Homework 5

Spring 2023

Tanner Huck

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 4 if you need a reminder

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

Assume a population of twins that consists of males and Females where each sex is equally likely and α_0 is the probability that the twins are identical. Identical twins means that the twins share the same sex, otherwise each sex is determined independently of each other. Hence using the using the law of total probability the probability of seeing male-male or female-female is $\pi_1 = P(MM) = \pi_2 = P(FF) = \frac{1+\alpha_0}{4}$ and the probability of seeing male-female is $\pi_3 = P(MF) = \frac{1-\alpha_0}{2}$. Our goal is to prove the result for $P(MM)$.

We can start by rewriting the event of "MM" as the union of two disjoint events, $MM = (T_1 \cap T_2 \cap I) \cup (T_1 \cap T_2 \cap I^c)$ where T_1 denotes the event that twin 1 is M, T_2 is the event that twin 2 is M, and I denotes the event where the twins are identical. Thus,

$$\begin{aligned}
P(MM) &= P(T_1 \cap T_2 \cap I) + P(T_1 \cap T_2 \cap I^c) \\
&\text{by the laws of conditional probability} \\
&= P(T_1 \cap T_2 | I)P(I) + P(T_1 \cap T_2 | I^c)P(I^c) \\
&= P(T_1 | I)P(T_2 | T_1 \cap I)P(I) + P(T_1 | I^c)P(T_2 | T_1 \cap I^c)P(I^c) \\
&\text{by our definition of } \alpha_0\text{, probability of M or F is identical,} \\
&\text{and sexes are determined independently if not itentical} \\
&= \frac{1}{2} \cdot \alpha_0 + \frac{1}{2} \cdot \frac{1}{2} \cdot (1 - \alpha_0) \\
&= \frac{2\alpha_0}{4} + \frac{1-\alpha_0}{4} \\
&= \frac{1+\alpha_0}{4}
\end{aligned}$$

Hence, $P(MM) = \frac{1+\alpha_0}{4}$.

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.

Assume that X_1, X_2, X_3 denote the number of MM, FF and MF out of n twins. Then $\langle X_1, X_2, X_3 \rangle \sim Multinom(n, \pi = (\pi_1, \pi_2, \pi_3))$ where π_1, π_2 and π_3 are functions of α_0. Our goal is to find the maximum likelihood estimate of α_0.

First, we know that the PMF of a multinomial distribution is given by $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1!x_2!x_3!}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3}$ where $n = x_1 + x_2 + x_3$. Thus the likelihood function is,

$$\begin{aligned}
L(\alpha) &= \frac{n!}{x_1!x_2!x_3!}\pi_1^{x_1}\pi_2^{x_2}\pi_3^{x_3}\pi_i^{x_i} \\
&= \frac{n!}{x_1!x_2!x_3!} \cdot \frac{(1+\alpha)^{x_1+x_2}}{4^{x_1+x_2}} \cdot \frac{(1-\alpha)^{x_3}}{2^{x_3}}.
\end{aligned}$$

Then finding $\ell\alpha)$,

$$\ell(\alpha) = lnn! - lnx_1! - lnx_2! - lnx_3! + (x_1 + x_2)ln(1+\alpha) - (x_1 + x_2)ln4 + x_3 ln(1-\alpha) - x_3 ln2.$$

Taking the derivative of the log-likelihood with respect to α,

$$\frac{d}{d\alpha} lnL(\alpha) = \frac{x_1 + x_2}{1+\alpha} - \frac{x_3}{1-\alpha}$$

Then setting the derivative equal to 0 and solving for α,

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

Hence $\widehat{\alpha}^{mle} = \frac{x_1+x_2-x_3}{n}$

```
c. Is $\widehat{\alpha}_0$ an unbiased estimator of $\alpha_0$? Yes or no and show your work.   (Hint: 
```

To see if $\widehat{\alpha}_0$ is an unbiased estimator of α_0 we can find the expected value of $\widehat{\alpha}_0$ and see if it is equal to α_0.

