Lab Week 3

Rosa Fallahpour

- **(1)** Assuming $Y|\theta \sim Bin(n, \theta)$, where y is the number of women who reported to be happy out of the sample of n women. We know that 118 out of 119 women reported to to be happy. So, in order to find out the maximum likelihood estimation of our parameter of interest which is θ, we have following procedure:

$$f(y_i; \theta) = \binom{129}{118}(1-\theta)^{11}\theta^{118}$$

We have the likelihood function as below:

$$L(\theta; y_i) = \prod_{i=1}^{129}\binom{129}{118}(1-\theta)^{11}\theta^{118},$$

taking log likelihood we will have:

$$l(\theta; y_i) = 129\big[ln\binom{129}{118} + 11ln(1-\theta) + 118ln\theta\big],$$

taking derivative with respect to θ and set it to zero, we will have MLE of θ as $\hat{\theta} \approx 0.91$. To calculate the 0.95 confidence interval from a normal distribution we have:

```
n <- 129
se <- sqrt(0.91*0.08/n)
CI <- c(0.91-qnorm(0.975)*se,0.91+qnorm(0.975)*se)
CI
```

```
## [1] 0.8634394 0.9565606
```

- **(2)** According to the lecture given $Y|\theta \sim Bin(n, \theta)$ and $\theta \sim Beta(1, 1)$ ($p(\theta)$), we will have the posterior distribution as $\theta|y \sim Beta(y + 1, n - y + 1)$, thus:

$$\theta|y \sim Beta(119, 12)$$

  We have the posterior mean for $\hat{\theta}$ and 95% credible interval as following:

```
Posterior_Mean <- 119/(119+12)
Posterior_Mean
```

```
## [1] 0.9083969
```

```
cred_int <- c(qbeta(0.025, 119, 12), qbeta(0.975, 119, 12))
cred_int
```

```
## [1] 0.8536434 0.9513891
```

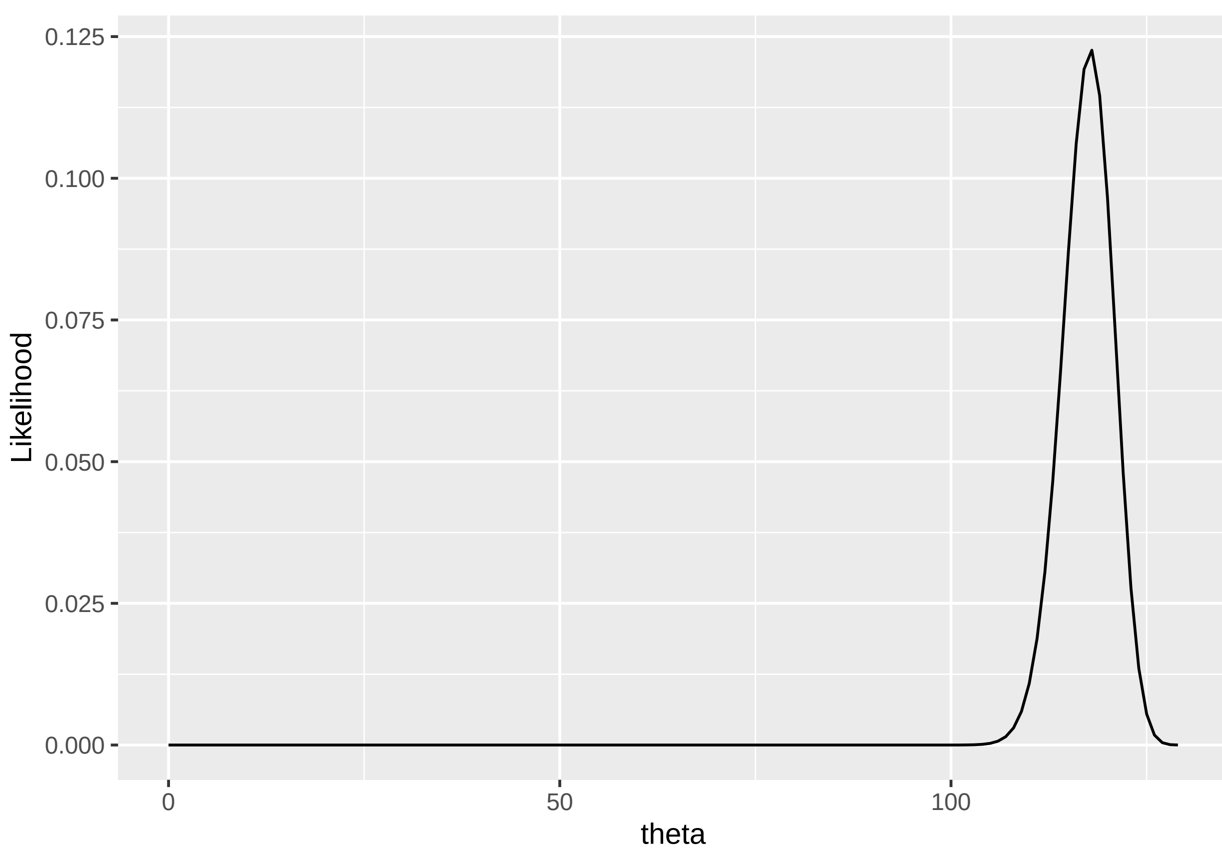
- **(3)** By considering the prior distribution $Beta(10, 10)$ We are assuming more information about θ. Since, the expected proportion of women 65+ being happy ($\frac{10}{10+10} = 0.5$), gains more weight in $Beta(10, 10)$ than $Beta(1, 1)$, which does not assume any preferences of the range for θ. We also can see that the expected prior value is centered in the range of $[Q1, Q3] = [0.44, 0.58]$, which is tighter than $Beta(1, 1)$ distribution in question 2.

```
summary(rbeta(100, 10, 10))
```

