Douglas Wei Jing Allwood (A0183939L)

ST4234

<u>Q1</u>   Note: I will use $\vec{}$ to denote vectors, e.g. $\vec{\theta} = (\theta_1, \ldots, \theta_J)$

(a) $P(\vec{\theta}, \mu, \tau \mid \vec{y}, \vec{\sigma}) \propto P(\vec{y} \mid \vec{\theta}, \vec{\sigma}) \cdot P(\vec{\theta} \mid \mu, \tau) \cdot P(\mu, \tau)$

$$\propto \prod_{j=1}^{J} \sigma_j^{-1} \exp\left\{ -\frac{(y_j - \theta_j)^2}{2\sigma_j^2} \right\} \tau^{-1} \exp\left\{ -\frac{(\theta_j - \mu)^2}{2\tau^2} \right\}$$

$$= \tau^{-J} \prod_{j=1}^{J} \sigma_j^{-1} \exp\left\{ -\left[\frac{(y_j - \theta_j)^2}{2\sigma_j^2} + \frac{(\theta_j - \mu)^2}{2\tau^2} \right] \right\}$$

(b) $P(\theta_j \mid \mu, \tau, \vec{y}, \vec{\sigma}) \propto \tau^{-1} \sigma_j^{-1} \exp\left\{ -\left[\frac{(y_j - \theta_j)^2}{2\sigma_j^2} + \frac{(\theta_j - \mu)^2}{2\tau^2} \right] \right\}$

$$\propto \exp\left\{ -\left[\frac{y_j^2 - 2\theta_j y_j + \theta_j^2}{2\sigma_j^2} + \frac{\theta_j^2 - 2\mu\theta_j + \mu^2}{2\tau^2} \right] \right\}$$

$$\propto \exp\left\{ -\left[\frac{\tau^2 y_j^2 - 2\tau^2\theta_j y_j + \tau^2\theta_j^2 + \sigma_j^2\theta_j^2 - 2\mu\sigma_j^2\theta_j + \sigma_j^2\mu^2}{2\sigma_j^2\tau^2} \right] \right\}$$

$$\propto \exp\left\{ -\left[\frac{(\tau^2 + \sigma_j)^2\theta_j^2 - 2(\tau^2 y_j + \mu\sigma_j^2)\theta_j}{2\sigma_j^2\tau^2} \right] \right\}$$

$$\propto \exp\left\{ -\left[\frac{\theta_j^2 - \frac{2(\tau^2 y_j + \mu\sigma_j^2)}{\tau^2 + \sigma_j^2}\theta_j}{2\frac{\sigma_j^2\tau^2}{\tau^2 + \sigma_j^2}} \right] \right\}$$

$$\propto \exp\left\{ -\frac{\left(\theta_j - \frac{\tau^2 y_j + \mu\sigma_j^2}{\tau^2 + \sigma_j^2}\right)^2}{2\frac{\sigma_j^2\tau^2}{\tau^2 + \sigma_j^2}} \right\}$$

$$\sim N\left(\frac{\tau^2 y_j + \mu\sigma_j^2}{\tau^2 + \sigma_j^2}, \quad \frac{\sigma_j^2\tau^2}{\tau^2 + \sigma_j^2} \right)$$

$$P(\theta_j \mid \mu, \tau, \vec{y}, \vec{\sigma}) \sim N\left(\frac{\tau^2 y_j + \mu \sigma_j^2}{\tau^2 + \sigma_j^2},\ \frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2}\right)$$

$$\text{Let } \hat{\theta}_j = \frac{\frac{y_j}{\sigma_j^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} = \left(\frac{\tau^2 y_j + \mu \sigma_j^2}{\sigma_j^2 \tau^2} \times \frac{\tau^2 \sigma_j^2}{\tau^2 + \sigma_j^2}\right)$$

$$= \frac{\tau^2 y_j + \mu \sigma_j^2}{\tau^2 + \sigma_j^2}$$

$$\text{Let } V_j = \frac{1}{\left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}\right)} = \frac{1}{\left(\frac{\tau^2 + \sigma_j^2}{\tau^2 \sigma_j^2}\right)} = \frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2}$$

$$\therefore P(\theta_j \mid \mu, \tau, \vec{y}, \vec{\sigma}) \sim N(\hat{\theta}_j, V_j)$$

$$P(\mu, \tau \mid \vec{y}, \vec{\sigma}) \propto \int P(\theta_j, \mu, \tau \mid \vec{y}, \vec{\sigma})\, d\theta_j$$

First find $P(\theta_j, \mu, \tau \mid \vec{y}, \vec{\sigma})$:

$$P(\theta_j, \mu, \tau \mid \vec{y}, \vec{\sigma}) \propto \tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{(y_j - \theta_j)^2}{\sigma_j^2} + \frac{(\theta_j - \mu)^2}{\tau^2}\right]\right\}$$

$$= \tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{y_j^2 - 2\theta_j y_j + \theta_j^2}{\sigma_j^2} + \frac{\theta_j^2 - 2\mu \theta_j + \mu^2}{\tau^2}\right]\right\}$$

$$= \tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{(\tau^2 + \sigma_j^2)\theta_j^2 - 2(y_j \tau^2 + \mu \sigma_j^2)\theta_j}{\tau^2 + \sigma_j^2} + \frac{\mu^2 \sigma_j^2 + \tau^2 y_j^2}{\tau^2 + \sigma_j^2}\right]\right\}$$

$$= \tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta_j - \frac{y_j \tau^2 + \mu \sigma_j^2}{\tau^2 + \sigma_j^2}\right)^2}{\left(\frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2}\right)}\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 \sigma_j^2 + \tau^2 y_j^2}{\tau^2 + \sigma_j^2} - \left(\frac{y_j \tau^2 + \mu \sigma_j^2}{\tau^2 + \sigma_j^2}\right)^2\right]\right\}$$

$$\int P(\theta_j, \mu, \tau \mid \vec{y}, \vec{\sigma})\, d\theta_j \propto \tau^{-1}\left(\frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 \sigma_j^2 + \tau^2 y_j^2}{\sigma_j^2 \tau^2} - \left(\frac{y_j \tau^2 + \mu \sigma_j^2}{\tau^2 + \sigma_j^2}\right)^2\right]\right\}$$

I'm integrating over this, the result is sqrt(var) V_j

$$\therefore P(\mu, \tau \mid \vec{y}, \vec{\sigma}) \propto \prod_{j=1}^{J} \tau^{-1}\left(\frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 \sigma_j^2 + \tau^2 y_j^2}{\tau^2 + \sigma_j^2} - \left(\frac{y_j \tau^2 + \mu \sigma_j^2}{\tau^2 + \sigma_j^2}\right)^2\right]\right\}$$

$$\propto \prod_{j=1}^{J} \phi\left(y_j \mid \mu, \sqrt{\sigma_j^2 + \tau^2}\right)$$

Douglas Allwood (A0183939L)

ST4234

<u>Q1(b) Part 2</u>

$$P(\mu,\tau|\vec{y},\vec{\sigma}) \propto \int\int\int\int\int\int\int\int P(\vec{\theta},\mu,\tau|\vec{y},\vec{\sigma})\, d\theta_1\, d\theta_2\, d\theta_3\, d\theta_4\, d\theta_5 d\theta_6 d\theta_7 d\theta_8$$

I will shorten the 8 integrals over θ_1 to θ_8 to just an integral over $\vec{\theta}$

$$P(\mu,\tau|\vec{y},\vec{\sigma}) \propto \int P(\vec{\theta},\mu,\tau|\vec{y},\vec{\sigma})\, d\vec{\theta}$$

$$\propto \int P(\vec{y}|\vec{\theta},\vec{\sigma}) \cdot P(\vec{\theta}|\mu,\tau) \cdot P(\mu,\tau)\, d\vec{\theta}$$

$$\propto \int \prod_{j=1}^{J} \sigma_j^{-1}\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{(y_j-\theta_j)^2}{\sigma_j^2} + \frac{(\theta_j-\mu)^2}{\tau^2}\right]\right\} d\vec{\theta}$$

$$\propto \int \prod_{j=1}^{J} \sigma_j\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\tau^2 y_j^2 - 2\tau^2 y_j\theta_j + \tau^2\theta_j^2 + \sigma_j^2\theta_j^2 - 2\sigma_j^2\theta_j\mu + \sigma_j^2\mu^2}{\sigma_j^2\tau^2}\right]\right\} d\vec{\theta}$$

$$\propto \int \prod_{j=1}^{J} \sigma_j\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{(\tau^2+\sigma_j^2)\theta_j^2 - 2(\tau^2 y_j + \sigma_j^2\mu)\theta_j + \tau^2 y_j^2 + \sigma_j^2\mu^2}{\sigma_j^2\tau^2}\right]\right\} d\vec{\theta}$$

$$= \int \prod_{j=1}^{J} \sigma_j\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\theta_j^2 - 2\frac{(\tau^2 y_j + \sigma_j^2\mu)}{\tau^2+\sigma_j^2}\theta_j + \frac{\tau^2 y_j^2 + \sigma_j^2\mu^2}{\tau^2+\sigma_j^2}}{\left(\frac{\sigma_j^2\tau^2}{\tau^2+\sigma_j^2}\right)}\right]\right\} d\vec{\theta}$$

$$= \int \prod_{j=1}^{J} \sigma_j\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta_j - \frac{\tau^2 y_j + \sigma_j^2\mu}{\tau^2+\sigma_j^2}\right)^2 - \left(\frac{\tau^2 y_j + \sigma_j^2\mu}{\tau^2+\sigma_j^2}\right)^2 + \frac{\tau^2 y_j^2 + \sigma_j^2\mu^2}{\tau^2+\sigma_j^2}}{\left(\frac{\sigma_j^2\tau^2}{\tau^2+\sigma_j^2}\right)}\right]\right\} d\vec{\theta}$$

$$= \int \prod_{j=1}^{J} \sigma_j\tau^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta_j - \frac{\tau^2 y_j + \sigma_j^2\mu}{\tau^2+\sigma_j^2}\right)^2}{\left(\frac{\sigma_j^2\tau^2}{\tau^2+\sigma_j^2}\right)}\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\frac{\tau^2 y_j^2 + \sigma_j^2\mu^2}{\tau^2+\sigma_j^2} - \left(\frac{\tau^2 y_j + \sigma_j^2\mu}{\tau^2+\sigma_j^2}\right)^2}{\left(\frac{\sigma_j^2\tau^2}{\tau^2+\sigma_j^2}\right)}\right]\right\} d\vec{\theta}$$

$$\propto \int \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_j^2\tau^2}} \exp\left\{-\frac{1}{2}\left[\frac{(\theta_j-\hat{\theta}_j)^2}{V_j}\right]\right\} \exp\left\{-\frac{1}{2}\left[\frac{\frac{\tau^2 y_j^2 + \mu^2\sigma_j^2}{\tau^2+\sigma_j^2} - \left(\frac{\tau^2 y_j + \sigma_j^2\mu}{\tau^2+\sigma_j^2}\right)^2}{\left(\frac{\sigma_j^2\tau^2}{\tau^2+\sigma_j^2}\right)}\right]\right\} d\vec{\theta}$$

Douglas Allwood (A0183939L)

ST4234

<u>Q16) part 3</u>

$P(\mu, \tau \mid \vec{y}, \vec{\sigma})$ &

$$\propto \left[\prod_{j=1}^{J} \left[\frac{1}{\sqrt{2\pi(\tau^2 + \sigma_j^2)}} \exp\left\{ -\frac{1}{2} \left[\frac{\frac{\tau^2 y_j^2 + \mu^2 \sigma_j^2}{\tau_j^2 + \sigma_j^2} - \frac{\tau^4 y_j^2 + 2\tau^2 y_j \sigma_j^2 \mu + \sigma_j^4 \mu^2}{(\tau_j^2 + \sigma_j^2)^2}}{\left(\frac{\sigma_j^2 \tau^2}{\tau^2 + \sigma_j^2} \right)} \right] \right\} \right.$$

$$\left. \times \int \frac{1}{\sqrt{2\pi \left(\frac{\tau^2 \sigma_j^2}{\tau^2 + \sigma_j^2} \right)}} \exp\left\{ -\frac{1}{2} \frac{(\theta_j - \hat{\theta}_j)^2}{V_j} \right\} d\theta \right]$$

$$\propto \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi(\tau^2 + \sigma_j^2)}} \exp\left\{ -\frac{1}{2} \left[\frac{\tau^2 \mu^2 \sigma_j^2 + \sigma_j^2 \tau^2 y_j^2 - 2\tau^2 y_j \mu \sigma_j^2}{\sigma_j^2(\tau^2 + \sigma_j^2)\tau^2} \right] \right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi(\tau^2 + \sigma_j^2)}} \exp\left\{ -\frac{1}{2} \left[\frac{\mu^2 + y_j^2 - 2\mu y_j}{\tau^2 + \sigma_j^2} \right] \right\}$$

$$\propto \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi(\tau^2 + \sigma_j^2)}} \exp\left\{ -\frac{1}{2} \frac{(y_j - \mu)^2}{\tau^2 + \sigma_j^2} \right\}$$

$$\propto \prod_{j=1}^{J} \phi\left(y_j \mid \mu, \sqrt{\tau^2 + \sigma_j^2} \right)$$

Douglas Wei Jing Allwood (A0183939L)

ST4234

Q2

(a) $P(\vec{\theta}, \mu, \tau^2, \sigma^2 | \vec{y}) \propto P(\vec{y} | \vec{\theta}, \sigma^2, \mu, \tau^2) \cdot P(\vec{\theta}, \sigma^2, \mu, \tau^2)$

$\propto P(\vec{y} | \vec{\theta}, \sigma^2) \cdot P(\vec{\theta} | \mu, \tau^2) \cdot P(\sigma^2) \cdot P(\mu) \cdot P(\tau^2)$

$\propto (\sigma^2)^{-\frac{a_2}{2} - 1} e^{-\frac{b_2}{2(\sigma^2)}} \cdot (\tau^2)^{-\frac{a_1}{2} - 1} e^{-\frac{b_1}{2(\tau^2)}} \cdot \exp\left\{ -\frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right\}$

$\times \prod_{j=1}^{J} \left[\tau^{-1} \exp\left\{ -\frac{(\theta_j - \mu)^2}{2\tau^2} \right\} \right] \times \prod_{j=1}^{J} \prod_{i=1}^{n_j} \left[\sigma^{-1} \exp\left\{ -\frac{(y_{i,j} - \theta_j)^2}{2\sigma^2} \right\} \right]$

$\propto (\sigma^2)^{-\frac{a_2}{2} - 1} (\tau^2)^{-\frac{a_1}{2} - 1} \exp\left\{ -\left(\frac{b_2}{2\sigma^2} + \frac{b_1}{2\tau^2} + \frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right) \right\}$

$\times \prod_{j=1}^{J} \left[\tau^{-1} \exp\left\{ \frac{(\theta_j - \mu)^2}{2\tau^2} \right\} \prod_{i=1}^{n_j} \left[\sigma^{-1} \exp\left\{ -\frac{(y_{i,j} - \theta_j)^2}{2\sigma^2} \right\} \right] \right]$

For clarity:

$P(\vec{\theta}, \mu, \tau^2, \sigma^2 | \vec{y}) \propto (\sigma^2)^{-\frac{a_2}{2} - 1} (\tau^2)^{-\frac{a_1}{2} - 1} \exp\left\{ -\left(\frac{b_2}{2\sigma^2} + \frac{b_1}{2\tau^2} + \frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right) \right\}$

$\times \prod_{j=1}^{J} \left[\tau^{-1} \exp\left\{ -\frac{(\theta_j - \mu)^2}{2\tau^2} \right\} \prod_{i=1}^{n_j} \sigma^{-1} \exp\left\{ -\frac{(y_{i,j} - \theta_j)^2}{2\sigma^2} \right\} \right]$

