Name: SOLUTIONS

Problem 1 (25 points)
Answer the following questions:

a. Suppose \( \bar{X} \) is unbiased estimator of \( \mu_X \). Is \( \bar{X}^2 \) an unbiased estimator of \( \mu_{\bar{X}}^2 \)?

\[
E \bar{X}^2 = \text{VAR}(\bar{X}) + (E\bar{X})^2 = \frac{\sigma^2}{n} + \beta^2 \sim N(0,1)
\]

b. Let \( X_1, X_2, \ldots, X_n \) be i.i.d. random variables with \( X_i \sim N(0, \sigma) \). Find \( E \frac{X_i^2}{\sigma^2} \).

\[
\left( \frac{X_i - \mu}{\sigma} \right)^2 \sim X^2, \quad \Rightarrow \quad E \frac{X_i^2}{\sigma^2} = 1
\]

c. Refer to part (b). What is the distribution of \( \frac{X_i}{\sqrt{\sum_{i=1}^{n-1} X_i^2}} \)? Please show all your work.

\[
\frac{X_i - \mu}{\sigma} \sim T_{n-1}
\]

\[
\sqrt{\frac{n-1}{\sum_{i=2}^{n} (X_i - \mu)^2}} \sim T_{n-1}
\]

d. Suppose the logarithm of a stock price in six months will follow the normal distribution with mean $50 and standard deviation $20. Find the probability that the stock price in six months will exceed $79. Note: It is given that \( \ln(X) \sim N(50, 20) \).

\[
P(\ln X > \ln 79) = P\left( Z > \frac{\ln 79 - 50}{20} \right) = P\left( Z > 2.18 \right)
\]

\[
P(Z > 2.18) = 0.0897
\]
Problem 2 (25 points)
Answer the following questions:

a. Let \( X \sim \Gamma(\alpha, \beta) \). Show that \( \frac{2}{3} X \sim \chi^2_{3\alpha} \).

\[
f_X(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)}
\]

\[
f_Y(y) = \frac{3}{2} f_X \left( \frac{3y}{\beta} \right) = \frac{3}{2} \frac{(\frac{3y}{\beta})^{\alpha-1} e^{-\frac{3y}{\beta}}}{\Gamma(\alpha) (\frac{3}{2})^\alpha} = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) 2^\alpha} = \frac{2}{3} f_X \left( \frac{2y}{\beta} \right) = f_X \left( \frac{2y}{\beta} \right)
\]

b. Refer to part (a). Suppose \( X_1, X_2, X_3, X_4 \) are i.i.d. random variables with \( X_i \sim \Gamma(3, 4) \). Find \( c \) such that \( P(\sum_{i=1}^4 X_i > c) = 0.05 \).

\[
P\left( \sum_{i=1}^4 X_i > \frac{3}{4} c \right) = 0.05 \Rightarrow P\left( \frac{3}{4} \sum_{i=1}^4 X_i > \frac{3}{4} c \right) = 0.05
\]

\[
P\left( \frac{2}{3} X_1 + \frac{2}{3} X_2 + \frac{2}{3} X_3 + \frac{2}{3} X_4 > \frac{3}{4} c \right) = 0.05
\]

\[
P\left( X_{24} > \frac{c}{2} \right) = 0.05
\]

\[
\begin{array}{c}
\text{Graph of } \chi^2_{24} \\
\end{array}
\]

c. Let \( X_1, X_2, \ldots, X_n \) be i.i.d. random variables with \( X_i \sim N(0, \sigma^2) \). Find \( \sigma^2 \), the mle of \( \sigma^2 \).

\[
l(\sigma^2) = \frac{1}{16\sqrt{\pi n}} \frac{e^{-\frac{1}{2} \sum_{i=1}^n X_i^2}}{\sigma^{2n}}
\]

\[
\ln l(\sigma^2) = \frac{1}{2} \sum_{i=1}^n X_i^2
\]

\[
\frac{\partial^2 l(\sigma^2)}{\partial \sigma^2} = -\frac{2n}{2\sigma^4} + \frac{\sum_{i=1}^n X_i^2}{2\sigma^4} = -\frac{n}{\sigma^4} + \frac{\sum_{i=1}^n X_i^2}{2\sigma^4}
\]

\[
\text{MUE with } \frac{\sum_{i=1}^n X_i^2}{\sigma^4}
\]

\[
\sigma^2 = \frac{\sum_{i=1}^n X_i^2}{\sigma^4}
\]

\[
\text{VAR}(\sigma^2) = \text{MSE} \left( \frac{\sum_{i=1}^n X_i^2}{\sigma^4} \right) = \text{MSE} \left( \frac{\sigma^2 \sum_{i=1}^n X_i^2}{\sigma^4} \right) = \frac{2n}{n^2} = \frac{2}{n}
\]

\[
\text{MSE}(\sigma^2) = \frac{2}{n}
\]

\[
\text{MSE}(\sigma^2) = \frac{2}{n} = \frac{2\sigma^4}{n}
\]

d. Refer to part (c). Is \( \sigma^2 \) MVUE of \( \sigma^2 \)?
Problem 3  (25 points)

Answer the following questions:

a. Let \( X_1, X_2, \ldots, X_n \) be i.i.d. random variables with \( X_i \sim N(\theta, 1) \). Let \( \hat{\theta} \) be the mle of \( \theta \). Is \( \hat{\theta} \) MVUE of \( \theta \)?

