Mathematical Statistics

Tutorial 1

1. Prove that the Multinomial distribution $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$ is a $\boldsymbol{k-1}$ **parameter exponential family**.

2. Let X be a random variable wit Probability Density Function (PDF) given by

$$f(x \mid \theta) = (1-\theta)e^{x} \cdot \mathbf{I}_{\{x<0\}} + \theta^2 e^{-\theta x} \cdot \mathbf{I}_{\{x \geq 0\}}, \quad \text{where } \theta \in (0,1).$$

   Prove that this distribution is a $2-$ parameter exponential family.

Excersice 1

Prove that the Multinomial distribution $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$ is a **$k-1$ parameter exponential family**. The **Probability Mass Function (PMF)** for $\mathbf{X} \sim \text{Multinomial}(n, \mathbf{p})$ is:

$$f(\mathbf{x}; n, \mathbf{p}) = \binom{n}{\mathbf{x}} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad \text{where } \binom{n}{\mathbf{x}} = \frac{n!}{x_1! x_2! \cdots x_k!}.$$

The Multinomial distribution is the generalization of the Binomial distribution. It models the counts resulting from a fixed number of independent trials (n), where each trial can result in one of k distinct categories.

- **Trials (n):** Fixed number of independent trials.
- **Probabilities ($\mathbf{p}$):** A vector $\mathbf{p} = (p_1, \ldots, p_k)$ where p_j is the probability of outcome j, constrained by $\sum_{j=1}^{k} p_j = 1$.
- **Variable ($\mathbf{X}$):** A random vector of counts $\mathbf{X} = (X_1, \ldots, X_k)$, where X_j is the count for category j, constrained by $\sum_{j=1}^{k} X_j = n$.

Tasks:

1. **Transform** the Multinomial PMF, $f(\mathbf{x}; n, \mathbf{p})$, into the canonical form of a $k-1$ parameter exponential family:
$$f(\mathbf{x}; \boldsymbol{\eta}) = h(\mathbf{x}) \exp\left\{\sum_{j=1}^{k-1} \eta_j T_j(\mathbf{x}) - A(\boldsymbol{\eta})\right\}.$$
2. **Explicitly identify** the following components of the canonical form based on the original parameters ($\mathbf{p}$) and variables ($\mathbf{x}$):
   - The base measure $h(\mathbf{x})$.
   - The vector of statistics $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), \ldots, T_{k-1}(\mathbf{x}))^\top$.
   - The vector of natural parameters $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{k-1})^\top$.
   - The log-normalizer function $A(\boldsymbol{\eta})$, expressed as a function of $\boldsymbol{\eta}$.

Using the k-th category as the reference (p_k and X_k), the Probability Mass Function (PMF) in the canonical form is:

$$\begin{aligned}
f(\mathbf{x}; \boldsymbol{\eta}) &= \binom{n}{\mathbf{x}} \exp\left\{\sum_{j=1}^{k-1} \log(p_j)\, x_j + x_k \log(p_k)\right\} \\
&= \binom{n}{\mathbf{x}} \exp\left\{\sum_{j=1}^{k-1} \log(p_j)\, x_j + (n - \sum_{j=1}^{k-1} x_j) \log(p_k)\right\} \\
&= \underbrace{\binom{n}{\mathbf{x}}}_{h(\mathbf{x})} \exp\left\{\sum_{j=1}^{k-1} \underbrace{\log\left(\frac{p_j}{p_k}\right)}_{\eta_j} \underbrace{x_j}_{T_j(\mathbf{x})} - \underbrace{[-n \log(p_k)]}_{A(\boldsymbol{\eta})}\right\}
\end{aligned}$$

with $\eta = (n, \mathbf{p})$

This is due to the inherent constraint $\sum p_j = 1$, meaning that only $k-1$ parameters are functionally independent.

1. **Vector of Statistics $\mathbf{T}(\mathbf{x})$:** The sufficient statistic is the vector of the first $k-1$ counts:

$$\mathbf{T}(\mathbf{x}) = (X_1, X_2, \ldots, X_{k-1})^\top.$$

2. **Vector of Natural Parameters $\boldsymbol{\eta}$:**

$$\boldsymbol{\eta} = \left(\log\left(\frac{p_1}{p_k}\right), \ldots, \log\left(\frac{p_{k-1}}{p_k}\right)\right)^\top.$$

3. **Log-Normalizer Function $A(\boldsymbol{\eta})$:** Expressing p_k in terms of $\boldsymbol{\eta}$ allows $A(\boldsymbol{\eta})$ to be written solely as a function of the natural parameters:

$$A(\boldsymbol{\eta}) = -n\log(p_k) = n\log\left(1 + \sum_{j=1}^{k-1} e^{\eta_j}\right).$$

since $e^{\eta_j} = \frac{p_j}{p_k}$ and $\sum_{j=1}^{k-1} e^{\eta_j} = \frac{1-p_k}{p_k}$, therefore $1+\sum_{j=1}^{k-1} e^{\eta_j} = \frac{1}{p_k}$ and $\log(1+\sum_{j=1}^{k-1} e^{\eta_j}) = -\log p_k$

Note that the Multinomial distribution can be written using k exponential terms. The PMF is written by incorporating all p_j's into the exponential term:

$$f(\mathbf{x} \mid \boldsymbol{\eta}) = \underbrace{\binom{n}{\mathbf{x}}}_{h(\mathbf{x})} \exp\left\{\sum_{j=1}^{k} \underbrace{\log(p_j)}_{\eta_j} \underbrace{x_j}_{T_j(\mathbf{x})}\right\}.$$

1. **Vector of Statistics $\mathbf{T}(\mathbf{x})$:** The vector includes all k counts:

$$\mathbf{T}(\mathbf{x}) = (X_1, X_2, \ldots, X_k)^\top.$$

2. **Vector of Natural Parameters $\boldsymbol{\eta}$:**

$$\boldsymbol{\eta} = (\log(p_1), \log(p_2), \ldots, \log(p_k))^\top.$$

This k-parameter form is correct but the $k-1$ form is preferred for theoretical statistical analysis.

Excersice 2

Let X be a random variable wit Probability Density Function (PDF) given by

$$f(x \mid \theta) = (1-\theta)e^x \cdot \mathbf{I}_{\{x<0\}} + \theta^2 e^{-\theta x} \cdot \mathbf{I}_{\{x\geq 0\}}, \quad \text{where } \theta \in (0,1).$$

Prove that this distribution is a 2− parameter exponential family.

We aim to write $f(x \mid \theta)$ in the canonical form:

$$f(x \mid \boldsymbol{\eta}) = h(x)\exp\{\eta_1 T_1(x) + \eta_2 T_2(x) - A(\boldsymbol{\eta})\}.$$

Note that

$$f(x \mid \theta) = [(1-\theta)e^x]^{\mathbf{I}_{\{x<0\}}} \left[\theta^2 e^{-\theta x}\right]^{\mathbf{I}_{\{x\geq 0\}}}$$

We use $\mathbf{I}_{\{x<0\}} = 1 - \mathbf{I}_{\{x\geq 0\}}$ and we have that

$$\begin{aligned} f(x \mid \theta) &= \left[(1-\theta)e^{x}\right]^{1-\mathbf{I}_{\{x\geq 0\}}} \left[\theta^2 e^{-\theta x}\right]^{\mathbf{I}_{\{x\geq 0\}}} \\ &= (1-\theta)e^{x} \left[(1-\theta)^{-1} e^{-x}\theta^2 e^{-\theta x}\right]^{\mathbf{I}_{\{x\geq 0\}}} \\ &= (1-\theta)e^{x - x\mathbf{I}_{\{x\geq 0\}}} e^{2\log(\theta/1-\theta)\mathbf{I}_{\{x\geq 0\}} - \theta x \mathbf{I}_{\{x\geq 0\}}} \\ &= e^{x - x\mathbf{I}_{\{x\geq 0\}}} e^{2\log(\theta/1-\theta)\mathbf{I}_{\{x\geq 0\}} - \theta x \mathbf{I}_{\{x\geq 0\}} - \log(1-\theta)} \end{aligned}$$

$$f(x \mid \theta) = \underbrace{e^{x - x\mathbf{I}_{\{x\geq 0\}}}}_{h(x)} e^{\underbrace{2\log(\theta/1-\theta)}_{\eta_1}\underbrace{\mathbf{I}_{\{x\geq 0\}}}_{T_1(x)} + \underbrace{-\theta}_{\eta_2}\underbrace{x\mathbf{I}_{\{x\geq 0\}}}_{T_2(x)} - \underbrace{\log(1-\theta)}_{A(\theta)}}$$

1. $h(x)$**:**

$$h(x) = \exp(x - x \cdot \mathbf{I}_{\{x\geq 0\}}) = \exp(x \cdot \mathbf{I}_{\{x<0\}})$$

2. **Statistics $\mathbf{T}(x)$:**

$$\mathbf{T}(x) = \begin{pmatrix} T_1(x) \\ T_2(x) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{\{x\geq 0\}} \\ x\mathbf{I}_{\{x\geq 0\}} \end{pmatrix}$$

3. **Natural Parameters $\boldsymbol{\eta}$:**

$$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 2\log\left(\frac{\theta}{1-\theta}\right) \\ -\theta \end{pmatrix}$$

4. $A(\boldsymbol{\eta})$:

$$A(\boldsymbol{\eta}) = \log(1-\theta)$$

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:

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

$\mathrm{MSE}(\hat{\theta}_n) < \mathrm{MSE}(\tilde{\theta}_n)$ 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 $\hat{\theta}_{\mathbf{10}} \approx \mathbf{3.880}$.

Mathematical Statistics

Tutorial 3

1. Let $X_1, \ldots, X_n$ be a random sample from a variable with the following probability mass function (PMF):
$$p(x \mid \theta) = \theta^{|x|}(1-2\theta)^{1-|x|} \cdot \cdot \mathbf{I}_{\{-1,0,1\}}(x).$$
Find the Maximum Likelihood Estimator (MLE) of θ.

2. Let $X_1, \ldots, X_n$ be a random sample whose probability density function (PDF) is given by:
$$f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$$
Find the Maximum Likelihood Estimator (MLE) of θ.

3. Let $X_1, \ldots, X_{n_1}$ be an independent random sample from a Normal distribution $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \ldots, Y_{n_2}$ be a second independent random sample, also from a Normal distribution $\mathcal{N}(\mu_2, \sigma^2)$. The two samples are mutually independent and share the same unknown variance σ^2, but have distinct unknown means μ_1 and μ_2.

   Find the Maximum Likelihood Estimators (MLEs) for μ_1, μ_2, and σ^2.

4. Let $X = (X_1, \ldots, X_n)$ be a random sample from a distribution with the following probability density function (PDF):
$$f(x \mid \theta_1, \theta_2) = \begin{cases} \frac{\theta_1}{\theta_2} e^{-x/\theta_2} & \text{if } x \geq 0 \\ \frac{1-\theta_1}{\theta_2} e^{x/\theta_2} & \text{if } x < 0 \end{cases}$$
where $0 < \theta_1 < 1$ and $\theta_2 > 0$.

   (a) Find the Maximum Likelihood Estimators (MLEs) for θ_1 and θ_2.

   (b) Find the MLE for the difference $\theta_1 - \theta_2$.

   (c) Prove the consistency of the MLE for $\theta_1 - \theta_2$.

Problem 1

Let $X_1, \ldots, X_n$ be a random sample from a variable with the following probability mass function (PMF):

$$p(x \mid \theta) = \theta^{|x|}(1-2\theta)^{1-|x|}. \cdot \mathbf{I}_{\{-1,0,1\}}(x).$$

Find the Maximum Likelihood Estimator (MLE) of θ.

Solution

Let n be the sample size. We define N_S as the number of observations where $|X_i| = 1$:

$$N_S = \sum_{i=1}^{n} \mathbf{I}_{\{|X_i|=1\}} = n_{-1} + n_1$$

where n_{-1} is the count of $X_i = -1$ and n_1 is the count of $X_i = 1$. The count of $X_i = 0$ is $n_0 = n - N_S$.

The likelihood function $L(\theta)$ is:

$$L(\theta) = \prod_{i=1}^{n} p(X_i \mid \theta) = \theta^{n_{-1}}(1-2\theta)^{n_0}\theta^{n_1} = \theta^{N_S}(1-2\theta)^{n-N_S}$$

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

$$\ell(\theta) = \ln L(\theta) = N_S \ln(\theta) + (n - N_S)\ln(1-2\theta)$$

To find the MLE $\hat{\theta}$, we differentiate $\ell(\theta)$ with respect to θ and set the derivative equal to zero:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{N_S}{\theta} + (n - N_S)\frac{-2}{1-2\theta} = 0$$

Solving for $\hat{\theta}$:

$$\frac{N_S}{\hat{\theta}} = \frac{2(n-N_S)}{1-2\hat{\theta}}$$

$$N_S(1-2\hat{\theta}) = 2\hat{\theta}(n - N_S)$$

$$N_S - 2N_S\hat{\theta} = 2n\hat{\theta} - 2N_S\hat{\theta}$$

$$N_S = 2n\hat{\theta}$$

The Maximum Likelihood Estimator for θ is:

$$\hat{\theta}_{MLE} = \frac{N_S}{2n}$$

where N_S is the total count of observations where $|X_i| = 1$.

Problem 2

Let $X_1, dots, X_n$ be a random sample whose probability density function (PDF) is given by:

$$f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$$

Find the Maximum Likelihood Estimator (MLE) of θ.

Solution

The likelihood function is the product of the individual PDFs:

$$L(\theta) = \prod_{i=1}^{n} f_{X|\theta}(x_i) = \prod_{i=1}^{n} \left(\frac{2x_i}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x_i) \right)$$

We separate the terms related to the parameter θ from the constant terms (those depending only on the observed sample $x_1, \ldots, x_n$).

$$L(\theta) = \left(\prod_{i=1}^{n} \frac{2x_i}{\theta^2} \right) \cdot \left(\prod_{i=1}^{n} \mathbf{I}_{(0,\theta)}(x_i) \right)$$

Let $C = 2^n \prod_{i=1}^{n} x_i$. The product of the indicator functions imposes the crucial constraint: for the likelihood to be non-zero, every observation x_i must be less than θ. This means θ must be greater than the maximum observation in the sample, $X_{(n)} = \max(X_1, \ldots, X_n)$.

$$L(\theta) = \frac{C}{\theta^{2n}}, \quad \text{subject to the constraint } \theta > X_{(n)}$$

The goal is to find the value of θ that maximizes $L(\theta)$ over the permissible domain $\theta \in (X_{(n)}, \infty)$.

