CHAPTER - III

SUFFICIENCY OF EXPERIMENTS

3.0 Introduction :

Consider a statistical decision problem in which Θ is a parameter and Θ takes values in parameter space Ω with the decision space D, and non-negative loss function defined on space Ω x D. Let $r_0(\eta)$ be the Bayes risk when η is prior distribution of Θ. Let $\mathcal{F}$ be the class of random variables X and $\{f_x(\cdot|\omega), \omega \in \Omega\}$ be the family of p.d.f.s where $f_x(\cdot|\omega)$ is p.d.f. of X under Θ. We also assume that Θ takes finite values and $\Omega = \{\omega_1, \omega_2, \ldots \omega_k\}$, Σ be the set of all probability distributions η on Ω Where $\eta_i \geq 0$ $\eta = \{\eta_1, \eta_2, \ldots, \eta_k\}$ with $\eta_1 + \eta_2 + \ldots + \eta_k = 1$. We also know that risk function r_o is a non-negative concave function on the set Σ, (M.H. DeGroot (1970)). In the following, by an "uncertainty function", we mean a non-negative concave function defined on Σ, set of all possible distributions of Ω.

Let the problem of interest be to find a procedure (of choosing random variables) that minimizes $E(H[\eta(X_1, X_2, \ldots, X_n)])$ for given η, H, n and f (the class of random variables). Since $r_0(\eta)$ is non-negative concave function on Σ, one can take $r_0(\eta)$ itself as

uncertainty function H , that is $H(\eta) = r_0(\eta)$.

Another, choice of H be the Shannon's entropy

$$\text{i.e. } H(\eta) = -\sum_{i=1}^{k} \eta_i \log \eta_i. \qquad (3.0.1)$$

Since the process of selecting random variables is carried out in finite n stages. As there are only n observations the optimal procedure can be determined by backward induction. Let $X_1, X_2, \ldots X_{n-1}$ are selected and observed with their values $x_1, x_2, \ldots x_{n-1}$. Let η_{n-1} be the posterior distribution of Θ and $\eta_{n-1}(X_n)$ be the posterior distribution of Θ when n observations have been observed. The optimal choice of the random variable X_n is made through

$$E\left(H[\eta_{n-1}(X_n)]\right) = \inf_{X \in \mathcal{F}} E\left(H[\eta_{n-1}(x)]\right) \qquad (3.0.2)$$

Let, $H = H_0$ and define the function H_1 on set Σ as follows:

$$H_1(\phi) = \inf_{X \in \mathcal{F}} E\left(H_0[\phi(X)]\right) \text{for } \phi \in \Sigma \qquad (3.0.3)$$

Here, we consider $H_1(\phi)$ as the minimum value of the expected uncertainty when one observation remains to be taken.

In general, let $H_1, H_2, \ldots, H_n$ be functions defined on Σ, the recurrance relation is given as :

$$H_{j+1}(\phi) = \inf_{X \in \mathcal{F}} E\left(H_j[\phi(X)]\right) \quad \text{for } \phi \in \Sigma \qquad (3.0.4)$$

Here ϕ is the posterior distribution of Θ, when j observations remain to be taken and $H_j(\phi)$ is the minimum value of the expected terminal uncertainty. In particular, η is the prior distribution of Θ and $H_n(\eta)$ is the minimum value of $E\left(H[\eta(X_1, X_2, \ldots, X_n]\right)$, which can be attained by any sequential procedure. An optimal procedure of selecting random variable $X_1 \in \mathcal{F}$ at first stage is given as

$$H_n(\eta) = E\left(H_{n-1}[\eta(Xp]\right) \qquad (3.0.5)$$

At $(j+1)^{th}$ state, the values $X_1 = x_1, \ldots X_j = x_j$ have been observed and the posterior distribution of these j variables is computed. Then the optimal procedure of selecting a random variable $X_{j+1} \in \mathcal{F}$ is given as

$$H_{n-j}(\eta) = E\left(H_{n-j-1}\left[\eta_j(X_{j+1})\right]\right) \qquad (3.0.6)$$

In this way the optimal sequential procedure can be carried out. But actual computation may be very difficult.

