PERMUTATION TESTS OF BIVARIATE INTERCHANGEABILITY

Approved by:

Dr. William R. Schucany

Dr. Richard F. Gunst

Dr. Rudy Guerra

Dr. Warren E. Ferguson, Jr.

PERMUTATION TESTS OF BIVARIATE INTERCHANGEABILITY

A Dissertation Presented to the Graduate Faculty of

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistical Science

by

Michael Duane Ernst

(B.A., St. Cloud State University)
(M.S., Southern Methodist University)

May 17, 1997

ACKNOWLEDGMENTS

This dissertation is the result of much more than just my time spent at SMU; it is the culmination of many years of hard work and persistence. I never would have accomplished this much without the support and encouragement of the many wonderful people I have encountered along the way.

First and foremost, I would like to thank my good friend and advisor Dr. William Schucany. His invaluable insight and advice as a colleague is surpassed only by his thoughtfulness and caring as a human being. It has been an immense pleasure to work with him.

I would also like to thank the members of my committee, Dr. Richard Gunst, Dr. Rudy Guerra, and Dr. Warren Ferguson, for their kind advice and thoughtful reading of this dissertation. In addition, the entire Department of Statistical Science deserves thanks for making my time at SMU so enjoyable. In particular, I am very grateful to Dr. Alan Polansky and Dr. Greg Miller for their friendship and encouragement.

I have had many teachers along the way that encouraged my mathematical interests and provided a wonderful environment to learn. Most prominent among this group are my parents, Charles and Patricia Ernst. They have my sincerest thanks. In addition, I would like to thank Dr. Robert Dumonceaux, Dr. Robert Earles, Dr. Gary Buls, and Dr. Richard Sundheim. Special thanks go to Dr. James Kepner, who is largely responsible for the direction my career has taken. For helping me keep a balance in my life, by encouraging my less mathematical interests, I wish to thank Dr. Kenton Frohrip and the rest of the Department of Music at St. Cloud State University.

I am deeply indebted to my parents and my entire family for providing me with such a supportive and loving environment in which to grow. Their support has made everything that I have accomplished possible. Finally, I wish to thank my wife, Bonnie. Her help during the stressful times in my life has been a blessing. As we look forward to the new challenges in our lives, I am glad that she will be there with me.

Ernst, Michael Duane B.A., St. Cloud State University, 1992
M.S., Southern Methodist University, 1994

Permutation Tests of Bivariate Interchangeability

Advisor: Professor William R. Schucany

Doctor of Philosophy degree conferred May 17, 1997

Dissertation completed May 9, 1997

Bivariate data are common in many settings, for example in measurements on pairs of siblings or when measurements are made before and after a treatment is applied to experimental units. Often it is of interest to determine if there is a location difference between the marginal variables, but occasionally we may wish to detect a difference in scale as well. Without relying on the usual parametric assumptions, we propose a set of permutation tests to simultaneously detect differences in either location or scale that is both exact and distribution-free.

The proposed permutation tests take advantage of the fact that only under the null hypothesis of equal means and variances are the pairwise differences symmetrically distributed about zero and the differences and pairwise sums uncorrelated. Two statistics that measure the marginal location and scale differences are combined in a quadratic form. The observed quadratic form is then compared to its value under all 2^n (conditionally) equally likely sign changes on the differences. The null hypothesis is rejected if the observed value is larger than a pre-specified proportion of the permutation distribution.

Several methods of estimating the covariance matrix of the quadratic form are considered including conditional and unconditional (or plug-in) approaches. Through a simulation study, these new tests are compared with the standard tests in the literature for several families of bivariate distributions and are found to perform quite favorably.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

For my parents

CHAPTER I

INTRODUCTION

1.1 Bivariate Interchangeability

The bivariate random variable $\mathbf{X} = (X_1, X_2)'$ is said to possess bivariate symmetry [also called bivariate interchangeability by Sen (1967)] if its cumulative distribution function (cdf) $F(\,\cdot\,,\,\cdot\,)$ satisfies

$$F(x_1, x_2) = F(x_2, x_1) \qquad [\![1.1]\!]$$

for every pair of real numbers (x_1, x_2). An important consequence of bivariate interchangeability is that X_1 and X_2 have the same marginal distributions, denoted by $X_1 \stackrel{d}{=} X_2$.

To introduce location and scale parameters, suppose that $\mathbf{X}$ has cdf $F\left(\frac{x_1-\mu_1}{\sigma_1}, \frac{x_2-\mu_2}{\sigma_2}\right)$. Without loss of generality, also assume that

$$\iint\limits_{\mathbb{R}^2} x_i \, dF(x_1, x_2) = 0 \text{ and } \iint\limits_{\mathbb{R}^2} x_i^2 \, dF(x_1, x_2) = 1$$

for $i = 1$ and 2. Then, the expectation and variance of $\mathbf{X}$ is $\mathbb{E}[\mathbf{X}] = \boldsymbol{\mu} = (\mu_1, \mu_2)'$ and $\mathbb{V}[\mathbf{X}] = \mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, respectively, where $|\rho| \leq 1$. The null hypothesis of interest is that the marginal distributions of $\mathbf{X}$ are identical (that is, $\mathbf{X}$ possesses bivariate interchangeability). This is stated as

$$\mathcal{H}_0\colon \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2, \text{ and } F(\,\cdot\,,\,\cdot\,) \text{ satisfies } [\![1.1]\!].$$

The alternative hypothesis is

$$\mathcal{H}_1\colon \mu_1 \neq \mu_2 \text{ or } \sigma_1^2 \neq \sigma_2^2,$$

that the marginals of **X** differ in either location or scale (or both). It is important to note that under $\mathcal{H}_0$, $X_1 \overset{d}{=} X_2$, yet X_1 and X_2 may still be (and probably are) correlated.

The approach taken here is to parameterize the problem in terms of location and scale parameters. Bivariate interchangeability is required to be true for $F(\,\cdot\,,\,\cdot\,)$ under $\mathcal{H}_0$. While $F(\,\cdot\,,\,\cdot\,)$ does not need to satisfy ⟦1.1⟧ under $\mathcal{H}_1$, the specific alternatives to bivariate interchangeability being tested are those of location and scale differences.

An alternative approach, given by Hollander (1971) and modified by Koziol (1979), tests the more general hypothesis ⟦1.1⟧ directly using empirical cdfs. This approach is less sensitive to the particular location and scale alternatives of interest in this dissertation (see Kepner and Randles 1984).

Most other approaches, including those given here, are based on transforming the paired data, $\mathbf{X}_i = (X_{1i}, X_{2i})'$, into their pairwise differences and sums,

$$\mathbf{Y}_i = (X_{1i} - X_{2i}, X_{1i} + X_{2i})' = (D_i, S_i)',$$

for $i = 1, \ldots, n$. Clearly, a suitable function of the D_is (such as $\overline{D}$) can be used to detect differences in marginal location since $\mathbb{E}[D_i] = \mu_1 - \mu_2$.

The difference in scale can be judged by measuring the covariance between the differences and sums since, for any bivariate distribution with finite second moments,

$$\begin{aligned}\mathbb{C}[D_i, S_i] &= \mathbb{C}[X_{1i} - X_{2i}, X_{1i} + X_{2i}] \\ &= \mathbb{C}[X_{1i}, X_{1i}] + \mathbb{C}[X_{1i}, X_{2i}] - \mathbb{C}[X_{2i}, X_{1i}] - \mathbb{C}[X_{2i}, X_{2i}] \\ &= \sigma_1^2 - \sigma_2^2.\end{aligned} \qquad ⟦1.2⟧$$

Hence, the marginal variances are equal if and only if the differences and sums are uncorrelated.

1.2 Tests Under Bivariate Normality

Taking very different paths, Pitman (1939) and Morgan (1939) each used ⟦1.2⟧ to suggest a test of equal variances for bivariate normal data that is essentially the test for zero correlation between the differences and sums.

Prompted by Morgan's use of "the likelihood ratio method … of Neyman and Pearson" (Morgan 1939, pp. 13-14), Hsu (1940) derived the likelihood ratio test, and its distribution, for several other hypotheses involving parameters of the bivariate normal distribution, including the hypothesis of interest here.

Bradley and Blackwood (1989) derived a test of $\mathcal{H}_0$ under bivariate normality by considering the regression of the differences on the sums. The F test on 2 and $n-2$ degrees of freedom that simultaneously tests the significance of the slope and intercept of this regression is a valid test of $\mathcal{H}_0$. It is shown in Section A.4 that this F test is equivalent to the likelihood ratio test of Hsu (1940), whose work has apparently been largely overlooked in the literature.

Bell and Haller (1969), also seemingly unaware of Hsu's work, proposed another approach for bivariate normal data which combines the results from two independent (under $\mathcal{H}_0$) tests. Their method rejects $\mathcal{H}_0$ if either the mean of the differences, called "D-bar" in Figure 1.1, or the sample correlation between the differences and sums is significantly different from zero (their expectation under $\mathcal{H}_0$). If these two tests are conducted at the α_1 and α_2 levels, respectively, then the overall level of the test can be calculated as $\alpha = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$. This approach leads to a rectangular non-rejection region with respect to the mean of the differences and the correlation between the differences and sums, whereas Hsu's likelihood ratio test has an elliptical non-rejection region. These non-rejection regions are shown in Figure 1.1, where the overall significance level of both tests is $\alpha = .05$. The significance levels for each of Bell and Haller's individual tests are $\alpha_1 = \alpha_2 = .02532$.

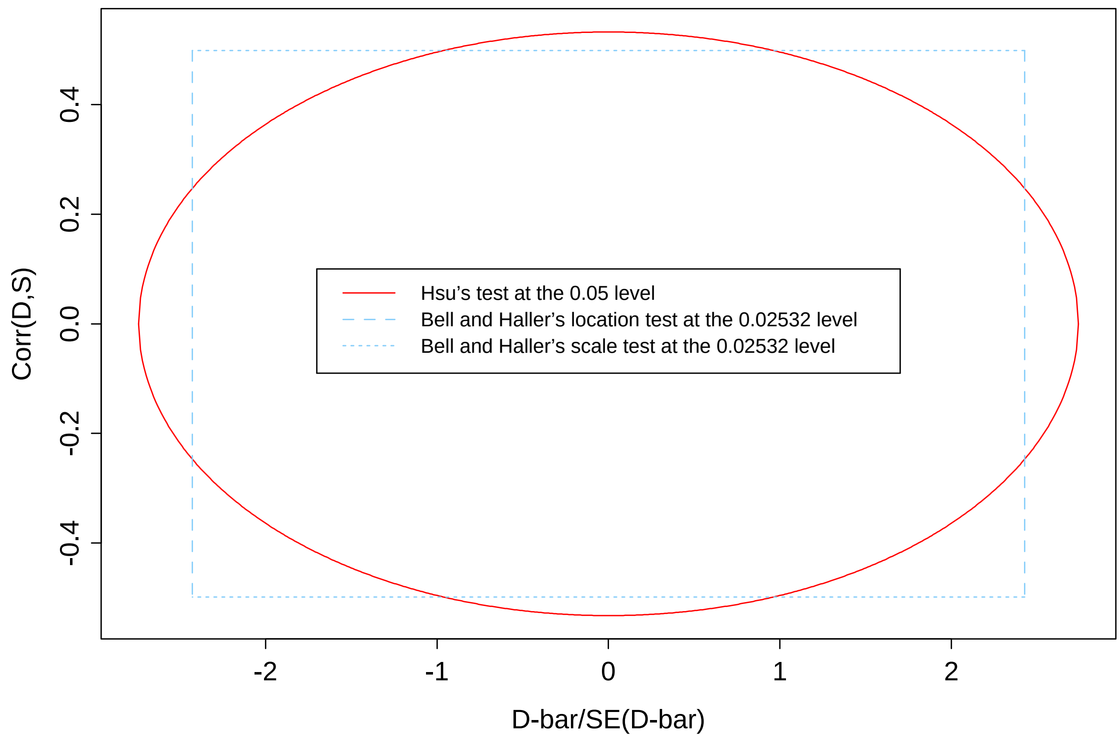

Figure 1.1. Non-rejection regions for Hsu's test (red) and Bell and Haller's test (blue), $\alpha = .05, \alpha_1 = \alpha_2 = .02532$.

1.3 Distribution-Free Tests

Conditionally distribution-free tests can be constructed by finding permutations of the observed data that, under the null, are equally likely. A test statistic can be calculated on all of the (conditionally) equally likely permutations of the data providing a reference distribution that can be used to determine the significance of the test statistic calculated on the observed permutation of the data. Fisher (1936, 1935, sec. 21) provides some outstanding discussions of this idea. A good treatment can be found in Bradley (1968, chap. 4) as well.

A simple example of this method is the well known sign test, which was used as early as 1710 when Arbuthnott (1710) attributed the rather equal proportion of male/female births not to chance, but to Divine Providence. The combinatorial mathematics needed for this argument can be traced back much earlier. A good historical discussion is given by Bradley (1968, chap. 1).

Several nonparametric tests have been proposed that exploit the fact that if $\mathcal{H}_0$ is true, then conditional on the n observed pairs, the value of $\mathbf{X}_i$ will be $(X_{1i}, X_{2i})'$ or $(X_{2i}, X_{1i})'$ with equal probability for each $i = 1, \ldots, n$ (see Kepner and Randles 1982). Hence, there are 2^n possible (conditionally) equally likely realizations of the observed data, two for each of the n pairs. This provides a permutation distribution with which the evidence about $\mathcal{H}_0$ can be evaluated. In other words, this is a frame of reference for a test statistic provided that $\mathcal{H}_0$ is true.

Sen (1967) proposed a class of rank tests that combine two test statistics in a quadratic form, one that measures the location difference and one that measures the scale difference. If R_{1i} is the rank of X_{1i} among all $2n$ observations, $X_{ji}, j = 1, 2,$ $i = 1, \ldots, n,$ then $S_1 = \sum_{i=1}^{n} R_{1i}$ measures the marginal location difference. To measure the marginal scale difference an appropriate rank statistic may be used. The statistic $S_2 = \sum_{i=1}^{n} \left| R_{1i} - \frac{2n+1}{2} \right|$ is due to Ansari and Bradley (among others) and $S_2 = \sum_{i=1}^{n} \left(R_{1i} - \frac{2n+1}{2} \right)^2$ is due to Mood. The test statistic is $S = \mathbf{S}' \boldsymbol{\Lambda}_S^{-1} \mathbf{S}$, where $\mathbf{S} = \binom{S_1}{S_2} - \mathbb{E}\left[\binom{S_1}{S_2} \middle| \mathcal{H}_0 \right]$ and $\boldsymbol{\Lambda}_S = \mathbb{V}[\mathbf{S} \,|\, \mathcal{H}_0]$. If the observed value of S is the m^{th} largest in the permutation distribution, then $\mathcal{H}_0$ can be rejected at the $m/2^n$ significance level.

Kepner and Randles (1984) proposed a test that is calculated from the pairwise differences and sums of the data. The appropriate permutation distribution is generated by noticing that $\mathbf{Y}_i$ will be $(D_i, S_i)'$ or $(-D_i, S_i)'$ with equal probability for each $i = 1, \ldots, n$. These two possibilities correspond directly to the interchange of X_{1i} and X_{2i}. Similar to Sen (1967), Kepner and Randles (1984) combine two appropriate statistics in a quadratic form. Their location statistic is the Wilcoxon signed-rank statistic,

$K_1 = \sum_{i=1}^{n} \phi(D_i) R_i$, where R_i is the rank of $|D_i|$ among $|D_1|, \ldots, |D_n|$ and ϕ is the indicator function

$$\phi(t) = \begin{cases} 0 & \text{for } t \leq 0 \\ 1 & \text{for } t > 0. \end{cases}$$

They combine this with a statistic designed to detect differences in scale. Utilizing $[\![1.2]\!]$, Kepner and Randles (1982) previously proposed using an estimate of Kendall's τ to test for zero correlation between the differences and sums as a test for equal marginal scales. Hence, their statistic for detecting scale differences is

$$K_2 = \sum_{1 \leq i < j \leq n} \sum \phi(D_j - D_i) \phi(S_j - S_i).$$

The quadratic form used to test $\mathcal{H}_0$ is $K = \mathbf{K}' \mathbf{\Lambda}_K^{-1} \mathbf{K}$, where $\mathbf{K} = \binom{K_1}{K_2} - \mathbb{E}\left[\binom{K_1}{K_2} \middle| \mathcal{H}_0\right]$ and $\mathbf{\Lambda}_K = \mathbb{V}[\mathbf{K} \,|\, \mathcal{H}_0]$. If the observed value of K is the m^{th} largest in the permutation distribution, then $\mathcal{H}_0$ can be rejected at the $m/2^n$ significance level.

Generally $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_K$ will need to be estimated. The most straightforward approach is to calculate the conditional estimates of $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_K$ under the permutation distribution. That is, calculate the sample versions of $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_K$ from the 2^n values of $\mathbf{S}$ and $\mathbf{K}$. Hence, if $\mathbf{S}_1, \ldots, \mathbf{S}_{2^n}$ are the 2^n values of $\mathbf{S}$ calculated from the permutation distribution, then define $\widehat{\mathbf{\Lambda}}_S = \frac{1}{2^n} \sum_{i=1}^{2^n} \mathbf{S}\mathbf{S}'$. Define $\widehat{\mathbf{\Lambda}}_K$ similarly.

These variance-covariance matrices can be estimated unconditionally as well. Using the asymptotic theory of *U*-statistics, Kepner and Randles (1984) give consistent estimates of $\mathbf{\Lambda}_S$ and $\mathbf{\Lambda}_K$. Unfortunately, these estimates do not always produce positive definite variance-covariance matrices.

1.4 Permutation Tests of Bivariate Interchangeability

Several of the existing tests of bivariate interchangeability assume bivariate normality of the data. This is often a very poor assumption and the need for distribution-free tests is clear. The (conditionally) distribution-free tests in the literature are rank based tests and their small sample properties are unknown. This dissertation introduces a new class of conditionally distribution-free tests of bivariate interchangeability. These tests are permutation tests based on the original data.

Chapter 2 presents the new tests and many of their theoretical properties. Chapter 3 gives some examples that provide insight into how the tests perform in practice. This chapter also includes the results of a simulation study comparing the small sample performance of the new permutation tests with the other tests in the literature. Chapter 4 draws some conclusions from the current research and points to some related topics for further investigation. The proofs of all results are relegated to the Appendix.

CHAPTER II

PERMUTATION TESTS OF BIVARIATE INTERCHANGEABILITY

This chapter introduces a class of permutation tests of bivariate interchangeability. Section 2.1 provides some motivation for these tests while Section 2.2 describes the new tests. Section 2.3 considers some problems in estimating nuisance parameters and Section 2.4 describes the large sample behavior of these tests.

2.1 Detecting Marginal Differences in Bivariate Data

There are several alternatives to $\mathcal{H}_0$ that could be considered when developing a test. It is instructive to consider some of the more standard approaches here.

2.1.1 Location Differences

If the alternative to $\mathcal{H}_0$ of interest is simply a difference in location ($\mu_1 \neq \mu_2$), then the appropriate permutation test would be to compare the observed value of $U_1 = \overline{D}$ to its values under all 2^n interchanges of X_{1i} and $X_{2i},\ i = 1, \ldots, n$ (see Fisher 1935, sec. 21). This corresponds to calculating the mean of the differences under all 2^n sign changes of the D_i, $i = 1, \ldots, n$. If U_1 is extreme in this permutation distribution, then $\mathcal{H}_0$ is rejected in favor of the alternative that $\mu_1 \neq \mu_2$.

Other tests commonly used in this situation include the sign test and the Wilcoxon signed-rank test. These are permutation tests based on transformations of the data, such as ranks. If the underlying distribution is assumed to be bivariate normal, the paired-sample t test is the appropriate procedure.

2.1.2 Scale Differences

If the alternative of interest is that the marginal variances differ ($\sigma_1^2 \neq \sigma_2^2$), then an appropriate permutation test can be constructed using $U_2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)\left(S_i - \overline{S}\right)$, the sample covariance between the differences and sums. It was shown in [[1.2]] that $\mathbb{C}[D_i, S_i] = \sigma_1^2 - \sigma_2^2$ so that under $\mathcal{H}_0$ $\mathbb{E}[U_2] = 0$. Hence, if U_2 is extreme in the permutation distribution found by recalculating U_2 for all 2^n sign changes of the differences, then $\mathcal{H}_0$ is rejected in favor of the alternative that $\sigma_1^2 \neq \sigma_2^2$.

Another approach, by Kepner and Randles (1982), uses an estimate of Kendall's τ to measure the degree of correlation between the differences and sums. The Pitman-Morgan test (Kotz, Johnson, and Read 1985, pp. 739-740) is designed for bivariate normal data. Morgan (1939) and Pitman (1939) each derived the distribution of the sample correlation coefficient between the differences and sums under bivariate normality.

2.1.3 Location and Scale Differences

As the previous two sections point out, the two statistics U_1 and U_2 are sensitive to differences in marginal location and scale, respectively. This pair, $\mathbf{U} = (U_1, U_2)'$, can be used to simultaneously detect differences in marginal location or scale. Recalculating $\mathbf{U}$ for each of the 2^n sign changes of the differences will result in a permutation distribution of 2^n points in $\mathbb{R}^2$. This permutation distribution is shown in Figure 2.1 for the data set consisting of ten pairs of observations displayed in Figure 2.2(a). This permutation distribution contains $2^{10} = 1024$ points. Figure 2.2(b) displays the differences and sums from the data in Figure 2.2(a) that are used to calculate U_1 and U_2.

