THE SMALL-SAMPLE POWER OF SOME NONPARAMETRIC TESTS

by

Jean Dickinson Gibbons, A.B., M.A.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

September 1962

Blacksburg, Virginia

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter I

INTRODUCTION

Although some nonparametric statistical techniques have a long history, most of the theoretical research in the discipline of nonparametric statistical inference is comparatively recent. Many nonparametric tests in common use have been advanced principally on intuitive grounds, and their properties are still not completely understood. In recent years, much interesting research has been done on the asymptotic properties of such tests, and methods of constructing tests with desirable large-sample properties have been developed. However, these investigations are of limited practical value unless it is known how soon asymptotic results provide a reasonable approximation in samples of moderate size.

In this dissertation, the emphasis is on exact small-sample properties, particularly on the power of a variety of nonparametric tests. Some asymptotic results are also obtained. Two new rank tests on the equality of two distributions are proposed and their properties examined.

The main difficulty in determining the power of nonparametric tests is brought about by their very generality and the consequent abundance of reasonable alternatives to whatever null hypothesis may be under consideration. In view of the central role played by the normal distribution in statistics, one approach has been to attempt the evaluation of power under suitable normal alternatives.

Most of the results presently available are for nonparametric tests of the null hypothesis that two samples of sizes m and n come from identical populations against the alternative of normal distributions differing only in location. Van der Waerden (1952, 1953) has found the exact power of the X test, Mann-Whitney, Maximum Absolute Deviation, and runs tests for isolated sample sizes. Dixon (1954) used numerical methods to find the power of the Mann-Whitney, Maximum Absolute Deviation, median, and runs two-sided tests for sample sizes $m = n = 3, 4, 5$. The exact power of the median test has been calculated by Barton (1957) for $m = n = 9$. Teichroew (1954) has computed mathematically the probabilities of the two most extreme rank orderings under normal alternatives, for a wide range of small sample sizes.

Since the extensive numerical work required for normal alternatives has greatly limited the range of results, some empirical and approximate calculations have been attempted. For $m = n = 10$, Epstein (1955) found the power of the Maximum Absolute Deviation, Epstein's Exceedances, median, and Mann-Whitney tests. Experimental comparisons of the latter test and the t test were made by Hemelrijk (1961) also for $m = n = 10$. The probabilities of all possible rank orderings under normal alternatives have been determined empirically by Teichroew (1955) for $(m,n) = (2,3)$, $(2,4)$, $(3,3)$ and $(3,4)$. The power of the c_1 test for $m = n = 4$ was determined by Terry (1952) using random sampling. Tsao (1957) used a polynomial method of approximation to the normal distribution to find all the rank order probabilities for equal sample sizes of 2 and 3. A few isolated rank order probabilities have been approximated by Sundrum (1954).

For alternatives of changes in location and scale in the uniform distribution and translation in the exponential distribution, exact power functions have been found by Leone et al. (1961) for the median test and Massey's test, and by Haynam et al. (1961) for the Mann-Whitney and median tests, for a variety of selected sample sizes.

The power function of the two-sided, two-sample sign test has been extensively tabulated by Dixon (1953) for α near .05 and .01, and p in intervals of .05, where p is the expected proportion of plus signs under the alternative. Previously, Walsh (1946) investigated the power of the one-sample, one-sided sign test and compared some results against normal alternatives with the power function of the t test for small sample sizes and isolated significance levels. He also found that performing the sign test on a distribution whose mean and median do not exactly coincide does not affect the significance level greatly.

Although a number of investigators have attempted the evaluation of power functions of nonparametric tests with parametric alternatives, the range of sample sizes and significance levels is very limited. Most of the results are for nonrandomized decision rules. The power functions are consequently difficult to compare.

The first portion of this paper (Chapter II) deals with power function comparisons between the one-sample, one-sided sign test and Student's t test, when the underlying distribution function is assumed to be close to normal, but deviating from the exact normal distribution either by skewness, kurtosis, or both. The Edgeworth-Cramér expansion is used to represent the population distribution function. Numerical results are given for samples of size 10, significance level .05. Srivastava (1958) has performed the power calculations for the t test under similar conditions.

The natural drawback to obtaining power functions against parametric alternatives is that the results apply only to the particular distribution assumed. Especially for comparisons among similar nonparametric tests, distribution-free properties would be desirable. If the power functions are independent of any specific distribution, the alternative can also be termed nonparametric. Consider the null hypothesis that two random samples come from the same population. Many nonparametric tests based on the ranks of the variables, obtained by combining the two samples and arranging the variables in ascending order, have been proposed for this hypothesis.

Lehmann (1953) has employed a nonparametric alternative to compare the small-sample power functions of six two-sample rank tests of this hypothesis when $m = n = 4, 6, \alpha = .10$. The alternative is that the random variables from the second sample, Y, are distributed as the largest of k of the variates from the first sample, X, where the distribution of X is not specified. Although the results are interesting from a theoretical point of view, the two alternative distributions are usually quite dissimilar unless k is large.

The second and major portion of this paper (Chapters III, IV and V) is concerned with power functions of two-sample rank tests of the hypothesis of identical populations. But here the alternative is that the random variables from the first sample, X, are distributed as the smallest of k random variables from some unspecified continuous distribution F, and the random variables from the second sample, Y, are distributed as the largest of these k. If the density function of F is symmetrical, then the densities of X and Y are mirror images. This is another nonparametric type of alternative for tests based on ranks, since the probability of any arrangement is independent of F.

Since the power functions of rank tests are found by summing the probabilities of the rejection orderings under the stated alternative, a practical method for their computation is needed. Hoeffding (1951) has introduced a general method of calculating the probability of any rank ordering, provided that the alternative expresses a functional relationship between the two distributions. However, extensive multiple integration is required to utilize this result. Here, more direct formulae are derived for these probabilities under the alternative of two extreme distributions, and a theorem is proved concerning the equivalence of certain pairs of order probabilities.

The power functions of eight nonparametric rank tests are tabulated for $k = 2$, 3 and 4. The cases included are equal sample sizes such that $m = n \leq 4$ with significance levels of .01, .05 and .10, and all unequal sample sizes such that $m + n \leq 8$ for significance levels which place only one or both of the two extreme orderings in the critical region. These results are compared with power functions against the alternative that both samples come from normal populations differing only in location. The standardized difference between the two normal populations corresponds to the standardized difference between the two extreme distributions assumed under the previous alternative.

The alternatives of one and two extreme distributions are special cases of an even more general type of alternative. This is that the X random variables are distributed as some arbitrary continuous function of the distribution of the Y random variables and a parameter θ. The locally most powerful rank test against this general alternative is derived. Its asymptotic properties are investigated by applying the general results obtained by Chernoff and Savage (1958), Uzawa (1960) and Capon (1961). Some specific alternatives are considered. Two of these alternatives, which are similar in spirit to the alternative of two extreme distributions, suggest new rank tests. One of them, which I have called the Psi test, is shown to have some desirable small-sample properties.

Since power function calculations for small samples are limited by the tedium of the computation of the order probabilities under the alternative, large sample power comparisons are often made when the asymptotic distributions are not too difficult to obtain. Large sample power results for the Wilcoxon and Wald-Wolfowitz runs tests are given for the alternative of two extreme distributions with $k = 2$, for equal sample sizes and significance level .05.

Chapter II

SMALL-SAMPLE POWER OF THE ONE-SAMPLE SIGN TEST FOR APPROXIMATELY NORMAL DISTRIBUTIONS

Suppose that we wish to test a hypothesis of location, where the data consist of a single sample of independent random variables from a population with a continuous distribution function. If the form of the population distribution function is known, a parametric test of location should be employed. On the other hand, the single-sample nonparametric tests of location require no further assumptions about the specific character of the distribution. But how powerful are these nonparametric tests?

When the common assumption of an underlying normal distribution is justified and the variance is unknown, Student's *t* test can be used to test the null hypothesis that the population mean μ is equal to some specified value μ_o. For a one-sided alternative, i.e., μ is equal to μ_1 which is greater than μ_o (or less than μ_o), the *t* test is the uniformly most powerful test of the null hypothesis. There is then no need to apply any other criterion since it is impossible to find a "better" test. In spite of this handicap, the power functions of one-sample nonparametric tests of location against a normal alternative are sometimes compared with the power function of the optimum test.

However, in the more usual situation, the investigator does not "know" that his data come from a normal population, but he may have good reason to believe that the frequency function is approximately normal. If the t test is applied anyway, how powerful are the results? Srivastava (1958) has tried to answer this query by computing the power function of the t test for an approximately normal distribution. He characterizes the deviation from the exact normal by degrees of skewness and kurtosis, as measured by the third and fourth cumulants, κ_3 and κ_4, and represents the distribution function by the Edgeworth-Cramér asymptotic expansion. The question naturally arises as to how powerful appropriate nonparametric tests are in comparison with the t test, when the population density function deviates from the normal.

In this chapter the power function of the single-sample, one-sided sign test will be determined for comparison with Srivastava's results. The main difficulty is the difference in hypotheses, since the theory of the sign test is based on testing a hypothesis on the location of the median. For an exact normal distribution, the mean and median coincide so that the hypotheses are the same. But if normality holds only approximately, the hypotheses will differ according to the degree of skewness in the distribution. Therefore, three different sets of hypotheses are considered:

(1) $H_o': M = M_o$, $H_1': M = M_1 > M_o$,

(2) $H_o'': M = M_o = \mu_o - \frac{1}{6}\kappa_3\sigma$, $H_1'': M = M_1 = \mu_1 - \frac{1}{6}\kappa_3\sigma > M_o$, and

(3) $H_o: \mu = \mu_o$, $H_1: \mu = \mu_1 > \mu_o$,

and their power functions tabulated. The results in all cases indicate that the gap between the power functions of the parametric and nonparametric tests is narrowed considerably when the population is not an exact normal distribution.

The calculations and conclusions are immediately applicable to a comparison between the paired <u>t</u> test and the paired sign test in a two-sample situation, when the underlying variable is taken to be the difference between corresponding measurements on the two populations.

2.1 The Edgeworth Expansion as Representation of an Approximately Normal Distribution

A distribution function F(y) for a random variable Y which is known to be approximately normal with mean 0 and variance 1 can be approximated by $\Phi(y)$ where

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2} \, dt \quad .$$

However, a more accurate representation of F(y) might be found in the form

$$F(y) = \Phi(y) + R(y) \quad ,$$

where R(y) is some convenient expansion of the remainder terms. We wish to obtain an expression for R(y) in terms of the cumulants of the distribution of Y.

If Y is considered to be a linear function of the sum of N independent random variables, Y_1, Y_2, ..., Y_N, drawn from a population with mean 0 and variance 1, and is defined by $Y = \sum_{i=1}^{N} Y_i/\sqrt{N}$, then we know by the Central Limit Theorem that Y is approximately normally distributed with mean 0 and variance 1 when N is large. Let $\Psi(t)$ denote the characteristic function of Y and $\Psi_1(t)$ the characteristic function of each of the Y_i. Then

$$\Psi(t) = [\Psi_1(t/\sqrt{N})]^N \quad .$$

The cumulant generating function of the Y's, where κ_r' denotes the rth cumulant of the random variables Y_i, is

$$\log \Psi(t) = N \log \Psi(t/\sqrt{N}) = N\Big[\frac{(it)^2}{2!N} + \frac{\kappa_3'(it)^3}{3!\ N^{3/2}} + \frac{\kappa_4'(it)^4}{4!\ N^2} + \ldots\Big]$$

$$= \frac{(it)^2}{2!} + \frac{\kappa_3'(it)^3}{3!\ N^{\frac{1}{2}}} + \frac{\kappa_4'(it)^4}{4!\ N} + \ldots \quad ,$$

so that $\Psi(t) = e^{-\frac{1}{2}t^2} \exp\Big[\frac{\kappa_3'(it)^3}{6N^{\frac{1}{2}}} + \frac{\kappa_4'(it)^4}{24N} + \ldots\Big]$,

or $$\Psi(t) = e^{-\frac{1}{2}t^2}\Big[1 + \frac{\kappa_3'(it)^3}{6N^{\frac{1}{2}}} + \frac{\kappa_4'(it)^4}{24N} + \frac{1}{2!}\frac{\kappa_3'^2(it)^6}{36N}\Big], \qquad (2.1)$$

the terms neglected are at most of order $1/N^{3/2}$.

Applying the Fourier Integral transform to $\Psi(t)$ in (2.1), we may write

$$f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}e^{-\frac{1}{2}t^2}\Big[1 + \frac{\kappa_3'(it)^3}{6N^{\frac{1}{2}}} + \frac{\kappa_4'(it)^4}{24N} + \frac{\kappa_3'^2(it)^6}{72N}\Big]dt \quad . \qquad (2.2)$$

Introducing the same transform for the normal distribution,

$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$, whose characteristic function is

$\Psi(t) = e^{-\frac{1}{2}t^2}$, we have

$$\varphi(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity} e^{-\frac{1}{2}t^2} dt \quad . \tag{2.3}$$

From (2.3) we observe that

$$\frac{\partial^r \varphi(y)}{\partial y^r} = \varphi^{(r)}(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (-it)^r e^{-ity} e^{-\frac{1}{2}t^2} dt \quad , \tag{2.4}$$

so that

$$f(y) = \varphi(y) - \frac{\kappa_3'}{6N^{\frac{1}{2}}}\varphi^{(3)}(y) + \frac{\kappa_4'}{24N}\varphi^{(4)}(y) + \frac{\kappa_3'^2}{72N}\varphi^{(6)}(y) \ , \tag{2.5}$$

from (2.2) and (2.4). The order of each term is evident. However, a more convenient expansion is in terms of the cumulants of Y, κ_r. The relationship is $\kappa_r = \kappa_r'/N^{\frac{r}{2}-1}$, so that (2.5) can be written

$$f(y) = \varphi(y) - \kappa_3\varphi^{(3)}(y)/6 + \kappa_4\varphi^{(4)}(y)/24 + \kappa_3^2\varphi^{(6)}(y)/72 \ , \tag{2.6}$$

or equivalently,

$$F(y) = \Phi(y) - \kappa_3\varphi^{(2)}(y)/6 + \kappa_4\varphi^{(3)}(y)/24 + \kappa_3^2\varphi^{(5)}(y)/72 \ . \tag{2.7}$$

This expression for F(y) is called the Edgeworth-Cramér asymptotic expansion (Edgeworth, 1905; Cramér, 1928, 1937). We will use it to represent the distribution function of a variable which is approximately normally distributed with zero mean and unit variance. Barton and Dennis (1952) have shown that the restrictions $0 \leq \kappa_4 \leq 2.4$ and $\kappa_3^2 \leq 0.2$ must be imposed to ensure a positive definite unimodal density function.

A somewhat more practical expression of the Edgeworth expansion is in terms of the Hermite polynomials. The general relationship between the coefficients $\varphi^{(r)}(y)$ and $H_r(y)$, the rth Hermite polynomial, is $\varphi^{(r)}(y) = (-1)^r H_r(y)\varphi(y)$. For example, $H_2(y) = y^2-1$, $H_3(y) = y^3-3y$, $H_5(y) = y^5-10y^3+15y$, etc. Substituting these coefficients into (2.7), we obtain

$$F(y) = \Phi(y) - \kappa_3(y^2-1)\varphi(y)/6 + \kappa_4(y^3-3y)\varphi(y)/24 + \kappa_3^2(y^5-10y^3+15y)\varphi(y)/72 \quad . \qquad (2.8)$$

2.2 Student's t test Applied to an Approximately Normal Distribution

Consider a sample of n independent random variables which are identically and approximately normally distributed with mean μ and variance σ^2. The standardized random variables will be denoted by $Y_1, Y_2, \ldots, Y_n$. With the density of Y represented by the Edgeworth-Cramér expansion (ignoring terms of order at most $1/N^{3/2}$), Srivastava (1958) has computed the power of the one-sample t test of the null hypothesis $H_o: \mu = \mu_o$ versus the one-sided alternative $H_1: \mu = \mu_1 > \mu_o$ for $\kappa_3 = -.6(.2).6$ and $\kappa_4 = -1(1)2$, $n = 10$, $\rho_n = \sqrt{n}\,(\mu_1-\mu_o)/\sigma = 0(1)4$. The upper five per cent point of the ordinary t distribution with nine degrees of freedom is used as the critical value of t. Thus the true significance level is .05 only for the case $\kappa_3 = \kappa_4 = 0$. His results (Table 2, p. 427) show that the power increases with κ_4 for all κ_3 when $\rho_n < 2$, and decreases with κ_4 when $\rho_n > 2$. The included values of κ_3 have more effect than κ_4 on the "true" Type I error probabilities, which are represented by the table entries for $\rho_n = 0$.

Since the true significance level changes with κ_3 and κ_4, the results are difficult to interpret. Therefore, Srivastava has computed the power of the t test for an exact .05 significance level for the one case $\kappa_3 = .6$, $\kappa_4 = .4$, with $\rho_n = 0(1)4$ and $\rho_n = -4(1)0$ (which would correspond to the alternative $\mu = \mu_1 < \mu_o$). The correct critical value of t is found using the inverse Cornish-Fisher expansion (Cornish and Fisher, 1937). The power functions are greatly affected by which value of t defines the critical region when ρ_n is small (see Table 2.5 *infra*). Both of these power functions for $\kappa_3 = .6$, $\kappa_4 = .4$, are compared with the power of the .05 level *t* test for an exact normal distribution, $\kappa_3 = \kappa_4 = 0$. The calculations indicate that the power of the *t* test for an approximately normal distribution with $\kappa_3 = .6$, $\kappa_4 = .4$, is considerably greater than for an exact normal distribution if $\rho_n > 0$ (and less if $\rho_n < 0$), when the correct critical region is employed for each distribution.

Although these latter comparisons made by Srivastava are interesting from a theoretical point of view, in actual practice the upper 5% point of the ordinary *t* distribution would generally be used to perform a test of the hypothesis. The investigator should be aware that the true significance level is .05 only if the distribution is exactly normal. In fact, unless the values of κ_3 and κ_4 can be assessed with reasonable accuracy for the given situation, the correct critical value of *t* for an approximately normal distribution cannot be determined.

2.3 Sign Test on the Population Median

A nonparametric test which is often used for a hypothesis of location when no assumptions are made about the underlying population distribution function is the one-sample sign test. However, here the null hypothesis to be tested is that the population median M is equal to M_o, H_o': $M = M_o$, where we have a sample of n random variables, $X_1, X_2, \ldots, X_n$, which are independent and identically distributed. Under the null hypothesis, $Pr(X_i > M_o) = Pr(X_i < M_o) = \frac{1}{2}$, regardless of the population density function, and we would expect half of the n differences $(X_i - M_o)$ to be positive and half to be negative. If the alternative is one-sided, H_1': $M = M_1 > M_o$, the sign test would require a rejection of the null hypothesis when the number of plus signs among the n differences $(X_i - M_o)$ is greater than r_o, where r_o is chosen so that $\sum_{r=r_o+1}^{n} \binom{n}{r} (\frac{1}{2})^n = \alpha$ (α being the probability of the Type I error). A randomized decision rule may be used when necessary to obtain the exact desired significance level α. Then r_o is chosen so that $\sum_{r=r_o+1}^{n} \binom{n}{r} (\frac{1}{2})^n$ is as close as possible to α, but still less than α. The test would be to reject always when $r > r_o$ and with probability p when $r = r_o$, where p satisfies

$$\sum_{r=r_o+1}^{n} \binom{n}{r} (\tfrac{1}{2})^n + p\binom{n}{r_o} (\tfrac{1}{2})^n = \alpha \quad . \qquad (2.9)$$

The power of this test is given by

$$\sum_{r=r_o+1}^{n} \binom{n}{r} p_1'^{\,r} q_1'^{\,n-r} + p\binom{n}{r_o} p_1'^{\,r_o} q_1'^{\,n-r_o} \quad , \qquad (2.10)$$

where $p_1' = Pr(X_i > M_o | H_1')$ and $q_1' = 1 - p_1'$.

Thus in order to calculate p_1' and evaluate the power function, some assumptions must be made about the population distribution under the alternative. If the assumed distribution is symmetrical about its mean, then mean and median coincide and $H_o = H_o'$, $H_1 = H_1'$. We shall assume only that, under the alternative, the $X_1, X_2, \ldots, X_n$ are independent random variables, identically distributed with mean μ_1 and variance σ^2, and the corresponding standardized variables are distributed according to the approximately normal density function defined in (2.6). The two sets of hypotheses will be identical if $\kappa_3 = 0$. But it is always true that $M_o > M_1$ if and only if $\mu_o > \mu_1$, and $M_o < M_1$ if and only if $\mu_o < \mu_1$. This implies that the alternative H_1' is true if and only if H_1 is true and similarly for the two null hypotheses, so that the results of the two tests will be approximately the same.

Under our assumptions, we have

$$p_1' = Pr[(X_i-\mu_1)/\sigma > (M_o-\mu_1)/\sigma | H_1'] = 1 - F[(M_o-\mu_1)/\sigma], \quad (2.11)$$

where $F(y)$ is given by (2.7). If we define $\rho_n' = \sqrt{n}\ (\mu_1-M_o)/\sigma$, then

$$p_1' = 1 - F(-\rho_n'/\sqrt{n}) \quad . \qquad (2.12)$$

The standardized difference for the t test of H_o was defined as $\rho_n = \sqrt{n}\ (\mu_1-\mu_o)/\sigma$, and the null hypothesis H_o is true whenever $\rho_n = 0$. However, $\rho_n' = 0$ does not indicate that the null hypothesis H_o' holds true unless the mean and median of f(x) coincide. For example, when $\kappa_3 = 0$, f(x) is symmetrical and thus $M_1 = \mu_1 = M_o$ for $\rho_n' = 0$. If we also assume that the X_i have mean μ_o and variance σ^2 under the null hypothesis, we can write $\rho_n' = \sqrt{n}\ (\mu_1-M_o)/\sigma = \rho_n + \sqrt{n}\ (\mu_o-M_o)/\sigma$, so that $\rho_n' = \rho_n$ only when $\mu_o = M_o$. The relationship between M_1 and M_o when $\rho_n' = 0$ depends upon the sign of κ_3. When a distribution is skewed to the right ($\kappa_3 > 0$), the median M is less than the mean μ, and the reverse inequality is true for a distribution skewed to the left ($\kappa_3 < 0$). For $\rho_n' = 0$ ($\mu_1 = M_o$), if $\kappa_3 > 0$, we have $M_1 < \mu_1 = M_o < \mu_o$ (and $p_1' < \frac{1}{2}$), and if $\kappa_3 < 0$, $\mu_o < M_o = \mu_1 < M_1$ ($p_1' > \frac{1}{2}$). Thus neither the null nor the alternative relationship remains valid for $\kappa_3 > 0$, $\rho_n' = 0$, and the alternative is true when $\kappa_3 < 0$, $\rho_n' = 0$. However, Table 2.1 shows that M_1 is close to μ_1 (which is equal to M_o), since p_1' is not too far from $\frac{1}{2}$ for all κ_3 and κ_4.

Table 2.1. $Pr(X_i > M_o | H_1') = Pr(X_i > \mu_1)$ for $\rho_n' = 0$, all values of κ_4

κ_3	-.6	-.4	-.2	0	.2	.4	.6
p_1'	.5399	.5266	.5133	.5000	.4867	.4734	.4601

For the <u>t</u> test, when $\rho_n > 0$, the one-sided alternative H_1 always holds. However, when $\rho_n' > 0$, the alternative H_1' may or may not be true, depending on the sign of κ_3 and the magnitude of ρ_n'. If $\rho_n' > 0$ and $\kappa_3 < 0$, we have $\mu_o < M_o < \mu_1 < M_1$ so that H_1' holds. However, if the distribution is skewed to the right, $\rho_n' > 0$ implies any one of three inequalities: $M_0 < \mu_o < M_1 < \mu_1$, $M_o < M_1 < \mu_o < \mu_1$, and $M_1 < M_o < \mu_1 < \mu_o$. The alternative H_1' holds in the first two cases only.

The power functions have been computed from (2.8), (2.9), (2.10), and (2.12) for $n = 10$, $\alpha = .05$, all combinations of $\rho_n' = 0(1)4$, $\kappa_3 = -.6(.2).6$, $\kappa_4 = -1(1)2$, and are presented in Table 2.2. The randomized decision rule for the sign test when $n = 10$, exact $\alpha = .05$ in all cases, is found from (2.9). We reject always when $r > 8$ and with probability 201/225 when $r = 8$. These computations of $F(y)$ and all further calculations involving the Edgeworth-Cramér

Table 2.2. Power of the sign test of H_o': $M = M_o$ versus H_1'': $M = M_1 > M_o$, for $n = 10$, $\rho_n' = \sqrt{n}\,(\mu_1 - M_o)/\sigma$, $\alpha = .05$

κ_4	ρ_n' \ κ_3	-.6	-.4	-.2	.0	.2	.4	.6
	0	.081	.070	.059	.050	.042	.035	.029
	1	.214	.202	.173	.169	.138	.130	.110
-1	2	.417	.416	.409	.395	.374	.348	.316
	3	.651	.662	.668	.668	.663	.652	.630
	4	.846	.857	.867	.875	.883	.889	.895
	0	.081	.070	.059	.050	.042	.035	.029
	1	.242	.229	.197	.194	.159	.151	.116
0	2	.480	.479	.471	.457	.434	.406	.372
	3	.714	.725	.730	.731	.725	.715	.699
	4	.877	.887	.896	.903	.910	.916	.921
	0	.081	.070	.059	.050	.042	.035	.029
	1	.273	.259	.224	.220	.183	.173	.148
1	2	.546	.545	.537	.521	.498	.469	.432
	3	.774	.785	.790	.790	.785	.775	.760
	4	.905	.914	.921	.928	.934	.939	.943
	0	.081	.070	.059	.050	.042	.035	.029
	1	.306	.291	.253	.249	.208	.198	.171
2	2	.613	.613	.605	.589	.565	.534	.496
	3	.831	.840	.845	.845	.840	.831	.818
	4	.929	.937	.944	.949	.954	.958	.962

expansion have been determined using the form (2.8). The Hermite polynomials have been evaluated with four decimal places in y when $y < 1$ and four significant figures when $y \geq 1$. $\varphi(y)$ and $\Phi(y)$ were obtained to five decimal places from tables (National Bureau of Standards, 1953). The actual power calculations from (2.10) were carried out on the IBM 1620 computer with four decimal places in p_1'. The program was written in FORTRAN language.

For all values of κ_4, the power increases as κ_3 increases when $\rho_n' = 4$, increases for increasing κ_3 for negative values of κ_3 and decreases with increasing κ_3 for positive κ_3 when $\rho_n' = 3$, and decreases with increasing κ_3 when $\rho_n' = 0,1$ and 2. As κ_4 increases, $F(x)$ defined by (2.7) decreases since $\varphi^{(3)}(x) < 0$ for $-\sqrt{3} < x < \sqrt{3}$; thus p_1' increases and therefore the power increases for all ρ_n' and all κ_3. The power for positive values of κ_3 is higher than the power for the corresponding negative value of κ_3 when $\rho_n' = 4$, for all values of κ_4. The opposite is true for all other values of ρ_n'.

