

DSL201: Statistical Programming

Dr. Anil Kumar Sao

Assignment 1

Deadline: 24/01/24 11:59pm

Instructions for Submission: You can submit your solution as a Jupyter Notebook/ Matlab file with comments and discussions on the results obtained in each step

1. Follow Standard Report Format: Include sections like Introduction, Data, Methodology, Results, Discussion, and Conclusion.
 2. **File Naming Convention:** Adhere to the specified naming convention for each file you submit (e.g. RollNumber_FirstName_Asg1).
 3. Refrain from using zip files. If necessary, submit multiple files.
 4. Include comments in the code explaining the logic and any assumptions made.
 5. Include References: Cite any external sources or references used in your assignment.
 6. Code Quality: Ensure your code follows best practices and is well-organized and **avoid plagiarism** as a plagiarism check will be conducted.
 7. Be aware that late submissions are not permitted; ensure timely submission.
-

1. A publishing firm wants to develop special printing machines for English. For this, they need to determine the probability of occurrence of specific letters and words. You are given two large text files (file A and file B).
 1. Determine the probability of each alphabet in the English language. Uppercase and lowercase alphabets are considered the same. Other characters like whitespace, punctuation, and numerals are to be omitted. List the top ten occurring alphabets. Is the result the same for file A and file B?
 2. The measure of uncertainty is determined by its entropy. Entropy should be computed as

$$H = \sum (-p_i \log p_i)$$

where p_i is the probability of event i. If we consider the occurrence of alphabets in English as events of interest, determine the entropy. In other words, determines the uncertainty of the alphabet in the English language. Use the results from the previous question.

3. How would the entropy change if all alphabets were equiprobable?

2. Consider a discrete random variable R with 16 states, denoted as r_0, r_1, \dots, r_{15} . The probabilities associated with each state are as follows:

[0.50, 0.25, 0.125, 0.0625, 0.03125, 0.015625, 0.03125, 0.015625, 0.0078125, 0.00390625, 0.001953125, 0.00390625, 0.001953125, 0.0009765625, 0.00048828125, 0.000244140625]

Your task is to write a program to transform the random variable R into a new random variable S in such a way that the probability mass function (PMF) of S is uniform.

Provide the code and explanation of your transformation method.

NOTE: Don't use the in-built function. Code from scratch.

3. Generate random numbers which has pdf shown in Figure 1.

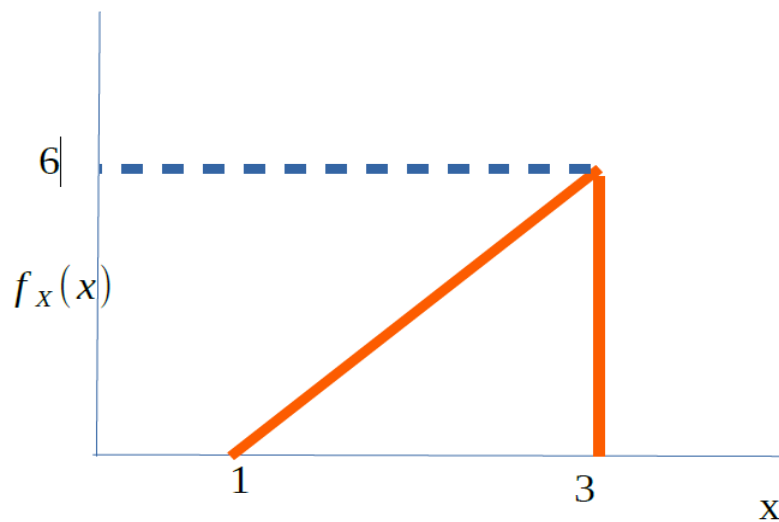


Figure 1: Figure for question 3

NOTE: Don't use the in-built function. Code from scratch.

BONUS QUESTION

4. Choose any grayscale image of size 256x256:
1. Compute the histogram of the image and adjust the bin size for the histogram to observe the impact on the visualization. Apply histogram equalization to the image and visualize the original and equalized histograms.
 2. Utilize the grayscale values of the chosen image from the previous problem. Apply a random variable transformation to the gray level values to obtain a uniform PMF.

Note: Choose the programming language of your choice and do not use the inbuilt function.