1 Chapter 3 Problem 2

If we assume independent noninformative uniform prior distributions on the multinomial parameters, then the posterior distributions are independent multinomials (i.e. independent Dirichlet distributions, demonstrated on page 69 in the text). Therefore, where

$$p(\pi_1, \pi_2, \pi_3|y) \sim \text{Dir}(295, 308, 39)$$
$$p(\pi_1^*, \pi_2^*, \pi_3^*|y) \sim \text{Dir}(289, 333, 20)$$

We are instructed to define α_1 and α_2 as follows:

$$\alpha_1 = \frac{\pi_1}{\pi_1 + \pi_2}, \alpha_2 = \frac{\pi_1^*}{\pi_1^* + \pi_2^*}$$

We let the prior distribution be defined as $p(\theta) \sim \text{Dir}(a_1, ..., a_n)$. Then we can use the known property of the Dirichlet distribution that the marginal posterior distribution of $(\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$ is also Dirichlet with parameters $(a_1 + y_1, a_2 + y_2, a_{rest} + y_{rest} - 1)$ where a_{rest} is the sum of all a's except the first two (seen in Appendix A). Therefore,

$$P(\theta_1, \theta_2|y) = \theta_1^{y_1+a_1-1}\theta_2^{y_2+a_2-1}(1 - \theta_1 - \theta_2)^{y_{rest}+a_{rest}-1}$$

Next we do a change of variable where $(\alpha, \beta) = (\frac{\theta_1}{\theta_1+\theta_2}, \theta_1 + \theta_2)$, so $\theta_1 = \alpha\beta$ and $\theta_2 = (1-\alpha)\beta$.

$$P(\alpha, \beta|y) \propto \beta(\alpha\beta)^{y_1+a_1-1}((1-\alpha)\beta)^{y_2+a_2-1}(1 - (\alpha\beta) - ((1-\alpha)\beta))^{y_{rest}+a_{rest}-1}$$
$$P(\alpha, \beta|y) = \beta(\alpha\beta)^{y_1+a_1-1}((1-\alpha)\beta)^{y_2+a_2-1}(1 - \beta)^{y_{rest}+a_{rest}-1}$$
$$P(\alpha, \beta|y) = \alpha^{y_1+a_1-1}(1-\alpha)^{y_2+a_2-1}\beta^{y_1+y_2+a_1+a_2 1}(1 - \beta)^{y_{rest}+a_{rest}-1}$$
$$P(\alpha, \beta|y) \propto \text{Beta}(\alpha|y_1 + a_1, y_2 + a_2)\text{Beta}(\beta|y_1 + y_2 + a_1 + a_2, y_{rest} + a_{rest})$$

Because the posterior density function decomposes for α and β, we see that there is independence between the two in the posterior. This allows us to consider the two separately and note that $\alpha|y \sim \text{Beta}(y_1 + a_1, y_2 + a_2)$, i.e. $\alpha_1|y \sim \text{Beta}(\pi_1, \pi_2)$ and $\alpha_2|y \sim \text{Beta}(\pi_1^*, \pi_2^*)$. Therefore,

$$p(\alpha_1|y) \sim \text{Beta}(295, 308)$$
$$p(\alpha_2|y) \sim \text{Beta}(289, 333)$$

Below you can see the histogram of 5000 samples pulled from the posterior distribution of $\alpha_2 - \alpha_1$. The code used to create the image is included after the figure. The difference is greater than 0 (i.e. we predict a shift towards Bush) about 19% of the time.

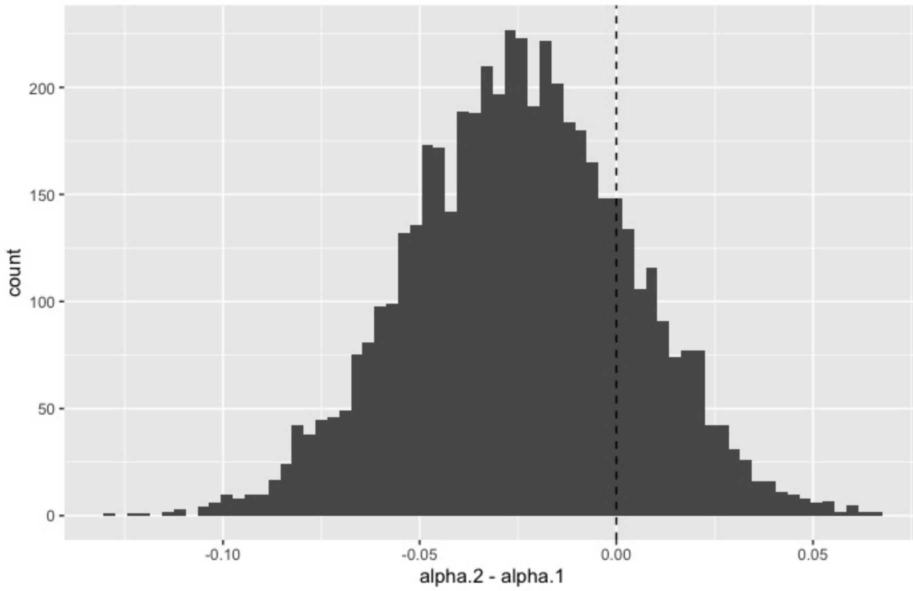

Figure 1: Chapter 3, Problem 2

```
library(ggplot2)

alpha.1 <- rbeta(5000, 295, 308)
alpha.2 <- rbeta(5000, 289, 333)

parameter <- data.frame(alpha.2 - alpha.1)
colnames(parameter) <- c("dif")

ggplot(data=parameter, aes(dif)) +
  geom_histogram(stat = "bin", binwidth = 0.003) +
  xlab("alpha.2 - alpha.1") + geom_vline(xintercept = 0, linetype="dashed")

mean(parameter>0)
```

2 Chapter 3 Problem 3

2.1 Part A

From the problem statement we are given that:

$$p(y|\mu_c, \mu_t, \sigma_c, \sigma_t) = \prod_{i=1}^{32} N(y_{ci}|\mu_c, \sigma_c^2) \prod_{i=1}^{36} N(y_{ti}|\mu_c, \sigma_t^2)$$

.

