CHAPTER - II

SUFFICIENCY AND INFORMATION

2.0 Introduction :

Suppose a certain experiment is performed, the main job of the statistician is to store and interpret data and draw valid conclusions. To store the raw data is very costly, since he likes to condense whole data in terms of a single figure, i.e. the figure represents whole data, that single figure is called 'statistic'.

In the problem of statistical inference one has to estimate an unknown parameter as to test certain hypothesis, based on the data collected for the purpose. Often the data can be simplified through the computation of a few numerical values (called statistic). The statistical analysis can be based on the above summarised data, just as effectively as an analysis that could be based on the original data. Such summaries are known as sufficient statistic. Loosely speaking, a statistic T is called sufficient it contains all the information that entire data contains. In the classical setup sufficient statistic is defined as follows :

Definition 2.0.1 : Let $X = (x_1, x_2, \ldots, x_n)$ be a sample from $\{F_\Theta : \Theta \in \Theta\}$ A statistic $T = T(x)$ is said to be sufficient statistic for the unknown parameter Θ, if and only if, the conditional distribution of X given $T = t$ is independent of Θ.

The entire sample is always sufficient for Θ. Thus the problem is to find a sufficient statistic of minimum dimensions. To find a sufficient statistic the factorization criterion can be used.

The Factorization Theorem 2.0.1 : Let, $x_1, x_2, \ldots, x_n$ be random variables with probability mass function $f_\Theta(x_1, x_2, \ldots, x_n), \Theta \in \Theta$. Then, $T(x_1, x_2, \ldots, x_n)$ is sufficient for Θ if and only if,

$$f_\Theta(x_1, x_2, \ldots, x_n, \Theta) = h(x_1, x_2, \ldots, x_n)\ f_\Theta(T(x_1, x_2, \ldots, x_n)),$$

Where h is independent of Θ and it is non-negative function of X and $f_\Theta(T(x_1, x_2, \ldots, x_n))$ is non-negative function of Θ and statistic T.

The discussion given above is interms of a classical setup. This chapter contains three sections. The Section 2.1 deals with sufficiency through information in which sufficiency of estimator is defined in terms of information theory. The Section 2.2 deals with minimum discrimination function in which

result...

related to the estimation is given, and $f_1(x)$, the density function which close to $f_2(x)$ is found out under certain conditions such that the discrimination function must be minimum. The section 2.3 deals with a variant of the fundamental lemma of Neyman and Pearson in which the Neyman-Pearson Lemma is proved in bayesian form (Renyi 1966).

2.1 Sufficiency Through Information :

Let, $(\Omega, \mathbb{F}, p)$ be a probability space. Where Ω is non-empty any arbitrary set, $\mathbb{F}$ be σ- field of subsets of Ω and p is a probability measure defined on $\mathbb{F}$. Let Θ be a discrete random variable in $(\Omega, \mathbb{F}, P)$. A function $\Theta = \Theta(\omega)$ where, $\omega \in \Omega$, taking finite number of different values $\Theta_1, \Theta_2, \ldots, \Theta_r$ $(r \geq 2)$ for $\omega \in \Omega$. We define event B_k such that,

$$B_k = \{\omega \mid \Theta(\omega) = \Theta_k\} \quad k = 1,2,\ldots r \qquad (2.1.1)$$

and $B_k \in \mathbb{F}$.

We interpret Θ as the parameter of probability distribution. And the event B_k as hypothesis that the true value of Θ is equal to Θ_k i.e.

$$p_k = p(B_k) = p(\Theta = \Theta_k) \quad k = 1,2,\ldots,r \qquad (2.1.2)$$

Thus the prior distribution of Θ be $P_1, P_2, \ldots, P_r$. Then the Shannon's entropy is defined as

$$H(\Theta) = -\sum_{k=1}^{r} p_k \log_2 p_k \qquad (2.1.3)$$

$H(\Theta)$ is interpreted as "the amount of missing information" about Θ when Θ is unknown except its prior distribution is given ($\Theta_1, \Theta_2, \ldots \Theta_k$ are known).

Let $X = X_1(\omega), X_2,(\omega), \ldots, X_n(\omega)$ are random variables which take only finite number of 's' different values. Let $X = X(\omega)$ be a random sample and $p(k|j)=p_k(x)$ be the posterior distribution of Θ after observing $X = x_j$ is given by $p(k|j) = p(\{\omega|\Theta = \Theta_k\}|\{\omega|X = x_j\})$ (2.1.4) where, $x_1, x_2, \ldots, x_s$ are the vector values of X. Then, the posterior entropy will be

$$H(\Theta|x_j) = -\sum_k p(k|j) \log_2 p(k|j) \qquad (2.1.5)$$

This can be interpreted as "the amount of the information concerning Θ still missing even observing x_j i.e. $X = x_j$. The average amount of information concerning Θ still missing" is given as

$$H(\Theta|X) = \langle H(\Theta|x_j)\rangle = -\sum_{k=1}^{r}\sum_{j=1}^{s} p\{\omega|X=x_j\} p(k|j) \log_2 p(k|j)$$
(2.1.6)

Which is interpreted as "the average amount of information still missing for Θ after observing random sample X".

