BAYESIAN STRATEGIES FOR DETECTING AND LOCATING TARGETS

by

John Yee-tseng Chu

A Thesis Submitted to the Faculty of the

DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE
WITH A MAJOR IN SYSTEMS ENGINEERING

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 7 3

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: John Y. Chu

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Lucien Duckstein
LUCIEN DUCKSTEIN
Professor of Systems and
Industrial Engineering

17 may 1973
Date

ACKNOWLEDGMENTS

The author wishes to express his gratitude to the entire faculty and fellow students of the Department of Systems and Industrial Engineering for their comments and criticisms during the course of this thesis. Particularly, the author would like to acknowledge his great and overwhelming debt to Drs. Lucien Duckstein, Chester C. Kisiel, and Donald R. Davis, whose technical assistance, advice, criticisms, foresight, and patience throughout the many revisions of this work have provided the guidance in keeping the author on the right track although he sometimes disregarded sound advice and pursued his own stubborn path.

The author gratefully acknowledges partial support from the National Science Foundation Grant GK 35791, "Practical Uses of Bayesian Decision Theory in Hydrology and Engineering."

Finally, the author would like to thank Mrs. Jessie Cude for her patience and professionalism in typing this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

ABSTRACT

This thesis defines a methodology for the detection of a transiting target intruding into an underwater hydrophone range. The detection problem is examined as a Bayesian sequential decision problem in which observations $x_1, x_2, \ldots$ can be taken one at a time from some distribution involving a parameter θ whose value is unknown. After each observation x_i, an evaluation of the information that has been obtained about θ from the observations $x_1, x_2, \ldots, x_i$ which have been taken up to that time is made. Then a decision whether to terminate the sampling process or to take another observation, x_{i+1}, is formulated based on the evaluated information.

A brief review of the history and development of sequential decision theory is given with emphasis on Wald's sequential probability ratio test (SPRT) and the Bayesian sequential decision theory. The Bayesian methodology is illustrated using a set of hypothesized data along with a couple of computation examples. The advantages of examining the target detection problem under the Bayesian decision theoretic framework are presented.

CHAPTER 1

INTRODUCTION

We shall be concerned in this paper with the investigation of how an optimum decision may be made based on a set of sequentially collected data. The decision that we want to make is to determine whether or not a transiting target has intruded into a defined area. To formulate our decision, we shall use some form of sequential analysis in which our conclusion depends on the data observed.

In discussing this investigation, we shall draw upon the theoretical works of others, particularly Wald (1945, 1950), Savage et al. (1962), Lindley (1965), Wetherill (1966), Aitchison (1970), and Winkler (1972).

Background of the Problem

The United States Navy operates an underwater tracking facility in the South Pacific. The tracking system is to provide the fleet operating forces with an antisubmarine warfare tactical evaluation capability. The underwater range is being used to evaluate the performance of various weapons, launched from submarine or surface vessels against surface or underwater targets. The range is capable of tracking objects which are equipped with an active sonic projector unit (pinger).

The in-water range area covers several square miles in a depth which varies from 400 to 1,000 fathoms (approximately 20,000 meters). The array of in-water sensors comprises 37 hydrophones, implanted on the ocean floor and geometrically configured to form 52 triangles, which lie adjacent to each other.

The tracking operation consists of launching a target equipped with a pinger unit near the edge of the hydrophone array at depths varying from the surface to 1000 feet from the ocean bottom. The pinger emits a coded pulse format, with pulse rate and pulse width preselected, which modulates a carrier frequency. Acoustic signals from the tracked object are received by the hydrophones, converted to electrical energy, and transmitted via coaxial cable to a shallow water junction point where conversion to balanced transmission is effected for final signal routing to the shore operations control center (OCC).

The OCC identifies individual targets on the basis of carrier frequency assignment, regenerates the envelope of the modulating pulse code format as received from each target and determines the apparent arrival time of each pulse relative to a master timing system. The central computing facility analyzes all of the processed data and establishes a position fix on each active target on the basis of pinger pulse transit time to each of three hydrophones or on the basis of the difference in arrival time of pinger pulse at each of four hydrophones. This computation also involves recognition and exclusion of false pulses, produced within the received signal structure, as a result of noise and acoustic reverberations.

A report filtering procedure is activated for the target so that it begins to look for coherent pulse trains on each of the hydrophones of the array. After allowing enough time (30 seconds, approximately) for the filter procedure to find coherence, tests are made to determine if any hydrophones have found coherent data and, if so, which ones. If none of the phones has coherent signals, the complementary "ACQUIRE" array is activated and the process repeats. This alternate action is to continue until coherent pinger signals are found. Neither array will find coherent pinging until the intruder enters the range volume. If one or more hydrophones within an acquire array finds coherent ping train, the procedure will select a tracking array based on which phones report, their relative loudness, and their relative time of arrival. The array selected will normally be close enough to allow a tracking fix to be made on the intruder so that normal tracking may begin.

On those occasions where the selected array does not allow enough hydrophones to acquire, the procedure will automatically select a new array which is centered around the phone which is nearest to the target (i.e., has the earliest ping time).

In all cases where decisions are based on what hydrophones acquire, time has been allowed for the report filtering procedures to acquire. Since those procedures depend mainly on autocorrelation of signal pulses, enough time has been allowed for this process to build up a high enough confidence factor before a decision is taken. This analysis takes approximately 28 to 42 seconds.

Motivation of the New Study

The primary reason that we choose a sequential method is that our information comes only now and then in small driblets and in a sequential manner. Also the cost of acquiring information is not so small that we can immediately ask for the maximum amount, since we are restricted by the amount of time we have to make a decision. We want to analyze our information as it comes in and make our decision when we have acquired the proper amount of information--no sooner, no later. All this is handled straightforwardly by Bayesian methods. We just keep using

$$\text{Posterior prob.} \propto \text{prior prob.} \times \text{likelihood},$$

where the prior probabilities represent our knowledge before the current observation.

Under the present operational procedures, a target cannot be tracked until it has been identified to be within the tracking range. This detection process is called acquisition of target. The sequence of procedures which will find and track such intruding targets proceeds as described below:

One of two complementary "ACQUIRE" hydrophone arrays will be assigned to each target. These arrays each consist of a group of widely spaced hydrophones so that any target within the range volume can be heard by at least one hydrophone in at least one of the arrays.

Since the present acquisition process described above is highly time consuming, the range's ability to track a rapidly transiting target will be greatly affected. The transiting target conceivably could have gone outside of the range volume before its presence inside the range is detected. In order to alleviate such difficulties a less time consuming detection decision process must be found and implemented when tracking fast-moving targets.

In this study, it is hoped that the sequential Bayesian decision theory will yield a solution to our problem, that is, a less time consuming decision on whether we have a target inside the range volume or not.

In the following chapters we will give a brief review of the history and development of sequential decision with emphasis on SPRT and BDT. We then will formulate and solve the detection problem under the framework of BDT. Application and computation of the scheme will also be presented.

CHAPTER 2

REVIEW OF LITERATURE

A special note is appropriate here about the contribution to the study of uncertainty by the Reverend Thomas Bayes and his work, "An Essay Towards Solving a Problem in the Doctrine of Chance" (Bayes 1763). His contribution to probability theory can be described as the discovery of one of the basic theorems of conditional probability. Bayes' theorem has been the center of controversy at various times, not over its mathematical correctness, but over its validity and appropriateness as a tool of statistical inference. This controversy has been quite disquieting in recent years. During the emerging years (1920-1945) of the main streams of modern statistical inference, a majority of statisticians adopted the prevailing frequentist attitude, with probability firmly associated to the notion of the limit of long run relative frequency. However, during this period of advances, there were statisticians adhering to the Bayesian principle. Jeffreys (1961), in his Theory of Probability and his earlier papers, and Keynes (1921), in his Treatise on Probability, propounded the viewpoint of probabilities as "degree of belief"; Ramsey (1964), in his "Truth and Probability," and de Finetti (1964), in his "Foresight: Its Logical Laws, Its Subjective Sources," presented the concepts of subjective probability in which each person must formulate his willingness to accept odds on imagined bets.

It has been only recently that the achievements of these pioneers have been met with full recognition. For the revival of interest during the last two decades in subjective probability and in the Bayesian argument, in general the main credit must go to L. J. Savage (1954). It is largely through his work that interest in the Bayesian form of analysis has developed to the extent that it is now recognized as an important practical tool.

Brief History of Sequential Analysis

In the field of statistical estimation the notion of sequential sampling was used by a number of statisticians during the decade 1933-1943 in somewhat unsophisticated forms. One of the earliest attempts was made by Thompson (1933, 1935). He considered a large population divided into two groups, the first group taking treatment I and the second group taking treatment II, such that the probabilities of occurrence of a certain critical event following the treatment are p_1 and p_2, respectively. In order to estimate the treatment effects (or to test the hypothesis of no difference between the treatments), Thompson suggested that small samples of size n_1 and n_2 be drawn in the first instance from the two groups. Let P be some estimate of the probability that $p_2 < p_1$ based on these $n_1 + n_2$ observations, and let g(P) be some suitable chosen monotonic increasing function of P. It was then argued that for subsequent sampling a fraction g(P) of individuals should be given treatment I while 1 - g(P) of individuals should be subjected to treatment II. The feeling behind such a sampling method is certainly sequential in that, instead of drawing m_1 and m_2

$(m_1 + m_2 = n)$ individuals arbitrarily from the two groups, one is drawing n_1 and n_2 arbitrarily at the first stage and then $(n - n_1 - n_2)g(P)$ and $(n - n_1 - n_2)(1 - g(P))$ at the second stage depending on the results of the first stage.

