A COMPARISON OF CLASSICAL AND BAYESIAN STATISTICAL ANALYSIS IN OPERATIONAL TESTING

A THESIS

Presented to

The Faculty of the Division of Graduate Studies and Research

By

Philip Vincent Coyle

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

June, 1977

A COMPARISON OF CLASSICAL AND BAYESIAN STATISTICAL ANALYSIS IN OPERATIONAL TESTING

Approved:

Douglas C. Montgomery, Chairman

Russell G. Heikes

Leslie G. Callahan, Jr.

Date approved by Chairman: May 30, 1977

ACKNOWLEDGMENTS

I would like to express my appreciation to my thesis advisor, Dr. Douglas C. Montgomery, for his encouragement and optimism in assisting me in the preparation of this thesis. My gratitude is also extended to Dr. Russell G. Heikes, for the many hours spent in consultation with this author, and to Dr. Leslie Callahan for his help in the preparation of this thesis.

A special note of thanks goes to my wife, Marianne, for her understanding and encouragement over the past two years.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

This research is devoted to investigating how Bayesian statistical analysis differs from classical statistical analysis in the context of operational testing. The specific aspects of operational testing which are considered are the power resulting from a hypothesis test and the expected loss, or risk, resulting from a decision.

First it is shown that it is quite difficult to develop a meaningful measure of comparison between Bayesian and classical analysis in the framework of hypothesis testing. Using the power of the hypothesis test as a measure of comparison, it is shown that under certain conditions classical statistical procedures lead to more powerful tests than Bayesian procedures. It is then shown that Bayesian statistical procedures are superior to classical procedures in the framework of minimizing expected loss or risk.

CHAPTER I

INTRODUCTION

Background

This study was prompted by the desire of the U. S. Army Operational Test and Evaluation Agency (OTEA) to compare Bayesian to classical statistical procedures for determining sample sizes for actual tests which have been conducted by OTEA. The objective of the comparison is to determine if smaller sample sizes can be obtained through the use of Bayesian procedures which yield inferences comparable to those drawn from classical procedures. To understand the procedures to be utilized in this study, one must be familiar with the nature of operational testing as performed by OTEA.

The purpose of operational testing is to provide data upon which to estimate a prospective system's military utility, operational effectiveness and suitability, and need for any modifications [2]. This data is obtained through a sequence of three operational tests (referred to as OT I, OT II, and OT III). Each test must be completed and analyzed prior to beginning the next test to determine if there is a need for the next test in the sequence. When possible the new system is tested alongside the existing system during each phase of testing to acquire data from both systems under identical conditions. At the end of each test, the data is collected and analyzed, and a decision is made to conduct the next test or to reject the new system [1].

The overall assessment procedure consists of identifying certain measures of effectiveness (MOE) which are critical to the system under consideration, such as, percent of target hits, mean miss distance, mean time between failure, and so on. Once identified, these MOE are incorporated into a test design which will provide for a side-by-side comparison of the competing systems with respect to each MOE. After all MOE of interest have been tested, the overall desirability of the system is then evaluated.

For a given test design, the problem at hand is one of determining the minimum number of replicates (sample size) required for each set of experimental conditions to achieve a specified level of confidence in the inference made as a result of the experiment. This sample size is currently being determined by classical statistical procedures [18]. As an example, suppose the random variable of interest is assumed to follow a normal distribution with unknown mean and variance, and the decision maker is interested in determining the expected value or mean of the random variable. In the classical sense, the mean is considered an unknown constant. The power of the test, or the probability of rejecting the hypothesized value of the mean, when the hypothesized value is inaccurate, is determined from the operating characteristic curves for the type of test conducted. The above theory of classical statistics will be important when compared to the Bayesian theory investigated in this study.

Objectives of Research

The objectives of this research are twofold. The first objective is to determine whether or not Bayesian methodology can be effectively applied to operational testing. As noted earlier, operational testing is conducted in three phases, and many times the same measures of effectiveness are examined in more than one phase. The current procedures used by OTEA consider each test in the sequence independently; i.e., the inferences made at the end of each test are based on the data obtained during that specific test only [18]. There is no attempt made to combine the data on a specific MOE measured in OT I and OT II, for example, to obtain a better estimate for the MOE from which better inferences can be made. Chapter III is devoted to developing a methodology which will apply Bayesian techniques to the combination of data from two phases of testing to determine the power of a hypothesis test for any specified sample size.

The second objective of this research is to determine under what conditions the Bayesian methodology will produce a "better" test than the classical methodology when considering the same sample size for both methods. Chapters III and IV are devoted to comparing the above methodologies in the context of an actual test conducted by OTEA.

Fundamentals of Bayesian Analysis

The discussion presented here will compare classical statistical theory to Bayesian statistical theory to demonstrate how OTEA's present concepts of testing would have to be altered to apply Bayesian techniques to operational testing. Presently, if OTEA is considering a data generating process which may be modeled by the normal process with unknown mean and variance, then the probability density function associated with the process is the normal density, with mean, μ, and variance, σ^2. These parameters would be viewed as unknown constants by the classical statistician. These constants are generally estimated by sampling from the data generating process and using the sample statistics $\overline{X}$ and s^2 to estimate μ and σ^2, respectively. If one is interested in μ, the mean of the process, $\overline{X}$ and s^2 could be used to construct a confidence interval on μ. For example, if $(1-\alpha)$ is the degree of confidence desired, then [12]

$$1 - \alpha = P[\overline{X} - (t_{\alpha/2,n-1})(s/\sqrt{n}) \leq \mu \leq \overline{X} + (t_{\alpha/2,n-1})(s/\sqrt{n})] \quad (1\text{-}1)$$

where n is the sample size and $t_{\alpha/2,n-1}$ is the percentage point of the central t-distribution with n-1 degrees of freedom such that $P(t > t_{\alpha/2,n-1}) = \alpha/2$. This confidence interval on μ would be interpreted in the relative frequency sense. That is, if repeated samples of size n were taken, each time computing new values of $\overline{X}$ and s^2, and a confidence interval on μ was constructed after each sample was taken, then it would be expected that $100(1-\alpha)\%$ of the confidence intervals so constructed would contain the "true" value of μ [12]. The Bayesian analyst would differ in several ways. He would consider the unknown parameters, $\tilde{\mu}$ and $\tilde{\sigma}^2$, as random variables. ("Tildes" will be used to indicate random variables throughout this study.) Since point estimates of random variables are useless, he would ascribe to them a probability distribution instead. If prior sampling information is not available, the analyst must use his subjective knowledge of the process to assess a probability distribution for the joint occurrence of $\tilde{\mu}$ and $\tilde{\sigma}^2$. This prior distribution can then be combined with sample information to produce new distributions for the unknown parameters, as will be demonstrated below. The conceptual differences between the classical and the Bayesian analyst play important roles in interpreting the results of a test [36].

The combination of a prior probability distribution of a random variable with sample information is achieved by use of Bayes' theorem. For a continuous random variable, $\tilde{\theta}$, Bayes' theorem may be written as

$$f''(\theta|y) = \frac{f'(\theta)f(y|\theta)}{\int_{-\infty}^{\infty} f'(\theta)f(y|\theta)d\theta}, \tag{1-2}$$

where a single prime superscript (') denotes a prior distribution or parameter, a double prime superscript (") denotes a posterior distribution or parameter, and no superscript denotes a sampling distribution on parameter.* Therefore, in equation (1-2), $f'(\theta)$ is the prior distribution of $\tilde{\theta}$ representing the analyst's beliefs regarding $\tilde{\theta}$ prior to sampling, $f(y|\theta)$ represents the likelihood function chosen to describe the sampling process, and $f''(\theta|y)$ is the posterior distribution of $\tilde{\theta}$ representing the analyst's beliefs regarding $\tilde{\theta}$ after sampling [36]. The theorem can also be applied to discrete random variables by substituting probability mass functions for probability density functions and a summation sign for the integral sign. Winkler [36] gives a

*Appendix 1 presents a detailed explanation of all notation in this study.

derivation of Bayes' theorem from conditional probability formulas. In applying Bayes' theorem, the major difficulties lie in assessing the prior distribution and likelihood function and in evaluating the integral in the denominator of the formula. Baker [4] has suggested methods for handling these difficulties which are discussed in the next chapter and which will be used in this study.

CHAPTER II

BAYESIAN DISTRIBUTION THEORY

In his thesis, Baker [4] considered a problem similar to the one addressed in Chapter I. He has proposed a methodology for combining data relative to a single MOE taken from one phase of testing with sample information on the same MOE taken in a later phase of testing. This procedure produces an estimate of the MOE for use in making decisions. The methodology applies to an operational test in which a proposed system is being tested side-by-side with the system it has been designed to replace, and a single MOE is under consideration. In general, this methodology uses the theory of selecting a prior distribution from the natural conjugate family of distributions which, when combined with the likelihood function in Bayes' theorem, produces a posterior distribution that will be of the same form as the prior. This will reduce the computational burden considerably in the sequential analysis used in this study. (For a complete discussion of natural conjugate distributions, see Raiffa and Schlaiffer [29], Chapter 3.)

In this study, the results of an actual operational test are supplied by OTEA. When considering a single MOE, OTEA assumes the univariate normal distribution with unknown mean and variance as the basic model for sample size determination for both measurement and attribute data [18]. The same function will, therefore, be used in this study as the likelihood function for the random variable under consideration.

The side-by-side nature of the operational tests under consideration suggests that inferences be drawn from the difference of performance characteristics of the systems rather than from the actual performance characteristics of a single system. Thus, if $\tilde{X}_1$ and $\tilde{X}_2$ represent the same MOE for systems one and two, respectively, $\tilde{D} = \tilde{X}_1 - \tilde{X}_2$ will represent the difference between the MOE of the two systems. Since $\tilde{X}_1$ and $\tilde{X}_2$ are assumed to follow the normal distribution with unknown mean and variance, $\tilde{D}$, which is just a linear combination of two independent, normally distributed random variables, can also be assumed to follow a normal distribution [12] with unknown mean and variance, say $\tilde{\mu}$ and $\tilde{\sigma}^2$, respectively. The variable of interest in this study will be $\tilde{\mu}$, the mean difference between the two systems.

In the classical sense μ, the mean of the distribution of $\tilde{D}$, is considered to be an unknown <u>constant</u>, and inferences are drawn from tests of hypothesized values of μ. Consequently, if μ can be shown to be equal to zero, one can conclude that there is no difference between the competing systems, whereas if μ is not equal to zero, then one can conclude that one system is better than the other.

In the Bayesian sense, since $\tilde{\mu}$ is considered as a random variable, tests on whether or not $\tilde{\mu}$ takes on a specific value are meaningless. One must consider tests where $\tilde{\mu}$ can take on a range of values; e.g., $\tilde{\mu} \leq \mu_0$, or one can consider a test on specific values of $\overline{\mu}$, the mean of $\tilde{\mu}$. If $\bar{\mu}$ can be shown to be equal to zero, one could reasonably conclude that there is no difference between competing systems, and if $\bar{\mu} \neq 0$, there is a difference.

It has been shown [29] that when $\tilde{\mu}$ is considered as a random variable, the distribution of $\tilde{\mu}$ is the Student's t distribution, represented by the density

$$f(\tilde{\mu}|m,v,n,\nu) = f_S(\tilde{\mu}|m,n/v,\nu), \qquad (2\text{-}1)$$

where (m,v,n,ν) is the statistic resulting from a sample of size n and is given by

$$m = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} D_i \qquad (2\text{-}2)$$

$$v = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - m)^2$$

$$\nu = n - 1 \, .$$

The parameters $(m,n/v,\nu)$ in the argument of f_S on the right side of equation (2-1) indicate the degree of non centrality of the distribution. The central or standard Student's t distribution would be given by $f_S(\tilde{\mu}|0,1,\nu)$. The distribution given in equation (2-1) can be standardized so that cumulative t tables can be used in computing probabilities as follows:

$$P(\tilde{\mu} < \mu|m,v,n,\nu) = F_{S^*}([\mu-m]\sqrt{n/v}|\nu),$$

where the subscript S^* indicates the standard Student's t distribution. It has also been shown [29] that the mean and variance of $\tilde{\mu}$ are given by

$$E(\tilde{\mu}|m,v,n,\nu) \equiv \bar{\mu} = m \qquad \nu > 1 \tag{2-3}$$

$$V(\tilde{\mu}|m,v,n,\nu) \equiv \breve{\mu} = \frac{v}{n}\frac{\nu}{\nu-2} \quad \nu > 2$$

The objective of this methodology is then to determine the minimum sample size which will produce a posterior distribution of $\tilde{\mu}$ that will enable the decision maker to achieve a specified level of confidence in the inference drawn concerning $\tilde{\mu}$.

Since the Department of the Army has imposed on OTEA the requirement that operational testing be independent of all other testing [2], it has been assumed that prior to OT I the state of knowledge concerning $\tilde{\mu}$ can be represented by a diffuse distribution for the normal-gamma family, as developed in Winkler [36]. Thus, when the prior distribution is combined with the sample information from OT I the resulting posterior distribution will also be normal-gamma [36]. When a measure of effectiveness that was considered in OT I is being reconsidered in OT II, it must be assumed that the posterior standard deviation of $\tilde{\mu}$, $\sqrt{\breve{\mu}''}$, determined in OT I was too large to reach a meaningful conclusion about $\tilde{\mu}$. The sequential nature of the testing then presents the opportunity to use the posterior distribution determined from OT I regarding $\tilde{\mu}$ as the prior state of knowledge of $\tilde{\mu}$ for OT II. The methodology now concentrates on developing a sample size for OT II which will produce a posterior standard deviation for $\tilde{\mu}$ equal to some fraction of the prior standard deviation; i.e., $\sqrt{\breve{\mu}''} = s\sqrt{\breve{\mu}'}$, where $0 < s \leq 1$.

Baker [4] has shown that a sample of size

$$n \cong (\frac{1}{s^2} - 1)n', \quad 0 < s \leq 1$$

where n' represents the sample size of the prior distribution, can be expected to reduce the prior standard deviation of $\tilde{\mu}$ by a factor s. He approximated $E(\sqrt{\tilde{\mu}''})$ with

$$E(\sqrt{\tilde{\mu}''}|m',v',n',\nu';\ n,\nu) = \sqrt{\frac{n\,\tilde{\mu}'}{n+n'}} \qquad (2\text{-}4)$$

Due to the approximations used in his formulation, Baker [4] has introduced an error into the expected posterior standard deviation which can be written as

$$\%\ \text{error} = 1 - \exp\left[-3/4\ \left(\left(\frac{1}{\nu'-2}\right) - \left(\frac{1}{\nu'+n-2}\right)\right)\right]. \qquad (2\text{-}5)$$

If this error is determined by the decision maker to be too large, then equation (2-4) cannot be used, and a more complex formula must be used to determine the sample size, n, which will produce a desired expected posterior standard deviation of $\tilde{\mu}$. This equation is

$$E[\sqrt{\tilde{\mu}''}\ |m',v',n',\nu';n,\nu] = \sqrt{(n'/n'')\tilde{\mu}'}\ \exp -\left[\frac{3}{4}\left(\left(\frac{1}{\nu'-2}\right) - \left(\frac{1}{\nu'+n-2}\right)\right)\right], \qquad (2\text{-}6)$$

where $n'' = n'+n$.

Although equation (2-6) cannot be solved explicitly for n, given a desired value of $E(\sqrt{\tilde{\mu}''})$, it can be solved iteratively. Baker has suggested a starting value of n to be that found by solving equation (2-4) for n.