$$\begin{aligned}
E[\widehat{\alpha}_0] &= E[\frac{X_1+X_2-X_3}{n}] \\
&\text{by linearity of expectation} \\
&= \frac{1}{n}E(X_1+X_2-X_3) \\
&\text{by how we defined each } X_i \text{ (from a multinomial dist.)} \\
&= \frac{1}{n}(n\pi_1 + n\pi_2 - n\pi_3) \\
&= \frac{1+\alpha_0}{4} + \frac{1+\alpha_0}{4} - \frac{1-\alpha_0}{2} \\
&= \frac{2+2\alpha_0}{4} - \frac{1-\alpha_0}{2} \\
&= \frac{1+\alpha_0}{2} - \frac{1-\alpha_0}{2} \\
&= \frac{2\alpha_0}{2} \\
&= \alpha
\end{aligned}$$

Hence $\widehat{\alpha}_0$ is an unbiased estimator of α_0.

```
d. The sampling distribution of $\alpha_0^{mle}$ is complicated because the counts $x_i$ are dependent 
\centering
```

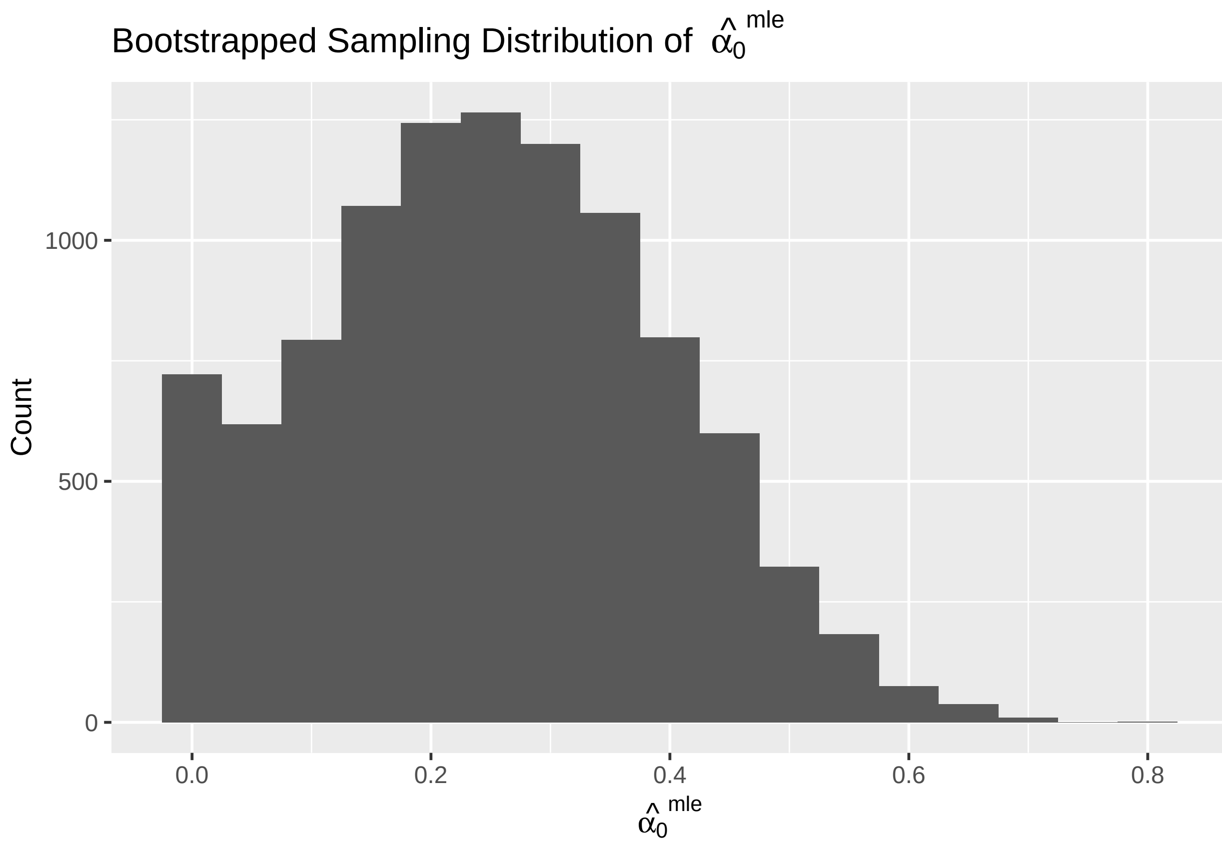
MM	FF	MF
10	15	15

\end{table}

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)

```
p1
```



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

The bootstrapped sampling sampling distribution is centered at about 0.25. This makes sense because in the original data, $\widehat{\alpha}^{mle} = \frac{10+15-15}{40} = 0.25$. Since in bootstrapping, we repeatedly resample the original dataset with replacement and we know that $\widehat{\alpha}^{mle}$ is an unbiased estimator of α we can expect the bootstrap distribution to be centered in relatively the same place. In bootstrapping, we are creating many new data that will still reflect the same underlying population.

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.

```
tableiii
```

	Lower	Upper
Percentile	0	0.55
Standard	-0.04	0.54

In this table we can we can see the 95% confidence intervals using the percentile and standard methods, which are about the same. The interval with the standard method is about $[-0.4, 0.54]$, and with the percentile method we get the interval $[0, 0.55]$. We can see that with a sample of $n = 1000$ both methods will yield intervals that will contain the true value of α_0 95% of the time, the probability of same sex twins being identical. Also note that the lower bound for the standard method is negative, however this is a probability, thus should be 0.

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 $s(\pi)$.

To find the maximum likelihood estimator for π_0 we need to find the log-likelihood function for our given geometric distribution and then set its derivative equal to 0. Thus for $0 \leq \pi < 1$,

