Bayesian Assignment #2

Anthony Sisti

3/3/2020

BDA Question 3.2

Let $Y_b = (Y_{b1}, Y_{b2}, Y_{b3})$, and $Y_a = (Y_{a1}, Y_{a2}, Y_{a3})$ be multinomial vectors of survey counts for Bush, Dukakis and "other", before and after the debate, respectively. Let π_b and π_a be their corresponding probability vectors.

Building off of the strategy employed in the pre-election polling example found on pages 69-70, we set a Dirichlet prior with $\alpha_1 = \alpha_2 = \alpha_3 = 1$ (non-informative uniform) on both π_b and π_a. Since we know Dirichlet is conjugate for multinomial distributions, we use established properties to determine

$$\begin{aligned} \pi_b|y &\sim \text{Dir}(295, 308, 39) \\ \pi_a|y &\sim \text{Dir}(289, 333, 20) \end{aligned}$$

For this question, we are interested in the quantity $\gamma_a - \gamma_b$ (re-labeled from α), the proportion of decided voters who supported Bush before the debate, subtracted from the proportion of decided voters who supported Bush after the debate. We define,

$$\begin{aligned} \gamma_a &= \frac{\pi_{a1}}{\pi_{a1} + \pi_{a2}} \\ \gamma_b &= \frac{\pi_{b1}}{\pi_{b1} + \pi_{b2}} \end{aligned}$$

In order to obtain this quantity, we sample 10,000 draws from the Dirichlet distributions above and compute $\gamma_a - \gamma_b$ for each draw. The posterior probability of a shift toward Bush, measured by the proportion of $\gamma_a - \gamma_b$ values greater than 0, is approximately 0.19. The histogram of these samples can be seen seen below.
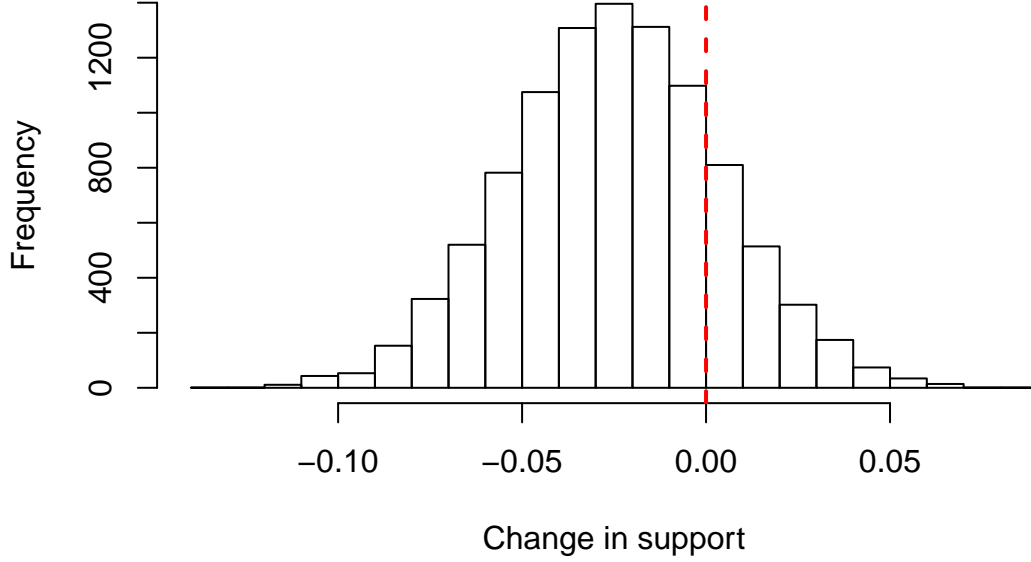
```
set.seed(32)

library(DirichletReg)

drawsb <- rdirichlet(10000, c(295,308,39))
drawsa <- rdirichlet(10000, c(289,333,20))

alpha_1 <- drawsb[,1]/(drawsb[,1] + drawsb[,2])
alpha_2 <- drawsa[,1]/(drawsa[,1] + drawsa[,2])

final <- alpha_2-alpha_1

hist(final, main = "Shift in Bush Support, 10K samples", xlab = "Change in support",
     breaks = 25)
abline(v = 0, col = "red", lty = 2, lwd =2)
```

Shift in Bush Support, 10K samples

BDA Question 3.3

a. We assume $p(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t) \propto 1$, so

$$p(\mu_c, \mu_t, \log \sigma_c, \log \sigma_t | \mathbf{y}) \propto p(\mathbf{y} | \mu_c, \mu_t, \log \sigma_c, \log \sigma_t) \prod_{i=1}^{N} p(y_i | \mu_c, \mu_t, \log \sigma_c, \log \sigma_t)$$

Where N is the total observations. Note, the control group is only dependent on μ_c and $\log \sigma_c$, and the treatment group on μ_t and $\log \sigma_t$. Thus, we can look at the treatment and control group separately and find marginal for μ_t and μ_c when $p(\mu_t, \log \sigma_t) \propto 1$ and $p(\mu_c, \log \sigma_c) \propto 1$. Page 66 of the text book explains that when $p(\mu, \log \sigma) \propto 1$, or equivalently $p(\mu, \sigma^2) \propto 1/\sigma^2$,

$$u | \mathbf{y} \sim t_{n-1}(\bar{y}, s^2/n)$$

.

