91

105300

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Some admissible nonparametric tests and a minimal complete class theorem

Li, Seung-Chun, Ph.D.

Iowa State University, 1990

.

Some admissible nonparametric tests and

a minimal complete class theorem

by

Seung-Chun Li

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

In Charge of Major Work

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa
1990

TABLE OF CONTENTS

iii

1. INTRODUCTION

1.1 Motivation

The problem of *statistical inference* arises when some aspect of the situation underlying the mathematical model is not known. The consequence of such a lack of knowledge is uncertainly as to the best behavior.

To formalize this, consider a sample space $\mathcal{X}$, a family $\{P_\theta : \theta \in \Theta\}$ of probability distributions on $\mathcal{X}$, and the set of all possible states of nature Θ. Typically, when experiments are performed to obtain information about θ, the experiments are designed so that the observations are distributed according to some probability distribution which has θ as an unknown parameter. In such situations, θ will be called the parameter and Θ the parameter space. Let $\mathcal{D}$ be the set of all possible decisions with a typical element denoted by d. We assume that a loss function $L(\theta, d)$ has been specified, where $L(\theta, d)$ represents the loss incurred if our decision is d when θ is the parameter of the distribution from which we sampled.

The problem is to determine a decision rule for each possible value of x. Mathematically such a rule is a function δ, which assigns a decision $d = \delta(x)$ to each possible value $x \in \mathcal{X}$, that is, a function whose domain is the set of values of $\mathcal{X}$ and whose range is the set of possible decisions, $\mathcal{D}$. If δ is the decision rule we used and θ is the true value of the parameter, our loss is the random variable $L(\theta, \delta(X))$.

We can not hope to make the loss small for every possible sample, but we can try to make the loss small on the average. Hence we measure the goodness of δ by the average loss, which we denote by $r(\theta, \delta)$. Thus

$$r(\theta, \delta) = E\left[L(\theta, \delta(X))\right].$$

We refer to $r(\theta, \delta)$ as the risk of δ at θ and the aim of *statistical decision theory* is the determination of a decision function δ which minimizes the risk function in some sense.

However, even for a given L, there will not, in general, exist a decision rule with uniformly smallest risk. So it is not clear what is meant by a best procedure. This kind of difficulty stems from the dependence of the risk function on θ. One possible way to avoid this difficulty is to remove the dependence by averaging out, in some sense, just as we average out the dependence on samples. Another way is to restrict the class of decision rules which possess a certain degree of impartiality and hope we can find the best procedure in this restricted class.

However, none of these approaches is reliable in the sense that the resulting procedure is necessarily satisfactory. This suggests the possibility, at least as a first step, of not insisting on a unique solution, but asking only how far a decision problem can be reduced without loss of relevant information.

A decision rule δ can be eliminated from consideration, if there exists a decision rule δ' which dominates δ in the sense that

$$r(\theta, \delta') \leq r(\theta, \delta) \text{ for all } \theta \in \Theta$$
$$r(\theta, \delta') < r(\theta, \delta) \text{ at least one } \theta \in \Theta$$

In this case δ is said to be inadmissible; δ is called admissible if no such dominating

δ' exists. Clearly, if a decision rule turned out to be inadmissible, one would usually not want to use it. So admissibility is a desirable property which a good decision rule should possess. However, the verification of admissibility or inadmissibility is very difficult, in general, especially in nonparametric problems. Actually, the admissibility of many nonparametric decision rules is not known, even for rules that are used frequently in practice.

Recently, Meeden, Ghosh and Vardeman [15] observed that the admissibility in nonparametric problems is closely related to the admissibility in finite population sampling, and they utilized this observation to prove the admissibility of many standard nonparametric estimators. Cohen and Kuo [4] also apply a similar argument in showing that the empirical distribution function is an admissible estimator of the distribution function in finite population sampling.

In Chapter 2, we will prove the admissibility of the rank-sum test, which may be the most frequently used test in two-sample nonparametric testing problems, as well as other well-known nonparametric decision rules using an argument similar to those mentioned above.

Chapter 3 is mainly devoted to the discussion of the minimal complete class for various decision problems with a finite parameter space.

A class C of decision procedures is said to be complete if for any δ not in C, there exists δ' in C dominating it. A complete class is minimal if no proper subset of the class is complete.

From the definition, it is clear that if we find a complete class for a certain decision problem, we do not need to look outside this class to find a decision rule, because we can just do well inside the class. Thus the minimal complete class, if

it exists, provides the maximum possible reduction of decision rules from which to search for a good rule.

As we noted earlier, admissibility is an optimal property, although in a very weak sense. So it is no surprise that if the minimal complete class exists, it is exactly the class of all admissible decision rules. That is, the collection of all admissible rules is the maximum possible reduction of decision rules.

Even though, proving admissibility is not easy, it is well known that there is a close relationship between admissibility and being Bayes. A common route to showing that a decision rule is admissible is to establish a prior distribution against which the decision rule is unique Bayes. Actually, in a certain decision problem, where Θ is finite and the risk set S is bounded from below and closed from below, any admissible rule is Bayes and the admissible Bayes rules form a minimal complete class. Thus in this situation, we can concentrate our attention on finding the admissible Bayes rules.

Recently a new mechanism, called *the stepwise Bayes approach*, was developed to find admissible rules, see Johnson [9], Hsuan [8], Meeden and Ghosh [14], and Brown [3]. When a prior distribution does not have support on the whole parameter space, the Bayes procedure often yields a collection of decision rules, rather than a unique decision rule. This collection is usually a mixture of admissible as well as inadmissible decision rules. The stepwise Bayes procedure applies the Bayes procedure in a stepwise manner to extract a subcollection at each step from a collection of decision rules which was obtained in the earlier step. They showed that this mechanism is successful in obtaining every admissible rule when the parameter space is finite and the loss function is strictly convex. However, the proof of admissibility is based on

the uniqueness at each step, which is guaranteed by strictly convex loss. Thus if we drop the assumption that the loss is strictly convex, then there is no way to prove that the resulting decision rule is admissible.

In Chapter 3, we will modify the stepwise Bayes procedure to present the minimal complete class without the assumption of a strictly convex loss function when the parameter space is finite and the risk set S is closed from below and bounded from below.

In Section 1.2 we review, without proofs, various complete classes and introduce the stepwise Bayes procedure. We will also give the necessary definitions.

1.2 Preliminary

1.2.1 Complete Class Theorem in Finite Problems

Decision theory consists of three basic elements: a nonempty set, Θ, here assumed to be finite, of possible states of nature, sometimes called the parameter space, a nonempty set, $\mathcal{D}$ of decisions, and a loss function $L(\theta, d)$, a real-valued function defined on $\Theta \times \mathcal{D}$. A statistical decision problem is a triple, $(\Theta, \mathcal{D}, L)$, coupled with an experiment involving a random variable X whose distribution P_θ depends on the unknown $\theta \in \Theta$. A statistician should choose decision rules from $\mathcal{D}^*$, the set of all possible decision rules.

Definition 1.1 A nonrandomized decision rule δ is a measurable function from $\mathcal{X}$ into $\mathcal{D}$. Using δ means that if $X = x$ is observed, then $\delta(x)$ is the decision which will be taken. A randomized decision rule $\delta^*(x, \cdot)$ is, for each $x \in \mathcal{X}$, a probability

distribution on $\mathcal{D}$. Using δ^* means that if $X = x$ is observed, then a decision $d \in \mathcal{D}$ will be chosen by the probability distribution $\delta^*(x, \cdot)$.

The risk function of a randomized decision rule δ^* is defined to be

$$r(\theta, \delta^*) = \int_{\mathcal{X}} \int_{\mathcal{D}} L(\theta, d) d\delta^*(x, d) dP_\theta(x)$$

Definition 1.2 A decision rule δ' is said to be *at least as good as* a rule δ, if $r(\theta, \delta') \leq r(\theta, \delta)$ for all $\theta \in \Theta$. A decision rule δ' is said to be *better* than δ, if $r(\theta, \delta') \leq r(\theta, \delta)$ for all $\theta \in \Theta$ and $r(\theta, \delta') < r(\theta, \delta)$ for at least one θ. A rule δ' is said to be *equivalent* to a rule δ, if $r(\theta, \delta') = r(\theta, \delta)$ for all $\theta \in \Theta$.

Definition 1.3 Suppose that Θ consists of k-points, $\Theta = \{\theta_1, \cdots, \theta_k\}$. The set $\mathcal{S}$, contained in k-dimensional Euclidean space $\mathcal{R}^k$, is called the risk set where

$$\mathcal{S} = \{(y_1, \cdots, y_k) : \text{ for some } \delta \in \mathcal{D}^*, y_i = r(\theta_i, \delta) \text{ for } i = 1, 2, \ldots, k\}$$

Definition 1.4 Let $\mathbf{u} = (u_1, \cdots, u_k)$ be in $\mathcal{R}^k$. We denote the set $Q_{\mathbf{u}}$ as

$$Q_{\mathbf{u}} = \left\{ (y_1, \cdots, y_k) \in \mathcal{R}^k, y_i \leq u_i \text{ for } i = 1, 2, \cdots, k \right\}$$

Definition 1.5 A point $\mathbf{u}$ is said to be a lower boundary point of a convex set $\mathcal{S} \subset \mathcal{R}^k$ if $Q_{\mathbf{u}} \cap \overline{\mathcal{S}} = \{\mathbf{u}\}$, where $\overline{\mathcal{S}}$ is the closure of $\mathcal{S}$. The set of lower boundary points of a convex set $\mathcal{S}$ is denoted by $\lambda(\mathcal{S})$.

Definition 1.6 A convex set $S \subset \mathcal{R}^k$ is said to be *closed from below* if $\lambda(S) \subset S$.

If we consider a nonrandomized decision rule $\delta(x)$ as a probability distribution which is degenerated at $d = \delta(x)$, a nonrandomized rule is a special case of a randomized decision rule. Thus we may drop the adjective "randomized" for the class of all possible decision rules. On the other hand, the randomized decision rules are normally of interest only from a theoretical point of view. The randomized rules will rarely be recommended for actual use.

Note that if a risk set S is a convex subset of $\mathcal{R}^k$, where the convexity of S is usually obtained by considering the randomized decision rules, and $\mathbf{u} \in S$, then $Q_{\mathbf{u}}$ is the set of risk points *as good as* $\mathbf{u}$, $Q_{\mathbf{u}} \sim \{\mathbf{u}\}$ is the set of risk points *better* than $\mathbf{u}$, and the elements of $\lambda(S)$ lead to admissible decision rules, that is, if the risk set is bounded from below and closed from below, a risk point $\{\mathbf{u}\}$ is admissible, if and only if $\mathbf{u} \in \lambda(S)$. Thus there is no reason to consider decision rules other than those corresponding to points in $\lambda(S)$. Hence the question is how we can find the decision rules whose risk points are in $\lambda(S)$. Theorem 1.1 gives a partial answer to this question.

Theorem 1.1 *If a risk set S is bounded from below and closed from below, then $\lambda(S)$ is a subset of risk points arising from Bayes decision rules.*

Note that not all Bayes risk points need be in $\lambda(S)$. However, Theorem 1.1 provides a useful tool for calculating $\lambda(S)$. That is, if a decision rule is admissible, then the rule is Bayes against some prior distribution. Thus we have the following corollary.

Corollary 1.1 (The Complete Class Theorem) *If, for a given decision problem* $(\Theta, \mathcal{D}, L)$ *with finite* Θ, *the risk set* $\mathcal{S}$ *is bounded from below and closed from below, then the class of all Bayes rules is complete and the admissible Bayes rules form a minimal complete class.*

Throughout this section, we assumed that the risk set was bounded from below and closed from below. Therefore it is of interest to observe the conditions which ensure the assumptions. The following Lemma gives a condition that guarantees $\mathcal{S}$ is indeed bounded from below.

Lemma 1.1 *A risk set* $\mathcal{S} \subset \mathcal{R}^k$ *is bounded from below if*

$$L(\theta_i, d) > -K > -\infty \text{ for all } d \in \mathcal{D} \text{ and } i = 1, 2, \ldots, k$$

The condition in Lemma 1.1 is always satisfied as long as we use nonnegative loss function. Hence we can always assume the risk set is bounded from below. The crucial assumption about the risk set is that it is closed. Note that if $\mathcal{S}$ is closed, then $\lambda(\mathcal{S})$ is a subset of $\mathcal{S}$. Thus one can verify that $\mathcal{S}$ is closed from below by showing that it is closed. To show that $\mathcal{S}$ is closed, it is useful to consider the set $\mathcal{S}_O = \{ (y_1, \cdots, y_k) : \text{ for some } d \in \mathcal{D}, y_i = L(\theta_i, d) \text{ for } i = 1, 2 \ldots, k \}$, set of loss points.

