Math 152 - Statistical Theory - Homework 3

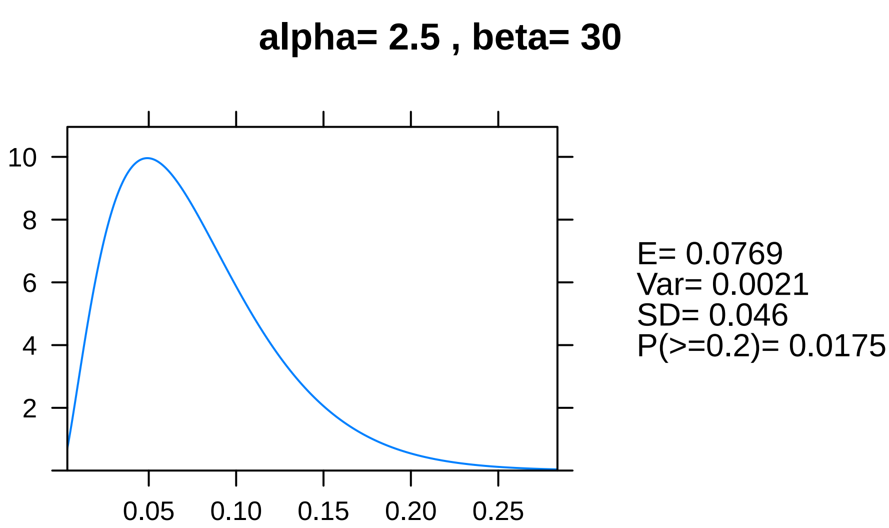
Ethan Ashby

Due: Friday, September 11, midnight PDT

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)), paste("P(>=0.2)=", round(pbeta(0.2, a, b, lower.tail=FALSE),
}
betaplot(a=2.5,b=30) # change a and b to something we did not do in class
```



I choose $\alpha_0 = 2.5$ and $\beta_0 = 30$ for it's center and spread. I'm skeptical that this player is capable of performing at a major league level. This distributions' center and (prior expected batting average=0.08) and probability of exceeded a barely adequate level (P(batting at 0.2)=0.0175) jive with my prior belief that this player would not succeed in the majors.

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.

(i)

$$MSE(\hat{\theta}_f) = E[(\hat{\theta}_f - \theta)^2] = Var(\hat{\theta}_f) + [E(\hat{\theta}_f) - \theta]^2$$

$$E(\hat{\theta}_f) = \frac{1}{n}E(x) = \frac{1}{n}(n\theta) = \theta$$

$$Var(E(\hat{\theta}_f)) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

$$\text{Bias} : [E(\hat{\theta}_f) - \theta]^2 = [\theta - \theta]^2 = 0$$

$$\text{Variance} : Var(E(\hat{\theta}_f)) = \frac{\theta(1-\theta)}{10}$$

(ii) **Note that X denotes the number of successful trials: i.e. base hits**

$$\hat{\theta}_b = \frac{X + \alpha}{n + \alpha + \beta}$$

$$E(\hat{\theta}_b) = \frac{E(X) + \alpha}{n + \alpha + \beta} = \frac{n\theta + \alpha}{n + \alpha + \beta}$$

$$Var(E(\hat{\theta}_b)) = Var(\frac{X + \alpha}{n + \alpha + \beta}) = \frac{Var(X)}{(n + \alpha + \beta)^2} = \frac{n\theta(1-\theta)}{(n + \alpha + \beta)^2}$$

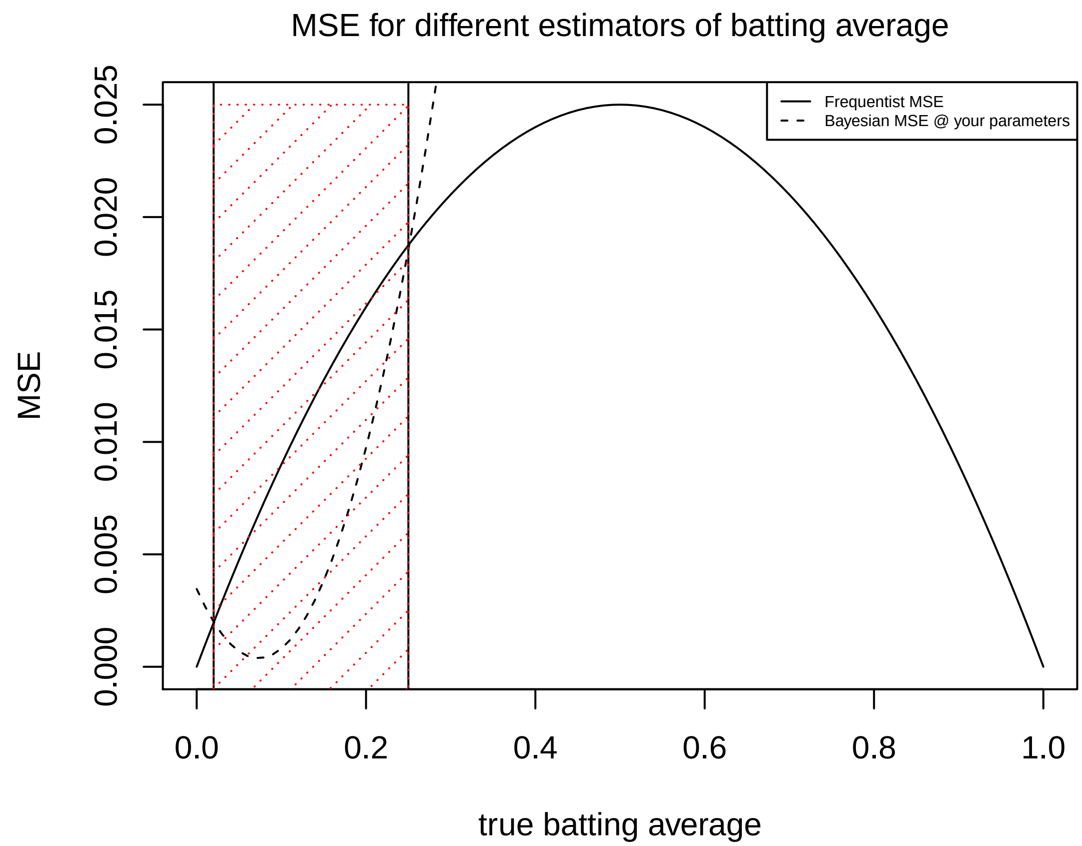
