MATH5945 Statistical Inference Assignment 2

Problem One

$$f(x;\theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, \quad x > 0, \theta > 0$$

(a) Maximum Likelihood Estimation for θ

Given the probability density function, for n i.i.d. observations $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \prod_{i=1}^{n} \left(\frac{2X_i}{\theta} e^{-\frac{X_i^2}{\theta}} \right)$$

Taking the natural logarithm:

$$\ln L(\theta) = \sum_{i=1}^{n} \left[\ln(2X_i) - \ln(\theta) - \frac{X_i^2}{\theta} \right]$$

Simplifying (removing terms constant in θ):

$$\ell(\theta) = -n\ln(\theta) - \frac{1}{\theta} \sum_{i=1}^{n} X_i^2$$

Differentiating with respect to θ:

$$\frac{d}{d\theta}\ell(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} X_i^2$$

Set the derivative to zero for maximization:

$$-\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} X_i^2 = 0$$

Multiply through by θ^2:

$$-n\theta + \sum_{i=1}^{n} X_i^2 = 0$$

The maximum likelihood estimator of θ is:

$$\boxed{\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i^2}$$

(b) Using the density transformation formula, show random variable $Y = \frac{2}{\theta}X_1^2$ has χ_2^2 distribution with d$f_Y(y) = \frac{1}{2}\exp\left(-\frac{y}{2}\right), y >$ 0. where X_1 is a random variable with density:$f(x;\theta) = \frac{2x}{\theta}e^{-\frac{x^2}{\theta}}, x > 0, \theta > 0$.

Given the transformation:

$$Y = \frac{2}{\theta}X_1^2$$

where X_1 has density:

$$f_{X_1}(x) = \frac{2x}{\theta}e^{-\frac{x^2}{\theta}}, \quad x > 0, \theta > 0$$

The transformation is:

$$Y = g(X_1) = \frac{2}{\theta}X_1^2$$

The inverse transformation is:

$$X_1 = g^{-1}(Y) = \sqrt{\frac{\theta Y}{2}}$$

derivative of the inverse transformation is:

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{2}\sqrt{\frac{\theta}{2y}}$$

Absolute value of the Jacobian is:

$$\left|\frac{d}{dy}g^{-1}(y)\right| = \frac{1}{2}\sqrt{\frac{\theta}{2y}}$$

Using the density transformation formula:

$$f_Y(y) = f_{X_1}\left(g^{-1}(y)\right) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|$$

Substitute the expressions:

$$f_Y(y) = \frac{2\sqrt{\frac{\theta y}{2}}}{\theta}e^{-\frac{\frac{\theta y}{2}}{\theta}} \cdot \frac{1}{2}\sqrt{\frac{\theta}{2y}}$$

Simplify:

$$f_Y(y) = \frac{2\sqrt{\frac{\theta y}{2}}}{\theta}e^{-\frac{y}{2}} \cdot \frac{1}{2}\sqrt{\frac{\theta}{2y}} = \frac{1}{2}e^{-\frac{y}{2}}$$

This is the density of a chi-squared distribution with 2 degrees of freedom:

$$f_Y(y) = \frac{1}{2}e^{-\frac{y}{2}}, \quad y > 0$$

Thus:

$$\boxed{Y \sim \chi_2^2}$$

(c) Do the MLE and the UMVUE of θ coincide for this family?

Maximum Likelihood Estimator (MLE)

From part (a), we derived:

$$\hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

UMVUE

$$L(\theta) = \left(\frac{2}{\theta}\right)^n \left(\prod_{i=1}^{n} X_i\right) e^{-\frac{1}{\theta}\sum X_i^2}$$

By the factorization theorem, $T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic for θ.

The density can be written in exponential family form:

$$f(x;\theta) = 2x.\frac{1}{\theta}.\exp\left(-\frac{1}{\theta}.x^2\right)$$

showing that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2$ is a complete sufficient statistic

We need an unbiased estimator based on T. Compute $E[X_i^2]$:

$$\begin{aligned} E[X_i^2] &= \int_0^\infty x^2 \frac{2x}{\theta} e^{-x^2/\theta} dx \\ &= \frac{2}{\theta}\int_0^\infty x^3 e^{-x^2/\theta} dx \\ &= \theta \quad \text{(via integration by parts or recognizing the Gamma distribution)} \end{aligned}$$

Thus, $E[\hat{\theta}_{\text{MLE}}] = \theta$, showing the MLE is unbiased. Thus $\frac{T}{n} = \frac{1}{n}\sum X_i^2$ is unbiased for θ.

