STAT 457: Homework No. 4

Tyler Maule

February 4th, 2022

Exercises without Code

Exercise 1

Let $Y_1, \ldots, Y_n$ be an *iid* sample from $Poisson(\lambda)$. Derive the Jeffrey's (noninformative prior). This prior corresponds to a gamma distribution with which parameters?

STAT 457
HW. #4

(1) $y_i \overset{iid}{\sim} \text{Pois}(\lambda)$, $i = 1, \dots, n$. Derive Jeffrey's Prior.

$$\pi_J(\theta) \propto I(\theta)^{1/2} \quad ; \quad I(\theta) = -E_\theta\left[\frac{d^2 \ell(\theta|\vec{x})}{d\theta^2}\right]$$

$$f_X(x|\lambda) = [\lambda^x \exp(-\lambda)]/x!$$

$$\Rightarrow L(\lambda|\vec{x}) = [\lambda^{\sum x_i} \exp(-n\lambda)] \cdot [\prod_{i=1}^n x_i!]^{-1}$$

$$\Rightarrow \ell(\lambda|\vec{x}) = (\textstyle\sum x_i) \log(\lambda) - n\lambda - \sum_{i=1}^n \log(x_i!)$$

$$\Rightarrow \frac{d}{\cancel{d\theta}\, d\lambda} \ell = (\textstyle\sum_i x_i/\lambda) - n$$

$$\Rightarrow \frac{d^2}{d\lambda^2} = -\lambda^{-2} \textstyle\sum_i x_i \Rightarrow I(\theta) = -E_\theta[-\lambda^{-2}\sum_i x_i]$$

$$I(\theta) = +\lambda^{-2} E[\textstyle\sum_i x_i] = \lambda^{-2} \sum_i E[x_i] = \lambda^{-2} \sum_i \lambda$$
$$= n/\lambda$$

so $\pi_J(\lambda) \propto (n/\lambda)^{1/2} \propto \lambda^{-1/2}$

In this case, Jeffrey's Prior is distributed as Gamma$(1/2, 0)$; $\alpha = 1/2$, $\beta = 0$

Figure 1: Exercise 1

Exercise 2

In the multivariate setting, $\overrightarrow{\theta} = (\theta_1, \ldots, \theta_d)^T$, $p(\theta) \propto |J(\theta)|^{(1/2)}$, provides an invariant prior, where the *ij*th entry of $J(\theta)$ equals $-E[\frac{\partial^2 l(\theta|Y)}{\partial\theta_i \partial\theta_j}]$ and $|X|$ is the determinant of the matrix X.

Let $Y_1, \ldots, Y_n$ be an *iid* sample from $N(\mu, \sigma^2)$, where both parameters are unknown. Derive the invariant prior. How does it compare with the prior $p(\theta) \propto (1/\sigma^2)$?

(2) $f(\vec{x} \mid \vec{\theta}) = (\sigma^2 2\pi)^{-n/2} \exp\left[(-1/2)(\sigma^{-2})\Sigma(x_i-\mu)^2\right]$

$\ell(\vec{\theta} \mid \vec{x}) = (-n/2)\log(\sigma^2 2\pi) + (-n/2)\log(\sigma^2) + (2\sigma^2)^{-1}\sum_i (x_i-\mu)^2$

$\propto (-n/2)\ln(\sigma^2) - (2\sigma^2)^{-1}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]$

$$\frac{\partial}{\partial\mu}\ell = (-2)(2\sigma^2)^{-1}\sum_i (x_i-\mu) = -\sigma^{-2}\left(\sum_i (x_i) - n\mu\right)$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$\frac{\partial^2}{\partial\mu^2}\ell = -(n/\sigma^2) \qquad \frac{\partial^2}{\partial\mu\,\partial\sigma^2}\ell = -\sigma^{-4}\left(\sum_i (x_i) - n\mu\right)$$

$$\frac{\partial}{\partial\sigma^2}\ell = (-n/2)\sigma^{+2^{-1}} + (n/2)\sigma^{+2^{-1}} - (1/2)(\sigma)^{+2^{-2}}\sum_i (x_i-\mu)^2$$

$$\Downarrow$$

$$\frac{\partial^2}{\partial\sigma^2{}^2} = (n/2)\sigma^{-4} - (\sigma)^{-6}\left[\sum_i (x_i-\mu)^2 + s^2(n-1)\right]$$

$$= \left(n + 4\left(\sum_i x_i - \mu\right)^2\right)\sigma^{-4}$$

$$\frac{\partial}{\partial\sigma^2\,\partial\mu} = -\sigma^{-4}\left(\sum_i (x_i) - n\mu\right)$$

E

So we have: $E[-n/\sigma^2] = -n/\sigma^2$

$$E\left[-\sigma^{-4}\left(\sum_i (x_i) - n\mu\right)\right] = n\mu\sigma^{-4} - \sigma^{-4} E\left[\sum_i X_i\right]$$

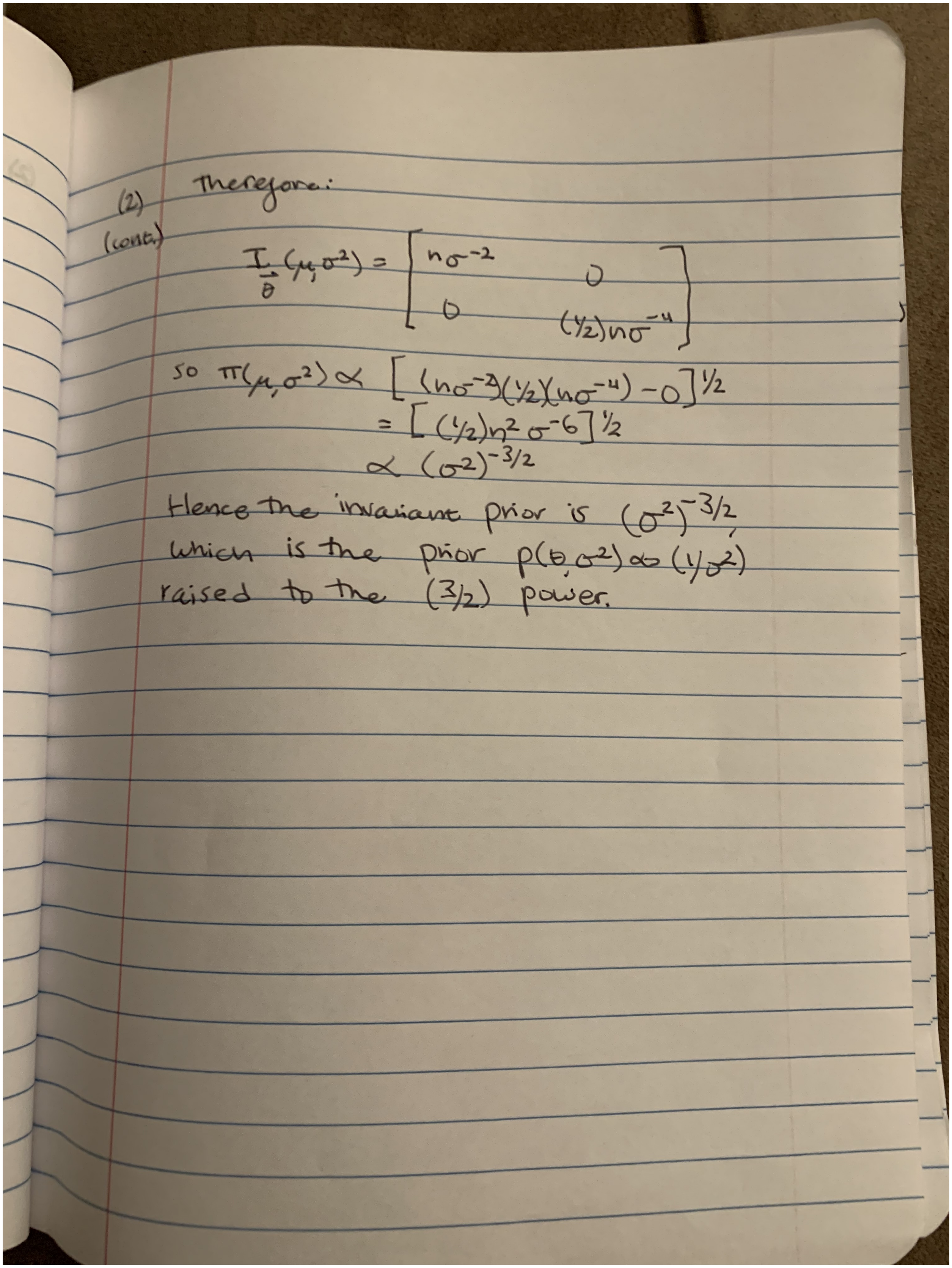
$$= n\mu\sigma^{-4} - n\mu\sigma^{-4} = 0$$

$$E\left[(n/2)\sigma^{-4} + 4\left(\sum_i (x_i-\mu)^2 + s^2(n-1)\right)\sigma^{-6}\right] = n/(2\sigma^4) + E\left[\sigma^{-6}\left((n-1)s^2 + (n-1)\sigma^2\right)\right]$$

$$= \left[n/(2\sigma^4) + 0\right]$$

$$= -n/(2\sigma^4)$$

Figure 2: Exercise 2 (1 of 2)



(2) (cont.) Therefore:

$$I_{\vec{\theta}}(\mu, \sigma^2) = \begin{bmatrix} n\sigma^{-2} & 0 \\ 0 & (1/2)n\sigma^{-4} \end{bmatrix}$$

so $\pi(\mu, \sigma^2) \propto \left[(n\sigma^{-2})(1/2)(n\sigma^{-4}) - 0\right]^{1/2}$

$= \left[(1/2)n^2\sigma^{-6}\right]^{1/2}$

$\propto (\sigma^2)^{-3/2}$

Hence the invariant prior is $(\sigma^2)^{-3/2}$, which is the prior $p(\theta, \sigma^2) \propto (1/\sigma^2)$ raised to the $(3/2)$ power.

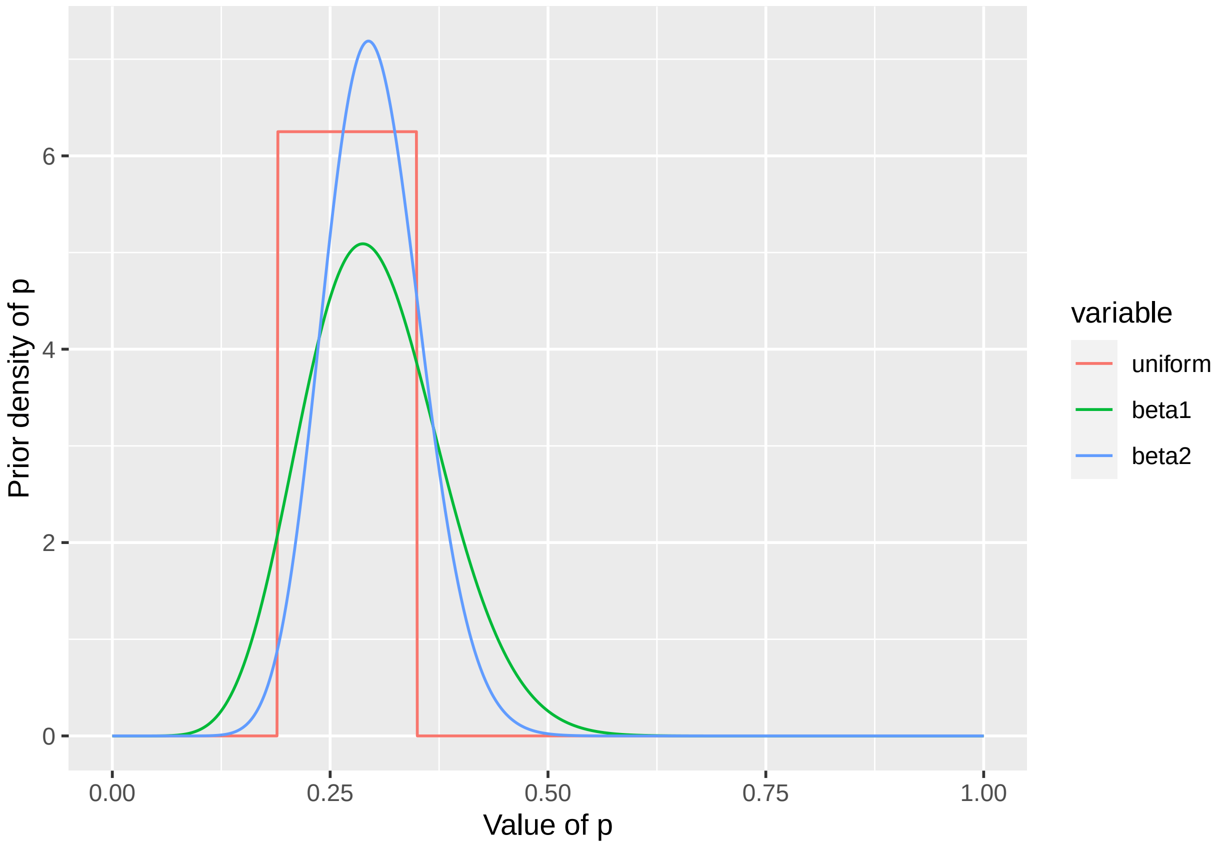
Figure 3: Exercise 2 (2 of 2)

Exercise 3

Let p denote the probability a specific major league baseball player gets a hit in a particular at-bat. Assume that batting averages usually fall in the range 0.19 to 0.35.

Exercise 3a

Consider priors $Unif(0.19, 0.35)$, $Beta(10.2, 23.8)$, and $Beta(20.4, 47.6)$. Plot these priors and discuss each choice.



The choice of a uniform distribution is unique in that it disallows values of p outside of the range 0.19 to 0.35, so in that sense this choice of a prior is quite informative. Since batting averages *usually* but don't always fit within that range, I believe it would be preferable to allow p to take values outside of it, as the two beta priors do. While both priors have a mode around 0.3, the second Beta prior places considerably more confidence in the "best guess" (mode) of 0.3 (it has lighter tails than the first choice of Beta priors). Personally, I'd be drawn to the use of the first Beta prior, simply because it (1) allows for a greater range of batting averages, and (2) places less weight on particular values. It's likely to be the most flexible in allowing a range of batting data to have a solid impact on the posterior.

Exercise 3b

Suppose a player gets 5 hits in 40 at-bats. For each of the above priors: plot likelihood, posterior, and prior; compute the probability that the player is better than a 0.200 hitter; compute your best guess as to the batting average of the player; compute a 95% credible interval for p.