Define,

$$R(\Theta, X) = H(\Theta) - H(\Theta|X) \qquad (2.1.7)$$

Where, $R(\Theta,X)$ is 'the amount of information in the observed sample X with respect to unknown parameter Θ'. It is the average decrease of the entropy of Θ after observing X.

Let the conditional distribution $p(k|j)$

$p(k|j) = (p_1(x), \dots p_r(x))$ of Θ is indentical with the prior distribution $(p_1,p_2,\dots,p_r)$; if and only if Θ and X are independent.

In this case,

$$H(\Theta|X) = H(\Theta). \qquad (2.1.8)$$

i.e. $R(\Theta,X) = 0$

i.e. the observations of X cannot give any information about Θ. From equation (1.2.10)

$$H(\Theta|X) \leqslant H(\Theta) \qquad (2.1.9)$$

i.e. $R(\Theta,X) > 0$.

Let $T(X)$, $X \in E_n$ be any k dimentional vector valued borel measurable function on n - dimentional space E_n. We call $T(X)$ as statistic because after observing the sample X, we can calculate $T(X)$.

<u>Theorem 2.1.1</u> We have,

$$R(\Theta, T(x)) = R(\Theta, X) \tag{2.1.10}$$

if and only if the conditional probability distribution of Θ given value of X is fixed, if it depends on the value of T(x) only,

$$\text{i.e.} \quad p(k|i) = p(k|j) \tag{2.1.11}$$

<u>Proof</u> - Let us define,

$$p(k,j) = p(\{\omega | \Theta = \theta_k, X = x_j\}) \tag{2.1.12}$$

$$q_j = p(\{\omega | X = x_j\})$$

Let $y_1, y_2, \ldots, y_m$ be the values taken by the function T on the set $\{x_1, x_2, \ldots x_s\}$

Let,

$$C_l = \{j | T(x_j) = y_l\} \cdot l = 1,2,\ldots,m \tag{2.1.13}$$

Now define,

$$h(x_j) = l, \; j \in C_l \; (l = 1,2,\ldots, m) \tag{2.1.14}$$

The probability,

$$\left.\begin{aligned} r(k, l) &= p(\{\omega \; \theta=\Theta_k, \; T(x)=y_l\}), \\ r(l) &= p(\{\omega | T(x) = y_l\}). \end{aligned}\right\} \tag{2.1.15}$$

The numbers,

$$u(k,j) = r(k,l) \, q_j \, / \, r(l) \tag{2.1.16}$$

Where $r(k, \ell) = \sum_{j \in C_\ell} p(k,j)$

i.e. $R(\Theta.X) - R(\Theta,T(x)) = H(\Theta) - (H(\Theta|X) - H(\Theta) + H(\Theta|T(X))$

$= H(\Theta|T(X)) - H(\Theta|X)$

i.e. $H(\Theta|T(x)) - H(\Theta|X)$

$$= -\sum_{k=1}^{r}\sum_{\ell=1}^{m} r(k,\ell) \log_2 [r(k,\ell)/r(\ell)] - $$

$$-\sum_{k=1}^{r}\sum_{j=1}^{s} p(k,j) \log_2 [p(k,j)/q_j]$$

$$= -\sum_{k=1} [p(k,1) \log_2 u(k\ 1) / q_1 +$$

$$+ p(k,2) \log_2 u(k,2)/q_2 + \ldots +] +$$

$$+ \sum_{k-1}\sum_{j=1} p(k\ j) \log_2 p(k\ j) / q_j$$

i.e. $$-\sum_{k=1}\sum_{j=1} p(k,j) \log_2 (u(k,j)/q_j) +$$

$$+ \sum_{k=1}\sum_{j=1} p(k,j) \log_2 (p(k,j) / q_j)$$

$$= \sum_{k=1}\sum_{j=1} [p(k,j) \log_2 (p(k,j) / u(k,j)) \geq 0 \qquad (2.1.17)$$

if and only if (2.1.10) satisfies

$$p(k,j) = u(k,j)$$

if $r(k,h(x_j))q_j / r(h(x_j)) = u(k,j)$

if and only if,

$$u(k,j) / q_j = u(k,i) / q_i \qquad (2.1.18)$$

i.e. $R(\Theta\ X) - R(\Theta,\ T(x) \geqslant 0$

i.e. $R(\Theta,\ T(X) \leqslant R(\Theta,X)$.

