CHAPTER- 0

INTRODUCTION

Statistical decision theory was introduced by Abraham Wald (1939) as a generalization of the classic statistical theories of hypothesis testing and estimation. After Wald's work important contributions have been made by Girshick and Savage (1951) and Stein (1956). A detailed introduction is given by Ferguson (1967) and Berger (1980) in their texts.

In a decision problem the statistician has to select a decision rule from a set of available decision rules. A strategy (also called a decision rule) is a plan that tells how to use the data to select a decision and is evaluated by reference to the cost of its expected consequences. The difficulty in selecting the best strategy derives from the fact that the consequences of a decision depend on the unknown state of nature.

'' The proper role of decision theory is the subject of considerable controversy. Because it makes essential use of costs, many statisticians feel that decision theory may be suitable for problems of the market place but not those of pure science. Another issue is the difficulty in assigning costs or values to the consequences'' (Encyclopedia PP 131).

Testing hypothesis, estimation and confidence intervals are the three main classical theories of statistical inference. A hypothesis testing problem corresponding to a two-action problem. The possible actions are to accept a hypothesis or to reject it.

In estimation problems one attaches a loss to estimating Θ by t. A popular special case is the loss $(t-\Theta)^2$. Then a good procedure tends to minimize $E(t-\Theta)^2$ (called the mean squared error of the estimator), which is the traditional measure of the goodness of an estimate.

The theory of confidence intervals is mathematically elegant, appealing, and popular but does not fit naturally in to the decision theory framework. The introduction of costs in a reasonable fashion would undoubtedly lead to substantial modification of the theory of confidence intervals.

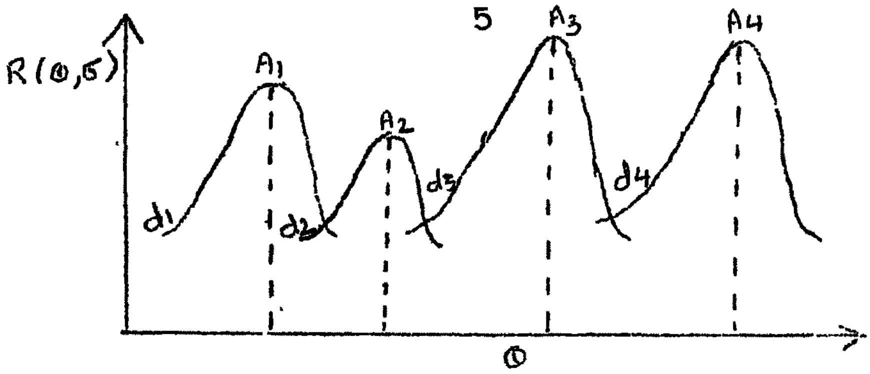
As we are interested in the study of admissible rules; so for this we introduce the preliminaries of this theory and are given in Chapter I of the dissertation. In the following we give the chapterwise summary of the dissertation.

A common formulation of statistical decision theory involves the sample space and the class of probability density functions (pdf) $f(x,\Theta)$ where Θ is a parameter, $\Theta \in \textcircled{H}$. A problem of interest is to choose an 'optimum' decision rule $d(x)$ (estimator) from the set of all possible decisions (Θ) based on a sample $X_1, X_2, \ldots, X_n$ drawn from the distribution with pdf $f(x,\theta)$. That is based on a random sample $X_1, X_2, \ldots, X_n$ we wish to estimate θ, corresponding to a specified loss function $L(\theta,d)$. The loss function $L(\theta,d)$ represents the loss incurred if we estimate θ by $d(x)$. Further $E_\theta\, L(\theta,d)$, the average loss of the estimator $d(x)$, $E_\theta\, L(\theta,d)$ called the risk of the estimator $d(x)$ and is denoted by $R(\theta,d)$, the related terms are explained in section 1.0. We further note that two estimators that is decision rules say $d_1(x)$ and $d_2(x)$, can be compared based on their corresponding risks; preference is to be given to that estimator with smaller risk. In general the risk functions of two estimators may cross, that is one risk function being smaller for some θ and the other smaller for other θ. Then since θ is unknown, it is not possible to make a choice between the two estimators. If one has some additional information about θ, let this information about θ (called the prior information) be expressed in the form of a density function $\pi(\theta)$ with the support Θ. We have a natural way of removing the dependence of the risk function on θ, namely by averaging out the θ, using the density $\pi(\theta)$. The Bayes risk of an estimator is an average risk, the average being taken over the parameter space Θ with respect to the prior density

$\pi(\Theta)$. For given loss function $L(\Theta,d)$ and prior density $\pi(\Theta)$ the Bayes risk of an estimator is a real number. So two competing estimators can be easily compared by comparing their repective Bayes risk, prefering that estimator with smaller Bayes risk. An estimator with smallest Bayes risk is called a Bayes estimator and this will be discussed in section 1.1. Even in those problems when the prior distribution is unknown the concept of Bayes estimation can benefit us. Each possible prior distribution has a corresponding estimator, whose merits can be judged by using our standard methods of comparison. Bayes estimation is useful in obtaining an estimator possessing some desirable property that does not depend on prior distribution. The property of minimax is such a property and this will be discussed in section 1.1. According to minimax principle we choose a decision rule for which the maximum of risks (over the parameter space Ⓗ) is minimum, that is a rule σ^* is minimax if,

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

where D* is the class of all randomized decision rules. Let the problem be to choose one among the rules d_1, d_2, d_3 and d_4 which have the risk function as shown in the following figure.



