STAT 457 Homework No. 2

Tyler Maule

1/13/2022

Exercises without Code

Exercise 1

1a

Letting X denote the random variable associated with drawing marbles from the first urn and Y refer to the process associated with the second urn, we know that $X|\pi \sim Bin(n = 18, p = \pi)$ and $Y|\psi \sim Bin(n = 6, p = \psi)$.

Haldane's prior is $\pi, \psi \sim Beta(0, 0)$. In this case, given binomial likelihood the posterior distribution will be $\pi|X \sim Beta(18\bar{x}, 18(1-\bar{x}))$ and $\psi|Y \sim Beta(6\bar{y}, 6(1-\bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{18\bar{x}-1}{16}$ and $\hat{\psi} = \frac{6\bar{y}-1}{4}$. After evaluating given the data, the modes are (13/16) and (1/4), respectively.

Jeffrey's noninformative prior is $\pi, \psi \sim Beta(0.5, 0.5)$. Given the binomial likelihood the posterior distribution will be $\pi|X \sim Beta((1/2) + 18\bar{x}, (1/2) + 18(1-\bar{x}))$ and $\psi|Y \sim Beta((1/2) + 6\bar{y}, (1/2) + 6(1-\bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{(1/2)+18\bar{x}-1}{17}$ and $\hat{\psi} = \frac{(1/2)+6\bar{y}-1}{5}$. After evaluating given the data, the modes are (27/34) and (3/10), respectively.

On the flat (uniform) prior, $\pi, \psi \sim Unif(0, 1)$ or $\pi, \psi \sim Beta(1, 1)$. Given the binomial likelihood the posterior distribution will be $\pi|X \sim Beta((1) + 18\bar{x}, (1) + 18(1-\bar{x}))$ and $\psi|Y \sim Beta((1) + 6\bar{y}, (1) + 6(1-\bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{(1)+18\bar{x}-1}{18}$ and $\hat{\psi} = \frac{(1)+6\bar{y}-1}{7}$. After evaluating given the data, the modes are (14/18) and (2/6), respectively.

Asymptotically, the estimated log odds ratio converges to a normal distribution with mean $log(\frac{\hat{\pi}}{1-\hat{\pi}}) - log(\frac{\hat{\psi}}{1-\hat{\psi}})$ and standard error estimate $\sqrt{(1/18) + (1/6)}$. Thus under the null hypothesis that the log odds are equal for both urns, our test statistic $\frac{log(\frac{\hat{\pi}}{1-\hat{\pi}}) - log(\frac{\hat{\psi}}{1-\hat{\psi}})}{\sqrt{(1/18)+(1/6)}}$ will be distributed as a standard normal variable.

```
## $haldane_prob
## [1] 1
##
## $jeffreys_prob
## [1] 0.9999984
##
## $flat_prob
## [1] 0.9999817
```

Alternatively, using Fisher's information as an approximation of variance, the normal approximations are $\pi|X \sim N(\hat{\pi}, \sqrt{\hat{\pi}(1-\hat{\pi})/18})$ and $\psi|Y \sim N(\hat{\psi}, \sqrt{\hat{\psi}(1-\hat{\psi})/6})$. Then $(\pi|x - \psi|Y) \sim N(\hat{\pi} - \hat{\psi}, \sqrt{\hat{\pi}(1-\hat{\pi})/18}) + \sqrt{\hat{\psi}(1-\hat{\psi})/6})$. Hence the probabilities follow as:

```
## $haldane_prob
```

```
## [1] 0.981818
##
## $jeffreys_prob
## [1] 0.9599222
##
## $flat_prob
## [1] 0.9370214
```

1b

```
##             Prior Iterations    Means Variances Normal_Probability Exact_Prob
## 1 Haldane's Prior      1e+05 2.175637 1.2855602          0.9999980  0.9798100
## 2 Jeffrey's Prior      1e+05 1.933145 1.0509656          0.9999794  0.9758400
## 3      Flat Prior      1e+05 1.751887 0.9050235          0.9998989  0.9719600
## 4 Haldane's Prior      1e+06 2.179504 1.2865365          0.9999981  0.9790400
## 5 Jeffrey's Prior      1e+06 1.937892 1.0598429          0.9999803  0.9755860
## 6      Flat Prior      1e+06 1.751118 0.9063931          0.9998983  0.9716570
## 7 Haldane's Prior      1e+07 2.180464 1.2870260          0.9999981  0.9793155
## 8 Jeffrey's Prior      1e+07 1.935822 1.0597797          0.9999799  0.9753214
## 9      Flat Prior      1e+07 1.751577 0.9067230          0.9998987  0.9716932
```

1c

```
##
##  Fisher's Exact Test for Count Data
##
## data:  urn_df
## p-value = 0.9931
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##   0.00000 63.37908
## sample estimates:
## odds ratio
##   6.334078
```

Under different priors and the normal approximation which takes advantage of the distribution of the log odds ratio, the probability that the log odds of drawing a blue marble from the first urn is greater than the log odds of drawing a blue marble from the second urn ranges from 0.9999817 to 1.

Under different priors and the normal approximation which uses Fisher's information, the estimates of the same probability range from 0.9370214 to 0.981818.

Then our simulation results yield probability estimates between 0.971 and 0.98.

Finally, the estimate given by Fisher's exact test p-value is 0.9931.

Therefore, estimates that approximate the normal distribution with Fisher's information give the most conservative (lowest) probability (with the flat prior and Jeffrey's prior). The simulation and Fisher's information-based normal approximation (with Haldane's prior) give the next most conservative results, followed by the exact test. Lastly the normal approximation using the distribution of the log odds ratio gives the most liberal (highest) probability estimates.

Exercise 2

See below the derivation of the requested asymptotic distribution.

STAT 457
HW #5

2 $\hat{p} \sim N(p,\ p(1-p)/n)$ asymptotically

Find the asymptotic dist. of $2\sin^{-1}\sqrt{\hat{p}}$

let $g(\hat{\theta}) = 2\sin^{-1}(\sqrt{\hat{\theta}})$

Then $g'(\hat{\theta}) = (1-\hat{\theta})^{-1/2}(\hat{\theta})^{-1/2}$; $g'(\hat{\theta})$ exists and $g'(\hat{\theta}) \neq 0$ for some $\hat{\theta}$

Since we are given algebraic rearrangement of:

$$(\sqrt{n})(\hat{p}-p) \xrightarrow{\text{Dist.}} N(0,\ \sigma^2 = p(1-p))$$

By the univariate delta method we find that:

$$(\sqrt{n})(g(\hat{p})-g(p)) \xrightarrow{\text{Dist.}} N(0,\ p(1-p)\cdot(1-p)^{-1}\cdot p^{-1})$$

or $(\sqrt{n})(g(\hat{p})-g(p)) \xrightarrow{\text{Dist}} N(0,1)$

so $g(\hat{p}) = 2\arcsin(\sqrt{\hat{p}}) \xrightarrow{\text{Dist.}} N(2\arcsin(\sqrt{p}),\ (1/n))$

as desired.

Figure 1: Exercise 2

Exercise 3

See below the requested derivation of the ratio distribution.

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

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

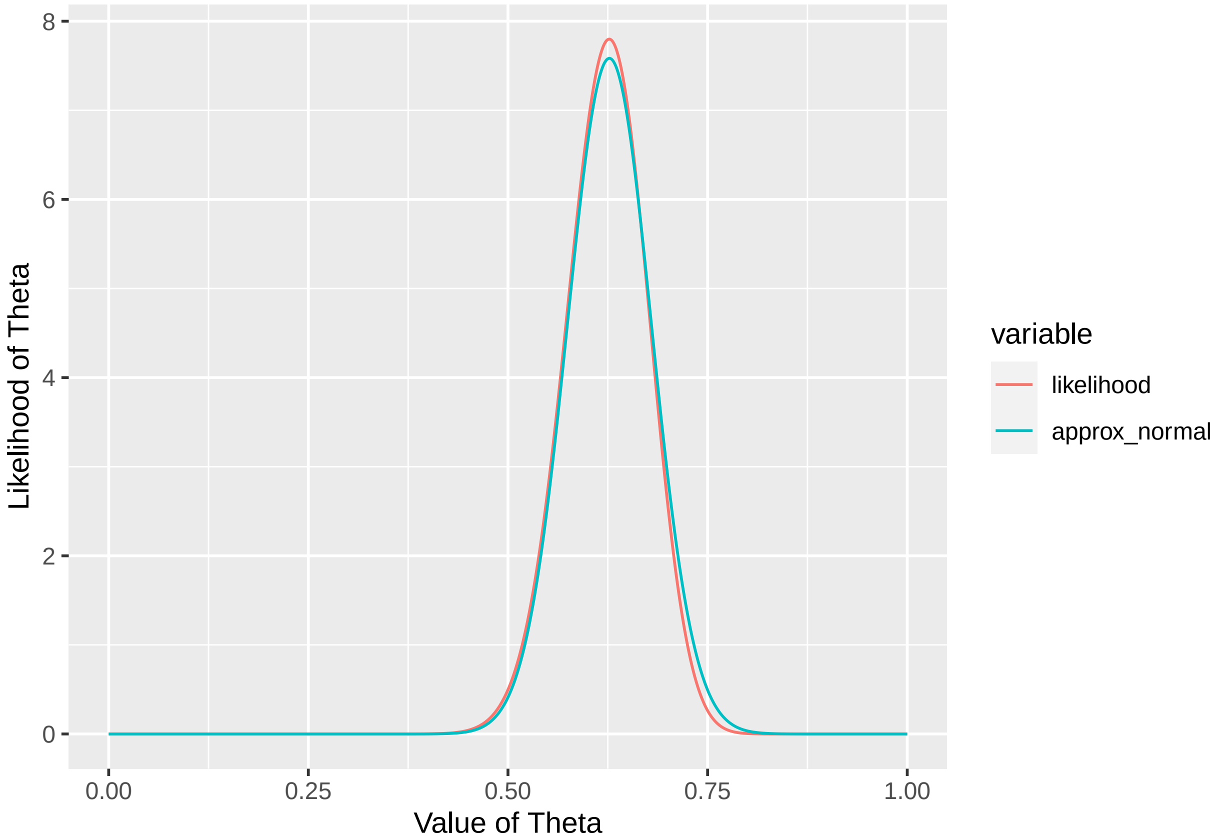
Figure 2: Exercise 3

Exercise 4

4a

See below the derivation of the normal approximation.

