Matlab Project 5: Analyzing Data and Using the Poisson Distribution

You are going to examine data using multiple techniques and try to answer some questions. This assignment is due **Sunday, November 14, 2021,** at ***11:59 PM***. It will be turned in electronically using Canvas.

Project To-Do List Summary

To summarize the long description below, do these following steps:

1. Write a CalcRelFreq function that
    a. Has one input: a data vector
    b. Has one ouput: a PMF calculated using the relative frequency approach
        i. Where the range of the PMF is the values 0, 1, …, max(inputData)
2. Write a CalcPoissonPMF function that
    a. Has two inputs: α and a maximum value
    b. Outputs the Poisson PMF calculated at points 0,1,…, maximum value
3. A main script "LastnameFirstname_Project_5.m" that
    a. Inputs the data file "Project_5_Data.mat"
    b. Calculates two possible PMFs derived from that data
        i. The first PMF is found using the CalcRelFreq function
        ii. The second PMF is found using the CalcPoissonPMF function, where α if found from the command "mean(userData);"
    c. Plots both PMFs on a single bar graph
4. A file LastnameFirstName_Project5Answers.pdf" that answers four questions

Introduction

You are working for a telecommunications company and are helping with a project to upgrade their base stations. You have been monitoring base station traffic, and the data is included in the file "Project_5_Data.mat", posted to Canvas. Each sample of that data file is an observation of the number of users that connected to the base station over a **5 second interval**. You will estimate the PMF of the number of users connected to the base station using two techniques: relative frequency and parametric.

Part 1: Estimating the PMF via Relative Frequency

First, assume that the range of the number of users that can connect to a base station is an integer value between 0 and the largest observation you have made. Create a function called CalcRelFreq that takes the input of userData and outputs an estimated PMF using the relative frequency approach. **Hint:** be careful that the range starts at zero and the index starts at the 1 (i.e., the value of the range+1).

Part 2: Estimating the PMF via the Parametric Approach

Note that if we *assume* the data is from a Poisson distribution, we only need the parameter α to find the PMF! Therefore, in a *parametric approach* we can just estimate the parameter α and then calculate the PMF from that one value. Create a CalcPoissonPMF function with two inputs: α (use a variable name other than alpha, say alphaHat or alphaVal or something like that) and a maximum value (you need to specific the maximum value as technically the range of the Poisson distribution is infinite!). The output of the function should be the Poisson PMF evaluated at 0,1,…,maxVal.

Hint: the Matlab function "exp(x)" evaluates e^x.

Important note: you may **NOT** use the poisspdf function included in Matlab. You must write this function yourself.

Part 3: Comparing PMFs

Recall that α is the average of the distribution. In your main Matlab file called "LastnameFirstname_Project_5.m" load the data and estimate α from the output of the command "mean(userData);". Then use the functions you created in parts 1 and 2 to obtain two different estimates of the PMF of the data. Note: both PMFs should have the same range – use the same procedure to find your maximum value in both PMFs**! Plot the estimated PMFS side by side in a bar graph.** Make sure to add a title and label your axes. Use the "legend" command to label each dataset.

Bar graph hint:

1. Say you have two $N \times 1$ vectors (i.e. each consists of single column of data with N rows) called vec1 and vec2 and you want to plot them side-by-side
    a. Important note: since these are PMFs they **have the exact same range x**. To repetitively reiterate: this is very important
2. You can plot them together with the command: "bar(x, [vec1, vec2]);"
3. Note: if they are not both column vectors this **will not work**
    a. To turn a row vector vec1 (i.e. the vector vec1 consists of a single row of data with N columns) into a column vector use the command "vec1 = vec1.';"

Part 4: Final Questions

Your boss hands you a specification that states that the base station will drop users if *more than* 10 users connect in a **one second interval**. He wants to know the probability that the base station will drop users.

In a document named "LastnameFirstName_Project5Answers.pdf" answer the following two questions:

1. Using the PMF obtained from the relative frequency approach, do you think you could give your boss an answer?
2. Using the parametric approach, what is your new estimated α for the question your boss asked you?
3. **Using this new α**, what is the probability that the base station will drop users?
4. What benefits do you see of using the parametric approach versus the relative frequency approach?

Comment all of your code. Comments will be a portion of your grade.

Turn in checklist:

4 separate files:

1. Main script called "LastnameFirstname_Project_5.m" (using your last name and first name)
2. A function file called "CalcRelFreq.m"
3. A function file called "CalcPoissonPMF.m"
4. A PDF called "LastnameFirstName_Project5Answers.pdf".

5. Please **zip** all files together (i.e. compress the files into a single .zip file called LastnameFirstname_Project_5.zip). and submit **electronically** using Canvas. The submission will **close at 11:59**, so please turn this in on time! Note: please do not .rar the files – make sure to use zip files.