2006

EMPIRICAL PROCESSES FOR ESTIMATED PROJECTIONS OF MULTIVARIATE NORMAL VECTORS WITH APPLICATIONS TO E.D.F. AND CORRELATION TYPE GOODNESS OF FIT TESTS

Christopher Paul Saunders
University of Kentucky, saunders@ms.uky.edu

ABSTRACT OF DISSERTATION

Christopher Paul Saunders

The Graduate School

University of Kentucky

2006

EMPIRICAL PROCESSES FOR ESTIMATED PROJECTIONS OF MULTIVARIATE NORMAL VECTORS WITH APPLICATIONS TO E.D.F. AND CORRELATION TYPE GOODNESS OF FIT TESTS

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By

Christopher Paul Saunders

Lexington, KY

Director: Dr. Constance L. Wood, Associate Professor of Statistics

Lexington, KY

2006

ABSTRACT OF DISSERTATION

EMPIRICAL PROCESSES FOR ESTIMATED PROJECTIONS
OF MULTIVARIATE NORMAL VECTORS WITH APPLICATIONS
TO E.D.F. AND CORRELATION TYPE GOODNESS OF FIT TESTS

Goodness-of-fit and correlation tests are considered for dependent univariate data that arises when multivariate data is projected to the real line with a data-suggested linear transformation. Specifically, tests for multivariate normality are investigated. Let $\{Y_i\}$ be a sequence of independent k-variate normal random vectors, and let d_0 be a fixed linear transform from R^k to R. For a sequence of linear transforms $\left\{\hat{d}\left(Y_1,\ldots,Y_n\right)\right\}$ converging almost surely to d_0, the weak convergence of the empirical process of the standardized projections from $\hat{d}$ to a tight Gaussian process is established. This tight Gaussian process is identical to that which arises in the univariate case where the mean and standard deviation are estimated by the sample mean and sample standard deviation (Wood, 1975). The tight Gaussian process determines the limiting null distribution of E.D.F. goodness-of-fit statistics applied to the process of the projections.

A class of tests for multivariate normality, which are based on the Shapiro-Wilk statistic and the related correlation statistics applied to the dependent univariate data that arises with a data-suggested linear transformation, is also considered. The asymptotic properties for these statistics are established.

In both cases, the statistics based on random linear transformations are shown to be asymptotically equivalent to the statistics using the fixed linear transformation. The statistics based on the fixed linear transformation have same critical points as the corresponding tests of univariate normality; this allows an easy implementation of these tests for multivariate normality.

Of particular interest are two classes of transforms that have been previously considered for testing multivariate normality and are special cases of the projections considered here. The first transformation, originally considered by Wood (1981), is based on a symmetric decomposition of the inverse sample covariance matrix. The asymptotic properties of these transformed empirical processes were fully developed using classical results. The second class of transforms is the principal components that arise in principal component analysis. Peterson and Stromberg (1998) suggested using these transforms with the univariate Shapiro-Wilk statistic.

Using these suggested projections, the limiting distribution of the E.D.F. goodness-of-fit and correlation statistics are developed.

KEYWORDS: Shapiro-Wilk Statistic, Multivariate Normality, Empirical Processes, Goodness-of-Fit, Asymptotic Properties

Christopher Paul Saunders

August 31, 2006

EMPIRICAL PROCESSES FOR ESTIMATED PROJECTIONS
OF MULTIVARIATE NORMAL VECTORS WITH APPLICATIONS
TO E.D.F. AND CORRELATION TYPE GOODNESS OF FIT TESTS

By

Christopher Paul Saunders

Constance L. Wood
(Director of Dissertation)

Arnold Stromberg
(Director of Graduate Studies)

August 31, 2006

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name	Date

DISSERTATION

Christopher Paul Saunders

The Graduate School

University of Kentucky

2006

EMPIRICAL PROCESSES FOR ESTIMATED PROJECTIONS
OF MULTIVARIATE NORMAL VECTORS WITH APPLICATIONS
TO E.D.F. AND CORRELATION TYPE GOODNESS OF FIT TESTS

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Christopher P. Saunders

Lexington, KY

Director: Dr. Constance L. Wood, Associate Professor of Statistics

Lexington, KY

2006

To Erika and Jacob

ACKNOWLEDGEMENTS

I am deeply indebted and very grateful for the kind help and support of the chairperson of my committee, Dr. Constance L. Wood. Her advice has always proved invaluable. Without her guidance, understanding, and patience this dissertation would not have been possible.

I am also extremely grateful to the other members of my committee: Dr. Arne Bathke, Dr. Larry Harris, Dr. Xin Ma, Dr. Peter Perry, Dr. Arnold Stromberg, and Dr. Mai Zhou. Their guidance and encouragement throughout my entire education at the University of Kentucky has been wonderful. A special thank you goes to Dr. Bathke for being a sounding board and always taking the time to listen to my ideas, no matter how far-fetched they appeared to be. I would also like to thank Dr. Harris for introducing me to functional analysis.

The rewarding experience as a statistician for the microarray research group, headed by Dr. Stromberg, has honed my skills as an applied statistician. I am grateful to Dr. Stromberg for making this experience possible. Both my time spent working for the microarray research group and the support of the Statistics Department of the University of Kentucky enabled me to focus on my studies.

I would like to thank my family and friends for their unfailing encouragement throughout this long, often difficult, process. I am grateful to my parents for their continued love, support, and encouragement. I would especially like to thank my wife for her unfailing faith and love, without her, I would not have completed this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FILES

Chapter I. Introduction and Overview

1.1. Introduction. Classical multivariate analysis relies on the assumption that the data is randomly selected from a population with a multivariate normal distribution. More specifically, a random vector $Y \in R^k$ is said to have a multivariate normal distribution, if and only if, for every $u \in R^k$, $u'Y$ is a univariate normal random variable. A natural and informative method for testing this assumption is to apply a linear transformation of the data from $R^k \mapsto R$, and then test the projections for univariate normality. The tests proposed in Wood (1981), Srivastava and Hui (1987), Mudholkar et al. (1995), Peterson and Stromberg (1998), and Royston (1983) are all examples of this class of tests. As Wood (1981) and Royston (1983) noted, if the linear transformation is fixed from the definition of multivariate normality, this is a standard test for univariate normality.

However, if the linear transformation is a function of the data, then the projections are not independent identically distributed normal random variables. Moreover, as noted by Gnanadeskan (1997), the asymptotic null distribution of the test statistics is unknown in most cases, with the test proposed in Wood (1981) being an exception. We derive asymptotic approximations to the sampling distribution of two broad classes of univariate goodness-of-fit test statistics applied to the projections from a data-suggested linear transformation of multivariate normal random vectors. Specifically, we consider Empirical Distribution Function (E.D.F.) and correlation type statistics for testing multivariate normality. Using the empirical process theory presented in van der Vaart and Wellner (1996), it is shown that the empirical process of the projections converges to a known tight Gaussian process with a specific covariance structure. This Gaussian process will determine the limiting behavior of statistics that are continuous functionals of the empirical process of the projections, such as the Cramer-von Mises, Kolmogorov-Smirnov, and other continuous E.D.F. type statistics. For the Shapiro-Wilk type correlation statistics applied to the projections, a different approach based on the rate theorems presented in Pollard (1984) is taken to derive the limiting behavior of these statistics.

1.2. <u>Previous Results.</u> During the last three decades, over fifty tests for multivariate normality have been proposed, however most of the procedures have unknown consistency properties and the asymptotic null distribution is rarely derived (Bogdan 1999, Koziol 1983). There have been two comprehensive review articles of tests for multivariate normality published in recent years; Henze (2002) and Mecklin and Mundfrom (2004). Henze (2002) surveys affine invariant tests and summarizes most of the current asymptotic theory for these tests. Mecklin and Mundfrom (2004) provide a more comprehensive survey in preparation for their later simulation study. Three major classes of tests for multivariate normality are considered in these papers.

- Mardia's Skewness and/or Kurtosis tests, Mardia (1970, 1974)
- Henze and Zirkler's Empirical Characteristic Function tests, Henze and Zirkler (1990)
- Univariate methods such as Royston's Shapiro-Wilk tests, Royston (1983)

Henze (2002) argues for use of affine invariant tests of multivariate normality with known consistency properties such as the Henze and Zirkler test. Mecklin and Mundfrom (2002, 2004) found the Henze and Zirkler test and the Royston Shapiro Wilk test to be powerful omnibus tests for multivariate normality, although they noted some concerns about the theoretical aspects of Royston's test.

Let Y_i, $i = 1,\ldots,n,$ be independent and identically distributed (i.i.d.) random vectors in R^k, $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ be the sample mean vector, and $S = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ be the sample covariance matrix. Then the Mardia skewness test statistic is

$$b_{1,k} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[(Y_i - \bar{Y})' S^{-1}(Y_j - \bar{Y})\right]^3$$

and the Mardia kurtosis test statistic is

$$b_{2,k} = n^{-1}\sum_{j=1}^{n}\left[(Y_i - \bar{Y})' S^{-1}(Y_i - \bar{Y})\right]^2.$$

The tests for multivariate normality based on these two statistics are only consistent against distributions that differ from the multivariate normal distribution in the third and fourth moments, respectively. A statistic, $T_n(Y_1,\ldots,Y_n)$, is said to affine invariant if, for every $b \in R^k$ and $A \in R^{k\times k}$, A nonsingular, $T_n(Y_1,\ldots,Y_n) = T_n(AY_1 + b,\ldots,AY_n + b)$.

The two Mardia test statistics are affine invariant, and considered in the Henze (2002) review paper, where the asymptotic properties of the statistics are summarized under the assumptions of normality and non-normality of Y_i 's.

The Henze and Zirkler test statistic measures the weighted L^2 distance between the empirical characteristic function of $\left\{S^{-1/2}\left(Y_i - \bar{Y}\right)\right\}_{i=1}^{n}$ and the characteristic function of a standard multivariate normal distribution. In the univariate setting, this statistic has been found to have power properties similar to the popular Shapiro-Wilk statistic. As mentioned in Henze (2002), these tests are consistent tests for multivariate normality with a known asymptotic null distribution. Due to the difficulty in interpreting these types of statistics, Csorgo (1989) recommends using these statistics in combination with other “less powerful but more transparent procedures”.

The third class of statistics reduces the multivariate problem to a univariate problem. The Shapiro-Wilk (1965) statistic has consistently been shown to be a powerful test for univariate normality against a wide range of alternatives. See Gan and Koeher (1990) and Seier (2002) for two recent simulations. Royston’s (1983) marginal method first tests each of the k variates for univariate normality with a Shapiro-Wilk statistic, then combines the k dependent tests into one omnibus test statistic for multivariate normality. Royston transforms the k -Shapiro-Wilk statistics into what he claims is an approximately Chi-squared random variable, with $m\left(m \le k\right)$ degrees of freedom. The degrees of freedom are estimated by taking into account possible correlation structures between the original k test statistics. This test has been found to behave well when the sample size is small and the variates are relatively uncorrelated (Mecklin and Mundfrom, 2002, 2005).

Srivastava and Hui (1987) along with Romeu and Ozturk (1993) noted that the correlation structure between the variates affects Royston’s test, even after correcting for the dependence between the k univariate test statistics. To account for the covariance dependence of Royston’s test, Srivastava and Hui (1987) and Peterson and Stromberg (1998) suggested using the k -eigenvectors of the sample covariance matrix, also known as the sample principle components, to project each of the original observations onto the real line. The k -univariate samples, one for each eigenvector, are then tested in turn for univariate normality with a Shapiro-Wilk or a related correlation test statistic. Each of the k test statistics will be asymptotically independent when the original vectors are from a multivariate normal distribution, which implies that they can easily be combined into an omnibus test statistic for multivariate normality with an asymptotic Type I error rate of alpha. However, the estimation of the principle components introduces dependence between the projections, which violates the assumptions under which the null distribution of the correlation statistics has been characterized. Peterson and Stromberg (1998) investigated this issue with a simulation study. They found that the null distribution of the Shapiro-Wilk statistic was not unduly affected by estimating the principle components.

Wood (1981) suggested a similar approach to Peterson and Stromberg (1998), using a symmetric decomposition of the inverse of the sample covariance matrix. She transformed a sample of random vectors into k approximately independent univariate samples, which can then be tested with an E.D.F. test such as a Cramer-von Mises or Kolmogorov -Smirnov type statistic. The k -linear transformations used are the rows of the symmetric decomposition of the inverse of the sample covariance matrix. Wood showed that the empirical process of the standardized projections converges weakly to the tight Gaussian process studied by Durbin (1973) and Wood (1975). This process determines the limiting distribution of the E.D.F. statistics such as the Cramer-von Mises or Kolmogorov-Smirnov type statistics.

The Shapiro-Wilk statistic is a powerful test for univariate normality that is commonly used to test the marginals or projections of multivariate vectors for normality (Seber, 1984, Gnanadeskan, 1997). Therefore, we provide an overview of univariate Shapiro-Wilk type statistics and the relevant asymptotic theory. Let $y_1,\ldots,y_n$ be independent and identically distributed (i.i.d.) univariate random variables and $y_{i:n}$ be the i^{th} order statistics from $\{y_i\}_{i=1}^{n}$. Let $z_1,\ldots,z_n$ be i.i.d. standard normal random variables with expectation of 0 and variance of 1. Let $z_{i:n}$ be the i^{th} order statistics from $\{z_i\}_{i=1}^{n}$. For two vectors U and V in R^n, let $\bar{U} = n^{-1}\sum_{i=1}^{n} U_i$ and $\bar{V} = n^{-1}\sum_{i=1}^{n} V_i$, then we define

$$(1.1) \quad r(U,V) = \sum_{i=1}^{n}\left(U_i - \bar{U}\right)\left(V_i - \bar{V}\right) \Bigg/ \sqrt{\sum_{i=1}^{n}\left(U_i - \bar{U}\right)^2 \sum_{i=1}^{n}\left(V_i - \bar{V}\right)^2}$$

to be the sample correlation between U and V. Let $m = \{E(z_{i:n})\}_{i=1}^{n}$ and V_0 be the covariance matrix of $z_{i:n}$, then the Shapiro-Wilk statistic is the square of sample correlation between the vector of order statistics and the vector, $V_0^{-1}m$.

Due to the difficulty in calculating m and V_0 for large samples, the test is somewhat impractical. To remedy this, various authors have suggested simplifications such as the Shapiro-Francia (1972) statistic, which replaces $V_0^{-1}m$ with m, or W_n, the de Wet and Venter statistic, which uses $\xi = \left(\Phi^{-1}(i/n+1)\right)_{n\times 1}$ in place of $V_0^{-1}m$, where $\Phi(t)$ is the cumulative distribution function of a standard normal random variable. These tests have slightly different power properties depending on the choice of plotting positions as illustrated in Looney and Gulledge (1985) or Brown and Hettmansperger (1996).

The de Wet and Venter statistic was the first of the Shapiro-Wilk type statistics for which the asymptotic distribution was derived. In de Wet and Venter (1972, 1973) it was shown that

$$2n\left(1-W_n^{1/2}\right)-a_n \rightsquigarrow \sum_{j=3}^{\infty}(X_i-1)/j, \text{ as } n\to\infty,$$

where $X_1,\dots,X_n$ are i.i.d. random variables with a Chi-squared distribution, one degree of freedom, and $a_n = (n+1)^{-1}\left\{\sum_{i=1}^{n}\left(\frac{i}{n+1}\right)\left(1-\frac{i}{n+1}\right)\left(\phi\left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right)^{-2}\right)\right\}-\frac{3}{2}$. Here $x_n \rightsquigarrow x$ denotes the weak convergence of x_n to x.

Extending the above result; Leslie, Stephens, and Fotopoulos (1986) showed that the Shapiro-Wilk statistic has the same limiting distribution as W_n. The following year Verrill and Johnson (1987) proved that for any vector of plotting coefficients, Ψ, of length n such that $\sum_{i=1}^{n}(\Psi_i-\xi_i)^2 = o\left((\log\log n)^{-1}\right)$, as $n\to\infty$, the square of the correlation statistic based on Ψ still has the same limiting distribution as W_n. All of these results assume that the observations are i.i.d. normal random variables.

In the case when the observations are not i.i.d., Wood (1984) derived the asymptotic properties of the standardized empirical process applied to the residuals from a fitted ridge regression model. Sen et al. (2003) proved a similar result for the de Wet-Venter statistic based on the residuals from a fitted regression model. Sen et al. (2003) proved that the de Wet-Venter statistic based on the fitted residuals is asymptotically equivalent to the de Wet and Venter statistic based the residuals when the regression model is known. This is the only case where the asymptotic null distribution of Shapiro-Wilk type statistic applied to approximately normal random variables has been derived.

1.3. __E.D.F. Approach.__ Our goal is to use E.D.F. and correlation type goodness-of-fit statistics for multivariate normal vectors. Our approach is to reduce the multivariate goodness-of-fit problem to a univariate goodness-of-fit test by using a procedure based on the definition of multivariate normality. Specifically, let $Y_1,\ldots,Y_n$ be k-dimensional multivariate normal vectors with mean, μ, and positive definite covariance matrix, Σ. Then $\{d_0(Y_i-\mu);\ i=1,\ldots,n\}$ is a univariate set of i.i.d. normal random variables, when d_0 is a fixed row vector in R^k. However, if d_0 is replaced by a random vector $\hat{d}=\hat{d}\left(Y_1,\ldots,Y_n\right)$ and μ is estimated with $\bar{Y}=\sum_{i=1}^{n}Y_i$, the corresponding univariate set of observations, $\{\hat{d}_n(Y_i-\bar{Y}),i=1,\ldots,n\}$, are only approximately i.i.d. normal random variables. Let

$$\hat{F}_n\left(t\right)=n^{-1}\sum_{i=1}^{n}I\left(\hat{d}(Y_i-\bar{Y})\le t\right),\text{ for } -\infty<t<\infty,$$

where, for a set A,

$$I\left(x\in A\right)=\begin{cases}1, & x\in A\\ 0, & x\notin A.\end{cases}$$

Let $\Phi(t)$ and $\phi(t)$, $-\infty<t<\infty$, be the cumulative distribution and probability density functions of a standard normal distribution, respectively. In Chapter III, we consider the univariate empirical process corresponding to $\{\hat{d}(Y_i-\bar{Y}),i=1,\ldots,n\}$,

(1.2) $\quad G_n(t)=n^{1/2}(\hat{F}\left(t\right)-\Phi(t)),\ -\infty<t<\infty,$

where $\hat{d}$ is related to the sample covariance matrix, $S=S\left(Y_1,\ldots,Y_n\right)$. In general, we will suppress the dependence on n. Specifically, in Chapter III, we consider $\hat{d}$, a sequence of row vectors in R^k, with the properties

(1.3) $\quad \hat{d}S\hat{d}'=1,$

and

(1.4) $\hat{d} \xrightarrow{p} d_0$, as $n \to \infty$, for $d_0 \in R^k$,

where

(1.5) $d_0 \Sigma d_0{}' = 1$.

Under these assumptions on $\hat{d}$, we show that

(1.6) $G_n \rightsquigarrow G$, as $n \to \infty$,

where G is a tight Gaussian process with covariance function

$$\Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - ts\phi(s)\phi(t)/2, \ -\infty < s, t < \infty.$$

It is important to note that $G(t)$ is the same process that arises in the univariate case when estimating the parameters in the normal distribution (Wood, 1975). By the Continuous Mapping Theorem, the Gaussian process, G, will determine the limiting behavior of continuous E.D.F. goodness-of-fit statistics applied to G_n, such as the Cramer-von Mises type statistics.

The theory presented in Chapter III is applicable to some interesting projections used for E.D.F. goodness of fit testing. The first was proposed by Wood in 1981. Let b_j be the j^{th} row of B, where $BB = \Sigma^{-1}$. Let $\hat{b}_j$ be the j^{th} row of $\hat{B}$, where $\hat{B}\hat{B} = S^{-1}$. Then, $\hat{b}_j$ has the properties (1.3) and (1.4), take $\hat{d} = \hat{b}_j$. Then the results presented in Chapter III demonstrate the weak convergence of G_n to G, as $n \to \infty$.

Since $b_j \Sigma b_k^t = 0$, if $k \neq j$, each of the k processes will be asymptotically independent of the $k-1$ other processes. The asymptotic properties of G_n, including the tightness of the limiting process, based on this set of data suggested linear transformations has been previously derived using the random change of time technique (Wood, 1981).

Peterson and Stromberg (1998) suggest k transformations for investigating multivariate normality in R^k. Each of the transformations is an eigenvector of the sample covariance matrix, S. Let $\{\hat{e}_j, \hat{\lambda}_j\}_{j=1}^k$ be the eigenvector/eigenvalue pairs of S and $\{e_j, \lambda_j\}_{j=1}^k$ be the eigenvector/eigenvalue pairs of Σ. Since, for fixed j, $\{e_j^t(Y_i - \mu), i = 1, \ldots, n\}$ are i.i.d. normal random variables with variance λ_j, Peterson and Stromberg (1998) consider the transformed data $\left\{\hat{e}_j'(Y_i - \overline{Y}), i = 1,\ldots,n\right\}$ for use with the univariate Shapiro-Wilk statistic. To apply the theorems in Chapter III for E.D.F. tests, let $\hat{d} = \hat{\lambda}_j^{-\frac{1}{2}}\hat{e}_j'$. The projections, $\{\hat{\lambda}_j^{-\frac{1}{2}}\hat{e}_j'\}_{j=1}^{k}$, have the properties (1.3) and (1.4). As above, the results in Chapter III demonstrate the weak convergence of G_n to G, $j = 1,\ldots,k$, as $n \to \infty$. Furthermore, since the $e_j'\Sigma e_k = 0$, if $k \neq j$, each process determined by an eigenvector, will be asymptotically independent of the $k-1$ other processes.

1.4. <u>Correlation Statistic Approach.</u> In Chapter IV, we also apply the de Wet and Venter statistic to the projections from a data suggested linear transformation. The correlation form of the modified de Wet and Venter statistic is the square of the sample correlation between the vector ξ and the vector of order statistics from $\left\{\hat{y}_i\right\}_{i=1}^{n}$. We will denote the modified de Wet and Venter statistic based on the projections $\{\hat{y}_i = \hat{d}(Y_i - \overline{Y}), i = 1,\ldots,n\}$ as $\hat{W}_n$. Similarly, we denote the de Wet and Venter statistic based on the projections from the fixed linear transformation $\{y_i = d_0(Y_i - \mu), i = 1,\ldots,n\}$ as W_n. In the case of correlation type statistics, we make the additional assumption that

(1.7) $$n^{1/2}\left\|\hat{d} - d_0\right\| = O_p\left(1\right), \text{ as } n \to \infty.$$

Our approach demonstrates for the de Wet and Venter statistic based the projections from $\hat{d}$ satisfying (1.7), that

(1.8) $$n\left(\hat{W}_n - W_n\right) = o_p\left(1\right), \text{ as } n \to \infty,$$

and

(1.9) $$n\left(1 - \hat{W}_n\right) - a_n \rightsquigarrow \sum_{j=3}^{\infty}\left(x_j - 1\right)\Big/ j, \text{ as } n \to \infty.$$

We prove similar results for all of the correlation type statistics considered in Verrill and Johnson (1987). It is shown in Chapter V that the two sets of projections mentioned above from Wood (1981) and Peterson and Stromberg (1998) satisfy the condition (1.7).

1.5. <u>Summary of Results.</u> In this dissertation, we derive the limiting distribution for E.D.F. and correlation type goodness-of-fit statistics applied to the projections from a data suggested linear transformation of multivariate normal random vectors. In particular, we derive the limiting distribution of these types of statistics for the projections from the sample principle components, as suggested by Peterson and Stromberg (1998) and Srivastava and Hui (1987). We also re-derive the result of Wood (1981) for E.D.F. goodness-of-fit statistics based on projections from the symmetric decomposition of the inverse of sample covariance matrix and derive the asymptotic distribution of the correlation goodness-of-fit statistic applied to Wood's projections.

In Chapter II, we provide a summary of the main results and definitions that we have used from Pollard (1984), van der Vaart (1998) and van der Vaart and Wellner (1996). One new result presented in this chapter is Lemma 2.12, this lemma is a generalization of the Lemma from Bahadur (1966) which yields a uniform result for $\left[F_n(t)-\Phi(t)\right]-\left[F_n(s)-\Phi(s)\right]$, $-\infty < t,\, s < \infty$, where F_n is the empirical C.D.F. of n i.i.d. standard normal random variables. Generally speaking, the results from van der Vaart and Wellner (1996) and van der Vaart (1998) are used in the derivation of the limiting process of G_n, while the results from Pollard (1984) are used in deriving the asymptotic distribution of the correlation statistic $\hat{W}_n$.

In Chapter III, we derive the convergence properties of empirical processes based on the projections from a data suggested linear transformation under the assumptions that the original observations have a multivariate normal distribution and that the linear transformation satisfies (1.2), (1.3), and (1.4). The proof is a straightforward application of the theory presented in Chapter II which generalizes the result from Wood (1981). The main result in this chapter is Theorem 3.11, where it is shown that the process G_n converges to the tight Gaussian process, G, which has a known covariance structure. By the continuous mapping theorem, G determines the limiting distribution of continuous functionals of G_n, such as most E.D.F. goodness-of-fit statistics.

In Chapter IV, we derive the limiting distribution of correlation statistics based on projections of multivariate normal random vectors from a data suggested linear transformation under the assumption of normality of the original observations and that the data suggested linear transformation satisfies condition (1.7). This chapter starts, by reducing the problem of showing (1.8) to bounding the difference between the normalized order statistics from $\{\hat{z}_i = \hat{d}(Y_i-\mu); i=1,\ldots,n;\, \hat{d}\Sigma\hat{d}'=1\}$ and the order statistics from

$\{z_i = d_0(Y_i - \mu); i = 1,\ldots,n;\ d_0\Sigma d_0' = 1\}$. In Lemma 4.14, we give a sequence of bounds on this difference, which holds uniformly in i and tends to zero at a specific rate. Let $\{i_n\}_{n=1}^{\infty}$ be a sequence of positive integers such that $n^{1/4} i_n^{-1} \log n = O(1)$ and $i_n n^{-1} = o(1)$, as $n \to \infty$. In Lemma 4.16, we give a sequence of bounds on the difference between the two sets of order statistics for $i_n \le i \le n - i_n$, where the actual rate at which these bounds tend to zero depends on the choice of the sequence $\{i_n\}_{n=1}^{\infty}$. Theorem 4.19, the main result of this chapter, is proven by combining these two rate theorems to show (1.8) which implies (1.9). Theorem 4.21 extends the results for the de Wet and Venter statistic to the other correlation goodness-of-fit statistics considered in Verril and Johnson (1981) and Corollary 4.18 specifically covers the Shapiro-Wilk statistic.

In Chapter V, we apply the theory from Chapter IV to derive the limiting distribution of correlation statistics from specific linear transformations. In Corollary 5.2, we apply our theory to $\left\{\hat{e}_j'(Y_i - \bar{Y}), i = 1,\ldots,n\right\}$, $j = 1,\ldots,k$, and derive the limiting distribution of the Srivastava and Hui (1987) and the Peterson and Stromberg (1998) statistics. In Corollary 5.4, we consider Wood's transformation and derive the limiting distribution of correlation type goodness of fit statistics applied to $\left\{\hat{b}_j(Y_i - \bar{Y}), i = 1,\ldots,n\right\}$. The proofs in this chapter make use the asymptotic results for sample principle components from a multivariate normal distribution in Anderson (1963).

In Chapter VI, we provide a simulation study of the power properties the proposed omnibus tests for normality. First, some problems associated with calculating the p-values for univariate correlation statistics is discussed. We review two methods presented in Peterson and Stromberg (1998) for combining k p-values for an omnibus test for multivariate normality.

In Chapter VII, we summarize our current research and discuss future directions.

Chapter II. Empirical Process Theory

2.1. <u>Introduction and Basic Definitions.</u> For completeness we will begin with a summary of the main results that we require from Billingsley (1968), Pollard (1984), van der Vaart (1998), and van der Vaart and Wellner (1996). In this section, we state some definitions for empirical processes from van der Vaart (1998). In Section 2.2, we review empirical processes indexed by classes of functions and some theorems concerning P-Donsker Classes of functions. In Section 2.3, we review VC-Classes of functions and their relationship to P-Donsker classes. Euclidean and Permissible Classes of functions are reviewed for use with Pollard's Rate Theorem in Section 2.4. Corollary 2.10 defines a class Euclidean class of functions that we will use in Sections 2.5 and 4.2. In Section 2.5, we state Pollard's Rate Theorem and use it to prove a generalization of the Lemma from Bahadur (1966) for the standard normal distribution.

Here we consider $X_1, X_2, \ldots$ an infinite sequence of random vectors on a probability space $(\Omega, \mathbb{B}, Q)$. Throughout, for i.i.d. random vectors, we work with the induced probability measure $P = Q\left(X_1^{-1}\right)$. A sequence of random elements X_n, taking values in the metric space $\mathcal{X}$, is said to converge in distribution (or weakly) to a random element X if $Ef\left(X_n\right) \to Ef\left(X\right),$ as $n \to \infty,$ for every bounded, continuous function $f : \mathcal{X} \mapsto R$. We will denote the weak convergence of X_n to X, as $n \to \infty$, by $X_n \rightsquigarrow X$. A Borel-measurable random element X into a metric space is tight if for every $\varepsilon > 0$ there exists a compact set K such that $P\left(X \notin K\right) < \varepsilon$.