Here are the derivations for posteriors in (3b):

(3b) likelihood: let y = # of hits in 40 at-bats

$y \sim \text{binomial}(40, p) \rightarrow p_{y|p}(y|p) = \binom{40}{y} p^{y} (1-p)^{40-y}$

(i) $\pi(p) \sim \text{Unif}(0.19, 0.35) \Rightarrow \pi(p) = 1/(0.35-0.19) = (1/0.16)$

$$p(p|y) = \frac{p(y|p)\pi(p)}{p(y)} = \frac{p(y|p)\pi(p)}{\int_{\Theta} p(y|p)\pi(p)\,dp}$$

$$= \frac{(1/0.16)\binom{40}{y} p^{y}(1-p)^{40-y}}{\int_{\Theta}\left(\frac{1}{0.016}\right)\binom{40}{y} p^{y}(1-p)^{40-y}\,dp}$$

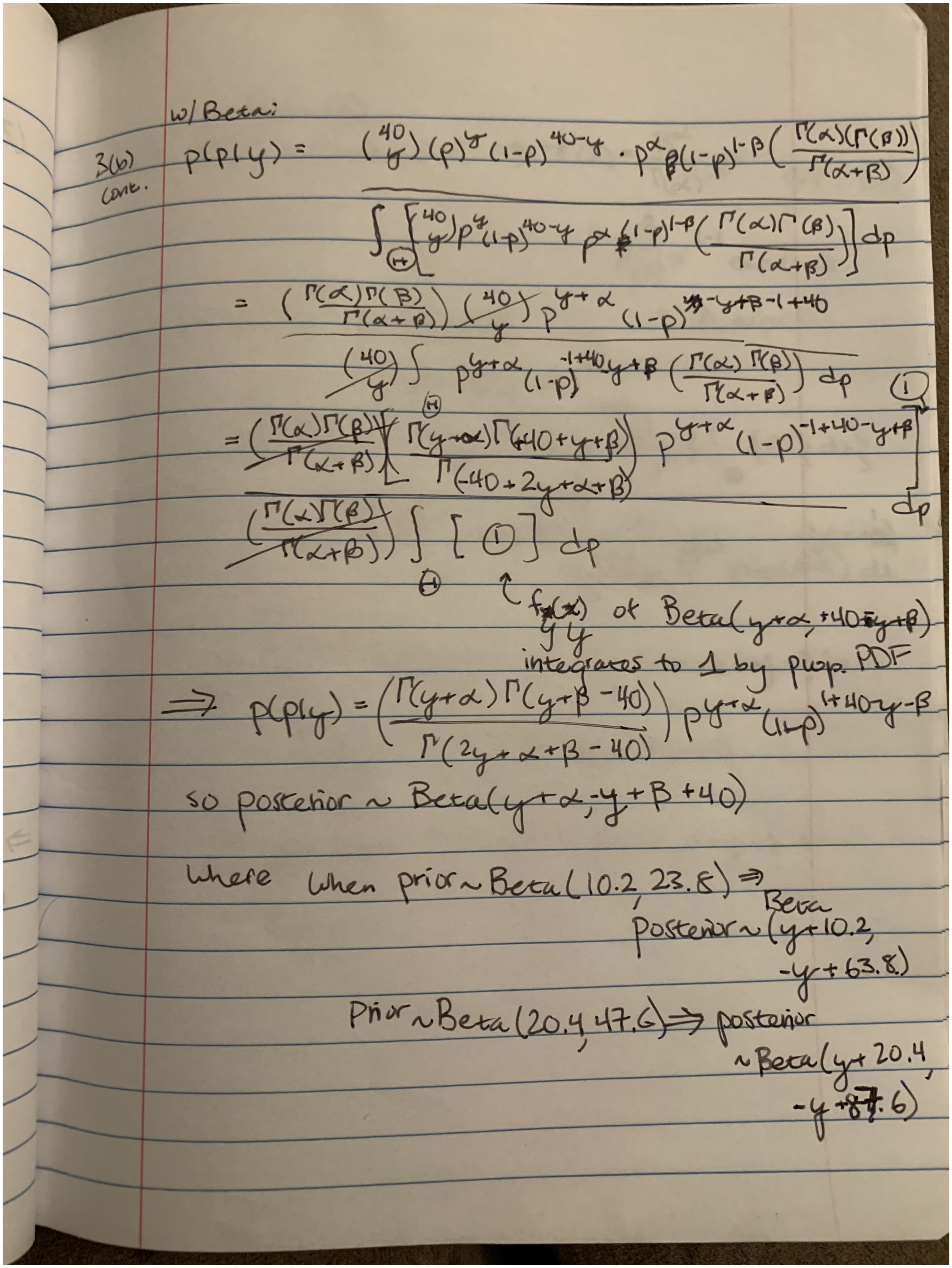
$$= \frac{\cancel{(1/0.16)}\binom{40}{y}\left[p^{y}(1-p)^{40-y}\right]}{\cancel{(1/0.16)}\int_{\Theta}\binom{40}{y} p^{y}(1-p)^{40-y}\,dp} = \frac{\binom{40}{y}\left[p^{y}(1-p)^{40-y}\right]}{1}$$

$$\Rightarrow p(p|y) = p(y|p) = \binom{40}{y}\left[p^{y}(1-p)^{40-y}\right]$$

$\Rightarrow$ reparameterize as $\text{Beta}(5+1, 40-5+1) = \text{Beta}(6, 36)$

(ii) $\pi(p) \sim \text{Beta}(10.3, 23.8) \Rightarrow \pi(p) = \dfrac{p^{\alpha-1}(1-p)^{\beta-1}}{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}$

Figure 4: Exercise 3 (1 of 2)



w/ Beta:

3(b) cont.

$$p(p|y) = \frac{\binom{40}{y}(p)^y(1-p)^{40-y} \cdot p^{\alpha}\cancel{\beta}(1-p)^{1-\beta}\left(\frac{\Gamma(\alpha)(\Gamma(\beta))}{\Gamma(\alpha+\beta)}\right)}{\int_{\Theta}\left[\binom{40}{y}p^y(1-p)^{40-y}p^{\alpha}\cancel{\beta}(1-p)^{1-\beta}\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)\right]dp}$$

$$= \frac{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)\cancel{\binom{40}{y}}\, p^{y+\alpha}(1-p)^{\cancel{40}-y+\beta-1+40}}{\cancel{\binom{40}{y}}\int_{\Theta} p^{y+\alpha}(1-p)^{-1+40-y+\beta}\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)dp} \quad (1)$$

$$= \frac{\cancel{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}\left[\left(\frac{\Gamma(y+\alpha)\Gamma(-40+y+\beta)}{\Gamma(-40+2y+\alpha+\beta)}\right)p^{y+\alpha}(1-p)^{-1+40-y+\beta}\right]dp}{\cancel{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}\int_{\Theta}\left[(1)\right]dp}$$

$f_{y|y}(x)$ of Beta$(y+\alpha, +40-y+\beta)$ integrates to 1 by prop. PDF

$$\Rightarrow p(p|y) = \left(\frac{\Gamma(y+\alpha)\Gamma(y+\beta-40)}{\Gamma(2y+\alpha+\beta-40)}\right)p^{y+\alpha}(1-p)^{1+40-y-\beta}$$

so posterior $\sim$ Beta$(y+\alpha, -y+\beta+40)$

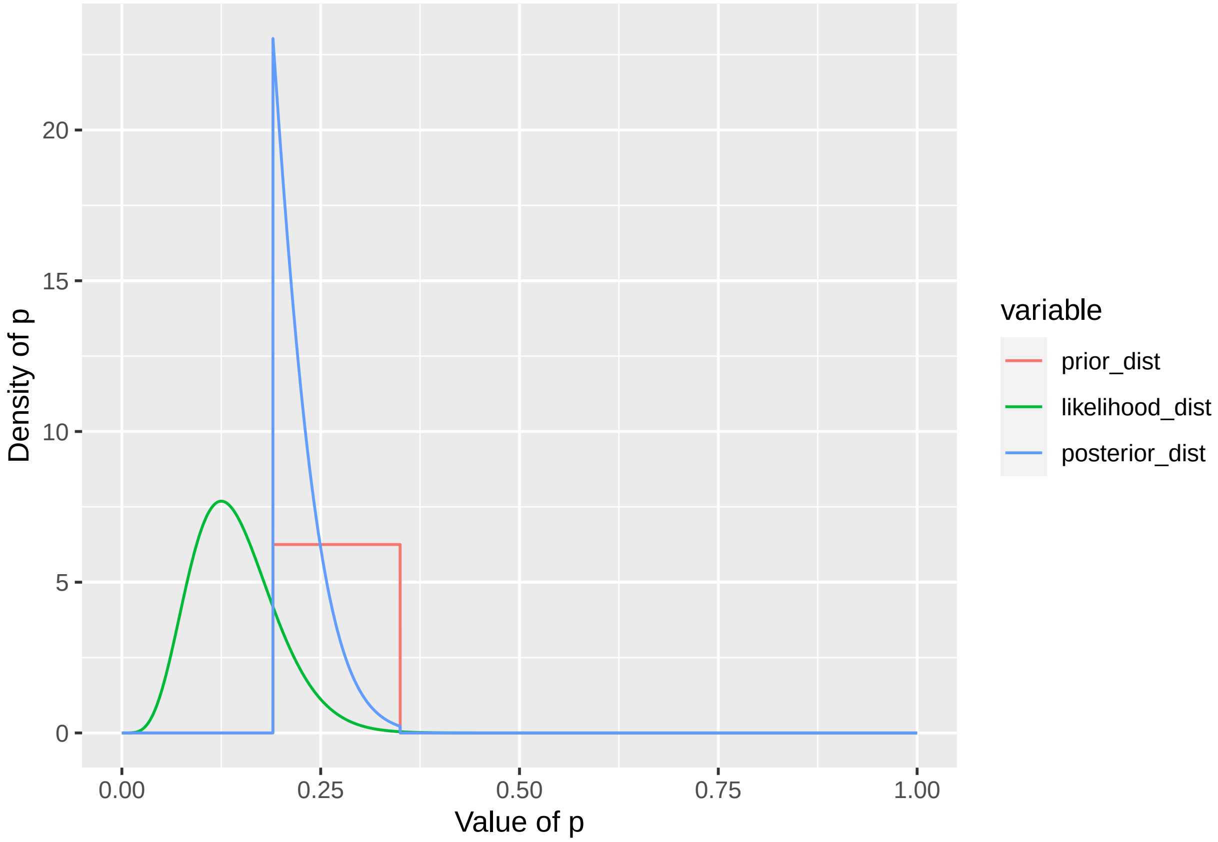
~~Where~~ When prior $\sim$ Beta$(10.2, 23.8) \Rightarrow$ posterior $\sim$ Beta$(y+10.2, -y+63.8)$

Prior $\sim$ Beta$(20.4, 47.6) \Rightarrow$ posterior $\sim$ Beta$(y+20.4, -y+87.6)$

Figure 5: Exercise 3 (2 of 2)

Results for $Unif(0.19, 0.35)$:

Likelihood is binomial, $Y \sim bin(n = 40, p)$ with Y being the number of hits in 40 at-bats. Prior is given as uniform, and posterior is calculated as $Beta(6, 36)$ (see derivations above).




```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.1900       6.25        4.185609       23.03652
## 2 0.1901       6.25        4.178540       22.99761
## 3 0.1902       6.25        4.171475       22.95873
```

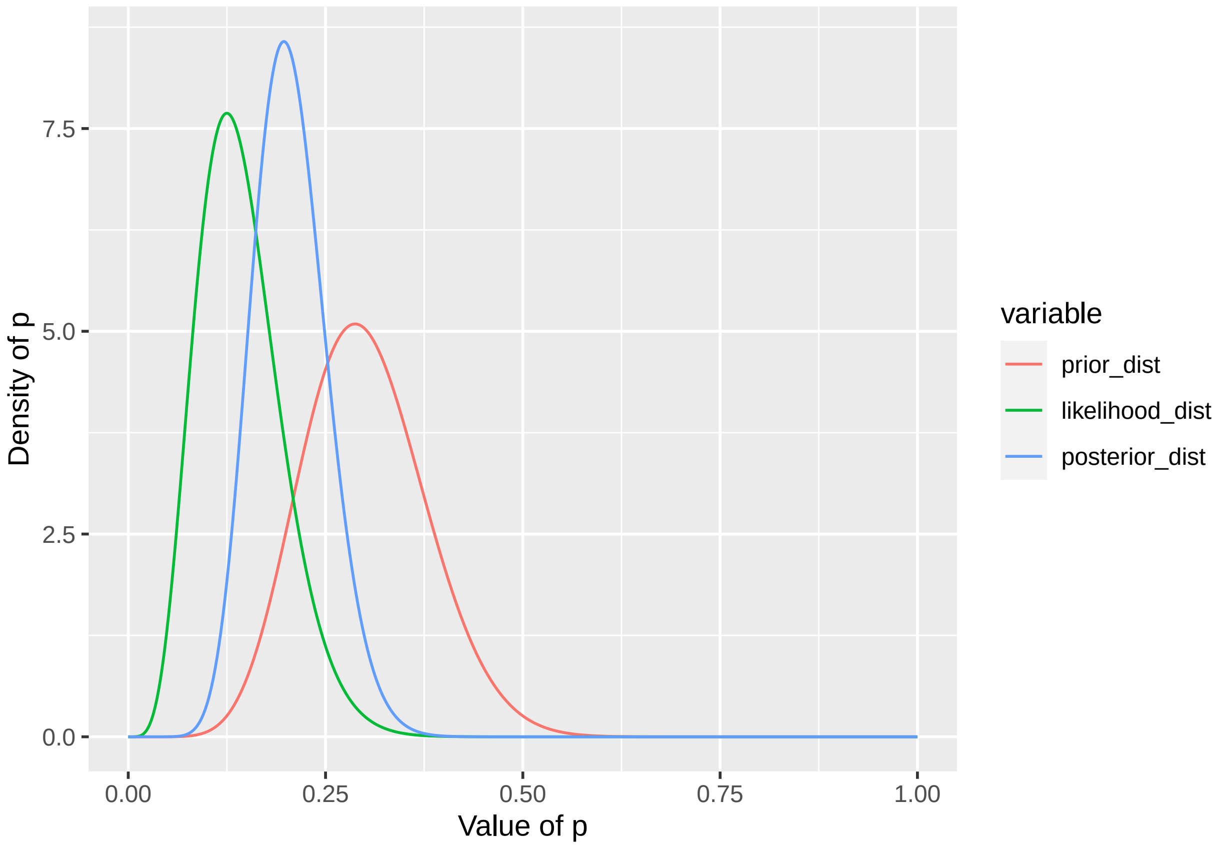
```
## [1] 0.7886796
```

```
## $unif_posterior_ci
##      2.5%     97.5%
## 0.1910947 0.3056695
```

My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.1900. The probability that the batter is better than a 0.200 hitter is 0.7886796. Finally, a 95% credible interval for the true value of p is [0.1910947, 0.3056695].

Results for $Beta(10.2, 23.8)$:

See above the derivation of the posterior distribution for p with prior distribution $Beta(10.2, 23.8)$ and the binomial likelihood. This posterior follows $p \sim Beta(y + 10.2, 23.8 + n - y) = Beta(15.2, 58.8)$.




```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.1972   2.409780        3.688139       8.572733
## 2 0.1973   2.414179        3.681414       8.572722
## 3 0.1971   2.405382        3.694870       8.572705
```

```
## [1] 0.5234115
```

```
## $beta_i_posterior_ci
##      2.5%     97.5%
## 0.1224982 0.3052243
```

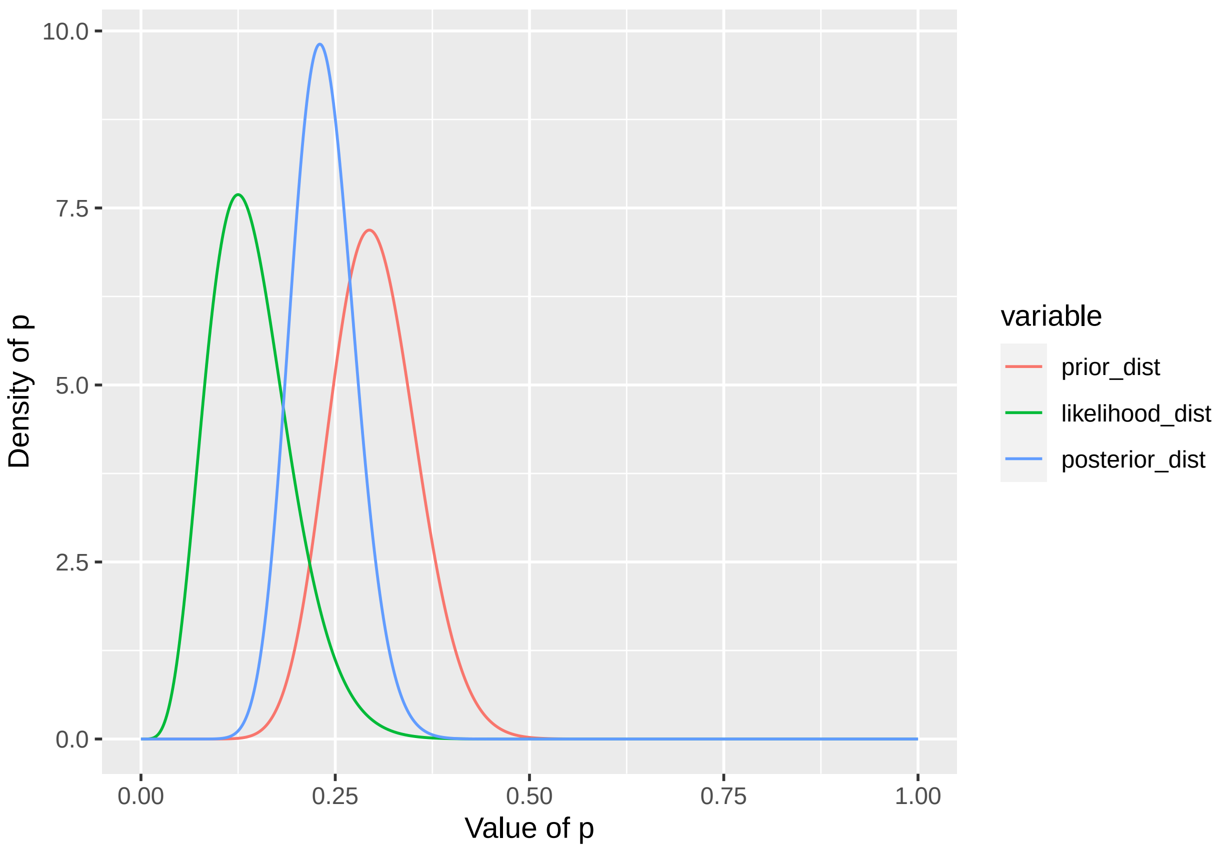
My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.1972. The probability that the batter is better than a 0.200 hitter is 0.5234115. Finally, a 95% credible interval for the true value of p is [0.1224982, 0.3052243].

