BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Statistical Inference-I

Course Code: (23ST06CC)

Unit-III

Point Estimation

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. N. Saranya

Guest Faculty

Department of Statistics

Unit-III

METHOD OF POINT ESTIMATION

* Method of moments
* Method of maximum likelihood estimation
* Method of Bayesian

Definition of Point Estimation:

* Method of minimum variance
* Method of least square
* Method of minimum chi-square
* Method of inverse probability

* Method of moments:

This method was discovered and studied in detailed by Karl Pearson. Let $f(x: \theta_1, \theta_2 \dots \theta_k)$ be the density function of the parent population with k parameters, $\theta_1, \theta_2, \dots \theta_k$. If μ_r' denotes the r^{th} moment about origin then

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

In general $\mu_1', \mu_2' \dots \mu_k'$ will be function of parameters $\theta_1, \theta_2 \dots \theta_k$. Let x_i, $i=1,2\dots n$ be the random sample of size n from the given population. Then $\theta_1, \theta_2, \dots \theta_k$ in terms of $\mu_1', \mu_2' \dots \mu_k'$ and then replacing these moments

$\mu_r' : r=1,2\dots k$ by the sample moments

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

where m_i' is the i^{th} 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:

Property 1:

Moment estimators are asymptotically unbiased. $N \to \infty$ as unbiased.

Property 2:

They are consistent estimators.

Property 3:

The general conditions for distribution are asymptotically normal.

Method of k^{th} Sample moment:

Let $(x_1, x_2 \ldots x_n)$ be a random sample the k^{th} sample moments

Raw: $M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$

Central: $M_k = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_i)^k$

Methods of k^{th} distribution moments:

Let $(x_1, x_2 \ldots x_n)$ be a random sample the k^{th} distribution moments.

Raw: $M_k = E(X^k)$

$$m_k = \begin{cases} \sum_i x_i^k p_i \,; & \text{if random variable is discrete} \\ \int_{-\infty}^{\infty} x^k f(x)\,dx \,; & \text{if random variable is continuous} \end{cases}$$

Central:

$$M_k = E\left[(X - E(X))^k\right]$$

$$M_k = \begin{cases} \sum_i (x_i - E(x))^k \cdot p_i \,; & \text{discrete} \\ \int_{-\infty}^{\infty} (x - E(x))^k f(x)\,dx \,; & \text{continuous.} \end{cases}$$

In the method of moments, we assume that a sample moment is equal to the moment from a distribution.

We uses many moments as unknown parameter, by solving the equation system we obtain estimator of these parameters.

Example 1:

To find the moment estimator of Bernoulli Population with parameter P

Soln:

The density function of Bernoulli distribution

$$P(X=x) = pq^{1-x} \,; \quad x = 0 \text{ (or) } 1 \,; \quad 0 \le p \le 1, \; p+q=1$$
$$= 0 \,; \quad \text{otherwise}$$

Raw moment of Bernoulli distribution

$\mu_1' = p$

Same moment $= m_1' = \sum_{i=1}^{1} x_i/n = \bar{x}$

The moment estimator

Example: 2

To find the moment estimator of Poisson Population with parameter λ.

Soln:

$$P(X=x) = P(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & ; x = 1, 2, \ldots \quad \lambda > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

Since $\mu_1' = \hat{\lambda}$

$$m_1' = \sum_{i=1}^{n} x_i/n = \bar{x}$$

$$\mu_1' = m_1' \Rightarrow \bar{x} = \hat{\lambda}.$$

Example 3:

To find the moment estimator of Exponential Distribution with parameter θ.

Soln:

$$f(x) = \theta e^{-\theta x} ; \theta > 0; x = 0, 1, \ldots$$

Since $\mu_1' = \theta$

$$m_1' = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$\mu_1' = m_1' \Rightarrow \bar{x} = \theta$$

Example 4:

To find the moment estimator of normal distribution with parameter μ and σ^2.