Douglas Allwood
ST4234
Q2(b) 1st part

(b) $P(\theta_j | \mu, \tau^2, \sigma^2, \vec{y}) \propto \exp\left\{-\frac{(\theta_j - \mu)^2}{2\tau^2}\right\} \prod_{i=1}^{n_j} \exp\left\{-\frac{(y_{i,j} - \theta_j)^2}{2\sigma^2}\right\}$

$\propto \exp\left\{-\frac{(\theta_j - \mu)^2}{2\tau^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_j}(y_{i,j} - \theta_j)^2\right\}$

$= \exp\left\{-\frac{1}{2}\left[\frac{(\theta_j - \mu)^2}{\tau^2} + \frac{\sum_{i=1}^{n_j}(y_{i,j} - \theta_j)^2}{\sigma^2}\right]\right\}$

$= \exp\left\{-\frac{1}{2}\left[\frac{\theta_j^2 - 2\mu\theta_j + \mu^2}{\tau^2} + \frac{\sum_{i=1}^{n_j}\left[y_{i,j}^2 - 2\theta_j y_{i,j} + \theta_j^2\right]}{\sigma^2}\right]\right\}$

$\propto \exp\left\{-\frac{1}{2}\left[\frac{\theta_j^2 - 2\mu\theta_j}{\tau^2} + \frac{n_j\theta_j^2 - 2\theta_j n_j \bar{y}_j}{\sigma^2}\right]\right\}$

$\propto \exp\left\{-\frac{1}{2}\left[\frac{\sigma^2\theta_j^2 - 2\mu\sigma^2\theta_j + n_j\tau^2\theta_j^2 - 2\theta_j n_j \bar{y}_j \tau^2}{\tau^2\sigma^2}\right]\right\}$

$\propto \exp\left\{-\frac{1}{2}\left[\frac{(\sigma^2 + n_j\tau^2)\theta_j^2 - 2(\mu\sigma^2 + n_j\bar{y}_j\tau^2)\theta_j}{\tau^2\sigma^2}\right]\right\}$

$\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta_j - \frac{\mu\sigma^2 + n_j\bar{y}_j\tau^2}{\sigma^2 + n_j\tau^2}\right)^2}{\left(\frac{\tau^2\sigma^2}{\sigma^2 + n_j\tau^2}\right)}\right]\right\}$

$\therefore \theta_j | \mu, \tau^2, \sigma^2, \vec{y} \sim N\left(\frac{n_j\tau^2}{n_j\tau^2 + \sigma^2}\bar{y}_j + \frac{\sigma^2}{n_j\tau^2 + \sigma^2}\mu, \frac{\tau^2\sigma^2}{\sigma^2 + n_j\tau^2}\right)$

Douglas Allwood
ST4234
Q2(b). 2nd part

$$p(\mu \mid \vec{\theta}, \tau^2, \sigma^2, \vec{y}) \propto \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\} \prod_{j=1}^{J} \exp\left\{-\frac{(\theta_j-\mu)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{\mu^2-2\mu\mu_0}{\sigma_0^2} + \frac{\sum_{j=1}^{J}(\mu^2-2\mu\theta_j)}{\tau^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{\tau^2\mu^2-2\mu\mu_0\tau^2 + \sigma_0^2 J\mu^2 - 2\mu\sigma_0^2\bar{\theta}J}{\sigma_0\tau^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{(\tau^2+\sigma_0^2 J)\mu^2-2(\mu_0\tau^2+\sigma_0^2\bar{\theta}J)\mu}{\sigma_0\tau^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\mu - \frac{\mu_0\tau^2+\sigma_0^2\bar{\theta}J}{\tau^2+\sigma_0^2 J}\right)^2}{\left(\frac{\sigma_0\tau^2}{\tau^2+\sigma_0^2 J}\right)}\right]\right\}$$

$$\therefore \mu \mid \vec{\theta}, \tau^2, \sigma^2, \vec{y} \sim N\left(\frac{\mu_0\tau^2+\sigma_0^2\bar{\theta}J}{\tau^2+\sigma_0^2 J}, \frac{\sigma_0\tau^2}{\tau^2+\sigma_0^2 J}\right)$$

$$p(\tau^2 \mid \vec{\theta}, \mu, \sigma^2, \vec{y}) \propto (\tau^2)^{-\frac{a_1}{2}-1} \exp\left\{-\frac{b_1}{2\tau^2}\right\} \times \prod_{j=1}^{J} \tau^{-1} \exp\left\{-\frac{(\theta_j-\mu)^2}{2\tau^2}\right\}$$

$$\propto (\tau^2)^{-\left(\frac{a_1}{2}+\frac{J}{2}\right)-1} \exp\left\{-\frac{1}{2}\left[\frac{b_1}{\tau^2} + \frac{\sum_{j=1}^{J}(\theta_j-\mu)^2}{\tau^2}\right]\right\}$$

$$\propto (\tau^2)^{-\left(\frac{a_1}{2}+\frac{J}{2}\right)-1} \exp\left\{-\frac{1}{2}\left[\frac{b_1+\sum_{j=1}^{J}(\theta_j-\mu)^2}{\tau^2}\right]\right\}$$

$$\therefore \tau^2 \mid \vec{\theta}, \mu, \sigma^2, \vec{y} \sim \text{Inv-Gamma}\left(\frac{a_1+J}{2}, \frac{b_1+\sum_{j=1}^{J}(\theta_j-\mu)^2}{2}\right)$$

Douglas Allwood
ST4234

Q2(b) 3rd Part

$$P(\sigma^2 \mid \vec{\theta}, \mu, \tau^2, \vec{y}) \propto (\sigma^2)^{-\frac{a_2}{2}-1} \exp\left\{-\frac{b_2}{2\sigma^2}\right\} \prod_{j=1}^{J} \prod_{i=1}^{n_j} \sigma^{-1} \exp\left\{-\frac{(y_{i,j}-\theta_j)^2}{2\sigma^2}\right\}$$

$$\propto (\sigma^2)^{-\left(\frac{a_2}{2}+\frac{\sum_{j=1}^{J} n_j}{2}\right)-1} \exp\left\{-\frac{1}{2}\left[\frac{b_2}{\sigma^2} + \frac{\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{\sigma^2}\right]\right\}$$

$$\propto (\sigma^2)^{-\left(\frac{a_2+\sum_{j=1}^{J} n_j}{2}\right)-1} \exp\left\{-\left[\frac{b_2+\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{2}\right](\sigma^2)^{-1}\right\}$$

$$\therefore \quad \sigma^2 \mid \vec{\theta}, \mu, \tau^2, \vec{y} \sim \text{Inv-gamma}\left(\frac{a_2+\sum_{j=1}^{J} n_j}{2}, \frac{b_2+\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{i,j}-\theta_j)^2}{2}\right)$$

ST4234 Data Analysis Project

Douglas Wei Jing Allwood (A0183939L)

Question 1

```r
# Load Libraries
library(LearnBayes)
library(coda)
library(invgamma)
```

1(c)

```r
set.seed(0183939)

y <- c(28, 8, -3, 7, -1, 1, 18, 12)
sigma <- c(15, 10, 16, 11, 9, 11, 10, 18)

logpost <- function(lambda, y, sigma) {
  mu <- lambda[1]
  tau <- exp(lambda[2])
  sd <- sqrt(sigma^2 + tau^2)

  sum(
    dnorm(y, mean = mu, sd = sd, log = TRUE)
  ) + log2(tau)  # Jacobian
}

lambda1.grid <- seq(from = -18, to = 37, by = 0.1)
lambda2.grid <- seq(from = -6, to = 4.1, by = 0.1)
grid_c <- expand.grid(lambda1.grid, lambda2.grid)

logpost.grid <- apply(grid_c, 1, logpost, y = y, sigma = sigma)
contour(x = lambda1.grid, y = lambda2.grid,
        z = matrix(logpost.grid, nrow = length(lambda1.grid),
                   ncol = length(lambda2.grid)),
        nlevels = 40,
        main = "Logposterior contour plot",
        xlab = expression(lambda[1]),
        ylab = expression(lambda[2]))
```

Logposterior contour plot
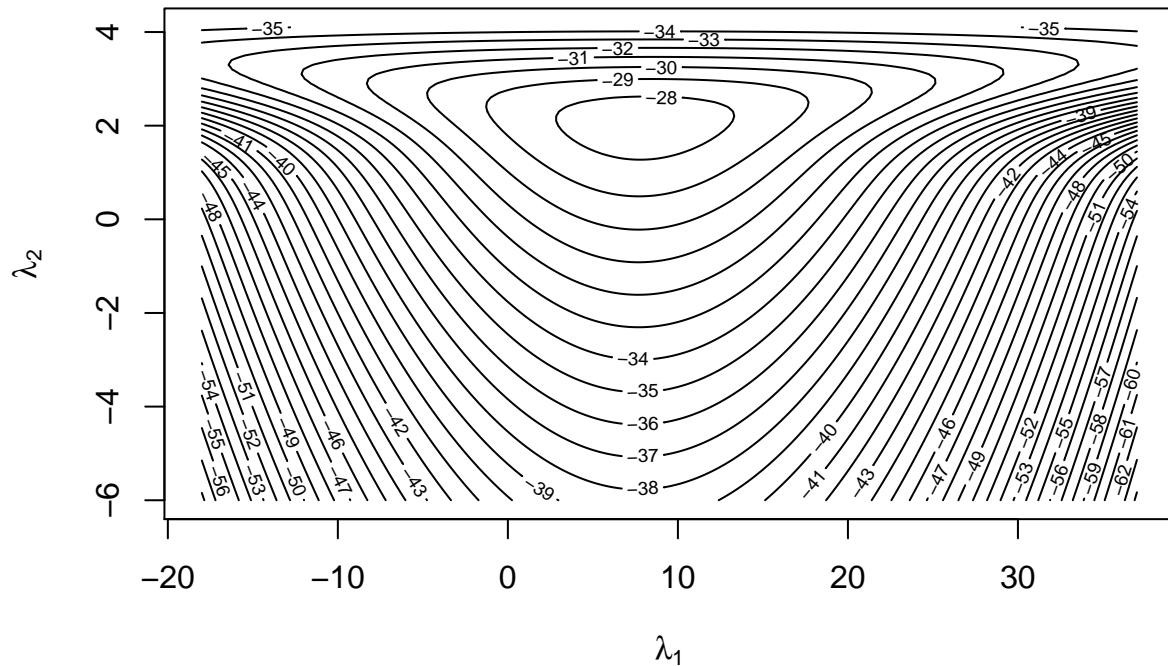

```
set.seed(0183939)
### Find the normal approximation to the posterior of (lambda1, lambda2)
out <- optim(par = c(9, 2), fn = logpost, hessian = TRUE,
             control = list(fnscale=-1), y = y, sigma = sigma)
post.mode <- out$par
post.cov <- -solve(out$hessian)

post.mode
```

```
## [1] 8.002639 2.044262
```

```
post.cov
```

```
##             [,1]       [,2]
## [1,] 25.0770290 0.1722124
## [2,]  0.1722124 0.3989440
```

Hence, the Normal Approximation of the posterior of (λ_1, λ_2) is :

$$\mu, \tau \,|\mathbf{y}, \sigma \sim N(\begin{pmatrix} 8.002639 \\ 2.044262 \end{pmatrix}, \begin{pmatrix} 25.0770290 & 0.1722124 \\ 0.1722124 & 0.3989440 \end{pmatrix})$$

1(d)

```
set.seed(0183939)

proposal <- list(var = post.cov, scale = 2)
```

```
fit1 <- rwmetrop(logpost, proposal, post.mode, 1e4, y = y, sigma = sigma)


lambda1.grid <- seq(from = -18, to = 37, by = 0.1)
lambda2.grid <- seq(from = -6, to = 4.1, by = 0.1)
grid_c <- expand.grid(lambda1.grid, lambda2.grid)

logpost.grid <- apply(grid_c, 1, logpost, y = y, sigma = sigma)
contour(x = lambda1.grid, y = lambda2.grid,
        z = matrix(logpost.grid, nrow = length(lambda1.grid),
                   ncol = length(lambda2.grid)),
        nlevels = 40,
        main = "Logposterior contour plot with samples overlayed",
        xlab = expression(lambda[1]),
        ylab = expression(lambda[2]))
points(fit1$par[5001:1e4,])
```

Logposterior contour plot with samples overlayed
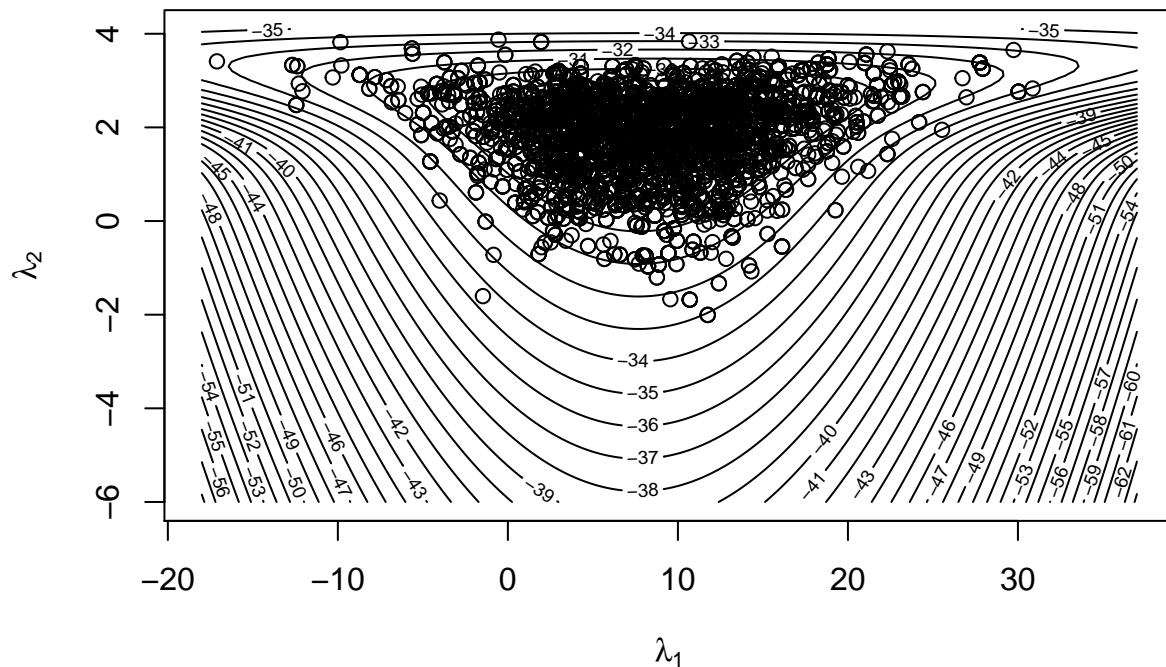

```
fit1$accept
```

```
## [1] 0.3207
```

The acceptance rate of the random walk Metropolis algorithm is 0.3207.

