A METHOD FOR DETERMINING THE SENSITIVITY OF A BAYES PROCEDURE TO THE PRIOR DISTRIBUTION

By

DONALD ALAN PIERCE

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1961

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1963

Submitted to the Faculty of the Graduate School
of the Oklahoma State University
in partial fulfillment of the
requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 1965

A METHOD FOR DETERMINING THE SENSITIVITY OF
A BAYES PROCEDURE TO THE
PRIOR DISTRIBUTION

Thesis Approved:

D. Leroy Folks
Thesis Adviser

R B Deal

Paul E Fergu

David L. Weeks

J H Boggs
Dean of the Graduate School

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. J. L. Folks for the guidance and criticism which he supplied daily during the preparation of this paper, and to the other members of my committee who were always willing to give any assistance that I requested.

My thanks go out also to those responsible for the National Science Foundation Graduate Fellowships, for it was through the help of that program that I was able to spend the time in extra-curricular study which led to this research.

I am also indebted to Miss Velda Davis for the correction of many errors and the expert typing of these pages.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

As the scope of problems to which statistical inference is applied becomes wider, it is apparent that it is sometimes necessary to use procedures which formally take into account information which is held prior to the formal experimentation. Rather than taking actions based solely on the outcome of the experiment, one wishes to combine in some manner the information from the experiment with the information already held, and then take action based on this combined information. The need for such a technique arises not only from the desire to make better inferences through the use of prior information, but even more urgently to protect the decision maker from making errors through the misuse of prior information. Such errors are quite apt to be made, because it is most unlikely that if substantial prior information exists, it will not be used in some way, particularly if it disagrees with the experimental results.

There are, of course, instances in decision making in which the prior information is quite nebulous, but attempts to define "complete ignorance" have pointed out some interesting and perhaps unexpected results. (Luce and Raiffa [8]). It is also easy to see that, in the final analysis, the choice of a decision procedure must depend upon the judgment of the decision maker. The alternatives seem to be either to use this judgment in a formal or informal way. The purpose here is not to justify the use of expert opinion, but to attempt to devise procedures to circumvent some of the difficulties of the so-called "Bayesian" method of using prior information.

Objections to Bayes Procedures

Many statisticians and logicians have taken exception to the use of Bayesian inference. Because of the nature of these objections, they can be discussed briefly here before outlining the procedures themselves. The dissension involves the idea of using a probability distribution to represent one's uncertainty about something, and the objections have been primarily of two types, which will be designated as "logical" and "practical."

The logical objection has been that the calculus of probabilities is not valid when the probabilities represent degree of belief rather than some relative frequency idea. For example, a person sympathetic with this objection would feel that to say "the probability that it is snowing in Moscow now is .8" is not an acceptable use of probability.

The practical objection is that in many cases a person's degree of belief is too "fuzzy" to be represented as a precise number. Anyone who has ever tried to represent a degree of belief with a probability will "probably" be sympathetic with this objection.

There has been considerable research directed toward resolving the logical objection. It is apparent that the resolution has not been obtained, but it is also apparent that the idea cannot be dismissed simply because it involves a different concept of probability than most statisticians have accepted in the past. There have been several axiomatic developments leading to weight functions which represent degrees of belief and also obey most* of the conditions for probability measures. Outstanding among these are the developments of Ramsey (12), De Finnetti (2), and Savage (13). It should also be pointed out that attempts to define probability in the relative frequency sense in a precise mathematical manner have not been entirely successful. (Jeffries [6]). As the concern here is with the practical difficulties, this issue will not be discussed further, but it does seem that at least an attempt to resolve some of the practical difficulties is justified.

It is significant that the objections of many of the leading statisticians, as reported in the literature (Pearson [11], Neyman [10]), are based not on the logical but the practical difficulties involved. In an attempt to overcome this difficulty, Lehman and Hodges (7) have proposed a technique called a modified Bayes procedure, which is a mixture between a Bayes procedure and a minimax

---

*There remains some question regarding complete additivity.

procedure. Smith (15) has suggested an idea, also mentioned by Good (3), that a person specify only a convex set of probabilities rather than a unique distribution. Good suggested that one might specify a probability distribution of probability distributions over that convex set. In a practical sense this seems to be asking a lot of the decision maker, and in a theoretical sense admission of this idea would apparently force one to admit the idea of probability distribution upon probability distribution ad infinitum. Smith suggests taking the minimax over this set, which seems to be a reasonable procedure, subject, of course, to the usual criticisms of minimax.

This paper is an attempt to make use of the fact that in many applications it is not necessary to know the prior probabilities exactly. This is to say that in some instances it is possible to change the prior distribution somewhat and not affect the inference or decision at all. Many of the authors who have presented applications of Bayes procedures have developed a sensitivity analysis regarding the prior distribution. All of these analyses, however, have assumed that the prior distribution belonged to some family indexed by a parameter and have examined the sensitivity to changes in this parameter. The basic idea here is to <u>not</u> make this restriction.

Isaacs (5) suggests an idea similar to the one to be developed here. The problem which he was considering was not in a Bayesian context, but in the area of decision making under uncertainty. One must make a decision which is related to a parameter θ. If the probability distribution of θ can be specified as $P(\theta)$, then the action with the smallest expected loss can be taken. To investigate how this decision depends on $P(\theta)$, one might find the distribution $Q(\theta)$, which is "closest" to $P(\theta)$, that results in a different optimal action.

The problem to be considered here is based on an idea quite similar to Isaacs'. The decision maker is to choose one of a finite number of actions. He decides on a prior distribution P. Now, considering the experiment to be conducted, what is the "closest" distribution to P which would result in a different procedure. This problem will not be stated more precisely until the elements of the Bayes procedure are presented. The approach and results here are not at all like Isaacs', and his results will not be presented in this paper.

Decision Problems and Bayes Procedures

The type of statistical decision problem presented here is a special case within the class named "partition problems" by Savage (13). Partition problems occur frequently in some fields of application, notably operations research activities. Other types of decision problems can often be approximated arbitrarily closely within this frame of reference.

Let the elements of the problem be:

A finite number of "states of nature,"

$$\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\},$$

a finite number of actions,

$$A = \{a_1, a_2, \ldots, a_n\},$$

a loss function on $A \times \Theta$,

$$L[a_i, \theta_j],$$

and an experiment, with outcomes

$$x \varepsilon X.$$

The decision maker is faced with the task of choosing an action from A. The proper action to be taken depends, in a way to be explained below, upon which state of nature actually holds true. Which state of nature this is is unknown to some degree to the decision maker, and a prior probability function on Θ,

$$P = \{p_1, p_2, \ldots, p_n\},$$

is an attempt by the decision maker to represent, before the experiment X, all of his information regarding the relative likelihood of the various states of nature.

Known to the decision maker is a loss function L, which is a measure of the desirability (or rather the undesirability) of any action, given that a certain state of nature is true. It will be assumed that the decision maker's primary motive is to minimize, in some sense, the expected loss. It should be pointed out that there are two common conceptions of a loss function. In one of them,

$$L[a, \theta] = -I[a, \theta],$$

where I is the income, or utility, (positive or negative) derived from taking action a when the state of nature is θ. The other conception is that it is more reasonable to use

$$L[a, \theta] = \max_{a^* \varepsilon A}\{I[a^*, \theta]\} - I[a, \theta].$$

That is, the loss values for each state of nature are measured from an origin corresponding to the very best that could be accomplished if the state of nature were known. Although this distinction is most important, it does not affect the mathematics of the procedure once the loss function is determined, so it will not be discussed here. Excellent discussions for both sides of the issue, and also developments of how utility is measured, are given in Savage (13), and in Luce and Raiffa (8).

The decision maker has the opportunity to observe, before choosing an action from A, an outcome $x \varepsilon X$ of an experiment. The probability distribution of x will depend upon the state of nature through a likelihood function on $\Theta \times X$, $f_i(x)$, determining a probability measure on X for each $\theta \varepsilon \Theta$. For a fixed $\theta \varepsilon \Theta$, the function of x, $f_i(x)$ can be either a probability function or a density function, depending on the structure of the space X. It will be assumed that the statistic x is sufficient for the family of experiments indexed by the state of the nature. It will also be assumed that $f_i(x) > 0$ for every x and for every i.

The decision maker's procedure can be represented mathematically by the specification of a class D of decision functions, with elements $d(x) \varepsilon D$, each decision function mapping X into A. If the experimental outcome is x, the decision maker chooses action d(x). The primary problem of decision theory is to investigate the choice of decision function. Clearly, the choice of a decision function should depend upon at least the prior distribution, the loss function, and the likelihood function.

It should be pointed out that except for the limitations made here on the structure of the spaces A and Θ, this is a very general formulation of the statistical inference problem. This formulation is primarily due to Wald and is presented elegantly in the first chapter of Wald (16). Excellent discussions of the application of this formulation are given by Schlaiffer (14), Savage (13), and Blackwell and Girshick (1), increasing in level of difficulty in the order that they are listed.

Consider now the problem of evaluating a decision function. If θ_i is the true state of nature and a decision function d(x) is used to choose an action, the <u>risk</u> <u>function</u>

$$R(d, \theta_i) = \int_X L[d(x), \theta_i]\ f_i(x)dx$$

is clearly the expected loss to the decision maker.

Because the decision maker does not know the true θ_i, however, this function does not serve to establish a preference of d_o over d_1 unless

$$R(d_o, \theta_i) \leqq R(d_1, \theta_i) \quad , \quad i = 1, 2, \ldots, n,$$

$$R(d, \theta_k) < R(d_1, \theta_k) \text{ for some } k = 1, 2, \ldots, n.$$

A Bayes procedure consists of using the prior distribution over Θ to obtain an expected risk, denoted by the <u>Bayes risk</u> function,

$$B(d, P) = \sum_{\theta_i \varepsilon \Theta} p_i R(d, \theta_i).$$

This provides, given a prior distribution, a procedure for ranking decision functions according to preference, and suggests the following definition.

1.1) <u>Definition</u>: A Bayes decision function against P, denoted by d_p, is a function in D which satisfies

$$B(d_p, P) \leqq B(d, P) \quad \text{for all } d \varepsilon D.$$

The use of a Bayes decision function is called a Bayes procedure.

It is interesting to note that if the decision maker has an order relation in D corresponding to his preference of one decision function over another, and if this order relation satisfies certain axioms which can be thought of as representative of the decision maker's "rationality," then there can be shown to exist a prior distribution P such that the ordering of the Bayes risk functions will correspond to the ordering in D. A proof of this is given in Blackwell and Girshick (1).

There are two approaches to the construction of a Bayes procedure. The decision maker can, before observing x, evaluate in some way the Bayes risk functions and choose from D the Bayes decision function, d_p. The other approach consists of first observing x, and then, instead of solving for the function d_p over its entire domain, he need only solve for $d_p(x)$ for the x observed.

Since

$$B(d, P) = \Sigma\, p_i\, R(d, \theta_i)$$

$$= \Sigma\, p_i \int_X L[d(x), \theta_i]\, f_i(x) dx$$

$$= \int_X [\Sigma L[d(x), \theta_i]\, p_i\, f_i(x)] dx$$

a necessary and sufficient condition for

$$B(d_o, P) \leqq B(d_1, P)$$

is for

$$\Sigma L[d_o(x), \theta_i]\, p_i\, f_i(x) \leqq \Sigma L[d_1(x), \theta_i] p_i\, f_i(x) \quad \text{for every } x \,\varepsilon\, X.$$

Thus, if x is the outcome observed, a Bayes procedure consists of choosing $a_p \,\varepsilon\, A$ such that

$$\Sigma L[a_p, \theta_i]p_i f_i(x) \leqq \Sigma L[a, \theta_i]p_i f_i(x) \text{ for every } a \in \mathbf{A}.$$

Since by assumption $\Sigma p_i f_i(x) > 0$ for any P and any x, this is equivalent to choosing a_p such that

$$\Sigma L[a_p, \theta_i] \frac{p_i f_i(x)}{\Sigma p_j f_j(x)} \leqq \Sigma L[a, \theta_i] \frac{p_i f_i(x)}{\Sigma p_j f_j(x)} \text{ for every } a \in \mathbf{A}.$$

Let the conditional distribution of θ given x be referred to as the posterior distribution over Θ, and denoted by

$$P(x) = \{p[\theta_1 | x], p[\theta_2 | x], \ldots, p[\theta_n | x]\}.$$

Thus, a Bayes procedure consists of choosing a_p such that

$$\Sigma L[a_p, \theta_i]P[\theta_i | x] \leqq \Sigma L[a, \theta_i]p[\theta_i | x] \text{ for every } a \in \mathbf{A}.$$

It will be convenient to write this equation in terms of the Bayes loss,

$$W[a, P(x)] = \Sigma L[a, \theta_i]\, p\, [\theta_i | x],$$

which is clearly the expected loss of any action under the posterior distribution. Finally, then, if the outcome $x \in X$ has been observed, a Bayes procedure consists of choosing a_p such that

$$W[a_p, P(x)] \leqq W[a, P(x)] \text{ for every } a \in \mathbf{A}.$$

If the action a_p satisfies this relation, it will be called the best action against P(x).

Sensitivity to the Prior Distribution

In the same manner that there are two approaches to finding Bayes procedures, there are two approaches to determining the sensitivity of the procedure to the prior distribution. The a priori approach is to determine, before the experiment is conducted, the effect of changes in the prior distribution on the choice of a decision function. The a posteriori method consists of determining, after the experiment has been performed, the effect of changes in the prior distribution on the choice of an action to be taken.

In the following chapters a procedure will be developed which can, to a certain extent, be applied to either an a priori or an a posteriori analysis. The primary emphasis will be placed on the a posteriori approach for reasons that will be discussed below. For the sake of clarity, the development will be presented in terms of the a posteriori analysis and the necessary changes to obtain an a priori analysis will be pointed out throughout this development.

The basic idea upon which the sensitivity analysis will be based was indicated previously. This idea is given that the decision maker has estimated the prior distribution to be P, what is the closest distribution Q which would result in a different decision? To state this idea more precisely, the distinction must be made between the two possible approaches pointed out above.

