

Univariate Likelihood Projections and Characterizations of the Multivariate Normal

Distribution

Albert Vexler

Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, U.S.A

avexler@buffalo.edu

Abstract

The problem of characterizing a multivariate distribution of a random vector using examination of univariate combinations of vector components is an essential issue of multivariate analysis. The likelihood principle plays a prominent role in developing powerful statistical inference tools. In this context, we raise the question: can the univariate likelihood function based on a random vector be used to provide the uniqueness in reconstructing the vector distribution? In multivariate normal (MN) frameworks, this question links to a reverse of Cochran's theorem that concerns the distribution of quadratic forms in normal variables. We characterize the MN distribution through the univariate likelihood type projections. The proposed principle is employed to illustrate simple techniques for assessing multivariate normality via well-known tests that use univariate observations. The displayed testing strategy can exhibit high and stable power characteristics in comparison to the well-known procedures in various scenarios when observed vectors are non-MN distributed, whereas their components are normally distributed random variables. In such cases, classical multivariate normality tests, such as Shapiro-Wilk's, Henze-Zirklers's and Mardia's tests, may break down completely.

Keywords: Characterization, Infinity divisible, Likelihood, Multivariate normal distribution, Projection, Quadratic form, Test for multivariate normality.

1. Introduction

In various theoretical and applied studies, multivariate analysis treats multivariate normally distributed data (e.g., Kotz et al. [16]). There is an extensive amount of fundamental results

related to characterizations of the multivariate normal distribution. In this context, characterizations of multivariate normality (MN) through univariate projections play fundamental roles, providing relatively simple procedures to assess the assumption of MN regarding a random vector distribution (e.g., Shao and Zhou [30]; Cuesta-Albertos et al. [8]; Looney [20]). Univariate characterizations of MN play an important part in different applications to multivariate data based comparisons, simulations and stochastic modeling (e.g., Nagaraja [22]). Perhaps, mostly addressed univariate characterization of MN employs that the random variables X_1, \dots, X_p are jointly normal if and only if every linear combination of them is a univariate normal. This property underlies many strategies of multivariate comparisons, stochastic modeling and testing for MN that have structures of powerful techniques developed in the univariate cases (e.g., Looney [20]; Zhu et al. [37]).

An important critical result is that the MN of all subsets ($r < p$) of the normal variables X_1, \dots, X_p together with the normality of an infinity number of linear combinations of them do not insure the joint normality of these variables, when $p > 2$ (e.g., Hamedani [10]). This raises a vital concern regarding the common statistical procedures, e.g., for assessing MN of a random vector by examining a limited number of linear combinations of its components (e.g., Shao and Zhou [30]). In practice, technical reasons restrict the number of the linear combinations to be considered.

The main aim of this paper is to introduce an alternative univariate projection of MN that is inspired by the following statements. The likelihood principle plays a prominent role in developing powerful statistical inference tools (e.g., Vexler and Hutson [35]). Oftentimes, likelihood functions assist to derive sufficient information regarding observed data. Then, one might ask: can a distribution of the likelihood function based on the vector $X = (X_1, \dots, X_p)^T$ be involved in complete reconstruction of X 's distribution? (In this context, the term

“likelihood function” is defined as the joint density function of X , say f , computed in the form $f(X_1, \dots, X_p)$. The likelihood function based on X is a univariate random variable.

In the case where X is MN distributed, the corresponding log likelihood function can be directly associated with so called quadratic forms (see Section 2 for details). According to Ruben [28], “from a substantive or statistical point of view the characterization of normality via quadratic forms must rank as of greater interest when one bears in mind that the core of statistical science, namely the entire vast area of regression analysis, including analysis of variance, is based on quadratic forms of the components of the observation vector.” Ruben [28] provided characterizations of normality, showing that, when X_1, \dots, X_p are symmetric, independently and identically distributed random variables with zero means and finite variances, the corresponding quadratic form has a chi-squared distribution if and only if X_1 is normal. This approach can characterize X_1, \dots, X_p as normally distributed random variables, but does not sufficiently imply that X is *MN*-distributed (Hamedani [10]). Indeed, it is of theoretical and applied interest to release the conditions regarding independence of X 's components and their symmetry.

In Section 2, we establish a new characterization of MN for a random vector by examining the relevant quadratic form. The obtained results can underlie a reverse of Cochran's theorem (e.g., Styan [32]) that concerns the distribution of quadratic forms in normal variables. It turns out that, in general cases, we can provide one-to-one mapping between the likelihood's and X 's distributions, using properties of infinity divisible (ID) distribution functions. For an extensive review and examples related to univariate and multivariate ID distributions, we refer the reader to Bose et al. [4]. We point out that the problem of univariate likelihood projections can be linked to the issue of reconstructing summands distributions by a distribution of their sum. In this context, the conclusions of Prokhorov and Ushakov [24] (see Theorem 1 and its

Corollary in [23]) show that, even in the simple case of independent X_1, \dots, X_p , the ID assumption applied in Section 2 cannot be significantly improved.