Thus,

$$u_c | \mathbf{y} \sim t_{31}\left(1.01, \frac{0.24^2}{32}\right)$$
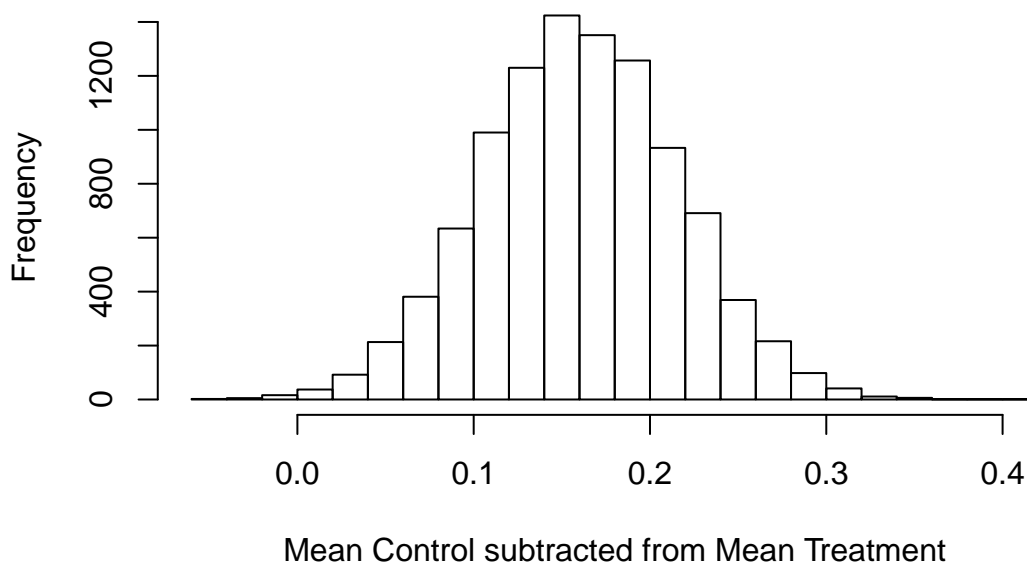$$u_t | \mathbf{y} \sim t_{35}\left(1.17, \frac{0.2^2}{36}\right)$$

b. Sampling from the t-distributions we found in (a), we can plot the following histogram for $\mu_t - \mu_c$ with corresponding 95% posterior interval $[0.05, 0.27]$.

```
set.seed(33)
mu_c <-  (0.24/sqrt(32))*rt(10000,31) +1.013
mu_t <- (0.20/sqrt(36))*rt(10000,35) + 1.173

hist(mu_t-mu_c, xlab = "Mean Control subtracted from Mean Treatment",
     main = "Histogram of Difference between Sampled Means", breaks = 20)
```


BDA Question 3.5

a. We assume $p(\mu, \log \sigma) \propto 1$. Then, referencing section 3.2, we have

$$p(\mu, \sigma^2 | \mathbf{y}) \propto p(\mathbf{y} | \mu, \sigma^2) p(\mu, \sigma^2)$$
$$\propto \sigma^{-n-2} \exp \left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right)$$

In the context of this problem, $n = 5, s^2 = 1.3$ and the mean of the observations is 10.4. However, we also know from 3.2 that

$$\mu | \sigma^2, \mathbf{y} \sim N(\bar{y}, \sigma^2/n)$$
$$\sigma^2 | \mathbf{y} \sim Inv - \chi^2(n-1, s^2),$$

b. We know the posterior will be of the form

$$p(\mu, \sigma^2|\mathbf{y}) \propto \frac{1}{\sigma^2} \prod_{i=1}^{n} p(y_i|\mu, \sigma^2).$$

In the rounded case, y_i can only be observed as an integer value so, with some abuse of notation, we write ,

$$p(y_i \text{rounded to} a|\mu, \sigma^2) = p(a - 0.5 \le y_i \le a + 0.5|\mu, \sigma^2),$$

where a is an integer. Recall, y_i is generated by a $N(\mu, \sigma^2)$ so we can write it as $\sigma z_i + \mu$ where $z_i \sim N(0, 1)$. Thus,

$$\begin{aligned} p(y_i = a|\mu, \sigma^2) &= p(a - 0.5 \le y_i \le a + 0.5|\mu, \sigma^2) \\ &= p\left(\frac{a - 0.5 - \mu}{\sigma} \le z_i \le \frac{a + 0.5 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{a + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{a - 0.5 - \mu}{\sigma}\right) \end{aligned}$$

Thus, from above,

$$\begin{aligned} p(\mu, \sigma^2|\mathbf{y}) &\propto \frac{1}{\sigma^2} \prod_{i=1}^{n} p(y_i|\sigma^2, \mu) \\ &\propto \frac{1}{\sigma^2} \prod_{i=1}^{n} \Phi\left(\frac{y_i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{y_i - 0.5 - \mu}{\sigma}\right) \end{aligned}$$

