Statistical Modelling HW5

Korawat Tanwisuth

October 20, 2018

1. Importance Resampling

To obtain independent samples from t_3 distribution, we use $N(0,3)$ as the proposal density and follow the following steps.

1. Sample $\theta_1, \theta_2, \ldots, \theta_S$ from $N(0,3)$ the proposal density.

2. Sample θ_s from the drawn samples $\theta_1, \theta_2, \ldots, \theta_S$ with probability equal to $w = \frac{q(\theta_s)}{g(\theta_s)}$ where q is the target density and g is the proposal density.

3. Repeat the 2nd step n times and exclude the value of θ that has been sampled.

Below are the results for $n = 100$ and $n = 10000$ respectively.

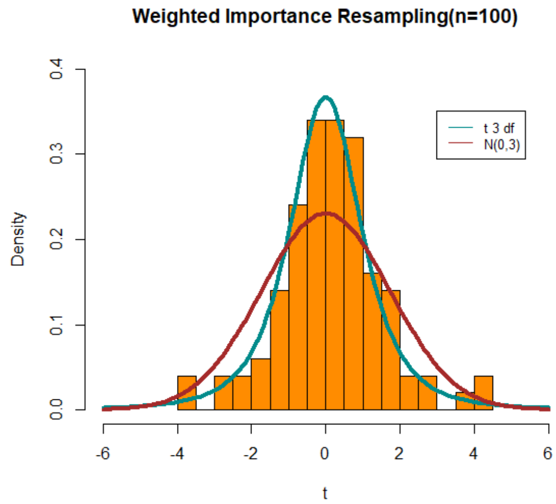


Figure 1: Histogram of Weighted Resampling Samples (n=100)

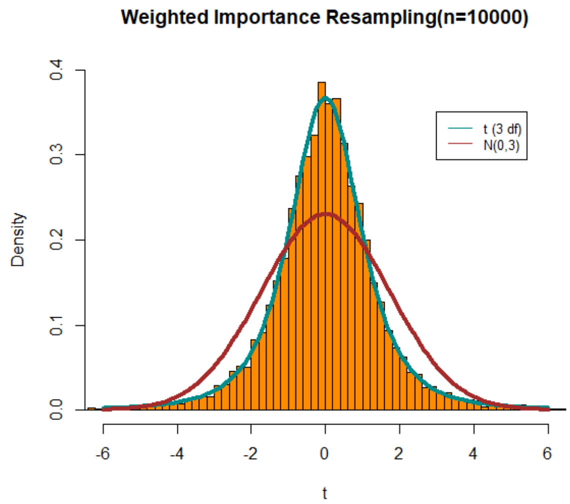


Figure 2: Histogram of Weighted Resampling Samples (n=10000)

We can see that for $n = 100$ the histogram still does not seem to approximate the density well. However, for $n = 10000$, we can see that theoretical density of t_3 aligns almost exactly with the histogram.

Importance Sampling

To calculate the mean and variance using importance sampling, we follow the following steps.

1. Sample $\theta_1, \theta_2, \ldots, \theta_S$ from $N(0, 3)$ the proposal density.

2. Calculate $\hat{\mu} = \dfrac{\sum_{s=1}^{S} \theta_s w_s}{S}$ where $w_s = \dfrac{q(\theta_s)}{g(\theta_s)}$.

3. Calculate $\hat{\sigma}^2 = \dfrac{\sum_{s=1}^{S} \theta_s^2 w_s}{S} - \dfrac{\sum_{s=1}^{S} \theta_s w_s}{S}$ where $w_s = \dfrac{q(\theta_s)}{g(\theta_s)}$.

For n=100,

$\hat{\mu} = 0.1469252$

$\hat{\sigma}^2 = 2.627491$

For n=10000,

$\hat{\mu} = -0.02077675$

$\hat{\sigma}^2 = 2.904507$

We expect $\hat{\mu} = 0$ and $\hat{\sigma}^2 = 3$. We can see that for small n, the estimate is not well very accurate. However, as n grows larger, we get a much better estimate.

2. To obtain independent samples from a bivariate normal distribution, we apply the Gibbs sampling algorithm.

1. Initialize $x_1^{(0)}$ and $x_2^{(0)}$ to some values. We simply initialize them to the mean 0 and 2 respectively.

2. Since $(x_1, x_2) \sim N(\mu, \Sigma)$ are jointly normal, the conditional distribution is also normal $x_1|x_2 \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2))$ and by symmetry $x_2|x_1 \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2))$.

Thus, at the nth iteration we sample $x_1^{(n)}$ from $N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2^{(n-1)} - \mu_2), \sigma_1^2(1-\rho^2))$ and $x_2^{(n)}$ from $N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1^{(n)} - \mu_1), \sigma_2^2(1-\rho^2))$ (Note that we use the most updated value of x_1 to sample x_2).