$$\text{Bias} : [E(\hat{\theta}_b) - \theta]^2 = \frac{n\theta + \alpha}{n + \alpha + \beta} - \theta = \frac{-(\alpha + \beta)\theta + \alpha}{n + \alpha + \beta} = \frac{-32.5\theta + 2.5}{42.5}$$

$$\text{Variance} : Var(E(\hat{\theta}_b)) = \frac{n\theta(1-\theta)}{(n + \alpha + \beta)^2} = \frac{10\theta - 10\theta^2}{42.5^2}$$

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?]

I would use the bayesian estimator in this case. The bayesian estimator has a lower MSE over the range $\theta \in [0.02, 0.25]$ which seems like a reasonable range of batting average for this not super accomplished batter.

```
theta<-seq(0,1,.01) # theata is his true batting average
plot(theta, theta*(1-theta)/10,type="l",lty=1,
     xlab="true batting average",ylab="MSE") #mse.f
lines(theta, (10*theta-10*theta^2)/42.5^2+((-32.5*theta+2.5)/42.5)^2,lty=2) #mse.b
abline(v=0.02)
abline(v=0.25)
rect(0.02, -0.1, 0.25, 0.025, col="red", density = 7, border = "red", lty=3)
legend(x="topright",c("Frequentist MSE", "Bayesian MSE @ your parameters"),
        lty=c(1:5), bg="white", cex=0.5)
mtext("MSE for different estimators of batting average",line=1,cex=1)
```



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

$\hat{\theta}_b = \frac{X+\alpha}{n+\alpha+\beta} = \frac{3+2.5}{10+2.5+30} = 0.129$

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

$$\hat{\theta}_b = \frac{X+\alpha}{n+\alpha+\beta}$$
$$\hat{\theta}_f = \frac{X}{n}$$
$$E[\xi(\theta)] = \frac{\alpha}{\alpha+\beta}$$

$$\hat{\theta}_b = \frac{n}{n+\alpha+\beta} * \frac{X}{n} + \frac{\alpha+\beta}{n+\alpha+\beta} * \frac{\alpha}{\alpha+\beta} = \gamma(\frac{X}{n}) + (1-\gamma)(\frac{\alpha}{\alpha+\beta})$$

This is a weighted average of the prior mean and frequentist estimator.