The amount of information contained in the observations of X is greater than the amount of information contained in T(X), i.e. $H(\Theta|X) \leqslant H(\Theta|T(x)$

If $R(\theta, T(x)) = R(\theta, x)$ with

$P(K|j) = P(K|i)$. a.e.

We call the function T(X) is sufficient for Θ. Now definition of sufficiency is given as

<u>Definition 2.1.1</u> The function T(x) of the observations is said to be sufficient if and only if the amount of information contained in the observations is equal to the amount of information contained in the function T(x) about the unknown parameter Θ.

Let the random vector X has density function $f_\Theta\ (x_1, x_2, \ldots, x_n)$ and $\phi_k(T(x))$ be the density function of T(x). Then,

$$f_\Theta\ (x_1, x_2, \ldots x_n) = \phi_k(T(x), \Theta)\ .\ h(x) \tag{2.1.20}$$

Where, h(x) is independent Θ. This is usual definition of sufficiency (through factorization theorem).

2.2 Minimum Discrimination Function :

In the previous chapter (Section 1.4) we have considered the Kullback-Leibler information measure I(1:2), and it is given as

$$I(1:2) = \int f_1(x) \log_2 [f_1(x)/f_2(x)] \, d\lambda (x) \qquad (2.2.1)$$

It is also considered as directed divergence.

Suppose that, $f_2(x)$ is given, then our aim is to select a probability measure P_1 close to P_2, i.e. we select $f_1(x)$ such that I(1:2) must be minimum. That is the mean information for discrimination in favour of H_1 against H_2 is minimum, where H_i, i=1,2 is the hypotheses that the density of X is $f_i(x)$.

Definition 2.2.1 The variation of the information when we pass from the initial probability measure p^* to the new probability measure p absolutely continuous to p^* is denoted by $H(p|p^*)$ is defined as

$$H(p|p^*) = \int_{\Omega} \phi(x) \log_e \phi(x) \, dp^* \qquad (2.2.2)$$

where, $\phi(x)$ is Randon-Nikodym derivative of p with respect to p^*.

Note that

$$\int \phi(x) \log_e \phi(x) \, dp^* = \int \log_e \phi(x) \, dp \qquad (2.2.3)$$

Let, $(\Omega, \mathbb{F}, \nu_2)$ be a probability space and $T(x)$ be a real valued $\mathbb{F}$ - measurable, non-degenerate bounded function. The probability measure ν^* is defined as

$$\nu^* (A) = \left[\int_A \exp(-\beta T(x)) d\nu_2 \right] / \phi(\beta) \qquad (2.2.4)$$

For every $A \in \mathbb{F}$, where,

$$\phi(\beta) = \int_\Omega \exp(-\beta T(x)) \, d\nu_2 (x) \qquad (2.2.5)$$

and

$$\int_\Omega T(x) d\nu^*(x) = \Theta. \qquad (2.2.6)$$

<u>Lemma 2.2.1</u> : The equation $\phi'(\beta) / \phi(\beta) = - \Theta$ (2.2.7) has unique solution.

<u>Proof</u> : Let, $g = \Theta - T$ (2.2.8)

Observe that, T is non-degenerate then,

$$\nu_2\{g > 0\} > 0, \quad \nu_2\{g < 0\} > 0$$

Define

$$G(\beta) = \int_{\Omega} (\theta - T) \exp [\beta(\theta - T)] d\nu_2 \qquad (2.2.9)$$

$$= \int_{A} (\theta - T) \exp [\beta(\theta - T)] d\nu_2 +$$

$$+ \int_{A^C} (\theta - T) \exp [\beta(\theta - T)] d\nu_2$$

Where,

$$A = \{ x | g(x) > 0 \}, \; A^C = \{ x | g(x) \leq 0 \}$$

we can write

$$\lim_{\beta \to \infty} \int_{\Omega} g(x) \exp [\beta g(x)] d\nu_2(x) =$$

$$= \lim_{\beta \to \infty} \int_{A} g(x) \exp [\beta g(x)] d\nu_2(x) +$$

$$+ \lim_{\beta \to \infty} \int_{A^C} g(x) \exp [\beta g(x)] d\nu_2$$

$$= \infty + 0$$

$$= + \infty$$

Similarly,

$$\lim_{\beta \to -\infty} \int_{\Omega} g(x) \exp [\beta g(x)] d\nu_2(x) =$$

$$= \lim_{\beta \to -\infty} \int_{A} g(x) \exp [\beta g(x)] d\nu_2(x) +$$

$$+ \lim_{\beta \to -\infty} \int_{A^C} g(x) \exp [\beta g(x)] d\nu_2(x)$$

$$= 0 - \infty = - \infty$$

i.e. $G(\beta)$ is the increasing function taking values $-\infty$ to ∞

Therefore,

$$G'(\beta) = \int_{\Omega} g^2(x) \exp[\beta g(x)]\, d\nu_2(x) > 0.$$

$G(\beta)$ exists, therefore, $G(\beta)$ is continuous

$$\Rightarrow G(\beta) = 0 \tag{2.2.10}$$

has unique solution.

