8410036

Rohrbach, Kermit John

MONETARY UNIT ACCEPTANCE SAMPLING: SEQUENTIAL AND FIXED SAMPLE SIZE PLANS FOR SUBSTANTIVE TESTS IN AUDITING

University of Illinois at Urbana-Champaign PH.D. 1983

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

MONETARY UNIT ACCEPTANCE SAMPLING: SEQUENTIAL AND FIXED SAMPLE SIZE PLANS FOR SUBSTANTIVE TESTS IN AUDITING

BY

KERMIT JOHN ROHRBACH

B.A., University of Kansas, 1967
A.M., Indiana University, 1974
M.B.A., Indiana University, 1974

THESIS

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Accountancy in the Graduate College of the University of Illinois at Urbana-Champaign, 1983

Urbana, Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

JULY 1983

WE HEREBY RECOMMEND THAT THE THESIS BY

KERMIT JOHN ROHRBACH

ENTITLED MONETARY UNIT ACCEPTANCE SAMPLING: SEQUENTIAL AND FIXED SAMPLE SIZE PLANS FOR SUBSTANTIVE TESTS IN AUDITING

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Donald M. Roberts
Director of Thesis Research

Frederick P. Neumann
Head of Department

Committee on Final Examination†

Joseph J. Schultz, Jr. Chairman

Frederick P. Neumann

Florence C. Sharp

Robert A. Wijsman

† Required for doctor's degree but not for master's

O 517

To my mother

and

To the memory of my father

I gratefully acknowledge the advice, criticism, and encouragement of my committee. In particular, Professors Donald Roberts and Robert Wijsman were extremely generous with both their time and knowledge. I have benefited from stimulating discussions, on auditing, with Professors Joe Schultz, Fred Neumann, Chuck Hamilton, and, on things in general, with Deepinder Sidhu, and from challenging, insightful courses conducted by Bob Bohrer, Song Kim, Jim McKeown, Lou Pondy, and Bob Wijsman. I owe a special debt to Professor Charlie Smith, without whose support and encouragement at various times I would doubtless have remained ABD forever. And, to my wife, Paula, who has suffered periodic disruptions in her personal and professional life on my account, my love and admiration.

I also gratefully acknowledge financial support from the Ernst and Whinney Foundation through much of the dissertation stage of my Ph.D. program.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Independent auditors are charged with the responsibility of deciding on the fairness of the financial statements of their clients. This decision process is admittedly complex. Within this process, however, there are many relatively routine decision problems. These routine problems are often amenable to statistical modeling. Conceptually, auditors perform two types of testing--compliance and substantive. Compliance tests are designed to provide evidence with regard to the functioning of certain control features of the system that generates the transactions and balances which ultimately appear on the financial statements. While compliance tests provide indirect evidence as to the fairness of these transactions and balances, auditors also perform substantive tests designed to provide direct evidence as to their fairness. In many of these testing situations, the auditor is confronted with a large group of reasonably homogeneous items that are susceptible of definition as a population. An audit test may then be realized as the examination of a sample from a given population. Typically, the auditor examines this sample with the aim of deciding whether or not the population as a whole is acceptable.

In compliance testing, the criterion of acceptability can, in some situations, be appropriately defined as the proportion of erroneous items in the population (i.e. the population error rate). In such cases, auditors have had available a powerful statistical model known as acceptance sampling. Both fixed sample size and sequential acceptance sampling plans have been proposed for audit use. Sequential plans have the advantage of lower sampling cost, on average, than fixed sample size plans with comparable decision risks.

In substantive testing, the criterion of acceptability is the fairness of the recorded monetary value (i.e. book value). While auditors have had available a large number of statistical procedures for substantive testing, most of these procedures are inferior to acceptance sampling in the sense that decision risks cannot be reliably controlled. To the extent that these procedures are derived from survey sampling methodology, they depend on the large-sample behavior of the estimators used rather than the validity of distributional assumptions incorporated in a model of the problem. Such procedures are distribution-free (or nonparametric) since they are designed without regard for the distribution of variables in the population. But the performance of these procedures has been shown (both in the audit literature and in research on finite population sampling theory) to be population-dependent. A given estimator may perform poorly on a given population. Furthermore, survey sampling methods are geared toward estimation rather than decision. Audit tests based on these methods tend to control only one of the two decision risks faced by the auditor.

Efforts have been made to model the substantive testing problem parametrically (i.e. impose distributional assumptions on the variables in the population). If such a model is appropriate, or robust against violations of the assumptions, the resulting tests should be superior, both conceptually and in terms of various performance measures, to distribution-free methods. In this thesis, the acceptance sampling model of compliance testing is extended for use in substantive testing. I call this extension "monetary unit acceptance sampling" (MUAS). While MUAS is not an exact test, it is designed to be conservative relative to ordinary acceptance sampling (here called "physical unit acceptance sampling" (PUAS)). Thus, under normal audit conditions, the decision risks of MUAS will be bounded by the decision risks of a corresponding PUAS model. One of the drawbacks of conserva ive tests is inefficiency, i.e. more sample information is obtained than is necessary to attain allowable risk levels. However, the extension from PUAS to MUAS includes, in particular, <u>sequential</u> plans. Sequential implementation can, under certain conditions, significantly reduce the inefficiency of MUAS. I present two sequential MUAS plans. One is derived from classical sequential acceptance sampling. The other is based on a new Bayesian sequential acceptance sampling model. It is hoped that MUAS will not only be applicable in audit testing but will also have positive pedagogical value by providing a unified framework (acceptance sampling) within in which to teach audit sampling.

The extension from PUAS to MUAS is contained in Chapter 4. Included in this chapter are the results of a Monte Carlo study on the performance of MUAS. Intervening chapters contain a review of the principal sources of the new models and the development of PUAS models, including the proposal of sequential plans appropriate for audit use.

CHAPTER 2

THE DEVELOPMENT OF STATISTICAL TESTING MODELS IN AUDITING

Statistical auditing in the United States has a history of some 50 years. It is not my purpose here to attempt a reconstruction of this history. Rather, I intend to recount elements of the research in audit sampling that are pertinent to the development of the monetary unit acceptance sampling (MUAS) models of Chapter 4. These models draw primarily upon three research strains in statistical auditing: classical acceptance sampling (both fixed sample size and sequential plans), Bayesian testing models, and monetary unit sampling (MUS) models. Although MUS models have proliferated in recent years, the essential contribution in MUS, for purposes of the research at hand, occurred in 1961. More recent work stems from Anderson and Teitlebaum (1973) and forms a body of work that is not particularly germane to the development of MUAS. Accordingly, we will not review much of the research in MUS.

The history of statistical auditing has not been without controversy, and the central controversy has involved the very purpose of statistics in auditing. Early advocates of statistical auditing tacitly or explicitly assumed that a statistical model for audit use should be designed to discriminate between acceptable and unacceptable values of some significant quantity, this discrimination being done with known risks of error. In statistics, such models are referred to as (hypothesis) <u>tests</u>. The earliest statistical tests proposed for audit use were variants of acceptance sampling, and the quantity being tested was the population error rate. Beginning in the late 1950's, this view of the function of statistics in auditing came under increasing attack (see, in particular, Trueblood and Cyert (1957)). The critics felt that statistical tests supplanted auditor judgment. They argued that statistical models should provide an <u>estimate</u> of the value of some significant quantity. The auditor was, then, free to use this estimate as he saw fit.

The watershed in this controversy came in 1956 with the publication of <u>Statistical Sampling for Auditors and Accountants</u> by Vance and Neter. The first half of this handbook is devoted to an exposition of acceptance sampling (both fixed sample size and sequential). The latter half is devoted to estimation. By 1962, a similar handbook (by Hill et al.) omitted acceptance sampling entirely. The estimation techniques were taken from survey sampling. While the advocates of estimation were unhappy with statistical tests, they provided their own framework within which statistical estimates were to be used in auditing. The statistical estimate of the true value should be used "to judge the reasonableness of the book figure" (Vance and Neter (1956, p. 169)), where book figure refers to the amount recorded by the client. This judgment was to be effected by means of a confidence interval for the estimate (and, thus, we will refer to such uses of statistical estimates as confidence procedures). A $100(1-\alpha)\%$ confidence interval is designed such that, on repeated trials of the same procedure on the same population, $100(1-\alpha)\%$ of the estimates will fall in the interval. If the estimator used is unbiased, we may conclude that, in $100(1-\alpha)\%$ of these trials, such a confidence interval constructed about the estimate will contain the true value.

The rule, then, was to construct a confidence interval about the estimate; if the book value fell in this interval, it was reasonable; otherwise it was not. And, if this was indeed the rule, we must ask precisely what Trueblood and Cyert (1957, p. 20) meant when they wrote, "There are no explicit rules for decision-making that are built into the sample, nor assumed for purposes of sample size computation."

The width (or precision) of the confidence interval depends on both α and the standard deviation of the estimate, which, in general, depends on sample size. Thus, if a certain precision is desired, it is necessary to set α beforehand and draw the necessary sample size to achieve this precision. (Technically, it may not be possible, with a single sample, to guarantee that a given precision will be attained. But sample size can subsequently be increased if the desired precision is not attained.) Now, Trueblood and Cyert must have meant that the α used for choosing sample size need not be the α used to construct the confidence interval. While this is literally true, the practical consequences regarding audit judgment are interesting. Consider, for example, an auditor who sets α at .05 to choose sample size in an estimate of inventory value. The book figure is, say, $500,000, and the estimate is $470,000 ± $20,000, if $\alpha = .05$ is used in constructing the interval. Assume that the auditor is satisfied with precision of $30,000. Apparently, he is f[illegible] to decrease α until the confidence interval just contains the book value. Since α represents (in part) the risk that the interval does not contain the true value, reducing that risk should not be open to criticism.

There are several problems here that were not addressed in the audit literature until 1972. In that year, Elliott and Rogers published an influential critique of the methods used to implement confidence procedures in auditing. They claimed that confidence procedures were being used to make decisions. As such, the auditor faced two risks. Not only did he face the risk that the confidence interval did not include the book value when it was reasonable (type I risk), he also faced the risk that it included the book value when it was unreasonable (type II risk). By decreasing α in order to accept the book value, our hypothetical auditor, in the inventory example above, increased one risk (type II) while decreasing the other (type I). Elliott and Rogers went on to show how confidence procedures could be implemented to control both risks. Although Elliott and Rogers argued for the duality of confidence procedures and hypothesis tests, they preferred the framework of the former. And, at least in part because of this, they introduced a new problem in the use of confidence procedures in auditing: materiality allocation. (Materiality, as an audit construct, refers to the auditor's assumption that some degree of error is serious enough to affect the financial decisions of a reasonably prudent investor. This degree of error is called material. A lesser degree of error is immaterial and would not affect those decisions.) Materiality allocation attempts to address the problem of setting desired precision when the results of more than one confidence procedure are going to be jointly considered. We will not pursue this matter here beyond the following remarks: (i) in the testing framework (which we will be adopting), if we combine the results of several tests, the quantity of concern is the risk of two or more incorrect decisions, not some measure of combined precision, and (ii) in MUS models (of which MUAS is one example), materiality is stated as a percentage of book value, and allocation of some absolute quantity is irrelevant. Some ten years later, the Elliott and Rogers position was, by and large, incorporated in the professional audit standards in the United States (SAS No. 39). Thus, MUAS, although cast entirely in the testing framework, is reasonably consistent with current audit standards.

There were, however, more fundamental problems with many of the confidence procedures advocated for audit use. The accuracy of the intervals depended on the large-sample behavior of the estimators used. Typical audit sample sizes were uncritically assumed to be "large enough" to insure that the estimator was normally distributed. Simulations conducted by Kaplan (1973) and Neter and Loebbecke (1975, 1977) provided evidence that this assumption was not necessarily warranted. This result must be compared with acceptance sampling (to which we turn shortly). With acceptance sampling, the auditor's problem was modeled such that the test statistic followed a known distribution--no large-sample assumptions were needed. Unfortunately, acceptance sampling had been applied successfully only for certain compliance tests. In 1961, van Heerden extended the acceptance sampling model for use in substantive testing. However, this accomplishment went unnoticed by the audit profession in the United States. We will return to van Heerden in the discussion of MUS below.

We now turn to the development of classical acceptance sampling in audit tests. Carman (1933) appears to have made the first contribution to statistical auditing in the United States. Carman proposed a discovery sampling model to detect the presence of fraud in a population of similar transactions (e.g. cash disbursements). By defining a fraudulent transaction as an error and sampling at random with replacement from the population, Carman showed that the total number of errors observed obeyed a binomial distribution. This distribution has two parameters: n (sample size) and p (error rate). The error rate is unknown. If $p>0$, then, no matter what n is, there is some risk that our sample does not contain an error, and hence we conclude, incorrectly, that $p=0$ (i.e. a type II decision error). If, however, we are willing to set some minimum error rate p_2, $0<p_2<1$, that we deem significant, we can control the risk of failing to detect this (or a higher) error rate by choosing the appropriate sample size. The test, then, is of the form

hypothesis: $p=0$

alternative: $p=p_2$

(We will consistently use simple hypotheses, i.e. those that specify only one point. In the classical construction, the simple alternative above is equivalent to the composite alternative $p\geqslant p_2$.) Carman adopted the decision rule that if we observe one or more errors, we reject the hypothesis, otherwise we accept. The critical value (the minimum number of errors needed to reject the hypothesis) need not be set higher than one, since, if even one error is observed, the hypothesis is certainly false. And, by requiring at least one error in order to reject, we face no type I risk. However, if we accept, there is some risk of having done so unfairly. Carman showed that this risk--type II risk--could be controlled by choosing sample size--the larger the sample size, the smaller the type II risk. Carman also observed that this plan may be implemented sequentially. If we observe an error, the test may be terminated and the hypothesis rejected. This procedure should reduce average sample size but there is no effect on decision risks (type I risk remains zero).

Although several articles in the late 1940's and early 1950's dealt informally with the use of acceptance sampling in auditing, the first formal exposition in the audit literature seems to have been Vance and Neter (1956). The hypothesis of a zero error rate (used in discovery sampling) is rarely justifiable in testing accounting controls since the auditor usually does not expect the control to function perfectly. If the auditor both expects a positive error rate and can tolerate a certain amount of error in the population, use of discovery sampling will result, more or less often, in rejection of the hypothesis when, in fact, the population error rate is at an acceptable level. While formally the auditor faces no type I risk, this is irrelevant because the problem has not been correctly modeled. Acceptance sampling is designed to discriminate between an acceptable (but positive) error rate and an unacceptable error rate. The test is of the form

hypothesis: $p=p_1$

alternative: $p=p_2$

where $0<p_1<p_2<1$ and p_1 is an acceptable error rate. We now face both type I risk (reject unfairly) and type II risk (accept unfairly). To control these risks, we now manipulate both sample size and critical value. (Critical value can no longer be independently set at one as in discovery sampling.)

The implementation of sequential acceptance sampling presents difficulties far beyond those of sequential discovery sampling. However, in the 1940's, Wald developed a sequential test of hypotheses, one form of which was sequential acceptance sampling (Wald (1947)). Vance (1950) adapted Wald's test to audit problems. In sequential acceptance sampling, we must decide at each sampling stage (e.g. after each observation) whether to accept, reject, or continue to make observations (because both the type I and II risks of an immediate decision are too high). This amounts to finding, at each sampling stage, an appropriate number of observed errors at which to accept and an appropriate number at which to reject. If the number of observed errors lies between these two numbers, we continue to make observations. The advantage over fixed sample size acceptance sampling is that, on average, we will make decisions at the same risks but with fewer observations. Further, as Vance was quick to recognize, most audit tests are, in fact, conducted sequentially. A sequential sampling plan represented a natural formulation of the audit problem.

Since the notion occurs in other discussions of sequential sampling, we should note that Vance committed a serious breach of the statistical testing paradigm. He suggested that one of the benefits of sequential testing was that it allowed the auditor to continue sampling if he was dissatisfied with the result at any given stage. This amounts to choosing the decision rule <u>after</u> the data have been observed. If the auditor wishes to control decision risks at stated levels, he is not free to adopt a new decision rule in the event that the results under the old rule are not to his liking. A correct formulation for the behavior suggested by Vance is a sequential plan in which type I risk is at lower than allowable levels at early sampling stages and rises gradually to the allowable level at late stages. Roberts (1976) proposed such a plan. It is a four-stage sampling plan, truncated at the fourth stage. One of the shortcomings of Vance's proposal was the absence of any truncation rule. Thus, at least occasionally, sample size could be quite large. Truncation, however, affects decision risks and complicates analysis of the behavior of the test. In part to overcome this difficulty and in part to simplify implementation, Roberts proposed grouping the observations to yield a four-stage test. (A more accessible source for this sampling plan is Roberts (1978) p. 57ff.) Implementation difficulties have, until recently, plagued sequential sampling. The advent of computer-assisted auditing, based primarily on microcomputers, has radically altered this situation.

Despite the impressive logic of these acceptance sampling models, it appeared for some time that they could not be applied to test the fairness of a monetary value (i.e. a substantive test). Vance (1950) had already recognized the desirability of such an extension but considered it impossible due to the absence of a necessary relationship between the occurrence of an error and the monetary value of that error. In 1961, van Heerden offered an ingenious solution to this dilemma. The monetary value of interest is typically contained in a balance composed of subunits defined by the audit client (e.g. an inventory balance composed of various items or parts). Van Heerden suggested that, instead of viewing this balance as a population of natural subunits, we view it as a population of monetary units (dollars, pounds, francs, marks, yen, etc.). For convenience, we will refer to these units as "dollars." We agree to classify a dollar as either fictitious (an error) or sound (a nonerror). The error rate now becomes an index of the reasonableness of the book value: a high error rate indicates material overstatement; a low error rate indicates immaterial overstatement. This general approach is called monetary unit sampling (MUS).

A difficulty arises when we actually attempt to identify a particular sample dollar as fictitious, because the client accounts for subunits rather than the individual dollars that comprise the subunits. If we are willing to adopt a discovery sampling model, this identification problem is not serious. If any of our sample dollars belong to a subunit that is overstated, we may safely reject the hypothesis that $p=0$. But once we adopt the more realistic acceptance sampling model, the identification problem is critical. Van Heerden not only solved this problem but solved it in such a way that the whole apparatus of acceptance sampling worked exactly as it had in the nonmonetary situation. In particular, the number of observed errors could be constrained to obey the binomial distribution. (Van Heerden's solution is formally considered in Chapter 4, including a proof of this latter claim.)

Lest I overstate van Heerden's contribution, let me add that, as written, van Heerden proposed an MUS <u>discovery</u> sampling plan. While he provided a methodology to implement MUS acceptance sampling, he does not make details of such an implementation clear, referring only to certain (unidentified) tables to aid the auditor if one or more errors are actually observed. It does appear that van Heerden used the discovery sampling model simply because it requires fewer observations than an acceptance sampling model with the same alternative. Thus, even if an error is observed, it is not clear that van Heerden is willing to reject the hypothesis. Similarly, in a reference to the sequential implementation of his plan (again, he provides no details), he repeats Vance's contention that the auditor can continue sampling if the initial result is "unsatisfactory."

The last research strain that we draw upon is the Bayesian testing framework. By and large, the work in Bayesian models in auditing has been in estimation and involves a reformulation of confidence procedures. In the Bayesian framework, estimation may naturally lead to considerably more complicated models than testing. The essential elements of Bayesian testing were introduced in the audit literature by Kinney (1975). Kinney assumed that there are two possible "states of nature" facing the auditor: (i) the book value is materially correct, and (ii) the book value is materially incorrect. The auditor must decide which of these states actually holds. As in the classical acceptance sampling framework, the auditor can make two decision errors--type I and type II. However, the risks of these errors are defined not as probabilities but as expected losses. That is, we define a loss function that specifies our losses for all possible outcomes (with two possible decisions and two possible states of nature, there are four possible outcomes). Kinney's loss function consists of a variable sampling cost, a fixed cost to access the sampling frame, and a fixed cost for an incorrect decision (which may vary as to the type of decision error). The auditor wishes, in some sense, to minimize his expected loss. However, of two competing decision rules (sampling plans), one may have lower expected loss under one state of nature and higher expected loss under the other. As it stands, these rules are noncomparable. The Bayesian approach solves this problem by requiring that we weight the expected losses using a prior distribution on the states of nature. Thus, if, before sampling, we feel that one state is more likely than another, the expected loss under this state plays a more significant role in our choice of decision rules. With the addition of a prior distribution and loss function, the acceptance sampling model goes through much as before.

Although the idea of choosing a decision rule with minimum risk is appealing, we may ask if there is any set of principles that requires us to do so. If we define loss as negative utility and if our utility function obeys the von Neumann-Morgenstern (1953, p. 23) axioms, then this question may be answered affirmatively. This defense of Bayesian procedures has been expounded at length by Savage (1972) and Lindley (1971).

In the following chapters, I present both classical and Bayesian testing models. I assume that all of the models can be usefully applied to assist the auditor in making certain routine (but nonetheless important) decisions. Juxtaposition of the two approaches to the same problem will, it is hoped, facilitate a reasoned choice between them.

CHAPTER 3

A STATISTICAL COMPLIANCE TESTING MODEL: PHYSICAL UNIT ACCEPTANCE SAMPLING

Auditors perform a variety of tests. Conceptually, two types of audit tests are defined in the professional audit standards in the United States (SAS No. 1): compliance tests and substantive tests. In compliance testing, it is often reasonable to identify the object of interest as an error rate in a population of similar transactions. An error in this situation is the failure of some control feature in the accounting system that generated the transactions. For example, a proper cash disbursement should exhibit, among other things, an authorized signature on the document effecting the disbursement. The lack of an authorized signature can be defined as an error. Typically, the auditor expects the population to contain some errors (i.e. controls are not expected to operate perfectly) and is interested in discriminating between an acceptably low error rate and an unacceptably high error rate.

Such situations correspond very closely with the quality control inspection setup, in which production lots are examined with the aim of discriminating between lots in which the rate of defective items is acceptably low and those in which it is unacceptably high. Acceptance sampling is a statistical procedure first designed to model the quality control inspection setup. Subsequently, acceptance sampling was adopted for use in audit testing.

In this chapter, we consider the acceptance sampling model in several forms. Our purposes are twofold. First, the development presented in Chapter 4 extends the use of acceptance sampling to substantive tests of the fairness of a monetary value. Thus, the models of this chapter are of broader applicability than may be immediately apparent. (In part to distinguish these models from the extension in Chapter 4 and in part because of the modifications cited below, we formally refer to the models of this chapter as <u>physical unit</u> acceptance sampling (PUAS). But, informally, we retain the general term acceptance sampling.) Second, I propose several modifications to acceptance sampling for audit use. These include (i) a simplified Bayesian framework for acceptance sampling that should prove easier to implement than previously proposed Bayesian models for audit tests, (ii) a new Bayesian sequential acceptance sampling model, and (iii) algorithms to compute the exact decision risks and approximate expected sample sizes of the proposed sequential tests--these algorithms should be efficient for typical audit sample sizes (say, $n \leq 200$). Before discussing PUAS as such, we briefly consider the general testing framework within which the PUAS models will be developed.

Acceptance sampling is one form of statistical test. To conduct any statistical test, we must model our problem along the following lines. We identify the characteristic of interest with the random variable (or vector) X. A realization of X will be denoted as x, and the set of all possible realizations will be denoted by $\mathcal{X}$, the sample space. We assume that the distribution of X (or of some function of X) is one of the family $\{P_p: p \in \mathcal{P}\}$ indexed by the parameter p. Two subsets of the parameter space $\mathcal{P}$ are of interest: $\mathcal{P}_1$ and $\mathcal{P}_2$. It is usual, but not necessary, that these subsets exhaust the parameter space. For reasonable tests, the subsets must be disjunctive. Two hypotheses, the null and the alternative, are entertained with regard to p, namely, H_1: $p \in \mathcal{P}_1$ and H_2: $p \in \mathcal{P}_2$. We will consider the case where these two subsets are restricted to one point each, i.e. a test of simple hypotheses:

$$\begin{aligned} H_1&: p=p_1 \\ H_2&: p=p_2 \end{aligned} \quad (1)$$

(We will occasionally refer to the null hypothesis, H_1, simply as the hypothesis. Furthermore, the terms accept and reject, when used alone, should be understood to refer to the null hypothesis rather than the alternative hypothesis.)

A finite action space $\mathcal{A}$ is available, usually consisting of

$$a_1 = \text{choose } H_1 \text{ (accept the null)} \qquad (2)$$
$$a_2 = \text{choose } H_2 \text{ (reject the null)}$$

For sequential tests, we must extend this space to include the nonterminal action

$$a_3 = \text{continue sampling} \qquad (3)$$

We seek a decision rule $d: \mathcal{X} \rightarrow \mathcal{A}$, in a given class $\mathcal{D}$ of rules, with minimum risk. Risk is defined differently in the classical and Bayesian testing frameworks. Two decision errors may occur:

$$\text{type I error: choose } H_2 \text{ when } H_1 \text{ is true} \qquad (4)$$
$$\text{type II error: choose } H_1 \text{ when } H_2 \text{ is true}$$

We will refer, on occasion, to the operating characteristic (OC) function and the power function. They are defined by

$$\text{OC function: } \alpha(p) = P_p\{d(X) = a_1\} \qquad (5)$$
$$\text{power function: } \beta(p) = P_p\{d(X) = a_2\}$$

The OC function gives, for all p, the probability of taking action a_1 (accept H_1). The power function gives, for all p, the probability of taking action a_2 (reject H_1). For proper tests, $\alpha(p)+\beta(p)=1$.

3.1 N-P Fixed Sample Size Acceptance Sampling

Neyman-Pearson (N-P) tests are an important subset, a most powerful subset, of likelihood ratio (LR) tests. My description of N-P testing follows Bickel and Doksum (1977).

In the N-P approach, risk is defined as the probability of decision error (error probability for short). There are two risks corresponding to the two types of decision error:

$$\begin{aligned} &\text{type I risk:} \quad P_{p_1}\{d(X)=a_2\} \\ &\text{type II risk:} \quad P_{p_2}\{d(X)=a_1\} \end{aligned} \tag{6}$$

We may also state these risks in terms of the OC and power functions:

$$\begin{aligned} &\text{type I risk:} \quad \beta(p_1)=1-\alpha(p_1) \\ &\text{type II risk:} \quad \alpha(p_2)=1-\beta(p_2) \end{aligned} \tag{7}$$

The type I risk of d is called the level of the test and is conventionally denoted α. Type II risk of d is conventionally denoted β, and $1-\beta$ is referred to as the power of the test. The N-P criterion is to find, within the class of all fixed sample size decision rules with level at least α, the most powerful rule. Thus, we minimize type II risk for a given type I risk. The N-P Lemma states that the most powerful rule for problem (1) is of the form:

$$d^n(x)=\begin{cases} a_1 & \text{if } f^n(x;p_2)/f^n(x;p_1) < D, \quad D \geq 0 \\ a_2 & \text{otherwise} \end{cases} \tag{8}$$

where $f^n(x;p)$ is the (conditional on p) frequency function (if X is discrete) or density function (if X is continuous) of $X=(X_1,\ldots,X_n)$, and D is some constant. The ratio of frequencies (or densities) is called the likelihood ratio (LR) and will be denoted by

$$l^n(x,p_1,p_2)=f^n(x;p_2)/f^n(x;p_1) \quad \text{for } x=(x_1,\ldots,x_n) \tag{9}$$

(If $x=x_i$, the LR will be denoted by $l(x_i,p_1,p_2)$.)

It is true in general that we can find a decision rule based on a sufficient statistic for p that is risk-equivalent to any rule using the sample information itself. Often, the decision rule can be simplified by finding a test statistic T(X) that is sufficient. The equivalent rule is

$$d^n(x) = \begin{cases} a_1 & \text{if } T_n(x) = t < C \\ a_2 & \text{otherwise} \end{cases} \tag{10}$$

where the constant C is called the critical value. The critical region, where $d^n(x) = a_2$, is $\{x: T_n(x) = t \geq C\}$. The decision rule is specified by choosing critical value C so as to attain a desired level and sample size n so as to attain a desired power. That is, we seek the smallest C and n such that the following conditions hold:

$$\begin{aligned} \beta(p_1) &= P_{p_1}\{T_n(X) \geq C\} \leq \alpha \\ \beta(p_2) &= P_{p_2}\{T_n(X) \geq C\} \geq 1-\beta \end{aligned} \tag{11}$$

(The rightmost inequalities reflect the possibility that exact level and power may not be attainable for discrete distributions, unless we randomize over decision rules.)

I present the following example, which may be construed as a compliance test, in some detail. The same example will be used for the alternative models discussed later. The use of one example should facilitate comparison of the models.

Example 3.1. An audit client maintains a purchased parts inventory on perpetual records. It is carefully controlled, and the client would prefer that the auditor rely on the perpetuals rather than require a complete count. The auditor agrees to test the perpetuals. One procedure in this test will be the comparison of recorded and on-hand quantities for a sample of items. In this procedure, the auditor is primarily concerned with the proportion of errors rather than the size of the errors, which he expects to be uniformly small. The auditor decides to model the problem statistically as follows:

(i) a difference between recorded and on-hand quantities will be treated as an error (all items are errors or nonerrors)

(ii) a counted item will be identified with the random variable X according to the rule:

$$X_i = \begin{cases} 1 \text{ if the ith item is an error} \\ 0 \text{ otherwise} \end{cases}$$

(iii) selection of items to count will be made randomly with replacement from the perpetual records (the sampling frame)

Under these conditions, the $\{X_i\}$ are independent and identically distributed (i.i.d.) binomial random variables with parameters 1 and p, where p is the (unknown) error rate. In simpler notation, $X_i \sim \text{binomial}(1,p)$. (See Appendix A for this and other distributions mentioned in this chapter.) Further, $S_n = \sum_{i=1}^{n} X_i \sim \text{binomial}(n,p)$ and is sufficient for p.

The auditor's problem is now transformed into a test for p. The client claims the error rate does not exceed .01. The auditor decides that an error rate of .05 or more is unacceptable. He proposes to test

H_1: p=.01

H_2: p=.05

The N-P decision rule is of the form (10):

$$d^n(x)=\begin{cases} a_1 & \text{if } s_n < C \\ a_2 & \text{otherwise} \end{cases}$$

where $s_n=\sum_{i=1}^{n} x_i$. To find the critical value and sample size, the auditor must specify the desired level and power of the test. He chooses .10 and .85 respectively. Thus, he wants

$$\beta(.01)=P_{.01}\{S_n \geq C\} \leq .10$$
$$\beta(.05)=P_{.05}\{S_n \geq C\} \geq .85$$

Using binomial tables, we find n=94 and C=3 is an acceptable test, with $\beta(.01)=.069$ and $\beta(.05)=.855$. The auditor proceeds to select randomly with replacement 94 items from the perpetuals. He then counts each item and records the errors observed. If these equal or exceed 3, H_1 is rejected and the error rate assumed to be .05.

As a practical matter, if tables are to be used, it is more convenient to use the Poisson approximation to the binomial distribution. I provide a short table of the cumulative Poisson distribution in Appendix B. To use the Poisson approximation, set q=np. In Example 3.1, $q_1=.01n$ and $q_2=.05n$, thus, $q_2=5q_1$. For any given q, find the smallest C that gives a level of .10 or less. Then check the power obtained with this C for 5q. For Example 3.1, we try, say, $q_1=1.0$. The smallest C is 3, giving a level of .080. The power of this test is found under $q_2=5$ with C=3. It is .875--slightly high. With C=3, the smallest q_2 possible is 4.70 with a power of .848 (assuming we are willing to round to .85). Then $q_1=4.70/5=.94$ and linear interpolation gives a level of .070. Thus, n=94 and C=3 is an acceptable test.

Some care must be taken in using tables of discrete distributions, since the underlying function is not smooth. For example, although n=94 was the best we could do if C=3, we have not yet ruled out the possibility that a smaller n with C=2 might work. In fact, n=68 with C=2 gives acceptable power but an unacceptable level. Nevertheless, this test would be preferable to any other using a sample size between 68 and 94 with C=2.

We now consider a post-experimental measure of risk. If we accept H_1, then less than C errors were observed, and we would have accepted H_1 even if C had been set as low as s+1, where we observed s errors. We define the achieved power of the test as $P_{p_2}\{S_n \geq s+1\}$. Similarly, if $s \geq C$, we would have rejected H_1 even if C had been set as high as s. We define the achieved level of the test as $P_{p_1}\{S_n \geq s\}$. (The achieved level is more commonly called the p-value of the test.) Now, the achieved power and level of the test will equal the desired power and level only if s=C-1 and s=C, respectively. When this is not so, the test has "overshot" its goal, and risk has been reduced below desired levels, at the expense of some unnecessary sampling. Assume, in Example 3.1, that we observe s=1 errors and accept H_1. Since we controlled power at .85, we know that the chance of this result if H_2 is true does not exceed .15, but apparently it is less. Referring to q_2=4.70 in Appendix B, we find achieved power of .991. That is, there is a chance of about .01 of this result if H_2 is true.

Alternatively, if we observe s=5 errors and reject H_1, we find under q_1=.94 an achieved level of about .003. Achieved power and level provide a post-experimental measure of our "confidence" in the decision.

Given a positive probability of "overshooting", the N-P test apparently can be improved upon by some procedure that "stops" nearer the goal. By the N-P Lemma, no fixed sample size procedure can improve upon the N-P test. However, Wald's sequential probability ratio test, which we will consider in the next section, is designed precisely to reduce this amount of "overshoot" and does improve on the N-P test.

The model presented in Example 3.1 is isomorphic with the (fixed sample size) acceptance sampling model in quality control inspection, where sample size is small relative to lot size or sampling is with replacement. We extend the model to discovery sampling in Example 3.2 below. This example will not be presented for alternative models discussed later.

Example 3.2. An audit client processes payroll on computer. The payroll register is generated under control of a program that has been in use for several years. One of the auditor's procedures tests the crossfooting accuracy of the register. The client claims that no crossfooting errors occur. The auditor will tolerate an error rate of less than .05. He proposes testing

H_1: p=.00

H_2: p=.05

The binomial distribution is degenerate at p=0, hence it would seem impossible to compute a level for this test. However, if the auditor chooses C=1, he cannot reject unfairly, i.e. he faces no type I risk. Thus, setting C=1, n may be found as before by controlling power. If desired power is .85, we find the smallest acceptable q_2 to be 1.90 giving n=38 and power of .850.

3.2 Wald Sequential Acceptance Sampling

Regardless of sampling plan, the audit of the sample (i.e. the fieldwork) proceeds sequentially in compliance tests. In fixed sample size tests, it is apparent that, as soon as a critical number of errors is found, auditing of the sample may stop and the null may be rejected. But there is no similar shortcut to accepting the null. In discovery sampling, this is reasonable, since acceptance requires an entirely error-free sample. But, in acceptance sampling, there may very well come a point during the test when one or the other action becomes highly improbable. It would be advantageous to have a rule that tells the auditor when a given action becomes sufficiently improbable, allowing him to terminate fieldwork on the test. More generally, the rule should indicate when the risk of a given action becomes acceptably low. Wald's sequential probability ratio test (SPRT) is such a rule.

Wald developed the SPRT during the 1940's. There have been extensions, but my description mainly follows Wald (1947). Wald improved on the N-P test by enlarging the class of procedures being considered. The additional procedures are those for which the number of observations is random. These procedures--sequential procedures--terminate when evidence for one hypothesis becomes persuasive. The improvement is in sample size: for given level and power, the SPRT has a significantly lower expected sample size than the optimal fixed sample size of the N-P test. For the problem given by (1), Wald and Wolfowitz (1948) proved that the SPRT has the lowest expected sample sizes (under H_1 and H_2) of all tests with level α and power $1-\beta$.

The N-P test for problem (1) rejects H_1 when the LR (9) equals or exceeds some positive constant. Wald suggested forming the LR after each observation. By appropriate choice of constants A and B ($0 < A < 1 < B < \infty$), a test of level α and power $1-\beta$ is

$$d^n(x) = \begin{cases} a_1 & \text{if } l^n(x, p_1, p_2) \leq A \\ a_2 & \text{if } l^n(x, p_1, p_2) \geq B \\ a_3 & \text{otherwise} \end{cases} \tag{12}$$

where action a_3 is "continue sampling." This is the extended action space given by (2) and (3).

As in (10) above, $T_n(X) = S_n$ is a sufficient statistic, and the decision rule given by (12) is risk-equivalent to

$$d^n(x)=\begin{cases} a_1 & \text{if } T_n(x)\leq a_n \\ a_2 & \text{if } T_n(x)\geq r_n \\ a_3 & \text{otherwise} \end{cases} \tag{13}$$

where a_n and r_n are, respectively, integer-valued acceptance and rejection numbers. These numbers may be determined from the LR bounds A and B as follows (see Wald (1947, p. 90ff) for the details of this derivation): let

$$\begin{aligned} w&=p_2/p_1 \\ y&=(1-p_2)/(1-p_1) \end{aligned} \tag{14}$$

and

$$\begin{aligned} u&=(\log A - n(\log y))/(\log w - \log y) \\ v&=(\log B - n(\log y))/(\log w - \log y) \end{aligned} \tag{15}$$

Then, a_n is the largest integer $\leq u$, and r_n is the smallest integer $\geq v$.

The test in (13) is completely specified once we have chosen the bounds A and B. Unfortunately, these bounds depend upon the sampling distribution of the LR and, so, may be difficult to determine. However, Wald proved that

$$\begin{aligned} A&\geq\beta/(1-\alpha)=A' \\ B&\leq(1-\beta)/\alpha=B' \end{aligned} \tag{16}$$

Replacing A and B with A' and B' results in a change in risks from α and β to α' and β'. Wald showed that $\alpha'+\beta'\leq\alpha+\beta$. He also obtained useful approximations for the OC function and the ASN (expected sample size) function of the SPRT. We will not pursue these results further due to considerations raised in the following paragraph.

The SPRT is only one of many possible sequential tests. Its distinguishing characteristic is the use of constant bounds for the LR (i.e. A and B do not depend on sampling stage n). Termination occurs only when one or the other bound is reached or exceeded. The SPRT possesses certain optimal properties--the result obtained by Wald and Wolfowitz has been noted. But the true SPRT has not been used extensively (see comments by Wetherill (1975, p. 24)). Presumably, the variability of sample size, with its detrimental effect on the planning of experiments, is an important factor. Although Wald proved that the SPRT terminates with probability one, the sample size will occasionally be large relative to the expected size. To guard against this eventuality, various truncation rules have been proposed. These rules do not allow the sample size to exceed some stated maximum. As a practical matter, I will assume that only truncated sequential procedures are acceptable for use in audit tests, and we will restrict our search for sequential procedures to the class of truncated SPRTs. (It should be noted that it is not clear that truncated SPRTs enjoy any optimal properties with respect to the class of truncated sequential procedures.) The Wald approximations for the OC and ASN functions are not useful for truncated SPRTs if truncation occurs at moderate sample sizes. The OC function of truncated SPRTs may be obtained exactly, and the ASN function may be approximated more closely.

The choice of truncation rule is not obvious. Wald speculated that, if truncation occurred somewhat beyond the optimal fixed sample size, the increase in decision risk would be moderate. In the spirit of this speculation, a reasonable truncation rule is the following: pursue the SPRT until a terminal decision is made or the N-P optimal fixed sample size is reached; if the latter occurs, abandon the SPRT and follow the N-P rule. More formally, the decision rule is

$$d(x)=\begin{cases} d^n(x) & \text{if } n<n^* \\ d^{n^*}(x) & \text{otherwise} \end{cases} \qquad (17)$$

where

$$d^n(x)=\begin{cases} a_1 & \text{if } T_n(x)\leq a_n \\ a_2 & \text{if } T_n(x)\geq r_n \\ a_3 & \text{otherwise} \end{cases} \qquad (18)$$

and

$$d^{n^*}(x)=\begin{cases} a_1 & \text{if } T_{n^*}(x)< C \\ a_2 & \text{otherwise} \end{cases} \qquad (19)$$

In the present case, $T_n(X)=\sum_{i=1}^{n} X_i=S_n$. As before, C is the critical value of the N-P test, and we will now refer to the optimal fixed sample size as n^*.

The test $d^n(x)$ is derived from $d^{n^*}(x)$ as follows: given desired risks of α and β, an optimal fixed sample procedure is selected with risks of α^* and β^* not exceeding the desired risks; the bounds (A',B') for the SPRT are computed using α^* and β^* and are converted into acceptance/rejection numbers by the relation given in (15), except that r_n cannot exceed C. (The relation in (15) may produce an $r_n>C$, however, once $S_n=C$, the test will reject at $n=n^*$ if not earlier. Hence, the restriction $r_n\leq C$ lowers ASN with no effect on risk.) We now turn to the OC and ASN functions of $d(x)$.

It will be necessary to take into account explicitly the randomness of n in sequential tests. To this end, let us denote the random stopping time (the value of n when the test terminates) by N.

In principle, the OC function of any truncated SPRT may be obtained by a method described by Aroian (1968). The method is based on the observation that the test can terminate in acceptance only at the acceptance points. Similarly, if the test accepts, then the test statistic at the termination point, S_N, can only be an acceptance number corresponding to the acceptance point N. More formally, let

$$\alpha_i(p)=P_p\{S_N=i \text{ and the test accepts } H_1\} \tag{20}$$

Then,

$$\alpha(p)=\sum_{i=0}^{C-1}\alpha_i(p) \tag{21}$$

where C is the critical value of the fixed sample size test at n^*. Note that N is a function of i if the test accepts. The summation in (21) runs only to $C-1$ since it is the largest acceptance number. Since all truncated SPRTs are proper tests, we have immediately $\beta(p)=1-\alpha(p)$.

<u>Example 3.3 (continued from Example 3.1)</u>. In Example 3.1 we found $n^*=94$, $C=3$, $\alpha=.070$, and $\beta=.152$. Substituting in (16),

$A'=.152/(1.0-.070)=.163$

$B'=(1.0-.152)/.070=12.114$

Using the relation in (15) and bearing in mind that we will truncate the test at $n^*=94$ if no decision is made earlier,

TABLE 3.1

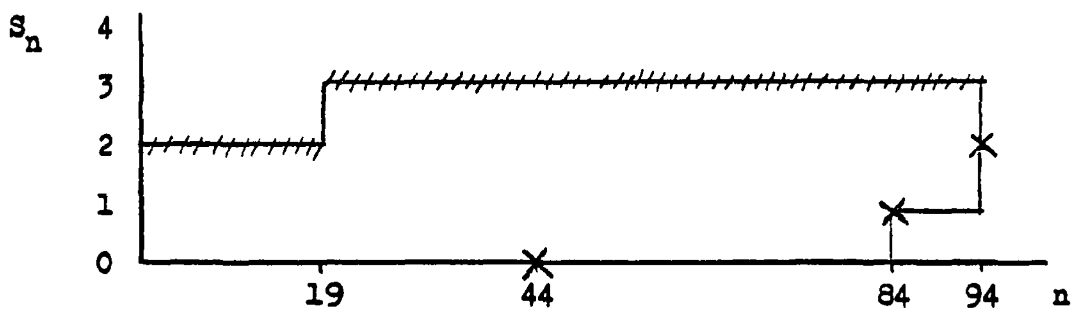
Acceptance/Rejection Numbers for the Test in Example 3.3

n	a_n
44	0
84	1
94	2

n	r_n
$2 \leq n \leq 19$	2
$20 \leq n \leq 94$	3

FIGURE 3.1

Acceptance/Rejection Regions for the Test in Example 3.3



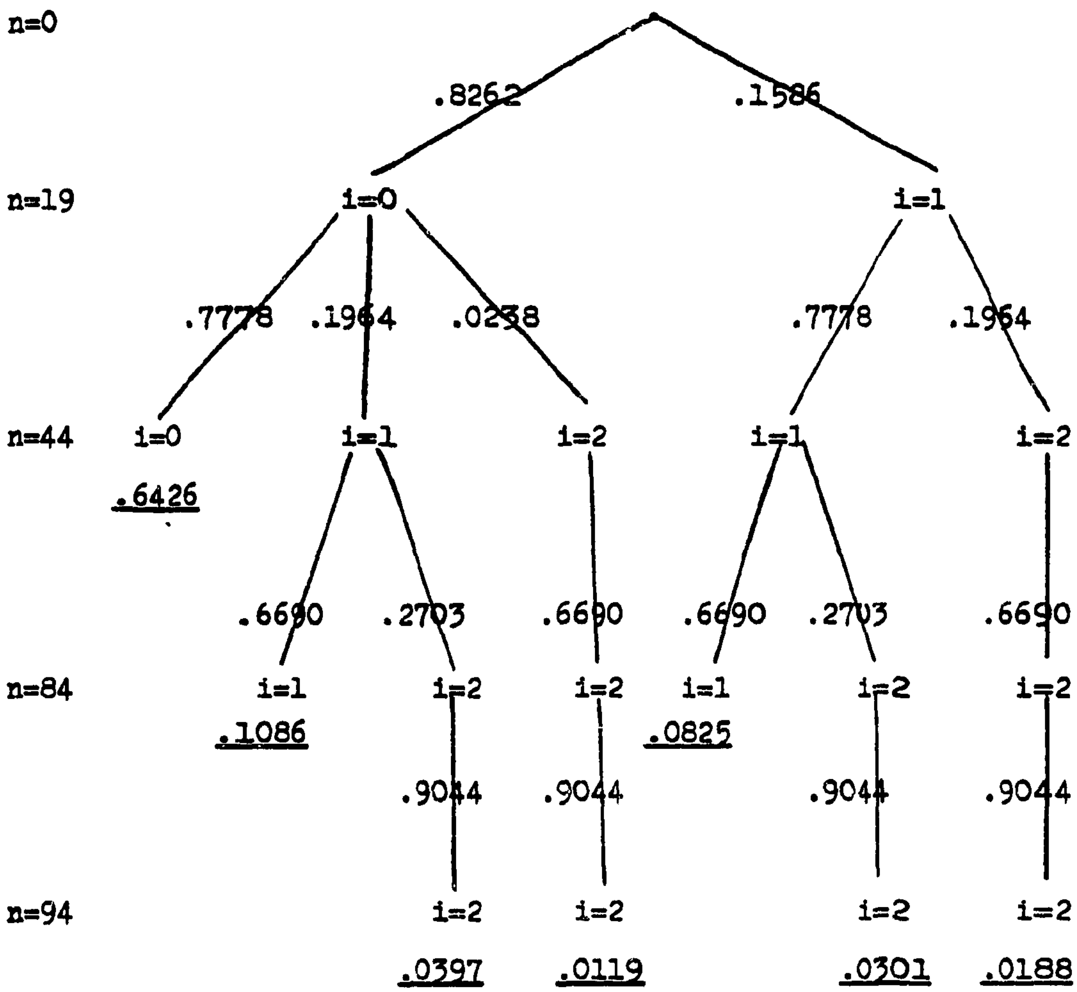
legend: hatched line=rejection boundary
"x"=acceptance point

note: the rejection boundary is not actually continuous but consists of 94 rejection points (one of which, $r_1=2$, cannot be achieved) the acceptance and rejection numbers are entered in Table 3.1. Note that, while acceptance can occur at only three points (n=44, 84, or 94), rejection can occur at any point except n=1. Recall also that no r_n is allowed to exceed C=3, even though r_n should increase to 4 at n=60 by the relation given in (15). We may graph this test as in Figure 3.1, where, for convenience, the rejection points are not shown exactly. Clearly, direct computation of the power function is not practicable. But the OC function is more tractable. We first identify the possible paths to acceptance and then compute the probability associated with that path. This method is best illustrated by means of a tree diagram, which we will call an acceptance tree. The acceptance tree for this test is presented in Figure 3.2 along with the branch and path probabilities assuming p=.01. I have used the binomial distribution here, rather than the Poisson, because some of the branches are quite short, if length is measured in number of observations, and the Poisson approximation becomes inaccurate.

