CHAPTER-I

BASIC CONCEPTS

1.0 Introduction :

To study the admissibility of a decision rule it is necessary to study the statistical decision theory. In the following we introduce the statistical decision theory along with the basic concepts. The statistical decision theory deals with the choice of a decision to be taken on the basis of some relevant information. Here decisions are not only based on the possible inferences that are listed, but also depend on

(i) the assigned loss resulting from wrong decisions.

(ii) the prior information about the true state.

Thus the statistical decision theory consists of three basic elements :

(I) In each decision problem there is a certain unknown quantity θ called the state of nature and this affects the decision procedure. The set of all apossible states of nature denoted by Θ is called the parameter space.

(II) In literature decisions are called actions, and a particular action will be denoted by 'a' while the set of all possible actions denoted by $\mathcal{A}$.

(III) A Key element of decision theory is the loss function. If nature chooses a point θ in Θ and the statistician chooses an action 'a' in $\mathcal{A}$, as the consequences of these two choices the statistician loses an amount $L(\Theta,a)$. This loss $L(\Theta,a)$ represents the loss to the statistician if he takes action a when Θ is the true state of nature, i.e. L is a non-negative real valued function defined on $\Theta \times \mathbb{A}$.

To obtain the information about Θ the statistical investigation is performed. The experiments are designed such that the observations X are distributed according to some probability distribution P_Θ, the state of nature Θ is called the parameter which is unknown. Then the outcome that is a random variable will be denoted by X and the particular value of X will be denoted by x. The set of all possible outcomes is the sample space, and denoted by $\mathfrak{X}$.

Thus a statistical decision problem is a triplet(Θ, $\mathbb{A}$, L) coupled with an experiment involving a random variable X whose distribution P_Θ depends on the state $\Theta \in \Theta$ chosen by nature. On the basis of the outcome of the experiment $X = x$, the statistician chooses an action $d(x) \in \mathbb{A}$. Such a function d which maps the sample space $\mathfrak{X}$ into $\mathbb{A}$. Corresponding to the decision d the loss is now the random quantity $L(\Theta, d(x))$. The expected value of $L(\Theta, d(x))$ when Θ is the true state of nature is called the risk function,

$$R(\Theta,d) = E_\Theta L(\Theta, d(x))$$

this represents the average loss to the statistician when the true state of nature is Θ and the statistician uses the decision d.

Definition (1.0.1)

Any function $d(.)$ that maps the sample space $\mathcal{X}$ into $\mathbb{A}$ is called a non-randomized decision rule or a non-randomized decision function; provided the risk function $R(\Theta,d)$ exists and is finite for all $\Theta \in \Theta$.

The class of all non-randomized decision rules is denoted by D.

Definition (1.0.2)

A randomized decision rule $\sigma^*(x,.)$ is for each x, a probability distribution on $\mathbb{A}$, with the interpretation that if x is observed, $\sigma^*(x,A)$ is the probability that an action in A (a subset of $\mathbb{A}$) will be chosen. The class of all randomized decision rules is denoted by D*.

1.1 Some optimal decision rules :

The aim of statistical decision theory is to determine the decision function σ that minimizes the risk function,

$$R(\Theta,\sigma) = E_\Theta [L(\Theta,\sigma(x)]$$

For each fixed state of nature, there is a decision rule for which the risk is small, so that the statistician take this decision. But this decision rule is differ for various values of Θ. So that no one action can be taken as a 'best decision rule' as compared to all other possible decision rules.

For example - Consider the problem of estimating the parameter Θ when the loss is squared error $L(\Theta,a)=(\Theta-a)^2$.

If Θ_0 is the true state of nature then the best action to be taken by the statistician is $a = \Theta_0$ for which the risk function is zero, and the best decision rule is the non-randomized decision rule $d_0(x) \equiv \Theta_0$. If Θ_1 is the true state of nature then the best action to take the statistician is $a = \Theta_1$, and the non-randomized decision rule $d_1(x) \equiv \Theta_1$. Thus for different values of Θ there may be different decision rules for which the risk is minimum. So that there does not exist best decision rule (best in the sense that for all Θ this decision has smallest risk as compared to any other decision rule).

Thus we have seen that a best rule usually does not exist. But for to get a better decision rule, we have to propose the two general methods so that a decision rules are satisfactory.

(1) Restriction to some classes of decision rules :

As described above uniformly best decision rule generally does not exist. Thus to choose a rule which is better (in some sense) than the other available decision rules, we need to put some restrictions on the available decision rules, so that the choice of best decision rule is meaningful.

By putting the appropriate restrictions the class of decision rules will be a smaller one and from this smaller class a best decision rule can be chosen. Commonly used restrictions (can also be viewed as desired properties) are (i) unbiasedness (II) invariance. In the following we describe these properties.

(I) Unbiasedness :

An estimate $\sigma(x)$ of $g(\theta)$ is said to be unbiased if, when θ is the true value of the parameter, the mean of the distribution of $\sigma(x)$ is $g(\theta)$.

$$E_{\theta}(\sigma(x)) = g(\theta) \qquad \text{for all } \theta.$$

Thus an unbiased estimate in a very weak sense treats all states of nature equally.

Hence we apply the principle of unbiasedness and restrict the available rules to be unbiased, it is then possible that a 'uniformly best unbiased estimate' of θ will exist.

(II) Invariance :

The invariance principle basically states that if two problems have identical formal structures (i.e. have the same sample space, parameter space, densities and loss functions) then the same decision rule should be used in each problem. This is called a principle of invariance. In this principle by considering the transformations, the given problem is transformed, and this transformed problem has the identical structure to the original problem. The decision rules in the original and transformed problems be the same, this leads to a restriction to so called as 'invariant' decision rules. This class of rules will be small so that a 'best invariant' decision rule will exist. To describe the above concepts we need to define some additional terms and we define these in the following.

Groups of Transformations :

Let $\mathfrak{X}$ denote an arbitrary space (in the present context $\mathfrak{X}$ is the sample space) and consider transformations of $\mathfrak{X}$ into itself. We will be concerned only with transformations that are one-to-one and onto.

A transformation g is said to be one-to-one if $g(x_1) = g(x_2) \Rightarrow x_1 = x_2$, and it is onto if the range of g is all of $\mathfrak{X}$.

If g_1 and g_2 are two transformations, the composition of g_2 and g_1, which is the transformation to be denoted $g_2 g_1$, which is defined by

$$g_2\, g_1(x) = g_2\,(g_1(x))$$

For $g \in G$, the inverse of g denoted by g^{-1} is defined as

$$g^{-1}(g(x)) = x$$

Definition (1.1.1) :

A group of transformations of $\mathfrak{X}$ to be denoted G, is a set of one-to-one and onto transformations of $\mathfrak{X}$ into itself, which satisfies the following conditions :

i) If $g_1 \in G$ and $g_2 \in G$ then $g_2 g_1 \in G$.

ii) If $g \in G$ then $g^{-1} \in G$

iii) The identity transformation e defined by $e(x) = x$, is in G.

Example (1.1.1)

Let $\mathfrak{X} = R$. Consider the group of transformations $G = \{ g_c : c > 0 \}$, where $g_c(x) = cx$.

It is easy to verify that G is a group of transformations. This is called the group of scale transformations.

Invariant Decision Problems :

Let X denote a random variable having density $f(x|\theta)$ with sample space $\mathfrak{X}$. Also $\mathbb{F}$ denote the class of all densities $f(x|\theta)$ for $\theta \in \Theta$. If G is a group of transformations of $\mathfrak{X}$ (which we call a group of transformations of X). We want to consider the problems based on observation of the random variables g(X), g is a specific member of G.

Definition (1.1.2) :

The family of densities $\mathbb{F}$ is said to be invariant under the group G if for every $g \in G$ and $\theta \in \Theta$, there exist a unique $\theta^* \in \Theta$ such that $Y = g(X)$ has density $f(y|\theta^*)$, and let θ^* be denoted by $\bar{g}(\theta)$.

Definition (1.1.3) :

A loss function $L(\theta,a)$ is said to be invariant under G, if for every $g \in G$ and $a \in \mathbb{A}$, there exist an $a^* \in \mathbb{A}$ such that $L(\theta,a) = L(\bar{g}(\theta), a^*)$ for all $\theta \in \Theta$ and let the action a^* denoted by $\tilde{g}(a)$.

Definition (1.1.4) :

The decision problem is said to be invariant under G, if (i) $\mathbb{F}$ is invariant under G, and

(ii) the loss function $L(\theta,a)$ is invariant under G.