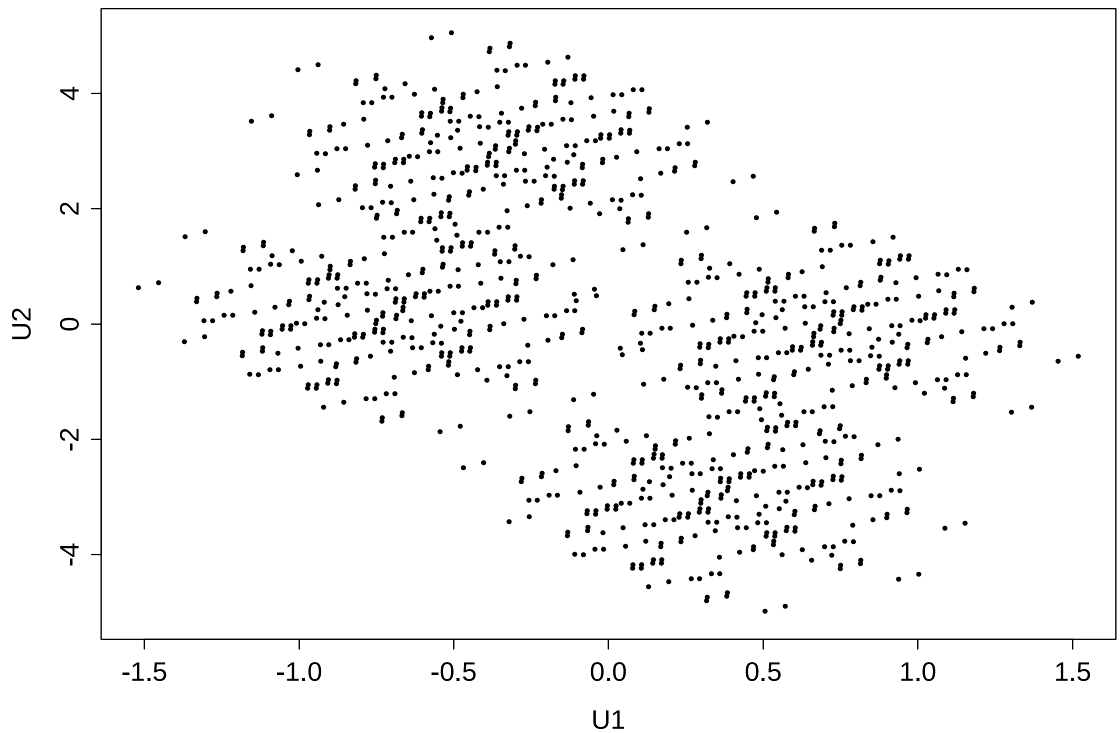

Figure 2.1. Permutation distribution of **U** for the data in Figure 2.2(a).

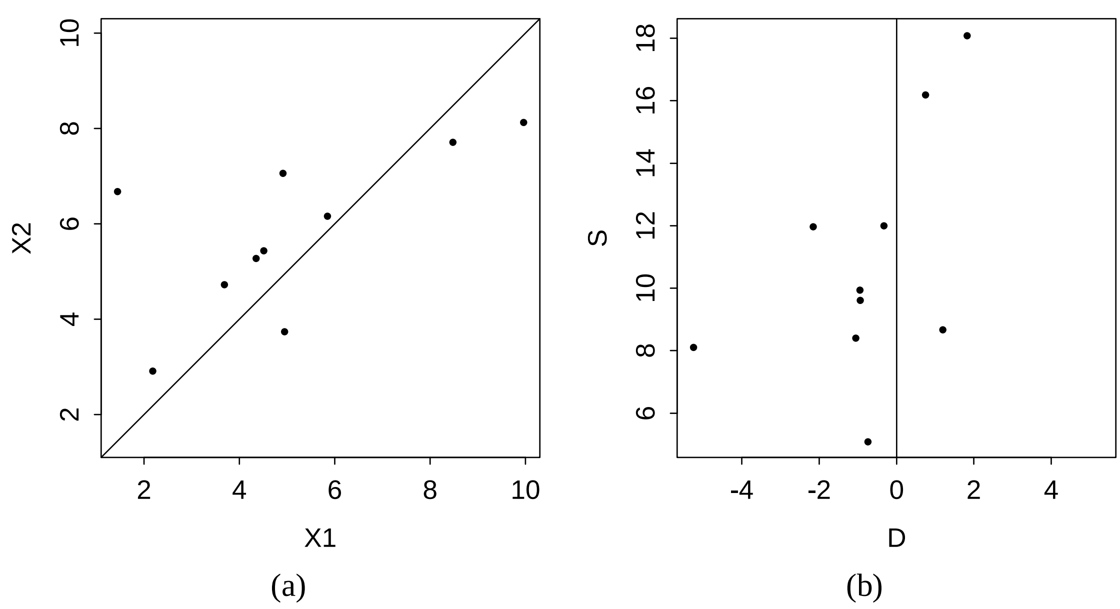

Figure 2.2. Scatterplots of (a) the random sample used to generate Figure 2.1 and (b) the same data transformed to differences and sums.

The same testing principle can be applied here; $\mathcal{H}_0$ is rejected if $\mathbf{U}$ is "extreme" in this permutation distribution. Under $\mathcal{H}_0$ $\mathbb{E}[\mathbf{U}] = \mathbf{0}$, hence the farther $\mathbf{U}$ is from $\mathbf{0}$ compared to the other points in the permutation distribution, the more "extreme" it is. The question that arises is how to measure each point's distance from the origin. This is addressed in the next section.

2.2 Combining U_1 and U_2

The two statistics U_1 and U_2 can be combined into one statistic by considering a distance measure from the origin. A statistical measure of distance, and the one employed here, is an estimated Mahalanobis distance given by the quadratic form $E = \mathbf{U}'\widehat{\mathbf{\Gamma}}^{-1}\mathbf{U}$, where $\widehat{\mathbf{\Gamma}}$ is an estimate of $\mathbf{\Gamma} = \mathbb{V}[\mathbf{U}]$, the 2×2 variance-covariance matrix of $\mathbf{U}$.

The following notation will prove helpful in constructing a reference distribution for E and will be used extensively throughout this dissertation. The binary expansion of an integer index between 0 and $2^n - 1$ consists of n digits, the i^{th} of which can be used to identify the sign of D_i. This representation is useful for identifying each of the 2^n permutations of the data.

Define $\mathfrak{b}_i^k$ to be the i^{th} digit (from the right) in the unique binary representation of the integer k. More specifically, $\mathfrak{b}_i^k = \omega(\lfloor k/2^{i-1} \rfloor)$, where $\lfloor x \rfloor$ is the integer part of x and ω is the indicator function

$$\omega(t) = \begin{cases} 0 & \text{for } t \text{ even} \\ 1 & \text{for } t \text{ odd.} \end{cases}$$

As a function of $\mathfrak{b}_i^k$, k can then be written as $k = \sum_{i=1}^{\infty} 2^{i-1}\mathfrak{b}_i^k$. For example, the binary representation of $k = 13$ is 1101. See Section A.1 for more details and illustrations.

Now, let

$$\mathbf{Y} = [\,\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_n\,] = \begin{bmatrix} D_1 & D_2 & \cdots & D_n \\ S_1 & S_2 & \cdots & S_n \end{bmatrix}$$

be the 2 n data matrix of observed differences and sums. In a similar fashion, define

$$\mathbf{Y}_{\langle k \rangle} = \begin{bmatrix} (-1)^{\mathfrak{b}_1^k} D_1 & (-1)^{\mathfrak{b}_2^k} D_2 & \cdots & (-1)^{\mathfrak{b}_n^k} D_n \\ S_1 & S_2 & \cdots & S_n \end{bmatrix}$$

for $k = 0, 1, \ldots, 2^n - 1$. Notice that $\mathbf{Y}_{\langle 0 \rangle}$ ($\mathfrak{b}_i^0 = 0, i = 1, \ldots, n$) is the observed value of the random variable $\mathbf{Y}$. Then $\mathcal{Y} = \{\mathbf{Y}_{\langle k \rangle} \mid k = 0, 1, \ldots, 2^n - 1\}$ forms the set of all equally likely data matrices conditional on the observed data matrix $\mathbf{Y}_{\langle 0 \rangle}$. It should be noted that the sums remain unchanged for every permutation of the data.

Next define $\mathbf{U}_{\langle k \rangle}$ to be the statistic $\mathbf{U}$ calculated from $\mathbf{Y}_{\langle k \rangle}$, the k^{th} permutation of the observed data. This can be written as

$$\mathbf{U}_{\langle k \rangle} = \begin{bmatrix} U_{1\langle k \rangle} \\ U_{2\langle k \rangle} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} (-1)^{\mathfrak{b}_i^k} D_i \\ \frac{1}{n-1} \sum_{i=1}^{n} (-1)^{\mathfrak{b}_i^k} D_i \left(S_i - \overline{S}\right) \end{bmatrix}. \qquad [\![2.1]\!]$$

Notice that $U_{2\langle k \rangle}$, the sample covariance between the permuted differences and sums, can be written without using the mean of the permuted differences. Now let $\mathcal{U} = \{\mathbf{U}_{\langle k \rangle} \mid k = 0, 1, \ldots, 2^n - 1\}$ be the set of 2^n conditionally equally likely outcomes of $\mathbf{U} = (U_1, U_2)'$. The 2^{10} points in Figure 2.1 form such a permutation set of values of $\mathbf{U}$. Again, $\mathbf{U}_{\langle 0 \rangle}$ is the observed value of the random variable $\mathbf{U}$.

In the next section, methods of estimating $\mathbf{\Gamma}$ will be considered. Estimates of $\mathbf{\Gamma}$ are denoted by $\widehat{\mathbf{\Gamma}}_M$, where M identifies the method used to estimate $\mathbf{\Gamma}$. Values of M will be discussed in the next section.

For a particular estimate of $\mathbf{\Gamma}$ the distance from a point in $\mathcal{U}$ to the origin is determined by $E_{M\langle k \rangle} = \mathbf{U}'_{\langle k \rangle} \widehat{\mathbf{\Gamma}}_M^{-1} \mathbf{U}_{\langle k \rangle}, k = 0, 1, \ldots, 2^n - 1$. The conditional permutation distribution of E_M is given by $\mathcal{E} = \{E_{M\langle k \rangle} \mid k = 0, 1, \ldots, 2^n - 1\}$. Again, $E_{M\langle 0 \rangle}$ is the observed value of the random variable E_M.

As noted earlier, values of $\mathbf{U}$ far from the origin are evidence against $\mathcal{H}_0$. Since E_M measures the Mahalanobis distance from $\mathbf{U}$ to the origin, large values of E_M are evidence in favor of $\mathcal{H}_1$. Conditional on the observed data, the elements of $\mathcal{Y}, \mathcal{U}$, and $\mathcal{E}$ are equally likely under $\mathcal{H}_0$. Hence for each $k = 0, 1, \ldots, 2^n - 1$, $\mathbb{P}_{\mathcal{E}}[E_M = E_{M\langle k \rangle} | \mathcal{H}_0] = 2^{-n}$, where the probability is taken with respect to $\mathcal{E}$, the permutation distribution of E_M. A conditionally distribution-free test of $\mathcal{H}_0$ with exact nominal level $\alpha = m/2^n$ is obtained by rejecting $\mathcal{H}_0$ when $E_{M\langle 0 \rangle}$ is one of the m largest elements in $\mathcal{E}$, where $m \in \{1, 2, 3, \ldots, 2^n\}$.

Another interpretation of this test can be seen by reconsidering Figure 2.1 which graphically depicts $\mathcal{U}$, the permutation distribution of $\mathbf{U}$. Out of the 2^n points in $\mathcal{U}$, if $\mathbf{U}_{\langle 0 \rangle}$, the observed value of $\mathbf{U}$, is one of the m furthest points from the origin (measured in Mahalanobis distance) then $\mathcal{H}_0$ is rejected.

2.3 Estimating $\boldsymbol{\Gamma}$

The test proposed in the previous section depends on $\boldsymbol{\Gamma}$, the variance-covariance matrix of $\mathbf{U}$. Since in practice $\boldsymbol{\Gamma}$ will not be known, it is necessary to estimate it from the data. In this section, several methods of estimating $\boldsymbol{\Gamma}$ will be considered and some properties of the test statistic E_M will be described.

2.3.1 A Conditional Estimate of $\boldsymbol{\Gamma}$

One estimate of $\boldsymbol{\Gamma}$ can be based on $\mathcal{U}$, the permutation distribution of $\mathbf{U}$. Conditional on the observed data, $\mathcal{U}$ provides a distribution from which an estimate of $\boldsymbol{\Gamma}$ can be calculated directly. Define the conditional estimate of $\boldsymbol{\Gamma}$ by

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}_C &= \mathbb{V}_{\mathcal{U}}[\mathbf{U}] = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \Big(\mathbf{U}_{\langle k \rangle} - \mathbb{E}_{\mathcal{U}}[\mathbf{U}] \Big) \Big(\mathbf{U}_{\langle k \rangle} - \mathbb{E}_{\mathcal{U}}[\mathbf{U}] \Big)' \\ &= \frac{1}{2^n} \sum_{k=0}^{2^n-1} \mathbf{U}_{\langle k \rangle} \mathbf{U}'_{\langle k \rangle}.\end{aligned} \qquad [\![2.2]\!]$$

By considering $\mathcal{U}$ as the conditional population from which $\mathbf{U}$ is drawn, $\widehat{\boldsymbol{\Gamma}}_C$ is the conditional population covariance matrix of $\mathbf{U}$. The fact that $\mathbb{E}_{\mathcal{U}}[\mathbf{U}] = \mathbf{0}$ can be shown easily with the following theorem.

Theorem 2.1

For any real numbers $a_1, a_2, \ldots, a_n, c_1, c_2, \ldots, c_n$,

(a) $\displaystyle\sum_{k=0}^{2^n-1} \sum_{i=1}^{n} a_i (-1)^{\mathfrak{b}_i^k} = 0,$

(b) $\displaystyle\sum_{k=0}^{2^n-1} \left[\sum_{i=1}^{n} a_i (-1)^{\mathfrak{b}_i^k} \right] \left[\sum_{j=1}^{n} c_j (-1)^{\mathfrak{b}_j^k} \right] = 2^n \sum_{i=1}^{n} a_i c_i.$

Proof: The proof of Theorem 2.1 is given in Section A.1.

Now it can be seen from ⟦2.1⟧ that

$$\mathbb{E}_{\mathcal{U}}[\mathbf{U}] = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \mathbf{U}_{\langle k \rangle} = \frac{1}{2^n} \begin{bmatrix} \frac{1}{n} \sum_{k=0}^{2^n-1} \sum_{i=1}^{n} (-1)^{\mathfrak{b}_i^k} D_i \\ \frac{1}{n-1} \sum_{k=0}^{2^n-1} \sum_{i=1}^{n} (-1)^{\mathfrak{b}_i^k} D_i \left(S_i - \overline{S} \right) \end{bmatrix} = \mathbf{0}.$$

The last equality comes from two applications of Theorem 2.1(a) to the components of the vector using $a_i = D_i$ and $a_i = D_i \left(S_i - \overline{S} \right)$.

The way that $\widehat{\boldsymbol{\Gamma}}_C$ is defined in ⟦2.2⟧ gives a computationally cumbersome method of calculating $\widehat{\boldsymbol{\Gamma}}_C$. This involves calculating a 2×2 covariance matrix from 2^n points. The following theorem gives an alternate and more efficient way to calculate $\widehat{\boldsymbol{\Gamma}}_C$.

Theorem 2.2

$\widehat{\mathbf{\Gamma}}_C$ in [[2.2]] can be written as

$$\widehat{\mathbf{\Gamma}}_C = \begin{bmatrix} \frac{1}{n^2}\sum_{i=1}^{n} D_i^2 & \frac{1}{n(n-1)}\sum_{i=1}^{n} D_i^2\left(S_i - \overline{S}\right) \\ \frac{1}{n(n-1)}\sum_{i=1}^{n} D_i^2\left(S_i - \overline{S}\right) & \frac{1}{(n-1)^2}\sum_{i=1}^{n} D_i^2\left(S_i - \overline{S}\right)^2 \end{bmatrix}. \qquad [\![2.3]\!]$$

Proof: The proof of Theorem 2.2 is given in Section A.2.

This simplifies the calculation of $\widehat{\mathbf{\Gamma}}_C$ greatly in that the covariance matrix can be calculated from n points rather than the much larger 2^n points.

2.3.2 Unconditional Estimates of $\mathbf{\Gamma}$

Another approach to estimating $\mathbf{\Gamma}$ is to express it in terms of the moments of the differences and sums and then estimate these moments from the data. The following theorem gives $\mathbf{\Gamma}$ in terms of these moments.

Theorem 2.3

If $\mathbf{\Gamma} = \mathbb{V}[\mathbf{U}] = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$ and $\mathbb{V}[\mathbf{Y}_i] = \begin{bmatrix} \sigma_D^2 & \sigma_{DS} \\ \sigma_{DS} & \sigma_S^2 \end{bmatrix}$, then $\gamma_{11} = \sigma_D^2/n$, $\gamma_{12} = \gamma_{21} = \delta_{21}/n$, and

$$\gamma_{22} = \frac{(n-1)\delta_{22} - (n-2)\sigma_{DS}^2 + \sigma_D^2\sigma_S^2}{n(n-1)},$$

provided $\delta_{22} < \infty$, where $\delta_{2a} = \mathbb{E}\left[(D_i - \mu_D)^2 (S_i - \mu_S)^a\right]$.

Proof: The proof of Theorem 2.3 is given in Section A.3.

Now an unconditional estimate of $\mathbf{\Gamma}$ can be obtained by replacing the unknown parameters in $\mathbf{\Gamma}$ by their appropriate sample estimates. This estimate is unconditional in the sense that $\mathbf{\Gamma}$ is derived without dependence on the conditional permutation distribution. This "plug-in" method will result in a different unconditional estimate of $\mathbf{\Gamma}$ for each $\mathbf{Y}_{\langle k\rangle} \in \mathcal{Y}$; call it $\widehat{\mathbf{\Gamma}}_{P\langle k\rangle}$. More formally, define the following estimates based on $\mathbf{Y}_{\langle k\rangle}$:

$$\begin{aligned}
\widehat{\mu}_{D\langle k\rangle} &= \frac{1}{n}\sum_{i=1}^{n}(-1)^{\mathfrak{b}_i^k} D_i = U_{1\langle k\rangle}, \\
\widehat{\sigma}^2_{D\langle k\rangle} &= \frac{1}{n-1}\sum_{i=1}^{n}\Big[(-1)^{\mathfrak{b}_i^k} D_i - \widehat{\mu}_{D\langle k\rangle}\Big]^2, \\
\widehat{\sigma}^2_{S} &= \frac{1}{n-1}\sum_{i=1}^{n}\big(S_i - \overline{S}\big)^2, \\
\widehat{\sigma}_{DS\langle k\rangle} &= \frac{1}{n-1}\sum_{i=1}^{n}\Big[(-1)^{\mathfrak{b}_i^k} D_i - \widehat{\mu}_{D\langle k\rangle}\Big]\big(S_i - \overline{S}\big) = U_{2\langle k\rangle}, \\
\widehat{\delta}_{2a\langle k\rangle} &= \frac{1}{n}\sum_{i=1}^{n}\Big[(-1)^{\mathfrak{b}_i^k} D_i - \widehat{\mu}_{D\langle k\rangle}\Big]^2\big(S_i - \overline{S}\big)^a, a = 1, 2.
\end{aligned} \qquad [\![2.4]\!]$$

Then the "plug-in" estimate of $\mathbf{\Gamma}$ is given by

$$\widehat{\mathbf{\Gamma}}_{P\langle k\rangle} = \frac{1}{n}\begin{bmatrix} \widehat{\sigma}^2_{D\langle k\rangle} & \widehat{\delta}_{21\langle k\rangle} \\ \widehat{\delta}_{21\langle k\rangle} & \widehat{\delta}_{22\langle k\rangle} - \big[(n-2)\widehat{\sigma}^2_{DS\langle k\rangle} - \widehat{\sigma}^2_{D\langle k\rangle}\widehat{\sigma}^2_{S}\big]/(n-1) \end{bmatrix}. \qquad [\![2.5]\!]$$

Thus the elements of $\mathcal{E}$ are calculated as $E_{P\langle k\rangle} = \mathbf{U}'_{\langle k\rangle}\widehat{\mathbf{\Gamma}}^{-1}_{P\langle k\rangle}\mathbf{U}_{\langle k\rangle}, k = 0, 1, \ldots, 2^n - 1.$

It should be noted that the conditional estimate of $\mathbf{\Gamma}$ in the previous section is invariant over $\mathcal{Y}$. That is, if any element of $\mathcal{Y}$ were the observed value of the data, the resulting permutation distribution, and thus $\widehat{\mathbf{\Gamma}}_C$, would be unchanged. Hence $\widehat{\mathbf{\Gamma}}_C$ only needs to be calculated once rather than the 2^n times that $\widehat{\mathbf{\Gamma}}_P$ does. This has obvious computational advantages. A similar invariant version of $\widehat{\mathbf{\Gamma}}_P$ can be obtained by more strongly imposing the null hypothesis in the estimation of $\mathbf{\Gamma}$.

Notice that under the null hypothesis $\mu_D = \mathbb{E}[U_1 | \mathcal{H}_0] = 0$ and $\sigma_{DS} = \mathbb{E}[U_2 | \mathcal{H}_0] = 0$. If these null values are used in the estimation of the parameters in $\mathbf{\Gamma}$, then the estimates in ⟦2.4⟧ become

$$\begin{aligned}
\widehat{\mu}_D &= 0, \\
\widehat{\sigma}_D^2 &= \frac{1}{n}\sum_{i=1}^{n} D_i^2, \\
\widehat{\sigma}_S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(S_i - \overline{S}\right)^2, \\
\widehat{\sigma}_{DS} &= 0, \\
\widehat{\delta}_{2a} &= \frac{1}{n}\sum_{i=1}^{n} D_i^2\left(S_i - \overline{S}\right)^a, a = 1, 2.
\end{aligned} \tag{2.6}$$

These estimates are invariant over $\mathcal{Y}$, hence, the invariant "plug-in" estimate of $\mathbf{\Gamma}$ is given by

$$\widehat{\mathbf{\Gamma}}_I = \frac{1}{n}\begin{bmatrix} \widehat{\sigma}_D^2 & \widehat{\delta}_{21} \\ \widehat{\delta}_{21} & \widehat{\delta}_{22} + \widehat{\sigma}_D^2\widehat{\sigma}_S^2/(n-1) \end{bmatrix} \tag{2.7}$$

and the elements of $\mathcal{E}$ are calculated as $E_{I\langle k\rangle} = \mathbf{U}'_{\langle k\rangle}\widehat{\mathbf{\Gamma}}_I^{-1}\mathbf{U}_{\langle k\rangle}, k = 0, 1, \ldots, 2^n - 1$.

