R Programming: Bayesian Inference for a Binomial Proportion

2026

The Beta–Binomial is a conjugate model for a binomial proportion π:

- **Prior:** $\pi \sim \text{Beta}(a_0, b_0)$
- **Likelihood:** $f(y \mid \pi) \propto \pi^y (1-\pi)^{n-y}$
- **Posterior:** $\pi \mid y \sim \text{Beta}(a_1,\ b_1)$, where $a_1 = a_0 + y$ and $b_1 = b_0 + n - y$

1 Function 1: Computation

Open a new R script and save it as `binbetaf.R`. The function below is partially written. Fill in the missing lines (marked with `???`) using the formulas above and the reference formulas below.

Reference formulas:

$$E(\pi \mid y) = \frac{a_1}{a_1 + b_1} \qquad \text{Var}(\pi \mid y) = \frac{a_1 b_1}{(a_1 + b_1)^2 (a_1 + b_1 + 1)}$$

$$w_{\text{prior}} = \frac{a_0 + b_0}{a_0 + b_0 + n} \qquad w_{\text{data}} = \frac{n}{a_0 + b_0 + n}$$

```
binbetaf <- function(a, b, n, y, alpha) {
  # Inputs:
  # a     : shape1 parameter of the Beta prior
  # b     : shape2 parameter of the Beta prior
  # n     : number of trials
  # y     : number of successes
  # alpha : credible interval level (e.g. 0.05 for 95% CI)

  # --- Posterior parameters ---
  a1 <- ???
  b1 <- ???

  # --- Posterior mean, variance, and SD ---
  m_pos   <- ???
  var_pos <- ???
  sd_pos  <- sqrt(var_pos)

```

[0]Email: paulius.kazlauskas@evaf.vu.lt

```
  # --- Weights of prior and data in the posterior mean ---
  w_prior <- ???
  w_data  <- ???

  # --- Credible interval bounds ---
  # qbeta(p, shape1, shape2) gives the p-th quantile of Beta(shape1, shape2)
  low <- qbeta(alpha / 2,     a1, b1)
  up  <- qbeta(1 - alpha / 2, a1, b1)

  # --- Print results (already written for you) ---
  print('===================================')
  print('Parameters of posterior distribution')
  print('a posterior'); print(a1)
  print('b posterior'); print(b1)
  print('===================================')
  print('Posterior Mean and Variance')
  print('Posterior Mean');     print(m_pos)
  print('Posterior Variance'); print(var_pos)
  print('Posterior SD');       print(sd_pos)
  print('===================================')
  print('Weights')
  print('Prior Weight'); print(w_prior)
  print('Data Weight');  print(w_data)
  print('===================================')
  print('Credible interval')
  print('Lower Bound'); print(low)
  print('Upper Bound'); print(up)

  return(list(apos = a1, bpos = b1))
}
```

Once you have written the function, open a **new script** in the same working folder and run:

```
# Load the function from its file
source("binbetaf.R")

# Run with example inputs: a=1, b=1, n=100, y=20, alpha=0.05
binbetaf(1, 1, 100, 20, 0.05)

# Store the result as an object and access individual elements
posterior_dist <- binbetaf(1, 1, 100, 20, 0.05)
posterior_dist$apos   # posterior a parameter
posterior_dist$bpos   # posterior b parameter
```

2 Function 2: Plot

Open a new script and save it as `binbetaf_plot.R`. The plotting function is partially written below. Fill in the missing lines.

```r
binbetaf_plot <- function(a, b, n, y, pp) {
  # Inputs:
  # a, b : Beta prior parameters
  # n    : number of trials
  # y    : number of successes
  # pp   : legend position (e.g. "topright", "topleft")

  # Create a fine grid of pi values between 0 and 1
  pi <- seq(0, 1, 0.0001)

  # Compute the prior density at each point
  # dbeta(x, shape1, shape2) evaluates the Beta pdf at x
  y_pri <- dbeta(???, a, b)

  # Compute the posterior parameters
  a1 <- ???
  b1 <- ???

  # Compute the posterior density
  y_pos <- dbeta(???, a1, b1)

  # Set the y-axis upper limit to fit both curves
  ma <- max(y_pri, y_pos)

  # Plot the prior (red) and posterior (blue)
  plot(pi, y_pri,
       col  = "red",
       ylab = "",
       ylim = c(0, ma),
       type = "l",
       lwd  = 1)
  lines(pi, y_pos,
        col  = "blue",
        type = "l",
        lwd  = 1)
  legend(pp,
         legend = c("Prior", "Posterior"),
         col    = c("red", "blue"),
         lty    = c(1, 1))
}
```