A similar type of method has been used by Neyman (1934) in the field of stratified sampling. Suppose that a population contains k groups or strata, the ith $(i = 1, 2, \ldots, k)$ stratum containing N_i individuals with a variance σ_i^2. It is known that, in order to estimate the mean of the entire population with minimum variance for the estimator, one should draw $n_i \propto N_i\sigma_i$ individuals from the ith $(i = 1, 2, \ldots, k)$ stratum. Since in many cases $\{\sigma_i\}$ are unknown, an element of sequential sampling is tacitly introduced by the suggestion that a pilot sample be drawn to estimate the $\{\sigma_i\}$. Subsequently, Sukhatme (1935) investigated the precision of such methods to attain the final goal of minimizing the variance of the estimator. Ten years later Stein (1945) developed an elegant method to achieve a similar goal. A problem in the same general area was discussed by Mahalanobis (1940) to determine the acreage under jute in Bengal. Also allied to these is Hotelling's (1941) discussion on the determination of the maximum of a regression function by a series of experiments.

The preceding survey has shown that up until 1943 statisticians were devising a wide variety of ad hoc sequential procedures for certain isolated problems on statistical testing and estimation. It remained largely for Wald (1943, 1945, 1947, 1950) to take the major step toward the formulation of a general theory of sequential tests when the probability model involves a real-value parameter. The actual technique proved to be highly successful in the sense that the average amount of data required to reach a decision turns out to be, in general, much less than that of the corresponding best fixed sample test. Immediately following Wald's work some applications of the theory were given by Freeman (1944). Credits should be given also to the works of the British school of thought led by Barnard (1946, 1947) and Stockman (1944) who independently developed similar sequential tests. However, Barnard's exposition was originally on rather different lines and concerned situations for which binomial (or Bernoulli processes) distributions usually provided the basic mathematical model. Most of these results are stated in Wald's (1947) book in full generality.

Background of Wald's Sequential Probability Ratio Test

The theory of games is a part of the rich mathematical legacy left by John von Neumann. Although others, notably Borel, preceded him in formulating a theory of games, it was von Neumann who with the publication in 1927 of a proof of the minimax theorem for finite games laid the foundation for the theory of games as it is known today. This work culminated in a book written in collaboration with Morgenstern entitled Theory of Games and Economic Behavior published in 1947 (von Neumann and Morgenstern).

At about the same time, statistical theory was being given an increasingly rigorous mathematical foundation in a series of papers by Neyman and Pearson (1928, 1936, 1938). Statistical theory until that time, as developed by Karl Pearson, R. A. Fisher, and others, had lacked the precise mathematical formulation supplied by Neyman and Pearson that allows the delicate foundation questions involved to be treated rigorously (Barnard 1958).

Apparently it was A. Wald who first appreciated the connection between the theory of games and statistical theory of Neyman and Pearson and who recognized the advantages of basing statistical theory on the theory of games. Wald's theory of statistical decisions, as it is called, generalizes and simplifies the Neyman-Pearson theory by unifying, that is, by treating problems considered as distinct in the Neyman-Pearson theory as special cases of the decision theory problem (Wald 1943).

Wald, following the Neyman-Pearson tradition of treating certain classes of statistical problems as action problems, went one step further than Neyman-Pearson. He incorporated such concepts as cost, loss, profit, value, and worth of consequences in the very formulation of a general statistical decision problem. He did not scale these economic factors in terms of utility values, but he did assume that it was appropriate to consider expected values of these economic factors. Also, Wald did not assign probabilities to the underlying states of nature, i.e., some prior probability distribution describing his degree of belief about the unknown true state.

Thus Wald is an objectivist (frequentist) or non-Bayesian because he refuses to make probability statement about the population parameter. He only admits probabilities that can be interpreted from a relative frequency point of view.

Wald's SPRT

We will now explore the central idea of Wald's (1947) sequential probability ratio test (SPRT), which is designed to deal with the testing of two simple hypotheses, and see how this is closely related to the Neyman-Pearson test. Suppose a random variable x has distribution $f(x,p)$ and we wish to test the null hypothesis that $p = p_1$ against the alternative hypothesis that $p = p_2$. The test constructed decides in favor of either p_1 or p_2 on the basis of observations (random variables) $x_1, x_2, \ldots$. We will suppose that if p_1 is true we wish to decide for p_1 (H_1) with probability $1 - \eta$, while if p_2 is true, we wish to decide for p_2 (H_2) with probability $1 - \lambda$.

For a fixed sample size (nonsequential) test, the optimum solution to this problem was provided by Neyman-Pearson (1928). They showed that for a given n, the test giving smallest λ (that is, the most powerful test) depends on the likelihood ratio L_n, where

$$L_n = \frac{\text{Probability of observed results given } p_2 \text{ true}}{\text{Probability of observed results given } p_1 \text{ true}}$$

$$= \prod_{i=1}^{n} \frac{f(x_i, p_2)}{f(x_i, p_1)} \tag{2.1}$$

and the test decides for or against p_1 according to whether L_n is less than or greater than a constant. The value of this constant can be chosen to give the test the correct size η, and in principle n can be chosen to give the test power $1 - \lambda$.

Wald's SPRT is analogous to this and has an analogous optimum property. It operates as follows.

Since SPRT is designed to deal with the testing of two simple hypotheses, we are working with a Bernoulli process. Suppose we know that the probability of an observed sequence of Bernoulli trials with s successes and r failures is

$$p^s (1 - p)^r$$

so

$$f \left(\frac{H_1}{H_2} \,\middle|\, \text{sequence}\right) = \left(\frac{p_1}{p_2}\right)^s \left(\frac{1 - p_1}{1 - p_2}\right)^r \tag{2.2}$$

where p_1 and p_2 are same as above and where H_1 denotes hypothesis 1, H_2 denotes hypothesis 2. We are continuing sampling as long as

$$B < f < A . \tag{2.3}$$

Choose H_2 when

$$f \leq B \quad . \tag{2.4}$$

Choose H_1 when

$$f \geq A \quad . \tag{2.5}$$

(The constants A and B can be chosen to obtain approximately the probabilities of error η and λ prescribed.)

Suppose that we can list all sequences of results that lead to the decision: choose H_1. Call this set X_I. Since we assume that the experiment will eventually end, then we have

$$P\ (X_I \mid H_1) = 1 - \eta\ ,\quad P\ (X_I \mid H_2) = \lambda\ , \tag{2.6}$$

because we err if a sample from H_2 leads us to X_I; we make the correct decision if a sample from H_1 leads us there. Similarly, suppose we can list all those sequences of results that lead to the decision: choose H_2. Call this set X_{II}. We have

$$P\ (X_{II} \mid H_1) = \eta\ ,\quad P\ (X_{II} \mid H_2) = 1 - \lambda\ , \tag{2.7}$$

again because we err if a sample from H_1 leads us to X_{II}; we act correctly if a sample from H_2 leads us there.

Now look at any one of the sequences of observations belonging to X_I. Call it x_j. It belongs to X_I because it led to the decision: choose H_1. This means that

$$A \leq f\ \left(\frac{H_1}{H_2} \mid x_j\right) = \frac{P(x_j \mid H_1)}{P(x_j \mid H_2)} \tag{2.8}$$

so

$$A \times P(x_j \mid H_2) \leq P(x_j \mid H_1)\ ,\quad j=1,2,\ldots\ . \tag{2.9}$$

Now sum over j. Since x_1, x_2, . . . exhaust the set X_I, the sum $\sum_j P\ (x_j \mid H_2)$ is just the probability of ever reaching X_I, when the sample really comes from H_2. This is the probability of an error of the second kind, i.e., $\lambda = \sum_j P\ (x_j \mid H_2)$. Next, $\sum_j P\ (x_j \mid H_1)$ is the probability of ever reaching X_I, when the sample is really from H_1, namely, $\sum_j P\ (x_j \mid H_1) = 1 - \eta$. Therefore substituting into (2.9) we have

$$A\beta \leq 1 - \eta. \tag{2.10}$$

By similar reasoning, we find that

$$\eta \leq \beta\ (1 - \lambda) \tag{2.11}$$

When the situation is investigated numerically, it turns out that most of the time these inequalities are very close to being equalities (Wald 1950), and we can set

$$A\lambda \simeq 1 - \eta\ , \quad \eta \simeq \beta\ (1 - \lambda)\ . \tag{2.12}$$

We will turn our attention at this point to investigate the Bayesian viewpoint.

Bayesian Sequential Analysis

In this section we investigate a method for applying Bayes' rules with respect to a given prior distribution, and we apply the method for finding the optimum sequential decision rules to our problem.

The Savage (1954) school extends Wald's results and concepts by formally introducing the subjective utilities for consequences and probability for states of the world. This is the approach which we will employ to solve our problem.

A sequential plan is said to be a Bayesian solution if it is derived by minimizing some overall risk with respect to a prior distribution. Here we examine a Bayesian solution for a test of two simple hypotheses, with the restriction that the prior distribution belonging to a class of families of distributions which possess natural conjugates, i.e., distribution closed under sampling to be defined below.

Before going on with the development of the method, we emphasize the elements involved in making a decision (according to Raiffa 1968). They are:

1. A predetermined number of possible acts. We need to choose one of these acts.
2. The situation at hand--a state of nature--is one of a list of known possibilities. But we do not know which one of the possibilities is our situation.
3. For every pair of action and state of nature, there is a consequence, measured in money or utility, which is known or estimated.
4. We have some estimates of the probabilities of the various states of nature.

We will adhere to this set of principles throughout the development of our model.

