

THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

A BAYESIAN TREATMENT OF A MULTIPLE COMPARISON

PROBLEM FOR BINOMIAL PROBABILITIES

by

Thomas Lester Bratcher

Technical Report No. 49
Department of Statistics THEMIS Contract

November 24, 1969

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

This document has been approved for public release
and sale; its distribution is unlimited.

DEPARTMENT OF STATISTICS
Southern Methodist University

A BAYESIAN TREATMENT OF A MULTIPLE COMPARISON
PROBLEM FOR BINOMIAL PROBABILITIES

Approved by:

A BAYESIAN TREATMENT OF A MULTIPLE COMPARISON
PROBLEM FOR BINOMIAL PROBABILITIES

A Dissertation Presented to the Faculty of the Graduate School

of

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistics

by

Thomas Lester Bratcher
(B.S., University of Texas at Arlington, 1963;
M.S., Southern Methodist University, 1965)

November 24, 1969

Bratcher, Thomas Lester

B.S., University of
Texas at Arlington,
1963
M.S., Southern Methodist
University, 1965

A Bayesian Treatment of a Multiple Comparison Problem for Binomial

Probabilities

Adviser: Associate Professor Richard P. Bland

Doctor of Philosophy degree conferred December 20, 1969

Dissertation completed November 24, 1969

The goal of this paper is to develop a usable procedure which will

solve a practical ranking problem for the probability parameters of binomial populations. The proposed procedure follows from an essentially

Bayesian approach to the multiple comparisons problem applied to the binomial model. The comparisons problem is formulated in decision theoretic terms as a multiple decision problem. The loss function is taken as

the sum of the losses of the component pair-wise comparison problems which generate the multiple comparisons problem. With the additive loss assumption, compatible Bayes rules for the component problems yield a minimum

risk solution to the ranking problem consequently reducing the complexity of the problem.

This approach to the multiple comparisons problem allows for the utilization of any prior information on the individual binomial probabilities.

The ranking of the parameters is to be based on sufficient statistics which are the numbers of successes in samples from each population. This procedure allows for decisions to be made on unequal sample sizes.

A related sequential problem is also considered. It is shown that a Bayes sequential test is truncated. An upper bound is found on the maximum sample size that a Bayes procedure can take so that its computation is possible by backward induction.

ACKNOWLEDGEMENTS

My personal gratitude is hardly adequate repayment for the profound effect which Dr. Richard P. Bland has had upon my achievements. His

guidance throughout my graduate studies and research not only has made this dissertation possible but has been the principal contribution in forming my professional direction. Another to whom I owe a great debt

of gratitude is Dr. Paul D. Minton who on innumerable occasions has come to my aid with personal, professional and financial assistance. Others whose contributions have enabled my graduate studies to culminate in

this dissertation include: Dr. John T. Webster, Dr. Donald B. Owen, Dr.

C. H. Kapadia, and Dr. Jean B. Richmond.

Also my thanks go to Mrs. Judy Ham for her very capable job of

typing.

TABLE OF CONTENTS

Page	
iv	ABSTRACT
vi	ACKNOWLEDGEMENTS
	Chapter
1	I. INTRODUCTION
1	1.1 Summary
2	1.2 Applications and Examples
5	1.3 Existing Methods
6	1.4 The Proposed Procedure
9	II. A DECISION THEORETIC FORMULATION
9	2.1 The Multiple Comparisons Problem
13	2.2 Additional Formulation
16	2.3 The Bayes Risk
19	III. A BAYES SOLUTION
19	3.1 Reduction of a Component Problem
23	3.2 A Solution to the Generating Problems
31	3.3 An Alternate Rule
36	IV. SEQUENTIAL PROBLEMS
36	4.1 Introduction
37	4.2 Terminology
38	4.3 Truncated Sequential Decision Rules
40	4.4 Symmetric Linear Loss
44	4.5 Linear Loss
47	V. DISCUSSION
47	5.1 Priors and Losses
50	5.2 Sequential Problems
51	BIBLIOGRAPHY

based on a sufficient statistic is found and compared to the solution given in Bland and Bratcher [4][†] which is not based on a sufficient statistic. Some sequential problems of related interest are considered in Chapter IV. Chapter V discusses loss functions, prior information and areas in which further work is necessary. The main contribution of this work is that it offers a usable technique with the properties studied herein (usable in that computer programs and tables are available for the practical implementation of the procedures).

1.2 Applications and Examples

The problem is introduced by the following three examples from the areas of research, training and reliability.

Example 1.--- Suppose that researchers infect a number of rats with some disease. The rats are then divided into, say, three groups and given one of the three treatments under investigation. The researchers may very well be interested in procedures to compare the proportion of successes for the various treatments, a success being defined as a test animal for which the disease has been arrested by the end of, say, one week after treatments begin. The three groups, after treatment, are considered as samples from three distinct populations.

Example 2.--- For the second example, suppose that a large corporation has just received a substantial government contract and must hire, say, two thousand additional skilled and semiskilled workers. Suppose also that

[†]The numbers in square brackets refer to the bibliography.

size n from a Bernoulli process can, with no loss of information about the great interest to the experimenter. (As is well-known a random sample of

in this work would offer a ranking of probabilities which should be of model, the development of such a procedure in Bland and Bratcher [4] and bution differs. Also with the wide applicability of the Bernoulli process themselves to the same type of solution in that only the underlying distribution is quite often chosen. It would appear that the above problems would lend to be parameters for normal populations), a multiple comparisons technique When an experimenter is faced with comparing means (usually assumed

probabilities.

ties containing the largest or 3) to choose some type of ranking of the the largest probability of success, 2) to choose a subset of probabilities compare the probabilities, could be one of the three goals: 1) to choose Just what the experimenter means when he states that he wishes to

the success probabilities.

observed each of the Bernoulli processes under consideration to compare samples $p=3, 2$ and 5 respectively). The problem in each case is having θ_1 is the probability of success for the i th process and in the ex- eral (say p) independent Bernoulli processes with parameters $\theta_1, \theta_2, \dots$ then all of the examples can be considered as cases where there are sev- ure of an item does not affect the success or failure of any other item in a process remains constant from item to item and the success or fail- order to make the comparison. Assuming that the probability of success experiment has been carried out on several items for each process in

parameter θ , be considered a sample from a binomial distribution or population with parameters n , known, and θ . Hence we will use the terminology of Bernoulli process and binomial distribution, or population, interchangeably in the following.)

1.3 Existing Methods

Sobel and Huyett [15] consider the type of problem being considered

here using an approach similar to that developed by Bechhofer [2]. The

goal of their procedure is to select only a single binomial population

as the best; that is to select the largest probability of success. For

their procedure, the experimenter must specify two constants, p^* and δ^* ,

then this "indifference" zone approach guarantees with probability p^*

that the largest probability will be selected whenever the second largest

is at least a distance δ^* away. The common sample size n is thus deter-

mined by p^* and δ^* . The decision rule is to choose that population (i.e.

probability) which generates the greatest number of successes in a sample

of size n .

A somewhat more flexible procedure is presented by Gupta and Sobel

[8] (more flexible in that decisions can be based on any sample size).

This procedure selects a subset of the given binomial probabilities which

includes the largest with the specified probability p^* . If one probability

must be selected as the best it may be chosen from the retained subset on

a non-statistical basis. Of course, the above is a screening procedure

and additional experimentation could be conducted on the retained subset. For

the former procedure, the distance (δ^*) between the largest and second largest probability is preassigned, the number of observations needed is tabled and the only information given is the assertion that a single population has the largest probability of success; for the latter procedure, the sample size is given, the critical difference is tabulated and the decision rule is to assert that a subset of the probabilities includes the largest. The number of probabilities in the selected subset, of course, depends on the data.

A ranking procedure is also discussed in Bechhofer, Kiefer, and

Sobel [3] utilizing an approach similar to the first procedure. Let

$\delta_{k-i+1, k-i}^*$ be the distance between the probability with rank $(k-i+1)$

and the probability with rank $(k-i)$, $i=1, 2, \dots, k$. Then the constants

$\{\delta_{k, k-1}^*, \delta_{k, k-2}^*, \dots, \delta_{2, 1}^*\}$ must be specified where the probability

of achieving a correct ranking is at least P^* whenever

$$\delta_{k-i+1, k-i}^* \geq \delta_{k-i+1, k-i}^*, i=1, 2, \dots, k-1.$$

1.4 The Proposed Procedure

The proposed procedure follows from an essentially Bayesian approach

to the multiple comparisons problem applied to the binomial model. The

multiple comparisons problem makes decisions about the differences among

several means by simultaneously ranking each and every pair of the means.

In order to use the Bayes criterion the problem is defined in decision-

theoretic terms; that is, the problem is formulated as a multiple decision

problem and loss functions are defined. Duncan [5 and 6] uses this approach

for comparing means of normal populations and Bland and Bratcher [4] con-

sider the binomial case. The work presented in Bland and Bratcher [4]

will be considered in more detail in Chapter III.

Many techniques exist for the multiple comparisons problem in the normal case, each having advantages and disadvantages. Duncan [6] discusses some of these techniques and compares them to his decision-theoretic approach. There can be no doubt about the importance of the multiple comparisons problem for as Jackson [10] states:

Most statisticians now realize that an analysis of variance table is only an intermediate step and that, ultimately, some sort of multiple comparison technique must be employed before a final solution to the problem has been achieved.

Anscombe [1] asserts that for an important class of problems for which the multiple comparisons approach is applicable, the problem can and should be formulated in decision-theoretic terms.

