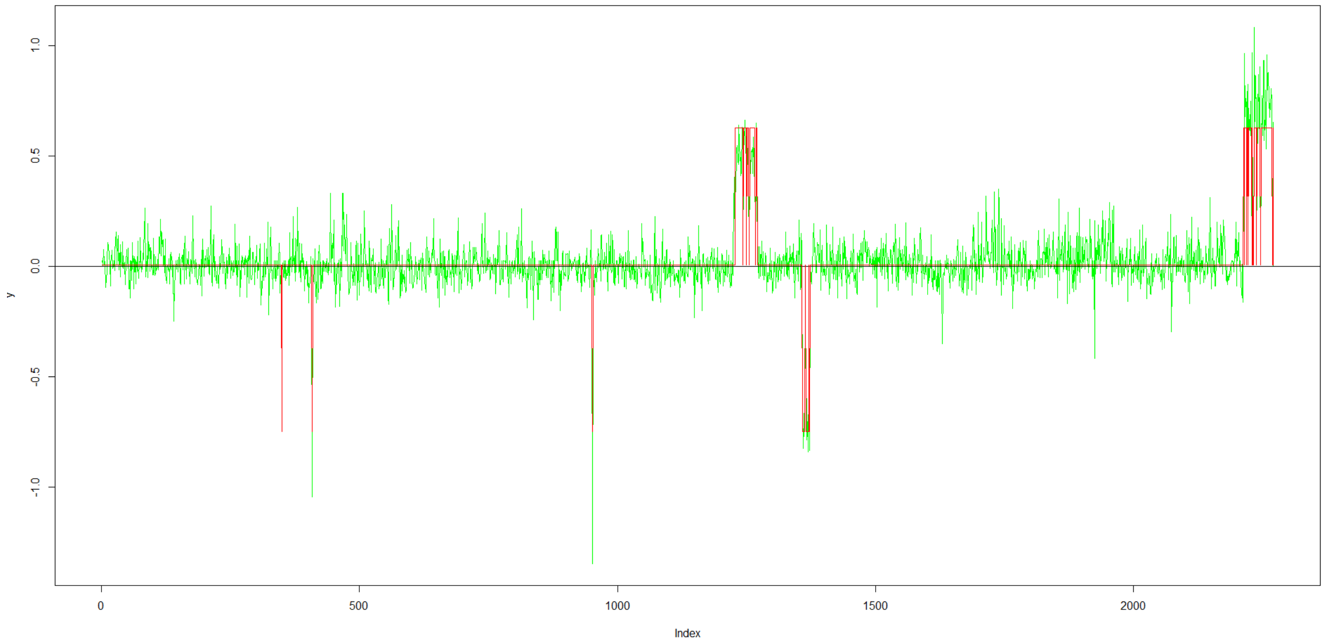

Figure 8: The plot shows the posterior mean superimposed on a plot of data points for 2271 data points