Since C and $2n$ are positive constants, maximizing $L(\theta)$ is equivalent to minimizing the denominator θ^{2n}. We analyze how $L(\theta)$ changes as θ increases in the valid domain:

$$L(\theta) \propto \frac{1}{\theta^{2n}}$$

As θ increases, θ^{2n} increases, and therefore $L(\theta)$ decreases. The function $L(\theta)$ is a **strictly decreasing function** for $\theta > 0$. Since $L(\theta)$ is strictly decreasing on its domain $(X_{(n)}, \infty)$, the maximum value must occur at the smallest possible value of θ, which is the boundary of the domain.

$$\max_{\theta > X_{(n)}} L(\theta) = \lim_{\theta \to X_{(n)}^+} L(\theta) = \frac{C}{(X_{(n)})^{2n}}$$

The value of θ that maximizes the likelihood function is the maximum order statistic.

The Maximum Likelihood Estimator (MLE) for θ is the maximum observation in the sample:

$$\hat{\theta}_{MLE} = X_{(n)} = \max(X_1, X_2, \ldots, X_n)$$

Problem 3

Let $X_1, \ldots, X_{n_1}$ be an independent random sample from a Normal distribution $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \ldots, Y_{n_2}$ be a second independent random sample, also from a Normal distribution $\mathcal{N}(\mu_2, \sigma^2)$. The two samples are mutually independent and share the same unknown variance σ^2, but have distinct unknown means μ_1 and μ_2.

Find the Maximum Likelihood Estimators (MLEs) for μ_1, μ_2, and σ^2.

Solution

The parameter vector to be estimated is $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma^2)$. The total sample size is $N = n_1 + n_2$.

The joint likelihood function is the product of the likelihoods of the two independent samples:

$$L(\boldsymbol{\theta}) = L_X(\mu_1, \sigma^2) \cdot L_Y(\mu_2, \sigma^2)$$

Where L_X and L_Y are:

$$L_X = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n_1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(x_i - \mu_1)^2\right\}$$
$$L_Y = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n_2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n_2}(y_j - \mu_2)^2\right\}$$

Multiplying them together, we get:

$$L(\boldsymbol{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n_1+n_2)/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \sum_{j=1}^{n_2}(y_j - \mu_2)^2\right]\right\}$$

Taking the natural logarithm (ln):

$$\ell(\boldsymbol{\theta}) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + \sum_{j=1}^{n_2}(y_j - \mu_2)^2\right]$$

We find the MLEs for μ_1 and μ_2 by setting the partial derivatives with respect to each mean equal to zero.

$$\frac{\partial \ell}{\partial \mu_1} = -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1} 2(x_i - \mu_1)(-1)\right] = \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(x_i - \mu_1) = 0$$
$$\sum_{i=1}^{n_1} x_i - n_1\hat{\mu}_1 = 0 \implies \hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$$
$$\frac{\partial \ell}{\partial \mu_2} = -\frac{1}{2\sigma^2}\left[\sum_{j=1}^{n_2} 2(y_j - \mu_2)(-1)\right] = \frac{1}{\sigma^2}\sum_{j=1}^{n_2}(y_j - \mu_2) = 0$$
$$\sum_{j=1}^{n_2} y_j - n_2\hat{\mu}_2 = 0 \implies \hat{\mu}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j$$

The MLEs for the means are the sample means: $\hat{\mu}_1 = \bar{X}$ and $\hat{\mu}_2 = \bar{Y}$.

We substitute the MLEs for the means ($\bar{X}$ and $\bar{Y}$) into the log-likelihood function and differentiate with respect to σ^2 (treating it as a single variable $v = \sigma^2$):

$$\frac{\partial \ell}{\partial v} = -\frac{N}{2v} + \frac{1}{2v^2}\left[\sum_{i=1}^{n_1}(x_i - \bar{X})^2 + \sum_{j=1}^{n_2}(y_j - \bar{Y})^2\right] = 0$$

Multiplying by $2v^2$ and solving for $\hat{\sigma}^2 = \hat{v}$:

$$-N\hat{v} + \left[\sum_{i=1}^{n_1}(x_i - \bar{X})^2 + \sum_{j=1}^{n_2}(y_j - \bar{Y})^2\right] = 0$$

$$\hat{\sigma}^2 = \frac{1}{N}\left[\sum_{i=1}^{n_1}(x_i - \bar{X})^2 + \sum_{j=1}^{n_2}(y_j - \bar{Y})^2\right]$$

The Maximum Likelihood Estimators (MLEs) for the parameters are:

$$\hat{\mu}_1 = \bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$$

$$\hat{\mu}_2 = \bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j$$

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2\right]$$

Problem 4

Let $X = (X_1, \dots, X_n)$ be a random sample from a distribution with the following probability density function (PDF):

$$f(x \mid \theta_1, \theta_2) = \begin{cases} \frac{\theta_1}{\theta_2} e^{-x/\theta_2} & \text{if } x \geq 0 \\ \frac{1-\theta_1}{\theta_2} e^{x/\theta_2} & \text{if } x < 0 \end{cases}$$

where $0 < \theta_1 < 1$ and $\theta_2 > 0$.

1. Find the Maximum Likelihood Estimators (MLEs) for θ_1 and θ_2.

2. Find the MLE for the difference $\theta_1 - \theta_2$.

3. Prove the consistency of the MLE for $\theta_1 - \theta_2$.

Solution

1) We partition the sample $X_1, \dots, X_n$ into two subsets based on the sign of the observations:

- S_A: Set of observations where $X_i \geq 0$. Let $n_A = |S_A|$.
- S_B: Set of observations where $X_i < 0$. Let $n_B = |S_B|$.

Note that $n_A + n_B = n$.

The Likelihood Function $L(\theta_1, \theta_2)$ is the product of the densities:

$$L(\theta_1, \theta_2) = \left(\prod_{i \in S_A} \frac{\theta_1}{\theta_2} e^{-X_i/\theta_2} \right) \cdot \left(\prod_{j \in S_B} \frac{1-\theta_1}{\theta_2} e^{X_j/\theta_2} \right)$$

The Log-Likelihood function $\ell(\theta_1, \theta_2) = \ln L(\theta_1, \theta_2)$ is:

$$\ell(\theta_1, \theta_2) = \sum_{i \in S_A} \ln \left(\frac{\theta_1}{\theta_2} e^{-X_i/\theta_2} \right) + \sum_{j \in S_B} \ln \left(\frac{1-\theta_1}{\theta_2} e^{X_j/\theta_2} \right)$$

Expanding the terms:

$$\begin{aligned} \ell(\theta_1, \theta_2) &= \sum_{i \in S_A} \left[\ln(\theta_1) - \ln(\theta_2) - \frac{X_i}{\theta_2} \right] + \sum_{j \in S_B} \left[\ln(1-\theta_1) - \ln(\theta_2) + \frac{X_j}{\theta_2} \right] \\ &= n_A \ln(\theta_1) - n_A \ln(\theta_2) - \frac{1}{\theta_2} \sum_{i \in S_A} X_i \\ &\quad + n_B \ln(1-\theta_1) - n_B \ln(\theta_2) + \frac{1}{\theta_2} \sum_{j \in S_B} X_j \\ &= n_A \ln(\theta_1) + n_B \ln(1-\theta_1) - (n_A + n_B) \ln(\theta_2) - \frac{1}{\theta_2} \left(\sum_{i \in S_A} X_i - \sum_{j \in S_B} X_j \right) \end{aligned}$$

Let $S = \sum_{i=1}^n |X_i| = \sum_{i \in S_A} X_i - \sum_{j \in S_B} X_j$. The log-likelihood simplifies to:

$$\ell(\theta_1, \theta_2) = n_A \ln(\theta_1) + n_B \ln(1-\theta_1) - n \ln(\theta_2) - \frac{S}{\theta_2}$$

We differentiate $\ell(\theta_1, \theta_2)$ with respect to θ_1 and set the derivative to zero:

$$\frac{\partial \ell}{\partial \theta_1} = \frac{n_A}{\theta_1} + n_B \left(\frac{-1}{1-\theta_1} \right) = 0$$

$$\frac{n_A}{\hat{\theta}_1} = \frac{n_B}{1-\hat{\theta}_1}$$

Solving for $\hat{\theta}_1$:

$$n_A(1-\hat{\theta}_1) = n_B\hat{\theta}_1 \implies n_A = (n_A+n_B)\hat{\theta}_1 = n\hat{\theta}_1$$

$$\hat{\theta}_1 = \frac{n_A}{n}$$

The MLE $\hat{\theta}_1$ is the sample proportion of non-negative observations.

We differentiate $\ell(\theta_1, \theta_2)$ with respect to θ_2 and set the derivative to zero:

$$\frac{\partial \ell}{\partial \theta_2} = -\frac{n}{\theta_2} + \frac{S}{\theta_2^2} = 0$$

Multiplying by θ_2^2:

$$-n\hat{\theta}_2 + S = 0$$

Solving for $\hat{\theta}_2$:

$$\hat{\theta}_2 = \frac{S}{n} = \frac{1}{n}\sum_{i=1}^{n} |X_i|$$

The MLE $\hat{\theta}_2$ is the sample mean of the absolute values of the observations.

The Maximum Likelihood Estimators for θ_1 and θ_2 are:

$$\hat{\theta}_1 = \frac{n_A}{n} \quad \text{and} \quad \hat{\theta}_2 = \overline{|X|} = \frac{1}{n}\sum_{i=1}^{n} |X_i|$$

2) The Invariance Property of the MLE states that for any measurable function $g(\boldsymbol{\theta})$, the MLE of $g(\boldsymbol{\theta})$ is $g(\hat{\boldsymbol{\theta}})$. While the function $g(\theta_1, \theta_2) = \theta_1 - \theta_2$ is not injective (since many pairs (θ_1, θ_2) can yield the same result), we can demonstrate its MLE using an injective reparameterization of the parameter space.

We define a new parameter vector $\boldsymbol{\eta}$ that includes the function of interest, $g(\boldsymbol{\theta}) = \theta_1 - \theta_2$, as one of its components, making the transformation $\boldsymbol{\theta} \to \boldsymbol{\eta}$ invertible (bijective/injective):

Let the transformation $h : \mathbb{R}^2 \to \mathbb{R}^2$ be:

$$\boldsymbol{\eta} = h(\boldsymbol{\theta}) = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \theta_1 - \theta_2 \\ \theta_2 \end{pmatrix}$$

- $\eta_1 = \theta_1 - \theta_2$ (The target quantity)
- $\eta_2 = \theta_2$ (An auxiliary component)

The transformation is easily inverted (and thus injective/bijective):

$$\theta_2 = \eta_2 \quad \text{and} \quad \theta_1 = \eta_1 + \eta_2$$

Since the transformation h is injective, the MLE for the new parameter vector $\boldsymbol{\eta}$ is obtained by applying the function h to the MLE of the original vector $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)$:

$$\hat{\boldsymbol{\eta}} = h(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \hat{\theta}_1 - \hat{\theta}_2 \\ \hat{\theta}_2 \end{pmatrix}$$

The MLE for the first component, $\eta_1 = \theta_1 - \theta_2$, is the first component of $\hat{\boldsymbol{\eta}}$:

$$\widehat{\theta_1 - \theta_2} = \widehat{\eta_1} = \hat{\theta}_1 - \hat{\theta}_2$$

Substituting the known MLEs: $\hat{\theta}_1 = n_A/n$ and $\hat{\theta}_2 = \overline{|X|}$:

$$\widehat{\theta_1 - \theta_2} = \frac{n_A}{n} - \overline{|X|}$$

By constructing an injective reparameterization that includes $\theta_1 - \theta_2$ as a component, we formally demonstrate that the MLE of the difference is simply the difference of the individual MLEs, confirming the general **Invariance Property**.

3) An estimator $\hat{\theta}_n$ is **consistent** for θ if it converges in probability to the true parameter value as $n \to \infty$ ($\hat{\theta}_n \xrightarrow{P} \theta$). We will use the **Law of Large Numbers (LLN)**.

The parameter θ_1 is the true probability of an observation being non-negative, $P(X \geq 0)$:

$$P(X \geq 0) = \int_0^\infty f(x)dx = \int_0^\infty \frac{\theta_1}{\theta_2} e^{-x/\theta_2} dx = \theta_1 \cdot \left[-e^{-x/\theta_2}\right]_0^\infty = \theta_1(0 - (-1)) = \theta_1$$

Let Y_i be an indicator variable for X_i:

$$Y_i = \mathbf{I}_{\{X_i \geq 0\}} \quad \text{such that} \quad \hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y}$$

Since $Y_1, \ldots, Y_n$ are i.i.d. random variables with finite mean $E[Y_i] = P(Y_i = 1) = P(X \geq 0) = \theta_1$, by the **Law of Large Numbers**:

$$\hat{\theta}_1 = \overline{Y} \xrightarrow{P} E[Y_i] = \theta_1$$

Therefore, $\hat{\theta}_1$ is a **consistent estimator** for θ_1.

The estimator $\hat{\theta}_2$ is the sample mean of $|X_i|$, so it must converge to the population mean $E[|X|]$. We need to show $E[|X|] = \theta_2$.

$$E[|X|] = \int_{-\infty}^\infty |x| f(x \mid \theta_1, \theta_2)dx = \int_{-\infty}^0 (-x)\frac{1-\theta_1}{\theta_2} e^{x/\theta_2} dx + \int_0^\infty x\frac{\theta_1}{\theta_2} e^{-x/\theta_2} dx$$

Using integration by parts (or recognizing that $\int_0^\infty u\frac{1}{\theta_2}e^{-u/\theta_2}du = \theta_2$, the mean of an Exponential distribution with scale θ_2):

- **First integral ($x < 0$):** Using $u = -x$: $\int_0^\infty u\frac{1-\theta_1}{\theta_2}e^{-u/\theta_2}du = (1-\theta_1)\left(\int_0^\infty u\frac{1}{\theta_2}e^{-u/\theta_2}du\right) = (1-\theta_1)\theta_2$
- **Second integral ($x \geq 0$):** $\int_0^\infty x\frac{\theta_1}{\theta_2}e^{-x/\theta_2}dx = \theta_1\left(\int_0^\infty x\frac{1}{\theta_2}e^{-x/\theta_2}dx\right) = \theta_1\theta_2$

Summing the terms:

$$E[|X|] = (1-\theta_1)\theta_2 + \theta_1\theta_2 = \theta_2 - \theta_1\theta_2 + \theta_1\theta_2 = \theta_2$$

Since $|X_1|, \ldots, |X_n|$ are i.i.d. random variables with finite mean $E[|X|] = \theta_2$, by the **Law of Large Numbers**:

$$\hat{\theta}_2 = \overline{|X|} \xrightarrow{P} E[|X|] = \theta_2$$

Therefore, $\hat{\theta}_2$ is a **consistent estimator** for θ_2.

