

# STA 602. HW02

Yicheng Shen

9/10/2022

## 1. PH 3.1

(a) The joint distribution is

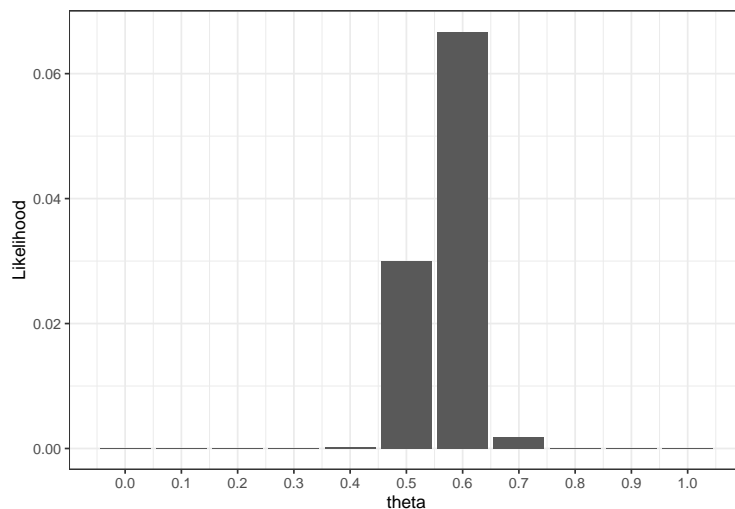
$$\begin{aligned} Pr(Y_1 = y_1, \dots, Y_{100} = y_{100} \mid \theta) &= \theta^{y_1 + y_2 + \dots + y_{100}} (1 - \theta)^{1 - y_1 + 1 - y_2 + \dots + 1 - y_{100}} \\ &= \theta^{\sum y_i} (1 - \theta)^{100 - \sum y_i} \\ Pr(\sum Y_i = y \mid \theta) &= \binom{100}{y} \theta^y (1 - \theta)^{100 - y} \end{aligned}$$

(b) The table and plot are shown below.

```
bayes_table <- data.frame(  
  theta = seq(0, 1, by = 0.1),  
  Prior = rep(1/11, 11)  
) %>%  
  mutate(Likelihood = dbinom(57, size = 100, prob = theta))  
round(bayes_table, 3)
```

##	theta	Prior	Likelihood
## 1	0.0	0.091	0.000
## 2	0.1	0.091	0.000
## 3	0.2	0.091	0.000
## 4	0.3	0.091	0.000
## 5	0.4	0.091	0.000
## 6	0.5	0.091	0.030
## 7	0.6	0.091	0.067
## 8	0.7	0.091	0.002
## 9	0.8	0.091	0.000
## 10	0.9	0.091	0.000
## 11	1.0	0.091	0.000

```
ggplot(bayes_table, aes(x = theta, y = Likelihood)) + geom_bar(stat = 'identity') +  
  scale_x_continuous(breaks = bayes_table$theta) + theme_bw()
```



(c) By Bayes's rule:

$$p(\theta \mid \sum_{i=1}^n Y_i = 57) = \frac{p(\sum Y_i = 57 \mid \theta)p(\theta)}{p(\sum Y_i = 57)}$$

Since we assume that that  $p(\theta)$  is the same across all  $\theta$ , we have

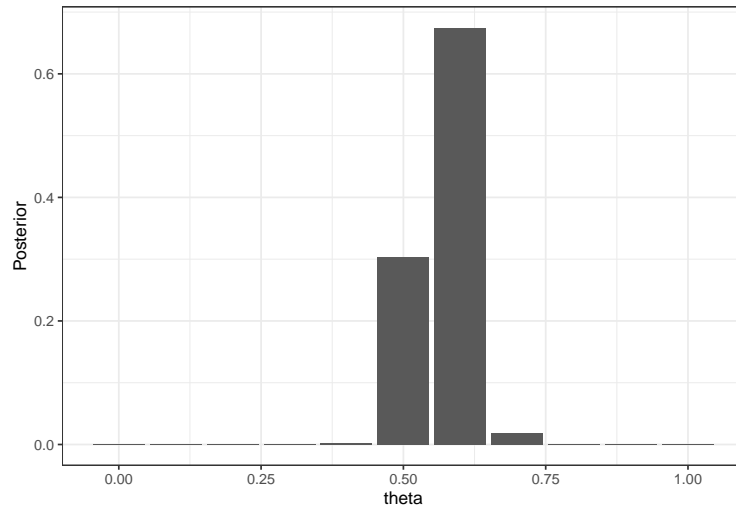
$$p(\theta \mid \sum Y_i = 57) \propto p(\sum Y_i = 57 \mid \theta)$$