Results for $Beta(20.4, 47.6)$:

The last prior yields a similar result (both beta). See above the derivation of the posterior distribution for p with prior distribution $Beta(20.4, 47.6)$ and the binomial likelihood. This posterior follows $p \sim Beta(y + 20.4, 47.6 + n - y) = Beta(25.4, 82.6)$.

Table 1: Results for Each Prior Distribution

Prior	Best_Guess	Prob_Greater_Than_1in5	CI_95_Lower_Bound	CI_95_Upper_Bound
Unif(0.19,0.35)	0.1900	0.7886796	0.1910947	0.3056695
Beta(10.2,23.8)	0.1972	0.5234115	0.1224982	0.3052243
Beta(20.4,47.6)	0.2302	0.8031853	0.1604733	0.3187223




```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.2302   3.518726        1.839790       9.813445
## 2 0.2301   3.510375        1.844163       9.813423
## 3 0.2303   3.527082        1.835425       9.813409
```

```
## [1] 0.8031853
```

```
## $beta_ii_posterior_ci
##      2.5%     97.5%
## 0.1604733 0.3187223
```

My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.2302. The probability that the batter is better than a 0.200 hitter is 0.8031853. Finally, a 95% credible interval for the true value of p is [0.1604733, 0.3187223].

Exercise 4

Data represents the number of arrivals for 45 minute time intervals of length 2 minutes at a cashier's desk at a supermarket:

```
## $arrivals_df
```

```
##  [1] 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3
## [39] 3 3 4 4 4 4 5
##
## $arrivals_mean
## [1] 1.733333
##
## $arrivals_sum
## [1] 78
```

Exercise 4a

For a $Gamma(2, 1)$ prior, obtain the posterior distribution under a $Poisson(\lambda)$ model for the data. Draw the prior and the posterior. Note on your plot the mean, variance, and mode of the posterior.

Here is the derivation of the posterior for (4a):

$y \sim \text{Poisson}(\lambda)$

(4a) $\pi(\lambda) \sim \text{Gamma}(2,1) \Rightarrow \pi(\lambda) = \frac{B^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-Bx)$

$$= \frac{(1)^2}{\Gamma(2)} x^{2-1} \exp(-x)$$

$$= \lambda \exp(-\lambda)$$

$$p(\lambda|y) = \frac{p(\vec{y}|\lambda)p(\lambda)}{\int p(y|\lambda)\pi(\lambda)\, d\lambda}$$

$$= \left\{ \left[\lambda^{\bar{y}n} \exp(-\lambda)\right](y!)^{-1} \lambda \exp(-n\lambda) \right\} \Big/ \left\{ \int_{+} \{1\}\, d\lambda \right\}$$

$$= \frac{\cancel{(y!)^{-1}} \left[\lambda^{n\bar{y}+1} \exp(-\lambda(n+1))\right]}{\cancel{(y!)^{-1}} \int_{+} \lambda^{n\bar{y}+1} \exp(-\lambda(n+1))\, d\lambda}$$

$$\propto \left[\lambda^{n\bar{y}+1} \exp(-\lambda(n+1))\right] \Big/ \int_{+} \text{Gamma}(n\bar{y}+2, 1+n)$$

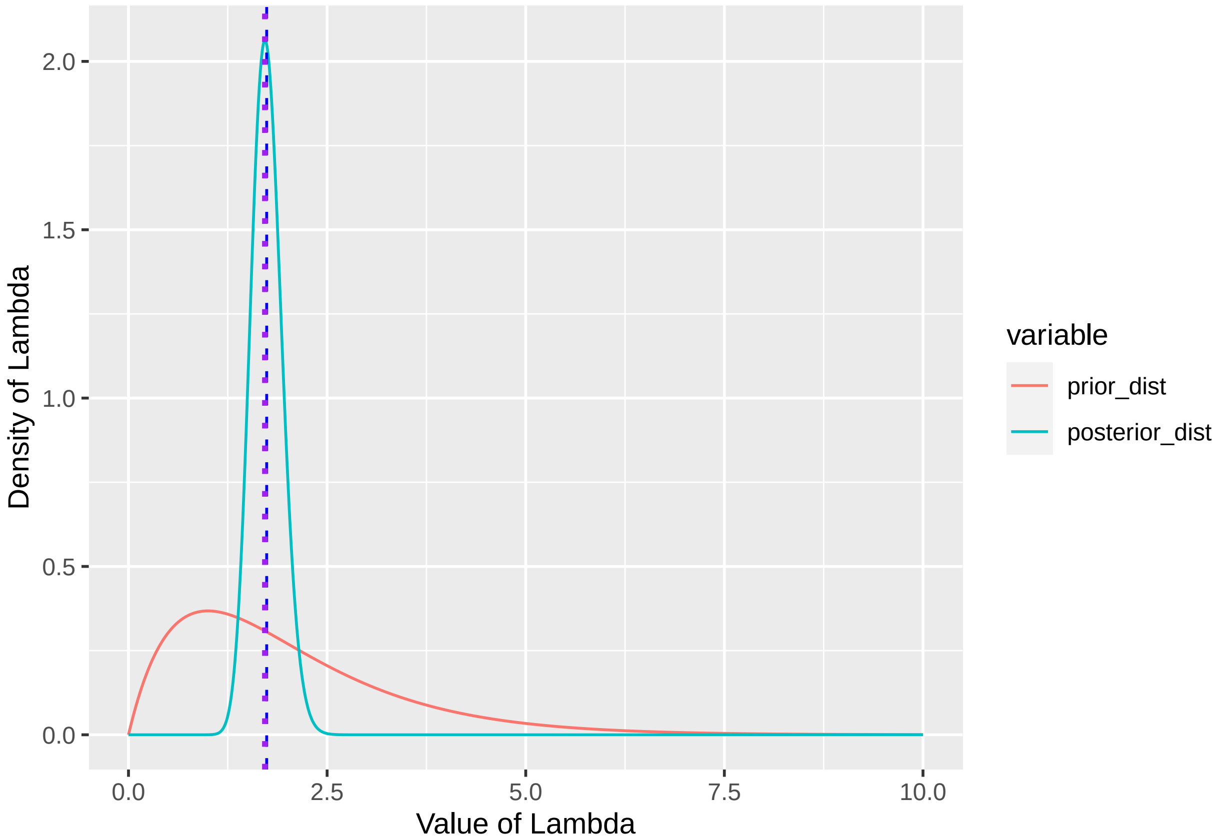
$$= [1]/1 = \lambda^{n\bar{y}+1} \exp(-\lambda(n+1))$$

So posterior $\pi(\lambda|\vec{y}) \sim \text{Gamma}(n\bar{y}+2, 1+n)$

Figure 6: Exercise 4

As pictured above, the derived posterior is $\pi(\lambda|Y) \sim Gamma(n\bar{y}+2, n+1) = Gamma(80, 46)$.

```
## $post_mean
## [1] 1.73913
##
## $post_mode
## [1] 1.717
##
## $post_var
## [1] 0.03780718
```



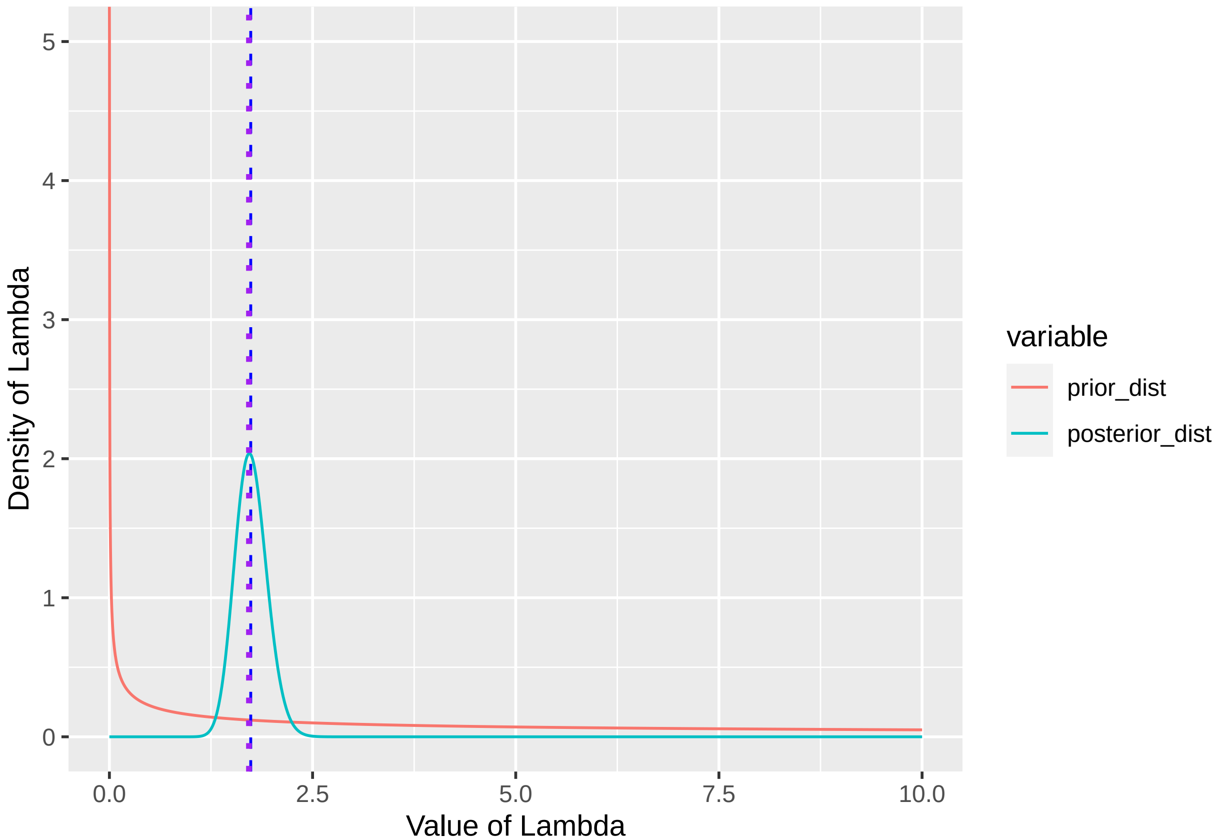
See above a plot of the posterior. Its mean is 1.73913, its mode 1.717, and its variance is 0.03780718.

Exercise 4b

For the noninformative prior, i.e. a $Gamma(?, ?)$, repeat (4a).

We found in exercise (1) that the noninformative (Jeffrey's) prior for a Poisson distribution is $\lambda \sim Gamma(0.5, 0)$. Then generalizing from (4a), we know that $\pi(\lambda|Y) \sim Gamma(n\bar{y}+\alpha, n+\beta)$, so the posterior in the case of (4b) is $\pi(\lambda|Y) \sim Gamma(78.5, 45)$

```
## $post_mean_b
## [1] 1.744444
##
## $post_mode_b
## [1] 1.722
##
## $post_var_b
## [1] 0.03876543
```



See above the plot of the posterior based on the noninformative prior. Here, the mean is 1.744444, the mode is 1.722, and the variance is 0.03876543.

Exercise 5

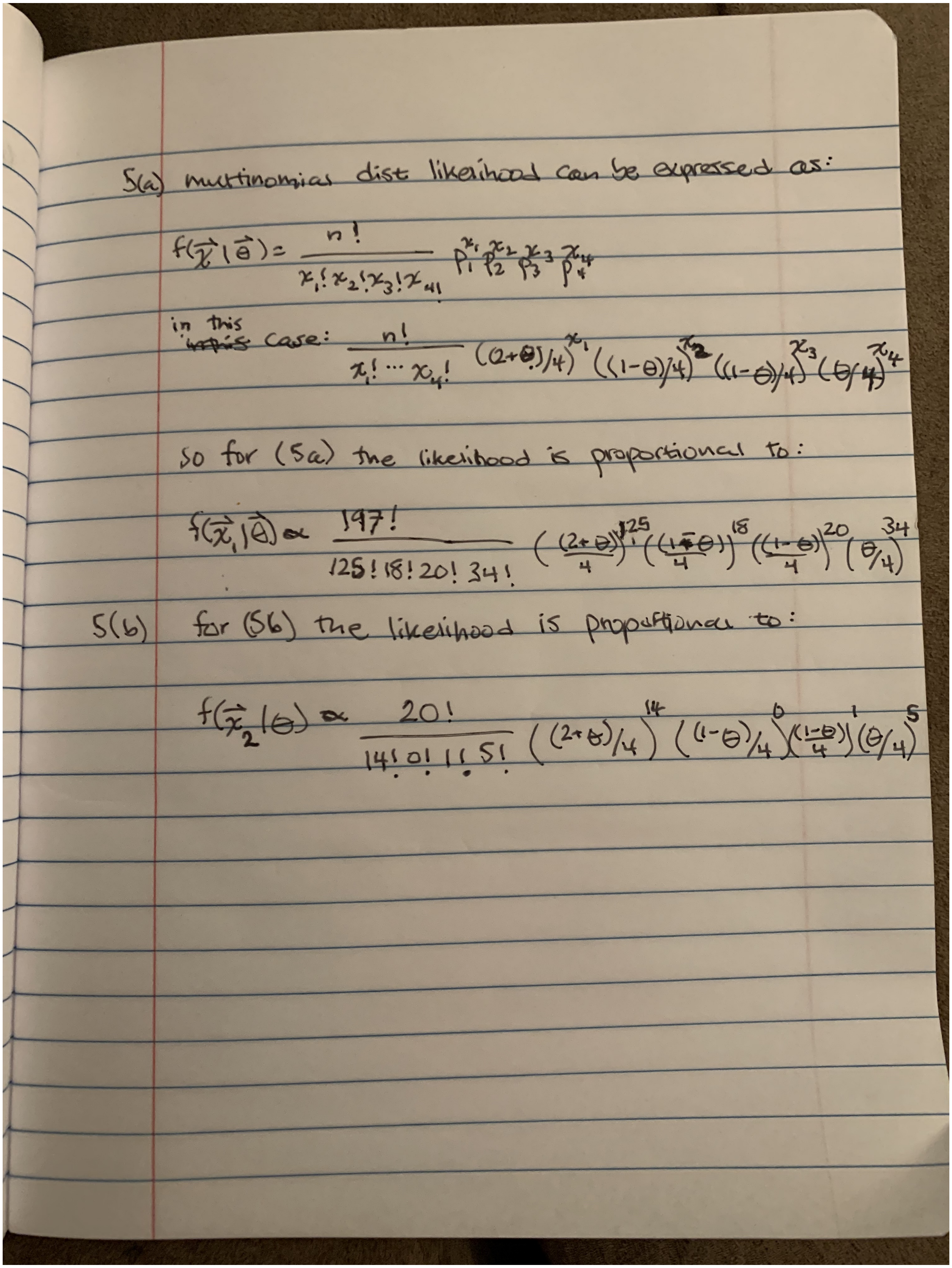
197 animals are distributed into four categories: $Y = (y_1, y_2, y_3, y_4)$ according to the genetic linkage model $((2+\theta)/4, (1-\theta)/4, (1-\theta)/4, \theta/4)$