$$\int_{\Omega} (\Theta - T) \exp[\beta(\Theta - T)]\, d\nu_2 = 0$$

$$\int_{\Omega} \Theta \exp[\beta(\Theta - T)] d\nu_2 - \int T \exp[\beta(\Theta - T)]\, d\nu_2 = 0$$

$$\int_{\Omega} \Theta \exp[\beta(\Theta - T)] d\nu_2 = \int_{\Omega} T \exp[\beta(\Theta - T)]\, d\nu_2$$

i.e. $\exp[\beta\Theta] \int_{\Omega} \Theta \exp[-\beta T] d\nu_2 = \exp[\beta\Theta] \int_{\Omega} T \exp[-\beta T] d\nu_2$

i.e. $\int_{\Omega} \Theta \exp[-\beta T]\, d\nu_2 = \int_{\Omega} T \exp[-\beta T]\, d\nu_2$

i.e. $-\Theta \int_{\Omega} \exp[-\beta T] d\nu_2 = -\int_{\Omega} T \exp[-\beta T]\, d\nu_2$

$$-\Theta\, \phi(\beta) = \phi'(\beta)$$

i.e. $\phi'(\beta) / \phi(\beta) = -\Theta$

Theorem 2.2.1 Let,

$$\int_{\Omega} T(x)d\nu_1(x) = \Theta \qquad (2.2.11)$$

then the inequality,

$$H(\nu_1 | \nu_2) \geq H(\nu^* | \nu_2) = \log_e[1/\phi(\beta)] - \beta\Theta \qquad (2.2.12)$$

where, the equality holds if and only if the probability measure ν_1 is equal to the probability measure ν^* on $\mathbb{F}$.

Proof : The above inequality holds if

$$H(\nu_1 | \nu_2) = \infty$$

Let us, ν_1 is absolutely continuous with respect to ν_2 which is initial probability measure. Therefore suppose that

$$H(\nu_1 | \nu_2) < +\infty$$

and there exists unique function by Randon-Nikodym theorem

$$\phi(x) = (d\nu_1 / d\nu_2)(x)$$

Then,

$$H(\nu_1 | \nu_2) = \int_{\Omega} \phi(x) \log_e \phi(x) d\nu_2(x) \qquad (2.2.13)$$

$$H(\nu^* | \nu_2) = \int_{\Omega} \phi^*(x) \log_e \phi^*(x) d\nu_2(x) \qquad (2.2.14)$$

Where, $\phi^*(x) = (d\nu^* / d\nu_2)(x) = \exp[-\beta T(x)] / \phi(\beta)$

By the equations (2.2.6) and (2.2.11) we can write

$$\int_{\Omega} \phi(x) \log_e \phi^*(x) d\nu_2(x) = \int_{\Omega} \phi^*(x) \log_2 \phi^*(x) d\nu_2(x) \tag{2.2.15}$$

But from the equations (2.2.13), (2.2.14) and (2.2.15) it is enough to show that

$$\int_{\Omega} \phi(x) \log_e \phi^*(x) d\nu_2(x) \leq \int_{\Omega} \phi(x) \log_e \phi(x) d\nu_2$$

therefore.

$$\int_{\Omega} \phi(x) \log_e [\phi^*(x) / \phi(x)] d\nu_2(x) \leq 0 \quad (2.2.16)$$

$= -I(\nu, \nu^*) \leq 0$ which is true.

i.e. $I\ (1:2) \geq 0$

equality holds if and only if $\nu_1 = \nu^*$.

Theorem 2.2.1 has the following interpretation :

Let, ν_2 be given probability measure and $\mathcal{P}$ be the class of all probability measures ν_1 satisfying the condition $\int T(x) d\nu_1(x) = \theta$.

Then, $H(\nu_1 | \nu_2)$ is minimum for $\nu_1 = \nu^*$ when minimum is taken over $\mathcal{P}$.

Equivalently,

$$\min_{\nu_1 \in \mathbf{P}} H(\nu_1 | \nu_2) = H(\nu^* | \nu_2)$$

The above theorem 2.2.1 can be used for the estimation problems also. The corresponding criterion is similar to that of maximum likelihood criterion. The criterion based on theorem (2.2.1) is a variant of the principle of maximum information applied to the discrimination function. The following result gives the similarity of the principle of maximum information and that of the principle of minimum discrimination function.