So the posterior distribution has the same shape as the Likelihood, but different scale. It can be calculated as the following:

```
# bayes_table %>%
#   mutate(Posterior = Likelihood*Prior/ sum(Likelihood*Prior))
posterior_table <- bayesian_crank(bayes_table) %>% dplyr::select(-Product)
round(posterior_table, 3)
```

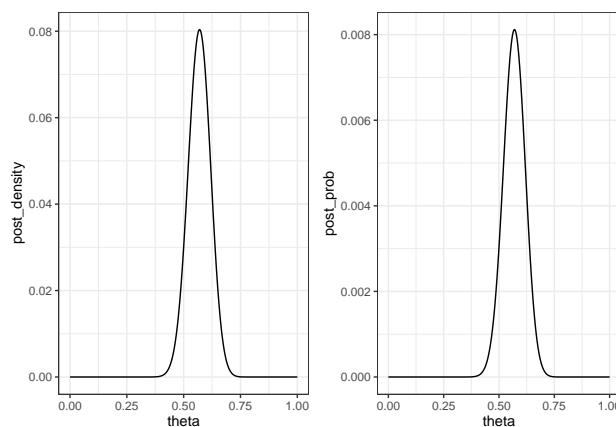
##	theta	Prior	Likelihood	Posterior
## 1	0.0	0.091	0.000	0.000
## 2	0.1	0.091	0.000	0.000
## 3	0.2	0.091	0.000	0.000
## 4	0.3	0.091	0.000	0.000
## 5	0.4	0.091	0.000	0.002
## 6	0.5	0.091	0.030	0.304
## 7	0.6	0.091	0.067	0.675
## 8	0.7	0.091	0.002	0.019
## 9	0.8	0.091	0.000	0.000
## 10	0.9	0.091	0.000	0.000
## 11	1.0	0.091	0.000	0.000

```
ggplot(posterior_table, aes(x = theta, y = Posterior)) +
  geom_bar(stat = 'identity') + theme_bw()
```



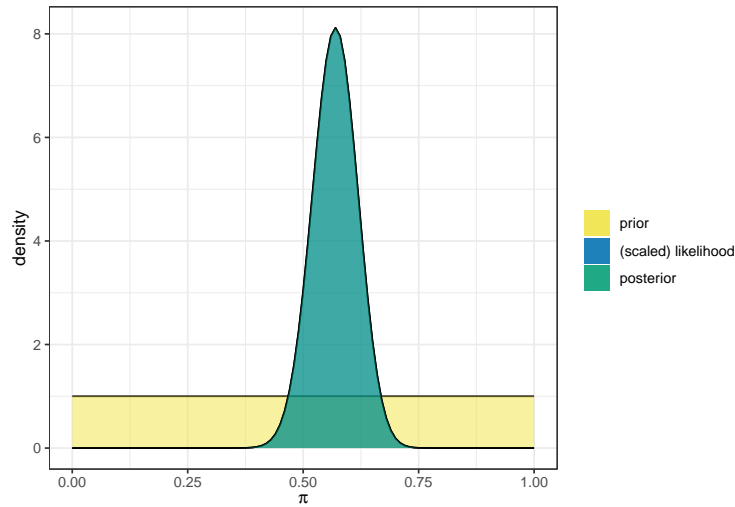
(d) Now we have the uniform distribution that describes the prior information. Since  $p(\theta) = 1$ , prior  $\times$  likelihood is  $p(\theta) \times P(\sum Y_i = 57 \mid \theta) = P(\sum Y_i = 57 \mid \theta)$ .

```
theta <- seq(0, 1, by = 0.001)
post <- dbinom(57, 100, theta)
bayestable <- data.frame(theta = theta, post_density = post) %>%
  mutate(post_prob = post_density / sum(post_density))
a <- ggplot(bayestable) + geom_line(aes(x = theta, y = post_density)) + theme_bw()
b <- ggplot(bayestable) + geom_line(aes(x = theta, y = post_prob)) + theme_bw()
grid.arrange(a,b, nrow = 1)
```