c. We sample from the posterior as given in part (a) to obtain summary statistics for μ and σ corresponding to when the measurements were assumed to be unrounded . We present them, and the contour plot, below.

```
posteriorNorm = function (theta, data)
{
    mu = theta[1]
    sig2 = theta[2]
    logf = function(y, mu, sig2) -(y - mu)^2/2/sig2 - log(sig2)/2
    z = sum(logf(data, mu, sig2))
    z = z - log(sig2)
    return(z)
}


numTicks = 500
XX = seq(3, 15, len = numTicks)
YY = seq(0.001,10,length=numTicks)
X = outer(XX, rep(1, numTicks))
Y = outer(rep(1, numTicks), YY)
numTicks2 = numTicks^2
Z = apply(cbind(X[1:numTicks2], Y[1:numTicks2]), 1, posteriorNorm, data=c(10,10,12,11,9))
```

```
Z = exp(Z - max(Z))
Z = matrix(Z, c(numTicks, numTicks))



pulls <- 2000
sampleZ <- apply(Z,1,sum)
meanI <- sample(1:length(XX), pulls, replace=T,
                prob = sampleZ)

mus <- XX[meanI]
vars <- rep(NA, pulls)
for (i in 1:pulls) {
  vars[i] <- sample(YY, 1, prob = Z[meanI[i],],replace = F)
}



summary_NR <- as.data.frame(rbind(c(mean(mus), sd(mus),
                                    quantile(mus, c(0.05,0.25,0.5,0.75,0.95))),
       c(mean(sqrt(vars)),sd(sqrt(vars)),quantile(sqrt(vars), c(0.05,0.25,0.5,0.75,0.95))))))

summary_NR <- as.data.frame(summary_NR)

rownames(summary_NR) <- c("mu", "sigma")
colnames(summary_NR)[1:2] <- c("mean", "sd")
summary_NR <- round(summary_NR,2)
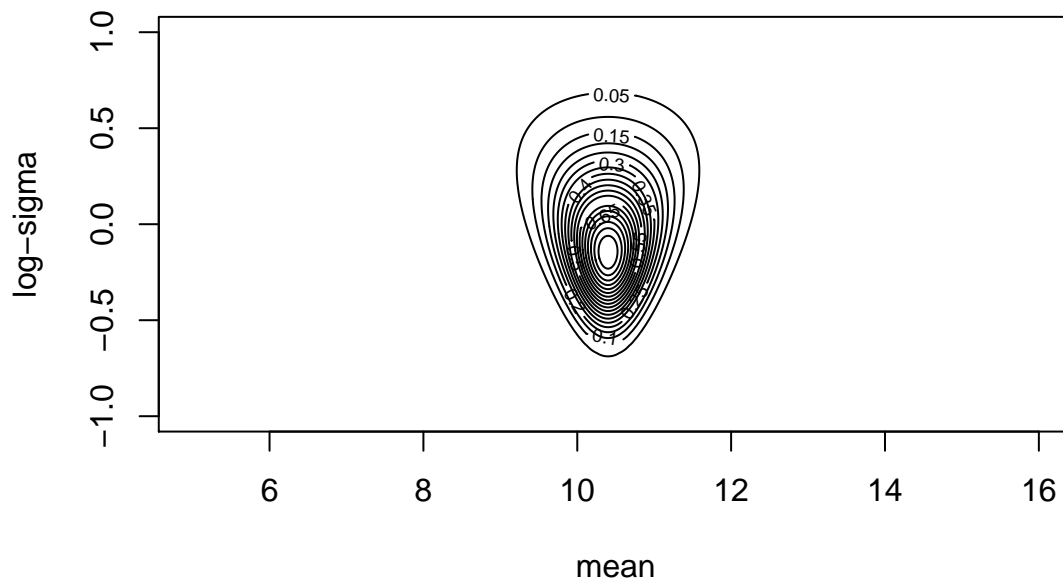```

## Posterior assuming no rounding

Table 1: Posterior assuming no rounding

|  | mean | sd | 5% | 25% | 50% | 75% | 95% |
|---|---|---|---|---|---|---|---|
| mu | 10.38 | 0.62 | 9.32 | 10.02 | 10.38 | 10.74 | 11.37 |
| sigma | 1.35 | 0.50 | 0.75 | 0.98 | 1.23 | 1.60 | 2.37 |

We now empirically sample $(\mu, \sigma)$ from the posterior that we wrote in part (b). We display the contour plot and summary statistics below.

```
posteriorNorm = function (theta, data)
{
    mu = theta[1]
    sig2 = theta[2]
    logf = function(y, mu, sig2){log(pnorm((y+0.5-mu)/sqrt(sig2)) -
                                    pnorm((y-0.5-mu)/sqrt(sig2)))}
    z = sum(logf(data, mu, sig2))
    z = z - log(sig2)
    return(z)
}



numTicks = 500
XX = seq(3, 18, len = numTicks)
YY = seq(0.1,12,length=numTicks)
X = outer(XX, rep(1, numTicks))
Y = outer(rep(1, numTicks), YY)
numTicks2 = numTicks^2
Z = apply(cbind(X[1:numTicks2], Y[1:numTicks2]), 1, posteriorNorm, data=c(10,10,12,11,9))
Z = Z - max(Z)
Z = matrix(exp(Z), c(numTicks, numTicks))




pulls <- 2000
sampleZ <- apply(Z,1,sum)
meanI <- sample(1:length(XX), pulls, replace=T,
                prob = sampleZ)

mus <- XX[meanI]
vars <- rep(NA, pulls)
for (i in 1:pulls) {
  vars[i] <- sample(YY, 1, prob = Z[meanI[i],],replace = F)
}


summary_R <- as.data.frame(rbind(c(mean(mus), sd(mus),
                                quantile(mus, c(0.05,0.25,0.5,0.75,0.95))),
     c(mean(sqrt(vars)),sd(sqrt(vars)),quantile(sqrt(vars), c(0.05,0.25,0.5,0.75,0.95)))))

summary_R <- as.data.frame(summary_R)
```

```r
rownames(summary_R) <- c("mu", "sigma")
colnames(summary_R)[1:2] <- c("mean", "sd")
summary_R <- round(summary_R,2)
```
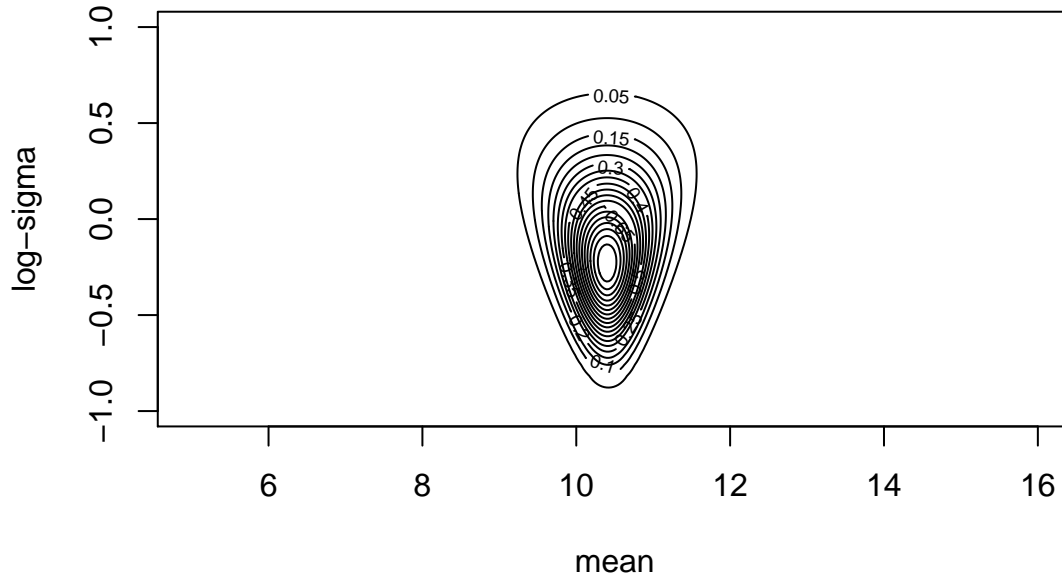
Posterior assuming the values are rounded

Table 2: Posterior assuming rounding

	mean	sd	5%	25%	50%	75%	95%
mu	10.41	0.63	9.34	10.06	10.42	10.79	11.45
sigma	1.31	0.55	0.68	0.90	1.18	1.57	2.45

When comparing the two distributions both visually and statistically, they are very similar. The empirical means and standard deviations for both μ and σ remain close regardless of whether rounding is assumed or not. The main visual difference is seen in the lower values for σ where, in the contour plot, it does appear that the distribution stretches to smaller values.

d. We know that
$$p(z_i|\mu,\sigma) \sim N(\mu,\sigma),$$
but given that we have observed the y_i values, which are rounded, we know that the z_i must have been in $[y_i - 0.5, y_i + 0.5]$. In order to determine the posterior mean of $(z_1 - z_2)^2$ we will sample z_i values from $[y_i - 0.5, y_i + 0.5]$ for for each draw of μ and σ that we sampled in part (c). Once we have these draws for z_1 and z_2, we will compute $(z_1 - z_2)^2$ for each pair.

