Applied_Stat_2_Lab3

Question 1

```r
mle <- 118/129

se <- sqrt(mle * (1 - mle)/129)
CI <- c(mle - 1.96 * se, mle + 1.96 * se)
CI
```

```
[1] 0.8665329 0.9629244
```

Question 2

```r
alpha_1 <- 1
beta_1 <- 1

post_alpha_1 <- alpha_1 + 118
post_beta_1 <- beta_1 + (129 - 118)

post_mean_1 <- post_alpha_1/(post_alpha_1 + post_beta_1)
post_mean_1
```

```
[1] 0.9083969
```

```r
bayes_CI_1 <- c(qbeta(0.025, post_alpha_1, post_beta_1), qbeta(0.975, post_alpha_1, post_b
bayes_CI_1
```

```
[1] 0.8536434 0.9513891
```

Question 3

```
alpha_10 <- 10
beta_10 <- 10

post_alpha_10 <-  alpha_10 + 118
post_beta_10 <- beta_10 + (129 - 118)

post_mean_10 <- post_alpha_10/(post_alpha_10 + post_beta_10)
post_mean_10
```

```
[1] 0.8590604
```

```
bayes_CI_10 <- c(qbeta(0.025, post_alpha_10, post_beta_10), qbeta(0.975, post_alpha_10, po
bayes_CI_10
```

```
[1] 0.7990363 0.9099708
```

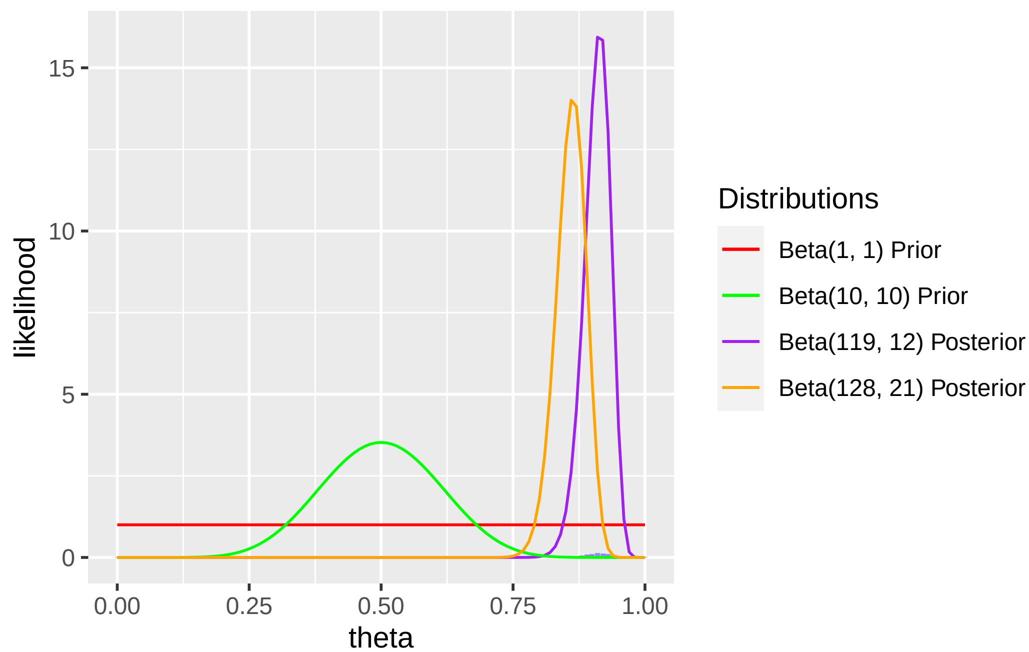
The prior $Beta(10, 10)$ is more informative than the prior $Beta(1, 1)$. It basically means we assume that there we observed 10 happy women and 10 unhappy women. Thus we assume more certainty that $\theta = 0.5$ and consequently also get a credible interval which is closer to 0.5.