Theorem 2.2.2 The probability function

$$f_1^*(x) \geq 0, \qquad \int_\Omega f_1^*(x)\,d\lambda(x) = 1 \qquad (2.2.17)$$

which minimizes the discrimination function

$$I(1:2) = \int_\Omega f_1(x) \log_e[f_1(x)/f_2(x)]\, d\lambda(x)$$

subject to constraint

$$\int_\Omega T(x) f_1(x)\, d\lambda(x) = \theta \qquad (2.2.18)$$

is given by

$$f_1^*(x) = f_2(x) \exp[-\beta T(x)] / \phi(\beta) \qquad (2.2.19)$$

Where, $\phi(\beta) = \int_{\Omega} f_2(x)\exp[-\beta T(x)]d\lambda(x)$ (2.2.20)

β being the unique solution of the equation

$$(d/d\beta)\ (\log_e \phi(\beta) = -\theta.$$

and T being a non-degenerate random variable.

<u>Proof</u> : We have to maximize the function

$$I = -I(1:2) = \int_{\Omega} f_1(x)\ \log_e[f_2(x)/f_1(x)]\ \ d\lambda(x)$$

using the Langragian multiplier method. and considering that

$$\log_e x < x-1 \quad \text{if} \quad x \neq 1$$

$$\text{and} \quad \log x = x-1 \quad \text{if} \quad x = 1.$$

We get,

$$I - \alpha - \beta\theta = \int_{\Omega} f_1(x)\log_e[f_2(x)/f_1(x)]d\lambda(x) - \alpha[\int_{\Omega} f_1(x)d\lambda(x)] -$$

$$-\beta\int_{\Omega} T(x)f_1(x)d\lambda(x)$$

$$= \int_{\Omega} f_1(x)[\log_e(f_2(x)/f_1(x)) - \alpha - \beta T(x)]d\lambda(x)$$

$$= \int_{\Omega} f_1(x)\log_e[(f_2(x)/f_1(x))\exp(-\alpha-\beta T(x)]d\lambda(x)$$

$$\leq \int_{\Omega} f_1(x)\log_e[(f_2(x)/f_1(x))(\exp[-\alpha-\beta T(x)])-1]d\lambda(x)$$

Where the equality holds if and only if

$$f_1(x) = f_2(x) \exp[-\alpha - \beta T(x)] \text{ a.e. } \lambda.$$

We get the constants α and β using the equations (2.2.17) and (2.2.18).

We get

$$\alpha = \log_e \phi(\beta).$$

and also we obtain the minimum value of the discrimination function. It is given as

$$I(* : 2) = -\Theta\beta - \log_e \phi(\beta) \qquad (2.2.21)$$

We know that Θ is the parameter if we take a sample of n observations and we estimate the value of Θ on the basis of the sample observations say $\hat{\Theta}(x)$. Then, we also estimate the value of $I(* : 2)$ and also $\beta(x) = \beta(\hat{\Theta}(x))$ such that,

$$T(x) = \hat{\Theta}(x) = -[(d/d\beta)(\log_e \phi(\beta))]_{\beta=\hat{\beta}(x)=\beta(\hat{\Theta}(x))}$$

$$\text{i.e.} \hat{I}(*:2) = -\hat{\Theta}(x)\,\hat{\beta}(x) - \log_e \phi(\hat{\beta}(x))$$

$$= -\hat{\Theta}\beta(\hat{\Theta}) - \log \phi\,(\hat{\beta}(\Theta))$$

The $\hat{I}(* : 2)$ is the minimum discrimination information function between the populations with the probability density function $f^*(x)$ with the parameter value Θ and its estimated value $\hat{\Theta}$ and with the density function $f_2(x)$. And

$$\hat{I}(* : 2) \geq 0$$

If equality holds, the estimated value $\hat{\Theta}$ is equal to the parameter value Θ and population having probability density $f_2(x)$; Also $I(* : 2)$ is interpreted as deviation between the sample and the population density function $f_2(x)$.

As application of above theorem (2.2.2) we consider following example :

Example 2.2.1 $f_2(x) \rightarrow N(\Theta,1)$ and $T(x) = \bar{x}$

$$T(x) = \bar{x} \quad \text{i.e. } \bar{x} \longrightarrow N\ (\Theta, 1/n)$$

$$f_2(x) = 1/\sqrt{2\pi}\ \exp\ [(-1/2)(x-\Theta)^2]$$

$$-\infty < x < \infty$$

We know that

$$\int_{\Omega} \bar{x}\ f_1^*(x)\ dx = \Theta'$$

$$\phi(\beta) = \int_{-\infty}^{\infty} \exp(-\beta\bar{x})(1/\sqrt{2\pi 1/n})\exp[(-n/2)(\bar{x}-\theta)^2]\ d\bar{x}$$

