1. Introduction

This lecture aims to introduce the Maximum Likelihood Estimation (MLE) method for determining the parameters of a normal distribution, specifically the mean μ and variance σ^2. MLE is a widely used statistical inference technique that finds the parameter values which maximize the probability of observing the given data.

2. The Normal Distribution Model

Assume that the random variable X follows a normal distribution with the probability density function:

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

For a set of independent and identically distributed (i.i.d.) samples $x_1, x_2, \ldots, x_n$, the joint probability density (likelihood function) is:

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

3. Log-Likelihood Function

To simplify the calculations, we take the logarithm of the likelihood function to obtain the log-likelihood:

$$\ell(\mu,\sigma^2) = \ln L(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

4. Derivation of the Maximum Likelihood Estimators

4.1 Estimation of the Mean μ

Differentiate the log-likelihood $\ell(\mu,\sigma^2)$ with respect to μ:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Setting the derivative equal to zero:

$$\sum_{i=1}^{n}(x_i-\mu) = 0$$

This gives the maximum likelihood estimator for the mean:

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

4.2 Estimation of the Variance σ^2

Differentiate the log-likelihood with respect to σ^2:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2$$

Set the derivative to zero:

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

Multiply both sides by $2\sigma^4$:

$$-n\sigma^2 + \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

Solving for σ^2 gives:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Note that this estimator for variance is biased in finite samples, though the bias diminishes as the sample size increases.