Lemma 1.2 *If* $\mathcal{S}_O \subset \mathcal{R}^k$ *is closed and bounded, then the risk set* $\mathcal{S}$ *is also closed and bounded*

There are many conditions under which the set of loss points is closed and bounded. For example, if $\mathcal{D}$ is finite, or $\mathcal{D}$ is compact subset of $\mathcal{R}$ and $L(\theta, d)$ is continuous in d for each $\theta \in \Theta$, then clearly $\mathcal{S}_0$ is closed and bounded, and hence $\mathcal{S}$ is closed and bounded. Indeed only in a few cases statistical decision problem fail to have a closed and bounded risk set. Note that the closed and bounded set has a nice property, namely, existence of Bayes rules.

Lemma 1.3 *Suppose that $\Theta = \{\theta_1, \cdots, \theta_k\}$ and the risk set S is closed from below and bounded. Then, for every prior distribution π on Θ, a Bayes rule with respect to π exists.*

In view of existence of Bayes rules, a lower semicontinuous loss function, that is, for each $\theta \in \Theta$, $\{d \in \mathcal{D} : L(\theta, d) \leq c\}$ is a closed subset of $\mathcal{D}$ for all real number c, is important in statistical decision problem. If Θ is finite, $\mathcal{D}$ is a compact subset of $\mathcal{R}$, and L is a lower semicontinuous function, then the risk set is closed and bounded from below, and for every prior distribution π, there exist a Bayes rule against π.

1.2.2 The Stepwise Bayes Procedure

Let X be a random variable which takes on values in some finite sample space $\mathcal{X}$. The σ-algebra of measurable sets is the power set of $\mathcal{X}$, that is, the collection of all subsets of $\mathcal{X}$. Let $\{f_\theta : \theta \in \Theta\}$ be a family of possible probability functions for X where $\Theta = \{\theta_1, \cdots, \theta_k\}$. We assume that for each $x \in \mathcal{X}$, there exists at least one $\theta_i \in \Theta$ such that $f_{\theta_i}(x) > 0$. Consider the problem of estimating some real valued function of θ, say $\gamma(\theta)$, with some nonnegative strictly convex loss function

$L(\theta, d)$. For this statistical decision problem, we take the decision space $\mathcal{D}$ to be some bounded interval of real numbers which contains the range space of γ.

The usual Bayes approach uses one prior distribution to obtain a decision rule. If the support of the prior distribution is the whole parameter space, we call it *regular* prior, then the Bayes rule is uniquely defined and admissible. However, if this is not the case, the Bayes approach can result in a collection of decision rules rather than a unique rule. This collection of Bayes rules usually consists of admissible as well as inadmissible decision rules.

The stepwise Bayes procedure extracts a subcollection from the collection of Bayes rules to find an admissible decision rule using the following simple observation. Let $\pi = \left\{\pi_1, \cdots, \pi_k\right\}$ be a prior distribution on Θ, and $g(x : \pi) = \sum_{i=1}^{k} f_{\theta_i}(x)\pi_i$ be the marginal distribution of X under π. Let $\Lambda = \left\{x \in \mathcal{X} : g(x, \pi) > 0\right\}, \Theta(\pi) = \left\{\theta_i \in \Theta : \pi_i > 0\right\}$, and $\Theta(\pi, \Lambda) = \left\{\theta \in \Theta - \Theta(\pi) : f_{\theta_i}(x) = 0 \text{ for all } x \in \mathcal{X} - \Lambda\right\}$. We assume that $\mathcal{X} - \Lambda$ is nonempty. Note that $\Theta - \Theta(\pi)$ is nonempty as well by the previous assumption. Then the risk of a decision rule δ is

$$r(\theta, \delta) = \sum_{x \in \Lambda} L(\theta, \delta(x))f_\theta(x) + \sum_{x \in \mathcal{X} - \Lambda} L(\theta, \delta(x))f_\theta(x).$$

Consider now the restricted problem where $x \in \mathcal{X} - \Lambda$ and $\theta \in \Theta - \Theta(\pi) - \Theta(\pi, \Lambda)$. For this restricted problem, the family of possible distributions is $\{f_\theta^* : \theta \in \Theta - \Theta(\pi) - \Theta(\pi, \Lambda)\}$ where for $x \in \mathcal{X} - \Lambda$,

$$f_\theta^*(x) = f_\theta(x)/c(\theta)$$

and

$$c(\theta) = \sum_{x \in \mathcal{X} - \Lambda} f_\theta(x) > 0.$$

Suppose that δ is a Bayes rule against π. We denote $\bar{\delta}$ the restriction of δ to $\mathcal{X} - \Lambda$. Then δ is admissible *if and only if* $\bar{\delta}$ is admissible for the restricted problem. Hence, if δ is admissible, then $\bar{\delta}$ is Bayes for the restricted problem with respect to a prior distribution which concentrates its mass on the set $\Theta - \Theta(\pi) - \Theta(\pi, \Lambda)$. Thus, if we again apply Bayes procedure for the restricted problem, we can extract a subcollection from the collection of the Bayes rules. The stepwise Bayes procedure utilizes this observation to obtain an admissible decision rule by applying the Bayes procedure in this manner.

Theorem 1.2 *If δ is admissible, then there exists a nonempty set of prior distributions $\pi^1 = \left\{ \pi_1^1, \cdots \pi_k^1 \right\}, \cdots, \pi^m = \left\{ \pi_1^m, \cdots \pi_k^m \right\}$ such that*

(i) $\Theta(\pi^i) \bigcap \Theta(\pi^j) = \emptyset$ *for all $i \neq j$.*

(ii) *If $\Lambda^1 = \left\{ x \in \mathcal{X} : g(x, \pi^1) > 0 \right\}$, and for $i = 2, \ldots, m, \Lambda^i = \left\{ x : x \notin \bigcup_{j=1}^{i-1} \Lambda^j \right.$ and $\left. g(x, \pi^i) > 0 \right\}$, then each Λ^i is nonempty and $\bigcup_{i=1}^m \Lambda^i = \mathcal{X}$.*

(iii) *For $x \in \Lambda^i, \delta(x)$ is the unique value of d which minimizes*

$$\sum_{\theta \in \Theta(\pi^i)} L(\theta, d) f_\theta(x) \pi^i(\theta) / g(x : \pi^i).$$

Conversely if there exists a set of prior distributions $\pi^1, \cdots, \pi^m$ which satisfies (i) *and* (ii), *then the decision rule δ given in* (iii) *is admissible.*

Essentially Theorem 1.2 says that the class of stepwise Bayes rules forms a minimal complete class under the conditions we stated above. It also should be noted

that ordinary Bayes rules can be treated as a special class of stepwise Bayes rules where $m = 1$, and a stepwise Bayes rule with respect to $(\pi^1, \cdots, \pi^m)$ is necessarily a Bayes rule with respect to π^1, but it need not be Bayes with respect to $\pi^i, i = 2, \ldots, m$.

Example 1.1 This example is due to Hsuan [8]. Let X be hypergeometrically distributed with population size $N = 3$, subpopulation size $M = 3\theta$ (where θ stands for the proportion of defectives), and sample size $n = 2$. Thus the sample space $\mathcal{X} = \{0, 1, 2\}$, and the parameter space $\Theta = \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}$. Assume that $\mathcal{D} = [0, 1]$, and our loss is the squared error loss. We pick a prior distribution $\pi^1 = \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right)$. Then $\Lambda^1 = \{0, 1\}$ and the posterior distribution is $\pi^1(\theta \,|\, x = 0) = \left(\frac{3}{4}, \frac{1}{4}, 0, 0\right)$ and $\pi^1(\theta \,|\, x = 1) = (0, 1, 0, 0)$. Hence any decision rule satisfying $\delta(x = 0) = \frac{1}{12}$, and $\delta(x = 1) = \frac{1}{3}$ is a Bayes rule against π^1. Now we put $\pi^2 = (0, 0, 1, 0)$, then $\Lambda^2 = \{2\}$ and $\pi^2(\theta \,|\, x = 2) = (0, 0, 1, 0)$. A Bayes rule against π^2 satisfies $\delta(x = 2) = \frac{2}{3}$. Therefore $\delta(x = 0) = \frac{1}{12}, \delta(x = 1) = \frac{1}{3}$ and $\delta(x = 2) = \frac{2}{3}$ is the stepwise Bayes rule against $\left(\pi^1, \pi^2\right)$ and it is admissible.

Throughout this section, we have assumed that the parameter space is finite and the loss is strictly convex. A general theory of the stepwise Bayes approach is not developed yet, but the assumptions can be relaxed in some cases. For example, Johnson [9] found the minimal complete class in the problem of estimating a binomial parameter with the squared error loss, when the parameter space is compact not finite, using the stepwise Bayes argument. We can also replace the assumption of strictly convex loss by the uniqueness of the stepwise Bayes rule. For example, we have the

following.

Theorem 1.3 *A unique stepwise Bayes rule is admissible.*

The following simple example shows that the uniqueness is essential in admissibility proof.

Example 1.2 Suppose $\mathcal{X} = \{x_1, x_2\}$ and $\Theta = \{0,1\}$. Let $f_{\theta=0}(x) = 1$ or 0 as $x = x_1$ or $x = x_2$, and $f_{\theta=1}(x) = \frac{1}{2}$ for all $x \in \mathcal{X}$. Consider the problem of estimating θ when $\mathcal{D} = [0,1]$ and the loss is

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| \leq \frac{1}{2} \\ 1 & \text{otherwise.} \end{cases}$$

Then the family of prior distributions $\pi^1 = (1,0)$ and $\pi^2 = (0,1)$ yields the family of stepwise Bayes rules $\left\{ \delta : 0 \leq \delta(x_1) \leq \frac{1}{2} \text{ and } \frac{1}{2} \leq \delta(x_2) \leq 1 \right\}$. In particular, $\delta^*(x) = 0$ or 1 as $x = x_1$ or $x = x_2$ is a stepwise Bayes rule against (π^1, π^2) and $r(\theta = 0, \delta^*) = 0, r(\theta = 1, \delta^*) = \frac{1}{2}$. Let $\delta_o = \frac{1}{2}$ for $x = x_1$ and $\delta_o = 1$ for $x = x_2$, then $r(\theta, \delta_o) = 0$ for all θ. Thus δ^* is dominated by δ_o.

2. ADMISSIBLE NONPARAMETRIC TESTS

2.1 Introduction

We recall the definition of admissibility of a statistical test. Let $\mathcal{X}$ be a sample space, $\mathcal{B}$ a σ-algebra of subsets of $\mathcal{X}$, and Θ a parameter space. Let P_θ be a probability measure on $\mathcal{B}$. We assume a random variable X is distributed in $\mathcal{X}$ according to P_θ, with θ an unknown element of Θ, and we want to test the hypothesis $H : \theta \in \Theta_o$ against the alternative $K : \theta \in \Theta - \Theta_o$ where Θ_o is a nonempty proper subset of Θ. A test ϕ is a $\mathcal{B}$-measurable function on $\mathcal{X}$ to the closed interval $[0,1]$, with the interpretation that if we observe X, we reject H with probability $\phi(X)$. The test ϕ_o is said to be admissible if there does not exist a test ϕ such that

$$r(\theta, \phi_o) \geq r(\theta, \phi) \quad \text{for all } \theta \in \Theta \tag{2.1}$$

with strictly inequality at least one $\theta \in \Theta$, where $r(\theta, \phi) = \int L(\theta, \phi) dP_\theta$ and L is our loss. If the loss is "0-1", then $r(\theta, \phi) = \int \phi dP_\theta$ for all $\theta \in \Theta_o$, and $1 - r(\theta, \phi) = \int \phi dP_\theta$ for all $\theta \in \Theta - \Theta_o$. Thus (2.1) becomes

$$\int \phi dP_\theta \leq \int \phi_o dP_\theta \quad \text{for all } \theta \in \Theta_o$$

$$\int \phi dP_\theta \geq \int \phi_o dP_\theta \quad \text{for all } \theta \in \Theta - \Theta_o$$

Now we consider a two-sample problem. Let $X_1, \cdots, X_m$, and $Y_1, \cdots, Y_n$ be random samples from unknown distributions F and G, respectively, which are as-

sumed to belong to Θ, some family of distributions. We wish to test the hypothesis $H : F(x) = G(x)$ for all x against the alternative $K : F(x) \neq G(x)$ for some x. If we have some information about F and G, for example, the two distributions are normal with a common variance, we may use the Uniformly Most Powerful Unbiased test, which is the usual *Student's t*-test. However, if the distributions are not normal, then this test may be a poor choice.