Observe that there is no rule which is uniformly better (smaller risk) than the others. However according to minimax principle d_2 is to be preferred, as it has smallest maximum of the risk.

Another procedure is by restricting to the class of estimators that are unbiasedness or/and invariant, which will be discussed in section 1.1 .

Bayes principle and minimax principle have there own limitations. As the Bayes rule depends very much on the prior distribution, it is not desirable to adopt Bayes principle when the choice of prior distributions is not strongly justified. In such situations one can adopt the criteria of admissibility wherein we compare the decision rules via risk functions. This concept of admissibility will be discussed in section 1.2. A rule σ is said to be admissible if there exists no rule better than σ. 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. In section 1.2 we will show that admissiblity

of $\hat{X}$ (with respect to the squared error loss) for m, the mean of the poisson distribution by using two different approaches. In section 1.4 the admissibility of linear estimates in case of exponential family will be discussed. In this case, we consider the admissibility of linear estimates aX+b of $E_\Theta(X)$ with respect to the squared error loss when the probability density function is of the form

$$P(x, \Theta) = \beta(\Theta)\ e^{\Theta\ T(x)} \qquad (1)$$

Let Ⓗ be the natural parameter space. Then Ⓗ is an interval with end points, say, $\underline{\Theta}$ and $\bar{\Theta}$ $(-\infty \leq \underline{\Theta} \leq \bar{\Theta} \leq \infty)$. We discuss the admissibility of such estimators aX+b for different values of 'a' and 'b'. It can be shown that if (i) $a > 1$ (ii) $a < 0$ (iii) $a = 1$, $b \neq 0$, the estimate aX+b is an inadmissible for estimating Θ when X is a random variable with mean Θ and variance σ^2, with respect to the squared error loss. Admissibility of linear estimate aX+b for $b = 0$ was first discussed by Karlin (1958). He considered the one parameter exponential family as in (1). His proof was extended to the other values of 'b' by Ping (1964). By the above result aX+b is inadmissible if $a < 0$ or $a > 1$ and is a constant for $a = 0$, To state Karlin's sufficient conditions in the remaining cases, it is convenient to write the estimator as

$$\frac{T}{1+u} + \frac{\gamma u}{1+u} = S(\text{say})$$

with $0 \leq u < \infty$ corresponding to $0 < a \leq 1$.

Under the above assumptions, a sufficient condition for the admissibility of the estimator S for estimating $E_\Theta(T) = g(\Theta)$ (say) with respect to the squared error loss the two integrals, that is

$$I_1 = \int_{\Theta_o}^{\Theta} \frac{e^{-\gamma u \Theta}}{[\beta(\Theta)]^u} \, d\Theta$$

and

$$I_2 = \int_{\Theta}^{\Theta_o} \frac{e^{-\gamma u \Theta}}{[\beta(\Theta)]^u} \, d\Theta$$

tend to infinity as Θ tends to $\bar{\Theta}$ and $\underline{\Theta}$ respectively. This result is given in the form of a theorem 1.4.2 due to Karlin (1958). An example is given to illustrate these results.

In continuation of Karlin's theorem we discuss in chapter II the admissibility of estimators for the exponential families with quadratic loss function. Suppose we have to estimate $g(\Theta)$ by $a(x)$ corresponding to the squared error loss. The risk for the estimate $a(x)$ when the true parameter value is Θ is calculated by using the formula,

