Charles Hmong
Dr. Matthews
STAT 488-001
22 June 2022

1 5.13A) $p(\theta,\alpha,\beta|y) = p(\alpha,\beta)\,p(\theta|\alpha,\beta)\,p(y|\theta,\alpha,\beta) \propto$

$$\frac{1}{(\alpha+\beta)^2}\left(\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1}\right)\left(\prod_{j=1}^{J}\theta_j^{y_j}(1-\theta_j)^{n_j-y_j}\right) \propto$$

$$\boxed{\frac{1}{(\alpha+\beta)^2}\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^J \prod_{j=1}^{J}\theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}}$$

5.13B) $p(\theta|\alpha,\beta,y) = \prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta+n_j)}{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}\theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1} \sim Beta(\alpha+y_j,\beta+n_j-y_j)$

$$p(\alpha,\beta|y) = \frac{p(\theta,\alpha,\beta|y)}{p(\theta|\alpha,\beta,y)} \propto \boxed{\frac{1}{(\alpha+\beta)^2}\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^J\prod_{j=1}^{J}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}} \propto$$

$$\frac{1}{(\alpha+\beta)^2}\prod_{j=1}^{J}\left(\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)^{-1}\cdot\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)} \propto$$

$$\frac{1}{(\alpha+\beta)^2}\prod_{j=1}^{J}\frac{B(\alpha+y_j,\beta+n_j-y_j)}{B(\alpha,\beta)}$$

where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function (for easier computing)

5.13F) Yes, it appears using a beta distribution for the θ_j's was a reasonable decision. Looking at the data and varias output from problem 5.13C, the data fit well.

2 5.14A) $p(\theta,\alpha,\beta|y) = p(\alpha,\beta)\,p(\theta|\alpha,\beta)\,p(y|\theta,\alpha,\beta) \propto$

$$p(\alpha,\beta)\left(\prod_{j=1}^{J}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta_j^{\alpha-1}e^{-\beta\theta_j}\right)\left(\prod_{j=1}^{J}\frac{\theta_j^{y_j}e^{-\theta_j}}{y_j!}\right) \propto$$

$$\boxed{\sqrt{\alpha}(\beta+1)\prod_{j=1}^{J}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta_j^{y_j+\alpha-1}e^{-(\beta+1)\theta_j}}$$

5.14B) $p(\theta|\alpha,\beta,y) = \prod_{j=1}^{J}\frac{(\beta+1)^{\alpha+y_j}}{\Gamma(\alpha+y_j)}\theta_j^{y_j+\alpha-1}e^{-(\beta+1)\theta_j}$

$$p(\alpha,\beta|y) = \frac{p(\theta,\alpha,\beta|y)}{p(\theta|\alpha,\beta,y)} \propto \boxed{\sqrt{\alpha}(\beta+1)\prod_{j=1}^{J}\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+y_j)}{(\beta+1)^{\alpha+y_j}}} \propto$$

$$\frac{\Gamma(\alpha+y_j)}{\Gamma(\alpha)} = \Gamma(y_j)\frac{\Gamma(\alpha+y_j)}{\Gamma(\alpha)\Gamma(y_j)} = \frac{\Gamma(y_j)}{B(\alpha,y_j)}$$

$$\sqrt{\alpha}(\beta+1)\prod_{j=1}^{J}\frac{\beta^\alpha}{(\beta+1)^{\alpha+y_j}}\frac{\Gamma(y_j)}{B(\alpha,y_j)}$$

where $B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function

5.14D) We can see from problem 5.14C that the posterior density is integrable.

Homework 3

Charles Hwang

6/23/2022

Charles Hwang

Dr. Matthews

STAT 488-001

23 June 2022

Problem 5.13

Problem 5.13(b)

