Naren Suri – Home Work -2 nsuri@iu.edu Bayesian statistics

3.3

a)

Tumor counts: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts for the two populations are

y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6);

y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7).

a)

Find the posterior distributions, means, variances and 95% quantilebased confidence intervals for _A and _B, assuming a Poisson sampling distribution for each group and the following prior distribution:

$\Theta_A \sim$ gamma(120,10),

$\Theta_B \sim$ gamma(12,1),

$p(\Theta_A, \Theta_B) = p(\Theta_A) \times p(\Theta_B)$

Given

For Θ_A

y_A=c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)

n_yA=length(y_A)

Prior_a_A=120

Prior_b_A=10

For Θ_B

y_B=c(11,11,10,9,9,8,7,10,6,8,8,9,7)

prior_a_B=12

prior_b_B=1

n_B=13

Code block

```
#3.3 a
#For Theta A

y_A=c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
n_yA=length(y_A)
Prior_a_A=120
Prior_b_A=10

posterior_mean_A=(Prior_a_A+sum(y_A))/(Prior_b_A+n_yA)
posterior_Var_A=(Prior_a_A+sum(y_A))/((Prior_b_A+n_yA)^2)
ConfIntervalA=qgamma( c (.025, .975),Prior_a_A+sum(y_A),Prior_b_A+n_yA)
```

```
#For Theta B
y_B=c(11,11,10,9,9,8,7,10,6,8,8,9,7)
prior_a_B=12
prior_b_B=1
n_B=13

#Applying the calculations

posterior_mean_B=(prior_a_B+sum(y_B))/(prior_b_B+n_B)
posterior_Var_B=(prior_a_B+sum(y_B))/((prior_b_B+n_B)^2)
```

```r
ConfintervalB=qgamma( c (.025, .975),prior_a_B+sum(y_B),prior_b_B+n_B)

#Reporting results for theta A and theta B

cat(sprintf("theta A \n posterior_mean_A= %f \n posterior_Var_A= %f \n Confin
tervalA= %f \n" , posterior_mean_A,posterior_Var_A,ConfIntervalA ))

## theta A
##  posterior_mean_A= 11.850000
##  posterior_Var_A= 0.592500
##  ConfintervalA= 10.389238
##  theta A
##  posterior_mean_A= 11.850000
##  posterior_Var_A= 0.592500
##  ConfintervalA= 13.405448

cat(sprintf(" theta B \n posterior_mean_B= %f \n posterior_Var_B= %f \n Confi
ntervalB= %f \n" , posterior_mean_B,posterior_Var_B,ConfintervalB ))

##  theta B
##  posterior_mean_B= 8.928571
##  posterior_Var_B= 0.637755
##  ConfintervalB= 7.432064
##   theta B
##  posterior_mean_B= 8.928571
##  posterior_Var_B= 0.637755
##  ConfintervalB= 10.560308
```

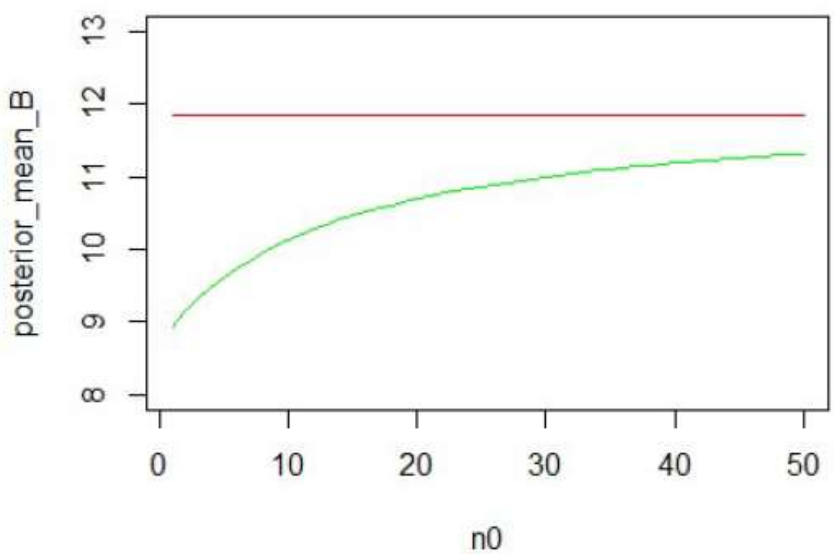
b)

Compute and plot the posterior expectation of Θ_B under the prior distribution
Θ_B~ gamma($12 \times n_0$, n_0)
for each value of n_0 €{1, 2, . . . , 50}.

Describe what sort of prior beliefs about Θ_B would be necessary in order for the posterior expectation of ΘB to be close to that of Θ_A.

```r
#For Theta A

y_A=c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
n_yA=length(y_A)
Prior_a_A=120
Prior_b_A=10

posterior_mean_A=(Prior_a_A+sum(y_A))/(Prior_b_A+n_yA)



# n0 C {1,2,.....,50}
```

```r
# For theta B
n0=seq(1,50)
prior_a_B=12*n0
prior_b_B=n0
y_B=c(11,11,10,9,9,8,7,10,6,8,8,9,7)
n_B=length(y_B)

sum_y_B=sum(y_B)

posterior_mean_B = (prior_a_B+sum_y_B)/(prior_b_B+n_B)


plot(posterior_mean_B, type = "l", col="green" , ylim = c(8,13) ,xlab="n0")
lines(n0,as.vector(rep(posterior_mean_A,50)),type = "l", col="red", xlab="n0"
)
```



The graph above shows the plot of posterior expectation of B. From the figure we see that the posterior expectation of B is approaching to that of A. For The posterior expectation of B to be close to that of A , n_0 should proceed further to 50. Or for the posterior expectation to approach to that of A it should have prior beliefs like A, when n_0 =10 the prior belief of B is similar to that of A.

c)

Should knowledge about population A tell us anything about population B?

