3282

NO AD NR

AFOSR - 3164

AF-AFOSR-62-7.

EXACT POWER OF SOME TWO-SAMPLE AND C-SAMPLE NON-PARAMETRIC STATISTICAL PROCEDURES

A Thesis Submitted to

Case Institute of Technology

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

George E. Haynam

1962

CASE INST

ABSTRACT

Expressions for the exact power of the two-sample Mann-Whitney Wilcoxon U test procedure against alternatives of exponential and rectangular populations have been derived. Several examples for total sample sizes of 11 and 15 have been compared with Mood's median test. Mood's test is more powerful than the U test in all instances in which the number of observations from the null population exceeds the number from the alternative population. The converse is true when the number of observations from the null population is less than the number from the alternative.

Expressions for the asymptotic efficiency of the Mann-Whitney-Wilcoxon U test relative to Mood's and Massey's tests and the likelihood ratio test have been derived for exponential populations. The asymptotic efficiency of the U test relative to the likelihood ratio test is zero.

Mood's and Massey's test procedures for two samples have been extended to the case of discriminating among c populations on the basis of c ordered samples. Expressions for the exact power have been derived for Mood's test with exponential and rectangular populations and for Massey's test with exponential populations. With exponential translation alternatives,the tests are biased.

The exact null distributions of goodness of fit tests for one-way and two-way contingency tables indicate that even for samples as small as ten, the exact distribution is closely approximated by a chi-square distribution with the appropriate degrees of freedom.

ACKNOWLEDGEMENTS

I am deeply grateful to Professor Fred C. Leone, my advisor, along with Professors Indra A. Chakravarti, and Z. Govindarajulu who have contributed a great deal to the execution of this thesis by providing both guidance and encouragement. In addition, I thank Professor S. N. Roy of the University of North Carolina for suggesting some of the topics included herein.

Also, some of the results presented in Chapter V could not have been obtained without the cooperation of Mr. Ken Powell of the International Business Machines Corporation and Mr. David Breedon of the Westinghouse Electric Corporation, Pittsburgh, who provided ample time on a 7090 computer. The remainder of the computations were performed on the Burroughs 220 computer at Case.

Appreciation is also due Mrs. Marie Fuller, who carefully typed the thesis.

This research was partially supported by the Air Force Office of Scientific Research.

SUMMARY

Many rank tests are available to discriminate between two populations on the basis of two ordered samples from the populations. Of them, Mood's test procedure [16] based on the median of the combined samples, Massey's extension of Mood's test [15] based on fractiles, and the Mann-Whitney-Wilcoxon U test procedure [14] based on the number of times an observation from the second sample exceeds an observation from the first sample, have much to commend them as quick tests.

The exact powers of Mood's and Massey's tests against alternatives of translation in normal and exponential populations and change in location and scale in a rectangular population have already been investigated by Barton [2] and Chakravarti, Leone, and Alanen [13]. Also, the exact power of the U test against the alternative of translation in the normal population has been computed by Dixon [6].

In Chapter I, expressions for the exact power of the two-sample Mann-Whitney-Wilcoxon U test procedure against alternatives of exponential and rectangular populations have been obtained. Several examples of the power for total sample sizes of 11 and 15 have been compared with similar results obtained from Mood's median test procedure. The results of the comparison indicate that for these two alternatives:

i) If the number of observations from the null population is less than the number from the alternative, the Mann-Whitney-Wilcoxon U test is more powerful than Mood's median test.

ii) If the number of observations from the null population is greater than the number from the alternative, then Mood's test is more powerful than the Mann-Whitney-Wilcoxon test.

iii) If the number of observations from both populations are the same, then both test procedures give approximately the same power.

In Chapter II, expressions for the asymptotic efficiency of the Mann-Whitney-Wilcoxon U test relative to Mood's and Massey's tests and the likelihood ratio test have been derived for exponential populations. The asymptotic efficiency of the Mann-Whitney-Wilcoxon test relative to the likelihood ratio test is zero, but in the case of Mood's and Massey's tests the resulting expressions are non-zero.

Chapter III is devoted to extending Mood's two-sample test procedure to the case of distinguishing among $c (c > 2)$ populations on the basis of c ordered samples from the populations. The appropriate expressions for the power functions for exponential and rectangular alternatives have been derived, and typical results for the case of three samples from exponential populations indicate that the test can be biased, especially when the level of significance is small.

Similarly, in Chapter IV, Massey's two-sample test procedure is extended to the case of distinguishing among $c(c > 2)$ populations on the basis of c ordered samples from the populations. Expressions for the exact power have been derived for the exponential translation alternatives, and again, typical results for the case of three samples indicate that the test can be biased, especially, when the level of significance is small.

In Chapter V, the exact null distribution of goodness of fit tests for one-way and two-way classifications is considered. Typical results are computed and are compared with the usual chi-square approximation. In general, the chi-square distribution with the appropriate degrees of freedom closely approximates the exact distribution, even for total sample sizes as small as ten. In addition, the exact power of the test statistic arising from a one-way classification has been computed for several alternatives, and the results have been compared with both non-central and central chi-square approximations. The results of the comparison indicate that both approximations tend to overestimate the power for small sample sizes, however, both approximations differ at most by one percent.

TABLE OF CONTENTS

CHAPTER I

EXACT POWER OF SOME TESTS BASED ON THE MANN-WHITNEY U STATISTIC

1.1 Introduction.

Many rank tests are available to discriminate between two populations on the basis of two ordered samples from the populations. Of them, Mood's test [16], based on the median of the combined samples, Massey's extension of Mood's test [15], based on fractiles, and Mann-Whitney's U test [14], based on the number of times an observation from the second sample exceeds an observation from the first sample, have much to commend them as quick tests.

The exact powers of Mood's and Massey's tests against alternatives of translation in the normal and exponential distributions and change in location and scale in the rectangular distribution have already been computed by Barton [2] and Chakravarti, Leone and Alanen [13]. Also the exact power of the Mann-Whitney U test against the alternative of translation in the normal distribution has been computed by Dixon [6].

The purposes of the investigation in this chapter are:

(i) To derive the exact power functions for the Mann-Whitney U test of two samples against alternatives of exponential and rectangular populations.

(ii) To tabulate and compare these results with those obtained for Mood's median test in order to evaluate if there is any resultant gain in the use of the Mann-Whitney U test. The latter is more elaborate than the former.

1.2.1 The Two Sample Problem - Mann-Whitney U Test.

Let $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be independently distributed with continuous cumulative distribution functions (cdfs) F and G respectively. We want to test the hypothesis

$$H_o : F(x) = G(x) ,$$

against the alternative H_1 given by

$$H_1 : F(x) > G(x) .$$

Let $n = n_1 + n_2$ denote the size of the combined sample and $Z_{(1)} < Z_{(2)} < \ldots < Z_{(n)}$ be the combined ordered X's and Y's. This ordering is unique with probability 1, since $\Pr\{X_j = X_{j'}\} = \Pr\{Y_i = Y_{i'}\} = \Pr\{X_j = Y_i\} = 0$ due to the assumption of continuity of F and G.

The test originally proposed by Wilcoxon [22] is based on the statistic T which is the sum of the ranks of the Y's in the combined ordered sample. A test of size α based on Wilcoxon's statistic is:

reject H_o if $T \geq t_\alpha$ and

accept H_o if $T < t_\alpha$, where $\Pr\{T \geq t_\alpha \mid H_o\} \leq \alpha$.

This test was modified by Mann and Whitney [14] by defining a statistic U which is equal to the number of times a Y precedes an X in the combined ordered sample. Then, a test of size α based on the Mann-Whitney U statistic is

reject H_o if $U \leq u_\alpha$ and

accept H_o if $U > u_\alpha$, where $\Pr\{U \leq u_\alpha \mid H_o\} \leq \alpha$.

This U statistic is related to Wilcoxon's T statistic by

$$U = n_1 n_2 + \tfrac{1}{2} n_2(n_2 + 1) - T \,, \tag{1.1}$$

which gives a simple way of computing U from the observed value of T. The exact distribution of U under the null hypothesis H_o has been tabulated by Mann and Whitney [14].

1.2.2 <u>The Null Distribution</u>.

Mann and Whitney have shown that the null distribution can be calculated recursively from

$$P_{n_1,n_2}(u) = \frac{n_1}{n_1 + n_2} P_{n_1-1,n_2}(u-n_2) + \frac{n_2}{n_1 + n_2} P_{n_1,n_2-1}(u) \,,$$

with

$$\left.\begin{aligned} P_{0,n_2}(u) &= 0 \\ P_{n_1,0}(u) &= 0 \end{aligned}\right\} \quad \text{if } u > 0 \,,$$

$$\left.\begin{aligned} P_{0,n_2}(u) &= 1 \\ P_{n_1,0}(u) &= 1 \end{aligned}\right\} \quad \text{if } u = 0 \,, \tag{1.2}$$

and

$$P_{n_1,n_2}(u) = 0 \qquad \text{if } u < 0 \,,$$

where $P_{n_1,n_2}(u) = \Pr\{U = u \mid H_o\}$ for samples of size n_1 and n_2. However, it would be desirable to be able to express the null distribution in closed form and to simultaneously derive a joint density function which could be used to calculate the exact power under fixed alternatives.

Let us first consider that the set $\{y_i \mid i = 1, 2, \ldots, n_2\}$ has been chosen from G noting that there are n_2 factorial ways of obtaining the set. Next we order the set and then compute the probability of choosing a set $\{x_j \mid j = 1, 2, \ldots, n_1\}$ from F such that a specific value for U is obtained. (That is, we want an expression for the joint distribution of U and the Y's.) For simplicity, we will first consider the special cases of $n_2 = 1, 2, 3$ and then generalize the results to the case of arbitrary n_2. For convenience, we define the following set of symbols to simplify the notation.

Let $\{i_\ell\}$ be an arbitrary set of integer variables. Define

(a) $\xi_k = u - \sum_{\ell=1}^{k-1} (\ell+1) i_\ell$ for $k > 1$, with $\xi_1 = u$,

(b) $\lambda_k = \xi_k + \sum_{\ell=1}^{k-1} i_\ell$ for $k > 1$, with $\lambda_1 = \xi_1$,

(c) $\alpha_k = n_1 - \lambda_k = n_1 - u + \sum_{\ell=1}^{k-1} \ell\, i_\ell$, for $k > 1$, $\alpha_1 = n_1 - \lambda_1$,

(d) $\beta_k = -\min(0, \alpha_k/k)$, where a fraction such as a/b denotes the largest integer contained in the quotient of a divided by b,

(e) $\binom{n_1}{i_1, i_2, \ldots, i_k} = n_1! \left[\prod_{\ell=1}^{k} i_\ell!\right]^{-1} \left[\left(n_1 - \sum_{\ell=1}^{k} i_\ell\right)!\right]^{-1}$

(f) $$\delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases} \qquad (1.3)$$

For simplicity throughout, $F(y_i)$ and $G(y_i)$ will be written as F_i and G_i respectively.

Let $n_2 = 1$, then we have one value of y say y_1, and we want to choose u values of x greater than y_1, and n_1-u values less than y_1. Since F is the cumulative distribution function of x, we get:

$$h(u, y_1) = 1!\binom{n_1}{u} F_1^{\,n_1-u} (1-F_1)^u \frac{dG_1}{dy_1}, \quad 0 \leq u \leq n_1 . \tag{1.4}$$

Using the special notation, this expression becomes

$$h(u, y_1) = \delta_{\beta_1,0}\binom{n_1}{\xi_1} F_1^{\,\alpha_1} (1-F_1)^{\xi_1} \frac{dG_1}{dy_1}, \tag{1.5}$$

where $\delta_{\beta_1,0}$ is the Kronecker delta defined by (1.3f) .

If $n_2 = 2$, we want to choose n_1 values of x from F such that the total number of x's greater than y_1 and y_2 is equal to u. This can be accomplished in several ways, noting that each value of x greater than y_2 is counted twice in generating the value of u. The resulting joint distribution of u, y_1, and y_2 is

$$h(u, y_1, y_2) = 2! \sum_{i_1} \binom{n_1}{u-2i_1,\ i_1} F_1^{\,n_1-u+i_1} (F_2-F_1)^{u-2i_1} \cdot$$

$$(1-F_2)^{i_1} \frac{dG_1}{dy_1} \frac{dG_2}{dy_2}, \tag{1.6}$$

where i_1 denotes the number of x's that are greater than y_2. The sum over i_1 includes all permissible values of i_1 such that none of the exponents in the expression become negative. Thus, these restrictions on the allowable values of i_1 can be restated in the following form:

(i) $u - 2i_1 \geq 0 \Rightarrow i_1 \leq u/2 = \xi_1/2$, and

(ii) $n_1 - u + i_1 \geq 0 \Rightarrow i_1 \geq -\min(0, n_1-u) = \beta_1$ or $\beta_2 = 0$.

These results may be combined together to yield

$$\beta_1 \leq i_1 \leq \xi_1/2 .$$

Recalling that $\beta_2 = 0$ whenever $i_1 \geq \beta_1$, (1.6) can be written in the following form:

$$h(u, y_1, y_2) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} F_1^{\alpha_2} (F_2-F_1)^{\xi_2} (1-F_2)^{i_1} \frac{dG_1}{dy_1} \frac{dG_2}{dy_2} . \tag{1.7}$$

Similarly for $n_2 = 3$, we want to choose n_1 values of x from F such that the total number of x's greater than y_1, y_2, and y_3 is equal to u. Again, those values of x between y_2 and y_3 are counted twice, while those greater than y_3 are counted three times. The resulting joint density function of u, y_1, y_2, and y_3 under these circumstances is

$$h(u, y_1, y_2, y_3) = 3! \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \delta_{\beta_3,0} \binom{n_1}{\xi_3;\ i_1;\ i_2} F_1^{\alpha_3} (F_2-F_1)^{\xi_3} .$$

$$(F_3-F_2)^{i_1} (1-F_3)^{i_2} \frac{dG_1}{dy_1} \frac{dG_2}{dy_2} \frac{dG_3}{dy_3}, \tag{1.8}$$

where i_1 denotes the number of x's greater than y_2 and less than y_3 and i_2 denotes the number of x's greater than y_3.

The joint density function for the general case can be found by using techniques similar to those used in the previous cases. This argument yields

$$h(u, y_1, \ldots, y_{n_2}) = n_2! \sum_{i_1=0}^{\xi_1/2} \cdots \sum_{i_{n_2-2}=0}^{\xi_{n_2-2}/(n_2-1)} \sum_{i_{n_2-1}=0}^{\xi_{n_2-1}/n_2} \delta_{\beta_{n_2},0}$$

$$\binom{n_1}{\xi_{n_2}, i_1, \ldots, i_{n_2-1}} F_1^{\alpha_{n_2}} (F_2-F_1)^{\xi_{n_2}} \cdot$$

$$(F_3-F_2)^{i_1} \cdots (F_{n_2}-F_{n_2-1})^{i_{n_2-2}} (1-F_{n_2})^{i_{n_2-1}} \cdot$$

$$\frac{dG_1}{dy_1} \frac{dG_2}{dy_2} \cdots \frac{dG_{n_2}}{dy_{n_2}} \quad . \tag{1.9}$$

Now the distribution of u under the null hypothesis: $F = G$, can be found by integrating the y's over the range $-\infty < y_1 < \ldots < y_{n_2} < \infty$. To simplify the integration, we transform the variables of integration from y_i to $F(y_i) = F_i$, and the new range of integration is $0 < F_1 < \ldots < F_{n_2} < 1$. We will first consider the special cases of $n_2 = 1, 2, 3$ and then extend the results to the general case.

For $n_2 = 1$, we substitute $F_1 = G_1$ in (1.4) and integrate.

This yields

$$\varphi_o(u) = \int_0^1 \delta_{\beta_1,0} \binom{n_1}{u} F_1^{n_1-u}(1-F_1)^u \, dF_1 = \delta_{\beta_1,0}/(n_1+1) \quad . \tag{1.10}$$

Again in the case $n_2 = 2$, we substitute $F_i = G_i$, $i = 1, 2$ into (1.7) and integrate. Thus

$$\varphi_o(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} \int_0^1 \int_0^{F_2} F_1^{\alpha_2}(F_2-F_1)^{\xi_2}(1-F_2)^{i_1} dF_1 \, dF_2 . \tag{1.11}$$

Letting $Q = \frac{F_1}{F_2}$ in the inner integral, (1.11) becomes

$$\varphi_o(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} \int_0^1 \int_0^1 Q^{\alpha_2}(1-Q)^{\xi_2} \, dQ \, F_2^{\alpha_2+\xi_2+1}(1-F_2)^{i_1} dF_2 , \tag{1.12}$$

which yields two complete Beta functions upon integration. The resulting expression in simplified form is

$$\varphi_o(u) = \frac{2!\ n_1!}{(n_1+2)!} \max\,[0,\ (\xi_1/2) - \beta_1 + 1] \quad . \tag{1.13}$$

Similarly the case for $n_2 = 3$ yields three complete Beta function integrals that can be simplified to

$$\varphi_o(u) = \frac{3!\ n_1!}{(n_1+3)!} \sum_{i_1=0}^{\xi_1/2} \max\,[0,\ (\xi_2/3) - \beta_2 + 1] \quad . \tag{1.14}$$

The general case for arbitrary n_2 can be developed in the same manner starting with (1.9). The resulting integrals simplify to n_2 complete Beta functions. These results can be simplified to yield

$$\varphi_o(u) = \frac{n_1!\, n_2!}{(n_1 + n_2)!} \sum_{i_1=0}^{\xi_1/2} \cdots \sum_{i_{n_2-2}=0}^{\xi_{n_2-2}/(n_2-1)} \max\,[0, (\xi_{n_2-1}/n_2) - \beta_{n_2-1} + 1] \,. \tag{1.15}$$

As a check, $\varphi_o(u)$ was evaluated for the cases $n_2 = 8$, $1 \le n_1 \le 8$, and $0 \le u \le n_1 n_2$, which showed complete agreement with the results given by Mann and Whitneys' recursive formulas (1.2).

1.3.1 <u>Power of U Test Against the Alternatives of Translation in the Exponential Population</u>.

Here the alternative hypothesis considered is

$$H_a \begin{cases} F(x) = 1 - e^{-x} & x \ge 0 \,, \\ \quad\;\; = 0 & x < 0 \,, \\ G(y) = 1 - e^{-(y-a)} & y \ge a \,, \\ \quad\;\; = 0 & y < a \,, \text{ where } a > 0 \,. \end{cases} \tag{1.15}$$

Let $\varphi_a(u)$ denote the probability of U taking on the value u given that H_a is true. Then

$$\varphi_a(u) = \int \cdots \int h(u, y_1, \ldots, y_{n_2})\, dy_1 \ldots dy_{n_2} \tag{1.16}$$

where h is given by (1.9). We will first consider the results for the special cases of $n_2 = 1, 2, 3$ and then will extend the results to the general case. For convenience of notation we further let

$$\eta = e^{-a} , \quad \gamma = 1 - \eta . \tag{1.17}$$

Then for $n_2 = 1$, the function to be evaluated under the alternative hypothesis H_a is

$$\varphi_a(u) = \int_{y_1=-\infty}^{y_1=\infty} \delta_{\beta_1,0} \binom{n_1}{\xi_1} F_1^{\alpha_1}(1-F_1)^{\xi_1} \, dG_1 . \tag{1.18}$$

Since $G_1 = 0$ for $y_1 < a$, the range of integration on y_1 can be reduced to $a \leq y_1 \leq \infty$, and we can substitute

$$F_1 = \gamma + \eta G_1 = 1 - \eta(1-G_1) \quad \text{valid for} \quad a \leq y_1 \leq \infty ,$$

into (1.18). This yields

$$\varphi_a(u) = \binom{n_1}{\xi_1} \eta^{\xi_1} \delta_{\beta_1,0} \int_0^1 (\gamma + \eta G_1)^{\alpha_1}(1-G_1)^{\xi_1} \, dG_1 . \tag{1.19}$$

Now if we expand $(\gamma + \eta G_1)^{\alpha_1}$ by the binomial theorem and interchange the order of summation and integration, we get

$$\varphi_a(u) = \binom{n_1}{\xi_1} \delta_{\beta_1,0} \sum_{v=0}^{\alpha_1} \binom{\alpha_1}{v} \gamma^{\alpha_1-v} \eta^{\xi_1+v} \int_0^1 G_1^{\,v} (1-G_1)^{\xi_1} \, dG_1 . \tag{1.20}$$

The resulting Beta function can be simplified to yield

$$\varphi_a(u) = n_1! \, 1! \, \delta_{\beta_1,0} \sum_{v=0}^{\alpha_1} \left[\gamma^{\alpha_1-v} \eta^{\xi_1+v} \right] / [(\alpha_1-v)!(\xi_1+v+1)!] . \tag{1.21}$$

Likewise in the case $n_2 = 2$, (1.16) becomes

$$\varphi_a(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} \int_{y_2=a}^{\infty} \int_{y_1=a}^{y_2} F_1^{\alpha_2} (F_2-F_1)^{\xi_2}$$

$$(1-F_2)^{i_1}\, dF_1\, dG_2\ , \qquad (1.22)$$

where the range of the y's has been reduced to $a \leq y_1 \leq y_2 < \infty$, since $G = 0$ for $y < a$. Now substituting

$$F_j = \gamma + \eta\, G_j = 1 - \eta(1 - G_j) \quad \text{valid for} \quad a \leq y_j \leq \infty \qquad (1.23)$$

in (1.22) and expanding the term $(\gamma + \eta\, G_1)^{\alpha_2}$ we get

$$\varphi_a(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \sum_{v=0}^{\alpha_2} \binom{n_1}{\xi_2,\ i_1} \binom{\alpha_2}{v} \gamma^{\alpha_2-v}\, \eta^{\xi_2+v+i_1}\ .$$

$$\int_0^1 \int_0^{G_2} G_1^{v} (G_2-G_1)^{\xi_2} (1-G_2)^{i_1}\, dG_1\, dG_2\ . \qquad (1.24)$$

By transforming the variables of integration we get two complete Beta functions which can be simplified to

$$\varphi_a(u) = n_1!\ 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \sum_{v=0}^{\alpha_2} \left[\gamma^{\alpha_2-v}\, \eta^{\xi_2+v+i_1} \right] .$$

$$[(\alpha_2-v)!(\xi_2+v+i_1+2)!]^{-1}\ . \qquad (1.25)$$

In the case $n_2 = 3$, we get three complete Beta functions. These can be simplified to yield

$$\varphi_a(u) = n_1!\ 3! \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \delta_{\beta_3,0} \sum_{v=0}^{\alpha_3} \gamma^{\alpha_3-v} \eta^{\lambda_3+v} \cdot$$

$$[(\alpha_3-v)!(\lambda_3+v+3)!]^{-1} . \tag{1.26}$$

Following a development along the lines used in the previous special cases, we get in the general case n_2 complete Beta functions. These simplify to give

$$\varphi_a(u) = n_1!\ n_2! \sum_{i_1=0}^{\xi_1/2} \cdots \sum_{i_{n_2-2}=0}^{\xi_{n_2-2}/(n_2-1)} \sum_{i_{n_2-1}=0}^{\xi_{n_2-1}/n_2} \delta_{\beta_{n_2},0}$$

$$\sum_{v=0}^{\alpha_{n_2}} \gamma^{\alpha_{n_2}-v} \eta^{\lambda_{n_2}+v} / [(\alpha_{n_2}-v)!(\lambda_{n_2}+v+n_2)!] . \tag{1.27}$$

The power of the test can be computed from (1.27) by evaluating

$$\Pr\{ U \leq u_\alpha \mid H_a \} = \sum_{u=0}^{u_\alpha} \varphi_a(u) \tag{1.28}$$

where u_α is determined from α, the level of significance, by evaluating

$$\Pr\{ U \leq u_\alpha \mid H_o \} \leq \alpha .$$

1.4 Power of U Test Against the Alternatives of Change in Location and Scale in the Rectangular Population.

The alternative hypothesis is given by

$$H_{a\theta}\left\{\begin{array}{llll} F(x) & = & x & 0 \le x \le 1\,, \\ & = & 0 & x < 0\,, \\ & = & 1 & x > 1\,, \\ G(y) & = & (y-a)/\theta & a \le y \le a+\theta\,, \\ & = & 0 & y < a\,, \\ & = & 1 & y > a+\theta\,, \text{ where } a > 0,\ \theta > 0\,. \end{array}\right. \qquad (1.29)$$

Let $\varphi_{a\theta}(u)$ denote the probability of U taking on the value u given that $H_{a\theta}$ is true. Then as in (1.16)

$$\varphi_{a\theta}(u) = \int \cdots \int h(u, y_1, \ldots, y_{n_2})\, dy_1 \ldots dy_{n_2}\,, \qquad (1.30)$$

where h is given by (1.9). For notational convenience let

$$b = 1 - a\,. \qquad (1.31)$$

There are two cases to be distinguished, namely:

(i) $a + \theta \le 1$, (ii) $a + \theta > 1$.