By the Lehmann-Scheffé theorem, since T is complete and sufficient, and $\frac{T}{n}$ is unbiased, it is the UMVUE:

$$\hat{\theta}_{UMVUE} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

The MLE and UMVUE are identical:

$$\hat{\theta}_{MLE} = \hat{\theta}_{UMVUE} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Reasons for Coincidence

- The model belongs to the exponential family with $T = \sum X_i^2$ as complete sufficient statistic
- The MLE is a function of T and happens to be unbiased
- When the MLE is unbiased and based on a complete sufficient statistic, it coincides with the UMVUE

(d) Prove that the family has a monotone likelihood ratio in the statistic T

For any $\theta_1 < \theta_2$, the likelihood ratio is:

$$\frac{L(\theta_2)}{L(\theta_1)} = \prod_{i=1}^{n} \frac{f(X_i;\theta_2)}{f(X_i;\theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^n \exp\left(\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)\sum_{i=1}^{n} X_i^2\right)$$

Let $T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2$. The ratio can be written as:

$$\frac{L(\theta_2)}{L(\theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^n \exp\left(\left(\frac{\theta_2 - \theta_1}{\theta_1\theta_2}\right) T(\mathbf{X})\right)$$

Since $\theta_1 < \theta_2$, $\frac{\theta_1}{\theta_2} < 1$ is a constant and $\frac{\theta_2-\theta_1}{\theta_1\theta_2} > 0$

The exponential function $\exp(kT)$ with $k = \frac{\theta_2-\theta_1}{\theta_1\theta_2} > 0$ is strictly increasing in T and the constant C doesn't affect the monotonicity

$$\frac{L(\theta_2)}{L(\theta_1)} = C \cdot e^{kT} \quad (C = (\frac{\theta_1}{\theta_2})^n, k > 0)$$

$$\boxed{\text{The family } \{f(x;\theta) : \theta > 0\} \text{ has MLR in } T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2.}$$

(e) Derive the uniformly most powerful α-size test for $H_0 : \theta \geq 2$ vs $H_1 : \theta < 2$

Neyman-Pearson Lemma Application for testing $\theta_0 \geq 2$ vs $\theta_1 < 2$, the most powerful test rejects when:

$$\frac{L(\theta_1)}{L(\theta_0)} > k \implies T < c$$

due to the MLR property.

Determining critical value, choose c such that:

$$P_{\theta=2}(T \leq c) = \alpha$$

Since $T \sim \chi^2_{2n}$ under $\theta = 2$ for n variables:

$$c = \chi^2_{2n,\alpha}$$

where $\chi^2_{2n,\alpha}$ is the α-quantile of χ^2_{2n}.

The UMP test function is:

$$\phi^*(T) = \begin{cases} 1 & \text{if } T \leq \chi^2_{2n,\alpha} \\ 0 & \text{otherwise} \end{cases}$$

Size of the test, for all $\theta \geq 2$:

$$P_\theta(T \leq c) \leq P_{\theta=2}(T \leq c) = \alpha$$

with equality at $\theta = 2$.

Power function, for $\theta < 2$:

$$\beta_\phi(\theta) = P_\theta(T \leq \chi^2_{2n,\alpha})$$

is maximized among all tests with size α.

The UMP α-size test is:

$$\boxed{\phi^*(T) = \mathbf{1} \qquad , T \leq \chi^2_{2n,\alpha\}}}$$

(f) Power Function Derivation

The power function is:

$$\begin{aligned}
\pi(\theta) &= E_\theta[\phi^*] = P_\theta(T \leq \chi^2_{2n,\alpha}) \\
&= P_\theta\left(\sum_{i=1}^n X_i^2 \leq \chi^2_{2n,\alpha}\right) \\
&= P\left(\frac{2}{\theta}\sum_{i=1}^n X_i^2 \leq \frac{2}{\theta}\chi^2_{2n,\alpha}\right) \\
&= F_{\chi^2_{2n}}\left(\frac{2\chi^2_{2n,\alpha}}{\theta}\right)
\end{aligned}$$

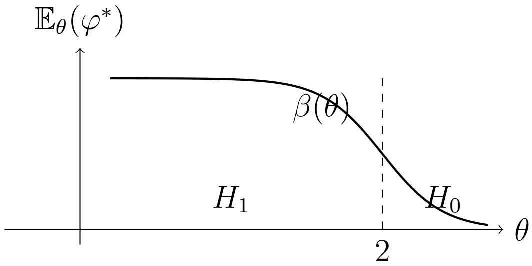
where $F_{\chi^2_{2n}}$ is the CDF of χ^2_{2n}.

Properties of the Power Function:

At $\theta = 2$: $\pi(2) = \alpha$ (exact size)

For $\theta < 2$: $\pi(\theta) > \alpha$ (increasing power as θ decreases)

For $\theta > 2$: $\pi(\theta) < \alpha$ (decreasing to 0 as $\theta \to \infty$)



Problem Two

Let $X = (X_1, X_2, \dots, X_n)$ be a random sample from a geometric distribution with probability mass function:

$$f(x,p) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

where $0 < p < 1$ is the unknown parameter.