$$= \sqrt{n}/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp\ [-\beta\bar{x} - (n/2)(\bar{x} - \theta)^2]\ d\bar{x}$$

$$= \sqrt{n}/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp\left\{(-n/2)[(\bar{x}-\theta)^2 + 2\beta\bar{x}/n]\right\} dx$$

$$= \sqrt{n}/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp\left\{(-n/2)[\bar{x}^2 - 2\bar{x}\theta + \theta^2 + 2\beta\bar{x}/n]\right\} dx$$

$$= \exp(-n\theta^2/2)\sqrt{n}/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp\left\{(-n/2)[\bar{x}^2 - 2(\theta-\beta/n)\bar{x} + (\theta - \beta/n)^2 - (\theta-\beta/n)^2]\right\} dx$$

$$= \exp[(-n\theta^2/2) + n(\theta-\beta/n)^2/2]\sqrt{n}/\sqrt{2\pi} \int_{-\infty}^{\infty} \exp\left\{-n/2[\bar{x}-(\theta-\beta/n)^2]\right\} dx$$

$$= \exp\ [-n\theta^2/2 + n(\theta-\beta/n)^2/2]$$

$$= \exp\ [-n\theta^2/2 + n/2(\theta^2 - 2\beta\theta/n + \beta^2/n^2)]$$

$$= \exp\ [-\beta\theta + \beta^2/2n].$$

Also, we know that,

$$((d/d\beta)\ [\log_e \phi(\beta)] = -\theta'$$

$$(d/d\beta)\ (-\beta\theta + \beta^2/2n) = -\theta'$$

i.e. $-\Theta + 2\beta \quad 2n \quad = -\Theta'$

i.e. $\beta = n(\Theta - \Theta')$

i.e. $f^*(x)$

$= f_2(x) \exp(-\beta\bar{x}) / \phi(\beta)$

$= (\sqrt{n}/\sqrt{2\pi}) \exp (\frac{-n}{2})(\bar{x}-\Theta)^2 \exp(-\beta\bar{x})/\exp[-\beta\Theta+\beta^2/2n]$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)[(\bar{x}-\Theta)^2+2\beta\bar{x}/n - 2\beta\Theta/n + B^2/n^2]\}$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)[\bar{x}^2-2\bar{x}\Theta+\Theta^2+2\beta\bar{x}/n-2\beta\Theta/n + B^2/n^2]$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)[\bar{x}^2-2(\Theta-\beta/n)\bar{x} + (\Theta-\beta/n)^2-(\Theta-\beta/n)^2+$

$+ \Theta^2 - 2\beta\Theta/n + B^2/n^2]\}$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)[(\bar{x} - \Theta - \beta/n)]^2-(\Theta-\beta/n)^2+(\Theta-\beta/n)^2\}$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)[\bar{x} - (\Theta - \beta/n)]^2\}$

$= (\sqrt{n}/\sqrt{2\pi}) \exp (-n/2)[\bar{x}-(\Theta-n(\Theta - \Theta')/n)]^2$

$= (\sqrt{n}/\sqrt{2\pi}) \exp\{(-n/2)(\bar{x} - \Theta')^2\}$

Thus $f^* \longrightarrow N(\Theta', 1/n)$

Therefore,

$$I(* : 2) = -\Theta'\beta - \log_2 \phi(\beta)$$

$$= -\Theta' n(\Theta-\Theta')+n\Theta(\Theta-\Theta')-n^2(\Theta-\Theta')^2/2n$$

$$= n(\Theta - \Theta')^2 / 2$$

2.3 A variant of the Fundamental Lemma of Neyman and Pearson :

Let, Θ be the unknown parameter and $X_1, X_2, \dots X_n$ be a sequence of random variables (say random sample X). The distribution of a random sample X depends on Θ. We also consider as Θ is a random variable which takes values Θ_o and Θ_1 with probabilities W_o and W_1 respectively. We also suppose that the random variables x_n independently and identically distributed under $\Theta = \Theta_o$ as well as $\Theta = \Theta_1$. The density functions of a random sample under $\Theta = \Theta_o$ and $\Theta = \Theta_1$ are $f_o(x)$ and $f_1(x)$ respectively with $f_o(x) \neq f_1(x)$.

We know that the amount of information $R(\Theta, X)$ containing in observing the sample is given as :

$$R(\Theta \quad X) = H(\Theta) - H(\Theta|X) \qquad (2.3.1)$$

Where, $H(\Theta)$ is the entropy of Θ.

i.e. $H(\Theta) \quad = \quad W_o \log_2 1/W_o + W_1 \log_2 1/W_1$

$$H(\Theta|x_n) \quad = \quad p(\Theta = \Theta_o \mid x_n) \log 1/p(\Theta = \Theta_o \mid x_n) +$$

$$p(\Theta = \Theta_1 | x_n) \log_e 1/p(\Theta = \Theta_1 \mid x_n)$$

Where, $H(\Theta|X)$ denotes the expection of $H(\Theta|x_n)$. Also $R(\Theta, X)$ is interpreted as the average information about Θ after observing sample X.