We are looking for the posterior distribution, written out as:

$$p(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t|y) = p(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t)p(y|\mu_c, \mu_t, \log\sigma_c, \log\sigma_t)$$

Recall our given prior: $p(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t) \propto 1$. Therefore,

$$p(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t|y) \propto 1 * \prod_{i=1}^{32} N(y_{ci}|\mu_c, \sigma_c^2) \prod_{i=1}^{36} N(y_{ti}|\mu_c, \sigma_t^2)$$

Which is just:

$$p(\mu_c, \mu_t, \log\sigma_c, \log\sigma_t|y) = \prod_{i=1}^{32} N(y_{ci}|\mu_c, \sigma_c^2) \prod_{i=1}^{36} N(y_{ti}|\mu_c, \sigma_t^2)$$

Because the posterior density function remains in this form, we see that there is independence between (μ_c, σ_c) and (μ_t, σ_t) in the posterior. This allows us to consider the two separately. Section 3.2 in the text allows us to write the marginal posterior distributions as:

$$P(\mu_t|y) \sim t_{31}\left(1.013, \frac{0.24^2}{32}\right)$$

$$P(\mu_c|y) \sim t_{35}\left(1.173, \frac{0.20^2}{36}\right)$$

2.2 Part B

My calculated 95% posterior interval for $\mu_t - \mu_c$ is [0.051, 0.268]. See my histogram and simulation code below.

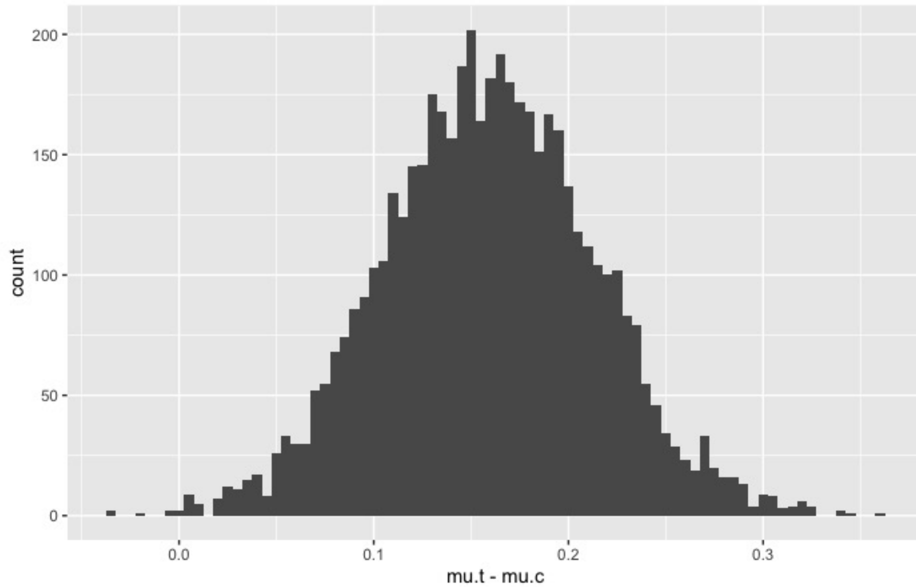

Figure 2: Chapter 3, Problem 3

```
mu.c <- 1.013 + (0.24/sqrt(32)) * rt(5000,31)
mu.t <- 1.173 + (0.20/sqrt(36)) * rt(5000,35)

parameter3.3 <- data.frame(mu.t - mu.c)
colnames(parameter3.3) <- c("dif")

ggplot(data=parameter3.3, aes(dif)) +
  geom_histogram(stat = "bin", binwidth = 0.003) +
  xlab("mu.t - mu.c")

quantile(parameter3.3$dif, c(0.025, 0.975))
```

3 Chapter 3 Problem 5

3.1 Part A

If we decide to treat the observations exact measurements and assume a noninformative prior then we can use the work done in Section 3.2 (page 65) of the text and write both the conditional posterior density and the marginal posterior density as:

$$p(\mu|\sigma^2, y) \sim N(\bar{y}, \frac{\sigma^2}{n})$$

$$p(\sigma^2|y) \sim \text{Inv} - \chi^2(n-1, s^2)$$

Where we calculate $\bar{y} = 10.4, n = 5, s^2 = 1.2$.

3.2 Part B

Recall that we have determined the non-informative prior of $p(\mu, \sigma^2) \propto 1/\sigma^2$. We wish to determine the posterior $p(\mu, \sigma^2|y)$, which we write as:

$$p(\mu, \sigma^2|y) = p(\mu, \sigma^2)p(y|\mu, \sigma^2)$$

Rounding implies adding a value between -0.5 and 0.5 to each measurement, such that the sum is the integer closest to the measurement obtained. We can therefore write the likelihood that the rounded observations, Y_i's, take the integer values of y_i's is the following (as described in the Heitjan, 1989 manuscript that they direct us to):

$$L(y|\theta) = \prod_{i=1}^{n}[F(y_i + 0.5, \theta) - F(y_i - 0.5, \theta)]$$

where F is the distribution function and θ are our set of parameters. Let us use the symbol Φ to represent the standard normal cumulative distribution function. Then,

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

3.3 Part C

I computed the contour plots on a grid with a $log(\sigma)$ scale. I sampled joint posteriors while sweeping over all possible values of sigma and mu in a 200 by 200 grid. I then numerical sampled the marginal posteriors of sigma and mu using the joint probabilities from this grid. I did this for each of the two scenarios (to by compared in Part D). My simulation code follows. First see below for information on the means and variances of the two posteriors. I only notice substantial differences when σ is small.

