Homework 5

Joseph Miller
Bayesian Data Analysis

March 30, 2016

3.11

In this question, I use a joint normal prior distribution on (α, β), with $\alpha \sim \mathrm{N}(0, 2^2)$, $\beta \sim \mathrm{N}(10, 10^2)$ and $\mathrm{corr}(\alpha, \beta) = 0.5$ and repeat the bioassay example of section 3.7 in the textbook.
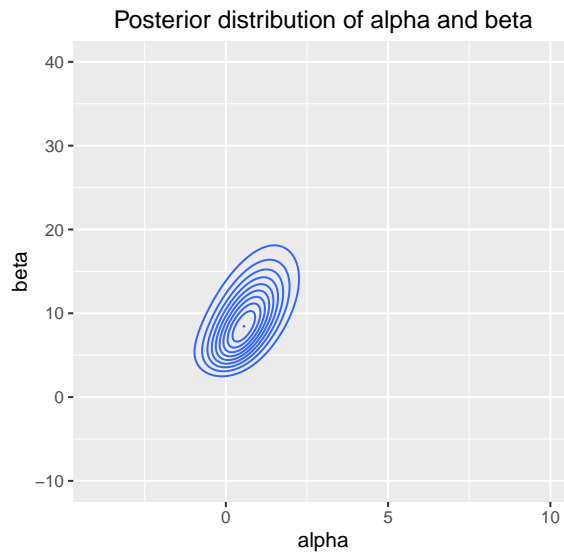
   a. Under the new prior, the posterior is as follows:

$$p(\alpha, \beta, |y) \propto p(\alpha, \beta)\Pi_{i=1}^{k}p(y_i|\alpha, \beta)$$

and is thus proportional to the form

$$\exp\left(-\frac{2(\alpha)^2}{2^2} + \frac{2\alpha(\beta - 10)}{20} - \frac{2(\beta - 10)^2}{10^2}\right)\Pi_{i=1}^{k}(\mathrm{logit}(\alpha + \beta x_i))^{y_i}(\mathrm{logit}(\alpha + \beta x_i))^{n_i - y_i}.$$

The corresponding contour plot is graphed below:

Posterior distribution of alpha and beta

And was sampled via the conditional method using $p(\alpha, \beta | y) = p(\beta | y) p(\alpha | \beta, y)$ implemented in the following R code snippet:

```
marginalBeta <- apply(pMat, MARGIN = 1, sum)
a <- vector(length=1000)
b <- vector(length=1000)
for (i in 1:1000) {
b[i] <- sample(size = 1, x = beta, prob = marginalBeta)
a[i] <- sample(size = 1, x = alpha, prob = pMat[which(beta==b[i]),])
}
# Adding an runif so the distribution is continuous.
a <- a + runif(n=1000, min = -0.01503006, max=0.01503006)
b <- b + runif(n=1000, min= -0.0501002, max=0.0501002)
```

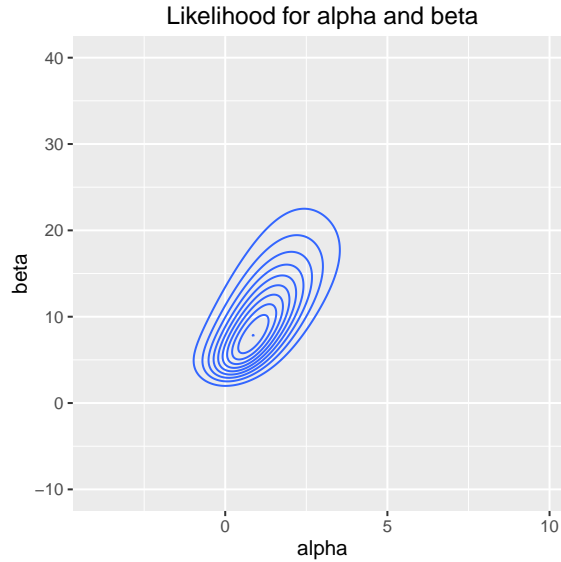where **pMat** is a matrix with entries as probabilities corresponding to all of the pairs of (α, β). The plot follows:
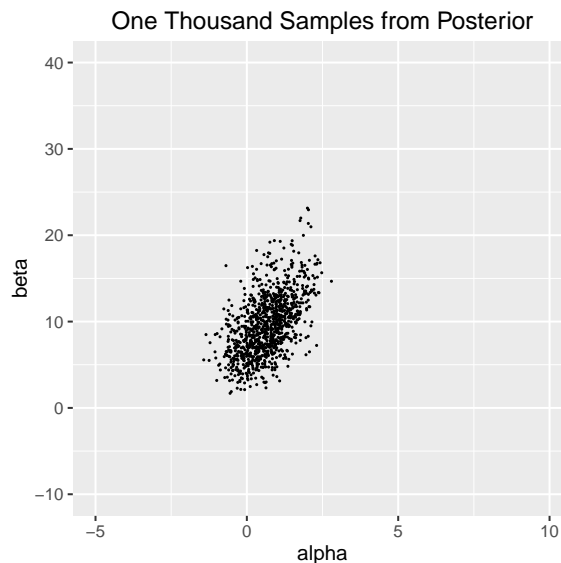
Finally, using the same sampled values, **a**, **b**, I simply plotted the distribution of **-a/b** yielding the final histogram:

Samples from Posterior of LD50
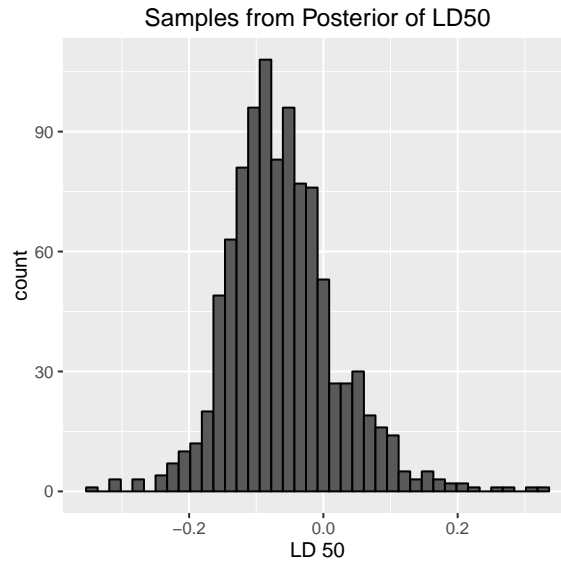
b. Contour plots of my prior and likelihood are as follows:

Prior distribution of alpha and beta

Likelihood for alpha and beta

When the contour plot and scatterplot of the posterior in the previous section, are compared to the prior and likelihood, it is easy to see that the mode of the posterior seems to lie somewhere between the maximum of the likelihood and the mode of the prior. Note that the contour lines in the different plots were not made to correspond to one another, so comparing contour lines is not very helpful.

c. This new prior is concentrated over positive values for β reflecting a prior belief that higher doses are more likely to cause death in the animals, rather than less likely. It also imposes a relatively narrow distribution on α, the zero-dose likelihood of death. Because the likelihood roughly conforms to this, the posterior distribution will be narrower than that of the non-informative prior.

3.14

Proof that the posterior density for the bioassay example has a finite integral over the range $(\alpha, \beta) \in (-\infty, \infty) \times (-\infty, \infty)$, assuming that $n_i > y_i$ for some i:

$$
\begin{aligned}
p(\alpha, \beta | n, y, x) &\propto p(\alpha, \beta) p(y | \alpha, \beta, n, x) \\
&\propto 1 \times \Pi_{i=1}^k p(y_i | \alpha, \beta, n_i, x_i) \\
&\propto \Pi_{i=1}^k \left(\text{logit}(\alpha + \beta x_i) \right)^{y_i} \left(1 - \text{logit}(\alpha + \beta x_i) \right)^{n_i - y_i} \\
&\propto \Pi_{i=1}^k \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{n_i - y_i} \\
&< \Pi_{i=1}^k \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{n_i - y_i} \quad \text{for } \alpha + \beta x_i \geq 0 \\
&< \Pi_{i=1}^k \left(\frac{1}{\exp(\alpha + \beta x_i)} \right)^{n_i - y_i} \quad \text{for } \alpha + \beta x_i \geq 0.
\end{aligned}
$$

We also have that line 4

$$
\begin{aligned}
&< \Pi_{i=1}^k \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \quad \text{for } \alpha + \beta x_i < 0 \\
&< \Pi_{i=1}^k \left(\frac{\exp(\alpha + \beta x_i)}{1} \right)^{y_i} \quad \text{for } \alpha + \beta x_i < 0 \\
&= \Pi_{i=1}^k \left(\frac{1}{\exp(\alpha + \beta x_i)} \right)^{y_i} \quad \text{for } \alpha + \beta x_i > 0
\end{aligned}
$$

which, just as in line 6, clearly has a finite integral in (α, β) over $\mathbb{R}^2$.

