

STA 601 Homework 6

Lingyun Shao

Oct. 22, 2018

5.3

Marginal distributions: Given observations $Y_1, \dots, Y_n \sim i.i.d. normal(\theta, \sigma^2)$ and using the conjugate prior distribution for θ and σ^2 , derive the formula for $p(\theta|y_1, \dots, y_n)$, the marginal posterior distribution of θ , conditional on the data but marginal over σ^2 . Check your work by comparing your formula to a Monte Carlo estimate of the marginal distribution, using some values of $Y_1, \dots, Y_n, \mu_0, \sigma_0^2, \nu_0$ and κ_0 that you choose. Also derive $p(\tilde{\sigma}^2|y_1, \dots, y_n)$, where $\tilde{\sigma}^2 = 1/\sigma^2$ is the precision.

We know the conjugate prior distributions for $\theta|\sigma^2$ and $1/\sigma^2$ are respectively Normal and Gamma. Let our conjugate prior distributions be

$$\begin{aligned}\theta|\sigma^2 &\sim N(\mu_0, \sigma^2/\kappa_0) \\ 1/\sigma^2 = \tilde{\sigma}^2 &\sim Gamma(\nu_0/2, \nu_0\sigma_0^2/2)\end{aligned}$$

a) Distribution of $\tilde{\sigma}^2|y_1, \dots, y_n$

We can first derive the marginal posterior distribution of $\tilde{\sigma}^2$.

$$\begin{aligned}p(\tilde{\sigma}^2|y_1, \dots, y_n) &\propto p(\tilde{\sigma}^2)p(y_1, \dots, y_n|\tilde{\sigma}^2) \\ &= p(\tilde{\sigma}^2) \int p(y_1, \dots, y_n|\theta, \tilde{\sigma}^2)p(\theta|\tilde{\sigma}^2)d\theta \\ &\propto p(\tilde{\sigma}^2) \int (\tilde{\sigma}^2)^{n/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} \sum (y_i - \theta)^2\right\} \times (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} \kappa_0(\theta - \mu_0)^2\right\} d\theta \\ &= p(\tilde{\sigma}^2)(\tilde{\sigma}^2)^{n/2} \int (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} [\sum (y_i - \theta)^2 + \kappa_0(\theta - \mu_0)^2]\right\} d\theta \\ &= p(\tilde{\sigma}^2)(\tilde{\sigma}^2)^{n/2} \int (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} (\sum y_i^2 - 2 \sum y_i \theta + n\theta^2 + \kappa_0\theta^2 - 2\kappa_0\mu_0\theta + \kappa_0\mu_0^2)\right\} d\theta \\ &= p(\tilde{\sigma}^2)(\tilde{\sigma}^2)^{n/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} (\sum y_i^2 + \kappa_0\mu_0^2)\right\} \int (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} (-2 \sum y_i \theta + n\theta^2 + \kappa_0\theta^2 - 2\kappa_0\mu_0\theta)\right\} d\theta \\ &= p(\tilde{\sigma}^2)(\tilde{\sigma}^2)^{n/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} (\sum y_i^2 + \kappa_0\mu_0^2)\right\} \exp\left\{-\frac{\tilde{\sigma}^2}{2} [-(n + \kappa_0)(\frac{\sum y_i + \kappa_0\mu_0}{n + \kappa_0})^2]\right\} \\ &\quad \times \int (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} [(n + \kappa_0)\theta^2 - 2(\sum y_i + \kappa_0\mu_0)\theta + (n + \kappa_0)(\frac{\sum y_i + \kappa_0\mu_0}{n + \kappa_0})^2]\right\} d\theta \\ &= p(\tilde{\sigma}^2)(\tilde{\sigma}^2)^{n/2} \exp\left\{-\frac{\tilde{\sigma}^2}{2} (\sum y_i^2 + \kappa_0\mu_0^2 - \frac{(\sum y_i + \kappa_0\mu_0)^2}{n + \kappa_0})\right\} \\ &\quad \times \underbrace{\int (\tilde{\sigma}^2)^{1/2} \exp\left\{-\frac{\tilde{\sigma}^2(n + \kappa_0)}{2} (\theta - \frac{\sum y_i + \kappa_0\mu_0}{n + \kappa_0})^2\right\} d\theta}_{\text{constant: it's the kernel of a normal distribution}}\end{aligned}$$

From the last row of equation, we know it's the kernel of $N(\frac{\sum y_i + \kappa_0 \mu_0}{n + \kappa_0}, \sigma^2 / (n + \kappa_0))$ inside the integral, so it's a constant number. Thus we have

$$\begin{aligned}
p(\tilde{\sigma}^2 | y_1, \dots, y_n) &\propto p(\tilde{\sigma}^2) (\tilde{\sigma}^2)^{n/2} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} \left(\sum y_i^2 + \kappa_0 \mu_0^2 - \frac{(\sum y_i + \kappa_0 \mu_0)^2}{n + \kappa_0} \right) \right\} \\
&\propto (\tilde{\sigma}^2)^{\nu_0/2-1} \exp \left\{ -(\nu_0 \sigma_0^2) \tilde{\sigma}^2 \right\} (\tilde{\sigma}^2)^{n/2} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} \left(\sum y_i^2 + \kappa_0 \mu_0^2 - \frac{(\sum y_i + \kappa_0 \mu_0)^2}{n + \kappa_0} \right) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 + \kappa_0 \mu_0^2 - \frac{(\sum y_i + \kappa_0 \mu_0)^2}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 + \frac{(n + \kappa_0) \kappa_0 \mu_0^2 - (\sum y_i)^2 - 2 \sum y_i \kappa_0 \mu_0 - \kappa_0^2 \mu_0^2}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 + \frac{n \kappa_0 \mu_0^2 - (\sum y_i)^2 - 2 \sum y_i \kappa_0 \mu_0}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 + \frac{n \kappa_0 (\mu_0^2 - 2 \mu_0 \bar{y} + \bar{y}^2) - n \kappa_0 \bar{y}^2 - (\sum y_i)^2}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 + \frac{n \kappa_0 (\mu_0^2 - 2 \mu_0 \bar{y} + \bar{y}^2) - (n + \kappa_0) n \bar{y}^2}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{(\nu_0+n)/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + \sum y_i^2 - n \bar{y}^2 + \frac{n \kappa_0 (\mu_0 - \bar{y})^2}{n + \kappa_0}) \right\} \\
&= (\tilde{\sigma}^2)^{\nu_n/2-1} \exp \left\{ -\frac{\tilde{\sigma}^2}{2} (\nu_0 \sigma_0^2 + (n - 1) S^2 + \frac{n \kappa_0 (\mu_0 - \bar{y})^2}{\kappa_n}) \right\}
\end{aligned}$$

Where $\kappa_n = \kappa_0 + n$, $\nu_n = \nu_0 + n$, $S^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$, $\bar{y} = \frac{1}{n} \sum y_i$. Judging from the kernel, we can easily find that $\tilde{\sigma}^2 | y_1, \dots, y_n \sim \text{Gamma}(\nu_n/2, \nu_n \sigma_n^2/2)$, where $\nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n - 1) S^2 + \frac{n \kappa_0 (\mu_0 - \bar{y})^2}{\kappa_n}$.

b) Distribution of $\theta | y_1, \dots, y_n$

It can be easily derived by Bayes theorem that $p(\theta | y_1, \dots, y_n) = \int p(\theta | y_1, \dots, y_n, \tilde{\sigma}^2) p(\tilde{\sigma}^2 | y_1, \dots, y_n) d\tilde{\sigma}^2$.

$$\begin{aligned}
p(\theta | y_1, \dots, y_n, \tilde{\sigma}^2) &\propto p(\theta | \tilde{\sigma}^2) p(y_1, \dots, y_n | \theta, \tilde{\sigma}^2) \\
&\propto \exp \left\{ -\frac{\kappa_0}{2\sigma^2} (\theta - \mu_0)^2 - \frac{1}{2\sigma^2} \sum (y_i - \theta)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} (\kappa_0 \theta^2 - 2 \kappa_0 \mu_0 \theta + n \theta^2 - 2 \sum y_i \theta) \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma^2} [(\kappa_0 + n) \theta^2 - 2(\kappa_0 \mu_0 + \sum y_i) \theta] \right\} \\
&\propto \exp \left\{ -\frac{\kappa_0 + n}{2\sigma^2} (\theta - \frac{\kappa_0 \mu_0 + n \bar{y}}{\kappa_0 + n})^2 \right\}
\end{aligned}$$

From the form of the kernel, we can easily find that $\theta | y_1, \dots, y_n, \tilde{\sigma}^2 \sim N(\mu_n, \sigma^2 / \kappa_n)$, where $\mu_n = \frac{\kappa_0 \mu_0 + n \bar{y}}{\kappa_n}$, $\kappa_n = \kappa_0 + n$.

