Matlab Assignment 3: Probability of Events

I've given you a datafile on Canvas called ECE2523_Project3_Data.mat. You are going to manipulate this dataset. This assignment is due **Friday, September 24, 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 main script called "LastnameFirstname_Project_3.m" that
    a. Loads a data file with site survey information
    b. Calls the function "CalcPowerDistribution" to calculate the desired event probabilities
2. Write a function "CalcPowerDistribution" that
    a. Takes a data matrix of potential energy output
    b. Finds the number of days each type of renewable energy farm output meets on of the three possible events: fails to meet specifications, meets specifications, or exceeds specifications
    c. Divides that matrix by the total number of days to find the probability of each event
    d. Returns that probability matrix as the output of the function

For further description, see the individual sections. **Read each section carefully!!**

Part 1: Loading the data

Create a new script in Matlab. Name it "LastnameFirstname_Project_3.m". Save this file, and in the same directory download the file "ECE2523_Project3_Data.mat" from Canvas. At the very top of your "LastnameFirstname_Project_3.m" script, put your name and the date as comments. Below those comments, on a single line, use the command "*load ECE2523_Project3_Data;*". This will load the data stored in that file into your workspace.

Hint: If Matlab can't find the data file, then that means that you didn't save it in the same directory (i.e. file folder) as your script is saved.

Part 2: Manipulating the data

Introduction

You work for a renewable energy company. Your company conducted a site survey for its next project, which is either going to be a new solar farm or a new wind farm. You have the data from this survey, stored in the variable *Site1Data*, which is contained in the file *ECE2523_Project3_Data*. The survey data consists of an $N \times 2$ matrix where each entry consists of the megawatt hours (MWh) that a renewable power farm *would have* generated on the day of observation. The first column of the matrix *Site1Data* is the MWh a wind farm would have generated at the site on each day, while the second column is the MWh a solar farm would have generated on each day. As these are renewable energy sources, their power output is not constant – and the solar and wind options will produce different power outputs! **Your company wants to know which choice would be better: to install a solar farm or to install a wind farm.**

Note that each observation is a particular outcome. However, there are many different outcomes! Consequently, it is difficult for the company to *make a decision* based on the raw observation data. As such the company wants **you** to convert these observations to more informative *events* of interest and find the probability of each event.

Events of Interest

The renewable energy farm has a specification that it must produce *at least* 3 MWh of energy in a day. However, if the energy farm produces *more than* 6 MWh of energy then the company will be able to sell the excess energy for a profit. Therefore, the company is interested in three events. First, the event that the power plant produces (strictly) less than 3 MWh of energy. In this case the power plant does not produce enough power to meet requirements. The second event is if the power plant produces between 3 and 6 MWh of energy. This event is the case where the company meets its specifications. Finally, the third event is when the power farm exceeds its specifications, producing (strictly) more than 6 MWh.

How to Access the data

The variable *Site1Data* is an $N \times 2$ matrix, which means that it has N rows and two columns. Each row is an estimate of the energy produced for a particular day. You will have to learn to access values in matrices for this project. The format for accessing data in a matrix is: "variableName(row, column)". Remember that **for this project** the row corresponds to the day and the column corresponds to the farm type. Some examples follow.

Example 1: the command "Site1Data(3,1)" would return the amount of energy (in MWh) that a *wind farm* (because we addressed the first column) would have produced on the *third day* (because we addressed the third row).

Example 2: the command "Site1Data(5,2)" would return the amount of energy (in MWh) that a *solar farm* (because we addressed the second column) would have produced on the *fifth day* (because we addressed the fifth row).

Example 3: the command "Site1Data(5,3)" would return **the error***: Index in position 2 exceeds array bounds (must not exceed 2).* This is because *Site1Data* is an $N \times 2$ matrix with two columns, and we tried to get the data in the third column (**which doesn't exist!**).

Task: Calculating the Probability of the Events

In a *new file* named "CalcPowerDistribution.m" make a function whose first line is "function [outputDist] = CalcPowerDistribution (siteSurvey)", where siteSurvey is a data vector (with **any length**).

Recall from the relative frequency definition of probability, that if there are N samples and a number a occurs m times, then the relative frequency of a is $\frac{m}{N}$. Your function should be able handle any value of N (hint: check out the "length()" function).

Therefore, your function must calculate **three probabilities** for **two different farm types**. Specifically, the probability of the site producing less than 3 MWh, between 3 and 6 MWh, and more than 6 MWh. Consequently, there will be many values of a that must be accounted for! You will need to use conditionals such as '<' and '>' to find the number of days for each event. You can then divide this count

by the total number of days to get the desired probabilities. The output probabilities must be assigned to the variable "outputDist". **Note that outputDist will be a *matrix* with 3 rows and 2 columns.**

In your main file (i.e. "LastnameFirstnameMatlab_Project_3.m") you will call the CalcPowerDistribution function to calculate the distribution of energy output for the variable *Site1Data*.

There are MANY ways to program the function CalcPowerDistribution – I can think of several right off the top of my head (and I will give you three different ways in my solutions!). If statements and for loops can make this a lot easier, but you don't have to use them. Therefore, write the function so that it mimics how you would solve this problem if you had to solve it by hand. As long as you get the correct answer, you will get full points – I'm not looking for a particular solution.

Part 3: Answer Questions

Write a brief report about the results of your data analysis. Please answer the following questions:

1. What *do you think* the sample space of the experiment is?
2. Make a table of the events and associated probabilities for both the wind and solar farm
    a. This table should have 6 entries
3. Which farm type would you recommend your company invest in? Why?

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

Turn in checklist:

2 separate files:

1. Main script called "LastnameFirstname_Project_3.m" (using your last name and first name)
2. A function file called "CalcPowerDistribution.m"
3. A final report called "LastnameFirstname_Project_3.pdf"

Please **zip** all files together (i.e. compress both files into a single .zip file called LastnameFirstname_Project_3.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.