CMPE 343: Introduction to Probability and Statistics for Computer Engineers (Fall 2021)

Homework #2
Due December 29, 2021, 11:59pm

Disclaimer

Please type your answers and submit your homework as PDF. To get full points, you need to show your steps clearly. If you use a theorem, rule, definition, derivation, etc. that were not covered in the lectures, you need to cite your resources. If you fail to cite your references, if you plagiarize, if you give your answers to another person, if you copy someone else's answers, your grade will be -100. If this homework is found online in homework help sites, or any other website, it will be subjected to cancellation for *all class*. The homework will be also subjected be converted into oral homework.

Problem 1

As a student of CMPE, your skills are not limited to academy related work. Now, Benan Köfte needs your help for choosing his new caravan. He wants to estimate the mean value for the number of daily customers. Depending on the number, he will choose the expensive big one, or the small cheaper one. Benan Köfte shares his data of daily number of customers. You should help him using the shared data.

a. They tell you that number of daily customers are Gaussian with $\sigma = 0.1$. Construct a 95% confidence interval for the mean value for the daily number of customers.

b. Benan Köfte claims that the true mean value for the number of daily customers is 55. State the null and alternative hypotheses. Test his claim with $\alpha = 0.05$.

c. Find also the p-value.

d. How are these three approaches **a**, **b**, and **c** are differ from each other?
e. You realize that his apprentice made a mistake in calculating the variance of the daily number of customers. Now, you don't know the variance but you still are sure of the fact that number of customers are Gaussian. Construct again a 95% confidence interval for the mean value for the number of customers. You should also comment on why the interval is wider than the one in **a**.

You can load the shared data as follows:

```
import numpy as np
path_to_data = "benan.npy"
data = np.load(path_to_data)
```

Problem 2

a. A random real number k is sampled from a uniform distribution on the interval $[0, \beta]$. You want to test the null hypothesis, $H_0, \beta = 3$ against the alternative hypothesis $H_A : \beta \neq 3$ by rejecting H_0 if $k \leq 0.1$ or $k \geq 2.9$. Compute the probability of Type I and Type II errors if the true value of β is 3.5.

b. (bonus with a very few points) Reflect and comment on this xkcd meme. Explain why testing the hypotheses multiple times is not a good idea.

Problem 3

Pseudo random number generators are able to generate sequences of numbers that statistically behave like uniform i.e. they appear to be sampled from $U[0,1]$. One natural question is: how can we transform these numbers to obtain number sequences from other distributions, for example the standard normal? To efficiently generate random numbers from the standard normal, Box-Muller suggested their now famous method where one can get two standard normal random numbers using two random uniform numbers with the following transform

$$Z_1 = \sqrt{-2\ln U_1} \cdot \cos 2\pi U_2$$

$$Z_2 = \sqrt{-2\ln U_1} \cdot \sin 2\pi U_2$$

where U_1 and U_2 are uniform random numbers.
a. Implement a linear congruential generator to obtain uniform random from $U[0,1]$.
b. Using the generator you have implemented, sample two sets of 10000 numbers from the uniform distribution $U[0,1]$. Apply the transformations to obtain Z_1 and Z_2 and plot two histograms and confirm visually that they are normally distributed. Now, repeat the process but this time obtain uniform random numbers from a prebuilt generator, you can use either Python's or Numpy's random modules, compare the results.
b. (Bonus) Why are Z_1 and Z_2 normally distributed? Argue analytically.

Problem 4

a. Use Monte-Carlo Simulation to integrate $\cos(x)$ from 0 to π. Give the result in 4 decimal point precision.
b. Use Monte-Carlo Simulation to estimate the probability $P(0 \leq Z \leq 1)$ where Z is a standard normal random variable.
c. Remember the Chebyshev's Inequality

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

for a continuous random variable X with mean μ and finite variance σ^2. Use Monte-Carlo Simulation to estimate the probability for $X \sim \mathcal{N}(0, 1)$ and for $k = 2, 5, 10$ and confirm that the inequality holds in all cases. For all questions, pick a relatively large sample size. You can experiment with different sample sizes and see how the estimations change.