MA 578: HW3

Benjamin Draves

10/9/2019

BDA (3.8)

(a)

Suppose that $\{y_i\}_{i=1}^{10}$ are the observed counts of bikes on residential roads with bike routes. Moreover, suppose that $\{z_i\}_{i=1}^{8}$ are the observed counts of bikes on residential roads with bike routes. Let $\{n_i\}_{i=1}^{10}$ and $\{m_i\}_{i=1}^{8}$ be the number of vehicles observed on the residential roads with and without bike routes, respectively. Let θ_y and θ_z be the probability of observing a bike on residential roads with and without bike routes, respectively. Then a reasonable model to consider is given by

$$y_i|\theta_y \overset{ind.}{\sim} \mathrm{Binom}(\theta_y, n_i) \qquad z_i|\theta_z \overset{ind.}{\sim} \mathrm{Binom}(\theta_z, m_i)$$

(b)

As $\theta_y, \theta_z \in [0,1]$ a prior that is independent in θ_y and θ_z is simply the flat prior $\theta_y \sim \mathrm{Unif}(0,1)$ and $\theta_z \sim \mathrm{Unif}(0,1)$. Written another way

$$\mathbb{P}(\theta_y, \theta_z) = \mathbb{P}(\theta_y)\mathbb{P}(\theta_z) \propto 1$$

(c)

We are now ready to derive the posterior distribution

$$
\begin{aligned}
\mathbb{P}(\theta_y, \theta_z|\mathbf{Y}, \mathbf{Z}) &\propto \mathbb{P}(\mathbf{Y}, \mathbf{Z}|\theta_y, \theta_z)\mathbb{P}(\theta_y, \theta_z) \propto \mathbb{P}(\mathbf{Y}|\theta_y)\mathbb{P}(\mathbf{Z}|\theta_z) \\
&= \prod_{i=1}^{10} \binom{n_i}{y_i}\theta_y^{y_i}(1-\theta_y)^{n_i-y_i} \prod_{j=1}^{8} \binom{m_i}{z_i}\theta_z^{m_i}(1-\theta_z)^{m_i-z_i} \\
&\propto \prod_{i=1}^{10} \theta_y^{y_i}(1-\theta_y)^{n_i-y_i} \prod_{j=1}^{8} \theta_z^{m_i}(1-\theta_z)^{m_i-z_i} \\
&= \theta_y^{\sum_{i=1}^{10} y_i}(1-\theta_y)^{\sum_{i=1}^{10}(n_i-y_i)}\theta_z^{\sum_{i=1}^{8} z_i}(1-\theta_z)^{\sum_{i=1}^{8}(m_i-z_i)}
\end{aligned}
$$

Marginally, we see that $\theta_y|\mathbf{Y} \sim \mathrm{Beta}(1 + \sum_{i=1}^{10} y_i, 1 + \sum_{i=1}^{10}(n_i - y_i))$ and that $\theta_z|\mathbf{Z} \sim \mathrm{Beta}(1 + \sum_{i=1}^{8} z_i, 1 + \sum_{i=1}^{8}(m_i - z_i))$. As $\theta_z|\mathbf{Z}$ and $\theta_y|\mathbf{Y}$ are independent, we sample each from this marginal posteriors. We plot the KDE of 1000 samples from $\theta_y|\mathbf{Y}$ in red and plot the KDE of 1000 samples from $\theta_z|\mathbf{Z}$ in blue. From here we see that the posterior is indicating the probability of seeing a biker on a residential road with a bike route is lower than similar roads without a bike route.

```
#number of simulations
B <- 1000

# Read in Data
bikes_y <- c(16, 9, 10, 13, 19,20, 18, 17, 35, 55)
cars_y <- c(58, 90, 48, 57, 103, 57, 86, 112, 273, 64)
n <- bikes_y + cars_y

bikes_z <- c(112, 1, 2, 4, 9,7, 9, 8)
cars_z <- c(113, 18, 14, 44, 208, 67, 29, 154)
m <- bikes_z + cars_z

#simulate from Beta(1 + Sy, S(n - bikes)) and Beta(1 + Sz, S(m - bikes))
sim_post_y <- rbeta(B, 1 + sum(bikes_y), 1 + sum(n - bikes_y))
sim_post_z <- rbeta(B, 1 + sum(bikes_z), 1 + sum(m - bikes_z))

#plot posterior densities
plot(density(sim_post_y), col = "red", main = "3.8: Posterior Densities")
points(density(sim_post_z), col = "blue", type = "l")
```
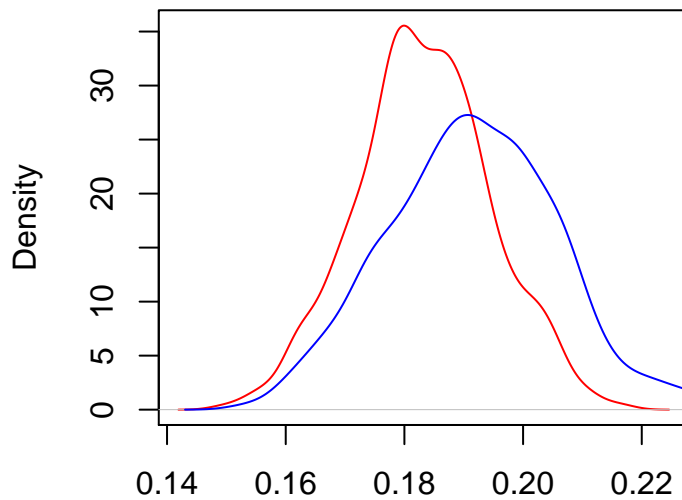
3.8: Posterior Densities