Then we can derive that

$$\begin{aligned}
p(\theta|y_1, \dots, y_n) &= \int p(\theta|y_1, \dots, y_n, \tilde{\sigma}^2) p(\tilde{\sigma}^2|y_1, \dots, y_n) d\tilde{\sigma}^2 \\
&= \int (2\pi)^{-1/2} (\kappa_n \tilde{\sigma}^2)^{1/2} \exp \left\{ -\frac{\kappa_n \tilde{\sigma}^2}{2} (\theta - \mu_n)^2 \right\} \times \frac{(\nu_n \sigma_n^2/2)^{\nu_n/2}}{\Gamma(\nu_n/2)} (\tilde{\sigma}^2)^{\nu_n/2-1} \exp \left\{ -\frac{\nu_n \sigma_n^2}{2} \tilde{\sigma}^2 \right\} d\tilde{\sigma}^2 \\
&= (2\pi)^{-1/2} (\kappa_n)^{1/2} \frac{(\nu_n \sigma_n^2/2)^{\nu_n/2}}{\Gamma(\nu_n/2)} \underbrace{\int (\tilde{\sigma}^2)^{(\nu_n+1)/2-1} \exp \left\{ -\frac{\kappa_n(\theta-\mu_n)^2 + \nu_n \sigma_n^2 \tilde{\sigma}^2}{2} \right\} d\tilde{\sigma}^2}_{\text{kernel of } \text{Gamma}(\frac{\nu_n+1}{2}, \frac{\kappa_n(\theta-\mu_n)^2 + \nu_n \sigma_n^2}{2})} \\
&= (2\pi)^{-1/2} (\kappa_n)^{1/2} \frac{(\nu_n \sigma_n^2/2)^{\nu_n/2}}{\Gamma(\nu_n/2)} \times \frac{\Gamma((\nu_n+1)/2)}{(\frac{\kappa_n(\theta-\mu_n)^2 + \nu_n \sigma_n^2}{2})^{(\nu_n+1)/2}} \\
&= \frac{\Gamma(\frac{\nu_n+1}{2})}{\sqrt{2\pi} \Gamma(\frac{\nu_n}{2})} (\kappa_n)^{1/2} (\nu_n \sigma_n^2/2)^{-1/2} \frac{(\nu_n \sigma_n^2/2)^{(\nu_n+1)/2}}{(\frac{\kappa_n(\theta-\mu_n)^2 + \nu_n \sigma_n^2}{2})^{(\nu_n+1)/2}} \\
&= \frac{\Gamma(\frac{\nu_n+1}{2})}{\sqrt{\pi} \nu_n \Gamma(\frac{\nu_n}{2}) (\frac{\sigma_n^2}{\kappa_n})^{1/2}} [1 + \frac{\kappa_n}{\nu_n \sigma_n^2} (\theta - \mu_n)^2]^{-\frac{\nu_n+1}{2}} \\
&= \underbrace{\frac{\Gamma(\frac{\nu_n+1}{2})}{\sqrt{\pi} \nu_n \Gamma(\frac{\nu_n}{2}) \lambda} [1 + \frac{1}{\nu_n} (\frac{\theta - \mu_n}{\lambda})^2]^{-\frac{\nu_n+1}{2}}}_{\text{let } \lambda = (\frac{\sigma_n^2}{\kappa_n})^{1/2}}
\end{aligned}$$

The result above is exactly the p.d.f. of a non-standardized Student's t-distribution, where $\lambda = (\frac{\sigma_n^2}{\kappa_n})^{1/2}$ is the scaling parameter of t-distribution, μ_n is the center parameter, $\nu_n = \nu_0 + n$ is the degree of freedom. If we scale θ , then we have $\frac{\theta - \mu_n}{\lambda} | y_1, \dots, y_n \sim t(\nu_n)$.

c) Check formula by Monte Carlo

To check the formula we derived, we can compare the Monte Carlo estimate of the distribution (empirical distribution) to the theoretical distribution from the formula we derived. I defined a function to compare them regarding different values of prior hyperparameters.

```
f_comp = function(n = 10, mu0 = 0, sig0_2 = 1, v0 = 2, k0 = 5) {
  nsam = 100000
  vn = v0 + n
  kn = k0 + n
  set.seed(1)
  Y = rnorm(n, 0, 2)
  y_bar = mean(Y)
  y_sum_sq = var(Y) * (n-1)
  mun = (mu0*k0 + n*y_bar)/kn
  vn_sign_2 = v0*sig0_2 + y_sum_sq + k0*n/(k0+n) *(y_bar-mu0)^2
  pre_samp = rgamma(nsam, vn/2, vn_sign_2/2) # sample of precision
  theta_samp = rnorm(nsam, mun, sqrt(1/pre_samp/kn))
  x_th = seq(-20, 20, by = 0.01)
  x_pre = seq(0.01, 5, by = 0.01)
  g = dgamma(x_pre, vn/2, vn_sign_2/2)
  t = dt(x_th, vn)
  # scaling the sample of theta to a standardized t distribution
  theta_samp_scaled = (theta_samp - mun)/sqrt(vn_sign_2/vn/kn)
```