(e) Given that prior is  $Beta(1, 1)$ , posterior is  $Beta(1 + 57, 1 + 100 - 57) = Beta(58, 44)$

```
# plot_beta(58, 44) # alternative, same posterior plot
plot_beta_binomial(1, 1, 57, 100) + theme_bw()
```

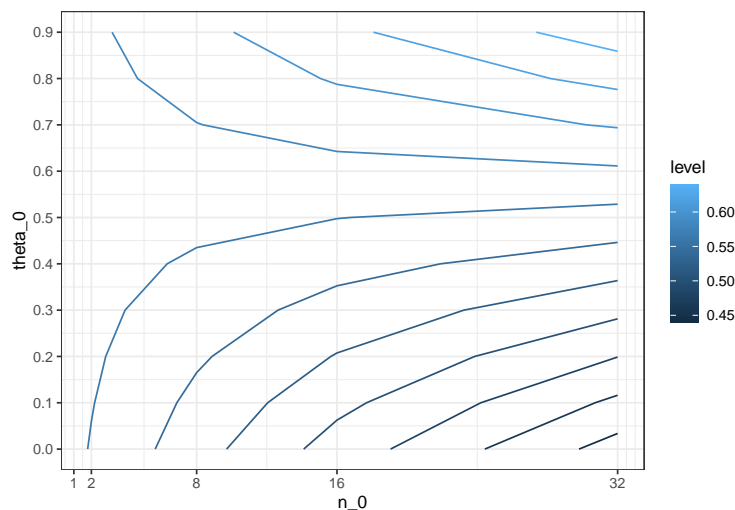


The plot in part (b) is the likelihood which describes the data; the plot in (c) is the shape of posterior that is a compromise of prior and likelihood. The plot in (e) is the posterior from a continuous uniform prior and dominated by the likelihood data.

## 2. PH 3.2

Based on the contour plot, because of evidence from sampled data, we could believe that  $\theta > 0.5$  during most of time, except prior info is really strong in the other direction, namely when the prior belief of  $\theta$  (expressed as  $\theta_0$ ) is very low and the prior sample size (expressed as  $n_0$ ) is large.

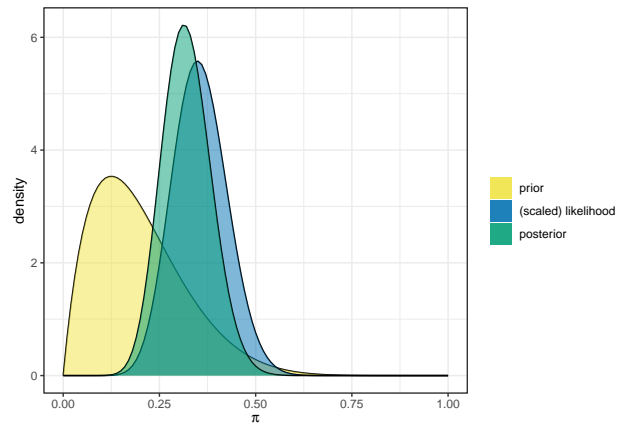
```
theta_0 <- seq(0, 0.9, by = 0.1)
n_0 <- c(rep(1, 10), rep(2, 10), rep(8, 10), rep(16, 10), rep(32, 10))
post_table <- data.frame(theta_0 = theta_0, n_0 = n_0,
  a = theta_0 * n_0, b = (1 - theta_0) * n_0) %>%
  mutate(theta = (a + 57) / (a + 57 + b + 100 - 57))
ggplot(post_table, aes(x = n_0, y = theta_0, z = theta)) +
  geom_contour(aes(color = ..level..)) + theme_bw() + scale_y_continuous(breaks = theta_0) +
  scale_x_continuous(breaks = c(1, 2, 8, 16, 32), labels = c(1, 2, 8, 16, 32))
```



