Assignment 2 - STAT20180

Ross Jackson, 19343511

2023-02-28

Load in necessary packages:

```
options(repos = "https://cran.rstudio.com/")
install.packages("ProbBayes")
```

```
##
## The downloaded binary packages are in
##  /var/folders/fm/x16t1xt921q_b292x5blbrt80000gn/T//Rtmp7FzSsf/downloaded_packages
```

```
install.packages("HDInterval")
```

```
##
## The downloaded binary packages are in
##  /var/folders/fm/x16t1xt921q_b292x5blbrt80000gn/T//Rtmp7FzSsf/downloaded_packages
```

```
library(ProbBayes)
```

```
## Loading required package: LearnBayes
## Loading required package: ggplot2
## Loading required package: gridExtra
## Loading required package: shiny
```

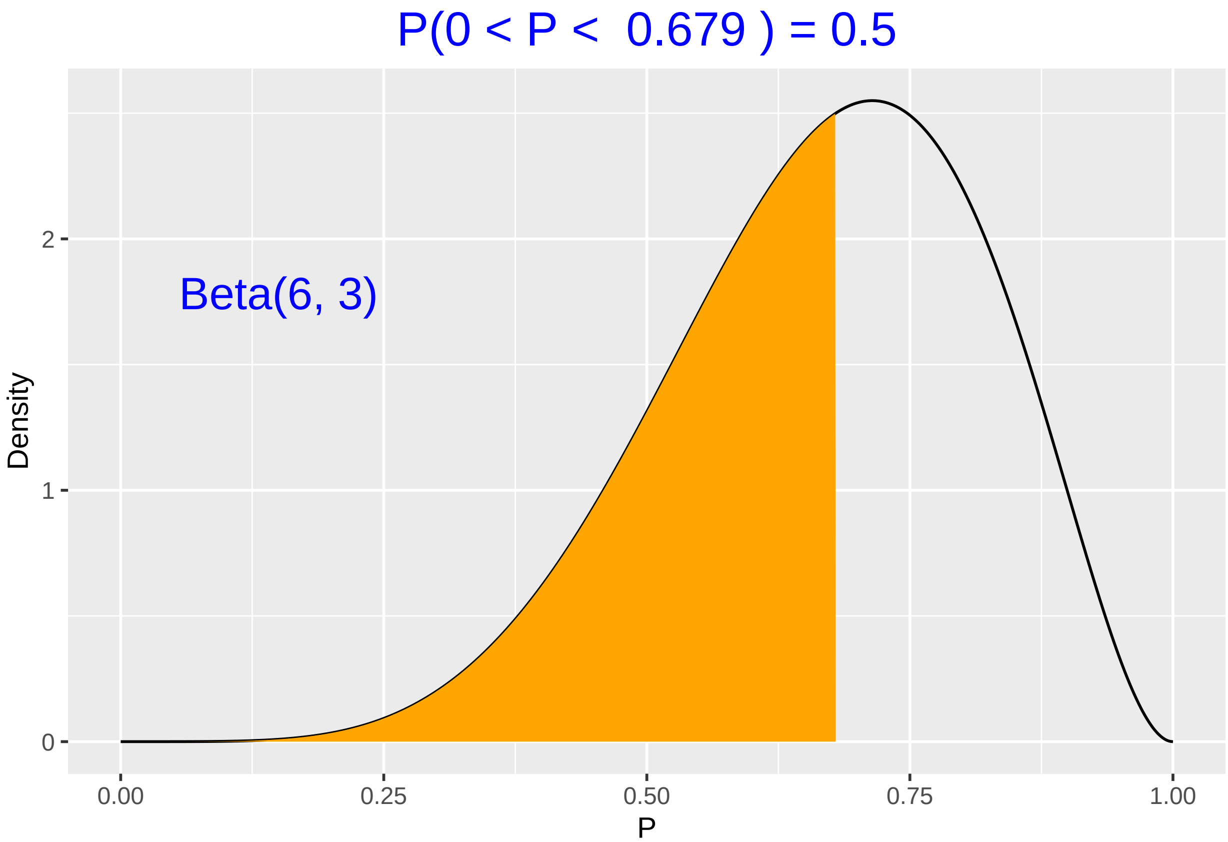
```
library(HDInterval)
```

Question 1

Consider a Beta(6,3) prior distribution for a proportion parameter :