Note that through out discussion, we consider the phrases 'performing experiment X' and 'taking observations on random variable X' have same meaning.

In this Chapter the Section 3.1 deals with sufficient experiments.

3.1 Sufficient Experiment :

In this section we shall consider the optimal sequential procedure for selecting n experiments. The observations are taken from the random variable $X \in \mathbb{F}$.

Theorem 3.1.1 : Suppose that there exists a random variable $X^* \in F$ such that for any distribution $\phi \in \Sigma$ and any random variable $X \in \mathbb{F}$,

$$E(H[\phi(X^*)]) \leq E(H[\phi(X)]) \qquad (3.1.1.)$$

Then, for any prior distribution η, a sequential procedure that minimizes the expected terminal uncertainty $E\,(H[\eta\,(X_1, X_2, \ldots X_n)]$ is to take all n observations on the random variable X^*.

Proof : From equations (3.1.1) and (3.02) we can say that the n^{th} observation must be taken on the random variable X^*, without considering which random variable is selected at

earlier stages and regardless of the observed values of random variable. Hence for any distribution, $\phi \in \Sigma$

$$H_1(\phi) = E\,(H[\phi\,(X^*)])$$

Suppose for any random variable $X \in \mathbb{F}$ and $\phi \in \Sigma$ it will be shown that

$$E\,(\,H_1[\phi(X^*)]\,) \leqslant E\,(\,H_1[\phi(X)]\,). \qquad (3.1.3)$$

We can say that all (n-1) observations are made on random variable X^* because of (3.1.1). When same argument is made on each stage, then the all n observations are carried on the random variable X^* by induction

$$\text{i.e.} \quad H_2(\phi) = E\,(\,H_1[\phi(X^*)]\,)$$

$$H_3(\phi) = E\,(\,H_2[\phi(X^*)]\,)$$

$$H_n(\phi) = E\,(\,H_{n-1}[\phi(X^*)]\,)$$

Now we will prove equation (3.1.3).

For any given distribution $\phi \in \Sigma$ Let $\phi_1^* = \phi(X^*)$ be the posterior distribution when single observation is made on the random variable X^*. Then, for any random variable $X \in \mathbb{F}$, then,

$$E(H[\phi(X,X^*)]) = E(E(H[\phi(X^*,X)] \mid X^*))$$

$$= E(E(H[\phi_1^*(X)] \mid \phi_1^*))$$

$$\geq E(E(H[\phi_1^*(X^*)] \mid \phi_1^*))$$

$$= E[H_1(\phi_1^*)] = E\{H_1[\phi(X^*)]\} \qquad (3.1.4)$$

Similarly, for any distribution $\phi \in \Sigma$ and for any random variable $X \in \mathcal{F}$, let $\phi_1 = \phi(X)$ be the posterior distribution when a observation is made on the random variable X.
We know that $\phi(X^*, X) = \phi(X, X^*)$ i.e. the posterior distribution will be same without considering order in which the observations are taken. Therefore,

$$E(H[\phi(X^*, X)]) = E(H[\phi(X, X^*)]) = E(E(H[\phi(X, X^*)] \mid X))$$

$$= E(E(H[\phi_1(X^*)] \mid \phi_1))$$

$$= E[H_1(\phi_1)] = E(H_1[\phi(X)]) \qquad (3.1.5)$$

From (3.1.4) and 3.1.5).

$$E(H_1[\phi(X)]) \geq E(H_1[\phi(X^*)])$$

Hence proved.

Now we will define the term sufficient experiment under condition (3.1.1).

