STAT 638 HW 2

Jack Cunningham

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.2      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ------------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becon
```

3.1)

a)

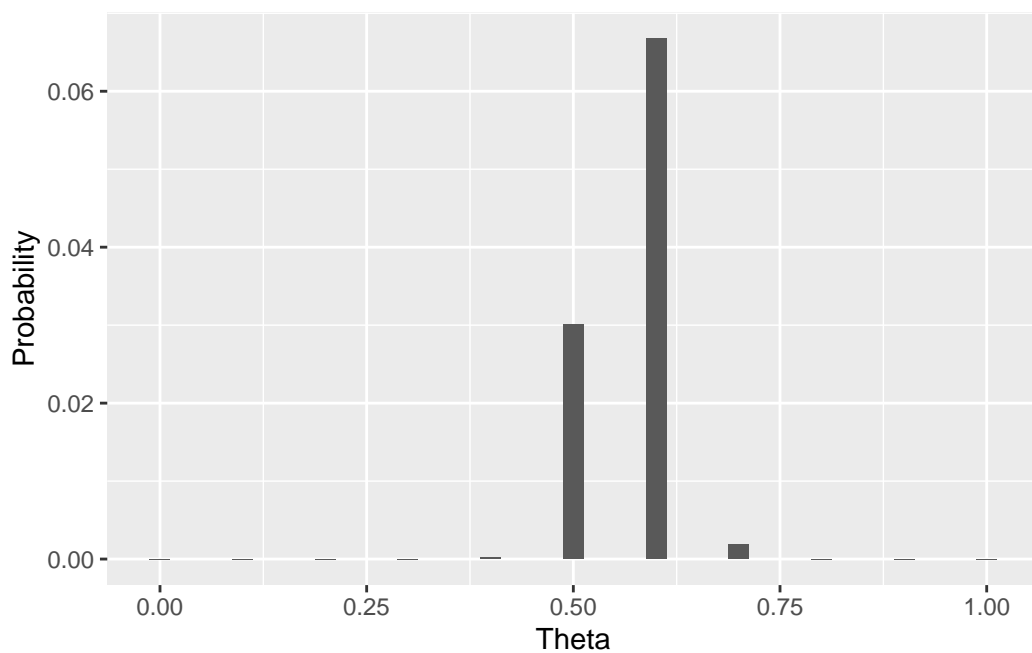$$Pr(Y_1 = y_1, ..., Y_{100} = y_{100}|\theta) = \theta^{\sum_{i=1}^{100} y_i}(1-\theta)^{100-\sum_{i=1}^{100} y_i}$$

$$Pr(\sum Y_i = y|\theta) = \binom{100}{y}\theta^y(1-\theta)^{100-y}$$

b)

```
theta = seq(0,1,.1)
y_sum = 57
n = 100
pr = dbinom(y_sum, size = 100,prob = theta)
df = tibble(theta,pr)
df
```

```
# A tibble: 11 x 2
   theta        pr
   <dbl>      <dbl>
 1   0    0
 2   0.1 4.11e-31
 3   0.2 3.74e-16
 4   0.3 1.31e- 8
 5   0.4 2.29e- 4
 6   0.5 3.01e- 2
 7   0.6 6.67e- 2
 8   0.7 1.85e- 3
 9   0.8 1.00e- 7
10   0.9 9.40e-18
11   1    0
```

```
ggplot(df,aes(x = theta, y = pr)) +
  geom_col(width = .025) +
  labs(x = "Theta", y = "Probability")
```


c)

We have:

$$p(\theta | \sum_{i=1}^{100} Y_i = 57) = \frac{Pr(\theta = \theta_i)Pr(\sum_{i=1}^{100} y_i = 57|\theta_i)}{Pr(\sum_{i=1}^{57} Y_i = 57)}$$

We can't find $Pr(\sum_{i=1}^{57} Y_i = 57)$ directly, but we can say that $Pr(\theta = \theta_i)Pr(\sum_{i=1}^{100} y_i = 57|\theta_i)$ is proportional to the sampling probability. So we can find those values, and then find the constant that allows for the probabilities to sum to 1.

Since we are assuming $Pr(\theta = \theta_i)$ is the same for all $\theta = 0, .1, ..., 1$ we can say that $Pr(\theta = 0, .1, .2, ..., 1) = 1/11$.