The power functions are not directly comparable with Srivastava's results (Table 2, p. 427) because of the difference in hypotheses, the difference in interpretation of ρ_n and ρ_n' , and the fact that

Srivastava's significance level varies with κ_3 and κ_4 whereas here the α level is constant. In general, the power of the sign test compares favorably with Srivastava's results for the t test under the same alternative assumptions. The graphical comparisons in Figure 2.1 of the power of the sign test versus the t test for the different values of ρ_n or ρ'_n for the four most extreme cases considered ($\kappa_4 = -1$, $\kappa_3 = -.6$; $\kappa_4 = 2$, $\kappa_3 = -.6$; $\kappa_4 = -1$, $\kappa_3 = .6$ and $\kappa_4 = 2$, $\kappa_3 = .6$) demonstrate the close agreement between them, especially for $\kappa_3 = -.6$, $\kappa_4 = 2$.

Due to the disparity between the hypotheses to be tested, the differences in power cannot be attributed entirely to a distinction between the t test and the sign test. However, even though the results are not directly comparable, they are enlightening; when we assume normality as a requisite for performing the t test, the coincidence of the means and medians is implied, making H_o and H'_o equivalent. In spite of the fact that exact knowledge of the population distribution is absent in most practical situations, normality is often assumed.

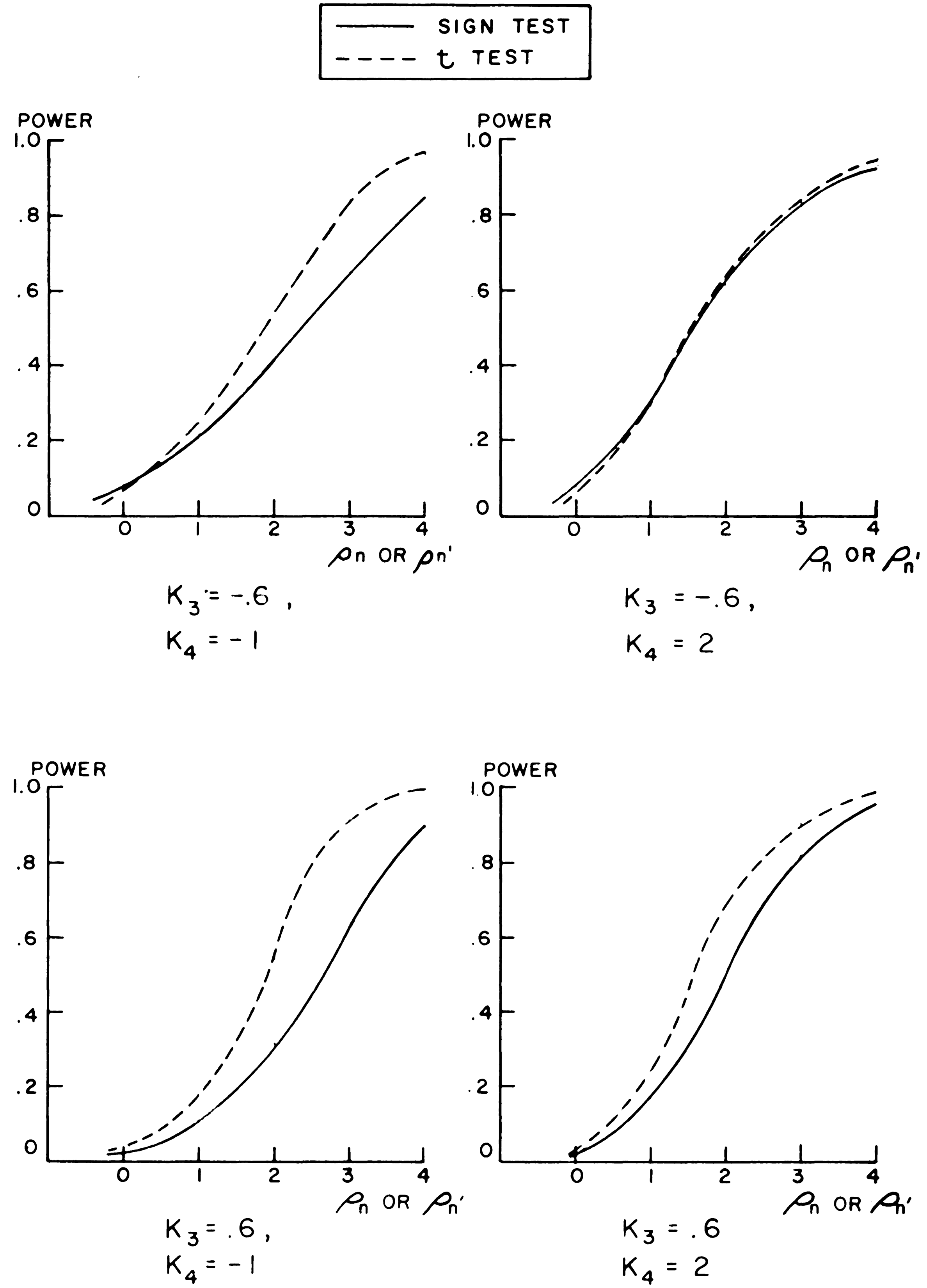

Figure 2.1. Comparisons of power between the t test and the sign test of H_0': $M=M_0$ versus H_1': $M=M_1>M_0$, n=10, $\rho_n'=\sqrt{n}(\mu_1-M_0)/\sigma$, α=.05

2.4 Sign Test on the Population Median as Approximated by the Population Mean and the Coefficient of Skewness

Comparisons of the power functions of the sign test and Student's t test for the approximately normal distribution defined by (2.7) might be more relevant theoretically if the difference between the two hypotheses could be minimized. One possible approach is to express the median as a function of the mean and the coefficient of skewness.

An approximately normal distribution with mean μ, variance σ^2, and standardized third cumulant κ_3 has the median

$$M = \mu - \kappa_3\sigma/6 \quad , \tag{2.13}$$

when terms of order at most 1/N are neglected. This result, due to Haldane (1942), may be established as follows. M is the solution of $F(M) = \frac{1}{2}$, where $F(M)$ is given by the first two terms of (2.8), i.e.,

$$\Phi(M) - \kappa_3(M^2-1)\varphi(M)/6 = \tfrac{1}{2} \quad .$$

Using the fact that

$$\Phi(M) = \Phi(0) + [\Phi(M) - \Phi(0)] = \tfrac{1}{2} + (\Phi(M) - \Phi(0)) \quad ,$$

we have

$$\int_0^M \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}\, dt - \kappa_3(M^2-1)\varphi(M)/6 = 0 \quad .$$

Then $e^{\frac{1}{2}M^2}\int_0^M e^{-\frac{1}{2}t^2}dt - \kappa_3(M^2-1)/6 = 0$.

Integrating by parts once, we obtain

$$M + e^{\frac{1}{2}M^2}\int_0^M t^2\, e^{-\frac{1}{2}t^2}dt = \kappa_3(M^2-1)/6 \quad . \qquad (2.14)$$

But since F(y) is approximately the normal distribution with zero mean and unit variance, the median M will be very close to zero. We will assume that M is at most of order $1/N^{\frac{1}{2}}$. Then, neglecting the terms of order at most 1/N in (2.14), we obtain $M = -\kappa_3/6$ for the median of the distribution of the standardized variable Y. The relationship (2.13) then follows immediately for the distribution of X.

Using the relationship given by (2.13), we would test the null hypothesis H_o'': $M = M_o = \mu_o - \kappa_3\sigma/6$ against the alternative H_1'': $M = M_1 = \mu_1 - \kappa_3\sigma/6$, $M_1 > M_o$, so that H_o'' and H_1'' are roughly the same as H_o and H_1. In a practical situation, then, some estimate of $\kappa_3\sigma$ is necessary, but this may be possible from previous experience with the same type of data. A comparison of the power functions for H_o and H_o'' is appropriate in the case where (1) the t test is used to test the null hypothesis H_o, even though we know, or are willing to assume, that the distribution is skewed but still approximately normal, in preference to some nonparametric test of location, and (2) the sign test is used to test the null hypothesis H_o'' so that no assumptions need be made about the distribution (except for the value of $\kappa_3\sigma$) to perform the test, but the same assumptions as for H_o are made to determine the power.

If M_o is the true median of the distribution, $Pr(X_i > M_o) = \frac{1}{2}$ regardless of the type of density function. However, M_o is only approximately equal to $\mu_o - \kappa_3\sigma/6$, where μ_o is the true mean, since the terms of order at most $1/N$ have been neglected in obtaining this approximation. Let

$$p_o'' = Pr(X_i > M_o | H_o'') = Pr[(X_i - \mu_o)/\sigma > (M_o - \mu_o)/\sigma | H_o'']$$

$$= Pr[(X_i - \mu_o)/\sigma > -\kappa_3/6 | H_o''] \ . \qquad (2.15)$$

If the null hypothesis is true, we would expect np_o'' of the n differences $(X_i - \mu_o + \kappa_3\sigma/6)$ to be positive and $n(1 - p_o'')$ to be negative. The sign test is to reject when $r > r_o$, where $\sum_{r=r_o+1}^{n} \binom{n}{r} p_o''^{r}(1-p_o'')^{n-r} = \alpha$. Now this implies that a different critical region is needed for each value of κ_3 to have an exact significance level of α. Furthermore, we must assume the population distribution function of the X_i under the null hypothesis in order to perform the sign test, which is typically a nonparametric test. If we assume that the standardized X_i are approximately normally distributed according to (2.7) under H_o'', $p_o'' = 1 - F(-\kappa_3/6)$. Table 2.3 shows that p_o'' ranges between .4925 and .5075 when $-1 \leq \kappa_4 \leq 2$, $-.6 \leq \kappa_3 \leq .6$, and is exactly equal to .5000 only when $\kappa_3 = 0$. Since these values of p_o'' are very close to ½, and since in actual practice the sign test would be performed assuming that $p_o'' = ½$ so that it is a nonparametric test, we will use the single critical region determined by (2.9). Then the test for $n = 10$, $\alpha = .05$, is exactly the same as for H_o', i.e., to reject always when $r > 8$ and reject with probability 201/225 if $r = 8$.

The power function of the sign test against the alternative H_1'' is given by (2.8) and (2.10), with p_1' replaced by p_1''. Here

$$p_1'' = Pr(X_i > M_o|H_1'') = Pr[(X_i-\mu_1)/\sigma > (M_o-\mu_1)/\sigma|H_1'']$$

$$= Pr[(X_i-\mu_1)/\sigma > (\mu_o-\mu_1)/\sigma-\kappa_3/6|H_1''] = 1-F[(\mu_o-\mu_1)/\sigma-\kappa_3/6]$$

$$= 1 - F(-\rho_n/\sqrt{n} - \kappa_3/6) \quad , \qquad (2.16)$$

with ρ_n defined as before for H_o. When $\rho_n = 0$, $p_1'' = p_o''$ so that the entries in Table 2.3 can also be interpreted as the values of $p_1'' = Pr(X_i > M_o|H_1'', \rho_n = 0)$. These p_1'' values are much closer to ½ than the corresponding entries for p_1' in

Table 2.3. $\Pr(X_i > M_o | H_o'')$ or $\Pr(X_i > M_o | H_1'', \rho_n = 0)$

		κ_3						
	κ_4	-.6	-.4	-.2	0	.2	.4	.6
p_1''	-1	.5074	.5040	.5018	.5000	.4982	.4960	.4926
	0	.5024	.5007	.5001	.5000	.4999	.4993	.4976
	1	.4975	.4974	.4984	.5000	.5016	.5026	.5025
	2	.4925	.4941	.4968	.5000	.5032	.5059	.5075

Table 2.1 for H_1'. When $\rho_n = 0$, $\mu_1 = \mu_o$ and $M_1 = M_o$, so that the null hypothesis H_o is true; H_o'' is approximately true, since M_o is quite close to the true median of the distribution $F(x)$. The power functions of the sign test of H_o'' versus H_1'' have been computed for the various combinations of κ_3 and κ_4, $n = 10$, $\rho_n = 0(1)4$, and are presented in Table 2.4. The exact significance level corresponds to the entries for $\rho_n = 0$. It ranges between .045 and .055.

The results in Table 2.4 seem to indicate that, in general, changes in κ_4 tend to have more effect on the power function than changes in κ_3. This is especially evident for small values of ρ_n. The power increases as κ_3 increases for all κ_4 and all ρ_n except $\rho_n = 0, 1$. The values for $\rho_n = 0$ are all close to .05 since p_1 is close to ½ when $\rho_n = 0$. Like the power function of H_o' in Table 2.2, the power

Table 2.4. Power of the sign test of H_o'': $M = M_o = \mu_o - \frac{1}{6}\kappa_3\sigma$ versus H_1'': $M = M_1 = \mu_1 - \frac{1}{6}\kappa_3\sigma > M_o$ for $n = 10$, $\rho_n = \sqrt{n}(\mu_1-\mu_o)/\sigma$, $\alpha = .05$

κ_4	ρ_n \ κ_3	-.6	-.4	-.2	0	.2	.4	.6
	0	.055	.053	.051	.050	.049	.047	.045
	1	.164	.166	.169	.169	.168	.165	.159
-1	2	.346	.366	.382	.395	.404	.408	.410
	3	.577	.612	.642	.668	.691	.713	.734
	4	.793	.823	.853	.875	.898	.920	.940
	0	.052	.050	.050	.050	.050	.050	.048
	1	.181	.186	.191	.194	.195	.193	.188
0	2	.401	.423	.442	.457	.467	.473	.475
	3	.645	.679	.707	.731	.751	.770	.788
	4	.835	.861	.883	.903	.922	.939	.955
	0	.048	.048	.049	.050	.051	.052	.052
	1	.199	.207	.214	.220	.224	.224	.221
1	2	.459	.484	.505	.521	.533	.540	.542
	3	.712	.743	.769	.790	.808	.823	.838
	4	.873	.894	.911	.928	.942	.956	.968
	0	.045	.046	.048	.050	.052	.054	.055
	1	.218	.229	.240	.249	.256	.258	.257
2	2	.519	.547	.571	.589	.601	.609	.611
	3	.777	.804	.827	.845	.859	.871	.883
	4	.908	.924	.935	.949	.960	.970	.979

increases for all ρ_n and all κ_3 as κ_4 increases, and for the same reason as stated there. The power functions in Table 2.4 are greater than the corresponding values in Table 2.2 in all cases for which $\kappa_3 > 0$, and less than the corresponding values for $\kappa_3 < 0$. They are, of course, the same when $\kappa_3 = 0$.

A comparison with Srivastava's results reveals that although the power of the sign test of H_o'' is much less than the power of the <u>t</u> test of H_o when $\kappa_3 < 0$, the gap is narrowed considerably when $\kappa_3 > 0$. Even these results are difficult to compare directly because of the larger variation in significance level for the <u>t</u> test. The close agreement of the power functions for $\kappa_3 > 0$ is demonstrated graphically in Figure 2.2 for the cases $\kappa_3 = .4$, $\kappa_4 = 2$ and $\kappa_3 = .6$, $\kappa_4 = 2$.

2.5 Sign Test on the Mean

There is still another possibility for minimizing the difference between the hypothesis H_o for the <u>t</u> test and the hypothesis to be tested using the sign test, and this is to make the two hypotheses exactly the same. This might be called a sign test on the mean. Thus we are testing the

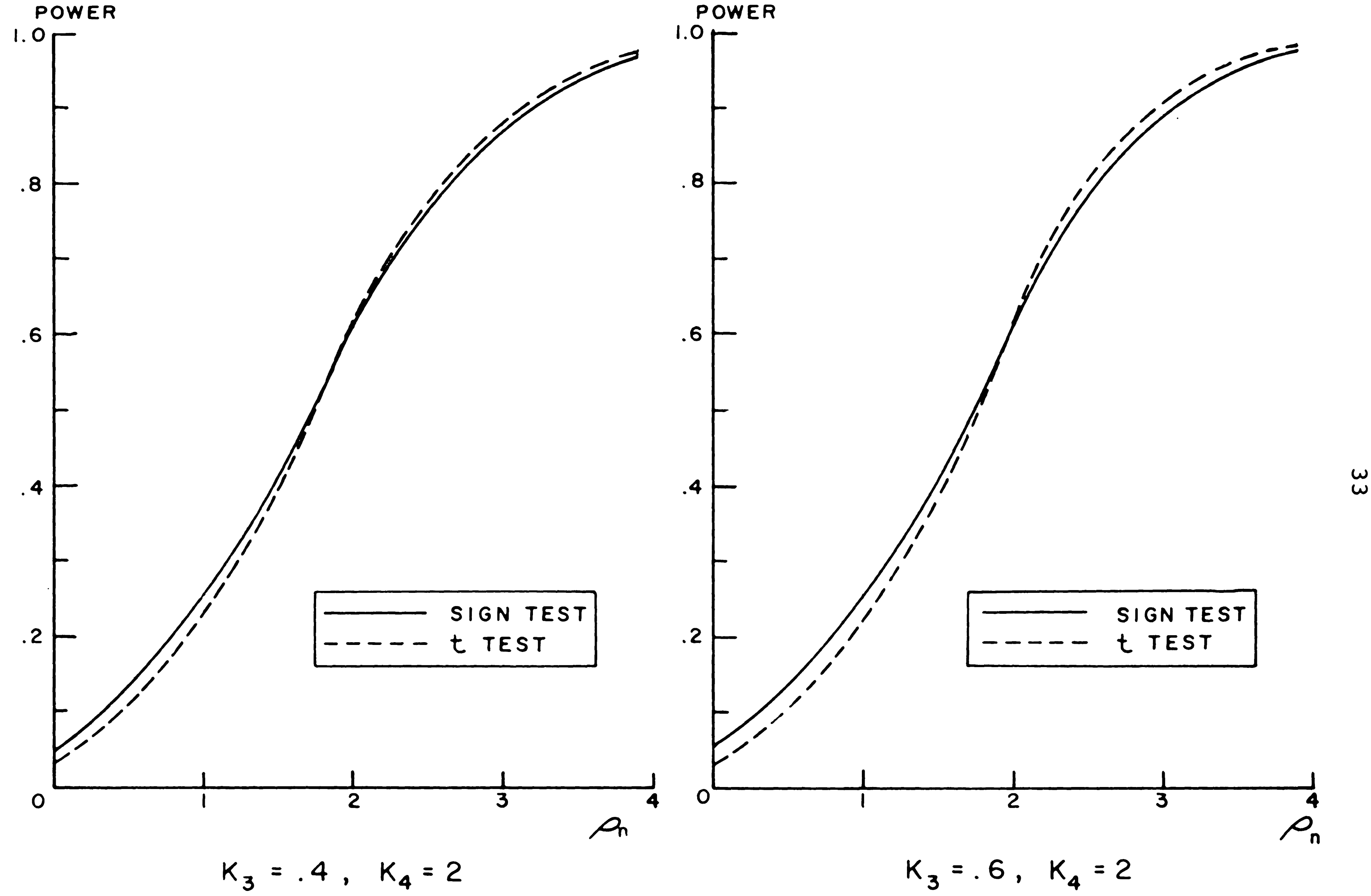

Figure 2.2. Comparisons of power between the t test and the sign test of H_o'': $M=M_o=\mu_o-\frac{1}{6}\kappa_3\sigma$ versus H_1'': $M=M_1=\mu_1-\frac{1}{6}\kappa_3\sigma>M_o$, $n=10$, $\rho_n=\sqrt{n}(\mu_1-\mu_o)/\sigma$, $\alpha=.05$

null hypothesis $H_o: \mu = \mu_o$ against the alternative $H_1: \mu = \mu_1 > \mu_o$. Since we must assume that the n random variables are independent and approximately normally distributed with mean μ and variance σ^2 even to perform the sign test of H_o, this test no longer belongs in the category of nonparametric tests. The sign test on the mean would be to reject if there are too many plus signs among the n differences $(X_i - \mu_o)$. The problem is how to define "too many" in order to determine the critical region.

Let us use the notation $p_o = Pr(X_i > \mu_o | H_o)$ and $p_1 = Pr(X_i > \mu_o | H_1)$. Then the p_o for the various combinations of κ_3 and κ_4 are the same as those given in Table 2.1 for H_o', and the critical region for an exact significance level α would differ for each value of κ_3. In a practical situation, we would probably use the sign test determined by $p_o = \frac{1}{2}$, realizing that this implies that the significance levels vary with the value of κ_3. The p_1 values are given by $p_1 = Pr[(X_i - \mu_1)/\sigma > (\mu_o - \mu_1)/\sigma] = 1 - F(-\rho_n/\sqrt{n})$. For $n = 10$, then, the power functions for this test are the same as those given by Table 2.2 for the various combinations of κ_3 and κ_4 with ρ_n' replaced by ρ_n, and the significance levels are represented by the table entries for $\rho_n = 0$. They range between .029 and .081.

The power function is still difficult to compare with Srivastava's results since now the significance level is varying for both tests.

2.6 Comparisons of Power for Equal Significance Levels

Another interesting theoretical comparison between the power functions of the sign test and the *t* test can be made by equalizing the significance levels as well as the hypotheses. Srivastava has calculated the power of the *t* test when the "true" significance level is .05 for $n = 10$, $\kappa_3 = .6$, $\kappa_4 = .4$, and $\kappa_3 = \kappa_4 = 0$, $\rho_n = 0(1)4$ and $\rho_n = -4(1)0$. The negative values of ρ_n apply if we are considering a one-sided alternative where $\mu_1 < \mu_o$. For the last test discussed, the sign test on the mean, $p_o = .4601$ when $\kappa_3 = .6$ and $\kappa_4 = .4$. The sign test for an exact .05 significance level when $\rho_n > 0$, found using (2.9), is to reject the null hypothesis H_o always when $r > 7$ and reject with probability .22141 if $r = 7$. The power is given by (2.10) and (2.8). When ρ_n is negative, we should reject the null hypothesis when there are too few plus signs (or, equivalently, too many minus signs) among the n differences $(X_i - \mu_o)$. Using the same notation as before, the test for $n = 10$, $\alpha = .05$, is to reject always if $r < 2$ and reject with probability .43474 if $r = 2$.

For the previously considered sign test of H_o'', $p_o'' = .4995$ when $\kappa_3 = .6$, $\kappa_4 = .4$, and the .05 level sign test would reject always if $r > 8$ and reject with probability .90064 if $r = 8$ for $\rho_n > 0$. When $\rho_n < 0$, the test is to reject if $r < 2$ and with probability .88608 if $r = 2$.

The results for both of these tests are presented in Table 2.5, along with the power of the t test of H_o when the population density is given by (2.6) with $\kappa_3 = .4$, $\kappa_4 = .6$. The power functions of the sign test and the t test for an exact normal distribution (i.e., $\kappa_3 = \kappa_4 = 0$) are also given.

The sign test compares quite favorably, especially the test of H_o''. The power for the sign test of H_o when ρ_n is negative is considerably lower than for positive ρ_n. However, the rejection region is much smaller, since the probability of a minus sign under the null hypothesis is .5399.

Table 2.5. Comparisons of power between t test and sign test when "true" $\alpha = .05$ and $n = 10$, for an exact normal distribution and for $\kappa_3 = .6$, $\kappa_4 = .4$

ρ_n or ρ'_n	$\kappa_3=\kappa_4=0$ t Test	Sign Test	$\kappa_3=.6,\ \kappa_4=.4$ t Test		Sign Test	
	$t_o=1.833$	of H_o, H'_o, H''_o	$t_o=1.833$	$t_o=1.627$	of H_o	of H''_o
0	.050	.050	.035	.051	.050	.050
1	.236	.194	.198	.259	.193	.202
2	.580	.457	.579	.663	.481	.504
3	.868	.731	.910	.940	.792	.810
4	.979	.903	.997	.9996	.958	.962
	$t_o=-1.833$	of H_o, H'_o, H''_o	$t_o=-1.833$	$t_o=-2.076$	of H_o	of H''_o
0	.050	.050	.070	.051	.050	.050
-1	.236	.194	.273	.219	.171	.187
-2	.580	.457	.585	.530	.375	.422
-3	.868	.731	.838	.784	.605	.670
-4	.979	.903	.957	.933	.790	.849

Chapter III

SMALL-SAMPLE POWER OF RANK TESTS ON THE EQUALITY OF TWO DISTRIBUTION FUNCTIONS

Nonparametric tests based on ranks are especially simple to use and can be applied even when no measurement is possible, since the ranks of the random variables constitute the new random variables on which the test is performed. Consider two samples of sizes m and n of independent random variables, $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$, where the X's and Y's are identically distributed with continuous cumulative distribution functions H and G respectively. The null hypothesis H_o to be tested is that $H = G$, their common cumulative distribution function being unspecified. A rejection of the null hypothesis might indicate that the densities differ in shape, location, scale, or a combination of these.

Many rank tests have been proposed to test the equality of two distribution functions. Which one of these is most appropriate in any given situation will depend partly on computational simplicity, but more basically upon what type of alternative is of interest to the investigator. The power of competing tests against the chosen alternative will therefore be important in the selection of a test. Often the class of relevant alternatives is one-sided in form, $H(a) \geq G(a)$ for all a, in which case the Y's are said to be stochastically larger than the X's. The null hypothesis H_o will then be rejected if most of the Y's are larger than most of the X's. For example, if X and Y represent measures on a control and a treated group, respectively, where the treatment is expected to increase the measure, or at least leave it unchanged, we would be interested in an alternative of this form.

The power of any nonparametric test based on order or ranks can be computed, at least in theory, for any alternative relating H and G. The probability of any particular ordering of the combined sample of m X's and n Y's arranged in ascending order is independent of the specific form of H or G, provided that G is a function of H, or both are functions of some common distribution function F. The alternative distribution functions must be specified completely enough to determine these probabilities. The power would be the sum of the probabilities for the orderings contained in the region of rejection. The rejection orderings are determined by the test, the significance level, and the fact that the probability of any arrangement of the $N = (m+n)$ random variables is equal to $1/\binom{N}{m}$ under the null hypothesis.

Lehmann (1953) has considered a nonparametric alternative of one extreme distribution (see also Savage, 1956). That is, the Y random variables are distributed as the largest of k of the X variates, where the meaning of k may be extended allowing it to be <u>any</u> positive number. Using this alternative, he has determined the power of six well-known two-sample rank tests -- the most powerful one-sided rank test, the Wilcoxon or Mann-Whitney U test, the one and two-sided median tests, the Wald-Wolfowitz runs test, and the two-sided Wilcoxon test. In this chapter a similar alternative will be considered -- the Y's are distributed as the largest of k variables from some unspecified distribution F, and the X's are distributed as the smallest of k from this same distribution function. This also belongs to the nonparametric class of alternatives, since the probability of any particular ordering under the alternative is independent of the specific form of F, provided that F is continuous. Methods for determining the probabilities of rank orders under this alternative will be found and used to calculate the small-sample power of Terry's c_1 test and a new test, called the Psi test, in addition to the six tests considered by Lehmann.

The other alternative to be considered in this chapter is that the X and Y random variables are both normally distributed with the same variance but different means. The power functions of the tests are given for those significance levels which require only one or both of the two extreme orderings in the critical region. Comparisons are made between power functions for these alternatives.

