ON MONTE CARLO DISTRIBUTION SAMPLING, WITH APPLICATION TO THE COMPONENT RANDOMIZATION TEST

A THESIS

Presented to

The Faculty of the Division of

Graduate Studies

By

Bruce Wayne Schmeiser

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Industrial and Systems Engineering

Georgia Institute of Technology

September, 1975

ON MONTE CARLO DISTRIBUTION SAMPLING, WITH APPLICATION

TO THE COMPONENT RANDOMIZATION TEST

Approved:

Stuart Jay Deutsch, Chairman

John J. Jarvis

Mary M. Mullet

Kenneth S. Stephens

Harrison M. Wadsworth

Date approved by Chairman: 9-29-75

ACKNOWLEDGMENTS

The chairman of my thesis committee, Dr. Stuart Jay Deutsch, provided guidance throughout this research. He both introduced me to randomization tests of hypotheses and suggested the examination of the properties of these tests as a topic. His suggestions and insight were very beneficial. His willingness to make himself available on short notice is also appreciated.

The other members of my committee, Dr. John J. Jarvis, Dr. Gary M. Mullet, Dr. Kenneth S. Stephens, and Dr. Harrison M. Wadsworth, examined the draft on short notice, despite their many other responsibilities. Their efforts and suggestions are appreciated.

Two other persons should be cited for their parts in enhancing my graduate education. Dr. W. W. Hines, Associate Director for Graduate Programs, has been most helpful with both advice and actions. Frank Alt, with whom I have shared an office for two years, has contributed his broad knowledge of statistics and an exposure to his high standards of excellence.

My family has also contributed much. My parents have always encouraged and aided my education. My wife, Donna, and son, Bryan, have spent much time without me. Donna's encouragement, good nature, and assistance are greatly appreciated.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure Page

SUMMARY

Component randomization tests of statistical hypotheses have good mathematical properties and are distribution-free. While the power of these tests is well known for asymptotically large samples, little is known for finite sample sizes. This research studies the power of the component randomization test of location in the paired sample design for finite sample sizes.

The component randomization test is computationally difficult, requiring approximately the effort of 2^n Student t tests, where n is sample size. Previously Monte Carlo studies were not applied directly since the thousands of samples necessary for adequate results would require an unreasonable amount of computer effort. Methods of Monte Carlo distribution sampling are developed to allow the Monte Carlo approach to be used. Areas of interest include quantile estimation, process generation, methods of performing the component randomization test efficiently, and variance reduction.

Quantile estimators are proposed for two cases: raw and grouped data. A linear combination of order statistics is recommended for raw data and interpolation within the group is recommended for grouped data. Confidence intervals and the effect of sample size are also treated.

Eight criteria for general process generation, the transformation of U(0,1) values to a distribution having any of a wide range of first four moments with only one functional form and four parameters,

are given. A transformation is given and developed which will attain any first four moments using only four parameters and one functional form.

A basic method and improvements are given to test the location hypothesis in the paired samples design using the component randomization test. The basic method allows generation of signs without tabled values through the use of modular arithmetic, while the improvements increase the computational speed.

The use of the "randomization sample," the 2^n dependent components arising from each sample in performing the component randomization test, as a variance reduction technique for estimating quantiles of the distribution of a sum of symmetric random variables is proposed and shown to be valid. The efficiency, examined in some detail, is shown to about 1000 times that attained when simply using a single value from each sample for n = 3 differences.

The power of the component randomization test and the corresponding, but usually unknown, parametric test is given in the form of operating characteristic curves for various underlying distributions and sample sizes. The power of the distribution-free component randomization test is seen to not be dominated by the corresponding parametric test.

CHAPTER I

INTRODUCTION

Digital computer simulation, widely applied to analyze problems too difficult for closed form solutions, has benefitted greatly from the increased capabilities of computers in the last few years. Faster computation times and increased memory have made possible the study of problems too complex for analysis a few years ago. New languages have eased the burden of coding the model.

But no matter how fast computer capabilities expand, the complexity of problems keeps pace. Therefore, digital computer simulation techniques have had to be developed to make more efficient use of the computer. Model building, through the development of simulation languages and packages, has been the focus of simulation methodology development until the last few years when emphasis shifted somewhat toward the probabilistic and statistical analysis associated with simulation. While developments in model building simplify placing the model on the computer, statistical methodology allows valid interpretation of the results (e.g. time series analysis), as well as more efficient use of the computer (e.g. variance reduction techniques). Probability is important in generating the random elements of the simulation (process generation), as well as sometimes providing a closed form analysis of a subsystem of the model (e.g., stochastic processes), thereby reducing the simulation effort.

The main focus of this work is to develop and apply the

probabilistic and statistical methodology necessary for the study of the power of several tests of hypothesis appropriate for the paired sample design, with particular emphasis on Fisher's component randomization test, using Monte Carlo techniques. While asymptotic properties of nonparametric permutation tests, of which component randomization tests are a special case, are well-known, little is known about the power of these tests for finite sample sizes. Puri and Sen [1971] point out

> The study of the power properties of nonparametric tests poses certain problems. First, . . . under suitable hypotheses of invariance, the distribution of the nonparametric test statistic does not depend on the parent population. But when the null hypothesis is not true, the sampling distribution of nonparametric statistics depends on the parent distribution in some way or other. Second, . . . permutation tests . . . are essentially conditional tests, and the study of their power properties demands the knowledge of the unconditional distribution of the allied test statistics. The evaluation of the exact unconditional non-null distribution often becomes quite a laborious task, if not impracticable. For this reason, the growth of the literature on the small-sample power properties of nonparametric tests is not at all satisfactory.

Since exact methods have not been successful in studying the small sample power properties of permutation tests, Monte Carlo analysis appears appropriate. While Monte Carlo methods are straightforward to apply to most permutation tests, the component randomization test poses computational problems which makes the direct application of these methods infeasible for all but the smallest of samples. In fact, while the component randomization test is known to have the best mathematical properties of the permutation tests (as discussed in the next chapter), its computational effort is so great (the equivalent of 2^n t tests, where n is sample size) that the test is not widely used. As an example, according to Hamm [1974] 116 CPU minutes were used to test

one hypothesis with n = 24 on a UNIVAC 1108 to obtain a result for Deutsch and Hamm [1975]. But to perform a Monte Carlo analysis, thousands of observations are necessary for reasonably accurate conclusions.

It is therefore necessary to develop efficient methodologies for many aspects of the Monte Carlo study. In particular, a technique for performing the component randomization test is needed which is very efficient and variance reduction techniques become very important. The estimation of quantiles of the distribution of the test statistic is a central problem, with accurate estimation of the quantiles at a premium due to the great expense of each observation. Since power is a function of the underlying distribution of the observations, techniques for generating random values from a wide variety of distributions are also of interest.

Following Chapter II, where background material and a literature survey are given, quantile estimation results are developed in Chapter III. Point estimators with associated variance formulae are proposed for two cases: raw and grouped data. The effect of sample size on estimation techniques is analyzed in some detail. In Chapter IV process generation, the transformation of uniformly distributed random variables to other distributions of interest, is considered. In particular, the problem of using only one functional form of two parameters to generate values having any given third and fourth moments is discussed. Criteria for general process generators are proposed and one such generator is given.

Chapters V, VI, and VII study the small-sample power properties of various tests of location shift for the paired sample case, with

special emphasis on Fisher's component randomization test. Chapter V develops a methodology for studying the component randomization test. An efficient algorithm for generating all 2^n combinations of n signs is developed, as well as the approach used to estimate the power of various paired sample tests. Chapter VI discusses the moments of the "unconditional distribution of the allied test statistics," in this case $\bar{X}\sqrt{n}/\sigma$. The moments are used as necessary conditions in estimating the distribution of $\bar{X}\sqrt{n}/\sigma$ for various underlying distributions and sample sizes. Chapter VII uses the methodologies of Chapters V and VI to evaluate the performance of the component randomization test compared to the appropriate parametric test and the inappropriate normal and t tests. Lastly, Chapter VIII discusses conclusions and directions for future research.

CHAPTER II

LITERATURE REVIEW

As discussed in the introduction, aspects of quantile estimation, process generation, and statistical hypothesis testing, with emphasis on Fisher's component randomization test, are the subject of this research. A general overview of hypothesis testing is given in Section 2.1 to lay the framework of the problem setting considered in subsequent chapters. This is followed by a discussion of the component randomization test in Section 2.2. A review of the literature related to Chapters V, VI, and VII (the paired sample randomization test, the distribution of $\overline{X}$ for various underlying distributions of X_i, and the power of various tests, respectively) is given in Section 2.3. The literature specifically associated with Chapter III (quantile estimation) and Chapter IV (process generation) is cited within those chapters as needed, allowing it to be discussed in full context after necessary background and notation have been established.

2.1 Statistical Hypothesis Testing

Statistical hypothesis testing has long been used for deciding whether or not to reject, based on a random sample of observations $(x_1, x_2, \ldots, x_n)$ a null hypothesis H_0 in favor of an alternative hypothesis H_1. If the sample observations fall in the critical region, a subset of the sample space, the null hypothesis H_0 is rejected, otherwise H_0 is not rejected. If H_0 is rejected when true, a type I error has occurred.

If H_0 is not rejected when false, a type II error has occurred. The significance level of a test, α, is the probability of a type I error and the power of a test, $(1-\beta)$, is the probability of rejecting H_0.

Ideally, test procedures would be designed so that the correct decision was always made. However, for a finite sample size, n, error can occur. It then becomes important to design the test procedure to make α and β as small as possible. The probabilities of error, α and β, are functions of H_0 and H_1, the sample size, the critical region, and the underlying distribution from which the sample is drawn.

The underlying distribution is usually determined by the situation, leaving the minimization of α and β dependent upon H_0 and H_1, sample size, and critical region. In general, the larger the sample size the less chance of error, if everything else is constant. Therefore, interest is restricted to some fixed sample size n. The choice of H_0 and H_1 depends heavily on the situation and is essentially fixed for a specific situation. By convention, H_0 and H_1 are chosen such that the strong statement is to reject H_0 at a significance level of α, so interest is constrained to a fixed value of α.

The problem then reduces to minimizing β through the selection of a critical region for fixed hypotheses, H_0 and H_1, significance level α, sample size n, and underlying distribution. This problem has a well known optimal solution for simple hypotheses through the application of the Neyman-Pearson Theorem, and a good, though not necessarily optimal, solution for composite hypotheses through the use of the likelihood ratio criterion.

However, a solution to any problem may be incorrect if any

parameters are incorrectly specified. In hypothesis testing there is little doubt about the hypotheses, α level and sample size n, since all are specified by the experimenter to fit the situation of interest. What may be of doubt is the underlying distribution. Other than in Monte Carlo simulation or areas in which much experimentation has already occurred, the experimenter may be able only to assume the distribution of the observations. The assumption made often reflects the availability, rather than appropriateness, of statistical techniques to analyze the data. If the assumed underlying distribution is incorrect, the experimenter may either over-estimate or under-estimate the power and the significance of the test. Since a basic purpose of statistical hypothesis testing is to quantify the probability of error, care needs to be taken to ensure that the underlying distribution assumption does not cause the experimenter to mislead himself or others.

Commonly, the underlying distribution of the observations is assumed to be the n dimensional multivariate normal with constant variance and zero covariance. There are three assumptions involved here: (1) normality of each observation, (2) independence between observations, and (3) constant variance.

Under these assumptions, and the additional assumption that σ^2 is a known constant, in a test of location H_0: $\mu = \mu_0$ vs. H_1: $\mu = \mu_1$, the best critical region is of the form $\Sigma_{i=1}^{n} x_i \geq c$ where the constant c is set for the desired significance level α. That is, if H_0 is true, c is the value such that $P\{\Sigma_{i=1}^{n} X_i \geq c\} = \alpha$. Under the above assumptions, $\Sigma_{i=1}^{n} X_i \sim N(n\mu_0, n\sigma^2)$ and therefore $\sqrt{n}(\bar{X} - \mu_0)/\sigma \sim N(0,1)$ where $\bar{X} = \Sigma_{i=1}^{n} X_i/n$. Here the reference distribution of the test statistic is

the standardized normal.

However, if any of the three assumptions are false, problems arise. If the underlying distribution is not normal, $\Sigma_{i=1}^{n} X_i$ is not normally distributed and the tabled values may be inappropriate, although the central limit theorem guarantees asymptotic normality. If the observations are not independent, the variance of the sum is not $n\sigma^2$, but rather $\sigma^2 n + 2\Sigma_{i=1}^{n}\Sigma_{j>i}^{n} \sigma_{ij}$ where σ_{ij} is the covariance of X_i and X_j. Non-zero σ_{ij} leads to under-estimation or over-estimation of the variance of the test statistic, and in turn the significance and power of the test. If the variance of each observation is not constant, then the true variance is $\Sigma_{i=1}^{n} \sigma_i^2$ where the variance of X_i is σ_i^2. These problems are compounded when more than one assumption is false.

With Student's development of the t distribution in 1908, the condition that σ^2 must be known was relaxed. In the t test, the statistic $\sqrt{n}(\bar{x} - \mu_0)/s$ is compared to a critical region defined in terms of the well-known t distribution, where $s^2 = (\Sigma_{i=1}^{n} x_i^2 - n\bar{x}^2)/(n-1)$ is an estimate of the unknown variance σ^2. The additional variability in the test statistic due to estimating σ^2 from the sample is reflected by using the t, rather then the normal, as the reference distribution. However, the assumptions of normality, independence, and constant variance affect the results here as in the previous case where σ^2 is known. Similar comments may be made about other <u>parameteric</u> tests of hypothesis such as those based on the chi-squared and F reference distributions.

In situations in which the underlying distribution is unknown, test procedures which are <u>robust</u> to incorrect assumptions become valuable. Ideally, procedures would be used which are <u>distribution-free</u>

but still minimize β, or, equivalently, maximize power. However, tests which make use of the exact form of the underlying distribution are generally more powerful. Thus there exists a trade-off between power and robustness.

This trade-off is made when distribution-free (or non-parametric) tests are used. These are tests in which the critical region is defined independently from considerations of underlying distribution, allowing the values of both α and β to be correctly evaluated. In a situation where the normality and independence assumptions do hold, the power of the parametric test is greater than the non-parametric test and the calculations of α and β are correct. However, the parametric test will be misleading in terms of α and β if the observations are not normally distributed, and in fact it may have less power.

Returning to the problem of minimizing β for given hypotheses, sample size, and significance level, but with no assumption about underlying distribution, the many distribution-free tests may be compared on the basis of their power. Many common tests replace the sample observation x_i by the rank of x_i. These rank tests lose power in this transformation, but are distribution free and easy to apply, needing only simple calculations and a tabled reference distribution.

2.2 Fisher's Component Randomization Test

A distribution-free test which is more powerful than other such tests, and asymptotically as powerful as the t test even under normality, is the component randomization test. Here the test statistic is calculated from the sample observations $x_1, x_2, \ldots, x_n$, not from their ranks. If under the null hypothesis any M possible values of the test

statistic are equally likely and N of these values are more indicative of H_1 being true than the observed value of the test statistic, H_0 may be rejected if α is greater than N/M.

Randomization tests may be used in many situations, including two sample tests of means and correlation, randomized blocks and other ANOVA situations, including multivariate applications. A special case of randomized blocks is the paired sample test of location shift. As an example, consider the comparison of the accuracy of two types of guns. Accuracy is measured in terms of distance (in feet) from the target, to the point of impact, and each gun is fired once at each of ten targets. If the targets vary in distance from the guns, the variance of each shot cannot be assumed constant. Let X_{ji} be the feet from the ith target for jth gun, j = 1,2. If the null hypothesis is H_0: $\mu_1 = \mu_2$ and the alternative hypothesis is H_1: $\mu_1 > \mu_2$, a paired test can be used on the data $d_i = x_{1i} - x_{2i}$, i = 1,2,...,10.

Consider the data shown in Table 1.

Table 1. Distance from Target in Feet

	Target									
	1	2	3	4	5	6	7	8	9	10
x_{1i}	7.40	6.88	7.12	7.42	7.18	6.64	6.98	7.13	6.93	7.36
x_{2i}	7.32	6.82	7.09	7.43	7.07	6.66	6.95	7.08	6.88	7.33
d_i	+.08	+.06	+.03	-.01	+.11	-.02	+.03	+.05	+.05	+.03

If H_0 is true, then each d_i could just as likely be positive or negative, since $d_i = x_{1i} - x_{2i}$ could just as well have been defined d_i =

$x_{2i} - x_{1i}$. If $\bar{d}$ is used as a test statistic, 2^{10} or 1024 possible values of $\bar{d}$ could have occurred with equal likelihood under H_0. Of these only three

+.08	+.06	+.03	+.01	+.11	+.02	+.03	+.05	+.05	+.03	$\bar{d}_1 = .047$
+.08	+.06	+.03	-.01	+.11	+.02	+.03	+.05	+.05	+.03	$\bar{d}_2 = .045$
+.08	+.06	+.03	+.01	+.11	-.02	+.03	+.05	+.05	+.03	$\bar{d}_3 = .043$

tend to indicate H_1 is true more than the observed value $\bar{d} = .041$. Therefore, H_0 may be rejected if α is greater than $3/1024 \approx .003$.

If the parametric test assumptions were made, the paired t test with 9 degrees of freedom indicates that H_0 may be rejected only if α is greater than $1 - .9957 = .0043$.

The disadvantage of component randomization tests which prevents them from being used extensively is that the randomization reference distribution is conditional upon the observed sample and therefore must be calculated each time the test is applied. Over many samples, this reference distribution is dependent upon two factors: the sample size and the underlying distribution from which the samples come. Parametric test reference distributions such as Student's t and Snedecor's F are also dependent on sample size and distribution, but the distribution is assumed to be the normal. Since the randomization reference distribution is calculated for each sample, it is automatically a function of the correct underlying distribution and therefore component randomization tests are distribution-free.

Three criteria are then important in comparing tests of hypothesis: (1) lack of assumptions, (2) power, and (3) ease of calculation. The

general types of tests discussed are each strong in two criteria and weak in a third. Parametric tests are dependent upon assumptions, rank randomization tests lack power, and component randomization tests require extensive calculations.

2.3 Synopsis of Technical Literature

The principle of randomization has been a primary consideration of experimentation since Fisher [1926] proposed its use to provide a valid estimate of error. Fisher [1935, 1936] suggested the use of randomization of components of an experiment for hypothesis testing in the matched pairs design and extended the idea to two independent samples. He advocated the use of randomization in hypothesis testing to the point of stating, "conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method."

Fisher's principle of randomization was used by Pitman [1937a] to test location in two samples of unequal size, Pitman [1937b] to test independence of two samples, Pitman [1937c] and Welch [1937] for ANOVA in randomized blocks experiments. Both Pitman and Welch used the moments of skewness and kurtosis to fit the beta distribution for use as their reference distribution, due to the extreme computation problem of applying the randomization test directly.

Nair [1940] considered the randomization test using statistics other than the mean. He found the distribution of the sample median (bell-shaped), midpoint (U-shaped), and range (J-shaped) and noted that these randomization tests could be performed without the calculation problems encountered using the sample mean. He suggested that of these

three statistics the median would in general have the best properties, although not as good as the mean. Earlier Pearson [1937] showed that for the rectangular distribution, the midpoint is more efficient than the mean, and therefore the unquestioning use of the mean in the randomization test should be avoided.

Attention then turned to the statistical properties of the tests. Almost invariably the test statistic considered was the sample mean, since most results were for asympotic large samples and were possible only through the use of the central limit theorem applied to the sample mean. Lehmann and Stein [1949] determined that the component randomization test is as efficient as the t test for normal populations. Hoeffding [1952] showed that the component randomization test is asymptotically as powerful as the corresponding parametric test.

Kempthorne [1955] discussed randomization theory in the design and analysis of experiments and concluded that at that time the gaps in the theory were:

(1) the accuracy of the approximation to randomization tests by F tests
(2) the rather stringest role of additivity (which is also present in the case of normal law inferences)
(3) the power of the randomization analysis of variance test
(4) the consideration of alternative test criteria.

Wilk [1955], Wilk and Kempthorne [1956], Zyskind [1963], Kempthorne [1966] and Kempthorne, Zyskind, Basson, Martin, Doerfler, and Carney [1966] considered randomization based ANOVA procedures in great detail. Collier and Baker [1963, 1966] generated 1000 random observations to study the power of the F test under non-normality for the randomized blocks and split-plot design, respectively.

Box and Anderson [1955] and Box and Watson [1962] used

randomization in studies of robustness and F distribution approximation, respectively.

Barton and David [1961] and Arnold [1964] extended the principle to the multivariate case. Arnold considered a statistic equivalent to Hotelling's T^2 under normality. Mardia [1971] gave a multivariate randomization test based on the Mahalonobis distance D^2. Chung and Fraser [1958] used randomization methods to develop an alternative for the multivariate two-sample problem, since T^2 is not applicable when there are only a few observations involving a large number of variables.

In addition, Chung and Fraser also suggested using subgroups of permutations (rather than all permutations) to reduce computations. The subgroup idea stemmed from Dwass [1957] who suggested using a random sample of the permutations. Heiler and Weichselberger [1969], considering the two independent sample location randomization test, developed algorithms to determine the combinations in the critical region and limit points of the confidence interval.

Kempthorne and Doerfler [1969] compared the randomization test to the sign test and the Wilcoxin test in the paired design. They concluded the randomization test is more powerful than the Wilcoxin test, which in turn is more powerful than the sign test. Bhattacharyya, Johnson, and Neave [1971] used Monte Carlo analysis to show that the power of the two sample Wilcoxin-type rank test "substantially exceeds" the power of Hotelling's T^2 for some normal shift alternatives, an interesting result because intuitively the parametric T^2 test should dominate the nonparametric Wilcoxin-type test.

Much of this research is concerned with the distribution of

$\bar{X}\sqrt{n}/\sigma$ and $\bar{X}\sqrt{n}/S$ corresponding to the normal test and Student's t test of the location shift, respectively. Much combinatorial work has been done on the problem of expressing sampling distributions in terms of cumulants of the underlying distribution. Fisher [1928] found the equations connecting the moment functions of the underlying and sampling distribution using cumulants for the case of $\bar{X}$. Thiele [1931] also used cumulants to relate the two distributions. Welker [1947] used the first four cumulants to provide a one-to-one relationship between the underlying distribution and the distribution of the sample mean for the specific case of Pearson curves. Reitsma [1963] developed approximations of means from non-normal populations by using a differential equation of the moment generating function of the sample mean and sample variance for samples from Perason type populations, both skewed and symmetric. Bradley [1971] used Monte Carlo techniques to take 10,000 samples each of size 2,4,8,...,1024 from both L-shaped and normal distributions to study convergence of the central limit theorem.

The distribution of the t statistic $\bar{X}\sqrt{n}/S$ was investigated by Sophister [1928] and Pearson [1929] using sampling experiments with random number tables. Bartlett [1935] showed that skewness affects the distribution of t more than kurtosis. Perlo [1933] derived the exact distribution of t for samples of size 3 from a rectangular distribution. Gayen [1949] derived the theoretical distribution of $\bar{X}\sqrt{n}/S$ from non-normal samples of any size, ignoring moments higher than order four, for any underlying distribution expressed by a number of terms of the Edgeworth series. Tiku [1963] obtained the distribution of t in terms of population cumulants up to the eighth order using Hermite and

Laguerre polynomials. Ractliffe [1968] used Monte Carlo techniques to conclude that

> extreme non-normality can as much as double the value of t at the 2-1/2 per cent (one tail) probability level for small samples, but increasing the sample size to 80, 50, 30, and 15 will for practical purposes remove the effect of extreme skewness, moderate skewness, extreme flatness, and moderate flatness, respectively.

CHAPTER III

ESTIMATION OF THE pTH QUANTILE

In Chapter VI the distribution of $\overline{X}$, denoted by F, for various underlying distributions, is estimated using Monte Carlo analysis. Since Monte Carlo analysis cannot yield a closed form expression for F, the pth quantile, $F^{-1}(p)$, must be estimated for any values of p of interest. This chapter addresses the problem of estimating the pth quantile given observations of the random variable. In addition to their use in Monte Carlo analysis, the results of this chapter may be used with data generated by a system simulation or by a physical process. In fact the results may be more important in the case of physical processes where observations are often more expensive and therefore fewer.

In Monte Carlo analysis a sample of X values is generated, the value of the random variable of interest is calculated from the sample, and a counter associated with the appropriate cell of a histogram is incremented. After repeating many times, the resulting histogram is used for estimation. In system simulation a similar approach is followed, except that one value of the random variable of interest results from one simulation run. Given the observed values, whatever their source, the method of combining them into quantile estimates is considered here for two all-encompassing cases: (1) large cell widths where the entries of the histogram have lost their identity through grouping, and (2) smaller cell widths where the entries of the histogram are still essentially separate entities. In the second case each cell

contains only zero or one entries with an occasional double or triple entry. This case arises when the cell widths are very small or when the number of observations are few, such as when estimating quantiles in the extreme tails of a distribution. In the first case cell widths are larger or the number of observations is relatively large, as is common in estimating quantiles in the body of the distribution. The methods of estimation for these two cases differ, since in the first case the problem is to minimize the loss of information caused by grouping and in the second case the problem is how to interpolate between sparse entries.

In this chapter, point estimators, with corresponding variance estimates, for the pth quantile are proposed for both cases. Section 3.1 considers estimators and confidence intervals for grouped data, the first case above. The second case, sparse entries, is treated in Section 3.2. The impact of sample size is discussed in some detail for both cases.

3.1 Quantile Estimation from Grouped Data

Let $F_n(s_n)$ be the cumulative distribution function, cdf, of S_n, evaluated at s_n, where S_n is a function of $X_1, X_2, \ldots, X_n$. Suppose k observations of S_n have been placed in a histogram with c_i observations in the ith cell, $i = 1,2,\ldots,Q$. The problem considered here is the estimation of the pth quantile of the distribution of S_n, $F_n^{-1}(p)$.

3.1.1 A Point Estimator

A reasonable, but crude, estimator of $F_n^{-1}(p)$ is

$$a + bq$$

where a = lower limit of the first cell,

b = cell width, and

q = smallest integer such that $\sum_{i=1}^{q} c_i \geq [p(k+1)] \equiv r$

where [X] indicates the largest integer less than or equal to X. The random variable q is simply the cell in which the rth order statistic falls. Figure 1 shows graphically the relationship of a, b, and q.

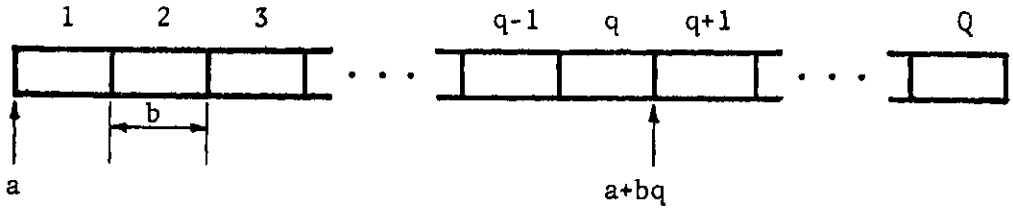

Figure 1. Relationship of a, b, and q

It is obvious from the figure that a + bq represents the upper boundary of the cell in which the rth sample order statistic lies and therefore tends to be biased high.

An alternative estimator is a + b(q-1), but similar reasoning indicates this estimate is biased low. A common compromise is to use the midpoint, a + b(q-.5), which is better but still leads to problems as discussed further in this chapter. These problems arise due to the effect of grouping and subsequent representation of the group by the midpoint (q-.5).

Since the cell q is simply the location of the rth order statistic, the impact of grouping on the properties of order statistics is of interest. David and Mishriky [1968] discuss the effect of grouping on the mean and variance of order statistics, concluding that "the effect of grouping . . . is . . . of minor importance for

h = 0.6 which represents quite course grouping." However, their results do indicate that the variance is sometimes reduced as much as 70% of its value. The effects of grouping on the mean and variance is discussed in more detail in Section 3.1.4. A possibly more important problem is the discreteness of (q-.5) caused by using only midpoint values. A discrete estimator (in this case the midpoints) of a continuous parameter (in this case a quantile) is consistent only if the true value of the parameter falls exactly at one of the possible estimator values, even though it may be essentially unbiased and possess the same variance as a related continuous (ungrouped) estimator.

These undesirable properties may be essentially eliminated by using the more complex estimator

$$\widehat{F_n^{-1}(p)} = a + b[q - (\sum_{i=1}^{q} c_i - p(k+1) + 1)/(c_q+1)] \equiv a + b[q-u]$$

The ratio u is always between zero and one. Its purpose is to reflect the degree to which the inequality $\Sigma_{i=1}^{q} c_i \geq r$ is satisfied in the definition of q. Let α denote $p(k+1) - [p(k+1)]$. If $\Sigma_{i=1}^{q} c_i = r$, then $\widehat{F_n^{-1}(p)} = a + b[q - (1-\alpha)/(c_q+1)]$ since the entire contents of the qth cell are needed to satisfy the inequality. On the other hand, if the first entry in the qth cell satisfies the inequality, then $\Sigma_{i=1}^{q} c_i - r = c_q - 1$ and the estimate is $a + b[q - (c_q-\alpha)/(c_q+1)]$--almost a whole cell width reduction compared to the previous case.

While $\widehat{F_n^{-1}(p)}$ appears quite a lot more complicated than a + b(q-.5), $\Sigma_{i=1}^{q} c_i$ is known from the calculation of q, as is p(k+1). Thus the only added complexity is a subtraction and a division. Note that the definition

of $F_n^{-1}(p)$ assumes the first case, grouped entries. It does not work well for the second case, wide-spread entries, since it is based on the idea of interpolating within a cell rather than between cells.

3.1.2 Expected Value and Variance of the pth Quantile Estimator

Ideally $\hat{F_n^{-1}}(p)$ would be defined such that $E\{\hat{F_n^{-1}}(p)\} = F_n^{-1}(p)$. Unbiasedness, however, depends on the underlying population. Specifically.

$$E\{\hat{F_n^{-1}}(p)\} = E\{a + b[q - u]\} \tag{1}$$

$$= a + bE\{q - u\}$$

$$= a + b\{\sum_{q=1}^{Q} [q - E\{(\sum_{i=1}^{q} c_i - p(k+1) + 1)/(c_q + 1)\}]P(q)\}$$

where P(q) is the probability the rth order statistic lies in the qth cell, a function of the underlying distribution.

As a special case, consider a = 0, b = 1, and a U(0,1) underlying distribution. P(1) = 1 since all observations fall in the first cell, and therefore

$$E\{\hat{F_n^{-1}}(p)\} = a + b\{\sum_{q=1}^{Q} [q - E\{(\sum_{i=1}^{q} c_i - p(k+1) + 1)/(c_q + 1)\}]P(q)\}$$

$$= 0 + 1\{[1 - (k - p(k+1) + 1)/(k+1)]1\}$$

$$= 1 - (1-p) = p = F_n^{-1}(p) \ .$$

While $F_n^{-1}(p)$ is unbiased in this case, it is not an unbiased estimator in general, since commonly $\Sigma_{i=1}^{q} c_i \neq c_q$ and also $p \neq F^{-1}(p)$ except for

the unit uniform distribution.

The variance of $\hat{F}_n^{-1}(p)$ may be calculated as

$$V\{\hat{F}_n^{-1}(p)\} = V\{a + b[q - u]\} \tag{2}$$

$$= b^2 V\{q - u\}$$

$$= b^2[V\{q\} + V\{u\} - 2\,Cov\{q,u\}]$$

$$= b^2\left[\sum_{q=1}^{Q} q^2 P(q) - \left\{\sum_{q=1}^{Q} q(P(q)\right\}^2 + \frac{1}{12}\right]$$

assuming $u \sim U(0,1)$ and independent of q.

Let $p_i = F_n(a+bi) - F_n(a + b(i-1))$, the probability that any one observation falls in the ith cell, $i = 1,2,\ldots,Q$. Numerical values for the p_i are not known, since the problem at hand is to determine $F_n^{-1}(p)$. However, each p_i may be estimated from the histogram cell counts using

$$\hat{p}_i = c_i/k$$

or more complicated estimators using smoothing rules to circumvent the problem of $c_i = 0$ not implying $p_i = 0$ in general.

Now q and the Q c_i values are random variables whose distributions depend on the underlying distribution and the definition of S_n through the p_i values. While the distribution of the c_i is simply multinomial with $N = k$ and probabilities p_i, $i = 1,2,\ldots,Q$, inclusion of this variability in the numerical calculation of the expected value or variance would require conditioning over all Q of the c_i values. Thus the simplifying assumptions of uniformity and independence were made to obtain eq. 2. If $P(q)$, $q = 1,2,\ldots,Q$ were known, $E\{a+bq\}$ and $V\{a+bq\}$

could be calculated using eqs. 1 and 2 to provide bounds on $E\{\hat{F_n^{-1}}(p)\}$ and $V\{\hat{F_n^{-1}}(p)\}$. A bound for the expected value is

$$E\{a + b(q-1)\} = a + bE\{q\} - b \leq E\{\hat{F_n^{-1}}(p)\} \leq a + bE\{q\} = E\{a+bq\}$$

A bound for the variance is

$$V\{a + bq\} = b^2V\{q\} \leq V\{\hat{F_n^{-1}}(p)\} \leq b^2[V\{q\} + 1/4]$$

since the minimum variance of the correction factor is zero and the maximum variance is 1/4.

3.1.3 Distribution of the Cell Containing the pth Quantile Estimator

To use the bounds of the last section for the expected value and variance, the values of P(q), the probability of the estimator falls in cell q, q = 1,2,...,Q, must be known in terms of the p_i values, the probability of any one observation falling in cell i, i = 1,2,...,Q.

For q = 1, at least r = [p(k+1)] observations must fall in the first cell. With a total of k observations, and c_i of them in cell i,

$$p(1) = \text{Prob}\{c_1 \geq r\}$$

$$= \sum_{i=r}^{k} \binom{k}{i} p_1^i (1-p_1)^{k-i}$$

since the independent observations lead to a sum of binomial probabilities. Now for $q \geq 2$

$$P(q) = \text{Prob}\{\sum_{i=1}^{q-1} c_i < r \leq \sum_{i=1}^{q} c_i\} \tag{3}$$

$$= \text{Prob}\{c_q \geq r - \sum_{i=1}^{q-1} c_i \text{ and } \sum_{i=1}^{q-1} c_i < r\}$$

$$= \sum_{j=0}^{r-1} [\text{Prob}\{c_q \geq r-j\}][\text{Prob}\{\sum_{k=1}^{q-1} c_i = j\}] \text{ by conditioning}$$

$$= \sum_{j=0}^{r-1}\left[\sum_{i=r-j}^{k-j}\binom{k-j}{i}\left(\frac{p_q}{1-\sum_{\ell=1}^{q-1} p_\ell}\right)^i\left(1-\frac{p_q}{1-\sum_{\ell=1}^{q-1} p_\ell}\right)^{k-j-i}\right]$$

$$\cdot\left[\binom{k}{j}\left(\sum_{\ell=1}^{q-1} p_\ell\right)^j\left(1-\sum_{\ell=1}^{q-1} p_\ell\right)^{k-j}\right]$$

$$= \sum_{j=0}^{r-1}\left[\sum_{i=r-j}^{k-j}\binom{k-j}{i}\binom{k}{j}\left(\frac{p_q}{1-\sum_{\ell=1}^{q-1} p_\ell}\right)^i\left(\sum_{\ell=1}^{q-1} p_\ell\right)^j\left(1-\frac{p_q}{1-\sum_{\ell=1}^{q-1} p_\ell}\right)^{k-j-i}\right.$$

$$\left.\cdot\left(1-\sum_{\ell=1}^{q-1} p_\ell\right)^{k-j}\right]$$

$$= \sum_{j=0}^{r-1}\left[\sum_{i=r-j}^{k-j}\binom{k}{ij} p_q^i\left(\sum_{\ell=1}^{q-1} p_\ell\right)^j\left(1-\sum_{\ell=1}^{q} p_\ell\right)^{k-j-i}\right]$$

where $\binom{k}{ij} = k!/[i!j!(k-i-j)!]$.

Appendix A contains a proof that $\sum_{q=1}^{Q} P(q) = 1$ and that eq. 3 is therefore a valid distribution.

A heuristic approach to determining P(q) is to note that each observation may fall to the left of cell q, in cell q, or to the right

of cell q. For k observations there are $\binom{k}{ij}$ ways that j observations are to the left of cell q, i observations are in cell q, and k-j-i observations are to the right of cell q. Equation 3 may therefore be viewed as trinomial probabilities summed over all i+j = r with $i \geq 0$ and $j \geq 0$.

Note that P(q) depends on p only through r = [p(k+1)]. Therefore, if a + b(q-.5) is the estimator used, for three places accuracy in p, k must be at least 1000 due to the discrete nature of (q-.5). The use of the correction factor regains some of this accuracy.

The above results may be used to calculate approximate confidence intervals on $F_n^{-1}(p)$. Assuming normality (this assumption is discussed in Section 3.2) and given a histogram with k observations, the p_i values could be estimated, P(q) values calculated, and bounds on $V\{\widehat{F_n^{-1}(p)}\}$ calculated in turn. An approximate confidence interval would be

$$\widehat{F_n^{-1}(p)} \pm Z_{\alpha/2}\sqrt{b^2[V\{q\} + \frac{1}{12}]}$$

where $b^2/12$ corresponds to an assumption of a uniform distribution of observation within the qth cell.

The analysis of the next section indicates that this approach may work well for small b, but for larger b values the confidence interval could be quite inaccurate.

3.1.4 Effect of Cell Width and Sample Size on Estimation

At first glance equations 1 and 2 may appear to indicate that as cell width b goes to zero

$$\lim_{b\to 0} E\{\widehat{F_n^{-1}(p)}\} = a$$

and

$$\lim_{b \to 0} V\{\hat{F_n^{-1}}(p)\} = 0 \ .$$

However, P(q) depends on the value of b. As b goes to zero the use of a histogram becomes equivalent to explicitly saving, rather than grouping, all values and using order statistics rather than cell counts. This case is considered in Section 3.2.

