Lab exercise3

Chenxi Liu 1010615050

2024-01-26

Question 1

```
n<- 129
y<- 118
p<- round(y/n, 3)
# Standard error
se <- sqrt(p * (1 - p) / n)
# Confidence Interval
CL_upper <- p+qnorm(0.975)*se
CL_lower <- p-qnorm(0.975)*se

p
```

```
## [1] 0.915
```

```
CL_lower
```

```
## [1] 0.8668747
```

```
CL_upper
```

```
## [1] 0.9631253
```

Question 2

```
alpha <- 1
beta <- 1
mean_theta <- round((alpha+y)/(alpha + beta + n),3)

# Confidence Interval
# Calculate the 95% credible interval
credible_interval <- qbeta(c(0.025, 0.975), alpha+y, beta + n - y)

mean_theta
```

```
## [1] 0.908
```

```
credible_interval
```

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

Question 3

Answer3:

Comparing it to the Beta(1,1) prior used in Question 2, the Beta(10,10) prior indicates that we are assuming more prior information about the distribution of theta. With alpha = 10 and beta = 10, we are expressing a more concentrated and specific belief about the proportion of happy women compared to the uniform distribution of the Beta(1,1) prior, we are more confident to assume the proportion of happy woman to be 0.5.

Question 4

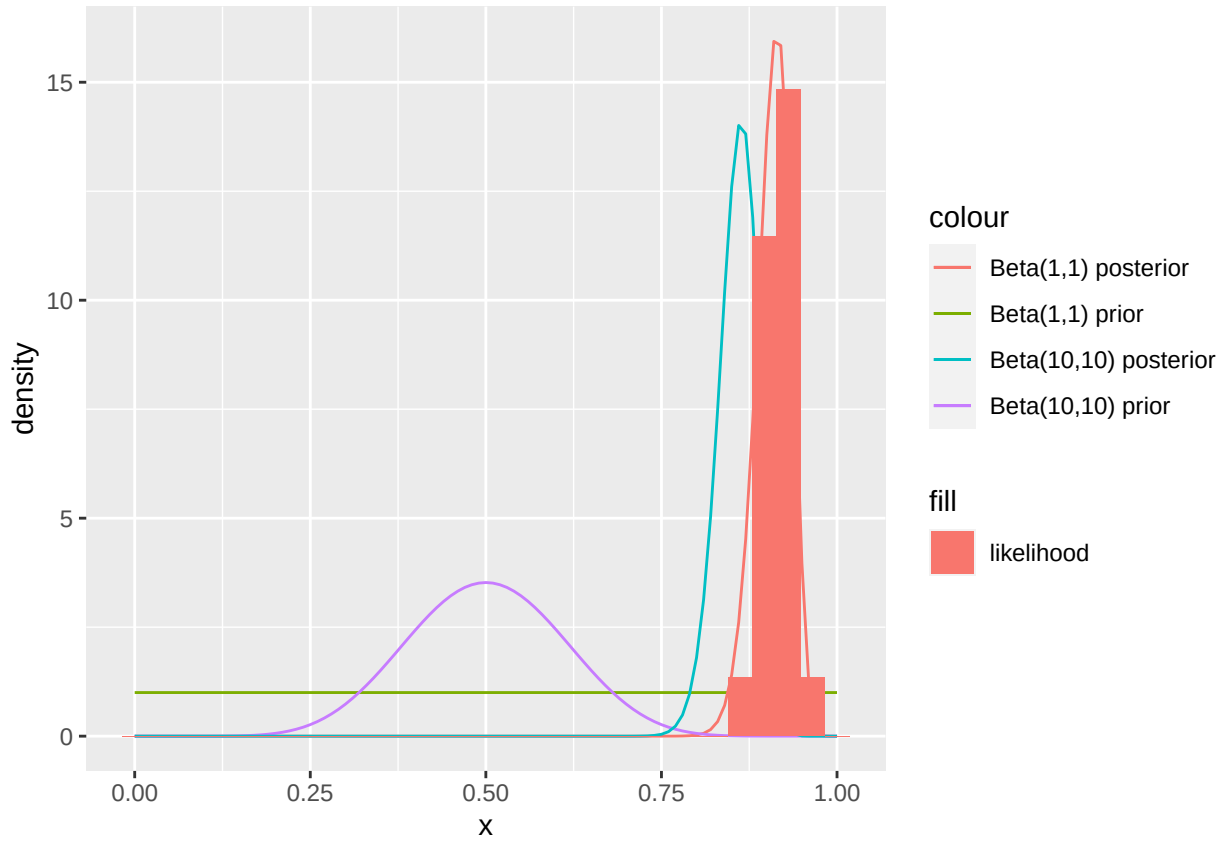
```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
plot <- ggplot(data = tibble(x = seq(0, 1, by = 0.01)), aes(x)) +
  stat_function(fun = dbeta, args = list(shape1 = 1, shape2 = 1), aes(color = "Beta(1,1) prior")) +
  stat_function(fun = dbeta, args = list(shape1 = y+1, shape2 = n-y+1), aes(color = "Beta(1,1) posterio
  stat_function(fun = dbeta, args = list(shape1 = 10, shape2 = 10), aes(color = "Beta(10,10) prior")) +
  stat_function(fun = dbeta, args = list(shape1 = y+10, shape2 = n-y+10), aes(color = "Beta(10,10) poste
  geom_histogram(data = tibble(x = rbinom(n, n = n, p)/(n)), aes(x = x, y = ..density.., fill = 'likeli
plot
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Question 5

```
boy <- 251527
girl <- 241945
n<- boy+girl
p_boy <- round(boy/(boy+girl),3)
alpha <- boy+1
beta <- girl+1
probability <- pbeta(0.5, alpha, beta)
probability
```

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

Question 6

Answer6:

A noninformative prior : it would be the uniform distribution U(-1,1)

A subjective/informative prior : the common sense would be that after practice, students will make progress, so the final proportion would be higher than the initial proportion. Based on my experience, I would assume the prior distribution be the right skewed distribution with mode around 0.2-0.4 (20%-40% improvement)