$$\begin{aligned} L(\pi) &= \pi(1-\pi)^x \\ ln(L(\pi)) &= ln(\pi(1-\pi)^x) \\ &= ln(\pi) + x_i ln(1-\pi) \\ \frac{d}{d\pi} ln(L(\pi)) &= \frac{1}{\pi} - \frac{x}{1-\pi} \end{aligned}$$

Thus $s(\pi) = \frac{d}{d\pi}\ell(\pi) = \frac{1}{\pi} - \frac{x}{1-\pi}$

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.

To perform Newton-Raphson, we can use the equation $\pi_{new} = \pi_{old} - \frac{s(\pi_{old})}{s'(\pi_{new})}$ where $s(\pi) = \frac{d}{d\pi}\ell(\pi)$. Then assuming we observe $x = 10$ and $\pi_{old} = 0.5$, $s(\pi) = frac1\pi - \frac{x}{1-\pi} = frac10.5 - \frac{10}{1-0.5} = -18$ and $s'(\pi) = \frac{d}{d\pi}\frac{1}{\pi} - \frac{x}{1-\pi} = -\frac{1}{\pi^2} - \frac{x}{(1-\pi)^2} = -\frac{1}{0.5^2} - \frac{10}{(1-0.5)^2} = -44$ for $0 \leq \pi \leq 1$. Thus calculating one update of the Newton Raphson method,

$$\pi_{new} = \pi_{old} - \frac{s(\pi_{old})}{s'(\pi_{new})} = 0.5 - \frac{-18}{-44} = \frac{1}{11} \approx 0.091$$

Thus from one update of Newton Raphson, $\widehat{\pi}_0^{mle} = \frac{1}{11}$

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, 0.27
```

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.

Given the beta distribution, we can calculate the log-likelihood function $\ell(\alpha, \beta)$ as,

$$\begin{aligned} \ell(\alpha,\beta) &= ln\prod_{i=1}^{n}[(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x_i^{\alpha-1}(1-x_i)^{\beta-1})] \\ &= \prod_{i=1}^{n} ln\Gamma(\alpha+\beta) - ln\Gamma(\alpha) - ln\Gamma(\beta) + (\alpha-1)ln(x_i) + (\beta-1)ln(1-x_i) \\ &= nln\Gamma(\alpha+\beta) - nln\Gamma(\alpha) - nln\Gamma(\beta) + (\alpha-1)\sum_{i=1}^{n} ln(x_i) + (\beta-1)\sum_{i=1}^{n} ln(1-x_i) \\ &\text{for } \alpha, \beta > 0 \end{aligned}$$

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)

From problem 1 on Homework 5 from STAT 341 we know that the method of moments estimators of α_0 and β_0 are:

$$\hat{\alpha}_0^{mom} = \bar{x}\left[\frac{\bar{x}-s}{s-\bar{x}^2}\right],$$
$$\hat{\beta}_0^{mom} = \hat{\alpha}_0^{mom}\frac{1-\bar{x}}{\bar{x}}$$

Where $s = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

Thus calculating $\widehat{\alpha}_0^{mom}$ and $\widehat{\beta}_0^{mom}$ for our data,

