SELF LEARNING MATERIAL

Decision Theory

PGSTAT-09 / MASTAT-09 (Old)

PGSTAT—13 / MASTAT- 13 (New)

Syllabus:

Block 1: Basic Elements and Bayes Rules

Unit 1: Basic Elements

Decision theoretic problem as a game, basic elements, optimal decision rules, unbiased ness, invariance, ordering.

Unit 2: Bayes and Minimax Rules

Bayes and minimax principles, generalized. Bayes rules, extended Bayes rules,

Block 2: Optimality of Decision Rules

Unit 3: Admissibility and Completeness

Admissibility, completeness, minimal complete class, separating and supporting hyper plane theorems,

Unit 4: Minimaxity and Multiple Decision Problems

Minimax theorem, complete class theorem, equalizer rules, examples, multiple decision problems.

BLOCK 1: BASIC ELEMENTS AND BAYES RULES

Unit 1: Basic Elements

Unit 2: Bayes and Minimax Rules

Suppose, you want to buy a new mobile phone. How do you decide which one is best for you and where to buy it? That is a decision problem. Now suppose you have, anyhow finalized the mobile you are willing to have. Then, **Decision theory** is the study of the reasoning underlying this decision. It is closely related to the well-known theory of games. In this chapter, firstly a decision problem has been explained as a game problem. Then it is explained from the perspective of a statistician. Various elements/components along with some other topics of importance have also been defined in this section. Next this chapter is focused on Bayes and minimax criteria and their description.

1. **Game Theory and Decision Theory**:

Basic Elements: the elements of decision theory are similar to those of the theory of games. In particular, decision theory may be considered as the theory of two-person game, in which nature takes the role of one of the players. The so-called normal form of a zero-sum two-person game, henceforth to be referred to as a *game,* consists of there basic elements:

1. A non empty set, Θ, of possible states of nature, sometimes referred to as the parameter space.
2. A non empty set, a, of action available to the statistician.
3. A loss function, L (θ, a), a real-valued function defined on Θ X a.

A game in mathematical sense is just such a triplet (Θ, a, L), and any such triplet defines a game, which is interpreted as follows.

Nature choose a point θ in Θ, and the statistician, without being informed of the choice nature has made, chooses an action a in a. as a consequence of these two choices, the statistician loses an amount L (θ, a).[the function L (θ, a) may take negative values. A negative loss may be interpreted as a gain, but throughout this book L (θ, a) represented the loss to the statistician if he takes action a when θ is the " *true state of nature*".] Simple through this definition may be, its scope is quite broad, as the following example illustrated.

<u>Example</u> 1.1: <u>Odd or even</u>: two contestants simultaneously put up either one or two fingers. One of the players, call him player I, wins if the sum of the digits showing is odd, and the other player, player II, wins if the sum of the digits showing is even. The winner in all cases receives in dollars the sum of the digits showing, this being paid to him by the loser.

To create a triplet (Θ, a, L), out of this game we give player I the label "nature" and the player II the label "statistician". Each of these players has two possible choices, so that Θ= {1, 2} =a, in which "1" and "2" stands for the decision to put up one and two fingers, respectively. The loss function is given by the table 1.1.

Thus L (1, 1) =-2

<u>Table 1.1</u>

	a	1	2
θ	1	-2	3
	2	3	-4

L (1, 2) =3, L (2, 1) =3 and L (2, 2) =-4 it is quite clear that this is a game in the sense described in the first paragraph. This example is discussed later, in which it is shown that one of the players has a distinct advantage over the other. Can you tell which one it is? Which player would you rather be?

<u>Example</u> 1.2: Consider the game (Θ, a, L) in which Θ (θ_1, θ_2), a =(a_1, a_2) and the loss function L is given by the table 1.2:

(Table 1.2)

"statistician"

		a_1	a_2
"nature"	θ_1	4	1
	θ_2	-3	0

In game theory, in which the player choosing a point from Θ is assumed to me intelligent and his winnings in the game are given by the function L (loss function of the statistician or gain function of the nature), the only "rational" choice for him is θ_1. No matter what his opponent does, he will gain more if he chooses θ_1 than if he chooses θ_2.thus it is clear that the statistician should choose action a_2 instead of actiona_1, for he will lose only one instead of four. This is the only reasonable things for him to do.

Now, suppose that the function L does not reflect the winning of nature or that nature chooses a state without any clear objective in mind. Then we can no longer state categorically that the statistician should choose action $a_2$if nature happens to chooses θ_2, the statistician will prefer take action a_1.

2. <u>Decision Function: Risk Function</u>:

To give a mathematical structure to this process of information gathering, we suppose that statistician before making a decision is allowed to look at the observed value of a random variable or vector, X, whose distribution depends on the true state of nature, θ. The sample space denoted as $\mathfrak{X}$is taken to be (a Borel subset of) a finite dimensional Euclidean space, and the probability distributions of X are supposed to be defined on the Borel subsets, β of $\mathfrak{X}$. thus for each θ ϵ Θ there is a probability measure $P_\theta$defined on β, a corresponding cumulative distribution function $F_X(x/\theta)$ which represents the distribution function of X when θ is the true state of the nature (the parameter)

A statistical decision problem or a statistical game is a game (Θ, a, L) coupled with an experiment involving a random variable X whose distribution $P_\theta$depends on the state θ∈Θ chosen by nature.

On the basis of the outcome of the experiment X=x (x is the observed value of X), the statistician chooses an action d(x)ϵ a .such a function d, which maps the sample space $\mathfrak{X}$ in to a, is an elementary strategy for the statistician in this situation .The loss is now the random quantity L (θ, d(x)).The expected value of L (θ, d(x)) when θ is the true state of nature is called the risk function.

$$R(\theta,d) \quad E\{L(\theta,d(x))\} \text{......................} \quad (2.1)$$

and represented the average loss to the statistician when the true state of nature θ and the statistician used the function d.

Defn. 2.1: Any function d(x) that maps the sample space 𝔛 in to a, is called a non-randomized decision rule or a non-randomized decision function, provided the risk function R (θ, d) exists and is finite for all θ∈Θ. The class of all non-randomized decision rules is denoted by D.

$$R(\theta,d) \quad E_\theta L(\theta,d(x)) \quad \int L(\theta,d(x))dP_\theta\,(x) \text{..............} \quad (2.2)$$

With such an understanding, D consists of those functions d for which $L(\theta,d(x))$ is for each θ∈Θ a Lebesgue integrable function of x. In particular, D contains all simple functions. On the other hand, the expectation in (2.2) may be taken as the Riemann or the Riemann-Stieltjes integral.

$$R(\theta,d) \quad E_\theta L(\theta,d(x)) \quad \int L\left(\theta,d(x_j)\right)dF_x\,(x/\theta) \text{........} \quad (2.2)$$

In that case D would contain only functions d for which $L(\theta,d(x))$ is for each θ∈Θ continuous on a set of probability one under $F_x(x/\theta)$.

Example 2.1: the game of "odd or even" may be extended to a statistical decision problem. Suppose that before the game is played the player called "the statistician" is allowed to ask the player called "nature" how many fingers he intends to put up and that nature must answer truthfully with probability 3/4. The statistician therefore observes a random variable X (the answer nature gives) taking the value 1 or 2. If θ= 1 is the true state of nature, $P_{\theta=1}^{[X=1]} \quad \frac{3}{4} \quad 1-P_{\theta=1}^{[X=2]}$. Similarly $P_{\theta=2}^{[X=1]} \quad 1/4 = 1-P_{\theta=2}^{[X=2]}$. There are exactly four possible functions from 𝔛 ={1,2} in to, a={1,2}. There are the four decision rules,

$$d_1(1) \quad 1 \qquad d_1(2) \quad 1 \;;$$
$$d_2(1) \quad 1 \qquad d_2(2) \quad 2 \;;$$
$$d_3(1) \quad 2 \qquad d_3(2) \quad 1 \;;$$
$$d_4(1) \quad 2 \qquad d_4(2) \quad 2 \;.$$

Rules d_1 and d_4 ignore the value of X , rule d_2 reflects the belief of the statistician that the nature is telling the truth, and rule d_3, that nature is not telling the truth. The risk table (2.1) is given as:

(Table 2.1)

D

Θ		d_1	d_2	d_3	d_4
	1	-2	$-3/4$	$7/4$	3
	2	3	$-9/4$	$5/4$	-4

$R(\theta, d)$

It is a custom, which we steadfastly observe, that the choice of a decision function should depend only on the risk function $R(\theta, d)$ and not other wise on the distribution of the random variable $L(\theta, d(X))$.

Notice that the original game (Θ, a, L) has been replaced by a new game (Θ, D, R), in which the space D and the function R have an underlying structure, depending on a, L, and the distribution of X, whose expectation must be the main objective of decision theory.

A "classical" mathematical statistics consists three important categories:

1. *a Consists of two points,a* $\{a_1, a_2\}$: decision theoretic problems in which a consists of exactly two points are called *problem of testing hypothesis*.

Consider the special case in which Θ is the real line and suppose that the loss function for some fixed number θ_0 given by the formulas:

$$L(\theta, a_1) \quad \begin{cases} l_1 & if\ \theta > \theta_0 \\ 0 & if\ \theta \leq \theta_0 \end{cases}$$ And

$$L(\theta, a_2) \quad \begin{cases} 0 & if\ \theta > \theta_0 \\ l_2 & if\ \theta \leq \theta_0 \end{cases}$$

Where l_1 and l_2 are positive numbers. Here we would like to take action $a_1\ if\ \theta \leq \theta_0$ and action $a_2\ if\ \theta > \theta_0$.the space D of decision rule consists of those functions d from the sample space in $\{a_1, a_2\}$ with the property that $P_\theta[d(x) \quad a_1]$ is well-defined for all values of θeΘ. The risk function in this case is ,

$$R(\theta, d) \quad EL(\theta, d(x))$$

$$l_1 P_\theta[d(x) \quad a_1] \qquad if\ \theta > \theta_0$$

$$l_2 P_\theta[d(x) \quad a_2] \qquad if\ \theta \leq \theta_0$$

In this case probabilities of making two types of error are involved. For $\theta > \theta_0$, $P_\theta[d(x) \quad a_1]$ is the probability of making the error of taking action a_1 when we should take action a_2 and θ is the true state of nature. Similarly, for $\leq \theta_0 \quad P_\theta[d(x) \quad a_2] \quad 1 - P_\theta[d(x) \quad a_1]$, is the probability of making the error of taking action a_2 when we should take action a_1 and θ is the true state of nature.

2. *a Consists of k points*, $\{a_1, a_2, \dots \dots a_k\}, k \geq 3$. these decision theoretic problems are called ***multiple decision problems***. For an example an experimenter is to judge which of treatments has a greater yield on the basis of an experiment.

He may (a) decide treatment 1 is better, (b) decide treatment 2 is better, or (c) withhold judgment until more data are available. In this exp. k=3

3. *a Consists of a real line, a* $\quad (-\infty, \infty)$.

such decision theoretic problems are referred to in a board sense as ***point estimation of a real parameter.*** Consider the special case in which Θ is also a real line and suppose that the loss function is given by the formula,

$$L(\theta, a) \quad c(\theta - a)^2 \ ,$$

Where, c is some positive constant. A decision function d, in this case a real-valued function defined on a sample space, may be considered as an "estimate" of the true unknown state of nature θ. It is the statistician desire to choose the function d to minimize the risk function.

$$R(\theta,d) \quad EL(\theta,d(x))$$

$$= cE_\theta(\theta - d(x))^2 ,$$

The criterion arrived here is that of choosing an estimate with a small mean squared error in some sense.

3. **Randomization:**

It is often useful to recognize explicitly that in any decision problem, the statistician may wish to choose a decision from D by means of an auxiliary randomization procedure of some short, such as by tossing a coin. In other words the statistician may wish to make a mixed or randomized decision δ by assigning probabilities $p_1, p_2, \ldots\ldots$ to the elements $d_1, d_2, \ldots\ldots$ of decisions from D and then one of the decisions δ on the basis of these probabilities is chosen.

More generally, a randomized decision for the statistician in a game (Θ, a, L) is a probability distribution over a (it is understood that a fixed σ-field of subsets of a containing the individual points of a is given). If P is probability distribution over a and Z is a random variable taking values is a.whose distribution is given by P, the expected or average loss in the use of randomized decision P is,

$$L(\theta,P) \quad E\,L(\theta,Z) \;\ldots\ldots\ldots\ldots\ldots\ldots\; (3.1)$$

Provided it exists. This formula is to be regarded as an extension of the domain of definition of the function $L(\theta,\cdot)$ from a to the sample space of randomized decisions, for each element a ϵ a may, and shall, be regarded as the probability distribution degenerate at a ,that is, the distribution giving probability one to point a. *the space of randomized decisions, P, for which* $L(\theta,P)$ *exists and is finite for all* $\theta \epsilon \Theta$ is denoted by a^* .

With this definition, the game (Θ, a^*, L) is to be considered as the game (Θ, a, L) in which the statistician is allowed randomization. a^*contains all the probability distributions giving mass one to a finite number of points of a.

By analogy, we may extend the game (Θ, D, R) to (Θ, D^*, R) where D^* is a space containing probability distribution over D. if δ denotes a probability distribution over D, R(θ ,δ) is defined analogously to (3.1) as,

$$R(\theta, \delta) \quad E\,R(\theta, Z) \text{ .... } (3.2)$$

Where Z is a rv taking values in D, whose distribution is given by δ.

Defn. 3.1: Any probability distribution δ on the space of non-randomized function, D, is called a randomized decision function or a randomized decision rule, provided the risk function (3.2) exists and is finite for all $\theta \epsilon \Theta$. The space of all randomized decision rule is denoted by D*. D* contains all the probability distributions giving mass one to a finite number of point of D.

The space D of non-randomized decision rules may, and shall, be considered as a subset of the space D* of randomized decision rules D ϵ D* by identifying a point d ϵ D with the probability distribution δ ϵ D* degenerate at point d.

One advantage in the extension of the definition of L (θ, $\cdot$) from a *to* a^* and the definition of R (θ, $\cdot$) from D to D* is that these functions become linear on a^* and D*, respectively. In other words, if $P_1 \epsilon a^*$, $P_2 \epsilon a^*$ and $0 \leq \alpha \leq 1$.

$$P \quad \alpha P_1 + (1-\alpha) P_2 \epsilon a^* \text{ and } L(\theta, \alpha P_1 + \overline{1-\alpha} P_2) \quad L(\theta, P) \quad E\,L(\theta, Z)$$

$$\alpha L(\theta, P_1) + (1-\alpha) L(\theta, P_2) \text{ .... } (3.3)$$

Similarly, if $\delta_1 \epsilon a^*$, $\delta_2 \epsilon a^*$ and $0 \leq \alpha \leq 1$.then

$$\delta \quad \alpha \delta_1 + (1-\alpha)\delta_2 \epsilon D^*$$

$$R(\theta, \delta) \quad E\,R(\theta, Z) \quad \alpha R(\theta, \delta_1) + (1-\alpha) R(\theta, \delta_2) \text{ .... } (3.4)$$

Example 3.1: Let the game be defined as,

		a_1	a_2	a_3
Θ	θ_1	4	1	3
	θ_2	1	4	3

If nature choosesθ_1, action a_3 is preferable to action a_1.if, on the other hand, nature choosesθ_2, action a_3 is preferable to action a_2. thus a_3 is preferred to either of the other action under the proper circumstances. However, suppose the statistician flips a fair coin to choose between actions a_1 and a_2; that is suppose the statistician's decision is to choose $a_1$if the coin comes up heads and choose $a_2$if the coin comes up tails. This decision, denoted by δ, is a *randomized decision*; such decisions allow the actual choice of the action in a to be left to a random mechanism and the statistician chooses only the probabilities of the various outcomes. In game theory δ would be called a *mixed strategy*. The randomized decision δ chooses action a_1 with probability ½, action a_2 with probability ½, action a_3 with probability zero. The expected loss in the use of δ is given by,

$$L(\theta,P) \quad E\,L(\theta,Z) \quad 1/2L(\theta,a_1)+1/2L(\theta,a_2)+0L(\theta,a_3)$$

$$\frac{1}{2}.4+\frac{1}{2}.1+0.3 \quad \frac{5}{2} \qquad if\theta \quad \theta_1$$

$$=\frac{1}{2}+\frac{4}{2}.1+0.3 \quad \frac{5}{2} \qquad if\theta \quad \theta_2$$

Because it is understood that the choice between strategies is to be made on the basis of expected loss only,δ is certainly to be preferred toa_3 for no matter what the true state of nature, the expected loss is smaller if we use δ than if we usea_3.

$$P_1 \quad \left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right), \qquad P_2 \quad \left(\frac{3}{8},\frac{5}{8},0\right)$$

$$L(\theta,P_1) \quad \frac{4}{4}+\frac{1}{2}+\frac{3}{4} \quad \frac{9}{4} \qquad if\theta \quad \theta_1$$

$$\frac{1}{4}+\frac{4}{2}+\frac{3}{4} \quad \frac{12}{4} \qquad if\theta \quad \theta_2$$

$$L(\theta,P_2) \quad \frac{3}{8}.4+\frac{5}{8}.1+0.3 \quad \frac{17}{8} \quad if\theta \quad \theta_1$$

$$\frac{3}{8}+\frac{5}{8}.4+0.3 \quad \frac{23}{8} \qquad if\theta \quad \theta_2$$

"If randomized decisions are allowed and the choice between strategies is based on expected loss only, the statistician should never take actiona_3."

<u>Optimal Decision Rules</u>: The fact that a best rule usually does not exist, two general methods, which have been proposed for arriving at a decision rule, are frequently satisfactory.

<u>Method I</u>: <u>Restricting the Available Rules</u>:

<u>Unbiasedness</u>: suppose the problem is such that for each θ there exist a unique correct decision and that each decision is correct for some θ. Assume further that $L(\theta_1, d) = L(\theta_2, d)$ for all d wherever the some decision is correct for both$\theta_1 and\ \theta_2$. Then the loss$L(\theta, d')$ depends only the actual decision taken, say d' and the correct decision d. thus the loss can be denoted by L(d, d') and this function measures how for a past d and d' are. Under these assumptions a decision function δ(x) is said to be unbiased w.r.t. the loss L if for all θ and d' $E_\theta L(d', \delta(x)) \geq E_\theta L(d, \delta(x))$ .... (3.5)

Where the subscript θ contains the distribution w.r.t. which the expectation can taken and where d is the correct decision for θ. Thus δ is unbiased if on the average δ(x) closer to the correct decision than to any wrong one. Extending this definition, δ is said to be L-unbiased for an arbitrary decision problem for all θ and θ'.

$$E_\theta L(\theta', \delta(x)) \geq E_\theta L(\theta, \delta(x)) \quad \text{............... (3.6)}$$

<u>Example</u> 3.2: In two decision problem, let ω_0 and $\omega_1$be the set of θ values for which $d_0$and d_1 are correct decisions. Assume that

$$L(\theta, d_0) \quad 0 \quad \theta\epsilon\omega_0 \qquad\qquad L(\theta, d_1) \quad b \quad \theta\epsilon\omega_0$$
$$= a \quad \theta\epsilon\omega_1 \qquad\qquad = 0 \quad \theta\epsilon\omega_1$$

$$E_\theta L(\theta', \delta(x)) \quad L(\theta', d_0)P_\theta[\delta(x) \quad d_0] + L(\theta', d_1)P_\theta[\delta(x) \quad d_1]$$

$$aP_\theta[\delta(x) \quad d_0] \qquad if\ \theta'\epsilon\omega_1$$

$$bP_\theta[\delta(x) \quad d_1] \qquad if\ \theta'\epsilon\omega_0$$

So that (3.6) reduced to

$$aP_\theta[\delta(x) \quad d_0] \geq bP_\theta[\delta(x) \quad d_1] \qquad for\ \ \theta\epsilon\omega_0$$

With reverse inequality holding for $\theta \epsilon \omega_1$

Since $P_\theta[\delta(x) \quad d_0] + P_\theta[\delta(x) \quad d_1] \quad 1$ the unbiasedness contains (3.6) reduces to, $P_\theta[\delta(x) \quad d_1] \leq \frac{a}{a+b} \quad for \;\; \theta \epsilon \omega_0$

And $\qquad P_\theta[\delta(x) \quad d_1] \geq \frac{a}{a+b} \quad for \;\; \theta \epsilon \omega_1$

<u>Example</u> 3.3: In the problem of estimating the real valued function $g(\theta)$ with square of the error as loss, the condition of unbiasedness become,

$E_\theta[\delta(x) - g(\theta')]^2 \geq E_\theta[\delta(x) - g(\theta)]^2$ For all θ and θ'............ (3.7)

$$E_\theta[\delta(x) + E_{\theta'}\delta(x) - E_{\theta'}\delta(x) - g(\theta')]^2 \geq E_\theta[\delta(x) + E_\theta\delta(x) - E_\theta\delta(x) - g(\theta')]^2$$

Let $E_\theta\delta(x) \quad h(\theta)$

$$E_\theta[\delta(x) - h(\theta) + h(\theta) - g(\theta')]^2 \geq E_\theta[\delta(x) - h(\theta) + h(\theta) - g(\theta)]^2$$

$[h(\theta) - g(\theta')]^2 \geq [h(\theta) - g(\theta)]^2 \qquad$ For all θ and θ'

If $g(\theta)$ is continuous over Ω and which is not continuous in any open subset of Ω, and that $h(\theta) = E_\theta\delta(x)$ is continuous function of θ for each estimate $\delta(x)$ of $g(\theta)$. Thus (3.2) reduces to,

$$g^2(\theta') - 2h(\theta)g(\theta) \geq g^2(\theta) - 2h(\theta)g(\theta)$$

Or $$g^2(\theta') - g^2(\theta) \geq 2h(\theta)\big(g(\theta') - g(\theta)\big)$$

$$[g(\theta) - g(\theta')][g(\theta') + g(\theta)] \geq 2h(\theta)[g(\theta') - g(\theta)]$$

If θ is neither a relative minimum or maximum of $g(\theta)$ it follows that there exist points θ' arbitrary chosen θ both such that,

$g(\theta') + g(\theta) \leq 2h(\theta)$ Hence $g(\theta) \quad h(\theta)$

Thus $\delta(x)$ is unbiased if$E_\theta\delta(x) = g(\theta)$. **<u>Proved</u>**

<u>Method II</u>: <u>Ordering the Decision Rule</u>: two important and useful principles are basic to the study of decision theory.

1. **<u>Bayes principle</u>:** The Bayes principle involves the notion of a distribution on the parameter space Θ called a <u>prior distribution</u>. Two things are needed of a prior distribution τ on Θ. First we may able to speak of the Bayes risk of a decision rule δ w.r.t. a prior distribution τ, namely

$$\square(\tau,\delta) \quad E\,R(T,\delta) \text{ .... } (3.8)$$

Where T is a r.v. over Θ having distribution τ. Second we need to be able to speak of the joint distribution T and X and of the conditional distribution of T, given X, the latter being called the posterior distribution of the parameter given the observations. We denote the space of prior distribution as θ^*.

<u>Defn.</u> 3.2: A decision rule δ_0 is said to be Bayes w.r.t. the prior distribution τ e θ^* if $\square(\tau,\delta_0) \quad \underset{\delta\epsilon\, D^*}{inf}\square(\tau,\delta)$ .... (3.9)

The value on the R.H.S. is known as the minimum Bayes risk. Bayes risk may not exist even if the minimum Bayes risk is defined and finite.

<u>Defn.</u> 3.3: Let $\epsilon > 0$. A decision rule δ_0 is said to be $\epsilon -$ Bayes w.r.t. the prior distribution τεθ^* if

$$\square(\tau,\delta_0) \leq \underset{\delta\epsilon D^*}{inf}\square(\tau,\delta) + \epsilon \text{ .... } (3.10)$$

2. **<u>Minimax principle</u>:** An essentially different type of ordering of the decision rule may be obtained by ordering the rules according to the worst that could happen to the statistician. In other words, a rule δ_1 is preferred to a rule δ_2 if

$$\underset{\theta}{sup} R(\theta,\delta_1) < \underset{\theta}{sup} R(\theta,\delta_2)$$

A rule that is most preferred in this ordering is called a minimax decision rule.

<u>Defn.</u> 3.4: A decision rule δ_0 is said to be minimax if

$$\underset{\theta\epsilon\Theta}{sup} R(\theta,\delta_0) \quad \underset{\delta\epsilon D^*}{inf}\,\underset{\theta\epsilon\Theta}{sup} R(\theta,\delta) \text{ .... } (3.11)$$

The value on the R.H.S. of (3.11) is called the minimax value or upper value of the game.

<u>Proposition</u> 3.1: A decision rule δ_0 is said to be minimax if and only if

$$R(\theta', \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta) \text{ .... } (3.12)$$

For all θ′εΘ and δεD^*

<u>Proof:</u> let $R(\theta', \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta)$ For all θ′εΘ and δεD^*

$$\sup_{\theta' \varepsilon \Theta} R(\theta', \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta) \quad \text{for } \delta \varepsilon D^*$$

Hence δ_0 minimizes the $\sup_{\theta \varepsilon \Theta} R(\theta, \delta)$ forδεD^*

Thus, $\sup_{\theta' \varepsilon \Theta} R(\theta', \delta_0) \quad \inf_{\delta \varepsilon D^*} \sup_{\theta \varepsilon \Theta} R(\theta, \delta)$ And δ_0 is minimax.

Conversely, let $\sup_{\theta \varepsilon \Theta} R(\theta, \delta_0) \quad \inf_{\delta \varepsilon D^*} \sup_{\theta \varepsilon \Theta} R(\theta, \delta)$

$$\Rightarrow \sup_{\theta \varepsilon \Theta} R(\theta, \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta) \quad \text{for } \delta \varepsilon D^*$$

$$\Rightarrow R(\theta', \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta_0) \leq \sup_{\theta \varepsilon \Theta} R(\theta, \delta) \quad \text{for all } \theta' \varepsilon \Theta, \delta \varepsilon D^*$$

<u>Proved</u>

<u>Defn.</u> 3.5: Let ϵ> 0. A decision rule δ_0 is said to be ϵ - minimax if

$$\sup_{\theta \varepsilon \Theta} R(\theta, \delta_0) \leq \inf_{\delta} \sup_{\theta} R(\theta, \delta) + \epsilon \text{ .... } (3.13)$$

More simply, δ_0 is ϵ-minimax if for all θ′εΘ and δεD^*

$$R(\theta', \delta_0) \leq \sup_{\theta} R(\theta, \delta) + \epsilon \text{ .... } (3.14)$$

<u>Defn.</u> 3.6: A distribution $\tau_0 \varepsilon\, \theta^*$ is said to be *least favorable* if

$$\inf_{\delta} \gamma(\tau_0, \delta) \quad \sup_{\tau} \inf_{\delta} \gamma(\tau, \delta) \text{ .... } (3.15)$$

The value on the R.H.S. of (3.15) is called the maximin value or lower value of the game.

Geometrical Interpretation for finite Θ: we give a geometric interpretation of the fundamental problem of decision theory in the case in which the parameter space Θ is finite.

Suppose that Θ contains k points, $\Theta \quad \{\theta_1, \theta_2, \dots\dots, \theta_k\}$ and consider the set S, to be called the *risk set*, contained in k-dimensional Euclidian space E_k of points of the form $\left(R(\theta_1, \delta), R(\theta_2, \delta), \dots\dots\dots\dots, R(\theta_k, \delta)\right)$, where δ ranges through D^*

$$S \quad \left\{(y_1, y_2, \dots\dots, y_k) for\ some\ \delta \epsilon D^*, y_j \quad R(\theta_j, \delta)\ for\ j \quad 1,2, \dots, k\right\}$$

.... (3.16)

If k=2 this set may easily be plotted in the plane.

Defn. **3.7:** A set S should be convex if when ever $y \quad (y_1, y_2, \dots\dots, y_k)$ $y' \quad (y'_1, y'_2, \dots\dots, y'_k)$ are elements of S, the point

$\alpha y + \overline{1-\alpha} y' \quad (\alpha y_1 + \overline{1-\alpha} y'_1, \dots\dots, \alpha y_k + \overline{1-\alpha} y'_k)$ are also elements of S , $0 \le \alpha \le 1$.

Lemma 3.1: The risk set S is convex subset of E_k.

Proof: Let y and y' be arbitrary point of S. according to the definition of S, there exist a decision rules δ and δ' in D^* for which$y_j \quad R(\theta_j, \delta)$

and $y'_j \quad R(\theta_j, \delta')$ j=1, 2,....., k .let α be an arbitrary number such that $0 \le \alpha \le 1$ and consider $\delta_\alpha \quad \alpha\delta + \overline{1-\alpha}\delta'$. Clearly $\delta_\alpha \epsilon D^*$. (as convex combination of d.f is also a d.f)

$$R(\theta_j, \delta_\alpha) \quad E\, L(\theta_j, \delta_\alpha) \quad \alpha\, E\, L(\theta_j, \delta) + \overline{1-\alpha} E\, L(\theta_j, \delta')$$

$$= \alpha R(\theta_j, \delta) + \overline{1-\alpha} R(\theta_j, \delta') \quad Z_j$$

$$Z \quad (Z_1, Z_2, \dots\dots, Z_k) \epsilon S$$

Proved

Defn. 3.8: let A be a set. The convex hull of a set A is the smallest convex set containing A or the intersection of all convex sets containing A.

