Homework 5

Spring 2023

Jaiden Atterbury

Instructions

- This homework is due in Gradescope on Wednesday May 10 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.

Exercises

1. (Twins) Suppose that in a population of twins consisting only of the two most common biological sexes[1], males (those with XY chromosomes) and females (those with XX chromosomes) are equally likely and that the probability that the twins are identical is α_0. If the twins are not identical, their biological sexes are independently determined. If they are identical, their biological sex is obviously the same.

   a. Denote

   $$\pi_1 = P(MM),$$
   $$\pi_2 = P(FF),$$
   $$\pi_3 = P(MF).$$

   where MM denotes the event that both are male, FF the event that both are female and MF is the event that one of the twins is female and the other is male.

   Then, using the rules of probability, we can say

   $$P(MM) = P(FF) = (1 + \alpha_0)/4$$

   and

   $$P(MF) = (1 - \alpha_0)/2.$$

   Prove this result for $P(MM)$. Be sure to show and justify your steps. **Since I am basically setting up the problem formulation for you, we are going to assess mastery of the concepts.**

   Hint: Label the members of a pair of twins as 1 and 2. Let T_1 denote the event that twin 1 is M, T_2 is the event that twin 2 is M. And let I denote the event that the twins are identical. Then the event "MM" is the union of two disjoint events:

   $$MM = (T_1 \cap T_2 \cap I) \cup (T_1 \cap T_2 \cap I^c)$$

   Review chapters 2 and 3 if you need a reminder

[1] see here for the six variations that are possible

Suppose that we are looking at a population of twins consisting only of the two most common biological males, those with XY chromosomes, and females, those with XX chromosomes. Furthermore, the probability of seeing either male or female is equally likely and the probability that the twins are identical is α_0. If the twins are not identical, their biological sexes are independently determined. If they are identical, their biological sex is the same.

If we let MM denote the event that both twins are male, it follows that $P(MM) = \frac{1+\alpha_0}{4}$. Our goal is to prove this result.

To start off if we label the members of a pair of twins as 1 and 2 and let T_1 denote the event that twin 1 is M, and let T_2 denote the event that twin 2 is M, and lastly let I denote the event that the twins are identical. Then the event "MM" is the union of two disjoint events: $MM = (T_1 \cap T_2 \cap I) \cup (T_1 \cap T_2 \cap I^c)$.

Below we will compute $P(MM)$.

$$\begin{aligned} P(MM) &= P((T_1 \cap T_2 \cap I) \cup (T_1 \cap T_2 \cap I^c)) \\ &= P(T_1 \cap T_2 \cap I) + P(T_1 \cap T_2 \cap I^c) \quad \text{(Definition 2.1.A3)} \end{aligned}$$

Finding $\mathbf{P(T_1 \cap T_2 \cap I)}$:
If we let $T_1 \cap I = E$, then it follows that $P(T_1 \cap T_2 \cap I) = P(T_1 \cap E)$. Hence by the chain rule of probability, $P(T_1 \cap E) = P(T_1|E) \cdot P(E)$. Replacing E with $T_1 \cap I$, we obtain $P(T_1 \cap T_2 \cap I) = P(T_1|T_1 \cap I) \cdot P(T_1 \cap I)$. Furthermore, multiplying and dividing the previous statement by $P(I)$, we obtain $\frac{P(T_1|T_1 \cap I) \cdot P(T_1 \cap I) P(I)}{P(I)}$. By Definition 4.1, this simplifies to $P(T_1|T_2 \cap I) \cdot P(T_2|I) \cdot P(I)$.

As stated in the problem statement $P(I) = \alpha_0$. Likewise, $P(T_2|I) = \frac{1}{2}$ since the probability of a given twin being male or female is equally likely and there are only two sexes to choose from. Lastly, $P(T_1|T_2 \cap I) = 1$, because if two twins are identical and one of them is male, then we know that the other ones sex is also male. Thus $P(T_1 \cap T_2 \cap I) = \frac{1}{2} \cdot 1 \cdot \alpha_0 = \frac{\alpha_0}{2}$

Finding $\mathbf{P(T_1 \cap T_2 \cap I^c)}$:
If we let $T_1 \cap I^c = E$, then it follows that $P(T_1 \cap T_2 \cap I^c) = P(T_1 \cap E)$. Hence by the chain rule of probability, $P(T_1 \cap E) = P(T_1|E) \cdot P(E)$. Replacing E with $T_1 \cap I^c$, we obtain $P(T_1 \cap T_2 \cap I^c) = P(T_1|T_1 \cap I^c) \cdot P(T_1 \cap I^c)$. Furthermore, multiplying and dividing the previous statement by $P(I^c)$, we obtain $\frac{P(T_1|T_1 \cap I^c) \cdot P(T_1 \cap I^c) P(I^c)}{P(I^c)}$. By Definition 4.1, this simplifies to $P(T_1|T_2 \cap I^c) \cdot P(T_2|I^c) \cdot P(I^c)$.

As stated in the problem statement $P(I) = \alpha_0$, hence by Theorem 2.1.a, $P(I^c) = 1 - \alpha_0$. Likewise, since the probability of a given twin being male or female is equally likely and there are only two sexes to choose from, it follows that $P(T_2|I^c) = \frac{1}{2}$. Similarly, since T_1 and T_2 are independent when the twins are fraternal/not identical, it turns out that $P(T_1|T_2 \cap I^c)$ simply turns into the probability of a given twin being male or female, which we know is equally likely since there are only two sexes to choose from, thus $P(T_1|T_2 \cap I^c) = \frac{1}{2}$. Thus $P(T_1 \cap T_2 \cap I^c) = \frac{1}{2} \cdot \frac{1}{2} \cdot \alpha_0 = \frac{1-\alpha_0}{4}$

Putting it all together:
Since $P((T_1 \cap T_2 \cap I) \cup (T_1 \cap T_2 \cap I^c)) = P(T_1 \cap T_2 \cap I) + P(T_1 \cap T_2 \cap I^c)$ by Definition 2.1.A3, then as computed in the above parts, $P(T_1 \cap T_2 \cap I) = \frac{\alpha_0}{2}$ and $P(T_1 \cap T_2 \cap I^c) = \frac{1-\alpha_0}{4}$. Hence $P(MM) = \frac{\alpha_0}{2} + \frac{1-\alpha_0}{4} = \frac{2\alpha_2}{4} + \frac{1-\alpha_0}{4} = \frac{1+\alpha_0}{4}$.