$$R(\Theta, a) = \int [a(x) - g(\Theta)]^2 \, P(x,\Theta) \, d\mu(x)$$

where $p(x,\Theta)$ is the density function of x with respect to a σ-finite measure μ. Our object is to select the estimate $a(x)$ which minimizes $R(\Theta, a)$. The quadratic loss as a measure of the discrepancy of an estimate from the two characteristics (i) when a(x) represents the unbiased estimateof $g(\Theta)$,then $R(\Theta,a)$ is the variance of a(x) and (ii) from a technical and mathematical view point squared error loss is easy to manipulation and computation. So for this we shall use the squared error loss for estimation of $g(\Theta)$. In section 2.1 for our convenience we take $\beta(\Theta)$ as devisor instead of multiplier as taken in (1). We have seen the estimate aX+b is inadmissible if $a < 0$ or $a > 1$. If $b = 0$ in the above, then aX is inadmissible if $a < 0$ or $a > 1$. Hence the admissibility of aX is to be discussed only for $0 \leq a \leq 1$. For our convenience $a = \frac{1}{1+u}$ and the admissibility of $x(1+u)^{-1}$ is to be discussed only for $u \geq 0$. Karlin has considered the admissibility of linear estimate $X(1+u)^{-1}$, $u \geq 0$ for estimating $E_\Theta(X)$ and proved the result as given in theorem 2.1.1. Karlin has conjectured that the conditions in theorem 2.1.1. are not merely sufficient, but are necessary also for the admissibility of $x(1+u)^{-1}$. Karlin has shown that $x(1+u)^{-1}$ is inadmissible for $g(\Theta)$ for certain values of u. In this respect he proposed a result as given in lemma 2.1.1. Further work is done by Joshi (1969) he gave the imprcved criteria for inadmissibility of the estimate $x(1+u)^{-1}$ in the form of lemma as given in 2.1.2. Also the necessary conditions for the convergence or divergence of the integrals in theorem 2.1.1 are easily obtained. Now using the improved criteria as in lemma 2.1.2, the range of values of u for which Karlin's conjectures remains open is narrowed down, which can be seen as given in the diagram page no. 77 . An example is given to illustrate these results.

The effect on admissibility due to truncation will be discussed in section 2.2 and the results are due to Kale (1964). Let Θ_T denotes the natural range of the parameter when the distribution is truncated and note that $\Theta_T \supseteq \Theta$. As the admissibility of an estimate is closely connected with the structure of the natural range of the parameter, the admissibility of an estimate may be destroyed by truncation. This will be shown by giving suitable examples. In lemma 2.2.3 we show that for any mode of truncation all the estimates γx , $\gamma > 1$ continue to be inadmissible even after truncation. Further an example 2.2.3 is given to show that an admissible estimator continue to be admissible one even after truncation.
Now consider the negation of the statement of Karlin's theorem, we have the following result:
If $x(1+u)^{-1}$ is inadmissible then one of the integrals must be convergent. If Karlin's conjecture be true then the convergence of atleast one of the integrals implies inadmissibility of $x(1+u)^{-1}$. Then, when $\Theta = \Theta_T$

it can be shown that (refer lemma 2.2.3) if Karlin's conjecture be true than an inadmissible estimate $x(1+u)^{-1}$, $u \geq 0$ continues to remain inadmissible after truncation. If in the non-truncated case certain estimate is inadmissible but in the truncated case, the estimator is admissible, such type of transition has not occured. If such a case is possible then from lemma 2.2.3 it implies that Karlin's conjecture is not true.

The admissibility of scale parameter in the exponential family will be discussed in section 2.3. Zidek (1969) has shown that when the estimation problem is invariant under a group of transformations and the induced group $\bar{G}$ acts transitively on the parameter space the best invariant estimator is formal Bayes. Portnoy (1971) has given sufficient conditions for the admissibility of a formal Bayes estimator when the loss is quadratic. So to begin with we give Portnoy conditions and then apply Portnoy conditions for estimating a power of the scale parameter by the best scale invariant estimator. For a ready reference we state the Portnoy (1971) theorem in which the formal Bayes estimator is admissible under certain conditions as in theorem 2.3.1. With this background we discuss the results of Sharma (1973) for admissibility of scale parameter. It is shown that under certain condition the formal Bayes estimator is admissible (see theorem 2.3.2), basically the result of Portnoy (1971) is used. An example is given to illustrate the result.

In chapter III we shall discuss inadmissibility of some standard estimate and the admissibility of confidence interval in presence of prior information. Consider the binomial distribution, the conventional estimate is $\frac{X}{n}$. Let Θ have a distribution which belongs to a subclass of the distributions on [0,1] as a prior information to the experimenter. Now we regard the binomial distribution to be conditional on Θ, the members of this subclass generate a family of joint distributions for X and Θ. With this as background we may view our problem as a special case of conventional prediction theory. We give a generalized maximum likelihood principle as applied to this example and investigate a class of predictors which it suggests. Under appropriate conditions, each of these has a uniformally smaller mean square error than the conventional estimate. These results are due to Skibinsky and Cote (1964), and these results will be discussed in section 3.1. In case of point estimation the problem of admissibility of a location parameter was treated by Blyth (1951), Blackwell (1951), Farrell (1964), Brown (1966). In each paper above, the admissibility require the existence of one more moment than what is needed for finite risk. Now we can show that, a unique best translation invariant estimate may be inadmissible if a certain moment condition fails to be satisfied. These results are due to Perng (1970), and this will be discussed in section 3.2.

Farrell (1964) has proved regarding the admissibility of estimators of the location parameter in a class of frequency functions. The analogous question regarding confidence intervals is considered in section 3.3. Joshi (1966) proved a theorem which gave a set of sufficient conditions for the admissibility of a certain confidence interval procedures for a location parameter. Instead of this theorem we state a simple theorem (due to Fox Martin, Perng (1970)) which includes moment condition. Now we show that a certain translation invariant confidence interval procedure may be inadmissible if a certain moment condition fails to hold, by giving an example in the form of theorem. These results are due to Perng (1970).