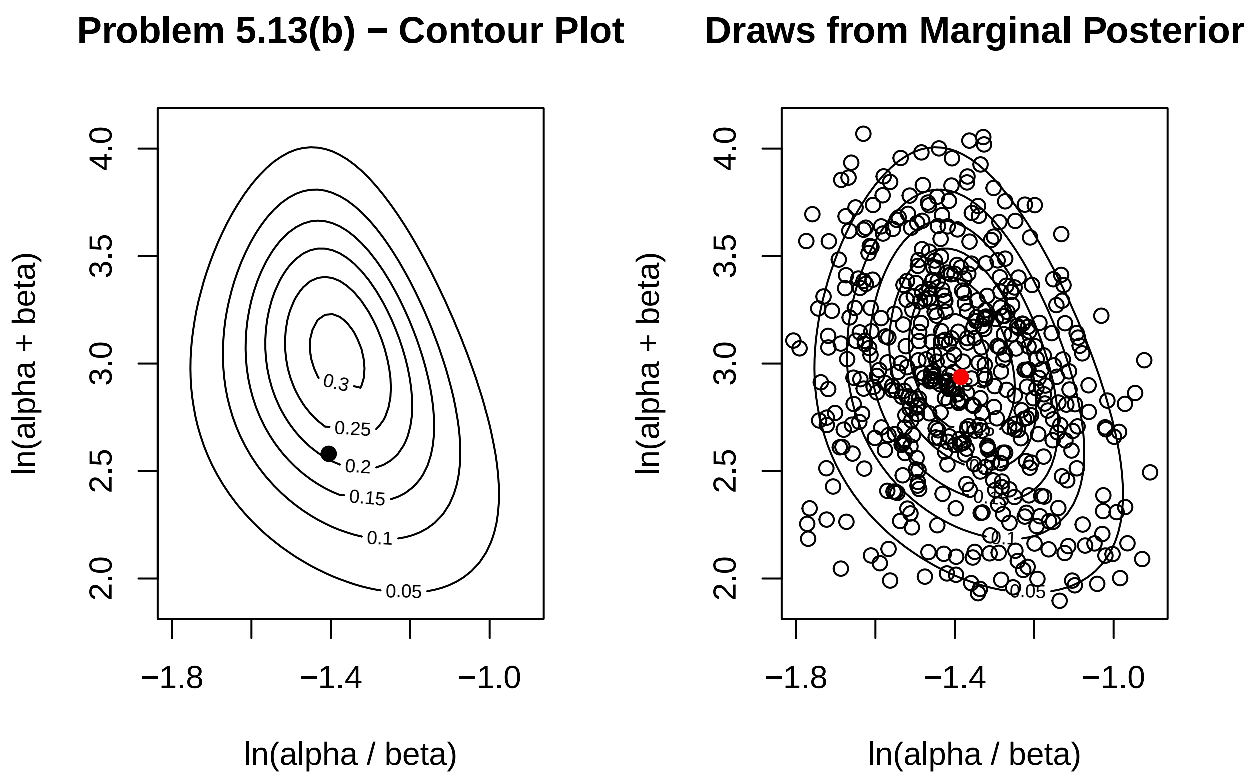
```
rm(list=ls())
y<-c(16,9,10,13,19,20,18,17,35,55)
o<-c(58,90,48,57,103,57,86,112,273,64)
n<-y+o
mean(y/n)    # mean = alpha / (alpha+beta)
```

```
## [1] 0.1961412
```