If the underlying distribution from which the data are generated is bivariate normal, then so too are the differences and sums. Since the differences and sums are uncorrelated under $\mathcal{H}_0$, bivariate normality implies that they are in fact independent. Therefore, under the null hypothesis and bivariate normality

$$\begin{aligned}
\delta_{21} &= \mathbb{E}\left[(D_i - \mu_D)^2 (S_i - \mu_S) | \mathcal{H}_0\right] \\
&= \mathbb{E}\left[(D_i - \mu_D)^2 | \mathcal{H}_0\right]\mathbb{E}[(S_i - \mu_S) | \mathcal{H}_0] = 0
\end{aligned}$$

and

$$\begin{aligned}
\delta_{22} &= \mathbb{E}\left[(D_i - \mu_D)^2 (S_i - \mu_S)^2 | \mathcal{H}_0\right] \\
&= \mathbb{E}\left[(D_i - \mu_D)^2 | \mathcal{H}_0\right]\mathbb{E}\left[(S_i - \mu_S)^2 | \mathcal{H}_0\right] = \sigma_D^2\sigma_S^2.
\end{aligned}$$

Thus, $\mathbf{\Gamma}$ simplifies to

$$\mathbf{\Gamma}_N = \begin{bmatrix} \sigma_D^2/n & 0 \\ 0 & \sigma_D^2\sigma_S^2/(n-1) \end{bmatrix},$$

which can be estimated by

$$\widehat{\mathbf{\Gamma}}_N = \begin{bmatrix} \widehat{\sigma}_D^2/n & 0 \\ 0 & \widehat{\sigma}_D^2\widehat{\sigma}_S^2/(n-1) \end{bmatrix}. \qquad [\![2.8]\!]$$

Even if the underlying distribution is not normal, the elements of $\mathcal{E}$, $E_{N\langle k\rangle} = \mathbf{U}'_{\langle k\rangle}\widehat{\mathbf{\Gamma}}_N^{-1}\mathbf{U}_{\langle k\rangle}, k = 0, 1, \ldots, 2^n - 1,$ are conditionally equally likely since they are functions of the elements of $\mathcal{U}$ and the mapping from $\mathcal{U}$ to $\mathcal{E}$ (through $\widehat{\mathbf{\Gamma}}_N$) does not depend on the permutation distribution. Hence, regardless of the underlying distribution, a test based on E_N is conditionally distribution-free.

The likelihood ratio test under bivariate normality was given by Hsu (1940) and, in a different form, by Bradley and Blackwood (1989). In Section A.4, it is shown that, not only are these two tests equivalent, but that they are each monotonic functions of E_N. Hence, E_N is a randomization version of the likelihood ratio test.

To calculate the test statistic E using any of the unconditional methods of estimating $\mathbf{\Gamma}$, it is necessary that $\widehat{\mathbf{\Gamma}}$ be invertable. This is addressed by the following theorem.

Theorem 2.4

For $M = C, P, I, N, \left|\widehat{\mathbf{\Gamma}}_M\right| \geq 0$.

Proof: The proof of Theorem 2.4 is given in Section A.5.

2.4 The Asymptotic Distribution of E_M

As the sample size increases, enumerating the entire permutation distribution becomes prohibitive. One way to get around this in practice is to take a random sample from the permutation distribution rather than enumerating it. This was first proposed by Dwass

(1957). The test remains exact and conditionally distribution-free, the only penalty being a loss of efficiency. Jöckel (1986) quantifies the loss of efficiency as a function of the resample size. For example, he shows that at the $\alpha = .05$ significance level, a sample of size 999 from the permutation distribution will be at least 94.5% as efficient as enumerating the entire permutation distribution.

Another method when the sample size is large is to use the large sample distribution of E_M to carry out the hypothesis test. The following theorem provides the appropriate asymptotic distribution.

Theorem 2.5

(a) As $n \longrightarrow \infty$,

$$E_P \xrightarrow{\mathcal{D}} {\chi'_2}^2(\lambda),$$

where $\lambda = \mu_D^2\sigma_D^2 + 2\mu_D\sigma_{DS}\delta_{21} + \sigma_{DS}^2(\delta_{22} - \sigma_{DS}^2)$. That is, E_P converges in distribution to a noncentral chi-squared distribution with 2 degrees of freedom and noncentrality parameter λ.

(b) When $\mathcal{H}_0$ is true, then as $n \longrightarrow \infty$,

$$E_M \xrightarrow{\mathcal{D}} \chi_2^2$$

for $M = C, P, I, N$.

Proof: The proof of Theorem 2.5 is given in Section A.6.

One interesting fact related to the asymptotic distribution of E_C is found by considering the expectation of E_C over the permutation distribution,

$$\mathbb{E}_{\mathcal{E}}[E_C] = \frac{1}{2^n}\sum_{k=0}^{2^n-1} E_{C\langle k\rangle} = \frac{1}{2^n}\sum_{k=0}^{2^n-1} \mathbf{U}'_{\langle k\rangle}\widehat{\mathbf{\Gamma}}_C^{-1}\mathbf{U}_{\langle k\rangle}.$$

If we let $\widehat{\gamma}_C^{(i,j)}$ be the ij^{th} element of $\widehat{\mathbf{\Gamma}}_C$ (see Section A.2), then by explicitly inverting $\widehat{\mathbf{\Gamma}}_C$ and writing out the quadratic form, we see that

$$\begin{aligned}\mathbb{E}_{\mathcal{E}}[E_C] &= \frac{1}{2^n}\sum_{k=0}^{2^n-1} \frac{U_{1\langle k\rangle}^2\widehat{\gamma}_C^{(2,2)} + U_{2\langle k\rangle}^2\widehat{\gamma}_C^{(1,1)} - 2U_{1\langle k\rangle}U_{2\langle k\rangle}\widehat{\gamma}_C^{(1,2)}}{\widehat{\gamma}_C^{(1,1)}\widehat{\gamma}_C^{(2,2)} - \widehat{\gamma}_C^{(1,2)}\widehat{\gamma}_C^{(1,2)}} \\ &= \frac{\widehat{\gamma}_C^{(1,1)}\widehat{\gamma}_C^{(2,2)} + \widehat{\gamma}_C^{(2,2)}\widehat{\gamma}_C^{(1,1)} - 2\,\widehat{\gamma}_C^{(1,2)}\widehat{\gamma}_C^{(1,2)}}{\widehat{\gamma}_C^{(1,1)}\widehat{\gamma}_C^{(2,2)} - \widehat{\gamma}_C^{(1,2)}\widehat{\gamma}_C^{(1,2)}} = 2.\end{aligned}$$

That is, regardless of the original data, the conditional expectation of E_C is always 2, which is also the mean of the limiting chi-squared distribution.

CHAPTER III

GRAPHICAL AND NUMERICAL COMPARISONS

This chapter describes some symmetry properties of the permutation distribution and gives some graphical interpretations as well. In addition, a simulation study compares the empirical power of the permutation tests with some of the tests described in Chapter I.

3.1. Graphical Properties of E_M

This section contains graphical interpretations and comparisons of our tests as well as comparisons with other tests. By considering the graphical properties of the permutation distribution, we can gain insight into the behavior of our test statistics.

3.1.1. Interpretation of the Rejection Region

Figure 2.1, discussed earlier, is an example of the graphical depiction of $\mathcal{U}$, the permutation distribution of **U**. Each point in the scatterplot represents a calculated value of U_1 and U_2 from one of the 2^n permutations of the observed data (elements of $\mathcal{Y}$). The rejection region for each test is determined by the m points corresponding to the largest values of $E_{M\langle k\rangle} = \mathbf{U}'_{\langle k\rangle}\widehat{\mathbf{\Gamma}}_M^{-1}\mathbf{U}_{\langle k\rangle}, k = 0, 1, \ldots, 2^n - 1,$ where $\alpha = m/2^n$ is the nominal level of the test. When $M = C, I, N,$ $\widehat{\mathbf{\Gamma}}_M$ is the same for all permutations of the data ($k = 0, 1, \ldots, 2^n - 1$), hence the distance from the origin to each point in $\mathcal{U}$ is being measured using the same metric. Therefore, constant values of E_M define equidistant points. Since constant values of a quadratic form define an ellipse in two dimensions, an ellipse can be drawn with constant distance equal to the $100(1-\alpha)^{\text{th}}$ percentile of the

$E_{M\langle k\rangle}$ distribution. Points outside this ellipse will correspond to the m largest values of $E_{M\langle k\rangle}$, hence these points constitute the rejection region for the test.

Different estimates of $\mathbf{\Gamma}$ will lead to different ellipses which define the appropriate rejection regions. The three ellipses for E_C, E_I, and E_N are shown in Figure 3.1 for the permutation distribution in Figure 2.1. These ellipses use $\alpha = 102/2^{10} = .0996$.

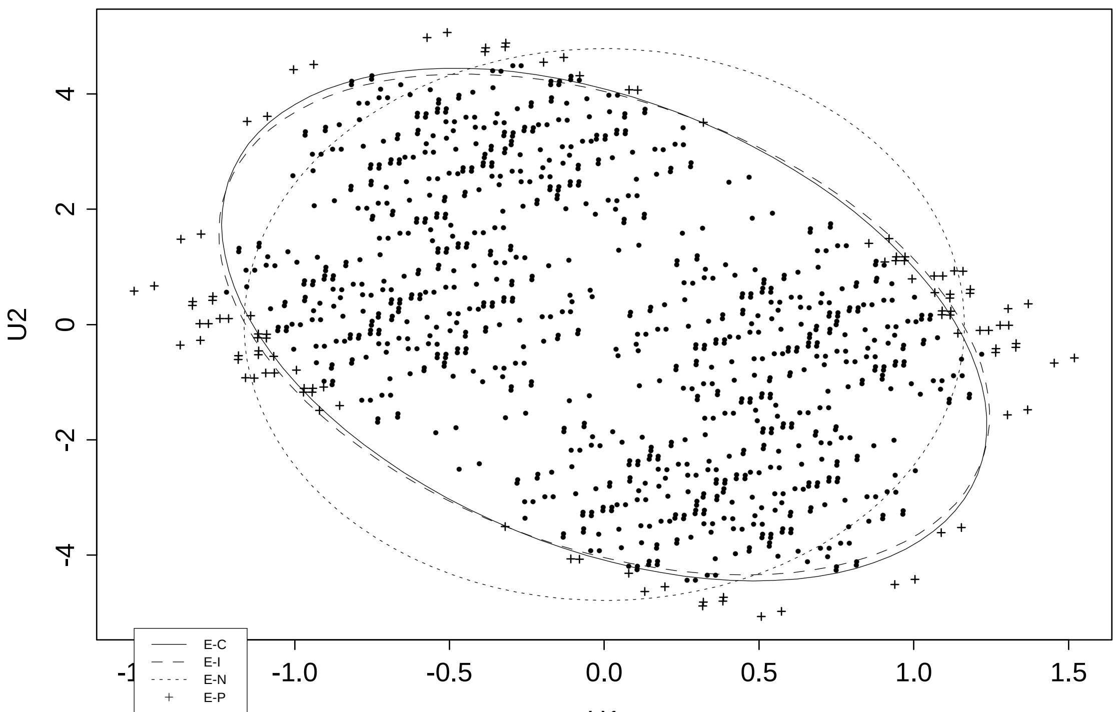

Figure 3.1. Permutation distribution of **U** with the rejection regions for E_C, E_P, E_I, and E_N depicted.

Since $\widehat{\mathbf{\Gamma}}_P$, the unconditional plug-in estimator, changes for each permutation of the data, an ellipse will not define the rejection region as for the other tests. But in Figure 3.1 the rejection region for E_P is depicted in another way; the points corresponding to the $m = 102$ largest values of $E_{P\langle k\rangle}$ are marked with a $+$.

This graphical representation allows us to see similarities and differences in the behavior of the different tests. Notice that the ellipse for E_N has axes that are parallel to the (U_1, U_2) axes. This is a result of the diagonality of $\widehat{\mathbf{\Gamma}}_N$. Also, the ellipses for E_C and E_I are very similar, almost overlapping. This similarity comes from the algebraic similarity of $\widehat{\mathbf{\Gamma}}_C$ and $\widehat{\mathbf{\Gamma}}_I$, which can be seen by comparing ⟦2.3⟧ with ⟦2.6⟧ and ⟦2.7⟧.

3.1.2. Comparisons With Other Distribution-Free Tests

Section 1.3 described two distribution-free tests of $\mathcal{H}_0$ due to Sen (1967) and Kepner and Randles (1984). Some preliminary simulations suggested that the version of Sen's test that uses Mood's statistic for detecting a difference in variances performs better than the one that uses the Ansari-Bradley statistic. This is supported by the Pitman asymptotic relative efficiency comparisons of the two in Kepner and Randles (1984). [Note that the column labels in Table I of Kepner and Randles (1984) should be reversed for the two versions of Sen's test (J. Kepner, personal communication, see Section A.7).] As a result, we consider only the version of Sen's test that incorporates Mood's statistic for the remainder of this dissertation.

The permutation distributions for Sen's test and Kepner and Randles's test, denoted by S and K, respectively, are calculated from the same permutations of the data as E_M, namely, the 2^n interchanges of X_{1i} and $X_{2i}, i = 1, \ldots, n$. Even though the points in Figure 3.1 represent values of U_1 and U_2, they also represent a particular permutation of the data. Hence, permutations of the data resulting in values of S and K in the rejection region of those tests can be color coded independently of the plotting symbol. For example, in Figure 3.2 the red points correspond to permutations of the data that lead to values of K that are in that test's rejection region, the blue points correspond to permutations of the data that lead to values of S that are in that test's rejection region, and the green points correspond to permutations of the data that lead to values of K and S that are in the rejection regions for both tests. Hence, there are $m = 102$ red and green points and $m = 102$ blue and green points so that both tests are conducted at the same $\alpha = 102/2^{10} = .0996$ level.

Now, the permutation distribution in Figure 3.2 is generated by one particular permutation of the data, namely, the permutation of the data that was observed. This observed value of $\mathbf{U} = (-.76, 3.86)'$ is marked with a $\square$. When this point falls outside one of the ellipses, then it is in the rejection region for the corresponding test. For each of the three ellipses, the legend shows (first) the value of E_M that the ellipse represents, so points outside the ellipse have larger values of E_M. The legend also shows (to the right of the equal sign) the observed value of E_M for each test. If the $\square$ is overlayed with a $+$, then it is in the rejection region for E_P. The appropriate color of the $\square$ indicates whether it is in the rejection regions for K or S. In Figure 3.2, the observed value falls in the rejection regions for E_N and K.

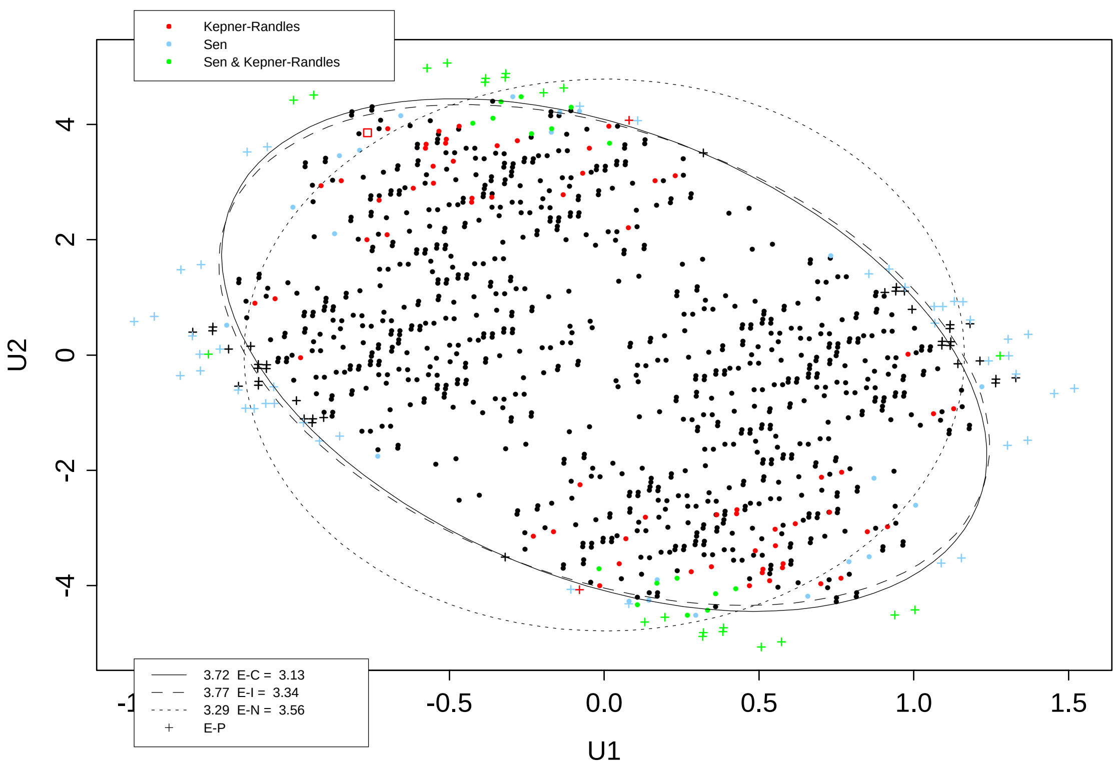

Figure 3.2. Permutation distribution of $\mathbf{U}$ with the rejection regions for E_C, E_P, E_I, and E_N depicted. The rejection regions for K and S are color coded.

Figures 3.3 to 3.8 show the permutation distributions for six random samples of size $n = 10$. Table 3.1 displays the underlying distributions from which these samples were drawn. Specific information about these distributions can be found in Section 3.2.2.

The rejection regions for each of the tests in Figures 3.3 to 3.8 were determined using a significance level of $\alpha = 102/2^{10} = .0996$. These figures were chosen to illustrate some of the distinguishing features of the permutation distribution. For example, with even these few repetitions we see that the ellipses for E_C and E_I behave quite similarly, as was noted earlier.

Another interesting feature is that the rejection region for K contains points that are in the interior of the permutation distribution of **U** for several of the scatterplots (namely Figures 3.4 to 3.7, and to a lesser extent Figure 3.3). Also, in Figures 3.3, 3.5, and 3.6, the points in the rejection region for K are not very extreme in the U_1 direction. This suggests that K is measuring "extremeness" in a somewhat fundamentally different way than our tests. (In fact, in Figures 3.4 to 3.7 the observed value is considered one of the 10% most extreme points according to K, while it is not very extreme in the **U** distribution.) Whether this is better or worse than our method cannot be answered here, but will be evaluated in the empirical power study in Section 3.2.

Table 3.1. The underlying distributions from which the samples of size $n = 10$ were drawn for Figures 3.3 to 3.9, and 3.11.

Figure	Bivariate Distribution	ρ	μ_1	μ_2	σ_1^2	σ_2^2
3.3	t	.5	0	.5	1	3
3.4	lognormal	.5	0	0	1	1
3.5	normal	0	0	0	1	1
3.6	generalized Laplace	.5	0	0	1	1
3.7	lognormal	.8	0	0	1	1
3.8	lognormal	.8	0	.5	1	3
3.9	lognormal	0	0	0	1	1
3.11	t	.8	0	0	1	1

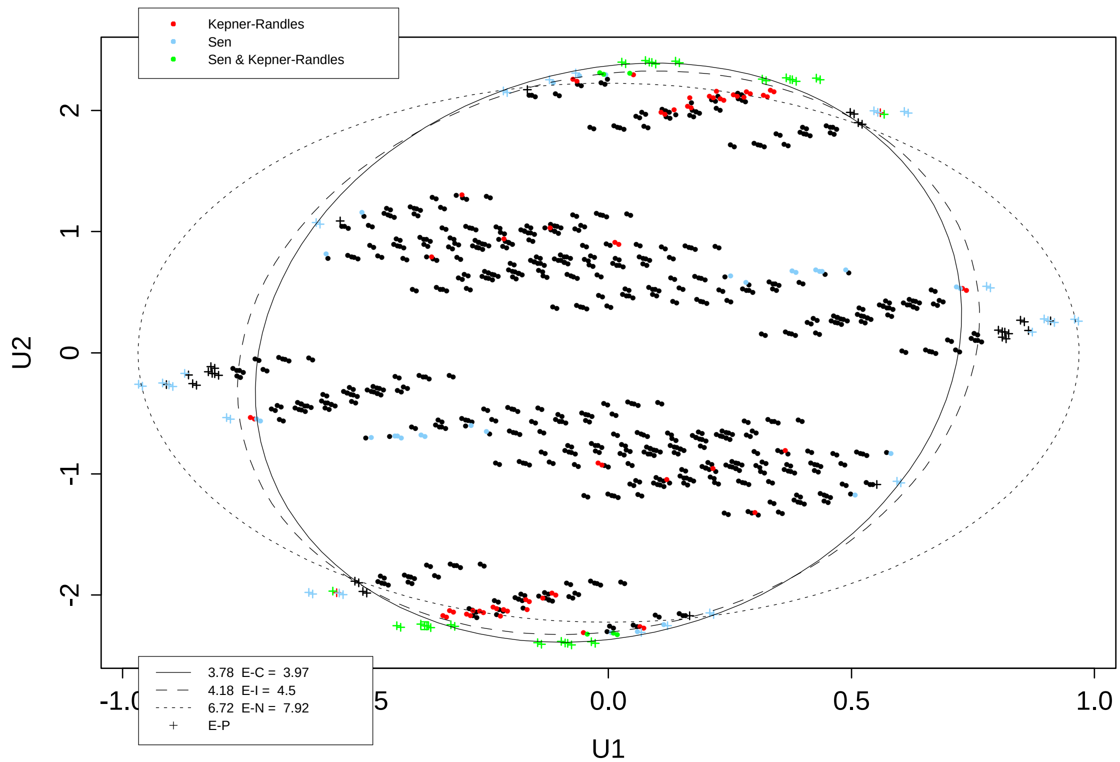

Figure 3.3. Permutation distribution for a sample from a t distribution.

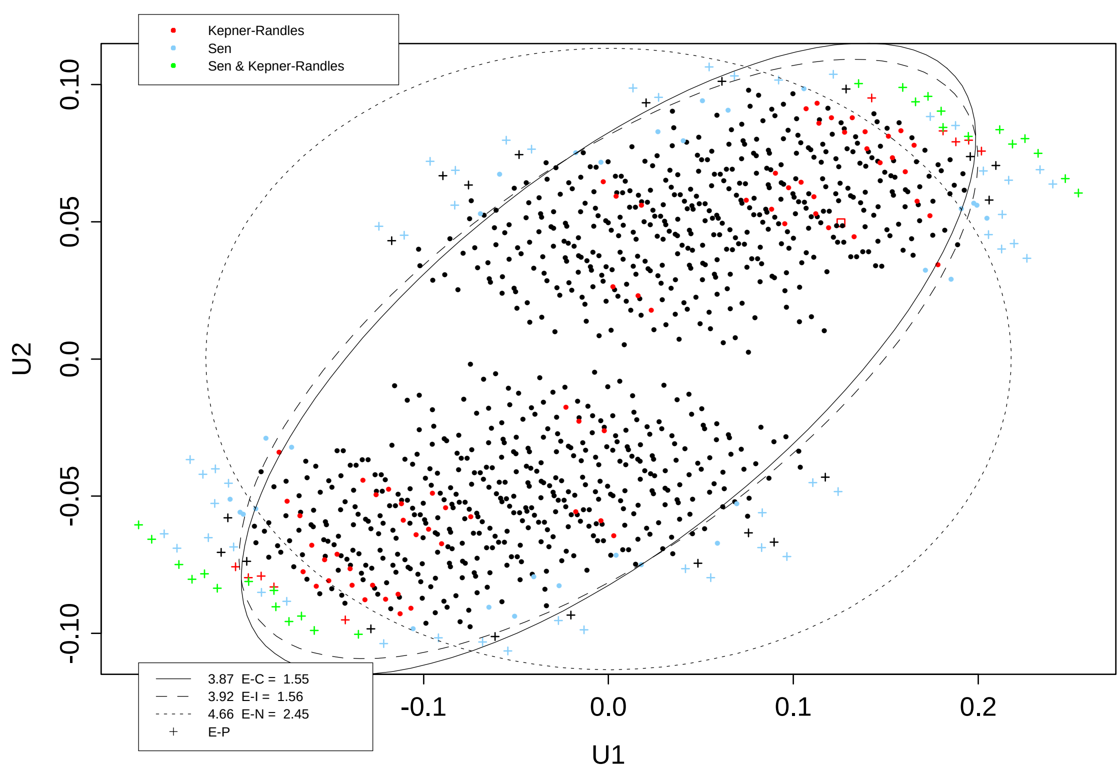

Figure 3.4. Permutation distribution for a sample from a lognormal distribution.

