Bayesian Methods - Assignment 3

Ryan Durfey

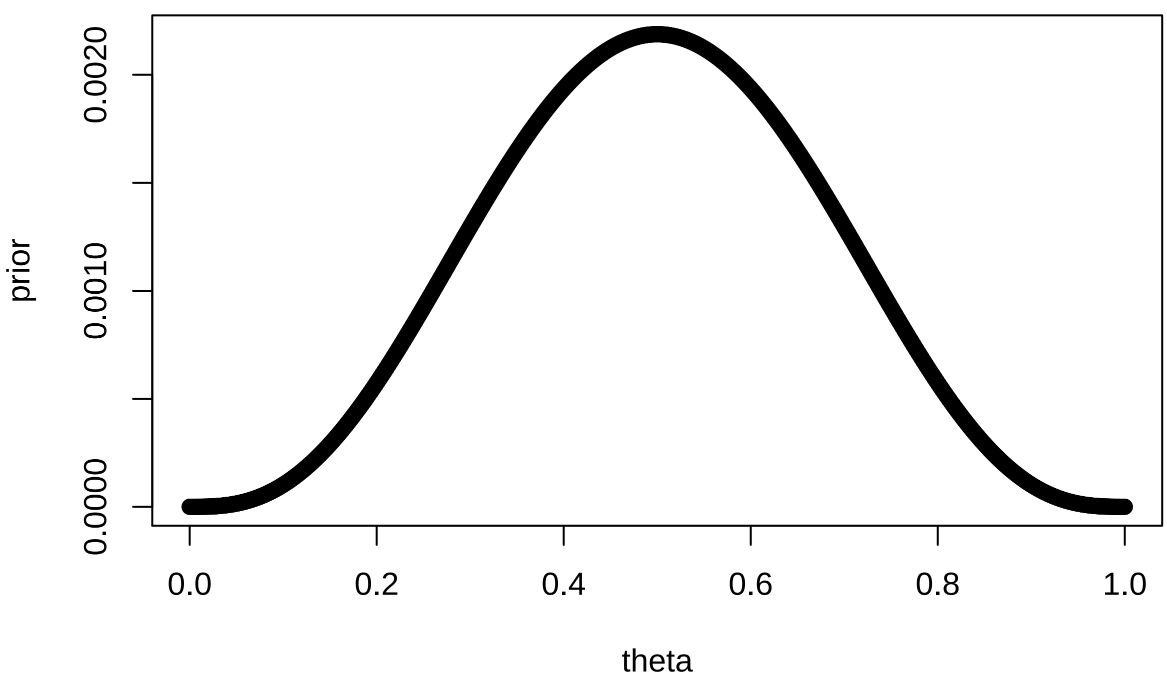
April 24, 2016

```
source('/Users/rdurfey/R_misc/BayesianMethods/DBDA2Eprograms/BernBeta.R')
opar<-par(no.readonly = TRUE)

# Likelihood function for calculating binomial likelihood as in the previous workshop.
likeli<-function(par,data){
  sdata<-sum(data)
  ldata<-length(data)
  return(par^sdata*(1-par)^(ldata-sdata))
}
```

1. Start with Prior Distirbution Beta(4,4). Plot it and interpret the shape

```
# initial parameters for prior
alpha.init <- beta.init <- 4 # initial shape parameters given
theta <- seq(0,1,length=1001) # initial grid of thetas (fine teeth)
prior <- dbeta(theta,alpha.init,beta.init) # Beta prior distribution
prior <- prior/sum(prior) # make prior sum to 1
plot(theta,prior)
```



The curve looks fairly similar to a normal bell-curve, but probably a bit wider. From Beta formulas, the mode = (4-1)/(4+4-2) = 0.5, and concentration = 4+4 = 8. With such a small concentration, I expect this initial Prior to be fairly weak, meaning that the Posterior will be influenced more by the Likelihood.

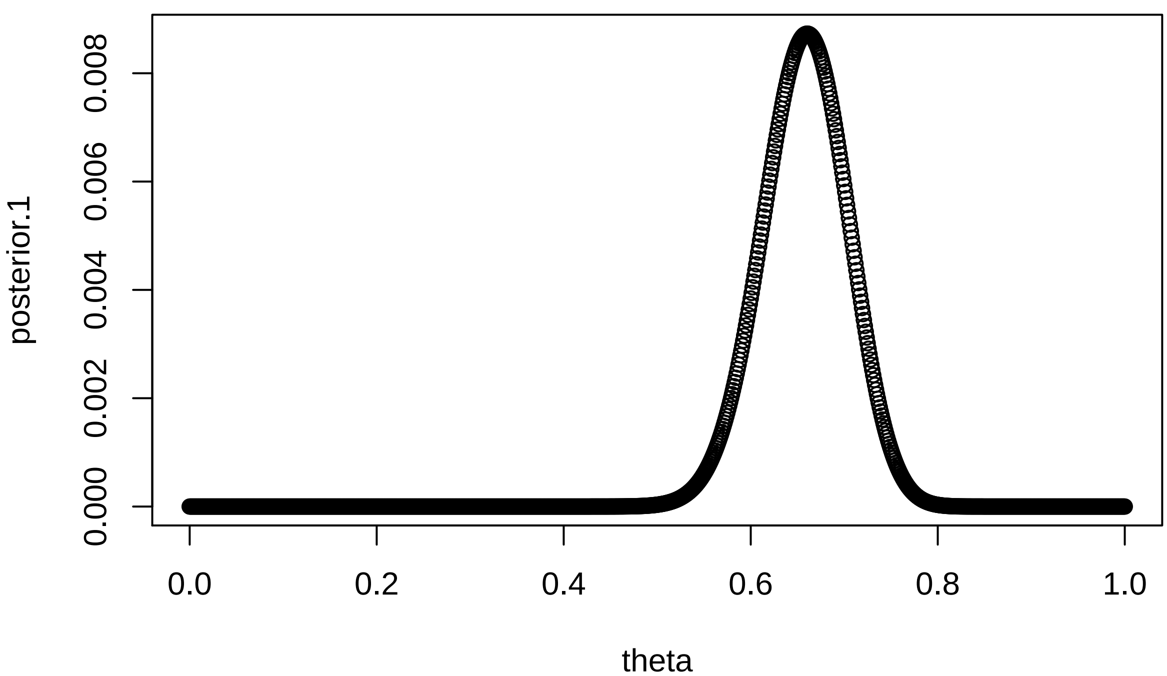
2. Select seed set.seed(13) and simulate 1 Bernoulli variable d1 with probability 0.62.

```
set.seed(13)
d1 <- rbinom(n=100,size=1,prob=0.62) # we'll use 100 simulated observations
N <- length(d1) # number of observed trials
z.1 <- sum(d1) # number of successes
z.1/N # yep, comes out to be about 0.62, which is the prob we set
```