Example (1.1.2)

Suppose $X \sim N(\theta,1)$, Ⓗ $= (-\infty,\infty)$, $\mathfrak{X} = (-\infty,\infty)$

consider the group of transformations on $\mathfrak{X}$, defined by,

$g_c(x) = x + c$ for all $x \in R$, all $c \in R$

$\therefore \quad P_\theta[X \leq x] = \int_{-\infty}^{x} f(t,\theta)\,dt \quad$ where $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta)^2}$

and $\quad \mathbb{F} = \left[f(t,\theta) : \theta \in Ⓗ \right]$

Now $P_\theta[g(X) \leq x] = P_\theta[X + c \leq x]$

$$= P_\theta[X \leq x - c]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x-c} e^{-\frac{1}{2}(t-\theta)^2}\,dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}[y-(\theta+c)]^2}\,dy, \quad y = x+c$$

$\therefore \quad Y = g_c(X) \sim N(\theta+c, 1)$

hence $f(y/\theta^*) \in \mathbb{F}$, for all $g_c \in G$, $\theta \in$ Ⓗ

hence $\mathbb{F}$ is invariant under G and $\bar{g}_c(\theta) = \theta+c$

Let $\mathbb{A} =$ Ⓗ that is the problem of interest is to estimate the parameter θ, and further $\tilde{g}_c(a) = c + a$ and $L(\theta,a)$ be invariant under G.

Thus the problem of estimating θ is invariant under G.

Remark:

In the above example if Ⓗ or/ and $\mathbb{A}$ is proper subset of R, then $\mathbb{F}$ will not be invariant under G.

Invariant Decision Rules

We have seen in an invariant decision problem, the formal structures of the problems involving X and $Y=g(X)$ are identical. Hence the invariance principle states that σ and σ^*, the decision rules used in the X and Y problems respectively should be identical.

Definition (1.1.5) :

If a decision problem is invariant under a group G of transformations, a (non-randomized) decision rule $\sigma(x)$ is invariant under G if for all $x \in \mathcal{X}$ and $g \in G$.

$$\sigma(g(x)) = \bar{g}(\sigma(x))$$

Definition (1.1.6) :

Two points Θ_1 and Θ_2 in Θ are said to be equivalent if $\Theta_2 = \bar{g}(\Theta_1)$ for some $\bar{g} \in \bar{G}$. An orbit in Θ is an equivalence class of such points. Thus the Θ_0-orbit in Θ, to be denoted $\Theta(\Theta_0)$, is the set

$$\Theta(\Theta_0) = [\,\bar{g}(\Theta_0) : \bar{g} \in \bar{G}\,]$$

Theorem (1.1.1) :

The risk function of an invariant decision rule σ is constant on orbits of Θ, or, equivalently,

$$R(\Theta, \sigma) = R(\bar{g}(\Theta), \sigma)$$

for all $\Theta \in \Theta$ and $\bar{g} \in \bar{G}$.

Proof :

By definition,

$R(\Theta,\sigma) = E_\Theta L(\Theta, \sigma(X)) = E_\Theta L[\bar{g}(\Theta), \tilde{g}(\sigma(X))]$; invariance of loss

$= E_\Theta L[\bar{g}(\Theta), \sigma(g(X))]$; (invariance of σ)

$= E_{\bar{g}(\Theta)} L[\bar{g}(\Theta), \sigma(X)]$ (invariance of distributions)

$= R(\bar{g}(\Theta), \sigma(X)]$ □

Definition (1.1.7) :

A group $\bar{G}$ of transformations of Ⓗ is said to be transitive if Ⓗ consists of a single orbit, or equivalently if, for any Θ_1 and Θ_2 in Ⓗ, there exists some $\bar{g} \in \bar{G}$ for which $\Theta_2 = \bar{g}(\Theta_1)$.

If $\bar{G}$ is transitive, then from theorem 1.1.1 it is clear that any invariant decision rule has a constant risk. An invariant decision rule which minimizes this constant risk will be called a best invariant decision rule.

Location Parameter Problems :

Consider the problem of estimating a parameter $\Theta \in$ Ⓗ, in which $\mathfrak{X}$ = Ⓗ = $\mathcal{A}$ = R, and Θ is a location parameter of the distribution of the observable random variable X. We assume that the loss function is a function of $(a-\Theta)$ alone, i.e. $L(\Theta, a) = h(a-\Theta)$.

This problem is clearly invariant under the group G of transformations $g_c : g_c(x) = x + c$ and $\bar{g}_c(\Theta) = \Theta + c$ and $\tilde{g}(a) = a+c$.

An invariant non-randomized estimate σ is such that

$$\sigma(g(x)) = \bar{g}(\sigma(x))$$

becomes,

$$\sigma(x + c) = \sigma(x) + c$$

$$\therefore \quad \frac{\sigma(x + c) - \sigma(x)}{c} = 1$$

$$\lim_{c \to 0} \quad \frac{\sigma(x + c) - \sigma(x)}{c} = 1$$

$$\therefore \quad \sigma'(x) = 1$$

$$\therefore \quad \sigma(x) = x + c' \qquad (1)$$

(where $c' = \sigma(0)$). Any invariant rule must be in the form of (1). The risk function of any invariant decision rule $\sigma \in D^*$ satisfies

$$R(\theta,\sigma) = R(\theta+c, \sigma) \quad \text{for all } \theta \in \text{Ⓗ} \quad \text{and} \quad \text{all } c.$$

Thus the risk is independent of θ.

The risk of the non-randomized rule is

$$R(\theta, \sigma) = E_{\theta} L(X + c' - \theta)$$

$$= E_{o} L(X + c') \qquad (2)$$

The best invariant decision rule is simply that rule of the form (1) for which c' minimizes (2).

<u>Example</u> (1.1.3)

Let $X_1, \ldots, X_n$ be a random sample of size n from the exponential distribution, whose density function given by

$$f(x|\theta) = e^{-(x-\theta)} I_{(\theta,\infty)}(x), \quad \text{Ⓗ} = R$$

Take $L(\theta,a) = (a-\theta)^2$.

Solution :

Here θ is a location parameter.

Now,

$$E_o L(X + c', 0) = \int_0^\infty (x+c')^2 e^{-x} dx$$

$$= c'^2 + 2c' + 2 \quad \text{(on simplification)}$$

Therefore, $E_o L(X + c', 0)$ exists for every c'.

To find minimum of $E_o L(X + C', 0)$, we find

$$\frac{d}{dC'} E_o L(X + C', 0)$$

$$\therefore \frac{d}{dC'} E_o L(X + C', 0) = 2 + 2C' = 0$$

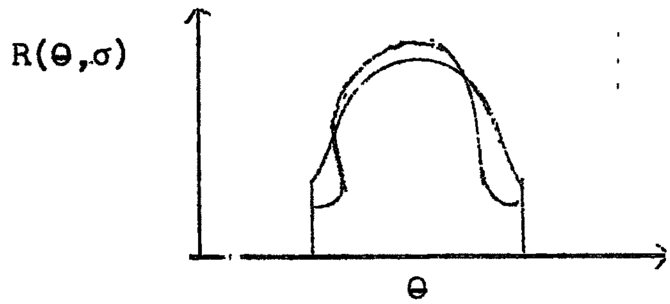
it implies that $C' = -1$

hence best invariant decision rule is $\sigma(x) = x-1$ and its risk is unity, which is constant

2. Ordering the Decision Rules :

Instead of restricting the class of procedures, one can approach the problem some what differently, consider the risk functions corresponding to two different decision rules σ_1 and σ_2. If $R(\theta,\sigma_1) < R(\theta, \sigma_2)$ for all θ, then σ_1 is clearly preferable to σ_2 since its use will lead to a smaller risk no matter what the true value of θ is.

However, the situation is not clear when the two risk functions intersect as in figure.



The statistician may introduce a principle by which he chooses a decision rule. Such a principle will lead to an ordering of the available decision rules, and any such ordering may be considered a principle.

There are two important and useful principles in the study of decision theory.

(A) : <u>The Bayes Principle</u> :

As usual the aim of statistical decision theory is to find estimator that minimize the risk $R(\Theta,\sigma)$ at every value of Θ and this is possible by restricting the available rules, by the use of unbiasedness or invariance principle. Now, we shall drop such restrictions and admitting all estimators into competition and we shall look for estimators that make the risk function $R(\Theta,\sigma)$ is small in over all sense.

Now, the problem of minimizing

$$\int R(\Theta, \sigma) \, d_{\Lambda} (\Theta) \qquad (3)$$

where we assume that the weights represented by Λ add up to one, that is,

$$\int d_{\Lambda} (\Theta) = 1$$

so that Λ is a probability distribution.

An estimater σ minimizing (3) is called a Bayes estimator with respect to Λ. The value of (3) is known as the (minimum) Bayes risk. Equivalently we have,

<u>Definition</u> (1.1.8) :

A decision rule σ_o is said to be Bayes with respect to the prior distribution, Λ defined on Ⓗ if,

$$r(\Lambda, \sigma_o) = \inf_{\sigma \in D^*} r(\Lambda, \sigma)$$

we use the notion Ⓗ to mean -

i) parameter space and

ii) The variable which takes the values in a parameter space, however the meaning of Ⓗ will be clear as per the context.

Thus we have seen that Λ is a probability distribution of Ⓗ and therefore, in Bayes principle involves the notion of a distribution on the parameter space Ⓗ , called a <u>prior distribution.</u>

A choice of prior distribution Λ is typically made like that of the distributions P_Θ by combining experience with convenience.