3. Repeat the second step N times (N is chosen such that the final samples after burned in and thinning is 1000).

4. Burn and thin the samples. We choose to burn 10 percent of the samples and thin every 5 samples.

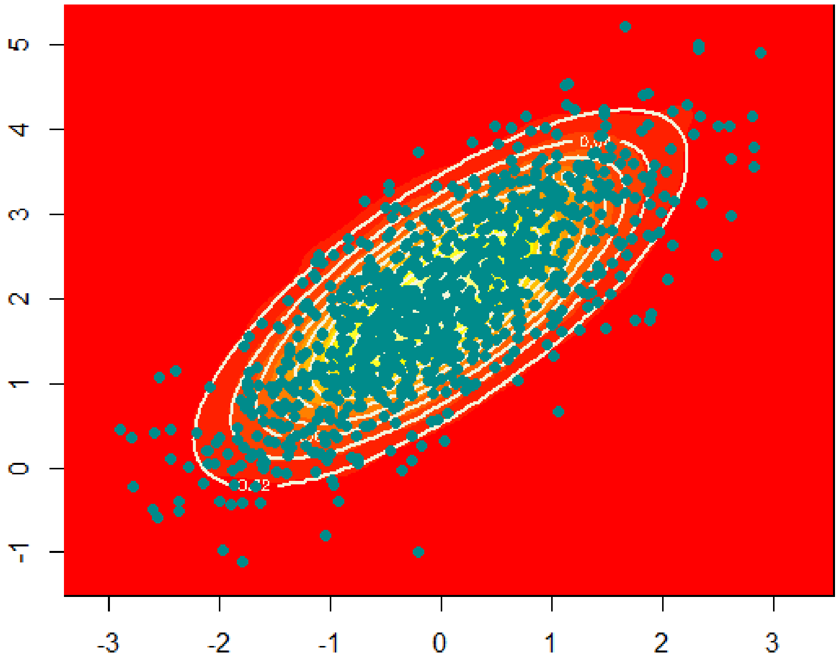
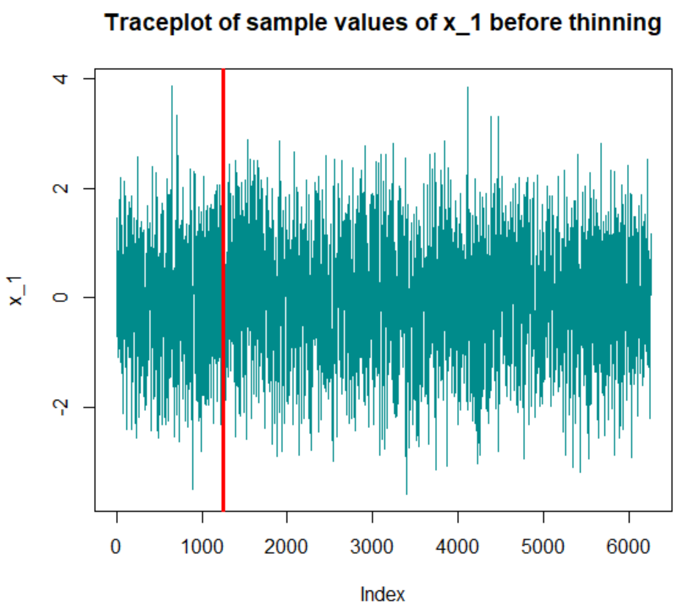
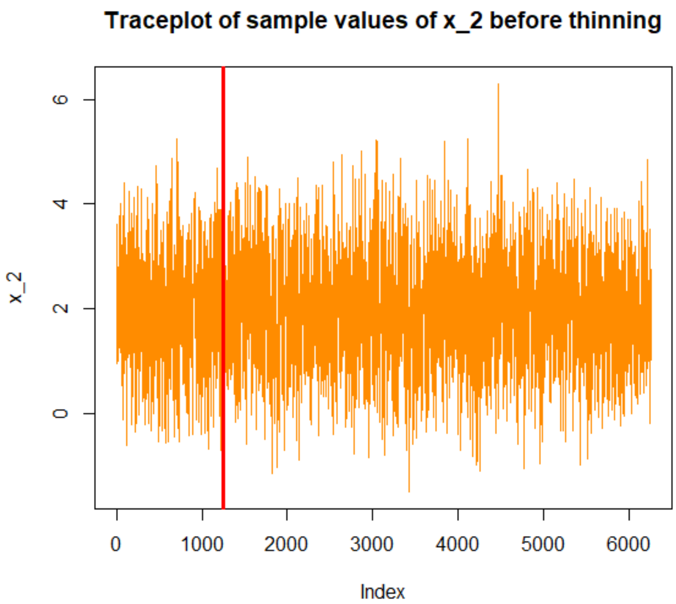
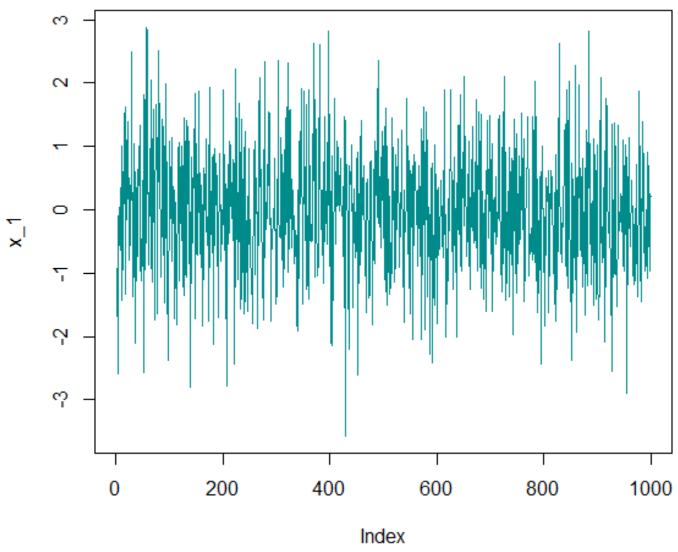


Figure 3: Contour Plot of Bivariate Normal ($\mu_1 = 0, \mu_2 = 2$),($\sigma_1 = 1, \sigma_2 = 1$), and $\rho = 0.75$. Blue dots represent 1000 Gibbs samples

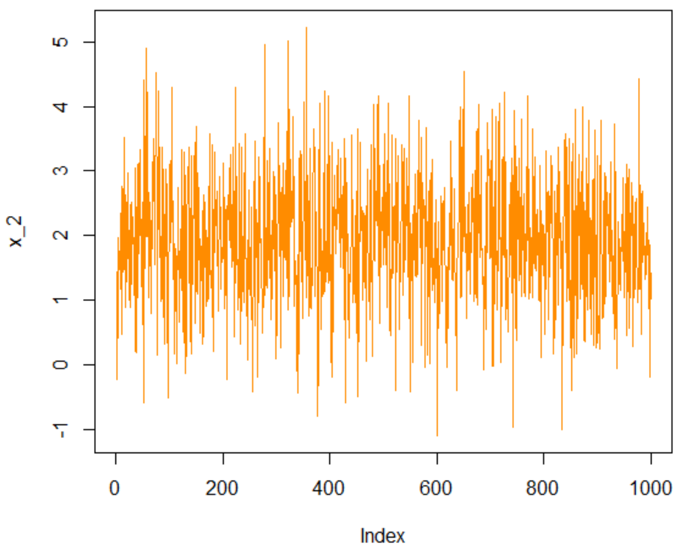

Traceplot of sample values of x_1 before thinning
x_1
Index
4
2
0
-2
0
1000
2000
3000
4000
5000
6000

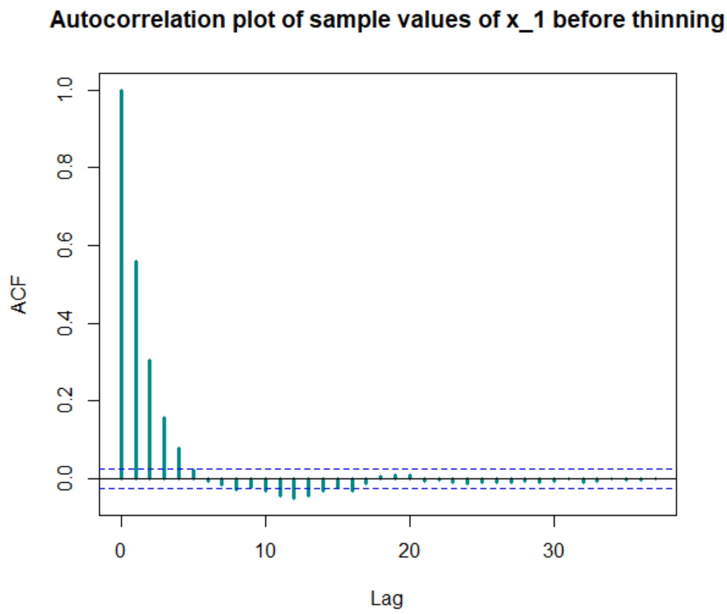

Traceplot of sample values of x_2 before thinning
x_2
Index
6
4
2
0
0
1000
2000
3000
4000
5000
6000

