Math 152 - Statistical Theory - Homework 3

write your name here

Due: Friday, September 11, midnight PDT

Important Note:

You should work to turn in assignments that are clear, communicative, and concise. Part of what you need to do is not print pages and pages of output. Additionally, you should remove these exact sentences and the information about HW scoring below.

Click on the *Knit to PDF* icon at the top of R Studio to run the R code and create a PDF document simultaneously. [PDF will only work if either (1) you are using R on the network, or (2) you have LaTeX installed on your computer. Lightweight LaTeX installation here: https://yihui.name/tinytex/]

> Either use the college's RStudio server (https://rstudio.pomona.edu/) or install R and R Studio on to your personal computer. See: https://research.pomona.edu/johardin/math152f20/ for resources.

Assignment

1: PodQ

Describe one thing you learned from someone in your pod this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

2: 7.3.21

Suppose that $X_1, \ldots, X_n$ form a random sample from the exponential distribution with parameter θ. Let the prior distribution of θ be improper with "p.d.f." $1/\theta$ for $\theta > 0$. Find the posterior distribution of θ and show that the posterior mean of θ is $1/\overline{X}$.

3: 7.4.3

Consider again the conditions of Exercise 2. Suppose that the prior distribution of θ is as given in Exercise 2, and suppose again that 20 items are selected at random from the shipment.

a. For what number of defective items in the sample will the mean squared error of the Bayes estimate be a maximum?

b. For what number will the mean squared error of the Bayes estimate be a minimum?

4: 7.4.6

Suppose that a random sample of size n is taken from a Poisson distribution for which the value of the mean θ is unknown, and the prior distribution of θ is a gamma distribution for which the mean is μ_0. Show that the mean of the posterior distribution of θ will be a weighted average having the form $\gamma\overline{X} + (1-\gamma)\mu_0$, and show that $\gamma \to 1$ as $n \to \infty$.

5: 7.4.9

Suppose that a random sample is to be taken from a normal distribution for which the value of the mean θ is unknown and the standard deviation is 2, the prior distribution of θ is a normal distribution for which the standard deviation is 1, and the value of θ must be estimated by using the squared error loss function. What is the smallest random sample that must be taken in order for the mean squared error of the Bayes estimator of θ to be 0.01 or less? (See Exercise 10 of Sec. 7.3.)

6: 7.4.10

Suppose that the time in minutes required to serve a customer at a certain facility has an exponential distribution for which the value of the parameter θ is unknown, the prior distribution of θ is a gamma distribution for which the mean is 0.2 and the standard deviation is 1, and the average time required to serve a random sample of 20 customers is observed to be 3.8 minutes. If the squared error loss function is used, what is the Bayes estimate of θ? (See Exercise 12 of Sec. 7.3.)

7: 7.4.12

Let θ denote the proportion of registered voters in a large city who are in favor of a certain proposition. Suppose that the value of θ is unknown, and two statisticians A and B assign to θ the following different prior p.d.f.s $\xi_A(\theta)$ and $\xi_B(\theta)$, respectively:

$$\begin{aligned} \xi_A(\theta) &= 2\theta \text{ for } 0 < \theta < 1, \\ \xi_B(\theta) &= 4\theta^3 \text{ for } 0 < \theta < 1. \end{aligned}$$

In a random sample of 1000 registered voters from the city, it is found that 710 are in favor of the proposition.

a. Find the posterior distribution that each statistician assigns to θ.

b. Find the Bayes estimate for each statistician based on the squared error loss function.

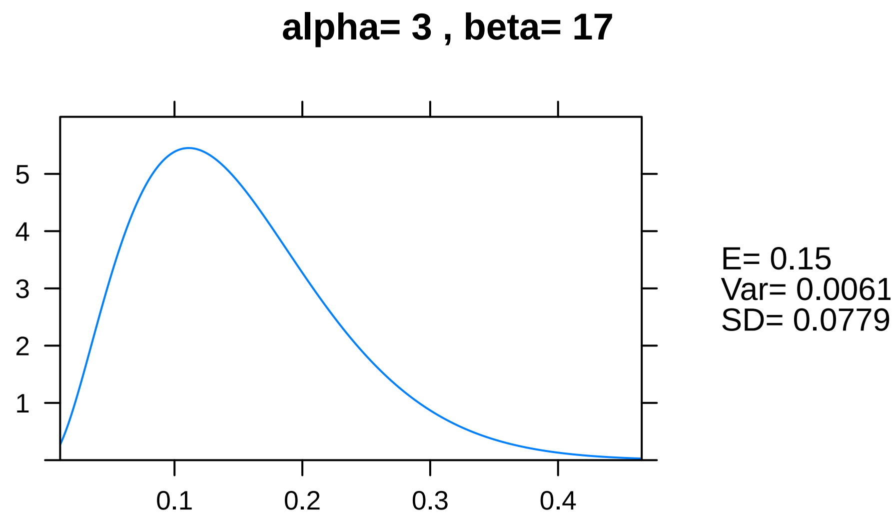
c. Show that after the opinions of the 1000 registered voters in the random sample had been obtained, the Bayes estimates for the two statisticians could not possibly differ by more than 0.002, regardless of the number in the sample who were in favor of the proposition.