The function $g(x, y) = x - y$ is a continuous function of its arguments. The **Continuous Mapping Theorem (CMT)** states that if a sequence of random variables converges in probability, any continuous function of that sequence also converges in probability.

Since $\hat{\theta}_1 \xrightarrow{P} \theta_1$ and $\hat{\theta}_2 \xrightarrow{P} \theta_2$:

$$\widehat{\theta_1 - \theta_2} = g(\hat{\theta}_1, \hat{\theta}_2) \xrightarrow{P} g(\theta_1, \theta_2) = \theta_1 - \theta_2$$

Therefore, the estimator $\widehat{\theta_1 - \theta_2}$ is a **consistent estimator** for $\theta_1 - \theta_2$.

Mathematical Statistics

Tutorial 4

1. Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a population X.

   Assume that the population variance is $V(X) = \sigma^2 < \infty$ and that the fourth central moment, μ_4, exists and is finite:
   $$\mu_4 = E\left[(X - E[X])^4\right] < \infty$$

   Then, the sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, is **asymptotically normally distributed** with a mean of $V(X)$ and an asymptotic variance given by $\frac{1}{n}(\mu_4 - \sigma^4)$.

   Formally, the standardized sequence converges in distribution to a Normal distribution:
   $$\sqrt{n}(S^2 - V(X)) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - (V(X))^2\right)$$
   where the asymptotic variance is often written as $\mu_4 - \sigma^4$.

2. Let $X_1, \ldots, X_n$ be a random sample from an Exponential distribution with mean parameter θ. The Probability Density Function (PDF) is given by $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$. We know that the variance of the distribution is $q(\theta) = \mathrm{Var}(X) = \theta^2$.

   (a) Calculate the Fisher Information per observation, $I(\theta)$.

   (b) Find the MLE of θ ($\hat{\theta}_n$) and its asymptotic distribution.

   (c) We are interested in estimating the variance $q(\theta) = \mathrm{Var}(X)$. Propose the MLE estimator of $q(\theta)$ and find its asymptotic distribution.

   (d) Another way to estimate $q(\theta)$ is with the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. Calculate the asymptotic distribution of S^2. If the sample size is large, which estimator for the variance would be preferred, and why?

3. Let $X_1, \ldots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$. Derive the $(1-\alpha)100\%$ confidence interval for the population variance σ^2 in two cases: (a) μ is known, and (b) μ is unknown.

Problem 1

Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a population X.

Assume that the population variance is $V(X) = \sigma^2 < \infty$ and that the fourth central moment, μ_4, exists and is finite:

$$\mu_4 = E\left[(X - E[X])^4\right] < \infty$$

Then, the sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, is asymptotically normally distributed with a mean of $V(X)$ and an asymptotic variance given by $\frac{1}{n}(\mu_4 - \sigma^4)$.

Formally, the standardized sequence converges in distribution to a Normal distribution:

$$\sqrt{n}(S^2 - V(X)) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - (V(X))^2\right)$$

where the asymptotic variance is often written as $\mu_4 - \sigma^4$.

Solutions:

We use the algebraic identity for the sum of squares:

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \sum_{i=1}^{n}(X_i - E[X])^2 - n(\bar{X}_n - E[X])^2$$

To find the asymptotic distribution of $\sqrt{n}(S^2 - \sigma^2)$, note that,

$$\begin{aligned}
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \sigma^2\right) &= \sqrt{n}\left(\frac{1}{n}\left[\sum_{i=1}^{n}(X_i - E[X])^2 - n(\bar{X}_n - E[X])^2\right] - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X])^2 - \sigma^2\right) - \sqrt{n}(\bar{X}_n - E[X])^2 \\
&= A + B
\end{aligned}$$

Let $Y_i = (X_i - E[X])^2$. Since X_i are i.i.d., Y_i are also i.i.d. We find the mean and variance of Y_i:

- **Mean of Y_i:** $E[Y_i] = E[(X_i - E[X])^2] = V(X) = \sigma^2$.
- **Variance of Y_i:** $\text{Var}(Y_i) = E[Y_i^2] - (E[Y_i])^2 = E[(X_i - E[X])^4] - (V(X))^2 = \mu_4 - \sigma^4$.

By the **Central Limit Theorem (CLT)** applied to the sample mean of Y_i, $\bar{Y}_n = \frac{1}{n}\sum Y_i$:

$$\sqrt{n}(\bar{Y}_n - E[Y_i]) = \sqrt{n}(\bar{Y}_n - \sigma^2) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X])^2 - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, \mu_4 - \sigma^4) \quad \text{(Term A)}$$

We analyze the convergence of the Residual Term (Term B):

$$\text{Term B} = \sqrt{n}(\bar{X}_n - E[X])^2$$

From the CLT applied to $\bar{X}_n$:

$$\sqrt{n}(\bar{X}_n - E[X]) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2)$$

We rewrite Term B by factoring $\frac{1}{\sqrt{n}}$:

$$\sqrt{n}(\bar{X}_n - E[X])^2 = \frac{1}{\sqrt{n}}\left[\sqrt{n}(\bar{X}_n - E[X])\right]^2$$

The first part, $\frac{1}{\sqrt{n}}$, converges to 0 in probability. The second part, $\left[\sqrt{n}(\bar{X}_n - E[X])\right]^2$, converges in distribution to Z^2 (which is finite almost surely).

By the continuous mapping theorem and properties of convergence, the product of a sequence converging to zero in probability and a sequence converging in distribution is a sequence that converges to zero in probability:

$$\sqrt{n}(\bar{X}_n - E[X])^2 \xrightarrow{p} 0$$

Therefore,by **Slutsky's Theorem**, the term that converges to zero is negligible in the limit:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \sigma^2\right) = \underbrace{\sqrt{n}(\bar{Y}_n - \sigma^2)}_{\xrightarrow{d}\mathcal{N}(0,\mu_4-\sigma^4)} - \underbrace{\sqrt{n}(\bar{X}_n - E[X])^2}_{\xrightarrow{p}0} \overset{d}{=\!\!\rightarrow} \mathcal{N}\left(0, \mu_4 - \sigma^4\right)$$

Finally, since $S^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and $\frac{n}{n-1} \to 1$, S^2 and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ are asymptotically equivalent, sharing the same limiting distribution.

$$\sqrt{n}(S^2 - \sigma^2) \xrightarrow{d} \mathcal{N}\left(0, \mu_4 - \sigma^4\right)$$

Problem 2

Probability Density Function (PDF) is given by $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$. We know that the variance of the distribution is $q(\theta) = \mathrm{Var}(X) = \theta^2$.

(a) Calculate the Fisher Information per observation, $I(\theta)$.

(b) Find the MLE of θ ($\hat{\theta}_n$) and its asymptotic distribution.

(c) We are interested in estimating the variance $q(\theta) = \mathrm{Var}(X)$. Propose the MLE estimator of $q(\theta)$ and find its asymptotic distribution.

(d) Another way to estimate $q(\theta)$ is with the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. Calculate the asymptotic distribution of $\sqrt{n}(S^2 - \sigma^2)$.

(e) If the sample size is large, which estimator for the variance would be preferred, and why?

Solution

The Fisher Information per observation is $I(\theta) = -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]$.

1. Log-Likelihood for one observation:

$$\log f(X;\theta) = -\log(\theta) - \frac{X}{\theta}$$

2. Second Derivative:

$$\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right) = \frac{1}{\theta^2} - \frac{2X}{\theta^3}$$

3. Fisher Information: Using $E[X] = \theta$:

$$I(\theta) = -E\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right] = -\left(\frac{1}{\theta^2} - \frac{2E[X]}{\theta^3}\right) = -\left(\frac{1}{\theta^2} - \frac{2\theta}{\theta^3}\right) = \frac{1}{\theta^2}$$

$$I(\theta) = \boldsymbol{\frac{1}{\theta^2}}$$

MLE of θ ($\hat{\theta}_n$): The MLE for the Exponential mean parameter is the sample mean:

$$\hat{\theta}_n = \bar{\mathbf{X}}_\mathbf{n}$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

Substituting $I(\theta) = 1/\theta^2$:

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\theta^2}\right)$$

We are estimating $q(\theta) = \theta^2$, by the invariance property of MLE, $\hat{q}_n = q(\hat{\theta}_n) = (\bar{X}_n)^2$. We use the Delta Method with $q(\theta) = \theta^2$ and $q'(\theta) = 2\theta$. The Asymptotic Variance (AV) is

$$\mathrm{AV}(\bar{X}_n^2) = \frac{(2\theta)^2}{1/\theta^2} = 4\theta^4$$

The asymptotic distribution is:

$$\sqrt{n}(\bar{X}_n^2 - \theta^2) \xrightarrow{d} \mathcal{N}\left(0, 4\theta^4\right)$$

Asymptotic Distribution of $\sqrt{n}(S^2 - \sigma^2)$

1. **Calculation of Asymptotic Variance (AV):**

   - Variance: $\sigma^2 = \theta^2 \implies \sigma^4 = \theta^4$.
   - Fourth Central Moment for Exp(θ): $\mu_4 = 9\theta^4$.

$$\text{AV}(S^2) = \mu_4 - \sigma^4 = 9\theta^4 - \theta^4 = 8\theta^4$$

$$\sqrt{n}(S^2 - \sigma^2) \xrightarrow{d} \mathcal{N}\left(0, 8\theta^4\right)$$

We compare the Asymptotic Variances:

Estimator	**Asymptotic Variance (AV)**
$\bar{X}_n^2$	$4\theta^4$
S^2	$8\theta^4$

Since $\text{AV}(\bar{X}_n^2) < \text{AV}(S^2)$, the estimator $\bar{X}_n^2$ is **asymptotically more efficient**. For a large sample, the estimator $\mathbf{\bar{X}_n^2}$ would be preferred, as it achieves a smaller asymptotic variance and is thus more precise.

Problem 3

Let $X_1, \ldots, X_n$ be a random sample from a Normal distribution $\mathcal{N}(\mu, \sigma^2)$. Derive the $(1-\alpha)100\%$ confidence interval for the population variance σ^2 in two cases: (a) μ is known, and (b) μ is unknown.

Solution

(a) Case 1: Mean μ is Known

- Pivot Quantity: $Y_1 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2_n$.
- Critical Values: We use the χ^2 quantiles $\chi^2_{\alpha/2,n}$ (lower) and $\chi^2_{1-\alpha/2,n}$ (upper, standard notation):

$$P\left(\chi^2_{\alpha/2,n} < Y_1 < \chi^2_{1-\alpha/2,n}\right) = 1 - \alpha$$

- Invert to find σ^2:

$$\mathrm{CI}(\sigma^2) = \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{1-\alpha/2,n}}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi^2_{\alpha/2,n}}\right]$$

(b) Case 2: Mean μ is Unknown

- Pivot Quantity: $Y_2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
- Critical Values: We use the χ^2 quantiles with $df = n - 1$:

$$P\left(\chi^2_{\alpha/2,n-1} < Y_2 < \chi^2_{1-\alpha/2,n-1}\right) = 1 - \alpha$$

- Invert to find σ^2:

$$\mathrm{CI}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right]$$

Mathematical Statistics

Tutorial 5

1. Let $X_1, \ldots, X_{n_1}$ be an independent random sample from $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \ldots, Y_{n_2}$ be a second independent random sample from $\mathcal{N}(\mu_2, \sigma^2)$. The samples are mutually independent and share the same unknown variance σ^2. Find the $(1-\alpha)100\%$ confidence interval for the difference of the means, $\mu_1 - \mu_2$.

2. Let $X_1, \ldots, X_n$ be a random sample whose PDF is $f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$. Find the $(1-\alpha)100\%$ confidence interval based on the MLE and MoM.

Problem 1

Let $X_1, \dots, X_{n_1}$ be an independent random sample from $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \dots, Y_{n_2}$ be a second independent random sample from $\mathcal{N}(\mu_2, \sigma^2)$. The samples are mutually independent and share the same unknown variance σ^2. Find the $(1-\alpha)100\%$ confidence interval for the difference of the means, $\mu_1 - \mu_2$.

Solution

The difference of the sample means is $\bar{X} - \bar{Y}$. Since X_i and Y_i are from Normal distributions, the difference of their means is also Normal:

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

Standardizing this difference gives a **Standard Normal** variable Z:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0,1)$$

If σ^2 were known, we would use Z. However, since σ^2 is unknown, we must estimate it.

The t-distribution is defined as the ratio of a Standard Normal variable (Z) and the square root of a Chi-Squared variable (W) divided by its degrees of freedom (ν):

$$T = \frac{Z}{\sqrt{W/\nu}}$$

We need an estimator for σ^2 that is independent of $\bar{X} - \bar{Y}$ and follows a χ^2 distribution when properly scaled.

The sum of the scaled sample variances follows a χ^2 distribution:

$$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1} \quad \text{and} \quad \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1}$$

Since the samples are independent, their sum is also χ^2:

$$W = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

The degrees of freedom are $\nu = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

The pooled variance S_p^2 is defined such that the numerator W is $\frac{(n_1+n_2-2)S_p^2}{\sigma^2}$:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

We construct the ratio $T = \frac{Z}{\sqrt{W/\nu}}$, substituting S_p^2 for σ^2 in the standard error:

$$T = \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}{\sqrt{\frac{\frac{(n_1+n_2-2)S_p^2}{\sigma^2}}{n_1+n_2-2}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Because the numerator is $\mathcal{N}(0,1)$ and the denominator is the square root of a χ^2 distribution (divided by its degrees of freedom, ν) and is independent of the numerator, the quantity T follows the **Student's t-distribution** with $\mathbf{df = n_1 + n_2 - 2}$.

