602_hw2

William Tirone

```r
library(VGAM)
```

```
Loading required package: stats4
```

```
Loading required package: splines
```

3.1

a)

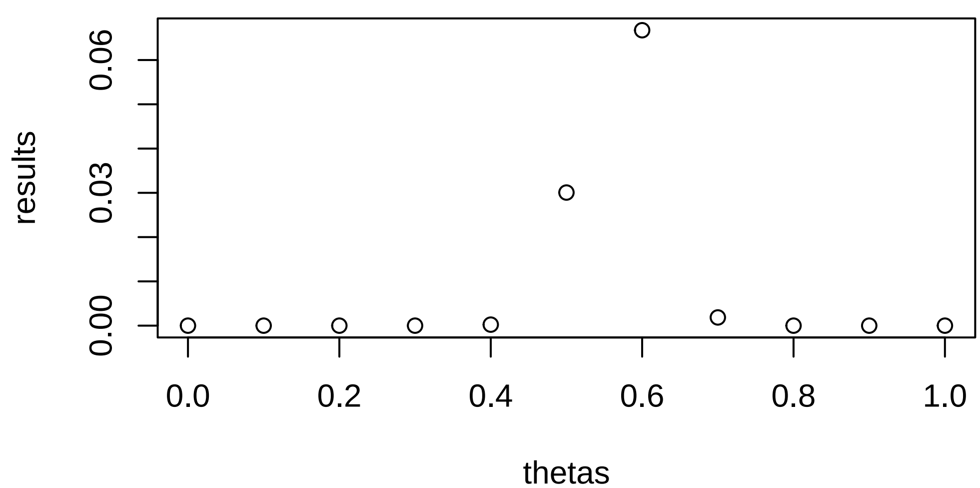
$$
\begin{aligned}
&P(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta) = \text{by independence} \\
&= \prod_{i=1}^{n} P(Y_i|\theta) \\
&= \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} \\
&= \theta^{\Sigma_{i=1}^{n} y_i}(1-\theta)^{100-\Sigma_{i=1}^{n} y_i}; y = 0, 1
\end{aligned}
$$

Finding the distribution of $P(\sum_{i=1}^{n} Y_i = y|\theta)$

$$
\begin{aligned}
&M_{\sum Y_i = y|\theta}(t) \\
&= \text{by independence} \\
&= \prod_{i=i}^{n} M_{Y_i|\theta}(t) = \\
&= \prod_{i=i}^{n} (1 - p + pe^t) \\
&= (1 - p + pe^t)^n \\
&= \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
&= \binom{100}{57} \theta^{57} (1-\theta)^{43}; \theta \in [0,1] \text{assuming a uniform prior}
\end{aligned}
$$

b)

```r
thetas = seq(0.0,1.0,by=0.1)
results = dbinom(57,100,thetas)

plot(thetas,results)
```

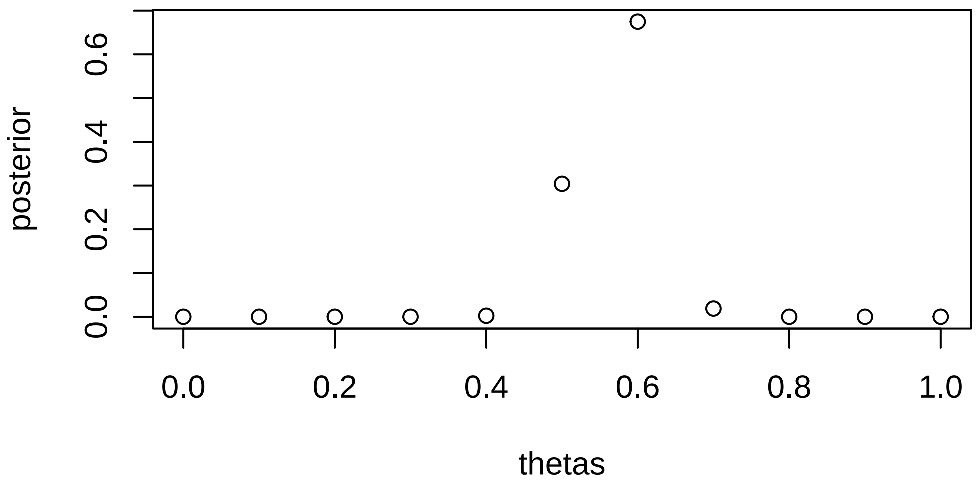


c)

$$p(\theta|\Sigma_{i=1}^{n} y_i = 57) = \frac{P(\Sigma_{i=1}^{n} y_i = 57|\theta)P(\theta)}{P(\Sigma_{i=1}^{n} y_i = 57)} \text{each } P(\Theta = \theta) = \frac{1}{11}$$