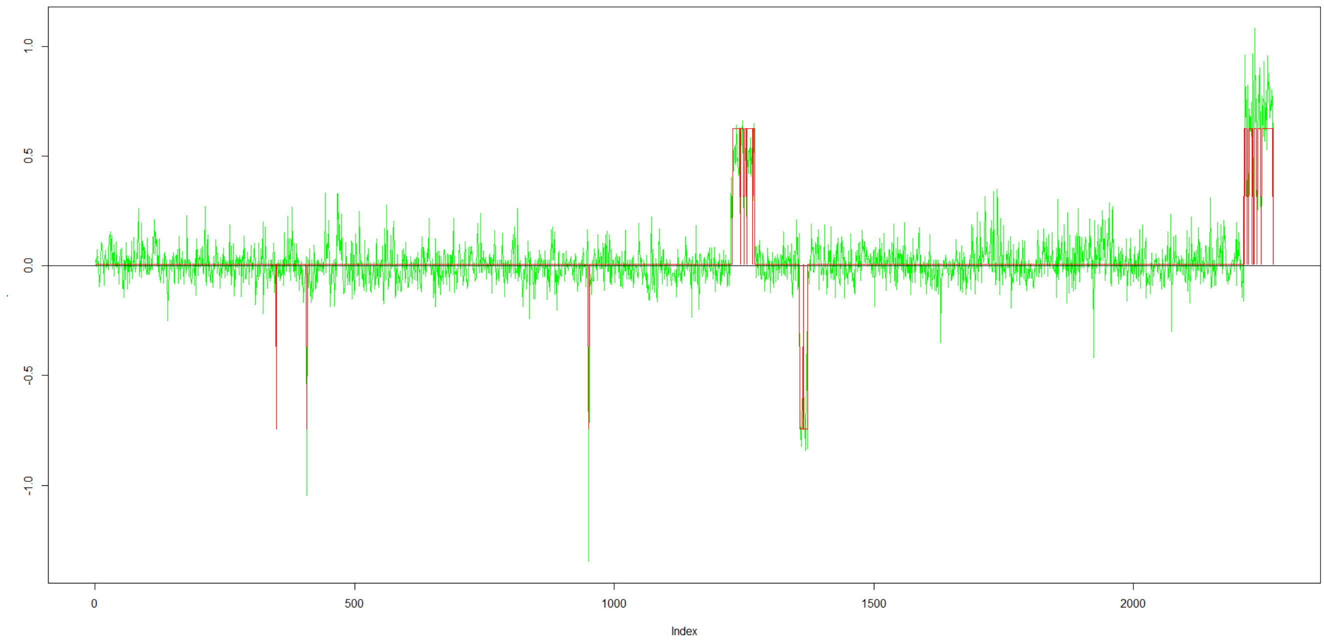

Figure 9: The plot shows the posterior mean superimposed on a plot of data points for 2270 data points

Appendix

Derivation for Full Conditional Problem 2

$$
\begin{aligned}
p(\mu_k|-) &\propto \prod_{t:z_t=k} \exp(-\frac{1}{2\sigma_k^2}(y_t-\mu_k)^2) \times \exp(-\frac{1}{2\sigma_0^2}(\mu_k-\mu_0)^2) \\
&= \exp(-\frac{1}{2\sigma_k^2}\sum_{t:z_t=k}(y_t-\mu_k)^2 - \frac{1}{2\sigma_0^2}(\mu_k-\mu_0)^2) \\
&\propto \{-\frac{1}{2}[(\frac{N_k}{\sigma_k^2}+\frac{1}{\sigma_0^2})\mu_k^2 - 2(\frac{\mu_0}{\sigma_0^2}+\frac{1}{\sigma_k^2}\sum_{t:z_t=k} y_t)\mu_k]\} \\
&= \exp\{-\frac{1}{2}(\frac{N_k}{\sigma_k^2}+\frac{1}{\sigma_0^2})[\mu_k^2 - 2(\frac{N_k}{\sigma_k^2}+\frac{1}{\sigma_0^2})^{-1}(\frac{\mu_0}{\sigma_0^2}+\frac{1}{\sigma_k^2}\sum_{t:z_t=k} y_t)\mu_k]\} \\
&\propto N(\frac{\sigma_0^2(\sum_{t:z_t=k} y_t)+\mu_0\sigma_k^2}{\sigma_0^2 N_k+\sigma_k^2}, \frac{\sigma_0^2\sigma_k^2}{N_k\sigma_0^2+\sigma_k^2})
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
p(\sigma_k^2|-) &\propto (\sigma_k^2)^{-a_0-1}\exp(-\frac{b_0}{\sigma_k^2})[\prod_{t:z_t=k}(\sigma_k)^{-\frac{1}{2}}\exp(-\frac{1}{2\sigma_k^2}(y_t-\mu_k)^2)] \\
&= (\sigma_k^2)^{-a_0-\frac{N_k}{2}-1}\exp(-(\frac{1}{\sigma^2})[b_0+\frac{1}{2}\sum_{t:z_t=k}(y_t-\mu_k)^2]) \\
&\propto \mathrm{InvGamma}(a_0+\frac{N_k}{2}, b_0+\frac{1}{2}\sum_{t:z_t=k}(y_t-\mu_k)^2)
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
p(\rho_k|-) &\propto [\prod_{k=1}^{K}\rho_k^{I(z_1=k)}][\prod_{k=1}^{K}\rho_k^{\frac{1}{K}}-1] \\
&\propto \mathrm{Dir}(I(z_1=1)+\frac{1}{K},\ldots,I(z_1=K)+\frac{1}{K})
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
p(A|-) &\propto [\prod_{j=1}^{K}\prod_{k=1}^{K} a_{ij}^{N_{ij}}][\prod_{j=1}^{K}\prod_{k=1}^{K} a_{ij}^{\frac{1}{K}-1}] \\
&\propto \prod_{j=1}^{K}\prod_{k=1}^{K} a_{ij}^{N_{ij}+\frac{1}{K}-1}
\end{aligned}
\tag{9}
$$

