STA 6351, Report.1.16

Carson Slater *Baylor University*

1.16

Suppose that $X \sim \text{Bin}(n, \theta)$ but that we only observe $0 \leq x \leq u < n$, for some constant, u and not the exact value of $X = x$. The likelihood is

$$L(\theta|u) = \sum_{x=0}^{u} \binom{n}{x} \theta^x (1-\theta)^{n-x}. \tag{1.13.9}$$

In his examples, Pawitan takes x to be the number of seeds, out of $n = 100$, that germinate and supposes that one observes $u = 10$. See Figure 1.4 for a graph of this likelihood, where it is contrasted with the conventional binomial likelihood. Note that the likelihood for the present case is not regular. The MLE is 0 and there is no distribution theory that will yield a reasonable value for the standard error of this estimator. However, one can easily construct a Bayesian model for this problem:

Suppose we take $\theta \sim \text{Beta}(a, b)$ as our prior. Then the posterior is

$$\begin{aligned} \pi(\theta|x \leq u) &\propto \left[\frac{1}{B(a,b)} \sum_{x=0}^{u} \binom{n}{x} \theta^{x+a-1} (1-\theta)^{n-x+b-1} \right] \\ &= \frac{\sum_{x=0}^{u} \binom{n}{x} \theta^{x+a-1}(1-\theta)^{n-x+b-1}}{B(a,b) \sum_{x=0}^{u} \binom{n}{x} B(x+a, n-x+b)} \end{aligned} \tag{1.13.10}$$

where $B(a, b)$ is the beta function and both a and b are positive. The marginal distribution is

$$\begin{aligned} m(u) &= \int_0^1 \frac{1}{B(a,b)} \sum_{x=0}^{u} \binom{n}{x} \theta^{x+a-1}(1-\theta)^{n-x+b-1} d\theta \\ &= \frac{1}{B(a,b)} \sum_{x=0}^{u} \binom{n}{x} \int_0^1 \theta^{x+a-1}(1-\theta)^{n-x+b-1} d\theta \\ &= \frac{1}{B(a,b)} \sum_{x=0}^{u} \binom{n}{x} B(x+a, n-x+b). \end{aligned} \tag{1.13.11}$$

Note that the marginal is for u, not x. Interval estimates derived from this posterior, such as equal-tailed credible sets, have a strictly probabilistic interpretation. That is, they contain θ with the specified probability.

Use the Bayesian model with $u = 10$, and $n = 100$. Construct a 95% equal-tailed credible interval for θ. Do this for $(\mathbf{a}, \mathbf{b}) = (\mathbf{1}, \mathbf{1}), (\mathbf{3}, \mathbf{3}), (\mathbf{1}, \mathbf{9})$.

The observed data is $X \sim \text{Bin}(n, \theta)$ right-censored at u, with $n = 100$ and $u = 10$. We assume a $\text{Beta}(a, b)$ prior for θ. The posterior distribution is a **mixture of Beta distributions**:

$$\pi(\theta|X \le u) = \sum_{x=0}^{u} w_x \cdot \text{Beta}(\theta|x+a, n-x+b)$$

where the normalized weights w_x are given by

$$w_x = \frac{\binom{n}{x} B(x+a, n-x+b)}{\sum_{k=0}^{u} \binom{n}{k} B(k+a, n-k+b)}.$$

The posterior is a mixture because conditioning on $X \le u$ requires averaging over all unobserved outcomes $x = 0, \ldots, u$. Each possible x would yield a Beta posterior Beta$(x+a, n-x+b)$, and the overall posterior integrates these components with weights proportional to their marginal likelihoods:

$$\pi(\theta|X \le u) = \sum_{x=0}^{u} w_x \, \text{Beta}(\theta|x+a, n-x+b),$$

where $w_x \propto \binom{n}{x} B(x+a, n-x+b)$. The 95% equal-tailed credible intervals $[\theta_L, \theta_U]$ are computed by finding the 0.025 and 0.975 quantiles of the posterior CDF $F(\theta|X \le u)$.

Table 1: 95% Equal-Tailed Credible Intervals for θ (Given $X \le 10$)

Prior a	Prior b	Lower (0.025)	Upper (0.975)	95% CI for θ
1	1	0.0027	0.1426	(0.0027, 0.1426)
3	3	0.0325	0.1736	(0.0325, 0.1736)
1	9	0.0018	0.1258	(0.0018, 0.1258)

Appendix

```r
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 = 'asis'
)

# 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 table generation
library(dplyr)
library(knitr)
library(kableExtra)

# Data from the Bayesian analysis (n=100, u=10)
results_data <- tibble::tribble(
  ~Prior_a,
  ~Prior_b,
  ~Lower_Quantile,
```

```
  ~Upper_Quantile,
  1,
  1,
  0.0027,
  0.1426,
  3,
  3,
  0.0325,
  0.1736,
  1,
  9,
  0.0018,
  0.1258
) |>
  # Create the final interval string for display
  mutate(
    `95% Credible Interval` = paste0(
      "(",
      sprintf("%.4f", Lower_Quantile),
      ", ",
      sprintf("%.4f", Upper_Quantile),
      ")"
    )
  )

# Define column names with LaTeX/Markdown formatting for the Greek letter the
col_names <- c(
  "Prior $a$",
  "Prior $b$",
  "Lower ($0.025$)",
  "Upper ($0.975$)",
  "95\\% CI for $\\theta$"
)

# Generate the kable table
results_table <- results_data |>
  # Select and order columns, including the calculated interval string
  select(
    Prior_a,
    Prior_b,
    Lower_Quantile,
    Upper_Quantile,
    `95% Credible Interval`
  ) |>
  # Use kable for base table generation
  kable(
    format = "latex", # Use 'html' or 'latex' depending on the output format
    align = "c",
```

```
    caption = "95\\% Equal-Tailed Credible Intervals for $\\theta$ (Given $X \
    col.names = col_names,
    booktabs = TRUE, # Looks better in LaTeX/PDF output
    escape = FALSE # Allows LaTeX commands in column names and caption
  ) |>
  # Enhance the table appearance using kableExtra
  kable_styling(
    latex_options = c("striped", "hold_position"),
    full_width = FALSE
  )

# Print the resulting LaTeX table code
print(results_table)
```