The nonparametric two-sample problem arises when we do not make any assumptions concerning the forms of the underlying distributions. That is, we assume that Θ is set of all possible distribution functions on the real line.

Meeden, Ghosh, and Vardeman [15] recently demonstrated that there is a close relationship between admissibility in nonparametric problems and admissibility in finite population sampling. They showed that both problems are related to admissibility for multinomial problems and to prove admissibility it was enough to consider the subfamily of Θ consisting of all discrete distributions with at most a finite number of jumps. This suggests that to find an admissible test for the two-sample nonparametric problem stated above we should first consider the problem of testing $H : p = q$ against the alternative $K : p \neq q$, where $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ are random samples from *multinomial*$(1, p)$ and *multinomial*$(1, q)$, respectively, and $p = (p_1, \cdots, p_k), q = (q_1, \cdots, q_k)$.

This multinomial testing problem was studied extensively by Matthes and Truax [13]. They characterized a complete class for testing the hypothesis that the parameter in the multivariate exponential distribution lies in a linear subspace of the natural parameter space, which is applicable to this multinomial testing problem. In Section 2.2, we review, without proof, these results.

For the two-sample nonparametric problem stated above, $(X_{(1)}, \cdots, X_{(m)})$, and $(Y_{(1)}, \cdots, Y_{(n)})$, the two vectors of order statistics for the two samples are sufficient and complete and we can restrict our attention to tests based on these order statistics. Note that if we assume that F and G are continuous, then the problem is invariant under the group $\mathcal{G}$ of all transformations $g_\varphi \left(x_{(1)}, \cdots, x_{(m)}, y_{(1)}, \cdots, y_{(n)} \right) = \left(\varphi(x_{(1)}), \cdots, \varphi(x_{(m)}), \varphi(y_{(1)}), \cdots, \varphi(y_{(n)}) \right)$ such that φ is continuous and strictly increasing. This follows from the fact that these transformations preserve both the continuity of the distributions and the property of two variables being either identically distributed or not. The maximal invariant under $\mathcal{G}$ is the set of ranks $(R_1, \cdots, R_m, R_{m+1}, \cdots, R_{m+n})$, where $R_1 < R_2 < \cdots < R_m$ are ranks of the order statistics of X_i's in the total sample of $N = m + n$ observations and $R_{m+1} < R_{m+2} < \cdots < R_{m+n}$ are the ranks of the order statistics of the Y_j's. That is $T(z_1) = T(z_2)$ implies $z_1 = g(z_2)$ for some $g \in \mathcal{G}$, where $z = (x_{(1)}, \cdots, x_{(m)}, y_{(1)}, \cdots, y_{(n)})$ and $T(z) = (r_1, \cdots, r_m, s_1, \cdots, s_n)$. So the invariance principle leads us to consider rank tests for this problem.

Let us consider a linear rank statistic $L = \sum_{i=1}^{N} c_i a(R_i)$, where $a(1), \cdots, a(N)$ and $c_1, \cdots, c_N$ are two sets of N constants such that the numbers within each set are not all the same. The constants $a(1), \cdots, a(N)$ are called the scores and $c_1, \cdots, c_N$ are termed the regression constants. We can generate many statistics by choosing $a(i)$'s and c_i's in a suitable manner. For example, if

$$c_i = \begin{cases} 0 & \text{for } i = 1, 2, \ldots, m \\ 1 & \text{for } i = m + 1, \ldots, m + n \end{cases} \tag{2.2}$$

and

$$a(i) = i \quad \text{for } i = 1, 2, \ldots, N, \tag{2.3}$$

then L becomes $\sum_{j=1}^{n} R_{j+m}$, which is the well-known Mann–Whitney–Wilcoxon test or rank sum test.

The linear rank tests which are most commonly applied to this two-sample non-parametric problem are the rank sum test and the Fisher-Yates test. It is well known that these tests have certain optimum properties in the class of all rank test, see e.g., Ferguson [6] and Lehmann [12].

In Section 2.3, we will discuss the admissibility of these tests as well as other linear rank statistics in the class of all test.

In Section 2.4, we will prove the admissibility of the Kruskal-Wallis test for the one-way layout problem.

2.2 Testing for the Multivariate Exponential Distribution

2.2.1 Complete Class Theorem

Let $X = (X_1, \cdots, X_k)$ be a k-dimensional random vector from the distribution

$$P_\theta(A) = \int_A e^{\theta x} \lambda(dx)$$

where λ is a finite measure on $\mathcal{R}^k$, $\theta = (\theta_1, \cdots, \theta_k)$, and $\theta x = \sum_{i=1}^{k} \theta_i x_i$. Let Θ denote the natural parameter space and Θ_o be a r-dimensional linear subspace($r < k$) of $\mathcal{R}^k$. In this section, we will consider the test of the null hypothesis $H : \theta \in \Theta \cap \Theta_o$. If we write the sample space as $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X} = \mathcal{R}^r, \mathcal{Y} = \mathcal{R}^{k-r}$, a sample point as (x,y), and the parameter point as the pair (θ,ω) where θ is an r-vector , and ω is a $(k-r)$-vector, the hypothesis can be put into the canonical form $H : \omega = 0$, by an orthogonal linear transformation. Hence we are interested in testing $H : \omega = 0$

against $K : \omega \neq 0$ where θ is considered to be a nuisance parameter. The goal is to find a complete class of tests for this problem.

The case of no nuisance parameters, that is $r = 0$, was considered by Birnbaum [2]. Although his proof involved the restriction that the probability distribution is absolutely continuous, he showed that the class of tests which accept the null hypothesis when Y is in some convex set in $\mathcal{Y}$ is a complete class. That is, Φ_C is a complete class where Φ_C is the collection of all tests such that

$$\phi(y) = \begin{cases} 0 & y \in intC' \\ 1 & y \in C' \end{cases},$$

and ϕ is randomized on the boundary of C', where C' is a convex set in $\mathcal{Y}$.

We now give a brief summary of the proof of this result. Let B denote the set of all Bayes tests; that is $\phi \in B$ if and only if there exists a prior distribution ξ such that ϕ is a Bayes solution with respect to ξ. Birnbaum then proved the following two results.

Theorem 2.1 B *is a subset of* Φ_C.

Lemma 2.1 *Let* $\phi_i(y) = 0$ *on a convex set* A_i, *and* $\phi_i(y) = 1$ *elsewhere. Let* $\phi_i \longrightarrow \phi_0$ *in the regular sense, that is* $\lim_{i \to \infty} \int_A \phi_i(y)dy = \int_A \phi_0(y)dy$ *for any bounded subset A of the sample space $\mathcal{Y}$. Then $\lim_{i \to \infty} \phi_i = \phi_0$ except on a set of Lebesgue measure zero. Furthermore $\phi_0 \in \Phi_C$, and except on a set of Lebesgue measure zero,*

$$\phi_0(y) = \begin{cases} 0 & on \ A_0 = \lim_{i \to \infty} \cap_{j \geq i} A_j \\ 1 & elsewhere \end{cases}.$$

Because of Wald [21], Theorem 2.1, together with Lemma 2.1 shows that the class Φ_C is essentially complete. From this result and the completeness property of the family of exponential distributions, we see that the class of tests which have convex acceptance region is a complete class.

The case $r > 0$ was considered by Matthes and Truax [13]. The presence of nuisance parameters complicates the problem, but we can use the conditional measure on Y determined by λ for a fixed $x \in \mathcal{X}$ to eliminate the nuisance parameters. Note that the resulting conditional measure also belongs to the exponential family. Thus applying Birnbaum's theorem, a collection of tests $\phi_x(y)$, whose acceptance region is convex in $\mathcal{Y}$ for each x-section of $\mathcal{Y}$, may form a complete class. However, the crux of the matter is the fact that considered as a function on $\mathcal{X} \times \mathcal{Y}, \phi$ may not be jointly measurable.

Matthes and Truax utilized these ideas to find a complete class for testing H : $\omega = 0$ against $K; \omega \neq 0$ and gave some additional results on the admissibility of certain tests within this class.

Let $\phi \in [0,1]$ be a measurable function on $\mathcal{X} \times \mathcal{Y}$. The test ϕ is said to have convex acceptance sections if there exists a measurable set $C \subset \mathcal{X} \times \mathcal{Y}$, each of whose x-sections are closed and convex in $\mathcal{Y}$, and

$$\phi(x,y) = \begin{cases} 0 & y \in intC(x) \\ 1 & y \in C'(x) \end{cases},$$

On the boundary of $C(x), \phi$ may be randomized. The family of all tests with convex acceptance section will be denoted by $\Phi_\mathcal{D}$. The marginal distribution of X and conditional distribution of Y given $X = x$ determined by λ will be denoted by ν and

$\mu(dy; x)$, respectively. Thus, the marginal density of X with respect to ν is given by

$$p_{\theta,\omega}(x) = c(\theta,\omega)e^{\theta x} \int_{\mathcal{Y}} e^{\omega y}\mu(dy; x), \qquad (2.4)$$

and the conditional density of Y given $X = x$, with respect to $\mu(dy; x)$ is

$$p_{\omega}(y|x) = e^{\omega y} \bigg/ \int_{\mathcal{Y}} e^{\omega y}\mu(dy; x). \qquad (2.5)$$

Theorem 2.2 *Let ψ be any test of $H : \omega = 0$. Then there exists a test $\phi \in \Phi_{\mathcal{D}}$ with the property that for each $\omega \in \Theta$,*

$$\int \phi(x,y)\ e^{\omega y}\mu(dy; x) \geq \int \psi(x,y)\ e^{\omega y}\mu(dy; x) \quad [\nu] \qquad (2.6)$$

with equality in case $\omega = 0$.

Note that if we multiple both sides of (2.6) by $[\int e^{\omega y}\mu(dy; x)]^{-1}$, then the integrals in (2.6) become the conditional powers of the two test, as the conditional density (2.5) shows. Thus, we can always find a test $\phi \in \Phi_{\mathcal{D}}$ which is at least as good as ψ. Eventually we have a complete class for the problem by the completeness property of the exponential family.

2.2.2 Admissibility

In general the verification of admissibility or inadmissibility of a test is very difficult. In particular the question of admissibility of tests in $\Phi_{\mathcal{D}}$ is quite complicated

as well, but there are a few cases in which the admissibility question can be satis-
factorily answered. In what follows we briefly summarize some of the discussion of
Matthes and Truax.

One such case is when ω is real. Suppose that there is a test ψ which is at least
as good as a test $\phi \in \Phi_{\mathcal{D}}$. Because of the Theorem 2.2, we may assume, without loss
of generality, $\psi \in \Phi_{\mathcal{D}}$. Then by the continuity property of the power function,

$$\int (\psi - \phi) \, e^{\theta x + \omega y} d\lambda \geq 0 \quad \text{for all } (\theta, \omega) \in \Theta \tag{2.7}$$

with equality whenever $\omega = 0$. Applying the completeness argument to (2.7) yields

$$\int_{\mathcal{Y}} (\psi - \phi) \mu(dy; x) = 0 \quad [\nu] \tag{2.8}$$

Then (2.8) together with the validity of the inequality (2.7) for all $\omega = 0$ in the
neighborhood of the origin shows that for each $\theta, \int_{\mathcal{Y}} y(\psi - \phi) \, e^{\theta x} d\lambda = 0$. Again,
completeness yields

$$\int_{\mathcal{Y}} y(\psi(x, y) - \phi(x, y)) \mu(dy; x) = 0 \quad [\nu]. \tag{2.9}$$

Since the conditional tests accepts the null hypothesis on the intervals of the real line
and the interval is determined by the size and "center of gravity", it must be the case
that $\psi \equiv \phi$. Thus we have the following Lemma.

Lemma 2.2 *Suppose that ω is real $(r = k - 1)$, then $\Phi_{\mathcal{D}}$ is the minimal complete
class for the problem.*

Another case is when λ has finite support. In this case, ν also has finite support
and it can be shown that, if every x-section of ϕ is an admissible test for the simple

hypothesis $\omega = 0$, then ϕ is admissible for the composite hypothesis as well. So to prove admissibility of a test $\phi \in \Phi_{\mathcal{D}}$, it suffices to show that each x-section of ϕ is admissible for testing the simple hypothesis when the dominating measure μ has finite measure, which is closely related to proving admissibility of tests whose acceptance regions are convex. Stein [20] gave the sufficient condition for admissibility of tests with closed convex acceptance regions. If we apply Stein's theorem to this case, when the parameter space for ω is the full $k - r$ Euclidean space, a test ϕ_O whose acceptance region is closed and convex for each its x-section is admissible. But Stein's theorem can not be applied to a convex acceptance region with randomization on the boundary, if the conditional measure does not assign measure zero to the boundary of the convex sets. However Matthes and Truax gave a more valuable result for this case.

