Methods of Estimation:

There are several methods of estimation, out of which we shall consider the following important methods.

(i) Method of maximum likelihood.
(ii) Method of moments

Method of maximum likelihood estimation

This is the most general method of estimation, which was initially formulated by C.F. Gauss and then was first introduced by Prof. R.A. Fisher. Before introducing the method we define likelihood function.

Likelihood Function:

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population with density function $f(x;\theta)$. The likelihood function of the sample values $x_1, x_2, \ldots, x_n$ usually denoted by $L(\theta)$ or L, is the joint density function given by

$$L = L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \ldots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

It gives the relative likelihood (or probability) that the random variables assume a particular set of values $x_1, x_2, \ldots, x_n$. For a given sample $x_1, x_2, \ldots, x_n$, $L(\theta)$ becomes a function of single variable θ, the parameter.

Scanned with CamScanner

The method (or principle) of maximum likelihood consists in finding an estimator of the parameter θ which maximizes the likelihood function $L(\theta)$ for variations in the parameter θ. Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ of the sample values which maximizes $L(\theta)$ for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ usually called Maximum Likelihood Estimator (MLE) of θ. Thus, $\hat{\theta}$ is the solution, if any, of

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L(\theta)}{\partial \theta^2} < 0$$

Since $L(\theta) > 0$, so is $\log L(\theta)$ which shows that $L(\theta)$ and $\log L(\theta)$ attain their extreme (maximum or minimum) at the same value of θ. Therefore, the first of the above two equations can be written as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial}{\partial \theta} \log L = 0, \quad \ldots (1) \text{ where } L = L(\theta).$$

a form which is more convenient from practical point of view. Equation (1) is usually called as the "Likelihood Equation"

If the maximum likelihood estimator (MLE) is not obtainable by the likelihood equation, the likelihood function is to be maximized by some other method.

Scanned with CamScanner

807
GK

Properties of Maximum Likelihood Estimators (MLE's)

We make the following assumptions:

i) The first and second order derivatives viz. $\frac{\partial}{\partial\theta}\log L(\theta)$ and $\frac{\partial^2}{\partial\theta^2}\log L(\theta)$ exist and are continuous functions of θ in a range R (including the true value θ_0 of the parameter) for almost all x. For every θ in R

$$\left|\frac{\partial}{\partial\theta}\log L(\theta)\right| < F_1(x) \text{ and } \left|\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right| < F_2(x),$$

where $F_1(x)$ and $F_2(x)$ are integrable functions over $(-\infty,\infty)$.

ii) The third order derivative $\frac{\partial^3}{\partial\theta^3}\log L(\theta)$ exist such that

$$\left|\frac{\partial^3}{\partial\theta^3}\log L(\theta)\right| < M(x),$$

where $E[M(x)] < K$, a positive quantity.

iii) For every θ in R,

$$E\left[-\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{-\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right\}.L(\theta)\,dx_1,\ldots,dx_n$$

is finite and non-zero.

iv) The range of integration is independent of θ. But if the range of integration depends on θ, then $f(x;\theta)$ vanishes at the extremes depending on θ.

This assumption is to make the differentiation under the integral sign valid.

Under the above assumptions MLE's possess a number of important properties, which are written below:

Scanned with CamScanner

Properties of MLE

1. (Cramer-Rao . Theorem).

With probability approaching unity as $n \to \infty$, the likelihood equation $\frac{\partial}{\partial \theta} \log L(\theta) = 0$, has a solution which converges in probability to the true value θ_0 of θ. In other words maximum likelihood estimators are consistent.

2. (Hazoor Bazar's Theorem)

Any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tending to unity as the sample size n tends to infinity.

3. (Assymptotic Normality of MLE's)

A consistent solution of the likelihood equation is assymptotically normally distributed about the true value θ_0. Thus $\hat{\theta}$ is assymptotically $N\left(\theta_0, \frac{1}{I(\theta_0)}\right)$ as $n \to \infty$.

Here, $V(\hat{\theta}) = \frac{1}{I(\theta)} = \frac{1}{E\left[-\frac{\partial^2}{\partial \theta^2} \log L(\theta)\right]}$.

$I(\theta)$ is known as the information on θ supplied by the sample $(x_1, x_2, \ldots, x_n)$.

4. If MLE exists, it is the most efficient in the class of such estimators.

5. If a sufficient estimator exists, it is a function of the Maximum Likelihood Estimator.

6. If for a given population with pdf $f(x;\theta)$ an MVB estimator T exists for θ, then the likelihood equation will have a solution equal to the estimator T.

7. Invariance Property: If T is the MLE of θ and $\psi(\theta)$ is any one-to-one function of θ, then $\psi(T)$ is the MLE of $\psi(\theta)$.

8. The MLE for a parametric function $\tau(\theta)$ may not be unique.

Scanned with CamScanner