Problem Section 7

Prior and Posterior

Exercises

1. The R function `leaflet_map` can be used to generate random locations on the globe and to use Google maps to show you where these locations are:

```
set.seed(3535)

leaflet_map(position=rgeo(n=20),mark=TRUE)
```

Use this to gather a sample of size 20 and calculate the proportion of the sample that represents a location which is covered with water. (You may need to zoom in for locations that are near where land and water meet)

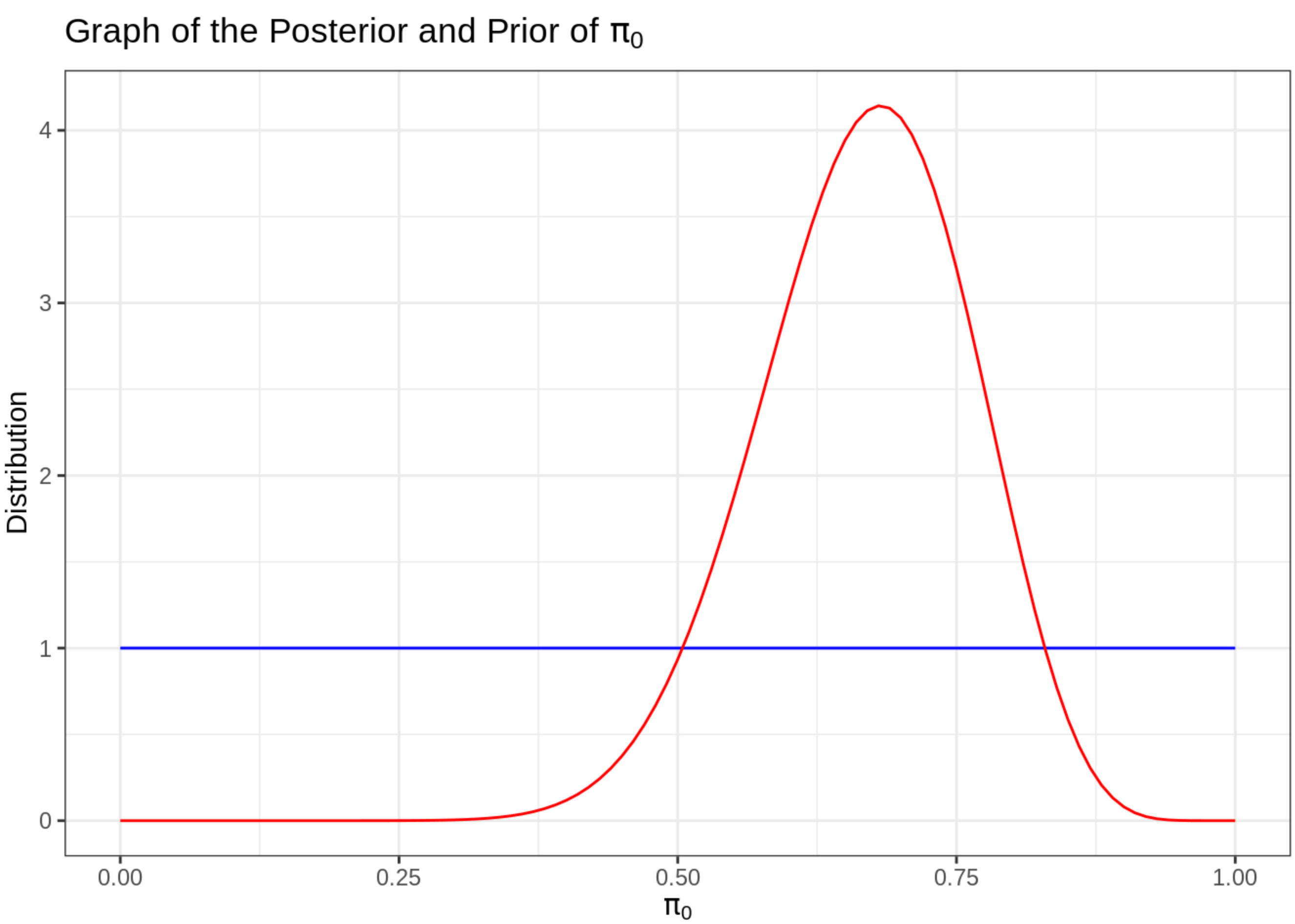
As seen above, the number of locations that land in the water is 14, and the number of locations that don't land in the water is 6. Thus the proportion of the sample that represents a location which is covered with water is hence $\frac{15}{20} = \frac{7.5}{10} = 0.75$.