Once a sample size has been determined and a sample has been taken, the statistic (m'',v'',n'',ν'') is determined [29] as follows:

$$m'' = \frac{n'm'+nm}{n'+n} \tag{2-7}$$

$$n'' = n' + n$$

$$v'' = \frac{[\nu'v'+n'(m')^2] + (\nu v+nm^2) - n''(m'')^2}{[\nu'+\delta(n')] + [\nu + \delta(n)] - \delta(n'')}$$

$$\nu'' = [\nu' + \delta(n')] + [\nu + \delta(n)] - \delta(n''),$$

where $\delta(\gamma) = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 & \text{if } \gamma > 0 \end{cases}$.

The mean and variance of the posterior distribution of $\tilde{\mu}$ are then

$$E(\tilde{\mu}'') \equiv \bar{\mu}'' = m'' \tag{2-8}$$

$$V(\tilde{\mu}'') \equiv \breve{\mu}'' = \frac{v''\nu''}{n''(\nu''-2)} .$$

In the case where the prior distribution is diffuse, as in OT I, $n' = \nu' = 0$, and the posterior parameter (m'',v'',n'',ν'') equals the sample statistic (m,v,n,ν) [29].

The above development is directed at producing a value of the posterior standard deviation of $\tilde{\mu}$ which will make the distribution of $\tilde{\mu}$ "tight" enough to enable the decision maker to make his decision concerning $\tilde{\mu}$ with a specified degree of confidence. However, the value of $\sqrt{\breve{\mu}''}$ which satisfies the above criterion is subjective in nature. The problem of determining values of $\sqrt{\breve{\mu}''}$ which meet certain criteria will be discussed in Chapter III.

CHAPTER III

CLASSICAL VS. BAYESIAN HYPOTHESIS TESTING

Introduction

In this chapter an attempt will be made to compare Bayesian and classical statistical methods in the context of hypothesis testing. One commonly accepted measure of comparison between methods of testing hypotheses is the power of each test. We shall define the power of a test as the probability of rejecting the null hypothesis when it is false, or, equivalently, the probability of not committing a type II error. The power of the test is an appropriate measure of comparison for this study because of the consequences of the decisions resulting when type II errors are made. In the case of operational testing, consider the null hypothesis: there is no difference between the standard equipment and its proposed replacement versus the alternate hypothesis: the proposed replacement is better than the standard. If the decision maker makes a type II error (i.e., the new equipment is better but it will not be purchased), he is denying the army the use of a better piece of equipment and thereby keeping the level of mission accomplishment lower than it could be.

In the case of a type I error, however, where the decision maker rejects the null hypothesis when it is true (i.e., there is no difference in equipment but the new equipment is purchased), the consequence would be that a probably more expensive piece of equipment would be purchased which would not improve the mission accomplishment of the army. A better piece of equipment would not have been overlooked, however.

In this example, a type II error could be more harmful to the army than a type I error. For this reason, the probability of not committing a type II error, or the power of the test, is considered of prime importance in this study.

The Two-tailed Hypothesis Test

To compare classical versus Bayesian tests in terms of power, the hypotheses of interest in both tests must be considered. In the classical two-tailed test, H_0: $\mu = 0$ vs H_1: $\mu \neq 0$ (μ is considered a constant), the type I error can be fixed at any desired level, and the type II error can be determined for any given sample size by use of the appropriate operating characteristic curves. However, since the Bayesian considers μ to be a continuous random variable, the probability that $\tilde{\mu} = 0$ will always be zero. In fact, Winkler [36] has stated that there is no logical Bayesian equivalent to the classical two-tailed test. Two modified Bayesian hypotheses will, therefore, be considered in this study. The first tests whether or not the mean of $\tilde{\mu}$, $\overline{\mu}$, equals zero; i.e., H_0: $\overline{\mu} = 0$ vs H_1: $\overline{\mu} \neq 0$. Since the variance of $\tilde{\mu}$ decreases as n increases, an infinite sample would yield exact knowledge of the true $\tilde{\mu}$. In the infinite sample case, the mean of $\tilde{\mu}$ would be the exact value of μ when the variance of $\tilde{\mu}$ is zero. It is, therefore, logical to compare the value of $\overline{\mu}$ in the Bayesian test to the value of μ in the classical test. This will be done in the next section.

The second modified Bayesian hypothesis tests whether or not $\tilde{\mu}$ lies in some interval about zero; i.e. H_0: $-a \leq \tilde{\mu} \leq a$ vs H_1: $-a > \tilde{\mu}$ or $a < \mu$, $a > 0$. Since the classical decision maker is really more interested in knowing whether μ is in some small interval about zero rather than if μ is exactly equal to zero, this Bayesian hypothesis would also serve as a valid comparative to the classical two-tailed test. This comparison will be discussed in connection with the one-tailed test later in this chapter.

<u>Solution Using Bayesian Prediction Interval</u>

In this section, the hypotheses H_0: $\bar{\mu} = 0$ vs H_1: $\bar{\mu} \neq 0$ in the Bayesian context will be compared with the hypotheses, H_0: $\mu = 0$ vs H_1: $\mu \neq 0$ in the classical context. The measure of comparison will be the power of the test. As stated in Chapter II, $\tilde{D}$, the difference between the same MOE of two competing systems, is assumed to follow a normal distribution with unknown mean, $\tilde{\mu}$, and unknown variance, $\tilde{\sigma}^2$. In the classical test, a sample size can be determined which will yield a specified power for the test for any fixed type I error, α. The rejection criteria for H_0, established from the α level desired, is [12]: reject H_0 if $|t_0| > t_{\alpha/2,n-1}$.

$$t_0 = \text{test statistic} = \frac{m-0}{v/\sqrt{n}}$$

$$m = \text{sample mean} = \frac{1}{n}\sum_{i=1}^{n} D_i$$

$$v = \text{sample variance} = \frac{1}{n-1}\sum_{i=1}^{n} (D_i-m)^2$$

$$t_{\alpha/2,n-1} = \text{value of } t \text{ such that } P(|t| > t_{\alpha/2,n-1}) = \alpha/2$$

The power of the test for various sample sizes and departures of μ from 0 are given by the appropriate operating characteristic curves for the 2-tailed t test in [12].

Before defining the power of the Bayesian test, some discussion of a Bayesian prediction interval is needed. A Bayesian prediction interval (BPI) is an interval having a stated probability, e.g., $(1-\gamma)$, of containing the variable of interest. In Figure 1, $\bar{\mu}''$ is the mean of the posterior distribution of $\tilde{\mu}$, a is the lower prediction limit, b is the upper prediction limit, and the shaded area is the probability that $a \leq \tilde{\mu}'' \leq b$.

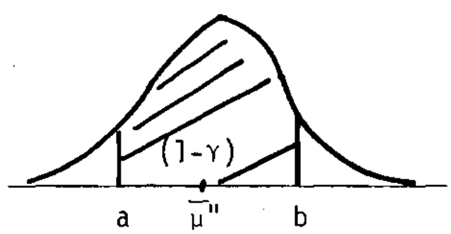

Figure 1. Generalized Bayesian Interval on $\tilde{\mu}''$.

If the interval is centered on $\bar{\mu}''$, the length of the interval, d", is given by [12]

$$d'' = 2t_{\gamma/2,\nu''} \sqrt{\breve{\mu}''} \,. \tag{3-1}$$

When considering the Bayesian hypotheses, H_0: $\bar{\mu} = 0$ vs H_1: $\bar{\mu} \neq 0$, the rejection criteria to be used in this section will be: reject H_0 if zero does not fall in the $(1-\gamma)$ BPI on $\tilde{\mu}''$. The type I error, α, is

$$\alpha = P\ (\text{rejecting } H_0 | \bar{\mu} = 0),$$

which can be restated as

$$\alpha = P(0 \text{ is not in } (1-\gamma)\text{BPI} | \bar{\mu} = 0). \tag{3-2}$$

The power of the test is defined as

$$\text{Power} = P\ (\text{rejecting } H_0 | \bar{\mu} = c \neq 0),$$

which can be restated as

$$\text{Power} = P\ (0 \text{ is not in } (1-\gamma)\text{BPI} | \bar{\mu} = c). \tag{3-3}$$

Using d" from equation (3-1), the power becomes

$$\text{Power} = P\ (0 \text{ is not in interval} [\bar{\mu}'' - d''/2,\ \bar{\mu}'' + d''/2] | \bar{\mu} = c). \tag{3-4}$$

Since $\bar{\mu}'' = m''$ (defined by equation (2-7)), equation (3-4) becomes

$$\text{Power} = P\ (0 \text{ is not in interval} [m'' - d''/2,\ m'' + d''/2] | \bar{\mu} = c). \tag{3-5}$$

Prior to sampling, m" and d" are random variables, denoted $\tilde{m}''$ and $\tilde{d}''$, which lead to

$$\text{Power} = P\ (0 \text{ is not in interval} [\tilde{m}'' - \tilde{d}''/2, \tilde{m}'' + \tilde{d}''/2] | \bar{\mu} = c). \tag{3-6}$$

Since zero will not be in the BPI only if the end points of the BPI have the same sign,

$$\text{Power} = P\ (\tilde{m}'' - \tilde{d}''/2 < 0 \text{ and } \tilde{m}'' + \tilde{d}''/2 < 0 | \bar{\mu} = c) + \tag{3-7}$$

$$P(\tilde{m}'' - \tilde{d}''/2 > 0 \text{ and } \tilde{m}'' + \tilde{d}''/2 > 0 | \bar{\mu} = c) .$$

Equation (3-7) is equivalent to

$$\text{Power} = P(|\tilde{m}''| > \tilde{d}''/2 \,|\, \bar{\mu} = c). \tag{3-8}$$

Substituting the value of d" given in equation (3-1),

$$\text{Power} = P(|\tilde{m}''| > t_{\gamma/2,\nu''}\sqrt{\tilde{\mu}''} \,|\, \bar{\mu} = c). \tag{3-9}$$

Since $\sqrt{\tilde{\mu}''}$ is always greater than 0,

$$\text{Power} = P(|\tilde{m}''|/\sqrt{\tilde{\mu}''} > t_{\gamma/2,\nu''} \,|\, \bar{\mu} = c). \tag{3-10}$$

It has been shown [29] that $\tilde{m}''$ follows a non central t distribution and $\sqrt{\tilde{\mu}''}$ follows an inverted beta 1 distribution. It is, therefore, very difficult to calculate the power of the test from the expression given in equation (3-10). To simplify the calculations, $\sqrt{\tilde{\mu}''}$ will be replaced by its expected value, as given in equation (2-4), and the resulting power computation is considered to be an approximation to the power in equation (3-10). After replacing $\sqrt{\tilde{\mu}''}$ with its expected value as given in equation (2-4) and letting $k = t_{\gamma/2,\nu''}\, E(\sqrt{\tilde{\mu}''})$, equation (3-9) becomes

$$\text{Power} = P(|\tilde{m}''| > k \,|\, \bar{\mu} = c) \tag{3-11}$$

$$= P(\tilde{m}'' > k \,|\, \bar{\mu} = c) + P(\tilde{m}'' < -k \,|\, \bar{\mu} = c) \tag{3-12}$$

Using $\tilde{m}''$ as given in equation (2-7), equation (3-12) becomes

$$\text{Power} = P\left(\frac{m'n' + mn}{n''} > k \,|\, \bar{\mu} = c\right) + P\left(\frac{m'n' + mn}{n''} < -k \,|\, \bar{\mu} = c\right) \tag{3-13}$$

Equivalently,

$$\text{Power} = P(\tilde{m} > \frac{kn'' - m'n'}{n} \mid \bar{\mu} = c) + P(\tilde{m} < \frac{-kn'' - m'n'}{n} \mid \bar{\mu} = c) \quad (3\text{-}14)$$

$$= 1 - P(\tilde{m} < \frac{kn'' - m'n'}{n} \mid \bar{\mu} = c) + P(\tilde{m} < \frac{-kn'' - m'n'}{n} \mid \bar{\mu} = c) . \quad (3\text{-}15)$$

It has been shown [29] that the distribution of $\tilde{m}$ is given by

$$D(\tilde{m}|m',v',n',\nu';n,\nu) = f_S(\tilde{m}|m',n_u/v',\nu'), \qquad (3\text{-}16)$$

where $n_u = \frac{n'n}{n+n'}$.

Strictly speaking, the Bayesian analyst does not consider $\bar{\mu}$ to be an unknown constant with some true value. Rather, he considers $\bar{\mu}$ to be a random variable also. However, in order to use Bayesian procedures to formulate a test which can be compared to the classical hypothesis test, it has been assumed that there is some true value for $\bar{\mu}$. With this assumption, the expected value of the sample mean, $E(\tilde{m})$, would then be equal to $\bar{\mu}$. Cumulative probabilities for $\tilde{m}$ would then be computed as given below.

$$P(\tilde{m} < a|\bar{\mu},n_u,v',\nu') = F_{S^*}([a - \bar{\mu}]\sqrt{n_u/v'}|\nu') \qquad (3\text{-}17)$$

Using equation (3-17) with $\bar{\mu} = c$, equation (3-15) can be rewritten as

$$\text{Power} = 1 - F_{S^*}(\frac{kn'' - m'n' - cn}{n\sqrt{v'/n_u}}|\nu') + F_{S^*}(\frac{-kn'' - m'n' - cn}{n\sqrt{v'/n_u}}|\nu'). \quad (3\text{-}18)$$

Summarizing the method for determining the power of the Bayesian test for a given sample size and a prior statistic (m',v',n',ν'):

1. Calculate $E(\sqrt{\tilde{\mu}''})$ from equation (2-4).

$$E(\sqrt{\tilde{\mu}''}) = \sqrt{\frac{\mu' n'}{n'+n}}$$

2. Calculate ν'' and n_u from equations (1-5) and (3-16).

$$\nu'' = n' + n - 1$$

$$n_u = \frac{n'n}{n+n'}$$

3. Calculate $k = t_{\gamma/2,\nu''} \; E(\sqrt{\tilde{\mu}''})$.
4. Calculate power from equation (3-18) for any value of c.

Illustrating the Procedure

In this section, the solution procedure described above will be illustrated in the context of an actual operational test conducted by OTEA. The test selected was an OT II for the Lightweight Company Mortar System (LWCMS), which is being considered as a replacement for the 81 mm mortar currently being used by the army. The purpose of the test was to provide data for a side-by-side comparison of the two mortars to assess the relative operational performance and military utility of the LWCMS [20]. One of the MOE which was considered in both OT I and OT II was the time required for an individual to complete the gunner's examination, which is a test designed to determine how quickly an individual can perform critical operations in preparing a mortar to fire. In OT I a sample size 14 was used to determine the distribution of times to perform the gunner's test. The results of this test are contained in Appendix 2. If $\tilde{X}_1$ and $\tilde{X}_2$ represent the times to perform the test on the old system and the new system, respectively, then $\tilde{D} = \tilde{X}_1 - \tilde{X}_2$ is the variable described in Chapter II. The mean of $\tilde{D}$, $\tilde{\mu}$, is the variable of interest in this study.

Using a diffuse prior distribution, Baker [4] determined the parameters of the posterior distribution of $\tilde{\mu}$ for OT I from equation (2-7) to be

$$m'' = m = 17.6 \text{ sec}$$

$$n'' = n = 14$$

$$v'' = 2040.5 \text{ sec}^2$$

$$\nu'' = 13$$

Since the same MOE was also tested in OT II, the above values will be used in the prior distribution of $\tilde{\mu}$ for OT II. The value of the prior variance of $\tilde{\mu}$ for OT II is computed from equation (2-3).

$$\breve{\mu}' = \frac{v'}{n'}\,\frac{\nu'}{(\nu'-2)} = \frac{(2040.5)(13)}{(14)(11)} = 172.25 \text{ sec}^2$$

In OT II, OTEA used a sample of 30 individuals, each of whom performed the gunner's test twice on each of the competing systems. The average times for each individual on each system are given in Appendix 3. The power curve for the classical two-tailed test with $n = 30$ will be compared to the power curve for the Bayesian test with $n = 30$. The step-by-step procedure for calculating the power using the statistic $(m', v', n', \nu') = (17.6,\ 2040.5,\ 14, 13)$, $n = 30$, and a 95% Bayesian prediction interval ($\gamma = .05$) is given below.