To illustrate the computations involved, the probability of the first branch is $P_{.01}\{S_{19}=0\}=.8262$. The probability of the leftmost branch at the second level is $P_{.01}\{S_{25}=0\}=.7778$, where the length of this branch is 44-19=25. This path now terminates, and, since it is the only path to acceptance with i=0 errors, we have $\alpha_0(.01)=(.8262)(.7778)=.6426$, the probability of this terminal path. There are two paths that terminate in acceptance at n=84 with i=1, hence $\alpha_1(.01)=.1086+.0825=.1911$. Proceeding in this way, we find $\alpha(.01)=.934$.

FIGURE 3.2

Acceptance Tree for the Test in Example 3.3



note: probabilities computed assuming p=.01

Thus, the level of the test is 1-.934=.066. Computing the same terminal paths assuming p=.05 gives $\alpha(.05)=.192$. Hence, the power of the test is 1-.192=.808.

Arcian's method becomes tedious when critical value and sample size become even moderately large. But the computation is amenable to computer solution, and an algorithm to perform this chore may be found in Appendix C, as well as an algorithm to compute acceptance/rejection numbers. For typical audit sample sizes (say, n=200 or less), this algorithm is efficient. We note that the OC function of the truncated SPRT differs from that of the N-P test. We will return to this question after considering the ASN function.

Although Wald provided an approximation for the expected sample size of the SPRT, this approximation is too conservative for truncated SPRTs when truncation occurs at moderate sample sizes. Moreover, the truncation rule we have adopted alters the rejection region, affecting the ASN function. (As was noted in Example 3.3, the rejection numbers for the true SPRT would have increased to $r_n=4$ at n=60.) For these reasons, I will derive a better approximation to the ASN function for the test given by (17). The method of derivation is due to Wald (1947, p. 52f).

Note that, due the truncation rule of (17), $N \lesssim n^*$, where n^* is the optimal fixed sample size. Partition the sum $S_{n^*}=X_1+\ldots+X_{n^*}$ as follows:

$$X_1+\ldots+X_{n^*}=(X_1+\ldots+X_N)+(X_{N+1}+\ldots X_{n^*}) \qquad (22)$$

Taking expectations and letting $E(X)=E(X_1)=\ldots=E(X_{n*})$,

$$n^*E(X)=E(X_1+\ldots+X_N)+E(X_{N+1}+\ldots+X_{n*}) \tag{23}$$

Since, for $m>N$, X_m is independent of N,

$$\begin{aligned} E(X_{N+1}+\ldots+X_{n*})&=E(n^*-N)E(X) \\ &=n^*E(X)-E(N)E(X) \end{aligned} \tag{24}$$

Substituting (24) in (23),

$$\begin{aligned} E(N)&=E(X_1+\ldots+X_N)/E(X) \\ &=E(S_N)/E(X) \end{aligned} \tag{25}$$

Thus, under p,

$$E_p(N)=E_p(S_N)/p \tag{26}$$

where we assume $p>0$.

Now, in the test given by (17), S_N can take on only the values $0,1,\ldots,C$, where C is the critical value of $d^{n^*}(x)$. To assess $E_p(S_N)$, we need the probabilities that S_N takes on these values. Define, analogously with (20),

$$\beta_i(p)=P_p\{S_N=i \text{ and the test rejects } H_1\} \tag{27}$$

then

$$\beta(p)=\sum_{i=1}^{C}\beta_i(p) \tag{28}$$

Now

$$E_p(S_N)=\sum_{i=0}^{C-1} i\,\alpha_i(p)+\sum_{i=1}^{C} i\,\beta_i(p) \tag{29}$$

The probabilities $\{\alpha_i(p)\}$ are provided by Aroian's method, but the $\{\beta_i(p)\}$ are not so easily assessed. However,

$$\sum_{i=1}^{C} i\,\beta_i(p)\leq\sum_{i=1}^{C} C\,\beta_i(p)=C\,\beta(p) \tag{30}$$

and $\beta(p)$ is known. Hence

$$E_p(S_N)\leq\sum_{i=0}^{C-1} i\,\alpha_i(p)+C\,\beta(p) \tag{31}$$

and

$$E_p(N)\leq\Big(\sum_{i=0}^{C-1} i\,\alpha_i(p)+C\,\beta(p)\Big)/p \tag{32}$$

The approximation given by (32) may be fairly good if $\beta(p)$ is not too large, thus it should be better under $p=p_1$ than $p=p_2$. We can easily improve on it for most tests.

Let m_1 be the first rejection point, let $j=r_{m_1}$ be the rejection number at this point, and let m_2 be the last rejection point for which the rejection number is j, i.e. $r_{m_1}=r_{m_1+1}=\ldots=r_{m_2}=j$. Then, if we assume that it is not possible to accept H_1 at or before m_2,

$$\beta_j(p)=P_p\{S_{m_2}\geqslant j\} \tag{33}$$

where the probability is based on the fixed sample size of m_2 observations. (Conceptually, we can extend any path for which $N\leqslant m_2$ to the point m_2. For any such hypothetical path, $S_{m_2}\geqslant j$ because, if $N\leqslant m_2$, we reject H_1 (by assumption, we cannot accept), and the smallest rejection number from m_1 to m_2 is j.) Using (33), we have the approximation

$$E_p(N)\leqslant(\textstyle\sum_{i=0}^{C-1} i\,\alpha_i(p)+j\,\beta_j(p)+C(\beta(p)-\beta_j(p)))/p \tag{34}$$

If it is possible to accept H_1 prior to making m_2 observations, then the test strongly favors H_1. In this case, the approximation given by (32) should be adequate, since interest will center on the ASN when $p=p_1$. For tests with only two rejection numbers, it should be noted that the approximation in (34) is exact.

Example 3.3 (continued). From the acceptance tree in Figure 3.2, we have

$\alpha_0(.01)=.6426$

$\alpha_1(.01)=.1086+.0825=.1911$

$\alpha_2(.01)=.0397+.0119+.0301+.0188=.1005$

The smallest rejection number (from Table 3.1) is j=2. It is sufficient for rejection through m=19 observations. Acceptance cannot occur prior to n=44 observations, hence

$\beta_2(.01)=P_{.01}\{S_{19}\geq 2\}=.0153$

(where the probability is based on a fixed sample size of 19). As found earlier, $\beta(.01)=.0658$. Hence, we have

$$E_{.01}(S_N)\leq (0).6426+(1).1911+(2).1005+(2).0153$$
$$+(3)(.0658-.0153)=.5742$$

and

$$E_{.01}(N)\leq .5742/.01=57.42$$

Proceeding in the same manner for p=.05, we find

$$E_{.05}(N)\leq 46.44$$

(Since there are only two rejection numbers, these results are actually exact.) Had we used the approximation in (29) we would have obtained 58.95 and 51.34, respectively. Note that the relative error is much larger when p=.05.

To carry out the truncated SPRT, the auditor must draw a sample of 94 items from the perpetual inventory listing. He audits these items sequentially in the order selected from the sampling frame. For each error observed, he increments the test statistic S_n by one. The test terminates when $s_n=a_n$ (accept) or $s_n=r_n$ (reject) or n=94. If the latter occurs, H_1 is accepted if there are no more than 2 sample errors.

We will pause briefly to compare the N-P and sequential tests of Examples 3.1 and 3.3. The principal results are:

	fixed sample size	sequential
level	.069	.066
power	.855	.808
$E_{.01}(N)$	94	57
$E_{.05}(N)$	94	46

Doubtless, if p=.01, the sequential test is superior to the N-P test, since we face, on average, lower decision risk <u>and</u> lower sampling cost. If p=.05, the situation is not clear. ASN has decreased even more than under H_1, but we have lost a considerable amount of power to detect p=.05. The classical model does not allow us to assess this tradeoff explicitly, and, so, we are unable to say which test is "better."

It is, of course, possible that, regardless of savings in sampling cost, the increase in type II risk in the sequential test is unacceptable to the auditor. In such a case, we have two possible approaches. The auditor may respecify desired risks and recalculate the sequential test, continuing until an acceptable test is found. In Example 3.3, for instance, desired risks were initially set at .10 and .15 for type I and II errors respectively. The auditor could try, say, .12 and .12, in light of the initial results. In this approach, the methodology of this section should be viewed as an iterative procedure designed to produce an acceptable, not necessarily optimal, sequential test.

The alternative approach is to relax the restriction to SPRT-type acceptance/rejection regions. We expand the class of procedures considered to include all those truncated at n* in accordance with the N-P rule. In this (very large) class we search for a "best" or, at least, an acceptable procedure. While this approach is conceptually appealing, it is fraught with practical difficulties. In the classical paradigm, the very definition of "best" is problematic for sequential procedures. However, granting that a reasonable definition is available (as is the case in the Bayesian framework discussed in section 3.4), implementation is contingent on the discovery of an efficient search algorithm. (Whether such an algorithm exists depends on the theoretical question of the existence/uniqueness of a "best" test.) We will return to this question after discussing Bayesian sequential procedures in section 3.4 below.

3.3 Bayesian Fixed Sample Size Acceptance Sampling

There are two principal objections to the optimality of N-P tests: (i) losses from decision errors and the cost of sampling are not incorporated in the analysis, and (ii) prior information (if any) as to the relative likelihood of the hypotheses is suppressed. Statistical decision theory (Wald (1950)) attempts to rectify the former omission, and Bayesian decision theory attempts to incorporate the latter. My presentation follows Berger (1980) for the most part.

Prior information may be incorporated via Bayes theorem if such information is summarized as a probability distribution. We will adopt the simplest approach to prior information under the problem given by (1). Our prior (distribution) is of the form

$$g(p_1)=g_1, \quad 0<g_1<1 \qquad (35)$$
$$g(p_2)=g_2=1-g_1$$

Thus $g(p)$ is a frequency function, placing all its mass at two points in the parameter space.

Following Wald (1950), we assume the existence of a loss function. Further, we assume that it is additive in decision error loss and sampling cost. In the testing framework, a natural loss function has the following form: apart from sampling cost, there is no loss for correct decisions, and losses for incorrect decisions may vary by type of error but are otherwise constant. We also assume that sampling cost is proportional to sample size. More than this, we take the constant of proportionality to be one. Thus, losses will be measured in unit sampling costs (USC). Alternatively, a USC may be interpreted as average audit time per sample item. Under these assumptions, our loss function is

$$L(p,a,n)=L(p,a)+n \qquad (36)$$

where the decision error loss is of the form

$$L(p_i,a_j)=\begin{cases}0 & \text{if } i=j\\ K_{ij} & \text{if } i\neq j\end{cases} \qquad (37)$$

Prior information summarized in a probability distribution $g(p)$ is incorporated with sample information as reflected in the likelihood function $f^n(x;p)$ by means of Bayes theorem to yield the posterior distribution $g^n(p;x)$ as follows:

$$g^n(p;x)=g(p)f^n(x;p)/m^n(x) \qquad (38)$$

where $m^n(x)$ is the marginal distribution (i.e. unconditional on p) of $X=(X_1,\ldots,X_n)$. In our case, given the discrete prior (35), (38) may be written

$$\begin{aligned} g^n(p_i;x) &= g(p_i)f^n(x;p_i)/\textstyle\sum_{j=1}^{2} g(p_j)f^n(x;p_j) \\ &= g_i f^n(x;p_i)/\textstyle\sum_{j=1}^{2} g_j f^n(x;p_j) \end{aligned} \tag{39}$$

for $i=1,2$.

Just as in the N-P framework, choice of decision rule in the Bayesian setting involves minimizing risk. But risk is now defined as expected loss. We temporarily assume that sample size is fixed, hence sampling cost is irrelevant in the choice of decision rule. I use the term "decision risk" to mean risk exclusive of sampling cost. The decision risk of a rule d^n, where $n>0$ is the fixed sample size, is defined as the expected loss from using d^n given p:

$$R(p,d^n)=E_p L(p,d^n(X)) \tag{40}$$

For our discrete parameter space, this may be written

$$R(p_i,d^n)=E_{p_i} L(p_i,d^n(X)) \quad \text{for } i=1,2 \tag{41}$$

The Bayes decision risk of d^n is defined as the decision risk weighted by one's prior beliefs as to p:

$$r(g,d^n)=E_g R(p,d^n)=E_g E_p L(p,d^n(X)) \tag{42}$$

In our case, this weighting is simply the sum over the discrete prior (35). The Bayes principle simply states that, in a given class $\mathcal{D}^n$ of decision rules, a rule with minimum Bayes decision risk should be used. That is, let

$$r(g)=\inf_{d^n \in \mathcal{D}^n} r(g,d^n) \tag{43}$$

If a decision rule with risk $r(g)$ exists, it is called a Bayes rule. (Bayes rules are not necessarily unique.)

To find a Bayes rule more explicitly, we rewrite the righthand side of (42) as follows:

$$\begin{aligned} E_g E_p L(p,d^n(X)) &= \sum_p \left\{ \sum_x L(p,d^n(x)) f^n(x;p) \right\} g(p) \\ &= \sum_p \left\{ \sum_x L(p,d^n(x)) g^n(p;x) m^n(x)/g(p) \right\} g(p) \\ &= \sum_x \left\{ \sum_p L(p,d^n(x)) g^n(p;x) \right\} m^n(x) \\ &= E_m E_{g;x} L(p,d^n(X)) \end{aligned} \tag{44}$$

$E_{g;x}L(p,d^n(X))$ is called the posterior decision risk of d^n, since, if we have already obtained sample information, we should take the action that minimizes this risk. Thus, we can find a Bayes rule by treating x as fixed and comparing the expected losses of the (two) possible actions. For a_1, we have

$$\begin{aligned} E_{g;x}L(p,a_1) &= \sum_{i=1}^{2} L(p_i,a_1) g^n(p_i;x) \\ &= L(p_1,a_1) g^n(p_1;x) + L(p_2,a_1) g^n(p_2;x) \\ &= 0 + K_{21} g^n(p_2;x) = K_{21} g^n(p_2;x) \end{aligned} \tag{45}$$

Similarly, for a_2, we find

$$E_{g;x}L(p,a_2) = K_{12} g^n(p_1;x) \tag{46}$$

Hence, a_1 is the Bayes action if $K_{21}g^n(p_2;x) < K_{12}g^n(p_1;x)$. Substituting for the posterior from (39), we can rewrite this as

$$f^n(x;p_2)/f^n(x;p_1) < K_{12}g_1/K_{21}g_2 \tag{47}$$

The lefthand side of (47) is simply the LR, so the Bayes rule is an LR test:

$$d^n(x) = \begin{cases} a_1 & \text{if } l^n(x,p_1,p_2) < D \\ a_2 & \text{otherwise} \end{cases} \tag{48}$$

where $D = K_{12}g_1/K_{21}g_2$.

We proceed to the more difficult question of finding an optimal fixed sample size n^*. The decision rule in (48) holds regardless of sample size, provided at least one observation is made, since no restrictions other than $n > 0$ were placed on n in deriving the rule. The optimal fixed sample size n^* is that n which minimizes overall risk:

$$r(g,d^n)=E_gE_pL(p,d^n(X),n)=E_gE_pL(p,d^n(X))+n \tag{49}$$

given our assumptions with regard to the loss function. Since the Bayes decision risk of d^n--the first term in the righthand side of (49)--is typically decreasing in n, and the sampling cost (here, simply n) is clearly increasing in n, the overall risk is typically strictly convex in n. Hence, there exists a unique n^* minimizing overall risk. The standard calculus approach to finding this minimum is to treat (49) as being continuous in n, differentiate, and set equal to zero. But this method often will fail to yield a closed-form result. Either an approximation to (49) may be found or numerical methods used.

To find n^*, we must specify (49) in terms of our problem. By (48), we take action a_1 if the LR is less than some constant D and take action a_2 otherwise. Specifying (49) from the inside out, we have

$$\begin{aligned} E_pL(p,d^n(X)) &= \begin{cases} L(p_1,a_2)P_{p_1}\{d^n(X)=a_2\} & \text{if } p=p_1 \\ L(p_2,a_1)P_{p_2}\{d^n(X)=a_1\} & \text{if } p=p_2 \end{cases} \\ &= \begin{cases} K_{12}P_{p_1}\{l^n(X,p_1,p_2)\geq D\} & \text{if } p=p_1 \\ K_{21}P_{p_2}\{l^n(X,p_1,p_2)<D\} & \text{if } p=p_2 \end{cases} \\ &= \begin{cases} K_{12}(1-\alpha^n(p_1)) & \text{if } p=p_1 \\ K_{21}\alpha^n(p_2) & \text{if } p=p_2 \end{cases} \end{aligned} \tag{50}$$

where $\alpha^n(p)$ is the OC function (5) of d^n. Using the discrete prior (35),

$$E_g E_p L(p,d^n(X)) = K_{12}(1-\alpha^n(p_1))g_1 + K_{21}\alpha^n(p_2)g_2 \qquad (51)$$

And for the overall risk of decision rule d^n, we add sampling cost:

$$r(g,d^n) = K_{12}(1-\alpha^n(p_1))g_1 + K_{21}\alpha^n(p_2)g_2 + n \qquad (52)$$

The OC function is determined by the sampling distribution of the LR, which is usually not tabled. However, we can minimize (52) by numerical methods, working out the sampling distribution at several points. We use this method in the next example.

Example 3.4 (continued from Example 3.1). The auditor specifies decision losses of $K_{12}=600$ and $K_{21}=1500$, measured in USCs. Thus a type II error is deemed more than twice as costly as a type I error. The auditor also specifies the following prior: $g_1=.8$ and $g_2=.2$. Thus, $D=600(.8)/1500(.2)=1.6$. The remaining elements of the problem are unchanged from Example 3.1. The overall risk is

$$r(g,d^n) = (480)P_{.01}\{l^n(X,.01,.05) \geq 1.6\}$$
$$+(300)P_{.05}\{l^n(X,.01,.05) < 1.6\} + n$$

To evaluate $P_{.01}\{l^n(X,.01,.05) \geq 1.6\}$, select an n, find the smallest C such that $l^n(x,.01,.05) \geq 1.6$, and find $P_{.01}\{S_n \geq C\}$. The probability under .05 is similarly found to be $P_{.05}\{S_n < C\}$ $=1-P_{.05}\{S_n \geq C\}$. To illustrate how this C is found, we use the Poisson LR, since S_n is approximately Poisson with q=np. We have, then,

$$(e^{-q_2}q_2^C/C!)/(e^{-q_1}q_1^C/C!) = e^{q_1-q_2}(q_2/q_1)^C \geq D$$

Since all terms are positive, this is equivalent to

$$q_1-q_2+(C)\log(q_2/q_1) \gtrless \log D$$

That is,

$$C \gtrless (\log D + q_2-q_1)/\log(q_2/q_1)$$

For D=1.6 and n=100, this gives

$$C \gtrless (\log 1.6 + 5 - 1)/\log 5 = 2.78$$

The smallest integral value, then, is C=3. $P_{.01}\{S_{100} \geq 3\}$ and $P_{.05}\{S_{100} < 3\}$ can be found (under q_1=1.0 and q_2=5.0) in the Poisson tables in Appendix B. They are, respectively, .080 and .125. Results of a search using various n are tabulated below:

n	100	120	60	80	90
q_1	1.00	1.20	.60	.80	.90
q_2	5.00	6.00	3.00	4.00	4.50
C	3	4	2	3	3
$\beta(.01)$	.080	.034	.122	.047	.063
$\alpha(.05)$	.125	.151	.199	.238	.174
$r(g,d^n)$	176	182	178	174	172

Thus, n* is about 90 with C=3 and Bayes risk of 172 USCs.

It is clear from Example 3.4 that, just as in the N-P case, the test in (48) may be restated, using the sufficient statistic $T_n(X)=S_n$, as

$$d^n(x)=\begin{cases} a_1 & \text{if } T_n(x) < C \\ a_2 & \text{otherwise} \end{cases} \qquad (53)$$

where C is the critical value of the test.

The minimization carried out in Example 3.4 is tedious, but it is, of course, amenable to computer solution, and an algorithm to find n^* and C is provided in Appendix D. Using this algorithm, we find, for this example, $n^*=88$ and $C=3$ with $r(g,d^{n^*})=r(g)=172.15$.

The sample size of 88 obtained in the foregoing example is not much different from that of the N-P test ($n^*=95$). But the Bayesian approach provides a considerably altered perspective. If the losses of 600 and 1500 are approximately correct, a sample size of 95 with critical value of 3 implies a strong disposition for H_1. This conclusion cannot be drawn from the type I risk of .07 and type II risk of .15 found in Example 3.1, although it would appear that H_1 is considered more likely. In the N-P test, the unstated prior and loss offset such that the auditor set desired risks at .10 and .15. While both of these factors--prior distribution and loss function--are dramatic simplifications of the decision-making process, the Bayesian construction is significantly richer in context detail.

3.4 Bayesian Sequential Acceptance Sampling

Conceptually, the Bayesian approach to sequential analysis is reasonably clear: at each stage of sampling (or "time") n, we compute the Bayes risk of an immediate decision; we then compute the Bayes risk at time n+1, n+2,...; if the Bayes risk of an immediate decision is no greater than the Bayes risk of going on (the minimum of the Bayes risks at times n+1, n+2, ...), then we should stop and make a decision. Unfortunately, when all possible sample sizes are admitted (i.e. an infinite horizon), the computations typically become unmanageable.

Note that the problem is usually not the terminal decision rule but the stopping rule. Once we have stopped sampling, the Bayes rule for the appropriate fixed sample size test is followed. The problem has been solved by limiting consideration to truncated procedures (those for which a maximum number of observations is allowed). Under certain conditions, the Bayes sequential procedure *is* a truncated procedure, and nothing is lost by this restriction. But, in general, the class of all truncated procedures is still too large. More restricted classes of procedures have been proposed, e.g. m-step look ahead, inner look ahead, and fixed sample size look ahead. Although Bayes procedures can often be found within these classes, they typically require considerable computation at each stage of sampling and, so, are not well-suited to audit situations.

The SPRT, with its constant bounds, is appropriate to audit situations, and it is possible to "rationalize" the classical SPRT to obtain the Bayes SPRT--the minimum risk SPRT. However, there are two drawbacks to implementing the Bayes SPRT for audit uses: (i) derivation of the bounds is rather complicated, and (ii) a reasonable truncation rule is not obvious (for example, there is no longer a necessary connection between the optimal fixed sample size procedure and the Bayes SPRT, and the *expected* sample size of the Bayes SPRT may well exceed the optimal fixed sample size). As in the classical case, we seek a sequential procedure that is tied to the optimal fixed sample size procedure. A Bayesian SPRT truncated at the optimal fixed sample size should be well-suited to audit needs, and I will propose such an SPRT below.

For various reasons (budgeting, cost to access the sampling frame, etc.), we decide on a fixed sample size procedure and select the optimal fixed sample size, n^*, from the sampling frame. We are, then, in effect, committed to the Bayes risk, r^*, of this procedure. But the observations will be made sequentially. If at any time $n < n^*$ the Bayes risk of an immediate decision does not exceed r^*, we should stop and make a decision. Otherwise, we continue sampling, eventually stopping at n^* if no decision has been made earlier.

We have already found the Bayes risk $r(g)=r^*$ and the sample size n^* of the optimal fixed sample size procedure. By the equivalence in (44), r^* may also be called the expected posterior Bayes risk at time n^*. We now need the Bayes risk of an immediate decision at time $n=1,2,\ldots$ (Assuming that $n^*>0$, the Bayes risk of a decision at $n=0$ will exceed r^*.) This risk is the posterior Bayes risk at time n. Let $g^n = g^n(p;x)$, the posterior at time n. Then the posterior risk of taking action a at time n is

$$r_0(g^n,a)=E_{g;x}L(p,a,n) \tag{54}$$

and the posterior Bayes risk of an immediate decision at time n is the minimum posterior risk:

$$r_0(g^n)=\inf_{a\in\mathcal{A}} E_{g;x}L(p,a,n) \tag{55}$$

Our rule is, then, to stop at the first n such that

$$r_0(g^n) \leq r^* \tag{56}$$

To implement this rule, we need to specify $r_0(g^n)$ in terms of our loss function in (36) and (37):

$$\begin{aligned} r_0(g^n, a_1) &= K_{21} g^n(p_2;x) + n \\ r_0(g^n, a_2) &= K_{12} g^n(p_1;x) + n \end{aligned} \tag{57}$$

Just as we found in (47) and (48), a_1 is the optimal action if the LR is less than $D = K_{12}g_1/K_{21}g_2$. So,

$$r_0(g^n) = \begin{cases} K_{21}g^n(p_2;x)+n & \text{if } l^n(x,p_1,p_2) < D \\ K_{12}g^n(p_1;x)+n & \text{otherwise} \end{cases} \tag{58}$$

In the development, we will rewrite the posterior, using (39), as

$$g^n(p_i;x) = \frac{g_i f^n(x;p_i)}{g_1 f^n(x;p_1) + g_2 f^n(x;p_2)} \quad \text{for } i=1,2 \tag{59}$$

By (58) and the stopping rule in (56), we take action a_1 at time $n < n^*$ if $l^n(x,p_1,p_2) < D$ and if

$$r_0(g^n) = K_{21} \frac{g_2 f^n(x;p_2)}{g_1 f^n(x;p_1) + g_2 f^n(x;p_2)} + n \leq r^* \tag{60}$$

That is, if

$$\frac{f^n(x;p_2)}{f^n(x;p_1)} \leq \frac{g_1}{g_2} \left[\frac{r^* - n}{K_{21} - r^* + n} \right] = A \tag{61}$$

Now the lefthand side of (61) is just the LR, and we will show that $A < D$, hence we may discard the condition $l^n(x,p_1,p_2) < D$.

Note that $r^* < \min(g_1K_{12}, g_2K_{21})$ if $n^* > 0$, otherwise the risk of going on equals or exceeds the risk of an immediate decision, and no sampling would be done. We first assume $g_1K_{12} < g_2K_{21}$, then, by the definition of g_2,

$$g_1K_{12} < (1-g_1)K_{21} \tag{62}$$

or

$$g_1 < K_{21}/(K_{12}+K_{21}) \tag{63}$$

so

$$g_1K_{12} < K_{12}K_{21}/(K_{12}+K_{21}) \tag{64}$$

since $K_{12}, K_{21} > 0$. And, since $0 < r^*-n < r^* < g_1K_{12}$,

$$r^*-n < K_{12}K_{21}/(K_{12}+K_{21}) \tag{65}$$

or

$$(r^*-n)(1+(K_{12}/K_{21})) < K_{12} \tag{66}$$

and

$$r^*-n < K_{12}-(K_{12}/K_{21})(r^*-n) \tag{67}$$

which we may factor into

$$r^*-n < (K_{12}/K_{21})(K_{21}-r^*+n) \tag{68}$$

Since $r^* < K_{21}$, $K_{21}-r^*+n > 0$ and

$$(r^*-n)/(K_{21}-r^*+n) < K_{12}/K_{21} \tag{69}$$

Multiplying both sides of (69) by $g_1/g_2 > 0$ gives the result. The same result obtains if $g_1K_{12} > g_2K_{21}$ (merely substitute g_2 for g_1 and interchange K_{12} and K_{21} in the first few steps). It is also easy to see that $g_1K_{12}=g_2K_{21}$ leads to the same result.

We now resume development of the test. We take action a_2 at $n < n^*$ if $l^n(x,p_1,p_2) \gtrless D$ and if

$$r_0(g^n)=K_{12}\frac{g_1f^n(x;p_1)}{g_1f^n(x;p_1)+g_2f^n(x;p_2)} + n \lesseqgtr r^* \tag{70}$$

That is, if

$$\frac{f^n(x;p_2)}{f^n(x;p_1)} \geq \frac{g_1}{g_2}\left[\frac{K_{12}-r^*+n}{r^*-n}\right] = B \tag{71}$$

The lefthand side of (71) is again the LR, and it may be shown, in a manner analogous with that of the proof that $A < D$, that $B > D$, hence we can discard the condition $l^n(x,p_1,p_2) \geq D$.

A and B are not the bounds of an SPRT since they widen slightly at each sampling stage, reflecting the decreased opportunity to save sampling cost in making an immediate decision. However, if we treat the sampling cost as foregone, we replace n with n* and obtain the constant bounds A' and B':

$$A' = \frac{g_1}{g_2}\left[\frac{r^*-n^*}{K_{21}-r^*+n^*}\right] \qquad B' = \frac{g_1}{g_2}\left[\frac{K_{12}-r^*+n^*}{r^*-n^*}\right] \tag{72}$$

This yields a more conservative sequential procedure, since $A' < A$ and $B' > B$ for all $n < n^*$. The stopping rule in (56), then, pertains to decision risk only, not overall risk. We may also justify the use of A' and B' on more substantive grounds. Assume that there is a significant cost attached to accessing the sampling frame and selecting the sample. To obtain a constant unit sampling cost, this fixed cost must be allocated on the basis of a known sample size, presumably n*. In this case, use of the variable bounds A and B would understate the risk faced.

We have arrived at the following sequential procedure:

$$d^n(x) = \begin{cases} a_1 & \text{if } l^n(x,p_1,p_2) \leq A' \\ a_2 & \text{if } l^n(x,p_1,p_2) \geq B' \\ a_3 & \text{otherwise} \end{cases} \tag{73}$$

Just as in the classical case, we can restate this test using the sufficient statistic $T_n(X)=S_n$ and replacing the bounds A' and B' with acceptance/rejection numbers determined by the relation in (15). Again, we truncate the test at $n=n^*$ and follow the optimal fixed sample size rule in (53) at this time. This leads to the following decision rule:

$$d(x)=\begin{cases} d^n(x) & \text{if } n < n^* \\ d^{n^*}(x) & \text{otherwise} \end{cases} \tag{74}$$

where

$$d^n(x)=\begin{cases} a_1 & \text{if } T_n(x) \leqslant a_n \\ a_2 & \text{if } T_n(x) \geqslant r_n \\ a_3 & \text{otherwise} \end{cases} \tag{75}$$

and

$$d^{n^*}(x)=\begin{cases} a_1 & \text{if } T_{n^*}(x) < C \\ a_2 & \text{otherwise} \end{cases} \tag{76}$$

where n^* is the sample size and C is the critical value of the Bayesian optimal fixed sample size procedure.

Note that (74) is a truncated SPRT. Hence we may compute the OC function using (21) and approximate the ASN function using (32) or (34).

We have found the posterior Bayes risk (the Bayes risk given x) of d. To obtain the Bayes risk, we must average the posterior Bayes risk over all possible x, i.e. take the expectation with respect to $m^n(x)$, the unconditional distribution of x. It is easier to reverse the order of expectations, finding the expected loss with respect to $f^n(x;p)$, the conditional (on p) distribution of x, and then averaging over p, i.e. taking the expectation with respect to the prior g(p). By (44), these two methods are equivalent.

The risk of d depends on the expected sample size as well as the decision loss:

$$\begin{aligned} R(p,d) &= E_p L(p,d(X),N) \\ &= E_p(L(p,d(X))+N) \\ &= E_p L(p,d(X))+E_p(N) \end{aligned} \tag{77}$$

given the form of loss function specified in (36). Hence,

$$\begin{aligned} R(p_1,d) &= K_{12}(1-\alpha(p_1))+E_{p_1}(N) \\ R(p_2,d) &= K_{21}\alpha(p_2)+E_{p_2}(N) \end{aligned} \tag{78}$$

where we have simplified notation by using the OC function. The Bayes risk of d is, then,

$$\begin{aligned} r(g,d) &= E_g R(p,d) \\ &= g_1 R(p_1,d)+g_2 R(p_2,d) \\ &= g_1(K_{12}(1-\alpha(p_1))+E_{p_1}(N))+g_2(K_{21}\alpha(p_2)+E_{p_2}(N)) \end{aligned} \tag{79}$$

<u>Example 3.5 (continued from Example 3.1)</u>. From the discussion just following Example 3.4, we have $r^*=172$, $n^*=88$, and $C=3$. The prior and loss are unchanged and are not restated here. Substituting in (72) gives

$A'=(.8/.2)(172-88)/(1500-172+88)=0.237$

$B'=(.8/.2)(600-172+88)/(172-88)=24.571$

Using the relation given by (15) and keeping in mind that $C=3$, the acceptance/rejection numbers are

n	a_n
35	0
75	1
88	2

n	r_n
$2 \leq n \leq 87$	2
88	3

The OC function is found, as before, using Aroian's (1968) method (21):

$$\alpha(.01)=0.954$$

$$\alpha(.05)=0.266$$

And, using the ASN approximation in (34),

$$E_{.01}(N)=47.28$$

$$E_{.05}(N)=47.20$$

and, in this case, is exact. The Bayes risk of d is

$$r(g,d)=.8[(1-.954)600+47.28]+.2[(.266)1500+47.20]$$

$$=.8(74.88)+.2(446.20)$$

$$=149.14$$

It should be noted that the Bayes risk of the truncated SPRT is less than that of the optimal fixed sample size procedure ($r^*=172$). This was, of course, the intention in deriving the bounds A' and B' for the sequential procedure. But the decrease in Bayes risk did not result from symmetric decreases in risk. We compare the fixed sample size and sequential procedures in the table below:

	fixed sample size	sequential
$R(.01,\cdot)$	123.76	74.88
$R(.05,\cdot)$	365.71	446.20
$E_{.01}(N)$	88	47
$E_{.05}(N)$	88	47

Here we have a result quite similar to the classical case in Example 3.4: one risk increased while the other decreased and ASN decreased in both cases. But here, as opposed to the classical situation, we have a criterion by which to judge this tradeoff: if we accept Bayes risk as the appropriate choice criterion for tests, the truncated SPRT is superior to the fixed sample size test. However, no claim is made that the truncated SPRT is optimal among the class of all procedures truncated at n^* using the optimal fixed sample size rule at that time. Conceptually, we would prefer to find a Bayes rule in this extended class of procedures. While the definition of a "best" procedure in this class is not problematic from a Bayesian perspective, the other objection raised at the end of section 3.2 still holds: finding this procedure is contingent on the existence and discovery of an efficient search algorithm.

3.5 Summary

The models presented in sections 3.1 through 3.4 are acceptance sampling models in which the sampling unit can be classified as an error or nonerror. They are isomorphic to quality control testing models in which the sampling unit can be classified as defective or effective. By analogy with the quality control situation, I refer to the models of this chapter as physical unit acceptance sampling (PUAS) models. (The motivation for this term will, it is hoped, become apparent in the following chapter.) Classical fixed sample size PUAS will refer to the test in (10), classical sequential PUAS will refer to the test in (17), Bayesian fixed sample size PUAS will refer to the test in (53), and, lastly, Bayesian sequential PUAS will mean the test in (74).

CHAPTER 4

A STATISTICAL SUBSTANTIVE TESTING MODEL: MONETARY UNIT ACCEPTANCE SAMPLING

The models presented in Chapter 3 were restricted to situations in which the auditor could classify the observations as errors or nonerrors. We have called these models, collectively, physical unit acceptance sampling (PUAS). But there are many audit situations for which a finer classification of the observations is needed. Notably, this occurs in direct tests of balances and transactions (i.e. substantive tests) where the natural measure of error is monetary, and the <u>degree</u> of error of each observation becomes critical. In the subsequent development, we will extend the PUAS models for use in substantive tests. I will refer to the proposed models, collectively, as monetary unit acceptance sampling (MUAS). The propriety of this name will become evident in the development. Except as noted in the sequel, we restrict the situation to a test for overstatement in an asset balance (e.g. inventory). Overstatement means that the book (or recorded) value exceeds the true value. The general hypotheses are, then,

$$\begin{aligned} &H_1\text{: the balance is correct} \\ &H_2\text{: the balance is overstated} \end{aligned} \tag{80}$$

The auditor, however, is willing to tolerate some degree of overstatement before deciding against H_1. In substantive testing, a tolerable degree of overstatement is termed <u>immaterial</u>. An intolerable degree is, then, <u>material</u>. Materiality as used here refers to the working assumption that some degree of overstatement in an asset balance has no effect on the decisions of a reasonably prudent user of the financial statements containing that balance. But some greater degree of overstatement will affect the decisions of such a user.

Materiality may be expressed in absolute terms, but it is naturally expressed as a percentage of the book value of the balance in question. Thus, for example, the auditor may expect an immaterial rate of overstatement of p=.01. And he may decide that the lowest material rate of overstatement is p=.05. In such a case, the general hypotheses in (80) may be operationalized as

$$\begin{aligned} &H_1\text{: } p=.01 \\ &H_2\text{: } p=.05 \end{aligned} \tag{81}$$

(I will consistently refer to materiality in percentage terms. And, to simplify usage, I will refer to the rate of overstatement as the error rate. This usage will be justified on its own merits when monetary unit sampling is introduced below.)

Given the hypotheses in (81), the PUAS models appear to be applicable. However, the natural sampling units of a balance are typically subunits of varying book value (e.g. the items or part numbers in an inventory balance). A classification of these subunits into "materially correct" and "materially overstated" is not sufficient for the decision required in (81). While it is true that, if no subunit is materially overstated, the balance is not materially overstated, and, if every subunit is materially overstated, the balance is materially overstated, the necessary relationship extends no further. The overstatement of just one subunit may be sufficient for material overstatement of the balance, provided this subunit is large enough (in book value) relative to the balance as a whole.

The traditional auditing approach to the problem in (81) has been the use of various survey sampling techniques to estimate the true value or, equivalently, the true error rate. This estimate is then compared to the book value by means of a confidence interval. These techniques, grounded in finite population sampling theory, are essentially nonparametric, relying on the large-sample behavior of the estimator to construct the confidence interval. (See Roberts (1978) for applications of this approach.) However, studies by Kaplan (1973b) and Neter and Loebbecke (1975,1977) provided evidence actual confidence levels could be significantly lower than nominal confidence levels for typical audit sample sizes in tests on typical accounting populations.

Another approach is a natural extension of the binomial model for compliance tests. Following on the notion that a finer classification of the observations is needed in substantive testing, Neter et al. (1978) proposed a multinomial model. While conceptually appealing, this model exhibits various difficulties attendant on moving from a univariate to a multivariate model. Among these are choice of test, power of the test (once chosen), and determination of necessary sample size.

An alternative, univariate, approach is based on monetary unit sampling (MUS). (For simplicity, we will refer to the monetary units in question as "dollars.") Rather than employ the natural sampling frame of subunits, MUS treats the balance as consisting of dollars. These dollars are labeled $1,2,\ldots,N$, where N is the total book value of the balance. This is an artificial sampling frame created by the auditor. It is usually created by ordering the subunits of the balance and identifying dollars $1,\ldots,N_1$ with the first subunit (where N_1 is the book value of the first subunit), identifying dollars $N_1+1,\ldots,N_1+N_2$ with the second subunit (where N_2 is the book value of the second subunit), and so forth. Other mappings are possible. The observations are now dollars, which are classified as errors ("defective" dollars) or nonerrors ("nondefective" dollars). The error rate is now simply the proportion of "defective" dollars in the balance. In such terms, the PUAS models appear applicable (i.e. each dollar becomes a physical unit).

The difficulty in applying the PUAS models lies in the determination of a defective dollar. Since clients account for subunits, not individual dollars, this will necessarily involve an audit of the subunit containing the dollar. (Thus, viewed as a method of selecting <u>subunits</u>, MUS is one form of probability proportional to size (pps) sample selection, where the measure of size is book value.) In just two cases can we be certain whether or not the dollar selected is defective: (i) the subunit containing the dollar is entirely fictitious, and (ii) the subunit containing the dollar is entirely sound. But the intermediate cases, in which the subunit containing the dollar is partially overstated, lead to an identification problem. For example, consider a dollar belonging to a subunit that is 10% overstated. The dollar selected apparently could be either one of the 10% that are defective or one of the 90% that are sound. Alternatively, our rationale in suggesting that PUAS might be applicable was grounded in the idea of defective dollars (errors) and nondefective dollars (nonerrors). Is it meaningful within the context of PUAS to speak of a <u>10% defective</u> dollar?

In section 4.1, we will restate, in somewhat altered form, the first solution proposed for this identification problem. In section 4.2, I offer an improvement on this solution and then, in section 4.3, present the results of a Monte Carlo study using the proposed MUAS models.

4.1 Conditional Randomization

The first solution to the identification problem in the case of partially overstated subunits was given by van Heerden (1961). To discuss his solution and the alternative, equivalent, solution that we call conditional randomization, we need additional notation. Recall that we now mean by "error" a defective or overstated or fictitious dollar and note that N has been redefined for use in this chapter. We will use the following notation:

for the population:

N = population size (in recorded dollars)

p = population error rate

K = total errors in the population

I = number of subunits in the population, $I \leqslant N$

for the ith subunit $(i=1,\dots,I)$:

N_i = size of the ith subunit (in recorded dollars)

p_i = error rate of the ith subunit

K_i = total errors in the ith subunit

for the sample:

n = sample size (number of dollars selected)

k = total errors in the sample

From these definitions, we have the following relations:

for the population:

$$N = \sum_{i=1}^{I} N_i$$

$$K = \sum_{i=1}^{I} K_i$$

$$K = Np \tag{82}$$

for the ith subunit:

$$K_i = N_i p_i$$

I will assume throughout that a random sample of size n is selected with replacement from a population of size N. Also, for any subunit i, p_i is known with certainty if, and only if, at least one dollar from the ith subunit is included in the sample.

We are now in a position to describe van Heerden's (1961) solution. Assume that we select the Jth dollar of the population ($1 \leq J \leq N$) and that this dollar is contained in the ith subunit ($1 \leq i \leq I$). The ith subunit contains K_i errors. If $K_i=0$ or N_i, there is no identification problem, hence I assume that $0 < K_i < N_i$. Van Heerden proposed that we identify these errors with the high-order dollars in the subunit. That is, let the ith subunit consist of dollars $M-N_i+1, M-N_i+2, \ldots, M$ ($N_i-1 \leq M \leq N$). We identify $M, M-1, \ldots, M-K_i+1$ as errors. If $M-K_i+1 \leq J \leq M$, we record an error for this observation and a nonerror otherwise.

Rather than work out the statistical implications of van Heerden's identification rule, we will consider an alternative solution based on conditional randomization. While these two solutions are probabilistically equivalent, the conditional randomization construction directly motivates the improvement offered in section 4.2.

The solution we consider consists of a conditional randomization device (crd) that records an error with conditional probability $p_i=K_i/N_i$ given that a dollar from the ith subunit has been selected, this selection having been made at random with replacement from the population of N dollars. The crd is invoked _after_ the dollar is selected and represents a second layer of randomization. Here is an example of the use of a crd. The jth dollar ($1 \leq j \leq n$) in our sample belongs to the ith subunit. We observe an error rate of $p_i=.5$ in this subunit. An appropriate crd is the toss of a fair coin, recording an error for heads and a nonerror for tails. We now examine the consequences of using conditional randomization.

Let Y_j represent the possible outcome of the crd for the jth sample dollar. More precisely, let

$$Y_j=\begin{cases}1 & \text{if the crd records an error for the jth dollar} \\ 0 & \text{otherwise}\end{cases} \tag{83}$$

for $j=1,\ldots,n$. We are interested in the distribution of the $\{Y_j\}$. Note that, since we are sampling at random with replacement, the $\{Y_j\}$ are independent, identically distributed random variables. Let Y be a random variable with the same distribution as Y_j, $j=1,\ldots,n$. And let A_i be the event that a dollar from the ith subunit is chosen and B be the event that the crd records an error. Then we have

$$\begin{aligned}P\{Y=1\} &= P\{\textstyle\bigcup_{i=1}^{I}(A_i \cap B)\} \\ &= \textstyle\sum_{i=1}^{I} P\{A_i \cap B\} \\ &= \textstyle\sum_{i=1}^{I} P\{B|A_i\}P\{A_i\} \\ &= \textstyle\sum_{i=1}^{I}(K_i/N_i)(N_i/N) \\ &= \textstyle\sum_{i=1}^{I} K_i/N \\ &= K/N=p\end{aligned} \tag{84}$$

The second step in (84) follows since the $\{A_i \cap B\}$ are pairwise disjoint events. That $P\{B|A_i\}=K_i/N_i$ follows from the definition of the crd. And $P\{A_i\}=N_i/N$ follows from the fact that we are sampling at random with replacement from a population of N dollars. It follows from (84) that

$$P\{Y=0\}=1-p \tag{85}$$

and, since Y is an indicator variable, we have immediately

$$E(Y)=P\{Y=1\}=p \tag{86}$$

with variance

$$\begin{aligned} Var(Y)&=E(Y^2)-(E(Y))^2 \\ &=p(1-p) \end{aligned} \tag{87}$$

The $\{Y_j\}$ are independent, identically distributed binomial(1,p) random variables. Hence,

$$S_n=\sum_{j=1}^{n} Y_j \sim \text{binomial}(n,p) \tag{88}$$

Thus, use of a crd conforming to our definition of such a device extends the PUAS models for use in substantive tests as characterized in (80). (To see that van Heerden's rule yields the same result, simply define B in (84) as the event that the dollar selected is defective.) It is of some importance to note that, by invoking the crd at each sampling stage n=1,2,..., the PUAS sequential plans may be implemented.

Before proceeding to discuss an improvement on this solution, we should pause to note that van Heerden's rule, or use of a crd, made available, for the first time, a parametric test of (80), with known risks under the control of the auditor and independent of any large-sample theory. It is a significant achievement in the history of audit sampling, for which van Heerden has not received due credit.

4.2 An Alternative to Conditional Randomization

There are both behavioral and statistical objections to the use of a crd. Behaviorally, there appears to be a general abhorrence of randomized rules for nontrivial decisions. While such a behavioral objection is of practical importance, there is a more substantive objection to the use of conditional randomization in the case at hand. If a crd is used, certain information is discarded. Prior to selecting a dollar from the ith subunit, the error rate p_i of that subunit is unknown. But, once we have selected a dollar that belongs to the ith subunit, p_i is known with certainty. The crd discards this information in favor of a 1 (with probability p_i) or a 0 (with probability $1-p_i$). Consider the degenerate case of a population with only one subunit ($I=1$). Here, $p_1=p$, and, after selecting one dollar, we know p with certainty. Using a crd, we will select n dollars, randomize for each, and record k errors. Unless $p=0$ or 1, use of the crd has introduced decision risk where there need be none (i.e. k/n is identically equal to p only if $p=0$ or 1). This argument suggests that we can improve on the crd by basing our decision on all the information available, i.e. all known $\{p_i\}$.

In the following construction, it will be necessary to modify the notation of section 4.1 slightly. We group together all subunits in the population with identical subunit error rates. We assume that there are $H \leq I$ distinct $\{p_i\}$. We label these q_h, $h=1,\ldots,H$. And we define I_h as the set of subscripts in $\{1,\ldots,I\}$ for which $p_i=q_h$. Then let

$$M_h = \sum_{i \in I_h} N_i$$
$$K_h = \sum_{i \in I_h} K_i \tag{89}$$

Note that $\sum_{h=1}^{H} M_h = N$ and $\sum_{h=1}^{H} K_h = K$ since the $\{I_h\}$ form a partition of $\{1,\ldots,I\}$.