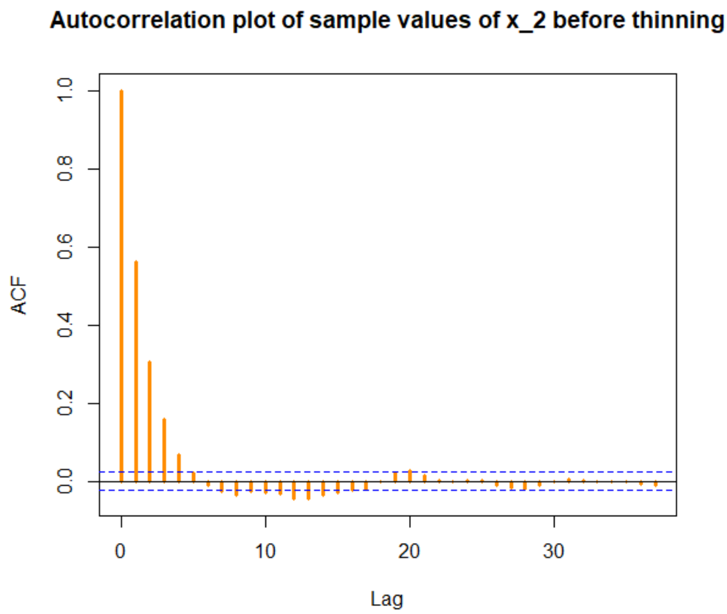
Traceplot of sample values of x_1 after thinning



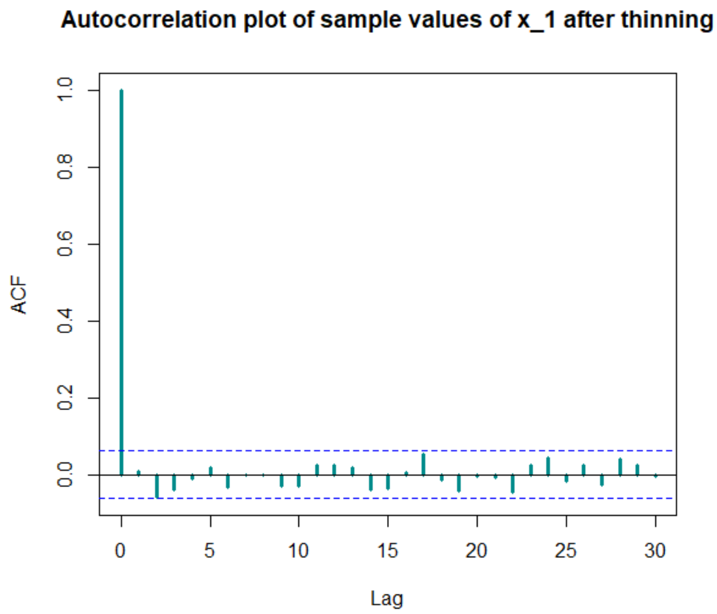
Traceplot of sample values of x_2 after thinning







Before thinning, we can see that the samples are correlated for small lags.



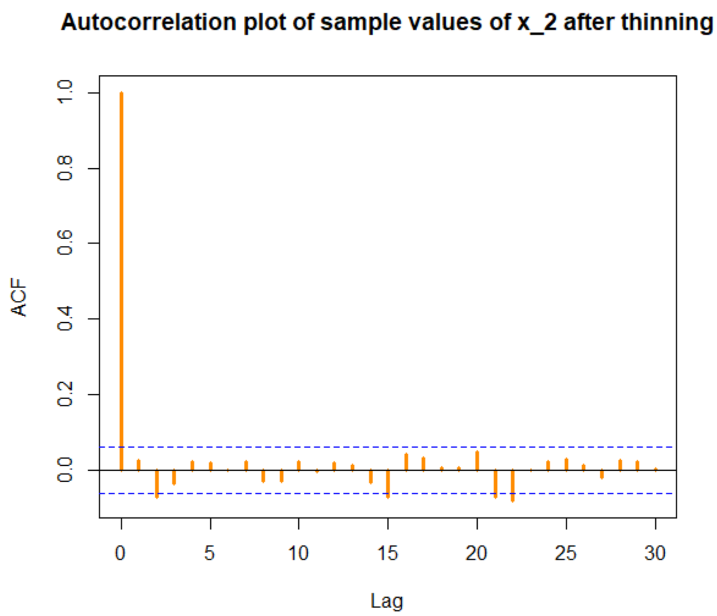


Figure 4:

We can see that the autocorrelation plot now looks better with thinning.

3. We assume a few assumptions here (Abhra approves).

1. $Var(y_i) = Var(y_j)$ for all $i, j = 1, 2, \ldots, n$
2. y_i, ϵ_j are independent for all $i < j$
3. $E(\epsilon_t) = 0$
4. $Var(\epsilon_t) = \sigma^2$
5. $E(y_i) = E(y_j)$ for all $i, j = 1, 2, \ldots, n$

a.)

$$
\begin{aligned}
Var(y_t) &= Var(\mu + \rho(y_{t-1} - \mu) + \epsilon_t) \\
&= \rho^2 Var(y_{t-1}) + Var(\epsilon_t) \text{ by } (2) \\
&= \rho^2 Var(y_t) + \sigma^2 \\
&\rightarrow Var(y_t) = \frac{\sigma^2}{(1-\rho^2)}
\end{aligned} \tag{1}
$$

b.)

$$
\begin{aligned}
y_t &= \mu + \rho(y_{t-1} - \mu) + \epsilon_t \\
&= \mu + \rho(\rho(y_{t-2} - \mu) + \epsilon_{t-1}) + \epsilon_t \\
&= \mu + \rho(\rho(\rho(y_{t-3} - \mu) + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\
&\dots \\
&= \mu + \rho^k y_{t-k} - \rho^k \mu + \rho^{(k-1)} \epsilon_{t-(k-1)} + \ldots + \rho^3 \epsilon_{t-3} + \rho^2 \epsilon_{t_2} + \rho \epsilon_{t-1} + \epsilon_t
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
Cov(y_t, y_{t-k}) &= Cov(\mu + \rho^k y_{t-k} - \rho^k \mu + \rho^{(k-1)} \epsilon_{t-(k-1)} + \ldots + \rho^3 \epsilon_{t-3} + \rho^2 \epsilon_{t_2} + \rho \epsilon_{t-1} + \epsilon_t, y_{t-k}) \\
&= Cov(y_{t-k}, \rho^k y_{t-k}) \\
&= \rho^k Var(y_{t-k})
\end{aligned} \tag{3}
$$

The second equality is justified since y_{t-k} is independent of all ϵ_j $j > t-k$ and $Cov(y_{t-k}, c) = 0$ where c is some constant.

$$
\begin{aligned}
Cor(y_t, y_{t-k}) &= \frac{Cov(y_t, y_{t-k})}{Var(\sqrt{y_t})\sqrt{Var(y_{t-k})}} \\
&= \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_{t-k})}\sqrt{Var(y_{t-k})}} \text{ by } (1) \\
&= \rho^k \frac{Var(y_{t-k})}{Var(y_{t-k})} \\
&= \rho^k
\end{aligned} \tag{4}
$$

c.)

$$\begin{aligned}
E(y_t) &= E(\mu + \rho(y_{t-1} - \mu) + \epsilon_t) \\
&= \mu + \rho E(y_{t-1}) - \rho\mu + 0 \\
&= \mu + \rho E(y_t) - \rho\mu \text{ by (5)} \\
&\rightarrow E(y_t) - \rho E(y_t) = \mu - \rho\mu \\
&\rightarrow E(y_t)(1-\rho) = \mu(1-\rho) \\
&\rightarrow E(y_t) = \mu \quad (\text{Since } |\rho| < 1) \\
E(\bar{y_n}) &= \frac{1}{n} E(\sum_i^n y_i) \\
&= \frac{1}{n}(\sum_i^n E(y_i)) \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned} \tag{5}$$

d.)