```r
set.seed(35)
round <- c(10,10,12,11,9)
pulls <- 2000
sds <- sqrt(vars)

z_1 <- rep(NA, pulls)
```

```
L <- pnorm(round[1]-.5, mus, sds)
U <- pnorm(round[1]+.5, mus, sds)
z_1 <- qnorm(L + runif(pulls, min = 0, max = U-L), mus, sds)

z_2 <- rep(NA, pulls)

L <- pnorm(round[2]-.5, mus, sds)
U <- pnorm(round[2]+.5, mus, sds)
z_2 <- qnorm(L + runif(pulls, min = 0, max = U-L), mus, sds)


mean((z_1-z_2)^2)
```

`## [1] 0.1603578`

As is seen above, the posterior mean of $(z_1 - z_2)^2$ is approximately 0.16 .

BDA Question 3.8

a. Since the y_i and z_i values are proportions, we choose a Beta distribution for modeling. In the Beta likelihood, $\theta_y = (\alpha_y, \beta_y)$ and $\theta_z = (\alpha_z, \beta_z)$. These are the parameters corresponding to y and z respectively. We have,

$$p(y_i|\theta_y) \propto y_i^{\alpha_y-1}(1-y_i)^{\beta_y-1}$$
$$p(z_i|\theta_z) \propto z_i^{\alpha_z-1}(1-z_i)^{\beta_z-1}$$

b. A variety of priors that restrict α and β to be positive are possible for this analysis. We decide to set

$$p(\alpha_y, \beta_y, \alpha_z, \beta_z) \propto \left(\frac{\mathbf{1}_{\{\alpha_y \in [0, 2.36]\}}}{2.36}\right)\left(\frac{\mathbf{1}_{\{\beta_y \in [0, 21.87]\}}}{21.87}\right)\left(\frac{\mathbf{1}_{\{\alpha_z \in [0, 2.88]\}}}{2.88}\right)\left(\frac{\mathbf{1}_{\{\beta_z \in [0, 11.38]\}}}{11.38}\right)$$

.

This indicates that each parameter is distributed uniformly and is mutually independent from the others. Each uniform distribution is bounded below by 0 and above by the method of moments estimators for the parameter, using the first and second moment from the sampled data, assuming a Beta distribution. For example, in order to obtain the upper bound for the uniform priors on α_y and β_y we solved the system.

$$\frac{\alpha_y}{\alpha_y + \beta_y} = \bar{y} = 0.1961$$

$$\frac{\alpha_y \beta_y}{(\alpha_y + \beta_y)^2(\alpha + \beta + 1)} + \left(\frac{\alpha_y}{\alpha_y + \beta_y}\right)^2 = \bar{y^2} = 0.04848$$

Using Wolphram Alpha, we obtain $\hat{\alpha}_y = 2.36$ and $\hat{\beta}_y = 21.87$. We followed a similar procedure to obtain the z parameters.

c / d . We combine parts (c) and (d). Given the prior we chose, the posterior has the form

$$p(\theta_y, \theta_z | \mathbf{y}, \mathbf{z}) \propto \left(\frac{\mathbf{1}_{\{\alpha_y \in [0,2.36]\}}}{2.36} \right) \left(\frac{\mathbf{1}_{\{\beta_y \in [0,21.87]\}}}{21.87} \right) \left(\frac{\mathbf{1}_{\{\alpha_z \in [0,2.88]\}}}{2.88} \right) \left(\frac{\mathbf{1}_{\{\beta_z \in [0,11.38]\}}}{11.38} \right) *$$

$$\left(\prod_{i=1}^{10} y_i^{\alpha_y - 1} (1 - y_i)^{\beta_y - 1} \prod_{j=1}^{8} z_i^{\alpha_z - 1} (1 - z_i)^{\beta_z - 1} \right)$$

$$\propto \left(\left(\frac{\mathbf{1}_{\{\alpha_y \in [0,2.36]\}}}{2.36} \right) \left(\frac{\mathbf{1}_{\{\beta_y \in [0,21.87]\}}}{21.87} \right) \prod_{i=1}^{10} y_i^{\alpha_y - 1} (1 - y_i)^{\beta_y - 1} \right) *$$

$$\left(\left(\frac{\mathbf{1}_{\{\alpha_z \in [0,2.88]\}}}{2.88} \right) \left(\frac{\mathbf{1}_{\{\beta_z \in [0,11.38]\}}}{11.38} \right) \prod_{j=1}^{8} z_i^{\alpha_z - 1} (1 - z_i)^{\beta_z - 1} \right)$$

$$\propto p(\mathbf{y}|\theta_y) p(\theta_y) p(\mathbf{z}|\theta_z) p(\theta_z)$$

In order to sample from this posterior, we calculated the posterior density at each possible point of α and β (separately for y and z). Once we had the density values over the (α, β) grid, we "integrated out" β by summing over the columns of the grid and sampling from the marginal posterior for α. Once we had these α samples, we sampled from the distribution on β corresponding to the α value we had obtained. We can write this as

$$p(\alpha_y, \beta_y | \mathbf{y}) \propto p(\beta_y | \alpha_y, \mathbf{y}) p(\alpha_y | \mathbf{y})$$

We know that

$$\mu_y = E[y|\theta_y] = \frac{\alpha_y}{\alpha_y + \beta_y}$$

and similarly for μ_z. Using the 1000 samples we drew from our posterior distribution, we compute $mu_y - \mu_z$ for each sample and plot them in a histogram below.