Soln:

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & ; -\infty < x < \infty ; \mu < \infty ; \sigma > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

Since,

$$\mu_1' = \mu$$

$$m_1' = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$\mu_2' = \mu^2 + \sigma^2$$

$$m_2' = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

$$\mu_1' = m_1' \Rightarrow \hat{\mu} = \bar{x}$$

$$\mu_2' = m_2' \Rightarrow \frac{\mu^2 + \sigma^2}{n} = \frac{\sum x_i^2}{n}$$

$$\sigma^2 = \frac{\sum x_i^2}{n} - \mu^2 \Rightarrow \sigma^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

Maximum likelihood Estimation:

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

$$L = f(x_1, \theta)\, f(x_2, \theta), \dots f_n(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2 \dots x_n$ for a given sample $x_1, x_2 \dots x_n$. L becomes a function of the variable, θ is the parameter.

The principles of maximum likelihood consists in binding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2 \dots \theta_k)$ say which maximize the likelihood function $L(\theta)$ for variations in parameter ie) we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots \hat{\theta}_k)$ so that

$$L(\hat{\theta}) > L(\theta) \ \forall\, \theta \in \Theta \quad \text{ie)} \quad L(\hat{\theta}) = \sup L(\theta) \ \forall\, \theta \in \Theta$$

Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots x_n)$ of a sample value which maximizes L for variation in θ, then $\hat{\theta}_1$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called maximum likelihood estimator.

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

Since, $L > 0$ and $\log L$ is an nondecreasing function of L then L and $\log L$ attain their extreme value of the $\hat{\theta}$.

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0.$$

from which is much more convenient from practical point of view.

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

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \dots \theta_k) = 0, \quad i = 1, 2 \dots k$$

The above equations are usually referred to as the likelihood equations for estimating the parameters.

Properties of MLE:

i) The first and second order derivatives is $\frac{\partial \log L}{\partial \theta}$ and $\frac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in Range R.

ii) The third order derivatives $\frac{\partial^3}{\partial \theta^3} \log L$ exists such that $\left| \frac{\partial^3}{\partial \theta^3} \log L \right| < M(x)$

where,

$E[M(x)] < k$, a positive quantity

iii) ~~For every~~ θ in R

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

is finite and non-zero.

iv) The range of integration is independent θ. But if the range of integration depends on θ, the $f(x, \theta)$ depending on also θ.

Example:

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

i) μ when σ^2 is known.

ii) σ^2 when μ is known and

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

Solution:

The density function of normal distribution is $f(x : \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{(-\frac{1}{2}\sigma^2)(x-\mu)^2}$ $\quad -\infty < x, \mu < \infty, \; \sigma > 0.$

and the likelihood function is,

$$L = \prod_{i=1}^{n} f(x : \mu, \sigma^2)$$

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

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

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\left(-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

$$\log L = \log\left\{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\left(-1/2\sigma^2 \sum_{i=1}^{n}(x_i-\mu)^2\right)}\right.$$

$$= \log\left\{\left(\frac{1}{\sigma^2(2\pi)}\right)^{n/2}\right\} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= n/2\left\{\log(1) - \log\left(\sigma^2(2\pi)\right)\right\} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

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

Case i) :-

When σ^2 is known, the likelihood eqn for estimating μ is,

$$\frac{\partial \log L}{\partial \mu} = \frac{\partial}{\partial \mu}\left[-n/2\log(2\pi) - n/2\log(\sigma^2) - 1/2\sigma^2\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\frac{\partial \log L}{\partial \mu} = 0 \Rightarrow \frac{-1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) = 0$$

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

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

$$\frac{1}{n}\sum_{i=1}^{n}x_i - \hat{\mu} = 0$$

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

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

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,

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-n/2\log(2\pi) - n/2\log(\sigma^2) - 1/2\sigma^2\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \Rightarrow -\frac{n(2\sigma)}{2\sigma^2} + \frac{2}{2\sigma^3}\sum_{i=1}^{n}(x_i-\mu) = 0$$