1. $E(\sqrt{\tilde{\mu}''}) = \dfrac{(172.25)(14)}{30+14} = 7.40$ sec

2. $\nu'' = 14 + 30 - 1 = 43$

$$n_u = \frac{30(14)}{30+14} = 9.55$$

3. $k = t_{.025,43}\, E(\sqrt{\tilde{\mu}''}) = 2.02(7.40) = 14.95$

4. $\text{Power} = 1 - F_{S^*}\left[\dfrac{(14.95)(44) - (17.6)(14) - 30c}{30\sqrt{2040.5/9.55}}\,\middle|\, 13\right]$

$$+ F_{S^*}\left[\frac{(-14.95)(44) - (17.6)(14) - 30c}{30\sqrt{2040.5/9.55}}\,\middle|\, 13\right]$$

The cumulative distribution for the standard Student's t distribution is given in <u>Biometrika Tables for Statisticians</u>, Volume 1, by Pearson and Hartley [22]. The power for c = 20 is calculated to be

$$\begin{aligned}\text{Power} &= 1 - F_{S^*}(-.2|13) + F_{S^*}(-3.9|13)\\ &= 1 - .42 + .0009\\ &= .58\end{aligned}$$

The power for other values of c is calculated in a similar manner. Since the value of γ in the $(1-\gamma)$ BPI is not the type I error for this test, the type I error must also be calculated for each value of n. The type I error, α, is given by

$$\begin{aligned}\alpha &= P(\text{rejecting } H_0 | \bar{\mu} = 0)\\ &= P(0 \text{ is not in } (1-\gamma)\text{BPI} | \bar{\mu} = 0)\end{aligned}$$

This formula is the same as the formula for the power with $\bar{\mu} = c = 0$. Therefore,

$$\alpha = 1 - F_{S^*}\left[\frac{(14.95)(44) - (17.6)(14)}{30\sqrt{2040.5/9.55}} \middle| 13\right]$$

$$+ F_{S^*}\left[\frac{(-14.95)(44) - (17.6)(14)}{30\sqrt{2040.5/9.55}} \middle| 13\right],$$

$$= 1 - F_{S^*}\,(.94|13) + F_{S^*}(-2.1|13)$$

$$= 1 - (.82) + .03 = .21$$

In order to fix α at a certain level, as is done in the classical case, the width of the $(1-\gamma)$ BPI must be changed with each value of n; i.e., γ must change with each n to keep α fixed. To calculate the value of γ which produces a given α, consider eq. (3-18) with c = 0.

$$\alpha = 1 - F_{S^*}\left(\frac{kn'' - m'n'}{n\sqrt{v'/n_u}} \middle| \nu'\right) + F_{S^*}\left(\frac{-kn'' - m'n'}{n\sqrt{v'/n_u}} \middle| \nu'\right) \tag{3-19}$$

For positive m', the last term in equation (3-19) is insignificant. Therefore, letting $\alpha = .05$ and dropping the last term yields

$$.95 = F_{S^*}\left(\frac{kn'' - m'n'}{n\sqrt{v'/n_u}} \middle| 13\right) . \tag{3-20}$$

The value of the argument in the right side of equation (3-20) which yields a probability of .95 is 1.8 [22]. Thus

$$\frac{kn'' - m'n'}{n\sqrt{v'/n_u}} = 1.8$$

Equivalently,

$$k = \frac{1.8\ n\sqrt{v'/n_u} + m'n'}{n''} \tag{3-21}$$

For n = 30,

$$k = \frac{1.8(30)\sqrt{2040.5/9.55} + (17.6)(14)}{30 + 14}$$

$$= 23.54\ .$$

Since $k = t_{\gamma/2,\nu''}\ E(\sqrt{\tilde{\mu}''})$ by definition,

$$t_{\gamma/2,43} = \frac{k}{E(\sqrt{\tilde{\mu}''})} = \frac{23.54}{7.40} = 3.18\ .$$

Thus, $\gamma \cong .005$, or a 99.5% BPI will produce a type I error of .05. Table 1 lists the values of k needed to produce $\alpha = .05$ for various values of n.

Table 1. Sample Size versus k, $\alpha = .05$

Sample Size	k
2	23.08
4	23.93
7	24.28
10	24.30
15	24.13
20	23.91
30	23.54
40	23.27

With the type I error fixed at .05, the power curves for the Bayesian and classical tests for $n = 30$ and $n = 4$ have been plotted in Figures 2 and 3, respectively. From the figures it can be seen that for $n = 4$, the Bayesian test is slightly more powerful but for $n = 30$ the classical test is much more powerful. Plots of the power versus sample size for each test for $|c| = 20$ and $|c| = 40$ are shown in Figures 4 and 5, respectively. There it is evident that the classical test is superior to the Bayesian test in detecting both small and large values of $|c|$, particularly when large values for the power are required.

We will now investigate the behavior of the power curves if γ in the $(1 - \gamma)$ BPI is held constant at $\gamma = .05$. In this case, the type I error will not remain fixed as it did in the previous calculations. The type I error can be computed from equation (3-19) for various sample sizes. The results of the calculations are given in Table 2.

Table 2. Type I Error Versus Sample Size

Sample Size	Type I Error
2	.01
4	.04
7	.08
10	.11
15	.14
20	.16
30	.21
40	.25

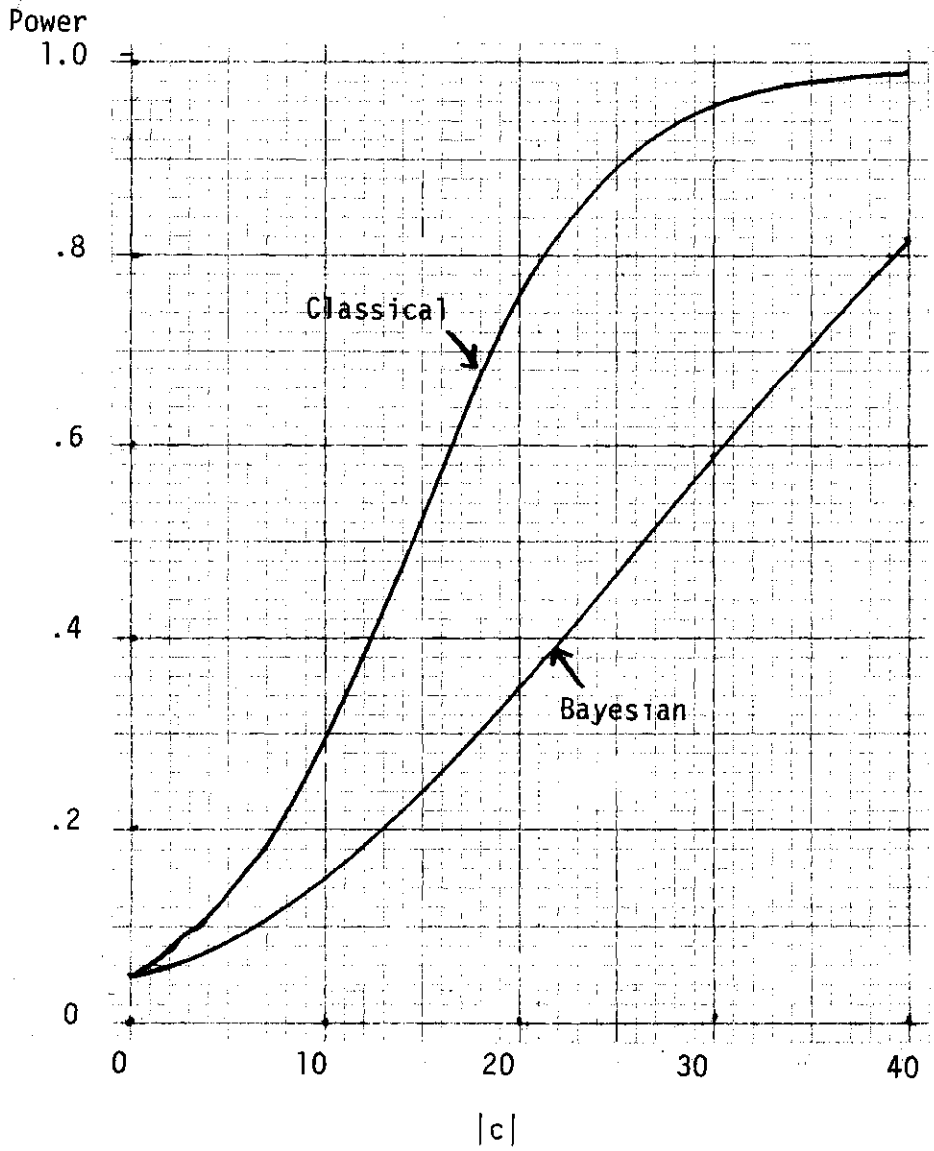

Figure 2. Power Curves, n = 30, α = .05.

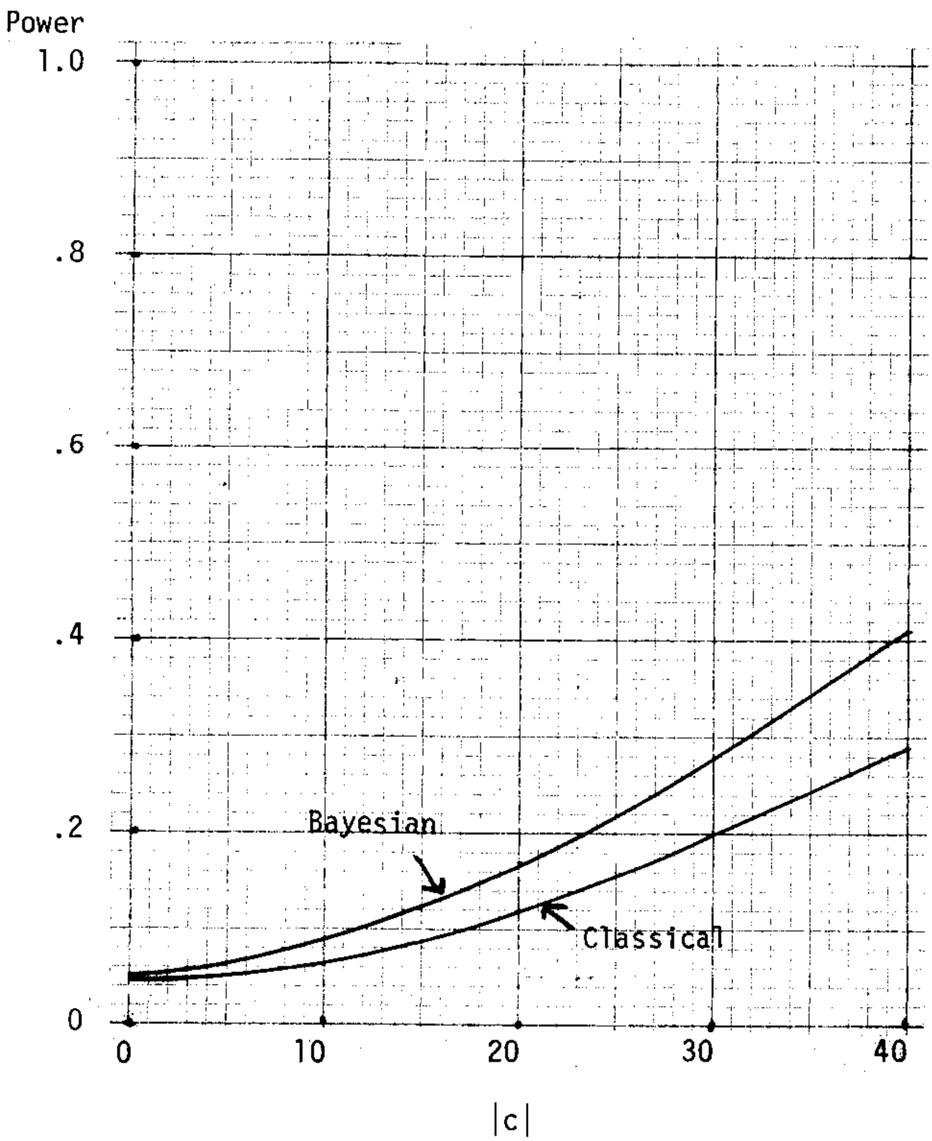

Figure 3. Power Curves, $n = 4$, $\alpha = .05$.

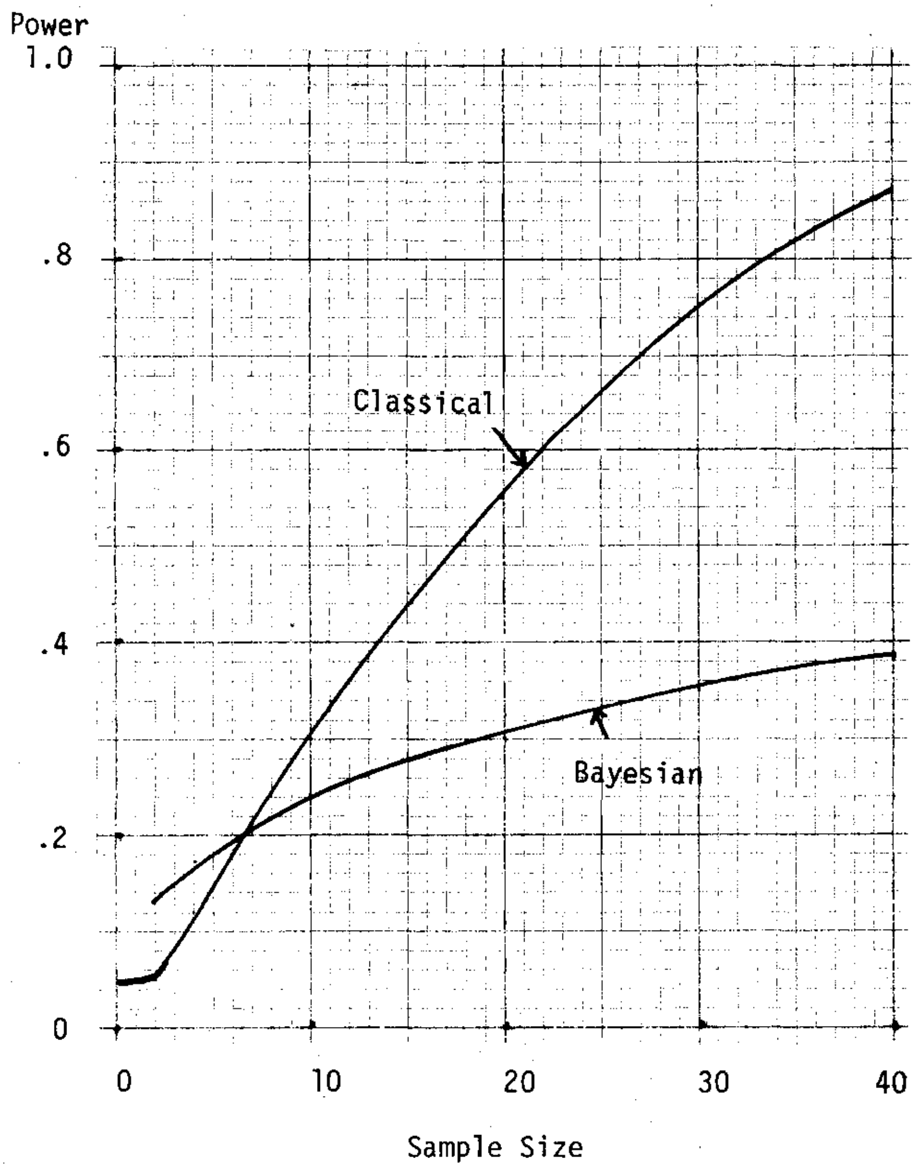

Figure 4. Power vs. Sample Size, $|c| = 20$, $\alpha = .05$.