3.1 The Alternative of One Extreme Distribution

If $P(x)$ is the cumulative distribution function of a variate X, then the cumulative distribution of the largest of k variables drawn from this distribution is $P^k(x)$. When $P(x)$ is known, an expression for $P^k(x)$ can always be found, although it may be complicated. The extreme distribution may differ from the original distribution by a shift in location, scale, form, or some combination of these. However, Fisher and Tippett (1928) have shown that the limiting cumulative distribution of the largest of k observations satisfies the functional equation $P^k(x) = P(a_k x + b_k)$, which has only three classes of solutions for $P(x)$. The argument here is that the limiting distribution $P(a_k x + b_k)$ of the largest of $n = km$ observations must be of the same form as the limiting distribution $P^k(x)$ of the largest of a sample of size k drawn from the largest of samples of size m, as m tends to infinity and k is held fixed, except for changes in location or scale. The solution which is called Type I has $a_k = 1$ so that $P^k(x) = P(x + b_k)$. Then P is an exponential-type distribution with $dP = \exp(-x - e^{-x})dx$. The limiting distribution of the largest value in a sample of k is of the same form but shifted a distance of log k. In the other two types of limiting extreme distributions, the scale is changed.

The normal distribution belongs to Type I, along with many other important distributions. In comparison with the normal distribution, the Type I density is slightly skewed to the right and platykurtic. Although the distribution of the extreme from a normal distribution converges slowly toward the Type I asymptotic distribution, the difference between the exact and asymptotic densities for a sample of size 100 is negligible (see, e.g., Graph 6.2.1(4), p. 222, Gumbel, 1958).

The alternative of one extreme distribution considered by Lehman expresses the relationship between H and G as

$G = H^k$. If H belongs to the Type I class, the asymptotic distribution of G is of the same form as H but shifted linearly, so that asymptotically the alternative expresses a change in location. Lehmann (1953, Table I, p. 29) has calculated the exact power functions of six rank tests when $m = n = 4$, $m = n = 6$, for $k = 2$, $k = 3$, with significance level .10. The one-sided tests which he considers are the most appropriate for this alternative since they are designed principally to detect the situation where the Y's are stochastically larger than the X's. Here $Pr(X<Y) = k/(k+1)$.

3.2 The Alternative of Two Extreme Distributions

Although the limiting distributions of H and G when $G = H^k$ may have the same shape, smaller values of k often effect a considerable change in the distribution functions. The differences between H and G are illustrated in Lehmann's article for $k = 2$, 3, and 6, when the density of H is normal, exponential, and uniform (pp. 26-28). In a practical situation, we usually like to think of H and G as being more similar under the alternative. An alternative which meets this requirement in many situations and has some desirable properties is the alternative H_1 of two extreme distributions, H and G both being functions of some common distribution function F. We will assume not only that the Y's have a distribution function $G = F^k$, i.e., Y is distributed as the largest of k variables from some unspecified distribution F, but also that the X's are distributed as the smallest of k from this same distribution, $H = 1 - (1-F)^k$. For any value of k, $Pr(X<Y) = 1 - kB(k+1,k)$, where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, which is a strictly increasing function of k. Thus the alternative expresses the fact that the Y's are stochastically considerably larger than the X's.

3.2.1 Properties

Both of the alternative distributions change their shape and location according to the value of k, but they remain "mutually symmetrical" whenever the density of F is symmetrical. We will say that two distribution functions H(x) and G(x) are mutually symmetric if there exists a constant <u>a</u> such that $H(x-a) = 1 - G(a-x)$ for all x. If the corresponding density functions h(x) and g(x) exist, an equivalent definition is that $h(x-a) = g(a-x)$ for some <u>a</u> and all x. We can assume without loss of generality that <u>a</u> is equal to zero.

For the particular case of two extreme distributions, let us assume that F is symmetrical about zero. The density function of H is

$$h(x) = k[1-F(x)]^{k-1}f(x) = k[F(-x)]^{k-1}f(-x) = g(-x). \quad (3.1)$$

Then $\mathcal{E}_H(x^p) = \mathcal{E}_G[(-x)^p]$, which implies that the even moments of H and G are equal, and the odd moments differ only with respect to sign. More descriptively, the density functions of the smallest and largest values are mirror images, and, regardless of the value of k, G and H are related to each other through their moments. Figure 3.1 illustrates the densities of F, G and H for k = 2 when f(x) is normal, uniform, and exponential.

Actually, the property of mutual symmetry applies for the more general case in which H is the cumulative distribution function of the r th order statistic of k, $r=1,2,\ldots,[\frac{1}{2}k]$, and G is the distribution of the (k-r+1)th order statistic of k. Suppose that the k independent random variables are identically distributed with density function f(x), and f(x) is symmetrical about $\mathcal{E}(x)$. We can assume without loss of generality that $\mathcal{E}(x) = 0$. Let $f_r(x)$ denote the density function of the r th order statistic. Then

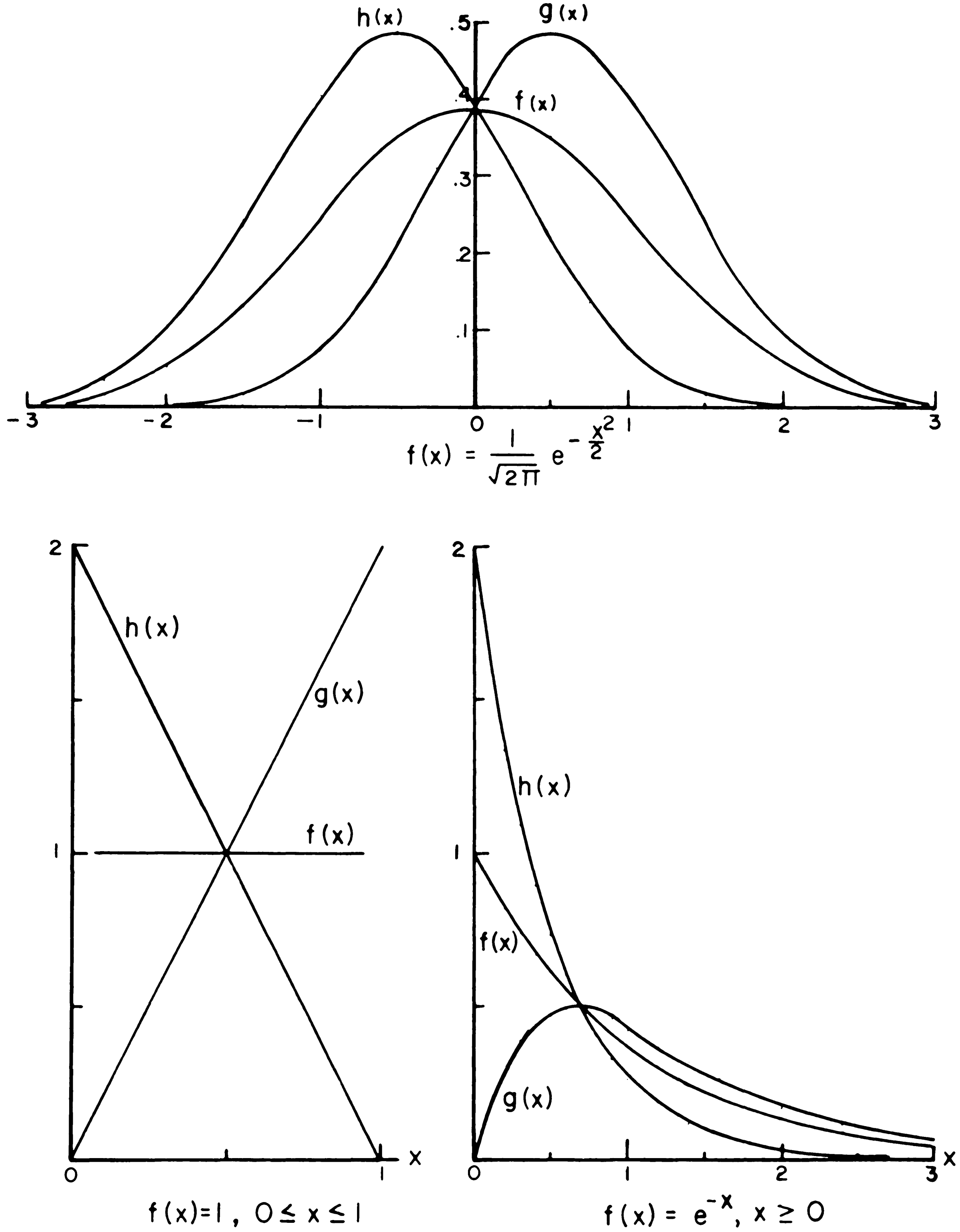

Figure 3.1. Density functions of F, $G=F^2$ and $H=1-(1-F)^2$ when F is normal, uniform, and exponential

$$f_r(x) = \frac{k!}{(r-1)!(k-r)!}[F(x)]^{r-1}[1-F(x)]^{k-r}f(x)$$

$$= \frac{k!}{(r-1)!(k-r+1-1)!}[1-F(-x)]^{r-1}[F(-x)]^{k-r+1-1}f(-x)$$

$$= f_{k-r+1}(-x) \quad . \tag{3.2}$$

This holds true theoretically even when k is not an integer provided that k is a positive number (the coefficient $k!/[(r-1)!(k-r)!]$ must be replaced by $1/B(r,k-r+1)$). However, the interpretation in terms of order statistics is not very meaningful.

For the case in which f is not symmetrical, relationships between the moments are more difficult to assess. It is interesting to note, however, that for $k = 2$, the pth moment of H plus the pth moment of G is equal to twice the pth moment of F, regardless of the character of F or the value of p, as long as the moments are all taken about some common point.

We can assume without loss of generality that f is a symmetrical density function. This will have no effect on the power functions, since the probability of any rank ordering is independent of the specific character of F as long as it is continuous.

3.2.2 Probabilities of Rank Orders

Under the null hypothesis $H = G$, all possible orderings of the m X's and n Y's ($m+n=N$) in the combined sample are equally likely and occur with probability $m!n!/N!$. The objective is to find a practical method of calculating the probability of any arrangement under the alternative H_1.

A very general theorem due to Hoeffding (1951) provides one expression for these order probabilities.

Theorem 3.1. Consider a sample of m X's and n Y's, each independent and identically distributed with continuous cumulative distribution functions H and G respectively. Let h and g denote their density functions. Under the alternative $H = Q(G)$, the probability of any arrangement of the X's and Y's is given by

$$\mathcal{E}[\prod_{j=1}^{m} q(U_{r_j})]/\binom{N}{m} \quad , \tag{3.3}$$

where q is the density function of Q, $0 \leq U_1 \leq U_2 \leq \ldots \leq U_N \leq 1$ are the order statistics of a random sample of size N ($m+n=N$) from the uniform distribution, and the r_j ($j=1,2,\ldots,m$) are the ranks of the X random variables in the combined sample. Two arrangements will be considered the same regardless of permutations among the X's or among the Y's.

Proof: We may assume without loss of generality that $g(x) = 1$, since if it is not, we can use the probability integral transformation to make it uniform. Then $H = Q$ and $h = q$. Let $s_1, s_2, \ldots, s_n$ denote the ranks of the Y's. One possible general ordering is

$$X_1 < \ldots < X_{s_1-1} < Y_1 < X_{s_1} < \ldots < X_{s_2-2} < Y_2 < \ldots$$
$$\ldots < Y_{n-1} < X_{s_{n-1}-n+2} < \ldots < X_{s_n-n} < Y_n < X_{s_n-n+1} < \ldots < X_m ,$$

and there are $m!n!$ of these same orderings corresponding to permutations among the m X's and among the n Y's. The probability of this arrangement then is

$$m!n! \Pr(X_1 < \ldots < X_{s_1-1} < Y_1 < X_{s_1} < \ldots < Y_n < X_{s_n-n+1} < \ldots < X_m)$$

$$= \frac{m!n!}{N!} \int_0^1 \int_0^{u_N} \int_0^{u_{N-1}} \ldots \int_0^{u_2} N! \prod_{j=1}^{m} h(u_{r_j})\, du_1 du_2 \ldots du_N \;, \quad (3.4)$$

since the density function is $g(u) = 1$ for each integral corresponding to a Y random variable. The multiple integral in (3.4) is $\mathcal{E}[\prod_{j=1}^{m} q(U_{r_j})]$ and thus the proof is complete.

Applying the theorem to the alternative $H = 1 - (1-F)^k$, $G = F^k$, let $u = F^k$ and $H(u) = 1 - (1-u^{1/k})^k$, then $h(u) = (1-u^{1/k})^{k-1}\, u^{(1/k)-1}$. For example, for $m = n = 2$, k = 2, we have

$$Pr(1010) = 2!2! \int_0^1 \int_0^{u_4} \int_0^{u_3} \int_0^{u_2} (1-u_1^{\frac{1}{2}}) u_1^{-\frac{1}{2}} du_1 du_2 (1-u_3^{\frac{1}{2}}) u_3^{-\frac{1}{2}} du_3 du_4 \ .$$

On the left hand side a 1 denotes an X random variable and 0 denotes a Y random variable, so that Pr(1010) indicates the probability that the X's and Y's alternate in the combined ordered sample.

For general k, m, and n, the calculations are considerably less tedious if we use the following formulae which are derived by changing the expressions $(1-F)^k$ to $\sum_{j=0}^{k} \binom{k}{j} (-F)^j$ (under the assumption that k is an integer), and integrating. The formulae can be extended with ease to the case of more than six groups of X's and Y's.

Let $m = a + c + e$, $n = b + d + f$.

Pr[a X's < b Y's < c X's < d Y's < e X's < f Y's] =

$$m!n!\,k^{m+n-1}\sum_{a_1=1}^{k}\binom{k}{a_1}(-1)^{a_1-1}\sum_{a_2=0}^{k-1}\frac{\binom{k-1}{a_2}(-1)^{a_2}}{a_1+a_2+1}\sum_{a_3=0}^{k-1}\frac{\binom{k-1}{a_3}(-1)^{a_3}}{a_1+a_2+a_3+2}$$

$$\cdot\ \ldots\ \cdot\sum_{a_a=0}^{k-1}\frac{\binom{k-1}{a_a}(-1)^{a_a}}{a_1+a_2+..+a_a+(a-1)}\cdot\frac{1}{a_1+a_2+..+a_a+(a-1)+k}$$

$$\cdot\frac{1}{a_1+a_2+..+a_a+(a-1)+2k}\ \cdots\ \cdot\frac{1}{\Sigma a_i+(a-1)+bk}$$

$$\cdot\sum_{c_1=0}^{k-1}\frac{\binom{k-1}{c_1}(-1)^{c_1}}{\Sigma a_i+(a-1)+bk+c_1+1}\sum_{c_2=0}^{k-1}\frac{\binom{k-1}{c_2}(-1)^{c_2}}{\Sigma a_i+(a-1)+bk+c_1+c_2+2}$$

$$\cdot\ \ldots\ \cdot\sum_{c_c=0}^{k-1}\frac{\binom{k-1}{c_c}(-1)^{c_c}}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c}\cdot\frac{1}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c+k} \tag{3.5}$$

$$\cdot\frac{1}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c+2k}\ \cdots\ \cdot\frac{1}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c+dk}$$

$$\cdot\sum_{e_1=0}^{k-1}\frac{\binom{k-1}{e_1}(-1)^{e_1}}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c+dk+e_1+1}\ \cdots\ \cdot\sum_{e_e=0}^{k-1}\frac{\binom{k-1}{e_e}(-1)^{e_e}}{\Sigma a_i+(a-1)+bk+\Sigma c_i+c+dk+\Sigma e_i+e}$$

$$\cdot\frac{1}{\Sigma a_i+(a-1)+bk+\Sigma c_i+dk+\Sigma e_i+e+k}\ \cdots\ \cdot\frac{1}{\Sigma a_i+(a-1)+bk+\Sigma c_i+dk+\Sigma e_i+e+fk}$$

Let $b + d + f = n$, $c + e + g = m$.

$Pr[b\ Y\text{'s} < c\ X\text{'s} < d\ Y\text{'s} < e\ X\text{'s} < f\ Y\text{'s} < g\ X\text{'s}]$

$$= m!n!\,\frac{k^{m+n}}{(k)(2k)\ldots(bk)} \sum_{c_1=0}^{k-1} \frac{\binom{k-1}{c_1}(-1)^{c_1}}{bk+c_1+1} \cdots \sum_{c_c=0}^{1} \frac{\binom{k-1}{c_c}(-1)^{c_c}}{bk+c_1+c_2+\ldots+c_c+c}$$

$$\cdot \frac{1}{bk+\Sigma c_i+c+k} \cdot \frac{1}{bk+\Sigma c_i+c+2k} \cdot \ldots \cdot \frac{1}{bk+\Sigma c_i+c+dk}$$

$$\cdot \sum_{e_1=0}^{k-1} \frac{\binom{k-1}{e_1}(-1)^{e_1}}{bk+\Sigma c_i+c+dk+e_1+1} \cdots \sum_{e_e=0}^{k-1} \frac{\binom{k-1}{e_e}(-1)^{e_e}}{bk+\Sigma c_i+c+dk+\Sigma e_i+e} \qquad (3.6)$$

$$\cdot \frac{1}{bk+\Sigma c_i+c+dk+\Sigma e_i+e+k} \cdot \ldots \cdot \frac{1}{bk+\Sigma c_i+c+dk+\Sigma e_i+e+fk}$$

$$\cdot \sum_{g_1=0}^{k-1} \frac{\binom{k-1}{g_1}(-1)^{g_1}}{bk+\Sigma c_i+c+dk+\Sigma e_i+e+fk+g_1+1} \cdots \sum_{g_g=0}^{k-1} \frac{\binom{k-1}{g_g}(-1)^{g_g}}{bk+\Sigma c_i+c+dk+\Sigma e_i+e+fk+\Sigma g_i+g}$$

The arithmetic operations can be simplified in four general cases by using the following formulae when applicable. In the initial statement of each of the probabilities, $F(x)$ will be represented by x and $G(y)$ by y.

$$\Pr(\text{all } m\ X\text{'s} < \text{all } n\ Y\text{'s}) = m\int_{-\infty}^{\infty}[1-G(x)]^{n}[H(x)]^{m-1}\,h(x)\,dx$$

$$= m\int_0^1 (1-x^k)^n[1-(1-x)^k]^{m-1}\,k(1-x)^{k-1}\,dx$$

$$= m\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\int_0^1(1-x^k)^n\,(1-x)^{ki}\,k(1-x)^{k-1}\,dx$$

$$= km\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\sum_{j=0}^{n}\binom{n}{j}(-1)^j\int_0^1 x^{jk}(1-x)^{ki+k-1}\,dx$$

$$= km\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\sum_{j=0}^{n}\binom{n}{j}(-1)^j\,B(jk+1,ki+k) \tag{3.7}$$

$$\Pr[(m-1)\ X\text{'s} < 1\ Y < (p-1)\ Y\text{'s} < 1\ X < (n-p)\ Y\text{'s}]$$

$$= mn\binom{n-1}{p-1}\int_{-\infty}^{\infty}\int_{-\infty}^{x}[H(y)]^{m-1}g(y)[G(x)-G(y)]^{p-1}\cdot h(x)[1-G(x)]^{n-p}\,dy\,dx$$

$$= k^2mn\binom{n-1}{p-1}\int_0^1\int_0^x[1-(1-y)^k]^{m-1}\,y^{k-1}(x^k-y^k)^{p-1}(1-x)^{k-1}\cdot(1-x^k)^{n-p}\,dy\,dx$$

$$= k^2mn\binom{n-1}{p-1}\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\int_0^1\int_0^x(1-y)^{ik}y^{k-1}(x^k-y^k)^{p-1}\cdot(1-x)^{k-1}(1-x^k)^{n-p}\,dy\,dx$$

$$= kmn\binom{n-1}{p-1}\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\int_0^1\int_0^{x^k}(1-v^{1/k})^{ik}(x^k-v)^{p-1}(1-x)^{k-1}\cdot(1-x^k)^{n-p}\,dv\,dx$$

$$= kmn\binom{n-1}{p-1}\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\sum_{j=0}^{ik}\binom{ik}{j}(-1)^j\int_0^1\int_0^{x^k}v^{j/k}(x^k-v)^{p-1}\cdot(1-x)^{k-1}(1-x^k)^{n-p}\,dv\,dx$$

$$= kmn\binom{n-1}{p-1}\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\sum_{j=0}^{ik}\binom{ik}{j}(-1)^j\,B[p,(j/k)+1]\sum_{w=0}^{n-p}\binom{n-p}{w}(-1)^w\cdot\int_0^1 x^{kw}(1-x)^{k-1}[x^k]^{p+(j/k)}\,dx$$

$$= kmn\binom{n-1}{p-1}\sum_{i=0}^{m-1}\binom{m-1}{i}(-1)^i\sum_{j=0}^{ik}\binom{ik}{j}(-1)^j\,B[p,(j/k)+1]\sum_{w=0}^{n-p}\binom{n-p}{w}(-1)^w\cdot B(k,j+kw+kp+1) \tag{3.8}$$

$$\Pr[(n-1)\ Y\text{'s} < 1\ X < (p-1)\ X\text{'s} < 1\ Y < (m-p)\ X\text{'s}]$$

$$= mn\binom{m-1}{p-1}\int_{-\infty}^{\infty}\int_{x}^{\infty}[G(x)]^{n-1}\ h(x)[H(y)-H(x)]^{p-1} \cdot g(y)[1-H(y)]^{m-p}\ dy\ dx$$

$$= k^2mn\binom{m-1}{p-1}\int_0^1\int_x^1 x^{k(n-1)}(1-x)^{k-1}[(1-x)^k-(1-y)^k]^{p-1} \cdot y^{k-1}(1-y)^{k(m-p)}\ dy\ dx$$

$$= kmn\binom{m-1}{p-1}\int_0^1\int_0^{(1-x)^k} x^{k(n-1)}(1-x)^{k-1}[(1-x)^k-v]^{p-1}\ v^{m-p} \cdot v^{(1/k)-1}(1-v^{1/k})^{k-1}\ dv\ dx$$

$$= kmn\binom{m-1}{p-1}\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\int_0^1\int_0^{(1-x)^k}[(1-x)^k-v]^{p-1} \cdot v^{(j/k)+(1/k)+m-p-1}(1-x)^{k-1}x^{k(n-1)}\ dv\ dx$$

$$= kmn\binom{m-1}{p-1}\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\ B[p,(j/k)+(1/k)+m-p) \cdot \int_0^1 (1-x)^{k-1}x^{k(n-1)}[(1-x)^k]^{m-1+(j/k)+(1/k)}\ dx$$

$$= kmn\binom{m-1}{p-1}\sum_{j=0}^{k-1}\binom{k-1}{j}(-1)^j\ B[p,(j/k)+(1/k)+m-p] \cdot B(km+j+1,kn-k+1) \qquad (3.9)$$

$$\Pr\ (\text{all } n\ Y\text{'s} < \text{all } m\ X\text{'s}) = n\int_{-\infty}^{\infty}[G(y)]^{n-1}g(y)[1-H(y)]^m\ dy$$

$$= kn\int_0^1 y^{k(n-1)}\ y^{k-1}\ (1-y)^{km}\ dy$$

$$= kn\ B(km+1,kn) \qquad (3.10)$$

In (3.8) and (3.9), $1 \leq p \leq (n-1)$, and in (3.7), (3.8), and (3.9), k must be a positive integer. The binomial coefficients

$\binom{a}{b}$ are defined to be equal to zero if $b > a$. Formula (3.10) is true as long as k is a positive number.

The computations of the order probabilities under the alternative H_1 are further simplified by the fact that certain orderings have equal probabilities for all values of k, m and n. This is the case whenever a given ordering is completely reversed (the original variables are now in a descending order) and the m X random variables are replaced by m Y random variables, and the n Y random variables become n X random variables (the new variables are in an ascending order). An arrangement for the combined sample of size N will be denoted by the vector $\bar{z} = [z_1, z_2, \ldots, z_N]$, where the z_i's are indicator variables, $z_i = 1$ if the ith ordered random variable in the combined sample is an X random variable, and $z_i = 0$ otherwise. Using this notation for a given ordering $\bar{z}$, the new ordering $\bar{z}'$ with the same probability is $\bar{z}' = [1-z_N, 1-z_{N-1}, \ldots, 1-z_1]$. For example, Pr(111010100) = Pr(110101000). The equivalence could be proved for each possible ordering using (3.7) - (3.10) and similar formulas for the special cases. An easier and more general proof can be accomplished using the form of the probabilities given by (3.3) for an arbitrary ordering and any H and G which are mutually symmetric.

<u>Theorem 3.2</u>. Let $\bar{z} = [z_1, z_2, \ldots, z_n]$ denote an arrangement of m random variables X from a population with distribution function H(x) and n random variables Y from G(x) (m+n=N), where $z_i = 1$ if the <u>i</u>th ordered random variable in the combined sample of N variables is an X random variable, and $z_i = 0$ otherwise. Let $\bar{z}' = [z_1', z_2', \ldots, z_N']$, where $z_i' = 1 - z_{N-i+1}$. If H and G are both continuous and are mutually symmetric such that H(x-a) = 1 - G(a-x) for some <u>a</u> and all x, then the probabilities of the two arrangements $\bar{z}$ and $\bar{z}'$ are equal.

Proof. The probability of the ordering $\bar{z}$ is

$$p(\bar{z}) = m!n!\int_{-\infty}^{\infty}\int_{-\infty}^{u_N}\cdots\int_{-\infty}^{u_2}\prod_{j=1}^{m} h(u_{r_j})\prod_{w=1}^{n} g(u_{s_w})\, du_1 du_2 \ldots du_N,$$

where the r_j (j=1,2,...,m) and s_w (w=1,2,...,n) are the ranks of the X's and Y's in the arrangement $\bar{z}$ of the combined sample. If we let $V_{N-i+1} = -U_i$ for i = 1,2,...,N, then

$-\infty \leq U_1 \leq U_2 \leq \ldots \leq U_N \leq \infty$ implies that

$-\infty \leq V_1 \leq V_2 \leq \ldots \leq V_N \leq \infty$, and

$$p(\bar{z}) = m!n!\int_{-\infty}^{\infty}\int_{-\infty}^{v_N}\cdots\int_{-\infty}^{v_2}\prod_{j=1}^{m} h(-v_{N-r_j+1})\prod_{w=1}^{n} g(-v_{N-s_w+1})\, dv_1 \ldots dv_N. \tag{3.11}$$

Since H and G are continuous and mutually symmetric,

$h(x-a) = g(a-x)$ for some $\underline{a}$ and all x. We can assume without loss of generality that $a = 0$. Then (3.11) becomes

$$p(\bar{z}) = m!n!\int_{-\infty}^{\infty}\int_{-\infty}^{v_N}\cdots\int_{-\infty}^{v_2}\prod_{j=1}^{m} g(v_{N-r_j+1})\prod_{w=1}^{n} h(v_{N-s_w+1})dv_1dv_2\cdots dv_N$$

$$= m!n!\int_{-\infty}^{\infty}\int_{-\infty}^{v_N}\cdots\int_{-\infty}^{v_2}\prod_{w=1}^{n} h(v_{r'_w})\prod_{j=1}^{m} g(v_{s'_j})\ dv_1dv_2\cdots dv_N$$

$$= p(\bar{z}')\quad ,$$

where the r'_w $(w=1,2,\ldots,n)$ and s'_j $(j=1,2,\ldots,m)$ are the ranks of the X's and Y's in the arrangement $\bar{z}'$.

<u>Corollary 3.2.1</u>. More generally, the theorem continues to be true for arrangements of random variables drawn from populations with cumulative distribution functions of the form $H[\Psi(x)]$ and $G[\Psi(x)]$, as long as $H(x) = 1 - G(-x)$, since the probability of any arrangement is unaffected if the same monotonic transformation is applied to all N random variables.