Exercise 5a

What is the likelihood for the data $Y = (125, 18, 20, 34)$?

```
## $Y_data
## [1] 125  18  20  34
##
## $Y_num_trials
## [1] 197
```

See below the derivations of the appropriate likelihood functions for both (5a) and (5b):



5(a) multinomial dist likelihood can be expressed as:

$$f(\vec{x}|\vec{\theta}) = \frac{n!}{x_1!x_2!x_3!x_4!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

in this case: $$\frac{n!}{x_1! \cdots x_4!} ((2+\theta)/4)^{x_1} ((1-\theta)/4)^{x_2} ((1-\theta)/4)^{x_3} (\theta/4)^{x_4}$$

so for (5a) the likelihood is proportional to:

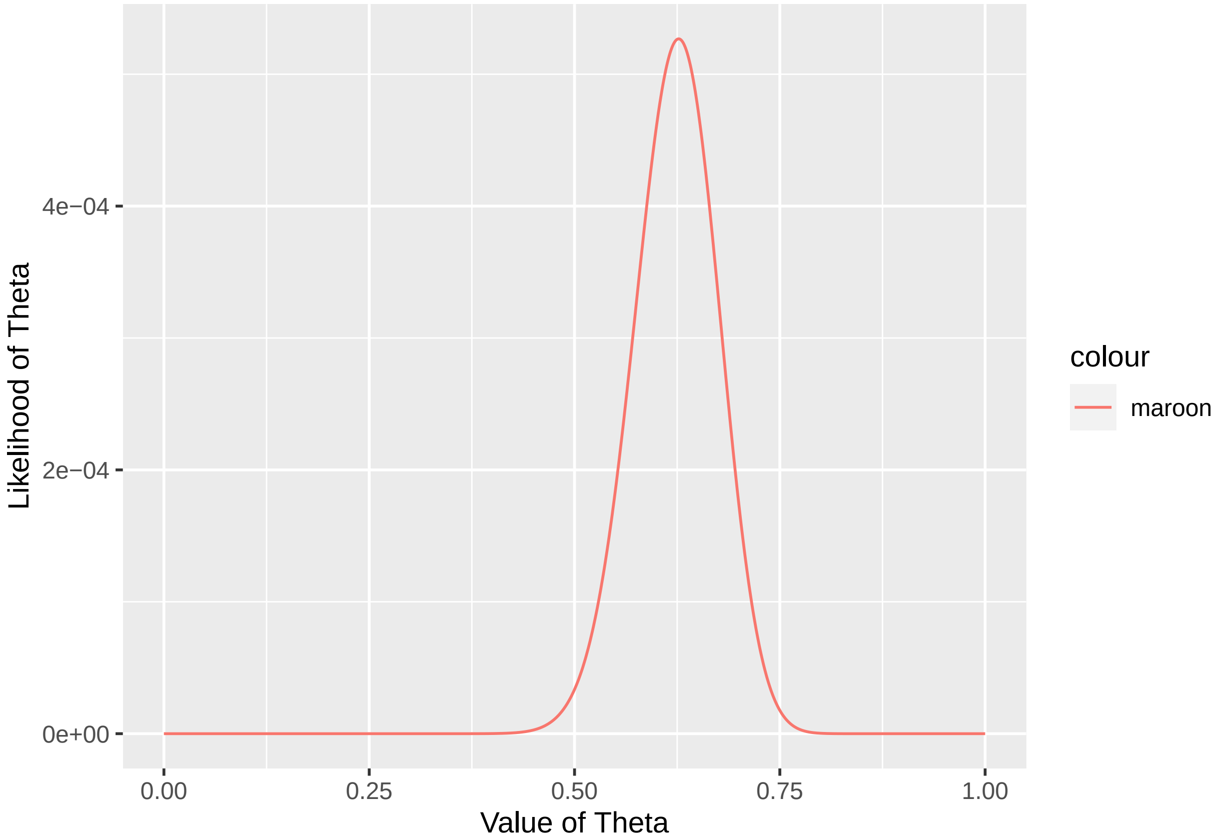
$$f(\vec{x}_1|\theta) \propto \frac{197!}{125!\,18!\,20!\,34!} \left(\frac{2+\theta}{4}\right)^{125} \left(\frac{1-\theta}{4}\right)^{18} \left(\frac{1-\theta}{4}\right)^{20} (\theta/4)^{34}$$

5(b) for (5b) the likelihood is proportional to:

$$f(\vec{x}_2|\theta) \propto \frac{20!}{14!\,0!\,1!\,5!} ((2+\theta)/4)^{14} ((1-\theta)/4)^{0} \left(\frac{1-\theta}{4}\right)^{1} (\theta/4)^{5}$$

Figure 7: Exercise 5

See below a plot of the likelihood function, followed by a table of the top three likelihood-maximizing θ values:




```
##   theta   prob1   prob2   prob3   prob4   likelihood
## 1 0.627 0.65675 0.09325 0.09325 0.15675 0.0005268137
## 2 0.626 0.65650 0.09350 0.09350 0.15650 0.0005267498
## 3 0.628 0.65700 0.09300 0.09300 0.15700 0.0005266786
```

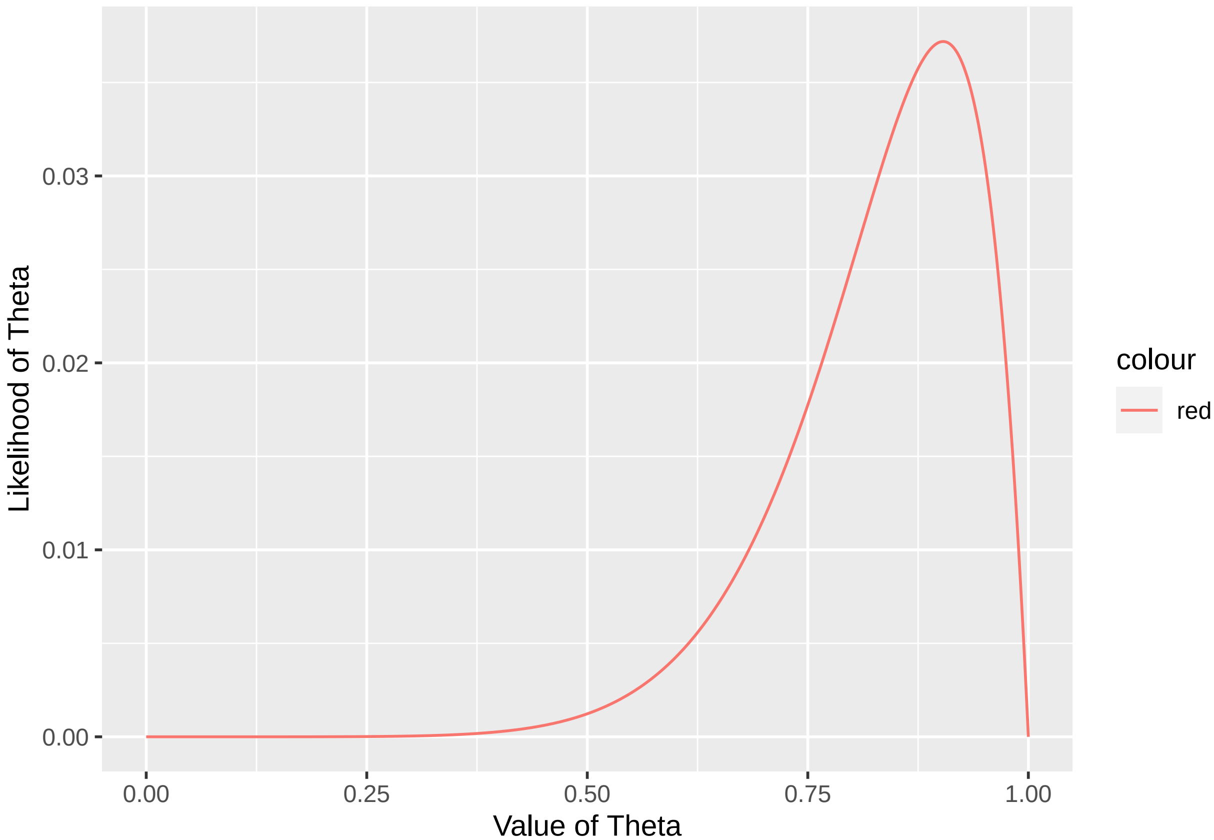
Exercise 5b

What is the likelihood for the data $Y = (14, 0, 1, 5)$?

```
## $Y_data
## [1] 14  0  1  5
##
## $Y_num_trials
## [1] 20
```

See above in the picture included in (5a) the appropriate likelihood function.

See below a plot of the likelihood function, followed by a table of the top three likelihood-maximizing θ values:




```
##   theta   prob1   prob2   prob3   prob4 likelihood
## 1 0.903 0.72575 0.02425 0.02425 0.22575 0.03718812
## 2 0.904 0.72600 0.02400 0.02400 0.22600 0.03718786
## 3 0.902 0.72550 0.02450 0.02450 0.22550 0.03718413
```

Exercise 5c

Use the Newton-Raphson algorithm to obtain the MLE ($\hat{\theta}$) of θ for $Y = (125, 18, 20, 34)$. Try starting your algorithm at $\theta = .1, .2, .3, .4, .6, .8$. How did you assess the convergence of the algorithm?

Given the likelihood derived in (5a), the log-likelihood is:

$((2+\theta)/4, (1-\theta)/4, (1-\theta)/4, \theta/4)$

$$log(p(\vec{x}|\theta)) = log(\frac{n!}{x_1!x_2!x_3!x_4!}) + x_1 log((2+\theta)/4) + x_2 log((1-\theta)/4) + x_3 log((1-\theta)/4) + x_4 log(\theta/4)$$

Hence the first derivative of the log-likelihood w.r.t. θ is:

$$\frac{\partial}{\partial\theta} log(p(\vec{x}|\theta)) = x_1(2+\theta)^{-1} - x_2(1-\theta)^{-1} - x_3(1-\theta)^{-1} + x_4\theta^{-1}$$

```
##   initial_x0    result no_iterations
## 1        0.1 0.6268215             6
## 2        0.2 0.6268215             5
## 3        0.3 0.6268215             5
## 4        0.4 0.6268215             4
## 5        0.6 0.6268215             4
## 6        0.8 0.6268215             5
```

Regardless of our initial guess for θ above, it always takes under 10 iterations for the algorithm to converge on 0.6268215 as the MLE. Based on our visualization in (5a), this result makes sense.

If the absolute difference between an estimate and the estimate in the prior iteration is less than 1e-6, we conclude that the algorithm has converged.

Exercise 5d

Repeat (5c) for $Y = (14, 0, 1, 5)$.

```
##   initial_x0              result no_iterations
## 1        0.1    0.903440114216673             9
## 2        0.2    0.903440114216673             6
## 3        0.3 N/A - No Convergence           544
## 4        0.4    0.903440114216814             6
## 5        0.6 N/A - No Convergence           544
## 6        0.8    0.903440114216679             8
```

Regardless of our initial guess for θ above, when the algorithm converges it does so in under 10 iterations and always converges to an MLE of about 0.903440114217. However, when we begin with an estimate of 0.3 or 0.6, the algorithm does not converge but instead gets stuck on the value of $-\infty$ around the 544th iteration. Regardless, the estimates of the MLE that *do* emerge from convergent algorithm runs align with the results from the visualization and table in (5b).

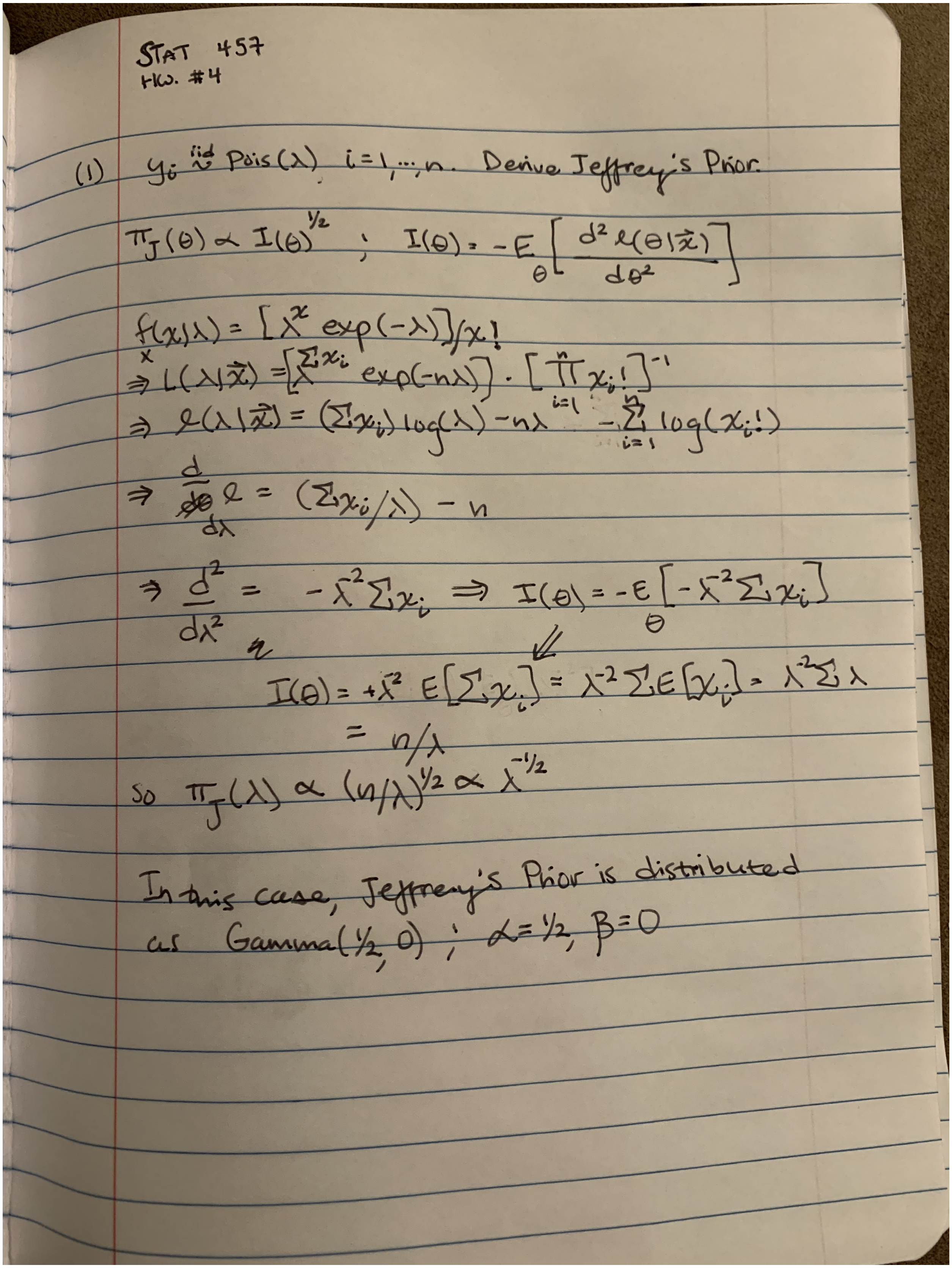
Again, if the absolute difference between an estimate and the estimate in the prior iteration is less than 1e-6, we conclude that the algorithm has converged.