(a) Maximum Likelihood Estimation for θ

The joint likelihood function for the sample is:

$$L(p) = \prod_{i=1}^{n} f(X_i, p) = \prod_{i=1}^{n} p(1-p)^{X_i - 1} = p^n (1-p)^{\sum_{i=1}^{n}(X_i - 1)}$$

Taking the natural logarithm to simplify:

$$\ln L(p) = n \ln p + \left(\sum_{i=1}^{n} X_i - n \right) \ln(1-p)$$

Differentiating with respect to p:

$$\frac{\partial \ln L(p)}{\partial p} = \frac{n}{p} - \frac{\sum_{i=1}^{n} X_i - n}{1-p}$$

Set the derivative equal to zero for maximization:

$$\frac{n}{p} - \frac{T-n}{1-p} = 0 \quad \text{where} \quad T = \sum_{i=1}^{n} X_i$$

$$n(1-p) = p(T-n)$$
$$n - np = pT - np$$
$$n = pT$$

Solving for p:

$$\hat{p} = \frac{n}{T} = \frac{1}{\bar{X}}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.
The second derivative is:

$$\frac{\partial^2 \ell(p)}{\partial p^2} = -\frac{n}{p^2} - \frac{T-n}{(1-p)^2} < 0$$

The maximum likelihood estimator for p is:

$$\boxed{\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}}$$

(b) Find the mean of this distribution and the maximum likelihood estimator of the mean

Mean of the Geometric Distribution

The expected value (mean) is calculated as:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x \cdot p(1-p)^{x-1}$$

using the infinite series standard formula for $|r| < 1$:

$$\sum_{x=1}^{\infty} x r^{x-1} = \frac{1}{(1-r)^2}$$

Let $r = 1 - p$. Then:

$$\mathbb{E}[X] = p \sum_{x=1}^{\infty} x(1-p)^{x-1} = p \cdot \frac{1}{[1-(1-p)]^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

$$\boxed{\mathbb{E}[X] = \frac{1}{p}}$$

Maximum Likelihood Estimator of the Mean

From part (a), we have the MLE for p:

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Using the **invariance property of MLEs**, the MLE of $\mathbb{E}[X] = \frac{1}{p}$ is:

$$\widehat{\mathbb{E}[X]} = \frac{1}{\hat{p}} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

where $\bar{X}$ is the sample mean. The sample mean $\bar{X}$ is a natural estimator for the population mean. As $n \to \infty$, by the Law of Large Numbers, $\bar{X} \to \mathbb{E}[X]$

$$\boxed{\widehat{\mathbb{E}[X]} = \bar{X}}$$

(c) Showing Monotone Likelihood Ratio in $T = -\sum_{i=1}^{n} X_i$

For any two parameters $p_1 < p_2$, the likelihood ratio is:

$$\frac{L(p_2)}{L(p_1)} = \frac{\prod_{i=1}^{n} p_2(1-p_2)^{X_i-1}}{\prod_{i=1}^{n} p_1(1-p_1)^{X_i-1}} = \left(\frac{p_2}{p_1}\right)^n \left(\frac{1-p_2}{1-p_1}\right)^{\sum_{i=1}^{n}(X_i-1)}$$

Let $T = -\sum_{i=1}^{n} X_i = -S$, where $S = \sum_{i=1}^{n} X_i$. Then:

$$\frac{L(p_2)}{L(p_1)} = \left(\frac{p_2}{p_1}\right)^n \left(\frac{1-p_2}{1-p_1}\right)^{-T-n}$$

Taking natural logarithms:

$$\ln\left(\frac{L(p_2)}{L(p_1)}\right) = n\ln\left(\frac{p_2}{p_1}\right) + (-T-n)\ln\left(\frac{1-p_2}{1-p_1}\right)$$

Since $p_2 > p_1$:

- $\frac{p_2}{p_1} > 1 \Rightarrow \ln\left(\frac{p_2}{p_1}\right) > 0$
- $\frac{1-p_2}{1-p_1} < 1 \Rightarrow \ln\left(\frac{1-p_2}{1-p_1}\right) < 0$

Let $k = -\ln\left(\frac{1-p_2}{1-p_1}\right) > 0$. Then:

$$\ln\left(\frac{L(p_2)}{L(p_1)}\right) = n\left[\ln\left(\frac{p_2}{p_1}\right) + \ln\left(\frac{1-p_2}{1-p_1}\right)\right] + kT$$

The term in brackets is constant with respect to T, and $k > 0$. Therefore:

$$\frac{L(p_2)}{L(p_1)} = C \cdot e^{kT}$$

where $C > 0$ is a constant. This expression is strictly increasing in T.