N = 1000   Bandwidth = 0.002433

(d)

For this question we use the simulated $B =1000$ draws from both $\theta_y|\mathbf{Y}$ and $\theta_z|\mathbf{Z}$, $\{\theta_y^{(b)}\}_{b=1}^B$ and $\{\theta_z^{(b)}\}_{b=1}^B$ generated in part (c). From this we generate a new data set $\{y_i^{(b)}\}_{i=1}^B$ and $\{z_i^{(b)}\}_{i=1}^B$ from the model used in From which we attain estimates $\hat{\delta}_i = (\hat{\mu}_y)_1 - (\hat{\mu}_z)_1$. We repeat this 1000 times to attain estimates of $\mu_y - \mu_z$, $\{\hat{\delta}_i\}_{i=1}^{1000}$. A histogram of these differences are given below.

```
#set up differences array
diff_means <- numeric(B)

#iterate
```
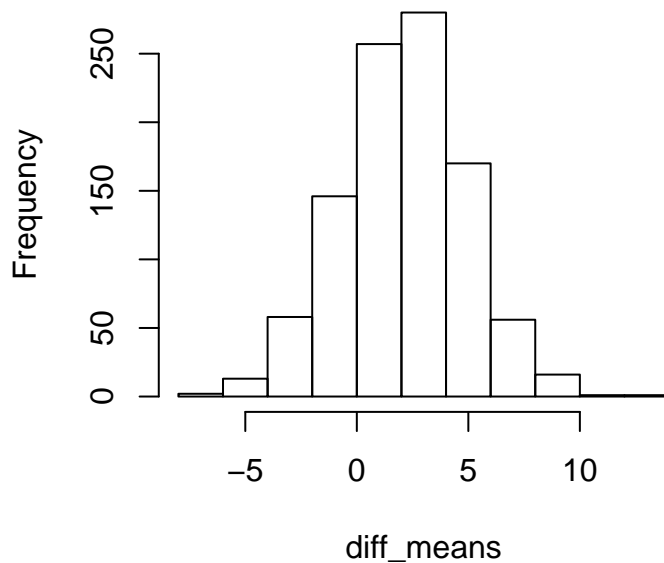
```
for(i in 1:B){

  #sample from binomial model
  ys <- sapply(n, function(x) rbinom(1, x, sim_post_y[i]))
  zs <- sapply(m, function(x) rbinom(1, x, sim_post_z[i]))

  #store mean differece
  diff_means[i] <- mean(ys) - mean(zs)
}


hist(diff_means,  main = "3.8: Difference in Means")
```

3.8: Difference in Means

From this histogram we see there is a slight with $\mu_y - \mu_z > 0$. Namely, this suggests that $\mu_y > \mu_z$ which would mean that the probability of observing a bike on a road without a bike route is lower than that of a road with a bike route. This is more in line with what I would expect - controlling for the exposures (n_i, m_i) we see that the final conclusion is reversed when compared to part (c).

BDA (3.9)

Suppose that $\mathbf{y}|\mu, \sigma^2 \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with $\mathbf{y}$ consisting of n samples. Moreover suppose that $\mu|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ and $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$. In this exercise we look to derive $p(\mu, \sigma^2|\mathbf{y})$. To that end consider the following

$$p(\mu, \sigma^2|\mathbf{y}) \propto p(\mu|\sigma^2, \mathbf{y})p(\sigma^2|\mathbf{y})$$

Therefore, we will first derive $p(\mu|\sigma^2, \mathbf{y})$ and then derive $p(\sigma^2|\mathbf{y})$.

$$p(\mu|\mathbf{y}, \sigma^2) \propto p(\mathbf{y}|\mu, \sigma)p(\mu|\sigma^2)$$

$$\propto \left[\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)\right] \exp\left(-\frac{1}{2\sigma^2/\kappa_0}(\mu - \mu_0)^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 - \frac{\kappa_0}{2\sigma^2}(\mu - \mu_0)^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 - 2n\bar{y}\mu + n\mu^2 + \kappa_0\mu^2 - 2\kappa_0\mu\mu_0 + \kappa_0\mu_0^2\right)\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\left(-2n\bar{y}\mu + n\mu^2 + \kappa_0\mu^2 - 2\kappa_0\mu\mu_0\right)\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\left((n + \kappa_0)\mu^2 - (2n\bar{y} + 2\kappa_0\mu_0)\mu\right)\right)$$