This value is between 20% and 50% and is hence acceptable for a random walk Metropolis algorithm.

```
set.seed(0183939)

mcmcobj1 <- mcmc(fit1$par[5001:1e4,1])
```

```
mcmcobj2 <- mcmc(fit1$par[5001:1e4,2])

par(mfrow = c(1,1), mar = c(5,4,2,2))
traceplot(mcmcobj1)
title("Traceplots for samples from posterior of Lambda1")
```

Traceplots for samples from posterior of Lambda1
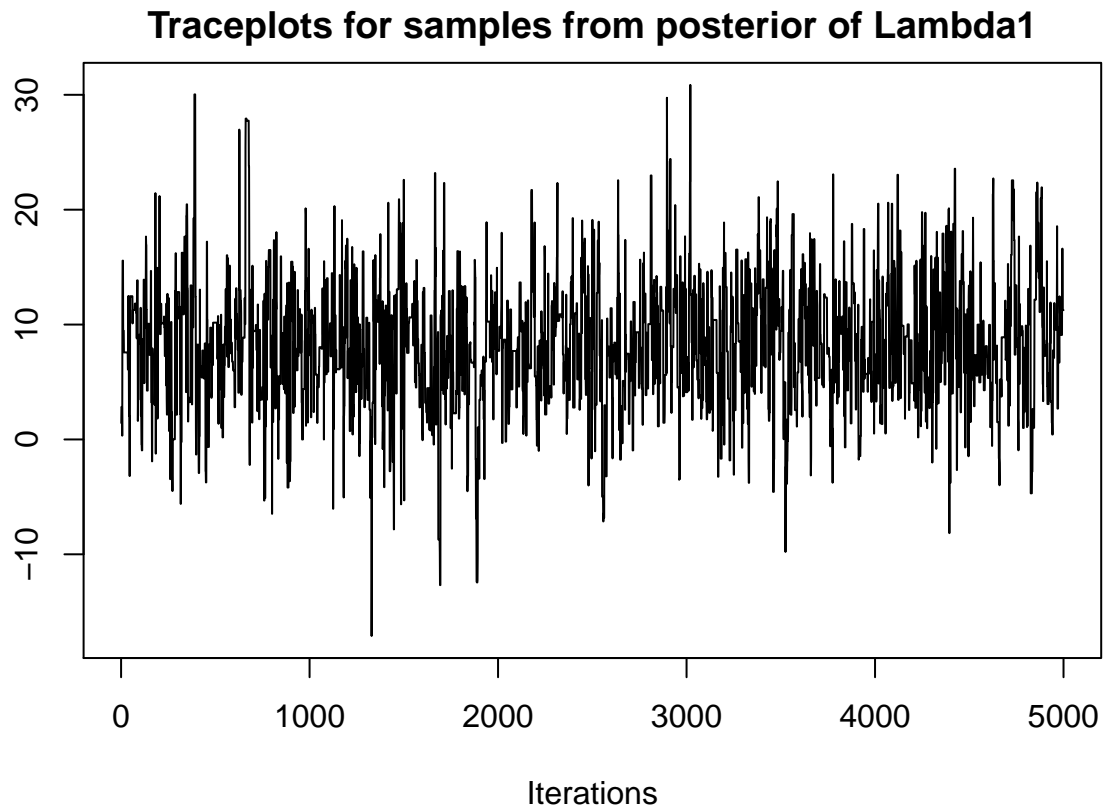

```
par(mfrow = c(1,1), mar = c(5,4,2,2))
traceplot(mcmcobj2)
title("Traceplots for samples from posterior of Lambda2")
```

Traceplots for samples from posterior of Lambda2

Iterations
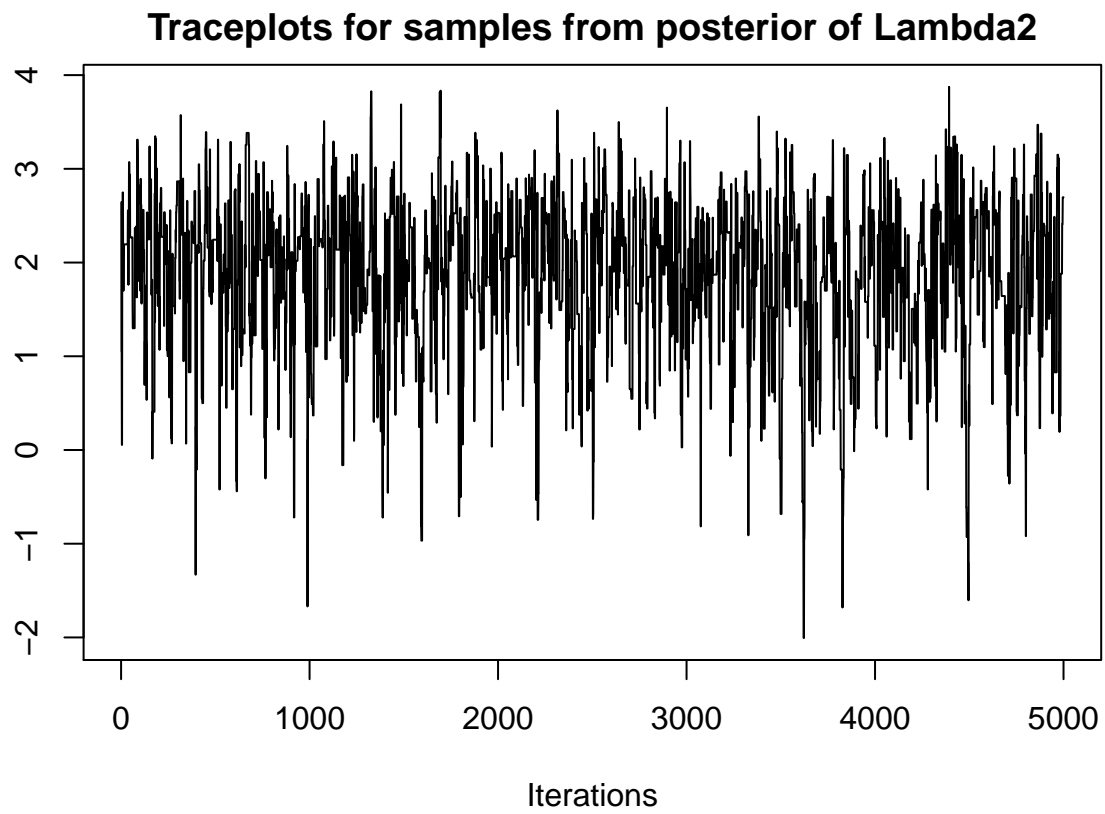
```r
autocorr.plot(mcmcobj1)
title(expression(lambda[1]))
```

λ_1

Autocorrelation

Lag
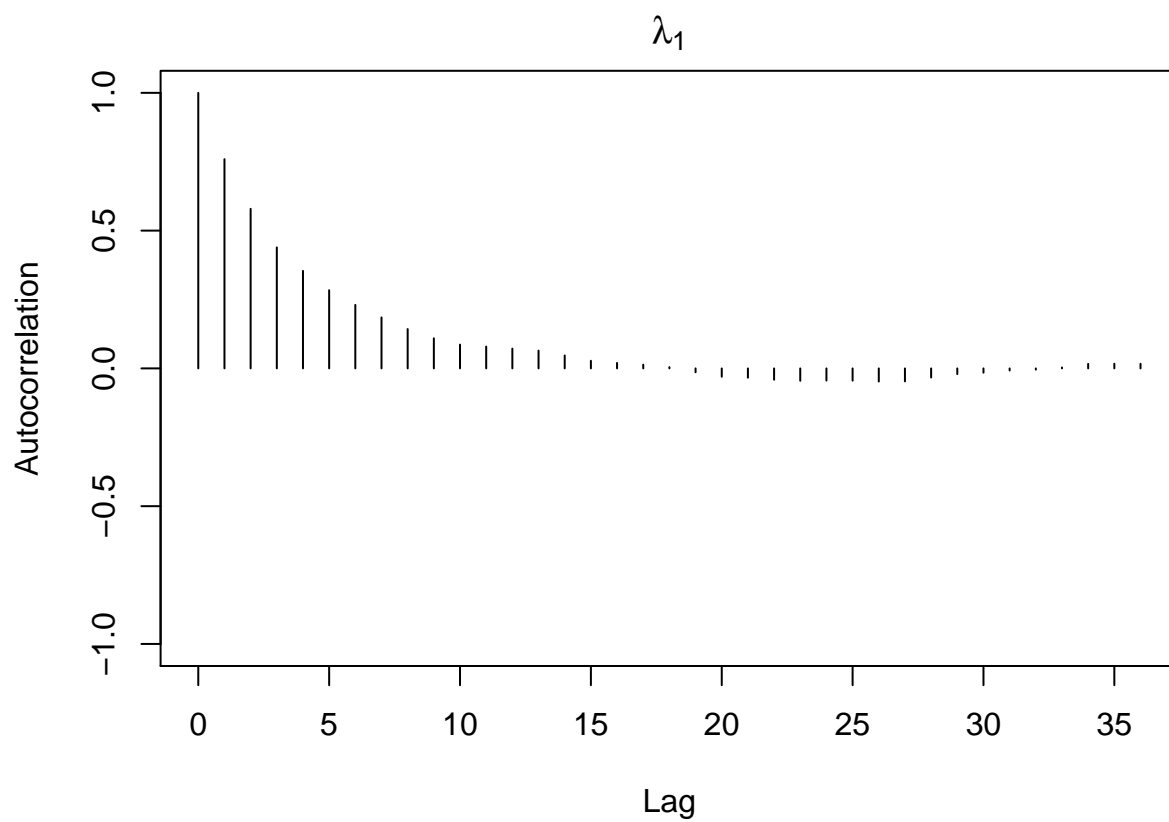
```r
autocorr.plot(mcmcobj2)
title(expression(lambda[2]))
```

λ_2
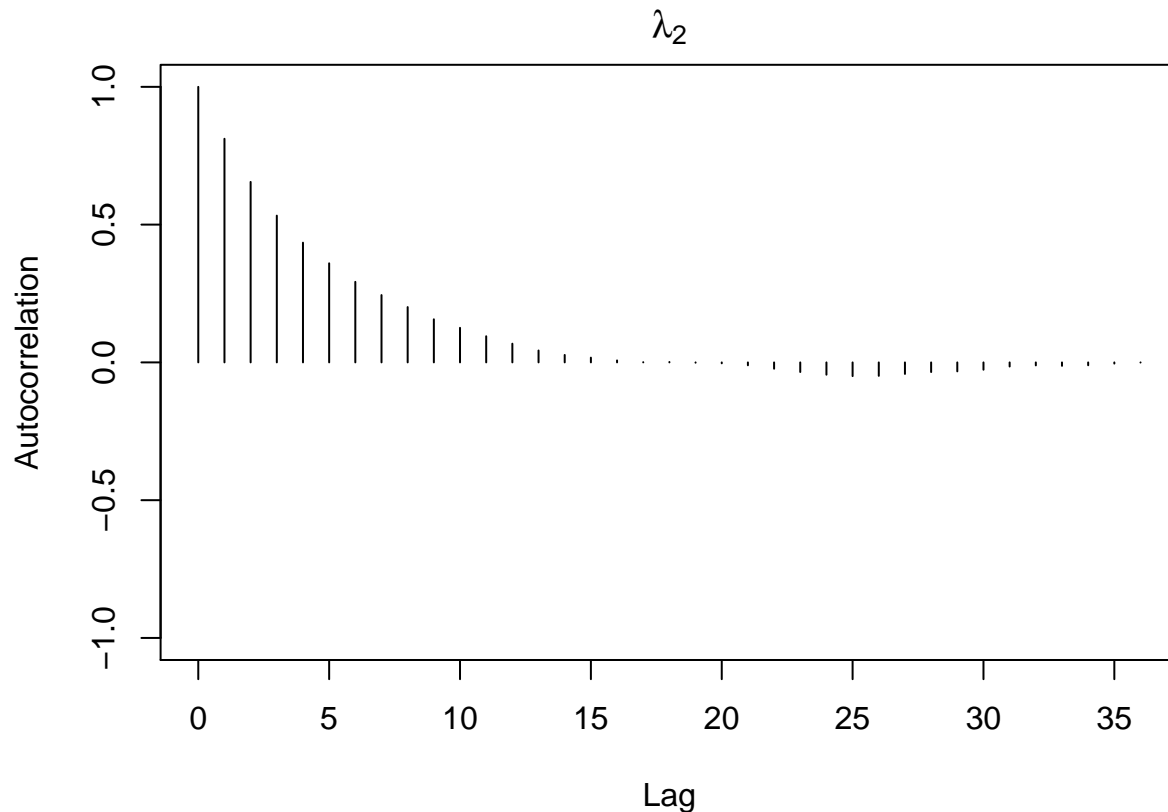
Above, we see that the traceplots both show random movements about the same range, indicating that a stationary distribution was likely to have been reached.

The contour plot with sampled points overlayed shows good mixing of the Metropolis algorithm. This combined with the acceptance rate of about 30% indicate that the covariance of the proposal is appropriate.

The autocorrelation plots show that autocorrelation decay takes about 15 samples, this is acceptable. However, if we want to reduce the autocorrelation in sampled data we could select every 15th point from the sampled data.

Overall, these results show us that the Markov Chain has likely converged to the stationary distribution and that mixing was good and little autocorrelation is present. Hence, the samples generated are appropriate.

1(e)

```r
set.seed(0183939)

mu <- fit1$par[5001:1e4,1]
tau <- exp(fit1$par[5001:1e4,2])

theta.samples <- matrix(0, nrow = 5000, ncol = 8)
for (i in 1:5000) {
  theta.hat <- ((y/sigma^2) + (mu[i]/tau[i]^2)) / ((1/sigma^2) + (1/tau[i]^2))
  v <- 1 / ((1/sigma^2) + (1/tau[i]^2))
  theta.samples[i,] <- rnorm(8, mean = theta.hat, sd = sqrt(v))
}
```

```
means <- apply(theta.samples, 2, mean)
sds <- apply(theta.samples, 2, sd)

## Report results
for (i in 1:8) {
  cat("Theta", i, "\tmean:", means[i], "\tstandard deviation:", sds[i], "\n")
}
```

```
## Theta 1  mean: 13.12913  standard deviation: 9.545343
## Theta 2  mean: 8.08929   standard deviation: 6.905731
## Theta 3  mean: 5.536705  standard deviation: 8.473469
## Theta 4  mean: 7.668079  standard deviation: 7.170273
## Theta 5  mean: 4.508205  standard deviation: 6.733229
## Theta 6  mean: 5.695994  standard deviation: 7.386265
## Theta 7  mean: 11.66636  standard deviation: 7.277727
## Theta 8  mean: 8.975628  standard deviation: 9.017865
```

1(f)(i)

```
set.seed(0183939)

b_coef <- function(tau, sigma) {
  return(tau^(-2) / (tau^(-2) + sigma^(-2)))
}

b.values <- numeric(8)
for (i in 1:8) {
  sig <- sigma[i]
  b.values[i] <- mean(b_coef(tau, sig))
}

for (i in 1:8) {
  cat("B_", i, ": ", b.values[i], "\n", sep = "")
}
```

```
## B_1: 0.7549597
## B_2: 0.6282458
## B_3: 0.7728801
## B_4: 0.6597959
## B_5: 0.5925985
## B_6: 0.6597959
## B_7: 0.6282458
## B_8: 0.8036067
```

1(f)(ii)

```
set.seed(0183939)

df <- data.frame(b = b.values, school = c('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'))
df[order(df$b, decreasing = TRUE),]
```

```
##           b school
## 8 0.8036067      H
## 3 0.7728801      C
```

```
## 1 0.7549597        A
## 4 0.6597959        D
## 6 0.6597959        F
## 2 0.6282458        B
## 7 0.6282458        G
## 5 0.5925985        E
```

The order of shrinkage from largest to smallest amongst the schools is:
H, C, A, D, F, B, G, E

The reason for the observed shrinkage values is due to the standard error of each observation. Observations with lower standard errors (like E, G, D, F) are observations whereby we are more certain of the mean, hence shrinkage is less pronounced for these observations. However, for observations with large standard error (e.g. H, A), there is less certainty about the sample mean and hence more strength is borrowed from the overall mean through the larger shrinkage value.

Hence, Shrinkage compensates for the lack of certainty by shrinking a sample mean closer to the overall mean when certainty of the value is low.

1(g)

```r
set.seed(0183939)

probabilities <- numeric(7)
for (i in 2: 8) {
  probabilities[i-1] <- mean(theta.samples[,1] > theta.samples[,i])
  cat("P(theta_1 > theta_", i, ") = ", probabilities[i-1], "\n", sep="")
}
```

```
## P(theta_1 > theta_2) = 0.6724
## P(theta_1 > theta_3) = 0.7266
## P(theta_1 > theta_4) = 0.6752
## P(theta_1 > theta_5) = 0.7678
## P(theta_1 > theta_6) = 0.7324
## P(theta_1 > theta_7) = 0.5518
## P(theta_1 > theta_8) = 0.6302
```

Quick note

Dear prof, apologies for the poor formatting (i.e. 1 graph per page), i am not too familiar with R graph margins.
I hope it is not too much of an annoyance.