The value of b does affect the expected value and variance, however. Using the formulae derived above, the effect of b on $E\{a + b(q-.5)\}$ and $V\{a + b(q-.5)\}$ may be examined. The special case of normally distributed observations with zero mean and unit variance is considered.

3.1.4.1 Effect on the Expected Value. Table 2 shows the effect on expected value for various numbers of observations k. Quantile values of .5, .7, .9, .95, and .99 are considered for each k. The two entries in each cell of the table corresponds to a = -4 with b = 1 and b = .1, respectively. The most important point to notice is that the position of $F_n^{-1}(p)$ in the histogram cell is the dominating factor. If the cells are labeled in ascending order from left to right, $F_n^{-1}(.7)$ = .5025 is very close to the center of the fifth cell for b = 1, but lies on the boundary of cells 45 and 46 for b = .1. For all values of k shown in the table, the estimate using the larger b is less biased than the estimate using the smaller b value for p = .7.

Of course, everything else constant, smaller b values are better. For example, $F_n^{-1}(.5) = 0.0$ which lies on cell boundaries for both b = 1 and b = .1. In this case b = .1 dominates b = 1 in terms of unbiasedness. Since the value of $F_n^{-1}(p)$ is not known in application, its

Table 2. Values of $E\{a + b(q-\frac{1}{2}) \mid k,p\}$ for b = 1 and b = .1

k \ p	.5	.7	.9	.95	.99
$F_n^{-1}(p)$	0.0	.5025	1.281	1.645	2.330
100	-.038	.500	1.435	1.526	2.175
	-.0126	.509	1.249	1.491	2.148
50	-.0561	.500	1.327	1.596	2.236
	-.025	.494	1.218	1.629	2.220
30	-.0722	.493	1.228	1.612	2.025
	-.0415	.473	1.179	1.616	2.025
20	-.0884	.471	1.146	1.425	1.859
	-.062	.448	1.131	1.408	1.856
15	0	.516	1.258	1.733	
	0	.516	1.258	1.727 →	
10	-.134	.388	.997	1.538	
	-.123	.376	1.001	1.533 →	
7	0	.357	1.350		
	0	.353	1.348 →		
5	0	.496	1.161		
	0	.495	1.160 →		
4	-.297	.297	1.028		
	-.297	.297	1.027		
3	0	.845			
	0	.845 →			
2	-.564	.564			
	-.564	.563 →			
1	0 →				
	0 →				

position in relation to cell boundaries is unknown. Certainly the effect is minimized in general by smaller values of b.

Smaller cells widths, however, lead to more cells and in turn more computer memory requirements. The value of b should be set as small as possible while still allowing the histogram to be kept in core. With large computers this is little problem. For example $b = .002\sigma$ requires 5000 cells for a range of ten standard deviations, which should be plenty for all but very heavy tailed distributions. However, with the expanding use of minicomputers, or for applications requiring many histograms to be stored at once, setting b to small values may be difficult. In these cases the correction factor is important, since its function can be seen to be the reduction of the cell border effect.

3.1.4.2 Effect on the Variance. The importance of the correction factor may also be seen in terms of $V\{a + b(q-.5)\}$. Table 3 is analogous to the previous table, but entries correspond to $k * V\{a + b(q-.5)\}$. The variance is multiplied by the number of observations k to eliminate the effect of k on the variance. Note again the irregularity for $p = .7$ due to $F_n^{-1}(p)$ falling at the center of a cell for $b = 1$ and on a cell boundary for $b = .1$. For $k > 20$ the correct cell is almost always selected for $b = 1$, while for $b = .1$ the cell containing $F_n^{-1}(.7)$ is far from certain even for $k = 100$ resulting in a higher variance for $b = .1$. Again the correction factor tends to eliminate the cell border effect.

A final point on the value of the correction factor can be made by examining the entries for $k = 4$ and $p = .5$ and .7 in both tables. Since $r = 2$ for either value of p, the entries are identical. The correction factor uses additional information in the histogram to provide

Table 3. Values of $k^*V\{a + b(q-\frac{1}{2})|k,p\}$ for b = 1 and b = .1

k \ p	.5	.7	.9	.95	.99
100	24.84 1.676	.0175 1.841	6.085 2.942	2.713 4.257	23.606 9.790
50	12.342 1.607	.341 1.753	7.251 2.765	5.286 4.329	19.438 17.384
30	7.345 1.577	1.067 1.708	6.222 2.620	5.270 4.160	11.345 9.551
20	4.862 1.557	1.719 1.670	4.985 2.475	4.423 3.210	7.440 6.406
15	3.804 1.539	2.018 1.691	4.121 2.700	5.820 → 4.986 →	
10	2.610 1.519	2.198 1.589	3.131 2.154	4.352 → 3.633 →	
7	2.165 .2113	2.078 1.544	3.385 → 2.827 →		
5	1.875 1.438	1.958 1.562	2.675 → 2.276 →		
4	1.776 1.445	1.776 1.445	2.309 → 1.990 →		
3	1.597 1.349	1.932 → 1.690 →			
2	1.529 1.365	1.531 → 1.368 →			
1	1.083 → 1.0008 →				

estimates which distinguish between $F_n^{-1}(.5)$ and $F_n^{-1}(.7)$.

3.2 Quantile Estimation Using Order Statistics

Section 3.1 discussed estimation of $F_n^{-1}(p)$ where observations were grouped in histogram cells of width b. Consider the limiting case as b approaches zero. The number of cells Q becomes large and the number of observations in the ith cell, c_i, becomes small. At the limit each observation is known exactly and the sample order statistics may be used directly to estimate $F_n^{-1}(p)$.

Recording each observation and later sorting into ascending order to determine the various order statistics is possible, but time consuming for large sample size k. Therefore, the value of the results of this section may lie more in the case where b is not zero, but rather where the value of b is so small that $E\{c_i\} < 1$ for almost all $i = 1,2,\ldots,Q$. In this case the effect of grouping is negligible. At the same time the correction factor $(\Sigma_{i=1}^{q} c_i - p(k+1) + 1)/(c_q+1)$ makes little sense and the formulae of the last section for confidence intervals cannot be used numerically since the calculation of P(q) for $q = 1,2,\ldots,Q$ requires too much effort for small values of b. Therefore a new estimator of $F_n^{-1}(p)$ and a different technique for estimating the variance of the estimate of $F_n^{-1}(p)$ is necessary for the case of small b values. Section 3.2.1 presents a point estimator, Section 3.2.2 discusses the effect of sample size on estimation, and Section 3.2.3 discusses the confidence interval for the pth quantile.

3.2.1 Point Estimation of $F_n^{-1}(p)$ Using Order Statistics

A common estimator for $F_n^{-1}(p)$ is the rth order statistic, denoted

$X_{(r)}$ or $X_{r:k}$, where $r = [p(k+1)]$. This estimator, however, is biased low as can be seen by considering a special case. Let $p = .8$ and $k = 10$. Then $r = 8$. The .8 quantile is the point where 80% of the distribution is to the left and 20% is to the right. In this case, however, the estimate $X_{(8)}$ is the point where 70% of the sample is to the left, and 20% is to the right. Crudely allocating the 10% represented by $X_{(8)}$ to the left and right of $X_{(8)}$ indicates that $X_{(8)}$ may be a better estimator of $F_n^{-1}(.75)$ than of $F_n^{-1}(.8)$.

An alternative is to use $X_{(r+1)}$ as the estimator, but a similar problem arises, but with positive bias. It should be noted that as the sample size increases, the bias in using either order statistic diminishes.

A better estimator may be based on

$$E\{X_{r:k}\} \approx F_n^{-1}(r/(k+1))$$

which is an exact equality for a uniform distribution [Gibbons, 1971, p. 36]. Define $\widehat{F_n^{-1}(p)}$ as $(1-\alpha)X_{r:k} + \alpha X_{r+1:k}$ where $r = [p(k+1)]$, $\alpha = p(k+1) - [p(k+1)]$ and $p \in [1/(k+1), k/(k+1)]$. Values of p outside this interval result in the indicated use of the non-existent zero and k+1 order statistics. Estimation of $F_n^{-1}(p)$ outside the above interval is possible by making assumptions on the form of F_n or increasing k.

$\widehat{F_n^{-1}(p)}$ is unbiased for uniformly distributed observations, since

$$E\{\widehat{F_n^{-1}(p)}\} = E\{(1-\alpha)X_{r:k} + \alpha X_{r+1:k}\}$$

$$= (1-\alpha)E\{X_{r:k}\} + \alpha E\{X_{r+1:k}\}$$

Now approximately for all distributions and exactly for the uniform

$$E\{\hat{F_n^{-1}}(p)\} = (1-\alpha)F_n^{-1}(r/(k+1)) + \alpha F_n^{-1}((r+1)/(k+1))$$

For the uniform (0,1) distribution

$$= (1-\alpha)(r/(k+1)) + \alpha((r+1)/(k+1))$$

$$= (r+\alpha)/(k+1) = p(k+1)/(k+1) = p = F_n^{-1}(p)$$

$\hat{F_n^{-1}}(p)$ is analogous to the grouped data estimator $a + b[q - (\Sigma_{i=1}^{q} c_i - p(k+1)+1)/c_q+1)]$ discussed in the last section in that both make the estimate less discrete by assuming, as an approximation, a uniform distribution of observations in the area of the estimate. However, here interpolation is being performed between cells rather than within cells.

3.2.2 Effect of Sample Size

The expected values of $X_{(r)}$ and $X_{(r+1)}$ become close for asymptotically large samples. Therefore there is some number K such that for $k > K$ the added accuracy of using the linear combination of order statistics, rather than one or the other, is small enough that the extra effort is not worthwhile.

Section 3.2.2.1 examines the effect of sample size on the expected difference of $X_{(r)}$ and $X_{(r+1)}$. Section 3.2.2.2 examines the effect on the variance of the order statistics. Techniques are developed in both sections to determine the value of K necessary to use only one order statistic for satisfactory results.

3.2.2.1 Effect on the Expected Difference of $X_{(r)}$ and $X_{(r+1)}$.

Table 4 shows the relationship of k and $E\{X_{(r)}\}$ and $F_n^{-1}(p)$ for p = .8, .9 in the special case of uniformly distributed observations. The weighting of the two order statistics eliminates the fluctuation in the estimates for both p values. The use of the weighted estimates in calculating the values of Table 2 would have eliminated the fluctuation under p = .5, where nonzero values occur for all even sample sizes and all odd sample sizes have zero entries due to using the r = [p(k+1)] order statistic rather than a weighted average of $X_{(r)}$ and $X_{(r+1)}$. Since $E\{X_{r:k}\} \approx F_n^{-1}(r/k+1)$ for all distributions, Table 4 also indicates less fluctuation and bias for arbitrary distributions.

The necessity of the α correction factor becomes less as k becomes larger for constant b, since as k increases, $E\{c_i\}$ increases for each i and the expected difference between $X_{(r)}$ and $X_{(r+1)}$ becomes small, making interpolation not worthwhile. It is of interest to study the minimum sample size K necessary for

$$|E\{X_{(r)}\} - E\{X_{(r+1)}\}| < \delta$$

for some given $\delta > 0$ and all $k > K$. K, of course, depends on both the value of p and the distribution of observations.

Figure 2 shows the relationship between sample size k and $E\{X_{(r)}\}$ for p = .9 and .99 and for the uniform and normal distributions. The uniform results were obtained directly using $E\{X_{r:k}\} = r/(k+1)$ while the normal distribution results were calculated using numerical integration as follows:

Table 4. Relationship of Sample Size, Expected Values of Order Statistics, and $\hat{F}_n^{-1}(p) = (1-\alpha)E\{X_{r:k}\} + \alpha E\{X_{r+1:k}\}$ for p = .8 and p = .9 and U(0,1) Observations

p=.8 \ k	8	9	10	11	12	13	14	15	16	17
pk	6.4	7.2	8	8.8	9.6	10.4	11.2	12	12.8	13.6
r = [p(k+1)]	7	8	8	9	10	11	12	12	13	14
r+1	8	9	9	10	11	12	13	13	14	15
$E\{X_{r:k}\} = r/(k+1)$	.778	.8	.727	.75	.769	.786	.8	.75	.765	.778
$E\{X_{r+1:k}\} = (r+1)/(k+1)$	.889	.9	.817	.833	.846	.857	.867	.813	.823	.833
$\alpha = p(k+1) - [p(k+1)]$	.2	0	.8	.6	.4	.2	0	.8	.6	.4
$\hat{F}_n^{-1}(p)$	.8	.8	.8	.8	.8	.8	.8	.8	.8	.8
p=.9 \ k	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
pk	7.2	8.1	9	9.9	10.8	11.7	12.6	13.5	14.4	15.3
r = [p(k+1)]	8	9	9	10	11	12	13	14	15	16
r+1	9	10	10	11	12	13	14	15	16	17
$E\{X_{r:k}\} = r/(k+1)$	.889	.9	.818	.833	.846	.857	.867	.875	.882	.889
$E\{X_{r+1:k}\} = (r+1)/(k+1)$	1	1	.909	.917	.923	.929	.933	.938	.941	.944
$\alpha = p(k+1) - [p(k+1)]$	.1	0	.9	.8	.7	.6	.5	.4	.3	.2
$\hat{F}_n^{-1}(p)$	.9	.9	.9	.9	.9	.9	.9	.9	.9	.9

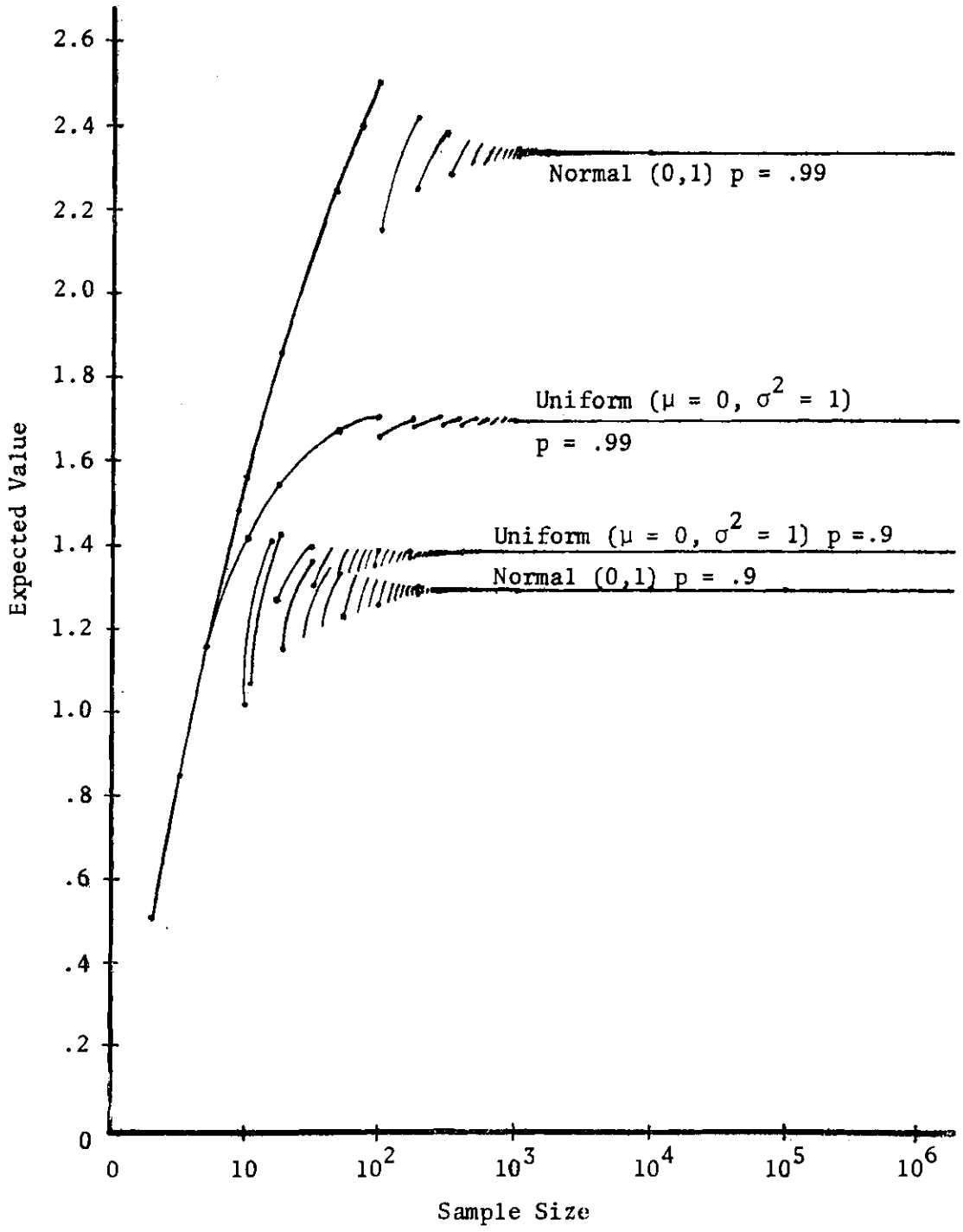

Figure 2. Sample Size vs. Expected Value Using Cell Midpoints

$$E\{X_{r:k}^{P}\} = \int_{-\infty}^{\infty} X^{P} f_{X_{r:k}}(X)\,dx$$

$$= \Delta\{\sum_{i=1}^{\infty} (X_{min} + i\Delta)^{P} f_{X_{r:k}} (X_{min} + i\Delta)\}$$

$$= \Delta\{\sum_{i=1}^{\infty} (X_{min} + i\Delta)^{P} [k!/(r-1)!(k-r)!][F_X (X_{min} + i\Delta)]^{r-1}$$

$$f_X (X_{min} + i\Delta)[1 - F_X (X_{min} + i\Delta)]^{k-r}\}$$

$$= \Delta r\binom{k}{r}\{\sum_{i=1}^{\infty} (X_{min} + i\Delta)^{P} [F_X (X_{min} + i\Delta)]^{r-1}$$

$$f_X (X_{min} + i\Delta)[1 - F_X (X_{min} + i\Delta)]^{k-r}\}$$

where Δ is a small positive value and X_{min} is $F_n^{-1}(.001)$ or less.

It is obvious from Figure 2 that the behavior of $E\{X_{(r)}\}$ depends heavily on both p and the distribution of the observations, with interaction effects. Consider first p = .99. The fluctuation in $E\{X_{(r)}\}$ is much greater for the normal distribution than for the uniform, because the short tails of the uniform distribution and high value of p provide little freedom for variation. On the other hand, for p = .9 the fluctuation depends much less on distribution type since the finite upper bound on the observations in the uniform case plays a lesser role with the smaller value of p.

It is desirable to have a simpler method of studying this fluctuation than the numerical integration and graphing technique just described. The minimum sample size K necessary to bound the

fluctuation below some given $\delta > 0$ may be calculated as follows:

Noting the longest fluctuations occur where k increases by one but r remains the same, consider

$$E\{X_{r:k-1}\} - E\{X_{r:k}\}$$

$$\approx F_n^{-1}(r/k) - F_n^{-1}(r/(k+1))$$

$$= F_n^{-1}(r/k) - F_n^{-1}(r/k - r/(k(k+1)))$$

$$\approx F_n^{-1}(\frac{p(k+1)}{k}) - F_n^{-1}(p(1 - 1/(k+1))$$

$$\approx F_n^{-1}(\frac{p(k+1)}{k}) - F_n^{-1}(p)$$

If this fluctuation is to be bounded by $\delta > 0$, then

$$F_n^{-1}(p(k+1)/k) < \delta + F_n^{-1}(p)$$

This relationship, though approximate, gives good numerical results, since the error in approximating $E\{X_{r:k-1}\}$ by $F_n^{-1}(r/k)$ is similar to the error in approximating $E\{X_{r:k}\}$ by $F_n^{-1}(r/(k+1))$, yielding an approximate difference with little error.

As an example, consider N(0,1), p = .99 and δ = .01. Then

$$F_n^{-1}(p(k+1)/k) < \delta + F^{-1}(p)$$

implies

$$F_n^{-1}(.99(k+1)/k) < + .01 + F^{-1}(.99)$$

$$F^{-1}(.99(k+1)/k) \leq .01 + 2.324 = 2.334$$

Solving by trial and error for the smallest k satisfying this relationship, gives $K \approx 3000$. For $\delta = .1$ a value of $K \approx 300$ is obtained. Reference to Figure 2 provides the same results. The conclusion is that for $k < K$, $\hat{F}_n^{-1}(p)$ should be used to estimate $F_n^{-1}(p)$ rather than $X_{(r)}$ or $X_{(r+1)}$.

3.2.2.2 Effect on the Difference of the Variance. If an estimate of the variance of the estimate of $F_n^{-1}(p)$ is desired, an analysis analogous to that for the expected value must be performed, since $V\{X_{r:k}\}$ can differ greatly from $V\{X_{r:k-1}\}$. The value of k must be large enough to reduce this fluctuation below some value $\delta > 0$ for the simpler order statistic estimators to be applicable.

Figure 3 is a plot of sample size k vs. $k*V\{X_{r:k}\}$. (The variance is multiplied by the sample size to eliminate the effect of k.) Here again the results are highly dependent upon both p and the distribution of the observations, with interaction effects. The least amount of fluctuation is for $p = .99$ and the uniform distribution, the greatest amount of fluctuation is for $p = .99$ and the normal distribution. The point here is not the specific fluctuations, but that the fluctuation depends heavily on both sample size and p. Calculations similar to those for expected value can be used to find K such that $|V\{X_{r:k-1}\} - V\{X_{r:k}\}| < \delta$ for a given $\delta > 0$ and all $k > K$. Using

$$V\{X_{r:k}\} \approx \frac{r(k-r+1)}{(k+1)^2(k+2)} \{f_X [F_X^{-1}(r/(k+1))]\}^{-2}$$

[Gibbons, 1971, p. 36] and $r = [p(k+1)]$, consider

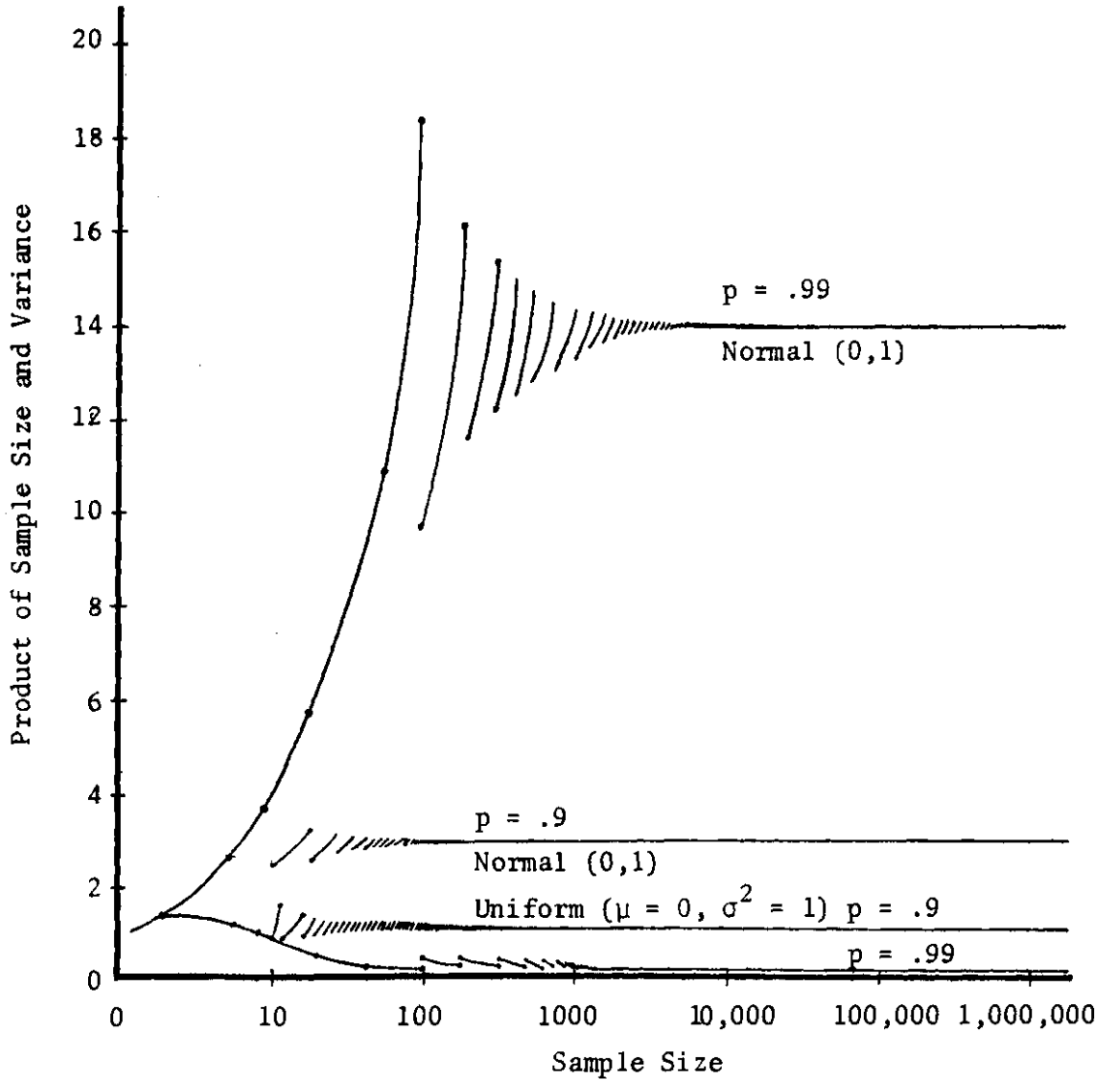

Figure 3. Sample Size vs. Standardized Variance Using Cell Midpoints

$$|V\{X_{r:k-1}\} - V\{X_{r:k}\}|$$

$$\approx \left|\frac{r(k-r)}{k^2(k+1)} \{f[F_n^{-1}(r/k)]\}^{-2} - \frac{r(k-r+1)}{(k+1)^2(k+2)} \{f[F_n^{-1}(\frac{r}{k+1})]\}^{-2}\right|$$

$$= \frac{r}{k+1} \left|\frac{(k-r)}{k^2} \{f[F_n^{-1}(r/k)]\}^{-2} - \frac{(k-r+1)}{(k+1)(k+2)} \{f[F_n^{-1}(\frac{r}{k+1})]\}^{-2}\right|$$

$$\approx \frac{p(k+1)}{k+1} \left|\frac{k-p(k+1)}{k^2} \{f[F_n^{-1}(\frac{p(k+1)}{k}]\}^{-2} - \frac{(k-p(k+1)+1)}{(k+1)(k+2)} \{f[F_n^{-1}(P)]\}^{-2}\right|$$

$$= p\left|\frac{k(1-p)-p}{k^2} \{f[F_n^{-1}(\frac{p(k+1)}{k})]\}^{-2} - \frac{(k+1)(1-p)}{(k+1)(k+2)} \{f[F_n^{-1}(P)]\}^{-2}\right|$$

$$\approx p(1-p)\left|\frac{1}{k} \{f[F_n^{-1}(\frac{p(k+1)}{k})]\}^{-2} - \frac{1}{k+2} \{f[F_n^{-1}(P)]\}^{-2}\right|$$

$$\approx \frac{p(1-p)}{k}\left|\{f[F_n^{-1}(\frac{p(k+1)}{k})]\}^{-2} - \{f[F_n^{-1}(P)]\}^{-2}\right|$$

Setting this quantity less than $\delta > 0$, the fluctuation is bounded to the desired level for all k larger than K, the smallest k satisfying the inequality. For example, if the observations are normally distributed, $p = .99$, and $\delta = .001$, trial and error indicates $k \approx 3800$. This value is verified by Figure 3.

3.2.3 Confidence Intervals Using Order Statistics

Blum and Rosenblatt [1963, 1963b] discuss the existence of multistage estimation procedures for finding a confidence interval of preassigned length and confidence for quantiles. Since multistage procedures do not exist for all distributions, the results are not of direct interest here.

Gibbons [1971, pp. 40-3], among others, discusses distribution-

free confidence intervals for quantiles based on order statistics. In particular

$$1 - \alpha = P(X_{r:k} < F_n^{-1}(p) < X_{s:k}) = \sum_{i=r}^{s} \binom{k}{i} p^i (1-p)^{k-i}$$

which is solved for r and s to minimize $X_{s:k} - X_{r:k}$ or s - r for a given α. This confidence interval has the four disadvantages that: (1) it is not unique, (2) it is determined by trial and error, (3) it assumes the order statistics are from independent observations, and (4) it is not as short as parametric confidence intervals. The first two disadvantages are obvious. The third is a disadvantage since in Monte Carlo studies many variance reduction techniques are based on causing correlation between the observations, yielding this confidence interval invalid. Observations from computer simulation or real world processes are often dependent also, due to their time series nature. The fourth disadvantage stems from the distribution-free nature of the interval which corresponds to the two-tailed sign test for quantiles. While the sign test is UMPU for an unspecified distribution [Kendall and Stuart, 1973, pp. 533-6], a valid normal theory confidence interval would certainly have better properties, corresponding to the greater power associated with knowledge of the distribution of the estimator of $F_n^{-1}(p)$.

Consider the possibility of using a normal theory confidence interval based on m independent estimates $\widehat{F_{ni}^{-1}(p)}$ i = 1,2,...,m, where each $\widehat{F_{ni}^{-1}(p)}$ is defined as the [pℓ]th sample order statistic of ℓ observations. Define $\widehat{F_n^{-1}(p)} = \Sigma_{i=1}^{m} \widehat{F_{ni}^{-1}(p)}/m$ based on k = mℓ

observations. Then if each $\widehat{F^{-1}_{ni}(p)}$ is normally distributed with unknown variance, a $100(1-\alpha)\%$ confidence interval on the pth quantile is

$$\overline{\widehat{F^{-1}_{n}(p)}} \pm t_{\alpha/2,m-1} \cdot S/m$$

where

$$S = \left[\sum_{i=1}^{m} (\widehat{F^{-1}_{ni}(p)} - \overline{\widehat{F^{-1}_{n}(p)})^2/(m-1)}\right]^{1/2}$$

This confidence interval is unique and closed form, obvious advantages over the distribution-free confidence interval. In addition, only the $\widehat{F^{-1}_{ni}(p)}$ values must be independent, rather than each observation, thereby allowing the use of variance reduction techniques or, in system simulation, the ith run. Finally, this confidence interval has better properties if, in fact, the normality assumption is valid.

Gibbons [1971, p. 40] shows that as $\ell \to \infty$ and $p = r/\ell$ remains fixed, the distribution of $X_{(r)}$ tends to normality with mean $F_n^{-1}(p)$ and standard deviation $[p(1-p)/\ell]^{1/2}\ [f(F_n^{-1}(p))]^{-1}$, where f is the density function corresponding to $F_n(\cdot)$. Thus for ℓ "large enough", the normal theory confidence interval based on m independent estimates is valid.

Consider the allocation of the k observations into m groups of ℓ for fixed $k = m\ell$. Asymptotically,

$$\overline{V\{\hat{F}_n^{-1}(p)\}} = V\{\hat{F}_{ni}^{-1}(p)\}/m$$

$$= [f^{-2}(F_n^{-1}(p))p(1-p)/\ell]/m$$

$$= f^{-2}(F_n^{-1}(p))p(1-p)/(m\ell)$$

Therefore the allocation does not affect the variance of the estimate for asymptotically large k. For purposes of calculating the normal theory confidence interval, however, the allocation is important. Certainly $m \geq 2$ is necessary to make possible the computation, and $m > 5$ is desirable since the value of $t_{\alpha/2,m-1}$ increases so rapidly for $m \leq 5$. The tradeoff, however, is that each $F_{ni}^{-1}(p)$ is more closely normal for large ℓ, made possible by small m.

If the confidence interval is being used as a stopping rule, approximate results are acceptable and some non-normality causes no harm. Especially if intervals are being calculated for several values of p at once for purposes of determining when sufficient accuracy has been obtained, approximate results are acceptable since the variability depends so much upon the value of p as shown in Figures 2 and 3.

In applications where valid confidence intervals are needed, the normal theory approach could be misleading. However, in Monte Carlo studies, k is usually at least in the thousands. The analysis of 3.2.1 indicated that large sample sizes are necessary to reduce fluctuation in expected value and variance of $X_{(r)}$. For the large samples, normality is probably not a bad assumption, since 100 observations is usually plenty for appeal to the central limit theorem for confidence intervals

on the mean. Therefore, even if accurate confidence intervals are needed, the normal theory approach could be considered.

CHAPTER IV

PROCESS GENERATION FROM (β_1,β_2)

In order to provide the observations used to estimate quantiles in Chapter III and to perform the Monte Carlo work of Chapters V, VI, and VII, a method must be available to generate random values having specific properties. Techniques for generating values from specific distributions are discussed in many places, for example Naylor et al. [1966] or Lewis [1975].

This chapter considers the problem of generating random variables having any specified first four central moments μ_1, μ_2, μ_3, and μ_4. While the first four moments do not completely characterize a distribution, different distributions with common lower order moments tend to have similar properties. Thus the capability of generating random variables having any specified first four moments would allow one technique to be used to approximate all common distributions. In addition sensitivity analysis would be simplified since generation of random values having moments $\mu_1 + \varepsilon_1$, $\mu_2 + \varepsilon_2$, $\mu_3 + \varepsilon_3$, and $\mu_4 + \varepsilon_4$ would be no more difficult than μ_1, μ_2, μ_3, and μ_4. The capability of generating a wide variety of random variables is used in the current work to generate observations having specified properties. The first four moments are used to measure the difference between underlying distributions in Chapters V, VI, and VII.

Section 4.1 discusses the well-known relationships between distributions in terms of the third and fourth moments, which are

measures of skewness and tailweight, respectively. Section 4.2 gives criteria for comparing process generation techniques. Section 4.3 discusses existing techniques of generating random variables having specified first four moments and compares them in terms of the criteria of Section 4.2. Finally, Section 4.4 presents a process generator designed to generate random values from a wide range of distributions. This generator is the one of several tried which satisfies the most criteria of Section 4.2.

4.1 The (β_1,β_2) Plane and Common Distributions

In this section the interpretation of the first four moments and their relationship to some common distributions is examined. Insight into this relationship is the basis for desiring a generator capable of producing values having any first four moments. Section 4.1.1 discusses standardized moments and their interpretation. Section 4.1.2 discusses the relationship of these moments to common distributions.

4.1.1 Interpretation of Moments

Any distribution with fixed parameters has a unique set of moments. That is, if μ'_r is the rth raw moment of a random variable X, then μ'_r is unique for all $r \geq 0$, which is obvious from the definition of the rth raw moment

$$\mu'_r \equiv \int_{-\infty}^{\infty} x^r f(x)\,dx$$

Thus, any distribution may be represented by the point $(\mu'_1,\mu'_2,\ldots,\mu'_s)$ in an s dimensional Euclidean space if the first s moments are finite.

The uniqueness does not follow in the reverse direction, however, since more than one distribution may fall at the point $(\mu_1', \mu_2', \ldots, \mu_s')$, even if s is arbitrarily large. A common example is the lognormal distribution, which cannot be defined by its moments, as shown by Heyde [1963]. Despite this nonuniqueness, insight into the relationship between distributions may be gained by observing where various distributions are located in terms of their moments. Commonly the first four moments are considered, since four is the maximum number of moments which may be analyzed graphically. The graphical analysis is made possible by using standardized moments

$$\alpha_r \equiv \mu_r / \mu_2^{r/2} \qquad r = 1,2,\ldots$$

where μ_r is the rth central moment. The standardized moments do not depend on either μ_1' (the mean) or μ_2 (the variance), as may be seen from the equivalent definition of α_r

$$\alpha_r \equiv \int_{-\infty}^{\infty} \left| \frac{(x-\mu_1')}{\mu_2^{1/2}} \right| f(x)\, dx \qquad r \geq 0$$

Thus any distribution with finite first four moments may be placed on a plane with axes α_3 and α_4, without regard to the first or second moments since $\alpha_1 = 0$ and $\alpha_2 = 1$ for all distributions. The third and fourth standardized moments are intimately related to the shape of the distribution. Both are invariant to variance (scale) and mean (location). The third standardized moment is often used as a measure of skewness. Its value is negative for a tail to the left and positive for a tail to the right. All symmetric distributions have

$\alpha_3 = 0$, but $\alpha_3 = 0$ is not a necessary condition for symmetry. However, $\alpha_3 = 0$ for asymmetric distributions only rarely. The fourth standardized moment is a measure of tail weight, a more difficult concept because it is so easily confused with variance, and is discussed in Section 4.1.2 in the context of specific examples.

4.1.2 Common Distributions on the (β_1,β_2) Plane

The interpretation of the standardized moments becomes more clear when considering the location of common distributions on the third and fourth standardized plane. From knowledge of the shapes of these common distributions (as discussed in Johnson and Kotz [1969, 1970a, 1970b], for example), a feeling for the relationship becomes more clear. Figure 4 shows the third and fourth moment plane, using $\beta_1 = \alpha_3^2$ and $\beta_2 = \alpha_4$ as the scales for the two axes. The positions of the common distributions placed on the plane indicate the relationship between the shape of a distribution and the values of its third and fourth moments.

Consider the line $\beta_1 = \alpha_3^2 = 0$. The beta distribution, with equal parameters, falls on this axis for β_2 such that $1 \leq \beta_2 \leq 3$. A special case is the equally likely Bernoulli trial which is located at $(\beta_1,\beta_2) = (0,1)$, corresponding to the limit as $p = q$ goes to zero. Another special case is the uniform distribution ($p = q = 1$) at $(\beta_1,\beta_2) = (0,1.8)$. The normal distribution falls at (0,3), the double exponential at (0,6), the t distribution with five degrees of freedom at (0.9), and the t distribution with four degrees of freedom at $(0,\infty)$.

