Homework 3

To get full credit you need to:

- Clearly state the correct answer.
- Justify and explain each step in your solution and the final answer.
- Make very clear that you understand the logic behind the steps.
- Using a symbolic algebra system to perform computations does not exonerate you to explain and motivate every step as mentioned above.

Very important: Honor code applies fully. You are required to sign the pledge. You must submit your own work only.

Problem 1

a. Reproduce the Glivenko-Cantelli plots from the slides `"FctsGaussian.pdf"` (Gaussian and Chi-squared cases).

b. Using the implementation of the empirical CDF that you developed in a., check the convergence of the empirical cdf of $\bar{X}$ to the corresponding Gaussian when the population distribution is the finite distribution:

x	80	100	120
$p(x)$	0.2	0.3	0.5

Problem 2

Suppose the expected tensile strength of type-A steel is 105 ksi and the standard deviation of tensile strength is 8 ksi. For type-B steel, suppose the expected tensile strength and standard deviation of tensile strength are 100 ksi and 6 ksi, respectively. Let $\bar{X}$ be the sample average tensile strength of a random sample of 40 type-A specimens, and Let $\bar{Y}$ be the sample average tensile strength of a random sample of 35 type-B specimens.

a. What is the approximate distribution of $\bar{X}$? Of $\bar{Y}$?

b. What is the approximate distribution of $\bar{X} - \bar{Y}$? Justify your answer.

c. Calculate (approximately) $\mathbb{P}\left(-1 \leq \bar{X} - \bar{Y} \leq 1\right)$.

d. Calculate $\mathbb{P}\left(\bar{X} - \bar{Y} \geq 10\right)$. If you actually observed $\bar{X} - \bar{Y} \geq 10$, would you doubt that $\mu_1 - \mu_2 = 5$?

Problem 3

Suppose the sediment density (g/cm) of a randomly selected specimen from a certain region is Gaussian distributed with mean 2.65 and standard deviation 0.85 (suggested in "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants," Water Research, 1984: 1169–1174).

a. If a random sample of 25 specimens is selected, what is the probability that the sample average sediment density is at most 3.00? Between 2.65 and 3.00?

b. How large a sample size would be required to ensure that the first probability in part (a) is at least 0.99?

Problem 4

We have seen that if $\mathbb{E}[X_1] = \mathbb{E}[X_2] = \cdots = \mathbb{E}[X_n] = \mu$, then $\mathbb{E}[X_1 + \cdots + X_n] = n\mu$. In some applications, the number of X_i's under consideration is not a fixed number n but instead is a random variable N. For example, let $N =$ the number of components that are brought into a repair shop on a particular day, and let X_i denote the repair shop time for the ith component. Then the total repair time is $X_1 + X_2 + \cdots + X_N$, the sum of a random number of rv's. When N is independent of the X_i's, it can be shown that

$$\mathbb{E}[X_1 + \cdots + X_N] = \mathbb{E}[N] \cdot \mu \,.$$

a. If the expected number of components brought in on a particularly day is 10 and expected repair time for a randomly submitted component is 40 min, what is the expected total repair time for components submitted on any particular day?

b. Suppose components of a certain type come in for repair according to a Poisson process with a rate of 5 per hour. The expected number of defects per component is 3.5. What is the expected value of the total number of defects on components submitted for repair during a 4-hour period? Be sure to indicate how your answer follows from the general result just given.

Problem 5

In a presidential election, 50% of the population supports Ponald Drump (candidate A), 20% supports Broe Jiden (candidate B), and the rest are divided between Camelot Harris (C), Barnie Senders (D), and Tedd Bruise (E). A survey asks 400 randomly selected people who they support.

a. Using the CLT, estimate the probability that at least 52.5% of the respondents prefer candidate A.

b. Using the CLT, estimate the probability that less than 25% of the respondents prefer candidates C, D or E.