Mean	StdDev	2.5%	25%	50%	75%	97.5%
10.40	0.76	8.96	10.00	10.39	10.80	11.93
1.46	0.82	0.69	0.99	1.26	1.66	3.32

Table 1: Exact Measures

Mean	StdDev	2.5%	25%	50%	75%	97.5%
10.41	0.70	9.11	10.01	10.39	10.76	11.82
1.37	0.73	0.63	0.93	1.19	1.60	3.21

Table 2: Rounded Measures

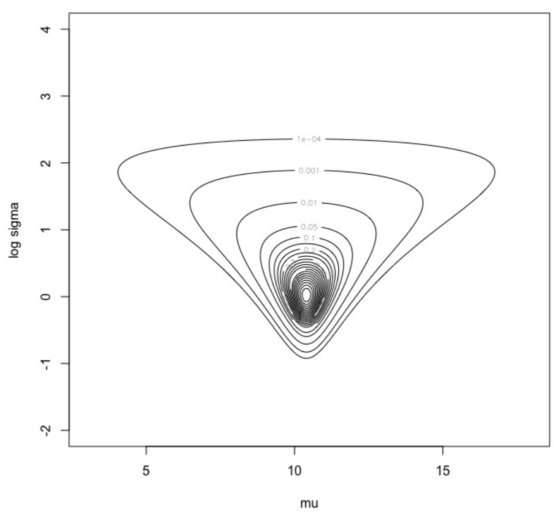

Figure 3: Exact Measures, Chapter 3, Problem 5

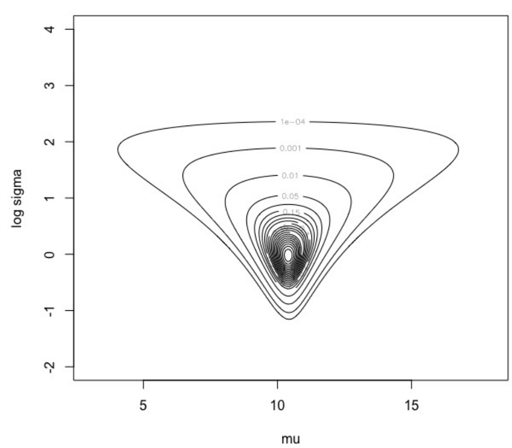

Figure 4: Rounded Measures, Chapter 3, Problem 5

```
# to construct contours of the posterior when we assume EXACT
#function for posterior when measurements are exact
exact.posterior <- function(mu,sd,y){
  ldens <- 0
  for (i in 1:length(y))
    ldens <- ldens + log(dnorm(y[i],mu,sd))
  ldens
}

#samples
y <- c(10,10,12,11,9)
#no of samples
n <- length(y)
#sample mean
ybar <- mean(y)
#sample variance
s2 <- sum((y-mean(y))^2)/(n-1)

#grids for potential mu and logsigma values
mu.vec <- seq(3,18,length=200)
logsd.vec <- seq(-2,4,length=200)
#storage for density values
dens.storage <- matrix(NA, 200, 200)
#pull values and store
for (i in 1:200) {
  for (j in 1:200) {
    dens.storage[i,j] <- exact.posterior(mu.vec[i], exp(logsd.vec[j]), y)
  }
}
#build contour plots
dens <- exp(dens.storage - max(dens.storage))
contours <- c(.0001,.001,.01, seq(.05,.95,.05))
contour(mu.vec, logsd.vec, dens, levels=contours, xlab="mu", ylab="log sigma",
    labex=0, cex=2)

#calculation of the posterior density from the Section 3.2 informed
#   conditional posterior density and the marginal posterior density
sd <- sqrt((n-1)*s2/rchisq(nsim,4))
mu <- rnorm(nsim,ybar,sd/sqrt(n))

#function to print mean, variance, quantiles of our samples
for.comparison <- function(x){c(mean(x),sqrt(var(x)), quantile(x,
    c(.025,.25,.5,.75,.975)))}
print(round((rbind(for.comparison(mu),for.comparison(sd))),2))
```

```
#function for posterior when measurements are rounded
rounded.posterior <- function(mu,sd,y){
  ldens <- 0
  for (i in 1:length(y))
    ldens <- ldens + log(pnorm(y[i]+0.5,mu,sd) - pnorm(y[i]-0.5,mu,sd))
  ldens}

# to construct contours of the posterior when we assume ROUNDED
#storage for density values
dens.storage <- matrix(NA, 200, 200)
#pull values and store
for (i in 1:200) {
  for (j in 1:200) {
    dens.storage[i,j] <- rounded.posterior(mu.vec[i], exp(logsd.vec[j]), y)
  }
}

#build contour plots
dens <- exp(dens.storage - max(dens.storage))
contour (mu.vec, logsd.vec, dens, levels=contours, xlab="mu", ylab="log sigma",
    labex=0, cex=2)

#calculation of the posterior, sample from the grid approximation to find
#   conditional posterior density and the marginal posterior density
#number of samples
nsim <- 2000
dens.mu <- apply(dens,1,sum)
mu.indices <- sample(1:length(mu.vec), nsim, replace=T, prob=dens.mu)
mu <- mu.vec[mu.indices]
sd <- rep(NA,nsim)
for (i in (1:nsim)) {
  sd[i] <- exp(sample(logsd.vec, 1, prob=dens[mu.indices[i],]))
}
#print mean, variance, quantiles of our samples
print(round((rbind(for.comparison(mu),for.comparison(sd))),2))
```