To complete this square, we multiple by $\exp\left(\frac{1}{2\sigma^2} * \frac{b^2}{4a}\right)$ where $b = 2(n\bar{y} + \kappa_0\mu_0)$ and $a = (n + \kappa_0)$. With this we arrive at

$$\exp\left(-\frac{1}{2\sigma^2}\left(\sqrt{n + \kappa_0}\mu - \frac{2(n\bar{y} + \kappa_0\mu_0)}{2\sqrt{n + \kappa_0}}\right)^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2/(n + \kappa_0)}\left(\mu - \frac{n\bar{y} + \kappa_0\mu_0}{n + \kappa_0}\right)^2\right)$$

Therefore we see that

$$\mu|\mathbf{y}, \sigma^2 \sim N\left(\frac{n}{n + \kappa_0}\bar{y} + \frac{\kappa_0}{n + \kappa_0}\mu_0, \frac{\sigma^2}{n + \kappa_0}\right)$$

Next we look to derive $p(\sigma^2|\mathbf{y})$. Well, consider

$$p(\sigma^2|\mathbf{y}) \propto p(\mathbf{y}|\sigma^2)p(\sigma^2) = p(\sigma^2)\int p(\mathbf{y}, \mu|\sigma^2)d\mu$$

$$= p(\sigma^2)\int p(\mathbf{y}|\mu, \sigma^2)p(\mu|\sigma^2)d\mu$$

$$= p(\sigma^2)\int (2\pi\sigma^2)^{-n/2}(2\pi\sigma^2/\kappa_0)^{-1/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)\exp\left(-\frac{1}{2\sigma^2/\kappa_0}(\mu - \mu_0)^2\right)d\mu$$

Notice by adding and subtracting $\bar{y}$ we see that

$$\sum_{i=1}^{n}(y_i - \mu)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + \sum_{i=1}^{n}(\bar{y} - \mu)^2 + \sum_{i=1}^{n}(y_i - \bar{y})(\bar{y} - \mu)$$

$$= (n - 1)s_y^2 + n(\bar{y} - \mu)^2$$

Therefore, we see that

$$p(\sigma^2) \int (2\pi\sigma^2)^{-n/2} (2\pi\sigma^2/\kappa_0)^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \exp\left(-\frac{1}{2\sigma^2/\kappa_0}(\mu - \mu_0)^2\right) d\mu$$

$$= p(\sigma^2) \int (2\pi\sigma^2)^{-n/2} (2\pi\sigma^2/\kappa_0)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s_y^2 + n(\bar{y} - \mu)^2 + \kappa_0(\mu - \mu_0)^2]\right) d\mu$$

$$= (2\pi\sigma^2)^{-\frac{\nu_0+n}{2}-1}(2\pi\sigma^2/\kappa_0)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s_y^2 + \nu_0\sigma_0^2]\right)$$

$$\times \int \exp\left(-\frac{1}{2\sigma^2}[n(\bar{y} - \mu)^2 + \kappa_0(\mu - \mu_0)^2]\right) d\mu$$

Focusing on this integral, we have

$$\exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 - 2n\bar{y}\mu + n\mu^2 + \kappa_0\mu^2 - 2\kappa_0\mu\mu_0 + \kappa_0\mu_0^2\right)\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 + \kappa_0\mu_0^2\right)\right) \exp\left(-\frac{1}{2\sigma^2}\left((n + \kappa_0)\mu^2 - 2(n\bar{y} + \kappa_0\mu_0)\mu\right)\right)$$

Multiplying this term by $\exp\left(-\frac{1}{2\sigma^2}\frac{(n\bar{y}+\kappa_0\mu_0)^2}{n+\kappa_0}\right)$ gives

$$\exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 + \kappa_0\mu_0^2\right)\right) \exp\left(\frac{1}{2\sigma^2}\frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0}\right) \exp\left(-\frac{1}{2\sigma^2}\left(\sqrt{n + \kappa_0}\mu - \frac{(n\bar{y} + \kappa_0\mu_0)}{\sqrt{n + \kappa_0}}\right)^2\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0}\right)\right) \exp\left(-\frac{1}{2\sigma^2}\left(\sqrt{n + \kappa_0}\mu - \frac{(n\bar{y} + \kappa_0\mu_0)}{\sqrt{n + \kappa_0}}\right)^2\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0}\right)\right) \exp\left(-\frac{1}{2\sigma^2/(n + \kappa_0)}\left(\mu - \frac{(n\bar{y} + \kappa_0\mu_0)}{n + \kappa_0}\right)^2\right)$$

Therefore, we see that

$$\int \exp\left(-\frac{1}{2\sigma^2/(n + \kappa_0)}\left(\mu - \frac{(n\bar{y} + \kappa_0\mu_0)}{n + \kappa_0}\right)^2\right) d\mu = \sqrt{\frac{2\pi\sigma^2}{n + \kappa_0}}$$