Discuss whether or not it makes sense to have
$p(\Theta_A, \Theta_B) = p(\Theta_A) \times p(\Theta_B)$.

From the problem we understand that

A and B are independent , and that prior knowledge about the population A will have nothing to influence to that of B. Hence it makes sense for us to have

$p(\Theta_A, \Theta_B) = p(\Theta_A) \times p(\Theta_B)$.

3.4)

Mixtures of beta priors: Estimate the probability of teen recidivism based on a study in which there were n = 43 individuals released from incarceration and y = 15 re-offenders within 36 months.

a) Using a beta(2,8) prior for Θ, plot p(Θ), p(y| Θ) and p(Θ|y) as functions of Θ. Find the posterior mean, mode, and standard deviation of Θ. Find a 95% quantile-based confidence interval.

Given

a=2, b=8, y=15 and n=43

Objective

To plot p(Θ), p(y| Θ) and p(Θ|y) as functions of Θ

The beta distribution

p(Θ) = dbeta(Θ; a; b) = dbeta(Θ; 2; 8)

The binomial distribution

A random variable Y has a binomial(n, Θ) distribution if

Pr(Y = y|Θ) = dbinom(y; n; Θ)) p(y|Θ) = dbinom(15; 43; Θ)

The posterior distribution

p(Θ|y) = dbeta(Θ; a + y; b + ny) = dbeta(Θ; 17; 36)

b)

Now for beta(8,2) prior for Θ

So by changing the values to a=8, b=2, we get

p(Θ) = dbeta(Θ, a, b) = dbeta(Θ; 8; 2)

p(y|Θ) = dbinom(n, y, Θ) = dbinom(15, 43, Θ)

p(Θ|y) = dbeta(Θ, a + y, b + ny) = dbeta(Θ, 17, 36)

c)

We consider the following prior distribution

$$P(\Theta) = \frac{1}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \; [\, 3\Theta \, (1-\Theta)^7 \; + \Theta^7 \; (1-\Theta)] \qquad \text{-------}(1)$$

Considering that (1) is a 75-25 % mixture of a beta(2,8) and beta (8,2) prior distribution

We can write (1) as

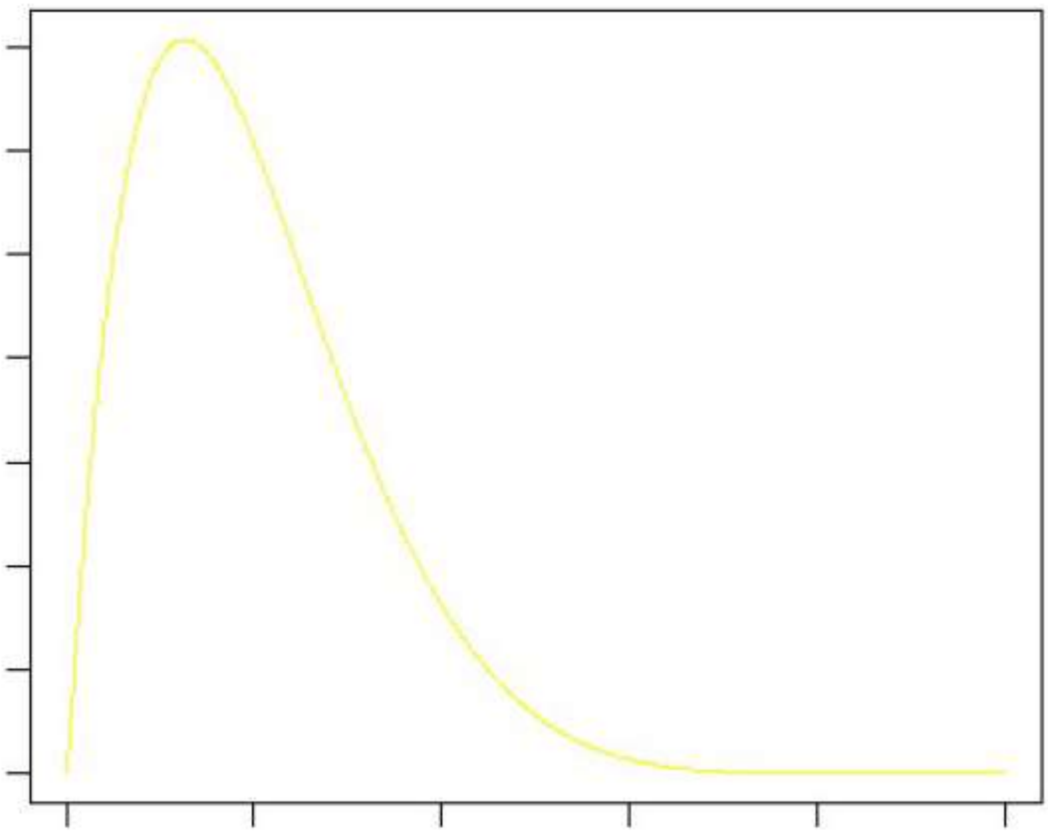
$$P(\Theta) = \frac{3}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \; [\, 3\Theta \, (1-\Theta)^7 \; + \Theta^7 \; (1-\Theta)]$$

$$= \frac{3}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \; [\, 3\Theta \, (1-\Theta)^7 \,] \; + \frac{1}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \; [\, \Theta^7 \; (1-\Theta)]$$

