IC-252 Lab Assignment 5

Question-1:

A data file named *'random_num'* has been provided containing an array of 10000 random real numbers. Write a Python code to:

a) Find the mean and standard deviation.

b) Plot the histogram (take bin size 0.02) of the given data and mention the most probable bin.

c) Fit gamma, log normal, beta, exponential, and normal distribution on given data using the hint given below. Based on the *ks_pvalue* report the distributions that our data follows.

d) From the histogram plot we can see that there are two distinct subgroups present in the data. For each subgroup find the mean and variance and also fit the distributions given in part c. For each subgroup based on the *ks_pvalue* report the distributions the subgroup follows.

Hint: To fit the distribution to a data set use following packages and functions

from fitter import Fitter, get_common_distributions, get_distributions
f = Fitter(data, distributions= a list containg distribution's name)
f.fit()
f.summary()

"For gamma, log normal, beta, exponential and normal we use following names in the code: *"gamma", "lognorm*", "*beta*","*expon*","*norm*" respectively. If *ks_pvalue* is less than *0.05* (level of significance) for any distribution, then we reject that distribution (means data do not follow that distribution). For example, if *ks_pvalue = .0006 < 0.05* and our distribution is *"norm"* then we can say that data do not follow the normal distribution. If *ks_pvalue > 0.05* for a particular distribution, then data follows that distribution.

Question-2:

Covid-19 proved to be a deadly pandemic which caused millions of deaths and destroyed the economy of various nations. Lockdowns were imposed to contain the spread of the virus. Mobility was significantly reduced during this period.

The Community Mobility Reports provided by google show movement trends by region, across different categories of places. These reports are created with aggregated, anonymized sets of data from users who have turned on the Location History setting, which is off by default. The data shows how visitors to (or time spent in) categorized places change compared to our baseline days. A baseline day represents a *normal* value for that day of the week. The baseline day is the median

value from the 5-week period Jan 3 – Feb 6, 2020.For each region-category, the baseline isn't a single value—it's 7 individual values. The same number of visitors on 2 different days of the week, result in different percentage changes.

You are given a dataset which contains the mobility data of India in 2021. You are expected to do the following for Mumbai City:

A. Plot the monthly graph (You have to essentially plot the mobility values for each day and display the month names on X-axis) for each category of mobility (grocery,retail,transit,parks,residential,workplaces). What do you infer from these graphs?

B. Calculate the variance for each of the 6 mobilities for 2 time periods:
   - During the lockdown period (consider only the months of April and May (till 20/05/2021)).
   - For the entire year (365 days).

   Compare these variances for each category. What do you infer?

C. Let's assume that the data given by Google is probabilistic in nature and each mobility has a certain probability associated with it. So on a particular day for a region, we compute the expected mobility using the formula:

$$E(Mobility) = \sum P(Mobility_i) \times Mobility_i$$

Consider the following distribution during the lockdown period (1/04/2021 to 20/05/2021):

Grocery / Pharma	Retail	Transport	Parks	Residential	Workplace
p=0.2	p=0.2	p=0.05	p=0.02	p=0.5	p=0.03

Compute the expected mobility for each day during the lockdown period and compare these values with the original mobilities for each category during the lockdown period. What do you infer?

Question 3:

Consider the following data for two stocks A and B:

Scenario	Probability	Asset A Return (%)	Asset B Return (%)
I	0.4	-10	5
II	0.2	10	20
III	0.4	20	10

1) Theoretically compute the expected return, variance, and standard deviation for each stock.

2) For each stock randomly sample N return values following the above given probability distribution. For simulation take N = 100,200,500, and 1000. For each value of N, from the sampled return values compute mean, variance and standard deviation and compare them with the theoretical results for each stock.

3) Suppose that the investor makes a portfolio which invests w1 proportion of the wealth in stock A and the remaining (let's say w2) in stock B. Vary the values of w1 in a range 0.1 to 0.9 and choose w2 = 1-w1 compute expected risk and return value for all combinations and plot risk vs return graph. Based on the expected risk and return value which (w1, w2) combination you will prefer for investment.

4) For your preferred (w1, w2) combination compute the portfolio return and risk from the sampled return data in step 2 for N=1000. Compare the theoretical and results from sampled data.