Math 152 - Statistical Theory - Homework 2

Ethan Ashby

Due: Friday, September 4, midnight PDT

10: R - beta-binomial family Consider the beta-binomal family (i.e., beta prior, binomial likelihood (with parameter theta), beta posterior). That is, the parameter of interest is theta, and both the prior and posterior distributions of theta are from the beta family.

(c) Using simulations, histograms, and means, **discuss the role of sample size** when using a prior and Bayesian inference. For the discussion:

i. give posterior histogram and sample means for the following combinations (12 histograms):

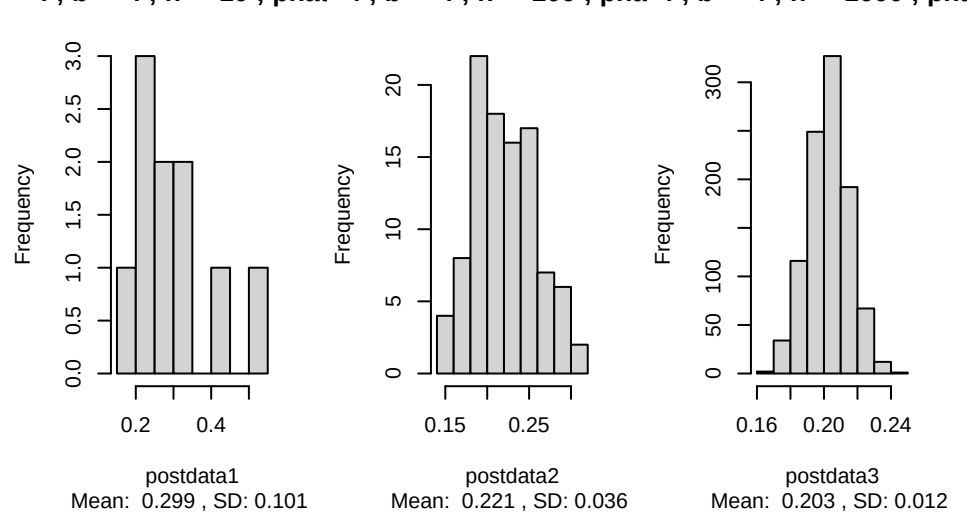
- (α, β): (4,4); (4,10)
- $\hat{p}$: 0.2, 0.5
- n: 10, 100, 1000

ii. Using your histograms and means above, discuss the role of sample size in determining the posterior distribution of the parameter.

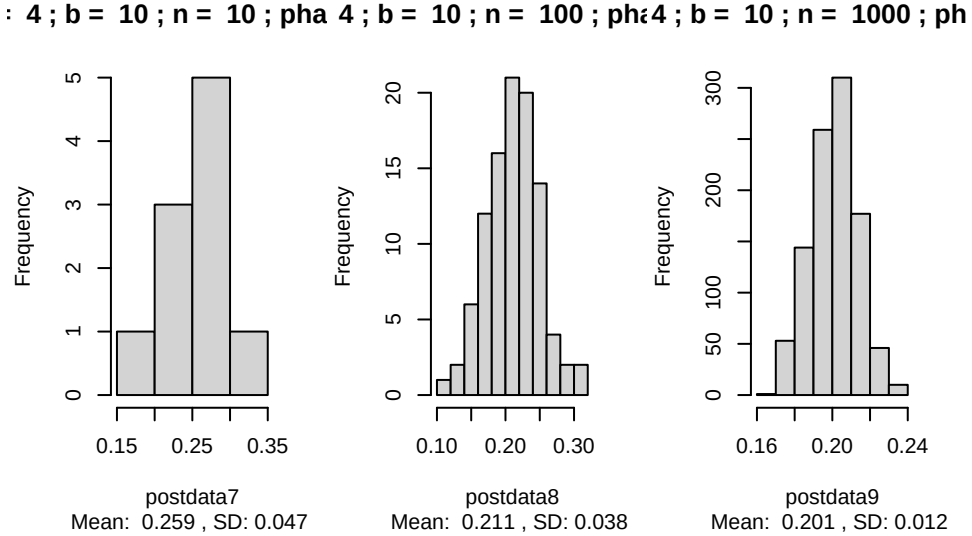
Some R code that might be helpful:

```
set.seed(11)

a1 = 4
b1 = 4
b2= 10
phat1=0.2
phat2=0.5
n1=10
n2=100
n3=1000

par(mfrow=c(1,3))

#alpha=4, beta=4, n=10, phat=0.2
postdata1 = rbeta(n1, a1+phat1*n1, b1+n1*(1-phat1))
hist(postdata1, main = paste("a = ", a1, "; b = ", b1, "; n = ", n1, "; phat = ", phat1), sub=paste("Mea

#alpha=4, beta=4, n=100, phat=0.2
postdata2 = rbeta(n2, a1+phat1*n2, b1+n2*(1-phat1))
hist(postdata2, main = paste("a = ", a1, "; b = ", b1, "; n = ", n2, "; phat = ", phat1), sub=paste("Mea
```