= 75 % of beta (2,8) prior + 25 % of beta(8,2) prior

CodeBlock

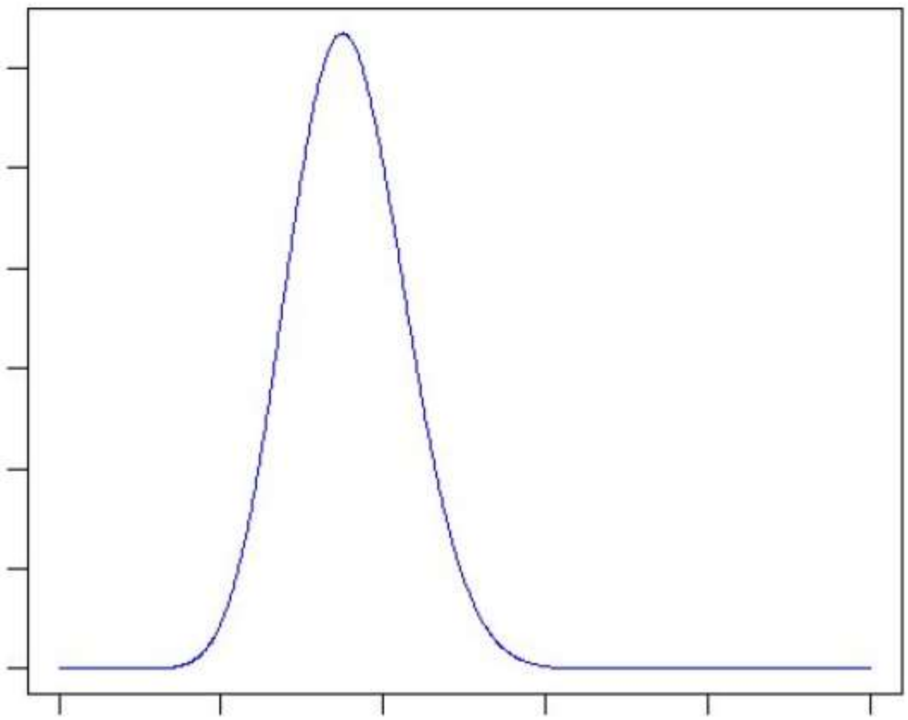
```
par("mar")  #Checking margins

## [1] 5.1 4.1 4.1 2.1

par(mar=c(1,1,1,1))  #change margins

a<-2
b<-8
y<-15
n<-43
theta<-seq(0,1,.001)
pr <- dbeta(theta, a, b)
plot(pr,type='l', col="yellow")
```



x- label – Theta

y – label - Density

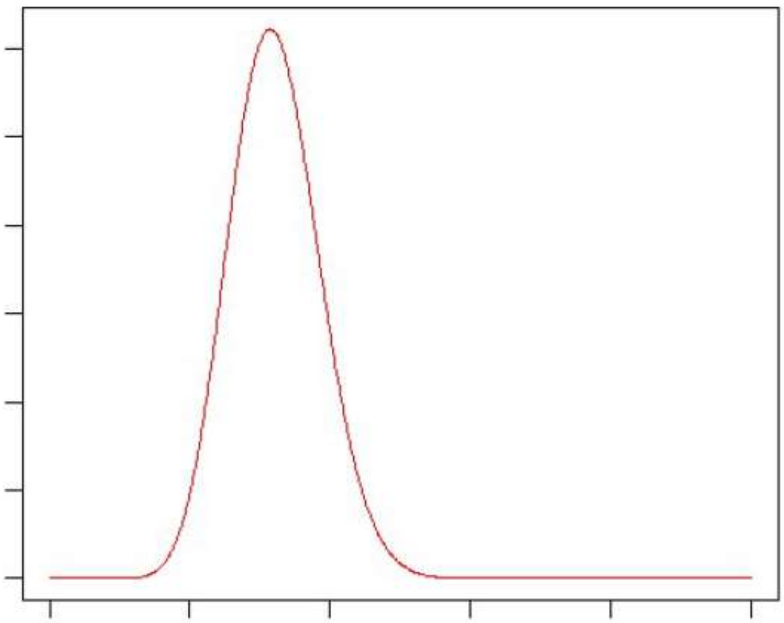
```r
lh <- dbinom(y,n,theta)
plot(lh,type='l', col="blue")
```