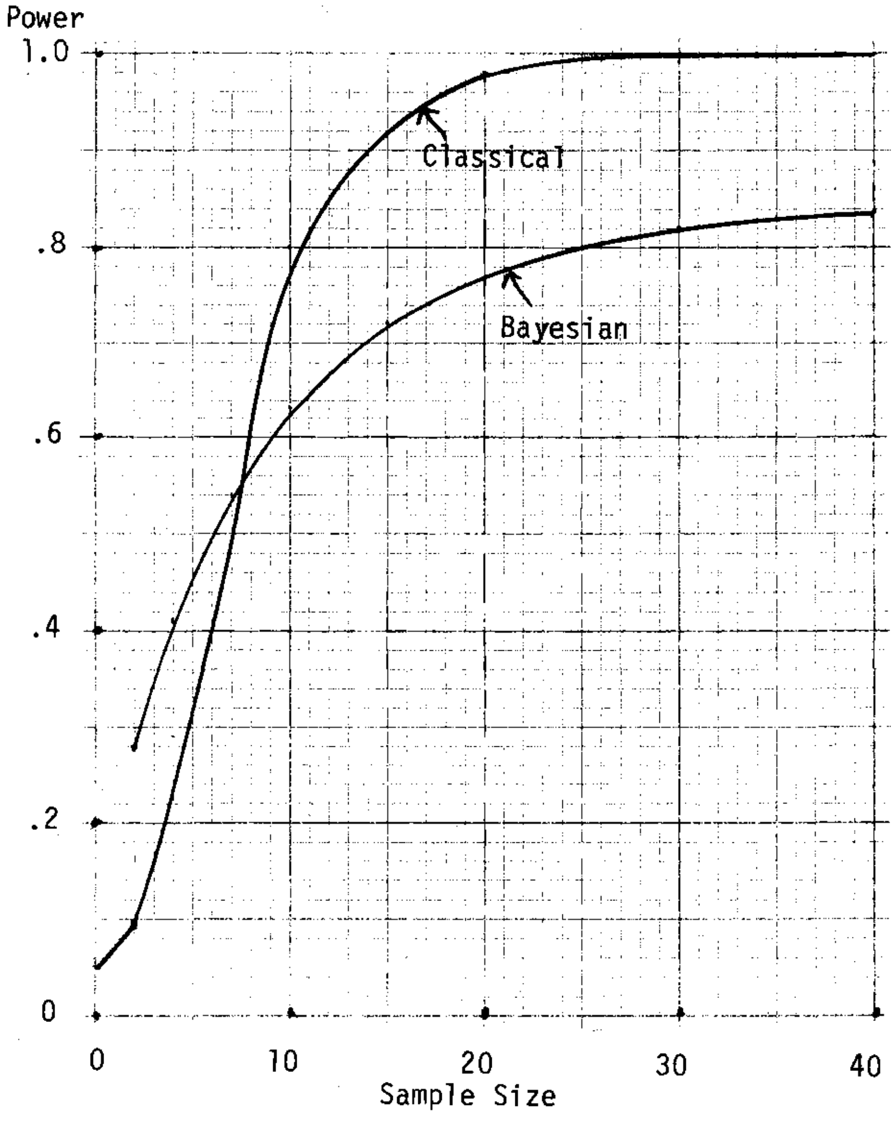

Figure 5. Power vs. Sample Size, $|c| = 40$, $\alpha = .05$.

Plots of the power versus sample size for the two tests for $|c| = 20$, 30, and 40 are given in Figures 6, 7, and 8, respectively. There it can be seen that as $|c|$ increases, the difference between the two curves decreases. However, as seen in Table 2, the type I error for the Bayesian test is greater than that for the classical test (.05) for sample sizes greater than four. Once again, the classical test appears to be superior, particularly when high values of the power are required.

In the foregoing example, a 95% BPI was utilized in computing the power for the Bayesian test. If a larger interval is used, both the power and the type I error will decrease. This is obvious from equations (3-3) and (3-2). If the length of the BPI is increased, the probability that the BPI will include zero must increase. Therefore, the probability that the BPI will not include zero (or power) must decrease. Similarly, decreasing the length of the BPI will increase the power and the type I error. Thus, various power and type I error combinations can be achieved by varying the width of the BPI.

In the above example, the variability of $\tilde{m}$ was affected by n_u, as defined in equation (3-16). In Table 3 below, the difference between n and n_u can be seen to increase as n increases. The parameter, n_u, takes into effect the variability of $\tilde{m}'$ in calculating the variability of m, which is given by [29]

$$V(\tilde{m}|m',v',n_u,\nu') = \frac{v'}{n_u}\frac{\nu'}{\nu'-2} , \qquad (3\text{-}22)$$

where $n_u = \frac{nn'}{n+n'}$.

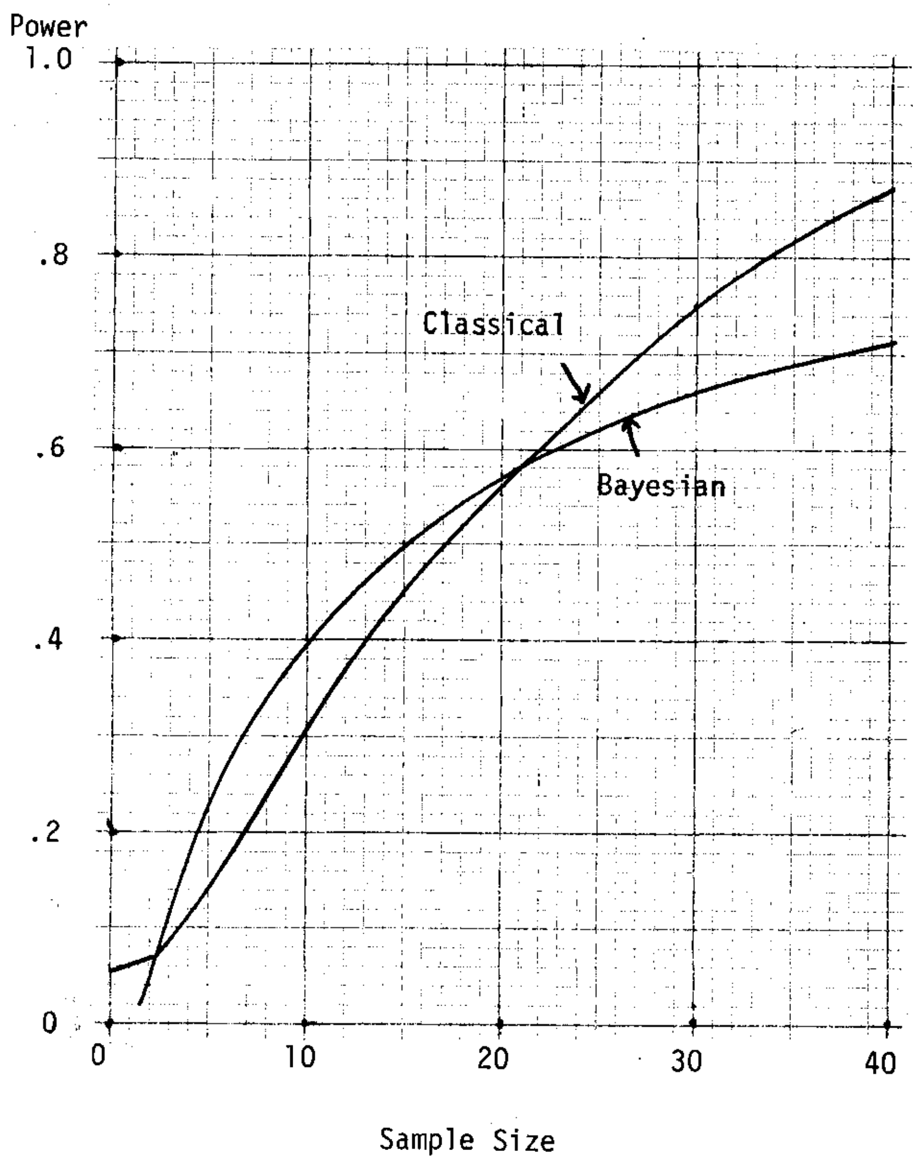

Figure 6. Power vs. Sample Size, $|c| = 20$, $\gamma = .05$.

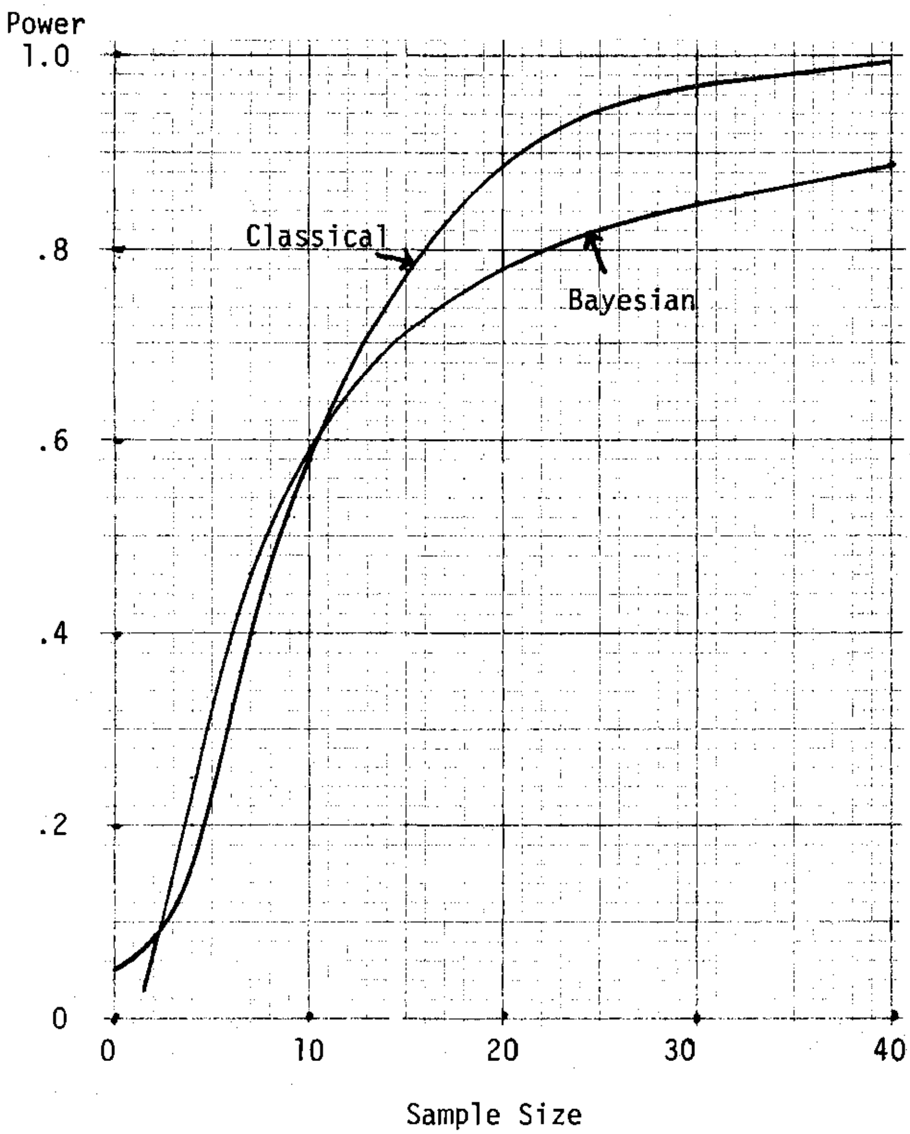

Figure 7. Power vs. Sample Size, $|c| = 30$, $\alpha = .05$.

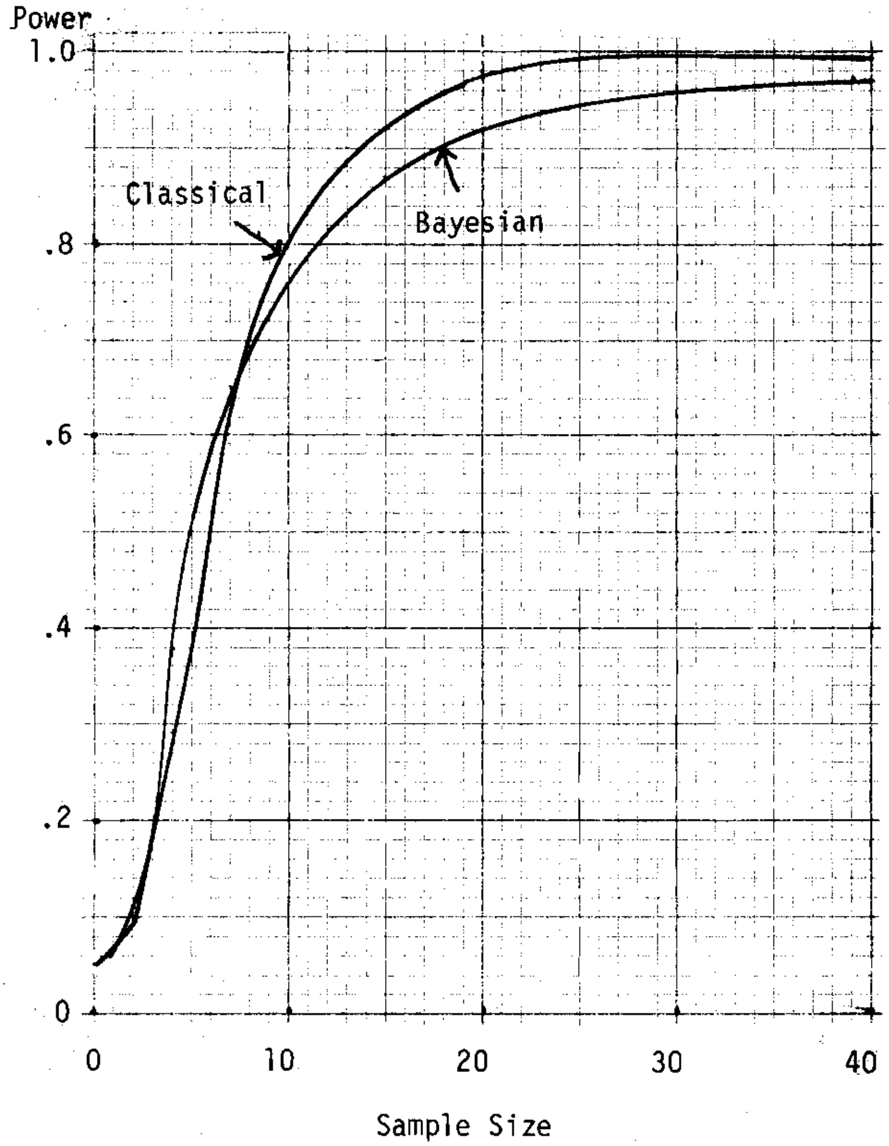

Figure 8. Power vs. Sample Size, $|c| = 40$, $\gamma = .05$.