```
mean_ba <- mean(ba)
var_ba <- var(ba)
s <- (1/length(ba)) * sum(ba^2)

alpha_mom <- (mean_ba * (mean_ba - s) ) / ( s - mean_ba^2)
beta_mom <- alpha_mom * ( 1- mean_ba ) / mean_ba

alpha_mom
```

```
## [1] 83.43861
```

```
beta_mom
```

```
## [1] 227.3197
```

Hence $\widehat{\alpha}_0^{mom}$ is about 83.44 and $\widehat{\beta}_0^{mom}$ is about 227.32.

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 we will use Newton Raphson with our method of moment estimators as starting values to fit the beta distribution by maximum likelihood.

```
loglik.beta_dist <- function(mom, x){
  ifelse((mom[1] | mom[2]) < 0, NA, sum(dbeta(x=x, mom[1], mom[2], log=T)))
}

mles <- maxLik(logLik = loglik.beta_dist, start=c(alpha_mom, beta_mom), method="NR", tol=1e-10, x=ba)

mles$estimate
```

```
## [1]  83.24594 226.79918
```

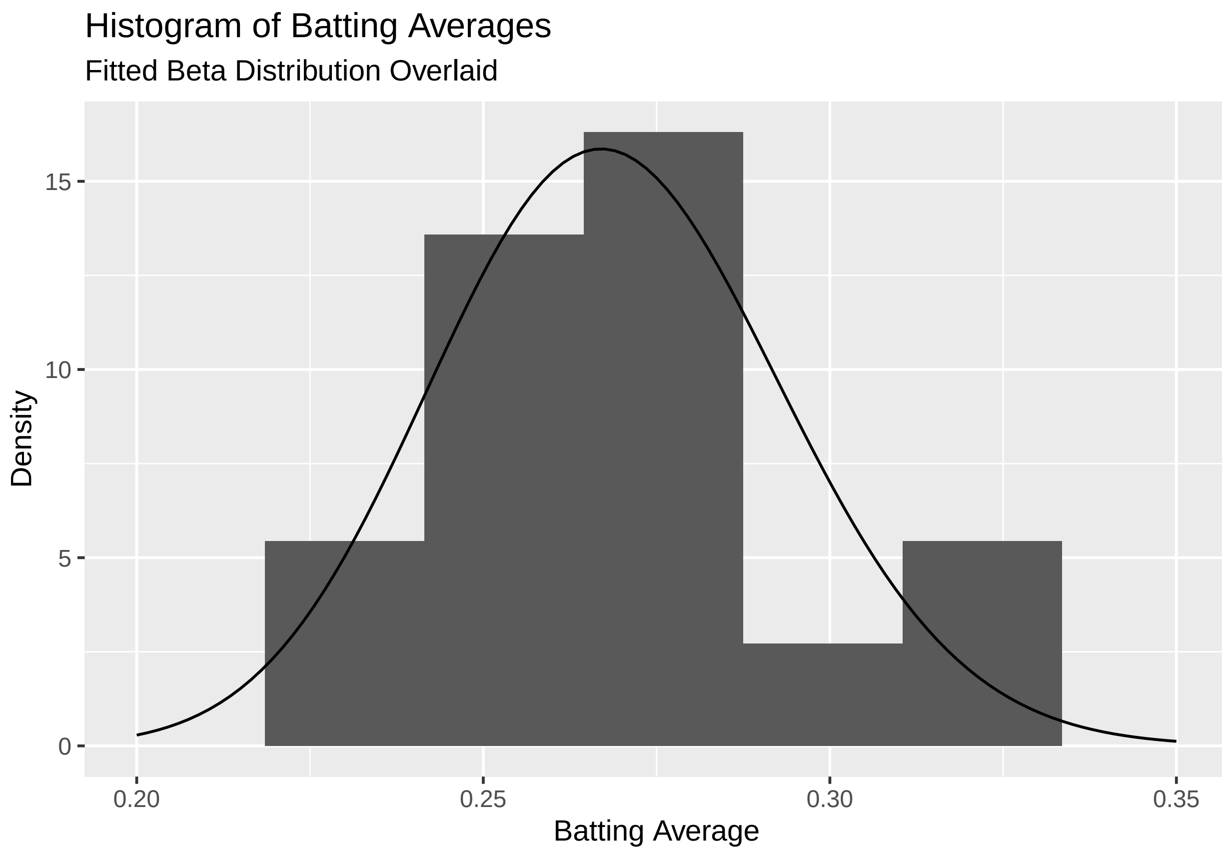
Hence the MLE for α is about 83.25 and the MLE for β is about 226.80.

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)

Here is a histogrm of the batting average data with our fitted beta distribution.

