Math329: Exam 04 (Out of Class)

NAME:

Please answer the following questions and bring your solution to the in-class exam. You may consult any class notes you would like, but may not search the internet or discuss the question with anyone else. You may typeset your solution in LaTeX, but this is not required. These will count for 50% of the exam.

1. Consider a random variable X that models whether a line of text uses the word *the* or not (it's one if it does and zero otherwise). Describe this random variable with a Bayesian model.

2. Make a reasonable guess for the hyperparameters for the model in question 1, given that a line of text has about 10-15 English words in it and *the* makes up about 4% of all the words in a given text.

3. Grab a novel off of your bookshelf and thumb through pages 80 to 100, counting how many times the first line of a page includes the word *the*.[1] Make sure to tell me the book you are using and the total count.

[1] If you don't own a novel, a very sad situation, you can borrow one from a friend or the library.

4. Compute the MLE, MME, and Bayes estimators based on your model from question 1, the hyperparameters from 2, and the data from 3. You do not need to re-derive these estimators but should make it clear what formulas you are plugging in to.

5. Now, count the number of times the word *the* is used in the first line of your novel for pages 101 to 150.[2] Which estimator from question 4 is the closest to this value?

[2] This should take not more than a couple of minutes. You do not need to write down each number; just count!