Note that β_2, the kurtosis, increases as the weight of the tails of the distribution increases. This is illustrated in Figure 5. Distributions with finite range, such as the beta, have kurtosis less than

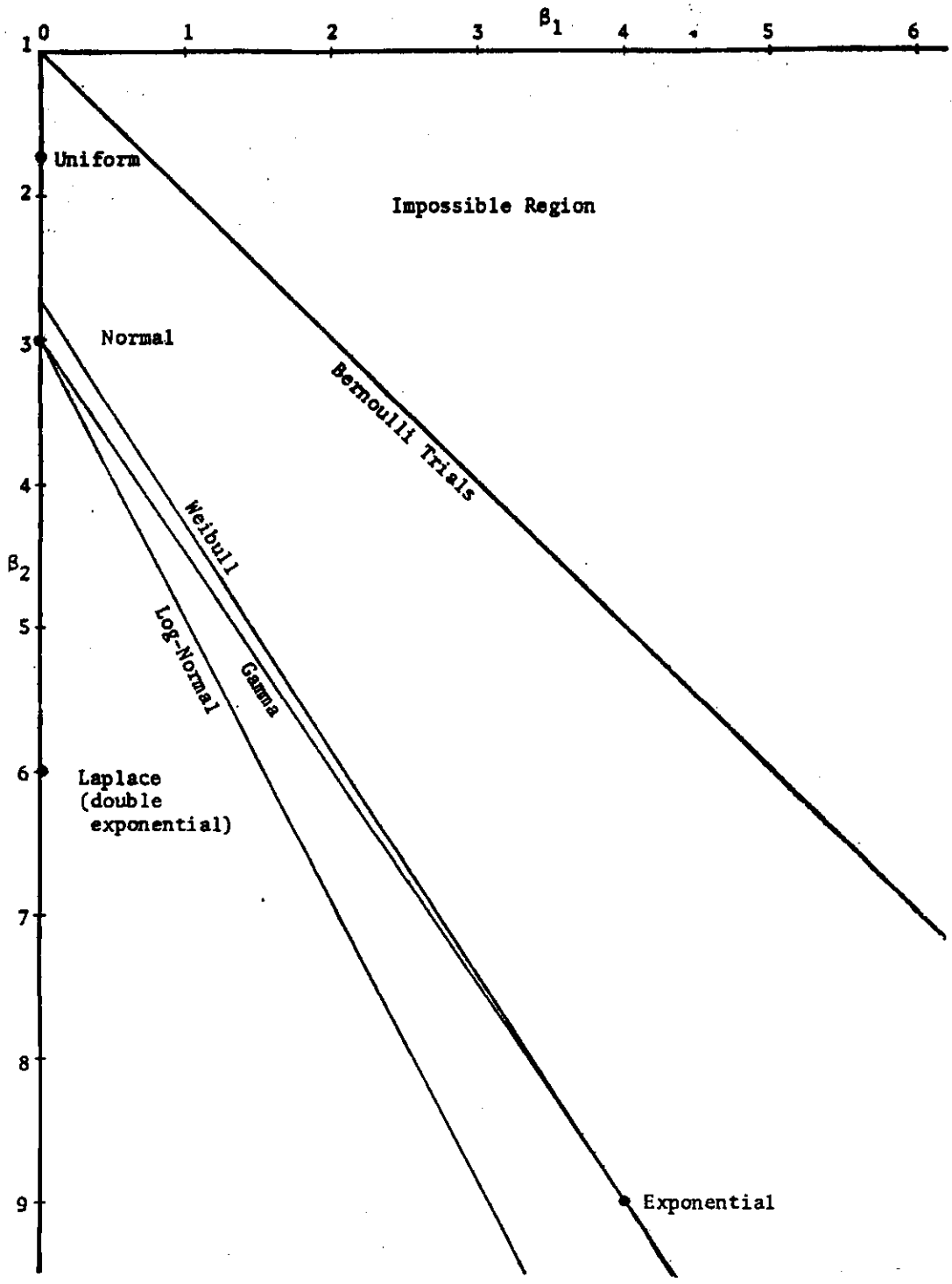

Figure 4. Common Distributions and the (β_1, β_2) Plane

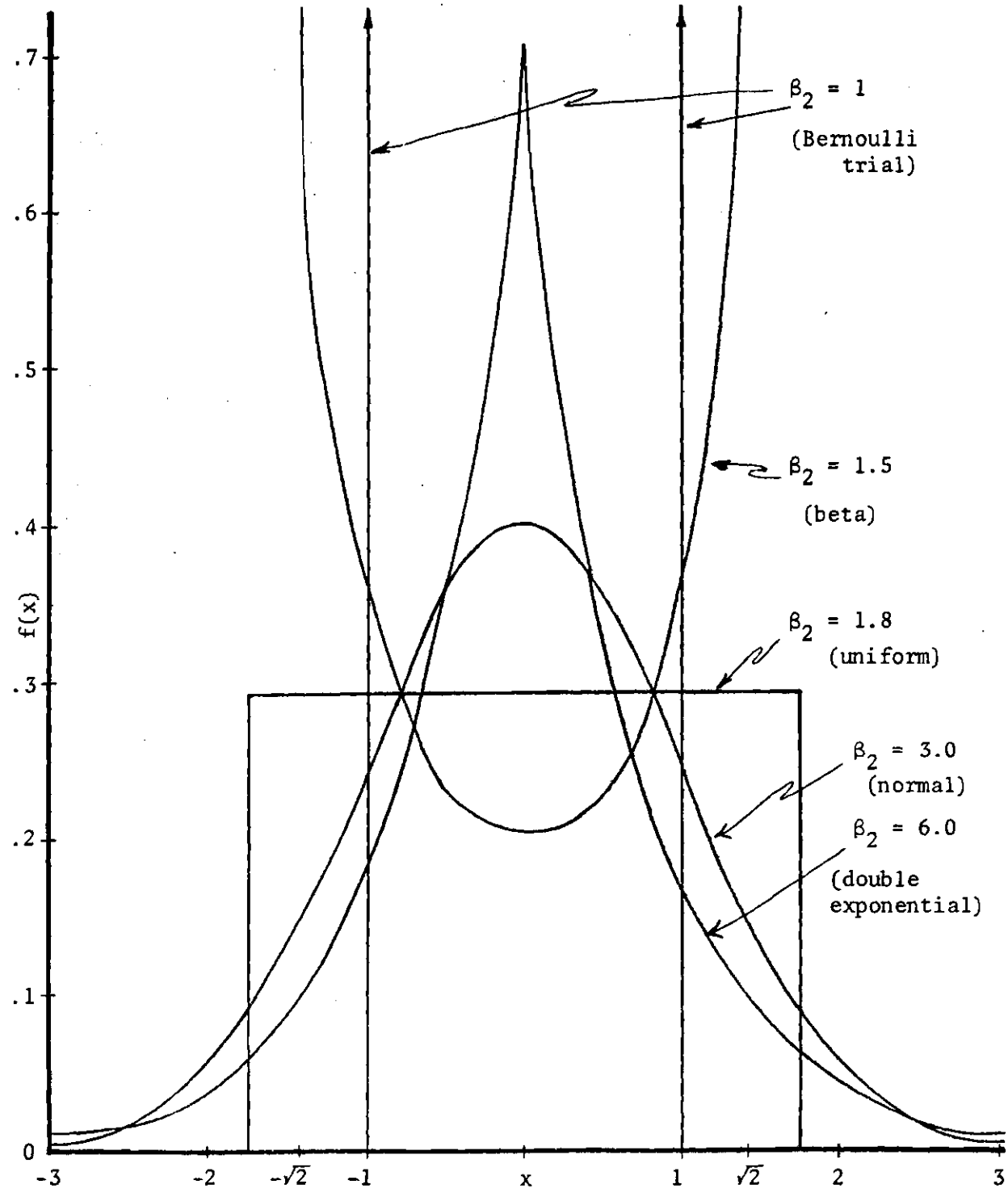

Figure 5. Relationship Between β_2 and Tail Weight or Peakedness

three. The normal distribution has the lightest tails possible with an infinite range at $\beta_2 = 3$. The double exponential has still thinner "shoulders" and heavier tails. Note that β_2 is also a measure of the "peakedness" of the distribution.

Just as β_2 measures tail weight or peakedness, β_1 is a measure of skewness. Several common skewed distributions are shown in Figure 4. The gamma, Weibull, and lognormal distributions correspond to straight lines. (Note that had α_3 rather than $\beta_1 = \alpha_3^2$ been used on the horizontal scale, these straight lines would be curves, thereby explaining the use of β_1.) The exponential distribution, located at $(\beta_1,\beta_2) = (4,9)$, is a special case of both the gamma and Weibull, so both of these lines pass through this point. The normal distribution is the limiting case of both the lognormal and the gamma distributions, forcing these lines to intersect at (0,3). The beta distribution, which is much more flexible, covers the entire area above the gamma line. Special cases include skewed Bernoulli trials on the line $\beta_2 = \beta_1 + 1$ (as p and q approach zero in a fixed ratio) and the gamma distribution (as p and q approach infinity at a fixed ratio).

From knowledge of the general shapes of these distributions it is seen that β_1 is an increasing function of skewness. Since skewness and kurtosis, as measured by β_1 and β_2, are so intimately related with the shape of the distribution, a random generator capable of producing values having any specified β_1 and β_2 is also capable of producing values from a very wide range of distribution shapes.

4.2 Criteria for General Process Generators

Any process generation technique oriented toward a wide range of (β_1,β_2) values, rather than a particular common distribution, will be called a "general process generator" in the remainder of this chapter. Without specifying the specific technique, a general process generator may be denoted

$$x = G(\lambda_1,\lambda_2,\ldots,\lambda_k,p_1,p_2,\ldots)$$

where the k λ_i values are parameters, $\{p_i\}$ is a sequence of U(0,1) values, and G is either a closed form expression or an iterative algorithm.

Eight criteria for comparing general process generators are given and discussed in this section. Not included are criteria common to all process generators, such as computational speed, core requirements, cycle length, and reproducibility. It is assumed that any technique considered will not require extensive tables, will have long cycles, and will be reproducible. Computational speed is almost always a consideration, especially in Monte Carlo applications, but does not need to be belabored here. The eight criteria pertinent to general process generation are now discussed.

4.2.1 One Functional Form

Ideally the process generator should have only one functional form. That is, G is the same for all (β_1,β_2), differing only in the values of the parameters λ_i. The use of several forms of G to obtain a wide range of (β_1,β_2) values complicates the logic, making error more likely. In addition, several forms necessitate several parameter

estimation methods, one for each form.

4.2.2 Number of Parameters

A technique should require only one parameter for each moment to be fit. That is, to obtain any first four moments the generator should be of the form

$$x = \lambda_1 + G(\lambda_3,\lambda_4,p_1,p_2,\ldots) \cdot \lambda_2$$

where λ_1 determines the mean (location), λ_2 determines the variance (scale), and λ_3 and λ_4 determine skewness and kurtosis. Finding a functional form for G using only two parameters to fit a wide range of (β_1,β_2) values is complicated by many of the following criteria.

4.2.3 Expressibility of F in Closed Form

Ideally the cdf of X, F, may be expressed in closed form. That is, G should be such that the pth quantile of X is given by

$$p = F(\lambda_3,\lambda_4,(x-\lambda_1)/\lambda_2)$$

Usually, although not always, this criterion implies that X is a function of only a single U(0,1) value p_1. That is,

$$x = \lambda_1 + G(\lambda_3,\lambda_4,p_1)*\lambda_2$$

Of course, the use of only one U(0,1) value also increases the computational speed of the technique. The value of being able to express the cdf, and the density function, in closed form arises in related statistical analyses, such as maximum likelihood estimation to estimate the parameter values or to determine the quantile of a particular value

of X.

4.2.4 Expressibility of F^{-1} in Closed Form

If G is a closed form function of only one U(0,1) value, then G is simply the inverse cdf, F^{-1}, of the standardized random variable $(X-\lambda_1)/\lambda_2$. When F^{-1} can be expressed in closed form the inverse transformation technique can be used. This technique has several advantages over the other three techniques mentioned earlier, including ability to evaluate the pth quantile directly, ease of coding, and ease of applying variance reduction techniques such as stratified sampling (see discussion of the subroutine STRAT in Chapter VI).

4.2.5 The Moments as Functions of the Parameters

The form of G should allow the moments to be expressed as a closed form expression of the parameters. The simpler the relationship, the better, since (1) it is often desirable to provide the parameter values to a computer routine to have the moment values calculated, and (2) roundoff error can be a problem in complicated functions.

4.2.6 The Parameters as Functions of the Moments

It is desirable that, given the specified first four moments, the appropriate parameter values may be easily determined. Ideally a closed form expression is obtained, although only seldom is this possible. Commonly, having determined λ_3 and λ_4 for specified (β_1,β_2) by nonlinear programming techniques, tables, or by trial and error, the appropriate λ_2 value for the desired variance and the appropriate λ_1 value for the desired means have closed form solutions.

4.2.7 Symmetry

Symmetric distributions are an important special case of random

variables. The general form of G should be such that some parameter values provide symmetric random variables. In particular, the specification of $\beta_1 = 0$ is often meant to imply symmetry. Therefore a favorable characteristic of a technique is that $\beta_1 = 0$ implies symmetry.

4.2.8 Shape

The distribution should have a "nice" shape. That is, given the conditions which the distribution must satisfy, such as certain specified moments,

1. the mode (or antimode) should be unique,
2. the range of X should be continuous,
3. f(x) should be continuous,
4. f(x)/dx should exist for all x, and
5. f(x)/dx should be continuous for all x.

These five properties are given in a rough order of importance. Obviously the conditions are not always desirable, since a Bernoulli trial, for example, satisfies none of them.

4.3 General Process Generation

Many techniques of process generation are available for generating values from a wide range of third and fourth moment values--some designed for this purpose and some not. In this section some of these techniques are compared in terms of the criteria of Section 4.2.

Uniformly distributed random values on the (0,1) interval are the basic elements of randomness in digital computer simulation and Monte Carlo work. Process generation is the transformation of these uniform (0,1) values to the distribution of interest. Common transformation techniques include:

1. rejection methods, where a random observation x is used with probability $f(x)/c$, where $c \geq \max_x f(x)$,
2. composition methods, where the cdf of the distribution of interest is approximated by piece-wise linear segments, which is equivalent to using a mixture of uniform distributions,
3. application of statistical relationships, such as summing several uniform random values to approximate the normal distribution using the central limit theorem, and
4. the inverse transformation technique, which uses the inverse cdf $x = F^{-1}(p)$ as a direct transformation from $p \sim U(0,1)$ to x having cdf F(x). For example,

$$x = - \ln(1-p)/\lambda$$

may be used to generate exponentially distributed random values with parameter λ.

In terms of the eight criteria of the last section, the inverse transformation technique appears to be the superior approach for general process generators, since all of the criteria are possible with this technique. Rejection methods do not work well with heavy tailed distributions due to the assumption of finite range. Composition methods require many parameters, in particular the position of each linear segment of the cdf approximation. The use of statistical relationships is by its nature applied to a specific distribution.

Despite this apparent superiority, many methods used to study distributions over a wide range of moments use other approaches. Most of these were not designed for random number generation but are sometimes

used. These methods are now studied in terms of the criteria of Section 4.2.

4.3.1 The Pearson System

The most well-known system of distributions covering the whole (β_1,β_2) plane is that due to K. Pearson [1895]. He used seven density functions $p(x)$, all satisfying a differential equation of form

$$\frac{1}{p}\frac{dp}{dx} = -\frac{a + x}{c_0 + c_1 x + c_2 x^2}$$

where the four parameters determine the shape of the distribution. Of course any mean and variance can be attained by the usual transformation $y = b(x+c)$. While seven types are used, there are no discontinuities in the system owing to their derivation.

In terms of the criteria of Section 4.2, however, this system does not fare too well. The method of estimating a, c_0, c_1, and c_2, given β_1 and β_2, differs by region of the plane and is complicated by the necessity of determining four parameters. The system was not devised for efficient generation of random values, although Cooper, Davis, and Dono [1965] have created a computer program for producing random variables having Pearson type distributions. Finally, in several cases neither the cdf nor inverse cdf have a closed form.

4.3.2 The Johnson System

Johnson [1949] used three transformations of a normal random variable to cover the (β_1,β_2) plane: one above the lognormal line, the lognormal, and one below the lognormal line. In terms of the criteria of Section 4.2, the disadvantages of the Johnson system for process

generation are the use of multiple functional forms, and the lack of a closed form for the cdf and the inverse cdf. Random values can be generated using any of several techniques to generate a normal random value and making the appropriate transformation, although the computational effort is high.

4.3.3 Bessel Function Distributions

McKay [1932] described two distributions, which are based on modified Bessel functions, that cover the whole (β_1,β_2) plane. The solution of a cubic equation leads to the appropriate parameter values given β_1 and β_2, a definite advantage to this approach. Disadvantages are multiple functional forms and lack of closed form cdf and inverse cdf. It would appear difficult to generate random values from these distributions.

4.3.4 The Burr Distribution

Burr [1942, 1973] considers the c.d.f.

$$1 - (1+y^c)^{-k} \quad (0 < y) \quad (k > 0)$$

where c and k are constants determining shape. The standard transformation of y gives the desired mean and variance. This approach is well suited to random value generation by setting

$$p = 1 - (1+y^c)^{-k} \quad \text{where } p \sim U(0,1)$$

to obtain

$$y = \sqrt[1/c]{p^{-1/k} - 1}$$

This expression may be used to generate values of y from values of p using the inverse transformation technique. Here the qth quantile of y is calculated using p = 1-q, a property which is good for checking the accuracy of approximations to data or known distributions.

Disadvantages of the distribution in terms of the criteria are few. The greatest is that the U-shaped ($\beta_1 + \beta_2 \lesssim 1.8$) region of the (β_1,β_2) plane is not accessible. Less important is that the distribution is always asymmetric and extensive tables are needed to determine c and k given β_1 and β_2.

4.3.5 The Generalized Lambda Distribution

Ramberg and Schmeiser [1972, 1974] proposed using

$$x = \lambda_1 + (p^{\lambda_3} - (1-p)^{\lambda_4})*\lambda_2$$

where $p \sim U(0,1)$, which is a generalization of Tukey's lambda distribution. All criteria are satisfied except:

1. the cdf does not exist in closed form,
2. light tailed distributions cannot be generated,
3. the parameters λ_3 and λ_4 are not closed form functions of the moments.

The generalized lambda distribution covers more of the (β_1,β_2) plane than the Burr distribution and includes symmetric distributions as a special case ($\lambda_3 = \lambda_4$), which the Burr distribution does not.

4.4 The Absolute Lambda Distribution

Several general inverse cdf's were analyzed in an attempt to find one which dominated the techniques of Section 4.3 in terms of the

criteria of Section 4.2. While none dominated, the one which performed best is described in this section. This generator has one functional form, the property that only four parameters are needed to provide any first four moments, a closed form cdf, a closed form inverse cdf, a relatively simple parameter-moment relationship, symmetry as a special case, and computational speed which is better than any technique of Section 4.3.

Section 4.4.1 discusses the distributional properties of the absolute lambda distribution, Section 4.4.2 discusses its relationship to the (β_1,β_2) plane, and Section 4.4.3 discusses a variation of the absolute lambda distribution with somewhat different relationship to the third and fourth moment plane.

4.4.1 Distributional Properties

The absolute lambda distribution is based on the inverse cdf

$$x = F^{-1}(p) = \lambda_1 + p'(\lambda_3,\lambda_4)*\lambda_2$$

where

$$p' = \begin{cases} -(\lambda_4 - p)^{\lambda_3} & \text{if } p \leq \lambda_4 \\ (p - \lambda_4)^{\lambda_3} & \text{if } p > \lambda_4 \end{cases}$$

where p is distributed U(0,1), $-\infty \leq \lambda_3 \leq \infty$, $0 \leq \lambda_4 \leq 1$, and λ_2 has the same sign as λ_3.

The density function, in terms of the inverse cdf, is $[dF^{-1}(p)/dp]^{-1}$; which may be easily seen from

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 \left[\frac{dF^{-1}(p)}{dp}\right]^{-1} \left[\frac{dF^{-1}(p)}{dp}\right] dp = \int_0^1 dp = 1$$

Therefore

$$f(x) = f(F^{-1}(p)) = \left[\frac{dF^{-1}(p)}{dp}\right]^{-1} = (\lambda_2\lambda_3)^{-1}|p-\lambda_4|^{1-\lambda_3}$$

The cdf, F, may be expressed in closed form as

$$F(x) = p = \begin{cases} \lambda_4 - \left[\dfrac{\lambda_1 - x}{\lambda_2}\right]^{1/\lambda_3} & \text{if} \quad \lambda_1 - \lambda_2\lambda_4^{\lambda_3} \le x \le \lambda_1 \\ \lambda_4 + \left[\dfrac{x-\lambda_1}{\lambda_2}\right]^{1/\lambda_3} & \text{if} \quad \lambda_1 \le x \le \lambda_1 + \lambda_2(1-\lambda_4)^{\lambda_3} \end{cases}$$

by simply solving for p in the expression $x = F^{-1}(p)$.

Therefore the density function of X can be expressed in terms of x, in addition to the previous expression in p. In particular

$$f(x) = \frac{dF(x)}{dx} = \left[\frac{|\lambda_1 - x|}{\lambda_2}\right]^{(1-\lambda_3)/\lambda_3}$$

for all $x \in [\lambda_1 - \lambda_2\lambda_4^{\lambda_3}, \lambda_1 + \lambda_2(1-\lambda_4)^{\lambda_3}]$.

For $\lambda_1 = 0$ the kth raw moment of the distribution is

$$\begin{aligned} E\{X^k\} &= \int_{-\infty}^{\infty} x^k f(x)\,dx \\ &= \int_0^1 [F^{-1}(p)]^k dp \\ &= \int_0^{\lambda_4} [-(\lambda_4 - p)^{\lambda_3}\lambda_2]^k dp + \int_{\lambda_4}^1 [(p-\lambda_4)^{\lambda_3}\lambda_2]^k dp \end{aligned}$$

$$= \lambda_2^k[(-1)^k \int_0^{\lambda_4} (\lambda_4-p)^{k\lambda_3}dp + \int_{\lambda_4}^{1} (p-\lambda_4)^{k\lambda_3}dp]$$

Then if $k\lambda_3 > -1$

$$E\{X^k\} = \lambda_2^k\left[\frac{(-1)^k(\lambda_4-p)^{k\lambda_3+1}}{k\lambda_3+1}\Bigg|_0^{\lambda_4} + \frac{(p-\lambda_4)^{k\lambda_3+1}}{k\lambda_3+1}\Bigg|_{\lambda_4}^{1}\right]$$

$$= \lambda_2^k\left[\frac{(-1)^k\lambda_4^{k\lambda_3+1}}{k\lambda_3+1} + \frac{(1-\lambda_4)^{k\lambda_3+1}}{k\lambda_3+1}\right]$$

$$= \frac{(-1)^k\lambda_4^{k\lambda_3+1} + (1-\lambda_4)^{k\lambda_3+1}}{\lambda_2^{-k}(k\lambda_3+1)}$$

for $\lambda_1 = 0$.

4.4.2 Relationship to the (β_1,β_2) Plane

The third and fourth standardized moments may be calculated by their definition

$$\sqrt{\beta_1} = \alpha_3 = \mu_3/\mu_2^{3/2}$$

and

$$\beta_2 = \alpha_4 = \mu_4/\mu_2^2$$

As illustrated in Figure 6 this distribution completely covers the (β_1,β_2) plane using only $\lambda_3 \geq 0$. The contour lines relate values of

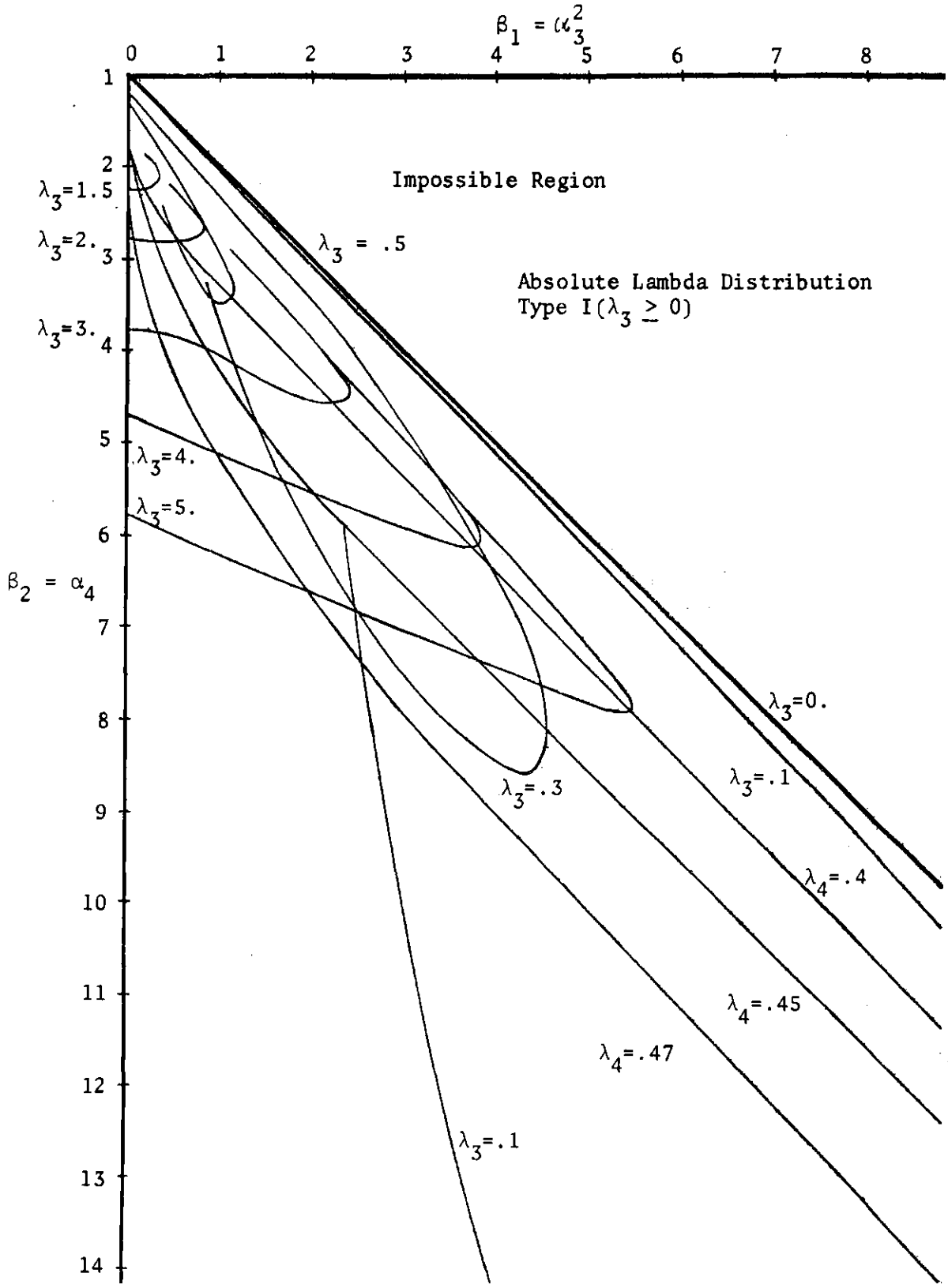

Figure 6. Absolute Lambda Distribution on the (β_1,β_2) Plane $(\beta_3 \geq 0)$

λ_3 and λ_4 to the corresponding moment values. Several properties of this distribution are illustrated in Figure 5.

1. The uniform distribution is a special case, corresponding to $\lambda_3 = 1$ and any $\lambda_4 \in [0,1]$. For this reason the contours collapse on (0,1.8) for all λ_4 as λ_3 approaches 1.
2. The Bernoulli distribution is a special case, corresponding to $\lambda_3 = 0$ and $\lambda_4 = p$. Thus $\lambda_3 = 0$ corresponds to points on the line $\beta_2 = \beta_1 + 1$.
3. The distribution is symmetric if and only if $\lambda_4 = 1/2$.
4. For $\lambda_3 \geq 1$, all points below the line $\beta_2 \approx (5/4)\beta_1 + 1$ can be obtained by adjusting λ_4.
5. For $\lambda_3 < 1$, all points above the line $\beta_2 \approx (5/4)\beta_1 + 1$ and much of the rest of the plane can be obtained by adjusting λ_4.

Thus (β_1,β_2) does not uniquely determine λ_3 and λ_4, but the parameters do uniquely determine the moments. The contours may be used to determine appropriate λ_3 and λ_4 values. More exact values may be found by trial and error or by solving

$$\text{Min}(\beta_1 - \beta_1^*)^2 + (\beta_2 - \beta_2^*)^2$$

using a two-dimensional unconstrained search technique, where β_1^* and β_2^* are the desired moments and β_1 and β_2 are functions of λ_3 and λ_4. Using the approximate graphical solution as a starting point, the objective value can be reduced as close to zero as desired.

4.4.3 The Absolute Lambda Distribution (Type II)

The above discussion has centered on the properties of the

distribution for $\lambda_3 \geq 0$. Figure 7 shows λ_3 and λ_4 contours on the (β_1,β_2) plane for $\lambda_3 \leq 0$. Here there is a one to one correspondence between parameter values and moments. In addition the contours are much straighter. Thus solving for the parameters in terms of the moments is easier for $\lambda_3 \leq 0$.

The results for $\lambda_3 \geq 0$ still apply, except that the range is now $(-\infty, \lambda_1 - \lambda_2\lambda_4^{\lambda_3})$ for $p \leq \lambda_4$ and $(\lambda_1 + \lambda_2(1-\lambda_4)^{\lambda_3},\infty)$ for $p \leq \lambda_4$. Thus the range of X for $\lambda_3 \leq 0$ has a gap corresponding exactly to the range of X for $\lambda_3 \geq 0$. In particular, $F^{-1}(0) = -\infty$, $F^{-1}(\lambda_4^-) = \lambda_1 - \lambda_2\lambda_4^{\lambda_3}$, $F^{-1}(\lambda_4^+) = \lambda_1 + \lambda_2(1-\lambda_4)^{\lambda_3}$, and $F^{-1}(1) = \infty$. Thus the distribution is bimodal. For $\lambda_1 = 0$, $\lambda_2 = 1/-1.57497$, $\lambda_3 = -.224745$, and $\lambda_4 = 1/2$, the first four moments match the normal distribution with zero mean, unit variance, zero skewness, and a kurtosis of three. Figure 8 shows the shape of the distribution for these parameter values. Table 5 gives the quantile values for the same distribution, the normal distribution, and for $\lambda_1 = 0$, $\lambda_2 = 1/.04119$, $\lambda_3 = 2.22474$, and $\lambda_4 = 1/2$. These latter parameter values also have the same first four moments as the normal distribution, but with $\lambda_3 \geq 0$. The similarity of the two λ_3 values appear to be coincidental.

Table 5 indicates that while $\lambda_3 = 2.22474$ may be an adequate approximation of the normal distribution for some applications, $\lambda_3 = -.224745$ is not even similar in shape to the normal distribution. Thus $\lambda_3 \leq 0$ has little application as a tool for approximating common distributions. On the other hand, $\lambda_3 \leq 0$ is excellent for checking the adequacy of the first four moments to measure the difference between the underlying distributions in Chapters V, VI, and VII. If

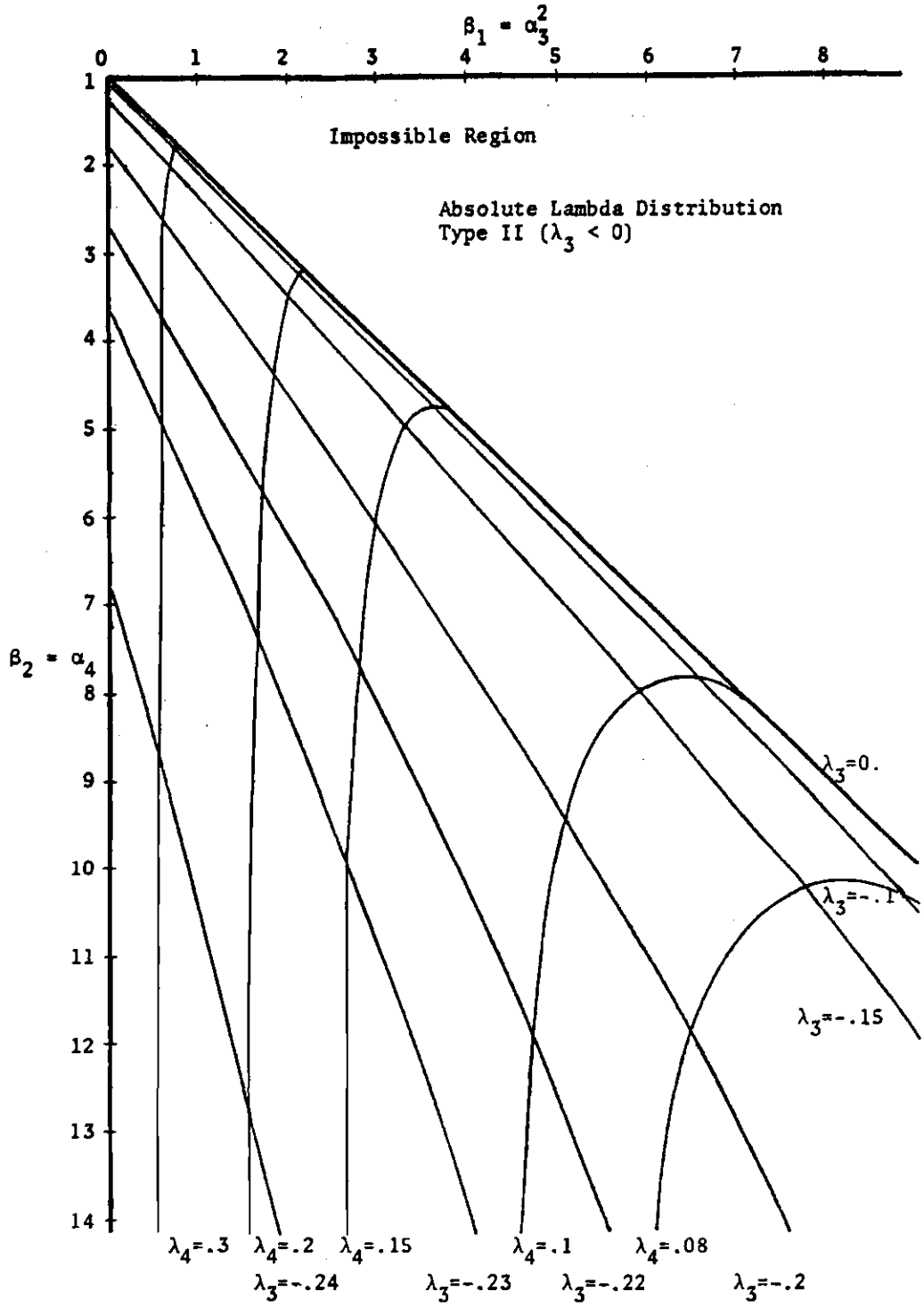

Figure 7. Parameter Contours on the (β_1,β_2) Plane $(\lambda_3 \leq 0)$

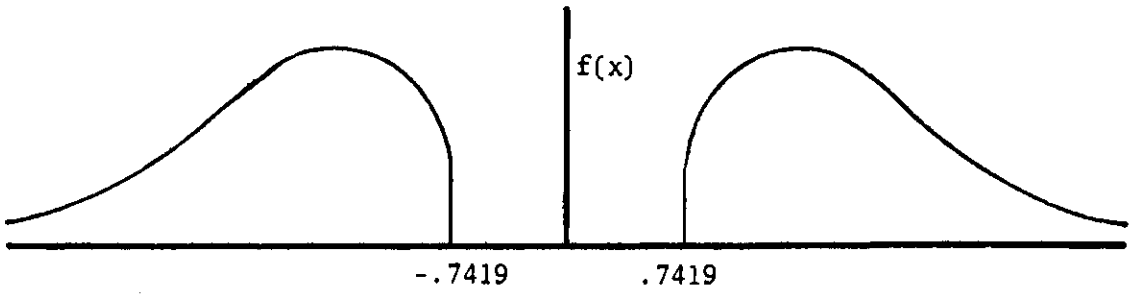

Figure 8. Shape of the ALD (II) Density with Same First Four Moments as the Normal Distribution

essentially the same results are obtained from the normal distribution and from the distribution corresponding to $\lambda_3 = -.224745$, then the first four moments would appear to capture the properties of the underlying distribution important in the current study.

Table 5. Comparison of Quantiles for the Standardized Normal and the Two Corresponding Absolute Lambda Distributions

P	N(0,1)	$\lambda_3 = 2.2247$	$\lambda_3 = -.2247$
.9999	3.75		5.03
.9995	3.29		3.50
.999	3.09		3.00
.995	2.58	2.33	2.09
.99	2.33	2.26	1.79
.975	1.96	2.09	1.45
.95	1.64	1.83	1.24
.9	1.28	1.38	1.06
.8	.84		.91
.7	.52		.83
.6	.26		.78
.55	.13		.76
.5	0	0	[-.7419, .7419]

CHAPTER V

THE RANDOMIZATION TEST

The component randomization test for the paired sample design is the subject of this chapter. Section 5.1 considers the computational problem of testing the hypothesis given a sample of size n, while Section 5.2 considers the problem of determining the power of the test under various conditions using Monte Carlo techniques. The computer program which performs the Monte Carlo analysis based on the results of this section and the techniques of Chapters III and IV is used to compare the power of the randomization test with the power of the corresponding parametric test in Chapter VII.

5.1 Testing the Paired Hypothesis

Despite the good mathematical properties of the component randomization test, other distribution-free tests are more often used due to their computational simplicity, since the computation of the 2^n values $\Sigma_{i=1}^{n} \pm d_i$ is formidable for even small samples. However, the values may be computed in a reasonable amount of time for samples of moderate size (about $n \leq 20$) with an efficient procedure. Such a procedure is discussed in this chapter.

Unfortunately, programs to perform this analysis are rare. None of the common statistical packages (SSP, SAS, Biomed, Univac STATPACK, or SPSS) have such a program. The only technique alluded to in the literature for generating the 2^n combinations of signs is reasonable

for only very small values of n, since it is quite slow and requires $n2^{n-1}$ tabled binary values. Therefore it appears than an efficient algorithm, in terms of both speed and memory, would be valuable. More importantly for the purposes at hand, the logic used in such an algorithm could serve as the nucleus of Monte Carlo analyses involving the paired randomization test.