In (5a) of homework 4, we found that the MLE for the first data set is $\hat{\theta}_{MLE} = 0.627$. We'll use that value to construct a normal approximation of the likelihood, as explained above.

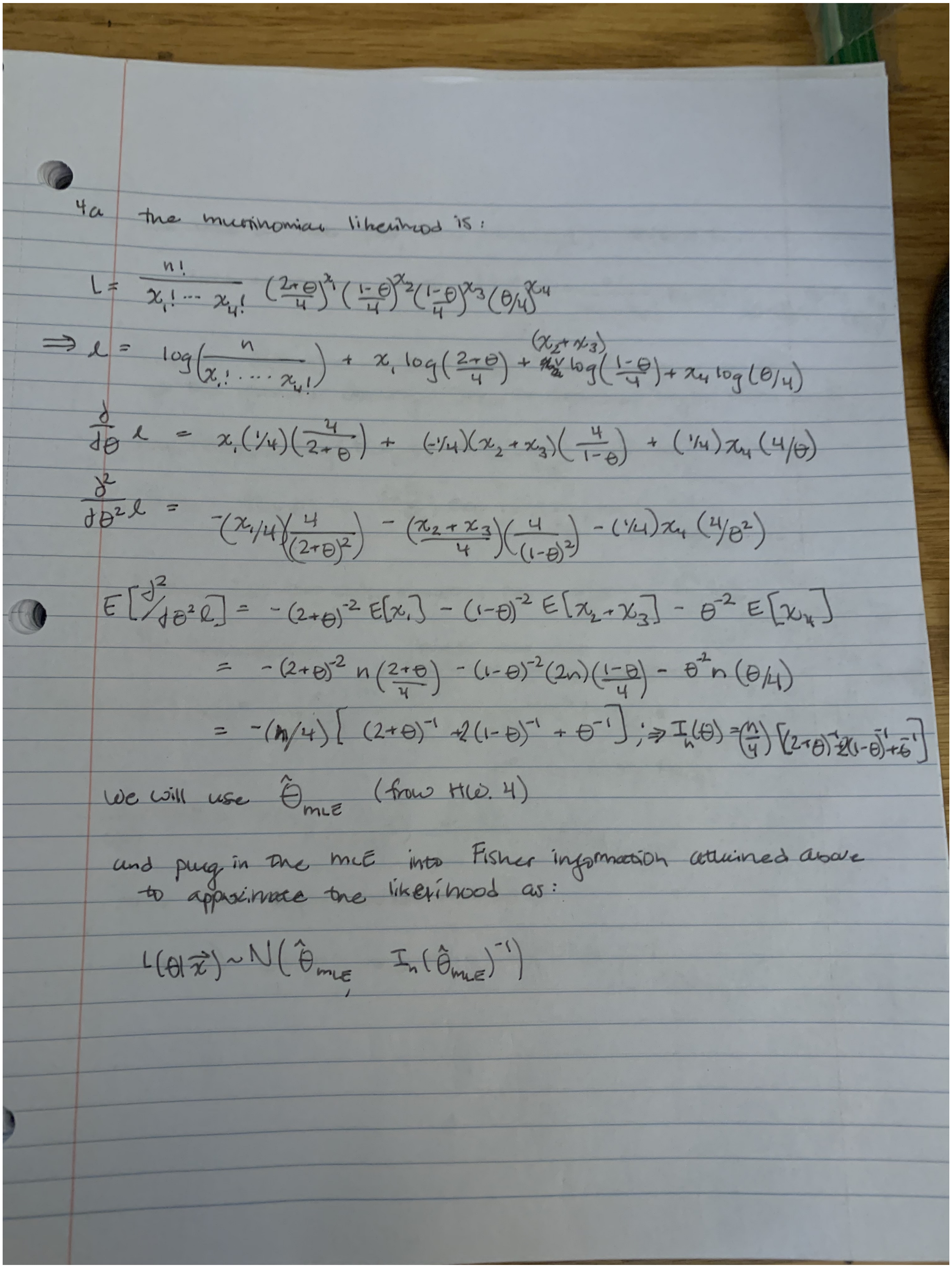



```
##   theta likelihood approx_normal
## 1 0.627   7.799261      7.583801
## 2 0.626   7.798315      7.582431
## 3 0.628   7.797261      7.582431
```

From the visualization, the normal approximation to the true likelihood appears to be a close match, although the approximation has slightly heavier tails (and thus a lower maximum likelihood). Still, both the true likelihood and the approximation estimate the MLE to be 0.627. So this seems like a great approximation.

4b

In (5b) of homework 4, we calculated that the MLE was $\hat{\theta}_{MLE} = 0.903$.



4a the multinomial likelihood is:

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

$$\Rightarrow \ell = \log\left(\frac{n}{x_1! \cdots x_4!}\right) + x_1 \log\left(\frac{2+\theta}{4}\right) + (x_2+x_3)\log\left(\frac{1-\theta}{4}\right) + x_4 \log(\theta/4)$$

$$\frac{\partial}{\partial\theta}\ell = x_1(1/4)\left(\frac{4}{2+\theta}\right) + (-1/4)(x_2+x_3)\left(\frac{4}{1-\theta}\right) + (1/4)x_4(4/\theta)$$

$$\frac{\partial^2}{\partial\theta^2}\ell = -(x_1/4)\left(\frac{4}{(2+\theta)^2}\right) - \left(\frac{x_2+x_3}{4}\right)\left(\frac{4}{(1-\theta)^2}\right) - (1/4)x_4(4/\theta^2)$$

$$E\left[\partial^2/\partial\theta^2 \ell\right] = -(2+\theta)^{-2}E[x_1] - (1-\theta)^{-2}E[x_2+x_3] - \theta^{-2}E[x_4]$$

$$= -(2+\theta)^{-2} n\left(\frac{2+\theta}{4}\right) - (1-\theta)^{-2}(2n)\left(\frac{1-\theta}{4}\right) - \theta^{-2} n(\theta/4)$$

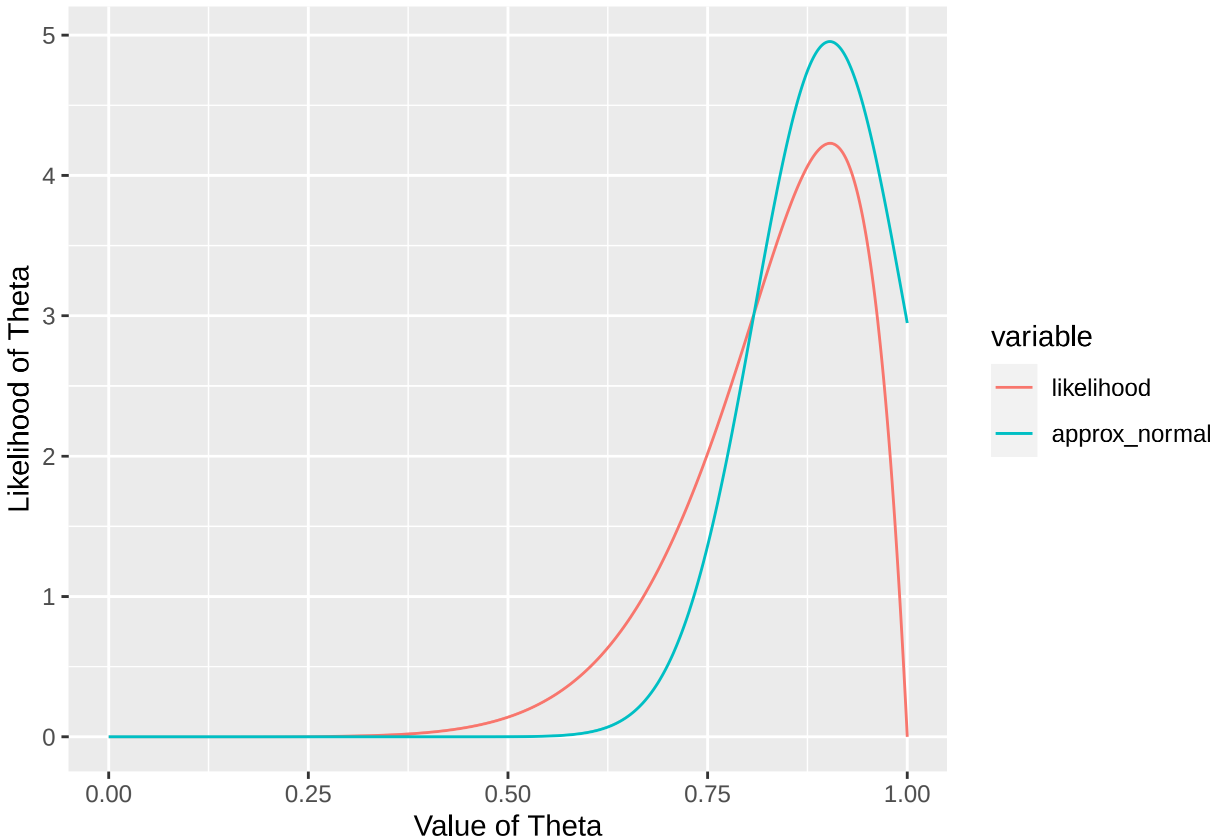
$$= -(n/4)\left[(2+\theta)^{-1} + 2(1-\theta)^{-1} + \theta^{-1}\right]; \Rightarrow I_n(\theta) = \left(\frac{n}{4}\right)\left[(2+\theta)^{-1} + 2(1-\theta)^{-1} + \theta^{-1}\right]$$