3.4 Part D

We will use our answer to part C. We know that the conditional distribution of z_i, given μ and σ, will again be $N(\mu, \sigma)$, where the samples fall between $(y_i - 0.5, y_i + 0.5)$ where the y_i's are the rounded values. So we use our posterior draws from part C to pull a sample based on the possible range of each y_i (assuming each independent). Lastly, we average the squared difference of z_1 and z_2. We discovered the posterior mean of $(z_1 - z_2)^2$ to be 0.156.

```
storage <- matrix (NA, nsim, length(y))
for (i in 1:length(y)){
  lower <- pnorm(y[i]-.5, mu, sd)
  upper <- pnorm(y[i]+.5, mu, sd)
  storage[,i] <- qnorm(lower + runif(nsim)*(upper-lower), mu, sd)}
mean((storage[,1]-storage[,2])^2)
```

4 Chapter 3 Problem 8

4.1 Part A

We will set up our model such that the y_i's and z_i's will be represented as beta distributions, where the given parameters θ_y and θ_z are $\{\alpha_y, \beta_y\}$ and $\{\alpha_z, \beta_z\}$, respectively.

$$p(y|\theta_y) \sim \text{Beta}(\alpha_y, \beta_y)$$

$$p(z|\theta_z) \sim \text{Beta}(\alpha_z, \beta_z)$$

4.2 Part B

I will use the uniform distribution for the priors of the α's and β's referenced above. It is logical that the minimum bound for each would be zero, as the lowest amount of bike crashes we could hypothesize seeing would be none (negative numbers are illogical). However, we will employ method of moments to determine the upper bounds on our respective uniform distributions.

$$\alpha_y \sim \text{Unif}(0, a_y)$$
$$\beta_y \sim \text{Unif}(0, b_y)$$

We will create a system of equations using what we know the variance and mean of the beta distribution to be and the method of moments. The method of moments tells us that:

$$E[X^j] = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

The first moment allows us to relate the sample mean.

$$E[y_i^1] = \frac{\alpha_y}{\alpha_y + \beta_y} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{y} = 0.2$$

We recall this property of the variance and the form of the variance of the beta distribution to find the second moment and avoid the full derivation:

$$\text{Var}[y_i] = E[y_i^2] - E[y_i^1]^2$$

$$\frac{\alpha_y \beta_y}{(\beta_y + \alpha_y)^2(\beta_y + \alpha_y + 1)} = E[y_i^2] - \left(\frac{\alpha_y}{\alpha_y + \beta_y}\right)^2$$

$$E[y_i^2] = \frac{(\alpha_y + 1)\alpha_y}{(\alpha_y + \beta_y)(\alpha_y + \beta_y + 1)}$$

$$s^2 = E[y_i^2] - E[y_i]^2 \rightarrow s^2 + E[y_i]^2 = E[y_i^2]$$

$$0.011 + 0.2^2 = E[y_i^2] = \frac{(\alpha_y + 1)\alpha_y}{(\alpha_y + \beta_y)(\alpha_y + \beta_y + 1)}$$

Now we have a system of equations to solve for α_y and β_y. The result is $\alpha_y = 2.709$ and $\beta_y = 10.84$. We use the same procedure for α_z and β_z where the sample mean is 0.097 and the sample variance is 0.004 and result is $\alpha_z = 2.027$ and $\beta_z = 18.87$. Therefore, we have fleshed out our uniform distributions for our parameters as follows:

$$\alpha_y \sim \text{Unif}(0, 2.709)$$

$$\beta_y \sim \text{Unif}(0, 10.84)$$

$$\alpha_z \sim \text{Unif}(0, 2.027)$$

$$\beta_z \sim \text{Unif}(0, 18.87)$$

4.3 Part C

As noted in the problem statement, we have constructed our model such that θ_y and θ_z are independent in the posterior so we can simulate them separately. Let us then start with the θ_y parameters. We wish to calculate the posterior distribution for the θ_y parameters, i.e.

$$p(\alpha_y, \beta_y | y) = p(\alpha_y, \beta_y) p(y | \alpha_y, \beta_y)$$

With the priors independent and the given data, we can write this as:

$$p(\alpha_y, \beta_y | y) = p(\alpha_y) p(\beta_y) \prod_{i=1}^{10} p(y_i | \alpha_y, \beta_y)$$

$$p(\alpha_y, \beta_y | y) = \left(\frac{1}{2.709}\right)\left(\frac{1}{10.84}\right) \prod_{i=1}^{10} y_i^{\alpha_y} (1 - y_i)^{\beta_y}$$

We can write the same for the posterior distribution for the θ_z parameters:

$$p(\alpha_z, \beta_z|z) = \left(\frac{1}{2.027}\right)\left(\frac{1}{18.87}\right)\prod_{i=1}^{8} z_i^{\alpha_z}(1-z_i)^{\beta_z}$$