As shown above $P(MM) = \frac{1+\alpha_0}{4}$.

b. Let $\{X_1, X_2, X_3\}$ denote the number (out of n twins) of MM, FF and MF. Then a reasonable model is

$$\langle X_1, X_2, X_3 \rangle \sim Multinom(n, \pi = (\pi_1, \pi_2, \pi_3))$$

where π_1, π_2 and π_3 are functions of α_0 as described earlier.

Based on observing x_1 MM twins, x_2 FF twins, and x_3 MF twins, show that the maximum likelihood estimate of α_0 is

$$\widehat{\alpha}_0^{mle} \quad = \quad (x_1 + x_2 - x_3)/n.$$

You do not need to verify the second order condition.

If we let $\langle X_1, X_2, X_3\rangle$ denote the number out of n twins that are of the sex MM, FF and MF respectively, then a reasonable model for these counts is $\langle X_1, X_2, X_3\rangle \sim Multinom(n, \pi = \langle \pi_1, \pi_2, \pi_3\rangle)$. Where π_1, π_2 and π_3 are functions of α_0 as described earlier. Based on observing x_1 MM twins, x_2 FF twins, and x_3 MF twins, we will show that show that the maximum likelihood estimate of α_0 is $\widehat{\alpha}_0^{mle} = \frac{x_1+x_2-x_3}{n}$.

Based on Theorem 14.1, the PMF of $\langle X_1, X_2, X_3\rangle$ can be written as $f(x_1, x_2, x_3) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3} = \frac{n!}{x_1!x_2!x_3!}\left(\frac{1+\alpha_0}{4}\right)^{x_1}\left(\frac{1+\alpha_0}{4}\right)^{x_2}\left(\frac{1-\alpha_0}{2}\right)^{x_3},\ 0 \le \alpha_0 \le 1$. Where $x_1 + x_2 + x_3 = n$ and $\pi_1 + \pi_2 + \pi_3 = 1$. Based on this following PMF definition, it follows that the likelihood function of α_0 is $L(\alpha) = \frac{n!}{x_1!x_2!x_3!}\left(\frac{1+\alpha}{4}\right)^{x_1+x_2}\left(\frac{1-\alpha}{2}\right)^{x_3},\ 0 \le \alpha \le 1$. We will now find the log-likelihood function by taking the natural log of the preceding likelihood function.

$$\begin{aligned}
\ell(\alpha) &= \ln(L(\alpha)) \\
&= \ln\left(\frac{n!}{x_1!x_2!x_3!}\left(\frac{1+\alpha}{4}\right)^{x_1+x_2}\left(\frac{1-\alpha}{2}\right)^{x_3}\right) \\
&= \ln\left(\frac{n!}{x_1!x_2!x_3!}\right) + \ln\left(\left(\frac{1+\alpha}{4}\right)^{x_1+x_2}\right) + \ln\left(\left(\frac{1-\alpha}{2}\right)^{x_3}\right) \\
&= \ln(n!) - \ln(x_1!x_2!x_3!) + (x_1+x_2)\ln(1+\alpha) - (x_1+x_2)\ln(4) + x_3\ln(1-\alpha) - x_3\ln(2),\ 0 \le \alpha \le 1
\end{aligned}$$

Hence as computed above, $\ell(\alpha) = \ln(n!) - \ln(x_1!x_2!x_3!) + (x_1+x_2)\ln(1+\alpha) - (x_1+x_2)\ln(4) + x_3\ln(1-\alpha) - x_3\ln(2),\ 0 \le \alpha \le 1$. Now in order to find the candidates for the MLE of α_0, we must take the first derivative of the log-likelihood function and set it to zero.

$$\begin{aligned}
\frac{d}{d\alpha}\ell(\alpha) &= \frac{d}{d\alpha}\left(\ln(n!) - \ln(x_1!x_2!x_3!) + (x_1+x_2)\ln(1+\alpha) - (x_1+x_2)\ln(4) + x_3\ln(1-\alpha) - x_3\ln(2)\right) \\
&= \frac{x_1+x_2}{1+\alpha} - \frac{x_3}{1-\alpha},\ 0 \le \alpha \le 1
\end{aligned}$$

Setting this expression equal to zero and solving for α we obtain will obtain the MLE of α_0 (since we are ignoring the second order condition).

$$\begin{aligned}
0 &= \frac{x_1+x_2}{1+\alpha} - \frac{x_3}{1-\alpha} \\
\frac{x_3}{1-\alpha} &= \frac{x_1+x_2}{1+\alpha} \\
x_3(1+\alpha) &= (x_1+x_2)(1-\alpha) \\
x_3 + x_3\alpha &= x_1 + x_2 - x_1\alpha - x_2\alpha \\
x_1\alpha + x_2\alpha + x_3\alpha &= x_1 + x_2 - x_3 \\
(x_1+x_2+x_3)\alpha &= x_1 + x_2 - x_3 \\
\alpha &= \frac{x_1+x_2-x_3}{x_1+x_2+x_3} \\
\alpha &= \frac{x_1+x_2-x_3}{n} \quad \text{(Since } x_1+x_2+x_3 = n\text{)}
\end{aligned}$$

Hence as shown above, $\widehat{\alpha}_0^{mle} = \frac{x_1+x_2-x_3}{n}$.

c. Is $\widehat{\alpha}_0$ an unbiased estimator of α_0? Yes or no and show your work. (Hint: read Theorem 14.1 from chapter 14)

In order to see if $\widehat{\alpha}_0^{mle}$ is an unbiased estimator of α_0, we must find the expected value of $\widehat{\alpha}_0^{mle}$ and see if it is equal to α_0. In this derivation we will use Theorem 14.1 which states that the marginal distribution of each X_i in a multinomial distribution is $Binom(n, \pi_1)$.

$$\begin{aligned} E[\widehat{\alpha}_0^{mle}] &= E\left[\frac{X_1 + X_2 - X_3}{n}\right] \\ &= \frac{1}{n}E[X_1 + X_2 - X_3] \quad \text{(Linearity of Expectation)} \\ &= \frac{1}{n}(E[X_1] + E[X_2] - E[X_3]) \quad \text{(Linearity of Expectation)} \end{aligned}$$

Since each of the $X_i \sim Binom(n, \pi_i)$, it follows that $X_1 \sim Binom\left(n, \frac{1+\alpha_0}{4}\right)$, $X_2 \sim Binom\left(n, \frac{1+\alpha_0}{4}\right)$, and lastly, $X_3 \sim Binom\left(n, \frac{1-\alpha_0}{2}\right)$. Furthermore, using the fact that if $X \sim Binom(n, \pi_0)$, then $E[X] = n\pi_0$, we can find $E[\widehat{\alpha}_0^{mle}]$.

