Statistica Sinica **13**(2003), 327-349

THE LIMIT DISTRIBUTION OF A TEST STATISTIC FOR BIVARIATE NORMALITY

Namhyun Kim and Peter J. Bickel

Hongik University and University of California, Berkeley

Abstract: Testing for normality has always been an important part of statistical methodology. In this paper we propose a test statistic for bivariate normality. We generalize the statistic proposed by de Wet and Venter to test bivariate normality using Roy's union-intersection principle. The generalized statistic is affine invariant. We show that the limit distribution of an approximation to the suggested statistic is representable as the supremum over an index set of the integral of a suitable Gaussian process. We also simulate the null distribution of the statistic, give some critical values of the distribution and make power comparisons to other procedures that have been proposed.

Key words and phrases: Bivariate normality, Brownian bridge, Gaussian process.

1. Introduction

Let $\mathbf{X}_1 = (X_{11}, X_{21})^T, \ldots, \mathbf{X}_n = (X_{1n}, X_{2n})^T$ be a sample of independent observations on a random 2-dimensional (column) vector $\mathbf{X} = (X_1, X_2)^T$, where T denotes transpose. Let us write $BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for a bivariate normal distribution with means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation coefficient ρ.

Most classical multivariate analysis techniques assume multivariate normality of a data set, and this is a natural assumption to test. In this paper we consider only bivariate normality H_0 : The law of $\mathbf{X}$ is $BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for some μ_1, μ_2, σ_1^2, σ_2^2, ρ. However it seems clear that our techniques should generalize.

In testing multivariate normality, a number of test procedures have been proposed in the literature. For a general review, see Gnanadesikan (1977), Mardia (1980), Cox and Small (1978), D'Agostino and Stephens (1986, Section 9.7). Mardia (1970, 1974, 1975), Mardia and Foster (1983), Malkovich and Afifi (1973) and Machado (1983) proposed multivariate measures of skewness and kurtosis and used them to develop tests for multinormality. Horsewell and Looney (1992) presented a comparison of tests that are based on measures of multivariate skewness and kurtosis. Csörgő (1986, 1989) proposed a test based on the empirical characteristic function. Baringhaus and Henze (1988) extended the approach of

Epps and Pulley (1983) to the multivariate case. Their test is based on a weighted integral of the squared modulus of the difference between the empirical characteristic function of the standardized residuals and its pointwise limit under the null hypothesis. This test was studied by Henze and Zirkler (1990), Henze and Wagner (1997). Zhu, Wong and Fang (1995) suggested a test for multinormality based on sample entropy and projection pursuit.

In this paper we begin by considering a test statistic for the simple hypothesis H_0^s: The law of $\mathbf{X}$ is $BVN(0,0,1,1,\rho)$,

$$P_n^o = \sup_{c_1,c_2.\ni.c_1^2+c_2^2+2\rho c_1c_2=1} \sum_{i=1}^n \left\{(c_1X_1+c_2X_2)_{(i)} - \Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2, \tag{1}$$

where $(\cdot)_{(i)}$ means the ith order statistic of the random variables in parentheses and Φ^{-1} denotes the inverse of the standard normal distribution function Φ. The natural extension of P_n^o for testing H_0 is

$$P_n = \sup_{c_1,c_2} \sum_{i=1}^n \left\{\frac{(c_1X_1+c_2X_2)_{(i)} - (c_1\bar{X}_1+c_2\bar{X}_2)}{sd(c_1X_1+c_2X_2)} - \Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2 \tag{2}$$

$$:= \sup_{c_1,c_2} \sum_{i=1}^n A(c_1,c_2)^2, \tag{3}$$

where $\bar{X}_k = n^{-1}\sum_{i=1}^n X_{ki}$, $sd^2(c_1X_1+c_2X_2) = c_1^2\hat{\sigma_1}^2 + c_2^2\hat{\sigma_2}^2 + 2c_1c_2\hat{\rho}\hat{\sigma}_1\hat{\sigma}_2$, $\hat{\sigma}_k^2 = n^{-1}\sum_{i=1}^n (X_{ki}-\bar{X}_k)^2$, $k=1,2$, and $\hat{\rho} = n^{-1}\sum_{i=1}^n (X_{1i}-\bar{X}_1)(X_{2i}-\bar{X}_2)/(\hat{\sigma}_1\hat{\sigma}_2)$. Note that P_n is affine invariant. Consequently, the distribution of P_n under H_0 does not depend on the parameters μ_1, μ_2, σ_1^2, σ_2^2, ρ. Hence in studying P_n and the approximation P_n^T to be introduced later, which is also affine invariant under H_0, it is enough to consider its behavior under H_0^s. It should be remarked that the statistics P_n^o and P_n can be readily generalized to the d-dimensional case for all $d \geq 2$. These statistics P_n^o, P_n generalize those introduced by de Wet and Venter (1972) for testing univariate normality using Roy's union-intersection principle (Roy (1953)).

Let $X_1,\ldots,X_n$ be independent and identically distributed (i.i.d.) random variables from a common distribution F. de Wet and Venter (1972) suggested a statistic $L_n^o = \sum_{i=1}^n (X_{(i)} - \Phi^{-1}(\frac{i}{n+1}))^2$ for testing the simple hypothesis $H_0^{\mathcal{S}}$: $F=\Phi$, and $L_n = \sum_{i=1}^n (\frac{X_{(i)}-\bar{X}}{S} - \Phi^{-1}(\frac{i}{n+1}))^2$ for testing the composite hypothesis $H_0^{\mathcal{C}}$: $F(x) = \Phi(\frac{x-\mu}{\sigma})$, where μ and σ are unknown. $X_{(1)},\ldots,X_{(n)}$ are the order statistics of $X_1,\ldots,X_n$ and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $S^2 = n^{-1}\sum_{i=1}^n (X_i-\bar{X})^2$. They proved (see de Wet and Venter (1972, 1973))

$$L_n^o - a_n^o \xrightarrow{\mathcal{D}} \sum_{i=1}^\infty \frac{Z_i^2-1}{i}, \tag{4}$$

where $\xrightarrow{\mathcal{D}}$ means convergence in distribution, $Z_1, Z_2, \ldots$ are i.i.d. standard normal random variables, and a_n^o, which is the approximate value of $E(L_n^o)$, is given by

$$a_n^o = \frac{1}{n}\sum_{i=1}^{n} \frac{i}{n+1}\left(1-\frac{i}{n+1}\right)/\phi^2\left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right) \quad (\phi(x) = d\Phi(x)/dx). \qquad (5)$$

They also showed the same result for the limit distribution of L_n. The L_n statistic is closely connected with Shapiro-Wilk's W statistic and Shapiro-Francia's W' statistic for testing normality. In fact, de Wet-Venter's L_n-statistic can be regarded as a simplified version of Shapiro-Francia's W' statistic, which is a modified version of Shapiro-Wilk's W statistic; see Shapiro and Wilk (1965), Shapiro and Francia (1972), de Wet and Venter (1972), D'Agostino and Stephens ((1986), Section 5.10). Moreover Shapiro-Wilk's W statistic and Shapiro-Francia's W' statistic are asymptotically equivalent to de Wet-Venter's L_n-statistic; see, Leslie, Stephens and Fotopolous (1986) for a rigorous proof.

Malkovich and Afifi (1973) and Fattorini (1986) generalized Shapiro-Wilk's W statistic to test multivariate normality based on Roy's union-intersection principle. Their procedures are briefly reviewed in Section 3. The generalized de Wet-Venter statistic P_n in (2) can be considered as a simplification of Malkovich and Afifi's (or Fattorini's) generalized Shapiro-Wilk statistic just as in the univariate case. Hence the results we give for the approximation P_n^T to P_n should be generalizable to the statistics of Malkovich and Afifi and Fattorini. Their limiting behavior is a well-known open question (see Henze and Zirkler (1990)).

In Section 2 we discuss the asymptotic distribution theory of P_n^o and P_n under H_0^s and H_0 respectively. In fact we represent the limit distribution as the supremum over an index set of the integral of a suitably defined Brownian bridge. In Section 3 we present the critical values of P_n and some power comparisons to other statistics.

2. The Limiting Null Distribution of P_n

Our goal is to characterize the asymptotic limit of P_n^o and P_n in (1), (2) as the supremum over an index set of the integral of a suitable Gaussian process. For this purpose we put $c_1 = \cos\theta - \frac{\rho}{\sqrt{1-\rho^2}}\sin\theta$, $c_2 = \frac{\sin\theta}{\sqrt{1-\rho^2}}$. Then P_n^o in (1) becomes

$$\begin{aligned} P_n^o &= \sup_{\theta\in[0,2\pi)} \sum_{i=1}^{n}\left\{\left(X_1\cos\theta + \frac{(X_2-\rho X_1)}{\sqrt{1-\rho^2}}\sin\theta\right)_{(i)} - \Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2 \\ &\stackrel{\mathcal{D}}{=} \sup_{\theta\in[0,2\pi)} \sum_{i=1}^{n}\left\{(Z_1\cos\theta + Z_2\sin\theta)_{(i)} - \Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2 \qquad (6) \end{aligned}$$

under H_0^s, where $X \overset{\mathcal{D}}{=} Y$ means X and Y have the same distribution and (Z_{1i}, Z_{2i}), $i = 1, \ldots, n$, are a random sample from $BVN(0,0,1,1,0)$. In the following, $B(y,\theta)$ is a Brownian bridge with the covariance structure

$$\begin{aligned}&\text{Cov}\,(B(y_1,\theta_1), B(y_2,\theta_2))\\ =&\Pr(Z_1\cos\theta_1+Z_2\sin\theta_1\le\Phi^{-1}(y_1)\ \text{ and }\ Z_1\cos\theta_2+Z_2\sin\theta_2\le\Phi^{-1}(y_2))-y_1y_2. \quad (7)\end{aligned}$$

In other words, $\text{Cov}\,(B(y_1,\theta_1),\ B(y_2,\theta_2)) + y_1y_2$ is bivariate normal probability of a quadrant under $BVN(0,0,1,1,\rho)$ with $\rho = \cos(\theta_1-\theta_2)$ and the covariance function of B is just that of the one dimensional Brownian bridge if $\theta_1 = \theta_2$.