1.4.1 $\underline{a + \theta \le 1}$.

As before, we will consider first the three special cases, $n_2 = 1, 2, 3$ and then will extend the results to the general case.

Then for $n_2 = 1$, the function to be evaluated under the alternative hypothesis $H_{a\theta}$ is:

$$\varphi_{a\theta}(u) = \int_{y_1=-\infty}^{y_1=\infty} \delta_{\beta_1,0} \binom{n_1}{\xi_1} F_1^{\alpha_1}(1-F_1)^{\xi_1}\, dG_1 . \qquad (1.32)$$

Since $G_1 = 0$ for $y_1 < a$ and $G_1 = 1$ for $y_1 > a + \theta$, the range of integration on y_1 can be reduced to $a \leq y_1 \leq a + \theta$, and we can substitute

$$F_1 = a + \theta G_1 \quad \text{valid for} \quad a \leq y_1 \leq a + \theta ,$$

in (1.32). This yields

$$\varphi_{a\theta}(u) = \binom{n_1}{\xi_1} \delta_{\beta_1,0} \int_0^1 (a+\theta G_1)^{\alpha_1}(b-\theta G_1)^{\xi_1}\, dG_1 . \qquad (1.33)$$

Now if we expand $(a + \theta\, G_1)^{\alpha_1}$ and $(b - \theta\, G_1)^{\xi_1}$ and interchange the order of summation and integration, we get

$$\varphi_{a\theta}(u) = \binom{n_1}{\xi_1} \delta_{\beta_1,0} \sum_{v=0}^{\alpha_1} \sum_{q=0}^{\xi_1} \binom{\alpha_1}{v}\binom{\xi_1}{q} (-1)^q\, a^{\alpha_1-v}\, b^{\xi_1-q}\, \theta^{v+q} \cdot \int_0^1 G_1^{v+q}\, dG_1 . \qquad (1.34)$$

This expression can be integrated and simplified to yield

$$\varphi_{a\theta}(u) = n_1!\, \delta_{\beta_1,0} \sum_{v=0}^{\alpha_1} \sum_{q=0}^{\xi_1} [(-1)^q a^{\alpha_1-v} b^{\xi_1-q} \theta^{v+q}][(\alpha_1-v)!(\xi_1-q)! \cdot v!\, q!\, (v+q+1)]^{-1} . \qquad (1.35)$$

Likewise in the case $n_2 = 2$, (1.30) becomes

$$\varphi_{a\theta}(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2, i_1} \int_{y_2=a}^{a+\theta} \int_{y_1=a}^{y_2} F_1^{\alpha_2} (F_2-F_1)^{\xi_2} (1-F_2)^{i_1} \, dG_1 dG_2 , \tag{1.36}$$

where the range of the y's has been reduced to $a \le y_1 \le y_2 \le a+\theta$ since $G = 0$ for $y < a$ and $G = 1$ for $y > a + \theta$. Now substituting

$$F_j = a + \theta G_j \quad \text{valid for } a \le y_j \le a + \theta \tag{1.37}$$

in (1.36) and expanding the terms $(a + \theta G_1)^{\alpha_2}$ and $\left(b - \theta G_2\right)^{i_1}$ we get

$$\varphi_{a\theta}(u) = 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \sum_{v=0}^{\alpha_2} \sum_{q=0}^{i_1} \binom{n_1}{\xi_2, i_1} \binom{\alpha_2}{v} \binom{i_1}{q} (-1)^q a^{\alpha_2-v} b^{i_1-q} \cdot$$

$$\theta^{\xi_2+v+q} \int_0^1 \int_0^{G_2} G_1^v (G_2 - G_1)^{\xi_2} G_2^q \, dG_1 \, dG_2 \quad . \tag{1.38}$$

Letting $Q = G_1/G_2$, the innermost integral in (1.38) yields a complete Beta function. The resulting expression when integrated and simplified becomes

$$\varphi_{a\theta}(u) = n_1! \; 2! \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \sum_{v=0}^{\alpha_2} \sum_{q=0}^{i_1} [(-1)^q a^{\alpha_2-v} b^{i_1-q} \theta^{\xi_2+v+q}] \cdot$$

$$[q! \; (i_1-q)!(\alpha_2-v)!(\xi_2+v+1)!(\xi_2+v+q+2)]^{-1} \quad . \tag{1.39}$$

In the case $n_2 = 3$, we get two complete Beta functions that can be simplified to yield

$$\varphi_{a\theta}(u) = n_1!\ 3! \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \delta_{\beta_3,0} \sum_{v=0}^{\alpha_3} \sum_{q=0}^{i_2} [(-1)^q a^{\alpha_3-v} b^{i_2-q} \theta^{\lambda_3-i_2+v+q}] \cdot$$

$$\alpha_3-v)!\ (i_2-q)!\ (\lambda_3-i_2+v+2)!\ (\lambda_3-i_2+v+q+3))]^{-1} \quad . \quad (1.40)$$

Following a development along the lines used in the previous special cases, we get in the general case $n_2 - 1$ complete Beta functions. The resulting integrated expression can be simplified to give

$$\varphi_{a\theta}(u) = n_1!\ n_2! \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \cdots \sum_{i_{n_2-1}=0}^{\xi_{n_2-1}/n_2} \delta_{\beta_{n_2},0} \sum_{v=0}^{\alpha_{n_2}} \sum_{q=0}^{i_{n_2-1}}$$

$$[(-1)^q a^{\alpha_{n_2}-v} b^{i_{n_2-1}-q} \theta^{\lambda_{n_2}-i_{n_2-1}+v+q}] \cdot$$

$$[q!(\alpha_{n_2}-v)!(i_{n_2-1}-q)!(\lambda_{n_2}-i_{n_2-1}+v+n_2-1)!(\lambda_{n_2}-i_{n_2-1}+v+q+n_2))]^{-1} \quad .$$

(1.41)

1.4.2 $\underline{a + \theta = 1}$.

In this special case (1.41) can be further simplified to yield

$$\varphi_{a\theta}(u) = n_1!\, n_2! \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \cdots \sum_{i_{n_2-1}=0}^{\xi_{n_2-1}/n_2} \delta_{\beta_{n_2},0} \sum_{v=0}^{\alpha_{n_2}} \sum_{q=0}^{i_{n_2-1}}$$

$$[(-1)^q a^{\alpha_{n_2}-v} b^{\lambda_{n_2}+v}]\,[q!(\alpha_{n_2}-v)!(i_{n_2-1}-q)!\,\cdot$$

$$(\lambda_{n_2}-i_{n_2-1}+v+n_2-1)!(\lambda_{n_2}-i_{n_2-1}+v+q+n_2)]^{-1} \quad . \qquad (1.42)$$

1.4.3 $\underline{a + \theta > 1}$.

This case must be further subdivided into two subcases, namely:

(i) $a < 1$,

(ii) $a \geq 1$.

1.4.3.1 $\underline{a < 1}$.

For $a < 1$, the range of integration for y can be split into four parts, namely: (1) $-\infty < y < a$, (2) $a \leq y \leq 1$, (3) $1 < y \leq a + \theta$, (4) $a + \theta < y < \infty$. Over parts (1) and (4) the value of the integral is zero since G is constant. Hence, we will consider only the ranges (2) and (3).

For the case of $n_2 = 1$, (1.30) becomes

$$\varphi_{a\theta}(u) = 1!\left[\int_{y_1=a}^{1} \delta_{\beta_1,0}\binom{n_1}{\xi_1}F_1^{\alpha_1}(1-F_1)^{\xi_1}\,dG_1 + \delta_{u,0}\int_{y_1=1}^{a+\theta} dG_1\right], \tag{1.43}$$

since $\xi_1 = u$ and $u = 0$ implies that $\beta_1 = 0$.

Consider $P_1(u)$ defined by

$$P_1(u) = (n_1!)^{-1}\int_{y_1=a}^{1} \delta_{\beta_1,0}\binom{n_1}{\xi_1}F_1^{\alpha_1}(1-F_1)^{\xi_1}\,dG_1 \; . \tag{1.44}$$

Perform the following substitution: $F_1 = a + \theta\, G_1$, $a \leq y_1 \leq 1$.

Then (1.44) becomes

$$P_1(u) = (n_1!)^{-1}\int_{0}^{b/\theta} \delta_{\beta_1,0}\binom{n_1}{\xi_1}(a + \theta\, G_1)^{\alpha_1}(b - \theta\, G_1)^{\xi_1}\,dG_1 \; . \tag{1.45}$$

If we expand both binomials and integrate the resulting expression, (1.45) takes the form:

$$P_1(u) = \delta_{\beta_1,0}\sum_{v=0}^{\alpha_1}\sum_{q=0}^{\xi_1}[(-1)^q\, a^{\alpha_1-v}\, b^{\xi_1+v+1}][\theta\; v!\; q!\; (\alpha_1-v)!\,\cdot$$

$$(\xi_1-q)!\;(v+q+1)]^{-1} \; . \tag{1.46}$$

If we define

$$P_0(u) = \delta_{u,0}/n_1! \;, \tag{1.47}$$

then since $G_1(1) = b/\theta$, (1.43) can be written as

$$\varphi_{a\theta}(u) = n_1!\; 1!\; [P_1(u) + P_0(u)(1 - b/\theta)] \;. \tag{1.48}$$

If $n_2 = 2$, the expression for $\varphi_{a\theta}(u)$ can be broken down into three integrals as follows:

$$\varphi_{a\theta}(u) = 2! \int_{y_2=a}^{1} \int_{y_1=a}^{y_2} \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\; i_1} F_1^{\alpha_2}(F_2-F_1)^{\xi_2}(1-F_2)^{i_1} dG_1 dG_2 +$$

$$2! \int_{y_2=1}^{a+\theta} \int_{y_1=a}^{1} \delta_{\beta_1,0} \binom{n_1}{\xi_1} F_1^{\alpha_1} (1-F_1)^{\xi_1} dG_1\, dG_2 +$$

$$2!\; \delta_{u,0} \int_{y_2=1}^{a+\theta} \int_{y_1=1}^{y_2} dG_1\, dG_2 \;. \tag{1.49}$$

Note that when $i_{k-1} = 0$, then $\xi_k = \xi_{k-1}$, $\alpha_k = \alpha_{k-1}$, $\lambda_k = \lambda_{k-1}$, $\beta_k = \beta_{k-1}$ $(k = 2, 3, \ldots, n_2)$. Also $u = 0$ implies that $\beta_1 = 0$.

Let us consider the first of the above three integrals by defining

$$P_2(u) = (n_1!)^{-1} \int_{y_2=a}^{1} \int_{y_1=a}^{y_2} \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} F_1^{\alpha_2} (F_2-F_1)^{\xi_2} (1-F_2)^{i_1}\, dG_1\, dG_2 \ . \tag{1.50}$$

We can substitute

$$F_i = a + \theta\, G_i \ ; \quad a \le y_i \le 1 \ , \tag{1.51}$$

into (1.50). Also we can expand the binomials and interchange the order of summation and integration. This yields

$$P_2(u) = (n_1!)^{-1} \sum_{i_1=0}^{\xi_1/2} \sum_{v=0}^{\alpha_2} \sum_{q=0}^{i_1} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} \binom{\alpha_2}{v} \binom{i_1}{q} (-1)^q \cdot$$

$$a^{\alpha_2-v}\, b^{i_1-q}\, \theta^{\xi_2+v+q} \int_{y_2=a}^{1} \int_{y_1=a}^{y_2} G_1^{v}\, (G_2-G_1)^{\xi_2} G_2^{q}\, dG_1\, dG_2 \ . \tag{1.52}$$

This expression integrates and simplifies to

$$P_2(u) = \sum_{i_1=0}^{\xi_1/2} \sum_{v=0}^{\alpha_2} \sum_{q=0}^{i_1} [(-1)^q\, a^{\alpha_2-v}\, b^{\xi_2+i_1+v+2}][\theta^2\, q!\, (i_1-q)! \cdot$$

$$(\alpha_2-v)!\,(\xi_2+v+1)!\,(\xi_2+v+q+2)]^{-1} \ . \tag{1.53}$$

With this notation (1.49) can be written as

$$\varphi_{a\theta}(u) = n_1!\ 2!\ [P_2(u) + P_1(u)(1 - b/\theta) + P_0(u)(1 - b/\theta)^2(2!)^{-1}]. \tag{1.54}$$

For $n_2 = 3$, the expression for $\varphi_{a\theta}(u)$ becomes

$$\varphi_{a\theta}(u) = 3! \int_{y_3=a}^{1} \int_{y_2=a}^{y_3} \int_{y_1=a}^{y_2} \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \delta_{\beta_3,0} \binom{n_1}{\xi_3,\ i_1,\ i_2} F_1^{\alpha_3} \cdot$$

$$(F_2-F_1)^{\xi_3}(F_3-F_2)^{i_1}(1-F_3)^{i_2}\ dG_1\ dG_2\ dG_3 +$$

$$3! \int_{y_3=1}^{a+\theta} \int_{y_2=a}^{1} \int_{y_1=a}^{y_2} \sum_{i_1=0}^{\xi_1/2} \delta_{\beta_2,0} \binom{n_1}{\xi_2,\ i_1} F_1^{\alpha_2}(F_2-F_1)^{\xi_2} \cdot$$

$$(1-F_2)^{i_1}\ dG_1\ dG_2\ dG_3 +$$

$$3! \int_{y_3=1}^{a+\theta} \int_{y_2=1}^{y_3} \int_{y_1=a}^{1} \delta_{\beta_1,0} \binom{n_1}{\xi_1} F_1^{\alpha_1}(1-F_1)^{\xi_1}\ dG_1\ dG_2\ dG_3 +$$

$$3! \int_{y_3=1}^{a+\theta} \int_{y_2=1}^{y_3} \int_{y_1=1}^{y_2} \delta_{u,0}\ dG_1\ dG_2\ dG_3 \quad . \tag{1.55}$$

Now, if we define

$$P_3(u) = (n_1!)^{-1} \int_{y_3=a}^{1} \int_{y_2=a}^{y_3} \int_{y_1=a}^{y_2} \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \delta_{\beta_3,0} \binom{n_1}{\xi_3,\ i_1,\ i_2} \cdot$$

$$F_1^{\alpha_3}(F_2-F_1)^{\xi_3}(F_3-F_2)^{i_1}(1-F_3)^{i_2}\, dG_1\, dG_2\, dG_3 \quad ,$$

which upon integration and simplification, becomes

$$P_3(u) = \sum_{i_1=0}^{\xi_1/2} \sum_{i_2=0}^{\xi_2/3} \sum_{v=0}^{\alpha_3} \sum_{q=0}^{i_2} \left[\delta_{\beta_3,0}(-1)^q\, a^{\alpha_3-v}\, b^{\lambda_3+v+3}\right] \cdot$$

$$[\theta^3\, q!\, (\alpha_3-v)!\, (i_2-q)!\, (\lambda_3-i_2+v+2)!\, (\lambda_3-i_2+v+q+3)]^{-1} \quad , \tag{1.56}$$

then (1.55) becomes

$$\varphi_{a\theta}(u) = n_1!\ 3!\ \sum_{j=0}^{3} \left\{P_j(u)\ (1-b/\theta)^{3-j}[(3-j)!]^{-1}\right\} \quad . \tag{1.57}$$

In an analogous fashion for the general case, we can define $P_j(u)$ and obtain

$$P_j(u) = \sum_{i_1=0}^{\xi_1/2} \cdots \sum_{i_{j-1}=0}^{\xi_{j-1}/j} \sum_{v=0}^{\alpha_j} \sum_{q=0}^{i_{j-1}} \left[\delta_{\beta_j,0}(-1)^q\, a^{\alpha_j-v}\, b^{\lambda_j+v+j}\right] \cdot \tag{1.58}$$

$$[\theta^j\, q!\, (\alpha_j-v)!\, (i_{j-1}-q)!(\lambda_j-i_{j-1}+v+j-1)!(\lambda_j-i_{j-1}+v+q+j)]^{-1},$$

for $j \geq 2$, with $P_0(u)$, $P_1(u)$ as previously defined.

Then

$$\varphi_{a\theta}(u) = n_1!\, n_2! \sum_{j=0}^{n_2} \left\{ P_j(u)\ (1-b/\theta)^{n_2-j} [(n_2-j)!]^{-1} \right\}. \quad (1.59)$$

1.4.3.2 a ≥ 1

In the case $a \geq 1$, the results are trivial, namely:

$$\varphi_{a\theta}(u) = n_2!\, \delta_{u,0} \int_{y_{n_2}=a}^{a+\theta} \int_{y_{n_2-1}=a}^{y_{n_2}} \cdots \int_{y_1=a}^{y_2} dG_1 dG_2 \ldots dG_{n_2} = \delta_{u,0} \ . \quad (1.60)$$

Using the above results for $\varphi_{a\theta}(u)$, the power of the U test under the alternative hypothesis can be calculated from

$$\Pr\{\, U \leq u_\alpha \mid H_{a\theta} \} = \sum_{u=0}^{u_\alpha} \varphi_{a\theta}(u) \quad , \quad (1.61)$$

where u_α is determined by the level of significance, α, from the relation

$$\Pr\{\, U \leq u_\alpha \mid H_0 \} \leq \alpha \quad .$$

1.5 Results

Tables 1.1 and 1.2 compare the powers of the Mann-Whitney U test against Mood's median test. The numerical results for Mood's test presented in Tables 1.1 and 1.2 were taken from Leone, Chakravarti and Alanen [13]. In Table 1.1, the exponential alternative is considered for various values of the location parameter, $\underline{a}$, for $\underline{a} = 0(0.1)1$, 1.5, 2, 3, for sample sizes 11 and 15. It should be noted that when the location parameter is zero, we get the null distribution with the power equal to the level of significance, α. Since the distributions of the test statistics are discrete, the values of α do not in general coincide for both the tests. Hence, although many different cases have been computed, only those values that are relatively close together and which indicate the general trend, have been tabulated in Table 1.1. The conclusions that can be drawn from this table (relative to the exponential alternative) are:

1) If n_1 is smaller than n_2, the Mann-Whitney test is more powerful than Mood's test. To note this increase of power, several cases were intentionally chosen where the level of significance for the Mann-Whitney test was slightly less than that of the Mood test. In these cases, the power of Mann-Whitney's test rapidly overtakes Mood's test as the location parameter, $\underline{a}$, increases.

2) If n_1 is larger than n_2, Mood's test is more powerful than the Mann-Whitney test. Likewise, to note this increase of power, several cases were intentionally chosen where the level of significance for Mood's test was slightly less than that of the Mann-Whitney test. In these cases, the power of Mood's test rapidly overtakes Mann-Whitney's test as $\underline{a}$

increases.

3) In those cases in which $n_1 \approx n_2$, the two test procedures seem to exhibit powers that are approximately the same.

The rectangular alternative for the special case in which $\theta = 1 - a$, is considered in Table 1.2. The values of the parameter, a, range between 0.0 and 0.9 with increments of 0.1 (where the value of 0.0 indicates the level of significance, α, of the test under the null hypothesis). The total sample sizes chosen are again, 11 and 15. As in the case of the exponential alternatives, the levels of significance do not in general coincide, since the distributions are discrete, but in those cases in which the levels are relatively close together, the results indicate that the conclusions drawn from the exponential data continue to hold in the rectangular case.

In both of these tables, a is non-negative. If the alternative hypothesis were for $a < 0$, the same situation would hold. That is, if $n_1 > n_2$, the Mood test would exhibit more power, while the Mann-Whitney test would be more powerful for $n_1 < n_2$.

These results (that is, with respect to the exponential and rectangular alternatives) indicate that in those cases when $n_1 \geq n_2$, it is preferable to use Mood's median test over the Mann-Whitney U test. A further advantage in the case of Mood's test is that the experiment needs to be run only until the median of the combined sample has been observed. In many experiments, this fact gives rise to a reduction in the cost, due to savings in time, experimental material, availability of equipment, and the like.

1.6 Further Extensions.

First, the results in Tables I and II should be extended to larger sample sizes to see if the previous results still hold. This extension will also indicate how rapidly the results approach the asymptotic situation, and the complete tables can be used to determine the sample size required to obtain a given power.

Second, tables similar to Tables I and II should be computed for comparing the Mann-Whitney U test with Massey's two sample test.

Third, exponential alternatives with a change in the scale parameter should be considered. Power functions for these alternatives can be developed for Mood's, Massey's and Mann-Whitney's two sample tests, and tables comparing these results can be computed.

Fourth, an attempt should be made to analytically compare the power functions of Mood's, Massey's and Mann-Whitney's tests independent of the computational results to see if the same conclusions are indicated.

Fifth, an attempt should be made to develop a class of functions to which the results of this chapter can be applied.

Sixth, the power function for the Mann-Whitney U test can be extended to the case of c samples for whatever tests are developed for the c sample case.