We will use $\hat{\theta}_{MLE}$ (from HW. 4)

and plug in the MLE into Fisher information attained above to approximate the likelihood as:

$$L(\theta|\vec{x}) \sim N\left(\hat{\theta}_{MLE}, I_n(\hat{\theta}_{MLE})^{-1}\right)$$

Figure 3: Exercise 4



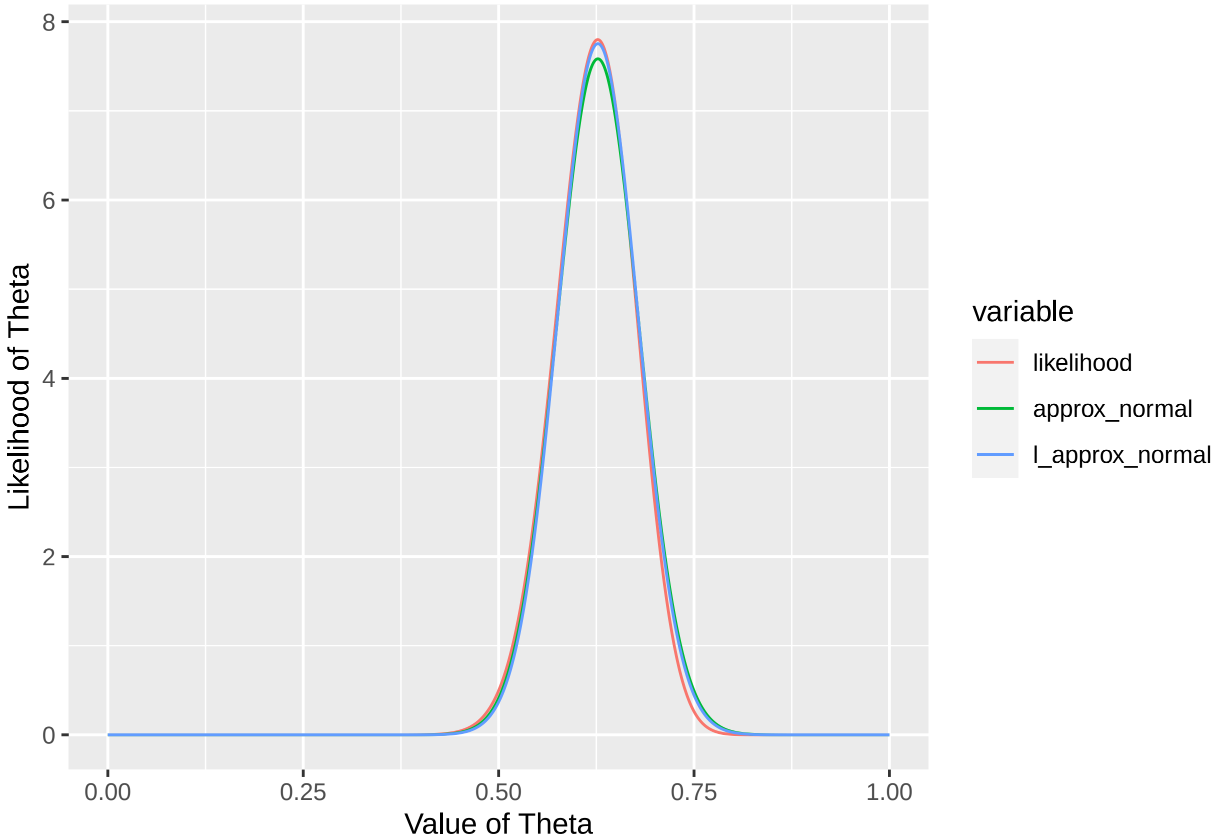

```
##   theta likelihood approx_normal
## 1 0.903   4.228978      4.954340
## 2 0.904   4.228948      4.954066
## 3 0.902   4.228525      4.954066
```

From the second visualization, the normal approximation to the true likelihood appears to be a poor match. In this case, the true likelihood has much heavier tails, whereas the approximation underestimates likelihood of lower theta values and overestimates likelihoods of higher theta values. Nevertheless, both the true likelihood and the approximation estimate the MLE to be 0.903. This seems like a much weaker approximation, and it seems that the sample size is to blame.

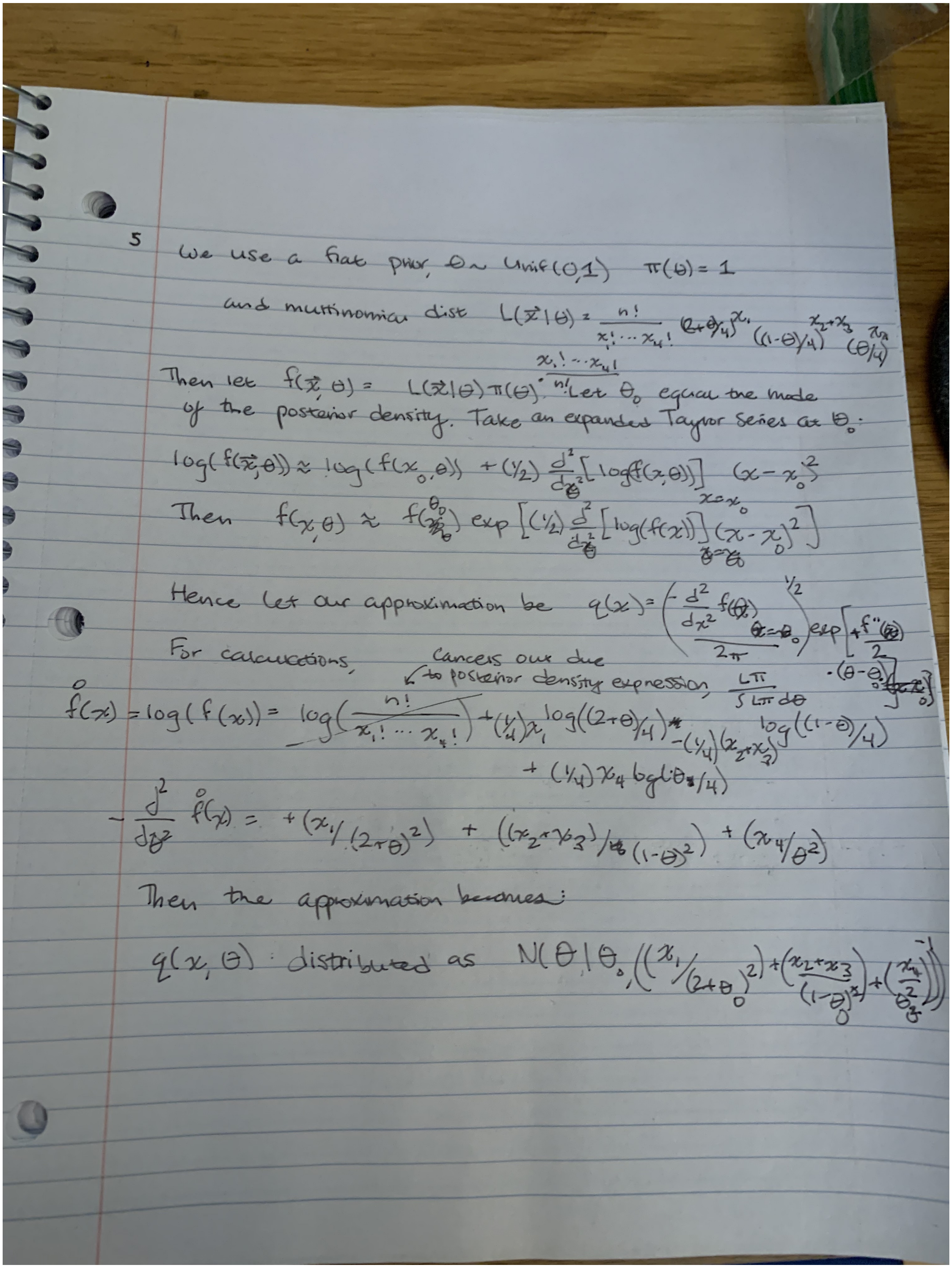
Exercise 5

See below the derivation related to Laplace's method (similar to the work shown in 4, above).