Thus S defined above is the convex hull of the set S_0, where

$S_0 \quad \{(y_1, y_2, \dots\dots, y_k)\ y_j \quad R(\theta_j, d), d\epsilon D, j \quad 1,2, \dots, k\}$ .... (3.17)

Because the risk function contains all the pertinent information about a decision rule as for as we concerned, the risk set S contains all the information about a decision problem. For a given decision problem (θ, D^*, R) for Θ finite the risk set S is convex; conversely, for any convex set S in k-dimensional space there is a decision problem, (θ, D^*, R) in which Θ consists of k points, whose risk set is the set S.

Bayes Rules:

let $(p_1, p_2, \dots\dots, p_k)$ be a probability distribution on Θ. See points that yield the same expected risk.

$\sum_{j=1}^{k} p_j R(\theta_j, \delta) \quad \sum p_j y_j \qquad , y_j \quad (\theta_j, \delta)$ .... (3.18)

are equivalent in the ordering given by the principle for the prior distribution$(p_1, p_2, \dots\dots, p_k)$. Thus all points on the plane $\sum p_j y_j = b$ for any real number b are equivalent. Every such plane is perpendicular to the vector from the origin to the points $(p_1, p_2, \dots\dots, p_k)$ and because $p_j$is non negative the slope of the line of the interaction of the plane $\sum p_j y_j = b$ with the coordinate planes cannot be positive. The quantity b can best be visualized by noting that the point of interaction of the diagonal line $y_1 = y_2 =.. = y_k$ with the plane $\sum p_j y_j = b$ must occur at $(b, b, \dots, b)$

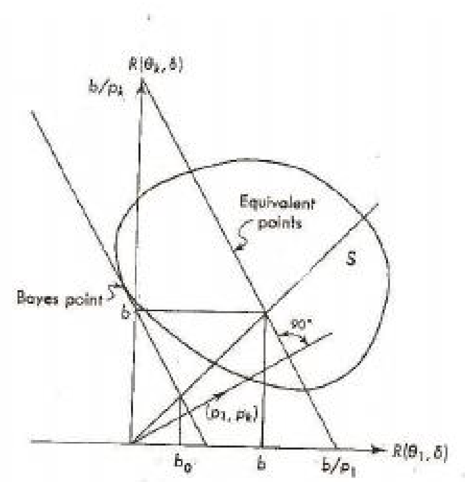


Fig (3.1)

To find the Bayes rules we find the infimum of those values of b, call it b_0, for which the plane $\sum p_j\, y_j = b$ intersected the set S. decision rule corresponding to points in the intersection are Bayes rule with respect to the prior distribution $(p_1, p_2, \dots\dots, p_k)$. There may be many Bayes rules or there may not be any Bayes rules.

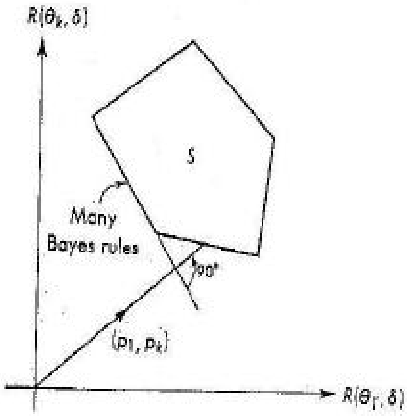


Fig (3.2)

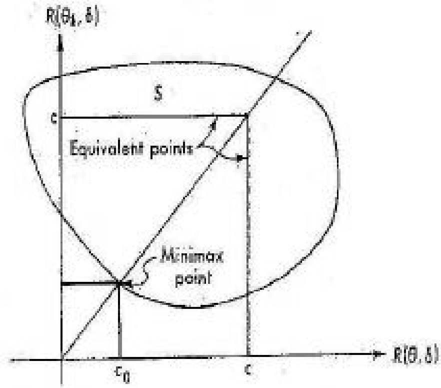


Fig (3.3)

Minimax Rules:

The minimax risk for a fixed δ is $max_j y_j$ $\quad max_j R(\theta_j, \delta)$. Any point y∈S that give rise to the same value of $max_j y_j$ are equivalent in the ordering given by minimax principle. Thus all points on the boundary of that set

$Q_c \quad \{(y_1, y_2, \dots\dots, y_k): y_j \le c \quad for\ i \quad 1, \dots\dots, k\}$ for any real number c are equivalent. To find the minimax rules we find the infimum of those values of c, call it c_0, such that the set Q_c intersects S. any decision rule δ, whose associated risk point is an element of the intersection $Q_{c0} \cap S$, is minimax decision rule. Of course, minimax decision rule do not exist when the set S does not contains its boundary points.

A minimax strategy for nature which is otherwise called a "least favorable distribution" may also be visualized geometrically. A strategy for nature is a prior distribution $\tau \quad (p_1, p_2, \dots\dots, p_k)$ Because the minimum Bayes risk $inf_\delta\ \Upsilon\ (\tau, \delta)$ is b_0, where $(b_0, b_0, \dots\dots, b_0)$ in the intersection of the line $y_1 \quad y_2 \quad \cdots \quad y_k$ and the plane, tangent to and below S, and perpendicular to $(p_1, p_2, \dots\dots, p_k)$, a least favorable distribution is the choice of $(p_1, p_2, \ldots \ldots, p_k)$ that makes this intersection as for up the line as possible. It is clear that b_0 is not greater than c_0, the minimax risk is c_0. This distribution must be least favorable.

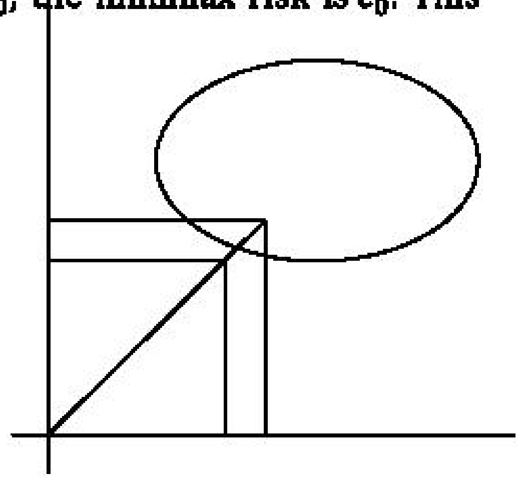


Fig (3.4)

Since

$R(\theta,\delta) \quad E\, R(\theta, Z)$ where Z is a r.v. taking values in D with d.f δ.

$if\ \delta_0$ is such that $R(\theta,\delta_0) \quad inf_{\delta \varepsilon D^*} R(\theta,\delta)$ then

$R(\theta,\delta_0) \quad E\, R(\theta, Z)$ where Z is a r.v. taking values in D with d.f δ_0.

Obviously $\int R(\theta,\delta_0)d\tau \leq \int R(\theta,d)d\tau$ for all dε D

$Y(\tau,\delta_0) \quad \int R(\theta,\delta_0)d\tau \leq inf_{d\varepsilon D}\square(\tau,d)$

$Y(\tau,\delta_0) \quad inf_{\delta\varepsilon D^*}\square(\tau,\delta) \leq inf_{d\varepsilon D}\square(\tau,d)$ ………………… (3.19)

Also $R(\theta,\delta_0) \quad E\, R(\theta, Z)$ Z is a r.v. taking values in D with d.f δ_0.

$$= \int R(\theta,Z)d\delta_0$$

$$\int R(\theta,\delta_0)d\tau \quad \int[\int R(\theta,Z)d\delta_0]d\tau$$

$$= \int[\int R(\theta,Z)d\,\tau]d\,\delta_0$$

$$Y(\tau,\delta_0) \quad \int[\int R(\theta,Z)d\,\tau]d\,\delta_0$$

$$\geq \int[inf_{d\varepsilon D}\int R(\theta,Z)d\,\tau]d\,\delta_0$$

$$= inf_{d\varepsilon D}\square(\tau,d)$$

$$Y(\tau,\delta_0) \geq inf_{d\varepsilon D}\text{?}(\tau,d) \text{ .... } (3.20)$$

From (4.19) and (4.20)

$$Y(\tau,\delta_0) \quad inf_{d\varepsilon D}\text{?}(\tau,d) \text{ .... } (3.21)$$

Equation (3.21) states that none of the mixed strategy (randomized decision rule) can reduce the risk below the minimum value which can be attained from the non-randomized decision D. if Bayes risk $Y(\tau,\delta_0)$ is finite and is attained for a randomized decision rulesδ_0, then it follows from the above comments that this risk must be attained for some non- randomized decision D.

Thus if a Bayes rule with respect to a prior distribution τ exits,

there exist a non- randomized Bayes rule w.r.t. τ. Therefore, one definite computational advantage that the Bayes approach has over the minimax approach to decision theory problem is that the search for good decision rules may be restricted to the class of non- randomized decision rules.

<u>Example</u> 3.4: Let $\theta = a = \{0,1\}$ and let the loss function be L(0,0)=L(1,1)=0 , L(1,0)=L(0,1)=1 Suppose that the statistician observes the r.v. X with discrete distribution

$$P[X \quad x/\theta] \quad 2^{-K} \qquad K \quad x+\theta \qquad k \quad 1,2,3,\ldots\ldots\ldots$$

(I) Describe the set of all non- randomized decision rules.
(II) Plot the risk set S in the plane.
(III) Find the minimax and Bayes decision rules.

<u>Sol</u>: $\mathfrak{X} \quad N$ set of all non- negative integers

Let A be any finite subset of N. $d{:}\mathfrak{X} \rightarrow a = \{0,1\}$

$$D \quad \{d{:} \quad d{:}\mathfrak{X} \rightarrow a\}$$

Thus D contains only two types of functions

$$d_1(x) \quad 1 \quad if\ x\varepsilon A \qquad\qquad d_2(x) \quad 1 \quad if\ x\varepsilon A'$$
$$= 0 \quad if\ x\varepsilon A' \qquad\qquad = 0 \quad if\ x\varepsilon A$$

The cardinality of D is C

$R(\theta, d) \quad E\,L(\theta, d(X))$ is risk function of d.

$R(0, d_1) \quad E\,L(0, d_1(X)) \quad P[X \varepsilon A]$ .... (3.22)

$R(1, d_1) \quad E\,L(1, d_1(X)) \quad P[X \varepsilon A']$ .... (3.23)

$R(0, d_2) \quad E\,L(0, d_2(X)) \quad P[X \varepsilon A']$ .... (3.24)

$R(1, d_1) \quad E\,L(1, d_2(x)) \quad P[X \varepsilon A]$ .... (3.25)

$R(\theta, \delta) \quad \int R(\theta, Z)\, d\delta$ Where Z is a r.v. taking values in D with d.f δ.

Let A {0} ,{0,1}, Φ

$R(0, d_1) \quad P[X \varepsilon A] \quad 0, 1/2, 0$

$R(0, d_2) \quad P[X \varepsilon A'] \quad 1, 1/2, 1$

$R(1, d_1) \quad P[X \varepsilon A'] \quad 1/2, 1/4, 1$

$R(1, d_2) \quad P[X \varepsilon A] \quad \frac{1}{2}, \frac{3}{4}, 0$

(0,1/2) , (1/2,1/4) , (0,1)

(1,1/2) , (1/2,3/4) , (1,0)

$S \quad \{(\alpha, \beta) : 0 \le \alpha \le 1, 0 \le \beta \le 1\}$

$y_1 =$ R(0,δ), $y_2 \quad R(1, d)$

$\alpha \quad R(0, d)\,, \beta \quad R(1, d) \qquad d \varepsilon D$

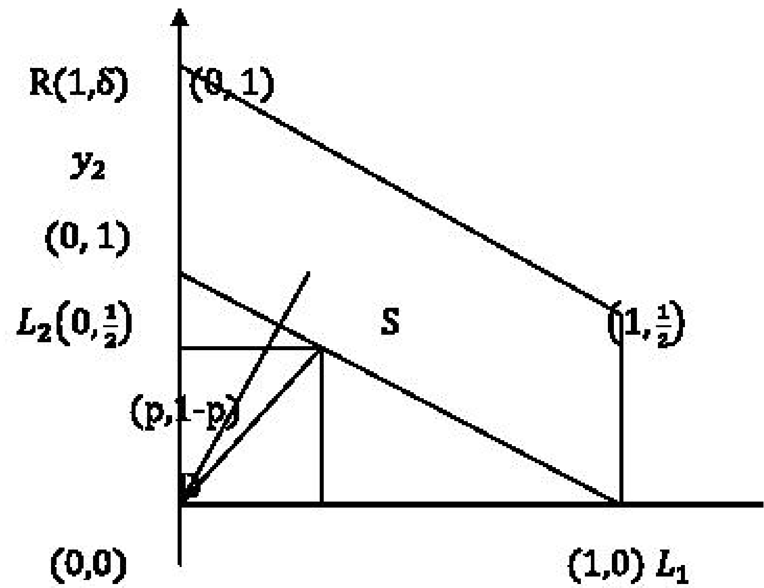


Fig (3.5)

Thus minimax decision rule δ_0 at point D

i.e line L_1 L_2 and intersection of $y_1 \quad y_2$

Line L_1 L_2 is $2y_2 + y_1 \quad 1$

Where $y_1 \quad y_2 \Rightarrow D \quad (\frac{1}{3}, \frac{1}{3})$

So corresponding to $\left(\frac{1}{3},\frac{1}{3}\right)$ *is* $\left(\frac{2}{3},\frac{1}{3}\right)$.

A Bayes decision rule which minimize (3.23) can be found.

__To find a non-randomized rule__:

Let A= {1, 3, 5, 7 ...} d(x) $\begin{matrix} 0 & x\epsilon A \\ 1 & x\epsilon A' \end{matrix}$

$R(0,d) \quad EL(0,d) \quad P[X\epsilon A'] \quad \sum_{x=2,4,6,\dots} 2^{-x}$

$$= \frac{1}{2^2}+\frac{1}{2^4}+\cdots \quad \frac{\frac{1}{4}}{1-\frac{1}{4}} \quad \frac{1}{3}$$

$R(1,d) \quad EL(1,d) \quad P[X\epsilon A] \quad \sum_{x=1,3,5,\dots} 2^{-(x+1)}$

$$\frac{1}{2^2}+\frac{1}{2^4}+\cdots \quad \frac{\frac{1}{4}}{1-\frac{1}{4}} \quad \frac{1}{3}$$

Thus there exist a non-randomized Bayes decision rule such that$\left(\frac{1}{3},\frac{1}{3}\right)$ *point D with probability* $\left(\frac{2}{3},\frac{1}{3}\right)$. A minimax decision rule is$\left(\frac{2}{3},\frac{1}{3}\right)$ choosing,

$d_1(x)$ 0 if x 0 with probability $\frac{2}{3}$ and

=1 if x≥ 1

$d_2(x)$ 1 x ≥ 0 with probability $\frac{1}{3}$

This rule is also Bayes rule with $(p_1,p_2) \quad \left(\frac{1}{3},\frac{2}{3}\right) \quad (p,1-p)$ *as*

$$\frac{1-p}{p}\left(-\frac{1}{2}\right) \quad -1 \Rightarrow 2p \quad 1-p \Rightarrow p \quad \frac{1}{3}$$

__Example__ **3.5:** consider the statistical decision problem.

$\Omega \quad (\theta_1,\theta_2) \quad D \quad (d_1,d_2) \;\; L(\theta,d)$ *as*

		d_1	d_2	
$L(\theta,d)$	θ_1	0	a_1	$a_i > 0 \; i \quad 1,2$
	θ_2	a_2	0	

Let $\alpha(\delta) \quad P[\delta(x) \quad d_2/\theta \quad \theta_1]$

and $\beta(\delta) \quad P[\delta(x) \quad d_1/\theta \quad \theta_2]$

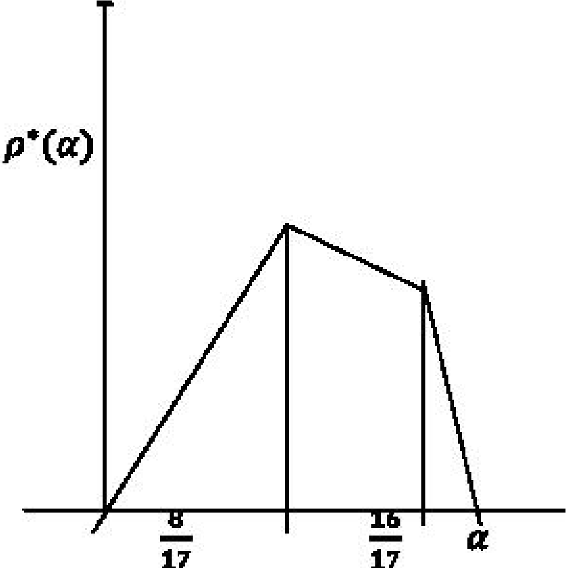


Fig (3.8)

$\alpha(\delta)$ *and* $\beta(\delta)$ *are* the probabilities that δ will lead to a decision when $\theta \quad \theta_1$ and $\theta \quad \theta_2$ respectively, suppose $P[\theta \quad \theta_1] \quad \xi$

$P[\theta \quad \theta_2] \quad 1-\xi$, $0<\xi<1$ is the prior probability.

$▯(\tau,\delta) \quad \iint L(\theta,\delta)\, dF(x/\theta) d\tau(\theta)$

$\int \{L(\theta,d_1)\, P[\delta(x) \quad d_1/\theta] + L(\theta,d_2) P[\delta(x) \quad d_2/\theta]\} d\tau(\theta)$

$[L(\theta_1,d_1)P[\delta(x) \quad d_1/\theta_1] + L(\theta_1,d_2)P[\delta(x) \quad d_2/\theta_1]]\xi$

$+[L(\theta_2,d_1)P[\delta(x) \quad d_1/\theta_2] + L(\theta_2,d_2)P[\delta(x) \quad d_2/\theta_2]](1-\xi)$

$L(\theta_1,d_2)P[\delta(x) \quad d_2/\theta_1]\xi + L(\theta_2,d_1)P[\delta(x) \quad d_1/\theta_2](1-\xi)$

$a_1\alpha(\delta)\xi + a_2\beta(\delta)(1-\xi)$

$a\alpha(\delta) + b\beta(\delta)$ …….. (3.33) Where, $a \quad a_1\xi$, b $\quad a_2(1-\xi)$

__Example__ **3.6:** $\theta \quad \{\theta_1,\theta_2\} \quad$ a $\quad \{a_1,a_2\}$

		a_1	a_2
$L(\theta,a)$	θ_1	-2	3
	θ_2	3	-4

A randomized strategy $\delta \epsilon a^*$ is represented as a number $0 \le q \le 1$, with understanding that

a_1 is taken with probability q and a_2 with 1-q

$S \quad \{(L(\theta_1,\delta), L(\theta_2,\delta)), \delta\epsilon a^*\}$

$L(\theta_1,\delta) \quad EL(\theta_1,z) \quad L(\theta_1,a_1)P_{\theta_1}[z \quad a_1] + L(\theta_1,a_2)P_{\theta_1}[z \quad a_2]$

$$= -2q + 3(1-q) \quad 3 - 5q$$

Similarly, $L(\theta_2,\delta) \quad EL(\theta_2,z) \quad 3q - 4(1-q) \quad 7q - 4$

$S \quad \{(3-5q, 7q-4), 0 \le q \le 1\}$ (Fig 3.6)

Which is nearly a line segment joining (-2, 3) and (3, -4) minimax strategy occurs when,

$3 - 5q \quad 7q - 4 \;\; or\; q \quad \frac{7}{12}$

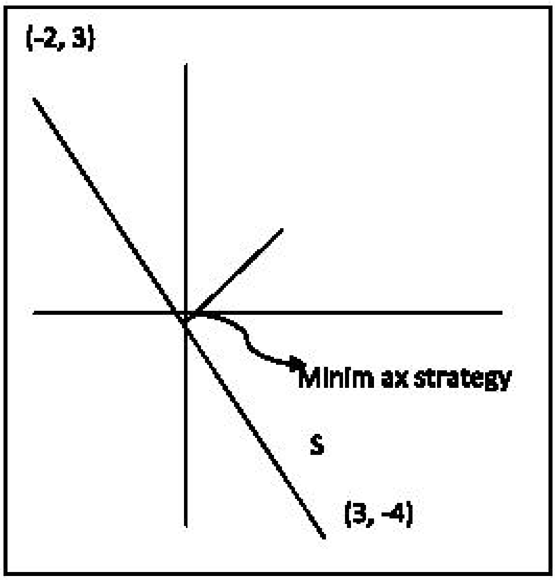
The minimax risk is $\left(\frac{1}{12},\frac{1}{12}\right)$

Thus minimax rule is $\left(\frac{7}{12},\frac{5}{12}\right)$

And this is also Bayes rule since,

$\frac{1-p}{p}\left(-\frac{7}{5}\right) \quad -1 \Rightarrow p \quad \frac{7}{12}$

If we choose θ_1 *with prob.* $\frac{7}{12}$



(Fig 3.6)

And θ_2 *with prob.* $\frac{5}{12}$. $\left(\frac{7}{12},\frac{5}{12}\right)$ *is prior probability.*

<u>Example</u> 3.7: $\theta \quad \{1,2\} \quad$ a

$d_1(1) \quad 1 \quad , d_1(1) \quad 1$

$d_2(1) \quad 1 \quad , d_2(1) \quad 2$

$d_3(1) \quad 2 \quad , d_3(1) \quad 1$

$d_4(1) \quad 2 \quad , d_4(1) \quad 2$

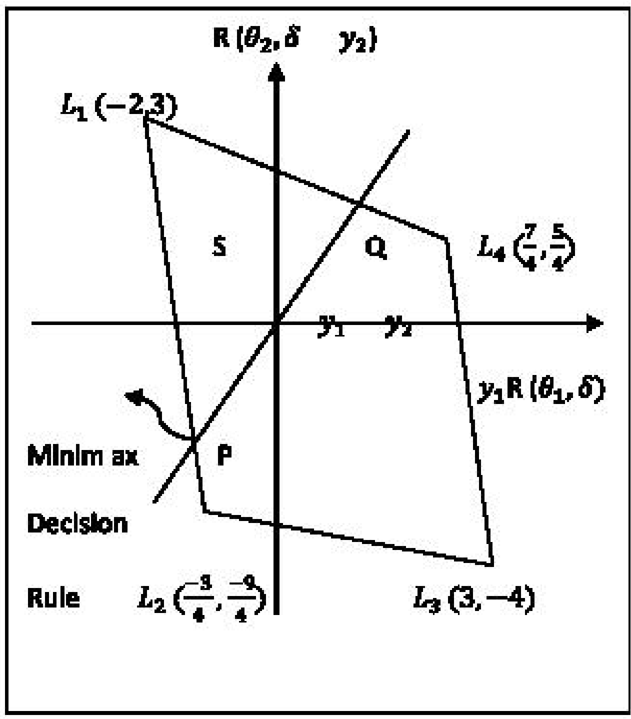
	d_1	d_2	d_3	d_4
1	-2	$-\frac{3}{4}$	$\frac{7}{4}$	3
2	3	$-\frac{9}{4}$	$\frac{5}{4}$	-4

$$R(\theta_1,\delta) \quad p_1R(\theta_1,d_1)+p_2R(\theta_1,d_2)+p_3R(\theta_1,d_3)+p_4R(\theta_1,d_4)$$

$$=-2p_1-\frac{3}{4}p_2+\frac{7}{4}p_3+3p_4, \sum p_i \quad 1$$

$$R(\theta_1,\delta) \quad \sum_{i=1}^{4} p_iR(\theta_2,d_i) \quad 3p_1-\frac{9}{4}p_2+\frac{5}{4}p_3-4p_4$$

$S \quad \{(R(\theta_1,\delta),R(\theta_2,\delta)):\delta\epsilon a^*\}$ (Fig 3.7)



Line L_1L_2 is $y_2 \quad -\frac{21}{5}y_1-\frac{27}{5}$

$$5y_2+21y_1+27 \quad 0$$

Line PQ intersects L_1L_2 at

$y_1 \quad -\frac{27}{26}, y_2 \quad (-27)/26$ Thus

The Minimax risk at $\left(\frac{-27}{26},\frac{-27}{26}\right)$

Thus δ_0 corresponding to this

Minimum is attained by

$\delta_0 \quad \left(\frac{3}{13},\frac{10}{13},0,0\right)$.

Thus δ_0 is also bayes w.r.to

$\left(\frac{21}{26},\frac{5}{26}\right) \quad \tau \;\; as \frac{1-p}{p}\left(-\frac{21}{5}\right) \quad -1 \Rightarrow (1-p)21 \quad 5p \Rightarrow p \quad \frac{21}{26}$

And minimum Bayes risk $\gamma(\tau,\delta_0) \quad \frac{21}{26}$

Also d_1 is non- randomized bayes rule w.r.to τ as

$$\gamma(\tau,d_1) \quad pR(\theta_1,d_1)+(1-p)R(\theta_2,d_1)$$

$$= \frac{21}{26}(-2) + \frac{5}{26}(3) \;= \frac{-42+15}{26} \quad -\frac{27}{26}$$

Thus $\delta_0 \quad \left(\frac{3}{13}, \frac{10}{13}, 0,0\right)$ is randomized Bayes rule and d_1 is non-randomized Bayes rule w.r.to $\tau \quad \left(\frac{21}{26}, \frac{5}{26}\right)$

Thus minimax Bayes risk is $-\frac{27}{26}$.

Given the prior distribution τ, we want to choose a non –randomized decision rule d∊D that minimizes Bayes risk,

$\gamma(\tau, d)$
$\int R(Z, d) d\tau$ *where, Z is a random variable taking values*

$R(\theta, d) \quad \int L(\theta, d(x))\, dF_X(x/\theta)$

A choice of θ by the distribution τ (θ), followed by a choice of X from the distribution$F_X(x/\theta)$, determines a joint distribution of θ and X, which in turn, can be determined by first choosing X according to its marginal distribution,

$$F_X(x) \quad \int F_X(x/\theta) d\tau(\theta) \; \text{.........................} \quad (3.26)$$

and then choosing θ according to the conditional distribution of θ, given X=x, $\tau(\theta/x)$. Hence by a change of integration we may write,

$$\gamma(\tau, d) \quad \int [\int L(\theta, d(x))\, \mathrm{d}\tau(\theta/x)]\, dF_X(x) \; \text{................} \quad (3.27)$$

Given that these operations are legal, it is easy to describe aBayes decision rule.

To find a function d(x) that minimizes the double integral (3.27), we may minimize the inside integral separately for each x; that is, we may find for each x the action, call it d(x), that minimizes

$\int L(\theta, d(x))\, \mathrm{d}\tau(\theta/x)$

Thus, the Bayes decision rule minimizes the posterior conditional expected loss, given the observations.

<u>Non–negative loss function</u>:

Suppose that the distribution of the parameter θ in some decision problem is τ (θ). Let a be a given constant (>0), and let $\lambda(\theta)$ be a real valued function over parameter space Θ=Ω, such that

$$\int_{\Omega} \lambda(\theta) d\tau(\theta) < \infty$$

Consider a new loss function L_0 which is defined in terms of the original loss function L by relation

$$L_0(\theta, d) \quad aL(\theta, d) + \lambda(\theta) \qquad \theta\epsilon\Omega\,, d\epsilon D \text{ .... } (3.28)$$

For any decision d∊D, let Υ (τ, d) denote the risk which results from the original loss function L.

$$\gamma(\tau, d) \quad \int R(\theta, d) d\tau \quad \int\int L(\theta, d)\, dF(x/\theta)\, d\tau(\theta) \text{ .... } (3.29)$$

And let $\gamma_0(\tau, d) \quad \int\int L_0(\theta, d)\, dF(x/\theta)\, d\tau(\theta)$ .... (3.30)

Then for any two decisions $d_1 and\ d_2 \epsilon D$

$$\gamma_0(\tau, d_1) \le \gamma_0(\tau, d_2) \Leftrightarrow \Upsilon(\tau, d_1) \le \Upsilon(\tau, d_2) \text{ .... } (3.31)$$

In particular, a decision d* is Bayes w.r.to τ in the original problem with loss function L (θ, d) if and only if d* is a Bayes w.r.to τ in the new problem with loss functionL_0.

Now consider $\lambda_0(\theta) \quad \inf_{d\epsilon D} L(\theta, d)$

If $\int_{\Omega} \lambda_0(\theta) d\tau(\theta) < \infty$, We can replace L now by a new loss function L_0 which is defined as,

$$L_0(\theta, d) \quad L(\theta, d) - \lambda_0(\theta)$$

Then loss function L_0 has the following property

$$\left.\begin{array}{l} L_0(\theta, d) \ge 0 \quad for\ all\ \theta\ and\ d\ and \\ \inf_{d\epsilon D} L_0(\theta, d) \quad 0 \end{array}\right\} \text{ .... } (3.32)$$

It has been found convenient in many problems to role with non-negative loss function of this type, although the use of such function makes it appear that the statistician must continually choose decisions from which he can never realize a positive gain.