TABLE 1.1

POWER OF MOOD'S MEDIAN TEST, MANN-WHITNEY'S U TEST -- EXPONENTIAL DISTRIBUTION

$F(x) = 1 - e^{-x}$, $G(y) = 1 - e^{-(y-a)}$, $a > 0$

n	n_1	a = 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.5	2.0	3.0
11	4	.0152	.0241	.0370	.0545	.0765	.1031	.1338	.1680	.2049	.2438	.2841	.4851	.6536	.8591*
		.0141	.0237	.0417	.0661	.0960	.1307	.1689	.2098	.2524	.2959	.3395	.5423	.6996	.8808**
	5	.0022	.0039	.0071	.0126	.0211	.0334	.0499	.0708	.0960	.1252	.1577	.3502	.5418	.8044
		.0021	.0039	.0071	.0125	.0210	.0333	.0499	.0708	.0960	.1251	.1577	.3502	.5418	.8044
		.0671	.0975	.1371	.1855	.2410	.3015	.3647	.4283	.4906	.5500	.6057	.8141	.9209	.9877
		.0627	.1018	.1522	.2114	.2764	.3441	.4120	.4781	.5407	.5990	.6522	.8423	.9344	.9900
	6	.0130	.0214	.0351	.0561	.0860	.1252	.1728	.2273	.2868	.3490	.4120	.6887	.8579	.9764
		.0151	.0249	.0399	.0610	.0885	.1219	.1604	.2030	.2484	.2955	.3431	.5644	.7276	.8992
	7	.0455	.0678	.1007	.1464	.2053	.2751	.3519	.4316	.5102	.5847	.6529	.8803	.9657	.9978
		.0545	.0808	.1163	.1609	.2136	.2724	.3351	.3993	.4630	.5245	.5825	.8027	.9163	.9870
15	5	.0070	.0121	.0200	.0314	.0467	.0663	.0900	.1178	.1491	.1833	.2200	.4175	.5979	.8320
		.0063	.0148	.0296	.0513	.0799	.1144	.1538	.1968	.2421	.2886	.3354	.5509	.7122	.8892
	7	.0089	.0158	.0271	.0447	.0699	.1033	.1450	.1939	.2485	.3072	.3679	.6503	.8354	.9717
		.0069	.0146	.0281	.0491	.0785	.1160	.1609	.2115	.2661	.3230	.3805	.6374	.8056	.9460
	8	.0012	.0025	.0050	.0100	.0189	.0336	.0556	.0858	.1243	.1704	.2227	.5190	.7561	.9549
		.0018	.0037	.0074	.0139	.0243	.0397	.0605	.0869	.1186	.1549	.1949	.4184	.6172	.8520
		.0137	.0513	.0806	.1215	.1746	.2385	.3105	.3870	.4645	.5397	.6101	.8579	.9577	.9972
		.0360	.0648	.1072	.1630	.2299	.3042	.3819	.4594	.5336	.6025	.6648	.8725	.9563	.9951
	10	.0186	.0307	.0506	.0823	.1295	.1932	.2713	.3593	.4511	.5412	.6251	.8958	.9786	.9994
		.0199	.0327	.0524	.0810	.1195	.1680	.2253	.2892	.3570	.4263	.4744	.7685	.9094	.9882

*First row — Mood's Median Test **Second row — Mann-Whitney's U Test

TABLE 1.2

POWER OF MOOD'S MEDIAN TEST, MANN-WHITNEY'S U TEST - - RECTANGULAR DISTRIBUTION

$F(x) = x, \quad G(y) = (y-a)/(1-a), \quad a > 0$

n	n_1	a = 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
11	4	.0152 .0121	.0247 .0245	.0406 .0468	.0664 .0824	.1063 .1346	.1665 .2070	.2503 .3030	.3677 .4264	.5257 .5811	.7332* .7709**
	5	.0022 .0021	.0041 .0040	.0082 .0081	.0170 .0169	.0350 .0349	.0693 .0692	.1303 .1302	.2323 .2322	.3941 .3941	.6400 .6400
		.0671 .0627	.0994 .1042	.1476 .1652	.2162 .2477	.3082 .3515	.4240 .4736	.5594 .6080	.7050 .7446	.8447 .8691	.9543 .9624
	6	.0130 .0259	.0220 .0434	.0392 .0729	.0720 .1194	.1299 .1880	.2234 .2827	.3593 .4048	.5357 .5512	.7351 .7127	.9155 .8718
	7	.0455 .0545	.0693 .0824	.1100 .1258	.1783 .1899	.2831 .2791	.4261 .3949	.5963 .5342	.7684 .6871	.9078 .8353	.9847 .9517
15	5	.0070 .0063	.0125 .0154	.0223 .0340	.0396 .0667	.0686 .1185	.1158 .1937	.1891 .2963	.2994 .4285	.4599 .5913	.6869 .7831
	7	.0089 .0069	.0162 .0151	.0306 .0322	.0580 .0647	.1075 .1206	.1903 .2079	.3170 .3324	.4911 .4941	.7001 .6821	.9008 .8685
	8	.0012 .0018	.0026 .0038	.0059 .0086	.0144 .0192	.0356 .0417	.0835 .0849	.1785 .1612	.3419 .2643	.5796 .4661	.8785 .7100
		.0317 .0360	.0526 .0667	.0890 .1189	.1502 .1998	.2460 .3125	.3817 .4541	.5515 .6131	.7332 .7696	.8895 .8986	.9809 .9775
	10	.0186 .0199	.0316 .0336	.0567 .0582	.1070 .1016	.2010 .1738	.3530 .2846	.5554 .4375	.7666 .6225	.9249 .8097	.9925 .9506

*First row — Mood's Median Test   **Second row — Mann-Whitney's U Test

CHAPTER II

ASYMPTOTIC RELATIVE EFFICIENCY OF THE MANN-WHITNEY U TEST AGAINST AN EXPONENTIAL ALTERNATIVE

2.1 Introduction.

In Chapter 1, the exact power of the Mann-Whitney two sample U test for discriminating between two populations was derived. Two types of alternatives were considered; namely, a change in location of an exponential population and a change in location and scale of a rectangular population.

The asymptotic relative efficiency of the Mann-Whitney U test against an alternative of a change in location of a normal population was shown to be $3/\pi$ [16], [1]. The asymptotic relative efficiencies of Mood's test based on the median, and Massey's test based on the first quartile and the median, when compared against the likelihood ratio test appropriate for detecting a shift in location of an exponential population, were found to be zero by Chakravarti, Leone, and Alanen [3].

In this chapter the asymptotic relative efficiencies of the Mann-Whitney U test, when compared with the likelihood ratio test, Mood's and Massey's tests for detecting a shift in location of an exponential population, are considered.

2.2 Limiting Distribution of the Mann-Whitney U Statistic.

Let us define the statistic V_{n_1,n_2} by

$$V_{n_1,n_2} = U/(n_1 n_2) \ , \tag{2.1}$$

then

$$E(V_{n_1,n_2}) = E(U) \ [n_1 n_2]^{-1} \ , \tag{2.2}$$

and

$$\mathrm{Var}(V_{n_1,n_2}) = \mathrm{Var}(U) \ [n_1 n_2]^{-2} \ . \tag{2.3}$$

It has been shown by Lehmann, [11] and [12], that

$$n_2^{\frac{1}{2}} \left(V_{n_1,n_2} - E(V_{n_1,n_2})\right) \tag{2.4}$$

has an asymptotic normal distribution, provided that as $n_1, n_2 \to \infty$,

$$(n_2/n_1) \to \text{constant} < \infty \ . \tag{2.5}$$

Furthermore, Mann and Whitney [14] have shown that

$$E(V_{n_1,n_2}) = \int G \ dF \ , \tag{2.6}$$

and

$$n_1 n_2 \mathrm{Var}(V_{n_1,n_2}) = [(n_1+n_2+1)/12]+[(n_1-1)(\lambda-\epsilon_1)]+[(n_2-1)(\lambda-\epsilon_2)]-[(n_1+n_2-1)\lambda^2] \ , \tag{2.7}$$

where

$$\lambda = \tfrac{1}{2} - \int G\,dF, \quad \epsilon_1 = (1/3) - \int G^2\,dF, \text{ and } \quad \epsilon_2 = (1/3) - \int (1-F)^2\,dG.$$

Thus, the expressions for the mean and variance of U can be written as

$$E(U) = n_1 n_2 \int G\,dF\,, \tag{2.8}$$

$$\mathrm{Var}(U) = (n_1 n_2)\{[(n_1+n_2+1)/12]+(n_1-1)(\lambda-\epsilon_1)+(n_2-1)(\lambda-\epsilon_2) - (n_1+n_2-1)\lambda^2\}\,. \tag{2.9}$$

With the exponential alternative considered in Chapter 1 (See Equation (1.15)), (2.8) and (2.9) become

$$E(U) = (n_1 n_2/2)\, e^{-a}\,, \tag{2.10}$$

and

$$\mathrm{Var}(u) = (n_1 n_2/12)[(n_1+n_2+1)+2(n_1-1)(1-e^{-a})+2(n_2-1)(1-e^{-a})(1-2e^{-a}) - 3(n_1+n_2-1)(1-e^{-a})^2]\,. \tag{2.11}$$

2.3 Asymptotic Relative Efficiency of the Mann-Whitney U Test.

Let θ, the parameter of interest, label the sequence of distributions. Consider the null hypothesis $H_0 : \theta = \theta_0$ and the sequence of alternatives $H_m : \theta_m = \theta_0 + dm^r$ for some positive r and d. Let $N(\delta)$ and $N^*(\delta)$ be respectively the sample sizes required by two test procedures τ and τ^*, to achieve the same power $(1-\beta)$ at the same level of significance α, where δ is the difference $\theta_m - \theta_0$. Then, the asymptotic efficiency of τ^* relative to

τ is defined as:

$$\mathrm{Eff}(\tau^*/\tau) = \lim_{\delta \to 0} [N(\delta)/N^*(\delta)] \quad . \tag{2.12}$$

Since for the Mann-Whitney U test,

$$\lim_{n_1, n_2 \to \infty} P_{\theta_{n_2}} \{[(n_2)^{\frac{1}{2}}(U/n_1 n_2 - E(U/n_1 n_2))][n_1 n_2 \mathrm{Var}(U/n_1 n_2)]^{-\frac{1}{2}} \leq x\} =$$

$$= \Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt \quad , \tag{2.13}$$

where $\theta_{n_2} = dn_2^{-\frac{1}{2}}$ and $\theta_o = 0$, the following theorem due to Hoeffding and Rosenblatt [9] can be applied:

<u>Theorem</u>: If for a sequence of test procedures $\{t_n\}$, where t_n is based on a random sample of size n, the following regularity conditions hold:

a) $\beta_n(\theta_o) \leq \alpha$, $\lim_{n \to \infty} \beta_n(\theta_o) = \alpha$, where $\beta_n(\theta)$ is the probability of rejecting the null hypothesis,

b) There exists a positive r and normalizing functions $\mu(\theta)$ and $\sigma(\theta)$ such that for any real x and any $d \geq 0$,

$$\lim_{n \to \infty} P_{\theta_n} \{n^r [(t_n - \mu(\theta_n))/\sigma(\theta_n)] \leq x\} = \Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt \ ,$$

where $\theta_n = \theta_o + dn^{-r}$,

c) $\mu(\theta)$ has a derivative $\mu'(\theta_o)$ at θ_o and $\mu'(\theta_o) > 0$,

d) $\sigma(\theta)$ is continuous and positive at $\theta = \theta_o$,

then

$$\lim_{n \to \infty} \beta_n(\theta_o + dn^{-r}) = \Phi[d\,\mu'(\theta_o)/\sigma(\theta_o) - \lambda_\alpha] , \qquad (2.14)$$

where $\Phi(-\lambda_\alpha) = \alpha$. The efficiency index $N(\delta)$ of the test based on t_n has the expression

$$N(\delta) = [(\lambda_\alpha + \lambda_\beta)\,\sigma(\theta_o)/(\delta\,\mu'(\theta_o))]^{1/r} . \qquad (2.15)$$

The Mann-Whitney test procedure with $\theta_o = 0$, satisfies all of the hypotheses of Hoeffding and Rosenblatt's theorem with the exception of part (c), $\mu'(\theta_o) > 0$. In this case

$$\mu'(\theta_o) = -(n_1 n_2/2)e^{-\theta_o} ,$$

which for $\theta_o = 0$ becomes

$$\mu'(0) = -(n_1 n_2/2) < 0 .$$

However, the restriction that $\mu'(\theta_o) > 0$ is not necessary in the case of $r = \frac{1}{2}$, since Pitman's original result for $r = \frac{1}{2}$ does not require this restriction on $\mu'(\theta_o)$, [18]. Since for the Mann-Whitney test procedure $r = \frac{1}{2}$, we can ignore the restriction on $\mu'(\theta_o)$ and apply the above theorem with $\theta_o = 0$. This yields

$$N_1(\delta) = \{(\lambda_\alpha + \lambda_\beta)[n_1 n_2 \mathrm{Var}(U/n_1 n_2)_o]^{\frac{1}{2}}[\delta E'(U/n_1 n_2)_o]^{-1}\}^2 , \qquad (2.16)$$

where $E'(U/n_1n_2)$ denotes differentiation with respect to θ and the subscript $_o$ means evaluate at $\theta = 0$.

Now, for the exponential alternatives with a shift in the location parameter, (2.16) becomes

$$N_1(\delta) = [(\lambda_\alpha+\lambda_\beta)2((n_1+n_2+1)/12)^{\frac{1}{2}}\ \delta^{-1}]^2$$

$$= [(\lambda_\alpha+\lambda_\beta)^2(n_1+n_2+1)(3\ \delta^2)^{-1}] \quad . \tag{2.17}$$

Similar results have been derived for Mood's and Massey's tests, and the likelihood ratio test by Chakravarti, Leone, and Alanen [3]. Their results are summarized below:

For Mood's test procedure based on the median

$$N_2(\delta) = [(\lambda_\alpha+\lambda_\beta)^2(n_1+n_2)(n_2\ \delta^2)^{-1}] \quad . \tag{2.18}$$

For Mood's test procedure based on the first quartile

$$N_3(\delta) = [(\lambda_\alpha+\lambda_\beta)^2(n_1+n_2)(3n_2\ \delta^2)^{-1}] \quad . \tag{2.19}$$

For Massey's test procedure based on the first quartile and median

$$N_4(\delta) = [(n_1+n_2)\ \Delta^2\ (3n_2\ \delta^2)^{-1}] \quad , \tag{2.20}$$

where Δ^2 is a solution of the equation

$$\int_{m_o}^{\infty} f(\chi^2,\ \Delta^2)\ d\chi^2 = 1 - \beta \ , \tag{2.21}$$

and f is the non-central chi-square density function with two degrees of freedom.

For the Likelihood Ratio test procedure

$$N^*(\delta) = D^*/\delta \ , \tag{2.22}$$

where D^* denotes the solution of $H(D^*) = 1-\beta$,

and $H(d) = \lim_{n \to \infty} \beta_n(\theta_o + dn^{-r})$.

It is easily seen that the asymptotic efficiency of all of the above tests relative to the likelihood ratio test is zero, since

$$N^*(\delta)/N_i(\delta) \to 0 \quad \text{as} \quad \delta \to 0 \text{ for } i = 1, 2, 3, 4 \ . \tag{2.23}$$

Likewise, the asymptotic efficiency of the Mann-Whitney test relative to the median test is

$$\text{Eff}(\tau_1, \tau_2) = N_2(\delta)/N_1(\delta) = 3(n_1+n_2)[n_2(n_1+n_2+1)]^{-1} \approx 3/n_2 \ . \tag{2.24}$$

The asymptotic efficiency of the Mann-Whitney test relative to the test based on the first quartile is

$$\text{Eff}(\tau_1, \tau_3) = N_3(\delta)/N_1(\delta) = (n_1+n_2)[n_2(n_1+n_2+1)]^{-1} \approx n_2^{-1} \ . \tag{2.25}$$

Also, the asymptotic efficiency of the Mann-Whitney test relative to Massey's test based on the first quartile and the median is

$$\text{Eff}(\tau_1, \tau_4) = N_4(\delta)/N_1(\delta) = (n_1+n_2)\Delta^2[n_2(n_1+n_2+1)(\lambda_\alpha+\lambda_\beta)^2]^{-1}$$

$$\approx \Delta^2 \ [n_2(\lambda_\alpha+\lambda_\beta)^2]^{-1} \ . \tag{2.26}$$

2.4 Further Extensions.

First, the results given in the above equations can be tabulated for various sample sizes, and a comparison can be made to determine the asymptotically most efficient test procedure among those considered.

Second, the above tests can be compared with the standard t test for detecting a shift in the location parameter, when the usual assumption of normality is made.

CHAPTER III

EXACT POWER OF SOME TESTS BASED ON A GENERALIZATION OF MOOD'S STATISTIC

3.1 Introduction.

In many practical situations, such as life testing, the sample observations arise in order of their magnitude, so that the first observation X_1 is always the smallest, the second observation X_2 is second smallest, and so on. To discriminate between two populations on the basis of two such ordered samples, many rank tests are available. Of them, Mood's test [16] based on the median of the combined samples and Massey's extension of Mood's test [15] based on fractiles, have much to commend themselves as quick tests. The exact power of these tests against the alternatives of exponential and rectangular populations for the case of two populations has been investigated in detail by Chakravarti, Leone, and Alanen [13]. The purpose of this investigation is to extend the results available for Mood's two sample test to the case of discriminating among c populations on the basis of c ordered samples. The corresponding extension of Massey's test is investigated in a subsequent chapter.

3.2 The c Sample Problem.

Let $\left\{ X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)} \right\}$ for $i = 1, 2, \ldots, c$ be c sets of independently distributed random variables with continuous cumulative distribution functions F_i, respectively. We wish to test the hypothesis

$$H_o : F_1(x) = F_2(x) = \dots = F_c(x) ,$$

against the alternative

$$H_a : F_1(x) > F_2(x) > \dots > F_c(x) .$$

Denote the size of the combined sample by $n = \sum_{i=1}^{c} n_i$ and for the sake of simplicity assume that $n = 2r + 1$, where r is an integer. Let $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ be the ordered combined sample. $Z = Z_{(r+1)}$ denotes the median of the combined sample, and u_i denotes the number of observations in the i^{th} sample less than $Z(i=1, 2, \dots, c)$.

Thus, the observations can be arranged to form a 2 by c contingency table as follows:

Number in Sample below and above Median.

Category	1^{st} Sample	2^{nd} Sample	...	c^{th} Sample	Total
Less than Z	u_1	u_2	...	u_c	r
Greater than or equal to Z	$n_1 - u_1$	$n_2 - u_2$		$n_c - u_c$	$r + 1$
Total	n_1	n_2	...	n_c	n

Subject to restrictions

$$\sum_{i=1}^{c} u_i = r ; \qquad \sum_{i=1}^{c} n_i = n \qquad . \tag{3.1}$$

We can define a statistic T by

$$T = \sum_{i=1}^{c} \left\{ (u_i - rn_i/n)^2 (rn_i/n)^{-1} + [n_i - u_i - (r+1)n_i/n]^2 [(r+1)n_i/n]^{-1} \right\}. \tag{3.2}$$

Then a test of size α based on the statistic T is as follows:

reject H_0 if $T \geq t_\alpha$,

accept H_0 if $T < t_\alpha$,

where t_α is defined by $\Pr\{ T \geq t_\alpha \mid H_0 \} \leq \alpha$.

This problem is equivalent to testing a 2 by c contingency table with fixed marginal sums for independence between columns.

3.3 The Null Distribution.

We need to develop an expression for the density function $h(u_1, \ldots, u_c, z)$ of U_i, $(i = 1, 2, \ldots, c)$, and Z. Let us assume that the median Z is from sample j, then the probability $P_j(u_1, \ldots, u_c, z)$ of obtaining these values in the contingency table is given by

$$P_j(u_1, \ldots, u_c, z) = (n_j - u_j) K \left\{ \prod_{i=1}^{c} [(F_i(z))^{u_i} (1 - F_i(z))^{n_i - u_i}] \right\} \cdot [1 - F_j(z)]^{-1} \frac{dF_j(z)}{dz} , \tag{3.3}$$

where $K = \prod_{i=1}^{c} \binom{n_i}{u_i}$.

From Equation (3.3), we obtain

$$h(u_1, \ldots, u_c, z) = \sum_{j=1}^{c} P_j(u_1, \ldots, u_c, z)$$

$$= \sum_{j=1}^{c} (n_j - u_j) K \{ \prod_{i=1}^{c} [(F_i)^{u_i} (1-F_i)^{n_i - u_i}] \} [1-F_j]^{-1} \frac{dF_j}{dz} . \tag{3.4}$$

The null distribution $\varphi_0(u_1, \ldots, u_c)$ of U_i, $(i = 1, 2, \ldots, c)$, under the null hypothesis is derived from $h(u_1, \ldots, u_c, z)$ by substituting

$$F_1(z) = F_2(z) = \ldots = F_c(z) = F(z)$$

in (3.4) and integrating the resulting expression over the range of z. This yields

$$\varphi_0(u_1, \ldots, u_c) = K(r+1) \int_0^1 F^r (1-F)^r \, dF$$

$$= \left[\prod_{i=1}^{c} \binom{n_i}{u_i} \right] \binom{n}{r}^{-1} . \tag{3.5}$$

This result is in agreement with that obtained by Chakravarti, Leone, and Alanen [13].

3.4 Power Function of Median T Test Against the Alternative of Translation in the Exponential Population.

Here the alternative hypothesis considered is:

$$H_a \begin{cases} F_i(x) = 1 - e^{-(x-a_i)}, & x \geq a_i , \\ \quad\;\; = 0 & x < a_i , \\ \text{where } a_{i+1} \geq a_i \quad (i = 1, 2, \ldots, c-1), & a_1 > 0 . \end{cases} \tag{3.6}$$

The only necessary requirement is that $a_c > a_i$ $(i = 1, 2, \ldots, c-1)$.

The joint distribution of the U_i's is obtained by substituting (3.6) in the expression for the joint density function $h(u_1, \ldots, u_c, z)$ given by (3.4) and integrating over the range of z. The range of z can be reduced to $a_1 \leq z < \infty$, since all of the F_i's are zero for $z < a_1$. This gives us

$$\Pr\{U_i = u_i \mid i = 1, 2, \ldots, c\} = \varphi_a(u_1, \ldots, u_c)$$

$$= \sum_{j=1}^{c} (n_j - u_j) K \int_{z=a_1}^{\infty} \{ \prod_{i=1}^{c} [F_i(z)^{u_i} (1-F_i(z))^{n_i-u_i}] \} [1-F_j(z)]^{-1} dF_j(z). \tag{3.7}$$

Now, we can write

$$\int_{a_1}^{\infty} = \int_{a_1}^{a_2} + \int_{a_2}^{a_3} + \ldots + \int_{a_c}^{a_{c+1}=\infty}$$

and obtain

$$\varphi_a(u_1, \ldots, u_c) = \sum_{t=1}^{c} S_t(u) \quad , \tag{3.8}$$

where

$$S_t(u) = \sum_{j=1}^{c} (n_j-u_j) K \int_{z=a_t}^{a_{t+1}} \{ \prod_{i=1}^{c} [F_i^{u_i}(1-F_i)^{n_i-u_i}]\}[1-F_j]^{-1} \, dF_j \quad . \tag{3.9}$$

Since $F_i = 0$ for $t+1 \leq i \leq c$, $S_t(u)$ will be zero unless $u_i = 0$. Also, all terms for $j > t$ will be zero, since $F_j = 0$. Thus, if we define $u_{c+1} \equiv 0$, (3.9) reduces to

$$S_t(u) = \sum_{j=1}^{t} (n_j-u_j) K (\prod_{i=t}^{c} \delta_{u_{i+1},0}) \int_{z=a_t}^{a_{t+1}} \{ \prod_{i=1}^{t} [F_i^{u_i}(1-F_i)^{n_i-u_i}]\} \cdot$$

$$[1-F_j]^{-1} \, dF_j \quad , \tag{3.10}$$

where $\delta_{i,j}$ is the Kronecker delta.