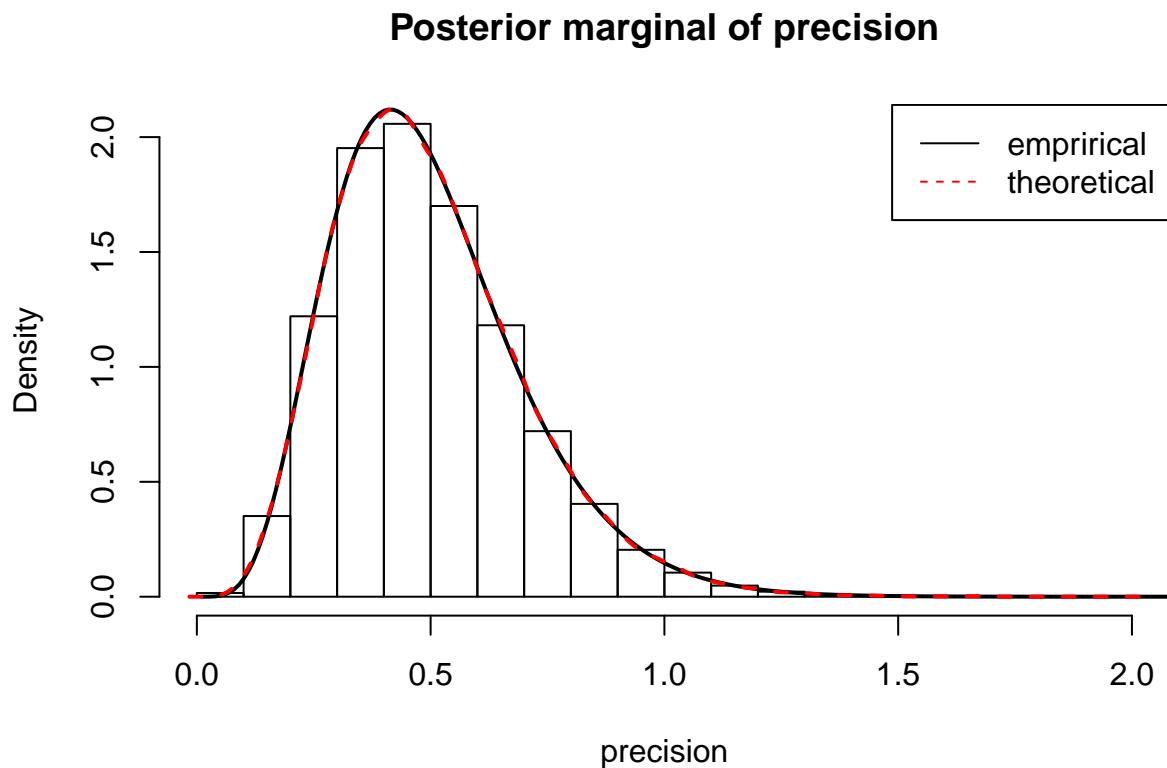
```

# par(mfrow = c(2,1))
hist(pre_samp, freq = FALSE, xlab = 'precision',
      main = 'Posterior marginal of precision')
lines(x_pre, g, lwd = 2)
lines(density(pre_samp), col = 2, lty = 2, lwd = 2)
legend('topright', legend = c('empirical', 'theoretical'), col = 1:2, lty = 1:2)
hist(theta_samp_scaled, freq = FALSE, xlab = 'scaled theta', ylim = c(0, 0.5),
      main = 'Posterior marginal of scaled theta')
lines(x_th, t, lwd = 2)
lines(density(theta_samp_scaled), col = 2, lty = 2, lwd = 2)
legend('topright', legend = c('empirical', 'theoretical'), col = 1:2, lty = 1:2)
}

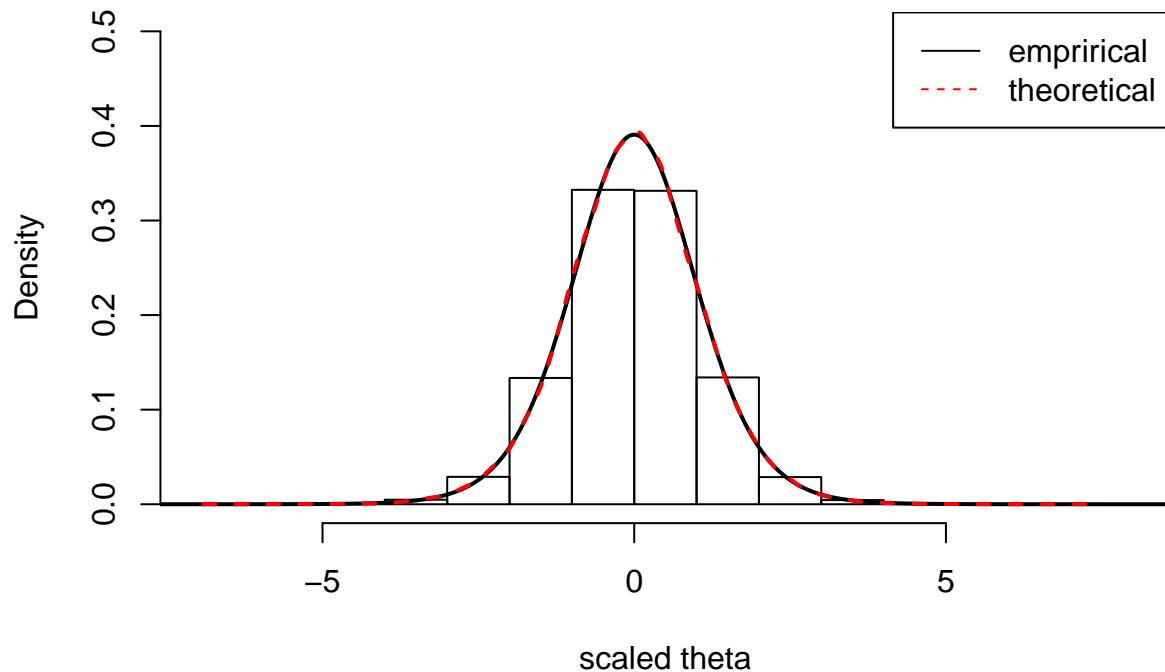
```

We set $\mu_0 = 0$, $\sigma_0^2 = 1$, $\nu_0 = 2$, $\kappa_0 = 5$ and $Y_1, \dots, Y_{10} \sim N(0, 2^2)$ with random seed set to be 1, then generated some Monte Carlo samples of θ and $\tilde{\sigma}^2$ from the posterior distribution. We can compare the histogram of our Monte Carlo samples and its kernel density estimate with the theoretical p.d.f. from the formula we derived. As is shown in the plot below, we can see that the empirical distribution from Monte Carlo approximation are exactly the same as the theoretical distribution as in our formula. So we know the posterior marginal distributions of θ and $\tilde{\sigma}^2$ we derived are correct

f_comp()



Posterior marginal of scaled theta



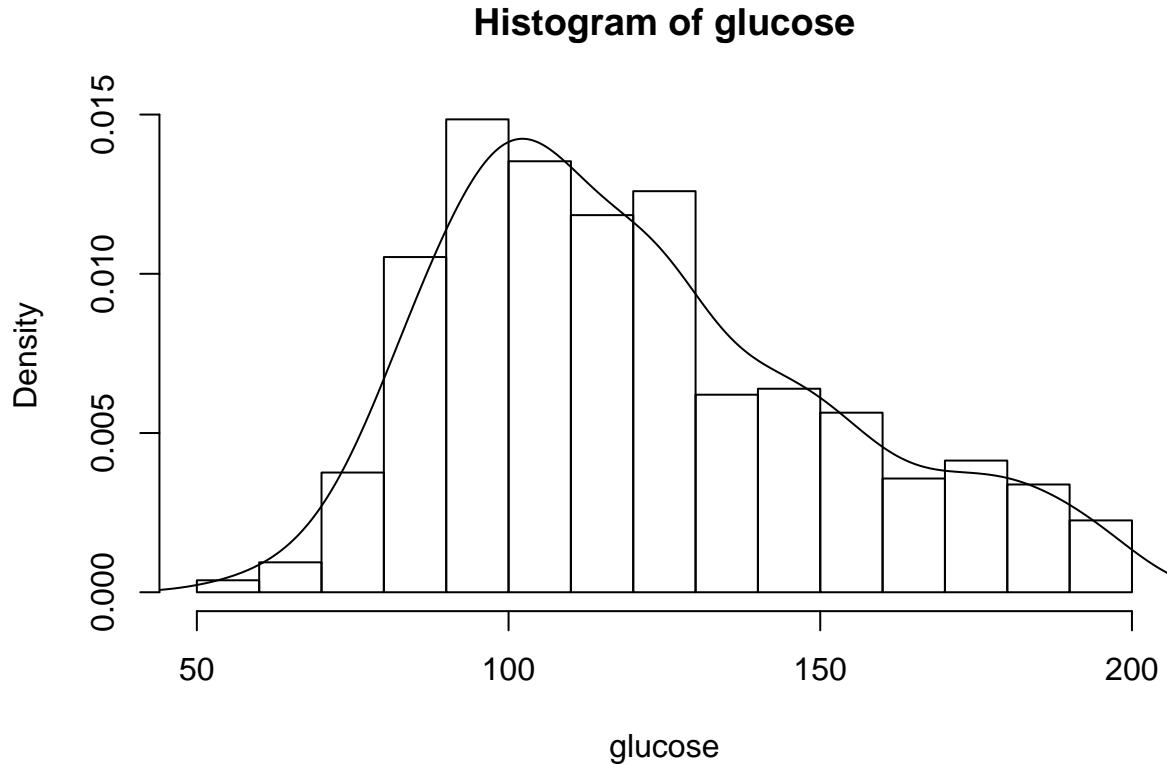
6.2

Mixture model: The file `glucose.dat` contains the plasma glucose concentration of 532 females from a study on diabetes (see Exercise 7.6).

- Make a histogram or kernel density estimate of the data. Describe how this empirical distribution deviates from the shape of a normal distribution.

```
glucose = read.table(file = 'http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/glucose.dat',
                     header = FALSE) %>% .\$V1

hist(glucose, freq = FALSE)
lines(density(glucose))
```



The empirical distribution of the data deviates from the shape of a normal distribution in that

1. Obviously it's not symmetric as is a normal distribution. It's a little positive-skewed.
 2. It has a much heavier tail on the right side than a normal distribution.
 3. It seems to have more than one peak judging from the kernel density plot while a normal distribution only has one.
- b) Consider the following mixture model for these data: For each study participant there is an unobserved group membership variable X_i which is equal to 1 or 2 with probability p and $1 - p$. If $X_i = 1$ then $Y_i \sim \text{normal}(\theta_1, \sigma_1^2)$, and if $X_i = 2$ then $Y_i \sim \text{normal}(\theta_2, \sigma_2^2)$. Let $p \sim \text{beta}(a, b)$, $\theta_j \sim \text{normal}(\mu_0, \tau_0^2)$ and $1/\sigma_j^2 \sim \text{gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$ for both $j = 1$ and $j = 2$. Obtain the full conditional distributions of $(X_1, \dots, X_n), p, \theta_1, \theta_2, \sigma_1^2$ and σ_2^2 .

Full conditional distribution of X_i

The full conditional distribution of X_i is given as

$$\begin{aligned}
& Pr(x_i = j | x_{(i)}, y_1, \dots, y_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) \\
&= Pr(x_i = j | y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) \\
&= \frac{p(y_i | x_i = j, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) Pr(x_i = j | \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p)}{\sum_{j=1,2} p(y_i | x_i = j, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) Pr(x_i = j | \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p)} \\
&= \frac{p(y_i | x_i = j, \theta_j, \sigma_j^2) Pr(x_i = j | p)}{\sum_{j=1,2} p(y_i | x_i = j, \theta_j, \sigma_j^2) Pr(x_i = j | p)}
\end{aligned}$$

Thus we have

$$\begin{aligned}
Pr(x_i = 1 | y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) &= \frac{p(y_i | \theta_1, \sigma_1^2) \times p}{p(y_i | \theta_1, \sigma_1^2) \times p + p(y_i | \theta_2, \sigma_2^2) \times (1 - p)} \\
&= \frac{p \times (\sigma_1^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_1^2} (y_i - \theta_1)^2 \right\}}{p \times (\sigma_1^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_1^2} (y_i - \theta_1)^2 \right\} + (1 - p) \times (\sigma_2^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_2^2} (y_i - \theta_2)^2 \right\}} \\
&= \frac{p \times dnorm(y_i, \theta_1, \sigma_1)}{p \times dnorm(y_i, \theta_1, \sigma_1) + (1 - p) \times dnorm(y_i, \theta_2, \sigma_2)}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
Pr(x_i = 2 | y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) &= \frac{(1 - p) \times dnorm(y_i, \theta_2, \sigma_2)}{p \times dnorm(y_i, \theta_1, \sigma_1) + (1 - p) \times dnorm(y_i, \theta_2, \sigma_2)} \\
&= \frac{(1 - p) \times (\sigma_2^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_2^2} (y_i - \theta_2)^2 \right\}}{p \times (\sigma_1^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_1^2} (y_i - \theta_1)^2 \right\} + (1 - p) \times (\sigma_2^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_2^2} (y_i - \theta_2)^2 \right\}}
\end{aligned}$$

Full conditional distribution of p

Notice the $p \sim beta(a, b)$ and $p(x_i | p) = p^{2-x_i} (1-p)^{x_i-1}$, $x_i = 1, 2$

$$\begin{aligned}
& p(p | x_1, \dots, x_n, y_1, \dots, y_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) \\
&= p(p | x_1, \dots, x_n) \\
&\propto p(p) \prod_{i=1}^n p(x_i | p) \\
&\propto p^{a-1} (1-p)^{b-1} p^{2n - \sum x_i} (1-p)^{\sum x_i - n} \\
&= p^{a+2n-\sum x_i-1} (1-p)^{b+\sum x_i-n-1}
\end{aligned}$$

From the kernel, we know that it's a $beta(a + 2n - \sum x_i, b + \sum x_i - n)$

Full conditional distribution of θ_1, θ_2

For θ_1 , we define $\{y_k^{(j)}\} = \{y_i : x_i = j\}$ is the set of all y' s whose corresponding x' s = j , and the size of this set is $n_j = \sum_{i=1}^n I_{(x_i=j)}$. The full conditional of θ_1 is

$$\begin{aligned}
& p(\theta_1 | x_1, \dots, x_n, y_1, \dots, y_n, \theta_2, \sigma_1^2, \sigma_2^2, p) \\
&= p(\theta_1 | \{y_k^{(1)}\}, \sigma_1^2) \\
&\propto p(\theta_1) \prod_{k=1}^{n_1} p(y_k^{(1)} | \theta_1, \sigma_1^2) \\
&\propto \exp \left\{ -\frac{1}{2\tau_0^2} (\theta_1 - \mu_0)^2 \right\} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{k=1}^{n_1} (y_k^{(1)} - \theta_1)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2A_1} (\theta_1 - B_1)^2 \right\}
\end{aligned}$$

where $A_1 = (\frac{1}{\tau_0^2} + \frac{n_1}{\sigma_1^2})^{-1}$, $B_1 = (\frac{1}{\tau_0^2}\mu_0 + \frac{n_1}{\sigma_1^2}\bar{y}^{(1)})A_1$ and $\bar{y}^{(1)} = \frac{1}{n_1} \sum_k y_k^{(1)}$, so the full conditional distribution of θ_1 is obviously a $N(B_1, A_1)$

Similarly, we can derive the full conditional distribution of θ_2 , which is $N(B_2, A_2)$, where $A_2 = (\frac{1}{\tau_0^2} + \frac{n_2}{\sigma_2^2})^{-1}$, $B_2 = (\frac{1}{\tau_0^2}\mu_0 + \frac{n_2}{\sigma_2^2}\bar{y}^{(2)})A_2$ and $\bar{y}^{(2)} = \frac{1}{n_2} \sum_k y_k^{(2)}$.

Full conditional distribution of $1/\sigma_1^2, 1/\sigma_2^2$

$$\begin{aligned}
& p(1/\sigma_1^2 | x_1, \dots, x_n, y_1, \dots, y_n, \theta_1, \theta_2, \sigma_2^2, p) \\
&= p(1/\sigma_1^2 | \{y_k^{(1)}\}, \theta_1) \\
&\propto p(1/\sigma_1^2) \prod_{k=1}^{n_1} p(y_k^{(1)} | \theta_1, \sigma_1^2) \\
&\propto (1/\sigma_1^2)^{\nu_0/2-1} \exp(-\frac{\nu_0 \sigma_0^2}{2}(1/\sigma_1^2)) \times (1/\sigma_1^2)^{n_1/2} \exp(-\frac{1}{2\sigma_1^2} \sum_{k=1}^{n_1} (y_k^{(1)} - \theta_1)^2) \\
&= (1/\sigma_1^2)^{(\nu_0+n_1)/2-1} \exp(-\frac{\nu_0 \sigma_0^2 + \sum (y_k^{(1)} - \theta_1)^2}{2}(1/\sigma_1^2))
\end{aligned}$$

From the form of the kernel, we can easily find that the full conditional distribution of $1/\sigma_1^2$ is $gamma(\frac{\nu_0+n_1}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{k=1}^{n_1} (y_k^{(1)} - \theta_1)^2}{2})$

Similarly, the full conditional distribution of $1/\sigma_2^2$ is $gamma(\frac{\nu_0+n_2}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{k=1}^{n_2} (y_k^{(2)} - \theta_2)^2}{2})$

- c) Setting $a = b = 1, \mu_0 = 120, \tau_0^2 = 200, \sigma_0^2 = 1000$ and $\nu_0 = 10$, implement the Gibbs sampler for at least 10,000 iterations. Let $\theta_{(1)}^{(s)} = \min \{\theta_{(1)}^{(s)}, \theta_{(2)}^{(s)}\}$ and $\theta_{(2)}^{(s)} = \max \{\theta_{(1)}^{(s)}, \theta_{(2)}^{(s)}\}$. Compute and plot the autocorrelation functions of $\theta_{(1)}^{(s)}$ and $\theta_{(2)}^{(s)}$, as well as their effective sample sizes.

```

a = b = 1
mu_0 = 120
tau_0sq = 200
sig_0sq = 1000
v_0 = 10
y = glucose
n = length(y)