$$\begin{aligned} \frac{1}{n}(E[X_1] + E[X_2] - E[X_3]) &= \frac{1}{n}(n\pi_1 + n\pi_2 - n\pi_3) \\ &= \pi_1 + \pi_2 - \pi_3 \\ &= \frac{1+\alpha_0}{4} + \frac{1+\alpha_0}{4} + \frac{1-\alpha_0}{2} \\ &= \frac{2+2\alpha_0}{4} - \frac{2-2\alpha_0}{4} \\ &= \frac{2+2\alpha_0 - 2 + 2\alpha_0}{4} \\ &= \alpha_0 \end{aligned}$$

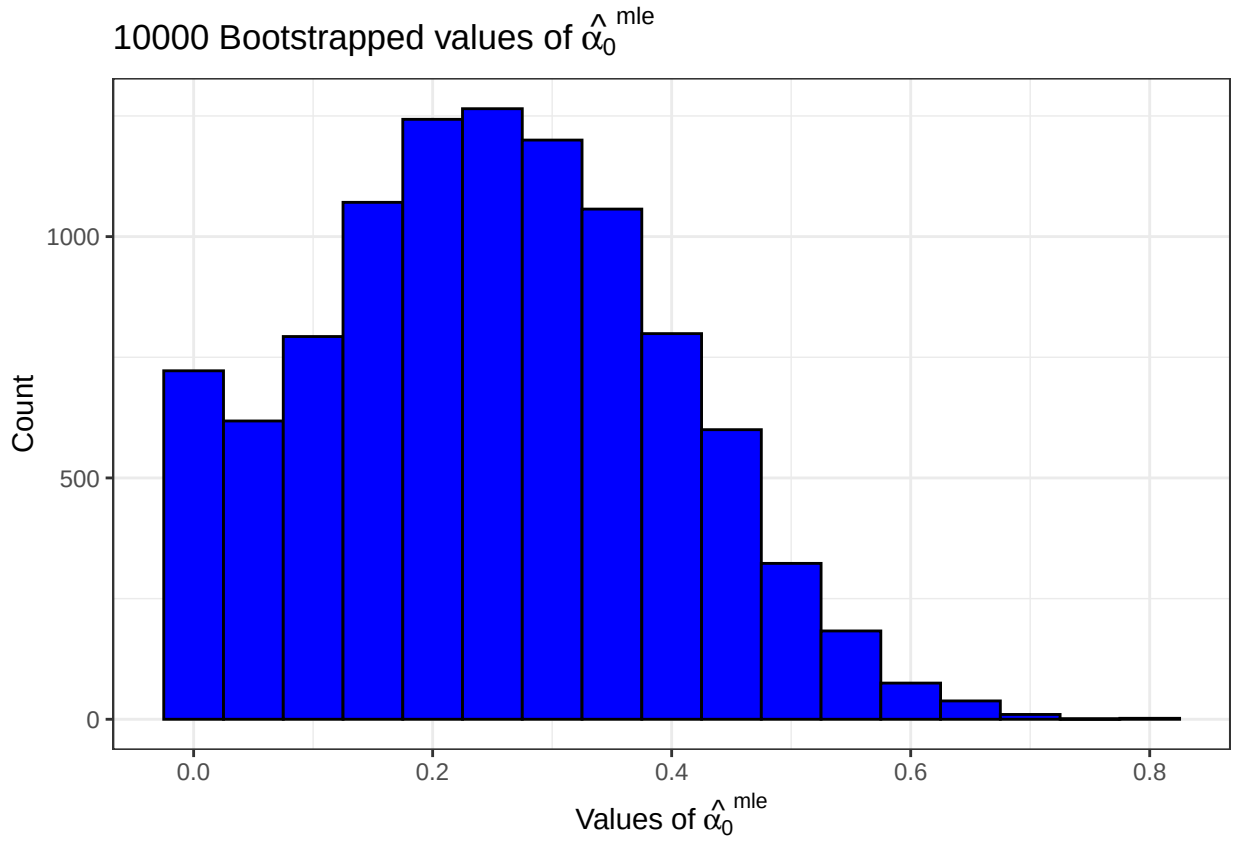
Since $E[\widehat{\alpha}_0^{mle}] = \alpha_0$ we can see that $\widehat{\alpha}_0^{mle}$ is an unbiased estimator of α_0.

d. The sampling distribution of α_0^{mle} is complicated because the counts x_i are dependent of each other as they have to sum to n. In this exercise, you will construct the distribution based on the following hypothetical data using the non-parametric bootstrap method.

MM	FF	MF
10	15	15

Fill in the partial code provided and echo it in the Appendix with clearly labeled section header.

i. Make a histogram of the bootstrapped sampling distribution of $\widehat{\alpha}_0^{mle}$. Don't forget those binwidths and label your axis/titles (using `expression` where necessary for math symbols. See Chapter 7 page 62 for a reminder)



As can be seen above, the distribution is slightly right skewed. However, if we allowed α_0 to take on negative values, then the bootstrapped sampling distribution of α_0 would be symmetric/approximately normal.

ii. Where is the bootstrapped sampling distribution centered? Does this make sense? Why?

As calculated above the bootstrapped sampling distribution is centered around 0.249645. Based on the data that we resampled from $\widehat{\alpha}_0^{mle} = 0.25$. Furthermore, in the previous part we showed that $\widehat{\alpha}_0^{mle}$ is an unbiased estimator of α_0. Hence, according to the bootstrapping idea, the bootstrapped sampling distribution represents the actual sampling distribution. Thus, since the actual sampling distribution is centered around α_0, we'd expect the bootstrapped sampling distribution to be around α_0. Therefore, since we took resamples from data with an estimated α_0 value of 0.25, we'd expect our sampling distribution to be centered around 0.25 (since the non-parametric bootstrap treats the sample as the population). Lastly, from a purely theoretical standpoint, we'd expect the probability of seeing identical twins to be much lower than that of seeing fraternal twins.

iii. Calculate and report (in context) a 95% bootstrap confidence interval for α_0. Calculate both types of intervals - standard and percentile - and report them in a neatly formatted table with headings and a title.

Below we will report (in context) a 95% bootstrap confidence interval for α_0 using both types of intervals - standard and percentile.