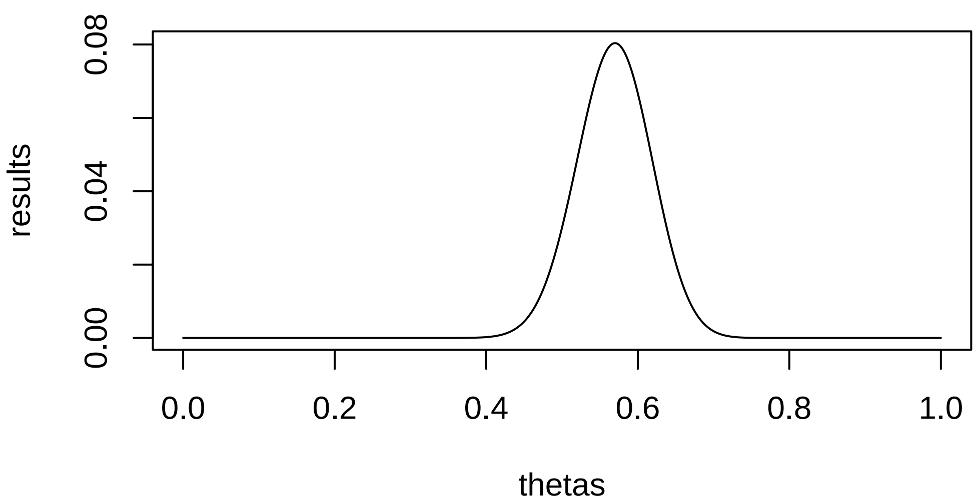
The posterior distribution and marginal distribution of Y are just scaling constants since the denominator does not depend on theta and we have equal belief for each of $P(\theta)$.

```
marginal_y = sum((1/11) * dbinom(57,100,thetas))
posterior = (results * (1/11))/marginal_y
plot(thetas,posterior)
```



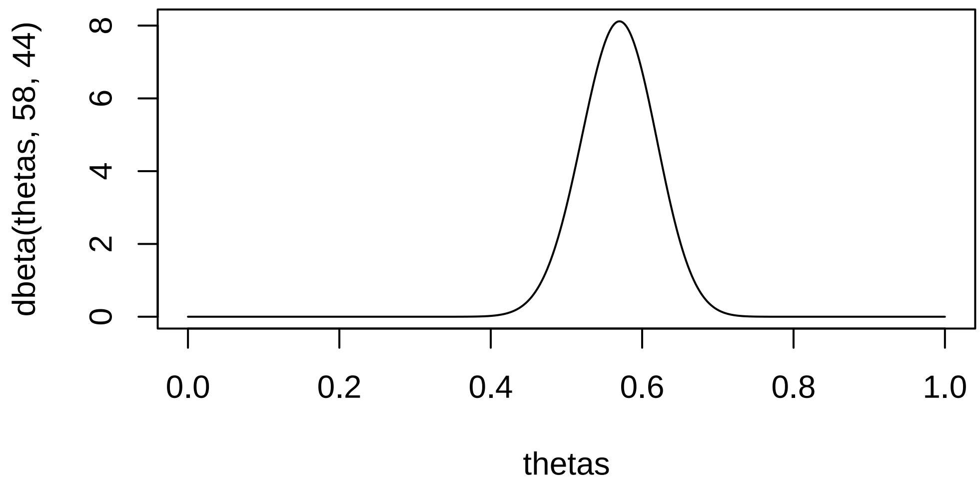
d)

```
thetas = seq(0,1,by=0.001) #U(0,1)
results = dbinom(57,100,thetas)
plot(thetas,results,type="l")
```



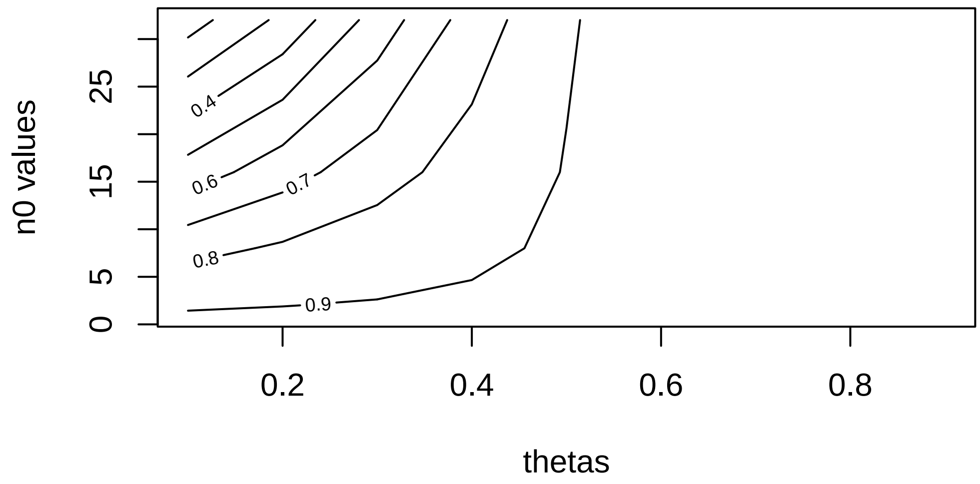
e)

```r
plot(thetas,dbeta(thetas,58,44),type='l')
```