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

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

Hence, M.L.E for σ^2 is the $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$

Case iii):

The likelihood equation for simultaneous estimation of μ and σ^2 are,

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

Thus, from 1 and 2 we get,

1 $\Rightarrow \hat{\mu} = \bar{x}$ and

2 $\Rightarrow \hat{\sigma}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$

Sub 1 in 2,

$$\hat{\sigma}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2$$

$\hat{\sigma}_2 = s^2$ (which is the sample variance).

Method of least squares:

For fitting a curve of the form,

$$y = f(x; b_0, b_1, \dots b_n) \dots \to ①$$

where $b_0, b_1 \dots b_n$ are unknown parameter to the observed sample observation (X_1, Y_1), $(X_2, Y_2), \dots (X_n, Y_n)$ by the principle of least squares we have to minimize

$$E = \sum_{i}^{n}\{y_i - f(x_i; b_0, b_1, b_3 \dots b_n)\}^2 \dots \to ②$$

with respect to the parameters $b_0, b_1 \dots b_n$

This is the same as to minimize. The Sum of squares of the distance of the observed points from the curve measured in the direction of y-axis.

* In case equation (1) is the regression equation of Y on $X, X_1, X_2 \dots X_n$ may be taken as observed values of the independent variable and

$e_i = y_i - f(x_i; b_0, b_1, \ldots b_n)$ are the residuals or errors.

If we assume that the errors are independently normally distributed with zero means and constant variance σ_e^2, then the joint probability density of the error or the likelihood function is given by,

$$L = \text{const.} \exp\left[-\frac{1}{2\sigma^2} \sum_i \{y_i - f(x_i; b_0, b_1, \ldots b_n)\}\right]$$

Hence the maximizing L amounts to minimizing

$$\sum_{i=1}^{n} \{y_i - f(x_i; b_0, b_1, \ldots b_n)\}^2$$

In case e_i's are independently normally distributed with zero means and variance $\sigma_{e_i}^2$, maximizing L will amount to minimizing

$$\sum_{i=1}^{n} \frac{1}{\sigma_{e_i}^2} \{y_i - f(x_i; b_0, b_1, \ldots b_n)\}^2$$

which is the sum of squares of residuals each weighted by the inverse of its variance.

* This may be called the weighted least squares method. In general we may consider the regression of Y on $X_1, X_2, \ldots X_p$ and the method of least square appropriate for this may be similarly reduced.

* The least-squares estimators do not have any optimum properties even asymptotically.

* However, in linear estimation, this method provides good estimators in small samples.

* When we are estimating $f(x_i; b_0, b_1, \ldots b_n)$ as linear function of parameters $b_0, b_1, \ldots b_n$.

* The x_i's being known, given values the least squares estimators obtained as linear functions of the Y's will be minimum variance unbiased estimator.

Method of Least Squares:

A method for obtaining estimators of the regression parameters α and β. It chooses an estimators those values that make the sum of the squares of the differences between the observed and the predicted responses as small as possible.

The principle of least squares is used to fit a curve of form:

$$y = f(x, a_0, a_1 \dots a_n)$$

where α_i's are unknown parameters, to the set of n sample observations (x_i, y_i); $i = 1, 2 \dots n$ from a bivariate population. It consists in minimising the sum of squares of residuals. viz.,

$$E = \sum_{i=1}^{n} \{y_i - f(x_i, a_0, a_1, \dots a_n)\}^2$$ subject to variations in $a_0, a_1, \dots a_n$.

The normal equations for estimating $a_0, a_1, \dots a_n$ are given by

$$\frac{\partial E}{\partial \alpha_i} = 0, \quad i = 1, 2 \dots n$$

Some assumptions of the least square method are

* Linear relationship between the variance.
* Observations are independent of each other.
* Variance of residual is constant with a mean of 0.
* Errors are distributed normally.

