

# MAS3301 Bayesian Statistics: Project Solutions

Semester 2

2008-9

## 1 Background

When analysing proportions (eg. with a binomial model) we can often use a beta prior distribution. However sometimes, particularly in more complicated cases, it is convenient to use a different form of prior distribution. Typically we transform the proportion to put it onto a  $(-\infty, \infty)$  scale, rather than  $(0, 1)$ , and then give the transformed proportion a normal distribution. In particular this makes it easy to give a relationship to two or more proportions in our prior distribution.

One form of transformation which is often used is known as “logits.” If our proportion is  $\theta$  then we transform this to

$$\eta = \log \left( \frac{\theta}{1 - \theta} \right) \quad (1)$$

and then give  $\eta$  a normal prior distribution.

This project is concerned with analyses of this type.

## 2 Data

Each student will use a different data set. Each data set is identified by a reference number. A list of students and reference numbers and a separate list giving the data set for each reference number are provided.

## 3 Tasks

1. Find an expression for the pdf of  $\theta$  when  $\eta$  is given by (1) and  $\eta \sim N(m, v)$ . You may wish to refer to Section 5.4 of the lecture notes.
2. Many people regard a beta(1,1) distribution as a “noninformative” prior since the pdf is a constant. It might be thought that, if we use a logit transformation and then give  $\eta$  a  $N(0, v)$  prior distribution, then this will be “noninformative” if we make the variance  $v$  large. However things are not as simple as this because, for large enough  $v$ , the prior distribution for  $\theta$  becomes bimodal.
  - (a) Use R to plot graphs of the pdf of  $\theta$  when  $\eta$  is given by (1) and  $\eta \sim N(0, v)$  for a range of values of  $v$  in  $1 \leq v \leq 4$  and hence deduce, at least approximately, the value of  $v$  at which the distribution of  $\theta$  becomes bimodal.
  - (b) Find analytically the exact value of  $v$  at which the distribution of  $\theta$  becomes bimodal when  $\eta \sim N(0, v)$ . *Hint: The density of  $\theta$  is symmetric about  $\theta = 1/2$  so look at the second derivative of the log density at this point.*
3. Each patient in a sample of  $n_1$  patients with a certain chronic illness is given a treatment. The number  $x_1$  who show a particular response is recorded. Your values of  $n_1$  and  $x_1$  are

in your data set. Given the value of a parameter  $\theta_1$ , we regard  $x_1$  as an observation from the binomial( $n_1, \theta_1$ ) distribution. The value of  $\theta_1$  is, however, unknown so we give it a prior distribution by first transforming to  $\eta_1$  using (1) and then giving  $\eta_1$  a normal  $n(m_1, v_1)$  prior distribution.

- (a) In our prior beliefs, the lower quartile of  $\theta_1$  is 0.4 and the upper quartile is 0.8. Use these to find the lower and upper quartiles of  $\eta_1$  and hence find the values of  $m_1$  and  $v_1$ .
  - (b) Use numerical methods to find the posterior density of  $\theta_1$  given your data and plot a graph showing both the prior and posterior densities.
  - (c) Find the posterior probability that  $\theta_1 > 0.75$ .
4. In another experiment with the same disease two further groups of patients are used. Group 2 contains  $n_2$  patients and is given treatment  $T_2$ . Group 3 contains  $n_3$  patients and is given treatment  $T_3$ . The number in Group  $j$  who show the response is  $x_j$ . Given the values of parameters  $\theta_2, \theta_3, x_j$  is regarded as an observation from the binomial( $n_j, \theta_j$ ) distribution and  $x_2, x_3$  are independent.

Our prior distribution for  $\theta_2, \theta_3$  is as follows. We transform  $\theta_2, \theta_3$  to  $\eta_2, \eta_3$  using (1) and then give  $\eta_1, \eta_2$  a bivariate normal prior distribution with parameters as follows.

$$\begin{aligned} E(\eta_2) = E(\eta_3) &= 0.4 \\ \text{var}(\eta_2) = \text{var}(\eta_3) &= 0.6 \\ \text{covar}(\eta_2, \eta_3) &= 0.3 \end{aligned}$$

- (a) Use R to find and plot the joint posterior density of  $\theta_2$  and  $\theta_3$ .
- (b) Find and plot the prior and posterior probability density functions of the log relative risk  $\gamma$  where

$$\begin{aligned} \gamma &= \log \left( \frac{\theta_2/(1-\theta_2)}{\theta_3/(1-\theta_3)} \right) \\ &= \log \left( \frac{\theta_2}{1-\theta_2} \right) - \log \left( \frac{\theta_3}{1-\theta_3} \right) \\ &= \eta_2 - \eta_3 \end{aligned}$$

- (c) Comment on your results.

