STA 6351, Report.1.14

Carson Slater *Baylor University*

1.14

Problem

Using a uniform prior for p on $[0, 1]$, construct an equal-tailed 95% credible interval for the binomial proportion p given $n = 20$ and $x = 1$. Contrast this with the frequentist intervals obtained in the text, including the interpretation of "95%".

1. Bayesian Approach: Derivation and Calculation

1.1. Prior and Likelihood

- **Data:** $n = 20$ trials, $x = 1$ success.
- **Likelihood:** Binomial $\mathcal{L}(p|x) \propto p^x(1-p)^{n-x}$.
- **Prior:** Uniform $\pi(p) = 1$ for $p \in [0, 1]$. The uniform prior is equivalent to a $\text{Beta}(\alpha = \mathbf{1}, \beta = \mathbf{1})$ distribution.

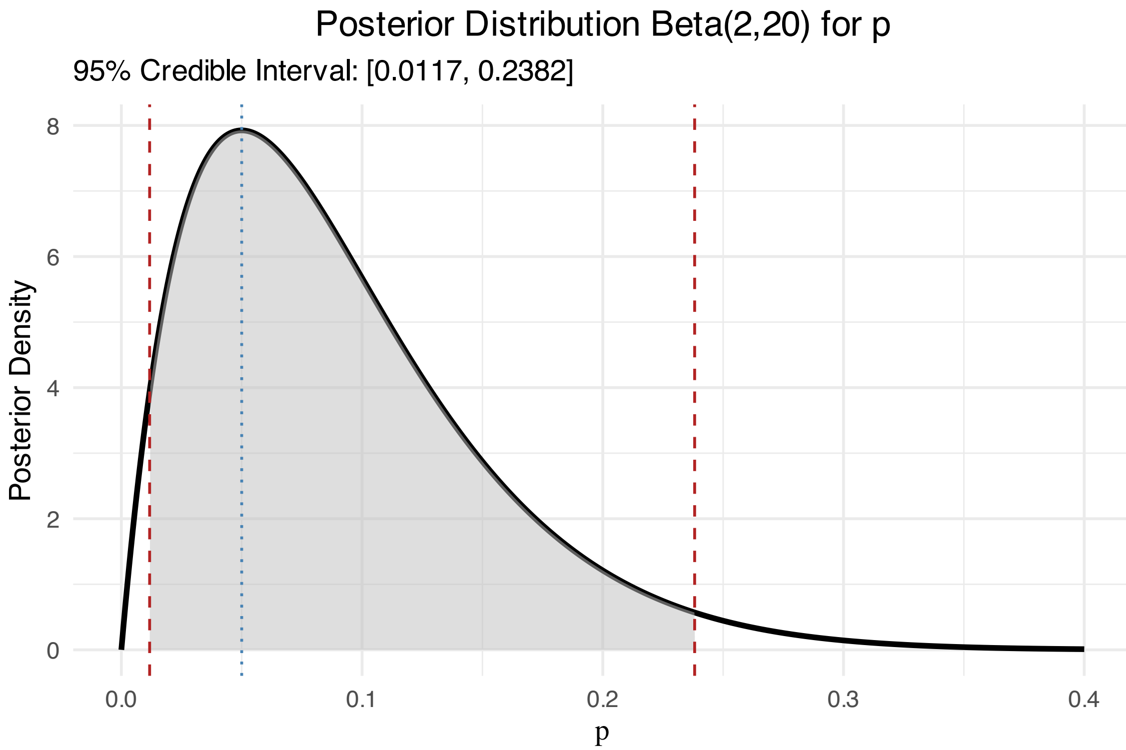
1.2. Posterior Distribution

The Beta distribution is the conjugate prior for the Binomial likelihood, meaning the posterior distribution is also a Beta distribution.

$$\pi(p|x) \propto L(p|x) \times \pi(p) \propto p^1(1-p)^{19} \times 1 = p^{2-1}(1-p)^{20-1}$$

The posterior distribution is $\mathbf{p}|\mathbf{x} \sim \text{Beta}(\alpha' = \mathbf{1} + \mathbf{x}, \beta' = \mathbf{1} + \mathbf{n} - \mathbf{x}) = \text{Beta}(\mathbf{2}, \mathbf{20})$.

