# Lecture 30 and 31

## Maximum Likelihood Estimation (MLE)

After covering the idea of estimating the CI of an unknown parameter, we will proceed to introduce the problem of Point Estimation. There are several procedures to find a point estimate of unknown parameter $\theta$ such as Method of Moments(MME), Unbiased Estimator, Bayes Estimators, Maximum Likelihood Estimation(MLE). Out of so many procedures available, we will derive the MLE of $\theta$. First, let us discuss some of its basic concepts.

# 1 Basic Concepts of Point Estimation

**Definition 1.1** Estimator:- *Any function of the random variables which is used to estimate the unknown value of the given parametric function $g(\theta)$ is called an estimator.*

If $\underline{X} = (X_1, \ldots, X_n)$ is a random sample from a population with the probability distribution $P_\theta$, a function $d(X)$ used for estimating $g(\theta)$ is known as an estimator. Let $\underline{x} = (x_1, \ldots, x_n)$ be a realization of $\underline{X}$, then $d(\underline{x})$ is called an estimate.

**Example 1.1** *Let $X$ be a random variable denoting the average height of Adult males in an ethnic group. We may use $\underline{X}$(Sample Mean) as an estimator. Now, if a random sample of* 50 *has a sample mean* 180, *then* $180cm$ *is an estimate of the average height.*

**Definition 1.2** *Let $\underline{x} = (x_1, \ldots, x_n)$ be an observed random sample. Define Likelihood function of $\underline{x}$ as*

$$L(\underline{\theta}, \underline{x}) = \prod_{i=1}^{n} f(x_i, \underline{\theta})$$

The value of $\theta$, say $\hat{\theta}(\underline{x})$, so that $L(\hat{\underline{\theta}}, \underline{x}) \geq L(\underline{\theta}, \underline{x})$ is called the MLE of $\theta$. It may be noted that, the idea of taking the likelihood function is that it contains all the informations of the parameters. In order to find out the MLE of an unknown parameter $\theta$ is to find out the value of $\theta$ for which the likelihood function is maximum.

**Example 1.2** *Consider a random sample $x_1, x_2, ..., x_n$ from a normal distribution $\mathcal{N}(\mu, \sigma)$. Find the maximum likelihood estimators for $\mu$ and $\sigma^2$.*

**Solution:-** The likelihood function for the normal distribution is

$$L(x_1, \cdots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left[ -\frac{1}{2}\sum_{i=1}^{n}\Big(\frac{x_i - \mu}{\sigma}\Big)^2 \right].$$

Taking logarithms gives us

$$lnL(x_1, \cdots, x_n; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\Big(\frac{x_i - \mu}{\sigma}\Big)^2.$$

Hence,

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

and

$$\frac{\partial \ln L}{\partial \mu} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\big(x_i - \mu\big)^2.$$

Setting both derivatives equal to 0, we obtain

$$\sum_{i=1}^{n} x_i - n\mu = 0 \text{ and } n\sigma^2 = \sum_{i=1}^{n}\big(x_i - \mu\big)^2.$$

Thus, the maximum likelihood estimator of $\mu$ is given by

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

which is a pleasing result since $\overline{x}$ has played such an important role in this chapter as a point estimate of $\mu$. On the other hand, the maximum likelihood estimator of $\sigma^2$ is

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Checking the second-order partial derivative matrix confirms that the solution results in a maximum of the likelihood function. Students may consider different cases to find out the MLE of $\mu$ when $\sigma^2$ is known vis-a-vis the MLE of $\sigma^2$ when $\mu$ is known.

**Example 1.3** *Suppose* 10 *rats are used in a biomedical study where they are injected with cancer cells and then given a cancer drug that is designed to increase their survival rate. The survival times, in months, are* $14, 17, 27, 18, 12, 8, 22, 13, 19$ *and* 12. *Assume that the exponential distribution applies. Give a maximum likelihood estimate of the mean survival time.*

**Solution:-** From Chapter 6, we know that the probability density function for the exponential random variable X is

$$f(x,\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & elsewhere. \end{cases}$$

Thus, the log-likelihood function for the data, given $n = 10$, is

$$\ln L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \ln \beta - \frac{1}{\beta} \sum_{i=1}^{10} x_i.$$

Setting

$$\frac{\partial \ln L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{10} x_i = 0$$

implies that

$$\hat{\beta} = \frac{1}{10} \sum_{i=1}^{10} x_i = \overline{x} = 16.2.$$

Evaluating the second derivative of the log-likelihood function at the value $\hat{\beta}$ above yields a negative value. As a result, the estimator of the parameter $\beta$, the population mean is the sample average $\overline{x}$.

**Example 1.4** *Q-81 Suppose that there are n trials $x_1, x_2, ..., x_n$ from a Bernoulli process with parameter p, the probability of a success. That is, the probability of r successes is given by $\binom{n}{r} p^r (1-p)^{n-r}$. Work out the maximum likelihood estimator for the parameter p.*

**Solution:-** Let $X = X_1 + \ldots + X_n$, then we see that it contains all the information of the unknown parameter $p$. Therefore $X$ alone is sufficient to estimate $p$. Also we observe that $X \sim B(n,p)$(Sum of $n$ Bernoulli random variables). So, the likelihood function can be written as

$$L(x,p) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n \text{ and } 0 < p < 1(\text{unknown}).$$

$$logL = log \binom{n}{x} + x \; logp + (n-x) \; log(1-p).$$

$$\frac{\partial L}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)} \begin{cases} > 0 & p < \frac{x}{n} \\ < 0 & p > \frac{x}{n}. \end{cases}$$

From the above expression we get $\hat{p}_{MLE} = \frac{X}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$.

**Example 1.5** *Q-85 Consider a random sample of $x_1, x_2, ..., x_n$ from a uniform distribution $U(0, \theta)$ with unknown parameter $\theta$, where $\theta > 0$. Determine the maximum likelihood estimator of $\theta$.*

**Solution:-** Let $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta), \theta$ is unknown. So, $\hat{\theta}_{MLE} = X_{(n)} = \max(X_1, \ldots, X_n)$. We observe that the likelihood function $L = \frac{1}{\theta^n}$ is a decreasing function of $\theta$ and hence the maximum occurs at the lower bound.

**Assignment:-** Consider a hypothetical experiment where a man with a fungus uses an antifungal drug and is cured. Consider this, then, a sample of one from a Bernoulli distribution with probability function $f(x) = p^x q^{1-x}, x = 0, 1$, where $p$ is the probability of a success (cure) and $q = 1 - p$. Now, of course, the sample information gives $x = 1$. Write out a development that shows that $\hat{p} = 1.0$ is the maximum likelihood estimator of the probability of a cure.