Lemma 2.3 *Let Y be a random variable having density*

$$p_{\omega}(y) = c(\omega)\, e^{\omega y}$$

with respect to a dominating measure μ in $\mathcal{Y}$. Suppose that μ has finite measure. Then, for testing the null hypothesis $\omega = 0$, a test ϕ is admissible if and only if the set $C = \{y \in \mathcal{Y} : \phi(y) < 1\}$ is convex, and for every y which is not an extreme point of $C, \phi(y) = 0$.

Returning to the case of a composite hypothesis, we suppose λ has finite support. Probably the most important case of this is the multinomial case.

Theorem 2.3 *Let ϕ be a test and let $C = \{(x,y) : \phi(x,y) < 1\}$. Then ϕ is admissible if and only if, every x-section of C is convex $[\mu(\cdot;x)]$ and $\phi(\cdot,x) = 0$ at all nonextreme point of the section C_x.*

Lemma 2.3, together with the fact that ν has finite support yields, Theorem 2.3 as an obvious result. Hence, when λ has finite support, $\Phi_{\mathcal{D}}$ is the minimal complete class for testing $H : \omega = 0$ against $K : \omega \neq 0$ in the multivariate exponential family.

In the next section we will use this result to prove the admissibility of some linear rank tests for the nonparametric problem discussed in Section 2.1.

2.3 Admissible Nonparametric Tests for the Two-sample Problem

Let random variables X and Y take on the k distinct values, $x_1, \cdots, x_k$, with probability $Pr(X = x_i) = p_i$ and $Pr(Y = x_i) = q_i, i = 1, 2, \cdots, k$, where $0 \leq p_i, q_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1, \sum_{i=1}^{k} q_i = 1$. If we select m and n samples from each population, $T_1, \cdots, T_{k-1}$ and $S_1, \cdots, S_{k-1}$, where T_i is the number of $X = x_i$ in m samples and S_i is the number of $Y = x_i$ in n samples, are jointly complete and sufficient for p and q, and have joint probability distribution,

$$f_{p,q}(t,s) = \begin{pmatrix} m \\ t_1 \cdots t_{k-1} \end{pmatrix} \begin{pmatrix} n \\ s_1 \cdots s_{k-1} \end{pmatrix} p_1^{t_1} \cdots p_{k-1}^{t_{k-1}} p_k^{m-t_1-\cdots-t_{k-1}}$$
$$\times \; q_1^{s_1} \cdots q_{k-1}^{s_{k-1}} q_k^{n-s_1-\cdots-s_{k-1}}$$

$$(2.10)$$

For testing $H : p = q$ against $K : p \neq q$, we will consider the linear rank statistics $L = \sum_{i=1}^{m+n} c_i a(R_i)$.

Note that (2.10) can be written as

$$f_{\omega,\theta}(t,u) = c(\omega,\theta;t,u)\ e^{t\omega+u\theta} \qquad (2.11)$$

where $u_j = t_j + s_j, t\omega = \sum_{j=1}^{k-1} t_j\omega_j, u\theta = \sum_{j=1}^{k-1} u_j\theta_j, \omega_j = \log\frac{p_j/q_j}{p_k/q_k}$, and $\theta_j = \log\frac{q_j}{q_k}$ and our hypothesis can then be stated using the ω_j parameters or equivalently using the odds ratios $\rho_j = \frac{p_j/q_j}{p_k/q_k}$. Since $p = q$ if and only if $\omega = 0$, the original hypothesis therefore is reduced to $H : \omega = 0$ against $K : \omega \neq 0$.

Lemma 2.4 *For testing $H : \omega = 0$ against $K : \omega \neq 0$, the linear rank statistic $L = \sum_{i=1}^{m+n} c_i a(R_i)$ is admissible if the c_i's are constants over each group of samples. i.e., $c_1 = c_2 = \cdots = c_m$ and $c_{m+1} = c_{m+2} = \cdots = c_{m+n}$.*

Proof: Without loss of generality, we assume that $x_1 = 0, x_1 = 1, \cdots, x_k = k - 1$. Then there are u_1 0's and u_2 1's and so on, and each $0, 1, \ldots, k - 1$ has rank $\frac{u_1+1}{2}, u_1 + \frac{u_2+1}{2}, u_1 + u_2 + \frac{u_3+1}{2}, \cdots, u_1 + \cdots + u_{k-1} + \frac{u_k+1}{2}$, respectively. Thus L becomes

$$
\begin{aligned}
L = {} & a\left(\frac{u_1+1}{2}\right)\left[\sum_{i=1}^{t_1} c_i + \sum_{i=m+1}^{m+s_1} c_i\right] \\
& + a\left(u_1 + \frac{u_2+1}{2}\right)\left[\sum_{i=t_1+1}^{t_1+t_2} c_i + \sum_{i=m+s_1+1}^{m+s_1+s_2} c_i\right] \\
& \qquad\qquad \vdots \\
& + a\left(u_1 + \cdots + u_{k-1} + \frac{u_k+1}{2}\right) \\
& \qquad \times \left[\sum_{i=t_1+\cdots+t_{k-1}+1}^{m} c_i + \sum_{m+s_1+\cdots+s_{k-1}+1}^{N} c_i\right]
\end{aligned}
$$

$$= a\left(\frac{u_1 + 1}{2}\right)[t_1 c_1 + (u_1 - t_1)c_{m+1})]$$

$$+ a\left(u_1 + \frac{u_2 + 1}{2}\right)[t_2 c_1 + (u_2 - t_2)c_{m+1}]$$

$$\vdots$$

$$+ a\left(u_1 + \cdots + u_{k-1} + \frac{u_k + 1}{2}\right)$$

$$\times \left[\left(m - \sum_{i=1}^{k-1} t_i\right) c_1 + \left(N - \sum_{i=1}^{k-1} u_i - \sum_{i=1}^{k-1} t_i\right) c_{m+1}\right] \quad (2.12)$$

Note that for each fixed $u_1, \cdots, u_{k-1}$, the a's are constant and L is a linear function of $t_1, \cdots, t_{k-1}$. Because we reject the null hypothesis if L is too large or too small, the acceptance region of the test is convex in $t_1, \cdots, t_{k-1}$, for each set of fixed u_i's. Therefore the linear rank test has convex acceptance section and is admissible. □

Lemma 2.5 *Let* $\phi_0, \phi \in \Phi_{\mathcal{D}}$ *and* $r(p, q, \phi_0) = r(p, q, \phi)$ *for all* p, q. *Then*

$$\phi_0 = \phi \quad [\lambda]$$

If $\phi_0, \phi \in \Phi_{\mathcal{D}}$, then both ϕ_0 and ϕ are functions of complete sufficient statistics T and U. Hence by the completeness property of underlying distributions, Lemma 2.5 is trivial.

Theorem 2.4 *Let* ϕ_0 *be an admissible test for the multinomial problem. Then* ϕ_0 *is admissible for the nonparametric problem as well.*

Proof :   We will assume ϕ_0 is not admissible for the nonparametric problem and get a contradiction.

If ϕ_0 is not admissible for the nonparametric problem. Then there exists a test ϕ such that

$$r(F, G, \phi) \leq r(F, G, \phi_0) \quad \text{for every } F, G \in \Theta \tag{2.13}$$

with strictly inequality for some $F, G \in \Theta$. Hence there exist $x_1, \cdots, x_m, y_1, \cdots, y_n$ such that

$$\phi_0(x_1, \cdots, x_m, y_1, \cdots, y_n) \neq \phi(x_1, \cdots, x_m, y_1, \cdots, y_n) \tag{2.14}$$

Let $\alpha_1, \cdots, \alpha_k$ be the k-distinct values which appear in the set $\{x_1, \cdots, x_m, y_1, \cdots, y_n\}$ and let $\Theta(\alpha_1, \cdots, \alpha_k)$ denote all distribution functions which concentrate all their mass on $\alpha_1, \cdots, \alpha_k$. We now consider the testing problem $H : F = G$ against $K : F \neq G$ where $F, G \in \Theta(\alpha_1, \cdots, \alpha_k)$. In this case $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ are the outcomes for the random samples from *multinomial* $(1, p_1, \cdots, p_k)$ and *multinomial* $(1, q_1, \cdots, q_k)$ where $p_i = Pr(X = \alpha_i)$ and $q_i = Pr(Y = \alpha_i)$ for $i = 1, 2, \ldots, k$. Note that $\Theta(\alpha_1, \cdots, \alpha_k)$ is equivalent to the $(k - 1)$-dimensional simplex

$$\Gamma = \left\{ p = (p_1, \cdots, p_k); 0 \leq p_i \leq 1 \quad \text{for } i = 1, \ldots, k \quad \text{and} \quad \sum_{i=1}^{k} p_i = 1 \right\}.$$

and each $p \in \Gamma$ determines a unique F, say F_p.

Since ϕ_0 is admissible for the multinomial problem, it must be the case that

$$r(F_p, G_q, \phi_0) = r(F_p, G_q, \phi) \quad \text{for all } p, q \in \Gamma. \tag{2.15}$$

and hence ϕ is admissible for the multinomial problem as well. Therefore both ϕ_0 and ϕ belong to $\Phi_{\mathcal{D}}$ and by Lemma 2.5, $\phi_0 = \phi[\lambda]$ which contradicts (2.14).   □

Note that the admissibility of the linear rank statistic depends on the regression constants $c_1, \cdots, c_N$. The choice of the regression constants for a linear rank statistic is usually dictated by the nature of the particular testing problem and hence are not controllable. On the other hand, we are at liberty to choose the scores so as to achieve desirable power properties. For example, (2.2), the two-sample regression constants, may be the unique reasonable choice of regression constants, if we exclude the similar selection of c_i's such as $c_i = -1$ for $i = 1, \ldots, m$ and $c_i = 1$ for $i = m+1, \ldots, m+n$. Thus the conclusion of Lemma 2.4 and Theorem 2.4 is that any linear rank statistic is admissible as long as their choice of regression constants is reasonable. Because we are considering a large nonparametric family of distributions, this result is not surprising.

Example 2.1 We can prove the admissibility of many well-known nonparametric tests using Lemma 2.4 and Theorem 2.4. The scores $a(1), \cdots, a(N)$ defined in (2.3) are called the Wilcoxon scores and together with (2.2) yield the Mann-Whitney-Wilcoxon test. A different choice of scores produces another admissible nonparametric test. For example, if $a(i) = E\left(\mathcal{N}_{(i)}^N\right)$ for $i = 1, 2, \ldots, N$, where $\mathcal{N}_{(i)}^N$ is the i-th order statistic in N-samples from $\mathcal{N}(0, 1)$, then the linear rank statistic becomes the Fisher-Yates test. The two-sample median test and the Savage test are the other examples of admissible nonparametric tests which can be generated by appropriate scores and two-sample regression constants.

2.4 Admissible Nonparametric Test for the One-way Layout Problem

The technique used in proving the admissibility of the linear rank statistics can be extended to multi-sample problems. Suppose $X_{11}, \cdots, X_{1n_1}, X_{21}, \cdots, X_{v1}, \cdots,$ X_{vn_v} are v independent random samples from $F_1, F_2, \cdots, F_v$, respectively. We will consider a test $H : F_1 = \cdots = F_v$ against $K : F_i \neq F_j$ for at least one $i \neq j$. One typical choice of the F_i's are $F(x - \theta_1), \cdots, F(x - \theta_v)$, where θ_i denotes the median of i-th population. Then our testing problem is equivalent to $H : \theta_1 = \cdots = \theta_v$ against $K : \theta_i \neq \theta_j$ for at least one $i \neq j$, which is commonly referred to as the one-way layout problem.

Let R_{ij} be the rank of X_{ij} in the combined samples and let $R_i. = \sum_{j=1}^{n_i} R_{ij}$ and $\overline{R}_i. = \frac{R_i.}{n_i}$. Then under the null hypothesis, it can be shown that $E(\overline{R}_i.) = \frac{N+1}{2}$ and $var(\overline{R}_i.) = \frac{(N-n_i)(N+1)}{12n_i}$, where $N = \sum_{i=1}^{v} n_i$. Thus the difference $\overline{R}_i. - \frac{N+1}{2}$ represent the departure from the expected value and we might reject the null hypothesis if the accumulated departure is too large. This suggests a test statistic of the form

$$W = \sum_{i=1}^{v} c_i \left\{ \frac{\overline{R}_i. - \frac{N+1}{2}}{\sqrt{var(\overline{R}_i.)}} \right\}^2 \qquad (2.16)$$

where $c_1, \cdots, c_v$ are constants which are chosen so that W has a convenient distribution. One such statistic is the Kruskal-Wallis statistic. Kruskal and Wallis [10] chose $c_i = 1 - \frac{n_i}{N}$ so that the limiting distribution of W would be chi-square with $v - 1$ degree of freedom.