Limitations of LSM:

The least squares method assumes that the data evenly distributed and doesn't contain any outliers for deriving a line of best fit. But, the method doesn't provide accurate results for unevenly distributed data or for data containing outliers.

Methods 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 the groups or classes and the unknown probabilities that f_i elements belong to the ith group or class be P_i $(i=1,2\dots k)$ p's are the functions of unknown parameters $\theta_1, \theta_2 \dots \theta_k$.

Then $P_i = P_i(\theta)$ 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

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

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

under the method of minimum chi square one has to choose $(\theta_1, \theta_2 \dots \theta_n)$ which minimize 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 differentive χ^2 statistic w.r.to θ_i $(i=1,2\dots n)$ successively and equate to zero. Also check that the standard deviations are non negative.

(ie) $\frac{\partial \chi^2}{\partial \theta_i} = 0$ for $i=1,2\dots m$ and $\frac{\partial^2 \chi^2}{\partial \theta_i^2} \geq 0$

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

Properties of minimum chi-square estimators:

* The minimum chi square estimators are consistent.
* The minimum chi square estimators are asymptotically normal.
* Minimum chi square estimator are efficient.
* Minimum χ^2 estimator are not necessarily unbiased.

Uses:

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

Modified x^2 statistic:

Let $x_1, x_2 \ldots x_n$ be the k^{th} sample observations with observed frequencies $O_1, O_2 \ldots O_k$ respectively. Assume that these observations are grouped into k classes.

Let $(P_1, P_2 \ldots P_k)$ be the k^{th} unknown probabilities for the k classes which are functions of r^{th} unknown parameters, $\theta = (\theta_1, \theta_2 \ldots \theta_r)$. Then $P_i = P_i(\theta)$ for $(i = 1, 2 \ldots k)$.

By definitions the expected frequencies for the k classes are respectively given by $e_1, e_2 \ldots e_k$ where $e_i = nP_i$ and $n = \sum_{i=1}^{k} O_i$

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

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

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

in similar manner, the modified chi-square statistics is given by,

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

Method of modified minimum chi square:

The minimum chi square method provide some computational difficulties for estimating the parameters since P_i's are occuring in the denominator of minimum x^2 equation.

In such cases one can use the method of modified minimum x^2

By definition, the modified χ^2 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

$$L = \frac{n!}{\prod_{i=1}^{k} O_i!} \prod_{i=1}^{k} P_i^{O_i}$$

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

$$\therefore P_i^{O_i} = \left(\frac{nP_i}{O_i} \cdot \frac{O_i}{n}\right)^{O_i}$$

Taking log on both sides we get,

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

where c is independent of P_i's.

For large sample. 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 $$\log L = c + \sum_{i=1}^{k} O_i \log\left[1 + \frac{c_i n^{1/2}}{O_i}\right]$$

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

$$= c + \sum_{i=1}^{k} c_i n^{1/2} - \sum_i n \frac{c_i^2}{2O_i} + \dots$$

$$= c - \frac{1}{2} \sum_i \frac{(nP_i - O_i)^2}{O_i} + o(n^{-1/2})$$

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

If we neglect the terms of orders $o(n^{-1/2})$, The maximum of $\log L$ amounts to the minimization of χ'^2.

Confidence Interval & Confidence Limits:

Let x_i $(i = 1, 2 \dots n)$ be a random sample of n observations from a population involving a single unknown parameter θ, (say) Let $f(x, \theta)$ be the probability function of the parent distribution from which the sample is drawn and let us suppose that this distribution is continuous.

Let $t = t(x_1, x_2 \ldots x_n)$ a function of the sample values be an estimate of the population parameter θ, with the sampling distribution given by $g(t, \theta)$.