### 3. PH 3.4

- (a) The prior follows  $p(\theta) \sim \text{Beta}(2, 8)$ , the sampling model follows  $p(y|\theta) \sim \text{Binomial}(15, 43)$  and the posterior follows  $p(\theta|y) \sim \text{Beta}(2 + 15, 8 + 43 - 15) = \text{Beta}(17, 36)$ .

```
plot_beta_binomial(2,8,15,43) + theme_bw()
```



$$\text{mean}(\text{Beta}(17, 36)) = \frac{a}{a+b} = (17)/(17+36) = 0.32$$

$$\text{mode}(\text{Beta}(17, 36)) = \frac{a-1}{a+b-2} = (17-1)/(17+36-2) = 0.313$$

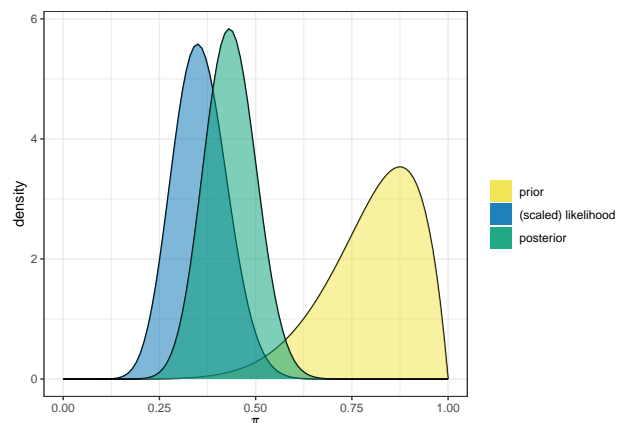
$$\text{sd}(\text{Beta}(17, 36)) = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}} = \sqrt{(17 \times 36)/[(17+36)^2(17+36+1)]} = 0.063519$$

```
cat("The 95% CI is (", qbeta(c(0.025, 0.975), 17, 36), ")\n")
```

```
## The 95% CI is ( 0.2032978 0.451024 )
```

- (b) The prior now follows  $p(\theta) \sim \text{Beta}(8, 2)$ , the sampling model follows  $p(y|\theta) \sim \text{Binomial}(15, 43)$  and the posterior follows  $p(\theta|y) \sim \text{Beta}(8 + 15, 2 + 43 - 15) = \text{Beta}(23, 30)$

```
plot_beta_binomial(8,2,15,43) + theme_bw()
```



$$\begin{aligned}\text{mean}(\text{Beta}(23, 30)) &= \frac{a}{a+b} = (23)/(23+30) = 0.434 \\ \text{mode}(\text{Beta}(23, 30)) &= \frac{a-1}{a+b-2} = (23-1)/(23+30-2) = 0.431 \\ \text{sd}(\text{Beta}(23, 30)) &= \sqrt{\frac{ab}{(a+b)^2(a+b+1)}} = \sqrt{(23 \times 30)/[(23+30)^2(23+30+1)]} = 0.067445\end{aligned}$$

```
cat("The 95% CI is (", qbeta(c(0.025, 0.975), 23, 30), ")\n")
```

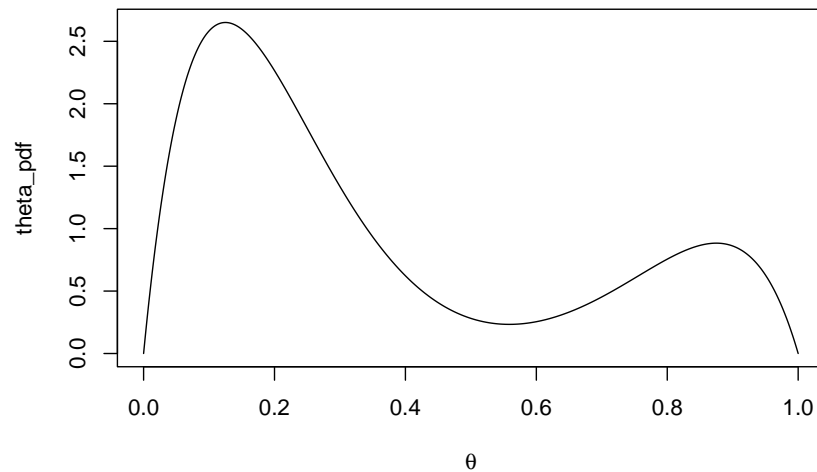
```
## The 95% CI is ( 0.3046956 0.5679528 )
```