3.2

```r
theta0 = seq(0.1,0.9,by=0.1)
n0 = c(1,2,8,16,32)

data=c()

for (i in theta0) {
for (j in n0) {

    a = i * j
    b = (1-i)*j

    p = pbeta(.5,a+57,b+43,lower.tail=FALSE) #posterior (theta > .5 | sum = 57)
    data = append(data,p)

}
}

probability_data = matrix(data,nrow=9,ncol=5,byrow=TRUE)
contour(theta0,n0,probability_data, xlab="thetas",ylab='n0 values')
```



3.4

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

a)

calculations for posterior with prior beta(2,8) and beta(8,2) are here, with plots for part a) and part b) following this chunk.

```
beta_mean = function(a,b){
  print("mean:")
  a / (a+b)
}

beta_mode = function(a,b){
  print('mode:')
  (a-1) / (a+b-2)
}

beta_sd = function(a,b){
  print('standard deviation:')
  var = (a*b) / ((a+b)^2 * (a+b+1))
  sd = sqrt(var)
  return(sd)
}
```

```r
CI_28 = c(qbeta(.025,17,36),qbeta(.975,17,36))
CI_82 = c(qbeta(.025,23,30),qbeta(.975,23,30))

#data for the posterior w/ 2,8 prior and posterior a = 17, posterior b = 36
print("using alpha = 17 and beta = 36 with beta(2,8) prior")
```

```
[1] "using alpha = 17 and beta = 36 with beta(2,8) prior"
```

```r
beta_mean(17,36)
```

```
[1] "mean:"
```

```
[1] 0.3207547
```

```r
beta_mode(17,36)
```

```
[1] "mode:"
```

```
[1] 0.3137255
```

```r
beta_sd(17,36)
```

```
[1] "standard deviation:"
```

```
[1] 0.0635189
```

```r
print(c("95% CI",CI_28))
```

```
[1] "95% CI"             "0.203297787819103" "0.451023982216632"
```

```r
#with 8,2 prior
print("using alpha = 23, beta = 30 with beta(8,2) prior")
```

```
[1] "using alpha = 23, beta = 30 with beta(8,2) prior"
```

```r
beta_mean(23,30)
```

```
[1] "mean:"
```

```
[1] 0.4339623
```

```r
beta_mode(23,30)
```

```
[1] "mode:"
```

```
[1] 0.4313725
```

```r
beta_sd(23,30)
```

```
[1] "standard deviation:"
```

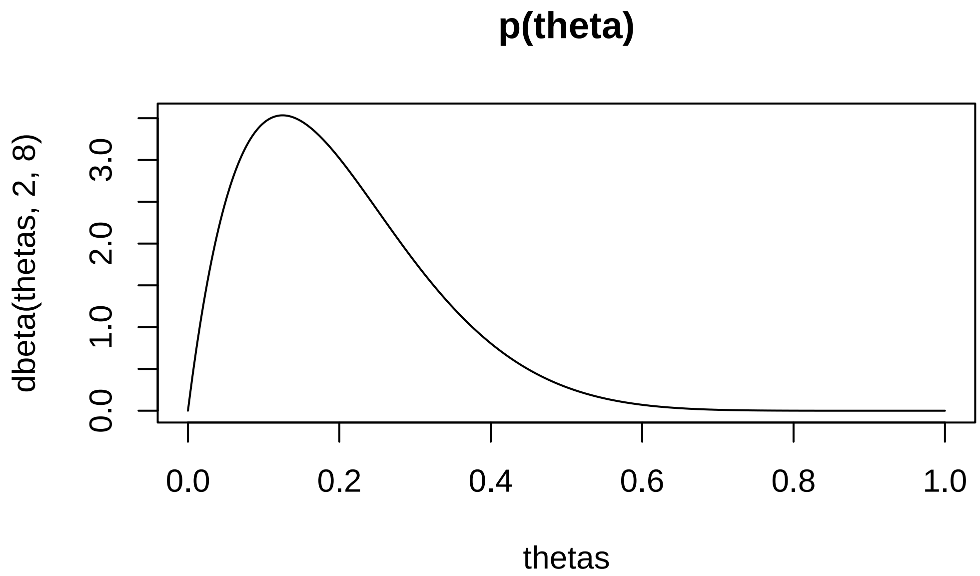
```
[1] 0.06744532
```

```r
print(c("95% CI",CI_82))
```