For technical reasons we introduce truncated versions of P_n^o, P_n which we denote by P_n^{oT} and P_n^T respectively. They are given by

$$P_n^{oT} = \sup_{c_1,c_2.\ni.c_1^2+c_2^2+2\rho c_1c_2=1} \sum_{i=I_n}^{n-I_n} \left\{(c_1X_1+c_2X_2)_{(i)} - \Phi^{-1}\left(\frac{i}{n+1}\right)\right\}^2 \quad (8)$$

with $I_n = [n^{1-\delta}]$, $0<\delta<1/8$, and

$$P_n^T = \sup_{c_1,c_2} \sum_{i=I_n}^{n-I_n} A(c_1,c_2)^2 = \sup_{\theta\in[0,2\pi)} \sum_{i=I_n}^{n-I_n} A(\cos\theta,\sin\theta)^2, \quad (9)$$

where $A(c_1,c_2)$ defined in (3) denotes the term inside the parentheses on the right side of (2) and $\cos\theta = c_1/(c_1^2+c_2^2)^{1/2}$, $\sin\theta = c_2/(c_1^2+c_2^2)^{1/2}$.

Theorem 1. *Under the simple hypothesis H_0^s: The law of $\mathbf{X}$ is $BVN(0,0,1,1,\rho)$,*

$$P_n^{oT} - a_n^T \xrightarrow{\mathcal{D}} \sup_{\theta\in[0,2\pi)} \int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)}dy, \quad (10)$$

where a_n^T is given by

$$a_n^T = \frac{1}{n}\sum_{i=I_n}^{n-I_n} \frac{i}{n+1}\left(1-\frac{i}{n+1}\right)/\phi^2\left(\Phi^{-1}(\frac{i}{n+1})\right). \quad (11)$$

Conjecture 1. Under the simple hypothesis H_0^s: The law of $\mathbf{X}$ is $BVN(0,0,1,1,\rho)$,

$$P_n^o - a_n^o \xrightarrow{\mathcal{D}} \sup_{\theta\in[0,2\pi)} \int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)}dy,$$

where a_n^o is given by (5).

Theorem 2. *Under the composite hypothesis H_0, the statistic P_n^T in (9) has the following limit distribution,*

$$
\begin{aligned}
&P_n^T - a_n^T \\
&\xrightarrow{\mathcal{D}} \sup_{\theta\in[0,2\pi)} \Bigg[\int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)}dy \qquad (12)\\
&\qquad - \Big(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))}dy\Big)^2 - \Big(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))}\Phi^{-1}(y)dy\Big)^2\Bigg].
\end{aligned}
$$

Conjecture 2. Under the composite hypothesis H_0, the statistic $P_n - a_n^o$ has the limit distribution on the right side of (12).

A heuristic argument for the limit distributions in Theorems 1 and 2 is given below. The rigorous proofs are quite complicated. As already mentioned, de Wet and Venter (1972) represented the limit null distribution of their statistics as infinite series of standard normal variables. Recall that their statistics are a one-dimensional version of P_n^o and P_n. However their approach does not help much in finding the asymptotic distributions of P_n^o and P_n. On the other hand, del Barrio, Cuesta, Matrán and Rodríguez (1999) represented the limit distribution of de Wet-Venter's L_n statistic as an integrable variable in terms of a Brownian bridge, and Theorem 2 generalizes their result.

Let $F_n(x)$ be the empirical distribution function of a sample $X_1,\dots,X_n$ defined by $F_n(x)=\frac{1}{n}\sum_{i=1}^n I(X_i\le x)$. Define the nth sample quantile function $Q_n(y)$ by

$$
Q_n(y)=\begin{cases} X_{(k)} \text{ if } \frac{k-1}{n+1}<y\le\frac{k}{n+1}, & k=1,\dots,n,\\ X_{(n)} \text{ if } \frac{n}{n+1}<y\le 1, \end{cases}
$$

and the normed quantile process $\{\rho_n(y):0<y<1\}$ by $\rho_n(y)=\sqrt{n}\phi(\Phi^{-1}(y))$ $(Q_n(y)-\Phi^{-1}(y))$. Then we have

$$
\begin{aligned}
L_n^o - a_n^o &= \frac{1}{n}\sum_{i=1}^n \frac{\left(\rho_n(\frac{i}{n+1})\right)^2-\frac{i}{n+1}(1-\frac{i}{n+1})}{\phi^2\left(\Phi^{-1}(\frac{i}{n+1})\right)}\\
&\approx \int_{-\infty}^{\infty}\frac{\{\rho_n(F_n(x))\}^2-F_n(x)(1-F_n(x))}{\phi^2\left(\Phi^{-1}(F_n(x))\right)}dF_n(x). \qquad (13)
\end{aligned}
$$

By Theorems 4.5.6 and 4.5.7 in Csörgő and Révész (1981), one can define, on the same probability space, a Brownian bridge $\{B_n(y):0\le y\le 1\}$ for each n such that

$$
\sup_{\delta_n\le y\le 1-\delta_n}|\rho_n(y)-B_n(y)| \overset{a.s.}{=} \mathcal{O}(n^{-1/2}\log n),
$$

where $\delta_n = 25n^{-1}\log\log n$. Recall that a separable Gaussian process $\{B(y) : 0 \le y \le 1\}$ is called a Brownian bridge if $E(B(y)) = 0$ and $E(B(y_1)B(y_2)) = y_1 \wedge y_2 - y_1 y_2$. Hence the representation of L_n^o in (13) suggests we should have

$$L_n^o - a_n^o \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(y) - y(1-y)}{\phi^2(\Phi^{-1}(y))}\, dy. \tag{14}$$

Since the sequence

$$\left\{ \int_{1/n}^{1-1/n} \frac{B^2(y) - y(1-y)}{\phi^2(\Phi^{-1}(y))}\, dy \right\}_n \tag{15}$$

is an L_2-Cauchy sequence, the right hand side of (14) is defined as the L_2-limit of the sequence in (15). For a rigorous proof of (14), see del Barrio, Cuesta, Matrán and Rodríguez (1999). Also see Csörgő ((1983), Chapter 7). The right side of (4) would give the infinite series representation of the random variable on the right side of (14). By (14) and (6) we would expect Theorem 1 and Conjecture 1 to hold.

As will become apparent in the Appendix, we cannot establish these heuristics fully and the relation between the obvious generalizations of the Shapiro-Wilk and Shapiro-Francia statistics for the same reason. Estimates of and bounds for multinomial tail probabilities are not as simple or available as for bivariate Gaussians.

To prove these theorems we need a series of lemmas. These lemmas and the proofs of our theorems are given in the Appendix. We define an empirical process and quantile processes in $\mathbf{R}^2$ that are closely related to (6), and show the processes can be approximated by a sequence of Brownian bridges with covariance structure (7). For this purpose, the theorems in Massart (1989) and Adler (1990) play an important role.

Define, for $0 < y < 1$, $0 \le \theta < 2\pi$,

$$S := S(y,\theta) := \{(u_1,u_2) \in [0,1]^2 : \Phi^{-1}(u_1)\cos\theta + \Phi^{-1}(u_2)\sin\theta \le \Phi^{-1}(y)\}, \tag{16}$$

$$\Im := \{S(y,\theta) \subset [0,1]^2 : 0 < y < 1, 0 \le \theta < 2\pi\}. \tag{17}$$

Consider the empirical measure $\mathbf{P}_n(S)$, $S = S(y,\theta) \in \Im$,

$$\begin{aligned} \mathbf{P}_n(S) := \mathbf{P}_n(S(y,\theta)) &= \frac{1}{n}\sum_{i=1}^n I\left((U_{1i},U_{2i}) \in S(y,\theta)\right) \\ &= \frac{1}{n}\sum_{i=1}^n I\left(\Phi^{-1}(U_{1i})\cos\theta + \Phi^{-1}(U_{2i})\sin\theta \le \Phi^{-1}(y)\right) \end{aligned} \tag{18}$$

and the family of empirical processes

$$\alpha_n(S) := \alpha_n(S(y,\theta)) = \sqrt{n}\,(\mathbf{P}_n - P)\,(S(y,\theta)), \tag{19}$$

where I is an indicator function, P is the uniform distribution on $[0,1]^2$ and U_{1i}, U_{2i}, $i = 1,\ldots,n$, are independent and identically distributed (i.i.d.) uniform random variables on $(0,1)$. Since an element $S := S(y,\theta) \in \Im$ defined in (16) is determined by (y,θ), we can identify a set $S := S(y,\theta)$ with the parameters $(y,\theta) \in (0,1) \times [0,2\pi) := \Theta$ without loss of generality. We define $\mathbf{P}_n(S) := \mathbf{P}_n(S(y,\theta)) := \mathbf{P}_n(y,\theta)$, $\alpha_n(S) := \alpha_n(S(y,\theta)) := \alpha_n(y,\theta)$. We see that P_n^o in (6) can be expressed in terms of a family of quantile processes that are closely related to empirical process α_n in (19). Therefore we need to show the empirical process $\alpha_n(\cdot,\cdot)$ can be approximated by a Gaussian process to prove Theorem 1 rigorously. Massart (1989) provided an essential tool for doing this.