```

```

# updating x
update_x = function(y, theta, sigma_sq, p) {
  # odds of x=2 vs x=1
  odds = (1-p)/p * 1/sqrt(sigma_sq[2]/sigma_sq[1]) *
    exp(-1/2/sigma_sq[2]*(y-theta[2])^2 + 1/2/sigma_sq[1]*(y-theta[1])^2)
  prob1 = 1/(odds + 1)
  u = runif(n, 0, 1)
  x = (u<prob1)*1 + (u>=prob1)*2
  return(x)
}

# updating p
update_p = function(x, a = 1, b = 1) {
  sum = sum(x)
  a_n = a+2*n-sum
  b_n = b+sum-n
  p = rbeta(1, a_n, b_n)
  return(p)
}

# updating theta
update_theta = function(x, y, sigma_sq) {
  y1 = subset(y, x==1)
  y2 = subset(y, x==2)
  n1 = length(y1)
  n2 = length(y2)
  y1.mean = mean(y1)
  y2.mean = mean(y2)
  A1 = (1/tau_0sq + n1/sigma_sq[1])^(-1)
  A2 = (1/tau_0sq + n2/sigma_sq[2])^(-1)
  B1 = (mu_0/tau_0sq + n1*y1.mean/sigma_sq[1]) * A1
  B2 = (mu_0/tau_0sq + n2*y2.mean/sigma_sq[2]) * A2
  theta1_raw = rnorm(1, B1, A1)
  theta2_raw = rnorm(1, B2, A2)
  theta1 = min(theta1_raw, theta2_raw)
  theta2 = max(theta1_raw, theta2_raw)
  return(c(theta1, theta2))
}

# updating sigma2
update_sigma_sq = function(x, y, theta) {
  y1 = subset(y, x==1)
  y2 = subset(y, x==2)
  n1 = length(y1)
  n2 = length(y2)
  y1.sse = sum((y1-theta[1])^2)
  y2.sse = sum((y2-theta[2])^2)
  alpha1 = (v_0+n1)/2
  alpha2 = (v_0+n2)/2
  beta1 = (v_0*sig_0sq+y1.sse)/2
  beta2 = (v_0*sig_0sq+y2.sse)/2
  prec1 = rgamma(1, alpha1, beta1)
  prec2 = rgamma(1, alpha2, beta2)
  return(c(1/prec1, 1/prec2))
}

# set starting point

```

```

theta = c(100, 100)
sigma_sq = c(100, 100)
p = 0.5
Theta = NULL
Sigma_sq = NULL
P = NULL
# X = NULL
iter = 10000
for(i in 1:iter) {
  x = update_x(y, theta, sigma_sq, p)
  p = update_p(x)
  theta = update_theta(x, y, sigma_sq)
  sigma_sq = update_sigma_sq(x, y, theta)
  Theta = rbind(Theta, theta)
  Sigma_sq = rbind(Sigma_sq, sigma_sq)
  P = c(P, p)
}

Theta1 = Theta[,1]
Theta2 = Theta[,2]
acf1 = acf(Theta1, plot = FALSE)
acf2 = acf(Theta2, plot = FALSE)
ef1 = effectiveSize(Theta1)
ef2 = effectiveSize(Theta2)
res = cbind(c(ef1, acf1$acf %>% as.vector()), c(ef2, acf2$acf %>% as.vector()))
rownames(res) = c('Eff_Size', paste('lag', 0:40))
colnames(res) = c('theta1', 'theta2')
kable(res)

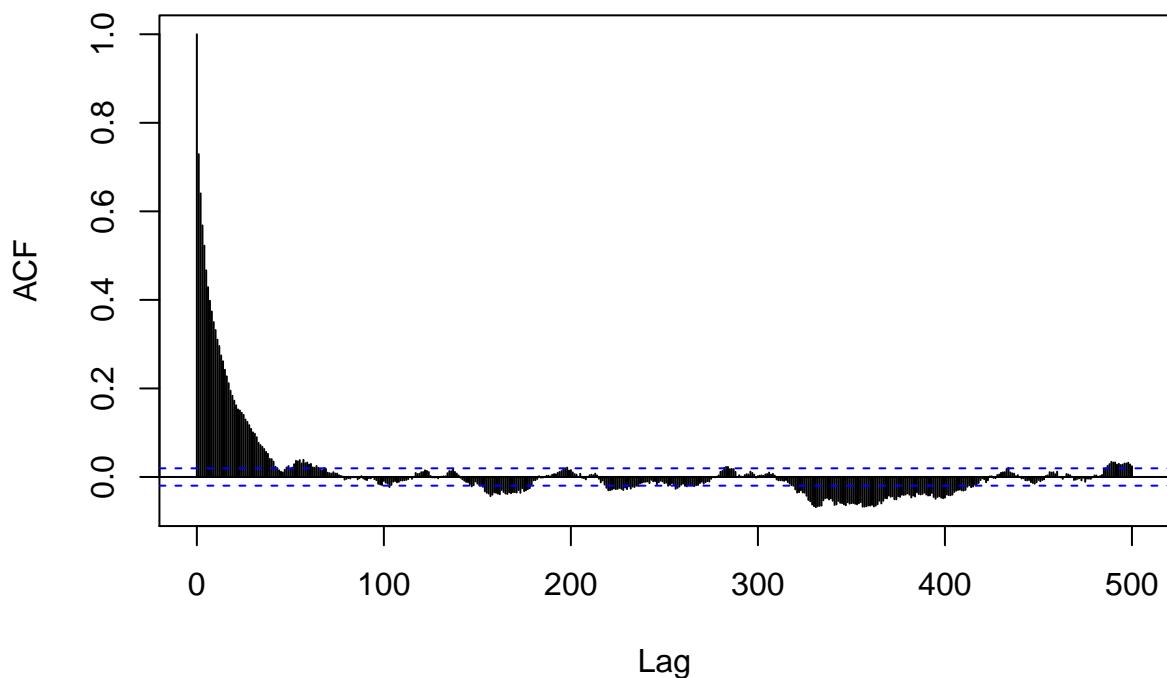
```

| | theta1 | theta2 |
|----------|-------------|-------------|
| Eff_Size | 523.4461039 | 307.2885412 |
| lag 0 | 1.0000000 | 1.0000000 |
| lag 1 | 0.7291226 | 0.8084071 |
| lag 2 | 0.6404988 | 0.7632015 |
| lag 3 | 0.5682432 | 0.7236061 |
| lag 4 | 0.5225357 | 0.6850249 |
| lag 5 | 0.4668355 | 0.6547380 |
| lag 6 | 0.4287836 | 0.6232033 |
| lag 7 | 0.3979353 | 0.5907409 |
| lag 8 | 0.3737579 | 0.5618481 |
| lag 9 | 0.3499174 | 0.5360065 |
| lag 10 | 0.3319488 | 0.5104468 |
| lag 11 | 0.3101475 | 0.4898972 |
| lag 12 | 0.2958843 | 0.4650899 |
| lag 13 | 0.2745091 | 0.4420891 |
| lag 14 | 0.2613426 | 0.4223793 |
| lag 15 | 0.2421093 | 0.4037158 |
| lag 16 | 0.2273257 | 0.3864897 |
| lag 17 | 0.2117969 | 0.3633936 |
| lag 18 | 0.1949140 | 0.3419098 |
| lag 19 | 0.1832846 | 0.3252244 |
| lag 20 | 0.1723358 | 0.3084172 |
| lag 21 | 0.1618997 | 0.2925920 |

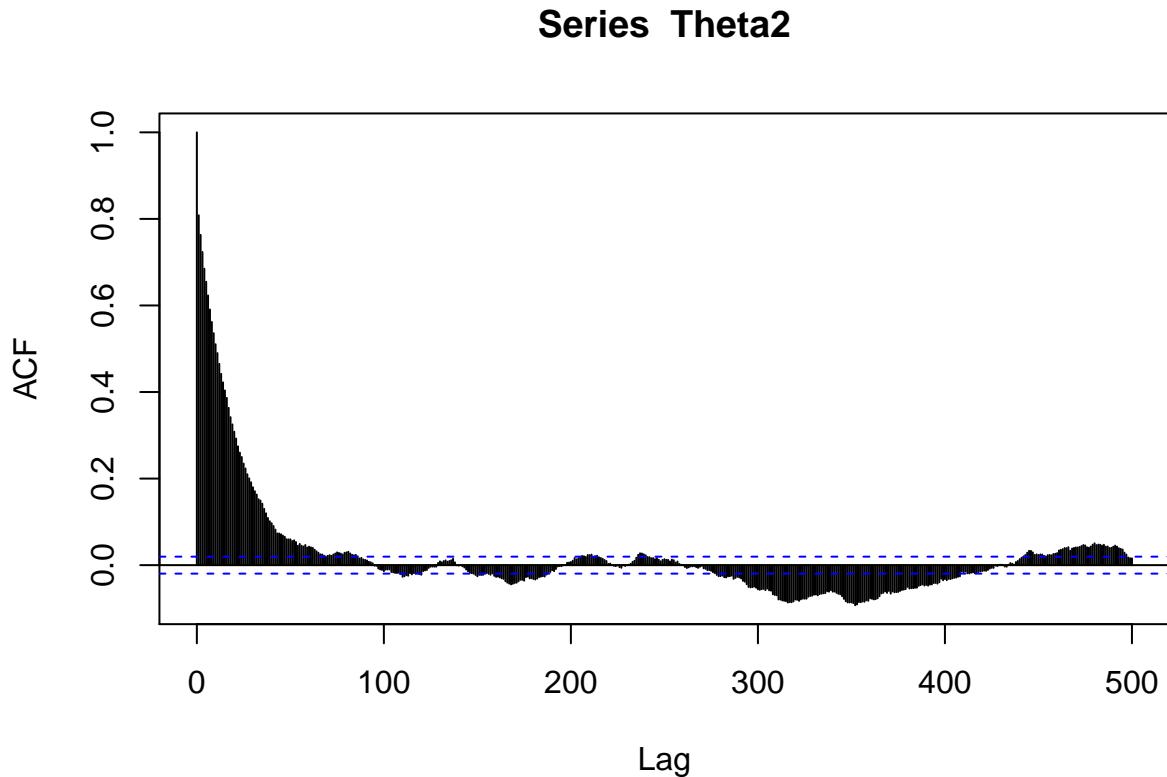
| | theta1 | theta2 |
|--------|-----------|-----------|
| lag 22 | 0.1525063 | 0.2744851 |
| lag 23 | 0.1500533 | 0.2600813 |
| lag 24 | 0.1449453 | 0.2498913 |
| lag 25 | 0.1402258 | 0.2347228 |
| lag 26 | 0.1298606 | 0.2233850 |
| lag 27 | 0.1243675 | 0.2100824 |
| lag 28 | 0.1171492 | 0.2006027 |
| lag 29 | 0.1086656 | 0.1915193 |
| lag 30 | 0.1007951 | 0.1799468 |
| lag 31 | 0.0977246 | 0.1708719 |
| lag 32 | 0.0898003 | 0.1631284 |
| lag 33 | 0.0771266 | 0.1522630 |
| lag 34 | 0.0721313 | 0.1492619 |
| lag 35 | 0.0681400 | 0.1419923 |
| lag 36 | 0.0631924 | 0.1300623 |
| lag 37 | 0.0567622 | 0.1198045 |
| lag 38 | 0.0522231 | 0.1089636 |
| lag 39 | 0.0408390 | 0.1013964 |
| lag 40 | 0.0402446 | 0.0969412 |

```
acf(Theta1, lag.max = 500)
```