There may possibly be some technical difficulties in the formulation, arising from the presence of many parameters, and, of course, there may be mathematical difficulties in the solution. But such difficulties should not blind us to the fact (if it is a fact) that the problem is decision-theoretic in principle; we should not pretend that some entirely different formulation is correct.

The author, fully realizing the controversy over both multiple comparisons and the decision-theoretic formulation, asserts that, if the multiple comparisons approach is desirable for comparing normal means, then a multiple comparisons approach (in decision-theoretic terms) is certainly appropriate for problems similar to the three examples pre-

sented in Section 1.2. The alternative procedures to the problem pre-
 sented in Section 1.3, while rivals of the present procedure in that
 they are concerned with comparing binomial probabilities, are essen-
 tially different in the approaches used and the decision which the
 experimenter may make at the conclusion of the experiment. Also, ex-
 periments of this kind are generally only conducted once, with any
 subsequent experiments being modified. Consequently, the experiment-
 er may prefer a procedure which attempts to minimize the average risk
 to one which guarantees a correct answer a certain percentage of the
 time. Of course, using the Bayesian criterion the minimum average
 risk is achieved exactly only if knowledge of the prior distribution
 is complete (i.e., the prior distribution is completely known).
 An interesting aspect of the problem being considered here is
 that only the functional form of the prior distribution need be spec-
 ified. The data can be used to estimate unspecified parameters. This
 is not completely in harmony with the subjective or personalistic inter-
 pretation of the prior distribution but it seems to be an interesting,
 if only partial answer, to the question of how the prior distribution
 is to be specified. More details on this aspect of the problem are
 given in Chapter V.

Let θ_0 be the set of order pairs (θ_i, θ_j) such that θ_i and θ_j are unranked
tain θ_i 's relation to the remaining $p-1$ probabilities for each $i=1,2,\dots,p$.
A natural step in seeking a solution would be to attempt to ascer-

for all the decisions.
decisions respectively). Furthermore, loss functions have to be defined
then two populations are to be compared (for $p = 2,3,4$ there are 3,19,183
a difficult task, for the number of possible decisions is large if more
this problem in decision-theoretic terms would at first seem to present
(a few results are obtained for sequential sampling). Just to formulate
size but some results are included for the case of unequal size samples
size. Most of the results obtained are for the case of a common sample
with the θ 's. The samples from the p Bernoulli processes are of fixed
from independent random samples from the Bernoulli processes associated
approach. The ranking of the θ 's is to be based on information gained
chosen in this paper is that which results from a multiple comparisons
the probabilities. As indicated in Section 1.4, the "ranking" to be
discussed in Section 1.2; that is, to choose some type of ranking of
the binomial probabilities $\theta_1, \theta_2, \dots, \theta_p$ in the sense of the third goal
The main problem of this work is to derive a procedure to compare

2.1 The Multiple Comparisons Problem

A DECISION THEORETIC FORMULATION

CHAPTER II

Thus a decision from (2.1.1) ranks θ_i greater than θ_j , ranks θ_j greater than θ_i or leaves θ_i and θ_j unranked.

There are $\frac{2}{p(p-1)}$ (to be denoted m) such pair-wise comparison problems which will be referred to as component problems. It is clear that the choice of a decision from each of the m component problems can be done in 3^m ways, thus considered simultaneously the m component problems define a multiple decision problem with 3^m decisions. Lehmann [11] denoted such a multiple decision problem the "product" of the component problems. Some of the 3^m decisions result from inconsistent component decisions (e.g., for $p=3$, d_{12}^{23} , d_{23}^{12} and d_{31}^{23} are inconsistent) and if these decisions are excluded we have what Lehmann termed the "restricted product" of the component problems. Thus the multiple comparison problem, as treated here, is the restricted product of m three-decision

$$w = \{ (\theta_i, \theta_j) \mid i=1, 2, \dots, p-1; j=2, 3, \dots, p; i < j \} \quad (2.1.2)$$

where

$$\begin{aligned} d_{1j}^1 : (\theta_i, \theta_j) \in \theta_0 \\ d_{1j}^2 : (\theta_i, \theta_j) \in \theta_1 \\ d_{1j}^3 : (\theta_i, \theta_j) \in \theta_2 \end{aligned} \quad (2.1.1)$$

(i.e. the whole parameter space), θ_1 be the set such that θ_i is larger than θ_j , and θ_2 be the set such that θ_j is larger than θ_i . Then one would choose one of the following decisions for each $(\theta_i, \theta_j) \in w$,

component problems and will be referred to as the comparison problem. The decisions of the comparison problem are thus defined only implicitly in terms of the decisions of the component problems (for $p=3$ the decisions are explicitly presented in Table 2.1, while for $p=2$ the decisions are just those in (2.1.1)). For $p=3$ (as in the medical research example), the decisions for the three component problems are given in (2.1.3), using the well established notation of Duncan:

$$\begin{array}{lll}
 d_{12}^1 = (1,2) & d_{13}^1 = (1,3) & d_{23}^1 = (2,3) \\
 d_{12}^2 = (2,1) & d_{13}^2 = (3,1) & d_{23}^2 = (3,2) \\
 d_{12}^3 = (1,2) & d_{13}^3 = (1,3) & d_{23}^3 = (2,3)
 \end{array}
 \quad (2.1.3)$$

Also for $p=3$, the 19 decisions of the comparison problem are given in Table 2.1 where the eight excluded decisions resulting from inconsistent component decisions have been indicated by 'dots'. The decisions of Table 2.1, and the comparison problem in general, rank the θ_i in ascending order from left to right but leave unranked the θ_i whose corresponding subscripts are underscored. For example, the decision $(\underline{3}, 1, 2)$, which is generated by d_{12}^1, d_{13}^1 and d_{23}^1 , asserts that θ_2 is larger than θ_3 but that the evidence is insufficient to order (θ_1, θ_2) and (θ_1, θ_3) . The loss for a particular decision in the comparison problem will be taken as the sum of the losses associated with the component decisions which generate that decision (termed "additive losses" by Lehmann). Consequently, only the losses for the component problems need be given ex-

TABLE 2.1

	d_{12}^1			d_{12}^2			d_{12}^3		
	d_{23}^1	d_{13}^1	d_{23}^2	d_{23}^3	d_{23}^1	d_{13}^2	d_{23}^1	d_{13}^3	d_{23}^3
$(\overline{1,2},3)$	$(\overline{1,2},3)$	$(\overline{3,1},2)$	$(\overline{2,1},3)$	$(\overline{3,2},1)$	$(\overline{3,1},2)$	$(\overline{2,1},3)$	$(\overline{2,3},1)$	$(\overline{1,2},3)$	$(\overline{1,2},3)$
d_{12}^1	$(\overline{1,2},3)$	$(\overline{3,1},2)$	$(\overline{2,1},3)$	$(\overline{3,2},1)$	$(\overline{3,1},2)$	$(\overline{2,1},3)$	$(\overline{2,3},1)$	$(\overline{1,2},3)$	$(\overline{1,2},3)$
d_{12}^2	$(\overline{2,3},1)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{2,3},1)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$
d_{12}^3	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$	$(\overline{1,3},2)$

plicity. The main justification for our considering the multiple comparisons problem as the restricted product of the component problems is given in Section 2.3. Briefly this is that, with additive losses, compatible Bayes solutions to the component problems yield a Bayes solution for the restricted product problem [where solutions for the component problems are said to be compatible if, when considered simultaneously, they never (i.e. with probability one) lead to inconsistent decisions, that is, they only lead to decisions in the restricted product problem].

2.2 Additional Formulation

In this section the structure of the component problems is given

along with other terminology.

Often the experimenter has prior information about the θ 's, this may be subjective feelings about the θ 's or past data giving information about the θ 's. This type of information is assumed to be such that it may be converted into a function with the mathematical properties of a distribution function, called the prior distribution of θ . Consequently, the θ 's will be treated as random variables in the following development. Also, the prior information is assumed to be such that the θ 's can be treated as independent, not necessarily identically distributed, random variables. For computational purposes the prior distribution for θ_i is assumed to be a member of the beta family with density function:

$$\xi_i(\theta_i) = [b(a_i, b_i)]^{-1} \theta_i^{a_i-1} (1-\theta_i)^{b_i-1}, \quad 0 \leq \theta_i \leq 1, \quad a_i, b_i > 0. \quad (2.2.1)$$

This family, which is conjugate to the binomial distribution (see Raftery and Schlaifer [12]), is generally considered to be rich enough to approximate most useful continuous distributions on the unit interval.

It is well known that solutions to decision problems need be based only on sufficient statistics so the principle of sufficiency is used to reduce the data from the sequence of successes and failures observed to the number of successes observed. Hence, the data utilized to help select a decision from (2.1.1) are the independent binomial variables X_1, X_j with parameters (θ_1, n_1) and (θ_j, n_j) , respectively, with probability functions:

$$f(x|\theta_k) = \binom{n_k}{x} \theta_k^x (1-\theta_k)^{n_k-x}, \quad x = 0, 1, \dots, n_k, \quad k = 1, j. \quad (2.2.2)$$

The losses associated with decision d_{ij}^k and $\bar{\theta} = (\theta_1, \theta_j)$ is denoted by $L_{ij}^k(\bar{\theta})$ and the form of this loss function is given by (2.2.3).

$$(2.2.3) \quad \begin{aligned} L_{1j}^1(\bar{\theta}) &= 0, & \theta_1 &= \theta_j, & c_1 \ell(\theta_1, \theta_j) &= c_1 \theta_1 \neq \theta_j, \\ L_{1j}^2(\bar{\theta}) &= 0, & \theta_1 &> \theta_j, & c_2 \ell(\theta_1, \theta_j) &= c_2 \theta_1 \leq \theta_j, \\ L_{1j}^3(\bar{\theta}) &= 0, & \theta_1 &> \theta_j, & c_2 \ell(\theta_1, \theta_j) &= c_2 \theta_1 \geq \theta_j, \end{aligned}$$

where ℓ is a non-negative monotonically non-decreasing function of

$|\theta_1 - \theta_j|, \ell(\theta, \theta) = 0$ and c_1 and c_2 are constants such that $c_2 \geq 2c_1 > 0$.

