Homework 4: Written Part (SOLUTIONS)

STAT 343: Mathematical Statistics

Details

How to Write Up

The written part of this assignment can be either typeset using latex or hand written.

Grading

5% of your grade on this assignment is for turning in something legible and organized.

An additional 15% of your grade is for completion. A quick pass will be made to ensure that you've made a reasonable attempt at all problems.

In grading these problems, an emphasis will be placed on full explanations of your thought process. You don't need to write more than a few sentences for any given problem, but you should write complete sentences! Understanding and explaining the reasons behind what you are doing is at least as important as solving the problems correctly.

Collaboration

You are allowed to work with others on this assignment, but you must complete and submit your own write up. You should not copy large blocks of code or written text from another student.

Sources

You may refer to our text, Wikipedia, and other online sources. All sources you refer to must be cited in the space I have provided at the end of this problem set.

Problem 1. Adapted from Chihara and Hesterberg, Example 6.15

Let $X \sim Binomial(n, \theta)$, where n is known and θ is unknown. Let's consider two potential estimators for θ, $\hat{\theta}_1 = \frac{X}{n}$ and $\hat{\theta}_2 = \frac{X+1}{n+2}$, which adds one artificial success and one artificial failure to the real data.

(a) Are these estimators unbiased for θ?

$$\begin{aligned} E\left(\hat{\theta}_1\right) &= E\left(\frac{X}{n}\right) \\ &= \frac{1}{n}E(X) \\ &= \frac{1}{n}(n\theta) \\ &= \theta \end{aligned}$$

Since $Bias\left(\hat{\theta}_1\right) = E\left(\frac{X}{n}\right) - \theta = 0$, $\hat{\theta}_1$ is unbiased for θ.

$$
\begin{aligned}
E\left(\hat{\theta}_2\right) &= E\left(\frac{X+1}{n+2}\right) \\
&= \frac{1}{n+2} E(X+1) \\
&= \frac{1}{n+2}\left(E(X)+1\right) \\
&= \frac{n\theta+1}{n+2}
\end{aligned}
$$

$\hat{\theta}_2$ is not unbiased for θ.

(b) Find the mean squared error for $\hat{\theta}_1$.

$$
\begin{aligned}
MSE\left(\hat{\theta}_1\right) &= Var\left(\hat{\theta}_1\right) + Bias\left(\hat{\theta}_1\right)^2 \\
&= Var\left(\hat{\theta}_1\right) \text{ since this estimator is unbiased by (a)} \\
&= Var\left(\frac{X}{n}\right) \\
&= \frac{1}{n^2} Var(X) \\
&= \frac{n\theta(1-\theta)}{n^2} \\
&= \frac{\theta(1-\theta)}{n}
\end{aligned}
$$

For unbiased estimators, the MSE and the variance are the same.

(c) Find the mean squared error for $\hat{\theta}_2$.

$$
\begin{aligned}
MSE\left(\hat{\theta}_2\right) &= Var\left(\hat{\theta}_2\right) + Bias\left(\hat{\theta}_2\right)^2 \\
&= Var\left(\frac{X+1}{n+2}\right) + \left(\frac{n\theta+1}{n+2} - \theta\right)^2 \\
&= \frac{1}{(n+2)^2}\left(Var(X) + Var(1)\right) + \left(\frac{1-2\theta}{n+2}\right)^2 \\
&= \frac{n\theta(1-\theta) + (1-2\theta)^2}{(n+2)^2}
\end{aligned}
$$

(d) Plot the mean squared errors you found in (b) and (c) as a function of θ. You may do this either in R or *carefully by hand.*