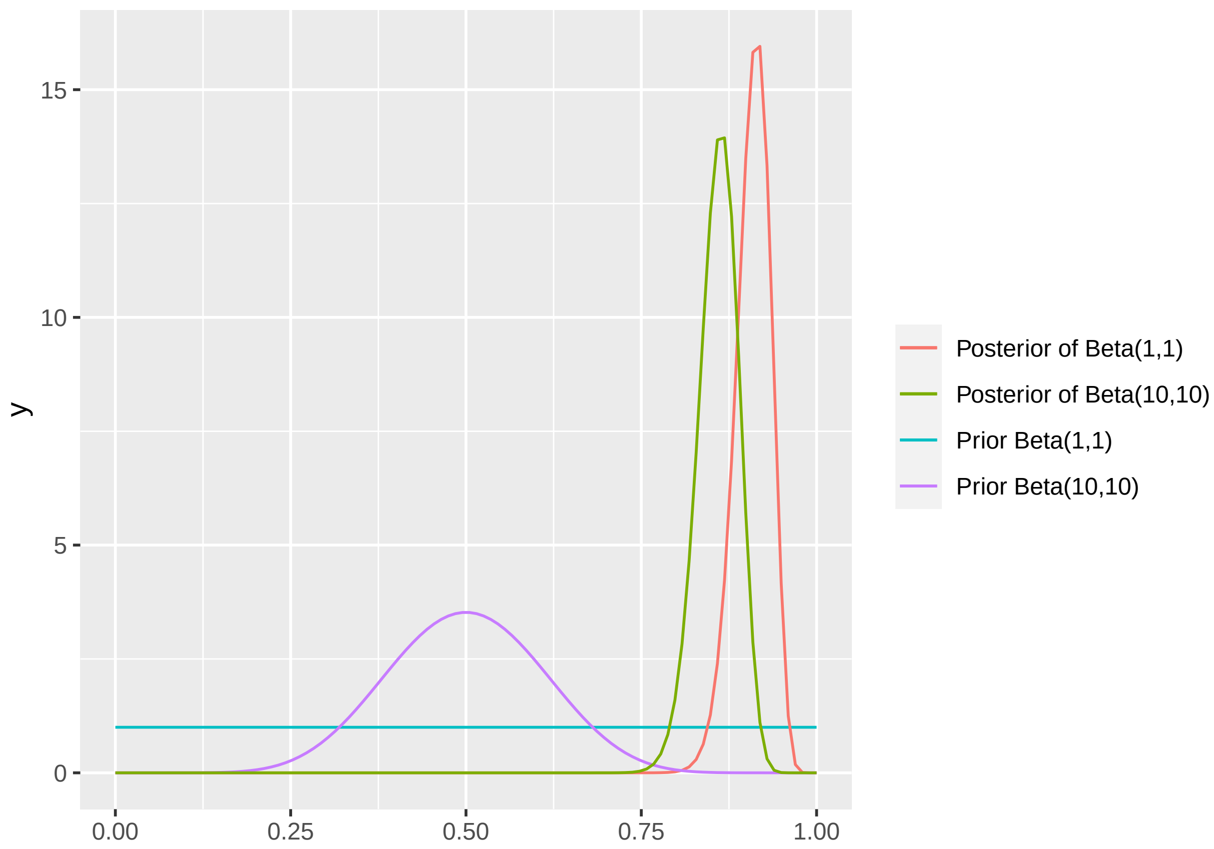
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2664  0.4000  0.4942  0.4911  0.5645  0.7751
```

- **(4)** As the following graphs display, the posterior distributions related to priors $Beta(1, 1)$ and $Beta(10, 10)$ are close. However, the posterior of $Beta(1, 1)$ shifts towards 1. In addition, the posterior mean of prior $Beta(1, 1)$ is closer to MLE of θ.

```
library(tidyverse)
theta <- seq(0,129)
likefun <- dbinom(theta,129,0.91)
df <- data.frame(x=theta,y=likefun)
df |> ggplot(aes(x=theta,y=likefun))+geom_line()+labs(x="theta",y="Likelihood")
```



```
df |> ggplot(aes())+stat_function(fun = dbeta, n = 100, args = list(shape1 = 1, shape2 = 1), aes(colour
  stat_function(fun = dbeta, n = 100, args = list(shape1 = 119, shape2 = 12), aes(colour = "Posterior o
  stat_function(fun = dbeta, n = 100, args = list(shape1 = 10, shape2 = 10), aes(colour = "Prior Beta(1
  stat_function(fun = dbeta, n = 100, args = list(shape1 = 128, shape2 = 21), aes(colour = "Posterior o
```



- **(5)** A noninformative prior distribution for θ could be uniform distribution over real line ($\theta \sim U(-\infty, \infty)$), which does not assign any weight on θ over this interval, thus does not provide any specific information about it.
  Since we have θ defined as average improvement in success probability, it takes value between 0 and 1. In addition, we expect an improvement in student's performance after practicing for one month, therefore, I would consider $Beta(2, 5)$ as the prior distribution for this parameter of interest.