Suppose first that the space D of possible decision functions contains only a finite number of elements. The a priori analysis in this case can consist of finding the distribution Q, nearest to P, for which the best decision function is different than the one which is best against P. The a posteriori analysis can consist of, given that $x \varepsilon X$ has been observed, finding the distribution Q, nearest to P, for which the best action against Q(x) is different from the best action against P(x).

A procedure will be developed in the next chapter which can be used in almost exactly the same manner to solve either problem. It is felt, however, that the results of the a posteriori analysis will have much more practical significance. The primary reason for this is that the decision maker is essentially interested in the decision function evaluated at only one point, the x which is observed in the outcome of the experiment. He will, in general over-estimate the sensitivity of the procedure with the a priori analysis because he will be concerning himself with changes in the decision function which will not actually make any difference, since many decision functions map the x which is actually observed into the same action. It will also be seen that as far as the action taken, the sensitivity to the prior distribution is very much dependent upon which experimental outcome obtains. That is to say, for some experimental outcomes the prior distribution plays very little part in determining the posterior distribution and for other outcomes the prior distribution is most important.

From one point of view, the two analyses can be compared in the following manner. The a priori analysis will divide the space of all possible prior distributions into equivalence classes, with $Q_1 \sim Q_2$ if the same decision function is best against each of them. For each experimental outcome $x \varepsilon X$, the a posteriori analysis will partition the same space into equivalence classes, with $Q_1 \sim Q_2$ if the same action is best against $Q_1(x)$ and $Q_2(x)$. The equivalence classes for the former case can be formed using the equivalence classes of the latter as follows.

Suppose that the space X contains a finite number of elements. Given any $x \varepsilon X$, the Bayes procedure for any given prior distribution Q can be found by minimizing over $a \varepsilon A$, the quantity

$$\sum_{\theta_i} L[a, \theta_i] f_i(x) q_i . \tag{1.2}$$

A priori, the Bayes decision function is found by minimizing over $d \varepsilon D$, the quantity

$$\sum_{x} \Big[\sum_{\theta_i} L[d(x), \theta_i] f_i(x) q_i \Big] , \tag{1.3}$$

but this can be done by solving (1.2) for each $x \varepsilon X$. Thus, given Q and the best action against Q(x) for each $x \varepsilon X$, the decision function can be built up.

As an example, let

$$A = \{a_1, a_2\}$$

$$X = \{x_1, x_2, x_3\}.$$

For each $x \varepsilon X$, the a posteriori partitions are symbolically indicated in Figure 1.

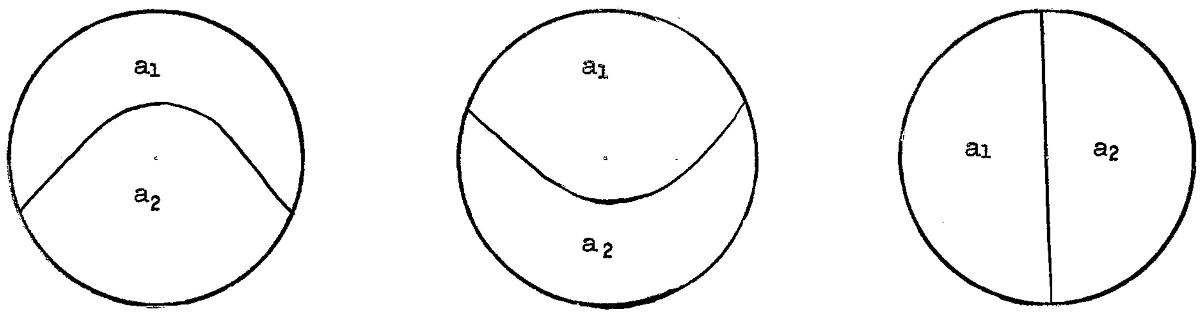

Figure 1. The A Posteriori Partitions of the Space of Prior Distributions

The a priori partition is indicated in Figure 2 where $\{a_i, a_j, a_k\}$ is the decision function for which $d(x_1) = a_i$, $d(x_2) = a_j$, and $d(x_3) = a_k$.

If P were as shown in Figure 2, then the a priori solution would be Q, whereas if x_2 were the experimental outcome, then Q' would be the nearest distribution to P which really made any difference. It can be seen that the entire a priori analysis hinges upon the experimental outcome for which the procedure will be the most sensitive to the prior distribution.

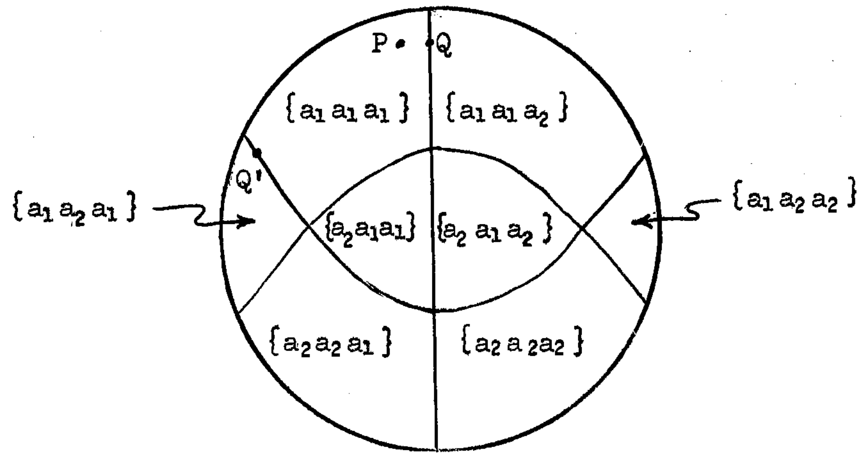

Figure 2. The A Priori Partition of the Space of Prior Distributions

The role of the experimental outcome in determining the sensitivity to the prior distribution can be illustrated in another way. When P and x are fixed,

$$p[\theta_i \mid x] \propto p_i f_i(x),$$

when considered as a function of θ_i, with the constant of proportionality determined by the condition that the posterior probabilities must add to one. It is seen, then, that $f_i(x)$, considered as a function of i is most important in determining the posterior distribution. In Figure 3(a), the posterior distribution would be quite sensitive to changes in the prior distribution; whereas, in Figure 3(b), the posterior distribution would not be sensitive at all to changes in the prior distribution.

In addition to the logical considerations, computational

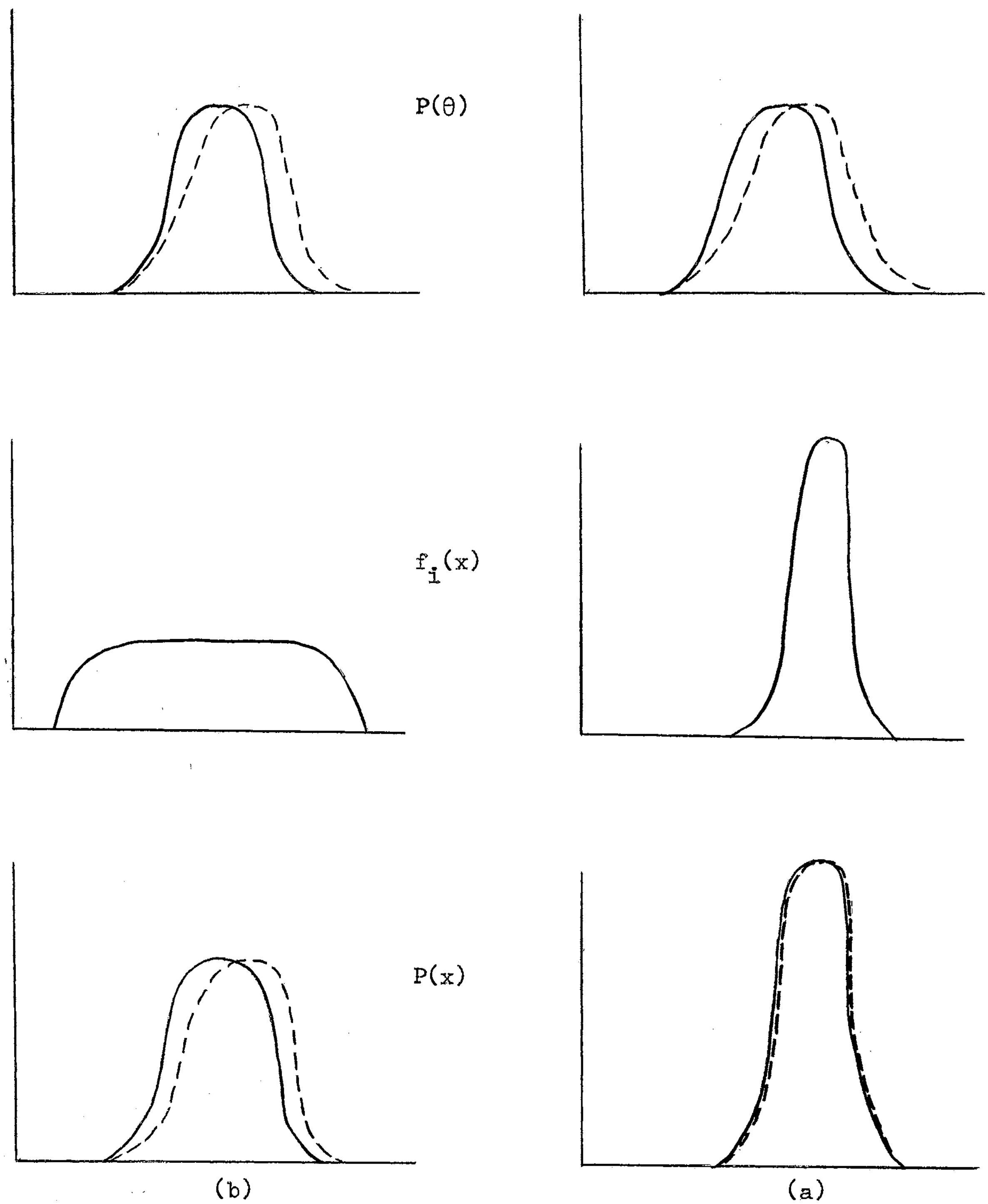

Figure 3. The Influence of the Likelihood Function in Determining the Posterior Distribution

and methodological factors indicate the desirability of the a posteriori analysis. In the case that x can take on infinitely many values, there are an infinite number of possible decision functions to consider. If x is a continuous variable and the class of decision functions is not restricted severely, then Q can be arbitrarily close to P and still result in a different Bayes decision function. Even if X is finite, there are m^k possible decision functions, where A has m elements and X has k elements. In the a posteriori analysis, the structure of the space X has no bearing on the problem since the only concern is with the value of x obtained.

There is at least one general drawback to the a posteriori analysis. In formulating a prior distribution, the decision maker is to act independently of the observation x. This is tacitly assumed in the procedure for computing the posterior distribution. The seriousness of this drawback is difficult to evaluate, but it might be advisable for the decision maker not to be told the value of x obtained, but only the solution Q to the sensitivity analysis procedure. Of course, the choice of the prior distribution should not depend upon Q either, but no alternative procedure has been developed. The decision maker could be told only the distance from P to Q, but it is doubtful if this information would be very meaningful to him.

Generally, it seems that the a posteriori analysis will be more valuable. There will be certain specialized situations, however, in which the a priori analysis will be needed. It might, for example, be necessary to evaluate the choice of a decision function from a very restricted class of functions for a situation in which circumstances prohibit an a posteriori analysis. This might be the situation if one were developing an automated decision making device for use in a guided missile. To allow for this possibility, the necessary procedural changes for this type of analysis will be pointed out throughout the remainder of the paper. An example of an a priori analysis will be given for a quality control problem in Chapter IV.

The following example will illustrate the nature of the Bayesian approach. The sensitivity of the solution to the prior distribution will be analyzed at the end of the next chapter.

Example 1: A manufacturer of small rockets has a contract to produce a considerable number of a newly designed model and estimates of the utility structure as a function of the probability of failure are as follows, where a_1 is the action to begin production:

θ_i	.1	.2	.3	.4	.5
$-I[a_1, \theta_i]$	0	50	100	500	1000

The engineering office has estimated the prior distribution

P of the probability of failure θ to be,

$$P = \{.1, .25, .35, .25, .05\}.$$

Five rockets have been tested and one failure observed. The following calculations have been made, using $f_i(x) = \binom{5}{x}\theta_i^x(1 - \theta_i)^{5-x}$, where x is the number of failures observed out of five trails.

	θ_1	θ_2	θ_3	θ_4	θ_5
$f_i(1)$	.328	.410	.360	.260	.156

The decision to be made is whether to begin production (a_1), redesign (a_3), or to test five more rockets (a_2). The redesign cost is 100 and it will be assumed that redesign will result in a reliability of at least .9. For simplicity, it will be assumed that in the event that five more are tested, more than one failure will result in a decision to redesign, and one or less failures will result in a decision to produce. The cost of the additional testing is 25.

In this case, the utility if it is decided to test should be calculated as

$$-I[a_2, \theta_i] = 25 + \left[\sum_{x=0}^{1}\binom{5}{x}\theta_i^x(1 - \theta_i)^{5-x}\right]\left[-I[a_p, \theta_i]\right]$$

$$+ \left[\sum_{x=2}^{5}\binom{5}{x}\theta_i^x(1 - \theta_i)^{5-x}\right]\left[100\right].$$

Thus, the following utility functions are obtained.

	θ_1	θ_2	θ_3	θ_4	θ_5
$-I[a_1, \theta_i]$	0	50	100	500	1000
$-I[a_2, \theta_i]$	33	86	125	260	294
$-I[a_3, \theta_i]$	100	100	100	100	100

The following loss functions are computed using

$$L[a, \theta] = \max_{a^* \varepsilon A} \{I[a^*, \theta]\} - I[a, \theta].$$

	θ_1	θ_2	θ_3	θ_4	θ_5
$L[a_1, \theta_i]$	0	0	0	400	900
$L[a_2, \theta_i]$	33	36	25	160	194
$L[a_3, \theta_i]$	100	50	0	0	0

The posterior distribution is computed from P and $f_i(1)$, $i = 1, 2, 3, 4, 5$,

$$P(1) = \{.097, .308, .378, .194, .023\}.$$

The Bayes loss for each action is computed using this posterior distribution.

$$W[a_1, P(1)] = 98.3,$$

$$W[a_2, P(1)] = 59.3,$$

$$W[a_3, P(1)] = 25.1.$$

The best action, then, is to redesign. Just how sensitive to the prior distribution this decision is will be considered at the end of the next chapter.