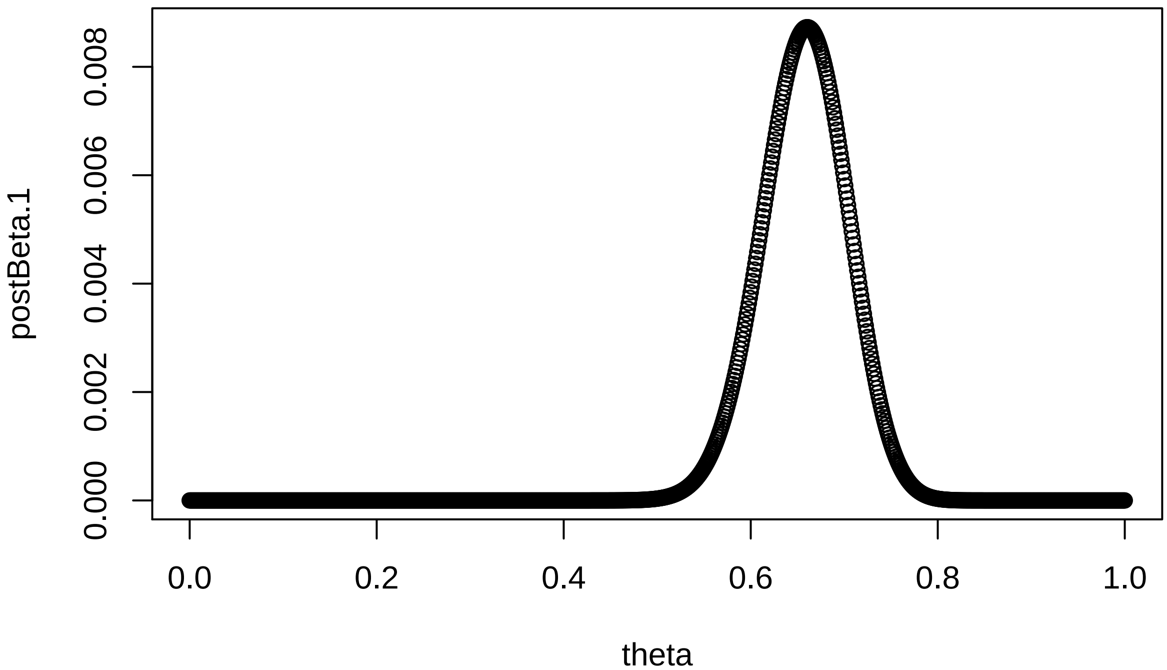
```
## [1] 0.67
```

3. Find posterior distribution by formula. Plot it. Calculate mode and mean value. Predict next observation.

```
# Posterior from formula:  prior * likelihood
posterior.1 <- prior * likeli(theta,d1)
posterior.1 <- posterior.1/sum(posterior.1)
plot(theta,posterior.1)
```

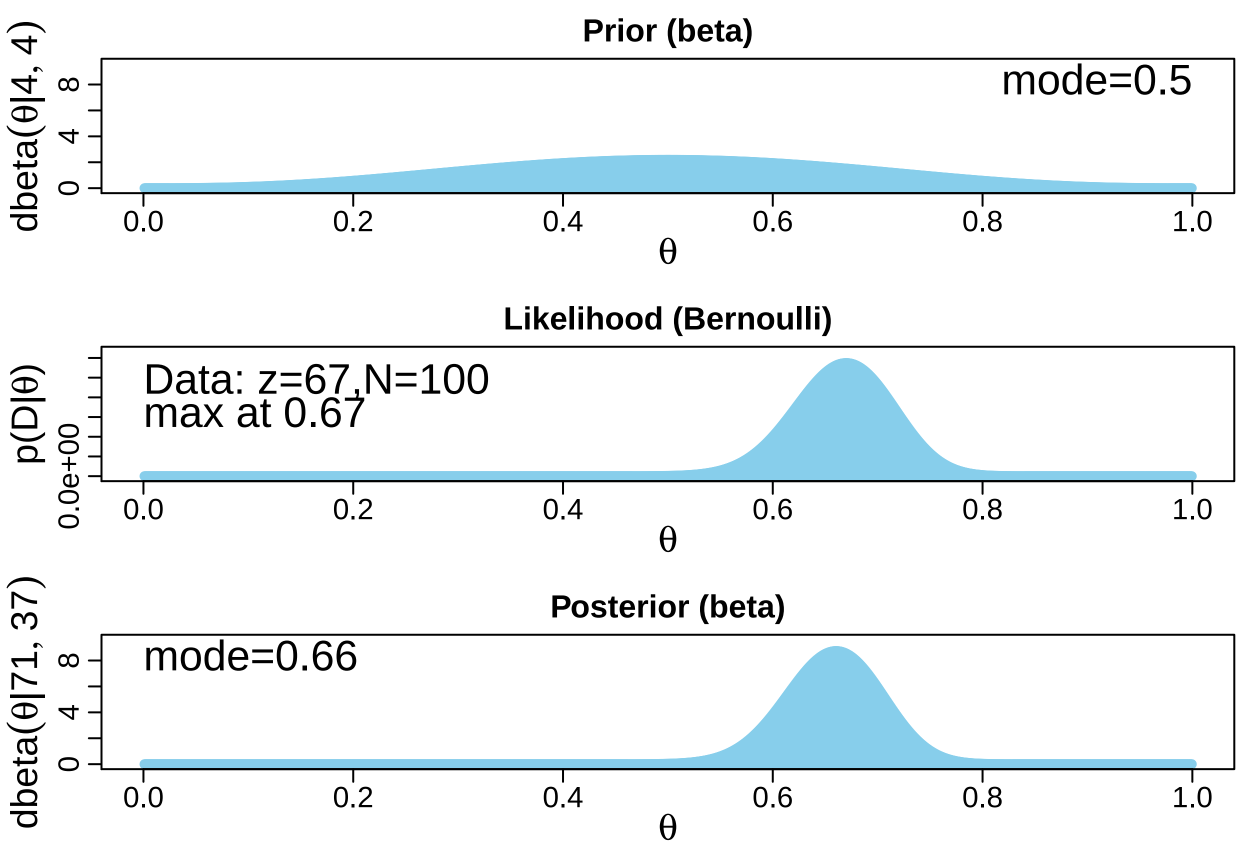



```
# Posterior from Beta shape parameters formula
postBeta.1<-dbeta(theta,shape1=z.1+alpha.init,shape2=N-z.1+beta.init)
postBeta.1<-postBeta.1/sum(postBeta.1)
plot(theta,postBeta.1)
```




```r
# NOTE: posteriors calculated from both methods are the same

# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.init,beta.init), Data=d1, plotType="Bars",
                       showCentTend="Mode" , showHDI=F , showpD=FALSE )
```



```r
par(opar)


# mode & mean
(mode.1<-theta[which.max(posterior.1)])
```

```
## [1] 0.66
```

```r
(mean.1<-theta%*%posterior.1)
```

```
##           [,1]
## [1,] 0.6574074
```

```r
# next observation prediction
alpha.post.1 <- alpha.init + z.1
beta.post.1 <- N - z.1 + beta.init
(mean.posterior.1 <- alpha.post.1/(alpha.post.1+beta.post.1))
```

```
## [1] 0.6574074
```

```r
round(mean.posterior.1)
```