According to Theorem 1 in Massart (1989), if a given collection of Borel sets $\Im^d$ in $\mathbf{R}^d$ is not too large in a suitable sense, then the centered and normalized empirical process with n independent observations with common law μ, the uniform distribution on the unit cube in $\mathbf{R}^d$, can be strongly approximated by a sequence of Brownian bridges indexed by $\Im^d$. A Gaussian process indexed by $\Im^d$, $\{B(S) : S \in \Im^d\}$, is called a Brownian bridge indexed by $\Im^d$ if $E(B(S)) = 0$, $E(B(S_1)B(S_2)) = P(S_1 \cap S_2) - P(S_1)P(S_2)$, $S_i \in \Im^d$, $i = 1,2$. Massart (1989) gave two conditions on $\Im^d$ for the strong invariance principle to hold. One is the uniform Minkowski condition, which is the smoothness condition on the boundaries of $S \in \Im^d$, and the other is a reasonable growth condition on $\Im^d$. He denoted the second condition by $H(\zeta)$, $0 \le \zeta < 1$. By applying Theorem 1 in Massart (1989), we have the following result.

Lemma 1. *The collection of sets $\Im$ in $\mathbf{R}^2$ defined in* (17) *satisfies the uniform Minkowski condition and $H(\zeta)$ with $\zeta = 1/2$ in Massart* (1989). *Hence, for the empirical process $\{\alpha_n(S) : S \in \Im\}$ in* (19), *there exists a sequence of Brownian bridges $\{B_n(S) : S \in \Im\}$ satisfying*

$$\sup_{S\in\Im} |\alpha_n(S) - B_n(S)| \overset{a.s.}{=} \mathcal{O}(n^{-1/8}\log n). \tag{20}$$

The centered Brownian bridge $B_n(S)$ has the covariance structure given in (7).

Now we define a uniform quantile process u_n corresponding to the empirical process α_n in (19) and establish a weak invariance principle for u_n. Let $U_n(y,\theta)$ be the uniform quantile function defined by

$$U_n(y,\theta) := U_n(S(y,\theta))$$

$$:= \left(\Phi\left(\Phi^{-1}(U_1)\cos\theta + \Phi^{-1}(U_2)\sin\theta\right)\right)_{(k)} := U_{(k)}(\theta) \tag{21}$$
$$\text{if } \frac{k-1}{n} < y \le \frac{k}{n}, \quad k = 1, \ldots, n,$$

where $(\cdot)_{(k)}$ denotes the kth order statistic of the random variables in parentheses, and $u_n(y, \theta)$ be the uniform quantile process defined by

$$u_n(y, \theta) := u_n(S(y, \theta)) := \sqrt{n}(U_n(y, \theta) - y). \tag{22}$$

Lemma 2. *For the quantile process* $u_n(\cdot)$ *defined in* (22), *there exists a sequence of Brownian bridges* B'_n *indexed by* $\Im$ *or* Θ *satisfying*

$$\sup_{(y,\theta)\in\Theta} |u_n(y, \theta) - B'_n(y, \theta)| = \mathcal{O}_p\left(n^{-1/8}\log n\right).$$

In fact, $B'_n(\cdot) = -B_n(\cdot)$ *with* $B_n(\cdot)$ *in* (20).

Define the quantile function $Q_n(y, \theta)$ by

$$Q_n(y,\theta) := \begin{cases} \left(\Phi^{-1}(U_1)\cos\theta + \Phi^{-1}(U_2)\sin\theta\right)_{(k)} := X_{(k)}(\theta) \\ \qquad \text{if } \frac{k-1}{n+1} < y \le \frac{k}{n+1}, \quad k = 1, \ldots, n, \\ \left(\Phi^{-1}(U_1)\cos\theta + \Phi^{-1}(U_2)\sin\theta\right)_{(n)} := X_{(n)}(\theta) \\ \qquad \text{if } \frac{n}{n+1} < y \le 1 \end{cases} \tag{23}$$

and the normed quantile process $\rho_n(y, \theta)$ by

$$\rho_n(y, \theta) := \phi\left(\Phi^{-1}(y)\right)\sqrt{n}\left(Q_n(y, \theta) - \Phi^{-1}(y)\right). \tag{24}$$

The following lemma establishes a weak invariance principle for ρ_n.

Lemma 3. *For* $0 < \delta < 1/4$,

$$\sup_{n^{-\delta}\le y\le 1-n^{-\delta},\, \theta\in[0,2\pi)} |\rho_n(y, \theta) - u_n(y, \theta)| = \mathcal{O}_p\left(n^{-(1/2-2\delta)}\right) = o_p(1).$$

Proof. The proof is almost the same as that of Theorem 4.5.6 in Csörgő and Révész (1981). See also Kim (1994) for details.

Lemma 4. *For the normed quantile process* $\rho_n(\cdot)$ *defined in* (24), *one can define a version of Brownian bridge* B'_n *indexed by* $\Im$ *or* Θ *for each* n *such that*

$$\sup_{n^{-\delta}\le y\le 1-n^{-\delta},\, \theta\in[0,2\pi)} |\rho_n(y, \theta) - B'_n(y, \theta)| = \mathcal{O}_p\left(n^{-1/8}\log n\right),$$

for any $0 < \delta \le 3/16$.

Proof. It follows immediately from Lemmas 2 and 3.

3. Simulation Studies

A Monte Carlo experiment was performed to determine approximate upper percentiles of the null distribution of P_n in (2), and to study the power of the test based on P_n. The critical values are given in Table 1 for sample sizes $n = 20$, 30, 50, 100 and the usual significance levels $\alpha = 0.01$, 0.05 and 0.10. Each empirical percentage point is based on 5000 pseudo-random realizations of P_n. Pseudo random numbers were generated using S-plus version 3.2. The results indicate somewhat slow convergence of the quantiles of P_n especially in the upper tail ($\alpha = 0.01$).

Table 1. Simulated critical values k_α of the statistic $P_n : \Pr(P_n - a_n^o \geq k_\alpha) = \alpha$.

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
20	0.4073	0.7897	1.7304
30	0.6089	1.0388	1.9452
50	0.7678	1.2270	2.4464
100	0.9617	1.4905	2.7415

When we take the supremum of the P_n statistic, we can assume $c_1^2 + c_2^2 = 1$ without loss of generality. Hence we put $\cos\theta = c_1/(c_1^2 + c_2^2)^{1/2}$, $\sin\theta = c_2/(c_1^2 + c_2^2)^{1/2}$ as in (9), and P_n becomes

$$P_n = \sup_{c_1,c_2 \;.\ni.\; c_1^2+c_2^2=1} \sum_{i=1}^{n} A(c_1, c_2)^2 = \sup_{\theta\in[0,2\pi)} \sum_{i=1}^{n} A(\cos\theta, \sin\theta)^2.$$

Now it is enough to change θ from 0 to 2π and θ is varied in increments of $2\pi/720$ in the simulation.

For sample sizes $n = 20$, 50, the power of P_n is investigated under the testing level $\alpha = 0.05$. A thousand Monte Carlo samples were generated from each of various alternative bivariate distributions. Henze and Zirkler (1990) performed a simulation experiment to assess the power performance of their proposed test in comparison with other invariant procedures for testing multinormality. The procedures compared are Mardia's (1970), Malkovich and Afifi's (1973), and Fattorini's (1986) test. The alternative distributions included in Table 2 are a part of their studies. They are (i) distributions with independent marginals, (ii) mixtures of normal distributions. The following notations are used. $N(0,1)$, $C(0,1)$, $Logis(0,1)$ and $\exp(1)$ denote the standard normal, Cauchy, logistic and exponential distributions; χ_k^2 and t_k are the Chi-square, student's t-distribution with k degrees of freedom; $\Gamma(a,b)$ is the Gamma distribution with density $b^{-a}\Gamma(a)^{-1}x^{a-1}\exp(-x/b)$, $x > 0$; $B(a,b)$ stands for the beta distribution with density $B(a,b)^{-1}x^{a-1}(1-x)^{b-1}$, $0 < x < 1$; $LN(a,b)$ is for the log-normal distribution with density $(\sqrt{2\pi}bx)^{-1}\exp(-(\log x - a)^2/2b^2)$, $x > 0$. $F_1 * F_2$

is the distribution having independent marginal distributions F_1 and F_2. The product of two independent copies of F_1 is denoted by F_1^2. $NMIX_2(\kappa, \delta, \rho_1, \rho_2)$ is the bivariate normal mixture $\kappa BVN(0, 0, 1, 1, \rho_1) + (1 - \kappa)BVN(\delta, \delta, 1, 1, \rho_2)$.

Table 2. Percentage of 1000 Monte Carlo samples declared significant by the test based on P_n, MA and FA for bivariate normality ($\alpha = 0.05$).