<u>Definition</u> (conjugate families) (1.1.9) :

Let $\mathbb{F}$ denote the class of density functions $f(x|\Theta)$, $\Theta \in$ Ⓗ . A class $\mathbb{P}$ of prior distributions is said to be a conjugate family for $\mathbb{F}$ if h (posterior distribution) of Ⓗ given X is in the class $\mathbb{P}$ for all $f \in \mathbb{F}$ and $h \in \mathbb{P}$.

We give below some conjugate prior distributions.

Distribution	Parameter	Conjugate prior distribution
Binomial	Probability of success	Beta
Poisson	Mean	Gamma
Exponential	Reciprocal of Mean	Gamma
Normal	Mean (variance known)	Normal
Normal	Variance (mean known)	Inverse Gamma

Definition (1.1.10) :

Let $\epsilon > 0$. A rule σ_ϵ is said to be ϵ-Bayes with respect to the prior distribution Λ if,

$$r(\Lambda, \sigma_\epsilon) \leq \inf_{\sigma \in D^*} r(\Lambda, \sigma) + \epsilon.$$

The following example shows that there exist ϵ-Bayes but not Bayes.

Example (1.1.4) :

Let $\Theta = A = R$, $L(\theta, a) = (\theta - a)^2$

Let the distribution of X given θ be normal with mean θ and variance unity, and the prior distribution of θ is normal with mean 0 and variance unity.

It can be easily shown that the posterior density of θ given X = x is,

$$g(\theta|x) = \frac{1}{\sqrt{\pi}} e^{-(\theta - \frac{x}{2})^2}$$

which is normal with mean $\frac{x}{2}$ and variance 1/2.

The Bayes rule is $d(x) = \frac{x}{2}$

which has Bayes risk, r(d) (say)

$r(d) = E\ [(\theta - d(x))^2/x] = \frac{1}{2}$

Let d(x) an estimator of the form ax.

If $D_1 = \{aX / a \in R\}$ then it is easy to verify that, $d(x) = \frac{1}{2} X$ is the Bayes estimator of θ.

If $D_2 = \{aX,\ a > 1/2\}$, is the class of decision rules then note that there does not exist a Bayes decision rule.

However,

The Bayes risk is,

$$r(d) = E[E[(aX - \theta)^2 / X]]$$

$$= a^2 + (a-1)^2$$

hence every decision rule bX ($\frac{1}{2} < b \le \frac{1}{2} + \sqrt{\frac{\epsilon}{2}}$) is not a Bayes rule and is a ϵ-Bayes rule.

To find the Bayes estimator one can use the following result :

<u>Theorem</u> (1.1.2) :

Let Θ have distribution Λ and given $\Theta = \theta$, let X have distribution P_θ. Suppose in addition, the following assumptions hold for the problem of estimating $g(\Theta)$ with non-negative loss function $L(\theta, d)$.

a) There exists an estimator σ_o with finite risk.

b) For almost all X, there exists a value $\sigma_\Lambda(x)$. minimizing

$$E[\,L(\Theta, \sigma(x)) / X=x]$$

then $\sigma_\Lambda(X)$ is a Bayes estimator.

For a proof see Lehmann(1982) pp 239.

As an application of the above theorem we have the following:

For various loss functions the Bayes decision is an estimator $\sigma_{\Lambda}(X)$ which minimizes the posterior risk. In the following we present some loss functions and the corresponding Bayes estimators.

I) If $L(\Theta, d) = [d - g(\Theta)]^2$ then the Bayes rule is,

$$\sigma_{\Lambda}(X) = E\,[g(\Theta)\,/\,X]$$

which is the mean of the posterior distribution of Θ given X.

II) If $L(\Theta,d) = w(\Theta)\,[d - g(\Theta)]^2$

then the Bayes rule is

$$\sigma_{\Lambda}(X) = \frac{\int w(\Theta)\; g(\Theta)\; d_{\Lambda}(\Theta/X)}{\int w(\Theta)\; d_{\Lambda}(\Theta/X)} = \frac{E\,[\,w(\Theta)\;\; g(\Theta)\,/\,X]}{E\,[\,w\,(\Theta)\,/\,X\,]}$$

III) If $L(\Theta,d) = |d - g(\Theta)|$ then the Bayes rule is any median of the conditional distribution of Θ given X.

IV)
$$L(\Theta,d) = \begin{cases} 0 & \text{when } |d-\Theta| \leq C \\ 1 & \text{when } |d-\Theta| > C \end{cases}$$

then the Bayes rule is the midpoint of the Interval I of length 2C which maximizes $P[\Theta \in I\,/X]$

We have seen that for various loss functions and prior distributions, we can find the Bayes estimators provided it exist. But untill we have not seen when and where the Bayes estimators are unique. The following lemma gives sufficient conditions for the Bayes estimator to be unique when the loss function is strictly convex.

Lemma (1.1.1)

If the loss function $L(\Theta,d)$ is squared error or more generally if it is strictly convex in d, a Bayes solution σ_Λ is unique (a.e. $\mathbb{P}$), where $\mathbb{P}$ is the class of distributions P_Θ, provided

i) its average risk with respect to Λ is finite and

ii) if Q is the marginal distribution of X given by

$$Q(A) = \int P_\Theta(X \in A)\ d\Lambda(\Theta)$$

then a.e. Q implies a.e. $\mathbb{P}$.

For a proof see Lehmann (1982) pp 240.

Definition (Formal Bayes rule) (1.1.11) :

A Bayes rule can be found by choosing for each x, an anction which minimizes the posterior expected loss or equivalently which minimizes

$$\int_{\Theta} L(\Theta,a)\ f(x/\Theta)\ d\Lambda(\Theta) \qquad (4)$$

If the Bayes risk is infinite, we define a Bayes rule as given by (4), such a rule is called as a formal Bayes rule.

Definition (Generalized Bayes rule) (1.1.12) :

If Λ is an improper prior in a decision problem with Loss L, a generalized Bayes rule, for given x, is an action which minizes

$$\int_{\Theta} L(\theta,a)\, f(x/\theta)\, d_{\wedge}(\theta)$$

that is, which minimizes the posterior expected loss.

We presented above the Bayes estimators, for specific prior distributions and loss functions. We proceed now to discuss minimax estimators. The objective is to derive an estimator which minimizes the maximum possible risk.

(B) Minimax Estimation :

A rule σ_1 is preferred to a rule σ_2 if,

$$\sup_{\theta \in \Theta} R(\theta,\sigma_1) < \sup_{\theta \in \Theta} R(\theta,\sigma_2)$$

A rule σ_o that is most preferred in this ordering (σ_o is preferred to any other rule $\sigma \in D^*$) is called a minimax decision rule.

Definition (1.1.13) :

A decision rule σ_o is said to be minimax if,

$$\sup_{\theta \in \Theta} R(\theta,\sigma_o) = \inf_{\sigma \in D^*} \sup_{\theta \in \Theta} R(\theta,\sigma)$$

Definition (1.1.14) :

Let $\epsilon > 0$. A decision rule σ_o is said to be ϵ- minimax if,

$$\sup_{\theta \in \Theta} R(\theta,\sigma_o) \leq \inf_{\sigma \in D^*} \sup_{\theta \in \Theta} R(\theta,\sigma) + \epsilon$$

Definition (1.115) :

A prior distribution Λ is said to be least favourable if,

$$r_\Lambda \geq r_{\Lambda'} \quad \text{for all prior distributions } \Lambda'.$$

where

$$r_\Lambda = \int R(\theta, \sigma_\Lambda)\, \sigma_\Lambda(\theta)$$

We have seen that for small risk a search for such estimator is only restricted to Bayes estimator and suitable limits of such Bayes estimators. But for what prior distribution Λ is the Bayes solution σ_Λ likely to be minimax ?

The following theorem provides a simple condition for a Bayes estimator σ_Λ to be minimax.

Theorem (1.1.3) :

Suppose that Λ is a distribution of Θ such that

$$\int R(\theta, \sigma_\Lambda)\, d\Lambda(\theta) = \sup_\theta R(\theta, \sigma_\Lambda)$$

Then

i) σ_Λ is minimax

ii) If σ_Λ is the unique Bayes solution with respect to Λ, it is the unique minimax procedure.

iii) Λ is least favourable.

For a proof see Lehmann (1982) pp 250.

(1.2) Admissibility :

The earlier two principles, Bayes principle and minimax principle have there own limitations. As the Bayes rule depends very much on the choice of prior distribution, it is not desirable to adopt Bayes principle when the choice of prior distribution is not strongly justified.

