Week 3: Intro to Bayes

27/01/23

Question 1

```r
library(ggplot2)

thetahat <- 118/129
se.thet <- sqrt(thetahat*(1-thetahat)/129)
ci.thet <- c(thetahat - 1.96*se.thet,thetahat +1.96*se.thet)
```

$\hat{\theta} = 118/129 = 0.915$ with 95% confidence interval (0.867, 0.963)

Question 2

Since the posterior distribution of $\theta|y \sim Beta(y+1, n-y+1)$, the posterior mean is given by $\frac{y+1}{y+1+n-y+1} = \frac{118+1}{129+2} = 0.908$.

```r
cred.int <- c(qbeta(0.025, 119, 129-118+1), qbeta(0.975, 119, 129-118+1))
```

The 95% credible interval is given by 0.854, 0.951

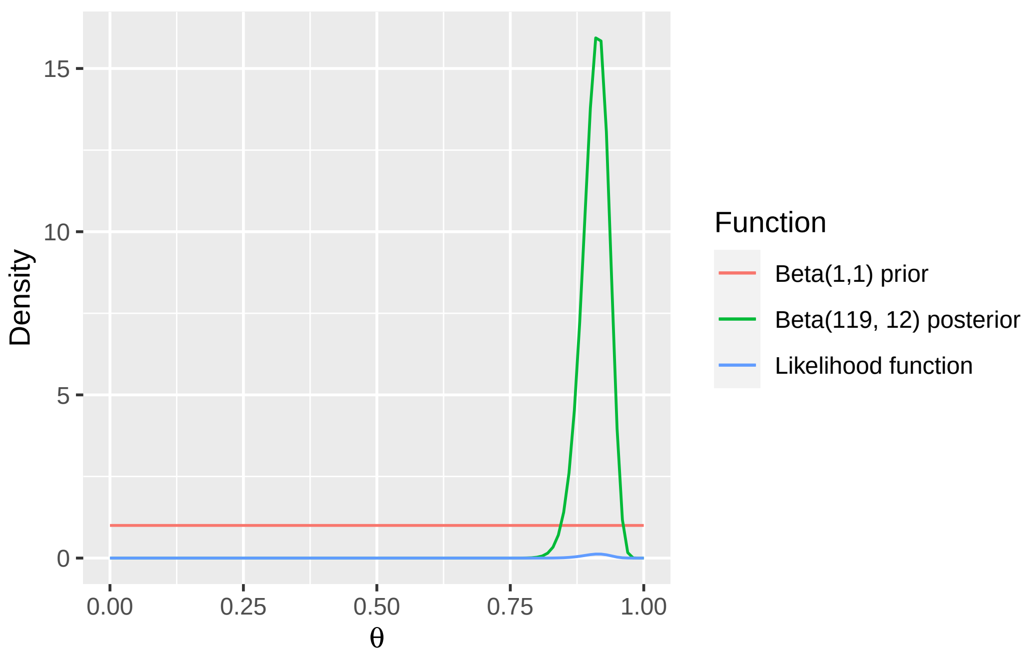
Question 3

Now assume a Beta(10,10) prior on θ. What is the interpretation of this prior? Are we assuming we know more, less or the same amount of information as the prior used in Question 2?