Let $X_1, \ldots, X_n$ be i.i.d. random elements with the induced probability distribution P on a measurable space $\{\mathcal{X}, \mathcal{A}\}$. Then the empirical measure P_n is defined by $P_n(\bullet) = n^{-1}\sum_{i=1}^{n}\delta_{X_i}(\bullet)$, where $\delta_x(\bullet)$ is the probability distribution that is degenerate at x. For $\mathcal{F}$ a class of measurable functions $f : \mathcal{X} \mapsto R$, we define the operators

(2.1) $Pf = \int f dP$,

(2.2) $P_n f = \int f dP_n = n^{-1}\sum f(X_i)$,

and

(2.3) $\mathbb{G}_n f = n^{1/2}(P_n - P)f$, for every $f \in \Im$.

Here $P_n f$ and $\mathbb{G}_n f$ are random and technically are $P_n f(\omega) = n^{-1}\sum_{i=1}^{n} f(X_i(\omega))$ and $\mathbb{G}_n f(\omega) = n^{1/2}\left[P_n f(\omega) - Pf\right]$. Here we suppress the dependence on ω. For a given arbitrary set T, let $\ell^\infty(T)$ be the collection of all bounded functions $z : T \mapsto R$. We seek conditions under which

(2.4) $\mathbb{G}_n \rightsquigarrow \mathbb{G}_p$, in $\ell^\infty(\Im)$, as $n \to \infty$,

where $\mathbb{G}_p$ is a tight Gaussian process in the space $\ell^\infty(\Im)$, with zero mean and covariance function

(2.5) $Pf_1 f_2 - Pf_1 Pf_2$, for f_1, $f_2 \in \Im$.

A set of functions is said to be totally bounded for a given semi-metric if, for every $\varepsilon > 0$, the set of functions can be covered with finitely many balls of radius ε. It is important to note that every totally bounded space is separable.

The covering number $N(\varepsilon, \mathcal{F}, \|\bullet\|)$ of a class of functions $\mathcal{F}$, with respect to the norm $\|\bullet\|$, is the minimal number of balls $\{g : \|g - f\| < \varepsilon\}$ of radius ε needed to cover $\mathcal{F}$. An envelope function of a class of real function $\mathcal{F}$ on a measurable space $(\mathcal{X}, \mathcal{A})$ is any function F on $\mathcal{X}$ such that $|f(x)| \le F(x)$ for all $x \in \mathcal{X}$ and all $f \in \mathcal{F}$.

For P be a probability distribution on a measurable space $(\mathcal{X}, \mathcal{A})$, the $L_r(P)$ norm of a real valued function f will be denoted by

$$\|f\|_{P,r} = \left(\int |f|^r \, dP\right)^{1/r}.$$

We will consistently use the notation from Serfling (1981). For two functions $u(x)$ and $v(x)$, the notation $u(x) = O(v(x))$, as $x \to L$, denotes that $|u(x)/v(x)|$ remains bounded as $x \to L$. The notation $u(x) = o(v(x))$, as $x \to L$, denotes that $\lim_{x \to L}|u(x)/v(x)| = 0$. A sequence of random variables, $\{X_n\}$, with respective distribution functions $\{F_n\}$, is said to be bounded in probability if for every $\varepsilon > 0$ there exists M_ε and N_ε such that $F_n(M_\varepsilon) - F_n(-M_\varepsilon) > 1 - \varepsilon$, for all $n > N_\varepsilon$. For two sequences of random variables $\{U_n\}$ and $\{V_n\}$, the notation $U_n = O_p(V_n)$, as $n \to \infty$, denotes that the sequence of random variables $\{U_n / V_n\}$ is bounded in probability. The notation $U_n = o_p(V_n)$, as $n \to \infty$, denotes that $U_n / V_n \xrightarrow{p} 0$, as $n \to \infty$. For the sequence of random variables $\{X_n\}$, the statement "with probability 1, $X_n = O(g(n))$, as $n \to \infty$" means that there exists a set Ω_0 such that $P(\Omega_0) = 1$ and for each $\omega \in \Omega_0$ there exists a constant $B(\omega)$ such that $|X_n(\omega)| \le B(\omega) g(n)$, for all n sufficiently large. The statement "with probability 1, $X_n = o(g(n))$, as $n \to \infty$" means that there exists a set Ω_0 such that $P(\Omega_0) = 1$ and for each $\omega \in \Omega_0$ $X_n(\omega)/g(n) = o(1)$, as $n \to \infty$.

Let $\{g_n\}$ be a sequence of real numbers. When considering a sequence of random elements, say $\{Y_n\}_{n=1}^{\infty}$, taking values in metric space, S, with a metric $\|\bullet\|_S$, the notation

$$Y_n = O_p(g_n), \text{ as } n \to \infty,$$

refers to the sequence of random variables $\{\|Y_n\|_S / g_n\}$ being bounded in probability, as $n \to \infty$. In addition, the notation

$$Y_n = o_p(g_n), \text{ as } n \to \infty,$$

denotes that the sequence of random variables $\{\|Y_n\|_S / g_n\}$ converges to zero, in probability, as $n \to \infty$.

2.2. <u>P-Donsker Classes and Tight Gaussian Processes.</u> A class of measurable functions, $\mathfrak{I}$, is said to be a P-Donsker class of functions if $\mathbb{G}_n \rightsquigarrow \mathbb{G}_p$ in $\ell^\infty(\mathfrak{I})$, as $n \to \infty$, where $\mathbb{G}_p$ is the tight Gaussian process with covariance defined in (2.5). A Gaussian process, $\mathbb{G}_p$, in $\ell^\infty(\mathfrak{I})$

is said to be tight if and only if $(\Im, \rho_p)$ is totally bounded and almost all paths $f \mapsto \mathbb{G}_p(f, \omega)$ are uniformly ρ_p-continuous for some p, where

$$\rho_p(f_1, f_2) = P(f_1 - Pf_1 - f_2 + Pf_2)^p, \text{where } f_1, f_2 \in \Im.$$

Here ρ_2 is a semi-metric on $\Im$. This is the natural semi-metric to be used with a Gaussian process and corresponds to the standard deviation metric (van der Vaart and Wellner, 1996, pg. 41).

An important result, due to van der Vaart (1998), is stated and proved here for completeness as Theorem 2.3. To prove Theorem 2.3, we make use of the following two results from Billingsley (1968), which are presented without proof. The first is the Extended Continuous Mapping Theorem. The second result is Slutsky's Lemma for the convergence of random elements, in the associated product space.

In general, a random element defined on $(\Omega, \mathscr{B}, Q)$, say Y, is a mapping taking values in a metric space $(S, \mathscr{L})$; i.e. $Y^{-1}\mathscr{L} \subset \mathscr{B}$. Let S be a metric space equipped with the metric ρ and the σ-field $\mathscr{L}$ of Borel sets. Let $\{Y_n\}$ be a sequence of random elements of S defined on the probability spaces $(\Omega, \mathscr{B}, Q_n)$. Let h be a measurable mapping from S into the metric space S' with the metric ρ' and the σ-field $\mathscr{L}'$ of Borel sets, then each probability measure P_Y on $(S, \mathscr{L})$ determines a unique probability measure $P_Y h^{-1}$, defined by $P_Y h^{-1} = P_Y\left(h^{-1}A\right)$ for $A \in \mathscr{L}'$. Let h_n and h be measurable mappings from S to $S,'$ S separable. Let $E \subset S$ be the set of y such that $h_n y_n \to hy$ fails to hold for some sequence $\{y_n\}$ approaching y. We will use the notation $P_n \rightsquigarrow P$ to denote the weak convergence of the sequence of probability measures $\{P_n\}$ to P.

Theorem 2.1. (Billingsley (1968), Theorem 5.5, pg. 33) *If $P_n \rightsquigarrow P$, as $n \to \infty$, and $P(E) = 0$, then $P_n h_n^{-1} \rightsquigarrow Ph^{-1}$, as $n \to \infty$.*

Lemma 2.2. (Billingsley (1968), Theorem 4.4, pg. 27) *If $Y_n \rightsquigarrow Y \in S$ and $X_n \rightsquigarrow c$, as $n \to \infty$, with S separable and c a constant, then $(X_n, Y_n) \rightsquigarrow (c, Y)$, as $n \to \infty$.*

The following theorem is contained in Theorem 19.26 of van der Vaart (1998).

Theorem 2.3. *Let* Θ *be a normed space and*

$$\Im_\delta = \{ f_{\theta,t} - f_{\theta_0,t} : \| \theta - \theta_0 \| \le \delta, \theta, \theta_0 \in \Theta, t \in R\}$$

be a P-Donsker class of functions, which map $\mathcal{X} \mapsto R$ *, for some* $\delta > 0$ *.*

If

$$\sup_{t\in R} \int (f_{\theta,t} - f_{\theta_0,t})^2 dP \to 0, \text{ as } \theta \to \theta_0 \text{, for } \theta, \ \theta_0 \in \Theta_\delta \text{, and } t \in R,$$

and

$$\hat{\theta} \xrightarrow{p} \theta_0, \text{ as } n \to \infty,$$

then

$$\sup_{t\in R} \left| \mathbb{G}_n (f_{\hat{\theta},t} - f_{\theta_0,t}) \right| = o_p(1), \text{ as } n \to \infty .$$

Proof. Let $\Theta_\delta = \{\theta : | \theta - \theta_0 | < \delta\}$ and consider $g : \ell^\infty(\Theta_\delta \times R) \mapsto \ell^\infty(R)$ by $g(z,\theta)\, t = z(\theta,t)$.

Note that $g(\cdot)$ is a continuous function for every point (z, θ_0), where

$$\sup_{t\in R} | z(\theta, t) - z(\theta_0, t) | \to 0, \text{ as } \theta \to \theta_0.$$

Define, for $-\infty < t < \infty$,

$$Z_n(\theta, t) = \mathbb{G}_n (f_{\theta,t} - f_{\theta_0,t})$$

and

$$Z_p(\theta, t) = \mathbb{G}_p (f_{\theta,t} - f_{\theta_0,t}) .$$

Since $\Im_\delta$ is a P-Donsker class of functions, $\mathbb{G}_n \rightsquigarrow \mathbb{G}_p$, as $n \to \infty$, in $\ell^\infty(\Im_\delta)$, where $\mathbb{G}_p$ is a tight Gaussian process defined on $\Im_\delta$. Therefore,

$$Z_n(\theta, t) = \mathbb{G}_n (f_{\theta,t} - f_{\theta_0,t}) \rightsquigarrow \mathbb{G}_p (f_{\theta,t} - f_{\theta_0,t}) = Z_p(\theta, t), \text{ as } n \to \infty .$$

In order to use Theorem 2.1 we need to show that $g(\cdot)$ is a continuous map at (Z_p, θ_0). Since $\mathbb{G}_p$ is a tight Gaussian process on $\Im_\delta$, by the characterization of a tight Gaussian process mentioned above, $\Im_\delta$ is totally bounded with respect to ρ and almost all sample paths on $\Im_\delta$ are uniformly $\rho-$continuous on $\Im_\delta$ with respect to the standard deviation metric on $\Im_\delta$, i.e.

$$\rho^2(f_{\theta_1,t_1}, f_{\theta_2,t_2}) = P(f_{\theta_1,t_1} - Pf_{\theta_1,t_1} - f_{\theta_2,t_2} + Pf_{\theta_2,t_2})^2, \text{ for } f_{\theta_2,t_2},\ f_{\theta_1,t_1} \in \Im_\delta.$$

Note that

$$\rho^2(f_{\theta_1,t_1} f_{\theta_2,t_2}) = P(f_{\theta_1,t_1} - f_{\theta_2,t_2})^2 - (P(f_{\theta_1,t_1} - f_{\theta_2,t_2}))^2$$
$$\leq P(f_{\theta_1,t_1} - f_{\theta_2,t_2})^2.$$

The uniform ρ-continuity of almost all sample paths implies that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $\rho(f_{\theta_1,t_1}, f_{\theta_2,t_2}) < \delta$, $|\mathbb{G}_p(f_{\theta_1,t_1} - f_{\theta_2,t_2})| < \varepsilon$.

To show that $g(\cdot)$ is continuous at (Z_p, θ_0), we need to show that for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\sup_{t\in R} |\mathbb{G}_p(f_{\theta,t} - f_{\theta_0,t})| < \varepsilon$, whenever $\|\theta - \theta_0\| < \delta$. By the uniform ρ-continuity of $\mathbb{G}_p$, there exists a $\delta_\Im > 0$ such that if $\sup_{t\in R} \rho(f_{\theta,t} - f_{\theta_0,t}, 0) < \delta_\Im$ then $|\mathbb{G}_p(f_{\theta,t} - f_{\theta_0,t})| < \varepsilon,$ for every $t \in R$. However, from the note above, $\sup_{t\in R} \rho^2(f_{\theta,t} - f_{\theta_0,t}, 0) \leq \sup_{t\in R} P(f_{\theta,t} - f_{\theta_0,t})^2$. Now by assumption, there exists a δ' such that if $\|\theta - \theta_0\| < \delta'$, then $\sup_{t\in R} P(f_{\theta,t} - f_{\theta_0,t})^2 < \delta_\Im^2$. Without loss of generality, assume $\delta' < \delta$. Therefore, if $\|\theta - \theta_0\| < \delta'$, then $\sup_{t\in R} \rho^2(f_{\theta,t} - f_{\theta_0,t}, 0) < \delta_\Im^2$. This implies $\rho^2(f_{\theta,t}, f_{\theta_0,t}) < \delta_\Im^2$, $\forall$ t $\in R$ and θ such that $\|\theta - \theta_0\| < \delta'$. Therefore, $|\mathbb{G}_p(f_{\theta,t} - f_{\theta_0,t})| < \varepsilon$, $\forall$ t$\in R$ and $\sup_{t\in R} |\mathbb{G}_p(f_{\theta,t} - f_{\theta_0,t})| \leq \varepsilon$. Which implies that $\sup_{t\in R} |Z_p(\theta, t)| = \sup_{t\in R} |g(Z_p, \theta)| \leq \varepsilon$ whenever $\|\theta - \theta_0\| < \delta'$. Therefore, $g(\cdot)$ is a continuous map at (Z_p, θ_0).

Next, we note that $\hat{\theta} \xrightarrow{p} \theta_0$ and $Z_n \rightsquigarrow Z_p$ and that $\Im$ is totally bounded with respect to ρ_2, which gives us that $(\theta_0,\ Z_p)$ takes values on a separable space with probability one. Now by Theorem 2.1, $(\hat{\theta},\ Z_n) \rightsquigarrow (\theta_0,\ Z_p)$ in $\Theta_\delta \times \ell^\infty(\Theta_\delta \times R)$, as $n \to \infty$. Now by applying the Continuous Mapping Theorem we get that $g(\hat{\theta},\ Z_n) \rightsquigarrow g(\theta_0,\ Z_p) = \mathbb{G}_p(f_{\theta_0,t} - f_{\theta_0,t}) = 0$, as $n \to \infty$, in $\ell^\infty(R)$. Therefore, $\sup_{t\in R} |\mathbb{G}_p(f_{\hat{\theta},t}(Y) - f_{\theta_0,t}(Y))| \xrightarrow{P} 0$. □

2.3. VC Classes and Uniform Entropy Integrals. There are two common methods used to show that a class of functions has the P-Donsker property,

(i) the bracketing integral is finite (For further information see the van der Vaart(1998) and van der Vaart and Wellner(1996))

or

(ii) the uniform entropy integral is finite and $\|F\|_{Q,2} < \infty$.

For completeness we will define uniform entropy integrals and list some results concerning them from van der Vaart and Wellner (1998). The norm that is used when defining uniform entropy is the $L_2(Q)$ norm. The uniform entropy integral is defined as

(2.6) $$J(\delta, \Im, L_2) = \int_0^{\delta} \sqrt{\log \sup_Q N(\varepsilon \|F\|_{Q,2}, \Im, L_2(Q))}\, d\varepsilon .$$

The next theorem from van der Vaart(1998) relates P-Donsker classes and uniform entropy integrals.

Theorem 2.4. (Theorem 19.14, van der Vaart, 1998, pg. 274) *Let $\Im$ be a suitably measurable class of functions with $J(1, \Im, L_2) < \infty$. If $PF^2 < \infty$, then $\Im$ is P-Donsker.*

One of the classes of functions considered form a Vapnik-Cervonenkis (VC) class of functions. A VC-class of functions has the property that there exists an upper bound on $N(\varepsilon \|F\|_{Q,2}, \Im, L_2(Q))$, which is a polynomial of $\frac{1}{\varepsilon}$. Theorem 2.5 below gives a bound on the covering number for a VC-class that is uniform for all probability measures, such that $\|F\|_{Q,r} > 0$. This bound is strong enough to guarantee that the uniform entropy integral is finite. Therefore, VC-classes of functions are P-Donsker classes of functions if they possess a finite envelope function.

A VC-class of functions is defined in terms of an associated VC-class of sets. Let us first define a VC-class of sets. A class of sets, B defined on X, will shatter another set, C of size k, contained in X, if every subset of C can be written as $b_i \cap C$, where $b_i \in B$. The VC-index of B, denoted V(B), is the size of a set for which any set of that size cannot be shattered by B. B is a VC-class of sets if $V(B)$ is finite. The subgraph of a function is the set $\{(x,t) : t < f(x)\}$. A class of functions, $\Im$, is a VC-class of functions if the class of sets generated from the subgraphs of the elements of $\Im$ is a VC-class of sets.

Theorem 2.5. (Theorem 2.6.7, van der Vaart and Wellner, 1996, pg. 141) *For a VC-class of function with a measurable envelope function F and* $r \geq 1$ *, one has for any probability measure* Q *with* $\|F\|_{Q,r} > 0$,

$$N(\varepsilon \|F\|_{Q,r}, \Im, L_r(Q)) \leq KV(\Im)(16e)^{V(\Im)} \left(\frac{1}{\varepsilon}\right)^{r(V(\Im)-1)},$$

for a universal constant K and $0 < \varepsilon < 1$.

There are three results concerning VC-Classes and P-Donsker classes, from van der Vaart and Wellner(1996), which are used in Chapter III, the lemmas are stated here for ease of reference.

Lemma 2.6. (Example 19.17, van der Vaart, 1998, pg.276) *Let* $\Im$ *be all linear combinations* $\sum_i^k \lambda_i f_i$ *of a given, finite set of functions* $f_1 \ldots f_k$ *on* $\mathcal{X}$. *Then* $\Im$ *is a VC class and hence has a finite uniform entropy integral. Furthermore, the same is true for the class of all sets* $\{f > c\}$ *if f ranges over* $\Im$ *and* c *over* R.

Lemma 2.7. (Lemma 2.6.18, van der Vaart and Wellner, 1996, pg.147) *Let* $\Im$ *be a VC-subgraph class of functions on a set* X *and* $\phi : R \mapsto R$ *be a monotonic fixed function. Then* $\phi \circ \Im$ *is VC-subgraph.*

Lemma 2.8. (Example 19.20, van der Vaart, 1998, pg. 277) *For any fixed Litschitz function* $\phi : R^2 \mapsto R$, *the class of all functions of the form* $\phi(f, g)$ *is Donsker, if* f *and* g *range over Donsker classes* $\Im$ *and* $\mathfrak{G}$ *with integrable envelope functions. For example, the class of all sums f+g, all minima* $f \wedge g$ *and the class of all maxima* $f \vee g$ *are Donsker.*

2.4. **Euclidean and Permissible Classes.** A class of functions is said to be a Euclidean class of functions, with respect to an envelope function F, if there exists A and V, not depending on the probability distribution Q, such that

$$N\left(\varepsilon\|F\|_{Q,1}, \mathcal{F}, L_1(Q)\right) \le A\varepsilon^{-V}, \quad \varepsilon > 0,$$

whenever $0 < \|F\|_{Q,1} < \infty$. In this Chapter and Chapter IV, Euclidean classes are used with Pollard's Rate theorem, which is summarized in Section 2.5. Wellner (2004) gives a complete discussion of Euclidean classes. See Wellner (2004), van der Vaart and Wellner (1996), or van der Vaart (1998) for a discussion of VC classes.

Lemma 2.9. (Wellner, 2004, Proposition 8.5) *Suppose that $\mathcal{F}$ and $\mathcal{G}$ are Euclidean classes of functions with envelopes F and G respectively, suppose that $QG^r < \infty$, for some $r \ge 1$, Then the class of functions*

(2.7) $\mathcal{F} + \mathcal{G} = \{ f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$

is Euclidean for the envelope $F + G$.

Next we will show that the class of functions on R^k

(2.8) $\mathcal{F} = \left\{ f_{c_1,c_2,a_1,a_2,s,t}(y) = I\left(c_1(y-a_1) \le t\right) - I\left(c_2(y-a_2) \le s\right); c_i', a_i \in R^k, s,t \in R\right\}$

is a Euclidean class of functions.

Corollary 2.10. *The class of functions, $\mathcal{F}$, defined in (2.8) is a Euclidean class of functions with an envelope function $F \equiv 1$.*

Proof. Define

$$\mathcal{F}_1 = \left\{ f_{c_1,a_1,t}(y) = I\left(c_1(y-a_1) \le t\right); c_1', a_1 \in R^k, t \in R\right\}$$

and

$$\mathcal{F}_2 = \left\{ f_{c_2,a_2,s}(y) = -I\left(c_2(y-a_2) \le s\right); c_2', a_2 \in R^k, s \in R\right\}.$$

By Lemma 2.6, both $\mathcal{F}_1$ and $\mathcal{F}_2$ are VC-Classes of functions and therefore Euclidean classes of functions. Note that $\mathcal{F} \subset \mathcal{F}_1 + \mathcal{F}_2$ and that F =1 is an envelope function for $\mathcal{F}_1$ and $\mathcal{F}_2$. By

Lemma 2.9, $\mathscr{F}_1 + \mathscr{F}_2$ is a Euclidean class of functions with respect to the envelope function $F_2 \equiv 2$. Now we need to show that $\mathscr{F}$ is a Euclidean class with respect to the envelope function $F \equiv 1$. Next, consider

$$N\left(\varepsilon \left\|F_2\right\|_{Q,1}, \mathscr{F}, L_1\left(Q\right)\right) \le A\varepsilon^{-V}, 0 < \varepsilon \le 1.$$

Let $\delta = 2\varepsilon$. Then

$$\begin{aligned} N\left(\delta, \mathscr{F}, L_1\left(Q\right)\right) &\le A2^V \delta^{-V}, 0 < \delta/2 \le 1, \\ &= A2^V \delta^{-V}, 0 < \delta \le 2. \end{aligned}$$

Note, for $f \in \mathscr{F}$, $\left|f\right| \le 1 = F$. Therefore,

$$\begin{aligned} N\left(\varepsilon \left\|F\right\|_{Q,1}, \mathscr{F}, L_1\left(Q\right)\right) &= N\left(\varepsilon, \mathscr{F}, L_1\left(Q\right)\right) \\ &\le A2^V \varepsilon^{-V}, 0 < \varepsilon \le 1. \quad \square \end{aligned}$$

In order to apply Pollard's Theorem, we need to show that the class of functions considered is also a permissible class of functions. This is used to insure the measurability of the space indexed by *c*, *a*, and *t*. The following definitions are found in Pollard (1984, Appendix C). Let $\xi_1, \xi_2, \ldots$ be measurable maps from a probability space $\left(\Omega, \mathscr{E}, P\right)$ into a set S equipped with the σ-field $\mathscr{L} / \mathscr{B}\left(R\right)$. Let $\mathscr{F}$ be a class of $\mathscr{L} / \mathscr{B}\left(R\right)-$measurable real valued functions on S. We consider the empirical measure P_n attached to each f in $\mathscr{F}$ to get the real number

$$P_n f = n^{-1} \sum_{i=1}^{n} f\left(\xi_i\right).$$

The definition of a permissible class of functions depends upon the underlying probability space $\left(\Omega, \mathscr{E}, P\right)$ and the set S equipped with the σ-field $\mathscr{L}$.

Suppose that a class $\mathscr{F}$ is indexed by a parameter t that ranges over some set T. Let $\mathscr{B}\left(T\right)$ be the Borel σ-field on T. Let $\mathscr{B}\left(T\right) \otimes \mathscr{L}$ be the product σ-field associated with $\mathscr{B}\left(T\right)$ and $\mathscr{L}$. Assume that T is a separable metric space. The class of functions $\mathscr{F}$ is said to be permissible if it can be indexed by a T in such a way that

(i) the function $f\left(\bullet, \bullet\right)$ is $\mathscr{L} \otimes \mathscr{B}\left(T\right)$ measurable as a function from $S \otimes T$ into the real line;

(ii) T is an analytic subset of a compact subset of compact metric space $\bar{T}$.

Note that the subset of (2.8) defined as

(2.9) $$\mathscr{F}^* = \left\{ f_{\substack{c_1, a_1, p_1 \\ c_2, a_2, p_2}}(y) = I\left(c_1(y-a_1) \le \Phi^{-1}(p_1)\right) - I\left(c_2(y-a_2) \le \Phi^{-1}(p_2)\right) \right\}$$

when indexed by the vector $(c_1, a_1, p_1, c_2, a_2, p_3) \in R^{4k+2}$, where $\|c_i\| \le L_c < \infty,\ \|a_i\| \le L_a < \infty,$ and $p_i \in (0,1)$ is a permissible class of functions. To see that (2.9) is a permissible class of functions, let

$$T = \left\{ (c_1, a_1, p_1, c_2, a_2, p_3) \in R^{4k+2} : \|c_i\| \le L_c < \infty, \|a_i\| \le L_a < \infty,\ \ p_i \in (0,1) \right\}$$

and

$\bar{T} = \left\{ (c_1, a_1, p_1, c_2, a_2, p_3) \in R^{4k+2} : \|c_i\| \le L_c < \infty, \|a_i\| \le L_a < \infty,\ \ p_i \in [0,1] \right\}$. Then $\bar{T}$ is a compact metric space. Since T is Lebsegue measurable with $T \subset \bar{T}$, T is analytic. Also $f(\bullet,\bullet)$ is a $\mathscr{L} \otimes \mathscr{B}(T)$ measurable function from $S \otimes T \mapsto R$.

2.5. <u>Pollard's Rate Theorem.</u>

In the representation theorems associated with the correlations statistics we will make use of the following result from Pollard (1984) and Lemmas 2.12 and 4.11, which are derived from Pollard's result.

Theorem 2.11. **(Pollard, 1984)** *For each n, let $\mathscr{F}_n$ be a permissible class of functions whose covering numbers satisfy*

$$\sup_Q N\left(\varepsilon, \mathscr{F}_n, L_1(Q)\right) \le A\varepsilon^{-W}, 0 < \varepsilon \le 1,$$

with constants A and W not depending on n. If $|f| \le 1$ and $\left(Pf^2\right)^{1/2} \le \delta_n$, for each f in $\mathscr{F}_n$, then for a non-increasing sequence of positive numbers $\{\alpha_n\}$ such that

$$\log n \big/ n\delta_n^2 \alpha_n^2 \to 0,\ \ as\ \ n \to \infty,$$

we have

$$\sup_{\mathscr{F}_n} \left| P_n f - Pf \right| \Big/ \delta_n^2 \alpha_n \to 0,\ \ w.p.1\ \ as\ n \to \infty.$$

Lemma 2.12 is a generalization of a Lemma from Bahadur (1966) for the standard normal distribution.