Let

$$\eta_{i,t} = e^{-(a_t-a_i)}$$

$$\gamma_{i,t} = 1 - \eta_{i,t} \quad . \tag{3.11}$$

Then we can substitute

$$F_i = 1 - \eta_{i,t}(1-F_t) = \gamma_{i,t} + \eta_{i,t}F_t, \text{ for } a_t \le z \le a_{t+1},\ 1 \le i \le t, \tag{3.12}$$

in (3.10). This gives us

$$S_t(u) = \sum_{j=1}^{t}(n_j-u_j)K\left(\prod_{i=t}^{c}\delta_{u_{i+1},0}\right)\int_0^{F_t(a_{t+1})}\left\{\prod_{i=1}^{t}[1-\eta_{i,t}(1-F_t)]^{u_i}\right\}\cdot$$

$$\left\{\prod_{i=1}^{t}\eta_{i,t}^{n_i-u_i}\right\}[1-F_t]^{\sum_{i=1}^{t}(n_i-u_i)-1}\, dF_t \quad . \tag{3.13}$$

Expanding $[1-\eta_{i,t}(1-F_t)]^{u_i}$ by the binomial theorem, we get:

$$S_t(u) = K\left[\sum_{j=1}^{t}(n_j-u_j)\right]\left[\prod_{i=t}^{c}\delta_{u_{i+1},0}\right]\sum_{q_1=0}^{u_1}\cdots\sum_{q_t=0}^{u_t}\left\{\prod_{i=1}^{t}[(-1)^{q_i}\cdot\right.$$

$$\left.\left(\eta_{i,t}^{n_i-u_i+q_i}\right)\binom{u_i}{q_i}]\right\}\int_0^{F_t(a_{t+1})}(1-F_t)^{\sum_{i=1}^{t}(n_i-u_i+q_i)-1}\, dF_t. \tag{3.14}$$

This expression can be integrated to yield:

$$S_t(u) = K\left[\sum_{j=1}^{t}(n_j-u_j)\right]\left[\prod_{i=t}^{c}\delta_{u_{i+1},0}\right]\sum_{q_1=0}^{u_1}\cdots\sum_{q_t=0}^{u_t}\left\{\prod_{i=1}^{t}[(-1)^{q_i}\cdot\right.$$

$$\left.\binom{n_i-u_i+q_i}{\eta_{i,t}}\binom{u_i}{q_i}]\right\}\left[1-\eta_{t,t+1}^{\sum_{i=1}^{t}(n_i-u_i+q_i)}\right]\left[\sum_{i=1}^{t}(n_i-u_i+q_i)\right]^{-1}.$$

(3.15)

For the case in which $t = 1$, (3.10) reduces to

$$S_1(u) = (n_1-u_1)K\left[\prod_{i=1}^{c}\delta_{u_{i+1},0}\right]\int_{z=a_1}^{a_2} F_1^{u_1}(1-F_1)^{n_1-u_1-1}\,dF_1 \quad , \qquad (3.16)$$

which can be integrated to yield

$$S_1(u) = K\left[\prod_{i=1}^{c}\delta_{u_{i+1},0}\right]\sum_{q_1=0}^{n_1-u_1}(-1)^{q_1}\binom{n_1-u_1}{q_1}(n_1-u_1-q_1)\gamma_{1,2}^{u_1+q_1+1}[u_1+q_1+1]^{-1}.$$

(3.17)

Consider the special case in which $a_1 = a_2 = \ldots = a_{c-1} = 0$ and $a_c = a > 0$. Let

$$\eta_{1,c} = \eta_{2,c} = \ldots = \eta_{c-1,c} = \eta = e^{-a}, \quad F_1 = F_2 = \ldots = F_{c-1} = F, \quad F_c = G,$$

$$\bar{U}_c = \sum_{i=1}^{c-1} u_i \; , \qquad \bar{N}_c = \sum_{i=1}^{c-1} n_i \; , \qquad (3.18)$$

then (3.7) becomes

$$\varphi_a(\bar{U}_c,u_c) = (\bar{N}_c-\bar{U}_c)K\int_{z=0}^{\infty} F^{\bar{U}_c}(1-F)^{\bar{N}_c-\bar{U}_c-1}G^{u_c}(1-G)^{n_c-u_c}\,dF +$$

$$(n_c-u_c)K\int_{z=0}^{\infty} F^{\bar{U}_c}(1-F)^{\bar{N}_c-\bar{U}_c}G^{u_c}(1-G)^{n_c-u_c-1}\,dG \quad . \tag{3.19}$$

This can be written as

$$\varphi_a(\bar{U}_c,u_c) = (\bar{N}_c-r)K\,\delta_{u_c,0}\int_{z=0}^{a} F^{r}(1-F)^{\bar{N}_c-r-1}\,dF +$$

$$(\bar{N}_c-\bar{U}_c)K\int_{z=a}^{\infty} F^{\bar{U}_c}(1-F)^{\bar{N}_c-\bar{U}_c-1}G^{u_c}(1-G)^{n_c-u_c}\,dF +$$

$$(n_c-u_c)K\int_{z=a}^{\infty} F^{\bar{U}_c}(1-F)^{\bar{N}_c-\bar{U}_c}G^{u_c}(1-G)^{n_c-u_c-1}\,dG \quad . \tag{3.20}$$

Substitute $F = 1 - \eta(1 - G)$ for $a \le z < \infty$, in (3.20) and obtain

$$\varphi_a(\bar{U}_c,u_c) = K\,\delta_{u_c,0}\sum_{q_1=0}^{\bar{N}_c-r}(-1)^{q_1}\binom{\bar{N}_c-r}{q_1}(\bar{N}_c-r-q_1)\int_0^{\gamma} F^{r+q_1}\,dF +$$

$$K(r+1)\sum_{q_1=0}^{\bar{U}_c}\sum_{q_2=0}^{r+q_1}(-1)^{q_1+q_2}\binom{\bar{U}_c}{q_1}\binom{r+q_1}{q_2}\int_0^{1} G^{u_c+q_2}\,dG \quad , \tag{3.21}$$

which upon integration yields

$$\varphi_a(\bar{U}_c,u_c) = K\,\delta_{u_c,0} \sum_{q_1=0}^{\bar{N}_c-r} (-1)^{q_1} \binom{\bar{N}_c-r}{q_1} (\bar{N}_c-r-q_1)\gamma^{r+q_1+1} [r+q_1+1]^{-1} +$$

$$K(r+1) \sum_{q_1=0}^{\bar{U}_c} \sum_{q_2=0}^{r+q_1} (-1)^{q_1+q_2} \binom{\bar{U}_c}{q_1} \binom{r+q_1}{q_2} [u_c+q_2+1]^{-1} \qquad (3.22)$$

If $c = 2$, this result agrees with that obtained by Leone, Chakravarti, and Alanen [13].

These results can be used to compute the power of the test for the case of the exponential alternative by first defining t_α by

$$\Pr\{ T \geq t_\alpha \mid H_o \} \geq \alpha \quad ,$$

where the probability is evaluated using (3.5). The power is calculated from $\Pr\{ T \geq t_\alpha \mid H_a \} = \sum_{u_i} \varphi_a(u_1, \ldots, u_c)$, such that $T \geq t_\alpha$.

3.5 Power Function of the Median T Test Against Alternatives of Change in Location and Scale of the Rectangular Population.

In this case, two sets of alternative hypotheses will be considered, namely: one in which the location parameter changes and another in which the scale parameter varies.

3.5.1 Change in Location of the Rectangular Population.

The alternative hypothesis is

$$H_a \begin{cases} F_i(x) = x - a_i & \text{for} \quad a_i \le x \le 1 + a_i\,, \\ = 0 & \text{for} \quad x < a_i\,, \\ = 1 & \text{for} \quad x > 1 + a_i\,, \end{cases} \tag{3.23}$$

where again the populations are ordered according to their location parameter:

$$0 < a_1 < a_2 < \ldots < a_c \quad .$$

The joint distribution of the U_i's is obtained by substituting (3.23) for F_i in the expression for the joint density function $h(u_1, \ldots, u_c, z)$ given in (3.4) and integrating with respect to z. The range of z can be reduced to $a_1 \le z \le 1 + a_c$, since all of the F_i's are zero for $z < a_1$ and one for $z > 1 + a_c$. This yields

$$\Pr\{ U_i = u_i \mid i = 1, 2, \ldots, c \} = \varphi_a(u_1, \ldots, u_c)$$

$$= \sum_{j=1}^{c}(n_j - u_j) K \int_{z=a_1}^{1+a_c} \{ \prod_{i=1}^{c} [F_i(z)^{u_i}(1-F_i(z))^{n_i-u_i}]\}[1-F_j(z)]^{-1} dF_j(z). \tag{3.24}$$

The method used to evaluate the integral in (3.24) depends upon the relative sizes of the a_i's. In this section we will consider only one situation, that is perhaps the most interesting. However, any other possible situations can be handled in an analogous fashion.

For the developments in this section we will assume that

$$a_c < a_1 + 1 \quad . \tag{3.25}$$

Then the integral in (3.24) can be partitioned by dividing the range of z into $2c-1$ pieces. This yields

$$\varphi_a(u_1, \ldots, u_c) = \sum_{t=1}^{c-1} \{ P_t(u) + R_t(u) \} + Q_c(u) \quad , \tag{3.26}$$

where

$$P_t(u) = \sum_{j=1}^{c} (n_j - u_j) K \int_{z=a_t}^{a_{t+1}} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j(z) \quad , \tag{3.27}$$

$$R_t(u) = \sum_{j=1}^{c} (n_j - u_j) K \int_{z=1+a_t}^{1+a_{t+1}} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j(z) \quad , \tag{3.28}$$

$$Q_c(u) = \sum_{j=1}^{c} (n_j - u_j) K \int_{z=a_c}^{1+a_1} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j(z) \quad , \tag{3.29}$$

Consider $P_t(u)$. Since $F_i = 0$ for $1 + t \le i \le c$, $P_t(u)$ will be zero unless $u_i = 0$. Further, all terms for $j > t$ will be zero. Thus for $1 < t < c$, (3.27) reduces to

$$P_t(u) = K\left(\prod_{i=t+1}^{c} \delta_{u_i,0}\right) \sum_{j=1}^{t} (n_j-u_j) \int_{z=a_t}^{a_{t+1}} \left\{ \prod_{i=1}^{t} \left[F_i^{u_i} (1-F_i)^{n_i-u_i} \right] \right\} \cdot$$

$$[1-F_j]^{-1} \, dF_j \quad . \qquad (3.30)$$

Let

$$\mu_{i,j} = a_j - a_i \;, \qquad \nu_{i,j} = 1-\mu_{i,j} \;, \qquad (3.31)$$

then we can substitute

$$F_i = \mu_{i,t} + F_t \;, \quad 1-F_i = \nu_{i,t} - F_t \;, \quad \text{for } a_t \le z \le a_{t+1} \;, \quad 1 \le i < t \;, \qquad (3.32)$$

in (3.30) and expand the resulting binomials. This give us

$$P_t(u) = K\left(\prod_{i=t+1}^{c} \delta_{u_i,0}\right) \sum_{j=1}^{t} (n_j-u_j) \sum_{q_1=0}^{u_1} \cdots \sum_{q_{t-1}=0}^{u_{t-1}} \sum_{v_1=0}^{n_1-u_1} \cdots \sum_{v_{j-1}=0}^{n_{j-1}-u_{j-1}}$$

$$\sum_{v_j=0}^{n_j-u_j-1} \sum_{v_{j+1}=0}^{n_{j+1}-u_{j+1}} \cdots \sum_{v_t=0}^{n_t-u_t} (-1)^{\sum_{i=1}^{t} v_i} \binom{u_1}{q_1} \cdots \binom{u_{t-1}}{q_{t-1}} \binom{n_1-u_1}{v_1} \cdots$$

$$\binom{n_{j-1}-u_{j-1}}{v_{j-1}} \binom{n_j-u_j-1}{v_j} \binom{n_{j+1}-u_{j+1}}{v_{j+1}} \cdots \binom{n_t-u_t}{v_t} \nu_{j,t}^{-1} \cdot$$

$$\left\{ \prod_{i=1}^{t-1} \left[\mu_{i,t}^{u_i-q_i} \; \nu_{i,t}^{n_i-u_i-v_i} \right] \right\} \int_0^{F_t(a_{t+1})} F_t^{\,u_t + \sum_{i=1}^{t-1}(q_i+v_i) + v_t} \, dF_t \;. \qquad (3.33)$$

By noting that

$$(n_j-u_j)\binom{n_j-u_j-1}{v_j} = (n_j-u_j-v_j)\binom{n_j-u_j}{v_j}, \text{ and } F_i(a_j)=\mu_{i,j} ,$$

(3.33) can be integrated and simplified. This yields

$$P_t(u) = K\left(\prod_{i=t+1}^{c} \delta_{u_i,0}\right) \sum_{q_1=0}^{u_1} \cdots \sum_{q_{t-1}=0}^{u_{t-1}} \sum_{v_1=0}^{n_1-u_1} \cdots \sum_{v_t=0}^{n_t-u_t} (-1)^{\sum_{i=1}^{t} v_i}$$

$$\left\{\prod_{i=1}^{t-1}\left[\binom{u_i}{q_i}\binom{n_i-u_i}{v_i}\right]\right\} \binom{n_t-u_t}{v_t}\left\{\prod_{i=1}^{t-1}\left[\mu_{i,t}^{u_i-q_i}\ \nu_{i,t}^{n_i-u_i-v_i}\right]\right\} .$$

$$\left\{\sum_{j=1}^{t}[(n_j-u_j-v_j)/\nu_{j,t}]\right\}\mu_{t,t+1}^{u_t+\sum_{i=1}^{t-1}(q_i+v_i)+v_t+1} \left[u_t+\sum_{i=1}^{t-1}(q_i+v_i)+v_t+1\right]^{-1}, \tag{3.34}$$

for $t = 2, 3, \ldots, c-1$. $P_1(u)$ is easy to handle and is given by

$$P_1(u) = K\left(\prod_{i=2}^{c} \delta_{u_i,0}\right) \sum_{v=0}^{n_1-r} \binom{n_1-r}{v}(-1)^v(n_1-r-v)\mu_{1,2}^{r+v+1}[r+v+1]^{-1} . \tag{3.35}$$

Similarly (3.28) can be simplified if we note that $F_i = 1$ for $i \le t$ and hence (3.28) will be zero unless $(n_i-u_i) = 0$ for $i \le t$. Also, all terms for $j \le t$ will be zero, and for $1 \le t \le c-2$, (3.28) will reduce to

$$R_t(u) = K\left(\prod_{i=1}^{t} \delta_{n_i-u_i,0}\right) \sum_{j=t+1}^{c} (n_j-u_j) \int_{z=1+a_t}^{1+a_{t+1}} \left\{ \prod_{i=t+1}^{c} [F_i^{u_i}(1-F_i)^{n_i-u_i}] \right\} \cdot$$

$$[1-F_j]^{-1}\, dF_j \quad . \tag{3.36}$$

If we substitute

$$F_i = \mu_{i,t+1} + F_{t+1} \ , \quad 1-F_i = \nu_{i,t+1} - F_{t+1} \ , \tag{3.37}$$

for $1+a_t \leq z \leq 1+a_{t+1}$, $t+1 < i \leq c$, in (3.36) and expand the binomials, we get:

$$R_t(u) = K\left(\prod_{i=1}^{t} \delta_{n_i-u_i,0}\right) \sum_{q_{t+2}=0}^{u_{t+2}} \cdots \sum_{q_c=0}^{u_c} \sum_{v_{t+1}=0}^{n_{t+1}-u_{t+1}} \cdots \sum_{v_c=0}^{n_c-u_c} (-1)^{\sum_{i=t+1}^{c} v_i} \cdot$$

$$\left[\prod_{i=t+2}^{c} \binom{u_i}{q_i}\binom{n_i-u_i}{v_i}\right] \binom{n_{t+1}-u_{t+1}}{v_{t+1}} \left\{ \prod_{i=t+2}^{c} \left[\mu_{i,t+1}^{u_i-q_i}\, \nu_{i,t+1}^{n_i-u_i-v_i}\right] \right\} \cdot$$

$$\left[\sum_{j=t+1}^{c} (n_j-u_j-v_j)/\nu_{j,t+1}\right] \int_{\nu_{t,t+1}}^{1} F_{t+1}^{u_{t+1}+\sum_{i=t+2}^{c}(q_i+v_i)+v_{t+1}}\, dF_{t+1} \quad , \tag{3.38}$$

since $\nu_{i,j} = F_j(1+a_i)$. This expression can be integrated to yield

$$R_t(u) = K\left(\prod_{i=1}^{t} \delta_{n_i-u_i,0}\right) \sum_{q_{t+2}=0}^{u_{t+2}} \cdots \sum_{q_c=0}^{u_c} \sum_{v_{t+1}=0}^{n_{t+1}-u_{t+1}} \cdots \sum_{v_c=0}^{n_c-u_c} (-1)^{\sum_{i=t+1}^{c} v_i} \cdot$$

$$\left[\prod_{i=t+2}^{c} \binom{u_i}{q_i}\binom{n_i-u_i}{v_i}\right]\binom{n_{t+1}-u_{t+1}}{v_{t+1}}\left\{\prod_{i=t+2}^{c}\begin{bmatrix} u_i-q_i & n_i-u_i-v_i \\ \mu_{i,t+1} & \nu_{i,t+1}\end{bmatrix}\right\}.$$

$$\left[\sum_{j=t+1}^{c}(n_j-u_j-v_j)/\nu_{j,t+1}\right]\left[1-\nu_{t,t+1}^{\,u_{t+1}+\sum_{i=t+2}^{c}(q_i+v_i)+v_{t+1}+1}\right][u_{t+1}+\sum_{i=t+2}^{c}(q_i+v_i)+$$

$$v_{t+1}+1]^{-1} \quad . \qquad (3.39)$$

Also

$$R_{c-1}(u) = K\left(\prod_{i=1}^{c-1}\delta_{n_i-u_i,0}\right)\sum_{v_c=0}^{r+1}(-1)^{v_c}\binom{r+1}{v_c}\left[1-\nu_{c-1,c}^{\,u_c+v_c+1}\right][u_c+v_c+1]^{-1} \cdot$$

$$[r+1-v_c] \quad . \qquad (3.40)$$

In a similar fashion using (3.31) and (3.32), (3.29) can be integrated and simplified to give:

$$Q_c(u) = K \sum_{q_1=0}^{u_1} \cdots \sum_{q_{c-1}=0}^{u_{c-1}} \sum_{v_1=0}^{n_1-u_1} \cdots \sum_{v_c=0}^{n_c-u_c} (-1)^{\sum_{i=1}^{c} v_i} \left[\prod_{i=1}^{c-1} \binom{u_i}{q_i} \binom{n_i-u_i}{v_i} \right] \cdot$$

$$\binom{n_c-u_c}{v_c} \left[\prod_{i=1}^{c-1} \mu_{i,c}^{u_i-q_i} \nu_{i,c}^{n_i-u_i-v_i} \right] \left[\nu_{1,c}^{u_c+\sum_{i=1}^{c-1}(q_i+v_i)+v_c+1} \right] \cdot$$

$$\left[u_c + \sum_{i=1}^{c-1} (q_i+v_i) + v_c + 1 \right]^{-1} \left[\sum_{j=1}^{c} (n_j-u_j-v_j)/\nu_{j,c} \right] . \qquad (3.41)$$

If we consider the special case in which $a_1 = \ldots = a_{c-1} = 0$ and $a_c = a$, where $0 < a < 1$, and let

$$\mu_{1,c} = \mu_{2,c} = \ldots = \mu_{c-1,c} = \mu = a \quad ,$$

$$F_1 = F_2 = \ldots = F_{c-1} = F \ , \quad F_c = G \ ,$$

$$\bar{U}_c = \sum_{i=1}^{c-1} u_i \ , \quad \bar{N}_c = \sum_{i=1}^{c-1} n_i \quad , \qquad (3.42)$$

then (3.24) can be written in the form:

$$\varphi_a(\overline{U}_c,u_c) = K(\overline{N}_c-r)\delta_{u_c,0}\int_{z=0}^{a} F^r\,(1-F)^{\overline{N}_c-r-1}\,dF +$$

$$K(\overline{N}_c-\overline{U}_c)\int_{z=a}^{1} F^{\overline{U}_c}(1-F)^{\overline{N}_c-\overline{U}_c-1}\,G^{u_c}(1-G)^{n_c-u_c}\,dF +$$

$$K(n_c-u_c)\int_{z=a}^{1} F^{\overline{U}_c}(1-F)^{\overline{N}_c-\overline{U}_c}\,G^{u_c}(1-G)^{n_c-u_c-1}\,dG +$$

$$K(n_c-u_c)\,\delta_{\overline{N}_c-\overline{U}_c,0}\int_{z=1}^{1+a} G^{u_c}(1-G)^{n_c-u_c-1}\,dG \quad . \tag{3.43}$$

Now following a procedure analogous to that used in the general case in which we substitute for F, expand and integrate, we get:

$$\varphi_a(\bar{U}_c,u_c) = K\delta_{u_c,0}\sum_{v_1=0}^{\bar{N}_c-r}(-1)^{v_1}\binom{\bar{N}_c-r}{v_1}a^{r+v_1+1}[\bar{N}_c-r-v_1][r+v_1+1]^{-1} +$$

$$K\sum_{q_1=0}^{\bar{U}_c}\sum_{v_1=0}^{\bar{N}_c-\bar{U}_c}\sum_{v_2=0}^{n_c-u_c}(-1)^{v_1+v_2}\binom{\bar{U}_c}{q_1}\binom{\bar{N}_c-\bar{U}_c}{v_1}\binom{n_c-u_c}{v_2}a^{\bar{U}_c-q_1} \cdot$$

$$(1-a)^{\bar{N}_c-\bar{U}_c+u_c+q_1+v_2+1}[u_c+q_1+v_1+v_2+1]^{-1}[(\bar{N}_c-\bar{U}_c-v_1)/(1-a) +$$

$$(n_c-u_c-v_2)] + K\delta_{\bar{N}_c-\bar{U}_c,0}\sum_{v_1=0}^{n_c-u_c}(-1)^{v_1}\binom{n_c-u_c}{v_1}[n_c-u_c-v_1] \cdot$$

$$[1-(1-a)^{u_c+v_1+1}][u_c+v_1+1]^{-1} \quad . \qquad (3.44)$$

3.5.2 <u>Change in Scale of the Rectangular Population</u>.