Based on the derivations calculated above, we find that the the appropriate approximations can be used as follows, with the approximation distributed as $N(\mu = 0.627, \sigma = 0.09362805)$:



The Laplace approximation shown above seems to perform a bit better than the normal approximation given in (4). Here our posterior mean is the mode of the likelihood, equaling 0.627.



5 We use a flat prior, $\theta \sim \mathrm{Unif}(0,1)$ $\quad \pi(\theta) = 1$

and multinomial dist $L(\vec{x}|\theta) = \frac{n!}{x_1! \cdots x_4!} ((2+\theta)/4)^{x_1} ((1-\theta)/4)^{x_2+x_3} (\theta/4)^{x_4}$

Then let $f(\vec{x},\theta) = L(\vec{x}|\theta)\pi(\theta) \cdot \frac{x_1! \cdots x_4!}{n!}$. Let θ_0 equal the mode of the posterior density. Take an expanded Taylor Series at θ_0:

$$\log(f(\vec{x},\theta)) \approx \log(f(x_0,\theta)) + (1/2)\frac{d^2}{d\theta^2}[\log f(x,\theta)]_{x=x_0} (x-x_0)^2$$

Then $$f(x,\theta) \approx f(\theta_0) \exp\left[(1/2)\frac{d^2}{d\theta^2}[\log(f(x))]_{\theta=\theta_0} (x-x_0)^2\right]$$

Hence let our approximation be $$q(x) = \left(\frac{-\frac{d^2}{dx^2} f(\theta)\big|_{\theta=\theta_0}}{2\pi}\right)^{1/2} \exp\left[+\frac{f''(\theta)}{2} \cdot (\theta-\theta_0)^2\right]$$

For calculations, cancels out due to posterior density expression, $\frac{L\pi}{\int L\pi\, d\theta}$

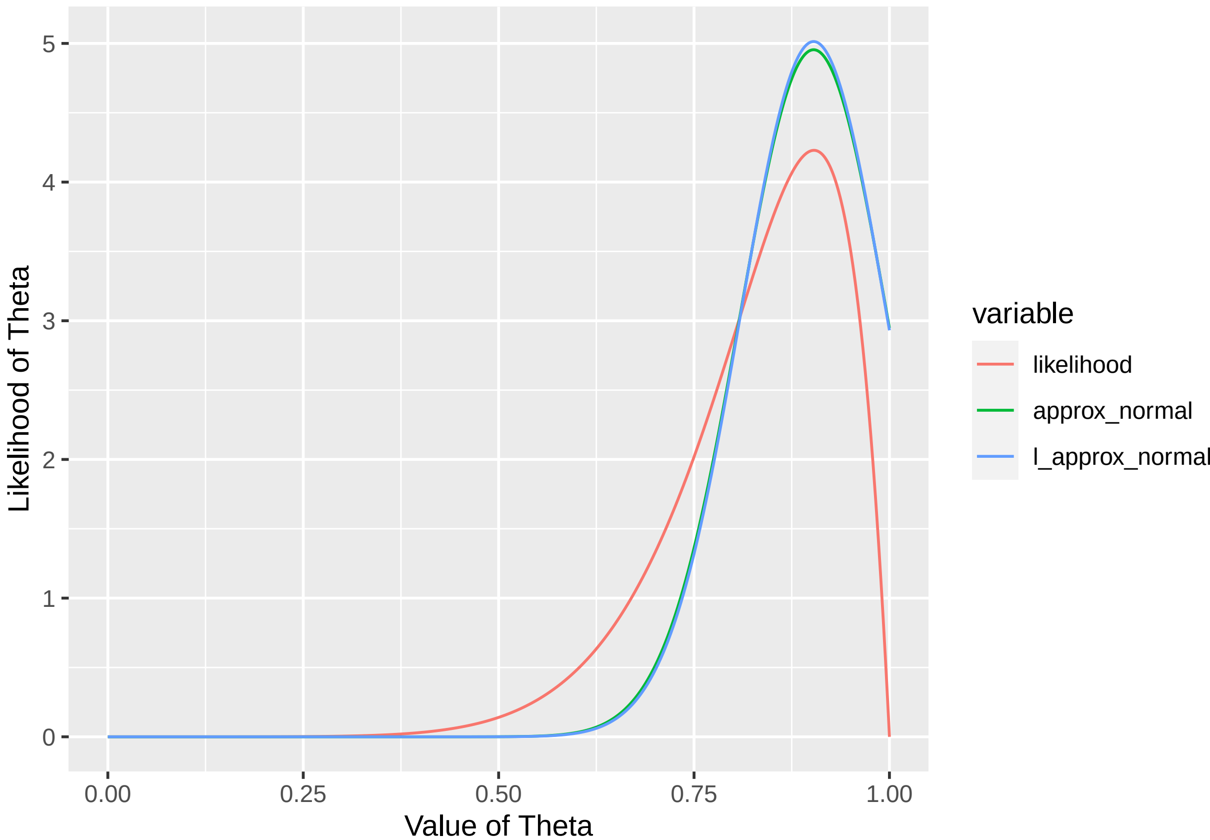
$$\mathring{f}(x) = \log(f(x)) = \cancel{\log\left(\frac{n!}{x_1! \cdots x_4!}\right)} + (1/4)x_1 \log((2+\theta)/4) - (1/4)(x_2+x_3)\log((1-\theta)/4) + (1/4)x_4 \log(\theta/4)$$

$$-\frac{d^2}{d\theta^2}\mathring{f}(x) = +(x_1/(2+\theta)^2) + ((x_2+x_3)/(1-\theta)^2) + (x_4/\theta^2)$$

Then the approximation becomes:

$q(x,\theta)$ distributed as $$N\left(\theta \,\middle|\, \theta_0, \left(\left(x_1/(2+\theta_0)^2\right) + \left(\frac{x_2+x_3}{(1-\theta_0)^2}\right) + \left(\frac{x_4}{\theta_0^2}\right)\right)^{-1}\right)$$

Figure 4: Exercise 5



The Laplace approximation shown above seems to perform just as poorly as the normal approximation given in (4). Here our posterior mean is the mode of the likelihood, equaling 0.903. Again, it seems like a higher sample size is key to getting a good approximation. No surprise there.

Exercises with Code

```
library(dplyr)
library(ggplot2)
library(reshape2)
```

Exercise 1

1a

Letting X denote the random variable associated with drawing marbles from the first urn and Y refer to the process associated with the second urn, we know that $X|\pi \sim Bin(n = 18, p = \pi)$ and $Y|\psi \sim Bin(n = 6, p = \psi)$.

Haldane's prior is $\pi, \psi \sim Beta(0, 0)$. In this case, given binomial likelihood the posterior distribution will be $\pi|X \sim Beta(18\bar{x}, 18(1 - \bar{x}))$ and $\psi|Y \sim Beta(6\bar{y}, 6(1 - \bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{18\bar{x}-1}{16}$ and $\hat{\psi} = \frac{6\bar{y}-1}{4}$. After evaluating given the data, the modes are (13/16) and (1/4), respectively.

Jeffrey's noninformative prior is $\pi, \psi \sim Beta(0.5, 0.5)$. Given the binomial likelihood the posterior distribution will be $\pi|X \sim Beta((1/2) + 18\bar{x}, (1/2) + 18(1 - \bar{x}))$ and $\psi|Y \sim Beta((1/2) + 6\bar{y}, (1/2) + 6(1 - \bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{(1/2)+18\bar{x}-1}{17}$ and $\hat{\psi} = \frac{(1/2)+6\bar{y}-1}{5}$. After evaluating given the data, the modes are (27/34) and (3/10), respectively.

On the flat (uniform) prior, $\pi, \psi \sim Unif(0,1)$ or $\pi, \psi \sim Beta(1,1)$. Given the binomial likelihood the posterior distribution will be $\pi|X \sim Beta((1) + 18\bar{x}, (1) + 18(1-\bar{x}))$ and $\psi|Y \sim Beta((1) + 6\bar{y}, (1) + 6(1-\bar{y}))$. Then the modes of these distributions are, respectively, $\hat{\pi} = \frac{(1)+18\bar{x}-1}{18}$ and $\hat{\psi} = \frac{(1)+6\bar{y}-1}{7}$. After evaluating given the data, the modes are (14/18) and (2/6), respectively.