Let X_j, $j=1,\ldots,n$, be the jth random subunit error rate. The $\{X_j\}$ are independent, identically distributed random variables. Let X be a random variable with the same distribution as X_j, $j=1,\ldots,n$. Then

$$P\{X=q_h\}=M_h/N \tag{90}$$

and the expected value of X is

$$\begin{aligned} E(X) &= \sum_{h=1}^{H} q_h(M_h/N) \\ &= \sum_{h=1}^{H} (K_h/M_h)(M_h/N) \\ &= \sum_{h=1}^{H} K_h/N \\ &= K/N = p \end{aligned} \tag{91}$$

with variance

$$\begin{aligned} Var(X) &= E(X^2)-(E(X))^2 \\ &= E(X^2)-p^2 \\ &= \sum_{h=1}^{H} q_h^2(M_h/N)-p^2 \end{aligned} \tag{92}$$

Since $0 \leq q_h \leq 1$ implies that $q_h^2 \leq q_h$, (92) implies that

$$Var(X) \leq p(1-p) \tag{93}$$

Let

$$S_n' = \sum_{j=1}^{n} X_j \tag{94}$$

Then we have

$$\begin{aligned} E(S_n') &= np \\ Var(S_n') &= n(Var(X)) \leq np(1-p) \end{aligned} \tag{95}$$

Thus, S_n' has the same expectation as S_n (defined in (88)) but has lower variance except in the extreme case of only two distinct $\{p_i\}$, 0 and 1. To see that S_n' achieves its maximum variance under these conditions, note that H=2 and let, say, $q_1=0$ and $q_2=1$. Substituting in (92) we have

$$\begin{aligned} Var(X) &= \sum_{h=1}^{2} q_h^2 (M_h/N) - p^2 \\ &= (M_2/N) - p^2 \\ &= (K_2/N) - p^2 \\ &= p - p^2 \\ &= p(1-p) \quad \text{if } p_i=0 \text{ or } 1 \text{ for } i=1,\dots,I \end{aligned} \tag{96}$$

Under these conditions, $S_n' \sim$ binomial(n,p), i.e. $S_n'=S_n$. Furthermore, S_n' achieves its minimum variance if all $\{p_i\}$ are equal, that is,

$$Var(X)=0 \quad \text{if } p_i=p \text{ for } i=1,\dots,I \tag{97}$$

4.2.1 Fixed Sample Size MUAS

These results suggest that we obtain a conservative fixed sample size test as follows. We derive necessary sample size n and critical value C based on the conditional randomization model, that is, we plan the test based on S_n. In conducting the test, however, we substitute $T_n'(X)=S_n'$ for $T_n(X)=S_n$. This is not quite as straightforward as it may appear. Since S_n' is a continuous random variable (except in the degenerate cases of (96) and (97)) and C is integer-valued, the decision rule

$$d''(x)=\begin{cases} a_1 & \text{if } s_n'<C \\ a_2 & \text{otherwise} \end{cases} \tag{98}$$

(where $s_n'=T_n'(x)$) is equivalent to the rule

$$d''(x)=\begin{cases} a_1 & \text{if } [s_n']<C \\ a_2 & \text{otherwise} \end{cases} \tag{99}$$

where $[w]$ is the largest integer $\leq w$. Since $[S_n']\leq S_n'$, $E([S_n'])\leq E(S_n')=np$, and we introduce systematic bias in our statistic. The appropriate continuity correction is .5, with the rule

$$d'(x)=\begin{cases} a_1 & \text{if } s_n'\leq C-.5 \\ a_2 & \text{otherwise} \end{cases} \tag{100}$$

or, equivalently,

$$d'(x)=\begin{cases} a_1 & \text{if } [s_n'+.5]<C \\ a_2 & \text{otherwise} \end{cases} \tag{101}$$

The continuity correction is usually associated with the normal approximation to the binomial distribution (see Bickel and Doksum (1977) p. 464). Here, we are discretizing S_n' and attempting to preserve (approximately) its unbiasedness. (If the density of S_n' is constant on each interval $(k,k+1)$, $k=0,\ldots,n-1$, then $E([S_n'+.5])=E(S_n')=np$. In this case, $E([S_n'])$ $=np-.5$.)

It may appear simpler to work with S_n' directly rather than substitute it in a test based on S_n. The difficulty is that the exact distribution of S_n' is not known. As the sum of independent, identically distributed random variables, the central limit theorem gives an approximate distribution. The quality of the approximation will depend on the population tested. However, we can use the normal approxiamtion to compare the nominal decision risks (of d(x) using S_n) and the true decision risks (of d'(x) using S_n').

In the following, Φ denotes the normal(0,1) distribution function, and z(b), $0<b<1$, denotes the value such that $\Phi(z(b))=b$. We first consider type I risk:

$$P_{H_1}\{S_n \geq C\}=P_{H_1}\left\{\frac{S_n-np_1}{\sqrt{Var_{p_1}(S_n)}} \geq \frac{C-np_1-.5}{\sqrt{Var_{p_1}(S_n)}}\right\}$$

$$\doteq 1-\Phi\left(\frac{C-np_1-.5}{\sqrt{Var_{p_1}(S_n)}}\right)=1-\Phi(z(1-\alpha))=\alpha \qquad (102)$$

(See, for example, Bickel and Doksum (1977) p. 170 for use of the continuity correction in this situation.)

$$P_{H_1}\{S_n' \geq C-.5\}=P_{H_1}\left\{\frac{S_n'-np_1}{\sqrt{Var_{p_1}(S_n')}} \geq \frac{C-np_1-.5}{\sqrt{Var_{p_1}(S_n')}}\right\}$$

$$\doteq 1-\Phi\left(\frac{C-np_1-.5}{\sqrt{Var_{p_1}(S_n')}}\right)=1-\Phi(z(1-\alpha'))=\alpha' \qquad (103)$$

Since $Var_{p_1}(S_n') \leq Var_{p_1}(S_n)$, and assuming $\alpha < .5$, $0 < z(1-\alpha) \leq z(1-\alpha')$, hence $\alpha' \leq \alpha$. Similarly,

$$P_{H_2}\{S_n < C\} \doteq \Phi\left(\frac{C-np_2-.5}{\sqrt{Var_{p_2}(S_n)}}\right) = \Phi(z(\beta))=\beta \qquad (104)$$

$$P_{H_2}\{S_n' < C-.5\} \doteq \Phi\left(\frac{C-np_2-.5}{\sqrt{Var_{p_2}(S_n')}}\right) = \Phi(z(\beta'))=\beta' \qquad (105)$$

Then, since $Var_{p_2}(S_n') \leq Var_{p_2}(S_n)$, and assuming $\beta < .5$, $z(\beta') \leq z(\beta) < 0$, hence $\beta' \leq \beta$.

If the normal approximations hold, these results establish the risk reduction claimed for $d'(x)$. We will call the test based on S_n' in (101) fixed sample size monetary unit acceptance sampling (MUAS). In the following section, we will extend MUAS to sequential sampling.

4.2.2 Sequential MUAS

The extension of fixed sample size MUAS to sequential testing is quite straightforward. At each sampling stage n, we have, in the sequential PUAS models, integer-valued acceptance and rejection numbers (a_n and r_n, respectively) such that we reject H_1 if $s_n \geq r_n$ and accept H_1 if $s_n \leq a_n$ and continue sampling otherwise (up to n^*). In replacing S_n with S_n', we make the following continuity corrections: reject H_1 if $s_n' \geq r_n - .5$ and accept H_1 if $s_n' \leq a_n + .5$ and continue sampling otherwise. Now, $s_n' \geq r_n - .5$ if and only if $[s_n' + .5] \geq r_n$. And $s_n' < a_n + .5$ if and only if $[s_n' + .5] \leq a_n$. (We are entitled to ignore the possibility that $s_n' = a_n + .5$.) Hence, at each stage n, we substitute $[S_n' + .5]$ for S_n as the test statistic. We have, then, the following decision rule for sequential MUAS:

$$d(x) = \begin{cases} d^n(x) & \text{if } n < n^* \\ d^{n^*}(x) & \text{otherwise} \end{cases} \tag{106}$$

where

$$d^n(x) = \begin{cases} a_1 & \text{if } [s_n' + .5] \leq a_n \\ a_2 & \text{if } [s_n' + .5] \geq r_n \\ a_3 & \text{otherwise} \end{cases} \tag{107}$$

and

$$d^{n^*}(x) = \begin{cases} a_1 & \text{if } [s_n' + .5] < C \\ a_2 & \text{otherwise} \end{cases} \tag{108}$$

where n^* is the sample size and C is the critical value of the optimal fixed sample size PUAS procedure.

We have obtained apparently conservative substantive procedures, sequential and fixed sample size, as follows: we derive necessary sample size and critical value based on $S_n \sim \text{binomial}(n,p)$; in performing the test, we substitute S_n' for S_n by discretizing S_n' according to the rule $[S_n'+.5]$. When $S_n'=S_n$ identically, as in compliance testing, $[S_n'+.5]=S_n$ identically, and, so, the test mechanics of MUAS can also be used for PUAS. The degree of conservatism depends, at least in part, upon the degree to which $Var(S_n)=np(1-p)$ overstates $Var(S_n')$. A Monte Carlo study was performed both to provide empirical support for the claim of conservatism and to assess the degree of conservatism under plausible audit circumstances. The study is described in detail, and the results reported, in section 4.3. These results indicate that MUAS is quite conservative under conditions that may well be considered typical, given our limited knowledge of audit populations in general. The principal drawback of conservative tests is inefficiency, i.e. excessive sample size. The use of sequential MUAS should serve to reduce this inefficiency to acceptable levels in many audit testing situations.

4.3 Monte Carlo Study of MUAS

The Monte Carlo results presented below provide some empirical support for the claim of conservatism for the monetary unit acceptance sampling (MUAS) plans, as well as some measure of the degree of conservatism under plausible audit substantive testing conditions. In the case of Bayesian MUAS, the study also provides some evidence for the adequacy of model construction. It should be emphasized that a systematic robustness, or sensitivity, analysis is not contemplated. Rather, the performance of MUAS under a plausible, but constrained, set of circumstances is examined.

4.3.1 Description of the Study

The study population used is an adaptation of Neter and Loebbecke's (1975) population 4. The principal characteristics of the study population are presented in Table 4.1. There appears to be only one characteristic typical of accounting populations: relative frequency is a decreasing function of subunit size (in monetary value). Although the study population is an abstraction of an actual accounts receivable population, it could easily represent inventory, fixed assets, or accounts payable. Actual accounting populations exhibit a wide variety of subunit sizes. Since MUAS places no constraint on subunit size, only nine sizes are used, thereby reducing the cost and time needed to generate test populations from the study population.

Test populations are created by randomly seeding relative errors in subunits of the study population in accordance with one of ten relative error distributions. We will need to distinguish the mean and variance of the relative error distribution from the mean and variance of the test population. The terms "relative error mean" and "relative error variance" will be reserved for the former quantities, and "error mean" and "error variance" will be used for the latter. The relative error distribution consists of positive relative errors only, while the error distribution (test population) is a mixture of positive relative errors (which follow the relative error distribution) and zero relative errors (a constant). For each relative error distribution, two test populations, with error means of .01 and .05, are generated. The following relative error distributions are used:

(1) reverse J--low and high variance (denoted by "low J" and "high J" respectively)

(2) reverse J with 100% relative errors--low and high variance ("low J-100" and "high J-100" respectively)

(3) unimodal--low and high variance

(4) uniform

(5) degenerate at .3, .5, and .8 (i.e. three distributions exhibiting constant relative errors)

In addition to these 10 distributions, a control distribution (in which all relative errors are 0 or 1, i.e. the relative error distribution is degenerate at 1) is used to provide empirical results on nominal risks, since, in this case, the error distribution (under MUAS) is truly binomial. In all cases, the desired error mean (.01 or .05) is attained by varying the proportion of subunits overstated.

There is limited empirical evidence on relative error distributions in accounting populations. Johnson et al. (1981) report a variety of distributions. Distributions (1)-(4) have been used in several Monte Carlo studies (e.g. Roberts et al. (1982) and Leitch et al. (1982)). The degenerate distributions have not been used in other audit studies and are discussed below.

Theoretical distributions are used to model the nondegenerate relative error distributions. The intent here is to produce an approximate shape and predictable properties rather than accurately simulate any given theoretical distribution. The test population generator developed for this study induces relative errors in accordance with the frequencies of a cumulative distribution function (cdf). The cdf may be specified more accurately by increasing the number of points, x_0, x_1,..., at which the cumulative frequency is given. Between any two such points, x_i and x_{i+1}, the relative errors are uniformly induced. The test population generator is listed in Appendix E, and input data for each test population is given in Appendix F.

The J distributions are modelled on gamma distributions. (See Appendix A for all of the theoretical distributions mentioned in this section.) The low J is approximately an exponential(10), i.e. a gamma(1,10). The high J is based on a gamma(.25,2.5). These theoretical distributions have a mean of .1 and variances of .01 and .04, respectively. Due to truncation at 1.0, the high J distributions have variances of about .03. The J-100 distributions are modelled in the same way but with the addition of independently induced 100% relative errors. For these distributions, about 20% of the total error is attributable to 100% relative errors. The choice of 20% is somewhat arbitrary. Johnson et al. (1981) do not report this statistic directly. However, they do report the proportion of relative errors that are 100% errors. Since they found no significant correlation between error amount and relative error, the proportion of 100% relative errors should be a reasonable surrogate for the proportion of total error due to 100% relative errors. (Parenthetically, the lack of significant correlation found in the Johnson study supports the random approach to relative error induction used in this study and others.) Of the high error populations, Johnson et al. report that 7 of 10 of the accounts receivable, and 10 of 10 of the inventory, populations exhibit 20% or less 100% relative errors. Thus, 20% appears to be a reasonable choice. The low unimodal is based on a normal(.5,.01), and the high unimodal is based on a normal(.5,.03). The uniform distribution is approximately a uniform(0,1).

These distributions form three mean-variance groups: J, J-100, and unimodal-uniform. Within each group, the relative error mean is approximately constant and the relative error variance increases. Between groups, the relative error mean increases. Histograms with summary statistics (relative error mean and variance) for these relative error distributions are presented in Figures 4.1-4.7. Each figure consists of two parts: part A depicts the distribution when the error mean is .01, and part B depicts the distribution when the error mean is .05. (Although the same cdf is used in both cases, there are slight differences because the distributions were independently induced in the two cases.) More detailed data on the resulting test populations is given in Table 4.2.

The degenerate distributions exhibit constant relative error of .3, .5, or .8. These distributions are discrete and may be transformed to obtain exact fixed sample size tests. They are included here to assess their impact on sequential MUAS.

All tests are of the following problem:

H_1: p=.01

H_2: p=.05

H_1 represents an immaterial (but positive) level of overstatement. H_2 represents the lowest level of overstatement considered material in the audit literature. Six classical tests are conducted. (The tests are labeled 1.1 through 1.6, where, if used, "F" refers to the fixed sample size test and "S" to its sequential counterpart.) These tests differ in level and power approximately as given in Table 4.3. Exact nominal level and power for each test are given in Table 4.4. (Nominal risks are computed assuming the maximum error variance. Table 4.3 gives the target level/power for the fixed sample size tests. The exact level/power given in Table 4.4 represents the best approximation to the target level/power without randomizing over decision rules.) The choices of level/power were influenced by Elliott and Rogers (1972). They recommend setting level from .05 to .10 and setting power at .95, .90, .85, .70, or .50, depending on the assessed quality of internal control. The low powers of .70 and .50 are not included in the classical tests. However, one of the Bayesian tests (2.6F) effectively has power of about .50 and provides some evidence for low power tests. Sample sizes for the sequential tests are given in Table 4.5. The theoretical values are based on the approximation in (33). Observed values are based on 2500 replications on the control distributions.

Six Bayesian tests (2.1-2.6) are conducted. These tests vary only in specification of the prior distribution as indicated in Table 4.6. Given the loss specification (discussed below), these tests cover the available range, since a prior of .3 or less for H_1 results in a no-sample decision to reject H_1. That is, the lowest prior, g_1=.4, is effectively as extreme as the highest, g_1=.9. The loss function is specified at K_{12}=600 (type I loss) and K_{21}=1500 (type II loss), where losses are measured in unit sampling costs. The particular loss specification used is not critical to this study (if it yields reasonable sample sizes). This is so because the performance of the Bayesian procedures is assessed in terms of average observed loss. Also, given the sample size, it is the <u>ratio</u> of losses that affects the decision. Type I loss of 600 based on the following reasoning. Examination of the 600 largest subunits of the study population will cover approximately 80% of total book value. I assume that, if an auditor rejects H_1 and fails to find a material error after examining 80% of book value, he will not pursue the matter further, concluding that H_1 was, in fact, true. Type II loss of 1500 was arrived at indirectly by answering the question of how much an auditor would be willing to do to forego a type II decision error. (Kinney (1975a) suggested this approach to type II loss specification.) Since a purposive examination of the largest 1500 subunits will cover about 95% of book value, an auditor would presumably be unwilling to do more than this, assuming a materiality level of 5%. On the other hand, he could not do less and still guarantee reduction of the error to an immaterial level, assuming no knowledge of the distribution of relative errors in the population. Implicit in this specification is the notion, generally accepted in the audit profession, that a type II decision error is more serious than a type I decision error.

Exact nominal risks for the Bayesian tests are given in Table 4.7. Theoretical and observed sample sizes for the sequential tests are given in Table 4.8.

All tests, except those on the control distributions, are replicated 500 times. (Control distribution tests, performed to obtain observed nominal values, are replicated 2500 times.) In general, this degree of replication allowed sufficient precision for the hypotheses of interest (discussed below). It should be noted that the tests were performed simultaneously on each of the 500 samples from the various test populations. This facilitates comparison among tests since differences observed from test to test are not caused by sampling variation. Furthermore, the fixed sample size tests are performed by carrying out the sequential tests to $n=n^*$. Thus, the fixed sample size results indicate precisely the risks that would have been incurred if we opted for the fixed sample size test instead of the sequential in each situation. This facilitates comparison between fixed sample size MUAS and its sequential counterpart.

Results are presented graphically and in tabular format for the relative conservatism of the various tests. Relative conservatism is defined as

$$RC_p=(\text{nominal risk}-\text{observed risk})/\text{nominal risk}$$

and measures the degree by which nominal risk overstates ($RC_p > 0$) or understates ($RC_p < 0$) actual risk when p is the error mean. For example, $RC_{.01}=.2$ indicates, in this study, that observed type I risk is 20% less than nominal risk. In the Bayesian tests, the measure of risk used to RC_p is $R(p,d)$ as defined in (78). For the sequential tests, expected sample size is not necessarily known exactly. To overcome this difficulty, observed ASN is used to calculate nominal risk.

The efficiency of sequential MUAS is also considered. Three measures of efficiency are presented. The first is relative efficiency, which is defined as

$$RE_p=(n^*-ASN)/n^*$$

where n*=optimal fixed sample size and ASN=(observed) average sample size. RE_p measures the expected savings in observations over the fixed sample size procedure when p is the error mean. Note that, given the truncation rule adopted in MUAS, $RE_p \geqslant 0$. Since sample size is quite variable in the sequential tests, two other measures of efficiency are presented. The second measure is $\max(RE_p)$, where ASN is replaced in the RE ratio by the minimum observed sample size for a correct decision. $\text{Max}(RE_p)$ is the upper bound on the relative efficiency of sequential MUAS. The lower bound is zero. A more informative statistic is the proportion of truncated decisions (PTD), which measures how often RE_p is zero. Since ASN did not vary significantly over distributions, only results for the control distributions are presented.

4.3.2 Hypotheses of Interest

We are primarily interested in the relative conservatism of MUAS when carried out on plausible error distributions. I have contended that the actual risks of MUAS are bounded by the nominal risks based on the maximum error variance distribution. Thus, we expect to reject, for all nondegenerate relative error distributions, the following hypotheses:

H1: $RC_{.01} < 0$

H2: $RC_{.05} < 0$

where RC_p is the relative conservatism of the MUAS procedure when p is the error mean.

The degenerate relative error distributions are special cases. These distributions violate the assumption of MUAS that the test statistic, S_n', is continuous. No particular hypotheses are entertained with respect to these distributions, and the results are discussed separately.

It should be noted that efficiency as well as conservatism should be considered in evaluating sequential MUAS: a gain in efficiency may offset a loss in conservatism. However, no formal hypotheses are entertained with respect to the efficiency of sequential MUAS. Descriptive statistics on relative efficiency are presented and discussed.

4.3.3 Discussion of the Results

The results of the Monte Carlo study are presented in several figures and tables. The first group--Figures 4.8-4.13 and Table 4.9--pertains to the classical tests on the nondegenerate relative error distributions (J, J-100, unimodal, uniform). Within this group, each figure is a graphical presentation of the results (based on 500 replications) for the relative conservatism of one test. Each figure has two parts. Part A reports sequential MUAS results, and part B reports fixed sample size MUAS results. Each part is divided into upper and lower sections. The upper section reports results when p=.01 (H_1 true), and the lower section reports results when p=.05 (H_2 true). Table 4.9 reports the numerical results that support these graphs.

The second group--Figures 4.14-4.19 and Table 4.10--presents results on the relative conservatism of Bayesian MUAS for the nondegenerate relative error distributions in the same format as that of classical MUAS.

The third group--Tables 4.11 and 4.12--present results on the efficiency of sequential MUAS. 4.11 pertains to classical, and 4.12 to Bayesian, MUAS.

The last group--Tables 4.13 and 4.14--present results on the relative conservatism of MUAS for the degenerate relative error distributions.

4.3.3.1 Conservatism of Classical MUAS

The results for H1 and H2 are presented in Figures 4.8-4.13 and Table 4.9. It should be noted that, for the one-sided hypotheses of interest, the "95%" lower confidence limit in the figures is actually a 97.5% confidence limit. If this limit does not include 0, the hypothesis may be rejected at least at the .025 level. More exact significance levels can be found from the data in Table 4.9. Based on the figures and Table 4.9, we conclude that

(i) H1 may be rejected (p-value $<$ .001) for all relative error distributions except the uniform;

(ii) for the uniform distribution, there is strong evidence (p-value $<$ .01) against H1 for tests with low nominal type I risk (tests 1.1-1.3) and some evidence (p-value $<$ .15) against H1 for tests with high nominal type I risk (tests 1.3-1.6); and

(iii) H2 may be rejected (p-value $<.001$) for all relative error distributions.

An important concern here is the effect, if any, of relative error distribution on RC_p. It is clear from the figures that RC_p tends to decline from the low J to the uniform distribution. Since the relative conservatism of MUAS is predicated on an error variance less than the maximum, we predict an inverse relationship between RC_p and error variance. The maximum error variances are .0099 and .0475 if p=.01 and p=.05, respectively. The error variances of the test populations are given in Table 4.2. The results tend to confirm the prediction. There are anomalies--such as the high $RC_{.01}$ associated with the low unimodal distribution--but the results are not inconsistent with an ordering based on error variance.

The varying test results suggest two conjectures. First, for any given distribution, RC_p is decreasing (not constant) in nominal risk. Thus, for high nominal risk, we may observe low relative conservatism. For $RC_{.01}$, we may compare tests 1.1 and 1.4, 1.2 and 1.5, or 1.3 and 1.6. For $RC_{.05}$, we may compare tests 1.1, 1.2, and 1.3, or 1.4, 1.5, and 1.6. (The nominal risks are given in Table 4.4.) The second conjecture is that RC_p is not symmetric in the hypotheses, that is, for equal nominal risks, $RC_{p_2} > RC_{p_1}$. Nominal risks are not exactly equal in any of the tests. But we may look at test 1.5F and 1.4S, where the nominal risks are quite close. For these two tests, we find only one case in which $RC_{.01} > RC_{.05}$. In most cases, $RC_{.05}$ appears to be significantly higher than $RC_{.01}$.

4.3.3.2 Conservatism of Bayesian MUAS

The results for H1 and H2 are presented in Figures 4.14 -4.19 and Table 4.10. The results are somewhat mixed. The following conclusions pertain to sequential MUAS. Slightly stronger conclusions may be drawn for fixed sample size MUAS, but the pattern is much the same.

(i) H1 can be rejected (p-value $<.001$) for all distributions except the high unimodal and uniform;

(ii) H1 can be rejected (p-value $<.001$) for the high unimodal distribution for all tests except 2.1S;

(iii) H1 cannot be rejected for the uniform distribution;

(iv) H2 can be rejected (p-value $<.001$) for all distributions for all tests except 2.6S; and

(v) there is at least weak evidence (p-value $<.04$) against H2 for test 2.6S for the low J and unimodal distributions, but H2 cannot be rejected for test 2.6S for other distributions.

As in the classical case, MUAS fared worst, in general, against the highest error variance, i.e. the uniform distribution. But with Bayesian MUAS we have two tests that did not exhibit conservatism on one or more distributions. Both of these tests have high nominal risks under one of the hypotheses. The OC functions of these two tests are given below:

	2.1S	2.1F	2.6S	2.6F
$\alpha(.01)$	.721	.754	.974	.954
$\alpha(.05)$	.049	.050	.670	.493

The level of test 2.1 considerably exceeds that of any of the classical tests in this study. Similarly, the power of 2.6 is far lower than any other test. (Note that, for 2.6S, it is merely .330.) Thus, these results are consistent with the conjecture that RC_p declines in nominal risk. It should be noted that, relative to nominal risk of 2.1F, the the increase in nominal risk of 2.1S for the uniform distribution is rather insignificant. However, for 2.6S, this is not the case, since the nominal type II risk of 2.6S is considerably in excess of that of 2.6F.

There is some evidence, then, that sequential MUAS may be more sensitive to prior misspecification than fixed sample size MUAS. In defense of sequential MUAS, it should be noted that both 2.2S (the minimax test) and 2.3S are relatively more conservative than their fixed sample size counterparts for all distributions but the uniform, for which there is no significant difference. And, since these two sequential procedures have lower nominal risks than the corresponding fixed sample size tests, they are clearly superior.

4.3.3.3 Efficiency of Sequential MUAS

Results on relative efficiency in Tables 4.11-12 are based on the control distributions. Observed ASNs for the other relative error distributions are within +11% of the control ASNs, and relative efficiency results are essentially the same.

In general, it is apparent that greater efficiency is attainable under H_2 than H_1. Under H_2, the saving can be dramatic, since rejection can occur after only a few observations. For the classical tests, RE_p is a decreasing function of nominal risk. For the Bayesian tests, RE_p increases as the prior is more correctly specified. When the the prior is significantly incorrect (e.g. test 2.1 for p=.01 and test 2.6 when p=.05), RE_p is no longer meaningful since the sequential test terminates with the incorrect decision too frequently. I have omitted the RE_p measure in these cases because the apparent savings are spurious.

Sample size of sequential MUAS varies rather broadly. Although not reported here, the standard deviation of the sample size is usually from 30% to 50% of ASN. However, the average savings of sequential MUAS over fixed sample size MUAS, when $p=p_1$ or $p=p_2$, appear to be significant (from 40% to 60%).

4.3.3.4 Results for the Degenerate Distributions

The degenerate distributions used in this study exhibit constant relative error of .3, .5, or .8. Results on the relative conservatism of MUAS for these distributions are reported in Tables 4.13 and 4.14.

Degeneracy in the relative error distribution violates the assumption of MUAS that the test statistic, S_n', is continuous and invalidates use of the continuity correction. For example, if the distribution is degenerate at .5, we would normally require two observations of overstatements before recording an error (.5+.5=1). However, with the continuity correction, one occurrence is sufficient to record an error.

For fixed sample size tests, it is easy to transform the problem to obtain an exact test. For sequential MUAS, the effects are not apparent.

To illustrate the needed transformation, consider the degenerate .5 distribution. Test 1.3F has a critical value of 4, hence it is necessary to observe 7 occurrences of overstatement in order to reject ((7)(.5)=3.5+.5 for the continuity correction). Transform the error rates as follows: $p_1'=p_1/.5=.01/.5=.02$ and $p_2'=p_2/.5=.05/.5=.10$. Then test 1.3F is risk equivalent to the following test 1.3F':

H_1: $p'=.02$

H_2: $p'=.10$

with $n^*=95$ and $C=7$. The nominal level and power of test 1.3F' are .012 and .954, respectively. Then the expected value of $RC_{.01}$ for the test is $(.034-.012)/.034=.647$. (The nominal value .034 is taken from Table 4.4.) This expected value is within half a standard deviation of the observed value of .585 in Table 4.13. Similarly, we find $RC_{.05}=.695$ which is within one and a half standard deviations of the observed value, .775.

By this method, we may assess the impact of degeneracy on fixed sample size MUAS. It is clear that it may cause RC_p to go negative (e.g. test 2.1F for the degenerate .5 distribution).

The error variances for test populations with degeneracies may be computed from (92) as $Var(X)=p(p'-p)$, where p' is the point of degeneracy. They are given in the table below for the degenerate distributions tested and for the control distributions (degenerate at 1.0):

	Point of Degeneracy			
	.3	.5	.8	1.0
p=.01	.0020	.0049	.0079	.0099
p=.05	.0125	.0225	.0375	.0475

By comparing these variances with those of Table 4.2, we see that a degeneracy at .3 is comparable to the low J distribution, .5 is comparable to the high J-100, and .8 is more variable than the uniform. However, except for p'=.3, the results here are rather different from those for the comparable nondegenerate distributions. We would expect RC_p to decline as p' increases. And this occurs in, say, test 1.1. However, a different pattern is obserred in 1.5 and 2.6. Furthermore, similar shifts in p' can produce both significant and insignificant changes in RC_p. For example, in test 1.2, for p=.01, the shift from .3 to .5 produces a dramatic drop in RC_p while the shift from .5 to .8 has an insignificant effect.

It would appear from Tables 4.13 and 4.14 that the effects of degeneracy on sequential MUAS tend to follow the effects on fixed sample size MUAS. If this is true in general, it is at least possible to predict the effect of any given degeneracy on sequential MUAS from the expected effect on its fixed sample size counterpart. It would, then, also appear that any significant increase in risk due to degeneracy will be in type I risk.

In developing MUAS, I have assumed that degeneracies of the type tested here do not occur in accounting populations. Although presumably rare, they could occur due to a systematic bookkeeping error. For example, a clerk could systematically understate purchase discounts in a scheme to abstract funds from the employer. This could result in a constant error rate for overstated inventory items. Under these circumstances, it would seem that, depending on the parameters of the MUAS test, the auditor's type I risk might exceed the nominal risk. This is not a particularly discouraging result, since, in the event of rejection, the auditor will, in fact, search for sources of systematic error. And, while the error may not be material in the current period, its discovery and correction may forestall a material error in future periods.

4.3.4 Other Considerations

In this section, we will consider the power function of MUAS and its implications with regard to choice of test. For the time being, I limit the discussion to classical tests.

We have restricted MUAS to the testing of simple hypotheses. This is an admitted simplification of the problem. The error rate p can lie anywhere in the interval from 0 to 1. Consider the alternative test

$$\begin{aligned} H_1' &: p < p^* \\ H_2' &: p \geq p^* \end{aligned} \qquad (109)$$

where p* is a material error rate. In the N-P framework, we cannot conduct a reasonable test of (109) as it stands. This so because, if type II risk at $p=p^*$ is β, then, as p approaches p^* from the left, type I risk approaches $1-\beta$. This difficulty is removed if we are willing to use an indifference zone. That is, we introduce a $p'<p^*$ such that, if $p'<p<p^*$, we are indifferent to the decision made. Thus, we control type I risk at p' and type II risk at p^*. Since the power function is monotonically increasing in p, these are the maximum risks we face for $p \leqslant p'$ and $p \geqslant p^*$, respectively. Thus, (109) is equivalent to

$$H_1'': p=p' \qquad H_2'': p=p^* \tag{110}$$

And, letting $p_1=p'$ and $p_2=p^*$, we arrive at the MUAS construction.

While use of simple hypotheses, if interpreted in this way, does not represent a constraint in the N-P framework, we must nevertheless consider the performance of MUAS when p is not one of the two hypothesized values.

We will consider the power function of only one test (1.1) for only two relative error distributions (low J and uniform). However, this should be adequate to indicate the general nature of the power function of MUAS. The empirical power of test 1.1 (fixed sample size and sequential) against various values of p from .005 to .07, based on 500 replications, is given in Table 4.15. The observed average sample sizes (ASN) are also given there. The theoretical power assumes a binomial error distribution (all relative errors equal 0 or 1). The empirical power functions of 1.1F are plotted against the theoretical power function in Figure 4.20. The power functions of 1.1S (which are not plotted) would be shifted slightly to the right.

From Figure 4.20, it is clear that the effect of nondegenerate relative error distributions is a rotation of the theoretical power counter-clockwise with, perhaps, a small shift to the left. (To simplify description, we will call the p such that $\beta(p)=.50$ the <u>midpoint</u> of the power function.) Based on the performance of MUAS at $p=p_1$ and $p=p_2$ for the various relative error distributions tested earlier, it is reasonable to conclude that the power functions for high J, low J-100, etc. lie between those for the low J and the uniform. It is also reasonable to conclude that a decreasing error variance tends to increase the slope of the power function near its midpoint. (In the limit, when the error variance is zero, the power function jumps from 0.0 to 1.0 in the vicinity of the theoretical midpoint.) The location of the midpoint, then, is of some importance in choosing an MUAS test. In addition, we observe that the ASN for nondegenerate distributions tends to <u>exceed</u> the theoretical bound when $p_1 < p < p_2$. For p near the midpoint, the sequential sample size will equal the optimal fixed sample size fairly often, particularly for low error variance distributions. While this has implications for the choice of sequential MUAS test, it must be kept in mind that sequential MUAS is being advanced as a means of early detection of outliers, i.e. $p < p_1$ or $p > p_2$.

In a recent paper, Duke et al. (1982) compared the power functions of several statistical substantive test procedures. Their results are not directly comparable with those in Table 4.15 because they test $p_1=.00$ versus $p_2=.02$ and control either type I risk at p_1 or type II risk at p_2 but not both. Further, for reasons to be discussed, p=.02 is usually an unrealistically low alternative. Temporarily adopting the notation used by Duke et al., let M be a material error rate. As constructed by these authors, a good test would exhibit a power function rising from α at p=M-e to $1-\beta$ at p=M for some small e. In particular, they require $e \leqslant .5M$. Unfortunately, for reasonable values of M, α, and β (say, $\leqslant .1$), this constraint will yield very large sample sizes. In fact, such a constraint may lead to the conclusion that a purposive sampling plan designed to cover 100(1-M)% of book value is preferable to a random sampling plan. (This is, for example, probably the case if we set M=.02, since then .5M=.01, and it is clear that very large sample sizes are needed to discriminate with reasonable accuracy between p=.01 and p=.02, if the sampling is at random.)

We will consider the power characteristics of MUAS along the lines of the Duke et al. construction but with the following modifications: (i) materiality will be treated as an interval, rather than point, concept, and (ii) purposive sampling of large subunits in the population will be allowed. The Duke et al. discussion is incomplete in these two areas. In addition, they do not address the impact of multiple tests on the design of a particular component test. This problem has two dimensions. The first is the impact of compliance tests on subsequent substantive tests. The second is the impact of other substantive tests on a particular substantive test. This area has been the subject of research. But it is a complex problem, discussion of which would carry us far afield, hence we, too, will consider only the isolated test.

All discussions of audit materiality with which I am familiar have recognized the difficulty of establishing a "threshold" of materiality. For example, Mautz and Sharaf (1964, p. 105) refer to "borderline assertions" that are "more than immaterial but less than definitely material." The reluctance of standard-setting bodies to incorporate quantitative materiality rules is an implicit recognition of this grey area (see FASB, 1980, Appendix C). For purposes of formal development, we have taken p_2 as <u>the</u> material error rate. But it is unrealistic to assume that an auditor is able to specify a material error rate M such that M-e is immaterial for some small e. Rather, it is reasonable to suppose that an auditor is able to specify, for a given population, an error rate M' that is marginally material and an error rate M* that is certainly material, with $M' < M^*$. There are at least two objective interpretations of these error rates. The first is that, in the auditor's judgment, the decisions of <u>some</u> reasonable users of the financial statements would be affected by knowledge of an undisclosed M' error rate in the population, while the decisions of <u>all</u> reasonable users would be affected by knowledge of an undisclosed M* error rate. A second interpretation, more in accord with current legal views on materiality, is that there is a <u>moderate likelihood</u> that the decisions of a reasonable user would be affected by knowledge of M' but <u>virtual certainty</u>

if the error rate is M*. In addition, I assume the auditor is able to specify an error rate m that is certainly immaterial. That is, an undisclosed m error rate would affect no reasonable user, or there is virtually no likelihood that it would affect a reasonable user. Although it is possible that $m > 0$, I will assume that $m=0$.

The value of this construction lies in its implications for the choice of test. We have immediately that $m \leq p_1 < M' < p_2 \leq M^*$. Further, we are able to characterize the desired power function to some degree. We require that (i) the power against m is quite low, (ii) the power against M' is moderate (since we are rather indifferent about detecting a marginally material error rate), and (iii) the power against M* is quite high. If, as agreed, we set $m=0$, then $\beta(m)=0.0$ in all MUAS tests, hence we need not be concerned with the power function at this point. (This is not true for all statistical substantive procedures.) A power function rising from about $\beta(M')=.50$ to $\beta(M^*)=.99$ might satisfy the remaining requirements. If we set equal decision risks (i.e. $\beta(p_1)=1-\beta(p_2)$), then choosing p_1 and p_2 equidistant from M' should yield $\beta(M') \doteq .50$. (Since there will typically be considerable latitude in the choice of p_1 and p_2, the use of equal decision risks is not particularly constraining, but, regardless, we are only suggesting one possibility for specifying the test in a reasonable manner.) Beyond this, choice of p_1 and p_2 represents a tradeoff. A relatively smaller indifference zone is usually preferable, especially for sequential implementation, but optimal fixed sample size is quite sensitive to the size of this zone. We will return to this question later.

In the following example, we take $M'=.02$ and $M^*=2M'=.04$. A reasonable choice for p_2 is $p_2=(M'+M^*)/2=1.5M'=.03$. Then if we set $p_1=.5M'=.01$, M' will be roughly the midpoint of the indifference zone, if equal decision risks are used. This corresponds to the Duke et al. setup except that we treat p=.02 as marginally material. If we set $\alpha=\beta(.01)\leq .1$ and $1-\beta=\beta(.03)\geq .9$, we arrive at n=320 and C=6 as the best test. The theoretical power function of the (fixed sample size) test at several points is presented below:

p	$\beta(p)$
.005	.006
.0075	.035
.01	.104
.015	.349
.02	.618
.025	.812
.03	.919
.035	.969
.04	.989
.05	.999

This plan provides considerable ultimate protection against $M^*=.04$ even if we face a binomial error distribution. If, on the other hand, a low error variance distribution is encountered, the power function will be quite steep in the vicinity of p=.02, $\beta(.01)$ will be significantly lower than .104, and $\beta(.03)$ will be significantly higher than .919.

We now allow purposive sampling of large subunits. We let q be the proportion of book value covered by the purposive sample. It is clear that, if M is a material error rate prior to purposive sampling, then $M''=M/(1-q)$ is material subsequent to the purposive sample (i.e. for the random sample of the remaining subunits). This can have a significant impact on the statistical test. For the example above, we transform the parameters and recompute the necessary sample size for various values of q. (For consistency, we also let $p_1''=p_1/(1-q)$ and similarly for p_2''.)

q	p_1	M'	p_2	M*	n	C
.00	.01	.02	.03	.04	320	6
.33	.015	.03	.045	.06	205	6
.50	.02	.04	.06	.08	160	6
.75	.04	.08	.12	.16	78	6

Now, q=.75 may seem unrealistically high. However, our study population is based on Neter and Loebbecke's (1975) population 4 in which the "very few" excluded subunits (those over $25,000) accounted for 75% of book value. Neter and Loebbecke excluded these subunits precisely because they assumed they would be purposively selected by an auditor (Neter and Loebbecke, 1975, p. 25). In the only other complete population used by these researchers, population 3, the excluded subunits accounted for 33% of book value.

It is clear that purposive sampling of large subunits can dramatically reduce the necessary size of the random sample. Furthermore, based on the high degree of skewness in the distribution of subunit size typically found in accounting populations (e.g. Neter and Loebbecke (1975), Johnson et al. (1981)), it would appear that purposive sampling of large subunits will often significantly impact the statistical test of the remaining subunits.

Before turning to Bayesian MUAS, we will pause to reconsider test 1.1 and the question of choice of p_1 and p_2. Given the theoretical power function in Table 4.15, test 1.1 is apparently appropriate if M'=.03 and M*=.06, which is the situation in our example for q=.33. However, in test 1.1, we set $\alpha \doteq \beta \doteq .05$ with p_1=.01 and p_2=.05 rather than $\alpha \doteq \beta \doteq .10$ with p_1=.015 and p_2=.045. For the latter test, the indifference zone is smaller, but the optimal fixed sample size is larger. A brief comparison of their theoretical power functions is given below:

p	$\beta(p)$ n=182	$\beta(p)$ n=205
.005	.002	.001
.01	.037	.018
.015	.142	.090
.02	.301	.229
.03	.640	.581
.04	.856	.832
.045	.916	.903
.05	.952	.946
.06	.986	.985

The latter test provides better protection against a type I error at the cost of larger sample sizes if $.015 < p < .045$. This kind of tradeoff must be assessed by the decision-maker.

We now consider the relation of Bayesian MUAS and the power function. First, the Bayesian framework does not help in the choice of p_1 and p_2. But, given p_1, p_2, and M', Bayesian MUAS gives an alternative, and perhaps superior, means of choosing sample size. Assume that, in the event of rejection, a purposive sample covering 100(1-M')% of book value will be taken. Assume further that this is also what the auditor would "pay" to forego a type II error. Thus, $K_{12}=K_{21}=K$.

Of course, K is decreasing in M'. Given the same parameters as our classical example, a table of the comparable Bayesian tests is given below. The value of K is based on our study population (Table 4.1), i.e. for M'=.02, it is necessary to examine approximately 2250 of the largest subunits to cover 98% of book value, etc. The prior for H_1 is .5 for all tests.

q	p_1	M'	p_2	M*	K	n	C
.00	.01	.02	.03	.04	2250	248	5
.33	.015	.03	.045	.06	2000	202	6
.50	.02	.04	.06	.08	1800	178	7
.75	.04	.08	.12	.16	1300	115	9

From a Bayesian perspective, it would appear that our classical test for q=.00 is too conservative and for q=.75 is too liberal, for our study population. Note, however, that the Bayesian construction is directly sensitive to the skewness of subunit size in the population through the specification of K (in USCs), regardless of the value of q. But the classical construction is sensitive to this skewness only indirectly through the specification of q.

Finally, it must be observed that the test in (109) does not reduce to that in (110), in the Bayesian approach, without some arbitrary simplification. To test (109) would require assessing a continuous prior (or reasonable discrete analog thereof). This constitutes a well-studied behavioral difficulty. Beyond this behavioral difficulty, there is a nontrivial increase in technical complexity. Given the uncertainty of the benefits to be derived, I have adopted the position that the simplified construction should be shown defective before the more realistic construction is embraced.

4.4 Summary

In this section, we will reiterate rather generally the strengths and weaknesses of MUAS and also discuss some issues that were deferred in order to keep the development reasonably uncluttered.

The Monte Carlo study tends to support the use of MUAS in substantive testing for overstatement in asset balances. In general, the claim that the actual risks of MUAS are bounded by the nominal risks based on a binomial error distribution holds for the nondegenerate relative error distributions considered in the study. Indeed, if the error variance is significantly less than that of the binomial error distribution (as would typically be the case for certain gamma-type relative error distributions), MUAS is quite conservative. That is, the nominal risks, based on the binomial error distribution, will significantly overstate the actual risks at the hypothesized error rates (p_1 and p_2). For other values of the error rate, the effect of low error variance distributions is essentially a counter-clockwise rotation of the power function for the binomial error distribution about the midpoint of the indifference zone, with the result that the true power function may be significantly steeper than the nominal power function in the vicinity of the midpoint.

There is both analytic and empirical evidence that gamma-type, low error variance relative error distributions occur frequently in accounting populations. The anlytic evidence is based on the following kinds of argument. Positive relative errors occur more or less uniformly on the unit interval in the accounting process, however, the effectiveness of accounting controls imposed by an entity's management is an increasing function of the magnitude of the relative error. Thus, such controls operate as a filter, converting, say, a uniform relative error distribution into a gamma-type distribution. Alternatively, the accounting process, with controls, may be viewed as yielding a normal (positive and negative) relative error distribution (truncated at 1 on the right), with zero mean and variance depending on control effectiveness. The positive relative errors, then, follow a gamma-type distribution. Empirical evidence for such distributions is mainly derived from the limited number of accounting populations described by Johnson et al. (1981).

However, from both analytic and empirical viewpoints, it would appear that 100% positive relative errors may be independently generated. Johnson et al. found several populations with high proportions of such errors. And Duke et al. (1982) suggest that one fraud strategy is the use of entirely fictitious subunits to achieve a material overstatement in the population. Thus, reliance on an assumption that relative errors follow a gamma-type distribution (e.g. Cox and Snell (1979)) does not appear warranted without considerable investigation of the robustness of such an assumption against high error variance populations. That is, it would appear that auditors must use procedures that are conservative under typical circumstances in order to obtain nominal protection in atypical circumstances. MUAS is such a procedure.

The principal drawback of conservative procedures is excessive sample size. Sequential MUAS has been advanced as a reasonable solution to this dilemma. When the true error rate p is either significantly better or worse than expected, sequential MUAS will typically detect this fact at moderate sample sizes. Furthermore, these moderate sample sizes will be attained without adopting an unrealistic model (e.g. use of discovery sampling when some positive error rate is both expected and tolerable) or sacrificing power against material error rates. Sequential MUAS, then, is best viewed as a scheme for the early detection of "outliers" (i.e. $p < p_1$ or $p > p_2$). (Elliott (1976) first advanced this view of sequential audit tests.) If $p < p_1$, the client should not be burdened with excessive sampling cost since he has performed better than auditor expectations. If $p > p_2$, excessive sampling is again unwarranted, but for the reason that audit resources are better expended to assist the client in remedial work on the balance in question. However, when $p_1 < p < p_2$, the situation is not so clear, and the auditor may very well need additional sample information in order to make a reasonable decision on how to proceed if indeed H_1 is rejected. A primary drawback of the SPRT is the potentially large sample size that may be required under these circumstances. Hence, the truncation rule adopted in sequential MUAS (stopping at the optimal fixed sample size if no decision is made earlier) is an important component in the applicability of sequential tests in auditing.

The performance of sequential MUAS is more or less sensitive to other factors considered in the Mónte Carlo study. The following matrix indicates, in a qualitative way, the utility of sequential MUAS.

		efficiency		effectiveness	
		$p \leq p_1$	$p \geq p_2$	$p \leq p_1$	$p \geq p_2$
low nominal risk	low error variance	good	excellent	excellent	excellent
	high error variance	good	excellent	good	good
high nominal risk	low error variance	fair	good	excellent	excellent
	high error variance	fair	good	fair	good

The availability of sequential implementation is, perhaps, the principal advantage of MUAS over current statistical audit methodology. However, there are other advantages. MUAS is the first statistical substantive procedure cast entirely in the testing framework. Although confidence procedures can be used to make decisions, the terminology and construction of statistical tests is a more natural framework for audit tests. Moreover, MUAS is derived from PUAS and thus unifies statistical auditing (compliance and substantive) conceptually in terms of a readily accessible discrete probability structure (the binomial distribution). Not only does this unification simplify implementing statistical tests in an audit, I hope that MUAS will significantly facilitate statistical audit pedagogy.