The likelihood ratio is a strictly increasing function of $T = -\sum_{i=1}^{n} X_i$ for any $p_2 > p_1$. Therefore:

$$\boxed{\text{The family } \{L(X,p)\}_{0<p<1} \text{ has a monotone likelihood ratio in the statistic } T = -\sum_{i=1}^{n} X_i.}$$

(d) Uniformly Most Powerful Test for $H_0 : p \leq 0.3$ vs $H_1 : p > 0.3$

For testing:

$$H_0 : p \leq 0.3 \quad \text{vs} \quad H_1 : p > 0.3$$

the UMP α-level test rejects H_0 for small values of $\sum X_i$.

The general form of the test function ϕ^* is:

$$\phi^*(X) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i < c \\ \gamma & \text{if } \sum_{i=1}^{n} X_i = c \\ 0 & \text{if } \sum_{i=1}^{n} X_i > c \end{cases}$$

(e) Large Sample Test Using CLT

For testing $\tilde{H}_0 : p = 0.3$ vs $\tilde{H}_1 : p \neq 0.3$,

The maximum likelihood estimator is:

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{30}{120} = 0.25$$

Under H_0, the standard error of $\hat{p}$ is:

$$\text{SE}(\hat{p}) = \sqrt{\frac{p_0^2(1-p_0)}{n}} = \sqrt{\frac{(0.3)^2(0.7)}{30}} \approx 0.0458$$

The test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0^2(1-p_0)}{n}}} = \frac{0.25 - 0.3}{\sqrt{\frac{(0.3)^2(1-0.3)}{30}}} \approx -1.09$$

For $\alpha = 0.05$ (two-tailed), the critical values are ± 1.96. Since $|Z| = 1.09 < 1.96$, we fail to reject $\tilde{H}_0$.

$$\boxed{\text{Do not reject } \tilde{H}_0 \text{ at } \alpha = 0.05.}$$

Problem Three

$$f(x, \theta) = \frac{1}{\beta} \exp\left(-\frac{x-\theta}{\beta}\right), \quad \theta < x < \infty$$

where $\beta > 0$ is known and θ is an unknown location parameter. We have a sample $\mathbf{X} = (X_1, \ldots, X_n)$ of n i.i.d. observations.

(i) Compute the CDF and PDF for $T = X_{(1)}$

The cumulative distribution function for one observation X is:

$$F_X(x) = P(X \leq x) = \int_\theta^x \frac{1}{\beta} \exp\left(-\frac{t-\theta}{\beta}\right) dt$$

after integration

$$F_X(x) = 1 - \exp\left(-\frac{x-\theta}{\beta}\right), \quad x \geq \theta$$

and $F_X(x) = 0$ for $x < \theta$.

For $T = X_{(1)} = \min(X_1, \ldots, X_n)$, the CDF is:

$$F_T(t) = 1 - \Pr(\text{all } X_i > t) = 1 - [1 - F_X(t)]^n$$

Substituting $F_X(t)$:

$$F_T(t) = \begin{cases} 1 - \exp\left(-\frac{n(t-\theta)}{\beta}\right) & \text{for } t \geq \theta \\ 0 & \text{for } t < \theta \end{cases}$$

Differentiating $F_T(t)$ with respect to t:

$$f_T(t) = \frac{d}{dt}F_T(t) = \begin{cases} \frac{n}{\beta}\exp\left(-\frac{n(t-\theta)}{\beta}\right) & \text{for } t \geq \theta \\ 0 & \text{for } t < \theta \end{cases}$$

The cumulative distribution function of $T = X_{(1)}$ is:

$$\boxed{F_T(t) = \begin{cases} 1 - \exp\left(-\frac{n(t-\theta)}{\beta}\right) & \text{for } t \geq \theta \\ 0 & \text{otherwise} \end{cases}}$$

The probability density function of $T = X_{(1)}$ is:

$$\boxed{f_T(t) = \begin{cases} \frac{n}{\beta}\exp\left(-\frac{n(t-\theta)}{\beta}\right) & \text{for } t \geq \theta \\ 0 & \text{otherwise} \end{cases}}$$

(ii) MLR Property in $T = X_{(1)}$

For a sample $\mathbf{X} = (X_1, \ldots, X_n)$, the joint density is:

$$L(\theta; \mathbf{X}) = \prod_{i=1}^{n} f(X_i, \theta) = \left(\frac{1}{\beta^n}\right)\exp\left(-\frac{1}{\beta}\sum_{i=1}^{n}(X_i - \theta)\right) \cdot I(X_{(1)} > \theta)$$

where $I(\cdot)$ is the indicator function.