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

As expected, the Posterior here is heavily influenced by the Data. We can see the distribution shifting a lot and looks like it could be nearly identical to the Likelihood in the BernBeta plot.

The prediction for the next observation is a probability of 0.657, which would translate into a 1 for binary prediction purposes.

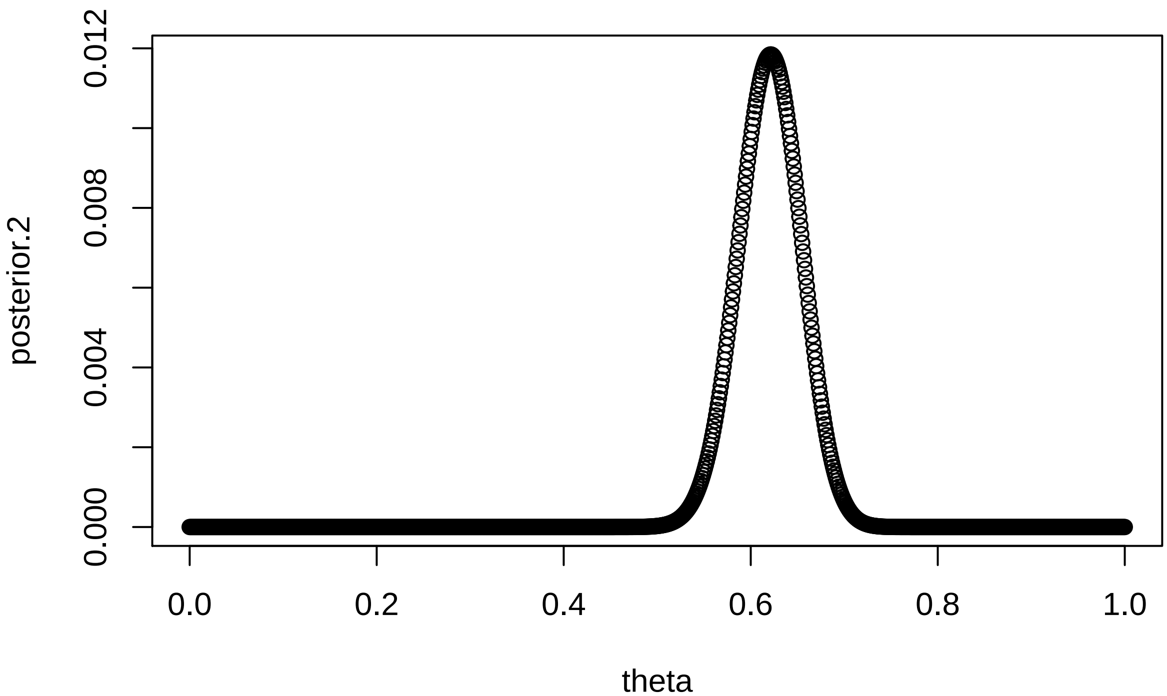
4. Select seed set.seed(917). Simulate second Bernoulli variable d2 with probability 0.62.

```r
set.seed(917)
d2 <- rbinom(n=100,size=1,prob=0.62) # we'll use 100 simulated observations
N <- length(d2)
z.2 <- sum(d2)
z.2/N
```

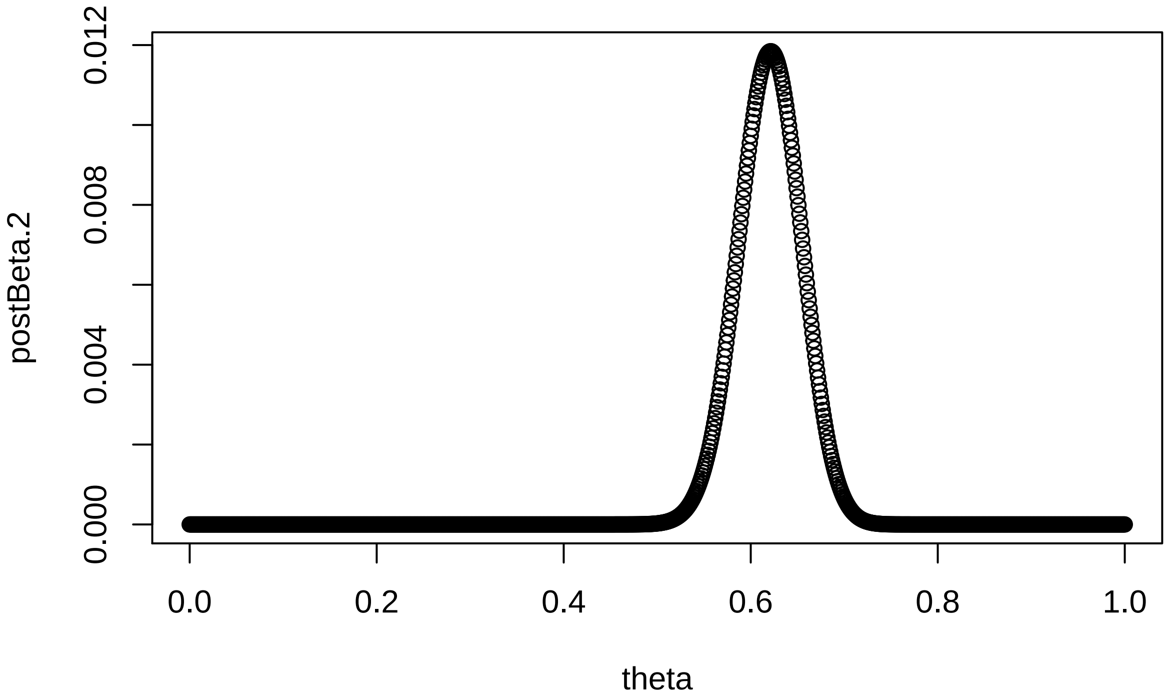
```
## [1] 0.58
```

5. Use parameters of the first posterior to define new prior. Find second posterior. Plot it. Calculate mode and mean value. Predict next observation.