Asymptotically, the estimated log odds ratio converges to a normal distribution with mean $log(\frac{\hat{\pi}}{1-\hat{\pi}}) - log(\frac{\hat{\psi}}{1-\hat{\psi}})$ and standard error estimate $\sqrt{(1/18) + (1/6)}$. Thus under the null hypothesis that the log odds are equal for both urns, our test statistic $\frac{log(\frac{\hat{\pi}}{1-\hat{\pi}}) - log(\frac{\hat{\psi}}{1-\hat{\psi}})}{\sqrt{(1/18)+(1/6)}}$ will be distributed as a standard normal variable.

```
haldane_test_stat <- (log((13/16)/(1-(13/16))) - log((1/4)/(1-(1/4))))/sqrt((1/18)+(1/6))
jeffreys_test_stat <- (log((27/34)/(1-(27/34))) - log((3/10)/(1-(3/10))))/sqrt((1/18)+(1/6))
flat_test_stat <- (log((7/9)/(1-(7/9))) - log((2/6)/(1-(2/6))))/sqrt((1/18)+(1/6))

list(haldane_prob =pnorm(haldane_test_stat),
     jeffreys_prob = pnorm(jeffreys_test_stat),
     flat_prob = pnorm(flat_test_stat))
```

```
## $haldane_prob
## [1] 1
##
## $jeffreys_prob
## [1] 0.9999984
##
## $flat_prob
## [1] 0.9999817
```

Alternatively, using Fisher's information as an approximation of variance, the normal approximations are $\pi|X \sim N(\hat{\pi}, \sqrt{\hat{\pi}(1-\hat{\pi})/18})$ and $\psi|Y \sim N(\hat{\psi}, \sqrt{\hat{\psi}(1-\hat{\psi})/6})$. Then $(\pi|x - \psi|Y) \sim N(\hat{\pi} - \hat{\psi}, \sqrt{\hat{\pi}(1-\hat{\pi})/18}) + \sqrt{\hat{\psi}(1-\hat{\psi})/6})$. Hence the probabilities follow as:

```
calc_prob <- function(pi_est,psi_est){
  return (pnorm((pi_est - psi_est)/(sqrt((pi_est*(1-pi_est))/18)+sqrt((psi_est * (1-psi_est))/6))))
}

list(haldane_prob=calc_prob((13/16),(1/4)),jeffreys_prob=calc_prob((27/34),(3/10)),flat_prob=calc_prob(
```

```
## $haldane_prob
## [1] 0.981818
##
## $jeffreys_prob
## [1] 0.9599222
##
## $flat_prob
## [1] 0.9370214
```

1b

```
logit <- function(x){
  return (log(x/(1-x)))
}

draw_bdeviates <- function(iterations){
  draws <- 1
```

```r
  haldane_istat <- numeric(iterations)
  jeffreys_istat <- numeric(iterations)
  flat_istat <- numeric(iterations)

  for (i in 1:iterations){
    haldane_post_pi <- rbeta(draws, 14, 4)
    haldane_post_psi <- rbeta(draws, 2, 4)
    haldane_istat[i] <- mean(logit(haldane_post_pi) - logit(haldane_post_psi))

    jeffreys_post_pi <- rbeta(draws, 14.5, 4.5)
    jeffreys_post_psi <- rbeta(draws, 2.5, 4.5)
    jeffreys_istat[i] <- mean(logit(jeffreys_post_pi) - logit(jeffreys_post_psi))

    flat_post_pi <- rbeta(draws, 15, 5)
    flat_post_psi <- rbeta(draws, 3, 5)
    flat_istat[i] <- mean(logit(flat_post_pi) - logit(flat_post_psi))
  }
  return(list(h=haldane_istat,j=jeffreys_istat,f=flat_istat))
}

se_estimate <- sqrt((1/18)+(1/6))
results_df <- data.frame(Prior=c("Haldane's Prior","Jeffrey's Prior","Flat Prior",
                                 "Haldane's Prior","Jeffrey's Prior","Flat Prior",
                                 "Haldane's Prior","Jeffrey's Prior","Flat Prior"),
                         Iterations=c(1e5,1e5,1e5,1e6,1e6,1e6,1e7,1e7,1e7),
                         Means = numeric(9), Variances = numeric(9), Normal_Probability =
                           numeric(9), Exact_Prob = numeric(9))

for (j in 0:2){
  k <- 1

  dresults <- draw_bdeviates(results_df[((j+1)*3),2])

  for (prior in c(dresults["h"],dresults["j"],dresults["f"])){
    results_df[(k + 3 * j),3] <- mean(prior)
    results_df[(k + 3 * j),4] <- var(prior)
    results_df[(k + 3 * j),5] <- pnorm(mean(prior)/se_estimate)
    results_df[(k + 3 * j),6] <- (length(prior[prior > 0]))/length(prior)

    k <- k + 1
  }
}

results_df
```

```
##             Prior Iterations    Means Variances Normal_Probability Exact_Prob
## 1 Haldane's Prior      1e+05 2.182147 1.2889313          0.9999982  0.9792600
## 2 Jeffrey's Prior      1e+05 1.934380 1.0459980          0.9999796  0.9749900
## 3      Flat Prior      1e+05 1.751947 0.9102346          0.9998990  0.9713100
## 4 Haldane's Prior      1e+06 2.180158 1.2877416          0.9999981  0.9791890
## 5 Jeffrey's Prior      1e+06 1.937805 1.0604034          0.9999803  0.9753590
## 6      Flat Prior      1e+06 1.750867 0.9073893          0.9998981  0.9715730
## 7 Haldane's Prior      1e+07 2.180084 1.2861561          0.9999981  0.9792738
```

```
## 8 Jeffrey's Prior      1e+07 1.936024 1.0590181          0.9999800  0.9754296
## 9      Flat Prior      1e+07 1.751139 0.9074028          0.9998983  0.9717326
```

1c

```r
urn_df <- data.frame(
  "Blue" = c(14, 2),
  "Not Blue" = c(4, 4)
)

fisher.test(urn_df,alternative="less")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  urn_df
## p-value = 0.9931
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##   0.00000 63.37908
## sample estimates:
## odds ratio
##   6.334078
```

Under different priors and the normal approximation which takes advantage of the distribution of the log odds ratio, the probability that the log odds of drawing a blue marble from the first urn is greater than the log odds of drawing a blue marble from the second urn ranges from 0.9999817 to 1.

Under different priors and the normal approximation which uses Fisher's information, the estimates of the same probability range from 0.9370214 to 0.981818.

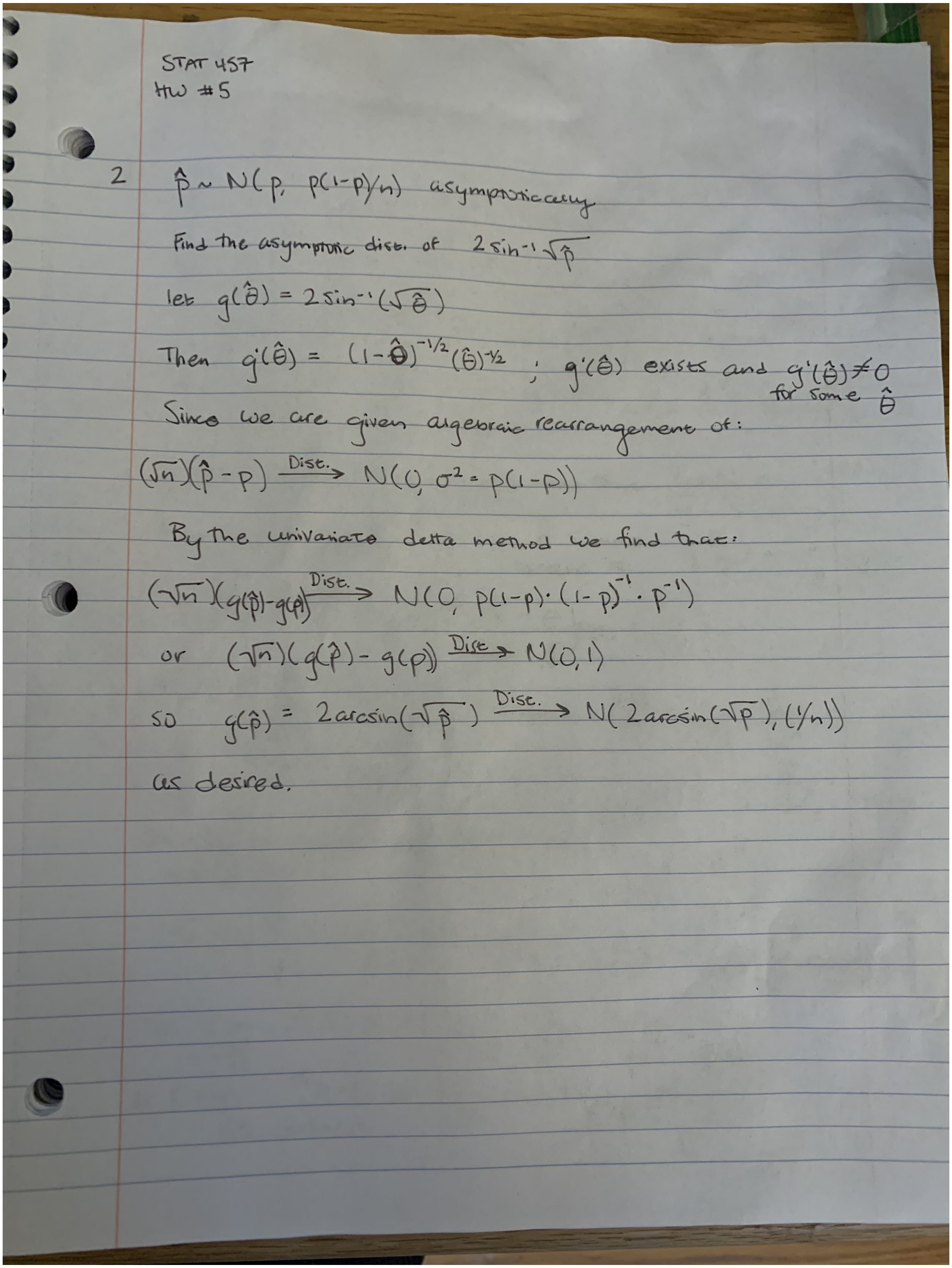
Then our simulation results yield probability estimates between 0.971 and 0.98.