Therefore, we are left with

$$p(\sigma^2|\mathbf{y}) \propto (2\pi\sigma^2)^{-\frac{\nu_0+n}{2}-1}(2\pi\sigma^2/\kappa_0)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s_y^2 + \nu_0\sigma_0^2]\right)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}\left(n\bar{y}^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0}\right)\right) \sqrt{2\pi\frac{\sigma^2}{n + \kappa_0}}$$

$$\propto (\sigma^2)^{-\frac{\nu_0+n}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s_y^2 + \nu_0\sigma_0^2 + n\bar{y}^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0}\right]\right)$$

5

Next notice,

$$n\bar{y}^2 + \kappa_0\mu_0^2 - \frac{(n\bar{y} + \kappa_0\mu_0)^2}{n + \kappa_0} = \frac{(n + \kappa_0)(n\bar{y}^2 + \kappa_0\mu_0^2) - n^2\bar{y}^2 - n\kappa_0\bar{y}\mu_0 - \kappa_0^2\mu_0^2}{n + \kappa_0}$$

$$= \frac{n^2\bar{y}^2 + \kappa_0 n\mu_0^2 + n\kappa_0\bar{y}^2 + \kappa_0^2\mu_0^2 - n^2\bar{y}^2 - n\kappa_0\bar{y}\mu_0 - \kappa_0^2\mu_0^2}{n + \kappa_0}$$

$$= \frac{\kappa_0 n\mu_0^2 + n\kappa_0\bar{y}^2 - n\kappa_0\bar{y}\mu_0}{n + \kappa_0}$$

$$= \frac{\kappa_0 n}{n + \kappa_0}(\bar{y} - \mu_0)^2$$

Therefore, we see that

$$\mathbb{P}(\sigma^2|\mathbf{y}) \propto (\sigma^2)^{-\frac{\nu_0 + n}{2} - 1}\exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s_y^2 + \nu_0\sigma_0^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2\right]\right)$$

From which we conclude that

$$\sigma^2|\mathbf{y} \sim \text{Inv} - \chi^2\left(\nu_0 + n, (n-1)s_y^2 + \nu_0\sigma_0^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2\right)$$

Define the following quantities

$$\mu_n = \frac{n}{n + \kappa_0}\bar{y} + \frac{\kappa_0}{n + \kappa_0}\mu_0$$

$$\kappa_n = n + \kappa_0$$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = (n-1)s_y^2 + \nu_0\sigma_0^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2$$

We conclude that $\mu, \sigma^2|\mathbf{y} \sim \text{N-Inv} - \chi^2(\mu_n, \sigma^2/\kappa_n; \mu_n, \sigma_n^2)$ thus concluding our derivation of the posterior.

BDA 3.12

(a)

In this problem, we consider the Poisson Regression model. While the book suggests modeling the mean as $\alpha + \beta t$, this does not necessitate that the mean be positive. For that reason we will be using the canonical link in this exercise. That is, we will model

$$\mathbf{y}_i|\lambda_i \sim \text{Pois}(\lambda_i) \qquad \lambda_i = \exp(\alpha + \beta * t_i)$$

The corresponding likelihood for this data is given by

$$p(\mathbf{y}|\alpha, \beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

Therefore, to derive a posterior distribution on (α, β) we would consider

$$
\begin{aligned}
p(\alpha, \beta|\mathbf{y}) &\propto \prod_{i=1}^{n} e^{-\lambda_i} \lambda_i^{y_i} p(\alpha, \beta) \\
&= \prod_{i=1}^{n} \exp\left(-\exp(\alpha + \beta t_i)\right) \exp\left(y_i(\alpha + \beta t_i)\right) p(\alpha, \beta) \\
&= \exp\left(-\sum_{i=1}^{n} \exp(\alpha + \beta t_i) + \alpha \sum_{i=1}^{n} t_i + \beta \sum_{i=1}^{n} y_i t_i\right) p(\alpha, \beta)
\end{aligned}
$$

Since we see there is no hope of achieving conjugacy at this point, we are free to choose any prior we wish as we will need to rely on computational methods to solve this problem. As we wish to choose a noninformative prior, we simply let $p(\alpha, \beta) \propto 1$.