Section 5.1.1 discusses the logic of the component randomization test in flow chart form. Section 5.1.2 discusses desirable and feasible values of sample size n for consideration. The basic proposed algorithm is developed in Section 5.1.3 and modifications are added in Section 5.1.4

5.1.1 The Component Randomization Test

This section discusses the logic of the component randomization test. Figure 9 is a flow chart of the logic, which at this level is straightforward. The n differences are read, the test statistic $\Sigma_{i=1}^{n} \pm d_i$ is calculated, the number of the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ less than or equal to the test statistic is counted, and the quantile of the test statistic in the randomization sample is p, the count divided by 2^n. If α is the preselected value of the Type I error, then the null hypothesis is rejected or not depending upon the value of α and the form of the alternative hypothesis H_1. In particular if

$$H_1: \mu < 0 \quad \text{then reject } H_0 \text{ if } p < \alpha,$$

if

$$H_1: \mu > 0 \quad \text{then reject } H_0 \text{ if } p > 1-\alpha,$$

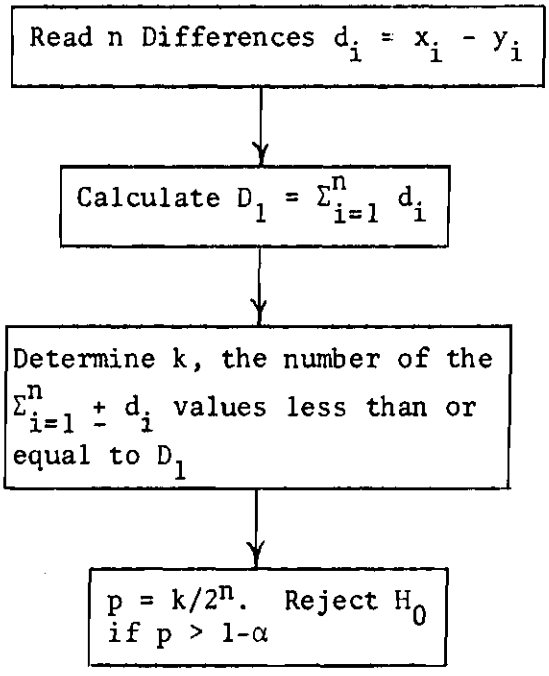

Figure 9. Logic Flow to Apply the Component Randomization Test for the Alternative Hypothesis H_1: $\mu > 0$

and if

H_1: $\mu \neq 0$ then reject H_0 if $p < \alpha/2$ or if $p > 1 - \alpha/2$.

5.1.2 Desirable and Feasible Sample Sizes

The only difficulty in testing the hypothesis of no treatment difference involves the third box in Figure 9: The 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ must be calculated. The manner in which this is performed determines the feasibility of applying the component randomization test, since this operation is the only one requiring nontrivial computer effort. The effort required doubles each time the sample size increases by one, so if a sample of size n requires one cpu second, a sample of size n+6 requires one minute, and a sample of size n+12 requires

about one hour. However, for large n, the central limit theorem applies, allowing the normal theory paired t test to be used. Thus a program capable of performing the randomization test for "small n" is sufficient.

The sufficient value of n is difficult to determine. Hines and Montgomery [1972] state that $n \geq 4$ is probably sufficient for bell-shaped distributions, $n \geq 12$ is probably sufficient for uniformly distributed observations, and $n \geq 100$ may be necessary for U-shaped distributions. However, these guidelines can be misleading in certain instances. For example, very heavy tailed bell-shaped distributions, such as the Cauchy, will not converge for any value of n. Thus the definition of "small n" must be determined with care. With respect to the current interest, a program to perform the randomization test must be able to handle as large an n as possible.

The most straightforward technique of generating the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$, next to explicitly coding 2^n statements, is to create an n by 2^n matrix A containing elements $a_{ji} = +1$ or $a_{ji} = -1$, with each column of A corresponding to one $\Sigma_{i=1}^{n} \pm d_i$ value. Figure 10 illustrates the matrix A for n = 4.

$$A = \begin{bmatrix} -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}$$

Figure 10. Matrix for Calculating $\Sigma_{i=1}^{n} \pm d_i$ for n = 4

Simplification results by noting that the jth and n-j+1th columns of A are exactly the reverse of each other. Therefore only $n2^{n-1}$ values are needed. Such a modification allows the sample size to be one larger for the same amount of effort.

Using the elements of A, the jth value of $\Sigma_{i=1}^{n} \pm d_i$ may be calculated directly as

$$\left[\sum_{i=1}^{n} \pm d_i\right]_j = \sum_{i=1}^{n} a_{ji} d_i$$

In actual implementation, the n multiplications can be eliminated by keying on the value of a_{ji} to branch either to a subtraction statement or to an addition statement. However, while conceptually simple, the matrix A approach is not feasible for sample sizes much larger than five, when the creation of the A matrix becomes a problem. Here the brute force creation by inputting the $n2^n$ values is time consuming and difficult to verify.

A gain in feasible sample size can be made by using some form of binary counting, such as the algorithm developed in the next section, to automatically generate the elements of the A matrix as needed. Core requirements and inputting effort then cease to be problems. Time of generation of the a_{ji} values then becomes the important criterion.

5.1.3 An Algorithm for Testing the Paired Hypothesis

This section develops an algorithm to perform the paired sample component randomization test of location shift. Since the computational feasibility of the test revolves around the generation of the conditional reference distribution (the 2^n equally likely values of the test

statistic), the development focuses on an efficient technique for generating the 2^n combinations of signs.

The algorithm, using modular arithmetic, is based on treating each -1 of the A matrix as a zero and each +1 as a one. The jth column may then be viewed as the binary representation of j-1. For example, column j = 4 is

$$(+1 \ +1 \ -1 \ -1)^T$$

which becomes binary three. Now the binary element in the ith row of the jth column is -1 if $2^{i-1} > \text{Mod}\ (j-1,2^i)$ and is +1 otherwise. This relationship may be seen to be valid by considering specific values of i.

For i = 1, the relationship is

$$-1 \quad \text{if} \quad 1 > \text{Mod}(j-1,2)$$

and

$$+1 \text{ otherwise.}$$

The sign changes for each increment of column j, as desired. For i = 2, the relationship is -1 if 2 Mod(j-1,4) and +1 otherwise, thereby changing the sign every second increment of j. The systematic nature of the sign change as a function of i and of j is similar for higher values of i and j.

An algorithm for generating the jth value of $\Sigma_{i=1}^{n} \pm d_i$ is shown in Figure 11. For each i = 1,2,...,n the relationship is checked and the logic branches to the appropriate subtraction or addition statement.

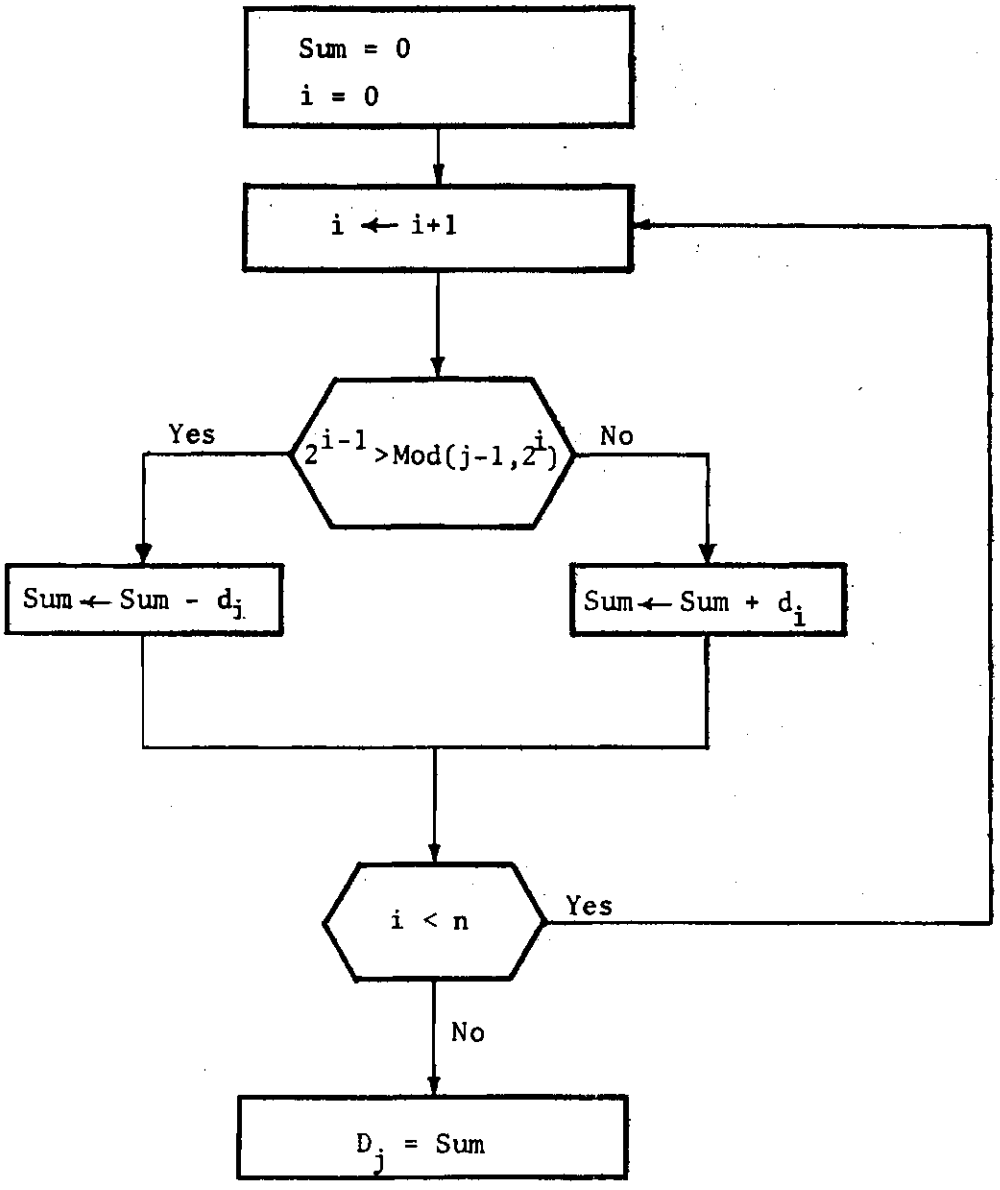

Figure 11. Logic to Calculate the jth of the 2^n $\Sigma_{i=1}^{n} \pm d_i$ Values Using Modular Arithmetic

5.1.4 Modifications to the Algorithm

Using the above technique the a_{ji} values are generated implicitly for any value of n. For $n < 10$ this approach works well. However, the computational effort of calculating the 2n exponentials and the n modular values for each of the 2^n component values becomes noticeably time consuming for larger samples.

This basic technique can be made significantly faster in several ways:

1. The powers of two are calculated using $2^i = 2^{i-1} + 2^{i-1}$ rather than using exponentials. Alternatively, the powers of two are predetermined and placed in a vector ITWO(I) = 2**(I-1) for I = 1,2,...,n+1.

2. The modular arithmetic is unnecessary, since

$$\text{Mod}(j-1,2^{i-1}) = \begin{cases} \text{Mod}(j-1,2^i) & \text{if } 2^{i-1} > \text{Mod}(j-1,2^i) \\ \text{Mod}(j-1,2^i) - 2^{i-1} & \text{otherwise.} \end{cases}$$

This is true since if 2^i goes into j-1 exactly m (integer) times, then 2^{i-1} goes into j-1 either 2m or 2m+1 times. The remainder, the value of the mod operation, remains the same if the quotient is 2m and decreases by 2^{i-1} if the quotient is 2m+1. For example,

$$\text{Mod}(12,2^5) = 12$$

$$\text{Mod}(12,2^4) = 12$$

$$\text{Mod}(12,2^3) = 4$$

$$\text{Mod}(12,2^2) = 0$$

$$\text{Mod}(12,2^1) = 0$$

This relationship is valid as i decreases, but not as i increases, so the logic of the algorithm is reversed to consider row i = n first. The first modular value needed is then

$$\text{Mod}(j-1,2^n) = j-1$$

which is always true since $j = 1,2,\ldots,2^n$. Thus no explicit modular arithmetic is needed.

3. As can be seen by examining Figure 10, the pattern of signs in the A matrix repeats itself 2^k times if only rows $1,2,\ldots,n-k$ are considered. The effort of generating the same pattern 2^k times can be reduced to one generation if the pattern is matched with all combinations of signs from the last k rows when it is first generated. For k = 1, this corresponds to noting the right half of A is the same as the left half, not considering the last row, the result used previously to eliminate need for half of the A matrix. For k = n, this modification corresponds to explicitly coding 2^n statements, one for each value of $\Sigma_{i=1}^{n} \pm d_i$.

Using k = 2 or k = 3 leads to significant savings without much extra coding effort. Consider k = 2. Four lines of code provide the 2^2 values corresponding to $p_j = (\Sigma_{i=1}^{n-2} \pm d_i)_j$. In particular, using FORTRAN-like equations,

$$p_j = p_j + d_{n-1} + d_n$$

$$p_j = p_j + d_{n-1} - d_n$$

$$p_j = p_j - d_{n-1} + d_n$$

and

$$p_j = p_j - d_{n-1} - d_n$$

This procedure involves the calculation of 2^{n-k} columns, each using only rows i = 1,2,...,n-k.

The value of k determines the tradeoff between coding simplicity and execution time as may be seen by examining Figure 12. Figure 12 is the FORTRAN code of the interactive program to perform the test of hypothesis given n differences. The logic follows that of Figure 9 exactly, implementing the techniques discussed here. Eight explicit lines of code, corresponding to k = 3 are used. Definition of variables include IEVEN = 2^i, IEVEN2 = 2^{i-1}, and ITERM = Mod(j-1,2^i).

Other techniques could doubtless be applied. An important method, for testing the hypothesis, is to apply heuristic rules to find sets of $\Sigma_{i=1}^{n} \pm d_i$ which are less than, or greater than the test statistic. For example, if $\Sigma_{i=1}^{n-j} \pm d_i + \Sigma_{i=n-j+1}^{n} |d_i| < \Sigma_{i=1}^{n} |d_i|$, then the counter may be incremented by 2^j, corresponding to the 2^j values of $\Sigma_{i=n-j+1}^{n} \pm d_i$. However, for a Monte Carlo study of the randomization sample, all 2^n values are needed explicitly, so techniques of grouping are not pursued further.

Figure 13 may shed some light on the quest for speed in a program performing 2^n operations. Using semi-logarithmic paper, the cpu seconds required to test an hypothesis vs. the sample size n is plotted. The times are for Univac 1108. Three plots are shown corresponding to

1. Explicit calculation of the modular values
2. Implicit calculation of the modular values and with k = 1

```
C
C        TO PERFORM THE COMPONENT RANDOMIZATION TEST OF HYPOTHESIS
C       ON N GIVEN DIFFERENCES.
      DIMENSION D(100), ITWO(100)
      DATA D(2)/0./,D(3)/0./,ITWO(3)/0/
C
C*****ENTER DATA, INITIALIZE, AND CALCULATE TEST STATISTIC
C
      WRITE (6,1)
    1 FORMAT (' ENTER SAMPLE SIZE N AND N DIFFERENCES.')
      READ (5,2) N, (D(I),I=1,N)
    2 FORMAT ()
      KOUNT = 0
      D22 = D(2) + D(2)
      D32 = D(3) + D(3)
      D23 = D22 + D32
      DSUM = -D(1) - D(2) - D(3)
      TWON = 2**N
      STAT = 0
      DO 100 I=1,N
      ITWO(I) = 2**(N-I)
  100 STAT = STAT + D(I)
      M = ITWO(3)
      IF (N .GT. 3)  GO TO 200
      J = TWON
      SUM = DSUM
      GO TO (10, 20, 30), N
  200 J = 0
C
C*****DETERMINE THE J TH COMBINATION OF SIGNS FOR COMPONENTS I=4,...,N
C
 1000 J = J + 1
      SUM = DSUM
      IEVEN2 = M
      ITERM = J
      DO 400 I=4,N
      IEVEN = IEVEN2
      IEVEN2 = ITWO(I)
      IF (IEVEN .LT. ITERM)  ITERM = ITERM - IEVEN
      IF (ITERM .GT. IEVEN2)  GO TO 300
      SUM = SUM - D(I)
      GO TO 400
  300 SUM = SUM + D(I)
  400 CONTINUE
C
C*****FOR THIS COMBINATION OF SIGNS, CONSIDER ALL 8 POSSIBILITIES
C
   30 X = SUM + D32
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
      X = SUM + D23
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
   20 X = SUM + D22
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
   10 IF (SUM .LE. STAT)  KOUNT = KOUNT + 1
      IF (-SUM .LE. STAT)  KOUNT = KOUNT + 1
      IF (J .LT. M)  GO TO 1000
C
C*****PRINT RESULTS
C
      RATIO = KOUNT / TWON
      WRITE (6,3) KOUNT, TWON, RATIO
    3 FORMAT (' KOUNT =', I15, / ' 2**N =', F17.0/' RATIO =', F10.7)
      STOP
      END
```

Figure 12. Program to Test the One-Sided Paired Sample Location Hypothesis Using Component Randomization

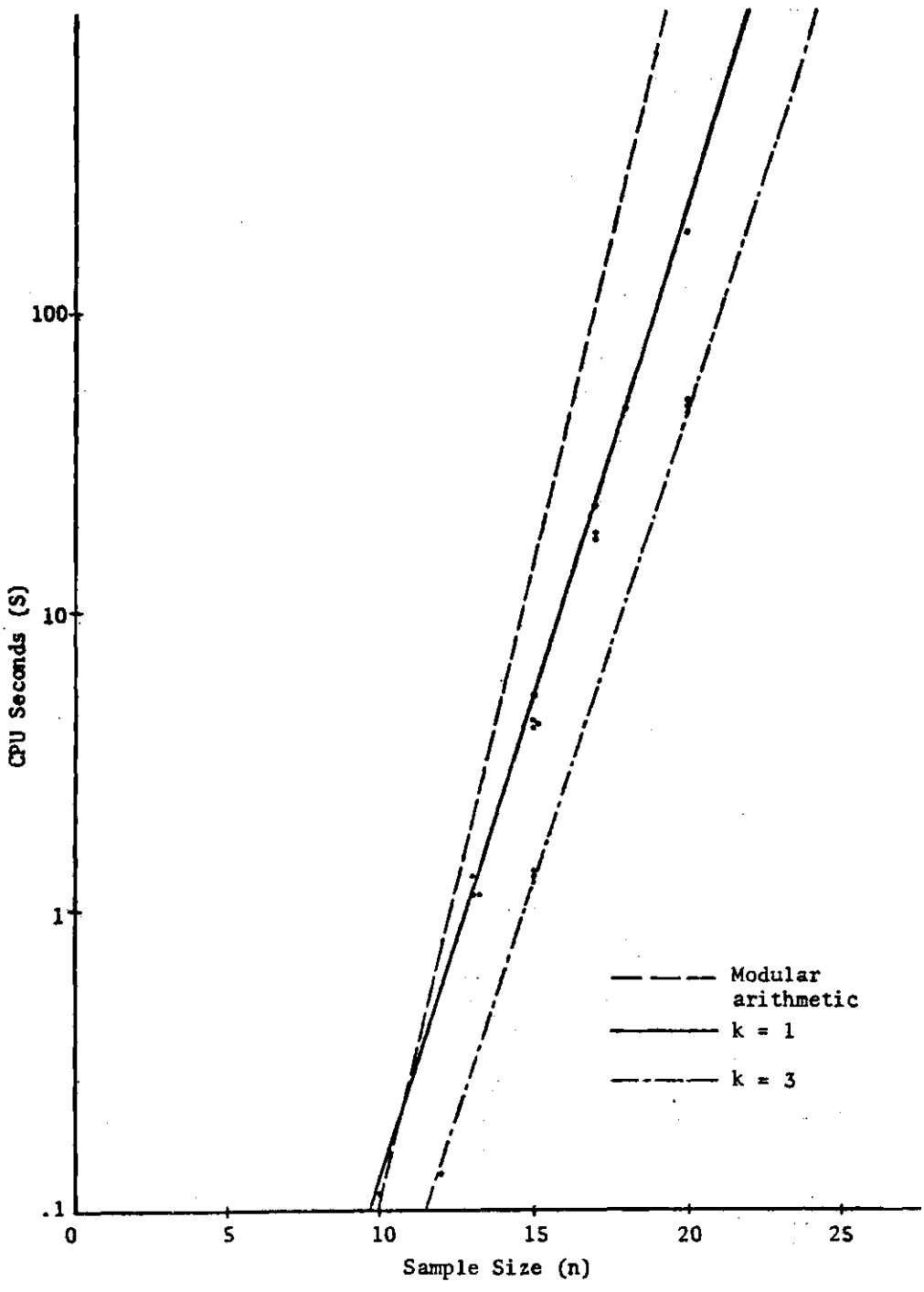

Figure 13. CPU Seconds to Perform the Component Randomization Test as a Function of Sample Size

corresponding to looping 2^{n-1} times, and

3. Implicit calculation of the modular values and with k = 2 corresponding to looping 2^{n-2} times.

From Figure 13 it is seen that n = 10 is a fairly trivial problem, requiring only .1 second even when explicitly calculating all modular values. Recall that n = 10 is quite a formidable problem using an explicit A matrix due to storage and inputting requirements. For $n \geq 15$ the comparison of times between the three variations of the algorithm is clear. For any n the ratio of the required time remains constant. Using the explicit calculation of all modular values as a standard, the substitution of implicit calculations with k still equal to one requires only one-third the time. Increasing k to three requires only 9% of the time. Larger values of k can obviously reduce the time still further, but the program quickly becomes long due to the necessity of explicitly considering 2^k values.

5.2 Determination of the Power of the Component Randomization Test

It is well known that the asymptotic relative efficiency of the component randomization test is one for the paired sample case. That is, for large n, the component randomization test rejects the null hypothesis with the same probability as the corresponding parametric test.

For finite sample sizes, however, a price is paid in terms of power for using the distribution-free randomization test. Kempthorne and Doerfloer [1969] show that the loss is less than in using the rank permutation or the sign permutation test. However, the exact power

loss for the randomization test is not known. The remainder of this chapter, and the next two chapters, considers the problem of determining the power loss for various sample sizes, underlying distributions, and alternative hypotheses. In this section the power of the randomization test is discussed, Chapter VI discusses the power of the corresponding parametric test, and Chapter VII brings together the techniques of Chapters V and VI to compare the power of the two tests.

In parametric tests the power may be determined by considering the distribution of the test statistic under the alternative hypothesis and evaluating the probability of the test statistic being in the critical region under H_1. But since the critical region for the component randomization test is not defined in terms of the distribution of the test statistic, the power must be evaluated differently. Section 5.2.1 is a general discussion of the power of the randomization test. Section 5.2.2 describes the Monte Carlo approach used to study the power in Chapter VII.

5.2.1 The Power of the Component Randomization Test

The power of a test is the probability of rejecting the null hypothesis when the alternative hypothesis is true, or equivalently, one minus the Type II error β, the probability of not rejecting H_0 when H_0 is false. In the case of the component randomization test for H_1: $\mu > 0$

$$\text{Power} = 1 - \beta$$

$$= \text{Prob} \left\{ \sum_{i=1}^{n} d_i > [(1-\alpha)2^n] \text{ of the } 2^n \sum_{i=1}^{n} \pm d_i \text{ values} \right\}$$

However, the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ are not independent (see Section 6.3.3), making difficult the calculation of the power in terms of this probability.

Some insight is gained by considering the power for small sample sizes (n = 1,2), but general results do not easily follow. Consider n = 1 and H_1: $\mu > 0$. Then $\Sigma_{i=1}^{n} d_i = d_1$ and the other of the $2^n = 2$ components is $-d_1$. The power is

$$\text{Power} = 1 - \beta$$

$$= \text{Prob}\{d_1 > [(1-\alpha)2^1] \text{ of the } 2^1 \sum_{i=1}^{n} \pm d_i \text{ values}\}$$

$$= \text{Prob}\{d_1 > d_1 \text{ or } d_1 > -d_1 \text{ or both}\} \text{ for any } \alpha \geq .5\}$$

$$= \text{Prob}\{d_1 > 0\}$$

The power is therefore simply the probability that the observed value d_1 is positive, a direct function of the underlying distribution.

Consider n = 2 and the one sided alternative hypothesis. The power is

$$\text{Power} = 1 - \beta$$

$$= \text{Prob}\{d_1 + d_2 > [(1-\alpha)2^2] \text{ of the } 2^2 \sum_{i=1}^{2} \pm d_i \text{ values}\}$$

$$= \text{Prob}\{d_1 + d_2 < -d_1 + d_2 \text{ or } d_1 + d_2 < d_1 - d_2 \text{ or}$$

$$d_1 + d_2 < -d_1 - d_2 \text{ or combinations}\} \text{ if } \alpha \leq .25$$

Since the sums are dependent upon each other, this probability is difficult to evaluate, even if the distribution of the differences

is known. For larger values of n, the direct determination of the power is not feasible because the number of dependent events that might occur increases exponentially.

5.2.2 The Monte Carlo Approach

Since closed form results are not attainable, the Monte Carlo approach seems appropriate to determine the power of the component randomization test. A discussion of the methodology is given in this section. The general logic is discussed in Section 5.2.2.1, the variance of the power estimate is given in Section 5.2.2.2 and a method of determining the number of samples N necessary for a given variance of the estimate is given in Section 5.2.2.3.

5.2.2.1 Program Logic. Given a sample size n, the underlying distribution, and the value of $E\{d\} = E\{d_i\}$ $i = 1,2,\ldots,n$, the Monte Carlo method can be used to determine the power of the component randomization test. Figure 14 is a flow chart of the necessary logic for H_1: $\mu > 0$.

The logic consists of two loops. The first loop, repeated N times, fills a histogram with the position j, $j = 1,2,\ldots,2^n$, of $\Sigma_{i=1}^{n} d_i$ by incrementing a counter c_j by one. The second loop then analyzes the histogram, for each given value of α, to calculate the estimate of the power, $1-\beta$, and the variance of this estimate. The formulae used are

$$\widehat{1 - \beta} = 1 - \sum_{j=1}^{[(1-\alpha)2^n]} c_j/N$$

and

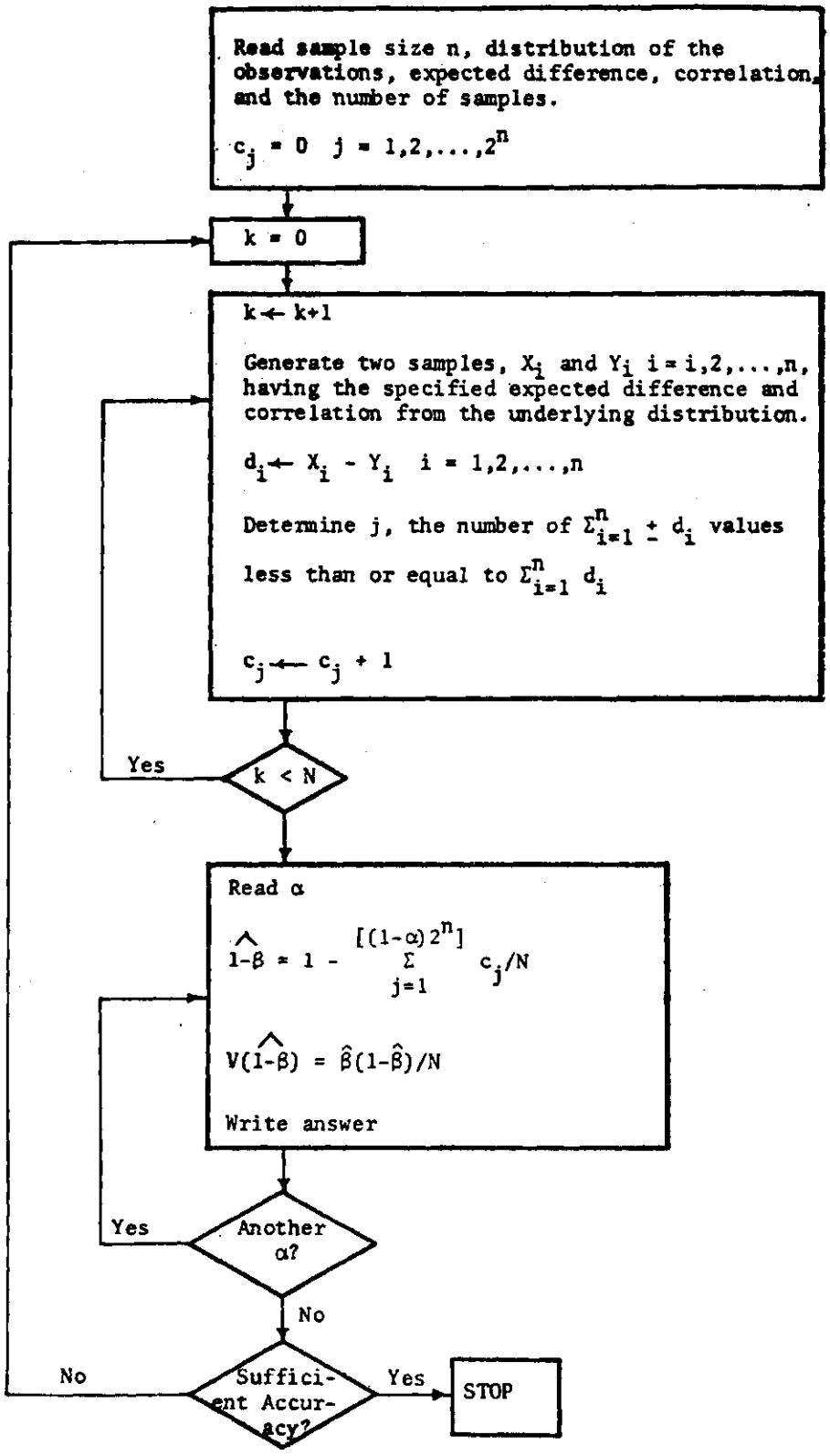

Figure 14. Monte Carlo Determination of Power for Component Randomization Test

$$V\{\widehat{1 - \beta}\} = V\{\hat{\beta}\} = \beta(1-\beta)/N$$

The formula for the power estimate is simply the number of times the test rejects H_0 divided by the number of tests performed, N. The formula for the variance of the estimate is derived in Section 5.2.2.2.

The logic assumes H_1: $\mu > 0$, but if H_1: $\mu < 0$ is the desired alternative hypothesis, the estimate of the power is simply one minus the estimate of the power provided by the program when the specified value of α is one minus the type I error. The variance of the estimate is not changed. For either alternative hypothesis, the process is repeated until the variance of the estimate is sufficiently small. A program to implement this logic is given in Appendix B, including a description of the required input.

5.2.2.2 Variance of the Estimate of the Power. It is important to have an estimate of the variance of the estimate of the power to determine the accuracy of the results. While replicating the analysis several times and averaging the power estimate of each replication may be used to estimate the variance and to calculate confidence intervals based on the central limit theorem, a simpler approach is to determine the variance analytically. The derivation of the variance is the subject of this section.

As discussed earlier, the obvious estimate of the power is

$$\widehat{1 - \beta} = 1 - \sum_{j=1}^{M} c_j/N$$

where $M = [(1-\alpha)2^n]$. The variance of this estimate is $(1-\beta)\beta/N$. This may be seen directly as follows:

$$\widehat{V\{1-\beta\}} = V\{\hat{\beta}\}$$

$$= V\{\sum_{j=1}^{M} c_j/N\}$$

$$= \frac{1}{N^2}\left[\sum_{j=1}^{M} V(c_j) + \sum_{\substack{i=1 \\ i\neq j}}^{M}\sum_{j=1}^{M} \mathrm{Cov}\{c_i, c_j\}\right]$$

$$= \frac{1}{N^2}\left[\sum_{j=1}^{M} Np_j(1-p_j) + \sum_{\substack{i=1 \\ i\neq j}}^{M}\sum_{j=1}^{M} (-Np_ip_j)\right]$$

where p_j is the probability of $\Sigma_{i=1}^{n} d_i$ being less than or equal to j of the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$, since the c_j's are multinomial random variables. Continuing

$$\widehat{V\{1-\beta\}} = \frac{1}{N}\left[\left(\sum_{j=1}^{M} p_j - \sum_{j=1}^{M} p_j^2\right) - \sum_{\substack{i=1 \\ i\neq j}}^{M}\sum_{j=1}^{M} p_ip_j\right]$$

$$= \frac{1}{N}\left[\sum_{j=1}^{M} p_j - \sum_{i=1}^{M}\sum_{j=1}^{M} p_ip_j\right]$$

$$= \frac{1}{N}\left[\left(1 - \sum_{j=1}^{M} p_j\right)\sum_{j=1}^{M} p_j\right]$$

$$= (1-\beta)\beta/N$$

and the result is shown. An estimate of the variance is then $(1-\hat{\beta})\hat{\beta}/N$. Note that under H_0, $V\{1-\hat{\beta}\} = \alpha(1-\alpha)/N$, and therefore the variance is known exactly.

5.2.2.3 The Required Number of Samples. In performing Monte Carlo studies it is convenient to be able to estimate, or place a bound, on the length of run necessary to obtain results of a specified accuracy. An upper bound for N, the number of samples, is derived in this section.

In general, β decreases as H_1 moves away from H_0. But if β decreases, $V\{\widehat{1-\beta}\} = \beta(1-\beta)/N$ decreases, if $\alpha \leq .5$. Therefore the maximum variability occurs under H_0 when $V\{\widehat{1-\beta}\} = \alpha(1-\alpha)/N$ as shown in the previous section. The variance for all H_1 may then be bounded by setting N such that

$$\alpha(1-\alpha)/N < k$$

or

$$N > \alpha(1-\alpha)/k$$

where k is a specified required variance of the estimate. If several α values are of interest, the largest value ($\alpha \leq .5$) may be used to determine N. The true probability of type I error, rather than the nominal alpha value, should be used. Misleading results may be obtained if $\alpha \neq m/2^n$ for some $m = 1,2,\ldots,2^n-1$, since the discrete reference distribution allows only these discrete α values. For example, α must be greater than 2^{-n} since rejecting H_0 only when $\Sigma_{i=1}^{n} d_i$ is the largest of the 2^n components $\Sigma_{i=1}^{n} \pm d_i$ is the extreme possibility.

As an example, if $\alpha = .05$ and $\alpha = .01$ are to be considered, and $V\{\widehat{1-\beta}\} < .0001 = k$ is desired, then

$$N > .05(.95)/.0001 = .0475/.0001 = 475$$

Variance reduction techniques may reduce the necessary value of N, although the above calculations still provide an upper bound.

CHAPTER VI

THE DISTRIBUTION OF $\Sigma_{i=1}^{n} d_i$

Having developed the methodology to determine the power of the component randomization test in Chapter V, the power of the corresponding parametric test is now considered. This test is performed by comparing the statistic $\Sigma_{i=1}^{n} d_i$ to its distribution under the null hypothesis. While such a test is conceptually straightforward and has more power than the randomization test, the sum of differences distribution (sdd) is not often known. An important special case where the sdd is known, of course, is normally distributed observations resulting in the sdd being normal. But if the underlying distribution is not normal, the parametric test incurs the combined problem of (1) uncertainty of the underlying distribution, and (2) if the underlying distribution is known, having to determine the corresponding sdd.

The analytical determination of the sdd from the underlying distribution is well known for only a few distributions other than the normal. If the differences are Cauchy, the sdd is Cauchy. If the differences are uniform, the sdd is known, but complicated. Differences with identical gamma distributions result in a gamma sdd, but since the differences are assumed symmetric, the Cauchy and uniform are the only common applicable results.

While there are doubtless other examples among less used distributions, it is obvious that even if the underlying distribution is known, application of the parametric test may be difficult. This

difficulty is often overcome by using the normal theory test, which is asymptotically valid as n becomes large by the central limit theorem, as discussed in Chapter V.

A general methodology for determining the sdd given a sample of size n and an underlying distribution would be valuable for (1) use of the parametric test when the underlying distribution is known, and (2) comparing the power of the randomization test (as discussed in Chapter V) to the power of the parametric test.

The power of the parametric test may be determined analytically if the sdd is known, since

$$\text{Power} = \text{Prob}\left\{\sum_{i=1}^{n} d_i \geq c \,|\, H_1: \; E(d_i) = \mu\right\}$$

$$= \text{Prob}\left\{\sum_{i=1}^{n} d_i + n\mu \geq c \,|\, H_0: \; E(d_i) = 0\right\}$$

$$= \text{Prob}\left\{\sum_{i=1}^{n} d_i \geq c - n\mu \,|\, H_0\right\}$$

$$= 1 - F(c - n\mu)$$

where F is the cdf of the sdd and c is the critical value.

This chapter develops a methodology of determining the sdd for an arbitrary underlying distribution and sample size. An approach, which is implemented via a computer program, is developed to determine any desired quantiles of the sdd, the form of the distribution needed for hypothesis testing. Section 6.1 discusses the moments of the sdd as functions of the moments of the differences, and the moments of the differences as functions of the underlying distribution. Section 6.2

discusses some well-known analytic, but approximate, methods of determining the sdd, including asymptotic expansions and Chebyshev type bounds on quantiles. The estimation of the sdd moments and quantiles using dependent, rather than independent, observations is developed in Section 6.3. The use of dependent observations as a variance reduction technique makes feasible the use of Monte Carlo techniques for determining the sdd. The Monte Carlo approach is developed in Section 6.4.

6.1 Moments of the sdd

In any analysis, closed form results are more desirable than Monte Carlo results, due to the generality of the analytical results and the cost in time and money of the Monte Carlo approach. In the context of determining the quantiles of the sdd, this implies that a closed form expression of the cdf, the inverse cdf, and/or the density function is desired.