Limit of Bayes rules, generalized Bayes rules and extended Bayes rules:

Defn. 3.9: A rule δ is said to be limit of Bayes rulesδ_n, if for almost all x

$\delta_n(x) \longrightarrow \delta(x)$ (In the sense of distribution) for non-randomized decision rules this definition becomes $d_n \longrightarrow d$ if $d_n(x) \longrightarrow d(x)$ for almost all **x.**

Def 3.10: A rule δ_0 is said to be generalized Bayes rules if there exist a measure τ on Θ (or non decreasing function on θ if Θ is real), such that $R(\tau,\delta) \quad \int\int L(\theta,\delta) f(x/\theta)\, d\tau(\theta)$ takes on a finite minimum value when $\delta \quad \delta_0$

Def 3.11: A rule δ_0 is said to be extended Bayes rules if δ_0 is ϵ- Bayes for every$\epsilon > 0$.

In other words, δ_0 is extended Bayes rules if for every$\epsilon > 0$ there exist a prior distribution τ such that δ_0 is ϵ- Bayes w.r.to τ i.e

$$\Upsilon(\tau,\delta_0) \le \inf_{\delta} \Upsilon(\tau,\delta)$$

Example 3.8: let $X \sim N\ (\theta, 1)$ and let $\tau(\theta) \quad N(0,\sigma^2)$

$L(\theta,d) \quad (\theta - d)^2$ The joint p.d.f of (θ, x)

$$h(\theta,x) \quad \frac{1}{2\pi\sigma}\exp\left[\frac{-(x-\theta)^2}{2} - \frac{\theta^2}{2\sigma^2}\right]$$

$$f_X(\mathrm{x}) \quad \frac{1}{2\pi\sigma}\int \exp\left[\frac{-(x-\theta)^2}{2} - \frac{\theta^2}{2\sigma^2}\right] d\theta$$

$$[2\pi(1+\sigma^2)]^{\frac{-1}{2}} \exp\left[\frac{x^2}{2(1+\sigma^2)}\right]$$

Posterior density of θ given x,

$$f(\theta/x) \quad \frac{(1+\sigma^2)^{\frac{-1}{2}}}{(2\pi\sigma^2)^{\frac{-1}{2}}}\exp\left[\frac{-1+\sigma^2}{2\sigma^2}\left(\theta - \frac{x\sigma^2}{1+\sigma^2}\right)^2\right]$$

$$\sim N\left(\frac{x\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$$

The Bayes rule w.r.to τ_σ is posterior mean i.e $d_\sigma(x) \quad \frac{x\sigma^2}{1+\sigma^2}$

The Bayes risk, $Y(\tau_\sigma, d_\sigma) \quad E[E(\theta - d_\sigma(x))^2/X] \quad \frac{\sigma^2}{1+\sigma^2}$

Thus d(x)=x is not Bayes.

But $d_\sigma(x) \longrightarrow$ d(x) as $\sigma \longrightarrow \infty$.

Theorem 3.1: for any constants a, b >0, let δ^* be a decision rule such that $\delta^*(x) \quad d_1 \qquad if\ af_1(x) > bf_2(x)$

$$= d_2 \qquad if\ af_1(x) < bf_2(x)$$

where f_i denote the conditional p.d.f of X for $\theta \quad \theta_i, i \quad 1,2$

The value of $\delta^*(x)$ may be either d_1 *or* d_2 if $af_1(x) \quad bf_2(x)$. Then for any other decision function δ we have

$$a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta)$$

Proof: let $S_1 \quad \{x: \delta(x) \quad d_1\}$, $S_2 \quad \{x: \delta(x) \quad d_2\} \quad {S_1}^c$

$$A \quad \{x: af_1(x) > bf_2(x)\} \qquad B \quad \{x: af_1(x) < bf_2(x)\}$$

Then $\qquad a\alpha(\delta) + b\beta(\delta) \quad a\int_{S_2} f_1 d\mu + b\int_{S_1} f_2 d\mu$

$a + \int_{S_1}(bf_2 - af_1)d\mu$ .... (3.34)

(3.34) will be minimum if $\int_{S_1}(bf_2 - af_1)d\mu < 0$

Thus $a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta)$.

Finding a decision function δ which minimize the linear combination

$a\alpha(\delta) + b\beta(\delta)$ is equivalent to finding a set S_1 for which the integral

$\int_{S_1}(bf_2 - af_1)d\mu$ is minimized. This integral will be minimized if the set S_1 includes every point x ε S (sample space) for which the integral is negative and excludes every point x ε S for which the integral is positive.

Remark: the posterior distribution of θ θ_1 given X=x , denoted as α(x) is given by,

$$\alpha(x) \quad P[\theta \quad \theta_1/X \quad x]$$

$$= \lim_{h\to 0} \frac{P[\theta=\theta_1, x-h<X\leq x+h]}{P[x-h<X\leq x+h]}$$

$$= \lim_{h\to 0} \frac{P[x-h<X\leq x+h/\theta=\theta_1]P(\theta=\theta_1)}{P[x-h<X\leq x+h]}$$

$$\frac{f(x/\theta_1)P(\theta=\theta_1)}{f_x(x)} \quad \frac{\alpha f(x/\theta_1)}{f_x(x)} \quad \frac{\alpha f_1(x)}{\alpha f_1(x)+1-\alpha f_2(x)}$$

Provided limit exists, where

$f_1(x) \quad f(x/\theta_1)$, $f_2(x) \quad f(x/\theta_2)$

Posterior risk of d_1 $\quad L(\theta_1, d_1)\alpha(x) + L(\theta_2, d_1)(1-\alpha(x))$

$$= a_2(1-\alpha(x)) \quad \text{Similarly, } d_2 \quad a_1\alpha(x)$$

We choose d_2 if (i.e d_2 is Bayes rule) posterior risk of d_2 < posterior risk of d_1. i.e

$$a_1\alpha(x) < a_2(1-\alpha(x)) \text{ or } a_1\alpha f_1(x) < a_2\overline{1-\alpha}f_2(x)$$

Thus $\delta^*(x) \quad d_2(x)$ $if\, a_1\alpha f_1(x) < a_2\overline{1-\alpha}f_2(x)$

Let $S_2 \quad \left\{x: \frac{f_2(x)}{f_1(x)} > \frac{a_1\alpha}{a_2(1-\alpha)}\right\}$ then, $\delta^*(x) \quad d_2(x) \qquad if\, x\varepsilon S_2$

$$d_1(x) \qquad if\, x\varepsilon S_2{}^c$$

For testing $\quad H_0{:}\,\theta \quad \theta_1\, against\ H_1{:}\,\theta \quad \theta_2$,

$$d_1 \quad accept H_0 , d_2 \quad reject H_0 ,$$

$\delta^*(x) \quad \{0,1\}$ i.e choosing d_1 with prob. 0 and d_2 with prob.1.

Or $\delta^*(x) \quad 1 \quad if\ x\varepsilon S_2$

$\quad\quad = 0 \quad if\ x\varepsilon S_2{}^c$

For each θ we have a d.f. of r.v. X as $F(x/\theta)$. Let G (θ) is the d.f. of r.v. θ. Then,

$$F(x/\theta) \quad \lim_{k\to 0} \frac{P[X\le x, \theta-K<\theta<\theta+K]}{P[\theta-K<\theta<\theta+K]} \quad \lim_{k\to 0} \frac{\int_{-\infty}^{x}\int_{\theta-k}^{\theta+k} f(t,v)dtdv}{\int_{\theta-k}^{\theta+k} f_\theta(v)dv}$$

Provided such $f(t,v)$, $f_\theta(v)$ exist and also limit exists. If $f(t,v)$ and $f_\theta(v)$ are continuous.

$$F(x/\theta) \quad \lim_{k\to 0} \frac{2K\int_{-\infty}^{x} f(t,v_0)dt}{2Kf_\theta(v_0)} \qquad \text{Where } v_0\varepsilon(\theta-k,\theta+k)$$

$$= \frac{\int_{-\infty}^{x} f(t,\theta)dt}{f_\theta(\theta)}$$

Since $f(t,v)$ is assumed to be continuous, then

$$F(x/\theta) \quad \frac{f(x,\theta)}{f_\theta(\theta)} \quad \frac{f(x,\theta)}{g(\theta)} \qquad\qquad g(\theta) \quad f_\theta(\theta)$$

Similarly $F(x/\theta) \quad \lim_{k\to 0} \frac{P[X\le x,\theta-K<\theta<\theta+K]}{P[\theta-K<\theta<\theta+K]} \quad = \frac{\int_{-\infty}^{x} f(x,v)dv}{f_X(x)}$

The posterior density of θ given x (when observation X=x is taken.)

$$F(x/\theta) \quad \frac{f(x,\theta)}{f_\theta(\theta)} \quad \frac{f(x,\theta)}{\int f(x,\theta)d\theta} \quad \frac{F(x/\theta)g(\theta)}{\int f(x/\theta)g(\theta)d\theta}$$

This is a continuous version of Bayes theorem.

<u>Bayes Estimation Problem</u>: the determination of a Bayes estimator is in principle simple. First we consider the estimation before any observation is taken.

Then θ has d.f. Λ and the Bayes estimator of g (θ) is any number which minimizes $EL(\theta,d) = \int L(\theta,d)\, d\Lambda(\theta)$. Once the data have been obtained and are given by the observed value x of X, the prior distribution Λ of θ is replaced by the posterior distribution i.e the conditional distribution of θ given X=x and the Bayes estimator is any number δ(x) which minimizes the posterior risk $E[L(\theta,\delta(X))/X \quad x]$.

Theorem 3.2: Let Θ has distribution Λ and given Θ=θ, let X have distribution $F(x/\theta)$ ($or\ p.d.f\ f(x/\theta)$), the following assumption hold for the problem of estimating $g(\theta)$ with non-negative loss function $L(\theta,d)$

(a) There exists an estimator δ_0 with finite risk.

(b) For almost all x, there exists a value $\delta_\Lambda(x)$ which minimizes $E[L(\theta,\delta(X))/X \quad x]$ .... (3.35)

Then, $\delta_\Lambda(x)$ is Bayes estimator.

Proof: let δ be any estimator with finite risk. Then (3.35) is finite a.e. since L is non-negative. i.e

$$\iint L(\theta,\delta(X))dF(x/\theta)d\Lambda < \infty$$

$$\int\left[\int L(\theta,\delta(X))\,d\Lambda(\theta/x)dF_X(x)\right] < \infty$$
$$\int L(\theta,\delta(X))\,d\Lambda(\theta/x) < \infty$$

$$E[L(\theta,\delta(X))/X \quad x] \geq E[L(\theta,\delta_\Lambda(X))/X \quad x]\ a.e$$

$$E[E\{L(\theta,\delta(X))/X \quad x\}] \geq E[E\{L(\theta,\delta_\Lambda(X))/X \quad x\}]\ a.e$$

⍰$(\Lambda,\delta) \quad E\,L(\theta,\delta(X)) \geq E\,L(\theta,\delta_\Lambda(X)) \quad$ ⍰(Λ,δ_Λ)

This is true for all δ for which risk is finite.

$inf_{\delta\varepsilon D^*}$⍰$(\Lambda,\delta) \quad$ ⍰(Λ,δ_Λ) *thus* δ_Λ is Bayes estimator for squared error loss function the non-randomized Bayes rule is the mean of the posterior distribution. For $L(\theta,d) = |\theta - d|$ the Bayes rule is median of the posterior distribution.

For $L(\theta,d) \quad 0 \qquad if|\theta - d| \leq c$

$\qquad\qquad\quad 1 \qquad if|\theta - d| > c$

Thus $\delta_\Lambda(x)$ is the mid point of the integral I of the length 2c which maximizes $P[\theta \varepsilon I/x]$

$$E[g(\theta)-\delta(x)]^2 \quad E[g(\theta)-E(g(\theta)/x)]^2+E[g(\theta)/x-\delta(x)]^2$$

$$\geq E[g(\theta)/x-\delta(x)]^2 \quad E\{[g(\theta)-\delta(x)]^2/x\}$$

L.H.S will be minimized when $\delta(x) \quad E[g(\theta)/x]$ similarly others.

Corollary: If the loss function $L(\theta, d)$ is squared error or more generally if it is strictly convex in d, a Bayes solution δ_Λ is unique a.e. P where P is the class of distributions P_θ, provided

1. Its average risk w.r.t. to Λ is finite.
2. If Q is marginal distribution of X given by

$$Q(A) \quad \int P_\theta(x \varepsilon A)\, d\Lambda(\theta)$$

Then a.e Q implies P.

Proof: To prove this, we have the following property of convex function. " every convex function in an open interval (a, b) are continuous. Let Φ be strictly convex function defined over an interval I (finite or infinite). If there exist a value a_0 in I minimize $\Phi(a)$, then a_0 is unique."

Let $L(\theta, d) \quad w(\theta)(g(\theta)-d)^2$ where, $w(\theta)>0$ for all θ ε Ω.

$$?(\tau, d) \quad E\, L(\theta, d) \quad \iint w(\theta)(g(\theta)-d)^2\, dF(x/\theta) d\tau(\theta)$$

$$= \iint w(\theta)(g(\theta)-d)^2\, d\tau(\theta/x) dF_X(x)$$

$$\int \left[\int w(\theta)(g(\theta)-d)^2\, d\tau(\theta/x)\right] dF_X(x)$$

Minimized[by Theorem (3.2)],if

$\int w(\theta)(g(\theta)-d)^2\, d\tau(\theta/x)$ is minimized. Let us define

$$h(d) \quad \int w(\theta)(g(\theta)-d)^2\, d\tau(\theta/x)$$
$$\int w(\theta) g^2(\theta) d\tau(\theta/x) + d^2 \int w(\theta) d\tau(\theta/x) \; - 2d \int w(\theta) g(\theta) d\tau(\theta/x)$$

For maxima and minima, we must have

$\frac{\partial h}{\partial d} \quad 0 \Rightarrow d \quad \frac{\int w(\theta)g(\theta)d\tau(\theta/x)}{\int w(\theta)d\tau(\theta/x)} \quad \frac{E[w(\theta)g(\theta)/x]}{E[w(\theta)/x]}$

Hence $\delta_\tau \quad \frac{E[w(\theta)g(\theta)/x]}{E[w(\theta)/x]}$

But, $w(\theta)(g(\theta)-d)^2 \quad w(\theta)g^2(\theta)+d^2w(\theta)-2dw(\theta)g(\theta)$ is convex (strictly) function in d as, $\frac{\partial^2}{\partial d^2}w(\theta)(g(\theta)-d)^2 \quad w(\theta)>0$ hence, by the property of convex function δ_τ is unique Bayes estimator for g (θ).

To see the a.e Q ⇒ a.e P. let the parameter Ω is an open set which is support of $\tau(\theta)(=\Lambda(\theta))$ and $P_\theta(X\varepsilon A)$ is a continuous function of θ for any A. Thus Q(N)=0⇒ $P_\theta(N)$=0 a.e Λ. If there exists $P_\theta(N)$>0,

then there exists a neighborhood w of θ_0 in which $P_\theta(N)$>0 for θεw.

By the support assumption $P_\Lambda(w)$>0 contradicts this assumption $P_\theta(N)$=0 a.e Λ. Hence Q(N)=0⇒ $P_\theta(N)$=0 for a.e θ.

<u>Theorem 3.3</u>: let Θ have a distribution function τ. Let F_θ denote the conditional d.f of X given θ. Consider the estimation of g(θ) when the loss function is squared error. Then no unbiased estimator $\delta(x)$ can be a Bayes solution unless

$$E[\delta(x)-g(\theta)]^2 \quad 0 \text{ .... } (3.36)$$

Where, the expectation is taken w.r.t. variations in both X & θ.

<u>Proof</u>: suppose δ(x) is Bayes estimator, and is unbiased for g (θ). Since δ is Bayes and loss is squared

$$\delta(x) \quad E[g(\theta)/X \quad x] \text{ with probability 1. .... } (3.37)$$

Since $\delta(x)$ is unbiased

$E_\theta[\delta(x)] \quad g(\theta)$ for all θ .... (3.38)

$E[g(\theta)\delta(x)] \quad E[\delta(x)g(\theta)/X \quad x] \quad E\delta^2(x)$ .... (3.39)

And also,

$E[g(\theta)\delta(x)] \quad E[g(\theta)\delta(x)/\theta] \quad Eg^2(\theta)$ .... (3.40)

Thus,

$E[g(\theta)-\delta(x)]^2 \quad Eg^2(\theta)+E\delta^2(x)-2E[g(\theta)\delta(x)]$

$= Eg^2(\theta)-E[g(\theta)\delta(x)]+E\delta^2(x)-E[g(\theta)\delta(x)]$ .... (3.41)

By (3.39) & (3.40) $\Rightarrow$ (3.41) =0.

#

<u>Def 3.12</u>: Denote the average Bayes risk of the Bayes decision rule δ_τ by

$\square(\tau,\delta_\tau) \quad \int R(\theta,\delta_\tau)\,d\tau(\theta) \quad \gamma_\tau$ .... (3.41)

A prior distribution τ is said to be least favorable if $\gamma_\tau \geq \gamma_{\tau'}$ for all prior distribution τ'.

<u>Theorem 3.4</u>: suppose that τ is a distribution of Θ such that,

$\int R(\theta,\delta_\tau)\,d\tau(\theta) \quad supR(\theta,\delta_\tau)$ .... (3.42)

Then,

1. δ_τ is minimax.
2. If δ_τ is the unique Bayes solution w.r.t. τ, it is unique minimax procedure.
3. τ is least favorable d.f of θ.

<u>Proof</u>: 1. Let δ be any other decision rule. Then

$supR(\theta,\delta_\tau) \geq \int R(\theta,\delta)\,d\tau(\theta)$

$\geq \int R(\theta,\delta_\tau)\,d\tau\theta$ As δ_τ is Bayes decision rule.

$sup_\theta R(\theta,\delta_\tau)$ by (3.11) δ_τ is minimax.

2. Follows by replacing ≥ by > in second inequality of the proof 1.

3. Let τ' be any other prior of θ then,

$\square(\tau',\delta_{\tau'}) \quad \int R(\theta,\delta_{\tau'})\,d\tau' \leq \int R(\theta,\delta_\tau)\,d\tau'(\theta)$

$\leq sup R(\theta,\delta_\tau)$ □(τ,δ_τ) Thus τ is least favorable.

<u>Corollary</u>: If δ_τ has constant risk$R(\theta,\delta_\tau)$, δ_τ is minimax.

<u>Ex 3.9</u>: suppose $X \sim b(n,\theta)$ and $L(\theta,\delta) \quad (\theta-\delta)^2$ find a minimax estimator of θ.

<u>Solution</u>: let the prior distribution of θ be B (a, b) (beta distribution) the posterior d.f of θ given x.

$$g(\theta/x) \quad \frac{f(\theta,x)}{f_X(x)} \quad \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}\theta^{a-1}(1-\theta)^{b-1}}{\binom{n}{x}\int_0^1 \theta^{x+a-1}(1-\theta)^{n-x+b-1}}$$

$$\frac{\theta^{x+a-1}(1-\theta)^{n-x+b-1}\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(n+b-x)}$$

Thus $\delta_B(x) \quad E(\theta/x) \quad \frac{a+x}{a+b+n}$

$$\left[if\ x \sim B(a,b), E(x) \quad \frac{a}{a+b}, var\ (x) \quad \frac{ab}{(a+b)^2(a+b+1)}\right]$$

$$R(\theta,\delta_B) \quad E_\theta(\theta-\delta_B)^2 \quad E_\theta\left(\theta-\frac{a+x}{a+b+n}\right)^2$$

$$E_\theta\frac{(a\theta+b\theta+n\theta-a-x)^2}{(a+b+n)^2} \quad E_\theta\frac{\left((a+b)\theta-(x-n\theta)-a\right)^2}{(a+b+n)^2}$$

$$\frac{(a+b)^2\theta^2+var(x)+a^2-2a(a+b)\theta}{(a+b+n)^2}$$

$$\frac{[n\theta(1-\theta)+a^2-2a^2\theta-2ab\theta+a^2\theta^2+b^2\theta^2+2ab\theta^2]}{(a+b+n)^2}$$

$$\frac{[n\theta(1-\theta)+a^2(1-2\theta+\theta^2)+b^2\theta^2-2ab\theta(1-\theta)]}{(a+b+n)^2}$$

$$\frac{[n\theta(1-\theta)+[a(1-\theta)-b\theta]^2]}{(a+b+n)^2}$$

Thus $R(\theta,\delta_B) \quad \frac{n\theta(1-\theta)+[a(1-\theta)-b\theta]^2}{(a+b+n)^2}$

If $R(\theta,\delta_B)$ is constant δ_B is minimax. Equivalently the coefficient of θ^2 and θ equal to zero we get the value of a and b for which $R(\theta,\delta_B)$ is constant.

$$-n + a^2 + b^2 + 2ab \quad 0 \Rightarrow a + b \quad \sqrt{n} \quad (as\ a > 0, b > 0)$$

$$n - 2a^2 - 2ab \quad 0 \Rightarrow 2a(a + b) \quad n \Rightarrow a \quad \frac{\sqrt{n}}{2} \quad b$$

Thus $R(\theta, \delta_B) \quad \frac{a^2}{(a+b+n)^2} \quad \frac{n}{4(\sqrt{n}+n)^2} \quad \frac{1}{4(1+\sqrt{n})^2}$ Thus δ_B has constant Bayes risk for this value of a & b. and the constant risk $\frac{1}{4(1+\sqrt{n})^2}$.Thus by corollary of theorem 3.4

$\delta_B \quad \frac{X+\frac{\sqrt{n}}{2}}{n+\sqrt{n}} \quad \frac{X}{n}\left(\frac{\sqrt{n}}{1+\sqrt{n}}\right) + \frac{1}{2(1+\sqrt{n})}$ is of constant risk and it is unique minimax Bayes estimator of θ.

<u>Def 3.13</u>: let τ_n be a sequence of prior distribution, and δ_n the Bayes estimator corresponding τ_n. Suppose that Bayes risk is $\int R(\theta, \delta_n)\, d\tau_n = □_n$ and that $□_n \to □$ .... (3.43)

Then the sequence $\{\tau_n\}$ is said to be least favorable if for every prior τ we have $□_\tau \leq □$.

<u>Theorem 3.5</u>:

Suppose that τ_n is a sequence of prior distribution with Bayes risk $□_n$ satisfying (3.43) and that δ is an estimator for which $sup_\theta R(\theta, \delta) \quad □$ then,

1. δ is minimax.
2. The sequence τ_n is least favorable.

<u>Proof</u>: 1. let δ' be any other estimator. Then,

$$sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta') d\tau_n \geq \gamma_n \quad for\ \forall\ n.$$

Hence $\quad sup_\theta R(\theta, \delta') \geq \lim_{n\to\infty} \gamma_n \quad \gamma \quad sup_\theta R(\theta, \delta)$

Thus δ is minimax.

2. If τ is any distribution then,

$$\gamma_\tau \quad \int R(\theta, \delta_\tau)\, d\tau \leq \int R(\theta, \delta)\, d\tau \leq sup_\theta R(\theta, \delta) \quad \gamma$$

Thus by definition $\{\tau_n\}$ is least favorable.

<u>Example</u> 3.10: let $X_1, X_2, \dots\dots, X_n$ be i.i.d r.v from N (θ, 1). Let $L(\theta, d)$ $(\theta - d)^2$. Let the prior distribution θ be N $(0, b^2)$. Then show that $\bar{X}$ is minimax estimator.

<u>Solution</u>: $\delta_b \quad \frac{n\bar{x}}{n+\frac{\sigma^2}{b^2}} \quad \frac{n\bar{x}}{n+\frac{1}{b^2}}$ *as* $\sigma^2 \quad 1$

$var(\theta/x) \quad \frac{1}{n+\frac{1}{b^2}} \qquad \gamma_b \quad \frac{1}{n+\frac{1}{b^2}} \to \frac{1}{n} \quad$ *as* $b \to \infty$

Thus $\gamma \quad \frac{1}{n} \quad R(\theta, \bar{x}) \quad E(\theta - \bar{x})^2 \quad \frac{1}{n} \quad sup_\theta R(\theta, \bar{x}) \quad \frac{1}{n}$

Thus by Theorem (3.5) $\bar{X}$ is minimax.

<u>Example</u> 3.11: let $\theta \quad \left\{\frac{1}{3}, \frac{2}{3}\right\}$, $L(\theta, d) \quad (\theta - d)^2 \; a \quad (-\infty, \infty)$ a coin is tossed once and the probability of 'H' is θ. Find a non-randomized Bayes decision rule which is minimax with prior distribution.

$$P[\theta \quad \theta_1] \quad p \quad 1 - P[\theta \quad \theta_2].$$

<u>Solution</u>: $d: x \to [0,1]$ where $x \quad \{H, T\}$

$D \quad \{(x, y): d\{H\} \quad x, d\{T\} \quad y\}$ *set of non randomized rules,*

$$0 \le x \le 1 \qquad 0 \le y \le 1$$

$d\{H\} \quad x$, *we estimate θ by x when H is observed*

d$\{T\} \quad y$, *we estimate θ by y when T is observed*

$$R(\theta_1, (x, y)) \quad R(\theta_1, d) \quad E_{\theta_1} L(\theta_1, d)$$

$$R\left(\frac{1}{3}, (x, y)\right) \quad E_{\frac{1}{3}} L\left(\frac{1}{3}, (x, y)\right)$$

$$\left(\frac{1}{3} - x\right)^2 P_{\frac{1}{3}}[H] + \left(\frac{1}{3} - y\right)^2 P_{\frac{1}{3}}[T]$$

$$= \frac{1}{3}\left(\frac{1}{3} - x\right)^2 + \frac{2}{3}\left(\frac{1}{3} - y\right)^2$$

$$R\left(\frac{2}{3},(x,y)\right) \quad E_{\frac{2}{3}}L\left(\frac{2}{3},(x,y)\right) \quad E_{\frac{2}{3}}L\left(\frac{2}{3}-(x,y)\right)^2$$

$$\left(\frac{2}{3}-x\right)^2 P_{\frac{2}{3}}[H]+\left(\frac{2}{3}-y\right)^2 P_{\frac{2}{3}}[T]$$

$$\frac{2}{3}\left(\frac{2}{3}-x\right)^2+\frac{1}{3}\left(\frac{2}{3}-y\right)^2$$

The Bayes risk, $\square\left(\tau,(x,y)\right) \quad E\, R\left(\theta,(x,y)\right)$

$$pR\left(\frac{1}{3},(x,y)\right)+(1-p)R\left(\frac{2}{3},(x,y)\right)$$

$$p\left[\frac{1}{3}\left(\frac{1}{3}-x\right)^2+\frac{2}{3}\left(\frac{1}{3}-y\right)^2\right]+(1-p)\left[\frac{2}{3}\left(\frac{2}{3}-x\right)^2+\frac{1}{3}\left(\frac{2}{3}-y\right)^2\right]$$

$$p\left[\frac{1}{9}+\frac{x^2}{3}-\frac{2x}{9}+\frac{2y^2}{3}-\frac{4y}{9}\right]+(1-p)\left[\frac{4}{9}+\frac{2x^2}{3}-\frac{8x}{9}+\frac{y^2}{3}-\frac{4y}{9}\right]$$

$$p\left[\frac{1}{9}+\frac{x^2}{3}-\frac{2x}{9}-\frac{4y}{9}+\frac{2y^2}{3}-\frac{4}{9}-\frac{2x^2}{3}-\frac{y^2}{3}+\frac{8x}{9}+\frac{4y}{9}\right]+\left[\frac{4}{9}+\frac{2x^2}{3}-\frac{8x}{9}+\frac{y^2}{3}-\frac{4y}{9}\right]$$

$$p\left[-\frac{x^2}{3}+\frac{6x}{9}-\frac{3}{9}+\frac{y^2}{3}\right]+\left[\frac{4}{9}+\frac{2x^2}{3}-\frac{8x}{9}+\frac{y^2}{3}-\frac{4y}{9}\right] \text{..........} \quad (3.44)$$

For that value of p foe which $\square\left(\tau,(x,y)\right)$ is constant, we have

$$\frac{y^2}{3}-\frac{x^2}{3}+\frac{6x}{9}-\frac{3}{9} \quad 0 \Rightarrow y^2-(x^2-2x+1) \quad 0$$

$$\Rightarrow y^2-(1-x)^2 \quad 0 \text{ or } y^2 \quad (1-x)^2$$

Now we have to find that x & y which minimizes (3.44)

$$p\left[-\frac{2x}{3}+\frac{6}{9}\right]+\frac{4x}{9}-\frac{8}{9} \quad 0 \Rightarrow x\left[\frac{4}{3}-\frac{2p}{3}\right] \quad \frac{8}{3}-\frac{6p}{9}$$

$$x[2-p] \quad \frac{4-3p}{3} \Rightarrow x \quad \frac{4-3p}{3(2-p)} \text{..................} \quad (3.45)$$

Similarly,

$$p\left[\frac{2y}{3}\right]+\frac{2y}{3}-\frac{4}{9} \quad 0 \Rightarrow \frac{2y}{3}[1-p] \quad \frac{4}{9} \Rightarrow y \quad \frac{2}{3(1-p)} \text{..................} (3.46)$$

The Bayes risk will be constant if only if,

$$y^2 \quad (1-x)^2 \quad > \left(\frac{2}{3(1-p)}\right)^2 \quad \frac{4}{(3(2-p))^2}\left[1-x \quad \frac{2}{3(2-p)}\right]$$

Or $2-p \quad \pm(1-p) \Rightarrow p \quad \frac{1}{2}$

Thus the prior distribution $\left(\frac{1}{2},\frac{1}{2}\right)$ *on* $\theta \quad \left(\frac{1}{3},\frac{2}{3}\right)$ is least favorable by Theorem (3.4) and non randomized Bayes rule is,

d: $x \quad 5/9$, $y \quad 4/9$ (That is estimate θ if 'H'turns up by 5/9, and by 4/9 if 'T' turns up.)