The ready availability of a "worst case" power function for MUAS is also a distinguishing feature. That is, in the event that a binomial error distribution is encountered (which is essentially the "worst case" for MUAS), the auditor can easily compute the power against any error rate or consult binomial or Poisson tables. This should be of assistance to the auditor in choosing an appropriate test. The power function of a sequential MUAS will be somewhat different than that of the corresponding fixed sample size MUAS test. However, the power of the sequential test can be computed, and I have provided an algorithm for this purpose. This algorithm should be efficient for typical audit sample sizes.

A major contribution of Bayesian MUAS is a new sequential procedure appropriate for audit usc. In addition, the Bayesian construction of MUAS incorporates certain simplifications over previous Bayesian models proposed for audit testing. In developing Bayesian MUAS, I have adopted the simple construction of a two-point parameter space and discrete prior under the assumption that a simple construction should be shown defective before more complicated constructions are espoused. The Monte Carlo study performed here does not directly address this question, but there is no evidence in the Monte Carlo results of defective construction. In fact, Bayesian MUAS appears reasonably robust against prior misspecification, a worrisome aspect of Bayesian models. Against values of p other than those hypothesized, Bayesian MUAS shares the power characteristics of classical MUAS.

Simplified construction is also evident in the choice of loss function and scale. Use of the unit sampling cost (USC) as the loss scale should ease the implementation of Bayesian MUAS over both different audit clients and different testing situations for the same client. (The usefulness of this scale was apparent in the discussion of Bayesian tests in section 4.3.4 above.) And we have excluded any cost to access the sampling frame (startup costs). This is a one-time fixed cost (not, as Kinney (1975, p. 123) claims, a fixed cost that will be incurred at each sampling stage). It will be incurred regardless of the decision taken, if any sampling is done. Hence it affects only the decision of whether or not to sample. This decision is based only in part on the startup costs. An attempt to formalize this decision at the testing level appears counterproductive.

We now consider some of the (real or apparent) deficiencies of MUAS. I have assumed throughout that, in the event of rejection, remedial work on the population will be performed by the auditor or the client (or both). Some auditors have advocated the use of <u>stochastic adjustments</u>, i.e. a proposed adjustment to the population book value based on a statistical estimate of the true value (see, e.g., Loebbecke and Neter (1975)). Although I do not advocate the use of stochastic adjustments, MUAS does provide an unbiased estimate of the population error rate, namely, s_n'/n. Further, an unbiased estimate of the variance of the estimator S_n'/n is available (Cochran (1977, p. 308)).

Large-sample confidence intervals using this variance estimator have not proved especially accurate when few errors are encountered (Neter and Loebbecke (1975)), primarily because the variance estimate is zero if no errors are found. However, a stochastic adjustment would be needed only if H_1 is rejected. This typically will require observing several errors. Thus, large-sample confidence intervals constructed only when H_1 is rejected may be rather accurate. These conditional confidence intervals will differ from the usual unconditional intervals which, if used in these circumstances, would have lower than nominal coverage probability. While it is possible to compute the appropriate conditional interval, if we are interested only in the upper confidence bound (UCB), then the unconditional UCB will lie to the right of the conditional UCB in MUAS tests, and, so, the unconditional upper confidence coefficient will be at least as large as the conditional coefficient. Hence, use of an unconditional $100(1-\alpha)\%$ UCB on p may be viewed as a conservative approximation to the conditional UCB. (See Meeks and D'Agostino, American Statistician (May 1983), p. 134-136. Note that their objections to the use of conditional intervals relates to the behavior of the lower confidence bound.)

Both in PUAS and in MUAS, we have used sampling with replacement. In a labeled finite population, sampling without replacement is generally superior. However, by assuring independent and identically distributed random variables, random sampling with replacement considerably simplifies the probability structure of a sampling plan. In fact, in sampling with unequal probabilities of selection (as in MUS viewed as a

<u>subunit</u> selection method), the analysis in the case of sampling without replacement becomes quite complex (Cochran (1977, p. 308ff)). This complexity has led some (e.g. Duke et al. (1982)) to use sampling with replacement for MUS procedures and others to use sampling without replacement but analyze the results as if the observations were independent (see discussion in Cox and Snell (1979)).

While I have used sampling with replacement primarily to simplify the analysis, I will offer an alternative defense for its use. In MUAS, if two or more dollars are selected from the same subunit, each dollar counts as a valid observation, but the subunit need be audited only once. Thus, it is only necessary to tag sample dollars from the same subunit at the time the sample selection is made. But, if we use sampling without replacement, this is precisely what we must do to avoid duplicate choices, if the sampling is at random. (Since the probabilities of selection are <u>unequal</u>, it is not sufficient to coerce the random number generator into passing over duplicates. That is, two different numbers may still select the same subunit.) Thus the cost of random sampling with and without replacement in MUAS is essentially the same. (A systematic sampling scheme does not require tagging, however, such a plan introduces additional analytic difficulty and the need for additional assumptions.) In PUAS, on the other hand, we have used sampling with replacement because the populations involved are usually large and the probability of duplicate selection is quite low. Here, although the preferable scheme is well understood (requiring use of the hypergeometric instead of the binomial distribution), the added complexity provides little benefit.

A final disadvantage of MUAS is its failure to address the problem of understatement in liabilities and assets. (Overstatement of liabilities, while not usually a concern of an independent auditor, may be treated by MUAS as it stands.) Understatement of liabilities, which leads to an overstatement of income, is a major concern of independent auditors. However, no statistical procedure currently available to auditors adequately deals with this problem. The difficulty is the lack of a reasonably complete sampling frame. In the case of accounts payable, for example, the balance itself cannot be assumed to be complete, since omissions of entire subunits are not only possible but probable. To apply MUAS we must find a reasonably complete frame. For example, if the client's payables turnover is about 6, and the cash disbursements system is reliable, the first 60 days' disbursements in the subsequent period may serve as a frame for the testing of accounts payable. In this situation, valid disbursements are those for debts arising subsequent to year-end or for debts recorded in accounts payable at year-end. Invalid ("overstated") disbursements are those for debts arising before year-end but not listed in accounts payable at year-end. These "overstatements" will lie in the unit interval, and the test may proceed as with asset balances. The understatement of assets, as the overstatement of liabilities, is usually not the concern of independent auditors. A statistical test again depends upon finding a reasonably complete frame (for example, the last 60 days' sales in the period for an accounts receivable balance). While any understatements observed in the course of an MUAS test for overstatement can be corrected, the theory does not permit netting these against observed overstatements.

TABLE 4.1

Study Population Characteristics

Total Book Value: 8,988,750

Number of Subunits: 4,000

Distribution of Subunits by Size:

Subunit Size (in dollars)	Frequency	Relative Frequency	Cumulative Value
75	1050	.26	78750
150	700	.18	183750
300	450	.11	318750
600	350	.09	528750
1200	450	.11	1068750
2400	400	.10	2028750
4800	150	.04	2748750
9600	250	.06	5148750
19200	200	.05	8988750
Totals	4000	1.00	

TABLE 4.2

Summary Statistics of the Test Populations

Distribution	Error*		Partial Tainting+		100% Tainting+	
	Mean	Variance	Dollars	Subunits	Dollars	Subunits
Low J	.0095	.0019	.0809	.0808	--	--
High J	.0100	.0031	.1024	.1160	--	--
Low J-100	.0104	.0035	.0928	.0723	.0023	.0015
High J-100	.0097	.0048	.0768	.0853	.0025	.0028
Low Unimodal	.0098	.0045	.0216	.0195	--	--
High Unimodal	.0103	.0049	.0247	.0193	--	--
Uniform	.0104	.0068	.0203	.0170	--	--
Low J	.0501	.0068	.5005	.4978	--	--
High J	.0498	.0166	.4801	.4980	--	--
Low J-100	.0496	.0160	.3800	.3793	.0115	.0110
High J-100	.0497	.0239	.4213	.4255	.0121	.0125
Low Unimodal	.0500	.0236	.0998	.1100	--	--
High Unimodal	.0497	.0253	.0986	.0925	--	--
Uniform	.0504	.0303	.1043	.1040	--	--

*error mean=error rate as given in (91), i.e. p=K/N
error variance=Var X as given in (92)

+a "tainted" subunit is one that is partially or 100% in error; these columns measure the proportion (relative to total book dollars) of dollars in tainted subunits and the proportion (relative to total subunits) of tainted subunits

TABLE 4.3

Classical Tests Performed in the Study

H_1: p=.01 vs. H_2: p=.05

Test	Level*	Power*	Sample Size	Critical Value
1.1	.05	.95	182	5
1.2	.05	.90	134	4
1.3	.05	.85	120	4
1.4	.10	.95	155	4
1.5	.10	.90	107	3
1.6	.10	.85	94	3

*target nominal risks; since the underlying distribution is discrete, these target risks are not exactly attainable (without randomizing over decision rules); exact nominal risks for the classical tests used are given in Table 4.4

TABLE 4.4

Nominal Risks of the Classical Tests

H_1: p=.01 vs. H_2: p=.05

Test[1]	Theoretical[2] Level	Theoretical[2] Power	Observed[3] Level	Observed[3] Power
1.1F	.038	.948	.041 (.004)	.950 (.004)
1.2F	.047	.901	.046 (.004)	.907 (.006)
1.3F	.034	.849	.036 (.004)	.847 (.007)
1.4F	.072	.950	.066 (.005)	.956 (.004)
1.5F	.094	.902	.088 (.006)	.895 (.006)
1.6F	.070	.848	.066 (.005)	.845 (.007)
1.1S	.039	.930	.040 (.004)	.926 (.005)
1.2S	.046	.870	.044 (.004)	.860 (.007)
1.3S	.031	.800	.035 (.004)	.787 (.008)
1.4S	.076	.935	.065 (.005)	.935 (.005)
1.5S	.092	.877	.087 (.006)	.861 (.007)
1.6S	.066	.808	.065 (.005)	.795 (.008)

(1) F=fixed sample size, S=sequential

(2) for fixed sample size tests, risks calculated using Poisson approximation to the binomial; for the sequential tests, risks calculated using the binomial by the method in (21)

(3) based on 2500 replications on the control distributions; the standard deviation is shown in parentheses

TABLE 4.5

Sample Sizes of the Classical Tests
H_1: p=.01 vs. H_2: p=.05

Note: observed average sample size (ASN) is based on 2500 replications on the control distributions; standard deviation of the ASN is less than 1.0 for all tests

		ASN			
		Bound+		Observed	
Test	n*	p=.01	p=.05	p=.01	p=.05
1.1	182	105	82	102	78
1.2	134	79	70	76	64
1.3	120	64	62	64	62
1.4	155	100	68	96	62
1.5	107	69	47	69	48
1.6	94	57	46	57	47

+computed using the approximation given by (33) in section 3.2; the results are exact for tests 1.3, 1.5, and 1.6

TABLE 4.6

Bayesian Tests Performed in the Study

H_1: p=.01 vs. H_2: p=.05

Test	Prior for H_1	Sample Size	Critical Value
2.1	.4	95	2
2.2	.5	120	3
2.3	.6	112	3
2.4	.7	102	3
2.5	.8	88	3
2.6	.9	34	2

Note: losses of K_{12}=600 and K_{21}=1500 are used in all tests; see Table 0.6 for the nominal risks of these tests

TABLE 4.7

Nominal Risks of the Bayesian Tests

H_1: p=.01 vs. H_2: p=.05

Test[1]	Theoretical[2] R(.01)	Theoretical[2] R(.05)	Observed[3] R(.01)	Observed[3] R(.05)
2.1F	242.51	169.62	234.92 (5.07)	171.80 (6.79)
2.2F	192.31	212.95	184.32 (3.71)	210.00 (7.12)
2.3F	174.19	235.58	168.40 (3.50)	232.00 (8.14)
2.4F	152.41	276.72	150.00 (3.26)	286.80 (10.53)
2.5F	123.76	365.71	122.56 (2.80)	375.40 (11.80)
2.6F	61.74	773.87	64.00 (2.61)	776.20 (15.00)
2.1S	237.89	105.93	230.20 (5.30)	114.40 (6.80)
2.2S	162.39	154.07	158.40 (3.93)	159.00 (7.77)
2.3S	135.42	206.36	129.28 (3.48)	215.60 (9.36)
2.4S	106.68	293.90	104.48 (3.08)	313.20 (11.35)
2.5S	75.50	449.88	77.20 (2.57)	477.40 (13.52)
2.6S	28.83	1024.03	29.24 (1.92)	1021.00 (14.13)

(1) F=fixed sample size, S=sequential

(2) R(p)=R(p,d) as given by (78); for fixed sample size tests, E(n)=n* and the Poisson approximation to the binomial is used; for sequential tests, the observed ASN is used for E(n) and the binomial distribution is used by means of (21)

(3) based on 2500 replications on the control distributions; the standard deviation, shown in parentheses, is computed assuming the ASN is fixed at the observed quantity

TABLE 4.8

Sample Sizes of the Bayesian Tests

H_1: p=.01 vs. H_2: p=.05

Note: observed average sample size (ASN) is based on 2500 replications on the control distributions; standard deviation of the ASN is less than 1.0 for all tests

		ASN			
		Bound+		Observed	
Test	n*	p=.01	p=.05	p=.01	p=.05
2.1	95	69	30	69	31
2.2	120	85	47	84	48
2.3	112	74	48	73	49
2.4	102	62	49	61	49
2.5	88	47	47	47	47
2.6	34	13	15	13	15

+computed using the approximation given by (33) in section 3.2; the results are exact for all tests

TABLE 4.9A

Relative Conservatism of Classical Sequential MUAS

H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk

		Distribution						
		J		J-100		Unimodal		Uniform
Test	p	Low	High	Low	High	Low	High	
1.1S	.01	1.000 (0.000)	0.900 (0.071)	0.849 (0.087)	0.799 (0.100)	0.900 (0.071)	0.749 (0.112)	0.498 (0.157)
	.05	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.939 (0.043)	0.816 (0.075)	0.785 (0.081)	0.693 (0.096)
1.2S	.01	1.000 (0.000)	1.000 (0.000)	0.870 (0.075)	0.783 (0.097)	0.913 (0.061)	0.696 (0.114)	0.479 (0.149)
	.05	1.000 (0.000)	0.969 (0.022)	0.954 (0.027)	0.877 (0.043)	0.724 (0.064)	0.678 (0.069)	0.678 (0.069)
1.3S	.01	1.000 (0.000)	1.000 (0.000)	0.871 (0.091)	0.806 (0.112)	0.871 (0.091)	0.612 (0.157)	0.548 (0.170)
	.05	0.990 (0.010)	0.830 (0.041)	0.920 (0.028)	0.710 (0.052)	0.710 (0.052)	0.579 (0.062)	0.519 (0.066)
1.4S	.01	0.974 (0.026)	0.816 (0.069)	0.816 (0.069)	0.657 (0.094)	0.868 (0.059)	0.578 (0.104)	0.157 (0.144)
	.05	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.784 (0.081)	0.814 (0.075)	0.753 (0.087)
1.5S	.01	0.978 (0.022)	0.761 (0.071)	0.718 (0.077)	0.718 (0.077)	0.718 (0.077)	0.653 (0.085)	0.176 (0.129)
	.05	1.000 (0.000)	0.968 (0.023)	0.951 (0.028)	0.838 (0.051)	0.773 (0.060)	0.773 (0.060)	0.643 (0.074)
1.6S	.01	1.000 (0.000)	0.848 (0.068)	0.696 (0.095)	0.696 (0.095)	0.878 (0.061)	0.605 (0.108)	0.240 (0.148)
	.05	0.990 (0.010)	0.875 (0.036)	0.886 (0.034)	0.761 (0.049)	0.688 (0.055)	0.657 (0.058)	0.553 (0.065)

TABLE 4.9B

Relative Conservatism of Classical Fixed Sample Size MUAS

H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk

		Distribution						
		J		J-100		Unimodal		Uniform
Test	p	Low	High	Low	High	Low	High	
1.1F	.01	1.000 (0.000)	0.894 (0.075)	0.789 (0.105)	0.736 (0.117)	0.894 (0.075)	0.789 (0.105)	0.578 (0.148)
	.05	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.923 (0.055)	0.845 (0.077)	0.729 (0.102)	0.729 (0.102)
1.2F	.01	1.000 (0.000)	0.958 (0.042)	0.830 (0.084)	0.788 (0.094)	0.915 (0.060)	0.703 (0.111)	0.491 (0.145)
	.05	1.000 (0.000)	0.960 (0.029)	0.939 (0.035)	0.858 (0.053)	0.696 (0.077)	0.696 (0.077)	0.717 (0.075)
1.3F	.01	1.000 (0.000)	0.941 (0.059)	0.822 (0.102)	0.822 (0.102)	0.882 (0.084)	0.704 (0.132)	0.645 (0.144)
	.05	0.987 (0.013)	0.788 (0.052)	0.894 (0.037)	0.643 (0.067)	0.709 (0.061)	0.563 (0.073)	0.563 (0.073)
1.4F	.01	0.972 (0.028)	0.806 (0.073)	0.806 (0.073)	0.722 (0.087)	0.861 (0.062)	0.584 (0.106)	0.167 (0.147)
	.05	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.800 (0.089)	0.761 (0.097)	0.800 (0.089)
1.5F	.01	0.979 (0.021)	0.765 (0.070)	0.658 (0.084)	0.594 (0.091)	0.722 (0.076)	0.615 (0.089)	0.145 (0.130)
	.05	1.000 (0.000)	0.959 (0.029)	0.939 (0.035)	0.817 (0.061)	0.776 (0.067)	0.735 (0.073)	0.653 (0.083)
1.6F	.01	1.000 (0.000)	0.828 (0.070)	0.684 (0.094)	0.655 (0.098)	0.828 (0.070)	0.655 (0.098)	0.253 (0.143)
	.05	0.987 (0.013)	0.856 (0.043)	0.856 (0.043)	0.737 (0.058)	0.724 (0.059)	0.632 (0.068)	0.606 (0.070)

TABLE 4.10A

Relative Conservatism of Bayesian Sequential MUAS

H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk

		Distribution						
		J		J-100		Unimodal		Uniform
Test	p	Low	High	Low	High	Low	High	
2.1S	.01	0.461 (0.029)	0.334 (0.037)	0.211 (0.043)	0.193 (0.044)	0.247 (0.042)	0.022 (0.049)	-0.124 (0.053)
	.05	0.776 (0.000)	0.737 (0.000)	0.730 (0.000)	0.702 (0.000)	0.653 (0.053)	0.616 (0.060)	0.550 (0.073)
2.2S	.01	0.424 (0.014)	0.293 (0.033)	0.270 (0.035)	0.270 (0.036)	0.327 (0.031)	0.214 (0.039)	0.040 (0.051)
	.05	0.732 (0.000)	0.702 (0.000)	0.702 (0.000)	0.628 (0.034)	0.560 (0.053)	0.596 (0.045)	0.536 (0.057)
2.3S	.01	0.422 (0.012)	0.311 (0.033)	0.286 (0.035)	0.274 (0.038)	0.309 (0.034)	0.275 (0.036)	0.063 (0.053)
	.05	0.771 (0.000)	0.738 (0.015)	0.741 (0.015)	0.658 (0.035)	0.598 (0.048)	0.642 (0.041)	0.553 (0.054)
2.4S	.01	0.415 (0.000)	0.315 (0.033)	0.290 (0.036)	0.281 (0.038)	0.343 (0.030)	0.250 (0.040)	0.066 (0.057)
	.05	0.808 (0.000)	0.786 (0.014)	0.773 (0.017)	0.681 (0.034)	0.607 (0.044)	0.568 (0.048)	0.489 (0.056)
2.5S	.01	0.381 (0.000)	0.319 (0.033)	0.246 (0.046)	0.282 (0.040)	0.315 (0.033)	0.205 (0.049)	0.062 (0.064)
	.05	0.845 (0.009)	0.675 (0.034)	0.743 (0.027)	0.539 (0.044)	0.470 (0.048)	0.452 (0.049)	0.394 (0.052)
2.6S	.01	0.538 (0.045)	0.344 (0.097)	0.519 (0.044)	0.344 (0.097)	0.562 (0.000)	0.374 (0.084)	0.116 (0.128)
	.05	0.155 (0.032)	0.015 (0.031)	-0.015 (0.031)	-0.088 (0.029)	0.114 (0.032)	0.058 (0.032)	0.032 (0.031)

TABLE 4.10B

Relative Conservatism of Bayesian Fixed Sample Size MUAS

H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk

		Distribution						
		J		J-100		Unimodal		Uniform
Test	p	Low	High	Low	High	Low	High	
2.1F	.01	0.435 (0.028)	0.311 (0.036)	0.163 (0.043)	0.183 (0.042)	0.242 (0.039)	0.054 (0.046)	-0.060 (0.049)
	.05	0.440 (0.000)	0.440 (0.000)	0.440 (0.000)	0.440 (0.000)	0.387 (0.031)	0.369 (0.035)	0.334 (0.043)
2.2F	.01	0.357 (0.011)	0.232 (0.029)	0.220 (0.030)	0.214 (0.031)	0.257 (0.027)	0.183 (0.034)	0.052 (0.043)
	.05	0.437 (0.000)	0.437 (0.000)	0.437 (0.000)	0.394 (0.024)	0.366 (0.031)	0.352 (0.034)	0.366 (0.031)
2.3F	.01	0.343 (0.010)	0.247 (0.027)	0.205 (0.032)	0.212 (0.031)	0.254 (0.026)	0.205 (0.032)	0.061 (0.043)
	.05	0.525 (0.000)	0.512 (0.013)	0.512 (0.013)	0.448 (0.031)	0.410 (0.038)	0.423 (0.036)	0.372 (0.044)
2.4F	.01	0.323 (0.008)	0.252 (0.025)	0.221 (0.029)	0.205 (0.031)	0.244 (0.026)	0.213 (0.030)	0.055 (0.045)
	.05	0.631 (0.000)	0.610 (0.015)	0.599 (0.019)	0.523 (0.034)	0.501 (0.037)	0.447 (0.044)	0.404 (0.049)
2.5F	.01	0.289 (0.000)	0.250 (0.019)	0.211 (0.027)	0.182 (0.032)	0.250 (0.019)	0.192 (0.030)	0.066 (0.045)
	.05	0.743 (0.012)	0.530 (0.042)	0.612 (0.034)	0.480 (0.046)	0.448 (0.049)	0.407 (0.051)	0.390 (0.052)
2.6F	.01	0.430 (0.019)	0.333 (0.047)	0.410 (0.027)	0.294 (0.055)	0.410 (0.027)	0.294 (0.055)	0.080 (0.083)
	.05	0.301 (0.041)	0.080 (0.043)	0.088 (0.043)	0.022 (0.043)	0.123 (0.043)	0.119 (0.043)	0.095 (0.043)

TABLE 4.11

Relative Efficiency of Classical Sequential MUAS
H_1: p=.01 vs. H_2: p=.05

Table entries: mean (AVG) and maximum (MAX) of the ratio RE=(n*-ASN)/n*, where n* is the optimal fixed sample size and ASN is the average sample size, and the proportion of truncated decisions (PTD), based on 2500 replications on the control distributions

		Test					
		1.1S	1.2S	1.3S	1.4S	1.5S	1.6S
p=.01	MAX	.610	.590	.625	.542	.495	.532
	AVG	.440	.433	.467	.381	.355	.394
	PTD	.098	.157	.132	.076	.114	.098
p=.05	MAX	.984	.985	.975	.987	.981	.979
	AVG	.570	.522	.421	.594	.542	.500
	PTD	.040	.070	.090	.030	.060	.078

TABLE 4.12

Relative Efficiency of Bayesian Sequential MUAS
H_1: p=.01 vs. H_2: p=.05

Table entries: mean (AVG) and maximum (MAX) of the ratio RE=(n*-ASN)/n*, where n* is the optimal fixed sample size and ASN is the average sample size, and the proportion of truncated decisions (PTD), based on 2500 replications on the control distributions

		Test					
		2.1S	2.2S	2.3S	2.4S	2.5S	2.6S
p=.01	MAX	.232	.425	.482	.539	.602	.676
	AVG	--+	.300	.348	.402	.466	.618
	PTD	.235	.127	.117	.102	.084	.082
p=.05	MAX	.989	.983	.982	.980	.977	.941
	AVG	.674	.600	.563	.510	.455	--+
	PTD	.041	.041	.050	.065	.084	.144

+efficiency measures omitted (see text)

TABLE 4.13

Relative Conservatism of Classical MUAS: Degenerate Distributions
H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk for sequential (S) and fixed sample size (F) tests

		Distribution					
		.3		.5		.8	
Test	p	S	F	S	F	S	F
1.1	.01	1.000 (0.000)	1.000 (0.000)	0.699 (0.122)	0.683 (0.129)	0.548 (0.149)	0.472 (0.165)
	.05	1.000 (0.000)	1.000 (0.000)	0.877 (0.061)	0.845 (0.077)	0.632 (0.105)	0.652 (0.115)
1.2	.01	1.000 (0.000)	1.000 (0.000)	0.479 (0.149)	0.534 (0.139)	0.436 (0.154)	0.491 (0.145)
	.05	0.985 (0.015)	1.000 (0.000)	0.831 (0.050)	0.798 (0.063)	0.616 (0.075)	0.534 (0.095)
1.3	.01	1.000 (0.000)	1.000 (0.000)	0.483 (0.181)	0.585 (0.156)	0.612 (0.157)	0.526 (0.166)
	.05	0.960 (0.020)	0.960 (0.023)	0.820 (0.042)	0.775 (0.054)	0.319 (0.077)	0.246 (0.094)
1.4	.01	0.921 (0.046)	0.917 (0.048)	0.499 (0.113)	0.473 (0.119)	0.262 (0.135)	0.445 (0.122)
	.05	1.000 (0.000)	1.000 (0.000)	0.907 (0.053)	0.880 (0.069)	0.722 (0.092)	0.681 (0.112)
1.5	.01	0.892 (0.048)	0.893 (0.048)	0.197 (0.127)	0.210 (0.125)	0.306 (0.119)	0.551 (0.096)
	.05	0.968 (0.023)	0.980 (0.020)	0.886 (0.043)	0.857 (0.054)	0.497 (0.088)	0.307 (0.115)
1.6	.01	0.939 (0.043)	0.943 (0.041)	0.392 (0.133)	0.368 (0.132)	0.361 (0.136)	0.569 (0.110)
	.05	0.969 (0.018)	0.974 (0.019)	0.823 (0.042)	0.790 (0.052)	0.314 (0.079)	0.199 (0.096)

TABLE 4.14

Relative Conservatism of Bayesian MUAS: Degenerate Distributions

H_1: p=.01 vs. H_2: p=.05

Table entries: mean and standard deviation (in parentheses) of the ratio RC=(nominal risk-observed risk)/nominal risk for both sequential (S) and fixed sample size (F) tests

		Distribution					
		.3		.5		.8	
Test	p	S	F	S	F	S	F
2.1	.01	0.098 (0.046)	0.079 (0.045)	-0.372 (0.056)	-0.213 (0.052)	-0.220 (0.054)	-0.268 (0.053)
	.05	0.793 (0.000)	0.440 (0.000)	0.786 (0.034)	0.387 (0.031)	0.604 (0.070)	0.369 (0.035)
2.2	.01	0.383 (0.019)	0.326 (0.018)	0.072 (0.047)	0.083 (0.041)	-0.070 (0.050)	0.183 (0.034)
	.05	0.742 (0.000)	0.437 (0.000)	0.683 (0.037)	0.394 (0.024)	0.447 (0.074)	0.239 (0.052)
2.3	.01	0.380 (0.018)	0.323 (0.015)	0.092 (0.049)	0.095 (0.041)	0.157 (0.047)	0.205 (0.032)
	.05	0.767 (0.015)	0.525 (0.000)	0.737 (0.031)	0.474 (0.025)	0.397 (0.073)	0.206 (0.062)
2.4	.01	0.365 (0.018)	0.307 (0.014)	0.072 (0.053)	0.079 (0.043)	0.145 (0.052)	0.181 (0.034)
	.05	0.788 (0.014)	0.621 (0.011)	0.730 (0.033)	0.523 (0.034)	0.267 (0.071)	0.122 (0.071)
2.5	.01	0.361 (0.000)	0.289 (0.000)	0.112 (0.053)	0.114 (0.040)	0.187 (0.051)	0.163 (0.035)
	.05	0.784 (0.022)	0.694 (0.023)	0.682 (0.035)	0.587 (0.037)	0.157 (0.060)	0.062 (0.069)
2.6	.01	0.562 (0.000)	0.449 (0.000)	0.191 (0.111)	0.158 (0.074)	-0.332 (0.183)	-0.173 (0.106)
	.05	0.297 (0.032)	0.456 (0.038)	0.352 (0.032)	0.406 (0.039)	0.169 (0.032)	0.266 (0.042)

TABLE 4.15

Empirical and Theoretical Power Functions of Test 1.1

		β(p)			ASN		
Test	p	(1)	(2)	(3)	(1)	(2)	(3)
1.1F	.005	0.000	0.000	0.002	182	182	182
	.01	0.000	0.016	0.037	182	182	182
	.02	0.116	0.274	0.301	182	182	182
	.025	0.534	0.478	0.479	182	182	182
	.03	0.794	0.700	0.640	182	182	182
	.04	1.000	0.926	0.856	182	182	182
	.05	1.000	0.986	0.952	182	182	182
	.06	1.000	0.998	0.986	182	182	182
	.07	---	0.998	0.996	---	182	182
1.1S	.005	0.000	0.000	0.003	82	88	88
	.01	0.000	0.020	0.039	102	113	105
	.02	0.118	0.276	0.287	166	137	124
	.025	0.532	0.462	0.454	173	134	123
	.03	0.794	0.692	0.607	159	131	117
	.04	1.000	0.914	0.824	113	99	100
	.05	1.000	0.980	0.930	72	75	82
	.06	1.000	0.998	0.973	53	61	67
	.07	---	0.998	0.990	--	49	55

Legend:
(1) low J relative error distribution
(2) uniform relative error distribution
(3) theoretical results, i.e. assuming a binomial error distribution; for ASN, the bound given by (33) in section 3.2 is used

Note: empirical results are based on 500 replications; standard deviations do not exceed .0225 and are quite low in the tails; results for the low J distribution for p=.07 were not obtained