*Hint: Part 4 of Practical 1 should help with Task 4. You will need to modify the R functions appropriately.*

## 4 Submission

Reports are to be submitted to the General Office of the School of Mathematics and Statistics no later than 4.00pm on **Thursday 30th April** (Week 10). As this counts for 10% of the module mark you will need to hand in your report at the reception window.

## 5 Solution

1. Find an expression for the pdf of  $\theta$  when  $\eta$  is given by (1) and  $\eta \sim N(m, v)$ .

We have

$$\eta = \log \left( \frac{\theta}{1-\theta} \right) \quad \text{so} \quad \theta = \frac{e^\eta}{1+e^\eta}$$

and

$$\frac{d\eta}{d\theta} = \left( \frac{1-\theta}{\theta} \right) \left\{ \frac{(1-\theta) + \theta}{(1-\theta)^2} \right\} = \frac{1}{\theta(1-\theta)} = \frac{(1+e^\eta)^2}{e^\eta}.$$

So the density of  $\theta$  is

$$\begin{aligned} f_\theta(\theta) &= f_\eta(\eta) \frac{d\eta}{d\theta} \\ &= (2\pi v)^{-1/2} \exp \left\{ -\frac{1}{2v}(\eta - m)^2 \right\} \frac{(1+e^\eta)^2}{e^\eta} \\ &= (2\pi v)^{-1/2} \exp \left\{ -\frac{1}{2v} \left( \log \left[ \frac{\theta}{1-\theta} \right] - m \right)^2 \right\} \theta^{-1} (1-\theta)^{-1} \end{aligned}$$

[5 marks]

2. (a) Use R to plot graphs of the pdf of  $\theta$  when  $\eta$  is given by (1) and  $\eta \sim N(0, v)$  for a range of values of  $v$  in  $1 \leq v \leq 4$  and hence deduce, at least approximately, the value of  $v$  at which the distribution of  $\theta$  becomes bimodal.

Figure 1 shows some R commands to do this.

The resulting graph is shown in Figure 2. The curve becomes bimodal at approximately  $v = 2.0$ .

[10 marks]

- (b) Find analytically the exact value of  $v$  at which the distribution of  $\theta$  becomes bimodal when  $\eta \sim N(0, v)$ .

The density is

$$f_\theta(\theta) = (2\pi v)^{-1/2} \exp \left\{ -\frac{1}{2v} \left( \log \left[ \frac{\theta}{1-\theta} \right] - m \right)^2 \right\} \theta^{-1} (1-\theta)^{-1}$$

The log density is

$$g(\theta) = \log[f_\theta(\theta)] = -\frac{1}{2} \log(2\pi v) - \frac{1}{2v}(\eta - m)^2 - \log \theta - \log(1-\theta).$$

The first derivative is

$$\frac{dg(\theta)}{d\theta} = -\frac{1}{v}(\eta - m) \frac{d\eta}{d\theta} - \frac{1}{\theta} + \frac{1}{1-\theta}.$$

The second derivative is

$$\frac{d^2g(\theta)}{d\theta^2} = -\frac{1}{v} \left( \frac{d\eta}{d\theta} \right)^2 - \frac{1}{v}(\eta - m) \frac{d^2\eta}{d\theta^2} + \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2}$$

where

$$\frac{d\eta}{d\theta} = \frac{1}{\theta(1-\theta)} \quad \text{and} \quad \frac{d^2\eta}{d\theta^2} = -\frac{1-2\theta}{\theta^2(1-\theta)^2}.$$

```

function(v)
{nv<-length(v)
 theta<-seq(0.001,0.999,0.001)
 eta<-log(theta/(1-theta))
 jac<-theta*(1-theta)
 thetadens<-matrix(nrow=999,ncol=nv)
 for (i in 1:nv)
   {sd<-sqrt(v[i])
    etadens<-dnorm(eta,0,sd)
    thetadens[,i]<-etadens/jac
   }
 top<-max(thetadens)
 plot(theta,thetadens[,1],ylim=c(0,top),type="l",xlab=expression(theta),ylab="Density")
 for (i in 2:nv)
   {lines(theta,thetadens[,i],lty=i)
   }
 }
}