- What is the 50th percentile for this distribution?

```
beta_quantile(0.5, c(6, 3))
```



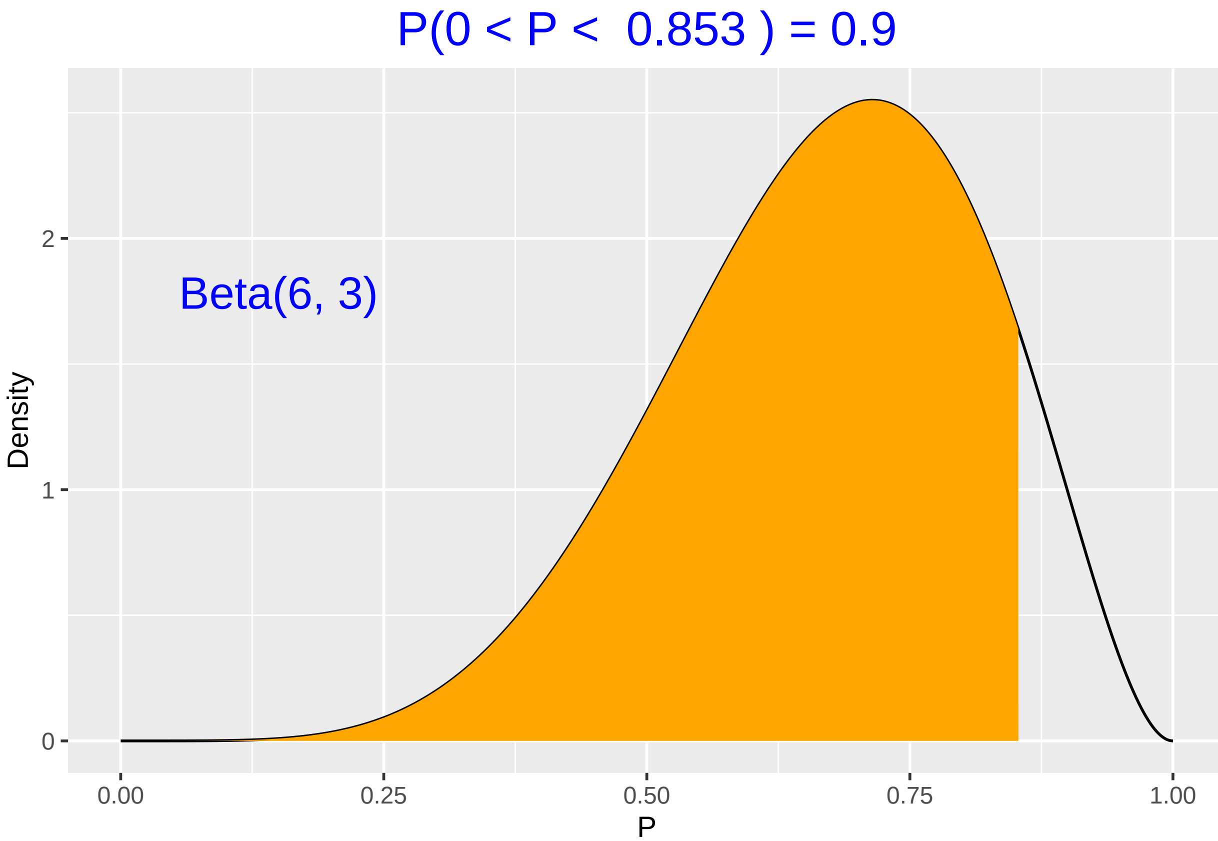

```
# Create Beta(6,3) prior distribution
prior <- dbeta(seq(0,1,length.out=1001), 6, 3)

# Compute the 50th percentile
quantile(prior, 0.5)
```

```
##       50%
## 0.7261673
```

From the distribution and the results we have obtained above, we can see that 50% of the Beta(6,3) distribution falls below the value 0.7261673.