Let's assume that X, the number out of the 20 randomly selected locations which are covered with water, is a binomial random variable with success probability π.

a. Calculate the posterior distribution for the proportion of the earth that is covered with water. Use a uniform prior for π. Make a plot of the uniform prior and the resulting posterior distribution on the same graph.

If we assume that X, the number out of the 20 randomly selected locations which are covered with water, is a binomial random variable with success probability π. Then the posterior distribution for the proportion of the earth that is covered with water using a uniform prior for π, given that we observed an X value of 14 out of 20 is
$\text{Beta}(x + 1,\ n - x + 1) = \text{Beta}(14 + 1,\ 20 - 14 + 1) = \text{Beta}(16, 8).$

```
ggplot() +
  stat_function(fun = dunif,
                args = list(min = 0, max = 1),
                color = "blue") +
  stat_function(fun = dbeta,
                args = list(shape1 = 16, shape2 = 8),
                color = "red") +
  labs(title = expression("Graph of the Posterior and Prior of" ~ pi[0]),
       x = expression(pi[0]),
       y = "Distribution") +
  theme_bw()
```



b. Now suppose a Bayesian wants to use a beta distribution which reflects the prior knowledge that a majority of the earth is covered with water. Specifically, they assume that the mean of the prior is 0.70 and the variance is 0.05.

i. Keeping in mind that for a $\text{Beta}(\alpha, \beta)$ distribution,

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

and the variance is

$$Var[X] = \left(\frac{\alpha}{\alpha + \beta}\right) \times \left(\frac{\beta}{\alpha + \beta}\right) \times \frac{1}{\alpha + \beta + 1}$$

elicit the values for α and β.
(Hint: Write the variance in terms of the mean as much as possible. It will help to remember that $\frac{\beta}{\alpha+\beta} = 1 - \frac{\alpha}{\alpha+\beta}$)

