Question 1 (20 points – 4 points for each part)

TRUE, FALSE, or UNCERTAIN. Explain your answers:

(a) If an estimator is consistent, it is also unbiased.

(b) An unbiased estimator is consistent only if the variance of the estimator converges to zero.

(c) The sample mean always has a normal distribution regardless of the underlying population distribution.

(e) We typically try to construct tests whose test statistics have known distributions under the null so that an expression for α can be written down (i.e. so that we can choose a critical region for a given α).

Question 2 (15 points)

Let $X_1, \ldots, X_n$ be a random sample drawn from a $N(\mu, \sigma^2)$ distribution, where μ is known and σ^2 is unknown.

(a) (5 points) Derive the maximum likelihood estimator for σ^2.

(b) (3 points) Is the maximum likelihood estimator unbiased? Is it consistent?

(c) (7 points) For $n = 10$ and $\widehat{\sigma}^2_{MLE} = 8$, construct a test of the following hypotheses at the 5% significance level:

$$\begin{aligned} H_0 &: \sigma^2 = 4 \\ H_A &: \sigma^2 \neq 4 \end{aligned}$$

Question 3 (20 points)

Let $X_1, \ldots, X_n$ be a random sample drawn from a distribution with pdf

$$f_X(x) = \begin{cases} e^{-(x-\alpha)} & \text{for } x > \alpha \\ 0 & \text{for } x \leq \alpha \end{cases}, \text{ where } \alpha \text{ is unknown.}$$

(a) (6 points) Derive the maximum likelihood estimator for α.

(c) (8 points) Show how you can use the maximum likelihood estimator to generate a 90% confidence interval for α. Provide as much detail as possible.

Question 4 (20 points)

Mega-Hertz Auto rental is trying to decide what rate to charge for its new fleet of hybrid cars. The company decides to poll a group of renters about how much they are able to pay. Specifcally, the company asks 25 people what they can pay per day, X_i, and finds the following:

$$\overline{X} = \frac{1}{25}\sum_{i=1}^{25} X_i = 50$$

$$s^2 = \frac{1}{24}\sum_{i=1}^{25}(X_i - \overline{X})^2 = 225$$

The company then infers that 95% of the renters in the area can pay (approximately) between \$20 and \$80 a month.

(a) (5 points) What assumptions underly this inference? What are the potential problems with the company's reasoning? How will the number of renters the company polls affect the width of the interval?

(b) (7 points) Now the company is informed that economy-wide, $X \sim N(\mu, \sigma^2)$. Given the company's sample, construct a 95% confidence interval for the mean of the underlying distribution of X. How many renters would the company need to poll to limit the width of the confidence interval to \$5?

(c) (4 points) Provide an intuition for your choice of distribution in constructing the confidence interval in part (b).

(d) (4 points) Suppose instead the company is informed that economy-wide, $X \sim$ Uniform $[0, \theta]$.Unfortunately, the company has already thrown away the original data and kept only the sample mean and sample variance. Can the company still construct a 95% confidence interval for θ? Explain.

Question 5 (25 points)

Suppose that the height of men in Boston is normally distributed with mean of 70 inches and variance of 10 inches. In a recent publication, a Harvard Medical School doctor claims that men who attend MIT are on average 4 inches shorter (but still have the same variance). You interview 10 MIT male students and their mean height turns out to be 67 inches.

(a) (3 points) Suppose you want to do a test of the null hypothesis that the MIT males have the same average height as Boston males in general against the alternative hypothesis that the Harvard doctor is right. Write down the null hypothesis and alternative hypothesis in mathematical terms.

(b) (7 points) Perform a 5% test of the hypothesis in part (a).

(c) (4 points) What is the power $(1-\beta)$ of the test in part (b) ?

(d) (7 points) Suppose you only knew the mean of the distribution of men's height, not the variance, but you did have an estimate of the variance, $s^2 = 15$. Perform a 5% test of the null against alternative.

(e) (4 points) How would your answer in part (b) change if your alternative hypothesis was that the mean height of male MIT students was different from the Boston mean, but not in a specified way (i.e., it could be higher or lower)?

