1. Introduction

Welcome to the series of E-learning modules on Methods of Estimation. In this module, we are going to cover the various methods of estimation- Point Estimation and Interval estimation. Methods of Point estimation – method of moments and method of maximum likelihood, their limitations, advantages and properties.

By the end of this session, you will be able to:

- Explain the types of estimation
- Explain the methods of point estimation
- Explain the method of moments
- Explain the methods of maximum likelihood
- List the limitations, applications and properties of these methods

Any statistical investigation aims at making generalisations from sample to population. Moreover, selecting a random sample is essential for drawing valid conclusions about the population. Methods of estimation develops theoretical basis of connection between sample information and population model. This in turn permits inference about the population.

Suppose in a specified population we are interested in the population average mu, then we might use the sample mean x bar to estimate the population mean. For example, we might wish to make a statement about mean cholesterol level of all men residing in India. It is then obvious to use a sample mean as an estimate of the population mean.

Similarly, the value of the sample proportion p is a point estimate of the population proportion P. Since a sample is only a part of the population, numerical value of the sample mean cannot be expected to give exact value of the parameter. Moreover, the value of the mean depends on the particular sample that happens to be selected.

For example, two different samples of same size from the same population will yield two different means. This is because a sample mean is a random variable. In general, if x one, x two, up to xn is a random sample, then any function of x one, x two, up to xn is a random variable.

True value of a parameter is an unknown constant that can be correctly ascertained only by an exhaustive study of the population, if indeed that were possible. Our objective may be to obtain a guess or an estimate of the unknown true value along with the determination of its accuracy. This type of inference is called estimation of parameters.

Estimation is the process by which sample data are used to indicate the value of an unknown quantity in a population.
Results of estimation can be expressed as a single value, known as a point estimate, or a range of values, known as an interval estimate.

Example
Suppose a manager of a bank wanted to know the average number of visits of customers to

the bank in the last year, he could calculate the average number of visits of the hundreds (or perhaps thousands) of customers who have transactions in the bank, that is, the population mean. Instead, he could use an estimate of this population mean by calculating the mean of a representative sample of customers. If this value was found to be sixty, then sixty would be his estimate.

2. Methods of Estimation

There are two methods of estimation, they are:

- **Point Estimation and**
- **Interval Estimation**

Point Estimation

An objective of point estimation is to compute a single value from the sample data that is likely to be close to the unknown value of the parameter. A statistic intended for estimating a parameter is called a point estimator or simply an estimator of the parameter.

A point estimate is often insufficient, because it is either right or wrong. If you are told only that, a point estimate of average number of students for a particular course is wrong, you do not know how wrong it is, and you cannot be certain of the estimates reliability. If you learn that, it is off by only ten students you can accept the closer figure of enrolment of students as a good estimate of future enrolment.

Without an assessment in accuracy, a single number quoted as an estimate may not serve a very useful purpose. We must indicate an estimate of variability in the distribution of an estimator. The standard error provides information about its variability.

Therefore, point estimate is much more useful if it is accompanied by an estimate of the error that might be involved.

Interval estimation

Assume that we have a sample (x one, x two, up to xn) from a given population. All parameters of the population are known except some parameter *theta*.

We want to determine from the given observations unknown parameter *theta*.

In other words, we want to determine a number or range of numbers from the observations that can be taken as a value of an unknown parameter *theta*.

Interval estimation is a process of defining two numbers, between which a population parameter is said to lie. For example, *a* **less than** x less than *b* is an interval estimate of the population mean mu. It indicates that the population mean is greater than *a* **but** less than **b**.

An interval estimate is a range of value used to estimate a population parameter.

Because of the presence of sampling error, sometimes it is more useful to start with the point estimate, and then establish a range of values both above and below the point estimate.

In practice, confidence interval estimates are used more commonly by far than point estimates. Nevertheless, since point estimates are used in certain important ways in statistics, and carry with them some important concepts and terms, we need to look at them briefly in this course.

3. Two Popular Methods of Point Estimation

Methods of Point Estimation:

There are several methods using which we can obtain the point estimates of the population parameters. Some of them are:

- Method of moments
- Method of maximum likelihood
- Method of least squares
- Method of minimum variance
- Method of Chi square etc

However, our study is concentrated only on two popular methods they are Method of moments and Method of maximum likelihood.

The method of moments
The method of moments is the oldest method of deriving point estimators. It usually produces some asymptotically unbiased estimators, although they may not be the best estimators.
The method of moments was discovered by Karl Pearson. The method is as follows.