Derivation for Full Conditional Problem 3

$$
\begin{aligned}
p(\mu_k|-) &\propto [\prod_{t:z_t=k}\exp(-\frac{1}{2\sigma^2})(y_t-\mu_k)^2](\exp(-\frac{1}{2\sigma_{0k}^2(\mu_k-\mu_{0k})^2})) \\
&= \exp(-\frac{1}{2\sigma^2}\sum_{t:z_t=k}(y_t-\mu_k)^2 - \frac{1}{2\sigma_{0k}^2}(\mu_k-\mu_{0k})^2) \\
&\propto \exp\{-\frac{1}{2}[(\frac{N_k}{\sigma^2}+\frac{1}{\sigma_{0k}^2})\mu_k^2 - 2(\frac{1}{\sigma^2}\sum_{t:z_t=k} y_t+\frac{\mu_{0k}}{\sigma_{0k}^2})\mu_k]\} \\
&\propto N(\frac{\sigma^2\mu_{0k}+\sigma_{0k}^2\sum_{t:z_t=k} y_t}{\sigma_{0k}^2 N_k\sigma^2}, \frac{\sigma^2\sigma_{0k}^2}{\sigma_{0k}^2 N_k+\sigma^2})
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
p(\sigma^2|-)&[\prod_{k=1}^{K}\prod_{t:z_t=k}\frac{1}{(\sigma^2)^{\frac{1}{2}}}\exp(-\frac{1}{2\sigma^2}(y_t-\mu_k)^2)][(\sigma^2)^{-1-1}\exp(-\frac{1}{sigma^2})] \\
&=(\sigma^2)^{\frac{T}{2}}\exp(-\frac{1}{2\sigma^2}\sum_{k=1}^{K}\sum_{t:z_t=k}(y_t-\mu_k)^2)[(\sigma^2)^{-1-1}\exp(-\frac{1}{\sigma^2})] \\
&=(\sigma^2)^{-\frac{T}{2}-1-1}\exp(-\frac{1}{\sigma^2}[1+\frac{1}{2}\sum_{k=1}^{K}\sum_{t:z_t=k}(y_t-\mu_k)^2]) \\
&\propto \text{InvGamma}(1+0.5T,1+0.5\sum_{k=1}^{K}\sum_{t:z_t=k}(y_t-\mu_k)^2)
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
p(\rho_k|-)&\propto[\prod_{k=1}^{K}\rho_k^{I(z_1=k)}][\prod_{k=1}^{K}\rho_k^{\frac{1}{K}}-1] \\
&\propto \text{Dir}(I(z_1=1)+\frac{1}{K},\dots,I(z_1=K)+\frac{1}{K})
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
p(A|-)&\propto[\prod_{j=1}^{K}\prod_{k=1}^{K}a_{ij}^{N_{ij}}][\prod_{j=1}^{K}\prod_{k=1}^{K}a_{ij}^{\frac{1}{K}-1}] \\
&\propto \prod_{j=1}^{K}\prod_{k=1}^{K}a_{ij}^{N_{ij}+\frac{1}{K}-1}
\end{aligned} \tag{13}
$$