Table 1: Bootstrapped 95% CIs for α_0

Method	Lower	Upper
Percentile	0	0.55
Standard	-0.035	0.535

As shown above, the 95% bootstrapped confidence interval of α_0 using the percentile method is [0, 0.55]. This means that over repeated samples of size 10000, 95% of intervals constructed in this fashion will contain the true probability that twins of the same sex are identical. Another possible interpretation is that, we are 95% confident that the true probability that twins of the same sex are identical is in the range [0, 0.55].

Furthermore, as shown above, the 95% bootstrapped confidence interval of α_0 using the standard bootstrap method is [-0.0352926, 0.5352926]. However, notice that the lower limit is negative, but α_0 is a probability and can't be negative. Thus our interval is actually, [0, 0.5352926]. This means that over repeated samples of size 10000, 95% of intervals constructed in this fashion will contain the true probability that twins of the same sex are identical. Another possible interpretation is that, we are 95% confident that the true probability that twins of the same sex are identical is in the range [0, 0.5352926].

It is important to note that in our reported 95% bootstrapped confidence interval of α_0, we did not bias correct because in the previous part we showed that $\widehat{\alpha}_0^{mle}$ was an unbiased estimator of α_0. However, we also added a caveat to the bootstrapped sampling distribution; we turned all negative numbers into 0. By doing this we are systematically overestimating all negative values, and on average all values as well. Hence it could be appropriate to bias correct our interval. By doing this we obtain a 95% bootstrapped confidence interval of α_0 using the standard bootstrap method with a bias correction of [-0.0349376, 0.5356476]. However, notice that the lower limit is negative, but α_0 is a probability and can't be negative. Thus our interval is actually, [0, 0.5356476]. This means that over repeated samples of size 10000, 95% of intervals constructed in this fashion will contain the true probability that twins of the same sex are identical. Another possible interpretation is that, we are 95% confident that the true probability that twins of the same sex are identical is in the range [0, 0.5356476].

Appendix

Code for problem 1

```
# Google "cache chunk option" to see what it does

# Calculate the MLE of the above sample:
mle_sample <- (10 + 15 - 15) / 40

set.seed(188)
x <- rep(c("MM","FF","MF"), times = c(10,15,15))
B = 10000

boot_sim <- lapply(1:B, FUN = function(i) {
    # Generate a resample from x:
    sample = sample(x, size = 40, replace = TRUE)

    # Count the number of MM, MF, FF:
    x_vals <- c(0, 0, 0)
    for (val in 1:40) {
      if (sample[val] == "MM") {
        x_vals[1] = x_vals[1] + 1
      }
      else if (sample[val] == "FF") {
        x_vals[2] = x_vals[2] + 1
      }
```

```
      else if (sample[val] == "MF") {
        x_vals[3] = x_vals[3] + 1
      }
    }

    # Calculate the bootstrapped MLE of alpha and return value as data frame:
    data.frame(alpha_0 = max(0, (x_vals[1] + x_vals[2] - x_vals[3]) / 40))
})

boot_sim_mle <- do.call(rbind, boot_sim)

# Write code to create a bootstrapped sampling distribution and save the figure
# to an object called p1 and display it in part i. below by typing p1 in a code
# chunk that is not echoed:
p1 <- ggplot(data = boot_sim_mle) +
  geom_histogram(mapping = aes(x = alpha_0), binwidth = 0.05,
                 color = "black", fill = "blue") +
  labs(title = expression("10000 Bootstrapped values of" ~ hat(alpha[0])^{mle}),
       x = expression("Values of" ~ hat(alpha[0])^{mle}),
       y = "Count") +
  theme_bw()

# Write code to calculate center of bootstrapped sampling distribution. Save
# your answer in a variable and reference its value with inline code for part
# ii. below:
center_boot <- mean(boot_sim_mle$alpha_0)
se_boot <- sd(boot_sim_mle$alpha_0)

# Write code to calculate bootstrap confidence intervals and reference the end
# points using inline code for part iii. below:
standard <- boot_sim_mle %>%
  summarise(lower = mle_sample - qnorm(0.975) * se_boot,
            upper = mle_sample + qnorm(0.975) * se_boot)

percentile <- boot_sim_mle %>%
  summarise(lower = quantile(alpha_0, 0.025),
            upper = quantile(alpha_0, 0.975))

# Extra: Bias corrected interval:
bias <- center_boot - mle_sample

standard_bias <- boot_sim_mle %>%
  summarise(lower = mle_sample - bias - qnorm(0.975) * se_boot,
            upper = mle_sample - bias + qnorm(0.975) * se_boot)
```

2. (Newton Raphson) Suppose $X \sim Geom(\pi_0)$, that is,

$$f(x) = (1-\pi_0)^x \, \pi_0, \quad 0 \leq \pi_0 \leq 1$$

a. We need to find the root of the equation $s(\pi) = \frac{d}{d\pi}\ell(\pi) = 0$ in order to find the MLE of π_0. Write $s(\pi)$.

Suppose we have a random variable $X \sim Geom(\pi_0)$ with PMF $f(x) = (1-\pi_0)^x \, \pi_0, \quad 0 \leq \pi_0 \leq 1$. Our main

goal is to find the root of the equation $s(\pi) = \frac{d}{d\pi}\ell(\pi) = 0$ in order to find the MLE of π_0, and in order to do that we must find $s(\pi)$.

From the PMF of a Geometric random variable, we find that we can write the likelihood function of π as $L(\pi) = (1-\pi)^x \, \pi, \ 0 \leq \pi \leq 1$. We will now take the natural log of both sides to find $\ell(\pi)$.

$$\begin{aligned}
\ell(\pi) &= \ln\left((1-\pi)^x \, \pi\right) \\
&= \ln\left((1-\pi)^x\right) + \ln(\pi) \\
&= x\ln(1-\pi) + \ln(\pi), \ 0 \leq \pi \leq 1
\end{aligned}$$

Hence, as calculated above, $\ell(\pi) = x\ln(1-\pi) + \ln(\pi), \ 0 \leq \pi \leq 1$. We will now find $s(\pi) = \frac{d}{d\pi}\ell(\pi)$, by doing this we will be able to find the root of $s(\pi)$ and thus find the MLE of π_0.

$$\begin{aligned}
s(\pi) &= \frac{d}{d\pi}\ell(\pi) \\
&= \frac{d}{d\pi}\left(x\ln(1-\pi) + \ln(\pi)\right) \\
&= \frac{-x}{1-\pi} + \frac{1}{\pi}, \ 0 \leq \pi \leq 1
\end{aligned}$$

Thus, as calculated above, $s(\pi) = \frac{-x}{1-\pi} + \frac{1}{\pi}, \ 0 \leq \pi \leq 1$.

b. Say we decide to find the MLE of π_0 using Newton Raphson. Suppose we observe $x = 10$. Calculate π_{new} assuming we begin the algorithm at $\pi_{old} = 0.5$. That is, perform one update of the Newton Raphson method.