```
ggplot() +
  geom_histogram(mapping=aes(x=ba, y=..density..),
                 bins=ceiling(log2(16) + 1)) +
  stat_function(fun=dbeta,
                args=list(mles$estimate[1],
                          mles$estimate[2]),
                xlim=c(0.2, 0.35)) +
```

```r
labs(title = "Histogram of Batting Averages",
     subtitle = "Fitted Beta Distribution Overlaid",
     x = "Batting Average",
     y = "Density")
```



From the histogram we can see that the beta distribution is not an appropriate choice for modeling the batting averages. The beta distribution overestimates how many players have a batting average around 0.3 and underestimates at other values.

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 $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ 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.)

Given that we have $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$ with μ and σ both unknown, we know the maximum likelihood estimator is denoted as $\widehat{\sigma^2_0}^{mle} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$. Furthermore, we know that the sample variance can be denoted as $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Since each X_i is independent and identically distributed from $Norm(\mu_0, \sigma_0)$, we know that the sample variance S^2 is an unbiased estimator of the population variance. ($E[S^2] = \sigma_0^2$). Thus we know that $Bias(S^2)^2 = 0$. Then from homework 3, we found that for $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(\mu_0, \sigma_0)$, $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$ where S^2 was the sample variance. Additionally, $Var(\frac{(n-1)S^2}{\sigma_0^2}) = 2(n-1)$, by non linearity of variance, we can simplify this to $Var[S^2] = \frac{2\sigma_0^4}{n-1}$, which follows from the fact that the variance of a chi square distribution is just twice its degrees of freedom.

Now putting this together to find the mean squared error of S^2,

$$MSE[S^2] = Bias[S^2]^2 + Var[S^2] = Var[S^2] = \frac{2\sigma_0^4}{n-1}$$

Hence, $MSE[S^2] = \frac{2\sigma_0^4}{n-1}$.

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.)