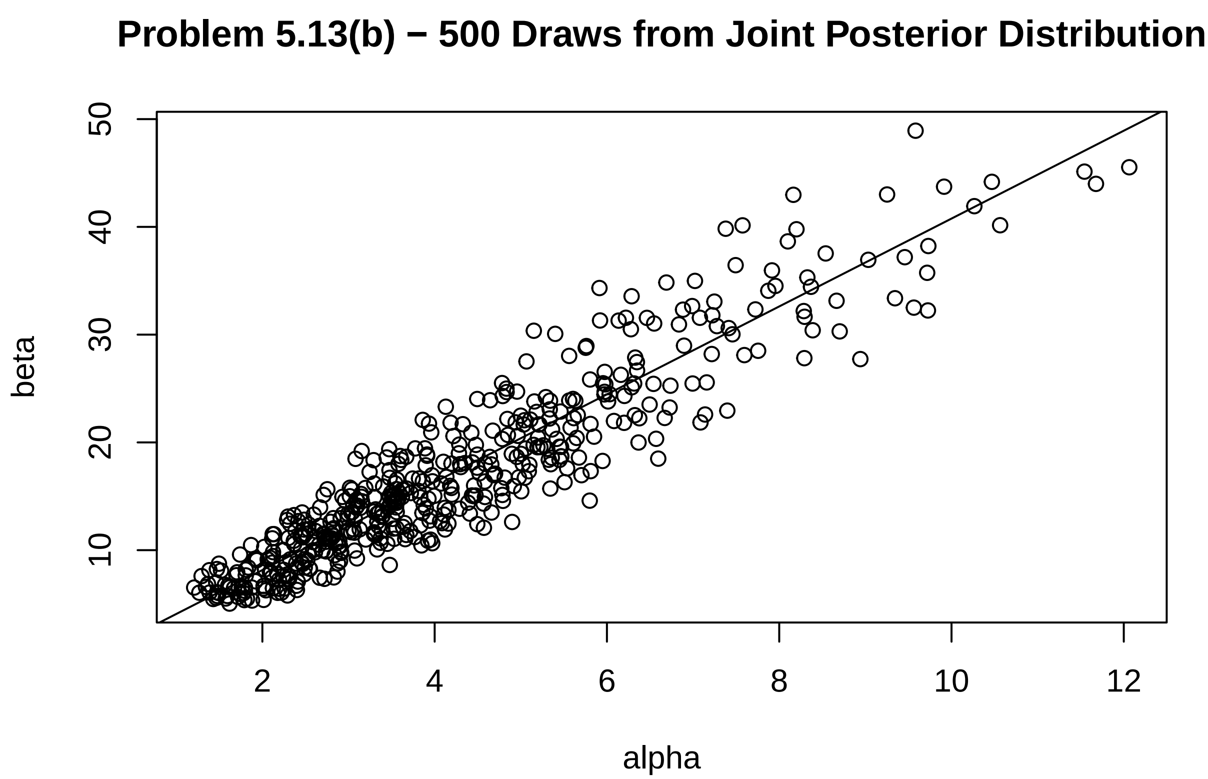
```
var(y/n) # variance = alpha*beta / ((alpha+beta)^2*(alpha+beta+1))
```

```
## [1] 0.01112067
```

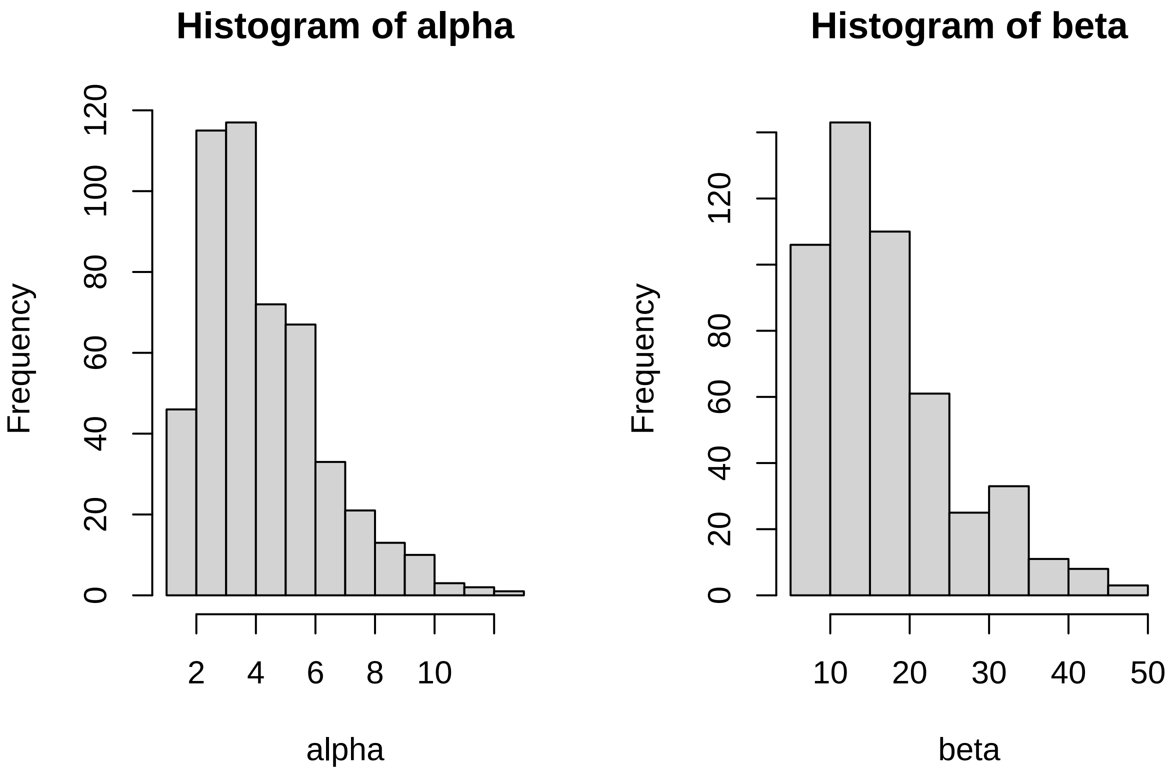
```
a<-2.6 # Solving for alpha and beta in Desmos and rounding to nearest tenth
b<-10.6
mp<-function(X,Y){a<-exp(Y)*exp(X)/(exp(X)+1)
  b<-exp(Y)/(exp(X)+1)
  10^230*(a+b)^(-2)*prod(beta(a+y,b+n-y)/beta(a,b))*exp(2*Y+X)/(exp(X)+1)^2} # Jacobian
X<-seq(-1.8,-0.9,length=50) # Setting bounds of contour plot via trial-and-error
Y<-seq(1.9,4.1,length=50)
z<-matrix(NA,length(Y),length(X))
for (i in 1:length(Y)){for (j in 1:length(X)){z[j,i]<-mp(X[j],Y[i])}}
par(mfrow=c(1,2))
contour(X,Y,z,xlab="ln(alpha / beta)",ylab="ln(alpha + beta)",main="Problem 5.13(b) - Contour Plot")
points(log(a/b),log(a+b),pch=19)
Xm<-Ym<-Xn<-Yn<-rep(NA,500)
set.seed(226)
for (k in 1:length(Xm)){Xm[k]<-sample(X,1,TRUE,apply(z,1,sum)/sum(apply(z,1,sum)))
  Ym[k]<-sample(Y,1,TRUE,z[which(Xm[k]==X),]/sum(z[which(Xm[k]==X),]))}
contour(X,Y,z,xlab="ln(alpha / beta)",ylab="ln(alpha + beta)",main="Draws from Marginal Posterior")
Xm<-Xm+runif(length(Xm),-diff(X)/2,diff(X)/2) # Adding random jitter
Ym<-Ym+runif(length(Ym),-diff(Y)/2,diff(Y)/2)
points(Xm,Ym)
points(mean(Xm),mean(Ym),pch=19,col="red") # Mean of samples
```




```
alpha<-exp(Ym)*(exp(Xm)/(exp(Xm)+1)) # Now drawing from joint posterior distribution
beta<-exp(Ym)/(exp(Xm)+1)
par(mfrow=c(1,1))
plot(alpha,beta,main="Problem 5.13(b) - 500 Draws from Joint Posterior Distribution")
abline(0,b/a) # Expected slope
```