The alternative hypothesis is

$$H_\theta \begin{cases} F_i(x) = x/\theta_i \ , & 0 \le x \le \theta_i \ , \\ \quad\ \ = 0 \ , & x < 0 \ , \\ \quad\ \ = 1 \ , & x > \theta_i \ , \end{cases} \qquad (3.45)$$

where the scale parameters are ordered as follows:

$$0 < \theta_1 < \theta_2 < \dots < \theta_c .$$

The joint distribution of the U_i's is obtained by substituting (3.45) in the expression for $h(u_1, \dots, u_c, z)$ given by (3.4) and integrating over the range of z. The actual range in this case is $0 \leq z \leq \theta_c$, since all of the F_i's are zero for $z < 0$ and one for $z > \theta_c$. Thus we obtain:

$$\Pr\{ U_i = u_i \mid i = 1, 2, \dots, c \} = \varphi_\theta(u_1, \dots, u_c)$$

$$= \sum_{j=1}^{c} (n_j - u_j) K \int_{z=0}^{\theta_c} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j \quad . \tag{3.46}$$

The range of integration on z can be broken into c pieces. This will yield c integrals, and (3.46) can be written as:

$$\varphi_\theta(u_1, \dots, u_c) = Q_c(u) + \sum_{t=1}^{c-1} R_t(u) \quad , \tag{3.47}$$

where

$$Q_c(u) = \sum_{j=1}^{c} (n_j - u_j) K \int_{z=0}^{\theta_1} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j \quad , \tag{3.48}$$

$$R_t(u) = \sum_{j=1}^{c} (n_j - u_j) K \int_{z=\theta_t}^{\theta_{t+1}} \left\{ \prod_{i=1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i - u_i} \right] \right\} [1-F_j]^{-1} \, dF_j \quad . \tag{3.49}$$

If we let

$$\epsilon_{i,j} = \theta_j/\theta_i \ , \tag{3.50}$$

then (3.48) can be simplified if we substitute

$$F_i = \epsilon_{i,1} F_1 \ , \quad \text{for} \quad 0 \le z \le \theta_1 \ , \quad 1 \le i \le c \ , \tag{3.51}$$

in (3.48), expand the binomials and integrate. This yields

$$Q_c(u) = K \sum_{v_1=0}^{n_1-u_1} \cdots \sum_{v_c=0}^{n_c-u_c} (-1)^{\sum_{i=1}^{c} v_i} \left[\prod_{i=1}^{c} \binom{n_i-u_i}{v_i} \epsilon_{i,1}^{u_i+v_i} \right] \cdot$$

$$\left\{ \sum_{j=1}^{c} [(n_j-u_j-v_j)\epsilon_{j,1}] \right\} [r + \sum_{i=1}^{c} v_i + 1]^{-1} \quad . \tag{3.52}$$

Following a similar development, (3.49) can be simplified, if we note that (3.49) will be zero unless $(n_i-u_i) = 0$, $1 \le i \le t$ since $F_i = 1$; $i = 1, 2, \ldots, t$. Thus for $1 \le t < c$, (3.49) reduces to

$$R_t(u) = \sum_{j=t+1}^{c} (n_j-u_j) K \left(\prod_{i=1}^{t} \delta_{n_i-u_i,0} \right) \int_{z=\theta_t}^{\theta_{t+1}} \left\{ \prod_{i=t+1}^{c} \left[F_i^{u_i} (1-F_i)^{n_i-u_i} \right] \right\} \cdot$$

$$[1-F_j]^{-1} \, dF_j \quad . \tag{3.53}$$

Again we can substitute

$$F_i = \epsilon_{i,t+1} F_{t+1} \ , \quad \theta_t \leq z \leq \theta_{t+1} \ , \quad t+1 \leq i \leq c \ , \qquad (3.54)$$

into (3.53), expand and integrate to yield:

$$R_t(u) = K\left(\prod_{i=1}^{t} \delta_{n_i-u_i,0}\right) \sum_{v_{t+1}=0}^{n_{t+1}-u_{t+1}} \cdots \sum_{v_c=0}^{n_c-u_c} (-1)^{\sum_{i=1}^{c} v_i} \left[\prod_{i=t+1}^{c} \binom{n_i-u_i}{v_i} \epsilon_{i,t+1}^{u_i+v_i}\right] \cdot$$

$$\left\{\sum_{j=t+1}^{c} [(n_j-u_j-v_j)\epsilon_{j,t+1}]\right\} \left[1-\epsilon_{t+1,t}^{\sum_{i=t+1}^{c}(u_i+v_i)+1}\right] \left[\sum_{i=t+1}^{c} (u_i+v_i)+1\right]^{-1} , \qquad (3.55)$$

where $F_{t+1}(\theta_t) = \epsilon_{t+1,t}$.

If we consider the special case in which

$$\theta_1 = \theta_2 = \ldots = \theta_{c-1} = 1 \quad \text{and} \quad \theta_c = \theta > 1 \ , \qquad (3.56)$$

and let

$$F_1 = F_2 = \ldots = F_{c-1} = F \ , \quad F_c = G \ ,$$

$$\bar{U}_c = \sum_{i=1}^{c-1} u_i \ , \qquad \bar{N}_c = \sum_{i=1}^{c-1} n_i \ , \qquad (3.57)$$

then (3.47) reduces to

$$\varphi_\theta(\bar{U}_c,u_c) = K \sum_{v_1=0}^{\bar{N}_c-\bar{U}_c} \sum_{v_2=0}^{n_c-u_c} (-1)^{v_1+v_2} \binom{\bar{N}_c-\bar{U}_c}{v_1} \binom{n_c-u_c}{v_2} (1/\theta)^{u_c+v_2} \cdot$$

$$[(\bar{N}_c-\bar{U}_c-v_1)+(n_c-u_c-v_2)/\theta][r+v_1+v_2+1]^{-1} +$$

$$K\delta_{\bar{N}_c-\bar{U}_c,0} \sum_{v_1=0}^{n_c-u_c} (-1)^{v_1} \binom{n_c-u_c}{v_1} (n_c-u_c-v_1)[1-(1/\theta)^{u_c+v_1+1}] \cdot$$

$$[u_c+v_1+1]^{-1} , \qquad (3.58)$$

using techniques similar to those used in the general case.

Again these results can be used to compute the power of the test by first defining t_α by

$$\Pr\{ T \geq t_\alpha \mid H_o \} \leq \alpha ,$$

then the power is computed from $\Pr \{ T \geq t_\alpha \mid H_a \} =$

$\sum_{u_i} \varphi_a(u_1, \ldots, u_c)$ such that $T \geq t_\alpha$.

3.6 Results.

The exact power of the T statistic has been computed for the case of three samples using various exponential alternatives. These results are presented in Table 3.1 for total sample sizes of 11 and 15. Since the computations were performed in single precision arithmetic, the values computed for the sample of size 15 may have errors as large as $\pm$ 2 in the third decimal place as indicated by the cumulative sum.

The power of the test, in general, increases with a positive shift in the locations, especially when the test is unbiased. However, several cases can be noted in which this trend fails to occur indicating that in these cases the test is biased. (For example, see Table 3.1 for $n = 11$, $n_1 = 5$, $n_2 = 4$, and $\alpha = 0.0476$.) This effect seems to occur frequently when α is very small. An analysis of the complete distribution of the T statistic indicates that the actual distribution becomes highly "peaked" in addition to shifting in the positive direction when the location parameters increase. This means that the actual tail area for small α's can decrease even though the distribution is shifting in the positive direction.

3.7 Further Extensions.

The computational results presented in Table 3.1 can be extended to larger sample sizes and to other combinations of n_1, n_2, and n_3. The accuracy of the results can also be improved by performing all of the computations in double precision arithmetic. Similar tables can be computed with rectangular alternatives. The results can also be extended to more than three samples when more efficient computational equipment is available.

Similar expressions for the power of the test can be derived for a set of exponential alternatives with a change in scale. Also, extensions similar to those indicated in Chapter I may be applied to this test.

TABLE 3.1

Exact Power of Mood's Test for Three Samples with Exponential Alternatives: $Pr\{ T \geq t \mid H \}$.

n	n_1	n_2	t	α	$a_1 = 0$, $a_2 = .1$, $a_3 = .2$	$a_1 = 0$, $a_2 = .2$, $a_3 = .5$	$a_1 = 0$, $a_2 = .5$, $a_3 = 1.0$	$a_1 = 0$, $a_2 = .1$, $a_3 = 1.0$	$a_1 = 0$, $a_2 = 1.0$, $a_3 = 2.0$	$a_1 = 0$, $a_2 = .5$, $a_3 = 2.0$	$a_1 = 0$, $a_2 = .1$, $a_3 = 2.0$
11	2	4	8.983	.0065	.0082	.0186	.0343	.1109	.0666	.1574	.2574
			7.975	.0152	.0260	.0729	.1890	.3692	.5511	.7476	.8410
			6.160	.0476	.0593	.1020	.2088	.3948	.5566	.7542	.8492
			4.748	.1126	.0978	.1164	.2126	.3958	.5568	.7542	.8492
	4	3	8.311	.0130	.0181	.0471	.1577	.1068	.4267	.2755	.1611
			6.967	.0563	.0660	.1174	.2501	.2736	.4946	.4110	.4046
			5.286	.1082	.1284	.2446	.5103	.6369	.8900	.9409	.9628
	5	4	8.983	.0065	.0073	.0129	.0492	.0115	.1885	.0532	.0118
			7.975	.0152	.0114*	.0142*	.0493	.0116*	.1886	.0532	.0118*
			6.160	.0476	.0397*	.0337*	.0572	.0380*	.1897	.0602	.0404*
			4.748	.1126	.1448	.2077	.3985	.2230	.6876	.4297	.2365
15	3	5	11.786	.0014	.0032	.0153	.0687	.1886	.3753	.6054	.7319
			8.878	.0112	.0127	.0284	.0807	.2266	.3825	.6204	.7528
			5.663	.0590	.0806	.1677	.3574	.5707	.7702	.9032	.9509
			4.898	.1133	.1388	.2207	.4001	.5863	.7816	.9052	.9516
	5	5	10.179	.0093	.0118	.0280	.1043	.1002	.3544	.2254	.1654
			8.571	.0210	.0257	.0503	.1430	.1152	.3749	.2289	.1665
			6.964	.0675	.0798	.1620	.4192	.5997	.8814	.9512	.9758
	7	5	11.786	.0014	.0009*	.0018	.0129	.0019	.0877	.0146	.0021
			8.878	.0112	.0124	.0154	.0245	.0194	.0891	.0183	.0205
			5.663	.0590	.0539*	.0664	.1766	.0741	.4668	.1984	.0814
			4.898	.1133	.1047*	.1246	.2203	.2302	.4802	.2649	.2627

* These values indicate that the test is biased.

CHAPTER IV

EXACT POWER OF SOME TESTS BASED ON A GENERALIZATION OF MASSEY'S STATISTIC

4.1 Introduction.

The exact power of some tests based on Massey's statistic for the case of two samples has been investigated by Chakravarti, Leone, and Alanen [13]. The purpose of the investigation in this chapter is to extend their results to the case of discriminating between c populations on the basis of c ordered samples.

4.2 The c Sample Problem.

Let $\{X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}\}$ for $i = 1, 2, \ldots, c$ be c sets of independently distributed random variables with continuous cumulative distribution functions F_i, respectively. We wish to test the hypothesis

$$H_o : F_1(x) = F_2(x) = \ldots = F_c(x) \quad ,$$

against the alternative

$$H_a : F_1(x) > F_2(x) > \ldots > F_c(x) \quad .$$

Let n_i denote the size of the i^{th} sample, and $n = \sum_{i=1}^{c} n_i$, the size of the combined sample. For simplicity, we will assume that $n = 4r + 1$, where r is an integer. Also let Z_1 and Z_2 denote respectively the first quartile and median of the combined sample. Let $U_{1,i}$ and $U_{2,i}$ denote respectively the number of observations in the i^{th} sample less than Z_1 and the number of observations in the i^{th} sample that are greater than or equal to Z_1 but less than Z_2, $i = 1, 2, \ldots, c$. The results for the combined sample can be arranged in a 3 by c contingency table showing U_{ti}, the number of values from the i^{th} sample in the t^{th} interval ($i = 1, 2, \ldots, c$; $t = 1, 2, 3$). This table will have the following form:

Number of observations less than the first quartile and between the first quartile and the median.

Intervals	1^{st} Sample	2^{nd} Sample	...	c^{th} Sample	Total
1. $x < Z_1$	$u_{1,1}$	$u_{1,2}$	...	$u_{1,c}$	$S_1 = r$
2. $Z_1 \leq x < Z_2$	$u_{2,1}$	$u_{2,2}$	...	$u_{2,c}$	$S_2 = r$
3. $x \geq Z_2$	$u_{3,1}$	$u_{3,2}$	...	$u_{3,c}$	$S_3 = 2r + 1$
Total	n_1	n_2	...	n_c	n

where

$$u_{3,i} = n_i - u_{1,i} - u_{2,i}\ , \quad i = 1, 2, \ldots, c\ ,$$

To test the null hypothesis

$$H_o : F_1(x) = F_2(x) = \ldots = F_c(x)\ ,$$

the usual chi-square statistic, T, based on the set $\{ u_{t,i} \}$ may be used. We reject H_o for large values of T, where the statistic T is defined as follows:

$$T = \sum_{i=1}^{c} \sum_{t=1}^{3} \{[u_{t,i} - (n_i S_t/n)]^2/(n_i S_t/n)\} \quad . \tag{4.1}$$

The test rule based on the statistic T is:

reject H_o if $T \geq t_\alpha$,

accept H_o if $T < t_\alpha$,

where t_α is chosen so that $Pr\{ T \geq t_\alpha \mid H_o \} \leq \alpha$, and α is the preassigned level of significance.

4.3 The Null Distribution.

Let $P_{ij}(\{U_{t,i} = u_{t,i}\}, Z_1 = z_1, Z_2 = z_2)$ denote the joint probability density of $\{U_{t,i}\}$, Z_1 and Z_2, when Z_1 belongs to the i^{th} sample, and Z_2 belongs to the j^{th} sample, $i,j = 1, \ldots, c$.

Then the expression for $P_{i,j}$ is given by:

$$P_{i,j} = \prod_{m=1}^{c} \left\{ \binom{n_m}{u_{1,m}, u_{2,m}} [F_m(z_1)]^{u_{1,m}} [F_m(z_2) - F_m(z_1)]^{u_{2,m}} [1 - F_m(z_2)]^{u_{3,m}} \right\} .$$

$$u_{2,i} u_{3,j} [F_i(z_2) - F_i(z_1)]^{-1} [1 - F_j(z_2)]^{-1} \frac{dF_i(z_1)}{dz_1} \frac{dF_j(z_2)}{dz_2} \quad . \tag{4.2}$$

Hence, the joint density of the $U_{i,j}$'s, Z_1 and Z_2 is given by

$$h(u_{i,j}, z_1, z_2) = \sum_{i=1}^{c} \sum_{j=1}^{c} P_{i,j} \quad . \tag{4.3}$$

The null distribution $\varphi_o(u_{i,j})$ of the $U_{i,j}$'s, $(i,j = 1, 2, \ldots, c)$, under the null hypothesis is derived from $h(u_{i,j},z_1,z_2)$ by substituting

$$F_1(z) = F_2(z) = \ldots = F_c(z) = F(z)$$

in (4.3) and integrating the resulting expression over the range of z_1 and z_2. This yields

$$\varphi_o(u_{i,j}) = \iint\limits_{-\infty<z_1<z_2<\infty} h(u_{i,j},z_1,z_2)\, dz_1\, dz_2$$

$$= Kr(2r+1)\int_0^1\int_0^{F(z_2)} [F(z_1)]^r [F(z_2)-F(z_1)]^{r-1} [1-F(z_2)]^{2r} dF(z_1)dF(z_2) \,, \tag{4.4}$$

where

$$K = \prod_{m=1}^{c} \binom{n_m}{u_{1,m},\ u_{2,m}} \,.$$

Letting $Q_1 = F(z_1)/F(z_2)$ in the innermost integral, (4.4) can be integrated. This yields:

$$\varphi_o(u_{i,j}) = Kr(2r+1)\ B(r+1,\ r)\ B(2r+1,\ 2r+1) \,, \tag{4.5}$$

where $B(i,j) = [(i+j-1)!]^{-1}[(i-1)!(j-1)!]$.

This expression can be rewritten in the following form:

$$\varphi_o(u_{i,j}) = \left[\prod_{m=1}^{c} \binom{n_m}{u_{1,m},\ u_{2,m}}\right]\binom{n}{r,\ r}^{-1} \,. \tag{4.6}$$

For the special case in which $c = 2$, (4.6) reduces to

$$\varphi_o(u_{1,1}, u_{2,1}) = \binom{r}{u_{1,1}}\binom{r}{u_{2,1}}\binom{2r+1}{n_1 - u_{1,1} - u_{2,1}}\binom{n}{n_1}^{-1} , \qquad (4.7)$$

which agrees with the result obtained by Chakravarti, Leone, and Alanen [13]. It should be noted that the statistic T defined by (4.1) under the null hypothesis is distributed approximately as chi-square with $2(c-1)$ degrees of freedom. However, the exact distribution may be calculated from (4.6) and (4.1).

4.4 Power Function of T Test Against the Alternative of Translation in the Exponential Population.

The alternative hypothesis considered here is the same as that in Section (3.4), namely:

$$H_a \left\{ \begin{array}{ll} F_i(x) = 1 - e^{-(x-a_i)}, & x \geq a_i , \\ \quad\quad\;\; = 0 & x < a_i , \\ \text{where } a_{i+1} \geq a_i \; (i = 1, 2, \ldots, c-1), \; a_1 \geq 0. & \end{array} \right. \qquad (4.8)$$

Let $\varphi_a(u_{i,j})$ denote the probability $Pr\{ U_{t,i} = u_{t,i}$, $t = 1, 2, 3,\; i = 1, 2, \ldots, c\}$, when the alternative hypothesis H_a is true. Then $\varphi_a(u_{i,j})$ is given by:

$$\varphi_a(u_{i,j}) = \iint_{-\infty<z_1<z_2<\infty} h(u_{i,j}, z_1, z_2)\, dz_1\, dz_2 , \qquad (4.9)$$

where $h(u_{i,j}, z_1, z_2)$ is given by (4.3) .

The integration takes place over the appropriate ranges of z_1 and z_2. These ranges can be reduced to $a_1 \leq z_1 \leq z_2$ and $a_1 \leq z_2 < \infty$, since all of the F_i's are zero for $z < a_1$. Thus, (4.9) can be written in the form:

$$\varphi_a(u_{i,j}) = K \sum_{i=1}^{c} \sum_{j=1}^{c} \int_{z_2=a_1}^{\infty} \int_{z_1=a_1}^{z_2} \prod_{m=1}^{c} \{[F_m(z_1)]^{u_{1,m}} [F_m(z_2)-F_m(z_1)]^{u_{2,m}-\delta_{i,m}} \cdot$$

$$[1-F_m(z_2)]^{u_{3,m}-\delta_{j,m}}\} u_{2,i}\, u_{3,j}\, dF_i(z_1)\, dF_j(z_2) \quad . \qquad (4.10)$$

Now, we can write

$$\int_{a_1}^{\infty}\int_{a_1}^{z_2} = \int_{a_1}^{a_2}\int_{a_1}^{z_2} + \int_{a_2}^{a_3}\left[\int_{a_1}^{a_2} + \int_{a_2}^{z_2}\right] + \dots + \int_{a_c}^{\infty}\left[\int_{a_1}^{a_2} + \dots + \int_{a_c}^{z_2}\right]$$

and obtain

$$\varphi_a(u_{i,j}) = \sum_{1 \leq t<s \leq c}\sum Q_{s,t} + \sum_{s=1}^{c} Q_{s,s} \qquad (4.11)$$

where

$$Q_{s,t} = K \sum_{i=1}^{c} \sum_{j=1}^{c} \int_{z_2=a_s}^{a_{s+1}} \int_{z_1=a_t}^{a_{t+1}} P_{i,j}\, dF_i(z_1)\, dF_j(z_2),\ 1 \leq t < s \leq c\ , \qquad (4.12)$$

$$Q_{s,s} = K \sum_{i=1}^{c} \sum_{j=1}^{c} \int_{z_2=a_s}^{a_{s+1}} \int_{z_1=a_s}^{z_2} P_{i,j}\, dF_i(z_1)\, dF_j(z_2),\ 1 \leq s \leq c \quad ,$$

and

$$P_{i,j} = \prod_{m=1}^{c} \{[F_m(z_1)]^{u_{1,m}} [F_m(z_2) - F_m(z_1)]^{u_{2,m}-\delta_{i,m}} \cdot$$

$$[1-F_m(z_2)]^{u_{3,m}-\delta_{j,m}}\}\, u_{2,i}\, u_{3,j} \quad ,$$

with $a_{c+1} = \infty$.