Theorem 2.5 *The Kruskal-Wallis test is admissible for the one-way layout problem.*

Proof: As before, we begin with the multinomial problem. For convenience we will only consider the case $v = 3$.

Let $X_{11}, \cdots, X_{1n_1}, X_{21}, \cdots, X_{2n_2}$, and $X_{31}, \cdots, X_{3n_3}$ be independent random samples from $multinomial(1, p_1)$, $multinomial\ (1, p_2)$, and $multinomial\ (1, p_3)$, respectively, where $p_i = (p_{i1}, \cdots, p_{ik})$ $i = 1, 2, 3$. We will test the null hypothesis $H : p_1 = p_2 = p_3$ against the alternative hypothesis $K : p_i \neq p_j$ for at least one $i \neq j$.

The joint probability distribution of the X_{ij}'s is

$$f(x_{11}, \cdots, x_{3n_3}; p_1, p_2, p_3) = c(n_1, n_2, n_3) p_{1k}^{n_1} p_{2k}^{n_2} p_{3k}^{n_3}$$
$$\times \exp\left\{ \sum_{i=1}^{3} \sum_{l=1}^{k-1} t_{il} \log \frac{p_{il}}{p_{ik}} \right\}. \qquad (2.17)$$

where t_{il} is the number of observations in the l-th category in the i-th sample. Since the exponent term can be written as

$$\sum_{i=1}^{3} \sum_{l=1}^{k-1} t_{il} \log \frac{p_{il}}{p_{ik}} = \sum_{i=1}^{2} \sum_{i=1}^{k-1} \left\{ t_{il} \left(\log \frac{p_{il}}{p_{ik}} - \log \frac{p_{3l}}{p_{3k}} \right) \right\}$$
$$+ \sum_{l=1}^{k-1} (t_{1l} + t_{2l} + t_{3l}) \log \frac{p_{3l}}{p_{3k}},$$

we can rewrite (2.17) as follows.

$$f(t, u; \omega, \theta) = c_0(n_1, n_2, n_3) \exp\left\{ t\omega + u\theta \right\}$$

where $t\omega = \sum_{i=1}^{2} \sum_{l=1}^{k-1} t_{il}\omega_{il}$, $\omega_{il} = \log \frac{p_{il}/p_{3l}}{p_{1k}/p_{3k}}$, $u\theta = \sum_{l=1}^{k-1} u_l\theta_l$, $u_l = t_{1l} + t_{2l} + t_{3l}$, and $\theta_l = \log \frac{p_{3l}}{p_{3k}}$.

Now the original testing problem is equivalent to $H : \omega = 0$ against $K : \omega \neq 0$ and by Theorem 2.3, $\Phi_{\mathcal{D}}$ forms the minimal complete class for this problem. Hence

it remains to show that the Kruskal-Wallis test belongs to this class. But this is true if we can show that W is a convex function of $t_{11}, \cdots, t_{1k-1}, t_{21}, \cdots, t_{2k-1}$ for each fixed $u_1, \cdots, u_{k-1}$, because the acceptance region of the test is given by $W < c$ for some constant c.

Note that

$$W = \frac{12}{N(N+1)} \sum_{i=1}^{3} n_i \left(\overline{R}_{i\cdot} - \frac{N+1}{2} \right)^2$$

$$= \frac{12}{N(N+1)} \sum_{i=1}^{3} \frac{R_{i\cdot}^2}{n_i} - 3(N+1) \qquad (2.18)$$

where $N = n_1 + n_2 + n_3$. Thus if we show that $\sum_{i=1}^{3} \frac{R_{i\cdot}^2}{n_i}$ is a convex function of $t_{11}, \cdots, t_{1k-1}, t_{21}, \cdots, t_{2k-1}$ for each fixed $u_l, \cdots, u_{k-1}$, then we are done. Since each $0, 1, \ldots, k-1$ has rank $\frac{u_1+1}{2}, u_1 + \frac{u_2+1}{2}, u_1 + u_2 + \frac{u_3+1}{2}, \cdots, \sum_{l=1}^{k-1} u_l + \frac{u_k+1}{2}$, respectively,

$$R_{i\cdot} = t_{i1} \left(\frac{u_1 + 1}{2} \right) + \cdots + t_{ik} \left(u_1 + \cdots + u_{k-1} + \frac{u_k + 1}{2} \right) \quad \text{for } i = 1, 2$$

and

$$R_{3\cdot} = (u_1 - t_{11} - t_{21}) \left(\frac{u_1 + 1}{2} \right) + \cdots$$
$$+ (u_k - t_{1k} - t_{2k}) \left(u_1 + \cdots + u_{k-1} + \frac{u_k + 1}{2} \right)$$

Clearly, for fixed $u_1, \cdots, u_{k-1}, R_{1\cdot}$ is a linear function of $t_{11}, \cdots, t_{1k-1}, R_{2\cdot}$ is a linear function of $t_{21}, \cdots, t_{2k-1}$ and $R_{3\cdot}$ is a linear function of $t_{11}, \cdots, t_{1k-1}$ and $t_{21}, \cdots, t_{2k-1}$. i.e., the $R_{i\cdot}$'s are convex functions.

Hence by a standard argument, see for example, A. W. Roberts and D. E. Varberg [19, pp. 15-16], we see that $\sum_{i=1}^{3} \frac{R_{i\cdot}^2}{n_i}$ is a convex function of t_{ij}'s for each

fixed u_j's. Hence Kruskal-Wallis test is admissible for the case $v = 3$. This argument can be extended to any finite v and hence admissibility follows for the multinomial problem. Thus by the previous argument, we see that the test is admissible for the one-way layout problem. □

3. A MINIMAL COMPLETE CLASS THEOREM

3.1 The Extended Stepwise Bayes Procedure

Consider a statistical decision problem $(\Theta, \mathcal{D}, L)$, with a random variable $X \in \mathcal{X}$ and $\{f_\theta : \theta \in \Theta\}$ a family of possible probability distributions on $\mathcal{X}$. We assume that Θ is finite and L is nonnegative. Let γ be a real-valued function of the unknown parameter $\theta \in \Theta$. To estimate the true, but unknown, value $\gamma(\theta)$, we must define a real-valued function δ on $\mathcal{X}$. Typically the decision space $\mathcal{D}$ is a bounded closed interval in real line which contains the range space of γ.

For this statistical decision problem, a Bayes procedure against a regular prior is commonly used to find an admissible decision rule. However Hsuan [8] pointed out the importance of nonregular prior distributions in decision theory and many authors, for example, Johnson [9], Hsuan [8], Meeden and Ghosh [14], and Brown [3], utilized nonregular prior distributions to prove the admissibility of well known decision rules. They developed the stepwise Bayes technique and showed that it is successful in finding a minimal complete class when the loss is strictly convex. The class of all stepwise Bayes rules however is not minimal complete if we drop the assumption of strictly convex loss.

In this chapter, we will modify the stepwise Bayes procedure so that we can find a minimal complete class without the assumption of strictly convex loss.

Definition 3.1 Let Γ be a collection of decision rules. Then a decision rule $\delta' \in \Gamma$ is said to be *Bayes within* Γ against π if,

$$\int_\Theta r(\theta, \delta')d\pi \le \int_\Theta r(\theta, \delta)d\pi \quad \text{for all } \delta \in \Gamma.$$

We denote the class of all Bayes rules within Γ against π by $\Gamma(\pi)$.

Definition 3.2 Let $\{\pi^\alpha : \alpha \in I\}$ be a family of prior distributions on Θ where I is a well-ordered set with the least element $\alpha(0)$. Given $\{\pi^\alpha : \alpha \in I\}$, we define $\{\Gamma_\alpha : \alpha \in I\}$ as follows.

(i) $\Gamma_{\alpha(0)} = \mathcal{D}^*(\pi^{\alpha(0)})$ where $\mathcal{D}^*$ is the class of all decision rules.

(ii) $\Gamma_{\alpha^*} = \left(\bigcap_{\alpha < \alpha^*} \Gamma_\alpha\right)(\pi^{\alpha^*})$ for $\alpha^* > \alpha(0)$.

Definition 3.3 We say that a decision rule δ is a extended stepwise Bayes against $\{\pi^\alpha : \alpha \in I\}$ if,

(i) $\{\Gamma^\alpha : \alpha \in I\}$ is a strictly decreasing sequence of sets of decision rules.

(ii) $\delta \in \bigcap_{\alpha \in I} \Gamma_\alpha$

(iii) No member of $\bigcap_{\alpha \in I} \Gamma_\alpha$ dominates any other member of $\bigcap_{\alpha \in I} \Gamma_\alpha$. *i.e.*, within $\bigcap_{\alpha \in I} \Gamma_\alpha$, every member is admissible.

Note that both a stepwise Bayes and an extended stepwise Bayes procedures utilize nonregular prior distributions. As in the case of stepwise Bayes rules, a different

ordering of the prior distributions may result in different extended stepwise Bayes decision rules also and an extended stepwise Bayes rule with respect to $\{\pi^\alpha : \alpha \in I\}$ is necessarily Bayes with respect to $\pi^{\alpha(0)}$, but it need not be Bayes with respect to $\pi^\alpha, \alpha > \alpha(0)$. However, the usual stepwise Bayes procedure employs mutually singular prior distributions to extract decision rules and at each step, say the i-th step, possible decisions are specified for each $x \in \Lambda^i$. Thus the procedure should terminate in a finite number of stages and the resulting decision rule is admissible if it is unique stepwise Bayes. But our procedure does not require the prior distributions to have mutually exclusive supports. We can even use an infinite sequence of prior distributions. Under the strictly convex loss assumption, it is easy to check that a stepwise Bayes rule is a special case of the extended stepwise Bayes rules.

The idea of an extended stepwise Bayes procedure is to reduce the class of possible decision rules at each step and continue the procedure until we find a unique decision rule or can not reduce the class of possible decision rules further.

Before proving a minimal complete class theorem we need some preliminary results. The first theorem yields a slight improvement to a standard result of decision theory, which will be necessary in what follows.

Recall that for a risk set S, the lower boundary of $S, \lambda(S)$ is just the class of admissible rules.

Theorem 3.1 *Let $S \subset \mathcal{R}^k$ be a risk set. Suppose S is convex, closed and bounded from below, and $\lambda(S)$ is a proper subset of S. If $u^* \in \lambda(S)$, then there exists a prior distribution π against which u^* is Bayes and the class of all Bayes rule against π is a proper subset of S.*

Proof: We will prove the theorem by induction. Assume $k = 2$. Let $\mathbf{u}^* = \left(u_1^*, u_2^*\right) \in \lambda(\mathcal{S})$, then $\mathbf{u}^*$ is admissible. Hence there exists a prior $\pi = (\pi_1, \pi_2)$ against which $\mathbf{u}^*$ is Bayes. If there exists $\mathbf{v} = (v_1, v_2) \in \mathcal{S}$ such that $\pi_1 u_1^* + \pi_2 u_2^* < \pi_1 v_1 + \pi_2 v_2$, we are done. Thus we assume this is not the case. *i.e.*, the risk set $\mathcal{S}$ is contained in the hyperplane

$$\pi_1 u_1^* + \pi_2 u_2^* = constant \doteq c \tag{3.1}$$

If both π_1 and π_2 are greater the zero, then given $\mathbf{v} \in \mathcal{S} - \lambda(\mathcal{S})$, there exists $\mathbf{u}' = \left(u_1', u_2'\right) \in \lambda(\mathcal{S})$ such that $u_i' \leq v_i$ for $i = 1, 2$ with strict inequality for some i. Hence $\pi_1 u_1' + \pi_2 u_2' < \pi_1 v_1 + \pi_2 v_2$, which contradicts (3.1). Thus we may assume, without loss of generality, $\pi_1 = 1$. Then the set of risk points corresponding to the Bayes rules against the prior is $\mathcal{S}' = \left\{(u_1, u_2) \in \mathcal{S} : u_1 = u_1^*\right\}$. If there exists $\mathbf{v} = (v_1, v_2) \in \mathcal{S}$ such that $v_1 \neq u_1^*$, we are done.

Assume $\mathcal{S}' = \mathcal{S}$. *i.e.*, for all $\mathbf{v} \in \mathcal{S}, v_1 = u_1^*$. Since $\mathbf{u}^*$ is admissible, it must be the case that $u_2^* = \inf_{(u_1, u_2) \in \mathcal{S}} u_2$ and hence $\mathbf{u}^*$ is also Bayes against $\pi' = (0, 1)$. If the class of Bayes rules against π' is not a proper subset of $\mathcal{S}$, then for all $\mathbf{v} \in \mathcal{S}, v_2 = u_2^*$. So $\mathcal{S}$ contains just one point $\mathbf{u}^* = \left(u_1^*, u_2^*\right)$, which contradicts our assumption that $\lambda(\mathcal{S})$ is a proper subset of $\mathcal{S}$.