Question 2

```r
set.seed(0183939)
# Load in the data
y <- list()
for (i in 1:8) {
  file <- paste0("../data/school", i, ".txt")
  y[[i]] <- as.vector(as.matrix(read.table(file)))
}
```

2(c)

```r
set.seed(0183939)
mu0 <- 7
sigma0sq <- 5
a1 <- 10
b1 <- 20
a2 <- 15
b2 <- 30
J <- 8
size <- 1e4

# vector of numbers representing number of samples in each dataset
n <- sapply(y, length)
n.sum <- sum(n)

# vector of numbers representing mean of samples in each dataset
y.bar <- sapply(y, mean)

# Storage for samples
theta <- matrix(0, nrow = size, ncol = 8)
mu <- numeric(size)
tausq <- numeric(size)
sigmasq <- numeric(size)

# Initial samples / "Current" values
theta.curr <- numeric(8)
mu.curr <- rnorm(1, mean = mu0, sd = sqrt(sigma0sq))
tausq.curr <- rinvgamma(1, shape = a1/2, rate = b1/2)
sigmasq.curr <- rinvgamma(1, shape = a2/2, rate = b2/2)

##### Gibbs Sampling
for (i in 1:size) {
  ### Sample Thetas
  # n.mean is a vector of 8 means for the 8 thetas to be drawn
  n.mean <- (n*tausq.curr*y.bar + sigmasq.curr*mu.curr) / (n*tausq.curr + sigmasq.curr)
```

```r
  # n.var is a vector of 8 vars for the 8 thetas to be drawn
  n.var <- (tausq.curr * sigmasq.curr) / (sigmasq.curr + n*tausq.curr)

  theta.curr <- rnorm(8, mean = n.mean, sd = sqrt(n.var))
  theta[i,] <- theta.curr


  ### Sample Mus
  # n.mean is a scalar mean for the mu to be drawn
  n.mean2 <- (mu0 * tausq.curr + sigma0sq * sum(theta.curr)) / (tausq.curr + sigma0sq * J)

  # n.mean is a scalar var for the mu to be drawn
  n.var2 <- (sqrt(sigma0sq) * tausq.curr) / (tausq.curr + sigmasq.curr * J)

  mu.curr <- rnorm(1, mean = n.mean2, sd = sqrt(n.var2))
  mu[i] <- mu.curr


  ### Sample Taus
  alpha <- (a1 + J)/2
  beta <- (b1 + sum((theta.curr - mu.curr)^2)) / 2

  tausq.curr <- rinvgamma(1, shape = alpha, rate = beta)
  tausq[i] <- tausq.curr


  ### Sample Sigma Squares
  alpha2 <- (a2 + n.sum) / 2

  beta_sum <- 0
  for (j in 1:J) {
    sch <- y[[j]]
    beta_sum <- beta_sum + sum((sch - theta.curr[j])^2)

  }

  beta2 <- (b2 + beta_sum) / 2
  sigmasq.curr <- rinvgamma(1, shape = alpha2, rate = beta2)
  sigmasq[i] <- sigmasq.curr
}

set.seed(0183939)
# Assess Gibbs samples
theta.mc <- mcmc(theta[5001:size,])
mu.mc <- mcmc(mu[5001:size])
tausq.mc <- mcmc(tausq[5001:size])
sigmasq.mc <- mcmc(sigmasq[5001:size])

for (i in 1:J) {
  traceplot(mcmc(theta[5001:size,i]))
  title(paste("Traceplot of theta ", i))
}
```