CHAPTER II

THE SENSITIVITY ANALYSIS PROCEDURE

This chapter will be devoted to developing a mathematical procedure for carrying out the sensitivity analysis procedure described in the introduction. Before this can be done, some attention must be given to defining what is meant by distance in the space of all possible prior distributions. For the first part of the chapter, it will only be assumed that the distance function $\delta(P, Q)$ is a non-negative, strictly convex function of Q and that $\delta(P, P) = 0$. Actually, then, $\delta(P, Q)$ need not be a distance function at all in the sense of defining a metric. Later in the chapter, considerable development will be done using the usual Euclidean norm as the distance function. In the next chapter, a discussion of the role of the distance function will be presented, and some alternative distance functions considered.

The basic problem presented in the introduction is, given a prior distribution P and an experimental outcome x, with a_p any best action against P(x), find Q_o such that:

$$W[a_k, Q_o(x)] < W[a_p, Q_o(x)] \quad \text{for some } k \neq p, \tag{2.1.1}$$

and such that, for any Q satisfying

$$W[a_i, Q(x)] < W[a_p, Q(x)], \quad \text{for any } i = 1, 2, \ldots, m, \tag{2.1.2}$$

$$\delta(P, Q_o) \leqq \delta(P, Q). \tag{2.1.3}$$

If it is desired to make an a priori analysis, the problem becomes, given a prior distribution P, with d_p the Bayes decision function against P, find Q_o such that:

$$B(d_k, Q_o) < B(d_p, Q_o) \quad \text{for some } k \neq p, \tag{2.2.1}$$

and such that, for any Q satisfying

$$B(d_i, Q) < B(d_p, Q), \quad \text{for any } i, \tag{2.2.2}$$

$$\delta(P, Q_o) \leqq \delta(P, Q). \tag{2.2.3}$$

Since

$$W[a_k, Q(x)] - W[a_p, Q(x)] = \frac{1}{\Sigma q_i f_i(x)}$$

$$\Sigma q_i \left\{f_i(x)\Big[L[a_k, \theta_i] - L[a_p, \theta_i]\Big]\right\},$$

and

$$B(d_k, Q) - B(d_p, Q) = \Sigma q_i \{R(d_k, \theta_i) - R(d_p, \theta_i)\},$$

the a priori analysis can be carried out through the same procedure as the a posteriori, provided that the space D contains a finite number of elements, by replacing the quantities $f_i(x)\{L[a_k, \theta_i] - L[a_p, \theta_i]\}$

by

$$R(d_k,\ \theta_i) - R(d_p,\ \theta_i), \quad i = 1,\ 2,\ \ldots,\ n.$$

The Existence of a Solution

The problem is stated above in what seems to be the most natural manner. It will be shown in this section that the problem as stated has no solution, but that a suitable modification can be made resulting in a problem that always has a unique solution.

The approach to solving problem (2.1) is to find, for each $k = 1, 2, \ldots, m$; $k \neq p$, Q_k such that:

$$W[a_k,\ Q_k(x)] < W[a_p,\ Q_k(x)], \tag{2.3.1}$$

and such that, for any Q satisfying

$$W[a_k,\ Q(x)] < W[a_p,\ Q(x)], \tag{2.3.2}$$

$$\delta(P,\ Q_k) \leqq \delta(P,\ Q). \tag{2.3.3}$$

Then, choosing Q_o from the set

$$\{Q_k\} \quad k = 1,\ 2,\ \ldots,\ m \quad ; \quad k \neq p,$$

such that

$$\delta(P,\ Q_o) \leqq \delta(P,\ Q_k) \quad k = 1,\ 2,\ \ldots,\ m \quad ; \quad k \neq p,$$

will give a solution to problem (2.1) if one exists. It is also clear that if none of the problems (2.3) have a solution, then problem (2.1) cannot have a solution.

Considering the problem (2.3) for a specific $k \neq p$, one must:

$$\text{Minimize} \quad \delta(P, Q), \tag{2.4}$$

subject to:

$$\sum_i L[a_k, \theta_i] \frac{q_i f_i(x)}{\Sigma q_j f_j(x)} < \sum_i L[a_p, \theta_i] \frac{q_i f_i(x)}{\Sigma q_j f_j(x)}, \tag{2.4.1a}$$

$$\Sigma q_i = 1, \tag{2.4.2}$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n. \tag{2.4.3}$$

The restriction (2.4.1a) can be replaced by

$$\sum_i q_i b_{ki} < 0, \tag{2.4.1}$$

where

$$b_{ki} = f_i(x)\{L[a_k, \theta_i] - L[a_p, \theta_i]\}.$$

2.5) Theorem: If $\delta(P, Q)$ is a non-negative, strictly convex function of Q and $\delta(P, P) = 0$, problem (2.4) has no solution.

Proof: It will be shown that if Q satisfies conditions (2.4.1) through (2.4.3), then there exists a probability distribution R on Θ, also satisfying those conditions with

$$\delta(P, R) < \delta(P, Q).$$

This can be shown by letting

$$R = \alpha P + (1 - \alpha) Q ,$$

where (dropping the subscript k)

$$\alpha = -\Sigma\, q_i\, b_i / 2(\Sigma\, p_i\, b_i - \Sigma\, q_i\, b_i).$$

Then

$$\begin{aligned}
\Sigma\, r_i\, b_i &= \alpha\, \Sigma\, p_i\, b_i + (1-\alpha)\, \Sigma\, q_i\, b_i \\
&= \alpha\, (\Sigma\, p_i\, b_i - \Sigma\, q_i\, b_i) + \Sigma\, q_i\, b_i \\
&= -\tfrac{1}{2} \Sigma\, q_i\, b_i + \Sigma\, q_i\, b_i \\
&= \tfrac{1}{2} \Sigma\, q_i\, b_i < 0 ,
\end{aligned}$$

since

$$\Sigma\, q_i\, b_i < 0.$$

Since a_p is a best action against P(x), $\Sigma\, p_i\, b_i \geqq 0$, and thus,

$$0 < \alpha \leqq 1.$$

Therefore, by the convexity of δ,

$$\begin{aligned}
\delta(P, R) &= \delta(P, \alpha P + (1 - \alpha) Q) \\
&\leqq \alpha \delta(P, P) + (1 - \alpha)\, \delta\, (P, Q) \\
&= (1 - \alpha)\, \delta\, (P, Q) \\
&< \delta(P, Q).
\end{aligned}$$

It is seen then that although the statement of the problem in (2.1) seems to be the most reasonable formulation, this problem has no solution. The statement of the problem will, therefore, be re-formulated, asking for the nearest distribution for which one is <u>indifferent</u> between a_p and some other action. A great deal of relatively useless qualification will be avoided throughout the remainder of this paper if it is assumed that there is only one best action against P(x), so this assumption will be made.

The problem as re-formulated is:

$$\text{Minimize} \quad \delta(P, Q), \tag{2.6}$$

subject to:

$$\Sigma\, q_i\, b_{ki} = 0, \tag{2.6.1}$$

$$\Sigma\, q_i = 1, \tag{2.6.2}$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n. \tag{2.6.3}$$

The solution to this problem will exist unless a_k is an inadmissible action according to the following definition.

2.7) <u>Definition</u>: An action a_k is inadmissible if for some j

$$L[a_k, \theta_i] \geqq L[a_j, \theta_i] \quad \text{for all } i = 1, 2, \ldots, n,$$

and

$$L[a_k, \theta_i] > L[a_j, \theta_i] \quad \text{for some } i = 1, 2, \ldots, n.$$

2.8) <u>Theorem</u>: If a_k is not an inadmissible action, and if $\delta(P, Q)$ is a non-negative, strictly convex function of Q with $\delta(P, P) = 0$, then,

problem (2.6) has a unique solution Q_k, (2.8.1)

if $W[a_k, Q(x)] < W[a_p, Q(x)]$, then $\delta(P, Q_k) < \delta(P, Q)$, (2.8.2)

$$\inf \{\delta(Q_k, Q) \mid W[a_k, Q(x)] < W[a_p, Q(x)]\} = 0. \quad (2.8.3)$$

<u>Proof</u>: It will be shown that there is at least one distribution Q which satisfies the constraints (2.6.1) through (2.6.3). Since the set of distributions which satisfy those constraints is closed and bounded, and $\delta(P, Q)$ by its convexity must be continuous, then the function $\delta(P,Q)$ must take on a minimum value on this set.

Since a_k is not inadmissible, either $b_{ki} = 0$ for all i or $b_{ki} < 0$ for some i. In the former case, any distribution satisfies the constraints. Because a_p is a Bayes solution against $P(x)$, $\Sigma p_i b_{ki} \geqq 0$. If $\Sigma p_i b_{ki} = 0$, then P satisfies the constraints. The only remaining case is where $b_{ki} < 0$ for some i and $\Sigma p_i b_{ki} > 0$. In this case, the must be some j for which $b_{kj} > 0$. The distribution Q with

$$q_i = b_{kj}/(b_{kj} - b_{ki}),$$

$$q_j = b_{ki}/(b_{kj} - b_{ki}),$$

$$q_r = 0 \quad , \quad r = 1, 2, \ldots, n \quad ;$$
$$r \neq i \quad , \quad r \neq j,$$

will then satisfy the constraints of problem (2.6).

To show uniqueness of the solution, suppose that Q and R are both solutions to the problem. Since the constraint set is convex, $\alpha Q + (1 - \alpha)R$ also satisfies the constraints for $0 < \alpha < 1$. But, due to the strict convexity of $\delta(P, Q)$, the contradiction

$$\delta(P, \alpha Q + (1 - \alpha)R) < \alpha\delta(P, Q) + (1 - \alpha)\delta(P, R) = \delta(P, Q)$$

is obtained.

To show that $\delta(P, Q_k) < \delta(P, Q)$ for the set $\{Q \mid W[a_k, Q(x)] < W[a_b, Q(x)]\}$, let Q satisfy $W[a_k, Q(x)] < W[a_b, Q(x)]$. Letting

$$R = \alpha P + (1 - \alpha) Q,$$

where

$$\alpha = -\Sigma q_i b_{ki} / (\Sigma p_i b_{ki} - \Sigma q_i b_{ki}),$$

then, it is easily seen that

$$\Sigma r_i b_{ki} = 0.$$

It was shown in Theorem 2.5 that $0 < \alpha \leqq 1$. Thus, R satisfies the constraints of problem (2.6) and

$$\delta(P, Q_k) \leqq \delta(P, R).$$

But, by the convexity of $\delta(P, Q)$,

$$\delta(P, R) \leqq (1 - \alpha)\delta(P, Q),$$

and, therefore,

$$\delta(P, Q_k) < \delta(P, Q).$$

That $\inf\{\delta(Q_k, Q) \mid W[a_k, Q(x)] < W[a_b, Q(x)]\} = 0$ is shown as follows. Due to the continuity of $\delta(P, Q)$, for any $\varepsilon > 0$ a Q such that $W[a_k, Q(x)] < W[a_b, Q(x)]$ can be found for which

$$\delta(Q_k, Q) < \varepsilon$$

by increasing one of the coordinates of Q_k associated with the smallest b_{ki} by a sufficiently small amount and decreasing by the same amount one of the coordinates associated with a larger b_{ki}. If all the b_{ki} are equal, they must all be zero and this case has been assumed not to occur.

The fundamental problem of this chapter, stated in equations (2.1) is now modified to read:

Find Q_o such that (2.9)

$$W[a_k, Q_o(x)] = W[a_p, Q_o(x)] \quad \text{for some } k \neq p, \tag{2.9.1}$$

and such that, for any Q satisfying

$$W[a_i, Q(x)] \leq W[a_p, Q(x)] \quad \text{for any } i = 1, 2, \ldots, n, \tag{2.9.2}$$

$$\delta(P, Q_o) \leqq \delta(P, Q). \tag{2.9.3}$$

Theorem 2.8 establishes that this problem has a solution unless all the actions other than a_p are inadmissible, and that there will always exist, arbitrarily close to Q_o, a distribution against which a_p is not the best action. Problem (2.9) will be solved as indicated in equations (2.3) with the inequalities replaced by equalities. It will be shown subsequently that it is generally not necessary to solve the problem for every action alternative to a_p, and that in many instances it is only necessary to solve the problem for one alternative action.

The idea of inadmissibility can be easily extended to decision functions rather than actions simply by replacing the loss function with the risk function. Theorems for the a priori analysis can then be obtained which are analogous to the two above.

Construction of a Solution

Attention will now be turned to finding a solution to the problem presented in equations (2.6) when $\delta(P, Q)$ is the usual Euclidean norm; that is,

$$\text{Minimize} \quad \delta(P, Q) = [\Sigma(p_i - q_i)^2]^{\frac{1}{2}}, \tag{2.10}$$

subject to

$$\Sigma q_i b_i = 0, \tag{2.10.1}$$

$$\Sigma q_i = 1, \tag{2.10.2}$$

$$q_i \geqq 0 \quad, \quad i = 1, 2, \ldots, n. \tag{2.10.3}$$

It will be assumed that $b_i > 0$ and $b_j < 0$ for some i and j.

Two methods of solving this problem will be presented. Method I is proved to result in a solution for every problem. Method II is considerable easier to apply, but has not been proved to be successful in every case. The difficulty will be discussed as the methods are presented.

To solve the problem, it will clearly be sufficient to minimize $[\delta(P, Q)]^2$. Neglecting the non-negativity constraint (2.10.3) for the moment, the method of La Grange multipliers can be applied.