Series Theta1



```
acf(Theta2, lag.max = 500)
```



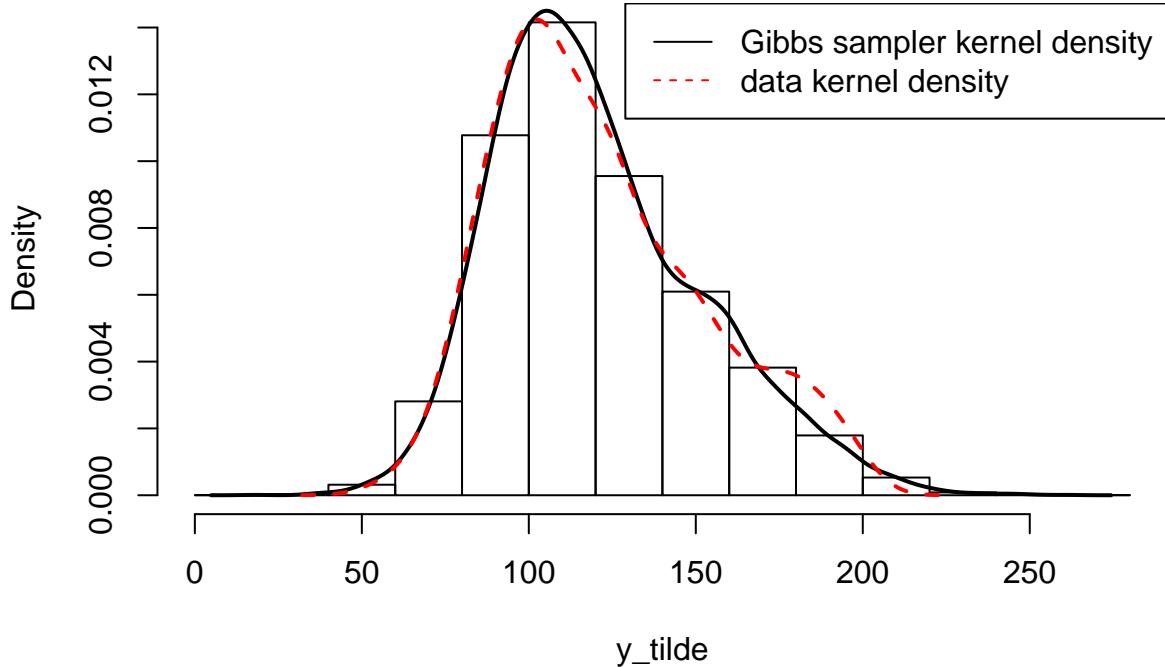
- d) For each iteration s of the Gibbs sampler, sample a value $x \sim \text{binary}(p^{(s)})$, then sample $\tilde{Y}^{(s)} \sim \text{normal}(\theta_x^{(s)}, \sigma_x^{2(s)})$. Plot a histogram or kernel density estimate for the empirical distribution of $\tilde{Y}^{(1)}, \dots, \tilde{Y}^{(s)}$, and compare to the distribution in part a). Discuss the adequacy of this two-component mixture model for the glucose data.

```

u = runif(iter, 0, 1)
x_samp = (u<P)*1 + (u>=P)*2
th_samp = Theta1*(x_samp==1) + Theta2*(x_samp==2)
sig2_samp = Sigma_sq[,1]*(x_samp==1) + Sigma_sq[,2]*(x_samp==2)
y_tilde = rnorm(iter, th_samp, sqrt(sig2_samp))
hist(y_tilde, freq=FALSE)
lines(density(y_tilde), lwd=2)
lines(density(y), col=2, lty=2, lwd=2)
legend('topright', legend = c('Gibbs sampler kernel density', 'data kernel density'), col = 1:2, lty = 1:2)

```

Histogram of $y_{\tilde{}}$



By comparing the kernel density of Gibbs samples and the kernel density of data, we found that these two matched fairly well. The Gibbs samples show a sign of asymmetric, positive-skewed and a little sign of two peaks, which means that the features of the data are well characterized by our two-component mixture model. So we can conclude that this model is adequate for the glucose data.

6.3

Probit regression: A panel study followed 25 married couples over a period of five years. One item of interest is the relationship between divorce rates and the various characteristics of the couples. For example, the researchers would like to model the probability of divorce as a function of age differential, recorded as the man's age minus the woman's age. The data can be found in the file `divorce.dat`. We will model these data with probit regression, in which a binary variable Y_i is described in terms of an explanatory variable x_i via the following latent variable model:

$$\begin{aligned} Z_i &= \beta x_i + \epsilon_i \\ Y_i &= \delta_{(c,\infty)}(Z_i), \end{aligned}$$

where β and c are unknown coefficients, $\epsilon_1, \dots, \epsilon_n \sim i.i.d. normal(0, 1)$ and $\delta_{(c,1)}(z) = 1$ if $z > c$ and equals zero otherwise.

- a) Assuming $\beta \sim normal(0, \tau_\beta^2)$ obtain the full conditional distribution $p(\beta|y, x, z, c)$.

$$\begin{aligned}
p(\beta|y, x, z, c) &= p(\beta|z, x) \\
&\propto p(\beta|x)p(z|\beta, x) \\
&\propto \exp\left(-\frac{1}{2\tau_\beta^2}\beta^2\right) \exp\left(-\frac{1}{2}\sum(z_i - x_i)^2\right) \\
&\propto \exp\left(-\frac{1}{2\tau_\beta^2/(1+\tau_\beta^2\sum x_i^2)}\left(\beta - \frac{\tau_\beta^2 \sum z_i x_i}{1+\tau_\beta^2 \sum x_i^2}\right)^2\right)
\end{aligned}$$

From the form of the kernel, we know the full conditional distribution of β is $N\left(\frac{\tau_\beta^2 \sum z_i x_i}{1+\tau_\beta^2 \sum x_i^2}, \frac{\tau_\beta^2}{1+\tau_\beta^2 \sum x_i^2}\right)$

- b) Assuming $c \sim \text{normal}(0, \tau_c^2)$, show that $p(c|y, x, z, \beta)$ is a constrained normal density, i.e. proportional to a normal density but constrained to lie in an interval. Similarly, show that $p(z_i|y, x, z_{-i}, \beta, c)$ is proportional to a normal density but constrained to be either above c or below c , depending on y_i .