4.1

Consider $\vec{y} = (-2, -1, 0, 1.5, 2.5)$ which are independent samples from a common Cauchy distribution with location parameter θ, a scale parameter of 1, and a unif(0,1) prior for θ.

   a. The posterior, log-posterior, and the first and second derivatives of the log-posterior

are as follows:

$$p(\theta|\vec{y}) \propto p(\theta)p(\vec{y}|\theta)$$

$$\propto 1 \times \Pi_{i=1}^{5} p(y_i|\theta)$$

$$\propto \left(\frac{1}{1 + (-2 - \theta)^2}\right)\left(\frac{1}{1 + (-1 - \theta)^2}\right)\left(\frac{1}{1 + (0 - \theta)^2}\right)$$

$$\left(\frac{1}{1 + (1.5 - \theta)^2}\right)\left(\frac{1}{1 + (2.5 - \theta)^2}\right)$$

$$\log p(\theta|\vec{y}) = C - \log\left(1 + (-2 - \theta)^2\right) - \log\left(1 + (-1 - \theta)^2\right) - \log\left(1 + (\theta)^2\right)$$

$$- \log\left(1 + (1.5 - \theta)^2\right) - \log\left(1 + (2.5 - \theta)^2\right)$$

$$\frac{d}{d\theta}\log p(\theta|\vec{y}) = -\frac{2(2 + \theta)}{1 + (2 + \theta)^2} - \frac{2(1 + \theta)}{1 + (1 + \theta)^2} - \frac{2(\theta)}{1 + (\theta)^2} - \frac{2(1.5 - \theta)}{1 + (1.5 - \theta)^2} - \frac{2(2.5 - \theta)}{1 + (2.5 - \theta)^2}$$

$$\frac{d^2}{d\theta^2}\log p(\theta|\vec{y}) = -\frac{2(1 + (2 + \theta)^2) - (2(2 + \theta))^2}{(1 + (2 + \theta)^2)^2} - \frac{2(1 + (1 + \theta)^2) - (2(1 + \theta))^2}{(1 + (1 + \theta)^2)^2}$$

$$- \frac{2(1 + (\theta)^2) - (2(\theta))^2}{(1 + (\theta)^2)^2} - \frac{2(1 + (1.5 - \theta)^2) - (2(1.5 - \theta))^2}{(1 + (1.5 - \theta)^2)^2}$$

$$- \frac{2(1 + (2.5 - \theta)^2) - (2(2.5 - \theta))^2}{(1 + (2.5 - \theta)^2)^2}$$

b. Setting the first derivative of the log-posterior to 0 and attempting to solve yields a quintic polynomial in θ, with roots (according to Wolfram Alpha) at -0.687, -2.466, and 2.410. Only one of which, $\theta = -0.687$ implies a local maximum, but is outside of the support of θ's posterior density. Evaluating the log-posterior at the endpoints of the interval, [0,1] yields a maximum (thus mode) at $\theta = 0$.

c. Because part (b) resulted in a boundary-solution, the normal approximation for this posterior will not be an adequate one. Nevertheless, I will pretend that the prior distribution had included the mode of -0.687 and continue from there.
The second derivative of the log-likelihood is given above. Evaluating this at $\theta = -0.687$ yields approximately -1.415. By the derivation in the textbook, we conclude that the posterior mode can be approximated by $\theta \sim \mathrm{N}(-0.687, 0.706)$ where $0.706 \approx 1/|-1.415|$.