8: R Baseball & Bayes

Consider the baselball problem we discussed in class. (See website for a copy of the handout.) Let α_0 and β_0 be your prior parameters.

a. What are your choices of α_0 and β_0? [Use α_0 and β_0 values that were **not** discussed in class.] What features of the plot of the prior density function made you think these were good choices? I wrote a function below that you can use to try out different options for the prior.

```
betaplot<-function(a,b){
ex<-a/(a+b)
varx<-a*b/((a+b)^2*(a+b+1))
plotDist('beta', params = list(a,b),
         main = paste("alpha=",a,", beta=",b),
         key=list(space="right",
         text=list(c(paste("E=",round(ex,4)),
                        paste("Var=",round(varx,4)),
        paste("SD=", round(sqrt(varx),4))))))
}
betaplot(a=3,b=17) # change a and b to something we did not do in class
```



b. Using properties of expectation [that is, consider both estimates as functions of X, not of θ], find the bias and variance of $\hat{\theta}_f$ and $\hat{\theta}_b$. You are a frequentist here, and your answers should both be functions of θ.

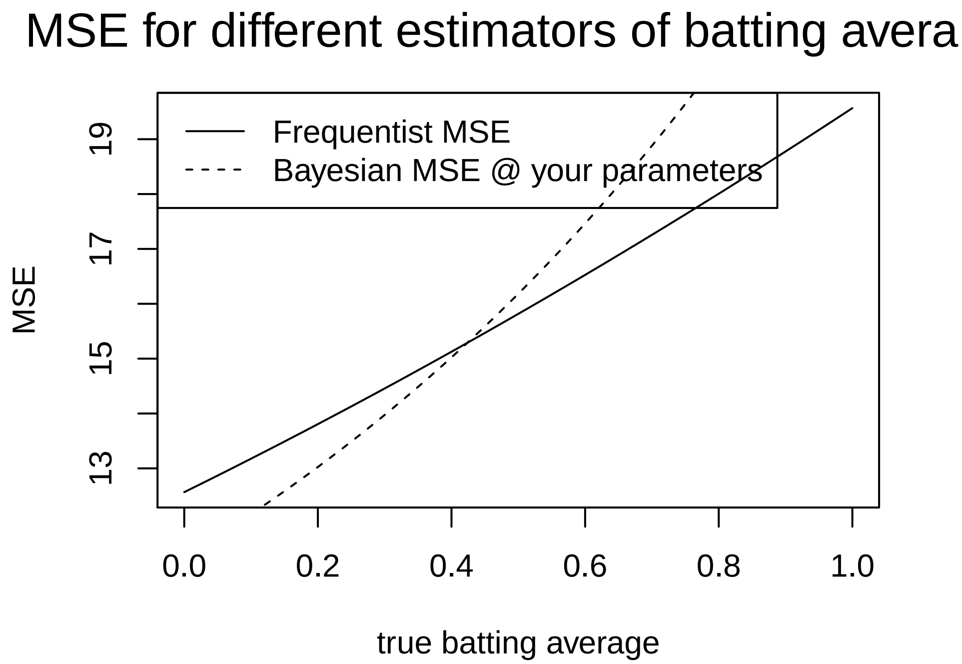
c. Based on your comparison of the MSE, do you recommend using $\hat{\theta}_f$ or $\hat{\theta}_b$? Explain. [Hint: first determine whether one estimator has a smaller MSE. Over what region?]

Hint1: first determine whether one estimator has a smaller MSE. Over what region? Remember from class that Ty Cobb has the best batting average **ever**, and his average was 0.366.

Hint2: say you think the mean squared error associated with $\hat{\theta}_f$ is $\theta^2 + 6\theta + 4\pi$, and you think that the mean squared error associated with $\hat{\theta}_b$ is $\exp(\theta + 2) + 4$. Those functions are totally wrong, but if I thought they were true, I could plot them using the code below.

Hint3: Note that in my totally made up graph, it seems like the Bayesian estimator has a smaller MSE whenever the true batting average is less than 0.4ish.

```
theta<-seq(0,1,.01) # theata is his true batting average
plot(theta,theta^2 + 6*theta + 4*pi,type="l",lty=1,
     xlab="true batting average",ylab="MSE") #mse.f
lines(theta,exp(theta + 2) + 4,lty=2) #mse.b
legend(x="topleft",c("Frequentist MSE", "Bayesian MSE @ your parameters"),
        lty=c(1:5))
mtext("MSE for different estimators of batting average",line=1,cex=1.5)
```



d. If John Spurrier gets three hits in ten at bats, what is your estimate of θ? (Given your answer to c.)

e. Show that in the beta-binomial family, $\hat{\theta}_b$ is a weighted average of $\hat{\theta}_f$ and the prior mean.