Lemma 2.12. *Let* $z_1, \ldots, z_n$ *be i.i.d. standard normal random variables and*

$$F_n(t) = n^{-1} \sum_{i=1}^{n} I(z_i \le t), \ -\infty < t < \infty .$$

Let $a_n(q)$ be a sequence of positive constants such that

$$a_n(q) \sim c_0 n^{-1/2} \log n^q, \text{ as } n \to \infty ,$$

for $q > 0$ *and* $c_0 > 0$. *Let* β_n *be a sequence of constants such that* $\beta_n \to \infty$, *as* $n \to \infty$, *and* $n^{-1/2} \log n^{1-q} \beta_n$ *is a non-increasing sequence of positive numbers. Then*

(2.10) $$\sup_{|s-t|<a_n(q)} \left| F_n(s) - F_n(t) - \Phi(s) + \Phi(t) \right| = o\left(n^{-3/4} \log n^{\frac{1+q}{2}} \beta_n \right), \text{ as } n \to \infty, \text{ almost surely.}$$

Proof. Define $\mathscr{G}_n = \left\{ f_{s,t}(z) = I(z \le t) - I(z \le s) : |s-t| \le a_n(q) \right\}$ and note that for $f_{s,t} \in \mathscr{G}_n$

$$(P_n - P) f_{s,t} = \left| F_n(s) - F_n(t) - \Phi(s) + \Phi(t) \right|.$$

Let $\mathfrak{F}^*$ be defined as in (2.9), then $\mathscr{G}_n \subseteq \mathfrak{F}^*$, for every n, then $N(\varepsilon, \mathscr{G}_n, L_1(Q)) \le N(\varepsilon, \mathfrak{F}^*, L_1(Q))$ and $\mathscr{G}_n$ is a permissible class of functions for every n. By Corollary 2.10,

$$\sup_Q \left| N(\varepsilon, \mathscr{G}_n, L_1(Q)) \right| \le A\varepsilon^{-W}, \text{ for } 0 < \varepsilon \le 1,$$

where A and W are constants not depending on n. Note that $\left| f_{s,t} \right| \le 1$, for every $f_{s,t} \in \mathscr{G}_n$. Consider

$$\begin{aligned} Pf_{s,t}^2 &= E_\Phi I(s < z \le t)^2 \\ &\le \Phi(t) - \Phi(t - a_n(q)) \\ &= \phi(t^*) a_n(q), \text{ for } t \in (t - a_n(q), t), \\ &\le (2\pi)^{-1/2} a_n(q), t \in R. \end{aligned}$$

Let $\delta_n^2 = (2\pi)^{-1/2} a_n(q)$ and $\alpha_n^2 = n^{-1/2} \log n^{1-q} \beta_n^2$ then, by Theorem 2.11,

(2.11) $$\sup_{|s-t| \le a_n(q)} \left| (P_n - P) f_{s,t} \right| = o(\delta_n^2 \alpha_n), \ w.p.1, \text{ as } n \to \infty.$$

Next, note

$$\delta_n^2 = (2\pi)^{-1/2} a_n(q) = c_0 n^{-1/2} \log n^q$$

and

$$\alpha_n^2 = n^{-1/2} \log n^{1-q} \beta_n^2 .$$

Therefore we can rewrite (2.10) as

$$\sup_{|s-t|<a_n(q)} \left| F_n(s) - F_n(t) - \Phi(s) + \Phi(t) \right| = o\left(n^{-3/4} \log n^{\frac{1+q}{2}} \beta_n \right), \ w.p.1, \text{ as } n \to \infty. \quad \square$$

Corollary 2.13. $\sup_{|s-t|<c_0 n^{-1/2} \log n^{1/2}} \left| F_n(s) - F_n(t) - \Phi(s) + \Phi(t) \right| = o\left(n^{-3/4} \log n \right), \ w.p.1, \text{ as } n \to \infty.$

Proof. Take $q = 1/2$, $\beta_n = \log n^{1/4}$, and apply Lemma 2.12. □

Chapter III. An Asymptotic Representation for Empirical Processes of Projections

3.1. **Basic Definitions and Introduction.** Let $Y_1,\ldots,Y_n$ be k-dimensional multivariate vectors with mean, μ and positive definite covariance matrix, Σ. Let S denote the sample covariance matrix. Let $\overline{Y}$ denote the sample mean. Let $\Phi(t)$ and $\phi(t)$, $-\infty < t < \infty$, be the cumulative distribution and probability density functions of a standard normal distribution, respectively.

Assumption 3.1. *For $Y_1,\ldots,Y_n$ be i.i.d. random vectors with a positive definite covariance matrix Σ and $\hat{d} = \hat{d}\left(Y_1,\ldots,Y_n\right)$ a sequence of row vectors,*

(i) $\hat{d}S\hat{d}' = 1;$

and

(ii) $\hat{d} \xrightarrow{P} d_0$*, as* $n \to \infty$*, for* $d_0 \in R^k$*;*

where

(iii) $d_0 \Sigma d_0' = 1.$

If $Y_1,\ldots,Y_n$ are i.i.d. multivariate normal vectors, then $\{d_0(Y_i - \mu);\ i = 1,\ldots,n\}$ are i.i.d. standard normal random variables. Here we consider the transformed data $\{\hat{d}(Y_i - \overline{Y}), i = 1,\ldots,n\}$ and the corresponding univariate empirical process,

(3.1) $G_n(t) = n^{1/2}(n^{-1}\sum_{i=1}^{i=n} I(\hat{d}(Y_i - \overline{Y}) \le t) - \Phi(t)),\ -\infty < t < \infty.$

In this chapter, it is shown under the assumption of multivariate normality that $G_n(t), -\infty < t < \infty,$ is asymptotically equivalent to

(3.2)
$$\begin{aligned} A_n(t) = \sqrt{n}\Big(n^{-1}\sum_{i=1}^{i=n}\big[&I(d_0(Y_i - \mu) \le t) - \Phi(t) \\ &+ t\phi(t)(d_0(Y_i - \mu))^2/2 - t\phi(t)/2 \\ &+ \phi(t)d_0(Y_i - \mu)\big]\Big),\ -\infty < t < \infty, \end{aligned}$$

uniformly over t, for estimated projections, $\hat{d}$, that satisfy Assumption 3.1. Then it is shown that A_n converges weakly to G, as n tends to infinity, where G is a tight zero mean Gaussian process with covariance function

$$E\left(G(t)G(s)\right) = \Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \frac{ts\phi(s)\phi(t)}{2}, \ -\infty < s,t,< \infty .$$

This implies the weak convergence of G_n to G, as $n \to \infty$, in $l^\infty\left(R\right)$.

3.2. <u>An application of Theorem 2.3.</u> In order to apply Theorem 2.3, it is first necessary to choose a proper class of functions. For $t \in R$, $c', a \in R^k$, define

(3.3) $\Im = \{ f_{c,a,t} = I(c(y-a) \le t) : t \in R, c^t, a \in R^k \}.$

Then

$$G_n(t) = n^{1/2}\left[P_n f_{\hat{d},\bar{Y},t} - Pf_{d_0,\mu,t}\right], \ -\infty < t < \infty.$$

In particular, we consider $Y_1,\ldots,Y_n$ i.i.d. on the measure space $\left(\mathcal{X}, \mathcal{A}, P\right)$ where $\mathcal{X}$ is restricted to R^k. Here, we show that $\Im$ is a P-Donsker class of functions. Our approach is to apply Theorem 2.3. This will result in a approximating process which is the usual empirical process plus a drift term. The majority of this chapter deals with representing the drift term as a sum of i.i.d. random elements, uniformly over the real line.

Lemma 3.1. Let $\Im$ be the class of functions defined in (3.3). Then $\Im$ is a P-Donsker class of functions.

Proof. By Corollary 2.10, $\Im$ is a Euclidean class of functions with an envelope function of 1, which implies that $J\left(1,\Im,L_2\right) < \infty$. Then, by Theorem 2.4, $\Im$ is a P-Donsker class of functions.

□

In Lemma 3.2, we verify the second assumption of Theorem 2.3 for the class of functions defined in (3.3) and $\hat{d}$ satisfying Assumption 3.1.

Lemma 3.2. *Let* $\Im$ *be the class of functions defined in (3.3) and* P *a continuous probability measure on* $\left(R^k, \mathcal{A}\right)$*. Then*

$$\sup_{t\in R}\left\| f_{c,a,t} - f_{c_0,a_0,t}\right\|_{L_2(P)} = \sup_{t\in R}\int (f_{c,a,t}(y) - f_{c_0,a_0,t}(y))^2\, dP(y) = o(1),$$

as $(c,a)\to(c_0,a_0)$.

Proof. Consider

$$\begin{aligned}\left\| f_{c,a,t} - f_{c_0,a_0,t}\right\|_{L_2(P)} &= \int (f_{c,a,t}(y) - f_{c_0,a_0,t}(y))^2\, dP(y)\\ &= E[I(c(Y-a)\le t) - I(c_0(Y-a_0)\le t)]^2\\ &= E[I(c(Y-a)\le t, c_0(Y-a_0)>t) + I(c(Y-a)>t, c_0(Y-a_0)\le t)].\end{aligned}$$

Let $M = (c-c_0)(Y-a_0) + c(a_0-a)$ and note that

$|M| \le \|c-c_0\|\|Y-a_0\| + \|c\|\|a_0-a\| = o(1)O_p(1) + O(1)o(1) = o_p(1)$, as $(c,a)\to(c_0,a_0)$.

Therefore, we have

$$\begin{aligned}\int (f_{c,a,t}(y) - f_{c_0,a_0,t}(y))^2\, dP(y) &= P[c_0(Y-a_0) + (c-c_0)(Y-a_0) + c(a_0-a)\le t, c_0(Y-a_0)>t]\\ &\quad + P[c_0(Y-a_0) + (c-c_0)(Y-a_0) + c(a_0-a)>t, c_0(Y-a_0)\le t]\\ &= P[M + c_0(Y-a_0)\le t, c_0(Y-a_0)>t]\\ &\quad + P[c_0(Y-a_0) + M > t, c_0(Y-a_0)\le t].\end{aligned}$$

First consider

$$\begin{aligned}P[M + c_0(Y-a_0)\le t, c_0(Y-a_0)>t] &= P[M + c_0(Y-a_0)\le t, c_0(Y-a_0)>t, |M|\le\delta]\\ &\quad + P[M + c_0(Y-a_0)\le t, c_0(Y-a_0)>t, |M|>\delta]\\ &\le P[t< c_0(Y-a_0)\le t-M, |M|\le\delta] + P[|M|>\delta]\\ &\le P[t< c_0(Y-a_0)\le t+\delta] + P[|M|>\delta], \forall\delta>0.\end{aligned}$$

By a similar process, we get that, for every $\delta>0$,

$$P[c_0(Y-a_0) + M > t,\ c_0(Y-a_0)\le t] \le P[t-\delta < c_0(Y-a_0)\le t] + P[|M|>\delta].$$

Now we have that

$$\int (f_{c,a,t}(y))^2\, dP(y) \le P[t-\delta < c_0(Y-a_0)\le t] + P[t< c_0(Y-a_0)\le t+\delta] + 2P[|M|>\delta],$$

for every $\delta>0$ and for every $t\in R$. Then $P[|M|>\delta] = o(1)$, for every $\delta>0$, as c approaches c_0 and a approaches a_0. Since $c_0(Y-d_0)$ is a continuous random variable, there exists a $\delta>0$ for every $\varepsilon>0$ such that $P[t< c_0(Y-d_0)\le t+\delta]\le\varepsilon$, for every $t\in R$. Therefore,

$\sup_{t\in R}\int (f_{c,a,t} - f_{d_0,\mu,t})^2 dP = o(1)$, as c approaches c_0 and a approaches a_0.

□

We will denote the common empirical process evaluated at $Pf_{d_0,\mu,t}$, $-\infty < t < \infty$, by $\mathcal{W}_n(t)$, $-\infty < t < \infty$, i.e.,

$$\mathcal{W}_n(t) = n^{1/2}(P_n - P) f_{d_0,\mu,t}, -\infty < t < \infty.$$

Note that under the assumption of normality of $Y_1,\ldots,Y_n$,

$$\mathcal{W}_n(t) = n^{1/2}\left[n^{-1}\sum_{i=1}^{n} I\left(d_0(Y_i - \mu) \le t\right) - \Phi(t)\right], -\infty < t < \infty.$$

Lemma 3.3. *Let $Y_1,\ldots,Y_n$ be k-dimensional multivariate normal random vectors with mean μ and positive definite covariance matrix Σ. Let $\hat{d}$ and d_0 satisfy Assumption 3.1. Then, for $-\infty < t < \infty$,*

$$G_n(t) = \mathcal{W}_n(t) + n^{-1/2}\left[\Phi\left(\frac{t + \hat{d}(\bar{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi(t)\right] + o_p(1), \textit{ as } n \to \infty.$$

Proof. By Assumption 3.1, we have that $\hat{d}_n$ is weakly consistent for d_0, as $n \to \infty$. By the Strong Law of Large numbers, we have that $\bar{Y}$ is strongly consistent for μ, as $n \to \infty$. Therefore, by Lemma 3.2, the assumptions to Theorem 2.3 are satisfied for the class of functions defined in (3.3), when $(c, a) = (\hat{d}, \bar{Y})$ and $(c_0, a_0) = (d_0, \mu)$.

Applying Theorem 2.3 we get that

$$n^{1/2}(P_n - P)(f_{\hat{d},\bar{Y},t} - f_{d_0,\mu,t}) = o_p(1), \text{ uniformly in } t \in R, \text{ as } n \to \infty.$$

Performing the above integrations, we get that

$$\text{(3.4)} \quad P_n f_{\hat{d},\bar{Y},t} - \Phi\left(\frac{t + \hat{d}(\bar{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - P_n f_{d_0,\mu,t} + \Phi(t) = o_p(n^{-\frac{1}{2}}),$$

uniformly in t, as $n \to \infty$. To complete the proof, rewrite (3.4) as

$$P_n f_{\hat{d},\bar{Y},t} - \Phi(t) = \left[P_n f_{d_0,\mu,t} - \Phi(t)\right]$$
$$+\left[\Phi\left(\frac{t+\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi(t)\right] + o_p(n^{-\frac{1}{2}}), \text{ as } n\to\infty,$$

uniformly in $t \in R$.

□

3.3. <u>The Characterization of the Drift Term.</u> In Lemma 3.3, G_n is shown to be asymptotically equivalent to the standard empirical process plus a drift term. In this section, we will characterize the drift term in terms of a sum off i.i.d. random variables.

The drift term has the form of $\Phi(\hat{t}) - \Phi(t)$, where $\hat{t} = \frac{t+\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}$, $-\infty < t < \infty$.

To deal with the drift term we first need to find the asymptotic properties of $n^{1/2}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}} - 1\right)$ and $n^{1/2}\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}$.

Lemma 3.4. *Let $Y_1,\ldots,Y_n$ be k-dimensional random vectors with mean vector μ, positive definite covariance Σ, and $E\left(\|Y_1\|^4\right) < \infty$. Let $\hat{d}$ and d_0 satisfy Assumption 3.1. Then*

$$\sqrt{n}(\hat{d}\Sigma\hat{d}' - 1) = -\sqrt{n}\left[n^{-1}\sum_{i=1}^n\left(d_0\left(Y_i-\mu\right)\right)^2 - 1\right] + o_p(1), \text{ as } n\to\infty.$$

Proof. First consider

$$\begin{aligned}\sqrt{n}(\Sigma - S) &= \sqrt{n}(\Sigma - n^{-1}\sum_{i=1}^n (Y_i-\mu-\bar{Y}+\mu)(Y_i-\mu-\bar{Y}+\mu)') \\ &= \sqrt{n}(\Sigma - n^{-1}\sum_{i=1}^n((Y_i-\mu)(Y_i-\mu)' - (\bar{Y}-\mu)(Y_i-\mu)' \\ &\quad -(Y_i-\mu)(\bar{Y}-\mu)' + (\bar{Y}-\mu)(\bar{Y}-\mu)')) \\ &= \sqrt{n}(\Sigma - n^{-1}\sum_{i=1}^n (Y_i-\mu)(Y_i-\mu)') - \sqrt{n}(\bar{Y}-\mu)(\bar{Y}-\mu)'.\end{aligned}$$

By the Central Limit Theorem, we have $n^{1/2}(\bar{Y}-\mu) \sim N(0,\Sigma)$ and $(\bar{Y}-\mu)' = o_p(1)$. Therefore, (3.5) $\sqrt{n}(\Sigma - S) = \sqrt{n}(\Sigma - n^{-1}\sum_{i=1}^{n}(Y_i - \mu)(Y_i - \mu)') + o_p(1),$ as $n \to \infty.$

Next consider

$$\begin{aligned}\sqrt{n}(\hat{d}\Sigma\hat{d}' - 1) &= \sqrt{n}(\hat{d}\Sigma\hat{d}' - \hat{d}S\hat{d}') \\ &= \hat{d}\sqrt{n}(\Sigma - S)\hat{d}' \\ &= (\hat{d} - d_0)\left[\sqrt{n}(\Sigma - S)\right]\hat{d}' + d_0\left[\sqrt{n}(\Sigma - S)\right]\hat{d}'.\end{aligned}$$

By the consistency of $\hat{d}_n$ for d_0 and $\sqrt{n}(\Sigma - S) = O_p(1)$, we have the following

$$\begin{aligned}(\hat{d} - d_0)\sqrt{n}(\Sigma - S)\hat{d}' + d_0\sqrt{n}(\Sigma - S)\hat{d}' &= o_p(1) + d_0\left[\sqrt{n}(\Sigma - S)\right]\hat{d}' \\ &= o_p(1) + d_0\left[\sqrt{n}(\Sigma - S)\right](\hat{d}' - d_0') \\ &\quad + d_0\left[\sqrt{n}(\Sigma - S)\right]d_0' \\ &= o_p(1) + o_p(1) + d_0\left[\sqrt{n}(\Sigma - S)\right]d_0' \\ &= d_0\left[\sqrt{n}(\Sigma - S)\right]d_0' + o_p(1), \text{ as } n \to \infty.\end{aligned}$$

Therefore,

(3.6) $\sqrt{n}(\hat{d}\Sigma\hat{d}' - 1) = d_0\left[\sqrt{n}(\Sigma - S)\right]d_0' + o_p(1),$ as $n \to \infty.$

To complete the proof, note that

$$\begin{aligned}d_0\left[\sqrt{n}(\Sigma - S)\right]d_0' &= \sqrt{n}(1 - d_0Sd_0') \\ &= \sqrt{n}\left[1 - d_0\left(n^{-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(Y_i - \bar{Y}\right)'\right)d_0'\right] \\ &= \sqrt{n}\left[1 - n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i - \bar{Y}\right)\right)^2\right] \\ &= \sqrt{n}\left[1 - n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i - \mu\right)\right)^2\right] + o_p(1), \text{ as } n \to \infty. \qquad \square\end{aligned}$$

Now, by the Central Limit Theorem, $\sqrt{n}(1 - n^{-1}\sum_{i=1}^{n}(d_0(Y_i - \mu))^2 \rightsquigarrow N(0,\eta),$ as $n \to \infty$, where η is the fourth moment of a standard normal distribution.

Lemma 3.5. *Let $Y_1, \ldots, Y_n$ be k-dimensional random vectors with mean vector μ and positive definite covariance matrix Σ. Let $\hat{d}$ and d_0 satisfy Assumption 3.1. Let $\bar{Y}$ be the sample mean. Then*

$$\sqrt{n}\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}} = \sqrt{n}d_0(\bar{Y}-\mu) + o_p(1), \text{ as } n \to \infty.$$

Proof. The result follows immediately by noting $\hat{d}\Sigma\hat{d}' \xrightarrow{p} 1$, as $n \to \infty$, and

$$\begin{aligned}\sqrt{n}\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}} &= \sqrt{n}\hat{d}(\bar{Y}-\mu) + \sqrt{n}\hat{d}(\bar{Y}-\mu)(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}} - 1) \\ &= \sqrt{n}\hat{d}(\bar{Y}-\mu) + o_p(1) \\ &= \sqrt{n}d_0(\bar{Y}-\mu) + (\hat{d}-d_0)\sqrt{n}(\bar{Y}-\mu) + o_p(1) \\ &= \sqrt{n}d_0(\bar{Y}-\mu) + o_p(1), \text{ as } n \to \infty. \end{aligned}$$

□

Lemma 3.6. *Let $Y_1, \ldots, Y_n$ be k-dimensional random vectors with mean vector μ, positive definite covariance matrix Σ, and $E\left(\|Y_1\|^4\right) < \infty$. Let $\hat{d}$ and d_0 satisfy Assumption 3.1. Then*

$$\sqrt{n}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}} - 1\right) = \frac{\sqrt{n}(1-\hat{d}\Sigma\hat{d}')}{2} + o_p(1), \text{ as } n \to \infty.$$

Proof. Consider

$$\begin{aligned}\sqrt{n}\left(1 - \frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) &= \sqrt{n}\left(1 - \frac{1}{\hat{d}\Sigma\hat{d}'}\right)\left(\frac{1}{1 + \frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}}\right) \\ &= \sqrt{n}\left(\hat{d}\Sigma\hat{d}' - 1\right)\left(\frac{1}{\hat{d}\Sigma\hat{d}' + \sqrt{\hat{d}'\Sigma\hat{d}'}}\right) \\ &= \sqrt{n}\left(\hat{d}\Sigma\hat{d}' - 1\right)\left(\frac{1}{\hat{d}\Sigma\hat{d}' + \sqrt{\hat{d}\Sigma\hat{d}'}} - \frac{1}{2}\right) + \sqrt{n}\frac{\hat{d}\Sigma\hat{d}' - 1}{2}. \end{aligned}$$

Note that $\left(\frac{1}{\hat{d}\Sigma\hat{d}' + \sqrt{\hat{d}\Sigma\hat{d}'}} - \frac{1}{2}\right) \xrightarrow{P} 0$, as $n \to \infty$, because $\sqrt{\hat{d}\Sigma\hat{d}'} \xrightarrow{P} \sqrt{d_0 \Sigma d_0{}'} = 1$, as $n \to \infty$. Therefore, we have that $\sqrt{n}\frac{(\hat{d}\Sigma\hat{d}' - 1)}{2} = \sqrt{n}\left(1 - \frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) + o_p(1)$. □

Now by combining Lemma 3.4 and Lemma 3.6, we get that

(3.7) $$\sqrt{n}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}} - 1\right) = \frac{\sqrt{n}\left[n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i - \mu\right)\right)^2 - 1\right]}{2} + o_p(1), \text{ as } n \to \infty.$$

In Lemma 3.7, we will approximate the drift term from Lemma 3.3,

(3.8) $$\Phi\left(\frac{t + \hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi(t), \ -\infty < t < \infty.$$

The remainder term to the approximation to (3.8) is shown to be $O_p(n^{-1})$, uniformly in t. This guarantees that the remainder is converging to zero sufficiently fast in probability.

Lemma 3.7. *Let $Y_1, \ldots, Y_n$ be k-dimensional random vectors with mean vector μ, positive definite covariance matrix Σ, and $E\left(\left\|Y_1\right\|^4\right) < \infty$. Let $\hat{d}$ satisfy Assumption 3.1. Then*

$$\Phi\left(\frac{t + \hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi(t) = \phi(t)\left(\frac{\hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) + t\phi(t)\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}} - 1\right) + o_p\left(n^{-1/2}\right), \text{ as } n \to \infty,$$

uniformly in $t \in R$.

Proof. Consider

$$\Phi\left(\frac{t + \hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) = \phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)\left(\frac{\hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) + \frac{\phi'\left(t^*\right)}{2}\left(\frac{\hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)^2,$$

where t^* is between $\frac{t + \hat{d}(\overline{Y} - \mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}$ and $\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}$. Note that $\frac{\phi'\left(t^*\right)}{2}$ is a bounded function of t^*.

Therefore, by Lemma 3.5, we have

$$\Phi\left(\frac{t+\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)=\phi(t)\left(\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)$$

$$(3.9)\qquad +\left(\phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\phi(t)\right)\left(\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)$$

$$+O_p\left(n^{-1}\right),\ \text{as } n\to\infty,$$

uniformly in $t\in R$. Note that $\sup_{t\in R}\left|\phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\phi(t)\right|=o_p(1)$, as $n\to\infty$. This implies the following modification of (3.9),

$$(3.10)\quad \Phi\left(\frac{t+\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)=\phi(t)\left(\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)+o_p\left(n^{-1/2}\right),\ \text{as } n\to\infty,$$

uniformly in $t\in R$. Now consider

$$(3.11)\quad \Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\Phi(t)=t\phi(t)\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right)+t^2\frac{\phi'(t^*)}{2}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right)^2,$$

where t^* is between t and $\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}$. Next we will show that $t^2\frac{\phi'(t^*)}{2}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right)^2=O_p\left(n^{-1}\right)$, uniformly for $t\in R$, as $n\to\infty$.

We will consider two cases, the first is when $t>0$ and the second is for $t<0$.

Case 1: Let $t>0$. If $0<t\big/\sqrt{\hat{d}\Sigma\hat{d}'}<t^*<t$, then

$$\begin{aligned}\left|t^2\phi'(t^*)\right|&=t^2t^*\phi(t^*)\\ &\le\left(\sqrt{\hat{d}\Sigma\hat{d}'}\right)^3\phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)^3\\ &\le\left(\sqrt{\hat{d}\Sigma\hat{d}'}\right)^3\sup_{t\in R}\left|t^3\phi(t)\right|=O_p(1),\ \text{as } n\to\infty.\end{aligned}$$

If $0<t<t^*<\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}$, then

$$\left|t^2\phi'\left(t^*\right)\right| = t^2 t^* \phi\left(t^*\right)$$

$$\le t^2\phi(t)\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)$$

$$\le \left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)\sup_{t\in R}\left|t^3\phi(t)\right| = O_p(1), \text{ as } n\to\infty.$$

Case 2: Let $t<0$. If $t\Big/\sqrt{\hat{d}\Sigma\hat{d}'} < t^* < t < 0$, then

$$\left|t^2\phi'\left(t^*\right)\right| = t^2\left|t^*\right|\phi\left(t^*\right)$$

$$\le t^2\phi(t)\left|\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right|$$

$$\le \left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)\sup_{t\in R}\left|t^3\phi(t)\right| = O_p(1), \text{ as } n\to\infty.$$

If $t < t^* < \dfrac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}} < 0$, then

$$\left|t^2\phi'\left(t^*\right)\right| = t^2\left|t^*\right|\phi\left(t^*\right)$$

$$\le \left(\sqrt{\hat{d}\Sigma\hat{d}'}\right)^3 \phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)\left|\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right|^3$$

$$\le \left(\sqrt{\hat{d}\Sigma\hat{d}'}\right)^3 \sup_{t\in R}\left|t^3\phi(t)\right| = O_p(1), \text{ as } n\to\infty.$$

Therefore,

(3.12) $$\sup_{t\in R}\left|t^2\frac{\phi'\left(t^*\right)}{2}\right| = O_p(1), \text{ as } n\to\infty,$$

this implies

(3.13) $$\sup_{t\in R}\left|t^2\frac{\phi'\left(t^*\right)}{2}\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right)^2\right| = O_p\left(n^{-1}\right), \text{ as } n\to\infty,$$

and

(3.14) $$\Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right) - \Phi(t) = t\phi(t)\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right) + O_p\left(n^{-1}\right), \text{ as } n\to\infty,$$

uniformly for $t\in R$. To finish the proof combine (3.10) and (3.14) to get

$$\Phi\left(\frac{t-\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\Phi(t)=\Phi\left(\frac{t}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)-\Phi(t)+\phi(t)\left(\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)+o_p\left(n^{-1/2}\right)$$

$$=t\phi(t)\left(\frac{1}{\sqrt{\hat{d}\Sigma\hat{d}'}}-1\right)+\phi(t)\left(\frac{\hat{d}(\bar{Y}-\mu)}{\sqrt{\hat{d}\Sigma\hat{d}'}}\right)+o_p\left(n^{-1/2}\right), \text{ as } n\to\infty,$$

uniformly for $t\in R$. □

3.4. <u>The Limit of the Empirical Process of the Projections.</u> In this section, we first show the asymptotic equivalence of A_n and G_n, as $n\to\infty$. Then we will show that $A_n \rightsquigarrow G$, as $n\to\infty$, where G is a zero mean, tight Gaussian process with covariance function

$$\Phi(min(t,s))-\Phi(t)\Phi(s)-\phi(s)\phi(t)-\frac{ts\phi(s)\phi(t)}{2}, -\infty<s,t<\infty .$$

Lemma 3.8. *Let $Y_1,\ldots,Y_n$ be k-dimensional multivariate normal vectors with mean, μ and positive definite covariance matrix, Σ. Let G_n and A_n be defined as in (3.1) and (3.2) respectively. Then*

$$\sup_{t\in R}|G_n(t)-A_n(t)|=o_p(1), \text{ as } n\to\infty .$$

Proof. By Lemma 3.3, we have,

$$G_n(t)=W_n(t)+n^{1/2}\left(\Phi\left(\frac{t+\hat{d}_n(\bar{Y}-\mu)}{\sqrt{\hat{d}_n\Sigma\hat{d}_n'}}\right)-\Phi(t)\right)+o_p(1), \text{ as } n\to\infty,$$

uniformly in $t\in R$. Now by applying Lemma 3.7 to the drift term, we have,

$$G_n(t)=W_n(t)+\phi(t)n^{1/2}\left(\frac{\hat{d}_n(\bar{Y}-\mu)}{\sqrt{\hat{d}_n\Sigma\hat{d}_n'}}\right)+t\phi(t)n^{1/2}\left(\frac{1}{\sqrt{\hat{d}_n\Sigma\hat{d}_n'}}-1\right)+o_p(1), \text{ as } n\to\infty,$$

uniformly in $t\in R$. Now using (3.7),

$$\left(\frac{1}{\sqrt{\hat{d}_n\Sigma\hat{d}_n'}}-1\right)=\frac{n^{-1}\sum_{i=1}^{n}(d_0(Y_i-\mu))^2-1}{2}+o_p(n^{-1/2}), \text{ as } n\to\infty,$$

and Lemma 3.5,

$$\frac{\hat{d}_n(\bar{Y}-\mu)}{\sqrt{\hat{d}_n\Sigma\hat{d}_n'}}=d_0(\bar{Y}-\mu)+o_p(n^{-\frac{1}{2}}), \text{ as n}\to\infty,$$

we have that

$$G_n(t) = \mathcal{W}_n(t) + t\phi(t)n^{1/2}\left(\frac{n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i-\mu\right)\right)^2 - 1}{2}\right) + \phi(t)n^{1/2}d_0\left(\overline{Y}-\mu\right) + o_p(1)$$
$$= A_n(t) + o_p(1), -\infty < t < \infty, \text{ as } n \to \infty. \qquad \square$$

Note that $z_i = d_0(Y_i - \mu)$ are i.i.d. standard normal random variables. For completeness, we next find the covariance structure for A_n.

