

# Week 8 Lab problems

EEB 429

Bhaskar Kumawat

Submit all your answers in a single .Rmd file (along with the knitted html).

## A. Calculating and maximizing likelihood

Aliens come to earth and hand you a mysterious coin with a certain probability (say  $p$ ) of turning up a head when flipped. Your task is to figure out the value  $p$  using the maximum likelihood method. The steps for doing so are as follows:

1. Write a function called `LogLikCoin` that takes in a vector of 1s and 0s (call it `test_data`), and a fractional number between 0 and 1 (call it `test_p`) as an input. In the vector, a 1 denotes a head and a 0 denotes a tail. This function then calculates the log-likelihood of observing that particular sequence of heads/tails given the coin has a probability `test_p` of flipping a head.
2. Write another function called `MaximizeCoinLikelihood` that takes in a vector of 1s and 0s (call it `data`). This function,
  - a. Creates a vector called `p_vals` that contains values from 0 to 1 with a step size of 0.001. Then,
  - b. It uses the `LogLikCoin` function to calculate the value of the log likelihood for the given data for each value of  $p$  in `p_vals`. These values must be stored in a vector. Then,
  - c. The function finds the maximum value of log likelihood in the resulting vector, and also the value of  $p$  that gives the maximum value (the `which()` function might be useful here). Finally,
  - d. It plots the vector of log-likelihoods against the vector `p_vals` and creates a vertical line at the value of  $p$  that gives the maximum value of log-likelihood. After plotting, it also returns the maximum likelihood estimate of  $p$ .
3. When you flip the coin a 100 times, you get the following sequence. Use your function to calculate the MLE for  $p$  with this data.

```
aliencoinflips <- c(0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1,
0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,
0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
1, 1, 1, 0, 0, 0)
```

For full credit, your file should include the functions for parts 1 and 2, the function `MaximizeCoinLikelihood` should plot the values and return the MLE, and the correct MLE obtained using your function in part 3 should be included.

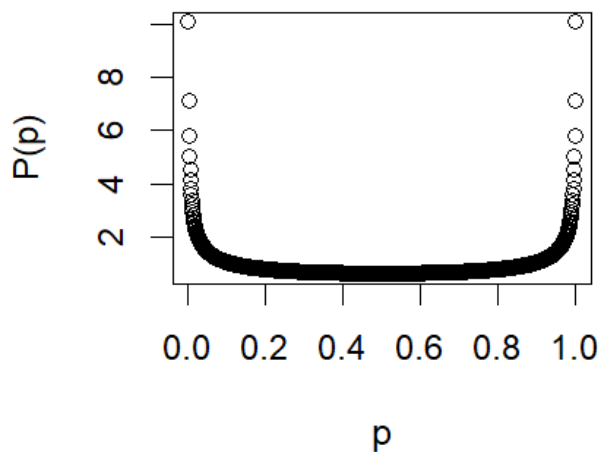
## B. Bayesian estimation of the distribution

Now, instead of using the maximum likelihood function, you decide to estimate the value of  $p$  using Bayes estimation. The maximum likelihood estimate is a special case of Bayesian estimation with a *uniform prior*—i.e., you assume all values of  $p$ , from 0 to 1, are equally likely.

Instead, now we assume that the aliens are tricking us with the coin and that  $p$  is more likely to be close to 0 or 1. You assume that the  $P(p)$  can be calculated using a function `prior()` defined as below,

```
prior <- function(test_p){  
  return(dbeta(test_p, 0.5, 0.5))  
}
```

This prior looks something like,



Recall that,

$$P(p|Data) = (P(Data|p) * P(p)) / P(Data)$$

Let us ignore the  $P(Data)$  term for now, and note that  $P(Data|p)$  is just the likelihood of the data given  $p$  (and you already have a function to calculate its logarithm!).

1. Write a function called `PosteriorEstimate` that takes in a vector of 1s and 0s (call it `data`). This function,
  - a. Generates a sequence called `p_vals` from 0 to 1 with a step size of 0.001.

- b. At each value of  $p$  in `p_vals`,
    - i. Calculates the Likelihood (NOT log likelihood) for the given data,
    - ii. Calculates the prior probability at the value of  $p$  using the function defined earlier, and
    - iii. Multiplies these numbers to get the posterior probability at that value of  $p$ , and stores these in a vector.
  - c. Plots the calculated posterior distribution against `p_vals`.
  - d. Returns the value of  $p$  that gives the maximum posterior probability.
2. Use the same data as in part A.3 to find the value of  $p$  that has the highest posterior probability given the observed sequence.

For full credit, your file should include the function for part 1, the function should plot the requested values and return the maximum posterior estimate, and the correct estimate of  $p$  obtained using your function in part 2 should be included.