Traceplot of theta 1
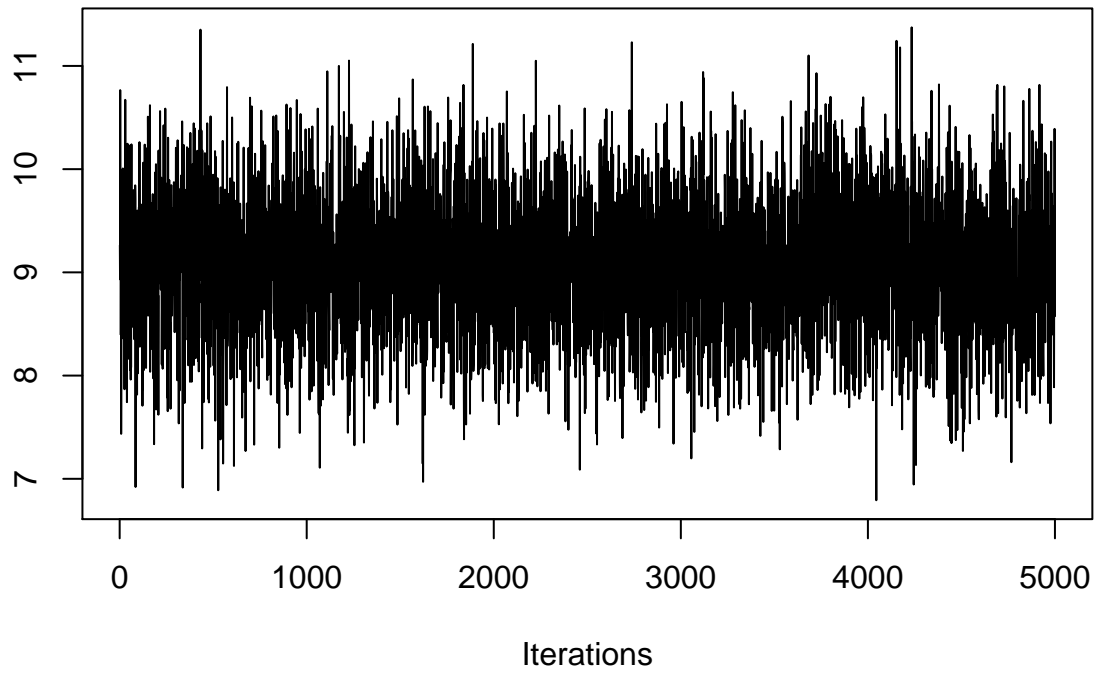
Iterations

Traceplot of theta 2
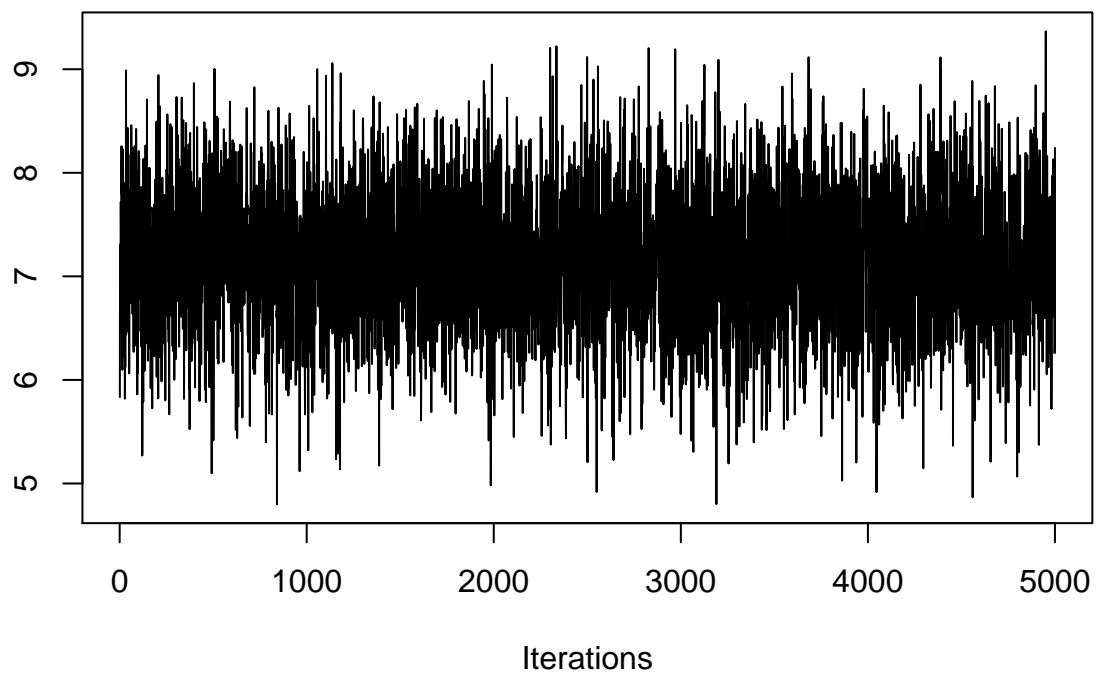
Iterations

Traceplot of theta 3
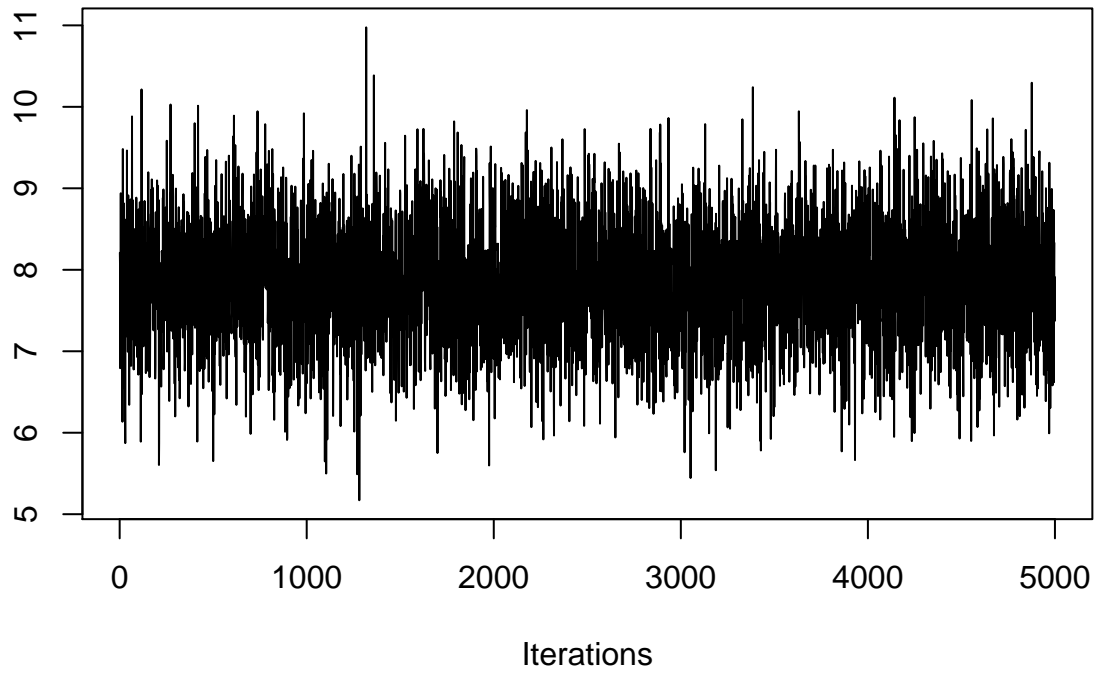
Iterations

Traceplot of theta 4

Iterations

Traceplot of theta 5

Iterations

Traceplot of theta 6

Iterations

Traceplot of theta 7

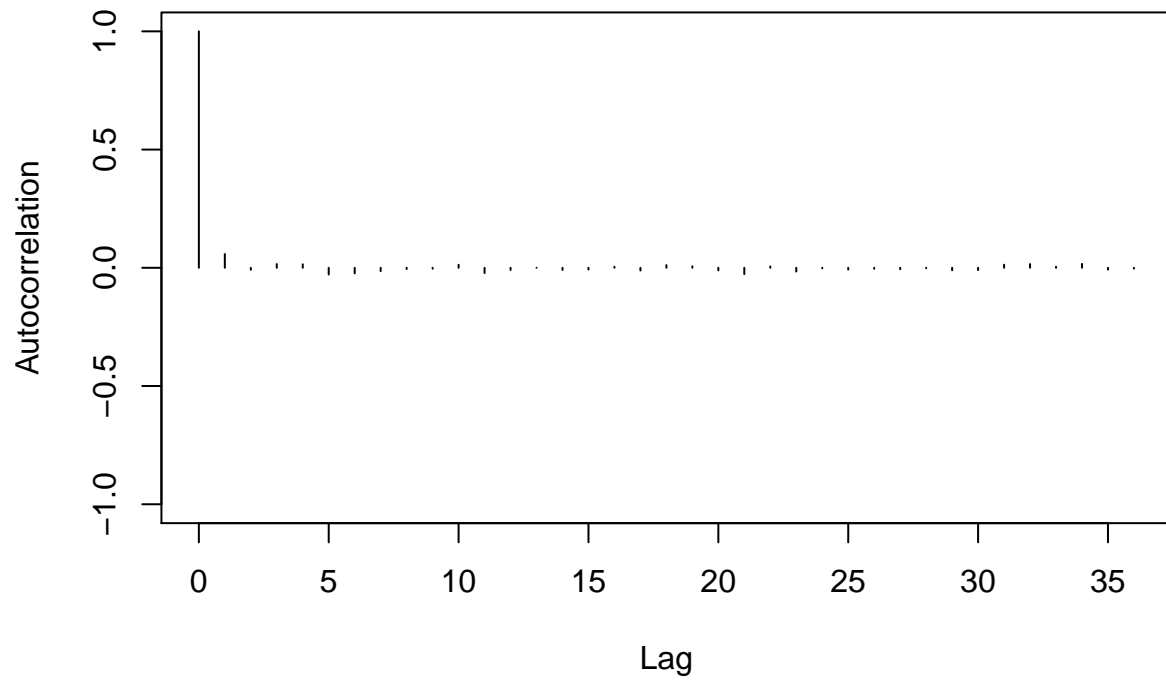Iterations
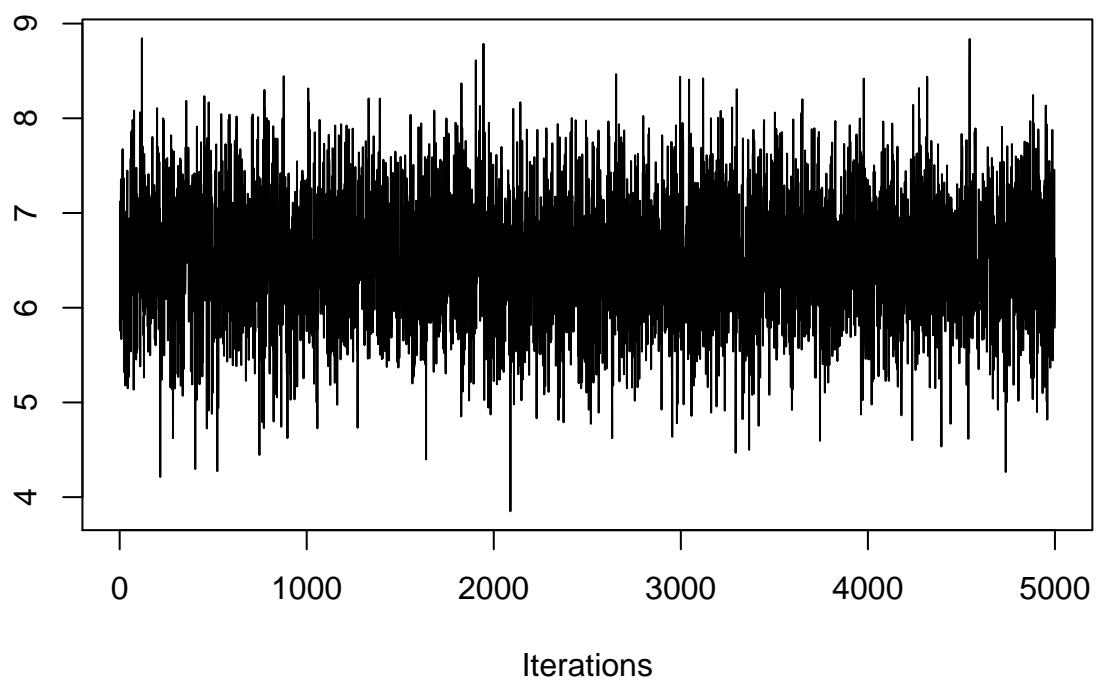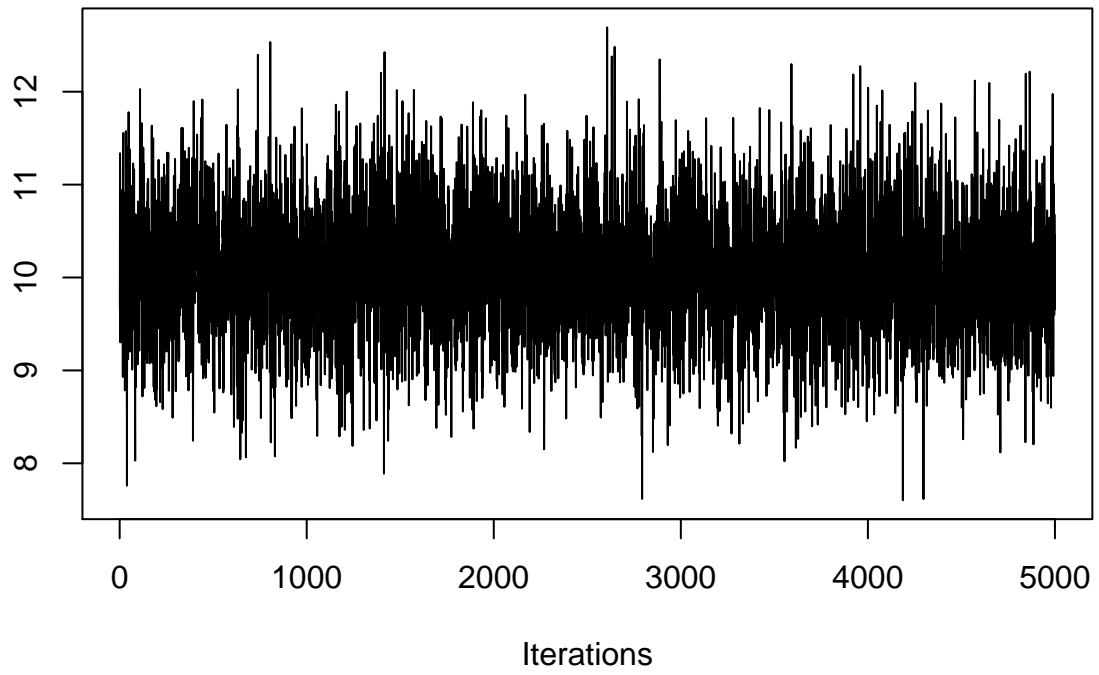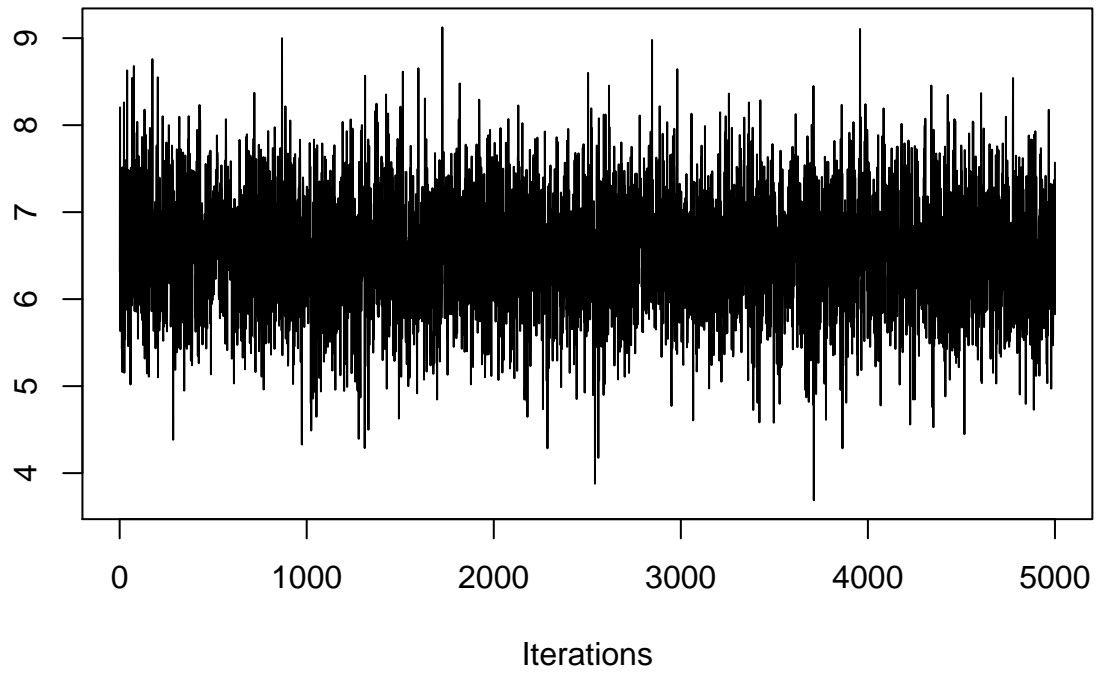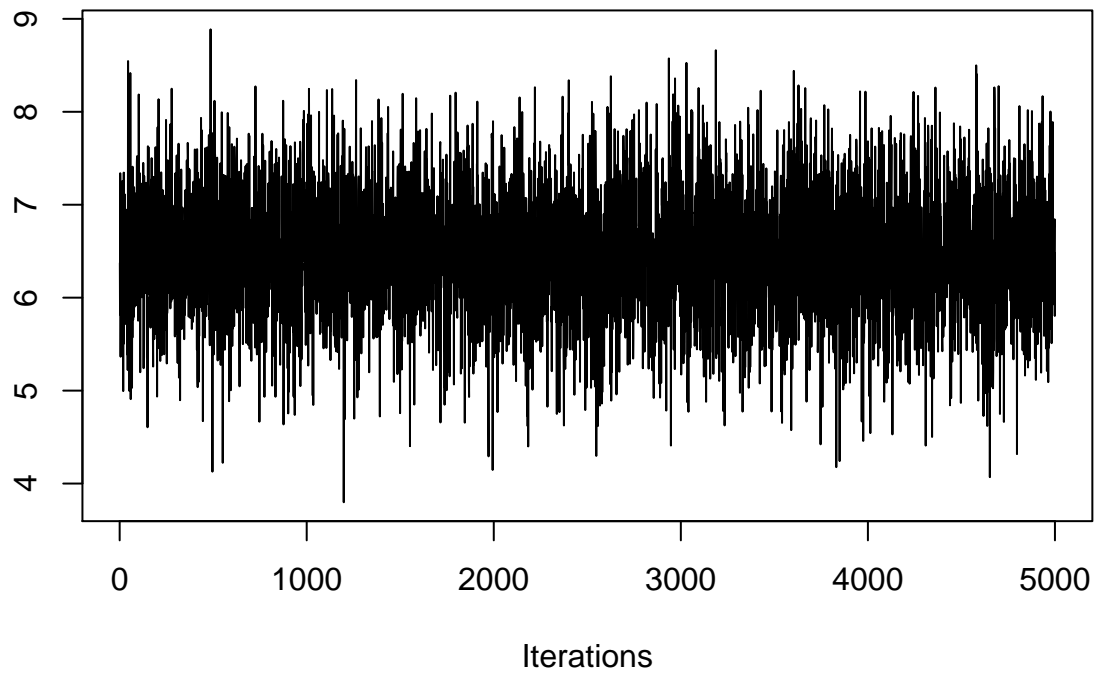
Traceplot of theta 8

Iterations
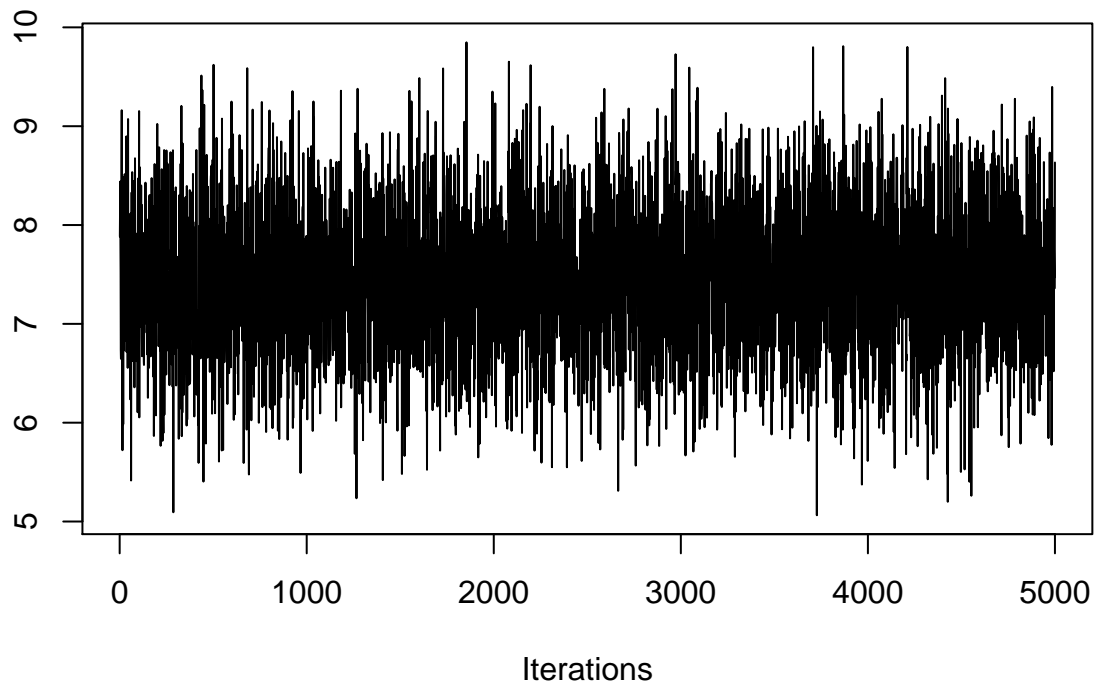
```r
for (i in 1:J) {
  autocorr.plot(mcmc(theta[5001:size,i]))
  title(paste("Autocorrplot of theta ", i))
}
```
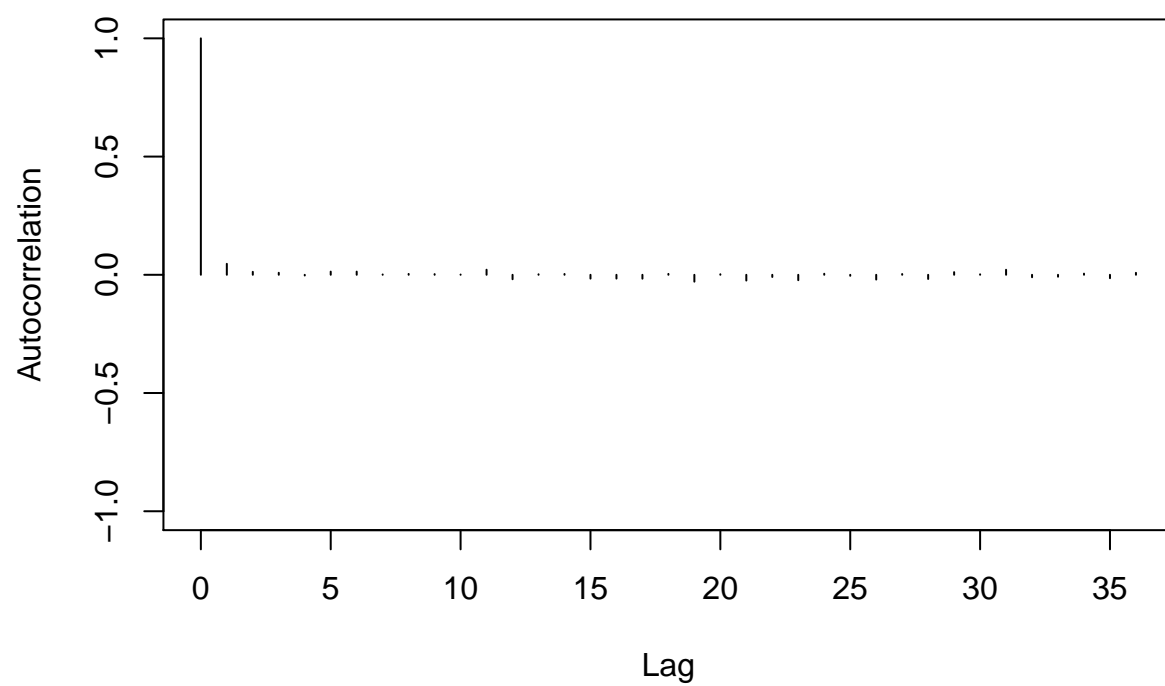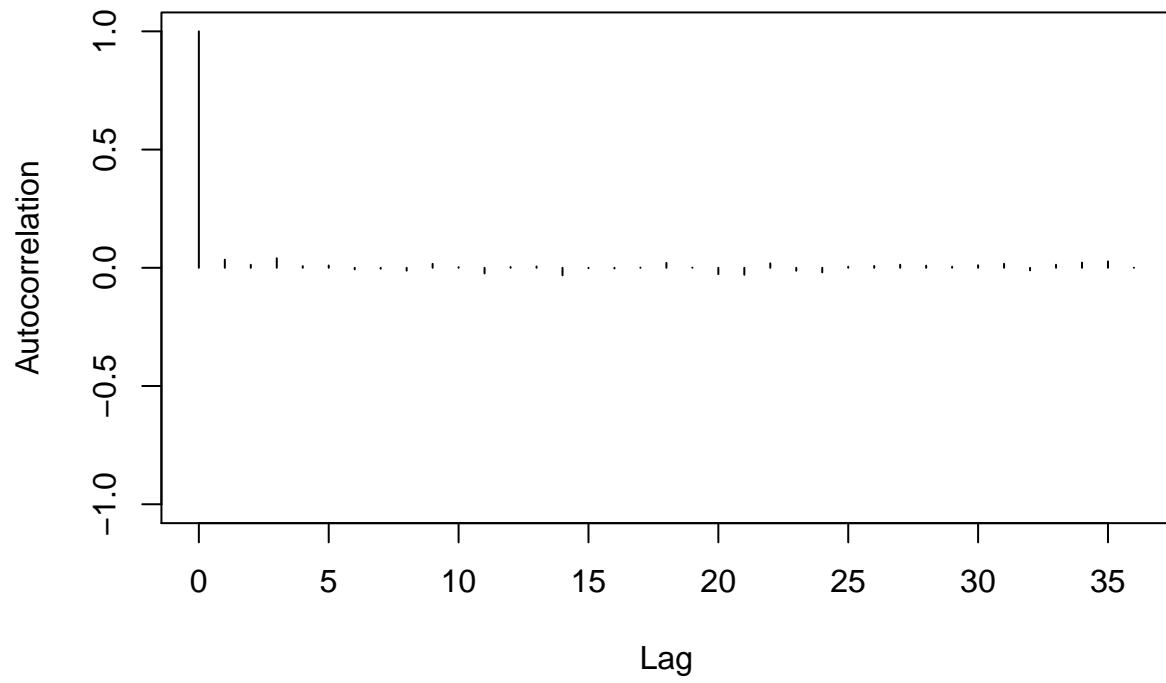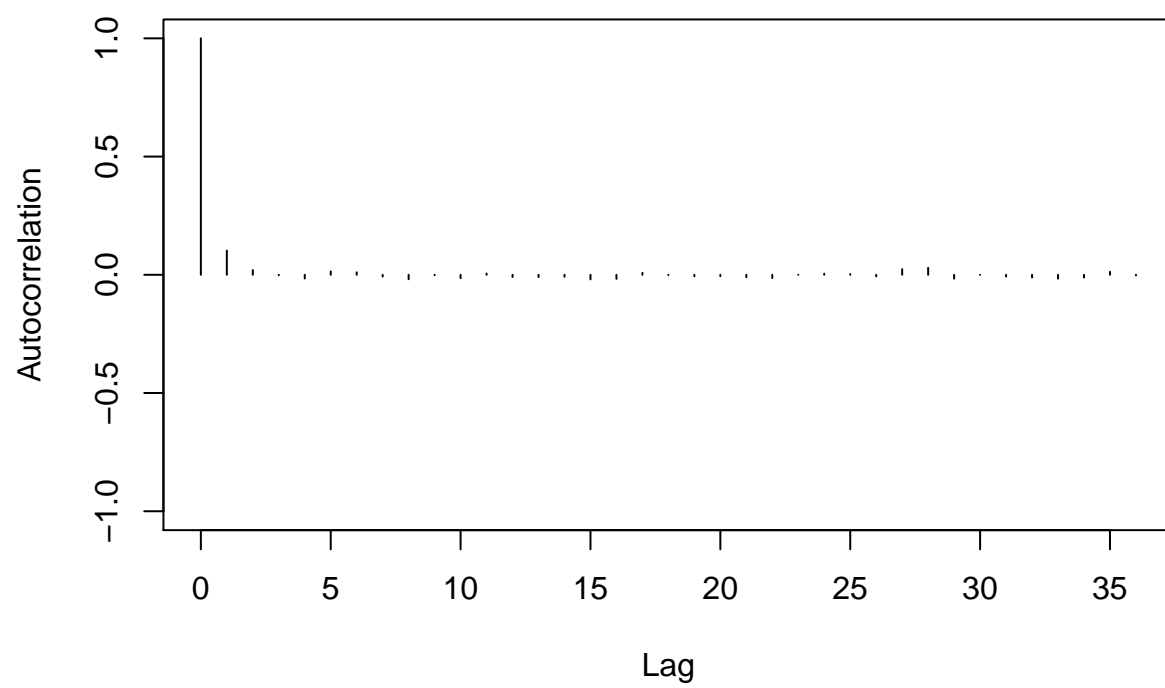
Autocorrplot of theta 1
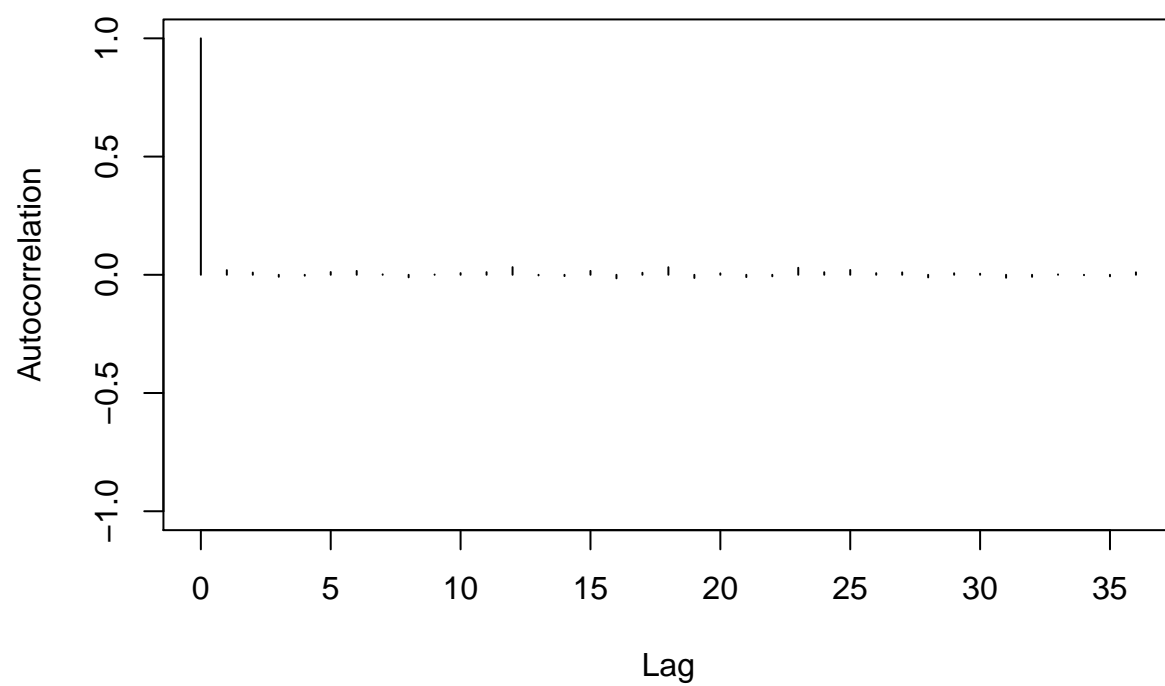
Autocorrplot of theta 2