Instead of finding the critical points of $s(\pi)$ to find the MLE of π_0, we will instead use the Newton Raphson method to find the MLE. Suppose we observe $x = 10$, our goal is to find π_{new} assuming we begin the algorithm at $\pi_{old} = 0.5$. That is, we will perform one update of the Newton Raphson method.

The fist step of the Newton Raphson method is to check and see if π_{old} isn't already considered the MLE of π_0 by calculating $s(\pi_{old})$ and seeing if it's within some reasonable tolerance of zero. Using our observed value of $x = 10$ and our assumed value of $\pi_{old} = 0.5$, we see that $s(0.5) = \frac{-10}{1-0.5} + \frac{1}{0.5} = -18$. Hence our value of $s(\pi_{old})$ is not within any reasonable tolerance of zero.

Since our value of $s(\pi_{old})$ is not within any reasonable tolerance of zero, we will linearize $s(\pi)$ around π_{old} by using a Taylor Expansion. This means we are looking to find the roots of $s(\pi) \approx s(\pi_{old}) + (\pi - \pi_{old})s'(\pi_{old})$. First we will calculate $s'(\pi_{old}) = \frac{d}{d\pi}s(\pi)$ evaluated at π_{old}.

$$\begin{aligned}
\frac{d}{d\pi}s(\pi) &= \frac{d}{d\pi}\left(\frac{-x}{1-\pi} + \frac{1}{\pi}\right) \\
&= \frac{-x}{(1-\pi)^2} - \frac{1}{\pi^2}, \ 0 \leq \pi \leq 1
\end{aligned}$$

Hence, it follows that $s'(\pi_{old}) = \frac{-10}{(1-0.5)^2} - \frac{1}{0.5^2} = -44$. Now we will set the Taylor expansion equal to zero and solve for π.

$$\begin{aligned}
s(\pi) &\approx -18 - 44(\pi - 0.5) \\
0 &= -18 - 44(\pi - 0.5) \\
&= -18 + 22 - 44\pi \\
44\pi &= 4 \\
\pi &= \frac{1}{11}
\end{aligned}$$

Hence as computed above, $\pi_{new} = \frac{1}{11}$, and it turns out that $s(\pi_{new}) = \frac{-10}{1-\frac{1}{11}} + \frac{1}{\frac{1}{11}} = 0$ which means $\widehat{\pi}_0^{mle} = \frac{1}{11}$.

3. (Batting averages) Recall that the beta distribution

$$f(x) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\,\Gamma(\beta_0)}\, x^{\alpha_0 - 1}\,(1-x)^{\beta_0 - 1} \quad 0 < x < 1$$

is a useful distribution for modeling proportions. The parameters α_0 and β_0 are both required to be non-negative in order for $f(x)$ to be a valid PDF.

Below are the batting averages for 16 randomly selected major league baseball players (from the 2015 season, minimum 200 at bats)

```
ba <- c(0.276, 0.281, 0.225, 0.283, 0.257, 0.250, 0.250, 0.261, 0.312, 0.259,
        0.273, 0.222, 0.314, 0.271, 0.294, 0.268)
```

a. Write the log-likelihood function $\ell(\alpha, \beta)$. Please leave the data as x's and n in your equation, do not plug in numbers. Don't forget the range for α and β.

The PDF of a beta distribution is $f(x) = \frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\,\Gamma(\beta_0)}\, x^{\alpha_0-1}\,(1-x)^{\beta_0-1}$, $0 < x < 1$, and it is a very useful distribution for modeling proportions. The parameters α_0 and β_0 are both required to be non-negative in order for $f(x)$ to be a valid PDF. In order to find the MLE of α_0 and β_0 and use it to analyze batting averages for 16 randomly selected major league baseball players, we will need to find the log-likelihood function $\ell(\alpha, \beta)$. We will start off by calculating the normal likelihood function of α and β below. Assuming the X_i's are independent and identically distributed from a beta distribution with α_0 and β_0 as parameters, we can write $L(\alpha, \beta)$ as:

$$\begin{aligned}
L(\alpha,\beta) &= f(x_1) \times f(x_2) \times \cdots \times f(x_n) \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x_1^{\alpha-1}\,(1-x_1)^{\beta-1} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x_2^{\alpha-1}\,(1-x_2)^{\beta-1} \times \cdots \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x_n^{\alpha-1}\,(1-x_n)^{\beta-1} \\
&= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n (x_1 x_2 \ldots x_n)^{\alpha-1}\,((1-x_1)(1-x_2)\ldots(1-x_n))^{\beta-1} \\
&= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n \left(\prod_i^n x_i\right)^{\alpha-1} \left(\prod_i^n (1-x_i)\right)^{\beta-1} \quad \alpha > 0,\ \beta > 0
\end{aligned}$$

As can be seen from the above calculation, the likelihood function of α and β is $L(\alpha,\beta) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n (\prod_i^n x_i)^{\alpha-1}\,(\prod_i^n (1-x_i))^{\beta-1}$ $\alpha > 0,\ \beta > 0$. Hence in order to find $ell(\alpha,\beta)$, we will take the natural log of both sides of this function. This calculation is shown below.

$$\begin{aligned}
\ell(\alpha,\beta) &= \ln(L(\alpha,\beta)) \\
&= \ln\left(\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n \left(\prod_i^n x_i\right)^{\alpha-1} \left(\prod_i^n (1-x_i)\right)^{\beta-1}\right) \\
&= \ln\left(\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^n\right) + \ln\left(\left(\prod_i^n x_i\right)^{\alpha-1}\right) + \ln\left(\left(\prod_i^n (1-x_i)\right)^{\beta-1}\right) \\
&= n\ln\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right) + (\alpha-1)\ln\left(\prod_i^n x_i\right) + (\beta-1)\ln\left(\prod_i^n (1-x_i)\right) \\
&= n\ln\left(\Gamma(\alpha+\beta)\right) - n\ln\left(\Gamma(\alpha)\,\Gamma(\beta)\right) + (\alpha-1)\ln\left(\prod_i^n x_i\right) + (\beta-1)\ln\left(\prod_i^n (1-x_i)\right) \\
&= n\ln\left(\Gamma(\alpha+\beta)\right) - n\ln\left(\Gamma(\alpha)\right) - n\ln\left(\Gamma(\beta)\right) + (\alpha-1)\ln\left(\prod_i^n x_i\right) + (\beta-1)\ln\left(\prod_i^n (1-x_i)\right) \quad \alpha > 0,\ \beta > 0
\end{aligned}$$