If $a \quad \theta \quad (0,1)$ in this case let the

$$R(\theta,(x,y)) \quad \theta(\theta-x)^2+(1-\theta)(\theta-x)^2$$

$$\theta x^2-2\theta^2 x+(1-\theta)y^2-2\theta(1-\theta)y+\theta^2$$

For any prior distribution over Θ. $\tau(\theta)$

$$\square(\tau,(x,y)) \quad x^2E(\theta)-2xE\theta^2+y^2E(1-\theta)-2yE\theta(1-\theta)+E\theta^2$$

Let $m_1 \quad E\theta$, $E\theta^2 \quad m_2$

$$\square(\tau,(x,y)) \quad x^2m_1-2xm_2+y^2-y^2m_1-2y(m_1-m_2)+m_2$$

$$x^2m_1-2xm_2+(1-m_1)y^2-2y(m_1-m_2)+m_2 \text{ .... } (3.47)$$

For minimum value of (3.47)

$$2xm_1-2m_2 \quad 0 \Rightarrow x \quad \frac{m_2}{m_1} \text{ .... } (3.48)$$

And $2y(1-m_1)-2(m_1-m_2) \quad 0 \Rightarrow y \quad \frac{m_1-m_2}{1-m_1}$ .... (3.49)

For the constant $R(\theta,(x,y))$ for all θ. We must have the coefficients of θ^2 and θ.

$$(-2x+2y+1) \quad 0 , \qquad x^2-y^2-2y \quad 0$$

Or $\quad 2x \quad 2y+1$, $\qquad x^2 \quad y^2+2y$

$\left(\frac{2y+1}{2}\right)^2 \quad y^2 + 2y \Rightarrow 4y^2 + 4y + 1 \quad 4y^2 + 8y$

$\Rightarrow y \quad \frac{1}{4} \,\&\, x \quad 3/4$

Thus for Bayes risk should be constant. $\frac{m_2}{m_1} \quad \frac{3}{4}$

Or $\quad \frac{m_1 - m_2}{(1-m_1)} \quad \frac{1}{4} \Rightarrow m_1 \quad \frac{1}{2} \,\&\, m_2 \quad \frac{3}{8}$

That a prior distribution for which, $E\theta \quad \frac{1}{2}, \quad E\theta^2 \quad \frac{3}{8}$

A distribution satisfying this property is, $\theta \sim B\left(\frac{1}{2},\frac{1}{2}\right) \quad E\theta \quad \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}} \quad \frac{1}{2}$

$E\theta^2 \quad var(\theta) + (E\theta)^2 \quad \frac{\left(\frac{1}{2}\right)^2}{2} + \left(\frac{1}{2}\right)^2 \quad \frac{1}{8} + \frac{1}{4} \quad \frac{3}{8}$

Thus if,

$\tau(\theta) \quad B\left(\frac{1}{2},\frac{1}{2}\right)$, Then the above estimator is minimax.

Example:3.12 Let X be a random variable with pmf

$P(X \quad x) \quad (1-\theta)\theta^{x-1}, x \quad 1, 2, \ldots, \infty, 0 \le \theta < 1.$

For the squared loss function $(d-\theta)^2$, find the minimax estimator for θ.

Solution: Let $d(1) \quad a, d(2) \quad d(3) \quad d(4) \quad \cdots. \quad b$. Then

$R(\theta, d) \quad E_\theta[d(X) - \theta]^2$

$= E_\theta[d(X)]^2 - 2\theta\, E_\theta[d(X)] + \theta^2$

$R(\theta, d) \quad [d(1)]^2 P(X \quad 1) + \sum_{x=2}^{\infty}[d(x)]^2 P(X \quad x) - 2\theta\{[d(1)]P(X \quad 1) + \sum_{x=0}^{\infty}[d(x)]\, P(X \quad x)\} + \theta^2$

$= a^2(1-\theta) + b^2\,\theta - 2\theta[a(1-\theta) + b\,\theta\,] + \theta^2$

$= a^2 + \theta(b^2 - a^2) - 2\theta[a + \theta(b-a)] + \theta^2$

$$= \theta^2[1-2(b-a)] + \theta[(b^2 - a^2) - 2a] + a^2$$

For a minimax estimator $R(\theta, d)$ must be a constant that is coefficients of θ^2 and θ must be zero. This implies that

$$(b-a) \quad \frac{1}{2}, b^2 \quad a^2 + 2a \Rightarrow a \quad \frac{1}{4}, b \quad \frac{3}{4}$$

Thus the minimax estimator is

Estimate θ by $\frac{1}{4}$ if $x \quad 1$ is observed and by $\frac{3}{4}$ if $x \geq 2$. Let τ be the prior distribution of θ. Then Bayes risk is

$r_\tau(d) \quad m_2[1-2(b-a)] + m_1[(b^2 - a^2) - 2a] + a^2$, where

$m_2 \quad E(\theta^2), m_1 \quad E(\theta)$.

For minima and maxima $\frac{\partial r_\tau}{\partial a} \quad 0 \; and \; \frac{\partial r_\tau}{\partial b} \quad 0 \Rightarrow a \quad \frac{m_2 - m_1}{m_1 - 1}, b \quad \frac{m_2}{m_1}$. Putting the values of $a \; and \; b$ just obtained, we will have $m_1 \quad \frac{1}{2}, m_2 \quad \frac{3}{8}$. The corresponding prior distribution of θ is choose $\theta \quad \frac{3}{4}$

With probability $\frac{2}{3}$ and $\theta \quad 0$ with probability $\frac{1}{3}$.

<u>Problems :</u>

1. Let X be a random variable with pmf

$P(X \quad x) \quad \theta(1-\theta)^{x-1}, x \quad 1, 2, \ldots, \infty, 0 < \theta \leq 1.$

For the squared loss function $(d-\theta)^2$, find the minimax estimator for θ.

2. Let X be a random variable with pmf

$P(X \quad x) \quad \theta(1-\theta)^{x}, x \quad 0, 1, 2, \ldots, \infty, 0 < \theta < 1.$

Under the squared loss function $\frac{(d-\theta)^2}{1-\theta}$, find the minimax estimator for θ.

<u>Example</u>: 3.13: Let X and Y be independent binomial variates respectively with $b(n,p)$ and $(n,1-p)$. Show that the minimax estimator of $p_1 - p_2$ is

$d^{**} \quad \frac{\sqrt{2n}}{1+\sqrt{2n}}\left[\frac{X}{n}-\frac{Y}{n}\right]$, where $p_1 \quad 1-p_2 \quad p \Rightarrow p_1-p_2 \quad 2p-1.$

<u>Solution</u>: $X{\sim}b(n,p), Y{\sim}b(n,1-p) \Rightarrow n-Y{\sim}b(n,p)$.Hence

$Z \quad X+n-Y{\sim}b(2n,p)$. We know that the minimax estimator having constant Bayes risk for $b(n,p)$ is (see example 4.9)

$d^{*} \quad \frac{1}{1+\sqrt{n}}\left[\frac{X}{\sqrt{n}}+\frac{1}{2}\right]$. Thus for $b(2n,p)$

$$d^{*} \quad \frac{1}{1+\sqrt{2n}}\left[\frac{Z}{\sqrt{2n}}+\frac{1}{2}\right]=\frac{1}{(1+\sqrt{2n})}\frac{Z}{\sqrt{2n}}+\frac{1}{2(1+\sqrt{2n})}$$

$$=\frac{\sqrt{2n}}{(1+\sqrt{2n})}\frac{Z}{2n}+\frac{1}{2(1+\sqrt{2n})} \Rightarrow 2d^{*} \quad \frac{\sqrt{2n}}{(1+\sqrt{2n})}\frac{Z}{n}+\frac{1}{(1+\sqrt{2n})}$$

$$2d^{*}=\frac{\sqrt{2n}}{(1+\sqrt{2n})}\frac{X+n-Y}{n}+\frac{1}{(1+\sqrt{2n})} \quad \frac{\sqrt{2n}}{(1+\sqrt{2n})}\left[\frac{X}{n}-\frac{Y}{n}\right]+\frac{\sqrt{2n}}{(1+\sqrt{2n})}+\frac{1}{(1+\sqrt{2n})}$$

$$2d^{*} \quad \frac{\sqrt{2n}}{(1+\sqrt{2n})}\left[\frac{X}{n}-\frac{Y}{n}\right]+1 \Rightarrow 2d^{*}-1 \quad \frac{\sqrt{2n}}{(1+\sqrt{2n})}\left[\frac{X}{n}-\frac{Y}{n}\right] \quad d^{**}$$

Thus, the minimax estimator of $2p-1$ is $2d^{*}-1 \quad d^{**}$. The risk of d^{**} under squared error loss function with $\theta \quad p_1-p_2$ is

$$R(d^{**}) \quad R(\theta,d^{**}) \quad E_\theta\left[\frac{\sqrt{2n}}{1+\sqrt{2n}}\left(\frac{X}{n}-\frac{Y}{n}\right)-(p_1-p_2)\right]^2$$

$$=E_\theta\left[\frac{\sqrt{2n}}{1+\sqrt{2n}}\left(\frac{X}{n}-\frac{Y}{n}\right)-\frac{\sqrt{2n}+1}{\sqrt{2n}+1}(p_1-p_2)\right]^2$$

$$=E_\theta\left[\frac{\sqrt{2n}}{1+\sqrt{2n}}\left\{\left(\frac{X}{n}-\frac{Y}{n}\right)-(p_1-p_2)\right\}-\frac{1}{\sqrt{2n}+1}(p_1-p_2)\right]^2$$

$$=\frac{2n}{(1+\sqrt{2n})^2}E_\theta\left[\left(\frac{X}{n}-\frac{Y}{n}\right)-(p_1-p_2)\right]^2+\frac{1}{(1+\sqrt{2n})^2}(p_1-p_2)^2.$$

The product term will be zero as

$$E\left(\frac{X}{n}-\frac{Y}{n}\right) \quad \frac{np}{n}-\frac{n(1-p)}{n} \quad 2p-1 \quad p_1-p_2.$$

$$R(d^{**}) \quad \frac{2n}{\left(1+\sqrt{2n}\right)^2}Var\left(\frac{X}{n}-\frac{Y}{n}\right)+\frac{1}{\left(1+\sqrt{2n}\right)^2}(p_1-p_2)^2$$

$$=\frac{1}{\left(1+\sqrt{2n}\right)^2}\left[(p_1-p_2)^2+\ 2nVar\left(\frac{X}{n}-\frac{Y}{n}\right)\right]$$

$$=\frac{1}{\left(1+\sqrt{2n}\right)^2}\left[(p_1-p_2)^2+\ 2pq+2pq\right]$$

$$=\frac{1}{\left(1+\sqrt{2n}\right)^2}\left[(p_1-p_2)^2+2p_1(1-p_1)+2p_2(1-p_2)\right]$$

$$=f(p_1,p_2)$$

For maxima and minima $\frac{\partial f}{\partial p_i} \quad 0\ for\ all\ i \quad 1,2$ and $\left.\frac{\partial 2f}{\partial p_i^2}\right/_{p_{1+p_2} \quad 1} < 0$

Thus $R(d^{**})$ is maximum when $p_1+p_2 \quad 1$. If $T \quad T(X,Y)$ is any other estimator of (p_1-p_2), then

$$\underset{p_1,p_2}{Sup}\,R(T) \geq \underset{p_1,p_2:p_1+p_2=1}{Sup}\,R(T) \geq \underset{p_1,p_2:p_1+p_2=1}{Sup}\,R(d^{**}) \quad \underset{p_1,p_2}{Sup}\,R(d^{**})$$

$$\underset{T}{inf}\,\underset{p_1,p_2}{Sup}\,R(T) \quad \underset{p_1,p_2}{Sup}\,R(d^{**})$$

Hence d^{**} is minimax estimator of p_1-p_2.

<u>Example</u>: 3.14: If $T(X)$ is a mionimax estimator of θ then show that $h(X) \quad a\,T(X)+b$ is a minimax estimator under the squared error loss function of $a\,\theta+b, a\neq 0, b\ being\ constant$.

<u>Solution</u>: Let $g(X)\not\equiv h(X)$ which is minimax estimator of θ.

$$\underset{\theta}{Sup}\,R(\theta,g)\leq \underset{\theta}{Sup}\,R(\theta,h)$$

$$\underset{\theta}{Sup}E_\theta[g(X)-a\theta-b]^2 \leq \underset{\theta}{Sup}E_\theta[h(X)-a\theta-b]^2$$

$$= \underset{\theta}{Sup}E_\theta[aT(X)+b-a\theta-b]^2$$

$$= \underset{\theta}{Sup}E_\theta[aT(X)-a\theta]^2$$

$$\underset{\theta}{Sup}E_\theta[g(X)-a\theta-b]^2 \leq a^2\, Sup_\theta\, R\big(\theta, T(X)\big)$$

$$a^2\underset{\theta}{Sup}E_\theta\left[\frac{g(X)-b}{a}-\theta\right]^2 \leq a^2\underset{\theta}{Sup}E_\theta[T(X)-\theta]^2$$

$$\underset{\theta}{Sup}E_\theta[\Phi(X)-\theta]^2 \leq a^2\underset{\theta}{Sup}E_\theta[T(X)-\theta]^2, \Phi(X) \quad \frac{g(X)-b}{a}$$

Thus,

$$\underset{\theta}{Sup}\, R(\theta,\Phi) \leq \underset{\theta}{Sup}\, R(\theta,T)$$

This contradicts the fact that $T(X)$ is minimax. Hence $h(X)$ is minimax estimator of $a\,\theta + b$.

BLOCK 2: OPTIMALITY OF DECISION RULES

UNIT 3: ADMISSIBILITY AND COMPLETENESS

UNIT 4: MINIMAXITY AND MULTIPLE DECISION PROBLEMS

4. Admissibility of Decision Rules

Def 4.1: **Natural Ordering**: A decision rule δ_1, is said to be as good as a rule δ_2, if $R(\theta,\delta_1) \leq R(\theta,\delta_2)$ for all θ εΘ. A rule δ_1 is said to be better than a rule δ_2, if $R(\theta,\delta_1) \leq R(\theta,\delta_2)$ for all θ εΘ and $R(\theta,\delta_1) < R(\theta,\delta_2)$ for at least one θ εΘ. A decision rule δ_1 is said to be equivalent to a rule δ_2, if $R(\theta,\delta_1) = R(\theta,\delta_2)$ for all θ εΘ.

Def 4.2: A rule δ is said to be *admissible* if there exist no rule better than δ. A rule δ is said to be in*admissible* if it is not admissible.

We should note that in a given problem every rule may be inadmissible. For example when the risk set S does not contain its boundary points.

Def 4.3: A class C of decision rules, $C \subset D^*$ is said to be complete if for given any rule $\delta \varepsilon D^*$ not in C, there exists a rule $\delta_0 \varepsilon C$ that is better than δ. A class C of decision rules is said to be *essentially complete,* if for given any rule δ (not in C), there exist a rule $\delta_0 \varepsilon C$ that is as good as δ.

Lemma 4.1: if C is a complete class, and A denotes the class of all admissible rules, then A⊂C.

Proof: let $\delta_0 \varepsilon A$ and $\delta \varepsilon D^*$. Since δ_0 is admissible ⇒

$$R(\theta,\delta_0) \leq R(\theta,\delta) \text{ for all } \theta\ \varepsilon\Theta (\text{for all } \delta\ \varepsilon D^* \neq \delta_0)$$

And $\quad R(\theta,\delta_0) < R(\theta,\delta)$ for some θεΘ

By definition of completeness, $\delta_0 \varepsilon C$.

Def 4.4: A class C of decision rules is said to be *minimal completes* if C is a complete class and if no proper subset of C is complete.

Similarly, a class C is minimal *essentially complete* class if C is essentially complete and no proper subset of C is essentially complete.

Theorem 4.1: If a minimal complete class exists, it consists of exactly the admissible rules.

Proof: let C denote the minimal complete class and A denote the class of all admissible rules. We are to show that C=A. Lemma 4.1 implies that A⊂C because minimal complete class is complete. We must show that C⊂A. This is done by assuming it to be false and arriving at a contradiction. Let $\delta_0 \varepsilon C$ and suppose that $\delta_0 \notin A$. We assert that there exists a $\delta_1 \varepsilon C$ that is better than δ_0. Because δ_0 is inadmissible, there exist a δ better than δ_0. If $\delta \varepsilon C$, we may $take, \delta_1 \quad \delta. If\ \delta \notin C, then$ because C is complete, there exists a, $\delta_1 \varepsilon C$ that is better than δ, hence better than δ_0. In either case our claim is verified. Now let $C_1 \quad C \sim \{\delta_0\}$. We will show that C_1 is complete, contradicting the fact that C is minimal. Let δ be an arbitrary rule not in C_1. If δ=δ_0, then $\delta_1 \varepsilon C_1$ is better than δ. If $\delta \neq \delta_0$, there exists $a\ \delta' \varepsilon C\ that$ is better than δ. if $\delta' = \delta_0$, then $\delta_1 \varepsilon C_1$ is better than δ. If $\delta' \neq \delta_0$, there exists $a\ \delta' \in C\ that$ is better than δ. In any case, there exists an element of C_1 better than δ. Which proves that C_1 is complete. This is a contradiction.

Admissibility of Bayes Rules

Theorem 4.2: Assume that $\theta \quad (\theta_1, \theta_2, \dots, \theta_k)$ and a Bayes rule δ_0 w.r.to the prior distribution $(p_1, p_2, \dots, p_k)$ exists. If $p_j > 0$ for j=1,2,...,k, then δ_0 is admissible.

Proof: Suppose that δ_0 is inadmissible, then there exist a $\delta' \varepsilon D^*$

which is better than δ_0. That is,

$$R(\theta_j, \delta') \leq R(\theta_j, \delta_0) \qquad \text{for all j}$$

$$R(\theta_j, \delta') < R(\theta_j, \delta_0) \qquad \text{for some j}$$

Because, all p_j are positive

$$\sum R(\theta_j, \delta')p_j < \sum p_j R(\theta_j, \delta_0)$$

The strict inequality showing that δ_0 is not Bayes w.r.to $(p_1, p_2, \dots, p_k)$. This is a contradiction.

The following counter example shows that δ_0 is not necessarily admissible if the hypothesis $p_j > 0$ for j=1,2,...,k is violated.

<u>Ex 4.1</u>: let $\Theta \quad \{\theta_1, \theta_2\}$, $L(\theta, a)$ as follows:

$L(\theta,a)$		a_1	a_2	a_3	a_4
	θ_1	1	1	2	2
	θ_2	0	1	0	1

$d(0) \quad a_1,\ d(0) \quad a_2,\ d(0) \quad a_3,\ d(0) \quad a_4$

$R(\theta_1, a_1) \quad 1,\ R(\theta_2, a_1) \quad 0, \dots\dots, R(\theta_1, a_4) \quad 2, R(\theta_2, a_4) \quad 1$

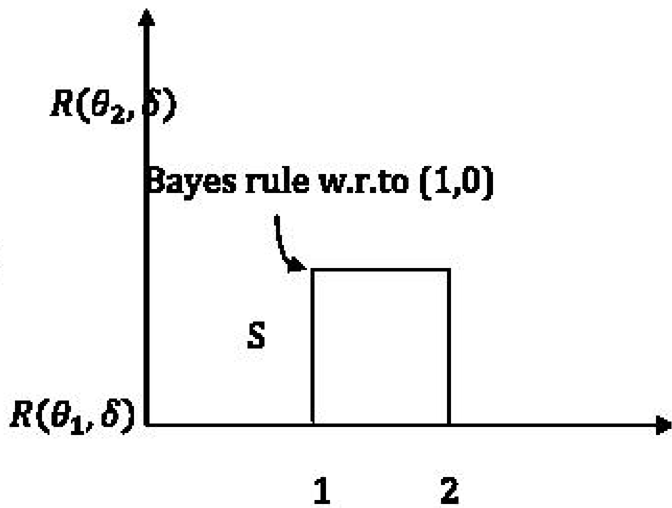
$R(\theta_1, \delta) \quad \sum_{i=1}^{4} \alpha_i R(\theta_1, a_1)$

$S \quad \{R(\theta_1, \delta), R(\theta_2, \delta): \quad \delta \in D^*\}$

$\{(y_1, y_2): \quad 1 \le y_1 \le 2; 0 \le y_2 \le 1\}$

Let the prior distribution, $p_1 \quad 1, p_2 \quad 0$

$\sum_{i=1}^{4} p_i R(\theta_i, \delta) \quad R(\theta_1, \delta) \quad y_1$



Thus, any decision rule that minimizes $\sum p_i R(\theta_i, \delta)$ and that achieved the minimum value =1=y_1 will be a Bayes rule w.r.to prior $(1, 0)$.

Thus the rule $R(\theta_1, \delta_0) \quad R(\theta_2, \delta_0) \quad 1$ is Bayes w.r.to $(1, 0)$.that a_2 and a_1 are Bayes rules w.r.to (1,0). But a_2 is not admissible since

$R(\theta_1, a_2) \le R(\theta_2, a_1)$ and $R(\theta_2, a_2) > R(\theta_2, a_1)$.

<u>Def 4.5</u>: A point θ_0 in E_1(one dimensional Euclidian space) is said to be in support of a distribution τ on the real line if for ∀ ε> 0 the interval $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ has positive probability,

$\tau(\theta_0 - \varepsilon, \theta_0 + \varepsilon) > 0$

Theorem 4.3: let $\theta \varepsilon E_1$ and assume that $R(\theta,\delta)$ is a continuous function of θ for all$\delta \varepsilon D^*$. If δ_0 is a Bayes rule w.r.to a probability distribution τ on the real line, for which ⊠(τ, δ_0) is finite and if the support of τ is the whole real line, then δ_0 is admissible.

Proof: As before, assume that δ_0 is not admissible. Then, there exists a $\delta' \varepsilon D^*$for which

$$R(\theta,\delta') \leq R(\theta,\delta_0) \qquad \text{for all } \theta.$$

$$R(\theta_0,\delta') < R(\theta_0,\delta_0) \qquad \text{for some } \theta_0 \varepsilon E_1.$$

Since $R(\theta,\delta)$ is continuous in θ for all δ. Let

$$\eta \quad R(\theta_0,\delta_0) - R(\theta,\delta') \text{ ....} (4.1)$$

For $|\theta - \theta_0| < \varepsilon \quad ,\varepsilon > 0$

$|R(\theta,\delta) - R(\theta_0,\delta)| < \frac{\eta}{4}$ whenever $|\theta - \theta_0| < \varepsilon \ for\ all\ \delta\varepsilon D^*$

Or $\quad -\frac{\eta}{4} \leq R(\theta,\delta) - R(\theta_0,\delta) \leq \frac{\eta}{4} \qquad |\theta - \theta_0| < \varepsilon$ .... (4.2)

Or $\quad R(\theta,\delta) \leq R(\theta_0,\delta) + \frac{\eta}{4}$

$$R(\theta,\delta') \leq R(\theta_0,\delta') + \frac{\eta}{4} \qquad \text{for all } |\theta - \theta_0| < \varepsilon$$

$$R(\theta,\delta_0) - R(\theta,\delta_0) + R(\theta_0,\delta') + \frac{\eta}{4}$$

$$R(\theta,\delta_0) - [R(\theta,\delta_0) - R(\theta_0,\delta_0) + R(\theta_0,\delta_0) - R(\theta_0,\delta')] + \frac{\eta}{4}$$

$$R(\theta,\delta_0) - [R(\theta,\delta_0) - R(\theta_0,\delta_0)] - [R(\theta_0,\delta_0) - R(\theta_0,\delta')] + \frac{\eta}{4}$$

$$\leq R(\theta,\delta_0) + \frac{\eta}{4} - \eta + \frac{\eta}{4} \quad R(\theta,\delta_0) - \frac{\eta}{2}$$

Thus, $R(\theta,\delta') \leq R(\theta,\delta_0) - \frac{\eta}{2}$ whenever $|\theta - \theta_0| < \varepsilon$

Letting T denote the r.v. whose d.f is τ

⊠(τ, δ_0) − ⊠(τ, δ') $\quad$ E $R(T,\delta_0)$ − E $R(T,\delta')$

$$\beta x_1 + \overline{1-\beta} x_2 \varepsilon\, S_1 \;, \beta y_1 + \overline{1-\beta} y_2 \,\varepsilon\, S_2 \Rightarrow S \text{ is convex.}$$

2. $0 \notin S$ For if $0 \in S$, there could be point x $\varepsilon\, S_1$, y $\varepsilon\, S_2$, such that (x-y)=0 ⇒ x=y contradicts that S_1 and S_2 are disjoint.
3. From Theorem (4.6) there exists a vector $P \neq 0$ such that $P^T Z \geq 0$ for all Z $\varepsilon\, S$. Thus$P^T(x-y) \geq 0$ *for all* x $\varepsilon\, S_1$, y $\varepsilon\, S_2$, completing the proof.

Lemma 4.4: If S is a convex sub set of E_k and Z is a k-dimensional random vector for which E (Z) exists and is finite, then EZ∈S.

Proof: Let Y=Z-EZ and let S′ be the translation of S by E Z, i.e S' $\{Y: Y \quad Z - EZ \;\; for\ all\ Z\varepsilon S\}$. Thus S′ is convex $P[Y\varepsilon S'] = 1$ and EY=0. We will show that$0 \in S'$. We prove by induction method. The Lemma is trivially true for k=0 in which case Y is degenerate at zero. Now suppose the Lemma is true for k-1. We are to show that Lemma is true for k≥ 1.

Suppose $0 \notin S'$ then by Theorem (4.6) there exists a vector $P \neq 0$ such that $P^T Y \geq 0$ for all Y ε S′. Let U=$P^T Y$. The r.v. U has expectation 0, and $P[U \geq 0] \quad 1 \Rightarrow P[U \quad 0] \quad 1$, then with probability one Y lies in the hyper plane$P^T Y \quad 0$. Let

$S'' \quad S' \cap \{y: P^T Y \quad 0\}$ Then S″ is convex subset of (k-1) dimensional Euclidian space for which $P[Y\varepsilon S''] \quad 1$ *and* $EY \quad 0$

By the induction, 0εS″. Since $S'' \subset S' \Rightarrow 0\varepsilon S'$ which is contradiction of the assumption 0∉ S′. #

Corollary: S is a convex hull of S_0.

Lemma 4.5: (Jensen's Inequality): Let f(x) be a convex real-valued function defined on a non empty convex subsets of E_k and let Z be a k-dimensional random-vector with finite expectation E Z for which $P[Z \in S] = 1$. Then E(Z)∈S and $f[E(Z)] \leq E[f(Z)]$ .... (4.8)

Proof: for k=1, the point $(EZ, f\,(EZ))$ is on the boundary of the convex set S_1.

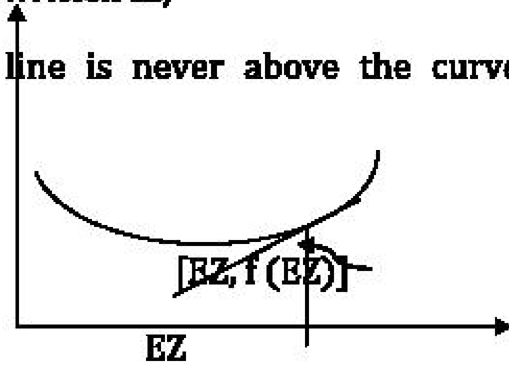
$$S_1 \quad \left\{ \begin{matrix} (Z_1, Z_2, \ldots, Z_{k+1})^T \; for\ some\ x\varepsilon S, x^T & (Z_1, Z_2, \ldots, Z_{k+1}) \\ & and\ f(x) \leq Z_{k+1} \end{matrix} \right\} . (4.9)$$

Hence there exists a supporting hyper plane (straight line) at

$(EZ, f\ (EZ))$. Call this $y \quad m\,x + \ c$

Because $(EZ, f\ (EZ))$ is on this line. It may be written as,

$Y \quad f(EZ) + m(x - EZ)$ And because this line is never above the curve y=f(x) we have, f(x)



$f(x) \geq f(EZ) + m(x - EZ)\ for\ all\ x.$

$f(Z) \geq f(EZ) + m(Z - EZ) \quad for\ Z \varepsilon\ S.$

$E(f(Z)) \geq f(EZ)$

Thus theorem is true for k=1. Suppose theorem is true for k-1, we prove for k≥ 1.