```
df <- df |>
  mutate(post_precorrection = 1/11*pr)
p_sum = sum(df$post_precorrection)
df <- df |>
  mutate(posterior = post_precorrection*1/p_sum)
ggplot(df,aes(x = theta, y = posterior)) +
  geom_col(width = .025) +
  labs(x = "Theta", y = "Probability", title = "Posterior")
```
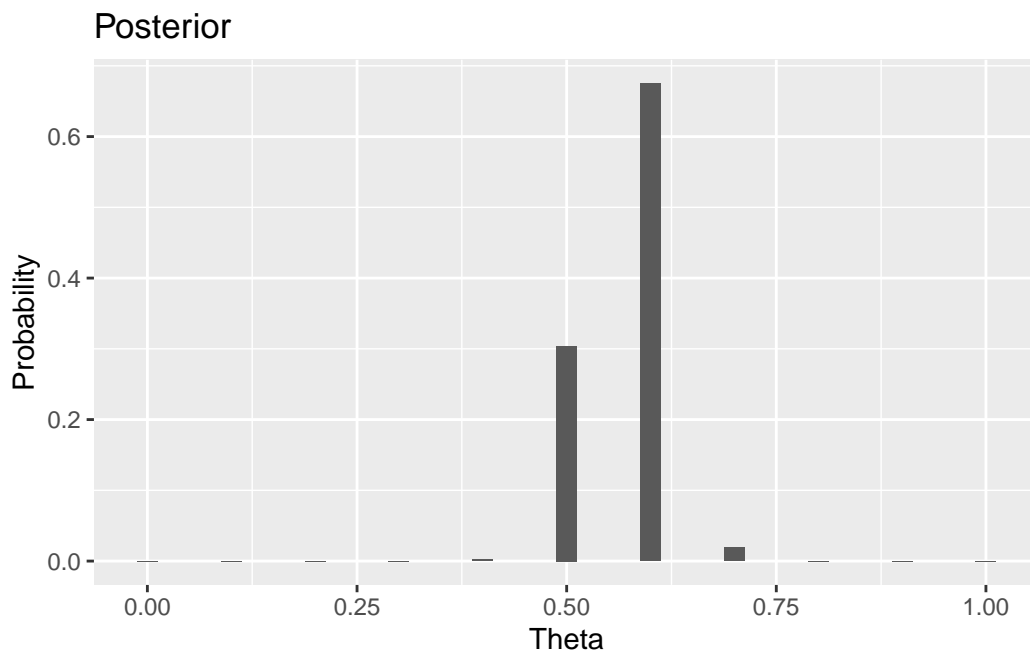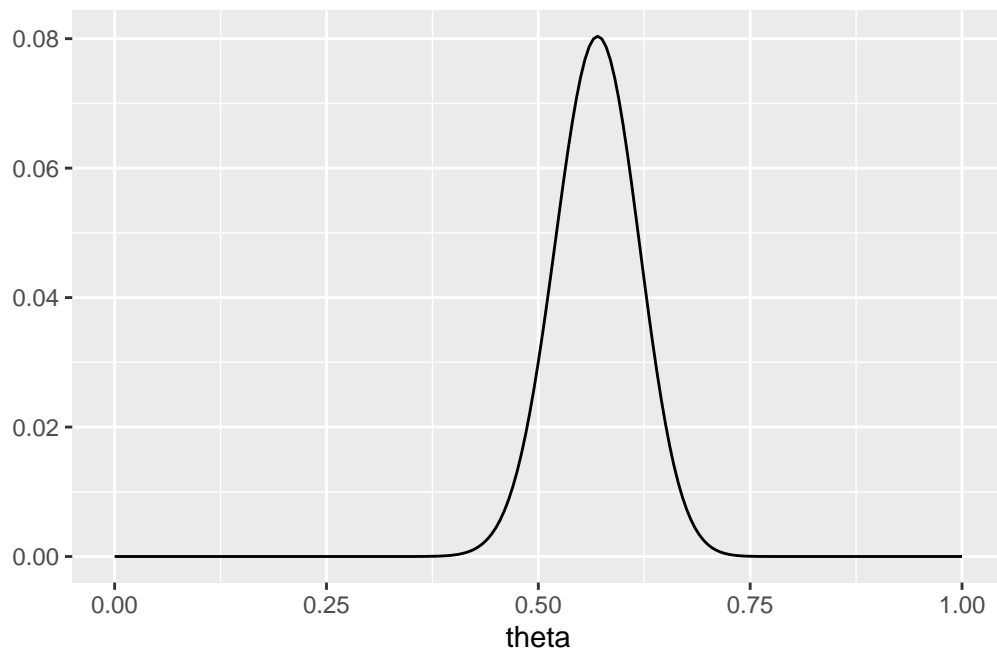

d)

Using the same idea, but this time with a uniform prior for θ we have:

3

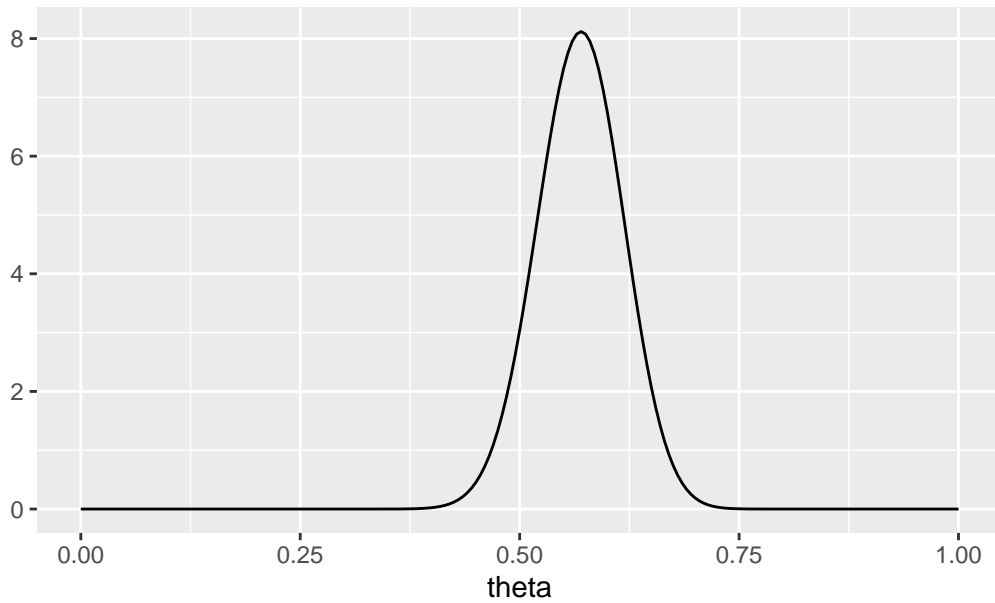$$p(\theta) \times Pr(\sum_{i=1}^{100} Y_i = 57|\theta) = 1 \times Pr(\sum_{i=1}^{100} Y_i = 57|\theta)$$

```r
theta_2 <- seq(0,1,.005)
pr_2 <- dbinom(y_sum, size = 100,prob = theta_2)
df_2 <- tibble(theta_2,pr_2)
ggplot(df_2, aes(theta_2, pr_2)) +
  geom_line() +
  labs(x = "theta", y = "")
```


e)

```r
df_2 <- df_2 |>
  mutate(true_post = dbeta(theta_2, 1 + 57, 1 + 100 - 57))
df_2 |>
  ggplot(aes(x = theta_2, y = true_post)) +
  geom_line() +
  labs(x = "theta", y = "", title = "True Posterior")
```

True Posterior

All the plots peak around the same area, but there are differences. When we assumed that θ had the same probability on $\theta = 0, .1, ..., 1$ and no probability elsewhere we are saying θ is a discrete random variable. As such we get back a posterior distribution that is a PMF. When we assume θ has a uniform distribution we get back a PDF. The differences between each group is the height on the y-axis. In some plots we are only looking at values proportional to the true posterior distribution. However, if our goal was to choose a particular θ that height wouldn't impact our decision.

3.3)

```
y_a = c(12,9,12,14,13,13,15,8,15,6)
y_b = c(11,11,10,9,9,8,7,10,6,8,8,9,7)
```

a)

We know that given a Gamma prior and a sampling distribution of Poisson we have:

$$\text{gamma}(a + \sum_{i=1}^{n} Y_i, b + n)$$

```
# A
y_a_sum = sum(y_a)
```

```
y_a_n = 10
y_a_a = 120
y_a_b = 10
y_a_posterior = c(y_a_a + y_a_sum, y_a_b + y_a_n)

# B
y_b_sum = sum(y_b)
y_b_n   = 13
y_b_a   = 12
y_b_b   = 1
y_b_posterior = c(y_b_a + y_b_sum, y_b_b + y_b_n)
```

The posterior distribution of type A is $\text{gamma}(237, 20)$ and type B is $\text{gamma}(125, 14)$.

The expected value and variance of the posterior is:

$$E[\theta] = \frac{a}{b}, Var[\theta] = \frac{a}{b^2}$$

```
y_a_posterior_exp = y_a_posterior[1]/y_a_posterior[2]
y_b_posterior_exp = y_b_posterior[1]/y_b_posterior[2]

y_a_posterior_var = y_a_posterior[1]/(y_a_posterior[2]^2)
y_b_posterior_var = y_b_posterior[1]/(y_b_posterior[2]^2)
```

The expected value of the posterior of type A is 11.85 and type B is 8.929 .

The variance of the posterior of type A is 0.593 and type B is 0.638 .