Let f of (x, theta one, theta two, up to theta k) be the density function of the population under consideration. To estimate the unknown parameters theta one, theta two up to theta k if mu r dash denotes the rth moment (about zero) then by definition,
Mu r dash is equal to integral of x to the power r into f of (x, theta one, theta two, up to theta k) dx
Where, r is equal to one, two,.. etc.

That is,
Mu 1 dash is equal to integral of x into f of (x, theta one, theta two, up to theta k) dx
Mu 2 dash is equal to integral of x to the power 2 into f of (x, theta one, theta two, up to theta k) dx
Up to
Mu k dash is equal to integral of x to the power k into f of (x, theta one, theta two up to theta k) dx
Mu 1 dash, mu 2 dash, up to mu k dash are in general functions of the parameters theta 1, theta 2, up to theta k. Thus, the above is a set of k equations involving k unknown parameters theta 1, theta 2, up to theta k.

Now solving the equations, theta 1, theta 2, up to theta k can be written as the functions of Mu 1 dash, mu 2 dash, etc mu k dash. But in general, Mu 1 dash, mu 2 dash, up to mu k dash are unknown and hence their estimators by sample moments m 1 dash, m 2 dash, up to m k dash respectively where mr dash is equal to summation xi to the power r by n, x one, x two, up to xn being sample observations are determined.

In case of frequency distribution of the sample observations, r^{th} sample moment m r dash is given by mr dash is equal to summation fi into xi to the power r by N. Thus, the method of moments consists in equating rth raw moments about the origin in the population to the rth raw moments about the origin in the sample by giving values r is equal to one, two, etc. And obtaining various equations containing parameters and solving these equations to obtain the estimate of the parameters.

Properties:

- Moment estimators are consistent provided the population moments exists
- Moment estimators need not be unbiased
- Under certain conditions, moment estimators have an asymptotic normal distribution
- Moment estimators are less efficient than maximum likelihood estimators

Method of maximum likelihood

In statistics, Maximum-Likelihood Estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.

The method of maximum likelihood corresponds to many well-known estimation methods in statistics. For example, one may be interested in the heights of adult female giraffes, but we are unable to measure the height of every single giraffe in a population due to cost or time constraints.

Assuming that the heights are normally or (Gaussian) distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable.

The method of maximum likelihood provides a basis for many of the techniques. The reasons are:

- The method has a good intuitive foundation. The underlying concept is that, the best estimate of a parameter is giving the highest probability that the observed set of measurements will be obtained
- The least-squares method and various approaches for combining errors or calculating weighted averages, etc. can be derived or justified in terms of the maximum likelihood approach

4. Applications, Principles and Properties of MLE

Applications of the method of maximum likelihood
Maximum likelihood estimation is used for a wide range of statistical models, including:

- **Linear Models** and **generalized linear models**
- **Exploratory** and **confirmatory factor analysis**
- **Structural Equation Modelling**
- Many situations in the context of **hypothesis testing** and **confidence interval** formation
- **Discrete Choice** models

These uses arise across applications in widespread set of fields, including:

1. **Communication Systems**
2. **Psychometrics**
3. **Econometrics**
4. Time-delay of arrival (TDOA) in acoustic or electromagnetic detection
5. Data modelling in nuclear and particle physics
6. Magnetic resonance imaging
7. **Computational Phylogenetics**
8. Origin/destination and path-choice modelling in transport networks
9. Geographical satellite-image classification

On the other hand, MLE is not as widely recognized among modellers in psychology. However, it is a standard approach to parameter estimation and inference in statistics. MLE has many optimal properties in estimation:

- Sufficiency - complete information about the parameter of interest contained in its MLE estimator
- Consistency - true parameter value that generated the data recovered asymptotically, i.e. for data of sufficiently large samples

- Efficiency - lowest-possible variance of parameter estimates achieved asymptotically and
- Parameterization invariance - same MLE solution obtained independent of the parameterisation used

The principle of Maximum Likelihood Estimation (MLE) originally developed by R. A. Fisher in the nineteen twenties, states that the desired probability distribution is the one that makes the observed data "most likely," which means that one must seek the value of the parameter vector that maximizes the likelihood function.

Principles
Suppose there is a sample x one, x two, up to xn of n independent and identically distributed observations, coming from a distribution with an unknown pdf f of dot. It is however surmised that the function f belongs to a certain family of distributions {f of (dot, theta), where theta belongs to the parameter space}, called the parametric model, so that f is

equal to ƒ of (dot, theta).