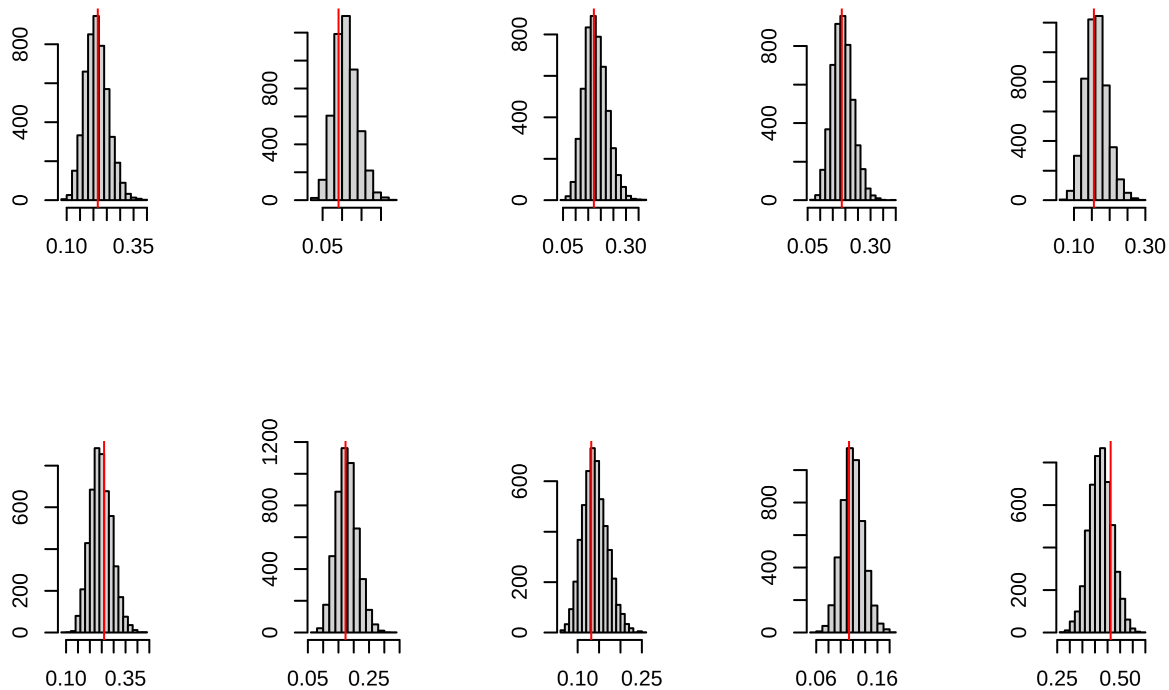
```r
par(mfrow=c(1,2))
hist(alpha)
hist(beta)
```



