Math 152 - Statistical Theory - Homework 3

write your name here

Due: 9/21/2018

Book problems (if you have the 3rd edition of the book, the problems will be the same unless they don't exist – that is, the 4th edition *added* problems but didn't change them). Stop by my office or ask me in class if you want to see the statement of the problem.)

- 7.3 – 21
- 7.4 – 3, 6, 9, 10, 12

n.b., remember that your book parameterizes both the exponential and the gamma differently from your sheet.

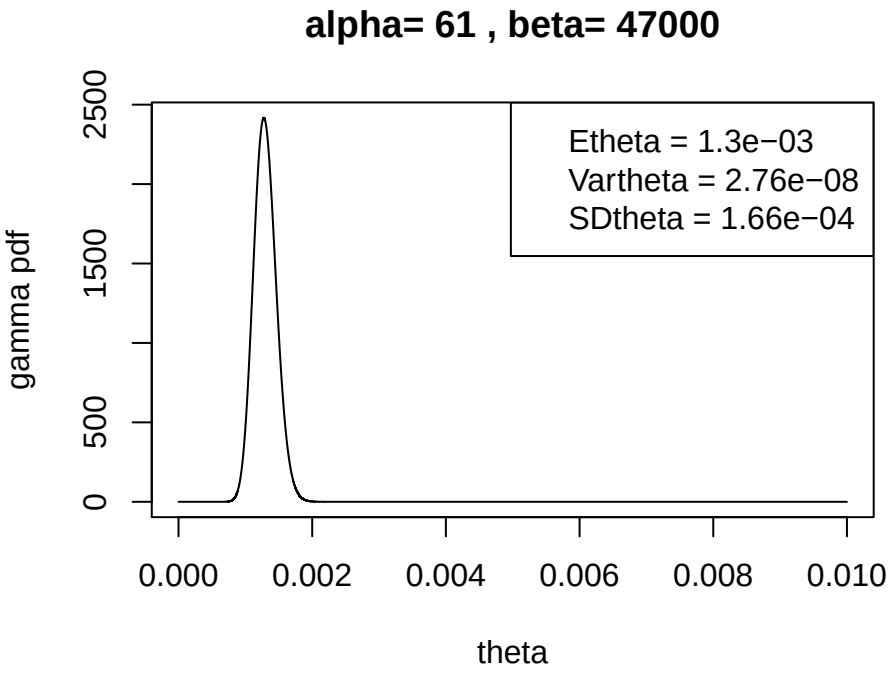
R problem

Consider the kidney cancer problem we discussed in class. (See class notes posted to Sakai for more information.) Let θ be the rate of some cancer in the US, and say you are only able to measure one county's incidence. Let α_0 and β_0 be your prior parameters for the gamma distribution.

(a) The data from class is from the 1980's and kidney cancer. Do some quick googling and see if you can come up with a prior related to a cancer of your choice [do not spend too much time trying to find a perfect prior, just find something reasonable]. 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 gamma density function made you think these were good choices?

```
possible_theta<-seq(0,.01,length=100000) # true cancer rate
a = 61
b = 47000
etheta<-a/b
vartheta<-a/b^2

plot(possible_theta, dgamma(possible_theta,a,b),type="l",xlab="theta",ylab="gamma pdf",
        main=paste("alpha=",a,", beta=",b))
legend(x="topright", c(paste("Etheta =", format(etheta, scientific = TRUE, digits = 3), sep=" "),
                       paste("Vartheta =", format(vartheta, scientific = TRUE, digits = 3), sep=" "),
                       paste("SDtheta =", format(sqrt(vartheta), scientific = TRUE, digits = 3), sep="
```



(b) Using frequentist 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 $\hat{\theta}_b$ (that is, you need to come up with two different estimators, one as a Bayesian, one as a frequentist). In *comparing* the two estimators, you are a frequentist, and your answers should be functions of θ.

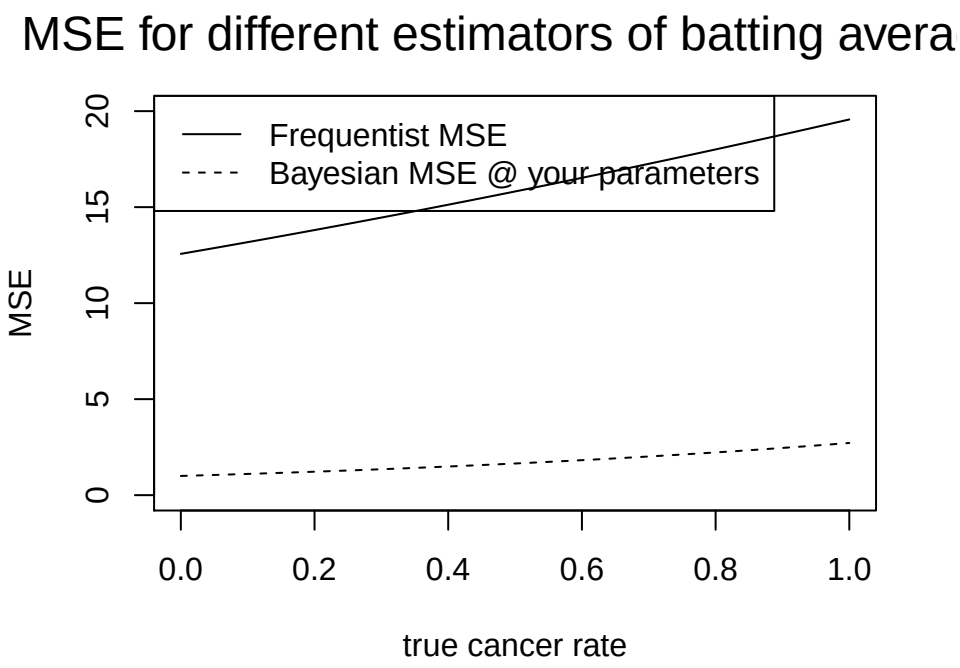
(c) Based on your comparison of the MSE, do you recommend using $\hat{\theta}_f$ or $\hat{\theta}_b$? Explain.

Hint1: first determine whether one estimator has a smaller MSE. Over what region? Remember from class that the county sizes varied from 55 people to almost 10 million people.

Hint2: Let's 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 $\exp(\theta)$.

Hint3: The plot below uses θ on the x-axis. Maybe your MSE is also a function of n? If so, try drawing multiple lines (`lines` function) at different values of n.

```
possible_theta<-seq(0,1,length = 10000) #true cancer rate
plot(possible_theta,possible_theta^2 + 6*possible_theta + 4*pi,type="l",lty=1,
     xlab="true cancer rate",ylab="MSE", ylim=c(0,20)) #mse.f
# you should change the y limits so that your graph is more readable
lines(possible_theta,exp(possible_theta),lty=2) #mse.b (your parameters)
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)
```


MSE for different estimators of batting avera
Frequentist MSE
Bayesian MSE @ your parameters
MSE
0
5
10
15
20
0.0
0.2
0.4
0.6
0.8
1.0
true cancer rate