I sampled these joint posteriors while sweeping over all possible values of alpha and beta in a 2000 by 2000 grid. I then numerically sampled the marginal posteriors of alpha and beta using the joint probabilities from this grid. My simulation code follows.

```
#function for theta_y's joint posterior
y.joint.posterior <- function(alpha.val,beta.val,y){
  ldens <- 1
  for (i in 1:length(y))
    ldens <- ldens * ((y[i]^alpha.val)*((1-y[i])^beta.val))
  return(ldens*(dunif(1, 0, 2.709))*(dunif(1, 0, 10.84)))}

# to calculate posterior at different values of alpha and beta
alpha.y.vec <- seq(0,2.709,length=2000)
beta.y.vec <- seq(0,10.84,length=2000)
#storage for density values
dens.storage <- matrix(NA, 2000, 2000)
#pull values and store
for (i in 1:2000) {
  for (j in 1:2000) {
    dens.storage[i,j] <- y.joint.posterior(alpha.y.vec[i], beta.y.vec[j],
        y.samples)
  }
}

#build contour plots
dens<-dens.storage
contours <- c(10^(-30:0))
contour (alpha.y.vec, beta.y.vec, dens, levels=contours, xlab="alpha",
   ylab="beta", labex=3, cex=10)

#calculation of the posterior, sample from a grid approximation to find
#   conditional posterior density and the marginal posterior density
#number of samples

nsim <- 1000
dens.alpha <- apply(dens,1,sum)
alpha.indices <- sample(1:length(alpha.y.vec), nsim, replace=T, prob=dens.alpha)
alpha.y <- alpha.y.vec[alpha.indices]
beta.y <- rep(NA,nsim)
for (i in (1:nsim)) {
```

```
  beta.y[i] <- sample(beta.y.vec, 1, prob=dens[alpha.indices[i],])
}
y.finals <- alpha.y/(alpha.y+beta.y)

###_______________________________________________________________

#function for theta_z's joint posterior
z.joint.posterior <- function(alpha.val,beta.val,y){
  ldens <- 1
  for (i in 1:length(y))
    ldens <- ldens * ((y[i]^alpha.val)*((1-y[i])^beta.val))
  return(ldens*(dunif(1, 0, 2.027))*(dunif(1, 0, 18.87)))}

# to calculate posterior at different values of alpha and beta
alpha.z.vec <- seq(0,2.027,length=2000)
beta.z.vec <- seq(0,18.87,length=2000)
#storage for density values
dens.storage <- matrix(NA, 2000, 2000)
#pull values and store
for (i in 1:2000) {
  for (j in 1:2000) {
    dens.storage[i,j] <- z.joint.posterior(alpha.z.vec[i], beta.z.vec[j],
        z.samples)
  }
}
dens.z<-dens.storage
#build contour plots
#dens <- exp(dens.storage - max(dens.storage))
contours <- c(10^(-30:0))
contour (alpha.z.vec, beta.z.vec, dens.z, levels=contours, xlab="alpha",
   ylab="beta", labex=3, cex=10)

#calculation of the posterior, sample from a grid approximation to find
#   conditional posterior density and the marginal posterior density
#number of samples
nsim <- 1000
dens.alpha <- apply(dens.z,1,sum)
alpha.indices <- sample(1:length(alpha.z.vec), nsim, replace=T, prob=dens.alpha)
alpha.y <- alpha.z.vec[alpha.indices]
beta.y <- rep(NA,nsim)
for (i in (1:nsim)) {
  beta.y[i] <- sample(beta.z.vec, 1, prob=dens.z[alpha.indices[i],])
}
z.finals <- alpha.y/(alpha.y+beta.y)
```

4.4 Part D

Finally, I used these values to calculate μ_y and μ_z and their difference. Below is my histogram of the posterior simulations of $\mu_y - \mu_z$. The mean was ~ 0.0956.

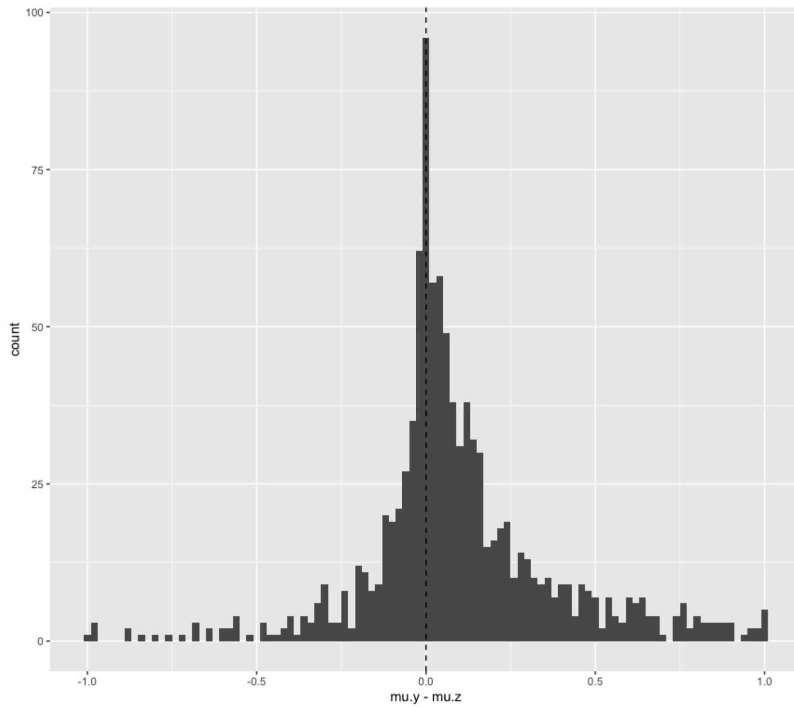

Figure 5: Chapter 3, Problem 8

```
## final difference of the means histogram
diff.finals.df <- data.frame(c(y.finals-z.finals))
colnames(diff.finals.df) <- c("z")
mean(diff.finals.df$z)
ggplot(data=diff.finals.df, aes(z)) +
  geom_histogram(stat = "bin", binwidth = 0.02) +
  xlab("diff.final") +
  xlab("mu.y - mu.z") + geom_vline(xintercept = 0, linetype="dashed")
```

5 Chapter 3 Problem 12

We let $\theta = \alpha + \beta t$.

5.1 Part A

Since θ is the mean of a Poisson distribution we know it must be greater than zero. Therefore, one possible prior is $p(\alpha, \beta) \propto 1_{[0,\text{inf}]}$. We could also consider Jeffreys' prior, $p(\alpha, \beta) \sim [J(\alpha, \beta)]^{1/2}$. However, as noted on page 53 of the text, the results of using Jeffrey's prior with multiparameter models can be inconsistent. I would therefore default to the uniform prior.