Lemma 3.9. *Let $Y_1, \ldots, Y_n$ be i.i.d. multivariate normal random vectors. The covariance function of A_n is given by*

$$Cov\left(A_n(t),\, A_n(s)\right) = \Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \frac{ts\phi(s)\phi(t)}{2}, \textit{ for } -\infty < s, t < \infty.$$

Proof. Let $X_t = I(Z \le t) + \frac{t\phi(t)Z^2}{2} - \Phi(t) - \frac{t\phi(t)}{2}$ and $Y_t = \phi(t)Z$, where $Z \sim N(0,1)$. Then

$$Cov\left(A_n(t),\, A_n(s)\right) = Cov(X_s + Y_s, X_t + Y_t)$$
$$= Cov(X_s, X_t) + Cov(Y_s, Y_t) + Cov(X_s, Y_t) + Cov(Y_s, X_t).$$

First consider $Cov(X_s, X_t)$ and note the following,

$$Cov(X_s, X_t) = E\left(I\left(Z \le t\right) + t\phi(t)\frac{Z^2}{2} - \Phi(t) - \frac{t\phi(t)}{2}\right)$$
$$\times\left(I\left(Z \le s\right) + s\phi(s)\frac{Z^2}{2} - \Phi(s) - \frac{s\phi(s)}{2}\right)$$
$$= \Phi(\min(t,s)) + \int_{-\infty}^{t} s\frac{z^2\phi(s)}{2}\phi(z)dz + \int_{-\infty}^{s} t\frac{z^2\phi(t)}{2}\phi(z)dz$$
$$+\frac{ts\phi(t)\phi(s)}{4}\int z^4\phi(z)dz - \Phi(t)\Phi(s) - \frac{s\phi(s)\Phi(t)}{2} - \frac{t\phi(t)\Phi(s)}{2} - \frac{ts\phi(s)\phi(t)}{4}.$$

Integrate the second and third terms by parts to get the following,

$$\int_{-\infty}^{s} z^2\phi(z)dz = -\int_{-\infty}^{s} z\phi'(z)dz$$
$$= -s\phi(s) + \int_{-\infty}^{s}\phi(z)dz$$
$$= -s\phi(s) + \Phi(s).$$

Note that $\int z^4\phi(z)dz$ =3, since the integral is the expectation of the square of a χ^2 random variable with one degree of freedom. Performing the above integrals we get the following,

$$\begin{aligned}
Cov(X_s, X_t) &= \Phi(min(t,s)) + \frac{s\phi(s)\Phi(t) - ts\phi(s)\phi(t)}{2} + \frac{t\phi(t)\Phi(s) - ts\phi(s)\phi(t)}{2} \\
&\quad + \frac{3ts\phi(s)\phi(t)}{4} - \Phi(t)\Phi(s) - \frac{t\phi(t)\Phi(s)}{2} - \frac{s\phi(s)\Phi(t)}{2} - \frac{ts\phi(s)\phi(t)}{4} \\
&= \Phi(\min(t,s)) - \Phi(t)\Phi(s) - \frac{ts\phi(s)\phi(t)}{2}.
\end{aligned}$$

Next note that $Cov(Y_s, Y_t) = \phi(s)\phi(t)$. Using calculations similar to those used for $Cov(X_s, X_t)$, we get the following,

$$\begin{aligned}
Cov(X_s, Y_t) &= E_Z\left(I(Z \le s) + s\phi(s)\frac{Z^2}{2} - \Phi(s) - \frac{s\phi(s)}{2} \right)\left(\phi(t)Z\right) \\
&= E_Z\left(ZI(Z \le s)\phi(t) + s\phi(s)\phi(t)\frac{Z^3}{2} - Z\phi(t)\Phi(s) - Z\frac{s\phi(s)\phi(t)}{2} \right) \\
&= -\phi(t)\phi(s).
\end{aligned}$$

Combine the above results to get the following,

$$\begin{aligned}
Cov\left(A_n(t), A_n(s)\right) &= Cov(X_s, X_t) + Cov(Y_s, Y_t) + Cov(X_s, Y_t) + Cov(Y_s, X_t) \\
&= \Phi(\min(t,s)) - \Phi(t)\Phi(s) - \frac{ts\phi(s)\phi(t)}{2} + \phi(s)\phi(t) - \phi(s)\phi(t) - \phi(t)\phi(s) \\
&= \Phi(\min(t,s)) - \Phi(t)\Phi(s) - \frac{ts\phi(s)\phi(t)}{2} - \phi(s)\phi(t). \qquad \square
\end{aligned}$$

The space $D[a,b]$ is the set of all Cadlag functions on an interval $[a,b] \subset \bar{R}$: functions $z : [a,b] \mapsto R$ that are continuous from the right and have limits from the left everywhere.

Theorem 3.10. *Let G_n be defined as in (3.1) and G be a tight Gaussian process with covariance* $\Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \dfrac{ts\phi(s)\phi(t)}{2}$.

Then

$G_n \rightsquigarrow G$, *in* $D[-\infty,\infty]$, as $n \to \infty$.

Proof. By Lemma 3.8, we have that $\sup_{t\in R} | G_n(t) - A_n(t) | = o_P(1)$, as $n \to \infty$, so it suffices to find the limiting distribution of A_n . Note that

$$\sup_{t\in R}\left|G_n(t)-A_n(t)\right| = \sup_{p\in(0,1)}\left|G_n\left(\Phi^{-1}(p)\right)-A_n\left(\Phi^{-1}(p)\right)\right|$$

and

$$\begin{aligned}A_n(\Phi^{-1}(p)) = {} & W_n\left(\Phi^{-1}(p)\right) \\ & +\Phi^{-1}(p)\phi\left(\Phi^{-1}(p)\right)n^{1/2}\left(\frac{n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i-\mu\right)\right)^2-1}{2}\right) \\ & +\phi\left(\Phi^{-1}(p)\right)n^{1/2}d_0\left(\bar{Y}-\mu\right), \quad \text{for } 0<p<1.\end{aligned}$$

First, we note that $W_n\left(\Phi^{-1}(p)\right), 0<p<1,$ converges to a uniform Brownian bridge in $D[0,1]$, as $n\to\infty$ (see Theorem 13.1, Billingsley, 1968). Next, note that $\Phi^{-1}(p)\phi\left(\Phi^{-1}(p)\right), 0<p<1,$ and $\Phi^{-1}(p)\phi\left(\Phi^{-1}(p)\right), 0<p<1,$ are bounded uniformly continuous on $[0,1]$. Therefore, by the Central Limit Theorem, the limiting distribution of the second and third terms is going to is going to be normal. This implies that

$$\Phi^{-1}(p)\phi\left(\Phi^{-1}(p)\right)n^{1/2}\left(\frac{n^{-1}\sum_{i=1}^{n}\left(d_0\left(Y_i-\mu\right)\right)^2-1}{2}\right)+\phi\left(\Phi^{-1}(p)\right)n^{1/2}d_0\left(\bar{Y}-\mu\right),$$

$0<p<1,$ converges to a tight Gaussian process as, in $D[0,1]$, $n\to\infty.$ The sum of two tight Gaussian processes is a tight Gaussian process. Therefore, $A_n\left(\Phi^{-1}(p)\right)$, $0<p<1$, converges weakly to a tight Gaussian process in $D[0,1]$, as $n\to\infty$. By Lemma 3.10, the covariance function of $A_n\left(\Phi^{-1}(p)\right)$, $0<p<1,$ is

$$\begin{aligned}Cov\left(G\left(\Phi^{-1}(p)\right),G\left(\Phi^{-1}(q)\right)\right) = {} & \Phi\left(min\left(\Phi^{-1}(p),\Phi^{-1}(q)\right)\right)-\Phi\left(\Phi^{-1}(p)\right)\Phi\left(\Phi^{-1}(q)\right) \\ & -\phi\left(\Phi^{-1}(p)\right)\phi\left(\Phi^{-1}(q)\right) \\ & -\frac{\Phi^{-1}(p)\Phi^{-1}(q)\phi\left(\Phi^{-1}(p)\right)\phi\left(\Phi^{-1}(q)\right)}{2},\end{aligned}$$

for $0<p,q<1$. To complete the proof, let $t=\Phi^{-1}(p)$. □

3.5. <u>The Asymptotic Distribution of Certain Functionals of the Empirical Process</u>. In Corollary 3.11, we apply the theory from the preceding sections to derive the distribution of continuous functionals applied to the empirical process of the estimated projections. Let T be a continuous functional from $\ell^{\infty}(R)$ to R. Two examples of continuous functionals of the empirical process are the Cramer-von Mises type statistic,

(3.15) $T_1\left[G_n\right] = \int \left[n^{1/2}(n^{-1}\sum_{i=1}^{i=n} I(\hat{d}(Y_i - \overline{Y}) \le t) - \Phi(t)) \right]^2 d\Phi(t)$,

and the Kolmogorov-Smirnov type statistic,

(3.16) $T_2\left[G_n\right] = \sup_{t \in R} \left| n^{1/2}(n^{-1}\sum_{i=1}^{i=n} I(\hat{d}(Y_i - \overline{Y}) \le t) - \Phi(t)) \right|$.

Corollary 3.12. *Let G_n be defined as in (3.1) and G be a tight Gaussian process with covariance $\Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \frac{ts\phi(s)\phi(t)}{2}$. Let T be a continuous functional from $\ell^{\infty}(R)$ to R.*

Then

$$T\left[G_n\right] \rightsquigarrow T\left[G\right], \text{ as } n \to \infty .$$

Proof. The result following immediately from Theorem 2.11 and Theorem 3.11. □

Chapter IV. Asymptotic Distribution Results for Correlation Tests

4.1. <u>Some Preliminary Results and Definitions.</u> Let $Y_1,\dots,Y_n$ be i.i.d. k-variate multivariate normal vectors with a mean, μ, and covariance matrix, $\Sigma_{k\times k}$, which is positive definite. Let $\hat{d}' \in R^k$ be a random vector which converges to $d_0{}' \in R^k$. Since d_0 is generally based upon the unknown parameters Σ we will consider

(4.1) $\hat{y}_i = \hat{d}\left(Y_i - \overline{Y}\right),\ i = 1,\dots,n$.

We are interested in deriving the asymptotic distribution of the DeWet and Venter statistic applied to $\left\{\hat{y}_i\right\}_{i=1}^n$. Let $r\left(U,V\right)$ be the sample correlation between two vectors U and V defined in (1.1). Then the correlation form of the modified de Wet and Venter statistic is

(4.2) $\hat{W}_n = r^2\left(\left(\hat{y}_{i:n}\right)_{n\times1},\xi\right)$,

where

(4.3) $\xi = \left(\Phi^{-1}\left(i/n+1\right)\right)_{n\times1}$

and $\hat{y}_{i:n}$ is the i^{th} order statistic from $\left\{\hat{y}_i\right\}_{i=1}^n$. Since $r^2(\cdot,\cdot)$ is location and scale invariant,

$$\begin{aligned}\hat{W}_n &= r^2\left(\left(\hat{y}_{i:n}\right)_{n\times1},\xi\right)\\ &= r^2\left(\left(\hat{y}_{i:n}\right)_{n\times1} - \hat{d}\left(\mu-\overline{Y}\right)1_{n\times1},\xi\right)\\ &= r^2\left(\left(\hat{d}\Sigma\hat{d}'\right)^{-1/2}\left(\left(\hat{y}_{i:n}\right)_{n\times1} - \hat{d}\left(\mu-\overline{Y}\right)1_{n\times1}\right),\xi\right).\end{aligned}$$

Let

(4.4) $\hat{c} = \left(\hat{d}\Sigma\hat{d}'\right)^{-1/2}\hat{d}$,

(4.5) $\hat{z}_i = \hat{c}(Y_i-\mu)$,

and

(4.6) $\hat{Z} = \left(\hat{z}_{i:n}\right)_{n\times1}$, where $\hat{z}_{i:n}$ is the i^{th} order statistic from $\{\hat{z}_i\}_{i=1}^{n}$. Then

$$\hat{Z} = \left(\hat{d}\Sigma\hat{d}'\right)^{-1/2}\left(\left(\hat{y}_{i:n}\right)_{n\times 1} - \hat{d}\left(\mu - \overline{Y}\right)1_{n\times 1}\right)$$

and

(4.7) $\quad \hat{W}_n = r^2\left(\hat{Z}, \xi\right)$.

Therefore, it is sufficient to consider

$$\hat{z}_i = \hat{c}\left(Y_i - \mu\right),\ i = 1, \ldots, n,$$

where $\hat{c}$ is given in (4.4) and satisfies $\hat{c}\Sigma\hat{c}' = 1$. Without loss of generality we may assume that $\mu = 0$. We will consider estimators, $\hat{d}$, which satisfy Assumption 4.1.

Assumption 4.1. *For $\hat{d} = \hat{d}\left(Y_1, \ldots, Y_n\right)$ a sequence of row vectors,*

i.) $\left\|\hat{d} - d_0\right\| = O_p\left(n^{-1/2}\right)$, *as* $n \to \infty$, *for* $d_0 \in R^k$,

where

ii.) $\left\|d_0\right\| > 0.$

It will also be convenient to let

$$c_0 = \left(d_0\Sigma d_0'\right)^{-1/2} d_0 .$$

Lemma 4.1. *Let $\hat{d}$ satisfy Assumption 4.1. Then*

i.) $\left(\hat{d}\Sigma\hat{d}'\right)^{-1/2} - \left(d_0\Sigma d_0'\right)^{-1/2} = O_p\left(n^{-1/2}\right)$, *as* $n \to \infty$,

and

ii.) $\left\|\hat{c} - c_0\right\| = O_p\left(n^{-1/2}\right)$, as $n \to \infty$.

Note that Lemma 4.1 **ii.)** implies $\hat{c} \xrightarrow{p} c_0$, as $n \to \infty$.

Proof. Let $\lambda_0 = \left(d_0\Sigma d_0'\right)^{1/2} > 0$ and $\hat{\lambda} = \left(\hat{d}\Sigma\hat{d}'\right)^{1/2}$. We will first consider (i), define $g : R^k \mapsto R$, by $g(x) = x'\Sigma x$, $x \in R^k$. Let $x_0 \in R^k$ be arbitrary but fixed. Let $x \in R^k$ such that $\left\|x - x_0\right\| < \delta$, $\delta > 0$.

Then

$$\begin{aligned}|g(x)-g(x_0)| &= |x'\Sigma x - x_0'\Sigma x_0| \\ &= \left|\left(x'-x_0'\right)\Sigma\left(x+x_0\right)\right|.\end{aligned}$$

Then, by the Cauchy –Schwartz inequality, we have

$$|g(x)-g(x_0)| = \left\|\left(x'-x_0'\right)\right\|\left\|\Sigma\left(x+x_0\right)\right\|.$$

By the matrix norm inequality, $\|\Sigma a\| \le \|\Sigma\|\|a\|$,

$$|g(x)-g(x_0)| \le \left\|\left(x'-x_0'\right)\right\|\left\|\Sigma\right\|\left\|\left(x+x_0\right)\right\|.$$

Note that

$$\begin{aligned}\|x+x_0\| &= \|x+x_0-x_0+x_0\| \\ &\le \|x-x_0\| + 2\|x_0\| \\ &\le \delta + 2\|x_0\|.\end{aligned}$$

Then $|g(x)-g(x_0)| \le \delta\|\Sigma\|\left(\delta+2\|x_0\|\right)$, which implies that if x is in a δ - neighborhood of x_0 then $g(x)$ is in a $\delta\|\Sigma\|\left(\delta+2\|x_0\|\right)$-neighborhood of $g(x_0)$. Therefore $g(\cdot)$ is a continuous function on R^k, which implies, by the convergence properties of transformed sequences (Serfling, 1980, pg. 24),

$$\hat{\lambda}^2 = \hat{d}\Sigma\hat{d}' = g\left(\hat{d}'\right) \xrightarrow{p} g\left(d_0'\right) = d_0\Sigma d_0' = \lambda^2, \text{ as } n\to\infty.$$

This implies that $\hat{\lambda}^{-1} \xrightarrow{p} \lambda_0^{-1}$, as $n\to\infty$, because $\lambda_0 > 0$.

Finally, consider

$$\begin{aligned}\|\hat{c}-c_0\| &= \left\|\hat{\lambda}^{-1}\hat{d} - \lambda_0^{-1}d_0\right\| \\ &= \left\|\hat{\lambda}^{-1}\hat{d} - \lambda_0^{-1}d_0 + \hat{\lambda}^{-1}d_0 - \hat{\lambda}^{-1}d_0\right\| \\ &= \left\|\hat{\lambda}^{-1}\left(\hat{d}-d_0\right) + \left(\hat{\lambda}_0^{-1}-\lambda_0^{-1}\right)d_0\right\| \\ &\le \left\|\hat{\lambda}^{-1}\left(\hat{d}-d_0\right)\right\| + \left\|\left(\hat{\lambda}_0^{-1}-\lambda_0^{-1}\right)d_0\right\| \\ &= \left|\hat{\lambda}^{-1}\right|\left\|\hat{d}-d_0\right\| + \|d_0\|\left|\hat{\lambda}_0^{-1}-\lambda_0^{-1}\right| \\ &= O_p\left(n^{-1/2}\right) + o_p(1) \\ &= o_p(1), \text{ as } n\to\infty.\end{aligned}$$

Next consider

$$
\begin{aligned}
\left|\hat{d}\Sigma\hat{d}' - \hat{d}\Sigma d_0{}'\right| &= \left|\hat{d}\Sigma\hat{d}' - d_0\Sigma\hat{d}' + \hat{d}\Sigma d_0{}' - d_0\Sigma d_0{}'\right| \\
&= \left|\hat{d}\Sigma\left(\hat{d} - d_0\right)' + d_0\Sigma\left(\hat{d} - d_0\right)'\right| \\
&= \left|\left(\hat{d} + d_0\right)\Sigma\left(\hat{d} - d_0\right)'\right| \\
&\le \left\|\hat{d} + d_0\right\|\left\|\Sigma\right\|\left\|\hat{d} - d_0\right\| \\
&= O_p(1)O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.
\end{aligned}
$$

This implies $\hat{\lambda}^2 - \lambda_0^2 = O_p\left(n^{-1/2}\right),$ as $n \to \infty,$

and

$$
\begin{aligned}
\hat{\lambda} - \lambda_0 &= \left(\hat{\lambda} + \lambda_0\right)^{-1}\left(\hat{\lambda}^2 - \lambda_0^2\right) \\
&= O_p(1)O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\hat{\lambda}^{-1} - \lambda_0^{-1} &= \left(\hat{\lambda}\lambda\ \right)^{-1}\left(\hat{\lambda} - \lambda_0\right) \\
&= O_p(1)O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.
\end{aligned}
$$

In other words

$$
\begin{aligned}
\left(\hat{d}\Sigma\hat{d}'\right)^{-1/2} - \left(d_0\Sigma d_0{}'\right)^{-1/2} &= \hat{\lambda}^{-1} - \lambda_0^{-1} \\
&= O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.
\end{aligned}
$$

Finally, consider

$$
\begin{aligned}
\left\|\hat{c} - c_0\right\| &= \left\|\hat{\lambda}^{-1}\hat{d} - \lambda_0^{-1}d_0\right\| \\
&= \left\|\hat{\lambda}^{-1}\hat{d} - \lambda_0^{-1}\hat{d} + \lambda_0^{-1}\hat{d} - \lambda_0^{-1}d_0\right\| \\
&= \left\|\left(\hat{\lambda}^{-1} - \lambda_0^{-1}\right)\hat{d} + \lambda_0^{-1}\left(\hat{d} - d_0\right)\right\| \\
&\le \left|\hat{\lambda}^{-1} - \lambda_0^{-1}\right|\left\|\hat{d}\right\| + \left|\lambda_0^{-1}\right|\left\|\hat{d} - d_0\right\| \\
&= O_p\left(n^{-1/2}\right)O_p(1) + O_p\left(n^{-1/2}\right), \text{ as } n \to \infty \\
&= O_p\left(n^{-1/2}\right), \text{ as } n \to \infty. \qquad \square
\end{aligned}
$$

Define

(4.8) $z_i = c_0 \left(Y_i - \mu \right)$

and

(4.9) $Z = \left(z_{i:n} \right)_{n \times 1}$,

where $z_{i:n}$ is the i^{th} order statistic from $\left\{ z_i \right\}_{i=1}^{n}$. Note that $z_1, \ldots, z_n$ are i.i.d. standard normal random variables. de Wet and Venter (1972) give the following result for the de Wet and Venter statistic,

(4.10) $W_n = r^2 \left(Z, \xi \right)$.

Theorem 4.2. (de Wet and Venter, 1972, Theorem 2) *If* $z_1, \ldots, z_n$ *are i.i.d. standard normal random variables,*
then

$$2n\left(1 - W_n^{1/2}\right) - a_n \rightsquigarrow \sum_{j=3}^{\infty} \left(X_i - 1 \right) / j, \text{ as } n \to \infty,$$

where $X_1, \ldots, X_n$ *are i.i.d. random variables with a Chi-squared distribution, one degree of freedom, and* $a_n = \left(n+1 \right)^{-1} \left\{ \sum_{i=1}^{n} \left(\frac{i}{n+1} \right) \left(1 - \frac{i}{n+1} \right) \left(\phi \left(\Phi^{-1} \left(\frac{i}{n+1} \right) \right)^{-2} \right) \right\} - \frac{3}{2}$.

Lemma 4.3 is a result due to Leslie et al. (1986) which gives the convergence rate of $\left\{ a_n \right\}$.

Lemma 4.3. (Leslie et al., 1986, Lemma) *There exists constants* c_1 *and* c_2, $0 < c_1 < c_2 < \infty$, *such that*

$$c_1 \log\log\left(n\right) < a_n < c_2 \log\log\left(n\right).$$

Using the above lemma and Theorem 4.2, we get Corollary 4.4 to Theorem 4.2. This result is found in the proof of Theorem 2 from de Wet and Venter (1972). Their proof is informative and is included for completeness.

Corollary 4.4. (de Wet and Venter, 1972, pg.145) *Let* $z_1, \ldots, z_n$ *are i.i.d. standard normal random variables. Then*

$$n\left(1-W_n\right)-a_n \rightsquigarrow \sum_{j=3}^{\infty}\left(X_i-1\right)/j, \text{ as } n\to\infty.$$

Proof. By Theorem 4.2, we have

$$2n\left(1-W_n^{1/2}\right)-a_n = O_p\left(1\right), \text{ as } n\to\infty.$$

This implies

$$2n^{1/2}\left(1-W_n^{1/2}\right)-n^{-1/2}a_n = O_p\left(n^{-1/2}\right), \text{ as } n\to\infty.$$

By Lemma 4.3, we have $n^{-1/2}a_n = o_p\left(1\right)$, as $n\to\infty$. Now consider

$$\begin{aligned} 2n\left(1-W_n^{1/2}\right)-n\left(1-W_n\right) &= n\left(1-W_n^{1/2}\right)^2 \\ &= \left(n^{1/2}\left(1-W_n^{1/2}\right)\right)^2 \\ &= o_p\left(1\right), \text{ as } n\to\infty. \end{aligned} \tag{4.11}$$

Finally, consider

$$n\left(1-W_n\right)-a_n = 2n\left(1-W_n^{1/2}\right)-a_n+\left[n\left(1-W_n\right)-2n\left(1-W_n^{1/2}\right)\right].$$

By Theorem 4.1, $2n\left(1-W_n^{1/2}\right)-a_n \rightsquigarrow \sum_{j=3}^{\infty}\left(X_i-1\right)/j$, as $n\to\infty$, and by (4.11) $2n\left(1-W_n^{1/2}\right)-n\left(1-W_n\right)=o_p\left(1\right)$, as $n\to\infty$. Therefore, apply Slutsky's Theorem to complete the proof. □

Our goal is to show

$$n\left(1-\hat{W}_n\right)-a_n \rightsquigarrow \sum_{j=3}^{\infty}\left(X_i-1\right)/j, \text{ as } n\to\infty. \tag{4.12}$$

As a first step, we consider

$$n\left(1-\hat{W}_n\right)-a_n = n\left(1-W_n\right)-a_n+n\left(\hat{W}_n-W_n\right).$$

However, since $\sum_{i=1}^{n}\xi_i = 0$, we have

(4.13) $n\left(\hat{W}_n - W_n\right) = \dfrac{n}{n^{-1}\sum_{i=1}^{n}\xi_i^{\,2}}\left[\left(1-\dfrac{\hat{T}_n^2}{\hat{s}_n^2}\right)\left(1-\dfrac{T_n^{\,2}}{s_n^2}\right)\right],$

where

(4.14) $T_n = n^{-1}\sum_{i=1}^{n} z_{i:n}\xi_i,$

(4.15) $\hat{T}_n = n^{-1}\sum_{i=1}^{n} \hat{z}_{i:n}\xi_i ,$

s^2 is defined as the sample variance of the z_i's, i.e.,

$$s^2 = n^{-1}\sum_{i=1}^{n} z_i^{\,2} - \left(n^{-1}\sum_{i=1}^{n} z_i\right)^2 ,$$

and $\hat{s}^2$ is defined as the sample variance of the $\hat{z}_i$'s, i.e.,

$$\hat{s}^2 = n^{-1}\sum_{i=1}^{n} \hat{z}_i^{\,2} - \left(n^{-1}\sum_{i=1}^{n} \hat{z}_i\right)^2 .$$

Let

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(Y_i - \overline{Y}\right)' .$$

Then

(4.16)
$$s^2 = n^{-1}\sum_{i=1}^{n}\left(c_0 Y_i\right)^2 - \left(c_0\overline{Y}\right)^2 = c_0\left[n^{-1}\sum_{i=1}^{n} Y_i Y_i' - \overline{Y}\overline{Y}'\right]c_0'$$
$$= c_0\hat{\Sigma}c_0'$$

and, by a similar argument,

(4.17) $\hat{s}^2 = \hat{c}\hat{\Sigma}\hat{c}' .$

Hoeffding (1953) showed that

$$n^{-1}\sum_{i=1}^{n}\Phi^{-1}\left(\frac{i}{n+1}\right)^2 \to \int_0^1\left[\Phi^{-1}\left(t\right)\right]^2 dt = 1, \text{ as } n\to\infty.$$

Therefore, in order to show (4.12) we will need to show

(4.18) $\hat{V}_n - V_n = o_p\left(1\right),$ as $n \to \infty,$

where

(4.19) $\hat{V}_n = n\left(1 - \frac{\hat{T}_n^{\,2}}{\hat{s}^2}\right)$

and

(4.20) $V_n = n\left(1 - \frac{T_n}{s^2}\right).$

In order to show (4.18), following del Barrio et al (2000), we will work with

(4.21) $\hat{U}_n = n\left(\hat{s}_n^2 - \hat{T}_n^2\right)$

and

(4.22) $U_n = n\left(s^2 - T_n^{\,2}\right)$

and show

(4.23) $\hat{U}_n - U_n = o_p(1)$, as $n \to \infty$.

The relationship between (4.18) and (4.23) is explained in the following lemma and corollary.

Lemma 4.5. *Let* $\hat{V}_n$, V_n, $\hat{U}_n$, *and* U_n *be defined as in (4.19), (4.20), (4.21), and (4.22), respectively. Let* $\hat{d}$ *satisfy Assumption 4.1. Then*

$$\hat{V}_n - V_n = \hat{s}_n^{\,-2}\left(\hat{U}_n - U_n\right) + o_p(1), \text{ as } n \to \infty.$$

Proof. Consider

$$\begin{aligned}
\hat{V}_n - V_n &= \hat{s}_n^{\,-2}\hat{U}_n - s_n^{-2}U_n \\
&= \hat{s}_n^{\,-2}\left(\hat{U}_n - U_n + U_n\right) - s_n^{-2}U_n \\
&= \hat{s}_n^{\,-2}\left(\hat{U}_n - U_n\right) + U_n\left(\hat{s}_n^{\,-2} - s_n^{-2}\right) \\
&= \hat{s}_n^{\,-2}\left(\hat{U}_n - U_n\right) + n\left(s^2 - T_n^{\,2}\right)\left(\hat{s}_n^{\,-2} - s_n^{-2}\right) \\
&= \hat{s}_n^{\,-2}\left(\hat{U}_n - U_n\right) + n\left(s^2 - T_n^{\,2}\right)\left(\hat{s}_n^{\,2} - s_n^2\right)\Big/\hat{s}_n^{\,2}s_n^2.
\end{aligned}$$

Now, since $\hat{s}_n^{\,2} - s_n^2 = \left(\hat{c} + c_0\right)\left(\hat{\Sigma} - \Sigma\right)\left(\hat{c} - c_0\right)' = O_p\left(n^{-1}\right)$, as $n \to \infty$, $s^2 \xrightarrow{as} 1$, and $T_n \xrightarrow{as} 1$, as $n \to \infty$, we have that

$$n\left(s^2 - T_n^{\,2}\right)\left(\hat{s}_n^{\,2} - s_n^2\right)\Big/\hat{s}_n^{\,2}s_n^2 = o_p(1), \text{ as } n \to \infty.$$

Therefore,

$$\hat{V}_n - V_n = \hat{s}_n^{-2}\left(\hat{U}_n - U_n\right) + o_p(1), \text{ as } n \to \infty. \qquad \square$$

Corollary 4.6. *If* $\hat{U}_n - U_n = o_p(1)$, *as* $n \to \infty$, *then* $\hat{V}_n - V_n = o_p(1)$, *as* $n \to \infty$.