Recall that Bayes' theorem states that

$$\text{Posterior density} = \frac{(\text{Prior density})\ (\text{Likelihood})}{K}\ . \tag{2.13}$$

where K is the normalizing constant. Let the probability density of observations, $x = \{x_1, x_2, \ldots, x_n\}$, be $\emptyset(x \mid \theta)$ is an unknown parameter, $0 \leq 0 \leq 1$, and let the prior distribution for θ have the form $g(\theta \mid \alpha)$, where α is a parameter which indexes a unique distribution from the family g. Then the family of prior distributions $g(\theta|x)$ is said to be closed under sampling if for all observations $x = \{x_1, x_2, \ldots, x_n\}$ for which $\emptyset(x_i \mid \theta) \neq 0$, there is a β, $\beta = \beta(\alpha, x_1, x_2, \ldots, x_n)$ such that

$$g(\theta \mid \beta) = g(\theta \mid \alpha_1 x_1, \ldots, x_n)$$

$$= \frac{g(\theta \mid \alpha)\, \phi(x \mid \theta)}{\int g(\theta \mid \alpha)\, \phi(x \mid \theta)\, d\theta} \tag{2.14}$$

$$= \frac{g(\theta \mid \alpha)\, \phi(x_1 \mid \theta) \ldots \phi(x_n \mid \theta)}{\int g(\theta \mid \alpha)\, \phi(x_1 \mid \theta) \ldots \phi(x_n \mid \theta)\, d\theta}$$

This is simply an application of Bayes' theorem for continuous random variables. The density $g(\theta \mid \beta)$ and the density $g(\theta \mid \alpha)$ represent the posterior distribution and the prior distribution, respectively, and the product $\phi(x_1 \mid \theta)\, \phi(x_2 \mid \theta) \ldots \phi(x_n \mid \theta)$ represents the likelihood function (since we assume the x_i to be independent).

Now suppose observations $x = \{x_1, x_2, \ldots, x_n\}$ are taken from an experiment and the observations are independently and identically distributed with a common probability density function $\phi(x_i \mid \theta)$ and the prior distribution of θ is closed under sampling in the sense described above. (We make this restriction to insure that the problem can be solved in closed form.) As a result of a sequence of observations, one of two terminal decisions is to be made, which we label decision 1 and decision 2. The cost of taking decision i when the uncertainty, state of nature, is given by $\theta = \theta_i$ is $W_i(\theta)$. The cost of taking an observation is supposed constant.

After n observations have been taken, the risk defined as the expected loss of making decision i, i = 1, 2, conditioned upon observations $x = \{x_1, x_2, \ldots, x_n\}$ is

$$R_{d_i}(\alpha,x) = \frac{\int [g(\theta\,|\alpha)\ \phi(x|\ \theta)]\ W_i(\theta)\ d\theta}{\int g(\theta\ |\alpha)\ \phi(x\ |\ \theta)\ d\theta}$$

$$= \frac{\int [g(\theta\ |\alpha)\ \ \phi(x_1\ |\ \theta) \ldots \phi(x_n\ |\ \theta)]\ W_i(\theta)\ d\theta}{\int g(\theta\,|\ \alpha)\ \phi(x_1|\ \ \theta) \ldots \phi(x_n|\ \ \theta)\ d\theta}$$

$$= \int g(\theta\ |\ \beta)\ W_i(\theta)\ d\theta \tag{2.15}$$

$$= R_{d_i}(\alpha,\ x_1,\ x_2,\ \ldots\ ,\ x_n)$$

$$= R_{d_i}(\beta)$$

where

$$i = 1,\ 2\ ,$$

and

$$\beta = \beta(\alpha,\ x_1,\ x_2,\ \ldots\ ,\ x_n)\ ,$$

and decision 1 is preferred to decision 2 if $R_{d_1}(\beta) < R_{d_2}(\beta)$. Thus we can see that the decision to be made is dependent only on the posterior distribution of θ.

In general any sequential sampling is essentially represented by a division of the g space into three regions, two terminal decision regions and a continuation region as shown in Figure 1. We can define the space g as follows. Suppose we have distributions closed under sampling, and the prior distribution is $g(\theta\,|\,\alpha)$, where $\alpha = (v_1,\ v_2,\ \ldots,\ v_m)$. The effect of making observations can be described by a random walk in this g space, in which the dimensions are components $v_1,\ v_2,\ \ldots\ ,\ v_m$ of α. Specifically, if we have parameter distribution $g(\theta\,|\,\alpha) \equiv \text{Beta}\ (\theta\ ;\ s,\ t)$, we can define a two-dimensional space for g, in which each member of the family of distributions of the

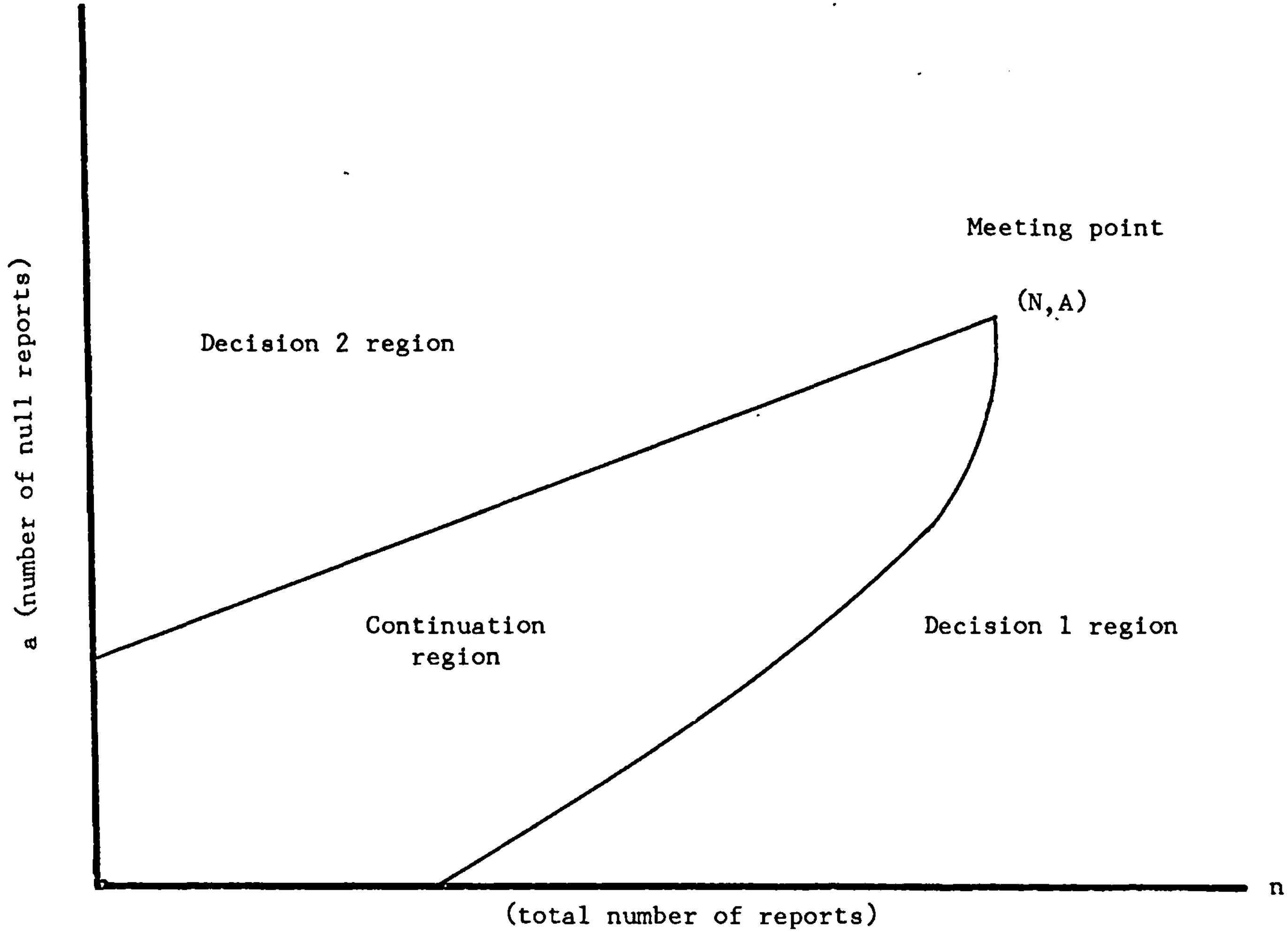

Figure 1. Decision regions and the meeting point of g-space.

type Beta (θ ; s, t) is represented by a point in this g space. Our total knowledge in the experiment is a point in this space. This results from the fact that the continuation risk at any point of the experiment depends only on the number of observations, n, taken and the decision boundaries for sequences of observations of length greater than n, and it does not depend on the particular path taken to reach that point.

At any point β in the g space, we can define the continuation risk $R_c(\beta)$ and the terminal risks, $R_{d_1}(\beta)$ and $R_{d_2}(\beta)$. Terminal risk is the expected loss associated with a terminal decision which results in taking an action after a sequence of n observations. Whereas, if the calculated risk dictates one additional observation to be taken before we make the decision, we call this a continuation decision and observations are continued. The point β is a continuation point or a decision 1 or 2 point, respectively, according to which of these three risks is the minimum. The Bayesian solution is completely described by the boundaries between these various regions. This brings us to the basic difficulty of this particular approach--that in order to define the continuation risk at β, it is necessary to know the decision boundaries for all points leading from β. There are several ways to circumvent this difficulty; they are:

1. The decision boundaries may meet, in which case the decision boundaries can be worked out backward from this point (backward induction approach) (Wetherill 1966, Aitchison 1970).

2. An approximate solution can be obtained by forcing the boundaries to meet at a very large sample size.

3. Even when the decision boundaries may not meet, it is still possible in certain cases to write the equations they satisfy and solve the equations.

We will use method 1 to approach our problem solution, since it produces the simplest (easiest to handle) closed form.

We can show that if one decision boundary meets the neutral boundary, i.e., the locus of points where the expected loss of taking either decision is equal, then the other decision boundary does so at the same point. Suppose all three boundaries, the two decision boundaries and the continuation boundary, meet at a point (N, A) in this g space, then at this point

Risk of making either terminal decision

= Risk of making one more observation and then making the relevant terminal decision.

The above equation gives us a mean of determining the meeting point boundary (N, A).