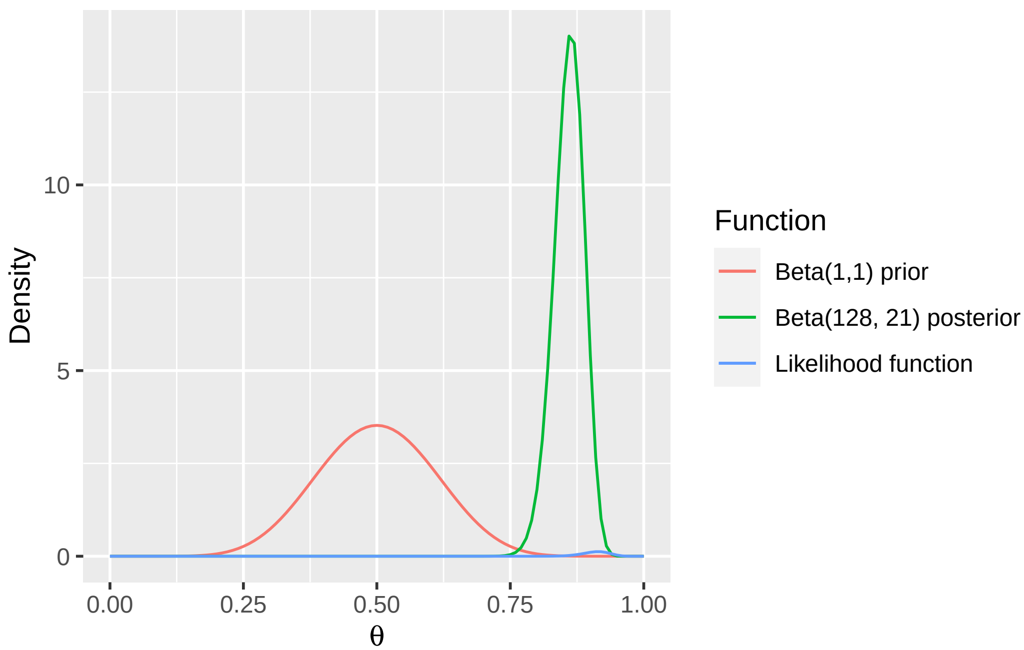
The prior assumes that the expected proportion of women aged 65+ who are happy is $(10)/(10+10) = 0.5$. We assume we know more information since the prior places more weight on certain values of θ (symmetric around 0.5) than others unlike Beta(1,1) which places equal amount of information over all possible values of θ.

Question 4

```
# Likelihood of binomial distribution
bin.ll <- function(theta){
  dbinom(118, size = 129, prob = theta)
}


colors <- c("Likelihood function" = "blue",
            "Beta(1,1) prior" = "red",
            "Beta(119, 12) posterior" = "orange")


ggplot() + xlim(0, 1) +
  geom_function(fun = dbeta, args = list(1,1),
                aes(col = "Beta(1,1) prior")) +
  geom_function(fun = dbeta, args = list(119,129-118+1),
                aes(col = "Beta(119, 12) posterior")) +
  geom_function(fun = bin.ll,
                aes(col = "Likelihood function")) +
  labs(x = bquote(theta), y = "Density", color = "Function" )
```



```r
colors <- c("Likelihood function" = "blue",
            "Beta(10,10) prior" = "red",
            "Beta(119, 12) posterior" = "orange")


ggplot() + xlim(0, 1) +
  geom_function(fun = dbeta, args = list(10,10),
                                    aes(col = "Beta(1,1) prior")) +
  geom_function(fun = dbeta, args = list(118 + 10 ,129-118+10),
                aes(col = "Beta(128, 21) posterior")) +
  geom_function(fun = bin.ll,
                aes(col = "Likelihood function")) +
  labs(x = bquote(theta), y = "Density", color = "Function" )
```



The posterior distribution shifts slightly to the left.

Question 5

Denote $\theta = p_2 - p_1$ to be the change in proportions where p_2 is the proportion after the training program, and p_1, before the training program.

- A noninformative prior could be a uniform prior on $\mathbb{R}$, $\theta \sim U(-\infty, \infty)$ which assigns equal probability to every possible value of θ.
- It can be expected that the performance of the students would improve (i.e $\theta > 0$) but that the change in improvement is less than 1 ($\theta < 1$). Further, we assume that the performance of most students will not drastically improve. Hence, we choose a Beta(1.5, 3) prior (density shown below).

```
ggplot() + xlim(0, 1) + geom_function(fun = dbeta, args = list(1.5,3))
```