```
y_a_int = qgamma(c(.025,.975), y_a_posterior[1], y_a_posterior[2])
y_b_int = qgamma(c(.025,.975), y_b_posterior[1], y_b_posterior[2])
```

The 95% quantile based confidence interval for θ_A is (10.389,13.405) and θ_B is (7.432,10.56)
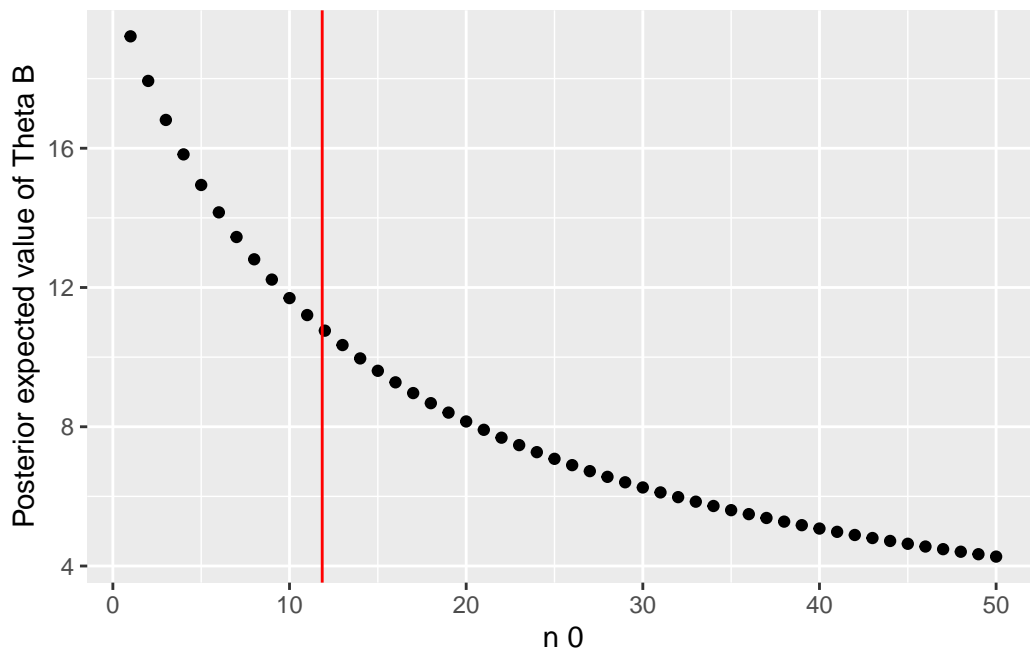
b)

```
y_b_n_seq = seq(1,50,1)
y_b_n_posterior_exp_seq = (12*y_b_n + y_b_sum)/(y_b_n_seq + y_b_n)
df = tibble(y_b_n_seq, y_b_n_posterior_exp_seq)
df |>
  ggplot(aes(y_b_n_seq, y_b_n_posterior_exp_seq)) +
  geom_point() +
  geom_vline(xintercept = y_a_posterior_exp, col = "red") +
```

```
labs(x = "n 0", y = "Posterior expected value of Theta B")
```


In order for the posterior expectation of θ_B to be close to that of θ_A would be a prior of gamma$((12)(17), 17)$. That prior distribution has an expected value of 12 with a strong weight being placed on the prior distribution (the equivalent of 17 prior observations with a mean of 12).

d)

Knowledge about population A mice could be used to tell us something about population B mice, but we would have to be careful not to overstate how related the two populations are. I think it makes good sense to consider the two populations independent in an initial analysis (as we have by assuming $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$.

3.4)

```
n = 43
y = 15
a = 2
b = 8
theta = seq(0,1,.005)
p_theta = dbeta(theta, a, b)
```

```
p_y_theta = dbinom(y, n, theta)
p_theta_y = dbeta(theta, a + y, b + n - y)
df = tibble(theta,p_theta,p_y_theta,p_theta_y)
```
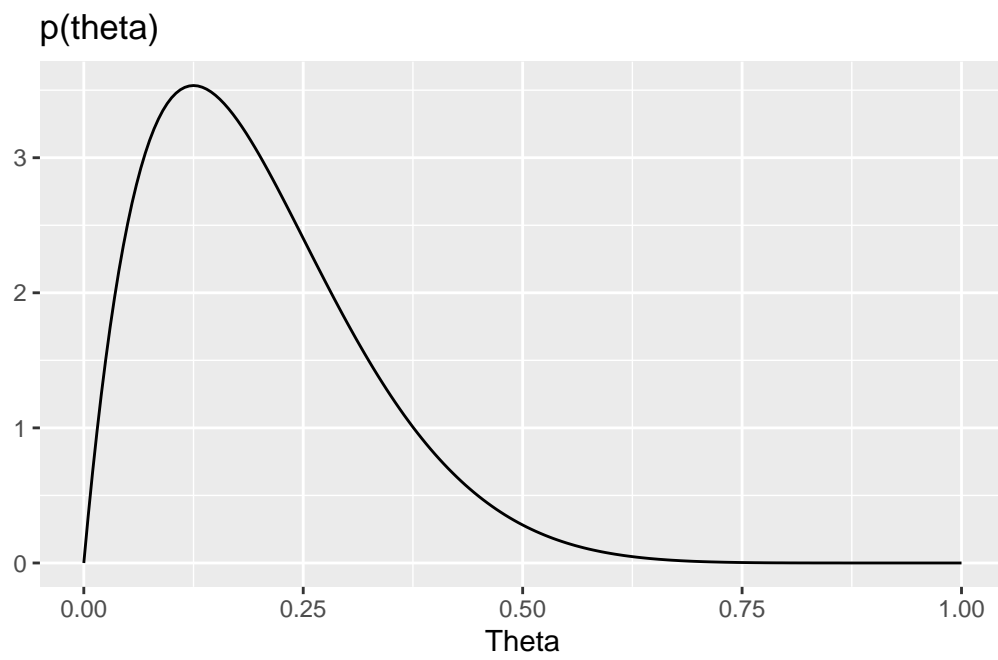
Plotting:

```
par(mfrow= c(2,2))
df |>
  ggplot(aes(x = theta, y = p_theta)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(theta)")
```
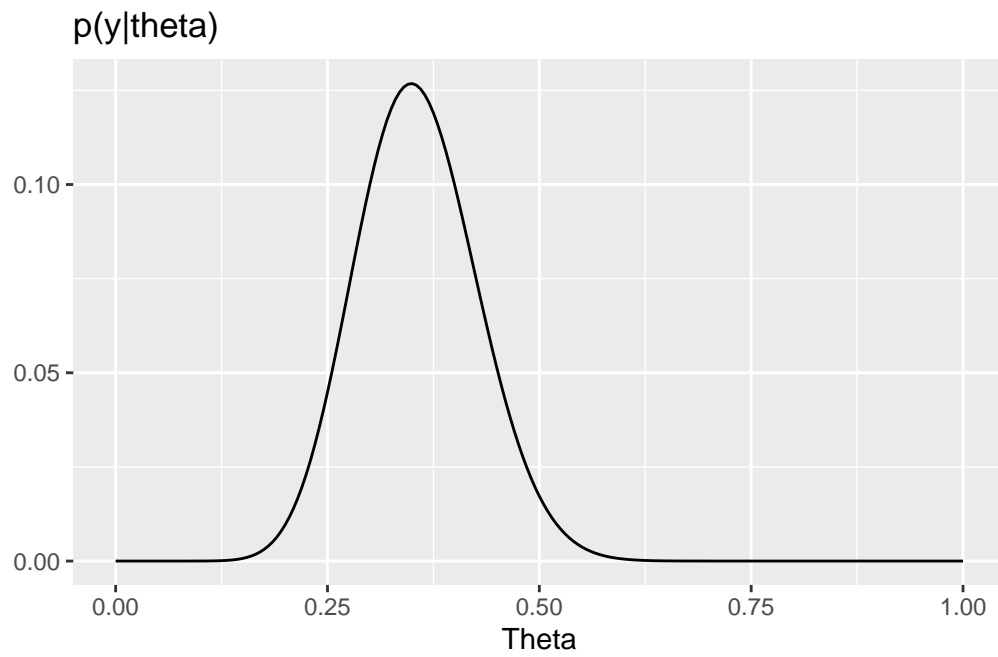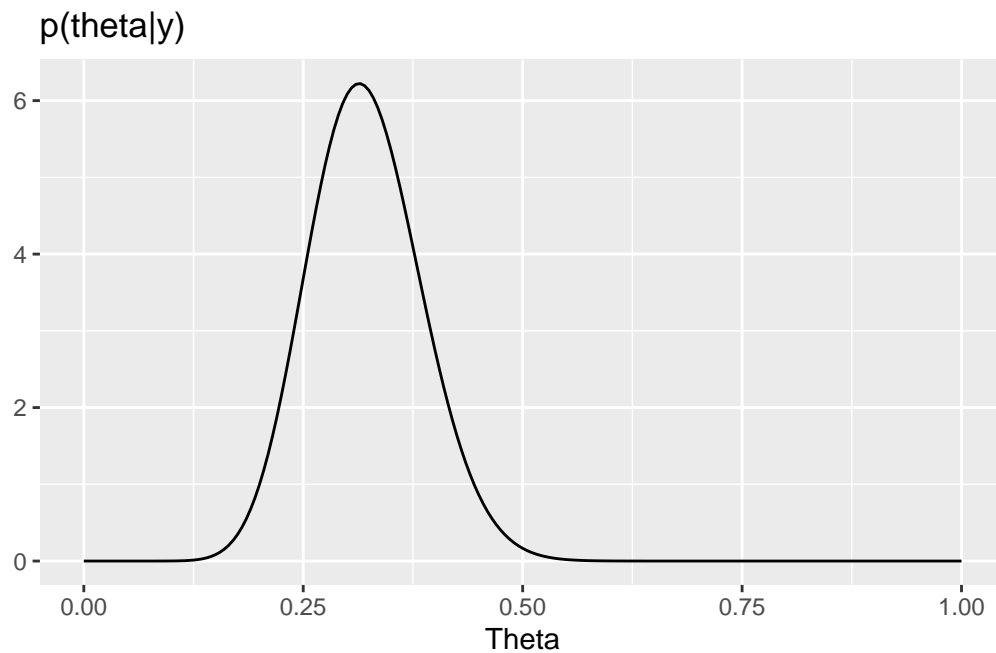


```
df |>
  ggplot(aes(x = theta, y = p_y_theta)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(y|theta)")
```

p(y|theta)

```
df |>
  ggplot(aes(x = theta, y = p_theta_y)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(theta|y)")
```

p(theta|y)

```
p_a = a + y
p_b = b + n - y
p_mode = (p_a - 1)/((p_a - 1) + (p_b - 1))
p_exp = p_a/(p_a + p_b)
p_sd = sqrt(p_a*p_b/((p_a+p_b+1)*(p_a+p_b)^2))
```

Mean	Mode	Standard Dev
0.321	0.314	0.064

```
p_int = qbeta(c(.025,.975), p_a, p_b)
```
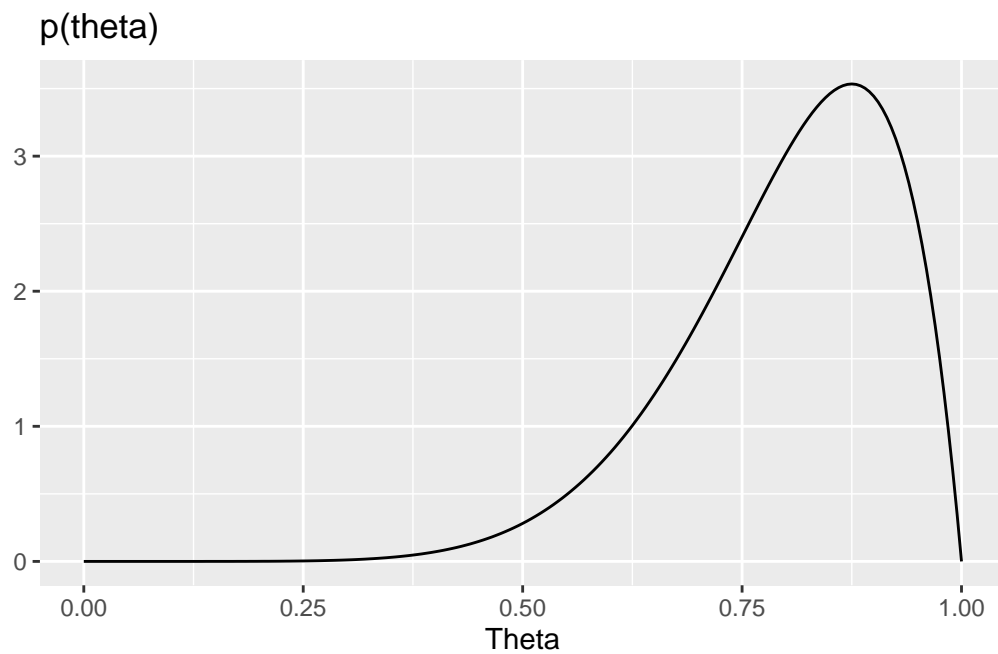
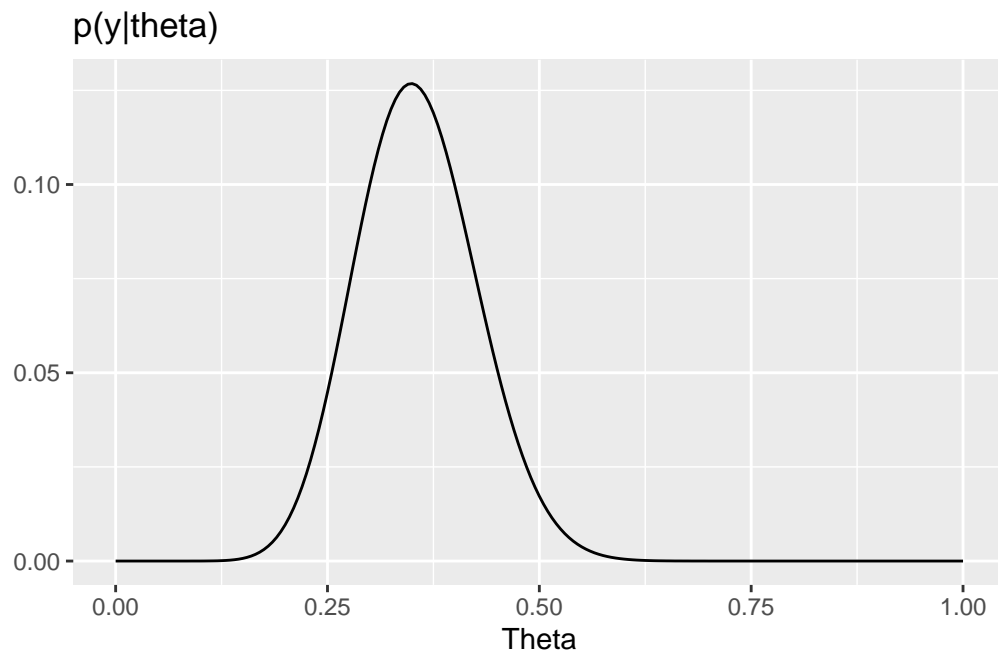The 95% confidence interval for θ is $(0.203, 0.451)$

b)