(b)

In settings where we wish to do variable selection or regularization, we can choose priors that place high mass around $(0, 0)$. For variable selection, this is reminiscent of the "spike and slab prior" where each variable (α, β) is associated with a latent variable that effectively turns that variable "on and off". This prior can be seen in red in the figure below. For general regularization, we simply place double exponential prior on each α, β encouraging these coefficients to be shrunk towards 0. We sketch these below as well as the contour of the double exponential.

```r
library(rmutil)
```

```
##
## Attaching package: 'rmutil'

## The following object is masked from 'package:stats':
##
##      nobs

## The following objects are masked from 'package:base':
##
##      as.data.frame, units
```

```r
plot(density(rnorm(1000, sd = 5)), type = "l", ylim = c(0 , 1), xlim = c(-10, 10),
     main = "Spike and Slab and the Double Exponential", col = "red")
points(density(rlaplace(1000, s= 1/2)), type = "l", col = "red")
points(density(rlaplace(1000, s= 1)), type = "l", col = "blue")
```
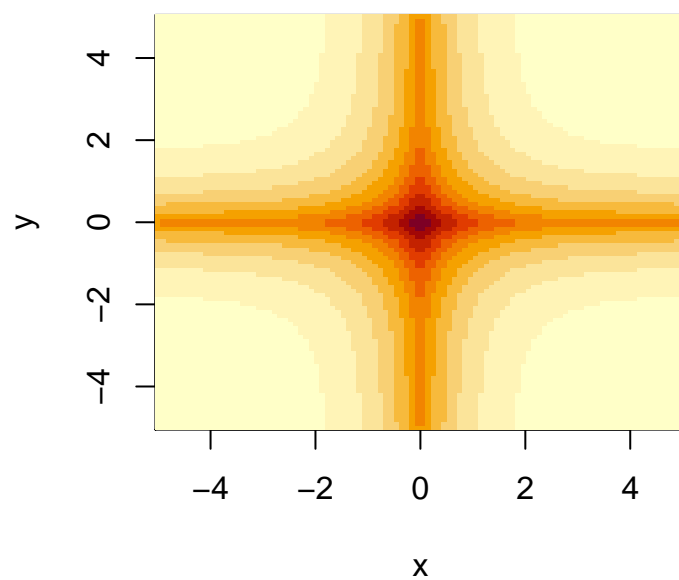
Spike and Slab and the Double Exponenti
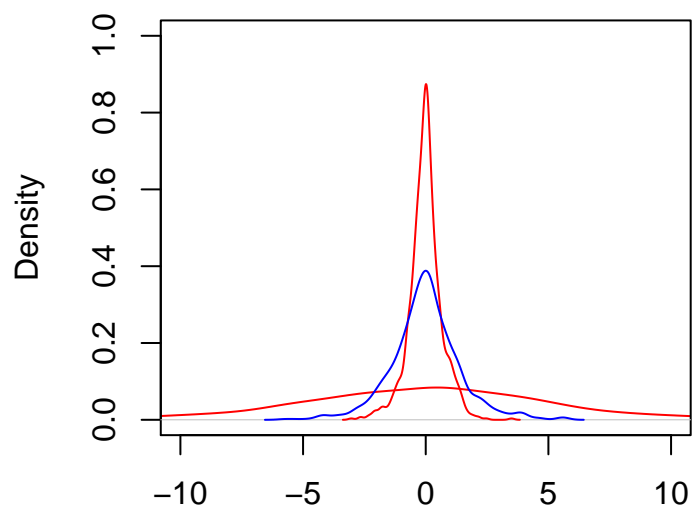
N = 1000   Bandwidth = 1.094

```
#contour
mat <- matrix(NA, ncol = 100, nrow = 100)
x <- y <- seq(-5, 5, length.out = 100)
for(i in 1:length(x))
  for(j in 1:length(y))
    mat[i,j] <- dlaplace(x[i], s = 1) + dlaplace(y[j], s = 1)

image(x, y, mat, main = "Double Exponential")
```

Double Exponential

c

We've effectively derived the posterior in part (a). Using the flat prior, the posterior is given by

$$p(\alpha, \beta | \mathbf{y}) \propto \exp\left(-\sum_{i=1}^{n} \exp(\alpha + \beta t_i) + \alpha \sum_{i=1}^{n} t_i + \beta \sum_{i=1}^{n} y_i t_i\right)$$

The sufficient statistics here are given by

$$\sum_{i=1}^{n} t_i, \sum_{i=1}^{n} y_i t_i, \sum_{i=1}^{n} \exp(\alpha + \beta t_i)$$

d

As this distribution is just the likelihood of a collection of Poisson data, this density is proper.

e

Here we get crude estimates of α, β using both linear model and estimates from Poisson regression models. Moreover, to avoid overflow errors, we center our t_i data. In that way, we do not lose precision in the computation schemes we use to generate samples from the posterior distribution.
We plot the contours of the posterior distribution as well as generate B samples $\{\hat{\alpha}^{(b)}, \hat{\beta}^{(b)}\}_{b=1}^{B}$

```r
# Read in Data
fatal_accidents <- c(24 , 25, 31, 31, 22, 21, 26, 20, 16, 22)
years <- 1976:1985
years.centered <- years - mean(years)

#crude linear models estimats
lm_coef <- coef(lm(fatal_accidents ~ years))
alpha_lm<- unname(lm_coef[1])
beta_lm<- unname(lm_coef[2])

#crude Poisson regression model estimates
pois_coef <- coef(glm(fatal_accidents ~ years, family = "poisson"))
alpha_pois<- unname(pois_coef[1])
beta_pois <- unname(pois_coef[2])

#print table
library(knitr)
df <- data.frame(alpha = c(alpha_lm, alpha_pois), beta = c(beta_lm, beta_pois))
rownames(df) <- c("Linear Model", "Poisson Model")
kable(df)
```