Plotting argument reference:

- `col = "red"` : choose the colour of the line
- `ylab = ""` : suppress the y-axis label
- `ylim = c(0, ma)` : set the y-axis range from 0 to `max(y_pri, y_pos)`
- `type = "l"` : draw a line (not individual points)

- `lwd` : control the thickness of the line
- `lty` : line type — 1 = solid, 2 = dashed

Run the function from a new script:

```r
source("binbetaf_plot.R")
binbetaf_plot(1, 1, 100, 20, "topright")
```

Extension: plotting with the credible interval. Write a third function `binbetaf_plot_CI(a, b, n, y, pp, alpha)` that extends `binbetaf_plot` by:

1. Computing the credible interval bounds `low` and `up` using `qbeta`.
2. Adding vertical dashed lines at those bounds using:
   `abline(v = c(low, up), col = "black", lty = 2, lwd = 1)`
3. Updating the legend to include a `"CI"` entry with `lty = 2`.

Test it with: `binbetaf_plot_CI(1, 1, 100, 20, "topright", 0.05)`

3 Application 1: How Much of the Earth is Covered in Water?

Suppose you have a globe small enough to hold in your hands. You toss it in the air and, each time you catch it, record whether your right index finger lands on water (W) or land (L). The sequence of nine tosses is:

$$\{W,\ L,\ W,\ W,\ W,\ L,\ W,\ L,\ W\}$$

Let π be the proportion of the Earth's surface covered in water. Assume initially the same plausibility to every proportion of water.

1. What is the posterior distribution of π after the **first** observation (W)? Write down the prior, the likelihood, and the posterior, identifying a_1 and b_1.
2. What is the posterior distribution of π after **all nine** observations? Identify a_1 and b_1.
3. What are the posterior mean and standard deviation of π? Round to 2 decimal places.

R Corner. Fill in the correct values of `n` and `y` for each call, then run both functions.

```r
source("binbetaf.R")
source("binbetaf_plot.R")

# Question 1a: after first observation only
binbetaf(1, 1, ???, ???, 0.05)
binbetaf_plot(1, 1, ???, ???, "topleft")

# Questions 1b and 1c: after all observations
binbetaf(1, 1, ???, ???, 0.05)
binbetaf_plot(1, 1, ???, ???, "topleft")
```

4 Application 2: Proportion of Women in Economic Science

In order to estimate the proportion of women π among economics professors, you plan to survey a random sample. Before seeing any data, you believe π has mean 0.5 and standard deviation 0.1.

1. Determine the Beta(a, b) prior that matches your prior belief. *Hint:* set up the two moment equations $\mu_0 = a/(a+b)$ and $\sigma_0^2 = ab/[(a+b)^2(a+b+1)]$ and solve for a and b.
2. Out of 1000 people in the sample, 250 are women. Determine the posterior distribution $\pi \mid y$.
3. Calculate the weights of the prior and of the data in the posterior mean. What do they tell you about the relative influence of your prior belief versus the data?
4. Find the normal approximation to the posterior distribution. Under what condition is this approximation valid? Is it satisfied here?
5. Compute a 95% credible interval for π using the normal approximation.

R Corner. Once you have determined a and b in question 1, fill them in below.

```r
source("binbetaf.R")
source("binbetaf_plot.R")

a     <- ???   # fill in from question 1
b     <- ???   # fill in from question 1
n     <- 1000
y     <- 250
alpha <- 0.05

binbetaf(a, b, n, y, alpha)
binbetaf_plot(a, b, n, y, "topright")
```

Material collected by: Ieva Mikaliunaite-Jouvanceau