x- label – Theta

y – label - Density

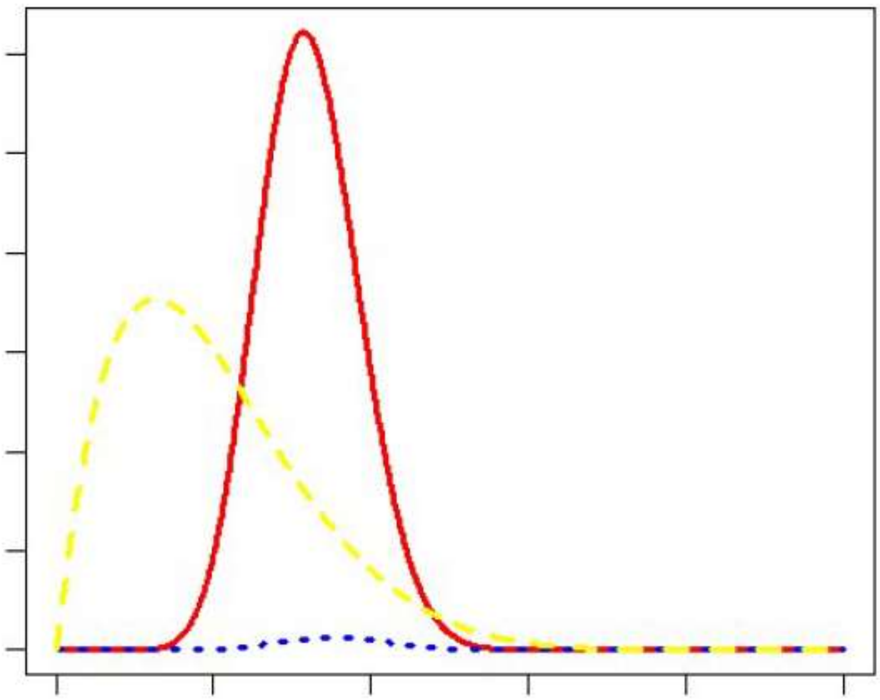
```r
pos <- dbeta(theta, a+y, b+n-y)
plot(pos,type='l', col="red")
```



x- label – Theta

y – label - Density

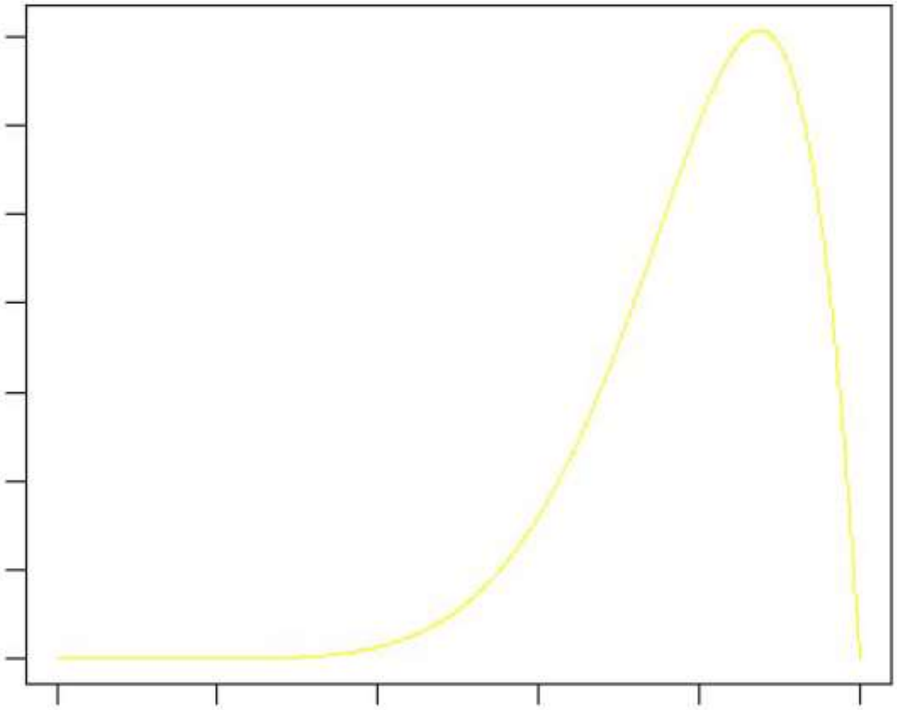
```
#Plotting against
plot(pos, col = "red",type='l', ylab = "Density", xlab = "theta",lty = 1, lwd
= 3)
lines(lh, type='l', col="blue", lty = 3, lwd = 3)
lines(pr, col = "yellow", lty = 2, lwd = 3)
```