- What is the 90th percentile?

```
beta_quantile(0.9, c(6, 3))
```




```r
# Compute the 90th percentile
quantile(prior, 0.9)
```

```
##     90%
## 2.44024
```

We can interpret from the result above that 90% of the values in a Beta(6,3) distribution are expected to be less than 2.44024. This is a significantly larger value than we obtained for the 50th percentile, and is explained by the fact that the Beta(6,3) distribution is clearly skewed to the right, as we can see fro the graph above. These large skewed values then push up the value of the 90th percentile, resulting in the value 2.44024.

- Consider again a Beta(6,3) prior distribution for . What is the probability that [0.6,0.8]?

```r
pbeta(0.8,6,3) - pbeta(0.6,6,3)
```

```
## [1] 0.4815232
```

From this, we can conclude that lies in the interval [0.6,0.8] with a probability of 0.4815232

Question 2:

A survey was conducted to assess if there was public support for the introduction of a local property tax. To do this, a polling company sampled 50 individuals at random and asked them to answer the question: Do you support the introduction of a local property tax? Yes or No. Of the 50 individuals surveyed, 16 answered Yes and 34 answered No. Of primary interest is to assess the uncertainty in, , the probability that an individual supports this bill. To do this, you should carry out a Bayesian analysis. You may assume that , the probability of success (support for the bill) was the same for every individual and that each response is independent of one and other, ie, a likelihood model. Further you can use a Beta(a,b) prior distribution for

- Your prior belief is that the 30th percentile for is 0.3. Also that the 60th percentile for is 0.5. What values of a and b should you choose to match your prior beliefs? Provide an interpretation of your prior beliefs.

```
beta.select(list(x = 0.3, p = 0.3),
            list(x = 0.5, p = 0.6))
```

```
## [1] 1.76 2.21
```

To match our prior beliefs I have used the beta.select function along with the percentile information provided in the question. The result of this chunk of code gives us a pair of values, 1.76 and 2.21, which when used as the parameters in a Beta distribution will match with the percentiles we were given in the question. To make sure my selection was correct, I then verified my results as can be seen below.

Verifying that 30th percentile and 60th percentile = 0.3 and 0.5:

```
qbeta(0.3, 1.76, 2.21)
```

```
## [1] 0.3005123
```

```
qbeta(0.6, 1.76, 2.21)
```