Table 3. Sample Size Versus n_u

Sample Size	n_u
2	1.75
4	3.11
7	4.67
10	5.83
20	8.24
30	9.55
40	10.37

Since we are considering the true value of $\bar{\mu}$ to be a constant, which is the expected value of $\tilde{m}$, we shall next investigate how the power of the Bayesian test is affected if the variability of $\tilde{m}'$ is not considered in the variance of $\tilde{m}$; i.e., n_u will be replaced by n in equation (3-18).

Solution Using Alternate Method

Replacing n_u with n in equation (3-18) yields

$$\text{Power} = 1 - F_{S^*}\left(\frac{kn'' - m'n' - cn}{n\sqrt{v'/n}}\middle|\nu'\right) + F_{S^*}\left(\frac{-kn'' - m'n' - cn}{n\sqrt{v'/n}}\middle|\nu'\right) \quad (3\text{-}23)$$

The type I error for this test is obtained from equation (3-23) with $c = 0$.

$$\alpha = 1 - F_{S^*}\left(\frac{kn'' - m'n'}{n\sqrt{v'/n}}\middle|\nu'\right) + F_{S^*}\left(\frac{-kn'' - m'n'}{n\sqrt{v'/n}}\middle|\nu'\right) \quad (3\text{-}24)$$

Illustrating the Procedure

Using the same sample data as in the previous sections, the values of k required to keep α = .05 are obtained from equation (3-21) with n_u = n and are shown below.

Table 4. Sample Size Versus k, α = .05

Sample Size	k
2	22.59
4	22.72
7	21.98
10	20.98
15	19.36
20	17.94
30	15.72
40	14.09

With the type I error fixed at .05, the power curves for n = 30 and n = 4 for $|c|$ = 20 are plotted in Figures 9 and 10, respectively. It can be seen that for n = 4 the Bayesian test is more powerful and for n = 30, the classical test is marginally more powerful. The plots of power versus sample size for $|c|$ = 20 and $|c|$ = 40 are given in Figures 11 and 12, respectively. From these curves, it can be seen that there is little difference between the two tests in terms of power. Thus, when the variability of $\tilde{m}'$ is not considered in the variability of $\tilde{m}$, there is no significant difference in the power of the two tests. There has been no evidence so far to justify using Bayesian instead of classical

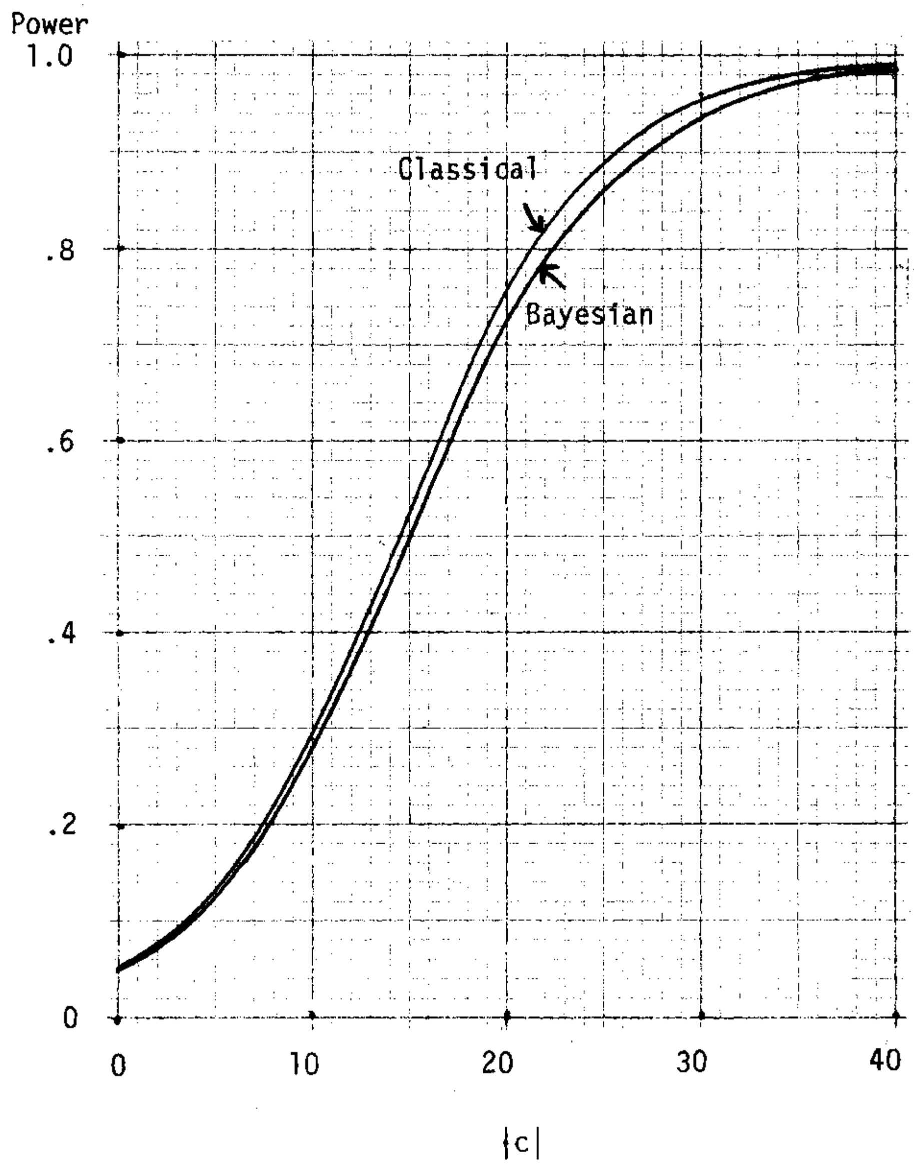

Figure 9. Power Curves, n = 30, α = .05.

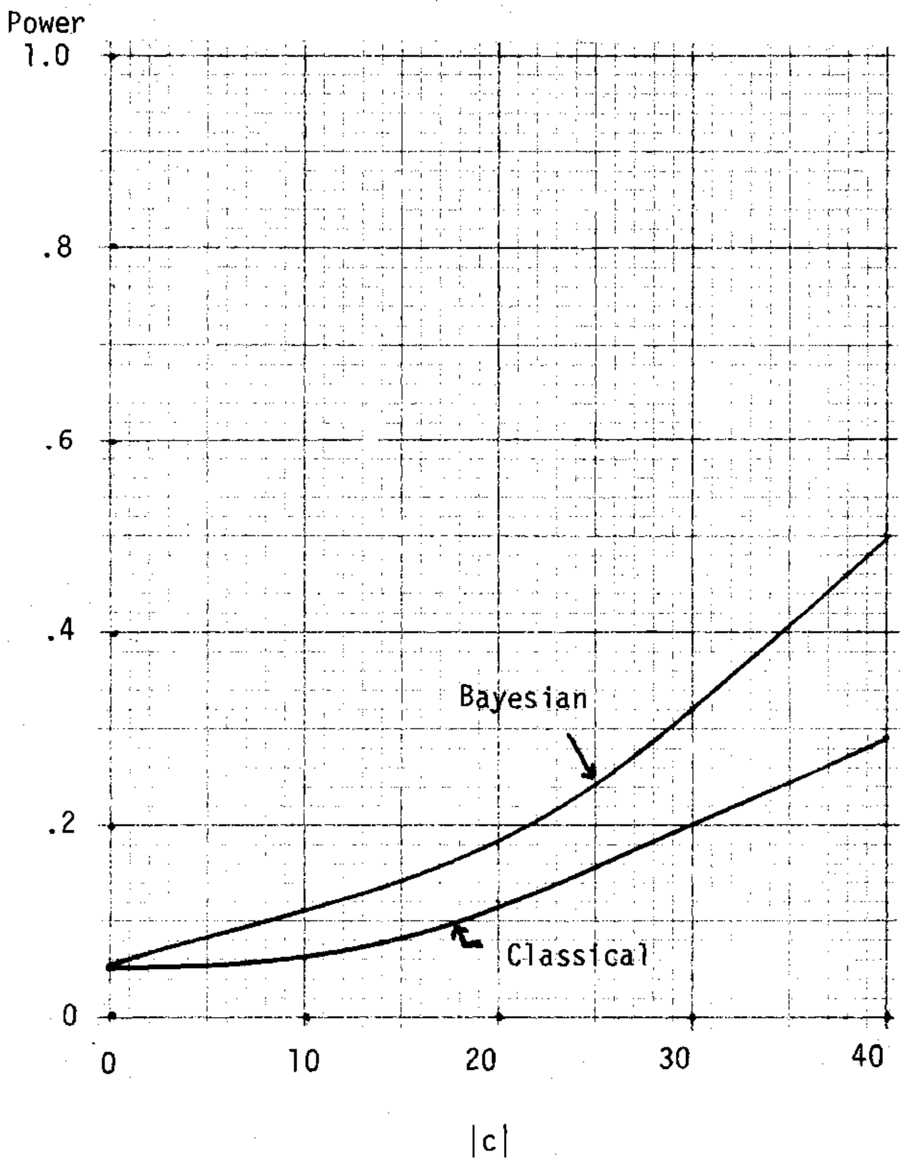

Figure 10. Power Curves, $n = 4$, $\alpha = .05$.

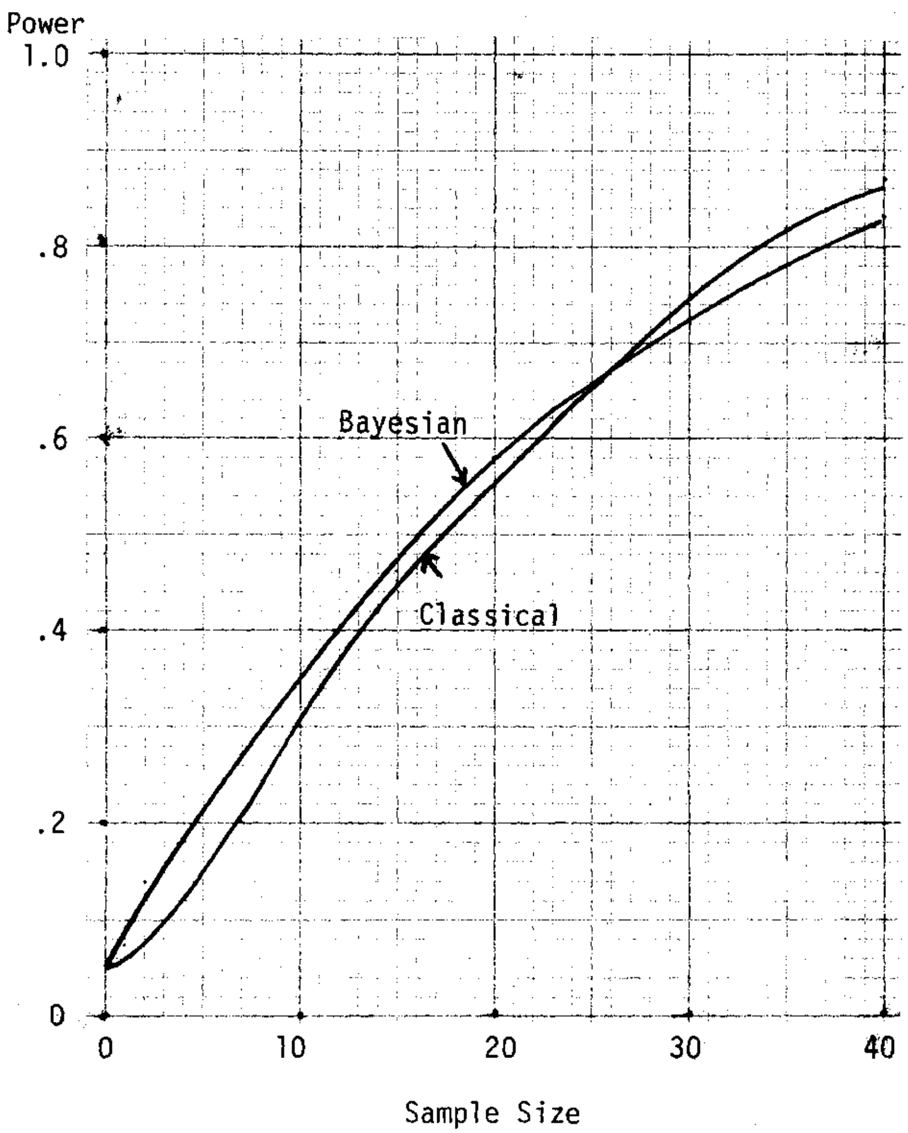

Figure 11. Power vs. Sample Size, $|c| = 20$, $\alpha = .05$.

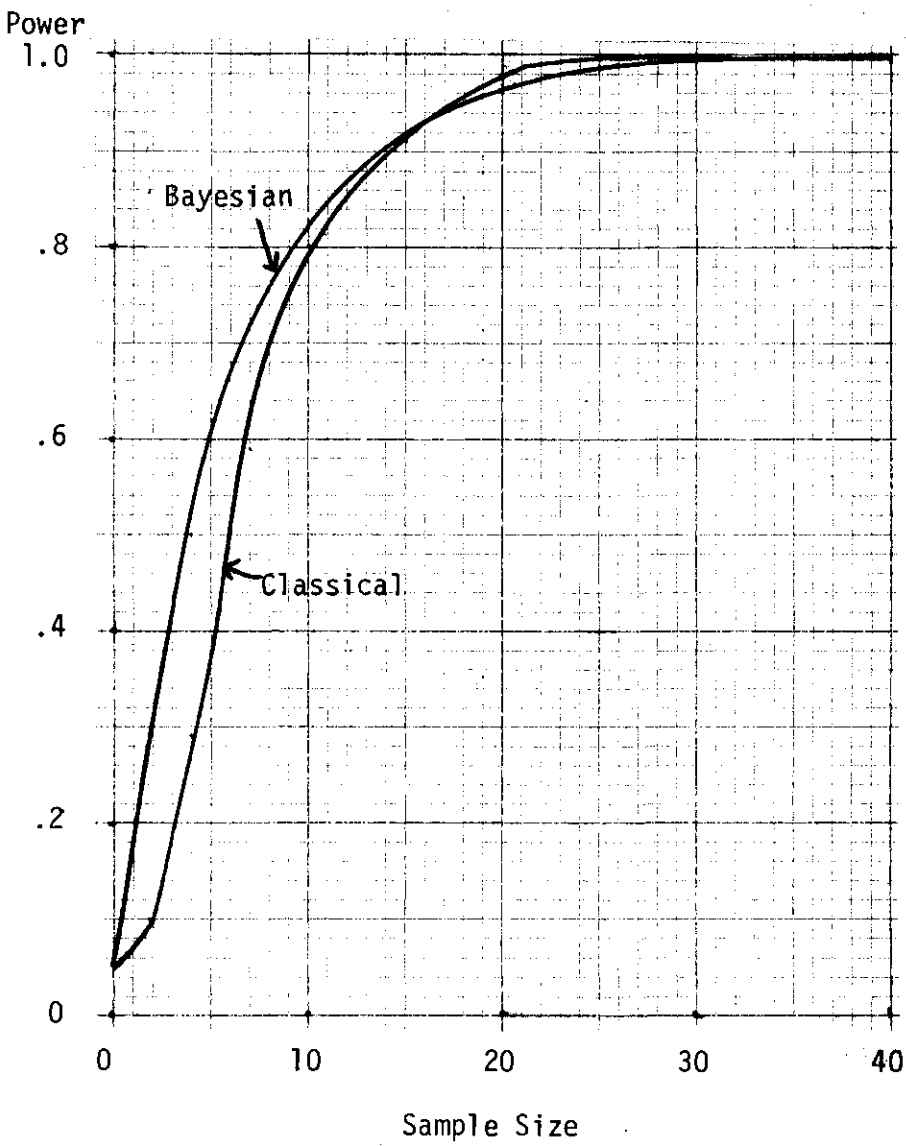

Figure 12. Power vs. Sample Size, $|c| = 40$, $\alpha = .05$.

procedures in the case of the two-tailed hypothesis test. Another method of treating the two-tailed test will be discussed in connection with the one-tailed test in the next section.