As described in Chapter IV, the moments, especially the mean, variance, standardized third moment, and standardized fourth moment, are effective in describing a distribution. Of course, the more moments known, theoretically the more known about the shape and quantiles of the distribution. The "problem of moments" is the determination of the cdf, inverse cdf, and/or density function in terms of the moments. Since this problem has been studied extensively, it seems reasonable to study the moments of the sdd even though quantiles are the ultimate goal.

The moments of the sdd are now discussed. Section 6.1.1 derives the standardized moments of the differences as functions of the underlying distribution moments. The standardized moments of the sdd are

then given in terms of the moments of the differences in Section 6.1.2.

6.1.1 Moments of the Differences

In this section the higher order moments of the differences d_i are derived as functions of the moments of the observations X_i and Y_i, where $d_i = X_i - Y_i$. This relationship is important because:

1. an experimenter sometimes has a better idea of the distribution of the observations than of the differences,
2. the differencing operation results in d_i being more normally distributed than either X_i or Y_i, and
3. the results are needed later in this chapter in the development of the Monte Carlo approach of determining the sdd.

Consider $d = X - Y$ where the subscript i has been suppressed. Now under the null hypothesis and assumptions of the randomization test,

$$E\{X\} = E\{Y\}$$

and either higher order moments of X and Y are identical or both X and Y have all zero odd moments. Let $\sigma_x^2, \alpha_{3x}, \alpha_{4x}, \ldots$ denote the variance and higher order standardized moments of X and of Y and let $\sigma^2, \alpha_3, \alpha_4, \ldots$ denote the same for the differences. Then

$$E\{d\} = E\{X\} - E\{Y\} = 0$$

and

$$\sigma^2 = V\{X-Y\} = V\{X\} + V\{Y\} + 2\,Cov\{X,Y\} = 2\sigma_x^2$$

The third and all higher odd moments are zero since d is symmetric about

zero, as discussed in Section 5.1.1.1.

If X and Y are identically distributed, the higher order even moments of d are

$$\alpha_4 = E\{d^4/\sigma^4\} \tag{4}$$

$$= E\{(X-Y)^4/(2\sigma_x^2)^2\}$$

$$= E\{(X^4 - 4X^3Y + 6X^2Y^2 - 4XY^3 + Y^4)/4\sigma_x^4\}$$

$$= \frac{1}{4}\,[\alpha_{4x} - 0 + 6(1)(1) - 0 + \alpha_{4x}]$$

$$= \frac{1}{2}\,[\alpha_{4x} + 3]$$

$$\alpha_6 = E\{d^6/\sigma^6\}$$

$$= E\{(X-Y)^6/(2\sigma_x^2)^3\}$$

$$= E\{(X^6 - 6X^5Y + 15X^4Y^2 - 20X^3Y^3 + 15X^2Y^4 - 6XY^5 + Y^6)/8\sigma_x^6\}$$

$$= \frac{1}{8}\,[\alpha_{6x} - 0 + 15\alpha_{4x}(1) - 20\alpha_{3x}\alpha_{3x} + 15\alpha_{4x}(1) - 0 + \alpha_{6x}]$$

$$= \frac{1}{4}\,[\alpha_{6x} + 15\alpha_{4x} - 10\alpha_{3x}^2]$$

and

$$\alpha_8 = E\{(X-Y)^8/\sigma^8\}$$

$$= \frac{1}{8}\,[\alpha_{8x} + 28\alpha_{6x} - 56\alpha_{3x}\alpha_{5x} + 35\alpha_{4x}^2]$$

An important point is that the differencing operation tends to normalize the differences; that is, the differences are more nearly normal than X and Y. For any normal distribution $\alpha_4 = 3$, $\alpha_6 = 15$,

$\alpha_8 = 105$ and all odd standardized moments are zero due to symmetry. The difference being symmetric, even if X and Y are skewed, certainly is a factor in making the differences more nearly normal.

But notice also the even moments. Clearly from equation 4 neither $3 \leq \alpha_4 \leq \alpha_{4x}$ or $\alpha_{4x} \leq \alpha_4 \leq 3$. Similarly, if $\alpha_4 = 3$ and $\alpha_3 = 0$ as for the normal, then either $15 \leq \alpha_6 \leq \alpha_{6x}$ or $\alpha_{6x} \leq \alpha_6 \leq 15$. Of course, $\alpha_i = \alpha_{ix}$ for $i = 3,4,\ldots$ if the α_{ix} values are those of the normal distribution.

6.1.2 Moments of the sdd

The moments of the sdd as functions of the moments of the differences are discussed in this section. Knowledge of this functional relationship is important since it appears to be the only closed form link between the underlying distribution and the ssd. These results are used for validation of the Monte Carlo approach developed later in this chapter.

Cumulants are an important tool when working with sums of random variables. The standardized moments of any sum may be found indirectly as a function of the underlying standardized moments by transforming the underlying moments to cumulants, using the relationship

$$\kappa_r \left(\sum_{j=1}^{n} X_j \right) = \sum_{j=1}^{n} \kappa_r (X_j) \quad \text{for all } r$$

where κ_r is the rth cumulant, and transforming back to moments using the relationships

$$\mu_1' = \kappa_1$$

$$\mu_2 = \kappa_2$$

$$\mu_3 = \kappa_3$$
$$\mu_4 = \kappa_4 + 3\kappa_2^2$$

and

$$\alpha_i = \mu_i/\mu_2^{i/2} \quad \text{for } i = 2,3,\ldots$$

A complete discussion of cumulants, including higher order transformation relationships, may be found in Kendall and Stuart [1969, Chapter 3].

Simplication of analyses results from being able to relate the underlying standardized moments directly to the sdd moments. Letting $\bar{\alpha}_i$ be the ith standardized moment of the sdd, Burr [1974] shows

$$\bar{\alpha}_3 = \alpha_3/\sqrt{n}$$

and

$$\bar{\alpha}_4 = (\alpha_3 - 3)/n + 3$$

but the relationships for higher order moments do not seem to appear in the literature. Since these higher order moments are needed for validation of the Monte Carlo approach, they are derived directly using straightforward, though tedious, algebra. The results are

$$\bar{\alpha}_5 = (\alpha_5 - 10\alpha_3)/n^{3/2} + 10\alpha_3/\sqrt{n}$$

$$\bar{\alpha}_6 = \frac{\alpha_6 - 15\alpha_4 - 10\alpha_3^2 + 30}{n^2} + \frac{15\alpha_4 + 10\alpha_3^2 - 45}{n} + 15$$

$$\bar{\alpha}_7 = \frac{\alpha_7 - 21\alpha_5 - 35\alpha_4\alpha_3 + 210\alpha_3}{n^{5/2}} + \frac{21\alpha_5 + 35\alpha_4\alpha_3 - 415\alpha_3}{n^{3/2}} + \frac{105\alpha_3}{n^{1/2}}$$

and

$$\bar{\alpha}_8 = (\alpha_8 - 28\alpha_6 - 56\alpha_3\alpha_5 - 35\alpha_4^2 + 420\alpha_4 + 560\alpha_3^2 - 630)/n^3$$
$$+ (28\alpha_6 + 56\alpha_3\alpha_5 + 35\alpha_4^2 - 630\alpha_4 - 840\alpha_3^2 + 1155)/n^2$$
$$+ (210\alpha_4 + 280\alpha_3^2 - 630)/n + 105$$

These results may be summarized by

$$\bar{\alpha}_p = n^{-p/2} \sum_{\ell=1}^{[p/2]} \left[\sum_{\substack{\text{all } \lambda_i \\ \sum_{i=1}^{\ell} \lambda_i = p \\ \lambda_i \geq 1}} \left\{ \left(\prod_{i=1}^{\ell} \alpha_{\lambda_i} \right) \binom{p}{\lambda_1 \lambda_2 \ldots \lambda_\ell} \binom{n}{n_1 n_2 \ldots n_p} \right\} \right]$$

where n_k is the number of times α_i appears in $\alpha_{\lambda_1}\ \alpha_{\lambda_2} \ldots \alpha_{\lambda_\ell}$. The general result eliminates the need for lengthy algebra which results from the complexity of expanding $(\sum_{i=1}^{n} d_i)^p$.

6.2 Quantiles of Distributions Having Known Moments

This section considers the possibility of using the moments of the sdd, developed in Section 6.1, to directly determine the quantiles of the sdd. It would be convenient if a closed form function g could be found such that

$$F^{-1}(p) = g(p, \mu, \sigma^2, \alpha_3, \alpha_4, \ldots)$$

where $\mu, \sigma^2, \alpha_3, \alpha_4, \ldots$ are the mean, variance, and standardized moments

of the sdd, and F is the cdf of the sdd.

Unfortunately some common distributions, such as the log normal, are not uniquely determined by their moments. Kendall and Stuart [1969] discuss this topic in some detail, including necessary conditions for uniqueness. Any distribution having finite range is uniquely characterized by its moments. As a rule of thumb, sets of moments less than or equal to the moments of the normal distribution uniquely determine a distribution. Due to the possibility of a given set of moments mapping onto several distributions, the general closed form function g described above does not exist.

While a set of moments may not specify a unique distribution, it does provide quite a lot of information about shape, scale, and location. Therefore it is reasonable to approximate quantiles using knowledge of some or all of the moments. Commonly only the first m moments are used. Two types of methods are commonly used to approximate $F^{-1}(p)$ given some or all of the moments: (1) asymptotic approximation, and (2) Chebyshev-type inequalities. Asymptotic approximations are discussed in Section 6.2.1 and inequalities are discussed in Section 6.2.2.

6.2.1 Asymptotic Approximations to Distributions

Let $F_n(x)$ be the cdf of $\Sigma_{i=1}^{n} d_i$. Then an asymptotic approximation of $F_n(x)$ is

$$F_n(x) \approx \sum_{i=0}^{r} A_i(x)/n^{i/2}$$

where the $A_i(x)$ terms are functions of moments or cumulants, the specific form of which arises from the specific expansion used. The relationship can be inverted to find $F_n^{-1}(p)$, again depending upon the specific

technique. The errors satisfy the condition

$$\left|F_n(x) - \sum_{i=0}^{r} A_i(x)/n^{i/2}\right| \leq C_r(x)/n^{(r+1)/2}$$

where the bound $C_r(x)$ may or may not depend upon x.

The use of asymptotic approximations to determine the sdd, or any distribution, has two disadvantages: (1) possible non-convergence and (2) no measure of error. Quoting from Wallace [1958] on convergence;

> For any fixed n, the infinite series may be convergent, but in statistical application usually is not. . . . though the addition of the next term will for sufficiently large n improve the approximation, for any prescribed n it may not do so. Typically the bounds $C_r(x)$ increase rapidly with r, and for small n only the first few terms are improvements.

Wallace also states that "only the order of magnitude of the errors is known and only rarely are explicit bounds known, and these are far from sharp."

To illustrate these problems consider the special case of Cornish-Fisher expansions, where an infinite series gives $F^{-1}(p)$ as a function of the pth quantile of the normal distribution U_p. Cornish and Fisher [1937, 1960] give detailed formulae through the eighth cumulant. The first few terms are

$$F^{-1}(p) = U_p + (1/6)(U_p^2-1)\kappa_3 + (1/24)(U_p^3-3U_p)\kappa_4$$

$$+ (1/36)(2U_p^3-5U_p)\kappa_3^2 + \ldots$$

where κ_i is the ith cumulant of the distribution. Kendall and Stuart [1969] and Johnson and Kotz [1970a] discuss this expansion in more detail.

The problem of determining the .95 quantile of the average n random variables $X_1, X_2, \ldots, X_n$ is considered, where each X_i is exponentially distributed with mean $1/\lambda$. Although the differences of the randomization test are symmetric by assumption, the exponential distribution is used here since the average (or sum of the X_i/n) has a known distribution. Table 6 shows the .95 quantiles of the distribution of $\Sigma_{i=1}^{n} X_i/n \equiv \overline{X}_n$ and the Cornish-Fisher expansion for various values of n.

Table 6. The Cornish-Fisher Expansion Approximation of the .95 Quantile of a Sum of Exponentially Distributed Random Variables

n	Actual	Cornish-Fisher
	$(\chi^2(2n)/2n)$	(29 terms--eight moments)
1	3.00	wild fluctuation
2	2.37	2.1-2.3
3	2.10	1.97
4	1.94	1.84
5	1.83	1.74
7	1.69	1.63
10	1.57	1.52
15	1.46	1.43
25	1.35	1.33
50	1.24	1.23

The values in the "actual" column are the .95 quantiles of the χ^2 distribution divided by $2n\lambda$ where $\lambda = 1$. That this is correct is seen as follows:

$$\sum_{i=1}^{n} X_i \sim \Gamma(n,\lambda)$$

$$\Rightarrow \quad \bar{X}_n \sim \Gamma(n,n\lambda)$$

$$\Rightarrow \quad 2n\lambda\bar{X}_n \sim \Gamma(n,.5)$$

$$\Rightarrow \quad \bar{X}_n \sim \chi^2(2n)/2n\lambda$$

The value in the "Cornish-Fisher" column for each n is the sum of the first 29 terms of the Cornish-Fisher expansion, which involves the first eight cumulants (moments) of the exponential distribution. Note that while the approximation is progressively better as n increases, for small n the results are not accurate. In fact, for n = 1 each term added changes the result drastically. The problems arising here are general to asymptotic expansions, as noted by Wallace. In addition, there is no theory available to indicate the error of the approximation.

6.2.2 Chebyshev-type Inequalities

Bounds may be placed on $F^{-1}(p)$ when the first m moments are known. A well-known example, corresponding to m = 2, is Chebyshev's inequality. If μ and σ^2 are finite and k is a positive number, then

$$\text{Prob}\{|X-\mu| \geq k\} \leq 1/k^2$$

whether X is discrete or continuous.

Royden [1953] discusses the general case of the first m moments given, citing the results of Markoff [1884] and Stieltjes [1884]. The technique given is in terms of finding bounds on the cdf, but by trial and error the technique could be inverted to place bounds on $F^{-1}(p)$.

The trial and error procedure is cumbersome, however, since the technique of bounding the cdf involves finding the zeroes of a 2n degree polynomial.

If the bounds obtained are inadequate for a given purpose, more moments may be used. However, the convergence of the bounds may be quite slow for the same reasons the asymptotic approximations sometimes converges slowly or not at all. Due to the disadvantages of the asymptotic approximations and the Chebyshev-type inequalities discussed here, Monte Carlo methods appear appropriate for the problem of determining the distribution of the sum of the differences. This approach is considered in Section 6.3.

6.3 Estimation of the sdd Using Dependent Observations

Monte Carlo techniques may be used to determine the sdd given any sample size n and underlying distribution. The most straightforward method is to generate many samples of size n from the underlying distribution, for each sample to calculate $\Sigma_{i=1}^{n} d_i$, and to use the appropriate order statistics to estimate desired quantiles of the sdd. By generating enough samples any degree of accuracy (within the limits of the computer) may be obtained and confidence intervals may be placed on the results, distinct advantages over the methods of Section 6.2.

The major disadvantage of this Monte Carlo approach is the computer effort involved. Many observations of $\Sigma_{i=1}^{n} d_i$ are needed for accurate estimation of sdd quantiles, with each one requiring the generation of a sample of size n. The computational effort may or may not be prohibitive, depending upon the complexity of generating each

of the n sample values, the sample size n, the required accuracy, and the quantiles of interest.

This section investigates the use of dependent, rather than independent, observations to estimate the sdd. The dependent observations are the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ associated with each sample. The genesis of the dependent observations is described in Section 6.3.1, the marginal distribution of each observation is discussed in Section 6.3.2, the form of dependence is derived in Section 6.3.3, the expected values of the sample moments are derived in Section 6.3.4, the variance of the second sample moment is derived in Section 6.3.5, and the validity of estimating sdd quantiles from the dependent observations is established in Section 6.3.6.

6.3.1 The Randomization Sample

Determining the sdd is a special case of the problem of determining the distribution of a sum of random variables. It is special in that for this sum of differences, it is known that

1. $E\{d_i\} = 0$ for $i = 1,2,\ldots,n$

and

2. Each d_i is a symmetric random variable.

From these assumptions of the randomization test it is clear that if $\Sigma_{i=1}^{n} d_i$ is observed, then each of the 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ could have occurred with equal probability. These 2^n values of $\Sigma_{i=1}^{n} \pm d_i$ may be thought of as components of a realization of a 2^n dimensional random vector.

An approach to determine the sdd using these components, shown to be valid in Section 6.3.2, is to use 2^n histograms of $\Sigma_{i=1}^{n} \pm d_i$

component values. Here each histogram contains one value of $\Sigma_{i=1}^{n} \pm d_i$ from each sample of size n. From each of the 2^n histograms estimates of the sdd moments and quantiles may be obtained using the sample moments and the results of Chapter III, since the values within each histogram are independent. While the 2^n estimators are dependent, they may still be averaged to provide one estimate which is biased as little as any of the individual estimators but uses much more information. In particular, the symmetry about zero is exploited.

While this approach is valid, the impracticality of using 2^n histograms makes it unappealing. Of course, $m \leq 2^n$ histograms could be used, but then available information in the form of the other 2^n-m components is lost, because their values are not used.

Define the "randomization sample" as the $m2^n$ dependent component values $\Sigma_{i=1}^{n} \pm d_i$ arising from m samples of size n. It is tempting to treat the randomization sample from m samples as independent observations of $\Sigma_{i=1}^{n} d_i$ by placing them into a single histogram, thereby using only a reasonable amount of core and all available information. Using the technique described in Section 5.1.2 the 2^n dependent values from one sample may be determined with much less computer effort than generating 2^n samples of size n to obtain the 2^n independent observations.

However, it is not true in general that a sample of independent observations has the same properties as a sample of dependent components of a multivariate random variable with dependencies among the components, even if the marginal distribution of all components is the same. For example, the properties of a sample of distances between random points within a circle differs depending on whether each observation is from

two independent points or the $\binom{m}{2}$ distances corresponding to m points. Thus it is not necessarily clear that treating the randomization sample as independent observations is a valid approach to determine the sdd.

But certainly the distribution of the test statistic $\Sigma_{i=1}^{n} d_i$, the sdd, and the randomization sample are closely related, since the randomization test has the mathematical properties of the parametric test as n becomes large. Therefore the next few sections examine the validity of using the randomization sample to estimate the moments and quantiles of the sdd.

6.3.2 Marginal Distribution of the Components $\Sigma_{i=1}^{n} \pm d_i$

That the components $D_k \equiv \Sigma_{i=1}^{n} \pm d_i$ for $k = 1,2,\ldots,2^n$ all have the same marginal distribution is important in showing the relationship of the randomization sample and the sdd in Section 6.3.5. The approach taken is to show that the moments of each D_k is the same as the moments of $D_1 \equiv \Sigma_{i=1}^{n} d_i$, the test statistic.

The definition of the null hypothesis and the assumptions of the randomization test lead trivially to all odd moments of each D_k being zero. Under H_0 treatment effects are zero and therefore the ith pair of observations X_i and Y_i both have the same mean. By assumption X_i and Y_i are either identically distributed or both symmetric. Therefore d_i is symmetric about zero. Symmetry follows since $d_i \equiv X_i - Y_i$ could have been $d_i \equiv Y_i - X_i$. That the $E\{d_i\} = 0$ is seen from

$$E\{d_i\} = E\{X_i - Y_i\} = E\{Y_i - X_i\} = 0$$

if the mean of X_i and Y_i exists. Since d_i is symmetric, all odd moments of d_i are zero. Therefore all odd moments of each $D_k = \Sigma_{i=1}^{n} \pm d_i$

are zero.

Now consider the pth central moment, where p is even, and an arbitrary component D_k having j minus signs (and n-j plus signs). For simplicity let the j minus signs be the last of the n positions. The value may then be represented by

$$\sum_{i=1}^{n-j} d_i - \sum_{i=n-j+1}^{n} d_i$$

The pth moment of the arbitrary component D_k is

$$E\left\{\left(\sum_{i=1}^{n-j} d_i - \sum_{i=n-j+1}^{n} d_i\right)^p\right\} \tag{5}$$

while the pth moment of the observed test statistic $D_1 \equiv \sum_{i=1}^{n} d_i$ is

$$E\left\{\left(\sum_{i=1}^{n} d_i\right)^p\right\}$$

If these moments are shown to be equal, then there is one common marginal distribution for all the D_k.

Now expression (5) may be expanded with each term being of the form

$$E\{a^{\lambda_1} b^{\lambda_2} \ldots e^{\lambda_\ell}\} \tag{6}$$

where $\ell = 1,2,\ldots,$ or p, each letter represents $\pm d_i$, and

$$\sum_{i=1}^{\ell} \lambda_i = p \text{ and } \lambda_i \geq 1 \text{ for } i = 1,2,\ldots,\ell$$

The expression (6) is equal to

$$[E\{a^{\lambda_1}\}E\{b^{\lambda_2}\}\ldots E\{e^{\lambda_\ell}\}]$$

since each of the d_i values are independent.

If any λ_i is odd, the associated expected value is zero, due to the symmetry of the d_i's. But whenever all λ_i values are even, the minus signs have no effect on the expected value of the term. Therefore the only nonzero terms in the expansion of expression (5) are the same whether $-d_i$ or $+d_i$ is used. Since the moments of each of the 2^n possible component values are the same, the 2^n marginal distributions are the same.

6.3.3 Dependence Among the Components D_k

As noted earlier, the component values D_k, $k = 1,2,\ldots,2^n$, corresponding to each sample of size n can be viewed as one realization of a 2^n dimensional multivariate random variable. In the last section the marginal distributions of the D_k's were shown to be identical. The form of dependence among the components is the other facet of the multivariable distribution necessary for this study. This dependence is developed in this section.

Three results pertaining to the dependence among the D_k's are important for current purposes:

1. $\Sigma_{k=1}^{2^n} D_k = 0$ for any sample

2. Corr $(D_j, D_k) = (n-2\ell)/n$

   where $\ell = 1,2,\ldots,$ or n is number of conflicting signs between D_j and D_k, and

3. $\Sigma_{j=1}^{2^n}\Sigma_{k=1}^{2^n}$ Corr $(D_j, D_k) = 0$ for any sample.

The first result is clearly true, since for each D_k there is a corresponding component $D_j = -D_k$ with n conflicting signs. Thus any sum of these pairs is zero, including the sum of all components from any sample.

The second result may be derived by considering any two arbitrary components, D_j and D_k, with signs conflicting in ℓ of the n positions. Let these positions be the last ℓ for convenience. Then

$$\begin{aligned}
\text{Cov}\ (D_j, D_k) &= E\{D_j D_k\} - E\{D_j\}E\{D_k\} \\
&= E\{D_j D_k\} \text{ since the components have mean zero under } H_0 \\
&= E\{\sum_{i=1}^{n-\ell} d_i^2 - \sum_{i=n-\ell+1}^{n} d_i^2 + \text{crossproducts}\} \\
&= (n - \ell)\sigma^2 - \ell\sigma^2 + 0 \\
&= (n - 2\ell)\sigma^2
\end{aligned}$$

where σ^2 is the variance of the common marginal distribution of the components; that is,

$$\sigma^2 \equiv V\{D_j\} = V\{D_k\}$$

Therefore the correlation between the two arbitrary components with conflicting signs is

$$\text{Corr}\ (D_j, D_k) = \frac{\text{Cov}\ (D_j, D_k)}{\sqrt{n\sigma^2}\ \sqrt{n\sigma^2}} = \frac{(n - 2\ell)\sigma^2}{n\sigma^2} = \frac{n - 2\ell}{n}$$

Note that D_k and $-D_k$ have perfect negative correlation of -1 since $\ell = n$. If n/2 signs conflict, the correlation is zero. In general each conflicting sign reduces the correlation by 2/n.

The third result involving the sum of all 2^{2n} correlations associated with any sample of size n,

$$\sum_{j=1}^{2^n} \sum_{k=1}^{2^n} \text{Corr}\,(D_j, D_k) = 0$$

may be shown to be true by noting this sum is equal to

$$2^n \left[\binom{n}{0} \frac{n-0}{n} + \binom{n}{1} \frac{n-2}{n} + \binom{n}{2} \frac{n-4}{n} + \ldots + \binom{n}{n} \frac{n-2n}{n} \right] \tag{7}$$

since the first summation only fixes D_j, thereby setting a reference for D_k. The correlation is dependent only upon the number of conflicting signs between D_j and D_k, hence the factor 2^n replaces the first summation. Now for each D_j, $\binom{n}{0}$ values D_k have a correlation with D_j of $\frac{n-0}{n}$, that one value being $D_k = D_j$. Likewise the $(\ell+1)$ th term $\binom{n}{\ell} \frac{n-2\ell}{n}$ results from D_k differing from D_j by exactly ℓ signs $\binom{n}{\ell}$ times.

Proceeding directly now

$$\sum_{j=1}^{2^n} \sum_{k=1}^{2^n} \text{Corr}\,(D_j, D_k) = 2^n \left[\sum_{\ell=0}^{n} \binom{n}{\ell} \frac{(n-2\ell)}{n} \right]$$

$$= \frac{2^n n!}{n} \left[\sum_{\ell=0}^{n} (n-2\ell)/\ell!(n-\ell)! \right]$$

$$= \frac{2^n n!}{n} \left[\sum_{\ell=0}^{[n/2]} \left(\frac{n-2\ell}{\ell!(n-\ell)!} + \frac{n-2(n-\ell)}{(n-\ell)!\ell!} \right) \right]$$

$$= \frac{2^n n!}{n} \left[\sum_{\ell=0}^{[n/2]} \frac{0}{\ell!(n-\ell)!} \right]$$

$$= 0$$

whereby proving the third result.

A more general result is that the sum of any m correlations with a total of mn/2 conflicting signs is zero since

$$\sum_{i=1}^{m} \text{Corr}\,(D_j, D_k) = \sum_{i=1}^{m} \frac{n-2\ell_i}{n} = \frac{1}{n}\left[mn - 2 \sum_{i=1}^{m} \ell_i\right]$$

$$= \frac{1}{n}\left[mn - 2(mn/2)\right] = 0$$

where ℓ_i is the number of conflicting signs in the ith correlation.

6.3.4 Properties of the Randomization Sample Moments

In this section it is shown that the randomization sample moments are unbiased estimators of the moments of the sdd. This result is used in the next section to show that sdd quantiles may be estimated from the randomization sample much as if the $n2^n$ component values are independent.

First consider the odd moments. Since the sdd is symmetric about zero, all odd distribution moments are zero. The odd moments of the randomization sample are very good estimators of the sdd odd moments (in a trivial sense), since the randomization sample is symmetric for any m samples. Letting $D_{\ell k}$ denote the kth component from the ℓth sample, the odd randomization moments may be seen to be zero since for each $D_{\ell k}$ observed, $D_{\ell j} = -D_{\ell k}$ is also observed, causing symmetry about zero in the sample. Having a symmetric sample to estimate properties of a

symmetric distribution is certainly appealing.

Now consider the even moments. The main result is that the central even moments of the randomization sample

$$\frac{\sum_{\ell=1}^{m}\sum_{k=1}^{2^n}\left(D_{\ell k}-\frac{\sum_{\ell=1}^{m}\sum_{k=1}^{2^n}D_{\ell k}}{m2^n}\right)^p}{m2^n} \qquad p = 2,4,6,\ldots$$

are unbiased estimators of the even moments of the sdd. Note that the denominator $m2^n$ would be $m2^n-1$ if $m2^n$ independent observations were being used. The degree of freedom for estimating the mean is not lost in the randomization sample since the sum of the differences is known to be exactly zero for each of the m samples.

The result is first proved for the variance for clarity and then proved for all general even moments for generality. If σ^2 is the variance of the sdd, then for each sample $\ell = 1,2,\ldots,m$

$$\sigma^2 = V\{\sum_{i=1}^{n} d_i\} = V\{D_{\ell 1}\} = V\{D_{\ell k}\} = E\{D_{\ell k}^2\}$$

and therefore

$$E\{\hat{\sigma}^2\} = E\left\{\frac{\sum_{\ell=1}^{m}\sum_{k=1}^{2^n}\left(D_{\ell k}-\frac{\sum_{\ell=1}^{m}\sum_{k=1}^{2^n}D_{\ell k}}{m2^n}\right)^p}{m2^n}\right\}$$

$$= E\left\{\frac{\sum_{\ell=1}^{m}\sum_{k=1}^{2^n}D_{\ell k}^2}{m2^n}\right\} \text{ since } \sum_{i=1}^{2^n} D_{\ell k} = 0$$

$$= \frac{\sum_{\ell=1}^{m} \sum_{k=1}^{2^n} E\{D_{\ell k}^2\}}{m2^n}$$

$$= \frac{\sum_{\ell=1}^{m} \sum_{k=1}^{2^n} \sigma^2}{m2^n} = \sigma^2$$

The result may be generalized to the higher even moments μ_p, $p = 2,4,6,\ldots$ as follows:

$$E\{\hat{\mu}_p\} = E\left\{\frac{\sum_{\ell=1}^{m} \sum_{k=1}^{2^n} \left(D_{\ell k} - \frac{\sum_{\ell=1}^{m} \sum_{k=1}^{2^n} D_{\ell k}}{m2^n}\right)^p}{m2^n}\right\}$$

$$= E\left\{\frac{\sum_{\ell=1}^{m} \sum_{k=1}^{2^n} D_{\ell k}^p}{m2^n}\right\}$$

$$= \sum_{\ell=1}^{m} \sum_{k=1}^{2^n} \mu_p/m2^n \quad \text{from Section 6.3.2}$$

$$= \mu_p$$

The dependent observations must be taken as a group, since otherwise $\Sigma_{k=1}^{2^n} D_{\ell k}$ does not vanish, thereby changing the moments. Dwass [1957] suggested taking a random sample of the 2^n observations when 2^n was so large that calculation of all values required too much effort. However, it is easy to see that the expected value of the even moments is affected by such a procedure, resulting in the reference distribution

being different than the distribution of $\Sigma_{i=1}^{n} d_i$, the test statistic. Chung and Fraser [1958] suggested using subgroups of the 2^n components whose sum is zero thereby leaving the moments unchanged.

6.3.5 Efficiency of the Randomization Sample Approach

Having established the validity of estimating moments of the sdd using the randomization sample rather than independent observations in Section 6.3.4, the value of randomization sample estimation is examined here. In particular, the value of calculating the 2^n-1 additional component values associated with each sample is quantified in Section 6.3.5.1 by deriving the variance of the randomization sample estimator of $\sigma^2 \equiv V\{\Sigma_{i=1}^{n} d_i\}$ and comparing to the corresponding independent observation variance. Further the randomization sample estimators are shown to be consistent. The tradeoff between the reduction in variance and the effort of calculating the additional component values for each sample is examined in some detail in Section 6.3.5.2, concluding that for samples of size $n \leq 15$ the randomization sample requires less computation for the same accuracy.

6.3.5.1 Variance and Consistency of the Randomization Sample Estimators. The variance of the unbiased estimator of σ^2 based on the randomization sample, denoted here by $\hat{\mu}_2$, is

$$V\{\hat{\mu}_2\} = [\mu_4 - (3-2n)\sigma^4]/m$$

This variance may be derived as follows:

$$V\{\hat{\mu}_2\} = V\left\{\frac{\sum_{i=1}^{m}\sum_{k=1}^{2^n} D_{jk}^2}{m2^n}\right\}$$

$$= \frac{m}{(m2^n)^2} V\{\sum_{k=1}^{2^n} D_k^2\}$$

$$= \frac{1}{m2^{2n}} \left[\sum_{k=1}^{2^n} V\{D_k^2\} + \sum_{\substack{j=1 \\ j\neq k}}^{2^n} \sum_{k=1}^{2^n} Cov\{D_j^2, D_k^2\}\right]$$

$$= \frac{1}{m2^{2n}} \left[2^n(\mu_4 - \sigma^4) + \sum_{\substack{j=1 \\ j\neq k}}^{2^n} \sum_{k=1}^{2^n} \mu_4 - \{\frac{8(n-\ell)\ell}{n^2} + 1\}\sigma^4\right]$$

where ℓ is the number of conflicting signs between D_j and D_k, by results of Appendix E

$$= \frac{1}{m2^{2n}} [2^n(\mu_4 - \sigma^4) + 2^n(2^n-1)(\mu_4 - \sigma^4) - \frac{2^n 8\sigma^4}{n^2} \sum_{\ell=1}^{n} \binom{n}{\ell}(n-\ell)\ell]$$

using logic similar to that used to derive expression (7)

$$= \frac{1}{m2^n} [(\mu_4 - \sigma^4) + (2^n-1)(\mu_4 - \sigma^4) - \frac{8\sigma^4}{n^2} (2^{n-2}n(n-1))]$$

by results of Appendix E

$$= \frac{1}{m} [(\mu_4 - \sigma^4) - \frac{2^{n+1}}{2^n} \frac{n(n-1)}{n^2} \sigma^4]$$

$$= \frac{1}{m} [(\mu_4 - \sigma^4) - 2(\frac{n-1}{n})\sigma^4]$$

$$= \frac{1}{m} [\mu_4 - (3-2/n)\sigma^4]$$

thereby proving the result. As the number of samples, m, becomes large the variance goes to zero. Recalling that the estimator is unbiased

establishes consistency since then

$$\lim_{m\to\infty} |\widehat{F_m^{-1}}(p) - F_m^{-1}(p)| < \varepsilon$$

which is the definition of a consistent sequence of estimators.

Note that m = 1 sample yields a valid estimate of the variance of the components and the variance of the estimate is defined for m = 1. This is not true for independent observations. Wilkes [1962] shows the variance of the estimate of the variance based on independent observations, denoted here by $\hat{\sigma}^2$, to be

$$V\{\hat{\sigma}^2\} = [\mu_4 - \frac{m-3}{m-1}\sigma^4]/m$$

which is undefined for m = 1.

Let m' be the number of randomization samples and let m be the number of samples using independent observations. The reduction in variance of the estimate due to calculating the additional 2^n-1 component values for each sample is

$$V\{\hat{\sigma}^2\} - V\{\hat{\mu}_2\} = \sigma^4[(3-2n) - \frac{m-3}{m-1}]/m'$$

which is always nonnegative. The impact of this variance reduction is that fewer samples are needed using the randomization sample than when using independent observations. The variances of the estimates are equal when

$$[\mu_4 - \frac{m-3}{m-1}\sigma^4]/m = [\mu_4 - (3-2n)\sigma^4]/m'$$

or when

$$\frac{m[\alpha_4 - (3-2/n)]}{[\alpha_4 - \frac{m-3}{m-1}]} = m'$$

where $\alpha_4 = \mu_4/\sigma^4$ is the fourth standardized moment.

Since m must be large for an accurate estimate of the variance,

$$\frac{m-3}{m-1} \approx 1$$

and

$$\frac{[\alpha_4 - (3-2/n)]}{\alpha_4 - 1} \approx \frac{m'}{m}$$

For normally distributed D_k's, $\alpha_4 = 3$ and

$$\frac{m'}{m} \approx 1/n$$

This ratio is smaller for light tailed distributions and larger for heavy tailed distributions, but never is greater than one.

6.3.5.2 Tradeoff Considerations. The tradeoff between the use of the randomization sample and independent observations is now examined. Let W_1 be the effort (work) of generating each difference d_i. Then the effort to generate one sample is nW_1. Let W_2 be the effort of calculating each $\pm d_i$ component and placing it in a histogram. Then $nW_1 + 2^nW_2$ is the effort associated with each sample using dependent observations and $nW_1 + W_2$ is the effort per sample using independent observations.

Now to obtain the same variance when the D_k's are normally

distributed $m' \approx m/n$. The total effort is then $\frac{m}{n}[nW_1 + 2^n W_2]$ for dependent components and $m[nW_1 + W_2]$ for independent observations. The dependent component approach requires less effort if

$$\frac{m}{n}[nW_1 + 2^n W_2] < m[nW_1 + W_2]$$

or

$$\frac{2^n - n}{n(n-1)} < \frac{W_1}{W_2}$$

The ratio W_1/W_2, the effort to generate one d_i value divided by the effort to deal with one $\pm d_i$ component, is much greater than one. W_1 is much larger than W_2 in general since the generation of random values requires the higher order operations of exponentiation and multiplication while only logic statements and addition are required for dealing with the components. Commonly the ratio W_1/W_2 is in the interval (200,2000).

The values of n corresponding to $(2^n-n)/(n(n-1))$ falling in this interval are approximately n = 16, 17, 18, and 19. Thus for the smaller values of n considered in this research the use of the randomization sample is more efficient. The impact of this increased efficiency is more apparent when considering specific values of n. For n = 3 and $W_1/W_2 = 1000$, the effort associated with the randomization sample is only .001 that of independent observations as can be seen by taking the ratio of the total effort formulae. For n = 7 the ratio of effort is about .003. Thus brute force application of the Monte Carlo technique, by using independent observations, would require orders of

magnitude more computer time for the same accuracy.

6.3.6 Estimation of sdd Quantiles Using the Randomization Sample

As mentioned in Section 6.3.1, it is not valid in general to use dependent observations as if they were independent, except for the mean. An argument is now advanced that the randomization sample of dependent components may be used to estimate quantiles of the sdd as if the components were independent observations, based on the validity of randomization sample moment estimation established in Section 6.3.4 and the consistency of the estimators established in Section 6.3.5.

Consider the $m2^n$ dependent components of the randomization sample arising from m samples of size n. As m approaches infinity the moments of the randomization sample converge to the moments of the sdd, since the randomization moments are unbiased and consistent estimators of the sdd moments. Then, for large m, the randomization sample has the same properties as the sample of independent observations. Therefore, for large m, the quantiles of the sdd may be estimated directly from the randomization sample. (The techniques of Chapter III may be used.)

Another argument, possibly more rigorous, is based on Cornish-Fisher expansions, discussed previously in Section 6.2. First note that since the randomization sample moments are consistent estimators of the sdd moments, simple transformations of these estimators are consistent estimators of the cumulants of the sdd, since cumulants and moments have a one-to-one relationship.