Exercises with Code

```
library(ggplot2)
library(reshape2)
library(truncdist)
library(dplyr)
library(TeachingDemos)
library(numDeriv)
library(knitr)
library(kableExtra)

set.seed(457)
```

Exercise 1

Let $Y_1, \ldots, Y_n$ be an *iid* sample from $Poisson(\lambda)$. Derive the Jeffrey's (noninformative prior). This prior corresponds to a gamma distribution with which parameters?



STAT 457
HW. #4

(1) $y_i \overset{iid}{\sim} \text{Pois}(\lambda)$, $i = 1, \dots, n$. Derive Jeffrey's Prior.

$\pi_J(\theta) \propto I(\theta)^{1/2}$; $I(\theta) = -E_\theta\left[\frac{d^2 \ell(\theta|\vec{x})}{d\theta^2}\right]$

$\underset{x}{f}(x|\lambda) = [\lambda^x \exp(-\lambda)]/x!$

$\Rightarrow L(\lambda|\vec{x}) = [\lambda^{\sum x_i} \exp(-n\lambda)] \cdot [\prod_{i=1}^n x_i!]^{-1}$

$\Rightarrow \ell(\lambda|\vec{x}) = (\sum x_i)\log(\lambda) - n\lambda - \sum_{i=1}^n \log(x_i!)$

$\Rightarrow \frac{d}{d\lambda}\ell = (\sum_i x_i/\lambda) - n$

$\Rightarrow \frac{d^2}{d\lambda^2} = -\lambda^{-2}\sum x_i \Rightarrow I(\theta) = -E_\theta[-\lambda^{-2}\sum_i x_i]$

$I(\theta) = +\lambda^{-2} E[\sum_i x_i] = \lambda^{-2}\sum_i E[x_i] = \lambda^{-2}\sum_i \lambda$
$= n/\lambda$

So $\pi_J(\lambda) \propto (n/\lambda)^{1/2} \propto \lambda^{-1/2}$

In this case, Jeffrey's Prior is distributed as Gamma(1/2, 0) ; $\alpha = 1/2$, $\beta = 0$

Figure 8: Exercise 1

Exercise 2

In the multivariate setting, $\overrightarrow{\theta} = (\theta_1, \ldots, \theta_d)^T$, $p(\theta) \propto |J(\theta)|^{(1/2)}$, provides an invariant prior, where the *ij*th entry of $J(\theta)$ equals $-E[\frac{\partial^2 l(\theta|Y)}{\partial\theta_i \partial\theta_j}]$ and $|X|$ is the determinant of the matrix X.

Let $Y_1, \ldots, Y_n$ be an *iid* sample from $N(\mu, \sigma^2)$, where both parameters are unknown. Derive the invariant prior. How does it compare with the prior $p(\theta) \propto (1/\sigma^2)$?

(2) $f(\vec{x}|\vec{\theta}) = (\sigma^2 2\pi)^{-n/2} \exp[(-1/2)(\sigma^{-2})\Sigma(x_i-\mu)^2]$

$\ell(\vec{\theta}|\vec{x}) = (-n/2)\log(\sigma^2 2\pi) + (-n/2)\log(\sigma^2) + (2\sigma^2)^{-1}\sum_i(x_i-\mu)^2$

$\propto (-n/2)\ln(\sigma^2) - (2\sigma^2)^{-1}[(n-1)s^2 + n(\bar{x}-\mu)^2]$

$$\frac{\partial}{\partial\mu}\ell = (-2)(2\sigma^2)^{-1}\sum_i(x_i-\mu) = -\sigma^{-2}(\sum_i(x_i)-n\mu)$$

$\swarrow$ $\Downarrow$

$$\frac{\partial^2}{\partial\mu^2}\ell = -(n/\sigma^2) \qquad \frac{\partial^2}{\partial\mu\partial\sigma^2}\ell = -\sigma^{-4}(\sum_i(x_i)-n\mu)$$

$$\frac{\partial}{\partial\sigma^2}\ell = (-n/2)\sigma^{+2^{-1}} + (n/2)\sigma^{+2^{-1}} - (1/2)(\sigma)^{+2^{-2}}\sum_i(x_i-\mu)^2$$

$\swarrow$

$$\frac{\partial^2}{\partial\sigma^{2^2}} = (n/2)\sigma^{-4} - (\sigma)^{-6}\left[\sum_i(x_i-\mu)^2 + s^2(n-1)\right]$$

$$= (n + 4(\sum x_i-\mu)^2)\,\sigma^{-4}$$

$$\frac{\partial}{\partial\sigma^2\partial\mu} = -\sigma^{-4}(\sum_i(x_i)-n\mu)$$

E

So we have: $E[-n/\sigma^2] = -n/\sigma^2$

$E[-\sigma^{4}(\sum_i(x_i)-n\mu)] = n\mu\sigma^{4} - \sigma^{4}E[\sum X_i]$

$= n\mu\sigma^{4} - n\mu\sigma^{4} = 0$

$E[(n/2)\sigma^{-4} + 4(\sum(x_i-\mu)^2 + s^2(n-1))\sigma^{-6}] = n/(2\sigma^4) + E[\sigma^{-6}((n-1)s^2 + (n-1)\sigma^2)]$

$= [n/(2\sigma^4) + 0]$

$= -n/(2\sigma^4)$

Figure 9: Exercise 2 (1 of 2)

(2) (cont.) Therefore:

$$I_{\vec{\theta}}(\mu, \sigma^2) = \begin{bmatrix} n\sigma^{-2} & 0 \\ 0 & (1/2)n\sigma^{-4} \end{bmatrix}$$

$$\text{so } \pi(\mu, \sigma^2) \propto \left[(n\sigma^{-2})(1/2)(n\sigma^{-4}) - 0 \right]^{1/2}$$
$$= \left[(1/2) n^2 \sigma^{-6} \right]^{1/2}$$
$$\propto (\sigma^2)^{-3/2}$$

Hence the invariant prior is $(\sigma^2)^{-3/2}$, which is the prior $p(\theta, \sigma^2) \propto (1/\sigma^2)$ raised to the $(3/2)$ power.

Figure 10: Exercise 2 (2 of 2)

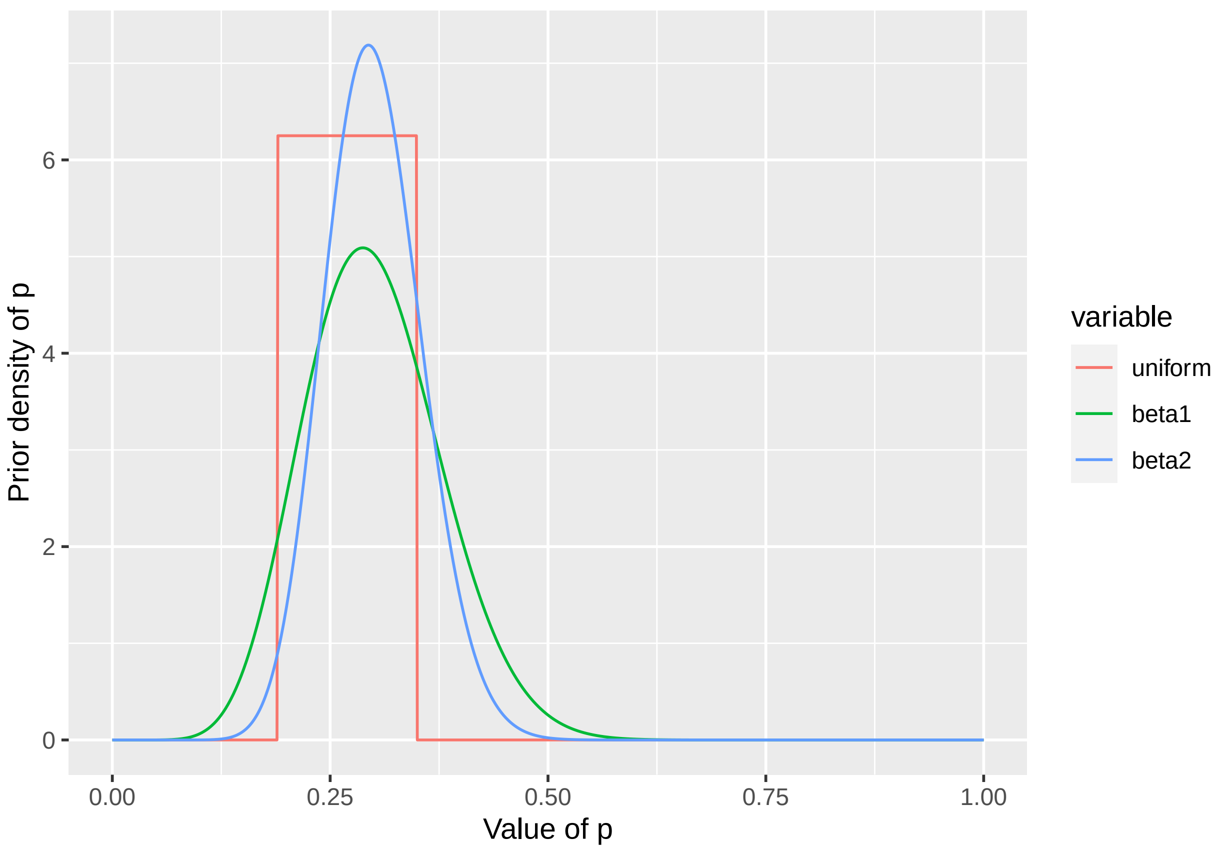
Exercise 3

Let p denote the probability a specific major league baseball player gets a hit in a particular at-bat. Assume that batting averages usually fall in the range 0.19 to 0.35.

Exercise 3a

Consider priors $Unif(0.19, 0.35)$, $Beta(10.2, 23.8)$, and $Beta(20.4, 47.6)$. Plot these priors and discuss each choice.

```
prior_df <- data.frame(x=seq(from=0,to=1,by=0.001))
prior_df$uniform <- dunif(prior_df$x,min=0.19,max=0.35)
prior_df$beta1 <- dbeta(prior_df$x,10.2, 23.8)
prior_df$beta2 <- dbeta(prior_df$x, 20.4,47.6)

prior_df_melted <- melt(prior_df,id="x")

ggplot(data=prior_df_melted,aes(x=x, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of p") +
  ylab("Prior density of p")
```



The choice of a uniform distribution is unique in that it disallows values of p outside of the range 0.19 to 0.35, so in that sense this choice of a prior is quite informative. Since batting averages *usually* but don't always fit within that range, I believe it would be preferable to allow p to take values outside of it, as the two beta priors do. While both priors have a mode around 0.3, the second Beta prior places considerably more confidence in the "best guess" (mode) of 0.3 (it has lighter tails than the first choice of Beta priors). Personally, I'd be drawn to the use of the first Beta prior, simply because it (1) allows for a greater range of batting averages, and (2) places less weight on particular values. It's likely to be the most flexible in allowing a range of batting data to have a solid impact on the posterior.

Exercise 3b

Suppose a player gets 5 hits in 40 at-bats. For each of the above priors: plot likelihood, posterior, and prior; compute the probability that the player is better than a 0.200 hitter; compute your best guess as to the batting average of the player; compute a 95% credible interval for p.

Here are the derivations for posteriors in (3b):

(3b) likelihood: let y = # of hits in 40 at-bats

$y \sim \text{binomial}(40, p) \rightarrow P_y(y|p) = \binom{40}{y} p^{y} (1-p)^{40-y}$

(i) $\pi(p) \sim \text{Unif}(0.19, 0.35) \Rightarrow \pi(p) = 1/(0.35-0.19) = (1/0.16)$

$$p(p|y) = \frac{p(y|p)\pi(p)}{p(y)} = \frac{p(y|p)\pi(p)}{\int_{\Theta} p(y|p)\pi(p)\,dp}$$

$$= \frac{(1/0.16)\binom{40}{y} p^{y}(1-p)^{40-y}}{\int_{\Theta} \left(\frac{1}{0.016}\right)\binom{40}{y} p^{y}(1-p)^{40-y}\,dp}$$

$$= \frac{\cancel{(1/0.16)}\binom{40}{y}[p^{y}(1-p)^{40-y}]}{\cancel{(1/0.16)}\int_{\Theta} \binom{40}{y} p^{y}(1-p)^{40-y}\,dp} = \frac{\binom{40}{y}[p^{y}(1-p)^{40-y}]}{1}$$

$$\Rightarrow p(p|y) = p(y|p) = \binom{40}{y}[p^{y}(1-p)^{40-y}]$$

$\Rightarrow$ reparameterize as Beta$(5+1, 40-5+1)$ = Beta$(6, 36)$

(ii) $\pi(p) \sim \text{Beta}(10.3, 23.8) \Rightarrow \pi(p) = \dfrac{p^{\alpha-1}(1-p)^{\beta-1}}{\left(\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}$

Figure 11: Exercise 3 (1 of 2)

3(b) cont.

w/ Beta:

$$p(p\mid y) = \frac{\binom{40}{y}(p)^{y}(1-p)^{40-y}\cdot p^{\alpha}(1-p)^{1-\beta}\left(\frac{\Gamma(\alpha)(\Gamma(\beta))}{\Gamma(\alpha+\beta)}\right)}{\int_{\Theta}\left[\binom{40}{y}p^{y}(1-p)^{40-y}p^{\alpha}(1-p)^{1-\beta}\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)\right]dp}$$

$$= \frac{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)\cancel{\binom{40}{y}}\,p^{y+\alpha}(1-p)^{-y+\beta-1+40}}{\cancel{\binom{40}{y}}\int_{\Theta}p^{y+\alpha}(1-p)^{-1+40-y+\beta}\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)dp} \quad (1)$$

$$= \frac{\cancel{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}\left[\left(\frac{\Gamma(y+\alpha)\Gamma(40+y+\beta)}{\Gamma(-40+2y+\alpha+\beta)}\right)p^{y+\alpha}(1-p)^{-1+40-y+\beta}\right]}{\cancel{\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}\int_{\Theta}\left[\,(1)\,\right]dp}\,dp$$

$f_{y|y}(x)$ of $\text{Beta}(y+\alpha, +40-y+\beta)$ integrates to 1 by prop. PDF

$$\Rightarrow p(p\mid y) = \left(\frac{\Gamma(y+\alpha)\Gamma(y+\beta-40)}{\Gamma(2y+\alpha+\beta-40)}\right)p^{y+\alpha}(1-p)^{-1+40-y-\beta}$$

so posterior $\sim \text{Beta}(y+\alpha, -y+\beta+40)$

Where when prior $\sim \text{Beta}(10.2, 23.8) \Rightarrow$ posterior $\sim \text{Beta}(y+10.2, -y+63.8)$