The One-Tailed Hypothesis Test

If the decision maker is interested in the classical one-tailed test, H_0: $\mu \leq 0$ vs H_1: $\mu > 0$, there is an equivalent Bayesian test; namely, H_0: $\tilde{\mu} \leq 0$ vs H_1: $\tilde{\mu} > 0$. In fact, an alternate method for testing the two-tailed hypothesis also falls into this category. Rather than test H_0: $\tilde{\mu} = 0$ vs H_1: $\tilde{\mu} \neq 0$, consider H_0: $-a \leq \tilde{\mu} \leq a$ vs H_1: $\tilde{\mu} < -a$ or $\tilde{\mu} > a$, $a > 0$. This really tests whether $\tilde{\mu}$ is in some interval about zero and can be treated as a special case of the one-tailed test discussed below.

As in the two-tailed test, the type I and type II errors for the classical one-tailed test can be determined for any distribution of the random variable of interest. However, in the Bayesian test once a posterior distribution for $\tilde{\mu}$ has been determined, the probabilities of H_0 and H_1 being true can be determined; i.e.,

$$P(\tilde{\mu} \leq 0|\text{sample data}) = \int_{-\infty}^{0} f(\tilde{\mu})d\tilde{\mu} .$$

If the density function of $\tilde{\mu}$ is known, the above integral can be computed. Additionally,

$$P(\tilde{\mu} > 0|\text{sample data}) = 1 - P(\tilde{\mu} \leq 0|\text{sample data}).$$

Equivalently, [36]

$$P(H_1 \text{ is true}) = 1 - P(H_0 \text{ is true}).$$

If one considers

$$\alpha = P(\text{rejecting } H_0 | H_0 \text{ is true}),$$

and one knows the probability that H_0 is true, it is difficult to justify any rejection criteria for H_0 which would lead to a meaningful calculation of α. Winkler [36] suggests that the significance level of the test can be determined by measuring how "unusual" the sample result obtained is, given that the null hypothesis is true. Equivalently, one could determine the chance of obtaining a sample result more "extreme" than the one observed, given H_0 is true. In the test considered in the previous section, for example, if $\tilde{\mu} = 0$, how "unusual" is the sample result of m = 72.95 sec? (See data in Appendix 3.) The standardized value corresponding to m = 72.95 is

$$t_0 = \frac{m-0}{s/\sqrt{n}} = \frac{72.95}{38.96/\sqrt{30}} = 10.26 \ .$$

Since H_1 is one-tailed to the right, the significance level is equal to the P $(t_0 \geq 10.26)$, which is less than .0001. The smaller the significance level, the less likely the sample result is, given that H_0 is true [36]. It can be seen, then, that the significance level as defined above cannot be fixed as in the classical test since it depends on the sample result. Additionally, there is no clear method for determining a power for the Bayesian test. As stated earlier in this section, the modified two-tailed test can be considered a special case of the one-tailed test. If the hypotheses of interest are H_0: $-a \leq \tilde{\mu} \leq a$ and H_1: $\tilde{\mu} < -a$ or $\tilde{\mu} > a$, $a > 0$, then the probability that H_0 is true is

$$P(H_0 \text{ is true}|\text{sample data}) = \int_{-a}^{a} f(\tilde{\mu})d\tilde{\mu}\ .$$

When the posterior distribution of $\tilde{\mu}$ is determined, the above integral can be computed. Obviously,

$$P(H_1 \text{ is true}|\text{sample data}) = 1 - P(H_0 \text{ is true}|\text{sample data}).$$

The arguments given for determining the significance level and power for the one-tailed test apply as well for the modified two-tailed test.

Since there is no meaningful definition of power available for the Bayesian one-tailed test, it is necessary to determine a different measure of comparison between the classical and Bayesian statistical procedures. The concept of minimum loss will, therefore, be considered in Chapter IV as the basis for comparison.

CHAPTER IV

CLASSICAL VS. BAYESIAN ANALYSIS WITH LINEAR LOSS FUNCTIONS

Introduction

In this chapter, a linear loss function will be utilized to compare the consequences of the decisions made under Bayesian and classical analyses of the same problem. In all real world problems, there are certain payoffs or losses associated with decisions made under uncertainty. When the decision maker is not sure of the value of a certain quantity, such as μ in the analysis in the last chapter, he is subject to making a decision which is based on the assumption of the wrong value of μ. For example, if the null hypothesis, $H_0: \mu \leq 0$, were accepted, causing the decision maker to reject the new equipment, when in fact the true μ is greater than 0, a certain "opportunity" loss is experienced. The army would be penalized, in that it would not have the opportunity to use a better piece of equipment. Even though it is not always possible to attach a monetary figure to the opportunity loss, some type of loss function must be considered by the classical decision maker, at least subjectively. When the decision maker determines maximum acceptable levels for the type I and type II errors for a test, he is indicating the relative importance of each type of error. For example, if .05 and .10 are the maximum levels for the type I and type II errors, respectively, the decision maker could be indicating that he considers the loss associated with a type I error to be twice as great as the loss associated with a type II error. In the classical analysis of a problem, however, a decision is based on the outcome of a hypothesis test on some central MOE, rather than on the possible losses resulting from each possible decision. Many times the type I error is arbitrarily set as some low value, say .05 or .01, and the power of the test is made as high as necessary by increasing the sample size. However, in considering actual loss functions formally, the decisions resulting from the classical and Bayesian approaches to the problem may differ considerably. The linear loss function will be considered in this chapter.

Linear Payoff Function

Before considering the linear loss function, a brief discussion of the linear payoff function is needed. In considering the two action problem of concern in this study, let a_1 denote the action of rejecting the new equipment in favor of the old, and let a_2 denote the action of purchasing the new equipment. Define linear payoff functions as in [36], say

$$R(a_1, \mu) = r_1 + s_1\mu \qquad (4\text{-}1)$$

$$R(a_2, \mu) = r_2 + s_2\mu$$

where r_i and s_i are constants and $s_2 > s_1$.

With these functions, the decision maker would consider the payoff of a certain action linear with respect to the actual state of the world, μ. In this case action a_1 would be optimal if

$$E[R(a_1)] > E[R(a_2)] \tag{4-2}$$

$$E(r_1 + s_1\mu) > E(r_2 + s_2\mu)$$

$$r_1 + s_1E(\tilde{\mu}) > r_2 + s_2E(\tilde{\mu}).$$

Subtracting r_2 and $s_1E(\tilde{\mu})$ from both sides we get

$$r_1 - r_2 > E(\tilde{\mu})(s_2 - s_1).$$

Since $s_2 > s_1$, dividing by $s_2 - s_1$ gives

$$\frac{r_1 - r_2}{s_2 - s_1} > E(\tilde{\mu}). \tag{4-3}$$

Therefore, if equation (4-3) is satisfied, action a_1 is optimal. If the inequality is reversed, action a_2 is optimal. For this decision making problem, μ_b is called the breakeven value of $\tilde{\mu}$:

$$\mu_b = \frac{r_1 - r_2}{s_2 - s_1} \tag{4-4}$$

Figure 13 displays μ_b pictorially.

If the expected value of $\tilde{\mu}$ is less than μ_b, action a_1 is optimal; if it is greater than μ_b, action a_2 is optimal; if it is equal to μ_b, the payoffs are equal, and the decision maker should be indifferent toward each action.

The Linear Loss Function

If action a_1 is chosen and the true value of μ is really greater than μ_b, then an opportunity loss has been suffered by not having chosen

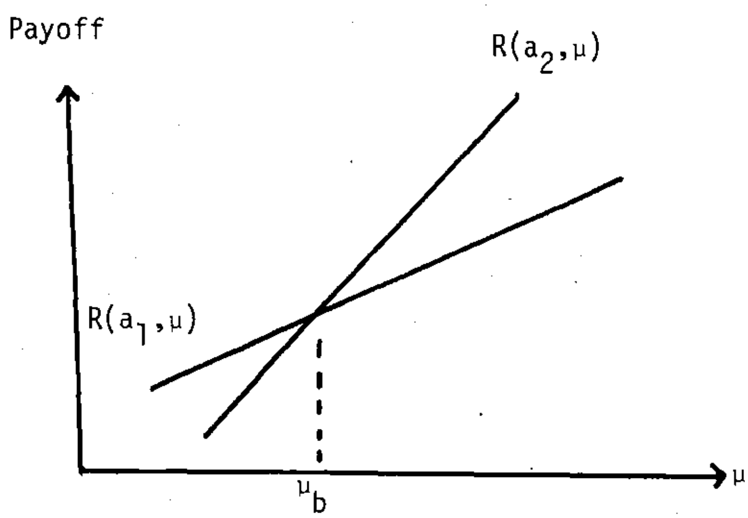

Figure 13. Payoff vs μ.

a_2 and is given by

$$L(a_1,\mu) = R(a_2,\mu) - R(a_1,\mu) \tag{4-5}$$

$$= r_2 + s_2\mu - (r_1 + s_1\mu)$$

$$= (r_2 - r_1) + (s_2 - s_1)\mu. \tag{4-6}$$

On the other hand, if action a_2 were chosen and the true value of μ is less than μ_b, then the opportunity loss is

$$L(a_2,\mu) = R(a_1,\mu) - R(a_2,\mu) \tag{4-7}$$

$$= r_1 + s_1\mu - (r_2 + s_2\mu)$$

$$= (r_1 - r_2) + (s_1 - s_2)\mu \tag{4-8}$$

If a_1 were chosen and the true value of μ is less than μ_b, then the opportunity loss would be 0. If a_2 were chosen and the true value of μ is greater than μ_b, the opportunity loss is also 0. The loss functions for a_1 and a_2 are summarized below:

$$L(a_1,\mu) = \begin{cases} 0 & \text{if } \mu < \mu_b \\ (r_2-r_1)+(s_2-s_1)\mu & \text{if } \mu \geq \mu_b \end{cases} \tag{4-9}$$

$$L(a_2,\mu) = \begin{cases} (r_1-r_2)+(s_1-s_2)\mu & \text{if } \mu < \mu_b \\ 0 & \text{if } \mu \geq \mu_b \end{cases} \tag{4-10}$$

The relationship between the payoff and loss functions is shown in Figure 14.

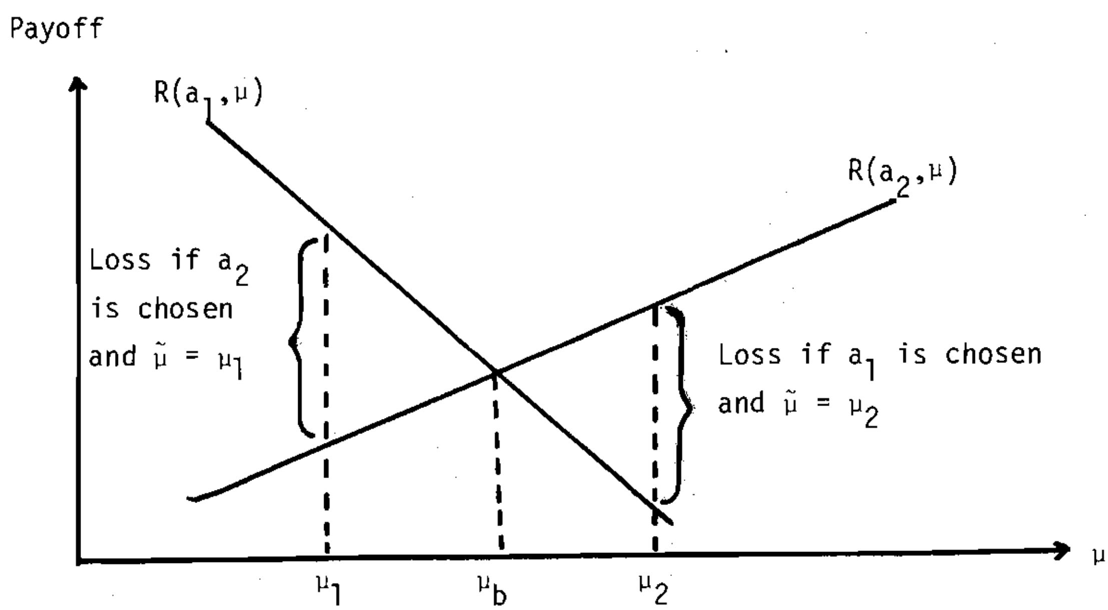

Figure 14. Payoff and Loss vs. μ.

The loss functions, $L(a_1,\mu)$ and $L(a_2,\mu)$ are shown in Figures 15 and 16. It is obvious from these figures that the loss functions are related to the value of the breakeven point as given in equation (4-4). If $\mu > \mu_b$,

$$L(a_1,\mu) = (r_2-r_1)+(s_2-s_1)\mu = (s_2-s_1)\frac{(r_2-r_1)}{s_2-s_1} + (s_2-s_1)\mu$$

$$= (s_2-s_1)(-\mu_b) + (s_2-s_1)\mu$$

$$= (s_2-s_1)(\mu - \mu_b) .$$

Similarly, for $\mu < \mu_b$

$$L(a_2,\mu) = (s_2-s_1)(\mu_b-\mu) .$$

Therefore, the loss functions can now be written

$$L(a_1,\mu) = \begin{cases} 0 & \mu \le \mu_b \\ (s_2-s_1)(\mu-\mu_b) & \mu \ge \mu_b \end{cases} \tag{4-11}$$

$$L(a_2,\mu) = \begin{cases} (s_2-s_1)(\mu_b-\mu) & \mu \le \mu_b \\ 0 & \mu \ge \mu_b \end{cases} . \tag{4-12}$$

The expected value of each loss function depends on the distribution of $\tilde{\mu}$ and is given by [36]

$$EL(a_1) = (s_2-s_1) \int_{\mu_b}^{\infty} (\mu-\mu_b)f(\mu)d\mu \tag{4-13}$$

$$EL(a_2) = (s_2-s_1) \int_{-\infty}^{\mu_b} (\mu_b-\mu)f(\mu)d\mu \tag{4-14}$$

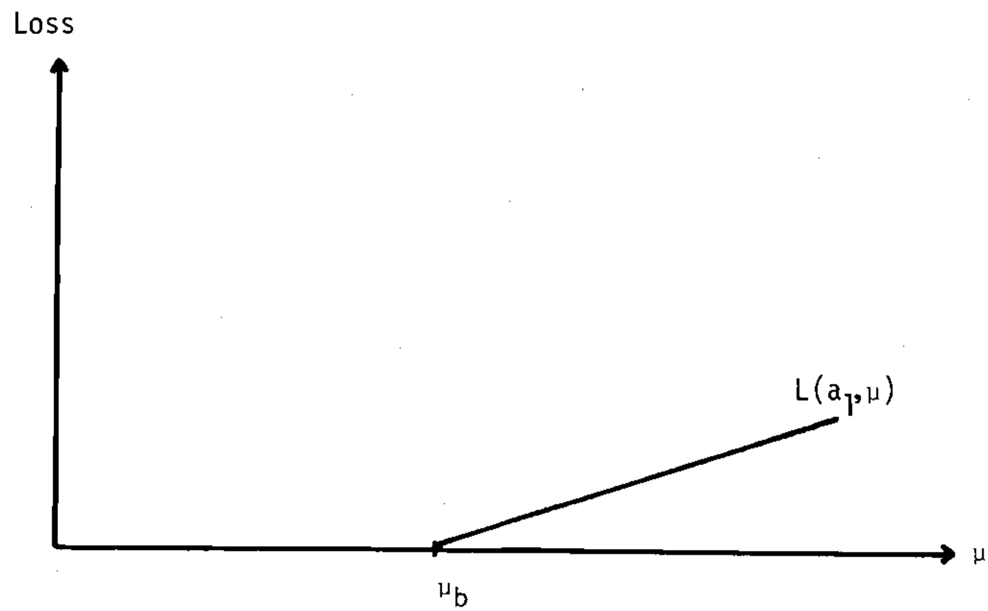

Figure 15. Loss vs. μ for Action a_1.