Since EZ ε S, the point $(EZ, f\ (EZ))$ is boundary point of the convex set S_1 defined (4.9) hence by supporting hyper plane theorem, there exists $a(k + 1)$-dimensional vector $P \neq 0$ such that,

$P^T Z \geq P^T(EZ, f\ (EZ)) \quad or$

$\sum_{j=1}^{k+1} p_j\, z_j \geq \sum_{j=1}^{k} p_j\, Ez_j + p_{k+1} f(Ez)\ \ for\ all (Z_1, \ldots, Z_k)^T \varepsilon S_1$. (4.10)

We note that; p_{k+1} can not be negative, for letting $Z_{k+1} \to \infty$ the inequality (4.10) will not be satisfied. Replacing Z_{k+1}

$with\ f(Z), Z \quad (Z_1, \ldots, Z_k) \varepsilon S$ and Z with random vector Z.

$p_{k+1} f(EZ) \leq \sum_{j=1}^{k+1} p_j\, (z_j - Ez_j) + p_{k+1} f(Z)$ .... (4.11)

If $p_{k+1} > 0$ taking the expectation.

$p_{k+1} f(EZ) \leq p_{k+1} Ef(Z) \Rightarrow f[E(Z)] \leq E[f(Z)]$

If $p_{k+1} \quad 0$ (4.11) ⇒ the random vector

$U \quad \sum p_j(z_j - Ez_j) \quad P^T(z - Ez)$ is non-negative and EU=0⇒P[U=0]=1 that gives all its mass to the (k-1) dimensional convex set $S' \quad S \cap \{Z: \sum p_j(z_j - Ez_j) \quad 0\}$ by induction method, theorem is proved.

Theorem 4.8: Let â be a convex subset of E_k and let L (θ, a) be a convex function of a ε â for all θ ε Θ there exist a ε> 0 and a c such that $L(\theta', a) \geq \varepsilon|a| + c$, then for every P ε â*, there exist *an* $a_0 \varepsilon$â such that $L(\theta, a_0) \leq L(\theta, P)$ for all θ ε Θ.

Proof: P ε â* and Z be a random vector with values in â when distribution is given by P. then EZ infinite since,

$\varepsilon E|Z| + c \leq EL(\theta', Z) \quad L(\theta', P) < \infty$ By definition of â*.

$L(\theta, P) \quad EL(\theta, Z) \geq L(\theta, EZ) \quad L(\theta, a_0)$ *Where*, $a_0 \quad EZ\varepsilon$â.

Remark: If the loss is convex we can always concerned with non-randomized decision rules. The non-randomized decision rules form a complete class.

Exp 4.2: $\theta =$ â $= [0,1]$, â is convex set.

$L(\theta, a) \quad (\theta - a)^2$ is convex loss function.

X has b= (2, θ)

$$P_\theta[X \quad x] \quad \binom{2}{x}\theta^x(1-\theta)^{2-x} \qquad x \quad 0,1,2$$

$$d_1(x) \quad \frac{x}{2} \quad d_2(x) \quad \frac{1}{2} \qquad for\ all\ x \quad 0,1,2$$

$$P[Z \quad d_1] \quad \frac{1}{2} \qquad P[Z \quad d_2] \quad \frac{1}{2}$$

$$E[Z] \quad \frac{d_1 + d_2}{2} \quad \frac{x+1}{4} \quad d$$

$$R(\theta, d) \quad EL(\theta . d(x)) \quad E(\theta - \tfrac{x+1}{4})^2$$

$$= \theta^2 + E(\tfrac{x+1}{4})^2 - 2\theta E(\tfrac{x+1}{4})$$

$$= \theta^2 + \tfrac{1}{16}[Ex^2 + 1 + 2Ex] - \tfrac{\theta}{2}(E(x) + 1)$$

$$= \theta^2 + \frac{1}{16}[2\theta(1-\theta) + 4\theta^2 + 1 + 2.2\theta] - \frac{\theta(2\theta+1)}{2}$$

$$= \frac{16\theta^2 + [2\theta - 2\theta^2 + 4\theta^2 + 1 + 4\theta] - 16\theta^2 - 8\theta}{16} \quad \frac{[2\theta^2 - 2\theta + 1]}{16}$$

Let d_0 be a randomized decision rule choosing d_1 with prob. $\frac{1}{2}$ and

d_2 with prob. $\frac{1}{2}$

$$R(\theta, d_0) \quad \frac{1}{2}[R(\theta, d_1) + R(\theta, d_2)]$$

$$= \frac{1}{2}\left[\frac{1}{2}\theta(1-\theta) + \frac{1}{4}(4\theta^2 - 4\theta + 1)\right] \quad \frac{1}{8}(2\theta^2 - 2\theta + 1)$$

Obvious, $R(\theta, d) \leq R(\theta, d_0)$ *as*

$$\frac{[2\theta^2 - 2\theta + 1]}{16} \leq \frac{(2\theta^2 - 2\theta + 1)}{8}$$

$2\theta^2 - 2\theta + 1 \geq 0 \qquad 1 - 2\theta(1-\theta) \geq 0$

as the maximum value of, $\theta(1-\theta)$ 1/4. Thus the inequality is always true.

<u>Complete class theorem</u>:

<u>Theorem 4.9</u>: (converse of theorem 4.2): If δ is admissible and Θ is finite, then δ is Bayes w.r.to some prior distribution τ.

<u>Proof</u>: If δ is admissible, then $Q_x \cap S$ $\{x\}$ where x $\{R(\theta_1, \delta), \ldots\ldots\ldots, R(\theta_k, \delta)\}$ as $S \subset \bar{S} \Rightarrow Q_x \cap S \subset Q_x \cap \bar{S}$ $\{x\}$. And x ε S. thus, because Q_x -{x} and S are disjoint convex sets, there exists a vector P≠ 0 such that $P^T y \leq P^T z$ for all $y \varepsilon Q_x$ -{x}, and z ε S. If some coordinate p_j of vector P were negative then by taking y so that y_j sufficiently negative, we would have $P^T y < P^T x$. Hence $p_j \geq 0$ for all j. we may normalize P so that $\sum p_j = 1$. Because P is now a probability

Distribution over Θ and $\sum p_j R(\theta_j, \delta) \leq P^T Z$ for all Z ε S, δ is a Bayes rule w.r.to P.

<u>Theorem 4.10</u>: (Complete class theorem): If for a given decision problem (Θ, D, R) with finite Θ, the risk set S is bounded from below and closed from below, then the class of all Bayes rules is complete and admissible Bayes rules form a minimal complete class.

<u>Exp 4.3</u>: $\theta \quad \{\theta_1, \theta_2\} \quad$ â $\quad [0,1]$

$$L(\theta_1, a) \quad a^2,\ L(\theta_2, a) \quad 1 - a$$

(Note that loss function is convex in a, for each Θ)

$$P_{\theta_1}(H) \quad \tfrac{1}{3} \qquad P_{\theta_2}(H) \quad \tfrac{2}{3}$$

1. Represent the class D rules as a subset of the plane.
2. Find the class of all non-randomized rules.
3. Find minimax Bayes rules.

<u>Solution</u>: $D \quad \{d: x \rightarrow [0,1]\}\ where\ x \quad \{H, T\}$

Let $d(H) = x$, $d(T) = y$ with the interpretation that we estimate Θ to be x when H is observed and y when T is observed.

$$D \quad \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

This is a square in the plane (x, y).

$$R(\theta_1, d) \quad EL(\theta_1, (x, y))$$

$$L(\theta_1, x)P\left[{}^{H}/_{\theta_1}\right] + L(\theta_1, y)P\left[{}^{T}/_{\theta_1}\right]$$

$$x^2\tfrac{1}{3} + y^2\tfrac{2}{3} \quad \tfrac{1}{3}(x^2 + 2y^2) \text{ .... } (4.12)$$

$$R(\theta_2, d) \quad EL(\theta_2, (x, y))$$

$$= L(\theta_2, x)P\left[{}^{H}/_{\theta_2}\right] + L(\theta_2, y)P\left[{}^{T}/_{\theta_2}\right]$$

$$= (1 - x)\tfrac{2}{3} + (1 - y)\tfrac{1}{3} \quad \tfrac{1}{3}(3 - 2x - y) \text{ .... } (4.13)$$

Let (p) and (1-p) be the probability distribution $\theta \quad \{\theta_1, \theta_2\}$ i.e choosing θ_1 with prob. (p) and choosing θ_2 with prob. (1-p).

$$\square\left(\tau,(x,y)\right) \quad ER(\theta,(x,y))$$

$$= pR\left(\theta_1,(x,y)\right) + 1 - pR(\theta_2,(x,y))$$

$$\tfrac{p}{3}(x^2 + 2y^2) + \tfrac{1-p}{3}(3 - 2x - y)$$

$$\tfrac{p}{3}(x^2 + 2y^2 + 2x + y - 3) + \tfrac{1}{3}(3 - 2x - y) \text{ ....} \quad (4.14)$$

Set of Bayes rules which minimizes (4.14) will be obtained as,

$$(2x + 2)\tfrac{p}{3} - \tfrac{2}{3} \quad 0 \Rightarrow x \quad \tfrac{1-p}{p} \quad \&$$

$$(4y + 1)\tfrac{p}{3} - \tfrac{1}{3} \quad 0 \Rightarrow y \quad \tfrac{1}{4}\left(\tfrac{1-p}{p}\right)$$

Then the set of Bayes rules are,

$$B \quad \left\{\left(\alpha, \tfrac{\alpha}{4}\right) : 0 \leq \alpha \leq 1\right\} \subset D.$$

Now to find minimax Bayes rule, we should have (4.12) = (4.13) for $\left(\alpha, \frac{\alpha}{4}\right) \in B \Rightarrow$

$$\frac{1}{3}(\alpha^2 + \tfrac{2\alpha^2}{16}) \quad \frac{1}{3}(3 - 2\alpha - \tfrac{\alpha}{4})$$

$$\tfrac{9\alpha^2}{18} \quad 3 - 2\alpha - \tfrac{\alpha}{4} \Rightarrow 9\alpha^2 + 18\alpha - 24 \quad 0 \Rightarrow 3\alpha^2 + 6\alpha - 2 \quad 0$$

$$\alpha \quad \tfrac{-6 \pm \sqrt{36+96}}{6} \quad -1 \pm \tfrac{5.74}{3} \quad 0.91, \qquad \text{as } \alpha \geq 0$$

$$\tfrac{1-p}{p} \quad 0.91 \Rightarrow p \quad 0.52\ (approx.)$$

Hence (0.52, 0.48) is prior distribution function (0.91, 0.23) is Bayes rule and since for this (x, y) risk is constant have (0.91, 0.23) is minimax Bayes rule.

<u>Example</u>: 4.4

<u>Admissibility of $\bar{X}$ for estimating normal mean:</u>

<u>First proof</u>: (the limiting Bayes method): Suppose $\bar{X}$ is not admissible, and without loss of generality we may assume σ=1. Then there exists δ^* such that

$$\left.\begin{aligned} R(\theta,\delta^*) &\le \tfrac{I}{n} \ \ for\ all\ \theta \\ &< \tfrac{I}{n} \ \ for\ some\ \theta \end{aligned}\right\} \ (under\ the\ square\ error\ loss\ function)$$

R (θ, δ) is a continuous function of θ for every δ, so that there exist

$\varepsilon > 0$ *and* $\theta_0 < \theta_1$ such that

$R(\theta,\delta^*) \le \frac{I}{n} - \varepsilon$ *for all* $\theta_0 < \theta < \theta_1$ (as in Theorem 4.3)

Let γ_T^* be the average Bayes risk of δ^* with respect to prior distribution $\tau \sim N(0,T^2)$ and let γ_T be the Bayes risk of the Bayes decision rule with respect to $N(0,T^2)$. Thus by exp. 3.11 for σ=1

$$\frac{\frac{1}{n}-\gamma_T^*}{\frac{1}{n}-\gamma_T} \quad \frac{\frac{1}{\sqrt{2\pi}T}\int_{-\infty}^{\infty}\left[\frac{1}{n}-R(\theta,\delta^*)\right]e^{\frac{-\theta^2}{2T^2}}d\theta}{\frac{1}{n}-\frac{T^2}{1+nT^2}}$$

$$\ge \frac{n(1+nT^2)\varepsilon}{T\sqrt{2\pi}}\int_{\theta_0}^{\theta_1} e^{\frac{-\theta^2}{2T^2}}d\theta \qquad \text{.................. (4.15)}$$

By Lebesgue dominated convergence theorem, as the integral

$e^{\frac{-\theta^2}{2T^2}} \to 1$ As T→ ∞, the integral converges to $(\theta_1-\theta_0)$ and the

R.H.S → ∞ ⇒ $\frac{\frac{1}{n}-\gamma_T^*}{\frac{1}{n}-\gamma_T} \to \infty$ thus there exist T_0 such that, $\gamma_{T_0}^* < \gamma_{T_0}$, which contradicts the fact that γ_{T_0} is the Bayes risk for $N(0,T_0^2)$.

<u>Second proof</u>: (the information inequality method):

$R(\theta,\delta) \quad E(\delta-\theta)^2 \quad var_\theta(\delta)+b^2(\theta)$, where $b(\theta) \quad E_\theta(\delta)-\theta$

$$\ge b^2(\theta)+\frac{[1+b'(\theta)]^2}{nI(\theta)} \quad \text{by F C R bound. .... (4.16)}$$

In the present case $\sigma^2 \quad 1, I(\theta) \quad 1$

Suppose now δ is any estimator satisfying

$$R(\theta,\delta) \le \tfrac{1}{n}\ For\ all\ \theta \text{.............. (4.17)}$$

and hence, $b^2(\theta) + \frac{[1+b'(\theta)]^2}{nI(\theta)} \le \frac{1}{n}$ *for all* θ .... (4.18)

We shall then show that (4.18) $\Rightarrow b(\theta) \equiv 0$ for all θ. i.e δ is unbiased.

1. Since $|b(\theta)| \le \frac{1}{\sqrt{n}}$ the function b is bounded.
2. From the fact that $1 + b'^2(\theta) + 2b'(\theta) \le 1 \Rightarrow b'(\theta) \le 0$ so that b is non-increasing.
3. Next, there exists a sequence of $\theta_i \to \infty$ and such that $b'(\theta_i) \to 0$

For suppose that $b'(\theta)$ were bounded away from 0 as $\theta \to \infty$, $say\, b'(\theta) \le -\varepsilon\, for\, all\, \theta$, then $b(\theta)$ can not be bounded

as θ→ ∞, which contradicts 1.

4. Analogically it is seen that there exist a square $\theta_i \to -\infty$ and such that $b'(\theta_i) \to 0$ .Thus $b(\theta) \to 0$ as $\theta \to \pm\infty$ with inequality (4.18). Thus $b(\theta) \equiv 0$ follows from 2.

   Since $b(\theta) \equiv 0 \Rightarrow b'(\theta) \quad 0\, for\, all\, \theta \Rightarrow (4.16)\, as\, R(\theta,\delta) \le \frac{1}{n}$ *For all* θ *and hence* $R(\theta,\delta) \equiv \frac{1}{n}$

This proves that $\bar{X}$ is admissible and minimax. This is unique admissible and minimax estimator. Because if δ' is any other estimator such that $R(\theta,\delta') \equiv \frac{1}{n}$. Then let $\delta^* \quad \frac{1}{2}(\delta + \delta')$

$$R(\theta,\delta^*) < \frac{1}{2}[R(\theta,\delta) + R(\theta,\delta')] \quad R(\theta,\delta)$$

Which contradicts that δ is admissible. Thus δ=δ' with prob. 1.

5. Conjugate family of prior distribution:

For the choice of prior densities on the parameter space, Raiffa & Schlaifer discuss as important class of densities called Natural Conjugates.

If the conditional distribution $F(x/\theta)$ has a mass function, we shall denote by $\rho(x/\theta)$, the prob. given θ that the experimental results in x. If the conditional distribution $F(x/\theta)$ has a density function, we shall denote by $\rho(x/\theta)$, the value of this density function at x for a given θ. In other case, we should use the word likelihood to denote $\rho(x/\theta)$ as a function of θ, for given x.

we assumed that, for any fixed x in the sample space $\mathfrak{X}$, $\rho(x/.)$ as a function on Θ is continuous except for at most a finite number of discontinuities (which may depend on x).

Def 5.1: The marginal likelihood of the experimental out come x given a particular prior density g (θ) is given by,

$$\rho^*(x/g) \quad \int_\Theta \rho(x/\theta) g(\theta)\, d\theta \quad \ldots\ldots\ldots\ldots\ldots\ldots (5.1)$$

and we shall say that x lies in the spectrum of g if $\rho^*(x/g) > 0$.

Def 5.2: If the density function of θ is g (θ) and if K is other function of θ such that g (θ) $\frac{K(\theta)}{\int_\Theta K(\theta)d\theta}$ .... (5.2)

That is if the ratio $\frac{K(\theta)}{g(\theta}$ is constant as regards θ, we shall write

$$g(\theta) \propto K(\theta) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5.3)$$

and say that K is a kernel of the density of θ.

Def 5.3: If the likelihood function of x given θ is $\rho(x/\theta)$ and if ρ and K are functions on $\mathfrak{X}$, such that for all x and θ,

$$\rho(x/\theta) \quad K(x/\theta)\rho(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots (5.4)$$

That is if the ratio $\frac{K(x/\theta)}{\rho(x/\theta)}$ is constant as regards θ, we shall say that $K(x/\theta)$ is a kernel of the likelihood function of x given θ, and $\rho(x)$is a residue of the likelihood function.

If the prior distribution of the random variable θ has a density function g (θ), and if the experimental out come x is finite in the spectrum of g, then it follows from Bayes theorem that the posterior distribution of θ has a density function $g(\theta/x)$ whose value at θ for given x is ,

$$g(\theta/x) = g(\theta)\rho(x/\theta)N(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots (5.4)$$

where, N(x) is simply the normalized constant defined by the condition,

$$\int_\Theta g(\theta/x)\, d\theta \quad N(x)\int_\Theta \rho(x/\theta) g(\theta)\, d\theta \quad 1 \ldots\ldots\ldots\ldots (5.5)$$

Letting K′ denote a kernel of the prior density of Θ, it follows from the definitions (5.1), that the Bayes formula (5.4) can be written as,

$$g(\theta / x) \quad g(\theta)\rho(x/\theta)N(x)$$

$$K'(\theta)\left[\int_{\Theta} K'(\theta)d\theta\right]^{-1} \; k(x/\theta)\rho(\theta)N(x)$$

$$\propto K'(\theta)k(x/\theta) \text{ .... } (5.6)$$

The value of the constant of probability for given x, that is,

$\rho(x)N(x)\left[\int_{\Theta} K'(\theta)d\theta\right]^{-1}$ can always be determined by the condition, $\int_{\Theta}$ g(θ/x) dθ 1

<u>Def 5.4</u>: A statistics T is said to be sufficient in Bayesian sense if for any prior density or mass function g (θ) and x in the spectrum of g,

$$\text{g}(\theta/T \quad t) \quad \text{g}(\theta/X \quad x) \quad □(T(x),\theta) \text{ Where T=T(x) ...... (5.7)}$$

<u>Theorem 5.1</u>: A statistics T is said to be sufficient in Bayesian sense for a family of generalized $p.d.f's\,\{f(./\theta),\theta\varepsilon\Theta\}$ if and only if $f(x/\theta)$, can be factored as follows for all values of x ε 𝔁 and θ ε Θ

$$f(x/\theta) \quad \rho(x/\theta) \quad K(t/\theta)\rho(x) \text{ .... } (5.8)$$

$$K(t,\theta)\rho(x)$$

where $K(t,\theta)$ $K(t/\theta)$ is a kernel of the likelihood function

$K(x/\theta)$ *and* $\rho(x)$ is the residue function.

<u>Proof</u>: First suppose that the factorization inducted in eqn. (5.8) is correct then for any generalized p.d.f(gpdf) $g\ (\theta)$ of Θ and point x ε 𝔁 and θ ε Θ the posterior gpdf of Θ is,

$$g(\theta/x) \quad g(\theta)\rho(x/\theta)N(x)$$

$$g(\theta)k(t,\theta)\rho(x)N(x) \qquad \frac{k(t,\theta)g(\theta)}{\int_{\Theta} g(\theta)k(t,\theta)d\theta}$$

$$as \quad \rho(x)N(x) \quad [\int_{\Theta} k(t,\theta)g(\theta)d\theta]^{-1}$$

$$g(\theta/x) \quad \frac{k(t,\theta)g(\theta)}{\int_{\Theta} k(t,\theta)g(\theta)d\theta} \quad \text{.......................} \; (5.9)$$

Since the R.H.S. of (5.9) depends on the observed value x only through the value T(x) =t, it follows that T is sufficient statistics in Bayesian sense i.e $g(\theta/x_1) \quad g(\theta/x_2)$ if $T(x_1) \quad T(x_2)$ $x_1, x_2 \varepsilon \mathfrak{x}$ and any prior gpdf $g\ (\theta)$.

Conversely that T is sufficient statistics. Let $g\ (\theta)$ be any gpdf of Θ such that $g\ (\theta) > 0$ at every point of Θ. The posterior gpdf $g(\theta/x)$ is specified at any point x ε $\mathfrak{x}$ and θ ε Θ as,

$$g(\theta/x) \quad g(\theta)\rho(x/\theta)N(x)$$

$$= \frac{g(\theta)\rho(x/\theta)}{\int_{\Theta} g(\theta)\rho(x/\theta)d\theta} \quad as\ N(x) \quad [\int_{\Theta} g(\theta)\rho(x/\theta)d\theta]^{-1}$$

$$\rho(x/\theta) \quad \frac{g(\theta/x)}{g(\theta)} \int_{\Theta} g(\theta)\rho(x/\theta)d\theta$$

$$\frac{■(T(x),\theta)}{g(\theta)} \int_{\Theta} g(\theta)\rho(x/\theta)d\theta$$

as T is sufficient in the Bayesian sense.

$$\rho(x/\theta) \quad k(t,\theta)\rho(x) \quad \#$$

<u>Def 5.5</u>: Let the kernel function k defined by (5.8) be a function $k(t/.)$ $k(t,.)$ with parameter t on the state space (parameter space) Θ. Let $\rho(\theta/t) \quad N(t)k(t/\theta)$ .... (5.10)

Where N is a function of t determined by

$\int_{\Theta} g(t/\theta)d\theta \quad 1$ The density function $g(./t)$ on Θ will be called Natural Conjugate Bayesian density (NCBD) with parameter t.

<u>Theorem 5.2</u>:

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a Bernoulli distribution with an unknown value of the parameter Θ. Suppose that the prior distribution of Θ is a beta distribution with parameter α and β such that α>0, β>0. Then the posterior distribution of θ when $X_i \quad x_i$ ($i \quad 1,2,\dots,n$) is a beta distribution with parameter $\alpha + y \; and \; \beta + n - y$, where $y \quad \sum x_i$.

<u>Proof:</u> $f(x_1, x_2, \dots, x_n/\theta) \quad \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$

The joint p.d.f of $(x_1, x_2, \dots, x_n)$ and θ is,

$$f(x_1, x_2, \dots, x_n, \theta) \quad f(x_1, x_2, \dots, x_n/\theta) g(\theta)$$

$$\propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^{\alpha+\sum x_i - 1}(1-\theta)^{\beta+n-\sum x_i - 1}$$

$f(x_1, x_2, \dots, x_n, \theta) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$ Where $y \quad \sum x_i$

The constant of probability is obtained as,

$$\int_\Theta f(\theta, x_1, \dots, x_n) d\theta \quad 1.$$

<u>Construction of the Conjugate family:</u>

Consider again the example summarized in theorem (5.2) consider for any positive constants α and β, $g(./\alpha, \beta)$ denote the p.d.f of a beta distribution with parameter α and β.

Consider any observational value $x_1, x_2, \dots, x_n$ of the variables $X_1, X_2, \dots, X_n$. The conditional joint p.d.f $f_n(x_1, x_2, \dots, x_n/\theta)$

of $X_1, X_2, \dots, X_n$ is specified by,

$f_n(x_1, \dots, x_n/\theta) \quad \theta^y(1-\theta)^{n-y}$ where, $y \quad \sum x_i$ ……….. (5.11)

If this function is regarded as a function of θ, then it follows from the distribution of a beta distribution that,

$f_n(x_1, \dots, x_n/\theta) \propto g(\theta/y+1, n-y+1)$ ………………. (5.12)

Therefore any observed value $x_1, x_2, \dots, x_n$ the function $f_n(x_1, \dots, x_n/\theta)$ is proportional to the p.d.f of beta distribution.

If $g(./\alpha_1,\beta_1)$ and $g(./\alpha_2,\beta_2)$ are the p.d.f's of any two beta distributions, then there is an another p.d.f of $g(./\alpha_3,\beta_3)$ such that for 0<θ<1,

$$g(\theta/\alpha_3,\beta_3) \propto g(\theta/\alpha_1,\beta_1)g(\theta/\alpha_2,\beta_2) \quad \dots\dots\dots\dots\dots (5.13)$$

Or $g(\theta/\alpha_3,\beta_3) \propto \theta^{\alpha_1+\alpha_2-2}(1-\theta)^{\beta_1+\beta_2-2}$ .... (5.14)

(By the def. of beta distribution)

$$\alpha_3 \quad \alpha_1+\alpha_2-1, \qquad \beta_3 \quad \beta_1+\beta_2-1$$

A family of prior distributions is taken to be beta distribution $g(\theta/\alpha,\beta)$ then the posterior distribution of θ will satisfy the relation,

$$g(\theta/x_1,x_2,\dots,x_n) \propto f_n(x_1,\dots,x_n/\theta)g(\theta/\alpha,\beta)$$

$$\propto \theta^{y+1-1}(1-\theta)^{n-y+1-1}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Or $\theta^{\alpha+y-1}(1-\beta)^{n+\beta-y-1}$

Or $g(\theta/\alpha+y,n+\beta-y)$

$i.e$ $g(\theta/x_1,x_2,\dots,x_n) \propto g(\theta/\alpha+\beta,n+\beta-y)$.

This development suggest a method for determining a conjugate family of distribution in any problem for which there exist a sufficient statistics of fixed dimensional. The statistician need only to determine a family of p.d.f's of the parameter θ such that,

1. For any sample size n and any observed value $x_1,x_2,\dots,x_n$ the conditional joint p.d.f $f_n(x_1,\dots,x_n/\theta)$ regarded as a function of θ is proportional to one of p.d.f's in the family.
2. The family is closed under the multiplication.

Let T$(x_1,x_2,\dots,x_n)$ be the sufficient statistics for the family of p.d.f's $\{f_n(./\theta)\}$ by factorization theorem,

$f_n(x_1,\dots,x_n/\theta) \propto k_n(t,\theta) \qquad t \quad T(x_1,x_2,\dots,x_n)$ .... (5.15)

And assume that $\int_{\theta} k_n(t,\theta)dv(\theta) < \infty$

Let $g(\theta/t,n) = \frac{k_n(t,\theta)}{\int_\theta k_n(t,\theta)dv}$

or there exists $g(\theta/t,n) \propto k_n(t,\theta)$ .... (5.16)

Consider the family of p.d.f's $g(./t,n)$. For all possible sample size n and all possible values of the statistics $T(x_1, x_2, \dots, x_n)$ it follows from the relation (5.15) and (5.16) that $f_n(x_1, \dots, x_n/\theta)$ must be proportional to one of p.d.f's in the family.

Now consider any two p.d.f's $g(./s,m)$ and $g(./t,n)$ which belong to the some family. Thus there must exist observed values $(x_1, x_2, \dots, x_n)$ and $(y_1, y_2, \dots, y_n)$ of the sample sizes m and n such that $T_m(x_1, x_2, \dots, x_n)$ s and$T_n(y_1, y_2, \dots, y_n) = t$. If the observations are combined, then form a sample size (m +n) and the p.d.f satisfy the equation.

$$f_{m+n}(x_1, \dots, x_n, y_1, y_2, \dots, y_n)/\theta) \quad f_m(x_1, \dots, x_n/\theta) f_n(y_1, \dots, y_n/\theta)$$

.... (5.17)

If we let, u $T_{m+n}(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$ then it follows from (5.15) to (5.17)

$$g(\theta/u, m+n) \propto g(\theta/s,m) g(\theta/t,n)$$

Therefore, family is closed under the multiplication.

<u>Theorem 5.3</u>:

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from a Poisson distribution with an unknown value of the mean θ, suppose also that the prior distribution of θ is a gamma distribution with parameters α and β, such that α> 0, β> 0. Then the posterior distribution of θ when, X_1 $x_1(z$ $1,2,\dots,n)$ is a gamma distribution with parameters $\alpha + \sum x_i$ and β +n.