Finally, the estimate given by Fisher's exact test p-value is 0.9931.

Therefore, estimates that approximate the normal distribution with Fisher's information give the most conservative (lowest) probability (with the flat prior and Jeffrey's prior). The simulation and Fisher's information-based normal approximation (with Haldane's prior) give the next most conservative results, followed by the exact test. Lastly the normal approximation using the distribution of the log odds ratio gives the most liberal (highest) probability estimates.

Exercise 2

See below the derivation of the requested asymptotic distribution.



STAT 457
HW #5

2 $\hat{p} \sim N(p, p(1-p)/n)$ asymptotically

Find the asymptotic dist. of $2\sin^{-1}\sqrt{\hat{p}}$

let $g(\hat{\theta}) = 2\sin^{-1}(\sqrt{\hat{\theta}})$

Then $g'(\hat{\theta}) = (1-\hat{\theta})^{-1/2}(\hat{\theta})^{-1/2}$; $g'(\hat{\theta})$ exists and $g'(\hat{\theta}) \neq 0$ for some $\hat{\theta}$

Since we are given algebraic rearrangement of:

$$(\sqrt{n})(\hat{p}-p) \xrightarrow{\text{Dist.}} N(0, \sigma^2 = p(1-p))$$

By the univariate delta method we find that:

$$(\sqrt{n})(g(\hat{p})-g(p)) \xrightarrow{\text{Dist.}} N(0, p(1-p)\cdot(1-p)^{-1}\cdot p^{-1})$$

or $(\sqrt{n})(g(\hat{p}) - g(p)) \xrightarrow{\text{Dist}} N(0,1)$

so $g(\hat{p}) = 2\arcsin(\sqrt{\hat{p}}) \xrightarrow{\text{Dist.}} N(2\arcsin(\sqrt{p}), (1/n))$

as desired.

Figure 5: Exercise 2

Exercise 3

See below the requested derivation of the ratio distribution.

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

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

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

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

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

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

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

Figure 6: Exercise 3

Exercise 4

4a

See below the derivation of the normal approximation.

In (5a) of homework 4, we found that the MLE for the first data set is $\hat{\theta}_{MLE} = 0.627$. We'll use that value to construct a normal approximation of the likelihood, as explained above.

```
n <- 197
mle <- 0.627
sd_approx <- sqrt(1/((n/4)*((1/(2+mle))+(2/(1-mle))+(1/mle))))

l_normalizer_func <- function(theta){
  ((prod(126:197))/(factorial(18)*factorial(20)*factorial(34)))*((2+theta)/4)**125 * ((1-theta)/4)**18 *
}

a_normalizer_c <- integrate(dnorm,0,1,mean=mle,sd=sd_approx)$value

l_normalizing_c <- integrate(l_normalizer_func,0,1)$value

like5a <- data.frame(theta = seq(0,1,0.001))
like5a <- mutate(like5a,likelihood=l_normalizer_func(theta))
like5a$likelihood <- like5a$likelihood/l_normalizing_c
like5a$approx_normal <- dnorm(like5a$theta,mean=0.627,sd=sd_approx)/a_normalizer_c

like5a_melted <- melt(like5a,id="theta")
like5a_melted <- like5a_melted %>% filter((like5a_melted$variable=="likelihood") | (like5a_melted$varial

ggplot(data=like5a_melted,aes(x=theta, y=value, group=variable)) +
  geom_line((aes(color=variable))) +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```

4a the multinomial likelihood is:

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

$$\Rightarrow \ell = \log\left(\frac{n}{x_1! \cdots x_4!}\right) + x_1 \log\left(\frac{2+\theta}{4}\right) + (x_2+x_3)\log\left(\frac{1-\theta}{4}\right) + x_4 \log(\theta/4)$$

$$\frac{\partial}{\partial\theta}\ell = x_1(1/4)\left(\frac{4}{2+\theta}\right) + (-1/4)(x_2+x_3)\left(\frac{4}{1-\theta}\right) + (1/4)x_4(4/\theta)$$

$$\frac{\partial^2}{\partial\theta^2}\ell = -(x_1/4)\left(\frac{4}{(2+\theta)^2}\right) - \left(\frac{x_2+x_3}{4}\right)\left(\frac{4}{(1-\theta)^2}\right) - (1/4)x_4(4/\theta^2)$$

$$E\left[\partial^2/\partial\theta^2 \ell\right] = -(2+\theta)^{-2}E[x_1] - (1-\theta)^{-2}E[x_2+x_3] - \theta^{-2}E[x_4]$$

$$= -(2+\theta)^{-2}n\left(\frac{2+\theta}{4}\right) - (1-\theta)^{-2}(2n)\left(\frac{1-\theta}{4}\right) - \theta^{-2}n(\theta/4)$$

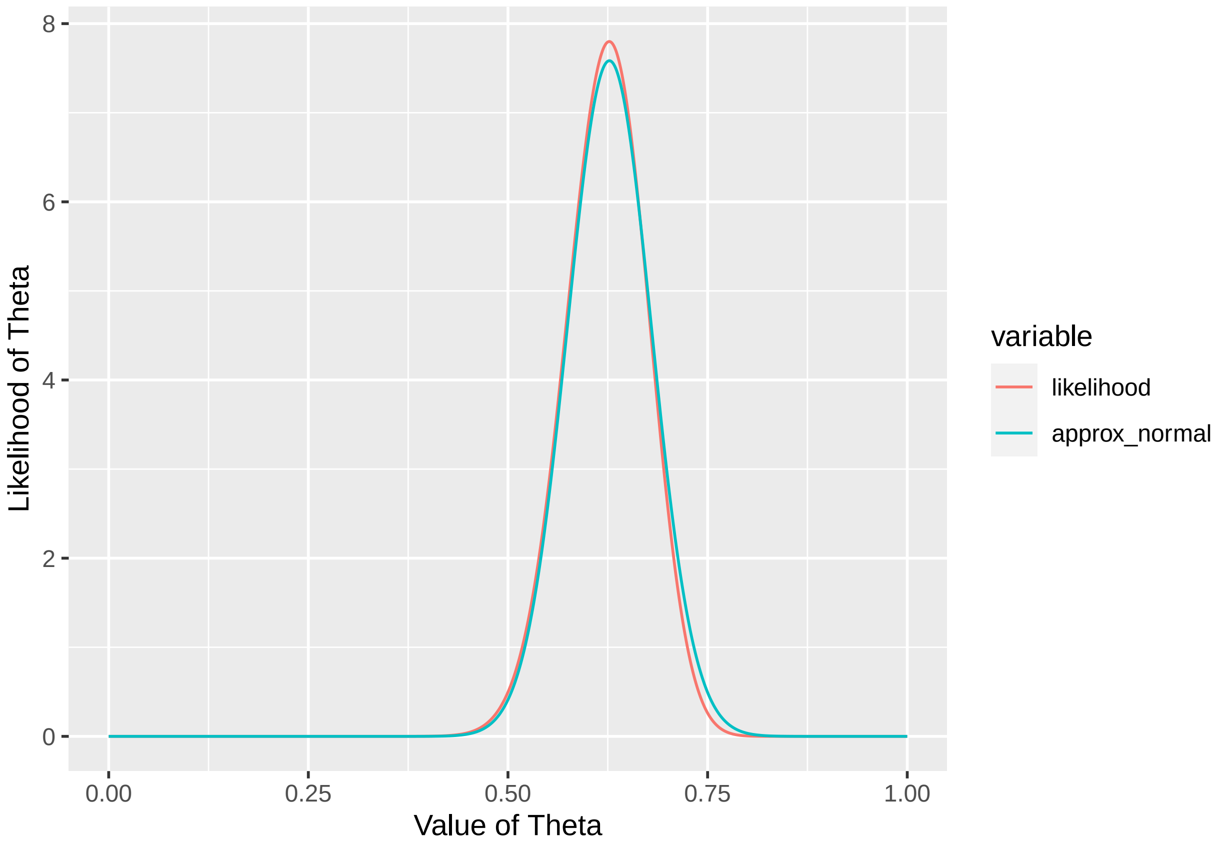
$$= -(n/4)\left[(2+\theta)^{-1} + 2(1-\theta)^{-1} + \theta^{-1}\right]; \Rightarrow I_n(\theta) = \left(\frac{n}{4}\right)\left[(2+\theta)^{-1} + 2(1-\theta)^{-1} + \theta^{-1}\right]$$

We will use $\hat{\theta}_{MLE}$ (from HW. 4)

and plug in the MLE into Fisher information attained above to approximate the likelihood as:

$$L(\theta|\vec{x}) \sim N\left(\hat{\theta}_{MLE},\ I_n(\hat{\theta}_{MLE})^{-1}\right)$$

Figure 7: Exercise 4




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