Let

$$G = \Sigma(p_i - q_i)^2 + 2\lambda \Sigma q_i b_i + 2\mu(\Sigma q_i - 1).$$

Then a solution must satisfy the $n+2$ equations,

$$q_i - p_i + \lambda b_i + \mu = 0 \quad , \quad i = 1, 2, \ldots, n, \qquad (2.11.1)$$

$$\Sigma q_i b_i = 0, \qquad (2.11.2)$$

$$\Sigma q_i = 1, \qquad (2.11.3)$$

for some λ and μ. It will be shown that if the solution to these equations does not violate the non-negativity conditions, then it is the solution to problem (2.10).

The method of incorporating conditions (2.10.3) into this problem is based on some results of Kuhn and Tucker, obtained in their work is the theory of games. This Kuhn-Tucker theory is that upon which some of the methods of quadratic programming have been based. Although these quadratic programming methods could be used directly for the problem at hand, the special nature of this problem enables one to apply the Kuhn-Tucker theory more directly. Because the results of Kuhn and Tucker are explained and proved very lucidly in Chapter 6 of Hadley (4), this book will be used as a primary reference.

The primary theorem will be proved here for any convex distance function so that the results can be used in a later chapter. It should be pointed out that the Kuhn-Tucker theory applies to much more general problems.

2.12) Theorem: Let $\delta(P, Q)$ be continuously differentiable strictly convex function of Q, and let the set E of vectors Q satisfying

$$\Sigma q_i b_i = 0,$$

$$\Sigma q_i = 1,$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n,$$

be non-empty. A necessary and sufficient condition that $\delta(P, Q)$ be the minimum of $\delta(P, R)$ for R in E, and for Q to be the only point in E at which $\delta(P, R)$ takes on this minimum value, is that there exist λ, μ such that,

$$\frac{\partial \delta}{\partial q_i}(P, Q) + \lambda b_i + \mu \geqq 0 \quad , \quad i = 1, 2, \ldots, n \tag{2.12.1}$$

$$\sum_i \left[\frac{\partial \delta}{\partial q_i}(P, Q) + \lambda\, b_i + \mu\right] q_i = 0 \tag{2.12.2}$$

$$Q \text{ is in } E. \tag{2.12.3}$$

Proof: Let $\frac{\partial \delta}{\delta q_i}$ be written δ^i. To show the sufficiency, it will first be shown that,

$$\delta(P, R) \geqq \delta(P, Q) + \sum_{i=1}^{n} \delta^i(P, Q)(r_i - q_i)$$

for any R in E.

For any $0 < \alpha \leqq 1$,

$$\delta[P, Q + \alpha(R - Q)] \leqq \alpha\delta(P, R) + (1 - \alpha)\delta(P, Q),$$

or

$$\frac{\delta[P, Q + \alpha(R - Q)] - \delta(P, Q)}{\alpha} \leqq \delta(P, R) - \delta(P, Q).$$

Then, Taylor's formula can be used to write

$$\Sigma\delta^i[P, Q + \theta\alpha(R - Q)](r_i - q_i) \leqq \delta(P, R) - \delta(P, Q),$$

where $0 \leqq \theta \leqq 1$, and taking the limit as α approaches zero,

$$\delta(P, Q) + \Sigma\delta^i(P, Q)(r_i - q_i) \leqq \delta(P, R).$$

Let Q, λ, and μ satisfy the conditions of the theorem. If Q and R each belong to E, then clearly,

$$\Sigma b_i(r_i - q_i) = 0$$

$$\Sigma(r_i - q_i) = 0.$$

Combining this with the inequality developed above,

$$\delta(P,R) \geqq \delta(P,Q) + \Sigma\delta^i(P,Q)(r_i - q_i) + \lambda\Sigma b_i(r_i - q_i)$$

$$+ \mu\Sigma(r_i - q_i)$$

$$= \delta(P,Q) + \Sigma[\delta^i(P,Q) + \lambda b_i + \mu](r_i - q_i),$$

and from condition (2.12.2) on Q,

$$\delta(P,R) \geqq \delta(P,Q) + \Sigma[\delta^i(P,Q) + \lambda b_i + \mu] r_i.$$

Since $r_i \geqq 0$, condition (2.12.1) on Q gives

$$\delta(P,R) \geqq \delta(P,Q),$$

and, thus, the sufficiency of the conditions. It was shown in Theorem 2.8 that the minimum can only be taken on at one point in E.

The necessity of the conditions of the theorem is very closely related to the duality theory of linear programming and this connection will be used in the proof.

Let Q be the point in E at which $\delta(P,R)$ takes on its minimum value. Let

$$\Delta' = \{\delta^1(P,Q), \delta^2(P,Q), \ldots, \delta^n(P,Q)\},$$

$$G = \begin{Bmatrix} b_1 & b_2 & \cdots & b_n \\ 1 & 1 & \cdots & 1 \end{Bmatrix}.$$

The proof hinges upon the fact that if

$$q_i = 0 \quad , \quad i = 1, 2, \ldots, k,$$

$$q_i > 0 \quad , \quad i = k+1, \ldots, n,$$

then, for any $V' = \{v_1, v_2, \ldots, v_n\}$ with

$$v_i \geqq 0 \quad , \quad i = 1, 2, \ldots, k,$$

for which

$$GV = \varphi ,$$

V must satisfy

$$\Delta' V \geqq 0 .$$

This is shown as follows. For any V satisfying the above conditions, there exists an $\alpha > 0$ such that for $0 \leqq \theta \leqq \alpha$,

$$q_i + \theta v_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n.$$

Since

$$GV = \varphi ;$$

$$\Sigma b_i [q_i + \theta v_i] = 0 ,$$

$$\Sigma [q_i + \theta v_i] = 1 ,$$

and, thus,

$$\delta(P, Q + \theta V) \geqq \delta(P, Q) .$$

If $\Delta' V < 0$, then $\Delta'(\theta V) < 0$, and a Taylor's expansion of $\delta(P, Q)$ about Q would show that there exists a $\theta_o \varepsilon (0, \alpha)$ such that

$$\delta(P, Q + \theta_o V) < \delta(P, Q).$$

Therefore, for any V such that

$$G\,V = \varphi,$$

$$[I_k\,N]\,V \geqq \varphi,$$

where I_k is a k-rowed identity matrix and N is a null matrix. Thus, the linear programming problem,

$$\text{Maximize} \quad -\Delta'\,V\,,$$

subject to:

$$G\,V = \varphi,$$

$$[I_k\,N]\,V \geqq \varphi,$$

with the v_i unrestricted in sign (except as restricted by the constraint matrix), has a solution $V = \varphi$. It follows that the dual problem,

$$\text{Minimize} \quad \varphi'\,U \quad , \quad U = \left\{u_1\,,\; u_2\,,\; \ldots,\; u_{k+2}\right\}\,,$$

subject to:

$$\left[G' \;\middle|\; \begin{matrix} I_k \\ N' \end{matrix}\right] U = -\Delta\;,$$

$$u_i \geqq 0 \quad , \quad i = 3,\, 4,\, \ldots,\, k + 2,$$

must have a solution. Note that the first two components of U are unrestricted in sign because of the equalities in the primal problem. The first two components of U are the quantities λ and μ needed for the conditions of the theorem, and the last k components insure the inequalities for condition (2.12.1).

2.13) Theorem: A necessary and sufficient condition for Q to be the unique solution to problem (2.10) is the existence of λ and μ satisfying the following conditions:

$$q_i - p_i + \lambda b_i + \mu \geqq 0, \quad i = 1, 2, \ldots, n, \tag{2.12.1}$$

$$\Sigma\, b_i\, q_i = 0, \tag{2.12.2}$$

$$\Sigma\, q_i = 1, \tag{2.12.3}$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n, \tag{2.12.4}$$

with equality in (2.12.1) for those i such that $q_i > 0$.

Proof: This theorem follows directly from Theorem 2.12.

The difficulty of the problem lies in obtaining Q, λ, and μ satisfying the above conditions. Both methods given here consist generally of solving equations (2.11) without the non-negativity conditions as an initial solution. If the non-negativity conditions are not violated by this solution, then this solution satisfies the conditions of Theorem 2.13. In the case that the non-negativity conditions are violated, both procedures will consist of adjusting the initial solution until the conditions of the theorem are met.

Method I consists of the use of the simplex method of linear programming to adjust the solution. The general idea of the linear programming problem will be to introduce artificial variables which will weaken the inequalities of Theorem 2.13. Under these weakened conditions, a solution will be found. Then, the simplex method will be used to force out the artificial variables, resulting in a solution to the original inequalities. One could use only artificial variables for the initial solution to the weakened inequalities, but the number of iterations will be far less if the solution to equations (2.11) is used for the initial solution.

To solve the equations

$$q_i - p_i + \lambda b_i + \mu = 0 \quad , \quad i = 1, 2, \ldots, n, \tag{2.11.1}$$

$$\Sigma\, b_i\, q_i = 0, \tag{2.11.2}$$

$$\Sigma\, q_i = 1, \tag{2.11.3}$$

the first n equations are summed, giving

$$\lambda \bar{b} + \mu = 0,$$

where

$$\bar{b} = \frac{1}{n} \Sigma\, b_i .$$

Hence,

$$q_i - p_i + \lambda (b_i - \bar{b}) = 0 \quad , \quad i = 1, 2, \ldots, n.$$

These n equations are each multiplied by their respective b_i, summed, and solved using (2.11.2), giving finally,

$$q_i = p_i - (b_i - \bar{b}) \frac{\Sigma p_j b_j}{\Sigma (b_j - \bar{b})^2} , \quad i = 1, 2, \ldots, n. \tag{2.14}$$

If each q_i is non-negative in this solution, then the problem has been solved. If this is not the case, let the q's be ordered so that

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, k,$$

$$q_i < 0 \quad , \quad i = k+1, \ldots, n.$$

Method I consists of solving the following linear programming problem:

$$\text{Minimize:} \quad \sum_{i=k+1}^{n} s_i \tag{2.15}$$

subject to:

$$q_i + \lambda b_i + \mu - u_i = p_i \quad , \quad i = 1, 2, \ldots, k,$$

$$q_i + \lambda b_i + \mu - u_i - s_i = p_i \quad , \quad i = k+1, \ldots, n,$$

$$\sum_{i=1}^{n} b_i q_i - \sum_{i=k+1}^{n} b_i s_i = 0 \quad ,$$

$$\sum_{i=1}^{n} q_i - \sum_{i=k+1}^{n} s_i = 1 \quad ,$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n,$$

$$u_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n,$$

$$s_i \geqq 0 \quad , \quad i = k+1, \ldots, n,$$

with the additional non-linear constraint,

$$\sum_{i=1}^{n} q_i u_i = 0.$$

2.16) <u>Theorem</u>: The solution to problem (2.15) satisfies the conditions of Theorem 2.13.

<u>Proof</u>: Since $s_i \geqq 0$, $i = k + 1, \ldots, n$, if the constraint set contained solutions with $s_i = 0$, $i = k+1, \ldots, n$, the solution of problem (2.15) would have to be of that form. When $s_i = 0$, $i = k+1, k+2, \ldots, n$, the constraints become

$$q_i - p_i + \lambda b_i + \mu - u_i = 0 \quad , \quad i = 1, 2, \ldots, n,$$

$$\Sigma b_i q_i = 0,$$

$$\Sigma q_i = 1,$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n,$$

$$u_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n,$$

and

$$\Sigma q_i u_i = 0.$$

These are precisely the conditions of Theorem 2.13, and thus have a unique solution, which, therefore, must be the solution to problem (2.15).

The initial basic feasible solution is found by evaluating equations (2.14), and setting

$$s_i = -q_i \quad , \quad i = k+1, \ldots, n$$

$$\lambda = -\Sigma p_i b_i / \Sigma (b_i - \overline{b})^2 ,$$

$$\mu = -\lambda \overline{b},$$

and using the original q_i values for i = 1, 2, ..., k. Thus, the basis variables are $\{q_1 , q_2 , \ldots, q_k, s_{k+1} , \ldots, s_n, \lambda , \mu\}$. That this is a basic feasible solution follows from the fact that equations (2.11) have a unique solution. Note that the variables λ and μ are unrestricted in sign in problem (2.15).

The constraint

$$\Sigma q_i u_i = 0$$

is imposed by not bringing q_i into the basis when u_i is already in and not bringing u_i into the basis when q_i is already in. Hadley describes a linear programming approach to this type of problem by using all artificial variables in the initial basic feasible solution and gives a proof that the additional nonlinear constraint will not affect the termination of the simplex procedure. This proof, given on pages 218-219 of Hadley (4), is quite long and will not be given here. It does not, however, depend on what basis is used for the initial solution, and thus applies to the procedure given here.

Method II also consists of evaluating equations (2.14) for an initial solution to the problem. Assume again that the first k components of Q are non-negative and the remainder negative. To find a solution to the conditions of Theorem 2.13, the negative probabilities are set to zero and the first k components are adjusted using the conditions of Theorem 2.13 in much the same manner that equations (2.11) were solved.

Assuming that the first k components of Q will remain non-negative and the remainder will be zero, Theorem 2.13 furnishes the following equations:

$$q_i - p_i + \lambda b_i + \mu = 0 \quad , \quad i = 1, 2, \ldots, k, \qquad (2.17.1)$$

$$\sum_{i=1}^{k} b_i q_i = 0 \quad , \qquad (2.17.2)$$

$$\sum_{c=1}^{k} q_i = 1. \qquad (2.17.3)$$

Summing the first n equations and using (2.17.3) gives

$$1 - \sum_{i=1}^{k} p_i + \lambda \sum_{i=1}^{k} b_i + k\mu = 0.$$

Let

$$A = 1 - \sum_{i=1}^{k} p_i \quad , \quad \overline{b}' = \frac{1}{k} \sum_{i=1}^{k} b_i \, ,$$

so that

$$q_i - p_i + \lambda (b_i - \overline{b}') - \frac{1}{k}A = 0 \quad , \quad i = 1, 2, \ldots, k.$$

These k equations are each multiplied by their respective b_i, summed, and solved with equation (2.17.2), giving

$$-\sum_{i=1}^{k} b_i p_i + \lambda \sum_{i=1}^{k} (b_i - \overline{b}')^2 - A\overline{b}' = 0 \quad ,$$

and thus,

$$q_i = p_i - (b_i - \overline{b}') \frac{\sum_{j=1}^{k} p_j b_j + A\overline{b}'}{\sum_{j=1}^{k} (b_j - \overline{b}')^2} + \frac{1}{k}A \quad , \quad i = 1, 2, \ldots, k. \tag{2.18}$$

If these k probabilities are non-negative and the remaining conditions of Theorem 2.13,

$$q_i - p_i + \lambda b_i + \mu \geqq 0 \quad , \quad i = k+1, \ldots, n,$$

are satisfied with $q_i = 0$, $i = k+1, \ldots, n$, then a solution has been obtained. It is important to note that the λ and μ used to check this condition must be those revised values calculated from equations (2.17).