```
a_2 = 8
b_2 = 2
p_theta_2 = dbeta(theta, a_2, b_2)
p_y_theta_2 = dbinom(y, n, theta)
p_theta_y_2 = dbeta(theta, a_2 + y, b_2 + n - y)
df_2 = tibble(theta,p_theta_2,p_y_theta_2,p_theta_y_2)
```

Plotting:

```
par(mfrow= c(2,2))
df_2 |>
  ggplot(aes(x = theta, y = p_theta_2)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(theta)")
```
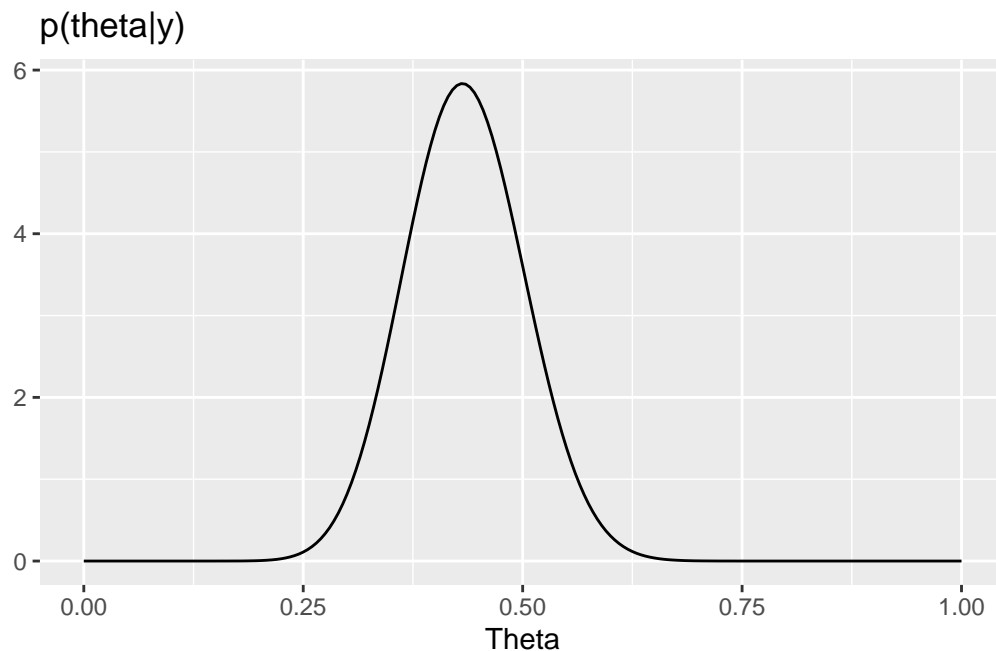

p(theta)

```
df_2 |>
  ggplot(aes(x = theta, y = p_y_theta_2)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(y|theta)")
```

p(y|theta)


```r
df_2 |>
  ggplot(aes(x = theta, y = p_theta_y_2)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "p(theta|y)")
```

p(theta|y)


```
p_a_2 = a_2 + y
p_b_2 = b_2 + n - y
p_mode_2 = (p_a_2 - 1)/((p_a_2 - 1) + (p_b_2 - 1))
p_exp_2 = p_a_2/(p_a_2 + p_b_2)
p_sd_2 = sqrt(p_a_2*p_b_2/((p_a_2+p_b_2+1)*(p_a_2+p_b_2)^2))
```

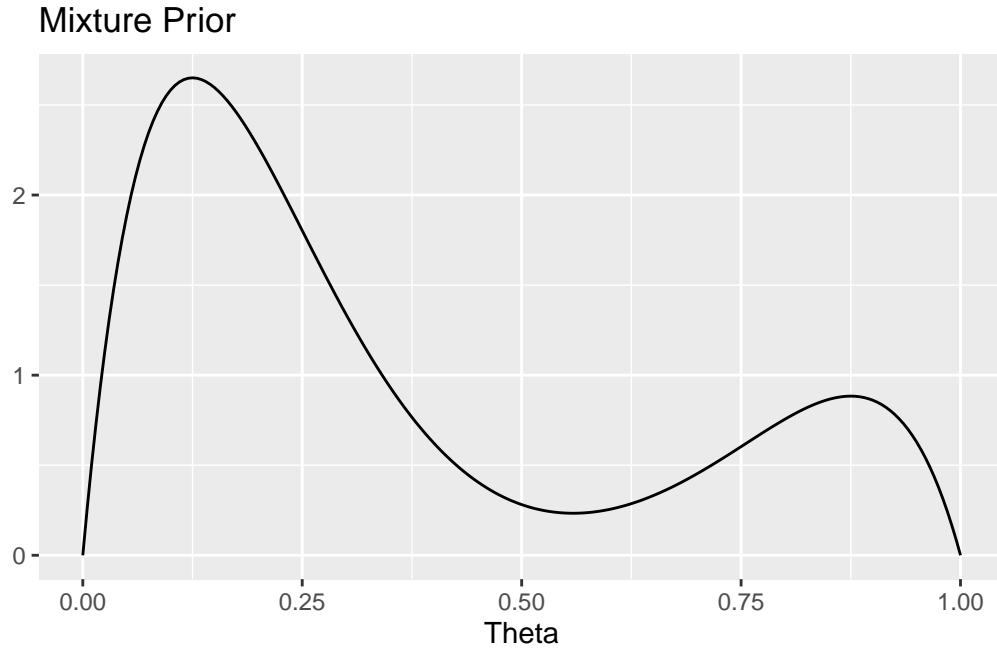| Mean	Mode	Standard Dev
0.434	0.431	0.067