	$n = 20$			$n = 50$		
alternative	P_n	MA	FA	P_n	MA	FA
$N(0,1)^2$	5	5	5	5	5	5
$\exp(1)^2$	81	76	86	100	100	100
$LN(0,.5)^2$	54	53	59	94	92	97
$C(0,1)^2$	98	96	96	100	–	–
$\Gamma(5,1)^2$	24	22	25	56	58	67
$(\chi_5^2)^2$	37	43	44	87	84	93
$(\chi_{15}^2)^2$	18	18	17	42	42	46
$(t_2)^2$	72	69	68	97	94	95
$(t_5)^2$	28	24	22	55	46	40
$B(1,1)^2$	1	2	6	6	4	77
$B(1,2)^2$	12	9	19	42	35	86
$B(2,2)^2$	1	2	3	1	2	15
$Logis(0,1)^2$	15	16	15	30	21	16
$N(0,1) * \exp(1)$	58	52	63	98	87	99
$N(0,1) * \chi_5^2$	25	26	28	65	61	73
$N(0,1) * t_5$	16	16	16	34	24	19
$N(0,1) * B(1,1)$	3	4	6	4	4	56
$NMIX_2(.5, 2, 0, 0)$	4	4	4	4	4	17
$NMIX_2(.5, 4, 0, 0)$	14	4	51	95	5	100
$NMIX_2(.5, 2, .9, 0)$	31	27	29	68	54	66
$NMIX_2(.5, .5, .9, 0)$	23	21	20	48	33	29
$NMIX_2(.5, .5, .9, -.9)$	51	47	51	93	76	83

Each number in Table 2 represents the percentage of 1000 Monte Carlo samples declared significant by the test based on P_n, rounded to the next integer. For comparison, we also put the power of Malkovich and Afifi's (MA) generalized Shapiro-Wilk's W statistic and that of Fattorini's adduced from Henze and Zirkler (1990). The Shapiro-Wilk's W statistic for testing univariate normality is

$$W(Z_1, \ldots, Z_n) = \frac{[\sum a_j(Z_{(j)} - \bar{Z})]^2}{\sum(Z_j - \bar{Z})^2}, \tag{25}$$

where $Z_{(j)}$'s are the univariate order statistics of $Z_1, \ldots, Z_n$, $\bar{Z} = n^{-1} \sum Z_j$, and a_j's are the coefficients tabulated in Shapiro and Wilk (1965). The test of Malkovich and Afifi (1973) accepts the hypothesis of multivariate normality if

$$\min_{\mathbf{c}} W(\mathbf{c}^T \mathbf{X}_1, \ldots, \mathbf{c}^T \mathbf{X}_n) \geq K_w, \tag{26}$$

where K_w is a constant. For numerical evaluation of the minimization in (26), Malkovich and Afifi (1973) proposed an approximate solution based on the observation that $W(\mathbf{c}^T \mathbf{X}_1, \ldots, \mathbf{c}^T \mathbf{X}_n)$ has a lower bound when $\mathbf{c}$ satisfies the conditions (Shapiro and Wilk (1965)):

$$\mathbf{c}^T (\mathbf{X}_l - \bar{\mathbf{X}}) = \frac{n-1}{na_1}, \quad \mathbf{c}^T (\mathbf{X}_j - \bar{\mathbf{X}}) = -\frac{1}{na_1}, \quad j = 1, \ldots, n, \ j \neq l,$$

where $\bar{\mathbf{X}}$ is the sample mean vector. Since a solution $\mathbf{c}$ to these equations does not exist, Malkovich and Afifi proposed to find a vector $\mathbf{c}$ which minimizes

$$\left[\mathbf{c}^T (\mathbf{X}_l - \bar{\mathbf{X}}) - \frac{n-1}{na_1}\right]^2 + \sum_{j \neq l}^{n} \left[\mathbf{c}^T (\mathbf{X}_j - \bar{\mathbf{X}}) + \frac{1}{na_1}\right]^2.$$

This vector is $\mathbf{c}^{(l)} = \frac{1}{a_1} A^{-1} (\mathbf{X}_l - \bar{\mathbf{X}})$ with $A = \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T$. Since $l \in \{1, \ldots, n\}$ is arbitrary, they proposed to choose $\mathbf{c}^{(m)} \in \{\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(n)}\}$ such that the denominator of $W(\mathbf{c})$ is maximized over these n choices, i.e., $(\mathbf{X}_m - \bar{\mathbf{X}})^T A^{-1} (\mathbf{X}_m - \bar{\mathbf{X}}) = \max_{1 \leq l \leq n} (\mathbf{X}_l - \bar{\mathbf{X}})^T A^{-1} (\mathbf{X}_l - \bar{\mathbf{X}})$. Hence their test statistic is

$$\begin{aligned} MA(\mathbf{X}_1, \ldots, \mathbf{X}_n) &= W({\mathbf{c}^{(m)}}^T \mathbf{X}_1, \ldots, {\mathbf{c}^{(m)}}^T \mathbf{X}_n) \\ &= \frac{[\sum_{j=1}^{n} a_j U_{(j)}]^2}{(\mathbf{X}_m - \bar{\mathbf{X}})^T A^{-1} (\mathbf{X}_m - \bar{\mathbf{X}})}, \end{aligned}$$

where $U_{(1)} \leq \cdots \leq U_{(n)}$ are the order statistics of $U_j = (\mathbf{X}_m - \bar{\mathbf{X}})^T A^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})$, $j = 1, \ldots, n$.

Since $MA(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ does not even minimize $W(\mathbf{c}^T \mathbf{X}_1, \ldots, \mathbf{c}^T \mathbf{X}_n)$ with respect to the n possible solutions $\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(n)}$, Fattorini (1986) proposed to consider the statistic $FA(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \min_{1 \leq l \leq n} W({\mathbf{c}^{(l)}}^T \mathbf{X}_1, \ldots, {\mathbf{c}^{(l)}}^T \mathbf{X}_n)$. Both the MA and the FA test reject the hypothesis of multivariate normality for small values of the test statistics. In fact $FA \leq MA$.

The conclusion that can be drawn from the power study in Table 2 is that, for sample size $n = 50$, MA is generally inferior to the P_n statistic and P_n is generally inferior to the procedure FA of Fattorini, especially against alternatives with shorter tailed marginals like $B(1,1)^2$, $B(1,2)^2$, $B(2,2)^2$ or $N(0,1) * B(1,1)$. However FA is slightly inferior to P_n against alternatives with symmetric longer tailed marginals like $(t_5)^2$, $Logis(0,1)^2$, $N(0,1) * t_5$. According to Shapiro and Francia (1972) and Looney and Gulledge (1985), the power comparison between Shapiro-Wilk's W test and Shapiro-Francia's W' test for testing univariate normality indicates that the W test is more powerful than the W' test especially against symmetric alternatives with shorter tails than normal. And the W' test appeared to be more sensitive than the W test when the alternative was continuous and symmetric with longer tails than normal. Since de Wet-Venter's L_n statistic is a simplified version of W', L_n and W' are expected to have similar behavior in power. The P_n statistic in (2) generalizes the L_n statistic to bivariate cases as we mentioned in Section 1. Thus the power results for dimension 2 are consistent with those for dimension 1.

4. Appendix. Proofs of Theorems

Proof of Lemma 1. The boundary of a set $S := S(y,\theta) \in \Im$, ∂S, is

$$\begin{aligned}\partial S &= \Big\{(u_1,u_2) : \Phi^{-1}(u_1)\cos\theta + \Phi^{-1}(u_2)\sin\theta = \Phi^{-1}(y)\Big\} \\ &= \Big\{(u_1,u_2) : u_1 = \Phi\Big(\frac{\Phi^{-1}(y)}{\cos\theta} - \Phi^{-1}(u_2)\,\tan\theta\Big) := f(u_2)\Big\}.\end{aligned}$$

Since the function $f(u_2)$ is a monotone function on (0,1), we easily see that the uniform Minkowski condition (the UM condition) is satisfied. Note that the UM condition is that for the class $\Im$, there exists a constant K such that $P((\partial S)^\epsilon) \le K\epsilon$ for any $\epsilon \in (0,1]$ and any $S \in \Im$, where A^ϵ denotes the set $A^\epsilon = \{y \in [0,1]^2 : |y-z| < \epsilon \text{ for some } z \in A\}$ for any $\epsilon(0,1]$. Since the length of ∂S for any fixed y and θ is $\int_0^1 \sqrt{\{f'(u_2)\}^2+1}du_2 \le \int_0^1 (|f'(s)|+1)\,ds = 2$, we have $P((\partial S)^\epsilon) \le 4\epsilon$ for any $S \in \Im$ and $\epsilon \in (0,1]$ and the UM condition is satisfied.

To show the condition $H(\zeta)$ for $\zeta = 1/2$, which is a minimal cardinality condition on the set $\Im$, recall that $S(y,\theta) \in \Im$ is determined by $(y,\theta) \in \Theta = (0,1) \times [0,2\pi)$. Hence it is enough to consider the cardinality of the set $\Theta = (0,1) \times [0,2\pi)$ instead of that of the set $\Im$. Although the rigorous proof is long and tedious, the basic idea is to show

$$P\left(S(y_1,\theta_1)\Delta S(y_2,\theta_2)\right) \le C\delta, \tag{27}$$

whenever $|y_1-y_2| \le \delta$ and $|\theta_1-\theta_2| \le \delta$, where Δ means the symmetric difference of two sets. Using (27), we can show $H(\zeta)$ is satisfied with $\zeta = 1/2$. In fact, $\mathcal{N}(\epsilon,\Im) \le (M/\epsilon)^2 \le e^{2\sqrt{M/\epsilon}} = e^{K\epsilon^{-1/2}}$, where $\mathcal{N}(\epsilon,\Im)$ denotes the minimal cardinality of a collection $\Im(\epsilon)$ of Borel sets such that for any $S \in \Im$ there exist $S^-(\epsilon)$, $S^+(\epsilon) \in \Im(\epsilon)$ with $S^-(\epsilon) \subseteq S \subseteq S^+(\epsilon)$ and $P(S^+(\epsilon) - S^-(\epsilon)) \le \epsilon$. For details of the rigorous proof, see Kim (1994).

To prove Lemma 2, we need to find the size of the modulus of continuity of a Brownian bridge B indexed by $\Im$ or Θ. The following lemmas are devoted to this goal.