\[
\begin{align*}
\ell(\theta) &= -\frac{1}{2} \ln(n) - \frac{1}{2} \sum_{i=1}^{n} (X_i - \theta)^2 \\
\ell_n(\theta) &= -\frac{1}{2} \ln(n) - \frac{1}{2} \sum_{i=1}^{n} (X_i - \hat{\theta})^2 \\
\frac{\partial^2 \ell_n}{\partial \theta^2} &= -\frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\theta})^2 \\
\text{and } \frac{\partial \ell_n}{\partial \theta} &= \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\theta}) \\
\text{Hence, } \text{Var}(\hat{\theta}) &= \frac{1}{n} \\
\text{and } \text{bias}(\hat{\theta}) &= 0 \\
\text{Therefore, } \hat{\theta} \text{ is unbiased.}
\end{align*}
\]

b. If \( X \) has a binomial distribution with parameters \( n \) and \( p \), then \( \hat{p}_1 := \frac{X}{n} \) is an unbiased estimator of \( p \). Another estimator of \( p \) is \( \hat{p}_2 := \frac{X + 1}{n + 2} \). Find the bias of \( \hat{p}_2 \).

\[
\begin{align*}
\text{E} \hat{p}_2 &= \frac{\text{E} X + 1}{n + 2} = \frac{np + 1}{n + 2} \\
\text{bias} \hat{p}_2 &= \frac{\text{E} X + 1 - np}{n + 2} = \frac{np - np - 2p}{n + 2} \\
\text{bias} \hat{p}_2 &= \frac{1 - 2p}{n + 2}
\end{align*}
\]

c. Refer to part (b). Find \( MSE(\hat{p}_1) \) and \( MSE(\hat{p}_2) \).

\[
\begin{align*}
\text{MSE} \hat{p}_1 &= \text{Var} \hat{p}_1 = \frac{n p (1 - p)}{n^2} = \frac{p(1 - p)}{n} \\
\text{MSE} \hat{p}_2 &= \text{Var} \hat{p}_2 + \text{bias}^2 \hat{p}_2 = \frac{n p (1 - p)}{(n + 2)^2} + \left( \frac{1 - 2p}{n + 2} \right)^2 \\
\end{align*}
\]

d. Assume that you have a sample mean \( \bar{X} \) and a sample variance \( S^2 \) based on a random sample \( X_1, X_2, \ldots, X_n \) from a normal distribution. Suppose you want to predict a new value \( X_p \). Find the distribution of \( X_p - \bar{X} \) and use it together with the distribution of \( S^2 \) to form a ratio that follows the \( t \) distribution. Note: \( X_p \) is independent of \( X_1, X_2, \ldots, X_n \).

\[
\begin{align*}
\text{E} (X_p - \bar{X}) &= 0 \\
\text{MSE} (X_p - \bar{X}) &= \frac{\sigma^2}{n - 1} \\
\frac{X_p - \bar{X}}{\sigma \sqrt{\frac{1 + \frac{1}{n}}{n - 1}}} &= \frac{X_p - \bar{X}}{\sigma \sqrt{\frac{n - 1}{S^2 / n}}} \\
\end{align*}
\]

\[
\text{Note: } X_p \text{ is independent of } X_1, X_2, \ldots, X_n.
\]
Problem 4  (25 points)
Consider \( k \) independent random variables that have normal distributions with unknown means \( \mu_1, \mu_2, \ldots, \mu_k \), respectively, and unknown but common variance \( \sigma^2 \). Random samples of size \( n_i, i = 1, 2, \ldots, k \) are selected from these \( k \) populations. Answer the following questions:

a. Write the likelihood and log-likelihood function that involves all the information above.

\[
L = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\sum n_i} \exp \left( -\frac{1}{2\sigma^2} \sum \left( \frac{x_i - \mu_i}{\sigma} \right)^2 \right)
\]

\[
\ln L = -\frac{1}{2\sigma^2} \sum \left( \frac{x_i - \mu_i}{\sigma} \right)^2
\]

b. Use the method of maximum likelihood to find estimates of \( \mu_i, i = 1, 2, \ldots, k \) and \( \sigma^2 \). Denote these estimates with \( \hat{\mu}_i, i = 1, 2, \ldots, k \) and \( \hat{\sigma}^2 \).

\[
\hat{\mu}_i = \overline{x}_i
\]

\[
\hat{\sigma}^2 = \frac{\sum (x_i - \overline{x}_i)^2 + \cdots + \sum (x_k - \overline{x}_k)^2}{n_1 + \cdots + n_k}
\]

\[\text{MLEs,} \quad \hat{\sigma}^2 = \frac{\sum (x_i - \overline{x}_1)^2 + \cdots + \sum (x_k - \overline{x}_k)^2}{n_1 + \cdots + n_k}
\]

c. Is \( \hat{\sigma}^2 \) unbiased estimator of \( \sigma^2 \)? If not, give the unbiased estimator of \( \sigma^2 \) and denote it with \( S_k^2 \).

\[
E \hat{\sigma}^2 = \frac{(n_1-1)\sigma^2 + \cdots + (n_k-1)\sigma^2}{n_1 + \cdots + n_k}
\]

That is, the unbiased estimator will be:

\[
S_k^2 = \frac{\sum (x_i - \overline{x}_1)^2 + \cdots + \sum (x_k - \overline{x}_k)^2}{n_1 + n_2 + \cdots + n_k - k}
\]

d. Find the distribution associated with \( S_k^2 \).

\[
\frac{n_1 + n_2 + \cdots + n_k - k}{\sigma^2} S_k^2 \sim \chi^2_{n_1 + n_2 + \cdots + n_k - k}
\]