As seen in Section 6.2, a quantile $F^{-1}(p)$ of an arbitrary distribution having known cumulants κ_i, $i = 1,2,\ldots$, may be expressed in terms of the corresponding normal quantile U_p by

$$F^{-1}(p) = U_p + (1/6)(U_p^2 - 1)\kappa_3 + (1/24)(U_p^3 - 3U_p)\kappa_4$$

$$+ (1/36)(2U_p^3 - 5U_p)\kappa_3^2 + \ldots$$

Since the randomization sample may be used to estimate each cumulant consistently, a consistent estimator of the pth quantile $F^{-1}(p)$ is obtained by substituting cumulant estimates for each κ_i.

Thus, for large m, quantiles, like moments, may be estimated using the randomization sample as if the components were independent observations.

6.4 Monte Carlo Determination of the sdd

This section discusses some aspects of the determination of the distribution of the sum of differences, the sdd, using Monte Carlo techniques. Variance reduction techniques are the topic of Section 6.4.1 and confidence intervals of moments and quantiles are discussed in Section 6.4.2. The computer program which implements the logic discussed here is described in more detail in Appendix D.

6.4.1 Variance Reduction Techniques

When the Monte Carlo approach is used to estimate a parameter, variance reduction techniques can often be used to reduce the variance of the estimate of the parameter. Since the variance can be reduced by simply lengthening the computer run, interest centers around minimizing the variance for a given run length.

The parameters being estimated in this study are the quantiles of the sdd. Three variance reduction techniques are applied to the estimation of the pth quantile: (1) use of the randomization sample rather

than independent observations, (2) stratified sampling, and (3) antithetic variables. The use of the randomization sample and the corresponding reduction of variance was discussed in Section 6.3. In Section 6.4.1.1 stratified sampling is discussed and antithetic variables are discussed in Section 6.4.1.2.

6.4.1.1 Stratified Sampling. Stratified sampling, a variance reduction technique commonly used in survey sampling, may be used in Monte Carlo studies to increase the accuracy of the estimate of a parameter without increasing the computer effort. While some increase in coding effort is necessary to implement stratified sampling logic in a computer program, the savings in computer execution time may be worth the effort. A discussion of stratified sampling in the context of survey sampling may be found in Cochran [1966]. Klienjen [1974] and Moy [1966] discuss stratified sampling and other variance reduction techniques in the context of simulation and Monte Carlo studies.

In this study the observations X_i and Y_i are stratified to ensure that each segment of the distribution is represented in the correct proportion. Each random variable is partitioned into K strata with probability $p_k = 1/K$ of any particular observation falling in the kth stratum. Every N observations exactly $n_k = N/K$ observations are selected from stratum k, k = 1,2,...,K.

Stratification causes dependence among the observations, since if X_i is known to have come from stratum k, then X_j has a lesser probability of coming from stratum k than if nothing is known about X_i. Care must be taken to ensure that the n observations of X_i and of Y_i in any given sample are taken independently (although X_i and Y_i may be

correlated).

The logic of the stratification process in subroutine STRAT (Appendix D) ensures independence. One initialization call to the subroutine is required at which time each p_k is set to 1/K and the number of observations to be taken in the kth stratum n_k is set to a predetermined value M = N/K and the total number of observations NT is set to N. After initialization, the subroutine is called once each time n independent observations are needed. For each of the n values, a U(0,1) value r is generated, the stratum is determined as the largest k such that $r \leq \Sigma_{i=1}^{k} p_i$. Then n_k is reduced by one to reflect that one less value is to be generated from stratum k in the future. The value r in the (0,1) interval is transformed to the same relative position in the kth stratum ((k-1)/K,k/K) by

$$r = [k - 1 + (r - \sum_{i=1}^{k-1} p_i)/p_k]/K$$

The n independent values of r are later passed to the process generator, resulting in n independent values of X_i or of Y_i having the desired distribution.

After the n independent values of r are generated, the number of observations yet to be taken NT is decreased by n and the stratum probabilities p_k are modified to reflect the number of observations still to be taken from stratum k by

$$p_k = n_k/NT \qquad k = 1,2,\ldots,K$$

These new p_k values are then used in the next call to the subroutine. If NT < n the p_k values are reinitialized to 1/K and NT is reset to N.

Subroutine STRAT in Appendix D performs this logic with the minor difference that $\Sigma_{i=1}^{k} p_i$ is stored as a variable rather than storing each p_i separately.

6.4.1.2 Antithetic Variables. Another variance reduction technique which is straightforward to apply is antithetic, or negatively correlated, variables. If two values $\hat{\delta}_1$ and $\hat{\delta}_2$ are unbiased estimates of a parameter δ, they may be combined to form one unbiased estimate

$$\hat{\delta} = (\hat{\delta}_1 + \hat{\delta}_2)/2$$

with

$$V\{\hat{\delta}\} = \frac{V\{\hat{\delta}_1\} + V\{\hat{\delta}_2\} + 2\ Cov\{\hat{\delta}_1,\hat{\delta}_2\}}{4}$$

The variance of the estimate $\hat{\delta}$ is then less if $\hat{\delta}_1$ and $\hat{\delta}_2$ are negatively correlated than if they are independent.

Negative correlation is commonly induced by making two Monte Carlo runs using the same random number seeds, but adding the additional logic

$$r = 1 - r$$

each time a U(0,1) value r is generated. Thus if the first run is based on a disproportionately large number of small values of r, the second run should be biased high. The averaged result is then more accurate than if two independent runs had been made.

A problem arises in the current study, however, since both r

and 1-r map onto the same 2^n randomization component values. This is true because X_i is symmetric about zero, implying that if r maps onto $+X_i$ then 1-r maps onto $-X_i$. The Y_i values are similarly reversed. Therefore the differences in the second run are exactly the negative of the differences in the first run. The 2^n component values are therefore identical.

Antithetic observations may be generated for randomization samples by using

$$r = \begin{cases} r - 1/2 \text{ if } r \geq 1/2 \\ r + 1/2 \text{ if } r < 1/2 \end{cases}$$

Then r = .4 becomes r = .9 and r = .6 becomes r = .1, resulting in antithetic observations. They have negative correlation because distance from the mean is the property of the random observations which affects the 2^n component values.

6.4.2 Confidence Intervals

Confidence intervals are needed for estimates to be properly interpreted. In Chapter V a closed form result was used to calculate the variance of the estimate of the power of the randomization test. No similar result is available here due to the use of variance reduction techniques.

The variance of the estimators of the quantiles and moments of the sdd are obtained by replication. That is, after N samples are generated and analyzed, one independent estimate $\hat{\delta}_i$ is calculated for each parameter from this replication. A single composite estimate is calculated after the Mth replication by averaging the previous

replication estimates

$$\hat{\delta} = \sum_{i=1}^{M} \hat{\delta}_i / M$$

The variance of the composite estimate is easily estimated by the sample variance S^2 since the replication estimates are independent. The averaged estimate is asymptotically normally distributed by the central limit theorem, so a $100(1-\alpha)\%$ confidence interval is

$$\hat{\delta} \pm t_{\alpha/2,M-1}(S/\sqrt{M})$$

The effects of the variance reduction techniques do not have to be dealt with directly using the replication technique since they are buried within the replications. The randomization components are independent between samples and therefore between replications. The stratification of observations begins anew each replication, so the observations are independent between replications. The antithetic variables effect on variance may be placed within each replication if, instead of using separate runs, immediately after a sample is generated and analyzed the antithetic sample is generated within the same replication.

Confidence intervals based on replications may be placed on both quantiles and moments. Although the moments are known from Section 6.1, confidence intervals may be placed on the known values to provide a mechanism for verification of the computer program and validation of the theory. The confidence intervals should bound the known values $100(1-\alpha)\%$ of the time. If they do not, an error in programming or theory is indicated, since the estimators of the moments were shown to

be consistent in Section 6.3.4. By definition, a consistent estimator must converge to the true value. The confidence interval provides the signal that convergence to the true value is or is not taking place.

CHAPTER VII

THE POWER OF PAIRED SAMPLE TESTS

In Chapters V and VI methods were developed to study the component randomization test and the corresponding parametric test, respectively. The methods are general in that they are appropriate for any underlying distribution and sample size. The techniques of these two chapters are implemented here, using also the methods of quantile estimation discussed in Chapter III and of process generation discussed in Chapter IV, to estimate the power of various tests of location for the paired sample design.

Specifically, the tests considered are the component randomization test, the corresponding parametric test, and the commonly used normal test. The "corresponding parametric test" is to reject H_0 if the test statistic is greater than the $(1-\alpha)$ quantile of the distribution of the test statistic. In the case of normally distributed observations, the parametric test and the normal test are one and the same.

Section 7.1 discusses these tests and their relationship to each other. Section 7.2 is an overview of the analysis performed for each underlying distribution and sample size in later sections. Section 7.3 contains analyses of the power of these tests under several specific distributions and sample sizes.

7.1 The Tests

As noted in the introduction, little study of the power of component randomization tests has been undertaken for finite sample sizes. Knowledge of the power is important for evaluating the tradeoffs between the robustness and computational ease of the various tests. The component randomization test, the most powerful distribution-free test, has the disadvantage of requiring great computational effort. The corresponding parametric test, which assumes knowledge of the exact form of the distribution of the test statistic, would be expected to have more power than distribution-free tests which do not make use of this information, but knowledge of this distribution is rare. The normal test is of interest because it is widely used and uniformly most powerful in the case of normally distributed observations.

Quantifying the difference in power between the first two tests is helpful in deciding which test to use in a particular situation. The power of the normal test under the null hypothesis, which is the probability of type I error, is important in determining the degree of inappropriateness of the normal test for various underlying distributions.

7.2 Analysis Methodology

Before looking at specific results, the steps necessary to determine the powers of the tests of interest are now discussed. Conceptually the approach is straightforward, requiring only two steps:

1. Determine $F_n^{-1}(p)$, the pth quantile of the reference distribution F for the parametric test. The power of the parametric

test or the normal test may then be determined from $1 - \beta = 1 - F(C_\alpha - \sqrt{n}\, E\{d_i/\sigma\})$ where C_α is the $(1-\alpha)$ quantile of F or the normal distribution.

2. Generate N samples of size n from the distribution of interest with specified expected differences. For each sample perform the four tests and update a counter for each test which rejects H_0. The count divided by N is then the estimate of the power of the test.

Thus the power of the parametric tests may be determined from either phase. Note that the power of the component randomization test may be determined only from Phase 2.

The complexity of implementation arises only because the straightforward technique requires an unreasonable amount of computational effort. The results of Chapter VI and Chapter III, may be used to perform Phase I more efficiently, while Chapter V may be used for Phase 2. Process generation, as discussed in Chapter IV, is used in both phases to generate random observations having desired properties. Phase 1 is implemented using the computer program in Appendix D and Phase 2 is implemented using the program in Appendix B.

7.3 The Power of Certain Tests

In this section the results arising from the techniques developed previously are presented. The probability of type II error, β, which is one minus the power, is given in both tabular and graphical form for each of the tests for various underlying distributions and sample sizes. The aim of this section is to examine some specific examples to determine the relationship of power to other factors in a general way. The

normal, exponential, uniform, and absolute lambda (as discussed in Chapter IV) cases are examined in Sections 7.3.1, 7.3.2, 7.3.3, and 7.3.4 respectively. These distributions were selected to illustrate the effect of tail weight and shape on the power of the tests.

In each section results are given in the form of operating characteristic (O.C.) curves. The curves for sample sizes 4, 7, 11, and 15 are grouped by underlying distribution and α value. The values of α considered are 1/128, 1/16, and 1/8, the closest values to .01, .05, and .1 of the form $m/2^n$ where n is integer. These values are necessary since α can take on only values of this form for permutation tests.

7.3.1 Normal Observations

In this section the power of the tests of interest, presented in the form of O.C. curves, are compared for the case of normally distributed observations. Since the appropriate parametric test is the normal test, only the component randomization and normal tests are of interest here.

Figure 15 shows graphically the results of this section, which are also presented in tabular form in Table 6. Several observations can be made from the graphs corresponding to α = 1/128, 1/16, and 1/8, respectively:

1. The value of β decreases as sample size increases.
2. The value of β decreases as $E\{d_i/\sigma\}$ increases.
3. The value of β decreases as α increases.
4. The normal test, which dominates the component randomization test in this case of normally distributed observations, has asymptotically the same power as the component randomization test. The dominance

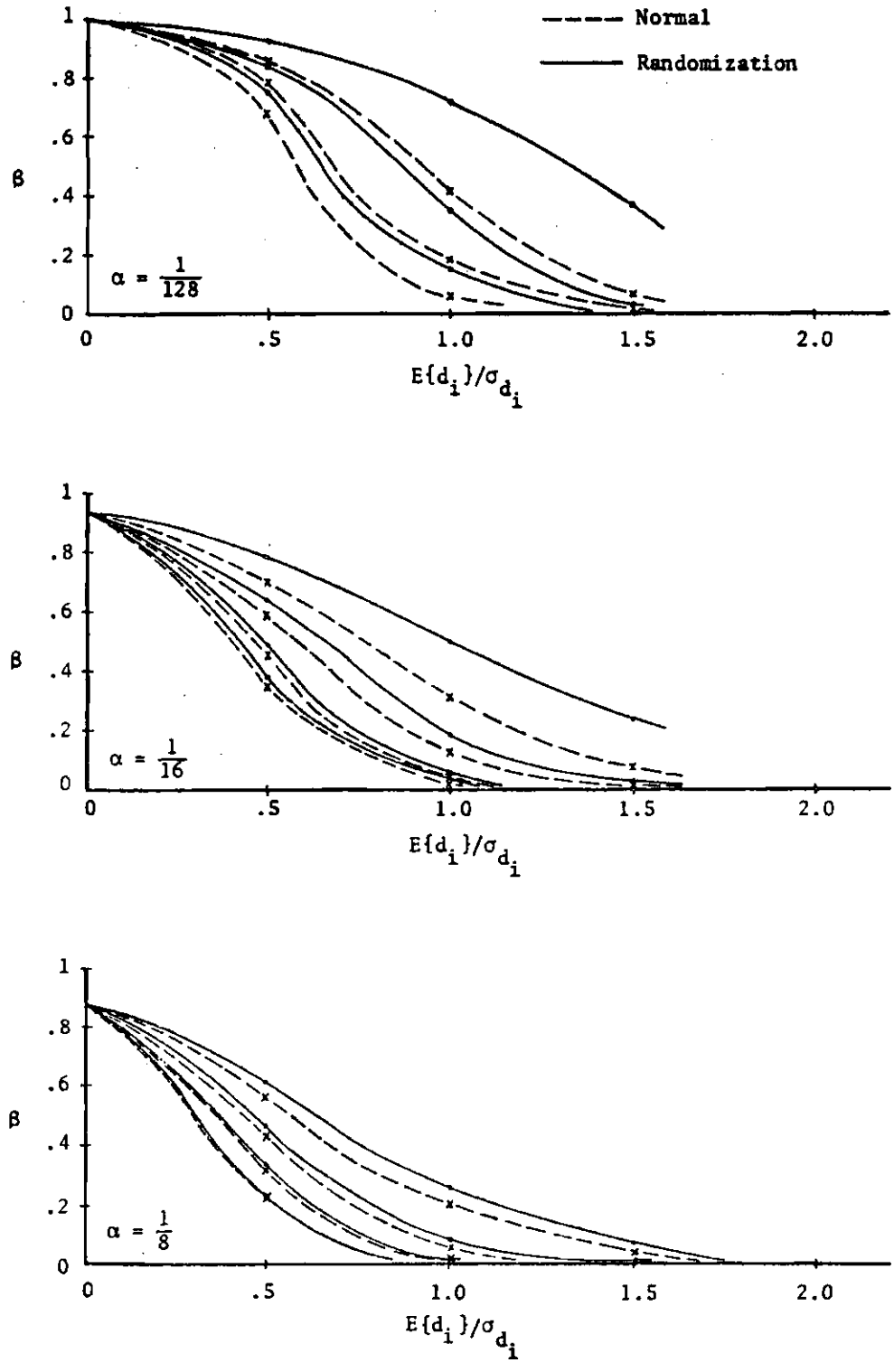

Figure 15. Operating Characteristic Curves from Normal Observations

Table 7. Type II Error (β) of Randomization and Parametric Test (Standard Deviations Are Shown in Parentheses)

n \ δ/σ	0	.5	1.0	1.5
	$X,Y \sim N(0,.5)$			
	$\alpha = 1/128 = .0078125$		$z_\alpha = 2.417$	
7	.991875 .991875	.93(.0046) .86	.71(.0083) .41	.38(.0088) .06
11	.991875 .991875	.84(.0081) .78	.35(.015) .18	.036(.0059) .005
15	.991875 .991875	.76 (.017) .69	.15(.013) .07	neg neg
	$\alpha = 1/16 = .0625$		$z_\alpha = 1.534$	
4	.9375 .9375	.77(.0066) .70	.50(.0079) .32	.24(.0067) .07
7	.9375 .9375	.64(.0088) .58	.19(.0072) .13	.02(.0026) .01
11	.9375 .9375	.49(.011) .45	.051(.0070) .04	neg neg
15	.9375 .9375	.38(.020) .35	.012(.0040) .01	neg neg
	$\alpha = 1/8 = .125$		$z_\alpha = 1.150$	
4	.875 .875	.61(.0077) .56	.27(.0069) .20	.075(.0042) .03
7	.875 .875	.46(.0091) .43	.09(.0051) .07	.005(.0013) .002
11	.875 .875	.33(.010) .31	.016(.0040) .015	neg neg
15	.875 .875	.22(.017) .22	.0038(.0022) .003	neg neg

of the normal test decreases as α increases and as n increases. This is logical since many observations are needed for the tail areas of the randomization reference distribution to be well developed. An α value far out in the tail requires more observations for the same power. Thus for normally distributed observations the largest loss of power in using the component randomization test is for small n and small α.

7.3.2 Exponential Observations

Results are given here for exponentially distributed observations. The form of the results is the same as for the normal case of the last section, except here the parametric test is distinct from the normal test.

Figure 16 shows the parametric reference distribution for the parametric test under exponential observations as determined by Phase 1. The upper half of the reference distribution is plotted on normal probability paper for sample sizes n = 4, 7, 11, 15, and infinity. The reference distribution as n approaches infinity is the normal distribution (the straight line in the figure) with the difference between the normal distribution and the parametric reference distribution for finite sample sizes being greatest for small samples. For all sample sizes the normal distribution lies above the parametric reference distributions above approximately the 95% quantile and below the parametric reference distribution otherwise.

Table 7 contains the results of this section in tabular form. Three tables are shown, corresponding to α = 1/128, 1/16, and 1/8, respectively. Each of the entries are the probability of type II error, β, for various values of n and $E\{d_i/\sigma\}$. Results for the component

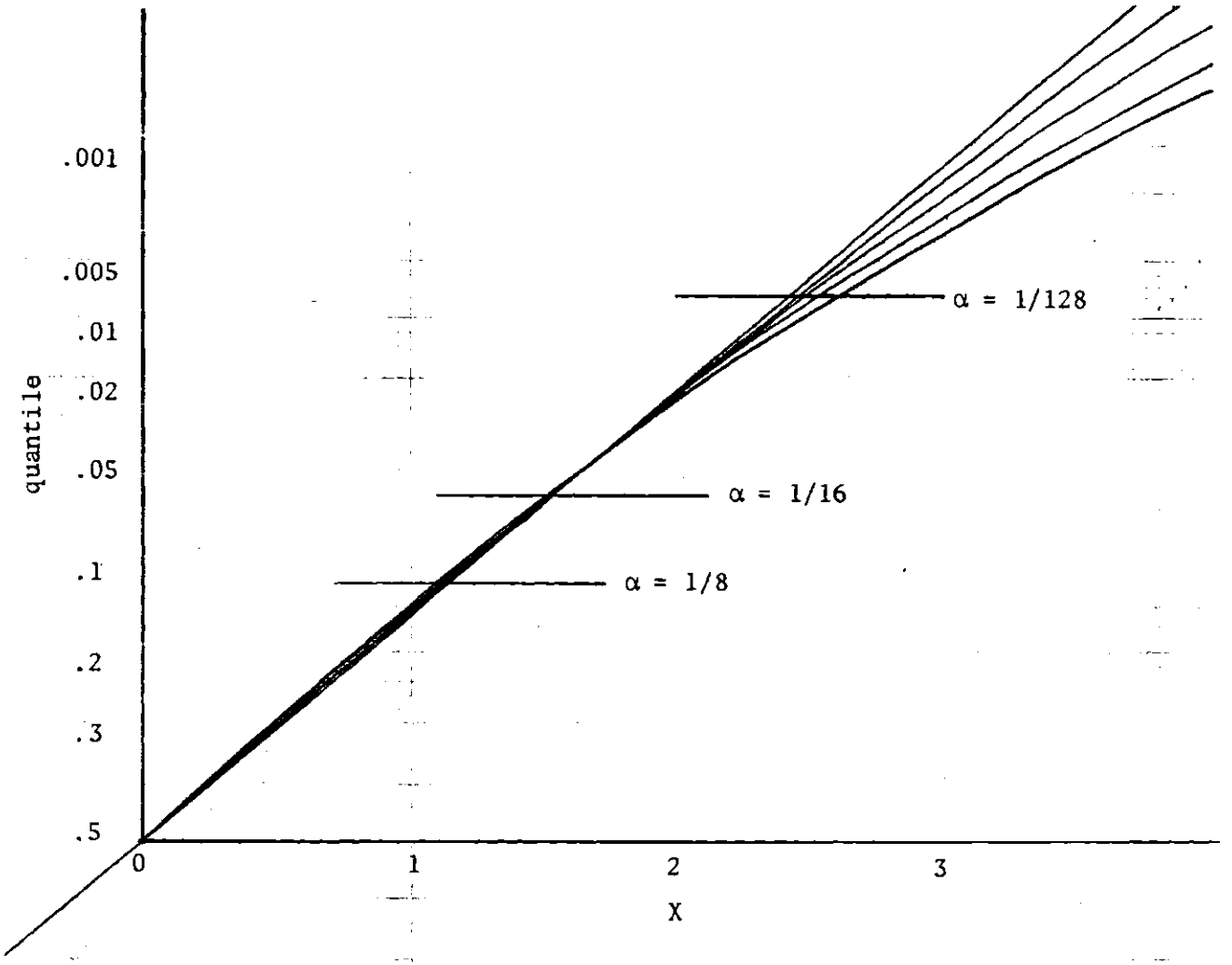

Figure 16. Reference Distributions Arising from Exponential Observations

Table 8. Type II Error (β) for Randomization (R), Parametric (P), and Normal (N) Tests Arising from Exponential Observations

n \ δ/σ	0			.5			1.0			1.5			C_α
	R	P	N	R	P	N	R	P	N	R	P	N	
	X,Y ~ Exp α = 1/128 = .0078125				z_α = 2.417								
7	$\frac{127}{128}$	$\frac{127}{128}$	.990	.86 (.0063)	.89	.86	.59 (.0090)	.48	.40	.34 (.0086)	.07	.06	2.54 (.02)
11	$\frac{127}{128}$	$\frac{127}{128}$	.990	.76 (.013)	.80	.79	.30 (.0089)	.20	.18	.07 (.0046)	.008	.007	2.49 (.01)
15	$\frac{127}{128}$	$\frac{127}{128}$	.991	.66 (.024)	.70	.69	.13 (.023)	.08	.07	neg	neg	neg	2.47 (.03)
	α = 1/16 = .0625				z_α = 1.534								
4	.9375	.9375	.940	.67 (.0074)	.70	.69	.40 (.0077)	.30	.31	.21 (.0065)	.06	.07	1.51 (.004)
7	.9375	.9375	.939	.54 (.0091)	.61	.61	.17 (.0069)	.12	.13	.035 (.0034)	.01	.01	1.52 (.007)
11	.9375	.9375	.938	.42 (.016)	.45	.45	.068 (.0046)	.04	.04	.005 (.0013)	.0005	.0005	1.53 (.004)
15	.9375	.9375	.938	.34 (.019)	.34	.34	.01 (.007)	.01	.01	neg	neg	neg	1.53 (.02)
	α = 1/8 = 1.25				z_α = 1.150								
4	.875	.875	.89	.51 (.0079)	.54	.56	.22 (.0065)	.16	.18	.08 (.0043)	.03	.035	1.09 (.004)
7	.875	.875	.88	.40 (.0090)	.43	.41	.09 (.0052)	.06	.07	.013 (.002)	.005	.005	1.11 (.006)
11	.875	.875	.88	.29 (.014)	.30	.30	.03 (.0031)	.01	.01	.002 (.0008)	neg	neg	1.13 (.006)
15	.875	.875	.88	.20 (.016)	.20	.20	neg	neg	neg	neg	neg	neg	1.14 (.01)

randomization test are in the columns labeled "R". Similarly the parametric test results are under "P" and the normal test results are under "N". For the randomization test the estimate of the standard deviation of the result is given in parentheses. The results for "P" and "N" are deterministic calculations from the results of Phase 1. The critical values used for the parametric test are determined from Figure 16 and are given in the right-hand column under C_α.

The most surprising aspect of these results is that the power of the parametric test does not dominate the power of the component randomization test. Especially for alternative hypotheses close to H_0, the component randomization test has more power for all n and values of α. At first glance this is counter-intuitive, since the usual circumstance is that power is lost in obtaining distribution-free properties. Note that over the range of values of $E\{d_i/\sigma\}$, however, the parametric test is indeed more powerful.

An intuitive rationale for the greater power of the component randomization test for alternative hypotheses close to the null hypothesis is as follows: the reference distribution for exponential observations has heavier tails than the normal ($\alpha_4 = 3.75$ for $n = 4$, and the normal value is $\alpha_4 = 3$) while the randomization reference distribution has light tails ($\alpha_4 < 3$) since its range is always finite. As discussed in Chapter IV, a distribution with a high value of α_4 is more peaked and has heavier tails than a distribution with smaller α_4. Small α_4 values indicate light tails and heavy shoulders. Now an alternative hypothesis close to the null hypothesis is more easily detected by the distribution with heavy shoulders, in this case the randomization

reference distribution, since many observations are close to the mean. On the other hand, alternative hypotheses in the tails are more easily detected by the parametric test due to its reference distribution having more observations in the tails.

Note that the power of the normal theory test is very similar to the parametric test. While not appropriate for exponential observations, the nominal value of α is not far from the true value, as indicated in the columns under $E\{d_i/\sigma\} = 0$. The power of the normal test is similar to the power of the parametric test for all alternative hypotheses and sample sizes, although the greatest difference is for small sample sizes where the parametric reference distribution differs the most from the normal distribution.

Note also that as the sample size increases the results of all three tests converge to the same values. This must be true for the parametric and normal tests due to the central limit theorem. This is equivalent to the reference distributions becoming the same in Figure 16 for large sample sizes. The component randomization test also converges to the same power for large sample sizes, since this test is asymptotically the normal theory test for large sample sizes. However, the nonnormality in terms of sample size is of interest here.

Figure 17 shows graphically the values of Table 7 for the parametric and component randomization tests. The normal test is not shown since it is only approximately correct for this case and since the results are so similar to the parametric test. All of the same general relationships between α, n, and $E\{d_i/\sigma\}$ are true here as for the normal case. β decreases as any of the three factors increase. The striking

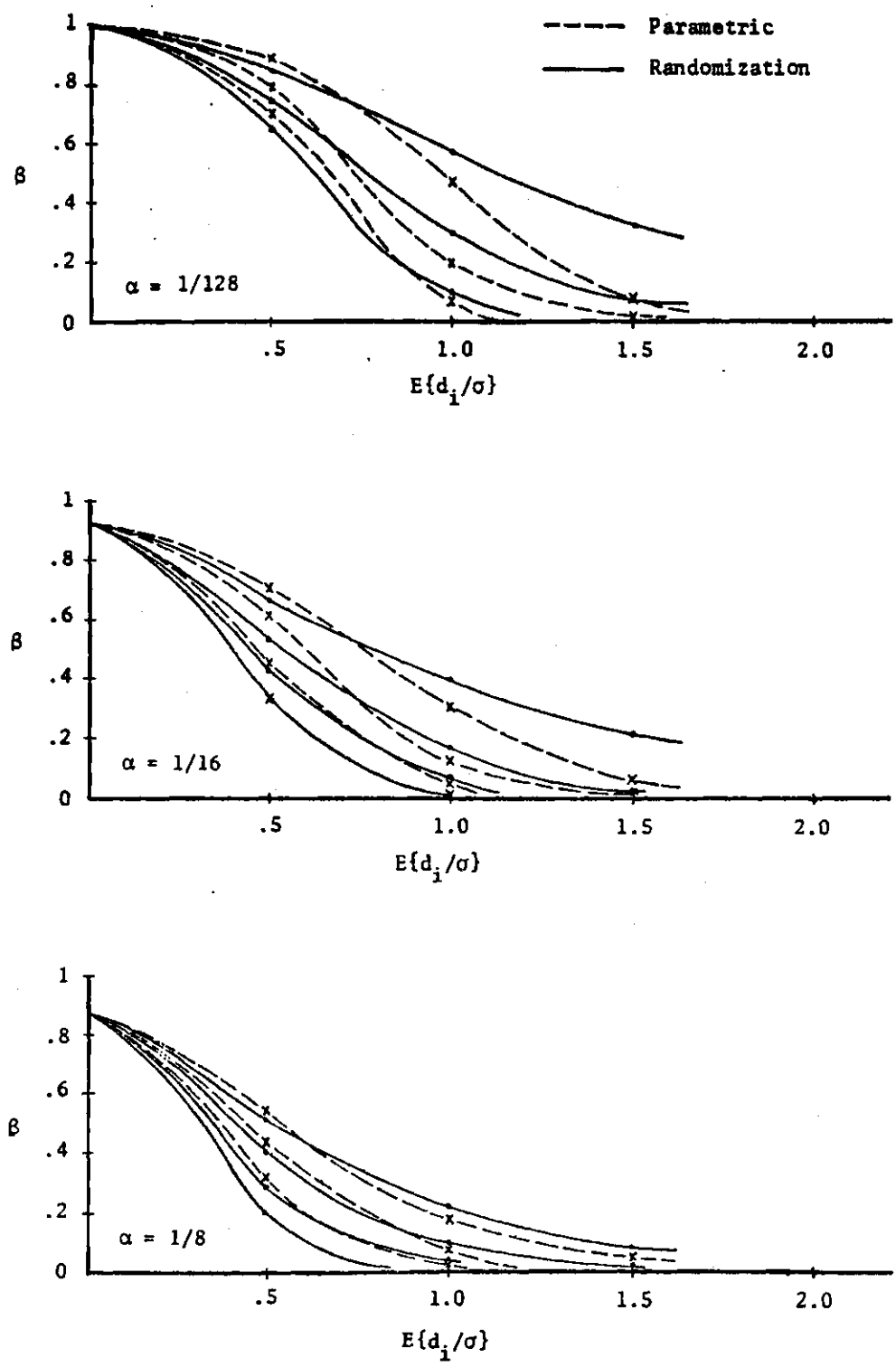

Figure 17. Operating Characteristic Curves from Exponential Observations

difference is that the parametric test does not dominate the component randomization test, as just discussed.

It is of interest to compare the results of this section to those of the last section for normal observations. The power of the component randomization test is generally greater for exponential observations than for normal observations, the greatest difference being .1 for $n = 4$, $\alpha = 1/16$, and $E\{d_i/\sigma\} = .5$. However, for large values of $E\{d_i/\sigma\}$ and/or for large n the normal observations lead to the greater power. In these cases the power is so great for both tests, however, that the difference is not usually important. Thus it appears that the component randomization test actually has better overall power for exponential observations than for normal observations.

This relationship does not hold true for the parametric test. For small α, the parametric test performs best under normal observations. For $\alpha = 1/16$ the test has similar power for both types of observations and for large α values the parametric test performs best for exponential observations. That this should be true is obvious from Figure 16. The reference distributions all cross the normal distribution around the .95 quantile. Since power of the parametric test depends on this distribution only through the $(1-\alpha)$ quantile, power should indeed be similar for $\alpha = 1/16$.

Note that the result of all the parametric reference distributions having the same .95 quantile is that the normal theory test is very close to exactly valid for exponential observations for any sample size if $\alpha \approx .05$. Smaller or larger values of α lead to inaccuracies in the normal test.

7.3.3 Uniform Observations

Results corresponding to those of the last section are given here for uniformly distributed observations. Figure 18 shows the reference distributions for sample sizes n = 4, 7, 11, 15, and infinity arising from uniform observations. Again the straight line, the normal distribution, is the limiting distribution as n approaches infinity. Not shown is the limiting case of n = 1, the distribution of the difference between two uniformly distributed random variables, which is the triangular distribution. All of these reference distributions have a finite upper bound and therefore lie above the normal line, whereas the exponential reference distributions were below the normal distribution. Note that, for the α values considered, the reference distribution is essentially identical with the normal. Only for α greater than .01 is the difference noticeable in Figure 18 and even there it is not as great as in the case of exponential observations.

Table 8 shows the probability of type II error for various sample sizes and alternative hypotheses $E\{d_i/\sigma\}$ for the component randomization and parametric tests. From the table it would appear that the normal test works well in this case as an approximation to the true parametric test, even for small sample sizes.

The results for uniform observations are similar to those of the prior two cases, with β decreasing as a function of n, α, and $E\{d_i/\sigma\}$. Again of interest is that the component randomization test has greater power (lesser β) for $E\{d_i/\sigma\} = .5$ and $\alpha = 1/16$ and $\alpha = 1/8$ than the parametric test. That is, even though the parametric test has better power for most combinations of n, α, and $E\{d_i/\sigma\}$, it does not dominate

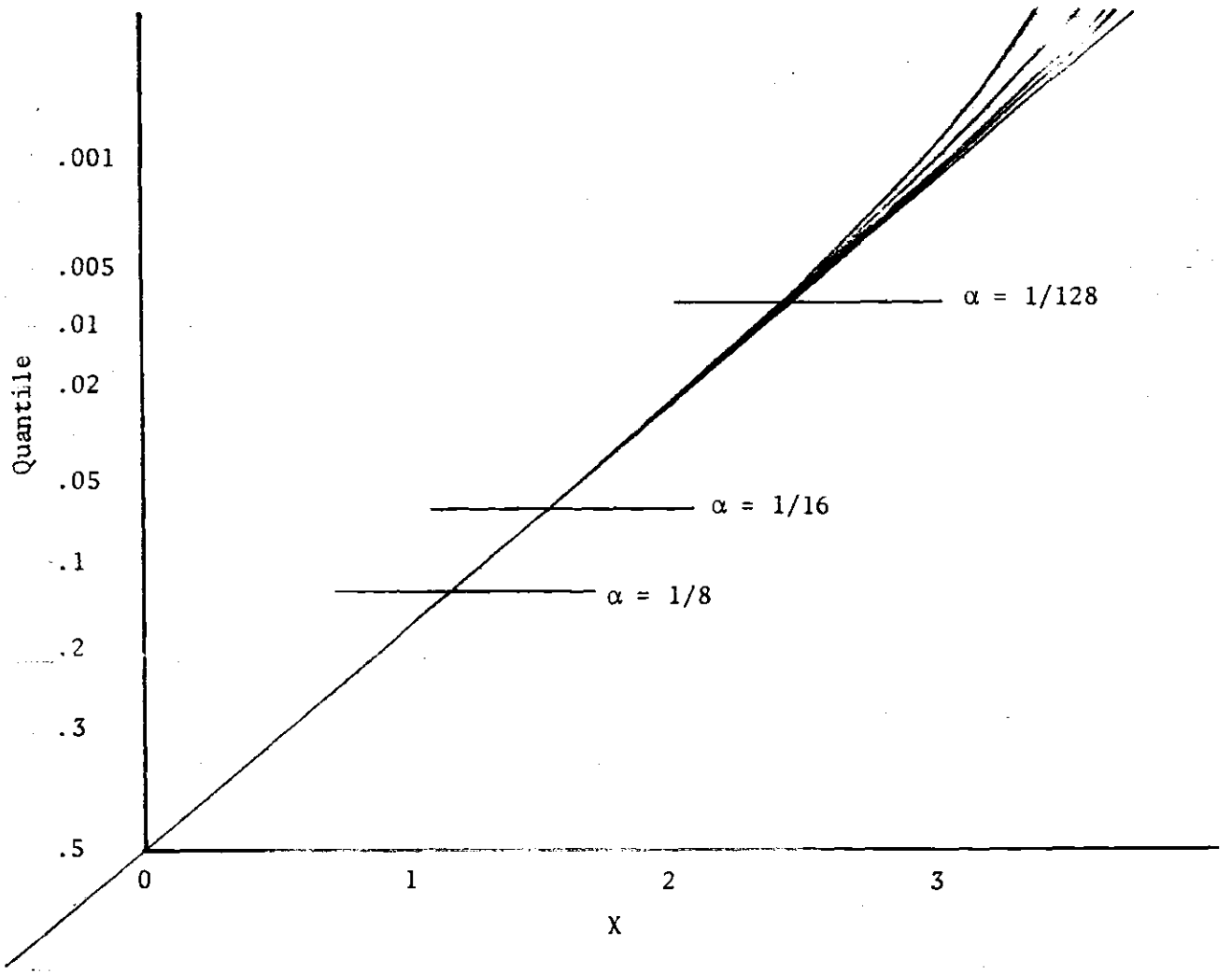

Figure 18. Reference Distributions Arising From Uniform Observations

Table 9. Type II Error (β) for Randomization (R), Parametric (P), and Normal (N) Tests Arising from Uniform Observations

n \ δ/σ	0			.5			1.0			1.5			C_α
	R	P	N	R	P	N	R	P	N	R	P	N	
	X,Y ~ Uniform $\alpha = 1/128 = .0078125$ $z_\alpha = 2.417$												
7	$\frac{127}{128}$	$\frac{127}{128}$	.99	.93 (.004)	.85	.85	.74 (.008)	.40	.41	.42 (.009)	.06	.06	2.39
11	$\frac{127}{128}$	$\frac{127}{128}$	.99	.87 (.01)	.80 (.01)	.80	.40 (.015)	.19 (.012)	.19	.03 (.005)	.004 (.002)	.004	2.40
15	$\frac{127}{128}$	$\frac{127}{128}$	.99	.81 (.03)	.67 (.04)	.67	neg			neg			2.41
	$\alpha = 1/16 = .0625$ $z_\alpha = 1.534$												
4	.9375	.9375	.94	.68 (.006)	.70	.70	.54 (.007)	.32 (.007)	.32	.27 (.006)	.07 (.003)	.07	1.53
7	.9375	.9375	.96	.65 (.009)	.57	.58	.21 (.007)	.14	.14	.02 (.002)	.005	.006	1.53
11	.9375	.9375	.96	.50 (.02)	.45 (.02)	.45	.05 (.006)	.04 (.005)		neg			1.53
15	.9375	.9375	.94	.34 (.04)	.29 (.04)	.29	neg			neg			1.53
	$\alpha = 1/8 = .125$ $z_\alpha = 1.150$												
4	.875	.875	.87	.52 (.006)	.56	.56	.30 (.006)	.19 (.006)	.19	.08 (.004)	.03 (.003)	.03	1.15
7	.875	.875	.87	.47 (.009)	.43	.43	.09 (.005)	.07	.07	neg	neg	neg	1.15
11	.875	.875	.87	.32 (.01)	.30 (.01)	.30	.01 (.004)	.01 (.004)		neg			1.15
15	.875	.875	.87	.19 (.03)	.18 (.03)	.18	neg			neg			1.15

the component randomization test.