For $\theta_1 < \theta_2$, the ratio is:

$$\frac{L(\theta_2; \mathbf{X})}{L(\theta_1; \mathbf{X})} = \exp\left(\frac{n(\theta_2 - \theta_1)}{\beta}\right) \cdot \frac{I(X_{(1)} > \theta_2)}{I(X_{(1)} > \theta_1)}$$

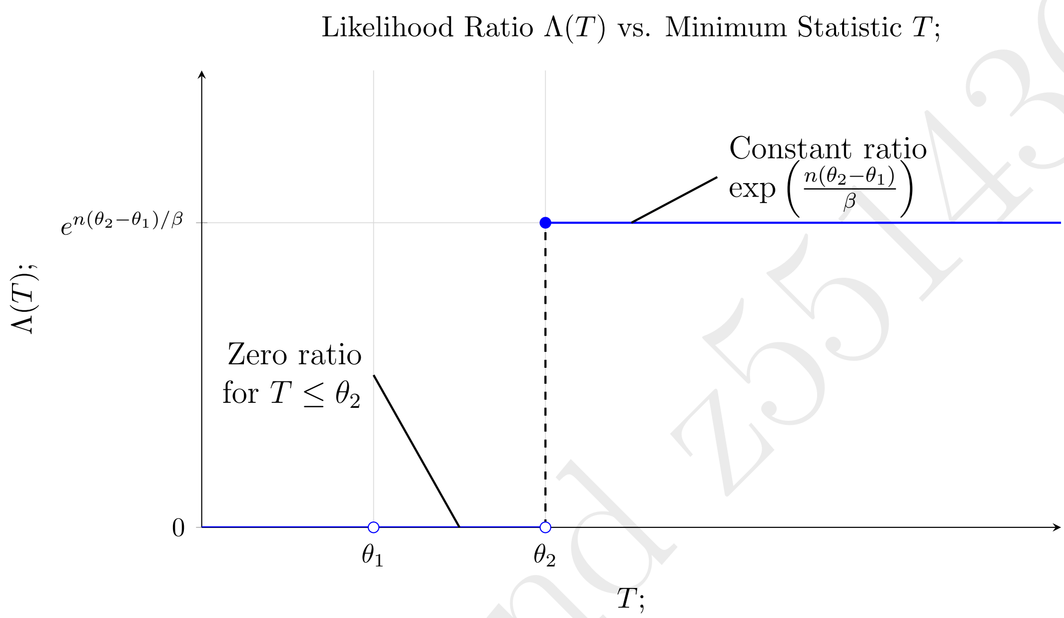
Possible cases:

- When $X_{(1)} > \theta_2$: $\frac{L(\theta_2;\mathbf{X})}{L(\theta_1;\mathbf{X})} = \exp\left(\frac{n(\theta_2-\theta_1)}{\beta}\right) > 1$
- When $\theta_1 < X_{(1)} \leq \theta_2$: $\frac{L(\theta_2;\mathbf{X})}{L(\theta_1;\mathbf{X})} = 0$
- When $X_{(1)} \leq \theta_1$: Undefined (zero support)

The likelihood ratio is:

$$\Lambda(T) = \begin{cases} \exp\left(\frac{n(\theta_2-\theta_1)}{\beta}\right) & T > \theta_2 \\ 0 & \theta_1 < T \leq \theta_2 \\ \text{undefined} & T \leq \theta_1 \end{cases}$$

This is a non-decreasing function of $T = X_{(1)}$, proving the MLR property.



$$\boxed{\text{The family has the MLR property in the statistic } T = X_{(1)}.}$$

(iii) Justification for UMP Test

The density $f(x, \theta) = e^{-(x-\theta)}$ (for $x > \theta$, $\beta = 1$) has the monotone likelihood ratio (MLR) property in the statistic $T = X_{(1)}$. If a family has MLR , then a uniformly most powerful (UMP) test exists for one-sided hypotheses $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$. The UMP α-size test rejects H_0 for small values of T, i.e., when $X_{(1)} < c$, where c is chosen such that:

$$\sup_{\theta \geq \theta_0} P(T < c \mid \theta) = \alpha.$$

The supremum is attained at $\theta = \theta_0$, since for $\theta > \theta_0$, $P(T < c \mid \theta) \leq P(T < c \mid \theta_0)$.