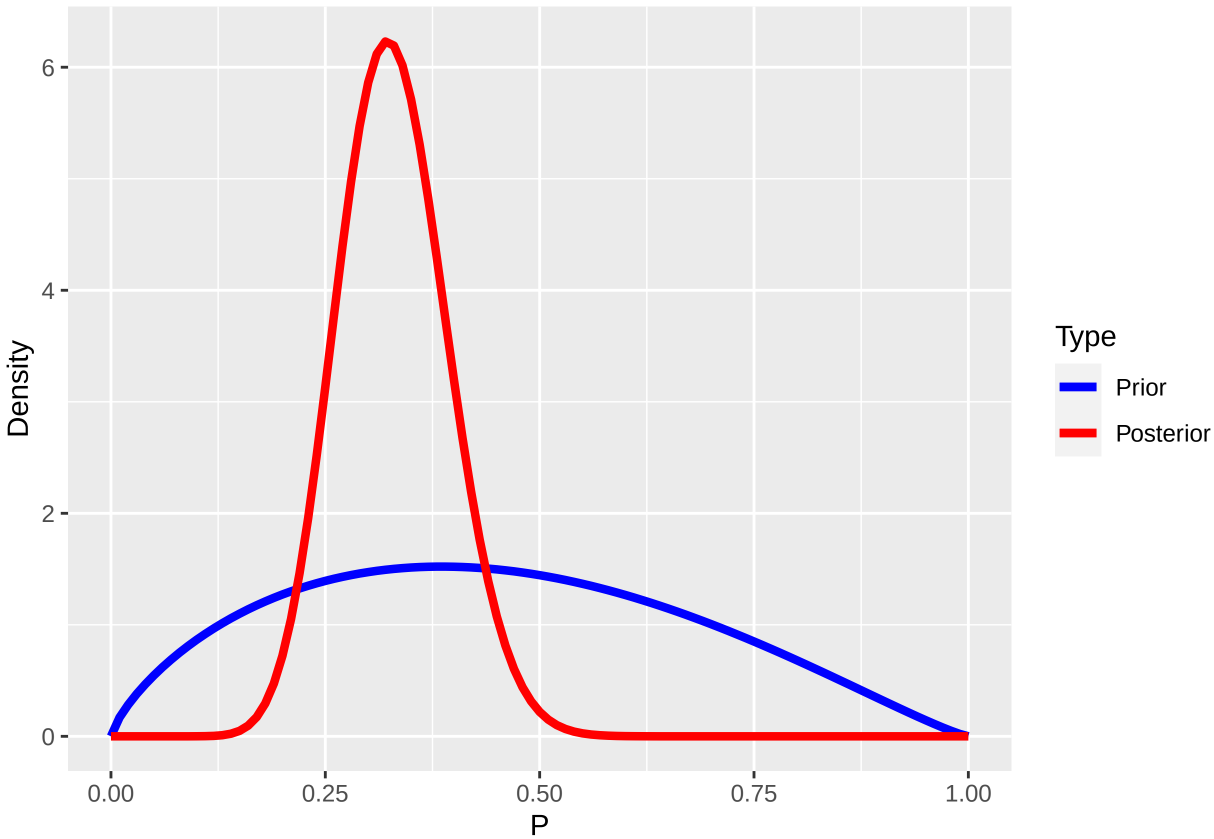
```
## [1] 0.5005008
```

Therefore our values for the prior are correct and are Beta(1.76,2.21)

- What is the posterior distribution? What is the posterior mean? What is the posterior variance? Plot the prior and posterior distributions. Provide a brief commentary

The posterior distribution is easily computed in this case. As the posterior is a Beta distribution, the posterior can be obtained by Beta(a + 16, b + 34), where a and b are the parameter of the prior Beta Distribution, of which we have calculated above.

```
ab <- c(1.76, 2.21)          # Parameters of the prior
ab_new <- c(17.76, 36.21)    # Parameters of the posterior
beta_prior_post(ab, ab_new)
```



As we can see here there is quite a contrast between the prior and posterior distributions. This suggests that the data provided in the question may have had a significant effect on the resulting posterior distribution, which may explain the clear contrasts in their respective shapes. The prior distribution is spread out across all probabilities, while the posterior seems to have a mean around 0.3, around where most of the proability density lies.

- What is the posterior mean?

```
post_mean <- 17.76/(17.76+36.21)
post_mean
```

```
## [1] 0.3290717
```