1. (15 points) Let X_1, X_2, X_3 be random variables such that $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, $\rho(X_{1,}X_2) = 0.4$, $\rho(X_{1,}X_3) = -0.1$, $\rho(X_{2,}X_3) = 0.2$.

a.(8 points) Find $E[(X_1 + X_2 + X_3)^2]$.

b.(7 points) Find $\mathrm{Cov}(X_1 + X_2, X_2 + X_3)$.

2. (15 points)

Let $X \sim \text{exponential}\,(3)$,

$Y = e^X$,

$Z \mid Y \sim \text{uniform}\,[0, Y]\,.$

a.(5 points) Calculate the pdf of Y.

b.(10 points) Calculate $E\,[Z]$ and $\mathrm{Var}\,(Z)$.

Hint: recall that $\mathrm{Var}\,(Z) = \mathrm{Var}\,(E[Z \mid Y]) + E\,[\mathrm{Var}\,(Z \mid Y)]\,.$

3. (30 points) Suppose that you are modeling the number of babies born per day in a particular hospital.

a.(5 points) Let n be a random variable representing the number of babies born in a given day. Which distribution that we discussed in class would be the most appropriate for modleing n? State the parameter(s) that you would need to estimate.

b.(5 points) Suppose that each baby has a probability p of needing intensive care. Let I represent the number of babies born on a particular day who will need intensive care. What is $E[I \mid n]$? What is $E[I]$?

c.(10 points) Conditional on n, what is the probability that at least one baby born on a given day will need intensive care? What is the unconditional probability?

d.(5 points) You are given data for the number of births over a ten day period: 1, 2, 4, 2, 3, 3, 1, 0, 2, 3. For each parameter of the distribution, explain how you would derive an unbiased estimator. Using the data provided, calculate an estimate for each parameter.

e.(5 points) Derive an unbiased estimator for the variance of n. Using the data provided, calculate an estimate for the variance.

4. (20 points) Let L be a random variable representing the wait time to get a driver's license renewed at the Boston Registry of Motor Vehicles, and let R represent the time to get a vehicle registration renewed. Suppose that you know that $E[L] = 20, E[R] = 16$, $\operatorname{Var}(L) = 100$, $\operatorname{Var}(R) = 150$, but you don't know how L and R are distributed. You take a survey of 25 customers of each type to obtain random samples $L_1, ..., L_{25}$ and $R_1, ..., R_{25}$. You may assume that the wait times for each customer are independent. How would you approximate the probability that $\overline{L}_{25} > \overline{R}_{25}$?

5. (20 points) Let $X_1, ..., X_n$ be a random sample from a distribution $X \sim \text{uniform}\,[0, \theta]$.

a.(5 points) Let $\widehat{\theta}_1 = 2\overline{X_n}$. Is $\widehat{\theta}_1$ an unbiased estimator for θ?

b.(5 points) Let $\widehat{\theta}_2 = \max\{X_1, ..., X_n\}$. Is $\widehat{\theta}_2$ an unbiased estimator for θ?

c.(10 points) Let $\widehat{\beta} = \frac{2}{n} \sum_{i=1}^{n} \frac{1}{X_i}$. Is $\widehat{\beta}$ an unbiased estimator for $\frac{1}{\theta}$? (Hint: let $Y = \frac{1}{X}$.)

1. Let $X_1, \ldots, X_3$ be Bernoulli(p). Perform a test of

$$\begin{aligned} H_0 &: p = \frac{1}{3} \\ H_A &: p = \frac{2}{3} \end{aligned}$$

2. You observe one value X drawn from one of the following distributions:

$$\begin{aligned} H_0 &: f_0(x) = 2x \\ H_A &: f_A(x) = 1 - 2x \end{aligned}$$

a. Find the most powerful test (smallest β) such that $\alpha \leq 0.1$. Would you reject H_0 for $x = 0.6$?

b. Find the testing procedure that minimizes $(\alpha + \beta)$ Would you reject H_0 for $x = 0.6$?