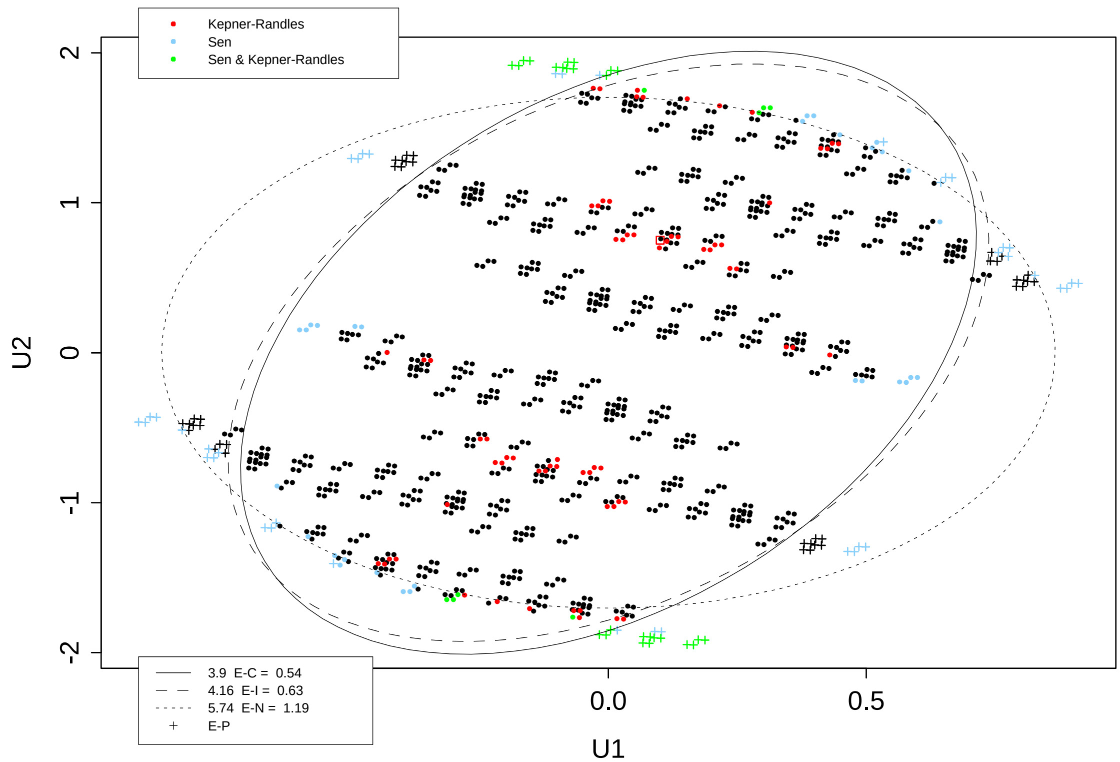

Figure 3.5. Permutation distribution for a sample from a normal distribution.

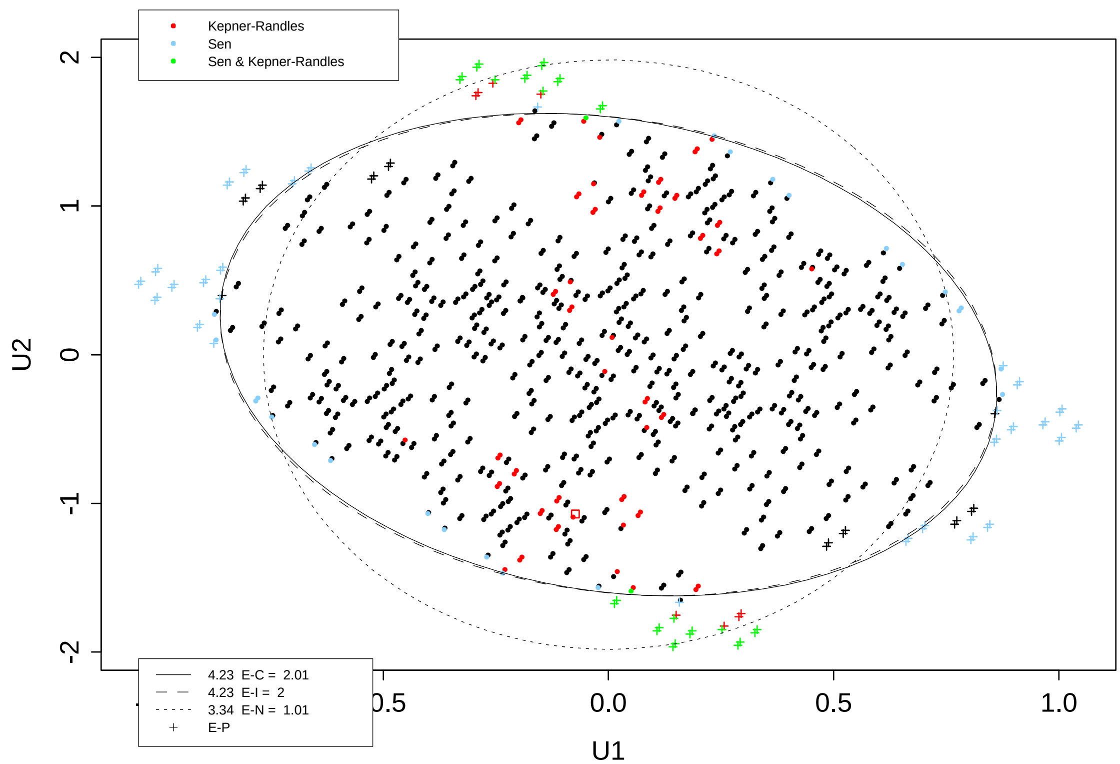

Figure 3.6. Permutation distribution for a sample from a generalized Laplace distribution.

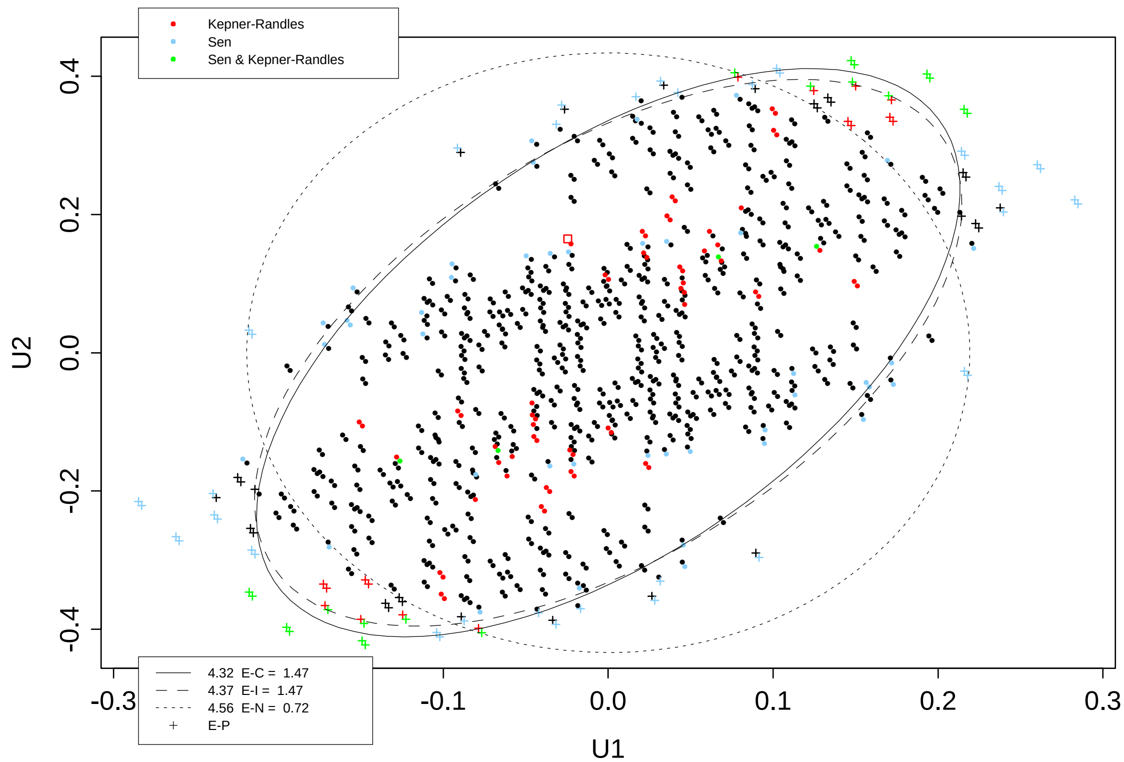

Figure 3.7. Permutation distribution for a sample from a lognormal distribution.

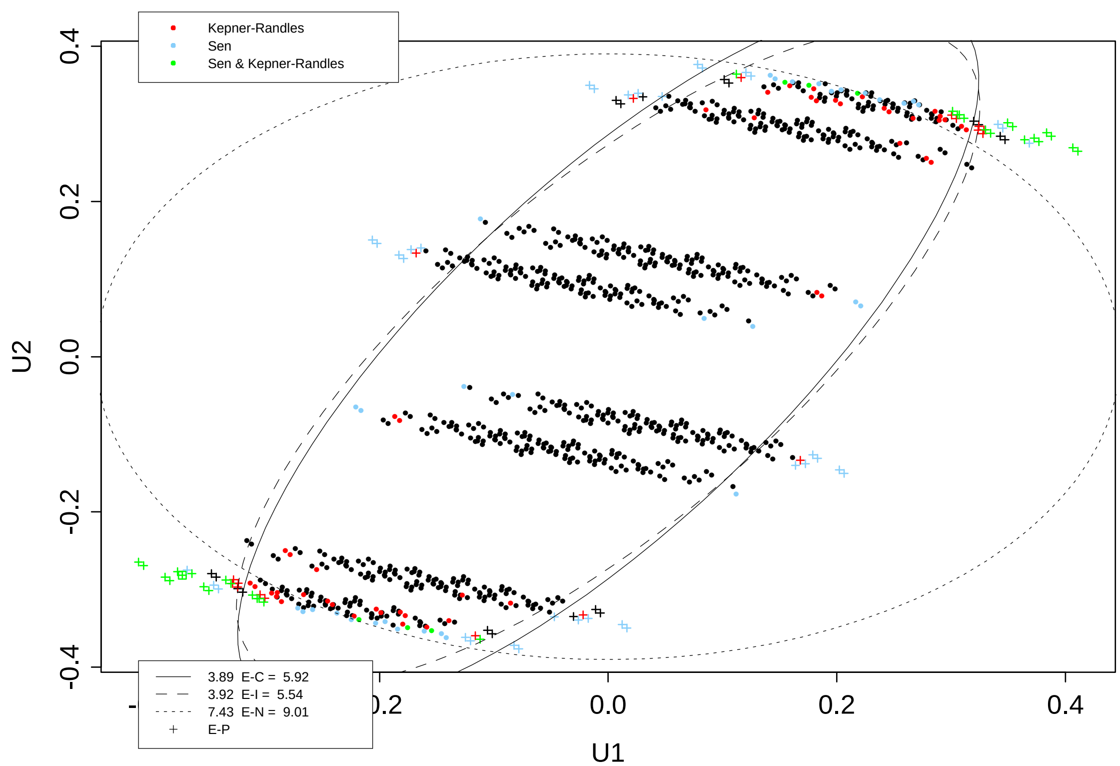

Figure 3.8. Permutation distribution for a sample from a lognormal distribution.

Figures 3.3 and 3.5 illustrate a major difference between E_N and the other permutation tests. Since $\widehat{\mathbf{\Gamma}}_N$ is diagonal, the ellipse for E_N does not follow the correlation in the permutation distribution as the other ellipses do. In these two examples, E_N will tend to reject more for points that are extreme in the U_2 direction than for extreme points in the U_1 direction. Figures 3.4 and 3.7 illustrate that since the ellipse for E_N does not follow the trend in the permutation distribution, when the observed value is extreme in both the U_1 and U_2 directions, E_N will tend to reject more than the other tests.

The permutation distributions in Figures 3.3 to 3.8 have some obvious symmetry patterns. In fact, each point in $\mathcal{U}$ has a partner that is obtained by interchanging X_{1i} and X_{2i} for all n pairs. This is equivalent to changing the sign on all of the differences. Now, it is easy to see that changing the sign on all of the differences will change the sign of both U_1 and U_2. Since, in our indexing scheme, the partner of $\mathbf{U}_{\langle k \rangle}$ is $\mathbf{U}_{\langle 2^n-1-k \rangle}$ (see Table A.1 as an example), it follows that $\mathbf{U}_{\langle k \rangle} = -\mathbf{U}_{\langle 2^n-1-k \rangle}$. Therefore, not only do we have symmetry about zero in each of the marginals, but we also have a rotational symmetry about the origin. That is, $\mathbf{U}_{\langle 2^n-1-k \rangle}$ is the image of $\mathbf{U}_{\langle k \rangle}$ after a $180°$ rotation about the origin. So, half of the points in $\mathcal{U}$ are obtained from the other half by a $180°$ rotation about the origin. This has obvious computational benefits since half of the permutation distribution can be obtained directly from the other half.

Another interesting characteristic can be observed in Figure 3.9. This is the permutation distribution obtained from the ten data points plotted in Figure 3.10(a). These data were generated from a bivariate lognormal distribution as indicated in Table 3.1. Notice the clustering that occurs in the permutation distribution; this is due to the outliers in Figure 3.10(a) that are farthest from the line $X_1 = X_2$. To see this, consider Figure 3.10(b) which is a scatterplot of the differences and sums of the same data. This is essentially a $45°$ counterclockwise rotation of Figure 3.10(a) about the origin. The 2^n permutations of the data are obtained by reflecting the points in Figure 3.10(a) over the line $X_1 = X_2$. This is equivalent to reflecting the points in Figure 3.10(b) over the line $D = 0$. Notice that when the points marked A and B in Figure 3.10(b) are on the right side of the line $D = 0$, which happens for one quarter of the permutations, $U_1 = \overline{D}$ and $U_2 = \mathbb{C}[D, S]$ will both be positive. This accounts for the cluster in the upper right corner of the permutation distribution. Similarly, when points A and B are on the left side of the line $D = 0$, U_1 and U_2 will be negative and we get the cluster in the lower left corner of the permutation distribution. For half the permutations, points A and B will be on opposite sides of the line $D = 0$ and both U_1 and U_2 will be near 0 giving us the cluster near the origin. We can see that the observed value of (U_1, U_2) is in this cluster.

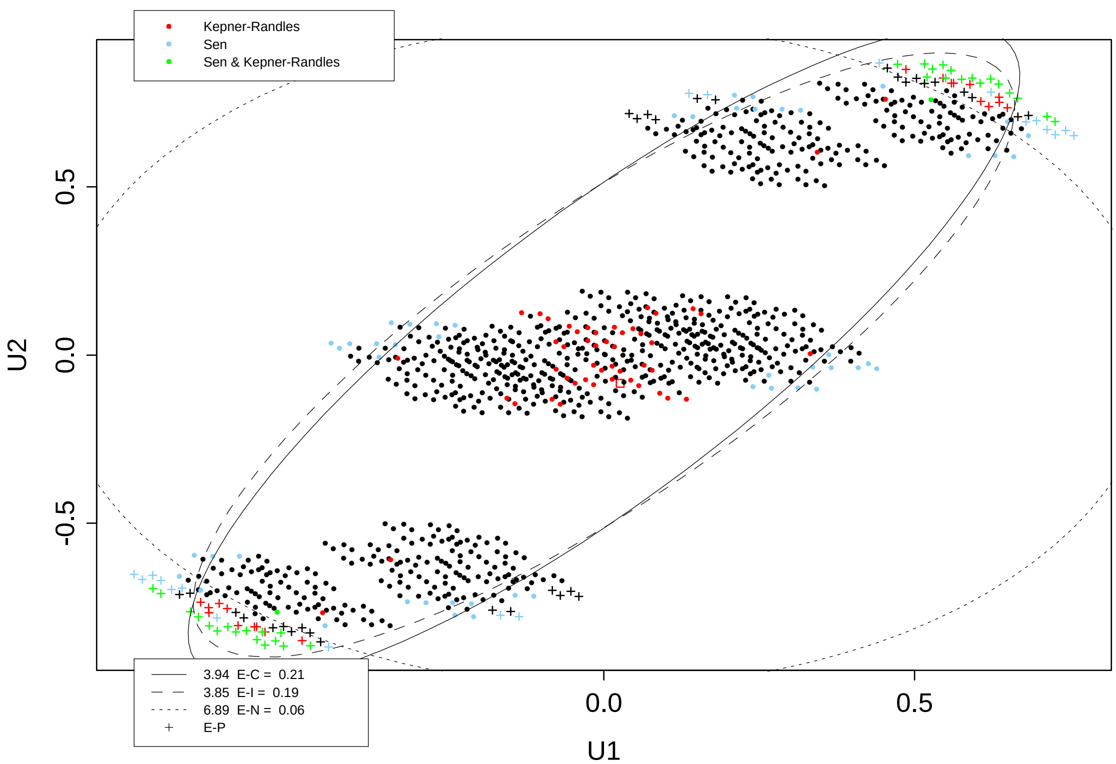

Figure 3.9. Permutation distribution for a sample from a lognormal distribution.

Within each cluster there is also some separation. This clustering is due to point C. For fixed values of points A and B, point C has much more influence on $\overline{D}$ than the remaining points causing half the points within a cluster to have generally larger values of U_1 that the other half.

A property of E_P that might seem reasonable is that if $\mathbf{U}_{\langle k \rangle}$ is in the rejection region for E_P and both components of $\mathbf{U}_{\langle l \rangle}$, $l \neq k$, are more extreme than the components of $\mathbf{U}_{\langle k \rangle}$, then $\mathbf{U}_{\langle l \rangle}$ will be in the rejection region of E_P as well. This can be stated more precisely by saying that if $E_{P\langle k \rangle}$ is in the rejection region and $|U_{i\langle l \rangle}| > |U_{i\langle k \rangle}|$ for $i = 1$ and 2, then $E_{P\langle l \rangle} > E_{P\langle k \rangle}$. It turns out that this is not necessarily true. A counterexample is given in Figure 3.11 which is the permutation distribution for a random sample of size $n = 10$ from a bivariate t distribution as indicated in Table 3.1. In this permutation distribution it can be clearly seen that there are points that are not in the rejection region of E_P that are more extreme in both the U_1 and U_2 directions than some points that are in the rejection region for E_P, which are denoted by a $+$.

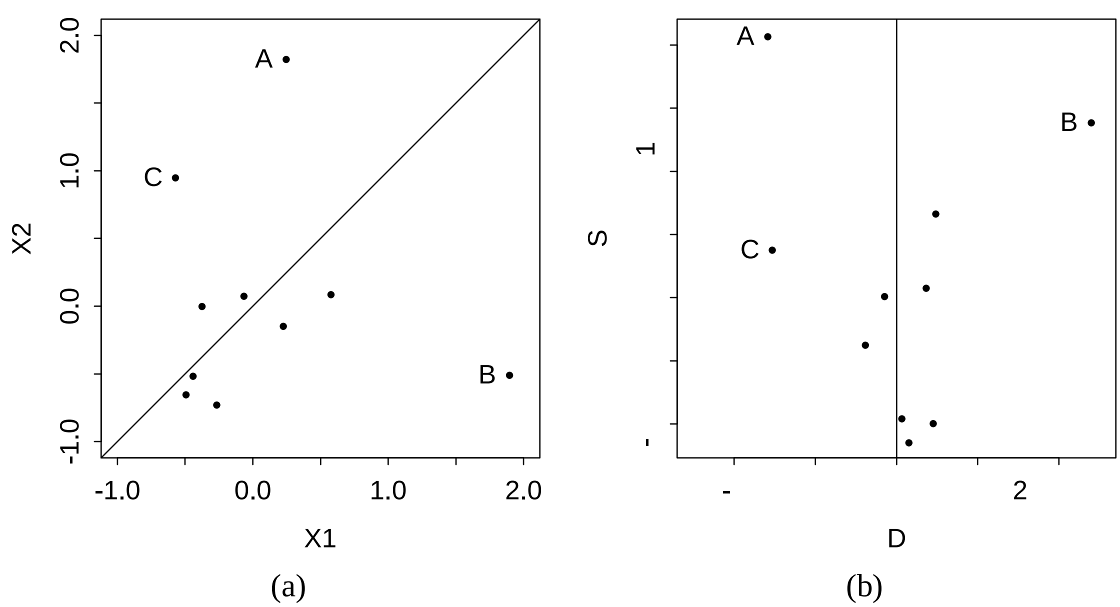

Figure 3.10. Scatterplots of (a) the random sample from a lognormal distribution used to generate Figure 3.9 and (b) the same data transformed to differences and sums.