Having obtained the value of the statistic t from a given sample, we can make some reasonable probability statements about the unknown parameter θ in the population from which the sample has been drawn by the technique of confidence interval due to Neyman and is obtained below.

for all small values of α (5% or 1%) and then determine two constants say C_1 & C_2 say that

$$P(C_1 < \theta < C_2 \mid t) = 1 - \alpha \rightarrow ①$$

[The quantities C_1 and C_2 so determined are known as confidence limits or fiducial limits and the interval $[C_1, C_2]$ within which the unknown value of the population parameter is expected to lie, is called the confidence ~~coefficient~~ interval and $(1-\alpha)$ is called confidence coefficient.]

Thus if we take $\alpha = 0.05$ [or 0.01) we shall get 95% (or 99%) confidence limits.

To find C_1 and C_2:

* Let T_1 and T_2 be two statistic such that

$$P(T_1 > \theta) = \alpha_1 \text{ and } \rightarrow ②$$

$$P(T_2 \leq \theta) = \alpha_2 \rightarrow ③$$

where α_1 and α_2 are constants independent of θ. Eqn ② & ③ can be contained to give

$$P(T_1 < \theta < T_2) = 1 - \alpha.$$

where $\alpha = \alpha_1 + \alpha_2$

Statistics T_1 and T_2 defined in ② & ③ may be taken as C_1 & C_2 defined in ①.

Example:

(4) obtain $100(1-\alpha)\%$ confidence intervals for the parameters (a) θ and (b) σ^2 of the normal distribution. $f(x;\theta,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\}$, $-\infty < x < \infty$.

Soln:

Let $x_i, (i=1,2\ldots n)$ be a random sample of size n from the density $f(x;\theta,\sigma)$ and

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

a) The statistic $t = \frac{\bar{x}-\theta}{S/\sqrt{n}}$ follows Student's t-distribution with $(n-1)$ degrees of freedom. Hence $100(1-\alpha)\%$ Confidence limits for θ are given by

$$P(|t| \le t_\alpha) = 1-\alpha$$

$$P\left(\left|\frac{\bar{x}-\theta}{S/\sqrt{n}}\right| \le t_\alpha\right) = 1-\alpha$$

$$P\left[|\bar{x}-\theta| \le S/\sqrt{n}\, t_\alpha\right] = 1-\alpha$$

$$P\left[\bar{x}-t_\alpha S/\sqrt{n} \le \theta \le \bar{x}+t_\alpha S/\sqrt{n}\right] = 1-\alpha \rightarrow (1)$$

where t_α is the tabulated value t for $(n-1)$ degrees of freedom at significance level "α".

Hence the required confidence interval for θ is

$$(\bar{x}-t_\alpha S/\sqrt{n},\ \bar{x}+t_\alpha S/\sqrt{n})$$

b) Case i) θ is known and equal to μ.

Then $\frac{\sum(x_i-\mu)^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n)} \rightarrow (2)$

If we define χ^2_α as the value of χ^2 such that

$$P(\chi^2 > \chi^2_\alpha) = \int_{\chi^2_\alpha}^{\infty} P(\chi^2)\, d\chi^2 = \alpha \rightarrow (3)$$

where $P(\chi^2)$ is the p.d.f. of χ^2- distribution with n.d.f then the required confidence interval is given by:

$$P\{\chi^2_{1-(\alpha/2)} \le \chi^2 \le \chi^2_{\alpha/2}\} = 1-\alpha$$

$$P\{\chi^2_{1-\alpha/2} \le \frac{ns^2}{\sigma^2} \le \chi^2_{\alpha/2}\} = 1-\alpha \quad [\text{from } ②] \rightarrow ④$$

Now

$$\frac{ns^2}{\sigma^2} \le \chi^2_{\alpha/2} \Rightarrow \frac{ns^2}{\chi^2_{\alpha/2}} \le \sigma^2 \text{ and}$$

$$\chi^2_{1-\alpha/2} \le \frac{ns^2}{\sigma^2} \Rightarrow \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2}}$$

Hence ④ eqn gives.

$$P\left\{\frac{ns^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2}}\right\} = 1-\alpha \rightarrow ⑤$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{(1-\alpha/2)}$ are obtained from ③ by using n d.f.