The points (N-1, A) all lead to points in (N, A), the risk of which can now be established. Thus the risk of taking either terminal decision, and the continuation risk can be calculated at all points (N-1, A) and the boundaries determined. In this way we can work backward and classify all points in this space (n, a) as terminal decisions or continuation points (see Figure 1).

CHAPTER 3

FORMAL SOLUTION: A BAYESIAN SEQUENTIAL APPROACH

In this chapter we will present a formal solution to the detection problem discussed in Chapter 1 using the method of Bayesian sequential analysis presented in Chapter 2.

We are concerned with an experiment in which we want to decide, after an arbitrarily long sequence of independent observations, whether or not a target is inside the bounds of the underwater range. An observation in this context is defined as an instantaneous pattern (a snapshot) from the array of detectors in the range. The data on the observations are derived from the responses of the individual detectors in the array. An observation is labeled positive if the target is within detecting range of at least two of the seven detectors composed of the array and is labeled a null otherwise. We wish to classify the experiment, after a sequence of observations become available, into one of two categories:

1. There is no target in the range, decision 1 (denoted d_1).
2. There is a target in the range, decision 2 (denoted d_2).

We assume that the probabilities of the reports are mutually independent. Each report has a constant cost. Furthermore, by reviewing records of previous experiments, we can derive an estimated prior distribution of θ, the probability of a null observation. We also impose the restriction that the prior distribution belongs to a family of distributions closed under sampling, i.e., the prior distribution has a natural conjugate.

Theoretical Model

We will organize our problem into the formalism of the Bayesian decision theory (BDT) framework (Winkler 1972). The problem involves two terminal decisions and the states of nature are indicated by a parameter θ, $0 \leq \theta \leq 1$. We will make the following assumptions about our problem:

1. It is possible to observe independent random variables x_1, x_2, ... The probability that $x_i = 1$ (null) is θ, and $x_i = 0$ (positive) is $1 - \theta$.
2. The cost of observing any x_i is a constant.
3. The losses are linear in θ, or powers of θ.[1]
4. There exists a prior distribution of θ which belongs to a family of distributions possessing natural conjugates.
5. A decision may be made after any number (including zero) of random variables have been observed either to make one of the two terminal decisions or to continue observing another x_i.

Under the conditions of our experiment, stated in Chapter 1, the best way in terms of minimum number of reports needed to make the decision involved with the least risk is to proceed sequentially, that

---

1. The restriction that the loss functions be linear or in powers of θ (in this case) is imposed to insure that we can handle the problem in closed form (Raiffa and Schlaifer 1961, pp. 214-220).

is, examine the reports one at a time, and after each observation deciding whether to continue observing or to make a terminal decision. After $x_1, x_2, \ldots, x_n$ have been observed, and letting $r = \Sigma x_i$, the likelihood of θ is proportional to $\theta^r(1 - \theta)^{n-r}$. Consequently, our process is Bernoulli.

Since Bayes' formula states that:

$$\text{posterior prob.} \propto \text{prior prob.} \times \text{likelihood},$$

we can assess a prior distribution and then find the posterior distribution following an observed datum. This could be a difficult task unless the prior distribution is a member of the class of families of distributions that is conjugate with respect to the Bernoulli process. The conjugate family in this case is the family of Beta distributions (see Appendix A concerning the properties and characteristics of the Beta distribution). The density function of the Beta distribution is very similar to the probability mass function of the binomial distribution. If the probability distribution of θ is a Beta distribution with parameter r and n as described above, where $0 < r < n$, then

$$f(\theta) = \frac{(n - 1)!}{(r - 1)!\,(n - r - 1)!}\,\theta^{r-1}\,(1 - \theta)^{n-r-1} \tag{3.1}$$

where $0 \leq \theta \leq 1$. We will now proceed to obtain a solution for our problem.

Solution

The problem has a single parameter, θ. The utilities are assumed to be as follows:

$$u_i(\theta) = \gamma_i + \tau_i\theta , \quad i = 1,2, \tag{3.2}$$

where $u_i(\theta)$ is the utility of the terminal decision d_i when the parameter has the value θ, and γ, τ are constants. With our problem it appears to be more convenient to define utility as loss. We can pose our problem in terms of losses without any loss of generality. We will denote these losses as $W_1(\theta)$ and $W_2(\theta)$, where

$$W_1(\theta) = k ; \tag{3.3}$$

$$W_2(\theta) = \theta . \tag{3.4}$$

We have as a parameter distribution $g(\theta \mid \alpha)$, the Beta distribution, where α is a vector, $\alpha = (s,t)$. This distribution, as we have shown, possesses a natural conjugate. We can define some space g in which each point belonging to g represents one member of the family of the distributions of the type $g(\theta \mid \alpha) = \text{Beta } (\theta ; s,t)$. Our accumulated knowledge at any stage in the experiment is a point in this space.

As stated earlier, in general, any sequential sampling plan is represented by a division of this g space into three regions, two terminal decision regions and a continuation region (which may be empty in some cases). This results from the fact that continuation risk at any point in the g space depends only on the length of sequence of the observations and not on the order of positive or null reports as they appeared in the sequence, i.e., the path that leads to the current point in the space. So at any point β in the g space, we can define the continuation risk $R_c(\beta)$ and the terminal decision risks $R_{d_1}(\beta)$ and $R_{d_2}(\beta)$. The point β is a continuation point or a decision 1 or 2 point, respectively, according to which of these three risks is the minimum. The Bayesian solution is completely described by the boundaries between these various regions.

We will describe the procedure to compute the optimum decision. Then we will show why we need to consider the various boundaries defined above,and demonstrate how they may be obtained.

Computation Scheme

The computation of the optimum decision begins by calculating the various expected losses corresponding to different actions and selecting that action which leads to the least expected loss. Suppose that our current total knowledge (our notion of likelihood) of θ has a probability distribution that is the Beta distribution parameterized by s and t. Then the expected loss if d_1 is taken is

$$E[W_1(\theta)] = E[k] = k , \tag{3.5}$$

since k is constant, and if d_2 is chosen, the expected loss is

$$E[W_2(\theta)] = E[\theta] = s/(s + t) , \tag{3.6}$$

since θ has a Beta distribution. Consequently, under the conditions of our experiment, d_1 is not chosen if $k > s/(s + t)$ and is chosen otherwise. With the Beta family tabulation for one particular member of the family effectively provides tabulation for all other members. Essentially, if the optimum scheme is tabulated for the special Beta distribution:

$$\frac{(s + t - 1)!}{(s - 1)!\ (t - 1)!}\ \theta^{s-1}\ (1 - \theta)^{t-1}\ ;$$

with $s = t = 1$ (i.e., for a uniform prior distribution of θ) then the optimum scheme is known for any Beta prior distribution with positive integer values of s and t. This follows since the tabulated scheme for $s = t = 1$ will tell one what to do if n samples have been taken and r found to be null reports. The likelihood of this is proportional to $\theta^r(1 - \theta)^{n-r}$ so the distribution of θ after the reports have been taken is proportional to the product of the original prior and this likelihood, namely, $\theta^r(1 - \theta)^{n-r}$. Consequently, the situation is exactly as if one was starting with $a = r$, $b = n - r$, and no further tabulation for this prior distribution is necessary. If the prior distribution has $s = t = 1$, then this current distribution is obtained by observing $r = s - 1$ null reports and $n - r = t - 1$ positive reports. This must be so, since currently we have $s = r$ null reports and $n - r = t - 1$ positive in the sequence of reports already, if a prior distribution having $s = t = 1$ brings us to this current position. In this case the decision maker would not choose a decision if the loss is such that $k < (r + 1)/(n + 2)$ and choose that decision if the loss is $k > (r + 1)/(n + 2)$. Therefore, the best we can do about the loss if a terminal decision is to be made at this point is [denoted as $R_d(s,t)$]

$$R_d(s,t) = \min[k, s/(s + t)]. \tag{3.7}$$

But stopping is not the only choice; there are other alternatives here. Suppose the decision maker desires not to make a terminal decision but wants to take one additional observation and then to select between the two terminal decisions. We will examine below how such a scheme may proceed.

Let $R(s,t)$ be expected loss [as shown in equations (3.5) and (3.6)] of the best (smallest expected loss) procedure when the prior distribution is indexed by the value (s,t). If $R_d(s,t)$ is equal to the smaller of the terminal losses k or $s/(s + t)$ then our best procedure is to stop observing and take that terminal decision that takes on the smaller loss. If, however, $R(s,t)$ is less than either of the two terminal decision losses, then the best procedure is not to make a terminal decision as yet but to make at least one further observation and then proceed in a scheme that we will present below.

We need first to compute the expected loss if one additional observation is taken followed by the optimum terminal decision. Suppose that the corresponding expected loss is $R_c(s,t)$. If the additional observation is a null report, then s will be increased by 1 as will $s+t$. The probability of a null report, given θ, is simply θ, and since θ is Beta-distributed, the expectation of a null report is $s/(s+t)$ as before. If the additional observation is a null report, the expected loss obtained by taking the optimum procedure after the observation will be $R(s+1,t)$; for a positive report it will be $R(s,t+1)$. The observation incurs an extra loss k_o, the cost of an observation. Therefore, if one additional observation is made when the state of knowledge is (s,t), we have

$$R_c(s,t) = k_o + \frac{s}{s+t} R(s+1,t) + \frac{t}{s+t} R(s,t+1). \tag{3.8}$$

Knowing $R_d(s,t)$, the stopping loss, and $R_c(s,t)$, the loss after continued onto one more observation, the equation for R(s,t), the expected loss for the best procedure is

$$R(s,t) = \min [R_d(s,t),\ R_c(s,t)] \tag{3.9}$$

since we have only the alternative of stop observing and make a decision or continue to make one more observation before deciding.