3. Let $X_1, \ldots, X_3$ be $U[0, \theta]$.. Perform a test of Y

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_A &: \theta \neq \theta_0 \end{aligned}$$

4. A study was conducted of a drug meant to decrease the number of asthma attacks in patients. For 9 patients participating in the trial, the number of attacks in the next month were reduced by the following amounts: 0, 9, -2, 4, 5, -1, 0, 9, 3. Test the following hypothesis

$$\begin{aligned} H_0 &: \text{decrease} \sim N(0, \sigma^2) \\ H_A &: \text{decrease} \sim N(\mu, \sigma^2), \text{ where } \mu > 0 \end{aligned}$$

Question 1:

Answer TRUE, FALSE, or UNCERTAIN. Correct answers without adequate explanation will receive no credit:

a) You construct 100 95% confidence intervals for β using 100 different random samples. You would expect 95 of them to contain the true value of β.

b) You construct a 95% confidence interval for β using a random sample. The interval you come up with is [-2.7,1.2]. The probability that the constructed confidence interval contains the true value of β is 0.95.

c) It is not possible to construct confidence intervals without knowing the underlying distribution of the sample.

Question 2:

Let $X_1, ..., X_n, X_{n+1}$ be a sample from a normal population having an unknown mean μ and variance 1. Let $\overline{X}_n$ be the average of the first n of them.

a) What is the distribution of $X_{n+1} - \overline{X}_n$?

b) If $\overline{X}_n = 4$, give an interval that, with 90% confidence, will contain the value of X_{n+1}.

Question 3:

Suppose the sample statistics for a random sample of ten observations from a $N(\mu, \sigma^2)$ population are the following:

$$\begin{aligned} \overline{x} &= 5 \\ s^2 &= 4 \end{aligned}$$

a) Construct a 95% confidence interval for μ.

b) Suppose you want the length of the confidence interval to be only 3.57; what is the confidence level?

Question 4:

Suppose a random sample is drawn from a $N(\mu, \sigma^2)$ population, with σ^2 known.

a) How large a sample must be drawn in order that a 95% confidence interval for μ has length less than 0.01σ?

b) Suppose σ^2 is known to be 2 and $\overline{x} = 5$.

Also, due to time and budget constraints, you are only able to come up with a sample half as large as what you found in part a. Holding the confidence interval length constant, what is the confidence level?

c) How wide is the 95% confidence interval for the sample of half the size?

1. Suppose you have a random sample of size n (i.e., X_i $i.i.d.$, $i = 1, 2, ..., n$) from a uniform distribution on $[\alpha, \beta]$.
(a) Assuming $\alpha = 0$ and $\beta = 4\theta$, propose an unbiased estimator for θ, and show that it is unbiased.
(b) Assume $\alpha = \theta - \frac{1}{2}$ and $\beta = \theta + \frac{1}{2}$. Derive the MLE for θ.
(c) Assume still $\alpha = \theta - \frac{1}{2}$ and $\beta = \theta + \frac{1}{2}$. Now however suppose that you cannot observe $X_i's$, and that instead you observe

$Y_1, ..., Y_n$,where

$$Y_i = \begin{cases} 0 & \text{if } X_i < 0 \\ 1 & \text{if } X_i > 0 \end{cases}.$$

Under what circumstances can you obtain an unbiased estimator for θ ? Describe specifically how you would do it.

2. Let $X_1, X_2, ..., X_n$ be a random sample where X_i ~exponential(β).
(a) Derive the MLE for β.
(b) Find the MLE for $\sqrt{\beta}$.
(c) Is your MLE in part (a) unbiased ? Formally prove or disprove.
(d) Is your MLE in part (a) consistent ? Formally prove or disprove using LLN.

3. Let $X_1, X_2, ..., X_n$ be a random sample from the distribution having pdf

$$f(x; \theta_1, \theta_2) = \begin{cases} (\frac{1}{\theta_2})e^{-(x-\theta_1)/\theta_2} & \theta_1 \leq x < \infty, -\infty < \theta_1 < \infty, 0 < \theta_2 < \infty \\ 0 & \text{elsewhere} \end{cases}$$

Find the MLE of θ_1 and θ_2. (Hint: First, find $\widehat{\theta}_{1,MLE}$ while paying careful attention to the support of the pdf)

5. Let Y_1 and Y_2 be two stochastically independent unbiased estimator of θ. Suppose that the variance of Y_1 is twice the variance of Y_2. Find the constants k_1 and k_2 so that $k_1 Y_1 + k_2 Y_2$ is an unbiased estimator with the smallest possible variance for such a linear combination. Provide an intuition about your answer.