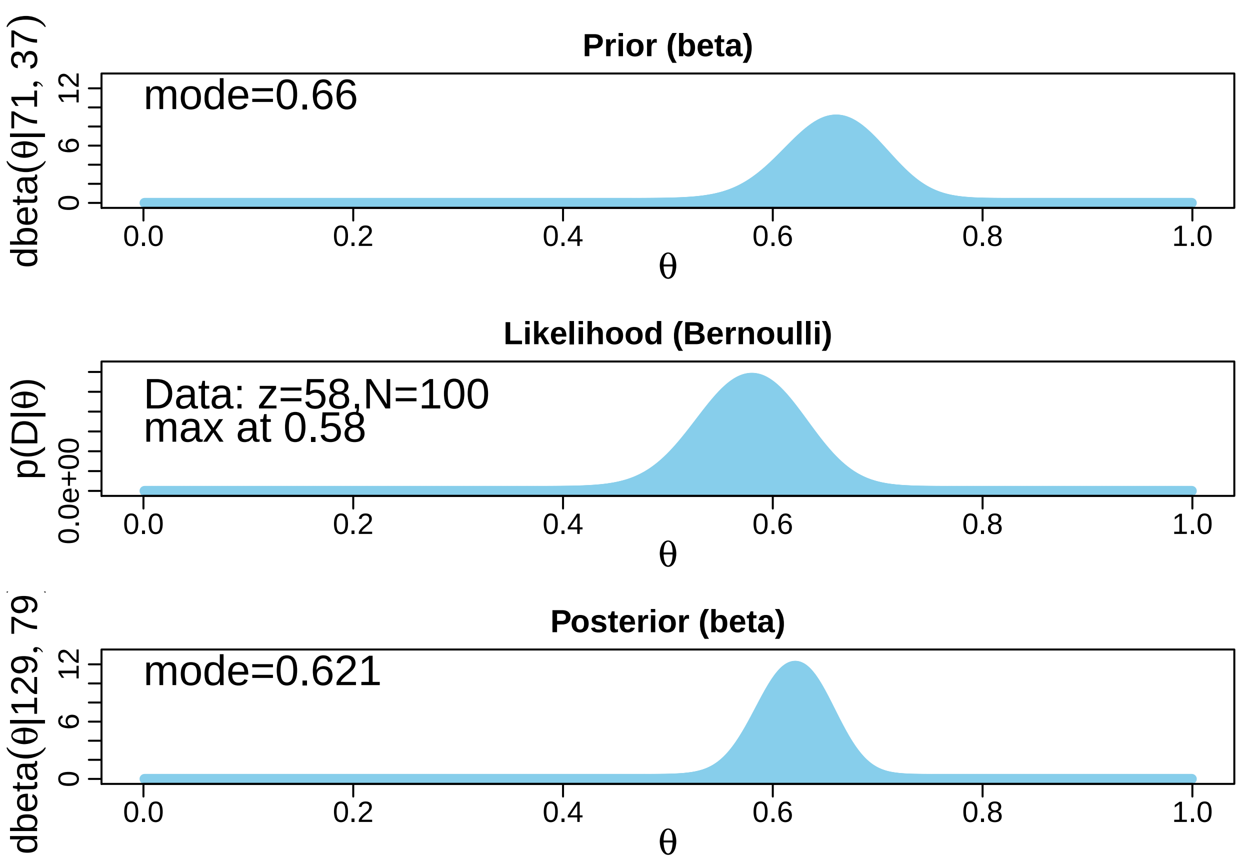
```r
# Posterior from formula:  prior * likelihood
posterior.2 <- posterior.1 * likeli(theta,d2)
posterior.2 <- posterior.2/sum(posterior.2)
plot(theta,posterior.2)
```




```r
# Posterior from Beta shape parameters formula
postBeta.2<-dbeta(theta,shape1=z.2+alpha.post.1,shape2=N-z.2+beta.post.1)
postBeta.2<-postBeta.2/sum(postBeta.2)
plot(theta,postBeta.2)
```




```r
# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.post.1,beta.post.1), Data=d2, plotType="Bars",
                       showCentTend="Mode" , showHDI=F , showpD=FALSE )
```



```r
par(opar)

# mode & mean
(mode.2<-theta[which.max(posterior.2)])
```

```
## [1] 0.621
```

```r
(mean.2<-theta%*%posterior.2/sum(posterior.2))
```

```
##           [,1]
## [1,] 0.6201923
```

```r
(theta%*%postBeta.2/sum(postBeta.2))
```

```
##           [,1]
## [1,] 0.6201923
```

```r
# next observation prediction
alpha.post.2 <- alpha.post.1 + z.2
beta.post.2 <- N - z.2 + beta.post.1
(mean.posterior.2 <- alpha.post.2/(alpha.post.2+beta.post.2))
```

```
## [1] 0.6201923
```

```r
round(mean.posterior.2)
```

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

With the previous Posterior being used as the new Prior here, it has a lot more influence on the new/updated Posterior. In the BernBeta plot, we see it that it is influenced by the Likelihood somewhat, but not nearly as much as it was earlier.

The prediction for the next observation has a probability of 0.620, which would also translate into a 1 for binary prediction purposes.