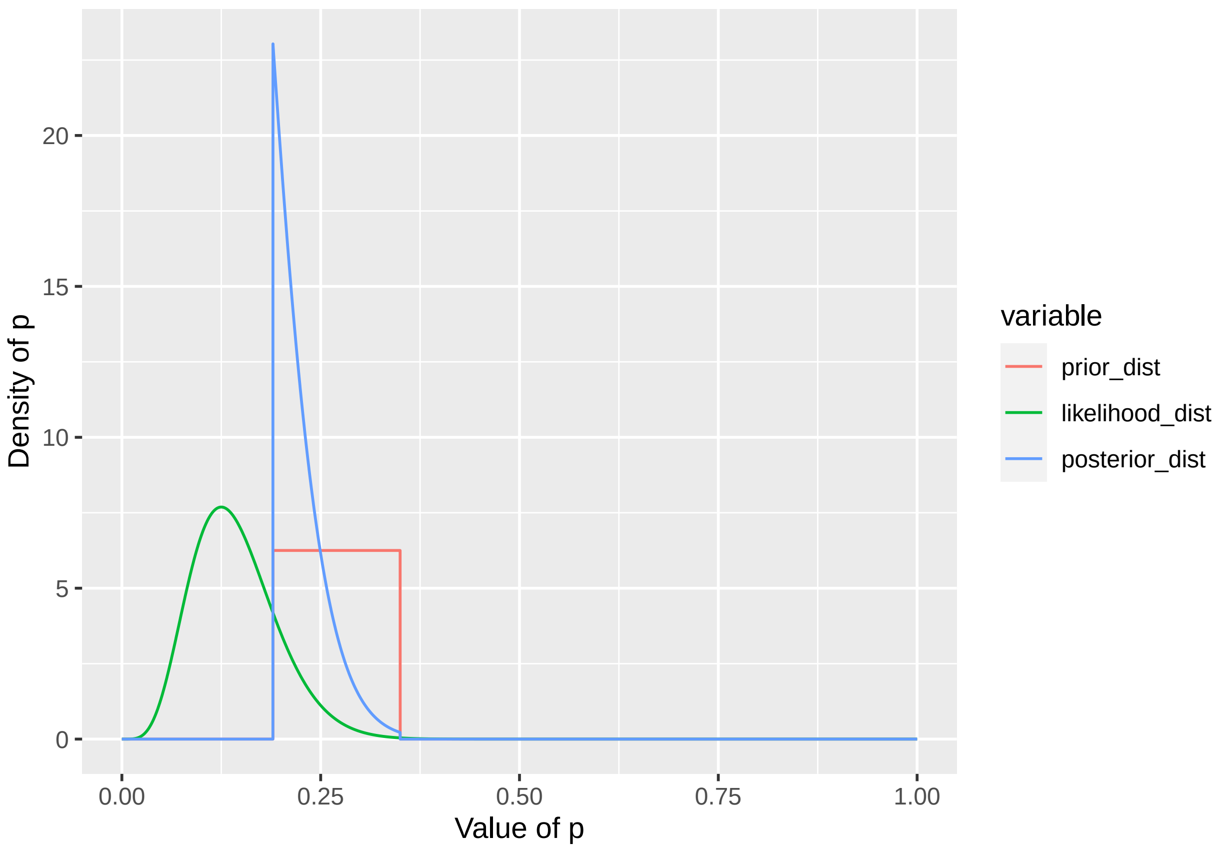
Prior $\sim \text{Beta}(20.4, 47.6) \Rightarrow$ posterior $\sim \text{Beta}(y+20.4, -y+87.6)$

Figure 12: Exercise 3 (2 of 2)

Results for $Unif(0.19, 0.35)$:

Likelihood is binomial, $Y \sim bin(n = 40, p)$ with Y being the number of hits in 40 at-bats. Prior is given as uniform, and posterior is calculated as $Beta(6, 36)$ (see derivations above).

```
binom_integrand <- function(p){
  p**5 * (1-p)**(35)
}
area_under_binom <- integrate(binom_integrand,0,1)$value

prior_df_unif <- data.frame(p=seq(from=0,to=1,by=0.0001))
prior_df_unif$prior_dist<- dunif(prior_df_unif$p,min=0.19,max=0.35)
prior_df_unif$likelihood_dist <- (prior_df_unif$p**5 * (1-prior_df_unif$p)**(35))/area_under_binom
prior_df_unif$posterior_dist <- dtrunc(prior_df_unif$p,spec="beta",a=0.19,b=0.35, 6,36)

prior_df_unif_melted <- melt(prior_df_unif,id="p")

ggplot(data=prior_df_unif_melted,aes(x=p, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of p") +
  ylab("Density of p")
```




```
prior_df_unif %>% arrange(desc(posterior_dist)) %>% head(3)
```

```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.1900       6.25        4.185609       23.03652
## 2 0.1901       6.25        4.178540       22.99761
## 3 0.1902       6.25        4.171475       22.95873
```

```r
1-ptrunc(0.2,spec="beta",a=0.19,b=0.35, 6,36)
```

```
## [1] 0.7886796
```

```r
punif_samples <- rtrunc(n=1e6,spec="beta",a=0.19,b=0.35, 6,36)
unif_posterior_ci <- quantile(punif_samples,c(0.025,0.975))
list(unif_posterior_ci=unif_posterior_ci)
```

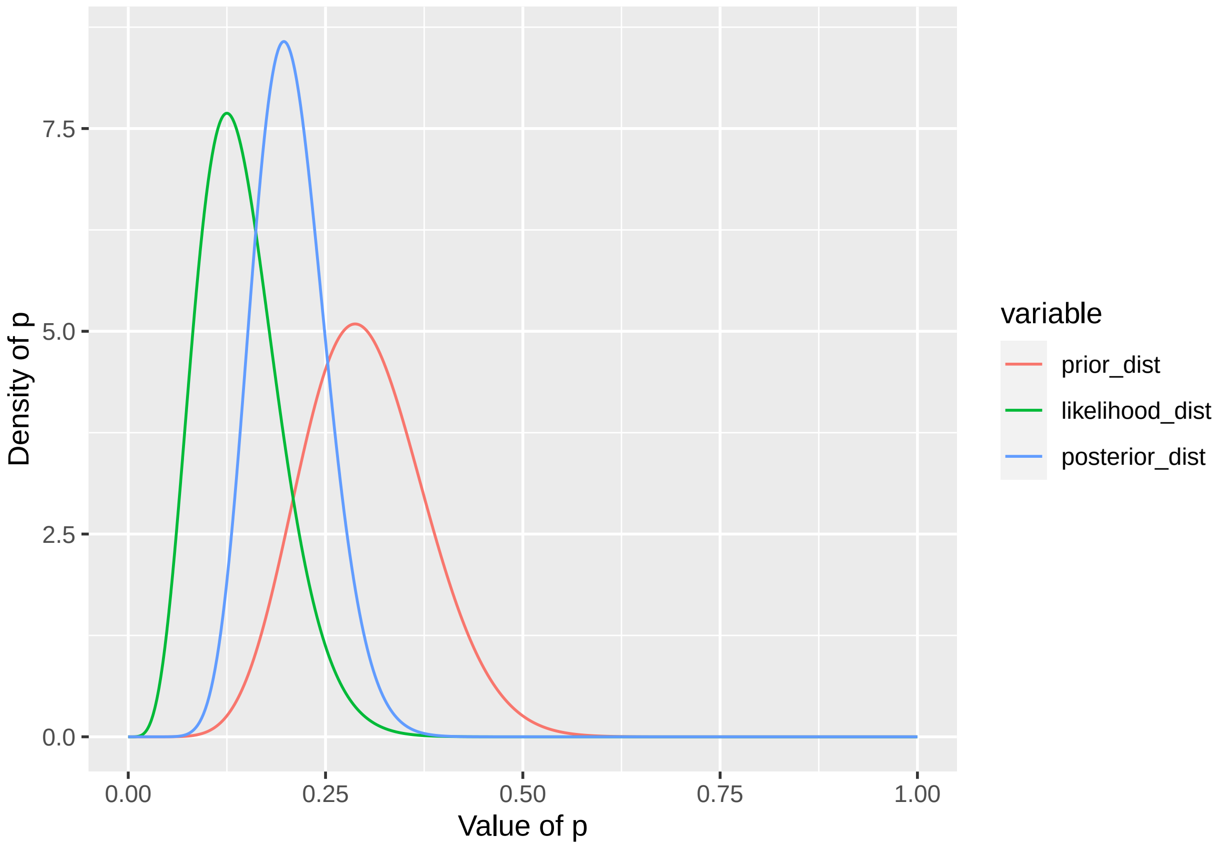
```
## $unif_posterior_ci
##      2.5%     97.5%
## 0.1910947 0.3056695
```

My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.1900. The probability that the batter is better than a 0.200 hitter is 0.7886796. Finally, a 95% credible interval for the true value of p is [0.1910947, 0.3056695].

Results for $Beta(10.2, 23.8)$:

See above the derivation of the posterior distribution for p with prior distribution $Beta(10.2, 23.8)$ and the binomial likelihood. This posterior follows $p \sim Beta(y + 10.2, 23.8 + n - y) = Beta(15.2, 58.8)$.

```r
prior_df_beta_i <- data.frame(p=seq(from=0,to=1,by=0.0001))
prior_df_beta_i$prior_dist<- dbeta(prior_df_beta_i$p,10.2, 23.8)
prior_df_beta_i$likelihood_dist <- (prior_df_beta_i$p**5 * (1-prior_df_beta_i$p)**(35))/area_under_binom
prior_df_beta_i$posterior_dist <- dbeta(prior_df_beta_i$p,15.2,58.8)

prior_df_beta_i_melted <- melt(prior_df_beta_i,id="p")

ggplot(data=prior_df_beta_i_melted,aes(x=p, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of p") +
  ylab("Density of p")
```




```
prior_df_beta_i %>% arrange(desc(posterior_dist)) %>% head(3)
```

```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.1972   2.409780        3.688139       8.572733
## 2 0.1973   2.414179        3.681414       8.572722
## 3 0.1971   2.405382        3.694870       8.572705
```

```
1-pbeta(0.2,15.2,58.8)
```

```
## [1] 0.5234115
```

```
pbeta_i_samples <- rbeta(1e6,15.2,58.5)
beta_i_posterior_ci <- quantile(pbeta_i_samples,c(0.025,0.975))
list(beta_i_posterior_ci=beta_i_posterior_ci)
```

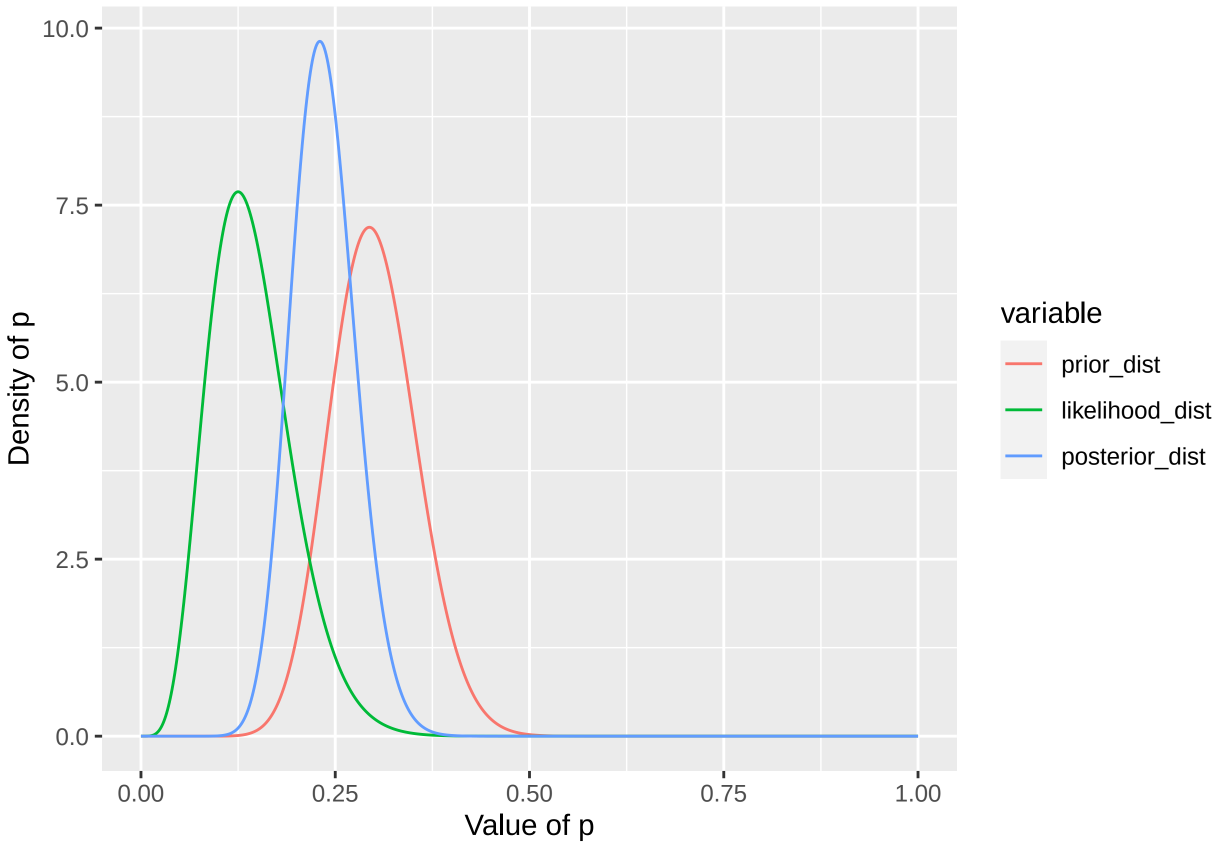
```
## $beta_i_posterior_ci
##      2.5%     97.5%
## 0.1224982 0.3052243
```

My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.1972. The probability that the batter is better than a 0.200 hitter is 0.5234115. Finally, a 95% credible interval for the true value of p is [0.1224982, 0.3052243].

Results for $Beta(20.4, 47.6)$:

The last prior yields a similar result (both beta). See above the derivation of the posterior distribution for p with prior distribution $Beta(20.4, 47.6)$ and the binomial likelihood. This posterior follows $p \sim Beta(y + 20.4, 47.6 + n - y) = Beta(25.4, 82.6)$.

```
prior_df_beta_ii <- data.frame(p=seq(from=0,to=1,by=0.0001))
prior_df_beta_ii$prior_dist<- dbeta(prior_df_beta_ii$p,20.4, 47.6)
prior_df_beta_ii$likelihood_dist <- (prior_df_beta_ii$p**5 * (1-prior_df_beta_ii$p)**(35))/area_under_bi
prior_df_beta_ii$posterior_dist <- dbeta(prior_df_beta_ii$p,25.4,82.6)

prior_df_beta_ii_melted <- melt(prior_df_beta_ii,id="p")

ggplot(data=prior_df_beta_ii_melted,aes(x=p, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of p") +
  ylab("Density of p")
```




```
prior_df_beta_ii %>% arrange(desc(posterior_dist)) %>% head(3)
```

```
##        p prior_dist likelihood_dist posterior_dist
## 1 0.2302   3.518726        1.839790       9.813445
## 2 0.2301   3.510375        1.844163       9.813423
## 3 0.2303   3.527082        1.835425       9.813409
```

```
1-pbeta(0.2,25.4,82.6)
```

```
## [1] 0.8031853
```

```
pbeta_ii_samples <- rbeta(1e6,25.4,82.6)
beta_ii_posterior_ci <- quantile(pbeta_ii_samples,c(0.025,0.975))
list(beta_ii_posterior_ci=beta_ii_posterior_ci)
```

```
## $beta_ii_posterior_ci
```

Table 2: Results for Each Prior Distribution

Prior	Best_Guess	Prob_Greater_Than_1in5	CI_95_Lower_Bound	CI_95_Upper_Bound
Unif(0.19,0.35)	0.1900	0.7886796	0.1910947	0.3056695
Beta(10.2,23.8)	0.1972	0.5234115	0.1224982	0.3052243
Beta(20.4,47.6)	0.2302	0.8031853	0.1604733	0.3187223

```
##      2.5%     97.5%
## 0.1604733 0.3187223
```

My best guess as to the batting average of the player would be the mode of the posterior distribution. For this Beta distribution, the mode is 0.2302. The probability that the batter is better than a 0.200 hitter is 0.8031853. Finally, a 95% credible interval for the true value of p is [0.1604733, 0.3187223].

```
batting_df <- data.frame(
  Prior = c("Unif(0.19,0.35)","Beta(10.2,23.8)","Beta(20.4,47.6)"),
  Best_Guess = c(0.1900,0.1972,0.2302),
  Prob_Greater_Than_1in5 = c(0.7886796,0.5234115,0.8031853),
  CI_95_Lower_Bound = c(0.1910947,0.1224982,0.1604733),
  CI_95_Upper_Bound = c(0.3056695,0.3052243,0.3187223)
)

#batting_df$Credible_Interval_Range = batting_df$CI_95_Upper_Bound - batting_df$CI_95_Lower_Bound
#batting_df$Credible_Interval_Mean = (batting_df$CI_95_Upper_Bound + batting_df$CI_95_Lower_Bound)/2

batting_df %>%
  kable(caption="Results for Each Prior Distribution") %>%
  kable_styling(font_size = 10) %>%
  kable_classic(full_width = F, position="left", html_font = "Cambria")
```

Exercise 4

Data represents the number of arrivals for 45 minute time intervals of length 2 minutes at a cashier's desk at a supermarket:

```
arrivals_df <- c(rep(0,6),rep(1,18),rep(2,9),rep(3,7),rep(4,4),5)
list(arrivals_df=arrivals_df, arrivals_mean=mean(arrivals_df), arrivals_sum=sum(arrivals_df))
```

```
## $arrivals_df
##  [1] 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3
## [39] 3 3 4 4 4 4 5
##
## $arrivals_mean
## [1] 1.733333
##
## $arrivals_sum
## [1] 78
```

Exercise 4a

For a $Gamma(2, 1)$ prior, obtain the posterior distribution under a $Poisson(\lambda)$ model for the data. Draw the prior and the posterior. Note on your plot the mean, variance, and mode of the posterior.