Proof. We need only to show that a monotonic transformation of the X and Y random variables drawn from $H(x)$ and $G(x)$ where $H(x) = 1 - G(-x)$ will yield random variables X' and Y' with cumulative distribution functions $H[\Psi(x)]$ and $G[\Psi(x)]$, respectively. Let $X' = \varphi(X)$ and $Y' = \varphi(Y)$. Then

$Pr(X' \leq x) = Pr[X \leq \varphi^{-1}(x)] = H[\varphi^{-1}(x)] = H[\Psi(x)]$ where $\Psi(x) = \varphi^{-1}(x)$.

Whereas Corollary 3.2.1 is a generalization of the theorem, the following are special cases.

<u>Corollary 3.2.2</u>. The theorem holds for any two symmetric probability distribution functions differing only in location, since they are mutually symmetrical about a vertical half-way between their means, and this line may be taken to be $x = 0$.

Proof. Let the two means be $\underline{a}$ and $\underline{b}$. Then $h(x-a)=g(x-b)=g(b-x)$, and $h(x - \frac{a+b}{2}) = g(b - x - \frac{b}{2} + \frac{a}{2}) = g(\frac{a+b}{2} - x)$.

<u>Corollary 3.2.3</u>. If $f(x)$ is a density function symmetric about zero, the theorem holds if H and G can be expressed as $G = \varphi(F)$, $H = 1 - \varphi(1-F)$.

Proof. $G(-x) = \varphi[F(-x)] = \varphi[1-F(x)] = 1 - H(x)$.

In view of Corollary 3.2.1, the result continues to hold for <u>any</u> density function $f(x)$. Let us assume, e.g., that F is uniform on $(-\frac{1}{2},\frac{1}{2})$, $F(x) = x+\frac{1}{2}$. Then

$G(-x) = \varphi(-x+\frac{1}{2}) = \varphi[1-F(x)] = 1 - H(x)$. With the transformation $X' = \Psi^{-1}(X+\frac{1}{2})$, $Y' = \Psi^{-1}(Y+\frac{1}{2})$, X', Y' have cumulative distribution functions $H[\Psi'(x)]$ and $G[\Psi'(x)]$ respectively, where $\Psi'(x) = \Psi(x) - \frac{1}{2}$.

Example. Let $\varphi(F) = F^k$. Then $H = 1 - (1-F)^k$ and $G = F^k$. If k is a positive integer, G and H are the distributions of the largest and smallest of k independent variates with cumulative distribution function F. In particular, if $F(x) = 1 - e^{-x}$ $(x > 0)$, then $G(x) = (1 - e^{-x})^k$ and $H(x) = 1 - e^{-kx}$.

Even though this theorem reduces the total number of order probability calculations, lengthy arithmetic operations are required to obtain the probability under the alternative of any individual arrangement. The formulae given by (3.4) - (3.10) are not readily adaptable to systematic calculation. Their practicability decreases as m, n, or k increases. Some type of recursive relationship would be desirable.

There is a back-recursive rule (Savage, 1960) which permits the calculation of the probability of any ordering for samples of sizes m and n, respectively, from the probabilities of orderings for samples of sizes (m+1) and n, regardless of the populations from which the two samples are drawn. The relationship is as follows:

Theorem 3.3. The probability of any given ordering $\overline{z} = [z_1, z_2, \ldots, z_{m+n}]$ of m X and n Y random variables,

$Pr_{m,n}(\bar{z})$, can be found by summing the probabilities of all possible orderings $\bar{z}'$ which can be obtained from $\bar{z}$ by placing one additional X random variable in every possible position in the original ordering, and then dividing by m + 1. That is,

$$Pr_{m,n}(\bar{z}) = \Sigma\, Pr_{m+1,n}(\bar{z}')/(m+1) \quad , \tag{3.12}$$

where $\bar{z}' = [z_1, z_2, \ldots, 1, z_j, \ldots, z_{m+n}]$, $j=1,2,\ldots,m+n+1$, and the sum is extended over the (m+n+1) orderings $\bar{z}'$, some of which will be equal.

Proof. Defining $u_o = -\infty$ and $u_{m+n+1} = \infty$, we can write

$$\Sigma\, Pr_{m+1,n}(\bar{z}')/(m+1) =$$

$$= \sum_{j=1}^{m+n+1} \frac{(m+1)!\,n!}{(m+1)} \int_{-\infty}^{\infty} \int_{-\infty}^{u_{m+n}} \cdots \int_{-\infty}^{u_2} \prod_{j=1}^{m} h(u_{r_j})$$

$$\cdot \prod_{w=1}^{n} g(u_{s_w})\ du_1 \ldots du_{m+n} \int_{u_{j-1}}^{u_j} h(x)dx$$

$$= m!\,n! \int_{-\infty}^{\infty} \int_{-\infty}^{u_{m+n}} \cdots \int_{-\infty}^{u_2} \prod_{j=1}^{m} h(u_{r_j}) \prod_{w=1}^{n} g(u_{s_w})\, du_1 \ldots du_{m+n} \int_{-\infty}^{\infty} h(x)dx$$

$$= Pr_{m,n}(\bar{z}) \quad .$$

For example,

$$Pr_{2,3}(10100) = [2Pr_{3,3}(110100) + 2Pr_{3,3}(101100)$$
$$+ Pr_{3,3}(101010) + Pr_{3,3}(101001)]/3 \quad .$$

The rule is of limited advantage for this alternative, since the complexity of calculations increases with k, m, and n, but it can be a check on exact numerical computations. It would be more useful if an electronic computer program could be set up to obtain the probabilities of all the orderings, either exactly or empirically, for a sufficiently large fixed m and n, since the probabilities of all the orderings for smaller sample sizes could be obtained by the recursive relation.

A forward-recursive scheme adaptable to systematic computer calculation (Klotz, 1962, pp. 501-502) would perhaps be more appropriate. If $\bar{z}$ denotes a particular ordering of the m X and n Y random variables, let $\bar{z}X$ and $\bar{z}Y$ denote the arrangements obtained by adjoining an X and a Y to the right, respectively. Let

$$P_{\bar{z}}(v) = \Pr(\text{all X's and Y's} \leq v \text{ and in the order } \bar{z})$$

$$= m!n! \int_{-\infty}^{v} \int_{-\infty}^{u_N} \cdots \int_{-\infty}^{u_2} \prod_{j=1}^{m} h(u_{r_j}) \prod_{w=1}^{n} g(u_{s_w})\, du_1 \cdots du_N \quad .$$

Then

$$P_X(v) = H(v) \quad \text{and} \quad P_Y(v) = G(v) \; ,$$

$$P_{\bar{z}X}(v) = (m+1) \int_{-\infty}^{v} P_{\bar{z}}(t)\, h(t)\, dt$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (3.13)

$$P_{\bar{z}Y}(v) = (n+1) \int_{-\infty}^{v} P_{\bar{z}}(t)\, g(t)\, dt \quad .$$

The probabilities of the respective orderings $\overline{Z}$, $\overline{Z}X$, and $\overline{Z}Y$ are $P_{\overline{Z}}(\infty)$, $P_{\overline{Z}X}(\infty)$, and $P_{\overline{Z}Y}(\infty)$.

3.3 Tests Used for Power Calculations

The rank tests which will be investigated for power calculations are:

- T_1 the most powerful rank test
- T_2 the Mann-Whitney or Wilcoxon test
- T_3 Terry's c_1 test
- T_4 the one-sided median test
- T_5 the two-sided Wilcoxon test
- T_6 the two-sided median test
- T_7 the Wald-Wolfowitz runs test
- T_8 the Psi test.

All of the tests except T_3 and T_8 were studied by Lehmann with the alternative of one extreme distribution. Each of these tests deserves individual discussion so that the power results will be more meaningful.

The most powerful nonparametric rank test T_1 rejects for those orderings with the largest probability under the alternative. Consider the problem of testing the composite hypothesis H_o: $H = G$ unspecified, where H and G denote the cumulative distribution functions of X and Y respectively, against the alternative H_1: $H \neq G$. Because any test based on ranks depends only on the arrangement $\bar{z}$ of 1's and 0's, where 1 and 0 are indicators for X and Y random variables, and all orderings are equally likely under the null hypothesis, all rank tests are similar tests. Then the problem can be reduced to testing the simple null hypothesis H_o: $H = G$, $G = G$ (where G can be assumed to be uniform without loss of generality) against the alternative H_1 above. A most powerful rank test of the simple hypothesis H_o against H_1 will be a most powerful rank test of the composite H_o versus H_1. The most powerful rank test then rejects for those orderings $\bar{z}$ for which

$$\frac{\text{probability of } \bar{z} \text{ under } H_1}{\text{probability of } \bar{z} \text{ under } H_o} = \frac{\text{probability of } \bar{z} \text{ under } H_1}{\binom{N}{m}} > c'_\alpha ,$$

i.e., when the probability of $\bar{z}$ under H_1 is greater than c_α, c_α being a constant determined by the desired significance level α. The rejection region for any rank test will consist of certain orderings, and there will be $\alpha\binom{N}{m}$ of them. If $\alpha\binom{N}{m}$ is not an integer, a randomized decision rule may be used. Thus for the most powerful rank test of H_o against H_1, the rejection region will consist of those orderings which have the largest probabilities under the alternative, and the power will be the sum of these probabilities.

For the one-sided Wilcoxon or Mann-Whitney U test T_2 (Mann and Whitney, 1947), the cases in the rejection region are those for which the sum of the ranks of the Y's, $s_1 + s_2 + \ldots + s_n$, is largest, where the number of cases in the rejection region is the smallest integer greater than or equal to $\alpha\binom{N}{m}$ for a randomized test.

Under the null hypothesis, the average (or expected) rank of any Y random variable is $(m+n+1)/2$ and thus the expected sum of the ranks of the Y's is $n(m+n+1)/2$. The two-sided Wilcoxon test T_5 (Wilcoxon, 1945) rejects when the absolute value of the difference between the observed sum of the ranks and the expected sum of the ranks is too large. The cases in the rejection region then are those with the largest values of

$$\left|\sum_{i=1}^{n} s_i - n(m+n+1)/2\right| \tag{3.14}$$

Let us define the median w of the combined sample as the variable with rank a, where $a = (m+n+1)/2$ if $(m+n)$ is odd and $a = (m+n)/2$ if $(m+n)$ is even. Under the null hypothesis, the probability of having u Y's and v X's greater than the median w is

$$f(u,v) = \binom{n}{u}\binom{m}{v}\Big/\binom{N}{a} \quad , \tag{3.15}$$

where $v = m+n-a-u$. The one-sided median test T_4 (Mood, 1950, pp. 394-395) then is to reject H_o always if $u > u_o$ and with probability p if $u = u_o$ where

$$\sum_{u=u_o+1}^{n} f(u,v) + p\, f(u_o,v) = \alpha \quad .$$

For the two-sided median test T_6 when $m = n$, we reject if $u > u_o$ or $v > u_o$ and with probability p if $u = u_o$ or $v = u_o$, where

$$\sum_{u=u_o+1}^{n} f(u,m+n-a-u) + \sum_{v=u_o+1}^{n} f(m+n-a-v,v) + p\, f(u_o,m+n-a-u_o) + p\, f(m+n-a-u_o,u_o) = \alpha \quad .$$

Although the power of both these median tests can be computed by adding up the corresponding probabilities under the alternative, it is simpler to use the formula

$$\begin{aligned} f(u,v,w) = {} & km\binom{m-1}{v}\binom{n}{u}[1-(1-F(w))^k]^{m-v-1} \\ & [1-F(w)]^{kv}\, F(w)^{k(n-u)}\, [1-F^k(w)]^u\, [1-F(w)]^{k-1}\, f(w) \\ & + kn\binom{m}{v}\binom{n-1}{u}[1-(1-F(w))^k]^{m-v}\, [1-F(w)]^{kv} \\ & F(w)^{k(n-u-1)}\, [1-F^k(w)]^u\, F^{k-1}(w)\, f(w) \quad . \qquad (3.16) \end{aligned}$$

When w has been integrated out of (3.16) and the substitution $u = m+n-a-v$ is made, the formula reduces to

$$f(v) = km\binom{m-1}{v}\binom{n}{\frac{m+n-2v-1}{2}}\sum_{j=0}^{m-v-1}\binom{m-v-1}{j}(-1)^j$$

$$\cdot \sum_{w=0}^{\frac{m+n-1-2v}{2}}\binom{\frac{m+n-1-2v}{2}}{w}(-1)^w\ \beta(kj+kv+k,\ kw+1+k(n+2v-m+1)/2)$$

$$+ kn\binom{m}{v}\binom{n-1}{\frac{m+n-2v-1}{2}}\sum_{j=0}^{m-v}\binom{m-v}{j}(-1)^j \qquad (2.17)$$

$$\cdot \sum_{w=0}^{\frac{m+n-2v-1}{2}}\binom{\frac{m+n-2v-1}{2}}{w}(-1)^w\ \beta(kj+kv+1,\ k+kw+k(n+2v-m-1)/2)$$

for the case where m+n is odd, a = (m+n+1)/2, and

$$f(v) = km\binom{m-1}{v}\binom{n}{\frac{m+n-2v}{2}}\sum_{j=0}^{m-v-1}\binom{m-v-1}{j}(-1)^j$$

$$\cdot \sum_{w=0}^{\frac{m+n-2v}{2}}\binom{\frac{m+n-2v}{2}}{w}(-1)^w\ \beta(kj+kv+k,\ kw+1+k(n+2v-m)/2)$$

$$+ kn\binom{m}{v}\binom{n-1}{\frac{m+n-2v}{2}}\sum_{j=0}^{m-v}\binom{m-v}{j}(-1)^j \qquad (3.18)$$

$$\cdot \sum_{w=0}^{\frac{m+n-2v}{2}}\binom{\frac{m+n-2v}{2}}{w}(-1)^w\ \beta(kj+kv+1,\ kw+k+k(n+2v-m-2)/2),$$

when m+n is even, a = (m+n)/2. Thus when m = n = m, u = m - v, we obtain

$$f(v) = km\binom{m-1}{v}\binom{m}{v} \sum_{j=0}^{m-v-1} \binom{m-v-1}{j}(-1)^j$$

$$\cdot \sum_{w=0}^{m-v} \binom{m-v}{w}(-1)^w \, \beta(kv+kw+1, \; kj+kv+k)$$

$$+ \; km\binom{m-1}{v-1}\binom{m}{v} \sum_{j=0}^{m-v} \binom{m-v}{j}(-1)^j \qquad (3.19)$$

$$\cdot \sum_{w=0}^{m-v} \binom{m-v}{w}(-1)^w \, \beta(kw+kv, \; kj+kv+1) \quad .$$

We must define $\binom{a}{b}$ equal to zero if $b < 0$ or $b > a$. All of the formulas hold for any positive integer k.

For the Wald-Wolfowitz runs test T_7 (Wald and Wolfowitz, 1940), under the null hypothesis we have

$$Pr(R = 2c) = 2\binom{m-1}{c-1}\binom{n-1}{c-1}/\binom{N}{m} \quad \text{and}$$

$$Pr(R = 2c+1) = [\binom{m-1}{c}\binom{n-1}{c} + \binom{m-1}{c-1}\binom{n-1}{c}]/\binom{N}{m} \quad , \qquad (3.20)$$

where R is the total number of runs of X's and Y's. The test is to reject if the observed $R < r_o$ and with probability p if $R = r_o$, where

$$\sum_{r=2}^{r_o-1} Pr(R = r) + p \; Pr(R = r_o) = \alpha \quad .$$

These six tests are all designed to detect any type of difference between the distributions of the two sets of random variables and are thus applicable to any alternative expressing the inequality, $H \neq G$. On the other hand, Terry's c_1 test T_3 (Terry, 1952) would be used principally when the relevant alternative is that the two populations are both normal with the same variance but different means. Terry has shown that the c_1 test is the locally most powerful rank test against this parametric alternative (see also Section 4.3.3). The test statistic is

$$c_1(P) = \sum_{j=1}^{n} \mathcal{E}(\xi_{s_j}) \quad , \tag{3.21}$$

where $\xi_1 \leq \xi_2 \leq \ldots \leq \xi_N$ are order statistics of a sample of size N from a normal distribution with mean zero and variance one, and $s_1, s_2, \ldots, s_n$ are the ranks of the Y random variables in the combined ordered sample. For a one-sided test, i.e., the alternative that the mean of the Y's is greater than the mean of the X's, we reject when $c_1 \geq c$ where c is a constant determined by the significance level. The cases in the rejection region are those arrangements with the largest values of c_1, and the power is the sum of the probabilities of these cases under the alternative. Terry's Table I (pp. 358-361) gives the exact distribution of $c_1(P)$ for all possible arrangements P of the m X's and nY's, and all combinations of m and n such that $m + n \leq 10$. (The notation in this paper is the opposite of that used by Terry, as he uses a 0 to represent an X and a 1 to represent a Y random variable.) For example, for $m = n = 4$, $c_1(11110000) = 2.89$, $c_1(11101000) = 2.59$, $c_1(11011000) = c_1(11100100) = 2.27$, and the probability of any arrangement under the null hypothesis is 1/70. Thus for significance level .05, the randomized test rejects for both of the first two orderings and with probability 3/4 for either of the last two, and the power is the sum of the probabilities of these orders. If $m + n > 10$, tables of the expected values of the order statistics from a normal distribution can be used to determine the critical orderings (Fisher and Yates, 1953; Harter, 1961).

The last test considered, T_8, the Psi test, is also a locally most powerful rank test. The one-sided test is to reject the null hypothesis for those orderings $\bar{z}$ for which

$$\sum_{i=1}^{N} [Y(N-i+1) - Y(i)] z_i > c \quad , \qquad (3.22)$$

where the z_i's are the indicator variables previously defined, c is a constant determined by the significance level, and $\Psi(x) = d[\log \Gamma(x)]/dx$, which has been tabulated by Davis (1933, pp. 291-367). Properties of the test are discussed in Section 4.3.2.

3.4 Power Functions of the Rank Tests

3.4.1 Results for the Alternative of Two Extreme Distributions

Power functions of the one and two-sided tests against the alternative H_1: $H = 1 - (1-F)^k$, $G = F^k$, F being unspecified are presented in Tables 3.1 - 3.6 for small sample sizes. The cases considered for equal sample sizes are all combinations of $m = n = 2, 3, 4$, $k = 2, 3, 4$, and significance levels .01, .05, .10. Power functions for unequal sample sizes are given for $k = 2$, 3 and 4, $\alpha \leq 1/\binom{N}{m}$ for the one-sided tests and $\alpha \leq 2/\binom{N}{m}$ in the case of two-sided tests. All combinations of m and n for which $m + n \leq 8$, plus $m = 1$, $n = 8$ or $m = 8$, $n = 1$, and $m = 1$, $n = 9$ or $m = 9$, $n = 1$ are considered. The significance levels are attained exactly in all cases by using a randomized decision rule.

Although it is difficult to draw any significant conclusions from results for such small sample sizes and limited ranges of α, the calculations of the order probabilities become extremely tedious for larger sample sizes and values of k. There are at least two practicable methods of enlarging the range of computations. First, if an electronic computer could be programmed to find the probabilities of all the orderings for sufficiently large sample sizes, Savage's back-recursive relationship (3.12) could be used to find the order probabilities for all smaller sample sizes. Second, starting with samples of size one, the forward-recursive scheme (3.13) could be programmed to build up to probabilities for larger sample sizes. The author plans to attempt this latter method on the IBM 7040 in the near future.

As an example of critical region and power function calculations, consider $m = n = 4$, $k = 2$, $\alpha = .05$. There are 70 possible arrangements of the random variables, each occurring with probability 1/70 under the null hypothesis. For the most powerful rank test, T_1, the four cases occurring with highest probability under the alternative belong to the rejection region since $\alpha \cdot \binom{N}{m} = 3.5$. These are (11110000), (11101000), (11011000), and (11100100). But the last two of the four cases have equal probabilities by Theorem 3.2. Thus the .05 level test rejects always for the orderings (11110000) and (11101000) and with probability 3/4 if either of the cases (11011000) or (11100100) occurs. The power is

$$.24357 + .13682 + \frac{3}{4} \cdot 2\ (.08038) = .5010\ .$$

Similarly for T_2, T_3, and T_8. The criterion for the one-sided median test T_4 is v, the number of X's larger than the median of the combined sample. From (3.15) we see that there is only one ordering with no X's larger than the median, and 16 cases with one X larger than the median. In order to achieve exact significance level .05, we reject always when v = 0 and with probability 5/32 when v = 1. The power is

$$.24357 + \frac{5}{32}\ (.57258) = .3330\ .$$

The following chart for equal sample sizes lists the critical orders and probabilities with which they must be rejected to attain exact significance level α. The listings for T_1, T_2, T_3, and T_8 are in descending order of probability except that certain rank orderings have equal probabilities by Theorem 3.2.

m=n	Tests	α=.01	α=.05	α=.10
2	T_1, T_2, T_3, T_4, T_8	(1100), .06	(1100), .3	(1100), .6
	T_5, T_6, T_7	(1100) (0011) } .03	(1100) (0011) } .15	(1100) (0011) } .3

m=n	Tests	α=.01	α=.05	α=.10
3	T_1, T_2, T_3, T_8	(111000),.2	(111000),1	(111000),(110100),1
	T_4	(111000),.2	(111000),1	(111000),1 (---100),(---010), (---001) } 1/9 where the blanks are to be filled by all possible arrangements of the remaining 0's and 1's (9 cases).
	T_5, T_6, T_7	(111000), (000111) } .1	(111000), (000111) } .5	(111000), (000111) } 1
4	T_1, T_2, T_3, T_8	(11110000),.7	(11110000), (11101000) } 1 (11100100), (11011000) } 3/4	(11110000),(11101000), (11100100),(11011000), (11010100),(10111000), (11100010), all with probability 1
	T_4	(11110000),.7	(11110000),1 (----1000), (----0100), (----0010), (----0001) } 5/32 (16 cases)	(11110000),1 the same 16 cases as for α = .05 but reject with probability 6/16

m=n	Tests	α=.01	α=.05	α=.10
4	T_5	(11110000) (00001111) } .35	(11110000) (00001111) } 1 (11101000) (00010111) } 3/4	(11110000), (00001111) (11101000), (00010111) } 1 (11011000), (00100111) (11100100), (00011011) } 3/4
	T_6	(11110000) (00001111) } .35	(11110000) (00001111) } 1 (----1000) (----0111) (----0100) (----1011) (----0010) (----1101) (----0001) (----1110) } 3/64 (32 cases)	(11110000) (00001111) } 1 the same 32 cases as for α = .05 but reject with probability 5/32
	T_7	(11110000) (00001111) } .35	(11110000) (00001111) } 1 (11100001) (00011110) (11000011) (00111100) (10000111) (01111000) } 1/4	(11110000) (00001111) } 1 the same 6 cases as for α = .05 but reject with probability 5/6

The probability of any arrangement of the $(m+n) = N$ random variables under the null hypothesis is $1/\binom{N}{m}$. For selected values of m, n, and N, these probabilities are:

N	m or n	Probability	N	m or n	Probability
4	1	1/4	7	1	1/7
4	2	1/6	7	2	1/21
5	1	1/5	7	3	1/35
5	2	1/10	8	1	1/8
6	1	1/6	8	2	1/28
6	2	1/15	8	3	1/56
6	3	1/20	8	4	1/70
			9	1	1/9
			10	1	1/10

Under the alternative H_1, the probability of any ordering may be found using (3.5) and (3.6), or (3.7), (3.8), (3.9) and (3.10) when applicable. An example of one of the calculations for $k = 2$, $m = n = 3$, using formula (3.5) is as follows:

$$\Pr(101100) = 3!\ 3!\ 2^5 \sum_{i=1}^{2} \frac{\binom{2}{i}(-1)^{i-1}}{(i+2)} \sum_{j=0}^{1} \frac{(-1)^j}{(i+j+3)}$$

$$\cdot \sum_{w=0}^{1} \frac{(-1)^w}{(i+j+w+4)(i+j+w+6)(i+j+w+8)} = .09355 \quad .$$

The following chart lists some of the rank orderings and their probabilities under the alternative. Those entries left blank have not been computed.

m=n	Order	Probability under H_1 k=2	k=3	k=4
2	1100	.58571	.84654	.95086
	1010	.21746		
	1001, 0110	.07778		
	0101	.02698		
	0011	.01429	.00108	.00008
3	111000	.38463	.72845	.90384
	110100	.18602	.15009	.06135
	110010, 101100	.09355		
	101010	.04616		
	110001, 011100	.03797		
	100110	.02786		
	101001, 011010	.01851		
	100101, 010110	.01101		
	011001	.00737		
	100011, 001110	.00680		
	010101	.00431		
	010011, 001101	.00264		
	001011	.00160	.00004	
	000111	.00108	.00002	.00001
	11110000	.24357	.61324	.85003
	11101000	.13682	.16563	.09489
	11011000, 11100100	.08038	.05715	.01922
	11010100	.04662	.02870	.00563
	10111000, 11100010	.04445	.01915	.00412
	11100001, 01111000	.01909	.00414	.00064
	11000011, 00111100	.00307		
	10000111, 00011110	.00055		
	00011011, 00100111	.00014	.00000	.00000
	00010111	.00010	.00000	.00000
	00001111	.00008	.00000	.00000

Since the critical regions for the two median tests, T_4 and T_6, contain so many orderings unique to these tests, the sum f(v), of the probabilities of all the orderings with the same number v or X's larger than the median, was found for equal sample sizes using (3.19). The results are as follows:

		f(v)		
m=n	v	k=2	k=3	k=4
2	0	.58571	.84654	.95086
	1	.40000	.15238	.04906
	2	.01429	.00108	.00008
3	0	.38463	.72845	.90384
	1	.53961	.26412	.09555
	2	.07468	.00741	.00061
	3	.00108	.00002	.00001
4	0	.24357	.61324	.85003
	1	.57258	.36389	.14787
	2	.17383	.02263	.00210
	3	.00995	.00025	.00001
	4	.00008	.00000	.00000

For unequal sample sizes, only the two most extreme order probabilities need be computed to find the power for the chosen significance levels. The number of calculations required for the power functions is reduced by the fact that the probabilities for the two extreme orderings are symmetric in m and n. That is, the probability that all observations in a sample of size m are less than all observations in a second sample of size n is equal to the probability that n observations from the first sample are less than m observations from the second sample. This is a special case of the equal probabilities of two orderings as defined by Theorem 3.2.

The following chart lists the extreme orderings and their probabilities as calculated from (3.7) and (3.10).

m,n or n,m	Orders	Probability under H_1 k=2	k=3	k=4
1,3	1000,1110	.66429	.88117	.96265
	0001,0111	.03571	.00455	.00055
1,4	10000,11110	.61270	.85544	.94268
	00001,01111	.02222	.00220	.00021
2,3	11000,11100	.49048	.79500	.86448
	00011,00111	.00476	.00020	.00001
1,5	100000,111110	.57215	.83336	.94422
	000001,011111	.01515	.00123	.00009
2,4	110000,111100	.42338	.75276	.91562
	000011,001111	.00202	.00005	.00000
1,6	1000000,1111110	.53913	.81403	.93618
	0000001,0111111	.01099	.00075	.00005
2,5	1100000,1111100	.37329	.76478	.89783
	0000011,0011111	.00100	.00002	.00000
3,4	1110000,1111000	.31375	.67488	.87974
	0000111,0001111	.00033	.00000	.00000
1,7	10000000,11111110	.51152	.79687	.92876
	00000001,01111111	.00833	.00049	.00003
2,6	11000000,11111100	.33435	.68654	.91346
	00000011,00111111	.00055	.00001	.00000
3,5	11100000,11111000	.26320	.63050	.85817
	00000111,00011111	.00012	.00000	.00000
1,8	100000000,111111110	.48797	.78048	.92185
	000000001,011111111	.00654	.00034	.00002
1,9	1000000000,1111111110	.46755	.76747	.91539
	0000000001,0111111111	.00526	.00025	.00001

The results of the power computations for the eight rank tests, when $m = n = 2$ and 3, $k = 2$, 3, and 4, are given in Tables 3.1 and 3.2.