Let σ_1 and σ_2 be such that

$$R(\Theta,\sigma_1) < R(\Theta,\sigma_2) \leq R(\Theta_0,\sigma_2) \leq R(\Theta_0,\sigma_1)$$

then according to minimax criteria one has to prefer σ_2 to σ_1. But however, if according to the prior distribution Θ_0 is very unlikely then the choice of σ_2 is not that desirable. A very satisfactory criteria to choose (if possible) one among the two possible decisions is by comparing there risk functions for all possible values of the parameter. i.e. prefer σ_1 to σ_2 if

$$R(\Theta,\sigma_1) \leq R(\Theta,\sigma_2) \qquad \text{for all } \Theta \in \text{Ⓗ}$$

If $R(\Theta,\sigma_1) = R(\Theta,\sigma_2)$ for all $\Theta \in$ Ⓗ then the performance of the two rules is the same, the choice is only the matter of convenience . If $R(\Theta,\sigma_1) < R(\Theta,\sigma_2)$ for some $\Theta \in$ Ⓗ then σ_1 is preferred to σ_2 and write it as '$\sigma_1 < \sigma_2$'. The above ordering is of course a partial ordering and however the class of decision rules will be significantly reduced by using the above partial ordering.

Definition (1.2.1) :

i) A decision rule σ_1 is said to be as good as a rule σ_2 if,

$R(\theta,\sigma_1) \leq R(\theta,\sigma_2)$ for all $\theta \in \Theta$

ii) A decision rule σ_1 is said to be better than a rule σ_2 if,

$R(\theta,\sigma_1) \leq R(\theta,\sigma_2)$ for all $\theta \in \Theta$

and $R(\theta,\sigma_1) < R(\theta,\sigma_2)$ for at least one $\theta \in \Theta$

iii) A rule σ_1 is said to be equivalent to a rule σ_2 if,

$R(\theta,\sigma_1) = R(\theta,\sigma_2)$ for all $\theta \in \Theta$.

Definition (1.2.2) :

A rule σ is said to be admissible if there exists no rule better than σ.

A rule is said to be inadmissible if it is not admissible.

Thus any admissible rule is one that cannot be dominated.

It is clear that an inadmissible decision rule should not be used, since a decision rule with smaller risk can be found.

Remark :

In the above definition 'there exists no rule' we mean there exists no rule in a specified class D of decision rules. If the class D is not specified it is understood to be the class of all possible decision rules. Sometimes D is specified by the form of the decision rules or by certain desired property of the rule e.g.

(I) $D = \{Cx : C > 0\}$

(II) $D = \{\sigma(x) : E(\sigma(x)) = \theta\}$, the class of all unbiased estimates.

Example (1.2.1):

$X \sim P(\theta)$, $\Theta = (0,\infty)$, $A = [0,\infty]$.

The loss function $L(\theta,a) = (\theta-a)^2$

Consider the decision rules of the form

$\sigma_C(x) = cx$.

Now $R(\theta,\sigma_c) = E_\theta^X L(\theta,\sigma_c(x))$

$= E_\theta^X (\theta-\sigma_c)^2$

$= E_\theta^X (\theta-cX)^2$

$= C^2\theta + \theta^2(1-C)^2$

$R(\theta,\sigma_1) = \theta$

$R(\theta,\sigma_1) = \theta < R(\theta,\sigma_c) = C^2\theta + \theta^2(1-C)^2$.

This implies, σ_C are inadmissible if $C > 1$.

On the other hand, for $0 \leq c \leq 1$ the rules are non-comparable. e.g. The risk functions of the rules σ_1 and σ_{γ_2} are graphed, they clearly cross. It is seen that for $0 \leq C \leq 1$, σ_C is admissible.

Thus the 'standard' estimator σ_1 is admissible.

From the above example it is clear that the admissibility gives no assurance that the decision rule is quite appropriate.

Note that

$R(\theta,\sigma_{\gamma_2}) < (>) R(\theta,\sigma_1)$, $0 < \theta < 3$, $(\theta > 3)$

<u>Definition</u> (1.2.3) :

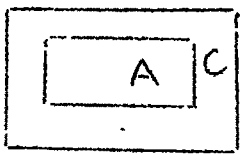
A class C of decision rules, $C \subset D^*$ is said to be complete, if given any rule $\sigma \in D^*$ not in C, there exists a rule $\sigma_0 \in C$ that is better than σ.

A class C of decision rules is said to be essentially complete if given any rule σ(not in C), there exists a rule $\sigma_0 \in C$ that is as good as σ.

<u>Lemma</u> (1.2.1) :

If C is a complete class, and A denotes the class of all admissible rules then $A \subset C$.

<u>Proof</u> :



Let σ be a rule such that $\sigma \notin C$, since C is complete there exists a rule $\sigma_0 \in C$ such that σ_0 is better than σ.

i.e. $R(\theta,\sigma_0) \leq R(\theta,\sigma)$ for all $\theta \in \Theta$

and $R(\theta,\sigma_0) < R(\theta,\sigma)$ for some $\theta \in \Theta$.

If σ were admissible then there should not be any rule better than σ.

Therefore, σ is inadmissible.

That is $\sigma \notin A$

Thus $\sigma \notin C \Rightarrow \sigma \notin A$

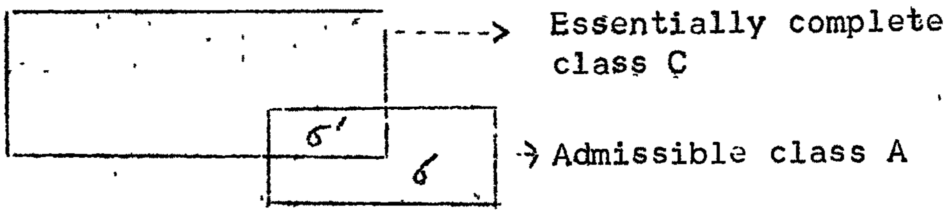
it follows that $\sigma \in A \Rightarrow \sigma \in C$

$\therefore A \subset C$ □

<u>Lemma</u> (1.2.2) :

If C is essentially complete class and there exists an admissible $\sigma \notin C$, there exists a $\sigma' \in C$ which is equivalent to σ.

<u>Proof</u> :



Since C is essentially complete and $\sigma \notin C$ then there exists a decision rule $\sigma' \in C$ which is as good as σ.

Further, since σ is admissible, σ' cannot be better than σ. i.e.

i.e. there exists $\sigma' \in C$ such that σ' is as good as σ and σ' is not better than σ.

i.e, σ and σ' are equivalent. □

<u>Definition</u> (1.2.4) :

A class C of decision rules is said to be <u>minimal complete</u> if C is complete and if no proper subclass of C is complete.

<u>Definition</u> (1.2.5) :

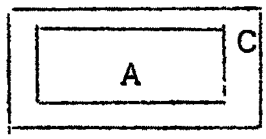
A class E of decision rules is said to be <u>minimal essentially complete</u> if C is essentially complete and if no proper subclass of C is essentially complete.

Note : $E \subset C$.

<u>Theorem</u> (1.2.1) :

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

<u>Proof</u> :



→ (Essentially complete)

As we have shown that $A \subset C$. It is enough to show that their exist no rule σ such that $\sigma \in C$ and $\sigma \notin A$.

If there exists a $\sigma \in C \cap A'$ then $C - \sigma$ will be complete, contradicting the fact that C is minimal complete. Hence $C \cap A'$ is empty. i.e. $C \cap A' = \emptyset \Longrightarrow C \subset A$

hence $A = C$ □

<u>Lemma</u> (1.2.3) :

Let σ be a Bayes (admissible) estimator of $g(\theta)$ for squared error loss. Then, $a\sigma+b$ is Bayes (admissible) for $a\,g(\theta)+b$.

<u>Proof</u> :

This follows immediately from the fact that

$$R(a\,g(\theta)+b,\ a\sigma+b) = a^2\,R(g(\theta),\sigma)$$ □

<u>Lemma</u>(1.2.4) :

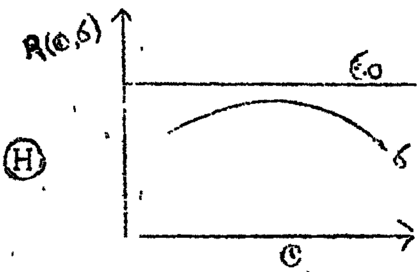
If an estimator has constant risk and is admissible, it is minimax.

<u>Proof</u> :

Let σ_o be admissible and has constant risk, i.e. $R(\Theta, \sigma_0) = C$ for all $\Theta \in \text{Ⓗ}$

Suppose σ_0 is not minimax, there exist a rule σ such that

$$\sup_{\Theta} R(\Theta,\sigma) < \sup_{\Theta} R(\Theta,\sigma_0)$$



Since $R(\Theta,\sigma_0) = C$ for all $\Theta \in \text{Ⓗ}$

$R(\Theta,\sigma) < R(\Theta,\sigma_0)$ for all $\Theta \in \text{Ⓗ}$

it implies that, σ_0 is not admissible, which is a contradiction.

Therefore, σ_0 is minimax. □

Hodges and Lehman (1951) have shown that for a sample of n independent observations from a univariate normal population the sample mean is an admissible estimator of the parent mean. Now, we will discuss, for a sample of n independent observations from a poisson population the sample mean ($\bar{X}$) is an admissible estimator of the parent mean. This problem is solved by two different procedures with respect to the squared error loss.