In Section 3, we exemplify an application of the proposed method, constructing simple tests for MN. Although many techniques for assessing MN have been proposed (e.g., Kim and Park [15]), there is still a paucity of genuine statistical tests for MN (e.g., Kotz et al. [16]). Taking into account the arguments presented by Looney [20], we demonstrate techniques for assessing MN based on well-known tests that use univariate observations. We experimentally show that the presented likelihood projection based testing strategy can exhibit high and stable power characteristics in comparison to the relevant well-known classical procedures in various scenarios when X is not MN-distributed, whereas X_1, \dots, X_p are dependent or independent normally distributed random variables (Stoyanov [31]). In such cases, the Shapiro-Wilk, Henze-Zirklers and the Mardia multivariate normality tests may break down completely. In Section 4 the proposed method is applied to a biomarker study associated with myocardial infarction disease. We conclude with remarks in Section 5.

2. Likelihood projections

We first introduce the basic notation regarding the statement of the problem. Then the main results are provided in Theorems 1 and 2 that establish univariate likelihood based characterizations of MN. The proofs of Theorems 1 and 2 are included for completeness and contain comments that assist to describe the obtained results. Important notes related to conditions used in the proposed technique are presented in Remarks.

Let $X = (X_1, \dots, X_p)^T$ be the p -dimensional random vector with mean vector $\mu = (\mu_1, \dots, \mu_p)^T$ and covariance matrix Σ . The covariance matrix Σ is positive-definite. Then we can use an orthogonal (orthonormal) matrix Q to present the diagonalizable form of Σ ,

$Q^T \Sigma Q = \Lambda$, where the matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$, $\lambda_i > 0, i = 1, \dots, p$, (e.g., Baldessari [1]).

Define the following matrices

$$\Delta = \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_p^{-1/2}), H = Q^T \Delta Q, \text{ and } z = H(X - \mu).$$

Obviously H is symmetric (e.g., $H = H^T$) and $H^T H = Q^T \Delta Q Q^T \Delta Q = Q^T \Lambda^{-1} Q = Q^T \Sigma^{-1} Q$,

since the equation $Q^T \Sigma Q = \Lambda$ provides $\Sigma = Q \Lambda Q^T$ after applying the inverse of both sides and

using that the inverse of an orthogonal matrix is equal to its transpose. Also we have

$$H \Sigma H^T = Q \Delta Q^T \Sigma Q \Delta Q^T = Q \Delta Q^T Q \Lambda Q^T Q \Delta Q^T = I_p,$$

where I_p is the identity matrix of size p .

Assuming that X is observed and follows a multivariate normal distribution, say

$X \sim N_p(\mu, \Sigma)$, we can write the conventional likelihood function

$L = (2\pi)^{-p/2} |A|^{1/2} \exp\{-0.5W\}$, where $A = \Sigma^{-1}$ is a real symmetric positively defined matrix

and the quadratic form $W = (X - \mu)^T A (X - \mu)$. It is clear that the distribution of W

determines the distribution of L and vice versa. Note that

$$W = (X - \mu)^T \Sigma^{-1} (X - \mu) = (X - \mu)^T H^T H (X - \mu) = [H(X - \mu)]^T H(X - \mu) = z^T z,$$

where $z = H(X - \mu)$.

Theorem 1 (Likelihood Projection and Characterization). The following two statements are equivalent:

(a) X is an infinitely divisible random (ID) vector (e.g., Bose et al. [4]), the random vector

z consists of p independent components and the random variable $z^T z = W$ has the chi-

square distribution with p degrees of freedom, say $z^T z \sim \chi_p^2$.

(b) $X \sim N_p(\mu, \Sigma)$.

Proof. Under Statement (a), we have $z = H(X - \mu) = (z_1, \dots, z_p)^T$, where z_i is a linear combination of X_1, \dots, X_p , $i = 1, \dots, p$. Therefore, for all $i \in \{1, \dots, p\}$, z_i is an ID random variable (e.g., Horn and Steutel [12]: Theorem 3.2; Rao [26: p. 66]). (Note that, in this case, the assumption: “ X is an ID random vector” is employed, whereas, in general, a linear combination of ID random variables can be not an ID random variable. Here, for example, in a particular case, we can regard a structure of the definition of normally distributed random vectors, comparing to that of normally distributed random variables, and refer to, e.g., Hamedani [10].) Then we apply the following result, focusing on $z^T z = z_1^2 + \dots + z_p^2$, where z_1, \dots, z_p are independent.