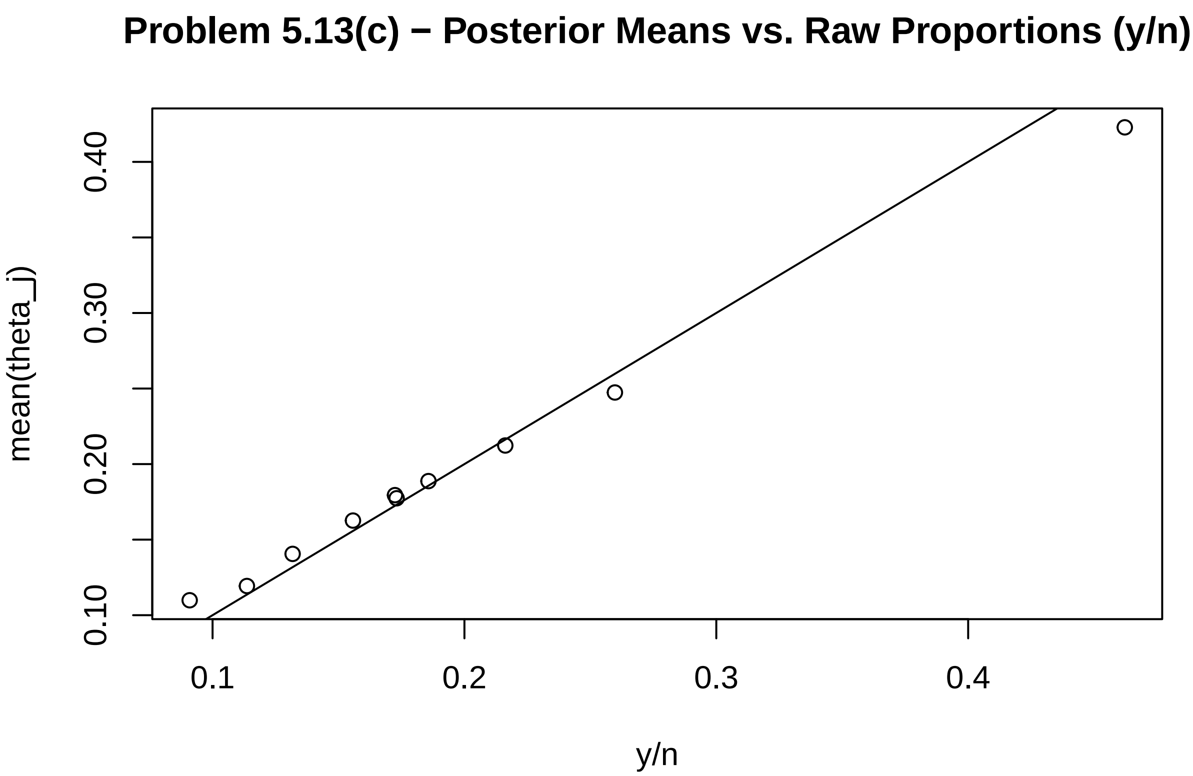
Problem 5.13(c)

```r
library(ggplot2)
t<-matrix(NA,5000,length(y))
pd<-data.frame()
set.seed(226)
for (i in 1:length(y)){t[,i]<-rbeta(5000,alpha+y[i],beta+n[i]-y[i])
pd<-rbind(pd,data.frame(t[,i],rep(i,5000)))}
names(pd)<-c("theta","i")
par(mfrow=c(2,5))
hist(subset(pd,i==1)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[1],col="red")
hist(subset(pd,i==2)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[2],col="red")
hist(subset(pd,i==3)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[3],col="red")
hist(subset(pd,i==4)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[4],col="red")
hist(subset(pd,i==5)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[5],col="red")
hist(subset(pd,i==6)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[6],col="red")
hist(subset(pd,i==7)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[7],col="red")
hist(subset(pd,i==8)$theta,xlab="",ylab="",main="")
```

```r
abline(v=(y/n)[8],col="red")
hist(subset(pd,i==9)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[9],col="red")
hist(subset(pd,i==10)$theta,xlab="",ylab="",main="")
abline(v=(y/n)[10],col="red")
```




```r
par(mfrow=c(1,1)) # Plotting posterior means vs. raw proportions
tj<-sapply(split(pd$theta,pd$i),mean)
plot(y/n,tj,ylab="mean(theta_j)",main="Problem 5.13(c) - Posterior Means vs. Raw Proportions (y/n)")
abline(0,1)
```