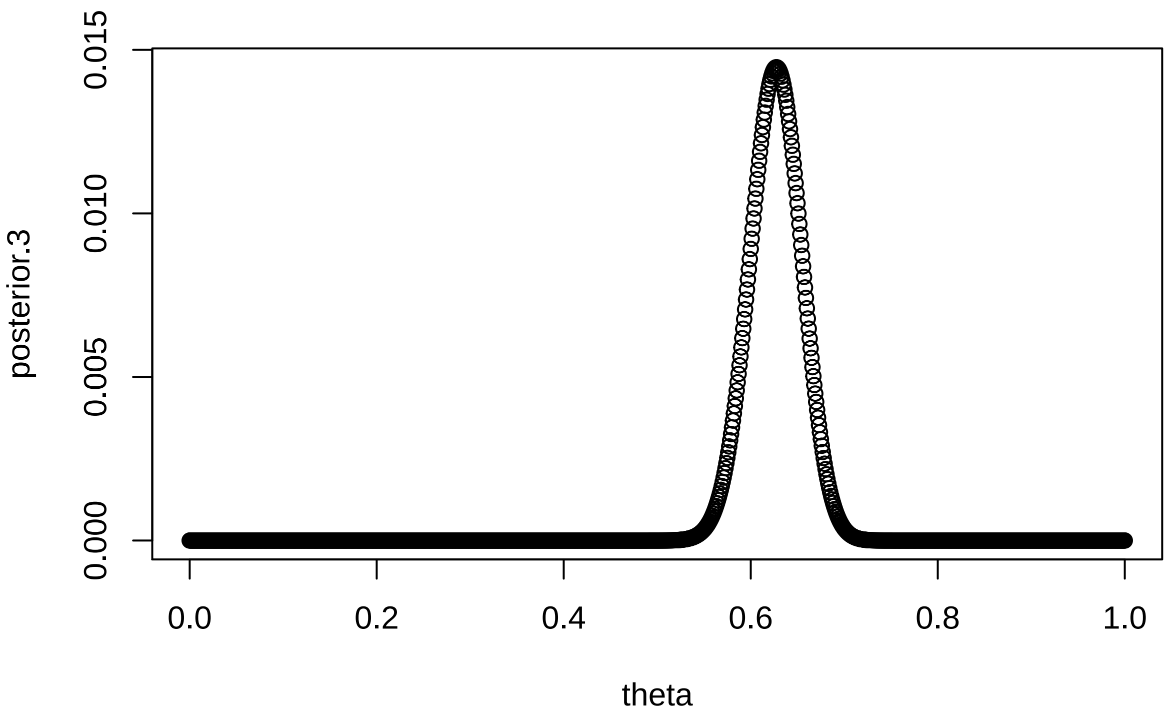
6. Select seed set.seed(1000). Simulate tird Bernoulli variable d3 with probability 0.62.

```r
set.seed(1000)
d3 <- rbinom(n=100,size=1,prob=0.62) # we'll use 100 simulated observations
N <- length(d3)
z.3 <- sum(d3)
z.3/N
```

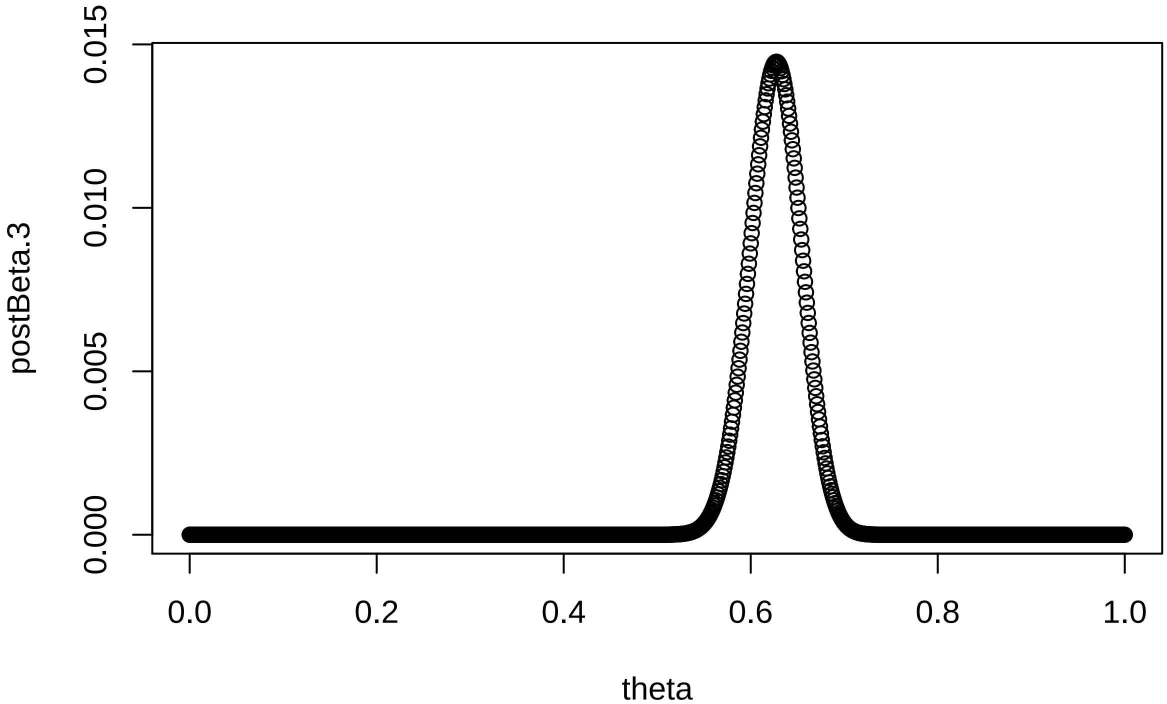
```
## [1] 0.64
```

7. Use parameters of the second posterior to define new prior. Find third posterior. Plot it. Calculate mode and mean value. Predict next observation.