Proposition 1 (Kruglov [17]). If Y_1, \dots, Y_p are independent ID random variables such that $Y_1^2 + \dots + Y_p^2$ has the chi-square distribution with p degrees of freedom then random variables Y_1, \dots, Y_p have the same standard normal distribution.

Thus, for all $i \in \{1, \dots, p\}$, $z_i \sim N_1(0, 1)$. Since $(X - \mu) = H^{-1}z$, for all $i \in \{1, \dots, p\}$, $X_i - \mu_i$ is a linear combination of independent identically $N_1(0, 1)$ -distributed z_1, \dots, z_p . Then, the simple use of a characteristic function of $X_i - \mu_i$ shows that $X_i - \mu_i \sim N_1$, $i \in \{1, \dots, p\}$.

(Note that, in this case, we use that z_1, \dots, z_p are independent and identically distributed, whereas, in general, a linear combination of normally distributed random variables can be non-normally distributed.) Now, Propositions 1 and 2 of Wesolowski [36] assist to conclude that the ID random vector $X \sim N_p(\mu, \Sigma)$ that is Statement (b).

Under Statement (b), it is clear that X is an ID random vector and we have the quadratic form $z^T z = W = (X - \mu)^T A(X - \mu) \sim \chi_p^2$ by virtue of Cochran’s theorem (e.g., Styan [32]). In this case, $E(z) = H[E(X - \mu)] = 0$ and $\text{var}(z) = H \text{var}(X - \mu) H^T = H \Sigma H^T = I_p$, and then

$z \sim N_p(0, I)$, since $X \sim N_p(\mu, \Sigma)$. These provide Statement (a) and then we complete the proof.

Remark 1. It seems that the ID requirement used in Theorem 1 can be substituted by a symmetric type restriction on z 's distributions (see the Introduction of Kruglov [17] as well as Ruben [28]). This approach leads to characterize X_1, \dots, X_p as normally distributed random variables, but cannot sufficiently assist to conclude that $X \sim N_p(\mu, \Sigma)$ (Hamedani [10]). This is one of reasons to require that X is an ID vector. In this case the ID restriction on z 's distributions is more profound than the symmetric distributions' considerations (Kruglov [17: p. 873]).

Remark 2. A set of results regarding situations when ID vectors are normally distributed can be found in, e.g., Wesolowski [36] and Bose et al. [4: p. 783]. Bose et al. [4] provided an extensive review and examples related to ID distributions.

The following proposition can get involved into the Theorem 1 structure instead of Proposition 1.

Proposition 2 (Golikova and Kruglov [7]). Let $Y_1, \dots, Y_p, p \geq 2$ be independent ID random variables. The random variable $\sum_{i=1}^2 \left(Y_i - \sum_{j=1}^2 Y_j / 2 \right)^2$ has the chi-square distribution with 1 degree of freedom if and only if (iff) Y_1 and Y_2 are Gaussian random variables with $EY_1 = EY_2$ and $E(Y_1 - EY_1)^2 + E(Y_2 - EY_2)^2 = 2$. In general for $p \geq 3$, if $EY_1 = \dots = EY_p$ and the random variable $\sum_{i=1}^p \left(Y_i - \sum_{j=1}^p Y_j / p \right)^2$ has the chi-square distribution with $p-1$ degrees of freedom then Y_1, \dots, Y_p are Gaussian random variables with $E(Y_1 - EY_1)^2 = \dots = E(Y_p - EY_p)^2 = 1$.

Since $E(z) = 0$ and $\text{var}(z) = I_p$, as an immediate modification of Theorem 1 we have:

Theorem 2. The following two statements are equivalent:

- (a) X is an ID random vector, the vector z consists of independent components and the random variable $\sum_{i=1}^p \left(z_i - \sum_{j=1}^p z_j / p \right)^2 = W - \left(\sum_{j=1}^p z_j \right)^2 / p$ has the chi-square distribution with $p-1$ degrees of freedom.
- (b) $X \sim N_p(\mu, \Sigma)$, $p \geq 2$.

Remark 3. Theorems 1 and 2 treat independent random variables z_1, \dots, z_p . In this context, assuming that $z = (z_1, \dots, z_p)^T$ is an ID random vector and z_1, \dots, z_p are from specific ID distributions with finite fourth moments, we have that z_1, \dots, z_p are independent if and only if $E(z_i^2 z_j^2) = E(z_i^2)E(z_j^2)$, $i \neq j$ for all $1 \leq i, j \leq p$ (see Pierre [23], for details). That is to say, a natural question is when are components of an ID vector independent? In this context, Pierre [23] and Veeh [34] discussed necessary and sufficient conditions in a parallel with those available in the normal case. It turns out that if the ID vector has finite fourth moment, then pairwise independence is equivalent to total independence.