FIGURE 4.1A

Low Variance J Distribution

p=.01

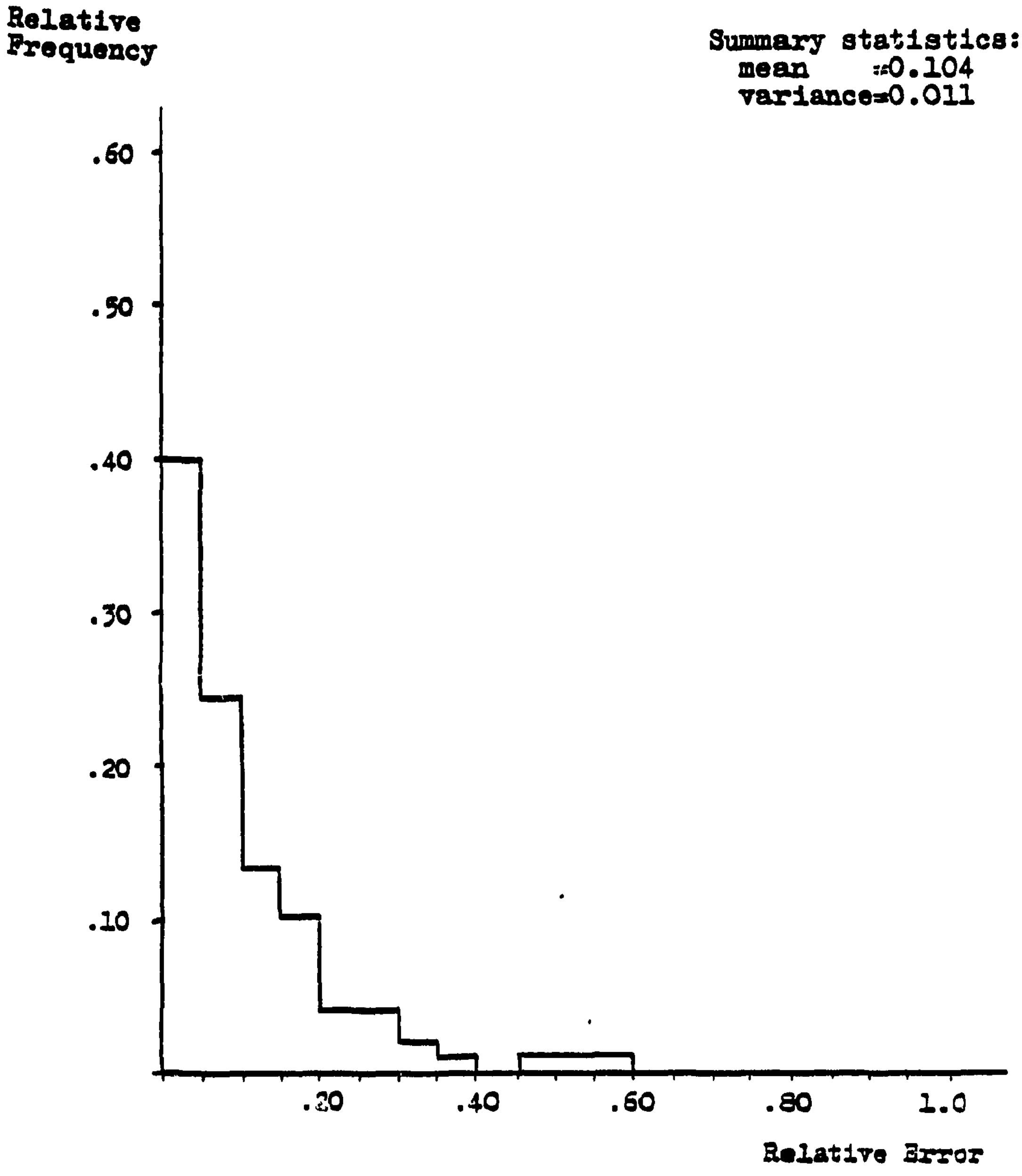


FIGURE 4.1B

Low Variance J Distribution

p=.05

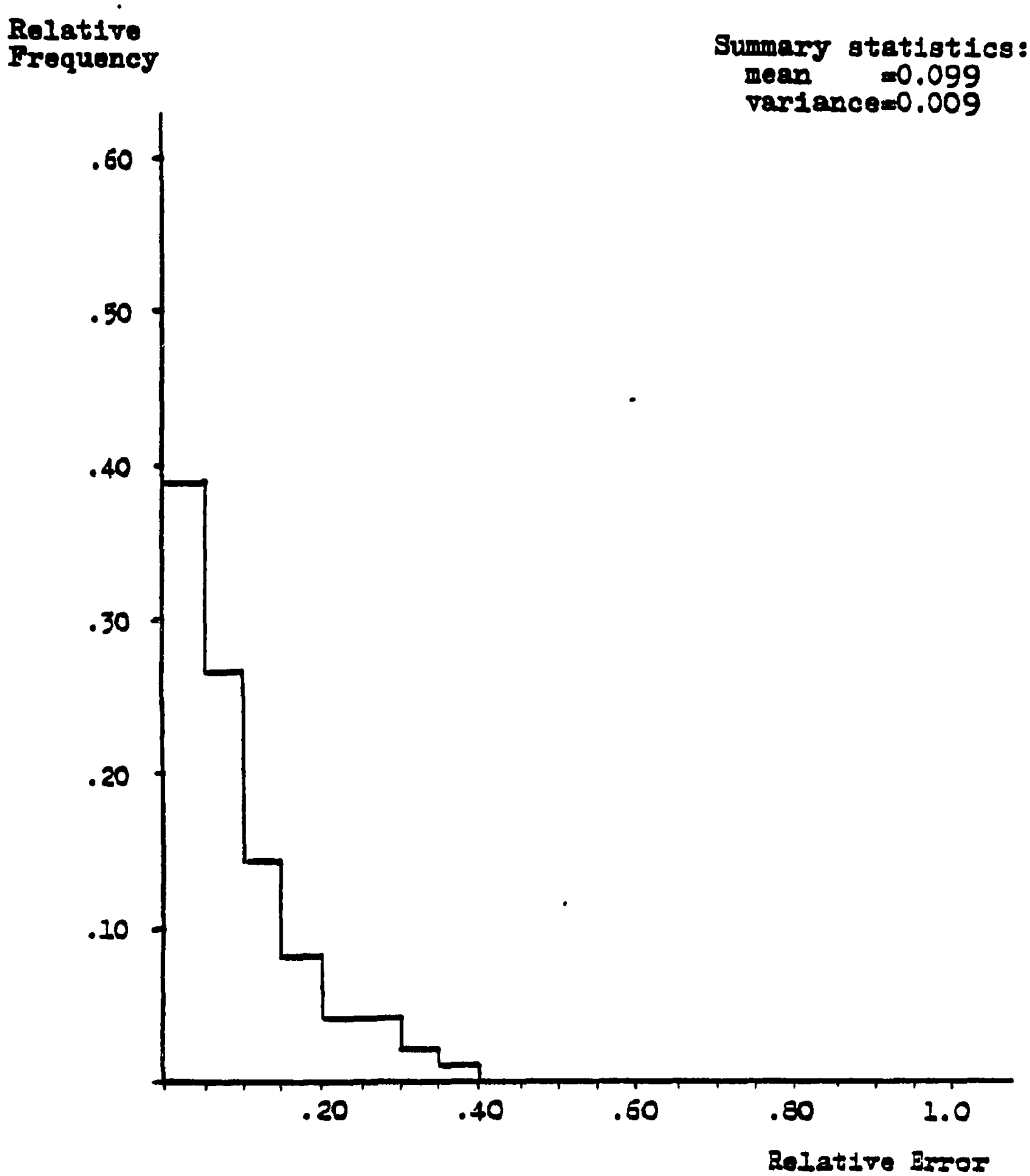


FIGURE 4.2A

High Variance J Distribution

p=.01

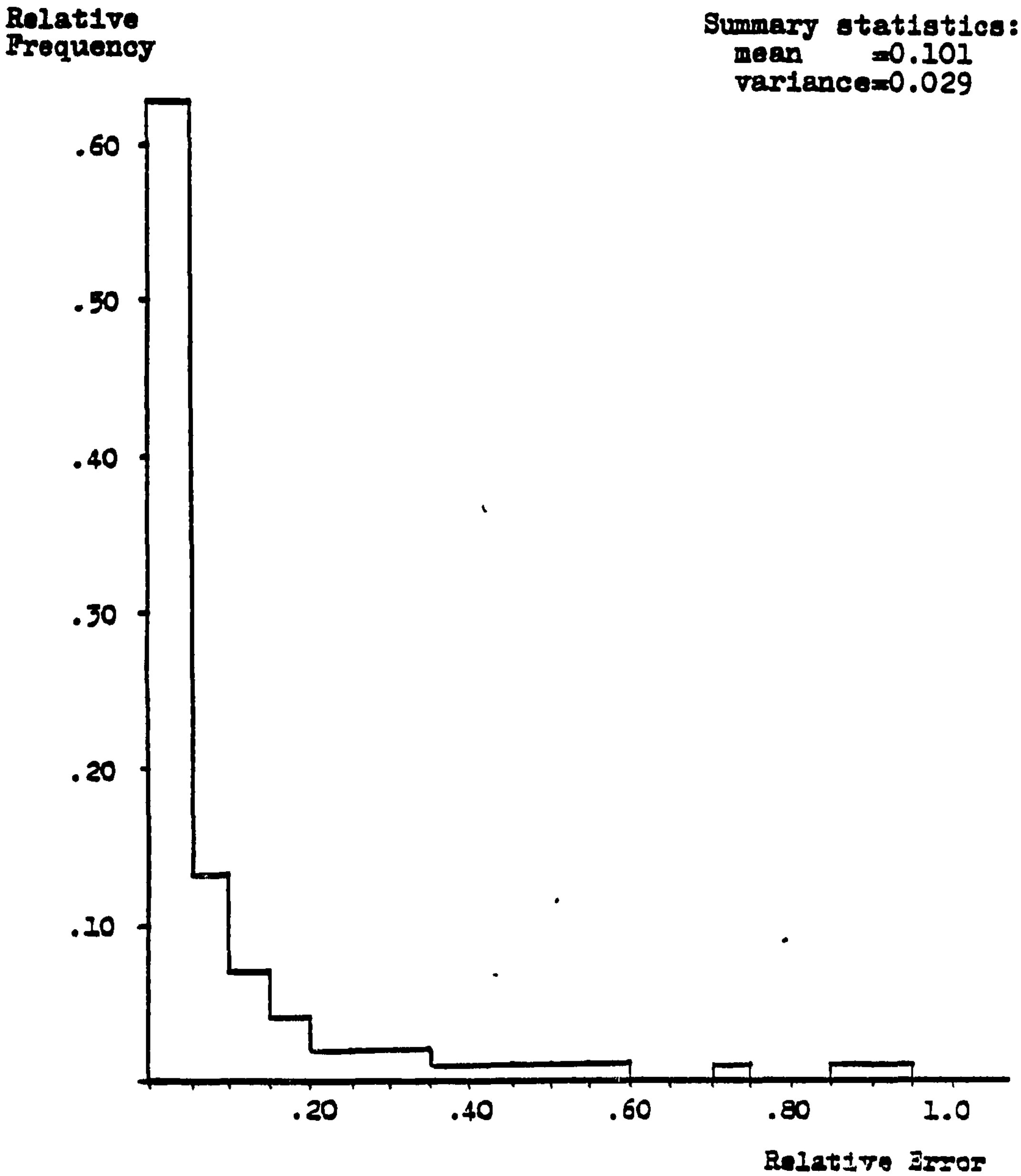


FIGURE 4.2B

High Variance J Distribution

p=.05

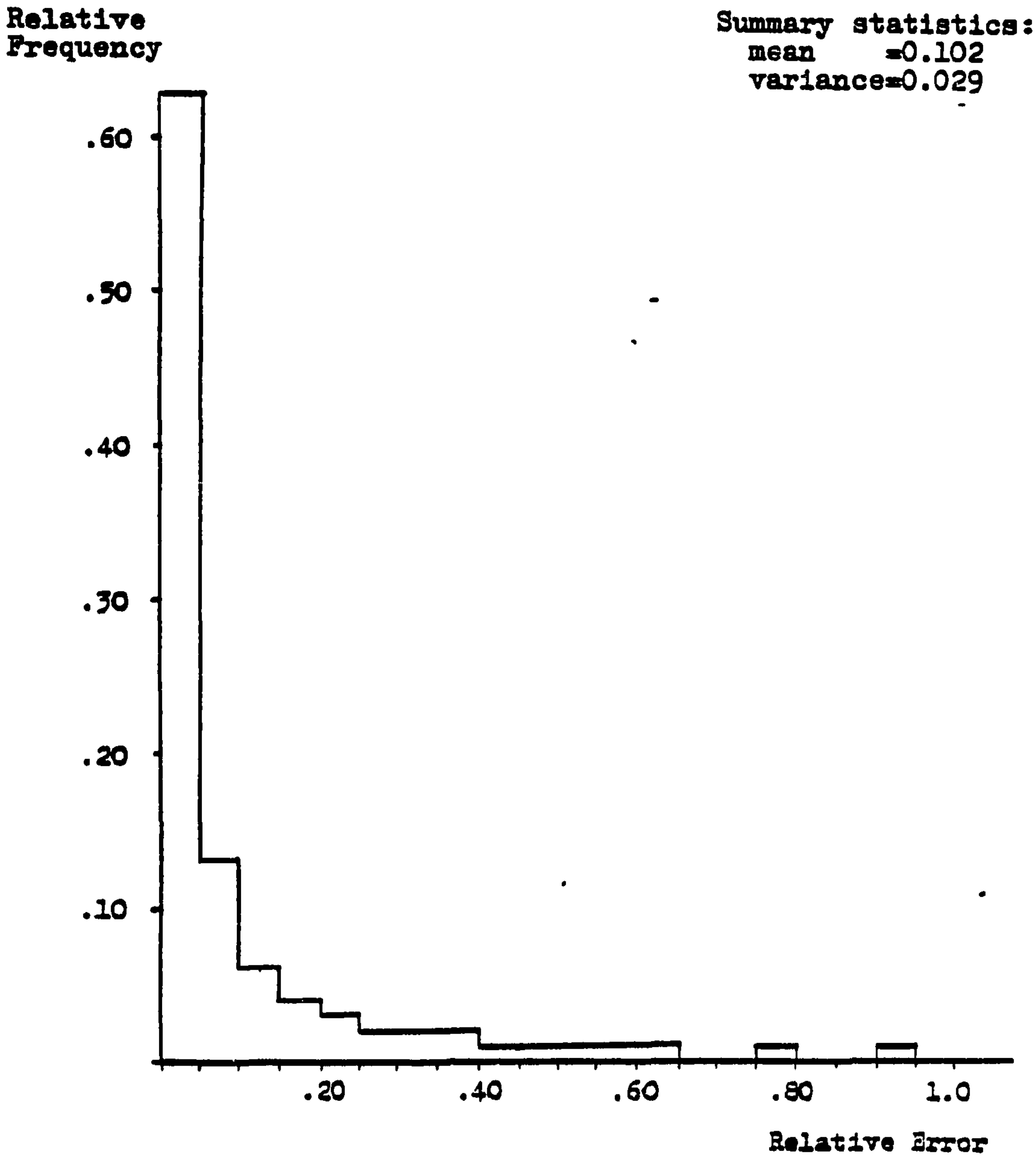


FIGURE 4.3A

Low Variance J-100 Distribution

p=.01

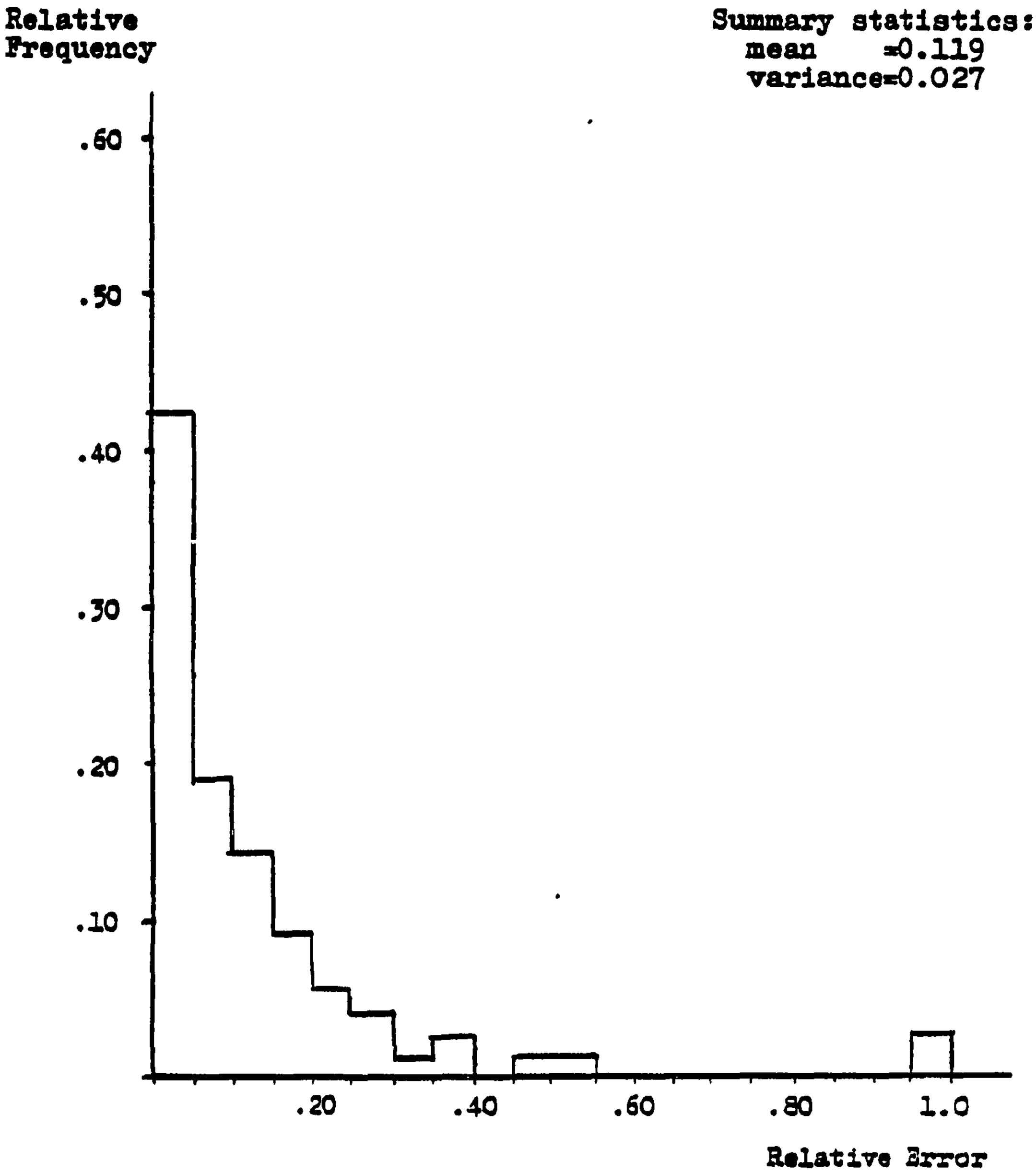


FIGURE 4.3B

Low Variance J-100 Distribution

p=.05

Summary statistics:
mean =0.131
variance=0.032

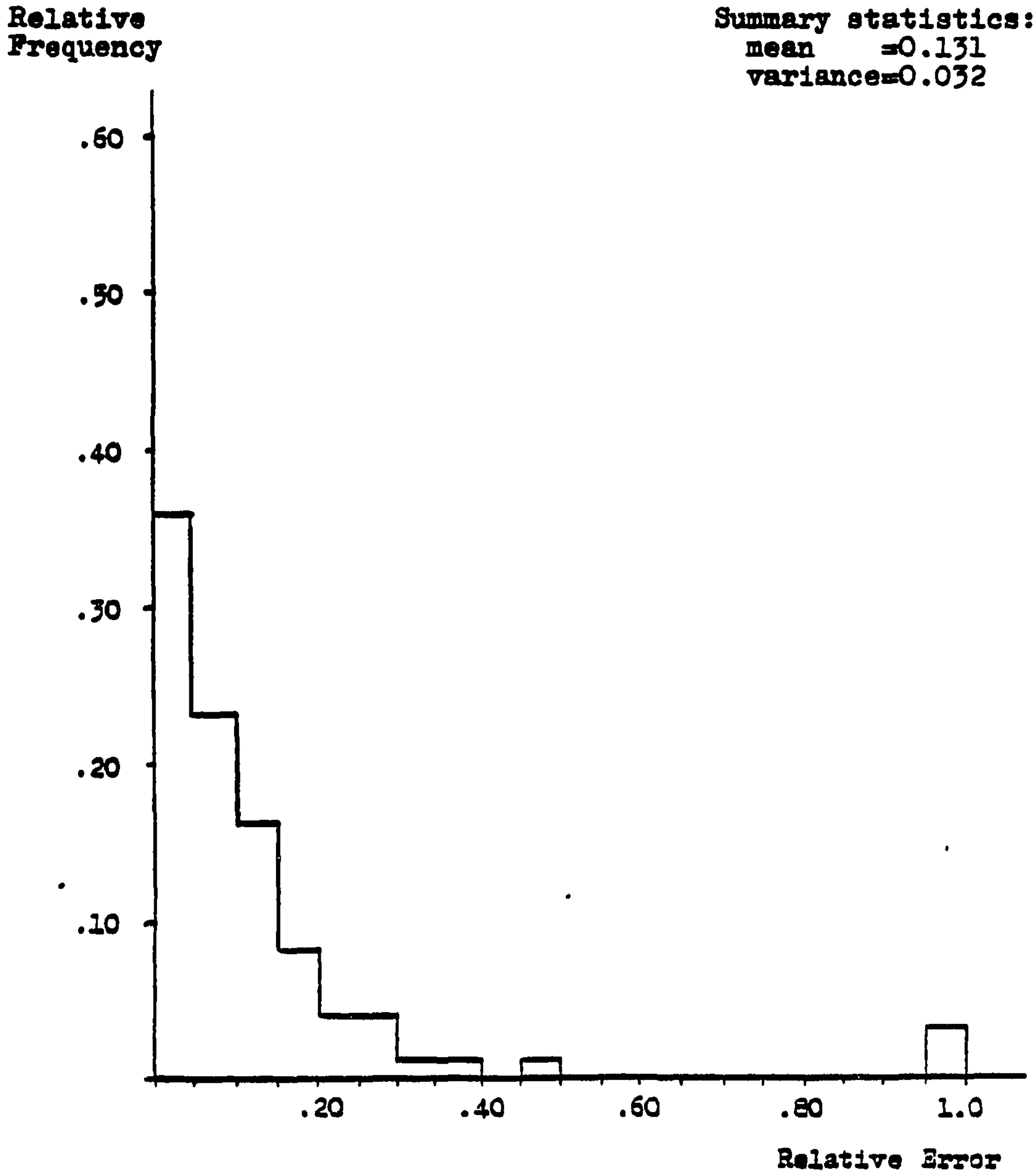


FIGURE 4.4A

High Variance J-100 Distribution

p=.01

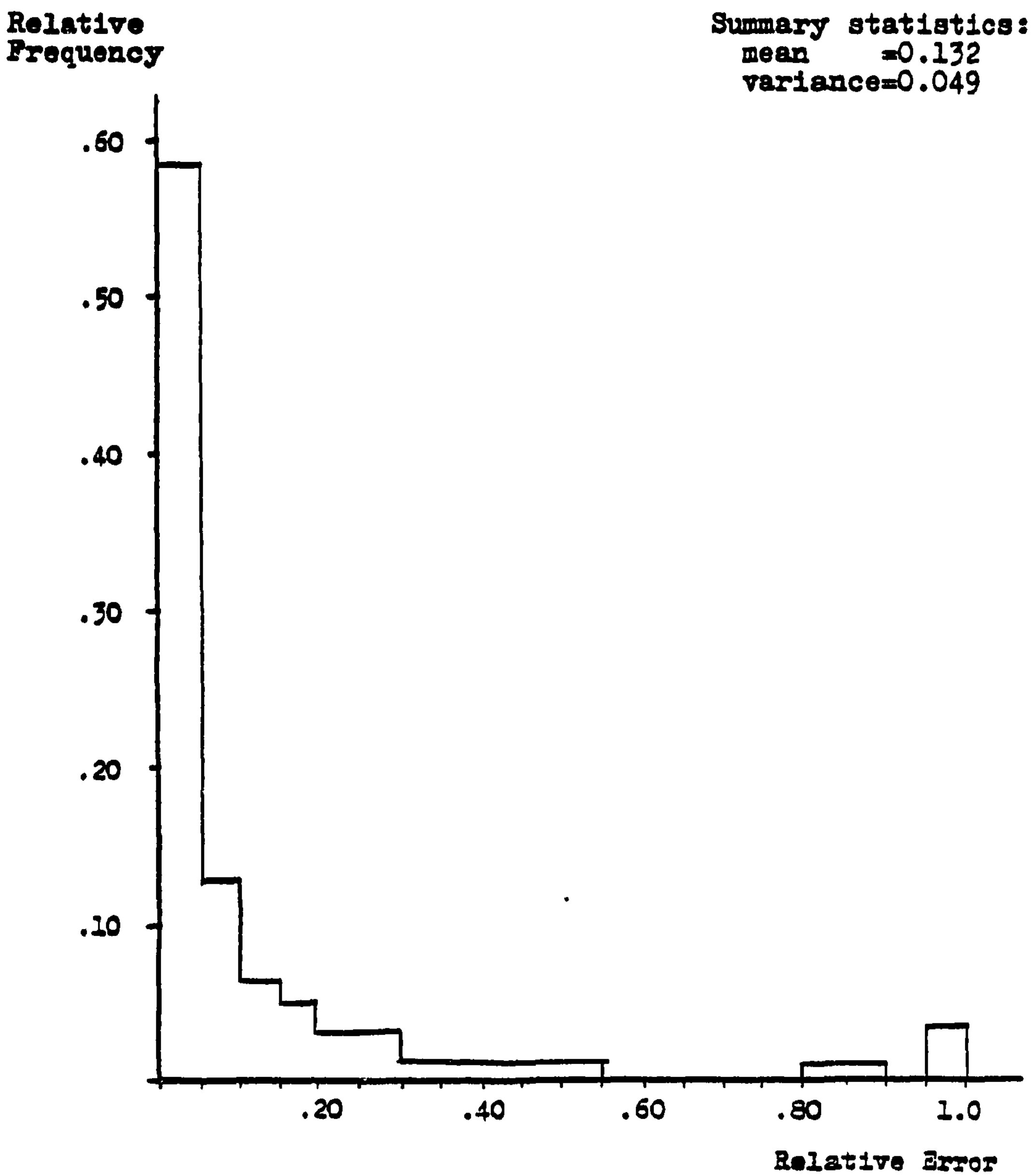


FIGURE 4.4B

High Variance J-100 Distribution

p=.05

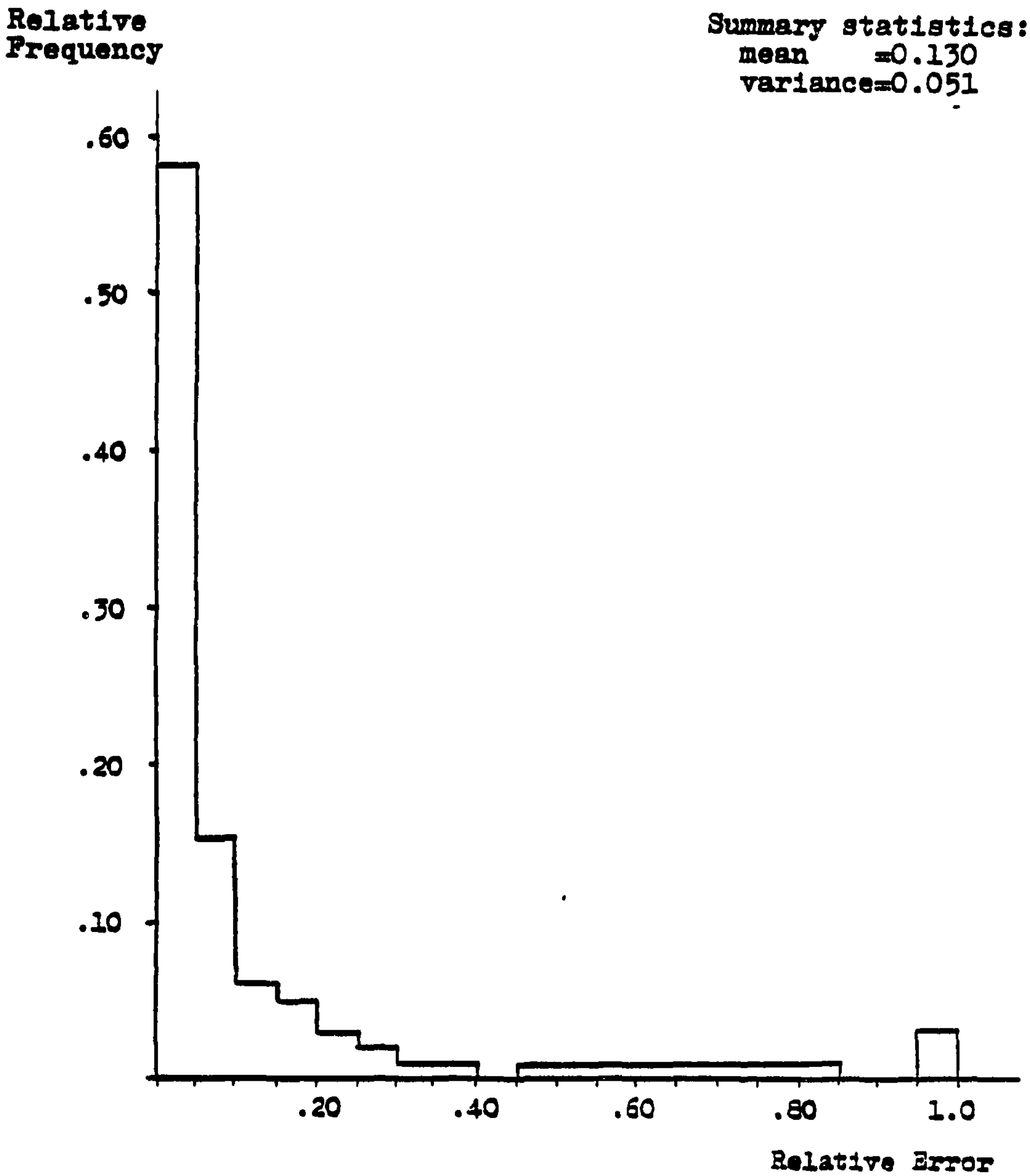


FIGURE 4.5A

Low Variance Unimodal Distribution

p=.01

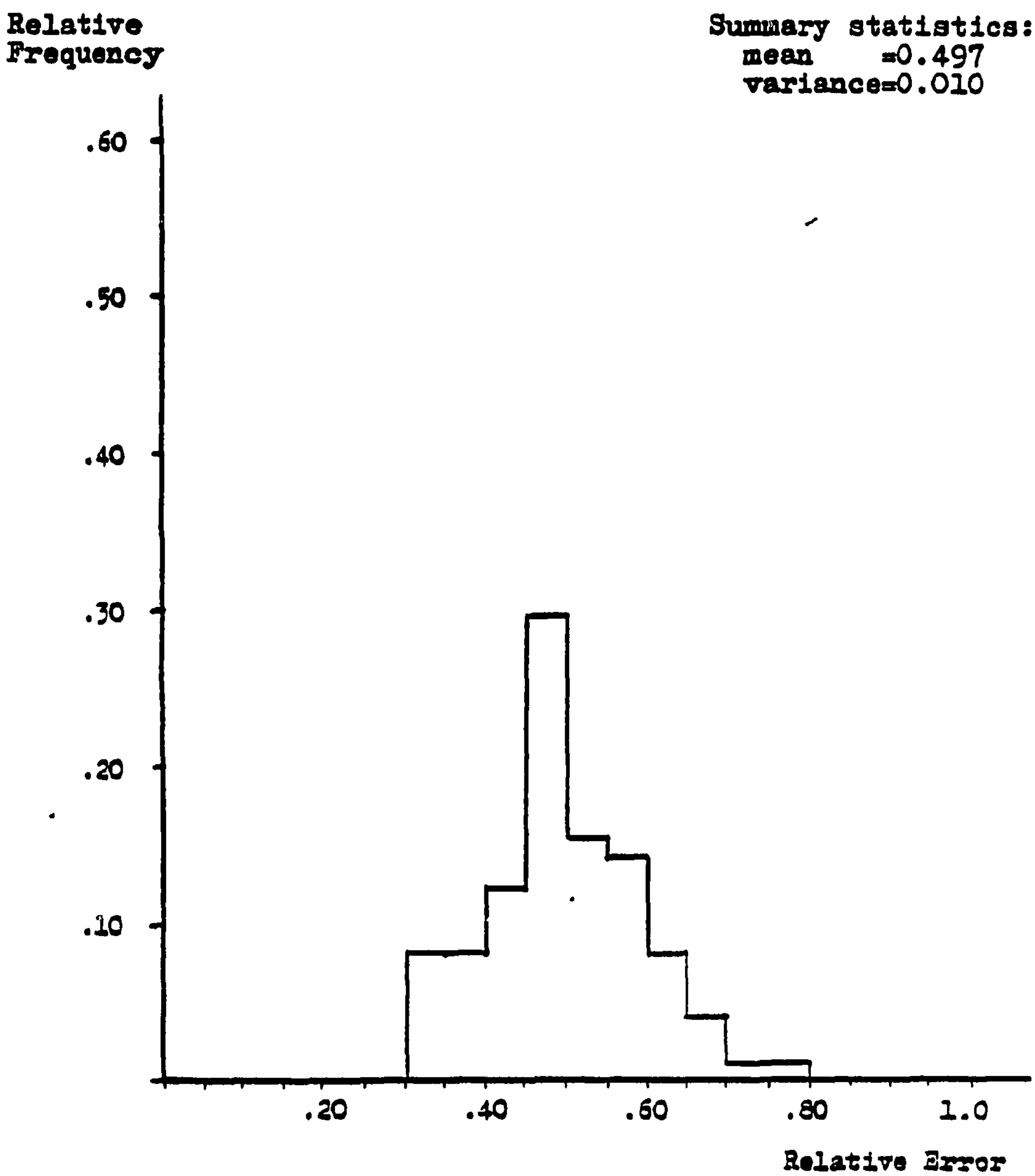


FIGURE 4.5B

Low Variance Unimodal Distribution

p=.05

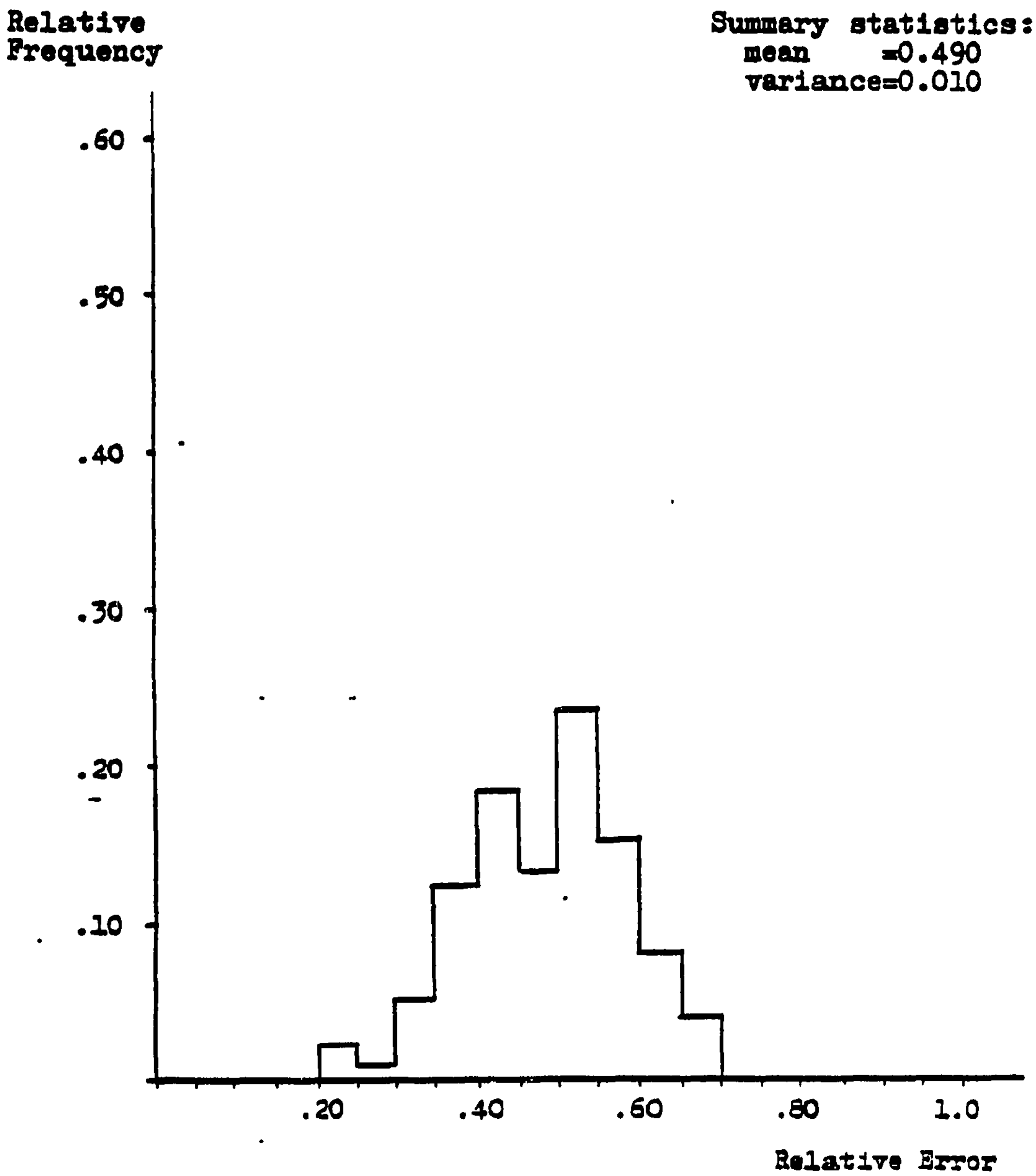


FIGURE 4.6A

High Variance Unimodal Distribution

p=.01

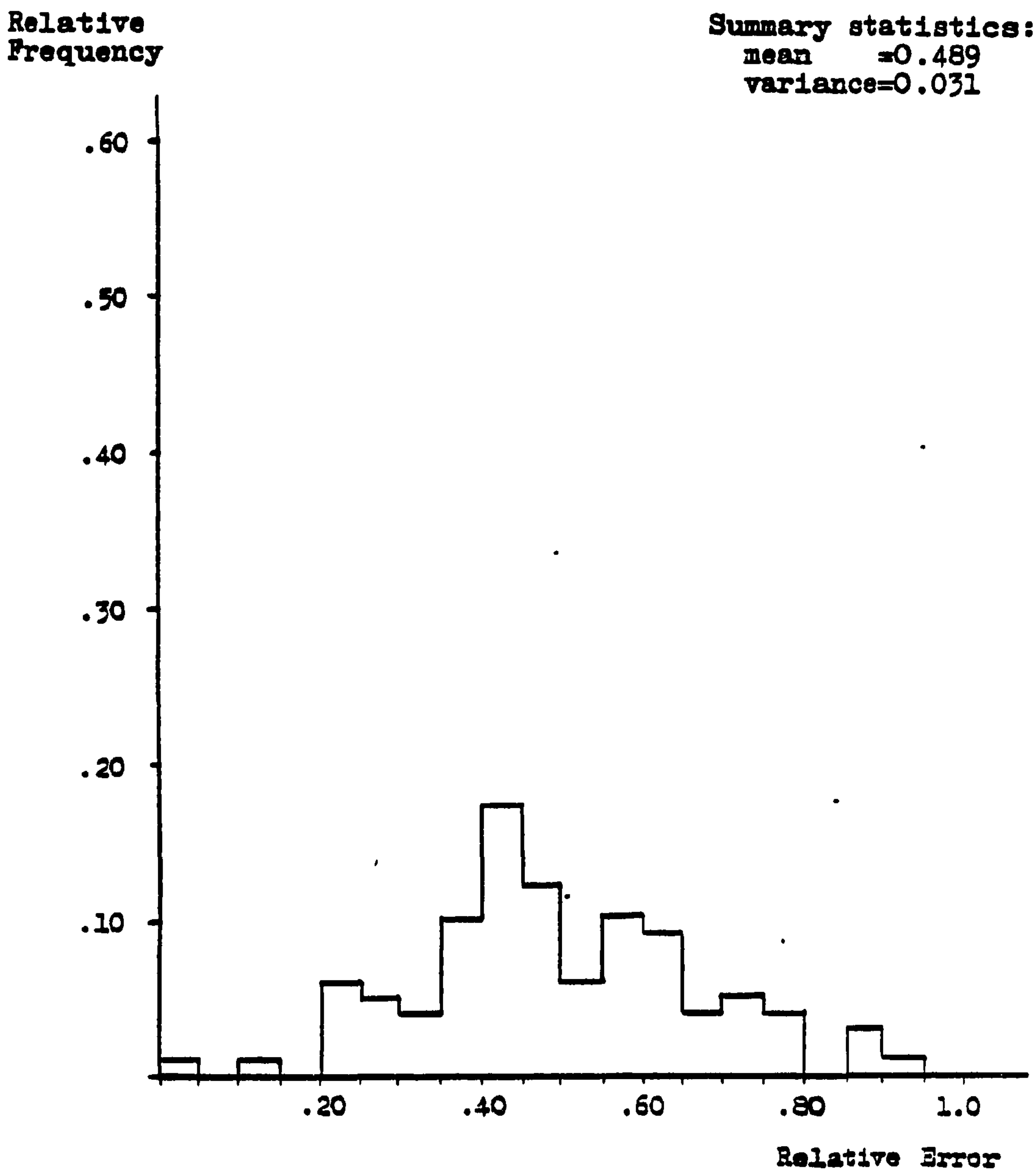


FIGURE 4.6B

High Variance Unimodal Distribution

p=.05

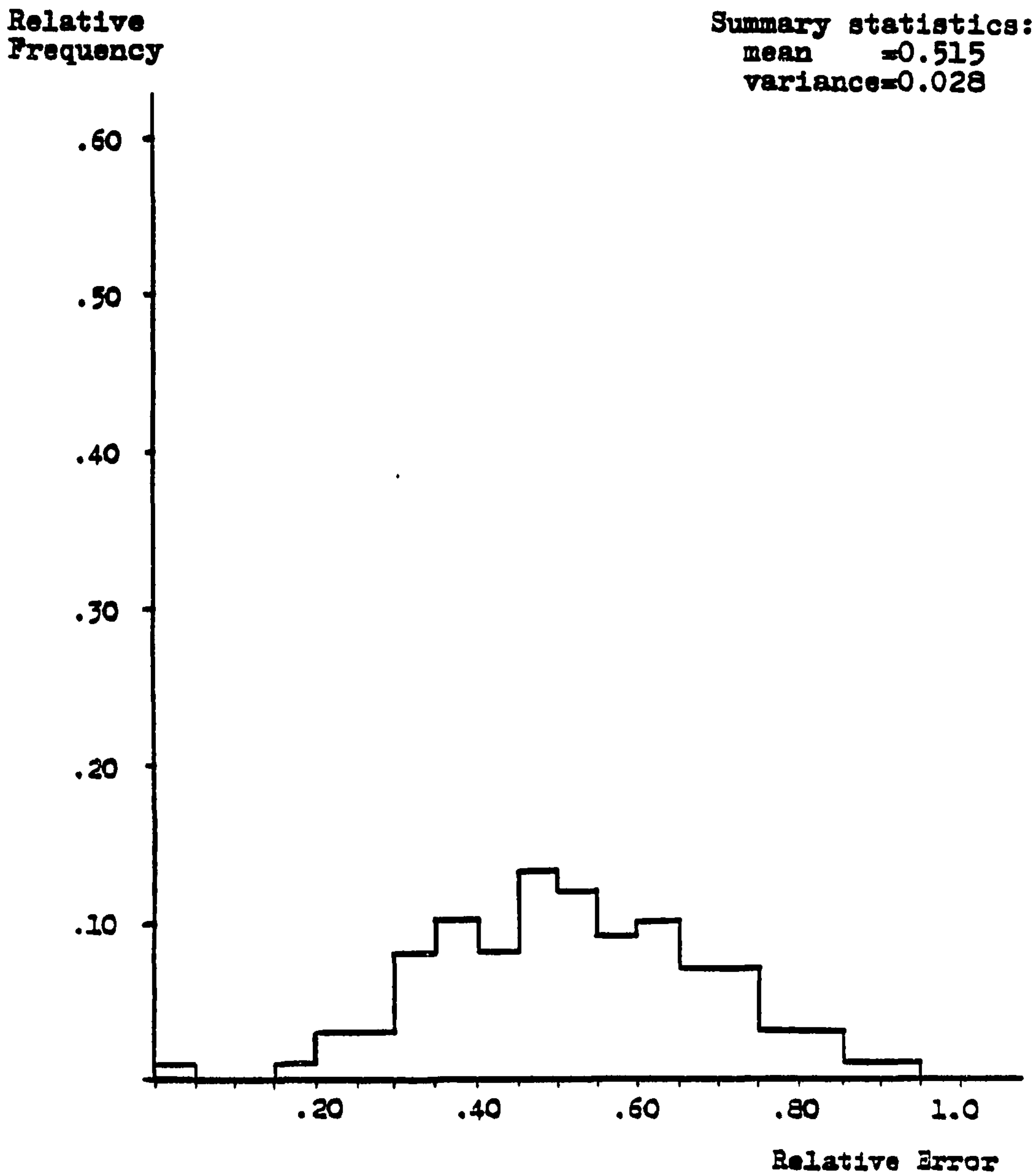


FIGURE 4.7A

Uniform Distribution

p=.01

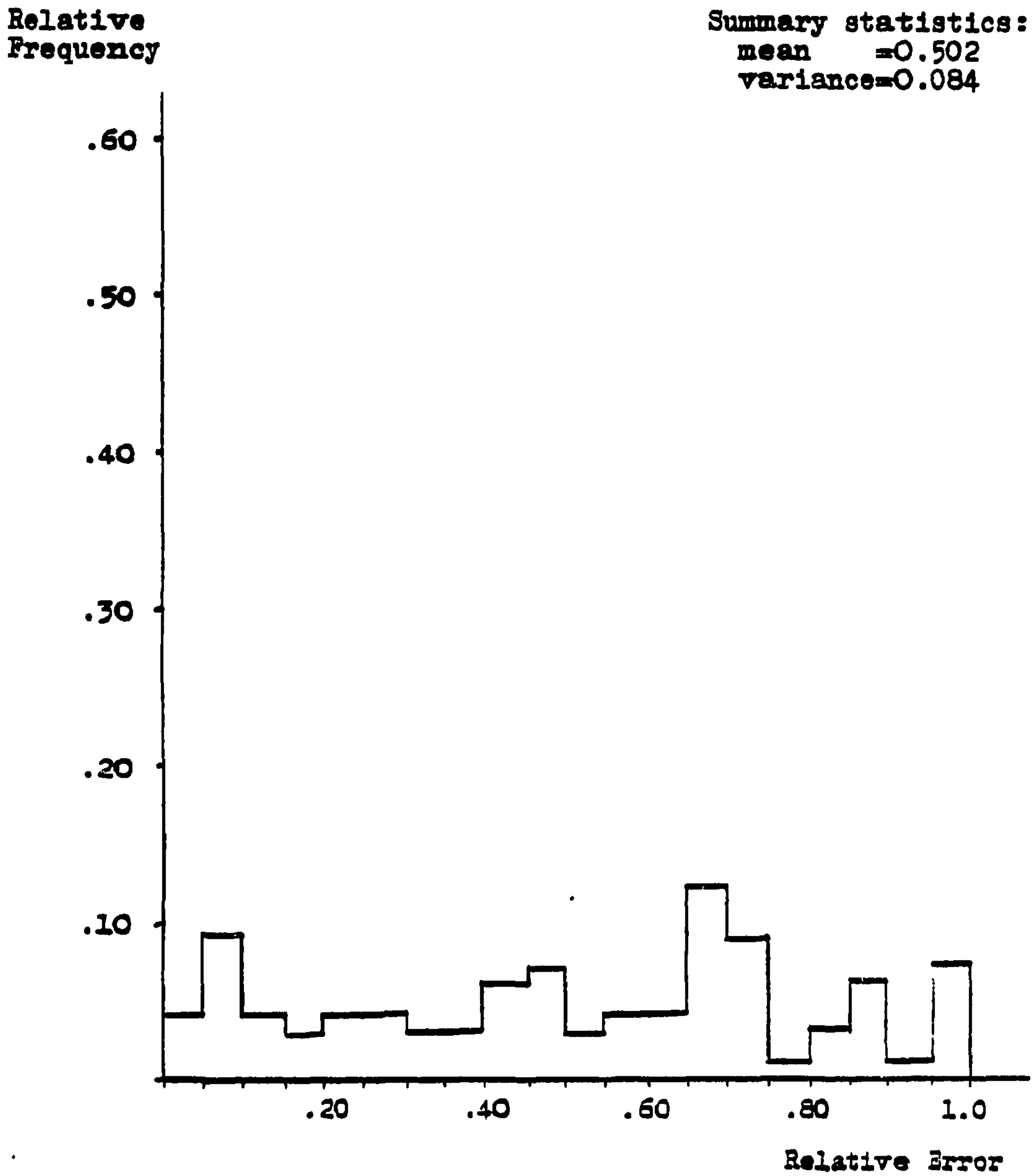


FIGURE 4.7B

Uniform Distribution

p=.05

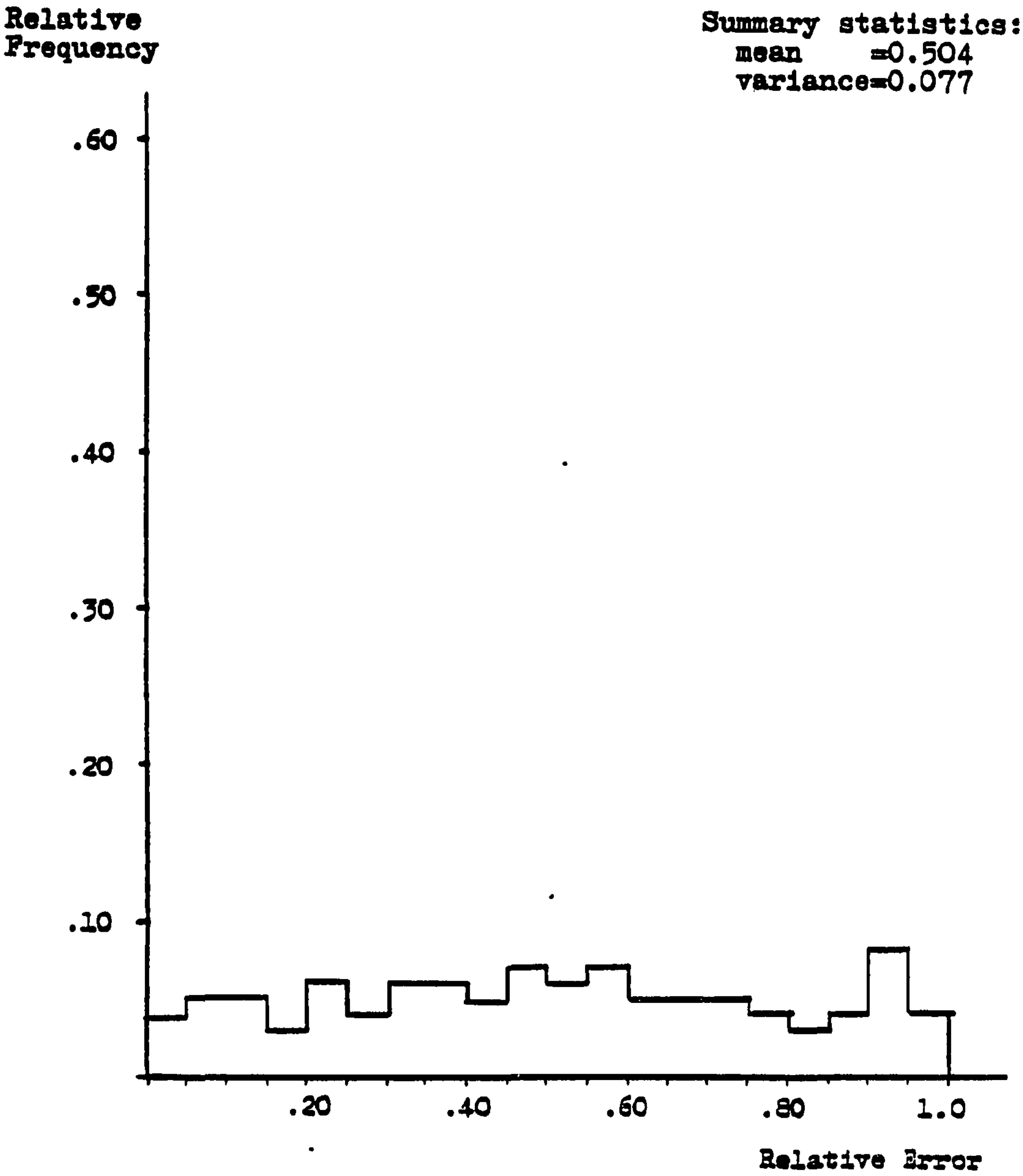
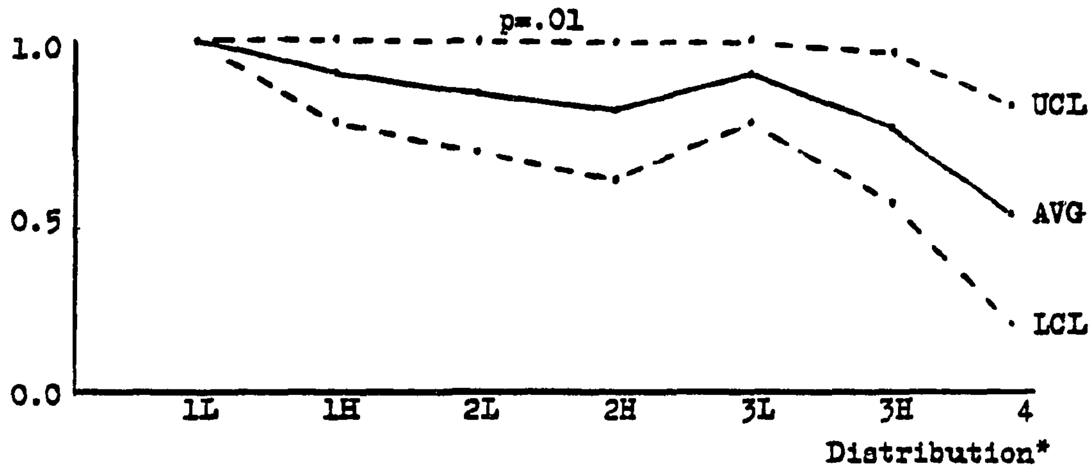
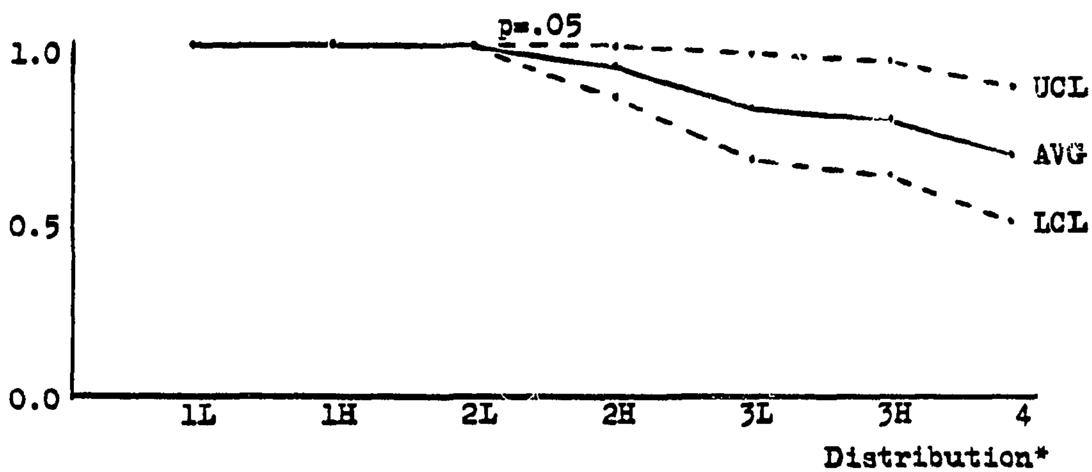


FIGURE 4.8A

Relative Conservatism of Classical MUAS: Test 1.1S
$\alpha=.040/\beta=.065$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



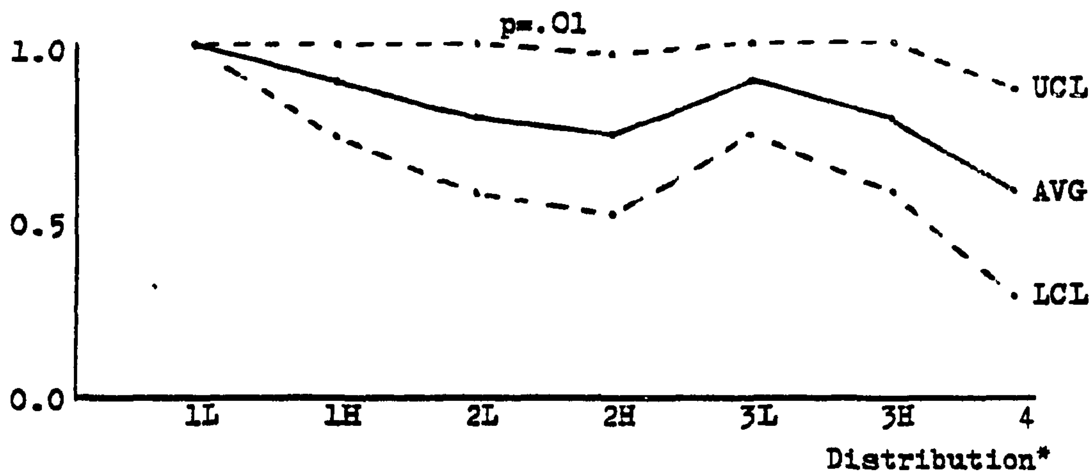


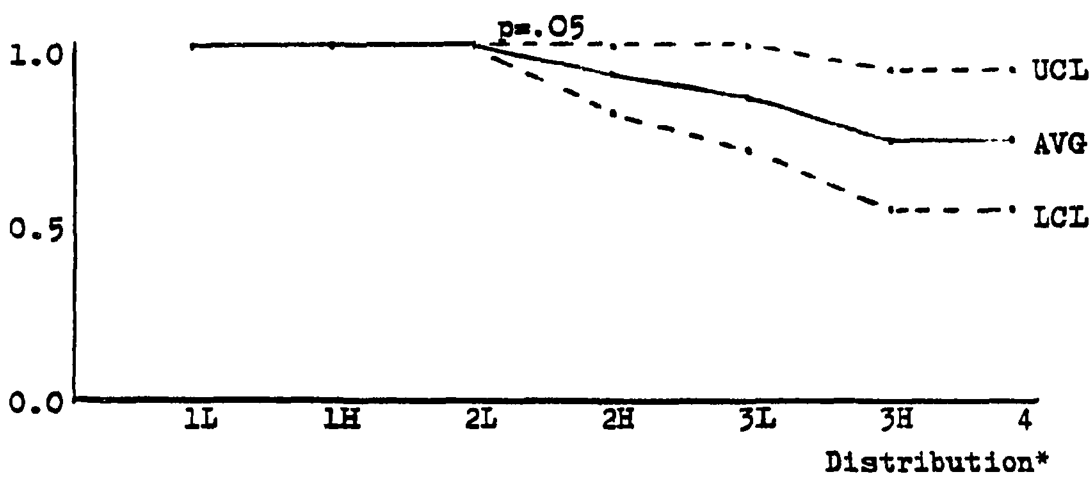
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.8B

Relative Conservatism of Classical MUAS: Test 1.1F
$\alpha=.038/\beta=.052$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



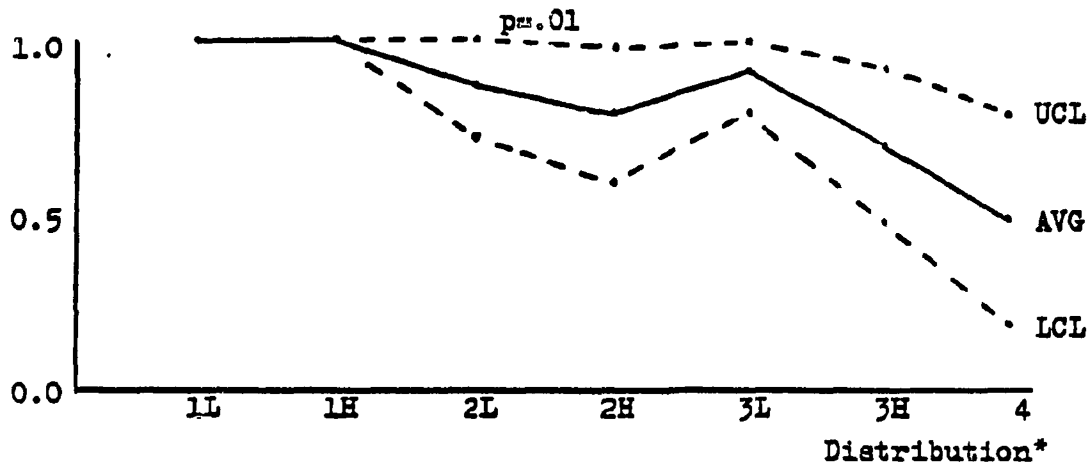


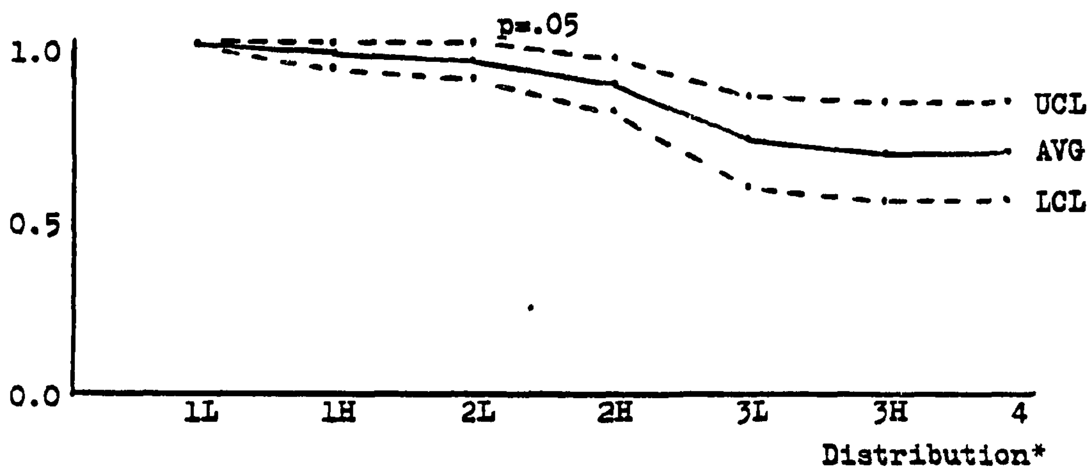
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.9A

