

# lab\_exercise3

## 1.

We know that the Maximum likelihood Estimator of the binomial distribution is  $\hat{\theta} = \frac{x}{n}$ . And 95% confidence interval for MLE is  $(\hat{\theta} - 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}, \hat{\theta} + 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}})$ .

```
x <- 118
n <- 129
mle <- x/n
se <- sqrt((mle*(1-mle))/n)
z <- 1.96
confidence_interval <- c(mle - z*se, mle + z*se)
```

```
# MLE
mle
```

```
[1] 0.9147287
```

```
# Confidence Interval
confidence_interval
```

```
[1] 0.8665329 0.9629244
```

## 2.

```
# Add the number of success
alpha <- 1 + x
# Add the number of failure
beta <- 1 + n-x
```

```
# Calculate posterior mean
posterior_mean <- alpha / (alpha + beta)
# Calculate credible interval (quantile-based)
credible_interval <- qbeta(c(0.025, 0.975), alpha, beta)
# Posterior mean
posterior_mean
```

```
[1] 0.9083969
```

```
# Credible interval
credible_interval
```

```
[1] 0.8536434 0.9513891
```

### 3.

If we use a Beta(10, 10) prior on  $\theta$ , we interpret this as the number of success (women who aged 65+ are happy) was the same as failure (women who aged 65+ are not happy) in the previous experiment (indicating  $\theta = 0.5$ ). Compared to the Beta(1,1) prior, which is a uniform distribution and implies that all values of  $\theta$  from 0 to 1 are equally likely, the Beta(10,10) prior carries more information. This is because the Beta(10,10) prior indicates a stronger belief that  $\theta$  is around 0.5, due to the higher concentration of probability density around the center of the distribution. We can check this by plotting distribution.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.2      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.0
v ggplot2    3.4.2      v tibble     3.2.1
v lubridate  1.9.2      v tidyr      1.3.0
v purrr      1.0.1
-- 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
```

```

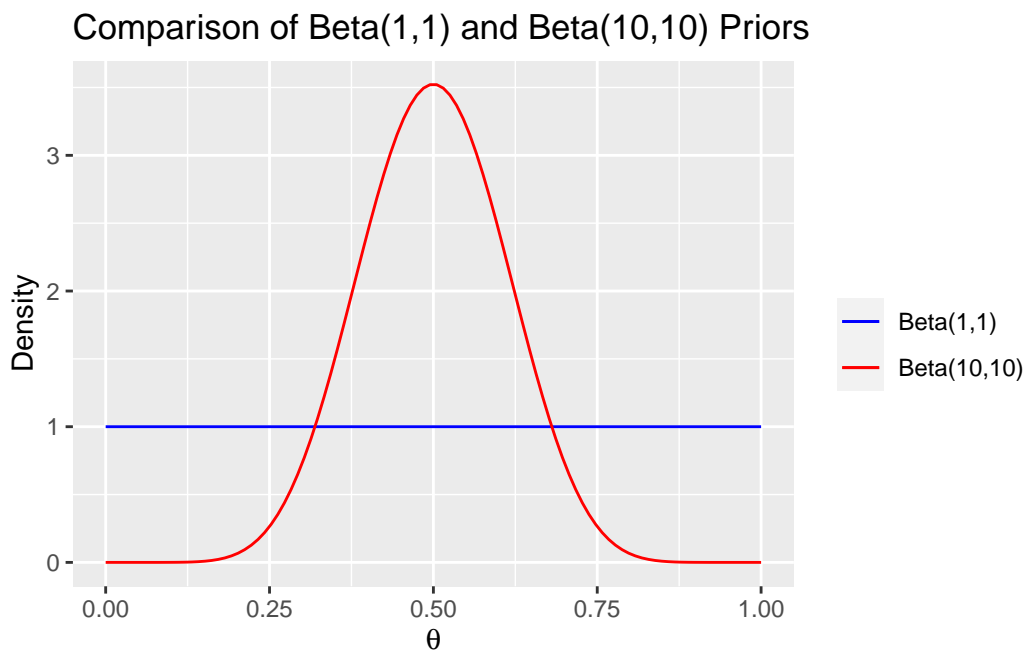
# Define the range of theta
theta <- seq(0, 1, length.out = 100)

# Compute the densities for Beta(1,1) and Beta(10,10)
density_beta_1_1 <- dbeta(theta, 1, 1)
density_beta_10_10 <- dbeta(theta, 10, 10)

# Create a data frame for plotting
df <- data.frame(theta, density_beta_1_1, density_beta_10_10)

# Plot the Beta distributions
ggplot(df, aes(x = theta)) +
  geom_line(aes(y = density_beta_1_1, colour = "Beta(1,1)")) +
  geom_line(aes(y = density_beta_10_10, colour = "Beta(10,10)")) +
  labs(title = "Comparison of Beta(1,1) and Beta(10,10) Priors",
       x = expression(theta),
       y = "Density") +
  scale_colour_manual("",
                     breaks = c("Beta(1,1)", "Beta(10,10)"),
                     values = c("blue", "red"))

```



4.