$$\begin{aligned}
Var(\bar{y_n}) &= Cov(\frac{\sum_{i=1}^n y_i}{n}, \frac{\sum_{j=1}^n y_j}{n}) \\
&= \frac{1}{n^2}[\sum_{i=1}^n Var(y_i) + 2\sum_{i<j} Cov(y_i, y_j)] \\
&= \frac{1}{n^2}[\sum_{i=1}^n Var(y_i) + 2\{Cov(y_1, y_2) + Cov(y_1, y_3) + \dots Cov(y_1, y_n) \\
&+ Cov(y_2, y_3) + Cov(y_2, y_4) + \dots + Cov(y_2, y_n) + \dots \\
&+ Cov(y_{n-2}, y_{n-1}) + Cov(y_{n-2}, y_n) + Cov(y_{n-1}, y_n)\}] \\
&= \frac{1}{n^2}[n\frac{\sigma^2}{1-\rho^2} + 2(\sum_{j=1}^{n-1}(n-j)\rho^j \frac{\sigma^2}{1-\rho^2})] \\
&= \frac{\sigma^2}{n^2(1-\rho^2)}[n + 2(\sum_{j=1}^{n-1}(n-j)\rho^j)]
\end{aligned} \tag{6}$$

The second to last equation is justified because there are $n-1$ terms with time lag 1 since the index runs from $1, 2, \dots, n$, and $n-2$ terms with time lag 2. In general, $n-j$ terms with time lag j. By part b, we found $Cov(y_t, y_{t-k})$ to be $\rho^k Var(y_{t-k}) = \rho^k \frac{\sigma^2}{1-\rho^2}$. Also, $Var(y_t) = \frac{\sigma^2}{1-\rho^2}$ from part 1.

e.)

$$nVar(\bar{y}) = \frac{\sigma^2}{n(1-\rho^2)}[n + 2(\sum_{j=1}^{n-1}(n-j)\rho^j)]$$

$$
\begin{aligned}
\lim_{n\to\infty} nVar(\bar{y}) &= \lim_{n\to\infty} \frac{\sigma^2}{n(1-\rho^2)}[n + 2(\sum_{j=1}^{n-1}(n-j)\rho^j)] \\
&= \frac{\sigma^2}{(1-\rho^2)}(1 + \lim_{n\to\infty} 2(\sum_{j=1}^{n-1}(1-\frac{j}{n})\rho^j)) \\
&= \frac{\sigma^2}{(1-\rho^2)}(1 + 2\lim_{n\to\infty}(\sum_{j=1}^{n-1}\rho^j - \sum_{j=1}^{n-1}\frac{j}{n}\rho^j)) \\
&= \frac{\sigma^2}{(1-\rho^2)}(1 + 2(\lim_{n\to\infty}\sum_{j=1}^{n-1}\rho^j - \lim_{n\to\infty}\sum_{j=1}^{n-1}\frac{j}{n}\rho^j)) \\
&= \frac{\sigma^2}{(1-\rho^2)}(1 + \frac{2\rho}{1-\rho} - 0) \quad \text{(See below)} \\
&= \frac{\sigma^2}{(1+\rho)(1-\rho)}(\frac{1-\rho+2\rho}{1-\rho}) \\
&= \frac{\sigma^2}{(1-\rho)^2}
\end{aligned} \tag{7}
$$

$\lim_{n\to\infty}\sum_{j=1}^{n-1}\rho^j = \sum_{j=1}^{\infty}\rho^j = \frac{\rho}{1-\rho}$ (limit of geometric series)

Now, we will find $\lim_{n\to\infty}\sum_{j=1}^{n-1}\frac{j}{n}\rho^j$.

Define $k_n = \sum_{j=1}^{n-1} j\rho^j$ and $S_n = \frac{1}{n}k_n$.

Now, we will show that $\lim_{n\to\infty} k_n$ exists by using the ratio test.

$\lim_{n\to\infty} k_n = \sum_{j=1}^{\infty} j\rho^j$

$L = \lim_{j\to\infty}|\frac{a_{j+1}}{a_j}| = \lim_{j\to\infty}|\frac{(j+1)\rho^{j+1}}{j\rho^j}| = \lim_{j\to\infty}|\frac{j\rho+\rho}{j}| = \lim_{j\to\infty}|\rho + \frac{\rho}{j}| = |\rho| < 1$

Since $L < 1$, $\lim_{n\to\infty} k_n$ exists.

$\lim_{n\to\infty}\sum_{j=1}^{n-1}\frac{j}{n}\rho^j = \lim_{n\to\infty} S_n = \lim_{n\to\infty}\frac{1}{n}k_n = \lim_{n\to\infty}\frac{1}{n}\lim_{n\to\infty}k_n = 0 * \lim_{n\to\infty}k_n = 0$ provided that both limit exists (we showed this in above).

f.)

$$
\begin{aligned}
E(s^2) = E(&\frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}) \\
&\frac{\sum_{i=1}^n E(y_i^2) - nE(\bar{y}^2)}{n-1} \\
&\frac{\sum_{i=1}^n (Var(y_i) + E(y_i)^2) - n(Var(\bar{y}) + E(\bar{y})^2)}{n-1} \\
&\frac{1}{n-1}(n(\frac{\sigma^2}{1-\rho^2} + \mu^2) - nVar(\bar{y}) - n\mu^2) \\
&\frac{n}{n-1}\frac{\sigma^2}{1-\rho^2} - \frac{n}{n-1}Var(\bar{y})
\end{aligned} \tag{8}
$$

$$\begin{aligned}
\lim_{n\to\infty} E(s^2) &= \lim_{n\to\infty}(\frac{n}{n-1}\frac{\sigma^2}{1-\rho^2} - \frac{n}{n-1}Var(\bar{y})) \\
&= \frac{\sigma^2}{1-\rho^2}\lim_{n\to\infty}\frac{n}{n-1} - \lim_{n\to\infty}\frac{n}{n-1}Var(\bar{y}) \\
&= \frac{\sigma^2}{1-\rho^2}(1) - (\lim_{n\to\infty} nVar(\bar{y})\lim_{n\to\infty}\frac{1}{n-1}) \\
&= \frac{\sigma^2}{1-\rho^2}(1) - \lim_{n\to\infty} nVar(\bar{y})(0) \\
&= \frac{\sigma^2}{1-\rho^2} \\
&= Var(y_t)
\end{aligned} \tag{9}$$

$\lim_{n\to\infty}\frac{n}{n-1}Var(\bar{y}) = (\lim_{n\to\infty} nVar(\bar{y})\lim_{n\to\infty}\frac{1}{n-1})$ since we showed in e) that the limit of $nVar(\bar{y})$ exists so the product of limit is the limit of product and the second term goes to 0 as n goes to ∞.

4. a.)