The fundamental lemma of Neyman and pearson is used in obtaining most powerful test for testing $\theta = \theta_o$ verses $\theta = \theta_1$. But here we consider bayesian form of Neyman Pearson Lemma. It is stated as 'decision procedure for which probability of an error is minimal'. Such decision is called as standard decision.

Let us, consider a decision $\Delta = \Delta(X)$ is a borel measurable function of a sample on the values of $\theta = \theta_o$ and $\theta = \theta_1$. If $\Delta = \theta_o$, we accept the hypothesis $\theta = \theta_o$ and if $\Delta = \theta_1$ we accept the hypothesis $\theta = \theta_1$. The error in the decision is defined as the decision which taken is not correct and it is denoted by ϵ.

$$\epsilon = p[\Delta \neq \theta]$$

i.e. $$\epsilon = W_o p(\Delta = \theta_1 | \theta = \theta_o) + W_1 p(\Delta = \theta_o | \theta = \theta_1) \qquad (2.3.2)$$

On the basis of the sample standard decision is :

$$\left.\begin{array}{lll} \text{if } p(\theta = \theta_o | X) > p(\theta = \theta_1 | X), & \text{accept} & \theta_o \\ \text{if } p(\theta = \theta_1 | X) > p(\theta = \theta_o | X), & \text{accept} & \theta_1 \\ \text{if } p(\theta = \theta_o | X) = p(\theta = \theta_1 | X), & & \end{array}\right\} \qquad (2.3.3)$$

The choice is made either Θ_o or Θ_1 with probability W_o and W_1 respectively for $p(\Theta = \Theta_o | X) = p(\Theta = \Theta_1 | X)$.

<u>Theorem 2.3.1</u> : No decision can have a smaller error than standard decision.

<u>Proof</u> : Consider the sample space 'S' and it is divided into three disjoint parts as S_o, S_1 and S_2 such that

$$\left.\begin{array}{l} X \in S_o \text{ if } p(\Theta_1 | X) < p(\Theta_o | X) \\ X \in S_1 \text{ if } p(\Theta_o | X) < p(\Theta_1 | X) \\ X \in S_2 \text{ if } p(\Theta_o | X) = p(\Theta_1 | X) \end{array}\right\} \qquad (2.3.4)$$

Let, $\bar{y} = (y_1, y_2, \ldots, y_n)$ and $f_i(\bar{y}) = \prod_{i=1}^{n} f(y_i)$

Define,

$$\delta(\bar{y}) = \begin{cases} 1 & \text{if } \bar{y} \in S_1 \\ 0 & \text{if } \bar{y} \in S_o \\ W_1 & \text{if } \bar{y} \in S_2 \end{cases} \qquad (2.3.5)$$

Equivalently δ is given as

$$\delta(\bar{Y}) = \begin{cases} 1 & \text{if } f_1(\bar{Y})W_1 > f_o(\bar{Y})\, W_o \\ 0 & \text{if } f_o(\bar{Y})W_o > f_1(\bar{Y})\, W_1 \\ W_1 & \text{if } f_o(\bar{Y})W_o = f(\bar{Y})\, W_1 \end{cases} \qquad (2.3.6)$$

The error in the standard decision is also given as

$$\epsilon = W_o \int \delta(\bar{y}) f_o(\bar{y}) d\bar{y} + w_1 \int [1 - \delta(\bar{y})] f_1(\bar{y}) \; d\bar{y}) \qquad (2.3.7)$$

$d\bar{y}$ denotes $dy_1 \; dy_2 \ldots, dy_n$ and integral is all over sample space.

Let us consider Δ^* be another different decision from standard decision Δ, then

$$\delta^*(\bar{y}) = \begin{cases} 1 & \text{if } \Delta^* = \theta_1 \\ 0 & \text{if } \Delta^* = \theta_o \end{cases} \qquad (2.3.8)$$

The error in decision Δ^* is denoted by ϵ^*.

$$\epsilon^* = W_o \int \delta^*(\bar{y}) f_o(\bar{y}) d\bar{y} + W_1 \int [1 - \delta^*(\bar{Y})] \; f_1(\bar{y}) d(\bar{y}) .$$

