CENG222 Probability and Statistics

HOMEWORK 2

DUE DATE: 01.06.2024 - 23:55

Submission Rules:

1. This is **NOT** a group assignment. All kinds of cheating and plagiarism will be penalized and such submissions will be graded as **0**.
2. You should submit your assignments through TEAMS until due date. There will be no extensions.
3. You have to submit one zipped file including one file for your code (**py**) and one file for your report (**pdf**).
4. Your homework should be named as **CENG222_HW2_studentID.zip**
5. You should write your **student ID** both in your code and report files**.**

You are a financial analyst working for a prestigious investment firm, Capital Analytics LLC. The firm is considering two different investment strategies for their clients: one based on high-frequency trading algorithms and the other on value investing principles. Your task is to analyze historical client data to determine which strategy, if any, has yielded superior returns.

Luckily, for returns gained from both strategies the same probability density function (pdf) is assumed:

$$f(x) = (k+1)x^k.$$

Your task is to simulate experiments in Python by filling in the provided skeleton code and documenting your findings in a report. The requirements for the report content is highlighted as **blue**. In total, you need to have **4 figures** in your report that have different histograms for both populations.

It is crucial to note that for random number generation, you are **ONLY** permitted to use the **random.random()** function. This function requires no parameters and yields values uniformly distributed in the range of 0 to 1.

Functions:

In this part, you are allowed to use numpy arrays if you like. From the numpy package you are **ONLY** allowed to use the **numpy.zeros** (for array allocation), **numpy.sum** (to compute the sum of the array without using any loops) and **numpy.log** (to compute the element-wise natural logarithm of the array without using any loops) functions.

You need to implement **10 functions** and their requirements are detailed in the skeleton code as line comments:

1. gen_inverse(k, M): Derive the **required algorithm for inverse transformation method** to generate samples from the given pdf and implement here.
2. gen_rejection(k, M): Derive the **required algorithm for rejection method** to generate samples from the given pdf and implement here.
3. calc_population_mean(k): Derive **the expected value of the given distribution** and implement here.
4. calc_population_variance(k): Derive **the varience of the given distribution** and implement here.
5. random_sample(population, N): Implement the function as instructed.

6. calc_sample_mean(sample): Implement the function as instructed.
7. calc_sample_variance(sample, unbiased=True): Implement the function as instructed.
8. estimate_k_mom(sample): Derive **the Method of Moments estimate of k for the given distribution** and implement here.
9. estimate_k_mle(sample): Derive **the Maximum Likelihood estimate of k for the given distribution** and implement here.
10. calc_conf_int_mean(sample, confidence_lvl, pop_std=0): Implement the function as instructed.

Experiments:

In this part, you are allowed to use numpy arrays if you like. From the numpy package you are **ONLY** allowed to use the **numpy.zeros** (for array allocation), **numpy.sum** (to compute the sum of the array without using any loops), **numpy.logical_and** (to compute the element-wise and of two arrays without using any loops) and **numpy.mean** (to compute the mean of the array without using any loops) functions. Obviously, you are also allowed to use the function you have implemented above.

1. You need to synthetically generate two populations with different k values given in the skeleton code. These populations will only be used to sample from. You also need to calculate and print their means and variances and plot the **population histograms**.

   What does these values and plots tell you about the two investment strategies?

2. You need to collect 100000 random samples of size 25 from both populations. Using these samples, you need to calculate sample means, biased and unbiased sample variances, MoM and MLE estimates of the parameter k and population mean intervals with 97% confidence with and without using the population standard deviation.
3. You need to calculate and print means of sample means, biased and unbiased sample variances and MoM and MLE estimates of the parameter k for both populations. You also need to plot the **histograms of sample means**, **histograms of k estimates using MoM** and **MLE** for both populations.

   What do you observe for sample mean distributions and the mean sample means?
   What do you observe for biased and unbiased sample variances?
   What do you observe for the distributions of MoM and MLE estimates of k? Are these estimators unbiased?

4. You need to calculate and print the ratio of confidence intervals computed with and without using the population standard deviation that contains the population mean for both populations.

   What do you observe for the confidence intervals computed with and without using the population standard deviation?

5. You need to collect a sample of length 100000*25 from both populations, calculate and print their sample means, biased and unbiased sample variances, MoM and MLE estimates of parameter k and confidence intervals with and without using the population standard deviation.

   What do you observe for all the sample statistics computed?
   What do you observe for the confidence intervals computed with and without using the population standard deviation?