CENG 222
Probability and Statistics
HOMEWORK 2
Central Limit Theorem

DUE DATE: 02.04.2022 - 23:59

Instructions:

If we add up a large number of values from almost any distribution, the distribution of the sum converges to normal. More specifically, if the distribution of the values has mean and standard deviation μ and σ, the distribution of the sum is approximately N(n*μ, n*(σ^2)). This is called the Central Limit Theorem.

Conditions:

- The values have to be drawn independently.
- The values have to come from the same distribution (although this requirement can be relaxed).
- The values have to be drawn from a distribution with finite mean and variance.
- The number of values you need before you see convergence depends on the skewness of the distribution.

The Central Limit Theorem explains, at least in part, the prevalence of normal distributions in the natural world. Most characteristics of animals and other life forms are affected by a large number of genetic and environmental factors whose effect is additive. The characteristics we measure are the sum of a large number of small effects, so their distribution tends to be normal. Another example can be the weight of a bag of potato chips which is actually the sum of the weights of many small independent contributions from each single chip.

In this assignment, you are going to simulate and observe the impact of number of values added up and the result when independent and identical sampling conditions are ignored. For this purpose, you need to implement 5 different experimental setups as the following:

Experiment 1: 2 values are sampled independently from a standard uniform distribution and summed.

Experiment 2: 4 values are sampled independently from a standard uniform distribution and summed.

Experiment 3: 50 values are sampled independently from a standard uniform distribution and summed

Experiment 4: 50 values are sampled dependently from a uniform distribution and summed. Dependence is introduced by the following rule: If a value is smaller than 99, the next value is sampled from a uniform distribution between 0 and 200, otherwise between 99 and 101.

Experiment 5: 50 values are sampled independently from different uniform distributions and summed. For each value generation, the uniform distribution parameters (a and b-a) should be sampled from a standard uniform distribution.

For each experiment, generate **200000** sums and create a **`pyplot`** figure which includes their normalized histograms and the theoretical normal distribution curve. Implement and use your own normal probability density function. For experiment 1, 2 and 3, calculate the theoretical mean and variance for the distribution of the sums and implement it explicitly. For experiment 4 and 5, estimate the mean and variance for the distribution of the sums by calculating them on the created sample set.

In order to draw samples from a uniform distribution, you can use **`numpy.random.uniform()`** function.

For clarification, the expected output for Experiment 3 is given below: