UNIT – III

METHODS OF ESTIMATION.

In the previous two units we have been discussing the ideas of a good estimator. Now we see the methods of estimation. Commonly used methods are

i) Method of maximum likelihood estimation.

ii) Method of minimum variance.

iii) Method of moments.

iv) Method of least squares.

v) Method of minimum Chi-square.

vi) Method of inverse probability.

1. Method of Maximum likelihood estimation.

This method was introduced by C.F. Gauss and developed by Prof. R.A. Fisher.

Def: Let $x_1, x_2 \ldots x_n$ be a random sample of size n taken from a population with density function $f(x; \theta)$. Then the likelihood function of the sample values $x_1, x_2 \ldots x_n$, usually denoted by $L = L(\theta)$ is their joint density function, given by

$$L = f(x_1, \theta) \cdot f(x_2, \theta) \ldots f(x_n, \theta)$$

$$= \prod f(x_i; \theta)$$

L gives the relative likelihood 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 becomes a function of the variable θ, the parameter.

The principle of maximum likelihood consists in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2 \ldots \theta_k)$, say, which maximizes the likelihood function $L(\theta)$ for variations in parameter.

That is we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_k)$

so that $L(\hat{\theta}) > L(\theta) \quad \forall \; \theta \in \Theta$

i.e. $L(\hat{\theta}) = \sup L(\theta), \quad \forall \; \theta \in \Theta$

Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2 \ldots x_n)$ of the sample values which maximizes L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is called Maximum Likelihood Estimator (M.L.E).

Then $\hat{\theta}$ is the solution, if any of

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0. \qquad \text{——— (1)}$$

Since $L > 0$, and $\log L$ is a non-decreasing function of L; L and $\log L$ attain their extreme value (maxima or minima) at the same value of $\hat{\theta}$.

$\therefore$ (1) can be written as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0 \qquad \text{——— (2)}$$

(2) is more convenient from practical point of view.

Note: If θ is vector valued parameter, then $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_k)$, is given by the solution of simultaneous equations:

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2 \ldots \theta_k) = 0$$

$$i = 1, 2 \ldots k.$$

Regularity conditions:

1. The first and second order derivatives $\frac{\partial \log L}{\partial \theta}$ and $\frac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in a range R (including the true value θ_0 of the parameter) and for almost all x. For every θ in R, $\left| \frac{\partial \log L}{\partial \theta} \right| < F_1(x)$ and $\left| \frac{\partial^2 \log L}{\partial \theta^2} \right| < F_2(x)$ where $F_1(x)$ and $F_2(x)$ are integrable functions over $(-\infty, \infty)$.

2. The third order derivative $\frac{\partial^3 \log L}{\partial \theta^3}$ exists such that $\left| \frac{\partial^3 \log L}{\partial \theta^3} \right| < M(x)$ where $E[M(x)] < K$, a positive quantity.

3. For every θ in R

$$E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) = \int_{-\infty}^{\infty} \left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) L \, dx = I(\theta) \text{ is}$$

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.

1. MLE and Sufficiency.

If T is a sufficient statistic for the parameter θ, then the solution of likelihood equation will be the function of sufficient statistic. T.

Proof: Let T be sufficient for θ.

According to Neyman factorization theorem,

$$L(\theta) = g(t,\theta).\ h(x) \quad \text{—— (1)}$$

where $h(x)$ is independent of θ.

According ML principle, the ML estimator of $\theta = \hat{\theta}$ satisfying

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$$ is the solution of the likelihood equation

$$\frac{\partial \log L}{\partial \theta} = 0 \quad \text{—— (2)}$$

Taking log on both sides of (1)

$$\log L(\theta) = \log g(t,\theta) + \log h(x)$$

Differentiating partially w.r.t. θ, we have

$$\frac{\partial \log L}{\partial \theta} = \frac{\partial \log g(t,\theta)}{\partial \theta} + 0 \quad \text{—— (3)}$$

Comparing (2) & (3) we have

$$\frac{\partial \log g(t,\theta)}{\partial \theta} = 0$$

Since $g(t,\theta)$ depends on θ, through the statistic t, any solution of θ, will be a function of t alone.

Hence $\hat{\theta} = f(t)$.

2. MLE and MVB.

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

Proof: If T is an MVB for θ, then according to the necessary and sufficient condition for the existence of MVB is

$$\frac{\partial \log L}{\partial \theta} = A(\theta)\ [T - \theta]. \quad \text{—— (1)}$$

According to MLE principle, the ML estimator of θ is $\hat{\theta}$ which is the solution of the likelihood equation

$$\frac{\partial \log L}{\partial \theta} = 0 \quad \text{--- (2)}$$

Comparing (1) and (2), we have

$$A(\theta)\,[T-\theta] = 0,$$

Since $A(\theta) \neq 0$, we have

$$T - \theta = 0 \Rightarrow T = \hat{\theta} = MVB.$$

3. (Cramer - Rao Theorem). 'With probability approaching unity as $n \to \infty$, the likelihood equation $\frac{\partial \log L}{\partial \theta} = 0$ has a solution which converges in probability to the true value θ_0'.

In otherwords M.L.E's are consistent.

Proof : Consider the likelihood function $L(\theta)$ at $\theta = \hat{\theta}$.

ie $L(\hat{\theta}) \geq L(\theta)$

or

$$L(x/\hat{\theta}) \geq L(x/\theta) \quad \text{--- (1)}$$

We consider the case of n independent observations from a distribution $f(x/\theta)$ and for each n, we choose the ML estimator $\hat{\theta}$ so that if θ is any admissible value of the parameter, we have

$$\log L(x/\hat{\theta}) \geq \log L(x/\theta) \quad \text{--- (2)}$$

We denote the true value of θ by θ_0 and let E_0 represent the operation of taking expectations when the true value θ_0 holds.

Consider the random variable

$$\frac{L(x/\theta)}{L(x/\theta_0)}$$

In virtue of the fact that the geometric mean of a distribution cannot exceed its arithmetic mean, we have for all $\theta^* \neq \theta_0$,

$$E_0 \log\left[\frac{L(x/\theta^*)}{L(x/\theta_0)}\right] < \log E_0\left[\frac{L(x/\theta^*)}{L(x/\theta_0)}\right] \quad \text{--- (3)}$$

Now the expectation on the RHS of (3) is

$$\int \cdots \int \frac{L(x/\theta^*)}{L(x/\theta_0)}\, L(x/\theta_0)\,.\, dx_1 \,.\, dx_2 \ldots dx_n = 1$$

Then equation (3) becomes

$$E_0 \log\left[\frac{L(x/\theta^*)}{L(x/\theta_0)}\right] < 1$$

or inserting the factor $1/n$

$$E_0 \frac{1}{n}\left[\log L(x/\theta^*)\right] < E_0 \frac{1}{n}\left[\log L(x/\theta_0)\right]$$

provided that the expectation of RHS is exist.

Now, for any value of θ,

$\frac{1}{n}\log L(x/\theta) = \frac{1}{n}\sum \log f(x_i/\theta)$ is the mean of a set of n independent random variables with expectation

$$E_0 \log f(x/\theta) = E_0\left[\frac{1}{n}\log L(x/\theta)\right]$$

By applying SLLN, we have

$\frac{1}{n}\log L(x/\theta) \rightarrow$ with probability unity to its expectation as n increases.

Thus as $n \to \infty$, we have from equation (3), with probability unity,

$$\frac{1}{n}\log L(x/\theta^*) < \frac{1}{n}\log L(x/\theta_0) \quad \text{—(4)}$$

or

$$\lim_{n\to\infty} \text{Prob}\left[\log L(x/\theta^*) < \log L(x/\theta_0)\right] = 1, \quad \theta^* \neq \theta_0.$$

On the other hand, from equation (2) with $\theta = \theta_0$, we have

$$\log L(x/\hat{\theta}) \geq \log L(x/\theta_0) \quad \text{—(5)}$$

Equations (4) & (5) imply that as $n \to \infty$, $L(x/\hat{\theta})$ cannot take any value other than $L(x/\theta_0)$. This implies that if $L(x/\theta)$ is a one-one function of θ,

$$P\left[\lim_{n\to\infty} \hat{\theta} = \theta_0\right] = 1$$

Asymptotic normality of MLEs.

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

OR

If the first derivative of likelihood function w.r.t θ exists, (it) is an interval of θ including the true value of θ_0

if $E\left(\frac{\partial \log L}{\partial \theta}\right) = 0$ —(1) and $R^2(\theta) = -E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right] = E\left[\frac{\partial \log L}{\partial \theta}\right]^2$ —(2)

exists and non-zero for all θ in the interval. The MLE $\hat{\theta}$ is asymptotically normally distributed as $\left(\theta_0, \frac{1}{R^2(\theta_0)}\right)$

Proof: Using Taylor's theorem, we have

$$\left.\frac{\partial \log L}{\partial \theta}\right|_{\hat{\theta}} = \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta_0} + (\hat{\theta} - \theta_0) \left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\hat{\theta}} \quad \text{— (3)}$$

where θ^* is some value in between θ_0 and $\hat{\theta}$. 6

Let $$R^2(\theta) = E\left[\frac{\partial \log L}{\partial \theta}\right]^2 = -E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right] \quad \text{— (4)},$$

which is the amount of information about θ contained in an observation.

We know that the likelihood equation

$$\frac{\partial \log L}{\partial \theta} = 0 \quad \text{at} \quad \theta = \hat{\theta}.$$

Hence from (3), we have

$$\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta = \theta_0} + (\hat{\theta} - \theta_0) \left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*} = 0$$

ie $$(\hat{\theta} - \theta_0) = \frac{-\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta = \theta_0}}{\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*}}$$

This can also be written as

$$\sqrt{n}\,(\hat{\theta} - \theta_0)\, R(\theta_0) = \frac{-\frac{1}{n}\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta = \theta_0} \cdot \sqrt{n} \Big/ R(\theta_0)}{\frac{1}{n}\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta = \theta^*} \Big/ R^2(\theta_0)} \quad \text{— (5)}$$

By WLLN, as $n \to \infty$, we have

$\frac{1}{n}\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*}$ tends to its expectation.

ie $$\frac{1}{n}\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*} \longrightarrow E\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta^*}$$

Hence from (4), we have

$$\frac{1}{n}\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*} \longrightarrow -R^2(\theta_0).$$

$$(5) \Rightarrow \sqrt{n}\,(\hat{\theta} - \theta_0) \cdot R(\theta_0) = \frac{-\frac{1}{n}\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta = \theta_0} \sqrt{n} \Big/ R(\theta_0)}{-R^2(\theta_0) \big/ R^2(\theta_0)}.$$

ie $\sqrt{n}\,(\hat{\theta} - \theta_0) \cdot R(\theta_0) = \left. \frac{\partial \log L}{\partial \theta} \right|_{\theta=\theta_0} \Big/ R(\theta_0)$.

ie $\sqrt{n}\,(\hat{\theta} - \theta_0) \cdot R(\theta_0) = \dfrac{\frac{1}{n} \cdot \left. \frac{\partial \log L}{\partial \theta} \right|_{\theta=\theta_0}}{R(\theta_0/n)}$

From this we can say that

$$\sqrt{n}(\hat{\theta} - \theta_0)\, R(\theta_0) \sim N[0, 1]$$

ie $\sqrt{n}(\hat{\theta} - \theta_0) \sim N\left[0, \dfrac{1}{R^2(\theta_0)}\right]$

ie $\sqrt{n}(\hat{\theta} - \theta_0) \sim N\left[0, \dfrac{1}{E\, \frac{\partial \log f(x, \theta)}{\partial \theta}}\right]_{\theta=\theta_0}$

From this we know

$$\hat{\theta} \sim N\left[\theta_0, \frac{1}{\frac{1}{n} E\, \frac{\partial \log f(x, \theta)}{\partial \theta}}\right]$$

5. MLE and Efficiency

If the first derivative of likelihood function with respect to θ exists in an interval (H) including the true value θ_0, and if $E\left[\frac{\partial \log L}{\partial \theta}\right] = 0$ — (1)

$$R^2(\theta) = E\left(\frac{\partial \log L}{\partial \theta}\right)^2 \quad \text{— (2)}$$

exists and is non-zero, $\forall\ \theta$ in (H), the MLE $\hat{\theta}$ is asymptotically distributed with mean θ_0 and variance $\dfrac{1}{R^2(\theta_0)}$.

Proof: using Taylor's theorem

$$\left. \frac{\partial \log L}{\partial \theta} \right|_{\theta=\hat{\theta}} = \left. \frac{\partial \log L}{\partial \theta} \right|_{\theta_0} + (\hat{\theta} - \theta_0) \left. \frac{\partial^2 \log L}{\partial \theta^2} \right|_{\hat{\theta}} \quad \text{— (3)}$$

where $\theta^* \neq \theta_0$ and it lies between θ_0 and $\hat{\theta}$.

Under regularity conditions, $\hat{\theta}$ is the root of $\frac{\partial \log L}{\partial \theta}$, the likelihood equation, so that the LHS of (3) satisfies $\left. \frac{\partial \log L}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$

we may write ③ as

$$(\hat{\theta} - \theta_0) R(\theta_0) = \frac{\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta_0} \Big/ R(\theta_0)}{\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta^*} \Big/ - R^2(\theta_0)} \quad \text{— (4)}$$

In the denominator on the RHS of ④, we have θ^*, since $\hat{\theta}$ is a consistent for θ_0 and θ^* lies between them, we have θ^* which is also consistent for θ_0.

Hence by applying previous theorem and equation ②, we have

$$\lim_{n \to \infty} P\left\{ \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]_{\theta^*} = - R^2(\theta_0) \right\} = 1 \quad \text{— ⑤}$$

so that the denominator on RHS converges to unity.

The numerator on RHS of ④ is a ratio to $R(\theta)$ of the sum of independent identical variables

$$\frac{\partial \log f(x_i, \theta_0)}{\partial \theta}.$$

The sum has zero mean by equation ① and variance defined to be $R^2(\theta_0)$. Hence by using the CLT, the numerator is a asymptotic standard normal variate exists. The same is true on the RHS as a whole. Thus the LHS of ④ is asymptotically standard normal.

In other words, the M.L.E $\hat{\theta}$ is asymptotically normally distributed with mean θ_0 and variance $\frac{1}{R^2(\theta_0)}$. Moreover, this result which gives the MLE and asymptotic variance equal to MVE, implies that under these regularity conditions MLE is efficient.

Since the MVE can only attained in the presence of a sufficient statistic, we are also justified in studying the MLE is asymptotically sufficient.

6. MLE and unbiasedness

MLEs need not necessarily be unbiased. That is because of the fact that if $\bar{x}$ or $\hat{\theta}$ is an unbiased estimate of θ, we cannot say that $g(\hat{\theta})$ is an unbiased estimate of $g(\theta)$.

For example,

i) in normal distribution, s^2 is an UE for σ^2. But we cannot say that s is an unbiased estimate for σ.

ii) in rectangular distribution, $f(x) = \frac{1}{\theta}$, $0 \le x \le \theta$, the largest order statistic $X_{(n)}$ is the MLE for θ. But $X_{(n)}$ is not unbiased estimator for θ.

This implies that MLE's need not necessarily be unbiased.

7. Invariance Property.

If T is the MLE of θ, and $\delta = \varphi(\theta)$ is a one-one function, then $\varphi(T)$ is MLE of $\varphi(\theta)$.

Proof: Let the likelihood function be $L(\theta)$. Since $\delta = \varphi(\theta)$ is a one-one function, $\exists$ a inverse function $\ni$

$$\theta = \varphi^{-1}(\delta)$$

Then $L(\theta)$ can be written as

$$L(\theta) = L[\varphi^{-1}(\delta)]$$
$$= L_1(\delta)$$

If $x_1, x_2 \ldots x_n$ are observations, the MLE T is a function of the observations and satisfies the inequality

$$L(T) \geq L(\theta)$$

ie $$L[\varphi^{-1}(d)] \geq L[\varphi^{-1}(\delta)]$$

where $d = \varphi(T)$

$$\Rightarrow T = \varphi^{-1}(d)$$

ie $$L_1(d) \geq L_1(\delta)$$

$\Rightarrow d = \varphi(T)$ is the MLE of δ

2. Method of Moments.

Let $f(x; \theta_1, \theta_2 \dots \theta_k)$ be the density function of the parent population with t parameters $\theta_1, \theta_2 \dots \theta_t$. If μ_r' denotes the r^{th} moment about origin, then

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x; \theta_1, \theta_2 \dots \theta_k)\, dx \quad ; \; r = 1, 2 \dots k. \qquad -(1)$$

In general, $\mu_1', \mu_2' \dots \mu_k'$ will be function of the parameter $\theta_1, \theta_2 \dots \theta_k$.

Let x_i, $i = 1, 2 \dots n$ be a random sample of size n taken from the given population. The method of moments consists in solving the k-equations in (1) for $\theta_1, \theta_2 \dots \theta_k$ in terms of $\mu_1', \mu_2' \dots \mu_k'$ and then replacing these moments by the sample moments $m_1', m_2' \dots m_k'$ respectively.

That is we get

$$\hat{\theta}_i = \theta_i(\hat{\mu}_1', \hat{\mu}_2' \dots \hat{\mu}_k')$$
$$= \theta_i(m_1', m_2' \dots m_k'), \quad i = 1, 2 \dots k.$$

Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_k$ are the required estimators of $\theta_1, \theta_2 \dots \theta_k$ respectively.

Properties.

1. According to Pearson, the moment method estimators are not generally efficient. But in the case of normal population, the moment method estimators $\bar{x}$, s^2 are efficient for μ and s^2 and both are consistent. But s^2 is biased. By making a simple correction, we can get an unbiased estimator. If we multiply s^2 by $\frac{n}{n-1}$, we can get an unbiased estimator.

2. Let $(x_1, x_2 \dots x_n)$ be a random sample of size n from a population with p.d.f $f(x, \theta)$. Then $x_i, (i = 1, 2 \dots n)$ are i.i.d rvs $\Rightarrow$ x_i^r are i.i.d. Hence if $E(x^r)$ exists, then WLLN, we get

$$\frac{1}{n} \sum x_i^r \xrightarrow{P} E(x_i^r) \Rightarrow m_r' \xrightarrow{P} \mu_r'$$

Hence the sample moments are consistent estimators of the corresponding population moments.

3. It has been shown that under quite general conditions, the estimates obtained by the method of moments are asymptotically normal but not in general efficient.

4. Generally the method of moments yields less efficient estimators than those obtained from the principle of maximum likelihood.

4. Method of Minimum chi - square

The method of minimum chi - square makes the use of the Pearson's chi - square statistic. This method can be used in case of discrete distributions or for grouped data from a continuous distribution.

Let $f_1, f_2 \dots f_k$ be the observed frequencies in k groups or classes and the unknown probabilities that f_i elements belong to the i th groups or class be p_i $(i = 1, 2 \dots k)$. p's are the functions of the unknown parameters $\theta_1, \theta_2 \dots \theta_k$. Thus

$$p_i = p_i(\theta), \text{ where } \theta = (\theta_1, \theta_2 \dots \theta_n).$$

Suppose the total sample size is n. Therefore $\Sigma f_i = n$. The expected frequencies are $np_1(\theta), np_2(\theta) \dots np_k(\theta)$. We know, Pearson's chi - square statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{[f_i - np_i(\theta)]^2}{np_i(\theta)}$$

$$= \Sigma \frac{f_i^2}{np_i(\theta)} - n$$

Under the method of minimum chi - square one has to chose $(\theta_1, \theta_2 \dots \theta_n)$ which minimizes χ^2. This will be minimum when $np_i(\theta)$ is as close as possible to f_i. So, to obtain the estimates of θ_i's, partially differentiate χ^2 statistic w.r.t θ_i $(i = 1, 2 \dots n)$ successively and equate to zero. Also check that the standard derivatives are non-negative.

i.e $\dfrac{\partial \chi^2}{\partial \theta_i} = 0$ for $i = 1, 2 \dots m$

and $\dfrac{\partial^2 \chi^2}{\partial \theta_i^2} \geq 0$

$\dfrac{\partial \chi^2}{\partial \theta_i} = 0$ provides m simultaneous equations in m unknowns. Solving these m equations for m unknown parameters, one gets the estimated values of $\theta_1, \theta_2 \dots \theta_m$ respectively.

Properties of minimum chi-square estimators.

1. The minimum chi-square estimators are consistent.
2. The minimum χ^2 estimators are asymptotically normal
3. Minimum χ^2 estimators are efficient.
4. Minimum χ^2 estimators are not necessarily unbiased.

Uses.

Minimum χ^2 method of estimation is rarely used in practice. It is used only when it is difficult to solve the likelihood simultaneous equations obtained under maximum likelihood estimation method.

Ex...

Modified Chi-square Statistic

Let $(x_1, x_2 \dots x_n)$ be the k sample observations with observed frequencies $O_1, O_2 \dots O_k$ respectively. Assume that these observations are grouped into k classes. Let $(p_1, p_2 \dots p_k)$ be the k unknown probabilities for the k classes, which are functions of r unknown parameters $\theta = (\theta_1, \theta_2 \dots \theta_r)$. By definition, Then $p_i = p_i(\theta)$, for $i = 1, 2, 3 \dots k$ for the k classes the expected frequencies $e_1, e_2 \dots e_k$ where are respectively given by

$$e_i = np_i \quad \text{and} \quad n = \sum_{i=1}^{k} O_i$$

A measure of the discrepancy between the observed and expected frequencies is supplied by the statistic χ^2 given by

$$\chi^2 = \frac{(O_1 - e_1)^2}{e_1} + \frac{(O_2 - e_2)^2}{e_2} + \dots + \frac{(O_k - e_k)^2}{e_k}$$

$$= \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$$

In a similar manner, the modified χ^2-statistic is given by

$$(\chi')^2 = \sum_{i=1}^{k} \frac{(e_i - O_i)^2}{O_i}$$

Method of Modified Minimum Chi-square

The minimum chi-square method provides some computational difficulties for estimating the parameters since p_i's are occurring in the denominator of minimum χ^2 equation. In such cases, one can use the method of modified minimum chi-square.

By definition, the modified χ-statistic is given by

$$\chi'^2 = \sum_{i=1}^{k} \frac{(np_i - o_i)^2}{o_i}$$

$$= \sum_{i=1}^{k} \frac{(np_i)^2}{o_i} - n$$

Consider the likelihood function

$$L = \frac{n!}{\prod_{i=1}^{k} o_i!} \prod_{i=1}^{k} p_i^{o_i}$$

$$= \frac{n!}{\prod_{i=1}^{k} o_i!} \prod_{i=1}^{k} \left(\frac{np_i}{o_i}\right)^{o_i} \prod_{i=1}^{k} \left(\frac{o_i}{n}\right)^{o_i}$$

$\left(\because p_i^{o_i} = \left(\frac{np_i}{o_i} \cdot \frac{o_i}{n}\right)^{o_i}\right)$

Taking log on both sides, one gets

$$\log L = c + \sum_{i=1}^{k} o_i \cdot \log\left(\frac{np_i}{o_i}\right)$$

where c is independent of p_i's.

For large samples, assume that

$$np_i = o_i + c_i\sqrt{n}$$

where c_i's are small compared to o_i's and $\sum_{i=1}^{k} c_i = 0$

Hence, $$\log L = c + \sum_{i=1}^{k} o_i \log\left[1 + \frac{c_i n^{1/2}}{o_i}\right]$$

$$= c + \sum_{i=1}^{k} o_i \left[\frac{c_i n^{1/2}}{o_i} - \frac{c_i^2 n}{2o_i^2} + \frac{c_i^3 n^{3/2}}{3o_i^3} + \cdots\right]$$

$$= c + \sum_{i=1}^{k} c_i n^{1/2} - \sum_i \frac{nc_i^2}{2o_i} + \cdots$$

$$= c - \frac{1}{2}\sum_i \frac{(np_i - o_i)^2}{o_i} + O(n^{-1/2})$$

$$= c - \tfrac{1}{2}(\chi')^2 + O(n^{-1/2})$$

$\therefore$, if we neglect terms of order $O(n^{-1/2})$, then maximization of $\log L$ amounts to the minimization of χ'^2.

Examples.

1. In random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimators for

i) μ when σ^2 is known ii) σ^2 when μ is known and

iii) the simultaneous estimation of μ and σ^2.

Sol: $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad -\infty < x < \infty$$

$$L = \prod f(x_i; \mu, \sigma^2)$$

$$= \prod \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\Sigma(x-\mu)^2}$$

Taking log on both sides

$$\log L = -n/_2 \log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad —(1)$$

case i) when σ^2 is known, the likelihood equation for estimating μ is

Differentiating (1) w.r.t μ

$$\frac{\partial \log L}{\partial \mu} = -\frac{1}{\cancel{2}\sigma^2}\cancel{2}\,\Sigma(x_i-\mu)(-1)$$

$$\frac{\partial \log L}{\partial \mu} = 0 \Rightarrow \frac{\Sigma(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \Sigma x_i = n\mu$$

$$\Rightarrow \hat{\mu} = \bar{x}$$

Hence M.L.E for μ is the sample mean $\bar{x}$.