Derivation for One-Step Ahead Predictive Density

$$
\begin{aligned}
p(z_{T+1}=j|y_1,\dots,y_T)&=\sum_{k=1}^{K}p(z_{T+1}=j|z_T=k)p(z_T=k|y_1,\dots,y_T) \\
&=\sum_{k=1}^{K}a_{kj}\frac{\alpha_T(z_T)}{\sum_{z_T}\alpha_T(z_T)} \\
&=\phi_T A_{.j}
\end{aligned} \tag{14}
$$

where ϕ_T denotes the normalized message at time T and $A_{.j}$ denotes the j th column of the transition matrix.

$$
\begin{aligned}
p(y_{T+1}|y_1,\dots,y_T)&=\sum_{j=1}^{K}p(y_{T+1}|z_{T+1}=j)p(z_{T+1}=j|y_1,\dots,y_T) \\
&=\sum_{j=1}^{K}N(y_{T+1}|\mu_j,\sigma_j^2)\phi_T A_{.j}
\end{aligned} \tag{15}
$$

Problem 2

$K = 2$

$\mu_1 = 56.54500$

$\mu_2 = 81.30917$

$\sigma_1^2 = 55.02263$

$\sigma_2^2 = 28.94764$

$\pi_1 = 0.4002496$

$\pi_2 = 0.5997504$

$K = 3$

$\mu_1 = 54.64694$

$\mu_2 = 80.06631$

$\mu_3 = 80.02590$

$\sigma_1^2 = 35.27935$

$\sigma_2^2 = 26.59226$

$\sigma_3^2 = 26.73240$

$\pi_1 = 0.3603346$

$\pi_2 = 0.3208458$

$\pi_3 = 0.3188196$

$K = 4$

$\mu_1 = 79.75442$

$\mu_2 = 68.05761$

$\mu_3 = 52.60266$

$\mu_4 = 83.03735$

$\sigma_1^2 = 14.21735$

$\sigma_2^2 = 33.10287$

$\sigma_3^2 = 21.63402$

$\sigma_4^2 = 20.94519$

$\pi_1 = 0.2530192$

$\pi_2 = 0.2301550$

$\pi_3 = 0.2589393$

$\pi_4 = 0.2578865$

$K = 5$

$\mu_1 = 55.30738$

$\mu_2 = 81.31030$

$\mu_3 = 55.81092$

$\mu_4 = 79.56099$

$\mu_5 = 80.75410$

$\sigma_1^2 = 25.74618$

$\sigma_2^2 = 17.92565$

$\sigma_3^2 = 26.72735$

$\sigma_4^2 = 15.87784$

$\sigma_5^2 = 17.47286$

$\pi_1 = 0.2049112$

$\pi_2 = 0.2039303$

$\pi_3 = 0.1939520$

$\pi_4 = 0.1928180$

$\pi_5 = 0.2043886$

Problem 3

With 2271 data points

$\mu_1 = -0.749583346$

$\mu_2 = 0.006652452$

$\mu_3 = 0.627342510$

$\sigma^2 = 0.008063701$

$\pi_1 = 0.007443937$

$\pi_2 = 0.953469925$

$\pi_3 = 0.039086139$

With 2270 data points

$\mu_1 = 0.006680238$

$\mu_2 = 0.627571703$

$\mu_3 = -0.744963893$

$\sigma^2 = 0.008075168$

$\pi_1 = 0.953620509$

$\pi_2 = 0.038830887$

$\pi_3 = 0.007548604$