First finding $Bias[\widehat{\sigma^2}_0^{mle}]^2$. From example 24.11 we showed that $E[\widehat{\sigma^2}_0^{mle}] = \frac{n-1}{n}\sigma_0^2$. Thus,

$$\begin{aligned} Bias[\widehat{\sigma^2}_0^{mle}] &= E[\widehat{\sigma^2}_0^{mle}] - \sigma_0^2 \\ &= \frac{n-1}{n}\sigma_0^2 - \sigma_0^2 \\ &= \frac{\sigma_0^2 n - \sigma_0^2}{n} - \frac{n\sigma_0^2}{n} \\ &= \frac{-\sigma_0^2}{n} \end{aligned}$$

Thus $Bias[\widehat{\sigma^2}_0^{mle}]^2 = (\frac{-\sigma_0^2}{n})^2$.

Lastly, we need to calculate the variance of $\widehat{\sigma^2}_0^{mle}$. We know that $\widehat{\sigma^2}_0^{mle} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$, thus

$$\begin{aligned} Var[\widehat{\sigma^2}_0^{mle}] &= Var[\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2] \\ &\text{substituting in } S^2 \text{ we get} \\ &= Var[\frac{(n-1)S^2}{n}] \\ &\text{by non linearity of varience} \\ &= (\frac{n-1}{n})^2 Var[S^2] \\ &\text{Substituting in } Var[S^2] \text{ from what we found in Hw 3, and part a} \\ &= (\frac{n-1}{n})^2 \cdot \frac{2\sigma_0^4}{n-1} \\ &= \frac{n-1 \cdot 2\sigma_0^4}{n^2} \end{aligned}$$

Now putting this together to find the mean squared error of $\widehat{\sigma^2}_0^{mle}$,

$$MSE[\widehat{\sigma^2}_0^{mle}] = Bias[\widehat{\sigma^2}_0^{mle}]^2 + Var[\widehat{\sigma^2}_0^{mle}] = (\frac{-\sigma_0^2}{n})^2 + \frac{n-1 \cdot 2\sigma_0^4}{n^2}$$
$$= \frac{\sigma_0^4 + n2\sigma_0^4 - 2\sigma_0^4}{n^2} = \frac{n2\sigma_0^4 - \sigma_0^4}{n^2}$$

Hence, $MSE[\widehat{\sigma^2}_0^{mle}] = \frac{n2\sigma_0^4 - \sigma_0^4}{n^2}$.

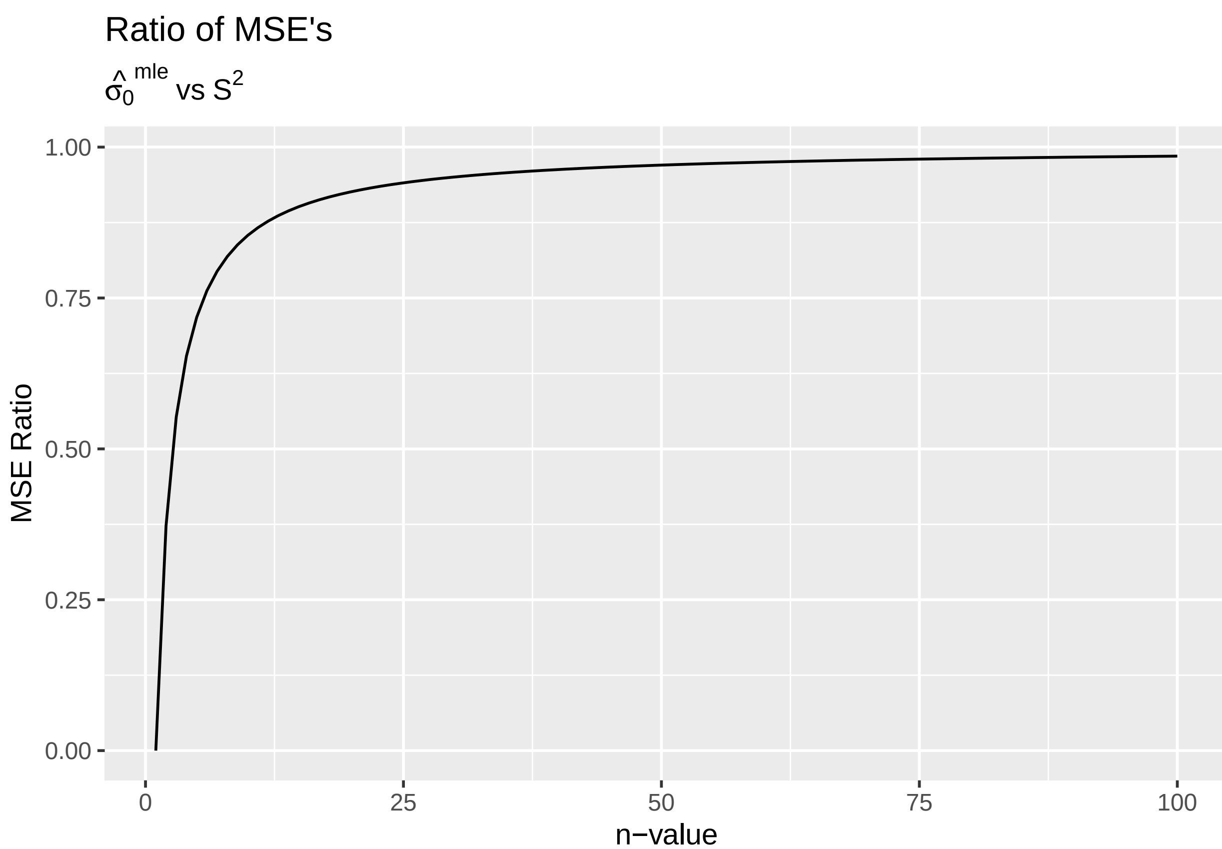
c. Make a plot of the ratio of the MSE of $\widehat{\sigma^2_0}^{mle}$ to the MSE of S^2 for n from 1 to 100. Write a couple of sentences with your conclusion.

From parts a and b, we can write the ratio,

$$
\begin{aligned}
\frac{MSE[\widehat{\sigma^2_0}^{mle}]}{MSE[S^2]} &= \frac{\frac{n2\sigma_0^4 - \sigma_0^4}{n^2}}{\frac{2\sigma_0^4}{n-1}} \\
&= \frac{(n-1)\cdot(n2\sigma_0^4 - \sigma_0^4)}{n^2 2\sigma_0^4} \\
&= \frac{n^2 2\sigma_0^4 - n\sigma_0^4 - n2\sigma_0^4 + \sigma_0^4}{n^2 2\sigma_0^4} \\
&= \frac{n^2 2 - n - n2 + 1}{n^2 2} \\
&= \frac{2n^2 - 3n + 1}{n^2 2}
\end{aligned}
$$

Then graphing this ratio from $n = 1$ to $n = 100$,