	alpha	beta
Linear Model	1848.26061	-0.9212121
Poisson Model	80.01559	-0.0388044

9

f

In the code below, we introduce the functions needed to make draws from the log posterior density.

```r
#here we set up the log posterior density
#this will only work for the scaled years! sum(exp(x) + exp(y)+ ... + exp(z)) giving overflows
lpost <- function(a, b) -sum(exp(a + b*years.centered))+ a*sum(fatal_accidents) + b*sum(years.centered*

lse <- function(x){
  m <- max(x); x <- x - m
  return(m + log(sum(exp(x[x > log(.Machine$double.eps)]))))
}

#grid around values of ML
int <- confint(glm(fatal_accidents~years.centered, family = "poisson"), level = .99)
```
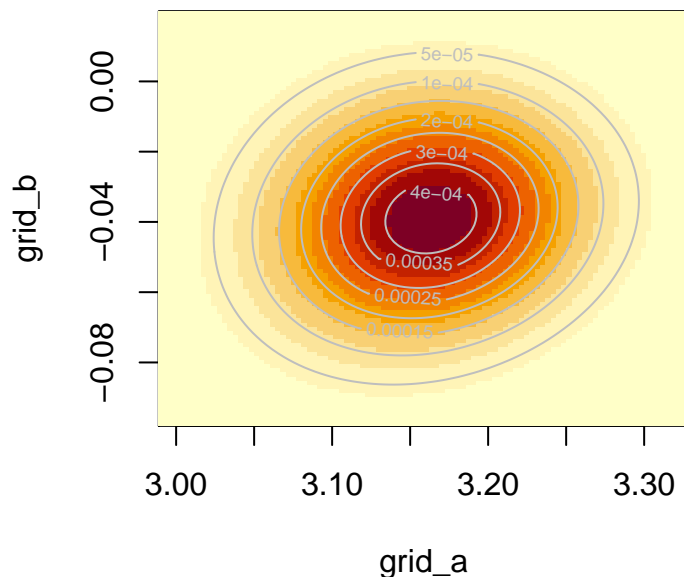
```
## Waiting for profiling to be done...
```

```r
M <- 100
grid_a <- seq(int[1,1],int[1,2], length.out = M)
grid_b <- seq(int[2,1],int[2,2], length.out = M)

#create contours
mat <- matrix(NA, ncol = M, nrow = M)
for(i in 1:M)
  for(j in 1:M)
    mat[i,j] <- lpost(grid_a[i],grid_b[j])

#normalize log_post
prob <- exp(mat - lse(mat))

#plot contours
image(grid_a, grid_b, prob)
contour(grid_a, grid_b, prob, add = TRUE, col = "grey")
```



```r
#sample from posterior grid
```

```
#get which spot on grid
ind_grid <- matrix(1:(M ^ 2), nrow = M)

#sampling function
sample_grid <- function(ns) {
  #apply over sampled indicies of the grid
  is <- sapply(sample.int(M ^ 2, ns, replace = TRUE, prob = prob),
              function (ind) which(ind_grid == ind, arr = TRUE)) # store grid indices
  cbind(grid_a[is[1,]], grid_b[is[2,]]) #return a and b values
}

#draw B samples
B <- 1000
samp <- sample_grid(B)
```

g

Next we use these samples to plot a histogram for the expected number of fatal accidents in the 1986. As we centered our data, this corresponds with $x = 5.5$. Moreover, our estimate of the expected number of fatal accidents is simply given by

$$\hat{\mu}|\hat{\alpha}, \hat{\beta} = \mathbb{E}[\mathbf{y}_{1986}|\alpha, \beta] = \exp(\hat{\alpha} + \hat{\beta} * 5.5)$$
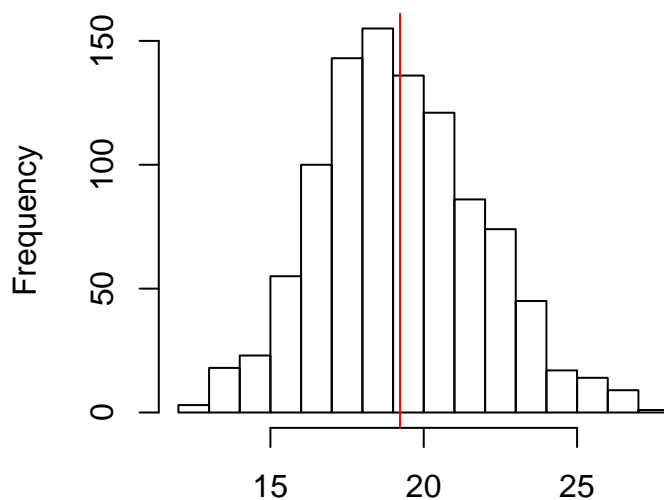
We plot the histogram of these estimates of each of our 1000 draws sampled in part f. In addition we add the posterior mean to this plot seen in red.

```
pred <- exp(samp[,1] + samp[,2]*5.5)
hist(pred, main = "Expected Fatal Accidents in 1986", xlab = " ")
abline(v = mean(pred), col = "red")
```