The $(1-\alpha)100\%$ confidence interval for $\mu_1 - \mu_2$ is found by solving the inequality $P(-t_{1-\alpha/2,df} < T < t_{1-\alpha/2,df}) = 1 - \alpha$:

$$\text{CI}(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm t_{1-\alpha/2, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $t_{1-\alpha/2,n_1+n_2-2}$ is the critical value leaving $\alpha/2$ in the upper tail of the t-distribution.

Problem 2

Let $X_1, \ldots, X_n$ be a random sample whose PDF is $f_{X|\theta}(x) = \frac{2x}{\theta^2} \cdot \mathbf{I}_{(0,\theta)}(x)$. Find the $(1-\alpha)100\%$ confidence interval based on the MLE and MoM.

Solution

The **MLE** is the maximum order statistic:

$$\hat{\theta}_n = X_{(n)}$$

The population mean is $E[X] = \int_0^\theta x \frac{2x}{\theta^2} dx = \frac{2\theta}{3}$. Setting the sample mean $\bar{X}$ equal to the population mean gives the MME $\tilde{\theta}_n$:

$$\bar{X} = \frac{2\tilde{\theta}_n}{3} \implies \tilde{\theta}_n = \frac{3}{2}\bar{X}$$

Exact Confidence Interval

Since the MLE is a function of the sufficient statistic $X_{(n)}$, we use the exact distribution of a pivot quantity to construct the confidence interval (CI). The CDF of the maximum order statistic is $F_{X_{(n)}}(t) = \left(\frac{t^2}{\theta^2}\right)^n = \frac{t^{2n}}{\theta^{2n}}$ for $0 < t < \theta$. The pivot quantity $Y = \frac{X_{(n)}^2}{\theta^2}$ follows the distribution $F_Y(y) = y^n$ for $0 < y < 1$, which is the Beta$(n, 1)$ distribution. For a $(1-\alpha)100\%$ CI, we find the quantiles $y_{\alpha/2}$ and $y_{1-\alpha/2}$ of Y:

$$y_{\alpha/2} = \left(\frac{\alpha}{2}\right)^{1/n} \quad \text{and} \quad y_{1-\alpha/2} = \left(1 - \frac{\alpha}{2}\right)^{1/n}$$

Inverting the statement $P(y_{\alpha/2} < X_{(n)}^2/\theta^2 < y_{1-\alpha/2}) = 1 - \alpha$ yields:

$$\text{CI}_{\text{EXACT}}(\theta) = \left[\frac{X_{(n)}}{\sqrt{y_{1-\alpha/2}}}, \frac{X_{(n)}}{\sqrt{y_{\alpha/2}}}\right] = \left[\frac{X_{(n)}}{\left(1-\frac{\alpha}{2}\right)^{1/(2n)}}, \frac{X_{(n)}}{\left(\frac{\alpha}{2}\right)^{1/(2n)}}\right]$$

Asymptotic Confidence Interval

The MME $\tilde{\theta}_n$ is based on the sample mean $\bar{X}$, which satisfies the conditions for the **Central Limit Theorem (CLT)**. The variance of X is $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{\theta^2}{2} - \left(\frac{2\theta}{3}\right)^2 = \frac{\theta^2}{18}$. Using the **Delta Method** on $\tilde{\theta}_n = g(\bar{X}) = \frac{3}{2}\bar{X}$:

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, [g'(E[X])]^2 \text{Var}(X)\right)$$

Since $g'(\cdot) = 3/2$, the asymptotic variance is:

$$V_{\text{ASYM}}(\tilde{\theta}_n) = \left(\frac{3}{2}\right)^2 \frac{\theta^2}{18} = \frac{\theta^2}{8}$$

We use the approximate normality $\tilde{\theta}_n \sim \mathcal{N}(\theta, V_{\text{ASYM}}(\tilde{\theta}_n)/n)$ and replace θ with the estimator $\tilde{\theta}_n$ in the variance:

$$\text{CI}_{\text{MME}}(\theta) \approx \tilde{\theta}_n \pm Z_{1-\alpha/2}\sqrt{\frac{\tilde{\theta}_n^2/8}{n}} = \tilde{\theta}_n \left[1 \pm \frac{Z_{1-\alpha/2}}{\sqrt{8n}}\right]$$

Asymptotic Confidence Interval for MLE

Since the support of the PDF depends on θ, the MLE $\hat{\theta}_n = X_{(n)}$ does not follow the standard asymptotic normality theorem. Instead, the asymptotic distribution is based on the limit of the maximum order statistic. The specialized result for this distribution, where the density is non-zero at θ, states that the quantity $n(\theta - X_{(n)})$ converges to a scaled Exponential distribution:

$$n(\theta - X_{(n)}) \xrightarrow{d} W, \quad \text{where } W \sim \text{Exponential}(2/\theta)$$

The CDF of W is $F_W(w) = 1 - e^{-2w/\theta}$ for $w > 0$. For large n, we use $n(\theta - X_{(n)}) \approx W$. We find the quantiles $w_{\alpha/2}$ and $w_{1-\alpha/2}$ such that $P(W > w_\gamma) = 1 - \gamma$.

$$w_\gamma = -\frac{\theta}{2}\ln(1-\gamma)$$

Substituting this into the confidence interval derivation (and using $\hat{\theta}_n$ for θ in the quantile terms for the final interval):

$$\text{CI}_{\text{MLE-ASYM}}(\theta) \approx \left[\hat{\theta}_n\left(\frac{1}{1+\frac{\ln(1-\alpha/2)}{2n}}\right), \hat{\theta}_n\left(\frac{1}{1+\frac{\ln(\alpha/2)}{2n}}\right)\right]$$

This interval is often simpler to calculate asymptotically than the exact one, but is based on the non-regular limit distribution of the MLE.

Mathematical Statistics

Tutorial 6

1. Let $X_1, \dots, X_n$ be an independent and identically distributed (i.i.d.) random sample from the density:
$$f(x, \theta) = \theta x^{\theta-1} I_{(0,1)}(x) \quad \text{where } \theta \in \Theta = (0, +\infty)$$
We are interested in estimating the parameter $\lambda = 1/\theta$.
The maximum Likelihood Estimator (MLE) of λ is $\hat{\lambda}_{MV} = -\frac{\sum_{i=1}^n \ln x_i}{n} = -\overline{\ln X}$. and the moment estimador based on the first population moment $\hat{\lambda}_{MO} = \frac{1-\bar{X}}{\bar{X}} = \frac{1}{\bar{X}} - 1$.
   (a) Compute the Cramér-Rao Lower Bound (CRLB) and compare it with the variance of the MLE estimators.
   (b) Compare the asymptotic variances of $\hat{\lambda}_{MV}$ and $\hat{\lambda}_{MO}$ and determine which is asymptotically more efficient.

2. Let $X_1, X_2, \dots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a Normal distribution $N(\mu, \rho^2)$, where $\mu \in \mathbb{R}$ and $\rho^2 \in \mathbb{R}^+$. Use the Neyman Factorization Theorem to find a sufficient statistic, $T(\mathbf{X})$, for the parameter(s) specified in each of the following cases:
   (a) Case 1: Both Parameters Unknown. Find a sufficient statistic for the parameter vector $\boldsymbol{\theta} = (\mu, \rho^2)$, where both the mean μ and the variance ρ^2 are unknown.
   (b) Case 2: Mean Unknown, Variance Known. Find a sufficient statistic for the parameter $\theta = \mu$, where the variance ρ^2 is known (a fixed constant).
   (c) Case 3: Mean Known, Variance Unknown. Find a sufficient statistic for the parameter $\theta = \rho^2$, where the mean μ is known (a fixed constant).

3. Let $X_1, X_2, \dots, X_n$ be an i.i.d. random sample from a Shifted Exponential distribution with the PDF:
$$f(x; \theta) = e^{-(x-\theta)} \mathbb{I}_{(\theta,+\infty)}(x), \quad \theta \in \mathbb{R}$$
Use the Neyman Factorization Theorem to find a sufficient statistic for the parameter θ.

Problem 1

Let $X_1, \ldots, X_n$ be an independent and identically distributed (i.i.d.) random sample from the density:

$$f(x, \theta) = \theta x^{\theta-1} I_{(0,1)}(x) \quad \text{where } \theta \in \Theta = (0, +\infty)$$

We are interested in estimating the parameter $\lambda = 1/\theta$.

The maximum Likelihood Estimator (MLE) of λ is $\hat{\lambda}_{MV} = -\frac{\sum_{i=1}^{n} \ln x_i}{n} = -\overline{\ln X}$. and the moment estimador based on the first population moment $\hat{\lambda}_{MO} = \frac{1-\bar{X}}{\bar{X}} = \frac{1}{\bar{X}} - 1$.

1. Compute the Cramér-Rao Lower Bound (CRLB) and compare it with the variance of the MLE estimators.
2. Compare the asymptotic variances of $\hat{\lambda}_{MV}$ and $\hat{\lambda}_{MO}$ and determine which is asymptotically more efficient.

Solution

We compute the Fisher Information for the parameter λ.

We compute the Fisher Information for θ ($I_1(\theta)$): The second derivative of the log-likelihood for one observation X is $\frac{d^2 \ln f(X|\theta)}{d\theta^2} = -\frac{1}{\theta^2}$.

$$I_1(\theta) = -\mathbb{E}\left[-\frac{1}{\theta^2}\right] = \frac{1}{\theta^2}$$

then $\lambda = q(\theta) = 1/\theta$. The Cramer-Rao Bound is

$$\frac{(q'(\theta))^2}{nI_1(\theta)} = \frac{(-\frac{1}{\theta^2})^2}{n\frac{1}{\theta^2}} = \frac{1}{n\theta^2}.$$

Also we can rewrite the bound in terms of λ. Using the transformation $\theta = 1/\lambda$,

$$\frac{(q'(\theta))^2}{nI_1(\theta)} = \frac{1}{n\theta^2} = \frac{\lambda^2}{n}.$$

Another posibility could be rewrite the density as a function of λ and compute the information number, $I_1(\lambda)$. We can see that the

$$I_1(\lambda) = I_1(\theta)\left(\frac{d\theta}{d\lambda}\right)^2 = \frac{1}{(1/\lambda)^2}\left(-\frac{1}{\lambda^2}\right)^2 = \lambda^2 \cdot \frac{1}{\lambda^4} = \frac{1}{\lambda^2}$$

then

$$\text{CRLB}(\lambda) = \frac{1}{nI_1(\lambda)} = \frac{1}{n/\lambda^2} = \frac{\lambda^2}{n}$$

Comparison with Estimator Variances of MLE ($\hat{\lambda}_{MV}$)

It is easy to see that $Y_i = -\ln X_i \sim \text{Exponential}(\theta)$, so $\hat{\lambda}_{MV} = \bar{Y}$. The variance of $\hat{\lambda}_{MV}$ is $\text{Var}(\hat{\lambda}_{MV}) = \frac{\text{Var}(Y_i)}{n}$. Since $\text{Var}(Y_i) = 1/\theta^2 = \lambda^2$:

$$\text{Var}(\hat{\lambda}_{MV}) = \frac{\lambda^2}{n}$$

The MLE is efficient for finite samples because $\text{Var}(\hat{\lambda}_{MV}) = \text{CRLB}(\lambda)$.

Asymptotic Distribution of $\hat{\lambda}_{MV}$ (MLE)

The asymptotic distribution is:

$$\sqrt{n}(\hat{\lambda}_{MV} - q(\theta)) \xrightarrow{D} N(0, (q'(\theta))^2/(I_1(\theta)))$$

then the asymprtotic variance is

$$\text{AVar}(\hat{\lambda}_{MV}) = \frac{1}{\theta^2} = \lambda^2$$

Asymptotic Distribution of $\hat{\lambda}_{MO}$ (MME)

We can compute the mean and the variance of X

- Mean: $\mu_1 = \mathbb{E}[X] = \frac{\theta}{\theta+1}$.
- Variance: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\text{Var}(X) = \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 = \frac{\theta(\theta+1)^2 - \theta^2(\theta+2)}{(\theta+2)(\theta+1)^2} = \frac{\theta}{(\theta+2)(\theta+1)^2}$$

Since the MME is $\hat{\lambda}_{MO} = \frac{1}{\bar{X}} - 1$, we define the function $g(m)$

$$g(m) = \frac{1}{m} - 1$$

Then $\hat{\lambda}_{MO} = \frac{1}{\bar{X}} - 1 = g(\bar{X})$ and $g(\frac{\theta}{\theta+1}) = \frac{1}{\theta}$ The using the Central Limit Theorem and applying the Delta Method, we have that

$$\sqrt{n}(\hat{\lambda}_{MO} - q(\theta)) \xrightarrow{D} N(0, \text{AVar}(\hat{\lambda}_{MO}))$$

where the asymptotic variance of $\hat{\lambda}_{MO}$ is given by:

$$\text{AVar}(\hat{\lambda}_{MO}) = \text{Var}(X) \cdot [g'(\mathbb{E}[X])]^2$$

First, we compute the derivative of $g(m)$ with respect to m:

$$g'(m) = \frac{d}{dm}\left(\frac{1}{m} - 1\right) = -\frac{1}{m^2}$$

Now we evaluate this derivative at the population mean $\mathbb{E}[X] = \frac{\theta}{\theta+1}$:

$$g'(\mathbb{E}[X]) = -\frac{1}{\left(\frac{\theta}{\theta+1}\right)^2} = -\frac{(\theta+1)^2}{\theta^2}$$

Substituting $\text{Var}(X)$ and $g'(\mathbb{E}[X])$ into the Delta Method formula:

$$\begin{aligned}
\text{AVar}(\hat{\lambda}_{MO}) &= \left(\frac{\theta}{(\theta+2)(\theta+1)^2}\right) \cdot \left(-\frac{(\theta+1)^2}{\theta^2}\right)^2 \\
&= \left(\frac{\theta}{(\theta+2)(\theta+1)^2}\right) \cdot \left(\frac{(\theta+1)^4}{\theta^4}\right) \\
&= \frac{(\theta+1)^2}{\theta^3(\theta+2)}
\end{aligned}$$

The Asymptotic Variance of the MME in terms of θ is:

$$\text{AVar}(\hat{\lambda}_{MO}) = \frac{(\theta+1)^2}{\theta^3(\theta+2)}$$

The asymptotic distribution is:

$$\sqrt{n}(\hat{\lambda}_{MO} - q(\theta)) \xrightarrow{D} N\left(0, \frac{(\theta+1)^2}{\theta^3(\theta+2)}\right)$$

Comparison of Asymptotic Variances

We compare $\text{AVar}(\hat{\lambda}_{MV}) = 1/\theta^2$ with $\text{AVar}(\hat{\lambda}_{MO}) = \frac{(\theta+1)^2}{\theta^3(\theta+2)}$.

$$\frac{\frac{(\theta+1)^2}{\theta^3(\theta+2)}}{1/\theta^2} = \frac{(\theta+1)^2}{\theta(\theta+2)}$$

Since $\theta > 0$, the ratio is greater than 1. Therefore, $\text{AVar}(\hat{\lambda}_{MV}) < \text{AVar}(\hat{\lambda}_{MO})$.

