Communications and Information Engineering Program
Probability and Stochastic Processes (CIE 327)
Project 1 (Random Variables)

STATISTICAL ANALYSIS OF TEXT FILES

It is required to create a GUI-based tool that allows a user to add any text file and results in some statistics related to the text written in the file. The GUI can be built using Matlab App Designer or any other software package.

GUI Description

The GUI should do the following:

1) Allow the user enter the path of a text file (.txt), in which the text is written in English letters.
2) Based on the added file, the GUI should display the following:
   - A plot showing the probability of each of the 26 English characters (case insensitive).
   - Display the most repeated 5 letters and the numbers of their occurrences in the file.
   - If a random variable X is defined as a numerical mapping to the English characters, such that $a \rightarrow a, b \rightarrow 2, \cdots, z \rightarrow 26$, the GUI should display:
     - Plot the PMF and the CDF of X.
     - The mean of X.
     - The variance of X
     - The skewness[1] of X
     - The kurtosis[2] of X

(Bonus) Find the probability distribution that most fits the produced PMF. Justify how and why it fits most. Note that this should be adaptive and dependent on the text file.

Testing your GUI

Test your GUI for the included text file (Sample Text.txt) and report the resulting output. However, your code and GUI will be tested for other text files. So, you must make sure the produced GUI works without errors on arbitrary text files.

Deliverable

Deliver the following:

1) An **executable file** for the GUI
2) All the **source codes** (**.m** files)
3) **Equations** used to define the mean, variance, skewness and kurtosis.
4) **Screenshots** for the output of the GUI for the test file.
5) A **5-minutes comprehensive recorded video** showing your functioning GUI and your comments on the results.

[1] Search, find and write the definition of skewness.
[2] Search, find and write the definition of kurtosis.

General Instructions & Grading Criteria

Instructions

1) This is an individual project.
2) Reports are not to be shared with others.
3) Any copied reports, either fully or partially, will receive 0 points. This applies to both the original and the copy.
4) Late submission will be penalized at the rate of 10% per day for a maximum of 5 days, after which no submissions will be considered.

Grading Criteria

Grading of each part will depend on:

- **60%:** Completeness and correctness of the deliverable.
- **10%:** Clarity of the GUI design and ease of use.
- **20%:** Report writing and organization.
- **10%:** Comprehensiveness and clarity of content in the recorded video.