We use normal-inverse gamma as the conjugate prior for this model.

$$\begin{aligned}
p(\rho,\sigma^2) &= (\sigma^2)^{-(\frac{v_0}{2}+1+\frac{1}{2})}\exp(\frac{-1}{2\sigma^2}(v_0\sigma_0^2 + k_0(\rho-\rho_0)^2)) \\
&= \text{NIG}(\rho_0, \frac{\sigma_0^2}{k_0}, v_0, \sigma_0^2)
\end{aligned} \tag{10}$$

$$\begin{aligned}
p(\rho,\sigma^2|y_{1:T}) &\propto \prod_{t=2}^{T} p(y_t|\rho,\sigma^2)p(\rho,\sigma^2) \\
&\propto (\sigma^2)^{-(\frac{T-1}{2})}\exp(\frac{-1}{2\sigma^2}\sum_{t=2}^{T}(y_t-\rho y_{t-1})^2)(\sigma^2)^{-(\frac{v_0}{2}+1+\frac{1}{2})}\exp(\frac{-1}{2\sigma^2}(v_0\sigma_0^2+k_0(\rho-\rho_0)^2)) \\
&= (\sigma^2)^{-(\frac{v_n}{2}+1+\frac{1}{2})}\exp(-\frac{1}{2\sigma^2}(k_n(\rho-\rho_n)^2 + v_n(\sigma_n^2))) \\
&= \text{NIG}(\rho_n, \frac{\sigma_n^2}{k_n}, v_n, \sigma_n^2)
\end{aligned} \tag{11}$$

$$\begin{aligned}
\rho_n &= \frac{\sum_{t=2}^{T} y_t y_{t-1} + k_0\rho_0}{\sum_{t=2}^{T} y_{t-1}^2 + k_0} \\
v_n &= T-1+v_0 \\
\sigma_n^2 &= \frac{-k_n\rho_n^2 + \sum_{t=2}^{T} y_t^2 + v_0\sigma_0^2 + k_0\rho_0^2}{v_n} \\
k_n &= \sum_{t=2}^{T} y_{t-1}^2 + k_0
\end{aligned}$$

The parameters above are obtained after completing the square and rearranging the terms. Steps for completing the square are shown below.

$$
\begin{aligned}
&\sum_{t=2}^{T}(y_t-\rho y_{t-1})^2+v_0\sigma_0^2+k_0(\rho-\rho_0)^2\\
&=\sum_{t=2}^{T}(y_t^2-2\rho y_t y_{t-1}+\rho^2 y_{t-1}^2)+v_0\sigma_0^2+k_0(\rho^2-2\rho\rho_0+\rho_0^2)\\
&=\sum_{t=2}^{T}y_t^2-2\sum_{t=2}^{T}\rho y_t y_{t-1}+\rho^2\sum_{t=2}^{T}y_{t-1}^2+v_0\sigma_0^2+k_0\rho^2-2k_0\rho\rho_0+k_0\rho_0^2\\
&=(\sum_{t=2}^{T}y_{t-1}^2+k_0)[\rho^2-2\rho(\frac{\sum_{t=2}^{T}y_t y_{t-1}+k_0\rho_0}{\sum_{t=2}^{T}y_{t-1}^2+k_0})+(\frac{\sum_{t=2}^{T}y_t y_{t-1}+k_0\rho_0}{\sum_{t=2}^{T}y_{t-1}^2+k_0})^2]-\frac{(\sum_{t=2}^{T}y_t y_{t-1}+k_0\rho_0)^2}{\sum_{t=2}^{T}y_{t-1}^2+k_0}\\
&+\sum_{t=2}^{T}y_t^2+v_0\sigma_0^2+k_0\rho_0^2\\
&=k_n(\rho-\rho_n)^2-k_n\rho_n^2+\sum_{t=2}^{T}y_t^2+v_0\sigma_0^2+k_0\rho_0^2\\
&=k_n(\rho-\rho_n)^2+v_n(\frac{-k_n\rho_n^2+\sum_{t=2}^{T}y_t^2+v_0\sigma_0^2+k_0\rho_0^2}{v_n})\\
&=k_n(\rho-\rho_n)^2+v_n(\sigma_n^2)
\end{aligned}
\tag{12}
$$

Thus,

$$p(\rho,\sigma^2|y_{1:T})=(\sigma^2)^{-(\frac{v_n}{2}+1+\frac{1}{2})}\exp(-\frac{1}{2\sigma^2}(k_n(\rho-\rho_n)^2+v_n(\sigma_n^2)))=\mathrm{NIG}(\rho_n,\frac{\sigma_n^2}{k_n},v_n,\sigma_n^2)$$

b.) By properties of Normal-Inverse Gamma, full conditionals are easily seen.

$p(\rho|\sigma^2)=\mathrm{N}(\rho_n,\frac{\sigma^2}{k_n})$

$p(\sigma^2|\rho)=\mathrm{InvGam}(\frac{v_n+1}{2},\frac{1}{2}(k_n(\rho-\rho_n)^2+v_n\sigma_n^2))$

c.)

To obtain independent samples from the joint distribution of ρ and σ^2, we apply the Gibbs sampling algorithm.

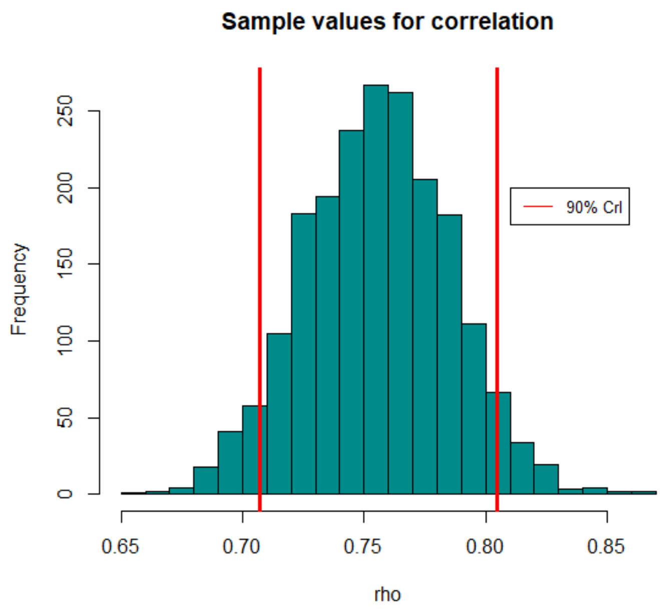
1. Initialize $\rho^{(0)}$ and $(\sigma^2)^{(0)}$ to some values. We simply initialize them to the sample correlation and sample variance. We also select the prior $v_0=1$,$k_0=5$ and σ_0^2 and ρ_0 to sample variance and correlation respectively.

2. At the nth iteration we sample $\rho^{(n)}$ from its full conditional $N(\rho_n,\frac{(\sigma^2)^{(n-1)}}{k_n})$ and $(\sigma^2)^{(n)}$ from its full conditional $\mathrm{InvGam}(\frac{v_n+1}{2},\frac{1}{2}(k_n(\rho^{(n)}-\rho_n)^2+v_n\rho_n^2))$ (Note that we use the most updated value of ρ to sample σ^2).