```

Figure 1: R commands for Task 2a

Put  $\theta = 1/2$  so  $\eta = 0$ ,

$$\frac{d\eta}{d\theta} = 4, \quad \frac{d^2\eta}{d\theta^2} = 0$$

and

$$\frac{d^2g(\theta)}{d\theta^2} = -\frac{16}{v} + 4 + 4.$$

Set this equal to zero and solve for  $v$ . This give  $v = 2$ .

The distribution becomes bimodal when  $v = 2$ .

[10 marks]

3. (a) *In our prior beliefs, the lower quartile of  $\theta_1$  is 0.4 and the upper quartile is 0.8. Use these to find the lower and upper quartiles of  $\eta_1$  and hence find the values of  $m_1$  and  $v_1$ .*  
The lower quartile of  $\eta_1$  is

$$\log\left(\frac{0.4}{1-0.4}\right) = -0.405465.$$

The upper quartile of  $\eta_1$  is

$$\log\left(\frac{0.8}{1-0.8}\right) = 1.386294.$$

The normal distribution is symmetric so the mean is half way between the quartiles. So  $m = 0.49041$ .

The upper and lower quartiles of a normal distribution are  $\pm 0.67449$  standard deviations from the mean. So the standard deviation is

$$\frac{1.386294 + 0.405465}{2 \times 0.67449} = 1.32823.$$

Hence the variance is  $v = 1.32823^2 = \underline{1.7642}$ .

[5 marks]