Thus as computed above, the log likelihood function of α and β is $\ell(\alpha,\beta) = n\ln(\Gamma(\alpha+\beta)) - n\ln(\Gamma(\alpha)) - n\ln(\Gamma(\beta)) + (\alpha-1)\ln\left(\prod_i^n x_i\right) + (\beta-1)\ln\left(\prod_i^n (1-x_i)\right),\ \alpha>0,\ \beta>0.$

b. Calculate $\widehat{\alpha}_0^{mom}$ and $\widehat{\beta}_0^{mom}$, the method of moments estimates of α_0 and β_0. Show your code and also print the answers. (See problem 1 on Homework 5 from STAT 341 for the formulas for the M.O.M. estimators. You can find it in the homework sub-folder for STAT 342)

As formulated in problem 1 on Homework 5 from STAT 341 the formulas for the M.O.M. estimators of α_0 and β_0 are: $\widehat{\alpha}_0^{mom} = \bar{x}\left[\frac{\bar{x}-s}{s-\bar{x}^2}\right]$ and $\widehat{\beta}_0^{mom} = \widehat{\alpha}_0^{mom}\frac{1-\bar{x}}{\bar{x}}$ where $s = \sum_{i=1}^n x_i^2$. Below we will calculate $\widehat{\alpha}_0^{mom}$ and $\widehat{\beta}_0^{mom}$, the method of moments estimates of α_0 and β_0.

```
# Find the sample size:
n <- length(ba)

# Find the sample mean:
xbar <- mean(ba)

# Find s:
s <- sum(ba^2) / n

# Find method of moments estimate of alpha:
alpha_mom <- xbar * ((xbar - s) / (s - xbar^2))

# Find method of moments estimate of beta:
beta_mom <- alpha_mom * ((1 - xbar) / xbar)
```

As calculated above, $\widehat{\alpha}_0^{mom} = 83.4386125$ and $\widehat{\beta}_0^{mom} = 227.3197208$.

c. We will now fit the beta distribution by maximum likelihood. Using the method of moments estimators as starting values for Newton Raphson, write code below to find the MLEs. (Show both code and output here)

Now that we have M.O.M estimates for α_0 and β_0, we will now fit the beta distribution by maximum likelihood. We will do this using the Newton Raphson method to find the MLE with the M.O.M estimates for α_0 and β_0 as starting values for the algorithm.

```
# Create the log-likelihood function:
loglik_beta <- function(params, x) {
  ifelse(params[1] < 0 | params[2] < 0,
         NA,
         sum(dbeta(x = x, shape1 = params[1], shape2 = params[2], log = TRUE))
  )
}

# Find the MLEs of alpha and beta using Newton Raphson:
beta_mle <- maxLik(logLik = loglik_beta, start = c(alpha_mom, beta_mom),
                   method = "NR", tol = 1e-8, x = ba)
beta_mle
```

```
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 2 iterations
```

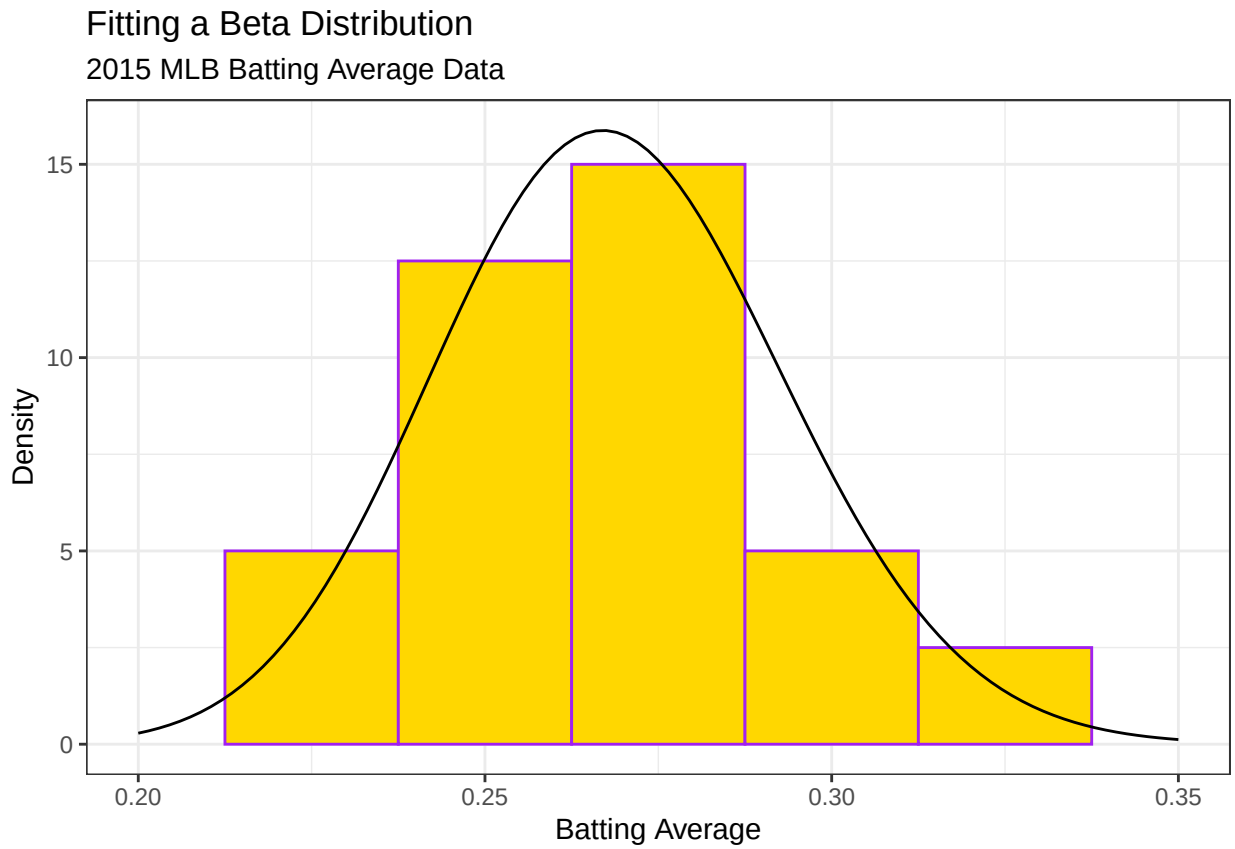
```
## Return code 2: successive function values within tolerance limit (tol)
## Log-Likelihood: 36.24572 (2 free parameter(s))
## Estimate(s): 83.39912 227.2415
```