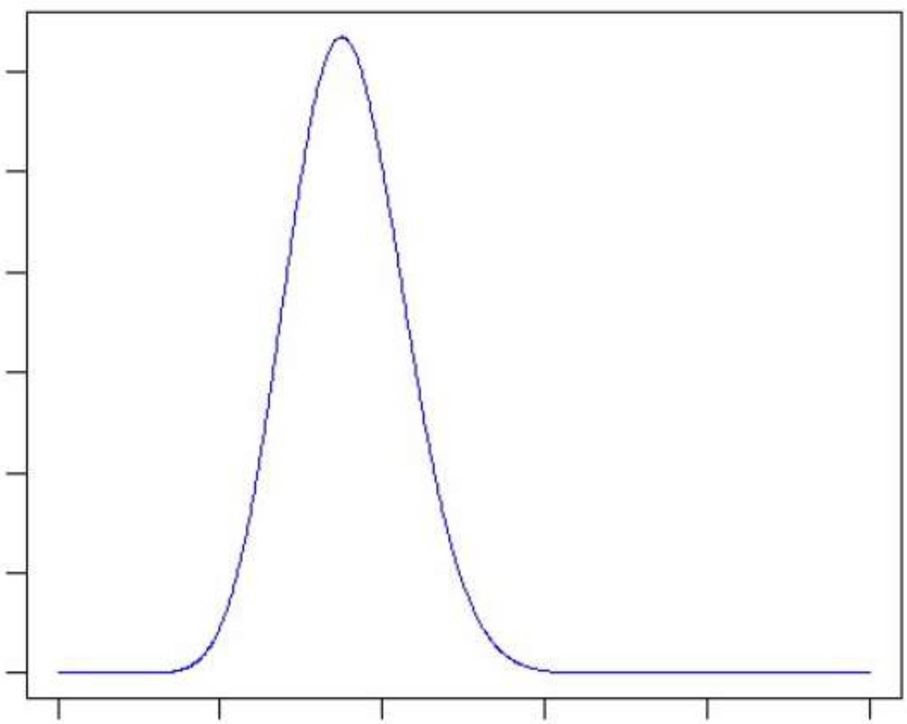
x- label – Theta

y – label - Density

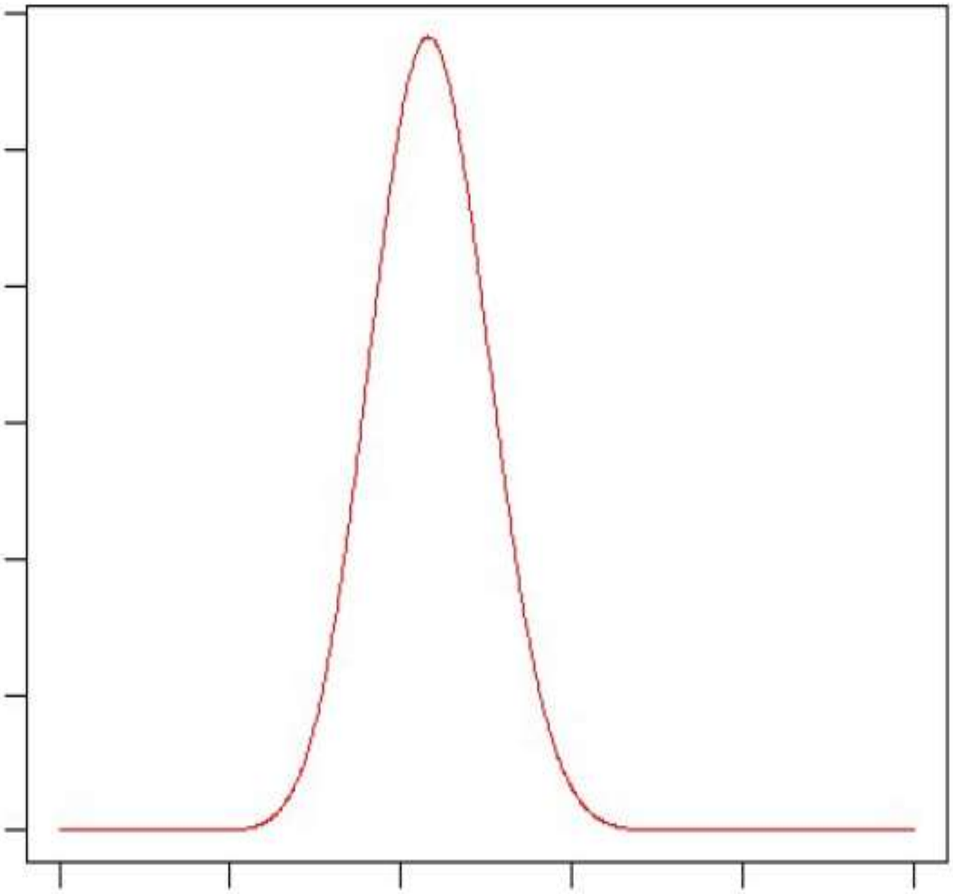
```
a<-8
b<-2
y<-15
n<-43
theta<-seq(0,1,.001)
pr <- dbeta(theta, a, b)
lh <- dbinom(y,n,theta)
pos <- dbeta(theta, a+y, b+n-y)

plot(pr,type='l', col="yellow")
```



```
plot(lh,type='l', col="blue")
```