Proof. Note that $\hat{s}_n^{\,2} = \hat{c}\hat{\Sigma}\hat{c}' \xrightarrow{p} 1$, as $n \to \infty$, and the result follows from Slutsky's Theorem.

$\square$

Corollary 4.6 allows us to work with

(4.24) $\hat{U}_n - U_n = n\left(\hat{s}_n^2 - s_n^2\right) - n\left(\hat{T}_n^{\,2} - T_n^{\,2}\right).$

To characterize $\left(\hat{T}_n^{\,2} - T_n^{\,2}\right)$ we will use a decomposition similar to Sen et al. (2003), which we state as Lemma 4.7. Let

(4.25) $B_1 = n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)^2$

and

(4.26) $B_2 = n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)\left(z_{i:n} - \xi_i\right).$

Lemma 4.7. *Let* T_n, $\hat{T}_n$, B_1, *and* B_2 *be defined as in (4.14), (4.15), (4.25), and (4.26), respectively. Then*

(4.27) $\hat{T}_n - T_n = \left(\hat{s}_n^2 - s_n^2\right)/2 + B_1/2 - B_2 + O_p\left(n^{-3/2}\right)$, as $n \to \infty$.

Proof. Consider

$$\begin{aligned}
\hat{T}_n - T_n &= n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)\xi_i \\
&= n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)\left\{\left(z_{i:n} + \hat{z}_{i:n}\right)/2 - \left(z_{i:n} + \hat{z}_{i:n}\right)/2 + \xi_i\right\} \\
&= n^{-1}\sum_{i=1}^{n} z_{i:n}^2/2 - n^{-1}\sum_{i=1}^{n}\hat{z}_{i:n}^2/2 + n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)\left\{-\frac{1}{2}\left(z_{i:n} - z_{i:n} + z_{i:n} + \hat{z}_{i:n}\right) + \xi_i\right\} \\
&= n^{-1}\sum_{i=1}^{n} z_{i:n}^2/2 - n^{-1}\sum_{i=1}^{n}\hat{z}_{i:n}^2/2 + n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)^2/2 - n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)\left(z_{i:n} - \xi_i\right).
\end{aligned}$$

Making use of the identities

$$n^{-1}\sum_{i=1}^{n} z_i^2 = s^2 + \left(c_0\overline{Y}\right)^2$$

and

$$n^{-1}\sum_{i=1}^{n} \hat{z}_i^2 = \hat{s}^2 + \left(\hat{c}\overline{Y}\right)^2,$$

we have $\hat{T}_n - T_n = s_n^2/2 - \hat{s}_n^2/2 + B_1/2 - B_2 + \left(c_0\overline{Y}\right)^2/2 - \left(\hat{c}\overline{Y}\right)^2/2$.

Note, that by Lemma 4.1 (ii), $\left\|c_0 - \hat{c}\right\| = O_p\left(n^{-1/2}\right)$, as $n \to \infty$, and, since $\mu = 0$, $\left\|\overline{Y}\right\| = O_p\left(n^{-1/2}\right)$, which implies that

$$\begin{aligned}\left|\left(\hat{c}\overline{Y}\right)^2 - \left(c_0\overline{Y}\right)^2\right| &= \left|\left(\hat{c}\overline{Y} - c_0\overline{Y}\right)\left(\hat{c} + c_0\right)\overline{Y}\right| \\ &= \left|\left(\hat{c} - c_0\right)\overline{Y}\left(\hat{c} + c_0\right)\overline{Y}\right| \\ &\leq \left\|\hat{c} - c_0\right\|\left\|\overline{Y}\right\|\left\|\hat{c} + c_0\right\|\left\|\overline{Y}\right\| \\ &= O_p\left(n^{-1/2}\right)O_p\left(n^{-1/2}\right)O_p\left(1\right)O_p\left(n^{-1/2}\right) \\ &= O_p\left(n^{-3/2}\right), \text{ as } n \to \infty.\end{aligned}$$

Therefore

$$\hat{T}_n - T_n = \left(\hat{s}_n^2 - s_n^2\right)/2 + B_1/2 - B_2 + O_p\left(n^{-3/2}\right), \text{ as } n \to \infty. \qquad \square$$

Next notice, by the Cauchy-Schwarz Inequality,

(4.28)
$$\begin{aligned}B_2 &\leq n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \hat{z}_{i:n}\right)^2 n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \xi_i\right)^2 \\ &= B_1\, n^{-1}\sum_{i=1}^{n}\left(z_{i:n} - \xi_i\right)^2.\end{aligned}$$

de Wet and Venter (1972) proposed $L_n^0 = \sum_{i=1}^{n}\left(z_{i:n} - \xi_i\right)^2$ as a test statistic to test the null hypothesis that $\left\{z_i\right\}_{i=1}^{n}$ have a standard normal distribution. Specifically, for

$$v(t) = t(1-t),\ 0 \leq t \leq 1,$$

$$H(t) = \Phi^{-1}(t),\ 0 \leq t \leq 1,$$

$$h(t) = H'(t), 0 \leq t \leq 1,$$

and

$$a_n^0 = (n+1)^{-1} \sum_{k=1}^{n} h\left(\tfrac{k}{n+1}\right) v\left(\tfrac{k}{n+1}\right),$$

they proved the following lemma.

Lemma 4.8. (de Wet and Venter, 1972, Thm. 1) *Let* z_i *be i.i.d.* $N(0,1)$, *for* $i = 1,\ldots,n$ *then*

$$L_n^0 - a_n^0 \rightsquigarrow \sum_{i=1}^{\infty} (X_i - 1)/j, \text{ as } n \to \infty,$$

where $X_1,\ldots,X_n$ *are i.i.d. random variables with a Chi-squared distribution, one degree of freedom.*

Note that

$$H'(t) = 1/\phi(\Phi(t)),\ 0 < t < 1$$

and

$$h(t) = \phi(\Phi(t))^{-2},\ 0 < t < 1.$$

Therefore

$$a_n^0 = (n+1)^{-1} \left\{ \sum_{j=1}^{n} \left(\frac{i}{n+1}\right)\left(1 - \frac{i}{n+1}\right)\left(\phi\left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right)^{-2}\right) \right\} - \frac{3}{2}$$

and

$$a_n^0 = a_n + \frac{3}{2}.$$

This immediately implies the following corollary to Lemma 4.3.

Corollary 4.9. *There exists constants* $c_1, c_2, 0 < c_1 < c_2 < \infty,$ *such that*

$$c_1 \log\log(n) + 3/2 < a_n^0 < c_2 \log\log(n) + 3/2.$$

Combining the results from Lemma 4.8 and Corollary 4.9, we have

$$n^{-1}\sum_{i=1}^{n}\left(z_{i:n}-\xi_i\right)^2 = n^{-1}L_n^0$$

$$(4.29)\qquad = n^{-1}O_p\left(\log\log(n)\right)$$

$$= O_p\left(n^{-1}\log\log(n)\right), \text{ as } n\to\infty.$$

Therefore, to prove (4.12) we need to show that $B_1 = o_p\left(n^{-(1+\delta)}\right)$, for some $\delta > 0$, as $n\to\infty$.

4.2. <u>Bounds on the difference between the two vectors of order statistics</u>. As noted above, to prove (4.12) we need to show that $B_1 = o_p\left(n^{-(1+\delta)}\right)$, for some $\delta > 0$, as $n\to\infty$. Since $B_1 = n^{-1}\sum_{i=1}^{n}\left(\hat{z}_{i:n}-z_{i:n}\right)^2$, our approach is to consider $\left|\hat{z}_{i:n}-z_{i:n}\right|$. The main results in this section are Lemma 4.14 and Lemma 4.16. In Lemma 4.14 we give a sequence of bounds on $\left|\hat{z}_{i:n}-z_{i:n}\right|$ which holds uniformly in i and tends to zero at a specific rate. Let $\{i_n\}_{n=1}^{\infty}$ be a sequence of positive integers such that $n^{1/4}i_n^{-1}\log n = O(1)$ and $i_n n^{-1} = o(1)$, as $n\to\infty$. In Lemma 4.16, we give a sequence of bounds on $\left|\hat{z}_{i:n}-z_{i:n}\right|$ for $i_n \le i \le n-i_n$, where the actual rate at which these bounds tend to zero depends on the choice of the sequence $\{i_n\}_{n=1}^{\infty}$. First we will show, in the corollary to Lemma 4.9 ,

$$(4.30)\quad \sup_t\left|F_n(t;\hat{c})-F_n(t;c_0)\right| = o_p\left(n^{-3/4}\log n\right), \text{ as } n\to\infty,$$

where

$$(4.31)\quad F_n(t;c) = n^{-1}\sum_{i=1}^{n} I\left(cY_i \le t\right),\ -\infty < t < \infty,\ c' \in R^k.$$

Next we prove Lemma 4.14, which, when combined with Corollary 2.13 and (4.30), implies Lemma 4.15, $\sup_i\left|\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right| = o_p\left(n^{-3/4}\log n\right)$, as $n\to\infty$. By taking advantage of the properties of $\Phi(t), -\infty < t < \infty$, Lemma 4.15 implies Lemma 4.16.

Lemma 4.9. *Let* $F_n(t;c)$ *be defined as in (4.31). Let* β_n *be a sequence of positive numbers such that* $\beta_n \to \infty$, $n \to \infty$, *such that* $n^{-1/2}(\log n)^{1/2}\beta_n^2$ *is a non-increasing sequence of positive numbers. Assume* $\|\hat{c}-c_0\| = O_p(n^{-1/2})$, $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, *and* $c_0\Sigma c_0{}' = 1$. *Then*

$$\sup_t \left|F_n(t;\hat{c}) - F_n(t;c_0)\right| = o_p\left(n^{-3/4}(\log n)^{3/4}\beta_n\right), \text{ as } n \to \infty.$$

Proof. First we will show that for $\|c-c_0\| \le Mn^{-1/2}$, and $c\Sigma c' = 1$,

$$\text{(4.32)} \quad \sup_{t,c}\left|F_n(t;c) - F_n(t;c_0)\right| = o\left(n^{-3/4}(\log n)^{3/4}\beta_n\right), \ w.p.1, \text{ as } n \to \infty.$$

To show (4.32) we will apply Theorem 2.11 to the class of function

$$\mathfrak{F}_n = \left\{ f_{c,t} = I(cY \le t) - I(c_0Y \le t) : t \in R, c\Sigma c' = 1, \|c-c_0\| = Mn^{-1/2}\right\},$$

where $c_0' \in R^k$ is fixed, $c_0\Sigma c_0' = 1$, and fixed $M > 0$. Note that $E\left[I(cY_1 \le t)\right] = \Phi(t)$, for every c such that $c\Sigma c' = 1$. Then, for every $f_{c,t} \in \mathfrak{F}_n$,

$$(P_n - P)f_{c,t} = F_n(t;c) - F_n(t;c_0) - \Phi(t) + \Phi(t), -\infty < t < \infty.$$

To show that

$$N\left(\varepsilon, \mathfrak{F}_n, L_1(Q)\right) \le A\varepsilon^{-W}, \ 0 < \varepsilon \le 1,$$

note that $\mathfrak{F}_n \subseteq \mathfrak{F}$, where $\mathfrak{F}$ is the class of functions defined in (2.8). Let $\mathfrak{F}^*$ be the permissible class of functions defined in (2.9). Then note that the bounds on $\mathfrak{F}^*$ can be chosen such that $\mathfrak{F}_n \subseteq \mathfrak{F}^*$, for every n. This also implies that $\{\mathfrak{F}_n\}_{n=1}^{\infty}$ is a sequence of permissible classes of functions. Now we will identify δ_n^2 such that $Pf_{c,t}^2 \le \delta_n^2$, for $f_{c,t} \in \mathfrak{F}_n$. Let $f_{c,t} \in \mathfrak{F}_n$ and consider

$$\begin{aligned}
Pf_{c,t}^2 &= E\left(I(cY_1 \le t) - I(c_0Y_1 \le t)\right)^2 \\
&= E\left[I(c_0Y_1 \le t)I(cY_1 > t)\right] + E\left[I(c_0Y_1 > t)I(cY_1 \le t)\right] \\
&= E\left[I(c_0Y_1 \le t)I(cY_1 > t)\right] + 1 - \Phi(t) - \left(1 - \Phi(t)\right) + E\left[I(c_0Y_1 > t)I(cY_1 \le t)\right] \\
&= 2E\left[I(c_0Y_1 \le t)I(cY_1 > t)\right] \\
&= 2E\left[I(c_0Y_1 \le t)I\left(c_0Y_1 > t + (c_0 - c)Y_1\right)\right] \\
&= 2E\left[I\left(t + (c_0 - c)Y_1 < c_0Y_1 \le t\right)\right].
\end{aligned}$$

Let $M^* = M\left\|\Sigma^{1/2}\right\|$, $z = c_0 Y_1 \sim N(0,1)$, and $\Sigma^{-1/2} Y_1 \sim MN(0, I)$. Note that $\left|(c_0 - c)\Sigma^{1/2}\Sigma^{-1/2}Y_1\right| \le \|c_0 - c\|\left\|\Sigma^{1/2}\right\|\left\|\Sigma^{-1/2}Y_1\right\|$ and $\|c_0 - c\| \le Mn^{-1/2}$. Therefore, we have

$$(4.33)\quad \begin{aligned} Pf_{c,t}^2 &= 2EI\left(t + (c_0 - c)\Sigma^{1/2}\Sigma^{-1/2}Y_1 < c_0 Y_1 \le t\right) \\ &\le 2EI\left(t - n^{-1/2}M^*\left\|\Sigma^{-1/2}Y_1\right\| < c_0 Y_1 \le t\right). \end{aligned}$$

Then, for every $B > 0$

$$(4.34)\quad \begin{aligned} &2EI\left(t - n^{-1/2}M^*\left\|\Sigma^{-1/2}Y_1\right\| < c_0 Y_1 \le t\right) \\ &\qquad = P\left(t - n^{-1/2}M^*\left\|\Sigma^{-1/2}Y_1\right\| < c_0 Y_1 \le t; \left\|\Sigma^{-1/2}Y_1\right\| \le B\right) \\ &\qquad\quad + P\left(t - n^{-1/2}M^*\left\|\Sigma^{-1/2}Y_1\right\| < c_0 Y_1 \le t; \left\|\Sigma^{-1/2}Y_1\right\| > B\right) \\ &\qquad \le P\left(t - n^{-1/2}M^* B < c_0 Y_1 \le t\right) + P\left(\left\|\Sigma^{-1/2}Y_1\right\| > B\right) \\ &\qquad = \Phi(t) - \Phi\left(t - n^{-1/2}M^* B\right) + P\left(\left\|\Sigma^{-1/2}Y_1\right\| > B\right) \\ &\qquad \le \left(\sup_t \phi(t)\right) n^{-1/2}M^* B + P\left(\left\|\Sigma^{-1/2}Y_1\right\| > B\right) \\ &\qquad = (2\pi)^{-1/2} n^{-1/2} M^* B + P\left(\left\|\Sigma^{-1/2}Y_1\right\| > B\right). \end{aligned}$$

Let z_i, $i = 1, \ldots k$, be i.i.d. standard normal random variables. Then

$$(4.35)\quad \begin{aligned} P\left(\|c_0 Y_1\| > B\right) &\le \sum_{i=1}^{k} P\left(|z_i| > B/k\right) \\ &= 2k\left(1 - \Phi(B/k)\right). \end{aligned}$$

By Serfling (1980, pg. 81)

$$(4.36)\quad 1 - \Phi\left((2\log n)^{1/2}\right) = O\left(n^{-1/2}\right), \text{ as } n \to \infty,$$

and (4.34) we have, for $B_n = k(2\log n)^{1/2}$,

$$Pf_{c,t}^2 \le (2\pi)^{-1/2} n^{-1/2} M^* B_n + 2k\left(1 - \Phi(B_n/k)\right).$$

Then, we have that

$$Pf_{c,t}^2 = O\left(n^{-1/2}(2\log n)^{1/2}\right) + O\left(n^{-1/2}\right), \text{ as } n \to \infty.$$

Then there exists a constant $M_1 > 0$, such that for $\delta_n^2 = M_1 n^{-1/2}(\log n)^{1/2}$, we have that $Pf_{c,t}^2 \le \delta_n^2$, for $f_{c,t} \in \mathcal{F}_n$. Choose β_n, an increasing sequence such that $\alpha_n = \beta_n n^{-1/4}(\log n)^{1/4}$ is a non-increasing sequence of positive real numbers. Then

$$\left(\log n\right)/n\delta_n^2\alpha_n^2 = \beta_n^{-2} \to 0, \text{ as } n \to \infty.$$

Then, by Theorem 2.11, we have

(4.37) $$\sup_{f_{c,t}\in\mathfrak{F}_n}\left|\left(P_n - P\right) f_{c,t}\right| = o\left(\beta_n n^{-3/4}\left(\log n\right)^{3/4}\right), \ w.p.1, \ n \to \infty,$$

this can be rewritten as (4.32),

$$\sup_{\substack{\|c-c_0\|\le Mn^{-1/2} \\ c\Sigma c'=1}}\ \sup_{t\in R}\left|F_n\left(t;c\right) - F_n\left(t;c_0\right)\right| = o\left(n^{-3/4}\left(\log n\right)^{3/4}\beta_n\right), \ w.p.1, \text{ as } n \to \infty.$$

Let

$$\Delta_n\left(c\right) = \left(n^{3/4}\left(\log n\right)^{-3/4}/\beta_n\right)\sup_{t\in R}\left|F_n\left(t;\, c\right) - F_n\left(t;\, c_0\right)\right|, \ n = 1, \ldots.$$

Then we need to show $\Delta_n\left(\hat{c}\right) = o_p\left(1\right)$, as $n \to \infty$. To do this, consider, for every $\delta > 0$ and $B > 0$,

(4.38) $$\begin{aligned} P\left(\Delta_n\left(\hat{c}\right) > \delta\right) &= P\left(\Delta_n\left(\hat{c}\right) > \delta; \left\|\hat{c} - c_0\right\| \le Bn^{-1/2}\right) + P\left(\Delta_n\left(\hat{c}\right) > \delta; \left\|\hat{c} - c_0\right\| > Bn^{-1/2}\right) \\ &\le P\left(\sup_{\|c-c_0\|\le Bn^{-1/2}} \Delta_n\left(c\right) > \delta\right) + P\left(\left\|\hat{c} - c_0\right\| > Bn^{-1/2}\right). \end{aligned}$$

Note that, since $\left\|\hat{c} - c_0\right\| = O_p\left(n^{-1/2}\right)$, for every $\varepsilon > 0$, there exists N_ε and B_ε such that

(4.39) $$P\left(\left\|\hat{c} - c_0\right\| > B_\varepsilon n^{-1/2}\right) < \varepsilon/2, \text{ for all } n > N_\varepsilon.$$

Now by (4.32), there exists $N_{\Delta,\varepsilon}$ such that

(4.40) $$P\left(\sup_{\|c-c_0\|\le B_\varepsilon n^{-1/2}} \Delta_n\left(c\right) > \delta\right) < \varepsilon/2, \text{ for all } n > N_{\Delta,\varepsilon}.$$

Combining (4.38), (4.39), and (4.40) we have, we have, for every $\delta > 0$, $\varepsilon > 0$, and $n > \max\left\{N_{\Delta,\varepsilon}, N_\varepsilon\right\}$,

(4.41) $$\begin{aligned} P\left(\left(n^{3/4}\left(\log n\right)^{-3/4}/\beta_n\right)\sup_{t\in R}\left|F_n\left(t;\, \hat{c}\right) - F_n\left(t;\, c_0\right)\right| > \delta\right) &= P\left(\Delta_n\left(\hat{c}\right) > \delta\right) \\ &\le P\left(\Delta_n\left(\hat{c}\right) > \delta\right) + P\left(\left\|\hat{c} - c_0\right\| > B_\varepsilon n^{-1/2}\right) \\ &\le \varepsilon/2 + \varepsilon/2 \\ &= \varepsilon. \end{aligned}$$

Therefore

$$\sup_t \left| F_n(t;\hat{c}) - F_n(t;c_0) \right| = o_p\left(n^{-3/4} (\log n)^{3/4} \beta_n \right), \text{ as } n \to \infty. \qquad \square$$

Corollary 4.10. *For* $Y_1, \ldots Y_n$ *, i.i.d. multivariate normal random vectors with* $\mu = 0$ *and positive definite covariance matrix* Σ. *Let* $F_n(t;c)$ *be defined as in (4.31).*

Assume $\left\| \hat{c} - c_0 \right\| = O_p\left(n^{-1/2} \right)$, *as* $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, *and* $c_0 \Sigma c_0{}' = 1$. *Then*

$$\sup_{t \in R} \left| F_n(t;\hat{c}) - F_n(t;c_0) \right| = o_p\left(n^{-3/4} \log n \right), \text{ as } n \to \infty.$$

Proof. Choose $\beta_n = (\log n)^{1/4}$ and apply Lemma 8. $\square$

Lemma 4.11. *For* $Y_1, \ldots Y_n$ *, i.i.d. multivariate normal random vectors with* $\mu = 0$ *and positive definite covariance matrix* Σ. *Assume* $\left\| \hat{c} - c_0 \right\| = O_p\left(n^{-1/2} \right)$, $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, *and* $c_0 \Sigma c_0{}' = 1$.

Then

(4.42) $$\max_{1 \le i \le n} \left| \hat{z}_i - z_i \right| = O_p\left((\log n)^{1/2} n^{-1/2} \right), \text{as } n \to \infty.$$

Proof. From (4.6) and (4.8), for $i = 1, \ldots, n$,

(4.43)
$$\begin{aligned} \left| \hat{z}_i - z_i \right| &= \left| (\hat{c} - c_0) Y_i \right| \\ &= \left| (\hat{c} - c_0) \Sigma^{1/2} \Sigma^{-1/2} Y_i \right| \\ &\le \left\| \hat{c} - c_0 \right\| \left\| \Sigma^{1/2} \right\| \left\| \Sigma^{-1/2} Y_i \right\|. \end{aligned}$$

Note that $\Sigma^{-1/2} Y_i$, $i = 1, \ldots, n$, are i.i.d. $MN(0, I_{n \times n})$ and that $\Sigma^{-1/2} Y_i = (z_{ij})_{k \times 1}$, where z_{ij} are i.i.d. $N(0,1)$, $j = 1, \ldots, k$ and $i = 1, \ldots, n.$ Then

$$\begin{aligned} \max_{1 \le i \le n} \left\| \Sigma^{-1/2} Y_i \right\| &= \max_{1 \le i \le n} \left(\sum_{j=1}^{k} z_{ij}^2 \right)^{1/2} \\ &\le \max_{1 \le i \le n} k^{1/2} \max_{1 \le j \le k} \left| z_{ij} \right| \\ &= k^{1/2} \max\left\{ z_{kn:kn}, - z_{1:kn} \right\}. \end{aligned}$$

By the symmetry of the standard normal distribution about zero, we have $z_{kn:kn} \overset{d}{=} -z_{1:kn}$.

Furthermore, by Serfling (1980, pg.91),

$$z_{kn:kn} \sim (2\log kn)^{1/2}, \text{ as } n \to \infty, \ w.p.1,$$

and since $(\log kn)/\log n \to 1$, as $n \to \infty$,

(4.44) $(2\log kn)^{-1/2} \max\{z_{kn:kn}, -z_{1:kn}\} \sim 1, \ w.p.1, \text{ as } n \to \infty.$

Combining (4.43) and (4.44) we have that

$$\begin{aligned}\max_{1\le i\le n} |\hat{z}_i - z_i| &\le \|\hat{c} - c_0\| \|\Sigma^{1/2}\| \|\Sigma^{-1/2} Y_i\| \\ &\le \|\hat{c} - c_0\| \|\Sigma^{1/2}\| k^{1/2} \max\{z_{kn:kn}, -z_{1:kn}\} \\ &= O_p(n^{-1/2}) o_p\left((\log n)^{1/2}\right), \text{ as } n \to \infty. \quad \square\end{aligned}$$

Lemma 4.12. *Let* a_1, a_2, b_1, *and* b_2 *be real numbers such that* $a_1 \le a_2$ *and* $b_1 \le b_2$. *Then*

$$\max\{|a_1 - b_1|, |a_2 - b_2|\} \le \max\{|a_1 - b_2|, |a_2 - b_1|\}.$$

Proof. Without loss of generality assume that

$$a_1 = \min\{a_1, a_2, b_1, b_2\}.$$

First assume $a_2 \le b_2$. Then either $a_1 \le a_2 \le b_1 \le b_2$ or $a_1 \le b_1 \le a_2 \le b_2$. Therefore $|a_1 - b_1| \le |a_1 - b_2|$ and $|a_2 - b_2| \le |a_1 - b_2|$. Secondly assume $b_2 \le a_2$. Then $a_1 \le b_1 \le b_2 \le a_2$ and $|a_1 - b_1| \le |a_1 - b_2|$ and $|a_2 - b_2| \le |b_1 - a_2|$. $\square$

Lemma 4.13. *Let* $x_1, \ldots, x_n$ *and* $y_1, \ldots, y_n$ *be any two sequences of real numbers. Then*

$$\max_{1\le i\le n} |x_{i:n} - y_{i:n}| \le \max_{1\le i\le n} |x_i - y_i|.$$

Proof. First we note that without loss of generality, we may assume that

(4.45) $x_1 \le x_2 \le \ldots \le x_n.$

Now we will use a proof by induction. Let $n = 1$, then

$$\max_{i=1} |x_{i:n} - y_{i:n}| = \max_{i=1} |x_i - y_i|.$$

Let $n = 2$, then, by Lemma 4.12,

$$\max_{1\le i\le 2}|x_{i:n}-y_{i:n}|\le \max_{1\le i\le 2}|x_i-y_i|.$$

Next assume that the result is true for all integers less than $n+1$ and we will prove

$$\max_{1\le i\le n+1}|x_{i:n}-y_{i:n}|\le \max_{1\le i\le n+1}|x_i-y_i|.$$

We will consider two cases. For case 1, assume that $y_{n+1:n+1}=y_{n+1}$. Then

$$\max_{1\le i\le n+1}|x_i-y_i| = \max\left\{\max_{1\le i\le n}|x_i-y_i|,\ |x_{n+1}-y_{n+1}|\right\}.$$

By the assumption of the induction hypothesis we have that

$$\begin{aligned}\max_{1\le i\le n+1}|x_i-y_i| &\ge \max\left\{\max_{1\le i\le n}|x_i-y_{i:n}|,\ |x_{n+1:n+1}-y_{n+1:n+1}|\right\}\\ &= \max_{1\le i\le n+1}|x_{i:n+1}-y_{i:n+1}|.\end{aligned}$$

For case 2, assume $y_{n+1:n+1}=y_j$, for some $j<n+1$. Then

$$\max_{1\le i\le n+1}|x_i-y_i| = \max\left\{\max_{\substack{1\le i\le n\\ i\ne j}}|x_i-y_i|,\ \max\left\{|x_j-y_j|,|x_{n+1}-y_{n+1}|\right\}\right\}.$$

Note that $y_{n+1}\le y_j$ and $x_j\le x_{n+1}$. Then, by Lemma 4.12,

$$\begin{aligned}\max_{1\le i\le n+1}|x_i-y_i| &\ge \max\left\{\max_{\substack{1\le i\le n\\ i\ne j}}|x_i-y_i|,\ \max\left\{|x_j-y_{n+1}|,\ |x_{n+1}-y_{n+1:n+1}|\right\}\right\}\\ &= \max\left\{\max_{\substack{1\le i\le n\\ i\ne j}}|x_i-y_i|,\ |x_j-y_{n+1}|,\ |x_{n+1}-y_{n+1:n+1}|\right\}.\end{aligned}$$

Note that by pairing x_i with y_i, for $i\ne n+1$ or j, and x_j with y_{n+1}, we can apply the induction hypothesis for this set of n pairs of observations to get

$$\max\left\{\max_{\substack{1\le i\le n\\ i\ne j}}|x_i-y_i|,\ |x_j-y_{n+1}|\right\}\ge \max_{1\le i\le n}|x_{i:n+1}-y_{i:n+1}|.$$

Therefore,

$$\begin{aligned}\max_{1\le i\le n+1}|x_i-y_i| &\ge \max\left\{\max\left\{\max_{\substack{1\le i\le n\\ i\ne j}}|x_i-y_i|,\ |x_j-y_{n+1}|\right\},\ |x_{n+1:n+1}-y_{n+1:n+1}|\right\}\\ &\ge \max\left\{\max_{1\le i\le n}|x_{i:n+1}-y_{i:n+1}|,|x_{n+1:n+1}-y_{n+1:n+1}|\right\}\\ &= \max_{1\le i\le n+1}|x_{i:n+1}-y_{i:n+1}|.\end{aligned}$$

This completes the proof by induction. □

Lemma 4.14. *Let* $Y_1,\ldots,Y_n$ *i.i.d. multivariate normal random vectors with* $\mu = 0$ *and positive definite covariance matrix* Σ. *Assume* $\left\|\hat{c} - c_0\right\| = O_p\left(n^{-1/2}\right)$, $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, *and* $c_0\Sigma c_0{}' = 1$. *Then*

(4.46) $\max_{1\le i\le n}\left|\hat{z}_{i:n} - z_{i:n}\right| = O_p\left(\left(\log n\right)^{1/2} n^{-1/2}\right)$, *as* $n \to \infty$.

