Maximum Likelihood Estimation (MLE) and Its Relationship with Ordinary Least Squares (OLS)

1 Introduction to MLE

Maximum Likelihood Estimation (MLE) is a powerful method for estimating the **parameters** of a presumed probability distribution based on observed data. The core idea is to maximize the likelihood function so that, given the assumed statistical model, **the observed data is the most probable outcome**.

1.1 Steps to Perform MLE

The typical steps involved in conducting MLE are:

- **Specify the Model:** Based on the data, specify the presumed probability distribution and identify the parameters of interest.
- **Formulate the Likelihood Function:** Write down the likelihood function using the probability density function (PDF) of the specified distribution, then take the natural logarithm of it.
- **Maximize the Log-Likelihood:** Obtain the parameter estimates by taking the derivative (first-order condition) of the log-likelihood and setting it to zero.

2 Illustrative Example: Estimating the Mean of a Normal Distribution

To illustrate MLE, consider a dataset $\{x_1, x_2, ..., x_n\}$ drawn from a normal distribution:

$$X_i \sim \mathcal{N}(\mu, \sigma^2). \tag{1}$$

2.1 Constructing the Likelihood Function

The probability density function (PDF) of a normal distribution is:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{2}$$

Thus, the likelihood function for a sample $X = \{x_1, x_2, ..., x_n\}$ is:

$$L(\mu,\sigma^2|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right). \tag{3}$$

2.2 Deriving the Log-Likelihood Function

Taking the natural logarithm of the likelihood:

$$\log L(\mu,\sigma^2|X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{4}$$

2.3 Maximizing the Log-Likelihood for Parameter Estimation

To find the MLE for μ, take the derivative with respect to μ and set it to zero:

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0. \tag{5}$$

Solving for μ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{6}$$

Thus, the MLE estimate for the mean is simply the sample mean.

3 Connecting MLE to OLS

Under certain conditions, MLE and OLS yield identical parameter estimates in linear regression models.

3.1 Simple Linear Regression Model

Consider the linear model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \text{for } i = 1, \dots, n \tag{7}$$

where:

- y_i is the dependent variable,
- x_i is the independent variable,

- n is the number of observations,
- β_0, β_1 are the parameters to be estimated,
- ϵ_i is the error term.

3.2 Assumptions for MLE and OLS Equivalence

The equivalence holds under the following assumptions:

- **Linearity:** $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.
- **i.i.d. Errors:** Errors ϵ_i are independently and identically distributed.
- **Normality:** $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
- **Homoscedasticity:** $\mathbb{E}[\epsilon_i^2] = \sigma^2$.
- **No Endogeneity:** $\mathbb{E}[\epsilon_i | x_i] = 0$.

3.3 Mathematical Justification

Comparing the objective functions of OLS and MLE reveals that they yield identical parameter estimates.

3.3.1 Ordinary Least Squares (OLS)

The OLS estimator minimizes the sum of squared residuals:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{8}$$

3.3.2 Maximum Likelihood Estimation (MLE)

Under the normality assumption, the likelihood function is:

$$L(\beta, \sigma^2 | X, y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right). \tag{9}$$

Taking the log-likelihood:

$$\log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{10}$$

Maximizing this function with respect to β_0 and β_1 is equivalent to minimizing the sum of squared residuals, which is exactly what OLS does.