Table 3.1. Power against H_1 for $m = n = 2$, $k = 2, 3, 4$ for the eight rank tests

	α=.01			α=.05			α=.10		
Test	k=2	k=3	k=4	k=2	k=3	k=4	k=2	k=3	k=4
T_1, T_2, T_3, T_4, T_8	.0351	.0508	.0571	.1757	.2540	.2853	.3514	.5079	.5705
T_5, T_6, T_7	.0180	.0254	.0285	.0900	.1271	.1426	.1800	.2543	.2853

The powers here are low, as is to be expected for such small sample sizes. The maximum possible power for $m = n = 2$ is .06 for $\alpha = .01$, .30 for $\alpha = .05$, and .60 for $\alpha = .10$ for the five one-sided tests, since there is only one case in the rejection region. With any two-sided test, the powers cannot exceed .03, .15, and .30 when $\alpha = .01$, .05, and .10 respectively. For $k = 4$, the results are quite close to these maximum values. As a result, considering larger values of k would have very little effect on the power functions.

Table 3.2. Power against H_1 for m = n = 3, k = 2, 3, 4 for the eight rank tests

Test	α=.01			α=.05			α=.10		
	k=2	k=3	k=4	k=2	k=3	k=4	k=2	k=3	k=4
T_1, T_2, T_3, T_8	.0769	.1457	.1808	.3846	.7285	.9038	.5706	.8785	.9652
T_4	.0769	.1457	.1808	.3846	.7285	.9038	.4446	.7578	.9145
T_5, T_6, T_7	.0386	.0728	.0904	.1929	.3642	.4519	.3857	.7285	.9038

For α = .01, the maximum possible power is .20 for m = n = 3 for any one-sided test based on ranks. For $\alpha \geq .05$, there is no limit on the powers. The power of any two-sided test cannot exceed .10 for α = .01 and .50 for α = .05.

In Table 3.3 the power functions of these same tests for m = n = 4, k = 2, α = .01, .05, and .10 are presented, as well as Lehmann's results for the alternative H = F, $G = F^2$, α = .10.

Table 3.3. Power against H_1 for m = n = 4, k = 2, for the eight rank tests

Test	$H=1-(1-F)^2$, $G=F^2$			$H=F$, $G=F^2$
	α=.01	α=.05	α=.10	α=.10
T_1, T_8	.1705	.5010	.6767	.32
T_2, T_3	.1705	.5010	.6767	.31
T_4	.1705	.3330	.4583	.23
T_5	.0853	.3463	.5014	.19
T_6	.0853	.2710	.3347	.15
T_7	.0853	.2550	.2815	.14

Since the power of the Wald-Wolfowitz runs test T_7 does not compare favorably with the other two-sided tests, and the rejection orderings are unique to this one test when $\alpha > 1/35$, it was omitted for further computations. The power functions for the other tests when $m = n = 4$, $k = 3$ and 4, are given in Tables 3.4 and 3.5 respectively, along with Lehmann's results for the alternative $H = F$, $G = F^3$, $\alpha = .10$.

Table 3.4. Power against H_1 for $m = n = 4$, $k = 3$, for the eight rank tests

	$H=1-(1-F)^3$, $G=F^3$			$H=F$, $G=F^3$
Test	$\alpha=.01$	$\alpha=.05$	$\alpha=.10$	$\alpha=.10$
T_1, T_8	.4293	.8646	.9602	.49
T_2, T_3	.4293	.8646	.9602	.47
T_4	.4293	.6701	.7497	.33
T_5	.2146	.7375	.8646	.32
T_6	.2146	.6303	.6701	.22

Table 3.5. Power against H_1 for $m = n = 4$, $k = 4$

	$H=1-(1-F)^4$, $G=F^4$		
Test	$\alpha=.01$	$\alpha=.05$	$\alpha=.10$
T_1, T_2, T_3, T_8	.5950	.9738	.9972
T_4	.5950	.8731	.9055
T_5	.2975	.9212	.9738
T_6	.2975	.8570	.8731

It should be noted that although the power functions against H_1 of the one-sided tests T_1, T_2, T_3 and T_8 are the same for the cases considered here with $m = n \leq 4$, this will not be true in general. In many cases for which the sums of the ranks of the Y's are equivalent, the probabilities of their occurrence under the alternative and/or the exact c_1 values may differ. For example, in the case $m = n = 3$, $k = 2$, Pr(101010) = .04616 and Pr(110001) = .03797, but the Mann-Whitney-Wilcoxon statistic $\sum_{i=1}^{3} s_i = 12$ in both cases. Also for $m = n = 3$, $\sum_{i=1}^{3} s_i = 9$ in both of the orderings (101010) and (011100), but the corresponding exact c_1 values are .83 and .64. For $m = n = 5$, the powers for T_2 and T_3 will be the same for $\alpha \leq 4/252$ only. It is evident from Terry's Table I (1952, pp. 358-361) that although the Wilcoxon and c_1 statistics are similar, the c_1 statistic is generally more sensitive than the Wilcoxon (or Mann-Whitney) statistic. A linear functional relationship between the two statistics does not exist. Terry has shown that the limit of the correlation coefficient between them under the null hypothesis is $(3/\pi)^{\frac{1}{2}}$, or approximately .9772, for large samples.

As k increases, the power functions increase very rapidly in every case. This is a result of the fact that for every $m = n$, although the ranking of the probabilities of the various orderings is the same, the concentration of probability is much higher for larger values of k. The probabilities of the extreme orderings for which all or almost all the Y's are less than the X's are negligible to four decimal places for $m = n \geq 3$, $k \geq 3$. For $m = n = 3$, 57.1%, 87.9% and 96.5% of the total probability is concentrated in the two cases (111000) and (110100) when $k = 2$, 3, and 4 respectively. When $m = n = 4$, however, (11110000) or (11101000) occur only 35.0% of the time for $k = 2$, with the frequency jumping to 77.9% and 94.5% (85.0% in the one case 11110000) when $k = 3$ and $k = 4$ respectively. Also for $m = n = 4$, the highest four cases account for 51.1% of the probability, and the highest seven cases account for 67.7% when $k = 2$, in contrast to 89.3% and 96.0% when $k = 3$, and 98.3% and 99.7% when $k = 4$. The concentration can be expected to be even more pronounced for larger k and larger sample sizes.

The tests T_1, T_2, T_3, and T_8 are the most powerful of the five one-sided tests, as is to be expected, since there are so many more cases to be considered in the rejection region for the one-sided median test. For larger m and n, it is reasonable to expect on the basis of these results that the test T_5 will be by far the most powerful of the two-sided tests considered, its power increasing rapidly as the significance level increases. It is to be noted that the two-sided Wilcoxon test is even more powerful than the one-sided median test for $m = n = 4$, $k = 2$, 3, and 4, $\alpha > 2/70$. This is due to the fact that the four cases with highest probability comprise such a large proportion of the total probability. For the test T_4, there are sixteen cases in the rejection region occurring with equal probability under the null hypothesis when $1/70 < \alpha \leq 17/70$, and only a few cases for T_5, some of which have negligible probabilities under the alternative. However, with T_5, those cases with very high probabilities are given full weight.

The results for unequal sample sizes are presented in Table 3.6. The two median tests have been eliminated from consideration since the cases in the critical regions are different, and the power will be much lower because of the larger number of rejection orderings.

Table 3.6. Power against H_1 for unequal sample sizes

m,n or n,m	One-sided tests T_1,T_2,T_3,T_8 α	k=2	k=3	k=4	Two-sided tests T_5,T_7 α	k=2	k=3	k=4
	.01	.0266	.0352	.0385	.01	.0140	.0177	.0193
1,3	.05	.1329	.1762	.1925	.05	.0700	.0886	.0963
	.10	.2656	.3525	.3851	.10	.1400	.1771	.1926
	.01	.0306	.0428	.0471	.01	.0159	.0214	.0236
1,4	.05	.1532	.2139	.2357	.05	.0794	.1072	.1179
	.10	.3063	.4277	.4713	.10	.1587	.2144	.2357
	.01	.0490	.0795	.0864	.01	.0248	.0398	.0432
2,3	.05	.2452	.3975	.4322	.05	.1238	.1988	.2161
	.10	.4905	.7950	.8645	.10	.2476	.3976	.4322
	.01	.0343	.0500	.0567	.01	.0176	.0250	.0283
1,5	.05	.1716	.2500	.2833	.05	.0881	.1252	.1416
	.10	.3433	.5000	.5665	.10	.1762	.2504	.2833
	.01	.0635	.1129	.1373	.01	.0319	.0565	.0687
2,4	.05	.3175	.5646	.6867	.05	.1595	.2823	.3434
	1/15	.4234	.7528	.9156	.10	.3190	.5646	.6867
	.01	.0377	.0570	.0655	.01	.0193	.0285	.0328
1,6	.05	.1887	.2849	.3277	.05	.0963	.1426	.1638
	.10	.3774	.5698	.6553	.10	.1925	.2852	.3277
	.01	.0784	.1606	.1885	.01	.0393	.0803	.0943
2,5	1/21	.3733	.7648	.8978	.05	.1965	.4015	.4714
					2/21	.3743	.7648	.8978

Table 3.6 continued

m,n or n,m	One-sided tests T_1,T_2,T_3,T_8 α	k=2	k=3	k=4	Two-sided tests T_5,T_7 α	k=2	k=3	k=4
	.01	.1098	.2362	.3079	.01	.0550	.1181	.1540
3,4	1/35	.3138	.6749	.8797	.05	.2748	.5905	.7698
					2/35	.3141	.6749	.8797
	.01	.0409	.0637	.0743	.01	.0208	.0319	.0372
1,7	.05	.2046	.3187	.3715	.05	.1040	.1595	.1858
	.10	.4092	.6375	.7430	.10	.2079	.3189	.3715
	.01	.0936	.1922	.2558	.01	.0469	.0961	.1279
2,6	1/28	.3343	.6865	.9135	.05	.2344	.4806	.6394
					2/28	.3349	.6866	.9135
3,5	.01	.1474	.3531	.4806	.01	.0737	.1765	.2403
	1/56	.2632	.6305	.8582	2/56	.2633	.6305	.8582
	.01	.0439	.0702	.0830	.01	.0223	.0351	.0415
1,8	.05	.2196	.1405	.4148	.05	.1113	.0703	.2074
	.10	.4392	.7024	.8297	.10	.2225	.3514	.4148
	.01	.0468	.0767	.0915	.01	.0236	.0768	.0915
1,9	.05	.2338	.3837	.4577	.05	.1182	.1919	.4577
	.10	.4675	.7675	.9154	.10	.2364	.3839	.9154

3.4.2 Comparisons Against Normal Alternatives

Terry's c_1 test T_3 is designed primarily for testing the null hypothesis H_o: $H = G$ against the specific alternative H_1':

$$H(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{1}{2\sigma^2}(t-\mu_X)^2}\,dt, \quad G(y) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{y} e^{-\frac{1}{2\sigma^2}(t-\mu_Y)^2}\,dt,$$

and it is the locally most powerful rank test against this alternative when $(\mu_Y - \mu_X) > 0$ (see Section 4.3.3). As a result, the power comparisons for this test would be more appropriate if a normal alternative were used instead of the previously considered definitions of H and G where F is unspecified. The power functions are calculated by summing the probabilities for the cases in the respective rejection regions, where the probabilities are computed under the assumption that the alternative H_1' applies. Thus, for example,

$$\Pr(\text{all } m\ X\text{'s} < \text{all } n\ Y\text{'s}) = m\int_{-\infty}^{\infty} [1-G(t)]^n[H(t)]^{m-1}\,h(t)\,dt$$

$$= m\int_{-\infty}^{\infty} [1-\Phi(\frac{t-\mu_Y}{\sigma})]^n\,[\Phi(\frac{t-\mu_X}{\sigma})]^{m-1}\,\varphi(\frac{t-\mu_X}{\sigma})\,dt\,, \qquad (3.23)$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}v^2}dv$. Letting $x = (t-\mu_X)/\sigma$ in (3.23), we obtain

$$\Pr(\text{all } m \text{ X's} < \text{all } n \text{ Y's}) = m\int_{-\infty}^{\infty} [1-\Phi(x-\frac{\mu_Y-\mu_X}{\sigma})]^n [\Phi(x)]^{m-1} \varphi(x)\,dx .$$

Let us define $\delta_k = (\mu_Y - \mu_X)/\sigma$, the standardized difference between the two distributions H and G above. Then the extreme order probability (3.23) is given by

$$\Pr(\text{all } m \text{ X's} < \text{all } n \text{ Y's}) = \int_{-\infty}^{\infty} [\Phi(\delta_k - x)]^n \, d([\Phi(x)]^m) \quad (3.24)$$

since $1 - \Phi(x-\delta_k) = \Phi(\delta_k - x)$. This integral must be evaluated numerically for specified values of m, n and δ_k. Teichroew (1954) has computed the value of (3.24) on the SWAC (National Bureau of Standards Western Automatic Computer) for $\delta_k = 0(.01)6.40$ and $\delta_k = -3.20(.10)0$ and forty-five combinations of m and n, by summing the products of the terms in the integrand evaluated at 160 values of x, $x = -8.0(.1)7.9$. Several systematic checks were performed throughout the program to ensure accuracy, and the results are believed to be correct to within one unit in the ninth decimal place. Direct linear interpolation for positive values of δ_k gives probabilities correct to within one unit in the fifth decimal place. When δ_k is negative, linearly interpolated values will be accurate to at least three decimal places.

Using Teichroew's calculations, the power of the one-sided test T_3 of the hypothesis H_o: $H = G$, unspecified, against the alternative H_1' can be calculated for any positive or negative value of δ_k within the tabulated range when $\alpha \leq 1/\binom{N}{m}$.

The power functions of the other tests considered are also of interest under the assumptions of the alternative H_1' and can be calculated from Teichroew's results. For a one-sided test where $(\mu_Y - \mu_X) > 0$, which corresponds to $H(a) \geq G(a)$ for all a, we would be interested only in positive values of δ_k. For two-sided tests, we need to compute

$$\Pr(\text{all } n \text{ Y's} < \text{all } m \text{ X's}) = m\int_{-\infty}^{\infty} [\Phi(x-\delta_k)]^n \varphi(x)[\Phi(-x)]^{m-1}\, dx$$

$$= \int_{-\infty}^{\infty} [\Phi(x-\delta_k)]^n \, d[\Phi(-x)^m] \quad . \tag{3.25}$$

But for any given positive value of δ_k, (3.25) is equivalent to (3.24) with $-\delta_k$ substituted for δ_k. Thus Teichroew's results yield the probabilities of both of the two most extreme orderings where all the observations in one sample are less than all observations in the other sample.

We will choose δ_k so that the power functions obtained can be compared with those presented in Tables 3.1 - 3.6

where F is not specified. Thus δ_k will correspond to the difference between H and G under the alternative H_1: $H=1-(1-F)^k$, $G=F^k$ where F is the normal distribution with mean 0 and variance 1. Since F is symmetrical about the origin, the even moments of H and G are equal and the odd moments are the negatives of each other. G is the distribution of the largest of k random variables from a sample from a standard normal population. The standardized distance between H and G is measured by

$$\delta_k = \frac{\mathcal{E}_G(x) - \mathcal{E}_H(x)}{\sigma_G(x)} = \frac{2\ \mathcal{E}_G(x)}{\sigma_G(x)} = \frac{2\ \mathcal{E}_F(x_{(k)})}{\sigma_F(x_{(k)})} \qquad (3.26)$$

where $x_{(k)}$ denotes the kth order statistic from a standardized normal population. The value of δ_k can be determined by using Ruben's Table 3 (1954, p. 226) to obtain $\sigma_F(x_{(k)})$ and Table 2 (p. 224, r = 1) to obtain $\mathcal{E}_F(x_{(k)})$ (or equivalently Harter, 1961, Table 1, p. 158). The values of δ_k are presented in Table 3.7 for k = 2(1)7.

Table 3.7. Numerical values of $\delta_k = \dfrac{2\ \mathcal{E}_F(x_{(k)})}{\sigma_F(x_{(k)})}$ for k=2(1)7

k	2	3	4	5	6	7
δ_k	1.367	2.263	2.936	3.477	3.930	4.320

The power functions for testing the null hypothesis H_o versus the normal alternative H_1' using the one-sided tests T_1, T_2, T_3, T_4, and T_8 when $\alpha \leq 1/\binom{N}{m}$ and the two-sided tests T_5, T_6, and T_7 when $\alpha \leq 2/\binom{N}{m}$ are given in Table 3.8 for m = n = 2, 3, and 4 and the six values of δ_k corresponding to k = 2(1)7. The power under H_1 (Tables 3.1 - 3.5) is, in most cases, slightly higher than when F is assumed to be normal. The difference in power appears to increase with k as well as with the sample size.

The power functions against H_1' (for all tests except the two median tests) for unequal sample sizes have also been computed from Teichroew's tables. The results are presented in Table 3.9 for $\alpha \leq 1/\binom{N}{m}$ for one-sided tests and $\alpha \leq 2/\binom{N}{m}$ for two-sided tests, and may be compared with the power functions in Table 3.6.

Table 3.8. Power function of H_o versus H_1' for equal sample sizes

m=n		k / α	2 δ=1.367	3 δ=2.263	4 δ=2.936	5 δ=3.477	6 δ=3.390	7 δ=4.320
2	One-Sided Tests	.01	.0352	.0501	.0562	.0585	.0594	.0597
		.05	.1762	.2505	.2810	.2926	.2970	.2987
		.10	.3524	.5010	.5620	.5851	.5840	.5974
	Two-Sided Tests	.01	.0180	.0251	.0281	.0293	.0297	.0299
		.05	.0902	.1254	.1405	.1463	.1485	.1494
		.10	.1804	.2509	.2810	.2926	.2970	.2987
3	One-Sided Tests	.01	.0777	.1424	.1758	.1900	.1958	.1982
		.05	.3875	.7119	.8788	.9500	.9790	.9909
	Two-Sided Tests	.01	.0389	.0712	.0879	.0950	.0979	.0991
		.05	.1943	.3560	.4394	.4750	.4895	.4955
		.10	.3886	.7119	.8788	.9500	.9790	.9909
4	One-Sided Tests	.01	.1728	.4156	.5701	.6438	.6757	.6893
		1/70	.2468	.5937	.8145	.9197	.9653	.9847
	Two-Sided Tests	.01	.0864	.2078	.2851	.3219	.3378	.3446
		2/70	.2469	.5937	.8145	.9197	.9653	.9847

Table 3.9. Power function of H_o versus H_1' for unequal sample sizes

m,n or n,m		k / α	2	3	4	5	6	7
			δ=1.367	δ=2.263	δ=2.936	δ=3.477	δ=3.930	δ=4.320
1,3	One-Sided Tests	.01	.0266	.0349	.0381	.0392	.0397	.0399
		.05	.1328	.1743	.1903	.1862	.1985	.1994
		.10	.2657	.3486	.3806	.3925	.3970	.3987
	Two-Sided Tests	.01	.0140	.0175	.0191	.0196	.0198	.0199
		.05	.0701	.0877	.0953	.0981	.0992	.0997
		.10	.1403	.1755	.1905	.1963	.1985	.1994
1,4	One-Sided Tests	.01	.0306	.0422	.0470	.0488	.0495	.0498
		.05	.1530	.2110	.2349	.2440	.2476	.2490
		.10	.3060	.4220	.4697	.4880	.4951	.4979
	Two-Sided Tests	.01	.0159	.0212	.0235	.0244	.0248	.0249
		.05	.0795	.1059	.1175	.1220	.1238	.1245
		.10	.1589	.2118	.2350	.2441	.2476	.2490
2,3	One-Sided Tests	.01	.0492	.0781	.0912	.0965	.0985	.0994
		.05	.2462	.3905	.4561	.4824	.4927	.4969
		.10	.4925	.7810	.9122	.9648	.9855	.9938
	Two-Sided Tests	.01	.0249	.0391	.0456	.0482	.0493	.0497
		.05	.1243	.1953	.2280	.2412	.2464	.2484
		.10	.2486	.3906	.4561	.4824	.4927	.4969

Table 3.9 continued

m,n or n,m		k / α	2 δ=1.367	3 δ=2.263	4 δ=2.936	5 δ=3.477	6 δ=3.930	7 δ=4.320
1,5	One-Sided Tests	.01	.0342	.0492	.0557	.0583	.0593	.0597
		.05	.1712	.2461	.2786	.2914	.2964	.2985
		.10	.3424	.4923	.5571	.5828	.5929	.5969
	Two-Sided Tests	.01	.0176	.0247	.0279	.0291	.0296	.0298
		.05	.0881	.1233	.1393	.1457	.1482	.1492
		.10	.1762	.2467	.2786	.2914	.2964	.2985
2,4	One-Sided Tests	.01	.0638	.1105	.1336	.1433	.1472	.1488
		.05	.3189	.5526	.6680	.7164	.7360	.7439
		1/15	.4253	.7368	.8907	.9552	.9813	.9919
	Two-Sided Tests	.01	.0320	.0553	.0668	.0716	.0736	.0744
		.05	.1602	.2763	.3340	.3582	.3680	.3720
		.10	.3205	.5527	.6680	.7164	.7360	.7439
1,6	One-Sided Tests	.01	.0376	.0560	.0643	.0677	.0690	.0696
		.05	.1879	.2799	.3215	.3484	.3451	.3479
		.10	.3759	.5599	.6430	.6768	.6903	.6958
	Two-Sided Tests	.01	.0192	.0280	.0322	.0338	.0345	.0348
		.05	.0961	.1402	.1608	.1692	.1726	.1740
		.10	.1923	.2804	.3215	.3384	.3451	.3479

Table 3.9 continued

m,n or n,m		k / α	2 δ=1.367	3 δ=2.263	4 δ=2.936	5 δ=3.477	6 δ=3.930	7 δ=4.320
2,5	One-Sided Tests	.01	.0787	.1469	.1830	.1987	.2052	.2079
		1/21	.3748	.6996	.8715	.9464	.9773	.9901
	Two-Sided Tests	.01	.0395	.0735	.0915	.0994	.1026	.1040
		.05	.1973	.3673	.4576	.4968	.5131	.5198
		2/21	.3758	.6997	.8715	.9464	.9773	.9901
3,4	One-Sided Tests	.01	.1109	.2298	.2974	.3278	.3406	.3459
		1/35	.3168	.6566	.8498	.9366	.9730	.9882
	Two-Sided Tests	.01	.0555	.1149	.1487	.1639	.1703	.1729
		.05	.2774	.5746	.7436	.8195	.8514	.8647
		2/35	.3171	.6566	.8498	.9366	.9730	.9822
1,7	One-Sided Tests	.01	.0407	.0625	.0727	.0770	.0787	.0795
		.05	.2034	.3126	.3637	.3850	.3937	.3973
		.10	.4068	.6251	.7275	.7701	.7874	.7945
	Two-Sided Tests	.01	.0207	.0313	.0364	.0385	.0394	.0397
		.05	.1036	.1565	.1819	.1925	.1968	.1986
		.10	.2072	.3129	.3638	.3850	.3937	.3973

Table 3.9 continued

m,n or n,m		k / α	2 δ=1.367	3 δ=2.263	4 δ=2.936	5 δ=3.477	6 δ=3.930	7 δ=4.320
2,6	One-Sided Tests	.01	.0939	.1870	.2392	.2695	.2726	.2768
		1/28	.3353	.6677	.8542	.9381	.9735	.9884
	Two-Sided Tests	.01	.0470	.0935	.1196	.1348	.1363	.1384
		.05	.2351	.4674	.5979	.6738	.6815	.6919
		2/28	.3359	.6677	.8542	.9381	.9735	.9884
3,5	One-Sided Tests	.01	.1489	.3422	.4615	.5175	.5417	.5519
		1/56	.2660	.6110	.8241	.9242	.9673	.9856
	Two Sided Tests	.01	.0745	.1711	.2307	.2588	.2708	.2760
		2/56	.2661	.6110	.8241	.9242	.9673	.9856
1,8	One-Sided Tests	.01	.0436	.0688	.0811	.0863	.0884	.0893
		.05	.2179	.3442	.4054	.4314	.4421	.4465
		.10	.4358	.6884	.8108	.8628	.8842	.8931
	Two-Sided Tests	.01	.0221	.0344	.0405	.0431	.0442	.0465
		.05	.1107	.1722	.2027	.2157	.2210	.2233
		.10	.2214	.3445	.4054	.4314	.4421	.4465

Table 3.9 continued

m,n or n,m		k / α	2 δ=1.367	3 δ=2.263	4 δ=2.936	5 δ=3.477	6 δ=3.930	7 δ=4.320
1,9	One-Sided Tests	.01	.0463	.0750	.0893	.0955	.0981	.0992
		.05	.2315	.3749	.4465	.4774	.4903	.4958
		.10	.4630	.7498	.8929	.9548	.9807	.9915
	Two-Sided Tests	.01	.0235	.0375	.0446	.0477	.0490	.0496
		.05	.1174	.1876	.2232	.2387	.2452	.2479
		.10	.2347	.3752	.4465	.4774	.4903	.4958

The alternative H_1' states a simple shift in location of two normal distributions, even though the hypothesis H_o does not specify the common distribution function. If a parametric test were to be employed with the alternative H_1' (with $\mu_Y - \mu_X > 0$), the hypothesis would be H_o': $\mu_Y - \mu_X = 0$. If we are willing to assume that both sets of random variables are independent and normally distributed with equal but unknown variances, the two-sample Student's t test is the uniformly most powerful test against the one-sided alternative H_1'. The test is to reject H_o' when

$$\frac{\bar{Y}-\bar{X}}{S(\frac{1}{m}+\frac{1}{n})^{\frac{1}{2}}} > t_{m+n-2,\alpha} \quad , \tag{3.27}$$

where $(m+n-2)S^2 = \Sigma(X_i-\bar{X})^2 + \Sigma(Y_i-\bar{Y})^2$, and $t_{m+n-2,\alpha}$ is the upper α point of the Student's t distribution with (m+n-2) degrees of freedom.

It is only natural to compare the power functions of the one-sided nonparametric tests of H_o versus H_1' with the power of this parametric test. Under the alternative H_1', the left-hand side of (3.27), denoted by t'_{m+n-2,ρ_k} is distributed as a noncentral t variable with m+n-2 degrees of freedom and noncentrality parameter $\rho_k = \frac{\delta_k \sqrt{mn}}{\sqrt{m+n}}$. The power for significance level α is $Pr[t'_{m+n-2,\rho_k} > t_{m+n-2,\alpha}]$. The power functions for $\alpha = .01$ and $\alpha = .05$ can be read off from Neyman's (1935, pp. 133-134) series of curves representing the probabilities of a Type II error. The results are presented in Table 3.10. The normal theory power is of course higher than the power functions given in Tables 3.1 - 3.5 and 3.8 for the one-sided nonparametric tests.