The non-decreasing nature of ℓ is to reflect the increasing seriousness of the error as $|\theta_1 - \theta_j|$ increases. Also, when incorrect, decisions d_2

and d_3 are more serious than decision d_1 so c_2 should be larger than c_1 . The reason for c_2 to be at least twice c_1 will be apparent later (see

following (3.1.6)). Finally, the loss functions for each component problem are the same, that is, ℓ, c_1 and c_2 are independent of i and j .

A non-randomized decision rule is a function which maps the sample space into the decision space and, for a k -decision problem, has the

following form:

$$\delta(\bar{x}) = [\delta_1(\bar{x}), \delta_2(\bar{x}), \dots, \delta_k(\bar{x})], \quad (2.2.4)$$

where

$$\delta_i(\bar{x}) = \begin{cases} 1, & \text{if decision } d_i \text{ is to be made,} \\ 0, & \text{if decision } d_i \text{ is not to be made.} \end{cases}$$

Thus, a non-randomized decision rule is a function whose values are k -dimensional vectors with $(k-1)$ components equal to zero and one component equal to one for all $\bar{x}$. The i th component function δ_i will be referred to as the indicator function for the i th decision.

As is well known, for finite decision problems non-randomized Bayes rules always exist, therefore there is no loss of generality in searching for Bayes rules in restricting ourselves to non-randomized decision rules.

2.3 The Bayes Risk

In this section we show that, with additive losses, compatible Bayes decision rules for the component problems generate a Bayes decision rule for the comparison problem.

Let k be the number of decisions in the comparison problem and as before $m = \binom{2}{k}$ is the number of component problems. Then,

$$(2.3.1) \quad \delta^*_i(\bar{x}) = \prod_{n=1}^m \delta^n_i(\bar{x}), \quad i=1,2,\dots,k,$$

gives a decision rule for the comparison problem where δ^n_i is an indicator function for decision d^n_i of the n th component problem and decision d_i is the "product" of the component decisions $d^n_1, d^n_2, \dots, d^n_m$. Of course, $\prod_{n=1}^m \delta^n_i = 0$ for every set of inconsistent decisions; and since there are 3^m decisions in the product decision problem and only k decisions in the component decision problems (the restricted product decision problem), there are $3^m - k$ sets of inconsistent decisions.

Hence the above product must be zero for $3^m - k$ sequences $\left\{ d^n_i \right\}_{n=1}^m$. The question at this point is whether or not, with the assumptions made in Section 2.1 and 2.2, the decision rule given by equation (2.3.1) is Bayes when the component decision rules are Bayes and compatible. The risk for δ^* and $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ is given by:

$$(2.3.2) \quad R(\delta^*, \bar{\theta}) = \sum_{k=1}^m \int_{\delta^*_k} L_k(\bar{\theta}) dF(\bar{x}|\bar{\theta}) = \sum_{k=1}^m \int_{\delta^n_k} L_k(\bar{\theta}) dF(\bar{x}|\bar{\theta}) = \int_{\delta^n_k} L_k(\bar{\theta}) dF(\bar{x}|\bar{\theta}).$$

rules for the generating problems yield decision rules for the three-

decision problems. Let the loss functions for the decisions in equation

(3.1.3) be given by:

$$\begin{aligned}
 L_1^1(\theta_1, \theta_2) &= 0, & \theta_1 \leq \theta_2, \\
 &= c_1 \ell(\theta_1, \theta_2), & \theta_1 > \theta_2, \\
 L_2^1(\theta_1, \theta_2) &= 0, & \theta_1 > \theta_2, \\
 &= c_2^* \ell(\theta_1, \theta_2), & \theta_1 \leq \theta_2,
 \end{aligned}
 \tag{3.1.5}$$

and

$$\begin{aligned}
 L_2^2(\theta_1, \theta_2) &= 0, & \theta_1 \leq \theta_2, \\
 &= c_1 \ell(\theta_1, \theta_2), & \theta_1 > \theta_2, \\
 L_1^2(\theta_1, \theta_2) &= 0, & \theta_1 > \theta_2, \\
 &= c_2^* \ell(\theta_1, \theta_2), & \theta_1 \leq \theta_2,
 \end{aligned}
 \tag{3.1.6}$$

where $c_2^* = c_2 - c_1$.

Since an incorrect ranking is a more serious error than failure to

rank, we have the natural restriction that $c_2^* > c_1$ which implies the re-

striction stated earlier that $c_2 > 2c_1$. This restriction is sufficient

to insure compatibility in the equal sample size, common prior case.

Expressing the ranking decisions (3.1.1) as the product of the de-

cisions (3.1.3) and with additive losses, we have from equations (3.1.4),

(3.1.5) and (3.1.6) that :

$$\begin{aligned}
 L_1(\theta_1, \theta_2) &= L_1^1(\theta_1, \theta_2) + L_2^1(\theta_1, \theta_2) \\
 L_2(\theta_1, \theta_2) &= L_1^2(\theta_1, \theta_2) + L_2^2(\theta_1, \theta_2) \\
 L_3(\theta_1, \theta_2) &= L_1^3(\theta_1, \theta_2) + L_2^3(\theta_1, \theta_2)
 \end{aligned}$$

In detail,

$$\begin{aligned}
 L_1(\theta_1, \theta_2) &= 0 + 0 \\
 &= c_1^1 \chi(\theta_1, \theta_2) + 0 \\
 &= c_1^1 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 < \theta_2 \\
 &= c_1^1 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 > \theta_2
 \end{aligned}$$

$$\begin{aligned}
 L_2(\theta_1, \theta_2) &= 0 + 0 \\
 &= c_2^2 \chi(\theta_1, \theta_2) + 0 \\
 &= c_2^2 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 = \theta_2 \\
 &= c_2^2 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 > \theta_2
 \end{aligned}$$

$$\begin{aligned}
 L_3(\theta_1, \theta_2) &= 0 + 0 \\
 &= 0 + c_2^2 \chi(\theta_1, \theta_2) \\
 &= c_2^2 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 = \theta_2 \\
 &= c_2^2 \chi(\theta_1, \theta_2) + c_1^2 \chi(\theta_1, \theta_2) \\
 &= c_2^2 \chi(\theta_1, \theta_2) \quad , \quad \theta_1 > \theta_2
 \end{aligned}$$

The above, indeed, gives the losses in (3.1.2). Thus the losses for the three-decision component problems are additive in terms of the losses for the two-decision generating problems.

Just as the comparison problem was formulated as the restricted product of three-decision component problems with additive losses, each component problem is expressed as the restricted product of two-decision generating problems with additive losses. Also the analysis of Section 2.3 easily carries over so that with additive losses a Bayes rule for a three-

decision component problem is obtained from compatible Bayes rules for

the two generating two-decision problems. For the case of a common

sample size and common prior in the comparison problem, compatibility

of the two-decision problems yields compatibility of the three-decision

component problems. Hence in this important special case a Bayes rule

for one two-decision generating problem yields a Bayes rule for the com-

parison problem, since the two-decision generating problems are the same

except for labeling. In all cases, a set of compatible Bayes rules for

the two-decision generating problems yields a Bayes rule for the three-

decision component problems, and, in turn, if they are compatible it yields

a Bayes rule for the comparison problem.

3.2 A solution to the Generating Problems

We turn now to the problem of finding Bayes rules for the two-decision

generating problems. Considering the first generating problem of (3.1.3)

with losses (3.1.5), the Bayes risk is:

$$r_1(\xi, \delta_1) = \int_2^{\theta} \sum_{i=1}^{\theta} \int_1^{\theta} \delta_1^i(\theta) dF(\bar{x}|\theta) d\xi(\bar{\theta}) \cdot \quad (3.2.1)$$

Since $\delta_1^i(x_1, x_2) = 1 - \delta_2^i(x_1, x_2)$ for all $x_i = 0, 1, \dots, n_i$, $(i=1, 2)$, (3.2.1) be-

comes:

$$r_1(\xi, \delta_1) = \int_1^{\theta} \int_1^{\theta} \delta_1^i(\theta) dF(\bar{x}|\theta) d\xi(\bar{\theta}) + \left\{ \int_1^{\theta} \int_2^{\theta} [\delta_1^i(\bar{\theta}) - \delta_2^i(\bar{\theta})] d\xi(\bar{\theta}) dF(\bar{x}) \right\} \delta_1^2(\bar{x}) dF(\bar{x}),$$

or

$$(3.2.2) \quad r_1(\xi, \delta_1) = \text{constant} + \int_X h(\bar{x}) \delta_1^2 \delta_2^2(\bar{x}) dF(\bar{x}),$$

where

$$(3.2.3) \quad h(\bar{x}) = E_{\xi} \{ [L_2(\bar{\theta}) - L_1(\bar{\theta})] | \bar{x} \}.$$

Clearly from (3.2.2), the Bayes risk, $r_1(\xi, \delta_1)$ is a minimum if

$$(3.2.4) \quad \delta_1^2(x_1, x_2) = \begin{cases} (1, 0) & , \text{ if } h(x_1, x_2) \geq 0 \\ (0, 1) & , \text{ if } h(x_1, x_2) < 0. \end{cases}$$

Thus a Bayes solution depends upon the set of $\bar{x}$'s where h changes signs.