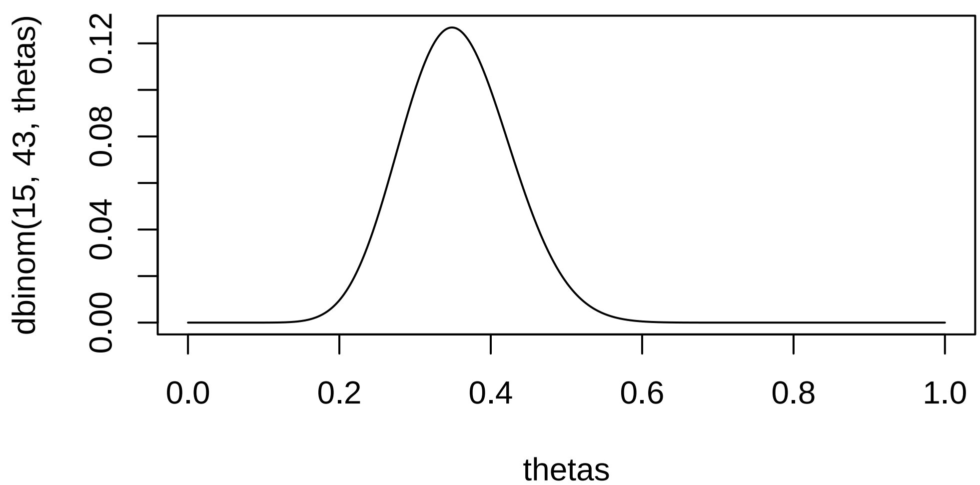
```
[1] "95% CI"            "0.304695624711747" "0.567952795996458"
```

plots for part a)

```r
#plotting prior p(\theta)
thetas = seq(0,1,by=0.001) #U(0,1)

plot(thetas, dbeta(thetas, 2,8), type='l',main="p(theta)")
```




```
#plotting p(y=15|\theta)
#plot a binomial here
plot(thetas, dbinom(15,43,thetas),type='l')
```



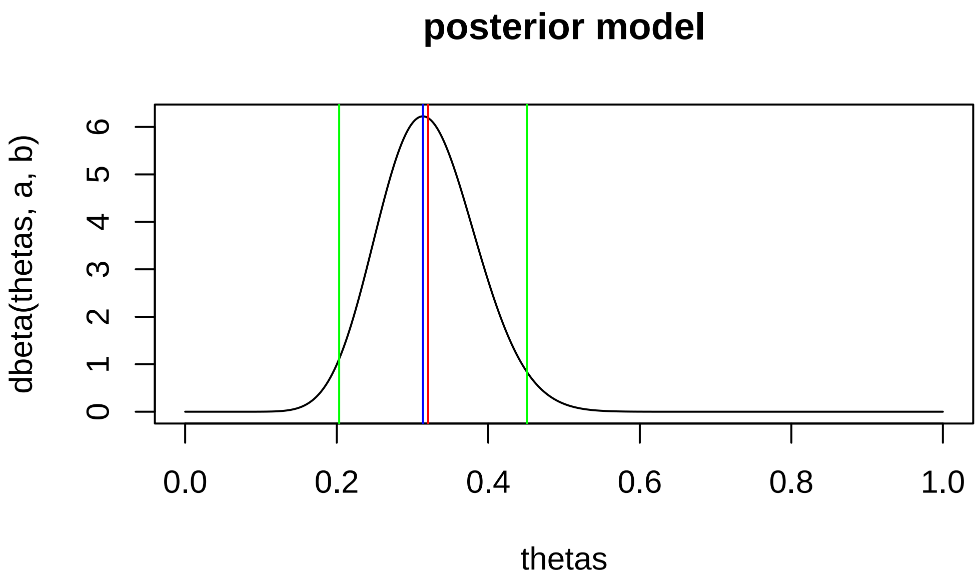

```
#posterior which is beta(2 + success, 8 + failure) = beta()
a=2+15
b=8+28
plot(thetas, dbeta(thetas,a,b),type='l',main="posterior model")
abline(v=beta_mean(a,b), col='red') #mean
```

```
[1] "mean:"
```

```
abline(v=beta_mode(a,b), col='blue') #mode
```

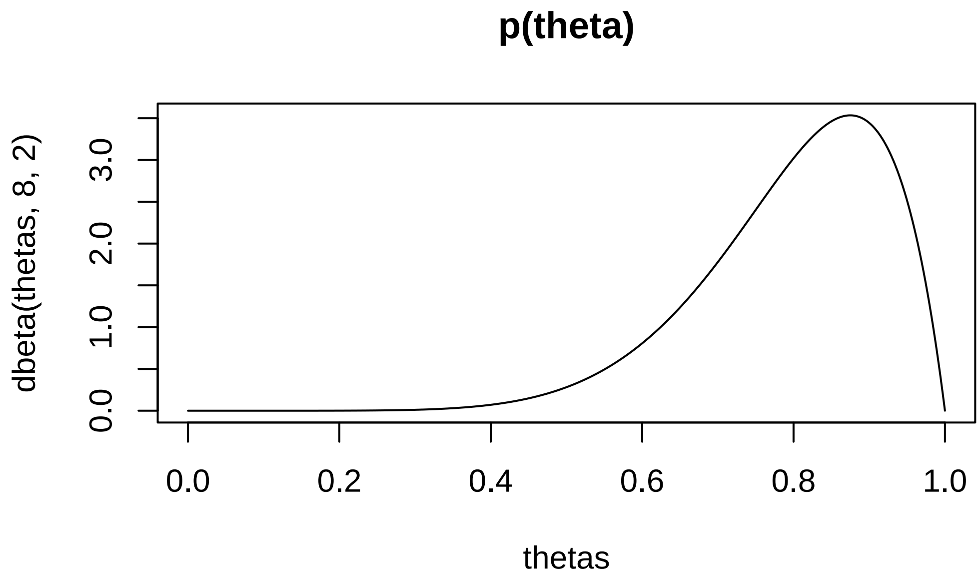
```
[1] "mode:"
```

```
# CI
abline(v=qbeta(.975,a,b),col='green') #lower bound
abline(v=qbeta(.025,a,b),col='green') #upper bound
```