Thus for 95% Confidence interval for σ^2 is

$$P\left(\frac{ns^2}{\chi^2_{0.05}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{0.975}}\right) = 0.95$$

Case ii) : θ is unknown

In this case the statistics

$$\frac{\Sigma(x_i-\bar{x})^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Here also confidence interval for σ^2 is given by (4) where now χ^2_α is the significant value of χ^2 for $(n-1)$ d.f at the significant level 'α'.

Asymptotic properties for Maximum Likelihood method:

By asymptotic properties, we mean Properties that are true when the sample size becomes large. Let $X_1, X_2 \dots X_n$ be a random sample from a distribution with a Parameter θ.

We will prove the MLE satisfies the following two properties called consistency and asymptotic normality.

1) Consistency, we say that as estimate θ^n is consistent if $\theta^n \to \theta_0$ in probability as $n \to \infty$ where θ_0 is the true unknown parameter of the distribution of the sample.

2) Asymptotic Normality:

We say that $\hat{\theta}$ is asymptotically normal if, $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \sigma_{\theta_0}^2)$

where, $\sigma_{\theta_0}^2$ is called the asymptotic variance of the estimate $\hat{\theta}$. Asymptotic normality says that the estimator not only converges to the unknown parameter but it converges fast enough at a rate $1/\sqrt{n}$.

Asymptotic Normality of MLE Fisher's Information:

We want to show the asymptotic normality of MLE i.e., to show that

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \sigma_{MLE}^2) \text{ for some } \sigma_{MLE}^2$$

fishers information of a random variable X with distribution P_{θ_0} from the family $\{P_\theta : \theta \in \Theta\}$ is defined by

$$I(\theta) = E_{\theta_0}[l'(x/\theta_0)]^2$$

$$= E_{\theta_0}\left[\frac{\partial}{\partial\theta} |\log f(x/\theta)|_{\theta=\theta_0}\right]^2$$

Consistency Asymptotic Normality and Efficiency:

Many of the proof will be rigorous to display more generally useful techniques also for later chapters.

We suppose that $X_n = (X_1, X_2, \ldots X_n)$ where the X_i's are i.i.d with common density.

$$P\{x; \theta_0\} \in T = \{P(x, \theta : \theta \in \Theta\}$$

We assume that θ_0 is identified in the sense that If $\theta \neq \theta_0$ and $\theta \in \Theta$, then $P(x;\theta) \neq P(x;\theta_0)$ with respect to the dominating measure μ.

Consistent, $\hat{\theta}(X_n) \xrightarrow{P} \theta_0$

Asymptotically Normal $\sqrt{n}(\hat{\theta}(X_n) - \theta_0) \xrightarrow{D(\theta_0)}$ Normal random variable.

Asymptotically efficient ($\mathring{\theta}$) if we want to estimate θ_0 by any other estimator within a "reasonable class" the MLE is the most precise.

LEMMA 1:

If θ_0 is identified and $E_{\theta_0}[|\log P(X;\theta)|] < \infty$ for all $\theta \in \Theta$, $Q_0(\theta)$ is uniquely maximized at $\theta = \theta_0$

Proof:

We know that

$E[g(Y)] > g[E(Y)]$, Take $g(y) = -\log(y)$ for $\theta \neq \theta_0$

$$E_{\theta_0}\left[-\log\left(\frac{P(X;\theta)}{P(X;\theta_0)}\right)\right] > -\log\left[E_{\theta_0}\left(\frac{P(X;\theta)}{P(X;\theta_0)}\right)\right]$$