3. Repeat the second step N times (N is chosen such that the final samples after burned in and thinning is 2000).

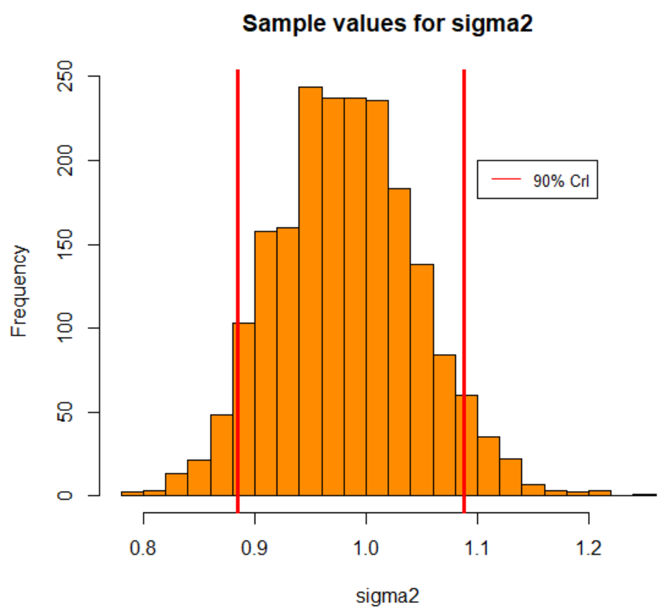
4. Burn and thin the samples. We choose to burn 10 percent of the samples and thin every 5 samples.

d.) Note that we are asked to report credible intervals and not credible regions.



90% Credible Interval

$$\rho \in (0.7069, 0.8046)$$



90% Credible Interval

$$\sigma^2 \in (0.8844, 1.0877)$$

5. Note that the derivation for full conditionals for this problem is a special case of number 6 (Please see number 6 for full derivation of the full conditionals) (also the formulas are derived in class).

We followed the following steps to obtain posterior densities for location-scale mixture model.

1. Initialize beginning values of the parameters and the prior hyperparameters.

$\pi^{(0)} = [1/K \ 1/K \ \dots \ 1/K]$

$z^{(0)} \sim \text{Multinom}(1, \pi)$

$\mu_k^{(0)} = \bar{y}$

$(\sigma^2)_k^{(0)} = \dfrac{s^2}{k}$

Prior Hyperparameters

$\mu_0 = \bar{y}$

$\sigma_0 = s^2$

$a_0 = 10$

$b_0 = 10$

2. At the nth step, we update the parameters according to the full conditionals.

We sample $\pi^{(n)}$ from the full conditional

$$p(\pi|-) = \text{Dir}(1 + n_1, 1 + n_2, \dots, 1 + n_K)$$

We sample $z^{(n)}$ from the full conditional

$$p(z_i = k|-) = \text{Mult}(1, \pi_{ik})$$

$$\pi_{ik} = \frac{\pi_k^{(n)} N(y_i|\mu_k^{(n-1)}, (\sigma_k^2)^{(n-1)})}{\sum_{j=1}^{K} \pi_j^{(n)} N(y_i|\mu_j^{(n-1)}, (\sigma_j^2)^{(n-1)})}$$

We sample $\mu_k^{(n)}$ from the full conditional

$$p(\mu_k|-) = N(\mu_{kn}, \sigma_{kn}^2)$$

$$\sigma_{kn}^2 = (\frac{1}{\sigma_0^2} + \frac{n_k}{(\sigma_k^2)^{(n-1)}})^{-1}$$

$$\mu_{kn} = \sigma_{kn}^2(\frac{\mu_0}{\sigma_0^2} + \frac{n_k \bar{y_k}}{(\sigma_k^2)^{(n-1)}})$$

$$\bar{y_k} = \frac{\sum_i \mathbf{I}(z_i = k) y_i}{n_k}$$

$$n_k = \sum_i \mathbf{I}(z_i = k)$$

We sample $(\sigma_k^2)^{(n)}$ from the full conditional

$$p(\sigma_k^2|-) = \text{InvGam}(a_{kn}, b_{kn})$$

$$a_{kn} = (a_0 + \frac{n_k}{2})$$

$$b_{kn} = b_0 + \frac{(n_k - 1)\sum_i \mathbf{I}(z_i = k)(y_i - \mu_k^{(n)})^2}{2} + \frac{n_k(\bar{y_k} - \mu_k^{(n)})^2}{2}$$

Please note that we always use the most updated values to samples from the full conditionals.

3. Repeat the second step N times (N is chosen such that we have the number of samples we want after burning and thinning)

4. Remove burned in samples and keep every 5 samples.

The plot below shows the posterior mean density and the 90 % credible interval.

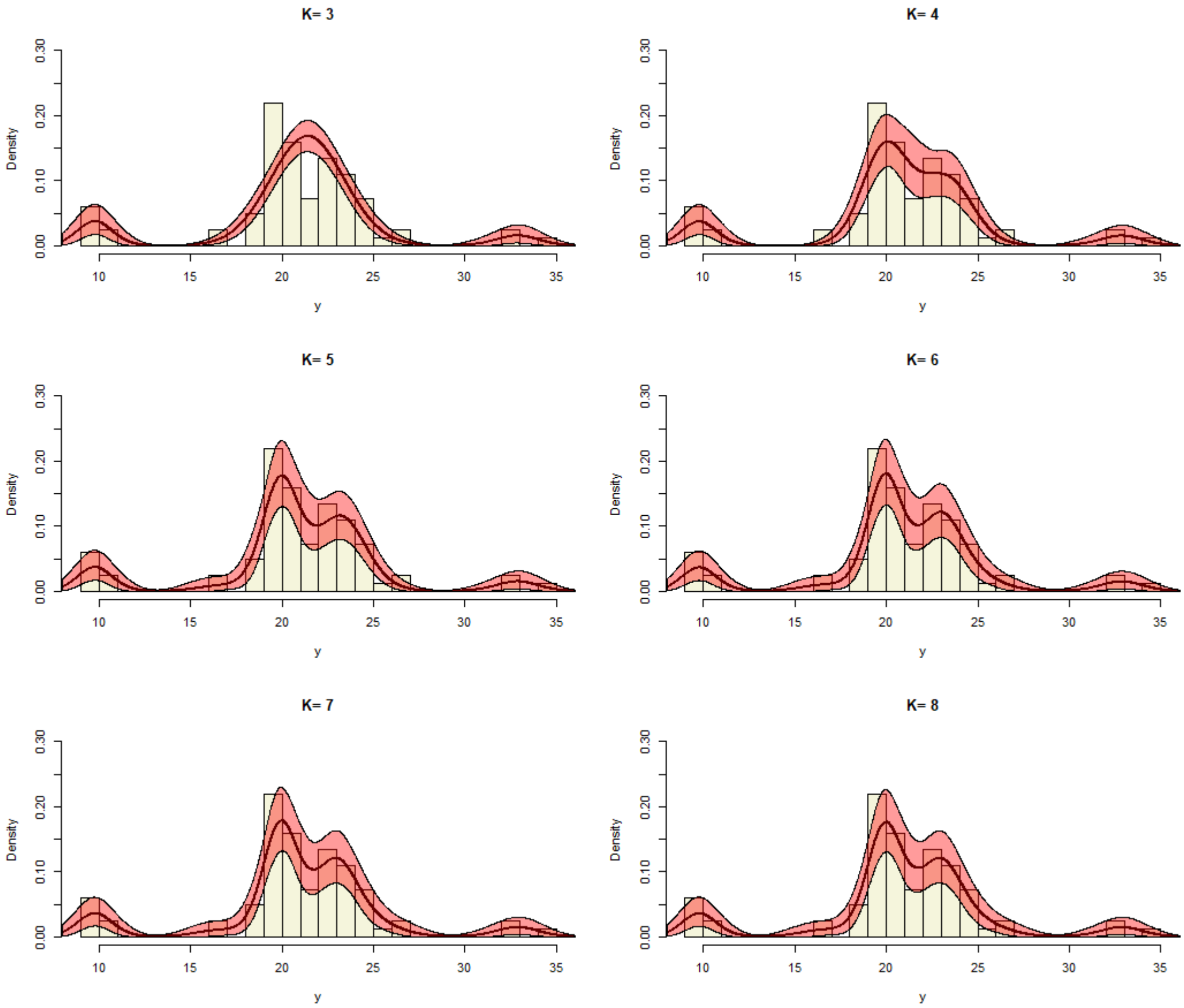


Figure 5: 90% point-wise credible interval and mean posterior density (black line)