Examining h in detail, we see that:

$$(3.2.5) \quad L_2(\bar{\theta}) - L_1(\bar{\theta}) = \begin{cases} (c_2 - c_1) \mathcal{L}(\theta_1, \theta_2) & , \text{ if } \theta_1 \leq \theta_2 \\ c_1 \mathcal{L}(\theta_1, \theta_2) & , \text{ if } \theta_1 > \theta_2, \end{cases}$$

and

$$(3.2.6) \quad \xi'(\bar{\theta} | \bar{x}) = \frac{a_1 + x_1 - 1}{b_1 + n_1 - x_1 - 1} \frac{a_1^{x_1} (1 - \theta_1)^{1 - x_1}}{a_2 + x_2 - 1} \frac{b_2^{x_2} (1 - \theta_2)^{1 - x_2}}{b_2^{n_2 - x_2 - 1}} \cdot$$

Using the fact that

$$\int_1^0 \int_1^0 g(\theta_1, \theta_2) d\theta_1 d\theta_2 = \int_1^0 \int_1^0 g(\theta_1, \theta_2) d\theta_1 d\theta_2 + \int_1^0 \int_1^0 g(\theta_1, \theta_2) d\theta_1 d\theta_2,$$

we may write $h(x_1, x_2)$ explicitly for the case of linear loss $(\ell(\theta_1, \theta_2) = |\theta_1 - \theta_2|)$ as being proportional to

$$h^*(x_1, x_2) = (c-1) \int_1^0 \theta^2 I_{\theta} (a_1 + x_1, b_1 + n_1 - x_1) d\theta \quad (3.2.7)$$

$$- \frac{a_1 + x_1}{a_1 + b_1 + n_1} I_{\theta} (a_1 + x_1 + 1, b_1 + n_1 - x_1) d\theta.$$

$$\frac{\theta^{a_2 + x_2 - 1} (1 - \theta)^{b_2 + n_2 - x_2} \beta(a_2 + x_2, b_2 + n_2 - x_2)}{d\theta}$$

$$+ \left[\frac{x_2 + a_2}{a_2 + b_2 + n_2} - \frac{x_1 + a_1}{a_1 + b_1 + n_1} \right]$$

where

$$I_{\theta}^*(u, v) = \int_0^1 \frac{\theta^{u-1} (1-\theta)^{v-1} \beta(u, v)}{d\theta}$$

and $c = \frac{c_1}{c_2} - 1 (= c_2^*/c_1)$. Hence, to find the set of (x_1^*, x_2^*) where h^* changes sign, only the loss ratio c_2/c_1 (or equivalently, c_2^*/c_1) needs to be specified rather than both c_1 and c_2 .

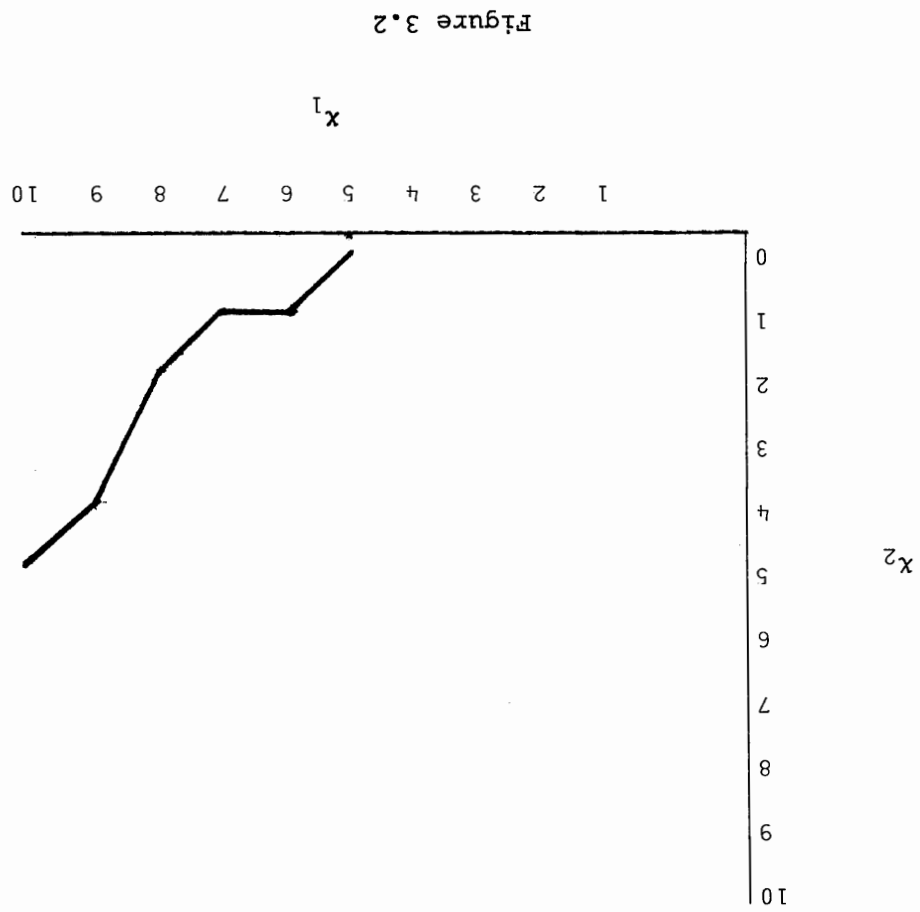
Clearly the Bayes rule for the second generating problem of (3.1.3) with losses (3.1.6) may be found by interchanging x_1 and x_2 , n_1 and n_2 , and (a_1, b_1) and (a_2, b_2) in the Bayes rule for the first generating problem. Let us now consider in some detail the important case of common sam-

ple size and common prior (i.e. $n_1 = n$ and $(a_1, b_1) = (a, b)$, $i = 1, 2, \dots, p$).

As an illustration of the Bayes rule for the first generating problem, consider Figure 3.2 for the case of $n=10$, $(a,b)=(1,1)$ and a loss ratio of $c=500$. If the data (x_1, x_2) falls on, below or to the right of the broken line (termed the critical difference line) the Bayes rule is to assert $\theta_1 > \theta_2$, otherwise θ_1 and θ_2 are unranked. The Bayes rule for the second generating problem is obtained by reflecting the critical difference line of Figure 3.2 about the line $x_1 = x_2$ or equivalently by interchanging the x_1 and x_2 axes (this is all that is necessary here since $n_1 = n_2$ and $(a_1, b_1) = (a_2, b_2)$). The rules for the generating problems are then seen to be compatible if the critical difference line of Figure 3.2 is below the line $x_1 = x_2$, since then the reflected critical difference line will not intersect the original critical difference line (otherwise there would be the possibility of inconsistent decisions). It is clear here that the rules are compatible and a condition to insure this in general is for the loss ratio to be greater than one ($c > 1$). Moreover, the compatibility of the rules for the generating problems leads to compatible rules for the component problems. An alternate way to carry out both rules simultaneously is to first order the x 's and use the critical difference line of Figure 3.2 with the larger x taking the role of x_1 . Another way to present this alternate way is by Table 3.2 (x_1 is the larger and x_2 the smaller observation). The table gives the difference $x_1 - x_2$ for each x_1 necessary to choose decision d_1^2 (asserting $\theta_1 > \theta_2$). For example, if $x_1 = 3$, the θ 's remain unranked since the difference cannot be large enough to assert that θ_1 is larger. If $x_1 = 8$, then d_1^2 is chosen if x_1 is such that $x_1 - x_2 \geq 6$; that is, if $x_2 = 0, 1$ or 2 .

larger of x_1 and x_2	larger minus smaller
0	10
1	9
2	8
3	7
4	6
5	5
6	4
7	3
8	2
9	1
10	0

Table 3.2



From Figure 3.2 we see that of the 66 possible data points, we assert $\theta_1 > \theta_2$ for only 19 points. The large loss ratio $c=500$ accounts for this. The number of points such that ranking occurs, called ranking points, decreases as c increases, as one would expect since c is a measure of how much more serious it is to incorrectly rank a pair of θ 's than to leave them unranked. Also the number of ranking points decreases as the prior information increases (i.e. variance of the prior distribution decreases). However, as the mean of the prior approaches an end point, 0 or 1, the number of ranking points increases.