5.2 Part B

I could use my prior knowledge about plane crashes and plane technology to hypothesize that planes will be getting safer over time. This would mean that β might trend negative over time. Additionally, we wouldn't expect frequent plane crashes, encouraging α to be small (although strictly non-negative). With this in mind, appropriate priors might be independent normal distributions, i.e. $p(\alpha) \sim N(30, 10)$ and $p(\beta) \sim N(-3, 5)$.

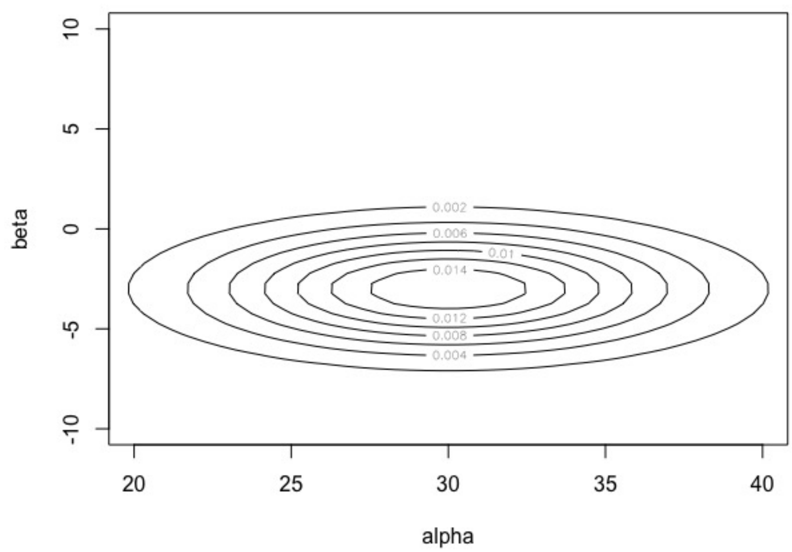

Figure 6: Chapter 3, Problem 12, Part B

```
storage <- matrix(NA, 100, 100)
alpha.vec <- seq(0, 50, length=100)
beta.vec <- seq(-30, 20, length=100)
for (i in 1:100) {
  for (j in 1:100) {
    storage[i,j] <- dnorm(alpha.vec[i], 30, 5)*dnorm(beta.vec[j], -3, 2)
```

```
  }
}
contour(alpha.vec, beta.vec, storage, xlab="alpha", ylab="beta", xlim=c(20, 40),
    ylim=c(-10, 10))
```

5.3 Part C

We will use the uninformative prior from Part A, $p(\alpha, \beta) \propto 1$, to find the posterior.

$$p(\alpha, \beta|y) \sim p(\alpha, \beta)p(y|\alpha, \beta)$$
$$p(\alpha, \beta|y) \propto 1 * \prod_{i=1}^{10} (\alpha + \beta t_i)^{y_i} e^{-(\alpha+\beta t_i)}$$
$$p(\alpha, \beta|y) \propto 1 * e^{-(n\alpha+\beta \sum t_i)} \prod_{i} (\alpha + \beta t_i)^{y_i}$$

Here the data are the sufficient statistics: $(y_1, t_1)...(y_{10}, t_{10})$.

5.4 Part D

To check if the posterior is proper we need to know if the following integral is finite:

$$\int \int p(\alpha, \beta|y) d\alpha d\beta < \infty$$
$$\int \int (\alpha + \beta t_i) e^{-(\alpha+\beta t_i)} d\alpha d\beta < \infty$$

Let us denote the space of possible values of the α, β parameters as Φ. We can bound the inside of this integral by noting:

$$sup_{(\alpha,\beta)\in\Phi}\{(\alpha + \beta t_i)e^{-(\alpha+\beta t_i)}\} \leq sup_{\theta\in[0,\infty)}\{\theta e^{-\theta}\}$$

And we know that for some constant C:

$$sup_{\theta\in[0,\infty)}\{\theta e^{-\theta}\} = C$$

Therefore,

$$sup_{(\alpha,\beta)\in\Phi}\{(\alpha + \beta t_i)e^{-(\alpha+\beta t_i)}\} \leq C$$

This lets us bound the integral by the largest value it can take given C:

$$\int \int C^9 (\alpha + \beta t_1) e^{-(\alpha+\beta t_1)} d\alpha d\beta$$

We note that this is indeed finite and we are done. $\square$

5.5 Part E

Using linear regression I find that $\alpha \sim N(28.87, 2.75)$ and $\beta \sim N(-0.92, 0.44)$.

```
accidents <- c(24, 25, 31, 31, 22, 21, 26, 20, 16, 22)
time <- c(1:10)
lm(accidents ~ time)
summary(lm(accidents ~ time))
```

5.6 Part F

Here are my contours of the the joint posterior density of (α, β) and the code I used to build them.