The inferences from the posterior distributions for the θ_j's are generally centered around the raw proportion $(\frac{y_j}{n_j})$. This is expected as $\theta_j|\alpha,\beta,y \sim Beta(\alpha+y_j,\beta+n_j-y_j)$ and it is apparent that 5,000 samples is sufficient. The center of the posterior distribution for θ_{10} appears to be considerably less (0.4228567) than the raw proportion $\frac{y_{10}}{n_{10}} = \frac{55}{55+64} = \frac{55}{119} \approx 0.4621849$, but we can see from the data that this particular location has a far greater proportion of bicycles than the other locations (its raw proportion is more than twice the next highest, $\theta_1 = \frac{y_1}{n_1} = \frac{16}{16+58} = \frac{16}{74} = \frac{8}{37} = 0.\overline{216}$) and could reasonably be considered an outlier location.

Problem 5.13(d)

```
quantile(alpha/(alpha+beta),c(0.05/2,1-0.05/2))
```

```
##      2.5%     97.5%
## 0.151536 0.265492
```

Problem 5.13(e)

```
set.seed(226) # Sampling 100 draws from beta (variance) and binomial distributions
quantile(rbinom(100,100,rbeta(100,alpha,beta)),c(0.05/2,1-0.05/2))
```

```
##   2.5%  97.5%
##  4.475 39.050
```

I do not *fully* trust this posterior interval in application because this new location may be correlated with other locations or not fully independent, or conversely could be much different than the original locations. We also do not know from the problem whether the number of bicycles at this location follows a binomial distribution, but that is our best guess. The interval is rather conservative (wide), but additional information on the location may be needed to use it in practice.

Problem 5.14

Problem 5.14(b)