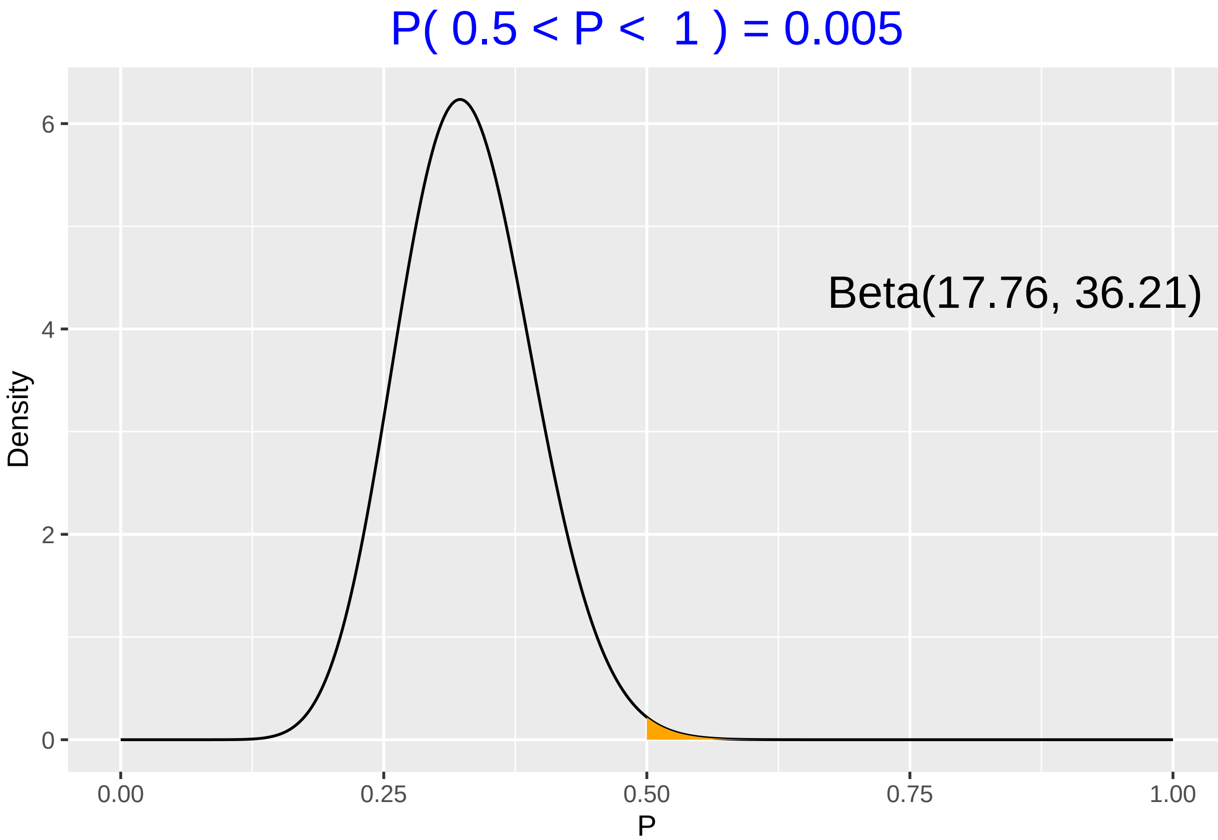
- What is the posterior variance?

```
a=17.76
b=36.21
post_var = (a*b)/((a+b)^2 *(a+b+1))
post_var
```

```
## [1] 0.004016437
```

- A politician claims that the proportion who support this bill is greater than 0.5. Use your posterior model to test this claim. Provide a brief comment

```
beta_area(lo = 0.5, hi = 1.0, shape_par = c(17.76, 36.21))
```