I) By Information Inequality Method.

II) By Limiting Bayes Method.

(I) <u>Admissibility of $\bar{X}$ for poisson distribution by Information Inequality Method</u> :

Let σ be any estimator of Θ, then

$$R(\Theta,\sigma) = E\, L(\Theta,\sigma)$$

$$= E\, (\Theta-\sigma)^2$$

$$= E\, (\sigma-\Theta)^2$$

$$R(\theta,\sigma) = E\,[\,\sigma(X) - E\,\sigma(X) + E\,\sigma(X)-\theta\,]^2$$
$$= var_\theta(\sigma) + [E(\sigma(x)) - \theta]^2$$
$$= var_\theta(\sigma) + (E(\sigma)-\theta)^2$$
$$= var_\theta\,(\sigma) + b^2(\theta)$$

where $b(\theta)$ is the bias of σ i.e. $b(\theta)=E_\theta(\sigma)-\theta$

The family of density functions $\{f(x,\theta);\ \theta \in \text{Ⓗ}\}$ with respect to μ, which satisfies the Cramer-Rao regularity conditions, so that

$$var_\theta(\sigma) \geq \frac{[b'(\theta)+1]^2}{n.\ I(\theta)}$$

where $I(\theta)$ is the Information about θ contained in X.

$$\therefore\quad R(\theta\ ,\sigma) \geq b^2(\theta) + \frac{[1+b'(\theta)]^2}{n.I(\theta)} \quad . \qquad (1)$$

Consider,

$$f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}$$

$$\therefore\quad \log f = -\theta + x \log\theta - \log x!$$

$$\frac{\partial}{\partial\theta} \log f = -1 + \frac{x}{\theta}$$

$$\frac{\partial^2}{\partial\theta^2} \log f = \frac{-x}{\theta^2}$$

Therefore,

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f\right) = -E\left(\frac{-x}{\theta^2}\right)$$
$$= \frac{1}{\theta^2}E(x) = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Therefore (1) becomes,

$$R(\theta,\sigma) \geq b^2(\theta) + \frac{[1+b'(\theta)]^2 . \theta}{n} \qquad (2)$$

Suppose that $\bar{X}$ is inadmissible, then there exists any estimator σ satisfying,

$$R(\theta,\sigma) \leq R(\theta,\bar{X}) \quad \text{for all } \theta \in \textcircled{H}$$
$$\leq E(\bar{X}-\theta)^2$$
$$\leq \operatorname{var}(\bar{X})$$
$$\leq \frac{\theta}{n} \quad \text{for all } \theta \in \textcircled{H} \qquad (3)$$

From (2) and (3) we have,

$$b^2(\theta) + \frac{[1+b'(\theta)]^2\theta}{n} \leq \frac{\theta}{n} \quad \text{for all } \theta \in \textcircled{H} \qquad (4)$$

We shall then show that (4) implies,

$$b(\theta) \equiv 0 \qquad (5)$$

i.e. σ is unbiased.

i) Since $|b(\theta)| \leq \sqrt{\theta/n}$, the function b is bounded.

ii) From the fact that

$1+2b'(\theta)+(b'(\theta))^2 \leq 1$

it follows that $b'(\theta) \leq 0$, so that b is non-increasing.

iii) Next, we shall show that there exists a sequence of values θ_i tending to ∞ and such that $b'(\theta_i) \to 0$. For suppose that $b'(\theta)$ were bounded away from 0 as $\theta \to \infty$, say $b'(\theta) \leq -\epsilon$ for all $\theta > \theta_0$. Then $b(\theta)$ cannot be bounded as $\theta \to \infty$, which contradicts (i).

iv) Analogously it is seen that there exists a sequence of values $\theta_i \to 0$ and such that $b'(\theta_i) \to 0$. Inequality (4) together with (iii) and (iv)

5678
A.

shows that $b(\theta) \to 0$ as $\theta \to 0$ or $\theta \to \infty$ and (5) follows from (ii).

Since (5) implies that $b(\theta) = b'(\theta) = 0$ for all θ, then by (2) we get

$$R(\theta,\sigma) \geq \frac{\theta}{n} \quad \text{for all } \theta$$

and hence from (3) that

$$R(\theta,\sigma) \equiv \frac{\theta}{n}$$

This proves that $\bar{X}$ is admissible for θ.

(II) <u>Admissibility of $\bar{X}$ by Limiting Bayes Method</u> :

Suppose that $\bar{X}$ is inadmissible then there exists an estimator σ^* such that

$$R(\theta,\sigma^*) \leq R(\theta,\bar{X}) \quad \text{for all } \theta \in \Theta$$

$$\leq E(\bar{X}- E(\bar{X}))^2 \quad \text{for all } \theta \in \Theta$$

$$\leq \text{var}\,(\bar{X})$$

$$\leq \frac{\theta}{n} \quad \text{for all } \theta \in \Theta$$

$$\therefore \quad R(\theta,\sigma^*) \leq \frac{\theta}{n} \quad \text{for all } \theta \in \Theta$$

and $$R(\theta,\sigma^*) < \frac{\theta}{n} \quad \text{for some } \theta \in \Theta$$

Now, $R(\theta,\sigma)$ is a continuous function of θ for every σ, so that there exists $\epsilon > 0$ and $\theta_o < \theta_1$ such that

$$R(\theta,\sigma^*) < \frac{\theta}{n} - \epsilon \quad , \quad \theta_o < \theta < \theta_1$$

$$R(\theta,\sigma^*) \leq \frac{\theta}{n} \quad , \quad (\theta_o < \theta < \theta_1)^C$$

Now we have to find the Bayes estimator.

Let $X_1,\ldots,X_n$ be i.i.d. random variables from the poisson population with parameter Θ, then

$$p(x_1,\ldots,x_n;\Theta) = \frac{e^{-n\Theta}\ \Theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i\,!}$$

Let us take the prior distribution for Θ is

$$g(\Theta) = \frac{1}{u}\ e^{-\Theta\cdot\frac{1}{u}} \quad , \quad \Theta > 0,\ u > 0$$

$$= \quad 0 \qquad\qquad , \text{ otherwise.}$$

The joint p.d.f. of X and Θ is

$$h(x,\Theta) = \frac{e^{-n\Theta}\ \Theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i\,!}\ \frac{1}{u}\ e^{-\Theta\frac{1}{u}}$$

$$h(x,\Theta) = \frac{e^{-\Theta(\frac{1}{u}+n)}\ \Theta^{\Sigma x_i}}{u\cdot\prod_{i=1}^{n} x_i\,!}$$

The marginal p.d.f. of X is

$$f(x) = \int_0^\infty h(x,\Theta)\ d\Theta$$

$$= \frac{1}{u\prod_{i=1}^{n} x_i!}\int_0^\infty e^{-\Theta(\frac{1}{u}+n)}\ \Theta^{\Sigma x_i}\ d\Theta$$

$$= \frac{1}{u\cdot\prod_{i=1}^{n} x_i!}\ \frac{\Gamma(\Sigma x_i + 1)}{(\frac{1}{u}+n)^{\Sigma x_i+1}}$$

The posterior distribution of Θ given X = x is

$$h(\Theta/X=x) = \frac{h(x,\Theta)}{f(x)}$$

$$= \frac{e^{-\Theta(\frac{1}{u}+n)}\ \Theta^{n\bar{x}}\ (\frac{1}{u}+n)^{n\bar{x}+1}}{n\bar{x}\ \Gamma n\bar{x}}$$

The Bayes estimator is $\sigma(x)$ = mean of the posterior distribution since loss function is squared error.

Therefore,

$$\sigma(x) = \int_0^\infty \frac{\Theta \; e^{-\Theta(\frac{1}{u}+n)} \; \Theta^{n\bar{x}} (\frac{1}{u}+n)^{n\bar{x}+1}}{n\bar{x} \; \Gamma n\bar{x}} \, d\Theta$$

$$= \frac{(n+\frac{1}{u})^{n\bar{x}+1}}{n\bar{x} \; \Gamma n\bar{x}} \int_0^\infty \Theta^{n\bar{x}+1} \, e^{-\Theta(\frac{1}{u}+n)} \, d\Theta$$

$$= \frac{(n+\frac{1}{u})^{n\bar{x}+1}}{n\bar{x} \; \Gamma n\bar{x}} \times \frac{\Gamma n\bar{x}+2}{(\frac{1}{u}+n)^{n\bar{x}+2}}$$

$$= \frac{(n\bar{x}+1) \; n\bar{x} \; \Gamma n\bar{x}}{n\bar{x} \; \Gamma n\bar{x} \; (\frac{1}{u}+n)} = \frac{n\bar{x}+1}{\frac{1}{u}+n}$$

$$\therefore \sigma(X) = \frac{u\,(n\bar{x}+1)}{1+nu}$$

$$= \frac{u\,n\bar{x}}{1+nu} + \frac{u}{1+nu}$$

$$\therefore \sigma(x) = \frac{\bar{x}}{1+\frac{1}{nu}} + \frac{1}{\frac{1}{u}+n}$$

It was seen above that $\bar{X} + \frac{1}{n}$ is the limit of the Bayes estimator as $u \to \infty$.