The value theta is unknown and is referred to as the "*true value*" of the parameter. It is desirable to find an estimator theta cap, which would be as close to the true value theta as possible. Both the observed variables x_i and the parameter theta can be vectors.
To use the method of maximum likelihood, one first specifies the joint density function for all observations.

For an independently identically distributed sample, this joint density function is
F of (x one, x two, up to xn , theta) is equal to f of (x one, theta) into f of (x two, theta) into up to f of (x n, theta)
Now, we look at this function from a different perspective by considering the observed values x one, x two, up to xn to be fixed "parameters" of this function, whereas theta will be the function's variable and allowed to vary freely. This function will be called the likelihood:
L of (theta, x one, x two, up to xn) is equal to f (x one, x two, up to xn , theta) which is equal to product of f of (x i, theta), i runs from 1 to n

In practice, it is more convenient to work with the logarithm of the likelihood function, called the log-likelihood.
Natural log of L of (theta, x one, x two, up to xn) is equal to summation, i runs from 1 to n natural log of f of (xi, theta)
We will get the point estimator of the parameter theta by maximizing the above function.
Hence, the method of maximum likelihood estimates theta by finding a value of theta that maximizes L of (xi, theta) or natural log of L of (xi, theta). This method of estimation defines a maximum-likelihood estimator (MLE) of theta.

An MLE estimate is the same regardless of whether we maximize the likelihood or the log-likelihood function, since as the likelihood function increases, decreases log likelihood also increases, or decreases. Log is a transformation. We can maximize the function by making use of principle of differentiation if the function is differentiable. In case the differentiation technique fails to get the maximum value of the function, then we have to make use of basic principle of MLE.
A maximum likelihood estimator coincides with the most probable Bayesian estimator given a uniform prior distribution on the parameters.

Properties
A maximum-likelihood estimator is an extremum estimator obtained by maximizing, as a function of theta, the *objective function.*
Maximum-likelihood estimators have no optimum properties for finite samples, in the sense that (when evaluated on finite samples) other estimators have greater concentration around the true parameter-value.

However, like other estimation methods, maximum-likelihood estimation possesses a number of attractive limiting properties. As the sample-size increases to infinity, sequences of maximum-likelihood estimators have these properties:
Consistency: A subsequence of the sequence of MLEs converges in probability to the value being estimated
Asymptotic normality: as the sample size increases, the distribution of the MLE tends to the

Gaussian distribution

Efficiency : MLE's are most efficient among the class of all consistent estimators. That is among the class of consistent estimators MLE has the minimum variance

5. Regularity Conditions of MLE

Maximum-likelihood estimators can lack asymptotic normality and can be inconsistent if there is a failure of one (or more) of the below regularity conditions:

Estimate on boundary

Sometimes the maximum likelihood estimate lies on the boundary of the set of possible parameters, or (if the boundary is not, strictly speaking, allowed) the likelihood gets larger and larger as the parameter approaches the boundary. Standard asymptotic theory needs the assumption that the true parameter value lies away from the boundary. If we have enough data, the maximum likelihood estimate will keep away from the boundary too.

Data boundary parameter-dependent

For the theory to apply in a simple way, the set of data values that has positive probability (or positive probability density) should not depend on the unknown parameter.

A simple example where such parameter-dependence does hold is the case of estimating theta from a set of independent identically distributed when the common distribution is uniform on the range (zero , theta). For estimation purposes, the relevant range of theta is such that theta cannot be less than the largest observation. Because the interval (zero, theta) is not compact, there exists no maximum for the likelihood function.

Nuisance parameters

For maximum likelihood estimations, a model may have a number of nuisance parameters. For the asymptotic behaviour outlined to hold, the number of nuisance parameters should not increase with the number of observations (the sample size).

Increasing information

For the asymptotic to hold in cases where the assumption of independently identically distributed observations does not hold, a basic requirement is that the amount of information in the data increases indefinitely as the sample size increases. Such a requirement may not be met if either there is too much dependence in the data (for example, if new observations are essentially identical to existing observations), or if new independent observations are subject to an increasing observation error.

Here’s a summary of our learning in this session, where we understood:

- The various methods of estimation
- The methods of point estimation
- The method of moments
- The method of maximum likelihood
- The limitations, applications and properties of the methods