case ii) when μ is known, the likelihood equation for estimating σ^2 is

Differentiating (1) w.r.t. σ^2

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \Rightarrow -\frac{n}{2}\times\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\Sigma(x_i-\mu)^2 = 0$$

$$\Rightarrow \frac{1}{\cancel{2}\sigma^{4 2}}\Sigma(x_i-\mu)^2 = \frac{n}{\cancel{2}\sigma^2} \Rightarrow \frac{\Sigma(x_i-\mu)^2}{\sigma^2} = n$$

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

case iii) The likelihood equations for simultaneous estimation of μ and σ^2 are

$$\frac{\partial \log L}{\partial \mu} = 0 \text{ and } \frac{\partial \log L}{\partial \sigma^2} = 0$$

$$\Rightarrow \hat{\mu} = \bar{x} \text{ and}$$

$$\hat{\sigma}^2 = \frac{\Sigma(x_i-\hat{\mu})^2}{n} = \frac{1}{n}\Sigma(x_i-\bar{x})^2.$$

2. Prove that the maximum likelihood estimate of the parameter α of a population having density function

$$f(x) = \frac{2}{\alpha^2}(\alpha - x); \quad 0 < x < \alpha$$

for a sample of unit size is $2x$, x being the sample value. Show also that the estimate is biased.

Sol: For a random sample of unit size $(n=1)$, the likelihood function is

$$L(\alpha) = f(x, \alpha) = \frac{2}{\alpha^2}(\alpha - x), \quad 0 < x < \alpha.$$

The likelihood equation gives

$$\frac{\partial \log L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[\log 2 - 2\log \alpha + \log(\alpha - x)\right] = 0$$

$$\Rightarrow \frac{-2}{\alpha} + \frac{1}{\alpha - x} = 0 \Rightarrow 2(\alpha - x) - \alpha = 0$$

$$\Rightarrow \alpha = 2x$$

Hence MLE of α is given by $\hat{\alpha} = 2x$.

$$E(\hat{\alpha}) = E(2x) = 2\int_0^\alpha x\, f(x, \alpha)\, dx$$

$$= \frac{4}{\alpha^2}\int_0^\alpha x(\alpha - x)\, dx$$

$$= \frac{4}{\alpha^2}\left[\frac{\alpha x^2}{2} - \frac{x^3}{3}\right] = \frac{2}{3}\alpha$$

Since $E(\hat{\alpha}) \neq \alpha$, $\hat{\alpha} = 2x$ is not an unbiased estimate of α.

3. Find the maximum likelihood estimate for the parameter λ of Poisson distribution.

Sol: The p.m.f of Poisson is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2 \ldots$$

Likelihood function of a random sample $x_1, x_2 \ldots x_n$ of n observations from the population is

$$L = \prod p(x_i) = \prod \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-n\lambda} \cdot \lambda^{\Sigma x_i}}{\prod x_i!}$$

$$\log L = -n\lambda + \Sigma x_i \log \lambda - c$$

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{\Sigma x_i}{\lambda} - 0$$

$$-n + \frac{\Sigma x_i}{\lambda} = 0$$

$$\Rightarrow \frac{n\bar{x}}{\lambda} = n$$

$$\Rightarrow \hat{\lambda} = \bar{x}$$

3. Estimate α and β in the case of Pearson's Type III distribution by the method of moments.

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x}, \quad 0 \le x \le \infty$$

Sol: we have

$$\mu_r' = \int x^r f(x)\, dx$$

$$= \frac{\beta^\alpha}{\Gamma\alpha} \int x^r x^{\alpha-1} e^{-\beta x} dx$$

$$= \frac{\beta^\alpha}{\Gamma\alpha} \frac{\Gamma\alpha+r}{\beta^{\alpha+r}} = \frac{\Gamma\alpha+r}{\Gamma\alpha\, \beta^r}$$

$$\therefore \mu_1' = \frac{\Gamma\alpha+1}{\Gamma\alpha\, \beta^1} = \frac{\alpha}{\beta}$$

$$\mu_2' = \frac{\Gamma\alpha+2}{\Gamma\alpha\, \beta^2} = \frac{(\alpha+1)\alpha}{\beta^2}$$

$$\frac{\mu_2'}{\mu_1'^2} = \frac{\alpha+1}{\alpha} = \frac{1}{\alpha} + 1$$

$$\Rightarrow \alpha = \frac{\mu_1'^2}{\mu_2' - \mu_1'^2}$$

$$\beta = \frac{\alpha}{\mu_1'} = \frac{\mu_1'}{\mu_2' - \mu_1'^2}$$

—— o ——