Remark 4. It is clear that the problem considered in Theorems 1 and 2 can be associated with the issue of reconstructing a summands distribution by a distribution of their sum. Even in the simple case of $X \sim N_p(\mu, I_p)$, it turns out that by virtue of the results of Prokhorov and Ushakov [24] (see Theorem 1 and its Corollary in [23]), the ID restriction on z 's distributions cannot be significantly improved. In this context, in a general case, the condition “ z_1, \dots, z_p are independent” seems to be essential.

3. Applications of the proposed technique to test for MN

In this section, we exemplify simple applications of the likelihood projection technique to test for MN, employing available software products. The demonstrated test procedures are experimentally evaluated. We first consider testing strategies when the parameters of the null

distribution are known. The analysis is relatively clear, and has the basic ingredients for more general cases studied in Sections 3 and 4 when the MN parameters are estimated.

Generally speaking, the univariate likelihood projections can yield easy ways to construct tests for MN, e.g., combining a test for $z^T z \sim \chi_p^2$ with a decision making rule for that the random vector z consists of p independent components. Designs, when test strategies combine statistics with structures based on related paradigms, can significantly simplify the development of the tests for MN. For example, taking into account the schematic rule “Likelihood($z^T z \sim \chi_p^2$, z_1, \dots, z_p are independent) = Likelihood($z^T z \sim \chi_p^2$ / z_1, \dots, z_p are independent) \times Likelihood(z_1, \dots, z_p are independent)”, one can employ a sum of test statistics that are based on log-likelihood type concepts.

Without loss of generality, we exemplify the proposed approach via testing of bivariate normality. (See Remark 5 and Section 4 below for testing of trivariate normality.) To this end, we transform the quadratic form $z^T z = W = (X - \mu)^T A (X - \mu)$ via $J = G(W)$, where

$G(x) = \int_0^x \exp(-u/2) du / 2$ is the chi-squared distribution function with two degrees of

freedom. Then, we can aim to test for $J \sim Unif[0,1]$, assessing that $W \sim \chi_2^2$. In this statement, the smooth Neyman test for uniformity (e.g., Ledwina [18]; Rayner et al. [27]), a log-likelihood structured decision making mechanism, uses the statistic

$$T_{1n} = \frac{1}{n} \sum_{j=1}^{k_{1n}} \left(\sum_{i=1}^n b_j(J_i) \right)^2,$$

where values J_1, \dots, J_n , independent realizations of J , are assumed to be observed; b_1, \dots, b_{k_1} are normalized Legendre polynomials on $[0,1]$; and k_{1n} is proposed to be chosen via the data-driven procedure, a modified Schwarz's rule, developed by Ledwina [18] and Inglot and Ledwina [13].

In order to obtain values of T_{1n} , we can employ the R-command (R Development Core Team

[25]): *ddst.uniform.test* that is contained in the R-package ‘*ddst*’ (<https://cran.r-project.org/web/packages/ddst/ddst.pdf>). To test for independence between z_1 and z_2 , we apply the data-driven rank strategy proposed by Kallenberg and Ledwina [14]. The log-likelihood type test statistic is

$$T_{2n} = \frac{1}{n} \sum_{j=1}^{k_{2n}} \left\{ \sum_{i=1}^n b_j \left(\frac{R_{1i} - 1/2}{n} \right) b_j \left(\frac{R_{2i} - 1/2}{n} \right) \right\}^2,$$

where we assume that samples (z_{j1}, \dots, z_{jn}) related to random variables $z_j, j \in \{1, 2\}$, are observed; R_{ji} denotes the rank of z_{ji} among (z_{j1}, \dots, z_{jn}) , $j \in \{1, 2\}$; and k_{2n} is chosen in the data-driven manner, a modified Schwarz's rule, shown in Kallenberg and Ledwina [14]. To implement this procedure, we can use the R-command *testforDEP* that is contained in the R-package ‘*testforDEP*’ (Miecznikowski et al. [21]). Thus, the test statistic for bivariate normality has the form $T_n = T_{1n} + T_{2n}$. The test statistic T_n is a sum of the powerful consistent test statistics.