The Maximum Likelihood Estimator ($\hat{\lambda}_{MV}$) is **asymptotically more efficient** than the Method of Moments Estimator ($\hat{\lambda}_{MO}$), as it achieves the minimum possible asymptotic variance (λ^2).

Problem 2

Let $X_1, X_2, \dots, X_n$ be an independent and identically distributed (i.i.d.) random sample from a Normal distribution $N(\mu, \rho^2)$, where $\mu \in \mathbb{R}$ and $\rho^2 \in \mathbb{R}^+$. Use the Neyman Factorization Theorem to find a sufficient statistic, $T(\mathbf{X})$, for the parameter(s) specified in each of the following cases:

1. Case 1: Both Parameters Unknown. Find a sufficient statistic for the parameter vector $\boldsymbol{\theta} = (\mu, \rho^2)$, where both the mean μ and the variance ρ^2 are unknown.

2. Case 2: Mean Unknown, Variance Known. Find a sufficient statistic for the parameter $\theta = \mu$, where the variance ρ^2 is known (a fixed constant).

3. Case 3: Mean Known, Variance Unknown. Find a sufficient statistic for the parameter $\theta = \rho^2$, where the mean μ is known (a fixed constant).

Solution

The probability density function (PDF) for a single observation X_i is:

$$f(x_i; \mu, \rho^2) = \frac{1}{\sqrt{2\pi\rho^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\rho^2}\right)$$

The likelihood function for the i.i.d. sample $\mathbf{x} = (x_1, \dots, x_n)$ is:

$$L(\mathbf{x}; \mu, \rho^2) = \prod_{i=1}^{n} f(x_i; \mu, \rho^2) = \left(\frac{1}{2\pi\rho^2}\right)^{n/2} \exp\left(-\frac{1}{2\rho^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

We expand the sum of squares in the exponent: $\sum(x_i - \mu)^2 = \sum x_i^2 - 2\mu\sum x_i + n\mu^2$.

$$L(\mathbf{x}; \mu, \rho^2) = \left(\frac{1}{2\pi\rho^2}\right)^{n/2} \exp\left(-\frac{1}{2\rho^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right)$$

According to the **Neyman Factorization Theorem**, $T(\mathbf{X})$ is a sufficient statistic for $\boldsymbol{\theta}$ if $L(\mathbf{x}; \boldsymbol{\theta})$ can be factored as $g(T(\mathbf{x}); \boldsymbol{\theta})h(\mathbf{x})$, where $h(\mathbf{x})$ does not depend on $\boldsymbol{\theta}$.

1. Case 1: Both Parameters Unknown ($\boldsymbol{\theta} = (\mu, \rho^2)$)

Since $h(\mathbf{x})$ must not depend on μ or ρ^2, we can set $h(\mathbf{x}) = 1$. The entire likelihood function must then be the function g.

$$L(\mathbf{x}; \mu, \rho^2) = \underbrace{\left(\frac{1}{2\pi\rho^2}\right)^{n/2} \exp\left(-\frac{1}{2\rho^2}\left(n\mu^2 - 2\mu\sum x_i + \sum x_i^2\right)\right)}_{g(T_1(\mathbf{x}); \mu, \rho^2)} \cdot \underbrace{1}_{h(\mathbf{x})}$$

The function g depends on the sample $\mathbf{x}$ only through the values of $\sum x_i$ and $\sum x_i^2$.
Sufficient Statistic:

$$\mathbf{T}_1(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$$

2. Case 2: μ Unknown ($\theta = \mu$), ρ^2 Known

Since ρ^2 is known, terms involving only ρ^2 can be placed in $h(\mathbf{x})$. We rearrange the exponent to separate terms containing μ from terms that do not:

$$L(\mathbf{x}; \mu) = \underbrace{\exp\left(\frac{\mu}{\rho^2}\sum x_i - \frac{n\mu^2}{2\rho^2}\right)}_{g(T_2(\mathbf{x}); \mu)} \cdot \underbrace{\left(\frac{1}{2\pi\rho^2}\right)^{n/2} \exp\left(-\frac{1}{2\rho^2}\sum x_i^2\right)}_{h(\mathbf{x})}$$

The function g depends on $\mathbf{x}$ only through $\sum x_i$ and depends on the unknown parameter μ. The function h depends on $\mathbf{x}$ but is independent of μ.

Sufficient Statistic:

$$T_2(\mathbf{X}) = \sum_{i=1}^{n} X_i \quad \text{or equivalently } \bar{X}$$

3. Case 3: ρ^2 Unknown ($\theta = \rho^2$), μ Known

Since μ is known, we use the initial unexpanded form of the exponent $\sum(x_i - \mu)^2$.

$$L(\mathbf{x}; \rho^2) = \underbrace{\left(\frac{1}{2\pi\rho^2}\right)^{n/2} \exp\left(-\frac{1}{2\rho^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)}_{g(T_3(\mathbf{x});\rho^2)} \cdot \underbrace{1}_{h(\mathbf{x})}$$

The function g depends on $\mathbf{x}$ only through the quantity $\sum_{i=1}^{n}(x_i - \mu)^2$ and depends on the unknown parameter ρ^2.

Sufficient Statistic:

$$T_3(\mathbf{X}) = \sum_{i=1}^{n}(X_i - \mu)^2$$

Problem 3

Let $X_1, X_2, \dots, X_n$ be an i.i.d. random sample from a Shifted Exponential distribution with the PDF:

$$f(x;\theta) = e^{-(x-\theta)}\mathbb{I}_{(\theta,+\infty)}(x), \quad \theta \in \mathbb{R}$$

Use the Neyman Factorization Theorem to find a sufficient statistic for the parameter θ.

Solution

The joint PDF for the sample $\mathbf{X}$ is:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} e^{-(x_i-\theta)}\mathbb{I}_{(\theta,+\infty)}(x_i) \\ &= e^{-\sum_{i=1}^{n}(x_i-\theta)} \prod_{i=1}^{n} \mathbb{I}_{(\theta,+\infty)}(x_i) \\ &= e^{n\theta} e^{-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \mathbb{I}_{(\theta,+\infty)}(x_i) \end{aligned}$$

The product of indicator functions is non-zero (equal to 1) if and only if $x_i > \theta$ for all $i = 1, \dots, n$, which is equivalent to requiring that the smallest observation is greater than θ: $\min\{x_i\} > \theta$.

$$\prod_{i=1}^{n} \mathbb{I}_{(\theta,+\infty)}(x_i) = \mathbb{I}_{(\theta,+\infty)}(\min\{x_i\})$$

Substituting this back, we factor the joint PDF:

$$f_{\mathbf{X}}(\mathbf{x};\theta) = \underbrace{e^{n\theta}\mathbb{I}_{(\theta,+\infty)}(\min\{x_i\})}_{k_1(t(\mathbf{x}),\theta)} \underbrace{e^{-\sum_{i=1}^{n} x_i}}_{k_2(\mathbf{x})}$$

The function k_1 depends on $\mathbf{x}$ only through $t(\mathbf{x}) = \min\{x_i\}$ and the parameter θ. The function k_2 depends on $\mathbf{x}$ but not on θ.

By the Neyman Factorization Theorem, the statistic $T = \min\{X_i\}$ is sufficient for θ.

Mathematical Statistics

Tutorial 7

1. We have a random sample $X_1, \ldots, X_n \overset{iid}{\sim}$ which are independent and identically distributed discrete random variables with the following probability mass function for an unknown parameter $\theta > 0$ and Support $\text{Supp}(X) = \{0, 1, 2, 3 \cdots\} = \mathbb{N}_0$:

$$P_\theta\left(X_i = x\right) = \left(1 - e^{-\theta}\right) \cdot e^{-\theta x}$$

You may use without proof that: $E(X) = \dfrac{1}{e^\theta - 1}$ and $\text{Var}(X) = \dfrac{e^\theta}{(e^\theta - 1)^2}$

   (a) Show that the Maximum Likelihood Estimator (MLE) of θ is:

$$\hat{\theta}_{MLE} = \ln\left(\frac{\overline{X}_n + 1}{\overline{X}_n}\right)$$

   (b) Calculate the Fisher information for a single sample unit, $I_1(\theta)$.

   (c) Compute the CRB for $1 + \overline{X}_n$ as an estimator of $q(\theta) = \dfrac{1}{1 - e^{-\theta}}$.

   (d) Calculate the Maximum Likelihood Estimator (MLE) of $\beta(\theta) = e^{-\theta}$. Call it $\widehat{\beta}_{MLE}$.

   (e) Assume that the regularity conditions hold under which:

$$\sqrt{n}(\widehat{\beta}_{MLE} - \beta(\theta)) \xrightarrow{D} N(0, W(\theta)). \qquad W(\theta) = \frac{(e^\theta - 1)^2}{e^{3\theta}}$$

   (f) Based on the answer from the previous part, provide the formula for an asymptotic confidence interval with approximate coverage probability $1 - \alpha$ for $\beta(\theta)$.

2. Suppose $X_i, i = 1, ..., n$ are independent random variables, each with density $f\left(x; \theta^*\right)$ where for any $\theta > 0$,

$$f\left(x; \theta\right) = \begin{cases} \theta x \text{ if } 0 < x \leq \sqrt{\frac{2}{\theta}} \\ 0 \text{ otherwise} \end{cases}$$

   (a) Prove that the Method of Moments Estimator (MME) for θ is equal to $\widetilde{\theta} = \frac{8}{9\left(\overline{X}\right)^2}$ where $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ and $\sqrt{n}\left(\widetilde{\theta} - \theta\right) \xrightarrow{L} N\left(0, v\left(\theta\right)\right)$ (provide the formula for $v\left(\theta\right)$).

   (b) Consider the estimator of θ defined as $\widehat{\theta} = \frac{2}{\max\left\{X_1^2, \ldots, X_n^2\right\}}$. Prove that the cumulative distribution function $F\left(u\right) = P\left(\widehat{\theta} \leq u\right)$ is equal to:

$$F\left(u\right) = \begin{cases} 1 - \left(\frac{\theta}{u}\right)^n & \text{if } u > \theta \\ 0 & \text{otherwise} \end{cases}$$

   (c) Is $\widehat{\theta}$ the Maximum Likelihood Estimator (MLE) for θ? If yes, prove it; otherwise, explain why.

   (d) Using part (b), calculate a confidence interval with exact coverage probability equal to 0.95.

   (e) Compute the biase and the MSE of $\widehat{\theta}$ the estimator defined in part (b).

Problem 1

We have a random sample $X_1, \ldots, X_n \overset{iid}{\sim}$ which are independent and identically distributed discrete random variables with the following probability mass function for an unknown parameter $\theta > 0$ and Support $\text{Supp}(X) = \{0, 1, 2, 3 \cdots\} = \mathbb{N}_0$:

$$P_\theta(X_i = x) = \left(1 - e^{-\theta}\right) \cdot e^{-\theta x}$$

You may use without proof that:

$$E(X) = \frac{1}{e^\theta - 1} \quad \text{and} \quad \text{Var}(X) = \frac{e^\theta}{(e^\theta - 1)^2}$$

(a) Show that the Maximum Likelihood Estimator (MLE) of θ is:

$$\hat{\theta}_{MLE} = \ln\left(\frac{\overline{X}_n + 1}{\overline{X}_n}\right)$$

(b) Calculate the Fisher information for a single sample unit, $I_1(\theta)$.

(c) Compute the CRB for $1 + \overline{X}_n$ as an estimator of $q(\theta) = \dfrac{1}{1 - e^{-\theta}}$.

(d) Calculate the Maximum Likelihood Estimator (MLE) of $\beta(\theta) = e^{-\theta}$. Call it $\widehat{\beta}_{MLE}$.

(e) Assume that the regularity conditions hold under which:

$$\sqrt{n}(\widehat{\beta}_{MLE} - \beta(\theta)) \xrightarrow{D} N(0, W(\theta)).$$

$$W(\theta) = \frac{(e^\theta - 1)^2}{e^{3\theta}}$$

(f) Based on the answer from the previous part, provide the formula for an asymptotic confidence interval with approximate coverage probability $1 - \alpha$ for $\beta(\theta)$.

Solution:

a. Maximum Likelihood Estimator (MLE)

The log-likelihood function is:

$$l(\theta) = n \ln(1 - e^{-\theta}) - \theta \sum x_i$$

Taking the first derivative and setting it to zero:

$$\frac{\partial l}{\partial \theta} = \frac{n e^{-\theta}}{1 - e^{-\theta}} - \sum x_i = 0$$

Dividing by n and replacing $\frac{1}{n}\sum x_i$ with $\overline{X}_n$:

$$\frac{e^{-\theta}}{1 - e^{-\theta}} = \overline{X}_n \implies e^{-\theta} = \overline{X}_n - \overline{X}_n e^{-\theta}$$

$$e^{-\theta}(1 + \overline{X}_n) = \overline{X}_n$$

Solving for $\hat{\theta}_{MLE}$:

$$\hat{\theta}_{MLE} = -\ln\left(\frac{\overline{X}_n}{1 + \overline{X}_n}\right) = \ln\left(\frac{\overline{X}_n + 1}{\overline{X}_n}\right)$$

b. Fisher Information $I_1(\theta)$

The Fisher Information $I_1(\theta) = -E\left[\frac{\partial^2 \ln p(x)}{\partial \theta^2}\right]$. The second derivative of the log-likelihood for a single observation is:

$$\frac{\partial^2 \ln p(x)}{\partial \theta^2} = -\frac{e^{-\theta}}{(1-e^{-\theta})^2}$$

The Fisher Information is:

$$I_1(\theta) = E\left[\frac{e^{-\theta}}{(1-e^{-\theta})^2}\right] = \frac{e^{-\theta}}{(1-e^{-\theta})^2}$$

Simplifying the expression:

$$I_1(\theta) = \frac{e^{-\theta}}{\left(\frac{e^{\theta}-1}{e^{\theta}}\right)^2} = \frac{e^{-\theta}e^{2\theta}}{(e^{\theta}-1)^2} = \frac{e^{\theta}}{(e^{\theta}-1)^2}$$

c.CRB $q(\theta) = \frac{1}{1-e^{-\theta}}$

Unbiasedness:

$$E(1+\overline{X}_n) = 1 + E(X) = 1 + \frac{1}{e^{\theta}-1} = \frac{e^{\theta}-1+1}{e^{\theta}-1} = \frac{e^{\theta}}{e^{\theta}-1}$$

The target parameter is:

$$q(\theta) = \frac{1}{1-e^{-\theta}} = \frac{1}{\frac{e^{\theta}-1}{e^{\theta}}} = \frac{e^{\theta}}{e^{\theta}-1}$$

Then $E(1+\overline{X}_n) = q(\theta)$, the estimator is **unbiased**.