```
set.seed(38)

y_props <- c(16/(16+58), 9/(9+90), 10/(10+48), 13/(13+57),
             19/(19+103), 20/(20+57), 18/(18+86), 17/(17+112),
             35/(35+273), 55/(55+64))

z_props <- c(12/(12+113), 1/(1+18), 2/(2+14), 4/(4+44),
             9/(9+208), 7/(7+67), 9/(9+29), 8/(8+154))



alpha_y <- seq(0, 2.88, length.out = 4000)
beta_y <- seq(0, 11.8378, length.out=4000)
post_y <- matrix(0,nrow=4000,ncol=4000)
for (i in 1:4000) {
for (j in 1:4000) {
  post_y[i,j] <- (1/(2.88))*(1/(11.8378))*prod(y_props^(alpha_y[i]-1))*
    prod((1-y_props)^(beta_y[j]-1))

  }
```

```r
}

pulls <- 1000
sampley <- apply(post_y,1,sum)
alphaI <- sample(1:length(alpha_y), pulls, replace=T,
                  prob = sampley)

alpha_ys <- alpha_y[alphaI]
beta_ys <- rep(NA, pulls)
for (i in 1:pulls) {
  beta_ys[i] <- sample(beta_y, 1, prob = post_y[alphaI[i],])
}



Post_mean_ys <- alpha_ys/(alpha_ys + beta_ys)



alpha_z <- seq(0, 2.36, length.out=4000)
beta_z <- seq(0, 21.8781, length.out=4000)
post_z <- matrix(0,nrow=4000,ncol=4000)
for (i in 1:4000) {
for (j in 1:4000) {
  post_z[i,j] <- (1/(2.36))*(1/(21.8781))*prod(z_props^(alpha_z[i]-1))*
    prod((1-z_props)^(beta_z[j]-1))

 }
}


pulls <- 1000
samplez <- apply(post_z,1,sum)
alphaI <- sample(1:length(alpha_z), pulls, replace=T,
                  prob = samplez)

alpha_zs <- alpha_z[alphaI]
beta_zs <- rep(NA, pulls)
for (i in 1:pulls) {
  beta_zs[i] <- sample(beta_z, 1, prob = post_z[alphaI[i],])
}

Post_mean_zs <- alpha_zs/(alpha_zs + beta_zs)



hist(Post_mean_ys-Post_mean_zs, breaks = 30,
     main = "Expected differnce in proportions of Bicycle Traffic",
     xlab = "mu_y - mu_z")
```
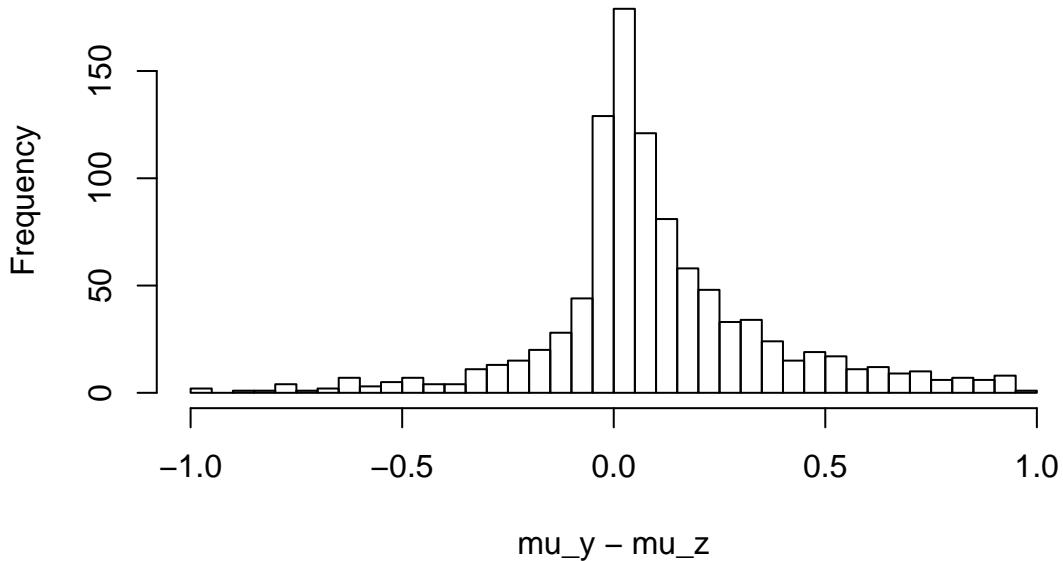
Expected differnce in proportions of Bicycle Traffic

The mean of this distribution is 0.101 and the proportion of $\mu_y > \mu_z$ is 70%.

BDA 3.12

a. The most obvious choice of a non-informative prior for (α, β) is

$$p(\alpha, \beta) \propto 1$$

.

This is a uniform prior on $(-\infty, \infty) \times (-\infty, \infty)$, indicating that all values on the real line are equally likely a priori. If we are working under the invariance principle we could use Jeffery's prior:

$$p(\alpha, \beta) \propto \det(I(\alpha, \beta))^{1/2}$$

This would allow us to maintain the same prior form on any one-to-one function of α and β. For this problem, we decide to choose $p(\alpha, \beta) \propto 1$.

b. In the case that we had knowledge about the possible values of (α, β) from past studies on fatal plane accidents, it would be reasonable to integrate this knowledge into a bivariate normal distribution. One could also add a correlation structure if this had been observed in the past. However, we have no prior knowledge on the subject, and especially these coefficients. In order to create an informative prior, we can use crude linear regression estimates and have

$$p(\alpha, \beta) \sim N\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}, \begin{pmatrix} S_\alpha^2 & 0 \\ 0 & S_\beta^2 \end{pmatrix} \right)$$

Where S represents the coefficient standard error estimates. In part (e) of this question, we run this regression, where `Year` is on a 1-10 scale rather than 1976-1985. Taking estimates from that regression output, we

11

find $\hat{\alpha} = 28.7$, $\hat{\beta} = -0.92$, $S_\alpha = 2.75$ and $S_\beta = 0.443$. Below, we create contours for this bivariate normal distribution.