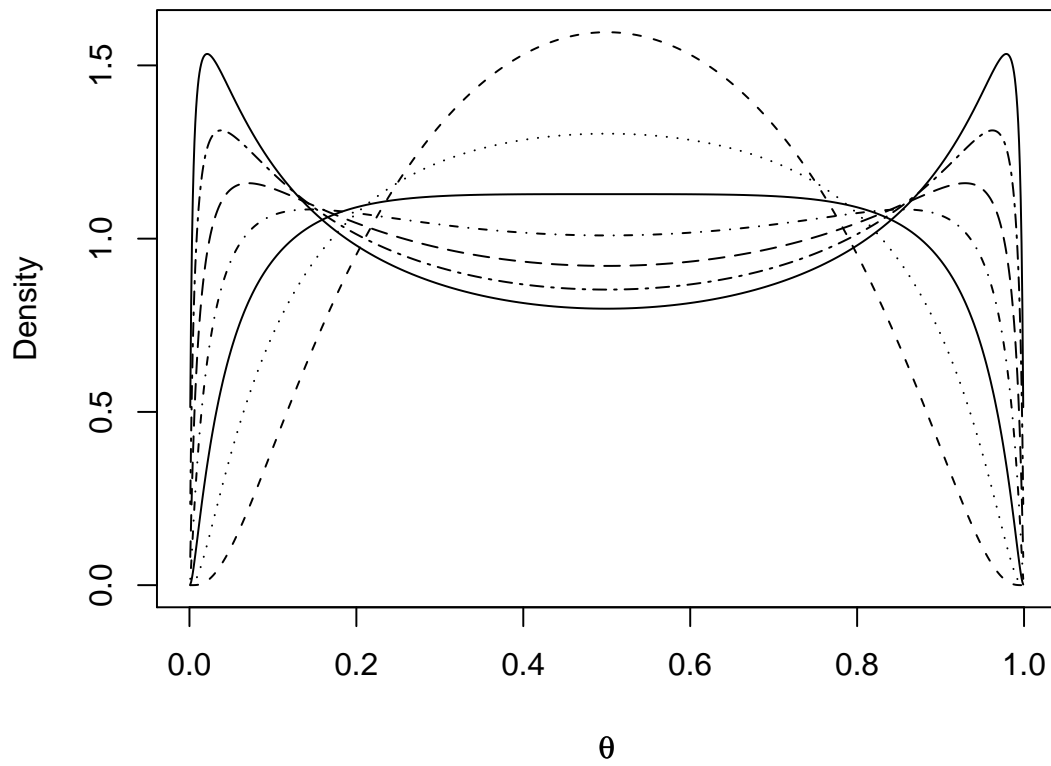


Figure 2: Densities of  $\theta$  for  $v = 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0$  (in descending order at  $\theta = 0.5$ ). The solid line is for  $v = 2.0$ .

```

task3<-function(n1,x1)
{q1<-log(0.4/(1-0.4))
q3<-log(0.8/(1-0.8))
m<-(q1+q3)/2
sd<-(q3-m)/qnorm(0.75,0,1)
v<-sd^2
theta<-seq(0.001,0.999,0.001)
eta<-log(theta/(1-theta))
etadens<-dnorm(eta,m,sd)
logprior<-log(etadens)-log(theta)-log(1-theta)
loglik<-x1*log(theta)+(n1-x1)*log(1-theta)
logpos<-logprior+loglik
logpos<-logpos-max(logpos)
pos<-exp(logpos)
pos<-pos/(sum(pos)*0.001)
prior<-exp(logprior)
plot(theta,pos,type="l",xlab=expression(theta[1]),ylab="Density")
lines(theta,prior,lty=2)
posprob<-sum(pos[theta>0.75])*0.001
ans<-list(m=m,v=v,theta=theta,prior=prior,posterior=pos,posprob=posprob)
ans
}

```

Figure 3: R function for Task 3.

- (b) *Use numerical methods to find the posterior density of  $\theta_1$  given your data and plot a graph showing both the prior and posterior densities.*

The log likelihood, apart from a constant, is

$$\log[\theta^{x_1}(1-\theta)^{n_1-x_1}] = x_1 \log(\theta) + (n_1 - x_1) \log(1 - \theta).$$

We can add this to the log prior density then exponentiate and normalise.

Figure 3 shows a R function to do the calculations.

Each student has different data but Figure 4 shows the graph for Data Set 1. The graph was made using the following R command.

```
test<-task3(25,18)
```

[15 marks]

- (c) *Find the posterior probability that  $\theta_1 > 0.75$ .*

We find this probability as

$$\int_{0.75}^1 f_{\theta_1}^{(1)} d\theta_1$$

where  $f_{\theta_1}^{(1)}$  is the posterior density of  $\theta_1$ . This is calculated (approximately) in the function `task3` in the line

```
posprob<-sum(pos[theta>0.75])*0.001
```

Each student's answer will be different but the answer for Data Set 1 is extracted as follows.

```
test$posprob
[1] 0.3371888
```

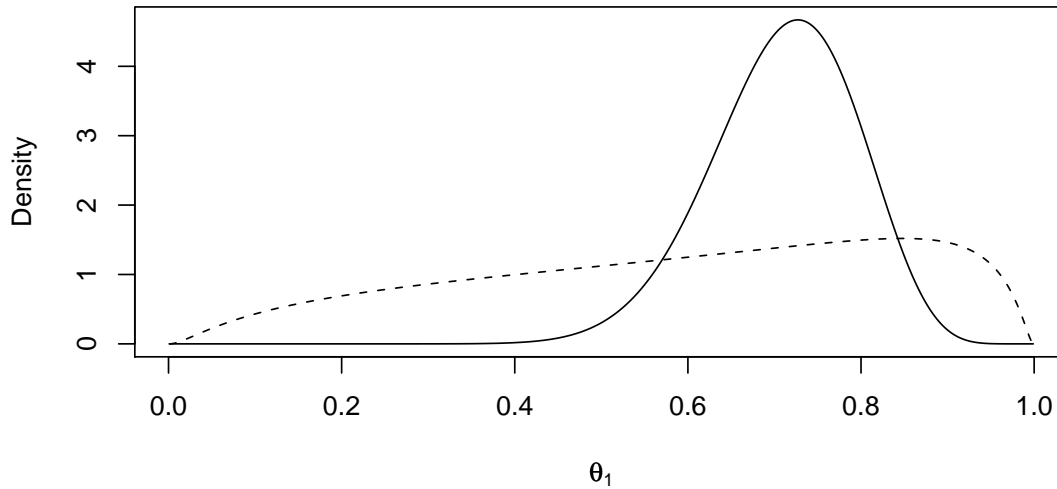


Figure 4: Prior (dashes) and posterior (solid) densities for  $\theta_1$  in Task 3. (Data Set 1).