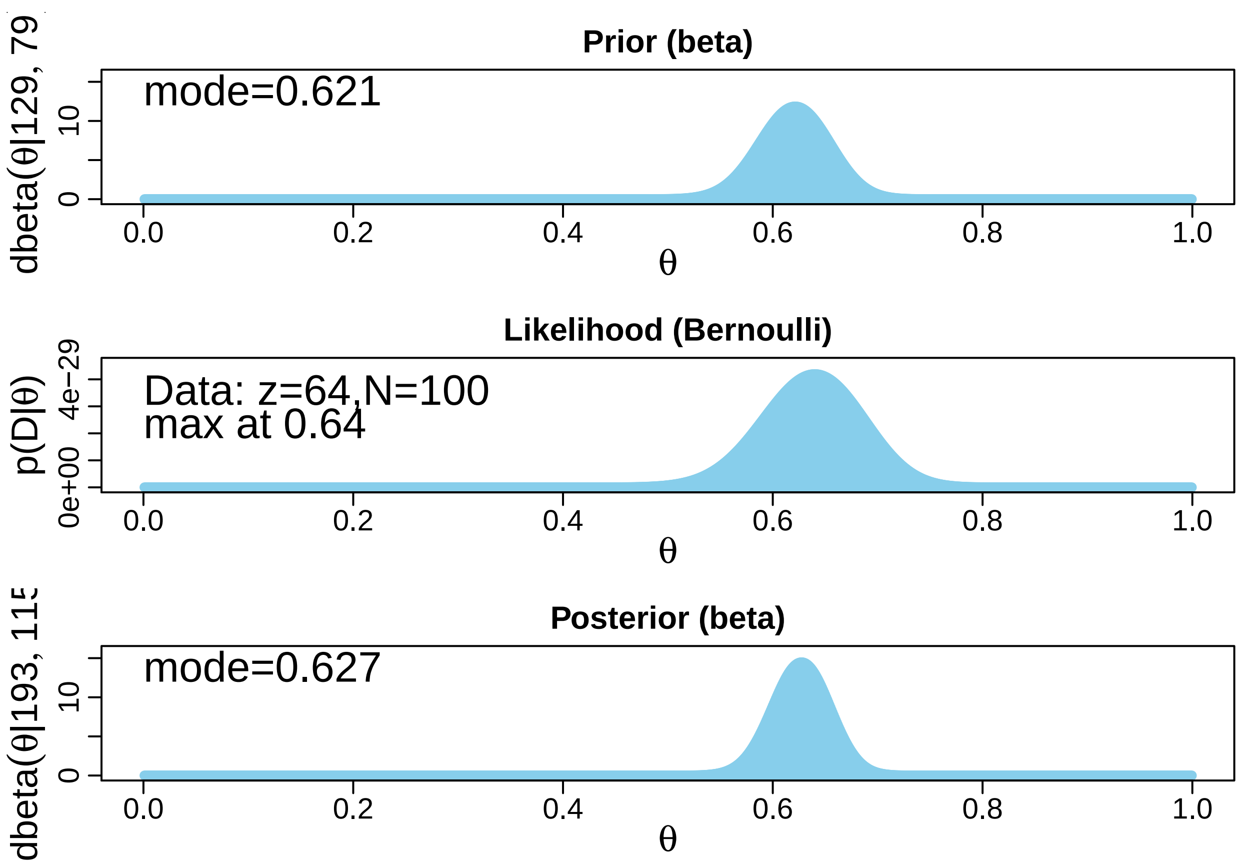
```
# Posterior from formula:  prior * likelihood
posterior.3 <- posterior.2 * likeli(theta,d3)
posterior.3 <- posterior.3/sum(posterior.3)
plot(theta,posterior.3)
```




```
# Posterior from Beta shape parameters formula
postBeta.3<-dbeta(theta,shape1=z.3+alpha.post.2, shape2=N-z.3+beta.post.2)
postBeta.3<-postBeta.3/sum(postBeta.3)
plot(theta,postBeta.3)
```




```
# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.post.2,beta.post.2), Data=d3, plotType="Bars",
                       showCentTend="Mode" , showHDI=F , showpD=FALSE )
```



```r
par(opar)

# mode & mean
(mode.3<-theta[which.max(posterior.3)])
```

```
## [1] 0.627
```

```r
(mean.3<-theta%*%posterior.3/sum(posterior.3))
```

```
##           [,1]
## [1,] 0.6266234
```

```r
(theta%*%postBeta.3/sum(postBeta.3))
```

```
##           [,1]
## [1,] 0.6266234
```

```r
# next observation prediction
alpha.post.3 <- alpha.post.2 + z.3
beta.post.3 <- N - z.3 + beta.post.2
(mean.posterior.3 <- alpha.post.3/(alpha.post.3+beta.post.3))
```

```
## [1] 0.6266234
```

```r
round(mean.posterior.3)
```

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

After this third iteration, we see the Prior influences the Posterior even more. This is because as we add more data, the concentration of the Beta Prior distribution gets stonger, making each new data point influence the Posterior less and less.

The prediction for the next observation has a probability of 0.627, which would translate into a 1 for binary prediction purposes.

```r
# Comparison of Modes and Means/Predictions of Posteriors
data.frame(Posterior_1=rbind(mode.1,mean.1),Posterior_2=rbind(mode.2,mean.2),Posterior_3=rbind(mode.3,m
```

```
##                 Posterior_1 Posterior_2 Posterior_3
## Mode              0.6600000   0.6210000   0.6270000
## Mean/Prediction   0.6574074   0.6201923   0.6266234
```

8. Start from the initial prior Beta(4,4) and repeat all calculations of posteriors using the sequence of data in reverse order: d3, d2, d1. Compare the models predictions and graphs of the distributions after observation received.

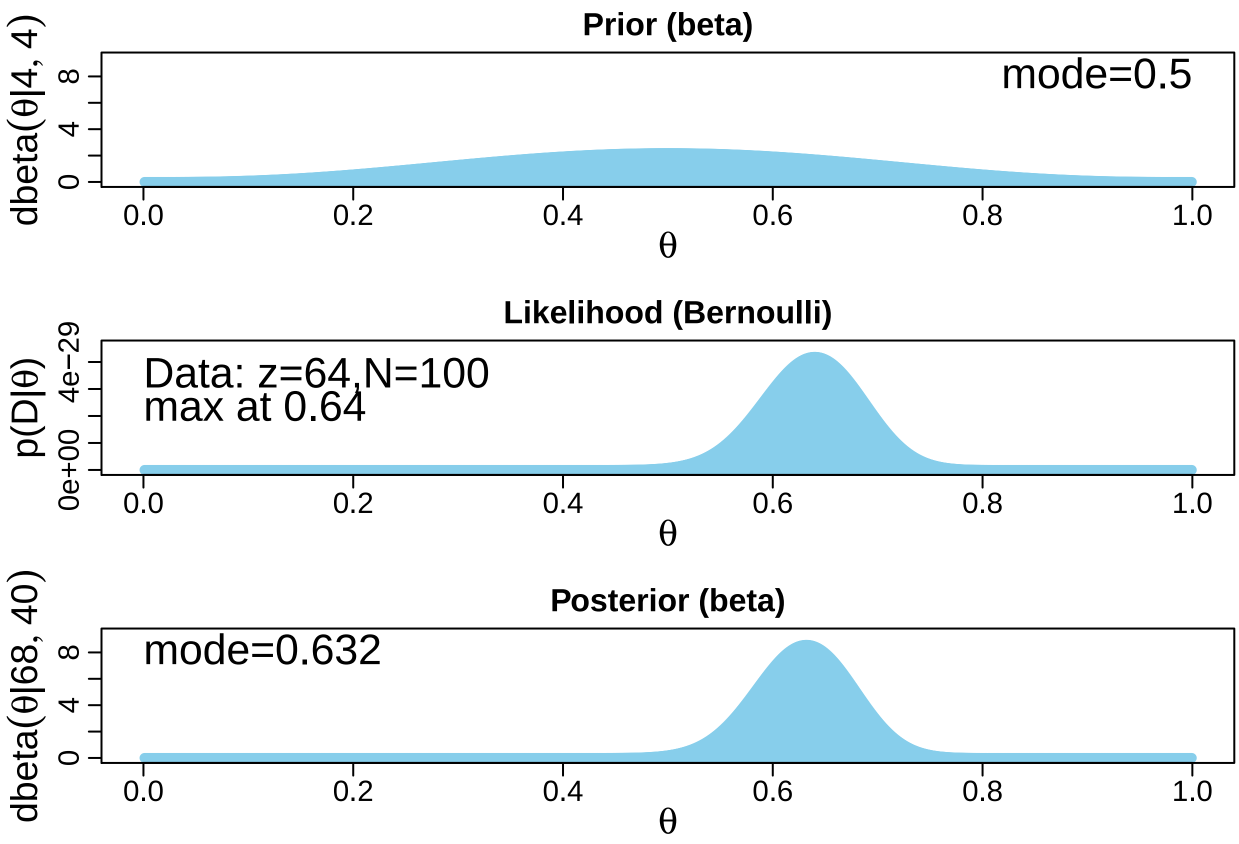
Note: To cut down on superfluous plots (and save space in the pdf output), only the BernBeta() plots are shown below.

```r
#### POSTERIOR 1 with Data D3 ####

# Posterior from formula:  prior * likelihood
posterior.1 <- prior * likeli(theta,d3)
posterior.1 <- posterior.1/sum(posterior.1)
# plot(theta,posterior.1)

# Posterior from Beta shape parameters formula
postBeta.1<-dbeta(theta,shape1=z.1+alpha.init,shape2=N-z.1+beta.init)
postBeta.1<-postBeta.1/sum(postBeta.1)
# plot(theta,postBeta.1)
# NOTE: posteriors calculated from both methods are the same

# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.init,beta.init), Data=d3, plotType="Bars",
                       showCentTend="Mode" , showHDI=F , showpD=FALSE )
```




```r
par(opar)


# mode & mean
(mode.1<-theta[which.max(posterior.1)])
```

```
## [1] 0.632
```

```r
(mean.1<-theta%*%posterior.1)
```

```
##            [,1]
## [1,] 0.6296296
```

```r
# next observation prediction
alpha.post.1 <- alpha.init + z.1
beta.post.1 <- N - z.1 + beta.init
(mean.posterior.1 <- alpha.post.1/(alpha.post.1+beta.post.1))
```

```
## [1] 0.6574074
```

```r
round(mean.posterior.1)
```