It is apparent by comparing Tables 6 and 8 that the normal test does indeed approximate the properties of the parametric test well in this case. The power never differs by more than .01 except for large values of $E\{d_i/\sigma\}$ where the difference in tail weights between the uniform and exponential distributions has an effect. If, however, an value of .0001 and a small sample size were used, Figure 18 shows that the normal test is not so good an approximation since $c_{.9999}$, equal to 3.35 for $n = 4$, does differ from the normal value $z_{.9999} = 3.45$, resulting in a biased indication of the true α value.

7.3.4 Absolute Lambda Observations

In the previous three sections the power of the component randomization test and the corresponding parametric test has been examined for normal, exponential, and uniform observations. These three distributions were selected as three distributions representing a wide range of tail weights. The uniform has the lightest tails ($\beta_2 = 1.8$), the normal has medium tails ($\beta_2 = 3$), and the exponential distribution has heavy tails ($\beta_2 = 9$). A question of interest is whether tail weight, as measured by the fourth standardized moment, is really the central factor in determining the effect of underlying distribution on the power of the tests.

An analysis similar to the last three sections is performed here to gain insight into the fourth moment's impact on power. In particular, observations are generated from the absolute lambda distribution (discussed in Chapter IV) with moments one through four identical to the standardized normal distribution. Despite having the same skewness

and kurtosis, this density function has a shape quite unlike the normal, as shown in Figure 8.

Figure 19 shows the reference distribution arising from the absolute lambda distribution with parameters $\lambda_1 = 0$, $\lambda_2 = -1.575$, $\lambda_3 = -.2247$, and $\lambda_4 = .5$ for a sample size of four. The reference distributions for larger sample sizes are not shown since in the figure they would not be distinguishable. Also included, for comparative purposes, is the normal distribution which appears as a straight line. While the difference between the reference distribution shown and the normal is slight, compared to previous examples, it is significant. The major difference between this and previous examples is that the nonnormality occurs in the body, rather than the tail, of the distribution.

Table 10 shows the results of the Monte Carlo determination of the power of the component randomization and the parametric tests for this underlying distribution. The results for the normal test and the parametric test are identical to two places of accuracy, and therefore the results are given only for the parametric test. For the cases studied, the power of the test does not differ noticeably from the results for normally distributed observations. Note, however, that the power of the component randomization test is somewhat less for these observations than for normal observations. The discontinuity of the underlying distribution seems to affect the component randomization test while not affecting the parametric test.

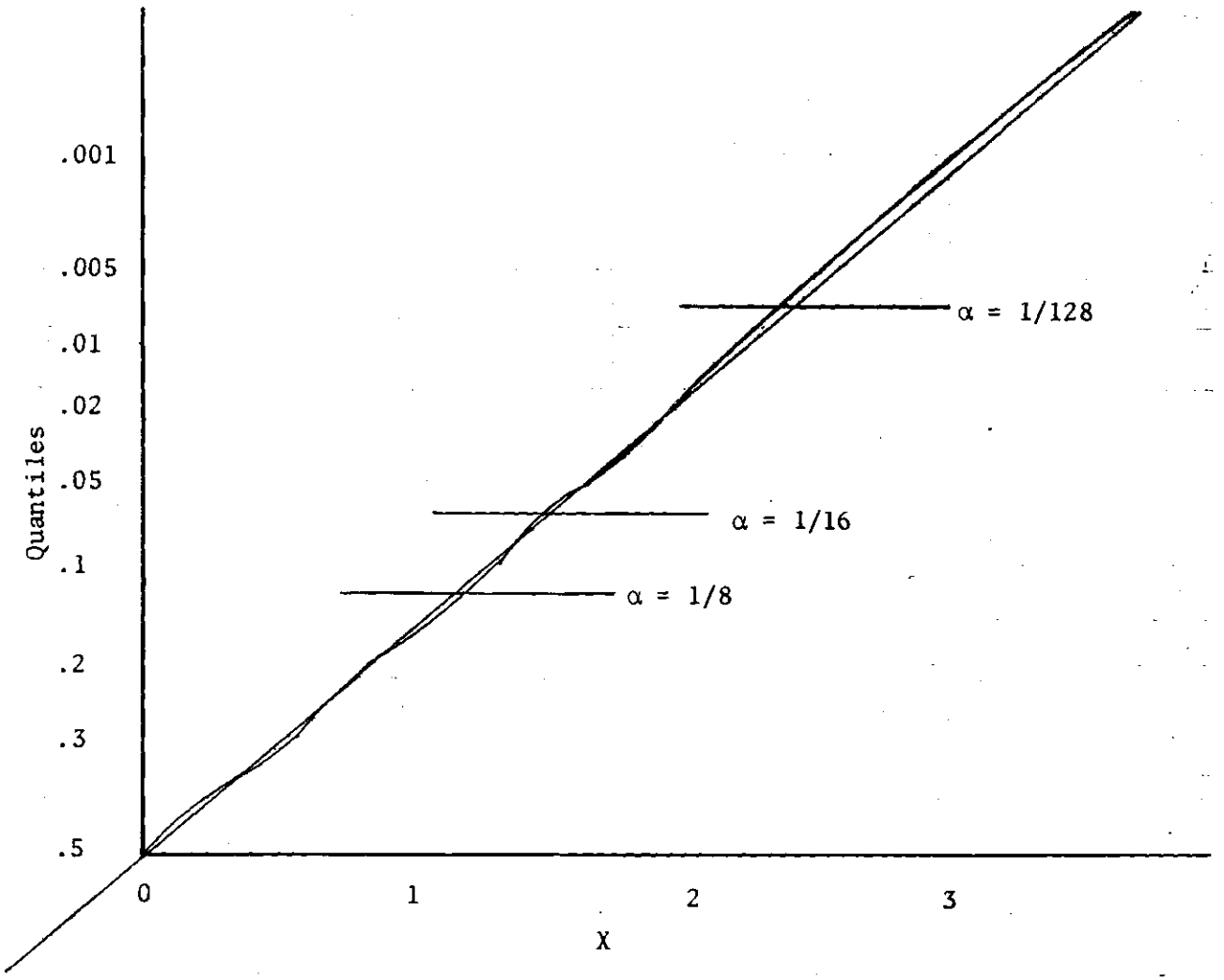

Figure 19. The Reference Distribution Arising from Absolute Lambda Observations for n = 4

Table 10. Type II Error (β) for Randomization (R), Parametric (P), and Normal (N) Tests Arising from Absolute Lambda Observations (0, 2.227, -.224745, .5)

δ/σ / n	.5		1.0		1.5	
	R	P	R	P	R	P
$\alpha = 1/128 = .0078125$				$z_\alpha = 2.417$		
7	.89	.87	.87	.42	.30	.06
11	.87	.78	.36	.19	neg	
15	.72 (.036)	.68	.14	.08 (.028)	neg	
$\alpha = 1/16 = .0625$				$z_\alpha = 1.534$		
4	.71	.67	.68	.32	.18	.06
7	.66	.59	.21	.14	.013	.007
11	.52	.47	.04	.03	neg	
15	.37 (.039)	.34	.013 (.009)	.013	neg	
$\alpha = 1/8 = .125$				$z_\alpha = 1.150$		
4	.66	.58	.29	.20	.06	.03
7	.49	.45	.09	.07	.003	.001
11	.33	.32	.01	.01	neg	
15	.29 (.035)	.23	.006 (.006)	0	neg	

CHAPTER VIII

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This research is directed toward studying the power of tests of location in the paired samples design, but also involves several aspects of Monte Carlo distribution sampling with wider applicability. For example, the quantile estimation procedures discussed in Chapter III are not limited to the study of power or even to data generated by Monte Carlo studies. In addition, the process generation, the topic of Chapter IV, is applicable in either systems simulation of Monte Carlo studies. However, Chapters V, VI, and VII, are applicable only to component randomization and as a set consider the component randomization test, the reference distribution of the corresponding parametric test, and the power of these tests. A more complete summary of the results of the research is given in Section 8.1, followed by conclusions in Section 8.2, and recommendations for future research in Section 8.3.

8.1 Summary of Results

A summary of the results of this research is given here. The results are listed in the same order as presented in the body of this work. The section upon which the result is based is given in parentheses.

8.1.1 Chapter III. Estimation of the pth Quantile

1. Given grouped data,

$$\hat{F^{-1}}(p) = a + b[q - (\sum_{i=1}^{q} c_i - p(k+1) + 1)/(c_q + 1)]$$

is recommended as an improvement over the midpoint of the qth cell as an estimator of the pth quantile of the distribution of (3.1.1) the cdf.

2. The expected value and variance of this estimator is derived in terms of P(q), the probability of the estimator falling in cell q (3.1.2).
3. P(q) is derived as a function of the probability of an arbitrary observation falling in cell i, i = 1,2,...,Q (3.1.3 and Appendix A).
4. The impact of cell width and sample size is analyzed. It is shown why the recommended estimator is less biased and provides a more accurate estimate of the variability (3.1.4).
5. For raw data,

$$F^{-1}(p) = (1-\alpha)X_{(r)} + \alpha X_{(r+1)}$$

where $r = [p(k+1)]$ and $\alpha = p(k+1) - r$

is recommended for estimating the pth quantile (3.2.1).

6. This estimator is shown to be unbiased assuming a uniform approximation in the region of the estimate (but not necessarily a uniform distribution of the observations) (3.2.1).
7. The effect of sample size on the performance of the quantile estimator is studied. Rules are given to determine the sample size necessary for the use of a simple order statistic to provide a desired accuracy (3.2.2).

8.1.2 Chapter IV. Process Generation from (β_1,β_2)

8. Criteria for comparing general process generators are proposed (4.2).
9. The absolute lambda distribution is presented, including distributional properties and its use in generating random values having any desired skewness and kurtosis (4.4).

8.1.3 Chapter V. The Randomization Test

10. A basic method and improvements are given to test the paired samples location hypothesis in the paired samples design using the component randomization test. The basic method allows generation of signs without tabled values through the use of modular arithmetic, while the improvements increase the computational speed by making the modular arithmetic implicit. A FORTRAN program to implement the method is given (5.1).
11. An approach for estimating the power of the component randomization test (using the last result) is developed, including the variance of the estimate and minimum sample size necessary for a desired accuracy (5.2).

8.1.4 Chapter VI. The Corresponding Parametric Test

12. The standardized moments of the differences are given as a function of the moments of the observations (6.1.1).
13. The standardized moments of the sum of differences are given as a function of the moments of the differences (6.1.2).
14. The inadequacy of asymptotic expansions and Chebyshev-type inequalities as general solution procedure is illustrated (6.2).

15. The use of the randomization sample, the 2^n dependent components arising from each sample, as a variance reduction technique for quantile estimation is proposed and shown to be valid (6.3).
16. The decrease in computational effort arising from this technique is quantified (6.3.5).

8.15. Chapter VII. The Power of Paired Sample Tests

17. The power of the component randomization test and the parametric test is given in the form of operating characteristic curves for normal, exponential, uniform, and absolute lambda observations (7.3).

8.2 Conclusions

Conclusions arising from the overall research are as follows:

1. The recommended estimators of the pth quantile reduce bias and provide a more accurate variance estimate than the common approach of using cell midpoints. The effect is greatest for small samples with raw data and large samples for grouped data.
2. Based on the criteria of Section 4.2, the Burr distribution, the generalized lambda distribution, and the absolute lambda distribution each have advantages not possessed by the others. The absolute lambda distribution, presented in Section 4.4, is the only technique known which will provide random variates having any desired skewness and kurtosis using only two parameters and one functional form.

3. The algorithm proposed for performing the component randomization test makes feasible the use of this test for samples as large as n = 25.
4. A useful variance reduction technique, using component randomization, for finding the distribution of a sum of symmetric random variables can reduce computational effort as much as 99.9%.
5. The power of the component randomization test and the parametric test may be studied using Monte Carlo techniques in a reasonable amount of computational effort by using the techniques of Chapters III, V, VI, and VII. In addition, the results of Chapter IV make the selection of underlying distribution simpler.
6. The power of the component randomization test is usually less than the power of the appropriate parametric test, but is <u>not</u> dominated for all alternative hypotheses. That is, in some cases, the distribution-free randomization test has more power than the parametric test which assumes knowledge of the exact distribution of the observations, including variance.
7. The robustness of the normal test of location, a well-known property, is observed in relation to the performance of the correct parametric test. The difference in reference distributions is seen to be greatest in the tails of the distributions for common distributions.
8. The power of the component randomization test is different

for normal and absolute lambda observations, even though both have identical first four moments. The power of the parametric test, on the other hand, is essentially the same for both types of observations.

9. Since in practice the appropriate parametric test is not known, the component randomization test is a viable alternative to the normal test for samples no larger than 25. Especially when the variance is unknown the component randomization test costs little in terms of power loss.

8.3 Recommendations for Future Research

The avenues for future research based on this work are perceived to be:

1. A distribution for general process generation satisfying all the criteria of Section 4.2 is needed.
2. The component randomization test in the paired samples design may be made still more computationally efficient by perhaps considering techniques which do not explicitly consider each of the 2^n components of the sample.
3. The method of generating all 2^n combinations of signs may have application to other areas. For example, the two sample component randomization test requires similar logic. Another area of application is seen by noting that the 2^n combinations of signs is an exact analogue of explicitly enumerating the solutions of an n item knapsack problem.
4. The procedures of Chapter VII, based on Chapters V and VI,

may be used to study the power of the tests for other specific underlying distributions and sample sizes.

APPENDIX A

VALIDITY OF THE DISTRIBUTION OF THE pTH QUANTILE ESTIMATOR CELL

APPENDIX A

VALIDITY OF THE DISTRIBUTION OF THE pTH QUANTILE ESTIMATOR CELL

In Chapter III the distribution of P(q) is derived. The proof that $\Sigma_{i=1}^{\infty} P(q) = 1$ given here uses mathematical induction on [p(k+1)]. To establish the initial condition, let $k' \equiv k+1 = [p(k+1)]$ using the notation of Chapter III.

$$\sum_{q=1}^{Q} P(q) = \sum_{q=1}^{Q} \left\{ \sum_{j=0}^{k'-1} \left[\sum_{i=r-j}^{k'-j} \binom{k'}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k'-j-i} \right] \right\}$$

$$= \sum_{q=1}^{Q} \left\{ \sum_{j=0}^{k'-1} \binom{k'}{j} (p_q)^{k'-j} \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \right\}$$

$$= \sum_{q=1}^{Q} \left\{ \left(p_q + \sum_{\ell=1}^{q-1} p_\ell \right)^{k'} - \left(\sum_{\ell=1}^{q-1} p_\ell \right)^{k'} \right\} \quad \text{since} \quad \sum_{j=0}^{k'} \binom{k'}{j} P^{k'-j} S^j = (P+S)^k$$

$$= \sum_{q=1}^{Q} \left(\sum_{\ell=1}^{q} p_\ell \right)^{k'} - \sum_{q=1}^{Q} \left(\sum_{\ell=1}^{q-1} p_\ell \right)^{k'}$$

$$= \left[p_1^{k'} + (p_1+p_2)^{k'} + \ldots + \left(\sum_{\ell=1}^{Q} p_\ell \right)^{k'} \right] - \left[0 + p_1^{k'} + (p_1+p_2)^{k'} + \ldots + \left(\sum_{\ell=1}^{Q-1} p_\ell \right)^{k'} \right]$$

$$= \left(\sum_{\ell=1}^{Q} p_\ell \right)^{k'} = 1^{k'} = 1$$

Now assume $\sum_{q=1}^{Q} P(q) = 1$ if $[pk] = m$. That is

$$\sum_{q=1}^{Q} P(q) = \sum_{q=1}^{Q} \left\{ \sum_{j=0}^{m-1} \left[\sum_{i=m-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right] \right\} = 1 \qquad (8)$$

Let $[pk] = m-1$. Then if

$$\sum_{q=1}^{Q} P(q) = \sum_{q=1}^{Q} \left\{ \sum_{j=0}^{m-2} \left[\sum_{i=m-1-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right] \right\} = 1 \qquad (9)$$

the proof is complete. This may be done by showing that the difference between the center expressions of eqs. 8 and 9 is zero as follows:

$$\sum_{q=1}^{Q} \left\{ \sum_{j=0}^{m-1} \left[\sum_{i=m-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right] \right.$$

$$\left. - \sum_{j=0}^{m-2} \left[\sum_{i=m-1-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right] \right\}$$

$$= \sum_{q=1}^{Q} \left\{ \sum_{i=m-(m-1)}^{k-m+1} \binom{k}{i\ (m-1)} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^{m-1} \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-(m-1)-i} \right.$$

$$+ \sum_{j=0}^{m-2} \left[\sum_{i=m-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right]$$

$$\left. - \sum_{i=m-1-j}^{k-j} \binom{k}{ij} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-j-i} \right] \right\}$$

$$= \sum_{q=1}^{Q} \left\{ \sum_{i=1}^{k-m+1} \binom{k}{i\ (m-1)} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell \right)^{m-1} \left(1 - \sum_{\ell=1}^{q} p_\ell \right)^{k-m-i+1} \right.$$

$$+ \sum_{j=0}^{m-2}\left[- \binom{k}{(m-1-j)} j \; p_q^{n-1-j} \left(\sum_{\ell=1}^{q-1} p_\ell\right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-j-(m-1-j)}\right]\Bigg\}$$

$$= \sum_{q=1}^{Q}\Bigg\{\sum_{i=1}^{k-m+1} \binom{k}{i\ (m-1)} p_q^i \left(\sum_{\ell=1}^{q-1} p_\ell\right)^{m-1}\left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m-i+1}$$

$$- \sum_{j=0}^{m-2} \binom{k}{(m-1-j)j} p_q^{m-1-j} \left(\sum_{\ell=1}^{q-1} p_\ell\right)^j \left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1}\Bigg\}$$

$$= \sum_{q=1}^{Q}\Bigg\{\left(\sum_{\ell=1}^{q-1} p_\ell\right)^{m-1} \frac{k(k-1)(k-2)\ldots(k-m+2)}{(m-1)!} \sum_{i=1}^{k-m+1} \left(\frac{(k-m+1)!}{i!(k-m+1-i)!}\right) p_q^i \cdot$$

$$\left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1-i}$$

$$- \left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1} \frac{k(k-1)\ldots m}{(k-m+1)!} \sum_{j=0}^{m-2} \binom{m-1}{j} p_q^{m-1-j} \left(\sum_{\ell=1}^{q-1} p_\ell\right)^j\Bigg\}$$

$$= \frac{k(k-1)\ldots(k-m+2)}{(m-1)!} \sum_{q=1}^{Q}\Bigg\{\left(\sum_{\ell=1}^{q-1} p_\ell\right)^{m-1}\left[\left(1 - \sum_{\ell=1}^{q-1} p_\ell\right)^{k-m+1} - \left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1}\right]$$

$$- \left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1}\left[\left(\sum_{\ell=1}^{q} p_\ell\right)^{m-1} - \left(\sum_{\ell=1}^{q-1} p_\ell\right)^{m-1}\right]\Bigg\}$$

$$= \frac{k(k-1)\ldots(k-m+2)}{(m-1)!} \sum_{q=1}^{Q}\Bigg\{\left(\sum_{\ell=1}^{q-1} p_\ell\right)^{m-1}\left(1 - \sum_{\ell=1}^{q-1} p_\ell\right)^{k-m+1}$$

$$- \left(\sum_{\ell=1}^{q} p_\ell\right)^{m-1}\left(1 - \sum_{\ell=1}^{q} p_\ell\right)^{k-m+1}\Bigg\}$$

$$= \frac{k(k-1)\ldots(k-m+2)}{(m-1)!}\left\{\left[0 - p_1^{m-1}\ (1-p_1)^{k-m+1}\right]\right.$$

$$+ \left[p_1^{m-1}\ (1-p_1)^{k-m+1} - (p_1+p_2)^{m-1}\ (1-p_1-p_2)^{k-m+1}\right]$$

$$\left. + \ldots + \left[\left(\sum_{\ell=1}^{Q-1} p_\ell\right)^{m-1}\left(1 - \sum_{\ell=1}^{Q-1} p_\ell\right)^{k-m+1} - \left(\sum_{\ell=1}^{Q} p_\ell\right)^{m-1}(0)^{k-m+1}\right]\right\}$$

$$= \frac{k(k-1)\ldots(k-m+2)}{(m-1)!}\{0\} = 0 \qquad \text{Q.E.D.}$$

APPENDIX B

PROGRAM LISTING FOR CHAPTER V

APPENDIX B

PROGRAM LISTING FOR CHAPTER V

```
C       TO PERFORM MONTE CARLO ESTIMATION OF THE POWER OF THE
C       RANDOMIZATION TEST
      DIMENSION ITWO(50), D(50), ALPHA(20),ICRIT(20),C(20),
     *  Z(20),T(20),IRC(20),IPC(20),INC(20),ITC(20)
C
C*****ENTER PARAMETERS AND INITIALIZE
C
   15 WRITE (6,1)
    1 FORMAT (' ENTER SAMPLE SIZE N, EXPECTED DIFFERENCE, CORRELATION,'
     * / ' # OF SAMPLES, RANDOM # SEED, DISTRIBUTION TYPE,'/
     * ' AND FOUR PARAMETERS. (I,X,X,I,I,I,X,X,X,X)')
      READ (5,2,END=99) N,XMU,RHO,NBIG,ISEED,IDIST,P1,P2,P3,P4
    2 FORMAT ()
      NTWO = 2**N
      NA = 0
   11 NA = NA + 1
      WRITE (6,3)
    3 FORMAT (' ENTER ALPHA, C, Z, AND T')
      READ (5,2,END=12) ALPHA(NA), C(NA), Z(NA), T(NA)
      IRC(NA) = 0
      IPC(NA) = 0
      INC(NA) = 0
      ITC(NA) = 0
      ICRIT(NA) = (1-ALPHA(NA)) * NTWO
      GO TO 11
   12 NA = NA - 1
      NTOTAL = 0
      FRHO = SQRT (1-RHO*RHO)
      SUMD2 = 0
      DO 50 I=1,N
   50 ITWO(I) = 2**(N-I)
C
C*****PERFORM TEST NBIG TIMES.
C*****FIRST GENERATE RANDOM DIFFERENCES AND CALCULATE TEST STATISTIC
C
   25 NTOTAL = NTOTAL + NBIG
      TOTAL = NTOTAL
      DO 500 K=1,NBIG
      STAT = 0
      TSUMD2 = 0
      DO 600 I=1,N
```

```
      CALL DATAIN (IDIST,P1,P2,P3,P4,ISEED,X)
      CALL DATAIN (IDIST,P1,P2,P3,P4,ISEED,Y)
      Y = Y * FRHO + X * RHO
      D(I) = X - Y + XMU
      TSUMD2 = TSUMD2 + D(I) * D(I)
  600 STAT = STAT + D(I)
      SUMD2 = SUMD2 + TSUMD2
C
C*****PERFORM THREE PARAMETRIC TESTS (IDEAL, NORMAL, AND T)
C
      S2 = (TSUMD2 - STAT*STAT/N) / (N-1)
      TSTAT = STAT / SQRT (S2*N)
      PSTAT = STAT / SQRT(N)
      ZSTAT = STAT / SQRT(N)
      DO 60 I=1,NA
      IF (PSTAT .LT. C(I))  IPC(I) = IPC(I) + 1
      IF (ZSTAT .LT. Z(I))  INC(I) = INC(I) + 1
      IF (TSTAT .LT. T(I))  ITC(I) = ITC(I) + 1
   60 CONTINUE
C
C*****INITIALIZE FOR PERFORMING TEST
C
      KOUNT = 0
      D22 = D(2) + D(2)
      D32 = D(3) + D(3)
      D23 = D22 + D32
      DSUM = -D(1) - D(2) - D(3)
      M = ITWO(3)
      IF (N .GT. 3)  GO TO 200
      J = NTWO
      SUM = DSUM
      GO TO (10, 20, 30), N
  200 J = 0
C
C*****DETERMINE THE J TH COMBINATION OF SIGNS FOR COMPONENTS I=4,...,N
C
 1000 J = J + 1
      SUM = DSUM
      IEVEN2 = M
      ITERM = J
      DO 400 I=4,N
      IEVEN = IEVEN2
      IEVEN2 = ITWO(I)
      IF (IEVEN .LT. ITERM)  ITERM = ITERM - IEVEN
      IF (ITERM .GT. IEVEN2)  GO TO 300
      SUM = SUM - D(I)
      GO TO 400
  300 SUM = SUM + D(I)
  400 CONTINUE
C
C*****FOR THIS COMBINATION OF SIGNS, CONSIDER ALL A POSSIBILITIES
C
```

```
   30 X = SUM + D32
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
      X = SUM + D23
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
   20 X = SUM + D22
      IF (X .LE. STAT)  KOUNT = KOUNT + 1
      IF (-X .LE. STAT)  KOUNT = KOUNT + 1
   10 IF (SUM .LE. STAT)  KOUNT = KOUNT + 1
      IF (-SUM .LE. STAT)  KOUNT = KOUNT + 1
      IF (J .LT. M)  GO TO 1000
C
C*****UPDATE ACCUMULATORS REFLECTING RESULTS OF THIS TEST
C
      DO 500 I=1,NA
      IF (KOUNT .LE. ICRIT(I))  IRC(I) = IRC(I) + 1
  500 CONTINUE
C
C*****DETERMINE AND PRINT RESULTS FOR DESIRED VALUES OF ALPHA
C
      VAR = SUMD2/(NTOTAL*N) -XMU*XMU
      WRITE (6,6) VAR
    6 FORMAT (' VARIANCE OF THE DIFFERENCES =', F10.4)
      DO 900 I=1,NA
      BR = IRC(I) / TOTAL
      SBR = SQRT (BR*(1-BR)/TOTAL)
      BC = IPC(I) / TOTAL
      SBC = SQRT (BC*(1-BC)/TOTAL)
      BN = INC(I) / TOTAL
      SBN = SQRT (BN*(1-BN)/TOTAL)
      BT = ITC(I) / TOTAL
      SBT = SQRT (BT*(1-BT)/TOTAL)
      CC = ICRIT(I)
  900 WRITE (6,4) ALPHA(I),CC,C(I),Z(I),T(I),BR,BC,BN,BT,
     *            SBR,SBC,SBN,SBT
    4 FORMAT (' ALPHA =', F10.6/
     * '          RANDOMIZATION    PARA    NORMAL          T'/
     *' C. VALUES',4F10.4/' BETAS     ',4F10.6/' STD BETAS',4F10.6/)
C
C*****ENTER NUMBER OF ADDITIONAL SAMPLES TO BE TESTED
C
      WRITE (6,5)
    5 FORMAT (' HOW MANY MORE SAMPLES? (I)')
      READ (5,2,END=15) NBIG
      IF (NBIG) 99,15,25
C
   99 STOP
      END
```

```
C       PROCESS GENERATOR FOR VARIOUS DISTRIBUTIONS
      SUBROUTINE DATAIN (IDIST,P1,P2,P3,P4,ISEED,X)
      GO TO (1,2,3,4), IDIST
C
C*****IDIST = 1 IMPLIES THE GENERALIZED LAMBDA DISTRIBUTION
C
    1 P = RANDOM(ISEED)
      X = P1 + (P**P3 - (1-P)**P4) / P2
      RETURN
C
C*****IDIST = 2 IMPLIES NORMAL (P1,STD=P2)
C
    2 PA = RANDOM(ISEED)
      PB = RANDOM(ISEED)
      X = (-2*ALOG(PA))**.5 * COS(6.28318*PB)
      X = P1 + X*P2
      RETURN
C
C*****IDIST = 3 IMPLIES THE ABSOLUTE LAMBDA DISTRIBUTION
C
    3 P = RANDOM(ISEED)
      IF (P .LT. P4)  X = - (P4-P)**P3
      IF (P .GE. P4)  X = (P-P4)**P3
      X = P1 + X/P2
      RETURN
C
C*****IDIST = 4 IMPLIES THE EXPONENTIAL DISTRIBUTION
C
    4 P = RANDOM(ISEED)
      X = -ALOG(P) * P1
      RETURN
      END
```

APPENDIX C

RESULTS FOR SECTION 6.3.5

APPENDIX C

RESULTS FOR SECTION 6.3.5

In Section 6.3.5 Cov $\{D_j^2, D_k^2\}$, where D_j and D_k have ℓ conflicting signs, was used in the derivation of the variance of the randomization sample estimator of σ^2. The value of this covariance is now derived.

$$\text{Cov}\{D_j^2, D_k^2\} = \text{Cov}\left\{\left(\sum_{i=1}^{n-\ell} d_i - \sum_{i=n-\ell+1}^{n} d_i\right)^2, \left(\sum_{i=1}^{n} d_i\right)^2\right\}$$

$$= E\left\{\left[\left(\sum_{i=1}^{n-\ell} d_i\right)^2 - 2\left(\sum_{i=1}^{n-\ell} d_i\right)\left(\sum_{i=n-\ell+1}^{n} d_i\right) + \left(\sum_{i=n-\ell+1}^{n} d_i\right)^2\right]\left(\sum_{i=1}^{n} d_i\right)^2\right\} - E\{D_j^2\}\, E\{D_k^2\}$$

$$= E\left\{\left[\left(\sum_{i=1}^{n-\ell} d_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n-\ell}\sum_{j=1}^{n-\ell} d_i d_j\right) - 2\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j + \left(\sum_{i=n-\ell+1}^{n} d_i^2 + \sum_{i=n-\ell+1}^{n}\sum_{j=n-\ell+1}^{n} d_i d_j\right)\right]\left[\sum_{i=1}^{n} d_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j\right]\right\} - \sigma^4$$

$$= E\left\{\left[\sum_{i=1}^{n} d_i^2 - 2\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j\right.\right.$$

$$\left.\left. + \left(\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j - 2\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j\right)\right]\right.$$

$$\left.\left[\sum_{i=1}^{n} d_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j\right]\right\} - \sigma^4$$

$$= E\left\{\left[\sum_{i=1}^{n} d_i^2 - 4\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j\right]\right.$$

$$\left.\left[\sum_{i=1}^{n} d_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j\right]\right\} - \sigma^4$$

$$= E\left\{\left[D_k^2 - 4\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j\right]\left[D_k^2\right]\right\} - \sigma^4$$

which, with sufficient insight, may be written directly.

$$= E\left\{D_k^4 - 4\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} d_i d_j\left(\sum_{i=1}^{n} d_i^2 + \sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{j=1}^{n} d_i d_j\right)\right\} - \sigma^4$$

$$= \mu_4 - 4E\left\{\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n}\sum_{k=1}^{n} d_i d_j d_k^2\right.$$

$$\left. + 2\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n}\sum_{k=1}^{n}\sum_{m=k+1}^{n} d_i d_j d_k d_m\right\} - \sigma^4$$

$$= \mu_4 - 8\left[\sum_{k=1}^{n}\sum_{m=k+1}^{n}\sum_{i=1}^{n-\ell}\sum_{j=n-\ell+1}^{n} E\{d_k d_m d_i d_j\}\right] - \sigma^4$$

since $E\{d_i d_j d_k\} = 0$ for $i \neq j$

$$= \mu_4 - 8\left[\sum_{k=1}^{n-\ell}\sum_{m=n-\ell+1}^{n} E\{d_k^2 d_m^2\}\right] - \sigma^4$$

$$= \mu_4 - 8[(n-\ell)\ell(\sigma^2/n)^2] - \sigma^4$$

$$= \mu_4 - [\frac{8(n-\ell)\ell}{n^2} + 1]\sigma^4$$

thereby showing the result used in Section 6.3.5.

The result

$$\sum_{\ell=1}^{n} \binom{n}{\ell} (n-\ell)\ell = 2^{n-2}\, n(n-1)$$

was also used in Section 6.3.5. The validity of this result is established as follows:

$$\sum_{\ell=1}^{n} \binom{n}{\ell} (n-\ell)\ell = \sum_{\ell=1}^{n-1} \frac{n!}{\ell!(n-\ell)!} (n-\ell)\ell$$

$$= n(n-1) \sum_{\ell=1}^{n-1} \frac{(n-2)!}{(\ell-1)!(n-\ell-1)!}$$

$$= n(n-1) \sum_{k=0}^{n-2} \frac{(n-2)!}{(k!)(n-k)!}$$

$$= n(n-1)2^{n-2}$$

since the summation is the number of all possible combinations of n-2 things.