(Note also that there may also be slight numerical differences if students have used a different set of points for evaluation in the numerical integration).

[5 marks]

4. (a) Use R to find and plot the joint posterior density of  $\theta_2$  and  $\theta_3$ .

Because we are given the prior in terms of  $\eta_2$ ,  $\eta_3$ , but we want the posterior for  $\theta_2$ ,  $\theta_3$  we need to transform between them and for this we need a Jacobian. So

$$\begin{aligned}
 J &= \left| \begin{array}{cc} \frac{\partial \eta_1}{\partial \theta_1} & \frac{\partial \eta_1}{\partial \theta_2} \\ \frac{\partial \eta_2}{\partial \theta_1} & \frac{\partial \eta_2}{\partial \theta_2} \end{array} \right| = \left| \begin{array}{cc} \frac{1}{\theta_1(1-\theta_1)} & 0 \\ 0 & \frac{1}{\theta_2(1-\theta_2)} \end{array} \right| \\
 &= \theta_1^{-1}(1-\theta_1)^{-1}\theta_2^{-1}(1-\theta_2)^{-1}
 \end{aligned}$$

Figure 5 shows a R function to do the calculations.

Each student has different data so the posterior distributions will be different. Figure 6 shows the plot for Data Set 1. I have superimposed the line  $\theta_2 = \theta_3$ .

The graph was produced with the following R commands.

```

> theta2<-seq(0.01,0.99,0.01)
> theta3<-seq(0.01,0.99,0.01)
> prior<-c(0.4,0.4,0.6,0.6,0.3)
> n<-c(55,20)
> x<-c(32,9)
> posterior<-logit1(theta2,theta3,n,x,prior)
> contour(theta2,theta3,posterior,xlab=expression(theta[2]),ylab=expression(theta[3]))
> abline(0,1)

```

Note that some students may have changed the ranges to avoid large areas of negligible density. This is fine. In fact it is a good idea in terms of the numerical integration.

[20 marks]

```

logit1<-function(theta1,theta2,n,x,prior)
{# Evaluates posterior density for logit example.
# prior is mean1, mean2, var1, var2, covariance
n1<-length(theta1)
n2<-length(theta2)
step1<-theta1[2]-theta1[1]
step2<-theta2[2]-theta2[1]
theta1<-matrix(theta1,nrow=n1,ncol=n2)
theta2<-matrix(theta2,nrow=n1,ncol=n2,byrow=T)
eta1<-log(theta1/(1-theta1))
eta2<-log(theta2/(1-theta2))
sd1<-sqrt(prior[3])
sd2<-sqrt(prior[4])
delta1<-(eta1-prior[1])/sd1
delta2<-(eta2-prior[2])/sd2
r<-prior[5]/(sd1*sd2)
d<-1-r^2
logprior<- -(delta1^2 + delta2^2 - 2*r*delta1*delta2)/(2*d)
logJ<- log(theta1)+log(1-theta1)+log(theta2)+log(1-theta2)
logprior<-logprior-logJ
loglik<-x[1]*log(theta1)+(n[1]-x[1])*log(1-theta1)+x[2]*log(theta2)+(n[2]-x[2])*log(1-theta2)
logpos<-logprior+loglik
logpos<-logpos-max(logpos)
posterior<-exp(logpos)
int<-sum(posterior)*step1*step2
posterior<-posterior/int
posterior
}

```

Figure 5: R function for Task 4a.