Relative Conservatism of Classical MUAS: Test 1.2S
α=.046/β=.130

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



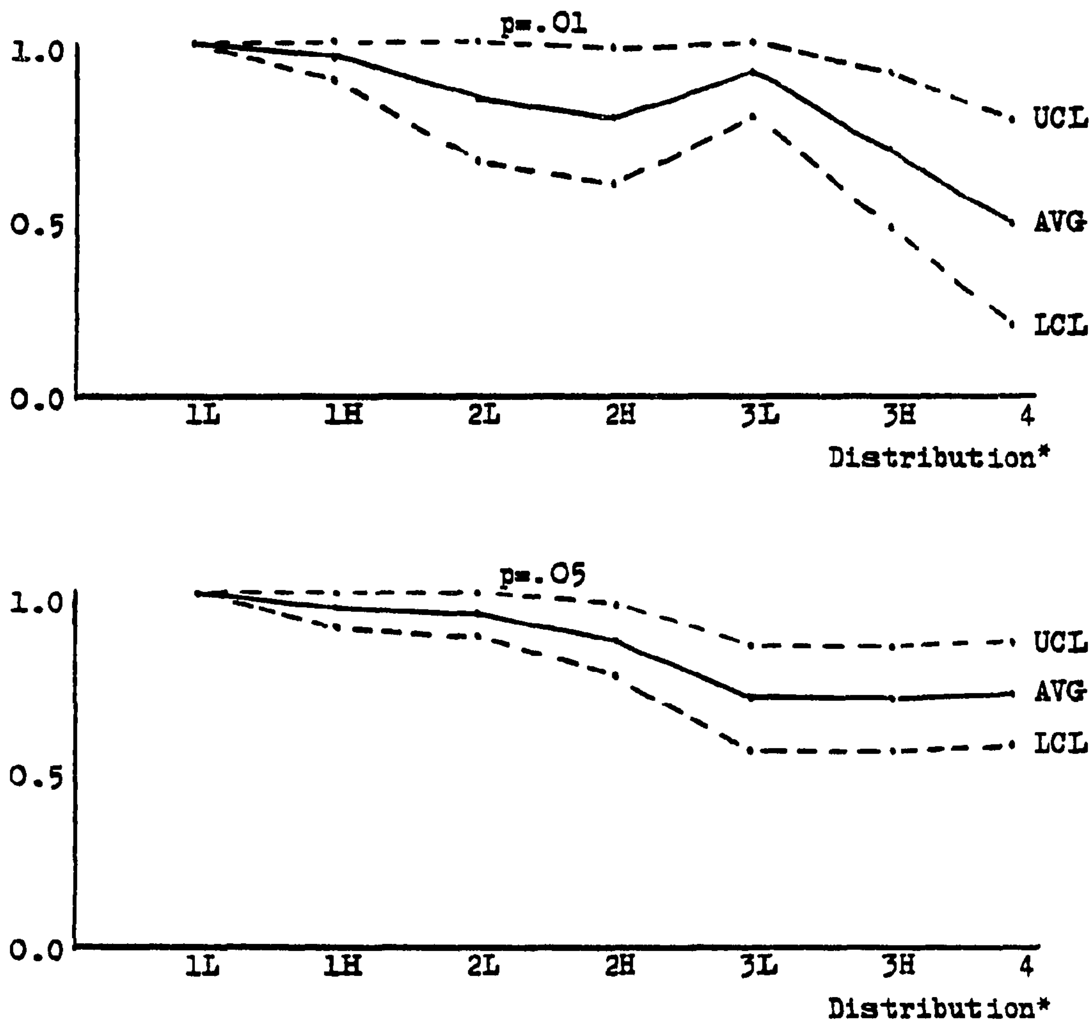


*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.9B

Relative Conservatism of Classical MUAS: Test 1.2F
$\alpha=.047/\beta=.099$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



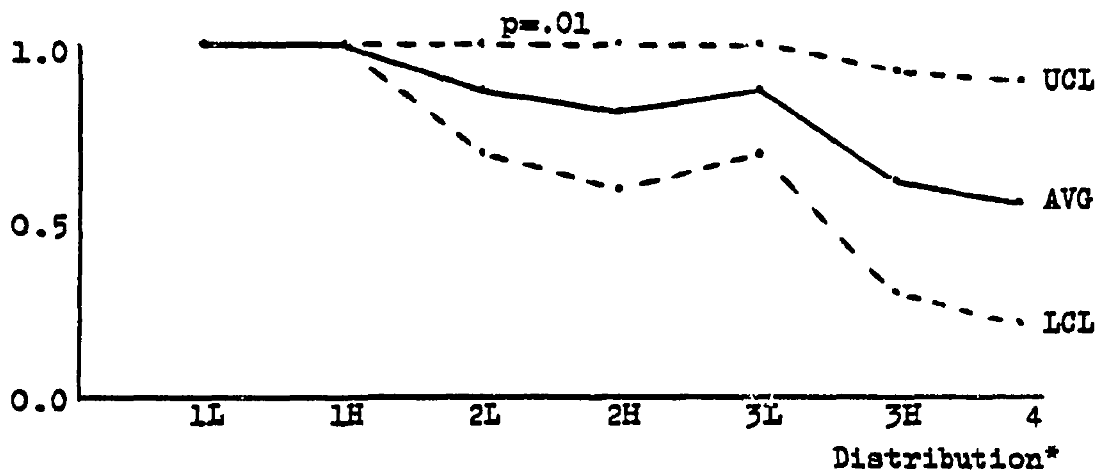
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

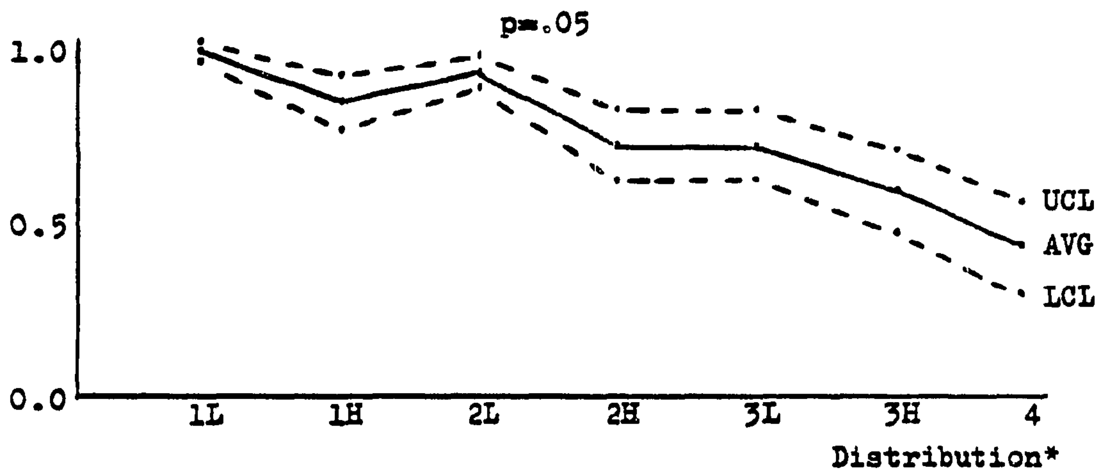
FIGURE 4.10A

Relative Conservatism of Classical MUAS: Test 1.3S

$\alpha=.031/\beta=.200$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



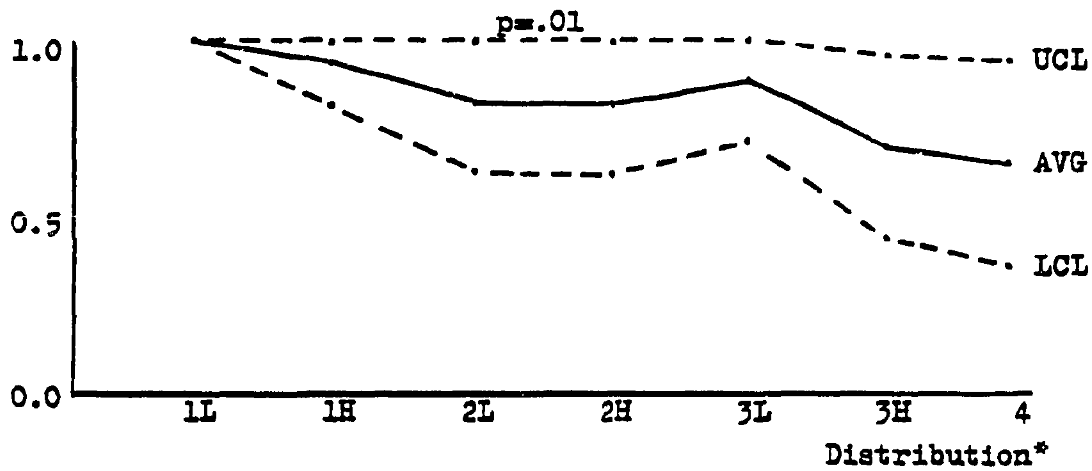


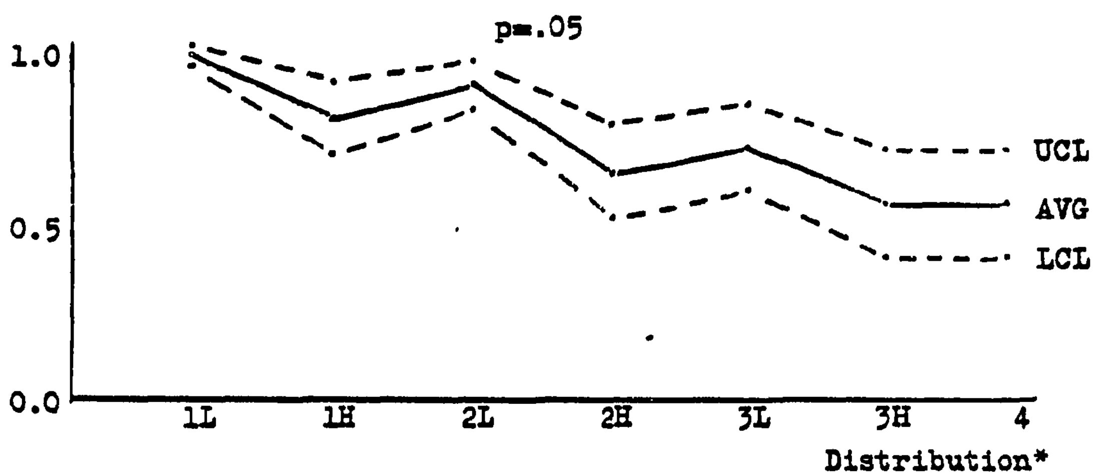
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.10B

Relative Conservatism of Classical MUAS: Test 1.3F
$\alpha=.034/\beta=.151$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



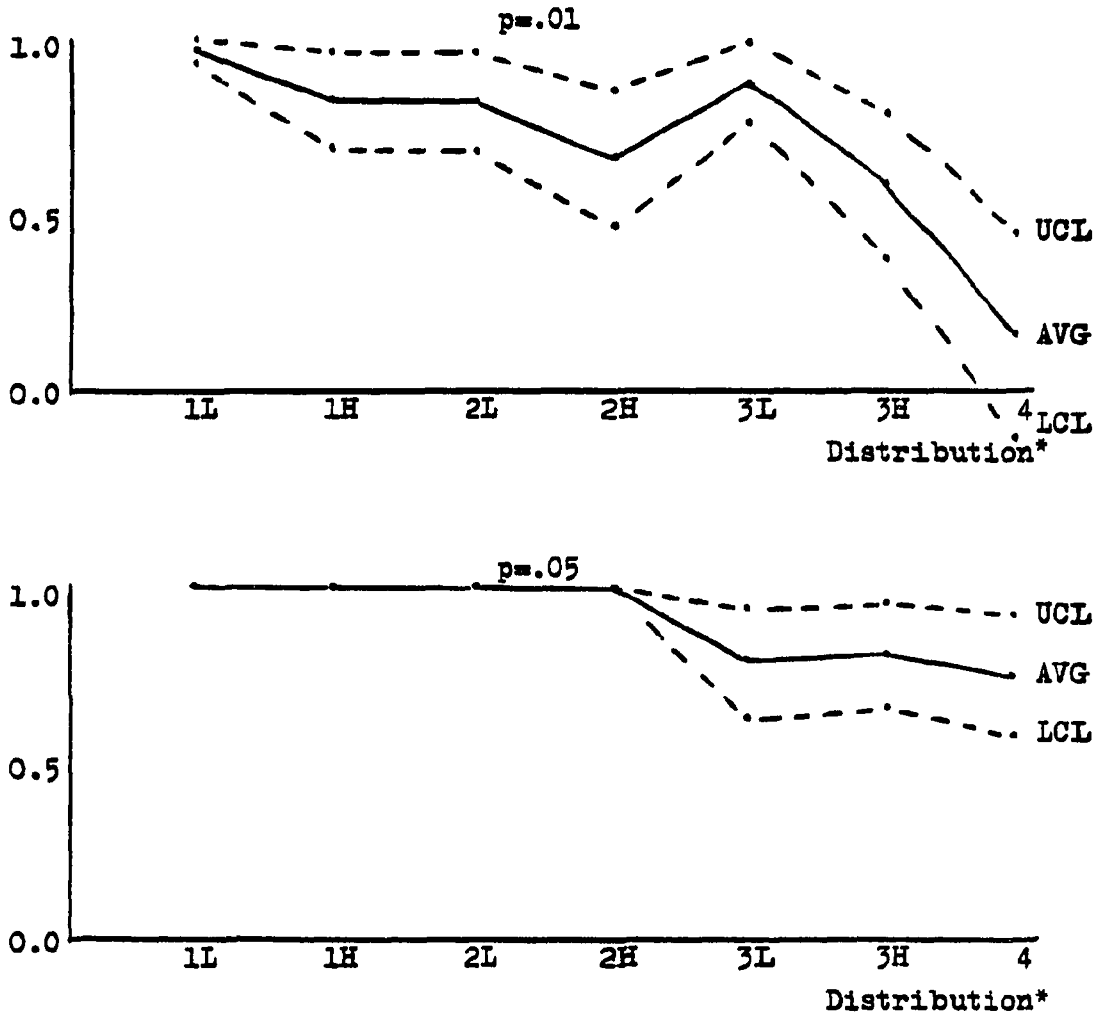


*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.11A

Relative Conservatism of Classical MUAS: Test 1.4S
α=.076/β=.065

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



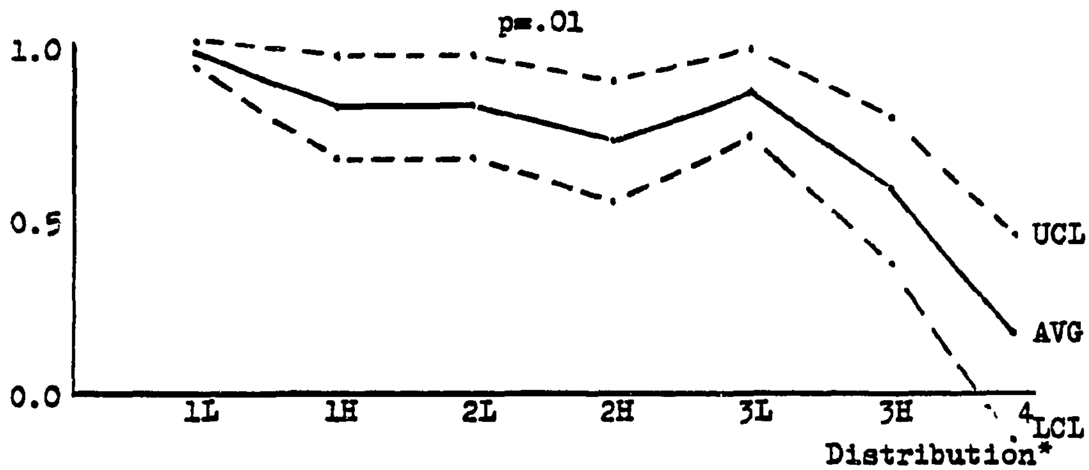
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

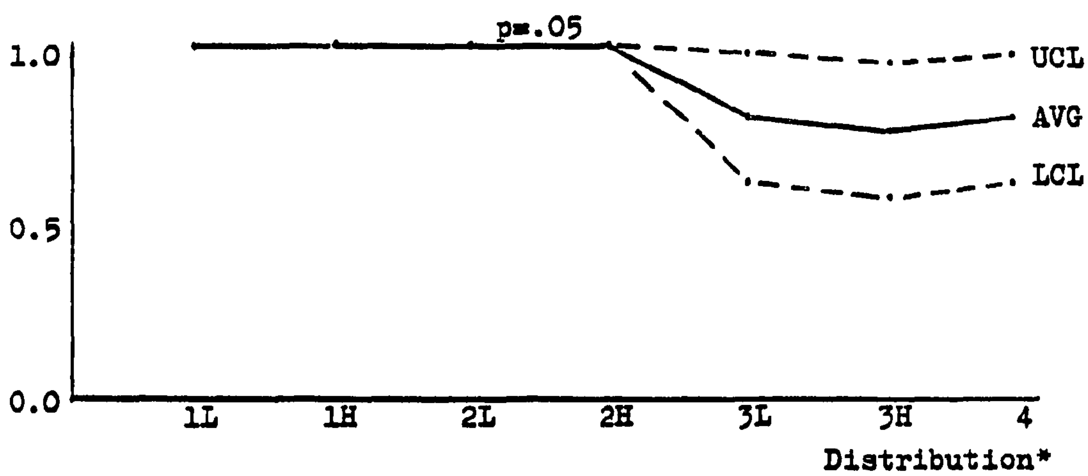
FIGURE 4.11B

Relative Conservatism of Classical MUAS: Test 1.4F

$\alpha=.072/\beta=.050$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk





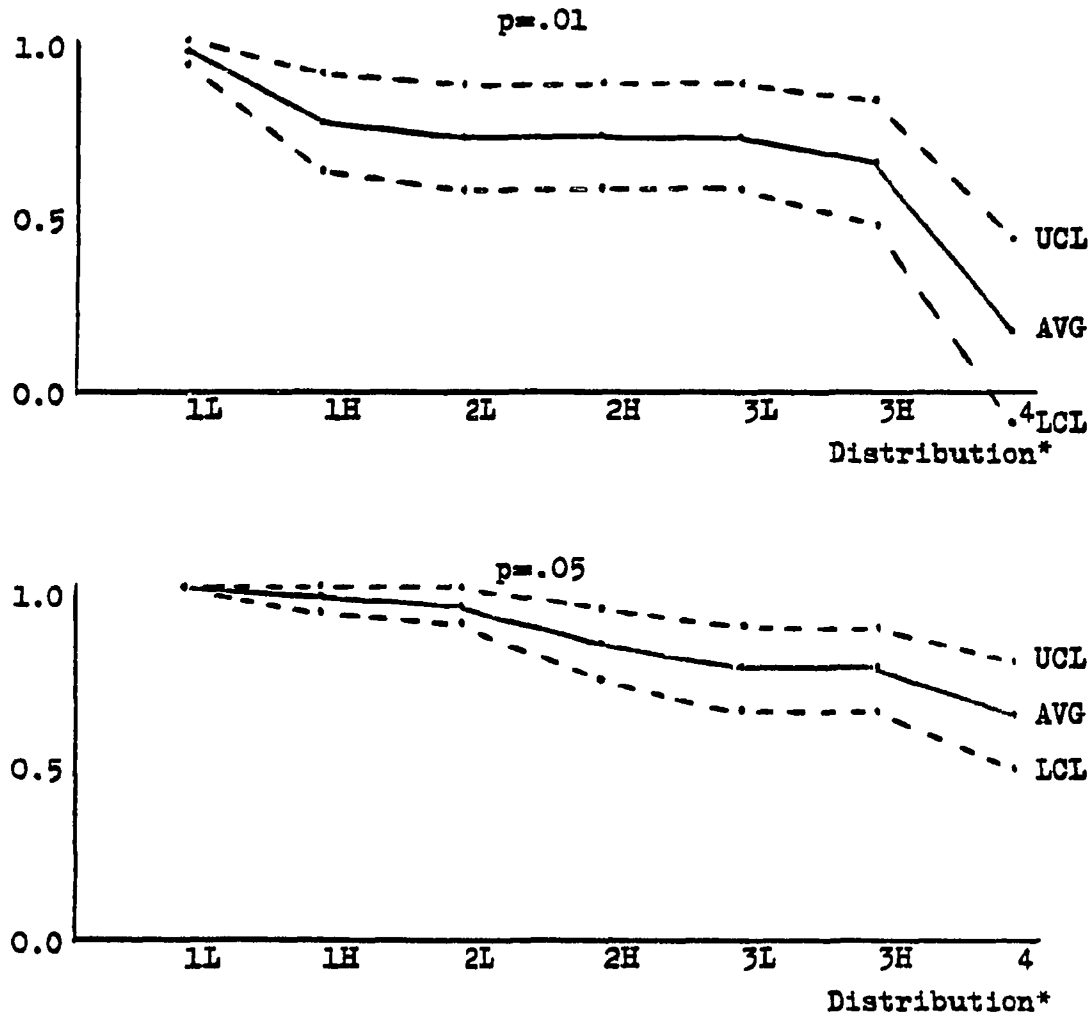
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.12A

Relative Conservatism of Classical MUAS: Test 1.5S

α=.092/β=.127

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



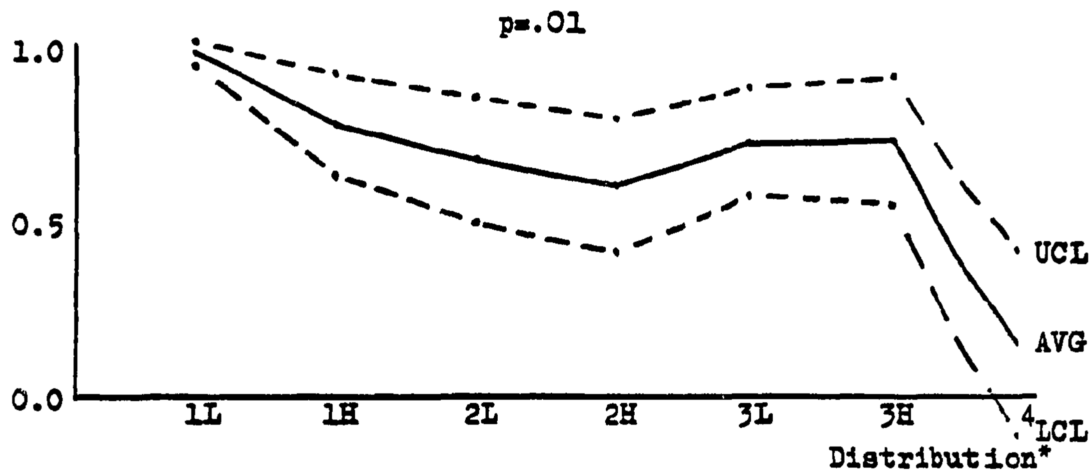
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

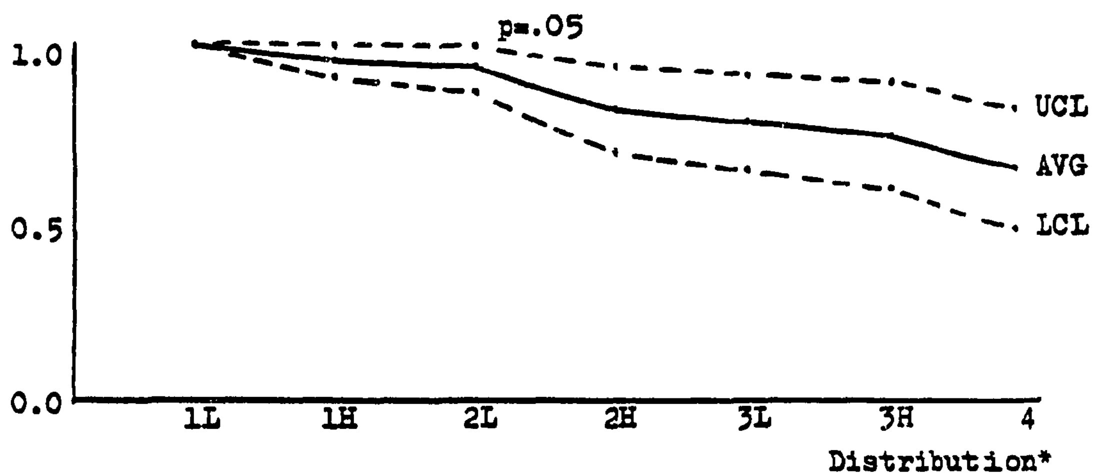
FIGURE 4.12B

Relative Conservatism of Classical MUAS: Test 1.5F

$\alpha = .094 / \beta = .098$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk





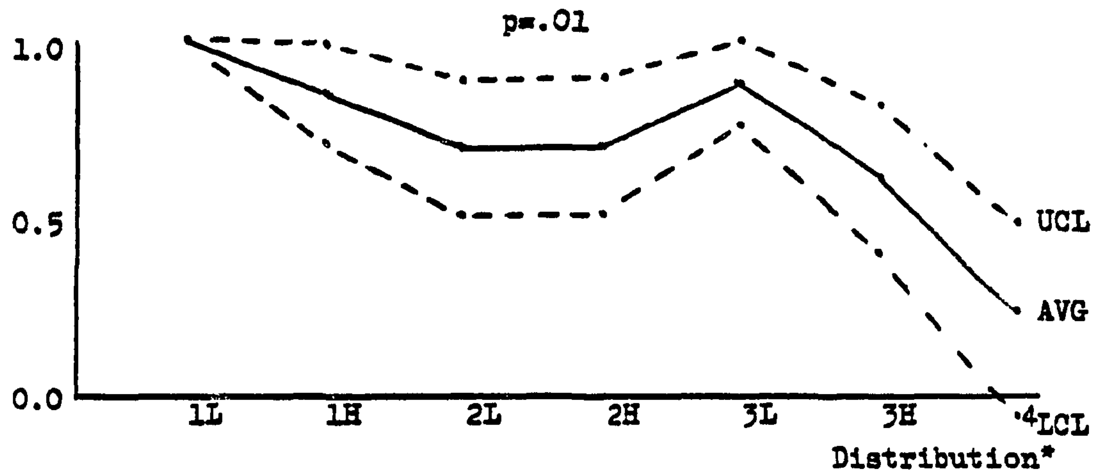
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

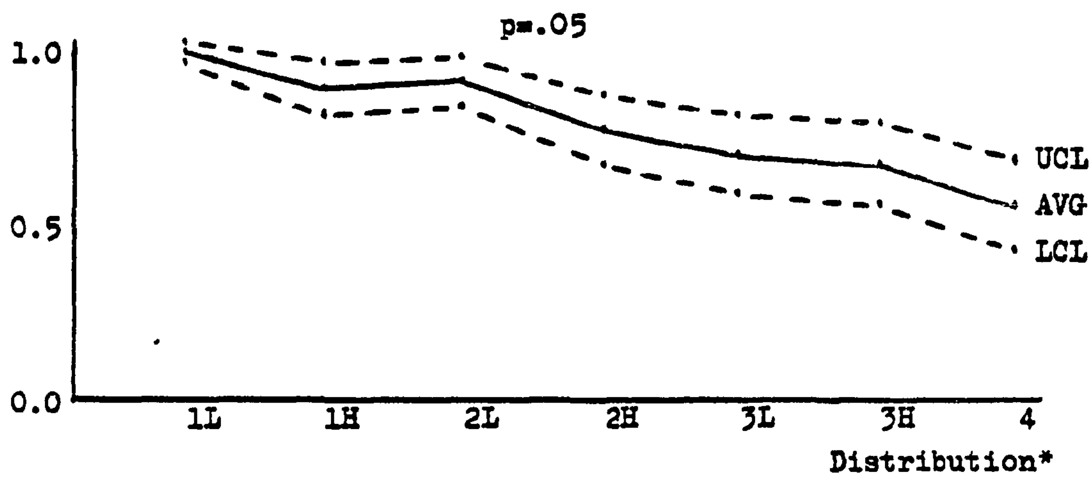
FIGURE 4.13A

Relative Conservatism of Classical MUAS: Test 1.6S

$\alpha = .066 / \beta = .192$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



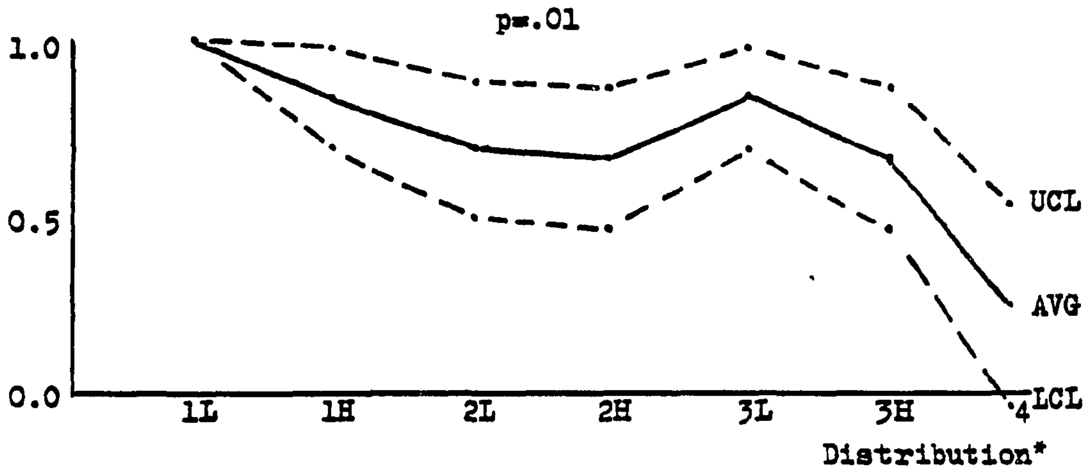


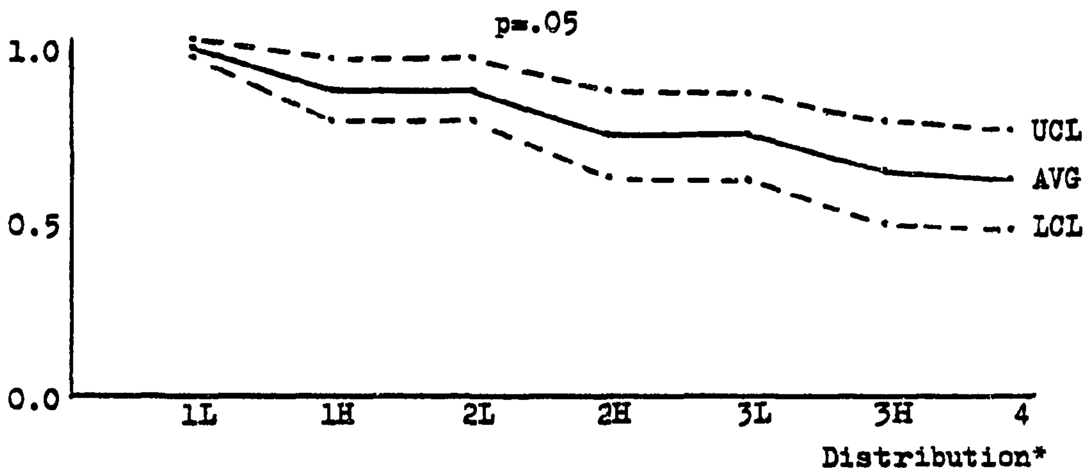
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.13B

Relative Conservatism of Classical MUAS: Test 1.6F
$\alpha=.070/\beta=.152$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk





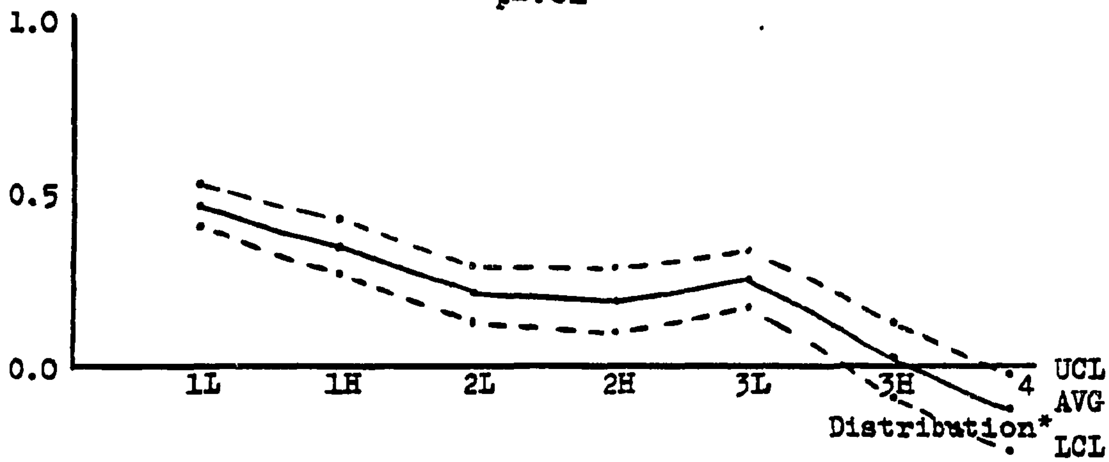
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.14A

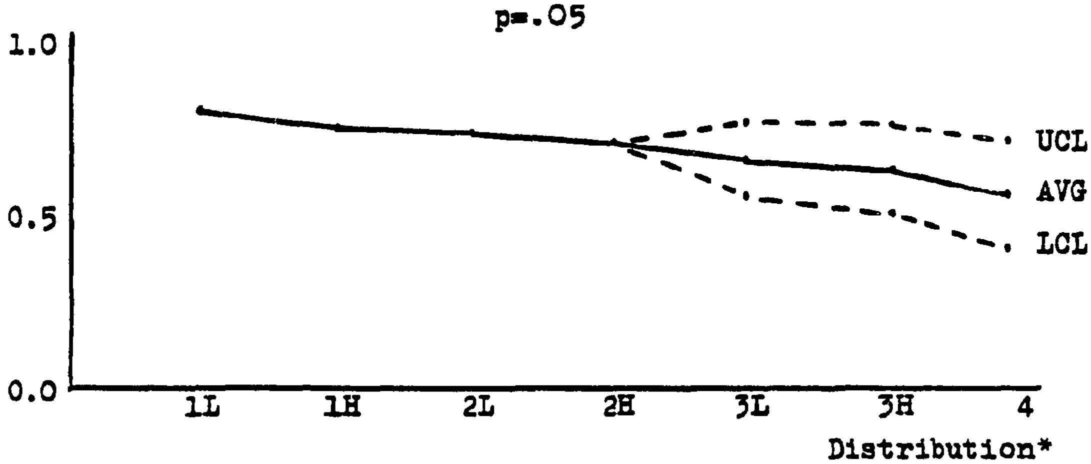
Relative Conservatism of Bayesian MUAS: Test 2.1S
g(.01)=.4/g(.05)=.6

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk

p=.01



p=.05



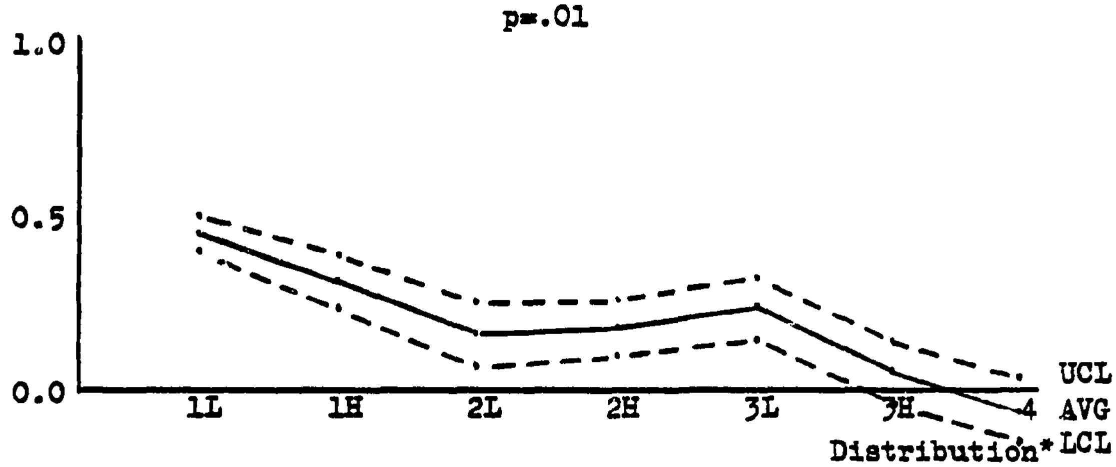
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.14B

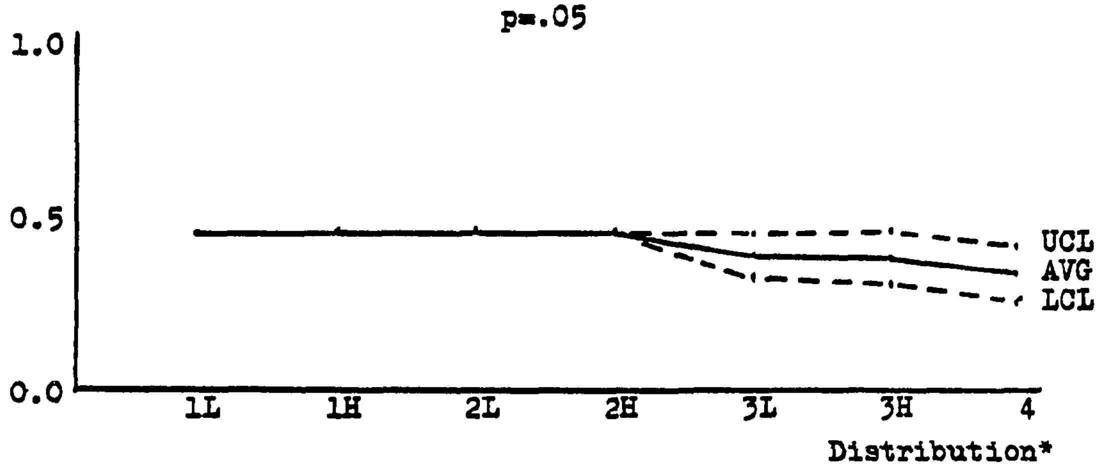
Relative Conservatism of Bayesian MUAS: Test 2.1F
g(.01)=.4/g(.05)=.6

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk

p=.01



p=.05

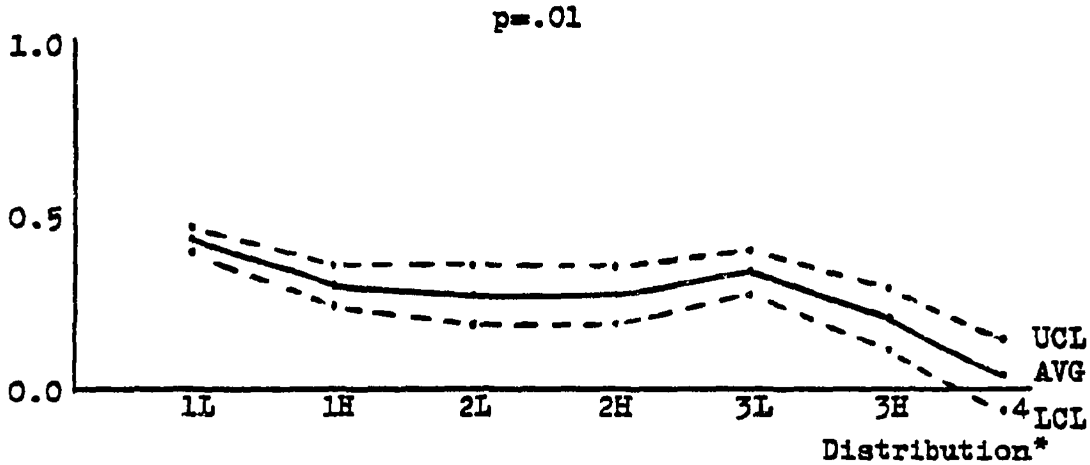


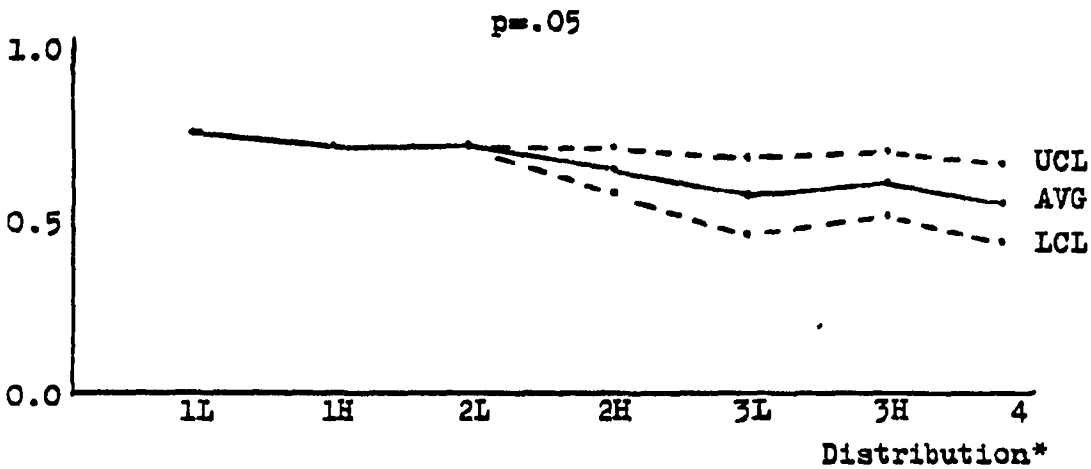
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.15A

Relative Conservatism of Bayesian MUAS: Test 2.2S
g(.01)=.5/g(.05)=.5

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



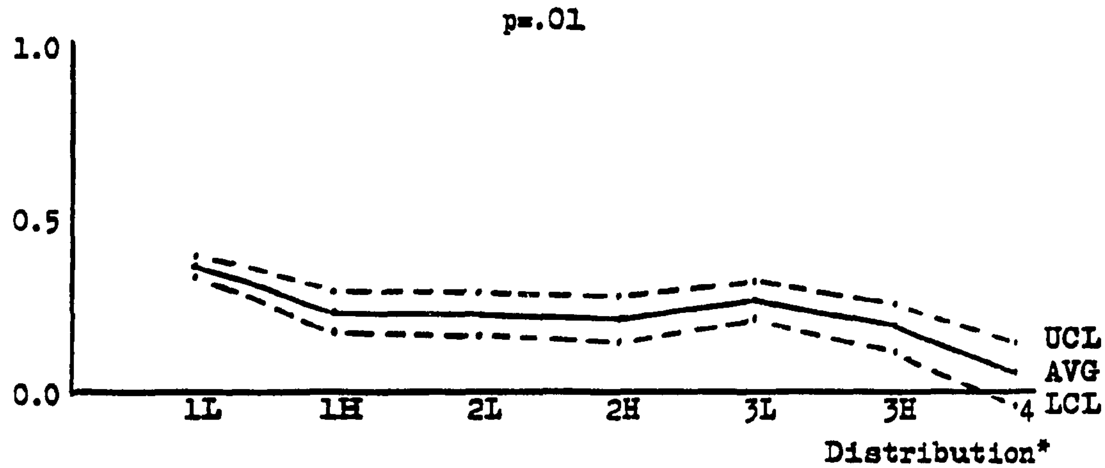


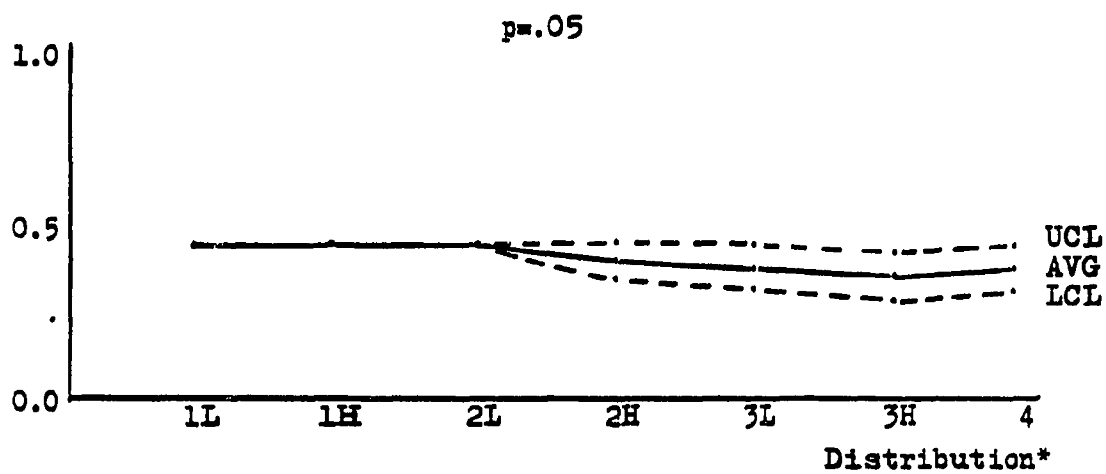
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.15B

Relative Conservatism of Bayesian MUAS: Test 2.2F
g(.01)=.5/g(.05)=.5

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk



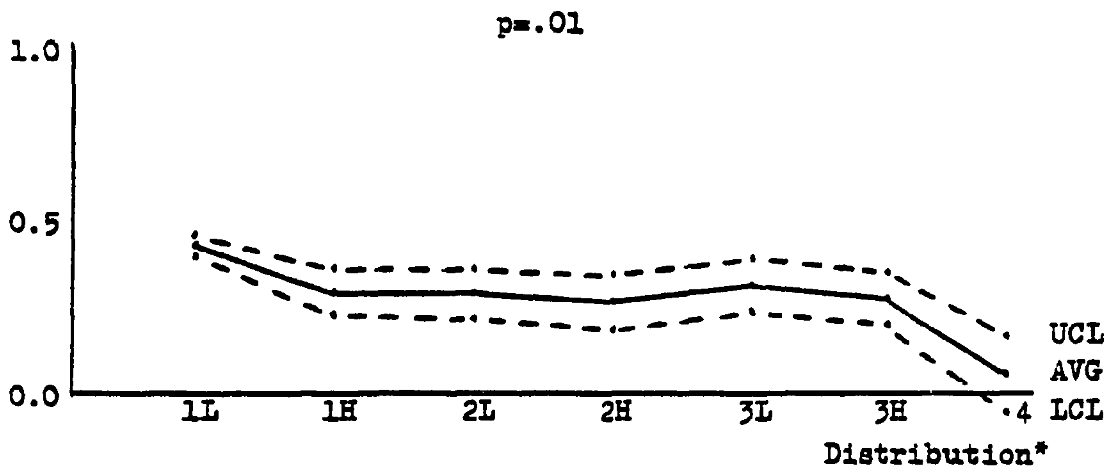


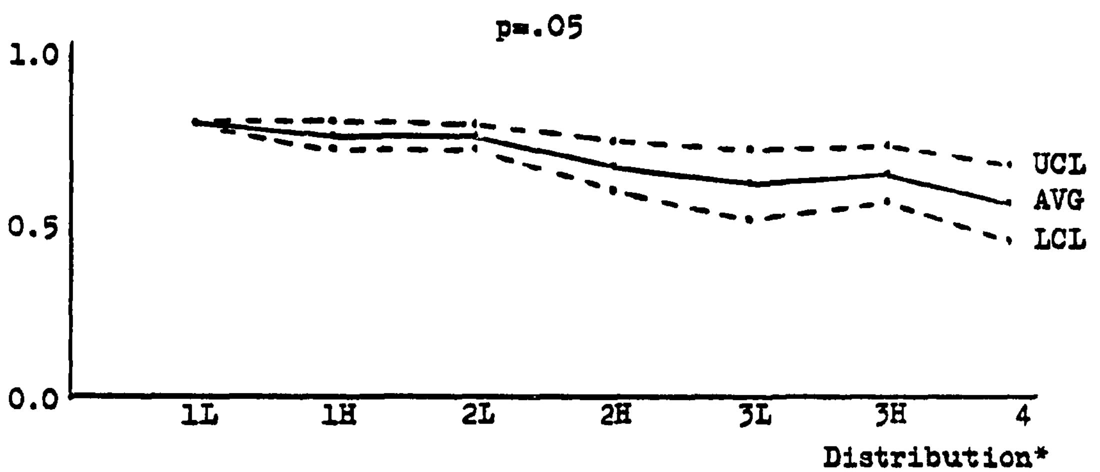
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.16A

