MA5750: APPLIED STATISTICS
ASSIGNMENT 3
E Naveen
ME16B077

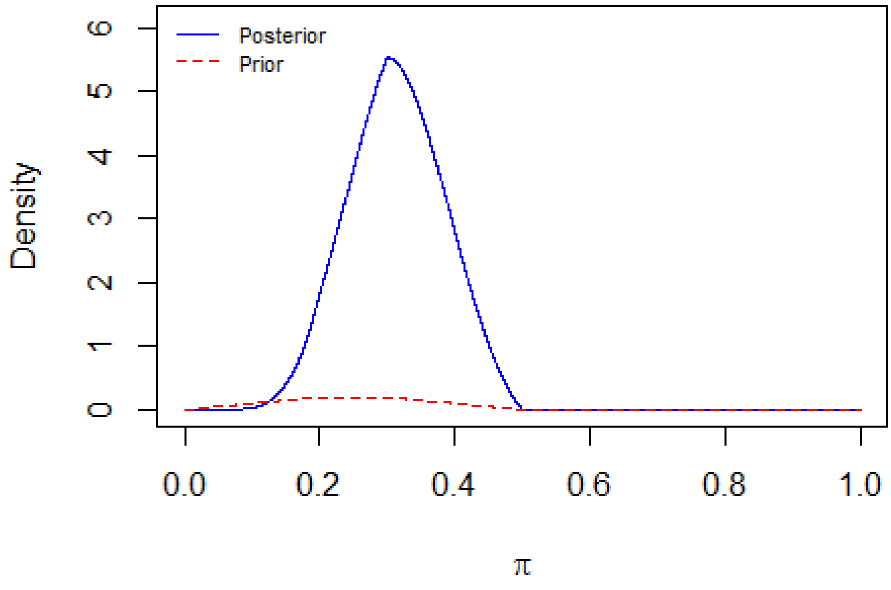
8.4 We will use the Minitab macro BinoGCPmac or the equivalent R function to find the posterior distribution of the binomial probability π when the observation distribution of Y|π is binomial(n,π) and we have a general continuous prior for π.

Suppose out of n = 20 independent trials, y = 7 successes were observed.

(a) Use BinoGCPmac or the equivalent R function to determine the posterior distribution g(π|y).Details for invoking BinoGCPmac and the equivalent R function are in Appendix 3 and Appendix 4, respectively.

R Code:

```
# Load the dependent libraries
library(Bolstad)

##
## Attaching package: 'Bolstad'

## The following objects are masked from 'package:stats':
##
##     IQR, sd, var

# (a)
pi_values = seq(from = 0, to = 1, by=0.0001)
pi_prior = rep(0,length(pi_values))
pi_prior[pi_values<=0.2] = pi_values[pi_values<=0.2]
pi_prior[pi_values>0.2 & pi_values<=0.3] = 0.2
pi_prior[pi_values>0.3 & pi_values<=0.5] = 0.5 - pi_values[pi_values>0.3 & pi_values<=
0.5]
pi_prior[pi_values>0.5] = 0
output_1 = binogcp(x = 7, n = 20, density="user", pi = pi_values, pi.prior = pi_prior)
```



(b) **Use tintegral.mac ind the posterior mean and posterior standard deviation of π. Details for invoking tintegral.mac and the equivalent R function are in Appendix 3 and Appendix 4, respectively.**