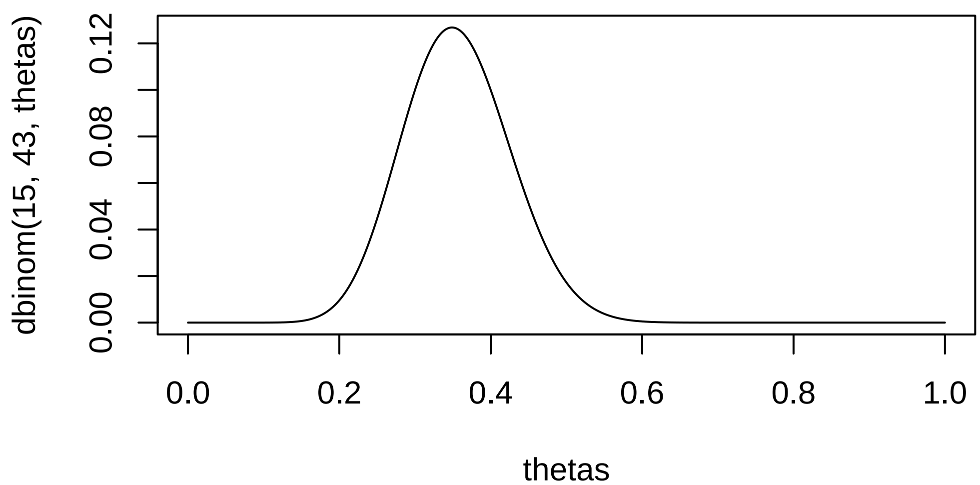


plots for part b)

```
#plotting prior p(\theta)
thetas = seq(0,1,by=0.001) #U(0,1)

plot(thetas, dbeta(thetas, 8,2), type='l',main="p(theta)")
```




```
#plotting p(y=15|\theta)
#plot a binomial here
plot(thetas, dbinom(15,43,thetas),type='l')
```



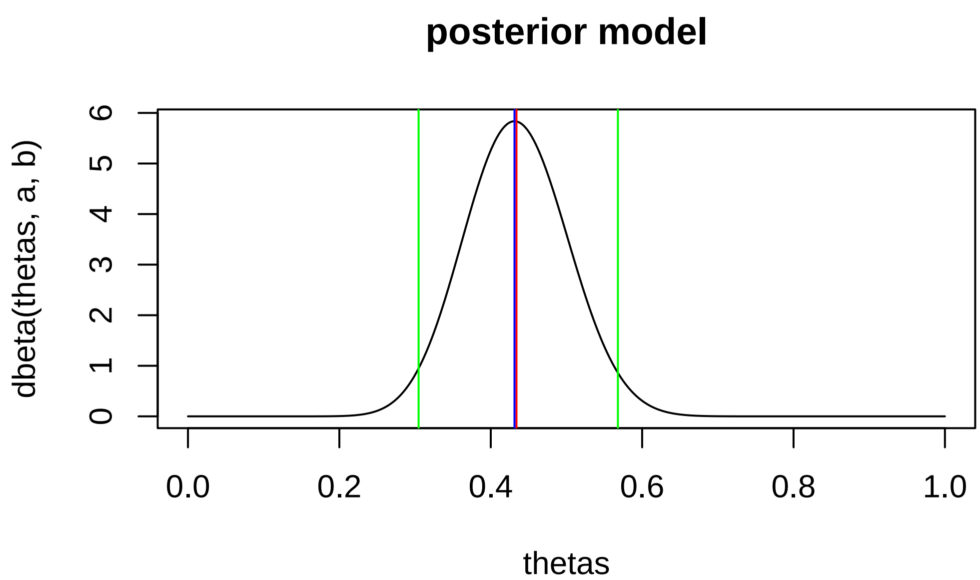

```
#posterior which is beta(2 + success, 8 + failure) = beta()
a=8+15
b=2+28
plot(thetas, dbeta(thetas,a,b),type='l',main="posterior model")
abline(v=beta_mean(a,b), col='red') #mean
```

```
[1] "mean:"
```

```
abline(v=beta_mode(a,b), col='blue') #mode
```

```
[1] "mode:"
```

```
# CI
abline(v=qbeta(.975,a,b),col='green') #lower bound
abline(v=qbeta(.025,a,b),col='green') #upper bound
```