Question 1:

Let $X_1, ..., X_n$ be a random sample (i.i.d.) of size n from a population $f(x)$ with mean μ and variance σ^2 (both finite). Prove the following:

a) $E(S^2) = \sigma^2$, where S^2 is the sample variance.

b) $S^2 \xrightarrow{p} \sigma^2$ (S^2 converges in probability to σ^2).

Question 2:

Let S_0 denote the price of a given stock today. Suppose the price of the stock evolves over time as follows:

$$S_t = S_{t-1} + X_t$$

where

$$Xt = \left\{ \begin{array}{c} 1 \text{ with probability } 0.39 \\ 0 \text{ with probability } 0.20 \\ -1 \text{ with probability } 0.41 \end{array} \right\}$$

a) Express the change in the price of the stock over the first 700 periods, $\Delta S = S_{700} - S_0$, in terms of the $X_t s$.

b) What is the approximate distribution of the average daily change in the stock's price?

c) What is the probability that the stock is up at least 10 after 700 time periods?

Question 3:

A manufacturer of booklets packages them in boxes of 100. It is known that, on average, each book weighs 1oz., with a standard deviation of 0.05oz. The manufacturer is interested in calculating the following probability: P(100 booklets weigh more than 100.4oz.).Approximate this probability, mentioning any relavent theorems you use.

Question 4:

The lifetime of a certain critical electrical part is a random variable with mean 100 hours and standard deviation 20 hours.

(i) If 16 such parts are tested, find the probability that the sample mean is:

a) less than 104 hours.

b) between 98 and 104 hours.

(ii) Given that immediate replacement of a faulty part is strictly necessary for operations, how many of the electrical parts must be in stock so that the probability that the system is in continual operation for the next 2000 hours is at least 0.95?

Question 5:

Suppose $X_1, ..., X_n$ are a random sample from a negative binomial (r, p) distribution.

a) How would you apply the Central Limit Theorem as n becomes sufficiently large?

b) If $r = 10$, $p = 0.5$, and $n = 60$, use the Central Limit Theorem to find $P(\overline{X} \leq 11)$.

Quesiton 6:

Provide short answers to the following:

a) Under what circumstances is the sample mean an unbiased estimator for the mean of a distribution? Under what circumstances does have a normal distribution?

b) What is a "standardized" random variable? The limit as $n \longrightarrow \infty$ of a standardized binomial random variable has what distribution?

c) Describe how choosing the minimum mean squared error estimator is a way to balance bias and efficiency of an estimator?

d) A random variable X has mean 0 and variance 20. How large can $P(|X| >$

$30)$ be? How large is it if X has a normal distribution?

Question 7:

Let $X_1, ..., X_n$ be a random sample of size n from a population with mean μ and variance σ^2. Consider a statistic formed by taking a linear combination of the $X_i s$, $c_1 X_1 + ... + c_n X_n$, where $c_i \geq 0$. For example, the sample mean, $\overline{X}$, is the linear combination with $c_i = \frac{1}{n}$.

(a) Under what condition(s), is the statistic an unbiased estimator of μ?

(b) What is the variance of the statistic?

Question 1:

PART A:

Answer TRUE, FALSE, or UNCERTAIN. Correct answers without adequate explanation will receive no credit:

a) You construct 100 95% confidence intervals for β using 100 different random samples. You would expect 95 of them to contain the true value of β.

b) You construct a 95% confidence interval for β using a random sample. The interval you come up with is [-2.7,1.2]. The probability that the constructed confidence interval contains the true value of β is 0.95.

c) It is not possible to construct confidence intervals without knowing the underlying distribution of the sample.