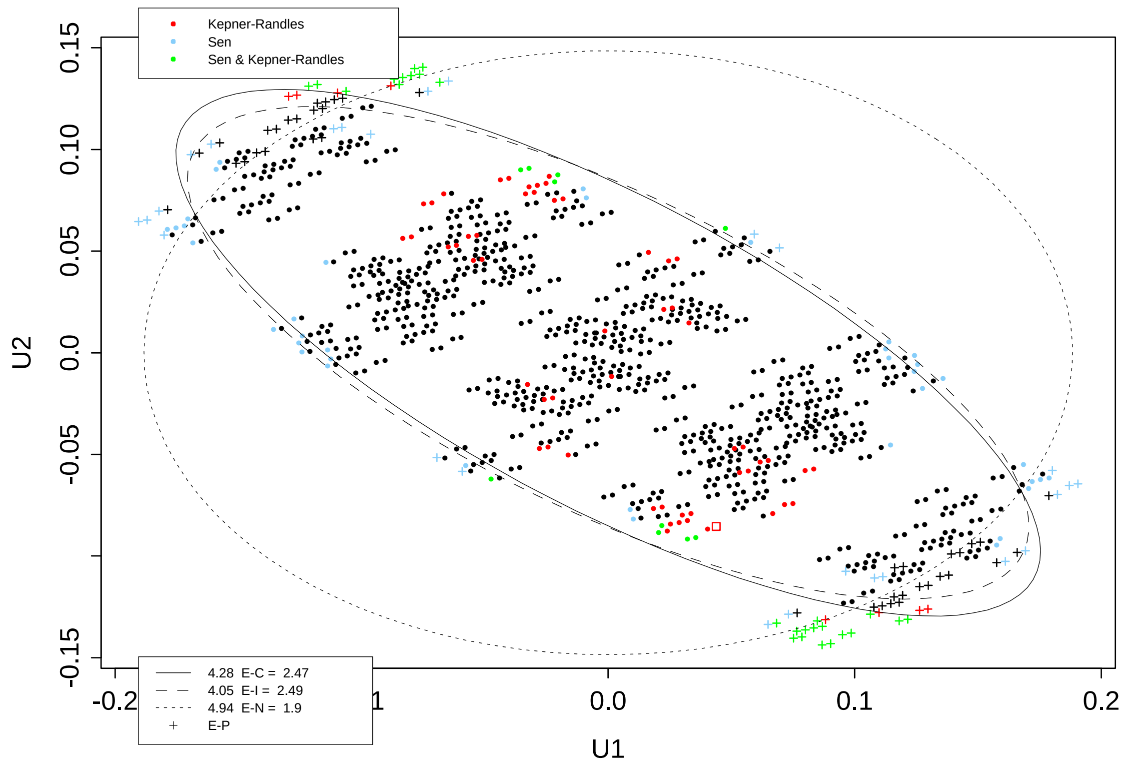

Figure 3.11. Permutation distribution for a sample from a t distribution.

3.1.3. An Illustration

As an example, consider the data in Table 3.2 showing the participation rate of women in the labor force in 18 U. S. cities and one state in 1968 and 1972 (obtained from the Data and Story Library maintained at StatLib on the World Wide Web at `http://lib.stat.cmu.edu/DASL/`). A scatterplot of the data is given in Figure 3.12. If these data can be considered representative of the labor situation in the U. S. as a whole (which is questionable, at best), we can test if the participation rate for women in the U. S. labor force was different in these two years. Besides determining if the participation rate went up or down from 1968 to 1972, it may also be of interest to determine if the variability in the participation rate was different for these two years. That is, it may be possible that the cities' participation rates became more or less variable over those four years.

Table 3.2. Participation rate of women in the labor force in 18 U. S. cities and one state in 1968 and 1972.

City/State	Participation Rate (%) 1968	Participation Rate (%) 1972
New York	42	45
Los Angeles	50	50
Chicago	52	52
Philadelphia	45	45
Detroit	43	46
San Francisco	55	55
Boston	45	60
Pittsburgh	34	49
St. Louis	45	35
Connecticut	54	55
Washington, DC	42	52
Cincinnati	51	53
Baltimore	49	57
Newark	54	53
Minneapolis/St. Paul	50	59
Buffalo	58	64
Houston	49	50
Paterson	56	57
Dallas	63	64

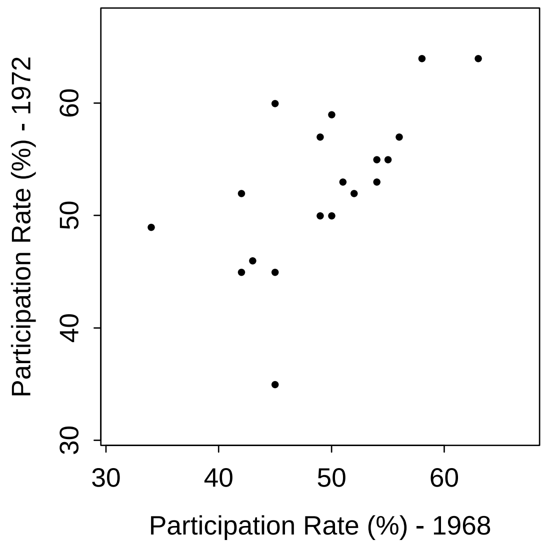

Figure 3.12. Scatterplot of data in Table 3.2.

The value of **U** calculated from the data in Table 3.2 is $\mathbf{U} = (-3.37, -3.89)'$. Figure 3.13 shows a random sample of 2,000 points from the permutation distribution of **U** with the rejection regions of all the tests set at the $\alpha = .05$ significance level. We see that E_N is the only test that does not reject $\mathcal{H}_0$. Figure 3.13 gives us an indication as to why most of the tests reject $\mathcal{H}_0$. Compared to the permutation distribution, **U** seems to be extreme in the negative U_1 direction but not extreme at all in the U_2 direction. This indicates that there is a location shift only, and that the participation rate for women in the labor force was higher in 1972 than in 1968.

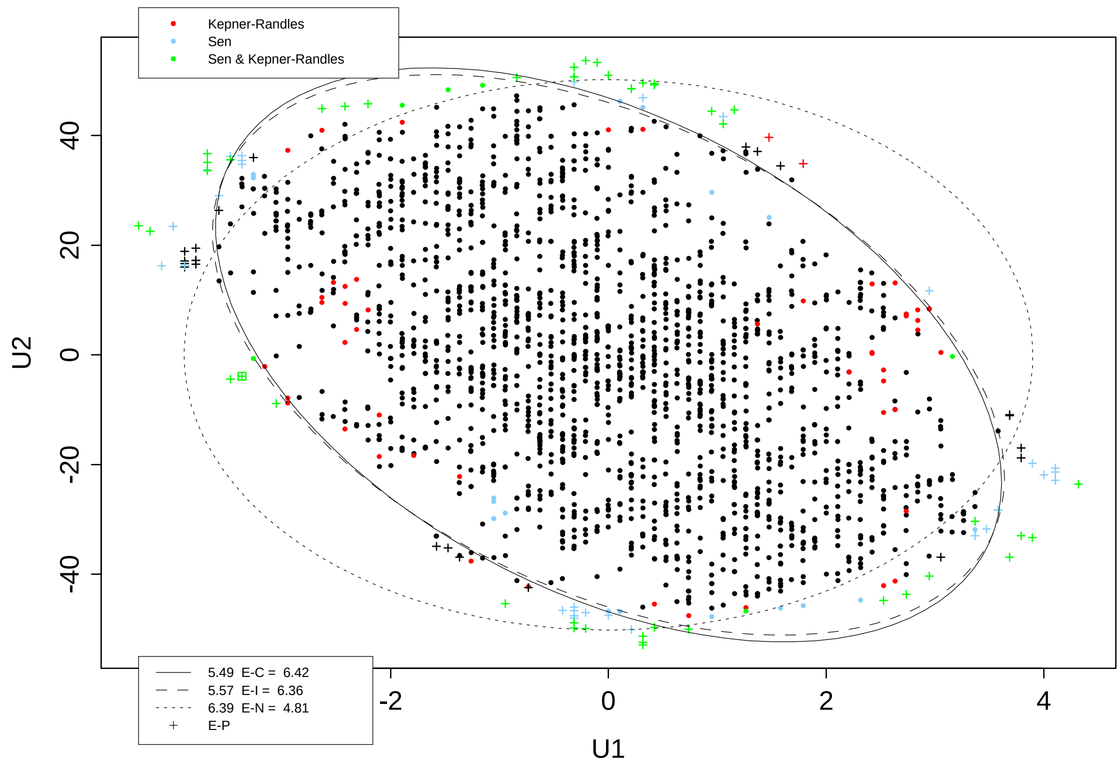

Figure 3.13. A random sample of 2,000 points from the permutation distribution of **U** for the Participation Rate data in Table 3.2.

3.2. Empirical Power Study

This section describes a simulation study that was carried out to compare the empirical powers of the proposed permutation tests among themselves and with the other tests described in Chapter I. Section 3.2.1 describes the factors involved in the study with the exception of the bivariate distributions; these are covered in Section 3.2.2 along with the methods used to generate random variates from these distributions. Section 3.2.3 discusses the results of the study and draws some conclusions.

3.2.1. Factors

There were seven test statistics considered in the power study: E_C, E_P, E_I, E_N, K, S, and F (the normal likelihood ratio test). A nominal significance level of $\alpha = .05$ was used throughout (except for $n = 8$, where $\alpha = 12/2^8 = .046875$). Each of the seven tests were carried out on 10,000 pseudo-random samples generated at each factor-level combination, where the five factors, and their levels, are described in Table 3.3. The empirical power of each test at each factor-level combination is the proportion of the 10,000 samples that the test rejected $\mathcal{H}_0$.

The five distributions considered in the power study are described in the next section. Four different small to moderate sample sizes were considered: $n = 8, 12, 16, 20$, as well as three different correlations: $\rho = 0, .5, .8$.

The mean and variance of one marginal variable, X_1, was set to $(\mu_1, \sigma_1^2) = (0, 1)$ in all cases while the mean and variance of X_2 were factors. Values of μ_2 used were $(\mu_{20}, \mu_{21}, \mu_{22}) = (0, .25, .5)$ and values of σ_2^2 were $(\sigma_{20}^2, \sigma_{21}^2, \sigma_{22}^2) = (1, 2, 3)$. Therefore, there was one value of (μ_2, σ_2^2) satisfying $\mathcal{H}_0$ and eight under $\mathcal{H}_1$.

For the six distribution-free tests, the entire permutation distribution was enumerated for $n = 8$. Rather than completely enumerating the permutation distribution for the larger sample sizes, a random sample of 499 elements was drawn. The same sample from the permutation distribution was used for all six tests.

Table 3.3. Factors and their levels used in the empirical power study.

Factor	# of Levels	Levels
Bivariate Distribution	5	normal, t, generalized Laplace, Cook-Johnson, lognormal
Sample Size	4	$n = 8, 12, 16, 20$
Correlation	3	$\rho = 0, .5, .8$
Mean Difference	3	$(\mu_{20}, \mu_{21}, \mu_{22}) = (0, .25, .5)$
Variance Difference	3	$(\sigma_{20}^2, \sigma_{21}^2, \sigma_{22}^2) = (1, 2, 3)$

3.2.2. Bivariate Distributions

There were five different families of bivariate distributions used in the power study. These were the normal, t, generalized Laplace, Cook-Johnson, and the lognormal. In what follows, each distribution is discussed and the method of variate generation is described. All pseudo-uniform variates on $(0,1)$, U_i, were obtained using the `ran2` random number generator from Press, Teukolsky, Vetterling, and Flannery (1992, p. 282).

Independent standard normal variates were generated using the method of Box and Muller (1958). Namely, a pair of independent standard normal variates was obtained by the transformation

$$\begin{aligned} V_1 &= \sqrt{-2\ln U_1}\cos(2\pi U_2), \\ V_2 &= \sqrt{-2\ln U_1}\sin(2\pi U_2), \end{aligned}$$

where U_1 and U_2 are independent uniform variates. The linear combinations

$$X_1 = \mu_1 + \sigma_1 V_1 \text{ and } X_2 = \mu_2 + \rho\sigma_2 V_1 + \sigma_2\sqrt{1-\rho^2}V_2 \qquad [\![3.1]\!]$$

were used to induce the desired means, variances, and correlation (see Johnson and Ramberg 1977).

The second distribution used was the bivariate t distribution. This is a special case of the multivariate Pearson Type VII distribution discussed in Johnson (1987, pp. 117-121). Variate generation in the bivariate case is similar to the normal distribution and is described in Johnson and Ramberg (1977). A pair of uncorrelated t variates with zero means, unit variances, and ν degrees of freedom are generated by the transformation

$$\begin{aligned} V_1 &= \sqrt{U_1^{-2/\nu}-1}\cos(2\pi U_2), \\ V_2 &= \sqrt{U_1^{-2/\nu}-1}\sin(2\pi U_2), \end{aligned}$$

where U_1 and U_2 are independent uniform variates. The linear combination $[\![3.1]\!]$ provides the correct means, variances, and correlation. In the power study, $\nu = 4$ degrees of freedom was used. This is a rather heavy-tailed bivariate elliptically contoured distribution with finite second moments.

The multivariate generalized Laplace distribution, described by Ernst (1997), is a family of elliptically contoured distributions that includes the multivariate Laplace, normal, and uniform distributions. In the bivariate case, its density function for uncorrelated components is

$$f(x_1, x_2; \lambda) = \frac{\lambda}{2\pi\Gamma(2/\lambda)} e^{-(x_1^2+x_2^2)^{\lambda/2}},$$

where $\lambda > 0$ is a shape parameter. Generating uncorrelated variates with zero means and unit variances is done by means of the transformation

$$V_1 = \sqrt{\frac{2\Gamma(2/\lambda)}{\Gamma(4/\lambda)}}\, Y^{1/\lambda} \cos(2\pi U),$$
$$V_2 = \sqrt{\frac{2\Gamma(2/\lambda)}{\Gamma(4/\lambda)}}\, Y^{1/\lambda} \sin(2\pi U),$$

where U is a uniform variate and Y is an independent gamma$\left(\frac{2}{\lambda}, 1\right)$ variate. The same linear combination ⟦3.1⟧ provides the correct means, variances, and correlation. The gamma variate Y was generated using Algorithm GT of Ahrens and Dieter (1973, p. 229). A value of $\lambda = 5$ was chosen for the power study to provide a fairly light-tailed bivariate elliptically contoured distribution.

The Cook-Johnson distribution, which is a type of multivariate uniform distribution, was introduced by Cook and Johnson (1981). In its bivariate form, the density function is

$$f(u_1, u_2; \alpha) = \frac{\alpha+1}{\alpha}(u_1 u_2)^{(-1/\alpha)-1}\left(u_1^{-1/\alpha} + u_2^{-1/\alpha} - 1\right)^{-(\alpha+2)}, \qquad [\![3.2]\!]$$

for $\alpha > 0$ and $0 < u_i \leq 1, i = 1, 2$. Since the marginals are uniform on $(0, 1)$, a variety of marginal distributions can be obtained by applying the appropriate inverse probability integral transformation. In this study, normal marginals were used. The algorithm used for calculating the inverse normal cdf is from Odeh and Evans (1974).

Cook and Johnson (1981) give a straightforward method of generating variates from the density ⟦3.2⟧. If Y_1 and Y_2 are independent gamma$(1, 1)$ variates and, independently, X is a gamma$(\alpha, 1)$ variate, then the joint density of

$$U_i = (1 + Y_i/X)^{-\alpha}, i = 1, 2,$$

is given by ⟦3.2⟧. The parameter α determines the correlation between the marginal variables. Johnson (1987, p. 176) provides a table to convert values of α to correlations when the marginals are normal.

If (Y_1, Y_2) is a bivariate normal random variable, then $(X_1, X_2) = (e^{Y_1}, e^{Y_2})$ has a bivariate lognormal distribution. Since generating bivariate normal variates was seen to be straightforward, this transformation was used to generate the lognormal variates. The bivariate lognormal distribution has two shape parameters, σ_1 and σ_2, that correspond to the standard deviations of the original normal marginals. Notice that if $\sigma_1 \neq \sigma_2$, our assumption of bivariate interchangeability under $\mathcal{H}_0$ would be violated. For the power study, we chose the shape parameters to be $\sigma_1 = \sigma_2 = 1$, giving us quite skewed marginal distributions.

The correlation between X_1 and X_2, ρ, is entirely dependent upon the correlation in the originally generated bivariate normal distribution, ρ_N, and the shape parameters, σ_1 and σ_2. The correct value of ρ_N can be calculated from the desired value of ρ by the relationship

$$\rho_N = \frac{\ln\left[\rho\sqrt{\left(e^{\sigma_1^2} - 1\right)\left(e^{\sigma_2^2} - 1\right)} + 1\right]}{\sigma_1\sigma_2}.$$

These five distributions provide a wide range of bivariate distributions. Besides the bivariate normal distribution, we have included a heavily skewed distribution, a non-elliptically contoured distribution with normal marginals, and two elliptically contoured distributions: one with heavier tails than the normal and one with lighter tails. Density plots of these five distributions are provided in Figure 3.14. For each of the densities, $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\rho = 0$, with the exception of the Cook-Johnson distribution for which $\rho = .5$ (corresponding to $\alpha = .93$). This displays its non-elliptical contours. The four distributions with symmetric marginals are also plotted on the same scale to allow easier comparisons. The other parameters used in these density plots were the same as those used in the power study. Namely, the t distribution has $\nu = 4$ degrees of freedom, the generalized Laplace distribution has a shape parameter of $\lambda = 5$, and the lognormal distribution uses the shape parameters $\sigma_1 = \sigma_2 = 1$.

More density plots of the generalized Laplace distribution can be found in Ernst (1997), while a plethera of density and contour plots of the other four distributions can be found in Johnson (1987, pp. 51-53, 64-69, 119-122, 168-169).

3.2.3. Results

The results of the power study are reported in Section A.8. Except where noted below, the tests performed similarly over all four sample sizes. Therefore, to simplify comparisons, only the results for $n = 20$ are reported. Each table in Section A.8 contains the empirical power based on 10,000 repeats for each of the seven tests at every one of the $3 \times 3 \times 3 = 27$ factor-level combinations of ρ, μ_2, and σ_2^2 in Table 3.3. Each table displays the results for one of the five distributions. At each of the 24 factor-level combinations under $\mathcal{H}_1$, the cell with the largest empirical power among the tests holding the nominal level of 5% (within three standard errors) is colored red, while the cell with the second highest power is colored yellow. This coloring facilitates interpretation of the results. The standard error of the empirical power ranges from $.2\%$ when $\mathcal{H}_0$ is true to a maximum of $.5\%$ when the true power is 50%.

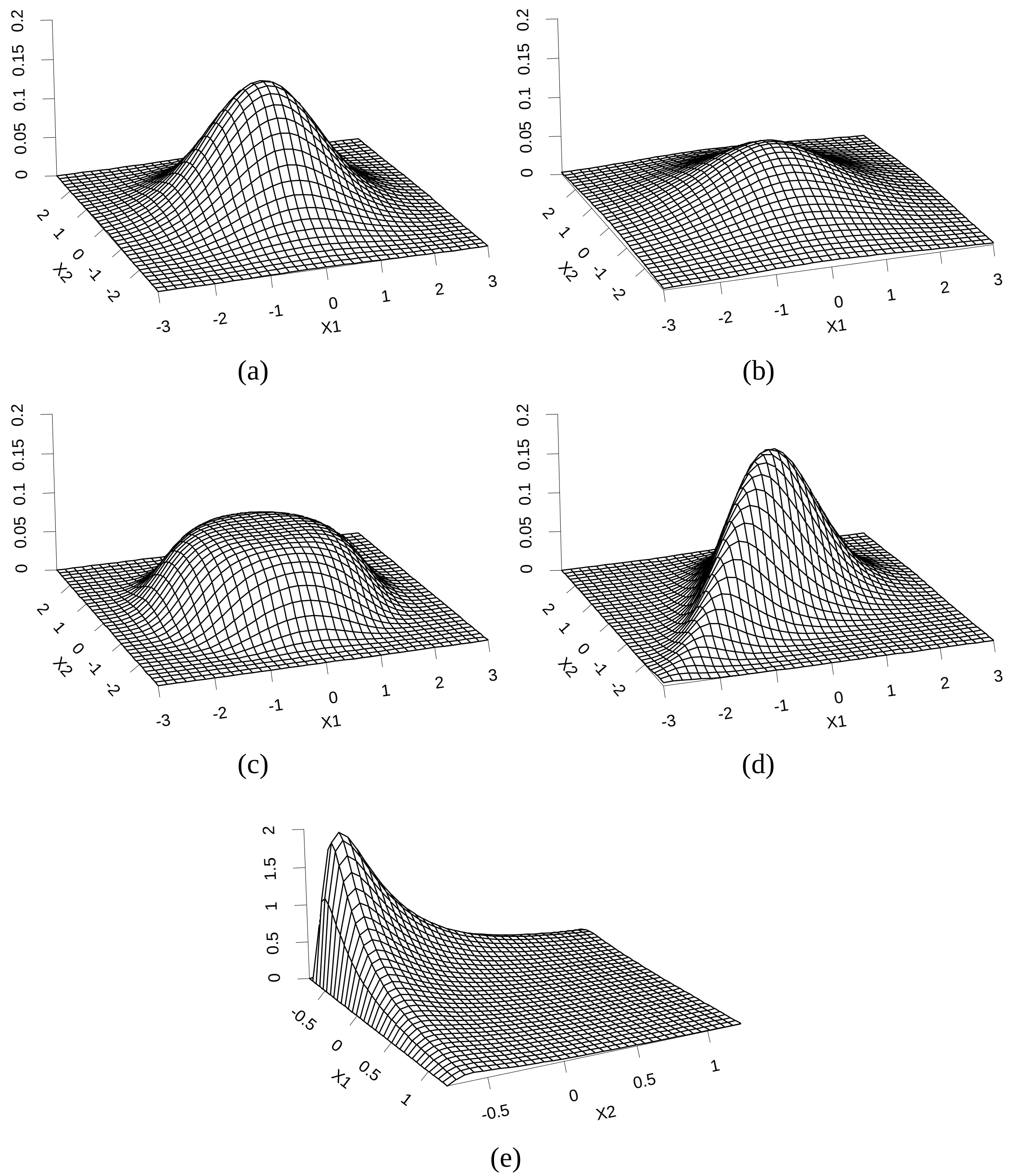

Figure 3.14. Density plots of the bivariate (a) normal, (b) t, (c) generalized Laplace, (d) Cook-Johnson (with normal marginals), and (e) lognormal distributions.

Table A.2(a) provides the results for the bivariate normal distribution. The likelihood ratio F test is the clear winner when there is some variance difference in the marginals, while E_N comes in second. When there is only a mean difference in the marginals, E_C and E_N both beat the F test, at times by as much as four standard errors.

Table A.2(b) displays the results for the bivariate t distribution on 4 degrees of freedom. When there is some difference in variances E_N tends to be the winner by a large margin, while E_C does well when there is only a mean difference. Although K is a strong second place, this is less pronounced in smaller samples where E_C tends to do better.

The results for the bivariate generalized Laplace distribution ($\lambda = 5$) are given in Table A.2(c). The clear winner here is E_N with the other variants of E_M claiming second place. One interesting feature of this table is that the F test cannot hold the nominal level of the test, as expected. But while it is on the conservative side, it is more powerful than the other tests in some cases. The empirical powers for F are colored red and yellow to indicate their place if the F test were considered with the others that held the nominal level. Similar to the bivariate normal distribution, the F test does best when there is some variance difference.