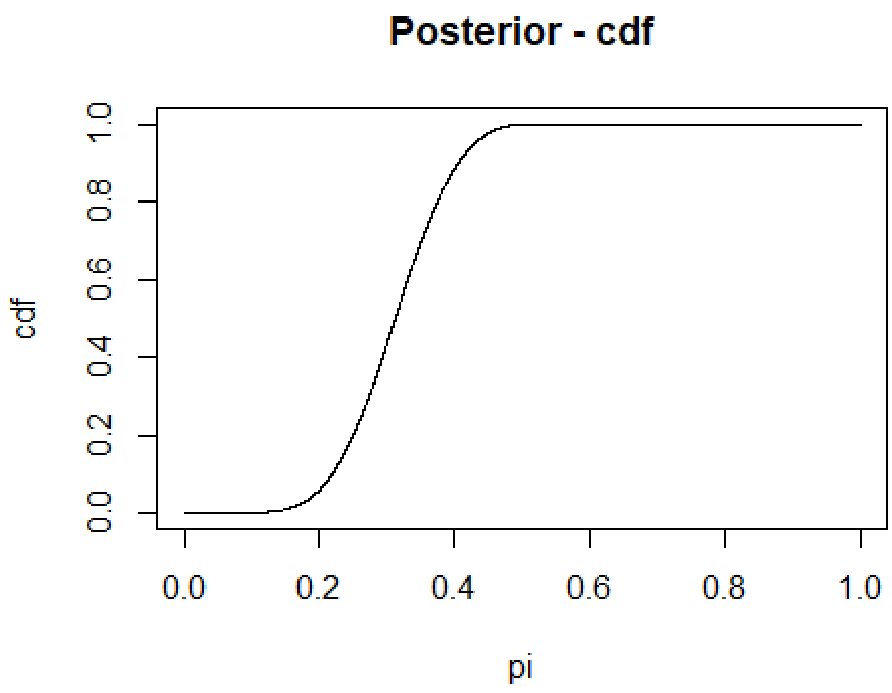
R Code:

```
# (b)

# Finding the posterior mean and standard deviation of pi using integration..
E_posterior = sintegral(pi_values,pi_values*output_1$posterior, n.pts=length(pi_values
))
mean_posterior = E_posterior$value
cat("Posterior mean of pi =", mean_posterior, '\n')

## Posterior mean of pi = 0.3120479

Variance_posterior = sintegral(pi_values,(pi_values-mean_posterior)^2*output_1$posteri
or, n.pts=length(pi_values))
stddev_posterior = Variance_posterior$value^0.5
cat("Posterior standard deviation of pi =", stddev_posterior, '\n')

## Posterior standard deviation of pi = 0.07070219
```

(c) **Find a 95% credible interval for π by using tintegral.mac or the equivalent R function.**

R Code:

```
# (c)
# Finding Lower and Upper bound for 95% Credible Interval by calculating cdf.
cdf_1 = sintegral(pi_values,output_1$posterior, n.pts=length(pi_values))
plot(cdf_1$cdf, type='l', xlab = 'pi', ylab = 'cdf', main = 'Posterior - cdf')
```



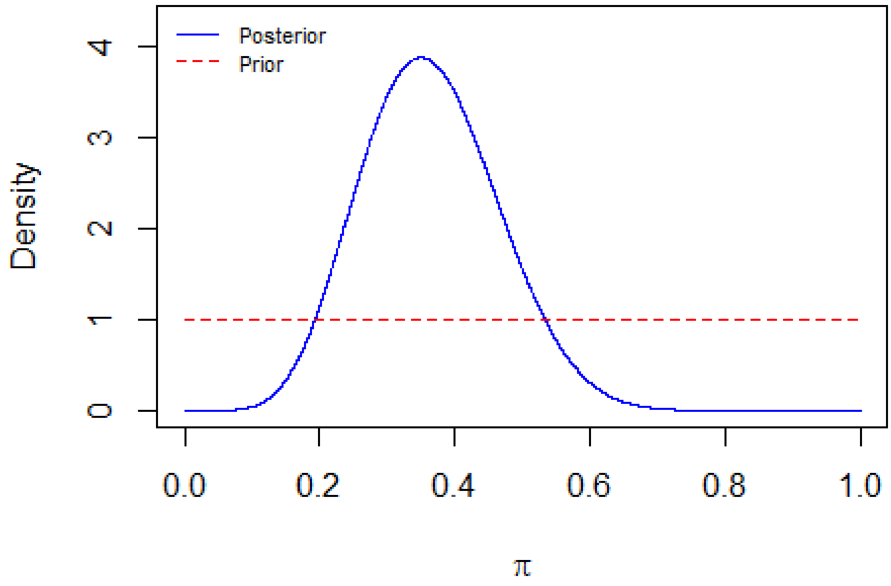

```
cdf = cdf_1$cdf
lower_bound = cdf$x[with(cdf,which.max(x[y<=0.025]))]
upper_bound = cdf$x[with(cdf,which.max(x[y<=0.975]))]
cat(paste("Approximate 95% credible interval: [", round(lower_bound, 4),
          " ", round(upper_bound, 4), "]\n", sep = ""))

## Approximate 95% credible interval: [0.175 0.4473]
```

It should be noted that the 95% credible interval is subject to slight changes depending upon the factor of discretization that we chose to split the interval [0,1] into.