Lemma 5. *For a fixed* $(y_1, \theta_1) \in \Theta$ *and* $0 < \eta \le \pi/2$, *let* $\Delta(\eta) := \Delta((y_1, \theta_1), \eta) := \{(y_2, \theta_2) \in \Theta : y_1 < y_2 < y_1 + \eta, \theta_1 < \theta_2 < \theta_1 + \eta\}$. *Then for a Brownian bridge* B *indexed by* Θ,

$$\sup_{(y_2,\theta_2)\in\Delta((y_1,\theta_1),\eta)} E\left(B(y_1, \theta_1) - B(y_2, \theta_2)\right)^2 \le C\eta$$

is satisfied for a constant C, *i.e.*,

$$\sup_{(y_2,\theta_2)\in\Delta((y_1,\theta_1),\eta)} d\left((y_1, \theta_1), (y_2, \theta_2)\right) \le \sqrt{C\eta}, \tag{28}$$

where d *is a canonical metric defined by*

$$d\left((y_1, \theta_1), (y_2, \theta_2)\right) := \left[E\left(B(y_1, \theta_1) - B(y_2, \theta_2)\right)^2\right]^{1/2}. \tag{29}$$

Proof. Let $(\xi_1^{(r)}, \xi_2^{(r)})$ be $BVN(0, 0, 1, 1, r)$ with joint density function $\phi(x_1, x_2; r)$. It is known that

$$\frac{\partial \phi(x_1, x_2; r)}{\partial r} = \frac{\partial^2 \phi(x_1, x_2; r)}{\partial x_1 \partial x_2}. \tag{30}$$

See for example Johnson and Kotz ((1972), Chapter 35). Assume $y_1 < y_2$ and $\rho = \cos(\theta_2 - \theta_1) > 0$. By (7) and (30),

$$\begin{aligned}
&E\left(B(y_1, \theta_1)B(y_2, \theta_2)\right) + y_1 y_2 = \Pr(\xi_1^{(\rho)} \le \Phi^{-1}(y_1), \xi_2^{(\rho)} \le \Phi^{-1}(y_2)) \\
&:= G(\rho) = G(1) - \int_\rho^1 G'(r)dr \\
&= y_1 - \int_\rho^1 \int_{-\infty}^{\Phi^{-1}(y_2)} \int_{-\infty}^{\Phi^{-1}(y_1)} \frac{\partial^2 \phi}{\partial x_1 \partial x_2} dx_1 dx_2 dr \\
&= y_1 - \int_\rho^1 \phi(\Phi^{-1}(y_1), \Phi^{-1}(y_2); r)\, dr \ge y_1 - \int_\rho^1 \frac{1}{2\sqrt{1 - r^2}} dr \\
&\ge y_1 - \int_\rho^1 \frac{1}{2\sqrt{1 - r}} dr = y_1 - \sqrt{1 - \rho} \\
&= y_1 - \sqrt{1 - \cos(\theta_2 - \theta_1)} \ge y_1 - |\theta_2 - \theta_1|.
\end{aligned}$$

Therefore $E(B(y_1, \theta_1) - B(y_2, \theta_2))^2 \le (y_2 - y_1) + 2(\theta_2 - \theta_1)$ and the lemma is proved.

Lemma 6. *For a Brownian bridge* B *indexed by* Θ, $N_d(\epsilon) \le K\epsilon^{-4}$, *where* $N_d(\epsilon)$ *denotes the smallest number of closed* d*-balls of radius* ϵ *that cover* Θ, d *the canonical metric defined in* (29).

Proof. By (28), one way to cover Θ by d-balls of radius ϵ is with the $(2 + C/\epsilon^2)(2 + 2\pi C/\epsilon^2)$ balls of radius ϵ^2/C centered at points of the form $(i_1\epsilon^2/C, i_2\epsilon^2/C)$, $i_1 = 0, 1, \ldots, [C/\epsilon^2]$, $i_2 = 0, 1, \ldots, [2\pi C/\epsilon^2]$. Hence we have $N_d(\epsilon) \leq (2 + C/\epsilon^2)(2 + 2\pi C/\epsilon^2) \leq K\epsilon^{-4}$.

Lemma 7. *A Brownian bridge B indexed by Θ is a.s. bounded on Θ.*

Proof. By Theorem 5.3 in Adler (1990) and Lemma 6, there exists a finite positive constant C such that for all $\lambda > 0$,

$$\Pr\Big(\sup_{(y,\theta)\in\Theta} B(y,\theta) > \lambda\Big) \leq C\lambda^4\Big(1 - \Phi\left(\lambda/\sigma_\Theta\right)\Big), \tag{31}$$

with $\sigma_\Theta^2 := \sup_\Theta E(B(y,\theta))^2$. Since $\sigma_\Theta^2 = \sup_\Theta y(1-y) \leq 1/4$, (31) implies the a.s. boundedness of B on Θ.

Lemma 8. *If we define a metric τ on Θ by $\tau((y_1,\theta_1),(y_2,\theta_2)) = \sup(|y_1 - y_2|, |\theta_1 - \theta_2|)$, then there exists an a.s. finite random variable $\delta = \delta(\omega)$ such that, for almost all ω,*

$$\sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} |B(y_1,\theta_1) - B(y_2,\theta_2)| \leq C\eta$$

for $\eta \leq \delta(\omega)$.

Proof. By Theorem 5.3 in Adler (1990) and Lemma 6, there exists a finite positive constant C such that for all $\lambda > 0$,

$$\Pr\Big(\sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} \Big(B(y_1,\theta_1) - B(y_2,\theta_2)\Big) > \lambda\Big) \leq C\lambda^4\Big(1 - \Phi\left(\lambda/\sigma_\eta\right)\Big), \tag{32}$$

where $\sigma_\eta^2 := \sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} E\left(B(y_1,\theta_1) - B(y_2,\theta_2)\right)^2$. Hence

$$\begin{aligned}
&E \sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} (B(y_1,\theta_1) - B(y_2,\theta_2)) \\
&\leq \int_0^\infty \Pr\Big(\sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} (B(y_1,\theta_1) - B(y_2,\theta_2)) > \lambda\Big) d\lambda \\
&\leq C\int_0^\infty \lambda^4\left(1 - \Phi(\lambda/\sigma_\eta)\right) d\lambda \leq C\int_0^\infty \lambda^4 \frac{\phi(\lambda/\sigma_\eta)}{\lambda/\sigma_\eta} d\lambda \\
&= C\sigma_\eta^5 \int_0^\infty \lambda^3\phi(\lambda)d\lambda \\
&= C\Big(\sup_{\tau((y_1,\theta_1),(y_2,\theta_2))<\eta} E\left(B(y_1,\theta_1) - B(y_2,\theta_2)\right)^2\Big)^{\frac{5}{2}} \\
&\leq C\eta, \quad \text{for} \quad 0 < \eta < 1,
\end{aligned} \tag{33}$$

by (32) and Lemma 5. Note that the generic constants C's appearing at different places need not be the same. By Lemma 7 and (33), all the conditions in Theorem 4.6 in Adler (1990) are satisfied and the lemma is proved.

Proof of Lemma 2. From the definition of $\boldsymbol{P}_n(\cdot)$ in (18) and the uniform quantile function $U_n(\cdot)$ in (21), $|\boldsymbol{P}_n(U_n(y,\theta),\theta) - y| \le 1/n$ and $u_n(y,\theta) = \sqrt{n}(U_n(y,\theta)-y) = -\sqrt{n}(\boldsymbol{P}_n(U_n(y,\theta),\theta)-U_n(y,\theta))+\sqrt{n}(\boldsymbol{P}_n(U_n(y,\theta),\theta)-y) = -\alpha_n(U_n(y,\theta),\theta)+\mathcal{O}(1/\sqrt{n})$. Hence

$$\begin{aligned}
&\sup_{(y,\theta)\in\Theta} |u_n(y,\theta) - B_n'(y,\theta)| \\
&\le \sup_{(y,\theta)\in\Theta} |\alpha_n\left(U_n(y,\theta),\theta\right) - \alpha_n(y,\theta)| + \sup_{(y,\theta)\in\Theta} |\alpha_n(y,\theta) - B_n(y,\theta)| + \mathcal{O}\left(1/\sqrt{n}\right) \\
&\le \sup_{(y,\theta)\in\Theta} |B_n\left(U_n(y,\theta),\theta\right) - B_n(y,\theta)| + \mathcal{O}\left(n^{-1/8}\log n\right) \quad a.s. \qquad (34)
\end{aligned}$$

by Lemma 1. Using Lemma 8, it follows

$$\begin{aligned}
&\sup_{(y,\theta)\in\Theta} |B_n(U_n(y,\theta),\theta) - B_n(y,\theta)| \le C/\sqrt{n} \sup_{(y,\theta)\in\Theta} |\sqrt{n}(U_n(y,\theta)-y)| \\
&= C/\sqrt{n} \sup_{(y,\theta)\in\Theta} |u_n(y,\theta)| = \mathcal{O}_p(1/\sqrt{n}) \qquad (35)
\end{aligned}$$

and the result follows by (34) and (35).

Lemma 9. *Under the simple hypothesis H_0^s: The law of $\mathbf{X}$ is $BVN(0,0,1,1,\rho)$,*

$$\left| (P_n^{oT} - a_n^T) - \sup_{\theta\in[0,2\pi)} \int_{I_n/n}^{1-I_n/n} \frac{B_n^2(y,\theta) - y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)} dy \right| \xrightarrow{p} 0$$

for I_n such that $I_n/n = n^{-\delta}$ with $0<\delta<1/8$, where P_n^{oT} is given in (8) *and a_n^T is in* (11).