Table 3.10. Power of the one-sided, two-sample Student's _t_ test of H_o' versus H_1'

m=n	α \ k	2 $\delta_2=1.367$	3 $\delta_3=2.263$	4 $\delta_4=2.936$	5 $\delta_5=3.477$	6 $\delta_6=3.930$	7 $\delta_7=7.320$
2	.01	.05	.12	.18	.24	.28	.33
	.05	.24	.47	.61	.72	.80	.84
3	.01	.13	.34	.52	.69	.79	.86
	.05	.42	.75	.90	.96	.98	.99
4	.01	.21	.59	.80	.92	.96	.99
	.05	.53	.87	.96	.99	-	-

Chapter IV

LOCALLY MOST POWERFUL RANK TESTS

"The idea of a statistical test of a hypothesis and the related concepts introduced by Neyman and Pearson have served as a model for much of modern statistics. In nonparametric work it is seldom possible to apply all of these concepts. This results from the fact that for most of the alternatives that have been considered there do not exist optimum critical regions or analytic tools for finding power functions. The sign test gives an illustration where it is possible to find the exact power function; on the other hand, this procedure is seldom optimum." (Savage, 1956, p. 590).

For the many two-sample nonparametric tests based on ranks, it is always possible, at least in theory, to find the probability under the alternative of any arrangement of the random variables by means of Hoeffding's theorem. The only conditions are that the two population distributions, H and G, should both be continuous and functionally related. Theoretically, then, the most powerful two-sample rank test of the hypothesis that both samples come from the same population can always be determined. However, at least some of the order probabilities must be calculated if one wishes to ascertain which orderings belong in the rejection region. The evaluation of the resulting multiple integral is often difficult. Also, one must have a relatively specific alternative in mind.

Another criterion which can easily be applied to two-sample nonparametric tests based on ranks is the concept of a "locally" most powerful test. Here the alternative distribution functions contain an unspecified parameter, say θ, which is equal to zero under the null hypothesis. The power of the rank test is maximized when this parameter θ is very close to zero by maximizing the slope of the power function at the point where θ is equal to zero. Thus a locally most powerful rank test admits a very general alternative.

In this chapter, we will show that a general test statistic can be obtained which will yield a rank test locally most powerful against any alternative expressing a functional relationship between the distribution functions of the two samples. The specific relationship between the distributions will depend on the parameter θ. Some of the properties of the resulting test statistic will be discussed.

The locally most powerful rank test will be found for two functional alternatives, both of which are similar to the earlier alternative of two extreme distributions. The resulting test statistics, which we will call the Gamma test and the Psi test, are considerably easier to apply than the most powerful rank test. Although the power functions of these tests would be expected to be lower than for the most powerful rank test, they are found to be quite close and indeed the same in several cases. The Psi Test was discussed briefly in Chapter III. Some of its properties will be investigated here.

The methods of this chapter can also be applied to an alternative specifying the two distribution functions. If we consider the alternative that both populations are normally distributed with the same variance but different means, we can let the parameter θ represent the difference between the population means. Terry's c_1 test is shown to be the locally most powerful rank test against this normal alternative, a result previously established by different methods (Terry, 1952).

4.1 Derivation of Test

We are again studying the situation in which there are two independent samples, $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$, of random variables, with continuous cumulative distribution functions H and G, respectively. For the null hypothesis H_o: H = G, a general alternative involving the parameter θ is H_a^*: $H = Q(G,\theta)$, G = G, where θ is restricted to lie in the interval $(0,\delta)$ for some $\delta > 0$. This is much more general than the alternative H = Q(G), G = G discussed previously, since H_a^* represents a large class of alternatives. The specific H function depends on the value assumed by θ. We will impose the restrictions that $Q(G,\theta)$ is a continuous cumulative distribution function for all θ in $(0,\delta)$ and that $Q(G,0) = G$ so that the null hypothesis H_o is true when θ is equal to zero.

We wish to derive (cf., Capon, 1961) the locally most powerful rank test of the hypothesis H_o against the alternative H_a^* for $\theta > 0$, i.e., the test which maximizes the slope of the power function at the point $\theta = 0$. Again we can assume without loss of generality that G is uniform, $G(u) = u$, and then write the Q function as $Q(u,\theta)$. The density function of Q will be denoted by $q(u,\theta)$, and $\partial q(u,\theta)/\partial\theta|\theta = \varphi$ by $q_\theta'(u,\varphi)$.

Let us make the following assumptions:

(i) For almost all u, the derivatives $q(u,\theta)$ and $q_\theta'(u,\theta)$ exist and are continuous with respect to θ in the interval $(0,\delta)$.

(ii) There exist functions $M_o(u)$ and $M_1(u)$, both integrable over (0,1) and independent of θ, such that

$$q(u,\theta) \leq M_o(u), \qquad |q_\theta'(u,\theta)| \leq M_1(u)$$

for $0 \leq \theta \leq \delta$.

By means of Hoeffding's theorem, the probability under the alternative H_a^* of any arrangement $\overline{z}$ of the N random variables is given by

$$p_\theta(\overline{z}) = \mathcal{E}_o[\prod_{j=1}^{m} q(U_{r_j},\theta)]/\binom{N}{m} , \tag{4.1}$$

where $0 \leq U_1 \leq U_2 \leq \ldots \leq U_N \leq 1$ are the order statistics for a random sample of size N from the uniform distribution, and the r_j $(j=1,2,\ldots,m)$ are the ranks of the X random variables in the combined sample. The expectation is taken under the null hypothesis $\theta = 0$. With assumption (i), we can form the Taylor's series expansion of $p_\theta(\overline{z})$ about the point $\theta = 0$ and obtain

$$p_\theta(\overline{z}) = p_0(\overline{z}) + \theta\partial p_\theta/\partial\theta|_{\theta=0} + R(u,\theta) \quad . \tag{4.2}$$

From (4.1), since $q(u,0) = 1$,

$$p_0(z) = \mathcal{E}_o[\prod_{j=1}^{m} q(U_{r_j},0)]/\binom{N}{m} = 1/\binom{N}{m} \quad . \tag{4.3}$$

Also,

$$\binom{N}{m} \frac{\partial p_\theta}{\partial \theta} \Big|_{\theta=0} = \frac{\partial}{\partial\theta} \{\mathcal{E}_o[\prod_{j=1}^{m} q(U_{r_j},\theta)]\}\Big|_{\theta=0} \quad .$$

The differentiation can be performed under the integral sign as a consequence of condition (ii) and a well-known theorem (Cramér, 1946, p. 67). Then

$$\begin{aligned} \binom{N}{m} \frac{\partial p_\theta}{\partial \theta} \Big|_{\theta=0} &= \mathcal{E}_o[\frac{\partial}{\partial\theta} \prod_{j=1}^{m} q(U_{r_j},\theta)\Big|_{\theta=0}] \\ &= \mathcal{E}_o[\sum_{j=1}^{m} q'_\theta(U_{r_j},0) \prod_{\substack{i=1 \\ i\neq j}}^{m} q(U_{r_i},0)] \\ &= \mathcal{E}_o[\sum_{j=1}^{m} q'_\theta(U_{r_j},0)] \quad . \end{aligned} \tag{4.4}$$

The remainder term $R(u,\theta)$ is of a smaller order than θ if the derivative $\partial p_\theta/\partial\theta$ is continuous for all θ in the admissible range (Cramér, 1946, p. 122). We have

$$\frac{\partial p_\theta}{\partial\theta} = m!n! \int_0^1\int_0^{u_N} \dots \int_0^{u_2} \sum_{j=1}^{m} q'_\theta(u_{r_j},\theta) \prod_{\substack{i=1 \\ i\neq j}}^{m} q(u_{r_i},\theta) \; du_1 du_2 \dots du_N \; ,$$

and condition (i) ensures the continuity of the integrand for almost all u. Then the integral is continuous for an arbitrary θ_o in $(0,\delta)$, and hence for all θ in the interval. Therefore, $\lim_{\theta\to 0} \frac{R(u,\theta)}{\theta} = 0$, and we write $R(u,\theta) = 0(\theta)$.

Substituting (4.2) and (4.4) in (4.1), we obtain

$$\binom{N}{m} p_\theta(\overline{z}) = 1 + \theta\, \mathcal{E}_o[\sum_{j=1}^{m} q_\theta'(U_{r_j},0)] + 0(\theta) \quad .$$

For θ sufficiently small, it follows that $p_\theta(\overline{z}) > p_\theta(\overline{z}')$ if and only if

$$\sum_{j=1}^{m} \mathcal{E}_o[q_\theta'(u_{r_j},0)] > \sum_{j=1}^{m} \mathcal{E}_o[q_\theta'(u_{r_j'},0)] \quad ,$$

where the r_j' are the ranks of the X random variables in the arrangement $\overline{z}'$. Thus we have proved the following theorem.

Theorem 4.1. The locally most powerful rank test of H_o: $H = G$ against the alternative H_a^*: $H = Q(u,\theta)$, $G = u$, $\theta > 0$, is to reject H_o when

$$\sum_{j=1}^{m} \mathcal{E}_o[q_\theta'(U_{r_j},0)] > c \quad , \tag{4.5}$$

where c is a constant determined by the size of the test.

A convenient notation for this type of test statistic is

$$T_N = \frac{1}{m} \sum_{i=1}^{N} a_i z_i \tag{4.6}$$

where the z_i's are the indicator variables previously defined, and the a_i are constants. T_N is known as a linear rank statistic.

4.2 Properties of the Test

For any linear rank statistic, it is easy to obtain exact moments under the null hypothesis. The first two of these can be determined from the following theorem (Savage, 1956).

Theorem 4.2. Let $T_N = \frac{1}{m} \sum_{i=1}^{N} a_i z_i$, where the a_i are constant, $z_i = 1$ if the ith ordered variable in the combined sample of size N corresponds to an X random variable, and $z_i = 0$ otherwise. The exact mean and variance of T_N under the null hypothesis are given by

$$\mathcal{E}_0(T_N) = \frac{1}{N} \sum_{i=1}^{N} a_i \quad , \tag{4.7}$$

$$\sigma_0^2(T_N) = \frac{n}{mN^2(N-1)} [N \sum_{i=1}^{N} a_i^2 - (\sum_{i=1}^{N} a_i)^2] \quad . \tag{4.8}$$

Proof. Since $\mathcal{E}_0(z_i) = m/N$, (4.7) is obvious. Also, $\mathcal{E}_0(z_i^2) = m/N$ and $\underset{i \neq k}{\mathcal{E}_0(z_i z_k)} = \frac{m(m-1)}{N(N-1)}$, so that $\sigma_0^2(z_i) = mn/N^2$ and $\text{cov}_0(z_i, z_k) = \frac{-mn}{N^2(N-1)}$. It follows that

$$\sigma_o^2(T_N) = \frac{1}{m^2}\left[\sum_{i=1}^{N} a_i^2\ \sigma_o^2(z_i) - \sum\sum_{i\neq k} a_i a_k \mathrm{cov}_o(z_i, z_k)\right]$$

$$= \frac{1}{m^2}\left[\frac{mn}{N^2}\sum_{i=1}^{N} a_i^2 - \frac{mn}{N^2(N-1)}\sum\sum_{i\neq k} a_i a_k\right]$$

$$= \frac{n}{mN^2(N-1)}\left[N\sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i\right)^2\right] .$$

Corollary 4.2.1. If $B_N = \frac{1}{m}\sum_{i=1}^{N} b_i z_i$ is another linear rank statistic, the covariance between B_N and T_N under the null hypothesis is given by

$$\mathrm{cov}_o(T_N, B_N) = \frac{n}{mN^2(N-1)}\left[N\sum_{i=1}^{N} a_i b_i - \sum_{i=1}^{N} a_i \sum_{i=1}^{N} b_i\right] . \quad (4.9)$$

We now turn to the asymptotic properties of the linear rank statistic (Chernoff and Savage, 19e8; Capon, 1961). For the purpose of the ensuing discussion only, we will consider the representation

$$T_N = \int_{u=0}^{1} J_N[R_N(u,\theta)]\ dQ_m(u,\theta) \quad (4.10)$$

where $Q_m(u,\theta)$ and $Q_n(u,0)$ are the empirical distribution functions of the X and Y random variables respectively, and

$$R_N(u,\theta) = mQ_m(u,\theta)/N + nQ_n(u,0)/N .$$

Then $R_N(u,\theta)$ is the proportion of variables in the combined sample which are less than or equal to u, and may assume the values 0, 1/N, 2/N, ..., N/N.

$Q_m(u,\theta)$ is a step function which can take jumps only at the points U_1, U_2, ..., U_N, and will increase by 1/m at U_i if U_i is an X random variable ($z_i = 1$). If U_i is a Y ($z_i = 0$), Q_m will remain constant. Then the representation of T_N in (4.10) means (von Mises, 1947)

$$T_N = \sum_{i=1}^{N} J_N[R_N(U_i,\theta)] \frac{z_i}{m} = \frac{1}{m} \sum_{i=1}^{N} J_N(i/N)\ z_i \quad .$$

We see that (4.10) is equivalent to (4.6) when $a_i = J_N(i/N)$.

Although J_N need be defined only at 0, 1/N, ..., N/N, we will assume that J_N is constant on $(\frac{i-1}{N}, \frac{i}{N}]$ so that its domain is the entire interval (0,1). Denote by I_N the interval where $0 < R_N(u,\theta) < 1$.

Chernoff and Savage have proved the following theorem concerning the asymptotic normality of linear rank statistics. If the linear rank statistic is the locally most powerful rank test, the two regularity conditions used in the proof of Theorem 4.1 must also be satisfied.

Theorem 4.3. If

(iii) $0 < \lim_{N\to\infty} m/n = r < \infty$,

(iv) $J(R) = \lim_{N\to\infty} J_N(R)$ exists for $0 < R < 1$ and is not constant, where $R = R_\theta(u) = mQ(u,\theta)/N + nQ(u,0)/N$,

(v) $\int_{I_N} [J_N(R_N) - J(R_N)]\, dQ_m(u,\theta) = o_p(N^{-\frac{1}{2}})$,

(vi) $J_N(1) = o(N^{\frac{1}{2}})$,

(vii) $|J^{(i)}(R)| = |\frac{d^i J}{dR^i}| \leq K[R(1-R)]^{-i-\frac{1}{2}+\delta}$ for $i = 0,1,2$ and some $\delta > 0$, where K is a constant independent of i,N,m,n, $Q(u,\theta)$, and u, then

$$\lim_{N\to\infty} \Pr[\frac{T_N - \mathcal{E}(T_N)}{\sigma(T_N)} \leq a] = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt \quad , \qquad (4.11)$$

where

$$\mathcal{E}(T_N) = \int_0^1 J[R_\theta(u)]\, q(u,\theta)\, du \qquad (4.12)$$

and

$$\sigma^2(T_N) = \frac{2n}{N^2}\left[\iint_{0<u<v<1} u(1-v) J'[R_\theta(u)] J'[R_\theta(v)] q(u,\theta) q(v,\theta)\, du dv + \frac{n}{m} \iint_{0<u<v<1} Q(u,\theta)[1-Q(v,\theta)] J'[R_\theta(u)] J'[R_\theta(v)]\, du dv\right] . \qquad (4.13)$$

The expression J(R) can easily be determined for any locally most powerful rank test statistic by the following corollary.

<u>Corollary 4.3.1</u>. For the linear rank statistic $T_N = \frac{1}{m} \sum_{i=1}^{N} a_i z_i$ where $a_i = \mathcal{E}_o[q'_\theta(U_i,0)]$, we have

$$J(R) = q'_\theta(u,0)\big|_{u=R} \quad . \tag{4.14}$$

Proof. $J_N(i/N) = a_i = \mathcal{E}_o[\, q'_\theta(U_i,0)] = \mathcal{E}_o[f(U_i)]$ say, where $f(u) = q'_\theta(u,0)$. Asymptotically, we have $\mathcal{E}[f(U)] = f[\mathcal{E}(U)]$. But since U_i is the <u>i</u>th smallest order statistic of a sample of size N from a uniform distribution,

$\mathcal{E}_o(U_i) = i/(N+1) = \dfrac{i/N}{1+(1/N)}$. Then

$$J(R) = \lim_{N\to\infty} J_N(R) = \lim_{N\to\infty} f\left(\frac{R}{1+(1/N)}\right) = f(R) = q'_\theta(u,0)\big|_{u=R} \quad .$$

<u>Corollary 4.3.2</u>. Under the null hypothesis, the mean and variance of T_N are given by

$$\mathcal{E}_o(T_N) = 0 \tag{4.15}$$

$$\sigma_o^2(T_N) = \frac{n}{mN} \{\mathcal{E}_o[q'_\theta(U,0)]^2\} \quad . \tag{4.16}$$

Proof. Under H_o, $\theta = 0$ and $Q(u,0) = u$, so that $R_0(u) = u$ and $J[R_0(u)] = J(u) = q'_\theta(u,0)$ by (4.14). From (4.12) and (4.13), we have

$$\mathcal{E}_o(T_N) = \int_0^1 q'_\theta(u,0)\,du = \frac{\partial}{\partial\theta}\int_0^1 q(u,\theta)\,du\Big|_{\theta=0} = 0 \quad .$$

$$\sigma_o^2(T_N) = \frac{2n}{N^2}\left(1+\frac{n}{m}\right)\iint_{0<u<v<1} u(1-v)J'(u)J'(v)du\,dv$$

$$= \frac{2n}{mN}\iiiint_{0<x<u<v<y<1} J'(u)J'(v)dx\,du\,dv\,dy$$

$$= \frac{2n}{mN}\int_0^1 dy\int_0^y dx\int_x^y J'(u)du\int_u^y J'(v)dv$$

$$= \frac{n}{mN}\int_0^1 dy\int_0^y [J(y) - J(x)]^2 dx$$

$$= \frac{n}{2mN}\left[\int_0^1 J^2(y)dy - 2\int_0^1 J(y)dy\int_0^1 J(x)dx + \int_0^1 J^2(x)dx\right]$$

$$= \frac{n}{mN}\left[\int_0^1 J^2(x)dx - \left(\int_0^1 J(x)dx\right)^2\right]$$

$$\sigma_o^2(T_N) = \frac{n}{mN}\int_0^1 J^2(x)dx \quad .$$

Using (4.14), this is equivalent to the form (4.16).

Since verification of the regularity conditions (iv) - (vii) for a particular R function may be a time-consuming operation, Chernoff and Savage have also proved the following useful theorem.

<u>Theorem 4.4</u>. If $J_N(i/N)$ is the expectation of the <u>i</u>th order statistic of a sample of size N from a population whose cumulative distribution function is the inverse function of J

and condition (vii) is satisfied, then (iv), (v) and (vi) are satisfied.

Obviously, if J is to have an inverse which is a cumulative distribution function, $J_N(i/N) = a_i$ must be a nondecreasing function of i. For an alternative of the form $H(a,\theta) \leq G(a,0)$ for all $\underline{a}$, where our test is to reject if $T_N > c$, a_i must increase with i. However, if the alternative states that the Y's are stochastically larger than the X's and we still wish to reject for $T_N > c$, a_i must be a nonincreasing function of i. Since this is equivalent to rejecting if $\sum_{i=1}^{N} -a_i z_i < c$, we can assume without loss of generality that $J_N(i/N)$ is a nondecreasing function of i for any alternative.

In many cases, Theorem 4.4 greatly simplifies the application of Theorem 4.3 to the locally most powerful rank test statistic. This is evident from the following corollary.

Corollary 4.4.1. Let $J_N(i/N) = \pm\, \mathcal{E}_0[q'_\theta(U_i,0)]$ (where the algebraic sign is determined so that $J_N(i/N)$ is a nondecreasing function of i). If $q'_\theta(u,0)$ has an inverse and condition (vii) of Theorem 4.3 is satisfied, then (iv), (v) and (vi) hold.

As an example of the use of the preceding theoretical discussion, consider the alternative $H = Q(u,\theta) = (1-\theta)u + \theta u^2$, $0 \leq \theta \leq 1$. This is a special case of the general alternative $H(a) \leq G(a)$ for all a. When $\theta = 1$, it reduces to Lehmann's alternative of one extreme distribution for the special case $k = 2$.

Since $q'_\theta(u,0) = 2u-1$ and $\mathcal{E}_0(U_i) = i/(N+1)$, the linear rank statistic in the form (4.6) is

$$T_N = \frac{1}{m} \sum_{i=1}^{N} [\frac{2i}{N+1} - 1] \; z_i \quad ,$$

and we reject H_0 if $T_N > c$. This statistic is a linear function of the Mann-Whitney or Wilcoxon statistic which rejects when

$$W'_N = \sum_{i=1}^{N} iz_i = m(N+1)(T_N+1)/2 > c \; .$$

Although the exact mean and variance of T_N can easily be computed, only the asymptotic properties will be discussed in this chapter.

From (4.14), we have $J(R) = 2R-1$ and the regularity condition (vii) is obviously satisfied. $J_N(i/N) = \frac{2i}{N+1} - 1$, an increasing function of i, can be considered as the expectation of the ith order statistic of a sample of size N from the distribution $F_X(x) = J^{-1}(x) = (x+1)/2$, $-1 \leq x \leq 1$. The other conditions of Theorem 4.3 are then satisfied. From (4.15) and (4.16), $\mathcal{E}_o(T_N) = 0$ and

$$\sigma_o^2(T_N) = \frac{n}{mN}\int_0^1 (2u-1)^2\, du = \frac{n}{3mN} ,$$

so that the asymptotic mean and variance of W_N' under the null hypothesis are $m(N+1)/2$ and $mn(N+1)^2/12N$. Under the alternative, since

$$R_\theta(u) = \frac{m}{N}\,[(1-\theta)u + \theta u^2] + \frac{n}{N}\, u , \quad \text{we have}$$

$$\mathcal{E}_a(T_N) = \int_0^1 \left[2\left\{\frac{m}{N}[(1-\theta)u+\theta u^2] + \frac{n}{N}\, u\right\} - 1\right](1-\theta+2u\theta)\,du$$

$$= \frac{\theta n}{3N} ,$$

and

$$\sigma_a^2(T_N) = \frac{2n}{N^2}\left[\begin{array}{l}\displaystyle\iint\limits_{0<u<v<1} 4u(1-v)(1-\theta+2u\theta)(1-\theta+2v\theta)\,du\, dv \\ \displaystyle + \frac{n}{m}\iint\limits_{0<u<v<1} 4[(1-\theta)u+\theta u^2][1-(1-\theta)v-\theta v^2]\,du\, dv\end{array}\right]$$

$$= \frac{n}{45N^2}\,[15 + \theta^2 + \frac{5n}{m}\,(3-\theta^2)]$$

from (4.12) and (4.13). The asymptotic power function could then easily be determined using (4.11) (cf., Section 5.1).

In general, if the conditions (i) - (vii) are satisfied, and the mean and variance of the linear rank statistic T_N

under both the hypothesis and the alternative exist and can be evaluated either exactly or asymptotically, the asymptotic power of any locally most powerful rank test can be determined from (4.11).

Capon has also demonstrated the asymptotic efficiency of the locally most powerful rank test based on a sequence of linear rank statistics T_N with a corresponding sequence of alternatives $Q(u,\theta_N)$, with θ_N getting closer and closer to zero as N gets large. That is, the asymptotic relative efficiency of the T_N test versus the likelihood ratio test of the same hypothesis and alternative is equal to one.

4.3 Application to Specific Alternatives

We will now consider various alternatives of the form $H = Q(G,\theta)$ and determine the locally most powerful rank test from Theorem 4.1. The exact power of the tests will be the sum of the probabilities under the alternative for those cases lying in the rejection region, which are determined by (4.5).

4.3.1 The Gamma Test

Our first alternative H_1^* is a mixture of the two distribution functions $H = 1 - (1-F)^k$, $G = F^k$, assumed under the earlier alternative H_1. We write

$$H_1^*: H = (1-\theta)F^k + \theta[1 - (1-F)^k], \; G = F^k \quad (0 \leq \theta \leq 1) .$$

Thus H_1^* reduces to H_o for $\theta = 0$ and to H_1 for $\theta = 1$. Taking $G(u) = F^k(u) = u$, the H function becomes

$$H = (1-\theta)u + \theta[1 - (1-u^{1/k})^k]$$

or $\quad H = Q(u,\theta) = u + \theta[1 - (1-u^{1/k})^k - u] \quad .$

Then $Q(u,\theta)$ satisfies the requirements of a cumulative distribution function for all θ as long as F is continuous. The density function is $q(u,\theta) = 1 + \theta[(1-u^{1/k})^{k-1}(u^{-1/k})^{k-1} - 1]$, and $q_\theta'(u,0) = (1-u^{1/k})^{k-1}(u^{-1/k})^{k-1} - 1$.

The locally most powerful rank test for a general k rejects when

$$\sum_{i=1}^{N} \mathcal{E}_o[(1-U_i^{1/k}) \; (U_i^{-1/k})]^{k-1} \; z_i > c \quad ,$$

where

$$\mathcal{E}_o[(1-u_i^{1/k})(u_i^{-1/k})]^{k-1} = \frac{N!}{(i-1)!(N-i)!}\int_0^1 (1-u^{1/k})^{k-1} \; u^{(1/k)-1} \cdot u^{i-1} \; (1-u)^{N-i} \, du$$

$$= \frac{N!}{(i-1)!(N-i)!}\sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \int_0^1 u^{(j/k)+i+(1/k)-2} \; (1-u)^{N-i} \, du$$

$$= \frac{N!}{(i-1)!} \sum_{j=0}^{k-1} \binom{k-1}{j}(-1)^j \; \Gamma(\frac{j+1}{k} + i-1)/\Gamma(N + \frac{j+1}{k}) \quad .$$

Neglecting constants, the Gamma test for any k is to reject when

$$\Gamma_N = \sum_{i=1}^{N} \frac{z_i}{(i-1)!} \sum_{j=0}^{k-2} \binom{k-2}{j} (-1)^j \; \Gamma\left(\frac{j+1}{k} + i-1\right) / \Gamma\left(N + \frac{j+1}{k}\right) > c \; .$$

In particular, the test rejects when

$$\sum_{i=1}^{N} \frac{z_i \; \Gamma(i-\frac{1}{2})}{(i-1)!} > c \quad \text{for } k = 2 \; , \tag{4.17}$$

$$\sum_{i=1}^{N} \frac{z_i}{(i-1)!} \left[\frac{\Gamma(i-\frac{2}{3})}{\Gamma(N+\frac{1}{3})} - \frac{2\Gamma(i-\frac{1}{3})}{\Gamma(N+\frac{2}{3})} \right] > c \quad \text{for } k = 3 \; , \tag{4.18}$$

$$\sum_{i=1}^{N} \frac{z_i}{(i-1)!} \left[\frac{\Gamma(i-\frac{3}{4})}{\Gamma(N+\frac{1}{4})} - \frac{3\Gamma(i-\frac{1}{2})}{\Gamma(N+\frac{1}{2})} + \frac{3\Gamma(i-\frac{1}{4})}{\Gamma(N+\frac{3}{4})} \right] > c \tag{4.19}$$

for $k = 4$, and so forth.