In practice, the parameters of the null distribution of the vector X are unknown. Thus, finally applying a common approach in assessing MN of underlying data distributions based on the residuals (e.g., Baringhaus and Henze [2]), we obtain the following decision making procedure. Let ${}_i X = ({}_i X_1, {}_i X_2)^T$, $i \in \{1, \dots, n\}$, be independent identically distributed bivariate random vectors that are realizations of $X = (X_1, X_2)^T$, with sample mean $\bar{X}_n = \sum_{i=1}^n ({}_i X) / n$ and sample covariance matrix $S_n = \sum_{i=1}^n ({}_i X - \bar{X})({}_i X - \bar{X})^T / n$. Assume ${}_i X \sim N_2(\mu, \Sigma)$ under the null hypothesis. Then, we can compute $S_n^{-1/2}$ that is (almost surely) the unique symmetric positive-definite square root of the inverse of S_n which is positive-definite with probability one (Eaton and Perlman [6]). Define the residuals $\tilde{z}_i = (\tilde{z}_{1i}, \tilde{z}_{2i})^T = S_n^{-1/2} ({}_i X - \bar{X})$ and the statistics $\tilde{J}_i = G(\tilde{z}_i^T \tilde{z}_i)$, $i \in \{1, \dots, n\}$. The null hypothesis is rejected for large values of

$$\tilde{T}_n = \frac{1}{n} \sum_{j=1}^{k_{1n}} \left(\sum_{i=1}^n b_j(\tilde{J}_i) \right)^2 + \frac{1}{n} \sum_{j=1}^{k_{2n}} \left\{ \sum_{i=1}^n b_j \left(\frac{\tilde{R}_{1i} - 1/2}{n} \right) b_j \left(\frac{\tilde{R}_{2i} - 1/2}{n} \right) \right\}^2,$$

where \tilde{R}_{ji} denotes the rank of \tilde{z}_{ji} among $(\tilde{z}_{j1}, \dots, \tilde{z}_{jn})$, $j \in \{1, 2\}$, k_{1n}, k_{2n} are chosen in the data-driven manner based on observations $\tilde{J}_i, \tilde{z}_i, i \in \{1, \dots, n\}$ (see the T_{1n}, T_{2n} -strategies above, respectively). To compute values of the test statistic \tilde{T}_n , one can use the R code:

```
library(ddst); library(testforDEP); zz<-z1^2+z2^2; J<-pchisq(zz,2); T<-ddst.uniform.test(J,
compute.p=FALSE)$statistic+testforDEP(z1,z2,test="TS2",num.MC = 100)@TS
```

Note that Neyman smooth tests for bivariate normality have been developed by Bogdan [3]. The data driven smooth tests proposed by Bogdan [3] have complicated structures that are difficult to generalize beyond the bivariate case. (Critical issues related to Bogdan's tests for bivariate normality are shown in Ducharme and Micheaux [9].) The likelihood projection based approach provides a simple method for testing MN. In the first stage of our development, it is assumed that the null distribution parameters are known. Then the parameters are replaced by their estimates. The proposed framework can be easily extended to higher dimension cases (see Remark 5 and Section 4 below for examples).

3.1. Null distribution

According to Szkutnik [33], the null distribution of the residuals based test statistic \tilde{T}_n does not depend on the parameters (μ, Σ) under the null hypothesis (see also, e.g., Baringhaus and Henze [2]). However Henze [11] provided concerns regarding this fact. We then present the critical values for the present test for different sample sizes using the Monte Carlo technique, and experimentally examine this result for different values of $\rho = \text{corr}(X_1, X_2)$, $i \in \{1, \dots, n\}$.

In order to tabulate the percentiles of the null distribution of the test statistic \tilde{T}_n , we drew

55,000 samples of $X_1, \dots, X_n \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \right)$ calculating values of \tilde{T}_n at each sample

size n . The generated values of the test statistic \tilde{T}_n were used to determine the critical values C_α of the null distribution of \tilde{T}_n at the significance levels α . The results of this Monte Carlo study are displayed in Table 1.

Table 1. Critical values of the proposed test statistic \tilde{T}_n defined in Section 3.

α					α				
n	0.2	0.1	0.05	0.01	n	0.2	0.1	0.05	0.01
25	0.9759	5.5513	9.3079	18.6332	60	0.6805	1.5788	6.7011	15.2855
30	0.8577	5.2503	9.1045	18.3795	80	0.6284	1.1314	5.9050	14.2164
35	0.8017	4.8672	8.3672	16.9611	90	0.6225	1.0960	5.9588	14.8072
45	0.7256	4.4877	7.7548	16.6368	100	0.6192	1.0682	5.8645	13.7717
50	0.7069	4.2035	7.1607	16.2087	125	0.6014	0.9961	5.5347	13.2184

In order to verify the results shown in Table 1, for different values of $\rho \in (-1,1)$ and n , we calculated the Monte Carlo approximations to

$$\Pr \left\{ \tilde{T}_n > C_{0.05} / ({}_i X_1, {}_i X_2)^T \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), i \in \{1, \dots, n\} \right\},$$

where $C_{\alpha=0.05}$'s are shown in Table 1. In this study, we also examined the Shapiro-Wilk test (SW), using the R-procedure “*mvShapiro.Test*”. For each value of ρ and n , the Type I error rates were derived using 75,000 samples of $({}_i X_1, {}_i X_2)^T \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), i \in \{1, \dots, n\}$. Table 2 presents the results of this Monte Carlo evaluation.