Let us now continue the illustration for the comparison problem. For ease of exposition here, suppose the processes are originally labeled $1', 2', \dots, p'$ and after observing the data $(x_1, x_2, \dots, x_p)$ the processes are relabeled or renumbered so the observations are ordered. Thus with the relabeling, process 1 (with parameter θ_1) has the smallest observation x_1 , and so on, to process p (with parameter θ_p) which has the largest observation x_p . (This is the usual way of ordering observations but opposite of that used just earlier here.) It doesn't matter how tied observations are treated just so long as their order with the remaining observations is preserved. First, compare x_p with $x_{p-1}, x_{p-2}, \dots, x_1$ in order using critical values, for a specified n , (a, b) and c (like those in Table 3.2). Let x_{k_1} be the smallest of these observations such that the decision to leave θ_p and θ_{k_1} unranked is to be made. If the decision is to be made that $\theta_p > \theta_{p-1}$ let $k_1=p$. If $k_1=1$, then the analysis is over. If $k_1 > 1$, after the comparison of x_p and x_{k_1} , no more comparisons of x_p

with the x 's smaller than $x_{k_1}^p$ need be made since θ^p will be ranked above the corresponding θ 's. To summarize these decisions take the vector of the subscripts of the θ 's, $(1, 2, \dots, p)$, and draw a line under p through k_1 to indicate the corresponding θ 's are unranked. No line need be drawn if $k_1 = p$. If $k_1 > 1$ then compare x_{p-1}^p with $x_{p-2}^p, \dots, x_1^p$ in order and let x_k^p be the smallest of these observations such that the decision is to leave θ_{k-1}^p and θ_k^p unranked is to be made, where if the decision is to be made that $\theta_{p-1}^p > \theta_{p-2}^p$ let $k_2 = p-1$. If $k_2 = 1$, then the analysis is over. If $k_2 > 1$ then, after the comparison of x_{p-1}^p and $x_{k_2}^p$, no more comparisons of x_{p-1}^p with the smaller x 's need be made since θ_{p-1}^p will be ranked above the corresponding θ 's. To summarize these decisions, if $k_2 = k_1$ no new line need be drawn under the vector $(1, 2, \dots, p)$, if $k_2 < k_1$ then draw a line, below the line drawn before, under $p-1$ through k_2 to indicate the corresponding θ 's are unranked. (Note that the case $k_2 > k_1$ cannot occur because of the 'one-sided' nature of the comparisons being made and the fact that $x_{p-1}^p < x_{p-2}^p$.) If $k_2 > 1$ then continue the comparisons with x_{p-2}^p and so on until no more comparisons can be made (either the x_i^p 's are exhausted or the corresponding θ 's have been underlined in previous comparisons). For the case of $c > 1$ the comparisons in the reverse order, starting with x_1 and comparing it with $x_2, \dots, x_p$ need not be made since if when comparing x_i and x_j ($x_i > x_j$) the decision is made to **assert** $\theta_i > \theta_j$ then when comparing x_j with x_i the decision will be made to leave θ_j and θ_i unranked (because for $c > 1$ the decision boundary is to the right of the line $x_i = x_j$, see Figures 3.2 and 3.3). Finally in the underlined vector

(1,2,...,p) replace θ_i by θ_i^* , (the corresponding θ in the original notation). The comparison decision is then, that the θ 's underlined are unranked while those not underlined are ranked in the order listed. To illustrate the above let us suppose that $n=10$, $(a,b)=(1,1)$, $c=500$ and the observations $(3,7,4,9,1)$. Ordered the observations are $(1,3,4,7,9)$ and comparing the largest, 9, with the smaller x 's, from Table 3.2 we see that a difference of 5 or more is necessary to assert $\theta_5 > \theta_1$. Therefore $k_1=4$ and 5 and 4 are underlined in the vector $(1,2,3,4,5)$. Next comparing $x_4=7$ with the smaller x 's we see from Table 3.2 that a difference of 6 or more is necessary to assert $\theta_4 > \theta_1$. Therefore for $k_2=2$ and 4, 3 and 2 are underlined (below the first line). Next comparing $x_3=4$ with the smaller x 's we see from Table 3.2 that no difference is large enough to rank the θ 's. Therefore $k_3=1$; 3, 2 and 1 are underlined and, since 2 has been underlined with 1, x_2 need not be compared with x_1 . Therefore the final decision is given by $(1,2,3,4,5)$ or in the original notation $(\theta_5, \theta_1, \theta_3, \theta_2, \theta_4)$. (It is clear that the relabeling is unnecessary and has been done here only to ease the notation.) Thus the final decision asserts that θ_4 and θ_2 are unranked but θ_4 is larger than θ_3 , θ_1 and θ_5 ; θ_2 , θ_3 and θ_1 are unranked but θ_2 is larger than θ_5 while θ_3 , θ_1 and θ_5 are unranked.

For the general case (not necessarily common sample size and common prior) a similar analysis can be made. It should be noted that in making the comparisons different critical differences will be necessary (since the critical difference depends on the sample sizes and parameters of the

priors, as well as the loss ratio). The question of compatibility is more difficult to treat in this case and will not be dealt with here except to comment that if the procedure outlined for the common sample size, common prior case is followed (insisting that $k_1 \geq k_2 \geq \dots$) then no inconsistent decisions will be made. However the procedure is then only approximately a Bayes rule.

3.3 An Alternate Rule

Bland and Bratcher [4] felt that little information concerning the difference between the θ 's ($\theta_1 - \theta_j$) would be lost by considering only those decision rules based on the difference between the observations ($x_i - x_j$). A restricted Bayes rule (restricted to depend only on the differences $x_i - x_j$) was found for the comparison problem and critical differences were tabulated for several sets of parameters (for the case of common sample size and common prior).

The above restricted Bayes rule (denoted δ^r) and the Bayes rule (denoted δ^*) in (3.2.4) are compared by means of the Bayes risks for the generating problems. To evaluate the Bayes risk for a rule δ for the first generating problem (comparing θ_1 and θ_2) let

$$S_1(\delta) = \{ (x_1, x_2) : \delta_1(\bar{x}) = 1 \}$$

and

$$S_2(\delta) = \{ (x_1, x_2) : \delta_2(\bar{x}) = 1 \},$$

then from (3.2.1) the Bayes risk is given by

$$(3.3.1) \quad r(\xi, \delta) = c_1 \int_0^{S_1(\delta)} [\theta_1 I_{\theta_1} (a+x_1, b+n-x_1) - \frac{a+b+n}{a+x_2} I_{\theta_1} (a+x_2+1, b+n-x_2)]$$

$$\cdot \binom{n}{x_1} \binom{n}{x_2} \frac{\beta(a+x_2, b+n-x_2)}{\beta(a+x_1-1, b+n-x_1-1)} \frac{\beta(a, b)}{\theta_1 (1-\theta_1)} d\theta_1$$

$$+ c_2 \int_0^{S_2(\delta)} [\theta_2 I_{\theta_2} (a+x_1, b+n-x_1) - \frac{a+b+n}{a+x_1} I_{\theta_2} (a+x_1+1, b+n-x_1)]$$

$$\cdot \binom{n}{x_1} \binom{n}{x_2} \frac{\beta(a+x_1, b+n-x_1)}{\beta(a+x_2-1, b+n-x_2-1)} \frac{\beta(a, b)}{\theta_2 (1-\theta_2)} d\theta_2.$$

The comparisons were made by means of $[r(\xi, \delta^t) - r(\xi, \delta^*)] / r(\xi, \delta^*)$, the relative reduction in the Bayes risk. For small sample sizes the two decision rules generally differed little for the cases considered and the relative reduction in the Bayes risk was unimportant in all cases (usually less than 1% of the Bayes risk). Tables 3.3.1 and 3.3.2 give some typical results. For example, with the loss ratio $c=100$, the common sample size $n=10$ and the prior parameters $(a, b)=(3, 12)$, δ^t says to choose the decision $\theta_1 > \theta_2$ only if $t=x_1-x_2 \geq 6$, so that θ_1 and θ_2 are unranked if $x_1=5$. Now δ^* differs from δ^t only for the one sample point $(x_1, x_2)=(5, 0)$, here the Bayes rule is to assert $\theta_1 > \theta_2$. The relative difference

in the Bayes risk is less than 0.005. Table 3.3.2 presents an example for which the two decision rules differ at 10 of 441, or about 2%, of the data points (see Figure 3.3). The relative difference in the Bayes risk is just over 3%.

Computationally, the Bayes rule δ^* is no more difficult than δ^t . The evidence indicates that little will be lost by using the tables of Bland and Bratcher [4]; however, if new values are to be computed, the Bayes rule δ^* would be preferred since it does have the smaller Bayes risk with little or no increase in computational difficulty. The only disadvantage of δ^* is with tabling the critical values, with δ^t only one critical difference is necessary (for a set of parameters) while with δ^* several (n) critical pairs are necessary.