Using (4.17), (4.18) and (4.19), we find that the critical regions for $m = n = 2$ and $m = n = 3$, $k = 2$, 3 and 4, $\alpha = .01$, .05 and .10, are identical to those obtained for the most powerful rank test against the alternative H_1. Similarly for $m = n = 4$, $\alpha = .01$, $k = 2$, 3 and 4. For $m = n = 4$, $\alpha = .05$ and .10, the arrangements in the rejection regions and their probabilities of rejection to attain exact significance level α are as follows:

	$\alpha = .05$	$\alpha = .10$
k=2	$\left.\begin{matrix}(11110000),(11101000)\\(11100100)\end{matrix}\right\}1$ (11100010), .5	$\left.\begin{matrix}(11110000),(11101000),(11100100)\\(11100010),(11011000),(11100001)\\(11010100)\end{matrix}\right\}1$
k=3 and k=4	$\left.\begin{matrix}(11110000),(11101000)\\(11100100)\end{matrix}\right\}1$ (11100010), .5	$\left.\begin{matrix}(11110000),(11101000),(11100100)\\(11100010),(11100001),(11011000)\\(11010100)\end{matrix}\right\}1$

The cases here are listed in descending order of the numerical values obtained for (4.17), (4.18) and (4.19). Note that the order differs slightly for $\alpha = .10$, although the seven cases included are identical. For the locally most powerful test Γ_N then, the cases in the critical regions for $4/70 < \alpha < 6/70$ will depend on the value of k. On the other hand, it may be noted that, for all cases with $\alpha \leq .10$ and $m = n \leq 4$, the most powerful rank test against H_1, test T_1 of Section 3.4.1, has the same critical region for all values of k = 2, 3 and 4. The critical orders for the other tests in Section 3.4.1 are always independent of k.

The power functions when $\theta = 1$ (then $H_1^* = H_1$) for $m = n = 2$ and 3 are identical to those given for the test T_1 in Tables 3.1 and 3.2. The results of the power computations for $m = n = 4$ are given in Table 4.1.

Table 4.1. Power of the locally most powerful rank test against H_1^* when $\theta = 1$, $m = n = 4$

k	α=.01	α=.05	α=.10
2	.1705	.4830	.6513
3	.4293	.8456	.9452
4	.5950	.9662	.9937

The power here is, of course, lower than the corresponding values for T_1 given in Tables 3.3, 3.4 and 3.5 when $\alpha > 1/70$. However, this is to be expected, since θ is not really sufficiently small. The test is considerably easier to perform, as it requires no probability calculations to determine the cases in the critical regions, and it is easy to predict which orders will yield the highest values of Γ_N.

Although the function $J(u) = -[(1-u^{1/k})(u^{-1/k})]^{k-1}$ does have an inverse, $J^{-1}(x) = [1 + (-x)^{1/(k-1)}]^{-k}$, $-\infty < x < \infty$, the condition (vii) of Theorem 4.3 is not satisfied for $k \geq 2$. For example, if $i = 0$, we must have

$$|(1-u^{1/k})^{k-1} u^{(1/k)-1}| \leq K[u(1-u)]^{-\frac{1}{2}+\delta}$$

or $$K \geq |u^{(1/k)-\frac{1}{2}-\delta} (1-u)^{\frac{1}{2}-\delta} (1-u^{1/k})^{k-1}| .$$

Since for any $k \geq 2$, $\delta > 0$, $(1/k)-\frac{1}{2}-\delta < 0$, the right hand side of the inequality increases without bound as $u \to 0$.

Thus the asymptotic properties of the Gamma test statistic cannot be assessed by the methods of this chapter.

4.3.2 <u>The Psi Test</u>

Another alternative depending on θ is

$$H_1^{**}: H = 1 - (1-F)^{1+\theta}, \; G = F^{1+\theta}, \; \theta \geq 0 \; ,$$

or

$$H_1^{**}: H = Q(u,\theta) = 1 - (1-u^{1/(1+\theta)})^{1+\theta}, \quad \theta \geq 0 \; ,$$

where we have taken $G(u) = F^k(u) = u$. This alternative is the result of substituting $k = 1+\theta$, $k \geq 1$, in the previous alternative H_1 and requiring $G(u) = F^{\theta+1}(u) = u$. As θ increases, the difference between the distributions H and G is magnified. Since $Q(u,0) = u$, the null hypothesis is true if $\theta = 0$. Differentiating, we obtain

$$q(u,\theta) = [u^{-1/(1+\theta)} - 1]^{\theta} \; , \quad \text{and}$$

$$q'_\theta(u,0) = \log\,[(1/u) - 1] = \log\,[(1-u)/u] \quad .$$

The test which maximizes the slope of the power function at $\theta = 0$ then is to reject when

$$\sum_{j=1}^{m} \mathcal{E}_o\{\log\,[(1-U_{r_j})/U_{r_j}]\} > c \quad \text{for any } k \geq 1 \quad .$$

We have

$$\begin{aligned} \mathcal{E}_o\{\log\,[(1-U_i)/U_i]\} &= \mathcal{E}_o[\log\,(1-U_i)] - \mathcal{E}_o(\log U_i) \\ &= \mathcal{E}_o(\log U_{N-i+1}) - \mathcal{E}_o(\log U_i) \\ &= -\mathcal{E}_o(V_i) + \mathcal{E}_o(V_{N-i+1}) \quad , \qquad (4.20) \end{aligned}$$

where $V = -\log U$ and $f(v) = e^{-v}$, $v \geq 0$.

The density of the *i*th order statistic of a sample of size N from an exponential population is

$$f_i(v) = \frac{N!}{(i-1)!(N-i)!} (1 - e^{-v})^{i-1} (e^{-v})^{N-i+1} , \quad v \geq 0 ,$$

with the moment generating function

$$\Psi_v(t) = \mathcal{E}(e^{tv}) = \binom{N}{i} i \int_0^{\infty} (1-e^{-v})^{i-1} (e^{-v})^{N-i-t+1} dv$$

$$= \binom{N}{i} i \int_0^{1} (1-z)^{i-1} z^{N-i-t} dz$$

$$= \binom{N}{i} i \ \beta(i, N-i-t+1) = \frac{N!}{(N-i)!} \frac{(N-i-t)!}{(N-t)!} .$$

The cumulant generating function is

$$\log \Psi_v(t) = \log N! - \log (N-i)! - \sum_{j=0}^{i-1} \log (N-t-j) .$$

By taking the first derivative of $\log \Psi_v(t)$ and setting $t = 0$, we obtain the first cumulant (or moment) as

$$\mathcal{E}(V_i) = \sum_{j=0}^{i-1} 1/(N-j) = \sum_{j=N-i+1}^{N} j^{-1} . \tag{4.21}$$

Substituting (4.21) into (4.20) we have

$$\mathcal{E}\{\log [(1-U_i)/U_i]\} = \sum_{j=1}^{N} j^{-1} - \sum_{j=N-i+1}^{N} j^{-1} . \tag{4.22}$$

This value of a_i is very easy to compute but can also be found, as is well known, from tables of the Psi Distribution, where $\Psi(x) = d[\log \Gamma(x)]/dx$. Since

$\Gamma(x+1) = x\Gamma(x) = x(x-1)\Gamma(x-1)$, etc., we can write

$$\Gamma(x+r+1) = (x+r)(x+r-1) \ldots x\Gamma(x)$$

and $\Psi(x+r+1) = (x+r)^{-1} + (x+r-1)^{-1} + \ldots + (x)^{-1} + d[\log \Gamma(x)]/dx$,

or $\Psi(x+r+1) - \Psi(x) = (x)^{-1} + (x+1)^{-1} + \ldots + (x+r)^{-1}$.

Using this result, $\sum_{j=i}^{N} j^{-1} = \Psi(i + (N-i) + 1) - \Psi(i)$,

$\sum_{j=N-i+1}^{N} j^{-1} = \Psi(N - i + 1 + (i-1) + 1) - \Psi(N-i+1)$, and

(4.22) can be expressed as $\Psi(N-i+1) - \Psi(i)$. $\Psi(x)$ has been tabulated by Davis (1933, Tables VII - XII, pp. 291-367).

The locally most powerful rank test against H_1^{**} then is to reject the null hypothesis when

$$\Psi_N = \frac{1}{m} \sum_{i=1}^{N} [\Psi(N-i+1) - \Psi(i)] z_i > c \quad . \tag{4.23}$$

We will call this the Psi test. The critical regions and power functions against H_1 were given in Section 3.4.1, and were seen to be identical with those of the most powerful rank test when $m = n \leq 4$. The descending orders of the values of (4.20) turn out to be equivalent to the orders of probabilities under H_1 for the cases considered. When two different arrangements are equally likely under H_1, the corresponding values of (4.20) are the same. It is rather surprising that the two tests are equally powerful for these small sample sizes. The Psi test has the advantage of ease in computation, and, of course, the critical regions are independent of the value of k.

The exact mean and variance of Y_N are given by (4.7) and (4.8) as

$$\mathcal{E}_0(Y_N) = \frac{1}{N} \sum_{i=1}^{N} [(\sum_{j=i}^{N} - \sum_{j=N-i+1}^{N}) j^{-1}] = 0$$

and $$\sigma_0^2(Y_N) = \frac{n}{mN(N-1)} \sum_{i=1}^{N} [(\sum_{j=i}^{N} - \sum_{u=N-i+1}^{N}) j^{-1}]^2 \quad .$$

Although $\sigma_0^2(Y_N)$ can easily be evaluated for small N, a convenient closed form for general N is not readily apparent.

For the asymptotic properties, we must first verify that the conditions of Theorem 4.3 are satisfied. Here

$$J_N(i/N) = -[\sum_{j=i}^{N} - \sum_{j=N-i+1}^{N}] j^{-1} = -\mathcal{E}_0\{\log[(1-U_i)/U_i]\} \quad ,$$

where the U_i are order statistics from a uniform distribution. Letting $X = -\log[(1-U)/U]$, $J_N(i/N)$ can be considered the expected value of X_i, the ith order statistic of a sample of size N from the distribution $F_X(x) = e^x/(e^x+1)$, $-\infty < x < \infty$. Since $J(u) = -\log[(1-u)/u]$, we have

$u = e^J/(e^J+1)$, so that F_X is the inverse of J. By Theorem 4.4, Theorem 4.3 can be applied if condition (vii) is satisfied. For $i = 0$, we need to verify that

$$|J(u)| = |\log[(1-u)/u]| \leq K[u(1-u)]^{-\frac{1}{2}+\delta}$$

or $K \geq [u(1-u)]^{\frac{1}{2}-\delta} \; | \; [\log(1-u) - \log u]| = K(u)$ say for $0 < u < 1$. We can assume that $\delta < \frac{1}{2}$. As $u \to 0$, $K(u) \to 0$; as $u \to 1$, $K(u) \to 0$; and $K(\frac{1}{2}) = 0$, $K(u+\frac{1}{2}) = K(\frac{1}{2}-u)$. The maximum value of $K(u)$ for $0 < u \leq \frac{1}{2}$ will occur for that u which satisfies $\log[(1-u)/u] = 2/[(1-2\delta)(1-2u)]$. Depending on the value of δ, the critical point lies in the interval (0,1/8). Since $K(u)$ has a finite maximum, the condition holds for $i = 0$.

For $i = 1$, we must show that

$$\left| \frac{-1}{u(1-u)} \right| \leq K[u(1-u)]^{(-3/2)+\delta}$$

or $K \geq [u(1-u)]^{(3/2)-\delta} \; [u(1-u)]^{-1} = [u(1-u)]^{\frac{1}{2}-\delta}$.

The maximum value of $u(1-u)$ is 1/4, so that if $K \geq 2^{2\delta-1}$, the inequality holds.

For $i = 2$, $\left| \frac{-(2u-1)}{u^2(1-u)^2} \right| \leq K[u(1-u)]^{(-5/2)+\delta}$

or $K \geq (1-2u)[u(1-u)]^{\frac{1}{2}-\delta}$.

The maximum value of the right hand side is again $2^{2\delta-1}$.

From (4.15), the asymptotic mean of Y_N is zero under H_o, and the variance can be determined from (4.16). We must evaluate

$$\sigma_o^2(Y_N) = \frac{n}{mN} \int_0^1 \{\log[(1-u)/u]\}^2 \, du$$

$$= \frac{n}{mN} \int_0^1 [\log(1-u) - \log u]^2 \, du \quad .$$

Then

$$\frac{mN}{n} \sigma_o^2(Y_N) = \int_0^1 [\log(1-u)]^2 \, du - 2 \int_0^1 \log(1-u) \log u \, du + \int_0^1 (\log u)^2 \, du$$

$$= 2 \int_0^1 (\log u)^2 \, du + 2 \sum_{r=1}^{\infty} \int_0^1 \frac{u^r}{r} \log u \, du \quad .$$

$$\frac{mN}{2n} \sigma_o^2(Y_N) = 2 + \sum_{r=1}^{\infty} \frac{1}{r} \int_{\infty}^{0} e^{-ry} \, y e^{-y} \, dy$$

$$= 2 - \sum_{r=1}^{\infty} \frac{1}{r} \cdot \frac{1}{(r+1)^2} \Gamma(2)$$

$$= 2 - \sum_{r=1}^{\infty} [(\frac{1}{r} - \frac{1}{r+1}) - \frac{1}{(r+1)^2}]$$

$$= 2 - [1 - (\pi^2/6 - 1)] = \pi^2/6 \quad .$$

Then $\qquad \sigma_o^2(Y_N) = n\pi^2/3mN \quad . \qquad (4.24)$

From (4.11), under the null hypothesis, the random variable $Y_N\sqrt{3mN}/\pi\sqrt{n}$ is asymptotically normally distributed with mean zero and variance one.

Terry (1952, Section 9) and Savage (1956, Section 7) have studied the correlation under the null hypothesis between the Wilcoxon (or Mann-Whitney) statistic and the c_1 and D_n statistics, respectively. Since the Psi test would be used for the same type of alternative as the Wilcoxon test, their correlation under H_o will also be of interest. Let

$$W_N = \frac{1}{m} \sum_{i=1}^{N} i(1-z_i) \quad \text{and} \quad \Psi_N = \frac{1}{m} \sum_{i=1}^{N} a_i z_i \ ,$$

where $a_i = (\sum_{j=1}^{N} - \sum_{j=N-i+1}^{N}) \, j^{-1}$, so that both tests are based on large values of the respective test statistics. It will be more convenient to find the correlation $\rho(W_N', \Psi_N)$ between $W_N' = \sum_{i=1}^{N} iz_i/m$ and Ψ_N. Since $W_N = N(N+1)/2 - W_N'$, we have $\rho(W_N, \Psi_N) = -\rho(W_N', \Psi_N)$.

Under H_o, the exact variance of W_N' is $\sigma_o^2(W_N') = n(N+1)/12m$, and the asymptotic variance of Ψ_N is $\sigma_o^2(\Psi_N) = n\pi^2/3mN$ from (4.24). The covariance between Ψ_N and W_N' can be found from (4.9). We have

$$\mathrm{cov}_o(W_N', \Psi_N) = \frac{n}{mN^2(N-1)}[N \sum_{i=1}^{N} a_i i - \sum_{i=1}^{N} a_i \sum_{i=1}^{N} i] \ . \qquad (4.25)$$

But $\sum_{i=1}^{N} a_i = 0$, and

$$\sum_{i=1}^{N} a_i i = \sum_{i=1}^{N} i[(\sum_{j=i}^{N} - \sum_{j=N-i+1}) j^{-1}] ,$$

where

$$\sum_{i=1}^{N} i \sum_{j=i}^{N} j^{-1} = \sum_{j=1}^{N} j^{-1} + 2 \sum_{j=2}^{N} j^{-1} + \ldots + N \sum_{j=N}^{N} j^{-1}$$

$$= \frac{1}{1} + (\frac{1}{2} + \frac{2}{2}) + (\frac{1}{3} + \frac{2}{3} + \frac{3}{3}) + \ldots + (\frac{1}{N} + \frac{2}{N} + \cdot\cdot + \frac{N}{N})$$

$$= \frac{2}{2} + \frac{3}{2} + \frac{4}{2} + \ldots + \frac{(N+1)}{2}$$

$$= N(N+3)/4 ,$$

and

$$\sum_{i=1}^{N} i \sum_{j=N-i+1}^{N} j^{-1} = \sum_{j=N}^{N} j^{-1} + 2 \sum_{j=N-1}^{N} j^{-1} + \ldots + N \sum_{j=1}^{N} j^{-1}$$

$$= (\frac{1}{N} + \frac{2}{N} + \ldots + \frac{N}{N}) + (\frac{2}{N-1} + \frac{3}{N-1} + \ldots + \frac{N}{N-1}) + \ldots + (\frac{N}{1})$$

$$= \frac{1}{2}(N+1) + \frac{1}{2}(N+2) + \ldots + \frac{1}{2}(N+N)$$

$$= \frac{1}{2}[N^2 + \frac{N(N+1)}{2}]$$

$$= N(3N+1)/4 .$$

It then follows that $\sum_{i=1}^{N} a_i i = -N(N-1)/2$. Substituting these results into (4.25), we have

$$\text{cov}_0(W_N', Y_N) = -n/2m \quad . \tag{4.26}$$

Hence, the correlation coefficient is asymptotically

$$\rho(W_N, Y_N) = \lim_{N\to\infty} \frac{n/2m}{\sqrt{\frac{n(N+1)}{12m} \cdot \frac{n\pi^2}{3mN}}} = 3/\pi \quad , \tag{4.27}$$

or .9550. This result is larger than $\rho(D_N, W_N) = .8660$ and slightly smaller than $\rho(c_1, W_N) = .9772$.

Evaluation of the asymptotic moments under the alternative from (4.12) and (4.13) is difficult with the complicated functions

$$R_\theta(u) = \frac{m}{N}[1 - (1-u^{1/(1+\theta)})^{1+\theta}] + \frac{n}{N}u \quad ,$$

$$J(R_\theta) = \log[(1-R)/R], \; J'(R_\theta) = -[R(1-R)]^{-1} \quad ,$$

and $Q(u,\theta) = 1 - (1-u^{1/(1+\theta)})^{1+\theta}$. For the mean,

$$\mathcal{E}_a(Y_N) = \int_0^1 \log[\frac{N-m+m(1-u^{1/(1+\theta)})^{1+\theta} - nu}{m-m(1-u^{1/(1+\theta)})^{1+\theta} + nu}](u^{-1/(1 \ \theta)} - 1)^\theta \, du$$

$$= \frac{N}{m}\int_0^1 \log[(1-x)/x] \, dx - \frac{n}{m}\int_0^1 \log[(1-x)/x] \, du$$

where $Nx = nu + m - m(1-u^{1/(1+\theta)})^{1+\theta}$. The first integral is equal to zero. When $m = n$,

$$\log(1-x) - \log x = \log[\tfrac{1}{2} - \tfrac{1}{2}u + \tfrac{1}{2}(1-u^{1/(1+\theta)})^{1+\theta}]$$

$$- \log[\tfrac{1}{2} + \tfrac{1}{2}u - \tfrac{1}{2}(1-u^{1/(1+\theta)})^{1+\theta}]$$

$$= \log(1-y) - \log(1+y)$$

where $y = u - (1-u^{1/(1+\theta)})^{1+\theta}$. But

$$\log(1-y) - \log(1+y) = -2 \sum_{r=0}^{\infty} \frac{y^{2r+1}}{2r+1} .$$

Then

$$\mathcal{E}_a(Y_N) = 2 \sum_{r=0}^{\infty} \frac{1}{2r+1} \int_0^1 [u-(1-u^{1/(1+\theta)})^{1+\theta}]^{2r+1}\, du$$

$$= -2 \sum_{r=0}^{\infty} \frac{1}{2r+1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (-1)^j \int_0^1 u^j (1-u^{1/(1+\theta)})^{(1+\theta)(2r+1-j)}\, du$$

$$= -2 \sum_{r=0}^{\infty} \frac{1+\theta}{2r+1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (-1)^j \int_0^1 v^{j(1+\theta)} (1-v)^{(1+\theta)(2r+1-j)} v^{\theta}\, dv .$$

Thus

$$\mathcal{E}_a(Y_N) = -2 \sum_{r=0}^{\infty} \frac{1+\theta}{2r+1} \sum_{j=0}^{2r+1} \binom{2r+1}{j} (-1)^j \beta[j(1+\theta)+\theta+1),\ (1+\theta) \cdot (2r+1-j)+1]$$

when $m = n$. For the variance when $m = n$, we need to find

$$\sigma_a^2(Y_N) = \frac{16}{N} \iint_{0<u<v<1} [1+u-(1-u^{1/(1+\theta)})^{1+\theta}]^{-1}$$

$$\cdot [1-u+(1-u^{1/(1+\theta)})^{1+\theta}]^{-1}[1+v-(1-v^{1/(1+\theta)})^{1+\theta}]^{-1}$$

$$\cdot [1-v+(1-v^{1/(1+\theta)})^{1+\theta}]^{-1}$$

$$\cdot \{u(1-v)(u^{-1/(1+\theta)}-1)^{\theta}(v^{-1/(1+\theta)}-1)^{\theta}$$

$$+ [1-(1-u^{1/(1+\theta)})^{1+\theta}](1-v^{1/(1+\theta)})^{1+\theta}\} du\, dv$$

$$= \frac{16(1+\theta)^2}{N} \int_0^1 \int_y^1 (1-x)^{\theta}\, y^{\theta}\, [1+(1-x)^{1+\theta}-x^{1+\theta}]^{-1}$$

$$\cdot [1-(1-x)^{1+\theta}+x^{1+\theta}]^{-1}[1+(1-y)^{1+\theta}-y^{1+\theta}]^{-1}[1-(1-y)^{1+\theta}+y^{1+\theta}]^{-1}$$

$$\cdot \{x^{\theta}(1-x)[1-(1-y)^{1+\theta}] + (1-x^{1+\theta})y(1-y)^{\theta}\}\, dx\, dy$$

$$= \frac{16(1+\theta)^2}{N} \int_0^1 \int_y^1 (1-x)^{\theta}\, y^{\theta}\, \{1-[x^{1+\theta}-(1-x)^{1+\theta}]^2\}^{-1}$$

$$\cdot \{1-[y^{1+\theta}-(1-y)^{1+\theta}]^2\}^{-1}\, \{x^{\theta}(1-x)[1-(1-y)^{1+\theta}]$$

$$+ (1-x^{1+\theta})y(1-y)^{\theta}\}\, dx\, dy \quad .$$

The integral could be evaluated by numerical methods.

It should be noted that the Psi test statistic is similar to the D_N statistic proposed by Savage. The D_N test rejects for large values of $\sum_{i=1}^{N} z_i \sum_{j=i}^{N} j^{-1}$, and is the locally most powerful rank test against the alternative $H = F^{\Delta_2-\Delta_1}$, $G = F^{\Delta_2}$, $\Delta_2 > \Delta_1 > 0$. Here $Q(u,\theta) = u^{1-\theta}$ where $\theta = \Delta_1/\Delta_2 \le 1$.

4.3.3 Terry's c_1 Test

These alternatives H_1^* and H_1^{**} are nonparametric in form since the only requisite assumption is that F is a continuous cumulative distribution function. Terry's c_1 test discussed previously (Section 3.3) is designed primarily for a parametric alternative expressing a change of location in the normal distribution. A uniformly most powerful rank test does not exist for this type of alternative, but Terry's c_1 test is the locally most powerful rank test. Since H can be expressed as a function of G, the methods of this chapter can be used to prove this important property.

Theorem 4.5. The locally most powerful rank test of the null hypothesis H_o: H = G versus the alternative

$$H_1': \; H(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(t-\mu_X)^2} \, dt \quad ,$$

$$G(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(t-\mu_Y)^2} \, dt \quad , \quad 0 < \mu_Y - \mu_X = \theta < \delta \; ,$$

is Terry's c_1 test. The rule is to reject when

$$c_1 = \sum_{j=1}^{n} \mathcal{E}(\xi_{s_j}) > c \; ,$$ where c is a constant depending on α, $\xi_1 \leq \xi_2 \leq \ldots \leq \xi_N$ are order statistics from a sample of size N drawn from a population which is normally distributed

with mean zero and variance one, and $s_1, s_2, \ldots, s_n$ are the ranks of the Y random variables in the combined sample.

Proof. Since G(y) is not the uniform distribution, make the probability integral transformation G(y) = u so that the random variable U is uniformly distributed on the interval (0,1). Then $y = G^{-1}(u)$, where the random variable Y is normally distributed with mean μ_Y and variance σ^2. We have

$$H(y) = \int_{-\infty}^{y+\mu_Y-\mu_X} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(t-\mu_Y)^2} dt = G(y + \mu_Y - \mu_X) \quad ,$$

and $Q(u,\theta) = G(G^{-1}(u) + \theta)$. Thus

$$q(u,\theta) = \partial Q(u,\theta)/\partial u = g(G^{-1}(u)+\theta) \cdot \partial G^{-1}(u)/\partial u$$

$$= g(G^{-1}(u)+\theta) \cdot \frac{1}{\partial u/\partial G^{-1}(u)} = \frac{g(G^{-1}(u)+\theta)}{g(G^{-1}(u))}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(G^{-1}(u)+\theta-\mu_Y)^2}}{e^{-\frac{1}{2\sigma^2}(G^{-1}(u)-\mu_Y)^2}} \quad ,$$

and $\partial q(u,\theta)/\partial\theta\big|_{\theta=0} = -\frac{1}{\sigma^2}(G^{-1}(u) - \mu_Y)$.

By (4.5), the locally most powerful rank test rejects when

$$-\frac{1}{\sigma^2}\sum_{j=1}^{m} \mathcal{E}(G^{-1}(U_{r_j}) - \mu_Y) > c \quad . \qquad (4.28)$$

Since $U_1 \leq U_2 \leq \ldots \leq U_N$ are order statistics from a uniform population, $Y_1 \leq Y_2 \leq \ldots \leq Y_N$, where $Y = G^{-1}(U_i)$, are order statistics from a normal population with mean μ_Y and variance σ^2. Hence, $\xi_1 \leq \xi_2 \leq \ldots \leq \xi_N$, where $\xi_i = (Y_i - \mu_Y)/\sigma$, are order statistics from a normal distribution with mean zero and variance one. Then (4.28) becomes $-\sum_{j=1}^{m} \mathcal{E}(\xi_{r_j}) > c$. Since $\sum_{j=1}^{m} \mathcal{E}(\xi_{r_j}) + \sum_{j=1}^{n} \mathcal{E}(\xi_{s_j}) = 0$, the test statistic can be written $c_1 = \sum_{j=1}^{n} \mathcal{E}(\xi_{s_j})$ and the theorem is proved.

Using the notation of (4.6), Terry's c_1 tests rejects when

$$\sum_{i=1}^{N} (1-z_i)\mathcal{E}(\xi_i) > c \quad ,$$

where c can be found in Terry's Table 1 (pp. 358–361) or by using tables of the expected values of the order statistics from a normal distribution (Harter, 1961).

The exact mean and variance of the c_1 statistic under the null hypothesis are zero and $\frac{mn}{N(N-1)} \sum_{i=1}^{N} [\mathcal{E}(\xi_i)]^2$ from Theorem 4.2 because $\sum_{i=1}^{N} \mathcal{E}(\xi_i) = 0$. Since the conditions of Theorem 4.3 are satisfied (Chernoff and Savage, 1958), the asymptotic distribution of c_1 is normal under any alternative as well as under the null hypothesis. Also, the c_1 test is at least as efficient as the Student's t test (Chernoff and Savage, 1958), which is the uniformly most powerful test for the alternative of a change of location in the normal distribution.

