CS F320 Assignment 1

Abhinav Sukumar Rao[1], Pratyush Banerjee[2], and Rohan Daniel[3]

[1]2018A7PS0172H
[2]2018A7PS0312H
[3]2018A7PS0584H

1 Introduction and Overview

This is our report on the Bayesian estimator for predicting coin tosses of a biased coin. We have generated 160 data points and have compared the results of two methods of training, featuring plots and GIF for the same.

2 Algorithm Implementation and Development

2.1 Data Generation

Data was generated by the `scipy.stats.math.bernoulli` in the notebook.

2.2 Bob's method: One by one

For every sample, we update our a and b as follows:

$$m = m + (x == 1)$$

$$a = a + (x == 1)$$

$$b = b + n - m$$

where,
n is the number of samples and
m is the number of positive results i.e. getting a head

In order to prevent problems with calculating the beta distribution for large values of a and b using the `scipy.special.gamma` function mentioned in the assignment, we have derived recurrence relations in a and b for the coefficient in the beta distribution:

$$\beta(x, a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1}$$

where,

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

The recurrence relation(s) are as follows:

$$\beta(a, b) = \frac{(a+b-1)}{(a-1)}\beta(a-1, b)x$$

and,

$$\beta(a,b) = \frac{(a+b-1)}{(b-1)}\beta(a,b-1)(1-x)$$

While this is not necessary for the purpose of this assignment, we can use the following function for much larger number of iterations, when we sample one-by-one.

2.3 Lisa's Method: All at once

Lisa's method involved just one computation, and for 160 datapoints, `scipy.special.gamma` did not overshoot over the maximum value possible, as a result, we just computed the posterior distribution:

$$\beta(x, a+m, b+n-m)$$

using the aforementioned function. It was much faster than Bob's method but can't scale up for really large values of n.

3 Computational Results

We notice that both Bob's and Lisa's method yield exactly the same result:

$$\mu_{bob} = \mu_{lisa} = p(Head) = 0.11$$

This is because all the terms that make up the posterior distribution are present at the end of both methods.
Here is the initial and final states for both methods. We have also attached a gif of Bob's method outside of this report.

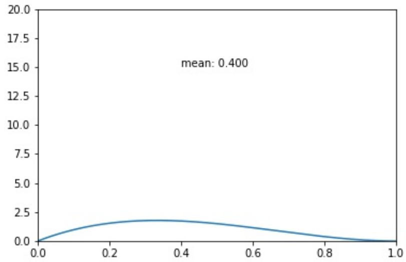

Figure 1: Prior

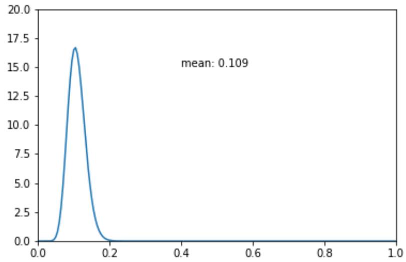

Figure 2: Posterior

4 Q and A

(i) For values of the order of 10^5 we see a very sharp and huge peak centered around almost exactly the true probability of getting head.

(ii) The beta distribution had a very convenient posterior for a variable that followed bernoulli distribution (i.e.) a coin toss. Calculating the posterior for Gamma and Pareto might lead to more complex equations that can be hard to deal with. Additionally, for Gamma and Pareto distributions, making such a recurrence as mentioned above may prove difficult