Question 4

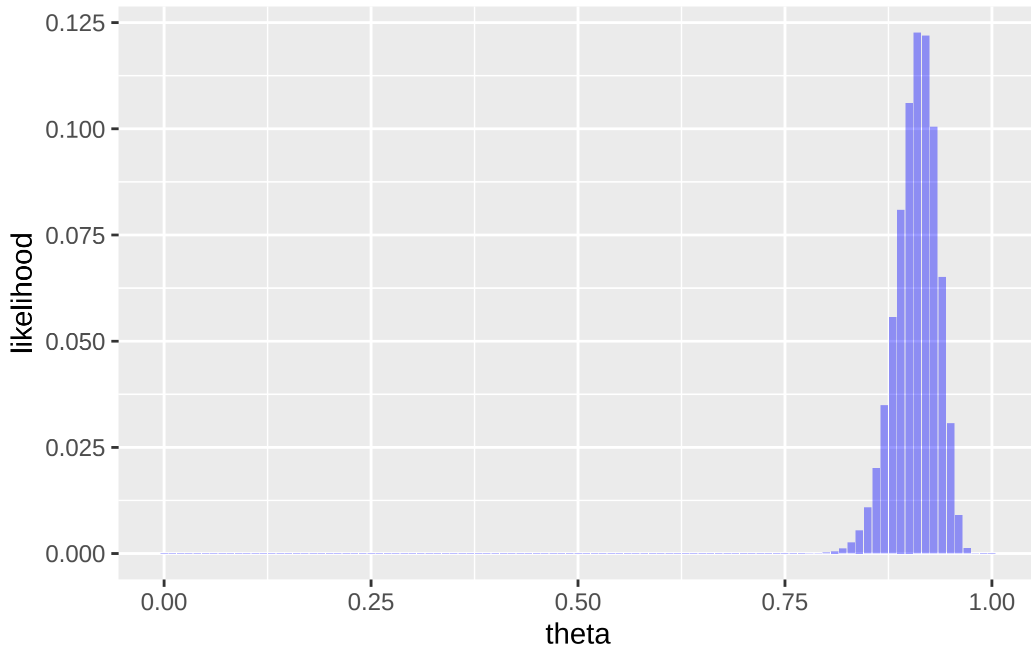
```
library(ggplot2)

theta_values <- seq(0, 1, by = 0.01)
  likelihood <- dbinom(118, size = 129, prob = theta_values)
  data <- data.frame(theta = theta_values, likelihood = likelihood)

p <- ggplot(data, aes(x = theta, y = likelihood)) +
  geom_histogram(stat = "identity", bins = 30, fill = "blue", alpha = 0.4)
```

```
Warning in geom_histogram(stat = "identity", bins = 30, fill = "blue", alpha =
0.4): Ignoring unknown parameters: `binwidth`, `bins`, and `pad`
```

```r
p + stat_function(fun = dbeta, args = list(shape1 = 1, shape2 = 1), aes(colour = "Beta(1,
  stat_function(fun = dbeta, args = list(shape1 = 1 + 118, shape2 = 1 + 11), aes(colour =
  stat_function(fun = dbeta, args = list(shape1 = 10, shape2 = 10), aes(colour = "Beta(10,
  stat_function(fun = dbeta, args = list(shape1 = 10 + 118, shape2 = 10 + 11), aes(colour
  labs(colour = "Distributions") +
  scale_colour_manual(values = c("red", "green", "purple", "orange"))
```




```r
# Display the plot
print(p)
```



We know that the $Beta(10, 10)$ prior has more certainty that $\theta = 0.5$. This is also reflected in the plot since the graph of the posterior with respect to the $Beta(10, 10)$ prior is centered a bit closer to 0.5 compared to the posterior of $Beta(1, 1)$ prior. Compared to this, the likelihood function seems to agree more with the $Beta(1, 1)$, centering around 85-90%.

Question 5

Assuming a $Beta(1, 1)$ prior we have the beta posterior resulting in the following probability:

```
x <- 251527
y <- 241945

prob <- pbeta(0.5, x+1 , y+1)
prob
```

```
[1] 1.146058e-42
```

Question 6

- Uninformative prior: In the case of an uninformative prior I would suggest a uniform distribution, since we don't assume any knowledge of the improvement rate after practice. Hence we consider all values of θ equally likely.
- Informative prior: In this case I would choose a distribution which centers around the improvement rate that I would expect from the training program. For example if I assume an improvement rate of 20% I would choose a α and β for a beta-distribution, such that the the mean $\alpha/(\alpha +) = 0.2$.