Equation (3.8) is the fundamental recursion scheme. If R(s,t) is known for all (s,t) with $s+t = n_o$, where n_o is the number of observations taken so far, then (3.8) enables R(s,t) to be found for all (s,t) with $s+t = n_o - 1$. Consequently once R(s,t) is found for all (s,t) with $s+t = n_o$, it is known for all (s,t) with $(n_o - s - t) > 0$, by the recursion equation (3.8). Once R(s,t) is known, the optimum procedure at (s,t) is to choose d_2 if $R(s,t) = s/(s+t)$, to choose d_1 if $R(s,t) = k$, and take yet an additional observation otherwise. Hence in the graph of the (s+t,s)-plane, each point is a "d_1," "d_2," or "continuation" point. The aggregate of points of the same type will form a region. The curves in the plane that separate any two regions will be the boundaries. The curve between the d_i, i = 1, 2, and the continuation regions will be the d_i boundary. The curve on which all three alternatives yield the same expected loss is the neutral boundary on which is the meeting point for all three alternatives, d_1, d_2, and continuation.

Boundaries and the Meeting Point

At this point we will examine the various boundaries which separate the decision regions, the meeting point, and the interrelationship between the meeting point and the decision regions.

The divisions of the g space into terminal regions and a continuation region do not take observations into account. We will have extremes where the expected quality of the experiment is so good or so poor that the decision maker can make his choice without any observation. In between there will be a region of doubt where it pays to take further observations. Hence we have three regions, a continuation region and two terminal decision regions, each of which is defined uniquely by its position in the g space.

The boundary in the g space between d_1 (or d_2) region and the other terminal decision region or continuation region will be called the decision 1 (or 2) boundary. Let the losses of taking decisions 1 and 2 be denoted W_1 (θ) and W_2 (θ), respectively [see equation (3.3)].

We consider the boundary where the three regions meet; this is often a point and, if so, will be called the meeting point. Our interest centers on this meeting point, because if it can be determined, the optimum terminal decision boundaries can be worked backwards from this meeting point by a backward induction approach.

Note that the neutral boundary is the locus of points such that the decision risk and the expected loss of taking either terminal decision are equal, and this is given by α satisfying

$$\int g(\theta \mid \alpha) [W_2(\theta) - W_1(\theta)] d\theta = 0 . \tag{3.10}$$

If the decision 1 and 2 boundaries meet, they must meet on this neutral boundary.

Suppose that the decision 1 and 2 boundaries meet the neutral boundary at exactly the same point; then at this meeting point the risk of making a terminal decision is equal to the cost of one additional observation plus the terminal decision risk associated with this extra observation (since one more observation from this meeting point can only lead to a terminal decision). If at the meeting point we take decision 2 (the resulting equation is the same whichever terminal decision is used here), then for a given θ the risk of a terminal decision at the meeting point is $W_2(\theta)$. So after one additional observation is taken, the terminal decision risks are

$$p(\alpha,\theta)\, W_1(\theta) + (1 - p(\alpha,\theta))W_2(\theta),$$

where $p(\alpha,\theta)$ is the probability that for an α on the meeting point boundary, one more observation leads to points in the terminal decision 1 region. Thus, for the meeting point, α needs to satisfy

$$\int [p(\alpha,\theta)W_1(\theta) + (1 - p(\alpha,\theta))W_2(\theta) + 1]\, g(\theta \mid \alpha)\, d\theta \tag{3.11a}$$

$$= \int g(\theta \mid \alpha) W_2(\theta)\, d\theta$$

which reduced to

$$\int p(\alpha,\theta)[W_2(\theta) - W_1(\theta)]\, g(\theta \mid \alpha)\, d\theta = 1 . \tag{3.11b}$$

Two points have to be clarified here before we proceed further. First is that the two terminal decision boundaries do in fact meet the neutral boundary at the same point, and second is that the locus of α- satisfying equations (3.10) and (3.11b) provides a complete definition of the meeting point boundary. Both of these propositions have been established in the following way by Wetherill (1966).

First we derive equations for the decision 1 and 2 boundaries. Simultaneous solution of these equations defines the boundary where the two terminal decision boundaries meet, and it will be shown that the resulting equations can be reduced to equations (3.10) and (3.11b).

Let points on the decision 1 and 2 boundaries be denoted by α' and α'', respectively. For observation starting from position α in the continuation region of this g space, with given boundaries and the value of θ specified, let $S(\alpha,\theta)$ and $A(\alpha,\theta)$ be the expected additional sample size and the probability that observations eventually lead to decision 1, respectively. For points α' and α'' on the terminal decision boundaries, $S(\alpha',\theta)$, $A(\alpha',\theta)$, etc. are defined to be values obtained if observation is continued from the points α', α'', but otherwise with the same terminal decision boundaries.

Consider now any point α' on the decision 1 boundary. Then the risk of continuing observing is equal to the risk of taking terminal decision 1 immediately. To decide right away has risk

$$\int g(\theta \mid \alpha') \, W_1(\theta) \, d\theta \, . \tag{3.12}$$

To continue observing has the risk

$$\int[S(\alpha',\theta) + (1 - A(\alpha',\theta))W_2(\theta) + A(\alpha',\theta)W_1(\theta)]\, g(\theta\,|\,\alpha')\, d\theta . \quad (3.13)$$

The equation for the value(s) of α' on the decision 1 boundary is obtained by equating these expressions (3.12) and (3.13) which simplifies to

$$\int [S(\alpha',\theta) + (1 - A(\alpha',\theta))\{W_2(\theta) - W_1(\theta)\}]\, g(\theta\,|\,\alpha')\, d\theta = 0. \quad (3.14)$$

Similarly for point α'' on decision 2 boundary we have

$$\int [S(\alpha'',\theta) + A(\alpha'',\theta)\ (W_1(\theta) - W_2(\theta))]\ g(\theta\,|\,\alpha'')\, d\theta = 0. \quad (3.15)$$

It was stated previously that the meeting point boundary will be defined by $\alpha = \alpha' = \alpha''$ simultaneously satisfying equations (3.14) and (3.15). If such a solution exists, then, provided the next step from points on this boundary does not land back into the continuation region, we have $S(\alpha,\theta) = 1$ and $A(\alpha,\theta) = p(\alpha,\theta)$. On substituting these values into equations (3.14) and (3.15), equation (3.15) becomes equation (3.11b) and equations (3.14) and (3.15) taken together can be reduced to equation (3.10). Therefore, if a solution to (3.10) and (3.11b) exists, then this solution defines a locus of points where both the terminal decision boundaries meet the neutral boundary.

If we consider a point β in the g space, then the terminal decision risk at this point is

$$R_d(\beta) = \min[\int g(\theta\,|\,\beta)W_1(\theta)\, d\theta,\ \ \int g(\theta\,|\,\beta)(\theta)\, d\theta] \quad (3.16)$$

and the continuation risk is

$R_c(\beta)$

$$= \int [S(\beta,\theta) + (1 - A(\beta,\theta))W_2(\theta) + A(\beta,\theta)W_1(\theta)]\ g(\theta \mid \beta)\ d\theta \qquad (3.17)$$

We will define $R(\beta)$ to be the risk at the point β, where

$$R(\beta) \;=\; \min\ [R_c(\beta),\ R_d(\beta)]\ . \qquad (3.18)$$

If $R(\beta) = R_c(\beta)$, then β is a continuation point, and if $R(\beta) = R_d(\beta)$, then β is a terminal decision point.

CHAPTER 4

COMPUTATIONAL PROCEDURES

We will present two examples in which we try to decide between two hypotheses. The data in both are sequences of Bernoulli trials. In the first example, the loss functions are linear; in the second example, one loss function is linear and one loss function is quadratic. We will apply the sequential analysis scheme, which we have developed in the preceding chapter, to our detection problem. Below we will demonstrate the computational procedures [using equations (3.14), (3.15), (3.16), and (3.17)] through the illustration of a couple of calculation examples. (Throughout the examples the cost of an observation $k_o = 1$ is assumed.)

Example 1 (linear case).

Suppose that the loss of deciding that there is no target in the range (d_1) is

$$W_1(\theta) = k(\theta - \theta_o) \quad \text{if } \theta \geq \theta_o \tag{4.1}$$
$$= 0 \quad \text{if } \theta \leq \theta_o$$

and the loss of deciding that there is a target in the range (d_2) is

$$W_2(\theta) = k(\theta_o - \theta) \quad \text{if } \theta \leq \theta_o \tag{4.2}$$
$$= 0 \quad \text{if } \theta \geq \theta_o$$

We will denote the number of null reports found, up to and including the current report, for the time during which the experiment is being monitored to be a and denote the number of positive reports by b, and our prior estimation of the uncertainty to be Beta(θ; s,t). Then applying equations (3.10) and (3.11b) we can find the meeting point by the following procedure:

$$\int_0^1 g(\theta \mid \alpha)\ [W_2(\theta) - W_1(\theta)]\ d\theta$$

$$= \int_{\theta_o}^1 \frac{(a+s+b+t-1)!}{(a+s-1)!\ (b+t-1)!}\ \theta^{a+s-1}\ (1-\theta)^{b+t-1}[0-k(\theta-\theta_o)]\ d\theta$$

$$+ \int_0^{\theta_o} \frac{(a+s+b+t-1)!}{(a+s-1)!\ (b+t-1)!}\ \theta^{a+s-1}\ (1-\theta)^{b+t-1}[k(\theta_o-\theta) - 0]\ d\theta$$

$$= \int_0^1 \frac{(a+s+b+t-1)!}{(a+s-1)!\ (b+t-1)!}\ \theta^{a+s-1}\ (1-\theta)^{b+t-1}[k(\theta-\theta_o)]\ d\theta$$

$$= \frac{(a+s)\ (1-\theta_o)}{\theta_o} = b+t$$

and

$$\int [S(\alpha,\theta) + p(\alpha,\theta)\ \{W_1(\theta) - W_2(\theta)\}]\ g(\theta \mid \alpha)\ d\theta = 0$$

$$\int_0^1 [1 + (1-\theta)\ (k\theta - k\theta_o)]\ \frac{(a+s+b+t-1)!}{(a+s-1)!\ (b+t-1)!}\ \theta^{a+s-1}(1-\theta)^{b+t-1}d\theta = 0$$

$$\frac{(a+s+b+t)\ (a+s+b+t+1)}{k} = (b+t)\ [\theta_o(b+t+1) - (1-\theta_o)(a+s)]$$

Solving the two equations (4.3) and (4.4) simultaneously we have the meeting point,

$$b = \frac{(a + s)\ (1 - \theta_o)}{\theta_o} - t \tag{4.5}$$

$$a = k(1 - \theta_o)\ \theta_o^2 - \theta_o - s \tag{4.6}$$

In the equations that have been derived, a and b can be any positive numbers. They can only change by integer values, but if we started the iterations with non-integers,[1] they would remain so. Since we deal only in complete reports we are only interested in a lattice of points in the plane and not all points of the plane.