Proof. By Lemma 4.11 and Lemma 4.13, we have

$$\begin{aligned}\max_{1\le i\le n}\left|\hat{z}_{i:n} - z_{i:n}\right| &\le \max_{1\le i\le n}\left|\hat{z}_i - z_i\right| \\ &= O_p\left(\left(\log n\right)^{1/2} n^{-1/2}\right), \text{ as } n \to \infty.\end{aligned}$$

□

Lemma 4.15. *Let* $z_{i:n}$ *and* $\hat{z}_{i:n}$ *be defined as in (4.6) and (4.8) respectively. Assume* $\left\|\hat{c} - c_0\right\| = O_p\left(n^{-1/2}\right)$, $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, *and* $c_0\Sigma c_0{}' = 1$. *Then*

(4.47) $\max_{1\le i\le n}\left|\Phi\left(z_{i:n}\right) - \Phi\left(\hat{z}_{i:n}\right)\right| = o_p\left(n^{-3/4}\log n\right)$, *as* $n \to \infty$.

Proof. Let $D_n = n^{3/4}\left(\log n\right)^{-1}\max_{1\le i\le n}\left|\Phi\left(z_{i:n}\right) - \Phi\left(\hat{z}_{i:n}\right)\right|$. Then, for every $\varepsilon > 0$, there exists $M_\varepsilon > 0$ and N_ε such that

$$P\left(\max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| > n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon\right) < \varepsilon/2, \text{ for all } n > N_\varepsilon,$$

we have

$$\begin{aligned}P\left(D_n > \varepsilon\right) &= P\left(D_n > \varepsilon; \max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| \le n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon\right) \\ &\quad + P\left(D_n > \varepsilon; \max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| > n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon\right) \\ &\le P\left(D_n > \varepsilon; \max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| \le n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon\right) \\ &\quad + P\left(\max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| > n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon\right).\end{aligned}$$

Note that $F_n\left(\hat{z}_{i:n};\hat{c}\right) = F_n\left(z_{i:n};c_0\right) = \dfrac{i}{n}$. Then, for $\max_{1\le i\le n}\left|z_{i:n} - \hat{z}_{i:n}\right| \le n^{-1/2}\left(\log n\right)^{1/2} M_\varepsilon$,

$$\max_{1\le i\le n}\left|\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right| = \max_{1\le i\le n}\left|F_n\left(\hat{z}_{i:n};\hat{c}\right)-F_n\left(\hat{z}_{i:n};c_0\right)+F_n\left(\hat{z}_{i:n};c_0\right)\right.$$
$$\left.-F_n\left(z_{i:n};c_0\right)+\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right|$$
$$\le \max_{1\le i\le n}\left|F_n\left(\hat{z}_{i:n};\hat{c}\right)-F_n\left(\hat{z}_{i:n};c_0\right)\right|$$
$$+\max_{1\le i\le n}\left|F_n\left(\hat{z}_{i:n};c_0\right)-F_n\left(z_{i:n};c_0\right)+\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right|$$
$$\le \sup_t\left|F_n\left(t;\hat{c}\right)-F_n\left(t;c_0\right)\right|$$
$$+\sup_{|s-t|<M_\varepsilon n^{-1/2}\log n^{1/2}}\left|F_n\left(t;c_0\right)-F_n\left(s;c_0\right)+\Phi\left(s\right)-\Phi\left(t\right)\right|.$$

Now apply Corollary 2.13 and Corollary 4.10 to complete the proof

$$P\left(D_n>\varepsilon;\max_{1\le i\le n}\left|z_{i:n}-\hat{z}_{i:n}\right|\le n^{-1/2}\left(\log n\right)^{1/2}M_\varepsilon\right)\le P\left(\sup_t\frac{\left|F_n\left(t;\hat{c}\right)-F_n\left(t;c_0\right)\right|}{n^{-3/4}\log n}>\frac{\varepsilon}{2}\right)$$
$$+P\left(\sup_{|s-t|<M_\varepsilon n^{-1/2}\log n^{1/2}}\frac{\left|F_n\left(t;c_0\right)-F_n\left(s;c_0\right)+\Phi\left(s\right)-\Phi\left(t\right)\right|}{n^{-3/4}\log n}>\frac{\varepsilon}{2}\right)\to 0,\text{ as } n\to\infty.\quad\square$$

Lemma 4.16. *Let* $z_{i:n}$ *and* $\hat{z}_{i:n}$ *be defined as in (4.6) and (4.8), respectively. Assume* $\left\|\hat{c}-c_0\right\|=O_p\left(n^{-1/2}\right)$, $n\to\infty$, $\hat{c}\Sigma\hat{c}'=1$, *and* $c_0\Sigma c_0{}'=1$. *Let* $\left\{i_n\right\}_{n=1}^{\infty}$ *be a sequence of positive integers such that* $n^{1/4}i_n^{-1}\log n=O(1)$ *and* $i_n n^{-1}=o(1)$, *as* $n\to\infty$. *Then*

$$\max_{i_n\le i\le n-i_n}\left|z_{i:n}-\hat{z}_{i:n}\right|=o_p\left(i_n^{-1}n^{1/4}\log n\right),\ \textit{as}\ n\to\infty.$$

Proof. First, we apply a first order Taylor series expansion to $\left|\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right|$ to get

(4.48) $\left|z_{i:n}-\hat{z}_{i:n}\right|=\left|\left(\Phi\left(z_{i:n}\right)-\Phi\left(\hat{z}_{i:n}\right)\right)\big/\phi(z_i^*)\right|,$

where z_i^* is between $\hat{z}_{i:n}$ and $z_{i:n}$. Choose p, $0<p<1/2$, then, by Theorem 2.3.1, Serfling (1980),

(4.49) $F_n^{-1}\left(p;c_0\right)\to\Phi^{-1}\left(p\right)<0,\ w.p.1,$ as $n\to\infty.$

Therefore, for $i/n\le p$,

$$z_{i:n}=F_n^{-1}\left(i/n;c_0\right)$$
(4.50) $$\le F_n^{-1}\left(p;c_0\right)$$
$$=\Phi^{-1}\left(p\right)+o(1),\ w.p.1,\text{ as } n\to\infty.$$

Since $\max_{1\le i\le n}\left|z_{i:n}-\hat{z}_{i:n}\right|=o_p(1),$ as $n\to\infty$, and $\Phi^{-1}(p)<0,$ we have that , given $\varepsilon>0,$ there exists N_ε, such that

(4.51) $P\left(\hat{z}_{i:n}\ge 0; i/n\le p\right)<\varepsilon, n\ge N_\varepsilon.$

Next we will make use of an inequality from Royden (1968), for a convex function $\varphi(t)$ on an interval $(a,b),$ for points x, y, x', y' of (a,b) such that $x\le x'<y$ and $x\le y<y'$

(4.52) $\dfrac{\varphi(y)-\varphi(x)}{y-x}\le\dfrac{\varphi(y')-\varphi(x')}{y'-x'}.$

First, we note that since $\Phi(t)$ is monotonically increasing and

$$\left|\left(\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})\right)/\left(z_{i:n}-\hat{z}_{i:n}\right)\right|=\frac{\Phi\left(\max\{z_{i:n},\hat{z}_{i:n}\}\right)-\Phi\left(\min\{z_{i:n},\hat{z}_{i:n}\}\right)}{\max\{z_{i:n},\hat{z}_{i:n}\}-\min\{z_{i:n},\hat{z}_{i:n}\}}.$$

Therefore, for $z_{i:n}<0$ and $\hat{z}_{i:n}<0$, noting $\Phi(t)$ is convex for $t\in(-\infty,0)$,

$\min\{z_{i-1:n},\hat{z}_{i-1:n}\}\le\min\{z_{i:n},\hat{z}_{i:n}\}\le\max\{z_{i:n},\hat{z}_{i:n}\},$ and

$\min\{z_{i-1:n},\hat{z}_{i-1:n}\}\le\max\{z_{i-1:n},\hat{z}_{i-1:n}\}\le\max\{z_{i:n},\hat{z}_{i:n}\},$

$$\left|\frac{\Phi(z_{i-1:n})-\Phi(\hat{z}_{i-1:n})}{z_{i-1:n}-\hat{z}_{i-1:n}}\right|\le\left|\frac{\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})}{z_{i:n}-\hat{z}_{i:n}}\right|.$$

This is implies, for $z_{i-1:n}<z_{i:n}<0$ and $\hat{z}_{i-1:n}<\hat{z}_{i:n}<0,$

$$\begin{aligned}\phi\left(z_{i-1}^*\right)&=\left|\frac{\Phi(z_{i-1:n})-\Phi(\hat{z}_{i-1:n})}{z_{i-1:n}-\hat{z}_{i-1:n}}\right|\\ &\le\left|\frac{\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})}{z_{i:n}-\hat{z}_{i:n}}\right| \qquad (4.53)\\ &=\phi\left(z_i^*\right).\end{aligned}$$

Now, by (4.48) and (4.53), we have, for $z_{[np]:n}\le 0$ and $\hat{z}_{[np]:n}\le 0,$

$$\begin{aligned}\max_{i_n\le i\le[np]}\left|z_{i:n}-\hat{z}_{i:n}\right|&=\max_{i_n\le i\le[np]}\left|\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})\right|/\phi(z_i^*) \qquad (4.54)\\ &\le\max_{i_n\le i\le[np]}\left|\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})\right|/\phi(z_{i_n}^*).\end{aligned}$$

Let $R_{1,n}=\max_{i_n\le i\le[np]}\left|\Phi(z_{i:n})-\Phi(\hat{z}_{i:n})\right|$ and apply Mill's Ratio for $t<0,$ $|t|\Phi(t)\le\phi(t),$ to get

(4.55) $$\max_{i_n \le i \le [np]} \left| z_{i:n} - \hat{z}_{i:n} \right| \le R_{1,n} \Big/ \left(\left| z_{i_n}^* \right| \Phi\left(z_{i_n}^* \right) \right).$$

Let

$$A_n = \Phi\left(z_{i_n}^* \right) - \Phi\left(z_{i_n} \right)$$

and

$$B_n = \Phi\left(z_{i_n:n} \right) - i_n / n,$$

then

(4.56) $$\Phi\left(z_{i_n}^* \right) = i_n / n + A_n + B_n.$$

Applying Lemma 4.15 to A_n, we have

(4.57) $$\begin{aligned} \left| A_n \right| &= \left| \Phi\left(z_{i_n}^* \right) - \Phi\left(z_{i_n:n} \right) \right| \\ &\le \left| \Phi\left(\hat{z}_{i_n:n} \right) - \Phi\left(z_{i_n:n} \right) \right| \\ &= o_p\left((\log n) n^{-3/4} \right), \text{ as } n \to \infty. \end{aligned}$$

There exists U_i, $i = 1, \ldots, n$, such that U_i are i.i.d. random variables with a uniform distribution on $[0,1]$ and $U_{i:n} = \Phi(z_{i:n})$; $i = 1, \ldots, n$. Therefore, we can rewrite

(4.58) $$\begin{aligned} B_n &= U_{i_n:n} - i_n / n \\ &= i_n / n - G_n\left(i_n / n \right) + R_{2,n}\left(i_n / n \right), \end{aligned}$$

where $G_n(t) = n^{-1} \sum_{i=1}^{n} I\left(U_i \le t \right)$ and $R_{2,n}(t)$ is the remainder term from the Bahadur Representation Theorem (1966). Then we have, by Kiefer (1967),

(4.59) $$\sup_{t \in (0,1)} \left| R_{2,n}(t) \right| = O_p\left(n^{-3/4} (\log n)^{1/2} \right), \text{ as } n \to \infty.$$

Next, note that $nG_n(t)$ has a binomial distribution with n trials and probability of success t. Therefore,

(4.60) $$\begin{aligned} \operatorname{var}\left(n^{1/2} \left(t - G_n(t) \right) / t^{1/2} \right) &= t(1-t)/t \\ &= (1-t) \\ &\le 1. \end{aligned}$$

Let $t = i_n/n$ and we have $\operatorname{var}\left(n^{1/2}\left(i_n/n - G_n\left(i_n/n\right)\right)\Big/\left(i_n/n\right)^{1/2}\right) = 1 - \left(i_n/n\right) \to 1$, as $n \to \infty$, which implies

(4.61) $\quad i_n/n - G_n\left(i_n/n\right) = O_p\left(i_n^{1/2}n^{-1}\right)$, as $n \to \infty$.

Consider

$$(4.62) \quad \begin{aligned} R_{1,n}\Big/\left(\left|z_{i_n}^*\right|\Phi\left(z_{i_n}^*\right)\right) &= R_{1,n}\left|z_{i_n}^*\right|^{-1}\left(i_n/n + A_n + B_n\right)^{-1} \\ &= \left(n/i_n\right)R_{1,n}\left|z_{i_n}^*\right|^{-1}\left(1 + \left(n/i_n\right)A_n + \left(n/i_n\right)B_n\right)^{-1}. \end{aligned}$$

By Lemma 4.15, we have

$$(4.63) \quad \begin{aligned} \left(n/i_n\right)R_{1,n} &= \left(n/i_n\right)\left(\log n/\log n\right)R_{1,n} \\ &= \left(n^{1/4}\left(\log n\right)/i_n\right)\left(n^{3/4}/\left(\log n\right)\right)o_p\left(n^{-3/4}\left(\log n\right)\right) \\ &= o_p\left(n^{1/4}\left(\log n\right)/i_n\right), \text{ as } n \to \infty. \end{aligned}$$

Now by (4.57), we have

$$(4.64) \quad \begin{aligned} \left(n/i_n\right)A_n &= \left(n/i_n\right)\left(\log n/\log n\right)A_n \\ &= \left(n^{1/4}\left(\log n\right)i_n^{-1}\right)\left(n^{3/4}/\left(\log n\right)\right)o_p\left(n^{-3/4}\log n\right) \\ &= o_p\left(n^{1/4}\left(\log n\right)i_n^{-1}\right), \text{ as } n \to \infty. \end{aligned}$$

By (4.58), (4.59), and (4.61),

$$(4.65) \quad \begin{aligned} \left(n/i_n\right)B_n &= \left(n/i_n\right)O_p\left(i_n^{1/2}n^{-1}\right) + \left(n/i_n\right)O_p\left(n^{-3/4}\left(\log n\right)^{1/2}\right) \\ &= i_n^{-1/2}O_p\left(1\right) + O_p\left(n^{-1/4}i_n^{-1}\left(\log n\right)^{1/2}\right) \\ &= o_p\left(1\right) + O_p\left(n^{-1/4}i_n^{-1}\left(\log n\right)^{1/2}\right), \text{ as } n \to \infty. \end{aligned}$$

Since $n^{-1/4}\log n\, i_n^{-1} = O\left(1\right)$, as $n \to \infty$, $\left(n/i_n\right)A_n = o_p\left(1\right)$ and $\left(n/i_n\right)B_n = o_p\left(1\right)$, as $n \to \infty$, we have

(4.66) $\quad \left(1 + \left(n/i_n\right)A_n + \left(n/i_n\right)B_n\right)^{-1} = O_p\left(1\right)$, as $n \to \infty$.

Therefore, noting that $\left|z_{i_n}^*\right|^{-1} = o_p\left(1\right)$, as $n \to \infty$, by (4.63), we have

$$(4.67) \quad \begin{aligned} R_{1,n}\Big/\left(\left|z_{i_n}^*\right|\Phi\left(z_{i_n}^*\right)\right) &= o_p\left(n^{1/4}\log n\, i_n^{-1}\right)o_p\left(1\right)O_p\left(1\right) \\ &= o_p\left(n^{1/4}\log n\, i_n^{-1}\right), \text{ as } n \to \infty. \end{aligned}$$

Let $D_n = \max_{i_n \le i \le [np]}\left|z_{i:n} - \hat{z}_{i:n}\right|$ and, for $\varepsilon > 0$, consider

$$P\left(i_n D_n / n^{1/4}(\log n) > \varepsilon\right) = P\left(i_n D_n / n^{1/4}(\log n) > \varepsilon; z_{i:n} < 0; \hat{z}_{i:n} < 0; i/n < p\right)$$

$$(4.68) \qquad + P\left(i_n D_n / n^{1/4}(\log n) > \varepsilon; \{z_{i:n} > 0 \text{ or } \hat{z}_{i:n} > 0\}; i/n < p\right)$$

$$\le P\left(i_n D_n / n^{1/4}(\log n) > \varepsilon; z_{i:n} < 0; \hat{z}_{i:n} < 0; i/n < p\right)$$

$$+ P\left(z_{i:n} > 0; i/n < p\right) + P\left(\hat{z}_{i:n} > 0; i/n < p\right).$$

By (4.54) and (4.67), we have

$$P\left(i_n D_n / n^{1/4}(\log n) > \varepsilon; z_{i:n} < 0; \hat{z}_{i:n} < 0; i/n < p\right)$$

$$(4.69) \qquad \le P\left(\left[i_n R_{1,n} \Big/ \left(\left|z_{i_n}^*\right| \Phi\left(z_{i_n}^*\right)\right) n^{1/4}(\log n)\right] > \varepsilon\right) \to 0, \text{ as n} \to \infty.$$

By (4.50) and (4.51), we have

(4.70) $P\left(z_{i:n} > 0; i/n < p\right) + P\left(\hat{z}_{i:n} > 0; i/n < p\right) \to 0$, as $n \to \infty$.

Combining (4.68), (4.69), and (4.70), we have

(4.71) $\max\limits_{i_n \le i \le [np]} \left|z_{i:n} - \hat{z}_{i:n}\right| = o_p\left(i_n^{-1} n^{1/4} \log n\right),$ as $n \to \infty$.

Next we note that $\Phi(t)$ is concave for $t \in [0, \infty)$. Therefore, we can use a similar argument, where $\phi\left(z_i^*\right) \ge \phi\left(z_{i+1}^*\right),$ for $z_i^* > 0$ and $z_{i+1}^* > 0$, and for $t > 0$, $t\left(1 - \Phi(t)\right) \le \phi(t),$ to get

(4.72) $\max\limits_{[n(1-p)] \le i \le n - i_n} \left|z_{i:n} - \hat{z}_{i:n}\right| = o_p\left(i_n^{-1} n^{1/4} \log n\right),$ as $n \to \infty$.

Finally, choose q, such that $0 < q < p$. Then by Theorem 2.3.1, Serfling(1980), for $p \le i/n \le 1 - p,$

(4.73) $\Phi^{-1}(q) < F_n^{-1}(p; c_0) \le z_{i:n} \le F_n^{-1}(1 - p; c_0) < \Phi^{-1}(1 - q),$ $w.p.1,$ as $n \to \infty$.

Since $\max\limits_{1 \le i \le n} \left|z_{i:n} - \hat{z}_{i:n}\right| = o_p(1),$ as $n \to \infty$, we have

(4.74) $P\left(\Phi^{-1}(q) < \hat{z}_{i:n} < \Phi^{-1}(1 - q);\ p \le i/n \le 1 - p\right) \to 1$, as $n \to \infty$.

For $\Phi^{-1}(q) < z_{[np]:n} \le z_{[n(1-p)]:n} < \Phi^{-1}(1 - q)$ and $\Phi^{-1}(q) < \hat{z}_{[np]:n} \le \hat{z}_{[n(1-p)]:n} < \Phi^{-1}(1 - q),$ we have

$$(4.75) \qquad \max_{[np] \le i \le [n(1-p)]} \left|z_{i:n} - \hat{z}_{i:n}\right| = \max_{[np] \le i \le [n(1-p)]} \left|\Phi\left(z_{i:n}\right) - \Phi\left(\hat{z}_{i:n}\right)\right| \Big/ \phi\left(z_i^*\right)$$

$$\le \max_{[np] \le i \le [n(1-p)]} \left|\Phi\left(z_{i:n}\right) - \Phi\left(\hat{z}_{i:n}\right)\right| \Big/ \phi\left(\Phi^{-1}(q)\right).$$

Combining (4.73), (4.74), (4.75), and applying Lemma 4.13 to the numerator of (4.75), we have

(4.76) $\max\limits_{[np] \le i \le [n(1-p)]} \left|z_{i:n} - \hat{z}_{i:n}\right| = o_p\left(n^{-3/4} \log n\right)$, as $n \to \infty$.

Finally, combine (4.71), (4.72), and (4.76) to get

$$\max_{i_n \le i \le n-i_n} \left| z_{i:n} - \hat{z}_{i:n} \right| = \left\{ \max_{[np] \le i \le [n(1-p)]} \left| z_{i:n} - \hat{z}_{i:n} \right|, \max_{[n(1-p)] \le i \le n-i_n} \left| z_{i:n} - \hat{z}_{i:n} \right|, \max_{i_n \le i \le [np]} \left| z_{i:n} - \hat{z}_{i:n} \right| \right\}$$
$$= o_p \left(i_n^{-1} n^{1/4} \log n \right), \text{ as } n \to \infty. \qquad \square$$

4.3. **Rates for B_1 and B_2.** In this section, we consider

$$B_1 = n^{-1} \sum_{i=1}^{n} \left(\hat{z}_{i:n} - z_{i:n} \right)^2$$

and

$$B_2 = n^{-1} \sum_{i=1}^{n} \left(z_{i:n} - \hat{z}_{i:n} \right) \left(z_{i:n} - \xi_i \right).$$

Using lemmas 4.11 and 4.16, we prove Lemma 4.17, that states

$$B_1 = o_p \left(n^{-7/6} \left(\log n \right)^2 \right), \text{ as } n \to \infty,$$

and Corollary 4.18, that states

$$B_2 = o_p \left(n^{-13/12} \log n \left(\log \log n \right)^{1/2} \right), \text{ as } n \to \infty.$$

As noted in Section 4.1, Lemma 4.17 and it's corollary will imply the main results presented in Section 4.4.

Lemma 4.17. *Let $z_{i:n}$ and $\hat{z}_{i:n}$ be defined as in (4.6) and (4.8), respectively. Assume $\left\| \hat{c} - c_0 \right\| = O_p \left(n^{-1/2} \right)$, $n \to \infty$, $\hat{c} \Sigma \hat{c}' = 1$, and $c_0 \Sigma c_0{}' = 1$. Let $B_1 = n^{-1} \sum_{i=1}^{n} \left(\hat{z}_{i:n} - z_{i:n} \right)^2$. Then*

$$B_1 = o_p \left(n^{-7/6} \left(\log n \right)^2 \right), \text{ as } n \to \infty.$$

Proof. Consider

$$B_1 = n^{-1} \sum_{i=1}^{n} \left(z_{i:n} - \hat{z}_{i:n} \right)^2$$

(4.77)
$$= n^{-1} \left[\sum_{i \le i_n} \left(z_{i:n} - \hat{z}_{i:n} \right)^2 + \sum_{i \ge n-i_n} \left(z_{i:n} - \hat{z}_{i:n} \right)^2 + \sum_{i_n < i < n-i_n} \left(z_{i:n} - \hat{z}_{i:n} \right)^2 \right]$$
$$\le n^{-1} \left[2 i_n \left(\max_{1 \le i \le n} \left| z_{i:n} - \hat{z}_{i:n} \right| \right)^2 + (n - 2 i_n) \left(\max_{i_n \le i \le n-i_n} \left| z_{i:n} - \hat{z}_{i:n} \right| \right)^2 \right].$$

We will apply Lemma 4.14 to the first term of (4.77) and letting $i_n = n^{3/4+\delta}$, $0 < \delta < 1/4$, to get

$$n^{3/4+\delta}\left(\max_{1\le i\le n}\left|z_{i:n}-\hat{z}_{i:n}\right|\right)^2 = n^{3/4+\delta}n^{-1}(\log n)\left[n^{1/2}(\log n)^{-1/2}\max_{1\le i\le n}\left|z_{i:n}-\hat{z}_{i:n}\right|\right]^2$$
$$= n^{\delta-1/4}(\log n)O_p(1),\ \text{as } n\to\infty. \tag{4.78}$$

Next, we apply Lemma 4.16, with $i_n = n^{3/4+\delta}$, $0<\delta<1/4$. Then

$$\left(n-2n^{3/4+\delta}\right)\left(\max_{n^{3/4+\delta}\le i\le n-n^{3/4+\delta}}\left|z_{i:n}-\hat{z}_{i:n}\right|\right)^2$$
$$=\left(n-2n^{3/4+\delta}\right)n^{-(1+2\delta)}\left(\log(n)\right)^2\left(n^{1/2+\delta}\left(\log(n)\right)^{-1}\max_{n^{3/4+\delta}\le i\le n-n^{3/4+\delta}}\left|z_{i:n}-\hat{z}_{i:n}\right|\right)^2 \tag{4.79}$$
$$= n^{-2\delta}(\log n)^2 o_p(1),\ \text{as } n\to\infty.$$

Let $\delta = 1/12$. Then, by (4.77), (4.78), and (4.79),

$$B_1 = n^{-1}\left[O_p\left(n^{-1/6}(\log n)\right)+o_p\left(n^{-1/6}(\log n)^2\right)\right]$$
$$= n^{-1}o_p\left(n^{-1/6}(\log n)^2\right) \tag{4.80}$$
$$= o_p\left(n^{-7/6}(\log n)^2\right),\ \text{as } n\to\infty. \qquad \square$$

Corollary 4.18. *Let* $z_{i:n}$ *and* $\hat{z}_{i:n}$ *be defined as in (4.6) and (4.8), respectively. Assume* $\left\|\hat{c}-c_0\right\| = O_p\left(n^{-1/2}\right)$, $n\to\infty$, $\hat{c}\Sigma\hat{c}'=1$, *and* $c_0\Sigma c_0{}'=1$. *Let* $B_2 = n^{-1}\sum_{i=1}^{n}\left(z_{i:n}-\hat{z}_{i:n}\right)\left(z_{i:n}-\xi_i\right)$.

Then

$$B_2 = o_p\left(n^{-13/12}\log n\left(\log\log n\right)^{1/2}\right),\ \text{as } n\to\infty.$$

Proof. By (4.28) we have

$$\left|B_2\right| \le \left(B_1 n^{-1}\sum_{i=1}^{n}\left(z_{i:n}-\xi_i\right)^2\right)^{1/2}. \tag{4.81}$$

Apply Theorem 4.8 and Lemma 4.18 to complete the proof,

$$\left|B_2\right| = o_p\left(n^{-7/12}\log n\right)O_p\left(n^{-1/2}\left(\log\log n\right)^{1/2}\right)$$
$$= o_p\left(n^{-13/12}\log n\left(\log\log n\right)^{1/2}\right),\ \text{as } n\to\infty. \qquad \square$$

4.4 <u>Asymptotic Equivalence for Correlation Statistics.</u> In this section, we derive the limiting distribution of correlation statistics based on projections of multivariate normal random vectors from a data suggested linear transformation under the assumption of normality of the original observations, and that the data suggested linear transformation satisfies condition Assumptions 4.1. Theorem 4.19, the main result of this chapter and section, gives the asymptotic distribution of the de Wet and Venter Statistic applied to the projections from the estimated linear transformation. Theorem 4.21 extends Theorem 4.19 to the other correlation goodness-of-fit statistics considered in Verril and Johnson (1981) and Corollary 4.22 specifically covers the Shapiro-Wilk statistic.

Theorem 4.19. *Let $Y_1,\ldots,Y_n$ be i.i.d. k-variate multivariate normal vectors with mean, μ, and positive definite covariance matrix, $\Sigma_{k\times k}$. Let $\hat{d}' \in R^k$ be a random vector satisfying Assumption 4.1. Let $\hat{W}_n$ and W_n be defined as in (4.2) and (4.10). Then*

$$i) \qquad n\left(\hat{W}_n - W_n\right) = o_p(1), \quad n \to \infty,$$

and

$$ii) \qquad n\left(1-\hat{W}_n\right) - a_n \rightsquigarrow \sum_{j=3}^{\infty} \left(x_j - 1\right)\Big/ j, \quad as\ n \to \infty.$$

Proof. By Lemma 4.7, Lemma 4.17, Corollary 4.18, we have

(4.82) $T_n - \hat{T}_n = \frac{\left(s_n^2 - \hat{s}_n^2\right)}{2} + o_p\left(n^{-13/12} \log n \left(\log\log n\right)^{1/2}\right)$, as $n \to \infty$.

