

CSE 4309 - Assignments - Optional Bayesian Estimation Assignment

This is an optional assignment that provides no extra credit, and that will not influence in any way the semester grade.

Task 1 (100 points)

In this task, you will estimate a probability distribution using a Bayesian approach.

Arguments

You must implement a Matlab function or a Python executable file called `bayes_estimation`. In particular, your function can be invoked as follows:

```
bayes_estimation(<text_file>)
```

If you use Python, just convert the Matlab function arguments to command-line arguments. The argument `text_file` will be the name a text file, saved locally on the computer. This text file will contain text whose characters are only 'a' and 'b' (don't worry about how your code behaves when that assumption is violated). The size of the file is not restricted.

Output

Each character in the text is independent of all other characters, and it is a random variable following a specific (but unknown) probability value $p(c = 'a')$. Obviously, $p(c = 'b') = 1 - p(c = 'a')$.

Let $p(c = 'a') = m$. You are given the following prior for m :

- $p(m = 0.1) = 0.9$
- $p(m = 0.3) = 0.04$
- $p(m = 0.5) = 0.03$
- $p(m = 0.7) = 0.02$
- $p(m = 0.9) = 0.01$

Obviously, all other possible values of m have probability 0. Your task is to compute the posterior probabilities of m , given the data in the text file.

At the end, your program needs to report the posterior distribution of m given the data. The program should also print $p(c = 'a')$, which can be computed using the sum rule, based on the posterior distribution of m . The program output should follow EXACTLY this format:

```
p(m = 0.1 | data) = %.4f
p(m = 0.3 | data) = %.4f
p(m = 0.5 | data) = %.4f
p(m = 0.7 | data) = %.4f
p(m = 0.9 | data) = %.4f
p(c = 'a' | data) = %.4f
```

Your program output should consist of only those lines, nothing else. Note that, in the above output sample, `%.4f` is a placeholder for a number with four decimal digits, following the Java and C `printf` conventions.

In your answers.pdf document, provide the complete output of your program when given [text1.txt](#) as the input test file.

[CSE 4309/6363/6364](#) - [Assignments](#) - Assignment 2b