Now assume that theorem is true for $k - 1$. Suppose $\mathbf{u}^* = \left(u_1^*, u_2^*, \cdots u_k^*\right) \in \lambda(\mathcal{S})$, then $\mathbf{u}^*$ is admissible and there exists $\pi = (\pi_1, \pi_2, \cdots, \pi_k)$ against which $\mathbf{u}^*$ is Bayes. We are done if there exists $\mathbf{v} = (v_1, v_2, \cdots, v_k) \in \mathcal{S}$ such that $\sum_{i=1}^{k} u_i^* \pi_i < \sum_{i=1}^{k} v_i \pi_i$. So we assume that the risk set $\mathcal{S}$ is contained in the hyperplane

$$\sum_{i=1}^{k} \pi_i u_i^* = constant = c \tag{3.2}$$

If each π_i is greater than zero, then given $\mathbf{v} \in \mathcal{S} - \lambda(\mathcal{S})$, there exists $\mathbf{u}' = \left(u_1', u_2', \cdots, u_k'\right) \in \lambda(\mathcal{S})$ such that $u_i' \leq v_i$ for $i = 1, 2, \ldots, k$ with strict inequality for some i, and

$$\sum_{i=1}^{k} \pi_i u_i' < \sum_{i=1}^{k} \pi_i v_i,$$

which contradicts (3.2). Hence we assume that there exists at least one $\pi_i = 0$. For convenience, say, $\pi_k = 0$ and assume that all the others are not zero. Then, for $\mathbf{u} \in \mathcal{S}$,

$$u_{k-1} = c' - \sum_{i=1}^{k-2} r_i u_i \tag{3.3}$$

by (3.2) where $c' = c/\pi_{k-1}$ and $r_i = \pi_i/\pi_{k-1} > 0$ for $i = 1, 2, \ldots, k - 2$. Thus if $\mathbf{u} \in \mathcal{S}$, then given $u_1, \cdots, u_{k-2}$, u_{k-1} is independent of u_k. Hence by (3.3) $\mathcal{S}$ is really a "$(k-1)$-dimensional set".

Let $\mathcal{S}_o = \left\{ (u_1, \cdots, u_{k-2}, u_k) : (u_1, \cdots, u_{k-2}, u_{k-1}, u_k) \in \mathcal{S} \right\}$. Note that $\mathcal{S}_o$ is convex and bounded below. If $\mathbf{u} \in \mathcal{S}$, let $\mathbf{u}^o$ be its corresponding member in $\mathcal{S}_o$.

Now we will show that $\mathcal{S}_o$ is closed from below. First we will prove that $\mathbf{u}_o = Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o$ if and only if $\mathbf{u} = Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$.

Assume $\mathbf{u}_o = Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o$. Then it is clear that $\mathbf{u} \in Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$. If there exists $\mathbf{v}$ such that $\mathbf{v} \neq \mathbf{u}$ and $\mathbf{v} \in Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$, then $\mathbf{v}_o \in Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o = \mathbf{u}_o$. Since $\mathbf{v} \in Q_{\mathbf{u}}$, it must be the case that $\mathbf{v} = (u_1, \cdots, u_{k-2}, u_{k-1} - \varepsilon, u_k)$ for some $\varepsilon > 0$, but $\mathbf{v} \in \overline{\mathcal{S}}$, which is impossible because of (3.3). Hence $\mathbf{u}_o = Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o \implies \mathbf{u} = Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$.

Suppose now $\mathbf{u} = Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$, then $\mathbf{u}_o \in Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o$. If there exists $\mathbf{v}_o$ such that $\mathbf{v}_o \neq \mathbf{u}_o$ and $\mathbf{v}_o \in Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o$, then we must have $v_i \geq u_i$ for $i = 1, 2, \ldots, k - 2$ and k, since $\mathbf{u} = Q_{\mathbf{u}} \cap \overline{\mathcal{S}}$, while $\mathbf{v}_o \in Q_{\mathbf{u}_o}$ implies that $v_i \leq u_i$ for $i = 1, 2, \ldots, k - 2$ and k, which is a contradiction. Thus $\mathbf{u} = Q_{\mathbf{u}} \cap \overline{\mathcal{S}} \implies \mathbf{u}_o = Q_{\mathbf{u}_o} \cap (\overline{\mathcal{S}})_o$. Hence we have

shown that $\mathbf{u}_O = Q_{\mathbf{u}_O} \cap (\overline{S})_O$ if and only if $\mathbf{u} = Q_O \cap \overline{S}$.

Note that $(\overline{S})_O = \overline{(S_O)}$, since by (3.3), we must have convergence in $(k-1)$-th coordinate if we have convergence in the other first $k - 2$ coordinate. Thus

$$\mathbf{u}_O \in \lambda(S_O) \iff \mathbf{u}_O = Q_{\mathbf{u}_O} \cap \overline{(S_O)}$$

$$\iff \mathbf{u}_O = Q_{\mathbf{u}_O} \cap (\overline{S})_O$$

$$\iff \mathbf{u} = Q_\mathbf{u} \cap \overline{S}$$

$$\iff \mathbf{u} \in \lambda(S),$$

and we have $\lambda(S_O) = (\lambda(S))_O \subset S_O$. *i.e.*, S_O is closed from below.

Now by assumption, $\lambda(S)$ is a proper subset of S. Let $\mathbf{v}$ be an inadmissible rule. Then there exists $\mathbf{u}' \in \lambda(S)$ such that $u'_i \leq v_i$ for $i = 1, 2, \ldots, k$ with strict inequality for at least one i. If $u'_i < v_i$ for some $i = 1, 2, \ldots, k - 2$ or k, then $\mathbf{u}'_O$ dominates $\mathbf{v}_O$. If $u'_i = v_i$ for $i = 1, 2, \ldots, k - 2$ and k, then $\mathbf{u}' = \mathbf{v}$ by (3.3), which is a contradiction. Thus if $\mathbf{v}$ is inadmissible, then $\mathbf{v}_O$ is inadmissible as well in S_O. Hence $\lambda(S_O)$ is a proper subset of S_O and we can apply the inductive argument to get the result.

This proof would work as long as at least 2 of π_i's are different from zero. So now suppose $\pi_k = 1$ and all the others are zero. *i.e.*, $\mathbf{u} \in S \implies u_k = constant$. Now let $S_O = \left\{ (u_1, \cdots, u_{k-1}) \,\middle|\, (u_1, \cdots, u_{k-1}, constant) \in S \right\}$. Just as before S_O is convex and bounded from below and $(\overline{S})_O = \overline{(S_O)}$, and by same argument, we can prove the theorem. $\qquad\Box$

Lemma 3.1 *Let $S \subset \mathcal{R}^k$ be a risk set which is convex, closed and bounded from below. Let $\{\pi^\alpha : \alpha \in I\}$ be a sequence of prior distributions and $S_{\Gamma_\alpha} = \{(u_1, \cdots, u_k) :$ for some $\delta \in \Gamma_\alpha,\ u_i = r(\theta_i, \delta)$ for $i = 1, 2, \ldots, k\}$. Then S_{Γ_α} is also convex, closed*

and bounded from below.

Proof : Note that If δ' and δ'' are Bayes with respect to π, then any convex linear combination of δ' and δ'' is also Bayes against π. So the convexity of $\mathcal{S}_{\Gamma_\alpha}$ is an obvious result from the property of Bayes rules.

Now will show that $\mathcal{S}_{\Gamma_\alpha}$ is closed from below, but it suffice to show that the theorem is true for the case $\alpha = \alpha(0)$.

If each of the π_i's are greater than zero, then $\mathcal{S}_{\Gamma_{\alpha(0)}}$ is contained in the hyperplane (3.2). Thus $\mathcal{S}_{\Gamma_{\alpha(0)}}$ is the intersection of a closed set and a set which is closed from below. Therefore it is closed from below.

Assume now that there exists at least one $\pi_i = 0$. For convenience, say, $\pi_k = 0$ and the others are not zero. Then $\mathcal{S}_{\Gamma_{\alpha(0)}} = \{(u_1, \cdots, u_k) \in \mathcal{S} : \sum_{i=1}^{k-1} \pi_i u_i = c\}$, where c is a constant such that $c = \inf_{u \in \mathcal{S}} \sum_{i=1}^{k-1} \pi_i u_i$. Let $\mathbf{u}^* \in \lambda(\mathcal{S}_{\Gamma_{\alpha(0)}})$. If $\mathbf{u}^* \in \mathcal{S}_{\Gamma_{\alpha(0)}}$, then we are done. Thus we assume that this is not the case. Then $\mathbf{u}^*$ is a limit point which does not belong to $\mathcal{S}_{\Gamma_{\alpha(0)}}$. Hence, for every $\varepsilon > 0$, there exists $\mathbf{u}' \in \mathcal{S}_{\Gamma_{\alpha(0)}}$ such that $\left| \mathbf{u}^* - \mathbf{u}' \right| < \varepsilon$. Therefore,

$$
\begin{aligned}
\left| \sum_i^{k-1} \pi_i u_i^* - c \right| &= \left| \sum_{i=1}^{k-1} \pi_i u_i^* - \sum_{i=1}^{k-1} \pi_i u_i' \right| \\
&= \left| \pi \cdot \mathbf{u}^* - \pi \cdot \mathbf{u}' \right| \\
&= \left| \pi \cdot (\mathbf{u}^* - \mathbf{u}') \right| \\
&\leq \left| \pi \right| \left| \mathbf{u}^* - \mathbf{u}' \right| \\
&\leq \left| \mathbf{u}^* - \mathbf{u}' \right| \\
&< \varepsilon \quad \text{for every } \varepsilon > 0
\end{aligned}
$$

and we see that $\sum_{i=1}^{k-1} \pi_i u_i^* = c$. If $\mathbf{u}^* \in \mathcal{S}$, then $\mathbf{u}^* \in \mathcal{S}_{\Gamma_{\alpha(0)}}$ and we are done.

Next we will show that $\mathbf{u}^* \in \mathcal{S}$. Note that $\mathbf{u}^* = Q_{\mathbf{u}^*} \cap \overline{\mathcal{S}_{\Gamma_{\alpha(0)}}} \subset Q_{\mathbf{u}^*} \cap \overline{\mathcal{S}}$. Assume that there exists $\mathbf{u}^o$ such that $\mathbf{u}^o \neq \mathbf{u}^*$ and $\mathbf{u}^o \in Q_{\mathbf{u}^*} \cap \overline{\mathcal{S}}$. If $\mathbf{u}^o \in \mathcal{S}$, then it must be the case that $u_i^o = u_i^*$, for $i = 1, 2, \ldots, k-1$, and $u_k^o = u_k^* - \varepsilon$ for some $\varepsilon > 0$, since $\mathbf{u}^o \in Q_{\mathbf{u}^*}$ and $\sum_{i=1}^{k-1} \pi_i u_i^o \geq c$. Thus $Q_{\mathbf{u}^o}$ is a proper subset of $Q_{\mathbf{u}^*}$ and $Q_{\mathbf{u}^o} \cap \overline{\mathcal{S}_{\Gamma_{\alpha(0)}}} \subset Q_{\mathbf{u}^*} \cap \overline{\mathcal{S}_{\Gamma_{\alpha(0)}}} = \mathbf{u}^*$, but $Q_{\mathbf{u}^o} \cap \overline{\mathcal{S}_{\Gamma_{\alpha(0)}}}$ is a nonempty set, which is a contradiction. Hence $\mathbf{u}^o \notin \mathcal{S}$.

Note that $\left| \mathbf{u}^* - \mathbf{u}' \right| \leq \left| \mathbf{u}^o - \mathbf{u}' \right|$ for every $\mathbf{u}' \in \mathcal{S}$. Let $\left| \mathbf{u}^* - \mathbf{u}^o \right| = \varepsilon$, then for $\mathbf{u}' \in \mathcal{S}$ such that $\left| \mathbf{u}^* - \mathbf{u}' \right| \leq \frac{\varepsilon}{2}$

$$\frac{\varepsilon}{2} \leq \left| \mathbf{u}^* - \mathbf{u}^o \right| - \left| \mathbf{u}^* - \mathbf{u}' \right| \leq \left| \mathbf{u}' - \mathbf{u}^o \right|$$

and for $\mathbf{u}' \in \mathcal{S}$ such that $\left| \mathbf{u}^* - \mathbf{u}' \right| \geq \frac{\varepsilon}{2}$,

$$\frac{\varepsilon}{2} \leq \left| \mathbf{u}^* - \mathbf{u}' \right| \leq \left| \mathbf{u}^o - \mathbf{u}' \right|$$

$\mathbf{u}^o$ can not be a limit point of $\mathcal{S}$ and hence $\mathbf{u}^o \notin \overline{\mathcal{S}}$ which is a contradiction. Thus we have $\mathbf{u}^* = Q_{\mathbf{u}^*} \cap \overline{\mathcal{S}}$, and $\mathbf{u}^* \in \lambda(\mathcal{S}) \subset \mathcal{S}$, since $\mathcal{S}$ is closed from below.

