Computational Statistics & Probability

Lab 1 - Bayesian Inference

Fall 2025

Learning Objectives

By the end of this lab, you will be able to: - Construct posterior distributions using grid approximation - Query posterior distributions to answer probability questions - Compare models with different priors - Understand how priors influence inference with limited data

The Globe Tossing Experiment

In lecture, we discussed estimating the proportion of Earth's surface covered by water using a simple experiment: toss a globe and record whether your right index finger lands on water (W) or land (L).

This lab walks you through the computational implementation of Bayesian updating for this problem.

1. Understanding Grid Approximation

Use the following R code to generate a set of `samples` from which to answer questions about its distribution.

```
set.seed(212)

p_grid <-seq( from=0 , to=1, length.out=1000 )
prior <- rep( 1 , 1000 )
likelihood <- dbinom( 6, size=9, prob=p_grid )
posterior <-likelihood * prior
posterior <- posterior / sum(posterior)

samples <- sample( p_grid , prob=posterior , size=1e4 , replace = TRUE)
```

a) How much posterior probability lies below $p = 0.25$?

```
# ANSWER 1a
sum(samples < 0.25) / length(samples)
```

```
## [1] 0.0025
```

b) How much posterior probability lies above $p = 0.75$?

c) How much posterior probability lies between $p = 0.25$ and $p = 0.75$?

d) 25% of the posterior probability lies below which value of p?

2 Prior Predictive Simulation

Before we fit models with real data, let's check what our priors predict.

a) Simulate predictions from the flat prior

The flat prior says all values of p (0 to 1) are equally plausible. What data should we expect if we truly believe this?

```r
# PARTIAL ANSWER FOR 2a

# Set random seed to 212
set.seed(212)

# Set number of simulations to 100
n_sims <- 100

# Draw 100 samples from [0,1]
p_samples <- runif(n_sims, 0, 1)  # Flat prior: any p equally likely

# For each p, simulate 9 globe tosses

# TODO: YOUR CODE HERE

# Visualize: What do you expect to observe?

# TODO: YOUR CODE HERE
```

b) Simulate predictions from the informative prior

```r
# PARTIAL ANSWER FOR 2b

# Set random seed to 212
set.seed(212)

# Set number of simulations to 100
n_sims <- 100

# Draw 100 samples from [0.5,1]

# p_samples_informed <- # TODO: YOUR CODE HERE

# For each p, simulate 9 globe tosses

# TODO: YOUR CODE HERE

# Visualize: What do you expect to observe?

# TODO: YOUR CODE HERE
```

3 Build Your Own

Suppose the globe tossing experiment yielded the following sequence of 15 observations,

$$[W, L, W, W, L, L, W, L, W, L, L, W, L, W, W]$$

where W denotes 'water' and L denotes 'land'.

Using grid approximation, construct the posterior using: - Grid approximation with 1000 points - A flat prior - The binomial likelihood

a) Write the code (i.e., modify the example from Part 1)

```r
# Construct the posterior distribution using grid approximation.

# STEP 1: Create the grid
# p_grid <- # TODO: YOUR CODE HERE

# STEP 2: Define the prior
# prior <- # TODO: YOUR CODE HERE (flat prior)

# STEP 3: Compute the likelihood
# likelihood <- # TODO: YOUR CODE HERE

# STEP 4: Compute the posterior
# posterior <- # TODO: YOUR CODE HERE

# STEP 5: Normalize
# posterior <- # TODO: YOUR CODE HERE

# STEP 6: Draw samples
# set.seed(212)
# samples_a <- # TODO: YOUR CODE HERE
```

b) Using grid approximation, construct the posterior distribution with a prior that is 0 below $p = 0.5$ and otherwise constant.

c) Explain the difference between model (a) and (b).

d) Which prior, (a) or (b), is better? Explain why.