<u>Proof</u>: $f_n(x_1, \dots, x_n/\theta) \quad \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}$

$$f_n(x_1, \dots, x_n/\theta) \propto e^{-n\theta}\theta^{\sum x_i} \quad e^{-n\theta}\theta^{y} \ , \ y = \sum x_i$$

Let $\quad g(\theta) \propto e^{-\beta\theta}\theta^{\alpha-1} \qquad \alpha, \beta > 0$

$$f_n(\theta/x_1, \dots, x_n) \propto f_n(x_1, \dots, x_n/\theta)g(\theta)$$

$$e^{-n\theta}\theta^{y}e^{-\beta\theta}\theta^{\alpha-1}$$

$$e^{-(n+\beta)\theta}\theta^{\alpha+y-1}$$

$$f(\theta/x_1, \dots, x_n) \propto e^{-(n+\beta)\theta}\theta^{\alpha+y-1} \sim G(\alpha + y, n + \beta) \quad \#$$

Theorem 5.4:

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from an exponential distribution with an unknown value of the parameter θ, suppose also that the prior distribution of θ is a $G(\alpha, \beta)$.

Then the posterior distribution of θ given, $X_1 \quad x_i(i \quad 1,2, \dots, n)$ is

$G(\alpha + x, \beta + y)$. $y=\sum x_i$

Proof: $f(x_1, \dots, x_n/\theta) \quad \prod_{i=1}^{n} e^{-\theta x_i}\theta \quad \theta^n e^{-\theta\sum x_i}$

$$f(x_1, \dots, x_n/\theta) \propto \theta^n e^{-\theta\sum x_i}$$

$$g(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}$$

$$f(\theta/x_1, \dots, x_n) \propto f_n(x_1, \dots, x_n/\theta)g(\theta)$$

$$\propto \theta^{\alpha+n-1}e^{-(\beta+\sum x_i)\theta} \sim G(\alpha + n, \beta + \sum x_i) \quad \#$$

Theorem 5.5:

Suppose that $X_1, X_2, \dots, X_n$ is a random sample from $N(\theta, \sigma_0^2)$, σ_0^2 is specified. Suppose also that the prior distribution of θ is $N(\mu, \tau^2)$.

Let $\frac{1}{\tau^2} \quad \tau', \frac{1}{\sigma_0^2} \quad$ ▯. Then the posterior distribution is $N(\mu', \tau' + n▯)$

Where, $\mu' \quad \frac{\tau'\mu + n▯\bar{x}}{\tau' + n▯}$

Proof: $f(x_1, \dots, x_n/\theta) \quad \exp\left[-\frac{▯}{2}\sum(x_i - \theta)^2\right]$

$$\propto \exp\left[-\frac{n\square}{2}(\theta-\bar{x})^2\right]$$

As $\quad \sum(x_i-\theta)^2 \quad \sum(x_i-\bar{x})^2+n(\bar{x}-\theta)^2]$

$$f(\theta/x_1,\ldots,x_n)\propto f_n(x_1,\ldots,x_n/\theta)g(\theta)$$

$$\propto \exp\left[-\frac{n\square}{2}(\theta-\bar{x})^2\right]\exp\left[-\frac{\tau'}{2}(\theta-\mu)^2\right.$$

$$\propto \exp\left[-\frac{n\square(\theta-\bar{x})^2+\tau'(\theta-\mu)^2}{2}\right] \qquad \text{But,}$$

$$n\square(\theta-\bar{x})^2+\tau'(\theta-\mu)^2 \quad (\tau'-n\square)+\tau'(\theta-\mu)^2+\frac{n\tau'\square(\bar{x}-\mu)^2}{\tau'+n\square}$$

$$f(\theta/x_1,\ldots,x_n)\propto \exp\left[-\frac{(\tau'-n\square)+\tau'(\theta-\mu)^2}{2}+\frac{n\tau'\square(\bar{x}-\mu)^2}{2(\tau'+n\square)}\right]$$

$$\propto \exp\left[-\frac{(\tau'+n\square)}{2}(\theta-\mu')^2\right] \qquad \#$$

6. Bayes Sequential Decision Problem

Consider a decision problem specified a parameter ϴ whose value are in ϴ (parameter space), a decision space D, and loss function L. we shall suppose that before the statistician chooses the decision in D, he will be permitted to observe sequentially the values of a sequence of r.v's $X_1, X_2, \ldots \ldots$ we shall suppose also that for any given value ϴ=θ, these observations are independent and identically distributed. It is then said that the observations are a *sequential random sample*. We shall suppose that the conditional p.d.f. of each observation X_i when ϴ=θ is $f(./\theta)$ and that the cost of observing the valuesX_i, in turn is C.

A sequential decision function or sequential decision procedure has two components. One component may be called as *sampling plan* or *stopping rule*. The statistician first specifies whether a decision should choose without any observations or whether at least one observation should be taken. If at least one observation is to be taken, the statistician specifies, for every possible set of observed values $X_1 \quad x_1, X_2 \quad x_2, X_n \quad x_n (n \geq 1)$

whether sampling should stop and a decision in D chosen without further observations or whether another value X_{n+1} should be observed.

The second component of sequential decision procedure may be called a *decision rule*. If no observations are to be taken, the statistician specifies a decision $d_0 \varepsilon D$that is to be chosen. If at least one observation is to be taken, the statistician specifies the decision $d_n(x_1, \ldots, x_n) \varepsilon D$that is to be chosen for each possible set of observed values $X_1 \quad x_1, X_2 \quad x_2, X_n \quad x_n$ after which the sampling might be terminated.

Let S denote the sample space of any particular observationX_1. For n=1, 2... We shall let $S^n \quad SxSx \ldots xS$ (with n factors) be the sample space of the n observations $X_1, X_2, \ldots, X_n$ and we shall let S^∞ be the sample space of the infinite sequence of observations $X_1, X_2, \ldots$

A sampling plan in which at least one observation is to be taken can be characterized by a sequence of subsets $B_n \varepsilon S^n$ (n=1, 2...) which have the following interpretations:

Sampling is terminated after the values $X_1 \quad x_1, \ldots, X_n \quad x_n$

have been observed if $(x_1, \ldots, x_n) \varepsilon B_n$. Another value x_{n+1} is observed if$(x_1, \ldots, x_n) \notin B_n$. If there is some value r for which $B_r \quad S^r$ or more generally if $P[(x_1, \ldots, x_n) \notin B_n]$ for n=1, 2... r] 0 then the sampling must stop after at most r observations have been taken. The specification of the sets B_n for any value of n such that $n > r$ then become irrelevant never the less, it is convenient to assume that the sets B_n will be defined for all values of n.

Each stopping sets B_n can be regarded not only as a subset of S^n but also as the subset of S^r for any value of r> n and as a subset ofS^∞. When B_n is regarded as a subset of$S^r, r > n$, B_n is a cylinder set. In other words if $(x_1, \ldots, x_n) \varepsilon B_n$ and if $(y_1, \ldots, y_r)$ is any other set in S^r such that, $y_i = x_i$, i=1,2........n then $(y_1, \ldots, y_r) \varepsilon B_n$ regarded as of the values of the final r-n components.

Suppose that at least one observation is to be taken with a given sampling plan, and let N denote the random total number of observations which will be taken before sampling is terminated. We shall [N=n] denote the set of points $(x_1, \dots, x_n) \varepsilon S^n$ for which [N=n]. in other words, suppose that the value $X_1 = x_1, \dots, X_n = x_n$ are observed in sequence,

then sampling will be terminated after the value x_n has been observed (and not before) if and only if $(x_1, \dots, x_n) \varepsilon [N \quad n]$. hence $[N=1] = B_1$ and for $n > 1$

$$[N \quad n] \quad (B_1 \cup B_2 \cup \dots \cup B_{n-1})^C \cap B_n$$

Similarly we shall let $[N \leq n] \quad \cup_{i=1}^{n} [N \quad i]$ denote the subset of $S^n$for which $N \leq n$ the events $[N \leq n]$ and [N=n] involve only the observations $X_1, X_2, \dots, X_n$. Hence these events are subset of S^n. Also they can be regarded as subsets of $S^r, r > n$. further more, events $[N > n] \quad [N \leq n]^C$ involve the observations $X_1, X_2, \dots, X_n$, and it can be regarded as subsets of S^r for any value of r, $r \geq n$.

For any prior p.d.f ξ of θ, we shall let $f_n(./\xi)$ denote the marginal p.d.f of the observations $X_1, X_2, \dots, X_n$

$$f(x_1, \dots, x_n/\xi) \quad \int_\theta f(x_1/\theta), \dots, f(x_n/\theta)\xi(\theta)dv(\theta) \text{ .... } (6.1)$$

Further more, we shall let $f_n(./\xi)$ denote the marginal joint d.f of $X_1, X_2, \dots, X_n$. Hence, for any event$A \subset S_n$,

$$P[x_1, \dots, x_n \varepsilon A] \quad \int_A df_n(x_1, \dots, x_n/\xi) \text{ .... } (6.2)$$

We can write the following equation:

$$P[N \leq n] \quad \int_A dF_n(x_1, \dots, x_n/\xi) =$$

$$\int_{[N=1]} dF_1(x_1/\xi) + \int_{[N=2]} dF_2(x_1, x_2/\xi) + \int_{[N=3]} dF_3(x_1, x_2, x_3/\xi) + \dots +$$
$$\int_{[N=n]} dF_n(x_1, x_2, \dots, x_n/\xi) \text{ .... } (6.3)$$

The decision rule of a sequential decision procedure is characterized by a decision rule $d_0 \varepsilon D$ and the sequence of functions $\delta_1, \delta_2, \ldots$ with the following property: for any point$(x_1, \ldots, x_n) \varepsilon S^n$, the function δ_n

satisfies a decision, $\delta_n(x_1, \ldots, x_n) \varepsilon D$. If the sampling plan specifies that an immediate decision in D is to be selected without any sampling then the decision $d_0 \varepsilon D$is chosen. If on the other hand, the sampling plan satisfies that at least one observation is to be taken and if the observed value $(x_1, \ldots, x_n)$ satisfies the condition $(x_1, \ldots, x_n) \varepsilon [N \quad n]$, then sampling is terminated and the decision , $\delta_n(x_1, \ldots, x_n) \varepsilon D$ is chosen. The value of the function , δ_n need only be specified on the subset [N=n] $\subset S^n$. A procedure involving a fixed number of observations n can always be obtained by adopting a sampling plan in which [N=j] =Φ, the empty set for j=1... n-1 and in which [N=n] $= S^n$. In general we can also consider sampling plans for which the probability is 1 that sampling will eventually be terminated. In other words, we shall assume that,

$$P[N < \infty] \quad \lim_{n \to \infty} P[N \leq n] \quad 1 \ldots\ldots\ldots\ldots\ldots (6.4)$$

[It need not be assumed that there is some finite upper bound n such that$P[N \leq n] \quad 1$]

Risk of a Sequential Decision Procedure

The total risk $\rho(\xi, d)$ of a sequential decision procedure which at least one observation is to be taken is,

$$\rho(\xi, \delta) \quad E\{L[\theta, \delta_N(X_1, \ldots, X_n)] + C_1 + C_2 + \cdots + C_N\}$$

$$\sum_{n=1}^{\infty} \int_{[N=n]} \int_{\theta} L[\theta, \delta_n(X_1, \ldots, X_n)]\, (\theta / x_1, \ldots, x_n) dv(\theta) dF_n(x_1, \ldots, x_n / \xi) + \sum_{n=1}^{\infty} (C_1 + C_2 + \cdots + C_N) P[N \quad n] \ldots\ldots\ldots\ldots\ldots (6.5)$$

Here $\xi(./ x_1, \ldots, x_n)$is posterior p.d.f of Θ after the values $X_1 \quad x_1, \ldots, X_n \quad x_n$ have been observed. Alternatively,

$$\rho(\xi, \delta) \quad \int_{\Omega} \left\{ \int_{[N=n]} L[\theta, \delta_n(X_1, \ldots, X_n)] \right\} [\prod_{i=1}^{n} f(x_i / \theta) d\mu(\mu)] \xi(\theta) dv(\theta) + \sum_{n=1}^{\infty} (C_1 + C_2 + \cdots + C_N) P[N \quad n] \ldots\ldots\ldots\ldots (6.6)$$

In the development of theory of sequential statistical decision problem we shall have little need to refer to any specified value $\xi(\theta/x_1,\dots,x_n)$ of the posterior p.d.f of Θ. However, we shall often have to refer to the entire posterior distribution as represented by its generalized p.d.f. therefore we shall denote the p.d.f simply by$\xi(x_1,\dots,x_n)$. *If* ξ *is prior distribution of* $\boldsymbol{\theta}$. Where$X_1 \quad x_1,\dots, X_n \quad x_n$ is$\xi(x_1,\dots,x_n)$.

For every p.d.f of θ. Let $\rho_0(\Phi)$ be defined as follows:

$$\rho_0(\Phi) \quad inf_{d\varepsilon D}\int_\Omega L[\theta,d]\Phi(\theta)dv(\theta) \text{ .... } (6.7)$$

In other words $\rho_0(\Phi)$ is the minimum risk from an immediate decision without any further observations when the p.d.f of θ is $\Phi(\theta)$.

A Bayes sequential decision procedure or an optimal sequential decision procedure is a procedure δ for which the risk ρ (ξ, δ) is minimized. Wherever a decision in D is chosen after sampling is terminated, that decision rule Bayes decision against the posterior distribution of Θ. For any such procedure δ which specifies that at least one observation is to be taken, we now have

$$\rho(\xi,\delta) \quad E[P_0[\xi(x_1,\dots,x_n).]+C_1+C_2+\cdots+C_N] \text{ .... } (6.8)$$

Further, more for the procedure δ_0 which specifies that can immediate decision in D should be chosen with out any observations we must have, $\rho(\xi,\delta_0) \quad \rho_0(\xi)$ .... (6.9)

<u>Exp 6.1</u>: $L(\theta_1,d_1) \quad L(\theta_2,d_2) \quad 0 \quad \theta \quad \{\theta_1,\theta_2\}, D \quad \{d_1,d_2\}$

$$L(\theta_1,d_2) \quad L(\theta_2,d_1) \quad b>0$$

Suppose X is discrete r.v.'s for which

$$f_i(x) \quad P[X \quad x/\theta \quad \theta_i]\; i \quad 1,2$$

$$f_1(1) \quad 1-\alpha, \qquad f_1(2) \quad 0, \qquad f_1(3) \quad \alpha \qquad 0<\alpha<1$$

$$f_2(1) \quad 0, \qquad f_2(2) \quad 1-\alpha, \qquad f_2(3) \quad \alpha$$

Suppose the cost per observation is C, let the prior distribution of θ is $P[\theta \quad \theta_1] \quad \xi \quad 1-P[\theta \quad \theta_2] \qquad \xi\le\frac{1}{2}$

<u>Solution</u>: $\xi(\theta/x) \quad \frac{f(x/\theta)P[\theta=\theta]}{P[X=x]}$

$\xi(\theta_1/1) \quad \frac{(1-\alpha)\xi}{(1-\alpha)\xi+0} \quad 1 \qquad \xi(\theta_1/1) \quad 0$

$$\xi(\theta_1/3) \quad \frac{f(3/\theta_1)P[\theta=\theta_1]}{f(3/\theta_1)P[\theta=\theta_1]+f(3/\theta_2)P[\theta=\theta_2]}$$

$$= \frac{\alpha\xi}{\alpha\xi+\alpha(1-\xi)} \quad \xi$$

Similarly, $\xi(\theta_2/1) \quad 0, \quad \xi(\theta_2/2) \quad 1, \quad \xi(\theta_2/3) \quad (1-\xi)$

Thus after an observation has been taken, either the value of θ becomes known or else the distribution of θ remains good as it was before the observation was taken.

$$\rho_0(\xi) \quad inf_d\{L(\theta_1,d_1)\xi + L(\theta_2,d_1)(1-\xi), L(\theta_1,d_2)\xi + L(\theta_2,d_2)(1-\xi)\}$$

$= inf_d\{b(1-\xi), b\xi\}$ Without any observation is taken.

$= b\xi \qquad since, \xi \leq \frac{1}{2}$

If the Bayes decision is chosen when $P[\theta \quad \theta_1] \quad \xi$, the expected loss is bξ.

If one observation is taken then the expected loss will be

$E\,\rho_0(\xi(X))$, where $\xi(X) \quad P[\theta \quad \theta/X \quad x]$

$\rho_0(1) \quad \rho_0(\xi(1))$

$inf_d\{L(\theta_1,\delta(1))P[\theta \quad \theta_1/X \quad 1] + L(\theta_2,\delta(1))P[\theta \quad \theta_2/X \quad 1]\}$

$inf_d\{0, \mathrm{b}\} \quad 0$

Now, $L(\theta_1,\delta(1))P[\theta \quad \theta_1/X \quad 1] + L(\theta_2,\delta(1))P[\theta \quad \theta_2/X \quad 1]$

$$\begin{array}{ll} 0 & if\,\delta(1) \quad d_1 \\ = b & if\,\delta(1) \quad d_1 \end{array}$$

Similarly, $\rho_0(2) \quad 0 \quad and \quad \rho_0(3) \quad b\xi$

$$E\rho_0(X) \quad 0P[X \quad 1] + 0P[X \quad 2] + \mathrm{b}\xi P[X \quad 3] \quad \mathrm{b}\xi\alpha$$

The expected loss $E\rho_0(X_1, \ldots, X_n) =$ b$\xi\alpha^n$ when the Bayes decision is chosen after n observations $X_1, \ldots, X_n$ have been taken,

ρ_n b$\xi\alpha^n$ + Cn Total risk for the optimal procedure when exactly n observations taken, assume ρ (1) <ρ (0)

$\frac{d}{dn}\rho(n) \quad 0 \Rightarrow n^* \quad [\log\frac{b\xi\log\left(\frac{1}{\alpha}\right)}{c}]\frac{1}{\log\left(\frac{1}{\alpha}\right)}$ ………………… (6.10)

and $\rho(n^*) \quad \frac{c}{\log\left(\frac{1}{\alpha}\right)}\left[1+\log\frac{b\xi\log\left(\frac{1}{\alpha}\right)}{c}\right]$ …………………… (6.11)

9. Wolfowitz Generalization of FCR bound and Sequential estimation and Testing:

A sequential provides a set of stopping rules $\{R_n(X_1, \ldots, X_n); n \quad 1,2 \ldots \ldots\}$ which are $\mathfrak{B}^{(n)}$ designate the Borel σ-field on $\mathfrak{x}^{(n)}$,

n-dimentional Euclidian space; assigning to $(X_1, \ldots, X_n)$ an integral value so that if $R_n(X_1, \ldots, X_n) \quad n$, we terminate sampling after the n^{th} observation otherwise, X_{n+1} is observed. Consider the σ-field $\mathfrak{B}_1 \subset \mathfrak{B}_2 \subset \cdots$ generated by $X_1, \ldots, (X_1, \ldots, X_n)$ a stopping rule R for a sequential procedure can be conveniently described by a sequence of sets $\{R_n; n = 1,2, \ldots.\}$ where, $R_n \varepsilon \mathfrak{B}_n$ for each n=1,2,….. Sampling is continued as by as consecutive vectors $(X_1, \ldots, X_n)$, n=1,2,….. do not enter one of the sets R_n. In another words, the sample size N (a random variable) is N= least integral n, n≥ 1 such that $(X_1, \ldots, X_n) \varepsilon R_n$

Define sets, $\overline{R_n} \quad \begin{array}{ll} R_1 & ifn \quad 1 \\ \overline{R_1} \cap \overline{R_2} \cap \ldots \cap R_n & ifn \geq 2 \end{array}$

The sets $\overline{R_n}$ is the set of all sample points which leads to stopping at N=n. The estimation rule for estimating a function g $(P_1, P_1, \ldots)$ is given by a srquence of functions $\widehat{g_1}, \widehat{g_2}, \ldots$ such that $\widehat{g_n} \varepsilon \mathfrak{B}_n$ for all n=1,2… and if N=n then the estimate of g is $\widehat{g_n}$.

<u>Lemma 9.1</u>: [*wald's equation*]: let$(X_1, \dots, X_n \dots)$ be a sequence of i.i.d random variables, distributed with some distribution, satisfying $E|X| < \infty$. For any sequential rule yielding EN$< \infty$

$$E\left(\sum_{i=1}^{N} X_i\right) \quad E(X)EN \text{ .... } (9.2)$$

<u>Proof</u>: let $(R_1, R_2, \dots)$ be the sequence of stopping regions. Then,

$$E\left(\sum_{i=1}^{N} X_i\right) \quad \sum_{n=1}^{\infty} \int_{R_n} \sum_{i=1}^{n} x_i \left(\prod_{i=1}^{n} dF(x_i)\right) \text{ .... } (9.2)$$

Now, $EX_i \quad \sum_{n=1}^{\infty} \int_{\overline{R_n}} (x_i) \prod_{i=1}^{n} dFx_i$

$$\sum_{n=1}^{i-1} \int_{\overline{R_n}} x_i \prod_{i=1}^{n} dF(x_i) + \sum_{n=i}^{\infty} \int_{\overline{R_n}} x_i \prod_{i=1}^{n} dF(x_i)$$

$$E\{X_i I[N < i]\} + E\{X_i I[N \geq i]\}$$

$$\sum_{n=i}^{\infty} \int_{\overline{R_n}} x_i \prod_{i=1}^{n} dF(x_i) \quad E\{X_i I[N \geq i]\} \quad P[N \geq i] E[X_i / N \geq i]$$

Since $[N \geq i]$ is $\mathfrak{B}_{i-1}$ measure and $\mathfrak{B}_0 \quad \mathfrak{B}$, therefore X_i is independent of $[N \geq i]$. thus

$$E[X_i / N \geq i] \quad E(X_i)$$

$$\sum_{n=i}^{\infty} \int_{\overline{R_n}} x_i \prod_{i=1}^{n} dF(x_i) \quad P[N \geq i] E(X_i)$$

$$P[N \geq i] E(X) \text{ .... } (9.3)$$

Now from (9.1)

$$\sum_{n=i}^{\infty} \int_{\overline{R_n}} \sum_{i=1}^{n} x_i \prod_{i=1}^{n} dF(x_i) \quad \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} \int_{\overline{R_n}} x_i \prod_{i=1}^{n} dF(x_i) \text{ .... } (9.4)$$

(This is permitted as $E|X| < \infty$)

$$\sum_{i=1}^{\infty} P[N \geq i] E(X) \quad \text{From } (9.3)$$

$$EX \sum_{i=1}^{\infty} P[N \geq i] \quad E(X)EN$$

$E\left(\sum_{i=1}^{n} X_i\right) \quad E(X)EN$ #

<u>Alternative Proof:</u> Define a r.v. Y_i such that

$Y_i \quad 1, \textit{if no decision is reached up to } (i-1)\textit{th stage, i.e. if } N > (i-1)$

$\quad 0 \;\; \textit{otherwise.}$

Clearly, Y_i depends only on $X_1 X_2, \ldots\ldots, X_{i-1}$ *and does not depend on* X_i. Also

$$S_N \quad \sum_{n=1}^{\infty} X_n Y_n$$

Hence $E(S_N) \quad E(\sum_{n=1}^{\infty} X_n Y_n)$ (9.5)

Now,

$$\sum_{n=1}^{\infty} E|X_n Y_n| \quad \sum_{n=1}^{\infty} E|X_n| E|Y_n| \; (\textit{because } X_n \textit{ and } Y_n \textit{ are independent})$$

$E|X_1| \sum_{n=1}^{\infty} E|Y_n| \quad E|X_1| \sum_{n=1}^{\infty} P[N \geq n]$ $(becasuse E|Y_n| \quad P[Y_n \quad 1]$
$P[N \geq n])$

$$E|X_1| \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P[N \quad k] \quad E|X_1| \sum_{n=1}^{\infty} nP[N \quad n]$$

$$E|X_1| |E(N)| < \infty$$

Therefore, $E(S_N)$ exists and we may change the order of operation of expectation and summation sign in (9.5). Hence,

$$E(S_N) \quad E\left(\sum_{n=1}^{\infty} X_n Y_n\right) \quad \sum_{n=1}^{\infty} E(X_n Y_n) \quad E(X_1) \sum_{n=1}^{\infty} E(Y_n)$$

$$E(X_1) \sum_{n=1}^{\infty} P[N \geq n] \quad E(X_1) E(N)$$

<u>Note:</u> Lemma 9.1 holds if only we assume $E(X_n) \quad \mu$ *and* $E(N) < \infty$ and the assumption that $X_i' s$ *are i.i.d. is not necessary.*

<u>Lemma 9.2</u>: Let$(X_1, \ldots, X_n)$ be a sequence of i.i.d random variables, having a common d.f. F(x) with mean zero and variance

$\sigma^2, 0 < \sigma^2 < \infty$ for any sequential stopping rule with E(N)< ∞ , if

$E\left\{\left(\sum_{i=1}^{N}|X_i|\right)^2\right\} < \infty$ *then,* $E\left\{\left(\sum_{i=1}^{N}X_i\right)^2\right\} \quad \sigma^2 EN$ .... (9.5)

Proof: As before,

$$E\left\{\left(\sum_{i=1}^{N}X_i\right)^2\right\} \quad \sum_{n=1}^{\infty}\int_{\overline{R_n}}\left(\sum_{i=1}^{n}x_i\right)^2\prod_{i=1}^{n}dF(x_i)$$

$$\sum_{n=1}^{\infty}\int_{\overline{R_n}}\left\{\sum_{i=1}^{n}x_i^{\,2} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_ix_j\right\}\prod_{i=1}^{n}dF(x_i)$$

$$\sum_{n=1}^{\infty}\int_{\overline{R_n}}\left(\sum_{i=1}^{n}x_i^{\,2}\right)\prod_{i=1}^{n}dF(x_i) + 2\sum_{n=1}^{\infty}\int_{\overline{R_n}}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_ix_j\right)\prod_{i=1}^{n}dF(x_i)$$

$$\sigma^2 EN + 2\sum_{n=1}^{\infty}\int_{\overline{R_n}}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_ix_j\right)\prod_{i=1}^{n}dF(x_i) \text{ By Lemma 9.1}$$

Now

$$\sum_{n=1}^{\infty}\int_{\overline{R_n}}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_ix_j\right)\prod_{i=1}^{n}dF(x_i) \quad \sum_{i=2}^{\infty}\sum_{j=1}^{i-1}\sum_{n=i}^{\infty}\int_{\overline{R_n}}x_ix_j\prod_{i=1}^{n}dF(x_i)$$

But $\sum_{n=i}^{\infty}\int_{\overline{R_n}}x_ix_j\prod_{i=1}^{n}dF(x_i) \quad P[N \geq i]E[X(x)/N \geq i]$ for j<i, (i=1, 2, 3...)
as X_i is independent $[N \geq i]$

$$P[N \geq i]EX_iE\left[X_j/N \geq i\right] \quad 0 \ \ for\ j < i, \ (i \quad 1, 2, 3 \ldots)$$

The rearrangement is guaranteed by condition $E\left\{\left(\sum_{i=1}^{N}|X_i|\right)^2\right\} < \infty$

Then $E\left\{\left(\sum_{i=1}^{N}X_i\right)^2\right\} \quad \sigma^2 EN$ #

Alternative Proof: Let Y_i be defined as in Alternative proof of Lemma 9.1. Then

$$E(S_N)^2 \quad E\{(\textstyle\sum_{i=1}^{\infty}X_iY_i)\}\{(\sum_{j=1}^{\infty}X_jY_j)\}$$

$$= E\left(\sum_{i=1}^{\infty}X_i^2Y_i^2 + \sum_{i\neq j}\sum_{j}X_iY_iX_jY_j\right) \qquad (9.6)$$

$$E|S_N^2| \quad E\left(\sum_{i=1}^{\infty}X_i^2Y_i^2 + \sum_{i\neq j}\sum_{j}|X_iX_j|Y_iY_j\right)$$

$$= E\left(\sum_{i=1}^{N}|X_i|\right)^2 < \infty \ (by\ assumption).$$

Hence the order of operation of summation and expectation in (9.6) can be interchanged. Now

$$E\left(\sum_{i=1}^{\infty} X_i^2 Y_i^2\right) \quad E(X_1^2)E\left(\sum_{i=1}^{\infty} Y_i^2\right) \quad \sigma^2 E\left(\sum_{i=1}^{\infty} Y_i\right) \quad \sigma^2 E(N)\ (by\ Lemma 9.1)$$

Again

$$E\left(\sum_{i\neq j}\sum_{j} X_i Y_i X_j Y_j\right) \quad 2E\left(\sum_{i>j}^{\infty}\sum_{j}^{i-1} X_i Y_i X_j Y_j\right) \quad 2\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} E(X_i X_j Y_i)$$

$$=$$

$$=2\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} E\left\{Y_i E\{X_i X_j/Y_i\}\right\} \quad 2\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} E\{Y_i E(X_i)E(X_j/Y_i)\} \quad 0$$

as X_j *and* Y_i *are independent of* X_i. #

Generalization of FCR bound for Sequential estimation

Theorem 9.1: [*wolfowitz*]: Let$(X_1, \dots, X_n, \dots)$ be a sequence of i.i.d random variables, whose common density $f(x;\ \theta)$ with respect to measure μ belong to a family $\psi \quad \{f(.;\ \theta): \theta\varepsilon\Theta\}$ on which the following regularity conditions are satisfied:

1. Θ contains an interval in a Euclidian k-space.
2. $f(x;\ \theta)$ is differentiable w.r.to θ on Θ.
3. $\int\left|\frac{\partial}{\partial\theta}f(x;\ \theta)\ \right|d\mu < \infty\ for\ all\ \theta\varepsilon\Theta.$
4. $0 < \int\left[\frac{\partial}{\partial\theta}log f(x;\ \theta)\right]^2 f(x;\ \theta)d\mu < \infty\ for\ all\ \theta\varepsilon\Theta.$
5. *For each* $n \quad 1,2,\dots\dots$ *and all* θ

$$\int\left[\sum_{i=1}^{n}\frac{\left|\frac{\partial}{\partial\theta}f(x_i;\theta)\right|}{f(x_i;\theta)}\right]^2\prod_{i=1}^{n}dF(x_i) < \infty$$

$$or\ \int\left[\sum_{i=1}^{n}\left|\frac{\partial}{\partial\theta}log f(x_i;\theta)\right|\right]^2\prod_{i=1}^{n}dF(x_i) < \infty$$

Let $(R_n, n \quad 1,2 \dots)$ be the sequence of stopping regions associated with a given sequential procedure. Let $g(\theta)$ be an estimable and differential function on Θ. Let $\hat{g}(X_1, \dots, X_n, \dots)$ be unbiased estimator of $g(\theta)$ satisfying the following conditions:

6. $\int |\hat{g}(x_1, \dots, x_n)| \frac{\partial}{\partial\theta} \prod_{v=1}^{n} f(x_v;\, \theta) \prod_{v=1}^{n} d\mu(x_v) < \infty$ *for each*
   $n \quad 1,2 \dots$
7. $\sum_{n=1}^{\infty} \frac{d}{d\theta} g_n(\theta)$ converges uniformly on Θ, where
   $$g_n(\theta) \quad \int_{R_n} \hat{g}(x_1, \dots, x_n) \prod_{v=1}^{n} dF(x_v)$$
   then $Var_\theta\{\hat{g}(X_1, \dots, X_n, \dots)\} \geq \frac{[g'(\theta)]^2}{I(\theta)E(N)}$ .... (9.6)
   for all θ, *provided* $EN < \infty$

Proof: Let N be the sample size associated with the given sequential procedure. Let $S(X_i;\, \theta) \quad \frac{d}{d\theta} \log f(X_i;\, \theta);\, i \quad 1,2 \dots$

These are i.i.d r.v's and 1-4 guarantee that E $S(X_i;\, \theta) \quad 0$ and $I(\theta)$ $E[S^2(X_i;\, \theta)] < \infty$ by condition 4 and the assumption $E(N) < \infty \Rightarrow$ *by Lemma* 9.1

$$E\left[\sum_{i=1}^{N} S(X_i;\, \theta)\right] \quad E(N)ES(X_i;\, \theta) \quad 0 \ \textit{for all}\ \theta \ \dots\dots\dots (9.7)$$

Furthermore according to condition 5

$$E\left[\sum_{i=1}^{N} |S(X_i;\, \theta)|\right]^2 < \infty \ \dots\dots\dots\dots\dots (9.8)$$

$$E\left[\left\{\sum_{i=1}^{N} S(X_i;\, \theta)\right\}^2\right] \quad E(N)ES^2(X, \theta) \quad E(N)I(\theta) \ \dots\dots\dots (9.8)$$

Consider the expectation,

$$E\left\{\hat{g}(X_1, \dots, X_n \dots) \sum_{i=1}^{N} S(X_i;\, \theta)\right\} \qquad \theta \varepsilon \Theta$$

Where $\hat{g}(X_1, \dots)$ is unbiased estimator of $g(\theta)$. According to (9.7) and by Schwartz inequality we have

$$E\left\{\hat{g}(X_1, \dots, X_N) \sum_{i=1}^{N} S(X_i;\, \theta)\right\} \leq$$

$$\left[E\left\{\left(\hat{g}(X_1, \dots, X_N) - g(\theta)\right)^2\right\} E\left\{\left(\sum_{i=1}^{N} S(X_i;\, \theta)\right)^2\right\}\right]^{\frac{1}{2}}$$

For all $\theta \,\varepsilon\, \Theta$ .... (9.10)

The quantity $E\left(\hat{g}(X_1,\dots,X_{N\dots}) - g(\theta)\right)^2$ is the variance of $\hat{g}(X_1,\dots,X_n \dots)$ under the sequential procedure. Further 6 & 7 allow the differentiation under the integral sign in,

$$g'(\theta) \quad \frac{d}{d\theta}\sum_{n=1}^{\infty}\int_{\bar{R}_n}\hat{g}(x_1,\dots,x_n)\prod_{v=1}^{n} dF(x_v)$$

$$\sum_{n=1}^{\infty}\int_{\bar{R}_n}\hat{g}(x_1,\dots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\,\theta)\prod_{i=1}^{n} d\mu(x_i)$$

$$\sum_{n=1}^{\infty}\int_{\bar{R}_n}\hat{g}(x_1,\dots,x_n)\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\,\theta)\right)\prod_{i=1}^{n} f(x_i;\,\theta)\,d\mu(x_i)$$

$$\sum_{n=1}^{\infty}\int_{\bar{R}_n}\hat{g}(x_1,\dots,x_n)\left(\sum_{i=1}^{n} S(x_i;\,\theta)\right)\prod_{i=1}^{n} dF(x_i)$$

$$E\left[\hat{g}(X_1,\dots,X_n \dots)\sum_{i=1}^{N} S(X_i;\,\theta)\right] \text{..........................} (9.11)$$

From (9.9) (9.10) & (9.11)

$$Var_\theta\hat{g}(X_1,\dots\dots) \geq \frac{E^2[\hat{g}(X_1,\dots,X_{n\dots})\sum_{i=1}^{Nn} S(X_i;\,\theta)]}{I(\theta)E(N)}$$

$$= \frac{[g'(\theta)]^2}{I(\theta)E(N)} \qquad \#$$

<u>**Optimality Criterion of Sequential Procedure**</u>

1. Subject to the condition $E_\theta(N) \leq m$ (m is a fixed integral bound) for all θ, minimize the variance of the best unbiased estimator that is, $E_\theta(\widehat{g_N} - g)^2$ uniformly in θ (if such an estimator exist.)
2. Subject to the condition $E(\widehat{g_N} - g)^2 \leq v < \infty (fixed\ finite\ positive\ value) for\ all\ \theta,$ minimize expected sample size $E_\theta(N)$.
3. Minimizes the expected cost of sampling plus expected loss, that is, $CE_\theta(N) + E_\theta(\widehat{g_N} - g)^2$

Generally there is no sequential estimator that can satisfy 3 uniformly in θ. In case 2, De*Groot*(1959) *and Wasan*(1964) have shown that a fixed sample size procedure in the binomial case does not minimize $E_\theta(N)$ w.r.to all sequential procedure uniformly in θ, $0< \theta < 1$ subject to the condition that $sup_{0<\theta<1} var_\theta(\hat{g}) \le \frac{1}{4m}$.

Sequential Estimation of the Mean of Normal Population

Let $(X_1, \dots, X_n)$ be i.i.d r.v's with mean μ and variance σ^2, both unknown as an estimate of μ, we choose $\bar{X}_n$, the sample mean. The problem now is to choose n. Let us assume that the loss incurred is $A|\bar{X}_n - \mu|$, where $A > 0$, is known constant and let each observation cost one unit. Then we wish to choose n to minimize,

$$EL(n) \quad E\{A|\bar{X}_n - \mu| + n\} \text{ .... } (9.12)$$

We have, $E\sqrt{n}\frac{|\bar{X}_n - \mu|}{\sigma} \quad \sqrt{\frac{2}{\pi}}$

So that $EL(n) \quad AE\left(\frac{\bar{X}_n - \mu}{\sigma}\sqrt{n}\right)\frac{\sigma}{\sqrt{n}} + n$

$$A\sqrt{\frac{2}{\pi}}\frac{\sigma}{\sqrt{n}} + n \text{ .... } (9.13)$$

Treating as continuous function n we have for minimax,

$$-A\sqrt{\frac{2}{\pi}}\frac{\sigma}{2(n)^{\frac{3}{2}}} + 1 \quad 0 \Rightarrow n_0 \quad \left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}} \text{ .... } (9.14)$$

At the value n that minimizes (9.13), for this value of n

$$v(\sigma) \quad EL(n_0) \quad A\sqrt{\frac{2}{\pi}}\sigma\left(\frac{\sqrt{2\pi}}{A\sigma}\right)^{\frac{1}{3}} + \left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}}$$

$$= A\sqrt{\frac{2}{\pi}}\sigma\left(\frac{\sqrt{2\pi}}{A\sigma}\right)^{\frac{-1}{3}} + \left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}}$$

$$\frac{A\sqrt{\frac{2}{\pi}}\sigma + \frac{A\sigma}{\sqrt{2\pi}}}{\left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}}} \quad 3\left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}} \quad 3n_0 \text{ .... } (9.15)$$

So that the loss due to the error of estimation is thrice the size of the sample, that is thrice the cost of sampling. Of course, this presupposes the knowledge of σ. If we do not know σ, we cannot compute n_0.

When σ is not known, we have the following sequential sampling procedure R:

$$N \quad least\ n\ , n \geq 2\ where \quad n \geq \left(\frac{As_n}{\sqrt{2\pi}}\right)^{\frac{2}{3}} \ldots\ldots\ldots\ldots\ (9.16)$$

Where, $s_n{}^2 \quad \frac{\Sigma(x_i-\overline{x_n})^2}{n-1} \quad , \overline{x_n} \quad \frac{1}{n}\sum_{i=1}^{n} x_i$

We may write this inequality,

$$N \quad first\ n\ , n \geq 2\ when\ \sum_{i=1}^{n}(x_i - \overline{x_n})^2 \leq \frac{2\pi}{A^2}(n-1)n^3 \ldots\ldots\ (9.17)$$

__Lemma 9.3__: Rule R terminates with probability 1.

__Proof__: It is sufficient to show that,

$$\left(\frac{As_n}{\sqrt{2\pi}}\right)^{\frac{2}{3}} \quad \overset{P}{\longrightarrow} n_0 \quad i.e\ \lim_{n\to\infty} P\left[\left| \left(\frac{As_n}{\sqrt{2\pi}}\right)^{\frac{2}{3}} - n_0\right| \leq \varepsilon\right] \quad 1$$

$$\text{Or} \qquad \lim_{n\to\infty} P\left[\left| \left(\frac{As_n}{\sqrt{2\pi}}\right)^{\frac{2}{3}} - n_0\right| > \varepsilon\right] \quad 0$$

$$\text{Now} \qquad \lim_{n\to\infty} P\left[\left| \left(\frac{As_n}{\sqrt{2\pi}}\right)^{\frac{2}{3}} - \left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}}\right| > \varepsilon\right]$$

$$\lim_{n\to\infty} P\left[\left| \left(\frac{s_n}{\sigma^2}\right)^{\frac{1}{3}} - 1\right| > \left(\frac{\sqrt{2\pi}}{A}\right)^{\frac{2}{3}} \varepsilon\right] \ldots\ldots\ldots\ldots\ (9.18)$$

$$\text{Since, } \lim_{n\to\infty} P\left[\left|\frac{s_n{}^2}{\sigma^2} - 1\right| > \left(\frac{\sqrt{2\pi}}{A}\right)^{\frac{2}{3}} \varepsilon\right] \leq \lim_{n\to\infty} \frac{2}{(n-1)} \varepsilon^2 \left(\frac{A\sigma}{\sqrt{2\pi}}\right)^{\frac{2}{3}} \quad 0$$

As $\frac{s_n{}^2}{\sigma^2} \sim \frac{\chi^2_{n-1}}{n-1}$ therefore (9.18) tends to zero as $n \to \infty$.

__Lemma 9.4__: For any fixed n, $\overline{X_n}$ is independent of $S_2^2, S_3^2, \ldots, S_n^2$ and hence,

$$P\left[\sqrt{n}\left(\frac{\overline{X_n}-\mu}{\sigma}\right) \le t/s_2^2, \dots, s_n^2\right] \quad \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \dots\dots\dots (9.19)$$

<u>**Proof:**</u> Define $U_i \quad \frac{X_i-\mu}{\sigma} \quad i \quad 1,2,\dots,n$

Then $U_i \sim N(0,1)$ r.v's and independent i=1,2.....

Let us write,

$$y_i \quad \frac{u_1+u_2+\cdots+u_i-iu_{i+1}}{\sqrt{i(i+1)}}, \text{i} \quad 1,2,\dots,\text{n}-1$$

$$y_n \quad \sqrt{\text{n}}\,\bar{\text{u}} \quad \text{where } \bar{\text{u}} \quad \frac{1}{\text{n}}\sum_{i=1}^{n} u_i$$

$$cov(Y_i, Y_j) \quad E\left[\frac{U_1+U_2+\cdots+U_i-iU_{i+1}}{\sqrt{i(i+1)}} \cdot \frac{U_1+U_2+\cdots+U_j-jU_{j+1}}{\sqrt{j(j+1)}}\right]$$

$$E\left[\frac{(U_1^2+U_2^2+\cdots+U_i^2)-iEU_{i+1}^2}{\sqrt{i(i+1)(j+1)}}\right] \quad \text{i}-\text{i} \quad 0$$

$$EY_i \quad 0, var(Y_i) \quad \frac{EU_1^2-i^2EU_{i+1}^2}{i(i+1)} \quad \frac{i+i^2}{i(i+1)} \quad 1$$

$$Y_i \text{ are } i.i.d\ N(0,1)\ i \quad 1,2,\dots,n$$

$$S_i^2 \quad \frac{1}{i-1}\sum_{j=1}^{n}(X_j-\bar{X})^2$$

$$\frac{\sigma^2}{i-1}\sum_{j=1}^{i-1} Y_j^{\,2} \quad \frac{\sigma^2}{i-1}\left(Y_1^{\,2}+\cdots+Y_{i-1}^{\,2}\right), i \quad 2,3,\dots,n$$

It follows that Y_n is independent of S_i^2 for i=2,...,n this is the same as saying $\overline{X_n}$ is independent of $S_2^2, S_3^2, \dots, S_n^2$.

Let us now compute the average loss for R.

$$L(N) \quad A\sqrt{N}\left|\frac{\overline{X_n}-\mu}{\sigma}\right| \frac{\sigma}{\sqrt{N}} + N$$

$$EL(N) \quad \sum_{n=2}^{\infty} P[N \quad n] E[L(N)/N \quad n]$$

$$= \sum_{n=2}^{\infty} P[N \quad n] E\left[A\sqrt{N}\left|\frac{\overline{X_n}-\mu}{\sigma}\right| \frac{\sigma}{\sqrt{N}} + N/N \quad n\right]$$

$$= \sum_{n=2}^{\infty} P[N \quad n] AE\left[\sqrt{N}\left|\frac{\bar{X}_n-\mu}{\sigma}\right|\frac{\sigma}{\sqrt{N}} + N/N \quad n\right] + E(N)$$

$$= \sum_{n=2}^{\infty} P[N \quad n](A\sqrt{\frac{2}{\pi}}\frac{\sigma}{\sqrt{N}}) + E(N)$$

$$= A\sqrt{\frac{2}{\pi}}\sigma E\left(N^{\frac{-1}{2}}\right) + E(N)$$

$$= 2n_0^{\frac{3}{2}} E\left(N^{\frac{-1}{2}}\right) + E(N)$$

Proposition: For large n_0 $P[N \leq n] \geq \frac{1}{2}$

Proof: We have, $P[N \leq n] \geq P\left[Y_1{}^2 + Y_2{}^2 + \cdots + Y_{n-1}{}^2 \leq \frac{(n-1)n^3}{n_0^3}\right]$ for $n \quad n_0$

$$P[N \leq n] \geq P\left[Y_1{}^2 + Y_2{}^2 + \cdots + Y_{n_0-1}{}^2 \leq n_0 - 1\right]$$

$$= P\left[\chi^2_{(n_0-1)} \leq n_0 - 1\right]$$

$$= P\left[\chi^2_{(n_0-1)} - \overline{n_0 - 1} \leq 0\right]$$

$$= P[Z \leq 0] \quad \frac{1}{2} \quad \text{Where, } Z \sim N(0,1) \qquad \#$$

Theorem 9.2: Let$(Z_1, \ldots, Z_n)$ be i.i.d r.v's such that $P[Z_j \quad 0] \neq 1$ set

$S_n \quad Z_1 + Z_2 \ldots + Z_n$ and for two constants C_1, C_2 *with* $C_1 < C_2$, define the random quantity N as the smallest n for which $S_n \leq C_1$ *or* $S_n \geq C_2$, *set* $N \quad \infty$ *if* $C_1 < S_n < C_2$ *for all* n. thus there exist $C > 0$ *and* $0 < \rho < 1$ such that,

$P[N > n] \leq C\rho^n$ *for all* n. .... (9.20)

Proof: The assumption $P[Z_j \quad 0] \neq 1$ implies that $P[Z_j > 0] > 0$. Let us suppose that $P[Z_j > 0] > 0$ then there exists $\varepsilon > 0$ such that $P\left[Z_j > \varepsilon\right] \quad \delta > 0$ in fact if $P\left[Z_j > \varepsilon\right] \quad 0$ *for* $\forall > \varepsilon$, then in particular $P\left[Z_j > \frac{1}{n}\right] \quad 0$ *for all* n.

but $P\left[Z_j > \frac{1}{n}\right] \uparrow P[Z_j > 0]$ and we have $0 \quad \lim_n P\left[Z_j > \frac{1}{n}\right] \quad P[Z > 0]$ which is a contradiction.

Thus for $P\left[Z_j > 0\right] > 0$ *we have* $P\left[Z_j > \varepsilon\right] \quad \delta > 0$ .... (9.21)

With C_1, C_2 and ε, there exist a positive integer m such that,

$$m\varepsilon > C_2 - C_1 \text{ .... (9.22)}$$

For such m we have,

$\bigcap_{j=k+1}^{k+m}[Z_j > \varepsilon] \subseteq [\sum_{j=k+1}^{k+m} Z_j > m\varepsilon] \subseteq [\sum_{j=k+1}^{k+m} Z_j > C_2 - C_1]$ .... (9.23)

$P[\sum_{j=k+1}^{k+m} Z_j > C_2 - C_1] \geq P\{\bigcap_{j=k+1}^{k+m}[Z_j > \varepsilon]\}$

$$\prod_{j=k+1}^{k+m} P\left[Z_j > \varepsilon\right] = \delta^m \text{, as } Z_j\text{'s are independent .}$$

Clearly,

$$S_{km} \quad \sum_{j=0}^{k-1}[Z_{jm+1} + \cdots + Z_{(j+1)m}]$$

Now we assert that, $C_1 < S_i < C_2, i \quad 1,2,\ldots,km \Rightarrow$

$$Z_{jm+1} + \cdots + Z_{(j+1)m} \leq C_2 - C_1, j \quad 1,2,\ldots,k-1 \text{ .... (9.24)}$$

This is because, if for some $j \quad 1,2,\ldots,k-1$ we suppose that $Z_{jm+1} + \cdots + Z_{(j+1)m} > C_2 - C_1$, this inequality together

$S_{jm} > C_1$ would imply $S_{(j+1)m} > C_2$, which is a contradiction to the first part of (9.24).

$$[N \geq km + 1] \subseteq \left[C_1 < S_j < C_2, j \quad 1,2,\ldots,km\right]$$

$$\subseteq \left[Z_{jm+1} + \cdots + Z_{(j+1)m} \leq C_2 - C_1\right]$$

$$P[N \geq km + 1] \leq \prod_{j=0}^{k-1}\left[Z_{jm+1} + \cdots + Z_{(j+1)m} \leq C_2 - C_1\right]$$

$$\leq (1 - \delta^m)^k$$

Thus, $P[N \geq km + 1] \leq (1 - \delta^m)^k \quad \frac{\left[(1-\delta^m)^{\frac{1}{m}}\right]^{mk+1}}{1-\delta^m} \quad C\rho^{mk+1}$

Put $C \quad \frac{1}{1-\delta^m} \quad ,\rho \quad (1 - \delta^m)^{\frac{1}{m}} \;, 0 < \rho < 1, C > 0$

thus, $\qquad P[N \geq n] \leq C\rho^n$ #

Theorem 9.3: Let $M_\theta(t) \quad M_\theta(e^{tz})$ be the m.g.f of Z, and let it be assumed to exist for all t, where $Z \quad log\frac{f(x,\theta_1)}{f(x,\theta_0)}$ then a necessary and sufficient condition that there exist a $(t \quad t_0 \neq 0)$ such that $M_\theta(t_0) \quad 1$ is that $E_\theta(Z) \neq 0$ and that Z takes on both positive and negative values with positive probability.

Proof: To prove the sufficiency, we observe that

$M_\theta''(t) \quad E_\theta(Z^2 e^{tZ}) > 0$ Unless Z=0 [since $M_\theta(t)$ exists for all t, it is differentiable any number of times]. Thus $M_\theta(t)$ is convex function of t. Now by assumption there exists a value Z'> 0 such that $P_\theta[Z > Z'] \quad u > 0$, therefore $t > 0$ implies

$$M_\theta(t) \quad E_\theta(e^{tZ}) > e^{tZ'} P_\theta[Z > Z'] \quad ue^{tZ'} \text{ .... } (9.25)$$

and *consequently* $M_\theta(t) \longrightarrow \infty$ *as* $t \to \infty$. *A* similar argument show that $M_\theta(t) \longrightarrow \infty$ *as* $t \to \infty$.

$$[M_\theta(t) > e^{tZ'} P_\theta[Z > Z'] \quad e^{tZ'} v$$

$$\textit{where } P_\theta[Z > Z'] \quad v > 0, Z' < 0]$$

The $M_\theta(t)$ assume a minimum value at the unique point t^* for which $M'_\theta(t^*) \quad 0$ *now* $M'_\theta(0) \quad E(Z) \neq 0$, so that $t^* \neq 0$ unless $E_\theta(Z) = 0$. Since $M_\theta(0) \quad 1$ *and* $M_\theta(t^*) < M_\theta(0) \quad 1$ wherever

$E_\theta(Z) \neq 0$ It must follow that there exist a $t_0 \neq 0$ such that $M_\theta(t_0) \quad 1$

To prove the condition is necessary, suppose that $P_\theta[Z \geq 0] \quad 1$ and let $P_\theta[Z = 0] = \alpha < 1$. Thus $P_\theta[Z > 0] = 1 - \alpha$, let $t < 0$ for any $0 < \varepsilon < 1 - \alpha$ we can find positive number C such that

$P_\theta[0 < Z < C] \leq \varepsilon$. Then,

$\alpha \leq M_\theta(t) \leq P_\theta[Z \quad 0] + \int_0^C e^{tZ} dF + \int_C^\infty e^{tZ} dF$

$$= \alpha + \epsilon + e^{tC}(1 - \alpha - \epsilon)$$

$as\ P[Z > C] \quad 1 - P[Z \leq C] \quad 1 - P[Z \quad 0] - P[0 < Z \leq C]$

$\therefore \quad \alpha \leq M_\theta(t) \leq [\alpha + \epsilon][1 - \alpha - \epsilon]e^{tC}$ .... (9.26)

And hence, $\alpha \leq \lim_{t\to\infty} M_\theta(t) \leq \alpha + \epsilon$

Since ε is arbitrary, $\lim_{t\to\infty} M_\theta(t) \quad \alpha$

We see that, $M'_\theta(t) \quad \lim_{t\to\infty} \frac{M_\theta(t+h) - M_\theta(t-h)}{h} > 0\ for\ all\ t < 0$

and hence $M_\theta(t) \quad 1$ has no solution other than t=0. A similar argument shows that, if $P_\theta[Z \leq 0] \quad 1$; $P_\theta[Z \quad 0] < 1$ then $M'_\theta(t) < 0$, for all t > 0, $M_\theta(t) \quad 1$ has no solution other than t=0. #

<u>Theorem 9.4</u>: [Fundamental Inequality]:

For a given θ and for all t such that $M_\theta(t) > \rho$, where ρ as in Theorem (9.2)

$E_\theta\left[e^{tS_N}\left(M_\theta(t)\right)^{-N}\right] \quad 1$ .... (9.27)

and if $P_\theta[Z > 0] > 0\ and\ P_\theta[Z < 0] > 0, where\ Z \quad log\frac{f(x,\theta_1)}{f(x,\theta_0)}$

$then$ (9.27) $holds\ for\ all\ t.$

<u>Proof</u>: Let the sequential procedure is defined in Theorem 9.2. Then since, $E_\theta e^{tS_n} \quad E_\theta e^{t(Z_1+\cdots+Z_n)}$

$= \prod_{i=1}^{n} E_\theta e^{tZ_i} \quad [M_\theta(t)]^n$ .... (9.28)

$E_\theta[e^{tS_n}[M_\theta(t)]^{-n}] \quad 1$

$1 \quad E_\theta[e^{tS_N}[M_\theta(t)]^{-N}]$

$\sum_{j=1}^{n} P_\theta[N \quad j] E[e^{tS_N}[M_\theta(t)]^{-n}/N \quad j] +$

$P_\theta[N > n] E_\theta[e^{tS_N}[M_\theta(t)]^{-n}/N > n]$

$\sum_{j=1}^{n} P_\theta[N \quad j]\, E\left[e^{tS_j}[M_\theta(t)]^{-j}/N \quad j\right] +$
$P_\theta[N > n] E_\theta[e^{tS_N}[M_\theta(t)]^{-n}/N > n]$ .... (9.29)

Since $E[e^{tS_N}[M_\theta(t)]^{-n}/N \quad j] \quad E\left[e^{tS_j}[M_\theta(t)]^{-j}/N \quad j\right]$ *as*

$\sum_{i=1}^{j} Z_i$ *is independent of* $\sum_{i=j+1}^{n} Z_i$

Since for N>n, $C_1 < S_n < C_2$ then by (9.29) and Theorem (9.2)

$0 \leq 1 - \sum_{j=1}^{n} P_\theta[N \quad j]\, E\left[e^{tS_j}[M_\theta(t)]^{-j}/N \quad j\right] \leq \frac{\rho^n}{[M_\theta(t)]^{-n}} E_\theta[e^{tS_n}/N > n]$
$\left(\frac{\rho}{M_\theta(t)}\right)^n k(t)$

Where k (t) is positive and for fixed θ depends only on t. Letting as $n \to \infty$ we see that for all real t such that $M_\theta(t) > \rho$ equation (9.27) holds.

Suppose now that Z takes on both positive and negative values so that $M_\theta(t)$ has a minimum value which is assumed at t=t* then it follows from (9.29) that for all t,

$P_\theta[N > n] < \frac{[M_\theta(t)]^n}{1<(t)}$ *and* $P_\theta[N > n] < \frac{[M_\theta(t^*)]^n}{1<(t^*)}$ .... (9.30)

And hence

$0 \leq 1 - \sum_{j=1}^{n} P_\theta[N \quad j]\, E\left[e^{tS_j}[M_\theta(t)]^{-j}/N \quad j\right] \leq \frac{[M_\theta(t^*)]^n}{1<(t^*)} \frac{k(t)}{k(t^*)}$

.... (9.31)

Thus $n \to \infty$ $0 \leq 1 - E_\theta[e^{tS_N}[M_\theta(t)]^{-N}] \leq 0$ *as* $\frac{M_\theta(t^*)}{M_\theta(t)} < 1$

Or $E_\theta[e^{tS_N}[M_\theta(t)]^{-N}] \quad 1$ #

<u>OC and ASN function of SPRT</u>

For brevity we denote by L (θ) the OC (*operating characteristic function)* of SPRT.