The CRB is

$$\frac{(q'(\theta))^2}{I(\theta)n} = \frac{\left(\frac{e^{\theta}}{(e^{\theta}-1)^2}\right)^2}{n\frac{e^{\theta}}{(e^{\theta}-1)^2}}$$

d. MLE of $\beta(\theta) = e^{-\theta}$

By the Invariance Property of the MLE, $\widehat{\beta}_{MLE} = \beta(\hat{\theta}_{MLE}) = e^{-\hat{\theta}_{MLE}}$. From the derivation in part 2, we have $e^{-\hat{\theta}_{MLE}} = \frac{\overline{X}_n}{1+\overline{X}_n}$.

$$\widehat{\beta}_{MLE} = \frac{\overline{X}_n}{1+\overline{X}_n}$$

e. Asymptotic Variance $W(\theta)$ for $\widehat{\beta}_{MLE}$

We use the Delta Method for $g(\overline{X}_n)$, where $g(u) = \frac{u}{1+u}$.

$E(X) = \frac{1}{e^{\theta}-1}$.

$\mathrm{Var}(X) = \frac{e^{\theta}}{(e^{\theta}-1)^2}$.

$$\sqrt{n}(\overline{X}_n - \frac{1}{e^{\theta}-1}) \to N(0, \frac{e^{\theta}}{(e^{\theta}-1)^2})$$

Derivative of $g(u)$:

$$g'(u) = \frac{\partial}{\partial u}\left(\frac{u}{1+u}\right) = \frac{1(1+u)-u(1)}{(1+u)^2} = \frac{1}{(1+u)^2}$$

Value of $g'(\mu)$:

$$1+\mu = 1+\frac{1}{e^{\theta}-1} = \frac{e^{\theta}}{e^{\theta}-1}$$

$$g'(\mu) = \left(\frac{e^{\theta}-1}{e^{\theta}}\right)^2$$

Asymptotic Variance $W(\theta)$:

$$W(\theta) = [g'(\mu)]^2 \cdot \mathrm{Var}(X) = \left[\left(\frac{e^{\theta}-1}{e^{\theta}}\right)^2\right]^2 \cdot \frac{e^{\theta}}{(e^{\theta}-1)^2}$$

$$W(\theta) = \frac{(e^{\theta}-1)^4}{e^{4\theta}} \cdot \frac{e^{\theta}}{(e^{\theta}-1)^2} = \frac{(e^{\theta}-1)^2}{e^{3\theta}}$$

f. Asymptotic Confidence Interval for $\beta(\theta)$

The Central Limit Theorem (CLT) establishes the asymptotic normality of the sample mean, $\overline{X}_n$. The **Delta Method** (used in part 6) ensures that the estimator $\widehat{\beta}_{MLE} = g(\overline{X}_n)$ is also asymptotically normal:

$$\sqrt{n}(\widehat{\beta}_{MLE} - \beta(\theta)) \xrightarrow{D} N(0, W(\theta))$$

For the confidence interval, we use the **Slutsky's Theorem** to replace the theoretical asymptotic variance $W(\theta)$ with its consistent estimator $\widehat{W}(\theta) = W(\hat{\theta}_{MLE})$ to form a standard normal pivot.

Formula for the Asymptotic Confidence Interval $(1-\alpha)$:

$$\widehat{\beta}_{MLE} \pm z_{\alpha/2}\sqrt{\frac{\widehat{W}(\theta)}{n}}$$

Where:

$\widehat{\beta}_{MLE} = \frac{\overline{X}_n}{1+\overline{X}_n}$

$\widehat{W}(\theta)$ is $W(\theta)$ evaluated at the MLE $\hat{\theta}_{MLE}$. Using $e^{-\hat{\theta}_{MLE}} = \frac{\overline{X}_n}{1+\overline{X}_n}$, it can be shown that $\widehat{W}(\theta) = \frac{\overline{X}_n^2}{(1+\overline{X}_n)^3}$.

$z_{\alpha/2}$ is the $(1-\alpha/2)$-th percentile of the standard normal distribution $N(0,1)$.

Problema 2

Suppose $X_i, i = 1, ..., n$ are independent random variables, each with density $f(x; \theta^*)$ where for any $\theta > 0$,

$$f(x;\theta) = \begin{cases} \theta x \text{ if } 0 < x \leq \sqrt{\frac{2}{\theta}} \\ 0 \text{ otherwise} \end{cases}$$

(a) Prove that the Method of Moments Estimator (MME) for θ is equal to $\widetilde{\theta} = \frac{8}{9(\overline{X})^2}$ where $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $\sqrt{n}\left(\widetilde{\theta} - \theta\right) \xrightarrow{L} N(0, v(\theta))$ (provide the formula for $v(\theta)$).

(b) Consider the estimator of θ defined as $\widehat{\theta} = \frac{2}{\max\left\{X_1^2,...,X_n^2\right\}}$. Prove that the cumulative distribution function $F(u) = P\left(\widehat{\theta} \leq u\right)$ is equal to:

$$F(u) = \begin{cases} 1 - \left(\frac{\theta}{u}\right)^n & \text{if } u > \theta \\ 0 & \text{otherwise} \end{cases}$$

(c) Is $\widehat{\theta}$ the Maximum Likelihood Estimator (MLE) for θ? If yes, prove it; otherwise, explain why.

(d) Using part (b), calculate a confidence interval with exact coverage probability equal to 0.95.

(e) Compute the biase and the MSE of $\widehat{\theta}$ the estimator defined in part (b).

Solution

a. Method of Moments Estimator (MME)

We equate the sample mean $\overline{X}_n$ to the theoretical expectation $E_\theta[X]$.

$$\begin{aligned} E_\theta[X] &= \int_0^{\sqrt{\frac{2}{\theta}}} \theta x^2 dx \\ &= \theta \frac{x^3}{3}\bigg|_0^{\sqrt{\frac{2}{\theta}}} = \theta \frac{\left(\sqrt{\frac{2}{\theta}}\right)^3}{3} = \theta \frac{\frac{2\sqrt{2}}{\theta\sqrt{\theta}}}{3} \\ &= \frac{2\sqrt{2}}{3\sqrt{\theta}} = \theta^{-1/2}\frac{2\sqrt{2}}{3} \end{aligned}$$

Equating $\overline{X}_n = E_\theta[X]$:

$$\overline{X}_n = \widetilde{\theta}^{-1/2}\frac{2\sqrt{2}}{3}$$

Solving for $\widetilde{\theta}$:

$$\widetilde{\theta}^{1/2} = \frac{2\sqrt{2}}{3\overline{X}_n} \implies \widetilde{\theta} = \frac{(2\sqrt{2})^2}{9(\overline{X}_n)^2} = \frac{8}{9(\overline{X}_n)^2}$$

We calculate the asymptotic variance $v(\theta)$ using the Delta Method for $g(\overline{X}_n)$, where $g(u) = \frac{8}{9u^2}$.

1. Population Mean: $\mu = E_\theta[X] = \theta^{-1/2}\frac{2\sqrt{2}}{3}$.
2. Population Variance: $\text{Var}_\theta(X) = E_\theta[X^2] - \mu^2$.

$$E_\theta\left[X^2\right] = \int_0^{\sqrt{\frac{2}{\theta}}} \theta x^3 dx = \theta \frac{x^4}{4}\bigg|_0^{\sqrt{\frac{2}{\theta}}} = \theta\frac{(2/\theta)^2}{4} = \frac{1}{\theta}$$

$$\text{Var}_\theta(X) = \frac{1}{\theta} - \left(\theta^{-1/2}\frac{2\sqrt{2}}{3}\right)^2 = \frac{1}{\theta} - \frac{8}{9\theta} = \frac{1}{9\theta}$$

$$\sqrt{n}(\overline{X}_n - \theta^{-1/2}\frac{2\sqrt{2}}{3}) \to N(0, \frac{1}{9\theta})$$

1. Derivative of $g(u)$: $g'(u) = -\frac{16}{9u^3}$.
2. Asymptotic Variance $v(\theta)$:

$$v(\theta) = [g'(\mu)]^2 \text{Var}_\theta(X)$$

   Substituting $\mu = \theta^{-1/2}\frac{2\sqrt{2}}{3}$:

$$\mu^3 = \theta^{-3/2}\frac{16\sqrt{2}}{27}$$

$$v(\theta) = \left(-\frac{16}{9 \cdot \theta^{-3/2}\frac{16\sqrt{2}}{27}}\right)^2 \cdot \frac{1}{9\theta} = \left(-\frac{3\theta^{3/2}}{\sqrt{2}}\right)^2 \cdot \frac{1}{9\theta} = \frac{9\theta^3}{2} \cdot \frac{1}{9\theta} = \frac{\theta^2}{2}$$

Result: $\sqrt{n}\left(\widetilde{\theta} - \theta\right) \xrightarrow{L} N\left(0, \frac{\theta^2}{2}\right)$, so $v(\theta) = \frac{\theta^2}{2}$.

b. CDF of $\widehat{\theta} = \frac{2}{\max\{X_i^2\}}$

Since $\max\{X_i^2\} \le \frac{2}{\theta}$, then $\widehat{\theta} = \frac{2}{\max\{X_i^2\}} \ge \theta$. Thus, $P\left(\widehat{\theta} \le u\right) = 0$ if $u \le \theta$.

For $u > \theta$:

$$\begin{aligned} P\left(\widehat{\theta} \le u\right) &= P\left(\frac{2}{\max\{X_i^2\}} \le u\right) = P\left(\max\{X_i^2\} \ge \frac{2}{u}\right) \\ &= 1 - P\left(\max\{X_i^2\} < \frac{2}{u}\right) \\ &= 1 - \prod_{i=1}^n P\left(X_i^2 < \frac{2}{u}\right) = 1 - \left(P\left(X_i < \sqrt{\frac{2}{u}}\right)\right)^n \end{aligned}$$

The simple probability is:

$$P\left(X_i < \sqrt{\frac{2}{u}}\right) = \int_0^{\sqrt{\frac{2}{u}}} \theta x dx = \theta \frac{x^2}{2}\bigg|_0^{\sqrt{\frac{2}{u}}} = \theta\frac{2/u}{2} = \frac{\theta}{u}$$

Substituting back:

$$F(u) = 1 - \left(\frac{\theta}{u}\right)^n \quad \text{if } u > \theta$$

c. Is $\widehat{\theta}$ the MLE?

Answer: Yes. The likelihood function is:

$$\mathcal{L}(\mathbf{X}, \theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^n \prod_{i=1}^n x_i \cdot \mathbb{I}_{\left(0, \sqrt{\frac{2}{\theta}}\right]}(\max\{X_i\})$$

The indicator function implies that θ must satisfy $\max\{X_i\} \le \sqrt{\frac{2}{\theta}}$, or $\theta \le \frac{2}{\max\{X_i^2\}}$.

$$\mathcal{L}(\mathbf{X}, \theta) = \theta^n \left(\prod_{i=1}^n x_i\right) \cdot \mathbb{I}_{\left(0, \frac{2}{\max\{X_i^2\}}\right]}(\theta)$$

The term θ^n is an increasing function of θ. The likelihood is zero for $\theta > \frac{2}{\max\{X_i^2\}}$. Therefore, the maximum is achieved at the largest possible value of θ in the parameter space:

$$\widehat{\theta}_{MLE} = \frac{2}{\max\{X_i^2\}}$$

d. Exact 0.95 Confidence Interval

We construct the interval from the exact probability $P(u_L \leq \widehat{\theta} \leq u_U) = 0.95$, where u_L and u_U are the 0.025 and 0.975 quantiles of $\widehat{\theta}$. From part 3, $P(\widehat{\theta} \leq u) = 1 - \left(\frac{\theta}{u}\right)^n$.

$$\begin{aligned}
P(\widehat{\theta} \leq u_L) = 0.025 &\implies 1 - \left(\frac{\theta}{u_L}\right)^n = 0.025 \implies \left(\frac{\theta}{u_L}\right)^n = 0.975 \\
&\implies u_L = \frac{\theta}{0.975^{1/n}} \\
P(\widehat{\theta} \leq u_U) = 0.975 &\implies 1 - \left(\frac{\theta}{u_U}\right)^n = 0.975 \implies \left(\frac{\theta}{u_U}\right)^n = 0.025 \\
&\implies u_U = \frac{\theta}{0.025^{1/n}}
\end{aligned}$$

The 0.95 probability statement is:

$$0.95 = P\left(\frac{\theta}{0.975^{1/n}} \leq \widehat{\theta} \leq \frac{\theta}{0.025^{1/n}}\right)$$

Solving for θ:

$$0.95 = P\left(0.025^{1/n}\widehat{\theta} \leq \theta \leq 0.975^{1/n}\widehat{\theta}\right)$$

The Exact Confidence Interval (0.95) is:

$$I_n = \left(0.025^{1/n}\widehat{\theta}, 0.975^{1/n}\widehat{\theta}\right)$$

e. Expected Values $E_\theta\left(\widehat{\theta}\right)$ and $Var_\theta\left(\widehat{\theta}^2\right)$

The probability density function is $f(u) = \frac{d}{du}F(u) = n\theta^n u^{-n-1}$ for $u > \theta$.