Full conditional distribution of c

Given everything else, the distribution of c will only have dependence on y and z . Given $Y = y$ and $Z = z$, we know that c must be higher than all z'_i 's for which $y_i = 0$ and lower than all z'_i 's for which $y_i = 1$. Let $a := \max\{z_i : y_i = 0\}$, $b := \min\{z_i : y_i = 1\}$. The full conditional of c is then proportional to $p(c)$ but constrained to (a, b) , then the full conditional distribution is given as

$$\begin{aligned}
p(c|y, x, z, \beta) &= p(c|y, z) \\
&\propto p(c)p(y|z, c) \\
&\propto \text{dnorm}(c, 0, \tau_c) \times \delta_{(a,b)}(c)
\end{aligned}$$

which means that $p(c|y, x, z, \beta)$ is a constrained normal density with a support on (a, b) , i.e. proportional to a normal density and lie in the interval (a, b) .

Further from chapter 12 of Hoff, we know that

$$N_{(a,b)}(0, \tau_c^2) = \frac{\phi(x/\tau_c)}{\tau_c(\Phi(b/\tau_c) - \Phi(a/\tau_c))}$$

where ϕ and Φ are the p.d.f. and c.d.f. of a standard normal distribution.

Full conditional distribution of Z_i

Under the sampling model, the conditional distribution of Z_i given β is $Z_i \sim \text{normal}(\beta x_i, 1)$. Given c , observing $Y_i = y_i$ tells us that Z_i must lie in the interval that y_i belongs to, i.e. $(-\infty, c)$ if $y_i = 0$ and (c, ∞) otherwise. The full conditional distribution of Z_i given $\{\beta, z_{-i}, x, y, c\}$ is then

$$p(z_i|y, x, z_{-i}, \beta, c) \propto p(z_i|\beta, x_i)p(y_i|z_i, c) \begin{cases} \text{dnorm}(z_i, \beta x_i, 1) \times \delta_{(c,\infty)}(z_i) & \text{if } y_i = 1, \\ \text{dnorm}(z_i, \beta x_i, 1) \times \delta_{(-\infty,c)}(z_i) & \text{if } y_i = 0 \end{cases}$$

- c) Letting $\tau_\beta^2 = \tau_c^2 = 16$, implement a Gibbs sampling scheme that approximates the joint posterior distribution of Z, β , and c (a method for sampling from constrained normal distributions is outlined in Section 12.1.1). Run the Gibbs sampler long enough so that the effective sample sizes of all unknown parameters are greater than 1,000 (including the Z'_i 's). Compute the autocorrelation function of the parameters and discuss the mixing of the Markov chain.

```

d = read.table(file = 'http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/divorce.dat',
               header = FALSE)
n = nrow(d)
x = d[,1]
y = d[,2]
tau_c_sq = tau_beta_sq = 16

update_beta = function(z, x) {
  M = tau_beta_sq* sum(z*x)/(1+tau_beta_sq*sum(x^2))
  Var = tau_beta_sq/(1+tau_beta_sq*sum(x^2))
  return(rnorm(1, M, sqrt(Var)))
}

update_c = function(y, z) {
  z0 = subset(z, y == 0)
  z1 = subset(z, y == 1)
  a = max(z0)
  b = min(z1)
  u = runif(1, pnorm(a/sqrt(tau_c_sq)), pnorm(b/sqrt(tau_c_sq)))
  return(0+sqrt(tau_beta_sq)*qnorm(u, 0, 1))
}

update_z = function(y, x, beta, c) {
  u0 = runif(n, 0, pnorm(c-x*beta))
  u1 = runif(n, pnorm(c-x*beta), 1)
  z0 = x*beta + qnorm(u0)
  z1 = x*beta + qnorm(u1)
  return(z0*(!y)+z1*y)
}

# starting point
z = rep(0, n)
beta = -2
c = 1
ite = 30000
Beta = NULL
C = NULL
Z = NULL
for(i in 1:ite) {
  beta = update_beta(z, x)
  c = update_c(y, z)
  z = update_z(y, x, beta, c)
  Beta = c(Beta, beta)
  C = c(C, c)
  Z = rbind(Z, z)
}
Sam = cbind(Beta, C, Z)
eff_size = Sam %>% apply(2, effectiveSize)
# effectiveSize
names(eff_size) = c('beta', 'c', paste('Z', 1:25, sep = ''))
eff_size

##      beta          c          Z1          Z2          Z3          Z4          Z5
## 1487.694  1184.267  5482.466 14896.039 16209.155  8532.180 13652.293
##      Z6          Z7          Z8          Z9          Z10         Z11         Z12

```

```

## 26344.131 4030.005 2105.074 15165.264 4942.487 24663.563 5239.534
##      Z13      Z14      Z15      Z16      Z17      Z18      Z19
## 3031.603 4498.543 3549.462 25415.787 3711.639 4513.922 16196.506
##      Z20      Z21      Z22      Z23      Z24      Z25
## 14290.753 4664.390 16611.790 3673.669 1744.137 9845.969
Acf = Sam %>% apply(., 2, function(x) acf(x, lag.max = 20, plot=FALSE) %>% .$acf)
res = rbind(eff_size, Acf)
rownames(res) = c('Eff_Size', paste('lag', 0:20))
res[,1:4] %>%
  round(4) %>%
  kable()

```

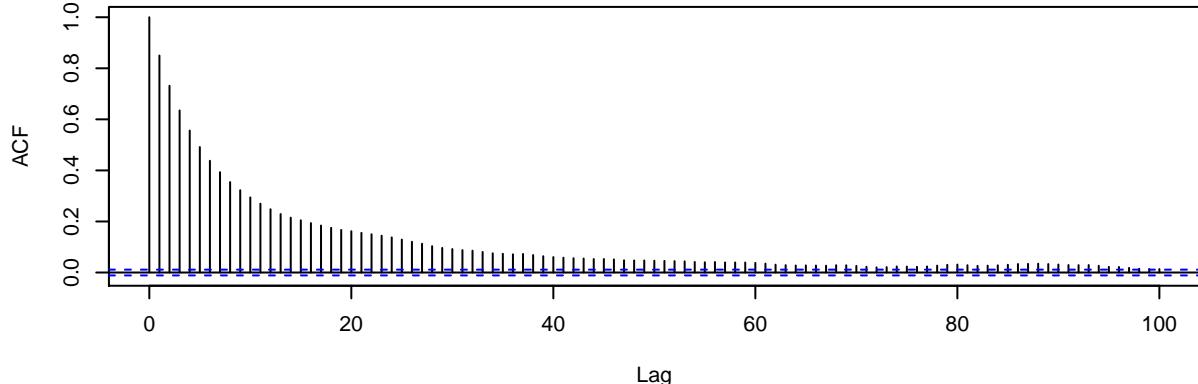
| | beta | c | Z1 | Z2 |
|----------|-----------|-----------|-----------|------------|
| Eff_Size | 1487.6936 | 1184.2671 | 5482.4655 | 14896.0392 |
| lag 0 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| lag 1 | 0.8500 | 0.9218 | 0.1450 | 0.0435 |
| lag 2 | 0.7315 | 0.8520 | 0.1504 | 0.0354 |
| lag 3 | 0.6349 | 0.7884 | 0.1303 | 0.0216 |
| lag 4 | 0.5561 | 0.7305 | 0.1206 | 0.0307 |
| lag 5 | 0.4916 | 0.6770 | 0.1171 | 0.0223 |
| lag 6 | 0.4377 | 0.6287 | 0.1209 | 0.0257 |
| lag 7 | 0.3932 | 0.5848 | 0.1072 | 0.0163 |
| lag 8 | 0.3543 | 0.5443 | 0.0997 | 0.0181 |
| lag 9 | 0.3225 | 0.5069 | 0.0922 | 0.0127 |
| lag 10 | 0.2944 | 0.4721 | 0.0792 | 0.0247 |
| lag 11 | 0.2698 | 0.4412 | 0.0801 | 0.0112 |
| lag 12 | 0.2479 | 0.4129 | 0.0687 | 0.0207 |
| lag 13 | 0.2291 | 0.3858 | 0.0672 | 0.0198 |
| lag 14 | 0.2150 | 0.3602 | 0.0566 | 0.0136 |
| lag 15 | 0.2047 | 0.3379 | 0.0742 | 0.0220 |
| lag 16 | 0.1936 | 0.3176 | 0.0562 | 0.0194 |
| lag 17 | 0.1841 | 0.2982 | 0.0544 | 0.0190 |
| lag 18 | 0.1751 | 0.2804 | 0.0541 | 0.0113 |
| lag 19 | 0.1670 | 0.2634 | 0.0415 | 0.0100 |
| lag 20 | 0.1619 | 0.2472 | 0.0482 | 0.0136 |

```

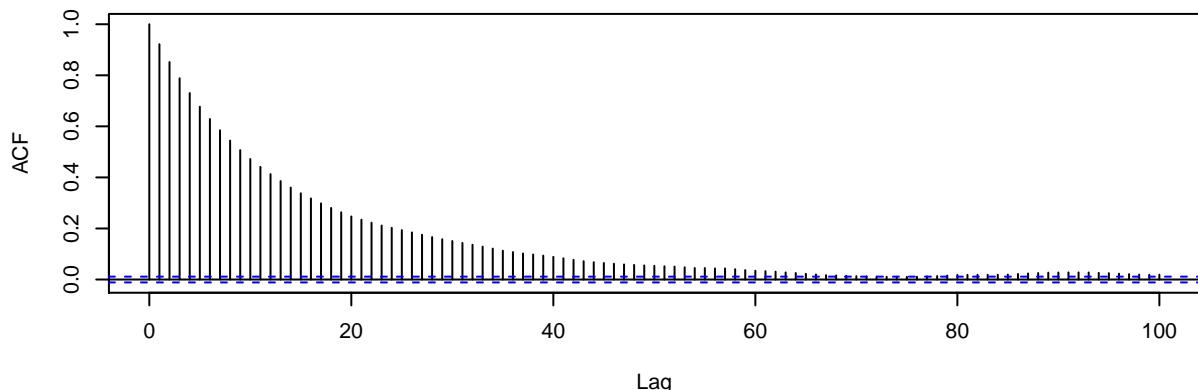
par(mfrow=c(3,1))
acf(Beta, lag.max = 100)
acf(C, lag.max = 100)
acf(Z[,1], lag.max = 100)