```
alpha_points <- seq(25,32, length.out=100)
beta_points <- seq(-3,1, length.out=100)
z <- matrix(0,nrow=100,ncol=100)
mu_prior <- c(28.7, -0.92)
sigma_prior <- matrix(c(2.75,0,0,0.443),nrow=2)
for (i in 1:100) {
for (j in 1:100) {
z[i,j] <- mvtnorm::dmvnorm(c(alpha_points[i],beta_points[j]),
mean=mu_prior,sigma=sigma_prior)
}
}
contour(alpha_points,beta_points,z)
title(xlab="Alpha",ylab="Beta", main = "Informative Prior Contour")
```

Informative Prior Contour
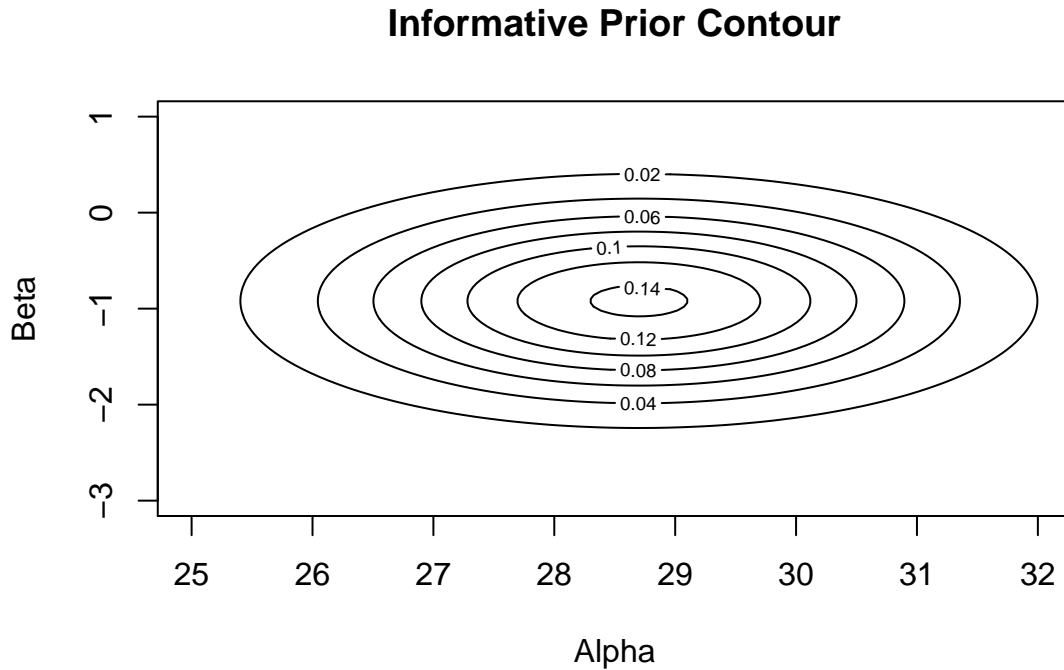
c. Assuming our non-informative prior,

$$
\begin{aligned}
p(\alpha,\beta|\mathbf{y}) &\propto p(\mathbf{y}|\alpha,\beta)p(\alpha,\beta) \\
&\propto p(\mathbf{y}|\alpha,\beta) \\
&\propto \prod_{i=1}^{10} \frac{(\alpha+\beta t_i)^{y_i}\, e^{-(\alpha+\beta t_i)}}{y_i!} \\
&\propto \prod_{i=1}^{10} (\alpha+\beta t_i)^{y_i}\, e^{-(\alpha+\beta t_i)}
\end{aligned}
$$

This density cannot be factored. Assuming that the associated time t_i is known for each value y_i, the sufficient statistic for (α,β) is the order statistics. If the t_i value associated with a given y_i is lost when reducing to the order statistics, no data reduction is possible because the (y_i, t_i) pairs are necessary.

d. We want

$$\int\int_{\alpha+\beta t_i>0} \prod_{i=1}^{10}(\alpha+\beta t_i)^{y_i}\,e^{-(\alpha+\beta t_i)}d\alpha d\beta < 0.$$

Let $Z_i = \alpha + \beta t_i \in (0,\infty)$. Then integrating each Z_i over the region where $\alpha + \beta t_i > 0$ is equivalent to finding

$$\int_0^\infty \ldots \int_0^\infty \prod_{i=1}^{10}(Z_i)^{y_i}\,e^{-(Z_i)}dZ_1...dZ_{10} < 0.$$

Notice that the density inside the integral is a product, where the ith term only depends on Z_i. Thus, we can rewrite this integral as

$$\int_0^\infty \ldots \int_0^\infty \prod_{i=1}^{10}(Z_i)^{y_i}\,e^{-(Z_i)}dZ_1...dZ_i = \prod_{i=1}^{10}\int_0^\infty (Z_i)^{y_i}\,e^{-(Z_i)}dZ_i$$

.

Note, each integral inside the product is the the kernel of a $\mathrm{Gamma}(y_i + 1, 1)$ density. Therefore, we know each of these 10 integrals will be finite, and so will their product. This proves that the posterior is proper.

e. Below, we run the crude regression we reference in part (a), the standard errors and means from this regression are re stated below.

```
Crash <- data.frame(accidents = c(24,25,31,31,22,21,26,20,16,22),
                    Year = 1:10)

crude_fit <- lm(accidents ~ Year, data = Crash)
```

Table 3: Crude Alpha and Beta estimates

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	28.867	2.749	10.499	0.000
Year	-0.921	0.443	-2.079	0.071

f. We sample from the posterior written in part (c) and draw its contours below.