This procedure can fail to terminate here for two reasons; either some of the probabilities in (2.18) are negative, or the condition (2.13.1) is not satisfied for those probabilities set to zero. In the former case, the new negative probabilities are set to zero and the adjustment procedure is repeated, leaving the original negative probabilities set to zero also. An example will be given of this procedure.

The author has been unable to construct an example in which the other failure occurs, and also unable to prove that it cannot occur. If it were to occur, it would seem reasonable to remove the q_i for which (2.13.1) is violated from the set forced to zero, and repeat the adjustment procedure.

Experience in trying to construct an example for which this procedure fails has led the author to believe that if such problems exist, that one is extremely unlikely to happen onto one in solving reasonably "well structured" problems. It is so much easier to use Method II than Method I that it is certainly recommended to try it first. It is important to note that if the procedure were to fail, the user would know that he had not reached a solution, since the procedure is not terminated until the conditions of Theorem 2.13 are met.

It was pointed out previously that it is not necessary to solve problem (2.10) for each alternative to a_p in order to find a solution to problem (2.1). If the solution to problem (2.10) is given by Q equations (2.14), then

$$\delta(P, Q) = [\Sigma(p_i - q_i)^2]^{\frac{1}{2}}$$

$$= \left\{\Sigma\left[p_i - p_i + (b_i - \overline{b})\frac{\Sigma p_j b_j}{\Sigma(b_j - \overline{b})^2}\right]^2\right\}$$

$$= \left\{\left[\frac{\Sigma p_i b_i}{\Sigma(b_i - \overline{b})^2}\right]^2 \Sigma(b_i - \overline{b})^2\right\}^{\frac{1}{2}}$$

$$= \frac{\Sigma p_i b_i}{\left[\Sigma(b_i - \overline{b})^2\right]^{\frac{1}{2}}} \qquad (2.19)$$

If the equations (2.14) do not give a non-negative solution, the quantity in equation (2.19) is a lower bound on $\delta(P, Q_o)$, since the imposition of the non-negativity conditions can only increase the minimum distance if they have an effect.

The procedure is to calculate the quantity in equation (2.19) for each alternative action and then solve for Q_k corresponding to the action resulting for the minimum of the quantities. If equations (2.14) give a solution for this Q_k, then this is the solution to problem (2.1). If equations (2.14) do not give a non-negative solution, then either method is used to find a solution. If the distance $\delta(P, Q_k)$ is still less than the next largest of the quantities from equation (2.19), then the solution to problem

(2,1) has been obtained. If neither of these is the case, then the Q corresponding to the next largest quantity from (2.19) must be found, and so forth, until it is assured that a solution has been obtained.

Example 2: This example is a continuation of Example 1 of the preceding chapter. From the loss functions and the quantities $f_i(1)$, i = 1, 2, 3, 4, 5, the values

$$b_{ki} = f_i(1)\{L[a_k, \theta_i] - L[a_3, \theta_i]\}$$

are computed.

	θ_1	θ_2	θ_3	θ_4	θ_5
b_{1i}	-32.8	-20.5	0.0	104.0	140.4
b_{2i}	-22.0	- 5.7	9.0	41.6	30.3

From the these values the following calculations are made.

$\bar{b}_1 = 38.2$ $\qquad\qquad$ $\bar{b}_2 = 9.0$

	θ_1	θ_2	θ_3	θ_4	θ_5
$b_{1i} - \bar{b}_1$	-71.0	-58.7	-38.2	65.8	102.2
$b_{2i} - \bar{b}_2$	-31.0	-14.7	0.0	32.6	21.3

$\Sigma(b_{1i} - \bar{b}_1)^2 = 24{,}727$ $\qquad$ $\Sigma p_i b_{1i} = 24.62$

$\Sigma(b_{2i} - \bar{b}_2)^2 = 2{,}693$ $\qquad$ $\Sigma p_i b_{2i} = 11.45.$

Since $11.45/2{,}693 > 24.62/24{,}727$, the solution for which a_1 and a_3 will be equally good is first found.

From equations (2.14), the following initial solution is obtained.

$$
\begin{aligned}
q_1 &= .1 + 71\ (24.62/24{,}727) \\
&= .1 + 71\ (.000995) && = .170 \\
q_2 &= .25 + 58.7\ (.000995) && = .308 \\
q_3 &= .35 + 38.2\ (.000995) && = .388 \\
q_4 &= .25 - 65.8\ (.000995) && = .185 \\
q_5 &= .05 - 102.2\ (.000995) && < 0
\end{aligned}
$$

Since $q_5 < 0$ in this solution, q_5 is set to zero and equations (2.18) are used to find the next solution.

$$A = 1 - \sum_{i=1}^{4} p_i = .05 \qquad \bar{b}'_1 = \tfrac{1}{4}\sum_{i=1}^{4} b_{1i} = 12.7$$

$$
\begin{array}{c|cccc|}
 & \theta_1 & \theta_2 & \theta_3 & \theta_4 \\
\hline
b_{1i} - \bar{b}'_1 & -45.5 & -33.2 & -12.7 & 91.3 \\
\hline
\end{array}
$$

$$\sum_{i=1}^{4} (b_{1i} - \bar{b}'_1) = 11{,}670$$

$$\sum_{i=1}^{4} p_i\, b_{1i} = 17.60.$$

$$q_1 = .1 + (45.5)\left(\frac{17.60 - .0125}{11{,}670}\right) + .0125$$

$$= .1 + (45.5)(.001507) + .0125 = .1810$$

$$q_2 = .25 + (33.2)(.001507) + .0125 = .3125$$

$$q_3 = .35 + (12.7)(.001507) + .0125 = .3816$$

$$q_4 = .25 - (91.3)(.001507) + .0125 = .1249$$

The remaining condition of Theorem 2.13.1,

$$q_5 - p_5 + \lambda b_5 + \mu \geqq 0 \ ,$$

is satisfied since

$$-.05 + (127.7)(.001507) - .0125 > 0.$$

It is now necessary to see if this solution is actually closer to P than the lower bound corresponding to a_2, given by equation (2.19). The distance from P to the above solution is computed to be

$$\delta(P, Q_1) = .172.$$

The lower bound for the distance to the solution corresponding to a_2 is

$$\delta(P, Q_2) = 11.45/\sqrt{2{,}693} = .221.$$

Thus, the solution to the problem has been obtained. A discussion of the use to be made of this solution will be given at the end of the following chapter.

<u>Example 3</u>: This example will illustrate the solution to

problem (2.10) by both Method I and Method II, and also the procedure for the case that Method II does not result in a solution the first time that it is applied. Let

$$P = \{.09,\ .01,\ .15,\ .75\},$$

$$b_1 = -100,\ b_2 = b_3 = b_4 = 100.$$

Then

$$\Sigma p_i b_i = 82 \quad , \quad \Sigma(b_i - \bar{b})^2 = 30{,}000$$

and equations (2.14) give the following solution:

$$q_1 = .09 + 150\ (82/30{,}000) = .500$$

$$q_2 = .01 - 50\ (82/30{,}000) = -.1267$$

$$q_3 = .15 - 50\ (82/30{,}000) = .0133$$

$$q_4 = .75 - 50\ (82/30{,}000) = .6133.$$

When q_2 is set to zero,

$$\sum_{i \neq 2} p_i b_i = 81 \quad , \quad \sum_{i \neq 2} (b_i - \bar{b}') = 26{,}400,$$

and equations (2.18) give the following solution:

$$q_1 = .09 + 133[(81 + .0033)/26{,}400] + .01/3 = .503$$

$$q_2 = 0$$

$$q_3 = .15 - 66[(81 + .0033)/26{,}400] + .01/3 = -.053$$

$$q_4 = .75 - 66[(81 + .0033)/26{,}400] + .01/3 = .550.$$

When q_3 is also set to zero,

$$\sum_{i \neq 2,3} p_i b_i = 66 \quad , \quad \sum_{i \neq 2,3} (b_i - \bar{b}')^2 = 20{,}000,$$

and equations (2.18) give the following solution.

$$q_1 = .09 + 100(66/20{,}000) + .16/2 = .50$$

$$q_2 = 0$$

$$q_3 = 0$$

$$q_4 = .75 - 100(66/20{,}000) + .16/2 = .50.$$

Since

$$-.01 + 100(66/20{,}000) - .16/2 > 0 \ ,$$

$$-.15 + 100(66/20{,}000) - .16/2 > 0 \ ,$$

the remaining conditions of Theorem 2.13 are satisfied and the solution has been obtained.

The linear programming problem which will solve this problem is as follows:

Minimize s_2 ,

subject to:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -100 & 100 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 100 & -100 & 1 & -1 & 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 100 & -100 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 100 & -100 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ -100 & 100 & 100 & 100 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -100 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ \lambda_1 \\ \lambda_2 \\ \mu_1 \\ \mu_2 \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ s_2 \end{bmatrix} = \begin{bmatrix} .09 \\ .01 \\ .15 \\ .75 \\ 1 \\ 0 \end{bmatrix}$$

$$q_i \geqq 0 \quad i = 1, 2, 3, 4$$
$$u_i \geqq 0 \quad i = 1, 2, 3, 4$$
$$s_2 \geqq 0$$

The variables λ_1, λ_2, and μ_1, μ_2 are necessary because the original λ and μ are unrestricted in sign. The quantities in the first solution above are used to determine the basis as $\{q_1, s_2, q_3, q_4, \lambda_1, \mu_2\}$, and the matrix associated with these variables must be inverted. The above set of equations is then multiplied by this inverse to put the equations in the canonical form for the simplex method. (Hadley [4]). The simplex procedure is then carried out in the usual manner, except for the imposition of the constraint $\Sigma\, q_i u_i = 0$, as described previously in this chapter. The tableaus are presented below in a form that should be self-explanatory.

q_1	q_2	q_3	q_4	λ_1	λ_2	μ_1	μ_2	u_1	u_2	u_3	u_4	s	
1	0	0	0	-100	100	1	-1	-1	0	0	0	0	.09
0	1	0	0	100	-100	1	-1	0	-1	0	0	-1	.01
0	0	1	0	100	-100	1	-1	0	0	-1	0	-1	.15
0	0	0	1	100	-100	1	-1	0	0	0	-1	0	.75
1	1	1	1	0	0	0	0	0	0	0	0	-1	1
-100	100	100	100	0	0	0	0	0	0	0	0	-100	0
												1	

1	0	0	0	0	0	0	0	0	0	0	0	0	.5
0	-1	0	0	0	0	0	0	0	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	1	.1267
0	0	1	0	0	0	0	0	0	$\frac{1}{3}$	$-\frac{2}{3}$	$\frac{1}{3}$	0	.0133
0	0	0	1	0	0	0	0	0	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{2}{3}$	0	.0133
0	0	0	0	1	-1	0	0	$\frac{1}{200}$	$-\frac{1}{600}$	$-\frac{1}{600}$	$-\frac{1}{600}$	0	.00273
0	0	0	0	0	0	-1	1	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	.1367
	1								$-\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$		-.1267

Figure 4. Simplex Tableau

q_1	q_2	q_3	q_4	λ_1	λ_2	μ_1	μ	u_1	u_2	u_3	u_4	s	
1	0	0	0	0	0	0	0	0	0	0	0	0	.50
0	-1	-2	0	0	0	0	0	0	0	1	-1	1	.10
0	0	3	0	0	0	0	0	0	1	-2	1	0	.0399
0	0	-1	1	0	0	0	0	0	0	1	-1	0	.60
0	0	$\frac{1}{200}$	0	1	-1	0	0	$\frac{1}{200}$	0	$-\frac{1}{200}$	0	0	.0028
0	0	$-\frac{1}{2}$	0	0	0	-1	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	.13
	1	2								-1	1	0	-.10
1	0	0	0	0	0	0	0	0	0	0	0	0	.50
0	-1	-2	0	0	0	0	0	0	0	1	-1	1	.10
0	-2	-1	0	0	0	0	0	0	1	0	-1	2	.2 99
0	1	1	1	0	0	0	0	0	0	0	0	-1	.50
0	$-\frac{1}{200}$	$-\frac{1}{200}$	0	1	-1	0	0	$\frac{1}{200}$	0	0	$-\frac{1}{200}$	$-\frac{1}{200}$	.00
0	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	-1	1	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$-\frac{1}{2}$	.08
												1	0

Figure 4. (Continued)

CHAPTER III

FURTHER ASPECTS OF THE SENSITIVITY ANALYSIS

Although Chapter II constitutes the basic development of the sensitivity analysis procedure, the results of that chapter provide only a framework within which to approach the problem of ascertaining the sensitivity to the prior distribution. Many of the details of the sensitivity analysis will necessarily depend upon the nature of the specific problem being considered. Some of the general aspects of the application of the procedure will be discussed in this chapter as guidelines for the decision maker.

The Use of the Results

The decision maker may find that it is his "good fortune" to have the solution Q_o be quite dissimilar to P. This should be taken to mean that no reasonable amount of reconsideration about P would lead him to Q_o as the prior distribution. It is important to realize that the "similarity" of probability distributions depends a great deal upon the nature of the information upon which the prior distributions are based. For example, one may feel much more certain about a prior distribution based upon a great amount of historical evidence than about one based entirely upon his "feelings."

In the case that the decision maker does feel that Q_o is quite dissimilar to P and if the distance function can be granted to be a reasonable one, then a_p is very likely the proper action. It is important that the distance function have this reasonability, for otherwise there might be some distribution Q with $\delta(P, Q) > \delta(P, Q_o)$ which would seem more similar to P than does Q_o, which would not have been considered for its optimal action.