The CDF of T is:

$$F_T(t \mid \theta) = 1 - e^{-n(t-\theta)} \quad \text{for } t \geq \theta$$

PDF is:

$$f_T(t \mid \theta) = ne^{-n(t-\theta)} \quad \text{for } t \geq \theta$$

Under H_0 with $\theta = \theta_0$, the CDF becomes:

$$P(T < c \mid \theta_0) = 1 - e^{-n(c-\theta_0)}.$$

To control the Type I error at $\alpha = 0.05$, solve:

$$1 - e^{-n(c-\theta_0)} = \alpha \implies e^{-n(c-\theta_0)} = 1 - \alpha$$

Taking the natural logarithm:

$$-n(c - \theta_0) = \ln(1 - \alpha) \implies c = \theta_0 - \frac{1}{n}\ln(1 - \alpha)$$

Substitute $\theta_0 = 2$, $n = 4$, $\beta = 1$, and $\alpha = 0.05$:

$$c = 2 - \frac{1}{4}\ln(1 - 0.05) = 2 - \frac{1}{4}\ln(0.95)$$

$$c = 2 - \frac{1}{4}(-0.051293) = 2 + 0.012823 \approx 2.012823.$$

The UMP α-size test rejects $H_0 : \theta \geq 2$ if the minimum observation $X_{(1)} < 2.013$. The threshold constant is:

$$\boxed{c \approx 2.013}$$

(iv) Power Function of UMP Test

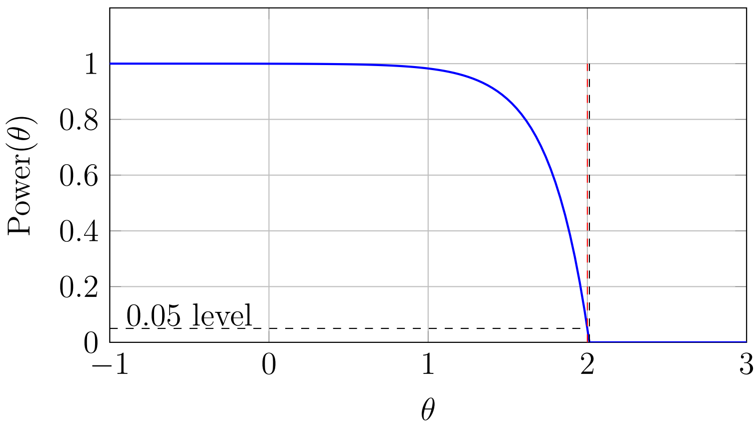
For testing $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$, we have:

Critical Region reject H_0 when $T < c$ where

$$c = \theta_0 - \frac{\beta}{n}\ln(1 - \alpha)$$

Power Function:

$$\begin{aligned}
\text{Power}(\theta) &= P(\text{Reject } H_0 \mid \theta) \\
&= P(T < c \mid \theta) \\
&= 1 - e^{-n(c-\theta)/\beta} \quad \text{for } \theta \leq c \\
&= 1 - (1 - \alpha)e^{-n(\theta_0-\theta)/\beta} \quad \text{for } \theta \leq \theta_0 \\
&= 0 \quad \text{for } \theta > c
\end{aligned}$$



(v) Hypothesis Test

The UMP test uses $T = X_{(1)} = \min(\mathbf{x})$. Observed value: $T_{\text{obs}} = \min(1.1, 2.0, 1.3, 3.1) = 1.1$

Under H_0 ($\theta = 1$), the CDF of T is:

$$P(T \leq t \mid \theta = 1) = 1 - e^{-\frac{n(t-\theta)}{\beta}} = 1 - e^{-2(t-1)} \quad \text{for } t \geq 1$$

Solve for c such that $P(T < c \mid \theta = 1) = 0.05$:

$$1 - e^{-2(c-1)} = 0.05 \implies e^{-2(c-1)} = 0.95 \implies -2(c-1) = \ln(0.95)$$

$$c = 1 - \frac{1}{2}\ln(0.95) \approx 1 - \frac{1}{2}(-0.0513) \approx 1.0257$$

Reject H_0 if $T_{\text{obs}} < c$ Observed: $1.1 \not< 1.0257 \implies$ Fail to reject H_0

$$p\text{-value} = P(T \leq 1.1 \mid \theta = 1) = 1 - e^{-2(1.1-1)} = 1 - e^{-0.2} \approx 0.1813$$

Since p-value > 0.05, we fail to reject H_0.

There is insufficient evidence at the 0.05 significance level to conclude that $\theta < 1$.

$$\boxed{\text{Fail to reject } H_0}$$

(vi) Distribution of $Z_n = n(X_{(1)} - \theta)$

The CDF of the minimum order statistic $X_{(1)}$ is:

$$F_{X_{(1)}}(x) = 1 - [1 - F_X(x)]^n = 1 - e^{-n(x-\theta)/\beta}, \quad x \geq \theta$$

with corresponding PDF:

$$f_{X_{(1)}}(x) = \frac{n}{\beta} e^{-n(x-\theta)/\beta}, \quad x \geq \theta$$

Let $Z_n = n(X_{(1)} - \theta)$. The CDF of Z_n is:

$$\begin{aligned} F_{Z_n}(z) &= P(Z_n \le z) \\ &= P(n(X_{(1)} - \theta) \le z) \\ &= P\left(X_{(1)} \le \theta + \frac{z}{n}\right) \\ &= 1 - e^{-z/\beta}, \quad z \ge 0 \end{aligned}$$

The PDF of Z_n is obtained by differentiating the CDF:

$$f_{Z_n}(z) = \frac{d}{dz} F_{Z_n}(z) = \frac{1}{\beta} e^{-z/\beta}, \quad z \ge 0$$

This is the PDF of an exponential distribution with rate parameter $1/\beta$.