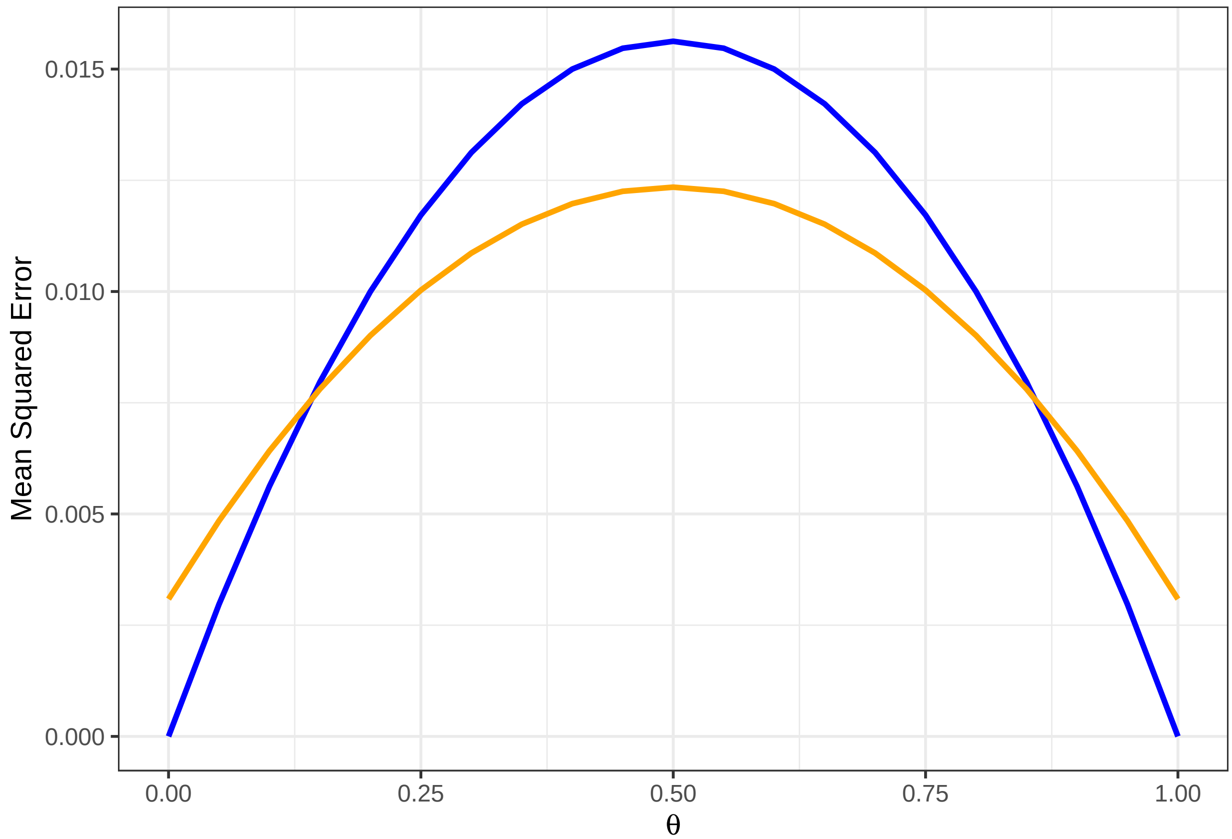
This needs to be done for some known n, so any n you chose is fine. I am going to do this for $n = 16$.

```r
mse_1 <- function(n, theta){
  theta*(1-theta)/n
}

mse_2 <- function(n, theta){
  (n*theta*(1-theta)+(1-2*theta)^2)/(n+2)^2
}
```

```r
plot_df <- data.frame(theta=seq(0,1,by=0.05),
                      mse1=mse_1(n=16, theta=seq(0,1,by=0.05)),
                      mse2=mse_2(n=16, theta=seq(0,1,by=0.05)))

ggplot(data=plot_df) +
  geom_line(aes(x=theta, y=mse1), color="blue", size=1) +
  geom_line(aes(x=theta, y=mse2), color="orange", size=1) +
  ylab("Mean Squared Error") +
  xlab(expression(theta)) +
  theme_bw()
```



You could also do it for multiple values of n if you were feeling ambitious (although this is not required).

(e) Why might someone choose to use $\hat{\theta}_2$ rather than $\hat{\theta}_1$ as an estimator? Refer to your plot in (d) as part of your answer.

For values that are closer to $\theta = 0.5$, the MSE is smaller for $\hat{\theta}_2$ than for $\hat{\theta}_1$.

Problem 2. Chihara and Hesterberg, Problem 6.30

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two different estimators for θ. Suppose $Var(\hat{\theta}_1) = 25$ and $Var(\hat{\theta}_2) = 4$.

(a) If $E(\hat{\theta}_1) = \theta$ and $E(\hat{\theta}_2) = \theta + 3$, which estimator has the smaller mean squared error?

$$\begin{aligned} MSE\left(\hat{\theta}_1\right) &= Var\left(\hat{\theta}_1\right) + Bias\left(\hat{\theta}_1\right)^2 \\ &= 25 + 0 \\ &= 25 \end{aligned}$$

$$\begin{aligned} MSE\left(\hat{\theta}_2\right) &= Var\left(\hat{\theta}_2\right) + Bias\left(\hat{\theta}_2\right)^2 \\ &= 4 + (\theta + 3 - \theta)^2 \\ &= 4 + 9 \\ &= 13 \end{aligned}$$

$\hat{\theta}_2$ has the smaller MSE.

(b) Suppose $E(\hat{\theta}) = \theta$ and $E(\hat{\theta}_2) = \theta + b$, for some positive number b. For what value (if any) does $\hat{\theta}_2$ have a smaller mean square error than $\hat{\theta}_1$?

In order for $\hat{\theta}_2$ to have a smaller mean square error than $\hat{\theta}_1$, we require that

$$4 + b^2 < 25$$

So,

$$b^2 < 21$$

Then,

$$b < \sqrt{21}$$

where $\sqrt{21}$ denotes the positive square root. So, for $0 < b < \sqrt{21}$, $\hat{\theta}_2$ has a smaller mean square error than $\hat{\theta}_1$.