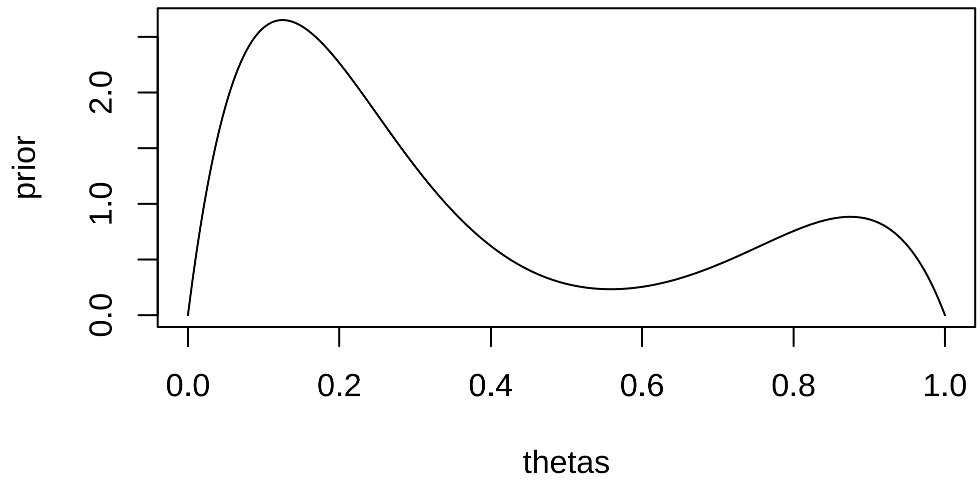


c)

This may represent that you have about 25% confidence that there are going to be 8 cases of recidivism and 2 cases of not, while the beta(2,8) represents you're 75% confident that there will be 2 cases of recidivism and 8 cases of failure respectively. This is if you've only seen 10 prior cases.

Or maybe there were two previous studies with 2 successes and 8 failures or 2 failures and 8 successes respectively.

```
prior = 0.75 * dbeta(thetas,2,8) + 0.25 * dbeta(thetas,8,2)
plot(thetas,prior,type="l")
```



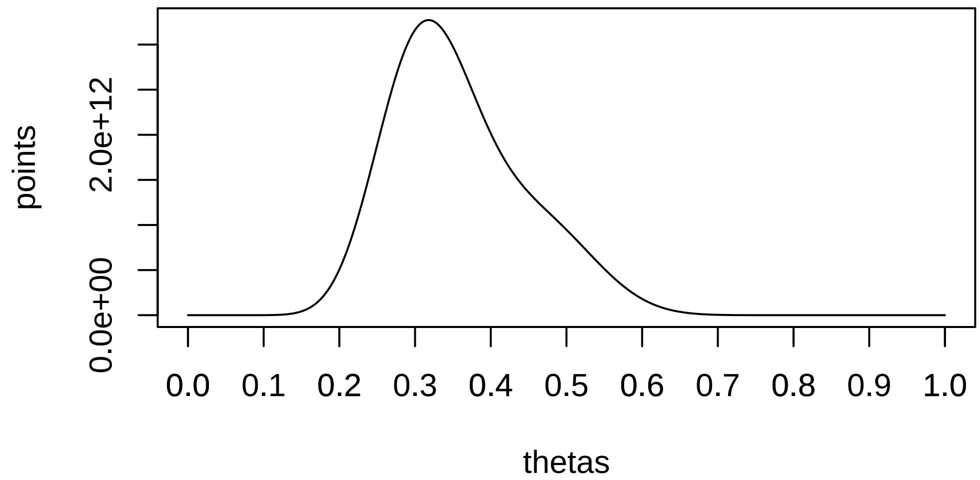
d) i)

$$\begin{aligned}&p(\theta) * p(y|\theta)\\&= \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\binom{43}{15}[3\theta^{16}(1-\theta^{35}) + \theta^{22}(1-\theta)^{25}]\end{aligned}$$

ii) This is a mixture of $beta(17, 36)$ and $beta(23, 26)$

iii) Plot of $p(\theta|y)$ is below. It looks like the mode is about 0.32 (approximately). Since this is more heavily weighted towards the prior of beta(2,8), it makes sense that this mode is closer to the mode of the previous example we saw with the same prior, though pulled slightly to the right by the beta(8.2) prior.

```
coefficient = .25 * 18 * choose(43,15)
thetas = seq(0,1,by=0.001)
points = coefficient* (0.75 * dbeta(thetas,17,36) + 0.25 * dbeta(thetas,23,26))
plot(thetas,points,type="l")
axis(1, at = seq(0.0,1,by=0.1))
```



e)

For a given distribution f(x):

$$f(x) = \sum_{i=1}^{n} w_i p_i(x) \text{where} \sum_{i=1}^{n} w_i = 1$$

This just means that each weight gets multiplied by its respective distribution, and they must sum up to 1 to ensure that the mixture is also a pdf.