Proof. By the same equation in (6),

$$P_n^{oT} \stackrel{\mathcal{D}}{=} \sup_{\theta\in[0,2\pi)} \sum_{i=I_n}^{n-I_n} \left\{ \left(\Phi^{-1}(U_1)\cos\theta + \Phi^{-1}(U_2)\sin\theta\right)_{(i)} - \Phi^{-1}\left(\frac{i}{n+1}\right) \right\}^2,$$

where U_{1i}, U_{2i}'s are i.i.d. random variables with the uniform $(0,1)$. By the definition of the normed quantile process $\rho_n(\cdot)$ in (24), P_n^{oT} can be written as

$$P_n^{oT} - a_n^T \stackrel{\mathcal{D}}{=} \sup_{\theta\in[0,2\pi)} \frac{1}{n} \sum_{i=I_n}^{n-I_n} \frac{\left(\rho_n\left(\frac{i}{n+1},\theta\right)\right)^2 - \frac{i}{n+1}\left(1-\frac{i}{n+1}\right)}{\phi^2\left(\Phi^{-1}(\frac{i}{n+1})\right)}.$$

By Lemma 4 and $\phi(\Phi^{-1}(y)) \approx y\sqrt{2|\log y|}$ as $y \to 0$,

$$\begin{aligned}
&\sup_{\theta\in[0,2\pi)} \sum_{i=I_n}^{n-I_n} \frac{\left|\left(\rho_n\left(\frac{i}{n+1},\theta\right)\right)^2 - B_n^2\left(\frac{i}{n+1},\theta\right)\right|}{n\phi^2\left(\Phi^{-1}(\frac{i}{n+1})\right)} \\
&= \sup_{\theta\in[0,2\pi)} \sum_{i=I_n}^{n-I_n} \frac{\left|\rho_n\left(\frac{i}{n+1},\theta\right) - B_n\left(\frac{i}{n+1},\theta\right)\right| \left|\rho_n\left(\frac{i}{n+1},\theta\right) + B_n\left(\frac{i}{n+1},\theta\right)\right|}{n\phi^2\left(\Phi^{-1}(\frac{i}{n+1})\right)} \\
&\le \sup_{n^{-\delta}\le y\le 1-n^{-\delta},\,\theta\in[0,2\pi)} |\rho_n(y,\theta) - B_n(y,\theta)| \\
&\quad \Big(\sup_{n^{-\delta}\le y\le 1-n^{-\delta},\,\theta\in[0,2\pi)} |\rho_n(y,\theta) - B_n(y,\theta)| + \sup_{(y,\theta)\in\Theta} 2|B_n(y,\theta)|\Big) \\
&\quad \sum_{i=I_n}^{n-I_n} \frac{1}{n\phi^2(\Phi^{-1}(\frac{i}{n+1}))} \\
&= \mathcal{O}_p(n^{-\frac{1}{8}}\log n)(\mathcal{O}_p(n^{-\frac{1}{8}}\log n) + \mathcal{O}_p(1)) \int_{n^{-\delta}}^{1-n^{-\delta}} 1/\phi^2(\Phi^{-1}(y))dy \\
&= \mathcal{O}_p(n^{-\frac{1}{8}}\log n)\mathcal{O}(n^{\delta}\log\log n) \\
&= \mathcal{O}_p(n^{-(\frac{1}{8}-\delta)}(\log n)(\log\log n)) \\
&= o_p(1) \qquad for \quad 0<\delta<1/8.
\end{aligned}$$

By Lemma 7, $B^2(y,\theta)$ is a.s. bounded on $[0,2\pi)$. Hence

$$\sup_{\theta\in[0,2\pi)} \int_{I_n/n}^{1-I_n/n} \frac{B_n^2(y,\theta)-y(1-y)}{\phi^2(\Phi^{-1}(y))}dy \le \int_{I_n/n}^{1-I_n/n} \sup_{\theta\in[0,2\pi)} \frac{B_n^2(y,\theta)-y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)}dy < \infty$$

and the lemma is proved.

Proof of Theorem 1. We show the existence of the integral $\sup_{\theta\in[0,2\pi)} \int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2(\Phi^{-1}(y))}dy$, suitably defined. Then it will follow that

$$(P_n^{oT} - a_n^T) \xrightarrow{\mathcal{D}} \sup_{\theta\in[0,2\pi)} \int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2\left(\Phi^{-1}(y)\right)}dy$$

by Lemma 9. del Barrio, et al. (1999) showed that $\int_0^1 \frac{B^2(y,0)-y(1-y)}{\phi^2(\Phi^{-1}(y))}dy$ is defined as an L_2-limit. To prove the existence, it is enough to show in view of the del Barrio et al. result that

$$\sup_{|\theta_1-\theta_2|<\eta} E\left(\int_0^{1/2} \frac{B^2(y,\theta_1)-B^2(y,\theta_2)}{\phi^2(\Phi^{-1}(y))}dy\right)^2 < C\eta^{1+\epsilon}$$

by Billingsley (1968, Theorem 12.3).

Now let us show

$$E\left(\int_0^\epsilon \frac{B^2(y,\theta+\eta)-B^2(y,\theta)}{\psi^2(y)}dy\right)^2 < C\eta^2|\log\eta| \tag{36}$$

with $\psi(y)=\phi(\Phi^{-1}(y))$. We can write

$$\begin{aligned}
&\int_0^\epsilon \frac{B^2(y,\theta+\eta)-B^2(y,\theta)}{\psi^2(y)}dy\\
&=\int_0^\epsilon \frac{((B(y,\theta+\eta)-B(y,\theta))^2-2\Pr(Z\le\Phi^{-1}(y),W(\eta)>\Phi^{-1}(y))}{\psi^2(y)}dy\\
&\quad -2\int_0^\epsilon \frac{B(y,\theta)(B(y,\theta)-B(y,\theta+\eta))-\Pr(Z\le\Phi^{-1}(y),W(\eta)>\Phi^{-1}(y))}{\psi^2(y)}dy\\
&:= I_1-2I_2,
\end{aligned}$$

where $W(\eta)=Z\cos\eta+Z'\sin\eta$ and Z, Z' are i.i.d. $N(0,1)$. Therefore $(Z,W(\eta))$ follows $BVN(0,0,1,1,\cos\eta)$. Note that

$$\begin{aligned}
&E[B(y,\theta)(B(y,\theta)-B(y,\theta+\eta))]\\
&=y(1-y)-\Pr(Z\le\Phi^{-1}(y),W(\eta)\le\Phi^{-1}(y))+y^2\\
&=\Pr(Z\le\Phi^{-1}(y),W(\eta)>\Phi^{-1}(y))
\end{aligned}\tag{37}$$

by (7). Let $B(y_j,\theta):=B_j$, $j=1,2$, $B(y_j,\theta+\eta):=B_j(\eta)$, $j=1,2$, $LU(y_1,y_2;\eta)$ $:=\Pr(Z\le\Phi^{-1}(y_1),\ W(\eta)>\Phi^{-1}(y_2))$. Then we also have

$$E[(B_1(\eta)-B_1)(B_2(\eta)-B_2)]=2LU(y_1,y_2;\eta). \tag{38}$$

Using (38) and the fact $E(Y_1^2Y_2^2)=\sigma_1^2\sigma_2^2+2(\rho\sigma_1\sigma_2)^2$ when (Y_1,Y_2) follows $BVN(0,0,\sigma_1^2,\sigma_2^2,\rho)$, we get

$$\begin{aligned}
&E(I_1^2)\\
&=2\int_0^\epsilon\int_{y_1}^\epsilon \frac{E[(B_1(\eta)-B_1)^2(B_2(\eta)-B_2)^2]-4LU(y_1,y_1;\eta)LU(y_2,y_2;\eta)}{\psi^2(y_1)\psi^2(y_2)}dy_2dy_1\\
&=16\int_0^\epsilon\int_{y_1}^\epsilon \frac{(LU(y_1,y_2;\eta))^2}{\psi^2(y_1)\psi^2(y_2)}dy_2dy_1.
\end{aligned}\tag{39}$$

Note that the following result (sometimes known as Wick's theorem) holds

$$E(X_1X_2X_3X_4)=E(X_1X_2)E(X_3X_4)+E(X_1X_3)E(X_2X_4)+E(X_1X_4)E(X_2X_3) \tag{40}$$

if (X_1, X_2, X_3, X_4) follows the multivariate normal with mean $\mathbf{0}$. By (37) and (40), we have

$$\begin{aligned}
&E(I_2^2)\\
&= 2\int_0^\epsilon \int_{y_1}^\epsilon \{E[B_1B_2(B_1 - B_1(\eta))(B_2 - B_2(\eta))]\\
&\qquad -E[B_1(B_1 - B_1(\eta))]E[B_2(B_2 - B_2(\eta))]\} /\psi^2(y_1)\psi^2(y_2)dy_2dy_1\\
&= 2\int_0^\epsilon \int_{y_1}^\epsilon \{E[B_2(B_1 - B_1(\eta))]E[B_1(B_2 - B_2(\eta))]\\
&\qquad +E(B_1B_2)E[(B_1 - B_1(\eta))(B_2 - B_2(\eta))]\} /\psi^2(y_1)\psi^2(y_2)dy_2dy_1\\
&= 2\int_0^\epsilon \int_{y_1}^\epsilon \frac{(LU(y_1, y_2; \eta))^2 + E(B_1B_2)E[(B_1-B_1(\eta))(B_2-B_2(\eta))]}{\psi^2(y_1)\psi^2(y_2)} dy_2dy_1. \quad (41)
\end{aligned}$$

To get (36), it is enough to show

$$L_1 := \int_0^\epsilon \int_{y_1}^\epsilon \frac{(LU(y_1, y_2; \eta))^2}{\psi^2(y_1)\psi^2(y_2)} dy_2dy_1 \le C\eta^2, \quad (42)$$

$$L_2 := \int_0^\epsilon \int_{y_1}^\epsilon \frac{E(B_1B_2)E[B_1 - B_1(\eta))(B_2 - B_2(\eta))]}{\psi^2(y_1)\psi^2(y_2)} dy_2dy_1 \le C\eta^2|\log \eta|, \quad (43)$$

by (39) and (41).