Autocorrplot of theta 3

Autocorrplot of theta 4

Autocorrplot of theta 5

Autocorrplot of theta 6
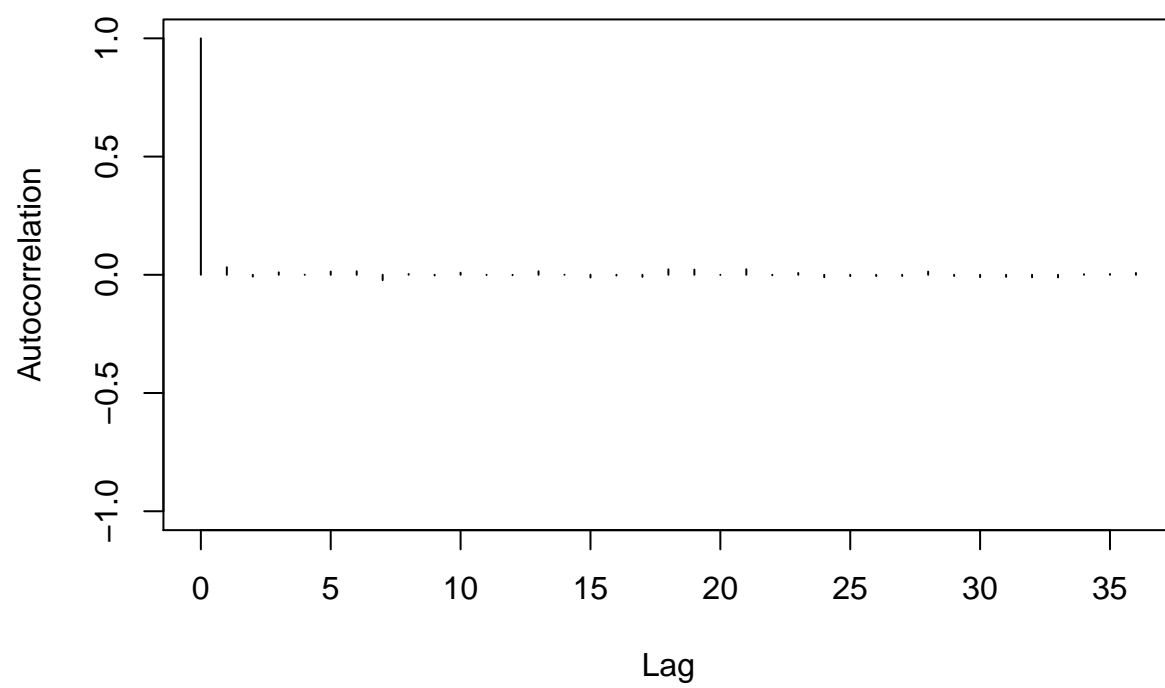
Autocorrplot of theta 7
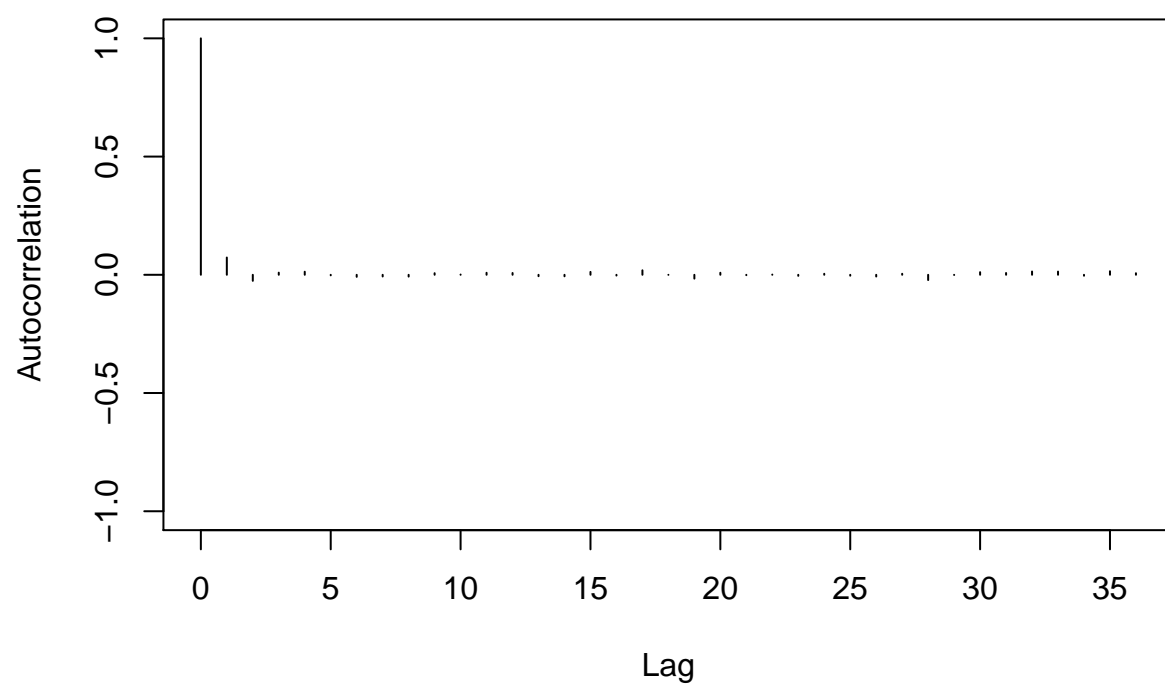
Autocorrplot of theta 8
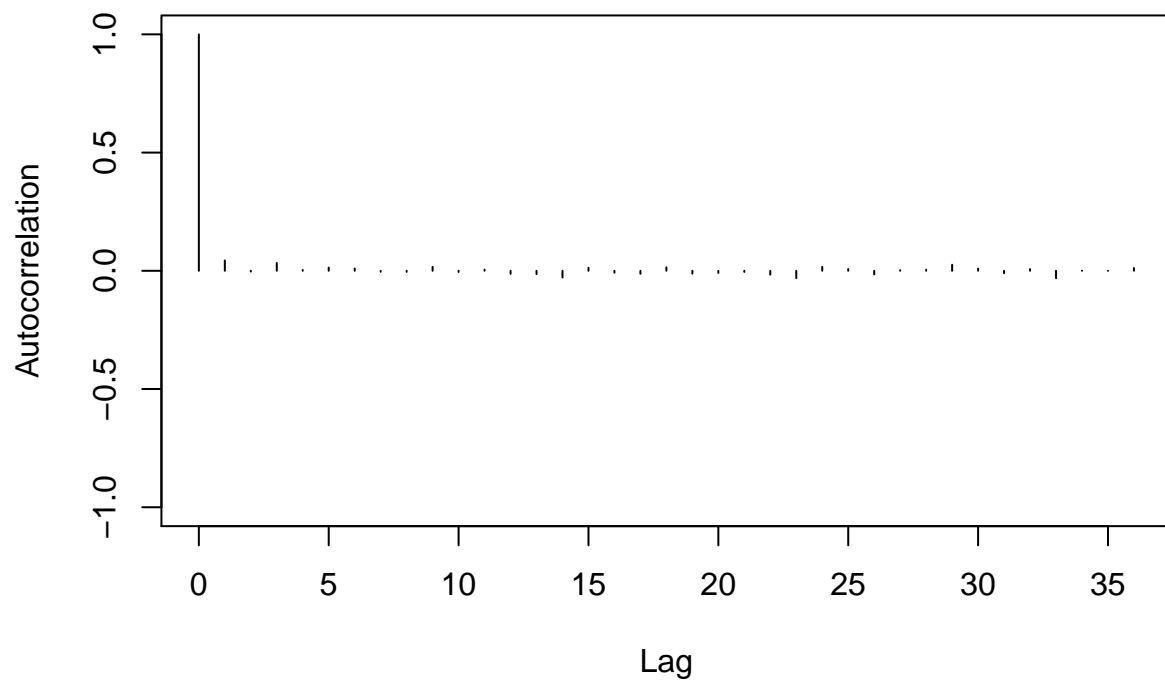

```r
traceplot(mu.mc)
title("Traceplot of Mu")
```
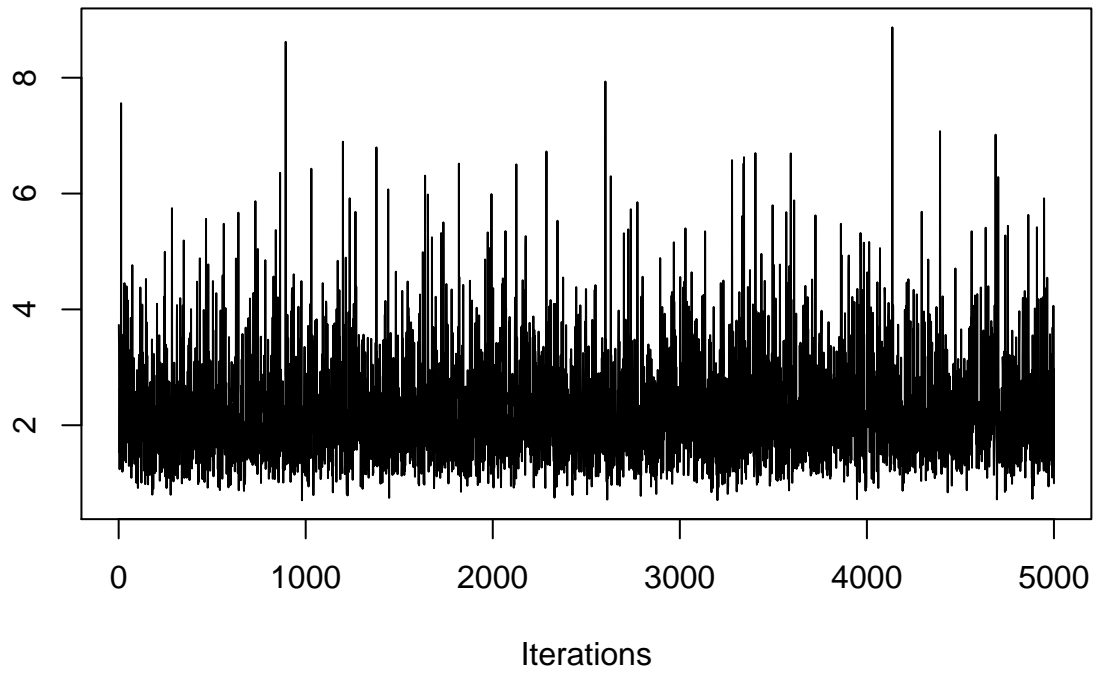
Traceplot of Mu