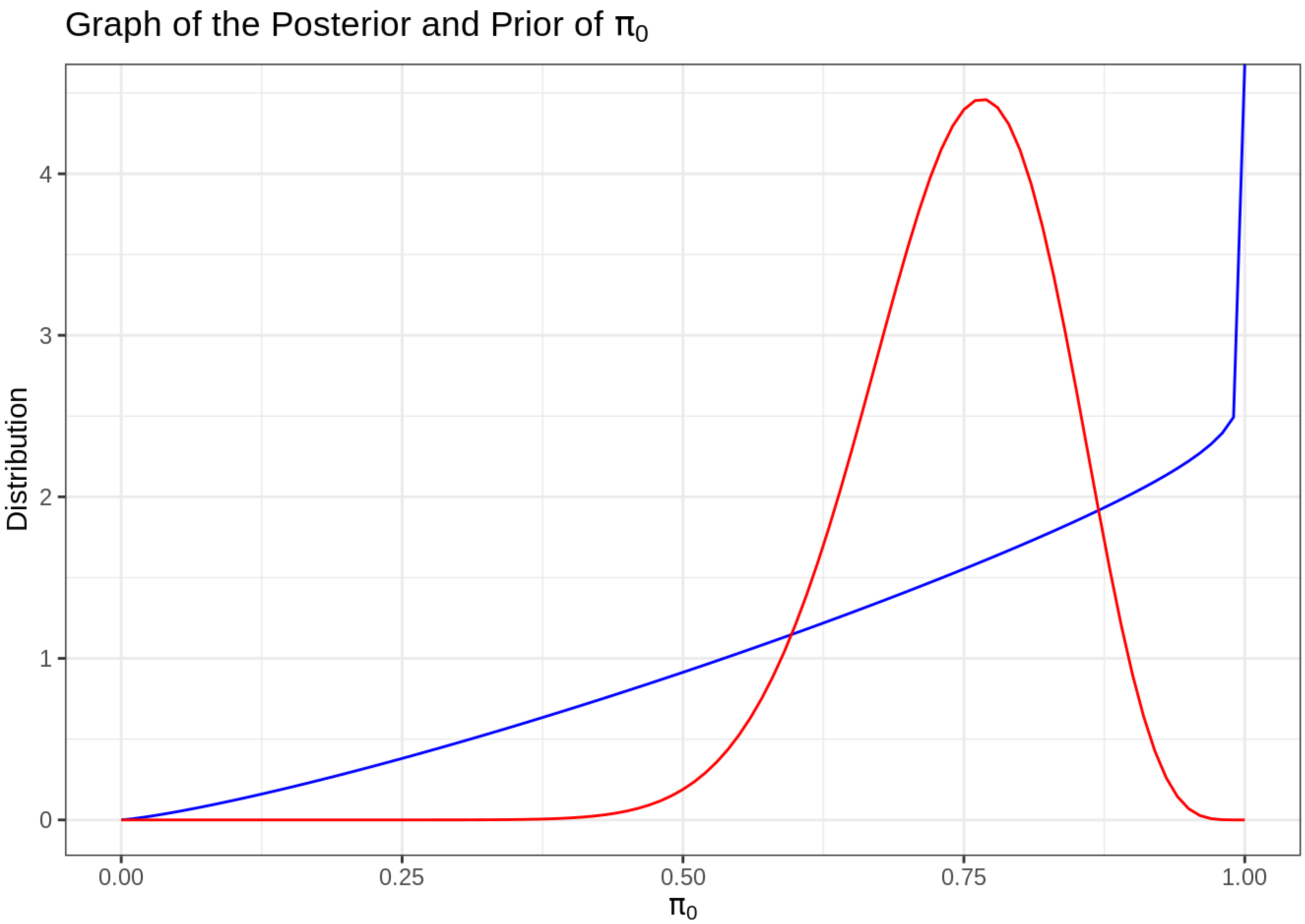
Based on my calculations: $\alpha = 2.24$ and $\beta = 0.96$.

ii. Now calculate the posterior distribution using this informative prior from part i.

Using the information from part i, the posteriror distribution is $\text{Beta}(17.24, 5.96)$.

iii. Make a plot of your prior and the resulting posterior distribution in one graph.

```
ggplot() +
  stat_function(fun = dbeta,
                args = list(shape1 = 2.24, shape2 = 0.96),
                color = "blue") +
  stat_function(fun = dbeta,
                args = list(shape1 = 17.24, shape2 = 5.96),
                color = "red") +
  labs(title = expression("Graph of the Posterior and Prior of" ~ pi[0]),
       x = expression(pi[0]),
       y = "Distribution") +
  theme_bw()
```



c. Now suppose you consider a prior with a median of 0.7 and 95% of the distribution lies above 0.68. Use the `beta.select` function from the **LearnBayes** package to find the values of α and β corresponding to these prior beliefs. Then repeat parts ii and iii from part b with this prior.