Table 2. The Monte Carlo Type I error probabilities of the proposed test, \tilde{T}_n , and the Shapiro-

Wilk test (SW), when $({}_i X_1, {}_i X_2)^T \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), i \in \{1, \dots, n\}$ and the anticipated

significance level is $\alpha = 0.05$.

ρ	$n = 35$		$n = 50$		$n = 100$	
	\tilde{T}_n	SW	\tilde{T}_n	SW	\tilde{T}_n	SW
-0.9	0.0505	0.0511	0.0499	0.0501	0.0499	0.0492
-0.7	0.0501	0.0515	0.0498	0.0513	0.0500	0.0497
-0.5	0.0498	0.0507	0.0500	0.0483	0.0500	0.0488
-0.3	0.0501	0.0501	0.0499	0.0489	0.0501	0.0505
-0.1	0.0506	0.0507	0.0495	0.0496	0.0496	0.0491
0	0.0510	0.0512	0.0501	0.0512	0.0500	0.0498
0.1	0.0495	0.0514	0.0502	0.0505	0.0499	0.0500
0.3	0.0500	0.0494	0.0498	0.0493	0.0500	0.0513
0.5	0.0510	0.0510	0.0499	0.0504	0.0499	0.0489
0.7	0.0495	0.0505	0.0507	0.0503	0.0499	0.0482
0.9	0.0497	0.0506	0.0498	0.0488	0.0500	0.0486

According to Table 2, the validity of the critical values related to the test statistic \tilde{T}_n is experimentally confirmed

3.2. Power

In general, in the considered goodness-of-fit framework, there are no most powerful decision making mechanisms. We examine the proposed approach in several scenarios, where decisions to reject MN can be anticipated to be difficult. Taking into account that “As recommended by many authors ..., a reasonable first step in assessing MVN is to test each variable separately for the univariate normality” (Looney [20]), we consider the designs displayed in Table 3, where X_1 and X_2 are normally distributed, whereas $X = (X_1, X_2)^T$ is not N_2 -distributed.

Table 3. Distributions for $X = (X_1, X_2)^T$ used in the power study

Alternative Designs	Models/Descriptions
---------------------	---------------------

- A₁ $X_1 = \xi_1$, $X_2 = |\xi_2| I(\xi_1 \geq 0) - |\xi_2| I(\xi_1 < 0)$, where $I(\cdot)$ is the indicator function and ξ_1, ξ_2 are independent random variables: $\xi_1 \sim N_1(0,1)$, $\xi_2 \sim N_1(0,1)$ (Stoyanov [31]: p. 88)
- A₂ X is from the two dimensional density function
 $f(x_1, x_2) = \varphi(x_1)\varphi(x_2)\{1 + \varepsilon(2\Phi(x_1) - 1)(2\Phi(x_2) - 1)\}$ with $\Phi(x) = \int_{-\infty}^x \varphi(u)du$ and $\varepsilon = 0.999$ (Stoyanov [31]: p. 89).
- A₃ $f(x_1, x_2) = \exp\{-\rho^{-2}(x_1^2 - 2\rho x_1 x_2 + x_2^2)/2\} / \{\pi(1 - \rho^2)^{1/2}\} I(x_1 x_2 \geq 0)$, $\rho = 0.9$
 (Stoyanov [31]: p. 89).
- A₄ $f(x_1, x_2) = \varphi_1(x_1, x_2)/2 + \varphi_2(x_1, x_2)/2$, where $\varphi_1(x_1, x_2)$ and $\varphi_2(x_1, x_2)$ are standard bivariate normal densities with correlation coefficients $\rho_1 = -0.5$ and $\rho_2 = 0.5$, respectively. In this case, X_1 and X_2 are uncorrelated (Stoyanov [31]: p. 93).
- A₅ $f(x_1, x_2) = [\exp\{-2(x_1^2 + x_1 x_2 + x_2^2)/3\} + \exp\{-2(x_1^2 - x_1 x_2 + x_2^2)/3\}] / (2\pi 3^{1/2})$. In this case, X_1 and X_2 are uncorrelated, but dependent (Stoyanov [31]: p. 93).
- A₆ $f(x_1, x_2) = \exp\{-(1+x_1^2)(1+x_2^2)\} / C$, where $C \approx 0.993795$. In this case, all the conditional distributions of X are normal (Stoyanov [31]: p. 97).
- A₇ $X = (\xi^{1/2}\eta_1 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_3 + (1-\xi)^{1/2}\eta_2)^T$, where $\xi \sim Unif[0,1]$ and η_1, η_2, η_3 are independent $N_1(0,1)$ distributed random variables (Stoyanov [31]: p. 97-98).