Suppose the meeting point is at (A, B). Then it may well be that at (A, B-1) the expected loss dictates the choice of decision 2, and at (A-1, B) it is more advantageous to take decision 1, while (A-1, B-1) is a continuation point. Then we see that it is impossible to reach (A, B) since if (A-1, B-1) is reached, the next report sampled will always lead to a decision. In order to find the last reachable point, we examine the points immediately off the neutral line, toward the origin from the meeting point of the boundaries, to see when they start being continuation points. If we are examining to see if (A, B) is the last reachable point, we compare the continuation and terminal decision risks for (A, B-1) and (A-1, B). If one or both of these points is a continuation point then (A, B) can be reached in the sampling.

---

1. This non-integer situation arises from the prior estimation of the uncertainty.

For any point, the terminal decision and continuation risks can be compared by comparing the left and right hand sides of the following equations:

$$\int g(\theta \mid \alpha)\, W_1(\theta)\, d\theta$$
$$= \int [S(\alpha,\theta) + (1 - A(\alpha,\theta))W_2(\theta) + A(\alpha,\theta)\, W_1(\theta)]\, g(\theta \mid \alpha)\, d\theta \qquad (4.7)$$

$$\int g(\theta \mid \alpha)\, W_2(\theta)\, d\theta$$
$$= \int [S(\alpha,\theta) + (1 - A(\alpha,\theta))W_2(\theta) + A(\alpha,\theta)\, W_1(\theta)]\, g(\theta \mid \alpha)\, d\theta \qquad (4.8)$$

Note that the left hand side of equations (4.7) and (4.8) are just risks of taking decision 1 or 2 at the chosen point and the right hand side of both equations is the risk of continuation of awaiting an extra report.

Risk for choosing d_1 is

$$\int g(\theta \mid \alpha) W_1(\theta)\, d\theta$$

$$= \int_{\theta_o}^{1} \frac{(a+b-1)!}{(a-1)!\,(b-1)!}\, \theta^{a-1}(1-\theta)^{b-1}[k(\theta-\theta_o)]\, d\theta$$

Risk for choosing d_2 is

$$\int g(\theta \mid \alpha) W_2(\theta)\, d\theta$$

$$\int_{0}^{\theta_o} \frac{(a+b-1)!}{(a-1)!\,(b-1)!}\, \theta^{a-1}(1-\theta)^{b-1}[k(\theta_o-\theta)]\, d\theta$$

Risk for continuing (awaiting one extra report) is

$$\int_0^1 g(\theta)\,|\,\alpha)\ [S(\alpha,\theta) + (1-A(\alpha,\theta))W_2(\theta) + A(\alpha,\theta)W_1(\theta)W_1(\theta)]d\theta$$

where $\theta_o = \frac{1}{2}$, $k = 200$ (see Figures 2 and 3).

Example 2.

In this example we have the identical conditions as example 1 except one of the loss functions will be quadratic.

The loss function for deciding that there is no target in the range (d_1) is

$$\begin{aligned} W_1(\theta) &= k(\theta_o - \theta)^2 && \text{if } \theta \geq \theta_o \\ &= 0 && \text{otherwise} \end{aligned}$$

and the loss for deciding that there is a target in the range (d_2) is

$$\begin{aligned} W_2(\theta) &= k(\theta - \theta_o) && \text{if } \theta \leq \theta_o \\ &= 0 && \text{otherwise.} \end{aligned}$$

The result of the computation is presented below with the graph showing the three decision regions (Figure 4).

Risk for choosing d_1 is

$$\int g(\theta\,|\,\alpha)W_1(\theta)d\theta$$

$$= \int_{\theta_o}^{1} \frac{(a+b-1)!}{(a-1)!\ (b-1)!}\ \theta^{a-1}(1-\theta)^{b-1}[k(\theta_o-\theta)^2]d\theta$$

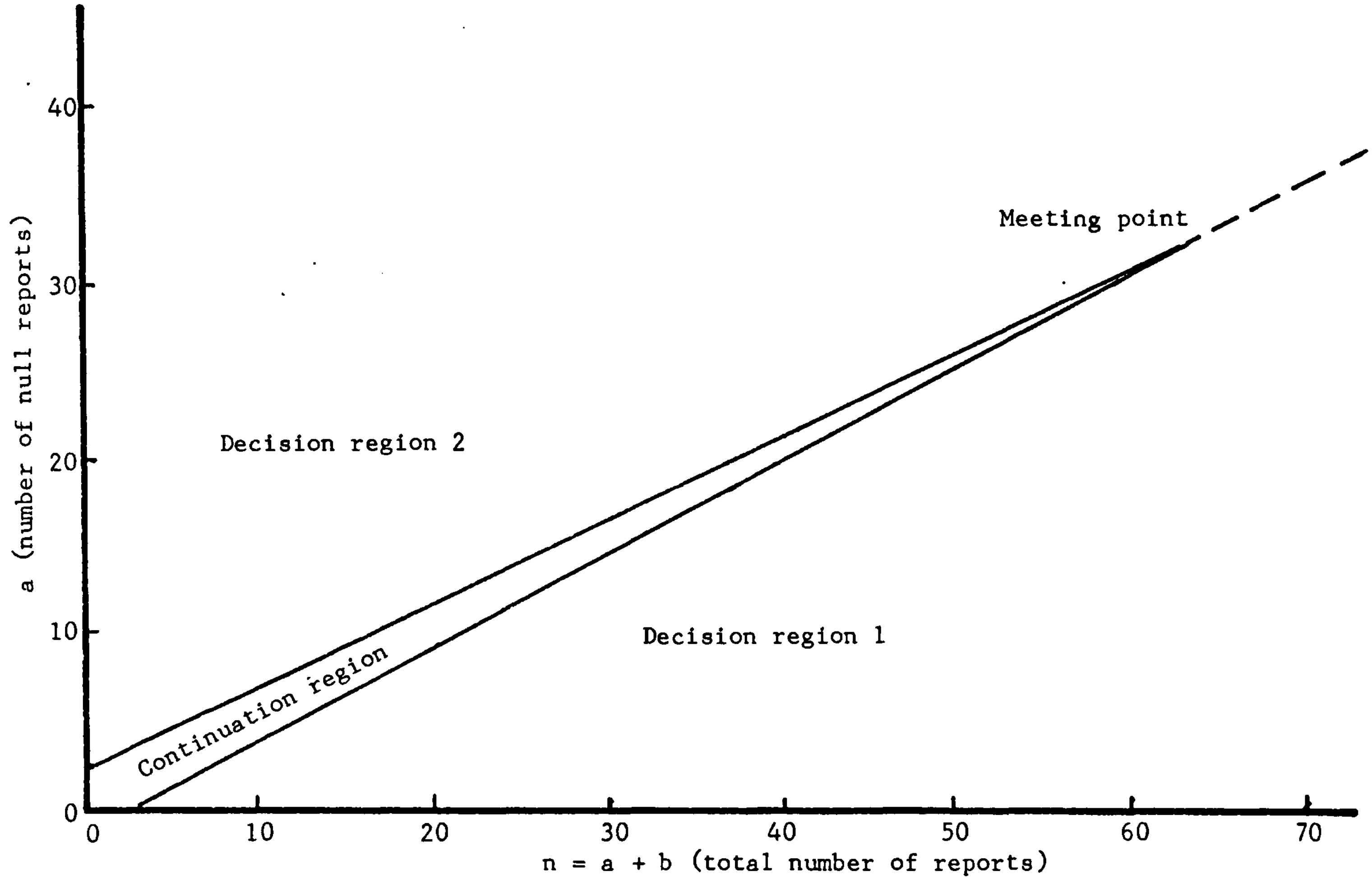

Figure 2. Decision regions of g-space.