Except for a subsequent discussion of the role of the distance function, little more will be said regarding the case that Q_o is quite dissimilar to P. This is not because this case is unimportant, for the possibility of this kind of information is the primary reason for conducting the sensitivity analysis. It merely seems that this case is quite easy to interpret and needs little clarification.

What, then, can be decided in the event that Q_o is quite similar to P? First, if nothing more were decided, the sensitivity analysis would not have been in vain. This type of information ranks in the same order of importance as determining a_p in the first place. The primary purpose of a mathematical model is to lend insight into the structure of a phenomenon, not merely to provide a procedure for attaining optimums.

In at least one sense, the outcome is easier to interpret in this case than when Q_o is dissimilar to P. The decision maker is assured that the sensitivity analysis has performed its function properly because it makes little difference whether the distance function is really a reasonable one or not. This is because the decision maker has exhibited a prior distribution resulting in a different action which, in his consideration, is near enough to P to warrant attention. Any other distance function which is more "reasonable" would have to yield a solution which is at least as "similar" to P as Q_o.

One of the first things that should be noted is that if Q_o is similar to P, then, under either of these distributions, the difference of the expected losses of the best actions against these two distributions is probably relatively small. If the discriminatory power of expected loss is not significant, the decision maker should probably consider other decision making criteria. It seems to the author that other criteria should be seriously considered in any case. In the first place, if θ is truly a random variable, the decision maker is quite likely not really interested in what happens in many repetitions, but what happens in the next realization of the random variable. In the second place, for many applications θ is not actually a random variable in the usual sense, but merely an unknown, the expectation being taken over the decision maker's uncertainty. In either case, some attention should be given to a minimax type consideration of the alternatives.

If the minimax criterion is used to decide between the two actions which are best against P and Q_o, a great deal of the objection to the minimax procedure is overcome, because the expected losses of the two actions do not differ drastically. The use of the minimax procedure in this situation would be a combination minimax and Bayes procedure. Such a philosophy of decision making has attracted the attention of several writers and the reader is referred to Lehmann and Hodges (7), Wesler (17), Good (3), and Smith (15) for presentation of their approaches. The procedure described above seems to be of a different nature than any presented by these authors. Further development of these ideas will not be attempted here except to point out that more than two actions can be compared in the same manner by finding the action which is best against the second nearest distribution to P and so forth. This would probably be a desirable procedure in the event that the second nearest "critical" distribution was also quite similar to P.

Further Experimentation

There is another action in addition to those normally in A which should be considered by the decision maker at this point. This is the option of performing additional experimentation before choosing a terminal action. It is sometimes possible to formally include this alternative in the action space and evaluate it along with the other actions. Generally, however, this approach results in computational procedures which are prohibitive.

The basic idea to be considered by the decision maker at this point is that the weight given to the prior distribution in the application of Bayes rule is inversely related to the amount of experimental evidence. This is because the likelihood function, in expectation, becomes more peaked about the true value of θ as experimentation is continued. Because of this phenomenon the sensitivity of the result to the prior distribution decreases with further experimentation.

A quantitative analysis of this phenomenon becomes very unwieldly because the experimental outcomes depend upon which state of nature is true and that, of course, is unknown. To take the expectation over the states of nature by using the prior distribution seems to miss the point entirely for this type of analysis, because the whole idea of the convergence with additional experimentation is based on the fact that there is one state of nature that is true. Even in the case that the state of nature is actually a random variable, one is usually concerned with the situation in which all of the experimentation is carried out under one realization of this random variable.

It will suffice, it is hoped, for the purpose at hand, to state the basic result that, whatever the prior distribution,

$$\lim_{n\to\infty} \Pr\left\{\frac{P[\theta_i \mid x_1, x_2, \ldots, x_n]}{P[\theta_j \mid x_1, x_2, \ldots, x_n]} > M\right\} = 1,$$

for any fixed M, where θ_i is the true state of nature and θ_j is any other. The proof of this result can be found in Savage (13). The convergence is based on the law of large numbers, with the rate of convergence depending on the quantities

$$E\left[\ln \frac{f_i(x)}{f_j(x)}\right] \qquad j = 1, 2, \ldots, n; \; j \neq i.$$

It appears that the examination of the rate of convergence can be used in the consideration of some problems, but such a procedure would be much too involved to investigate in a general manner here.

The Role of the Distance Function

Some discussion has been presented in preceding sections regarding the role of the distance function; namely, that the importance of the function depends upon whether the solution to the problem is similar to P. The author has been unable to locate in the literature work of a general nature which would give insight into the question of choosing a reasonable distance function.

The primary weakness of the Euclidean norm is that a change in probability from .00 to .05 weighs exactly as heavily as a change from .55 to .60. It appears to be agreed upon by all who have considered the problem that it would be more reasonable to have a distance function based more upon some sort of relative change, weighing a change from .00 to .05 more heavily than a change from .55 to .60. Functions based upon relative change, however, are generally not symmetric, and it is questionable whether the change should be relative to P or relative to the solution.

A solution procedure will now be given for a distance function which is a significant generalization of the Euclidean norm, and which would seem to meet nearly any need that might arise in practice. This function is

$$\delta(P, Q) = \left[\sum \frac{(p_i - q_i)^2}{c_i}\right]^{\frac{1}{2}}, \tag{3.1}$$

where c_i, $i = 1, 2, \ldots, n$ is any set of positive numbers.

This function has many merits. If it is desired to make the function depend upon the relative change in the probabilities, then the c_i can be chosen such that

$$\delta(P, Q) = \left[\sum \frac{(p_i - q_i)^2}{p_i}\right]^{\frac{1}{2}}. \tag{3.2}$$

In the event that the decision maker has different feelings of confidence in his estimate of the probability for different states of nature, he can assign the c_i accordingly. It may also be helpful in interpreting the results of the analysis to vary the c_i. Suppose, for example, that the solution using the Euclidean norm is such that $|p_k - q_k|$ is large for some k and $|p_i - q_i|$ is small for $i \neq k$. By solving the problem again with $c_k < 1$, the decision maker can effectively restrict the size of $|p_k - q_k|$ and see what effect this will have on the other deviations. It might also be pointed out that assigning relatively small c_i for those p_i which are small will have the effect of making the first trial solution less likely to violate the non-negativity conditions.

Since the function (3.1) is convex, Theorem 2.8 can be applied, resulting in the formulation of the problem as,

$$\text{Minimize} \quad \delta(P, Q) \tag{3.3}$$

subject to:

$$\Sigma\, b_i\, q_i = 0 \,, \tag{3.3.1}$$

$$\Sigma\, q_i = 1 \,, \tag{3.3.2}$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n. \tag{3.3.3}$$

Theorem 2.12 can be applied to this problem, giving as a necessary and sufficient condition for Q to be the solution the existence of λ and μ such that

$$\frac{q_i - p_i}{c_i} + \lambda b_i + \mu \geqq 0 \quad , \quad i = 1, 2, \ldots, n, \tag{3.4.1}$$

$$\Sigma\, b_i\, q_i = 0 \,, \tag{3.4.2}$$

$$\Sigma\, q_i = 1 \,, \tag{3.4.3}$$

$$q_i \geqq 0 \quad , \quad i = 1, 2, \ldots, n, \tag{3.4.4}$$

with equality in (3.4.1) for those i such that $q_i > 0$. Assuming that equality holds in (3.4.1) for each i, these equations can be solved in a manner similar to equations (2.11), giving

$$q_i = p_i - c_i\left(b_i - \frac{\Sigma c_j b_j}{C}\right) \frac{\Sigma p_j b_j}{\Sigma C_j b^2_j - \frac{1}{C}(\Sigma c_j b_j)^2} ,$$

$$i = 1, 2, \ldots, n, \tag{3.5}$$

where $C = \Sigma c_i$. If in this solution, some of the non-negativity conditions are violated, methods nearly identical with either Method I or Method II of Chapter II can be used to obtain a solution. Since the c_i are fixed, the simplex procedure can be carried out with merely a modification of the coefficients. Method II is carried out in exactly the manner as in Chapter II. A lower bound on the distance from P to the solution is given by

$$\delta(P, Q) = \frac{\Sigma p_i b_i}{\left[\Sigma c_i b^2_i - \frac{1}{C}(\Sigma c_i b_i)^2\right]^{\frac{1}{2}}} . \tag{3.6}$$

No extensive comparison has been carried out of the results of using various distance functions. Example 2 of Chapter II has been solved using (3.2), giving the following comparison. If

$$\delta(P, Q) = \left[\Sigma (p_i - q_i)^2\right]^{\frac{1}{2}},$$

then

$$Q_o = \{.18,\ .31,\ .38,\ .13,\ .00\} ,$$

whereas, if

$$\delta(P, Q) = \left[\sum \frac{(p_i - q_i)^2}{p_i} \right]^{\frac{1}{2}} ,$$

then

$$Q_o = \{.14,\ .33,\ .42,\ .10,\ .01\} .$$

Other Types of Decision Problems

The actual construction of Bayes procedures is quite dependent upon the structure of the spaces A and Θ, and this dependency carries over into the sensitivity analysis. In this paper, the primary consideration was intended to be for the case in which both of these spaces are finite. As pointed out in the introduction, however, it is often possible to approximate other types of problems arbitrarily closely within this frame of reference. Consideration will be given in the following chapters to the cases in which A is finite and Θ is an interval on the real line, and to a case in which both A and Θ are intervals.

CHAPTER IV

APPROXIMATE SOLUTIONS FOR CONTINUOUS PARAMETERS

It has been pointed out that the procedure of Chapter II can be used to obtain arbitrarily close approximations to solutions in certain problems in which the state of nature is represented by a continuous variable. This chapter will be devoted to describing the approximation procedure and proving the above statement.

The Convergence Theorem

Because the approximation procedure will be applied to several situations, the primary theorem will be proved for a rather general problem. This problem is:

$$\text{Minimize} \quad \int_a^b [p(\theta) - q(\theta)]^2 \, d\theta \tag{4.1}$$

subject to:

$$\int_a^b q(\theta)b(\theta)d\theta = 0 \quad ,$$

$$\int_a^b q(\theta)d\theta = 1 \quad ,$$

$$q(\theta) \geqq 0 \quad [a.e.] \quad ,$$

where

$$\int_a^b p(\theta)d\theta = 1 \quad , \; p(\theta) \geqq 0 \quad , \int_a^b p(\theta)b(\theta)d\,\theta > 0 \; ,$$

$b(\theta)$ is continuous and $p(\theta)$ is continuous almost everywhere on $[a,b]$; there exist θ_1, $\theta_2\,\varepsilon[a,b]$ such that $b(\theta_1) > 0$, $b(\theta_2) < 0$; and integration is in the Lebesgue sense.

Consider the sequence of partitions

$$[a = a_{no} < a_{n1} < \ldots < a_{n2^n} = b] \quad , \quad n = 2, 3, \ldots,$$

each dividing $[a,b]$ into equal sub-intervals. Let

$$p_n(\theta) = \int_{a_{n,i-1}}^{a_{ni}} p(\theta)d\theta \; , \qquad a_{n,i-1} \leqq \theta < a_{ni},$$

$$n = 2, 3, \ldots,$$

$$b_n(\theta) = \int_{a_{n,i-1}}^{a_{ni}} b(\theta)d\,\theta , \qquad i = 1, 2, \ldots, 2^n .$$

Since $\int_a^b p_n b_n$ is a continuous linear functional on L_2 space (Munroe [9]), it follows

$$\lim_n \int_a^b p_n(\theta)b_n(\theta)d\theta = \int_a^b p(\theta)b(\theta)d\theta \; .$$

Let r be such that $\int_a^b p_n b_n > 0$ for any $n > r$.

It follows from the continuity of b that there exists an s such that, for any $n > s$, there exist i and j such that

$$b_n(\theta) > 0 \quad , \quad a_{n,i-1} \leqq \theta < a_{ni} \quad ,$$

$$b_n(\theta) < 0 \quad , \quad a_{n,j-1} \leqq \theta < a_{nj} \; .$$

Let t be the larger of r and s and

$$p_{ni} = p_n(\theta) \quad , \qquad a_{n,i-1} \leqq \theta \leqq a_{ni} \; ,$$

$$n = t, t+1, \ldots ,$$

$$b_{ni} = b_n(\theta) \quad , \qquad i = 1, 2, \ldots, 2^n \, .$$

Then the problem

$$\text{Minimize} \quad \sum_i (p_{ni} - q_{ni})^2$$

subject to:

$$\sum_i q_{ni} b_{ni} = 0 \; ,$$

$$\sum_i q_{ni} = 2^n \; ,$$

$$q_{ni} \geqq 0 \quad , \quad i = 1, 2, \ldots, 2^n \; ,$$

satisfies, for any $n > t$, the hypotheses of Theorem 2.12, with the minor modification that $\Sigma\, q_{ni} = 2^n$, and can be solved by the methods of Chapter II. The solution will also be the solution to the problem

$$\text{Minimize} \int_a^b [p_n(\theta) - q_n(\theta)]^2\, d\theta \tag{4.2}$$

subject to:

$$\int_a^b b_n(\theta) q_n(\theta) d\theta = 0 \quad ,$$

$$\int_a^b q_n(\theta) = 1 \quad ,$$

$$q_n(\theta) \geqq 0 \quad .$$

The sequence of solutions to the above problem will be denoted $\langle q_n \rangle$, $n = t, t+1, \ldots$.

4.3) Lemma: For the sequence $\langle p_n \rangle$, $\langle q_n \rangle$ defined above, $\lim\limits_n \delta(p_n, q_n)$ exists, where

$$\delta(p, q) = \left[\int_a^b [p(\theta) - q(\theta)]^2 \ d\theta\right]^{\frac{1}{2}} .$$

Proof: Due to the way in which b_{ni} was defined, it can be easily verified that, for any $n = t, t+1, \ldots,$ the solution q_n satisfies the constraints imposed upon the solution q_{n+1}. Thus, for any m, n; $m > n$,

$$\delta(p_m, q_m) \leqq \delta(p_m, q_n) \quad .$$

Thus,

$$\delta(p_m, q_m) \leqq \delta(p_n, q_n) + \delta(p_n, p_m) \ ;$$

and since

$$\lim_{n,m} \delta(p_n, p_m) = 0$$

and for any p and q

$$\delta(p, q) \geqq 0 \ ,$$

then the limit in question exists.