Table A.2(d) displays the results for the bivariate Cook-Johnson distribution with normal marginals. When $\rho = 0$, the marginals are uncorrelated and the Cook-Johnson distribution is a bivariate normal distribution. The results for that special case are consistent with the results in Table A.2(a).

For $\rho = .5$, E_N is the clear winner when there is a marginal difference in variances, while E_P, in second place, is also significantly ahead of the others in most cases. When there is only a difference in marginal means, E_C does best and is followed by E_I, although the significance of these is only marginal. While there is no theoretical reason for F to achieve the nominal level of $\alpha = .05$, it appears to do so here. It is also significantly more powerful than the other tests when there is no mean difference in the marginals, as is indicated by the red entries in the table.

For $\rho = .8$, E_N is again the convincing winner when there is a marginal difference in variances, with the other permutation tests picking up the slack when there is only a difference in the marginal means. When there is a difference in marginal variances, K makes a strong second place showing, although it is less efficient at the smaller sample sizes where E_P makes a stronger showing.

The results for the bivariate lognormal distribution ($\sigma_1 = \sigma_2 = 1$) are given in Table A.2(e). The clear winner in most cases is S, although E_C and E_N tend to win on the diagonal of each of the three 3×3 sub-tables. Running a strong second is K, although this is slightly better than in the smaller sample sizes. In reality, none of the tests do very well when there is both a mean and variance difference. This can be seen in Figure 3.15 which shows the empirical power of each test for the lognormal distribution with $\rho = .5$, conditional on the mean of X_2 (rows, green strip labels), the variance of X_2 (columns, orange strip labels), and the sample size (colored dots). The center panel of this trellis plot makes it clear how poorly all the tests do when there is some difference in both means and variances.

By changing the conditioning in the trellis plot we can see different features of the simulation results. Figure 3.16 shows the empirical power of each test for the Cook-Johnson distribution with $\rho = .8$, conditional on the mean of X_2 (rows, green strip labels), the sample size (columns, orange strip labels), and the variance of X_2 (colored dots). The confounding between the mean and variance difference can be seen by noticing that in any panel in rows two or three (when there is a mean difference) there are tests whose power decreases as the variance difference increases (colored dots). In many cases in row three the tests are most powerful when there is no variance difference. This pattern appears for all five distributions, especially for $\rho = .8$.

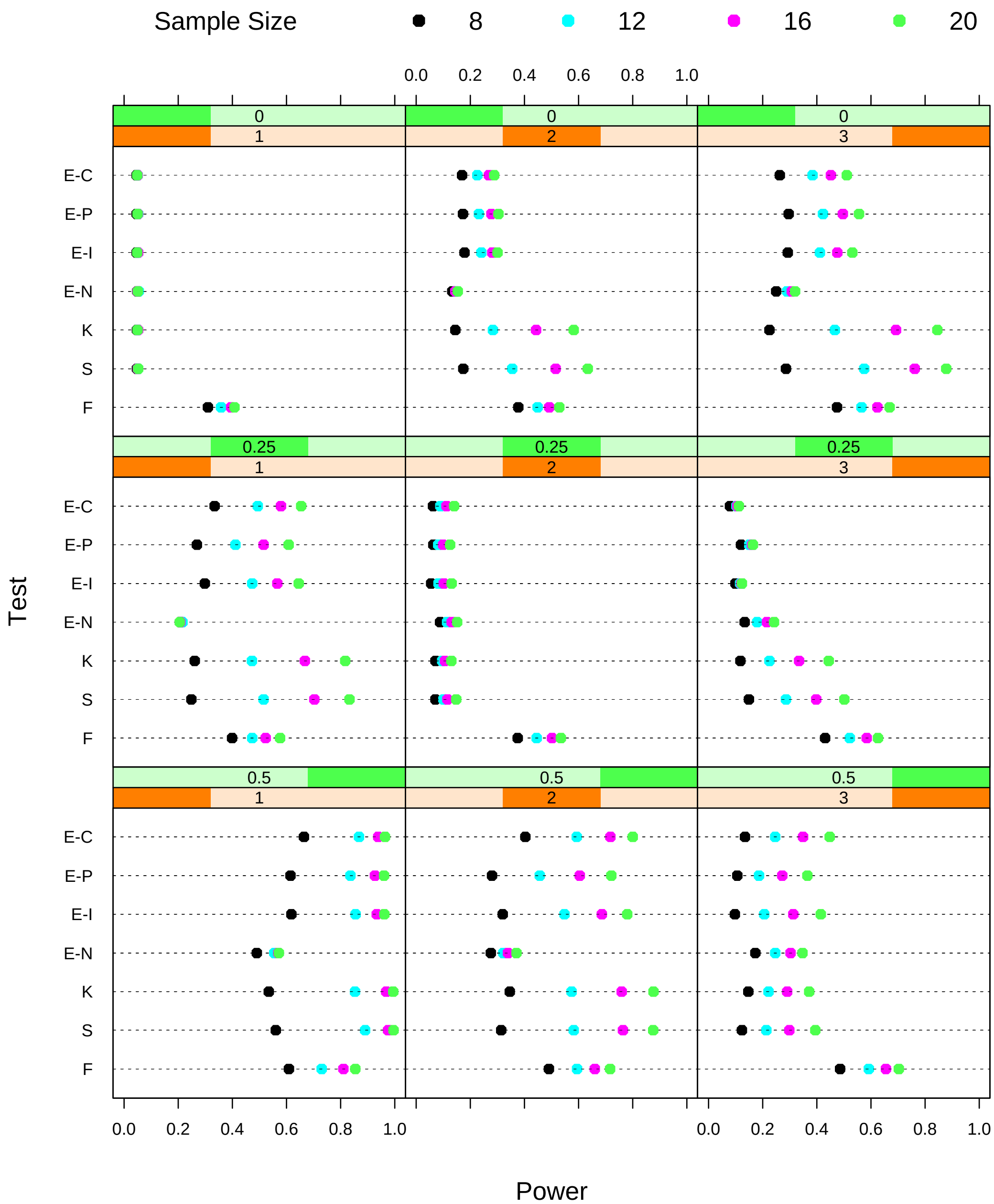

Figure 3.15. Empirical power of each test for the lognormal distribution with $\rho = .5$, conditional on the mean of X_2 (rows, green strip labels), the variance of X_2 (columns, orange strip labels), and the sample size (colored dots).

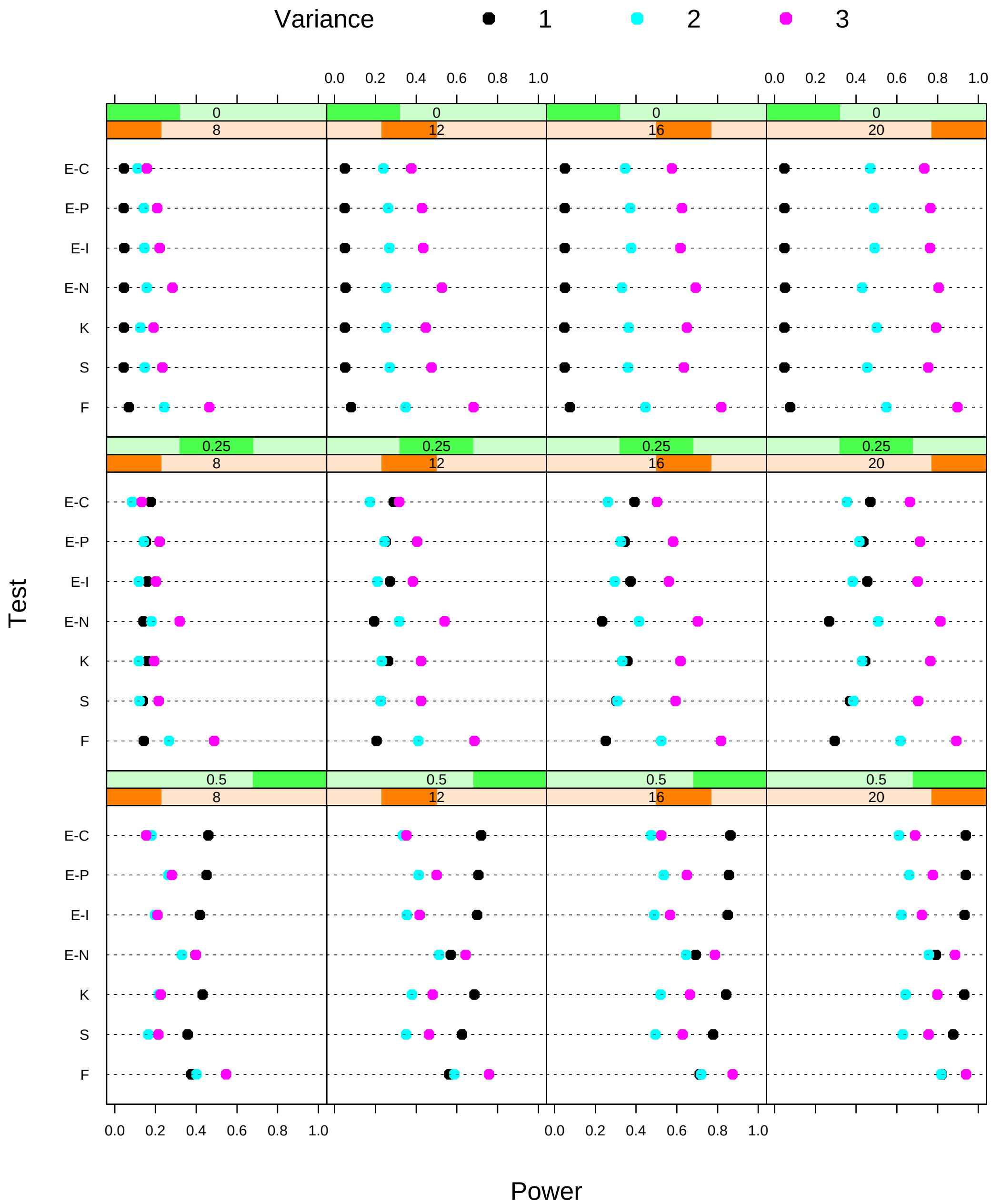

Figure 3.16. Empirical power of each test for the Cook-Johnson distribution with $\rho = .8$, conditional on the mean of X_2 (rows, green strip labels), the sample size (columns, orange strip labels), and the variance of X_2 (colored dots).

Another interesting feature of Figure 3.16 is that when there is only a difference in the marginal means (black dots in rows two and three) E_N tends to fall behind, but when there is a variance difference (blue and red dots) E_N is usually the winner. We noticed this behavior earlier in both the t and the Cook-Johnson distributions. It appears that in these distributions E_N tends to give up some power when there is only a mean difference in favor of increased power when there is a difference in both means and variances.

CHAPTER IV

CONCLUSIONS

4.1. Discussion

The work here provides some new methods for testing the hypothesis of bivariate interchangeability with a location and scale parameterization. The graphical representation of the permutation distribution was offered to give a better understanding of these tests and how they compare to the existing distribution-free tests of Sen (1967) and Kepner and Randles (1984). Also, we have recognized the overlooked work of Hsu (1940) for the bivariate normal distribution as well as showing the equivalence of his test with the test proposed by Bradley and Blackwood (1989).

Some data analysts may shy away from permutation tests because of the potentially large amount of computing involved in enumerating the entire permutation distribution. It is urged here not to avoid these methods simply because of the computational intensity. A good reason for this can be found in the work of Jöckel (1986) who quantifies the minimal loss of power when sampling a few thousand elements from the permutation distribution. What is more natural than a statistician taking a sample? Relying instead on the asymptotic χ_2^2 distribution is only an approximation that has no guarantees for small and moderate sample sizes. Sampling from the permutation distribution is still an exact distribution-free procedure.

As the empirical power study in Section 3.2 showed, the proposed tests do quite well under all of the distributions except the heavily skewed lognormal distribution. In fact, E_N does especially well in the elliptically contoured distributions, namely the normal, t, and generalized Laplace. This should not be surprising for the normal distribution since $\mathbf{\Gamma}_N$ was derived under bivariate normality. It turns out that $\delta_{21} = 0$, and hence $\mathbf{\Gamma}$ is diagonal, for every distribution in the broad class of elliptically contoured distributions (see Anderson 1993, eq. 3.18). This explains the success of E_N in the t and generalized Laplace distributions as well.

An attempt was made to expand the permutation distribution by a factor of $n!$ by considering all possible pairings of the differences and sums for each of the 2^n sign changes on the differences. The idea was that since the differences and sums are uncorrelated under $\mathcal{H}_0$, it may be possible to break the pairings to obtain $n!\,2^n$ equally probable permutations of the data. If this was not workable for every distribution, then possibly it would be advantageous for the class of elliptically contoured distributions. An early simulation study revealed that this was not the case. It seems that, although independence of the differences and sums is a sufficient condition for breaking the pairs, uncorrelatedness is not.

4.2. Further Work

Noted economist Thorstein Veblen (1961, p. 33) wrote that "the outcome of any serious research can only be to make two questions grow where one question grew before." This is the nature of any discovery process; it is self-sustaining. While it is hoped that this dissertation has cast some new light on the problem at hand, it is by no means intended as a definitive solution and it suggests some further work to be done. Some possible alternatives and extensions are discussed here.

The various ways to estimate $\mathbf{\Gamma}$ can really be viewed as different methods of combining U_1 and U_2 (and U_1U_2). Other methods of determining these weights could be considered. For instance, there may be reasons to apply more weight to one component or the other depending on the specifics of the problem. Methods for this could be developed.

Rather than using Mahalanobis distance to determine the rejection region of $\mathcal{U}$, it may be useful to consider other methods of partitioning $\mathcal{U}$ into the rejection and non-rejection regions. For example, pairs of horizontal and vertical lines could be drawn that form a box that contains the non-rejection region. Whatever method is developed will have its own distance measure, the only requirement being that the rejection region contain $m = \alpha 2^n$ points of $\mathcal{U}$ so that the level of the test is α.

The symmetry displayed in the permutation distribution of **U** and discussed in Section 3.1.2 suggests that a method for testing more specific directional alternatives than $\mathcal{H}_1$ could be developed. These would be similar to a one-sided test in the more traditional one parameter problem. It is not readily obvious how this should be carried out.

It is clear from Hettmansperger (1984, pp. 279-280) that the ellipse that determines the rejection region, re-centered at the observed value of **U**, provides an approximate $100(1-\alpha)$% confidence region for $(\mu_1 - \mu_2, \sigma_1^2 - \sigma_2^2)$. A more challenging problem is to determine how the permutation test can be inverted to obtain an exact confidence region, perhaps similar to an approach by Maritz (1981, pp. 222-223).

It also remains to be determined how to extend this method beyond the bivariate case to multivariate data. This extension to a repeated measures design does not seem trivial at all and opens up new problems of its own. The fundamental idea of transforming the data to differences and sums will need to be reassessed in favor of an alternative approach.

It is hoped that what has been learned here, especially about the null distribution of our statistics, could be helpful in problems in which bivariate interchangeability is inherent to the problem. See Ernst, Guerra, and Schucany (1996) for an example involving unordered pairs.

More important than the specific extensions of this work is the vast applicability of permutation ideas to a wide variety of problems. While permutation and randomization ideas have been around for decades, they have been woefully under-used due to the lack of computing power. With that barrier vanishing, permutation methods will be developed for every sort of problem that requires only minimal assumptions about the underlying distribution of the data.

APPENDIX

A.1 Proof of Theorem 2.1

This section provides a proof of Theorem 2.1. The key to this proof is understanding the patterned behavior of $\mathfrak{b}_i^k$. Notice that since $k < 2^n$, $\mathfrak{b}_i^k = 0$ for $i > n$, in which case considering $i \in \{1, 2, \ldots, n\}$ is sufficient in finding the binary representation of k. That is, there is an n bit expansion for every integer that we need to index the permutation distribution. For illustration, consider Table A.1 which contains values of $\mathfrak{b}_i^k$ for $i \in \{1, 2, 3, 4\}$ and $k \in \{0, 1, \ldots, 2^4 - 1\}$. Each column gives the binary representation of k.

Table A.1. Values of $\mathfrak{b}_i^k$ for $i = 1, 2, 3, 4$.

		k															
i	2^{i-1}	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2	2	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
3	4	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
4	8	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

Next, consider the behavior of $\mathfrak{b}_i^k = \omega(\lfloor k/2^{i-1} \rfloor)$ for fixed i as k goes from 0 to $2^n - 1$. That is, consider a fixed row in Table A.1. It can be observed that incrementing k by 2^{i-1} will change the value of $\mathfrak{b}_i^k$ (from zero to one, or one to zero). The result is that as k goes from 0 to $2^n - 1$, $\mathfrak{b}_i^k$ exhibits an alternating pattern of 2^{i-1} zeros followed by 2^{i-1} ones. This pattern is repeated 2^{n-i} times. This patterned behavior of $\mathfrak{b}_i^k$ results in the following lemma.

Lemma A.1

Let $i, j \in \{1, 2, \ldots, n\}, i \neq j,$ be fixed. Then as k goes from 0 to $2^n - 1$,

$$\mathfrak{b}_i^k = \begin{cases} 0 & 2^{n-1} \text{ times} \\ 1 & 2^{n-1} \text{ times} \end{cases}$$

and

$$\left(\mathfrak{b}_i^k, \mathfrak{b}_j^k\right) = \begin{cases} (0,0) & 2^{n-2} \text{ times} \\ (0,1) & 2^{n-2} \text{ times} \\ (1,0) & 2^{n-2} \text{ times} \\ (1,1) & 2^{n-2} \text{ times.} \end{cases}$$

We are now in a position to prove the theorem. The proof to each part of the theorem utilizes Lemma A.1. Part (a) can be written as

$$\begin{aligned} \sum_{k=0}^{2^n-1} \sum_{i=1}^{n} a_i (-1)^{\mathfrak{b}_i^k} &= \sum_{i=1}^{n} \left[a_i \sum_{k=0}^{2^n-1} (-1)^{\mathfrak{b}_i^k} \right] = \sum_{i=1}^{n} a_i \left[2^{n-1}(-1)^0 + 2^{n-1}(-1)^1 \right] \\ &= \sum_{i=1}^{n} a_i \left(2^{n-1} - 2^{n-1} \right) = 0. \end{aligned}$$

In a similar fashion for part (b), write

$$\begin{aligned} \sum_{k=0}^{2^n-1} \left[\sum_{i=1}^{n} a_i (-1)^{\mathfrak{b}_i^k} \right] \left[\sum_{j=1}^{n} c_j (-1)^{\mathfrak{b}_j^k} \right] &= \sum_{k=0}^{2^n-1} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i c_j (-1)^{\mathfrak{b}_i^k + \mathfrak{b}_j^k} \\ &= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i c_j \left[\sum_{k=0}^{2^n-1} (-1)^{\mathfrak{b}_i^k + \mathfrak{b}_j^k} \right] + \sum_{k=0}^{2^n-1} \sum_{i=1}^{n} a_i c_i (-1)^{2\mathfrak{b}_i^k} \\ &= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i c_j \left[2^{n-2}(-1)^0 + 2 \cdot 2^{n-2}(-1)^1 + 2^{n-2}(-1)^2 \right] + 2^n \sum_{i=1}^{n} a_i c_i \\ &= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i c_j \left(2^{n-2} - 2^{n-1} + 2^{n-2} \right) + 2^n \sum_{i=1}^{n} a_i c_i = 2^n \sum_{i=1}^{n} a_i c_i \end{aligned}$$

and the proof is complete. ■

A.2 Proof of Theorem 2.2

This section provides a proof of Theorem 2.2. The quantity $\widehat{\mathbf{\Gamma}}_C$ is the population covariance matrix of the elements of $\mathcal{U}$. If we let the ij^{th} element of $\widehat{\mathbf{\Gamma}}_C$ be $\widehat{\gamma}_C^{(i,j)}$, then we can calculate each of the elements directly as

$$\begin{aligned}
\widehat{\gamma}_C^{(1,1)} &= \frac{1}{2^n}\sum_{k=0}^{2^n-1} U_{1\langle k\rangle}^2 = \frac{1}{2^n}\sum_{k=0}^{2^n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(-1)^{\mathfrak{b}_i^k} D_i\right]\left[\frac{1}{n}\sum_{j=1}^{n}(-1)^{\mathfrak{b}_j^k} D_j\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} D_i^2, \\
\widehat{\gamma}_C^{(1,2)} = \widehat{\gamma}_C^{(2,1)} &= \frac{1}{2^n}\sum_{k=0}^{2^n-1} U_{1\langle k\rangle}U_{2\langle k\rangle} \\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(-1)^{\mathfrak{b}_i^k} D_i\right]\left[\frac{1}{n-1}\sum_{j=1}^{n}(-1)^{\mathfrak{b}_j^k} D_j\left(S_j-\overline{S}\right)\right] \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right),
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\gamma}_C^{(2,2)} &= \frac{1}{2^n}\sum_{k=0}^{2^n-1} U_{2\langle k\rangle}^2 \\
&= \frac{1}{2^n}\sum_{k=0}^{2^n-1}\left[\frac{1}{n-1}\sum_{i=1}^{n}(-1)^{\mathfrak{b}_i^k} D_i\left(S_i-\overline{S}\right)\right]\left[\frac{1}{n-1}\sum_{j=1}^{n}(-1)^{\mathfrak{b}_j^k} D_j\left(S_j-\overline{S}\right)\right] \\
&= \frac{1}{(n-1)^2}\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right)^2,
\end{aligned}$$

where the last equality in each case comes from an application of Theorem 2.1(b). ■

A.3 Proof of Theorem 2.3

To derive $\boldsymbol{\Gamma}$, notice that U_1 and U_2 are bivariate one-sample U-statistics of degrees 1 and 2, respectively. The symmetric kernel for U_1 is $h^{(1)}(\mathbf{Y}_i) = D_i$ so that $U_1 = \binom{n}{1}^{-1}\sum_{i=1}^{n} h^{(1)}(\mathbf{Y}_i) = \frac{1}{n}\sum_{i=1}^{n} D_i = \overline{D}$. The symmetric kernel for U_2 is $h^{(2)}(\mathbf{Y}_i, \mathbf{Y}_j) = \frac{1}{2}(D_i - D_j)(S_i - S_j)$ so that

$$\begin{aligned}
U_2 &= \binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} h^{(2)}(\mathbf{Y}_i, \mathbf{Y}_j) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2}(D_i - D_j)(S_i - S_j) \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (D_i S_i - D_i S_j) = \frac{1}{n(n-1)} \left[n \sum_{i=1}^{n} D_i S_i - n^2 \overline{D}\,\overline{S} \right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(D_i - \overline{D}\right)\left(S_i - \overline{S}\right).
\end{aligned}$$