Example 1. $\theta_o = \frac{1}{2}$, k = 200

```
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 N 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 1 C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
C C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
C C 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

a (number of null reports)

f   b (number of positive reports)

Figure 3. Decision regions - positive vs. nulls.

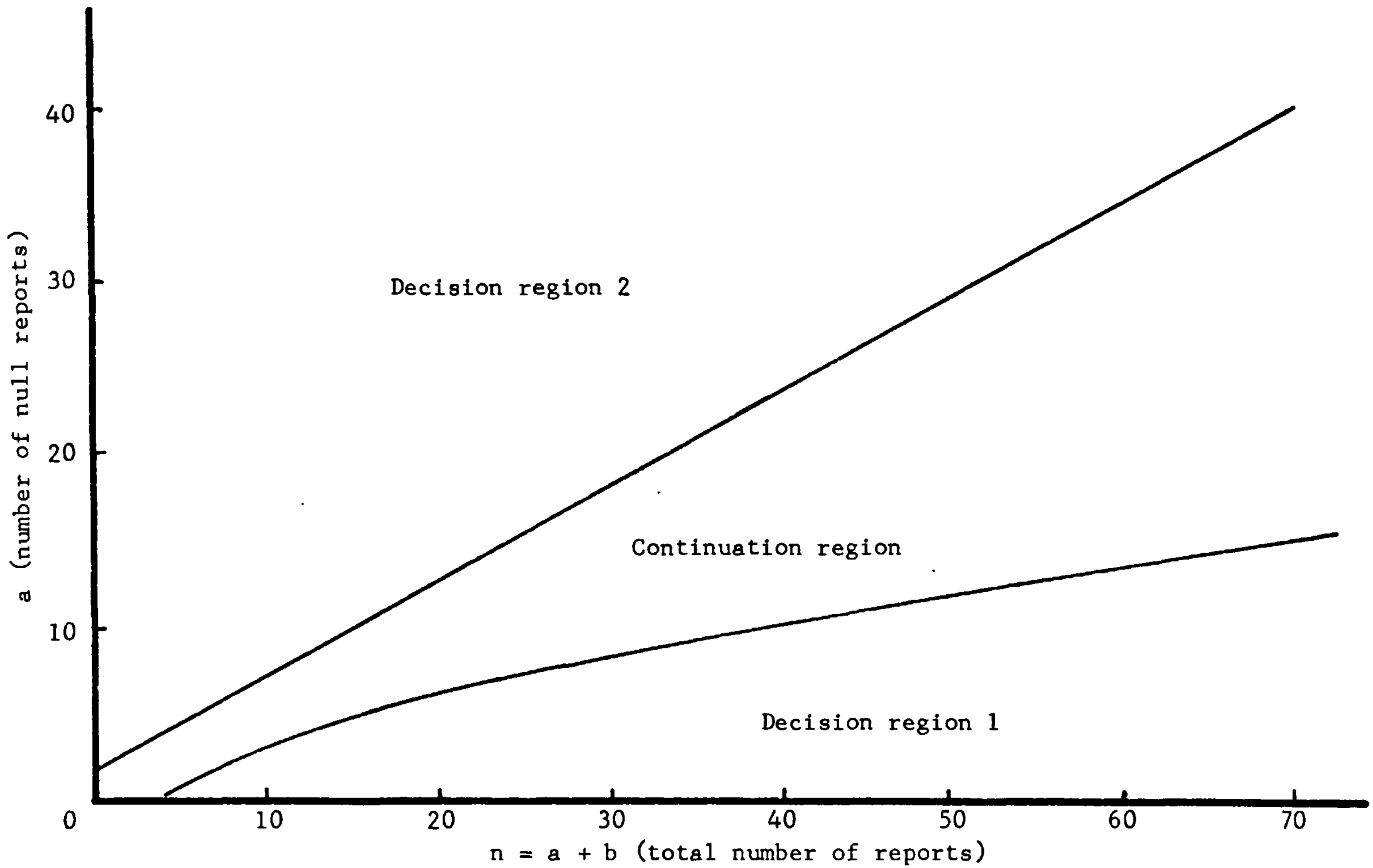

Figure 4. Decision regions - nonlinear case.

Example 2. $\theta_o = 2/3$, $k = 200$

Risk for choosing d_2 is

$$\int g(\theta \mid \alpha) W_2(\theta) d\theta$$

$$= \int_0^{\theta_o} \frac{(a+b-1)!}{(a-1)!\ (b-1)!}\ \theta^{a-1}(1-\theta)^{b-1}[k(\theta-\theta_o)]d\theta$$

Risk for continuing (awaiting one extra report) is

$$\int_0^1 g(\theta \mid \alpha)[S(\alpha,\theta) + A(\alpha,\theta)W_1(\theta) + (1-A(\alpha,\theta))W_2(\theta)]d\theta$$

CHAPTER 5

SUMMARY AND CONCLUSIONS

In this thesis we have discussed sequential decision schemes for the binomial situation with a definite loss and cost structure. We have presented the theoretical works of others showing that the inferential methods in Bayesian statistics are based on the posterior distribution, which is a combination of the prior distribution, representing the prior state of information, and the likelihood function, representing the information of an observation. We have demonstrated how we can, from an initial position of uncertainty about an unknown state of nature, take steps to remove some of that uncertainty by using the data obtained from an informative experiment and a prior distribution in the repeated applications of the conditional probability argument of Bayes' theorem. The method showed the change in the decisions involved as information was received. We also looked at another type of statistical inference whereby two states of nature were distinguished (hypothesis-testing) by the sequential probability ratio test (SPRT).

Finally, we synthesized the loss specification, the probability model-building, and the statistical inference into a statistical decision theoretic framework. The synthesis was completed by the principle of selecting as best action the one which minimized the expected loss with respect to the posterior distribution. We saw that this was a relatively simple concept and there were easy schemes within this framework of following a choice of alternatives in informative experiments and of permitting the loss structure to depend on the observation of the informative experiment.

The most useful value of the decision theoretic approach to the detection problem is most probably that it is a technique for obtaining decision boundaries whenever possible relatively simply and quickly. Once the boundaries are defined, decisions about the states of nature of the range can be made rapidly and automatically. This is a highly essential feature in the detection problem since quick identification of a target is most desirable. Also with the boundaries defined, we can represent the decision procedure in a convenient graphic form. In the application of Figure 2 we can move horizontally one step if a positive report is recorded and move one step diagonally upward if a null report is recorded. We start at the origin and as long as we stay in the region marked "continuation" we move onto the next observation. If we cross one of the boundaries we stop and take the action indicated. While decision problem may seem quite a difficult one and may require careful analysis, it is easy to give instruction in this graphic form.

Whereas a nonsequential scheme requires a fixed time requirement in that such a scheme must observe a prescribed number of reports, the sequential scheme has the feature wherein the course of the experiment depends in some way upon the results obtained. This implies a variable time requirement which hopefully is less than the fixed time requirement. Finally, the decision theoretic approach allows the use of loss function to be a means of balancing the cost of sampling reports against the costs of making the wrong decision. This is most valuable in the detection problem where the costs of data collection and wrong decision are substantial and commensurate.

The viewpoint of risk, throughout the thesis, is taken as loss averaged over a prior distribution for the uncertainty θ, the expected value point of view, rather than the minimax approach. Wald (1950) proposed that as an alternative to assuming a prior distribution (the expected value principle), first the maximum of the loss function $W_i(\theta)$ with respect to θ is found and then the ith function of the set of loss functions is chosen which minimizes this maximum,

$$\min_{i} \max_{\theta} \; W_i(\theta) \; .$$

Wald showed that under the minimax solution under certain conditions is a Bayes' solution with respect to the least favorable prior distribution, and this fact often provides a technique to obtain minimax solution [see Wald (1950), pp. 16-18].

The minimax solution is criticized by many statisticians as being over-pessimistic. The value of θ for which $W_i(\theta)$ achieves a maximum may be known to occur only very rarely and in such a situation the minimax solution is governed completely by possibilities which are extremely unlikely to occur.

Suggestions for Future Studies

Because of the limited nature of this thesis, the position estimation of the target is not included. In future studies, such localization of the target can be implemented with the Bayesian decision theoretics approach or Bayesian estimation technique. Sensitivity analysis of BDT approach should be undertaken. Also one should examine the worth of additional data in the entire tracking problem (i.e., detection and localization).

We should also examine how readily adaptable the model formulated in this work applies to other situations. One specific application in which this model may be useful is in the area of water quality control. Chemical sensors can be implanted in various arrays and at critical points along a river or seashore to detect pollution or poison. Another application may be in the area of fishery. Detectors can be placed along the continental shelves to aid the fishing vessels to locate the schools of fishes and to determine whether or not it is profitable for the fishermen to lower the nets once a school of fish is located.

Conclusions

An efficiency comparison analysis between the existing method of target detection and Bayesian decision theory is not possible at this time because such an analysis would require an actual on-site test run; however, the advantages of the Bayesian method appear to be as follows:

1. The Bayesian decision theory minimizes the expected loss of a wrong decision and has the advantage that if all costs involved are accurately represented, a unique optimum procedure results, and subjective judgment involved in choice of arbitrary restrictions is removed.

2. Bayesian decision theory does not depend on the master timing system, without which the existing method cannot operate.

3. Bayesian decision theory does not need to be re-initiated as with the existing method, whenever there are missing pulse frames in the highly time-sensitive ping-train. The theory requires only a running total count of the number of positive and null reports during which the experiment is monitored.

4. Bayesian decision theory involves less bookkeeping computation and therefore requires less computer time and memory for operation.

APPENDIX A

PROPERTIES OF BETA DISTRIBUTION

We note here that the Beta distribution possesses the desirable properties for conjugate families of distribution; that is, the Beta distribution possesses the following characteristics:

1. Mathematical tractability (i.e., the computational burdens associated with Bayesian decision theory) since it is relatively easy to specify the posterior distribution given the prior distribution and the likelihood function; it results in a posterior distribution that is also a member of the same conjugate family, so that successive applications of the Bayes' theorem are not difficult; and it is feasible to calculate expectations from the prior distribution.

2. Richness because the Beta family includes distributions with different locations, dispersions, and shapes which are capable of representing a wide variety of states of prior information.

3. Ease of interpretation because it contains information of previous sample results.