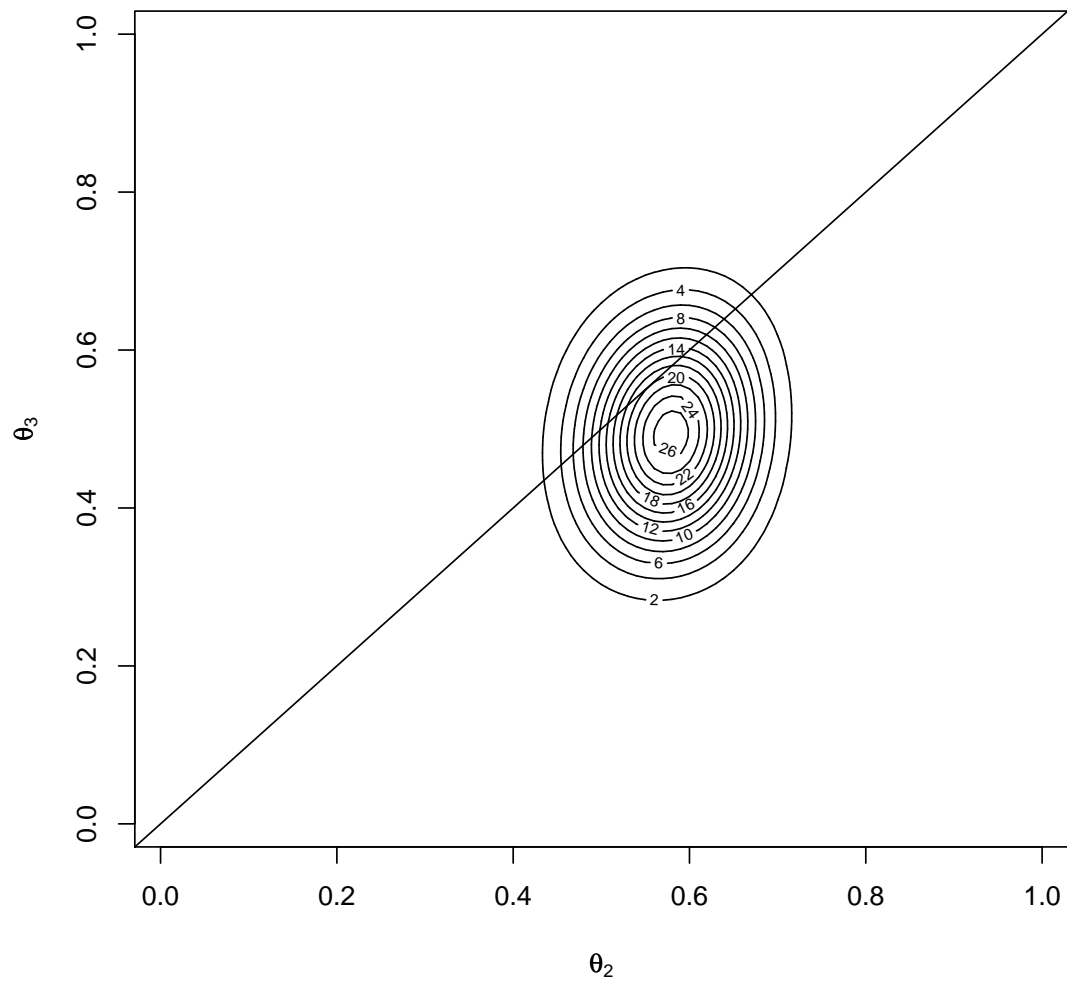


Figure 6: Joint posterior density of  $\theta_2$  and  $\theta_3$  in Task 4. (Data Set 1).

```

logit2<-function(eta2,gamma,n,x,prior)
{# Evaluates posterior density for 2009 project.
# prior is mean1, mean2, var1, var2, covariance for eta2, eta3, as in logit1
n1<-length(eta2)
ng<-length(gamma)
step1<-eta2[2]-eta2[1]
stepg<-gamma[2]-gamma[1]
eta2<-matrix(eta2,nrow=n1,ncol=ng)
gamma<-matrix(gamma,nrow=n1,ncol=ng,byrow=T)
eta3<-eta2-gamma
theta2<-exp(eta2)/(1+exp(eta2))
theta3<-exp(eta3)/(1+exp(eta3))
sd2<-sqrt(prior[3])
sdg<-sqrt(prior[3]+prior[4]-2*prior[5])
delta2<-(eta2-prior[1])/sd2
deltag<-(gamma-(prior[1]-prior[2]))/sdg
r<-(prior[3]-prior[5])/(sd2*sdg)
d<-1-r^2
logprior<- -(delta2^2 + deltag^2 - 2*r*delta2*deltag)/(2*d)
loglik<-x[1]*log(theta2)+(n[1]-x[1])*log(1-theta2)+x[2]*log(theta3)+(n[2]-x[2])*log(1-theta3)
logpos<-logprior+loglik
logpos<-logpos-max(logpos)
posterior<-exp(logpos)
int<-sum(posterior)*step1*stepg
posterior<-posterior/int
marginal<-colSums(posterior)*step1
ans<-list(joint=posterior,marginal=marginal)
ans
}

```

Figure 7: R function for Task 4b.