Let $\mathbf{y}_i = (d_i, s_i)'$ be an arbitrary fixed vector. These are involved in standard decompositions of *U*-statistics. Using the notation of Randles and Wolfe (1979), define the conditional expectations

$$\begin{aligned}
h_1^{(1)}(\mathbf{y}_i) &= \mathbb{E}\left[h^{(1)}(\mathbf{Y}_i) \middle|\, \mathbf{Y}_i = \mathbf{y}_i \right] = d_i, \\
h_1^{(2)}(\mathbf{y}_i) &= \mathbb{E}\left[h^{(2)}(\mathbf{Y}_i, \mathbf{Y}_j) \middle|\, \mathbf{Y}_i = \mathbf{y}_i \right] = \mathbb{E}\left[\frac{1}{2}(d_i - D_j)(s_i - S_j) \right] \\
&= \frac{1}{2}[(d_i - \mu_D)(s_i - \mu_S) + \sigma_{DS}],
\end{aligned}$$

and

$$h_2^{(2)}(\mathbf{y}_i, \mathbf{y}_j) = \mathbb{E}\left[h^{(2)}(\mathbf{Y}_i, \mathbf{Y}_j) \middle|\, \mathbf{Y}_i = \mathbf{y}_i, \mathbf{Y}_j = \mathbf{y}_j \right] = \frac{1}{2}(d_i - d_j)(s_i - s_j).$$

Introducing the centered differences and sums $D_i^* = D_i - \mu_D$ and $S_i^* = S_i - \mu_S$, define the quantities

$$
\begin{aligned}
\zeta_1^{(1)} &= \mathbb{V}\Big[h_1^{(1)}(\mathbf{Y}_i)\Big] = \sigma_D^2, \\
\zeta_1^{(2)} &= \mathbb{V}\Big[h_1^{(2)}(\mathbf{Y}_i)\Big] = \frac{1}{4}\mathbb{V}[(D_i-\mu_D)(S_i-\mu_S)] \\
&= \frac{1}{4}\big\{\mathbb{E}\big[D_i^{*2}S_i^{*2}\big] - \mathbb{E}[D_i^*S_i^*]^2\big\} = \frac{1}{4}\big(\delta_{22}-\sigma_{DS}^2\big), \\
\zeta_2^{(2)} &= \mathbb{V}\Big[h_2^{(2)}(\mathbf{Y}_i,\mathbf{Y}_j)\Big] = \frac{1}{4}\mathbb{V}[(D_i-D_j)(S_i-S_j)] \\
&= \frac{1}{4}\mathbb{V}\big[\big(D_i^*-D_j^*\big)\big(S_i^*-S_j^*\big)\big] \\
&= \frac{1}{4}\mathbb{V}\big[D_i^*S_i^* - D_i^*S_j^* - D_j^*S_i^* + D_j^*S_j^*\big] \\
&= \frac{1}{4}\Big\{\mathbb{E}\Big[\big(D_i^*S_i^* - D_i^*S_j^* - D_j^*S_i^* + D_j^*S_j^*\big)^2\Big] \\
&\qquad - \mathbb{E}\big[D_i^*S_i^* - D_i^*S_j^* - D_j^*S_i^* + D_j^*S_j^*\big]^2\Big\} \\
&= \frac{1}{4}\big\{\mathbb{E}\big[D_i^{*2}S_i^{*2} - 2D_i^{*2}S_i^*S_j^* - 2D_i^*D_j^*S_i^{*2} + 4D_i^*S_i^*D_j^*S_j^* + D_i^{*2}S_j^{*2} \\
&\qquad - 2D_i^*D_j^*S_j^{*2} + D_j^{*2}S_i^{*2} - 2D_j^{*2}S_i^*S_j^* + D_j^{*2}S_j^{*2}\big] - (2\sigma_{DS})^2\big\} \\
&= \frac{1}{4}\big(2\delta_{22} + 2\sigma_D^2\sigma_S^2 + 4\sigma_{DS}^2 - 4\sigma_{DS}^2\big) = \frac{1}{2}\big(\delta_{22}+\sigma_D^2\sigma_S^2\big),
\end{aligned}
$$

and

$$
\begin{aligned}
\zeta_1^{(1,2)} &= \mathbb{C}\Big[h_1^{(1)}(\mathbf{Y}_i), h_1^{(2)}(\mathbf{Y}_i)\Big] = \mathbb{C}\Big[D_i, \frac{1}{2}(D_i^*S_i^*+\sigma_{DS})\Big] \\
&= \frac{1}{2}\mathbb{C}[D_i^*, D_i^*S_i^*] = \frac{1}{2}\big\{\mathbb{E}\big[D_i^{*2}S_i^*\big] - \mathbb{E}[D_i^*]\mathrm{E}[D_i^*S_i^*]\big\} \\
&= \frac{1}{2}\delta_{21}.
\end{aligned}
$$

Lee (1990) gives the variance of a *U*-statistic (Theorem 3, p. 12) and the covariance between two *U*-statistics (Theorem 2, p. 17) in terms of these ζs [similar results may be found in Randles and Wolfe (1979)]. Using these theorems, the elements of $\mathbf{\Gamma}$ can be found to be

$$\gamma_{11} = \mathbb{V}[U_1] = \binom{n}{1}^{-1}\binom{1}{1}\binom{n-1}{1-1}\zeta_1^{(1)} = \frac{\sigma_D^2}{n},$$

$$\begin{aligned}
\gamma_{22} &= \mathbb{V}[U_2] = \binom{n}{2}^{-1}\sum_{c=1}^{2}\binom{2}{c}\binom{n-2}{2-c}\zeta_c^{(2)} \\
&= \frac{2}{n(n-1)}\left[2(n-2)\frac{1}{4}\left(\delta_{22} - \sigma_{DS}^2\right) + \frac{1}{2}\left(\delta_{22} + \sigma_D^2\sigma_S^2\right)\right] \\
&= \frac{1}{n(n-1)}\left[(n-1)\delta_{22} - (n-2)\sigma_{DS}^2 + \sigma_D^2\sigma_S^2\right],
\end{aligned}$$

and

$$\gamma_{12} = \mathbb{C}[U_1, U_2] = \binom{n}{1}^{-1}\binom{2}{1}\binom{n-2}{1-1}\zeta_1^{(1,2)} = \frac{\delta_{21}}{n}. \qquad \blacksquare$$

A.4 Relationships Between Tests Assuming Bivariate Normality

This section establishes the equivalence of the two tests for equal means and variances in bivariate normal data proposed by Hsu (1940) and Bradley and Blackwood (1989). In addition, it is shown that these two test statistics are monotonic functions of E_N.

We will proceed by showing the relationship between E_N and each of the two test statistics. This will allow us to then show the equivalence of the two normal theory tests. Before proceeding, it can be easily seen from ⟦2.8⟧ that

$$E_N = \mathbf{U}'\widehat{\mathbf{\Gamma}}_N^{-1}\mathbf{U} = \frac{nU_1^2}{\widehat{\sigma}_D^2} + \frac{(n-1)U_2^2}{\widehat{\sigma}_D^2\widehat{\sigma}_S^2},$$

where $\widehat{\sigma}_D^2 = \frac{1}{n}\sum_{i=1}^{n} D_i^2$ and $\widehat{\sigma}_S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(S_i - \overline{S}\right)^2$.

Hsu (1940) gives a monotone function of the likelihood ratio, λ, and derives its distribution. More specifically, after correcting a typographical error in his equation (9),

$$L_5 = \lambda^{2/n} = \frac{4s_1^2s_2^2(1-r_{12}^2)}{\left[s_1^2 + s_2^2 + \frac{1}{2}(\overline{x}_1 - \overline{x}_2)^2\right]^2 - \left[2s_1s_2r_{12} - \frac{1}{2}(\overline{x}_1 - \overline{x}_2)^2\right]^2},$$

where $(\overline{x}_1, \overline{x}_2, s_1^2, s_2^2, r_{12})$ are the MLEs of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. To facilitate the change to our notation, notice that

$$
\begin{aligned}
s_1^2 + s_2^2 - 2r_{12}s_1s_2 &= \frac{1}{n}\sum_{i=1}^{n}\bigl(D_i - \overline{D}\bigr)^2 = \frac{1}{n}\sum_{i=1}^{n} D_i^2 - \overline{D}^2 = \widehat{\sigma}_D^2 - U_1^2,\\
s_1^2 + s_2^2 + 2r_{12}s_1s_2 &= \frac{1}{n}\sum_{i=1}^{n}\bigl(S_i - \overline{S}\bigr)^2 = \frac{n-1}{n}\widehat{\sigma}_S^2,
\end{aligned}
$$

and

$$
\begin{aligned}
s_1^2 - s_2^2 &= \frac{1}{n}\sum_{i=1}^{n}\Bigl[\bigl(X_{1i} - \overline{X}_1\bigr)^2 - \bigl(X_{2i} - \overline{X}_2\bigr)^2\Bigr]\\
&= \frac{1}{n}\sum_{i=1}^{n}\bigl(X_{1i} - \overline{X}_1 - X_{2i} + \overline{X}_2\bigr)\bigl(X_{1i} - \overline{X}_1 + X_{2i} - \overline{X}_2\bigr)\\
&= \frac{1}{n}\sum_{i=1}^{n}\bigl(D_i - \overline{D}\bigr)\bigl(S_i - \overline{S}\bigr) = \frac{n-1}{n}U_2.
\end{aligned}
$$

Now, by expanding the numerator and denominator of L_5 and adding and subtracting $(s_1^2 + s_2^2)^2$ in the numerator, we get

$$
\begin{aligned}
L_5 &= \frac{(s_1^2+s_2^2)^2 - 4s_1^2s_2^2r_{12}^2 - (s_1^2+s_2^2)^2 + 4s_1^2s_2^2}{(s_1^2+s_2^2)^2 + \frac{1}{4}U_1^2 + U_1^2(s_1^2+s_2^2) - 4s_1^2s_2^2r_{12}^2 - \frac{1}{4}U_1^2 + 2s_1s_2r_{12}U_1^2}\\
&= \frac{\Bigl[(s_1^2+s_2^2)^2 - 4s_1^2s_2^2r_{12}^2\Bigr] - \Bigl[(s_1^2+s_2^2)^2 - 4s_1^2s_2^2\Bigr]}{(s_1^2+s_2^2)^2 - 4s_1^2s_2^2r_{12}^2 + U_1^2(s_1^2+s_2^2+2s_1s_2r_{12})}\\
&= \frac{(s_1^2+s_2^2+2s_1s_2r_{12})(s_1^2+s_2^2-2s_1s_2r_{12}) - (s_1^2-s_2^2)^2}{(s_1^2+s_2^2+2s_1s_2r_{12})(s_1^2+s_2^2-2s_1s_2r_{12}) + U_1^2(s_1^2+s_2^2+2s_1s_2r_{12})}\\
&= \frac{\frac{n-1}{n}\widehat{\sigma}_S^2\bigl(\widehat{\sigma}_D^2 - U_1^2\bigr) - \bigl(\frac{n-1}{n}U_2\bigr)^2}{\frac{n-1}{n}\widehat{\sigma}_S^2\bigl(\widehat{\sigma}_D^2 - U_1^2\bigr) + \frac{n-1}{n}U_1^2\widehat{\sigma}_S^2} = \frac{\widehat{\sigma}_S^2\bigl(\widehat{\sigma}_D^2 - U_1^2\bigr) - \frac{n-1}{n}U_2^2}{\widehat{\sigma}_D^2\widehat{\sigma}_S^2}\\
&= 1 - \frac{U_1^2}{\widehat{\sigma}_D^2} - \frac{(n-1)U_2^2}{n\widehat{\sigma}_D^2\widehat{\sigma}_S^2} = 1 - \frac{1}{n}\left[\frac{nU_1^2}{\widehat{\sigma}_D^2} + \frac{(n-1)U_2^2}{\widehat{\sigma}_D^2\widehat{\sigma}_S^2}\right] = 1 - \frac{E_N}{n}.
\end{aligned}
$$

Therefore, L_5 is a monotone function of E_N.

The test proposed by Bradley and Blackwood (1989) comes from regressing the differences on the sums. They show that testing $\mathcal{H}_0'\colon \beta_0 = \beta_1 = 0$, where β_0 and β_1 are the intercept and slope coefficients, respectively, is equivalent to testing $\mathcal{H}_0$. The statistic for testing $\mathcal{H}_0'$ has an F distribution with 2 and $n-2$ degrees of freedom and is given by

$$F = \frac{\left(\sum_{i=1}^{n} D_i^2 - \text{SSE}\right) \Big/ 2}{\text{SSE}/(n-2)},$$

where $\text{SSE} = \sum_{i=1}^{n} \left(D_i - \widehat{\beta}_0 - \widehat{\beta}_1 S_i\right)^2$ is the error sum of squares, $\widehat{\beta}_0 = \overline{D} - \widehat{\beta}_1 \overline{S}$, and $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})(S_i - \overline{S})}{\sum_{i=1}^{n}(S_i - \overline{S})^2} = U_2/\widehat{\sigma}_S^2$ (see for example Mason, Gunst, and Hess 1989, p. 450).

Now, we can write SSE as

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n} \left[D_i - \left(\overline{D} - \widehat{\beta}_1 \overline{S}\right) - \widehat{\beta}_1 S_i\right]^2 = \sum_{i=1}^{n} \left[D_i - \overline{D} - \widehat{\beta}_1 \left(S_i - \overline{S}\right)\right]^2 \\
&= \sum_{i=1}^{n} \left[\left(D_i - \overline{D}\right)^2 + \widehat{\beta}_1^2 \left(S_i - \overline{S}\right)^2 - 2\widehat{\beta}_1 \left(S_i - \overline{S}\right)\left(D_i - \overline{D}\right)\right] \\
&= \sum_{i=1}^{n} D_i^2 - n\overline{D}^2 + \widehat{\beta}_1^2 (n-1)\widehat{\sigma}_S^2 - 2\widehat{\beta}_1 (n-1) U_2 \\
&= n\widehat{\sigma}_D^2 - nU_1^2 - \frac{(n-1)U_2^2}{\widehat{\sigma}_S^2} = \widehat{\sigma}_D^2 (n - E_N).
\end{aligned}$$

Therefore, the F statistic is

$$F = \frac{n-2}{2}\left[\frac{n\widehat{\sigma}_D^2 - \widehat{\sigma}_D^2(n - E_N)}{\widehat{\sigma}_D^2(n - E_N)}\right] = \frac{n-2}{2}\left(\frac{E_N}{n - E_N}\right),$$

which is a monotone function of E_N.

Now, substituting $E_N = n(1 - L_5)$ into the previous result, we see that

$$F = \frac{n-2}{2}\left(\frac{1 - L_5}{L_5}\right).$$

Hsu (1940) showed that L_5 has a beta distribution with parameters $\frac{1}{2}(n-2)$ and 1. The relationship between a beta random variable, $B_{\omega,\nu}$, and the F distribution can be written as

$$\frac{\omega(1 - B_{\omega,\nu})}{\nu B_{\omega,\nu}} \sim F_{2\nu,2\omega}$$

(see Evans, Hastings, and Peacock 1993, p. 35). Applying this to L_5, we see that

$$\frac{n-2}{2}\left(\frac{1-L_5}{L_5}\right) \sim F_{2,n-2}.$$

Therefore, the likelihood ratio test of Hsu (1940) and the F test given by Bradley and Blackwood (1989) are equivalent tests. ■

A.5 Proof of Theorem 2.4

This section provides a proof of Theorem 2.4. Each of the four methods of estimating $\mathbf{\Gamma}$ will be considered separately.

Considering $M = N$ first, it is clear from ⟦2.8⟧ that $\left|\widehat{\mathbf{\Gamma}}_N\right| = \frac{1}{n(n-1)}\widehat{\sigma}_D^4\widehat{\sigma}_S^2$, which is clearly nonnegative.

Next for $M = C$, ⟦2.3⟧ and an application of the Cauchy-Schwarz inequality implies that

$$\left|\widehat{\mathbf{\Gamma}}_C\right| = \frac{1}{n^2(n-1)^2}\left\{\left[\sum_{i=1}^{n} D_i^2\right]\left[\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right)^2\right] - \left[\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right)\right]^2\right\} \geq 0.$$

The proof for $\widehat{\mathbf{\Gamma}}_I$ follows easily from that for $\widehat{\mathbf{\Gamma}}_C$. Simply notice from ⟦2.7⟧ that

$$\begin{aligned}
\left|\widehat{\mathbf{\Gamma}}_I\right| &= \frac{1}{n^2}\left\{\widehat{\sigma}_D^2\left[\widehat{\delta}_{22} + \widehat{\sigma}_D^2\widehat{\sigma}_S^2/(n-1)\right] - \widehat{\delta}_{21}^2\right\} \geq \frac{1}{n^2}\left(\widehat{\sigma}_D^2\widehat{\delta}_{22} - \widehat{\delta}_{21}^2\right) \\
&= \frac{1}{n^4}\left\{\left[\sum_{i=1}^{n} D_i^2\right]\left[\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right)^2\right] - \left[\sum_{i=1}^{n} D_i^2\left(S_i-\overline{S}\right)\right]^2\right\} \\
&= \frac{(n-1)^2}{n^2}\left|\widehat{\mathbf{\Gamma}}_C\right| \geq 0.
\end{aligned}$$

The proof for $\widehat{\mathbf{\Gamma}}_{P\langle k\rangle}$ is a bit more complicated. The method of proof is valid for all 2^n permutations of the data, therefore we concentrate on $\widehat{\mathbf{\Gamma}}_{P\langle k\rangle}$ calculated from the observed data, namely $\widehat{\mathbf{\Gamma}}_{P\langle 0\rangle}$.

Begin by defining $A_i = \left(D_i - \overline{D}\right)\left(S_i - \overline{S}\right), i = 1, \ldots, n$. Recalling the notation used in ⟦2.4⟧, notice that $\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)\left(S_i - \overline{S}\right) = \frac{n}{n-1}\widehat{\sigma}_{DS\langle 0\rangle}$. Define $\widehat{\mathbf{\Theta}}$ to be the maximum likelihood estimate of the covariance matrix of the D_is and A_is. Then the elements of $\widehat{\mathbf{\Theta}}$ are

$$\begin{aligned}
\widehat{\theta}_{11} &= \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2 = \frac{n}{n-1}\widehat{\sigma}^2_{D\langle 0\rangle}, \\
\widehat{\theta}_{12} &= \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)\left(A_i - \overline{A}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)A_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2\left(S_i - \overline{S}\right) = \widehat{\delta}_{21\langle 0\rangle},
\end{aligned}$$

and

$$\widehat{\theta}_{22} = \frac{1}{n}\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}A_i^2 - \overline{A}^2 = \widehat{\delta}_{22\langle 0\rangle} - \frac{n^2}{(n-1)^2}\widehat{\sigma}^2_{DS\langle 0\rangle}.$$

Now, by applying the Cauchy-Schwarz inequality as before, we see that

$$\begin{aligned}
\left|\widehat{\mathbf{\Theta}}\right| &= \widehat{\theta}_{11}\widehat{\theta}_{22} - \widehat{\theta}^2_{12} \\
&= \frac{1}{n^2}\left\{\left[\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2\right]\left[\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2\right] - \left[\sum_{i=1}^{n}\left(D_i - \overline{D}\right)\left(A_i - \overline{A}\right)\right]^2\right\} \\
&\geq 0.
\end{aligned}$$

But we can also write

$$\begin{aligned}
\left|\widehat{\mathbf{\Theta}}\right| &= \widehat{\theta}_{11}\widehat{\theta}_{22} - \widehat{\theta}^2_{12} = \frac{n}{n-1}\widehat{\sigma}^2_{D\langle 0\rangle}\left[\widehat{\delta}_{22\langle 0\rangle} - \frac{n^2}{(n-1)^2}\widehat{\sigma}^2_{DS\langle 0\rangle}\right] - \widehat{\delta}^2_{21\langle 0\rangle} \\
&= \widehat{\sigma}^2_{D\langle 0\rangle}\left[\frac{n}{n-1}\widehat{\delta}_{22\langle 0\rangle} - \frac{n^3}{(n-1)^3}\widehat{\sigma}^2_{DS\langle 0\rangle}\right] - \widehat{\delta}^2_{21\langle 0\rangle} \\
&\leq \widehat{\sigma}^2_{D\langle 0\rangle}\left[\frac{n}{n-1}\widehat{\delta}_{22\langle 0\rangle} - \frac{n-2}{n-1}\widehat{\sigma}^2_{DS\langle 0\rangle}\right] - \widehat{\delta}^2_{21\langle 0\rangle},
\end{aligned} \qquad ⟦A.1⟧$$

where the inequality results from the fact that $\frac{n^3}{(n-1)^3} \geq \frac{n-2}{n-1}$. Another application of the Cauchy-Schwarz inequality shows that $\frac{1}{n-1}\left(\widehat{\sigma}^2_{D\langle 0\rangle}\widehat{\sigma}^2_S - \widehat{\delta}_{22\langle 0\rangle}\right) \geq 0$. Adding this quantity inside the brackets in ⟦A.1⟧ and comparing the result to ⟦2.5⟧ leads to the required string of inequalities

$$0 \leq \left|\widehat{\mathbf{\Theta}}\right| \leq \widehat{\sigma}^2_{D\langle 0\rangle}\left[\widehat{\delta}_{22\langle 0\rangle} - \frac{n-2}{n-1}\widehat{\sigma}^2_{DS\langle 0\rangle} + \frac{\widehat{\sigma}^2_{D\langle 0\rangle}\widehat{\sigma}^2_S}{n-1}\right] - \widehat{\delta}^2_{21\langle 0\rangle} = n^2\left|\widehat{\mathbf{\Gamma}}_{P\langle 0\rangle}\right|.$$

Therefore, $\left|\widehat{\mathbf{\Gamma}}_M\right| \geq 0$ for $M = C, P, I, N$. ■

A.6 Proof of Theorem 2.5

This section provides a proof of Theorem 2.5 establishing the asymptotic distribution of E_M. Recall that in Section A.3 it was noted that U_1 and U_2 are bivariate one-sample *U*-statistics of degrees 1 and 2, respectively. Then it follows from the work there and Theorem 3.6.9 of Randles and Wolfe (1979, p. 107) concerning the asymptotic normality of *U*-statistics that

$$\sqrt{n}\mathbf{U} \xrightarrow{\mathcal{D}} N_2(\boldsymbol{\nu}, \mathbf{\Delta}),$$

as $n \longrightarrow \infty$, where $\boldsymbol{\nu} = (\mu_D, \sigma_{DS})'$ and $\mathbf{\Delta} = \begin{bmatrix} \sigma^2_D & \delta_{21} \\ \delta_{21} & \delta_{22} - \sigma^2_{DS} \end{bmatrix}$.