Here is the derivation of the posterior for (4a):

(4a) $y \sim \text{Poisson}(\lambda)$

$$\pi(\lambda) \sim \text{Gamma}(2,1) \Rightarrow \pi(\lambda) = \frac{B^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-Bx)$$

$$= \frac{(1)^2}{\Gamma(2)} x^{2-1} \exp(-x)$$

$$= \lambda \exp(-\lambda)$$

$$p(\lambda|y) = \frac{p(\vec{y}|\lambda)p(\lambda)}{\int p(y|\lambda)\pi(\lambda)\, d\lambda}$$

$$= \left\{[\lambda^{\bar{y}n} \exp(-\lambda)](y!)^{-1} \lambda \exp(-n\lambda)\right\} \Big/ \left\{\int_{(+)} \{①\}\, d\lambda\right\}$$

$$= \frac{\cancel{(y!)^{-1}}\, [\lambda^{n\bar{y}+1} \exp(-\lambda(n+1))]}{\cancel{(y!)^{-1}} \int_{(+)} \lambda^{n\bar{y}+1} \exp(-\lambda(n+1))\, d\lambda}$$

$$\propto \left[\lambda^{n\bar{y}+1} \exp(-\lambda(n+1))\right] \Big/ \int_{(+)} \text{Gamma}(n\bar{y}+2, 1+n)$$

$$= [①]/1 = \lambda^{n\bar{y}+1} \exp(-\lambda(n+1))$$

So posterior $\pi(\lambda|\vec{y}) \sim \text{Gamma}(n\bar{y}+2, 1+n)$

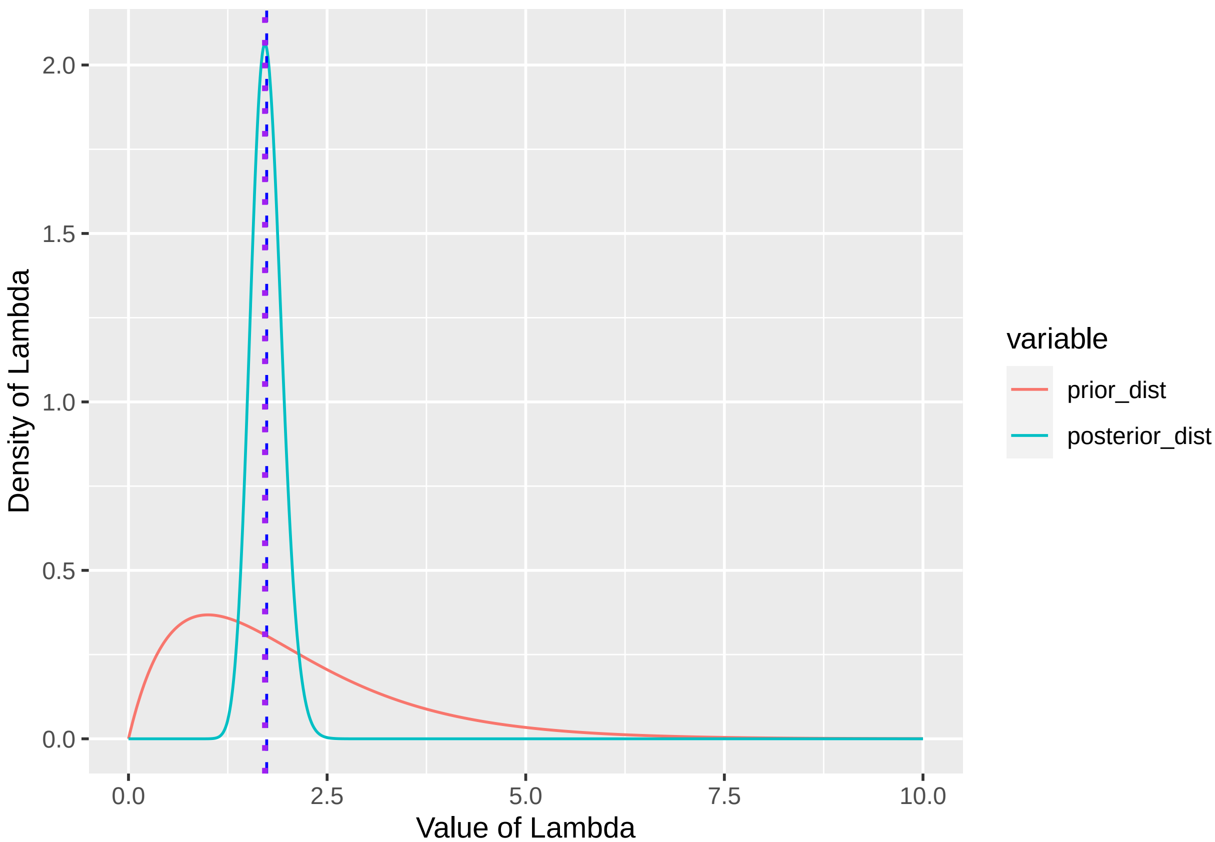
Figure 13: Exercise 4

As pictured above, the derived posterior is $\pi(\lambda|Y) \sim Gamma(n\bar{y} + 2, n + 1) = Gamma(80, 46)$.

```
dists4 <- data.frame(l=seq(from=0,to=10, by=0.001))
dists4$prior_dist <- dgamma(dists4$l,2,1)
dists4$posterior_dist <- dgamma(dists4$l,80,46)

dists4_melted <- melt(dists4,id="l")

post_mean <- 80/46
post_mode <- dists4 %>% arrange(desc(posterior_dist)) %>% select(l) %>% slice(1) %>% pull()
post_var <- 80/(46**2)
print(list(post_mean=post_mean,post_mode=post_mode,post_var=post_var))
```

```
## $post_mean
## [1] 1.73913
##
## $post_mode
## [1] 1.717
##
## $post_var
## [1] 0.03780718
```

```
ggplot(data=dists4_melted,aes(x=l, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of Lambda") +
  ylab("Density of Lambda") +
  geom_vline(aes(xintercept=post_mean),color="blue", linetype="dashed", size=0.5)+
  geom_vline(aes(xintercept=post_mode),color="purple", linetype="dotted", size=1)
```



See above a plot of the posterior. Its mean is 1.73913, its mode 1.717, and its variance is 0.03780718.

Exercise 4b

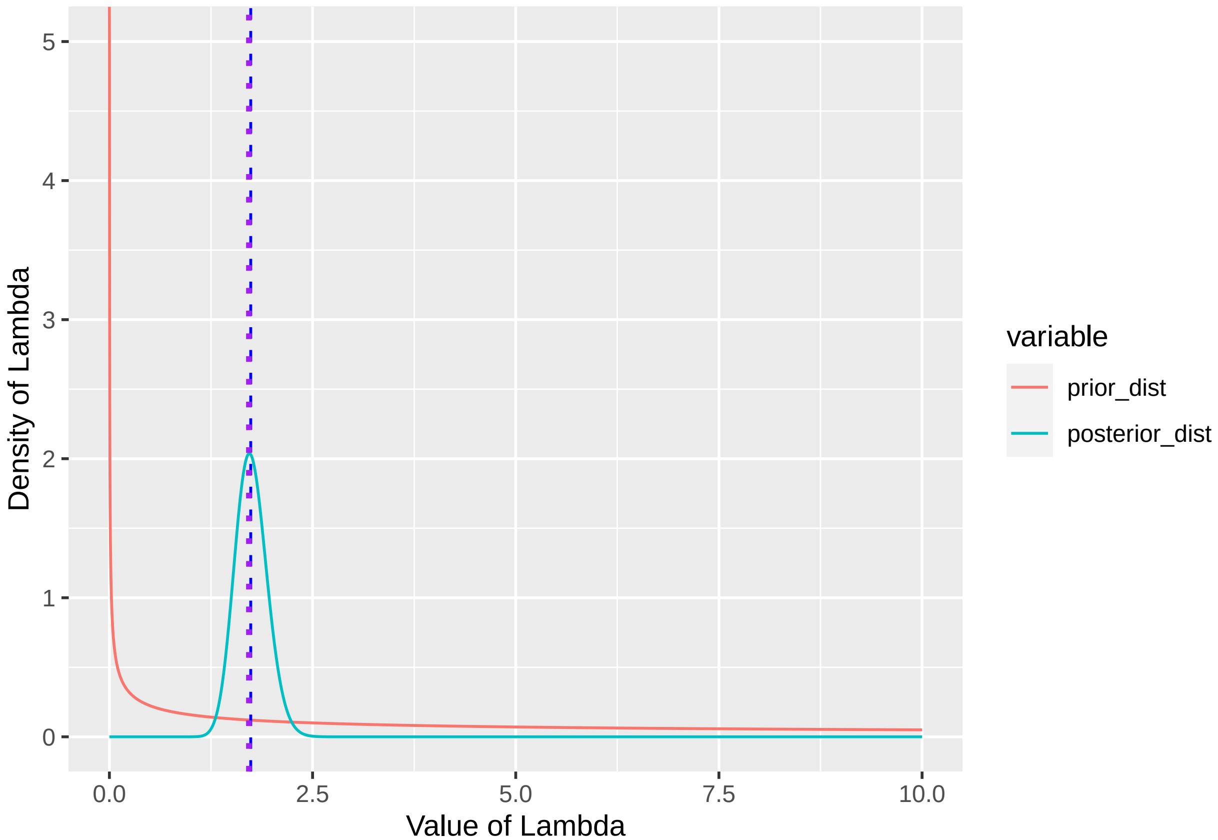
For the noninformative prior, i.e. a $Gamma(?, ?)$, repeat (4a).

We found in exercise (1) that the noninformative (Jeffrey's) prior for a Poisson distribution is $\lambda \sim Gamma(0.5, 0)$. Then generalizing from (4a), we know that $\pi(\lambda|Y) \sim Gamma(n\bar{y} + \alpha, n + \beta)$, so the posterior in the case of (4b) is $\pi(\lambda|Y) \sim Gamma(78.5, 45)$

```
gamma_integrand <- function(l){
  l**(-1/2)
}
area_under_gamma<- integrate(gamma_integrand,0,10)$value

dists4b <- data.frame(l=seq(from=0,to=10, by=0.001))
dists4b$prior_dist <- (dists4b$l**(-1/2))/area_under_gamma
dists4b$posterior_dist <- dgamma(dists4b$l,78.5,45)

dists4b_melted <- melt(dists4b,id="l")

post_mean_b <- 78.5/45
post_mode_b <- dists4b %>% arrange(desc(posterior_dist)) %>% select(l) %>% slice(1) %>% pull()
post_var_b <- 78.5/(45**2)
print(list(post_mean_b=post_mean_b,post_mode_b=post_mode_b,post_var_b=post_var_b))
```

```
## $post_mean_b
## [1] 1.744444
##
## $post_mode_b
## [1] 1.722
##
## $post_var_b
## [1] 0.03876543
```

```
ggplot(data=dists4b_melted,aes(x=l, y=value, colour=variable)) +
  geom_line() +
  xlab("Value of Lambda") +
  ylab("Density of Lambda") +
  geom_vline(aes(xintercept=post_mean),color="blue", linetype="dashed", size=0.5)+
  geom_vline(aes(xintercept=post_mode),color="purple", linetype="dotted", size=1)
```



See above the plot of the posterior based on the noninformative prior. Here, the mean is 1.744444, the mode is 1.722, and the variance is 0.03876543.

Exercise 5

197 animals are distributed into four categories: $Y = (y_1, y_2, y_3, y_4)$ according to the genetic linkage model $((2+\theta)/4, (1-\theta)/4, (1-\theta)/4, \theta/4)$

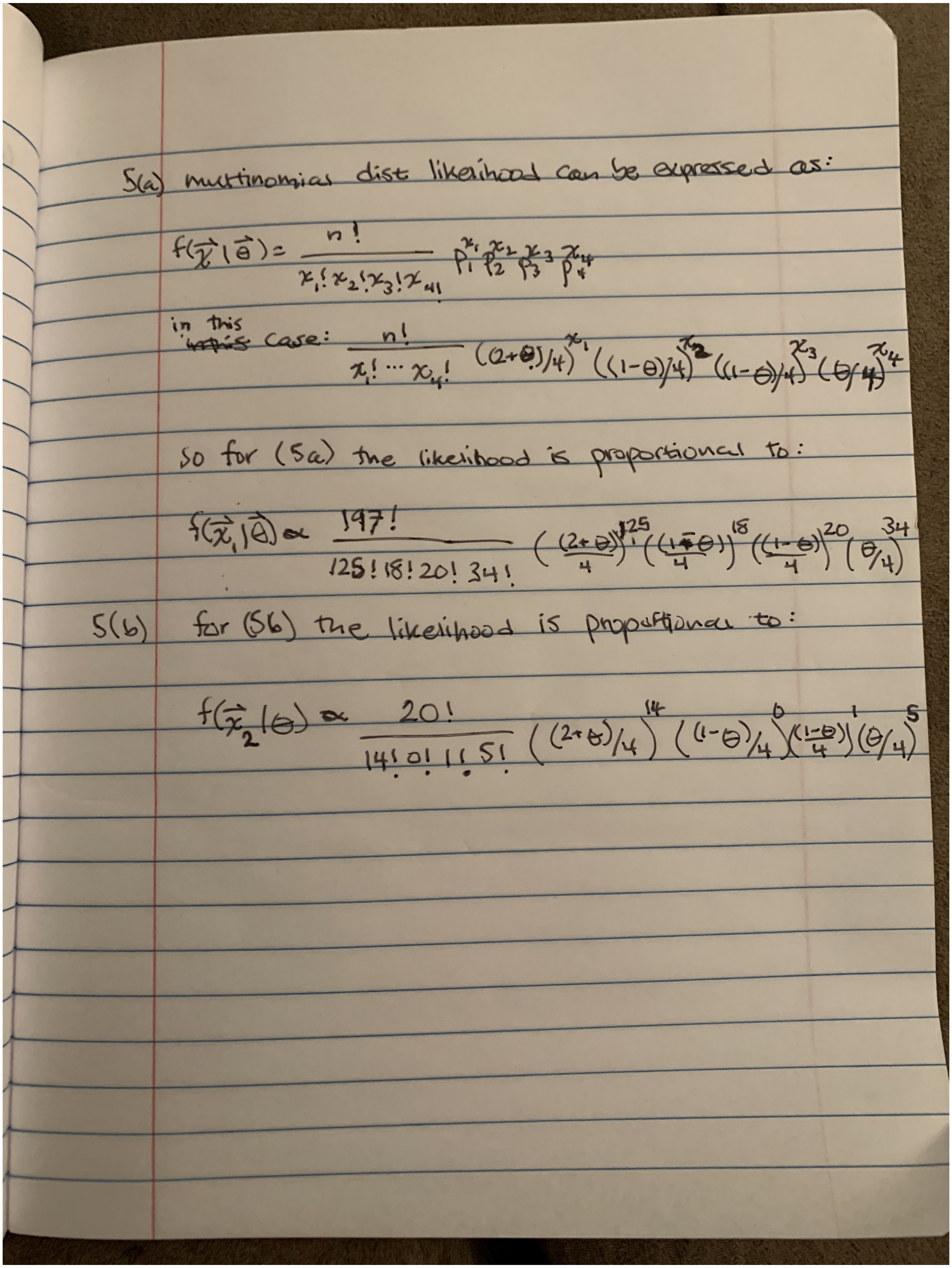
Exercise 5a

What is the likelihood for the data $Y = (125, 18, 20, 34)$?

```
data5a <- c(125,18,20,34)
list(Y_data=data5a,Y_num_trials=sum(data5a))
```

```
## $Y_data
## [1] 125  18  20  34
##
## $Y_num_trials
## [1] 197
```