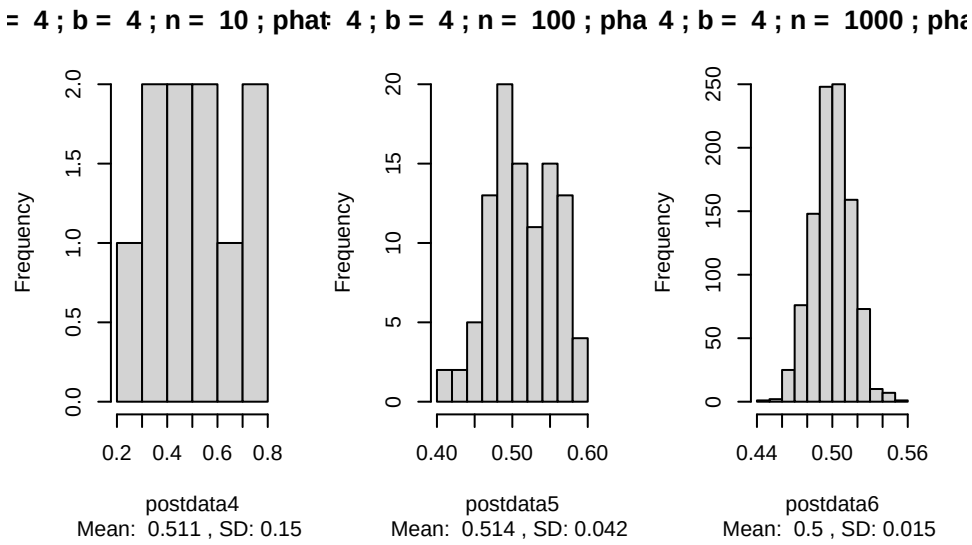
```r
#alpha=4, beta=4, n=1000, phat=0.2
postdata3 = rbeta(n3, a1+phat1*n3, b1+n3*(1-phat1))
hist(postdata3, main = paste("a = ", a1, "; b = ", b1, "; n = ", n3, "; phat = ", phat1), sub=paste("Mea
```