(c) For the new prior distribution:

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

Its density plot looks like a mixture of both Beta(2,8) and Beta(8,2), with Beta(2,8) being the stronger one of the mixture. This kind of bimodal prior represents that we believe for some teens the probability of recidivism is high while for many others the probability of recidivism is pretty low.

```
theta <- seq(from = 0, to = 1, length.out = 1000)
theta_pdf <- 1/4 * gamma(10)/(gamma(2) * gamma(8)) *
  (3 * theta * (1 - theta) ^ 7 + theta ^ 7 * (1 - theta))
plot(theta, theta_pdf, type = "l", xlab = expression(theta))
```



(d) i. The derivation is shown as the following:

$$\begin{aligned}
p(\theta) &= \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} [3\theta(1-\theta)^7 + \theta^7(1-\theta)] \\
&= \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} [3\theta^{2-1}(1-\theta)^{8-1} + \theta^{8-1}(1-\theta)^{2-1}] \\
&= \frac{3}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \theta^{2-1}(1-\theta)^{8-1} + \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \theta^{8-1}(1-\theta)^{2-1} \\
&= \frac{3}{4} p_1(\theta) + \frac{1}{4} p_2(\theta)
\end{aligned}$$

$$\text{Where } p_1(\theta) = \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \theta^{2-1}(1-\theta)^{8-1}, \quad p_2(\theta) = \frac{1}{4} \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)} \theta^{8-1}(1-\theta)^{2-1}$$

$$\begin{aligned}
p(\theta|y_i) &\propto p_{mix}(\theta) \times p(y_i|\theta) \\
&\propto \left[ \frac{3}{4} p_1(\theta) + \frac{1}{4} p_2(\theta) \right] \times p(y_i|\theta) \\
&= \frac{3}{4} p_1(\theta) \times p(y_i|\theta) + \frac{1}{4} p_2(\theta) \times p(y_i|\theta)
\end{aligned}$$

$$\text{Where } p_1(\theta) \sim \text{Beta}(2, 8), \quad p_2(\theta) \sim \text{Beta}(8, 2)$$

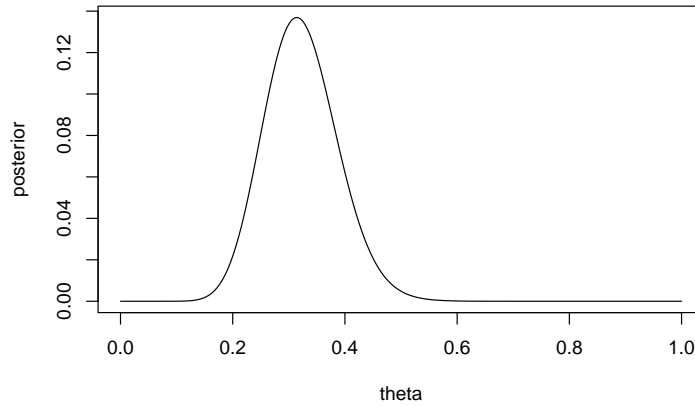
ii. As stated above, the posterior follows

$$\begin{aligned}
p(\theta|y_i) &\propto \frac{3}{4} p_1(\theta) \times p(y_i|\theta) + \frac{1}{4} p_2(\theta) \times p(y_i|\theta) \text{ where } p_1(\theta) \sim \text{Beta}(2, 8), \quad p_2(\theta) \sim \text{Beta}(8, 2) \\
&\propto \frac{3}{4} p_1(\theta) \times \left[ \binom{43}{15} \theta^{15} (1-\theta)^{28} \right] + \frac{1}{4} p_2(\theta) \times \left[ \binom{43}{15} \theta^{15} (1-\theta)^{28} \right]
\end{aligned}$$

The posterior is proportional to a mixture of Beta(17, 36) and Beta(23, 30).