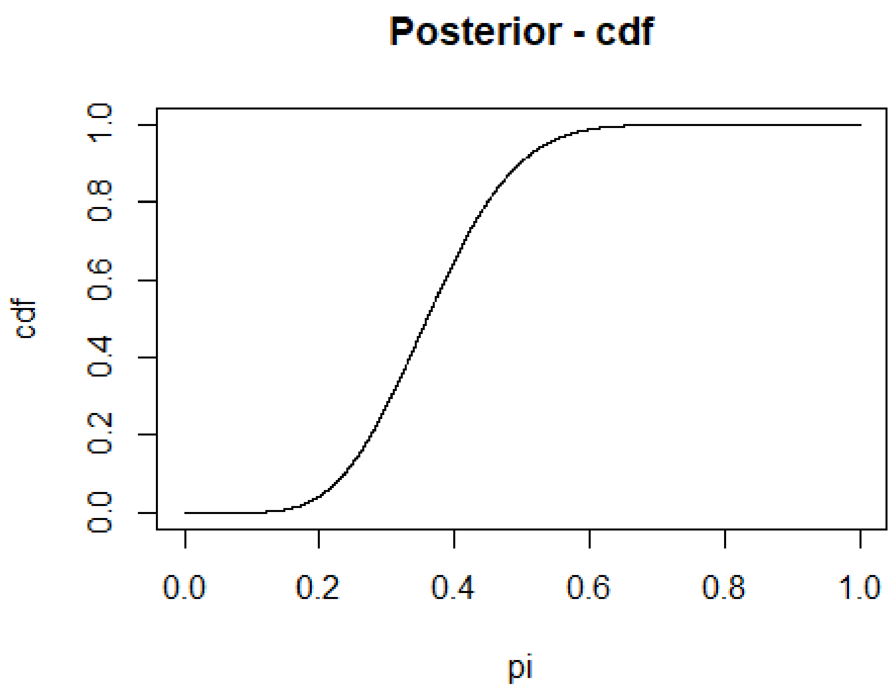
8.5 Repeat the previous question with a uniform prior for π.

R Code:

```
# Load the dependent libraries
library(Bolstad)

##
## Attaching package: 'Bolstad'

## The following objects are masked from 'package:stats':
##
##     IQR, sd, var

# (a)
pi_values = seq(from = 0, to = 1, by=0.0001)
pi_prior = rep(1,length(pi_values))
output_1 = binogcp(x = 7, n = 20, density="user", pi = pi_values, pi.prior = pi_prior)
```




```
# (b)

# Finding the posterior mean and standard deviation of pi using integration..
E_posterior = sintegral(pi_values,pi_values*output_1$posterior, n.pts=length(pi_values
))
mean_posterior = E_posterior$value
cat("Posterior mean of pi =", mean_posterior, '\n')

## Posterior mean of pi = 0.3636364

Variance_posterior = sintegral(pi_values,(pi_values-mean_posterior)^2*output_1$posteri
or, n.pts=length(pi_values))
stddev_posterior = Variance_posterior$value^0.5
cat("Posterior standard deviation of pi =", stddev_posterior, '\n')

## Posterior standard deviation of pi = 0.100305

# (c)
# Finding Lower and Upper bound for 95% Credible Interval by calculating cdf.
cdf_1 = sintegral(pi_values,output_1$posterior, n.pts=length(pi_values))
plot(cdf_1$cdf, type='l', xlab = 'pi', ylab = 'cdf', main = 'Posterior - cdf')
```