Definition 3.1.1 Let X and Y are random variables or experiments in class $\mathbb{F}$. Let S_x and S_y are sample spaces corresponding variables X and Y respectively. If the experiment Y is sufficient for the experiment X, there exists a non-negative function h on the space $S_X \times S_Y$, for which the following conditions are satisfied.

$$f_X(x|\omega) = \int_{S_Y} h(X,Y)\, f_Y(y|\omega)\, d\mu(y) \quad \text{for } \omega \in \Omega, \; x \in S_X \qquad (3.1.6)$$

$$\int_{S_X} h(x,Y)\, d\mu(x) = 1 \qquad y \in S_Y \qquad (3.1.7)$$

$$0 < \int_{S_y} h(x,y)\, d\mu(y) < \infty \quad \text{for } x \in S_X \qquad (3.1.8)$$

Intuitively it is clear that, by observing Y instead of X no information is lost. i.e. observing Y is equivalent to observing X.

Let A be the set of all vectors $a=(a_1,;..,a_k)$ such that $a_i \geq 0$ $(i = 1,2...k)$. The set $\sum$ is the subset of set A with $\sum_{i=1}^{k} a_i = 1$. For any vectors $a \in A$, $b \in A$ we define

$a.b = \sum_{i=1}^{k} a_i b_i$. And for $a.b > 0$ we define $a \otimes b$ as

$$a \otimes b = (a_1 b_1, \ldots a_k b_k) \,/\, a.b. \qquad (3.1.9)$$

Where $a \otimes b \in \Sigma$

If $a.b = 0$ we take $a \otimes b$ as any arbitary vector Σ.

Now we will prove following lemma. It is used in next theorem (3.1.2).

Lemma:3.1.1 Let H be an uncertainty function, and let $\phi \in \Sigma$ any fixed vector. Let ν be defined on the set A by

$$\nu(a) = (\phi.a)\, H(\phi \otimes a) \quad \text{for } a \in A \qquad (3.1.10)$$

Then ν is a concave function on set A.

Proof : Let any vectors $a \in A$ and $b \in A$ and positive constants α and β such that $\alpha + \beta = 1$ It will be shown that

$$\nu(\alpha a + \beta b) \geq \alpha\, \nu(a) + \beta \nu(b) \qquad (3.1.11)$$

$$\nu(\alpha a + \beta b)$$

$$= [\ \phi.(\alpha a + \beta b)\ ]\ H[\phi \otimes (\alpha a + \beta b)]$$

$$= [\ \alpha(\phi.a)+\beta(\phi.b)\ ]H([\alpha\phi_1 a_1+\beta\phi_1 b_1,\ldots,\alpha\phi_k a_k+\beta\phi_k b_k]/[\alpha(\phi.a) + \beta(\phi.b)]\ )$$

$$= [\alpha(\phi.a)+\beta(\phi.b)]H\ [\alpha(\phi_1 a_1 \phi_2 a_2 \cdots \phi_k a_k) + \beta(\phi_1 b_1 \cdots \phi_k b_k)]\ /\ \alpha(\phi.a)+\beta(\phi.b)$$

$$= [\alpha(\phi.a)+\beta(\phi.b)]H\{\alpha(\phi.a)(\phi\otimes a)\ /\ [\alpha(\phi.a) + \beta(\phi.b)] + \beta(\phi.b)(\phi\otimes b)\ /\ [\alpha(\phi.a)+\beta(\phi.b)]\ \}$$

$$= [\alpha(\phi.a)+\beta(\phi.b)]H[\alpha^*(\phi\otimes a) + \alpha^*(\phi\otimes b)] \qquad (3.1.12)$$

Where,

$$\alpha^* = \alpha(\phi.a)\ /\ [\alpha(\phi.a) + \beta(\phi.b)],\quad \beta^* = 1 - \alpha^* \qquad (3.1.13)$$

Here, we let $\phi.a > 0$, $\phi.b > 0$

Since H is concave function from equation (3.1.12) and (3.1.13) follow the relation.