In $Q_{s,t}$, $F_i(z_1) = 0$ for $t+1 \leq i \leq c$ and $F_j(z_2) = 0$ for $s+1 \leq j \leq c$. Hence, for $t+1 \leq i \leq c$ and $s+1 \leq j \leq c$, $Q_{s,t}$ will be zero unless $u_{1,i} = 0$ and $u_{2,j} = 0$. Thus, if we define $u_{1,c+1} = u_{2,c+1} = 0$, (4.12) and (4.13) reduce to

$$Q_{s,t} = K\left[\prod_{m=t}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=s}^{c} \delta_{u_{2,m+1},0}\right] \sum_{i=1}^{t} \sum_{j=1}^{s} \int_{z_2=a_s}^{a_{s+1}} \int_{z_1=a_t}^{a_{t+1}}$$

$$\prod_{m=1}^{t} \{F_m(z_1)^{u_{1,m}} [F_m(z_2)-F_m(z_1)]^{u_{2,m}-\delta_{m,i}}\} \prod_{m=t+1}^{s} [F_m(z_2)^{u_{2,m}}] \cdot$$

$$\prod_{m=1}^{s} \{[1-F_m(z_2)]^{u_{3,m}-\delta_{m,j}}\}\, u_{2,i} u_{3,j}\, dF_i(z_1)\, dF_j(z_2) \quad ,$$

(4.14)

$$Q_{s,s} = K\left[\prod_{m=s}^{c} \delta_{u_{1,m+1},0}\ \delta_{u_{2,m+1},0}\right] \sum_{i=1}^{s} \sum_{j=1}^{s} \int_{z_2=a_s}^{a_{s+1}} \int_{z_1=a_s}^{z_2}$$

$$\prod_{m=1}^{s} \left\{F_m(z_1)^{u_{1,m}}[F_m(z_2)-F_m(z_1)]^{u_{2,m}-\delta_{m,i}}[1-F_m(z_2)]^{u_{3,m}-\delta_{m,j}}\right\} \cdot$$

$$u_{2,i}\ u_{3,j}\ dF_i(z_1)\ dF_j(z_2) \quad . \qquad (4.15)$$

First, consider (4.14) and let

$$\eta_{i,t} = e^{-(a_t-a_i)} ,$$

$$\gamma_{i,t} = 1 - \eta_{i,t} \quad .$$

Then, we can substitute

$$F_i(z_1) = 1 - \eta_{i,t}(1-F_t(z_1)) = \gamma_{i,t} + \eta_{i,t}F_t(z_1),\ a_t \le z_1 \le a_{t+1} \quad ,$$

$$F_i(z_2) = 1 - \eta_{i,t}(1-F_t(z_2)) = \gamma_{i,t} + \eta_{i,t}F_t(z_2),\ a_t \le z_2 \le a_{t+1} \quad ,$$

$$1 \le i \le t \quad . \qquad (4.16)$$

in (4.14). This give us

$$Q_{s,t} = K\left[\prod_{m=t}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=s}^{c} \delta_{u_{2,m+1},0}\right] \sum_{j=1}^{s}\left\{\prod_{m=1}^{t} (\eta_{m,t}^{u_{2,m}})\right\}\left[\sum_{i=1}^{t} u_{2,i}\right] \cdot$$

$$\int_{z_2=a_s}^{a_{s+1}} \int_{0}^{\gamma_{t,t+1}} \left\{\prod_{m=1}^{t-1} [\gamma_{m,t}+\eta_{m,t}F_t(z_1)]^{u_{1,m}}\right\} F_t(z_1)^{u_{1,t}} \cdot$$

$$[F_t(z_2)-F_t(z_1)]^{\sum_{m=1}^{t} u_{2,m}-1} \left\{\prod_{m=t+1}^{s} F_m(z_2)^{u_{2,m}}\right\} u_{3,j} \cdot$$

$$\left\{\prod_{m=1}^{s} [1-F_m(z_2)]^{u_{3,m}-\delta_{m,j}}\right\} dF_t(z_1)\ dF_j(z_2) \quad . \qquad (4.17)$$

Assuming that $t > 1$, we can expand $[\gamma_{m,t} + \eta_{m,t}F_t(z_1)]^{u_{1,m}}$ for $1 \leq m < t$, $[F_t(z_2)-F_t(z_1)]^{\sum_{m=1}^{t} u_{2,m}-1}$, and perform the innermost integration.

This yields

$$Q_{s,t} = K\left[\prod_{m=t}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=s}^{c} \delta_{u_{2,m+1},0}\right] \sum_{j=1}^{s} \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{t-1}=0}^{u_{1,t-1}} \sum_{w=0}^{A_{2,t}}$$

$$(-1)^w \binom{A_{2,t}}{w} [A_{2,t}-w] \left\{ \prod_{m=1}^{t-1} \left[\eta_{m,t}^{u_{2,m}+q_m} \gamma_{m,t}^{u_{1,m}-q_m} \binom{u_{1,m}}{q_m} \right] \right\} u_{3,j} \ ,$$

$$[u_{1,t} + \sum_{m=1}^{t-1} q_m + w+1]^{-1} \ \gamma_{t,t+1}^{u_{1,t} + \sum_{m=1}^{t-1} q_m + w+1} \int_{z_2=a_s}^{a_{s+1}} \left[F_t(z_2) \right]^{A_{2,t}-w-1} \cdot$$

$$\left\{ \prod_{m=t+1}^{s} F_m(z_2)^{u_{2,m}} \right\} \left\{ \prod_{m=1}^{s} [1-F_m(z_2)]^{u_{3,m}-\delta_{m,j}} \right\} dF_j(z_2) \quad , \tag{4.18}$$

where $A_{i,j} = \sum_{m=1}^{j} u_{i,m}$. When $s > t+1$, we can substitute

$$F_j(z_2) = \gamma_{j,s} + \eta_{j,s} F_s(z_2), \text{ for } a_s \le z_2 \le a_{s+1}, \ 1 \le j < s \quad , \tag{4.19}$$

in (4.18), expand the binomials, and integrate. Thus, we get

$$Q_{s,t} = K\left[\prod_{m=t}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=s}^{c} \delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{t-1}=0}^{u_{1,t-1}} \sum_{w=0}^{A_{2,t}}$$

$$\sum_{q_t=0}^{A_{2,t}-w} \sum_{q_{t+1}=0}^{u_{2,t+1}} \cdots \sum_{q_{s-1}=0}^{u_{2,s-1}} \sum_{v=0}^{A_{3,s}} (-1)^{v+w} \binom{A_{2,t}}{w} \binom{A_{3,s}}{v} \binom{A_{2,t}-w}{q_t}$$

$$[A_{2,t}-w-q_t][A_{3,s}-v]\left\{\prod_{m=1}^{t-1}\left[\eta_{m,t}^{u_{2,m}+q_m}\ \gamma_{m,t}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\} \cdot$$

$$\left\{\prod_{m=t+1}^{s-1}\left[\eta_{m,s}^{q_m}\ \gamma_{m,s}^{u_{2,m}-q_m}\binom{u_{2,m}}{q_m}\right]\right\}\left\{\prod_{m=1}^{s-1}\eta_{m,s}^{u_{3,m}}\right\}\eta_{t,s}^{q_t} \cdot$$

$$\gamma_{t,s}^{A_{2,t}-w-q_t-1}\ \gamma_{t,t+1}^{u_{1,t}+\sum_{m=1}^{t-1} q_m+w+1}\left[u_{1,t}+\sum_{m=1}^{t-1} q_m+w+1\right]^{-1}$$

$$\gamma_{s,s+1}^{u_{2,s}+\sum_{m=t}^{s-1} q_m+v+1}\left[u_{2,s}+\sum_{m=t}^{s-1} q_m+v+1\right]^{-1} . \qquad (4.20)$$

Starting from (4.17) the results for the cases in which $s = t+1$

and $t = 1$ can be obtained in a similar fashion and are summarized below:

$$Q_{t+1,t} = K\left[\prod_{m=t}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=t+1}^{c} \delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{t-1}=0}^{u_{1,t-1}} \sum_{w=0}^{A_{2,t}}$$

$$\sum_{q_t=0}^{A_{2,t}-w} \sum_{v=0}^{A_{3,t+1}} (-1)^{v+w} \binom{A_{2,t}}{w} \binom{A_{3,t+1}}{v} \binom{A_{2,t}-w}{q_t} (A_{3,t+1}-v) \cdot$$

$$(A_{2,t}-w-q_t) \left\{ \prod_{m=1}^{t-1} \left[\eta_{m,t}^{u_{2,m}+q_m} \gamma_{m,t}^{u_{1,m}-q_m} \binom{u_{1,m}}{q_m} \right] \right\} \cdot$$

$$\gamma_{t,t+1}^{u_{1,t}+A_{2,t}+\sum_{m=1}^{t-1} q_m - q_t} \left\{ \prod_{m=1}^{t} \eta_{m,t+1}^{u_{3,m}} \right\} \eta_{t,t+1}^{q_t} [u_{1,t} + \sum_{m=1}^{t-1} q_m + w + 1]^{-1} \cdot$$

$$\gamma_{t+1,t+2}^{u_{2,t+1}+q_t+v+1} [u_{2,t+1}+q_t+v+1]^{-1} \quad , \tag{4.21}$$

$$Q_{s,1} = K\left[\prod_{m=1}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=s}^{c} \delta_{u_{2,m+1},0}\right] \sum_{w=0}^{u_{2,1}} \sum_{q_1=0}^{u_{2,1}-w} \sum_{q_2=0}^{u_{2,2}} \cdots \sum_{q_{s-1}=0}^{u_{2,s-1}}$$

$$\sum_{v=0}^{A_{3,s}} (-1)^{v+w} \binom{u_{2,1}}{w}\binom{A_{3,s}}{v}\binom{u_{2,1}-w}{q_1} (u_{2,1}-w-q_1)(A_{3,s}-v) \quad \cdot$$

$$\left\{\prod_{m=1}^{s-1} \eta_{m,s}^{u_{3,m}+q_m}\right\}\left\{\prod_{m=2}^{s-1} \left[\gamma_{m,s}^{u_{2,m}-q_m}\binom{u_{2,m}}{q_m}\right]\right\} \gamma_{1,s}^{u_{2,1}-w-q_1-1} \quad \cdot$$

$$\gamma_{1,2}^{u_{1,1}+w+1} [u_{1,1}+w+1]^{-1}\ \gamma_{s,s+1}^{u_{2,s}+\sum_{m=1}^{s-1} q_m+v+1} [u_{2,s}+\sum_{m=1}^{s-1} q_m+v+1]^{-1} \quad , \tag{4.22}$$

$$Q_{2,1} = K\left[\prod_{m=1}^{c} \delta_{u_{1,m+1},0}\right]\left[\prod_{m=2}^{c} \delta_{u_{2,m+1},0}\right] \sum_{w=0}^{u_{2,1}} \sum_{q_1=0}^{u_{2,1}-w} \sum_{v=0}^{A_{3,2}} (-1)^{v+w} \quad \cdot$$

$$\binom{u_{2,1}}{w}\binom{A_{3,2}}{v}\binom{u_{2,1}-w}{q_1} (u_{2,1}-w-q_1)(A_{3,2}-v)\,\eta_{1,2}^{u_{3,1}+q_1}[u_{1,1}+w+1]^{-1} \cdot$$

$$\gamma_{1,2}^{u_{1,1}+u_{2,1}-q_1}\ \gamma_{2,3}^{u_{2,2}+q_1+v+1} [u_{2,2}+q_1+v+1]^{-1} \quad . \tag{4.23}$$

If we substitute (4.19) for z_1 and z_2 in (4.15) and expand, we get

$$Q_{s,s} = K\left[\prod_{m=s}^{c} \delta_{u_{1,m+1},0}\,\delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{s-1}=0}^{u_{1,s-1}}$$

$$\left\{\prod_{m=1}^{s-1}\left[\eta_{m,s}^{u_{2,m}+u_{3,m}+q_m}\ \gamma_{m,s}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\} A_{2,s}\ A_{3,s}\ \cdot$$

$$\int_{z_2=a_s}^{a_{s+1}} \int_{z_1=a_s}^{z_2} F_s(z_1)^{u_{1,s}+\sum_{m=1}^{s-1} q_m}\ [F_s(z_2)-F_s(z_1)]^{A_{2,s}-1}\ \cdot$$

$$[1-F_s(z_2)]^{A_{3,s}-1}\ dF_s(z_1)\ dF_s(z_2)\quad . \qquad 4.24$$

Now, if we let $L = F_s(z_1)[F_s(z_2)]^{-1}$ in the innermost integral and integrate, we obtain

$$Q_{s,s} = K\left[\prod_{m=s}^{c} \delta_{u_{1,m+1},0}\,\delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{s-1}=0}^{u_{1,s-1}}$$

$$\left\{\prod_{m=1}^{s-1}\left[\eta_{m,s}^{u_{2,m}+u_{3,m}+q_m}\;\gamma_{m,s}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\} A_{2,s}\;A_{3,s}\;\cdot$$

$$B\left(u_{1,s}+\sum_{m=1}^{s-1} q_m+1, A_{2,s}\right)\int_{z_2=a_s}^{a_{s+1}} F_s(z_2)^{A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1} q_m}\;\cdot$$

$$[1-F_s(z_2)]^{A_{3,s}-1}\; dF_s(z_2)\quad . \tag{4.25}$$

The factor $[1-F_s(z_2)]^{A_{3,s}-1}$ can be expanded and the resulting expression integrated to yield

$$Q_{s,s} = K\left[\prod_{m=s}^{c} \delta_{u_{1,m+1},0}\,\delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{s-1}=0}^{u_{1,s-1}}$$

$$\left\{\prod_{m=1}^{s-1}\left[\eta_{m,s}^{u_{2,m}+u_{3,m}+q_m}\,\gamma_{m,s}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\} A_{2,s} \cdot$$

$$B\left(u_{1,s}+\sum_{m=1}^{s-1} q_m+1, A_{2,s}\right)\sum_{v=0}^{A_{3,s}}(-1)^v\binom{A_{3,s}}{v}(A_{3,s}-v)\cdot$$

$$\gamma_{s,s+1}^{A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1} q_m+v+1}\,[A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1} q_m+v+1]^{-1}\,, \qquad (4.26)$$

or rewriting the Beta function in terms of factorials one obtains

$$Q_{s,s} = K\left[\prod_{m=s}^{c} \delta_{u_{1,m+1},0}\delta_{u_{2,m+1},0}\right] \sum_{q_1=0}^{u_{1,1}} \cdots \sum_{q_{s-1}=0}^{u_{1,s-1}}$$

$$\left\{\prod_{m=1}^{s-1}\left[\eta_{m,s}^{u_{2,m}+u_{3,m}+q_m}\ \gamma_{m,s}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\} \cdot$$

$$\left(u_{1,s}+\sum_{m=1}^{s-1}q_m\right)!\,A_{2,s}!\,[(A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1}q_m)!]^{-1}\sum_{v=0}^{A_{3,s}}(-1)^v\binom{A_{3,s}}{v}(A_{3,s}-v)\cdot$$

$$\gamma_{s,s+1}^{A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1}q_m+v+1}\,[A_{2,s}+u_{1,s}+\sum_{m=1}^{s-1}q_m+v+1]^{-1} \quad . \qquad (4.27)$$

For the case in which $s = 1$, (4.27) becomes

$$Q_{1,1} = K\left[\prod_{m=1}^{c} \delta_{u_{1,m+1},0}\delta_{u_{2,m+1},0}\right] u_{1,1}!\,u_{2,1}!\,[(u_{1,1}+u_{2,1})!]^{-1} \quad .$$

$$\sum_{v=0}^{u_{3,1}}(-1)^v\binom{u_{3,1}}{v}\left(u_{3,1}-v\right)\ \gamma_{1,2}^{u_{2,1}+u_{1,1}+v+1}\,[u_{2,1}+u_{1,1}+v+1]^{-1} .$$

(4.28)

Also in the case $s = c$, (4.20) and (4.27) can be simplified since the outermost integral becomes a complete Beta function. These results are summarized below.

$$Q_{c,t}=K\left[\prod_{m=t}^{c}\delta_{u_{1,m+1},0}\right]\sum_{q_1=0}^{u_{1,1}}\cdots\sum_{q_{t-1}=0}^{u_{1,t-1}}\ \sum_{w=0}^{A_{2,t}}\ \sum_{q_t=0}^{A_{2,t}-w}\ \sum_{q_{t+1}=0}^{u_{2,t+1}}\cdots\sum_{q_{c-1}=0}^{u_{2,c-1}}(-1)^{w}\ \cdot$$

$$\binom{A_{2,t}}{w}\binom{A_{2,t}-w}{q_t}(A_{2,t}-w-q_t)\left\{\prod_{m=1}^{t-1}\left[\eta_{m,t}^{u_{2,m}+q_m}\ \gamma_{m,t}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\}\cdot$$

$$\left\{\prod_{m=t+1}^{c-1}\left[\eta_{m,c}^{q_m}\ \gamma_{m,c}^{u_{2,m}-q_m}\binom{u_{2,m}}{q_m}\right]\right\}\left\{\prod_{m=1}^{c-1}\eta_{m,c}^{u_{3,m}}\right\}\eta_{t,c}^{q_t}\ \gamma_{t,c}^{A_{2,t}-w-q_t-1}\ \cdot$$

$$\gamma_{t,t+1}^{u_{1,t}+\sum_{m=1}^{t-1}q_m+w+1}\left[u_{1,t}+\sum_{m=1}^{t-1}q_m+w+1\right]^{-1}(2r+1)!\left(u_{2,c}+\sum_{m=t}^{c-1}q_m\right)!\ \cdot$$

$$\left[\left(u_{2,c}+2r+\sum_{m=t}^{c-1}q_m+1\right)!\right]^{-1}\ . \qquad (4.29)$$

$$Q_{c,c-1} = K\,\delta_{u_{1,c},0}\sum_{q_1=0}^{u_{1,1}}\cdots\sum_{q_{c-2}=0}^{u_{1,c-2}}\sum_{w=0}^{A_{2,c-1}}\sum_{q_{c-1}=0}^{A_{2,c-1}-w}(-1)^w\binom{A_{2,c-1}}{w}\binom{A_{2,c-1}-w}{q_{c-1}}\cdot$$

$$(A_{2,c-1}-w-q_{c-1})\left\{\prod_{m=1}^{c-2}\left[\eta_{m,c-1}^{u_{2,m}+q_m}\ \gamma_{m,c-1}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\}\cdot$$

$$\left\{\prod_{m=1}^{c-1}\eta_{m,c}^{u_{3,m}}\right\}\eta_{c-1,c}^{q_{c-1}}\ \gamma_{c-1,c}^{A_{2,c-1}+u_{1,c-1}+\sum_{m=1}^{c-2}q_m-q_{c-1}}\ .$$

$$\left[u_{1,c-1}+\sum_{m=1}^{c-2}q_m+w+1\right]^{-1}(2r+1)!(u_{2,c}+q_{c-1})!\ \cdot$$

$$[(u_{2,c}+2r+q_{c-1}+1)!]^{-1}\ . \qquad (4.30)$$

$$Q_{c,c} = K\sum_{q_1=0}^{u_{1,1}}\cdots\sum_{q_{c-1}=0}^{u_{1,c-1}}\left\{\prod_{m=1}^{c-1}\left[\eta_{m,c}^{u_{2,m}+u_{3,m}+q_m}\ \gamma_{m,c}^{u_{1,m}-q_m}\binom{u_{1,m}}{q_m}\right]\right\}\cdot$$

$$r!(2r+1)!\left(u_{1,c}+\sum_{m=1}^{c-1}q_m\right)![(u_{1,c}+3r+\sum_{m=1}^{c-1}q_m+1)!]^{-1}\ . \qquad (4.31)$$

4.5 Results.

The exact power of the T statistic has been computed for three samples using several sets of exponential alternatives. These results are presented in Table 4.1 for total sample sizes of 9 and 13. Since the computations were performed in single precision arithmetic, the error in the computations is rather severe, and at times, amounts to $\pm$ 1 in the second decimal place as indicated by the cumulative sum.

As in the case of Mood's test, the power, in most cases, increases with a positive shift in the location parameters, but, again, several cases can be noted in which the increase does not occur. This indicates that the test is biased under these circumstances. As stated in Chapter III, this effect can be attributed to a sharp "peaking" in the shape of the distribution of the T statistic.

4.6 Further Extensions.

Expressions similar to those developed in this chapter can be derived for a change in scale and location of a rectangular distribution and a change in scale of an exponential distribution. The results presented in Table 4.1 can be extended to larger samples and other alternatives. The accuracy of the computations can be improved by using double precision arithmetic. Also, some extensions similar to those indicated in Chapter I can be applied to this test.

TABLE 4.1

Exact Power of Massey's Test for Three Samples with

Exponential Alternatives: $Pr\{ T \geq t \mid H \}$.

n	n_1	n_2	t	α	$a_1 = 0$, $a_2 = .1$, $a_3 = .2$	$a_1 = 0$, $a_2 = .1$, $a_3 = .5$	$a_1 = 0$, $a_2 = .2$, $a_3 = .5$
9	2	3	13.800	.008	.016	.043	.044
			11.700	.024	.039	.063	.071
			9.075	.056	.091	.128	.154
			8.850	.079	.119	.197	.204
			7.500	.127	.159	.237	.238
	3	3	9.600	.071	.084	.110	.115
	4	3	13.800	.008	.004*	.002*	.002*
			11.700	.024	.020*	.010*	.012*
			9.075	.056	.042*	.022*	.025*
			8.850	.079	.070*	.052*	.070*
			7.500	.127	.135	.149	.166
13	2	4	16.050	.002	.004	.012	.008
			11.718	.007	.015	.053	.046
			11.099	.019	.030	.072	.063
			9.728	.054	.080	.150	.139
			8.932	.073	.102	.205	.178
			8.269	.112	.163	.294	.280
	4	4	16.714	.001	.002	.009	.010
			13.929	.006	.008	.015	.017
			12.071	.013	.021	.047	.050
			9.657	.055	.070	.111	.127
			8.976	.078	.104	.176	.189
			7.738	.135	.179	.312	.306
	7	4	16.050	.002	.001*	.000*	.000*
			11.718	.007	.004*	.002*	.002*
			11.099	.019	.014*	.007*	.008*
			9.728	.054	.048*	.041*	.043*
			8.932	.073	.070*	.061*	.081
			8.269	.112	.100*	.076*	.097*

*These values indicate a bias in the test.

CHAPTER V

ANALYSIS OF CATEGORICAL DATA

5.1 Introduction.

The usual method of testing a hypothesis concerning categorical data is to compute a statistic T, which is distributed approximately as χ^2. Then the hypothesis is accepted or rejected depending upon the relationship of the observed value of T to a predetermined critical value obtained from the χ^2 distribution. Since T is only approximately distributed as χ^2, the exact distribution of T under several different null hypotheses has been computed for both one-way and two-way classifications, and these results have been compared with those obtained from the χ^2 approximation, in order to determine when the approximation is valid. Also, the exact power of the T test has been computed for a one-way classification with fixed alternatives, and these results compare favorably with those obtained from a non-central χ^2 approximation.

Extensive tables of the non-central chi-square distribution given by

$$F(\lambda,\nu,y) = \int_0^y \frac{e^{-\frac{x}{2}} e^{-\frac{\lambda}{2}}}{2^{\nu/2}} \sum_{j=0}^{\infty} \frac{x^{\nu/2+j-1} \lambda^j}{\Gamma(\nu/2+j) 2^{2j} j!} \, dx \quad ,$$

where λ denotes the non-centrality parameter and ν the degrees of freedom, have been computed at Case Institute of Technology [8].

The computational algorithm suggested by Professor N.L. Johnson [10] reduces the integral to a double Poisson sum:

$$F(\lambda,\nu,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{i} S_{2i+\nu}\left(\frac{y}{2}\right) P_j\left(\frac{\lambda}{2}\right) ,$$

where

$$\left.\begin{aligned} S_{2j}\left(\frac{y}{2}\right) &= P_j\left(\frac{y}{2}\right) \\ S_{2j-1}\left(\frac{y}{2}\right) &= Q_j\left(\frac{y}{2}\right) \end{aligned}\right\} \quad j \geq 1 ,$$

and

$$P_j(x) = \frac{e^{-x}\, x^j}{j!} , \qquad j \geq 0 ,$$

$$Q_j(x) = \frac{e^{-x}\, x^{j-\frac{1}{2}}}{\Gamma(j+\frac{1}{2})} , \qquad j \geq 1 .$$

5.2 One-way Classification.

In this case the set of n observations is partitioned into k cells $\{ A_i \mid i = 1, 2, \ldots, k \}$ with n_i observations in cell A_i. We will consider the null hypothesis given by

$$H_0 : \{ \pi_i \mid i = 1, 2, \ldots, k \} \text{ such that } \sum_{i=1}^{k} \pi_i = 1 , \tag{5.1}$$

where π_i represents the probability that an observation will fall in cell A_i. The usual statistic T is defined in this case by

$$T = \sum_{i=1}^{k} (n_i - n\pi_i)^2 \, (n\pi_i)^{-1} \quad . \tag{5.2}$$

The observed frequencies, n_i, will be distributed as the multinomial distribution

$$\varphi_o(n_1, \ldots, n_k) = (n!) \left(\prod_{i=1}^{k} n_i! \right)^{-1} \left(\prod_{i=1}^{k} \pi_i^{\,n_i} \right) \quad . \tag{5.3}$$

Hence the exact distribution of T can be computed from

$$\Pr\{ T \geq t \mid H_o \} = \sum_{n_i} \varphi_o(n_1, \ldots, n_k), \text{ such that } T \geq t\,, \tag{5.4}$$

where t is a fixed value of T. The results for the exact distribution of T as a function of the sample size, n, are displayed in figures (5.1) and (5.2) for two different null hypotheses. The approximating chi-square distribution follows the exact distribution very closely even for small values of n.