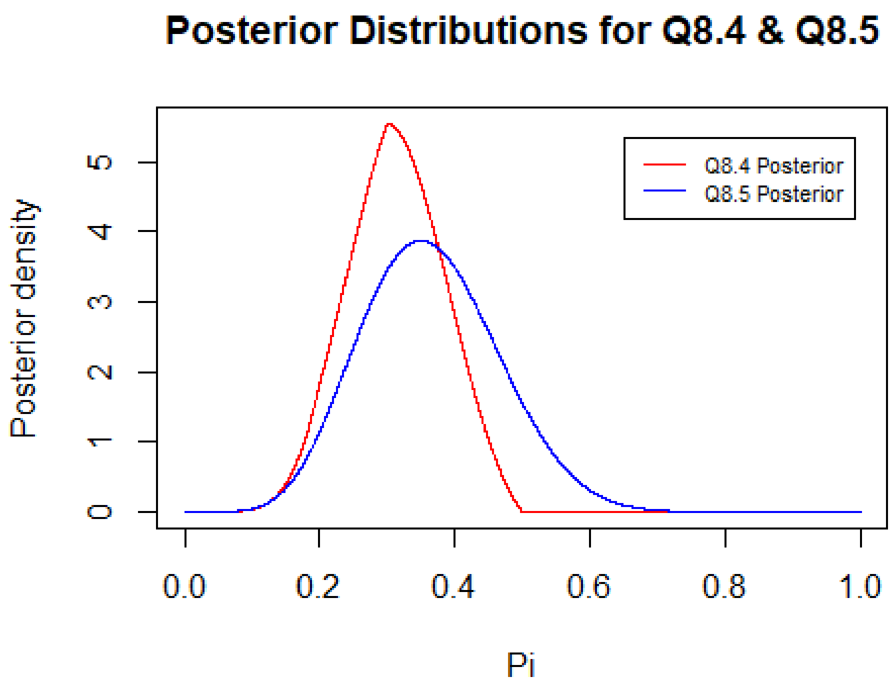

```
cdf = cdf_1$cdf
lower_bound = cdf$x[with(cdf,which.max(x[y<=0.025]))]
upper_bound = cdf$x[with(cdf,which.max(x[y<=0.975]))]
cat(paste("Approximate 95% credible interval: [", round(lower_bound, 4),
          " ", round(upper_bound, 4), "]\n", sep = ""))

## Approximate 95% credible interval: [0.1809 0.5696]
```

8.6 Graph the two posterior distributions on the same graph. What do you notice?

```
# Load the dependent libraries
library(Bolstad)

##
## Attaching package: 'Bolstad'

## The following objects are masked from 'package:stats':
##
##     IQR, sd, var

pi_values = seq(from = 0, to = 1, by=0.0001)

pi_prior = rep(0,length(pi_values))
pi_prior[pi_values<=0.2] = pi_values[pi_values<=0.2]
pi_prior[pi_values>0.2 & pi_values<=0.3] = 0.2
pi_prior[pi_values>0.3 & pi_values<=0.5] = 0.5 - pi_values[pi_values>0.3 & pi_values<=
0.5]
pi_prior[pi_values>0.5] = 0
output_1 = binogcp(x = 7, n = 20, density="user", pi = pi_values, pi.prior = pi_prior)
```

```
pi_prior = rep(1,length(pi_values))
output_2 = binogcp(x = 7, n = 20, density="user", pi = pi_values, pi.prior = pi_prior)
```

```
# Plotting the two posterior distributions in same graph.
plot(pi_values, output_1$posterior, type='l', xlab = 'Pi', ylab = 'Posterior density',
col="red",
     main = 'Posterior Distributions for Q8.4 & Q8.5')
points(pi_values, output_2$posterior, type='l', xlab = 'Pi', ylab = 'Posterior density
', col="blue")
legend(x = 0.65, y = 5.35, legend=c("Q8.4 Posterior", "Q8.5 Posterior"), col = c("red"
, "blue"), lty=1:1, cex=0.7)
```



From the graphs we can say that the posterior distribution for both cases tend to distribute near lower values of π and the shapes are similar. For the first case (Q8.4) the distribution is centered at π value lesser than the later (Q8.5). Also, the variance in the first case is lower than the later.

What do you notice about the two posterior means and standard deviations?

The posterior means and standard deviations calculated from previous stages are,

(Q8.4) Mean = `0.3120479 ; Standard deviation = 0.07070219`

(Q8.5) Mean = `0.3636364 ; Standard deviation = 0.100305`

As noticed from the graph as well, the mean for the posterior distribution in Q8.4 is lesser than that in Q8.5. This could be due to the fact that we have chosen a prior which has 0 density for the second half of π values and hence the posterior too has values of 0 for $\pi \geq 0.5$. The variance for the posterior in Q8.4 is lesser than that in Q8.5. This could also arise because the prior is null for $\pi \geq 0.5$ in Q8.4.

What do you notice about the two credible intervals for π?

The two credible intervals as calculated earlier are given by,

(Q8.4) Approximate 95% credible interval: [0.175 0.4473]

(Q8.5) Approximate 95% credible interval: [0.1809 0.5696]

The credible intervals here are a reflection of what the posterior means and standard deviations suggest. The interval for the first case is shorter and is more towards the lower side than the later.