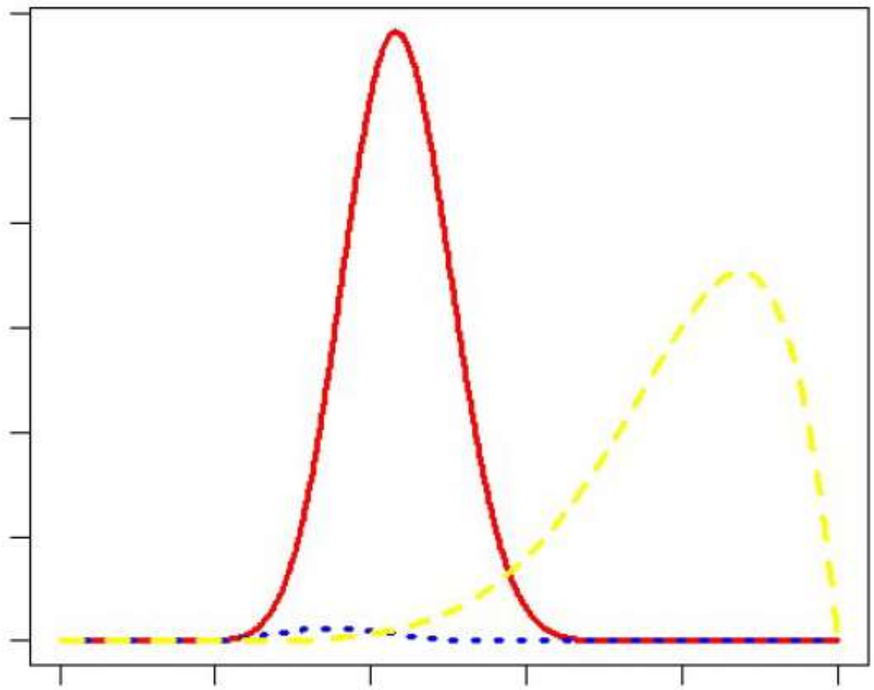
```
plot(pos,type='l', col="red")
```



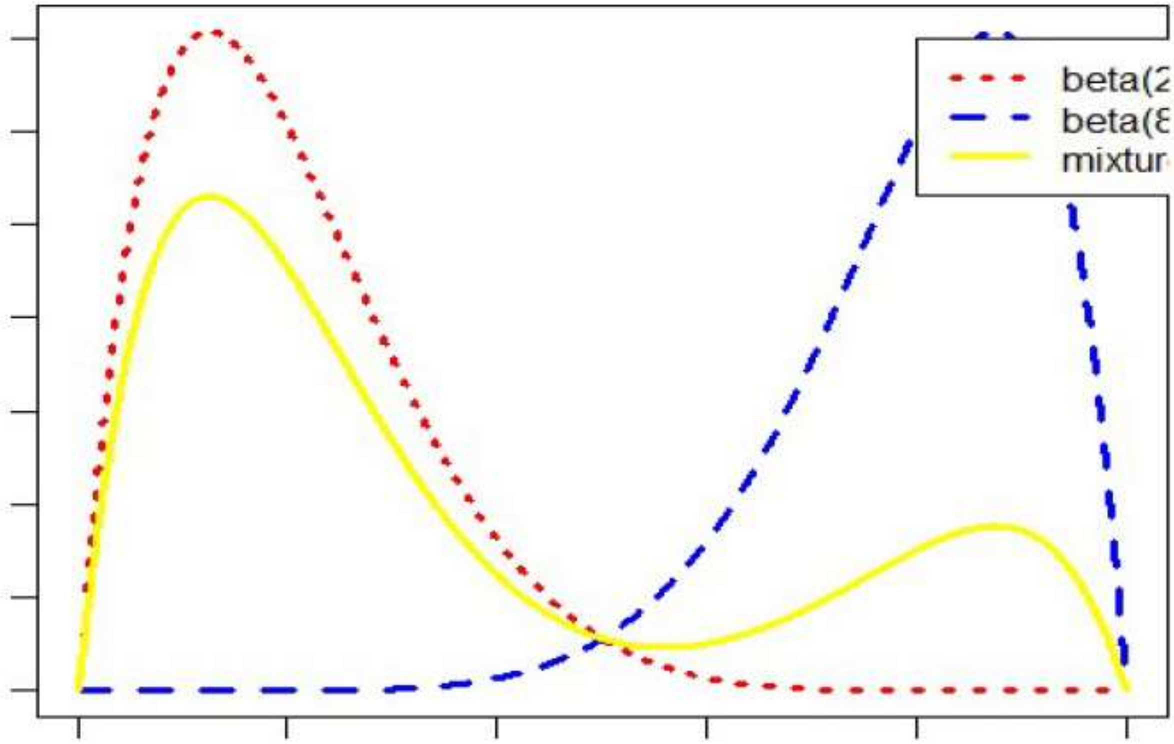
x- label – Theta

y – label - Density

```
plot(pos, col = "red",type='l', ylab = "Density", xlab = "theta",lty = 1, lwd
= 3)
lines(lh, type='l', col="blue", lty = 3, lwd = 3)
lines(pr, col = "yellow", lty = 2, lwd = 3)
legend(x=1500,y=5.0, c("Prior", "Likelihood", "Posterior"), lty = c(2, 3, 1),
       lwd = c(3, 3, 3), col = c("yellow", "blue", "red"))
```



```
a<-2
b<-8
y<-15
n<-43
theta<-seq(0,1,.001)
pr1 <- dbeta(theta, a, b)
a<-8
b<-2
y<-15
n<-43
theta<-seq(0,1,.001)
pr2 <- dbeta(theta, a, b)

plot(pr1, type='l', col="red", lty = 3, lwd = 3, ylab = "Density", xlab = "th
eta")
lines(pr2, col = "blue", lty = 2, lwd = 3)
lines((((3/4)*pr1)+((1/4)*pr2)),type='l', col="yellow", lty = 1, lwd = 3)
legend(x=800,y=3.5, c("beta(2,8) prior", "beta(8,2) prior", "mixture prior"),
lty = c(3, 2, 1),
       lwd = c(3, 3, 3), col = c(  "red","blue","yellow"))
```



75% Beta (2,8) 25% Beta (8,2)

x- label – Theta

y – label - Density

From the above plot we can see that there is a clear inclination in the mixture curve towards 75% Beta(2,8). This represents that the density is higher on the side with the greater concentration of theta.

d)

From equation (1) we know that

$P(\Theta) = \frac{1}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}$ [$3\Theta\ (1-\Theta)^7 + \Theta^7\ (1-\Theta)$] -------------(3)

$P(y|\Theta) = \Theta^y (1-\Theta)^{n-y}$

$P(y|\Theta) = \Theta^7\ (1-\Theta)^{28}$ ----------------------------(2)

From (3) and (2)

$P(\Theta) \times P(y|\Theta) = \frac{1}{4} \quad \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} [3\ \Theta^{16} (1-\Theta)^{35} + \Theta^{22} (1-\Theta)^{29}]$

ii)

We know

Y € binomial (n, Θ)

Hence the binomial distribution -> $P(y|\Theta) = \Theta^y (1-\Theta)^{n-y}$

And also the beta prior P(Θ) = d beta(Θ,a,b) . These identify the distributions

iii)