```
ratio_graph
```



From this graph we can see that the MSE ratio grows very quickly as n starts small and increases, meaning the difference between the MSE's is decreasing quickly. Then when n takes values of around 15 and higher, the ratio increases at a much slower rate, as the difference between the two MSE's becomes closer to 0 and the ratio approaches 1. Note that the ratio is always less than 1, hence the MSE for S^2 is always larger than $\widehat{\sigma^2_0}^{mle}$.

Appendix

Code for problem 1

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


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

  counts <- c(0, 0, 0)
  counts[1] <- sum(sample == "MM")
  counts[2] <- sum(sample == "FF")
  counts[3] <- sum(sample == "MF")

 data.frame(alpha = max(0, (counts[1]+counts[2]-counts[3])/40))})

boot_sim_mle <- do.call(rbind, boot_sim)

#i)
p1 <- ggplot(data=boot_sim_mle) +
  geom_histogram(aes(x=alpha),binwidth=0.05) +
  labs(title=expression("Bootstrapped Sampling Distribution of " ~ hat(alpha[0])^{mle}),
       x=expression(hat(alpha[0])^{mle}),
       y = "Count")

#ii)
center_samp <- mean(boot_sim_mle$alpha)

#iii)
stand_err <- sqrt(var(boot_sim_mle$alpha))

ci_standard_method <- boot_sim_mle %>% summarise(
  lower=center_samp - qnorm(0.975) * stand_err,
  upper=center_samp + qnorm(0.975) * stand_err)

ci_percentile_method <- boot_sim_mle %>%
  summarise(lower=quantile(alpha, 0.025),
            upper=quantile(alpha, 0.975))

table_data <- matrix(c(round(ci_percentile_method[1], 2),
                       round(ci_percentile_method[2], 2),
                       round(ci_standard_method[1], 2),
                       round(ci_standard_method[2], 2)),
                     nrow = 2, byrow = TRUE)
colnames(table_data) <- c("Lower", "Upper")
rownames(table_data) <- c("Percentile", "Standard")

tableiii <- kable(table_data)
```

Code for graph in problem 4

```
ratio_graph <- ggplot()+
  geom_function(fun=function(n){(2*n^2 - 3*n + 1)/(2*n^2)},
                xlim=c(1,100) ) +
  labs(title = "Ratio of MSE's",
       subtitle = hat(sigma[0])^{mle} ~ "vs" ~ S^2,
       y = "MSE Ratio",
       x = "n-value")
```