```
posteriorPois = function(theta, data){
    alpha = theta[1]
    beta = theta[2]
    check = alpha + beta*data[,2]
    if(min(check) < 0 ){
      return(0)
    }else{
    f = function(y, t, a, b){return((a+b*t)^y * exp(-(a+b*t)))}
    z = 1
    for(i in 1:nrow(data)){
      z = z * f(data[i,1], data[i,2], alpha, beta)
```

```
    }
    return(z)}
}



numTicks = 125
XX = seq(20, 40, len = numTicks)
YY = seq(-3,1,length=numTicks)
X = outer(XX, rep(1, numTicks))
Y = outer(rep(1, numTicks), YY)
numTicks2 = numTicks^2
Z = apply(cbind(X[1:numTicks2], Y[1:numTicks2]), 1, posteriorPois, data=Crash)
Z = matrix(Z, c(numTicks, numTicks))
contour(XX, YY, Z, lwd = 1)
title(ylab="Beta",xlab="Alpha", main = "Posterior for Non-informative Prior")
```

Posterior for Non−informative Prior
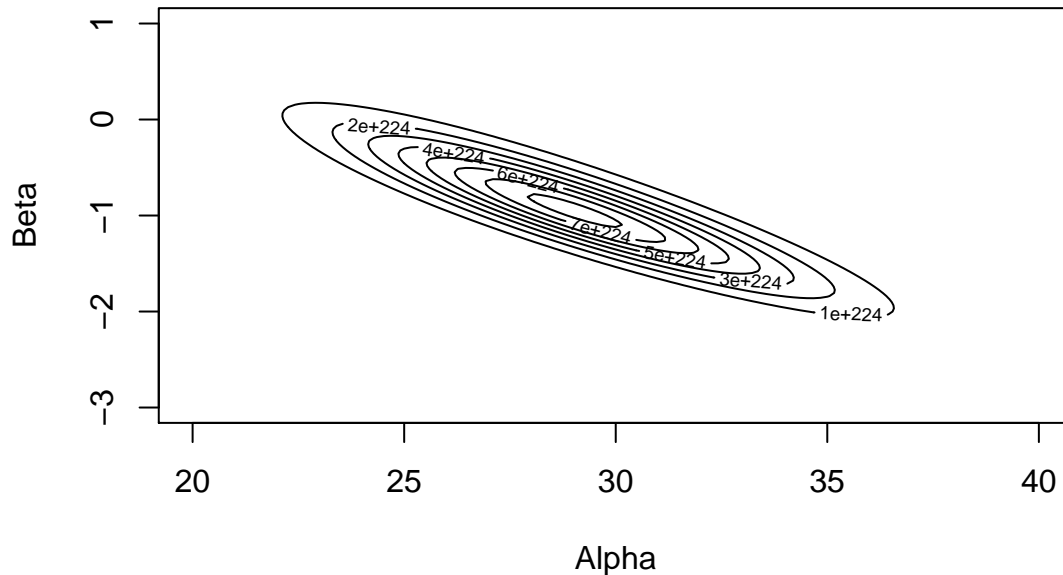

```
set.seed(312)

pulls <- 2000
sampleZ <- apply(Z,1,sum)
alphaI <- sample(1:length(XX), pulls, replace=T,
                 prob = sampleZ)

alphas <- XX[alphaI]
betas <- rep(NA, pulls)
for (i in 1:pulls) {
  betas[i] <- sample(YY, 1, prob = Z[alphaI[i],],replace = F)
}
```

g. Using these samples drawn from the posterior we plot a histogram of $\alpha + \beta * 11$ which corresponds to the the number of fatal accidents in 1986.

```
hist(alphas+betas*11, breaks = 20,main =  "Expected number of Fatal Accidents in 1986", xlab = "Expecte
```

Expected number of Fatal Accidents in 1986

The mean of this distribution is 18.8.

h. In order to create a simulation of draws and obtain a 95% predictive interval, we sample α and β values from the posterior and plug them into our likelihood, and pull a random sample from the conditional likelihood with those values plugged in. We do this 3000 times. We plot the histogram of this distribution below.
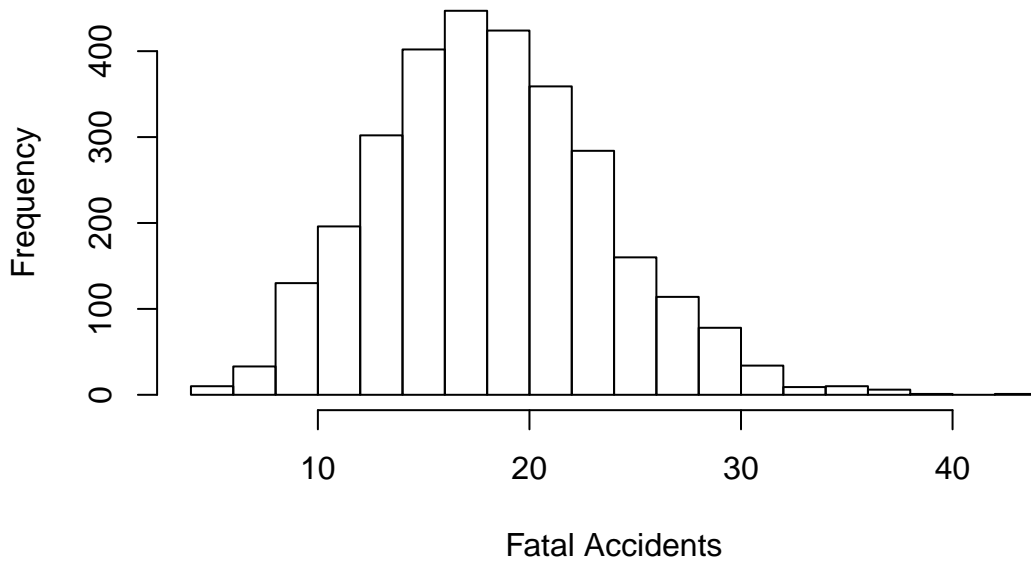
```
set.seed(312)


pulls <- 3000
sampleZ <- apply(Z,1,sum)
alphaI <- sample(1:length(XX), pulls, replace=T,
                 prob = sampleZ)

alphas <- XX[alphaI]
betas <- rep(NA, pulls)
for (i in 1:pulls) {
  betas[i] <- sample(YY, 1, prob = Z[alphaI[i],],replace = F)
}