6. We followed the following steps to obtain posterior densities for location-scale mixture model.

1. Initialize beginning values of the parameters and the prior hyperparameters.

$\pi^{(0)} = [1/K \ 1/K \ \dots \ 1/K]$

$z^{(0)} \sim \text{Multinom}(1, \pi)$

$\mu_k^{(0)} = [\bar{y} \ \bar{y} \ \dots \ \bar{y}]^T$

$(\Sigma)_k^{(0)} = \hat{\Sigma}$

Prior Hyperparameters

$\mu_0 = [\bar{y} \ \bar{y} \ \dots \ \bar{y}]^T$

$\Sigma_0 = \hat{\Sigma}$

$v_0 = 5$

$\Psi_0 = \hat{\Sigma}$

2. At the nth step, we update the parameters according to the full conditionals.

We sample $\pi^{(n)}$ from the full conditional

$$p(\pi|-) = \text{Dir}(1 + n_1, 1 + n_2, \dots, 1 + n_K)$$

We sample $z^{(n)}$ from the full conditional

$$p(z_i = k|-) = \text{Mult}(1, \pi_{ik})$$

$$\pi_{ik} = \frac{\pi_k^{(n)} N(y_i|\mu_k^{(n-1)}, (\Sigma_k^2)^{(n-1)})}{\sum_{j=1}^K \pi_j^{(n)} N(y_i|\mu_j^{(n-1)}, (\Sigma_j^2)^{(n-1)})}$$

We sample μ_k from the full conditional

$$p(\mu_k|-) = N(\tilde{\mu_k}, \tilde{\Sigma_k})$$

$$\tilde{\Sigma_k} = (\Sigma_0^{-1} + n_k(\Sigma_k^{-1})^{(n-1)})^{-1}$$

$$n_k = \sum_i \mathbf{I}(z_i = k)$$

$$\tilde{\mu_k} = \tilde{\Sigma_k}(\Sigma_0^{-1}\mu_0 + (\Sigma_k^{-1})^{(n-1)} \sum_i \mathbf{I}(z_i = k) y_i)$$

We sample Σ_k from the full conditional

$$p(\Sigma_k|-) = \text{InvWish}(v_0 + n_k, \Psi + C)$$

$$C = \sum_i \mathbf{I}(z_i = k)(y_i - \mu_k^{(n)})(y_i - \mu_k^{(n)})^T$$

$$n_k = \sum_i \mathbf{I}(z_i = k)$$

Please note that we always use the most updated values to samples from the full conditionals.

3. Repeat the second step N times (N is chosen such that we have the number of samples we want after burning and thinning)

4. Remove burned in samples and keep every 5 samples.

The plot below shows the posterior mean densities and the actual data points.

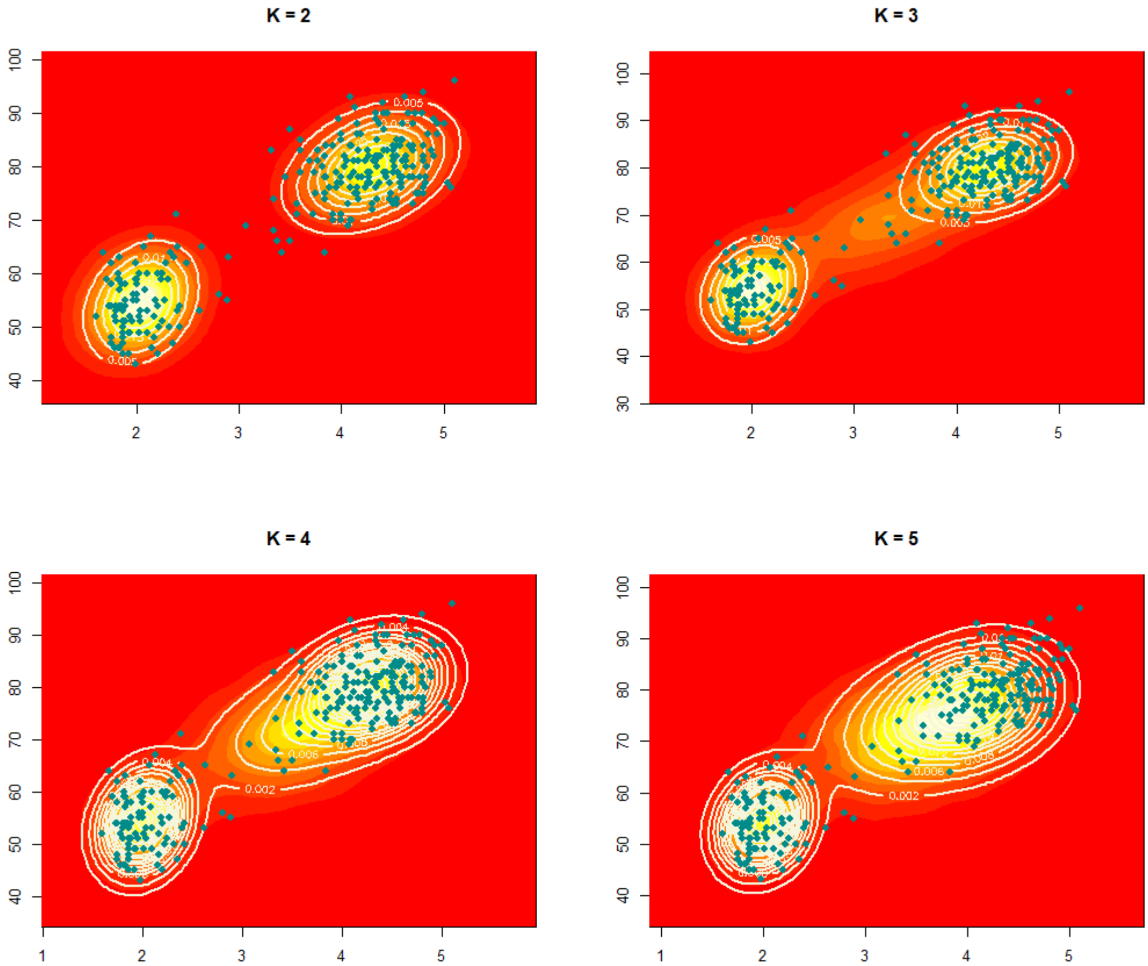


Figure 6: Posterior mean density superimposed with scatter plot of the data

Derivation for full conditionals. Here X_n denotes the nth data point.