Now, consider $\widehat{\mathbf{\Gamma}}_P$ in ⟦2.5⟧. It is easily seen from Sections 2.2.2 and 2.2.3 of Serfling (1980, pp.68-69) that all the sample estimates in $\widehat{\mathbf{\Gamma}}_P$ converge in probability to their population counterparts and hence $n\widehat{\mathbf{\Gamma}}_P \xrightarrow{\mathcal{P}} \mathbf{\Delta}$ as $n \longrightarrow \infty$. Furthermore, $\frac{1}{n}\widehat{\mathbf{\Gamma}}^{-1}_P \xrightarrow{\mathcal{P}} \mathbf{\Delta}^{-1}$. Therefore, an application of Theorem 3.4.8 in Sen and Singer (1993, pp. 137-138) implies that

$$E_P = \left(\sqrt{n}\mathbf{U}\right)'\left(\frac{1}{n}\widehat{\mathbf{\Gamma}}^{-1}_P\right)\left(\sqrt{n}\mathbf{U}\right) \xrightarrow{\mathcal{D}} \chi'^2_2(\boldsymbol{\nu}'\mathbf{\Delta}\boldsymbol{\nu}),$$

as $n \longrightarrow \infty$, where $\boldsymbol{\nu}'\mathbf{\Delta}\boldsymbol{\nu} = \mu^2_D\sigma^2_D + 2\mu_D\sigma_{DS}\delta_{21} + \sigma^2_{DS}(\delta_{22} - \sigma^2_{DS})$ is the noncentrality parameter.

When $\mathcal{H}_0$ is true, $\mu_D = \sigma_{DS} = 0$ and hence, as $n \longrightarrow \infty$,

$$\sqrt{n}\mathbf{U} \xrightarrow{\mathcal{D}} N_2(\boldsymbol{\nu}_0, \boldsymbol{\Delta}_0),$$

where $\boldsymbol{\nu}_0 = \mathbf{0}$ and $\boldsymbol{\Delta}_0 = \begin{bmatrix} \sigma_D^2 & \delta_{21} \\ \delta_{21} & \delta_{22} \end{bmatrix}$. Again it is clear from [[2.3]], [[2.5]], and [[2.7]] that for $M = C, P, I,\ \ n\widehat{\boldsymbol{\Gamma}}_M \xrightarrow{\mathcal{P}} \boldsymbol{\Delta}_0$ and therefore $\frac{1}{n}\widehat{\boldsymbol{\Gamma}}_M^{-1} \xrightarrow{\mathcal{P}} \boldsymbol{\Delta}_0^{-1}$, as $n \longrightarrow \infty$. The result of another application of Theorem 3.4.8 in Sen and Singer (1993, pp. 137-138) is that

$$E_M \xrightarrow{\mathcal{D}} \chi_2^2$$

for $M = C, P, I,$ as $n \longrightarrow \infty$.

It only remains to be shown that under $\mathcal{H}_0$, $E_N \xrightarrow{\mathcal{D}} \chi_2^2$ as $n \longrightarrow \infty$. The proof depends on the additional assumption that $F(\,\cdot\,,\,\cdot\,)$, the underlying cdf of **X**, satisfies [[1.1]] under $\mathcal{H}_1$ as well as $\mathcal{H}_0$. In Section A.4, the relationship between E_N and λ, the likelihood ratio under bivariate normality, was established, namely $\lambda = (1 - E_N/n)^{n/2}$. Since the number of constraints imposed by $\mathcal{H}_0$ is 2 ($\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$), then Wilk's likelihood ratio statistic, $-2\ln\lambda$, is asymptotically χ_2^2. But

$$\begin{aligned} -2\ln\lambda &= -2\ln\left[\left(1 - \frac{E_N}{n}\right)^{n/2}\right] = -n\ln\left(1 - \frac{E_N}{n}\right) = -n\sum_{i=1}^{\infty}\frac{1}{i}\left(-\frac{E_N}{n}\right)^i \\ &= E_N + O_p\left(n^{-1}\right), \end{aligned}$$

and therefore $E_N \xrightarrow{\mathcal{D}} \chi_2^2$. ■

A.7. Personal Communication With James L. Kepner

This section contains a partial text of a personal communication with James L. Kepner on January 31, 1996, via e-mail. This communication is referred to in Section 3.1.2 and clarifies a typographical error in Kepner and Randles (1984).

```
From: "Dr. James Kepner" <jim@pogstat1.pog.ufl.edu>
To: "Mike Ernst" <mernst@post.cis.smu.edu>
Subject: Re: Kepner and Randles (1984)
Date: Wed, 31 Jan 96 13:49:41 -0500
```

```
Mike:

So far, your problem has been my most interesting of the day.  It
turns out, the column on P. 923 labeled ARE(S*AB,K) should have
been labeled ARE(S*MD,K).   The reverse is true for the next
column to the right.   So, the columns in the table are
incorrectly labeled, but the text on the following pages is
correct.  Good eyes!!  The errors are in everything that I wrote
relating to this topic, including the UF Tech Report #160, EXCEPT
my Ph.D. thesis, where at last I found it written correctly!
```

A.8. Results of Empirical Power Study

This section contains the numerical results from part of the empirical power study discussed in Section 3.2. Tables A.2(a) to A.2(e) each display the empirical power (in percent) based on 10,000 samples of size $n = 20$ from one of the five bivariate distributions. The seven tests were all conducted at the $\alpha = .05$ significance level.

Table A.2(a). Empirical power (in percent) based on 10,000 samples of size $n = 20$ from the bivariate normal distribution.

		$\rho = 0$			$\rho = .5$			$\rho = .8$		
		σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}
E_C	μ_{20}	4.8	16.8	36.1	4.8	21.5	45.7	5.0	37.8	70.6
	μ_{21}	9.2	19.8	38.9	14.0	26.6	49.1	29.4	47.4	73.3
	μ_{22}	24.4	30.0	44.9	45.0	45.3	57.8	83.9	77.1	82.5
E_P	μ_{20}	4.8	17.3	38.4	4.6	22.5	47.8	5.1	40.1	74.2
	μ_{21}	9.2	21.4	41.6	13.6	29.0	53.2	28.0	52.6	79.2
	μ_{22}	23.8	32.5	49.5	43.5	49.4	64.9	81.9	82.1	89.4
E_I	μ_{20}	4.8	17.6	38.5	4.8	22.7	48.3	5.0	40.2	73.4
	μ_{21}	9.1	20.5	40.7	13.7	27.7	51.6	28.7	49.0	76.1
	μ_{22}	23.8	30.4	46.9	44.0	45.7	59.7	83.2	77.6	84.2
E_N	μ_{20}	4.6	19.6	46.1	4.7	25.3	56.8	5.1	47.8	85.6
	μ_{21}	9.5	23.8	48.0	14.8	32.8	62.1	30.7	59.6	88.7
	μ_{22}	25.4	35.6	56.8	46.2	53.7	72.7	83.9	86.7	95.3
K	μ_{20}	4.6	17.1	38.4	4.8	21.6	47.0	5.0	39.2	75.2
	μ_{21}	8.6	19.9	40.0	13.5	27.2	51.4	27.5	48.9	78.3
	μ_{22}	23.2	29.8	46.6	43.0	44.9	61.2	81.7	77.6	87.3
S	μ_{20}	5.1	16.6	36.5	4.6	20.7	44.5	5.0	35.9	69.4
	μ_{21}	8.4	20.0	38.9	12.9	25.8	48.7	24.0	45.1	73.6
	μ_{22}	22.5	29.0	45.6	39.6	42.8	58.5	75.5	73.5	84.4
F	μ_{20}	4.9	23.5	54.1	4.7	30.4	65.7	5.1	56.8	91.6
	μ_{21}	9.1	27.4	56.2	14.2	37.4	70.4	28.9	66.6	93.7
	μ_{22}	24.2	38.3	64.0	44.0	57.1	79.8	83.1	89.7	97.3

Table A.2(b). Empirical power (in percent) based on 10,000 samples of size $n = 20$ from the bivariate t distribution.

		$\rho = 0$			$\rho = .5$			$\rho = .8$		
		σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}
E_C	μ_{20}	5.3	11.7	23.9	4.8	14.4	29.3	5.0	24.2	47.0
	μ_{21}	12.0	22.5	31.6	34.8	35.1	42.4	68.5	61.9	64.8
	μ_{22}	60.0	51.8	53.1	85.3	75.8	72.6	99.0	96.2	92.6
E_P	μ_{20}	5.2	13.3	27.0	4.9	16.1	33.6	50.1	28.3	53.6
	μ_{21}	19.2	24.5	35.5	33.5	37.6	48.9	66.8	66.6	73.2
	μ_{22}	58.4	53.3	58.2	83.9	76.8	77.0	98.8	96.7	95.1
E_I	μ_{20}	5.3	12.2	26.0	4.8	15.3	31.8	5.1	26.5	50.3
	μ_{21}	19.4	22.7	33.1	34.2	35.3	44.2	67.6	62.2	67.0
	μ_{22}	59.1	51.6	53.8	84.6	75.6	72.9	99.0	96.2	92.9
E_N	μ_{20}	5.6	18.0	39.8	4.8	22.2	48.2	5.2	40.1	76.2
	μ_{21}	16.7	27.2	45.3	28.8	39.7	59.9	59.4	69.6	85.5
	μ_{22}	51.3	50.8	62.9	77.8	73.5	79.6	97.1	95.3	96.5
K	μ_{20}	5.3	15.5	34.7	4.9	19.3	42.7	5.0	35.2	69.4
	μ_{21}	18.8	25.3	40.8	32.9	38.3	53.6	67.0	68.0	81.9
	μ_{22}	59.1	53.2	59.9	84.9	77.0	78.8	99.3	97.0	96.4
S	μ_{20}	5.3	15.3	33.5	5.0	19.0	40.9	4.9	32.1	64.6
	μ_{21}	18.2	24.7	39.5	30.2	36.4	52.1	62.5	63.7	77.5
	μ_{22}	57.2	52.1	58.8	82.6	74.7	77.1	98.5	95.4	95.4
F	μ_{20}	13.1	31.4	56.1	11.9	36.6	64.1	12.0	56.0	86.4
	μ_{21}	23.8	39.3	59.3	36.4	51.5	72.0	65.8	79.3	92.2
	μ_{22}	57.8	59.8	73.2	82.6	80.6	87.1	98.3	97.4	98.3

Table A.2(c). Empirical power (in percent) based on 10,000 samples of size $n = 20$ from the bivariate generalized Laplace distribution.

		$\rho = 0$			$\rho = .5$			$\rho = .8$		
		σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}
E_C	μ_{20}	5.2	21.2	48.0	4.6	27.4	58.9	5.0	49.3	84.9
	μ_{21}	8.0	23.4	48.7	12.2	31.9	59.0	25.6	53.8	84.2
	μ_{22}	20.7	30.6	51.3	39.9	43.4	64.3	82.3	75.3	87.4
E_P	μ_{20}	5.3	21.1	48.1	4.7	27.4	59.0	4.8	49.0	85.7
	μ_{21}	8.2	24.2	49.7	12.1	34.0	61.0	25.0	58.8	87.3
	μ_{22}	20.4	33.2	55.1	38.4	48.4	70.4	79.3	81.9	92.8
E_I	μ_{20}	5.2	22.1	50.0	4.6	28.7	61.0	5.0	51.0	86.8
	μ_{21}	7.9	24.3	50.6	12.0	32.8	61.1	25.1	55.6	85.9
	μ_{22}	20.4	31.2	53.1	39.2	44.0	66.1	81.5	76.1	89.2
E_N	μ_{20}	5.3	21.4	51.5	4.8	27.3	63.8	4.8	52.5	91.3
	μ_{21}	9.7	25.5	52.9	14.7	36.2	67.2	30.3	64.4	92.3
	μ_{22}	25.2	36.8	59.8	45.6	53.4	76.1	85.1	86.9	96.8
K	μ_{20}	5.2	19.7	44.1	4.5	24.7	54.7	5.0	45.0	82.0
	μ_{21}	8.0	21.2	44.3	11.8	29.5	56.5	24.3	52.0	83.0
	μ_{22}	19.5	29.1	48.8	37.5	41.5	62.1	78.4	73.3	88.5
S	μ_{20}	5.0	18.3	41.1	4.8	23.5	50.8	5.0	40.6	76.8
	μ_{21}	7.7	21.3	41.7	11.0	28.2	52.6	21.2	48.0	78.0
	μ_{22}	19.5	28.9	46.9	34.5	40.4	60.2	72.1	71.1	85.7
F	μ_{20}	3.7	21.1	54.2	3.4	27.6	67.7	3.6	55.4	94.7
	μ_{21}	7.1	24.8	55.6	11.4	35.9	70.6	25.7	67.9	95.4
	μ_{22}	20.6	35.8	62.9	40.4	53.8	79.9	82.3	89.5	98.5

Table A.2(d). Empirical power (in percent) based on 10,000 samples of size $n = 20$ from the bivariate Cook-Johnson distribution with normal marginals.

		$\rho = 0$			$\rho = .5$			$\rho = .8$		
		σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}
E_C	μ_{20}	5.0	17.8	36.6	4.9	27.1	53.0	4.8	47.1	73.5
	μ_{21}	8.8	20.5	39.6	16.5	24.9	50.0	47.1	35.5	66.4
	μ_{22}	24.8	29.2	44.5	51.3	37.2	53.0	93.8	61.0	69.0
E_P	μ_{20}	5.0	18.4	38.6	5.1	27.9	55.4	4.8	48.8	76.5
	μ_{21}	8.4	22.0	42.1	15.7	27.7	53.3	43.5	41.6	71.5
	μ_{22}	23.8	31.9	49.1	49.8	41.5	59.5	93.9	66.1	77.7
E_I	μ_{20}	5.0	18.7	38.9	4.9	28.4	55.8	4.9	49.1	76.3
	μ_{21}	8.5	21.2	41.7	16.1	26.3	52.7	45.5	38.3	70.2
	μ_{22}	24.1	29.7	46.2	50.4	38.1	55.5	93.3	62.2	72.3
E_N	μ_{20}	4.8	20.2	45.6	5.0	26.2	59.6	5.0	43.0	80.5
	μ_{21}	9.0	24.4	49.3	15.7	33.0	61.5	26.7	50.8	81.3
	μ_{22}	25.4	35.0	55.7	46.9	50.3	70.0	79.2	75.7	88.5
K	μ_{20}	5.1	18.1	37.7	4.9	26.5	55.0	4.8	50.1	79.3
	μ_{21}	8.5	21.1	40.7	15.8	25.3	53.2	44.4	43.0	76.5
	μ_{22}	23.4	29.2	45.5	49.7	37.9	57.3	93.1	64.3	79.9
S	μ_{20}	5.2	17.5	35.9	5.2	24.5	50.5	4.8	45.4	75.4
	μ_{21}	8.3	20.3	39.0	14.3	24.8	49.8	37.0	38.7	70.5
	μ_{22}	23.1	28.9	44.9	43.7	36.8	54.1	87.7	62.9	75.6
F	μ_{20}	5.0	23.9	54.2	5.0	30.6	68.2	7.6	54.9	89.7
	μ_{21}	8.7	27.9	57.5	13.9	37.7	70.3	29.6	61.7	89.2
	μ_{22}	24.1	37.9	62.7	44.3	54.2	77.0	82.2	81.9	94.0

Table A.2(e). Empirical power (in percent) based on 10,000 samples of size $n = 20$ from the bivariate lognormal distribution.

		$\rho = 0$			$\rho = .5$			$\rho = .8$		
		σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}	σ^2_{20}	σ^2_{21}	σ^2_{22}
E_C	μ_{20}	5.0	18.6	34.3	4.9	28.8	51.1	5.0	47.6	66.7
	μ_{21}	37.9	8.2	8.9	65.4	14.0	11.2	90.4	30.6	17.4
	μ_{22}	84.4	45.3	20.5	96.4	80.0	44.8	99.9	98.8	84.6
E_P	μ_{20}	5.2	19.4	37.4	5.0	30.4	55.7	4.9	51.8	72.9
	μ_{21}	33.4	8.2	12.2	60.8	12.5	16.4	89.0	25.9	28.9
	μ_{22}	81.6	37.3	17.6	96.1	72.1	36.6	99.9	97.6	74.9
E_I	μ_{20}	4.9	19.6	36.4	4.9	30.0	53.1	5.0	49.6	69.2
	μ_{21}	37.5	7.9	9.8	64.5	13.1	12.4	89.8	28.1	19.8
	μ_{22}	83.5	43.3	19.1	96.2	78.0	41.5	99.8	98.5	81.9
E_N	μ_{20}	5.3	10.4	19.5	5.2	15.4	32.0	5.5	30.2	61.4
	μ_{21}	9.9	9.8	15.7	20.6	15.1	24.2	46.3	30.4	47.3
	μ_{22}	28.3	17.4	19.3	57.2	37.0	34.8	84.8	71.4	66.7
K	μ_{20}	4.9	38.6	68.1	5.0	58.2	84.6	4.8	81.3	95.3
	μ_{21}	50.2	8.7	29.4	81.7	13.0	44.4	97.6	26.6	71.2
	μ_{22}	93.7	49.4	17.5	99.5	87.6	37.2	99.9	99.8	74.6
S	μ_{20}	4.9	49.5	78.4	5.3	63.4	87.9	5.0	85.0	96.9
	μ_{21}	60.7	9.2	38.4	83.3	14.7	50.2	97.4	29.6	76.5
	μ_{22}	96.8	57.7	18.9	99.5	87.5	39.5	99.9	99.7	78.9
F	μ_{20}	43.6	50.1	58.0	40.8	52.9	66.9	40.9	63.6	83.1
	μ_{21}	48.4	49.1	56.9	57.7	53.5	62.6	77.0	66.1	76.8
	μ_{22}	68.1	56.6	60.1	85.5	71.6	70.3	97.5	92.6	88.3

REFERENCES

Anderson, T. W. (1993). "Nonnormal Multivariate Distributions: Inference Based on Elliptically Contoured Distributions," in *Multivariate Analysis: Future Directions*, ed. C. R. Rao, Amsterdam: North-Holland, pp. 1-24.

Ahrens, J. H. and Dieter, U. (1973). "Computer Methods for Sampling from Gamma, Beta, Poisson and Binomial Distributions," *Computing*, **12**, 223-246.

Arbuthnott, J. (1710). "An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes," *Philosophical Transactions*, **27**, 186-190.

Bell, C. B. and Haller, H. S. (1969). "Bivariate Symmetry Tests: Parametric and Nonparametric," *The Annals of Mathematical Statistics*, **40**, 259-269.

Box, G. E. P. and Muller, M. E. (1958). "A Note on the Generation of Random Normal Deviates," *The Annals of Mathematical Statistics*, **29**, 610-611.

Bradley, E. L. and Blackwood, L. G. (1989). "Comparing Paired Data: A Simultaneous Test for Means and Variances," *The American Statistician*, **43**, 234-235.

Bradley, J. V. (1968). *Distribution-Free Statistical Tests*, Englewood Cliffs, New Jersey: Prentice-Hall.

Cook, R. D. and Johnson, M. E. (1981). "A Family of Distributions for Modelling Non-elliptically Symmetric Multivariate Data," *Journal of the Royal Statistical Society*, Ser. B, **43**, 210-218.

Dwass, M. (1957). "Modified Randomization Tests for Nonparametric Hypotheses," *The Annals of Mathematical Statistics*, **28**, 181-187.

Ernst, M. D. (1997). "A Multivariate Generalized Laplace Distribution," *Computational Statistics*, **12**, in press.

Ernst, M. D., Guerra, R., and Schucany, W. R. (1996). "Scatterplots for Unordered Pairs," *The American Statistician*, **50**, 260-265.

Evans, M., Hastings, N. A. J., and Peacock, B. (1993). *Statistical Distributions* (2^{nd} ed.), New York: John Wiley & Sons, Inc.

Fisher, R. A. (1935). *The Design of Experiments*, Edinburgh: Oliver and Boyd.

Fisher, R. A. (1936). "'The Coefficient of Racial Likeness' and the Future of Craniometry," *Journal of the Royal Anthropological Institute of Great Britain and Ireland*, **66**, 57-63.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*, Malabar, FL: Krieger.

Hollander, M. (1971). "A Nonparametric Test for Bivariate Symmetry," *Biometrika*, **58**, 203-212.

Hsu, C. T. (1940). "On Samples From a Normal Bivariate Population," *The Annals of Mathematical Statistics*, **11**, 410-426.

Jöckel, K.-H. (1986). "Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests," *The Annals of Statistics*, **14**, 336-347.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*, New York: John Wiley.

Johnson, M. E. and Ramberg, J. S. (1977). "Elliptically Symmetric Distributions: Characterizations and Random Variate Generation," *American Statistical Association Proceedings of the Statistical Computing Section*, pp. 262-265.

Kepner, J. L. and Randles, R. H. (1982). "Detecting Unequal Marginal Scales in a Bivariate Population," *Journal of the American Statistical Association*, **77**, 475-482.

Kepner, J. L. and Randles, R. H. (1984). "Comparison of Tests for Bivariate Symmetry Versus Location and/or Scale Alternatives," *Communications in Statistics – Theory and Methods*, **13**, 915-930.

Kotz, S., Johnson, N. L., and Read, C. B. (eds.) (1985). *Encyclopedia of Statistical Science* (Vol. 6), New York: John Wiley.

Koziol, J. A. (1979). "A Test for Bivariate Symmetry Based on the Empirical Distribution Function," *Communications in Statistics – Theory and Methods*, **8**, 207-221.

Lee, A. J. (1990). *U-Statistics: Theory and Practice*, New York: Marcel Dekker.

Maritz, J. S. (1981). *Distribution-Free Statistical Methods*, London: Chapman and Hall.

Mason, R. L., Gunst, R. F., and Hess, J. L. (1989). *Statistical Design and Analysis of Experiments*, New York: John Wiley.

Morgan, W. A. (1939). "A Test for the Significance of the Difference Between the Two Variances in a Sample from a Normal Bivariate Population," *Biometrika*, **31**, 13-19.

Odeh, R. E. and Evans, J. O. (1974). "The Percentage Points of the Normal Distribution," *Applied Statistics*, **23**, 96-97.

Pitman, E. J. G. (1939). "A Note on Normal Correlation," *Biometrika*, **31**, 9-12.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C* (2nd ed.), Cambridge: Cambridge University Press.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to The Theory of Nonparametric Statistics*, New York: John Wiley.

Sen, P. K. (1967). "Nonparametric Tests for Multivariate Interchangeability. Part 1: Problems of Location and Scale in Bivariate Distributions," *Sankhyā*, Ser. A, **29**, 351-372.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics*, New York: Chapman & Hall.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Veblen, T. (1961). *The Place of Science in Modern Civilisation and Other Essays*, New York: Russell & Russell.