```
##   theta likelihood approx_normal
## 1 0.627   7.799261      7.583801
## 2 0.626   7.798315      7.582431
## 3 0.628   7.797261      7.582431
```

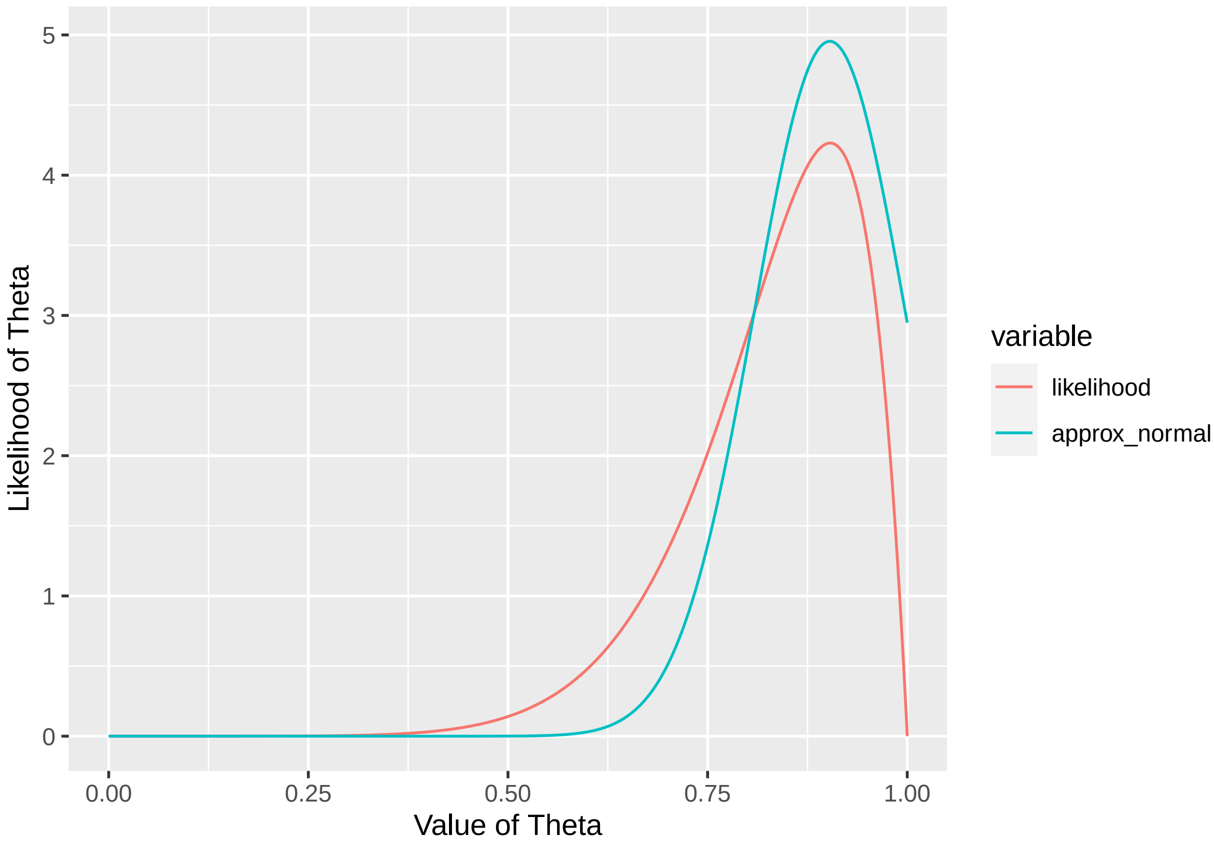
From the visualization, the normal approximation to the true likelihood appears to be a close match, although the approximation has slightly heavier tails (and thus a lower maximum likelihood). Still, both the true likelihood and the approximation estimate the MLE to be 0.627. So this seems like a great approximation.

4b

In (5b) of homework 4, we calculated that the MLE was $\hat{\theta}_{MLE} = 0.903$.

```
n <- 20
mle <- 0.903
sd_approx <- sqrt(1/((n/4)*((1/(2+mle))+(2/(1-mle))+(1/mle))))

l_normalizer_func <- function(theta){
  ((factorial(20))/(factorial(14)*factorial(0)*factorial(1)*factorial(5)))*((2+theta)/4)**14 * ((1-thet
}

a_normalizer_c <- integrate(dnorm,0,1,mean=mle,sd=sd_approx)$value

l_normalizing_c <- integrate(l_normalizer_func,0,1)$value

like5b <- data.frame(theta = seq(0,1,0.001))
```

```r
like5b <- mutate(like5b,likelihood=l_normalizer_func(theta))
like5b$likelihood <- like5b$likelihood/l_normalizing_c
like5b$approx_normal <- dnorm(like5b$theta,mean=mle,sd=sd_approx)/a_normalizer_c

like5b_melted <- melt(like5b,id="theta")
like5b_melted <- like5b_melted %>% filter((like5b_melted$variable=="likelihood") | (like5b_melted$varia

ggplot(data=like5b_melted,aes(x=theta, y=value, group=variable)) +
  geom_line((aes(color=variable))) +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```




```r
like5b %>% arrange(desc(likelihood,approx_normal)) %>% head(3)
```

```
##   theta likelihood approx_normal
## 1 0.903   4.228978      4.954340
## 2 0.904   4.228948      4.954066
## 3 0.902   4.228525      4.954066
```

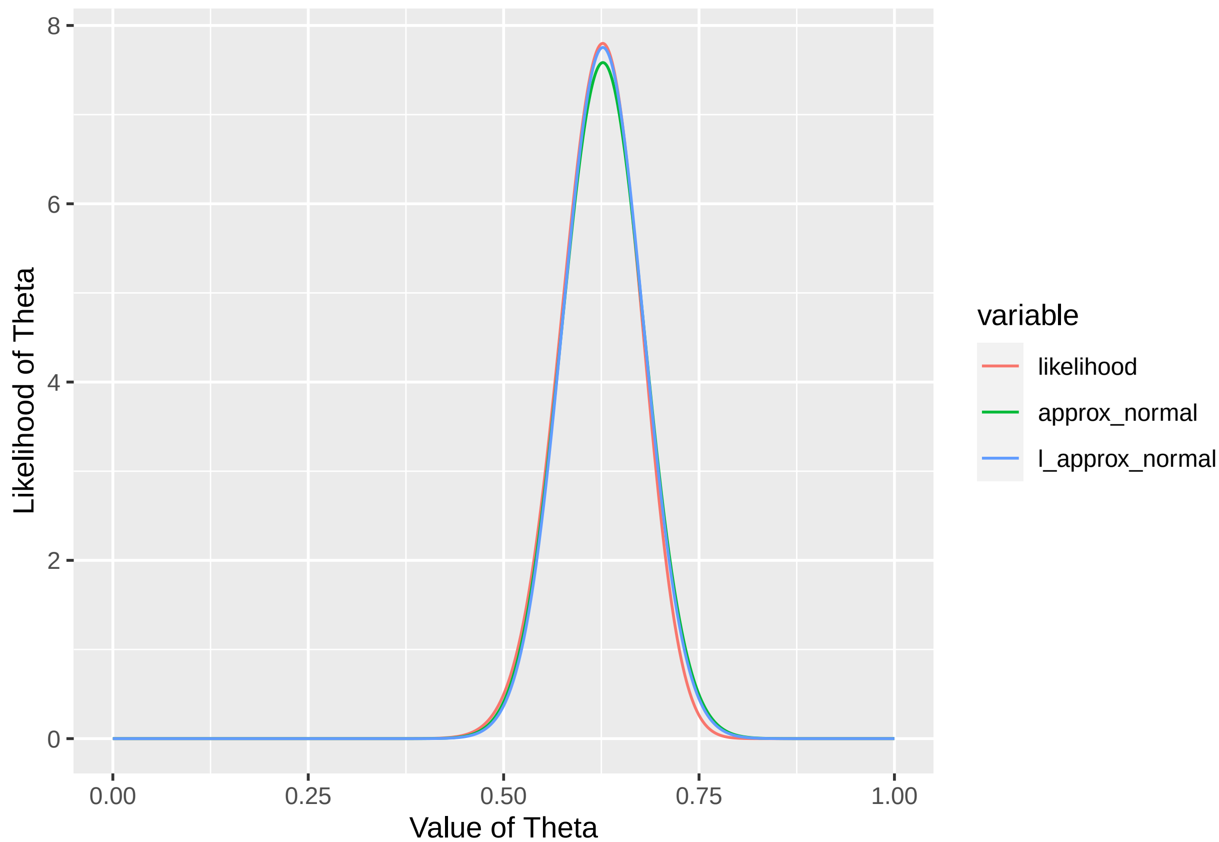
From the second visualization, the normal approximation to the true likelihood appears to be a poor match. In this case, the true likelihood has much heavier tails, whereas the approximation underestimates likelihood of lower theta values and overestimates likelihoods of higher theta values. Nevertheless, both the true likelihood and the approximation estimate the MLE to be 0.903. This seems like a much weaker approximation, and it seems that the sample size is to blame.