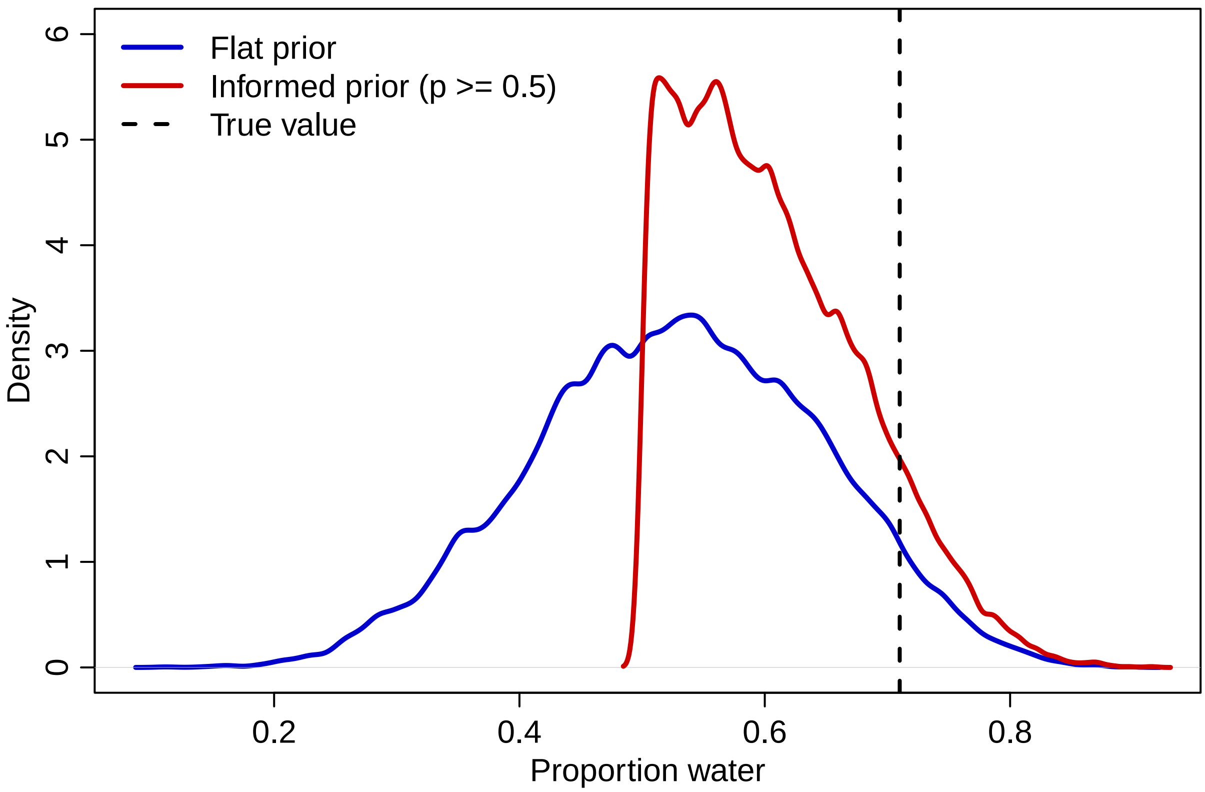
e) Compare the two posteriors visually

```r
#
dens(samples_a,
     xlab = "Proportion water",
     ylab = "Density",
     ylim = c(0, 6),
     col = "blue3",
     lwd = 2.5)

dens(samples_b, add = TRUE, col = "red3", lwd = 2.5)

abline(v = 0.71, lty = 2, lwd = 2)

legend("topleft",
       legend = c("Flat prior", "Informed prior (p >= 0.5)", "True value"),
       col = c("blue3", "red3", "black"),
       lty = c(1, 1, 2),
       lwd = c(2.5, 2.5, 2),
       bty = "n")
```



4 Posterior Predictive Check

Now that we've seen the data (8 waters in 15 tosses), does our model make sensible predictions?

```r
# HINT
set.seed(212)

posterior_predictions <- rbinom(n = 1e4, size = 15, prob = samples_a)

# Visualize posterior distribution

# How often does the model predict exactly 8 waters?

# Is 8 in a reasonable range? (Requires your interpretation and analysis)
```