PART B:

The capacities (in ampere-hours) of 10 batteries were recorded as follows:

140, 136, 150, 144, 148, 152, 138, 141, 143, 151

a) Esitmate the population variance.

b) Compute a 99% two-sided confidence interval for the population variance.

c) Compute a value v that enables us to state, with 90% confidence, that the population variance is less than v.

Question 2:

Let $X_1, ..., X_n, X_{n+1}$ be a sample from a normal population having an unknown mean μ and variance 1. Let $\overline{X}_n$ be the average of the first n of them.

a) What is the distribution of $X_{n+1} - \overline{X}_n$?

b) If $\overline{X}_n = 4$, give an interval that, with 90% confidence, will contain the value of X_{n+1}.

Question 3:

Suppose the sample statistics for a random sample of ten observations from a $N(\mu, \sigma^2)$ population are the following:

$$\begin{aligned} \overline{x} &= 5 \\ s^2 &= 4 \end{aligned}$$

a) Construct a 95% confidence interval for μ.

b) Suppose you want the length of the confidence interval to be only 3.57; what is the confidence level?

Question 4:

Suppose a random sample is drawn from a $N(\mu, \sigma^2)$ population, with σ^2 known.

a) How large a sample must be drawn in order that a 95% confidence interval for μ has length less than 0.01σ?

b) Suppose σ^2 is known to be 2 and $\overline{x} = 5$.

Also, due to time and budget constraints, you are only able to come up with a sample half as large as what you found in part a. Holding the confidence interval length constant, what is the confidence level?

c) How wide is the 95% confidence interval for the sample of half the size?

Question 5:

Suppose you are going to draw a random sampel from a Bernoulli distribution in hopes of estimating the parameter p. How large a sample will you need in order to guarantee that the lenght of a 95% confidence interval is at most 0.01? (Hint: Assume n will be large and consider bounds on the quantity $p(1-p)$.)

Question 1:
Let $X_1, X_2, ..., X_n$ be a random sample where X_i ~exponential(β), that is: $f_{X_i}(x_i) = \frac{1}{\beta}e^{-x_i/\beta}, x_i > 0$ and 0 elsewhere.

a. Derive the MLE for β.

b. Find the MLE for $\sqrt{\beta}$.

c. Is your MLE in part a. unbiased? Formally prove or disprove.

d. Is your MLE from part a. consistent?

Question 2:
Assume a sample of continuous random variables: $X_1, X_2, ..., X_n$, where $E[X_i] = \mu, Var[X_i] = \sigma^2 > 0$. Consider the following estimators: $\widehat{\mu}_{1,n} = X_n, \widehat{\mu}_{2,n} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$.

a. Are $\widehat{\mu}_{1,n}$ and $\widehat{\mu}_{2,n}$ unbiased?

b. Are $\widehat{\mu}_{1,n}$ and $\widehat{\mu}_{2,n}$ consistent?

c. What do you conclude about the relation between unbiased and consistent estimators?

Question 3: Consider a sample of random variables: $X_1, X_2, ..., X_n$, where $n > 10, E[X_i] = \mu, Var[X_i] = \sigma^2 > 0$ and the estimator: $\widehat{\mu}_n = \frac{1}{n-10}\sum_{i=11}^{n} X_i$. [Hint: note that the sum is over $(n-10)$ random variables].

a. Calculate the bias of $\widehat{\mu}_n$

b. Calculate the variance of $\widehat{\mu}_n$.

c. Calculate the MSE of $\widehat{\mu}_n$.

d. Is $\widehat{\mu}_n$ efficient in a finite sample?

e. Can you think of a scenario where you might want to use $\widehat{\mu}_n$?

Question 6:

.

b. Assume $(X_1, X_2, ..., X_n)$ are a random sample from the distribution with a cdf: $F_X(x) = \begin{cases} 1 - (\theta_1/x)^{\theta_2} & \text{if } x \geq \theta_1 \\ 0 \text{ otherwise} \end{cases}$, where $\theta_1 > 0, \theta_2 > 0$. Find the Maximum Likelihood estimator of $\theta = (\theta_1, \theta_2)$.