Here we compute the MLE for p as 0.05. The posterior is a Beta(2, 20). From this we compute the 95% credible interval to be [0.0117, 0.2382]. See the credible interval in the figure.



All three intervals demonstrate **strong asymmetry** around the MLE ($\hat{p} = 0.05$) stretching much further to the right. The Bayesian interval, like the Logit-Wald interval, closely matches the Likelihood Ratio interval, indicating a good approximation of the skewed likelihood.

Interval Type	**Calculated Interval**	**Interpretation of "95%"**
Bayesian Credible Set	[0.0117, 0.2382]	There is a 0.95 **probability** that the true parameter p lies within this interval.
Frequentist Logit-Wald	[0.007, 0.283]	95% of such random intervals would contain the true parameter p in repeated sampling.
Frequentist Likelihood Ratio	[0.003, 0.227]	Same interpretation as above.

Code Appendix

```
knitr::opts_chunk$set(
  dev = "cairo_pdf",
  fig.width = 5,
  fig.height = 5,
  fig.align = 'center',
  echo = FALSE,
  message = FALSE,
  warning = FALSE,
  error = FALSE,
  results = 'markup'
)

# Load required libraries
library("tidyverse")
library("patchwork")
library("glue")
library("scales", warn.conflicts = FALSE)
library("extrafont")
library("tinytex")
library("knitr")
library("tidyr")
library("latex2exp")
library("MASS")
library("kableExtra")

theme_set(theme_minimal(base_family = "Roboto Condensed"))

conflicted::conflicts_prefer(
  readr::col_factor(),
  purrr::discard(),
  dplyr::lag(),
  readr::parse_date(),
  kableExtra::group_rows(),
  dplyr::select
)
# Load necessary libraries for plotting
library(tidyverse)
# Observed data
n <- 20
x <- 1

# Logit-Wald Interval (from text)
wald_L <- 0.007
wald_U <- 0.283
```

```r
# Posterior parameters (Uniform prior Beta(1,1))
alpha_prime <- 1 + x
beta_prime <- 1 + n - x

# Calculate the 95% equal-tailed credible interval
p_L <- qbeta(0.025, alpha_prime, beta_prime)
p_U <- qbeta(0.975, alpha_prime, beta_prime)
# Create a sequence of p values
p_values <- seq(0, 0.4, length.out = 500)

# Calculate the posterior density (Beta(2, 20))
posterior_density <- dbeta(p_values, alpha_prime, beta_prime)

# Create a data frame for plotting
plot_data <- data.frame(p = p_values, density = posterior_density)

# Plot the posterior density
ggplot(plot_data, aes(x = p, y = density)) +
  geom_line(linewidth = 1) +
  # Add the credible interval shading
  geom_area(
    data = subset(plot_data, p >= p_L & p <= p_U),
    aes(x = p, y = density),
    fill = "grey",
    alpha = 0.5
  ) +
  # Add the interval bounds
  geom_vline(
    xintercept = c(p_L, p_U),
    linetype = "dashed",
    color = "firebrick"
  ) +
  # Add MLE
  geom_vline(xintercept = x / n, linetype = "dotted", color = "steelblue") +
  labs(
    title = paste0(
      "Posterior Distribution Beta(",
      alpha_prime,
      ",",
      beta_prime,
      ") for p"
    ),
    subtitle = paste0(
      "95% Credible Interval: [",
      round(p_L, 4),
      ", ",
      round(p_U, 4),
      "]"
```

```r
  ),
  x = "p",
  y = "Posterior Density"
) +
theme_minimal() +
# Use an axis label that supports LaTeX formatting
theme(
  plot.title = element_text(hjust = 0.5),
  axis.title.x = element_text(family = "serif")
)
```