Let us show (43). By (38) and $\frac{\cos\eta}{\sin\eta} = \frac{(1-\xi^2)^{1/2}}{\xi} \le \frac{1}{\xi} - \frac{\xi}{2}$ with $\xi = \sin\eta$,

$$\begin{aligned}
L_2 &= 2\int_0^\epsilon \int_{y_1}^\epsilon \frac{y_1(1 - y_2)LU(y_1, y_2; \eta)}{\psi^2(y_1)\psi^2(y_2)} dy_2dy_1\\
&\le 2\int_{-\infty}^{-\delta} \int_{x_1}^{-\delta} \frac{\Phi(x_1)}{\phi(x_1)\phi(x_2)} \int_{-\infty}^{x_1} \phi(z)\bar\Phi\left(\frac{x_2 - z}{\xi} + \frac{\xi z}{2}\right) dzdx_2dx_1\\
&:= 2\Delta(\xi)
\end{aligned}$$

with $\Phi^{-1}(\epsilon) = -\delta$ and $\bar\Phi(x) = 1 - \Phi(x)$. To prove (43), it is enough to show

$$|\Delta'(\xi)| \le C\xi|\log \xi|. \quad (44)$$

By differentiating with respect to ξ, we get

$$\Delta'(\xi) = \int_{-\infty}^{-\delta} \int_{x_1}^{-\delta} \frac{\Phi(x_1)}{\phi(x_1)\phi(x_2)} \int_{-\infty}^{x_1} \phi(z)\left(\frac{x_2-z}{\xi^2} - \frac{z}{2}\right)\phi\left(\frac{x_2-z}{\xi} + \frac{\xi z}{2}\right) dzdx_2dx_1.$$

By change of variables to $w = (x_2 - z)/\xi$ and the fact $x_2w \le 0$, we can easily show

$$\phi(z)\phi\left(\frac{x_2 - z}{\xi} + \frac{\xi z}{2}\right) \le \phi(w)\phi(x_2)e^{-\xi^2x_2^2/8}.$$

Therefore the inner integral in $\Delta'(\xi)$ becomes

$$\begin{aligned}&\int_{-\infty}^{x_1} \phi(z)\left(\frac{x_2-z}{\xi^2}-\frac{z}{2}\right)\phi\left(\frac{x_2-z}{\xi}+\frac{\xi z}{2}\right)dz\\ &\le e^{-\xi^2 x_2^2/8}\phi(x_2)\int_{(x_2-x_1)/\xi}^{\infty}\left(1+\frac{\xi^2}{2}\right)w\phi(w)-\frac{\xi x_2}{2}\phi(w)dw\\ &= e^{-\xi^2 x_2^2/8}\phi(x_2)\left[\left(1+\frac{\xi^2}{2}\right)\phi\left(\frac{x_2-x_1}{\xi}\right)-\frac{\xi x_2}{2}\bar{\Phi}\left(\frac{x_2-x_1}{\xi}\right)\right].\end{aligned}$$

Using $\frac{\Phi(x)}{\phi(x)} < \frac{1}{|x|}$, $x \le -\delta$ we get

$$\begin{aligned}|\Delta'(\xi)| \le\ & C\int_{-\infty}^{-\delta}\int_{x_1}^{-\delta}\frac{1}{|x_1|}e^{-\xi^2 x_2^2/8}\phi\left(\frac{x_2-x_1}{\xi}\right)dx_2dx_1\\ &-\frac{\xi}{2}\int_{-\infty}^{-\delta}\int_{x_1}^{-\delta}\frac{1}{|x_1|}e^{-\xi^2 x_2^2/8}x_2\bar{\Phi}\left(\frac{x_2-x_1}{\xi}\right)dx_2dx_1 := J_1 - J_2/2.\end{aligned}$$

By change of variables to $(x_2-x_1)/\xi = v$, J_1 becomes

$$J_1 = C\xi\int_{-\infty}^{-\delta}\int_0^{(-\delta-x_1)/\xi}\frac{1}{|x_1|}\phi(v)\exp[-\xi^2(\xi^2v^2+2x_1\xi v+x_1^2)/8]dvdx_1.$$

Since the inner integral in J_1 is

$$\int_0^{(-\delta-x_1)/\xi}\phi(v)\exp[-\xi^4v^2/8 - x_1\xi^3 v/4]dv \le \int_{-\infty}^{\infty}\phi(v)e^{-x_1\xi^3v/4}dv = e^{\xi^6x_1^2/32},$$

we have $J_1 \le C\xi\int_{-\infty}^{-\delta}\frac{1}{|x_1|}e^{-(\xi^2/8)(1-\xi^4/4)x_1^2}dx_1$. By change of variables to $\xi x_1 = t$, we get

$$\begin{aligned}J_1 &\le C\xi\int_{-\infty}^{-\xi\delta}\frac{1}{|t|}e^{-(1-\xi^4/4)(t^2/8)}dt\\ &\le C\xi + C\xi\int_{-1}^{-\xi\delta}\frac{1}{|t|}dt = O(\xi|\log\xi|).\end{aligned} \tag{45}$$

In the same way, we can show

$$J_2 = O(\xi|\log\xi|). \tag{46}$$

Then (44) follows from (45) and (46), and (43) follows from (44). We can show (42) in the exactly same way and the claim (36) follows. This proves the existence of the integral.

Proof of Theorem 2. The affine invariance of P_n in (2) allows us to assume that the law of $\mathbf{X}$ is $BVN(0,0,1,1,0)$. Let $Q_n(y,\theta)$ be the sample quantile function of $X_{1i}\cos\theta + X_{2i}\sin\theta$ defined as in (23) and $\hat{\sigma}^2(\theta) := \hat{\sigma}_1^2\cos^2\theta + \hat{\sigma}_2^2\sin^2\theta + 2\hat{\rho}\hat{\sigma}_1\hat{\sigma}_2\cos\theta\sin\theta = sd^2(X_1\cos\theta + X_2\sin\theta)$. Putting $\widetilde{Q}_n(y,\theta) := (Q_n(y,\theta) - (\bar{X}_1\cos\theta + \bar{X}_2\sin\theta))/\hat{\sigma}(\theta)$, let us show the result for the truncated statistic

$$\widetilde{P}_n^T = \sup_{\theta\in[0,2\pi)} \int_{n^{-\delta}}^{1-n^{-\delta}} n(\widetilde{Q}_n(y,\theta) - \Phi^{-1}(y))^2 dy \tag{47}$$

with $0 < \delta < 1/8$. Let $A_n(y,\theta) := Q_n(y,\theta) - \Phi^{-1}(y) + \Phi^{-1}(y)(1 - \hat{\sigma}(\theta)) - (\bar{X}_1\cos\theta + \bar{X}_2\sin\theta)$ and

$$M_n^T := \sup_{\theta\in[0,2\pi)} \int_{n^{-\delta}}^{1-n^{-\delta}} nA_n^2(y,\theta)dy.$$

Then we can easily show the equality $\widetilde{Q}_n(y,\theta) - \Phi^{-1}(y) = A_n(y,\theta) + A_n(y,\theta)$ $(1 - \hat{\sigma}(\theta))/\hat{\sigma}(\theta)$. Since

$$(\widetilde{Q}_n(y,\theta) - \Phi^{-1}(y))^2 \le A_n^2(y,\theta) + A_n^2(y,\theta)\left(\frac{1-\hat{\sigma}(\theta)}{\hat{\sigma}(\theta)}\right)^2 + 2A_n^2(y,\theta)\left|\frac{1-\hat{\sigma}(\theta)}{\hat{\sigma}(\theta)}\right|,$$

$\widetilde{P}_n^T$ in (47) becomes

$$\begin{aligned}
&|\widetilde{P}_n^T - M_n^T| \\
&\le \sup_{\theta\in[0,2\pi)} \int_{n^{-\delta}}^{1-n^{-\delta}} nA_n^2(y,\theta)dy\Big(\frac{(1-\hat{\sigma}(\theta))^2}{\hat{\sigma}^2(\theta)} + 2\Big|\frac{1-\hat{\sigma}(\theta)}{\hat{\sigma}(\theta)}\Big|\Big) \\
&\le C\Big(\sup_{\theta\in[0,2\pi)} \int_{n^{-\delta}}^{1-n^{-\delta}} n(Q_n(y,\theta) - \Phi^{-1}(y))^2 dy\Big(\sup_{\theta\in[0,2\pi)} \frac{(1-\hat{\sigma}(\theta))^2}{\hat{\sigma}^2(\theta)} \\
&\quad + \sup_{\theta\in[0,2\pi)} 2\Big|\frac{1-\hat{\sigma}(\theta)}{\hat{\sigma}(\theta)}\Big|\Big) + \int_0^1 (\Phi^{-1}(y))^2 dy\Big(\sup_{\theta\in[0,2\pi)} \frac{n(1-\hat{\sigma}(\theta))^4}{\hat{\sigma}^2(\theta)} + \sup_{\theta\in[0,2\pi)} \frac{2n|1-\hat{\sigma}(\theta)|^3}{|\hat{\sigma}(\theta)|}\Big) \\
&\quad + \sup_{\theta\in[0,2\pi)} n(\bar{X}_1\cos\theta + \bar{X}_2\sin\theta)^2\Big(\sup_{\theta\in[0,2\pi)} \frac{(1-\hat{\sigma}(\theta))^2}{\hat{\sigma}^2(\theta)} + \sup_{\theta\in[0,2\pi)} 2\Big|\frac{1-\hat{\sigma}(\theta)}{\hat{\sigma}(\theta)}\Big|\Big)\Big)
\end{aligned} \tag{48}$$

by the inequality $(x + y + z)^2 \le C(x^2 + y^2 + z^2)$. Since the terms on the right hand side of (48) are all $o_p(1)$, we have

$$\Big|\widetilde{P}_n^T - M_n^T\Big| \xrightarrow{p} 0. \tag{49}$$

Note that

$$\sqrt{n}(\bar{X}_1\cos\theta + \bar{X}_2\sin\theta) = \int_0^1 \sqrt{n}(Q_n(y,\theta) - \Phi^{-1}(y))dy. \tag{50}$$