N(0.687, .707) distribution
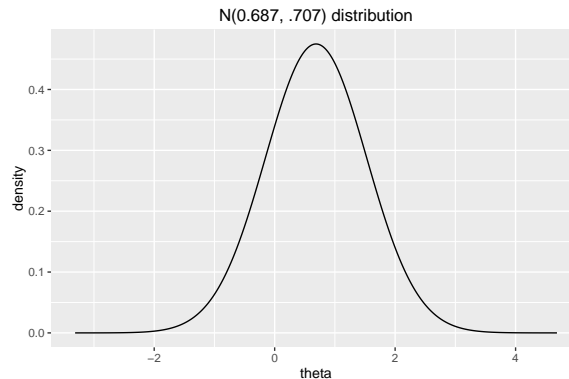
5.9

Question relating to noninformative hyperprior distribution from the example in Section 5.3.

a. A uniform prior density on $(\log(\alpha/\beta), \log(\alpha+\beta))$ implies the following joint distribution on (α, β):

$$p_{\alpha,\beta}(\alpha, \beta) = 1 \times |J| = 1 \times \begin{vmatrix} \frac{1}{\alpha} & \frac{-1}{\beta} \\ \frac{1}{\alpha+\beta} & \frac{1}{\alpha+\beta} \end{vmatrix} = \frac{\alpha+\beta}{\alpha\beta(\alpha+\beta)} = \frac{1}{\alpha\beta}$$

Looking at the integral under consideration

$$p(\alpha, \beta | y) \propto \frac{1}{\alpha\beta} \Pi_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}$$

$$> \frac{1}{\alpha\beta} \Pi_{j=1}^{J} \left(\frac{1}{\alpha+\beta+n_j-1} \right)^{n_j} \left(\frac{(\alpha+y_j-1)^{y_i}(\beta+n_j-y_j-1)^{n_j-y_j}}{1} \right)$$

$$= \frac{1}{\alpha\beta} \Pi_{j=1}^{J} \left(\frac{\alpha+y_j-1}{\alpha+\beta+n_j-1} \right)^{y_j} \left(\frac{\beta+n_j-y_j-1}{\alpha+\beta+n_j-1} \right)^{n_j-y_j}$$

where considering n_j and y_j as constants, it is clear that for any fraction α/β the function is proportional to $1/(\alpha\beta)$, which has an infinite integral.

b. A uniform prior on $(\alpha + \beta)^{-1/2}$ and on $\alpha/(\alpha + \beta)$ is equivalent to the following prior on (α, β):

$$p_{\alpha,\beta}(\alpha, \beta) = 1 \times |J| = 1 \times \begin{vmatrix} -\frac{1}{2}(\alpha+\beta)^{-3/2} & -\frac{1}{2}(\alpha+\beta)^{-3/2} \\ \beta/(\alpha+\beta)^2 & -\alpha/(\alpha+\beta)^2 \end{vmatrix} \propto (\alpha+\beta)^{-5/2}$$

5.12

Using the laws of total variance and expectation, one can find that

$$E[\theta_j|\tau, y] = E_\mu[E_{\theta_j}[\theta_j|\mu, \tau, y]|\tau, y] = E_\mu \left[\frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \bigg| \tau, y \right] = \frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau^2}\hat{\mu}}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$

$$\text{var}[\theta_j|\tau, y] = E_\mu[\text{var}_{\theta_j}[\theta_j|\mu, \tau, y]|\tau, y] + \text{var}_\mu[E_{\theta_j}[\theta_j|\mu, \tau, y]|\tau, y] = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} + \left(\frac{\frac{1}{\tau^2}}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \right)^2 V_\mu$$