$$\nu(\alpha a+\beta b) \geq [\alpha(\emptyset.a)+\beta(\emptyset.b)][\alpha^* H(\emptyset \otimes a) + \beta^* H(\emptyset \otimes b)]$$

$$= \alpha\,\alpha^*(\emptyset.a)\, H(\emptyset \otimes a) + \alpha\beta^*(\emptyset.a)\, H(\emptyset \otimes b) +$$

$$+ \alpha^*\beta(\emptyset.b)(\emptyset \otimes a) + \beta\beta^*(\emptyset.b)(\emptyset \otimes b)$$

$$= \alpha\,(1-\beta^*)\,(\emptyset.a)\, H(\emptyset \otimes a) + \alpha\beta^*(\emptyset.a)\, H\,(\emptyset \otimes b) +$$

$$+ \alpha^*\beta\,(\emptyset.b)(\emptyset \otimes a) + \beta(1-\alpha^*)(\emptyset.b)\,(\emptyset \otimes b)$$

$$= [\alpha\,(\emptyset.a)\, H(\emptyset \otimes a) + \beta(\emptyset.b)(\emptyset \otimes b)] - \alpha\beta^*(\emptyset.a)H(\emptyset \otimes a) +$$

$$+ \alpha\beta^*(\emptyset.a)\, H(\emptyset \otimes b) + \alpha^*\beta(\emptyset.b)\, H\,(\emptyset \otimes a) -$$

$$- \alpha^*\beta(\emptyset.b)\, H(\emptyset \otimes b)$$

$$= [\alpha(\emptyset.a)\, H(\emptyset \otimes a) + \beta(\emptyset.b)\,(\emptyset \otimes b) +$$

$$+ \alpha\beta^*(\emptyset.a)[H(\emptyset \otimes b) - H(\emptyset \otimes a)] -$$

$$- \alpha^*\beta(\emptyset.b)[\, H(\emptyset \otimes b) - H(\emptyset \otimes a)]$$

$$= [\alpha(\emptyset.a)\, H(\emptyset \otimes a) + \beta(\emptyset.b)\,(\emptyset \otimes b)] +$$

$$+ [H(\emptyset \otimes b) - H(\emptyset \otimes a)].[\alpha\beta^*(\emptyset.a)-\alpha^*\beta\,(\emptyset.b)]$$

$$= [\alpha(\emptyset.a)H(\emptyset \otimes a)+ \beta(\emptyset.b)(\emptyset \otimes b)] +$$

$$+ [H(\emptyset \otimes b)-H(\emptyset \otimes a)].[\beta^*\alpha^* c - \alpha^*\beta^*\, c]$$

$$= [\ \alpha(\emptyset.a)H(\emptyset \otimes a) + \beta(\emptyset.b)\,(\emptyset \otimes b)\ ]$$

From (3.1.13) $C\,\alpha^* = \alpha(\emptyset.a)$

$C\,\beta^* = \alpha(\emptyset.b)$

Where $C = \alpha(\emptyset.a) + \beta(\emptyset.b)$

Hence relation (3.1.11) is satisfied.

If both $\phi.a = 0$ and $\phi.b = 0$ each side of relation vanishes.

Theorem 3.1.2 : Suppose that the experiment Y is sufficient for the experiment X, when X and Y belongs to $\mathcal{F}$. Then, for any uncertainty function H and any distribution $\phi \in \Sigma$,

$$E\{H[\phi(X)]\} \geq E\{H[\phi(Y)]\} \qquad (3.1.14)$$

Proof : Any non-negative function g(.) is defined on parameter space Ω can be considered as a vector $(g_1, g_2, \ldots, g_k)$ in the set A. Let $g_i = g(\omega_i)$ for $i = 1,2,\ldots k$. We will use this convention in the proof.

