Mathematical Statistics

Tutorial 2

1. Given a random sample (r.s.) $X_1, \ldots, X_n$ from a Uniform distribution $\mathcal{U}(0, \theta)$, let $\hat{\theta}_n$ be $\hat{\theta}_n = \max(X_1, \ldots, X_n) = X_{(n)}$, and $\tilde{\theta}_n = 2\bar{X}_n$.

   (a) Prove that $\tilde{\theta}_n$ is unbiased and that $\hat{\theta}_n$ is asymptotically unbiased.

   (b) Calculate the MSE (Mean Squared Error) for both estimators. Which estimator would you prefer based on the MSE criterion?

2. Let X denote the proportion of assigned time that a randomly selected student spends working on a certain aptitude test, and suppose the probability density function (PDF) of X is:

$$f(x; \theta) = \begin{cases} (\theta + 1) x^{\theta} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

   where θ has an unknown value known to be > -1.

   A random sample of ten students yields the following information: 0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.73, 0.97, 0.94, 0.77.

   Use the method of moments and maximum likelihood to obtain two estimator for θ and then calculate the value of the estimator for the observed data.

Given a random sample (r.s.) $X_1, \ldots, X_n$ from a Uniform distribution $\mathcal{U}(0, \theta)$, let $\hat{\theta}_n$ be $\hat{\theta}_n = \max(X_1, \ldots, X_n) = X_{(n)}$, and $\tilde{\theta}_n = 2\bar{X}_n$.

1. Prove that $\tilde{\theta}_n$ is unbiased and that $\hat{\theta}_n$ is asymptotically unbiased.
2. Calculate the MSE (Mean Squared Error) for both estimators. Which estimator would you prefer based on the MSE criterion?

Given a random variable X from a Uniform distribution $\mathcal{U}(0, \theta)$,he population mean is $E[X] = \frac{0+\theta}{2} = \frac{\theta}{2}$. The population variance is $\text{Var}(X) = \frac{(\theta-0)^2}{12} = \frac{\theta^2}{12}$.

$\tilde{\theta}_n$ is Unbiased

An estimator is unbiased if $E[\tilde{\theta}_n] = \theta$. Using the linearity of expectation:

$$E[\tilde{\theta}_n] = E\left[2\bar{X}_n\right] = 2E\left[\bar{X}_n\right]$$

Since $E[\bar{X}_n] = E[X]$:

$$E[\tilde{\theta}_n] = 2E[X] = 2\left(\frac{\theta}{2}\right) = \theta$$

Since $E[\tilde{\theta}_n] = \theta$, $\tilde{\theta}_n$ is **unbiased**.

$\hat{\theta}_n$ is Asymptotically Unbiased

Step 1: Determine the PDF of $X_{(n)}$ For a continuous distribution, the PDF of the maximum order statistic $X_{(n)}$ is given by:

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$$

For $X \sim \mathcal{U}(0, \theta)$:

- The CDF is: $F(x) = P(X \leq x) = \frac{x}{\theta}$, for $0 < x < \theta$.
- The PDF is: $f(x) = \frac{1}{\theta}$, for $0 < x < \theta$.

Substituting these into the formula:

$$f_{X_{(n)}}(x) = n\left(\frac{x}{\theta}\right)^{n-1}\left(\frac{1}{\theta}\right) = \frac{nx^{n-1}}{\theta^n}, \quad \text{for } 0 < x < \theta$$

Step 2: Calculate the Expected Value $E[\hat{\theta}_n]$

$$\begin{aligned} E[\hat{\theta}_n] &= \int_0^\theta x \cdot f_{X_{(n)}}(x) dx = \int_0^\theta x \cdot \frac{n}{\theta^n} x^{n-1} dx = \frac{n}{\theta^n} \int_0^\theta x^n dx \\ &= \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{\theta^n}\left(\frac{\theta^{n+1}}{n+1}\right) = \frac{n}{n+1}\theta \end{aligned}$$