Let us consider the sequence Z_i of independent r.v's defined by Z_i $log\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)}$ i 1,2, ... satisfying the assumption of theorem (9.2) them if EZ≠0, there exist one and only $h_0 \neq 0$ such that $E(e^{h_0 z})$ $1; if\ E(Z) = 0$, this condition hold only for h_0 0 let us assume that $E(Z) \neq 0$. Since the distribution of Z depends on θ. Thus let us h_0 $h_0(\theta)$.

$$M_\theta(h_0) \quad M(h_0(\theta)) \quad Ee^{Zh_0(\theta)} \quad 1 \ \ldots\ldots\ldots\ldots\ (9.32)$$

$$= \int e^{Zh_0} f(Z,\theta)\, dZ \quad 1$$

Or $\sum e^{Zh_0}\, p(Z,\theta) \quad 1$ .... (9.33)

$$E_\theta e^{S_N h_0(\theta)} \quad \prod_{i=1}^{N} E\, e^{Z_i h_0(\theta)} \quad 1 \ \ldots\ldots\ldots\ldots\ (9.34)$$

$1 \quad E_\theta e^{S_N h_0(\theta)}$

$$L(\theta) E_\theta\left(e^{S_N h_0(\theta)} / S_N \leq \log B\right) + 1 - L(\theta) E_\theta\left(e^{S_N h_0(\theta)} / S_N \leq \log A\right) \ \ldots\ldots\ldots\ldots\ (9.35)$$

$$1 = L(\theta)\, E_\theta^{*} + [1 - L(\theta)] E_\theta^{**} \ \ldots\ldots\ldots\ldots\ (9.36)$$

Where $E_\theta^*, E_\theta^{**}$ represent the conditional expectations when we accept and reject the hypothesis respectively,

$$L(\theta) \quad \frac{E_\theta^{**} - 1}{E_\theta^{**} - E_\theta^{*}} \ \ldots\ldots\ldots\ldots\ (9.37)$$

We now find the approximate expression for $L(\theta)$. Let us consider, S_N $logB\ and\ S_N$ $logA$ instead of inequality $S_N \leq logB\ and\ S_N \geq logA$. Thus if S_N $logB$

$$E_\theta^*[expS_N h_0(\theta)] \approx E_\theta^*[exp(logB) h_0(\theta)]$$

$$\approx E_\theta^*[B]^{h_0(\theta)} \approx [B]^{h_0(\theta)}$$

Similarly, $E_\theta^{**}[expS_N h_0(\theta)] \approx E_\theta^{**}[exp(logA) h_0(\theta)] \approx [A]^{h_0(\theta)}$

$$\therefore \quad L(\theta) \quad \frac{[A]^{h_0(\theta)} - 1}{[A]^{h_0(\theta)} - [B]^{h_0(\theta)}}$$

When, $E_\theta(Z) \quad 0, then\ h_0(\theta') \quad 0$ where θ' is value of θ for which $E_\theta(Z) \quad 0 \quad 0$. Then,

$$\lim_{\theta\to\theta'} L(\theta) \quad L(\theta') \quad \lim_{\theta\to\theta'} \frac{[A]^{h_0(\theta)}-1}{[A]^{h_0(\theta)}-[B]^{h_0(\theta)}}$$

$$\lim_{\theta\to\theta'} \frac{\frac{A^{h_0(\theta)}-1}{\theta}}{\frac{A^{h_0(\theta)}-B^{h_0(\theta)}}{\theta}} \quad \frac{logA}{logA-logB}$$

For any real $h_0(\theta)$, we can determine the point in the plane with co-ordinate (θ, L (θ)). The locus of these points will be approximate graph of the **OC function.**

<u>Expected value of N i.e $E_\theta N$ or ASN</u> (<u>*Average Sampling Number*</u>):

We know that for

EZ≠0 $E_\theta[e^{S_N h}[M_\theta(h)]^{-N}] \quad 1$ differentiating w.r.to h at h=0

$$E_\theta\{S_N e^{S_N h}[M_\theta(h)]^{-N} - N e^{S_N h}[M_\theta(h)]^{\overline{-N-1}}(M''_\theta(h))\}_{h=0} \quad 0$$

$$E_\theta\{S_N - N E_\theta Z\} \quad 0 \quad E_\theta(N) \quad \frac{E_\theta(S_N)}{E_\theta(Z)}$$

$E^*_\theta[S_N]$ Denote the conditional expectation of the r.v's provided $S_N \le logB$ and $E^{**}_\theta[B_N]$ the conditional expectation of S_N provided $S_N \ge logA$.

$$E_\theta(S_N) \quad L(\theta)E^*_\theta(S_N) + \big(1 - L(\theta)\big)E^{**}_\theta(S_N)$$

$$E_\theta(N) \quad \frac{L(\theta)E^*_\theta S_N + (1-L(\theta))E^{**}_\theta(S_N)}{E_\theta(Z)}$$

If $S_N \quad logB\ or S_N \quad logA$ according as accepting and rejecting hypothesis.

$$E_\theta(N) \quad \frac{L(\theta)logB + (1-L(\theta))logA}{E_\theta(Z)}$$

If $E_\theta(Z) \quad 0$ we differentiate the fundamental Identity twice, we have,

$$E'_\theta\left[\left\{\left(S_N - N\frac{M'_\theta(h)}{M_\theta(h)}\right)^2 - \frac{NM''_\theta(h)M_\theta(h) - N\left(M'_\theta(h)\right)^2}{\left(M_\theta(h)\right)^2}\right\} e^{S_N h}[M_\theta(h)]^{-N}\right] \quad 0$$

Taking the derivative at h=0 and using

$M_\theta(0) \quad 1, M'_\theta(0) \quad E_\theta(Z) \quad 0$ And $M''_\theta(0) \quad E_{\theta_1}(Z^2) \neq 0$ we have

$$E_{\theta'}(S_N^2 - NE_{\theta'}Z^2) \quad 0$$

Or $E_{\theta'}(N) \quad \frac{E_{\theta'}S_N^2}{E_{\theta'}(Z^2)} \quad \frac{L(\theta')S_N^2 + (1-L(\theta'))E_\theta^{**}(S_N^2)}{E_{\theta'}(Z^2)}$

$$= \frac{L(\theta')(logB) + \left(1-L(\theta')\right)(logA)^2}{E_{\theta'}(Z^2)}$$

$$= \frac{logA}{logA-logB}(logB)^2 + \left(1-\frac{logA}{logA-logB}\right)(logA)^2/E_\theta(Z^2)$$

$$= -\frac{logAlogB}{E_{\theta'}(Z^2)}$$

<u>**Theorem 9.5**</u>**:** [$wald$] If SPRT is defined by $(logB, logA)$, where

$0 < B < 1, 0 < A < 1$, then the error probabilities α, β satisfy,

$A \leq \frac{1-\beta}{\alpha}, B \geq \frac{\beta}{1-\alpha}$ Where, $\alpha \quad P\theta_1[S_N \geq A], \beta \quad P\theta_0[S_N \geq B]$

If we set, $A' \quad \frac{1-\beta}{\alpha}, B' \quad \frac{\beta}{1-\alpha}$ then corresponding error probabilities α', β' satisfy, $\alpha' \leq \frac{\alpha}{1-\beta}, \beta' \geq \frac{\beta}{1-\alpha}$, and if $\alpha + \beta \leq 1, then$

$\alpha' + \beta' \leq \alpha + \beta$

<u>**Exp 9.1**</u>**:** Let$(X_1, \ldots, X_n)$ be i.i.d r.v's having N (θ, 1). The two simple hypotheses are, $H_0: \theta \quad -1, H_1: \theta \quad 1$

$$Z \quad log\frac{f(X,1)}{f(X,-1)} \quad loge^{\frac{(x-1)^2}{2}}e^{\frac{(x+1)^2}{2}} \quad loge^{2x} \quad 2X$$

m.g.f of X is, $G_\theta^{(t)} \quad \exp\left(\frac{t^2}{2} + \theta t\right)$

m.g.f of 2X is, $M_\theta^{(t)} \quad e^{2t^2+2\theta t}$

It follows that, $h_0(\theta) \quad -\theta$ thus,

$$L(\theta) \quad \frac{e^{-\theta a}}{e^{-\theta a} - e^{\theta b}} \quad where, -b \quad logB, a \quad logA$$

$$E_\theta(N) \quad \frac{1}{2\theta}\left[a\frac{1-e^{\theta b}}{e^{-\theta a} - e^{\theta b}} + b\frac{e^{-\theta a} - 1}{e^{-\theta a} - e^{\theta b}}\right]$$

For $H_0: \theta \quad \theta_0, H_1: \theta \quad \theta_1,$

$$\lambda_n \quad \prod_{i=1}^{n}\frac{f(X_i, \theta_1)}{f(X_i, \theta_0)} \; or \; log\lambda_n \quad \sum_{i=1}^{n}\frac{f(X_i, \theta_1)}{f(X_i, \theta_0)} \quad \sum Z_i$$

$$\sum_{i=1}^{n}\frac{f(X_i - \theta_1)^2}{2} + \sum_{i=1}^{n}\frac{f(X_i - \theta_0)^2}{2}$$

$$(\theta_1 - \theta_0)\sum X_i + \frac{(\theta_0^2 - \theta_1^2)n}{2} \quad \sum Z_i$$

We continue sampling as long as,

$$A < \sum Z_i < B \; or \; \frac{A}{(\theta_1 - \theta_0)} + \frac{(\theta_0^2 - \theta_1^2)n}{2(\theta_1 - \theta_0)} < \sum X_i < \frac{B}{(\theta_1 - \theta_0)} + \frac{n(\theta_0^2 - \theta_1^2)}{2(\theta_1 - \theta_0)}$$

$$Z_1 \quad (\theta_1 - \theta_0)X_1 + \frac{(\theta_0^2 - \theta_1^2)}{2}$$

$$E_{\theta_i}(Z_1) \quad (\theta_1 - \theta_0)\theta_i + \frac{(\theta_0^2 - \theta_1^2)}{2} \; , i \quad 0{,}1$$

If $\alpha \quad .01, \beta \quad .95$

$$A \approx loga' \; where, a' \quad \frac{1-\beta}{1-\alpha}$$

$$A \approx loga' \quad -1.29667$$

$$B \approx logb' \quad log\frac{\beta}{\alpha} \quad log\frac{.95}{.01} \quad log95 \quad 1.97772$$

$$E_0Z_1 \quad -\frac{1}{2} \quad -.5 \; , E_1Z_1 \quad .5$$

$$E_0N \approx \frac{(1-\alpha)A + \alpha B}{E_0Z_1} \quad \frac{.99(-1.29667) + .01(1.97772)}{-.5} \quad 2.53$$

$$E_1N \approx \frac{(1-\beta)A + \beta B}{E_1Z_1} \quad 3.63$$

<u>Non informative Priors</u>

Because of the compelling reasons to perform a conditional analysis and the alternatives of using Bayesian machinery to do so there have been attempts to use the Bayesian approach even when no (or minimal) prior information is available. What is needed in such situation is a <u>Non informative prior</u>, by which is meant a prior which contains no information about θ (or more crudely which 'faros' no possible values of θ over others.) for example, in testing between two simple hypothesis, the prior which gives probability ½ to each of the hypothesis is clearly non-informative.

<u>Exp</u>: suppose the parameter of interest is normal mean θ, so that the parameter space$\theta = \{-\infty, \infty\}$. If non-informative prior density is desired, it seems reasonable to give equal weights to all possible values of θ. unfortunately, if $\pi(\theta)\quad c > 0$ is chosen, the π has infinite mean $i.e \int \pi(\theta)d\theta \quad \infty$ and is not proper density. Nevertheless, such π can be successfully worked with the choice of c is unimportant, so that typically the non-informative prior clearly for this problem is chosen to be π(θ)=1 this is often called the informative density on R and was intersected and used by Laplace(1812).

As in the above example, it will frequently happen that natural non-informative prior is an <u>improper prior</u>, namely which has infinite mass.

<u>Exp</u>: instead of considering θ, suppose the problem has been parameterized in terms of$\eta = e^{\theta}$, this is one-to-one information and should have no bearing on the ultimate answer.

But if π (θ) is the density of θ, then the correspondently for η is,

$\pi^*(\eta)\quad \eta^{-1}\pi(log\eta)$ Hence if the non-informative prior of θ is chosen to be constant, we should choose the non-informative prior of η to be conditional to η^{-1} to maintain consistency. Thus we maintain consistency and choose both the non-informative prior

<u>Non Informative Priors for location and scale parameters:</u>

<u>Exp:</u> suppose that x and Θ are subsets ofR^k, and that the density of <u>X</u> is of the form $f(\underline{x} - \underline{\theta})$ $i.e$ depend on$(\underline{x} - \underline{\theta})$. The density then said to be a <u>location</u>

density, and θ is called a location parameter. (Some times a location vector when$k \geq 2$). The $N(\theta,\sigma^2),\sigma^2$ fixed, is an example of location density.

To derive a non-informative prior for this situation, imagine that, insisted of observing X, we observe the random variable Y=X+C. $C\epsilon R^k$. Define $\underline{\eta} = \underline{\theta} + \underline{C}$ it is clear that Y has density$f(\underline{y} - \eta)$. If now

$x \quad \theta \quad R^k$ Thus the sample space and parameter space for (Y, η) problem are alsoR^k. The (X, Θ) &(Y, η) problems are identical and sensitive and it seems reasonable to in sets that they have the same non-informative prior.

Letting π and π* denote the non-informative priors in the (X, Θ) and (Y, η) problems respectively, the above arguments implies that π and π* should be equal i.e

$$p^{\pi}[\theta\epsilon A] \quad p^{\pi^*}[\eta\epsilon A]$$

For any set A inR^k. Since η=θ+C, it should be true that

$$p^{\pi^*}[\eta\epsilon A] \quad p^{\pi}[\theta + C\epsilon A] \quad p^{\pi}[\theta\epsilon A - C]$$

$$A - C \quad \{Z - C: Z\epsilon A\} \quad then,$$

$$p^{\pi}[\theta\epsilon A] \quad p^{\pi}[\theta\epsilon A - C] \quad for\ all\ \theta\epsilon R^k \quad \text{.......................} \quad (1)$$

Any π satisfying relation (1) is said to be location invariant prior.

Assuming that the prior has a density then,

$$\int_A \pi(\theta)d\theta \quad \int_{A-C}\pi(\theta)d\theta \quad \int_A \pi(\theta - C)d\theta \qquad for\ all\ A\epsilon R^k$$

$$\pi(\theta) \quad \pi(\theta - C) \quad for\ all\ \theta\epsilon\Theta, or\ \ \pi(C) \quad \pi(0) \quad for\ all\ C\epsilon R^k$$

This conclusion is that π must be constant function. It convenient to choose the constant to be 1, so the non-informative prior density for a location parameter is $\pi(\theta) \quad 1$

A one dimensional scale density is a density of the form, $\alpha^{-1}f(\frac{x}{\alpha})$ where$\alpha > 0$. The parameter $\alpha > 0$ is called a scale parameter. The

$N(0,\sigma^2)G(\alpha,\beta)$, α known as scale density.

To derive a non-informative prior for this situation, imagine that, instead of observing X, we observe the random variable Y=CX C> 0.

Define $\eta \quad C\alpha$, can easy calculation shows that the density of Y is

$\eta^{-1}f(\frac{y}{\eta})$. If x=R or $(0,\infty)$ then the sample and parameter space for the (X, α) problems are the same as there for the (Y, η) problem. The two problems are thus identical in structure, which again indicates that they should have the same non-informative prior. Letting π and π* denote the priors in the (X, α) and (Y, η) problem, respectively, this means that the equality,

$$p^{\pi}[\alpha\epsilon A] \quad p^{\pi^*}[\eta\epsilon A]$$

Should for all $A \subset (0,\infty)$. Since η= $C\alpha$, it should also be true that

$$p^{\pi^*}[\eta\epsilon A] \quad p^{\pi}[\alpha\epsilon C^{-1}A],$$

$C^{-1}A \quad \{C^{-1}Z : Z\epsilon A\}$. Putting these together, it follows that π should satisfy,

$$p^{\pi}[\alpha\epsilon A] \quad p^{\pi}[\alpha\epsilon C^{-1}A] \qquad \text{for all } C > 0$$

And any distribution π for which this is true is called scale invariant.

$$\int_A \pi(\alpha)d\alpha \quad \int_{C^{-1}A}\pi(\alpha)d\alpha \quad \int_A \pi(C^{-1}\alpha)C^{-1}d\alpha \qquad \text{for all } A \subset (0,\infty) \Rightarrow$$
$$\pi(\alpha) \quad C^{-1}\pi(C^{-1}\alpha) \qquad \text{for all } \alpha.\ \text{let } \alpha \quad C$$

$\pi(C) \quad C^{-1}\pi(1)$. Setting for convenience, and nothing that above equality must hold for all $C > 0$, it follows that a reasonable non-informative for a scale parameter is $\pi\alpha \quad \alpha^{-1}$.

Non-informative prior in general setting:

For more general problem, various (some what ad hoe) suggestive have been advance for determining a non-informative prior. The most widely used method is that of Jeffrey's method which is as follows:

If $\underline{\theta} \quad (\theta_1, \dots, \theta_k)'$ is a vector, Jeffrey's suggest the use of

$\pi(\underline{\theta}) \quad [\det I(\underline{\theta})]^{\frac{1}{2}}$ *'det' determinant;*

Where $I(\underline{\theta}) \quad [I_{ij}(\underline{\theta})] \Rightarrow I_{ij}(\underline{\theta}) \quad -E_{\underline{\theta}}[\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log f(x/\underline{\theta})]$

__Exp:__ A location-scale density is a density of the form $\sigma^{-1} f(\frac{x-\theta}{\sigma})$ where $\theta \epsilon R, \sigma > 0$ are the unknown parameters. $N(\theta, \sigma^2)$ is crucial example of location-scale density Working with$N(\theta, \sigma^2)$, $\underline{\theta}$= (θ, σ). Fisher informative matrix is,

$$I(\underline{\theta}) \quad -E_{\underline{\theta}} \begin{pmatrix} \frac{\partial^2}{\partial\theta^2} \frac{(x-\theta)^2}{2\sigma^2} & \frac{\partial^2}{\partial\theta\,\partial\sigma} \frac{(x-\theta)^2}{2\sigma^2} \\ \frac{\partial^2}{\partial\theta\,\partial\sigma} \frac{(x-\theta)^2}{2\sigma^2} & \frac{\partial^2}{\partial\theta^2} \frac{(-(x-\theta)^2)}{2\sigma^2} \end{pmatrix}$$

$$-E_{\underline{\theta}} \begin{pmatrix} \frac{-1}{\sigma^2} & \frac{2(\theta-x)}{\sigma^3} \\ \frac{2(\theta-x)}{\sigma^3} & \frac{-3(x-\theta)^2}{\sigma^4} \end{pmatrix} \quad \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{3}{\sigma^2} \end{pmatrix}$$

$$\pi(\underline{\theta}) \quad [\frac{1}{\sigma^2}, \frac{3}{\sigma^2}]^{\frac{1}{2}} \alpha \frac{1}{\sigma^2}$$

This is actually the non-informative prior ultimately recommended by Jeffrey's non-informative prior is that it is not affected by restriction on the parameter space. Thus if it is known that θ> 0 , the Jeffrey's non-informative prior is still π (θ) =1.

__Exp:__ let $(X_1, \ldots, X_n)$ be a random sample from N (θ_1, θ_2) let the non-informative prior of (θ_1, θ_2) be $(\theta_1, \theta_2) \propto \frac{1}{\theta^2}$ and θ_1& θ_2 assumed to be independent. Find the posterior $.d.f$ of $f(\theta_1/\underline{x})$ & $f(\theta_2/\underline{x})$.

__Solution:__ $f(x_1, \ldots, x_n/\theta_1, \theta_2) \propto \frac{1}{(\theta_2)^{\frac{n}{2}}} \exp - \frac{\Sigma(x_i-\theta_1)^2}{2\theta^2}$

$$f(\theta_1, \theta_2/x_1, \ldots, x_n) \propto \frac{1}{(\theta_2)^{\frac{n}{2}}} \exp - \frac{\Sigma(x_i-\theta_1)^2}{2\theta^2} \frac{1}{\theta^2}$$

$$= \frac{1}{(\theta_2)^{\frac{n}{2}}\theta_2} \exp - \frac{\Sigma(\bar{x}-\theta_1)^2}{2\theta^2} \exp - \frac{n(\bar{x}-\theta_1)^2}{2\theta^2}$$

$$= \frac{1}{(\theta_2)^{\frac{n+2}{2}}} \exp - \frac{S^2 n-1}{2\theta^2} \exp - \frac{n(\bar{x}-\theta_1)^2}{2\theta^2}$$

$$f(\theta_1/\underline{x}) \propto \int_0^\infty \frac{1}{(\theta_2)^{\frac{n+2}{2}}} \exp - \frac{\Sigma(x_i-\theta_1)^2}{2\theta^2} d\theta_2 \quad \text{Put} \frac{1}{2\theta^2} \quad t \Rightarrow -\frac{d\theta_2}{\theta_2^2} \quad 2dt$$

$$\propto \int_0^\infty t^{\frac{n+2}{2}} \exp - \sum(x_i - \theta_1)^2 t \, \frac{1}{t} \, dt$$

$$= \int_0^\infty t^{\frac{n}{2}-1} \exp - t \sum(x_i - \theta_1)^2 \, dt$$

$$\propto \frac{1}{[\Sigma(x_i-\theta_1)^2]^{\frac{n}{2}}} \quad \frac{1}{[\Sigma(x_i-\bar{x})^2 + n(\bar{x}-\theta_1)^2]^{\frac{n}{2}}}$$

$$\propto \frac{1}{[1+\frac{n(\bar{x}-\theta_1)^2}{\Sigma(x_i-\bar{x})^2}]^{\frac{n}{2}}} \quad \frac{1}{\left[1+\frac{T^2}{n-1}\right]^{\frac{n-1}{2}}}$$

Where, $T \sim t -$ *distribution with* $(n-1)$ *degree of freedom.*

$$f(\theta_2/\underline{x}) \propto \frac{1}{(\theta_2)^{\frac{n+2}{2}}} \exp - \frac{\overline{n-1}\, s^2}{2\theta^2} \int_{-\infty}^{\infty} \exp - \frac{n(\bar{x}-\theta_1)^2}{2\theta^2} d\theta_1$$

$$\propto \frac{(\theta_2)^{\frac{1}{2}}}{(\theta_2)^{\frac{n+2}{2}}} \exp - \frac{\overline{n-1}\, s^2}{2\theta^2}$$

$$= \frac{1}{(\theta_2)^{\frac{n+1}{2}}} \exp - \frac{\overline{n-1}\, s^2}{2\theta^2}$$

Let $w \quad \frac{\overline{n-1}\, s^2}{\theta^2} \quad dw \quad \frac{-\overline{n-1}\, s^2}{\theta_2^2} d\theta_2$

$$f(w/\underline{x}) \propto \frac{(\theta_2)^{\frac{1}{2}}}{(\theta_2)^{\frac{n+1}{2}}} \exp - \frac{w}{2} \quad \frac{1}{(\theta_2)^{\frac{n-3}{2}}} \exp - \frac{w}{2}$$

$$= \frac{1}{(\theta_2)^{\frac{n-1}{2}-1}} \exp - \frac{w}{2} \propto \chi^2_{n-1}$$

<u>Highest Posterior Density Regions</u>: (HPD Regions)

<u>Def</u>: A $100(1-\alpha)\%$ credible set for θ is subset of Θ such that,

$$1-\alpha \le P[C/x] \quad \int_C dF^{\pi/(\theta/x)}(\theta)$$

$$\int_C \pi/(\theta/x)\, d\theta \quad \textit{for continuous case}$$

$$\sum_{\theta \epsilon C} \pi/(\theta/x) \quad \textit{for discrete case}$$

Since the posterior distribution is an actual prob. distribution on Θ, one can speak of the probability that θ is C. this is in contrast to classical confidence procedures, which can only be interpreted in term of coverage probability that is the probability that the random variable X will be such the confidence set C(X) contains θ.

In choosing a credible set for θ, it is usually describe to try to minimize its size. To do this one should include in the set only those points with the largest posterior density i.e the most likely values of θ.

Def: The $100(1-\alpha)\%$ HPD credible set (HPD region) for θ is the subset C of Θ of the form

$$C \quad \{\theta \epsilon \Theta : \pi(\theta/x) \geq K(\alpha)\}$$

Where $K(\alpha)$ Is the largest constant such that,

$$P[C/x] \geq 1-\alpha.$$

Exp: let $(X_1, \dots, X_n)$ be a random sample from N $(\theta, 1)$. Let the prior p.d.f of θ be N (μ, τ^2). Find the HDD regions for θ.

Solution: $f(\theta/x_1, \dots, x_n) \quad \frac{f(x_1,\dots x_n/\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(x_1,\dots x_n/\theta)\pi(\theta)d\theta}$

$$\frac{\exp\frac{\Sigma(x_i-\bar{x})^2}{2} \frac{n(\bar{x}-\theta)^2}{2} \exp\frac{(\theta-\mu)^2}{2\tau^2}}{\exp\frac{\Sigma(x_i-\bar{x})^2}{2}\int_{-\infty}^{\infty}\exp\frac{n(\bar{x}-\theta)^2}{2}\exp\frac{(\theta-\mu)^2}{2\tau^2}d\theta} = \frac{\exp\frac{n(\bar{x}-\theta)^2}{2}\exp\frac{(\theta-\mu)^2}{2\tau^2}}{\int_{-\infty}^{\infty}\exp-\left[\frac{n(\bar{x}-\theta)^2}{2}\exp\frac{(\theta-\mu)^2}{2\tau^2}\right]d\theta}$$

$$\int_{-\infty}^{\infty}\exp-\left[\frac{n(\bar{x}^2+\theta^2-2\bar{x}\theta)}{2}+\frac{(\theta^2+\mu^2+2\theta\mu)}{2\tau^2}\right]d\theta$$

$$\exp\left(-\left(\frac{n\bar{x}^2}{2}+\frac{\mu^2}{2\tau^2}\right)\int_{-\infty}^{\infty}\exp-\frac{1}{2}\left[\theta^2-2\theta\left(\bar{x}+\frac{\mu}{\tau^2}\right)+\frac{\mu^2}{\tau^2}\right]d\theta\right)$$

$$\exp\left(-\left(\frac{n\bar{x}^2}{2}+\frac{\mu^2}{2\tau^2}\right)\int_{-\infty}^{\infty}e^{-\frac{1}{2}}\left[\theta^2-2\theta\left(\bar{x}+\frac{\mu}{\tau^2}\right)+\frac{\mu^2}{\tau^2}\right]d\theta\right)$$

$$\exp\left(-\left(\frac{n\bar{x}^2}{2}+\frac{\mu^2}{2\tau^2}\right)\int_{-\infty}^{\infty} e^{-\frac{1}{2}}\left[\theta^2-2\theta\left(\bar{x}+\frac{\mu}{\tau^2}\right)+\left(\bar{x}^2+\frac{\mu}{\tau^2}\right)-\left(\bar{x}^2+\frac{\mu}{\tau^2}\right)+\frac{\mu^2}{\tau^2}\right]d\theta\right)$$

$$\exp\left(-\frac{n\bar{x}^2\tau^2+\mu^2}{2\tau^2}-\frac{1}{2}\left(\bar{x}+\frac{\mu}{\tau^2}\right)^2-\frac{\mu^2}{2\tau^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}}\left[\theta-\left(\bar{x}+\frac{\mu}{\tau^2}\right)\right]^2 d\theta\right)$$

Let $\mu \quad 0$

$$\int_{-\infty}^{\infty}\exp-\left[\frac{n\bar{x}^2+\theta^2-2\bar{x}\theta}{2}+\frac{\theta^2}{2\tau^2}\right]d\theta$$

$$=\exp\left(\frac{-n\bar{x}^2}{2}-\frac{\bar{x}^2}{2}\right)\int_{-\infty}^{\infty}\exp-\frac{1}{2}[\theta-\bar{x}]^2 d\theta$$

$$=\sqrt{2\pi}\exp-\frac{1}{2}(-n\bar{x}^2+\bar{x}^2)$$

$\therefore \quad \pi(\theta/x) \quad \frac{1}{\sqrt{2\pi}}\exp-\frac{(n\bar{x}-\theta)^2}{2}-\frac{\theta^2}{2\tau^2}+\frac{1}{2}(-n\bar{x}^2+\bar{x}^2)$

$$=\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2}\left[n\bar{x}^2+n\theta^2-2n\bar{x}\theta+\frac{\theta^2}{\tau^2}-n\bar{x}^2-\bar{x}^2\right]$$

$$=\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2\tau^2}[n\theta^2\tau^2-2n\bar{x}\theta\tau^2+\theta^2-\bar{x}^2\tau^2]$$

$$=\frac{1}{\sqrt{2\pi}}\exp-\frac{1}{2\tau^2}[\theta^2(1+n\tau^2)-2n\bar{x}\theta\tau^2-\bar{x}^2\tau^2]$$

$\pi(\theta/x) \quad N(\mu(\bar{x}), P^{-1})$

$\mu(\bar{x}) \quad \frac{\tau^2\bar{x}}{\tau^2+\frac{\sigma^2}{n}}, \; P \quad \frac{n\tau^2+\sigma^2}{\tau^2\sigma^2} \quad , \frac{1}{P} \quad \frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}$

Suggested Reading

1. Berger, J.O. (1993) Statistical Decision Theory and Bayesian Analysis, Springer Verlag.
2. Bernando, J.M. and Smith, A.F.M. (1994). Bayesian Theory, John Wiley and Sons.
3. Robert, C.P. (1994). The Bayesian Choice: A Decision Theoretic Motivation, Springer.
4. Berger, J.O. (1985). Statistical decision theory-Fundamental concepts and methods, Springer Verlag.
5. Degroot, M. H. (1971). Optimal statistical decisions, McGraw-Hill.
6. Ferguson, T.S. (1967). Mathematical statistics- A decision theoretic approach, Academic press.