Relative Conservatism of Bayesian MUAS: Test 2.3S
$g(.01)=.6/g(.05)=.4$

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



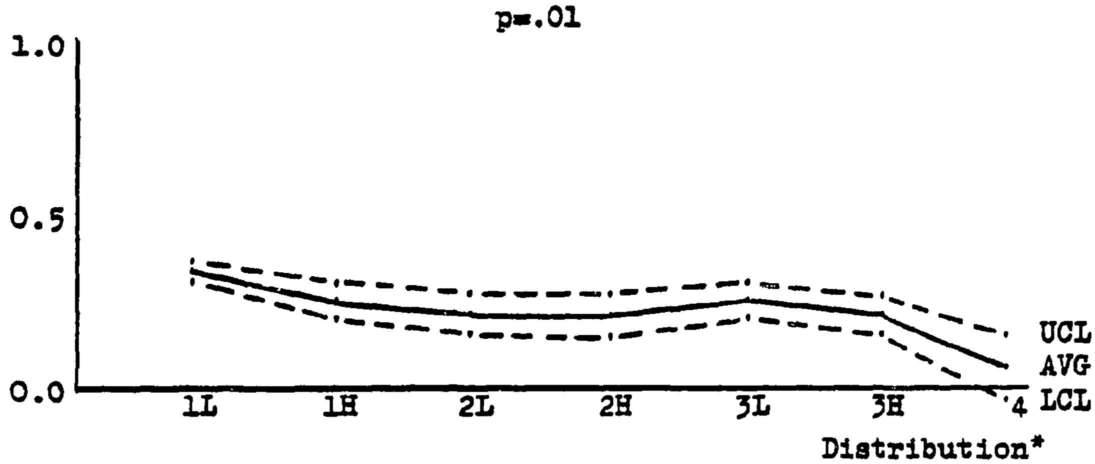


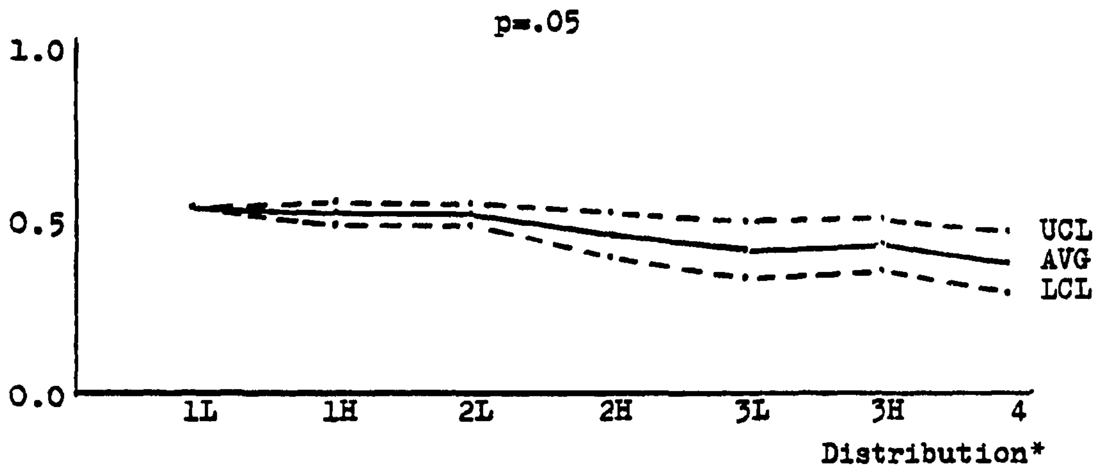
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.16B

Relative Conservatism of Bayesian MUAS: Test 2.3F
g(.01)=.6/g(.05)=.4

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk





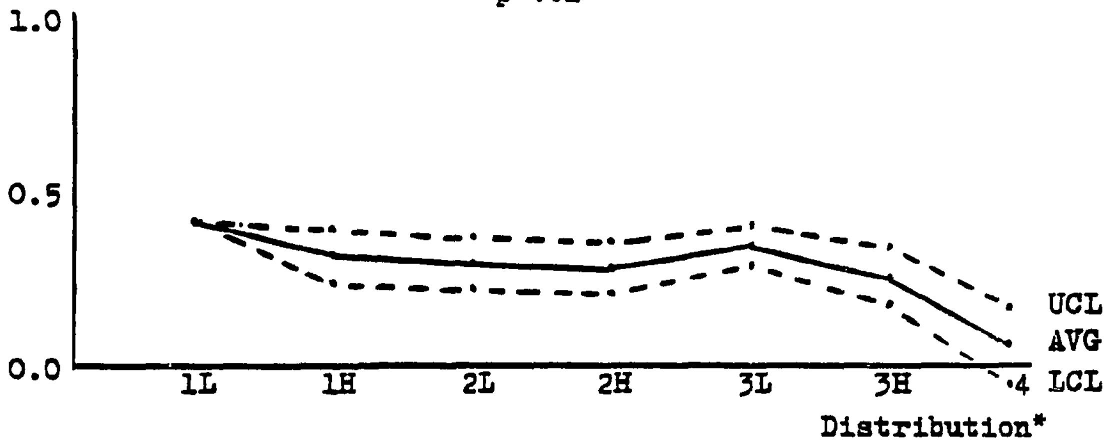
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.17A

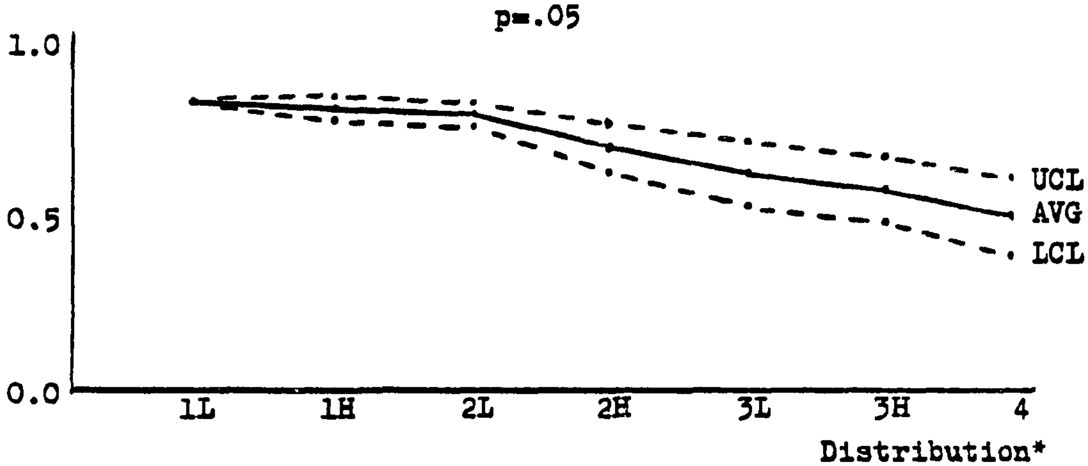
Relative Conservatism of Bayesian MUAS: Test 2.4S
g(.01)=.7/g(.05)=.3

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk

p=.01



p=.05



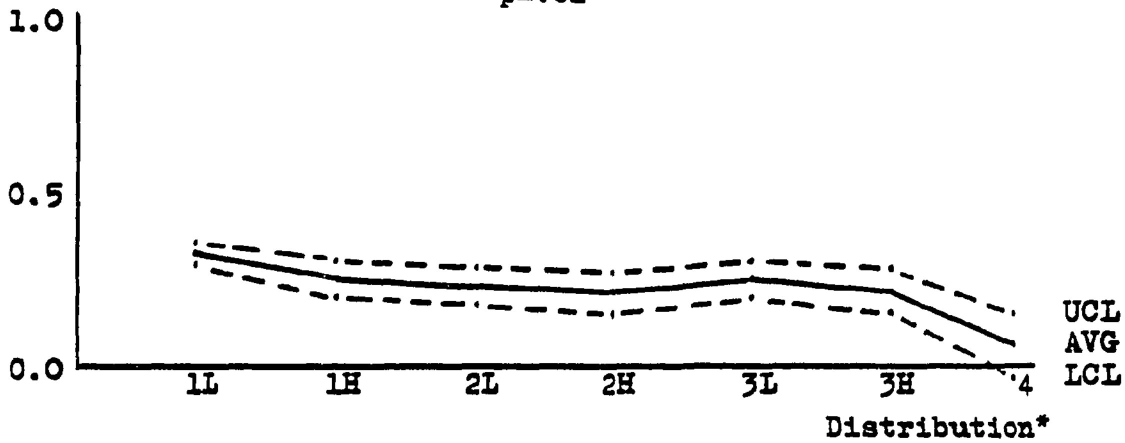
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.17B

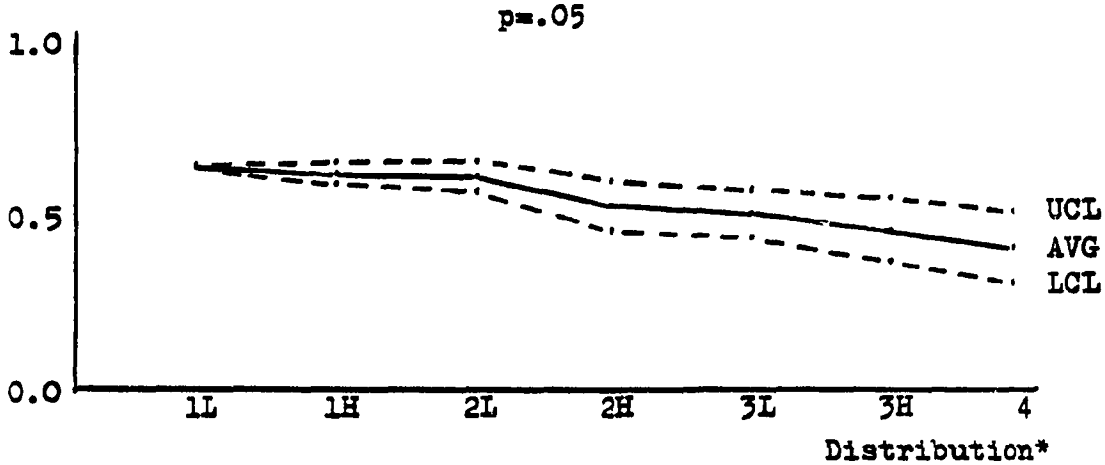
Relative Conservatism of Bayesian MUAS: Test 2.4F
g(.01)=.7/g(.05)=.3

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk

p=.01



p=.05

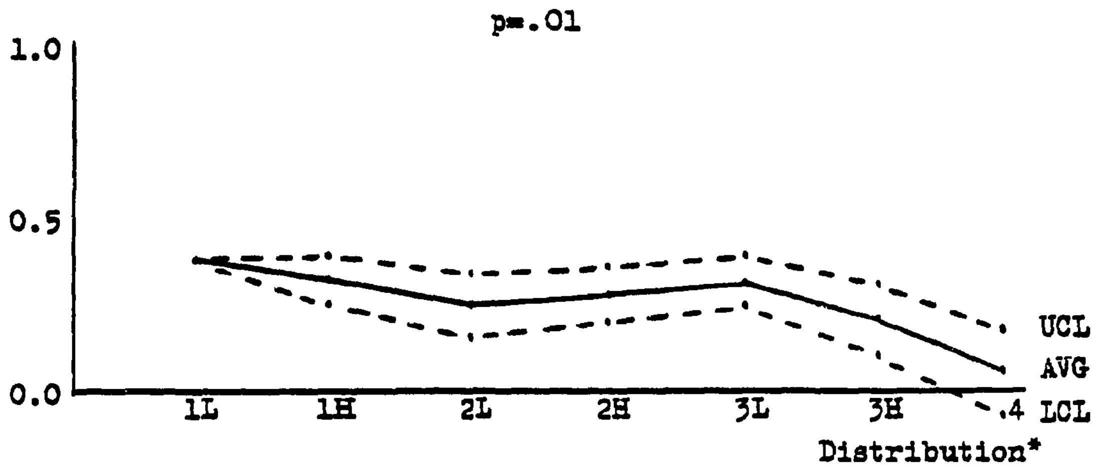


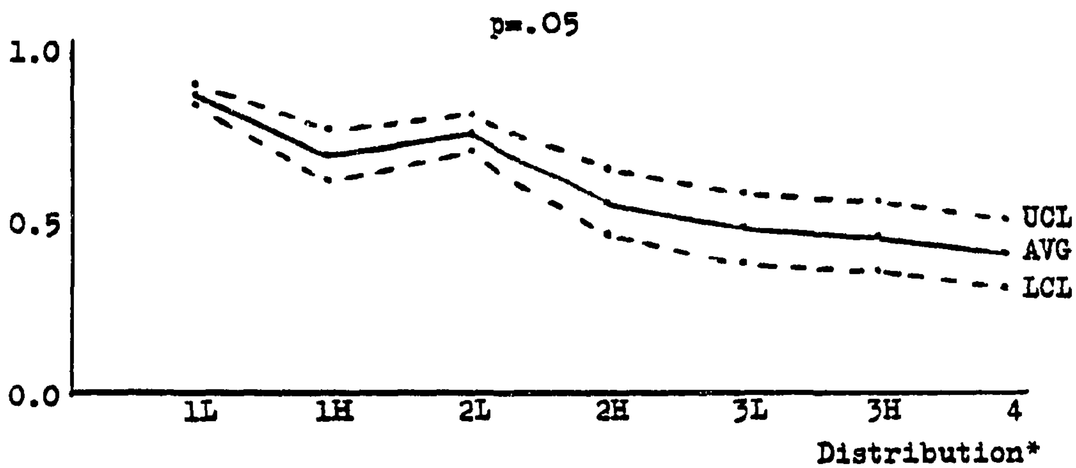
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.18A

Relative Conservatism of Bayesian MUAS: Test 2.5S
g(.01)=.8/g(.05)=.2

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



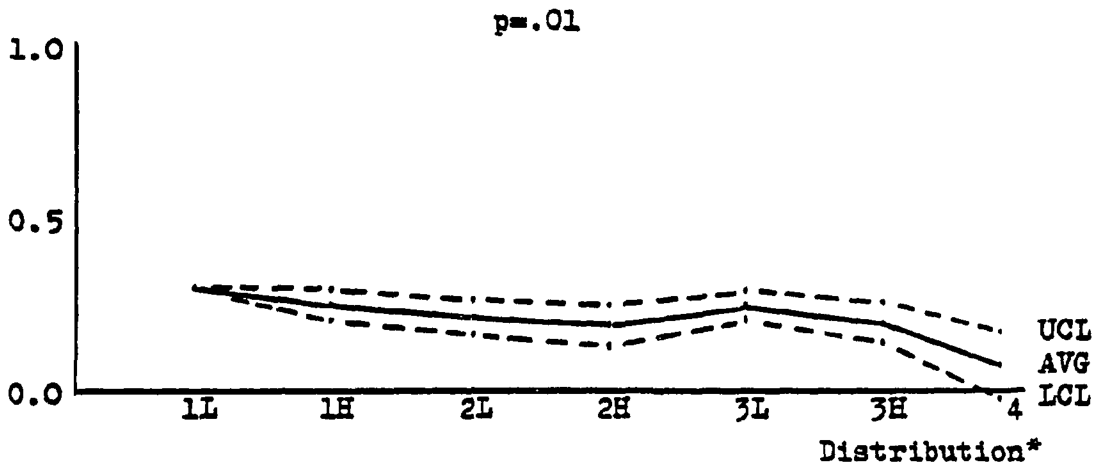


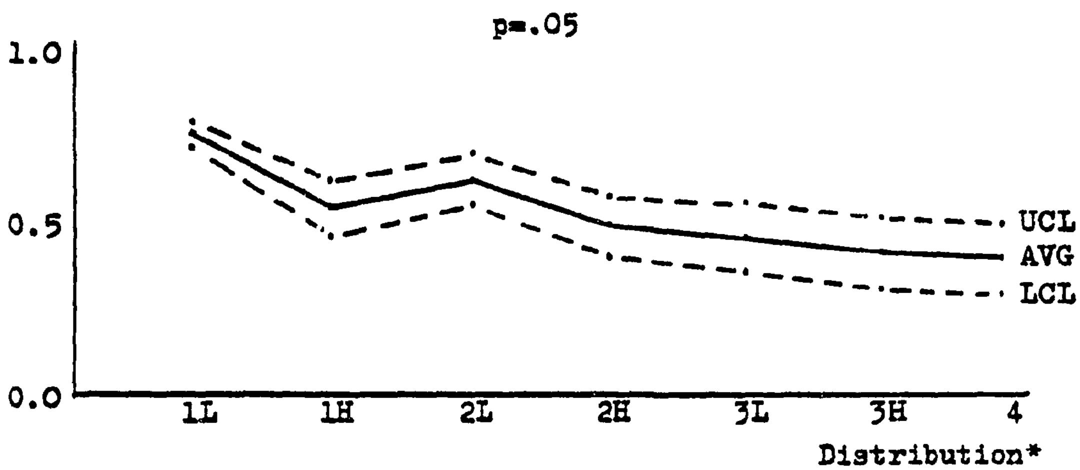
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.18B

Relative Conservatism of Bayesian MUAS: Test 2.5F
g(.01)=.8/g(.05)=.2

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk



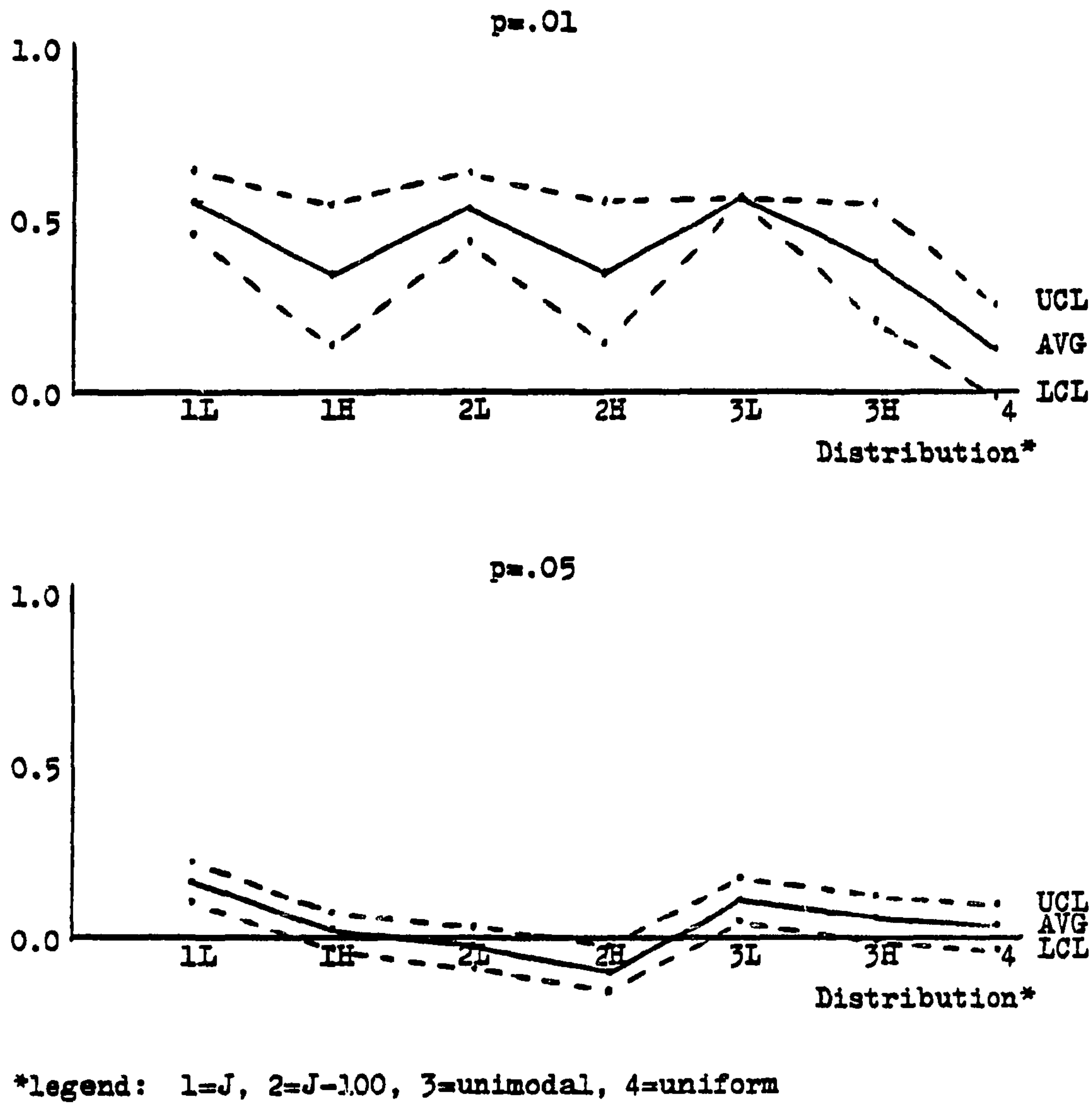


*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.19A

Relative Conservatism of Bayesian MUAS: Test 2.6S
g(.01)=.9/g(.05)=.1

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for
the ratio RC=(nominal risk-observed risk)/nominal risk

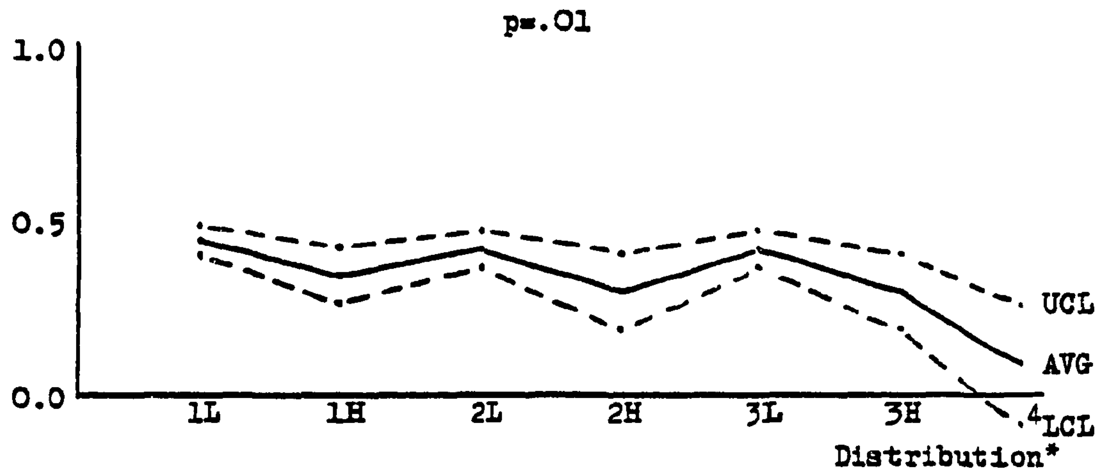


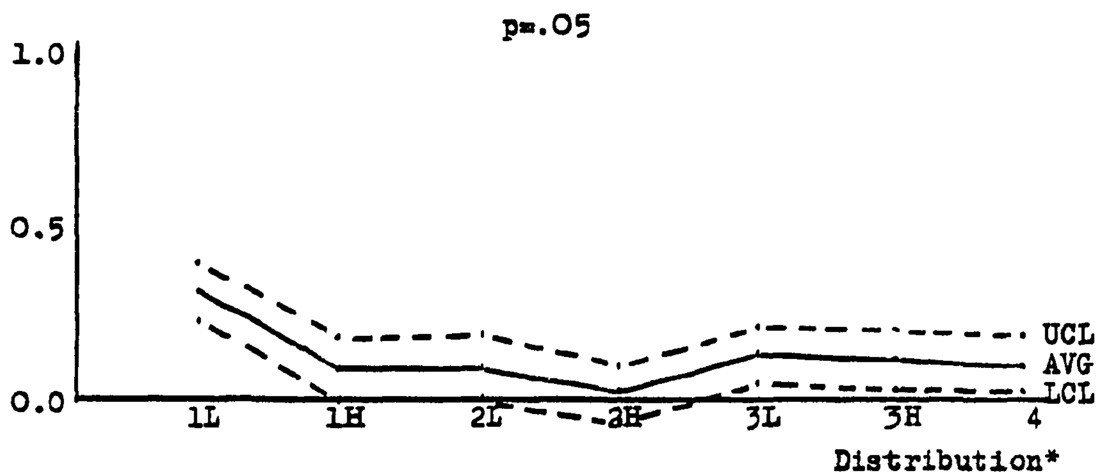
*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.19B

Relative Conservatism of Bayesian MUAS: Test 2.6F
g(.01)=.9/g(.05)=.1

Graph: mean (AVG) and 95% confidence limits (UCL,LCL) for the ratio RC=(nominal risk-observed risk)/nominal risk

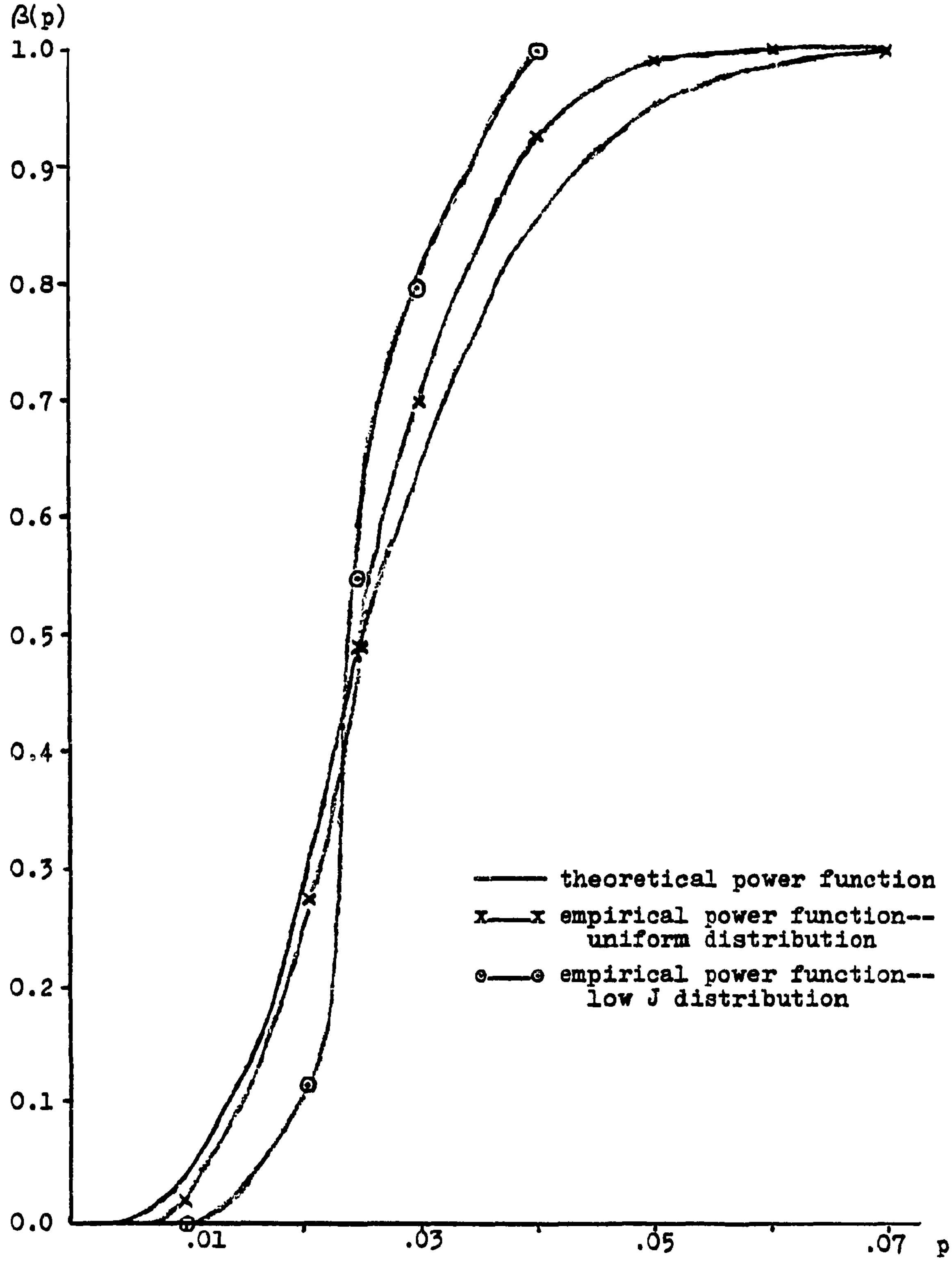




*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance

FIGURE 4.20

Empirical Versus Theoretical Power Functions for Test 1.1F



APPENDIX A

STATISTICAL FREQUENCY AND DENSITY FUNCTIONS

1. Binomial Distribution. The binomial(n,p) frequency function is given by

$$f^n(x;p)=\binom{n}{x}p^x(1-p)^{n-x} \quad x=0,1,\ldots,n$$

where $0<p<1$ and n is a positive integer. If X_i $(i=1,\ldots,n)$ are independent, identically distributed binomial(1,p) random variables (more commonly called Bernoulli random variables), then $S=\sum_{i=1}^{n} X_i \sim$ binomial(n,p) with E(S)=np and Var(S)=np(1-p).

2. Poisson Distribution. The Poisson(q) frequency function is given by

$$f(x;q)=e^{-q}q^x/x! \quad x=0,1,2,\ldots$$

where $q>0$. E(X)=Var(X)=q. For p small and np moderate, the binomial(n,p) may be approximated by the Poisson(np).

3. Normal Distribution. The normal(a,b^2) density function is given by

$$f(x;a,b^2)=(\sqrt{2\pi}\, b)^{-1}\exp\{-(x-a)^2/2b^2\}$$

where $b>0$. E(X)=a and Var(X)=b^2. The normal(0,1) distribution is called the standard normal distribution. Its (cumulative) distribution function is denoted by $\Phi(\cdot)$.

4. Gamma Distribution. The gamma(r,s) density function is given by

$$f(x;r,s)=s^r x^{r-1}e^{-sx}/\Gamma(r) \quad x>0$$

where $r,s>0$ and $\Gamma(\cdot)$ is the Euler gamma function. E(X)=r/s and Var(X)=r/s^2. The gamma(1,s) is called the exponential(s) distribution, with density given by

$$f(x;s)=se^{-sx} \quad x>0$$

APPENDIX B

TABLES OF THE CUMULATIVE POISSON DISTRIBUTION

$$P_q\left\{X \geqslant x\right\} = F(x;q) = \sum_{k=0}^{x} e^{-q} q^k / k!$$

$$q = 0.1(0.1)20.0$$

	q=0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
x=0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.095	.181	.259	.330	.393	.451	.503	.551	.593	.632
2	.005	.018	.037	.062	.090	.122	.155	.191	.228	.264
3	.000	.001	.004	.008	.014	.023	.034	.047	.063	.080
4		.000	.000	.001	.002	.003	.006	.009	.013	.019
5				.000	.000	.000	.001	.001	.002	.004
6							.000	.000	.000	.001
7										.000

	q=1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
x=0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.667	.699	.727	.753	.777	.798	.817	.835	.850	.865
2	.301	.337	.373	.408	.442	.475	.507	.537	.566	.594
3	.100	.121	.143	.167	.191	.217	.243	.269	.296	.323
4	.026	.034	.043	.054	.066	.079	.093	.109	.125	.143
5	.005	.008	.011	.014	.019	.024	.030	.036	.044	.053
6	.001	.002	.002	.003	.004	.006	.008	.010	.013	.017
7	.000	.000	.000	.001	.001	.001	.002	.003	.003	.005
8				.000	.000	.000	.000	.001	.001	.001
9								.000	.000	.000

	q=2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.878	.889	.900	.909	.918	.926	.933	.939	.945	.950
2	.620	.645	.669	.692	.713	.733	.751	.769	.785	.801
3	.350	.377	.404	.430	.456	.482	.506	.531	.554	.577
4	.161	.181	.201	.221	.242	.264	.286	.308	.330	.353
5	.062	.072	.084	.096	.109	.123	.137	.152	.168	.185
6	.020	.025	.030	.036	.042	.049	.057	.065	.074	.084
7	.006	.007	.009	.012	.014	.017	.021	.024	.029	.034
8	.001	.002	.003	.003	.004	.005	.007	.008	.010	.012
9	.000	.000	.001	.001	.001	.001	.002	.002	.003	.004
10			.000	.000	.000	.000	.001	.001	.001	.001
11							.000	.000	.000	.000

	q=3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90	4.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.955	.959	.963	.967	.970	.973	.975	.978	.980	.982
2	.815	.829	.841	.853	.864	.874	.884	.893	.901	.908
3	.599	.620	.641	.660	.679	.697	.715	.731	.747	.762
4	.375	.397	.420	.442	.463	.485	.506	.527	.547	.567
5	.202	.219	.237	.256	.275	.294	.313	.332	.352	.371
6	.094	.105	.117	.129	.142	.156	.170	.184	.199	.215
7	.039	.045	.051	.058	.065	.073	.082	.091	.101	.111
8	.014	.017	.020	.023	.027	.031	.035	.040	.045	.051
9	.005	.006	.007	.008	.010	.012	.014	.016	.019	.021
10	.001	.002	.002	.003	.003	.004	.005	.006	.007	.008
11	.000	.000	.001	.001	.001	.001	.002	.002	.002	.003
12			.000	.000	.000	.000	.000	.001	.001	.001
13								.000	.000	.000

x \ q	4.10	4.20	4.30	4.40	4.50	4.60	4.70	4.80	4.90	5.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.983	.985	.986	.988	.989	.990	.991	.992	.993	.993
2	.915	.922	.928	.934	.939	.944	.948	.952	.956	.960
3	.776	.790	.803	.815	.826	.837	.848	.857	.867	.875
4	.586	.605	.623	.641	.658	.674	.690	.706	.721	.736
5	.391	.410	.430	.449	.468	.487	.505	.524	.542	.560
6	.231	.247	.263	.280	.297	.314	.332	.349	.366	.384
7	.121	.133	.144	.156	.169	.182	.195	.209	.223	.238
8	.057	.064	.071	.079	.087	.095	.104	.113	.123	.133
9	.024	.028	.032	.036	.040	.045	.050	.056	.062	.068
10	.010	.011	.013	.015	.017	.020	.022	.025	.028	.032
11	.003	.004	.005	.006	.007	.008	.009	.010	.012	.014
12	.001	.001	.002	.002	.002	.003	.003	.004	.005	.005
13	.000	.000	.001	.001	.001	.001	.001	.001	.002	.002
14			.000	.000	.000	.000	.000	.000	.001	.001
15									.000	.000

x \ q	5.10	5.20	5.30	5.40	5.50	5.60	5.70	5.80	5.90	6.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.994	.994	.995	.995	.996	.996	.997	.997	.997	.998
2	.963	.966	.969	.971	.973	.976	.978	.979	.981	.983
3	.884	.891	.898	.905	.912	.918	.923	.928	.933	.938
4	.749	.762	.775	.787	.798	.809	.820	.830	.840	.849
5	.577	.594	.610	.627	.642	.658	.673	.687	.701	.715
6	.402	.419	.437	.454	.471	.488	.505	.522	.538	.554
7	.253	.268	.283	.298	.314	.330	.346	.362	.378	.394
8	.144	.155	.167	.178	.191	.203	.216	.229	.242	.256
9	.075	.082	.089	.097	.106	.114	.123	.133	.143	.153
10	.036	.040	.044	.049	.054	.059	.065	.071	.077	.084
11	.016	.018	.020	.023	.025	.028	.031	.035	.039	.043
12	.006	.007	.008	.010	.011	.012	.014	.016	.018	.020
13	.002	.003	.003	.004	.004	.005	.006	.007	.008	.009
14	.001	.001	.001	.001	.002	.002	.002	.003	.003	.004
15	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
16					.000	.000	.000	.000	.000	.001
17										.000

x \ q	6.10	6.20	6.30	6.40	6.50	6.60	6.70	6.80	6.90	7.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.998	.998	.998	.998	.998	.999	.999	.999	.999	.999
2	.984	.985	.987	.988	.989	.990	.991	.991	.992	.993
3	.942	.946	.950	.954	.957	.960	.963	.966	.968	.970
4	.857	.866	.874	.881	.888	.895	.901	.907	.913	.918
5	.728	.741	.753	.765	.776	.787	.798	.808	.818	.827
6	.570	.586	.601	.616	.631	.645	.659	.673	.686	.699
7	.410	.426	.442	.458	.473	.489	.505	.520	.535	.550
8	.270	.284	.298	.313	.327	.342	.357	.372	.386	.401
9	.163	.174	.185	.197	.208	.220	.233	.245	.258	.271
10	.091	.098	.106	.114	.123	.131	.140	.150	.160	.170
11	.047	.051	.056	.061	.067	.073	.079	.085	.092	.099
12	.022	.025	.028	.031	.034	.037	.041	.045	.049	.053
13	.010	.011	.013	.014	.016	.018	.020	.022	.024	.027
14	.004	.005	.005	.006	.007	.008	.009	.010	.011	.013
15	.002	.002	.002	.003	.003	.003	.004	.004	.005	.006
16	.001	.001	.001	.001	.001	.001	.002	.002	.002	.002
17	.000	.000	.000	.000	.000	.001	.001	.001	.001	.001
18						.000	.000	.000	.000	.000

	q=7.10	7.20	7.30	7.40	7.50	7.60	7.70	7.80	7.90	8.00
x= 0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	.999	.999	.999	.999	.999	.999	1.00	1.00	1.00	1.00
2	.993	.994	.994	.995	.996	.996	.996	.996	.997	.997
3	.973	.975	.976	.978	.980	.981	.983	.984	.985	.986
4	.923	.928	.933	.937	.941	.945	.948	.952	.955	.958
5	.836	.844	.853	.860	.868	.875	.882	.888	.894	.900
6	.712	.724	.736	.747	.759	.769	.780	.790	.799	.809
7	.565	.580	.594	.608	.622	.635	.649	.662	.674	.687
8	.416	.431	.446	.461	.475	.490	.504	.519	.533	.547
9	.284	.297	.311	.324	.338	.352	.366	.380	.393	.407
10	.180	.190	.201	.212	.224	.235	.247	.259	.271	.283
11	.106	.113	.121	.129	.138	.146	.155	.165	.174	.184
12	.058	.063	.068	.074	.079	.085	.091	.098	.105	.112
13	.030	.033	.036	.039	.043	.046	.050	.055	.059	.064
14	.014	.016	.018	.020	.022	.024	.026	.029	.031	.034
15	.006	.007	.008	.009	.010	.011	.013	.014	.016	.017
16	.003	.003	.004	.004	.005	.005	.006	.007	.007	.008
17	.001	.001	.001	.002	.002	.002	.003	.003	.003	.004
18	.000	.000	.001	.001	.001	.001	.001	.001	.001	.002
19			.000	.000	.000	.000	.000	.000	.001	.001
20									.000	.000

	q=8.10	8.20	8.30	8.40	8.50	8.60	8.70	8.80	8.90	9.00
x=0-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	.997	.997	.998	.998	.998	.998	.998	.999	.999	.999
3	.987	.988	.989	.990	.991	.991	.992	.993	.993	.994
4	.960	.963	.965	.968	.970	.972	.974	.976	.977	.979
5	.906	.911	.916	.921	.926	.930	.934	.938	.942	.945
6	.818	.826	.835	.843	.850	.858	.865	.872	.878	.884
7	.699	.710	.722	.733	.744	.754	.765	.774	.784	.793
8	.561	.575	.588	.601	.614	.627	.640	.652	.664	.676
9	.421	.435	.449	.463	.477	.491	.504	.518	.531	.544
10	.296	.308	.321	.334	.347	.360	.373	.386	.399	.413
11	.194	.204	.215	.226	.237	.248	.259	.271	.282	.294
12	.119	.127	.135	.143	.151	.160	.169	.178	.187	.197
13	.069	.074	.079	.085	.091	.097	.103	.110	.117	.124
14	.037	.040	.044	.048	.051	.055	.060	.064	.069	.074
15	.019	.021	.023	.025	.027	.030	.033	.035	.038	.041
16	.009	.010	.011	.013	.014	.015	.017	.018	.020	.022
17	.004	.005	.005	.006	.007	.007	.008	.009	.010	.011
18	.002	.002	.002	.003	.003	.003	.004	.004	.005	.005
19	.001	.001	.001	.001	.001	.001	.002	.002	.002	.002
20	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
21					.000	.000	.000	.000	.000	.000

x	q=9.10	9.20	9.30	9.40	9.50	9.60	9.70	9.80	9.90	10.0
x=0-1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	.999	.999	.999	.999	.999	.999	.999	.999	.999	1.00
3	.994	.995	.995	.995	.996	.996	.996	.997	.997	.997
4	.980	.982	.983	.984	.985	.986	.987	.988	.989	.990
5	.948	.951	.954	.957	.960	.962	.965	.967	.969	.971
6	.890	.896	.901	.907	.911	.916	.921	.925	.929	.933
7	.802	.811	.819	.827	.835	.843	.850	.857	.863	.870
8	.688	.699	.710	.721	.731	.742	.752	.761	.771	.780
9	.557	.570	.583	.596	.608	.620	.632	.644	.656	.667
10	.426	.439	.452	.465	.478	.491	.504	.517	.529	.542
11	.306	.318	.330	.342	.355	.367	.379	.392	.404	.417
12	.207	.217	.227	.237	.248	.259	.270	.281	.292	.303
13	.132	.139	.147	.155	.164	.172	.181	.190	.199	.208
14	.079	.084	.090	.096	.102	.108	.115	.121	.128	.136
15	.045	.048	.052	.056	.060	.064	.069	.073	.078	.083
16	.024	.026	.028	.031	.033	.036	.039	.042	.045	.049
17	.012	.013	.015	.016	.018	.019	.021	.023	.025	.027
18	.006	.007	.007	.008	.009	.010	.011	.012	.013	.014
19	.003	.003	.003	.004	.004	.005	.005	.006	.007	.007
20	.001	.001	.002	.002	.002	.002	.002	.003	.003	.003
21	.001	.001	.001	.001	.001	.001	.001	.001	.001	.002
22	.000	.000	.000	.000	.000	.000	.000	.001	.001	.001
23								.000	.000	.000

x	q=10.1	10.2	10.3	10.4	10.5	10.6	10.7	10.8	10.9	11.0
x=0-2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	.997	.998	.998	.998	.998	.998	.998	.999	.999	.999
4	.990	.991	.992	.992	.993	.993	.994	.994	.995	.995
5	.973	.974	.976	.977	.979	.980	.982	.983	.984	.985
6	.937	.940	.943	.947	.950	.952	.955	.958	.960	.962
7	.876	.882	.888	.893	.898	.903	.908	.913	.917	.921
8	.789	.797	.806	.814	.821	.829	.836	.843	.850	.857
9	.678	.689	.700	.710	.721	.731	.740	.750	.759	.768
10	.555	.567	.579	.591	.603	.615	.626	.637	.649	.659
11	.429	.442	.454	.467	.479	.492	.504	.516	.528	.540
12	.315	.326	.338	.350	.361	.373	.385	.397	.409	.421
13	.218	.228	.238	.248	.258	.268	.279	.290	.300	.311
14	.143	.151	.158	.166	.175	.183	.192	.201	.210	.219
15	.089	.094	.100	.106	.112	.118	.125	.132	.139	.146
16	.052	.056	.060	.064	.068	.073	.077	.082	.087	.093
17	.029	.032	.034	.037	.040	.043	.046	.049	.052	.056
18	.016	.017	.019	.020	.022	.024	.026	.028	.030	.032
19	.008	.009	.010	.011	.012	.013	.014	.015	.016	.018
20	.004	.004	.005	.005	.006	.006	.007	.008	.008	.009
21	.002	.002	.002	.003	.003	.003	.003	.004	.004	.005
22	.001	.001	.001	.001	.001	.001	.002	.002	.002	.002
23	.000	.000	.000	.001	.001	.001	.001	.001	.001	.001
24				.000	.000	.000	.000	.000	.000	.000

x	q=11.1	11.2	11.3	11.4	11.5	11.6	11.7	11.8	11.9	12.0
x=0-2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	.999	.999	.999	.999	.999	.999	.999	.999	.999	.999
4	.995	.996	.996	.996	.997	.997	.997	.997	.998	.998
5	.986	.987	.988	.988	.989	.990	.991	.991	.992	.992
6	.965	.967	.969	.971	.972	.974	.975	.977	.978	.980
7	.925	.929	.933	.936	.940	.943	.946	.949	.952	.954
8	.863	.869	.875	.881	.886	.892	.897	.901	.906	.910
9	.777	.785	.794	.802	.809	.817	.824	.831	.838	.845
10	.670	.681	.691	.701	.711	.721	.730	.740	.749	.758
11	.552	.564	.575	.587	.598	.609	.621	.631	.642	.653
12	.433	.445	.456	.468	.480	.492	.504	.515	.527	.538
13	.322	.333	.345	.356	.367	.378	.390	.401	.413	.424
14	.228	.238	.247	.257	.267	.277	.287	.298	.308	.318
15	.153	.161	.169	.177	.185	.193	.202	.210	.219	.228
16	.098	.104	.109	.115	.122	.128	.135	.141	.148	.156
17	.060	.064	.068	.072	.076	.081	.086	.091	.096	.101
18	.035	.037	.040	.043	.046	.049	.052	.056	.059	.063
19	.019	.021	.022	.024	.026	.028	.030	.033	.035	.037
20	.010	.011	.012	.013	.014	.016	.017	.018	.020	.021
21	.005	.006	.006	.007	.008	.008	.009	.010	.011	.012
22	.003	.003	.003	.003	.004	.004	.005	.005	.006	.006
23	.001	.001	.001	.002	.002	.002	.002	.002	.003	.003
24	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
25	.000	.000	.000	.000	.000	.000	.000	.001	.001	.001
26								.000	.000	.000

x	q=12.1	12.2	12.3	12.4	12.5	12.6	12.7	12.8	12.9	13.0
x=0-3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	.998	.998	.998	.998	.998	.999	.999	.999	.999	.999
5	.993	.993	.994	.994	.995	.995	.995	.996	.996	.996
6	.981	.982	.983	.984	.985	.986	.987	.988	.989	.989
7	.957	.959	.961	.963	.965	.967	.969	.971	.973	.974
8	.915	.919	.923	.927	.930	.934	.937	.940	.943	.946
9	.851	.858	.864	.869	.875	.880	.886	.891	.896	.900
10	.766	.775	.783	.791	.799	.806	.813	.821	.827	.834
11	.663	.673	.683	.693	.703	.712	.722	.731	.740	.748
12	.550	.561	.572	.583	.594	.605	.616	.626	.637	.647
13	.435	.447	.458	.470	.481	.492	.504	.515	.526	.537
14	.329	.340	.350	.361	.372	.383	.394	.405	.416	.427
15	.237	.246	.256	.265	.275	.285	.295	.305	.315	.325
16	.163	.170	.178	.186	.194	.202	.210	.219	.228	.236
17	.107	.113	.118	.124	.131	.137	.144	.150	.157	.165
18	.067	.071	.075	.080	.084	.089	.094	.099	.104	.110
19	.040	.043	.046	.049	.052	.055	.059	.062	.066	.070
20	.023	.025	.027	.029	.031	.033	.035	.037	.040	.043
21	.013	.014	.015	.016	.017	.019	.020	.022	.023	.025
22	.007	.007	.008	.009	.009	.010	.011	.012	.013	.014
23	.003	.004	.004	.004	.005	.005	.006	.006	.007	.008
24	.002	.002	.002	.002	.002	.003	.003	.003	.004	.004
25	.001	.001	.001	.001	.001	.001	.001	.002	.002	.002
26	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
27					.000	.000	.000	.000	.000	.000