Expectation $E\left(\widehat{\theta}\right)$ (for $n > 1$):

$$\begin{aligned}
E\left(\widehat{\theta}\right) &= \int_\theta^\infty u \cdot n\theta^n u^{-n-1} du = n\theta^n \int_\theta^\infty u^{-n} du \\
&= n\theta^n \left[\frac{u^{-n+1}}{-n+1}\right]_\theta^\infty = n\theta^n \left(0 - \frac{\theta^{-n+1}}{1-n}\right) \\
&= \frac{n}{n-1}\theta
\end{aligned}$$

Expectation $E\left(\widehat{\theta}^2\right)$ (for $n > 2$):

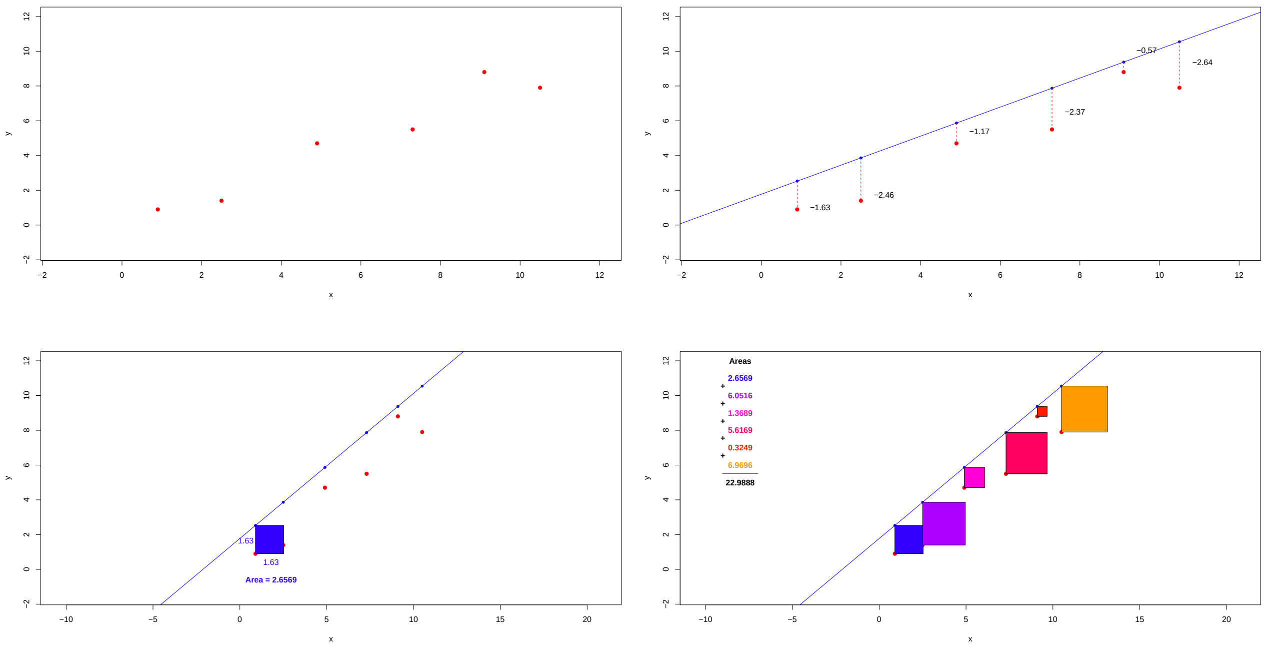
$$\begin{aligned}
E\left(\widehat{\theta}^2\right) &= \int_\theta^\infty u^2 \cdot n\theta^n u^{-n-1} du = n\theta^n \int_\theta^\infty u^{-n+1} du \\
&= n\theta^n \left[\frac{u^{-n+2}}{-n+2}\right]_\theta^\infty = n\theta^n \left(0 - \frac{\theta^{-n+2}}{2-n}\right) \\
&= \frac{n}{n-2}\theta^2
\end{aligned}$$

$$Var\left(\widehat{\theta}\right) = \frac{n}{n-2}\theta^2 - \frac{n^2}{(n-1)^2}\theta^2 = (\frac{n}{n-2} - \frac{n^2}{(n-1)^2})\theta^2$$

Mathematical Statistics

Tutorial 8

Simple linear regresion : Toy Example



Exercise: Soft Drink Delivery Time

The **Time** required for a worker to service and stock a soft drink dispensing machine as a function of the variables **Number of Cases** and **Distance**.

The data is available in the `MPV` package. The following R code loads the data and displays the first six observations. A total of 20 observations are available.

1. Explore the relationships between the variables using a pairwise scatterplot matrix. and a **3D scatterplot** .

2. Based on the 3D scatterplot, the model to be fitted is shown below:

$$\text{Time}_i = \beta_0 + \beta_1 \text{Cases}_i + \beta_2 \text{Distance}_i + \varepsilon_i$$

   Obtain the OLS of the unknown parameters.

3. Plot the estimated regression plane within the 3D scatterplot.

4. Predict the **average service time** ($\hat{\mu}$) for a hypothetical new observation where the worker has to handle **10 Cases** and the travel **Distance** is **400 feet**.

5. Calculate the 95% Confidence Intervals (CI) for all model coefficients ($\beta_0, \beta_1, \beta_2$).

6. Based on the CI for β_1 (**Cases**), does the interval contain zero? What is the statistical implication of this result regarding the variable's effect on **Time**?