The exact power of the test with the significance level at α, can be evaluated by considering an alternative hypothesis such as:

$$H_a : \{ p_i \mid i = 1, 2, \ldots, k \}, \text{ where } \sum_{i=1}^{k} p_i = 1 \quad . \tag{5.5}$$

Then the power of the test against this alternative will be given by

$$\Pr\{ T \geq t_\alpha \mid H_a \} = \sum_{n_i} \varphi_a(n_1, \ldots, n_k), \text{ such that } T \geq t_\alpha\,, \tag{5.6}$$

where φ_a is defined by

$$\varphi_a(n_1, \ldots, n_k) = (n!)\left(\prod_{i=1}^{k} n_i!\right)^{-1}\left(\prod_{i=1}^{k} p_i^{\,n_i}\right) , \quad (5.7)$$

and t_α is defined implicitly by

$$\Pr\{ T \geq t_\alpha \mid H_o \} \leq \alpha . \quad (5.8)$$

If we let X be a χ^2 variable with $k-1$ degrees of freedom, then T will be approximately distributed as X [4]. Patnaik [19] has shown that the distribution of T under the alternative hypothesis can be approximated by the non-central χ^2 distribution. This fact when applied to (5.6) yields:

$$\Pr \{T \geq x_\alpha \mid H_a\} \sim \int_{x_\alpha}^{\infty} f(\chi^2)\, d(\chi^2) , \quad (5.9)$$

where

$$f(\chi^2) = e^{-\frac{1}{2}\chi^2} e^{-\frac{1}{2}\lambda} 2^{-\frac{1}{2}\nu} \sum_{j=0}^{\infty} \{(\chi^2)^{\frac{1}{2}\nu+j-1}\lambda^j[\Gamma(\tfrac{1}{2}\nu+j)2^{2j}j!]^{-1}\} ,$$

is the non-central chi-square distribution with ν degrees of freedom and λ is the non-centrality parameter. In our case

$$\nu = k - 1, \quad \lambda = n \sum_{i=1}^{k} (p_i - \pi_i)^2(\pi_i)^{-1} . \quad (5.10)$$

This function has been extensively tabulated in [8]. Also, Patnaik [19] has shown that the non-central χ^2 distribution can be approximated by a central χ^2 distribution by equating the first two moments of the distributions. This yields a further approximation to (5.6), namely:

$$\Pr\{ T \geq x_\alpha \} \sim \int_{x_\alpha/\rho}^{\infty} g(y)\, dy \quad , \qquad (5.11)$$

where

$$g(y) = e^{-\frac{1}{2}y} (y)^{\frac{1}{2}\nu' -1}\, 2^{-\frac{1}{2}\nu'} [\Gamma(\tfrac{1}{2}\nu')]^{-1} \quad , \text{ and}$$

$$y = \chi^2/\rho \, , \; \rho = (\nu+2\lambda)(\nu+\lambda)^{-1}, \; \nu' = (\nu+\lambda)^2(\nu+2\lambda)^{-1} \quad .$$

Tables (5.1) and (5.2) compare these results of equations (5.6), (5.9), and (5.11) for two different sets of hypotheses. The approximations to the exact power given by (5.9) and (5.11) are close to each other. They both overestimate the exact power of T for $n \leq 20$ in example 1 presented in Table 5.1 and for $n \leq 15$ in example 2 presented in Table 5.1. However, for $n = 20$ in example 2, they underestimate the exact power.

5.3 Two-way Classification.

In this case, the set of observations form a two-way contingency table in which three different subcases can be distinguished, namely:

(i) Neither set of marginal sums fixed.

(ii) One set of marginal sums fixed.

(iii) Both sets of marginal sums fixed.

In each of the subcases, the expression for the test statistic T is:

$$T = \sum_{i=1}^{r} \sum_{j=1}^{s} (n_{ij} - n_{i.}n_{.j}/n)^2/(n_{i.}n_{.j}/n) \quad , \qquad (5.12)$$

where r denotes the number of rows, s the number of columns, n_{ij} the frequency in the ij^{th} cell, $n_{i.}$ the marginal i^{th} row sum, and $n_{.j}$ the marginal j^{th} column sum. Also, n denotes the grand total.

In order to obtain the null distribution of T, it will be necessary to consider each of the subcases separately.

5.3.1 Neither Sets of Marginal Sums Fixed.

In this subcase, we will consider the following null hypothesis:

$$H_o : \{ p_{ij} = p_{i.}p_{.j} \mid i = 1, \ldots, r ; j = 1, \ldots, s \} \quad , \qquad (5.13)$$

where p_{ij} is the probability that an observation will fall in the ij^{th} cell, $p_{i.}$ are the marginal row probabilities, and $p_{.j}$ the marginal column probabilities. Then

$$\sum_{i=1}^{r} p_{i.} = \sum_{j=1}^{s} p_{.j} = 1 \quad . \qquad (5.14)$$

The observed cell frequencies, n_{ij}, will be distributed under the null hypothesis as:

$$\varphi_o(n_{ij}) = n! \left[\prod_{i=1}^{r} \prod_{j=1}^{s} n_{ij}! \right]^{-1} \left[\prod_{i=1}^{r} \prod_{j=1}^{s} (p_{i.}p_{.j})^{n_{ij}} \right] \quad , \qquad (5.15)$$

where $\varphi_o(n_{ij})$ will denote the probability of observing the $n_{ij}(i=1, 2, \ldots, r;\ j = 1, 2, \ldots, s)$, and the null distribution of T will be given by

$$\Pr\{ T \geq t_\alpha \mid H_o \} = \sum_{n_{ij}} \varphi_o(n_{ij}) \text{ , such that } T \geq t_\alpha \tag{5.16}$$

The results for the cases a) $r = 2$, $s = 3$ and b) $r = 3$, $s = 3$ are summarized in figures (5.3) and (5.4). Since T is asymptotically distributed as the central χ^2 distribution with $(r - 1)(s - 1)$ degrees of freedom, this curve is also plotted on the graphs for comparison purposes. It should be noted that the exact distribution of T is fairly well approximated by the χ^2 distribution for relatively small sample sizes.

5.3.2 Only Row Marginal Sums Fixed.

For this subcase the null hypothesis becomes

$$H_o : \{ p_{ij} = p_{.j} \mid j = 1, 2, \ldots, s \} \text{ , with} \tag{5.17}$$

$$\sum_{j=1}^{s} p_{.j} = 1 , \quad \sum_{j=1}^{s} n_{ij} = n_{i.} \quad \text{(fixed)} . \tag{5.18}$$

The distribution of the observed cell frequencies, n_{ij}, under the null hypothesis is:

$$\varphi_o(n_{ij}) = \prod_{i=1}^{r} \left[n_{i.}! \left(\prod_{j=1}^{s} n_{ij}! \right)^{-1} \left(\prod_{j=1}^{s} p_{.j}^{\ n_{ij}} \right) \right] , \tag{5.19}$$

and the null distribution of T will again be given by (5.16). Also T is asymptotically distributed as χ^2 with $(r-1)(s-1)$ degrees of freedom. These results are displayed for several different sets of marginal sums, in figures (5.5) and (5.6).

5.3.3 Both Sets of Marginal Sums Fixed.

It has been shown by Mood [16] that the distribution of the cell frequencies, n_{ij}, does not depend upon the cell probabilities, namely, p_{ij}, but is dependent only on the fixed marginal sums. The distribution of the cell frequencies is given by the hypergeometric distribution:

$$\varphi(n_{ij} \mid n_{i.}, n_{.j} \text{ fixed}) = \left[\prod_{i=1}^{r} n_{i.}!\right]\left[\prod_{j=1}^{s} n_{.j}!\right]\left[n!\left(\prod_{i=1}^{r}\ \prod_{j=1}^{s} n_{ij}!\right)\right]^{-1}, \tag{5.20}$$

where

$$\sum_{i=1}^{r} n_{ij} = n_{.j} \text{ (fixed)}, \quad \sum_{j=1}^{s} n_{ij} = n_{i.} \text{ (fixed)} . \tag{5.21}$$

The exact distribution of T is given by (5.16), and T is asymptotically distributed as χ^2 with $(r-1)(s-1)$ degrees of freedom. Typical results are given in figures (5.7) and (5.8) for several different sets of marginal sums.

5.4 Further Extensions.

The exact power of the T test for the case of a two-way classification has yet to be investigated. Also, both the null distribution and power calculations can be done for other classifications such as a three-way table. However, present computing equipment is inadequate for extending most of the above results, since even a 3×3 contingency table with neither margins fixed requires a considerable amount of computing time on an IBM 7090. The actual number of combinations that were investigated for a sample size of 15 was 490,314, and this number increases rapidly with n.

TABLE 5.1

Power of the T Test for a One-way Classification

			Example 1				Example 2			
			H_o: $\pi_1=1/4$, $\pi_2=1/4$, $\pi_3=1/4$, $\pi_4=1/4$				H_o: $\pi_1=9/16$, $\pi_2=3/16$, $\pi_3=3/16$, $\pi_4=1/16$			
			H_1: $p_1=1/4$, $p_2=3/16$, $p_3=1/4$, $p_4=5/16$				H_1: $p_1=7/12$, $p_2=2/12$, $p_3=2/12$, $p_4=1/12$			
				Exact	Non-cent Chi-sq	Chi-sq		Exact	Non-cent Chi-sq	Chi-sq
α	x_α	n	t_α	$\Pr[T \geq t_\alpha]$	$\Pr[X \geq x_\alpha]$	$\Pr[Y \geq x_\alpha]$	t_α	$\Pr[T \geq t_\alpha]$	$\Pr[X \geq x_\alpha]$	$\Pr[Y \geq x_\alpha]$
.01	11.30	10	13.20	.0052	.0163	.0164	12.40	.0119	.0125	.0125
		15	11.93	.0130	.0198	.0199	12.85	.0152	.0137	.0137
		20	11.20	.0196	.0235	.0237	12.53	.0198	.0149	.0150
.05	7.81	10	8.40	.0512	.0692	.0690	8.84	.0533	.0575	.0574
		15	8.20	.0602	.0792	.0789	8.11	.0587	.0612	.0612
		20	8.00	.0747	.0896	.0891	8.09	.0655	.0650	.0650
.10	6.25	10	7.60	.0690	.1291	.1288	6.00	.1032	.1114	.1114
		15	6.60	.1361	.1440	.1433	6.45	.1166	.1172	.1170
		20	6.40	.1481	.1590	.1579	6.49	.1370	.1229	.1227

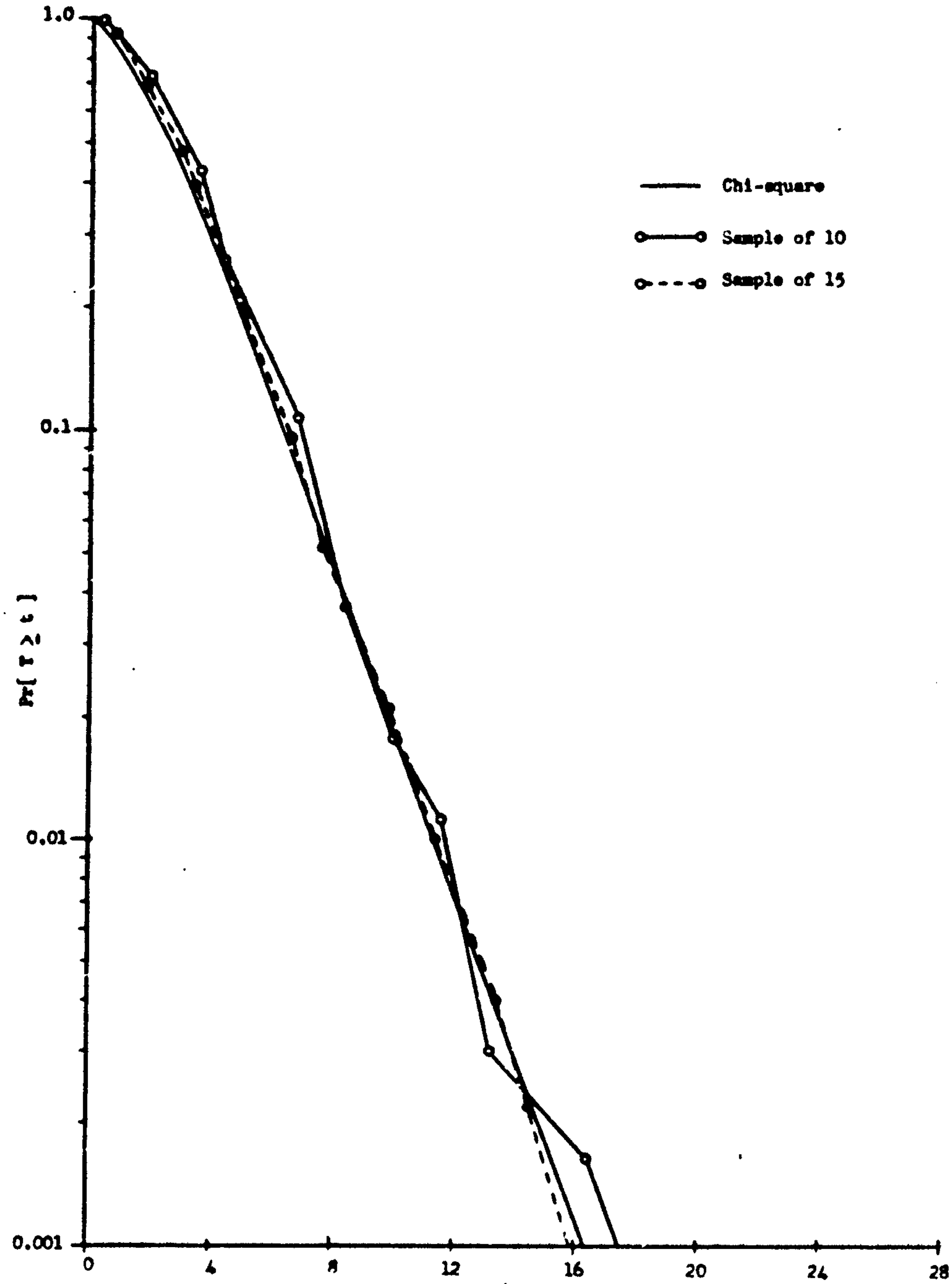

Figure 5.1 Exact null distribution for a one-way classification.
H_0: { p_i = 1/4 , i = 1, 2, 3, 4 } 3 degrees of freedom.

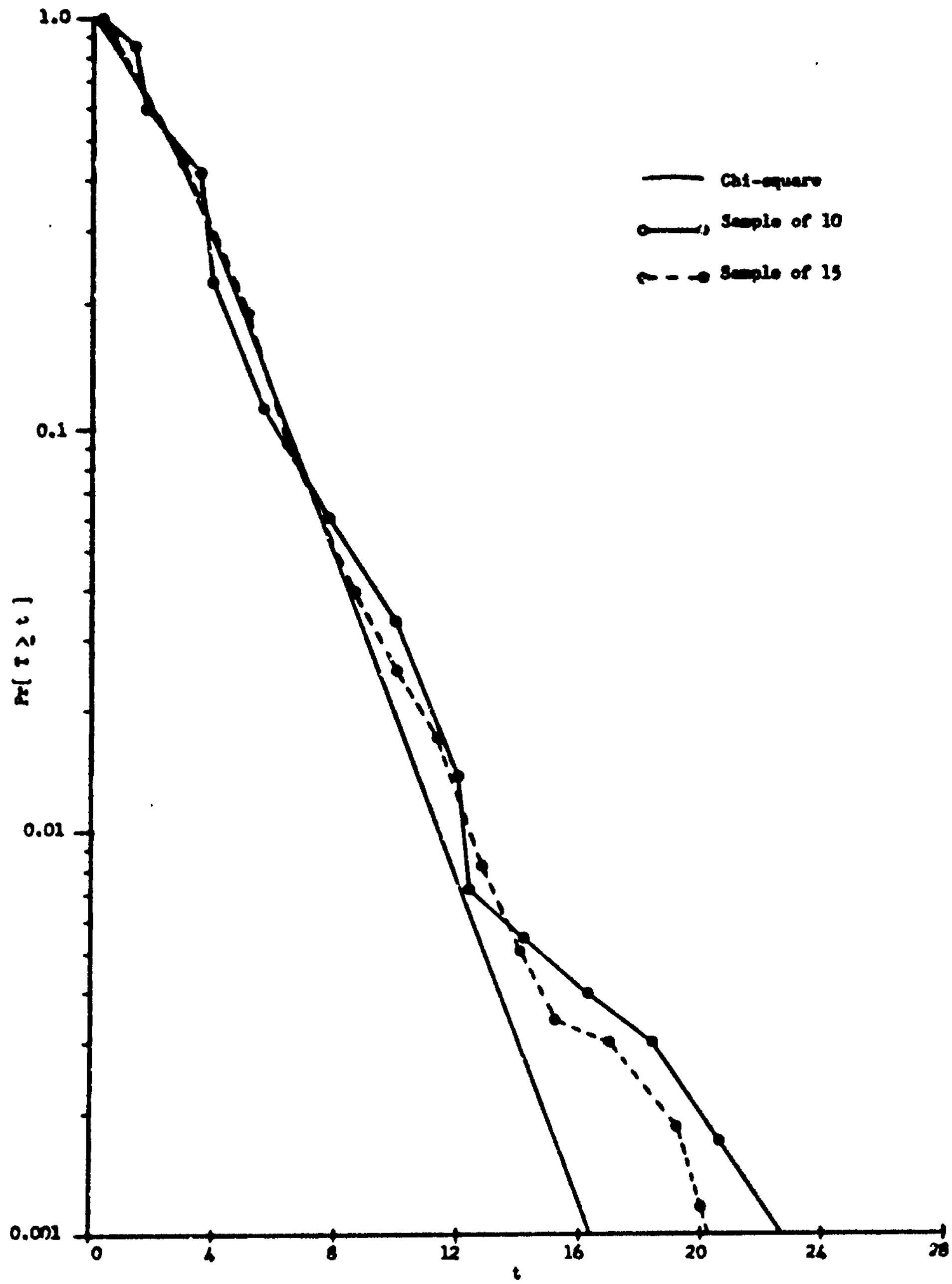

Figure 5.2 Exact null distribution for a one-way classification.
H_0: { $p_1 = 1/4$, $p_2 = 3/16$, $p_3 = 1/4$, $p_4 = 5/16$ }
3 degrees of freedom.

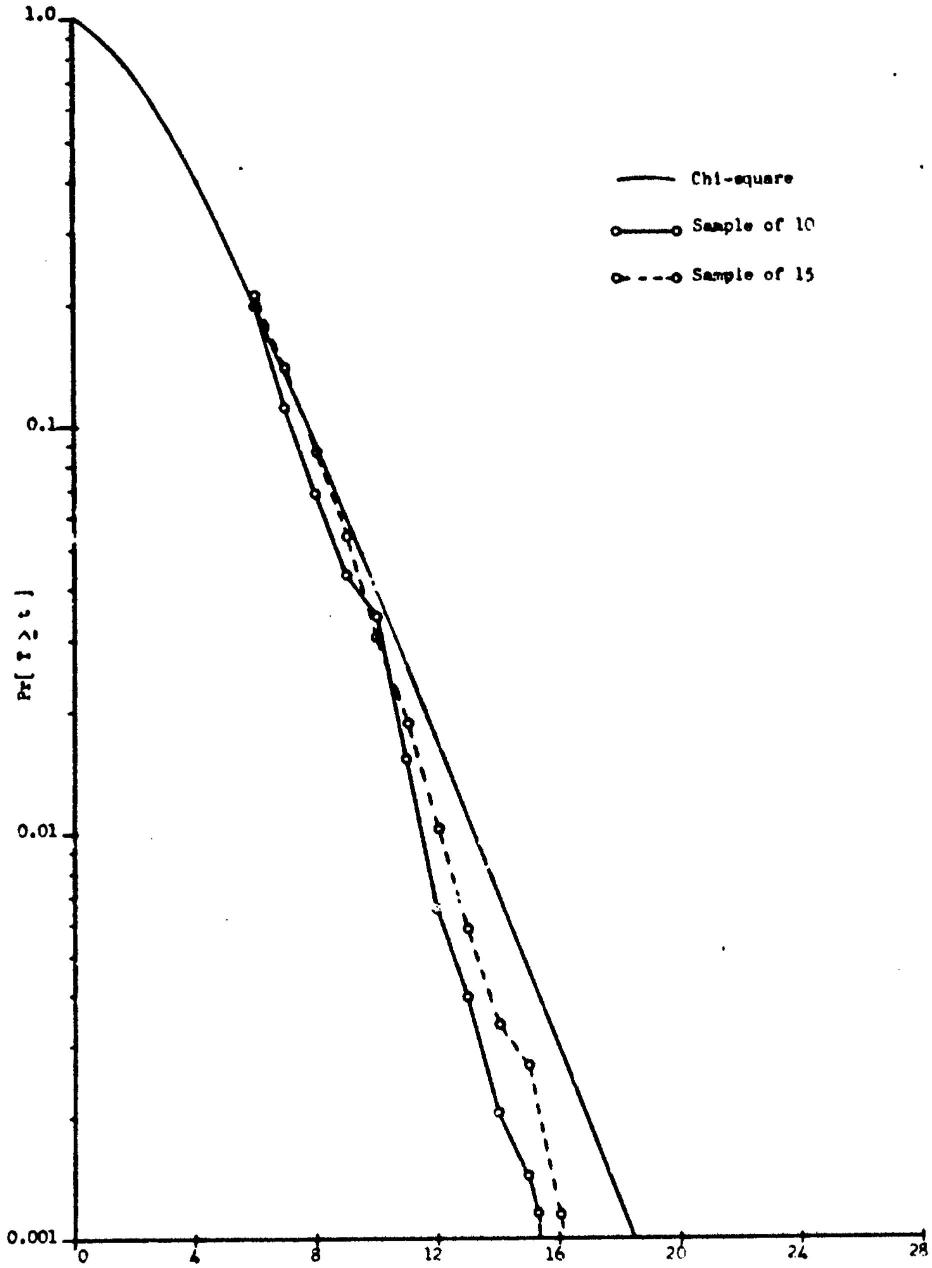

Figure 5.3 Exact null distribution for a 3×3 contingency table.
H_o: $\{ p_{ij} = 1/9,\ i = 1, 2, 3;\ j = 1, 2, 3 \}$
4 degrees of freedom and neither marginal sum fixed.