The expected value of the maximum order statistic $X_{(n)}$ is:

$$E[\hat{\theta}_n] = E[X_{(n)}] = \frac{n}{n+1}\theta$$

Since $E[\hat{\theta}_n] \neq \theta$, the estimator is biased. However, we check for asymptotic unbiasedness by taking the limit as $n \to \infty$:

$$\lim_{n\to\infty} E[\hat{\theta}_n] = \lim_{n\to\infty}\left(\frac{n}{n+1}\theta\right) = \lim_{n\to\infty}\left(\frac{1}{1+1/n}\theta\right) = \theta$$

Since $\lim_{n\to\infty} E[\hat{\theta}_n] = \theta$, $\hat{\theta}_n$ is **asymptotically unbiased**.

Calculation of MSE and Comparison

The Mean Squared Error (MSE) is $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$.

MSE for MME ($\tilde{\theta}_n$)

- **Bias**: $\text{Bias}(\tilde{\theta}_n) = 0$
- **Variance**:

$$\text{Var}(\tilde{\theta}_n) = \text{Var}(2\bar{X}_n) = 4\text{Var}(\bar{X}_n) = 4\left(\frac{\text{Var}(X)}{n}\right) = 4\left(\frac{\theta^2/12}{n}\right) = \frac{\theta^2}{3n}$$

- **MSE**:

$$\text{MSE}(\tilde{\theta}_n) = \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n}$$

MSE for MLE ($\hat{\theta}_n$)

- **Bias**:

$$\text{Bias}(\hat{\theta}_n) = E[\hat{\theta}_n] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1}$$

- **Variance**: First, we calculate $E[\hat{\theta}_n^2]$:

$$E[\hat{\theta}_n^2] = \int_0^\theta x^2 \cdot \frac{n}{\theta^n} x^{n-1} dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2}$$

  Then, the Variance is:

$$\text{Var}(\hat{\theta}_n) = E[\hat{\theta}_n^2] - (E[\hat{\theta}_n])^2 = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}$$

- **MSE**:

$$\begin{aligned}
\text{MSE}(\hat{\theta}_n) &= \text{Var}(\hat{\theta}_n) + (\text{Bias}(\hat{\theta}_n))^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{\theta^2}{(n+1)^2} \\
&= \frac{\theta^2}{(n+1)^2}\left[\frac{n}{n+2} + 1\right] \\
&= \frac{2\theta^2}{(n+1)(n+2)}
\end{aligned}$$

Comparison and Preference

We compare the two MSE values:

$$\text{MSE}(\tilde{\theta}_n) = \tfrac{\theta^2}{3n} \quad \text{versus} \quad \text{MSE}(\hat{\theta}_n) = \tfrac{2\theta^2}{(n+1)(n+2)}$$

Ignoring the common term θ^2, we compare the coefficients:

$$\frac{1}{3n} \quad \text{versus} \quad \frac{2}{(n+1)(n+2)}$$

$\text{MSE}(\hat{\theta}_n) < \text{MSE}(\tilde{\theta}_n)$ if and only if:

$$\begin{aligned}
\frac{2}{(n+1)(n+2)} &< \frac{1}{3n} \\
6n &< (n+1)(n+2) \\
6n &< n^2 + 3n + 2 \\
0 &< n^2 - 3n + 2 \\
0 &< (n-1)(n-2)
\end{aligned}$$

This inequality holds true for any sample size $n > 2$.

Let X denote the proportion of assigned time that a randomly selected student spends working on a certain aptitude test, and suppose the probability density function (PDF) of X is:

$$f(x;\theta) = \begin{cases} (\theta+1)\,x^{\theta} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where θ has an unknown value known to be > -1.

A random sample of ten students yields the following information: 0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.73, 0.97, 0.94, 0.77.

Use the method of moments and maximum likelihood to obtain two estimator for θ and then calculate the value of the estimator for the observed data.