iii.

```
posterior <- theta_pdf * choose(43, 15) * theta ^ 15 * (1 - theta) ^ 28
plot(theta, posterior, type = "l")
```



```
cat("Mode:", theta[which.max(posterior)])
```

```
## Mode: 0.3143143
```

We can calculate the modes of Beta(a,b) following  $\frac{a-1}{a+b-2}$

$$\text{mode}(\text{Beta}(17, 36)) = (17 - 1)/(17 + 36 - 2) = 0.313$$

$$\text{mode}(\text{Beta}(23, 30)) = (23 - 1)/(23 + 30 - 2) = 0.431$$

So the mode in this question is between these two values, although much closer to the mode of Beta(17, 36).

(e) The general formula is

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} [x\theta^{a-1}(1-\theta)^{b-1} + y\theta^{b-1}(1-\theta)^{a-1}]$$

$$p(\theta|y_i) \propto xp_1(\theta) \times p(y_i|\theta) + yp_2(\theta) \times p(y_i|\theta)$$

$$\text{where } x + y = 1 \text{ and } p_1(\theta) \sim \text{Beta}(a, b), p_2(\theta) \sim \text{Beta}(b, a)$$

Interpretation: as long as we can distinguish the pdf of  $\theta$  as the above form, we can simplify its posterior into a mixture of two Beta distribution, with weights specified above.

## 4. PH 3.7

(a) First, a uniform prior means a Beta(1,1) prior.

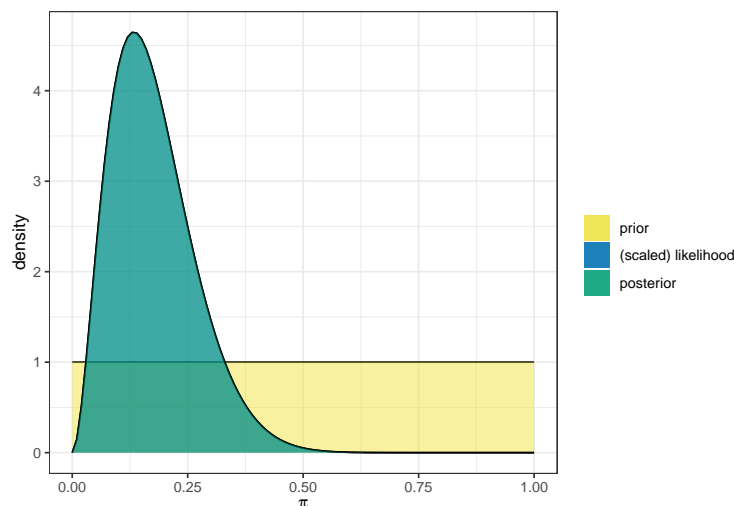
By class notes, we could identify the posterior as a Beta distribution as well, which is Beta(1+2, 1+15-2) = Beta(3,14)

$$\text{mean}(\text{Beta}(3, 14)) = \frac{3}{17}$$

$$\text{mode}(\text{Beta}(3, 14)) = \frac{a-1}{a+b-2} = \frac{2}{15} = 0.133$$

$$\text{sd}(\text{Beta}(3, 14)) = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}} = \sqrt{(3*14)/[3+14]^2(3+14+1)} = 0.089854$$

```
plot_beta_binomial(1,1,2,15) + theme_bw()
```





- (b) i. In order for the following to be true:

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

We are assuming that the 15 children in the pilot study are representative and close enough to the 278 children in the latter long-term study, alternatively speaking, the pilot and long-term studies are taking samples from the same population.