That is $\sigma(X) = \bar{X} + \frac{1}{n}$

$$\therefore R(\Theta,\sigma) = E(\Theta-\sigma)^2 = E(\Theta - \bar{x} - \frac{1}{n})^2$$

$$= \frac{1}{n^2} E(\,n\Theta - n\bar{x} - 1)^2$$

$$= \frac{n\Theta + 1}{n^2}$$

The Bayes risk is $r(\theta,\sigma) = r_\sigma$ say $= \int R(\theta,\sigma)g(\theta)d\theta$

$$= \int_0^\infty \left(\frac{n\theta + 1}{n^2}\right) \cdot \frac{1}{u} e^{-\theta \frac{1}{u}} d\theta$$

$$= \frac{1}{n} \int_0^\infty \theta \frac{1}{u} e^{-\theta \cdot \frac{1}{u}} d\theta + \frac{1}{n^2} \int_0^\infty \frac{1}{u} e^{-\theta \frac{1}{u}} d\theta$$

$$= \frac{1}{n}u + \frac{1}{n^2} = \frac{nu + 1}{n^2}$$

$$\therefore \quad r\sigma = \frac{nu + 1}{n^2} \qquad \text{(A)}$$

Let $r\sigma^*$ be the average risk of σ^* with respect to the prior distribution we have,

$$r\sigma^* = \int_0^\infty R(\theta,\sigma^*)\, g(\theta)\, d\theta$$

$$= \int_{\theta_0}^{\theta_1} \left(\frac{\theta}{n} - \epsilon\right) g(\theta)d\theta + \int_{(\theta_0,\theta_1)^C} \left(\frac{\theta}{n}\right) \cdot g(\theta)\, d\theta$$

$$= \int_{\theta_0}^{\theta_1} \frac{\theta}{n} g(\theta)d\theta - \epsilon \int_{\theta_0}^{\theta_1} g(\theta)d\theta + \int_{(\theta_0,\theta_1)^C} \frac{\theta}{n} g(\theta)\, d\theta$$

Combining first and third term we get,

$$r\sigma^* = \int_{\theta_0}^{\theta_1} \frac{\theta}{n} g(\theta)\, d\theta + \int_{(\theta_0,\theta_1)^C} \frac{\theta}{n} g(\theta) d\theta - \epsilon \int_{\theta_0}^{\theta_1} g(\theta)\, d\theta$$

$$= \int_0^\infty \frac{\theta}{n} g(\theta)\, d\theta - \epsilon \int_{\theta_0}^{\theta_1} g(\theta)\, d\theta$$

$$\therefore r\sigma^* = \frac{1}{n}\int_0^\infty \theta \frac{1}{u} e^{-\theta \frac{1}{u}} d\theta - \epsilon \int_{\theta_0}^{\theta_1} \frac{1}{u} e^{-\theta \frac{1}{u}} d\theta$$

$$= \frac{u}{n} - [e^{-\theta_0/u} - e^{-\theta_1/u}]\epsilon \qquad (B)$$

Now,

$$\frac{B}{A} = \frac{u}{n}\frac{n^2}{nu+1} - \epsilon\frac{[e^{-\theta_0/u} - e^{-\theta_1/u}] \cdot n^2}{nu+1}$$

$$= \frac{nu}{1+nu} - \left(\frac{n^2}{nu+1}\right) \cdot \epsilon [e^{-\theta_0/u} - e^{-\theta_1/u}]$$

$$< 1$$

i.e. $B < A$

which contradicts the fact that $r\sigma$ is the Bayes risk. Thus $\bar{X}$ is admissible estimator of θ. □

(1.3) Admissible decision rule :

Bayes rules with proper priors are uirtually always admissible. The basic reason is that if a rule with better risk $R(\theta,\sigma)$ existed, that rule would also have better Bayes risk i.e. $E\, R(\theta, \sigma)$.

We discuss the following :

I) Admissible rules need not be Bayes rules.

II) A Bayes rule (if exists) may be inadmissible.

III) Generalized Bayes rule need not be admissible.

I) <u>Admissible rules need not be Bayes rules</u> :

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables distributed like $N(\theta,1)$. Then the sample mean $\bar{X}$ is a admissible estimator of θ, with respect to the squared error loss (Ref. Lehmann(1982) pp 265).

Let the prior distribution of θ is $N(0,K)$, $K=1,2,\ldots$ The Bayes estimator with respect to the squared error loss is,

$$\sigma^*(X) = \bar{X}\left(1 + \frac{1}{nK}\right)^{-1}, \quad K = 1,2,\ldots.$$

The Bayes risk of $\sigma^*(X)$ is

$$r\,(\theta,\sigma^*(X)) = \frac{1}{n}\left(1+ \frac{1}{nK}\right)^{-2} + \frac{K}{(1+nK)^2}$$

Therefore,

$$r\,(\theta,\sigma^*(X)) = \frac{1}{n}\left(1+ \frac{1}{nK}\right)^{-2} + \frac{K}{(1+nK)^2} < r(\theta,\sigma) = \frac{1}{n}$$

where $\sigma(X) = \bar{X}$

it implies that, σ is not Bayes that is $\sigma(X) = \bar{X}$ is not Bayes.

From above it is clear that, admissible estimator need not be Bayes rule. □

II) <u>A Bayes rule (if exists) may be inadmissible</u> :

<u>Theorem</u> (1.3.1) :

Let $\Theta = (-\infty, \infty)$. If the risk function $R(\theta,\sigma)$ is continuous in θ for each σ, and if $\Lambda(\theta)$ is a prior distribution over Θ, whose support is Θ, then the

Bayes estimator against Λ , σ_Λ is admissible.

<u>Proof</u> :

If σ_Λ is inadmissible, then there exists an estimator σ^* satisfying

$$R(\theta,\sigma^*) \leq R(\theta,\sigma_\Lambda) \quad \text{for all } \theta \in \text{Ⓗ}$$

and $R(\theta_1,\sigma^*) < R(\theta_1,\sigma_\Lambda)$ for some $\theta_1 \in$ Ⓗ.

There exists a positive real ϵ_1, $\epsilon_1 > 0$ such that

$$R(\theta_1,\sigma^*) \leq R(\theta_1,\sigma_\Lambda) - \epsilon_1$$

Since $R(\theta,\sigma^*)$ is continuous in θ, there exists a θ - neighborhood of θ_1, say $N(\theta_1)$, such that

$$R(\theta,\sigma^*) \leq R(\theta,\sigma_\Lambda) - \epsilon_1 \quad \text{for all } \theta \in N(\theta_1)$$

Finally,

$$r(\Lambda,\sigma^*) = \int_{N(\theta_1)} R(\theta,\sigma^*)\, d\Lambda(\theta) + \int_{\bar{N}(\theta_1)} R(\theta,\sigma^*)\, d\Lambda(\theta)$$

$$\leq \int_{N(\theta_1)} [R(\theta,\sigma_\Lambda) - \epsilon_1]\, d\Lambda(\theta) + \int_{\bar{N}(\theta_1)} R(\theta,\sigma_\Lambda)\, d\Lambda(\theta)$$

$$\leq \int_{N(\theta_1)} R(\theta,\sigma_\Lambda)\, d\Lambda(\theta) + \int_{\bar{N}(\theta_1)} R(\theta,\sigma_\Lambda)\, d\Lambda(\theta) - \epsilon_1 \int_{N(\theta_1)} d\Lambda(\theta)$$

$$\leq r(\Lambda,\sigma_\Lambda) - \epsilon_1 \Lambda(N,(\theta_1))$$

But since $\Lambda(\theta) > 0$ for all $\theta \in$ Ⓗ , $\Lambda(N,(\theta_1)) > 0$

hence,

$$r(\Lambda,\sigma^*) < r(\Lambda,\sigma_\Lambda)$$

This contradicts the fact that σ_Λ is Bayes.

hence σ_Λ is admissible. ▯

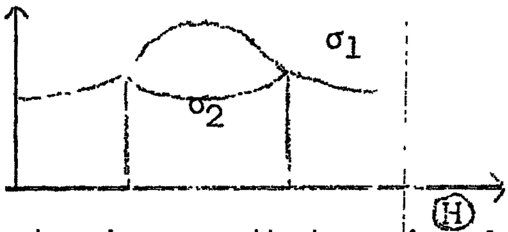
We note that if the support of prior distribution is not entire the parameter space then the Bayes estimator need not be admissible. This fact is clear from the following :

Let σ_1 and σ_2 be such that

i) σ_2 is Bayes with respect to a prior distribution $\Lambda(\theta)$, whose support is $Ⓗ_1$, proper subset of Ⓗ.

ii) $r(\Lambda, \sigma_1) = r(\Lambda, \sigma_2)$, $\theta \in Ⓗ_1$ and

$r(\Lambda, \sigma_1) > r(\Lambda, \sigma_2)$, $\theta \in Ⓗ - Ⓗ_1$



It is easy to observe that σ_1 is also a Bayes rule with respect to $\Lambda(\theta)$ and σ_1 is not admissible.