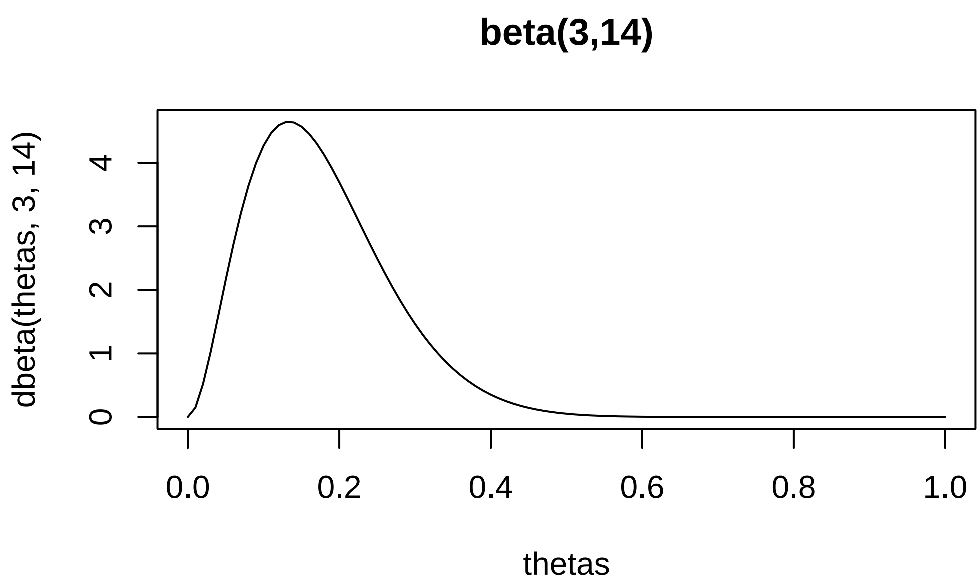
3.7

a)

$$
\begin{aligned}
p(\theta) &= U(0,1) \\
p(\theta|X) &\propto p(X|\theta)p(\theta) \\
&= c(x)\theta^2(1-\theta)^{13} \\
&\text{where c(x) is a proportionality constant} \\
p(\theta|Y_1 = 2) &\sim beta(3,14)
\end{aligned}
$$

$$
\begin{aligned}
mean &= \frac{3}{17} = .176 \\
mode &= \frac{2}{15} = .133 \\
sd &= \sqrt{\frac{42}{(17)^2(18)}} = 0.0899
\end{aligned}
$$

```
thetas = seq(0,1,0.01)
plot(thetas,dbeta(thetas,3,14),type='l',main="beta(3,14)")
```



b)

i) The key assumption is that $Y_2 = y_2|\theta$ is independent of $Y_1 = 2$. (I believe this also assumes exchangeability?)

ii)

$$\int_0^1 \binom{278}{y_2}\theta^{y_2}(1-\theta)^{278-y_2}\frac{1}{B(3,14)}\theta^2(1-\theta)^{13}d\theta$$

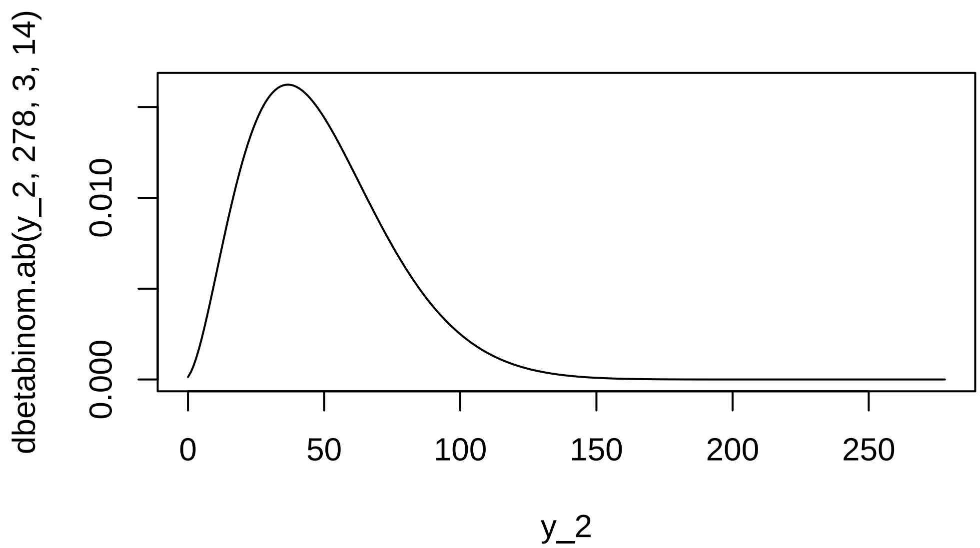
iii) using the kernel trick above:

$$= \frac{\binom{278}{y_2}}{B(3,14)}\frac{\Gamma(y_2+3)\Gamma(292-y_2)}{\Gamma(295)} \sim BetaBinomial(278,3,14)$$

c)

$$mean = \frac{278*3}{17} = 49 sd = \sqrt{\frac{278*3*14*295}{17^2*18}} = 25.732$$

```r
y_2 = seq(0,278,by=1)
plot(y_2,dbetabinom.ab(y_2,278,3,14), type="l")
```