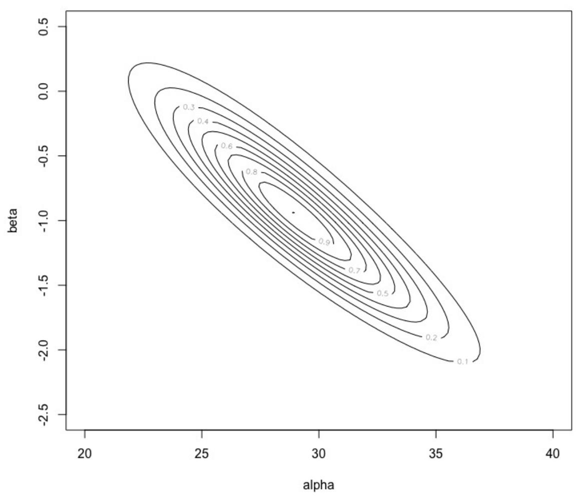

Figure 7: Chapter 3, Problem 12, Part F

```
storage <- matrix(NA, 100, 100)
#create alpha and beta grids
alpha.vec <- seq(20, 40, length=100)
beta.vec <- seq(-3, 1, length=100)
for (i in 1:100){
  for (j in 1:100){
    sum <- 0
    #calculate sum (log(product)) term
    for (ii in (1:10)) {
      sum <- sum+log(((alpha.vec[i]+beta.vec[j]*time[ii])^(accidents[ii]))
```

```
    }
    #calculate and sum with the rest of the expression
    storage[i,j] <- sum +
      log(exp(-((length(time)*alpha.vec[i])+(beta.vec[j]*sum(time)))))
  }
}
#scale density
dens <- exp(storage - max(storage))
#plot contours
contour(alpha.vec, beta.vec, dens, xlim=c(20, 40), ylim=c(-2.5, 0.5),
    xlab="alpha", ylab="beta")
```

5.7 Part G

I randomly sampled from the joint posterior probabilities from Part F for both alpha and beta and then calculated $\alpha + 11\beta$ (11 because I have been using 1:10 for my time variable). The mean settles around 18.82. See the histogram and code below:

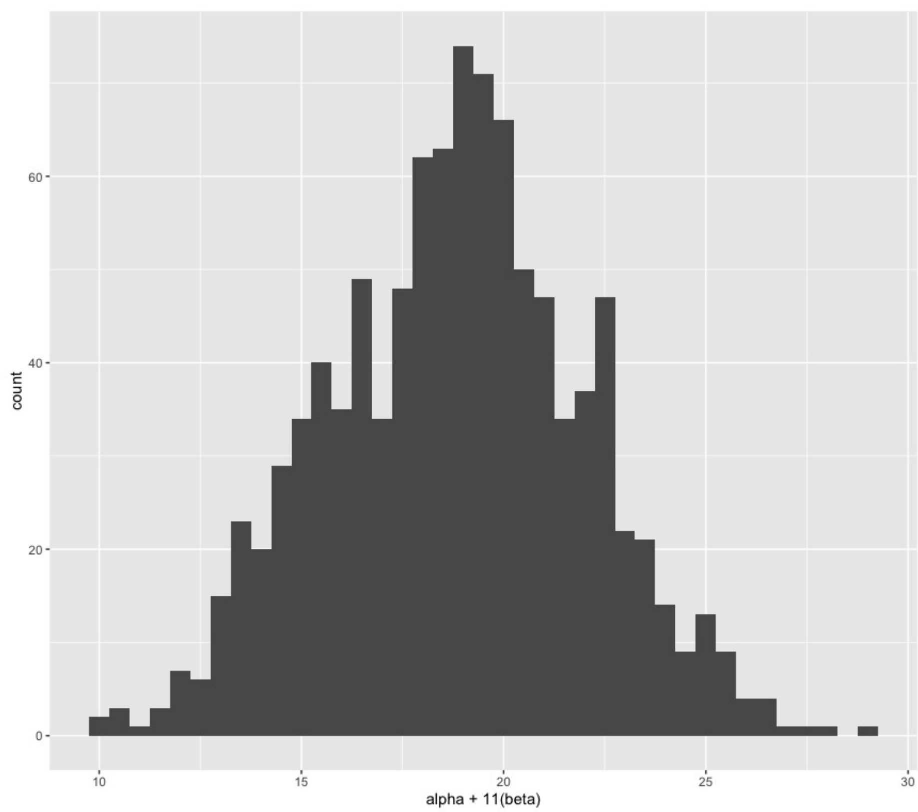

Figure 8: Chapter 3, Problem 12, Part G

```
#calculation of the posterior, sample from a grid approximation to find
#   conditional posterior density and the marginal posterior density
#number of samples
nsim <- 1000
dens.alpha <- apply(dens,1,sum)
alpha.indices <- sample(1:length(alpha.vec), nsim, replace=T, prob=dens.alpha)
alpha <- alpha.vec[alpha.indices]
beta <- rep(NA,nsim)
for (i in (1:nsim)) {
  beta[i] <- sample(beta.vec, 1, prob=dens[alpha.indices[i],])
}

a.11.b.df <- data.frame(c(alpha+11*beta))
colnames(a.11.b.df) <- c("val")
mean(a.11.b.df$val)
ggplot(data=a.11.b.df, aes(val)) +
  geom_histogram(stat = "bin", binwidth = 0.5) +
  xlab("alpha + 11(beta)")
```

5.8 Part H

I created simulation draws generating 1000 new values of θ using the pulled values of α and β. Lastly I calculated the 95% predictive interval to be [9,29].

```
num.accidents=rpois(1000, alpha+beta*11)
quantile(num.accidents, c(0.025, 0.975))
```

5.9 Part I

The informative prior does indeed vary from the posterior distribution that I found in Part F and Part G. The informative prior assumes that α and β are independent as you can see in the contour plot. Whereas the plot of the posterior of α and β show them strongly affecting one another–clearly dependent.