Solution

```
## Call:
## lm(formula = Time ~ Cases + Distance, data=softdrink)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7880 -0.6629  0.4364  1.1566  7.4197
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.341231   1.096730   2.135 0.044170 *
## Cases       1.615907   0.170735   9.464 3.25e-09 ***
## Distance    0.014385   0.003613   3.981 0.000631 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.259 on 22 degrees of freedom
## Multiple R-squared:  0.9596, Adjusted R-squared:  0.9559
## F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

The estimated model based on these results is:

$$\widehat{\text{Time}}_i = 2.341 + 1.616 \cdot \text{Cases}_i + 0.014 \cdot \text{Distance}_i,$$

Interpretation

- **Distance ($\hat{\beta}_2 = 0.014$):** If the truck is moved one foot farther from the machine, the expected **average time** for maintenance is expected to increase by 0.014 minutes, assuming the number of cases is unchanged.
- **Distance (scaled):** If the truck is moved 100 feet farther, the expected average time is expected to increase by 1.4 minutes (100×0.014).
- **Cases ($\hat{\beta}_1 = 1.616$):** For each additional case of soft drink that must be carried, the expected **average time** is expected to increase by 1.616 minutes, assuming the distance is unchanged.
- **Intercept ($\hat{\beta}_0 = 2.341$):** If the truck is 0 feet away from the machine and zero cases need to be carried, the expected **average time** for maintenance is 2.341 minutes.

The prediction

$$\widehat{\text{Time}}_i = 2.341 + 1.616 \cdot 10 + 0.014 \cdot 400$$

An approximate 0.95 confidence interval for a specific coefficient β_j is given by:

$$\hat{\beta}_j \pm z_{\alpha/2}\hat{\text{se}}(\hat{\beta}_j)$$

where $\hat{\text{se}}^2(\hat{\beta}_j)$ is the j-th diagonal element of the estimated covariance matrix, $\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}$.

Mathematical Statistics

Tutorial 9

Consider the Multiple Linear Regression Model, which includes an intercept term (β_0) and k explanatory variables ($X_1, \dots, X_k$).

$$\mathbf{Y} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

The model is estimated using the Ordinary Least Squares (OLS) method. Prove the following properties satisfied by the OLS residuals ($\hat{\epsilon}_i = Y_i - \hat{Y}_i$) .

1. **Mean of Residuals is Zero:** $\sum_{i=1}^{n} \hat{\epsilon}_i = 0$
2. **Orthogonality to Predicted Values:** ($\hat{Y}_i$). $\sum_{i=1}^{n} \hat{\epsilon}_i \hat{Y}_i = 0$
3. **Sum of Squares Decomposition:** The Total Sum of Squares (SSTo) equals the Regression Sum of Squares (SSReg) plus the Residual Sum of Squares (SSRes).

$$\text{SSTo} = \text{SSReg} + \text{SSRes}$$

Where:

- $\text{SSTo} = \sum (Y_i - \bar{Y})^2$ (Total Sum of Squares)
- $\text{SSReg} = \sum (\hat{Y}_i - \bar{Y})^2$ (Regression Sum of Squares)
- $\text{SSRes} = \sum (Y_i - \hat{Y}_i)^2$ (Residual Sum of Squares)

Solution

The minimization objective for OLS is: $\text{SSRes} = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots)^2$.

1. Proof that $\sum \hat{\epsilon}_i = 0$

This property is derived from the FOC with respect to the **intercept** ($\hat{\beta}_0$).

$$\frac{\partial \text{SSRes}}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots) = 0$$

Dividing by -2 yields:

$$\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \dots) = 0$$

Since the term in the parenthesis is the residual $\hat{\epsilon}_i$:

$$\sum_{i=1}^{n} \hat{\epsilon}_i = 0$$

2. Proof that $\sum \hat{\epsilon}_i \hat{Y}_i = 0$

We substitute the definition of the predicted value, $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}$, into the cross-product term:

$$\sum \hat{\epsilon}_i \hat{Y}_i = \sum \hat{\epsilon}_i (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})$$

Expanding the summation:

$$\sum \hat{\epsilon}_i \hat{Y}_i = \hat{\beta}_0 \sum \hat{\epsilon}_i + \hat{\beta}_1 \sum \hat{\epsilon}_i X_{i1} + \cdots + \hat{\beta}_k \sum \hat{\epsilon}_i X_{ik}$$

- **Intercept Term:** By property (1), $\sum \hat{\epsilon}_i = 0$. Thus, $\hat{\beta}_0 \sum \hat{\epsilon}_i = 0$.
- **Slope Terms:** The FOCs for the slope coefficients ($\hat{\beta}_j$) require that the residuals be orthogonal to all explanatory variables: $\sum \hat{\epsilon}_i X_{ij} = 0$ for all $j = 1, \ldots, k$.

Substituting these zeros back into the expanded equation:

$$\sum \hat{\epsilon}_i \hat{Y}_i = 0 + 0 + \cdots + 0 = 0$$

$$\sum \hat{\epsilon}_i \hat{Y}_i = 0$$

3. Proof of SSTo = SSReg + SSRes

We start by decomposing the total deviation $(Y_i - \bar{Y})$ into the explained part and the unexplained part, using the identity $Y_i = \hat{Y}_i + \hat{\epsilon}_i$:

$$Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) = \hat{\epsilon}_i + (\hat{Y}_i - \bar{Y})$$

Squaring both sides and summing over i:

$$\sum (Y_i - \bar{Y})^2 = \sum \left[\hat{\epsilon}_i + (\hat{Y}_i - \bar{Y})\right]^2$$

Expanding the right-hand side:

$$\sum (Y_i - \bar{Y})^2 = \sum \hat{\epsilon}_i^2 + \sum (\hat{Y}_i - \bar{Y})^2 + 2 \sum \hat{\epsilon}_i (\hat{Y}_i - \bar{Y})$$

We analyze the **cross-product term**: $2 \sum \hat{\epsilon}_i (\hat{Y}_i - \bar{Y})$.

$$2 \sum \hat{\epsilon}_i (\hat{Y}_i - \bar{Y}) = 2 \left[\sum \hat{\epsilon}_i \hat{Y}_i - \bar{Y} \sum \hat{\epsilon}_i\right]$$

Using the OLS properties proven above: $\sum \hat{\epsilon}_i \hat{Y}_i = 0$ (Property 2) and $\sum \hat{\epsilon}_i = 0$ (Property 1).

$$2 \sum \hat{\epsilon}_i (\hat{Y}_i - \bar{Y}) = 2[0 - \bar{Y}(0)] = 0$$

Substituting this back, the main equation becomes:

$$\sum (Y_i - \bar{Y})^2 = \sum \hat{\epsilon}_i^2 + \sum (\hat{Y}_i - \bar{Y})^2$$

Substituting the definitions of the Sums of Squares:

$$\text{SSTo} = \text{SSRes} + \text{SSReg}$$

Interpretation

To evaluate how well a linear model fits the observed data, we decompose the total variability in the dependent variable (Y) into two components: the variability explained by the model and the unexplained (residual) variability.

Total Sum of Squares (SSTo or SST) represents the total variation in the response variable (Y) around its mean ($\bar{Y}$). It is the measure of variability we aim to explain.

$$\text{SSTo} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

Residual Sum of Squares (SSRes or SSE) represents the variation in Y that **is not** explained by the regression model. It is the sum of the squared differences between the observed values (Y_i) and the values predicted by the model ($\hat{Y}_i$).

$$\text{SSRes} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \hat{\epsilon}_i^2$$

In matrix notation, where $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ is the projection matrix, we have:

$$\text{SSRes} = \hat{\boldsymbol{\epsilon}}^\top\hat{\boldsymbol{\epsilon}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^\top[\mathbf{I} - \mathbf{P}]\mathbf{Y}$$

Regression Sum of Squares (SSReg or SSM) represents the variation in Y that **is** explained by the regression model. It is the sum of the squared differences between the predicted values ($\hat{Y}_i$) and the mean of the dependent variable ($\bar{Y}$).

$$\text{SSReg} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$$

The Coefficient of Determination (R^2)

The coefficient of determination, R^2, is defined as the ratio of the variability explained by the model (SSReg) to the total variability in the response variable (SSTo). It measures the proportion of the total variation in Y that is accounted for by the regression model.

$$R^2 = \frac{\text{SSReg}}{\text{SSTo}} = 1 - \frac{\text{SSRes}}{\text{SSTo}}$$

R^2 represents the proportion of the total variability in Y that is **explained** by the set of independent variables (X's) in the regression model.

Properties of R^2

- $0 \le R^2 \le 1$.
- It is a dimensionless measure, independent of the units of measurement of Y.
- A higher R^2 indicates a stronger relationship between the dependent and independent variables, and better predictive power. $R^2 = 1$ implies a perfect fit (all observations lie exactly on the regression surface), and $R^2 = 0$ indicates no linear association.
- The square root of R^2, denoted by R (the **Multiple Correlation Coefficient**), is the Pearson correlation coefficient between the observed values of the response variable (Y_i) and the values predicted by the model ($\hat{Y}_i$).

$$R = \text{Corr}(Y_i, \hat{Y}_i) = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}}$$

In Simple Linear Regression, R is also equal to the absolute value of the correlation coefficient between Y_i and the single covariate X_i: $R = |r_{XY}|$.

- A crucial property of the standard R^2 is that **adding explanatory variables (X's) to the regression model can only increase R^2 or keep it the same; it can never decrease it.** This is because the SSRes can never increase with more covariates, while SSTo remains constant for a fixed set of Y observations.

  Because R^2 does not penalize for the inclusion of non-significant variables, it is **unsuitable for comparing models** with a different number of covariates.

The Adjusted Coefficient of Determination (R^2 Adjusted)

The Adjusted R^2 corrects the standard R^2 by dividing the sums of squares by their corresponding. This introduces a penalty for adding extra explanatory variables ($p = k+1$ the numer of parameters including the intercept increases).

$$R_a^2 = 1 - \frac{\text{SSRes}/(n-p)}{\text{SSTo}/(n-1)} = 1 - \frac{\text{MSRes}}{\text{MSTo}}$$

The Adjusted R^2 can **decrease** when a new covariate is added to the model if that variable does not sufficiently reduce the SSRes to compensate for the loss of a degree of freedom in the denominator $(n-p)$. **Interpretation for Model Comparison:** When comparing two nested models (e.g., Model 1 with $X_1, \ldots, X_k$ and Model 2 with $X_1, \ldots, X_k, X_{k+1}$), an increase in R_a^2 indicates that the added covariate (X_{k+1}) is important for predicting Y. Conversely, if R_a^2 remains the same or decreases, it signals that the new variable does not significantly contribute to explaining Y and should not be included in the model.

Mathematical Statistics

Tutorial 10

Exercise 1: Empirical Distribution and Plug-in Estimation

1. Program the Empirical Cumulative Distribution Function (ECDF) $F_n(x)$ as a function of the dataset and a value of x.
2. A dataset containing Total Family Income (ITF) data from Argentina (2023). Consider only the positive observations.
   (a) Plot the $F_n(x)$ for the income data.
   (b) Using the Plug-in Principle and the estimated $F_n(x)$, estimate the proportion of Argentine families whose total income X falls strictly between \$40,000 and \$50,000 Pesos. Express this as the estimate of $P(40,000 < X \leq 50,000)$.
   (c) Using the Plug-in Principle and the estimated $F_n(x)$, estimate the proportion of Argentine families whose total income X is strictly greater than \$100,000 Pesos. Express this as the estimate of $P(X > 100,000)$.
   (d) Calculate the Plug-in Estimators for the population's two central tendency measures: the **Median** (0.5 Quantile) and the **Mean** income.

Exercise 2: Parametric vs. Non-Parametric Bootstrap

A sample dataset containing the lifetimes of $n = 30$ lamps is generated from a Gamma distribution. The sample median is the chosen estimator for the population median. Use $B = 1000$ bootstrap repetitions for both procedures.

Perform the following two distinct procedures to estimate the Standard Error (SE) of the median estimator $\hat{\theta}$:

1. **Parametric Bootstrap:**
   - Estimate the shape and rate parameters $(\hat{\alpha}, \hat{\beta})$ of the Gamma distribution from the original sample.
   - In each bootstrap iteration, generate a new sample from the **estimated Gamma distribution**, re-estimate the Gamma parameters from the bootstrap sample, and calculate the median based on these new estimated parameters ($Q_{0.5}$ from Gamma$(\hat{\alpha}^*, \hat{\beta}^*)$).
   - Report the resulting Bootstrap Standard Error $\text{SE}_B(\hat{\theta})$.
2. **Non-Parametric Bootstrap:**
   - Generate $B = 1000$ bootstrap samples by **resampling with replacement** directly from the original data.
   - Calculate the median directly from each bootstrap sample ($Q_{0.5}$ from empirical distribution).
   - Report the resulting Bootstrap Standard Error $\text{SE}_B(\hat{\theta})$.

Mathematical Statistics

Tutorial 11

Exercise 1

A dataset containing Total Family Income (ITF) data from Argentina (2023). Consider only the positive observations. The median is the estimator $\hat{\theta}$ chosen for the center of the Total Family Income (ITF) population distribution. Construct the Approximate Normal 95% Confidence Interval (CI) for the true population median θ using the formula.

Exercise 2

1. Plot the theoretical Cumulative Distribution Functions (CDF) for a Normal distribution ($\mu = 4, \sigma = 2$) and a Uniform distribution ($min = 1, max = 7$) on a single plot using the `pnorm` and `punif` functions.

2. Generate a random sample of $n = 1000$ for both distributions. How do you visualize their Empirical Cumulative Distribution Functions using the `ecdf()` function in a single graphic?

3. Plot the theoretical PDFs for the Normal(4,2) and Uniform(1,7) distributions.

4. Display the frequency histograms for both simulated datasets side-by-side using a 1x2 layout grid.

5. Compare the frequency histograms for both simulated datasets using diferents samples size.

Exercise 3

Using the `buffalo_snow` data generates three different histograms using 5, 15, and 40 breaks respectively to demonstrate the effect of "oversmoothing" and "undersmoothing".

Plot a histogram for the Buffalo data (using 15 breaks) that overlaps a Kernel Density Estimate (KDE) line and includes a `rug()` plot at the bottom to show individual data points?

Mathematical Statistics

Homework 1

1. Show that the following distributions belong to the exponential family (you may need to hold one of their parameters fixed):
    - The Poisson distribution
    - The geometric distribution
    - The negative binomial distribution
    - The exponential distribution
    - The gamma distribution
    - The chi-square distribution

2. Linear Estimators and Minimum Variance: Let $X_1, \ldots, X_n$ be a random sample (r.s.) from a distribution such that $E(X_1) = \mu$ exists. Consider the class of estimators:

$$\mathcal{D} = \left\{ \delta(\mathbf{X}) = \sum_{i=1}^{n} \alpha_i X_i : \alpha_i \in R, \ \sum_{i=1}^{n} \alpha_i = 1 \right\}.$$

    (a) Prove that every $\delta \in \mathcal{D}$ is an unbiased estimator for μ.

    (b) Assume $V(X) = \sigma^2$ exists. Show that the sample mean $\bar{X}$ is the estimator that minimizes the variance within this class: $\bar{X} = \arg\min_{\delta \in \mathcal{D}} V(\delta)$.

3. Let $X_1, \ldots, X_n$ be a r.s. from a distribution such that $V(X) = \sigma^2$ exists.

    (a) Prove that the sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ is an unbiased estimator for σ^2.

    (b) Assume $X_1, \ldots, X_n$ is a r.s. from a Normal distribution $N(\mu, \sigma^2)$.

        i. Calculate the MSE for s^2 and for $\widetilde{\sigma}^2$, where $\widetilde{\sigma}^2 = \frac{n-1}{n+1} s^2$.

        ii. Prove that, although $\text{MSE}(s^2) > \text{MSE}(\tilde{\sigma}^2)$, the ratio of the MSE's tends to 1 as $n \to \infty$.

        iii. Show that the contribution of the bias of $\tilde{\sigma}^2$ to $\text{MSE}(\tilde{\sigma}^2)$ is negligible as $n \to \infty$ (in the sense that the ratio between $b(\tilde{\sigma}^2)^2$ and $\text{MSE}(\tilde{\sigma}^2)$ tends to 0).

Mathematical Statistics

Homework 2

1. Let $X_1, \ldots, X_n$ be a random sample (r.s.) from a shifted exponential distribution, whose density is $f(x) = e^{-(x-\theta)} I_{[\theta,\infty)}(x)$.

    (a) Find the Maximum Likelihood Estimator (MLE) of θ.

    (b) Find the Method of Moments Estimator (MME) of θ.

2. Let $X_1, \ldots, X_n$ be a r.s. from a $N(\mu_1, \sigma^2)$ distribution and $Y_1, \ldots, Y_m$ be a r.s. from a $N(\mu_2, \sigma^2)$ distribution, independent of each other.

    (a) Find the MLE of $\theta = (\mu_1, \mu_2, \sigma^2)$.

    (b) Find the MLE of $\alpha = \mu_1 - \mu_2$.

3. The number of microorganisms found per cluster on a surface follows the following distribution:

$$p(x) = \theta I_{\{1\}}(x) + \frac{1-\theta}{k-1} I_{\{2,3,\ldots,k\}}(x)$$

where $0 < \theta < 1$. Suppose n clusters are examined independently and the number of microorganisms $X_1, \ldots, X_n$ is counted in each one. Calculate the Maximum Likelihood Estimator (MLE) of θ and the Moments Estimator (MME) of θ.

Mathematical Statistics

Homework 3

1. Let $X_1, \ldots, X_n$ be a random sample (r.s.) from an exponential distribution $\mathcal{E}(\theta)$ (where θ is the rate parameter).
   (a) Prove that the MLE of θ is strongly consistent and asymptotically efficient.
   (b) Obtain a Confidence Interval (CI) of level $1-\alpha$ for θ, specifying if it is exact or asymptotic.

2. Let $X_1, \ldots, X_n$ be a r.s. from a Poisson distribution $\mathcal{P}(\lambda)$. Consider $\delta_n = \bar{X}$ and the estimator

$$\delta_n^* = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{n}\sum_{i=1}^{n} X_i^2}$$

   (a) Analyze whether these estimators of λ are strongly consistent.
   (b) Find their asymptotic distributions. State if either is asymptotically efficient.
   (c) Obtain a Confidence Interval (CI) of level $1-\alpha$ for λ based on the most efficient estimator, specifying if it is exact or asymptotic.

   (Note: $E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$)

3. Let $X_1, \ldots, X_n$ be a r.s. from a shifted exponential distribution, whose density is

$$f(x) = e^{-(x-\theta)} I_{[\theta,\infty)}(x).$$

   (a) Prove that the MLE of θ is weakly and strongly consistent and find its asymptotic distribution.
   (b) Obtain a Confidence Interval (CI) of level $1-\alpha$ for θ based on the MLE, specifying if it is exact or asymptotic.
   (c) Prove that the Method of Moments Estimator (MME) for θ is consistent and find its asymptotic distribution.
   (d) Obtain a Confidence Interval (CI) of level $1-\alpha$ for θ based on the MME, specifying if it is exact or asymptotic.
   (e) Which of these two estimators would you prefer? Why?

Mathematical Statistics

Homework 4

1. Let $X_1, \ldots, X_n$ be a random sample (r.s.) from a Poisson distribution $\mathcal{P}(\lambda)$, and let $T = \sum_{i=1}^{n} X_i$ be the sufficient statistic.

   (a) Let $\mu = e^{-\lambda}$. Show that $\hat{\mu} = I\{X_1 = 0\}$ is an unbiased estimator of μ.

   (b) Apply the Rao–Blackwell Theorem to obtain an unbiased estimator $\delta^*(T)$ that is better than $\hat{\mu}$, and compare the MSEs (ECMs) of both estimators.

   (c) Find the UMVUEs for λ, λ^2, and μ.

2. Let $X_1, \ldots, X_n$ be a r.s. from a Normal distribution $N(\mu, \sigma_0^2)$, with σ_0^2 known.

   (a) Show that the sample mean $\bar{X}$ is a UMVUE for μ, using the Rao–Cramér inequality.

   (b) Show that $\bar{X}^2 - \sigma_0^2/n$ is a UMVUE for μ^2, although it does not attain the Rao–Cramér bound.

3. Let $X_1, \ldots, X_n$ be i.i.d. with density $f(x;\theta) = \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{\theta^{\alpha}} \mathrm{e}^{-\frac{x}{\theta}}$. Assume θ is unknown but α is known.

   (a) Demonstrate that $T = \frac{1}{n\alpha} \sum_{i=1}^{n} X_i$ is the Maximum Likelihood Estimator (MLE) of θ.

   (b) Demonstrate that T is an unbiased estimator of the scale parameter θ.

   (c) Calculate the Cramér-Rao bound and determine if T has minimum variance among all unbiased estimators.

Mathematical Statistics

Homework 5

1. Consider the Multiple Linear Regression model with a design matrix $\mathbf{X} \in \mathbb{R}^{n\times 2}$ consisting of a first column of ones and a second column containing the observations of the explanatory variable $X \in \mathbb{R}$.

   Prove that the Least Squares Estimators obtained from the matrix formulation coincide with the estimators for the case of Simple Linear Regression.

2. Suppose the following Simple Linear Regression Model (SLR) where the explanatory variable (X_i) takes values $X_i = 1$ or $X_i = 0$.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

   Where Y is the Salary and $X = 1$ if the individual has a university degree, $X = 0$ if they do not. Let n_1 be the number of observations with $X = 1$, n_0 be the number of observations with $X = 0$, and $n = n_0 + n_1$.

   (a) Prove that the sample mean of the residuals is zero in each of the two groups defined by X.

   (b) Prove that the Least Squares Estimators estimator for the slope, $\hat{\beta}_1$, is equivalent to the difference between the sample means of Y for the two groups:

$$\hat{\beta}_1 = \bar{Y}_1 - \bar{Y}_0$$

   (c) How is the assumption of Homoskedasticity $(Var(\epsilon_i) = \sigma^2)$ translated in this context? What is the implication if this assumption is violated because the $X = 1$ group has a much larger variance than the $X = 0$ group? Do the estimators remain unbiased?

Mathematical Statistics

Homework 6

1. **Expectation of the Uniform Kernel Density Estimator**
   Consider the Kernel Density Estimator (KDE) $\hat{f}_n(x)$ with bandwidth $h > 0$ and the Uniform (or Rectangular) kernel K:
   $$K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
   Obtain the expected value $E[\hat{f}_n(x)]$.

2. **Optimal Bandwidth Selection (Rule of Thumb)**
   Obtain the Asymptotically Optimal Bandwidth (h_{opt}) when the density estimation is computed using the Epanechnikov kernel, assuming the target density is $N(0, \sigma^2)$ (i.e., $X_1, \ldots, X_n$ are i.i.d. following a Normal distribution).

3. **Comparison of Distribution Function Estimators**
   Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with an unknown cumulative distribution function (CDF) $F(x)$ and a continuous density function $f(x)$. We consider two estimators for $F(x)$:

   - **The Empirical Distribution Function (EDF):** $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \leq x)$
   - **The Kernel Distribution Estimator (KDE):** $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}_n(u)du$

   where $\hat{f}_n(u)$ is the kernel density estimator with a symmetric kernel K and bandwidth $h > 0$.

   (a) Prove that the KDE can be expressed as:
   $$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{x - X_i}{h}\right)$$
   where $G(u) = \int_{-\infty}^{u} K(t)dt$.

   (b) Show that the Kernel Distribution Estimator ($\hat{F}$) is biased and derive its leading bias term using a Taylor expansion.

   (c) Discuss the advantages and disadvantages of using $\hat{F}$ over F_n. In which scenarios might one prefer the smooth estimator?