Note also that

$$\begin{aligned}\hat{\sigma}^2(\theta) &= \int_0^1 (Q_n(y,\theta))^2 dy - \Big(\int_0^1 Q_n(y,\theta)dy\Big)^2 \\ &= 1 + 2\int_0^1 (Q_n(y,\theta) - \Phi^{-1}(y))\Phi^{-1}(y)dy \\ &\quad + \int_0^1 (Q_n(y,\theta) - \Phi^{-1}(y))^2 dy - \Big(\int_0^1 Q_n(y,\theta) - \Phi^{-1}(y)dy\Big)^2,\end{aligned}$$

$$\begin{aligned}\sqrt{n}(\hat{\sigma}(\theta) - 1) &\approx \sqrt{n}(\hat{\sigma}^2(\theta) - 1)/2 \\ &\approx \sqrt{n}\int_0^1 (Q_n(y,\theta) - \Phi^{-1}(y))\Phi^{-1}(y)dy. \qquad (51)\end{aligned}$$

By expanding the square in M_n^T and using (50) and (51), we can easily show

$$\begin{aligned}&\Big|M_n^T - \sup_{\theta\in[0,2\pi)}\Big\{\int_{n^{-\delta}}^{1-n^{-\delta}} n(Q_n(y,\theta) - \Phi^{-1}(y))^2 dy \\ &\quad -\Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \sqrt{n}(Q_n(y,\theta) - \Phi^{-1}(y))dy\Big)^2 \\ &\quad -\Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \sqrt{n}(Q_n(y,\theta) - \Phi^{-1}(y))\Phi^{-1}(y)dy\Big)^2\Big\}\Big| \\ =\ &\Big|M_n^T - \sup_{\theta\in[0,2\pi)}\Big\{\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n^2(y,\theta)}{\phi^2(\Phi^{-1}(y))}dy \\ &-\Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n(y,\theta)}{\phi(\Phi^{-1}(y))}dy\Big)^2 - \Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n(y,\theta)}{\phi(\Phi^{-1}(y))}\Phi^{-1}(y)dy\Big)^2\Big\}\Big| \\ \xrightarrow{p}\ & 0,\end{aligned}$$

where $\rho_n(y,\theta)$ is defined in (24). Therefore our claim reduces to showing that on the same probability space one can define a Brownian bridge such that

$$\begin{aligned}&\Big|\sup_{\theta\in[0,2\pi)}\Big\{\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n^2(y,\theta)}{\phi^2(\Phi^{-1}(y))}dy \\ &-\Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n(y,\theta)}{\phi(\Phi^{-1}(y))}dy\Big)^2 - \Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{\rho_n(y,\theta)}{\phi(\Phi^{-1}(y))}\Phi^{-1}(y)dy\Big)^2\Big\} \\ -\ &\sup_{\theta\in[0,2\pi)}\Big\{\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{B_n^2(y,\theta)}{\phi^2(\Phi^{-1}(y))}dy \\ &-\Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{B_n(y,\theta)}{\phi(\Phi^{-1}(y))}dy\Big)^2 - \Big(\int_{n^{-\delta}}^{1-n^{-\delta}} \frac{B_n(y,\theta)}{\phi(\Phi^{-1}(y))}\Phi^{-1}(y)dy\Big)^2\Big\}\Big| \\ :=\ &|S\rho_n - SB_n| \xrightarrow{p} 0.\end{aligned}$$

This can be shown easily by Lemma 4 and Lemma 9. By (49), we have $|\widetilde{P}_n^T - SB_n| \xrightarrow{p} 0$. Since $P_n^T = \widetilde{P}_n^T + o_p(1)$, we still have $|P_n^T - SB_n| \xrightarrow{p} 0$.

To complete the proof we should show the existence of the right hand side of (12). The existence of

$$\sup_{\theta\in[0,2\pi)} \int_0^1 \frac{B^2(y,\theta)-y(1-y)}{\phi^2(\Phi^{-1}(y))}dy$$

is given in the proof of Theorem 1. Note that $\sup_{\theta\in[0,2\pi)} \sqrt{n}(\bar{X}_1\cos\theta+\bar{X}_2\sin\theta) \leq \sqrt{n}\sqrt{\bar{X}_1^2+\bar{X}_2^2} = O_p(1)$. Therefore $\sup_{\theta\in[0,2\pi)} \left(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))}dy\right)^2 < \infty$. Also note that $\sup_{\theta\in[0,2\pi)} n(\hat{\sigma}^2(\theta)) \leq \sup_{\theta\in[0,2\pi)} n(\hat{\sigma}_1^2+\hat{\sigma}_2^2) \leq \sup_{\theta\in[0,2\pi)} 2n\max(\hat{\sigma}_1^2,\hat{\sigma}_2^2) = O_p(1)$ and therefore $\sup_{\theta\in[0,2\pi)} \left(\int_0^1 \frac{B(y,\theta)}{\phi(\Phi^{-1}(y))}\Phi^{-1}(y)dy\right)^2 < \infty$.

To prove Conjecture 2 we need to show the tails of $P_n - a_n^o$ are negligible for large n.

Acknowledgements

The authors are grateful to an associate editor and two referees for their careful review and helpful comments which improved the presentation of this paper.

References

Adler, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes.* Lecture Notes **12**, Institute of Mathematical Statistics.

Baringhaus, L. and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* **35**, 339-348.

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

Cox, D. R. and Small, N. J. H. (1978). Testing multivariate normality. *Biometrika* **65**, 263-272.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* CBMS-NSF Regional Conference Series in Applied Mathematics.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.* **14**, 708-732.

Csörgő, S. (1989). Consistency of some tests for multivariate normality. *Metrika* **36**, 107-116.

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit Techniques.* Marcel Dekker, New York.

de Wet, T. and Venter, J. H. (1972). Asymptotic distributions of certain test criteria of normality. *S. Afr. Statist. J.* **6**, 135-149.

de Wet, T. and Venter, J. H. (1973). Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.* **1**, 380-387.

del Barrio, E., Cuesta, J. A., Matrán, C. and Rodríguez, J. M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.* **27**, 1230-1239.

Epps, T. W. and Pulley, I. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70**, 723-726.

Fattorini, L. (1986). Remarks on the use of the Shapiro-Wilk statistic for testing multivariate normality. *Statistica* **46**, 209-217.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

Henze, N. and Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *J. Multivariate Anal.* **62**, 1-23.

Henze, N. and Zirkler, H. (1990). A class of invariant and consistent tests for multivariate normality. *Comm. Statist. Theory Methods* **19**, 3595-3617.

Horsewell, R. L. and Looney, S. W. (1992). A comparison of tests for multivariate normality that are based on measure of multivariate skewness and kurtosis. *J. Statist. Comput. Simulation* **42**, 21-38.

Johnson, N. L. and Kotz, S. (1972). *Continuous Multivariate Distributions.* John Wiley, New York.

Kim, N. (1994). Goodness of fit tests for bivariate distributions. Ph.D. dissertation, University of California, Berkeley.

Leslie, J. R., Stephens, M. A. and Fotopolous, S. (1986). Asymptotic distribution of the Shapiro-Wilk W for testing for normality. *Ann. Statist.* **14**, 1497-1506.

Looney, S. W. and Gulledge, T. R. Jr. (1985). Use of the correlation coefficient with normal probability plots. *Amer. Statist.* **39**, 75-79.

Machado, S. G. (1983). Two statistics for testing multivariate normality. *Biometrika* **70**, 713-718.

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68**, 176-179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhyā A* **36**, 115-128.

Mardia, K. V. (1975). Assessment of multivariate normality and the robustness of Hotelling's T^2 test. *Appl. Statist.* **24**, 163-171.

Mardia, K. V. (1980). Tests of univariate and multivariate normality. In *Handbook in Statistics* (Edited by P. R. Krishnaiah), 279-320. Amsterdam, North-Holland.

Mardia, K. V. and Foster, K. (1983). Omnibus test of multinormality based on skewness and kurtosis. *Comm. Statist. Theory Methods* **12**, 207-221.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab.* **17**, 266-291.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220-238.

Shapiro, S. S. and Francia, R. S. (1972). An approximate analysis of variance test for normality. *J. Amer. Statist. Assoc.* **67**, 215-216.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591-611.

Zhu, L., Wong, H. L. and Fang, K. (1995). A test for multivariate normality based on sample entropy and projection pursuit. *J. Statist. Plann. Inference* **45**, 373-385.

Department of Applied Mathematics, Hongik University, Seoul, Korea.

E-mail: nhkim@wow.hongik.ac.kr

Department of Statistics, University of California, Berkeley, U.S.A.

E-mail: bickel@stat.berkeley.edu

(Received May 2000; accepted September 2002)