4.4) <u>Lemma</u>: For the sequence $\langle q_n \rangle$ and the distance function described above $\lim\limits_{m,n} \delta(q_n , q_m) = 0$.

<u>Proof</u>: It will be convenient to use the notation

$$\delta_m(P, Q) = \left[\sum_{i=1}^{m} (p_i - q_i)^2\right]^{\frac{1}{2}} ,$$

where P and Q are vectors in m-space. It follows that if

$$p(\theta) = p_i \ ,$$

$$q(\theta) = q_i \ , \quad \frac{b-a}{m}(i-1) \leqq \theta < \frac{b-a}{m} i \ ; \ i = 1, 2, \ldots, m,$$

then

$$\delta(p, q) = \sqrt{\frac{b-a}{m}} \ \delta_m(P, Q).$$

Let $\lim\limits_{n} \delta(p_n, q_n) = d$, and for an arbitrary positive ε, let

N be such that for $m > N$, $n > N$,

$$|\,\delta(p_n, q_n) - d\,| < \varepsilon/2 \ ,$$

$$\delta(p_n, p_m) < \varepsilon/2 \ .$$

Then, for $m > n$,

$$\delta(p_m, q_n) \leqq \delta(P_m, P_n) + \delta(p_n, q_n)$$
$$< d + \varepsilon,$$

and, as pointed out in the preceding proof,

$$\delta(p_m, q_m) \leqq \delta(p_m, q_n) \ .$$

Define the vectors in 2^m-space,

$$Q_m = \{q_{m1}, q_{m2}, \ldots, q_{m2^m}\} \ ,$$

$$P_m = \{p_{m1}, p_{m2}, \ldots, p_{m2^m}\} \ ,$$

and Q'_n is such that the first 2^{m-n} components are q_{n1}, the second 2^{m-n} components are q_{n2}, and so forth. The plane determined by these three vectors is represented in Figure 5, and the distances indicated are in terms of the Euclidean norm in 2^m-space.

The circle of radius $\delta_{2^m}(P_m, Q'_n)$ is to represent the possible locus of the vector Q'_n. The proof of the lemma consists of showing that due to the manner in which these vectors are obtained, Q'_n must lie on the arc AB, where the line AB is orthogonal to $Q_m - P_m$. If Q'_n were not on the arc

AB, there would be a convex linear combination R of Q'_n and Q_m with $\delta_{2m}(P_m, R) < \delta_{2m}(P_m, Q_m)$. This would contradict the minimality of $\delta_{2m}(P_m, Q_m)$ because, as pointed out previously, Q'_n satisfies the constraints for the solution of Q_m, and since the set of vectors satisfying the constraints is convex, R also satisfies the constraints for the solution of Q_m.

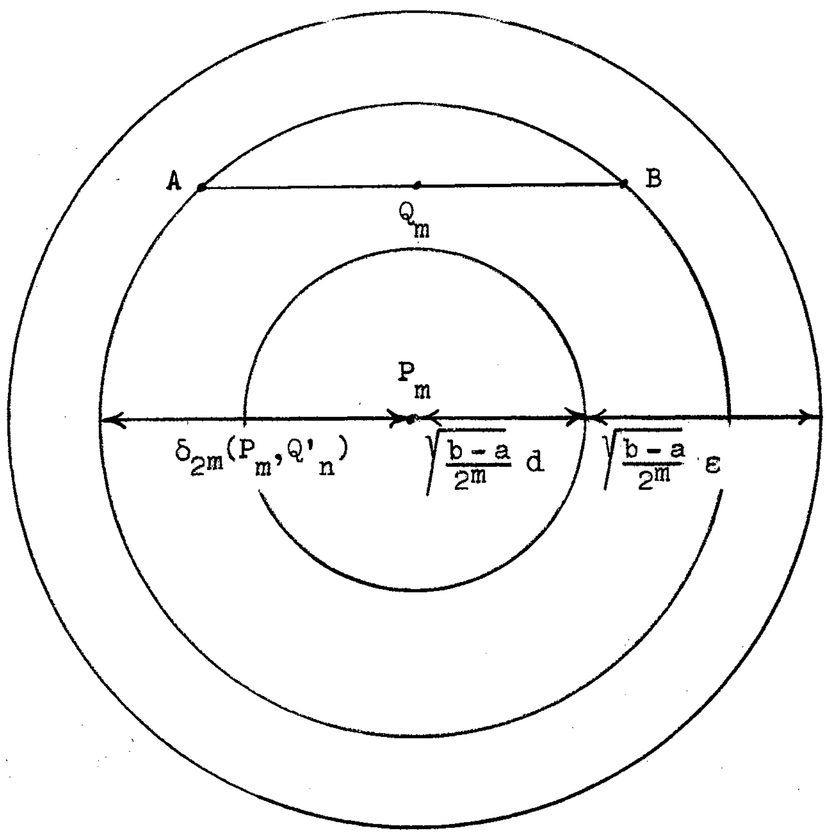

Figure 5. A Plane in 2^m-Space

The greatest possible distance from Q_m to a point on the arc AB would be the distance from Q_m to A when $| \delta_{2m}(P_m, Q_m) - \delta_{2m}(P_m, Q'_n) | = \varepsilon$. Thus

$$\delta_{2m}(Q_m, Q'_n) \leqq \left\{\frac{b-a}{2^m}\left[(a+\varepsilon)^2 - \varepsilon^2\right]\right\}^{\frac{1}{2}}$$

$$= \left\{\frac{b-a}{2^m}\left[\varepsilon^2 + 2a\varepsilon\right]\right\}^{\frac{1}{2}}$$

and

$$\delta(q_m, q_n) \leqq [\varepsilon^2 + 2a\varepsilon]^{\frac{1}{2}}.$$

Thus, by choosing ε small enough, $\delta(q_n, q_m)$ can be made arbitrarily small.

4.5) Lemma: There exists a function $q(\theta)$ such that $\lim_n \delta(q_n, q) = 0$ and q satisfies the constraints of problem (4.1).

Proof: The preceding lemma, due to the completeness of L_2 space, shows that the function q exists. Using again the continuity of linear functionals on L_2 space,

$$\int_a^b b(\theta)q(\theta)d\theta = \lim_n \int_a^b b_n(\theta)q_n(\theta)d\theta = 0 \quad ,$$

$$\int_a^b q(\theta)d\theta = \lim_n \int_a^b q_n(\theta)d\theta = 1 \ .$$

To show that $q(\theta) \geqq 0$ [a.e.], use is made of the fact that

$\lim_n \delta(q_n, q) = 0$ implies that there exists a subsequence of $\langle q_n \rangle$ that converges to q[a.e.], and since $q_n \geqq 0$ for every n, then $q(\theta) \geqq 0$ [a.e.].

4.6) Theorem: Let the sequence of solutions to problem (4.2) be $\langle q_n \rangle$, $n = t, t+1, \ldots$. Then

4.6.1) there exists a function $q(\theta)$ such that $\lim_n q_n = q$ [mean square], and

4.6.2) this function $q(\theta)$ is a solution to problem (4.1).

Proof: The result (4.6.1) has been established through the preceding lemmas. In light of Lemma 4.5, the remainder of the theorem can be proved by showing that q is actually the closest distribution that satisfies the constraints. Suppose that there exists a q_o satisfying the constraints of problem (4.1) and

$$\delta(p, q_o) < \delta(p, q) .$$

Consider the sequence of functions $\langle q_{on} \rangle$, $n = t, t+1, \ldots$,

$$q_{on}(\theta) = \int_{a_{n,i-1}}^{a_{ni}} q_o(\theta) d\theta \quad , \quad a_{n,i-1} \leqq \theta < a_{ni} \; ,$$
$$i = 1, 2, \ldots, 2^n.$$

Clearly

$$\int_a^b q_{on}(\theta) d\theta = 1 \quad ,$$

$$q_{on}(\theta) \geqq 0 \quad , \quad n = t,\ t+1,\ \ldots$$

and again using the continuity of linear functionals, for any $\varepsilon > 0$, there exists an N such that for $n > N$

$$\left| \int_a^b q_{on}(\theta) b(\theta) d\theta \right| < \varepsilon .$$

For any $n > N$, the simple function q'_{on} can be found such that

$$q'_{on}(\theta) = q_{on}(\theta) \quad , \quad a_{n,i-1} \leqq \theta < a_{n_i} \ ; \ i \neq j,\ k$$

and

$$\int q'_{on}(\theta) b(\theta) d\theta = 0$$

$$\int q'_{on}(\theta) d\theta = 1$$

by solving for $q'_{on}(\theta)$, $a_{n,i-1} \leqq \theta < a_{ni}$, $i = j, k$. For a simple function this is merely a matter of solving for these two unknowns the two linear equations represented by the two integral equations above. It can be easily shown that for any $\theta \varepsilon\ [a,b]$,

$$| q'_{on}(\theta) - q_{on}(\theta) | < \frac{\varepsilon}{|b_{nj} - b_{nk}|}$$

in this solution. Thus, by choosing N sufficiently large, $\delta(q_{on}, q'_{on})$ for $n > N$ can be made arbitrarily small.

Let

$$\varepsilon = \delta(p, q) - \delta(p, q_o)$$

and N be large enough that for n > N,

$$\delta(q_o, q_{on}) < \varepsilon/5$$

$$\delta(q_{on}, q'_{on}) < \varepsilon/5$$

$$\delta(p, p_n) < \varepsilon/5$$

$$\delta(q, q_n) < \varepsilon/5.$$

Then,

$$\begin{aligned} \delta(p_n, q'_{on}) &< \delta(p_n, q_{on}) + \varepsilon/5 \\ &< \delta(p, q_{on}) + 2\varepsilon/5 \\ &< \delta(p, q_o) + 3\varepsilon/5 \\ &= \delta(p, q) - 2\varepsilon/5 \\ &< \delta(p, q_n) - \varepsilon/5 \\ &< \delta(p_n, q_n) \end{aligned}$$

which contradicts the minimality of $\delta(p_n, q_n)$.

It is rather unfortunate that the b_{ni} had to be so carefully chosen in order to prove the theorem. It is clear, however, from consideration of the conditions of Theorem 2.12, that the solution is a continuous function of the b_{ni} in some neighborhood of the correct values, and thus approximate values can probably be used without serious consequences.

Application of the Approximation Procedure

The decision maker does not, in general, actually compute a sequence of solutions in finding an approximate solution. He simply divides the interval [a,b] into subintervals small enough to achieve what he considers a satisfactory approximation. The above theorem establishes that this approximation will be a reasonable one.

In the case that the action space is finite and the state of nature space is represented by an interval, the approximation procedure is applied in essentially the same manner as the procedure of Chapter II. The loss function and the likelihood function are continuous functions of a parameter representing the state of nature, and are used to form the function $b(\theta)$ in the same manner that the quantities b_i were determined in Chapter II. That is,

$$b(\theta) = \{L[a_k, \theta] - L[a_p, \theta]\} f(x; \theta) .$$

If the procedure is to be used for an a priori analysis, the function $b(\theta)$ is the risk function. For example, consider the choice between sampling plans A and B with OC curves as shown in Figure 6.

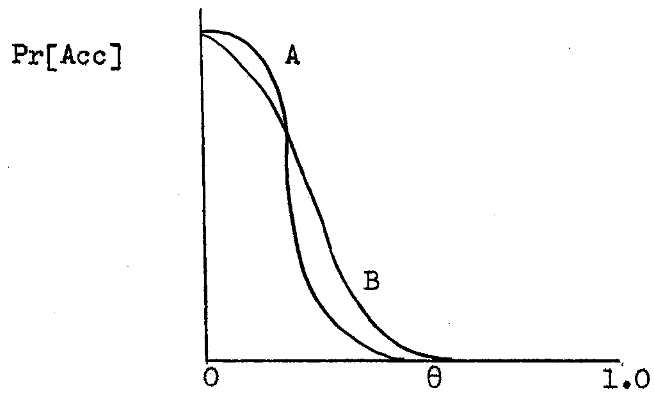

Figure 6. The OC Curves for Two Sampling Plans

An OC curve is a risk function in which L[accept, θ] = 1, and L[reject, θ] = 0 for all θ. If the decision maker wishes to accept lots with $\theta < \theta_o$ and reject lots with $\theta \geqq \theta_o$, then a reasonable (and Bayes) criterion for choosing between **A** and B would be to minimize

$$k_1 \int_0^{\theta_o} \{1 - \Pr[acc, \theta]\}\, p(\theta) d\theta + k_2 \int_0^1 P_r[Acc, \theta] p(\theta) d\theta \ ,$$

where $p(\theta)$ is the prior distribution. The risk function should be, then,

$$R(d, \theta) = k_1\{1 - Pr_d[Acc, \theta]\} \ , \quad \theta < \theta_o \ ,$$

$$= k_2 Pr_d[Acc, \theta] \ , \qquad \theta \geqq \theta_o \ , \ d = A,B.$$

It might be the usual procedure for the decision maker to use for the prior distribution a Beta distribution with parameters based upon past performance of the production facilities. The approximate sensitivity analysis procedure could be used to determine how much the prior distribution would have to change to result in the choice of a different sampling plan.

The remaining type of decision problem to which the approximate procedure will be applied is the case in which both the spaces A and Θ are intervals. The only problem within this situation which will be considered is the estimation problem with a quadratic loss function, and this will be the topic of the next chapter.

CHAPTER V

THE SENSITIVITY OF THE POSTERIOR MEAN

In many applications of decision theory, the choice of a best action depends only upon the mean of the posterior distribution. These applications include problems in which the loss functions are linear in a parameter representing the state of nature, and estimation problems where the loss function is proportional to the mean squared error of the estimate. The purpose of this chapter is to develop a procedure for determining the sensitivity of the posterior mean to the prior distribution. It will be shown that the procedure of Chapter II can, with minor modifications, be applied to this problem.