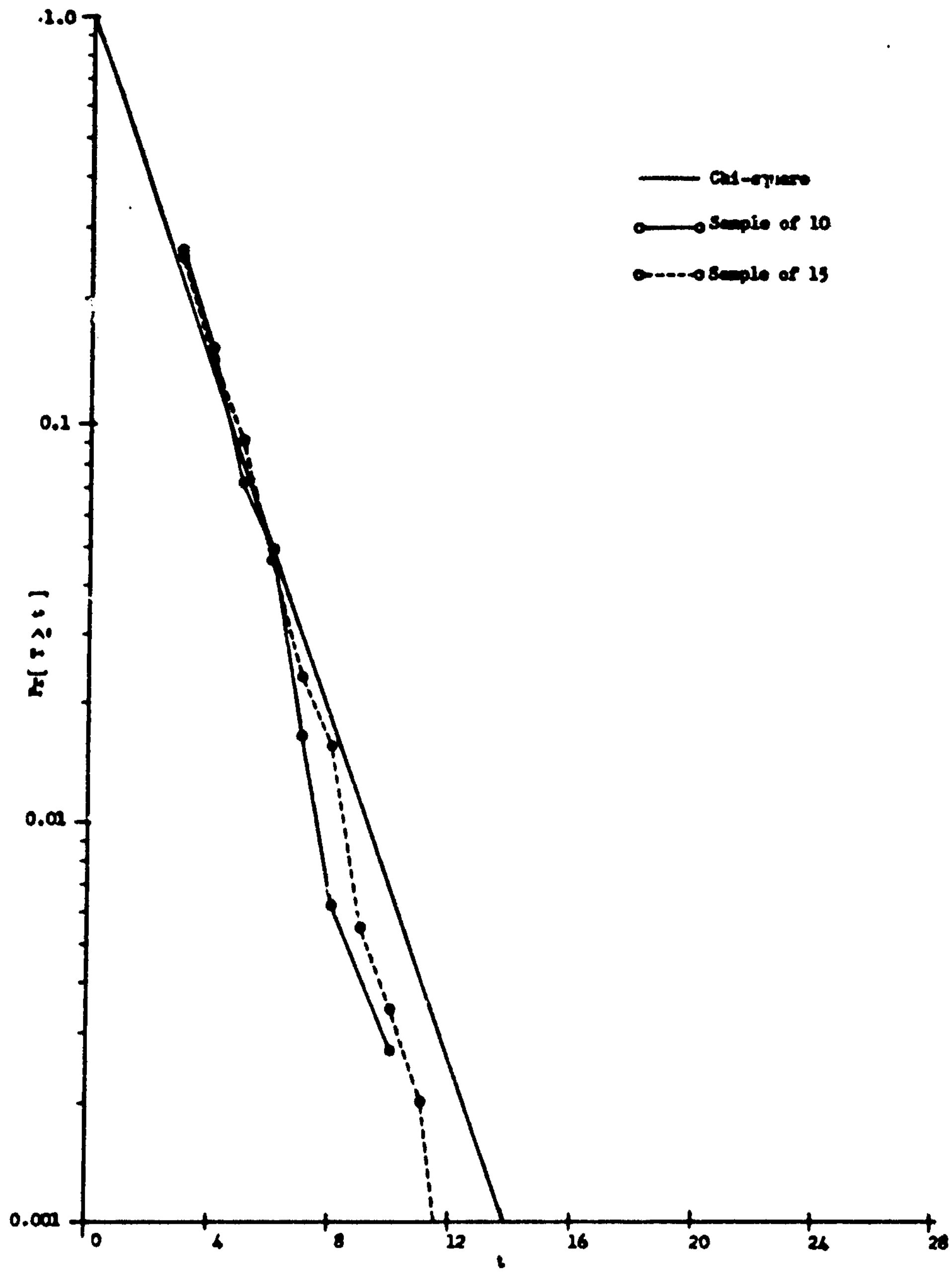

Figure 5.4 Exact null distribution for a 2×3 contingency table.
H_0: $\{ p_{ij} = 1/6,\ i = 1, 2;\ j = 1, 2, 3 \}$
2 degrees of freedom and neither marginal sum fixed.

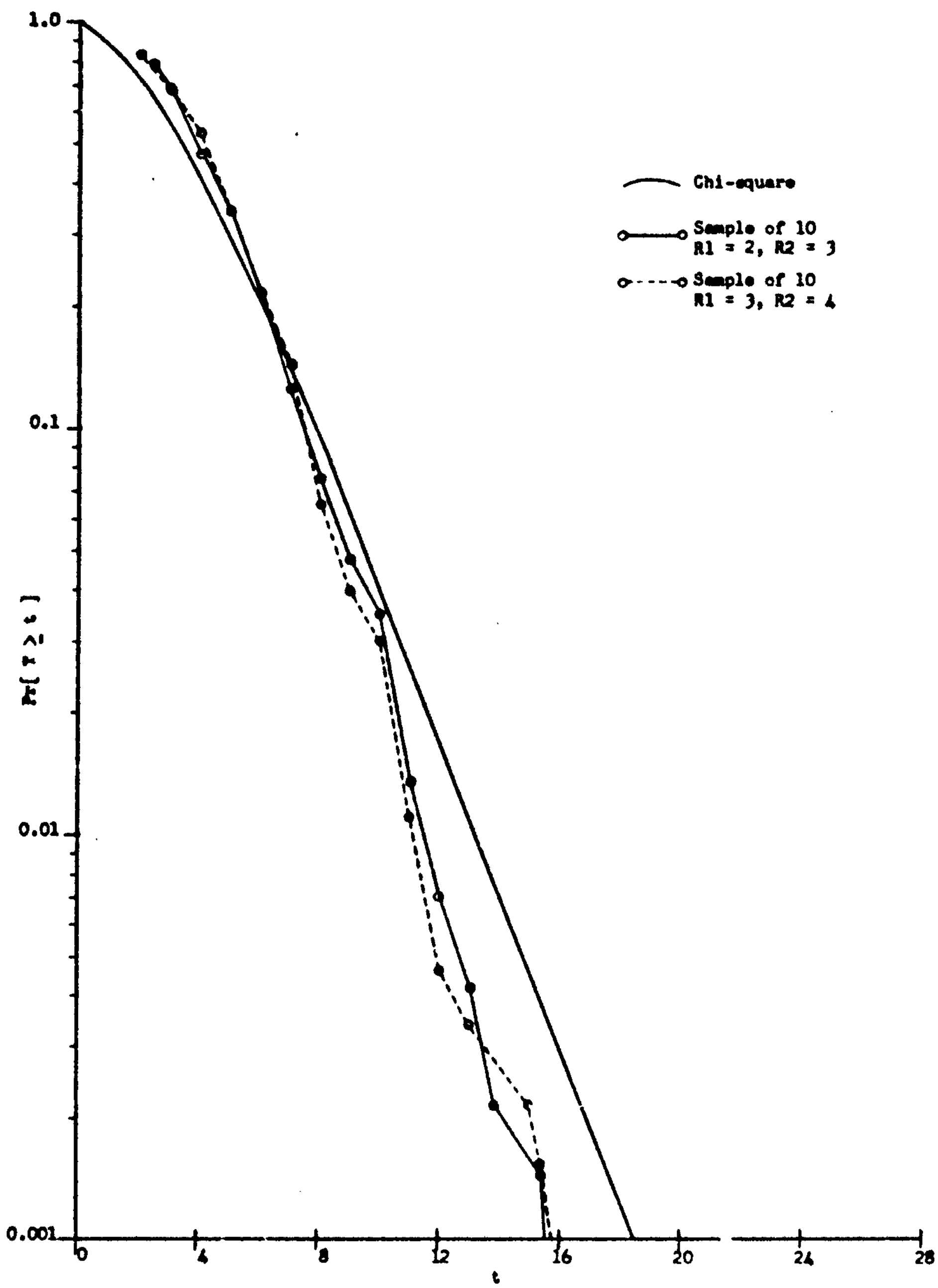

Figure 5.5 Exact null distribution for a 3×3 contingency table.
H_0: $\{ p_{.j} = 1/3,\ j = 1, 2, 3 \}$
4 degrees of freedom and row marginal sums are fixed.

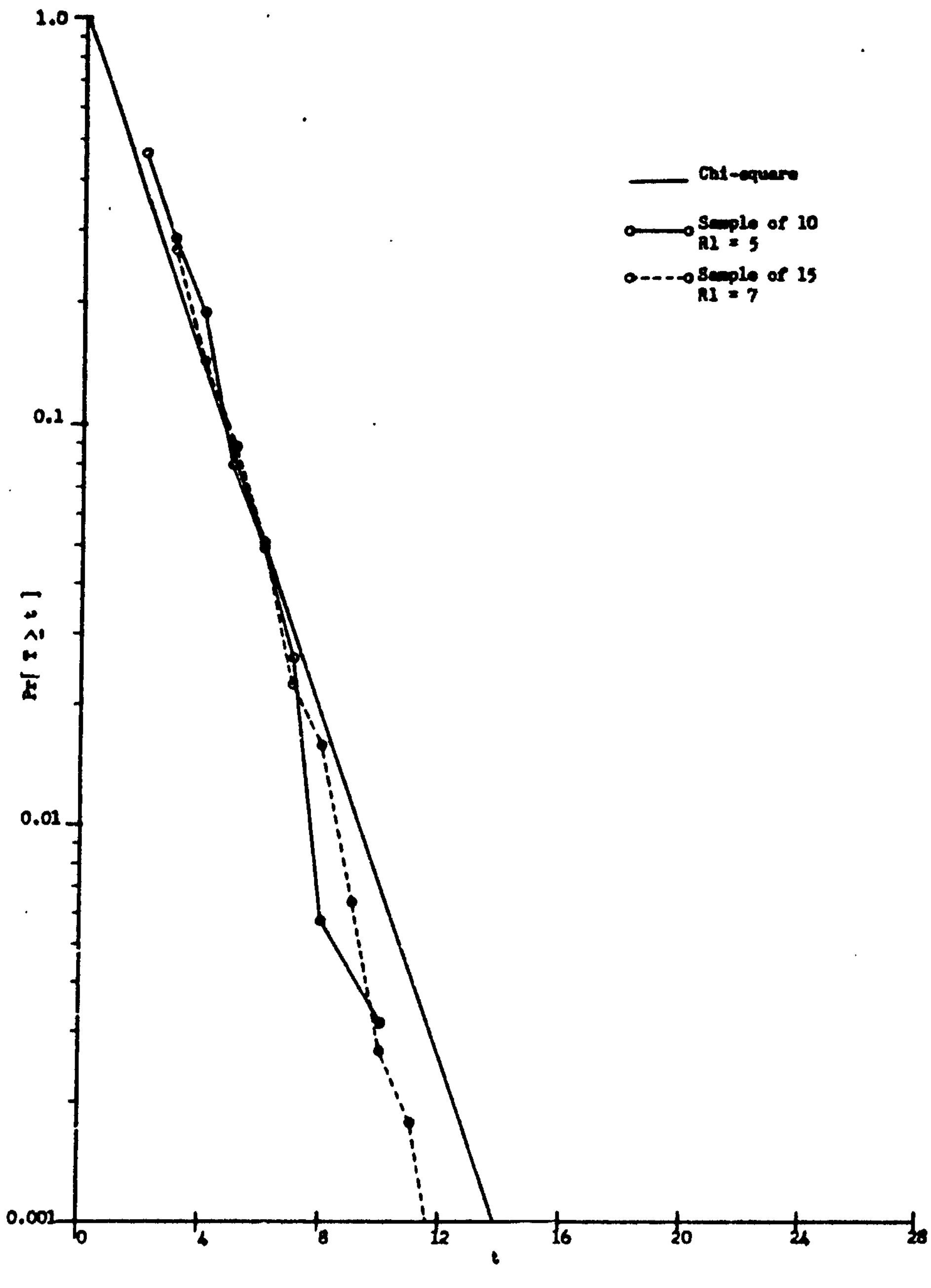

Figure 5.6 Exact null distribution for a 2×3 contingency table.
H_0: $\{ p_{\cdot j} = 1/3,\ j = 1, 2, 3 \}$
2 degrees of freedom and row marginal sums are fixed.

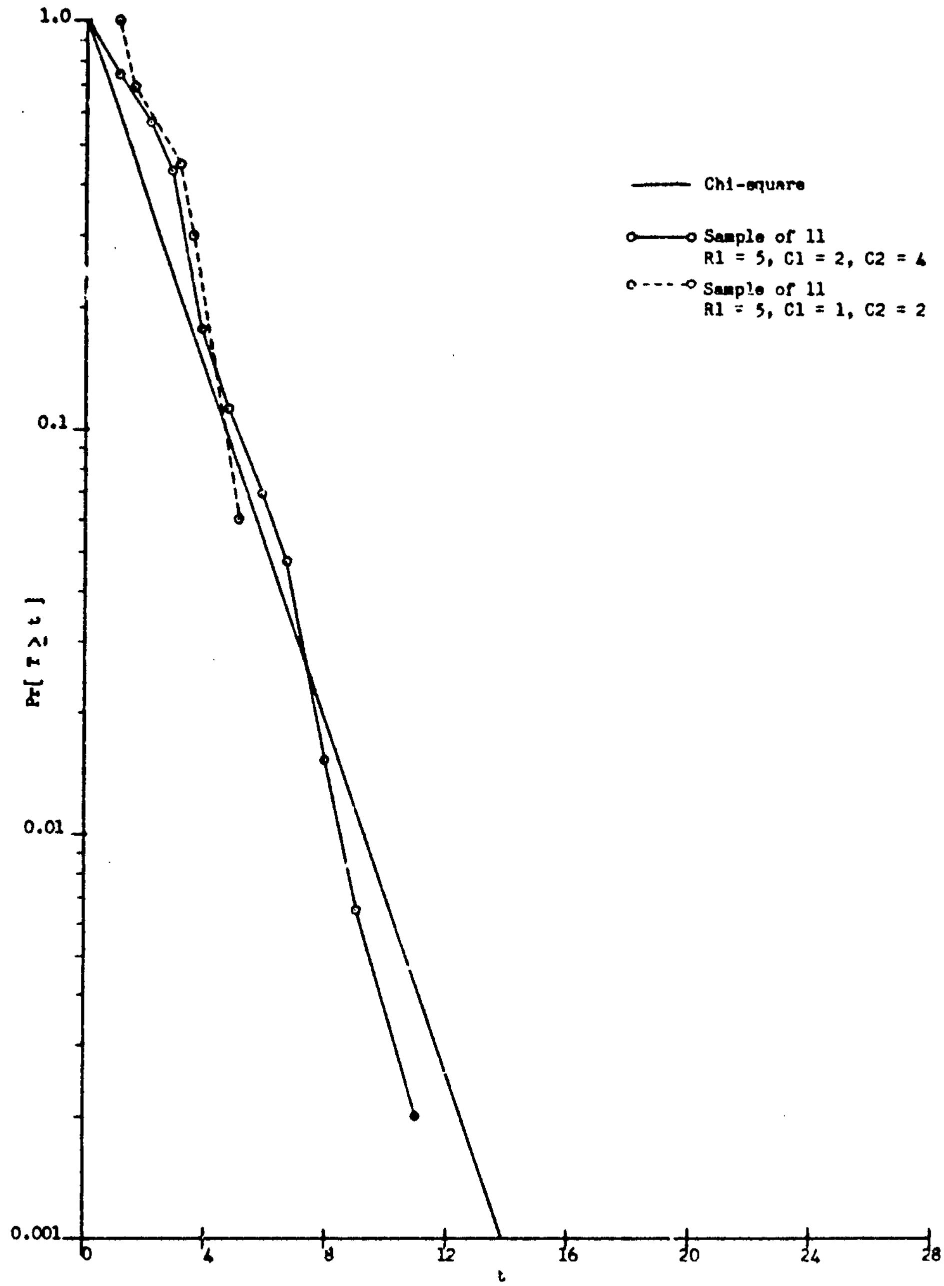

Figure 5.7 Exact null distribution for a 2 x 3 contingency table. 2 degrees of freedom and both row and column marginal sums are fixed.

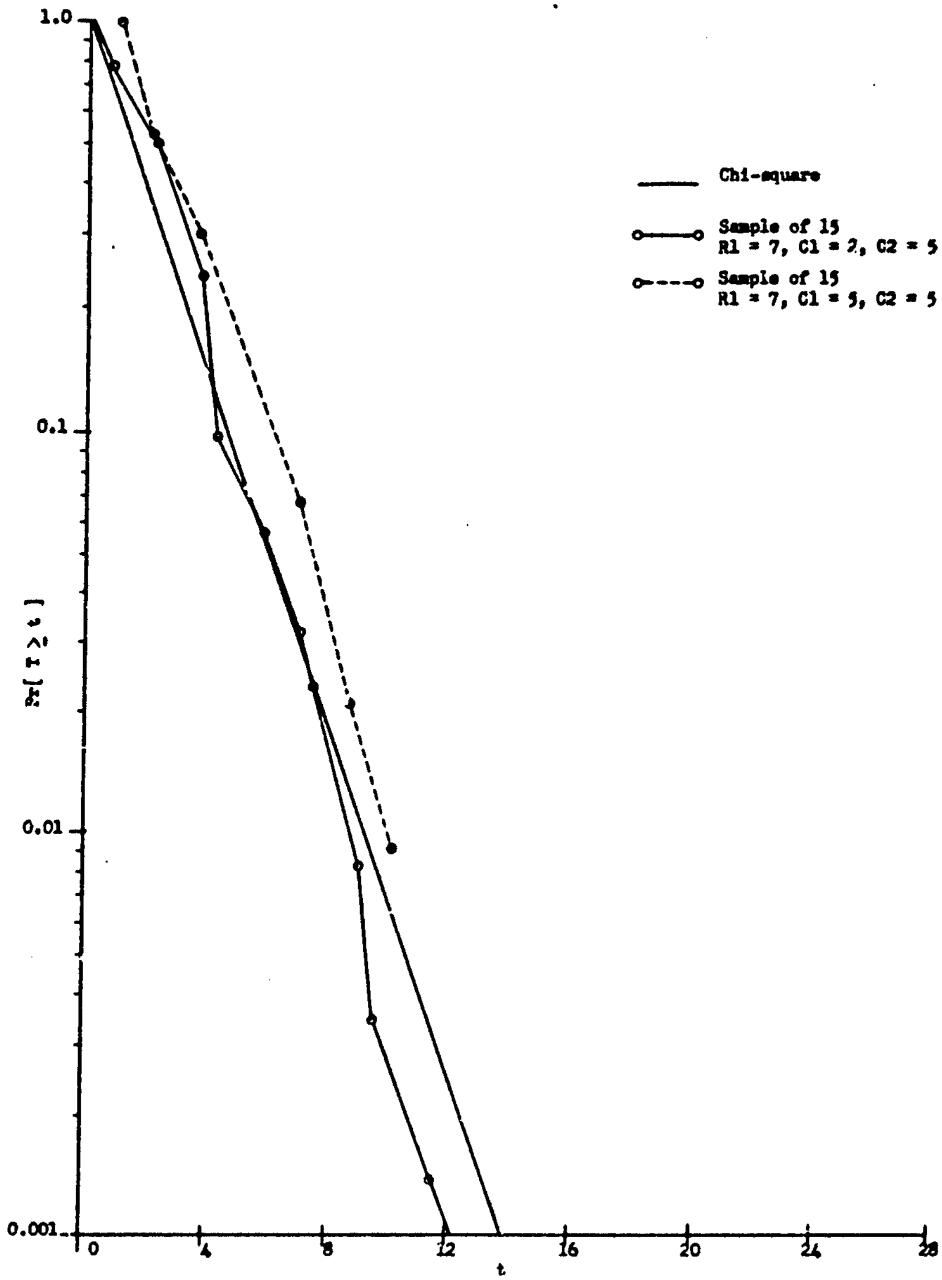

Figure 5.8 Exact null distribution for a 2 x 3 contingency table. 2 degrees of freedom and both row and column marginal sums are fixed.

REFERENCES

[1] ANDREWS, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance. Ann. Math. Statist. 25 724-735.

[2] BARTON, D. E. (1957). A comparison of two sorts of test for a change of location applicable to truncated data. J. Roy. Statist. Soc. Ser. B 19 119-124.

[3] CHAKRAVARTI, I. M., LEONE, F. C. and ALANEN, J. D. (1961). Relative efficiency of Mood's and Massey's two-sample tests against some parametric alternatives. Bull. Inst. Internat. Statist. 33rd Session (Paris) 110 1-9.

[4] COCHRAN, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Statist. 23 315-345.

[5] CRAMER, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton.

[6] DIXON, W. J. (1954). Power under normality of several non-parametric tests. Ann. Math. Statist. 25 610-614.

[7] FRAZER, D. A. S. (1957). Nonparametric Methods in Statistics, John Wiley and Sons, Inc., New York.

[8] HAYNAM, G. E. and LEONE, F. C. (1962). Tables of the non-central chi-square distribution. Computing Center and Statistical Laboratory Report No. 1060, Case Institute of Technology, Cleveland.

[9] HOEFFDING, W. and ROSENBLATT, J. R. (1955). The efficiency of tests. Ann. Math. Statist. 26 52-63.

[10] JOHNSON, N. L. (1959). On an extension of the connection between Poisson and χ^2 distributions. Biometrika 46 352-363.

[11] LEHMANN, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. Ann. Math. Statist. 22 165-179.

[12] LEHMANN, E. L. (1953). The power of rank tests. Ann. Math. Statist. 24 23-43.

[13] 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. Inst. Internat. Statist. 33rd Session (Paris) 115 1-11.

[14] 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.

[15] MASSEY, F. J. (1951). A note on a two sample test. Ann. Math. Statist. 22 304-306.

[16] MOOD, A. M. (1951). Introduction to the Theory of Statistics, McGraw Hill, New York.

[17] MOOD, A. M. (1954). On the asymptotic efficiency of certain nonparametric two sample tests. Ann. Math. Statist. 25 514-521.

[18] NOETHER, G. E. (1950). Asymptotic properties of the Wald-Wolfowitz test of randomness. Ann. Math. Statist. 21 231-246.

[19] PATNAIK, P. B. (1949). The noncentral χ^2 and F distributions and their applications. Biometrika 36 202-232.

[20] PEARSON, KARL (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag. Ser. 5 50 157-172.

[21] ROY, S. N. and MITRA, S. K. (1956). An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis. Biometrika 43 361-376.

[22] WILCOXON, F. (1945). Individual comparisons by ranking methods. Biometrics 1 80-83.