III) <u>Generalized Bayes rule need not be admissible</u> :

Let $X \sim G(\alpha,\beta)$ (α known) is observed, and that it is desired to estimate β under loss $L(\beta,a) = (\beta-a)^2\beta^{-2}$. It isdecided to use the improper prior density

$g(\beta) = \frac{1}{\beta^2}$.

$$f(x / \alpha,\beta) = \frac{1}{\Gamma\alpha\beta^{\alpha}} \quad x^{\alpha-1} e^{-x/\beta} \quad , x \geq 0, \alpha > 0, \beta > 0$$

$$= \quad 0 \qquad , \text{otherwise}$$

and

$$g(\beta) = \frac{1}{\beta^2}$$

The joint p.d.f. of X and β is

$$h(x,\beta) = \frac{1}{\Gamma\alpha\ \beta^{\alpha+2}}\ x^{\alpha-1}\ e^{-x/\beta}$$

The marginal density function of X is

$$f(x) = \frac{x^{\alpha-1}}{\Gamma\alpha} \int_0^\infty e^{-xy}\ y^{\alpha}\ dy \qquad (\text{ when } y = \frac{1}{\beta}\ )$$

$$= \frac{\alpha}{x^2}$$

The posterior distribution of β given X = x is

$$h(\beta/X = x) = \frac{1}{\alpha\,\Gamma\alpha\ \beta^{\alpha+2}}\ x^{\alpha+1}\quad e^{-x/\beta}$$

when the loss is weighted squared error, the Bayes estimator of β is given by,

$$\sigma(x) = \frac{x}{\alpha+2}$$

consider $\sigma_C(x) = C\ x$

The risk of the estimator $\sigma_C(x)$ is,

$$R(\beta,\sigma_C) = E(C\ x - \beta)^2\ \beta^{-2}$$

$$= \beta^{-2}\ E^{X}\ [\ C(X - \alpha\beta) + (C\alpha-1)\ \beta\ ]^2$$

$$= \beta^{-2}\ [\ \alpha\beta^2C^2 + (C\alpha-1)^2\beta^2\ ]$$

$$= C^2\alpha+(\alpha C-1)^2 \quad (\because E(X) = \alpha\beta,\ var(x)= \alpha\beta^2\ )$$

Differentiating with respect to C and setting equal to zero shows that the value of C minimizing this expression is unique and is given by,

$$2 C \alpha + 2 \alpha \ (C\alpha - 1) = 0$$

$$= C_o(\text{say}) = \frac{1}{1+\alpha}$$

It follows that if $C \neq C_o$, then

$$R(\beta, \sigma C_o) < R(\beta, \sigma_C) , \quad \text{for all } \beta$$

it implies that, $\sigma_C(X)$ is inadmissible.

Therefore,

$$\sigma(X) = \frac{x}{\alpha + 2} \quad \text{is inadmissible.}$$

Thus generalized Bayes rule need not be admissible. □

Lemma (1.3.1) :

Any unique Bayes estimator is admissible.

Proof :

If σ is unique Bayes with respect to the prior distribution Λ, and is dominated by σ', then

$$\int R(\theta,\sigma') \, d\Lambda(\theta) \leq \int R(\theta,\sigma) \, d\Lambda(\theta)$$

i.e. $r(\sigma') \leq r(\sigma)$

Further, since σ is Bayes we have

$$r(\sigma) \leq r(\sigma')$$

hence $r(\sigma) = r(\sigma')$

which contradicts the uniqueness. □

(1.4) Inadmissibility of linear estimates for exponential families :

In this section we consider the admissibility of a linear estimator a X+b of E(X) with respect to the squared error loss when the probability density function is given by

$$P_\theta(x) = \beta(\theta)\, e^{\theta\, T(x)}$$

We obtain sufficient conditions for admissibility of the above estimator. These results are obtained by Karlin (1958).

Theorem (1.4.1) :

Let X be a random variable with mean θ and variance σ^2 and the loss function is squared error. Then aX+b is an inadmissible estimator of θ whenever

i) $a > 1$, or

ii) $a < 0$, or

iii) $a = 1,\ b \neq 0$.

Proof :

Let the risk of aX+b be

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

i) If $a > 1$ then

$$R(a,b) \geq a^2\sigma^2 \geq \sigma^2 = R(1,0)$$

that is $R(1,0) \leq R(a,b)$

so that aX+b is inadmissible.

ii) If $a < 0$, then $(a-1)^2 > 1$

hence

$$R(a,b) \geq [(a-1)\theta + b]^2$$
$$\geq (a-1)^2 \left(\theta + \frac{b}{a-1}\right)^2$$
$$\geq \left(\theta + \frac{b}{a-1}\right)^2$$
$$\geq R\left(0, -\frac{b}{a-1}\right)$$

So that $aX+b$ is dominated by the constant estimator $\sigma \equiv -b / a-1$

iii) If $a = 1$, $b \neq 0$ we have

$$R(a,b) = \sigma^2 + b^2 \geq \sigma^2 = R(1,0)$$

$\therefore$ $R(1,0) \leq R(a,b)$

hence

$aX+b$ is inadmissible. $\square$

Theorem (Karlin 1958) (1.4.2) :

Let X have probability density

$$P_\theta(x) = \beta(\theta)\, e^{\theta T(x)} \qquad (\theta, T \text{ real valued}) \qquad (1)$$

with respect to μ and let $\textcircled{H}$ be the natural parameter space with end points say $\underline{\theta}$ and $\bar{\theta}$ $(-\infty \leq \underline{\theta} \leq \bar{\theta} \leq \infty)$.

Let θ_0 be a point in $(\underline{\theta}, \bar{\theta})$

i.e. $\underline{\theta} < \theta_0 < \bar{\theta}$ and u, $0 \leq u < \infty$ a value for which

$$\lim_{\theta \to \bar{\theta}} \int_{\theta_0}^{\theta} \frac{e^{-\gamma u \theta}}{[\beta(\theta)]^u}\, d\theta = \infty \qquad (2)$$

and

$$\lim_{\Theta \to \underline{\Theta}} \int_{\Theta}^{\Theta_0} \frac{e^{-\gamma u \Theta}}{[\beta(\Theta)]^u} \, d\Theta = \infty \qquad (3)$$

Then

$\frac{T}{1+u} + \frac{\gamma u}{1+u} = S$ (say) is an admissible estimator for estimating $g(\Theta) = E_\Theta(T)$ with respect to the squared error loss.

Proof :

$$p(x,\Theta) = \beta(\Theta)\, e^{\Theta T(x)}$$

$$\therefore \quad \log p = \log \beta(\Theta) + \Theta . T$$

$$\frac{\partial}{\partial\Theta} \log p = \frac{\partial}{\partial\Theta} \log \beta(\Theta) + T$$

$$E\left(\frac{\partial}{\partial\Theta} \log p\right) = E\left(\frac{\partial}{\partial\Theta} \log \beta(\Theta) \right) + E(T) = 0$$

$$\therefore \quad E_\Theta(T) = - \frac{\partial}{\partial\Theta} \log \beta(\Theta)$$

$$\therefore \quad E_\Theta(T) = \frac{-\beta'(\Theta)}{\beta(\Theta)} = g(\Theta) \text{ (say)} \qquad (4)$$

$$\begin{aligned} \text{var}_\Theta(T) &= E\,(T - E_\Theta(T))^2 \\ &= E\left(T + \frac{\partial}{\partial\Theta} \log \beta(\Theta)\right)^2 \\ &= E\left(\frac{\partial}{\partial\Theta} \log p \right)^2 \\ &= -E\left(\frac{\partial^2}{\partial\Theta^2} \log p\right) \\ &= -\frac{\partial^2}{\partial\Theta^2} \log p \\ &= \frac{\partial}{\partial\Theta}\left(-\frac{\partial}{\partial\Theta} \log p\right) \end{aligned}$$

$$\therefore \quad \text{var}_\Theta(T) = g'(\Theta) = I(\Theta) \qquad (5)$$

where $I(\Theta)$ is the Fisher information.