```
p_int_2 = qbeta(c(.025,.975), p_a_2, p_b_2)
```

The 95% confidence interval for θ is (0.305,0.568)

c)

```
p_theta_mix = function(theta){
  return(1/4*(gamma(10)/(gamma(2)*gamma(8)))*(3*theta*(1-theta)^7 +
                                          theta^7*(1-theta)))
}
p_theta_mix_seq = p_theta_mix(theta)
```

13

```
df_3 = tibble(theta,p_theta_mix_seq)
df_3 |>
  ggplot(aes(x = theta, y = p_theta_mix_seq)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "Mixture Prior")
```


Mixture Prior

This prior distribution takes aspects from the priors in a and b. The two peaks are in the same location as a and b, with more weight being placed on the prior in part a (beta(2,8)). This prior could be used to express the opinion that there is either a low proportion of re-offenders or a high proportion, with little chance of the proportion being in between.

d)

i)

We have $p(y|\theta) = \theta^{15}(1 - \theta)^{28}$ so:

$$p(\theta)p(y|\theta) = \frac{3}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\theta^{16}(1 - \theta)^{35} + \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\theta^{22}(1 - \theta)^{29}$$

ii)

We can identify the roots of two beta distributions, however their coefficients are unclear.

My initial idea is this being proportional to $\frac{3}{4}$beta(17,36)$+\frac{1}{4}$beta(23, 30), but the weights seem off.

iii)

```
p_post = function(theta){
   return(1/4*(gamma(10)/(gamma(2)*gamma(8)))*(3*theta^(16)*(1-theta)^(35) +
                                      theta^(22)*(1-theta)^(29)))}
p_post_seq = p_post(theta)
df_4 = tibble(theta, p_post_seq)
df_4 |>
   ggplot(aes(x = theta, y = p_post_seq)) +
   geom_line() +
   labs(x = "Theta", y = "", title = "Mixed Dist p(theta) x p(y|theta)")
```
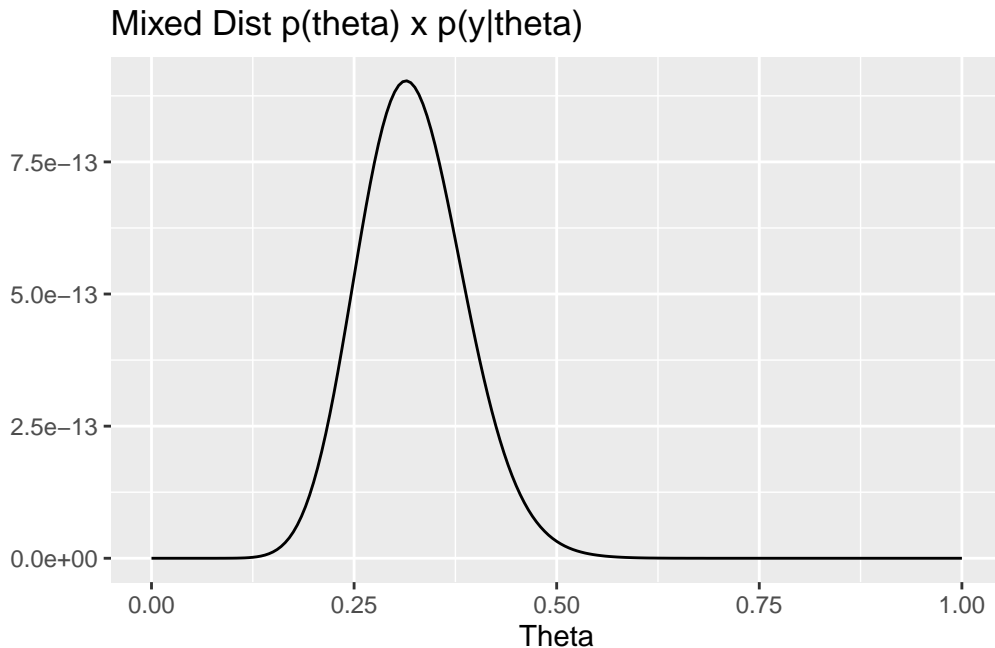

The approximate posterior mode is 0.315.

The mode is close to the mode from the beta(2,8) prior, this makes sense because the data favors a value of $\theta < .5$ and we have more weight placed on the beta(2,8) than beta(8,2).

e)

A general formula would be, working off the notes where it covered a poisson example:

Given a prior of $p(\theta) = \sum_{j=1}^{m} k_j f_j(\theta)$ where $k_1, ... k_m$ are the weights we would have a posterior of:

$$p(\theta|Y) = \sum_{j=1}^{m} \tilde{k}_j f_j(\tilde{\theta}), \tilde{k}_j = k_j c_j/c, c = \sum_{j=1}^{m} k_j c_j, f_j(\theta) p(\tilde{Y}|\theta)/c_j, c_j = \int f_j(\theta) p(y|\theta) d\theta$$

This part confused me a bit, I'm not super sure how the mixture is updated after data is observed.

3.7)

a)

Given a sampling distribution being binomial, a uniform prior distribution leads to a posterior distribution of:

$$p(\theta|y) \sim \text{beta}(1 + y, 1 + n - y), \text{beta}(3, 14)$$

Then the mean, mode and s.d. of θ is:

$$E[\theta|y] = \frac{a + y}{a + b + n} \quad \text{mode}[\theta|y] = \frac{a + y - 1}{a + b + n - 2} \quad SD[\theta|y] = \sqrt{\frac{E[\theta|y]E[1 - \theta|y]}{a + b + n + 1}}$$