Since $\hat{T}_n + T_n \xrightarrow{P} 2$, as $n \to \infty$, $\hat{c}\Sigma\hat{c}' = 1$, and $c_0 \Sigma c_0{}' = 1$, by (4.24) we have

$$\hat{U}_n - U_n = n\left(\hat{s}_n^2 - s_n^2\right) - n\left(\hat{T}_n + T_n\right)\left(\hat{T}_n - T_n\right)$$

$$= n\left(\hat{s}_n^2 - s_n^2\right) - \left(\hat{T}_n + T_n\right) n\left(\frac{\left(s_n^2 - \hat{s}_n^2\right)}{2} + o_p\left(n^{-1}\right)\right)$$

$$= n\left(\hat{s}_n^2 - s_n^2\right)\left[1 - \frac{\left(\hat{T}_n + T_n\right)}{2}\right] + \left(\hat{T}_n + T_n\right) o_p(1)$$

$$= n\left(\hat{s}_n^2 - s_n^2\right) o_p(1) + o_p(1) \tag{4.83}$$

$$= n\left(\hat{c} - c_0\right)\left(\hat{\Sigma} - \Sigma\right)\left(\hat{c} + c_0\right)' o_p(1) + o_p(1)$$

$$= nO_p\left(n^{-1}\right) o_p(1) + o_p(1)$$

$$= o_p(1), \text{ as } n \to \infty.$$

By Corollary 4.6, (4.13), and (4.83), we have

$$n\left(\hat{W}_n - W_n\right) = o_p(1), \text{as } n \to \infty.$$

Furthermore, by Slutsky's Theorem,

$$n\left(1 - \hat{W}_n\right) - a_n = n\left(1 - \hat{W}_n\right) - a_n + n\left(W_n - \hat{W}_n\right)$$

$$= n\left(1 - \hat{W}_n\right) - a_n + o_p(1)$$

$$\rightsquigarrow \sum_{j=3}^{\infty}\left(X_j - 1\right)\Big/ j, \text{ as } n \to \infty. \qquad \square$$

Now we will consider the Shapiro-Wilk statistic and the related statistics considered in Verrill and Johnson (1987). First, we state some results we will need from this paper. Define Ψ to be a vector of length n such that

$$\sum_{i=1}^{n}\left(\Psi_i - \xi_i\right)^2 = o\left(\left(\log\log n\right)^{-1}\right), \text{ as } n \to \infty, \tag{4.84}$$

where ξ is defined in (4.3). For ease of reference, we state the complete sample version of Verrill and Johnson's (1987) Theorem 3.1 as Theorem 4.20.

Theorem 4.20. (Verrill and Johnson, 1987, Theorem 3.1) *Let $y_1, \ldots, y_n$ be i.i.d. normal random variables. Let ξ be defined as in (4.3), and Ψ be a vector of constants that satisfies (4.84). Then*

$$n\left[r\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right) - r\left(\left(y_{i:n}\right)_{1\times n}, \xi\right)\right] = o_p(1), \text{ as } n \to \infty.$$

Let

(4.85) $\alpha = \frac{\xi}{\|\xi\|}$,

and

(4.86) $\beta_{\Psi} = \left(\Psi - n^{-1} \sum_{i=1}^{n} \Psi_i 1_{1\times n} \right) \Big/ \left\| \Psi - n^{-1} \sum_{i=1}^{n} \Psi_i 1_{1\times n} \right\|$.

In the proof of Verrill and Johnson's Theorem 3.1 the following two results were shown

(4.87) $\|\alpha - \beta_{\Psi}\|^2 = o\left(n^{-1} (\log \log n)^{-1} \right)$, as $n \to \infty$,

and

(4.88) $(\alpha - \beta_{\Psi})' \xi = o\left(n^{-1/2} \right)$, as $n \to \infty$.

Theorem 4.21. *Let* $Y_1, \ldots, Y_n$ *be i.i.d. k-variate multivariate normal vectors with mean,* $\underset{\sim}{\mu}$ *, and positive definite covariance matrix,* $\Sigma_{k\times k}$ *. Let* $\hat{d}$ *be a random vector and* d_0 *be a fixed vector which jointly satisfy Assumption 4.1. Let* $(\hat{y}_{i:n})_{1\times n}$ *and* $(y_{i:n})_{1\times n}$ *be the vectors of univariate order statistics based on* $\left\{ \hat{y}_i = \hat{d}\left(Y_i - \overline{Y} \right) \right\}_{i=1}^{n}$ *and* $\left\{ y_i = d_0 \left(Y_i - \mu \right) \right\}_{i=1}^{n}$*, respectively. Let* ξ *be defined as in (4.3), and* Ψ *be a vector of constants that satisfies (4.84). Then*

$$n\left[r^2\left((\hat{y}_{i:n})_{1\times n}, \Psi \right) - r^2\left((\hat{y}_{i:n})_{1\times n}, \xi \right) \right] = o_p(1), \text{ as } n \to \infty,$$

and

$$n\left[r^2\left((\hat{y}_{i:n})_{1\times n}, \Psi \right) - r^2\left((y_{i:n})_{1\times n}, \Psi \right) \right] = o_p(1), \text{ as } n \to \infty.$$

Proof. Let $\hat{Z}$ and Z be defined as in (4.7) and (4.9), respectively. Then, by the location-scale invariance of $r^2(\bullet,\bullet)$,

$$r^2\left((\hat{y}_{i:n})_{1\times n}, \Psi \right) = r^2\left(\hat{Z}, \Psi \right),$$

$$r^2\left((\hat{y}_{i:n})_{1\times n}, \xi \right) = r^2\left(\hat{Z}, \xi \right),$$

$$r^2\left((y_{i:n})_{1\times n}, \xi \right) = r^2\left(Z, \xi \right),$$

and

$$r^2\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right) = r^2\left(Z, \Psi\right).$$

For α and β_Ψ defined as in (4.85) and (4.86),

$$\begin{aligned} r\left(\hat{Z}, \xi\right) &= \left(\sum_{i=1}^{n} \hat{z}_{i:n}\xi_i\right)\left(n\hat{s}^2\sum_{i=1}^{n}\xi_i^{\,2}\right)^{-1/2} \\ &= \hat{Z}'\xi n^{-1/2}\hat{s}^{-1}\left\|\xi\right\|^{-1} \\ &= n^{-1/2}\hat{s}^{-1}\hat{Z}'\alpha \end{aligned} \tag{4.89}$$

and

$$\begin{aligned} r\left(\hat{Z}, \xi\right) &= \left(\sum_{i=1}^{n} \hat{z}_{i:n}\xi_i\right)\left(n\hat{s}^2\sum_{i=1}^{n}\xi_i^{\,2}\right)^{-1/2} \\ &= \hat{Z}'\xi n^{-1/2}\hat{s}^{-1}\left\|\xi\right\|^{-1} \\ &= n^{-1/2}\hat{s}^{-1}\hat{Z}'\alpha \end{aligned} \tag{4.90}$$

and

$$\begin{aligned} r\left(\hat{Z}, \Psi\right) &= \left(\sum_{i=1}^{n} \hat{z}_{i:n}\left(\Psi_i - n^{-1}\sum_{i=1}^{n}\Psi_i\right)\right)\left(n\hat{s}^2\sum_{i=1}^{n}\left(\Psi_i - n^{-1}\sum_{i=1}^{n}\Psi_i\right)^2\right)^{-1/2} \\ &= \hat{Z}'\left(\Psi - n^{-1}\sum_{i=1}^{n}\Psi_i 1_{1\times n}\right)n^{-1/2}\hat{s}^{-1}\left\|\Psi - n^{-1}\sum_{i=1}^{n}\Psi_i 1_{1\times n}\right\|^{-1} \\ &= n^{-1/2}\hat{s}^{-1}\hat{Z}'\beta_\Psi. \end{aligned} \tag{4.91}$$

Therefore, we have

$$\begin{aligned} \left(r\left(\hat{Z},\xi\right) - r\left(\hat{Z},\Psi\right)\right)^2 &= n^{-1}\hat{s}^{-2}\left[\hat{Z}'\alpha - \hat{Z}'\beta_\Psi\right]^2 \\ &= n^{-1}\hat{s}^{-2}\left[\hat{Z}'\left(\alpha - \beta_\Psi\right)\right]^2 \\ &= n^{-1}\hat{s}^{-2}\left[\left(\alpha - \beta_\Psi\right)'\left(\hat{Z} - \xi\right) + \left(\alpha - \beta_\Psi\right)'\xi\right]^2 \\ &\le n^{-1}\hat{s}^{-2}\left[\left\|\alpha - \beta_\Psi\right\|^2\left\|\hat{Z} - \xi\right\|^2 + \left|\left(\alpha - \beta_\Psi\right)'\xi\right|^2\right] \end{aligned} \tag{4.92}$$

Apply Lemma 4.17 and Lemma 4.8 to get

$$\begin{aligned} \left\|\hat{Z} - \xi\right\|^2 &\le nB_1 + L_n^0 \\ &= n o_p\left(n^{-7/6}\left(\log n\right)^2\right) + O_p\left(\log\log n\right) \\ &= O_p\left(\log\log n\right), \text{ as } n \to \infty. \end{aligned} \tag{4.93}$$

Note that $\hat{s}^{-2} \xrightarrow{p} 1$, as $n \to \infty$. To prove the first part of the theorem, combine (4.86), (4.88), (4.92), and (4.93) to get

$$\left(r\left(\hat{Z},\xi\right)-r\left(\hat{Z},\Psi\right)\right)^2 = O_p\left(n^{-1}\right)\left[o\left(n^{-1}\left(\log\log n\right)^{-1}\right)O_p\left(\log\log n\right)+o\left(n^{-1/2}\right)^2\right]$$
$$= o_p\left(n^{-2}\right), \text{ as } n \to \infty,$$

and note that

$$\left|r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \xi\right)-r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \Psi\right)\right| = \left|r^2\left(\hat{Z},\xi\right)-r^2\left(\hat{Z},\Psi\right)\right|$$
$$= \left\|\left[r\left(\hat{Z},\xi\right)+r\left(\hat{Z},\Psi\right)\right]\left[r\left(\hat{Z},\xi\right)-r\left(\hat{Z},\Psi\right)\right]\right\|$$
$$\le 2\left\|\left[r\left(\hat{Z},\xi\right)-r\left(\hat{Z},\Psi\right)\right]\right\|$$
$$= o_p\left(n^{-1}\right), \text{ as } n \to \infty.$$

To prove the second part of the theorem, note that

$$\left|r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \Psi\right)-r^2\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right)\right| \le \left|r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \Psi\right)-r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \xi\right)\right|$$
$$+\left|r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \xi\right)-r^2\left(\left(y_{i:n}\right)_{1\times n}, \xi\right)\right|$$
$$+\left|r^2\left(\left(y_{i:n}\right)_{1\times n}, \xi\right)-r^2\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right)\right|$$
$$= o_p\left(n^{-1}\right)+o_p\left(n^{-1}\right)+o_p\left(n^{-1}\right), \text{ as } n \to \infty. \quad \square$$

In Lemma 3.3 of Verril and Johnson (1987) different sets of vectors are shown to satisfy (4.84), including the vector from Shapiro-Francia statistic, $m = E(Z)$, where Z is defined in (4.9). The Shapiro-Wilk statistic uses the vector $V_0^{-1}m$, where V_0 is the covariance matrix of the standard normal order statistics. In Leslie (1984) it is shown that

(4.94) $\left\|V_0^{-1}m - 2m\right\| = O\left(\left(\log n\right)^{-1/2}\right), \text{ as } n \to \infty.$

These results imply the following Corollary to Lemma 4.20.

Corollary 4.22. *Let $Y_1,\ldots,Y_n$ be i.i.d. k-variate multivariate normal vectors with mean, μ, and positive definite covariance matrix, $\Sigma_{k\times k}$. Let $\hat{d}$ be a random vector and d_0 be a fixed vector which jointly satisfy Assumption 4.1. Let $(\hat{y}_{i:n})_{1\times n}$ and $(y_{i:n})_{1\times n}$ be the vectors of univariate order statistics based on $\left\{\hat{y}_i = \hat{d}\left(Y_i - \overline{Y}\right)\right\}_{i=1}^n$ and $\left\{y_i = d_0\left(Y_i - \mu\right)\right\}_{i=1}^n$, respectively. Then*

$$n\left[r^2\left((\hat{y}_{i:n})_{1\times n}, V_0^{-1}m\right) - r^2\left((y_{i:n})_{1\times n}, V_0^{-1}m\right)\right] = o_p(1),\ \textit{as}\ n\to\infty.$$

Proof. Note that, by the location-scale invariance of $r^2(\bullet,\bullet)$ it will suffice to show that

$$n\left[r^2\left((\hat{y}_{i:n})_{1\times n}, \tfrac{1}{2}V_0^{-1}\underset{\sim}{m}\right) - r^2\left((y_{i:n})_{1\times n}, \tfrac{1}{2}V_0^{-1}\underset{\sim}{m}\right)\right] = o_p(1),\ \text{as}\ n\to\infty. \tag{4.95}$$

Consider

$$\begin{aligned}\left\|\frac{1}{2}V_0^{-1}m - \xi\right\|^2 &= \left\|\frac{1}{2}V_0^{-1}m - m + m - \xi\right\|^2 \\ &\le \left\|\frac{1}{2}V_0^{-1}m - m\right\|^2 + \left\|m-\xi\right\|^2.\end{aligned} \tag{4.96}$$

By (4.94)

$$\begin{aligned}\left\|\frac{1}{2}V_0^{-1}m - m\right\|^2 &= O\left((\log n)^{-1/2}\right) \\ &= \frac{\log\log n}{(\log n)^{1/2}}\left(\frac{1}{\log\log n}O(1)\right) \\ &= o(1)O\left(\frac{1}{\log\log n}\right) \\ &= O\left((\log\log n)^{-1}\right),\ \text{as}\ n\to\infty.\end{aligned}$$

Since m satisfies (4.84), we have

$$\left\|m-\xi\right\|^2 = o\left((\log\log n)^{-1}\right),\ \text{as}\ n\to\infty.$$

To complete the proof, note that $\dfrac{1}{2}V_0^{-1}m$ satisfies (4.84) and invoke Lemma 4.21. □

Chapter V. Specific Tests for Multivariate Normality

5.1. **Introduction.** In Chapter IV, we showed, under the assumption of multivariate normality, that the correlation statistic, $\hat{W}_n$, based on the projections from an estimated linear transformation, $\hat{d}$, is asymptotically equivalent to a correlation statistic, W_n, based on the projections from a fixed linear transformation, d_0, provided that

(5.1) $\left\|\hat{d} - d_0\right\| = O_p\left(n^{-1/2}\right)$, as $n \to \infty$,

and

(5.2) $\left\|d_0\right\| > 0.$

In the case of continuous functionals of the empirical process of the projections from the estimated linear transformation, we require the weaker assumption that $\left\|\hat{d} - d_0\right\| = o_p\left(1\right)$, as $n \to \infty$, for $\left\|d_0\right\| > 0.$ Therefore, if we show that (5.1) and (5.2) are satisfied, we can apply Corollary 3.12 and Theorem 4.21 to get the asymptotic properties of both correlation and E.D.F. type goodness of fit statistics for multivariate normality. We specifically consider E.D.F. and correlation Goodness-of-Fit tests applied to projections from the linear transformations suggested by Peterson and Stromberg (1998) and Wood (1980). As in the preceding chapters, let ξ be the plotting scores associated with the de Wet and Venter statistic, specifically

(5.3) $\xi = \left(\Phi^{-1}\left(i/n+1\right)\right)_{n\times 1}.$

5.2. **Tests Based on Sample Principal Components.** Testing the marginal distribution for univariate normality is a standard practice for investigating multivariate normality. Royston (1983) proposed a method for combining the k-dependent tests into one omnibus test by transforming the k -Shapiro Wilk statistics into an approximately Chi-squared random variable, with $m \le k$ degrees of freedom. The degrees of freedom are then estimated taking into account possible correlation structures between the original m test statistics. This test has been found to behave well when the sample size is small and the variates are uncorrelated.

However, as the correlation increases between variates the type 1 error increases. To improve on this methodology, Srivastava and Hui (1987) and Peterson and Stromberg (1998) suggested using the k-eigenvectors of the sample covariance matrix, also known as the sample principal components, to create k-univariate samples that are then tested, in turn, for univariate normality. For univariate normality, they suggested using with a Shapiro-Wilk or a related correlation test statistic. Each of the k-test statistics will be asymptotically independent when the original vectors are from a multivariate normal distribution. The independence implies that that the tests can easily be combined into an omnibus test statistic for multivariate normality with an asymptotic type 1 error rate of alpha. However, the estimation of the principal components introduces dependence between the observations, which violates the assumptions under which the null distribution of the correlation statistics has been characterized. Peterson and Stromberg (1998) investigated these statistics with a simulation study. Corollary 5.2 proves Peterson and Stromberg's hypothesis, that using the sample eigenvectors does not unduly affect the null distribution of the test statistics for large samples. As in the preceding chapters, let $\bar{Y}$ be the sample mean,

(5.4) $$\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i,$$

and S be the sample covariance matrix,

(5.5) $$S = (n-1)^{-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(Y_i - \bar{Y}\right)'.$$

Let $\left\{\hat{e}_j, \hat{\lambda}_j\right\}_{j=1}^{k}$ be the eigenvector/eigenvalue pairs of the sample covariance matrix, S, and $\left\{e_j, \lambda_j\right\}_{j=1}^{k}$ be the eigenvector/eigenvalue pairs of the population covariance matrix, $\Sigma_{k\times k}$. Furthermore, we will assume that $\Sigma_{k\times k}$ has k distinct non-zero eigenvalues. We make use of the following result from Flurry (1988), which is based on an earlier result by Anderson (1984).

Theorem 5.1. (Flurry, 1988) *Let S denote a random symmetric $p\times p$ matrix, distributed as a Wishart distribution with $n-1$ degrees of freedom and parameter matrix $(n-1)^{-1}\Sigma$, where Σ is positive definite and symmetric. Let $S = \hat{E}\hat{\Lambda}\hat{E}'$ and $\Sigma = E\Lambda E'$ be the spectral decompositions of*

S and Σ *, where* $\hat{\Lambda} = diag\left(\hat{\lambda}_1,\ldots,\hat{\lambda}_p\right)$, $\Lambda = diag\left(\lambda_1,\ldots,\lambda_p\right)$, $\hat{E} = \left(\hat{e}_1,\ldots,\hat{e}_p\right)_{p\times p}$ *, and* $E = \left(e_1,\ldots,e_p\right)_{p\times p}$ *. Assume that all* λ_j *are distinct. Then*

i. The asymptotic distribution of $(n-1)^{1/2}\begin{pmatrix}\hat{\lambda}_1-\lambda_1\\ \vdots\\ \hat{\lambda}_p-\lambda_p\end{pmatrix}$ *as n tends to infinity is p-variate normal with mean zero and covariance matrix* $diag\left(2\lambda_1^2,\ldots,2\lambda_p^2\right)$ *, and the* $\hat{\lambda}_j$ *are independent of* $\hat{E}$ *.*

ii. The asymptotic distribution of $(n-1)^{1/2}\begin{bmatrix}\hat{e}_1-e_1\\ \vdots\\ \hat{e}_p-e_p\end{bmatrix}$ *is* p^2 *-variate normal with mean zero and covariance matrix* V_E *.*

Under the assumption that $Y_1,\ldots,Y_n$ are i.i.d. k-variate multivariate normal vectors with mean vector, μ, and positive definite covariance matrix, $\Sigma_{k\times k}$, the sample covariance matrix S has a Wishart distribution with n degrees of freedom and parameter matrix $(n-1)^{-1}\Sigma$ (Seber, 1984). Theorem 5.1 then implies the following two results

(5.6) $\left\|\hat{e}_j - e_j\right\| = O_p\left(n^{-1/2}\right),$ for $j = 1,\ldots k,$ as $n \to \infty$

and

(5.7) $\left|\hat{\lambda}_j - \lambda_j\right| = O_p\left(n^{-1/2}\right),$ for $j = 1,\ldots,k,$ as $n \to \infty$.

Also by definition $\left\|e_j\right\| = 1$, therefore (5.1) and (5.2) are satisfied. Then, Theorem 4.21 immediately implies the following corollary.

Corollary 5.2 *Let* $Y_1,\ldots,Y_n$ *be i.i.d. k-variate multivariate normal vectors with mean,* $\underset{\sim}{\mu}$ *, and positive definite covariance matrix,* $\Sigma_{k\times k}$ *, with* k *unique eigenvalues. Let* $\hat{e}_j \in R^k$ *be the* j^{th} *eigenvector of the sample covariance matrix,* S *, and* $e_j \in R^k$ *be the* j^{th} *eigenvector of the population covariance matrix,* $\Sigma_{k\times k}$ *. Let* $\left(\hat{y}_{i:n}\right)_{1\times n}$ *and* $\left(y_{i:n}\right)_{1\times n}$ *be the vectors of univariate order* *statistics based on* $\left\{\hat{y}_i = \hat{e}_j'\left(Y_i - \overline{Y}\right)\right\}_{i=1}^{n}$ *and* $\left\{y_i = e_j'\left(Y_i - \mu\right)_i\right\}_{i=1}^{n}$*, respectively. Let* Ψ *be a vector of constants that satisfies* $\sum_{i=1}^{n}\left(\Psi_i - \xi_i\right)^2 = o\left(\left(\log\log n\right)^{-1}\right)$*, as* $n \to \infty$*, where* ξ_i *is defined in (5.3). Then*

$$n\left[r\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \Psi\right) - r\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right)\right] = o_p(1), \text{ as } n \to \infty.$$

By noting that $\hat{\lambda}_j^{-1/2}\hat{e}_j' S \hat{e}_j \hat{\lambda}_j^{-1/2} = 1$, Theorem 5.1 and Corollary 3.12 immediately imply the Corollary 5.3.

Corollary 5.3. *Let* $Y_1, \ldots, Y_n$ *be i.i.d. k-variate multivariate normal vectors with mean,* $\underset{\sim}{\mu}$*, and positive definite covariance matrix,* $\Sigma_{k\times k}$*, with* k *unique eigenvalues. Let* $\{\hat{e}_j, \hat{\lambda}_j\}_{j=1}^{k}$ *be the eigenvector/eigenvalue pairs of* S*. Let*

$$G_n(t) = n^{1/2}\left[n^{-1}\sum_{i=1}^{n} I\left(\hat{\lambda}_j^{-1/2}\hat{e}_j'\left(Y_i - \overline{Y}\right) \le t\right) - \Phi(t)\right], \ -\infty < t < \infty,$$

and G *be a tight Gaussian process with covariance function*

$$\Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \frac{ts\phi(s)\phi(t)}{2}, -\infty < t < \infty.$$

Let T *be a continuous functional from* $\ell^{\infty}(R)$ *to* R*. Then*

$$T\left[G_n\right] \rightsquigarrow T\left[G\right], \text{ as } n \to \infty.$$

It is worth noting that in the case of common principle components, the large sample properties presented in Flurry (1988, Theorem 4.4) are sufficient to demonstrate that the assumptions (5.1) and (5.2) are satisfied, under appropriate assumptions on the rate at which the sample sizes of the different groups tend to infinity. This suggests that similar results to corollaries 5.2 and 5.3 will hold for these linear transformations.

5.3. Tests Based on Wood's Symmetric Decomposition. In Wood (1981), the rows of the symmetric decomposition of the inverse of the sample covariance matrix are used to create k univariate samples that are each tested for univariate normality with an empirical cumulative distribution function Goodness-of-Fit test, such as a Cramer-von Mises statistic.

Let S^{-1} of the inverse of the sample covariance matrix. Let $S^{-1/2}$ and $\Sigma^{-1/2}$ be symmetric positive definite matrices such that

(5.8) $S^{-1/2}S^{-1/2} = S^{-1}$,
and

(5.9) $\Sigma^{-1/2}\Sigma^{-1/2} = \Sigma^{-1}$.
Let

(5.10) $\hat{b}_j$ be the j^{th} row of $S^{-1/2}$,
and

(5.11) b_j be the j^{th} row of $\Sigma^{-1/2}$.

We then consider the k-correlation statistics applied $\left(\hat{y}_{i:n}\right)_{1\times n}$, in this case the vector of order statistics from $\left\{\hat{b}_j\left(Y_i - \overline{Y}\right)\right\}_{i=1}^{n}$. Before proceeding further we will show that the set of k linear transformations in (5.9) satisfy conditions (5.1) and (5.2).

Lemma 5.4. *Let S be the sample covariance matrix from an i.i.d. sample of n multivariate normal random vectors with a mean vector* μ *and positive definite covariance matrix* Σ. *Let* $S^{-1} = S^{-1/2}S^{-1/2}$ *and* $\Sigma^{-1} = \Sigma^{-1/2}\Sigma^{-1/2}$. *Then*

$$\left\| S^{-1/2} - \Sigma^{-1/2} \right\| = O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.$$

Proof. Let Γ be the matrix of eigenvectors of Σ and Δ the diagonal matrix of eigenvalues of Σ. We assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_k$ and that

$$\Delta = \begin{pmatrix} \lambda_1 I_{q_1\times q_1} & 0 & 0 & 0 \\ 0 & \lambda_2 I_{q_2\times q_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_k I_{q_k\times q_k} \end{pmatrix}$$

where q_i is the multiplicity associated with λ_i. Let D be the matrix of eigenvalues associated with S and C the matrix of eigenvectors of S such that $D = diag\left(D_i\right)$, where

$$D = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_k \end{pmatrix}_{p \times p},$$

$$D_i = \begin{pmatrix} d_{q_{i-1}+1} & & 0 \\ & \ddots & \\ 0 & & d_{q_{i-1}+q_i} \end{pmatrix},$$

$$d_1 > d_2 > d_3 > \ldots > d_p,$$

and

$$S = CDC'.$$

Note, by definition,

$$CC' = I$$

and

$$\Gamma\Gamma' = I.$$

Let

$$E = \begin{pmatrix} E_{11} & E_{12} & \cdots & E_{1r} \\ E_{21} & E_{22} & & \vdots \\ \vdots & & \ddots & \\ E_{r1} & \cdots & & E_{rr} \end{pmatrix}_{p \times p}$$

be the matrix of eigenvectors of $\Gamma' S \Gamma$. Then

$$T = \Gamma' S \Gamma$$

is the sample covariance matrix of a sample of n i.i.d. multivariate normal random vectors with covariance matrix Δ. Note that D is the matrix of of sample eigenvectors of T as well as S.

Anderson (1963) gives the following results concerning D and E,

(5.12) $\|D_k - \lambda_k I\| = O_p\left(n^{-1/2}\right)$, as $n \to \infty$,

and

(5.13) $\|E_{kl}\| = O_p\left(n^{-1/2}\right)$, $k \neq l$, as $n \to \infty$.

Let d_{ki} be the i^{th} diagonal element of D_k. Note that (5.12) implies

(5.14)
$$\begin{aligned} d_{ki}^{-1/2} - \lambda_k^{-1/2} &= \left(d_{ki}\lambda_k\right)^{-1/2}\left(\lambda_k^{1/2} - d_{ki}^{1/2}\right) \\ &= \left(d_{ki}\lambda_k\right)^{-1/2}\left(d_{ki}^{1/2} + \lambda_k^{1/2}\right)^{-1}\left(\lambda_k - d_{ki}\right) \\ &= O_p(1)O_p(1)O_p\left(n^{-1/2}\right) \\ &= O_p\left(n^{-1/2}\right), \text{ as } n \to \infty, \end{aligned}$$

which implies

(5.15) $D_k^{-1/2} - \lambda_k^{-1/2} I = O_p\left(n^{-1/2}\right),$ as $n \to \infty.$

Next, note that $C = \Gamma E$ and consider

$$\begin{aligned} S^{-1/2} - \Sigma^{-1/2} &= CD^{-1/2}C' - \Gamma\Delta^{-1/2}\Gamma' \\ &= \Gamma E D^{-1/2} E'\Gamma' - \Gamma\Delta^{-1/2}\Gamma' \\ &= \Gamma\left(T^{-1/2} - \Delta^{-1/2}\right)\Gamma'. \end{aligned}$$

Therefore it will suffice to show that

$$\left\|T^{-1/2} - \Delta^{-1/2}\right\| = O_p\left(n^{-1/2}\right), \text{ as } n \to \infty.$$

Consider

(5.16) $\left(T^{-1/2} - \Delta^{-1/2}\right) = \left(ED^{-1/2}E' - \Delta^{-1/2}\right) = \left(N_{ij}\right),$

where

$$N_{jj} = \sum_{i=1}^{r} E_{ji} D_i^{-1/2} E'_{ji} - \lambda_j I,$$

and

$$N_{kl} = \sum_{i=1}^{r} E_{ki} D_i^{-1/2} E'_{li}, \; k \neq l.$$

By (5.13) and (5.15) we have, for $k \neq l$,

(5.17) $\left\|E_{ki} D_i^{-1/2} E'_{li}\right\| = O_p\left(n^{-1/2}\right),$ as $n \to \infty,$

which implies

(5.18) $\left\|N_{kl}\right\| = O_p\left(n^{-1/2}\right),$ as $n \to \infty.$

Consider

$$\begin{aligned} \sum_{l=1}^{k} E_{il} D_l^{-1/2} E'_{il} &= E_{ii} D_i^{-1/2} E'_{ii} + \sum_{i \neq l} E_{il} D_l^{-1/2} E'_{il} \\ &= E_{ii} D_i^{-1/2} E'_{ii} + O_p\left(n^{-1}\right) \\ &= E_{ii}\left(D_i^{-1/2} - \lambda_i^{-1/2} I\right) E'_{ii} + \lambda_i^{-1/2} E_{ii} E' + O_p\left(n^{-1}\right) \\ &= \lambda_i^{-1/2} E_{ii} E'_{ii} + O_p\left(n^{-1/2}\right) \\ &= \lambda_i^{-1/2} E_{ii} E'_{ii} + \lambda_i^{-1/2} \sum_{i \neq l} E_{il} E'_{il} - \lambda_i^{-1/2} \sum_{i \neq l} E_{il} E'_{il} + O_p\left(n^{-1/2}\right) \\ &= \lambda_i^{-1/2} I + O_p\left(n^{-1/2}\right), \text{ as } n \to \infty. \end{aligned} \tag{5.19}$$

Which implies

$$(5.20)\quad \left\|N_{jj}\right\| = \left\|\sum_{i=1}^{r} E_{ji} D_i^{-1/2} E'_{ji} - \lambda_j I\right\|$$
$$= O_p\left(n^{-1/2}\right),\ \text{as } n \to \infty.$$

Therefore

$$\left\|T^{-1/2} - \Delta^{-1/2}\right\| = O_p\left(n^{-1/2}\right),\ \text{as } n \to \infty,$$

and

$$\left\|S^{-1/2} - \Sigma^{-1/2}\right\| = \left\|\Gamma\left(T^{-1/2} - \Delta^{-1/2}\right)\Gamma'\right\|$$
$$\le \left\|\Gamma\right\|\left\|T^{-1/2} - \Delta^{-1/2}\right\|\left\|\Gamma'\right\|$$
$$= O_p\left(n^{-1/2}\right),\ \text{as } n \to \infty. \qquad \square$$

Since $\Sigma^{-1/2}$ is positive definite, $\left\|b_j\right\| > 0$, for $j = 1,\ldots,k$, condition (5.2) is satisfied. By Lemma 5.3, condition (5.1) is satisfied. Corollary 5.4 will follow directly from Theorem 4.21**.**

Corollary 5.5. *Let* $Y_1,\ldots,Y_n$ *be i.i.d. k-variate multivariate normal vectors with mean,* μ*, and positive definite covariance matrix,* $\Sigma_{k\times k}$*. Let* $\hat{b}_j$ *and* b_j *be defined as in (5.10) and (5.11). Let* $\left(\hat{y}_{i:n}\right)_{1\times n}$ *and* $\left(y_{i:n}\right)_{1\times n}$ *be the vectors of univariate order statistics based on* $\left\{\hat{b}_j\left(Y_i - \overline{Y}\right)\right\}_{i=1}^{n}$ *and* $\left\{b_j\left(Y_i - \mu\right)\right\}_{i=1}^{n}$*, respectively. Let* Ψ *be a vector of constants that satisfies* $\sum_{i=1}^{n}\left(\Psi_i - \xi_i\right)^2 = o\left(\left(\log\log n\right)^{-1}\right)$, as $n \to \infty$, $\sum_{i=1}^{n}\left(\Psi_i - \xi_i\right)^2 = o\left(\left(\log\log n\right)^{-1}\right)$*, as* $n \to \infty$*, where* ξ_i *is defined in (3.3). Then*

$$n\left[r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n}, \Psi\right) - r^2\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right)\right] = o_p(1),\ \text{as } n \to \infty.$$

As with Corollary 5.3, by noting that $\hat{b}_j S \hat{b}'_j = 1$, Theorem 5.1 and Corollary 3.12 immediately imply Corollary 5.6.