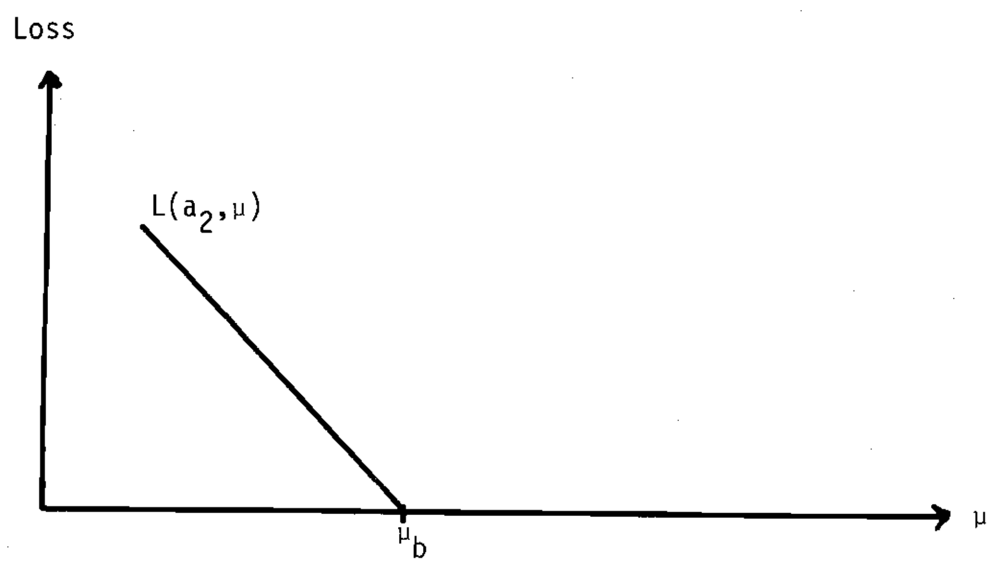

Figure 16. Loss vs. μ for Action a_2.

The integrals in equations (4-13) and (4-14) are called right hand and left hand linear loss integrals, respectively. Formulas for tabulation of the above integrals are given in [36] for various conjugate distributions. The loss functions given in equations (4-11) and (4-12) are valid for both the classical and Bayesian analyses. The difference in the two approaches arises from the differing decision criteria in each analysis.

Comparison of Decisions

In the classical analysis, action a_1 (reject new equipment) is taken if the null hypothesis, H_o: $\mu \leq 0$, is accepted, while action a_2 (purchase new equipment) is taken if the null hypothesis is rejected. No formal consideration is given to the value of μ_b or to the loss function. In the Bayesian analysis, however, action a_1 is taken if the expected loss due to a_1 is less than the expected loss due to a_2, and a_2 is taken if the expected loss due to a_2 is less than that due to a_1; i.e., expected loss is minimized [36]. Consider Figure 17, where two typical linear loss functions are graphed.

In the case where the classical analyst accepted the null hypothesis, resulting in action a_1, if $\mu_b < 0$, then the loss given by $L(a_1\mu)$ would still be incurred for values of μ between 0 and μ_b, <u>even though H_o is true</u>. If H_o were actually false, and the true μ is greater than 0 (i.e. a type II error), then the losses are even greater. If, however, $\mu_b > 0$, then a loss is incurred by choosing a_1 only if the true μ is greater than μ_b. This would also be a type II error. Thus, the classical analyst may incur a loss $L(a_1,\mu)$ by accepting H_o, if he has made a type

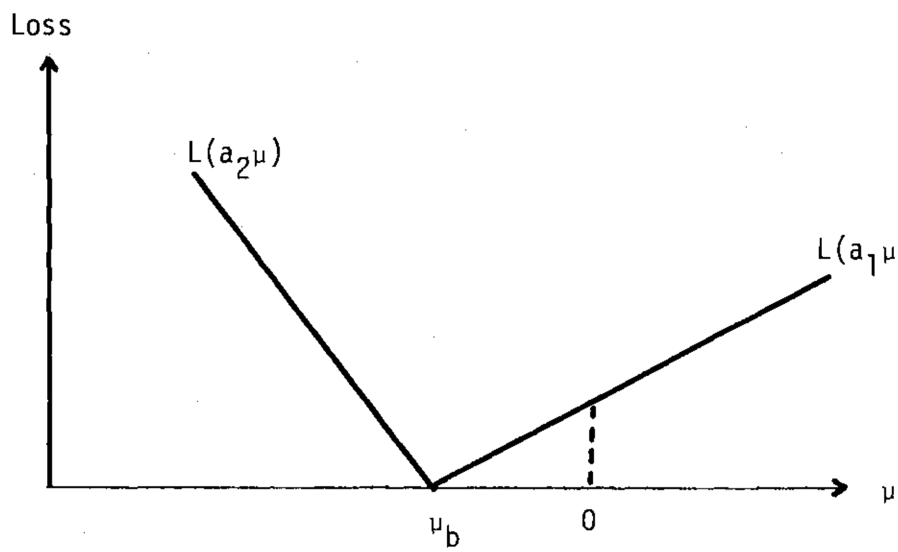

Figure 17. Loss vs. μ.

II error or no error at all, in terms of hypothesis testing.

Similarly, if the classical analyst rejected the null hypothesis, he would choose action a_2. If $\mu_b < 0$, the loss given by $L(a_2,\mu)$ would be incurred if the true value of μ is less than μ_b (a type I error). If $\mu_b < 0$ then a loss given by $L(a_2,\mu)$ is incurred for values of μ between 0 and μ_b even though he correctly rejected H_0. Thus, a loss given by $L(a_2,\mu)$ may be incurred by making a type I error or no error at all.

The above discussion points out that by not considering the break-even point or loss function in his analysis, the classical analyst is very likely to incur higher losses, even when he chooses the hypothesis which is true, than the Bayesian who chooses the action with the least expected loss.

The LWCMS OT II problem will again be used to demonstrate the above procedures.

Illustrating the Procedure

Consider the payoff functions given by

$$R(a_1,\mu) = -100 - 20\ \mu$$

$$R(a_2,\mu) = -250 + 10\ \mu$$

A reasonable explanation of such payoff functions could be as follows. Action a_1 corresponds to rejecting the new equipment. If testing the equipment costs 100 units and the decision maker considers a penalty cost of 20 units for each unit of μ above 0, he would be expressing the importance he attaches to the actual mean difference, μ, between the MOE of the competing systems. As μ becomes more positive, the new piece of equipment becomes much better than the old and the more costly (negative payoff) becomes the decision of having chosen action a_1.

Action a_2 indicates that the new system has been chosen. The cost of sampling plus purchase is equal to 250 units, and the decision maker attaches a payoff of 10 units per unit of μ.

Using equation (4-4),

$$\mu_b = \frac{r_1-r_2}{s_2-s_1} = \frac{-100-(-250)}{10-(-20)} = \frac{150}{30} = 5 \qquad (4\text{-}15)$$

From equations (4-11) and (4-12)

$$L(a_1,\mu) = \begin{cases} 0 & \mu \leq 5 \\ 30\ (\mu-5) & \mu \geq 5 \end{cases} \qquad (4\text{-}16)$$

$$L(a_2,\mu) = \begin{cases} 30(5-\mu) & \mu \leq 5 \\ 0 & \mu \geq 5 \end{cases} \tag{4-17}$$

From equations (4-13) and (4-14)

$$EL(a_1) = 30 \int_5^\infty (\mu-5)f(\mu)d\mu \tag{4-18}$$

$$EL(a_2) = 30 \int_{-\infty}^5 (5-\mu)f(\mu)d\mu \tag{4-19}$$

It has been shown [29] that if $\tilde{\mu}$ follows the student density, as it does in this example, then

$$\int_\mu^\infty (z-\mu)f_S(z|m,n/v,\nu)dz = L_{S^*}(t|\nu)\sqrt{v/n} \tag{4-20}$$

and

$$\int_{-\infty}^\mu (\mu-z)f_S(z|m,n/v,\nu)dz = L_{S^*}(-t|\nu)\sqrt{v/n}\ , \tag{4-21}$$

where

$$t = (\mu-m)\sqrt{n/v}$$

$$L_{S^*}(t|\nu) \equiv \frac{\nu + t^2}{\nu - 1} f_{S^*}(t|\nu) - tG_{S^*}(t|\nu)$$

$$G_{S^*}(t|\nu) = 1 - F_{S^*}(t|\nu)\ .$$

Values of $f_{S^*}(t|\nu)$ are given in [29] Table I.

The expected losses given in equations (4-18) and (4-19) could be computed from either the prior or posterior distributions for $\tilde{\mu}$. Since the decision will be made in the classical case after the sample has been taken, the posterior distribution will be used.

As given in Chapter III the prior distribution of $\tilde{\mu}$ before testing in OT II has parameters

$$(m',v',n',\nu') = (17.6,\ 2040.5,\ 14,\ 13).$$

The sample data given in Appendix 3 for OT II produced the statistic $(m,v,n,\nu) = (72.95,\ 1517.9,\ 30,\ 29)$. Thus the parameters of the posterior distribution of $\tilde{\mu}$, as given by equation (2-7), are

$$m'' = \frac{n'm' + nm}{n + n'} = \frac{(14)(17.6) + (30)(72.95)}{30 + 14} = 55.34$$

$$n'' = n + n' = 30 + 14 = 44$$

$$v'' = \frac{[\nu'v' + n'(m')^2] + (\nu v + nm^2) - n''(m'')^2}{[\nu v' + \delta(n')] + [\nu + \delta(n)] - \delta(n'')}$$

$$= \frac{(13)(2040.5)+(14)(17.6)^2 + (29)(1517.9)+(30)(72.95)^2-(44)(55.34)^2}{13 + 1 + 29 + 1 - 1}$$

$$= 2320.5$$

$$\nu'' = [\nu' + \delta(n')] + [\nu + \delta(n)] - \delta(n'')$$

$$= 13 + 1 + 29 + 1 - 1 = 43$$

where $\delta(\chi) = \begin{cases} 0 & x = 0 \\ 1 & x > 0 \end{cases}$

Thus $(m'',v'',n'',\nu'') = (55.34,\ 2320.5,\ 44,\ 43)$.

To evaluate the linear loss integrals in equations (4-18) and (4-19),

$$t = (\mu_b - m'')\sqrt{n''/v''} = (5 - 55.34)\sqrt{44/2320.5} = -6.9$$

$$G_{S^*}(t|\nu'') = 1 - F_{S^*}(t|\nu'') = 1 - F_{S^*}(-6.9|43) = 1$$

$$G_{S^*}(-t|\nu'') = 1 - F_{S^*}(-t|\nu'') = 1 - F_{S^*}(6.9|43) = 0$$

$$L_{S^*}(t|\nu'') = \frac{\nu''+t^2}{\nu''-1} f_{S^*}(t|\nu'') - t\, G_{S^*}(t|\nu'')$$

$$L_{S^*}(-6.9|43) = \frac{43+(-6.9)^2}{43-1} f_{S^*}(-6.9|43) - (6.9)G_{S^*}(-6.9|43)$$

$$= (2.16)(-.00000015) + (6.9)1$$

$$= 6.9 \tag{4-22}$$

$$L_{S^*}(-t|\nu'') = L_{S^*}(6.9|43)$$

$$= \frac{43 + (6.9)^2}{42} f_{S^*}(6.9|43) - 6.9\, G_{S^*}(6.9|43)$$

$$= (2.16)(.00000015) - 6.9(0)$$

$$L_{S^*}(6.9|43) \simeq 0 \tag{4-23}$$

Using the L_{S^*} calculated in (4-22) and (4-23)

$$\int_5^\infty (\mu - 5)f(\mu)d\mu = L_{S^*}(-6.9|43)\sqrt{v''n''}$$

$$= 6.9\ \sqrt{2320.5/44}$$

$$= 50.1$$

$$\int_{-\infty}^{5} (5 - \mu)f(\mu)d\mu = L_{S^*}(6.9|43)\sqrt{v''/n''}$$

$$= 0$$

Thus, from equations (4-18) and (4-19)

$$EL(a_1) = 30(50.1) = 1500.3$$

$$EL(a_2) = 30(0) = 0$$

The Bayesian would, therefore, choose action a_2 and buy the new equipment.

In the classical analysis, using H_0: $\mu \leq 0$ vs H_1: $\mu > 0$, the statistic $t_0 = \frac{\overline{X} - 0}{s/\sqrt{n}}$ would be computed and the null hypothesis would be rejected if $t_0 > t_{\alpha,n-1}$ [12]. In this example,

$$t_0 = \frac{72.95 - 0}{38.96/\sqrt{30}} = 10.26$$

$$t_{.05,29} = 1.697.$$

Therefore, the classical analyst would reject the null hypothesis and also choose action a_2. Since the data for this particular problem has a mean so much greater than 0, one should expect both methods to reach the same decision. A better comparison would result from a sample with a mean closer to zero. Consider the case where the sample results in a mean of $\overline{X} = 10$, with the same sample variance. Now, the classical analyst would not reject H_0 since $t_0 = \frac{\overline{X} - 0}{s/\sqrt{n}} = \frac{10}{38.96/\sqrt{30}} = 1.41$, which is less than $t_{.05,29} = 1.697$. The classical analyst would then choose action a_1 and reject the new equipment. On the other hand, the Bayesian would recompute $EL(a_1)$ and $EL(a_2)$. The new parameters of the posterior distribution of $\tilde{\mu}$ are

$$m'' = \frac{(14)(17.6) + (30)(10)}{30 + 14} = 12.42$$

$$n'' = 44$$

$$v'' = \frac{(13)(2040.5)+(14)(17.6)^2 + (29)(1517.9)+(30)(10)^2 - (44)(12.42)^2}{43}$$

$$= 1653.37$$

$$\nu'' = 43$$

The t value in equations (4-18) and (4-19) is

$$t = (5 - 12.42)\sqrt{44/1653.37} = -1.21$$

$$G_{S^*}(-1.21|43) = 1 - F_{S^*}(-1.21|43) = .8814$$

$$G_{S^*}(1.21|43) = 1 - F_{S^*}(1.21|43) = .1186$$

$$L_{S^*}(-1.21|43) = \frac{43 + (-1.21)^2}{42} f_{S^*}(-1.21|43) - (-1.21)G_{S^*}(-1.21|43)$$

$$= 1.059\ (-.19) + 1.21(.8814)$$

$$= .865$$

$$L_{S^*}(1.21|43) = \frac{43 + (1.21)^2}{42} f_{S^*}(1.21|43) - (1.21)G_{S^*}(1.21|43)$$

$$= (1.059)(.19) - 1.21(.1186)$$

$$= .058$$

$$EL(a_1) = 30 \int_5^\infty (\mu-5)f(\mu)d\mu$$

$$= 30\ L_{S^*}(-1.21|43)\sqrt{1653.37/43}$$

$$= 30(.865)(\sqrt{1653.37/43})$$

$$= 160.91$$

$$EL(a_2) = 30 \int_{-\infty}^{5} (5-\mu)f(\mu)d\mu$$

$$= 30\ L_{S^*}(1.21|43)\sqrt{1653.37/43}$$

$$= 30\ (.058)\sqrt{1653.37/43}$$

$$= 10.79$$

Since $EL(a_2) < EL(a_1)$, the Bayesian would choose action a_2 and buy the new equipment. In this example, the classical analyst chose the decision which had the higher expected loss. This resulted from considering only the true value of μ and not the effect of the value of μ on the loss which could be incurred from each decision.

