Mathematical Statistics

Tutorial 3

1. Let $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$ 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.

3. Let $X_1, \dots, X_{n_1}$ be an independent random sample from a Normal distribution $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \dots, Y_{n_2}$ 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 $X = (X_1, \dots, X_n)$ 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 $\theta_2 > 0$.

   (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 $X_1, \dots, X_{n_1}$ be an independent random sample from a Normal distribution $\mathcal{N}(\mu_1, \sigma^2)$. Let $Y_1, \dots, Y_{n_2}$ 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 $X = (X_1, \ldots, X_n)$ 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 $\theta_2 > 0$.

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 $X_1, \ldots, X_n$ 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 $\theta_1 - \theta_2$.