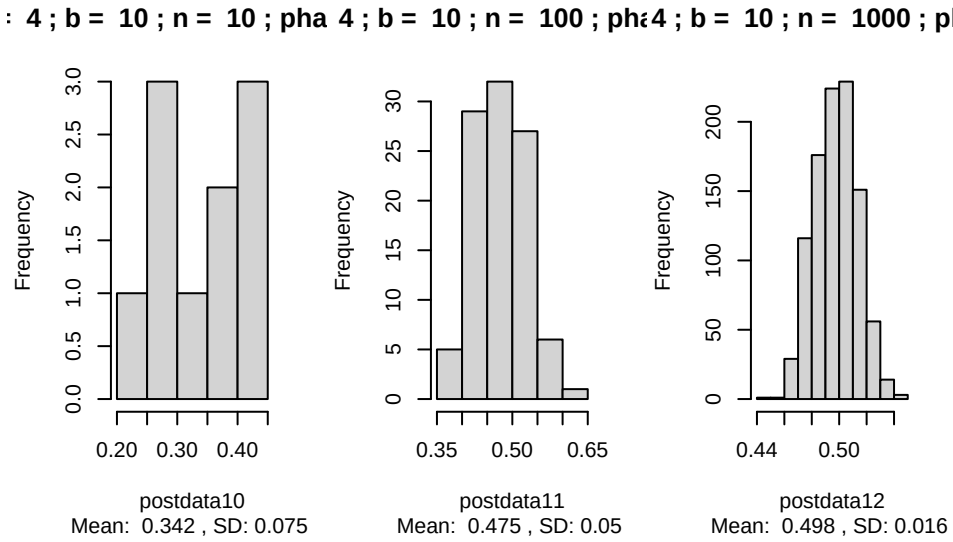

```r
par(mfrow=c(1,3))

#alpha=4, beta=10, n=10, phat=0.2
postdata7 = rbeta(n1, a1+phat1*n1, b2+n1*(1-phat1))
hist(postdata7, main = paste("a = ", a1, "; b = ", b2, "; n = ", n1, "; phat = ", phat1), sub=paste("Mea

#alpha=4, beta=10, n=100, phat=0.2
postdata8 = rbeta(n2, a1+phat1*n2, b2+n2*(1-phat1))
hist(postdata8, main = paste("a = ", a1, "; b = ", b2, "; n = ", n2, "; phat = ", phat1), sub=paste("Mea

#alpha=4, beta=10, n=1000, phat=0.2
postdata9 = rbeta(n3, a1+phat1*n3, b2+n3*(1-phat1))
hist(postdata9, main = paste("a = ", a1, "; b = ", b2, "; n = ", n3, "; phat = ", phat1), sub=paste("Mea
```



```r
par(mfrow=c(1,3))

#alpha=4, beta=4, n=10, phat=0.5
postdata4 = rbeta(n1, a1+phat2*n1, b1+n1*(1-phat2))
hist(postdata4, main = paste("a = ", a1, "; b = ", b1, "; n = ", n1, "; phat = ", phat2), sub=paste("Mea

#alpha=4, beta=4, n=100, phat=0.5
postdata5 = rbeta(n2, a1+phat2*n2, b1+n2*(1-phat2))
hist(postdata5, main = paste("a = ", a1, "; b = ", b1, "; n = ", n2, "; phat = ", phat2), sub=paste("Mea

#alpha=4, beta=4, n=1000, phat=0.5
postdata6 = rbeta(n3, a1+phat2*n3, b1+n3*(1-phat2))
hist(postdata6, main = paste("a = ", a1, "; b = ", b1, "; n = ", n3, "; phat = ", phat2), sub=paste("Mea
```




```r
par(mfrow=c(1,3))

#alpha=4, beta=10, n=10, phat=0.5
postdata10 = rbeta(n1, a1+phat2*n1, b2+n1*(1-phat2))
hist(postdata10, main = paste("a = ", a1, "; b = ", b2, "; n = ", n1, "; phat = ", phat2), sub=paste("M

#alpha=4, beta=10, n=100, phat=0.5
postdata11 = rbeta(n2, a1+phat2*n2, b2+n2*(1-phat2))
hist(postdata11, main = paste("a = ", a1, "; b = ", b2, "; n = ", n2, "; phat = ", phat2), sub=paste("M

#alpha=4, beta=10, n=1000, phat=0.5
postdata12 = rbeta(n3, a1+phat2*n3, b2+n3*(1-phat2))
hist(postdata12, main = paste("a = ", a1, "; b = ", b2, "; n = ", n3, "; phat = ", phat2), sub=paste("M
```



For (a,b)=(4,4) and $\hat{p} = 0.2$, increasing the sample size decreased the mean and spread.

For (a,b)=(4,10) and $\hat{p} = 0.2$, increasing the sample size decreased the mean and spread.

For (a,b)=(4,4) and $\hat{p} = 0.5$, increasing the sample size increased the mean and decreased the spread.

For (a,b)=(4,10) and $\hat{p} = 0.5$, increasing the sample size increased the mean and decreased the spread.