Note that,

$$E_{\theta_0}\left[\frac{P(X;\theta)}{P(X;\theta_0)}\right] = \int \frac{P(x;\theta)}{P(x;\theta_0)} P(x;\theta_0).\, d\mu(x)$$

$$= \int P(x;\theta) = 1.$$

So,

$$E_{\theta_0}\left[-\log\left(\frac{P(x;\theta)}{P(x;\theta_0)}\right)\right] > 0$$

(or)

$$Q_0(\theta_0) = E_{\theta_0}[\log P(x;\theta_0)] > E_{\theta_0}[\log P(x;\theta)] = Q_0(\theta)$$

This inequality holds for all $\theta \neq \theta_0$.

ASYMPTOTICS OF MLE IN GENERAL CASE:

In the general case we give a heuristic argument for the fact that under suitable regularity conditions, it holds that

$$\hat{\theta}_n \overset{a}{\sim} N_d\{\theta, i(\theta)^{-1}/n\},$$

where,

$i(\theta)$ is the Fisher information for an individual observation.

As usual we $l(\theta)$ be the log-likelihood functions and $S(\theta)$ the score statistic. We also obtain,

$$J'(\theta) = -l''(\theta) = -S'(\theta)$$

We have that,

$$S(\theta) = \sum_i S(x_i;\theta)$$

$$J'(\theta) = \sum_i J'(x_i;\theta)$$

and

$$V\{S(\theta)\} = E\{J'(\theta)\} = n\,i(\theta)$$

Now, Use Taylor's formula to write,

$$0 = S(\hat{\theta}) = S(\theta) - J'(\theta)(\hat{\theta} - \theta) + R(x, \theta, \hat{\theta})$$

Solve this equation to yield

$$\sqrt{n}\,(\hat{\theta} - \theta) = \frac{S(\theta)/\sqrt{n}}{J'(\theta)/n}$$

Convergence in Probability and in Distribution for MLE:

A sequence of random variable $Y_1, Y_2 \ldots$ is said to converge in probability to a random variables Y if for all $\varepsilon > 0$.

$$\lim_{n \to \infty} P\{|Y_n - Y| > \varepsilon\} = 0$$

and then we write.

$$Y_n = Y \text{ (or) } Y_n \xrightarrow{P} Y$$

$Y = c$ is constant.

$Y_1, Y_2 \ldots$ converges in distribution to Y if for all continuity points y of the distribution function.

$F(y) = P(Y \leq y)$ of y it holds that

$$\lim_{n \to \infty} P(Y_n \leq y) = F(y)$$

We then write $\lim Y_n = Y$ (or) $Y_n \xrightarrow{D} Y$

Asymptotics of MLE in canonical exponential families:

Consider a sample of $(X_1, X_2, \ldots X_n)$ of size n from a d. dimensional canonical exponential family with individual densities

$$f(x;\theta) = b(x)\, e^{\theta^T t(x) - c(\theta)}, \quad \theta \in \Theta \subseteq R^d$$

we have seen that the MLE of θ is given as,

$$\hat{\theta} = \hat{\theta}_n = T^{-1}(\bar{T}_n)$$

where T is then mean value mapping

and $$\bar{T}_n = \frac{t(X_1) + \ldots + t(X_n)}{n}$$

we will know that the MLE is asymptotically normally distributed and asymptotically unbiased and efficient

(i.e) $\hat{\theta}_n \overset{a}{\sim} N_d \{\theta, i(\theta)^{-1}/n\}$

The central limit theorem yields for r.

