Question 1

Say that you are investigating the number of cases of a certain disease in a population each month. Let $X_1, \ldots, X_n$ be the data collected over the past n months, and assume $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Pois}(\lambda)$.

1. Find the MLE for the mean number of cases per month.
2. Compute the bias, variance and MSE of the above MLE as a function of λ and n.
3. A researcher believes that λ is approximately 3, so they suggest using the estimator which is the average between the MLE and 3: $T = \frac{\bar{X}_n + 3}{2}$. Compute the MSE of T as a function of λ and n.
4. Say that you have a sample of size $n = 10$. For which values of λ is T better than the MLE?

Question 2

You are (once again) analyzing the performance of cloud computing infrastructure, and are interested in the time until the failure of a machine. The time until failure is modeled as an exponential random variable. For n independent machines, you have a sample $X_1, \ldots, X_n$ of the time until their failure.

1. Find the maximum likelihood estimate for the parameter of the exponential distribution λ and for the expected time until failure.
2. Using the method of moments, estimate λ and the expected time until failure.

HINT: To compute the MLE for λ, follow the steps as in the example for Bernoulli probability in the lecture notes, slide 47. In this case, the likelihood function is given by $L(\lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n X_i\right)$.

Question 3

You are analyzing the weight of a certain product produced in a factory. Assume that the weight of the product is normally distributed, and you are interested in the expected value of the weight. You have a sample of 12 products, and the weight (in Kg) of each product is as follows:

$$53.8, 67.34, 51.7, 52, 58.9, 74, 45.3, 53, 62.5, 48.87, 49, 55.6$$

1. Assuming that the variance is known and equals 1.5 Kg, calculate the confidence interval for the expected value of the weight with confidence level 95%.
2. Repeat part 1, this time for a confidence level of 90%. What can you say about the difference between the results?
3. Assuming that the variance is known and equals 2 Kg, calculate the confidence interval for the expected value of the weight with confidence level 95%. What can you conclude from the result?
4. Repeat part 1, assuming that the variance is unknown.

Question 4

In a random sample of $n = 100$ students, it was found that 30 like Bamba.

1. Compute a confidence interval for the proportion of Bamba lovers among the students, with confidence level 95%.
2. Find the minimal sample size n for which the length of the CI will be at most 0.02.

HINT: Recall that the length of the CI for proportion depends on $\frac{\hat{p}(1-\hat{p})}{n}$, and note that $\hat{p}(1-\hat{p}) \leq 0.25$.