Indian Institute of Technology Mandi
भारतीय प्रौद्योगिकी संस्थान मण्डी

IC252-Data Science 2

Assignment– 03

General instructions:

- Utilize Python programming language for implementation.
- Ensure the program is well-documented to enhance comprehension.
- Employ functions and loops for efficient code organization.
- Implement error handling to manage invalid inputs or unexpected scenarios.
- Optimize the code for performance and readability where applicable.

Question 1: The exponential distribution is ideal to model waiting times (e.g, time to failure of sensors in a system, flood occurrence, etc.). We will be using the exponential distribution to model time to the arrival of the next confirmed case of Covid-19 in India. Based upon data of confirmed cases from the source `https://www.covid19india.org/`, between 17th April 2020 and 23rd April 2020, there were on average 1373 confirmed cases per day, i.e. on average around 57 cases per hour.

(a) Write a program to plot the probability density function of the wait time for the next Covid-19 confirmed case, where the X axis is the wait time in hours and Y axis is the probability density.

(b) Write a program to find the probability of the wait time for the next Covid-19 confirmed case to be less than or equal to 1 minute (Hint: convert minutes into hours before using it in the cumulative density function).

(c) Write a program to find the probability of the wait time for the next Covid-19 confirmed case to be between 1 minute and 2 minutes.

(d) Now, write a program to find the probability of the wait time for the next Covid-19 confirmed case to be more than 2 minutes.

(e) Suppose the average number of cases per hour doubled. Write a program to find the probability of wait time for the next Covid-19 confirmed case to be between 1 minute and 2 minutes.

Question 2: A course instructor has brought a lie detector to check how many of his students lie. Let $X =$ number of times the lie detector buzzes for a student. He randomly chooses 50 students for the

test and the following information was obtained: $X = 0$ for 2 students, $X = 1$ for 11 students, $X = 2$ for 23 students, $X = 3$ for 9 students, $X = 4$ for 4 students, $X = 5$ for 1 student.

(a) Plot the PDF and CDF.

(b) Simulate such an experiment in python (for $n = 50, 500$ and 5000 times) and calculate the average number of lies a student tells and the standard deviation. Compare it with the theoretical values. (Calculate theoretical values using python)

(c) Pick one value of n, repeat n experiments 1000 times (try playing with this value). Plot a histogram of these mean values. What is the mean and variance of the sample means? Does the histogram seem to follow a type of distribution? (CLT foreshadowing). Can you comment on the following statement -"The sample mean is a Random Variable."

Question 3: A manufacturing company produces widgets, and the weights of these widgets follow a normal distribution with a mean weight of 150 grams and a standard deviation of 10 grams. The quality control team is responsible for ensuring that the weights of the widgets meet certain specifications. Simulate 100 such samples in python and answer the following questions.

(a) What is the probability that a randomly selected widget weighs less than 140 grams?

(b) The company has a policy that only the top 5% of the heaviest widgets will be shipped as premium products. What is the minimum weight a widget must have to qualify as a premium product?

(c) If the company wants to set a weight limit so that only the lightest 10% of widgets are considered defective and need further inspection, what is the maximum weight allowed for a widget to be considered within the acceptable range?

(d) What is the probability that a product is neither premium nor defective?

Validate and compare the answer of each subpart with its theoretically calculated answer.
Hint: To calculate theoretically you may explore the statistics. `NormalDist` library.

Question 4: In poker, a full house is a hand that contains three cards of one rank and two cards of another rank, such as $3 \leftrightarrow 3 * 6 * 6*$. (Wilkipedia)

(a) What is the probability of getting a full house?

(b) Simulate a python experiment and compare with the theoretical value.

(c) What is the probability that in 1000 trials one will get at least 2 full houses? You may assume that the hands were dealt independently from one another. Compare this value after simulating it in python. Calculate mean and variance.