Corollary 5.6. *Let* $Y_1,\ldots,Y_n$ *be i.i.d. k-variate normal vectors with mean,* μ*, and positive definite covariance matrix,* $\Sigma_{k\times k}$*. Let* $\hat{b}_j$ *and* b_j *be defined as in (5.10) and (5.11). Let*

$$G_n(t) = n^{1/2}\left[n^{-1}\sum_{i=1}^{n} I\left(\hat{b}_j\left(Y_i - \overline{Y}\right) \le t\right) - \Phi(t)\right],\ -\infty < t < \infty,$$

and G *be a tight Gaussian process with covariance function*

$$\Phi(min(t,s)) - \Phi(t)\Phi(s) - \phi(s)\phi(t) - \frac{ts\phi(s)\phi(t)}{2}, -\infty < t < \infty.$$

Let T *be a continuous functional from* $\ell^{\infty}(R)$ *to* R*. Then*

$$T\left[G_n\right] \rightsquigarrow T\left[G\right], \text{ as } n \to \infty .$$

Chapter VI. Simulations

6.1. **Introduction.** Up to this point we have considered univariate tests for normality applied to the projections determined by a single estimated linear transformation. We proved that the estimation of the fixed linear transformation by a data suggested linear transformation does not affect the limiting distribution of the test statistic under the assumption of multivariate normality. As mentioned in the sections 1.3 and 1.4, if we choose k fixed linear transformations such that they are orthogonal to each other with respect to the population covariance matrix, then the corresponding k standardized univariate test statistics based on the estimated projections will be asymptotically mutually independent. In Section 6.2, we present our method for simulating the univariate null distribution for correlation statistics. In sections 6.3, we review two strategies for combining the k p-values into one omnibus test statistic for multivariate normality. The first was proposed by Peterson and Stromberg (1998) and the second is a version of the method proposed by Royston (1983). For the omnibus tests of multivariate normality based on combining the k univariate tests, Section 6.3 has a Type I error simulation study and Section 6.4 has a power simulation study.

6.2. **Monte Carlo Simulation of the Null Distribution of Univariate Correlation Test Statistics.** In the preceding chapters it was shown that the correlation statistic determined by the estimated linear transformation, $r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n},\Psi\right)$, is asymptotically equivalent to the correlation statistic determined by the fixed linear transformation, $r^2\left(\left(y_{i:n}\right)_{1\times n},\Psi\right)$. The correlation statistic, $r^2\left(\left(y_{i:n}\right)_{1\times n},\Psi\right)$, is just the corresponding test statistic for testing univariate normality in the direction of d_0. Therefore, it is reasonable to use the distribution of the univariate correlation statistic to calculate a p-value for correlation statistic determined by the estimated linear transformation.

The Shapiro-Wilk statistic is the most extensively studied of the univariate correlation statistics. Unfortunately, the exact null distribution of the univariate Shapiro-Wilk statistic is unknown for sample sizes greater than three, and the convergence rate of the sampling distribution of the test statistic to its asymptotic distribution appears to be rather slow (Verrril and Johnson, 1988). The common approach to calculating the p-value for a univariate Shapiro-Wilk statistic is to transform the statistic such that it has an approximate normal distribution (Royston, 1995). Unfortunately these approximations only work well for a range of sample sizes, where the range depends upon the type of approximation (Royston, 1992). Royston's approximation is used in such packages as R (Royston,1995), S-Plus (Royston, 1992), and the univariate procedure in SAS (Royston, 1992). The sampling distributions of the other correlation statistics are not as extensively studied as the Shapiro-Wilk statistic's sampling distribution. Therefore, as recommended by Verril and Johnson (1988), we take advantage location scale invariance of the correlation statistics to use Monte Carlo methods to approximate the null distribution of the test statistic, $r^2\left(\left(y_{i:n}\right)_{1\times n}, \Psi\right)$.

In the following simulations, the null distribution of the univariate correlation statistics is determined by a simulation of 30,000 samples of n normal random variables with mean equal to zero and variance equal to one, where n is chosen appropriately for the application.

6.3. Combining P-Values from k-independent Tests. We will denote the approximate p-values based off of a univariate test of normality applied to the estimated projections as $\hat{p}_i\left(Y_1,\ldots,Y_n\right)$, $i=1,\ldots,k$. For the tests considered in Chapter V, $\hat{p}_i\left(Y_1,\ldots,Y_n\right) \rightsquigarrow p_i$, as $n\to\infty$, where p_i are i.i.d. uniform random variable on the interval $\left(0,1\right)$. In this section we review two methods for combining the k approximate p-values into one omnibus test statistic for multivariate normality.

The first method is based on the idea that we will reject the assumption of multivariate normality if at least one of the k directions suggests non-normality. Noting that $\min\left\{p_i\right\}_{i=1}^k$ is a beta random variable with parameters 1 and k. We will denote this random variable as $\beta\left(1,k\right)$. Therefore, for finite n, we will compare $\min\left\{\hat{p}_i\right\}_{i=1}^k$ with the α-quantile of a $\beta\left(1,k\right)$ random variable. This will give us an asymptotic Type I error of α. We will refer to this method of combining p-values as the minimum p-value omnibus test. This is a transformation of the "upper bound p-value" suggested in Peterson and Stromberg (1998).

Let Ψ_j be the C.D.F. of a chi-squared random variable with j degrees of freedom. Then $\sum_{i=1}^{k} -2\log(p_i)$ is a chi-squared random variable with $2k$ degrees of freedom and $\sum_{i=1}^{k} -2\log(\hat{p}_i) \rightsquigarrow \sum_{i=1}^{k} -2\log(p_i),$ as $n \to \infty$. Therefore, we will reject the assumption that the original vectors have a multivariate normal distribution when $\sum_{i=1}^{k} -2\log(\hat{p}_i) > \Psi_{2k}^{-1}(1-\alpha)$. This will give us an asymptotic Type I error of α. This is the method suggested by Fisher (1946) for combining the k independent p-values. We will refer to this method of combining p-values as the Fisher omnibus test.

6.4. **Type I Error Simulation.** To study the effect on the Type I error of using a set of data suggested linear transformations in place of a fixed linear transformation, we perform a short simulation. For correlation statistics, we will simulate the null distribution of the univariate test statistic under the assumption of normality. For the E.D.F. goodness of fit statistics we will use the tabulated values coded in most basic software packages. Due to the simulation of the univariate null distribution for the correlation statistics, the variability of the Type I error estimate for the correlation statistics will be greater than the variance of the E.D.F. goodness of fit Type I error estimates. The tests considered in the simulation study are summarized in Table 6.1.

Table 6.1- Tests of Multivariate Normality considered in Simulations

Test	Omnibus Test	Univariate Goodness of Test	Linear Transformation
1	*minimum p-value*	Shapiro-Wilk	Sample Eigenvectors
2	*minimum p-value*	de Wet and Venter	Sample Eigenvectors
3	*minimum p-value*	Shapiro-Wilk	Wood's Symmetric Decomposition
4	*minimum p-value*	de Wet and Venter	Wood's Symmetric Decomposition
5	*minimum p-value*	Kolmogorov-Smirnov	Sample Eigenvectors
6	*minimum p-value*	Cramer-von Mises	Sample Eigenvectors
7	*minimum p-value*	Kolmogorov-Smirnov	Wood's Symmetric Decomposition
8	*minimum p-value*	Cramer-von Mises	Wood's Symmetric Decomposition
9	*Fisher*	Shapiro-Wilk	Sample Eigenvectors
10	*Fisher*	de Wet and Venter	Sample Eigenvectors
11	*Fisher*	Shapiro-Wilk	Wood's Symmetric Decomposition
12	*Fisher*	de Wet and Venter	Wood's Symmetric Decomposition
13	*Fisher*	Kolmogorov-Smirnov	Sample Eigenvectors
14	*Fisher*	Cramer-von Mises	Sample Eigenvectors
15	*Fisher*	Kolmogorov-Smirnov	Wood's Symmetric Decomposition
16	*Fisher*	Cramer-von Mises	Wood's Symmetric Decomposition

The m^{th} sample in the simulation was generated as follows; first Σ_m, a random $k \times k$ Wishart matrix with k degrees of freedom and expectation $.1I_{kxk} + .9(1)_{kxk}$, is generated, then a sample of n zero mean multivariate normal random vectors with covariance Σ_m. Each sample generated in the simulation was tested for normality with the test in Table 6.1. The results of the Type I error simulation are summarized in Table 6.2, for $k = 2$, and Table 6.3, for $k = 5$. We consider sample sizes ranging from 20 to 1000.

Table 6.2- Type One Error Simulations in two dimensions

Test/Sample Size	20	40	100	250	500	1000
Test 1	0.048	0.048	0.049	0.050	0.051	0.050
Test 2	0.047	0.050	0.049	0.050	0.050	0.052
Test 3	0.049	0.048	0.049	0.050	0.055	0.050
Test 4	0.048	0.052	0.051	0.049	0.052	0.050
Test 5	0.048	0.050	0.050	0.052	0.048	0.047
Test 6	0.050	0.051	0.050	0.051	0.048	0.051
Test 7	0.047	0.050	0.050	0.051	0.047	0.048
Test 8	0.050	0.050	0.053	0.049	0.049	0.052
Test 9	0.049	0.051	0.051	0.053	0.051	0.051
Test 10	0.047	0.049	0.050	0.050	0.049	0.051
Test 11	0.048	0.049	0.051	0.052	0.054	0.050
Test 12	0.047	0.052	0.052	0.049	0.051	0.050
Test 13	0.047	0.050	0.051	0.053	0.048	0.049
Test 14	0.052	0.053	0.049	0.051	0.048	0.051
Test 15	0.047	0.048	0.051	0.051	0.048	0.048
Test 16	0.051	0.051	0.052	0.049	0.049	0.052

Table 6.3- Type One Error Simulations in five dimensions

Test/Sample Size	20	40	100	250	500	1000
Test 1	0.044	0.048	0.048	0.044	0.051	0.049
Test 2	0.047	0.050	0.047	0.043	0.053	0.051
Test 3	0.042	0.049	0.048	0.043	0.048	0.047
Test 4	0.044	0.052	0.049	0.044	0.050	0.050
Test 5	0.050	0.047	0.050	0.051	0.047	0.046
Test 6	0.050	0.048	0.049	0.049	0.049	0.050
Test 7	0.049	0.050	0.052	0.051	0.050	0.046
Test 8	0.049	0.049	0.050	0.050	0.050	0.048
Test 9	0.047	0.048	0.048	0.044	0.051	0.048
Test 10	0.048	0.048	0.049	0.046	0.052	0.049
Test 11	0.045	0.049	0.050	0.045	0.049	0.048
Test 12	0.045	0.049	0.050	0.046	0.052	0.047
Test 13	0.047	0.047	0.050	0.051	0.052	0.049
Test 14	0.052	0.050	0.049	0.049	0.050	0.050
Test 15	0.048	0.048	0.053	0.050	0.052	0.049
Test 16	0.050	0.051	0.051	0.049	0.051	0.048

This simulation suggests that all of the tests have reasonable Type I error control.

6.5. Power Simulation Study. In Rizzo and Szekely (2005) a simulation for a test of multivariate normality is performed based on using mixtures of multivariate normal random variables. We replicate their simulation study for a sample size of 50 in five dimensions. This simulation will illustrate the dependence of these tests on the choice of the linear transformation. Let $\mu_1 = (0)_{k\times1}$, $\mu_2 = (3)_{k\times1}$, $\mu_3 = (6.708204, 0, \ldots, 0)_{k\times1}$, $\Sigma_1 = I_{kxk}$ and $\Sigma_2 = .1I_{kxk} + .9(1)_{kxk}$. The mixtures of multivariate normal distributions considered in the power simulation are summarized in Table 6.4. The powers of the tests in Table 6.1 are summarized in Table 6.5.

In the case of the mixture distribution, the eigenvector transformation tests are more powerful than the symmetric decomposition test due to the eigenvectors tending to point in the directions of maximum variability, which the cases considered is also the direction of the departure from normality.

Table 6.4- Mixtures of Multivariate Normal Distributions

	Mixing Proportion	**First Normal Distribution**	**Second Normal Distribution**
Mixture 1	0.5	$N(\mu_1, \Sigma_1)$	$N(\mu_2, \Sigma_1)$
Mixture 2	0.79	$N(\mu_1, \Sigma_1)$	$N(\mu_2, \Sigma_1)$
Mixture 3	0.9	$N(\mu_1, \Sigma_1)$	$N(\mu_2, \Sigma_1)$
Mixture 4	0.5	$N(\mu_1, \Sigma_1)$	$N(\mu_1, \Sigma_2)$
Mixture 5	0.9	$N(\mu_1, \Sigma_1)$	$N(\mu_1, \Sigma_2)$
Mixture 6	0.5	$N(\mu_3, \Sigma_2)$	$N(\mu_1, \Sigma_2)$
Mixture 7	0.9	$N(\mu_3, \Sigma_2)$	$N(\mu_1, \Sigma_2)$

Table 6.5- Power Study for Mixtures of Multivariate Normal Distributions, for k=5 and n=50.

Test	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5	Mixture 6	Mixture 7
1	1.000	1.000	0.992	0.750	0.125	0.921	0.882
2	0.999	1.000	0.992	0.836	0.152	0.669	0.843
3	0.044	0.052	0.069	0.232	0.047	1.000	0.995
4	0.041	0.051	0.075	0.298	0.047	1.000	0.996
5	1.000	1.000	0.958	0.645	0.073	0.874	0.711
6	1.000	1.000	0.976	0.818	0.090	0.967	0.825
7	0.047	0.048	0.053	0.102	0.048	1.000	0.995
8	0.046	0.049	0.057	0.144	0.050	1.000	0.995
9	1.000	1.000	0.991	0.881	0.127	0.845	0.886
10	0.986	1.000	0.992	0.959	0.172	0.552	0.846
11	0.045	0.053	0.071	0.277	0.050	1.000	0.995
12	0.040	0.053	0.087	0.451	0.051	0.996	0.995
13	0.998	1.000	0.947	0.832	0.079	0.807	0.709
14	1.000	1.000	0.972	0.926	0.098	0.939	0.830
15	0.048	0.048	0.054	0.134	0.049	1.000	0.995
16	0.047	0.052	0.060	0.195	0.051	1.000	0.995

Chapter VII. Discussion and Future Research

7.1. **Summary.** In this dissertation, we considered univariate goodness-of-fit tests applied to projections of multivariate normal random vectors. The large sample behavior of these statistics was suspected to the same as the univariate statistics, however there has yet to be a formal derivation of the asymptotic properties of these statistics. The theorems presented in Chapter III, provide the general theory for the weak convergence the standardized empirical process of the projections from a data suggested linear transformation of multivariate normal random vectors. The limiting process determines the asymptotic behavior of continuous functions of the empirical process, such as E.D.F. goodness of fit statistics. The theorems presented in Chapter IV provide an asymptotic representation of univariate correlation goodness of fit statistics applied to projections from a data suggested linear transformation of multivariate normal random vectors, which determines the asymptotic distribution of the statistics. In Chapter V, we demonstrated that our theorems apply to some commonly used tests for multivariate normality that currently lack a formal derivation of the null distribution. Specifically, for the tests proposed by Srivastava and Hui (1987), Peterson and Stromberg (1998), and correlation tests based on the projections from Wood (1981), we derived the asymptotic distribution of the tests under the null hypothesis. In Chapter VI, we investigated the small sample behavior of the projection tests. We found that the Type I error is maintained at a reasonable rate even when the sample size is as small as 20 and the variates are highly correlated. However, the power of projection based tests was demonstrated to be highly dependent upon the type of departure from normality. This is a common problem in testing for normality, even in the univariate case.

7.2. **Future research** The popular Royston (1983) adaptation of Shapiro-Wilk test to multivariate normality has commonly been found to have a high power when compared to other tests for multivariate normality (Mecklin and Mundfrom, 2005; Romeu and Ozturk, 1993). This test, with the correction proposed by Srivastava and Hui (1987) and Peterson and Stromberg (1998), now has a known asymptotic null distribution. To date the consistency properties of the test are unknown, although we suspect that the test will not be consistent against alternatives that have normal distributions for projections determined by the eigenvectors of the covariance matrix. This suspicion leads to the following conjecture.

Conjecture 7.1. *Let* $Y_1,\ldots,Y_n$ *be i.i.d. k-variate vectors with mean,* μ *, such that* $\max_{1\le i\le n}\|Y_i\| = O_p\left(\log(n)\right)$*, as* $n\to\infty$ *. Let* c_0 *be a fixed row vector such that* $c_0\left(Y_i-\mu\right)$ *are i.i.d. standard normal random variables,* $i=1,\ldots,n$ *. Let* $\hat{c}$ *be a row vector such that* $\|c_0-\hat{c}\| = O_p\left(n^{-1/2}\right)$*, as* $n\to\infty$*. Let* $\left(\hat{y}_{i:n}\right)_{1\times n}$ *and* $\left(y_{i:n}\right)_{1\times n}$ *be the vectors of univariate order statistics based on* $\left\{\hat{c}\left(Y_i-\bar{Y}\right)\right\}_{i=1}^{n}$ *and* $\left\{c_0\left(Y_i-\mu\right)\right\}_{i=1}^{n}$*, respectively.*

$$n\left[r^2\left(\left(\hat{y}_{i:n}\right)_{1\times n},\xi\right)-r^2\left(\left(y_{i:n}\right)_{1\times n},\xi\right)\right]=o_p(1),\ as\ n\to\infty.$$

In the proof of Theorem 4.19, we assume that the difference between the estimated linear transformation and the fixed linear transformation was $O_p\left(n^{1/2}\right)$, as $n\to\infty$. For the eigenvectors of the sample covariance matrix this assumption is not satisfied when the eigenvalues of the population covariance matrix have a multiplicity greater than one. It is suspected that this assumption can be relaxed by using the method of proof used in Wood (1981).

BIBLIOGRAPHY

Anderson, T.W. (1963). Asymptotic Theory for Principal Component Analysis. *Annals of Mathematical Statistics*. 34:122-148.

Bahadur, R. R. (1966). A Note on Quantiles in Large Samples. *Annals of Mathematical Statistics*. 37:577-580.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Bogdan, M. (1999). Data Driven Smooth Tests for Bivariate Normality. *Journal of Multivariate Analysis*. 68:26-53.

Brown, B.M. and Hettmansperger, T.P. (1996). Normal Scores, Normal Plots, and Tests for Normality. *Journal of the American Statistical Association*. Vol. 91.

Csorgo, S. (1986). Testing for Normality in Arbitrary Dimension. *Annals of Statistics*. 14:708-723.

Csorgo, S. (1989). Consistency of Some Tests for Multivariate Normality. *Metrika*. 36: 107-116.

de Wet, T. and Venter, J. H. (1973). Asymptotic Distributions for Quadratic Forms with Applications to Tests of Fit. *Annals of Statistics*. 1:380-386.

de Wet, T., Venter, J. H. (1972). Asymptotic distributions of certain test criteria of normality. *South African Statistical Journal*. 6:135-149.

del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2000). Contributions of empirical and Quantile Processes to the Asymptotic Theory of Goodness-of-Fit Tests. Test. 9:1-96.

Durbin, J. (1973). Weak Convergence of the Sample Distribution Function when Parameters are Estimated. *Annals of Statistics*, 1: 279-290.

Flurry, B. (1988). Common Principal Components and Related Multivariate Models. New York: Wiley.

Gnanadesikan, R. (1997). Methods for Statistical Data Analysis of Multivariate Observations. 2nd ed. New York : Wiley.

Henze, N. (2002). Invariant tests for Multivariate Normality: a Critical Review. *Statistical Papers*. 43:467-506.

Henze, N. and B. Zirkler (1990). A class of invariant consistent tests for multivariate normality. Communications in Statistics: Theorem and Methods. 19:3595–3618.

Hoeffding, W. (1953). On the Distribution of the Expected Values of the Order Statistics. *Annals of Mathematical Statistics*. 24:93-100.

Koziol, J. A. (1983). On Assessing Multivariate Normality, *Journal of the Royal Statistical Society. Series B (Methodological)*. 45:358-361.

Leslie, J. R., Stephens, M. A., and Fotopoulos, S. (1986). Asymptotic Distribution of the Shapiro-Wilk W for Testing for Normality. *Annals of Statistics*. 14:1497-1506.

Looney, S. W. and Gulledge, T. R. (1985). Use of the Correlation Coefficient with Normal Probability Plots. *American Statistician*. 39:75-79.

Mardia, K.V. (1974). Applications of some measures of Multivariate Skewness and Kurtosis in Testing Normality and Robust Studies. *Sankhya: The Indian Journal of Statistics*. 36.B: 115-128.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 57519-530.

Mecklin, C. and J., Mundfrom, D. J. (2004). An Appraisal and Bibliography of Tests for Multivariate Normality. *The International Statistical Review*. 72, no. 1: 123–138.

Mecklin, C. and J. and Mundfrom, D. J. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *Journal of Statistical Computation and Simulation*. 75: 93 – 107.

Mudholkar, G. S., Srivastava, D. K., and Lin, C. T. (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality. *Communications in Statistics: Theory and Methods*. 24:953-985.

Peterson, P. and Stromberg, A. (1998). A Simple Test for Departures from Multivariate Normality. *University of Kentucky, Department of Statistics. Technical Report*: 373.

Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlag: New York.

Romeu, J. and A. Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *Journal of Multivariate Analysis*. 46:309–334.

Royden, H. L. (1968). Real Analysis. 2nd edition. Macmillan Publishing Co. Toronto, Ontario.

Royston, P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing*. 2:17 - 119.

Royston, J. P., (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Statistics and Computing*. 2:117-119.

Royston, J. P. (1995). Shapiro-Wilk normality test and P-values. *Applied Statistics.* 31:176-180

Seber, G. A. F. (1984). Multivariate Observations. New York: Wiley.

Seier, E. (2002). Comparison of Tests for Univariate Normality. *Interstat*. January Issue.

Sen, P. K., Jureˇckov´a, J., and Picek J. (2003). Goodness-of-Fit Test of Shapiro-Wilk Type with Nuisance Regression and Scale. *Austrian Journal of Statistics*, 32:163-177.

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

Shapiro, S. S. and Francia, R. S. (1972). An Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*. 67:215-216.

Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance test for normality(complete samples), Biometrika, 52, pp. 591-611.

Srivastava, M. S. and Hui, T. K. (1987). On Assessing Multivariate Normality based on the Shapiro-Wilk W Statistic. *Statistics and Probability Letters*. 5:15-18.

Szekely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*. 93:58-80

Vaart and Wellner (1996). Weak Convergence and Empirical Processes: With Applications to Statistics. Springer-Verlag, New York.

van der Vaart (1998). Asymptotic Statistics. Cambridge University Press.

Verril, S. and Johnson, R. A. (1987). The Asymptotic Equivalence of Some modified Shapiro-Wilk Statistics-Complete and Censored Sample Cases. *The Annals of Statistics.* 15:413-419.

Verril, S. and Johnson, R. A.(1988). Tables and Large Sample Distribution Theory for Censored Data Correlation Statistics for Testing Normality. *Journal of the American Statistical Association*. 83:1192-1197.

Wellner, J. A. (2005). Empirical Processes: Theory and Applications. Lecture Notes. Summer School on Statistics and Probability. Bocconi University. Milan, Torgnon. Valle d'Aosta, June 30 - July 19, 2003.
http://www.stat.washington.edu/jaw/RESEARCH/TALKS/talks.html.

Wood, C. L. (1975). Weak convergence of a modified empirical stochastic process with applications to Kolmogorov-Smirnov statistics. Ph. D. dissertation. Florida State University.

Wood, C. L. (1981). Goodness-of-Fit for Multivariate Normality. Technical Report 182. Department of Statistics. University of Kentucky.

Wood, C. L. (1984). On test of normality of experimental error in ridge regression. *Journal of statistical Planning and Inference*. 9:367-374.

Vita

DATE AND PLACE OF BIRTH January 4, 1978 Ukiah, California,

EDUCATION

2006	Ph.D. Candidate in Statistics, University of Kentucky
2002	M.S. in Statistics University of Kentucky
2000	B.S. in Mathematics California State University, Chico

PROFESSIONAL EXPERIENCE

08/03-08/06	*Research Assistant* University of Kentucky, Statistics
06/01-08/03	*Teaching Assistant /Instructor* University of Kentucky, Statistics
08/00-06/01	*Teaching Assistant* University of Kentucky, Statistics

PRESENTATIONS

Jan., 2005 "The Weak Convergence of Empirical Processes from Multivariate Normal Vectors," Hawaii International Conference on Statistics, Honolulu, HI.

Aug., 2005 "The Weak Convergence of Empirical Processes from Multivariate Normal Vectors for Goodness –of-fit Tests," Joint Statistical Meeting, Minneapolis, MN.

March, 2006 "The Asymptotic Distribution of Modified Shapiro-Wilk Statistics for Testing Multvariate Normality," The Eastern North American Region of the International Biometric Society Spring Meeting, Tampa, FL.

PAPERS AND REPORTS

"A novel group of genes regulate susceptibility to anti-neoplastic drugs in highly tumorigenic breast cancer cells." Mallory JC, Crudden G, Oliva A, Saunders C, Stromberg A, Craven RJ. Mol Pharmacol. 2005 Sep 8.

"Host Gene Expression in Local Tissues in Response to Periodontal Pathogens." Ebersole J, Meka A, Arnold Stromberg A, Saunders C, and Kesavalu L. Oral Biosciences & Medicine. 2 (2/3): 175-184. 2005.

"Leptin regulates olfactory-mediated behavior in ob/ob mice." Thomas V. Getchell, Kevin Kwong, Christopher P. Saunders, Arnold J. Stromberg, and Marilyn L. Getchell, Physiol Behav. May 30;87(5):848-56. Epub 2006 Mar 20.

"Statistical and graphical identification of functional gene categories in microarray experiments." Hua Liu, Christopher P Saunders, Aaron S Borders, Thomas V Getchell, Marilyn L Getchell, Arne Bathke, and Arnold J Stromberg (Submitted)

"Gene Expression Patterns that Predict Sensitivity to EGFR Tyrosine Kinase Inhibitors in Lung Cancer Cell Lines: Implications for Prediction of Sensitivity in Human Lung Tumors." Justin M. Balko, Christopher Saunders, Arnold Stromberg, Eric B. Haura and Esther P. Black (Submitted)

Membership in Professional Societies

American Statistical Association
Institute of Mathematical Statistics