```
#data
n = 15
y = 2
#priors
a = 1
b = 1


#posterior statistics
expected = (a + y)/(a + b + n)
mode     = (a + y - 1)/(a + b + n - 2)
sd       = sqrt(expected*(1-expected)/(a + b + n + 1))
```
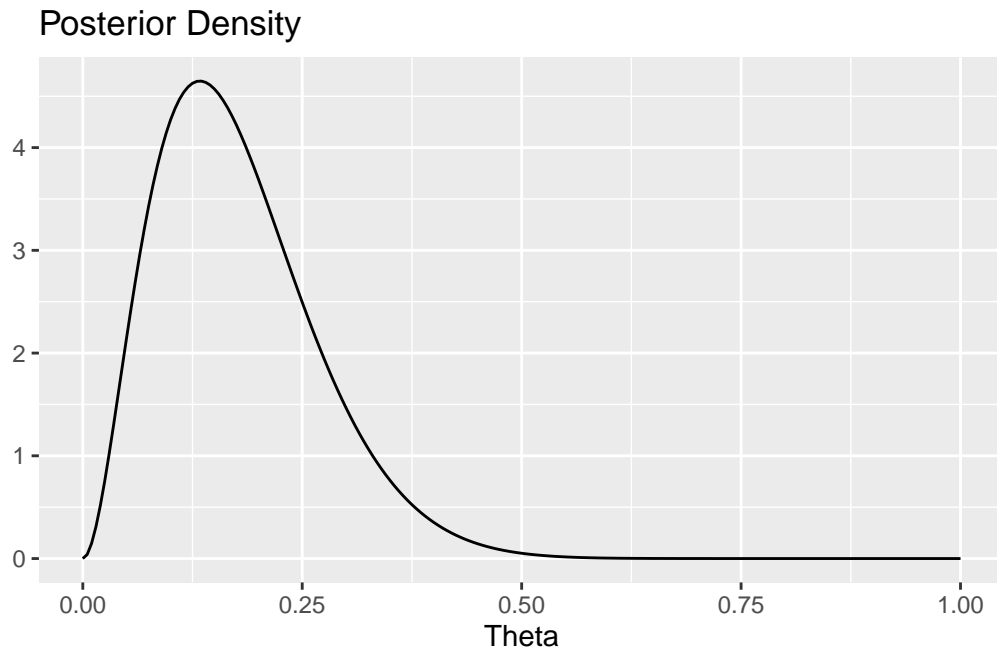
Mean	Mode	Standard Dev
0.176	0.133	0.09

16

```
theta = seq(0,1,.005)
post_denstiy = dbeta(theta, shape1 = a + y, shape2 = b + n - y)
df = tibble(theta,post_denstiy)
df |>
  ggplot(aes(x = theta, y= post_denstiy)) +
  geom_line() +
  labs(x = "Theta", y = "", title = "Posterior Density")
```


b)

i.

We need to assume that Y_1 and Y_2 are independent given θ.

ii.

We assume that the sampling distribution $Pr(Y_2|\theta)$ is binomial.

$$Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 \binom{n_2}{y_2}\theta^{y_2}(1-\theta)^{n_2-y_2}\frac{\Gamma(17)}{\Gamma(3)\Gamma(14)}\theta^2(1-\theta)^{13}d\theta$$

$$= \binom{n_2}{y_2}\frac{\Gamma(17)}{\Gamma(3)\Gamma(14)}\int_0^1 \theta^{y_2+2}(1-\theta)^{n_2+13-y_2}d\theta$$

17

iii.

Then I believe since that integral must integrate to 1 we can use the calculus trick from page 33 of the textbook, then:

$$p(Y_2 = y_2|Y_1 = 2) = \binom{n_2}{y_2}\frac{\Gamma(17)}{\Gamma(3)\Gamma(14)}\int_0^1 \theta^{y_2+2}(1-\theta)^{n_2+13-y_2}\,d\theta = \binom{n_2}{y_2}\frac{\Gamma(17)}{\Gamma(3)\Gamma(14)}\frac{\Gamma(y_2+3)\Gamma(n_2+14-y_2)}{\Gamma(n_2+14-y_2)}$$

c)

```
n_2 = 278
p_y2_given_y1 = function(y_2){
  #return(gamma(y_2 + 1)* gamma(n_2 + 13 - y_2)/(gamma(n_2 + 14)))
  return(choose(n_2,y_2)*(1/beta(3,14))*beta(y_2 + 3, n_2 + 14 - y_2))
}
y_2_seq = seq(0,278,1)
p_y2_given_y1_seq = p_y2_given_y1(y_2_seq)

df = tibble(y_2_seq,p_y2_given_y1_seq)

df |>
  ggplot(aes(x = y_2_seq, y = p_y2_given_y1_seq)) +
  geom_col() +
  labs(x = "y2", y = "", title = "Pr(Y2 = y2|Y1 = 2)")
```
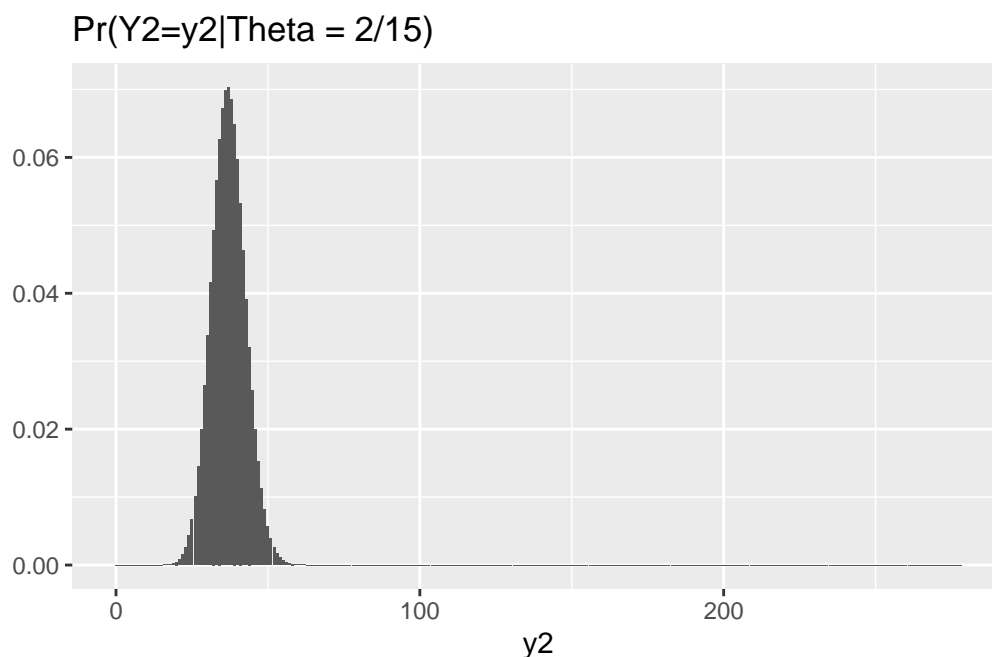
Pr(Y2 = y2|Y1 = 2)
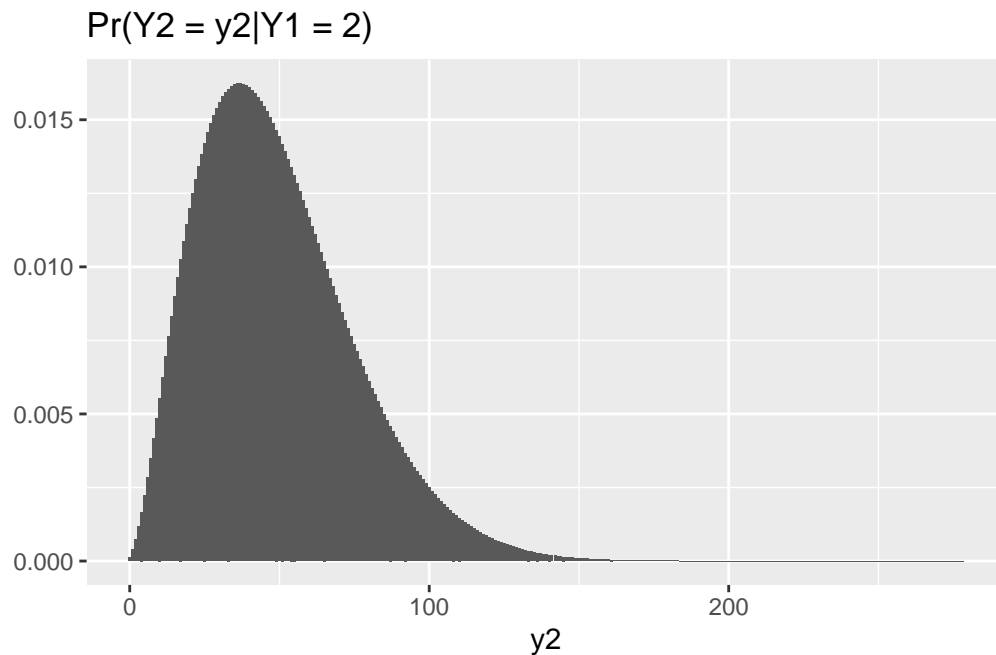