TABLE 3.3.1

CRITICAL VALUES OF THE BAYES RULES δ^* AND δ^t

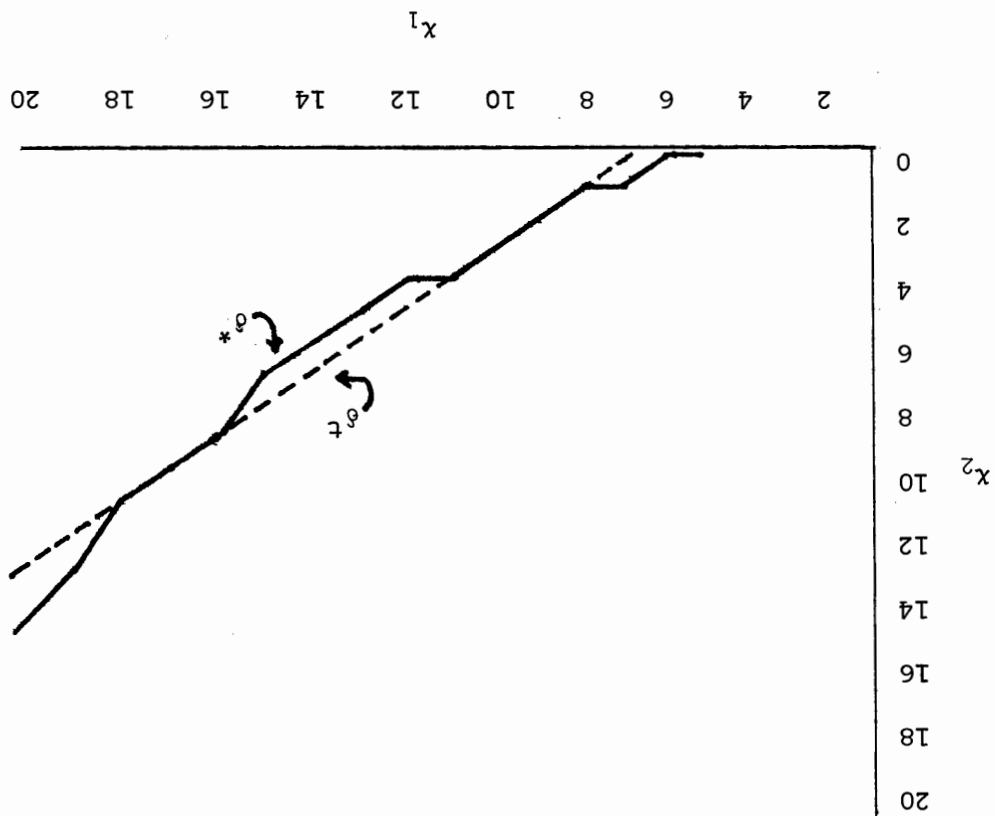
$c=50$		$c=100$		$c=500$	
(a,b)	x_1	$(3,12)$		$(1,1)$	
		$x_1 - x_2$	$x_1 - x_2$	$x_1 - x_2$	$x_1 - x_2$
$n=5$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	5	3	4	4
	6	4	3	4	4
	7	.	3	.	.
	8	.	.	.	.
	9	.	.	.	.
$n=10$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	.	5	5	5
	6	6	6	6	6
	7	7	6	6	6
	8	7	6	6	6
	9	7	6	6	6
$n=10$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	.	5	5	5
	6	6	6	6	6
	7	7	6	6	6
	8	7	6	6	6
	9	7	6	6	6
$n=10$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	.	5	5	5
	6	6	6	6	6
	7	7	6	6	6
	8	7	6	6	6
	9	7	6	6	6
$n=10$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	.	5	5	5
	6	6	6	6	6
	7	7	6	6	6
	8	7	6	6	6
	9	7	6	6	6
$n=10$	0	.	.	.	.
	1	.	.	.	.
	2	.	.	.	.
	3	.	.	.	.
	4	.	.	.	.
	5	.	5	5	5
	6	6	6	6	6
	7	7	6	6	6
	8	7	6	6	6
	9	7	6	6	6

TABLE 3.3.2
CRITICAL VALUES OF THE BAYES RULES δ^* AND δ^t

$n=20, c=500$			
x_1		$(a,b)=(1.1)$	x_1-x_2
		$t=7$	
0		5	•
1		•	•
2		•	•
3		•	•
4		•	•
5		6	•
6		6	•
7		6	•
8		7	•
9		7	•
10		7	•
11		7	•
12		8	•
13		8	•
14		8	•
15		8	•
16		8	•
17		8	•
18		8	•
19		8	•
20		7	•

FIGURE 3.3

COMPARISON OF THE BAYES RULES δ^* AND δ^t
($n=20, (a,b)=(1.1), c=500$)



is to be made based on sequential data from each of two Bernoulli processes with probabilities of success θ_1 and θ_2 , respectively. This problem is also important in that there are many direct applications for it. Similar problems are found in the literature (see Ray [13] for a detailed discussion of many of these papers). Nevertheless, there seem to be no straight forward sequential Bayes procedures available for comparing two binomial probabilities, θ_1 and θ_2 , with losses in terms of θ_1 and θ_2 . The problem here, as in the earlier chapters, is to find a Bayes decision rule with linear losses, except that now the sampling is done sequentially.

$$\begin{aligned} d_1: & \theta_1 \leq \theta_2, \\ d_2: & \theta_1 > \theta_2, \end{aligned} \quad (4.1.1)$$

The topics considered in this chapter are what are hoped to be first steps in leading to a solution for the problem of ranking two probabilities (as defined in (3.1.1) and (3.1.2)) where the observations are taken sequentially in pairs. The problem considered is similar to the generating two-decision problem discussed in Chapter III. The selection of one of the two decisions:

4.1 Introduction

SEQUENTIAL PROBLEMS

CHAPTER IV

quentially. Often, in the sequential situation, a minimum average risk decision rule is quite difficult to obtain directly. Consequently, the possible decision rules are sometimes restricted to include only those rules which allow no more than, say, n' observations. The minimum average risk decision rule for this restricted problem may be found, and

under the proper conditions the minimum Bayes risk for the original non-truncated problem is approached as n' becomes large. In fact, a Bayes rule for a sequential decision problem quite often turns out to be a truncated rule. A truncated decision rule forces a decision before more

than n^* observations are taken, where n^* is said to be the point of truncation. Indeed, the Bayes rule is a truncated rule for the case with

symmetric linear loss as considered in Section 4.4. At this point the

problem becomes that of finding an upper bound, say n' , for the point of

truncation, n^* , given the loss function and prior distributions. Now if

it is known that a Bayes rule is a truncated decision rule and n' is known

the Bayes rule may be found according to Section 4.3. In order to discuss

the method we must have the terminology of Section 4.2.

4.2 Terminology

A sequential decision rule is defined as a pair (ϕ, δ) such that ϕ is

a stopping rule and δ is a terminal decision rule. The Bayes terminal de-

cision rule is independent of the stopping rule and hence is the same as

a fixed sample size decision rule. The stopping rule ϕ is a sequence of

functions, $\phi(\bar{x}) = \{\phi_0, \phi_1(\bar{x}_1), \phi_2(\bar{x}_2), \dots\}$ where $\phi_k(\bar{x}_k)$ is the conditional

probability that no additional observations will be taken, given that the

first k observations have led to the sufficient statistic $X_k = (X_k^{(1)}, X_k^{(2)})$ and $X_k^{(1)}$ has the binomial distribution with parameters (k, θ_1) . The risk function is the expected value of the loss plus cost when θ_1, θ_2 are the

binomial probabilities. The Bayes risk, $r(\xi, (\phi, \delta))$, then is the expected value of the risk function with respect to the prior beta distribution, ξ . Let $p(\xi) = \inf \{r(\xi, (\phi, \delta))\}$ be the minimum Bayes risk. The posterior distribution, based on X_k , is denoted by ξ_k and the class of sequential decision rules based on at most the first n observations is denoted by Δ_n .

The Bayes risk in the class Δ_n , for which (ϕ, δ) takes at most n observations, is denoted by $p_n(\xi) = \inf_{\Delta_n} r(\xi, (\phi, \delta))$, thus $p_0(\xi) =$

$\inf_{\Delta} E_{\xi}[L_1(\bar{\theta})], E_{\xi}[L_2(\bar{\theta})]$. Also, for any distribution, ξ , of $\bar{\theta}$ let $\lambda(\xi) = p_0(\xi) - E_{\xi}[p_0(\xi_{\bar{X}})]$, where $\bar{X} = (X_1, X_2)$ represents a pair of independent Bernoulli random variables with probabilities of success θ_1 and θ_2 , respectively. The function L is defined as $L(\bar{\theta}) = L_1(\bar{\theta}) - L_2(\bar{\theta})$.

Theorem 4.2

In order that

$$\lim_{n \rightarrow \infty} p_n(\xi) = p(\xi) \quad (4.2.1)$$

it is sufficient that the loss is bounded.

This theorem is due to Hoeffding [9] and will be needed in later

sections.

4.3 Truncated Sequential Decision Rules

A sequential decision problem is said to be truncated at n if the

sampling must stop with probability 1 by the time that n' observations are taken. A Bayes stopping rule for a truncated sequential decision problem may be found using backward induction and may also be Bayes for the nontruncated problem. Theorem 4.2 and Theorem 4.4 (yet to be stated) provide tools to help show whether a Bayes rule for a truncated sequential decision problem is also Bayes for the associated nontruncated decision problem.

The minimum conditional expected loss plus cost after observing $\bar{x}_j$ is denoted for our problem by

$$U_j(\bar{x}_j) = p_0(\bar{x}_j) + c.$$

If $\bar{x}_{n'-1}$ has been observed and sampling stops before the next observation, $\bar{x}$ is observed, the loss plus cost is $U_{n'-1}(\bar{x}_{n'-1})$. If the last observation is taken, the expected loss plus cost is $E[U_{n'}(\bar{x}_{n'-1} + \bar{x})]$. Thus sampling should stop at $(n'-1)$ if the former is not greater than the latter and continue otherwise. Hence, $\phi_{n'-1}^*$ of a Bayes stopping rule, truncated at n' , is given by

$$\phi_{n'-1}^*(\bar{x}_{n'-1}) = \begin{cases} 1, & U_{n'-1}(\bar{x}_{n'-1}) < E[U_{n'}(\bar{x}_{n'-1} + \bar{x})] \\ 0, & \text{otherwise.} \end{cases}$$

The minimum Bayes risk, given $\bar{x}_{n'-1} = \bar{x}_{n'-1}$, is

$$V_{n'-1}(\bar{x}_{n'-1}) = \min_{U_{n'-1}} \{ U_{n'-1}(\bar{x}_{n'-1}) + E[U_{n'}(\bar{x}_{n'-1} + \bar{x})] \}.$$

Now if $\bar{X}_{n,-2}$ has been observed, sampling would stop if

$$U_{n,-2}(\bar{X}_{n,-2}) \leq E[V_{n,-1}^{n'}(\bar{X}_{n,-2} + \bar{X})].$$

This procedure can be continued. Inductively, the minimum Bayes risk is

given by

$$V_{n'}^{j-1}(\bar{X}_{j-1}^{n'}) = \min(U_{j-1}^{n'}(\bar{X}_{j-1}^{n'}), E[V_{n'}^j(\bar{X}_{j-1}^{n'} + \bar{X})]) \quad j=1, 2, \dots, n'$$

with $V_{n'}^{n'} = U_{n'}$ and $V_{n'}^0 = p_{n'}$. (defined in Section 4.2). Thus we have a

truncated Bayes stopping rule, with truncation at n' , given by

$$\phi^* = (\phi_0^*, \phi_1^*, \dots, \phi_{n'}^*) \text{ where}$$

$$\phi_{j-1}^* = \begin{cases} 1, & U_{j-1}^{n'}(\bar{X}_{j-1}^{n'}) \leq E[V_{n'}^j(\bar{X}_{j-1}^{n'} + \bar{X})] \\ 0, & \end{cases}$$

$j=1, 2, \dots, n'$ and of course $\phi_{n'}^*(\bar{X}_{n'}) = 1$ for all $\bar{X}_{n'}$.