$\eta = T(\theta)$ that

$$T_n \overset{a}{\sim} N_d \{\eta, 1/n \, i(\theta)\}$$

using inverse function for $g = T^{-1}$ gives

$$g'(\eta) = \frac{d\theta}{d\eta} \cdot \left\{\frac{d\eta}{d\theta}\right\}^{-1} = \{T'(\theta)\}^{-1} = i(\theta)^{-1}$$

The delta method now yields,

$$\hat{\theta}_n \overset{a}{\sim} N_d \{g(\eta), 1/n \; i(\theta)^{-1} i(\theta) \, i(\theta)^{-1}\}$$

$$= N_d \{\theta, i(\theta)^{-1}/n\}$$

Theorem 1:

We have $\sqrt{n}(\hat{\theta} - \theta_0) \to N\left(0, \frac{1}{I(\theta_0)}\right)$

as we can see that the asymptotic variance dispersion of the estimate around the true parameter will be smaller than Fisher information is large.

Proof:

Since MLE $\hat{\theta}$ is maximizer of

$$L_n(\theta) = 1/n \sum_{i=1}^{n} \log f(x_i/\theta)$$

we have

$$\frac{f(a) - f(b)}{a-b} = f'(c) \text{ or } f(a) + f'(c)(a-b) \text{ for } c \in [a, b]$$

with $f(\theta) = L_n'(\theta)$

$a = \hat{\theta}$

and $b = \theta$.

we can write,

$$0 = L_n'(\hat{\theta}) = L_n'(\theta_0) + L_n''(\hat{\theta}_1)(\hat{\theta} - \theta_0)$$

$$\hat{\theta}_1 \in [\hat{\theta}, \theta_0]$$

we get that

$$\hat{\theta} - \theta_0 = \frac{L_n'(\theta)}{L''(\hat{\theta}_1)} \quad \text{and}$$

$$\sqrt{n}\,(\hat{\theta} - \theta_0) = \frac{\sqrt{n}\, L_n'(\theta)}{L_n''(\hat{\theta}_1)}$$

therefore

$$\sqrt{n}\, L_n'(\theta_0) = \sqrt{n}\left(\tfrac{1}{n}\sum_{i=1}^{n} l'(x_i/\theta_0) - 0\right) - \sqrt{n}\left(\tfrac{1}{n}\sum_{i=1}^{n} l'(x_i - \theta_0) - E_\theta\, l'(x_i - \theta_0)\right)$$

$$\rightarrow N(0, \operatorname{var}\theta_0(l'(x_1/\theta_0)))$$

$$L_n''(\theta) = \tfrac{1}{n}\sum l''(x_i/\theta) \rightarrow E\theta_0\, l''(x_i/\theta)$$

$$L_n''(\hat{\theta}_1) = E\theta_0\, l''(x_i/\theta_0) = -I(\theta_0)$$ be lemma

above combining this equation

$$-\frac{\sqrt{n}\, L_n'(\theta_0)}{L_n''(\hat{\theta}_1)} \xrightarrow{d} N\left(0, \frac{\operatorname{Var}\theta_0(l'(x_i/\theta_0))}{(I(\theta_0))^2}\right)$$

Finally the variance,

$$V\theta_0(l'(x_i/\theta_0)) = E\theta_0(l'(x/\theta_0))^2 - (-\theta_0\, l'(x/\theta_0))^2$$

$$= I(\theta_0) - 0$$

Confidence Interval for Large Sample:

It has been proved that under certain regularity conditions the first derivative of the logarithm of the likelihood function w.r.to parameter θ $\partial/\partial\theta \log L$ is asymptotically normal with zero and variance is

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

$$= E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)$$

Hence for large n,

$$Z = \frac{\partial/\partial\theta \log L}{\sqrt{\operatorname{var}(\partial/\partial\theta \log L)}} \sim N(0,1)$$

Thus for large samples the confidence interval for θ with confidence coefficient $(+ (1-\alpha))$ is obtained by converting the inequalities in $P(|Z| \leq \lambda_\alpha) = 1-\alpha$

where

λ_α is given by $\frac{1}{\sqrt{2\pi}} \int_{-\lambda_\alpha}^{\lambda_\alpha} \exp\left(-u^2/2\right) du = 1-\alpha$