As computed by the code from above, the Newton Raphson method's estimates for α_0 and β_0 are: $\widehat{\alpha}_0^{mle} = 83.39912$ and $\widehat{\beta}_0^{mle} = 227.2415$.

d. Make a histogram of the batting average data along and overlay the fitted beta distribution from part c. (Show both code and output here)

```
# Use Sturge's Rules to calculate the number of bins, num_bins = 5:
num_bins <- ceiling(log(x = length(ba), base = 2) + 1)

# Plot the histogram of batting average data with the beta distribution:
# with the MLE estimates overlaid:
ggplot() +
  geom_histogram(mapping = aes(x = ba, y = after_stat(density)),
                 binwidth = 0.025,
                 color = "purple",
                 fill = "gold") +
  stat_function(fun = dbeta,
                args = list(shape1 = beta_mle$estimate[1],
                            shape2 = beta_mle$estimate[2]),
                xlim = c(0.2, 0.35)) +
  labs(title = "Fitting a Beta Distribution",
       subtitle = "2015 MLB Batting Average Data",
       x = "Batting Average",
       y = "Density") +
  theme_bw()
```



As can be seen by the above histogram of the 16 batting average data points from a random sample of MLB players in 2015, the fitted beta distribution with the alpha and beta parameters as the MLE of α_0 and β_0 found in part c. fits the data pretty well. Although there is a little underestimation at the tails, the beta distribution fits the data surprisingly well given the small amount of data points.

4. (Bias/variance tradeoff) Suppose $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$ where both parameters are unknown. Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ denote the usual sample variance. The maximum likelihood estimator is

$$\widehat{\sigma^2}_0^{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

a. Recall the mean squared error (MSE) of an estimator is defined as

$$\text{MSE} = \text{Bias}^2 + \text{Var}.$$

Write the mean squared error of S^2. (You do not need to prove results we have already proved in class or ones that you have proved on past homework. Just cite them with a reference.)

Setup:
If we suppose $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$ where both parameters are unknown, and let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ denote the usual sample variance, then the maximum likelihood estimator of σ_0^2 is $\widehat{\sigma^2}_0^{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Furthermore, the mean squared error (MSE) of an estimator is defined as $\text{MSE} = \text{Bias}^2 + \text{Var}$. Our goal in this part is to find the mean squared error of S^2.

Finding Bias^2 of S^2:
As was proven in Theorem 22.1 if $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$, then the sample variance is an unbiased estimator of the population variance. Thus $\big(\text{Bias}[S^2]\big)^2 = 0$, and hence $\text{MSE}[S^2] = \text{Var}[S^2]$.

Finding Var of S^2:
Furthermore, we proved in Homework 3 Question 3 that if $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$, then $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$ and thus $Var\left[\frac{(n-1)S^2}{\sigma_0^2}\right] = 2(n-1)$. Therefore we conclude that $Var[S^2] = \frac{2\sigma_0^4}{n-1}$.

Putting it all together:
Since MSE = Bias^2 + Var and in this case Bias$^2 = 0$ and $Var[S^2] = \frac{2\sigma_0^4}{n-1}$, we conclude that $\text{MSE}[S^2] = \frac{2\sigma_0^4}{n-1}$

b. Find the MSE of $\widehat{\sigma^2}_0^{mle}$. (You do not need to prove results we have already proved in class or ones that you have proved on past homework. Just cite them with a reference.)

Setup:
As mentioned above, the maximum likelihood estimator of σ_0^2 is $\widehat{\sigma^2}_0^{mle} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Furthermore, the mean squared error (MSE) of an estimator is defined as MSE = Bias^2 + Var. Our goal in this part is to find the mean squared error of $\widehat{\sigma^2}_0^{mle}$.

Finding Bias^2 of $\widehat{\sigma^2}_0^{\mathbf{mle}}$:
As was shown in example 24.11, $E\left[\widehat{\sigma^2}_0^{mle}\right] = \frac{n-1}{n}\sigma_0^2$. With this result we can see that $\text{Bias}\left[\widehat{\sigma^2}_0^{mle}\right] = E\left[\widehat{\sigma^2}_0^{mle}\right] - \sigma_0^2 = \frac{n-1}{n}\sigma_0^2 - \sigma_0^2 = \frac{-1}{n}\sigma_0^2$. Hence we can see that the Bias^2 part of the MSE of $\widehat{\sigma^2}_0^{mle}$ is $\text{Bias}^2 = \left(\frac{-\sigma_0^2}{n}\right)^2 = \frac{\sigma_0^4}{n^2}$

Finding Var of $\widehat{\sigma^2}_0^{\mathbf{mle}}$:
Below we will use the fact that $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2$ to find $Var\left[\widehat{\sigma^2}_0^{mle}\right]$.

$$\begin{aligned}
Var\left[\widehat{\sigma^2}_0^{mle}\right] &= Var\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= Var\left[\frac{n-1}{n}S^2\right] \\
&= \left(\frac{n-1}{n}\right)^2 Var[S^2] \quad \text{(Non-linearity of Variance)} \\
&= \left(\frac{n-1}{n}\right)^2 \cdot \left(\frac{2\sigma_0^4}{n-1}\right) \quad \text{(Homework 3 Problem 3)}
\end{aligned}$$

Hence as computed above, $Var\left[\widehat{\sigma^2}_0^{mle}\right] = \left(\frac{n-1}{n}\right)^2 \cdot \left(\frac{2\sigma_0^4}{n-1}\right)$

Putting it all together:
Since MSE = Bias^2 + Var and in this case $\text{Bias}^2 = \frac{\sigma_0^4}{n^2}$ and $Var\left[\widehat{\sigma^2}_0^{mle}\right] = \left(\frac{n-1}{n}\right)^2 \cdot \left(\frac{2\sigma_0^4}{n-1}\right)$, we conclude that $\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right] = \frac{\sigma_0^4}{n^2} + \left(\frac{n-1}{n}\right)^2 \cdot \left(\frac{2\sigma_0^4}{n-1}\right)$. Below we will simplify this expression to get a cleaner form of

$\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right]$, which will make finding the ratio of the two MSEs in part c much easier.

$$\begin{aligned}
\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right] &= \frac{\sigma_0^4}{n^2} + \left(\frac{n-1}{n}\right)^2 \cdot \left(\frac{2\sigma_0^4}{n-1}\right) \\
&= \frac{\sigma_0^4}{n^2} + \frac{2(n-1)\sigma_0^4}{n^2} \\
&= \frac{(2n-2)\sigma_0^4 + \sigma_0^4}{n^2} \\
&= \frac{(2n-1)\sigma_0^4}{n^2}
\end{aligned}$$

Hence, as computed above, $\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right] = \frac{(2n-1)\sigma_0^4}{n^2}$.

c. Make a plot of the ratio of the MSE of $\widehat{\sigma^2}_0^{mle}$ to the MSE of S^2 for n from 1 to 100. Write a couple of sentences with your conclusion.