$$
\begin{aligned}
p(\pi|-) &= p(\pi|z_1, z_2, \ldots, z_n) \\
&\propto p(z_1, z_2, \ldots, z_n|\pi)p(\pi) \\
&= \prod_{i=1}^{n} p(z_i|\pi)p(\pi|\alpha) \\
&\propto (\pi_1^{\mathbf{I}(z_1=1)}\pi_2^{\mathbf{I}(z_1=2)} \ldots \pi_K^{\mathbf{I}(z_1=K)}) \ldots (\pi_1^{\mathbf{I}(z_n=1)}\pi_2^{\mathbf{I}(z_n=2)} \ldots \pi_K^{\mathbf{I}(z_n=K)})\pi_1^{\alpha_1-1}\pi_2^{\alpha_2-1} \ldots \pi_K^{\alpha_K-1} \\
&= \pi_1^{\alpha_1+n_1-1}\pi_2^{\alpha_2+n_2-1} \ldots \pi_K^{\alpha_2+n_K-1} \\
&\propto \mathrm{Dir}(\alpha_1 + n_1, \alpha_2 + n_2, \ldots, \alpha_K + n_K)
\end{aligned}
\tag{13}
$$

$n_k = \sum_{i=1}^{n} \mathbf{I}(z_i = k)$

$$
\begin{aligned}
p(\mu_k|-) &= p(\mu_k|\mu_0, \Sigma_0, \Sigma_k, X_n : Z_n = k) \text{ (by conditional independence)} \\
&\propto p(X_n : Z_n = k|\mu_k, \Sigma_k)p(\mu_k|\mu_0, \Sigma_0) \\
&= \prod_{n:z_n=k} \exp\{-\frac{1}{2}(X_n - \mu_k)^T\Sigma_k^{-1}(X_n - \mu_k)\} \exp\{-\frac{1}{2}(\mu_k - \mu_0)^T\Sigma_0^{-1}(\mu_k - \mu_0)\} \\
&= \exp\{-\frac{1}{2}\sum_{n:Z_n=k}(X_n - \mu_k)^T\Sigma_k^{-1}(X_n - \mu_k)\} \exp\{-\frac{1}{2}(\mu_k - \mu_0)^T\Sigma_0^{-1}(\mu_k - \mu_0)\} \\
&\propto \exp\{-\frac{1}{2}(\sum_{n:z_n=k} X_n^T\Sigma_k^{-1}\mu_k - \mu_k^T\Sigma_k^{-1}\sum_{n:z_n=k} X_n + n_k\mu_k^T\Sigma_k^{-1}\mu_k \\
&+ \mu_k^T\Sigma_0^{-1}\mu_k - \mu_k^T\Sigma_0^{-1}\mu_0 - \mu_0^T\Sigma_0^{-1}\mu_k)\} \\
&= \exp\{-\frac{1}{2}(\mu_k^T\Sigma_k^{-1}\sum_{n:z_n=k} X_n - \mu_k^T\Sigma_k^{-1}\sum_{n:z_n=k} X_n + n_k\mu_k^T\Sigma_k^{-1}\mu_k \\
&+ \mu_k^T\Sigma_0^{-1}\mu_k - \mu_k^T\Sigma_0^{-1}\mu_0 - \mu_k^T\Sigma_0^{-1}\mu_0)\} \quad \text{(quadratic form is symmetric)} \\
&= \exp\{-\frac{1}{2}(\mu_k^T(\Sigma_0^{-1} + n_k\Sigma_k^{-1})\mu - 2\mu_k^T(\Sigma_k^{-1}\sum_{n:z_n=k} X_n + \Sigma_0^{-1}\mu_0))\} \\
&= \exp\{-\frac{1}{2}(\mu_k^T(\tilde{\Sigma_k}^{-1})\mu - 2\mu_k^T\tilde{\Sigma_k}^{-1}\tilde{\Sigma_k}(\Sigma_k^{-1}\sum_{n:z_n=k} X_n + \Sigma_0^{-1}\mu_0))\} \\
&= \exp\{-\frac{1}{2}(\mu_k^T(\tilde{\Sigma_k}^{-1})\mu - 2\mu_k^T\tilde{\Sigma_k}^{-1}\tilde{\mu_k})\} \\
&\propto \exp\{-\frac{1}{2}(\mu_k^T(\tilde{\Sigma_k}^{-1})\mu - 2\mu_k^T\tilde{\Sigma_k}^{-1}\tilde{\mu_k} + \tilde{\mu_k}^T\tilde{\Sigma_k}^{-1}\tilde{\mu_k}\} \\
&= \exp\{-\frac{1}{2}((\mu_k - \tilde{\mu_k})^T(\tilde{\Sigma_k}^{-1})(\mu_k - \tilde{\mu_k}))\} \\
&= N(\tilde{\mu_k}, \tilde{\Sigma_k})
\end{aligned}
\tag{14}
$$

where we replace $(\Sigma_0^{-1} + n_k\Sigma_k^{-1})$ by $\tilde{\Sigma_k}^{-1}$ and $\tilde{\Sigma_k}(\Sigma_k^{-1}\sum_{n:z_n=k} X_n + \Sigma_0^{-1}\mu_0)$ by $\tilde{\mu_k}$ so

$$\tilde{\Sigma_k} = (\Sigma_0^{-1} + n_k\Sigma_k^{-1})^{-1}$$

$$\tilde{\mu_k} = \tilde{\Sigma_k}(\Sigma_k^{-1} \sum_{n:z_n=k} X_n + \Sigma_0^{-1}\mu_0)$$

$$\begin{aligned}
p(\Sigma_k|-) &= p(\Sigma_k|\Psi_0, v_0, \mu_k, X_n : z_n = k) \\
&\propto \prod_{n:z_n=k} p(X_n|\Sigma_k, \mu_k) p(\Sigma_k|\Psi_0, v_0) \\
&= \frac{1}{|\Sigma_k|^{\frac{n_k}{2}}} \exp\{\sum_{n:z_n=k} -\frac{1}{2} tr[\Sigma_k^{-1}(X_n - \mu_k)(X_n - \mu_k)^T]\} \\
&|\Psi_0|^{\frac{v_0}{2}} |\Sigma_k|^{-\frac{v_0+p+1}{2}} \exp(-\frac{1}{2} tr(\Psi_0 \Sigma_k^{-1})) \\
&= \exp\{\sum_{n:z_n=k} -\frac{1}{2} tr[((X_n - \mu_k)(X_n - \mu_k)^T + \Psi_0)\Sigma_k^{-1}]\} \\
&|\Psi_0|^{\frac{v_0}{2}} |\Sigma_k|^{-\frac{n_k+v_0+p+1}{2}} \quad \text{tr(AB)=tr(BA) and tr is linear} \\
&= \text{InvWish}(n_k + v_0 \ , \ (X_n - \mu_k)(X_n - \mu_k)^T + \Psi_0)
\end{aligned} \tag{15}$$

$$\begin{aligned}
p(z_n = k|-) &= p(z_n = k|X_n, \mu, \Sigma) \\
&= \frac{p(z_n = k) p(X_n|\mu_k, \Sigma_k, z_n = k)}{\sum_{j=1}^K p(z_n = j) p(X_n|\mu_j, \Sigma_j, z_n = j)} \\
&= \frac{\pi_k N(X_n|\mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(X_n|\mu_j, \Sigma_j)}
\end{aligned} \tag{16}$$