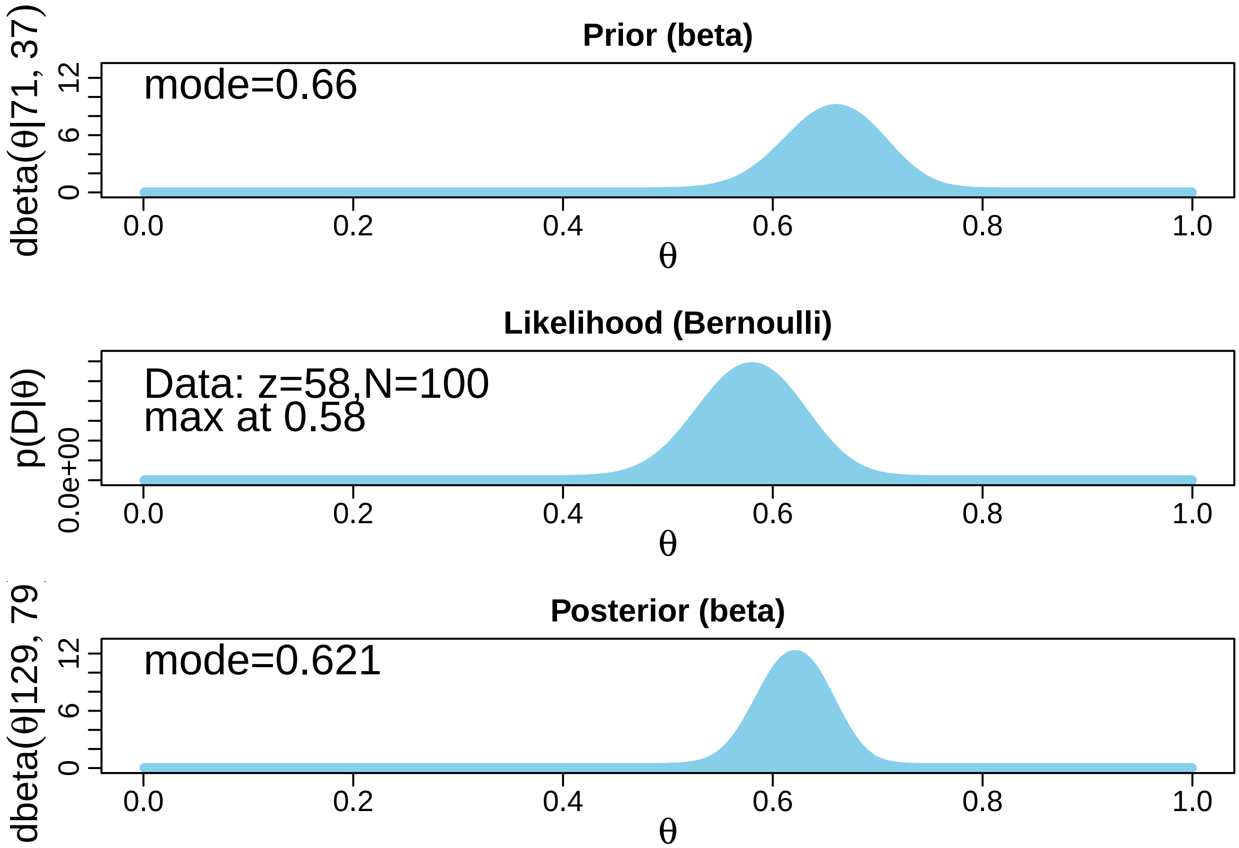
```
## [1] 1
```

```r
#### POSTERIOR 2 with Data D2 ####

# Posterior from formula:  prior * likelihood
posterior.2 <- posterior.1 * likeli(theta,d2)
posterior.2 <- posterior.2/sum(posterior.2)
# plot(theta,posterior.2)

# Posterior from Beta shape parameters formula
postBeta.2<-dbeta(theta,shape1=z.2+alpha.post.1,shape2=N-z.2+beta.post.1)
postBeta.2<-postBeta.2/sum(postBeta.2)
# plot(theta,postBeta.2)

# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.post.1,beta.post.1), Data=d2, plotType="Bars",
                        showCentTend="Mode" , showHDI=F , showpD=FALSE )
```




```
par(opar)

# mode & mean
(mode.2<-theta[which.max(posterior.2)])
```

```
## [1] 0.607
```

```
(mean.2<-theta%*%posterior.2/sum(posterior.2))
```

```
##           [,1]
## [1,] 0.6057692
```

```
(theta%*%postBeta.2/sum(postBeta.2))
```

```
##           [,1]
## [1,] 0.6201923
```

```
# next observation prediction
alpha.post.2 <- alpha.post.1 + z.2
beta.post.2 <- N - z.2 + beta.post.1
(mean.posterior.2 <- alpha.post.2/(alpha.post.2+beta.post.2))
```

```
## [1] 0.6201923
```

```
round(mean.posterior.2)
```