6 Chapter 3 Problem 14

We aim to show that the posterior density of the bioassay example is proper, i.e. has a finite integral over the given range. The posterior distributions is as follows:

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

We are suggested in the text to use a uniform prior, $p(\alpha, \beta) \propto 1$. Taking into account this prior and the given likelihood, we can write:

$$p(\alpha, \beta|y, n, x) \propto \prod_{i=1}^{k} [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}$$

Recall:

$$\text{logit}^{-1}(u) = \frac{e^u}{1 + e^u} = \text{expit}(u) = \frac{1}{1 + e^{-u}}$$

$$p(\alpha, \beta|y, n, x) \propto \prod_{i=1}^{k} [\text{expit}(\alpha + \beta x_i)]^{y_i} [1 - \text{expit}(\alpha + \beta x_i)]^{n_i - y_i}$$

We need to show the following:

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

Let us call $(\alpha + \beta x_i) = u$ for visual ease and let's focus on the inside of the product.

$$[\text{expit}(\alpha + \beta x_i)]^{y_i} [1 - \text{expit}(\alpha + \beta x_i)]^{n_i - y_i} = [\text{expit}(u)]^{y_i} [1 - \text{expit}(u)]^{n_i - y_i}$$

$$= \left[\frac{1}{1 + e^{-u}}\right]^{y_i} \left[1 - \frac{1}{1 + e^{-u}}\right]^{n_i - y_i}$$

$$= \left[\frac{1}{1 + e^{-u}}\right]^{y_i} \left[\frac{1 + e^{-u}}{1 + e^{-u}} - \frac{1}{1 + e^{-u}}\right]^{n_i - y_i}$$

$$= \left[\frac{1}{1 + e^{-u}}\right]^{y_i} \left[\frac{e^{-u}}{1 + e^{-u}}\right]^{n_i - y_i}$$

$$= \left[\frac{1}{(1 + e^{-u})^{y_i}}\right] \left[\frac{(e^{-u})^{n_i - y_i}}{(1 + e^{-u})^{n_i - y_i}}\right]$$

$$= \frac{(e^{-u})^{n_i - y_i}}{(1 + e^{-u})^{n_i}}$$

$$= \left[\frac{1}{(1 + e^{-u})}\right]^{n_i} \left[(e^{-u})^{n_i - y_i}\right]$$

$$= [\text{expit}(u)]^{n_i} \left[(e^{-(n_i - y_i)u})\right]$$

Let us denote the space of possible values of the α, β parameters as Φ. Also note that expit functions are restricted to between (0,1) no matter the values of α, β. Therefore, we can now bound the inside of this integral by noting:

$$sup_{(\alpha,\beta)\in\Phi}\{e^{-(n_i-y_i)(\alpha+\beta t_i)}\} \leq sup_{\theta\in[0,\infty)}\{e^{-\theta}\}$$

And we know that for some constant C:

$$sup_{\theta\in[0,\infty)}\{e^{-\theta}\} = C$$

Therefore,

$$sup_{(\alpha,\beta)\in\Phi}\{e^{-(\alpha+\beta t_i)}\} \leq C$$

This lets us bound the integral by the largest value the integral can take given C:

$$\propto \int\int C^{k-1}e^{-(\alpha+\beta t_1)}d\alpha d\beta$$

We note that this is indeed finite and we are done. $\square$

7 Chapter 3 Problem 15

7.1 Part A

We wish to investigate the distribution of y_t, given the observations at all other integer time points t, i.e.

$$p(y_t|...y_{t+1}, y_{t-1}, ..., y_1, y_0) = \frac{p(...y_{t+1}, y_t, y_{t-1}, ..., y_1, y_0)}{p(...y_{t+1}, y_{t-1}, ..., y_1, y_0)}$$

Because the marginal (the denominator) is not dependent on the value of y_t we can identify it as a constant and write "proportional to" instead of "equal." Therefore,

$$p(y_t|...y_{t+1}, y_{t-1}, ..., y_1, y_0) \propto p(...y_{t+1}, y_t, y_{t-1}, ..., y_1, y_0)$$

We can rewrite this as:

$$\propto ...p(y_{t+2}|y_{t+1}, y_t...)p(y_{t+1}|y_t, y_{t-1}...)p(y_t|y_{t-1}, y_{t-2}...)p(y_{t-1}|y_{t-2}, y_{t-3}...)...$$

Using the given property of $(y_t|y_{t-1}, y_{t-2}, ...) \sim N(0.8y_{t-1}, 1)$, we can write:

$$\propto ...e^{-\frac{(y_{t+2}-0.8y_{t+1})^2}{2}}e^{-\frac{(y_{t+1}-0.8y_t)^2}{2}}e^{-\frac{(y_t-0.8y_{t-1})^2}{2}}e^{-\frac{(y_{t-1}-0.8y_{t-2})^2}{2}}...$$

However again we note that terms that do not include any y_t's are simply constants and can be disregarded as such. Therefore, the only terms we care about are the following:

$$p(y_t|...y_{t+1}, y_{t-1}, ..., y_1, y_0) \propto e^{-\frac{(y_{t+1}-0.8y_t)^2}{2}}e^{-\frac{(y_t-0.8y_{t-1})^2}{2}}$$

Indeed, we have shown that $p(y_t)$ given all observations at other integer time points is only dependent on y_{t+1}, y_{t-1}.

7.2 Part B

Now, we wish to determine the distribution of (y_t) given y_{t+1}, y_{t-1}. We can start from where we left off above.

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

$$\propto \exp\left(\frac{-(y_{t+1}-0.8y_t)^2-(y_t-0.8y_{t-1})^2}{2}\right)$$

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

Again we can remove the constant terms and combine the others to give:

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

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

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

We can recognize this as almost the form of another Normal distribution, we just need to complete the square. We can avoid doing this by recognizing that the coefficient of y_t is the "b" value in the formula and this gets transformed into $\frac{b}{2a}$. We know that $a = 1$. Therefore, the final y_t coefficient will be $\frac{1.6}{1.64*2}$. The extra squared term drops as a constant. Therefore,

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

$$p(y_t|y_{t+1}, y_{t-1}) \propto N\left(0.489(y_{t+1}+y_{t-1}), \frac{1}{1.64}\right)$$