	q=13.1	13.2	13.3	13.4	13.5	13.6	13.7	13.8	13.9	14.0
x=0-3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	.999	.999	.999	.999	.999	.999	.999	.999	.999	1.00
5	.997	.997	.997	.997	.997	.998	.998	.998	.998	.998
6	.990	.991	.991	.992	.992	.993	.993	.994	.994	.994
7	.976	.977	.978	.980	.981	.982	.983	.984	.985	.986
8	.949	.951	.954	.956	.959	.961	.963	.965	.967	.968
9	.905	.909	.913	.917	.921	.925	.928	.932	.935	.938
10	.841	.847	.853	.859	.865	.870	.876	.881	.886	.891
11	.757	.765	.773	.781	.789	.796	.804	.811	.818	.824
12	.657	.667	.677	.686	.696	.705	.714	.723	.731	.740
13	.548	.559	.569	.580	.591	.601	.611	.622	.632	.642
14	.438	.449	.460	.471	.482	.493	.503	.514	.525	.536
15	.335	.345	.356	.366	.377	.387	.398	.408	.419	.430
16	.245	.254	.264	.273	.282	.292	.301	.311	.321	.331
17	.172	.179	.187	.195	.202	.211	.219	.227	.235	.244
18	.115	.121	.127	.133	.139	.146	.152	.159	.166	.173
19	.074	.078	.082	.087	.092	.096	.101	.107	.112	.117
20	.045	.048	.051	.055	.058	.061	.065	.069	.072	.077
21	.027	.029	.031	.033	.035	.037	.040	.042	.045	.048
22	.015	.016	.018	.019	.020	.022	.024	.025	.027	.029
23	.008	.009	.010	.011	.011	.012	.013	.014	.016	.017
24	.004	.005	.005	.006	.006	.007	.007	.008	.009	.009
25	.002	.002	.003	.003	.003	.004	.004	.004	.005	.005
26	.001	.001	.001	.001	.002	.002	.002	.002	.002	.003
27	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
28	.000	.000	.000	.000	.000	.000	.000	.001	.001	.001
29								.000	.000	.000

	q=14.1	14.2	14.3	14.4	14.5	14.6	14.7	14.8	14.9	15.0
x=0-4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	.998	.998	.999	.999	.999	.999	.999	.999	.999	.999
6	.995	.995	.995	.996	.996	.996	.997	.997	.997	.997
7	.987	.987	.988	.989	.990	.990	.991	.991	.992	.992
8	.970	.972	.973	.975	.976	.977	.979	.980	.981	.982
9	.941	.944	.947	.949	.952	.954	.956	.958	.961	.963
10	.895	.900	.904	.908	.912	.916	.920	.923	.927	.930
11	.831	.837	.843	.849	.855	.861	.866	.871	.877	.882
12	.748	.756	.764	.772	.780	.787	.795	.802	.809	.815
13	.651	.661	.670	.680	.689	.698	.707	.715	.724	.732
14	.546	.557	.567	.577	.587	.598	.608	.617	.627	.637
15	.440	.451	.461	.472	.482	.493	.503	.514	.524	.534
16	.341	.351	.361	.371	.381	.391	.401	.411	.422	.432
17	.253	.262	.271	.280	.289	.298	.307	.317	.326	.336
18	.180	.187	.195	.203	.210	.218	.226	.234	.243	.251
19	.123	.129	.135	.141	.147	.153	.160	.167	.174	.181
20	.081	.085	.089	.094	.099	.104	.109	.114	.119	.125
21	.051	.054	.057	.060	.064	.067	.071	.075	.079	.083
22	.031	.033	.035	.037	.040	.042	.045	.047	.050	.053
23	.018	.019	.021	.022	.024	.025	.027	.029	.031	.033
24	.010	.011	.012	.013	.014	.015	.016	.017	.018	.019
25	.005	.006	.006	.007	.008	.008	.009	.010	.010	.011
26	.003	.003	.003	.004	.004	.004	.005	.005	.006	.006
27	.001	.002	.002	.002	.002	.002	.003	.003	.003	.003
28	.001	.001	.001	.001	.001	.001	.001	.001	.002	.002
29	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
30					.000	.000	.000	.000	.000	.000

	q=15.1	15.2	15.3	15.4	15.5	15.6	15.7	15.8	15.9	16.0
x=0-4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	.999	.999	.999	.999	.999	.999	.999	1.00	1.00	1.00
6	.997	.998	.998	.998	.998	.998	.998	.998	.999	.999
7	.993	.993	.994	.994	.994	.995	.995	.995	.996	.996
8	.983	.984	.985	.986	.987	.987	.988	.989	.989	.990
9	.964	.966	.968	.970	.971	.973	.974	.975	.977	.978
10	.933	.936	.939	.942	.945	.947	.950	.952	.955	.957
11	.886	.891	.895	.900	.904	.908	.912	.916	.919	.923
12	.822	.828	.834	.840	.846	.852	.857	.863	.868	.873
13	.741	.749	.756	.764	.772	.779	.786	.793	.800	.807
14	.646	.656	.665	.674	.683	.692	.700	.709	.717	.725
15	.545	.555	.565	.575	.585	.594	.604	.614	.623	.632
16	.442	.452	.463	.473	.483	.493	.503	.513	.523	.533
17	.346	.355	.365	.375	.385	.394	.404	.414	.424	.434
18	.260	.268	.277	.286	.295	.304	.313	.322	.331	.341
19	.188	.195	.202	.210	.218	.225	.233	.241	.249	.258
20	.130	.136	.142	.148	.154	.161	.167	.174	.181	.188
21	.087	.092	.096	.101	.106	.111	.116	.121	.126	.132
22	.056	.059	.063	.066	.070	.073	.077	.081	.085	.089
23	.035	.037	.039	.042	.044	.047	.049	.052	.055	.058
24	.021	.022	.024	.025	.027	.029	.031	.033	.035	.037
25	.012	.013	.014	.015	.016	.017	.018	.020	.021	.022
26	.007	.007	.008	.008	.009	.010	.011	.011	.012	.013
27	.004	.004	.004	.005	.005	.005	.006	.006	.007	.007
28	.002	.002	.002	.002	.003	.003	.003	.003	.004	.004
29	.001	.001	.001	.001	.001	.002	.002	.002	.002	.002
30	.000	.001	.001	.001	.001	.001	.001	.001	.001	.001
31		.000	.000	.000	.000	.000	.000	.000	.001	.001
32									.000	.000

	q=16.1	16.2	16.3	16.4	16.5	16.6	16.7	16.8	16.9	17.0
x=0-5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	.999	.999	.999	.999	.999	.999	.999	.999	.999	.999
7	.996	.996	.997	.997	.997	.997	.997	.998	.998	.998
8	.991	.991	.992	.992	.993	.993	.993	.994	.994	.995
9	.979	.980	.981	.982	.983	.984	.985	.986	.987	.987
10	.959	.961	.963	.965	.966	.968	.970	.971	.972	.974
11	.926	.929	.932	.935	.938	.941	.944	.946	.949	.951
12	.878	.883	.887	.892	.896	.900	.904	.908	.912	.915
13	.813	.820	.826	.832	.838	.844	.849	.855	.860	.865
14	.734	.741	.749	.757	.764	.772	.779	.786	.792	.799
15	.642	.651	.660	.669	.677	.686	.695	.703	.711	.719
16	.543	.553	.563	.572	.582	.591	.601	.610	.619	.629
17	.444	.454	.464	.474	.484	.493	.503	.513	.523	.532
18	.350	.359	.369	.378	.388	.398	.407	.417	.426	.436
19	.266	.274	.283	.292	.300	.309	.318	.327	.336	.345
20	.195	.202	.209	.217	.224	.232	.240	.248	.256	.264
21	.137	.143	.149	.155	.162	.168	.174	.181	.188	.195
22	.094	.098	.103	.107	.112	.117	.122	.128	.133	.139
23	.061	.065	.068	.072	.075	.079	.083	.087	.091	.095
24	.039	.041	.044	.046	.049	.051	.054	.057	.060	.063
25	.024	.025	.027	.029	.030	.032	.034	.036	.038	.041
26	.014	.015	.016	.017	.018	.020	.021	.022	.024	.025
27	.008	.009	.009	.010	.011	.012	.012	.013	.014	.015
28	.004	.005	.005	.006	.006	.007	.007	.008	.008	.009
29	.002	.003	.003	.003	.003	.004	.004	.004	.005	.005
30	.001	.001	.001	.002	.002	.002	.002	.002	.003	.003
31	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
32	.000	.000	.000	.000	.000	.001	.001	.001	.001	.001

x	q=17.1	17.2	17.3	17.4	17.5	17.6	17.7	17.8	17.9	18.0
x=0-5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	.999	.999	.999	.999	1.00	1.00	1.00	1.00	1.00	1.00
7	.998	.998	.998	.998	.999	.999	.999	.999	.999	.999
8	.995	.995	.995	.996	.996	.996	.996	.997	.997	.997
9	.988	.989	.989	.990	.991	.991	.992	.992	.993	.993
10	.975	.976	.978	.979	.980	.981	.982	.983	.984	.985
11	.953	.955	.957	.959	.961	.963	.965	.967	.968	.970
12	.919	.922	.925	.929	.932	.935	.937	.940	.943	.945
13	.870	.875	.879	.884	.888	.893	.897	.901	.905	.908
14	.806	.812	.818	.824	.830	.836	.841	.847	.852	.857
15	.727	.735	.743	.750	.757	.765	.772	.779	.785	.792
16	.638	.646	.655	.664	.672	.681	.689	.697	.705	.713
17	.542	.551	.561	.570	.580	.589	.598	.607	.616	.625
18	.446	.455	.465	.474	.484	.494	.503	.513	.522	.531
19	.354	.363	.373	.382	.391	.400	.410	.419	.428	.438
20	.272	.280	.288	.297	.305	.314	.323	.331	.340	.349
21	.201	.209	.216	.223	.231	.238	.246	.254	.261	.269
22	.144	.150	.156	.162	.168	.174	.181	.187	.194	.201
23	.100	.104	.109	.114	.118	.124	.129	.134	.139	.145
24	.067	.070	.073	.077	.081	.085	.089	.093	.097	.101
25	.043	.045	.048	.050	.053	.056	.059	.062	.065	.068
26	.027	.028	.030	.032	.034	.036	.038	.040	.042	.045
27	.016	.017	.018	.020	.021	.022	.024	.025	.027	.028
28	.009	.010	.011	.012	.013	.013	.014	.015	.016	.017
29	.005	.006	.006	.007	.007	.008	.008	.009	.010	.010
30	.003	.003	.003	.004	.004	.004	.005	.005	.006	.006
31	.002	.002	.002	.002	.002	.002	.003	.003	.003	.003
32	.001	.001	.001	.001	.001	.001	.001	.002	.002	.002
33	.000	.000	.001	.001	.001	.001	.001	.001	.001	.001
34			.000	.000	.000	.000	.000	.000	.000	.000

x	q=18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9	19.0
x=0-6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	.999	.999	.999	.999	.999	.999	.999	.999	.999	.999
8	.997	.997	.998	.998	.998	.998	.998	.998	.998	.998
9	.993	.994	.994	.994	.995	.995	.995	.996	.996	.996
10	.985	.986	.987	.988	.988	.989	.990	.990	.991	.991
11	.971	.973	.974	.975	.976	.978	.979	.980	.981	.982
12	.948	.950	.952	.954	.956	.958	.960	.962	.964	.965
13	.912	.915	.919	.922	.925	.928	.931	.934	.937	.939
14	.862	.867	.872	.877	.881	.885	.890	.894	.898	.902
15	.798	.805	.811	.817	.823	.829	.834	.840	.845	.850
16	.721	.729	.736	.744	.751	.758	.765	.772	.779	.785
17	.634	.642	.651	.659	.668	.676	.684	.692	.700	.708
18	.541	.550	.559	.568	.577	.586	.595	.604	.613	.622
19	.447	.456	.466	.475	.484	.494	.503	.512	.521	.531
20	.358	.367	.376	.385	.394	.403	.412	.421	.430	.439
21	.277	.285	.294	.302	.310	.319	.327	.336	.344	.353
22	.208	.215	.222	.229	.236	.244	.251	.259	.267	.275
23	.151	.156	.162	.168	.174	.181	.187	.194	.200	.207
24	.106	.110	.115	.120	.125	.130	.135	.140	.145	.151
25	.072	.075	.079	.082	.086	.090	.094	.098	.102	.107
26	.047	.050	.052	.055	.058	.061	.064	.067	.070	.073
27	.030	.032	.033	.035	.037	.039	.042	.044	.046	.049

	q = 18.1	18.2	18.3	18.4	18.5	18.6	18.7	18.8	18.9	19.0
x=28	.018	.020	.021	.022	.023	.025	.026	.028	.030	.031
29	.011	.012	.013	.013	.014	.015	.016	.017	.018	.020
30	.006	.007	.007	.008	.008	.009	.010	.010	.011	.012
31	.004	.004	.004	.005	.005	.005	.006	.006	.007	.007
32	.002	.002	.002	.003	.003	.003	.003	.003	.004	.004
33	.001	.001	.001	.001	.001	.002	.002	.002	.002	.002
34	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
35	.000	.000	.000	.000	.000	.000	.000	.001	.001	.001
36								.000	.000	.000

	q = 19.1	19.2	19.3	19.4	19.5	19.6	19.7	19.8	19.9	20.0
x=0-7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	.999	.999	.999	.999	.999	.999	.999	.999	.999	.999
9	.996	.997	.997	.997	.997	.997	.997	.998	.998	.998
10	.992	.992	.993	.993	.993	.994	.994	.994	.995	.995
11	.983	.983	.984	.985	.986	.987	.987	.988	.989	.989
12	.967	.968	.970	.971	.973	.974	.975	.976	.978	.979
13	.942	.944	.947	.949	.951	.953	.955	.957	.959	.961
14	.905	.909	.912	.916	.919	.922	.925	.928	.931	.934
15	.855	.860	.865	.870	.874	.879	.883	.887	.891	.895
16	.792	.798	.804	.810	.816	.822	.827	.833	.838	.843
17	.716	.723	.731	.738	.745	.752	.759	.766	.772	.779
18	.630	.639	.647	.655	.664	.672	.680	.688	.695	.703
19	.540	.549	.558	.567	.575	.584	.593	.602	.610	.619
20	.449	.458	.467	.476	.485	.494	.503	.512	.521	.530
21	.362	.370	.379	.388	.397	.405	.414	.423	.432	.441
22	.282	.290	.298	.306	.315	.323	.331	.339	.348	.356
23	.214	.221	.228	.235	.242	.249	.257	.264	.272	.279
24	.157	.162	.168	.174	.180	.187	.193	.199	.206	.213
25	.111	.116	.121	.125	.130	.135	.141	.146	.151	.157
26	.077	.080	.084	.087	.091	.095	.099	.103	.108	.112
27	.051	.054	.056	.059	.062	.065	.068	.071	.075	.078
28	.033	.035	.037	.039	.041	.043	.045	.048	.050	.052
29	.021	.022	.023	.025	.026	.028	.029	.031	.033	.034
30	.013	.013	.014	.015	.016	.017	.018	.019	.021	.022
31	.007	.008	.009	.009	.010	.010	.011	.012	.013	.013
32	.004	.005	.005	.005	.006	.006	.007	.007	.008	.008
33	.002	.003	.003	.003	.003	.004	.004	.004	.004	.005
34	.001	.001	.002	.002	.002	.002	.002	.002	.002	.003
35	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
36	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
37					.000	.000	.000	.000	.000	.000

APPENDIX C

ALGORITHMS TO COMPUTE THE ACCEPTANCE/REJECTION REGIONS, POWER, AND ASN OF SEQUENTIAL MUAS

(Note: The following algorithms are written in Pascal, and, except for the nonstandard file "input/" and its operator "readln", syntax and usage comply with the Jensen-Wirth standard.)

```
program accrej(input/,output,tree,param,arfile);
(*computes the acceptance/rejection regions for sequential MUAS
  output to file "tree"; similar information, formatted for
  use in the program "direct", is output to the file "arfile".
  Input file "param" must contain parameters in the following
  format:
    low error rate   high error rate
    log bound for acceptance    log bound for rejection
    fixed sample size    critical value
         (other bounds/sample sizes for the same error
          rates may follow)                               *)
const
  maxlevel=50;
  maxbranch=100;
var
  h,i,j,k,  m,n,     cv,levela,levelr,levels:integer;
  loga,logb,p1,p2,denom,num,fact2,fact1a,fact1b:real;
  a,r:array[1..maxlevel,1..2] of integer;
  ar:array[0..maxlevel,1..3] of integer;
  tree,param,arfile:text;
  flag          :boolean;
begin
  rewrite(tree);
  rewrite(arfile);
  reset(param);
  readln(param,p1,p2);
  while not eof(param) do
    begin
      readln(param,loga,logb);
      readln(param,n,cv);
      denom:=ln(p2/p1)-ln((1.0-p2)/(1.0-p1));
      num:=ln((1.0-p1)/(1.0-p2));
      fact2:=num/denom;
      fact1a:=loga/denom;
      fact1b:=logb/denom;
      i:=0;
      j:=0;
      m:=0;
      while m<n do
        begin
          j:=j+1;
          m:=trunc((i-fact1a)/fact2)+1;
          a[j,1]:=m;
          a[j,2]:=i;
          i:=i+1;
        end;
```

```
levela:=j;
a[levela,1]:=n;
if levela>1
  then a[levela,2]:=a[levela-1,2]+1
  else a[levela,2]:=0;
i:=1;
j:=0;
m:=0;
while (m<n) and (i<=cv) do
  begin
    m:=trunc((i-fact1b)/fact2);
    if m>=i
      then
        begin
          j:=j+1;
          r[j,1]:=m;
          r[j,2]:=i;
        end;
    i:=i+1;
  end;
if j>=1
  then levelr:=j
  else levelr:=1;
r[levelr,1]:=n;
r[levelr,2]:=cv;
writeln(tree);
writeln(tree,'acceptance numbers:');
writeln(tree,'  m  =  a(m)');
for i:=1 to levela-1 do
  writeln(tree,a[i,1]:5,'   ',a[i,2]:5);
write(tree,a[levela,1]:5,'   ',a[levela,2]:5);
for i:=a[levela,2]+1 to cv-1 do
  write(tree,',',i:1);
writeln(tree);
writeln(tree);
writeln(tree,'rejection numbers:');
writeln(tree,'  m  =  r(m)');
for i:=1 to levelr do
  writeln(tree,r[i,1]:5,'   ',r[i,2]:5);
writeln(tree);
writeln(tree,'input data:');
writeln(tree,'p1=',p1:8:6,'  p2=',p2:8:6);
writeln(tree,'log a=',loga:8:6,'  log b=',logb:8:6);
writeln(tree,'n=',n:8,'  cv=',cv:8);
writeln(tree);
i:=1;
k:=1;
ar 0,1 :=0;
ar 0,2 :=0;
ar 0,3 :=0;
flag:=true;
```

```
      for j:=1 to levela do
        begin
          while r[i,1] < a[j,1] do
            begin
              ar[k,1]:=r[i,1];
              if flag=false
                then ar[k,2]:=ar[k-1,2]+1
                else ar[k,2]:=ar[k-1,2];
              ar[k,3]:=r[i,2]-1;
              if ar[k,1]=ar[k-1,1]
                then
                  begin
                    ar[k-1,3]:=ar[k,3];
                    k:=k-1;
                  end;
              i:=i+1;
              k:=k+1;
              flag:=true;
            end;
          ar[k,1]:=a[j,1];
          ar[k,2]:=a[j,2];
          if flag=true
            then ar[k,3]:=ar[k-1,3]+1
            else ar[k,3]:=ar[k-1,3];
          if i=1
            then ar[k,3]:=r[i,2]-1;
          k:=k+1;
          flag:=false;
        end; (*j*)
      levels:=k-1;
      for i:=1 to levels do
        if ar[i,3]>=cv-1
          then ar[i,3]:=cv-1;
      writeln(arfile);
      writeln(arfile,levels:2);
      for i:=1 to levels do
        writeln(arfile,ar[i,1]:3,' ',ar[i,2]:3,' ',ar[i,3]:3);
    end;  (*while*)
end.  (*accrej*)


program direct(input/,output,power,arfile);
(*computes power and asn for sequential muas given one or
  more acceptance/rejection regions in "arfile" as generated
  by the program "accrej"--warning: these regions are not
  ordinary regions and only output from "accrej" should be
  used.  output is to the file "power".                   *)
const
  maxlevel=50;
  maxbranch=100;
var
  i,j,k,m,n,cv,levels:integer;
  index:array[0..maxbranch] of integer;
  br,s:array[0..maxbranch] of real;
```

```
  ar:array[0..maxlevel,1..3] of integer;
  alpha,beta,en,es,x,p:real;
  power,arfile:text;
  flag,cont:boolean;
  ch:char;
function comb(n,k:integer):real;
(*computes combinations of n things taken k at a time;
  returns real value to avoid integer overflow problems*)
  var
    i,j:integer;
    tot:real;
  begin
    if (k<0) or (n<0)
      then comb:=0.0
      else if k=0
        then comb:=1.0
        else
          begin
            tot:=1.0;
            i:=n-k+1;
            j:=1;
            while (i<=n) and (j<=k) do
              begin
                tot:=tot*(i/j);
                i:=i+1;
                j:=j+1;
              end;
            comb:=tot;
          end; (*else*)
  end; (*comb*)
function biprob(n,k:integer;p:real):real;
(*computes binomial probability of k occurrences with
  parameters n and p*)
  begin
    if (p<=0.0) or (p>=1.0)
      then biprob:=0.0
      else biprob:=comb(n,k)*exp(k*ln(p))*exp((n-k)*ln(1.0-p));
  end;  (*biprob*)
begin
  rewrite(power);
  reset(arfile);
  while not eof(arfile) do
    begin
      readln(arfile,levels);
      for i:=1 to levels do
        readln(arfile,ar[i,1],ar[i,2],ar[i,3]);
      ar[0,1]:=0;
      ar[0,2]:=0;
      ar[0,3]:=0;
      ar[levels+1,1]:=ar[levels,1]+1;
      ar[levels+1,2]:=ar[levels,3]+1;
      ar[levels+1,3]:=ar[levels,3];
      n:=ar[levels,1];
      cv:=ar[levels,3]+1;
```

```
cont:=true;
while cont=true do
  begin
    for i:=0 to maxbranch do
      s[i]:=0.0;
    alpha:=0.0;
    index[0]:=0;
    index[1]:=ar 1,2 ;
    br[0]:=1.0;
    writeln('enter p, e.g. 0.05');
    readln;
    read(p);
    i:=1;
    repeat
      m:=ar[i,1]-ar[i-1,1];
      k:=index[i]-index[i-1];
      br[i]:=biprob(m,k,p)*br[i-1];
      i:=i+1;
      index[i]:=index[i-1];
      if index[i]<ar[i,2]
        then
          begin
            i:=i-1;
            alpha:=alpha+br[i];
            s[index[i]]:=s[index[i]]+br[i];
            index[i]:=index[i]+1;
            while (index[i] > ar[i,3]) and (i>0) do
              begin
                i:=i-1;
                index[i]:=index[i]+1;
              end;
        end; (*then*)
    until i=0;
    beta:=1.0-alpha;
    es:=0.0;
    for i:=1 to cv-1 do
      es:=es+(i*s[i]);
    x:=0.0;
    if (ar[1,3]<cv-1) and (ar[1,2]=0)
      then
        begin
          for i:=0 to ar[1,3] do
            x:=x+biprob(ar[1,1],i,p);
          x:=1.0-x;
        end;
    es:=es+((ar[1,3]+1)*x)+(cv*(beta-x));
    en:=es/p;
    writeln(power);
    writeln(power,'expected sample size and power:');
    writeln(power,'  p=',p:8:6,'  E(N)=',en:5:2,
      'beta(p)=',beta:8:6);
    writeln(power);
    writeln('continue on this test for another p? y=yes,
      n=no');
    readln;
```

```
            read(ch);
            if ch='y'
              then cont:=true
              else cont:=false;
          end;  (*while*)
        writeln(power,'input data:');
        writeln(power,'  m  a(m)*  r(m)-1');
        for i:=1 to levels do
          writeln(power,ar[i,1]:4,' ',ar[i,2]:4,' ',ar[i,3]:4);
        writeln(power);
        writeln(power,'* x is acceptance number only for the
          highest m such that x=a(m)');
      end;  (*eof*)
end.  (*direct*)
```

APPENDIX D

ALGORITHM TO FIND MUAS BAYES RULE

(Note: The following algorithm is written in Pascal, and, except for the nonstandard file "input/" and its operator "readln", usage conforms with the Jensen-Wirth standard. A graph of expected loss versus sample size for n*-50 to n*+50 is produced if desired.)

```
program baysamp(input/,output,loss);
  (*finds optimal fixed sample size for Bayesian MUAS and
    the corresponding sequential bounds*)
  const
    min=20;
    max=500;
    width=50;
    scale=5.0;
  var
    i,j,k,m,n,q,kstar,nstar,start,stop:integer;
    a,b,p1,p2,alpha,beta,lamb1,lamb2,c,q1,q2,Lstar,x,y,Lo,
      hi:real;
    L:array[min..max] of real;
    loss:test;
    ch:char;
    continue:boolean;
  function prob(q:integer;r:real):real;
    (*computes Poisson probability of X> =q, where q< =300
      and the Poisson parameter is r*)
    var
      i:integer;
      p:real;
      s:array[0..300] of real;
    begin
      s[0]:=exp(-r);
      p:=s[0];
      for i:=1 to q-1 do
        begin
          s[i]:=s[i-1]*r/i;
          p:=p+s[i];
        end;
      prob:=1.0-p;
    end; (*prob*)
  procedure header;
    var
      i,j:integer;
    begin
      write(loss,'  L:');
      for i:=1 to 10 do
        begin
          j:=round(10*i*scale);
          write(loss,'      ',j:4);
        end;
    end; (*header*)
```

```
begin  (*baysamp*)
  rewrite(loss);
  continue:=true;
  while continue=true do
    begin
      writeln('enter Lo and hi error rates, e.g. 0.01 0.05');
      readln;
      read(p1,p2);
      writeln('enter type I and II losses, e.g. 1000 2000');
      readln;
      read(alpha,beta);
      writeln('enter prior for Lo error rate, e.g. 0.75');
      readln;
      read(q1);
      q2:=1.0-q1;
      Lstar:=(q1*alpha)+(q2*beta);
      c:=(q1*alpha)/(q2*beta);
      for n:=min to max do
        begin
          lamb1:=n*p1;
          lamb2:=n*p2;
          x:=lamb1-lamb2;
          y:=lamb2/lamb1;
          k:=trunc((ln(c)-x)/ln(y))+1;
          L[n]:=(q1*prob(k,lamb1)*alpha)+(q2*(1.0-prob(k,lamb2))
            *beta)+n;
          if L[n]<=Lstar
            then
              begin
                Lstar:=L[n];
                nstar:=n;
                kstar:=k;
              end;
        end; (*for*)
      a:=ln((q1/q2)*(Lstar-nstar)/(beta-Lstar+nstar));
      b:=ln((q1/q2)*(alpha-Lstar+nstar)/(Lstar-nstar));
      Lo:=ln((1.0-p2)/(1.0-p1));
      hi:=ln(p2/p1);
      writeln(loss);
      writeln(loss,'test of ',p1:5:3,' vs ',p2:5:3,':');
      writeln(loss,'  prior for low rate=',q1:5:3);
      writeln(loss,'  losses: K12=',alpha:10:1,' K21=',
        beta:10:1);
      writeln(loss,'    L=',Lstar:6:1);
      writeln(loss,'    n=',nstar:6);
      writeln(loss,'    C=',kstar:6);
      writeln(loss);
      writeln(loss,'  sequential test:');
      writeln(loss,'    bounds: ',a:6:3,' ',b:6:3);
      writeln(loss,'    incrmn: ',Lo:6:3,' ',hi:6:3);
      writeln('graph of losses? y=yes, n=no');
      readln;
      read(ch);
```

```
        if ch='y'
          then
            begin
              writeln(loss);
              header;
              writeln(loss);
              writeln(loss,'n:');
              start:=nstar-width;
              if start< min
                then start:=min;
              stop:=nstar+width;
              if stop> max
                then stop:=max;
              for i:=start to stop do
                begin
                  write(loss,i:4);
                  q:=round(L[i]/scale);
                  if q< 100
                    then
                      begin
                        for j:=1 to q do
                          write(loss,' ');
                        writeln(loss,'*');
                      end
                    else
                      begin
                        for j:=1 to 99 do
                          write(loss,' ');
                        writeln(loss,'x');
                      end;
                end; (*i*)
              header;
            end; (*then*)
        writeln(loss);
        writeln('continue? y=yes, n=no');
        readln;
        read(ch);
        if ch='y'
          then continue:=true
          else continue:=false;
      end; (*while*)
end. (*baysamp*)
```

APPENDIX E

TEST POPULATION GENERATOR

(Note: The following algorithm is written in Pascal, and, except for the nonstandard file "input/" and its operator "readln", usage conforms with the Jensen-Wirth standard.)

```
program population(input/,output,dist,errpop);
  (*to generate an error population with a given random
    error pattern; output is written to the file "errpop";
    the cumulative distribution function of the desired
    relative error pattern must be input on a file called
    "dist" with the following format:
      n
      x1  F(x1)
      x2  F(x2)
      ........
      xn  F(xn)
    where xi<=1.0 for all i, F(x1)=0.0, and F(xn)=1.0,
    and n<=100*)
  const
    emax=2000;
    fmax=21;
    cmax=100;
  var
    h,i,j,k,L,over,bover,cover,under,bunder,cum,run,test,
      cellcount:integer;
    a,b,c,d,u,w,z,lo,hi,xbar,sampvar,wtvar,taint,p1,p2,
      seed,mean,variance:real;
    pop:array[0..9,1..2] of integer;
    ep:array[1..emax,1..3] of real;
    cell:array[1..fmax] of real;
    jdist:array[1..cmax,1..2] of real;
    freq:array[1..fmax] of integer;
    errpop,dist:text;
  function random(x:real):real;
    (*for 0<=x<=1 returns pseudorandom uniform(0,1) variable
      using D. Malm's generator--HP-67 Users' Library*)
    var
      y:real;
    begin
      y:=(9821*x)+0.211327;
      z:=y-trunc(y);
      random:=z;
    end; (*random*)
  function uniform(a,b:real):real;
    (*returns pseudorandom uniform(a,b) variable*)
    begin
      uniform:=((b-a)*random(z))+a;
    end; (*uniform*)
  begin (*population*)
    pop[0,1]:=0;
```

```
pop[0,2]:=0;
pop[1,1]:=1050;
pop[2,1]:=1750;
pop[3,1]:=2200;
pop[4,1]:=2550;
pop[5,1]:=3000;
pop[6,1]:=3400;
pop[7,1]:=3550;
pop[8,1]:=3800;
pop[9,1]:=4000;
pop[1,2]:=75;
for i:=2 to 9 do
  pop[i,2]:=pop[i-1,2]*2;
cell[1]:=0.0;
for i:=2 to fmax-1 do
  cell[i]:=0.05*(i-1);
cell[fmax]:=1.001;
for i:=1 to fmax do
  freq[i]:=0;
writeln ('enter run number, e.g. 1');
readln;
read(run);
reset(dist);
readln(dist,cellcount);
for i:=1 to cellcount do
  readln(dist,jdist[i,1],jdist[i,2]);
writeln('enter proportion of items in error p1');
writeln('  and proportion of 100% errors p2, p2  =p1');
writeln('  e.g. 0.05 0.01');
readln;
read(p1,p2);
writeln('enter seed, 0<seed<1, e.g. 0.4433');
readln;
read(seed);
z:=seed;
k:=0;
h:=0;
cum:=0;
cover:=0;
xbar:=0.0;
sampvar:=0.0;
wtvar:=0.0;
for j:=1 to 9 do
  begin
    for i:=pop[j-1,1]+1 to pop[j,1] do
      begin
        w:=random(z);
        if w<=p1
          then
            begin
              k:=k+1;
              ep[k,2]:=((i-pop[j-1,1])*pop[j,2])+cum;
              ep[k,1]:=ep[k,2]-pop[j,2]+1.0;
              L:=2;
              u:=random(z);
```

```
                while u>jdist[L,2] do
                  L:=L+1;
                ep[k,3]:=uniform(jdist[L-1,1],jdist[L,1]);
                if w<=p2
                  then
                    begin
                      ep[k,3]:=1.00;
                      cover:=cover+round(ep[k,2]-ep[k,1]+1.0);
                      h:=h+1;
                      freq[fmax]:=freq[fmax]+1;
                    end
                  else
                    begin
                      xbar:=xbar+ep[k,3];
                      sampvar:=sampvar+sqr(ep[k,3]);
                      L:=1;
                      while ep[k,3]>cell[L] do
                        L:=L+1;
                      freq[L]:=freq[L]+1;
                    end;
                wtvar:=wtvar+(sqr(ep[k,3])*(ep[k,2]-ep[k,1]+1.0));
              end;
      end; (*i*)
    cum:=cum+((pop[j,1]-pop[j-1,1])*pop[j,2]);
  end; (*j*)
if h<k
  then
    begin
      xbar:=xbar/(k-h);
      sampvar:=(sampvar/(k-h))-sqr(xbar);
    end;
over:=0;
bover:=0;
under:=0;
bunder:=0;
rewrite(errpop);
writeln(errpop,'run no. ',run:3);
writeln(errpop);
for i:=1 to k do
  begin
    writeln(errpop,i:4,' ',ep[i,1]:12:2,' ',ep[i,2]:12:2,
      ' ',ep[i,3]:12:4);
    taint:=ep[i,2]-ep[i,1]+1.0;
    w:=taint*ep[i,3];
    if w>0.0
      then
        begin
          over:=over+round(w);
          bover:=bover+round(taint);
        end
      else
        begin
          under:=under+round(w);
          bunder:=bunder+round(taint);
        end;
  end; (*i*)
```

```
bover:=bover-cover;
wtvar:=(wtvar/cum)-sqr((over+under)/cum);
writeln(errpop,'summary of errop pop ',run:3);
writeln(errpop);
writeln(errpop,'  parent pop:  items ',pop 9,1 :10);
writeln(errpop,'               dollars ',cum:10);
writeln(errpop);
writeln(errpop,'  error distribution:');
writeln(errpop,'      x  =',cell[1]:4:2,' ',freq[1]:4);
for i:=2 to fmax do
  writeln(errpop,' ',cell[i-1]:4:2,'  x  =',cell[i]:4:2,
    ' ',freq[i]:4);
L:=0;
for i:=2 to fmax do
  L:=L+freq[i];
writeln(errpop);
writeln(errpop,'  error mean (excl 100% over)=',xbar:8:6);
writeln(errpop,'  error var (excl 100% over) =',sampvar:8:6);
writeln(errpop,'  population var/n (eq. 92)  =',wtvar:8:6);
writeln(errpop);
a:=L/pop[9,1];
b:=over/cum;
c:=cover/cum;
d:=bover/cum;
writeln(errpop,'  overstatement:');
writeln(errpop,'    number of items(% of total)     ',L:10,
  '(',a:8:6,')');
writeln(errpop,'    book value of items overstated:');
writeln(errpop,'      partially overstated(% of total) ',
  bover:10,'(',d:8:6,')');
writeln(errpop,'      100% overstated(% of total)      ',
  cover:10,'(',c:8:6,')');
writeln(errpop,'    overstatement(% of total)          ',
  over:10,'(',b:8:6,')');
a:=freq[1]/pop[9,1];
b:=under/cum;
c:=bunder/cum;
writeln(errpop);
writeln(errpop,'  understatement:');
writeln(errpop,'    number of items(% of total)     ',
  freq 1 :10,'(',a:8:6,')');
writeln(errpop,'    book value of items(% of total) ',
  bunder:10,'(',c:8:6,')');
writeln(errpop,'    understatement(% of total)      ',
  under:10,'(',b:8:6,')');
writeln(errpop);
writeln(errpop,'  input data:');
writeln(errpop,'    seed=',seed:10:8);
writeln(errpop,'    p1  =',p1:5:4);
writeln(errpop,'    p2  =',p2:5:4);
writeln(errpop);
writeln(errpop,'  relative error cum dist (from file "dist"):');
writeln(errpop);
writeln(errpop,'      x         F(x)');
```

```
  for i:=1 to cellcount do
    writeln(errpop,'    ',jdist[i,1]:7:5,'   ',jdist[i,2]:7:5);
end.  (*population*)
```

APPENDIX F

INPUT DATA FOR TEST POPULATIONS

Test Population*	Proportion of Items in Error: p1(total)	p2(100%)	Seed
1L/.01	.0850	.0000	.3584
1H/.01	.1100	.0000	.6523
2L/.01	.0800	.0020	.0620
2H/.01	.0900	.0012	.1736
3L/.01	.0190	.0000	.6801
3H/.01	.0250	.0000	.8614
4 /.01	.0132	.0000	.7403
1L/.05	.5000	.0000	.5472
1H/.05	.4950	.0000	.7210
2L/.05	.4000	.0120	.6247
2H/.05	.4300	.0120	.8482
3L/.05	.1130	.0000	.0864
3H/.05	.0950	.0000	.1397
4 /.05	.1050	.0000	.4593

Pop.		Cumulative Distribution Function+													
1L	x:	.025	.05	.10	.15	.20	.30	.40	.50	.60					
	F_x:	.22	.39	.63	.78	.86	.95	.98	.99	1.0					
1H	x:	.025	.05	.10	.15	.20	.30	.40	.60	.80	1.0				
	F_x:	.43	.61	.74	.80	.85	.90	.93	.96	.98	1.0				
3L	x:	.30	.35	.40	.45	.50	.55	.60	.65	.70	.80				
	F_x:	.02	.07	.16	.31	.50	.69	.84	.93	.98	1.0				
3H	x:	.10	.20	.30	.35	.40	.45	.50	.55	.60	.65	.70	.80	.90	1.0
	F_x:	.01	.04	.12	.19	.28	.39	.50	.61	.72	.81	.88	.96	.99	1.0
4	x:	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.0				
	F_x:	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.0				

*legend: 1=J, 2=J-100, 3=unimodal, 4=uniform
L=low variance, H=high variance
.01 and .05 refer to the target error rates

+the c.d.f.s for populations 1 and 2 are the same; all c.d.f.s begin at x=.00 F_x=.00 except 3L which is x=.20 F_x=.00

REFERENCES

Note: the following abbreviations appear in the citations:

AMS = Annals of Mathematical Statistics
AR = Accounting Review
JAR = Journal of Accounting Research (Supp = Supplement)
JASA = Journal of the American Statistical Association
JoA = Journal of Accountancy

Abramowitz, M. and I. A. Stegun, eds. (1972), Handbook of Mathematical Functions (Dover, 1972).

American Institute of Certified Public Accountants (1981), Audit Sampling, Statement on Auditing Standards No. 39 (AICPA, June 1981).

Amster, S. J. (1963), "A Modified Bayes Stopping Rule," AMS (December 1963), pp. 1404-1413.

Anderson, R. J. and A. D. Teitlebaum (1973), "Dollar-unit Sampling," CA Magazine (April 1973), pp. 30-38.

Aroian, L. A. (1968), "Sequential Analysis--Direct Method," Technometrics (February 1968), pp. 125-132.

Basu, D. (1971), "An Essay on the Logical Foundations of Survey Sampling, Part One," in Godambe and Sprott (1971), pp. 203-233.

_____(1978), "Relevance of Randomization in Data Analysis," in Namboodiri (1978), pp. 267-292.

Berger, J. O. (1980), Statistical Decision Theory (Springer, 1980).

Bickel, P. J. and K. A. Doksum (1977), Mathematical Statistics (Holden-Day, 1977).

Birnbaum, A. (1962), "On the Foundations of Statistical Inference," JASA (June 1962), pp. 293-306.

Birnberg, J. G. (1964), "Bayesian Statistics: A Review," JAR (Spring 1964), pp. 108-116.

Carman, L. A. (1933), "The Efficiency of Tests," American Accountant (December 1933), pp. 360-366.

Cochran, W. G. (1977), Sampling Techniques (Wiley, 1977).

Cox, D. R. and E. J. Snell (1979), "On Sampling and the Estimation of Rare Errors," 66 (1979), pp. 125-132.

Cushing, B. E. (1982), "Decision-Theoretic Estimation Methods in Accounting and Auditing," in Symposium on Auditing Research IV (1982), pp. 1-52.

Demski, J. S. (1980), Information Analysis, 2d ed. (Addison-Wesley, 1980).

Duke, G. L., J. Neter, and R. A. Leitch (1982), "Power Characteristics of Test Statistics in the Auditing Environment: An Empirical Study," JAR (Spring 1982), pp. 42-67.

Elliott, R. K. (1976), comments on a paper by Roberts, in Symposium on Auditing Research I (1976), pp. 169-178.

_____ and J. R. Rogers (1972), "Relating Statistical Sampling to Audit Objectives," JoA (July 1972), pp. 46-55.

FASB (1980), Qualitative Characteristics of Accounting Information, Statement of Financial Accounting Concepts No. 2 (FASB, 1980).

Felix, W. L. and R. A. Grimlund (1977), "A Sampling Model for Audit Tests of Composite Accounts," JAR (Spring 1977), pp. 23-41.

Feller, W. (1968), An Introduction to Probability Theory and Its Applications, v. I, 3d ed. (Wiley, 1968).

Garstka, S. J. and P. A. Ohlson (1977), "Ratio Estimation in Accounting Populations With Probabilities of Sample Selection Proportional to Size of Book Values," JAR (Spring 1977), pp. 23-59.

Ghosh, B. K. (1970), Sequential Tests of Statistical Hypotheses (Addison-Wesley, 1970).

Godambe, V. P. and D. A. Sprott, eds. (1971), Foundations of Statistical Inference (Holt, Rinehart & Winston, 1971).

Govindarajulu, Z. (1981), The Sequential Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory (American Sciences Press, 1981).

Harvard University Computation Laboratory (1955), Tables of the Cumulative Binomial Probability Distribution (Harvard University, 1955).

Hill, H. P., J. L. Roth, and H. Arkin (1962), Sampling in Auditing (Ronald, 1962).

Ijiri, Y. and R. A. Leitch (1980), "Stein's Paradox and Audit Sampling," JAR (Spring 1980), pp. 91-108.

Johnson, J. R., R. A. Leitch, and J. Neter (1981), "Characteristics of Errors in Accounts Receivable and Inventory Audits," AR (April 1981), pp. 270-293.

Kaplan, R. S. (1973), "Statistical Sampling in Auditing With Auxiliary Information Estimators," JAR (Autumn 1973), pp. 238-258.

_____ (1975), "Sample Size Computations for Dollar-unit Sampling," JAR Supp (1975), pp. 126-133.

Kinney, W. R. (1975), "A Decision-Theory Approach to the Sampling Problem in Auditing," JAR (Spring 1975), pp. 117-132.

Leitch, R. A., J. Neter, R. Plante, and P. Sinha (1982), "Modified Multinomial Bounds for Larger Numbers of Errors in Audits," AR (April 1982), pp. 384-400.

Lindley, D. V. (1971), Bayesian Statistics: A Review (Society for Industrial and Applied Mathematics, 1971).

Loebbecke, J. K. and J. Neter (1975), "Considerations in Choosing Statistical Sampling Procedures in Auditing," JAR Supp (1975), pp. 38-52.

Mautz, R. K. and H. A. Sharaf (1961), The Philosophy of Auditing (American Accounting Association, 1961).

Namboodiri, N. K., ed. (1978), Survey Sampling and Measurement (Academic Press, 1978).

Neter, J. (1976), comments on a paper by Roberts, in Symposium on Auditing Research I (1976), pp. 179-184.

_____, R. A. Leitch, and S. E. Fienberg (1978), "Dollar-unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors," AR (January 1978), pp. 77-93.

_____ and J. K. Loebbecke (1975), Behavior of Major Statistical Estimators in Sampling Accounting Populations (AICPA, 1975).

_____ (1977), "On the Behavior of Statistical Estimators When Sampling Accounting Populations," JASA (September 1977), pp. 501-507.

Pratt, J. W., H. Raiffa, and R. Schlaifer (1964), "The Foundations of Decision Under Uncertainty: An Elementary Exposition," JASA (June 1964), pp. 353-375.

Roberts, D. M. (1976), "A Proposed Sequential Sampling Plan for Compliance Testing," in Symposium on Auditing Research I (1976), pp. 159-168.

_____ (1978), Statistical Auditing (AICPA, 1978).

_____, M. J. MacGuidwin, and M. D. Shedd (1982), "The Behavior of Selected Upper Bounds of Monetary Error Using PPS Sampling," in Symposium on Auditing Research IV (1982), pp. 351-378.

Savage, L. J. (1972), The Foundations of Statistics, 2d ed. (Dover, 1972).

Scott, W. R. (1973), "A Bayesian Approach to Asset Valuation and Audit Size," JAR (Autumn 1973), pp. 304-330.

_____(1975), "Auditor's Loss Functions Implicit in Consumption-Investment Models," JAR Supp (1975), pp. 98-117.

Smith, T. M. F. (1976), "The Foundations of Survey Sampling: A Review," Journal of the Royal Statistical Society, Series A (Part 2, 1976), pp. 183-195.

Symposium on Auditing Research I (1976), (University of Illinois, 1976).

Symposium on Auditing Research IV (1982), (University of Illinois, 1982).

Trueblood, R. M. and R. M. Cyert (1957), Sampling Techniques in Accounting (Prentice-Hall, 1957).

van Heerden, A. (1961), "Randomtests as a Method of Auditing," translation of "Steekproeven als middel an accountants-controle," Maandblad voor Accountancy en Bedrijfshuishoud-kunde (December 1961), pp. 453-475, provided by Prof. J. McCray in 1981.

Vance, L. L. (1950), Scientific Method for Auditing (University of California, 1950).

_____and J. Neter (1956), Statistical Sampling for Auditors and Accountants (Wiley, 1956).

von Neumann, J. and O. Morgenstern (1953), Theory of Games and Economic Behavior, 3d ed. (Princeton University, 1953).

Wald, A. (1947), Sequential Analysis (Dover, 1947).

_____(1950), Statistical Decision Theory (Wiley, 1950).

_____and J. Wolfowitz (1948), "Optimum Character of the Sequential Probability Ratio Test," AMS (September 1948), pp. 326-339.

Wetherill, G. B. (1975), Sequential Methods in Statistics, 2d ed. (Chapman & Hall, 1975).

VITA

Kermit John Rohrbach was born in California, Missouri, on 18 May 1945. He graduated from California R-1 school in 1963 and attended the University of Kansas, where he received a B.A. in Chinese language in 1967. After continuing his study of Chinese at Indiana University (where he was an I. U. Fellow, 1968-69), he entered the U. S. Army in 1969 and served in Vietnam, where he was awarded the Bronze Star Medal. Following separation from the Army at grade E-5, he resumed graduate study in Chinese at Indiana University and then entered the M.B.A. program at that university in 1972. From 1973 to 1974, he taught introductory accounting courses in the Indiana University Business School. He received an A.M. in Chinese and an M.B.A. with concentration in accounting in 1974 and passed the Uniform CPA Examination in May of that year. He joined Touche Ross and Co. in Chicago, Illinois, in 1974 as a staff auditor and was certified to practice public accounting in Indiana in 1976 after completing the two-year work experience requirement of that state. He entered the Ph.D. program in accountancy at the University of Illinois in 1976, where he taught courses in auditing for several semesters, and anticipates receiving the Ph.D. degree in fall 1983. He will join the faculty of the University of Oklahoma as an assistant professor of accounting in September 1983. He is a member of the Indiana Society of CPAs, the American Instute of CPAs, and the American Accounting Association.