Table 4 shows the results of the power evaluations of the present test \tilde{T}_n , the SW test, Henze-Zirklers's MN test (HZ) and the classical Mardia's MN test (M) via the Monte Carlo study based on 55,000 replications of the independent identically distributed bivariate random

vectors ${}_1X, \dots, {}_nX$ for designs A_1 - A_7 at each sample size n . To implement the HZ test (Baringhaus and Henze [2]), we used the R-procedure $mvn(X, mvnTest="hz")$ from the package *MVN*. The R- command $mardia(X, plot=FALSE)$ was employed to conduct the M test. The significance level of the tests was fixed at 5%.

Table 4. *The Monte Carlo power of the tests.*

Tests/ n	Design A ₁				Design A ₂			
	25	50	100	125	25	50	100	125
\tilde{T}_n	0.194	0.639	0.979	0.998	0.057	0.062	0.096	0.120
HZ	0.171	0.446	0.946	0.991	0.049	0.056	0.064	0.071
M	0.065	0.134	0.209	0.236	0.021	0.046	0.069	0.076
SW	0.175	0.304	0.558	0.676	0.056	0.057	0.058	0.061
Tests/ n	Design A ₃				Design A ₄			
	25	50	100	125	25	50	100	125
\tilde{T}_n	0.201	0.501	0.875	0.945	0.075	0.144	0.335	0.434
HZ	0.141	0.376	0.778	0.888	0.065	0.078	0.101	0.114
M	0.054	0.306	0.684	0.796	0.054	0.119	0.201	0.236
SW	0.036	0.049	0.096	0.131	0.044	0.045	0.045	0.047
Tests/ n	Design A ₅				Design A ₆			
	25	50	100	125	25	50	100	125
\tilde{T}_n	0.073	0.145	0.337	0.444	0.153	0.401	0.840	0.921
HZ	0.067	0.077	0.102	0.109	0.092	0.205	0.458	0.571
M	0.053	0.117	0.201	0.237	0.037	0.182	0.452	0.548
SW	0.043	0.041	0.046	0.045	0.048	0.053	0.062	0.066
Tests/ n	Design A ₇							
	25	50	100	125				

\tilde{T}_n	0.0730	0.123	0.243	0.304
HZ	0.073	0.097	0.159	0.193
M	0.051	0.110	0.187	0.218
SW	0.054	0.058	0.066	0.067

This study demonstrates that the likelihood projection based test is significantly superior to the considered classical tests in all scenarios A₁-A₇. Specifically, the presented test clearly outperforms the classical tests in terms of the power properties when detecting MN based on vectors with uncorrelated $N_1(0,1)$ -distributed components. It seems that the SW test is biased under A₃ ($n=25, 50$), A₄, A₅ and A₆ ($n=25$) and inconsistent under design A₅. The M test is biased under A₂ ($n=25, 50$).

Based on the Monte Carlo results, we conclude that the proposed test exhibits high and stable power characteristics in comparison to the well-known classical procedures.

Remark 5. Assume, for example, we observe trivariate independent identically distributed vectors ${}_iX = ({}_iX_1, {}_iX_2, {}_iX_3)^T$, $i \in \{1, \dots, n\}$ that are realizations of $X = (X_1, X_2, X_3)^T$. In a similar manner to the bivariate case considered above, we may define the residuals

$\tilde{z}_i = (\tilde{z}_{1i}, \tilde{z}_{2i}, \tilde{z}_{3i})^T$. By Remark 3, in order to test for $X \sim N_3$, we can construct the test statistic

$$\hat{T}_n = \frac{1}{n} \sum_{j=1}^{k_m} \left(\sum_{i=1}^n b_j(\hat{J}_i) \right)^2 + \sum_{(s,r)=(1,2),(1,3),(2,3)} \sum_{j=1}^{k_{sm}} \left\{ \sum_{i=1}^n b_j \left(\frac{\hat{R}_{si} - 1/2}{n} \right) b_j \left(\frac{\hat{R}_{ri} - 1/2}{n} \right) \right\}^2 / n,$$