- (b) Find and plot the prior and posterior probability density functions of  $\gamma = \eta_2 - \eta_3$ . (Note, I know that this is actually the log odds, not the log relative risk. That was a mistake when I typed the question sheet. It does not affect the solution though).

Since  $\gamma = \eta_2 - \eta_3$  it has a normal prior distribution. In fact  $\eta_2$  and  $\gamma$  have a bivariate normal prior distribution (as do  $\eta_3$  and  $\gamma$ ). The prior mean of  $\gamma$  is  $E_0(\gamma) = 0$ . The prior variance of  $\gamma$  is  $\text{var}_0(\gamma) = 0.6 + 0.6 - 2 \times 0.3 = 0.6$ . The prior covariance of  $\eta_2$  and  $\gamma$  is

$$\text{covar}_0(\eta_2, \gamma) = \text{var}_0(\eta_2) - \text{covar}_0(\eta_2, \eta_3) = 0.6 - 0.3 = 0.3.$$

There are various different ways to calculate the marginal posterior of  $\gamma$ . Figure 7 shows a R function for one method. This function returns the joint posterior density of  $\eta_2$  and  $\gamma$  as well as the marginal posterior density of  $\gamma$ . This is so that we can check whether we have chosen appropriate ranges for  $\eta_2$  and  $\gamma$ .

Each student has different data so the posterior distributions will be different. Figure 8 shows the joint density plot for Data Set 1. Figure 9 shows the marginal density plot for Data Set 1. I have superimposed a line at  $\gamma = 0$  corresponding to  $\theta_2 = \theta_3$ .

The graphs were produced with the following R commands.

```

> eta2<-seq(-1.5,1.5,0.03)
> gamma<-seq(-2,2,0.04)
> test2<-logit2(eta2,gamma,n,x,prior)