The random variable Z_n follows an exponential distribution that does not depend on n:

$$\boxed{Z_n \sim \text{Exponential}(\beta)}$$

(vii) Consistency of $X_{(1)}$ as an Estimator for θ

The CDF of $X_{(1)}$ is:

$$F_{X_{(1)}}(x) = 1 - [1 - F_X(x)]^n = 1 - e^{-n(x-\theta)/\beta}, \quad x \ge \theta,$$

where $F_X(x) = 1 - e^{-(x-\theta)/\beta}$ is the CDF of each X_i.

We show that $X_{(1)}$ converges in probability to θ:

$$X_{(1)} \xrightarrow{P} \theta \quad \text{as } n \to \infty.$$

For any $\epsilon > 0$:

$$P\left(|X_{(1)} - \theta| > \epsilon\right) = P\left(X_{(1)} > \theta + \epsilon\right),$$

since $X_{(1)} \ge \theta$

Using the CDF:

$$P\left(X_{(1)} > \theta + \epsilon\right) = 1 - F_{X_{(1)}}(\theta + \epsilon) = e^{-n\epsilon/\beta}.$$

As $n \to \infty$:

$$e^{-n\epsilon/\beta} \to 0.$$

Thus:

$$\lim_{n\to\infty} P\left(|X_{(1)} - \theta| > \epsilon\right) = 0.$$

Since $X_{(1)}$ converges in probability to θ, it is a consistent estimator:

$$\boxed{X_{(1)} \text{ is a consistent estimator of } \theta.}$$

Problem 4

1. The density of the midrange B

For ordered statistics $X_{(1)} < X_{(2)} < X_{(3)}$, the joint density is:

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x, x_2, y) = 3! \cdot f(x)f(x_2)f(y) = 6e^{-x}e^{-x_2}e^{-y}$$

where $0 < x < x_2 < y$.The factor $6 = 3!$ accounts for all possible orderings of the three observations

To find the joint density of just $X_{(1)}$ and $X_{(3)}$, we integrate over x_2:

$$f_{X_{(1)},X_{(3)}}(x, y) = \int_x^y 6e^{-x}e^{-x_2}e^{-y}\,dx_2 = 6e^{-x}e^{-y}\int_x^y e^{-x_2}\,dx_2$$

Compute the integral:

$$\int_x^y e^{-x_2}\,dx_2 = \left[-e^{-x_2}\right]_x^y = e^{-x} - e^{-y}$$

Substituting back:

$$f_{X_{(1)},X_{(3)}}(x, y) = 6e^{-x}e^{-y}(e^{-x} - e^{-y}) = 6(e^{-2x-y} - e^{-x-2y})$$

Joint Density of $X_{(1)}$ and $X_{(3)}$:

$$f_{X_{(1)},X_{(3)}}(x, y) = 6\left(e^{-2x-y} - e^{-x-2y}\right) \quad \text{for } 0 < x < y$$

Let $B = \frac{x+y}{2}$, which implies $y = 2B - x$. The limits for x become $0 < x < B$. Substitute $y = 2B - x$ into the joint density and integrate over x:

$$f_B(u) = 2\int_0^u 6\left(e^{-2x-(2u-x)} - e^{-x-2(2u-x)}\right)\,dx.$$

Simplify the exponents:

$$f_B(u) = 12e^{-2u}\int_0^u e^{-x}\,dx - 12e^{-4u}\int_0^u e^{x}\,dx.$$

integrating:

$$f_B(u) = 12e^{-2u}\left(1 - e^{-u}\right) - 12e^{-4u}\left(e^{u} - 1\right).$$

simplify:

$$f_B(u) = 12e^{-2u}\left(1 - e^{-u}\right) - 12e^{-3u} + 12e^{-4u}.$$

Factor to recognize the square:

$$f_B(u) = 12e^{-2u}\left(1 - 2e^{-u} + e^{-2u}\right) = 12e^{-2u}\left(1 - e^{-u}\right)^2.$$