Iterations
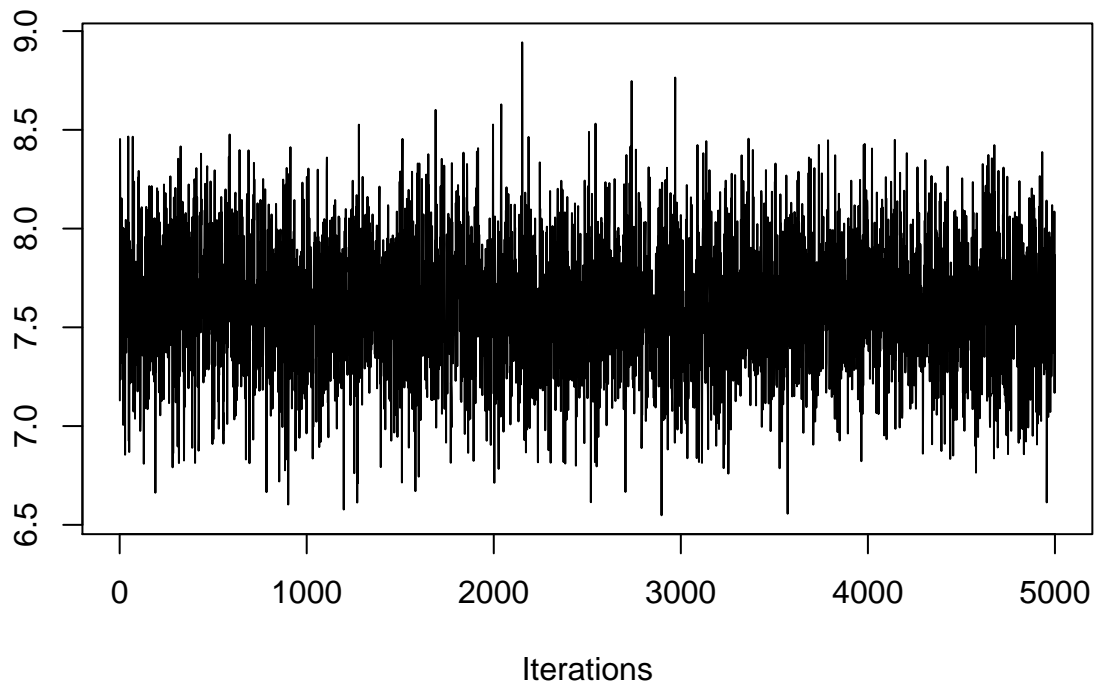
```r
autocorr.plot((mu.mc))
title("Autocorr of Mu")
```
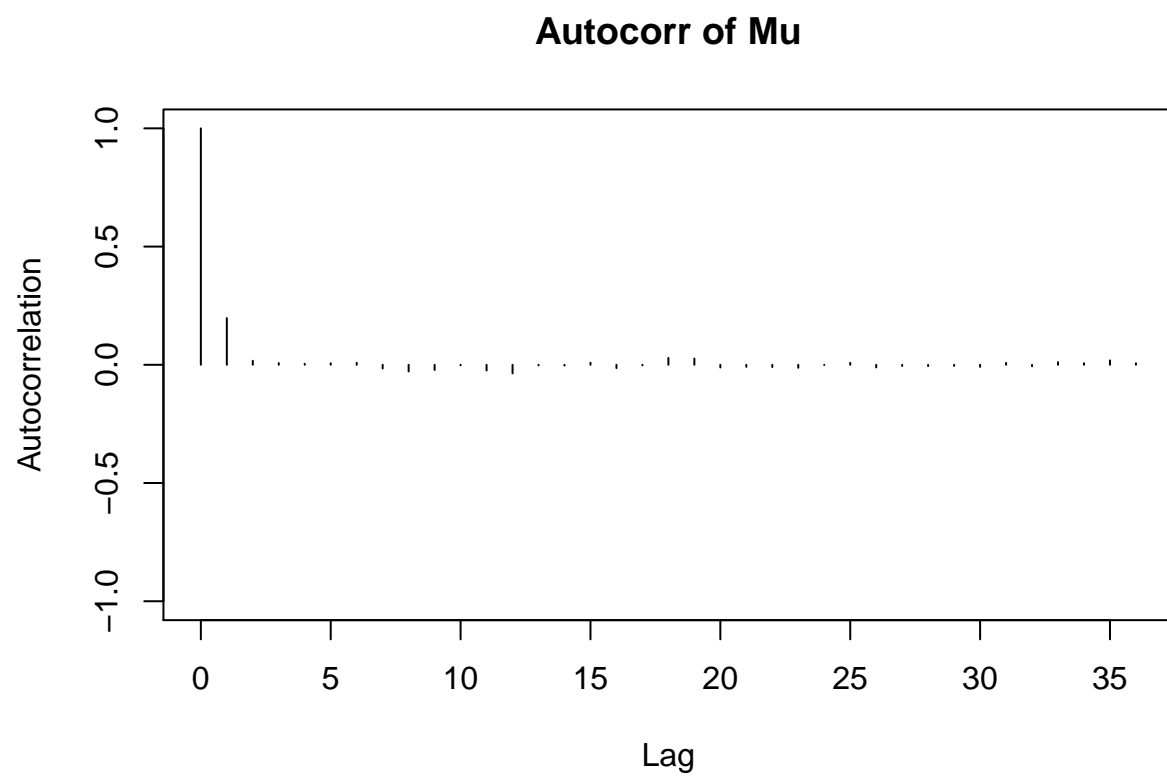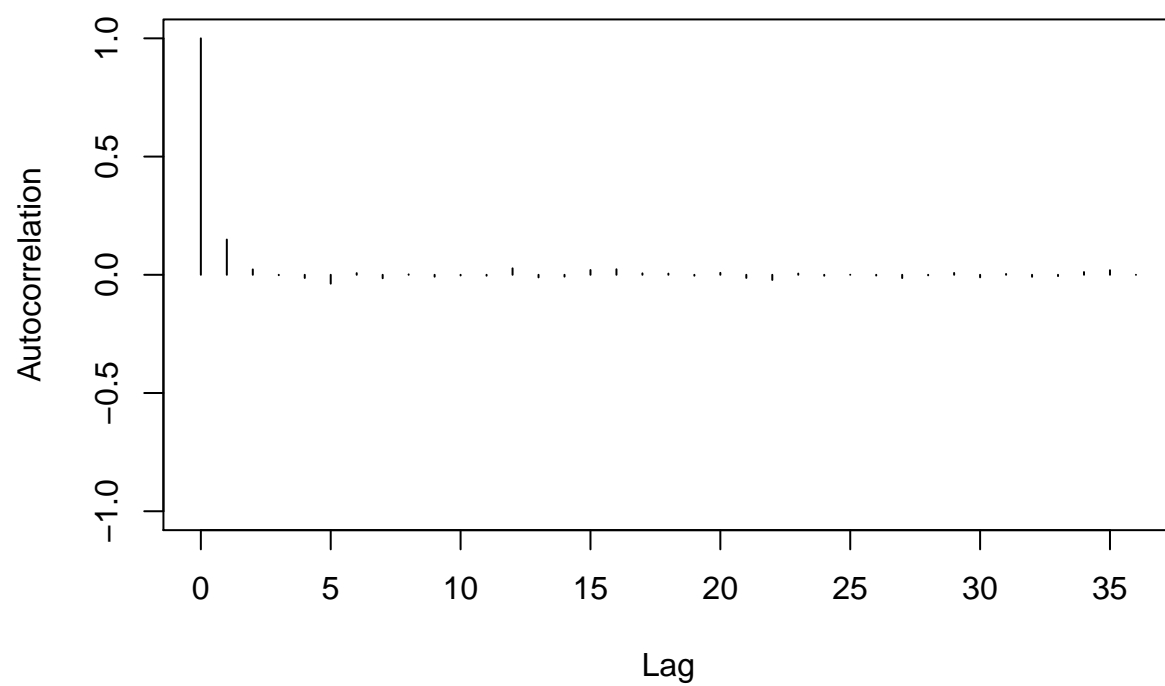
Autocorr of Mu


```
traceplot(tausq.mc)
title("Traceplot of TauSquare")
```
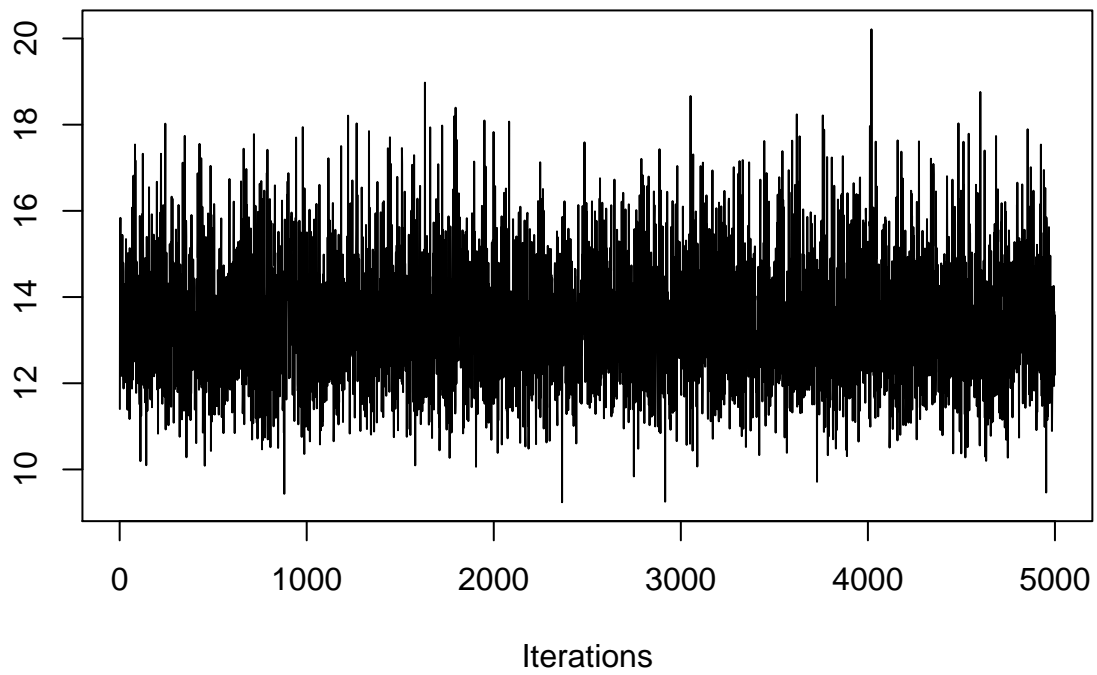
Traceplot of TauSquare

Iterations

```r
autocorr.plot(tausq.mc)
title("Autocorr of TauSquare")
```
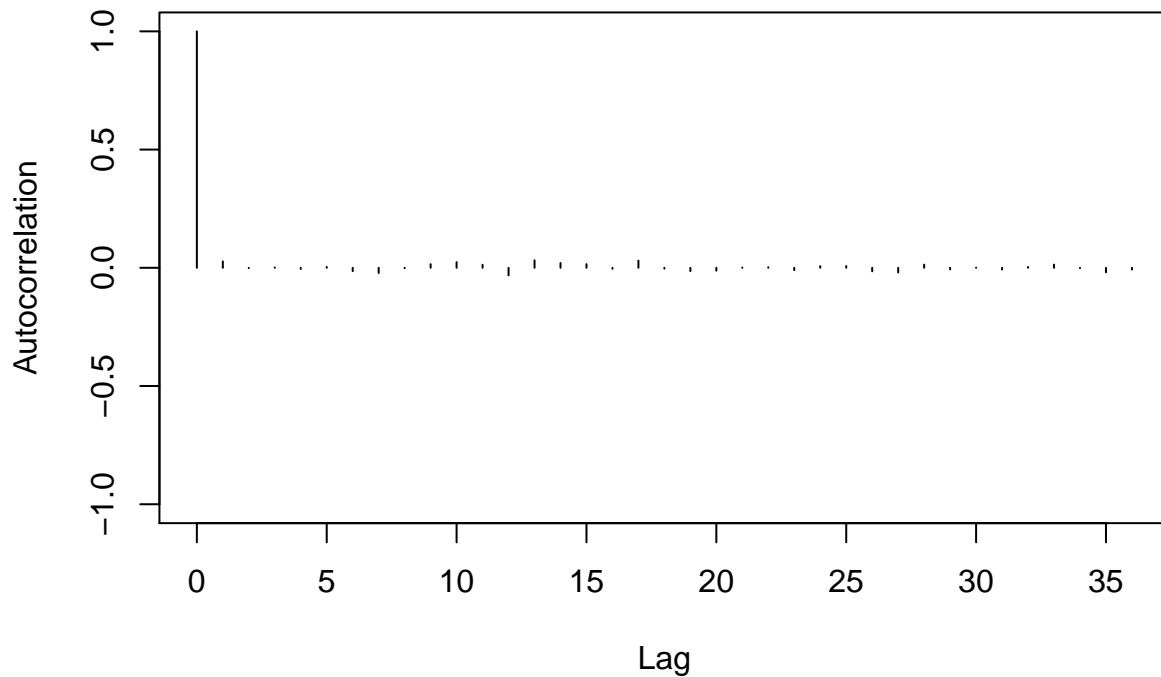
Autocorr of TauSquare


```
traceplot(sigmasq.mc)
title("Traceplot of SigmaSquare")
```

Traceplot of SigmaSquare

Iterations

```
autocorr.plot(sigmasq.mc)
title("Autocorr of SigmaSquare")
```

Autocorr of SigmaSquare

As seen in all the above traceplots and autocorrelation plots, the traces seem to steadily fluctuate about their means indicating that the Markov Chains have converged to their stationary distributions.
The autocorrelation plots show quick autocorrelation decay, usually in less than 5 draws distance. This means that the sampled data likely has little autocorrelation and is hence appropriate.

Overall, the consistent spread in traceplots, stability of traceplots, and low autocorrelation suggest that the Gibbs sampler has indeed converged to the stationary distribution and has generated samples that mix well.

2(d)

μ

```
set.seed(0183939)
hpd.mu <- HPDinterval(mu.mc, prob = 0.95)
cat("The posterior mean of mu is: ", mean(mu.mc), "\n")
```
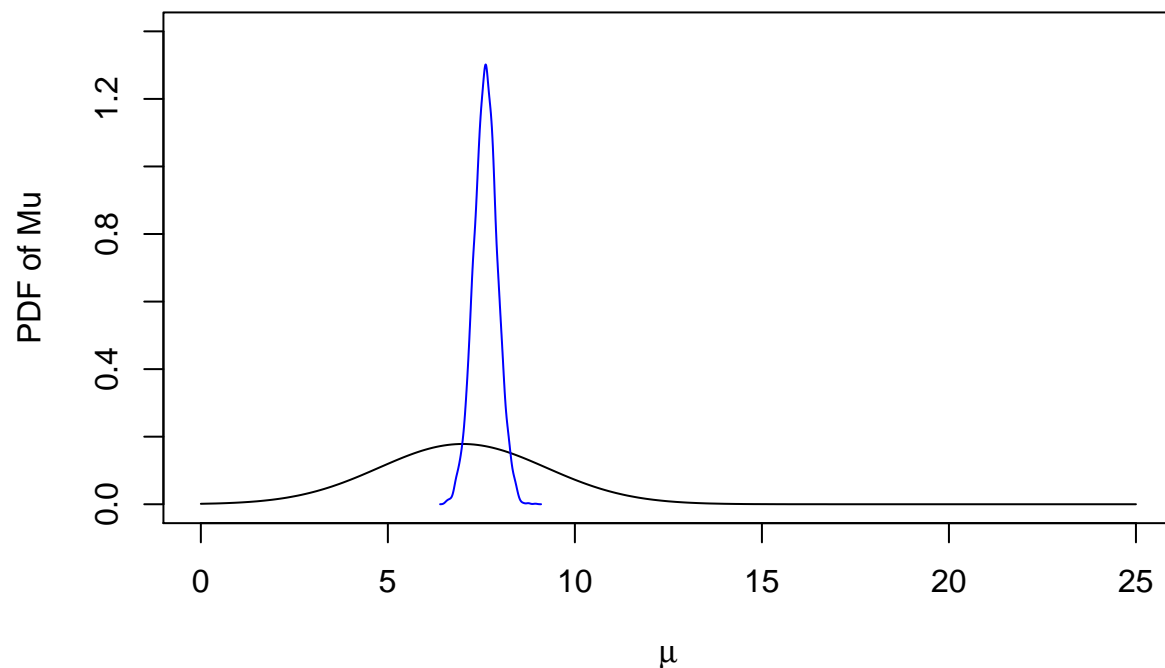
```
## The posterior mean of mu is:  7.610699
```

```
cat("The 95% HPD interval for mu is: (",
    hpd.mu["var1", "lower"], ", ",
    hpd.mu["var1", "upper"], ")\n", sep="")
```

```
## The 95% HPD interval for mu is: (6.992263, 8.248969)
```

```
# Plot Densities
x <- seq(0, 25, by = 0.1)
plot(x, dnorm(x, mean = mu0, sd = sqrt(sigma0sq)), type = "l",
     ylab = "PDF of Mu", xlab = expression(mu), ylim = c(0, 1.4))
lines(density(mu.mc), col = "blue")
```

Please note that the Posterior is in Blue

What is learned from the data:

The data has significantly reduced the variability of our estimation of Mu, producing a posterior with a tall and thin peak.