Exercise 5

See below the derivation related to Laplace's method (similar to the work shown in 4, above).

Based on the derivations calculated above, we find that the the appropriate approximations can be used as follows, with the approximation distributed as $N(\mu = 0.627, \sigma = 0.09362805)$:

```
theta0 <- 0.627
x1 <- 125
x2 <- 18
x3 <- 20
x4 <- 34


sd_lapprox <- sqrt(1/((x1/((2+theta0)**2))+((x2+x3)/((1-theta0)**2)) + (x4/(theta0**2))))

laplace_nc <- integrate(dnorm,0,1,mean=theta0,sd=sd_lapprox)$value
like5a5 <- like5a
like5a5$l_approx_normal <- dnorm(like5a5$theta,mean=theta0,sd=sd_lapprox)/laplace_nc

like5a5_melted <- melt(like5a5,id="theta")
#like5a5_melted <- like5a5_melted %>% filter((like5a_melted$variable=="likelihood") | (like5a5_melted$v

ggplot(data=like5a5_melted,aes(x=theta, y=value, group=variable)) +
  geom_line((aes(color=variable))) +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```



The Laplace approximation shown above seems to perform a bit better than the normal approximation given

5 We use a flat prior, $\theta \sim \text{Unif}(0,1)$ $\quad \pi(\theta) = 1$

and multinomial dist $L(\vec{x}\mid\theta) = \frac{n!}{x_1! \cdots x_4!} \left((2+\theta)/4\right)^{x_1} \left((1-\theta)/4\right)^{x_2+x_3} \left(\theta/4\right)^{x_4}$

Then let $f(\vec{x},\theta) = L(\vec{x}\mid\theta)\pi(\theta) \cdot \frac{x_1! \cdots x_4!}{n!}$. Let θ_0 equal the mode of the posterior density. Take an expanded Taylor Series at θ_0:

$$\log(f(\vec{x},\theta)) \approx \log(f(x_0,\theta)) + (1/2)\frac{d^2}{d\theta^2}\left[\log f(x,\theta)\right]\Big|_{x=x_0} (x-x_0)^2$$

Then $$f(x,\theta) \approx f(\theta_0) \exp\left[(1/2)\frac{d^2}{d\theta^2}\left[\log(f(x))\right]\Big|_{\theta=\theta_0} (x-x_0)^2\right]$$

Hence let our approximation be $$q(x) = \left(\frac{-\frac{d^2}{dx^2} f(\theta)\big|_{\theta=\theta_0}}{2\pi}\right)^{1/2} \exp\left[+\frac{f''(\theta)}{2} \cdot (\theta-\theta_0)^2\right]$$

For calculations, cancels out due to posterior density expression, $\frac{L\pi}{\int L\pi \, d\theta}$

$$\mathring{f}(x) = \log(f(x)) = \cancel{\log\left(\frac{n!}{x_1! \cdots x_4!}\right)} + (1/4)x_1 \log((2+\theta)/4) - (1/4)(x_2+x_3)\log((1-\theta)/4) + (1/4)x_4 \log(\theta/4)$$

$$-\frac{d^2}{d\theta^2}\mathring{f}(x) = +\left(x_1/(2+\theta)^2\right) + \left((x_2+x_3)/(1-\theta)^2\right) + \left(x_4/\theta^2\right)$$

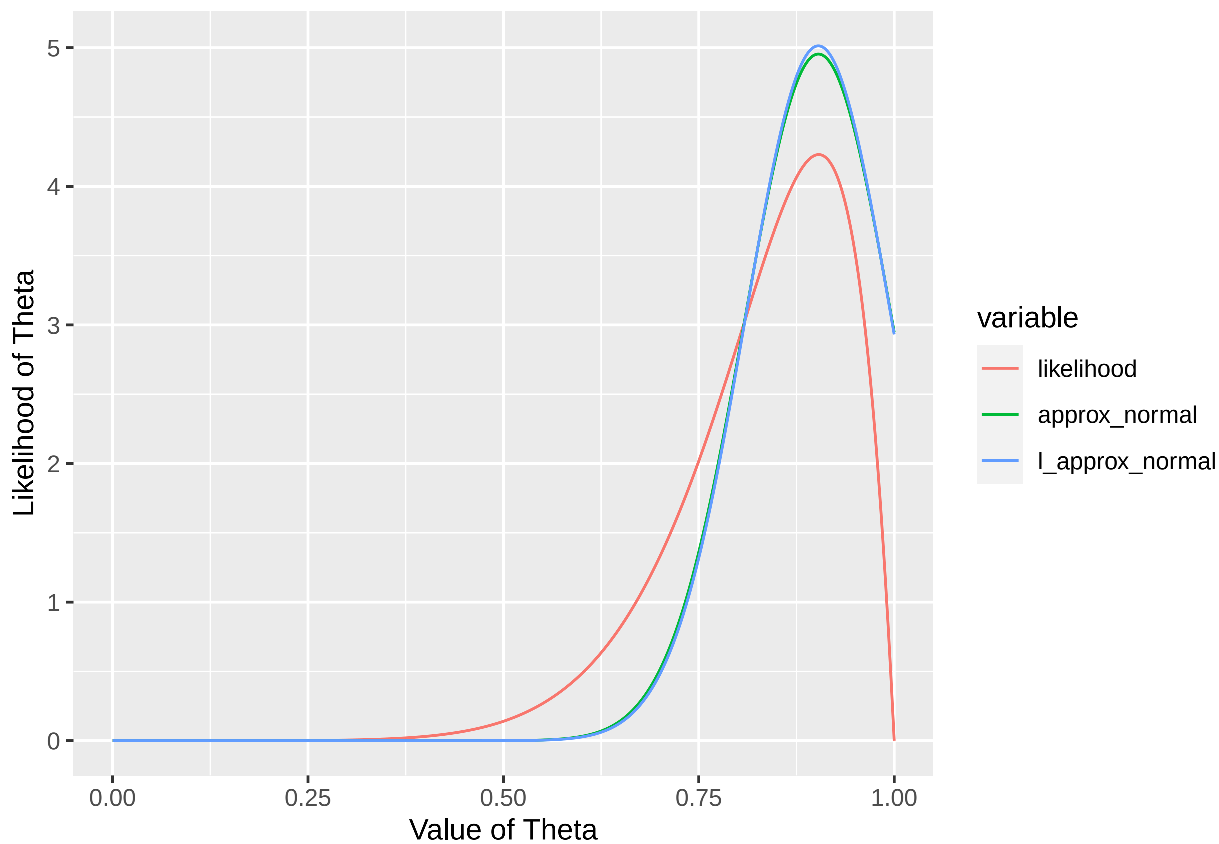
Then the approximation becomes:

$$q(x,\theta) \text{ distributed as } N\left(\theta \mid \theta_0, \left(\left(x_1/(2+\theta_0)^2\right) + \left(\frac{x_2+x_3}{(1-\theta_0)^2}\right) + \left(\frac{x_4}{\theta_0^2}\right)\right)^{-1}\right)$$

Figure 8: Exercise 5

in (4). Here our posterior mean is the mode of the likelihood, equaling 0.627.

```
theta0 <- 0.903
x1 <- 14
x2 <- 0
x3 <- 1
x4 <- 5


sd_lapprox <- sqrt(1/((x1/((2+theta0)**2))+((x2+x3)/((1-theta0)**2)) + (x4/(theta0**2))))

laplace_nc <- integrate(dnorm,0,1,mean=theta0,sd=sd_lapprox)$value
like5b5 <- like5b
like5b5$l_approx_normal <- dnorm(like5b5$theta,mean=theta0,sd=sd_lapprox)/laplace_nc

like5b5_melted <- melt(like5b5,id="theta")
#like5b5_melted <- like5b5_melted %>% filter((like5b_melted$variable=="likelihood") | (like5b5_melted$v

ggplot(data=like5b5_melted,aes(x=theta, y=value, group=variable)) +
  geom_line((aes(color=variable))) +
  xlab("Value of Theta") +
  ylab("Likelihood of Theta")
```



The Laplace approximation shown above seems to perform just as poorly as the normal approximation given in (4). Here our posterior mean is the mode of the likelihood, equaling 0.903. Again, it seems like a higher sample size is key to getting a good approximation. No surprise there.