```
y_2_given_y_1_mean = sum(df$p_y2_given_y1_seq*y_2_seq)
y_2_given_y_1_sd   = sqrt(sum(df$p_y2_given_y1_seq*y_2_seq^2)
                         - y_2_given_y_1_mean^2)
```

Mean	SD
49.059	25.732

d)

```
theta_hat = 2/15
mle_seq = dbinom(y_2_seq,n_2,theta_hat)
df = tibble(theta_hat, mle_seq)
df |>
  ggplot(aes(x = y_2_seq, y = mle_seq)) +
  geom_col() +
  labs(x = "y2",y = "", title = "Pr(Y2=y2|Theta = 2/15)")
```

Pr(Y2=y2|Theta = 2/15)


```
mle_mean = theta_hat*n_2
mle_sd   = sqrt(n_2*theta_hat*(1 - theta_hat))
```

Mean	SD
37.067	5.668

The biggest difference between the two methods is the variability of our predictions.

In the Bayesian approach we are factoring in the very small sample size used to give an initial idea of what θ could be. This allows us to properly allow for much more error in the prediction of y_2.

In the approach where we accept the MLE of θ we end up with a much tighter distribution of predicted y_2 values.

If I was to choose a distribution for Y_2 to make predictions I would certainly choose the Bayesian approach, the initial proportion was based on such a small group of students that it would be inappropriate to not factor in the large amount of error we have in our estimate of θ , the population proportion.

3.12)

a)

When the sampling distribution of Y is binomial(n, θ) the Jeffery's prior is Beta$(.5, .5)$ as developed in the notes on slide 37. Which is:

$$p_J(\theta) = \frac{\Gamma(1)}{\Gamma(.5)\Gamma(.5)}\theta^{-.5}(1-\theta)^{-.5}$$

b)

We start with:

$$p(y|\psi) = \binom{y}{n}e^{\psi y}(1+e^{\psi})^{-n}$$

Then we take the log (dropping terms we know will drop out after the first derivative):

$$\log[p(y|\psi)] = \psi y - n\log(1+e^{\psi})$$

Then we take the first derivative with respect to ψ:

$$\log(\frac{\delta}{\delta\psi}p(y|\psi)) = y - n\frac{e^{\psi}}{1+e^{\psi}}$$

Then we take the second derivative with respect to ψ:

$$\log(\frac{\delta}{\delta\psi^2}p(y|\psi)) = -n\frac{e^{\psi}}{(1+e^{\psi})^2}$$

Then we take the negative expected value:

$$-E[\log(\frac{\delta}{\delta\psi^2}p(y|\psi))] = n\frac{e^{\psi}}{(1+e^{\psi})^2}$$

We then get the Jeffery's prior for ψ by taking the square root:

$$p_J(\psi) = [n\frac{e^{\psi}}{(1+e^{\psi})^2}]^{1/2}$$

c)

The prior distribution from a) is:

$$p_J(\theta) = \frac{\Gamma(1)}{\Gamma(.5)\Gamma(.5)}\theta^{-.5}(1-\theta)^{-.5}$$

We then apply the change of variables formula:

$$p_\psi(\psi) = p_J(h(\psi)) \times |\frac{dh}{d\psi}|$$

Where $h(\psi) = \frac{e^\psi}{1+e^\psi}$, $|\frac{dh}{d\psi}| = \frac{e^\psi}{(1+e^\psi)^2}$ and $p_J(h(\psi)) = \frac{\Gamma(1)}{\Gamma(.5)\Gamma(.5)}[\frac{e^\psi}{1+e^\psi}(1 - \frac{e^\psi}{1+e^\psi})]^{-1/2}$.

Then (using c as a constant to avoid repedition):

$$p_\psi(\psi) = c[\frac{e^\psi}{1+e^\psi}(1 - \frac{e^\psi}{1+e^\psi})]^{-1/2}\frac{e^\psi}{(1+e^\psi)^2}$$

Then we manipulate the right most term:

$$p_\psi(\psi) = c[\frac{e^\psi}{1+e^\psi}(1 - \frac{e^\psi}{1+e^\psi})]^{-1/2}[\frac{(1+e^\psi)^4}{e^{2\psi}}]^{-1/2}$$

$$p_\psi(\psi) = c[\frac{(1+e^\psi)^3}{e^\psi}(1 - \frac{e^\psi}{1+e^\psi})]^{-1/2}$$

$$p_\psi(\psi) = c[\frac{(1+e^\psi)^3}{e^\psi} - \frac{1}{(1+e^\psi)^2}]^{-1/2}$$

$$p_\psi(\psi) = c[\frac{(1+e^\psi)^5 - e^\psi}{e^\psi(1+e^\psi)^2}]^{-1/2}$$

$$p_\psi(\psi) = c[\frac{e^\psi(1+e^\psi)^2}{(1+e^\psi)^5 - e^\psi}]^{1/2}$$

I'd continue further, but I'm not sure how I'd be able to get rid fo the $e^{5\psi}$ that will pop up, I think I made an algebra mistake somewhere.