The density of the midrange B is:

$$\boxed{12e^{-2u}(1 - e^{-u})^2} \quad \text{for } u > 0.$$

2. Showing that $P(B > 1) = 0.4687$.

The density of B is:

$$f_B(u) = 12e^{-2u}(1 - e^{-u})^2 \quad \text{for } u > 0.$$

Compute $P(B > 1)$:

$$P(B > 1) = \int_1^\infty f_B(u)\, du = 12 \int_1^\infty e^{-2u}(1 - e^{-u})^2\, du.$$

Expand $(1 - e^{-u})^2$:

$$(1 - e^{-u})^2 = 1 - 2e^{-u} + e^{-2u},$$

so the integer and becomes:

$$e^{-2u} - 2e^{-3u} + e^{-4u}.$$

Integrate term by term:

$$\int_1^\infty e^{-2u}\, du = \frac{e^{-2}}{2},$$
$$\int_1^\infty e^{-3u}\, du = \frac{e^{-3}}{3},$$
$$\int_1^\infty e^{-4u}\, du = \frac{e^{-4}}{4}.$$

Combine the results:

$$P(B > 1) = 12\left(\frac{e^{-2}}{2} - 2 \cdot \frac{e^{-3}}{3} + \frac{e^{-4}}{4}\right) = 6e^{-2} - 8e^{-3} + 3e^{-4}.$$

Sum:

$$P(B > 1) = 0.8118 - 0.3984 + 0.0549 = 0.4683$$

$$\boxed{P(B > 1) = 0.4687}$$

3. Show $\mathbf{Cov}(X_{(1)}, X_{(3)}) = \frac{1}{9}$

The covariance between the minimum $X_{(1)}$ and maximum $X_{(3)}$ is given by:

$$\text{Cov}(X_{(1)}, X_{(3)}) = E[X_{(1)}X_{(3)}] - E[X_{(1)}]E[X_{(3)}]$$

From Problem 1, we have the joint density of $X_{(1)}$ and $X_{(3)}$:

$$f_{X_{(1)},X_{(3)}}(x, y) = 6\left(e^{-2x-y} - e^{-x-2y}\right), \quad 0 < x < y$$

$$E[X_{(1)}X_{(3)}] = \int_0^\infty \int_x^\infty xy \cdot 6\left(e^{-2x-y} - e^{-x-2y}\right) dydx$$

We'll evaluate this as the difference of two integrals:

$$= 6\left(\underbrace{\int_0^\infty \int_x^\infty xye^{-2x-y}dydx}_{I_1} - \underbrace{\int_0^\infty \int_x^\infty xye^{-x-2y}dydx}_{I_2}\right)$$

For I_1 first integrate with respect to y:

$$\int_x^\infty ye^{-y}dy = e^{-x}(x+1)$$

Then the integral becomes:

$$I_1 = \int_0^\infty xe^{-2x} \cdot e^{-x}(x+1)dx = \int_0^\infty (x^2+x)e^{-3x}dx$$

Using integration by parts:

$$\int_0^\infty x^2e^{-3x}dx = \frac{2}{27}, \quad \int_0^\infty xe^{-3x}dx = \frac{1}{9}$$

Thus:

$$I_1 = \frac{2}{27} + \frac{1}{9} = \frac{5}{27}$$

For I_2 First integrate with respect to y:

$$\int_x^\infty ye^{-2y}dy = \frac{e^{-2x}(2x+1)}{4}$$

Then the integral becomes:

$$I_2 = \int_0^\infty xe^{-x} \cdot \frac{e^{-2x}(2x+1)}{4}dx = \frac{1}{4}\int_0^\infty (2x^2+x)e^{-3x}dx$$

Again using integration by parts:

$$\int_0^\infty 2x^2e^{-3x}dx = \frac{4}{27}, \quad \int_0^\infty xe^{-3x}dx = \frac{1}{9}$$

Thus:

$$I_2 = \frac{1}{4}\left(\frac{4}{27} + \frac{1}{9}\right) = \frac{7}{108}$$

$$E[X_{(1)}X_{(3)}] = 6\left(\frac{5}{27} - \frac{7}{108}\right) = 6\left(\frac{20}{108} - \frac{7}{108}\right) = \frac{13}{18}$$

For exponential order statistics:

$$E[X_{(1)}] = \frac{1}{3}, \quad E[X_{(3)}] = \frac{1}{3} + \frac{1}{2} + 1 = \frac{11}{6}$$

$$\mathrm{Cov}(X_{(1)}, X_{(3)}) = \frac{13}{18} - \left(\frac{1}{3} \times \frac{11}{6}\right) = \frac{13}{18} - \frac{11}{18} = \frac{2}{18} = \frac{1}{9}$$

$$\boxed{\mathrm{Cov}(X_{(1)}, X_{(3)}) = \frac{1}{9}}$$