It is important to keep in mind that a conjugate prior distribution is conjugate only with respect to a given statistical model. For additional information, see Raiffa and Schlaifer (1961).

APPENDIX B

DISCRIMINATION BETWEEN TWO POPULATIONS WHICH ARE GAUSSIAN-DISTRIBUTED

In discriminating a positive or negative response from a hydrophone, we use a continuous measurement to choose among a discrete set of alternatives. A criterion for making such a choice must be applied. This is an important part of our problem, since the number of positive responses from the array of hydrophones governs the attribute of a report from the array and since the discrimination of responses has to be done extremely frequently, rapidly, and automatically. In the digital transmission of the responses, the electronic circuits represent the negatives and positives by two levels of voltage; the observed voltages show variations about the mean levels. These variations come from instrumentation inadequacy and imperfection.

We want to establish a dividing line between the two populations of responses for decision purposes. When many decisions are made in a very short period of time, the dividing line must be calculated in advance. The variation is represented by a Gaussian distribution (see Figure 5).

We start with two strong assumptions.

1. The probability densities of the two populations are supposedly known. (We are not trying to identify new response and determine the population at the same time.)

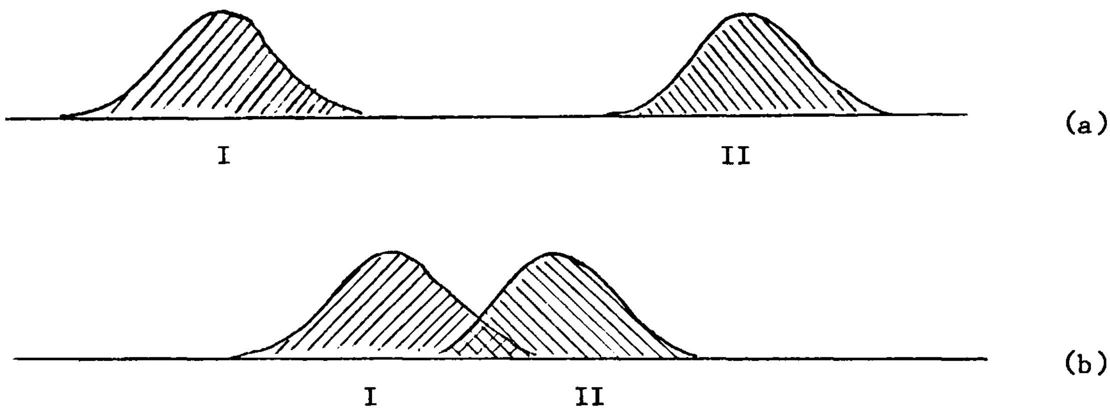

Figure 5. Probability densities representing the variation within two populations.

2. The populations are Gaussian with equal variance. Suppose our two populations are

$$\text{Population I:} \quad N(* \mid \mu_1, \sigma^2)$$

$$\text{Population II:} \quad N(* \mid \mu_2, \sigma^2)$$

(see Figure 5), and the measurement of a new response is w. Then, applying Wald's probability ratio test, we have

$$f\left(\frac{I}{II} \mid w\right) = \frac{N(w \mid \mu_1, \sigma^2)}{N(w \mid \mu_2, \sigma^2)}$$

$$= \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{w-\mu_1}{\sigma})^2]}{\frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(\frac{w-\mu_2}{\sigma})^2]}$$

$$= \exp\left[\left(\frac{\mu_1-\mu_2}{\sigma}\right)\left(\frac{w - \frac{\mu_1+\mu_2}{2}}{\sigma}\right)\right]$$

We can see that $\frac{\mu_1 + \mu_2}{2}$ is the point midway between the two means, so

$$\frac{w - \frac{\mu_1 + \mu_2}{2}}{\sigma}$$

is the distance the response is from the midpoint, measured in standard deviations. The other term, $\frac{\mu_1 - \mu_2}{\sigma}$, is the distance between the population mean, also measured in standard deviations. If w lies on the μ_1 side of the midpoint, the exponent is greater than 0 and

$$f(\tfrac{I}{II} \mid w) > 1 \;;$$

if w lies on the μ_2 side of the midpoint, then the factor is less than 1. This is as would be expected: an observation closer to μ_1 is evidence for population I; an observation closer to μ_2 is evidence for population II. The probability that the observation representing a negative response is on the wrong side of the midpoint is $1 - N(*|\mu_1, \sigma^2)$. For additional information, see Raiffa and Schlaifer (1961) and Ferguson (1967).

SELECTED BIBLIOGRAPHY

Aitchison, J., Choice Against Chance, Addison-Wesley Publishing Co., Reading, Mass., 1970, pp. 166-179, 244-247.

Barking Sands Tactical Underwater Range Systems Test Report, prepared by ITT Avionics Div. for the U. S. Navy, Oct. 1967.

Barnard, G. A., Sequential test in industrial statistics, J. Royal Stat. Soc., Supplement 8:1-26, 1946.

Barnard, G. A., Review of A. Wald's sequential analysis, J. Amer. Stat. Assn., 42:658-664, 1947.

Barnard, G. A., Studies in the history of probability and statistics: IX, Biometrika 45:293-315, 1958.

Bayes, T., An essay towards solving a problem in the doctrine of chances, Phil. Trans. Royal Soc. London 53:370-418, 1763.

Chernoff, H., and Ray, S. N., A Bayesian sequential sampling inspection plan, Annals Math. Stat. 36:1387-1407, 1965.

Davis, D. R., C. C. Kisiel, and L. Duckstein, Bayesian decision theory applied to design in hydrology, Water Resources Res. 8(1):33, 1972.

de Finetti, B., Foresight: Its Logical Laws, Its Subjective Sources, Translated by H. E. Kyburg, in Studies in Subjective Probability, H. E. Kyburg and H. E. Smokler (eds.), Wiley, N. Y., 1964.

de Groot, M. H., Optimal Statistical Decisions, McGraw-Hill Book Co., New York, 1970, pp. 267-299.

Ferguson, T. S., Mathematical Statistics: A Decision Theoretical Approach, Academic Press, New York, pp. 309-369, 1967.

Freeman, H. A., Sequential Analysis of Statistical Data: Applications, Stat. Res. Group Report 255, Columbia Univ., 1944.

Hotelling, H., Experimental determination of the maximum of a function, Annals of Math. Stat., 12:20-45, 1941.

Howard, R. A., The foundations of decision analysis, IEEE Trans. on Sys. Sci. and Cyber., ssc-4(3):211-219, Sept. 1968.

Jeffreys, H., Theory of Probability (3rd ed.), Clarendon Press, Oxford, U. K., 1961.

Keynes, J. M., A Treatise on Probability, Macmillan Co., London, U. K., 1921.

Kyburg, H. E., and H. E. Smokler (eds.), Studies in Subjective Probability, Wiley, New York, 1964.

Lindley, D. V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Cambridge Univ. Press, Cambridge, U. K., 1965.

Lindley, D. V., Bayesian Statistics, A Review, SIAM, Phila., Pa., 1970, pp. 17-34.

Mahalanobis, P. C., A sample survey of the acreage under jute in Bengal, Sankhya 4:511-531, 1940.

Neyman, J., On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, J. Royal Stat. Soc., 97:558-625, 1934.

Neyman, J., and E. S. Pearson, On the use and interpretation of certain test criteria for purposes of statistical inference, Biometrika 20A:175-240, 263-294, 1928.

Neyman, J., and E. S. Pearson, On the problem of the most efficient test of statistical hypotheses, Phil. Trans. Royal Soc. London Series A 231:289-337, 1933.

Neyman, J., and E. S. Pearson, Contributions to the theory of testing statistical hypotheses, Stat. Res. Mem., Parts I and II, 1936, 1938.

North, W., A tutorial introduction to decision theory, IEEE Trans. on Sys. Sci. and Cyber., ssc-4(3):200-210, Sept. 1968.

Pratt, J. W., H. Raiffa, and R. Schlaifer, The foundations of decision under uncertainty: an elementary exposition, J. Am. Stat. Assn. 59:353-375, 1964.

Raiffa, H., Decision Analysis, Addison-Wesley, Reading, Mass., 1968.

Raiffa, H., and R. Schlaifer, Applied Decision Theory, Harvard Business School, Boston, Mass., 1961, pp. 48-53, 124-131, 216-220.

Ramsey, F. P., Truth and Probability, The Foundations of Mathematics and Other Logical Essays, in Studies in Subjective Probability, H. E. Kyburg and H. E. Smokler (eds.), Wiley, N. Y., 1964.

Savage, L. J., _The Foundations of Statistics_, Wiley, New York, 1954.

Savage, L. J., et al., _The Foundations of Statistical Inference_, Methuen, London, U. K., 1962.

Schmitt, S. A., _Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics_, Addison-Wesley, Reading, Mass., 1969, pp. 231-270.

Stein, C., A two-sample test for a linear hypothesis whose power is independent of variance, Annals Math. Stat., _16_:243-258, 1945.

Stockman, C. M., A method of obtaining an approximation for the operating characteristic of a Wald sequential probability test applied to a binomial distribution, Brit. Min. Supply Adv. Serv., Stat. Meth. and Qual. Cont., Tech. Report QC/R/19, 1944.

Sukhatme, P. V., Contribution to the theory of the representative method, J. Royal Stat. Soc., Supplement _2_:253-268, 1935.

Thompson, W. R., On the likelihood that one unknown probability exceeds another in view of the evidence of two samples, Biometrika _25_: 285-294, 1933.

Thompson, W. R., On the theory of apportionment, Amer. J. Math., _57_:450-456, 1935.

von Neumann, J., and O. Morgenstern, _Theory of Games and Economic Behavior_, 2nd ed., Princeton University Press, Princeton, N. J., 1947.

Wald, A., Sequential analysis of statistical data: theory, Stat. Res. Group Report 75, Columbia Univ., 1943.

Wald, A., Sequential tests of statistical hypothesis, Annals Math. Stat. _16_:117-186, 1945.

Wald, A., _Sequential Analysis_, Wiley, New York, 1947.

Wald, A., _Statistical Decision Functions_, Wiley, New York, 1950, pp. 89-98, 151-155.

Wetherill, G. B., _Sequential Methods in Statistics_, Methuen, London, U. K., 1966, pp. 85-107.

Winkler, R. L., _Introduction to Bayesian Inference and Decision_, Holt, Rinehart, and Winston, Inc., 1972, pp. 132-181, 415-433.