Crash_preds <- rpois(alphas+betas*11,alphas+betas*11)

hist(Crash_preds, breaks = 20, main = "Predicted Fatal Accidents 1986", xlab = "Fatal Accidents")
```

Predicted Fatal Accidents 1986

The 95% predictive interval is $[9, 30]$ and the mean is 18.7.

i. The informative prior distribution, a bivariate normal, is symmetric, while the posterior distribution using the non-informative prior is skewed. There appears to be a correlation between the alpha and beta values in the posterior, that as alpha increases, beta must decrease. This makes intuitive sense as the value of $\alpha + \beta * t$ is restricted above zero, and should be correlated as they are being added to predict a single rate parameter for a Poisson distribution.

BDA Question 3.14

For the bioassay example $p(\alpha, \beta) \propto 1$. Thus,

$$p(\alpha, \beta | \mathbf{y}, \mathbf{n}, \mathbf{x}) \propto p(\mathbf{y} | \alpha, \beta, \mathbf{n}, \mathbf{x})$$

$$\propto \prod_{i=1}^{k} p(y_i | \alpha, \beta, n_i, x_i)$$

$$\propto \prod_{i=1}^{k} \left[\text{logit}^{-1}(\alpha + \beta x_i) \right]^{y_i} \left[1 - \text{logit}^{-1}(\alpha + \beta x_i) \right]^{n_i - y_i}.$$

We want,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k} \left[\text{logit}^{-1}(\alpha + \beta x_i) \right]^{y_i} \left[1 - \text{logit}^{-1}(\alpha + \beta x_i) \right]^{n_i - y_i} d\alpha d\beta < \infty$$

This is equivalent to

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{i=1}^{k} \left[\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \right]^{y_i} \left[\frac{1}{1+e^{\alpha+\beta x_i}} \right]^{n_i-y_i} d\alpha d\beta \ < \infty$$

Now define,

$$U_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \in (0,1)$$

Integrating these U_i values over $(\alpha, \beta) \in (-\infty, \infty) \times (-\infty, \infty)$ is equivalent to integrating each U_i from $(0,1)$. We re-write the above integral as,

$$\int_0^1 \cdots \int_0^1 \prod_{i=1}^{k} [U_i]^{y_i} [1-U_i]^{n_i-y_i} \, dU_1...dU_k$$

But, since each term inside the product is only dependent on the ith U-value, we can factor the product density inside the integral and rewrite it as

$$\prod_{i=1}^{k} \int_0^1 [U_i]^{y_i} [1-U_i]^{n_i-y_i} \, dU_i.$$

Notice that this is the product of k integrated Beta kernels, where the ith integral is a $\text{Beta}(y_i+1, n_i-y_1+1)$. Each of these integrals will obviously be finite, so their product will be finite as well, showing that the posterior is proper.

BDA Question 3.15

a. We can write

$$\begin{aligned}
p(y_t|y_0, y_1, ..., y_{t-1}, y_{t+1}, ...) &\propto p(..y_{t+1}, y_{t-1}, ...y_1, y_0|y_t)p(y_t) \\
&\propto ...p(y_{t+1}|y_{t-1}, ...y_1, y_0, y_t)p(y_{t-1}|y_{t-2}, ...y_1, y_0, y_t)...p(y_t) \\
&\propto e^{-(y_{t+1}-0.8y_t)^2/2} p(y_{t-1}|y_{t-2}, ...y_1, y_0, y_t)p(y_t) \\
&\propto e^{-(y_{t+1}-0.8y_t)^2/2} p(y_t|y_{t-2}, ...y_1, y_0, y_{t-1})p(y_{t-1}) \\
&\propto e^{-(y_{t+1}-0.8y_t)^2/2} p(y_t|y_{t-2}, ...y_1, y_0, y_{t-1}) \\
&\propto e^{-(y_{t+1}-0.8y_t)^2/2} e^{-(y_t-0.8y_{t-1})^2/2}
\end{aligned}$$

Notice that the last line is only dependent on y_{t+1} and y_{t-1} as desired.

b. We can show the following

$$\begin{aligned}
p(y_t|y_{t+1}, y_{t-1}) &\propto p(y_{t+1}, y_t, y_{t-1}) \\
&\propto p(y_{t+1}|y_t, y_{t-1})p(y_t|y_{t-1})p(y_{t-1}) \\
&\propto p(y_{t+1}|y_t, y_{t-1})p(y_t|y_{t-1}) \\
&\propto e^{-(y_{t+1}-0.8y_t)^2/2} e^{-(y_t-0.8y_{t-1})^2/2}
\end{aligned}$$

We work with the last line in the string of proportions to determine the form of the distribution.

$$p(y_t|y_{t+1}, y_{t-1}) \propto e^{-(y_{t+1}-0.8y_t)^2/2} e^{-(y_t-0.8y_{t-1})^2/2}$$

$$\propto \exp\left\{-\frac{1}{2}\left(y_{t+1}^2 - 1.6y_t y_{t+1} + 0.64y_t + y_t^2 - 1.6y_t y_{t-1} + 0.64y_{t-1}^2\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(1.64y_t^2 - 1.6y_t(y_{t+1} + y_{t-1})\right)\right\}$$

$$\propto \exp\left\{-\frac{1.64}{2}\left(y_t^2 - \frac{40}{41}y_t(y_{t+1} + y_{t-1})\right)\right\}$$

$$\propto \exp\left\{-\frac{1}{2\left(\frac{1}{1.64}\right)}\left(y_t - \frac{20}{41}(y_{t+1} + y_{t-1})\right)^2\right\}$$

We obtain the last line by completing the square. This is the kernel of a $N\left(\frac{20}{41}(y_{t+1} + y_{t-1}), \frac{1}{1.64}\right)$ distribution.