1 Appendix: R code

```
# Homework 5, BDA
setwd("/home/joey/Desktop/Bayesian Data Analysis/")
library(mvtnorm)
library(ggplot2)
```

```
alpha <- seq(from=-5, to=10, length.out = 500)
beta <- seq(from=-10, to=40, length.out = 500)
meanV <- c(0,10)
covM <- matrix(data = c(4, 10, 10, 100), ncol = 2)
y <- c(0,1,3,5)
x <- c(-0.86, -0.3, -0.05, 0.73)
n <- c(5,5,5,5)

lgpost <- function(alpha, beta) {
z <- -(1/2)*alpha^2+alpha*(beta-10)/10-(1/50)*(beta-10)^2 + alpha*(sum(y)) +
beta*(sum(x*y)) - 5*log(1+exp(alpha+beta*x[1])) - 5*log(1+exp(alpha+beta*x[2])) -
5*log(1+exp(alpha+beta*x[3])) - 5*log(1+exp(alpha+beta*x[4]))
return(z)
}

# Compute vectors of all combinations of alpha, beta in grid
alphaVec <- rep(alpha, times=500)
betaVec <- rep(beta, each=500)

df <- data.frame("alpha"=alphaVec, "beta"=betaVec)

lz <- lgpost(alphaVec, betaVec)
z <- exp(lz)
df <- cbind.data.frame(df, z)
names(df)[3] <- "prob"

df$prob <- df$prob/sum(df$prob)

View(df)

ggplot(data=df, aes(x=alpha, y=beta, z=prob)) +
geom_contour(bins=10) +
xlim(c(-4,10)) +
ylim(c(-10,40)) +
labs(title="Posterior distribution of alpha and beta")

dev.print(pdf, "PostAlphaBeta")

# Sample from posterior, using conditional sampling.
# First, create matrix of densities:
pMat <- t(matrix(df$prob, nrow=500))


marginalBeta <- apply(pMat, MARGIN = 1, sum)
a <- vector(length=1000)
```

```r
b <- vector(length=1000)
for (i in 1:1000) {
b[i] <- sample(size = 1, x = beta, prob = marginalBeta)
a[i] <- sample(size = 1, x = alpha, prob = pMat[which(beta==b[i]),])
}
# Adding an runif so the distribution is continuous.
a <- a + runif(n=1000, min = -0.01503006, max=0.01503006)
b <- b + runif(n=1000, min= -0.0501002, max=0.0501002)

dfSamples <- data.frame(a,b)

ggplot(dfSamples, aes(x=a, y=b)) +
geom_point(size=0.1) +
labs(x="alpha", y="beta", title="One Thousand Samples from Posterior") +
xlim(c(-5,10)) + ylim(c(-10,40))

dev.print(pdf, "PostSamples")

# Now, plotting -alpha/beta for the LD50:
ggplot() +
geom_histogram(aes(x=-(a/b)), color="black", bins = 40) +
labs(x="LD 50", title="Samples from Posterior of LD50")

dev.print(pdf, "LD50Samples")

# Contour plot for Prior:

lgprior <- function(alpha, beta) {
z <- -(1/2)*alpha^2+alpha*(beta-10)/10-(1/50)*(beta-10)^2
return(z)
}

alphaVec <- rep(alpha, times=500)
betaVec <- rep(beta, each=500)

df <- data.frame("alpha"=alphaVec, "beta"=betaVec)

lz <- lgprior(alphaVec, betaVec)
z <- exp(lz)
df <- cbind.data.frame(df, z)
names(df)[3] <- "prob"

df$prob <- df$prob/sum(df$prob)

View(df)
```

```r
ggplot(data=df, aes(x=alpha, y=beta, z=prob)) +
geom_contour(bins=10) +
xlim(c(-4,10)) +
ylim(c(-10,40)) +
labs(title="Prior distribution of alpha and beta")

dev.print(pdf, "PriorAlphaBeta")

# Contour plot for likelihood:

lgliklihood <- function(alpha, beta) {
z <- alpha*(sum(y)) + beta*(sum(x*y)) - 5*log(1+exp(alpha+beta*x[1])) -
5*log(1+exp(alpha+beta*x[2])) -
5*log(1+exp(alpha+beta*x[3])) - 5*log(1+exp(alpha+beta*x[4]))
return(z)
}

alphaVec <- rep(alpha, times=500)
betaVec <- rep(beta, each=500)

df <- data.frame("alpha"=alphaVec, "beta"=betaVec)

lz <- lgliklihood(alphaVec, betaVec)
z <- exp(lz)
df <- cbind.data.frame(df, z)
names(df)[3] <- "prob"

df$prob <- df$prob/sum(df$prob)

View(df)

ggplot(data=df, aes(x=alpha, y=beta, z=prob)) +
geom_contour(bins=10) +
xlim(c(-4,10)) +
ylim(c(-10,40)) +
labs(title="Likelihood for alpha and beta")

dev.print(pdf, "LikelihoodAlphaBeta")

# Approximation to the posterior in problem 4.1
theta <- seq(from = -3.313, to=4.687, length.out = 1000)
dens <- dnorm(theta, mean=0.687, sd = sqrt(.706))
df <- cbind.data.frame(theta, dens)
```

```
ggplot(data=df, aes(x=theta, y=dens)) +
geom_line() +
labs(title="N(0.687, .707) distribution", y="density")

dev.print(pdf, "DensNormalApprox.pdf")
```