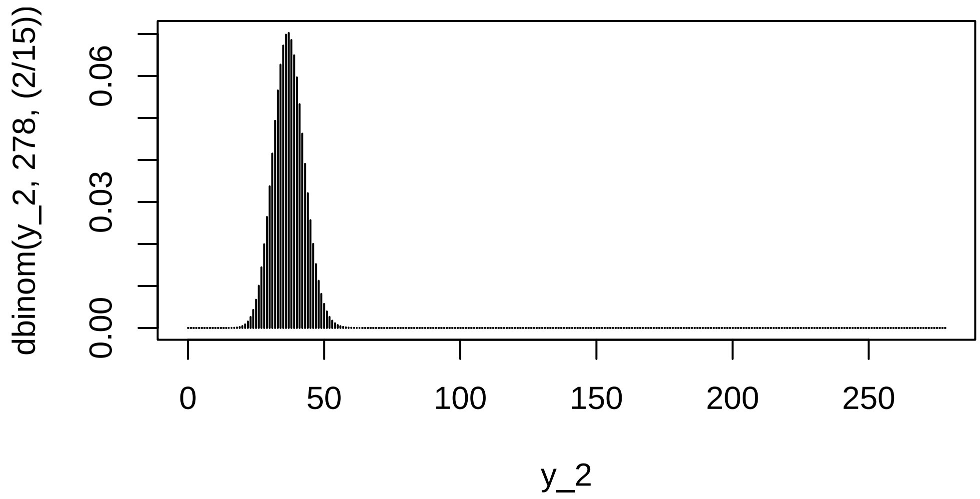


d)

$$P(Y_2 = y_2|\theta = \frac{2}{15}) = \binom{278}{y_2}\frac{2}{15}^{y_2}\frac{13}{15}^{278-y_2} \sim BIN(278,\frac{2}{15}) mean = 37.07 sd = \sqrt{278*\frac{2}{15}*\frac{13}{15}} = 5.668$$

The means are somewhat similar, in the range of 37-50. However, the plot from part c is much heavier tailed and skewed to the right. Considering we're Bayesian, it probably makes sense to incorporate previous data into our model, like we did with the predictive distribution, to make actual predictions. The standard deviation though, in part c, is much higher so that seems to indicate that using our MLE-based model in part d is more accurate.

```
plot(y_2, dbinom(y_2,278,(2/15)),type='h')
```



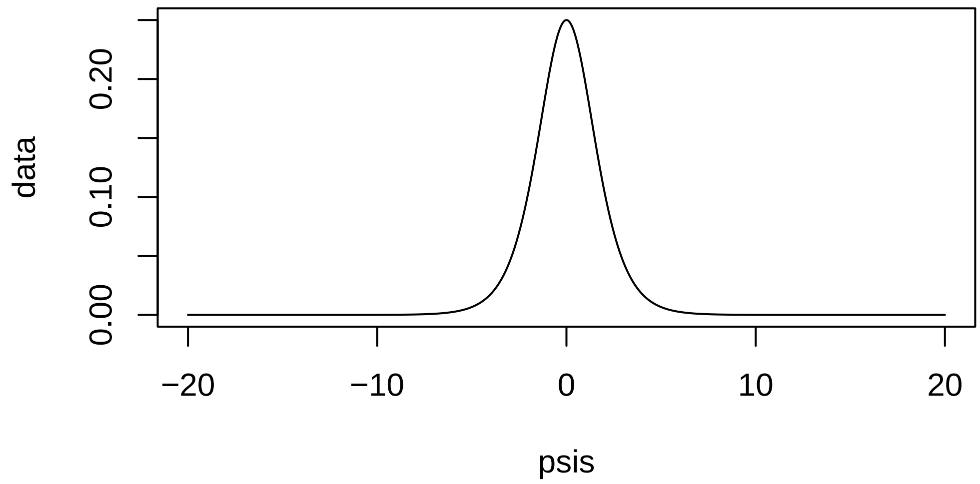
3.10

a)

$$p(\theta) = beta(1,1) = UNIF(0,1)$$
$$p(\psi) = 1\frac{1}{(e^{\psi}+e^{\psi})^2} = \frac{e^{\psi}}{(1+e^{\psi})^2}$$

note: worked out the above inverse and derivative
by hand, but please let me know if that should be typed up next time

```
psis = seq(-20,20,by=.01)
data = exp(1)**psis / (1+exp(1)**psis)**2
plot(psis,data,type="l")
```



b)

$e^{\psi} = \theta$

after simplification and plugging in the above using formula in the book

$p(\psi) = e^{\psi} e^{-e^{\psi}}; -\infty < \psi < \infty$

```
psis = seq(-20,20,by=.01)
data = exp(1)**(psis) * exp(1)**(-exp(1)**psis)
plot(psis,data,type='l')
```