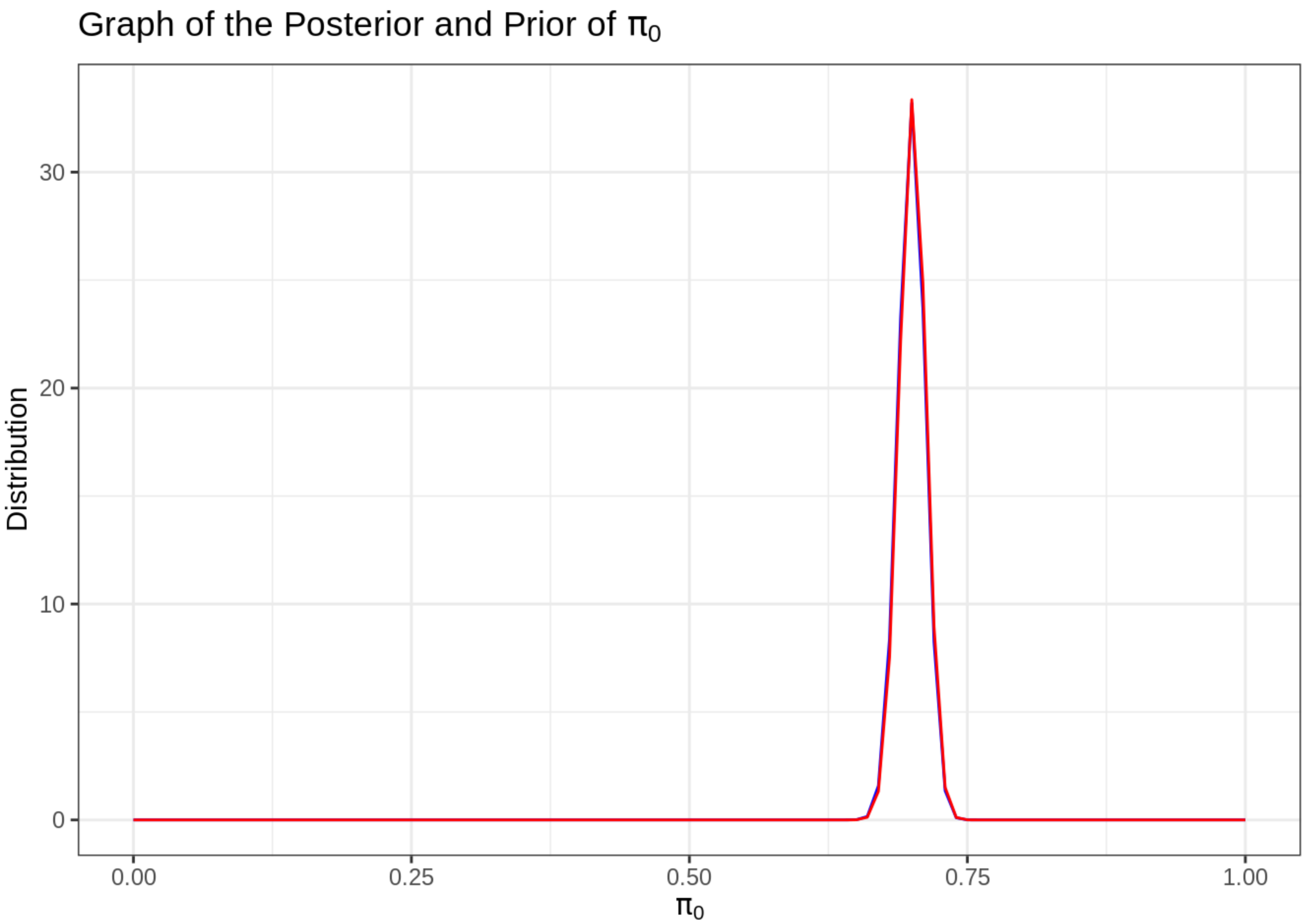
```
beta.select(quantile1 = list(p = 0.05, x = 0.68), quantile2 = list(p = 0.5, x = 0.7))
```

```
## [1] 1017.74  436.34
```

Based on this information, the prior is $\text{Beta}(1017.74, 436.34)$

Hence the posterior is $\text{Beta}(1032.74, 441.34)$

```
ggplot() +
  stat_function(fun = dbeta,
                args = list(shape1 = 1017.74, shape2 = 436.34),
                color = "blue") +
  stat_function(fun = dbeta,
                args = list(shape1 = 1032.74, shape2 = 441.34),
                color = "red") +
  labs(title = expression("Graph of the Posterior and Prior of" ~ pi[0]),
       x = expression(pi[0]),
       y = "Distribution") +
  theme_bw()
```


Graph of the Posterior and Prior of π0
Distribution
30
20
10
0
0.00
0.25
0.50
0.75
1.00
π0