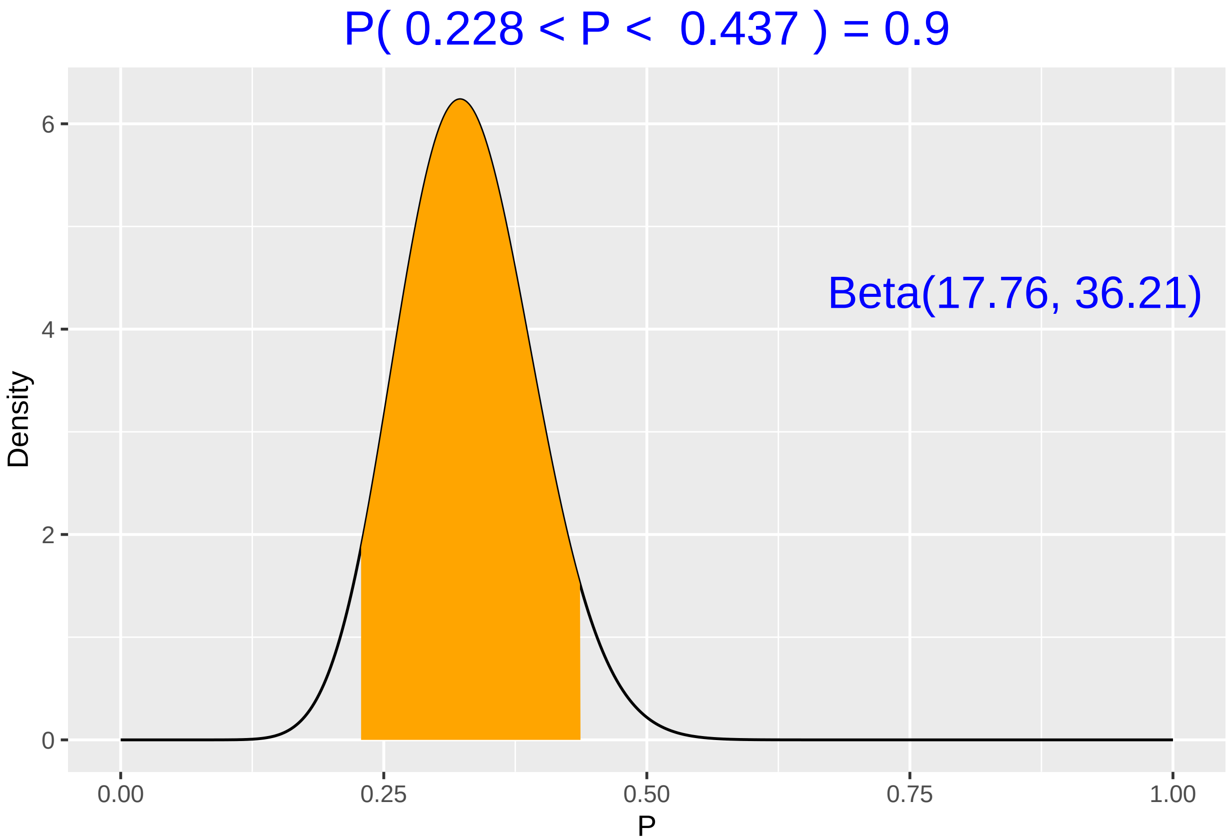

```
1 - pbeta(0.5, 17.76, 36.21)
```

```
## [1] 0.005269747
```

From this graph and along with the corresponding probability of 0.005269747, we can conclude that that this politicians claim is likely to be incorrect.

- Evaluate a 90% credible interval for

```r
beta_interval(0.9, c(17.76, 36.21))
```




```r
qbeta(c(0.05,0.95), 17.76, 36.21)
```

```
## [1] 0.2284594 0.4369530
```

From this resulting interval and graph, we can conclude that we are 90% confident that the true value of lies within the interval [0.2284594, 0.4369530].

- Similarly, evaluate a 90% HPD regional interval.

```r
HDInterval::hdi(qbeta, 0.90, shape1=17.76, shape2=36.21)
```

```
##     lower     upper
## 0.2243944 0.4324057
## attr(,"credMass")
## [1] 0.9
```

From this interval, [0.2243944, 0.4324057], we can firstly see that it slightly disagrees with the 90% credible interval from before. We can conclude that this is the narrowest interval that contains 90% of the posterior probability density. In contrast to the credible interval, this range is more consistent with our prior beliefs and the data.