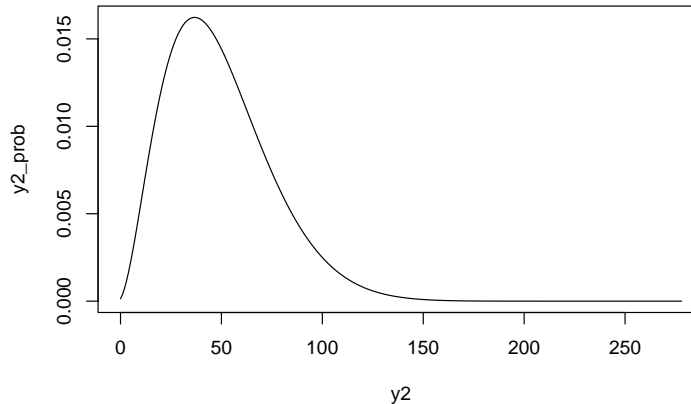
- ii. From part (a), we know that  $p(\theta | Y_1 = 2) \sim \text{Beta}(3, 14)$  So we plug things in

$$\begin{aligned} Pr(Y_2 = y_2 | Y_1 = 2) &= \int_0^1 Pr(Y_2 = y_2 | \theta) p(\theta | Y_1 = 2) d\theta \\ &= \int_0^1 \binom{278}{y_2} \theta^{y_2} (1 - \theta)^{278 - y_2} \frac{\Gamma(17)}{\Gamma(3)\Gamma(14)} \theta^2 (1 - \theta)^{13} d\theta \\ &= \frac{\Gamma(17)}{\Gamma(3)\Gamma(14)} \binom{278}{y_2} \int_0^1 \theta^{y_2 + 2} (1 - \theta)^{278 - y_2 + 13} d\theta \\ &= \frac{\Gamma(17)}{\Gamma(3)\Gamma(14)} \binom{278}{y_2} \int_0^1 \theta^{y_2 + 2} (1 - \theta)^{291 - y_2} d\theta \\ &= \frac{\Gamma(17)}{\Gamma(3)\Gamma(14)} \binom{278}{y_2} \frac{\Gamma(y_2 + 3) \Gamma(292 - y_2)}{\Gamma(295)} \\ &= \frac{16!}{2!3!} \frac{278!}{y_2! (278 - y_2)!} \frac{(y_2 + 2)! (291 - y_2)!}{294!} \\ &= \frac{1}{\binom{294}{278}} \binom{y_2 + 2}{2} \binom{291 - y_2}{13} \end{aligned}$$

iii.

(c) The plot is shown below

```
y2 <- seq(0, 278, by = 1)
y2_prob <- 1/choose(294, 278) * choose(y2+2, 2) * choose(291-y2, 13)
plot(y2, y2_prob, type = "l")
```



```
data.frame(y2,y2_prob) %>%
  mutate(expected = y2*y2_prob) %>%
  summarize(mean = sum(expected))
```

```
##           mean
## 1 49.05882
```

```
sqrt((y2 - rep(49.05882,length(y2)))^2 %*% y2_prob)
```

```
##           [,1]
## [1,] 25.73196
```

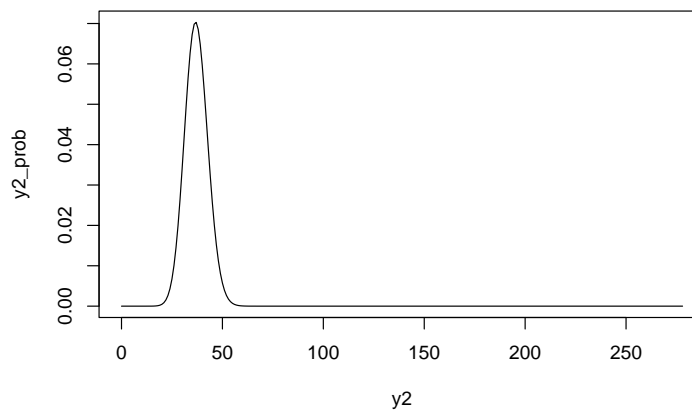
Mean is 49.05882 and SD is 25.73196.

(d) We now have

$$Pr(Y_2 = y_2 | \theta = \hat{\theta} = 2/15) = \binom{278}{y_2} \left(\frac{2}{15}\right)^{y_2} \left(1 - \frac{2}{15}\right)^{278-y_2}$$

which can be drawn as

```
y2_prob <- choose(278, y2) * (2/15)^{y2} * (1-2/15)^{278-y2}
plot(y2, y2_prob, type = "l")
```



This is a Binomial distribution of  $\text{Binomial}(\frac{2}{15}, 278)$ , so its mean is 37.067, SD is  $\sqrt{np(1-p)} = 5.67$ . Although the mean does not move too much, the distribution is now much narrower and tighter.