```
theta_values <- seq(0, 1, length.out = 1000)

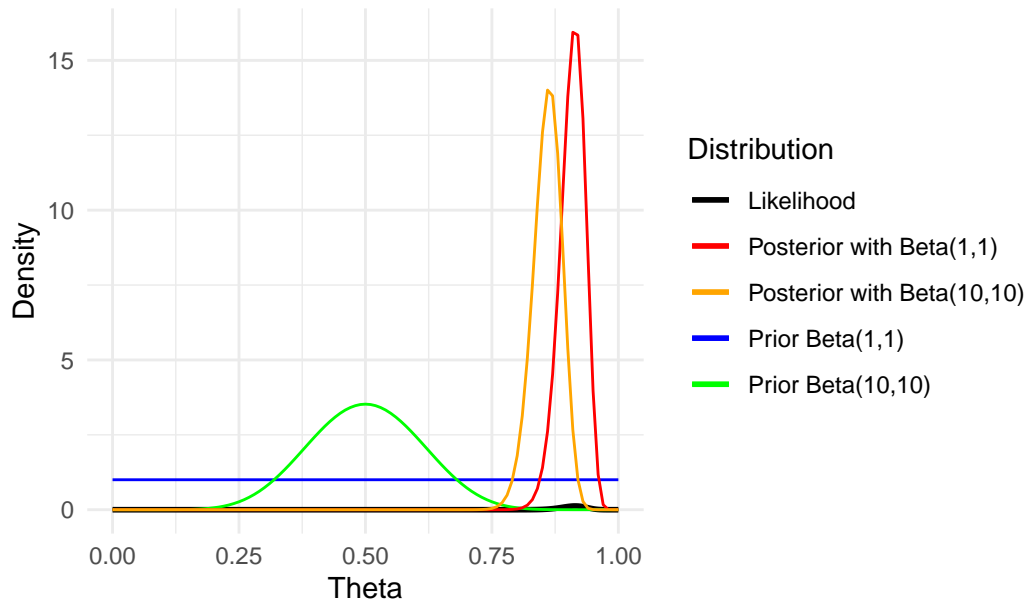
# Likelihood (Binomial)
likelihood <- dbinom(x, size = n, theta_values)

# Define the beta distribution function
beta_distribution <- function(theta, alpha, beta) {
  dbeta(theta, alpha, beta)
}

# Plotting
ggplot(data = data.frame(theta = theta_values, likelihood = likelihood), aes(x = theta)) +
  geom_line(aes(y = likelihood, color = "Likelihood"), size = 1) +
  stat_function(fun = beta_distribution, args = list(alpha = 1, beta = 1), aes(color = "Prior Beta(1,1)")) +
  stat_function(fun = beta_distribution, args = list(alpha = 10, beta = 10), aes(color = "Prior Beta(10,10)")) +
  stat_function(fun = beta_distribution, args = list(alpha = 1 + x, beta = 1 + n - x), aes(color = "Posterior with Beta(1,1)")) +
  stat_function(fun = beta_distribution, args = list(alpha = 10 + x, beta = 10 + n - x), aes(color = "Posterior with Beta(10,10)")) +
  scale_color_manual(values = c("Likelihood" = "black",
                                "Prior Beta(1,1)" = "blue",
                                "Prior Beta(10,10)" = "green",
                                "Posterior with Beta(1,1)" = "red",
                                "Posterior with Beta(10,10)" = "orange"),
                    ) +
  labs(color = "Distribution", title = "Likelihood, Prior and Posterior for Women Aged 65+",
       theme_minimal()
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
i Please use `linewidth` instead.

## Likelihood, Prior and Posterior for Women Aged 65+



We get following interpretations: 1. Likelihood suggests that around 90% of women aged 65+ shows happiness since it is centered around 90%. 2. As we can see Prior and posterior distributions, Posterior with Beta(1,1) shows similar trend to likelihood whereas Posterior Beta(10, 10) moves towards to the 0.5. We interpret this as Prior Beta(10,10) centers around 0.5 so there was already a belief that the true proportion was around 0.5, and this prior belief has pulled the posterior away from the likelihood's peak towards the center.

## 5.

```
# Sex ratio at birth in Paris (1781)
male <- 251527
female <- 241945

# Uniform prior
alpha <- 1
beta <- 1

# Posterior
alpha_posterior <- alpha + male
beta_posterior <- beta + female

# Calculate the probability that observing a male birth was less than 0.5
```

```
prob <- pbeta(0.5, alpha_posterior, beta_posterior)
```

```
prob
```

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

## 6.

Noninformative prior: A noninformative prior for  $\theta$  would be one that does not provide any additional information. It essentially says, “we do not know anything about the improvement students might experience by practicing”, which allows the data to speak for itself. An example could be a uniform distribution, such as  $\text{Beta}(1,1)$ , which suggests that any value of  $\theta$  is equally likely.

Subjective/informative prior: A subjective or informative prior for  $\theta$  incorporates specific expectations or expert knowledge about the improvement. For example, if a basketball coach believes that training typically yields a modest improvement in free-throw success, a Beta distribution such as  $\text{Beta}(10,10)$  could represent this belief. This prior suggests that the coach expects improvements to center around 50% but also indicates some degree of uncertainty or variability in that belief.