Method of Moments Estimator

The MOM is found by equating the first population moment ($E[X]$) to the first sample moment ($\bar{X}$).

Derivation of $E[X]$

$$\begin{aligned} E[X] &= \int_0^1 x \cdot f(x;\theta)dx \\ &= \int_0^1 x(\theta+1)x^{\theta}dx \\ &= (\theta+1)\int_0^1 x^{\theta+1}dx \\ &= (\theta+1)\left[\frac{x^{\theta+2}}{\theta+2}\right]_0^1 \\ E[X] &= \frac{\theta+1}{\theta+2} \end{aligned}$$

Setting $E[X] = \bar{X}$:

$$\bar{X} = \frac{\theta+1}{\theta+2}$$

Solving for θ:

$$\begin{aligned} \bar{X}(\theta+2) &= \theta+1 \\ \bar{X}\theta + 2\bar{X} &= \theta+1 \\ \bar{X}\theta - \theta &= 1 - 2\bar{X} \\ \theta(\bar{X}-1) &= 1 - 2\bar{X} \\ \tilde{\theta}_n &= \frac{1-2\bar{X}}{\bar{X}-1} = \frac{2\bar{X}-1}{1-\bar{X}} \end{aligned}$$

Numerical Calculation of MME

First, calculate the sample mean ($\bar{x}$):

$$\sum x_i = 0.92 + 0.79 + 0.90 + 0.65 + 0.86 + 0.47 + 0.73 + 0.97 + 0.94 + 0.77 = 8.00$$

$$\bar{x} = \frac{8.00}{10} = 0.80$$

Now, substitute $\bar{x}$ into the MME formula:

$$\tilde{\theta}_{10} = \frac{2(0.80) - 1}{1 - 0.80} = \frac{1.60 - 1}{0.20} = \frac{0.60}{0.20} = 3.00$$

The MME estimate is $\tilde{\theta}_{\mathbf{10}} = \mathbf{3.00}$.

Maximum Likelihood Estimator

The MLE is found by maximizing the likelihood function, which is often easier by maximizing the log-likelihood function.

Derivation of $\ln L(\theta)$

The likelihood function $L(\theta)$ is the product of the PDFs:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} (\theta+1)x_i^{\theta} = (\theta+1)^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}$$

The log-likelihood function $\ln L(\theta)$ is:

$$\ln L(\theta) = n\ln(\theta+1) + \theta \sum_{i=1}^{n} \ln(x_i)$$

MLE Formula $\hat{\theta}_n$

We maximize $\ln L(\theta)$ by setting its derivative with respect to θ to zero:

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n} \ln(x_i) = 0$$

Solving for θ:

$$\begin{aligned}
\frac{n}{\hat{\theta}_n + 1} &= -\sum_{i=1}^{n} \ln(x_i) \\
\hat{\theta}_n + 1 &= \frac{-n}{\sum_{i=1}^{n} \ln(x_i)} \\
\hat{\theta}_n &= -1 - \frac{n}{\sum_{i=1}^{n} \ln(x_i)} \\
\hat{\theta}_n &= -1 + \frac{n}{-\sum_{i=1}^{n} \ln(x_i)}
\end{aligned}$$

Let $T = -\frac{1}{n}\sum_{i=1}^{n} \ln(x_i)$ be the sample mean of the negative log-likelihood terms.

$$\hat{\theta}_n = \frac{1}{T} - 1$$

Numerical Calculation of MLE

First, calculate $\sum \ln(x_i)$:

$$\sum \ln(x_i) \approx \ln(0.92) + \ln(0.79) + \cdots + \ln(0.77) \approx -2.0494$$

Now, substitute the value into the MLE formula ($n = 10$):

$$\hat{\theta}_{10} \approx -1 - \frac{10}{-2.0494} \approx -1 + 4.8795$$

$$\hat{\theta}_{10} \approx 3.8795$$

The MLE estimate is $\hat{\theta}_{\mathbf{10}} \approx \mathbf{3.880}$.