h

Finally, with the sampled $\hat{\lambda}_{1986}^{(b)} = \exp(\hat{\alpha}^{(b)} + 5.5 * \hat{\beta}^{(b)})$ for $b = 1, 2, \ldots 1000$, we can create a predictive distribution of the number of fatal accidents by sampling $\tilde{y}^{(b)}|\theta^{(b)} \sim \text{Pois}(\theta^{(b)})$. From here, we plot a histogram of the collection $\{\tilde{y}^{(b)}\}_{b=1}^{B}$.

```
#calculate different lambdas
lambdas <- apply(samp, 1, function(x) exp(x[1] + 5.5 * x[2]))

#sample y tildes
y_tilde <- sapply(lambdas, function(x) rpois(1, x))

#create histogram
hist(y_tilde, main = "Predictive Distribution of Number of Fatal Accidents", xlab = "", breaks = 15)
abline(v = mean(y_tilde), col = "red")
```

edictive Distribution of Number of Fatal Acc
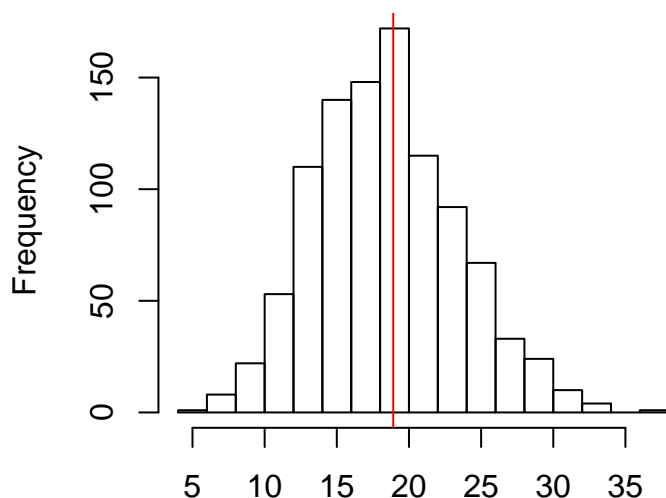
In addition, we get the 95% predictive interval as $(10, 30)$.

```
quantile(y_tilde, probs = c(.025, .975))
```

```
##  2.5% 97.5%
##    10    29
```

i

In this model, we use a flat prior so our posterior mimics the likelihood almost exactly. Indeed, we see that the model estimates cluster almost perfectly around the ML estimates. Moreover, we see that the contours are mostly elliptical as guaranteed by GLM theory. If we were to use another prior, the shape of this distribution may change. In particular, for priors encouraging sparse solutions, as those discussed in (b), we may expect that the model estimates would center closer to the mean. This may also have effects on the variance of our final posterior distribution.

Exercise (4)

Suppose that $\mathbf{X}|\mu, \sigma^2 \sim N(\mu \mathbf{1}_n, \sigma^2 \mathbf{I}_{n \times n})$ and assume $\mu|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ and $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$.

(a)

In this exercise we look to derive the form of $p(\sigma^2|\mu)$. To that end, consider

$$p(\sigma^2|\mu) \propto p(\mu|\sigma^2)p(\sigma^2)$$

$$\propto (\sigma^2/\kappa_0)^{-1/2}\exp\left(-\frac{1}{2\sigma^2/\kappa_0}(\mu-\mu_0)^2\right)(\sigma^2)^{-\frac{\nu_0}{2}-1}\exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right)$$

$$= (\sigma^2)^{-\frac{\nu_0+1}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left(\kappa_0(\mu-\mu_0)^2+\nu_0\sigma_0^2\right)\right]$$

$$= (\sigma^2)^{-\frac{\nu_0+1}{2}-1}\exp\left[-\frac{1}{2\sigma^2}\left(\frac{\kappa_0(\mu-\mu_0)^2+\nu_0\sigma_0^2}{\nu_0+1}(\nu_0+1)\right)\right]$$

Therefore, we conclude

$$\sigma^2|\mu \sim \text{Inv-}\chi^2\left(\nu_\mu := \nu_0+1, \sigma_\mu^2 := \frac{\kappa_0(\mu-\mu_0)^2+\nu_0\sigma_0^2}{\nu_0+1}\right)$$

(b)

Next, we look to derive $p(\mathbf{X}|\mu)$. We do so by introducing σ^2 and then marginalizing it out of the appropriate intergral.