Hence, if an upper bound is known for the point of truncation of a

Bayes sequential decision rule it may be possible to program the backward

induction method according to the above criterion.

4.4 Symmetric Linear Loss

Consider the sequential problem of choosing a decision from (4.1.1)

with constant cost per observation, $c=1$, and losses given by

$$L_1(\bar{\theta}) = \begin{cases} 0 & \theta_1 \leq \theta_2 \\ k(\theta_1 - \theta_2) & \theta_1 > \theta_2 \end{cases} \quad , \quad \theta_1 > \theta_2$$

$$L_2(\bar{\theta}) = \begin{cases} k(\theta_2 - \theta_1) & \theta_1 \leq \theta_2 \\ 0 & \theta_1 > \theta_2 \end{cases} \quad , \quad \theta_1 > \theta_2$$

(4.4.1)

We need the following theorem and lemma.

Theorem 4.4

With the assumption that equation (4.2.1) holds, there is a Bayes sequential procedure with respect to ξ that is truncated if there exists an integer n' such that

$$\lambda(\xi_{\bar{X}_n}) \leq E_{\xi} (c) = 1 \quad \text{a.s.} \quad (4.4.2)$$

for every $n \geq n'$.

Lemma 4.4

For any ξ ,

$$2\lambda(\xi) = E_{\xi} [E_{\xi} [L(\bar{\theta})] - |E_{\xi} [L(\bar{\theta})]|] \quad (4.4.3)$$

The left side of (4.4.2), $\lambda(\xi_{\bar{X}_n})$, represents the advantage of stopping after one free observation rather than stopping with the n observations

already taken. If, at some point in the sampling, say n' , the reduction in the stopping risk caused by taking one more observation is larger than the cost of an observation then the Bayes rule is to stop. The theorem

also states that only the decision problem truncated at some $n \geq n'$ needs to be considered since a Bayes rule for the truncated problem has the minimum Bayes risk for the nontruncated problem.

From (4.4.1) we have $L(\bar{\theta}) = k(\theta_1 - \theta_2)$. The posterior distribution of

(θ_1, θ_2) given the observation $\bar{X}_n$, denoted $\xi_{\bar{X}_n}$, is given by the density function,

$$\xi'_{\bar{X}+X}(\bar{\theta}) = \theta_1 \frac{a+X(1)+X_1-1}{a+X(1)+X_1+X_2-1} \frac{\beta(a+X(1)+X_1+X_2, b+n-X(1)-X_1+1)}{\beta(a+X(2)+X_2, b+n-X(2)-X_2+1)} \cdot$$

$$\text{Let } D(\bar{X}) = E_{\xi} \bar{X}^{n+\bar{X}} [L(\bar{\theta})]$$

$$= \frac{a+b+n+1}{k} [(X(1)-X(2))^n + (X_1-X_2)^n] \cdot$$

We need $E_{\xi} [D(\bar{X})] = E_{\xi} \{E[D(\bar{X})|\bar{\theta}]\}$ to use Lemma 4.4. Since X_1 has the

Bernoulli distribution with parameter θ_1 and X_1 and X_2 are independent,

we have

$$E[D(\bar{X})|\bar{\theta}] = D(0,0)P_{\bar{\theta}}[\bar{X}=(0,0)] + D(1,0)P_{\bar{\theta}}[\bar{X}=(1,0)]$$

$$+ D(0,1)P_{\bar{\theta}}[\bar{X}=(0,1)] + D(1,1)P_{\bar{\theta}}[\bar{X}=(1,1)].$$

Hence

$$E[D(\bar{X})|\bar{\theta}] = \frac{k|X(1)-X(2)|^n}{k|X(1)-X(2)+1|} \frac{a+b+n+1}{(1-\theta_1)(1-\theta_2)} + \frac{k|X(1)-X(2)|^n}{k|X(1)-X(2)+1|} \frac{a+b+n+1}{\theta_1(1-\theta_2)} \cdot$$

Taking the expected value of (4.4.4), we have

$$E_{\xi} \bar{X}_n | E_{\xi} \bar{X}_n + \bar{X} = \frac{a+b+n+1}{k(a+b+n)-2} \left\{ |X_{(1)}^n - X_{(2)}^n| (a+X_{(1)}^n) + 1 | (a+X_{(1)}^n) (b+n-X_{(2)}^n) \right. \\ \left. + |X_{(1)}^n - X_{(2)}^n| - 1 | (b+n-X_{(1)}^n) (a+X_{(2)}^n) \right\} + |X_{(1)}^n - X_{(2)}^n| [(a+X_{(1)}^n) (a+X_{(2)}^n) + (b+n-X_{(1)}^n) (b+n-X_{(2)}^n)]$$

Also we have

$$E_{\xi} [L(\bar{\theta})] = k(X_{(1)}^n - X_{(2)}^n) \frac{a+b+n}{a+b+n} \quad (4.4.5)$$

Then with some algebra we see that

$$\lambda(\xi_{\bar{X}_n}) = 0 \quad \text{whenever} \quad |X_{(1)}^n - X_{(2)}^n| \geq 1.$$

Thus λ has its maximum on the set of points $\{\bar{X}_n : X_{(1)}^n = X_{(2)}^n\}$, where the

stopping risk is the same for either decision (called the "neutral bound-

ary"); that is, where

$$E_{\xi} [L(\bar{\theta})] = 0.$$

For $X_{(1)}^n = X_{(2)}^n = X_0$, λ takes the value

$$\lambda_0 = \frac{k(a+X_0)(b+n-X_0)}{(a+b+n+1)(a+b+n)^2} \quad (4.4.6)$$

From (4.4.6) we see that λ_0 is quadratic in X_0 , λ_0 takes a maximum

value at $\chi_0 = (b-a+n)/2$, and

$$\lambda(\xi_{\chi}^{-n}) < \frac{4(a+b+n+1)}{k} \quad \text{for all } \bar{\chi}_n. \quad (4.4.7)$$

Applying Theorem 4.4, we see that the right side of (4.4.7) is decreasing in n and that (4.4.2) is satisfied for every n such that

$$n \geq \frac{4}{k} - (a+b+1). \quad (4.4.8)$$

For example, if $k=100$ and $(a,b)=(3,12)$ then we only need to consider the sequential problem truncated at $n'=9$, as there would be at most 9 observations in the nontruncated problem.

4.5 Linear Loss

In this section we consider the problem with linear but non-symmetric

losses.

$$L(\bar{\theta}) = k_2(\theta_1 - \theta_2), \quad \theta_1 \leq \theta_2, \quad (4.5.1)$$

$$= k_1(\theta_1 - \theta_2), \quad \theta_1 > \theta_2.$$

This situation is more difficult than that in Section 4.4. In applying Theorem 4.4, the problem was found to be quite involved; and, consequently, the following theorem is used.

Theorem 4.5

A sufficient condition for a Bayes sequential rule with respect to ξ to be truncated is that there exists a non-negative integer n' such

that

$$p_0(\xi) \leq E_{\xi} \bar{\chi}_n \quad (c) = 1 \quad \text{a.s.} \quad (4.5.2)$$

for every $n \geq n'$.

Lemma 4.5

For any ξ ,

$$2p_0(\xi) = E_{\xi} |L(\bar{\theta})| - |E_{\xi} [L(\bar{\theta})]|. \quad (4.5.3)$$

The above theorem states that if the stopping risk must be less than or equal to the cost of one additional observation for every observation after the n' th, regardless of the data observed, then the Bayes rule is

truncated at n' .

Let

$$S_1 = \{(\theta_1, \theta_2) : 0 \leq \theta_2 < \theta_1 \leq 1\}$$

and

$$S_2 = \{(\theta_1, \theta_2) : 0 \leq \theta_1 < \theta_2 \leq 1\},$$

then

$$E_{\xi} \bar{\chi}_n [L(\bar{\theta})] = k_1 \int_1^S (\theta_1 - \theta_2) d\xi \bar{\chi}_n(\bar{\theta}) - k_2 \int_2^S (\theta_2 - \theta_1) d\xi \bar{\chi}_n(\bar{\theta}). \quad (4.5.4)$$

also

$$E_{\bar{\chi}_n}^{\varepsilon} |L(\bar{\theta})| = k_1 \int_{S_1} (\theta_1 - \theta_2) d\bar{\chi}_n + k_2 \int_{S_2} (\theta_2 - \theta_1) d\bar{\chi}_n.$$

Now if (4.5.4) is less than zero, p_0 has the values

$$(4.5.5) \quad p_0(\bar{\chi}_n) = k_1 \int_{S_1} (\theta_1 - \theta_2) d\bar{\chi}_n, \quad (\bar{\theta}),$$

and if (4.5.4) is greater than zero, p_0 takes the values

$$(4.5.6) \quad p_0(\bar{\chi}_n) = k_2 \int_{S_2} (\theta_2 - \theta_1) d\bar{\chi}_n. \quad (\bar{\theta}).$$

On the neutral boundary where $E_{\bar{\chi}_n}^{\varepsilon} [L(\bar{\theta})] = 0$, (4.5.5) and (4.5.6) are equal.