```

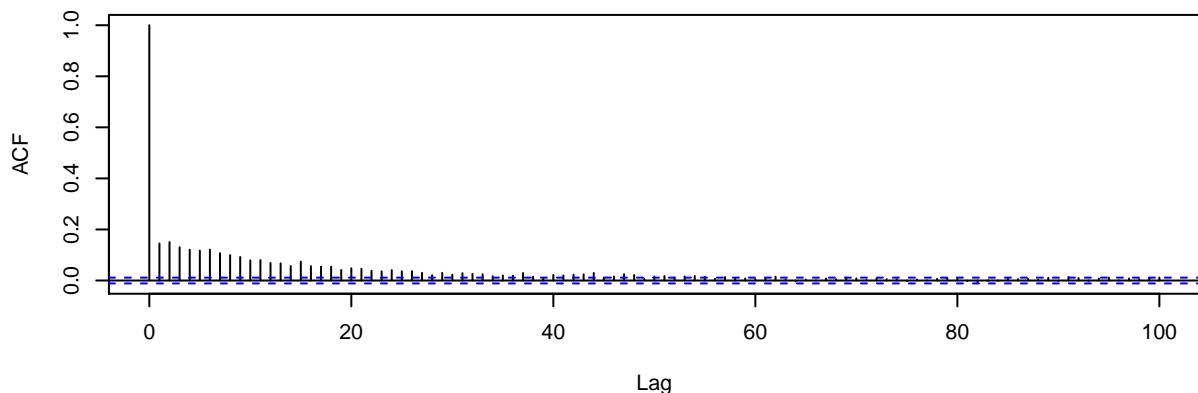
Series Beta



Series C

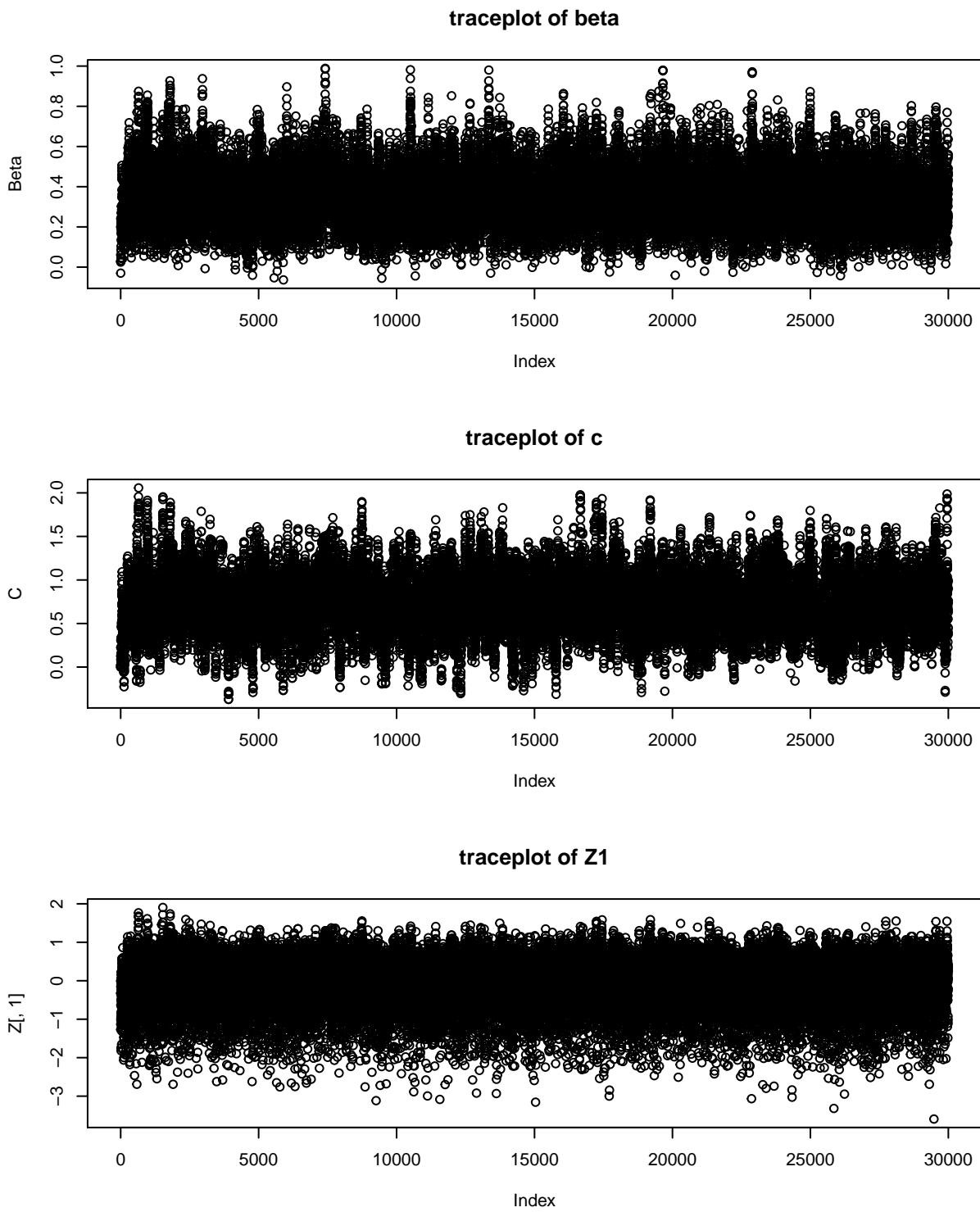


Series Z[, 1]



From the acf values and acf plots we can observe that the mixing of Markov Chain is fairly good in this Gibbs sampler. From around lag-50, the autocorrelation function values are satisfying for β and c , with the effective sample size being around 2000 and 1000 out of 30000 Gibbs samples. The Gibbs sampler did especially well for Z as is shown both in the plot and in the table.

```
par(mfrow=c(3,1))
plot(Beta, main = 'traceplot of beta')
plot(C, main = 'traceplot of c')
plot(Z[,1], main = 'traceplot of Z1')
```



Apart from acf and effective sample size, we can also see from the traceplot of our 30000 Gibbs samples that there is no sign of samples being stuck in some region of parameter space, which means that the Markov Chain have a good mixing.

d) Obtain a 95% posterior confidence interval for β , as well as $Pr(\beta > 0|y, x)$.

```
# 95% CI for beta
quantile(Beta, c(0.025, 0.975))

##      2.5%    97.5%
## 0.1060307 0.6552044

# Pr(beta>0|y,x)
mean(Beta>0)

## [1] 0.9992
```