See below the derivations of the appropriate likelihood functions for both (5a) and (5b):



5(a) multinomial dist likelihood can be expressed as:

$$f(\vec{x}\,|\,\vec{\theta}) = \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\, p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

in this case: $$\frac{n!}{x_1! \cdots x_4!} \left((2+\theta)/4\right)^{x_1} \left((1-\theta)/4\right)^{x_2} \left((1-\theta)/4\right)^{x_3} \left(\theta/4\right)^{x_4}$$

so for (5a) the likelihood is proportional to:

$$f(\vec{x}_1\,|\,\vec{\theta}) \propto \frac{197!}{125!\,18!\,20!\,34!} \left(\frac{2+\theta}{4}\right)^{125} \left(\frac{1-\theta}{4}\right)^{18} \left(\frac{1-\theta}{4}\right)^{20} \left(\theta/4\right)^{34}$$

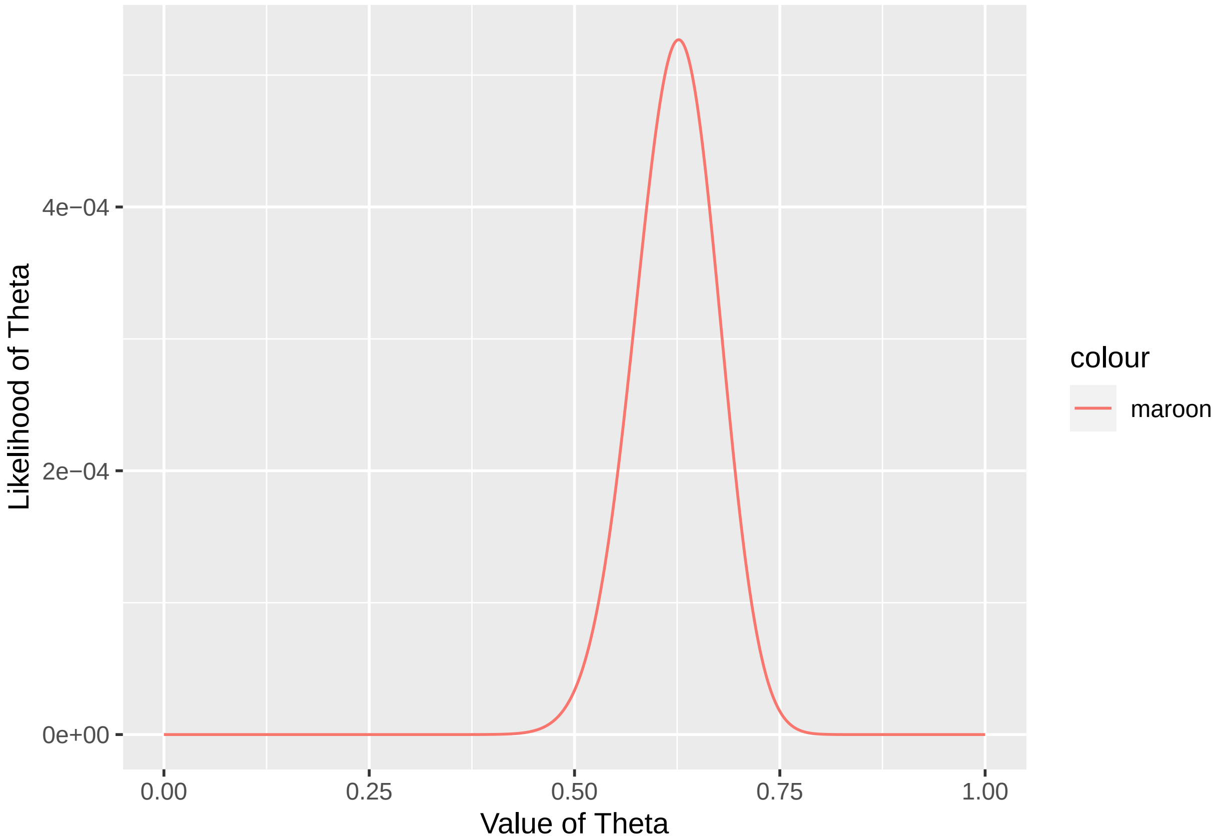
5(b) for (5b) the likelihood is proportional to:

$$f(\vec{x}_2\,|\,\theta) \propto \frac{20!}{14!\,0!\,1!\,5!} \left((2+\theta)/4\right)^{14} \left((1-\theta)/4\right)^{0} \left(\frac{1-\theta}{4}\right)^{1} \left(\theta/4\right)^{5}$$

Figure 14: Exercise 5

See below a plot of the likelihood function, followed by a table of the top three likelihood-maximizing θ values:

```
like5a <- data.frame(theta = seq(0,1,0.001))
like5a$prob1 <- ((2+like5a$theta)/4)
like5a$prob2 <- ((1-like5a$theta)/4)
like5a$prob3 <- ((1-like5a$theta)/4)
like5a$prob4 <- ((like5a$theta)/4)
like5a$likelihood <- ((prod(126:197))/(factorial(18)*factorial(20)*factorial(34)))*(like5a$prob1)**125 *

ggplot(data=like5a,aes(x=theta, y=likelihood, colour="maroon")) +
  geom_line() +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```




```
like5a %>% arrange(desc(likelihood)) %>% head(3)
```

```
##   theta   prob1   prob2   prob3   prob4   likelihood
## 1 0.627 0.65675 0.09325 0.09325 0.15675 0.0005268137
## 2 0.626 0.65650 0.09350 0.09350 0.15650 0.0005267498
## 3 0.628 0.65700 0.09300 0.09300 0.15700 0.0005266786
```

Exercise 5b

What is the likelihood for the data $Y = (14, 0, 1, 5)$?

```
data5b <- c(14,0,1,5)
list(Y_data=data5b,Y_num_trials=sum(data5b))
```

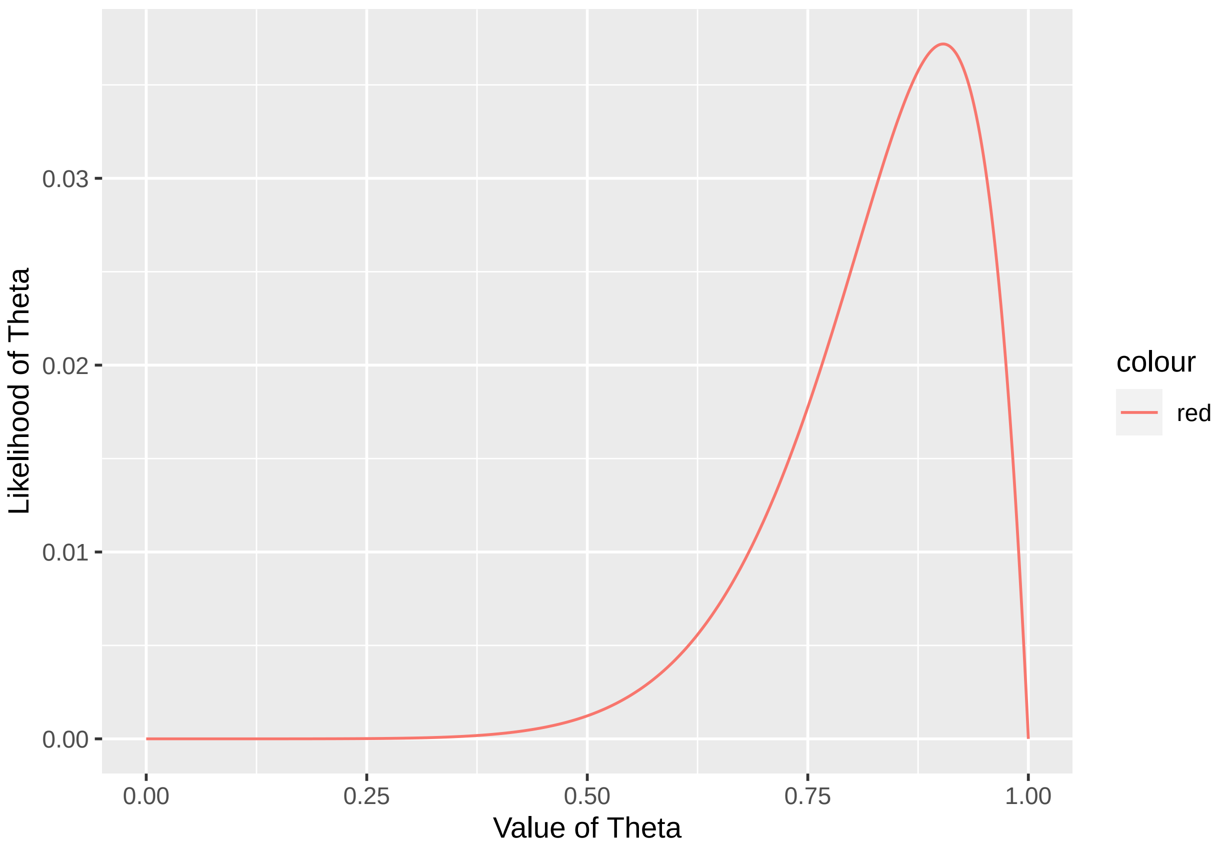
```
## $Y_data
```

```
## [1] 14  0  1  5
##
## $Y_num_trials
## [1] 20
```

See above in the picture included in (5a) the appropriate likelihood function.

See below a plot of the likelihood function, followed by a table of the top three likelihood-maximizing θ values:

```
like5b <- data.frame(theta = seq(0,1,0.001))
like5b$prob1 <- ((2+like5b$theta)/4)
like5b$prob2 <- ((1-like5b$theta)/4)
like5b$prob3 <- ((1-like5b$theta)/4)
like5b$prob4 <- ((like5b$theta)/4)
like5b$likelihood <- ((factorial(20))/(factorial(14)*factorial(0)*factorial(1)*factorial(5)))*(like5b$p

ggplot(data=like5b,aes(x=theta, y=likelihood, colour="red")) +
  geom_line() +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```




```
like5b %>% arrange(desc(likelihood)) %>% head(3)
```

```
##   theta   prob1   prob2   prob3   prob4 likelihood
## 1 0.903 0.72575 0.02425 0.02425 0.22575 0.03718812
## 2 0.904 0.72600 0.02400 0.02400 0.22600 0.03718786
## 3 0.902 0.72550 0.02450 0.02450 0.22550 0.03718413
```

Exercise 5c

Use the Newton-Raphson algorithm to obtain the MLE ($\hat{\theta}$) of θ for $Y = (125, 18, 20, 34)$. Try starting your algorithm at $\theta = .1, .2, .3, .4, .6, .8$. How did you assess the convergence of the algorithm?

Given the likelihood derived in (5a), the log-likelihood is:

$((2+\theta)/4, (1-\theta)/4, (1-\theta)/4, \theta/4)$

$$log(p(\vec{x}|\theta)) = log(\frac{n!}{x_1!x_2!x_3!x_4!}) + x_1 log((2+\theta)/4) + x_2 log((1-\theta)/4) + x_3 log((1-\theta)/4) + x_4 log(\theta/4)$$

Hence the first derivative of the log-likelihood w.r.t. θ is:

$$\frac{\partial}{\partial\theta} log(p(\vec{x}|\theta)) = x_1(2+\theta)^{-1} - x_2(1-\theta)^{-1} - x_3(1-\theta)^{-1} + x_4\theta^{-1}$$

```
dlogl_c <- function(theta){

  dlogl_result <- (125 * (2 + theta)**(-1)) - (18 * (1 - theta)**(-1)) - (20 * (1 - theta)**(-1)) + (34
  return(dlogl_result)
}

dlogl_d <- function(theta){

  dlogl_result <- (14 * (2 + theta)**(-1)) - (0 * (1 - theta)**(-1)) - (1 * (1 - theta)**(-1)) + (5/the
  return(dlogl_result)
}

nr_algorithm <- function(dlogl, init_val, tol_level = 1e-6, iters = 1000) {

  x0 <- init_val
  vals_record <- numeric(iters)

  for (i in 1:iters) {
    #if (i %% 100 == 0){
    #print(list(Iteration_Number=i))
    #}

    d2logl <- genD(func = dlogl, x = x0)$D[1] #take second derivative of l
    x1 <- x0 - (dlogl(x0) / d2logl) #take N-R iteration step
    vals_record[i] <- x1

    #checking for convergence
    if (is.na(x1)) {
      return(list(result="N/A - No Convergence", vals_record=vals_record[1:i], no_iterations=i))

    }
    else if (abs(x1 - x0) < tol_level) {
      result <- x1
      return(list(result=result, vals_record=vals_record[1:i], no_iterations=i))
    }

    #if convergence has not been achieved, iterate again
    x0 <- x1
```

```r
  }
  return(list(result="N/A - No Convergence", vals_record=vals_record[1:i], no_iterations=i))
}
```

```r
mle_df_c <- data.frame(initial_x0 = c(0.1,0.2,0.3,0.4,0.6,0.8),
                       result = rep(0,6),
                       no_iterations = rep(0,6))

for (i in 1:6){
  output <- nr_algorithm(dlogl_c,mle_df_c[i,1])
  mle_df_c[i,2] = output$result
  mle_df_c[i,3] = output$no_iterations
}

mle_df_c
```

```
##   initial_x0    result no_iterations
## 1        0.1 0.6268215             6
## 2        0.2 0.6268215             5
## 3        0.3 0.6268215             5
## 4        0.4 0.6268215             4
## 5        0.6 0.6268215             4
## 6        0.8 0.6268215             5
```

Regardless of our initial guess for θ above, it always takes under 10 iterations for the algorithm to converge on 0.6268215 as the MLE. Based on our visualization in (5a), this result makes sense.

If the absolute difference between an estimate and the estimate in the prior iteration is less than 1e-6, we conclude that the algorithm has converged.

Exercise 5d

Repeat (5c) for $Y = (14, 0, 1, 5)$.

```r
mle_df_d <- data.frame(initial_x0 = c(0.1,0.2,0.3,0.4,0.6,0.8),
                       result = rep(0,6),
                       no_iterations = rep(0,6))

for (i in 1:6){
  output <- nr_algorithm(dlogl_d,mle_df_d[i,1])
  mle_df_d[i,2] = output$result
  mle_df_d[i,3] = output$no_iterations
}

mle_df_d
```

```
##   initial_x0               result no_iterations
## 1        0.1    0.903440114216673             9
## 2        0.2    0.903440114216673             6
## 3        0.3 N/A - No Convergence           544
## 4        0.4    0.903440114216814             6
## 5        0.6 N/A - No Convergence           544
## 6        0.8    0.903440114216679             8
```

Regardless of our initial guess for θ above, when the algorithm converges it does so in under 10 iterations and always converges to an MLE of about 0.903440114217. However, when we begin with an estimate of 0.3 or 0.6, the algorithm does not converge but instead gets stuck on the value of $-\infty$ around the 544th iteration.

Regardless, the estimates of the MLE that *do* emerge from convergent algorithm runs align with the results from the visualization and table in (5b).

Again, if the absolute difference between an estimate and the estimate in the prior iteration is less than 1e-6, we conclude that the algorithm has converged.