## 5. PH 3.10

- (a) Given  $\psi = g(\theta) = \log \frac{\theta}{1-\theta}$ , so we can calculate  $\theta = h(\psi) = \frac{e^\psi}{1+e^\psi}$ . Then we derive  $p_\psi(\psi)$  based on the formula given:

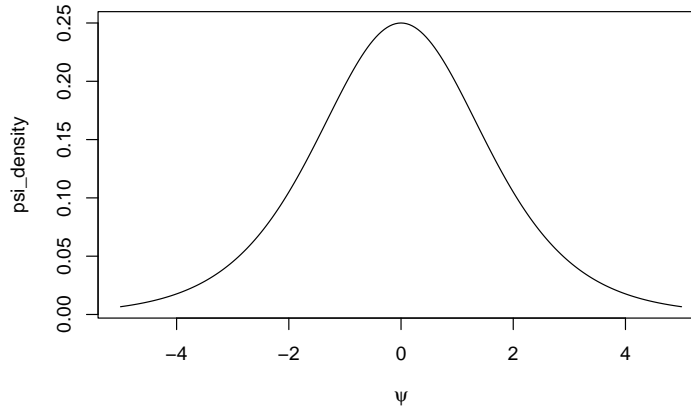
$$\begin{aligned}
 p_\psi(\psi) &= p_\theta(h(\psi)) \times \left| \frac{dh}{d\psi} \right| \\
 &= \left[ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} (h(\psi))^{a-1} (1-h(\psi))^{b-1} \right] \times \left| \frac{dh}{d\psi} \right| \\
 &= \left[ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left( \frac{e^\psi}{1+e^\psi} \right)^{a-1} \left( 1 - \frac{e^\psi}{1+e^\psi} \right)^{b-1} \right] \times \left| \frac{e^\psi}{(e^\psi + 1)^2} \right| \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[ \left( \frac{e^\psi}{1+e^\psi} \right)^{a-1} \left( \frac{1+e^\psi - e^\psi}{1+e^\psi} \right)^{b-1} \right] \times \frac{e^\psi}{(e^\psi + 1)^2} \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[ \left( \frac{e^\psi}{1+e^\psi} \right)^a \left( \frac{1+e^\psi}{e^\psi} \right) \left( \frac{1}{1+e^\psi} \right)^b \left( \frac{1+e^\psi}{1} \right) \right] \times \frac{e^\psi}{(e^\psi + 1)^2} \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[ \left( \frac{e^\psi}{1+e^\psi} \right)^a \left( \frac{1}{1+e^\psi} \right)^b \frac{(e^\psi + 1)^2}{e^\psi} \right] \times \frac{e^\psi}{(e^\psi + 1)^2} \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left( \frac{e^\psi}{1+e^\psi} \right)^a \left( \frac{1}{1+e^\psi} \right)^b
 \end{aligned}$$

The density plot is shown below

```

psi <- seq(-5, 5, by = 0.05)
psi_density <- gamma(2)/gamma(1)/gamma(1) *(exp(psi)/(1+exp(psi)))*(1/(1+exp(psi)))
plot(psi, psi_density, type = "l", xlab = expression(psi)) # set a and b to 1

```



(b) Given  $\psi = g(\theta) = \log \theta$ , we have  $\theta = h(\psi) = e^\psi$ . Then we derive  $p_\psi(\psi)$  based on the formula:

$$\begin{aligned} p_\psi(\psi) &= p_\theta(h(\psi)) \times \left| \frac{dh}{d\psi} \right| \\ &= \left[ \frac{b^a}{\Gamma(a)} (e^\psi)^{a-1} e^{-be^\psi} \right] \times e^\psi \\ &= \frac{b^a}{\Gamma(a)} \exp(a\psi - \psi - be^\psi + \psi) \\ &= \frac{b^a}{\Gamma(a)} \exp(a\psi - be^\psi) \end{aligned}$$

```
psi_density <- 1/gamma(1) * exp(1 * psi - 1*exp(psi)) # set a and b to 1
plot(psi, psi_density, type = "l", xlab = expression(psi))
```