Although this example considers only the linear loss function, the conclusions resulting from the example are valid for all loss functions. Since the decision maker is ultimately concerned with choosing the action which will minimize his losses (or maximize his payoffs), it is imperative for him to formally assess his loss or payoff function. Once this is done, he can base his decision on the action which has the least expected loss or greatest expected payoff, rather than on the true value of some statistic.

It can be seen from equations (4-18) through (4-21) that there is a relationship between the sample size and the expected loss from each action. The sample size affects both the degrees of freedon, ν, and the value of t, as well as the values of the integrals in equations (4-20) and (4-21). It is possible that a sample size could be determined which would minimize the expected loss of each action, but such a determination is beyond the scope of this study.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The conclusions of this study must be considered from two distinct viewpoints. The first is that of hypothesis testing. If the decision maker is interested purely in testing one hypothesis against another, such as H_0: $\mu = 0$ vs H_1: $\mu \neq 0$, there are several disadvantages to utilizing Bayesian statistical procedures.

The hypotheses of interest may not be meaningful from a Bayesian viewpoint, particularly for the two-tailed test. In fact, to utilize Bayesian statistical procedures, the decision maker must alter his conception of the mean and variance of a distribution of a random variable as discussed in Chapter I. With the Bayesian conception of a random variable in mind, the decision maker must formulate a new hypothesis to be tested which he feels will provide him with information equivalent to that which he would have obtained from the classical hypothesis test. An example of this was given in Chapter III with H_0: $\bar{\mu} = 0$ vs H_1: $\bar{\mu} \neq 0$. Once the alternate hypotheses have been formulated, they can be tested using Bayesian statistical procedures. However, it was shown in Chapter III that when the probability of a type I error was held constant, the Bayesian test was less powerful than the classical in the meaningful range of values for the power. When the BPI was kept constant, the Bayesian test was also less powerful than the classical test for large values of the power with the additional disadvantage that the probability of a type I error increased with the sample size. When the variability of the sample mean was assumed to be independent of the variability of the prior mean, it was shown that there is little difference between the two types of tests in terms of power.

In the case of the one-tailed test, there are nearly equivalent hypotheses which can be investigated with Bayesian and classical procedures; e.g., H_0: $\tilde{\mu} \leq 0$ vs. H_1: $\tilde{\mu} > 0$ and H_0: $\mu \leq 0$ vs H_1: $\mu > 0$, respectively. Although the probability of a type I error can be determined, it cannot be fixed in the Bayesian test. Also, the power of the Bayesian test cannot be meaningfully defined, as discussed in Chapter III. Therefore, the two types of procedures cannot be meaningfully compared for the one-tailed test.

The second viewpoint from which the conclusions must be considered is that of the decision criteria. If the decision maker can formally describe the loss function in relation to each of the possible decisions he may make, Bayesian statistical procedures have been developed which will enable him to make the decision which has the least expected loss. In Chapter IV an example was provided to demonstrate the procedures in the case of a linear loss function. Since the classical decision maker does not formally consider a loss function and bases his decision on the result of a hypothesis test, he may make a decision which would not minimize his expected loss. From this viewpoint, therefore, Bayesian statistical procedures are far superior to classical statistical procedures.

Therefore, if the decision maker is interested purely in testing one hypothesis against another, he should use classical statistical procedures. However, if he is interested in making a decision which has the least expected loss, he should use Bayesian statistical procedures.

Recommendations

In Chapter II it was stated that one of the objectives of the Bayesian methodology was to determine the minimum sample size from which meaningful probability statements could be made regarding $\tilde{\mu}$. In this study an attempt was made to determine the sample size which would produce a desired power. It is recommended that some other measure of a "meaningful probability statement" be investigated to reduce the sample size now being used by OTEA.

It is also recommended that the Bayesian methodology presented in Chapter IV be investigated to determine the effect of sample size on the decision to be made.

Finally, it is recommended that Bayesian statistical procedures be applied to a problem in which more than one MOE is under investigation since the procedures in this study apply to a situation in which only one MOE is being considered.

APPENDICES

APPENDIX I

EXPLANATION OF NOTATION

Chapter I

μ	mean of normal density function
σ^2	variance of normal density function
$\overline{X}$	sample mean
s^2	sample variance
$f'(\theta)$	prior distribution of $\tilde{\theta}$
$f(y\|\theta)$	likelihood function for $\tilde{y}$ given θ
$f''(\theta\|y)$	posterior distribution of $\tilde{\theta}$

Chapter II

$f_S(\mu\|m,n/v,\nu)$	density function for Student's t-distribution
m',v',n',ν'	prior parameters for Student's t-density function (these are interpreted on page 10)
m'',v'',n'',ν''	posterior parameters for Student's t-density function (these are defined mathematically on page 9)
m,v,n,ν	parameters of a normal sampling distribution (these are defined mathematically on page 11)
$F_{S^*}(\cdot\|\nu)$	left tail cumulative distribution function for standard Student's density function with ν degrees of freedom
$\overline{\mu}$	expected value of $\tilde{\mu}$
$\check{\mu}$	variance of $\tilde{\mu}$

$\check{\mu}'$	prior variance of $\tilde{\mu}$
$\sqrt{\check{\mu}'}$	prior standard deviation of $\tilde{\mu}$
$\bar{\mu}'$	prior mean of $\tilde{\mu}$
$\sqrt{\check{\mu}''}$	posterior standard deviation of $\tilde{\mu}$
s	ratio of expected posterior standard deviation of $\tilde{\mu}$ to prior standard deviation of $\tilde{\mu}$
$\check{\mu}''$	posterior variance of $\tilde{\mu}$
$\bar{\mu}''$	posterior mean value of $\tilde{\mu}$

Chapter III

n_u	$\frac{n'n}{n+n'}$
α	type I error
β	type II error
t_o	test statistic for classical hypothesis test
d''	length of a $(1-\gamma)$ Bayesian prediction interval on the posterior distribution of $\tilde{\mu}$

Chapter IV

$R(a_i,\mu)$	payoff function of the decision, a_i, and the true value of $\tilde{\mu}$, μ
μ_b	breakdown value of $\tilde{\mu}$
$L(a_i,\mu)$	loss function of the decision, a_i, and the true value of $\tilde{\mu}$, μ
$EL(a_i)$	expected loss if action a_i is chosen
$G_{S^*}(\cdot\|\nu)$	right tail cumulative distribution function for the standard Student's density function with ν degrees of freedom

$f_{S^*}(\cdot\,|\,\nu)$	standard Student density function with ν degrees of freedom
$L_{S^*}(\cdot\,|\,\nu)$	partial evaluation of linear loss integral for standardized Student density function with ν degrees of freedom

APPENDIX II

LIGHTWEIGHT COMPANY MORTAR SYSTEM OT I TEST DATA

Gunner's Examination Times [19]

	System		
Test Participant	81 mm (sec)	LWCMS (sec)	Difference in Performance
1	358.0	303.4	54.6
2	367.0	350.8	16.2
3	299.0	330.0	-31.0
4	261.0	147.5	113.5
5	380.0	313.0	67.0
6	226.8	250.0	-23.2
7	272.0	247.0	25.0
8	239.8	273.0	-33.2
9	235.0	258.0	-23.0
10	247.5	244.8	2.7
11	279.1	242.7	36.4
12	303.0	234.2	68.8
13	240.9	250.7	-9.8
14	279.0	296.9	-17.9

APPENDIX III

LIGHTWEIGHT COMPANY MORTAR SYSTEM OT II TEST DATA

Gunner's Examination Times [20]

Test Participant	Systems: 81 mm (sec)	Systems: LWCMS (sec)	Difference in Performance
1	321.5	225.5	96.0
2	310.0	194.5	115.5
3	314.0	248.0	66.0
4	293.0	272.5	20.5
5	304.5	259.0	45.5
6	256.0	173.0	83.0
7	321.5	224.0	97.5
8	397.5	256.0	141.5
9	297.5	282.0	15.5
10	254.5	220.0	34.5
11	258.0	262.0	-4.0
12	294.5	177.5	117.0
13	279.0	255.0	24.0
14	316.0	186.0	130.0
15	288.0	216.0	72.0
16	317.5	204.5	113.0
17	325.0	245.0	80.0
18	326.0	289.5	36.5
19	321.5	269.5	52.0
20	308.5	205.5	103.0
21	311.5	211.0	100.5
22	322.0	213.5	108.5
23	297.0	200.0	97.0
24	316.0	272.5	43.5
25	261.0	208.5	52.5
26	335.0	208.5	126.5
27	274.5	243.5	31.5
28	270.0	200.0	70.0
29	342.5	257.5	85.0
30	314.5	280.5	34.5

$$\text{Sample mean} = \frac{1}{30}\sum_{i=1}^{30} D_i = m = 72.95 \text{ sec}$$

$$\text{Sample variance} = \frac{1}{29}\sum_{i=1}^{30} (D_i - m)^2 = 1517.88 \text{ sec}^2$$

BIBLIOGRAPHY

1. AR 10-4, dated 30 December 1974, Organization and Functions of U.S. Army Operational Test and Evaluation Agency, Headquarters, Department of the Army, Washington, D.C.

2. AR 1000-1, dated 5 November 1974, Basic Policies for Systems Acquisition by the Department of the Army, Headquarters, Department of the Army, Washington, D.C.

3. Atchinson, J., "Bayesian Tolerance Regions," Journal of the Royal Statistical Society B, Vol. 26, 1964, pp. 161-175.

4. Baker, R. M., An Application of Bayesian Statistical Methods in the Determination of Sample Size for Operational Testing in the U. S. Army, Unpublished Master's Thesis, Georgia Institute of Technology, Atlanta, Georgia, 1976.

5. Birnbaum, A., "On the Foundations of Statistical Inference," Journal of the American Statistical Association, Vol. 57, 1962, pp. 269-306.

6. Box, G. E. P., and Tiao, G., "A Bayesian Approach to the Importance of Assumptions Applied to the Comparison of Variances," Biometrika, Vol. 51, 1964, pp. 153-167.

7. Box, G. E. P., and Tiao, G., "A Further Look at Robustness via Bayes Theorem," Biometrika, Vol. 49, 1962, pp. 419-432.

8. Cordova, R. L., An Application of Bayesian Analysis in Determining Appropriate Sample Sizes for Use in U. S. Army Operational Tests, Unpublished Master's Thesis, Georgia Institute of Technology, Atlanta, Georgia, 1975.

9. DeGroot, M. H., Optimal Statistical Decisions, McGraw-Hill Book Company, New York, 1970.

10. Draper, N. R., and Hunter, W. G., "The Use of Prior Distributions in the Design of Experiments for Parameter Estimation in Non-linear Situations," Biometrika, Vol. 54, 1967, pp. 147-153.

11. Fishburn, P. C., Murphy, A. H., and Isaacs, H. H., "Sensitivity of Decisions to Probability Estimation Errors: A Reexamination," Operations Research, Vol. 16, 1968, pp. 254-267.

12. Hines, W. W., and Montgomery, D. C., Probability and Statistics in Engineering and Management Science, The Ronald Press, Co., New York, 1972.

13. Isaacs, H. H., "Sensitivity of Decisions to Probability Estimation Errors," Operations Research, Vol. 11, 1963, pp. 536-552.

14. Lindley, D. V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Cambridge University Press, 1965.

15. Lindley, D. V., "The Use of Prior Probability Distributions in Statistical Inference and Decision," in J. Newman (ed.), Proceedings of the Fourth Berkley Symposium on Mathematics Statistics and Probabilities, Vol. 1, 1961 b, pp. 453-468.

16. Novick, M. R., and Grizzle, J. E., "A Bayesian Approach to the Analysis of Data from Clinical Trials," Journal of the American Statistical Association, Vol. 60, 1965, pp. 81-96.

17. Novick, M. R., and Hall, W. J., "A Bayesian Indifference Procedure," Journal of the American Statistical Association, Vol. 60, 1965, pp. 1104-1117.

18. Operations Research Associates, "Techniques and Guidelines," Operational Test Methodology Guide, Vol. II, January, 1974.

19. OTEA, "Lightweight Company Mortar System (LWCMS) Operational Test I (OT I) Final Report," U. S. Army Operational Test and Evaluation Agency, Ft. Belvoir, Va., September, 1973.

20. OTEA, "Lightweight Company Mortar System, XM 224E3, Operational Test II (LWCMS OT II) Test Report," U. S. Army Operational Test and Evaluation Agency, Falls Church, Va., 1976.

21. Pearson, E. S., "Bayes Theorem, Examined in Light of Experimental Sampling," Biometrika, Vol. 17, 1925, pp. 388-442.

22. Pearson, E. S., and Hartley, H. O., Biometrika Tables for Statisticians, Vol. 1, Cambridge University Press, 1954.

23. Peers, H. W., "On Confidence Points and Bayesian Probability Points in the Case of Several Parameters," Journal of the Royal Statistical Society B, Vol. 27, 1965, pp. 9-16.

24. Pfanzel, J., "Sampling Procedures Based on Prior Distributions and Costs," Technometrics, Vol. 5, 1963, pp. 47-61.

25. Pierce, D. A., and Folks, J. L., "Sensitivity of Bayes Procedures to the Prior Distribution," Operations Research, Vol. 17, 1969, pp. 344-350.

26. Plackett, R. L., "Current Trends in Statistical Inference," Journal of the Royal Statistical Society A, Vol. 129, 1966, pp. 249-267.

27. Pratt, J. W., Raiffa, H., and Schlaiffer, R., "The Foundations of Decision Under Uncertainty: An Elementary Exposition," Journal of the American Statistical Association, Vol. 59, 1964, pp. 353-375.

28. Pratt, J. W., Raiffa, H., and Schlaiffer, R., Introduction to Statistical Decision Theory, McGraw-Hill Book Co., New York, 1965.

29. Raiffa, H., and Schlaiffer, R., Applied Statistical Decision Theory, the MIT Press, Boston, 1961.

30. Roberts, H. V., "Probabilistic Predictions," Journal of the American Statistical Association, Vol. 60, 1965, pp. 50-62.

31. Savage, L. J., "Reading Suggestions for the Foundations of Statistics," American Statistician, Vol. 24, 1970, pp. 23-27.

32. Suppes, P., "The Role of Subjective Probability and Utility in Decision Making," in J. Neyman (ed.), Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 5, 1956, pp. 61-73.

33. Thatcher, A. R., "Relationships Between Bayesian and Confidence Limits for Predictions," Journal of the Royal Statistical Society B, Vol. 26, pp. 176-192.

34. Winkler, R. L., "The Assessment of Prior Distributions on Bayesian Analysis," Journal of the American Statistical Association, Vol. 62, 1967, pp. 776-800.

35. Winkler, R. L., "Probabilistic Predictions: Some Experimental Results," Journal of the American Statistical Association, Vol. 26, 1971, pp. 675-685.

36. Winkler, R. L., Introduction to Bayesian Inference and Decision, Holt, Rinehart, and Winston, Inc., New York, 1972.