Plotting p(Θ) x p(y| Θ) for a variety of values

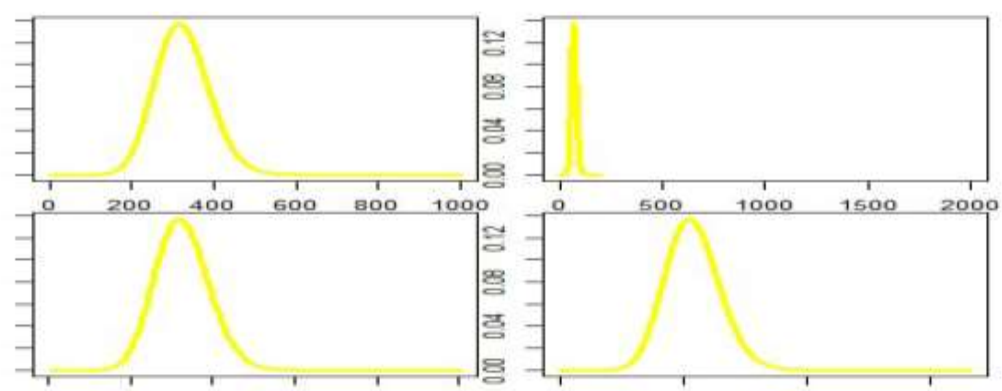
Code Block

```
attach(mtcars)
par(mfrow=c(2,2))
```

```
a<-2
b<-8
y<-15
n<-43
theta<-seq(0,1,.001)
pr1 <- dbeta(theta, a, b)

a<-8
b<-2
y<-15
n<-43
theta<-seq(0,1,.001)
pr2 <- dbeta(theta, a, b)

lh <- dbinom(y,n,theta)
plot( ( (lh)*(((3/4)*pr1)+((1/4)*pr2)) ),type='l', col="yellow", lty = 1, lwd
= 3,ylab = "Density", xlab = "theta")



a<-2
b<-8
y<-15
n<-43
theta<-seq(0,10,.005)
pr1 <- dbeta(theta, a, b)
a<-8
b<-2
y<-15
n<-43
theta<-seq(0,10,.005)
pr2 <- dbeta(theta, a, b)
lh <- dbinom(y,n,theta)

## Warning in dbinom(y, n, theta): NaNs produced

plot( ( (lh)*(((3/4)*pr1)+((1/4)*pr2)) ),type='l', col="yellow", lty = 1, lwd
= 3,ylab = "Density", xlab = "theta")
```

```
a<-2
b<-8
y<-15
n<-43
theta<-seq(0,1,.004)
pr1 <- dbeta(theta, a, b)
a<-8
b<-2
y<-15
```

```
n<-43
theta<-seq(0,1,.004)
pr2 <- dbeta(theta, a, b)
lh <- dbinom(y,n,theta)
plot( ( (lh)*(((3/4)*pr1)+((1/4)*pr2)) ),type='l', col="yellow", lty = 1, lwd = 3,ylab = "Density", xlab = "theta")


a<-2
b<-8
y<-15
n<-43
theta<-seq(0,1,.00006)
pr1 <- dbeta(theta, a, b)
a<-8
b<-2
y<-15
n<-43
theta<-seq(0,1,.00006)
pr2 <- dbeta(theta, a, b)
lh <- dbinom(y,n,theta)
plot( ( (lh)*(((3/4)*pr1)+((1/4)*pr2)) ),type='l', col="yellow", lty = 1, lwd = 3,ylab = "Density", xlab = "theta")
```



X label – Theta

Y label- Density

e)
The weights of the mixture distribution from ii) is given by their sums
Interpreting their values shows that the value for density is higher for greater concentration of theta.

Problem 3

Hoff 3.7 3.7 Posterior prediction: Consider a pilot study in which $n_1 = 15$ children enrolled in special education classes were randomly selected and tested for a certain type of learning disability. In the pilot study, $y_1 = 2$ children tested positive for the disability.

a) Using a uniform prior distribution, find the posterior distribution of θ, the fraction of students in special education classes who have the disability.
Find the posterior mean, mode and standard deviation of θ, and plot the posterior density.
[*Solution* : Given: $n_1 = 15$ & $y_1 = 2$
and we know that for $\theta \sim beta(1,1)(uniform)$,

$$(\theta|Y = y) \sim beta(1 + y,\ 1 + n - y)$$

to get the likelihood, assuming this is an i.i.d

$Pr(Y = y|\theta) = \theta^y\ (1-\theta)^{n-y} = dbinom(y, n, \theta) = dbinom(2, 15, \theta)$ *which is a* **binomial distribution that gives the**

From the prior inference condition of the uniform distribution, we know that

$$p(\theta|Y) = beta(1 + y,\ 1 + n - y) = beta(3, 14)$$

The above equations demonstrates that for any i.i.d prior, the $\sum y_i$ is the sufficient data to predict the posterior distribution.

Posterior Mean: It is given by the formula

$$E[\theta|y] = \frac{a+y}{a+b+n} = \frac{3}{17} = 0.1764$$

Posterior Mode:

$$mode[\theta|y] = \frac{a+y-1}{a+b+n-2} = \frac{2}{15} = 0.1333$$

Posterior Standard Deviation:

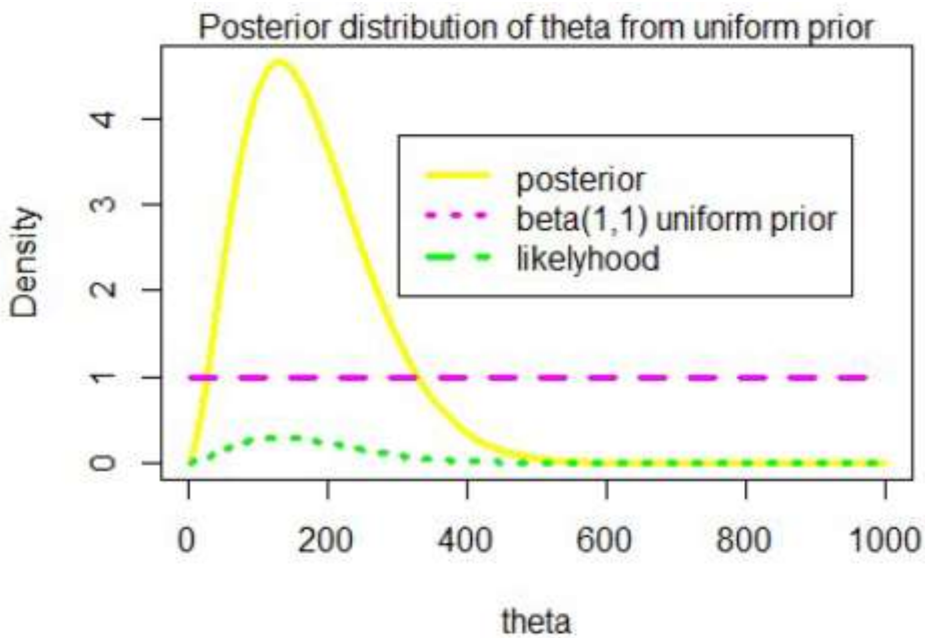
$$Var[\theta|y] = \frac{E[\theta|y]\ E[1-\ \theta|y]}{a+b+n+1} =$$

[0.176 * (1 – (3/17))] / 18 = 0.0080568

Sd = 0.0049

Codeblock

```
#3.7 a)
a=3
b=14
theta = seq(0.005, 0.995, length = 1000)
posterior = dbeta(theta, a, b)
prior = dbeta(theta,1,1)
likelihood = dbinom(2,15,theta)
plot(posterior, col = "Yellow",typ='l', ylab = "Density", xlab = "theta",lty
= 1, lwd = 3)
lines(prior, col = "magenta", lty = 2, lwd = 3)
lines(likelihood, col = "green", lty = 3, lwd = 3)
legend(x=300,y=3.8, c("posterior","beta(1,1) uniform prior","likelyhood"), lt
y = c(1, 3, 2),
       lwd = c(3, 3, 3), col = c(  "Yellow","Magenta","green"))
mtext("Posterior distribution of theta from uniform prior")
```



Researchers would like to recruit students with the disability to partici- pate in a long-term study, but first they need to make sure they can recruit enough students. Let n2 = 278 be the number of children in special edu- cation classes in this particular school district, and let Y2 be the number of students with the disability.

b) Find $Pr(Y2 = y2|Y1 = 2)$, the posterior predictive distribution of $Y2$, as follows:
i. Discuss what assumptions are needed about the joint distribution of $(Y1, Y2)$ such that the following is true:

[*Solution* :

$$Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 Pr(Y_2 = y_2|\theta)\ p(\theta|Y_1 = 2)\ \ d\theta$$

Using class room Notations:

For this to be true, in general when Y = y1……yn are considered in the denominator of p(y~ / Y), then Y = y1…yn should follow iid. And y1 .. yn are conditionally independent.

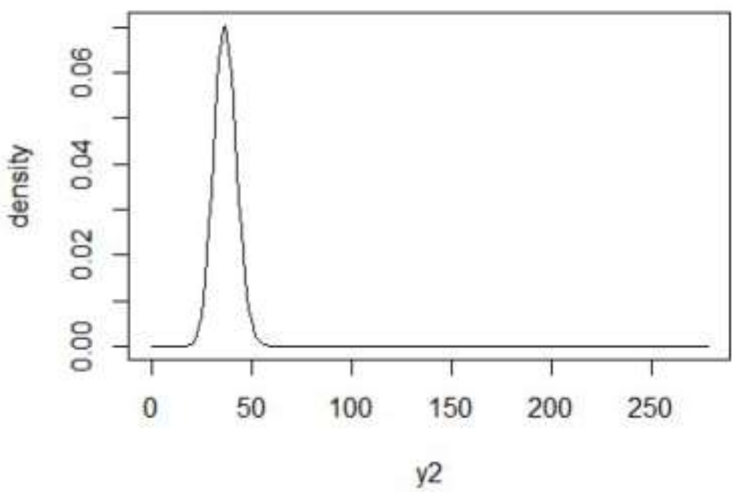
But here we are asked about, y2=y~ and Y=y_i . SO this makes y1and y2 are dependent. Because, y2 dependent is dependent on Theta and Theta is dependent on y1.

Based on what text book says below: y1 and y2 are dependent.

1. The predictive distribution does not depend on any unknown quantities. If it did, we would not be able to use it to make predictions.
2. The predictive distribution depends on our observed data. In this distribution, ˜ Y is not independent of $Y_1, \ldots, Y_n$ (recall Section 2.7). This is because observing $Y_1, \ldots, Y_n$ gives information about _, which in turn gives information about ˜ Y . It would be bad if ˜ Y were independent of $Y_1, \ldots, Y_n$ - it would mean that we could never infer anything about the un-sampled population from the sample cases.

d)

```r
mean<-278 * 2 /15
sd=sqrt(mean * (1 - (2/15) ))
y2=seq(0,278)
plot(y2,dbinom(x = y2, size = 278, prob = (2/15)), ylab = "density", type="l"
)
```




```r
print(paste0("mean  = ",mean,"  sd  = ",sd ))
## [1] "mean = 37.0666666666667  sd  = 5.66784301515528"
```

inference: I will use the posterior calculated in c for this. At lower sample size, the MLE estimates are not good in predictions. With increase in sample size