$$\mathbb{P}(\mathbf{X}|\mu) = \int \mathbb{P}(\mathbf{X},\sigma^2|\mu)d(\sigma^2) = \int \mathbb{P}(\mathbf{X}|\mu,\sigma^2)p(\sigma^2|\mu)d(\sigma^2)$$

From here, we can use the results of part (a) to evaluate $p(\sigma^2|\mu)$ and then the full integral.

$$\mathbb{P}(\mathbf{X}|\mu) = \int \mathbb{P}(\mathbf{X}|\mu,\sigma^2)p(\sigma^2|\mu)d(\sigma^2)$$

$$\propto \int (\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{X}-\mu)^T(\mathbf{X}-\mu)\right\}(\sigma^2)^{-\frac{\nu_\mu}{2}-1}\exp\left(-\frac{\nu_\mu\sigma_\mu^2}{2\sigma^2}\right)d(\sigma^2)$$

$$= \int (\sigma^2)^{-\frac{n+\nu_\mu}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu)+\nu_\mu\sigma_\mu^2\right]\right\}d(\sigma^2)$$

We recognize this intergral as the kernel of an Inverse-χ^2 distribution with parameters

$$\text{Inv}-\chi^2\left(\nu_x = n+\nu_\mu, \sigma_x^2 := \frac{(\mathbf{X}-\mu)^T(\mathbf{X}-\mu)+\nu_\mu\sigma_\mu^2}{n+\nu_\mu}\right)$$

Therefore, this integral integrates to one over its integrating constant and can be written as

$$\int (\sigma^2)^{-\frac{n+\nu_\mu}{2}-1} \exp\left\{ -\frac{1}{2\sigma^2}\left[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \nu_\mu \sigma_\mu^2 \right] \right\} d(\sigma^2) = \left[\frac{(\sigma_x^2 \nu_x/2)^{\nu_x/2}}{\Gamma(\nu_x/2)} \right]^{-1}$$

$$\propto (\sigma_x^2 \nu_x)^{-\nu_x/2}$$

$$= \left[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \nu_\mu \sigma_\mu^2 \right]^{-\frac{n+\nu_\mu}{2}}$$

$$= \left[\frac{(\mathbf{X}-\mu)^T(\mathbf{X}-\mu)}{\nu_\mu \sigma_\mu^2} + 1 \right]^{-\frac{n+\nu_\mu}{2}}$$

Thus $\mathbf{X}|\mu \sim t_{\nu_\mu}(\mu \mathbf{1}_n, \sigma_\mu^2 \mathbf{1}_{n\times n})$.

(c)

Finally, we look to derive $p(\mu|\mathbf{X})$. To that end consider the following

$$p(\mu|\mathbf{X}) = \int p(\mu, \sigma^2|\mathbf{X}) d(\sigma^2)$$

$$\propto \int p(\mathbf{X}|\mu,\sigma^2) p(\mu|\sigma^2) p(\sigma^2) d(\sigma^2)$$

$$= \int (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) \right) (\sigma^2/\kappa_0)^{-1/2} \exp\left(-\frac{1}{2\sigma^2/\kappa_0}(\mu-\mu_0)^2 \right) (\sigma^2)^{-\frac{\nu_0}{2}-1} \exp(-\frac{\nu_0 \sigma_0^2}{2\sigma^2}) d(\sigma^2)$$

$$= \int (\sigma^2)^{-\frac{n+1+\nu_0}{2}-1} \exp\left(-\frac{1}{2\sigma^2}\left[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2 \right] \right) d(\sigma^2)$$

We recogonize this as the kernel of an Inverse-χ^2 distribution with parameters

$$\text{Inv} - \chi^2\left(\nu_m := n+1+\nu_0, \sigma_m^2 = \frac{(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2}{n+1+\nu_0} \right)$$

Following the same process as above, this kerenel integrates to the inverse of its normalizing constant.

$$p(\mu|\mathbf{X}) \propto \left[\frac{(\sigma_m^2 \nu_m)^{\nu_m/2}}{\Gamma(\nu_n/2)} \right]^{-1} \propto (\sigma_m^2 \nu_m)^{-\nu_m/2}$$

$$\propto \left[(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2 \right]^{-\frac{n+1+\nu_0}{2}}$$

$$\propto \left[\frac{(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \kappa_0(\mu-\mu_0)^2}{\nu_0\sigma_0^2} + 1 \right]^{-\frac{n+1+\nu_0}{2}}$$

Through expanding $(\mathbf{X}-\mu)^T(\mathbf{X}-\mu) + \kappa_0(\mu-\mu_0)^2$ we can complete the square to see that $\mu|\mathbf{X} \sim t_{\nu_v}(\mu_n, \sigma_n^2/\kappa_n)$
for $\kappa_n = n + \kappa_0$, $\mu_n = \frac{n\bar{x}+\kappa_0\mu_0}{\kappa_n}$, $\sigma_n^2 = \frac{S_x^2 + \frac{n\kappa_0}{\kappa_n}(\bar{x}-\mu_0)^2 + \nu_0\sigma_0^2}{\kappa_n}$.