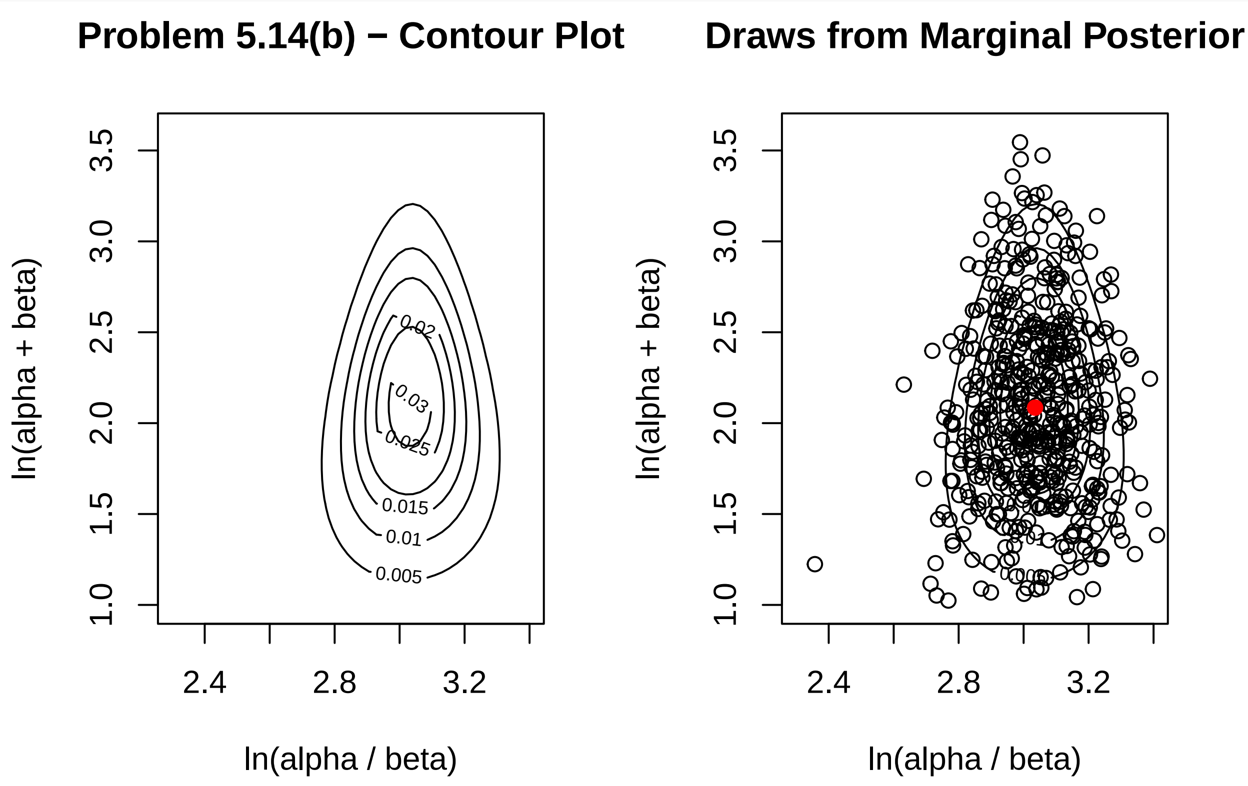
```
mean(y/n)              # mean = alpha / beta
```

```
## [1] 0.1961412
```

```
var(y/n)           # variance = alpha / beta^2
```

```
## [1] 0.01112067
```

```
b<-mean(y/n)/var(y/n) # beta = mean / variance
a<-b*mean(y/n)        # alpha = beta * mean
mpd<-function(X,Y){a<-exp(Y)*exp(X)/(exp(X)+1)
  b<-exp(Y)/(exp(X)+1)
  10^-200*sqrt(a)*(b+1)*prod(b^a/(b+1)^(y+a)*gamma(y)/beta(a,y))*exp(2*Y+X)/(exp(X)+1)^2}
X<-seq(2.3,3.4,length=50) # Setting bounds of contour plot via trial-and-error
Y<-seq(1,3.6,length=50)
z<-matrix(NA,length(Y),length(X))
for (i in 1:length(Y)){for (j in 1:length(X)){z[j,i]<-mpd(X[j],Y[i])}}
par(mfrow=c(1,2))
contour(X,Y,z,xlab="ln(alpha / beta)",ylab="ln(alpha + beta)",main="Problem 5.14(b) - Contour Plot")
set.seed(226)
for (k in 1:length(Xn)){Xn[k]<-sample(X,1,TRUE,apply(z,1,sum)/sum(apply(z,1,sum)))
  Yn[k]<-sample(Y,1,TRUE,z[which(Xn[k]==X),]/sum(z[which(Xn[k]==X),]))}
contour(X,Y,z,xlab="ln(alpha / beta)",ylab="ln(alpha + beta)",main="Draws from Marginal Posterior")
Xn<-Xn+runif(length(Xn),-diff(X)/2,diff(X)/2) # Adding random jitter
Yn<-Yn+runif(length(Yn),-diff(Y)/2,diff(Y)/2)
points(Xn,Yn) # One extreme outlier at (~2.3,~1.5)
points(mean(Xn),mean(Yn),pch=19,col="red") # Mean of samples
```



Problem 5.14(c)

Yes, the posterior density $(p(\theta, \alpha, \beta|y) \propto p(\alpha, \beta) \prod_{j=1}^{J} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta_j^{y_j+\alpha-1} e^{-(\beta+1)\theta_j})$ appears to be integrable. We can see from the plots in problems 5.14(e) and 5.14(b) that the joint posterior and contours are finite and do not diverge. Any function can be integrated if you try hard enough and have enough computing power.

Problem 5.14(e)

```
alpha<-exp(Yn)*(exp(Xn)/(exp(Xn)+1))
beta<-exp(Yn)/(exp(Xn)+1)
plot(alpha,beta,main="Problem 5.14(e) - 500 Draws from Joint Posterior Distribution")
abline(0,0.05) # Unknown where the slope comes from?
```