Formulation of the Problem

In this chapter the state of nature will be taken as some value of a real variable in an interval [a,b]. If this variable can take on only one of a finite number of values, a modification of the procedure of Chapter II can be applied directly. If the parameter representing the state of nature is considered as a continuous variable, the approximation procedure of the preceding chapter can be applied to reduce the problem to one in which there are only a finite number of states of nature.

The posterior mean is the expected value under the posterior distribution of the state of nature parameter. An analogy will be made between the possible values of the posterior mean and the action space considered in the earlier chapters. Because, however, the posterior mean can take on any value in the interval [a,b], the approach will have to be considerably modified, since the finiteness of the action space was essential to the previous approach.

As an approach to this situation, the following problem will be considered. Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ such that

$$a = \theta_1 < \theta_2 < \ldots < \theta_n = b \ .$$

Considering the experimental outcome as fixed, let the posterior mean be denoted by μ_q for any prior distribution Q. As before let the estimated prior distribution be P. The problem to be considered is:

5.1) Given some $\Delta > 0$,

$$\text{Minimize} \quad \delta(P, Q)$$

subject to the condition on the prior distribution Q that

$$\mu_q \geqq \mu_p + \Delta \ . \qquad (5.1.1)$$

The condition (5.1.1) can be written as

$$\Sigma\, b_i q_i \geqq 0 \ ,$$

where

$$b_i = f_i(x)(\theta_i - \mu_p - \Delta) ,$$

since

$$\sum \frac{q_i f_i(x)}{\sum_j q_j f_j(x)} \theta_i \geqq \mu_p + \Delta$$

is equivalent to

$$\sum q_i[f_i(x)(\theta_i - \mu_p - \Delta] \geqq 0 .$$

The following theorem will show that the above problem can be solved by replacing the inequality in (5.1.1) with equality, and that the procedures of Chapter II can be used to solve the problem.

5.2) Theorem: If $\delta(P, Q)$ is a non-negative, strictly convex function of Q, with $\delta(P, P) = 0$, problem (5.1) has the same solution as the problem:

$$\text{Minimize} \quad \delta(P, Q)$$

subject to:

$$\Sigma b_i q_i = 0 ,$$

$$\Sigma q_i = 1 ,$$

$$q_i \geqq 0 , \quad i = 1, 2, \ldots, n ,$$

where

$$b_i = f_i(x)(\theta_i - \mu_p - \Delta) .$$

<u>Proof</u>: If $\mu_r > \mu_q > \mu_p$ implies $\delta(P,R) > \delta(P,Q)$, where R and Q are the distributions nearest P having posterior means μ_r and μ_q, respectively, then the theorem is true. Let

$$S = \alpha P + (1 - \alpha) R,$$

where

$$\frac{\alpha \Sigma p_i f_i(x)}{\alpha \Sigma p_i f_i(x) + (1 - \alpha) \Sigma r_i f_i(x)} = \frac{\mu_r - \mu_q}{\mu_r - \mu_p} .$$

Then

$$\mu_s = \sum_i \frac{[\alpha p_i + (1 - \alpha) r_i] f_i(x) \theta_i}{\sum_j [\alpha p_j + (1 - \alpha) r_j] f_j(x)}$$

$$= \left[\frac{\alpha \Sigma p_j f_j(x)}{\Sigma [\alpha p_j + (1 - \alpha) r_j] f_j(x)}\right] \frac{\Sigma p_i f_i(x) \theta_i}{\Sigma p_j f_j(x)}$$

$$+ \left[\frac{(1 - \alpha) \Sigma r_j f_j(x)}{\Sigma [\alpha p_j + (1 - \alpha) r_j] f_j(x)}\right] \frac{\Sigma r_i f_i(x) \theta_i}{\Sigma r_j f_j(x)}$$

$$= \left[\frac{\mu_r - \mu_q}{\mu_r - \mu_p}\right] \mu_p + \left[1 - \frac{\mu_r - \mu_q}{\mu_r - \mu_p}\right] \mu_r$$

$$= \mu_q .$$

But, from the minimality of $\delta(P,Q)$, and since $0 < \alpha < 1$,

$$\delta(P,Q) \leqq \delta(P,S) < \delta(P,R) ,$$

and thus the theorem.

The problem is thus reduced to the same form as that considered previously in this paper, and all the theorems and procedures apply. It might be noted that the sufficient condition for the above problem to have a solution, i.e., that there are b_i both positive and negative, simply means in this application that $\mu_p + \Delta < \theta_n$. There is also, of course, an analogous problem and theorem for decreasing the posterior mean, which can be obtained by letting $\theta'_i = - \theta_i$, $i = 1, 2, \ldots, n$.

Application of the Procedure

So far nothing has been indicated regarding the choice of Δ. In the case that the application is to a problem with a finite number of actions, each with a linear loss function, the interval [a,b] will be divided into a finite number of sub-intervals such that if the posterior mean falls in a particular sub-interval, then the corresponding action is best. In this case, the choice of Δ is clearly the distance from μ_p to the next sub-interval on either side.

One of the most important decision problems based upon the posterior mean is the estimation of the parameter θ. A Bayes estimate is one which minimizes the expected loss under the posterior distribution, where the loss function depends on the true value of θ and the estimated value. If the loss function is proportional to the mean squared error of the estimate, then the Bayes estimate is obviously the mean of the posterior distribution.

In this application the choice of Δ is not so easy. In choosing Δ the decision maker is being called upon to specify how much the estimate would have to change in order to make a significant difference to him. This may or may not be an easy question, but it is an age-old question which the statistician must face in nearly all problems.

Some results have been obtained in studying the solution as a function of Δ. The equations are rather unwieldly and may not be useful. If the trouble were justified by the problem, however, the following results could be used to graph each q_i as a function of Δ in some neighborhood of zero.

It is clear that if $Q(\Delta)$ is the solution to problem (5.1), considered as a function of Δ, that

$$\lim_{\Delta \to 0} \delta[P, Q(\Delta)] = 0 .$$

It can also be easily seen that the distance from P to the nearest boundary of the space of all possible prior distributions is

$$\min_i \left\{ \sqrt{\frac{n}{n-1}}\, p_i \right\} .$$

Thus, there exists a Δ_0 such that for any $\Delta \varepsilon\ (0, \Delta_0)$, the solution for $Q(\Delta)$ can be found from

$$q_i = p_i - (b_i - \bar{b}) \frac{\Sigma\, p_j\, b_j}{\Sigma\, (b_j - \bar{b})^2}\ , \quad i = 1,\ 2,\ \ldots,\ n\ ,$$

where

$$b_i = f_i(x)(\theta_i - \mu_p - \Delta).$$

That is to say that none of the q_i in these equations will be negative.

This solution can be written as a function of Δ, giving

$$q_i = p_i - (\alpha_i - \Delta\, \beta_i) \frac{\Delta\, \Sigma\, p_j\, f_j\, (x)}{\Sigma\, (\alpha_j - \Delta\, \beta_j)^2}\ ,$$

where

$$\alpha_j = f_j(x)(\theta_i - \mu_p) - \frac{1}{n} \sum_k f_k(x)(\theta_k - \mu_p)\ ,$$

$$\beta_j = f_j(x) - \frac{1}{n} \sum_k f_k(x)\ .$$

By applying equation (2.19), it can be seen that this solution is non-negative provided that

$$\frac{\Delta\, \Sigma\, p_i\, f_i(x)}{\sqrt{\Sigma(\alpha_i - \Delta\, \beta_i)^2}} \leq \min_i \left\{ \sqrt{\frac{n}{n-1}}\, p_i \right\}\ .$$

The theory of this chapter can be used to investigate the sensitivity of other aspects of the posterior distribution by replacing the θ_i with "pseudo variables," whose posterior means are of interest. For example, the sensitivity of Bayesian confidence intervals can be studied through the use of characteristic functions of certain subsets of Θ. That is, if

$$\varphi_A(\theta) = 1 \quad \text{if} \quad \theta \,\varepsilon\, A \subset \Theta \,,$$

$$\varphi_A(\theta) = 0 \quad \text{if} \quad \varepsilon\, \Theta - A \,,$$

then

$$E\ \varphi_A(\theta) = \text{Prob. } [\theta \,\varepsilon\, A].$$

Thus, the posterior mean of φ_A is the confidence level that the true parameter lies in the set A.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The primary idea upon which this work has been based is that in many applications of Bayesian inference, it is not necessary to know the prior distribution exactly. The space of all possible prior distributions is divided, according to the nature of the problem, into mutually exclusive subsets, and it is necessary only to know in which subset the prior distribution is contained. Because, however, it is difficult to characterize or describe, in a meaningful way for practical considerations, subsets of an n-dimensional space when n is more than two or three, little use has been made of this idea. The approach used in this paper is to find the "distance" from an estimated prior distribution P to the nearest boundary of the subset containing it.

For certain distance functions in the space of all possible prior distributions, a procedure for solving the above problem has been developed, and this procedure is relatively easy to apply. Exact solutions are obtained when the number of states of nature is finite, and basically the same procedure gives apparently satisfactory approximations when the state of nature space is an interval on the real line. It is implicit in the approach to the problem that the number of possible actions is finite, but certain other cases, such as estimation problems, can be approached by approximating the action space with a finite space. The approximations, both in the action space and the parameter (state of nature) space, do not appear to be really serious drawbacks. For a great many problems in which continuous variables are used, the continuity is merely an abstraction, introduced for convenience. When this abstraction no longer offers convenience, but rather is a handicap, then it can well be discarded. This is particularly true for the decision maker who has access to a high speed computing machine, for this computing power alleviates many of the difficulties which have been in the past responsible for the introduction of continuous variables.

Ideas for Further Research

Sensitivity analysis of Bayes procedures to the various inputs seems to offer many possibilities for further research. In addition to the sensitivity to the prior distribution, there are also important considerations of a similar nature regarding the conditional distribution of the experimental outcome given the state of nature, and regarding the loss function used. To a large extent, the prior distribution, the conditional distribution, and the loss function enter the computations in a similar way. Thus, it would be expected that any progress in analyzing the sensitivity to one of these inputs might also be used in consideration of the others.

It should be clear that the approach to the sensitivity analysis used here is an approach and not the approach. The basic problem is simply studying the Bayes procedure as a function of the prior distribution, and there are undoubtedly many other approaches which would be profitable.

Within the framework used here, there are several avenues of further research. Further investigation of the role of the distance function is certainly one of these. Also, the same basic approach to the sensitivity analysis might be used to develop sensitivity analysis procedures for the loss functions or the conditional probabilities. Some new ideas would, no doubt, be obtained through applying the procedure of this paper to some special problem areas, such as statistical quality control, or analysis of variance. It might also be worthwhile to investigate more carefully the effect of further experimentation on the sensitivity to the prior distribution. Finally, it appears that a very fruitful area for further research is the combination of Bayesian and minimax procedures suggested in Chapter III.

BIBLIOGRAPHY

1. Blackwell, D., and M. Girshick. Theory of Games and Statistical Decisions. New York: John Wiley and Sons, 1954.

2. DeFinnetti, B. "Foresight: Its Logical Laws, Its Subjective Sources," Studies in Subjective Probability (ed. Kyburg, Smokler). New York: John Wiley and Sons, 1963.

3. Good, I. "Rational Decision," Royal Statistical Society, Series B, 14 (1952), pp. 107-114.

4. Hadley, G. Nonlinear and Dynamic Programming. Reading, Pennsylvania: Addison-Wesley, 1964.

5. Isaacs, H. "Sensitivity of Decisions to Probability Estimation Errors," Operations Research, 13 (1963), pp. 536-551.

6. Jeffries, H. Theory of Probability. Third ed. Oxford: Oxford University Press, 1961

7. Lehmann, E., and J. Hodges. "The Use of Previous Experience in Reaching Statistical Decisions," Annals of Mathematical Statistics, 23 (1952), pp. 396-407.

8. Luce, R., and H. Raiffa. Games and Decisions. New York: John Wiley and Sons, 1957.

9. Munroe, M. Introduction to Measure and Integration. Reading, Pennsylvania: Addison-Wesley, 1953.

10. Neyman, J. Lectures and Conferences on Mathematical Statistics and Probability. Washington, D. C.: U. S. Department of Agriculture, 1952.

11. Pearson, E. "Some Thoughts on Statistical Inference," Annals of Mathematical Statistics, 33 (1962), pp. 394-414.

12. Ramsey, F. The Foundations of Mathematics and Other Logical Essays (ed. Braithwaite). New York: Humanities Press, 1950.

13. Savage, L. The Foundations of Statistics. New York: John Wiley and Sons, 1954.

14. Schlaiffer, R. Probability and Statistics for Business Decisions. New York: McGraw-Hill, 1959.

15. Smith, C. "Consistency in Statistical Inference and Decision," Royal Statistical Society, Series B 23 (1961), pp. 1-25.

16. Wald, A. Statistical Decision Functions. New York: John Wiley and Sons, 1950.

17. Wesler, O. "Invariance Theory and a Modified Minimax Principle," Annals of Mathematical Statistics 30 (1959), pp. 1-20.

VITA

Donald Alan Pierce

Candidate for the Degree of

Doctor of Philosophy

Thesis: A METHOD FOR DETERMINING THE SENSITIVITY OF A BAYES PROCEDURE TO THE PRIOR DISTRIBUTION

Major Field: Mathematics and Statistics

Biographical:

Personal Data: Born in Enid, Oklahoma, October 19, 1939, the son of Mr. and Mrs. Gerald Pierce.

Education: Graduated from Enid High School, Enid, Oklahoma, in May 1957; attended University of Oklahoma in 1957 and 1958; received the Bachelor of Science degree from Oklahoma State University in 1961, with a major in Industrial Engineering; received the Master of Science degree from Oklahoma State University in 1963, as a National Science Foundation Fellow, with a major in Industrial Engineering; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May 1965, as a National Science Foundation Fellow.

Professional Experience: Project Engineer, U. S. Bureau of Mines, Summer 1960; Applied Science Representative, International Business Machines Corporation, Summer 1961; graduate teaching assistant, School of Industrial Engineering, Oklahoma State University, 1962-1963; Senior Operations Analyst, General Dynamics/Fort Worth, Summer 1964.