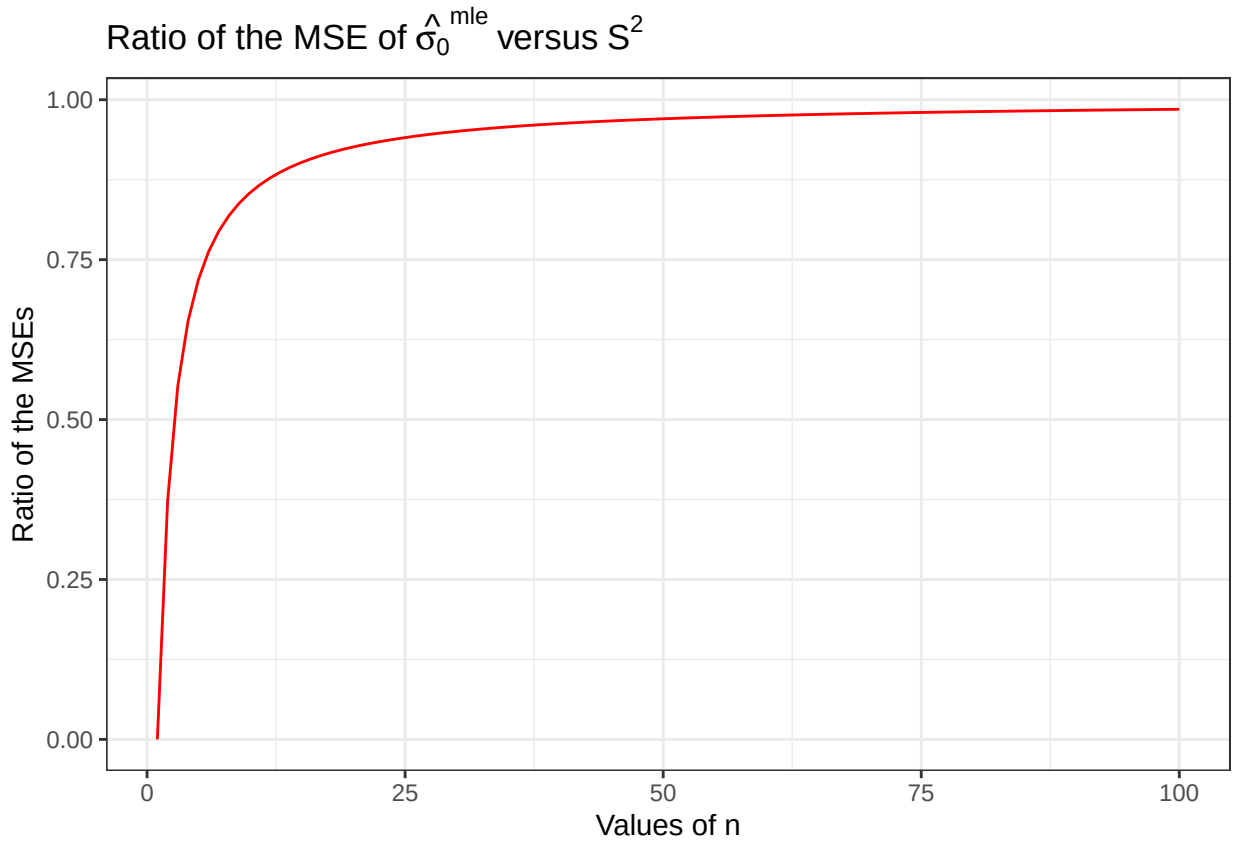
As computed in part a, $\text{MSE}[S^2] = \frac{2\sigma_0^4}{n-1}$. Likewise, as computed in part b, $\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right] = \frac{(2n-1)\sigma_0^4}{n^2}$. Thus the ratio of the MSE of $\widehat{\sigma^2}_0^{mle}$ to the MSE of S^2 is:

$$\begin{aligned}
\frac{\text{MSE}\left[\widehat{\sigma^2}_0^{mle}\right]}{\text{MSE}[S^2]} &= \frac{\frac{(2n-1)\sigma_0^4}{n^2}}{\frac{2\sigma_0^4}{n-1}} \\
&= \frac{(2n-1)(n-1)\sigma_0^4}{2n^2\sigma_0^4} \\
&= \frac{(2n-1)(n-1)}{2n^2}
\end{aligned}$$

Hence, as computed above the ratio of the MSE of $\widehat{\sigma^2}_0^{mle}$ to the MSE of S^2 is $\frac{(2n-1)(n-1)}{2n^2}$.

Below is a plot of the ratio of the MSE of $\widehat{\sigma^2}_0^{mle}$ to the MSE of S^2 for n from 1 to 100.

```
ggplot() +
  geom_function(fun = function(n){((2*n - 1) * (n - 1)) / (2*n^2)},
                color = "red",
                xlim = c(1, 100)) +
  labs(title = expression("Ratio of the MSE of" ~ hat(sigma[0])^{mle} ~ "versus" ~ S^2),
       y = "Ratio of the MSEs",
       x = "Values of n") +
  theme_bw()
```



From the above plot we can see that $MSE[S^2]$ is always larger than $MSE\left[\widehat{\sigma^2}_0^{mle}\right]$. However, as n gets larger and larger the difference between these two values converges to zero, hence why we see the ratio of these two values converging to 1. This happens because the bias in $MSE\left[\widehat{\sigma^2}_0^{mle}\right]$ goes to zero as n goes to infinity. Similarly, the difference between $Var\left[\widehat{\sigma^2}_0^{mle}\right]$ and $Var[S^2]$ goes to zero as n goes to infinity.