Therefore,

$$\epsilon^* - \epsilon = \int [\delta^*(\bar{y}) - \delta(\bar{y})][W_o f_o(\bar{Y})] \; d\bar{y} - \int [\delta^*(\bar{y}) - \delta(\bar{y})] W_1 f_1(\bar{y})$$

$$\geq 0$$

$$\epsilon^* \geq \epsilon$$

By likelihood ratio test we accept hypothesis

$$\theta = \theta_1, \qquad \text{if } [f_1(\bar{Y}) \; / \; f_o(\bar{Y})] > [W_o \; / \; W_1]$$

and the hypothesis $\Theta = \Theta_o$ is accepted if

$$[f_1(\bar{Y}) / f_o(\bar{Y})] < [W_o/W_1]$$

$$\text{if} \quad [f_1(\bar{Y}) / f_o(\bar{Y})] = [W_o/W_1]$$

The random choice is made between $\Theta = \Theta_o$ and $\Theta = \Theta_1$ with probabilities W_1 and W_o respectively.

Theorem 2.3.2 There exist constants A and λ with $A > 0$ and $0 < \lambda < 1$ depending on $f_o(x)$, $f_1(x)$ and W_o, such that

$$0 < H(\Theta) - R(\Theta, X) \leq A\lambda^n \quad (n = 1, 2, \ldots)$$

For λ we may take the value

$$\lambda = \inf_{0 \leq \alpha \leq 1} \left(\int_{-\infty}^{\infty} f_1^{\alpha}(x) f_o^{(1-\alpha)}(x) \, dx \right)$$

Theorem 2.3.3 Let ϵ denote the error of standard decision, Then,

$$\epsilon \leq [H(\Theta) - R(\Theta, X)]$$

From these both theorems, the ϵ_n is error in the standard decision after observing sample X. Therefore, $\epsilon_n \leq A\lambda^n$, $n = 1, 2, \ldots$

This implies the series $\sum_{n=0}^{\infty} \epsilon_n$ is convergent. Then by Borel cantelli lemma, if we take samples indefinitely in number and make standard decision for each n with probabilityl, the situation will occur that our all decisions are correct. Here, to find out Δ_n, we consider following example :

Example 2.3.1 : If $f_{\Theta/x}(x) = \exp(\Theta)$ based on single observation $p(\Theta = \Theta_1) = 1/3 = W_o$,

$$p(\Theta = \Theta_2) = 2/3 = W_o \quad \text{with } H_o : \Theta_1 = 1$$

$$H_1 : \Theta_2 = 2$$

$$\delta(x) = 1$$

Let,

$$\Delta_n : d = \Theta_1 \text{ if } p[\Theta = \Theta_1 | X] > p[\Theta = \Theta_2 | X]$$

$$d = \Theta_2 \text{ if } p[\Theta = \Theta_1 | X] < p[\Theta = \Theta_2 | X]$$

$$d = \Theta_1 \text{ or } \Theta_2 \text{ if } p[\Theta = \Theta_1 | X] = p[\Theta = \Theta_2 | X]$$

$p[\Theta = \Theta_1 | X]$

$$= f(x|\Theta_1).p(\Theta = \Theta_1) / [f(x|\Theta)p(\Theta = \Theta_1) + f(x|\Theta_2).p(\Theta = \Theta_2)]$$

$$= \Theta_1 \exp(-\Theta x)(1/3) / [\Theta \exp(-\Theta x) 1/3 + \Theta_2 \exp(-\Theta_2 x)(2/3)]$$

Similarly,

$p[\Theta = \Theta_2 | X]$

$$= \Theta_2 \exp(-\Theta_2 x).2/3 / [\Theta_1 \exp(-\Theta_1 x).1/3 + \exp(-\Theta_2 x) 2/3]$$

If $p[\Theta = \Theta_1 \mid X] > p[\Theta = \Theta_2 \mid X]$

$$\Theta_1 \exp[-\Theta_1 x](1/3) / [\Theta_1 \exp(-\Theta_1 x)(1/3) + \Theta_2 \exp[-\Theta_2 x](2/3)]$$

$$> \Theta_2 \exp(-\Theta_2 x)(2/3) / [\Theta_1 \exp(-\Theta_1 x)(1/3) + \Theta_2 \exp(-\Theta_2 x)(2/3)]$$

i.e. $\Theta_1 \exp(-\Theta_1 x)(1/3) > \Theta_2 \exp(-\Theta_2 x)(2/3)$

i.e. $\exp[(\Theta_2 - \Theta_1) x] > 2\, \Theta_2/\Theta_1$

$\exp(x) > 4$

$x > \log_e 4$

Thus, Accept H_o if $x > \log_e 4$

Accept H_1 if $x < \log_e 4$

Accept Θ_1 or Θ_2 if $x = \log_e 4$.

$$\delta(x) = \begin{cases} \text{Accept } H_o & \text{if } x > \log_e 4 \\ \text{Reject } H_o & \text{if otherwise} \end{cases}$$