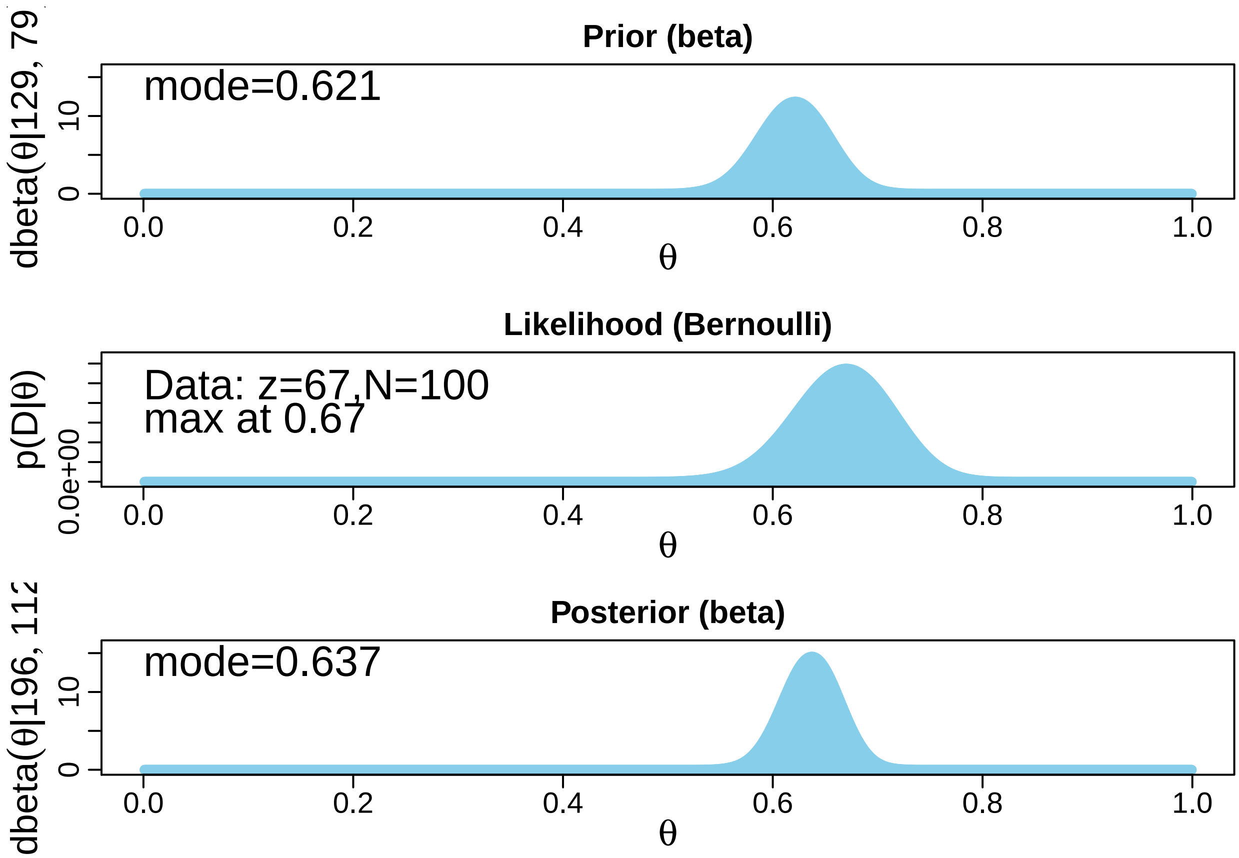
```
## [1] 1
```

```
#### POSTERIOR 3 with Data D1 ####

# Posterior from formula:  prior * likelihood
posterior.3 <- posterior.2 * likeli(theta,d1)
posterior.3 <- posterior.3/sum(posterior.3)
# plot(theta,posterior.3)

# Posterior from Beta shape parameters formula
postBeta.3<-dbeta(theta,shape1=z.3+alpha.post.2, shape2=N-z.3+beta.post.2)
postBeta.3<-postBeta.3/sum(postBeta.3)
# plot(theta,postBeta.3)

# double-check using BernBeta function
post.check <- BernBeta(priorBetaAB=c(alpha.post.2,beta.post.2), Data=d1, plotType="Bars",
                       showCentTend="Mode" , showHDI=F , showpD=FALSE )
```




```
par(opar)

# mode & mean
(mode.3<-theta[which.max(posterior.3)])
```

```
## [1] 0.627
```

```
(mean.3<-theta%*%posterior.3/sum(posterior.3))
```

```
##           [,1]
## [1,] 0.6266234
```

```
(theta%*%postBeta.3/sum(postBeta.3))
```

```
##           [,1]
## [1,] 0.6266234
```

```
# next observation prediction
alpha.post.3 <- alpha.post.2 + z.3
beta.post.3 <- N - z.3 + beta.post.2
(mean.posterior.3 <- alpha.post.3/(alpha.post.3+beta.post.3))
```

```
## [1] 0.6266234
```

```
round(mean.posterior.3)
```

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

Here, too, the predictions are all between 0.6 and 0.66. As such, a binary prediction in all cases would be 1.

Conclusions:

The above plots are similar to the ones observed in the previous part of the assignment. In the first Posterior, it is most influenced by the Likelihood. Then, in the second and third Posteriors, the Prior has an increasing amount of sway.

However, the first two Posteriors are not identical to those in the earlier part of the assignment. The differences are easily exhibited in the Modes and Means of each subsequent Posterior Distribution. Below is a table comparing the Modes and Means from this part. Notice that the values from Posterior 1 and 2 are different here than they were in the first part. But, Posterior 3 (the final Posterior) is identical to that observed earlier. This is because at this point in the analyses, all of the data has been included. This further shows that it does not matter in which order data is received.

```
# Comparison of Modes and Means/Predictions of Posteriors
data.frame(Posterior_1=rbind(mode.1,mean.1),Posterior_2=rbind(mode.2,mean.2),Posterior_3=rbind(mode.3,me
```

```
##                 Posterior_1 Posterior_2 Posterior_3
## Mode              0.6320000   0.6070000   0.6270000
## Mean/Prediction   0.6296296   0.6057692   0.6266234
```