The computer seems to be necessary to find an n satisfying (4.5.2). There remains a great deal of work to be done on the problem considered here, some of which will be mentioned in Chapter V.

5.1 Priors and Losses

The translation of subjective knowledge into a mathematical function (the prior distribution) is of critical importance in the application of Bayesian procedures. Winkler [17] presents a simple study of various techniques for extracting a function reflecting a person's subjective feelings concerning the proportion of a population possessing a certain characteristic. The author has considered the problem of aiding geologists to transmit their feelings concerning an oil prospect through a distribution function. A great more work is needed in this area.

An interesting aspect of the comparison problem as formulated in Chapters II and III is that for the case of common sample size and common prior it is possible to estimate the parameters of the prior distribution from the data. This will be referred to as an empirical Bayes aspect of the comparison problem. With a common sample size of n and common beta prior distribution with parameter (a, b) the χ^2 's have a beta-binomial distribution with mean and variance (μ, σ^2) :

$$\mu = \frac{a+b}{n} \quad \sigma^2 = \frac{nab(a+b+n)}{(a+b)^2(a+b+1)} \quad (5.1.1)$$

Solving for a and b, we have:

$$(5.1.2) \quad \begin{aligned} a &= -\mu \left[\frac{\bar{\mu} - \sigma^2 / (n - \mu)}{\bar{\mu} - n\sigma^2 / (n - \mu)} \right] \\ b &= -(n - \mu) \left[\frac{\bar{\mu} - \sigma^2 / (n - \mu)}{\bar{\mu} - n\sigma^2 / (n - \mu)} \right] \end{aligned} \quad .$$

Then a and b can be estimated by the method of moments simply by replacing μ and σ^2 by $\bar{x}$ and s^2 in (5.1.2). Preliminary studies indicate that this

method may not be completely satisfactory for the present problem. Another method of estimating a and b more in keeping with the Bayesian approach

would be to place a 'diffuse' prior on a and b and then find Bayes estimates for a and b. Of course, the prior need not be a beta distribution

for this empirical Bayes aspect of the problem to be applicable. Also,

this aspect of the problem makes the task of formulating the prior easier since only a parametric form of the prior need be specified prior to observing the data and then the parameters can be estimated from the data.

Further investigation of the above could prove worthwhile.

Another problem deserving further work is the case of non-independence of the θ 's. The comparison problem comparing probabilities of success

should be feasible using a multivariate beta distribution as the prior.

In order that minimum risk procedures may be utilized more, the experimenter needs to begin to think more in terms of losses than the traditional Type I and Type II error probabilities exclusively. Perhaps exam-

ples of practical problems with losses could be emphasized more in elementary statistics courses.

For the comparison problem a slightly different loss function can be used with no further analysis being necessary. If differences be-

tween the probabilities of success are more important near zero and one the loss function

$$L(\theta) = \frac{\theta^1 \theta^2 (1-\theta^1)(1-\theta^2)}{|\theta^1 - \theta^2|}$$

might be better than that considered earlier. The only change this makes is to use $a' = a-1$ and $b' = b-1$ in the decision rule. (Where a and b are the parameters of the prior.) It is clear that further changes in the

loss function could be made, for example

$$L(\bar{\theta}) = (\theta^1 \theta^2)^{m_1} [(1-\theta^1)(1-\theta^2)]^{m_2} |\theta^1 - \theta^2|$$

could be handled by letting $a' = a+m_1$ and $b' = b+m_2$ (where m_1 and m_2 are integers, either positive or negative).

The constant k in the symmetric loss function of Section 4.4 can be thought of as an index on the number of future patients who will be the potential users of the recommended treatment in the medical research example. Anscombe [1] discusses how to make some guess about this constant as related to the clinical trial situation.

Finally, we note that the comparison problem is invariant under the transformation from X to $n-X$ (i.e. number of failures instead of number of successes) and it might be possible to use this to further simplify the analysis. The Bayes rule found in Chapter III is not invariant but

the rule in Bland and Bratcher [4] is, which might make the earlier work more appealing where invariance is desired.

5.2 Sequential Problems

From (3.2.7) we see that, for a common fixed sample size n and common prior distribution, the Bayes solution trivially depends on the sign of

$$[E(\theta_1 | x_1) - E(\theta_2 | x_2)] \text{ when } c = 1 \text{ (i.e. symmetric linear loss). However, as}$$

shown in Sections 4.3 and 4.4, the sequential problem with symmetric linear losses is quite involved. The difficulty increases when the general linear loss function is used. This problem certainly warrants further

consideration. Once a usable solution is derived with $k_1 \neq k_2$, the sequential problem of comparing probabilities could be structured similar

to the fixed sample size problem considered in this work. With further

work it is believed that Theorem 4.4 may be applied to the problem.

In using Theorem 4.4, it would be helpful to know just where λ takes

its maximum for problems similar to the ones considered. Ray [13] presents a theorem which states that for the one dimensional exponential family

with absolutely continuous distributions the maximum of λ is attained on

the neutral boundary if $L(\theta)$ is non-decreasing in θ . Perhaps this theorem could be extended.

BIBLIOGRAPHY

1. Anscombe, F.J. (1965). Comments on paper by Kurtz, Link, Tukey, and Wallace, Technometrics, 7, pp. 167-168.
2. Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances, Ann. Math. Statist., 25, pp. 16-39.
3. Bechhofer, R.E., Kiefer, J., and Sobel, M. (1968). Sequential Identification and Ranking Procedures, The University of Chicago Press, Chicago and London.
4. Bland, R.P. and Bratcher, T.L. (1968). A Bayesian approach to the problem of ranking binomial probabilities, SIAM J. Appl. Math., 16, pp. 843-850.
5. Duncan, D.B. (1961). Bayes rules for a common multiple comparisons problem and related Student-t problems, Ann. Math. Statist., 32, pp. 1013-1033.
6. Duncan, D.B. (1965). A Bayesian approach to multiple comparisons, Technometrics, 7, pp. 171-222.
7. Ferguson, T.S. (1967). Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York and London.
8. Gupta, S.S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations, I. Olkin, S.G. Churye, W. Hoeffding, W.G. Madow, and H.B. Mann, editors, Contributions to Probability and Statistics, Stanford University Press, pp. 224-248.
9. Hoeffding, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. Ann. Math. Statist., 31, pp. 352-368.
10. Jackson, J.E. (1965). Comments on paper by Kurtz, Link, Tukey and Wallace, Technometrics, 7, pp. 163-165.
11. Lehmann, E.L. (1957). A theory of some multiple decision problems I and II, Ann. Math. Statist., 28, pp. 1-25 and pp. 547-572.

12. Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory, Harvard Business School, Boston.
13. Ray, S.N. (1963). Some sequential Bayes procedures for comparing two binomial parameters when observations are taken in pairs. Inst. Stat. Mimeo. Ser. No. 369. Inst. of Statist., Univ. of N. C., Chapel Hill, N. C.
14. Ray, S.N. (1965). Bounds on the maximum sample size of a Bayes sequential procedure. Ann. Math. Statist., 36, pp. 859-878.
15. Sobel, M. and Huyett, M. (1957). Selecting the best one of several binomial populations. Bell System Tech J., 36, pp. 537-576.
16. Weiler, H. (1965). The use of incomplete beta functions for prior distributions in binomial sampling. Technometrics, 7, pp. 335-348.
17. Winkler, R.L. (1967). The assessment of prior distributions in Bayesian analysis. J. Am. Statist. Assoc., 62, pp. 776-800.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

UNCLASSIFIED

SOUTHERN METHODIST UNIVERSITY

3. REPORT TITLE

A Bayesian Treatment of a Multiple Comparison Problem for Binomial Probabilities

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Thomas Lester Bratcher

6. REPORT DATE

November 24, 1969

8a. CONTRACT OR GRANT NO.

N00014-68-A-0515

b. PROJECT NO.

NR 042-260

d.

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Office of Naval Research

13. ABSTRACT

The goal of this paper is to develop a usable procedure which will solve a practical ranking problem for the probability parameters of binomial populations. The proposed procedure follows from an essentially Bayesian approach to the multiple comparisons problem applied to the binomial model. The comparisons problem is formulated in decision theoretic terms as a multiple decision problem. The loss function is taken as the sum of the losses of the component pair-wise comparison problems which generate the multiple comparisons problem. With the additive loss assumption, compatible Bayes rules for the component problems yield a minimum risk solution to the ranking problem consequently reducing the complexity of the problem.

This approach to the multiple comparisons problem allows for the utilization of any prior information on the individual binomial probabilities. The ranking of the parameters is to be based on sufficient statistics which are the numbers of successes in samples from each population. This procedure allows for decisions to be made on unequal sample sizes.

A related sequential problem is also considered. It is shown that a Bayes sequential test is truncated. An upper bound is found on the maximum sample size that a Bayes procedure can take so that its computation is possible by backward induction.