Here, experiment Y is sufficient for experiment X. Then, there exists a non-negative function h on the space $S_X \times S_Y$ which satisfies the conditions (3.1.6), (3.1.7) and (3.1.8). For every $x \in S_X$ we shall define the function $\psi(\omega_i | x)$ for every point of Ω as follows :

$$\psi(\omega_i | x) = \left[\int_{S_Y} h(x,Y)\, f_{Y|\omega_i}(y|\omega_i)\, d\mu(y)\right] / \int_{S_Y} h(x\ Y)\, d\mu(y) \qquad (3.1.15)$$

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

From equation (3.1.14) and $\phi(X)$ is the posterior distribution. It can be shown that the following relation must be satisfied.

$$E(H[\phi(X)]) = \int_{S_X} \nu\,[f_X(x|\omega_i)]\, du\,(x). \qquad (3.1.16)$$

From the equations (3.1.6) and (3.1.15)

$$E\,(H[\phi(X)]) = \int_{S_X} \nu[\psi(\omega_i|x)] \left[\int_{S_Y} h(x,y)\, du(y)\right] du(x) \qquad (3.1.17)$$

In equation (3.1.17) we used the fact that

$\nu(\alpha a) = \alpha\nu(a)$, $\alpha \geq 0$ for all vectors $a \in A$.

For every point $x \in S_X$, let $f^*(y|x)$ denote the p.d.f. on the sample S_Y defined as follows :

$$f^*(y|x) = h(x,y) \,/\, \left[\int_{S_Y} h(x,y)\,du(y)\right] \text{ for } y \in S_Y \qquad (3.1.18)$$

From equation (3.1.15) $i = 1,2,\ldots,k$ we can say that $\psi(\omega_i|x)$ is nothing but the expection of $f_Y(y|\omega_i)$ when the random variable Y has p.d.f. $f^*(y|x)$. By Lemma 3.1.1 ν is a concave function on set A. Therefore, using Jensen's inequality for a concace function of a k dimentional random vector, we get following :

$$\nu\,[\psi(\omega_i|x)\,] \geq \int_{S_Y} \nu\,[f_Y\,(y|\omega_i)]\, f^*(y|x)\, du(y) \qquad (3.1.19)$$

Integrating both sides of equation (3.1.19) and using equation (3.1.17),

With respect to X, we get

$$\int_{S_X} \nu[\Psi(\omega_i|\, x)].[\int_{S_Y} h(x,y)\, d\mu(y)]\, d\mu(x)$$

$$\geq \int_{S_X} \int_{S_Y} \nu[f_Y(y|\omega_i)]h(x,y).\int_{S_Y} h(x,y))d\mu(y)/[\int_{S_Y} h(x,y)d\mu(y)]$$

i.e.

$$E\,(H[\phi(X)] \geq \int_{S_X} \int_{S_Y} \nu[\phi_Y(y|\omega_i)]h(x,y)\, d\mu(x)d\mu(y) \qquad (3.1.20)$$

Reversing the order of integration, we get,

$$E\,(H[\phi(X)] \geq \int_{S_Y} \int_{S_X} \nu[f_Y(y|\omega_i)]h(x,y)\, d\mu(x)\, d\mu(y)$$

$$E\,(H[\phi(X)]) \geq \int_{S_Y} \nu\,[\phi_Y(y|\omega_i)]\, d\mu(y) \int_{S_X} h(x,y)\, d\mu(x) \quad .$$

By equation (3.1.7) we get,

$$E\,(H[\phi(X)]) \geq E\,(H\,[\phi(Y)]) \qquad (3.1.21)$$

Hence proof :

The final result, which says that performing n replications of a sufficient experiment must be an optimal procedure.

Corollary 3.1.1 :

Corollary 3.1.1 : Suppose that there exists a random variable $X^* \in \mathbb{F}$ which is sufficient for any other random variable $X \in \mathbb{F}$. Then, for any uncertainty function H on Σ for any prior distribution $\eta \in \Sigma$, and for any positive integer n, the sequential procedure that minimizes the expected terminal uncertainty

$E\{H[\eta(X_1, X_2, \ldots, X_n)]\}$ is to take all n observations on the random variable X^*.