It remains to show that $\mathcal{S}_{\Gamma_{\alpha(0)}}$ is bounded from below, but this is true, since $\mathcal{S}_{\Gamma_{\alpha(0)}} \subset \mathcal{S}$. $\qquad\qquad\square$

Assume now that the risk set $\mathcal{S}$ is closed and bounded from below, and $\lambda(\mathcal{S})$ is a proper subset of $\mathcal{S}$. Then as we remarked earlier, δ is admissible. Conversely suppose a decision rule δ is admissible and for the decision problem $\lambda(\mathcal{S})$ is a proper subset of $\mathcal{S}$. Then there exists a prior distribution $\pi^{\alpha(0)}$ against which δ is Bayes and $\Gamma_{\alpha(0)} = \mathcal{D}^* \left(\pi^{\alpha(0)} \right)$ is a proper subset of $\mathcal{S}$. If no member of $\Gamma_{\alpha(0)}$ dominates any other member of $\Gamma_{\alpha(0)}$, we stop. If this is not the case, then δ is admissible within

$\Gamma_{\alpha(0)}$ and we can repeat the process. That is there exists a prior $\pi^{\alpha(1)}$ such that δ is Bayes within $\Gamma_{\alpha(0)}$ against $\pi^{\alpha(1)}$ and $\Gamma_{\alpha(1)} = \Gamma_{\alpha(0)}\left(\pi^{\alpha(1)}\right)$ is a proper subset of $\Gamma_{\alpha(0)}$. If no member of $\Gamma_{\alpha(1)}$ dominates any other member of $\Gamma_{\alpha(1)}$, we are done. Otherwise we repeat the process. If after a finite number of steps we obtain a $\Gamma_{\alpha(n)}$ for which no member dominates any other member we are done. Suppose this is not the case, then there exists a sequence of priors $\left\{\pi^{\alpha(n)}\right\}$ such that $\Gamma_{\alpha(n)}$ is strictly decreasing and $\delta \in \bigcap_{n=0}^{\infty} \Gamma_{\alpha(n)}$. If no member of the intersection dominates any other member, we are done. If not, it is easy to check that $\bigcap_{n=0}^{\infty} S_{\Gamma_{\alpha(n)}}$ is closed and bounded from below where $S_{\Gamma_{\alpha(n)}}$ is the risk set generated by $\Gamma_{\alpha(n)}$, and there exists a prior π^{ω} such that $\delta \in \left(\bigcap_{n=0}^{\infty} \Gamma_{\alpha(n)}\right)(\pi^{\omega})$ where ω is the final ordinal number after the positive integers and this set is a proper subset of $\bigcap_{n=0}^{\infty} \Gamma_{\alpha(n)}$. If no member of $\left(\bigcap_{n=0}^{\infty} \Gamma_{\alpha(n)}\right)(\pi^{\omega})$ dominates any other member, we are done. Otherwise we must repeat the process. Eventually we must find a set which contains δ and for which no member dominates any other member. Hence we have proved the following theorem.

Theorem 3.2 *Suppose that Θ is finite. If a risk set $S \subset \mathcal{R}^k$ is convex, closed from below, and bounded from below, then the collection of all extended stepwise Bayes rules is a minimal complete class.*

The assumptions of this theorem can be weakened in terms of loss function and decision space. For example, if loss function is lower semicontinuous and decision space $\mathcal{D}$ is a compact subset of $\mathcal{R}$, then for every prior distribution π on Θ, a Bayes rule against π exists and we can apply the same argument. Thus we have the following

corollaryies.

Corollary 3.1 *Suppose that Θ is finite and $\mathcal{D}$ is a compact subset of $\mathcal{R}$. If the loss function is lower semicontinuous, then the collection of all extended stepwise Bayes rules is a minimal complete class.*

Corollary 3.2 *If both Θ and $\mathcal{D}$ is finite, then the collection of all extended stepwise Bayes rules is a minimal complete class.*

Example 3.1 Consider the Example 1.2. Let $\pi^1 = (1,0)$ and $\pi^2 = (0,1)$. Then the collection of Bayes rules with respect to π^1 is $\left\{\delta : 0 \leq \delta(x_1) \leq \frac{1}{2}\right\}$ and a Bayes rule δ within $\mathcal{D}^*(\pi^1)$ with respect to π^2 satisfies $\delta(x_1) = \frac{1}{2}$ and $\frac{1}{2} \leq \delta(x_2) \leq 1$. *i.e.*, $\Gamma_2 = \left\{\delta : \delta(x_1) = \frac{1}{2}, \ \frac{1}{2} \leq \delta(x_2) \leq 1\right\}$.

Note that any $\delta \in \Gamma_2$, $r(\theta,\delta) = 0$ for all $\theta \in \Theta$ and we can not extract a subcollection from Γ_2 by applying the Bayes procedure. Thus a decision rule $\delta \in \Gamma_2$ is a extended stepwise Bayes rule. It also should be noted that Γ_2 is the minimal complete class for this statistical decision problem.

Example 3.2 Let X be hypergeometrically distributed with parameters N, M and n, where N is the number of population, M is the unknown number of defective items and n is the sample size. Let $\theta = \frac{M}{N}$, the proportion of the defective items. Thus $\Theta = \left\{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1\right\}$ and $\mathcal{X} = \{0, 1, \cdots, n\}$. We will estimate θ with $\mathcal{D} = [0,1]$

and a loss function $L(\theta, d)$ where L is such that $L(\theta, d) = 0$ if $\theta = d$, $L(\theta, d) > 0$ if $\theta \neq d$, and continuous in d for each $\theta \in \Theta$.

Let $\pi^0 = \{1, 0, \cdots, 0\}, \pi^1 = \{0, 1, 0, \cdots, 0\}$ and so on. Then $\delta(x) = \frac{x}{N}$ is the unique extended stepwise Bayes rule against $\left\{\pi^0, \pi^1, \cdots, \pi^n\right\}$. It may be a silly estimator of θ, but it is still admissible.

Example 3.3 Let $\mathcal{X} = \{1, 2, \cdots, N\}, \Theta = \{1, 2, \cdots, N\}$ and $X_i, i = 1, 2, \ldots, n$ be random samples having probability distribution function

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{if } x = 1, 2, \ldots, \theta \\ 0 & \text{otherwise.} \end{cases}$$

We will estimate θ with $\mathcal{D} = [1, N]$ and a loss function $L(\theta, d)$ where L is such that $L(\theta, d) = 0$ if $\theta = d$, $L(\theta, d) > 0$ if $\theta \neq d$, and continuous in d for each $\theta \in \Theta$.

Let $T = \max_{i=1,\ldots,n} X_i$, then the p.d.f. of T is given by

$$p_\theta(t) = \begin{cases} \left(\frac{t}{\theta}\right)^n - \left(\frac{t-1}{\theta}\right)^n & \text{if } t = 1, 2, \ldots, \theta \\ 0 & \text{otherwise.} \end{cases}$$

Let $\pi^1 = \{1, 0, \cdots, 0\}, \pi^2 = \{0, 1, 0, \cdots, 0\}, \cdots; \pi^N = \{0, \cdots, 0, 1\}$. Then $\delta(t) = t$ is the unique extended stepwise Bayes rule against $\left\{\pi^1, \cdots, \pi^N\right\}$ and hence it is admissible.

3.2 Set Estimation

3.2.1 Introduction

In many statistical decision problem, instead of specifying a point estimate for a unknown parameter θ, it may be sufficient to provide an interval or set within which θ may be expected to lie. There are a wide variety of approaches to this statistical decision problem. For example, frequentists, Bayesians, empirical Bayesians and fiducialists, each have their own viewpoints.

Like point estimation, the problem of set estimation is two fold. First, there is the problem of finding confidence procedures, and, second, there is the problem of determining good, or optimum, confidence procedures. Even though statisticians may have different viewpoints, they would not want to use a confidence procedure if there were another with at least as large probability of containing θ and no larger expected set size for all θ. Thus we might use the pair, noncoverage probability and expected set size, as our risk in the set estimation problem.

Let $X \in \mathcal{X}$ be a random variable and $\{f_\theta : \theta \in \Theta\}$ a family of possible probability distributions on $\mathcal{X}$. We denote ν and μ as σ-finite measures on Θ and $\mathcal{X}$, respectively, and

$$Pr_\theta[X \in H] = \int_H f_\theta(x)d\mu(x)$$

where H is a measurable set on $\mathcal{X}$. Assume that f_θ is jointly measurable and for each $x \in \mathcal{X}$, there is at least one θ with $f_\theta(x) > 0$.

A nonrandomized confidence procedure T is a jointly measurable set on $\Theta \times \mathcal{X}$. If a statistician use the procedure T and observe $x \in \mathcal{X}$, then $T(x)$ is his set estimate of θ where $T(x)$ is the x-section of T.

For this nonrandomized confidence procedure T, define

$$\eta_T(\theta) = E\left[\nu\left(T(X)\right)\right] = \int_{\mathcal{X}} \nu\left(T(x)\right) f_\theta(x) d\mu(x)$$

and

$$p_T(\theta) = Pr_\theta\left[\theta \notin T(X)\right] = \int_{\mathcal{X}} I\left[\theta \notin T(x)\right] f_\theta(x) d\mu(x)$$

where I is the indicator function. We also define a randomized confidence procedure T_r, and its risk $\eta_{T_r}(\theta)$ and $p_{T_r}(\theta)$ in a similar manner as in a point estimation problem.

In what follows, unless it is necessary to distinguish between nonrandomized and randomized procedures, T will refer to both a nonrandomized and a randomized confidence procedure.

Definition 3.4 A confidence procedure T' is called *admissible* if there is no other procedure T with (a) $\eta_T(\theta) \leq \eta_{T'}(\theta)$ and (b) $p_T(\theta) \leq p_{T'}(\theta)$ for all $\theta \in \Theta$, where for some θ at least one of inequalities (a) and (b) is strict.

This definition of admissibility of a confidence procedure has, for example, been used by Cohen and Sackrowitz [5], Meeden and Vardeman [16], and Meeden [17]. Meeden and Vardeman also gave three different notions of being Bayes for set estimate. One of them is $c(\theta)$-Bayes.

Definition 3.5 Let Π be a distribution on Θ with a density with respect to ν given by $\pi(\theta)$, and $c(\theta)$ be a function taking values in $[0, 1]$. Then a confidence procedure T' is called $c(\theta)$-*Bayes versus* Π if there is no other confidence procedure T such

that

$$\int \left[c(\theta)\eta_T(\theta) + (1 - c(\theta))p_T(\theta) \right] \pi(\theta)d\nu(\theta)$$

$$< \int \left[c(\theta)\eta_{T'}(\theta) + (1 - c(\theta))p_{T'}(\theta) \right] \pi(\theta)d\nu(\theta)$$

In the decision theoretical point of view, the concept of $c(\theta)$-Bayesness is important because there is a close relationship between admissibility and being $c(\theta)$-Bayes. That is, Meeden and Vardeman showed that, when Θ is finite and a confidence procedure T is admissible, there exist a prior Π and a function $c(\theta)$ such that T is a $c(\theta)$-Bayes versus Π. The concept of $c(\theta)$-Bayesness may be somewhat strange. But the function $c(\theta)$ has an intuitive interpretation, that is, for a given θ, $c(\theta)$ is the statistician's relative weight of the set size against the noncoverage probability. For example, if $c(\theta) = 1$, then for that θ, the statistician is only concerned with controlling $\eta_T(\theta)$, the average size of confidence procedure T.