For any estimator $\sigma(X)$ we have

$$E_\Theta[\sigma(X) - g(\Theta)]^2 \geq \text{var}(\sigma(X)) + b^2(\Theta)$$

where

$b(\Theta) = E_\Theta(\sigma(X)) - g(\Theta)$ is the bias of σ.

$$\therefore \quad E_\Theta[\sigma(X)-g(\Theta)]^2 \geq b^2(\Theta) + \frac{[I(\Theta)+b'(\Theta)]^2}{I(\Theta)} \qquad (6)$$

Suppose that there exists an estimator σ_0 such that

$$E_\Theta\left[\frac{T+\gamma u}{1+u} - g(\Theta)\right]^2 \geq E_\Theta[\sigma_0(X) - g(\Theta)]^2 \text{ for all } \Theta \qquad (7)$$

L.H.S. of (7) is,

$$E_\Theta\left[\frac{T+\gamma u}{1+u} - g(\Theta)\right]^2 = \frac{1}{(1+u)^2} E_\Theta[T+\gamma u - g(\Theta) - u\, g(\Theta)]^2$$

$$= \frac{1}{(1+u)^2} E_\Theta[(T-g(\Theta)) - u(g(\Theta)-\gamma)]^2$$

$$= \frac{1}{(1+u)^2} E_\Theta[(T-g(\Theta))^2 + u^2(g(\Theta)-\gamma)^2 - 2u(g(\Theta)-\gamma)(T-g(\Theta))]$$

$$= \frac{1}{(1+u)^2} E_\Theta(T-g(\Theta))^2 + \frac{u^2}{(1+u)^2} E(g(\Theta)-\gamma)^2$$

$$- \frac{2u}{(1+u)^2}(g(\Theta)-\gamma) E_\Theta(T-g(\Theta))$$

using (4) and (5) we have,

$$E_\Theta\left(\frac{T+\gamma u}{1+u} - g(\Theta)\right)^2 = \frac{I(\Theta)}{(1+u)^2} + \frac{u^2(g(\Theta)-\gamma)^2}{(1+u)^2} \qquad (8)$$

Application of (6) to σ_o,

$$E_\theta(\sigma_o(X) - g(\theta))^2 \geq b_o^2(\theta) + \frac{[I(\theta)+b_o'(\theta)]^2}{I(\theta)} \qquad (9)$$

Thus from (7) we have

$$b_o^2(\theta) + \frac{[I(\theta)+b_o'(\theta)]^2}{I(\theta)} \leq \frac{I(\theta)}{(1+u)^2} + \frac{u^2[g(\theta)-\gamma]^2}{(1+u)^2} \qquad (10)$$

Let us now choose σ to be the estimator S, put

$$h(\theta) = b_o(\theta) - b(\theta) \qquad (11)$$

Now,

$$b(\theta) = E_\theta(\sigma(X)) - g(\theta)$$

$$= E_\theta\left(\frac{T+\gamma u}{1+u}\right) - g(\theta)$$

$$= \frac{u}{1+u}(\gamma - g(\theta)) \qquad (12)$$

$$\therefore \quad b'(\theta) = -\frac{u}{1+u} g'(\theta) \qquad (13)$$

using (11), (10) can be written as,

$$[h(\theta)+b(\theta)]^2 + \frac{[I(\theta) + h'(\theta) + b'(\theta)]^2}{I(\theta)}$$

$$\leq \frac{I(\theta)}{(1+u)^2} + \frac{u^2[g(\theta)-\gamma]^2}{(1+u)^2}$$

$$\therefore h^2(\theta) + 2h(\theta)b(\theta) + b^2(\theta) + \left[\frac{h'(\theta)+I(\theta)+b'(\theta)}{I(\theta)}\right]^2$$

$$\leq \frac{I(\theta)}{(1+u)^2} + \frac{u^2(g(\theta)-\gamma)^2}{(1+u)^2}$$

using (12) we have,

$$h^2(\theta) - 2h(\theta)\frac{u}{1+u}(g(\theta)-\gamma) + \frac{[h'(\theta)+(I(\theta)+b'(\theta))]^2}{I(\theta)} - \frac{I(\theta)}{(1+u)^2} \leq 0$$

$$h^2(\theta) - \frac{2u}{1+u}\, h(\theta)(g(\theta)-\gamma) + \frac{h'^2(\theta)}{I(\theta)} + \frac{2h'(\theta)(I(\theta)+b'(\theta))}{I(\theta)}$$

$$+ \frac{[I(\theta)+b'(\theta)]^2}{I(\theta)} - \frac{I(\theta)}{(1+u)^2} \leq 0 \qquad (14)$$

using (5) and (13)

consider

$$\frac{I(\theta)+b'(\theta)}{I(\theta)} = \frac{I(\theta) - \frac{u}{1+u}g'(\theta)}{I(\theta)} = \frac{1}{1+u}$$

Therefore, (14) reduces to,

$$h^2(\theta) - \frac{2u}{1+u}\, h(\theta)[g(\theta)-\gamma] + \frac{[h'(\theta)]^2}{I(\theta)} + \frac{2h'(\theta)}{1+u} + \frac{I(\theta)}{(1+u)^2} - \frac{I(\theta)}{(1+u)^2} \leq 0$$

$$\therefore\ h^2(\theta) - \frac{2u}{1+u}\, h(\theta)[g(\theta)-\gamma) + \frac{2}{1+u}\, h'(\theta) + \frac{[h'(\theta)]^2}{I(\theta)} \leq 0$$

which implies

$$h^2(\theta) - \frac{2u}{1+u}\, h(\theta)[g(\theta)-\gamma] + \frac{2}{1+u}\, h'(\theta) \leq 0 \qquad (15)$$

Finally let

$$K(\theta) = h(\theta)\, \beta^u(\theta)\, e^{\gamma u \theta} \qquad (16)$$

Differentiation of $K(\theta)$ and use of (4), (15)

reduces to,

$$K^2(\theta)\, \beta^{-u}(\theta)\, e^{-\gamma u \theta} + \frac{2}{1+u}\, K'(\theta) \leq 0 \qquad (17)$$

We shall show that $K(\theta) \geq 0$ for all θ.

Suppose that $K(\theta_o) < 0$ for some θ_o. Then $K(\theta) < 0$ for all $\theta \geq \theta_o$ and for $\theta > \theta_o$ we can write (17) as,

$$\frac{d}{d\theta}\left[\frac{1}{K(\theta)}\right] \geq \frac{1+u}{2}\, \beta^{-u}(\theta)\, e^{-\gamma u \theta} \qquad (18)$$

integrating (18) both sides with respect to θ from θ_o to θ we get

$$\frac{1}{K(\theta)} - \frac{1}{K(\theta_o)} \geq \frac{1+u}{2} \int_{\theta_o}^{\theta} \beta^{-u}(\theta)\; e^{-\gamma u \theta}\, d\theta \qquad (19)$$

As $\theta \rightarrow \bar{\theta}$, the right hand side tends to ∞, while the left hand side of (19) is bounded by $-\frac{1}{K(\theta_o)}$,

Which is a contradiction this implies $K(\theta) \geq 0$ for all θ.

Similarly $K(\theta) \leq 0$ for all θ

It follows that $K(\theta) = 0$, for all θ

and hence $h(\theta) = 0$ for all θ.

This shows that for all θ equality holds in (15), (10), and thus (7).

Therefore, there is no estimator σ_o which is better than S, hence S is admissible. □

Example (1.4.1) :

We know that for poisson distribution ; f(x,m) the probability density function with respect to the counting measure defined on the set of non-negative integers is given by ,

$$f(x, m) = \frac{e^{-m}\, m^{x}}{x!}\ , \quad x = 0,1,2,..,\infty, \quad m > 0$$

However, if we take the σ-finite measure $\mu(x) = \frac{1}{x!}$, $x = 0,1,2,\ldots$ as dominating measure for poisson distribution the corresponding density function is given by

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

$$= e^{-m} e^{x \log m}$$

putting $\theta = \log m \therefore m = e^{\theta} \Rightarrow -\infty < \theta < \infty$

Therefore,

$$P_{\theta}(x) = e^{-e^{\theta}} e^{x\theta} \; ; \; \theta \in (-\infty,\infty) \qquad (20)$$

compairing (20) with (1) we get

$$\beta(\theta) = e^{-e^{\theta}} \quad \text{and } T = X$$

From Karlin's theorem (1.4.2) we know that $S = \frac{T}{1+u} + \frac{\gamma u}{1+u}$ $= \frac{X}{1+u} + \frac{\gamma u}{1+u}$ is admissible for its expected value provided the two integrals (2) and (3) are satisfied. Note that this condition is satisfied for $u = 0$

i.e.

$$\int_{\theta_0}^{\bar{\theta}} \frac{e^{-\gamma u \theta}}{(\beta(\theta))^u} d\theta = \int_{\theta_0}^{\bar{\theta}} d\theta = \bar{\theta} - \theta_0 = \infty$$

and

$$\int_{\underline{\theta}}^{\theta_0} \frac{e^{-\gamma u \theta}}{(\beta(\theta))^u} d\theta = \int_{\underline{\theta}}^{\theta_0} d\theta = \theta_0 - \underline{\theta} = \infty$$

hence S is admissible that is X is admissible for $E(X) = m$.

Since $T(X) = \sum_{i=1}^{n} X_i \sim P(nm)$

$T(X)$ is admissible for nm and hence from Lemma (1.2.3),

$\frac{T(X)}{n} = \bar{X}$ is admissible for m.

Thus the sample mean $\bar{X}$ of the poisson distribution is admissible for estimating m.