APPENDIX D

PROGRAM LISTING FOR CHAPTER VI

APPENDIX D

PROGRAM LISTING FOR CHAPTER VI

```
C       MAIN PROGRAM FOR PERFORMING MONTE CARLO SIMULATION
C       TO DETERMINE THE DISTRIBUTION OF (SUM D(I)/SQRT(N))
C
      DIMENSION ICHIST(5000),IHIST(5000),XCMOM(8),XMOM(8),
     *   A(100),B(100),D(100),P(50),AA(8),RA(8),AADIF(8),AMOM(8,2),
     *   AHIST(50,3),ITWO(100)
C
      NCELLS = 5000
      HT1 = -5.
      HW = .002
      WRITE (6,3)
    3 FORMAT('0TO PERFORM MONTE CARLO SIMULATION TO DETERMINE THE '/
     * 'DISTRIBUTION OF (SUM D(I) / SQRT(N))')
  100 WRITE (6,18)
   18 FORMAT ( ' ENTER THE PARENT POPULATION INDICATOR,'/
     * '       FOUR PARAMETER VALUES,'/
     * '       AND THE INITIAL RANDOM NUMBER  (I,X,X,X,X,I)'/)
      READ (5,1,END=99) IDIST,P1,P2,P3,P4,ISEED
    1 FORMAT ()
      IISEED = ISEED
      INITS = ISEED
C
  200 WRITE (6,4)
    4 FORMAT('0ENTER SAMPLE SIZE (NO. OF OBSERVATION DIFFERENCES) (I)'/)
      READ (5,1,END=100) NO
C
C     CLEAR CUMULATIVE MOMENT ACCUMULATORS AND HISTOGRAM
C
      SQRTNO = SQRT(NO)
      DO 300 I=1,NCELLS
  300 ICHIST(I) = 0
      NS = 0
      DO 400 I=1,8
      DO 410 J=1,2
  410 AMOM(I,J) = 0.
  400 XCMOM(I) = 0.
      WRITE (6,35)
   35 FORMAT ('0CUMULATIVE HISTOGRAM AND MOMENTS ARE ZEROED.'/)
      DO 420 I=1,NO
  420 ITWO(I) = 2**(NO-I)
C
```

```
  450 WRITE (6,19)
   19 FORMAT (' TO CHANGE PERCENTILE TO BE EVALUATED,'/' ENTER NEW'
     * ' NUMBER OF PERCENTILES. (I)   ELSE HIT RETURN.')
      READ (5,1,END=200) II
      IF (II .LE. 0)  GO TO 465
      NP = II
       WRITE (6,17) NP
   17 FORMAT ('0ENTER', I3, ' NEW PERCENTILE VALUES  X,X,...')
      READ (5,1,END=200) (P(I),I=1,NP)
  465 DO 460 I=1,NP
      AHIST(I,1) = 0.
      AHIST(I,2) = 0.
  460 AHIST(I,3) = 0.
C
  470 WRITE (6,33)
   33 FORMAT ('0DO YOU WANT TO USE ANTITHETIC VARIABLES?')
      READ(5,11,END=450) I
      IANTIT = -1
      IF (I .EQ. 'YES   ')  IANTIT = 1
C
  500 WRITE (6,5)
    5 FORMAT ('0ENTER # SAMPLES / MONTE CARLO REPLICATION. (I)'/)
      READ (5,1,END=470) NR
      K2NR = NR
      IF (IANTIT .EQ. 1)  K2NR = 2*NR
C 600 WRITE (6,6) K2NR
    6 FORMAT (' ENTER ''YES'' TO GENERATE ANOTHER',I6,' SAMPLES.')
      READ (5,11,END=200) I
   11 FORMAT (A6)
      IF (I .EQ. 'YES   ')  GO TO 690
      WRITE (6,6) K2NR
      READ(5,11,END=200) I
      IF (I .NE. 'YES   ') GO TO 200
C
C     CLEAR HISTOGRAM AND MOMENT ACCUMULATORS FOR THIS REPLICATION
C
  690 DO 700 I=1,NCELLS
  700 IHIST(I) = 0
      DO 800 I=1,8
  800 XMOM(I) = 0.
C
C     GET NR MORE DIFFERENCE SAMPLES, EACH OF SIZE NO
C
  610 NS = NS + 1
      IF (IANTIT .LT. 0)  GO TO 990
      IF (IANTIT .EQ. 0)  GO TO 995
      IANTIT = 0
      IISEED = ISEED
      GO TO 990
  995 IANTIT = 1
      ISEED = IISEED
```

```
  990 CALL STRAT(NR,NU,NO,ISEED,1,P)
      DO 900 K=1,NR
      CALL DATAIN (IDIST,P1,P2,P3,P4,ISEED,A,NO,NS,NR,IANTIT,AA)
      CALL DATAIN (IDIST,P1,P2,P3,P4,ISEED,B,NO,NS,NR,IANTIT,AA)
      DO 1000 I=1,NO
 1000 D(I) = A(I) - B(I)
      IF (NS .GT. 1 .OR. K .GT. 1)  GO TO 1050
C
C     CALCULATE MOMENTS FOR THE DIFFERENCES, AADIF, FROM
C     THE MOMENTS OF THE VALUES RETURNED FROM DATAIN, AA
C
      AADIF(1) = 0.
      AADIF(2) = SQRT(2*AA(2)*AA(2))
      AADIF(3) = 0.
      AADIF(4) = .5*(AA(4)+3)
      AADIF(5) = 0.
      AADIF(6) = .25*(AA(6)+15*AA(4)-10*AA(3)*AA(3))
      AADIF(7) = 0.
      AADIF(8) = .125*(AA(8)+28*AA(6)-56*AA(3)*AA(5)+35*AA(4)*AA(4))
C
C     CALCULATE THE MOMENTS OF THE RANDOMIZATION DISTRIBUTION, RA,
C     FROM THE MOMENTS OF THE DIFFERENCES, AADIF
C
      RA(2) = AADIF(2)
      RA(3) = AADIF(3) / SQRT(NO)
      RA(4) = 3. + (AADIF(4)-3.) / NO
      RA(5) = (AADIF(5)-10*AADIF(3))/NO**1.5 + (10*AADIF(3))/ SQRT(NO)
      RA(6) = 15 + (15*AADIF(4)+10*AADIF(3)*AADIF(3)-45)/NO
     *       + (AADIF(6)-15*AADIF(4)-10*AADIF(3)*AADIF(3)+30)/(NO*NO)
      RA(7) = (AADIF(7)-21*AADIF(5)-35*AADIF(4)*AADIF(3)+210*AADIF(3))
     *      / NO**2.5
     *        + (21*AADIF(5)+35*AADIF(4)*AADIF(3)-415*AADIF(3))/NO**1.5
     *        + (105*AADIF(3))/SQRT(NO)
      RA(8) = (AADIF(8)-28*AADIF(6)-56*AADIF(3)*AADIF(5)-35*AADIF(4)
     *     *AADIF(4) + 420*AADIF(4)
     *  +560*AADIF(3)*AADIF(3)-630) / NO**3
     *+(28*AADIF(6)+56*AADIF(3)*AADIF(5)+35*AADIF(4)*AADIF(4)-
     *  630*AADIF(4)-840**AADIF(3)*AADIF(3)
     *       + 1155)/(NO*NO)
     * + (210*AADIF(4)+280*AADIF(3)*AADIF(3)-630) / NO + 105
      WRITE (6,31) (I,AA(I),AADIF(I),I=1,8)
   31 FORMAT ('0MOMENTS . . .'/' PARENT POPULATION    DIFFERENCES'/
     *  (I3,2F15.4))
C
C     CALCULATE THE CORNISH-FISHER EXPANSION OF THIS DISTRIBUTION
C
      WRITE (6,112)
  112 FORMAT (' DO YOU WANT THE CORNISH-FISHER APPROXIMATION?')
      READ (5,11,END=1050) I
      IF (I .NE. 'YES   ')  GO TO 1050
      WRITE (6,1039)
 1039 FORMAT ('0CORNISH-FISHER EXPANSION . . .'/
```

```
     * ' PERCENTILE   N(0,1)    THRU 4    THRU 6    THRU 8'/)
      XK4 = RA(4) - 3
      XK6 = RA(6) - 15*RA(4) + 30
      XK8 = RA(8) - 28*RA(6) - 35*RA(4)*RA(4) + 420*RA(4) - 630
      DO 1040 I=1,NP
      CALL NDTRI(P(I),XN,C,IER)
      XN3 = XN*XN*XN
      XN5 = XN3*XN*XN
      XN7 = XN5*XN*XN
      FIN4 = XN + XK4*(XN3-3*XN) / 24
     *          - XK4*XK4*(3*XN5-24*XN3+29*XN) / 384
     *          + XK4**3*(9*XN7-131*XN5+45 *XN3-321*XN) / 3072
      FIN6 = FIN4 + XK6*(XN5-10*XN3+15*XN) / 720
     *            + XK4*XK6*(XN7-17*XN5+69 XN3-57*XN) / 1152
      FIN8 = FIN6 + XK8*(XN7-21*XN5+105*XN3-105*XN) / 40320
 1040 WRITE (6,1041) P(I), XN, FIN4, FIN6, FIN8
 1041 FORMAT (5F10.6)
C
 1050 M = ITWO(1)
      DO 1100 J=1,M
      SUM = -D(1)
      IEVEN2 = M
      ITERM = J
      DO 1200 I=2,NO
      IEVEN = IEVEN2
      IEVEN2 = ITWO(I)
      IF (IEVEN .LT. ITERM) ITERM = ITERM - IEVEN
      IF (ITERM .GT. IEVEN2)  GO TO 1280
      SUM = SUM - D(I)
      GO TO 1200
 1280 SUM = SUM + D(I)
 1200 CONTINUE
      SUM = SUM / SQRTNO
      SUM2 = - SUM
C     IF (NR .EQ. 1 .AND. NS .EQ. 1)  WRITE (6,29) SUM, SUM2
   29 FORMAT (2F10.4)
C
C     UPDATE THE SAMPLE HISTOGRAM AND SAMPLE MOMENT ACCUMULATORS
C
      INTERV = (SUM-HT1)/HW + 2.
      IF (INTERV .LT. 1) INTERV = 1
      IF (INTERV .GT. NCELLS) INTERV = NCELLS
      IHIST(INTERV) = IHIST(INTERV) + 1
      INTERV = (SUM2-HT1)/HW + 2.
      IF (INTERV .LT. 1)  INTERV = 1
      IF (INTERV .GT. NCELLS)  INTERV = NCELLS
      IHIST(INTERV) = IHIST(INTERV) + 1
      X = SUM*SUM
      XMOM(2) = XMOM(2) + X + X
```

```
      Y = X * X
      XMOM(4) = XMOM(4) + Y + Y
      X = X * Y
      XMOM(6) = XMOM(6) + X + X
      Y = Y * Y
 1100 XMOM(8) = XMOM(8) + Y + Y
  900 CONTINUE
      IF (IANTIT .EQ. 0)  GO TO 610
C
C     UPDATE CUMULATIVE HISTOGRAM AND MOMENT ACCUMULATORS
C
      DO 1300 I=1,NCELLS
 1300 ICHIST(I) = ICHIST(I) + IHIST(I)
      XCMOM(2) = XCMOM(2) + XMOM(2)
      XCMOM(4) = XCMOM(4) + XMOM(4)
      XCMOM(6) = XCMOM(6) + XMOM(6)
      XCMOM(8) = XCMOM(8) + XMOM(8)
C
C     CALCULATE SAMPLE MOMENTS
C
      NTOTAL = K2NR *(M+M)
      SSTD = SQRT(XMOM(2)/(NTOTAL-1))
      X = XMOM(2) * XMOM(2)
      SA4 = XMOM(4) * NTOTAL / X
      X = X * XMOM(2)
      SA6 = XMOM(6) * NTOTAL * NTOTAL / X
      X = X * XMOM(2)
      SA8 = XMOM(8) * NTOTAL*NTOTAL*NTOTAL / X
C
C     CALCULATE SUM AND SUM OF SQUARES FOR EACH SAMPLE MOMENT
C
      AMOM(2,1) = AMOM(2,1) + SSTD
      AMOM(2,2) = AMOM(2,2) + SSTD*SSTD
      AMOM(4,1) = AMOM(4,1) + SA4
      AMOM(4,2) = AMOM(4,2) + SA4*SA4
      AMOM(6,1) = AMOM(6,1) + SA6
      AMOM(6,2) = AMOM(6,2) + SA6*SA6
      AMOM(8,1) = AMOM(8,1) + SA8
      AMOM(8,2) = AMOM(8,2) + SA8*SA8
C
      WRITE (6,14)
   14 FORMAT (' PRINT RESULTS?')
      READ (5,11,END=600) IPRINT
      IF (IPRINT .NE. 'YES   ')  GO TO 139
C
C     CALCULATE THE CUMULATIVE MOMENTS
C
      NTOTA2 = NS * NTOTAL
      IF (IANTIT .GE. 0)  NTOTA2 = NTOTA2 * 2
      CMSTD = SQRT(XCMOM(2)/(NTOTA2-1))
      X = XCMOM(2) * XCMOM(2)
```

```
      CMA4 = XCMOM(4) * NTOTA2/ X
      X = X * XCMOM(2)
      CMA6 = XCMOM(6) * NTOTA2 * NTOTA2 / X
      X = X * XCMOM(2)
      CMA8 = XCMOM(8) * NTOTA2*NTOTA2*NTOTA2 / X
C
C     PRINT RESULTS
C
      NSAM = NS * NR
      NDIF = NO * NSAM
      NSH = NS
      NR2 = NR
      IF (IANTIT .EQ. -1)  GO TO 129
      NSH = NS / 2
      NR2 = NR + NR
  129 WRITE (6,20) IDIST,P1,P2,P3,P4,INITS,NO,NR2,NSH,NSAM,NDIF,NTOTA2
   20 FORMAT ('0PARENT POPULATION                =',I10/
     *          ' WITH PARAMETERS', 4F10.6/
     *          ' USING INITIAL RANDOM # SEED    =', I11/
     *          ' SAMPLE SIZE                    =', I10/
     *          ' SAMPLES / MONTE CARLO REPL.    =', I10/
     *          ' REPLICATION JUST COMPLETED     =', I10/
     *          ' TOTAL # SAMPLES GENERATED      =', I10/
     *          ' TOTAL # DIFFERENCES GENERATED =', I10/
     *          ' TOTAL # TEST STAT. VALUES GEN.=', I10/)
      IF (IANTIT .GE. 0)  WRITE (6,34)
   34 FORMAT (' ANTITHETIC VARIABLES USED.'/)
C
C     CALCULATE MEAN AND STD OF SAMPLE MOMENT VALUES
C
      A2STD = 0.
      A2A4 = 0.
      A2A6 = 0.
      A2A8 = 0.
C
      NSH2 = (NSH-1)*NSH
      ASTD = AMOM(2,1) / NSH
      IF (NSH .GT. 1) A2STD = SQRT((AMOM(2,2)-NSH*ASTD*ASTD)/NSH2)
      AA4 = AMOM(4,1) / NSH
      IF (NSH .GT. 1) A2A4 = SQRT((AMOM(4,2)-NSH*AA4*AA4)/NSH2)
      AA6 = AMOM(6,1) / NSH
      IF (NSH .GT. 1) A2A6 = SQRT((AMOM(6,2)-NSH*AA6*AA6)/NSH2)
      AA8 = AMOM(8,1) / NSH
      IF (NSH .GT. 1) A2A8 = SQRT((AMOM(8,2)-NSH*AA8*AA8)/NSH2)
C
      WRITE(6,21) RA(2),CMSTD,ASTD,A2STD,RA(4),CMA4,AA4,A2A4,
     *  RA(6),CMA6,AA6,A2A6,RA(8),CMA8,AA8,A2A8
   21 FORMAT (' RANDOMIZATION DISTRIBUTION'/' MOMENTS    TRRD
     * '      ALL REPLICATIONS'/
     * 24X, 'CUM       MEAN        STD'/
     * '     STD',4F10.3/
```

```
     * '     A4', 4F10.3/
     * '     A6', 4F10.3/
     * '     A8', 4F10.3//
     * ' PERCENTILES   R(0,1)', 30X, 'STAN STD'/)
C
C     DETERMINE AND PRINT PERCENTILE INFORMATION FROM THE TWO HISTOGRAMS
C
      IISEED = ISEED
 1390 ISUM = 0
      J = 1
      ISUM2 = 0
      J2 = 1
      DO 1400 I=1,NP
      THRESH = P(I) * NTOTAL
 1405 IOLD = J
 1410 J = J + 1
      IF (J .LE. NCELLS)  GO TO 1440
      X = 0.
      GO TO 1450
 1440 IF (IHIST(J) .LE. 0)  GO TO 1410
      ISUM = ISUM + IHIST(J)
      IF (ISUM .LT. THRESH)  GO TO 1405
      X = (IOLD*IHIST(IOLD) + J*IHIST(J)) / (IHIST(IOLD)+IHIST(J))
      X = (X - 2) * HW + HT1
      AHIST(I,1) = AHIST(I,1) + X
      AHIST(I,2) = AHIST(I,2) + X*X
      AHIST(I,3) = AHIST(I,3) + 1.
      IF (IPRINT .NE. 'YES   ')  GO TO 1400
      AMEAN = 0.
      ASTD = 0.
      IF (AHIST(I,3) .LT. .5)  GO TO 1395
      AMEAN = AHIST(I,1) / AHIST(I,3)
      IF (AHIST(I,3) .GT. 1.5)  ASTD = SQRT((AHIST(I,2)
     * - AHIST(I,3)*AMEAN*AMEAN) / (AHIST(I,3)-1.))
      ASTD = ASTD / SQRT(AHIST(I,3))
C
 1450 THRESH = P(I) * NTOTA2
 1415 IOLD2 = J2
 1420 J2 = J2 + 1
      IF (J2 .LE. NCELLS)  GO TO 1460
      Y = 0.
      AMEAN = 0.
      ASTD = 0.
      GO TO 1395
 1460 IF (ICHIST(J2) .LE. 0)  GO TO 1420
      ISUM2 = ISUM2 + ICHIST(J2)
      IF (ISUM2 .LT. THRESH)  GO TO 1415
      Y = (IOLD2*ICHIST(IOLD2)+J2*ICHIST(J2))/(ICHIST(IOLD2)+ICHIST(J2))
      Y = (Y-2)*HW + HT1
```

```
C
 1395 Z = Y / CMSTD
      XX = ASTD / CMSTD
      IF (IPRINT .EQ. 'YES   ') WRITE (6,22) P(I), Z, Y, AMEAN, ASTD, XX
   22 FORMAT (F10.6,5F10.5)
 1400 CONTINUE
      GO TO 600
C
   99 STOP
      END



C     TO GENERATE RANDOM SAMPLES OF SIZE N  FROM THE DISTRIBUTION
C     INDICATED BY IDIST, WITH PARAMETERS P1,P2,P3,P4, USING
C     RANDOM NUMBER SEED ISEED.  THE RANDOM VALUES ARE PLACED IN X.
C
      SUBROUTINE DATAIN(IDIST,P1,P2,P3,P4,ISEED,X,NO,NS,NR,IANTIT,AA)
      DIMENSION X(NO),AA(8),P(1000),RAW(8)
      IMPLICIT DOUBLE PRECISION (B,D,R)
      DP3 = P3
      DP4 = P4
C
      GO TO (1,2,3,4),IDIST
C
C     IDIST=1 INDICATES THE GENERALIZED LAMBDA DISTRIBUTION
C
    1 NNO = NO
      CALL STRAT(NR,NO,NNO,ISEED,0,P)
      DO 10 I=1,NO
      IF (IANTIT .LE. 0)  GO TO 10
      IF (P(I) .LE. .5)  P(I) = .5 - P(I)
      IF (P(I) .GT. .5) P(I) = 1.5 - P(I)
   10 X(I) = P1 + (P(I)**DP3 - (1.-P(I))**DP4) / P2
      IF (NS .GT. 1)  RETURN
C
C     CALCULATE THE FIRST EIGHT RAW MOMENT
C
      B80 = 1 / (8*DP3+1)
      B71 = BETA(7*DP3+1,DP4+1)
      B62 = BETA(6*DP3+1,2*DP4+1)
      B53 = BETA(5*DP3+1,3*DP4+1)
      B44 = BETA(4*DP3+1,4*DP4+1)
      B35 = BETA(3*DP3+1,5*DP4+1)
      B26 = BETA(2*DP3+1,6*DP4+1)
      B17 = BETA(DP3+1,7*DP4+1)
      B08 = 1 / (8*DP4+1)
      B70 = 1 / (7*DP3+1)
      B61 = BETA (6*DP3+1,DP4+1)
      B52 = BETA(5*DP3+1,2*DP4+1)
      B43 = BETA(4*DP3+1,3*DP4+1)
```

```
      B34 = BETA(3*DP3+1,4*DP4+1)
      B25 = BETA(2*DP3+1,5*DP4+1)
      B16 = BETA(DP3+1,6*DP4+1)
      B07 = 1 / (7*DP4+1)
      B60 = 1 / (6*DP3+1)
      B51 = BETA(5*DP3+1,DP4+1)
      B42 = BETA(4*DP3+1,2*DP4+1)
      B33 = BETA(3*DP3+1,3*DP4+1)
      B24 = BETA(2*DP3+1,4*DP4+1)
      B15 = BETA(DP3+1,5*DP4+1)
      B06 = 1 / (6*DP4+1)
      B50 = 1 / (5*DP3+1)
      B41 = BETA(4*DP3+1,DP4+1)
      B32 = BETA(3*DP4+1,2*DP4+1)
      B23 = BETA(2*DP3+1,3*DP4+1)
      B14 = BETA(DP3+1,4*DP4+1)
      B05 = 1 / (5*DP4+1)
      B40 = 1 / (4*DP3+1)
      B31 = BETA (3*DP3+1,DP4+1)
      B22 = BETA (2*DP3+1,2*DP4+1)
      B13 = BETA (DP3+1,3*DP4+1)
      B04 = 1 / (4*DP4+1)
      B30 = 1 / (3*DP3+1)
      B21 = BETA (2*DP3+1,DP4+1)
      B12 = BETA (DP3+1,2*DP4+1)
      B03 = 1 / (3*DP4+1)
      B20 = 1 / (2*DP3+1)
      B11 = BETA (DP3+1,DP4+1)
      B02 = 1 / (2*DP4+1)
      B10 = 1 / (DP3+1)
      B01 = 1 / (DP4+1)
C
      R1 = B10 - B01
      R2 = B20 - 2*B11 + B02
      R3 = B30 - 3*B21 + 3*B12 - B03
      R4 = B40 - 4*B31 + 6*B22 - 4*B13 + B04
      R5 = B50 - 5*B41 + 10*B32 - 10*B23 + 5*B14 - B05
      R6 = B60 - 6*B51 + 15*B42 - 20*B33 + 15*B24 - 6*B15 + B06
      R7 = B70 - 7*B61 + 21*B52 - 35*B43 + 35*B34 - 21*B25
     *     + 7*B16 - B07
      R8 = B80 - 8*B71 + 28*B62 - 56*B53 + 70*B44 - 56*B35
     *     + 28*B26 - 8*B17 + B08
C
C
C     CALCULATE THE SKEWNESS AND KURTOSIS FROM THE RAW MOMENTS
C
  100 AA(1) = P1 + R1/P2
      AVAR = R2 - R1*R1
      AA(2) = SQRT(AVAR/(P2*P2))
      AA(3) = (R3 - 3*R2*R1 + 2*R1**3) / AVAR**1.5
      AA(4) = (R4 - 4*R3*R1 + 6*R2*R1*R1 - 3*R1**4) / AVAR**2
```

```
      AA(5)=(R5-5*R4*R1 + 10*R3*R1*R1 - 10*R2*R1**3 + 4*R1**4)/AVAR**2.5
      AA(6)=(R6-6*R5*R1+15*R4*R1*R1-20*R3*R1**3+15*R2*R1**4-5*R1**5)
     *     / AVAR**3.
      AA(7)=(R7-7*R6*R1+21*R5*R1*R1-35*R4*R1**3+35*R3*R1**4-
     *     21*R2*R1**5+6*R1**7) / AVAR**3.
      AA(8) = (R8-8*R7*R1+28*R6*R1*R1-56*R5*R1**3+70*R4*R1**4
     *     -56*R3*R1**5+28*R2*R1**6-7*R1**8) / AVAR**4.
      RETURN
C
C     IDIST = 2 FOR NORMAL RANDOM VARIABLE USING BOX-MULLER
C
    2 CALL STRAT (NR,NO,2*NO,ISEED,0,P)
      DO 20 I=1,NO
      IF (IANTIT .LE. 0)  GO TO 15
      P(2*I-1) = 1. - P(2*I-1)
      P(2*I) = 1. - P(2*I)
   15 X(I) = (-2*ALOG(P(2*I-1)))**.5*COS(6.28318*P(2*I))
   20 X(I) = P1 + P2*X(I)
      AA(1) = P1
      AA(2) = P2
      AA(3) = 0.
      AA(4) = 3.
      AA(5) = 0.
      AA(6) = 15.
      AA(7) = 0.
      AA(8) = 105.
      RETURN
C
C     GENERATE USING THE ABSOLUTE LAMBDA DISTRIBUTION
C
    3 CALL STRAT(NR,NO,NO,ISEED,0,P)
      DO 40 I=1,NO
      IF (IANTIT .LE. 0)  GO TO 45
      IF (P(I) .LE. .5)  P(I) = .5 - P(I)
      IF (P(I) .GT. .5)  P(I) = 1.5 - P(I)
   45 IF (P(I) .LT. DP4) PP = - (DP4-P(I))**DP3
      IF (P(I) .GE. DP4) PP = (P(I)-DP4)**DP3
   40 X(I) = P1 + PP/P2
      IF (NS .GT. 1)  RETURN
C
      DO 50 I=1,8
      Y = I*DP3+1
      ISIGN = -1
      IF (MOD(I,2) .EQ. 0)  ISIGN = 1
   50 RAW(I) = (ISIGN*DP4**Y + (1-DP4)**Y) / Y
      R1 = RAW(1)
      R2 = RAW(2)
      R3 = RAW(3)
      R4 = RAW(4)
      R5 = RAW(5)
      R6 = RAW(6)
```

```
      R7 = RAW(7)
      R8 = RAW(8)
      GO TO 100
C
C     GENERATE EXPONENTIAL RANDOM VALUES WITH MEAN P1
C
    4 CALL STRAT(NR,NU,NO,ISEED,0,P)
      DO 60 I=1,NO
      IF (IANTIT .LE. 0)  GO TO 60
      IF (P(I) .LE. .5)  P(I) = .5 - P(I)
      IF (P(I) .GT. .5)  P(I) = 1.5 - P(I)
   60 X(I) = -ALOG(P(I)) * P1
      IF (NS .GT. 1)  RETURN
      AA(1) = P1
      AA(2) = P1
      AA(3) = 2
      AA(4) = 9
      AA(5) = 44
      AA(6) = 265
      AA(7) = 1854
      AA(8) = 14833
      RETURN
      END


C     TO PERFORM STRATIFED SAMPLING GIVEN A U(0,1) RANDOM GENERATOR
C     K = # OF STRATA
C     M = # OBSERVATIONS PER STRATA
C     N = NUMBER OF RANDOM VALUES TO BE GENERATED THIS CALL
C          PROBABILITIES ARE UPDATED ONLY AT END OF ROUTINE,
C          SO THE N VALUES GENERATED ARE INDEPENDENT.
C     IFIRST = 1 FOR INITIALIZATION
C            = 0 FOR GENERATION OF A RANDOM VALUE
C     R = THE RANDOM VALUE GENERATED
C
      SUBROUTINE STRAT(K,M,N,ISEED,IFIRST,VALUE)
      DIMENSION P(2000),NOBS(2000),VALUE(2000)
      IF (IFIRST .EQ. 0)  GO TO 100
C
C*****INITIALIZE PROBABILITIES P(I) OF THE RANDOM VALUE COMING
C     FROM SUBINTERVALS 1,2,...,I
C     AND THE NUMBER OF OBSERVATIONS NOBS(I) STILL TO COME
C     FROM SUBINTERVAL I, I=1,2,...,K
C
    5 XK = K
      IF (K .LE. 2000 .AND. N .LE. 2000)  GO TO 15
      WRITE (6,1) K,N
    1 FORMAT ('0IN THE STRATIFIED SAMPLING SUBROUTINE ''STRAT'','/
     * ' THE NUMBER OF STRATA MUST NOT EXCEED 2000.  YOU HAVE',I10/
     *' THE NUMBER OF RANDOM VALUES MUST NOT EXCEED 2000. YOU HAVE',I10,
     * /' EXECUTION TERMINATED.')
```

```
      STOP 777
   15 DO 10 I=1,K
      P(I) = I/XK
   10 NOBS(I) = M
      TOTOBS = M*XK
      RETURN
C
C*****GENERATE ONE VALUE P DISTRIBUTED U(0,1)
C     THEN DETERMINE WHICH SUBINTERVAL I THE RANDOM VALUE
C     R WILL COME FROM, USING CURRENT P(I) VALUES
C
  100 DO 300 JJ=1,N
      R = RANDOM(ISEED)
      DO 20 J=1,K
      IF (R .GT. P(J))  GO TO 20
      I = J
      GO TO 200
   20 CONTINUE
      I = K
C
C     CALCULATE WHERE R FALLS IN THE I TH SUBINTERVAL
C
  200 NOBS(I) = NOBS(I) - 1
      Y = 0.
      IF (I .GT. 1)  Y = P(I-1)
      R = (R-Y) / (P(I)-Y)
      VALUE(JJ) = (I+R-1.) / XK
      IF (VALUE(JJ)  .LE. 0) VALUE(JJ) = .1E-10
      IF (VALUE(JJ) .GE. 1)  VALUE(JJ) = 1. - .1E-10
  300 CONTINUE
C
C*****REVISE THE SUBINTERVAL PROPERTIES FOR THE NEXT CALL
C
      IF (TOTOBS .LE. N+.1)  GO TO 5
      TOTOBS = TOTOBS - N
      P(1) = NOBS(1) / TOTOBS
      DO 30 I=2,K
   30 P(I) = P(I-1) + NOBS(I) / TOTOBS
      RETURN
      END


C     TO GENERATE UNIFORM (0,1) RANDOM NUMBERS ON THE UNIVAC 1108
C
      FUNCTION RANDOM(ISEED)
      ISEED = ISEED * 131075
      IF (ISEED .LE. 0)  ISEED = ISEED + 34359738367 + 1
      RANDOM = ISEED * .2910383E-10
      RETURN
      END
```

REFERENCES

Arnold, H. J. (1964) "Permutation support for multivariate techniques," Biometrika, Vol. 51, pp. 65-70.

Bartlett, M. S. (1935) "The effect of non-normality on the t distribution," Proc. Camb. Phil. Soc., Vol. 31, part 2, pp. 223-231.

Barton, D. C. and F. N. David. (1961) "Randomization bases for multivariate tests," Bull. Int. Stat. Inst., Vol. 39, part 2, pp. 455-467.

Bhattacharyya, G. K., Richard A. Johnson, and Henry R. Neave. (1971) "A comparative study of the bivariate rank sum test and T^2," Technometrics, Vol. 13, No. 1, pp. 191-198.

Box, G. E. P. and S. L. Anderson. (1955) "Permutation theory in the derivation of robust criteria and the study of departures from assumption," Journal of Roy. Stat. Soc., Series B, 17,1.

Box, G. E. P. and G. S. Watson. (1962) "Robustness to non-normality of regression tests," Biometrika, Vol. 49, No. 1, pp. 93-106.

Bradley, R. A., K. M. Patel, and D. D. Wackerly. (1971) "Approximate small-sample distributions for multivariate two-sample nonparametric tests," Biometrics, Vol. 27, pp. 515-530.

Burr, I. W. (1973) "Parameters for a general system of distributions to match a grid of α_3 and α_4," Comm. in Stat., Vol. 2, pp. 1-21.

Burr, Irving W. (1974) Applied Statistical Methods, Academic Press, New York.

Chung, J. H. and D. A. S. Fraser. (1958) "Randomization tests for a multivariate two-sample problem," J. Am. Stat. Assn., Vol. 53, pp. 729-735.

Cochran, W. G. (1966). Sampling Techniques, Wiley, New York.

Collier, Raymond O. and Frank B. Baker. (1966) "Some Monte Carlo results on the power of the F-test under permutation in the simple randomized block design," Biometrika, Vol. 53, pp. 199-203.

Collier, R. O. and Frank B. Baker. (1963) "The randomization distribution of F-ratios for the split plot design--an empirical investigation," Biometrika, Vol. 50, Nos. 3 & 4, pp. 431-438.

Cooper, J. D., S. A. Davis, and N. R. Dano. (1965) "Pearson Universal Distribution Generator (PURGE)," Proceedings of the 19th Annual Conference, American Society for Quality Control, pp. 402-411.

Cornish, E. A. and R. A. Fisher. (1937) "Moments and cumulants in the specification of distributions," Institut International de Statistique, Vol. 4, pp. 1-14.

David, H. A. and R. S. Mishriky. (1968) "Order statistics for discrete populations and for grouped samples," J. Am. Statist. Assn., Vol. 63, pp. 1390-1398.

Deutsch, S. J. and S. DeSetto. (1971) "Computations of the Randomization Reference Distribution," Unpublished Tech. Rept., Dept. of O.R., Polytechnic Institute of Brooklyn.

Deutsch, S. J. and W. Hamm. (1975) "Measuring consistency and convergence of opinion in Delphi studies," ORSA/TIMS National Meeting, Fall.

Dwass, Meyer. (1957) "Modified randomization tests for nonparametric hypotheses," Ann. Math. Statist., Vol. 28, pp. 181-187.

Fisher, R. A. (1926) "The arrangement of field experiments," Journal of Ministry of Agriculture, Vol. 33, pp. 503-513.

Fisher, R. A. (1929) "Moments and product moments of sampling distributions," Proceedings of the London Mathematical Society, Series 2, Vol. 30, pp. 199-238.

Fisher, R. A. (1935) The Design of Experiments, Oliver & Boyd, Ltd., Edinburgh & London.

Fisher, R. A. (1936) "The coefficient of racial likeness and the future of craniometry," J. Royal Anthropological Inst. of Great Britain, Vol. 66, pp. 57-63.

Gayen, A. K. (1949) "The distribution of "Student's" t in random samples of any size drawn from non-normal universes," Biometrika, Vol. 36, pp. 353-369.

Gibbons, J. D. (1971) Nonparametric Statistical Inference, McGraw-Hill, New York.

Hamm, W. (1974) Personal communication.

Heiler, S. and K. Weichselberger. (1969) "On the permutation test and a confidence interval derivable from it," Metrika, Vol. 14, pp. 232-248.

Heyde, C. C. (1963) "On a property of the lognormal distribution," J. R. Statist. Soc., B, Vol. 25, p. 392.

Hines, W. W. and D. C. Montgomery. (1972) Probability and Statistics in Engineering and Management Science, Ronald, New York.

Hoeffding, W. (1952) "The large sample power of tests based on permutations of observations," Annals Math. Stat., Vol. 23, pp. 169-192.

Johnson, N. L. (1949) "Systems of frequency curves generated by methods of translation," Biometrika, Vol. 36, pp. 149-176.

Johnson, N. L. and S. Kotz. (1969, 1970a, 1970b, 1972) Distributions in Statistics, Wiley, New York.

Kempthorne, Oscar. (1966) "Some aspects of experimental inference," Jour. Am. Statist. Assn., Vol. 61, No. 313, pp. 11-34.

Kempthorne, O. (1955) "The randomization theory of experimental inference," Jour. Am. Statist. Assn., Vol. 50, pp. 946-967.

Kempthorne, O. and T. E. Doerfloer. (1969) "The behavior of some significance tests under experimental randomization," Biometrika, Vol. 56, pp. 231-248.

Kempthorne, O., J. Zyskind, R. P. Basson, F. B. Martin, T. E. Doerfler, and E. J. Carney. (1966) "Research on analysis of variance and data interpretation," ARL Tech. Rep., Wright-Patterson Air Force Base, No. 66-0240.

Kendall, M. G. and A. Stuart. (1967, 1968, 1969) The Advanced Theory of Statistics. Vol. I (1969), Vol. II (1967), Vol. III (1968), Hafner, London.

Klienjen, Jack, P.C. (1974) Statistical Techniques in Simulation, Part I, Marcel Dekker, New York.

Lehmann, E. L. and C. Stein. (1949) "On the theory of some nonparametric hypotheses," Annals of Math. Stat., Vol. 20, pp. 28-45.

Lewis, T. G. (1975) Distribution Sampling for Computer Simulation, Heath, Lexington, Mass.

Mardia, K. V. (1971) "The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model," Biometrika, Vol. 58, pp. 105-121.

Markoff, A. (1884) "Demonstration de certaines inégalites de Tchebycheff," Math. Ann., Vol. 24, pp. 172-180.

McKay, A. T. (1932) "A Bessel Function distribution," Biometrika, Vol. 24, pp. 39-44.

Moy, W. A. (1966) Practical Variance-Reducing Procedures for Monte Carlo Simulations, University of Wisconsin, 7 Oct. 1966.

Nair, K. R. (1940) "The median in tests by randomization," Sankhya, Vol. 4, pp. 543-550.

Naylor, Thomas H., J. L. Balintfy, D. S. Burdick, and K. Chu. (1966) Computer Simulation Techniques, Wiley, New York.

Pearson, E. S. (1937) "Some aspects of the problem of randomization," Biometrika, Vol. 29, pp. 53-64.

Pearson, E. S. (1929) "Small samples from non-normal populations," Biometrika, Vol. 21, pp. 259-286.

Pearson, K. (1895) "Contributions to the mathematical theory of evolution. II. Skew variations in homogeneous material," Philosophical Transactions of the Roy. Soc. of London, Series A, Vol. 186, pp. 343-414.

Perlo, V. P. (1933) "Student's ratio for samples of 3 from a rectangular population," Biometrika, Vol. 25, pp. 203-204.

Pitman, E. J. G. (1937) "Significance tests which may be applied to samples from any population," J. Roy. Stat. Soc., Suppl., Vol. 4, pp. 119-130.

Pitman, E. J. G. (1937) "Significance tests which may be applied to samples from any population. II. The correlation coefficient test," J. Roy. Stat. Soc., Suppl., Vol. 4, pp. 225-232.

Pitman, E. J. G. (1937) "Significance tests which may be applied to samples from any population. III. The analysis of variance test," Biometrika, Vol. 29, pp. 322-335.

Puri, M. L. and P. K. Sen. (1971) Nonparametric Methods in Multivariate Analysis, Wiley, New York.

Ractliffe, J. F. (1968) "The effect on the t-distribution of non-normality in the sampled population," Applied Statistics, Vol. 17, pp. 42-48.

Ramberg, J. S. and B. W. Schmeiser. (1972) "An approximate method for generating symmetric random variables," Communications of the ACM, Vol. 15, No. 11, pp. 987-990.

Ramberg, J. S. and B. W. Schmeiser. (1974) "An approximate method for generating asymmetric random variables," Communications of the ACM, Vol. 17, No. 2, pp. 78-82.

Reitsma, A. (1963) "On approximations to sampling distributions of the mean for samples from non-normal populations," Ann. Math. Statist., Vol. 34, No. 4, pp. 1308-1314.

Royden, H. L. (1953) "Bounds on a distribution function when its first moments are given," Annals of Math. Statist., Vol. 34, pp. 361-376.

Sophister. (1928) "Samll samples from an infinite skew population," Biometrika, Vol. 20a, pp. 389-423.

Stieltjes, T. J. (1884) "Quelques recherches sur les quadratures dites mécaniques," Ann. Sci. Ecole Norm. Sup., Vol. 3, pp. 409-426.

Thiele, T. N. (1931) "The theory of observations," Annals of Math. Stat., Vol. 2, pp. 165-307.

Tiku, M. L. (1963) "Students' t distribution in terms of Hermite and Laguerre polynomials," J. Indian Math. Soc., Vol. 27, p. 91.

Wallace, David L. (1958) "Asymptotic approximations to distributions," Annals Math. Statist., Vol. 29, pp. 635-654.

Welker, E. L. (1947) "The distribution of the mean," Annals of Math. Statist., Vol. 18, pp. 111-117.

Wilk, M. B. and O. Kempthorne. (1956) "Derived linear models and their use in the analysis of randomized experiments," WADC Technical Report, Vol. II, No. 55-244.

Wilk, M. B. (1955) "The randomization analysis of a generalized randomized block design," Biometrika, Vol. 42, pp. 70-79.

Wilks, S. S. (1962) Mathematical Statistics, Wiley, New York.

Zyskind, George. (1963) "Some consequences of randomization in a generalization of the balanced incomplete block design," Ann. Math. Statist., Vol. 34, No. 4, pp. 1569-1581.

VITA

Bruce W. Schmeiser was born in Burlington, Iowa on January 17, 1947, the first son of L. Maurice and Maurine Schmeiser. He attended public schools in Wapello, Iowa and graduated from Wapello High School in June, 1965.

He enrolled at The University of Iowa in the fall of 1965 in the Honors Program and graduated with distinction in September, 1969 with a Bachelor of Arts degree in the Mathematical Sciences. During his senior year Mr. Schmeiser was employed by Westinghouse Learning Corporation as a computer programmer in the area of statistical analysis.

Immediately after graduation, Mr. Schmeiser began graduate study in Industrial and Management Engineering at The University of Iowa. With the acceptance of his thesis, A General Algorithm for Generating Random Variables, under the direction of Dr. John S. Ramberg, Mr. Schmeiser completed the requirements for the Master of Science Degree in 1971. Two papers, each based on his master's thesis and co-authored by Dr. Ramberg, were published in Communications of the ACM in November, 1972 and February, 1974, respectively.

In September, 1970, he joined Electronic Data Systems as a systems engineer in the Management Systems Group. During his employment, Mr. Schmeiser directed implementation of a computerized production, inventory, and distribution control system for a regional beer brewer, performed digital computer simulation studies for several organizations, designed and implemented a system for performance analysis of the

branches of a large brokerage firm, and consulted in the areas of operations research and statistics.

Leaving Electronic Data Systems to pursue a doctorate in Industrial and Systems Engineering, Mr. Schmeiser enrolled at The Georgia Institute of Technology in September, 1972. As a graduate student, he taught courses in probability, statistics, computer simulation, and operations research. In November, 1975, he will join the faculty of the Department of Industrial Engineering and Operations Research at Southern Methodist University in Dallas, Texas as an Assistant Professor.

In October, 1971, Mr. Schmeiser and Donna Lea Nelmes of Lafayette, Illinois were married. A son, William Bryan, was born in January, 1975.