Theorem 3.3 *A nonrandomized confidence procedure T is a $c(\theta)$-Bayes versus Π if and only if, for a set of x receiving marginal probability one:*

(i) $\int c(\theta)\pi(\theta|x)d\nu(\theta) = 0$ *implies that* $\{\theta : \pi(\theta|x) > 0\} - T(x)$ *has probability 0 under the conditional distribution of $\theta|X = x$.*

(ii) $\int c(\theta)\pi(\theta|x)d\nu(\theta) > 0$ *implies that the set* $T(x) - \{\theta : \pi(\theta|x) > 0\}$ *has ν measure 0 and both*

$$\left\{ \theta : (1 - c(\theta))\pi(\theta|x) > \int c(\theta)\pi(\theta|x)d\nu(\theta) \right\} - T(x)$$

and

$$\left\{ \theta : (1 - c(\theta))\,\pi(\theta|x) < \int c(\theta)\pi(\theta|x)d\nu(\theta) \right\} \cap T(x)$$

have probability 0 under the conditional distribution of $\theta|X = x$.

Theorem 3.3 says, essentially, a confidence procedure that is $c(\theta)$-Bayes versus II can be obtained as

$$T^{\mathrm{II}} = \left\{ (\theta, x) : \pi(\theta|x)(1 - c(\theta)) > \int c(\theta)\pi(\theta|x)d\nu(\theta) \right\}, \tag{3.4}$$

and we may randomize on the region $\{(\theta, x) : \pi(\theta|x)(1 - c(\theta)) = \int c(\theta)\pi(\theta|x)d\nu(\theta)\}$. This theorem is due to Meeden and Vardeman [16].

3.2.2   A Minimal Complete Class in Set Estimation Problems

In the decision theoretical framework, the set estimation problem is a triple $(\Theta, \mathcal{D}, L)$, with a random variable $X \in \mathcal{X}$, where $\mathcal{D}$ is a collection of subsets of Θ and $L(\theta, d) = (\nu(d), I(\theta \notin d))$ is a vector valued loss function. Thus a confidence procedure is a decision function which assigns a set in $\mathcal{D}$ for each $x \in \mathcal{X}$. However we will use the term *"confidence procedure"* to distinguish this from point estimation.

Note that, when Θ is finite, say $\Theta = \{\theta_1, \cdots, \theta_k\}$, the subset of $\mathcal{R}^{2k}, \mathcal{C} = \{(\eta_T(\theta_1), \cdots, \eta_T(\theta_k), p_T(\theta_1), \cdots, p_T(\theta_k)) : T$ is a confidence procedure $\}$ plays the role of the risk set in the point estimation. Thus under certain conditions, we can apply the Theorem 3.1 and Lemma 3.1 to the set $\mathcal{C}$.

We will now define the extended stepwise Bayes confidence procedure in a similar manner, but unlike point estimation, we need an ordered set of pairs (π^α, c^α) instead of an ordered set of prior distributions.

Definition 3.6 Let Γ be a collection of confidence procedures. Let $\pi(\theta)$ be a prior distribution and $c(\theta)$ be a function taking value in $[0,1]$. Then a confidence procedure T' is said to be *Bayes within* Γ with respect to (π,c) if,

$$\int_\Theta \left[c(\theta)\eta_{T'} + (1-c(\theta))p_{T'} \right] d\pi(\theta)$$

$$\leq \int_\Theta \left[c(\theta)\eta_{T} + (1-c(\theta))p_{T} \right] d\pi(\theta) \quad \text{for all } T \in \Gamma$$

we denote the class of all Bayes within Γ against (π,c) by $\Gamma(\pi,c)$

Definition 3.7 Let I be a well-ordered set with the least element $\alpha(0)$. Given $\{(\pi^\alpha,c^\alpha) : \alpha \in I\}$, we define $\{\Gamma_\alpha : \alpha \in I\}$ as follows.

(i) $\Gamma_{\alpha(0)} = \mathcal{D}^* \left(\pi^{\alpha(0)}, c^{\alpha(0)} \right)$ where $\mathcal{D}^*$ is the class of all confidence procedures.

(ii) $\Gamma_{\alpha^*} = \left(\bigcap_{\alpha < \alpha^*} \Gamma_\alpha \right) \left(\pi^{\alpha^*}, c^{\alpha^*} \right)$ for $\alpha^* > \alpha(0)$.

Definition 3.8 We say that a confidence procedure T is a extended stepwise Bayes against $\{(\pi^\alpha,c^\alpha) : \alpha \in I\}$ if,

(i) $\{\Gamma^\alpha : \alpha \in I\}$ is a strictly decreasing sequence of sets of confidence procedures.

(ii) $T \in \bigcap_{\alpha \in I} \Gamma_\alpha$

(iii) No member of $\bigcap_{\alpha \in I} \Gamma_\alpha$ dominates any other member of $\bigcap_{\alpha \in I} \Gamma_\alpha$. *i.e.*, within $\bigcap_{\alpha \in I} \Gamma_\alpha$, every member is admissible.

Note that $\mathcal{D}^*(\pi, c)$ is the collection of procedures that are $c(\theta)$-Bayes versus π. However, $c(\theta)$ is closely related to $\pi(\theta)$ in our discussion and for notational convenience, we will call it Bayes with respect to (π, c).

Theorem 3.4 *Suppose that $C \subset \mathcal{R}^{2k}$ is convex, closed from below and bounded from below, and $\lambda(C)$ is a proper subset of C. If $\mathbf{u}^* \in \lambda(C)$, then there exists a pair (π, c) against which $\mathbf{u}^*$ is Bayes and the class of all Bayes procedures with respect to (π, c) is a proper subset of C.*

Proof :   The convexity and boundedness is trivial.

In view of Theorem 3.1, if $\mathbf{u}^* \in \lambda(C)$, then there exists numbers $\phi_1, \cdots, \phi_k$ and $\rho_1, \cdots, \rho_k$ such that each $\phi_i \geq 0$, each $\rho_i \geq 0$, $\sum_{i=1}^{k} \phi_i + \sum_{i=1}^{k} \rho_i = 1$,

$$constant = \sum_{i=1}^{k} \phi_i u_i^* + \sum_{i=1}^{k} \rho_i u_i^* \leq \sum_{i=1}^{k} \phi_i u_i' + \sum_{i=1}^{k} \rho_i u_i' \quad \text{for all } u' \in C,$$

and the set $\left\{ \mathbf{u} \in C : \sum_{i=1}^{k} \phi_i u_i + \sum_{i=1}^{k} \rho_i u_i = constant \right\}$ is a proper subset of C.

Let $\pi_i = \pi(\theta_i) = \phi_i + \rho_i$ for $i = 1, 2, \ldots, k$, and

$$c_i = c(\theta_i) = \begin{cases} \frac{\phi_i}{\pi_i} & \text{if } \pi > 0 \\ \frac{1}{2} & \text{if } \pi = 0 \end{cases}$$

Then $\mathbf{u}^*$ minimizes

$$\sum_{i}^{k} \left[c_i u_i + (1 - c_i) u_{k+i} \right] \pi_i \tag{3.5}$$

over all $\mathbf{u} \in C$, and the Bayes procedures with respect to (π, c) is a proper subset of C. $\qquad \square$

Lemma 3.2 *Let $C \subset \mathcal{R}^{2k}$ be a risk set which is convex, closed and bounded from below, and $C_{\Gamma_\alpha} = \{(u_1, \cdots, u_{2k}) \in C : for\ some\ T \in \Gamma_\alpha\left((\pi, c)\right), u_i = \eta\ T(\theta_i), u_{k+i} = p\ T(\theta_i)\ for\ i = 1, 2, \ldots, k\}$. Then C_{Γ_α} is convex, closed and bounded from below.*

Proof :   As before it suffice to show that $\Gamma_{\alpha(0)}$ is convex, closed and bounded from below. Let $\mathbf{u}^* \in C$ be a Bayes with respect to (π, c). Then $\mathbf{u}^*$ minimizes (3.5) over all $\mathbf{u} \in C$. Let $\phi_i = c_i \pi_i$ and $\rho_i = \left(1 - c_i\right) \pi_i$ for $i = 1, 2, \ldots, k$. Then $\phi_i, \rho_i \geq 0$ for all i and $\sum_{i=1}^{k} \left(\phi_i + \rho_i\right) = 1$. Thus $\mathbf{u}^*$ is a Bayes with respect to $(\phi_1, \cdots, \phi_k, \rho_1, \cdots, \rho_k)$. Hence by Lemma 3.1, C_{Γ_α} is convex, closed and bounded from below.   □

Thus, under certain condition, the extended stepwise Bayes technique yields a minimal complete class in the set estimation problem as well. Suppose, for example that $\mathcal{D}$ is the power set of Θ. Then $\mathcal{D}$ is finite and the risk set C is compact. Thus we have a minimal complete class.

Example 3.4 Consider the Example 1.1. Let ν is the counting measure on $\mathcal{B}_\Theta$ and $\pi^1 = (\frac{1}{2}, \frac{1}{2}, 0, 0), c^1 = (\frac{1}{6}, \frac{1}{6}, \frac{1}{2}, \frac{1}{2})$. By (3.4), a Bayes confidence procedure with respect to $\left(\pi^1, c^1\right)$ satisfies $T(x = 0) = \left\{0, \frac{1}{3}\right\}$ and $T(x = 1) = \left\{\frac{1}{3}\right\}$. Let $\pi^2 = (0, 0\frac{1}{2}, \frac{1}{2})$ and $c^2 = (\frac{1}{2}, \frac{1}{2}, \frac{1}{6}, \frac{1}{6})$. Then T' is the unique member of Γ_2 and hence admissible where $T'(x = 0) = \left\{0, \frac{1}{3}\right\}, T'(x = 1) = \left\{\frac{1}{3}\right\}$ and $T'(x = 2) = \left\{\frac{2}{3}, 1\right\}$.

4. BIBLIOGRAPHY

[1] Berger, J. O. (1985). *Statistical decision theory and Bayes analysis.* Springer-Verlag, New York.

[2] Birnbaum, A. (1955). *Characterizations of complete classes of some multiparameter hypothesis, with application to likelihood ratio tests.* Annals of Mathematical Statistics 26: 21-36.

[3] Brown, L. D. (1981). *A complete class theorem for statistical problems with finite sample space.* Annals of Statistics 9: 1289-1300.

[4] Cohen, M. P. and Kuo, L. (1985). *The admissibility of the empirical distribution function.* Annals of Statistics 13: 262-271.

[5] Cohen, A. and Sackrowitz, H. B. (1984). *Decision theory results for vector risks with applications.* Statistics and Decisions Supplement Issue No. 1, Recent Results in Estimation Theory and Related Topics: 159-176.

[6] Ferguson, T. S. (1967). *Mathematical Statistics.* Academic Press, New York.

[7] Hettmansperger, T. P. (1984). *Statistical inference based on ranks.* John Wiley & Sons, Inc., New York.

[8] Hsuan, F. C. (1979). *A stepwise Bayesian procedure*. Annals of Statistics 7: 860-868.

[9] Johnson, B. M. (1971). *On the admissible estimators for certain fixed sample binomial problems*. Annals of Mathematical Statistics 42: 1579-1587.

[10] Kruskal, W. H. and Wallis, W. A. (1952). *Use of ranks in one-criterion variance analysis*. Journal of American Statistical Association 47: 583-621.

[11] Lehmann, E. L. (1983). *Theory of Point estimation*. John Wiley & Sons, Inc., New York.

[12] Lehmann, E. L. (1986). *Testing statistical hypotheses*. John Wiley & Sons, Inc., New York.

[13] Matthes, T. K. and Truax, D. R. (1967). *Tests of composite hypotheses for the multivariate exponential family*. Annals of Mathematical Statistics 38: 681-697.

[14] Meeden, G. and Ghosh, M. (1981). *Admissibility in finite problems*. Annals of Statistics 9: 846-852.

[15] Meeden, G., Ghosh, M. and Vardeman, S. (1985). *Some admissible nonparametric and related finite population sampling estimators*. Annals of Statistics 2: 811-817.

[16] Meeden, G., Vardeman, S. (1985). *Bayes and admissible set estimation*. Journal of American Statistical Association 80 465-471.

[17] Meeden, G. (1990). *Admissible contour credible sets*. Statistics and Decisions 8: 1-10.

[18] Randles, R. H., Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics.* John Wiley & Sons, Inc., New York.

[19] Roberts, A. W., and Varberg, D. E. (1973). *Convex functions.* Academic Press, New York.

[20] Stein, C. (1956). *The admissibility of Hotelling's T^2-test.* Annals of Mathematical Statistics 27: 616-623.

[21] Wald, A. (1950). *Statistical decision functions.* John Wiley & Sons, Inc., New York.

5. ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my major professor Dr. Glen D. Meeden for his considerable comments, advice, guidance, and constant encouragement throughout this dissertation.

Also I would like to thank all the professors in the Department of Statistics, Iowa State University, for their valuable teaching and advice during my graduate studies.

Special thanks are due to my parents, Hong-Joo Li and Sook-Ja Kim, and my wife Hyun-Gyung for their constant encouragement during the preparation of this dissertation. I dedicate this dissertation to them.