BDA - Assignment 2

Anonymous

Inference for binomial proportion

a)

Likelihood can be expressed as function of π:

$$p(y|\pi, n) = \binom{n}{k} \pi^y (1-\pi)^{n-y} = Binomial(n, \pi).$$

The prior $p(\pi)$ can be expressed as: $Beta(2, 10)$ And the posterior $p(\pi|y)$ as $Beta(2 + y, 10 + n - y)$. We can compute n and y from our data

```
n <- length(algae)
y <- length(which(algae == 1))
cat("n: ", n, "\n")
```

```
## n:  274
```

```
cat("y: ", y)
```

```
## y:  44
```

Therefore, we can express posterior in its final form as $Beta(46, 240)$.

b)

```
beta_point_est <- function(prior_alpha, prior_beta, data){
  return ((prior_alpha + length(which(data == 1))) / (prior_alpha + prior_beta + length(data)))
}

beta_interval <- function(prior_alpha, prior_beta, data, prob){
  n <- length(data)
  y <- length(which(data == 1))
  interval_bottom <- qbeta((1 - prob) / 2, prior_alpha + y, prior_beta + n - y)
  interval_top <- qbeta(prob + (1 - prob) / 2, prior_alpha + y, prior_beta + n - y)
  return (c(interval_bottom, interval_top))
}
beta_point_est(prior_alpha = 2, prior_beta = 10, data = algae)
```

```
## [1] 0.1608392
```

```
beta_interval(prior_alpha = 2, prior_beta = 10, data = algae, prob=0.9)
```

```
## [1] 0.1265607 0.1978177
```

From the code above, we have that $E(\pi|y) \approx 0.16$. And that the 90 posterior interval spans between 0.13 and 0.2.

c)

```r
beta_low <- function(prior_alpha, prior_beta, data, pi_0){
  n <- length(data)
  y <- length(which(data == 1))
  return (pbeta(pi_0, prior_alpha + y, prior_beta + n - y))
}
beta_low(prior_alpha = 2, prior_beta = 10, data = algae, pi_0 = 0.2)
```

```
## [1] 0.9586136
```

The probability that the proportion of monitoring sites with detectable algae levels π is smaller than $\pi_0 = 0.2$ is 95.86.

d)

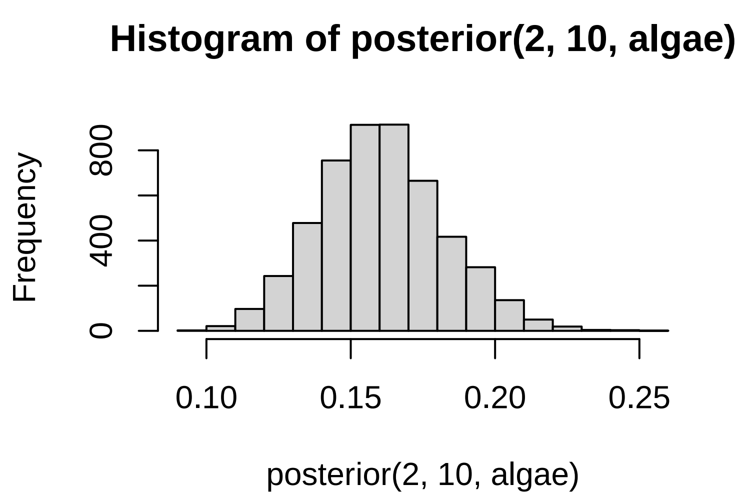
For a binomial model of this kind we are assuming that the events, i.e. detectable algae levels are independent and that the probability of presence equal to π for all events.

e)

```r
posterior <- function(prior_alpha, prior_beta, data){
  n <- length(data)
  y <- length(which(data == 1))
  return (rbeta(5000, prior_alpha + y, prior_beta + n - y))
}
```

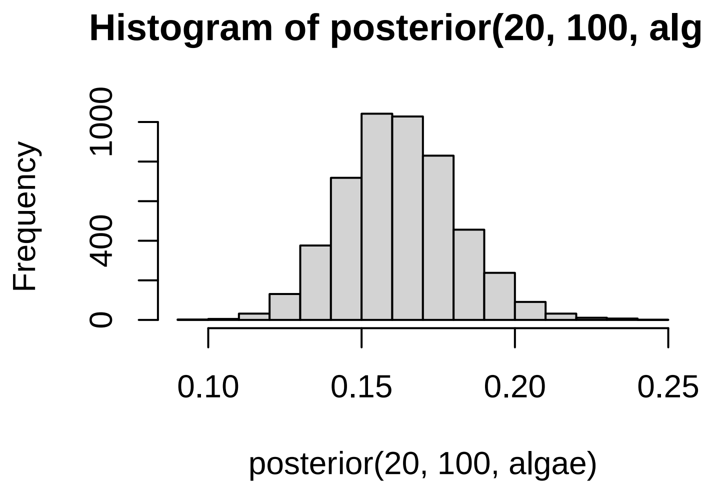
First we can plot and inspect the baseline prior $Beta(2, 10)$ where we have 11 prior observations.

```r
hist(posterior(2, 10, algae))
```



From previous subtask we know that the $E(\pi|y) \approx 0.16$. Now we can increase the number of prior observations to 118 total observations, i.e. $Beta(20, 100)$.

```r
hist(posterior(20, 100, algae))
```



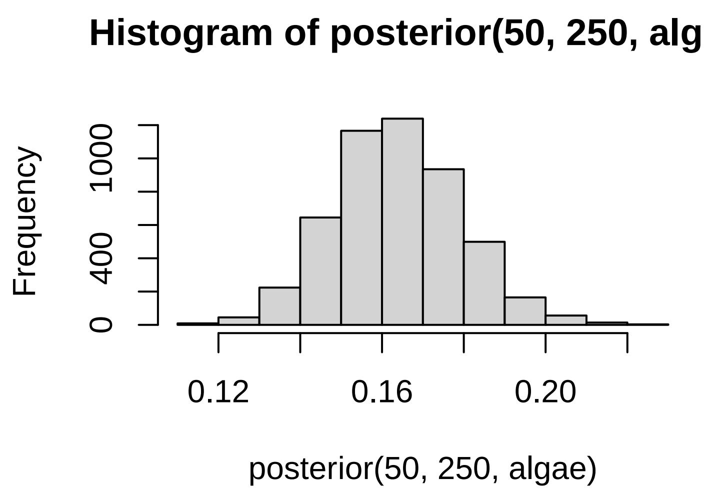
And we can compute $E(\pi|y)$ once again.

```
beta_point_est(prior_alpha = 20, prior_beta = 100, data = algae)
```

```
## [1] 0.1624365
```

Lastly, we can try another set of parameters, with even more prior observations, e.g. $Beta(50, 250), which brings us to the total of 298 prior observations.

```
hist(posterior(50, 250, algae))
```



And compute $E(\pi|y)$ once again

```
beta_point_est(prior_alpha = 50, prior_beta = 250, data = algae)
```

```
## [1] 0.1637631
```

We can see that the posterior is very robust to different priors as while we move the prior (its mean) around quite a lot, the posterior's expectations stays around the same area, i.e. 0.16.