Chapter V

LARGE-SAMPLE POWER

Although comparisons of power functions for small samples may be considered the most realistic appraisal of the performance of tests, the computations are usually tedious. Probably for this reason, many gaps are present in the information available on exact power computations. However, considerable progress has been made with respect to asymptotic power results. In many cases, the large sample power of nonparametric tests is easily computed because the test statistic is asymptotically normally distributed under the alternative. For example, any linear rank statistic (4.6) is asymptotically normal under both the hypothesis and the alternative, subject to certain regularity conditions.

Although this paper is primarily concerned with small-sample power, some asymptotic power results are desirable for comparative purposes. Since the first two moments of the one-sided Mann-Whitney or Wilcoxon test and the Wald-Wolfowitz runs test are easily computed and the asymptotic distributions of these test statistics have been determined, the approximate power functions of these two tests will be presented in this chapter for the previously considered alternative H_1 of two extreme distributions, when $k = 2$. In addition to the factor of ease in computation, these two tests were chosen since they are perhaps the most widely used of the two-sample rank tests for the purpose of testing whether two samples come from the same population.

5.1 Mann-Whitney or Wilcoxon Test

The Mann-Whitney or Wilcoxon test T_2 has been shown to be consistent against any alternative of the form $H(a) \geq G(a)$ for all a (Mann and Whitney, 1947; Lehmann, 1951) which certainly applies to the alternative $H = 1 - (1-F)^k$, $G = F^k$ for any value of $k > 1$. The test statistic is the sum of the ranks of the Y random variables (Wilcoxon test), or, equivalently, U, the number of pairs (X_i, Y_j) for which $X_i < Y_j$ (Mann-Whitney test). Thus we can define the statistic $U = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}$ where $X_{ij} = 1$ if $X_i < Y_j$ and $X_{ij} = 0$ otherwise. The test statistics are equivalent since

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} = (s_1 - 1) + (s_2 - 2) + \ldots + (s_n - n) = \sum_{j=1}^{n} (s_j - j)$$

$$= \sum_{j=1}^{n} s_j - n(n+1)/2, \quad \text{where } s_1 \leq s_2 \leq \ldots \leq s_n \text{ are the}$$

ranks of the Y random variables in the combined sample.

Mann and Whitney (1947) have shown that the distribution of the random variable $[U - \mathcal{E}(U)]/\sigma(U)$ tends to normality with zero mean and unit variance under the null hypothesis $H = G$ as $N \to \infty$, m/n constant. Lehmann (1951) has shown that the approximation is still good when the first two moments under the alternative are substituted.

To compute the exact first two moments, we can use

$$\mathcal{E}(U) = mn \Pr(X_{ij} = 1) = mn \Pr(X_i < Y_j) \quad , \qquad (5.1)$$

and $$\sigma^2(U) = \mathcal{E}(U^2) - [\mathcal{E}(U)]^2 \quad , \quad \text{where} \qquad (5.2)$$

$$\mathcal{E}(U^2) = mn\, \mathcal{E}(X_{ij}^2) + mn(n-1) \underset{j\neq k}{\mathcal{E}} (X_{ij}X_{ik}) + m(m-1)n \underset{i\neq h}{\mathcal{E}} (X_{ij}X_{hj})$$
$$+ m(m-1)n(n-1) \underset{i\neq h,\, j\neq k}{\mathcal{E}} (X_{ij}X_{hk}) \quad . \qquad (5.3)$$

Under the null hypothesis $H = G$, $\Pr(X_i < Y_j) = ½$, so that $\mathcal{E}_0(U) = mn/2$, $\mathcal{E}_0(X_{ij}^2) = ½$, $\underset{j\neq k}{\mathcal{E}_0}(X_{ij}X_{ik}) = \underset{i\neq h}{\mathcal{E}_0}(X_{ij}X_{hj}) = 1/3$, $\underset{i\neq h,\, j\neq k}{\mathcal{E}_0}(X_{ij}X_{hk}) = 1/4$. Then $\sigma_0^2(U) = mn(m+n+1)/12$ by (5.2) and (5.3). Thus, for example, the α level test rejects when

$$\frac{U/mn - ½}{[(m+n+1)/(12mn)]^{½}} > z_\alpha \quad , \qquad (5.4)$$

where z_α is defined by $\Phi(z_\alpha) = \int_{-\infty}^{z_\alpha} \frac{1}{\sqrt{2\pi}} e^{-½t^2} dt = 1 - \alpha$.

Under the alternative H_1: $H = 1 - (1-F)^k$, $G = F^k$, we have $\Pr(X_i < Y_j) = 1 - k\,\beta(k,k+1)$ from (3.7). Then

$$\mathcal{E}_A(U/mn) = 1 - k\ \beta(k,k+1) \tag{5.5}$$

from (5.1). Also from (3.7), it may be easily verified that

$$\underset{j\neq k}{\mathcal{E}_A}(X_{ij}X_{ik}) = \underset{i\neq h}{\mathcal{E}_A}(X_{ij}X_{hj}) = 1 - 2k\ \beta(k,k+1) + k\ \beta(k,2k+1) \quad , \tag{5.6}$$

and $$\underset{i\neq h,\, j\neq k}{\mathcal{E}}(X_{ij}X_{hk}) = 1 - 2k\ \beta(k,k+1) + k^2[\beta(k,k+1)]^2 . \tag{5.7}$$

Upon substituting (5.5) - (5.7) into (5.3) and (5.2), we have

$$\sigma_A^2(U/mn) = \frac{k}{mn}\{\beta(k,k+1) + (m+n-2)\ \beta(k,2k+1) - (m+n-1)k[\beta(k,k+1)]^2\}. \tag{5.8}$$

The power for any value of k, m, and n is

$$\Pr\{\frac{U/mn - \mathcal{E}_A(U/mn)}{\sigma_A(U/mn)} \geq c\} = 1 - \Phi(\frac{z_\alpha[12mn(m+n+1)]^{-\frac{1}{2}} + \frac{1}{2} - \mathcal{E}_A(U/mn)}{\sigma_A(U/mn)}) \tag{5.9}$$

where c is determined by (5.4).

The power curve $\beta(T_2)$ for m = n, k = 2, α = .05, as computed from (5.9) is shown in Figure 5.1. The first two moments under the alternative when k = 2 are 5/6 and (7N+11)/(180mn). For small samples, the approximate power function yields .19 for m = n = 2, .31 for m = n = 3, and .43 for m = n = 4. Thus, at least for these three cases, the power is underestimated when the asymptotic theory is used.

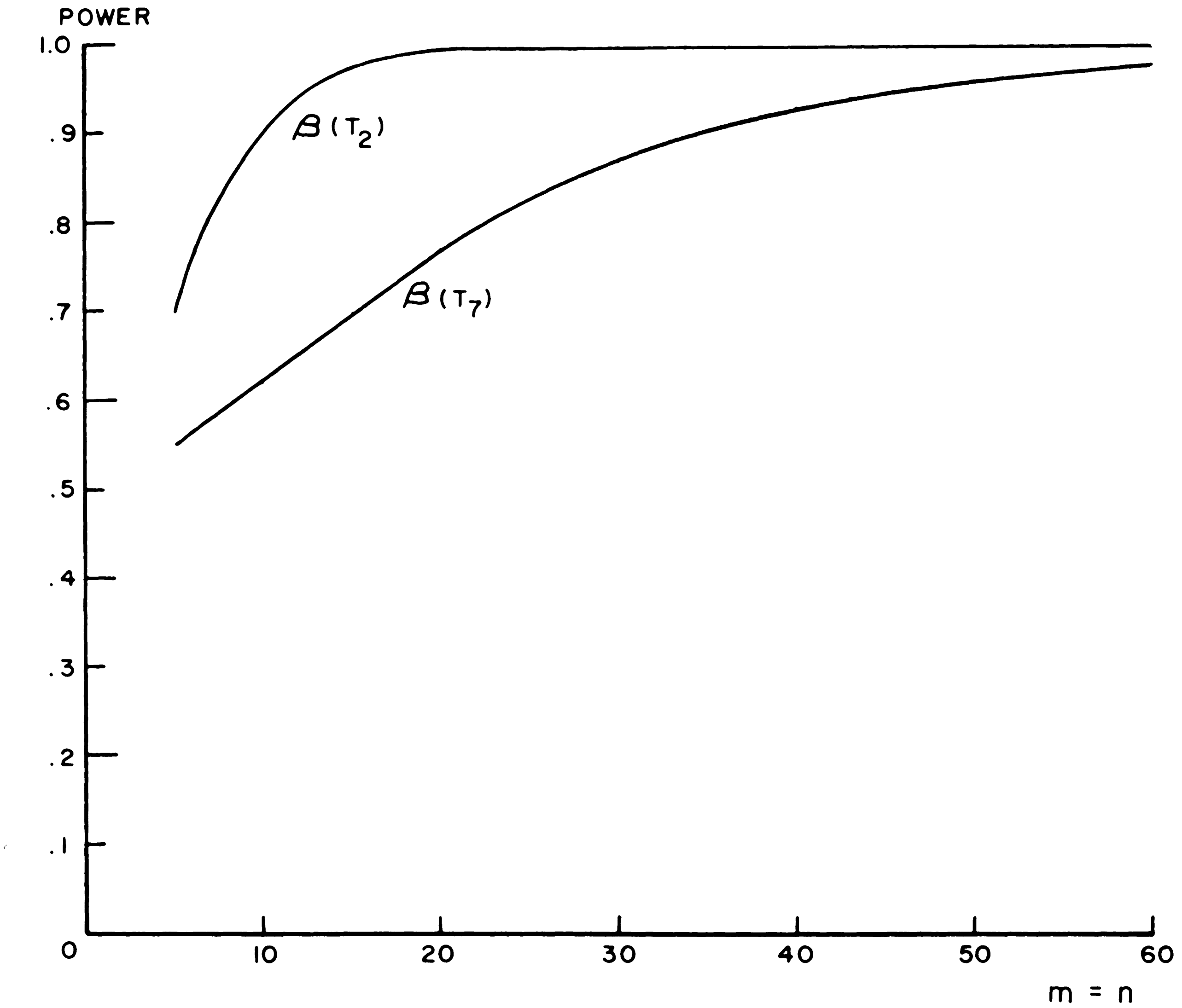

Figure 5.1. Approximate power functions of Mann-Whitney and Wald-Wolfowitz runs tests against H_1: $H=1-(1-F)^2$, $G=F^2$ for equal sample sizes, $\alpha=.05$

5.2 Wald-Wolfowitz Runs Test

The Wald-Wolfowitz runs test T_7 (Wald and Wolfowitz, 1940) is also consistent against any alternative. The random variable $\underline{r}$, the number of runs of both X's and Y's in the combined sample, is asymptotically normal under the null hypothesis H = G as well as under the alternative as $N \to \infty$, $m/n = c$, a constant. The first two moments of $\underline{r}$ under any alternative are given by the following theorem due to Wolfowitz (1949).

Theorem 5.1. Let R(x) and Q(x) be the cumulative distribution functions from which m and n observations respectively are obtained.

Let

(a) $$R(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

(b) $$Q(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$$

(c) the derivative q(x) of Q(x) exist and be continuous and positive in the interval $0 \leq x \leq 1$.

(d) $m/n = c$, a constant.

Then $$\lim_{m\to\infty} \mathcal{E}_A(r/m) = 2\int_0^1 \frac{q(x)}{c+q(x)}\,dx \quad , \tag{5.10}$$

$$\lim_{m\to\infty} \sigma_A^2(r/\sqrt{m}) = 4\{\int_0^1 \frac{cq^2}{(c+q)^3}\,dx + \int_0^1 \frac{q(c^3+q^3)}{(c+q)^4}\,dx - [\int_0^1 \frac{q^2}{(c+q)^2}\,dx]^2 - c^3[\int_0^1 \frac{q}{(c+q)^2}\,dx]^2\} \quad . \tag{5.11}$$

For the alternative H_1 with $k = 2$, we must have $x = 2F(x) - F^2(x)$ so that $R(x) = x$; $Q(x) = F^2(x)$. But $1 - x = 1 - 2F(x) + F^2(x)$, or $F(x) = 1 - \sqrt{1-x}$. Then $Q(x) = (1 - \sqrt{1-x}\,)^2$, and $q(x) = [1 - \sqrt{1-x}\,]/\sqrt{1-x}$. Thus $\lim_{m\to\infty} \mathcal{E}_A(r/m) = 2/3$, $\lim_{m\to\infty} \sigma_A^2(r/\sqrt{m}\,) = 26/45$, and

$$\lim_{m\to\infty} \Pr[\frac{U/m - 2/3}{(26/45m)^{\frac{1}{2}}} \leq x] = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}dt \quad . \tag{5.12}$$

Under the null hypothesis $H = G$, the exact values of the first two moments are $\mathcal{E}_0(r) = \frac{2mn}{m+n} + 1$, $\sigma_0^2(r) = \frac{2mn(2mn - m - n)}{(m+n)^2\,(m+n-1)}$ (Mood, 1950, p. 393), and the test for $m = n$ at significance level $\alpha = .05$, for example, rejects when

$$\frac{U/m - [1+(1/m)]}{[(m-1)/m(2m-1)]^{\frac{1}{2}}} \leq -1.645 \quad . \tag{5.13}$$

Thus for $k = 2$, we can easily compute the approximate power for any α and any $m = n$ using (5.13) and (5.12). The results for $m = n$, $k = 2$, $\alpha = .05$, are given in Figure 5.1, labeled $\beta(T_7)$. For $m = n \leq 4$, the asymptotic power function does not give reasonable results.

The Mann-Whitney test is seen to be considerably more powerful than the Wald-Wolfowitz runs test, although both have high power functions for this type of alternative.

Chapter VI

SUMMARY AND CONCLUSIONS

Although nonparametric statistical techniques have generally achieved widespread acceptance, more information is needed concerning their performance in various situations. The amount of research activity in nonparametric statistical inference has grown rapidly each year, but no major breakthroughs have yet been achieved in determining small-sample power functions. Most of the results available are highly peculiar to the conditions assumed -- usually normal alternatives, isolated sample sizes, and "convenient" significance levels. In this paper, comparisons are made under as similar and general conditions as possible and for the more usual significance levels by employing randomized tests.

The problem of dealing with different conditions is particularly evident with the sign test and its parametric analogue, the $\underline{t}$ test, for density functions whose mean and median do not coincide. This is dealt with here (Chapter II) by considering three different sets of hypotheses of location for the sign test. The power functions of the one-sided, one-sample sign test are computed for $n = 10$, $\alpha = .05$, against a wide range of approximately normal alternatives. The significance level is found to be not greatly affected by noncoincidence of the mean and median. The difference between the power functions of the t test and the ordinary sign test under similar conditions is slight for highly leptokurtic distributions. When the sign test is performed on the median as approximated by the mean and coefficient of skewness, the difference is almost negligible for positively skewed, highly leptokurtic distributions.

The remainder of the thesis (Chapters III, IV, and V) is concerned with a different class of nonparametric techniques -- two-sample tests based on the ranks of the variables in the combined sample. Many rank tests have been proposed to test the null hypothesis that the two samples of sizes m and n are drawn from identical populations, but their power functions have been compared almost exclusively against parametric alternatives.

First, the nonparametric alternative of two extreme distributions has been considered, H_1: $H = 1 - (1-F)^k$, $G = F^k$, with F being unspecified. If F has a symmetrical density function, H and G are mutually symmetric distributions. More descriptively, their densities are mirror images. However, since F need not be specified, we can assume without loss of generality that F is symmetric. Formulae are presented for the calculations of the probabilities under this alternative of any of the $\binom{N}{m}$ arrangements of the $N = m + n$ random variables and used to calculate the small-sample power functions of nine two-sample rank tests. The cases included are all combinations of $k = 2$, 3, and 4, $m = n \leq 4$ for $\alpha = .01$, .05 and .10, and unequal sample sizes such that $m + n \leq 8$ for $\alpha \leq 1/\binom{N}{m}$ for one-sided tests and $\alpha \leq 2/\binom{N}{m}$ for two-sided tests. Two of the test statistics, the Gamma test and Psi test, are new. The six one-sided tests considered, the most powerful rank test, Mann-Whitney or Wilcoxon, Terry's c_1 test, the Psi test, Gamma test, and median test, are the most appropriate for this alternative. The first four tests have the same power for all cases considered. The one-sided median test has by far the least power. Of the three two-sided tests, the median, runs, and Wilcoxon tests, only the latter has high power.

These results are compared with power functions of the same tests against normal alternatives such that the standardized difference corresponds to the difference between the two extreme distributions. The power against normal alternatives is found to be slightly lower in most cases.

No clear-cut and final conclusions can be drawn from the limited power functions calculated in this paper. A need for more computations for larger sample sizes is clearly indicated. Although extensive numerical work is involved, a complete picture of the performance of two-sample rank tests against these alternatives would seem worthwhile.

The locally most powerful rank test against general functional alternatives has been derived and its properties studied. The technique is applied to two specific alternatives, H_1^*: $H = (1-\theta)F^k + \theta[1 - (1-F)^k]$, $G = F^k$, and H_1^{**}: $H = 1 - (1-F)^{\theta+1}$, $G = F^{\theta+1}$, where $\theta \geq 0$. Both are similar in spirit to the alternative of two extreme distributions. The resulting test statistics are called the Gamma test and the Psi test, respectively. The latter test rejects the null hypothesis when $\sum_{j=1}^{m} [\Psi(N-r_j+1) - \Psi(r_j)] > c$, where the r_j are the ranks of the X random variables and $\Psi(x) = d\,[\log \Gamma(x)]dx$. The power of the Psi test turns out to be the same as that of the most powerful rank test for all cases considered. The critical regions are independent of any parameters, and the test is very simply performed. Under the null hypothesis, the correlationbetween the Psi and Wilcoxon test statistics is asymptotically $3/\pi$, or .9550.

Although the Gamma test is mainly of theoretical interest, the Psi test seems to merit further investigation.

Determination of its power function for larger sample sizes would be desirable, as well as studies of its performance against other alternatives. Its asymptotic properties should be more completely examined. Further, an attempt should be made to determine how soon asymptotic properties provide a reasonable approximation to the test's behavior in moderate-sized samples.

In the final chapter, approximate power functions of the Wilcoxon and runs tests have been examined for equal sample sizes against the alternative of two extreme distributions when $k = 2$. The Wilcoxon test is seen to be considerably more powerful against this alternative, and its asymptotic power function provides a good approximation to the exact power for smaller samples. A similar study for the other tests would also be of interest.

ACKNOWLEDGMENTS

The author wishes to express her appreciation for the helpful suggestions, guidance and encouragement given by Dr. Herbert A. David in the preparation and writing of this thesis.

This research was supported in part by a grant from the Office of Ordnance Research. The author is also grateful to the Southern Fellowships Fund, National Science Foundation, and National Institutes of Health, for their financial assistance during various phases of her graduate studies.

BIBLIOGRAPHY

Barton, D. E. (1957). A comparison of two sorts of test for a change of location and applicable to truncated data. J. R. Statist. Soc. (B) 19, 119-124.

Barton, D. E. and Dennis, K. E. (1952). Conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 39, 425-427.

Capon, Jack (1961). Asymptotic efficiency of certain locally most powerful rank tests. Ann. Math. Statist. 32, 88-100.

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. 29, 972-994.

Cornish, E. A. and Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. Revue de l'Institut International de Statistique 5, 307-320.

Cramér, H. (1928). On the composition of elementary errors. Skand. Aktuartidskr. 11, 141-180.

Cramér, H. (1937). Random Variables and Probability Distributions. Cambridge Tracts in Mathematics, No. 36. Cambridge University Press.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

Davis, Harold T. (1933). Tables of the Higher Mathematical Functions, Vol. I. Principia Press, Bloomington (Ind.).

Dixon, W. J. (1953). Power functions of the sign test and power efficiency for normal alternatives. Ann. Math. Statist. 24, 467-473.

Dixon, W. J. (1954). Power under normality of several nonparametric tests. Ann. Math. Statist. 25, 610-614.

Edgeworth, F. Y. (1905). The law of error. Trans. Camb. Phil. Soc. 20, 36-65, 113-141.

Epstein, Benjamin (1955). Comparison of some non-parametric tests against normal alternatives with an application to life testing. _J. Amer. Statist. Ass._ 50, 894-900.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. _Proc. Camb. Phil. Soc._ 24, 180-190.

Fisher, R. A. and Yates, F. (1953). _Statistical Tables for Biological, Agricultural and Medical Research._ Oliver and Boyd, Edinburgh.

Gumbel, E. J. (1958). _Statistics of Extremes._ Columbia University Press, N.Y.

Haldane, J. B. S. (1942). The mode and median of a nearly normal distribution with given cumulants. _Biometrika_ 32, 294-299.

Harter, H. Leon (1961). Expected values of normal order statistics. _Biometrika_ 48, 151-165.

Haynam, G. E., Leone, F. C. and Chakravarti, I. M. (1961). Exact power of some tests based on the Mann-Whitney-Wilcoxon U statistic. Unpublished report.

Hemelrijk, J. (1961). Experimental comparison of Student's and Wilcoxon's two sample tests. _Quantitative Methods in Pharmacology,_ North-Holland Publishing Co., Amsterdam, 118-134.

Hoeffding, W. (1951). "Optimum" nonparametric tests. _Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,_ University of California Press, Berkeley, 83-92.

Klotz, Jerome (1962). Nonparametric tests for scale. _Ann. Math. Statist._ 33, 498-508.

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. _Ann. Math. Statist._ 22, 165-179.

Lehmann, E. L. (1953). The power of rank tests. _Ann. Math. Statist._ 24, 23-43.

Leone, F. C., Chakravarti, I. M. and Alanen, J. D. (1961). Exact power of some quick tests based on Mood's and Massey's statistics. Bull. de l'Inst. Intern. de Stat., 33e Session, Paris, No. 115.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Statist. 18, 50-60.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. 18, 309-348.

Mood, A. M. (1950). Introduction to the Theory of Statistics. McGraw-Hill Book Co., New York.

National Bureau of Standards (1953). Tables of Normal Probability Functions. Applied Mathematics Series No. 23.

Neyman, J., Iwaszkiewicz, K. and Kolodziejczyk, St. (1935). Statistical problems in agricultural experimentation. J. R. Statist. Soc. Supp. 2, 107-154.

Ruben, H. (1954). On the moments of order statistics in samples from normal populations. Biometrika 41, 200-227.

Savage, I. Richard (1956). Contributions to the theory of rank order statistics -- the two-sample case. Ann. Math. Statist. 27, 590-615.

Savage, I. Richard (1960). Contributions to the theory of rank order statistics: Computation rules for probabilities of rank orders. Ann. Math. Statist. 31, 519-520.

Srivastava, A. B. L. (1958). Effect of nonnormality on the power function of the t test. Biometrika 45, 421-429.

Sundrum, R. M. (1954). A further approximation to the distribution of Wilcoxon's statistic in the general case. J. R. Statist. Soc. (B) 16, 255-260.

Teichroew, Daniel (1954). A table giving the probability associated with order statistics in samples from two normal populations which have the same variance but different means. Technical report. U.S. Dept. of Commerce, National Bureau of Standards.

Teichroew, Daniel (1955). Empirical power functions for non-parametric two-sample tests for small samples. Ann. Math. Statist. 26, 340-344.

Terry, Milton E. (1952). Some rank order tests which are most powerful against specific parametric alternatives. Ann. Math. Statist. 23, 346-366.

Tsao, Chia Kuei (1957). Approximations to the power of rank tests. Ann. Math. Statist. 28, 159-172.

Uzawa, H. (1960). Locally most powerful rank tests for two-sample problems. Ann. Math. Statist. 31, 685-702.

van der Waerden, B. L. (1952). Order tests for the two-sample problem and their power. Proc. Kon. Ned. Akad. Wetensch. (A) 55, 453-458. Errata: Ibid. (1953) 80.

van der Waerden, B. L. (1953). Order tests for the two-sample problem (sedond and third communications). Proc. Kon. Ned. Akad. Wetensch. (A) 56, 303-316.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. Ann. Math. Statist. 11, 147-162.

Walsh, John E. (1946). On the power function of the sign test for slippage of means. Ann. Math. Statist. 17, 358-362.

Wilcoxon, Frank (1945). Individual comparisons by ranking methods. Biometrics 1, 80-83.

Wolfowitz, J. (1949). Nonparametric statistical inference. Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, 93-113.

The two page vita has been removed from the scanned document.

Page 1 of 2

The two page vita has been removed from the scanned document.

Page 2 of 2

ABSTRACT

I. Small-Sample Power of the One-Sample Sign Test for Approximately Normal Distributions. The power function of the one-sided, one-sample sign test is studied for populations which deviate from exact normality, either by skewness, kurtosis, or both. The terms of the Edgeworth asymptotic expansion of order more than $N^{-3/2}$ are used to represent the population density. Three sets of hypotheses and alternatives, concerning the location of (1) the median, (2) the median as approximated by the mean and coefficient of skewness, and (3) the mean, are considered in an attempt to make valid comparisons between the power of the sign test and Student's t test under the same conditions. Numerical results are given for samples of size 10, significance level .05, and for several combinations of the coefficients of skewness and kurtosis.

II. Power of Two-Sample Rank Tests on the Equality of Two Distribution Functions. A comparative study is made of the power of two-sample rank tests of the hypothesis that both samples are drawn from the same population. The general alternative is that the variables from one population are stochastically larger than the variables from the other.

One of the alternatives considered is that the variables in the first sample are distributed as the smallest of k variates with distribution F, and the variables in the second sample are distributed as the largest of these k -- H_1: $H = 1 - (1-F)^k$, $G = F^k$. These two alternative distributions are mutually symmetric if F is symmetrical. Formulae are presented, which are independent of F, for the evaluation of the probability under H_1 of any joint arrangement of the variables from the two samples. A theorem is proved concerning the equality of the probabilities of certain pairs of orderings under assumptions of mutually symmetric populations. The other alternative is that both samples are normally distributed with the same variance but different means, the standardized difference between the two extreme distributions in the first alternative corresponding to the difference between the means. Numerical results of power are tabulated for small sample sizes, k = 2, 3 and 4, significance levels .01, .05 and .10. The rank tests considered are the most powerful rank test, the one and two-sided Wilcoxon tests, Terry's c_1 test, the one and two-sided median tests, the Wald-Wolfowitz runs test, and two new tests called the Psi test and the Gamma test.

The two-sample rank test which is locally most powerful

against any alternative expressing an arbitrary functional relationship between the two population distribution functions and an unspecified parameter θ is derived and its asymptotic properties studied. The method is applied to two specific functional alternatives, H_1^*: $H = (1-\theta)F^k + \theta[1 - (1-F)^k]$, $G = F^k$, and H_1^{**}: $H = 1 - (1-F)^{1+\theta}$, $G = F^{1+\theta}$, where $\theta \geq 0$, which are similar to the alternative of two extreme distributions. The resulting test statistics are the Gamma test and the Psi test, respectively. The latter test is shown to have desirable small-sample properties.

The asymptotic power functions of the Wilcoxon and Wald-Wolfowitz tests are compared for the alternative of two extreme distributions with $k = 2$, equal sample sizes and significance level .05.