```

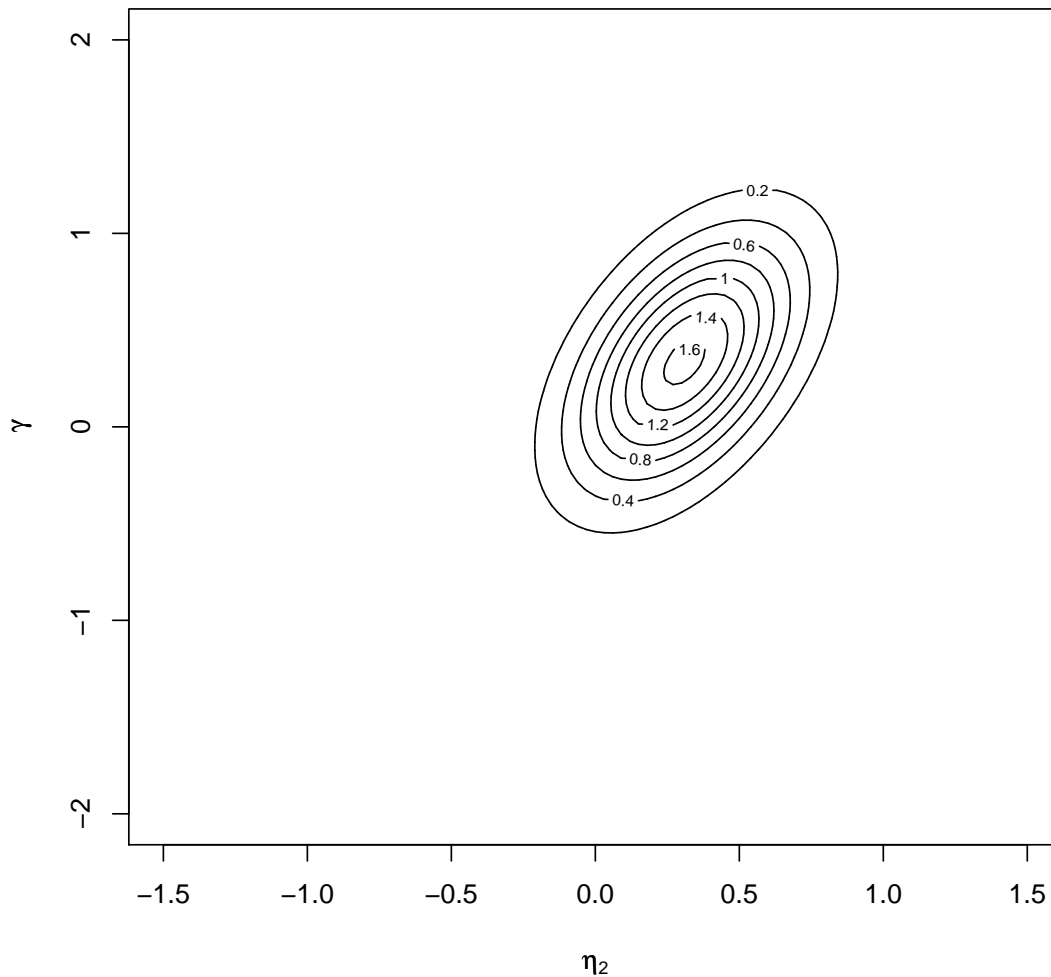


Figure 8: Joint posterior density of  $\eta_2$  and  $\gamma$  in Task 4b. (Data Set 1).

```
> post2<-test2$joint
> contour(eta2,gamma,post2,xlab=expression(eta[2]),ylab=expression(gamma))
> gammapost<-test2$marginal
> gammaprior<-dnorm(gamma,0,sqrt(0.6))
> plot(gamma,gammapost,type="l",xlab=expression(gamma),ylab="Density")
> lines(gamma,gammaprior,lty=2)
> abline(h=0)
> abline(v=0,lty=3)
```

[20 marks]

(c) *Comment on your results.*

Results will vary from one data set to another. In the case of Data Set 1, most of the posterior probability is in the region where  $\gamma > 0$ , that is where  $\theta_2 > \theta_3$ . This shows that, after seeing the data, it appears likely that a greater proportion of patients give the response when given treatment  $T_2$  than when given treatment  $T_3$ . (It would not be

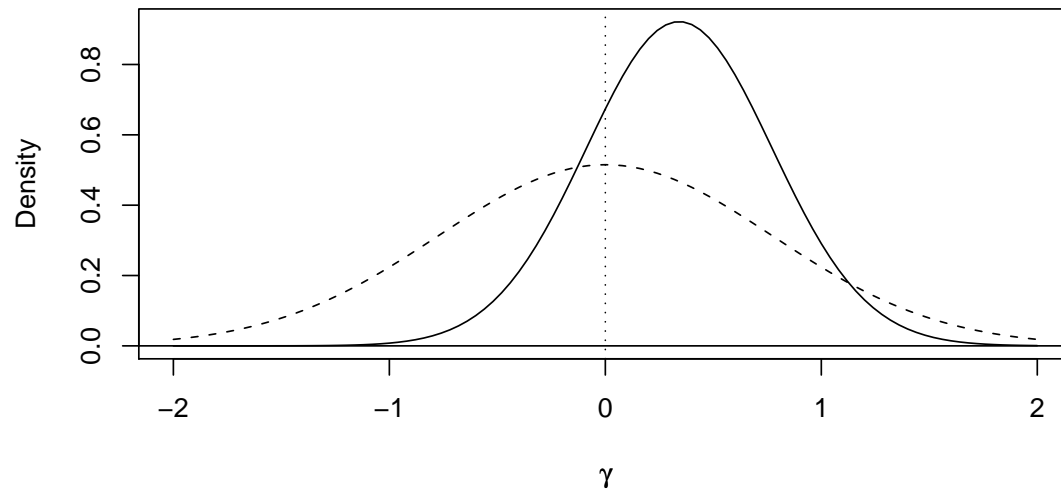


Figure 9: Prior (dashes) and posterior (solid) marginal densities for  $\gamma$  in Task 4b. (Data Set 1). The vertical dotted line shows where the two treatments are equal.

difficult to calculate the posterior probability of this). Students should be given marks for comments which show that they understand the meaning of the graphs.

[10 marks]