This posterior means that we are much more sure about where the mean of Mu is.
Notably, the peak of the posterior is somewhat close to the peak of the prior, suggesting that our initial guess of the mode of the prior was somewhat close.

Overall, we have learnt more about where the value of mu is likely to be, improving our certainty.

τ^2

```
set.seed(0183939)
hpd.tausq <- HPDinterval(tausq.mc, prob = 0.95)
cat("The posterior mean of TauSquare is: ", mean(tausq.mc), "\n")
```

```
## The posterior mean of TauSquare is:  2.229522
```
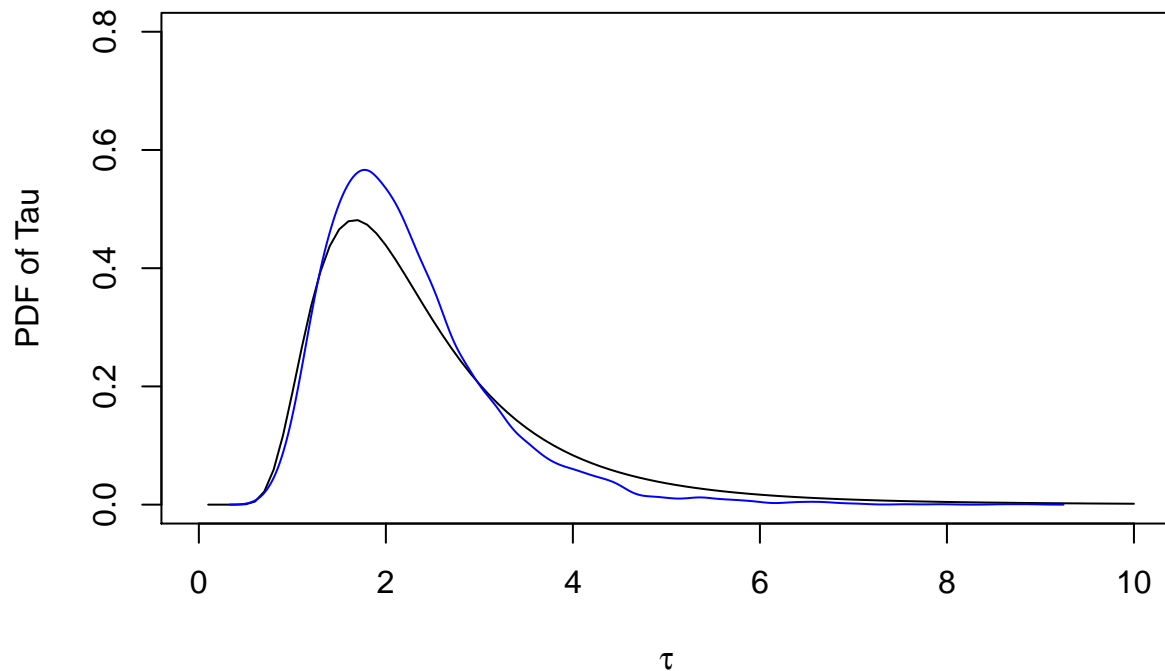
```
cat("The 95% HPD interval for TauSquare is: (",
    hpd.tausq["var1", "lower"], ", ",
    hpd.tausq["var1", "upper"], ")\n", sep="")
```

```
## The 95% HPD interval for TauSquare is: (0.9164617, 4.062765)
```

```
# Plot Densities
x <- seq(0, 10, by = 0.1)
plot(x, dinvgamma(x, a1/2, b1/2), type = "l",
```

```
      ylab = "PDF of Tau", xlab = expression(tau), ylim = c(0, 0.8))
lines(density(tausq.mc), col = "blue")
```


Please note that the Posterior is in Blue

What is learned from the data:

The data has slightly improved the variance by producing a posterior that is slightly more peaked than the prior.
However, the difference is not that drastic, and the mode has barely shifted.
This suggests that our initial guess of the prior was decent.

Overall, we have improved certainty of Tau and would believe that it is likely to be slightly higher than when only considering our prior beliefs.

σ^2

```
set.seed(0183939)
hpd.sigmasq <- HPDinterval(sigmasq.mc, prob = 0.95)
cat("The posterior mean of SigmaSquare is: ", mean(sigmasq.mc), "\n")
```

```
## The posterior mean of SigmaSquare is:  13.47517
```

```
cat("The 95% HPD interval for SigmaSquare is: (",
    hpd.sigmasq["var1", "lower"], ", ",
    hpd.sigmasq["var1", "upper"], ")\n", sep="")
```

```
## The 95% HPD interval for SigmaSquare is: (10.74263, 16.22769)
```

```
# Plot Densities
x <- seq(0, 25, by = 0.1)
plot(x, dinvgamma(x, a2/2, b2/2), type = "l",
     ylab = "PDF of SigmaSquare", xlab = expression(sigma[2]), ylim = c(0, 0.7))
lines(density(sigmasq.mc), col = "blue")
```
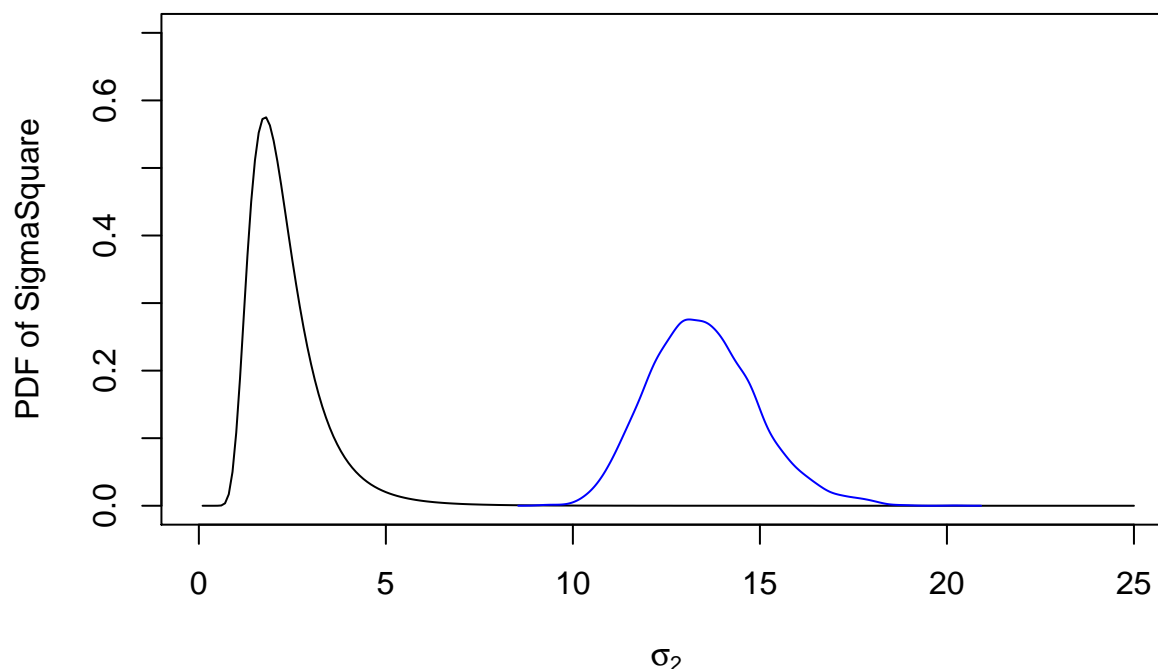

Please note that the Posterior is in Blue

What is learned from the data:

The posterior has shifted some distance away from the prior. This shows us that the prior was likely to have been quite a bad guess. Additionally, the posterior is less peaked, and hence more spread out, than the prior. This suggests that we have more uncertainty about the location of σ^2 after observing our data.

Overall, we have learnt that the location of σ^2 is likely to be much higher than initially believed, and that the uncertainty about its location is also higher than initially believed.

2(e)

```
set.seed(0183939)
# Prior samples
tausq.prior <- rinvgamma(5000, shape = a1/2, rate = b1/2)
sigmasq.prior <- rinvgamma(5000, shape = a2/2, rate = b2/2)
R.prior <- tausq.prior / (tausq.prior + sigmasq.prior)

# Posterior samples
```
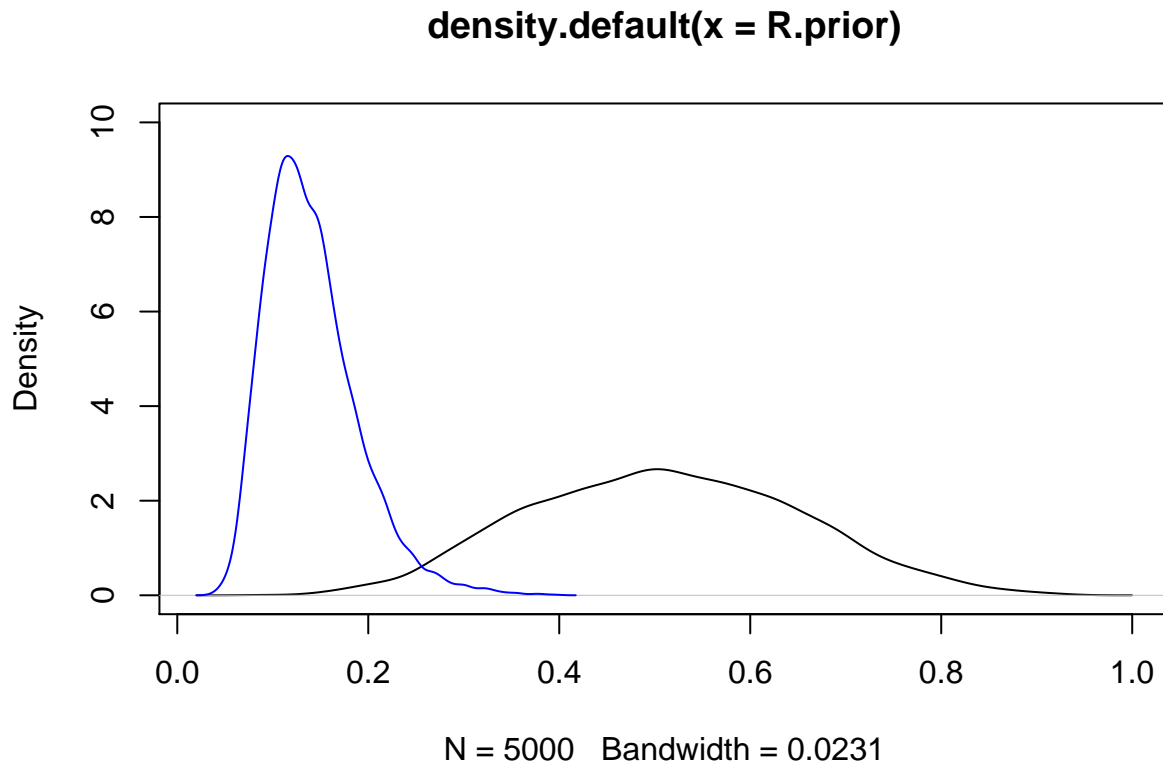
```
R.post <- tausq.mc / (tausq.mc + sigmasq.mc)

# Plot prior and posteriors
plot(density(R.prior), ylim = c(0, 10))
lines(density(R.post), col = "blue")
```

density.default(x = R.prior)

N = 5000   Bandwidth = 0.0231

Please note that the Posterior is in Blue

τ^2 is a measure of between-group variance whereas σ^2 is a measure of within-group variance. Therefore, $R = \frac{\tau^2}{\tau^2 + \sigma^2}$ is a measure of the ratio between-group variance and the overall variance.

As seen in the graphs above, our prior belief for the distribution of R is much further to the right (i.e. larger values) than our posterior distribution. Additionally, the prior has much more spread than the posterior which is more peaked and has smaller variability.

What this means is that the ratio of between-group variance to overall variance is lower, and has less variability, than originally believed. This would suggest that that 8 schools are not as different as initially thought.
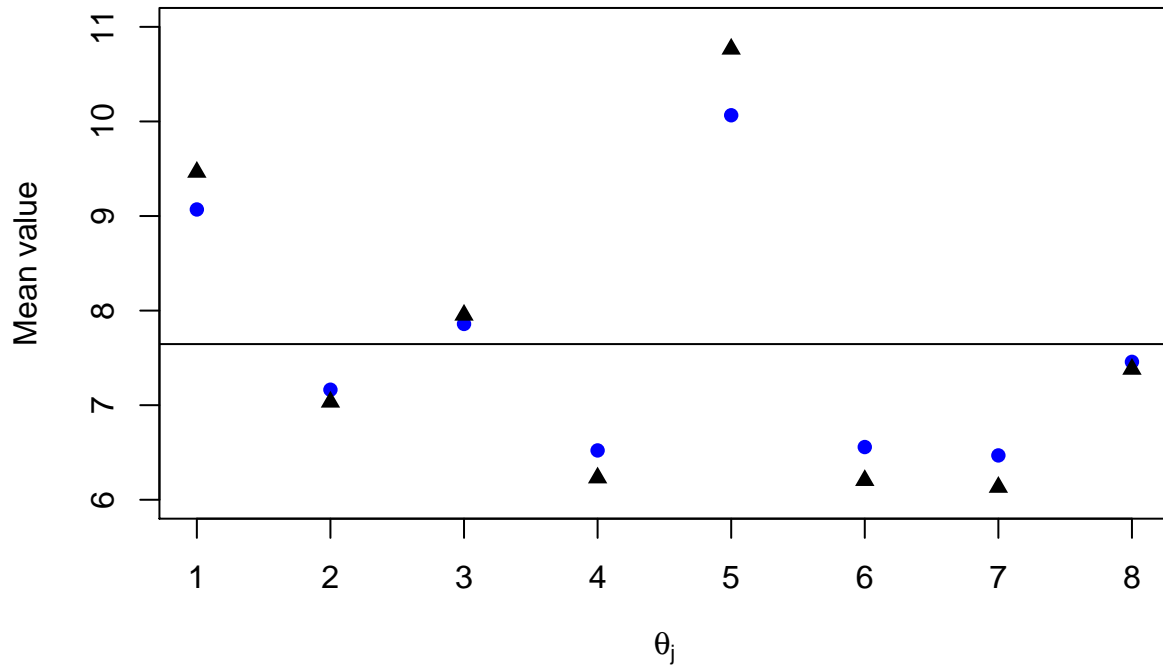
2(f)

```
set.seed(0183939)
theta.means <- apply(theta.mc, 2, mean)
plot(1:8, theta.means, col="blue", pch = 16, cex = 1, ylim = c(6, 11),
     xlab = expression(theta[j]), ylab = "Mean value")
points(1:8, y.bar, pch = 17, cex = 1)
```

```r
abline(h = mean(y.bar))
```



```r
# Note that the overall mean is 7.64589
mean(y.bar)
```

```
## [1] 7.64589
```

Please note that the Theta means are in Blue

The horizontal line shows that overall mean of the samples (i.e. $\bar{\mathbf{y}}$).

The relationship seen is that shrinkage occurs for each of the sample means whereby the posterior is closer to the overall mean (7.6458903), than the original sample mean. This is expected from a bayesian result as it produces compromise estimates of each θ_j.

Overall, the shrinkage apprears to be smaller in effect for sample means that were already close to the overall mean, and larger in effect for sample means that were far from the overall mean.