where $\hat{J}_i = G(\tilde{z}_i^T \tilde{z}_i)$, \hat{R}_{ji} is the rank of \tilde{z}_{ji} among $(\tilde{z}_{j1}, \dots, \tilde{z}_{jn})$, $j \in \{1, 2, 3\}$, k_{1n}, k_{sm} are chosen in the data-driven manner based on observations $\hat{J}_i, \tilde{z}_i, i = 1, \dots, n$. For example, for $n = 250$, using 55,000 replications of ${}_1X, \dots, {}_nX$, we computed the critical value $C_\alpha = 7.0513$ of the null distribution of \hat{T}_n at the significance level $\alpha = 0.05$. The corresponding Monte Carlo powers of

the \widehat{T}_n -based test and the HZ, M, SW tests were obtained as 0.485 and 0.239, 0.227, 0.069, respectively, when $X = \left(\xi^{1/2}\eta_1 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_3 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_4 + (1-\xi)^{1/2}\eta_5 \right)^T$, where $\xi \sim Unif[0,1]$ and $\eta_j, j \in \{1, \dots, 5\}$ are independent $N_1(0,1)$ -distributed random variables. In this design, X_1, X_2 and X_3 are $N_1(0,1)$ -distributed, X_3 , conditionally on X_1, X_2 , has a normal distribution $N_1(0,1)$, however X cannot have a trivariate normal distribution (Stoyanov [31]: p. 97-98). Note also that in the Monte Carlo experiments based on generations of

${}_i X = ({}_i X_1, {}_i X_2, \dots, {}_i X_p)^T, i \in \{1, \dots, n = 250\}, p \in \{5, 7\}$, we obtained $C_{0.05} = 11.655$ and

$C_{0.05} = 12.910$ related to the test statistic

$$\widehat{T}_n = \frac{1}{n} \sum_{j=1}^{k_{in}} \left(\sum_{i=1}^n b_j(\widehat{J}_i) \right)^2 + \sum_{(s,r) \in Q_p} \sum_{j=1}^{k_{sm}} \left\{ \sum_{i=1}^n b_j \left(\frac{\widehat{R}_{si} - 1/2}{n} \right) b_j \left(\frac{\widehat{R}_{ri} - 1/2}{n} \right) \right\}^2 / n$$

with $Q_{p=5} = \{(1,2), (1,3), (1,4), (1,5), (2,3), (2,4), (2,5), (3,4), (3,5), (4,5)\}$ and

$Q_{p=7} = \{(1,2), \dots, (1,7), (2,3), \dots, (2,7), (3,4), \dots, (3,7), (4,5), (4,6), (4,7), (5,6), (5,7), (6,7)\}$,

respectively. In these cases, under the alternatives

$X = \left(\xi^{1/2}\eta_1 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_3 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_4 + (1-\xi)^{1/2}\eta_5, \xi^{1/2}\eta_6 + (1-\xi)^{1/2}\eta_5, \xi^{1/2}\eta_7 + (1-\xi)^{1/2}\eta_8 \right)^T$ and $X = \left(\xi^{1/2}\eta_1 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_3 + (1-\xi)^{1/2}\eta_2, \xi^{1/2}\eta_4 + (1-\xi)^{1/2}\eta_5, \xi^{1/2}\eta_6 + (1-\xi)^{1/2}\eta_5, \xi^{1/2}\eta_7 + (1-\xi)^{1/2}\eta_8, \xi^{1/2}\eta_9 + (1-\xi)^{1/2}\eta_7, \xi^{1/2}\eta_{10} + (1-\xi)^{1/2}\eta_{11} \right)^T$, where $\xi \sim Unif[0,1]$ and $\eta_j \sim N_1(0,1), j \in \{1, \dots, 11\}$ are independent

random variables, the Monte Carlo powers of the \widehat{T}_n -based test and the HZ, M, SW tests

defined in Section 3.2 were obtained as 0.315, 0.119, 0.158, 0.068, for $p = 5$, and 0.271,

0.085, 0.126, 0.060, for $p = 7$. Then, in these cases, the proposed test demonstrated robust

power characteristics with respect to the values of p .

Remark 6. It was observed that the developed test based on \tilde{T}_n showed good power levels in different scenarios that are relatively simple in comparison to those shown in Table 3. For example, considering the regression type model: $X = (X_1, X_2)^T$, $X_1 = \eta \sim Unif[-10,10]$, $X_2 = X_1 + \xi$, $\xi \sim N_1(0,1)$, the Monte Carlo powers of the \tilde{T}_n -based test and the HZ, M, SW tests were computed as 0.320, 0.317, 0.151,0.041, when $n = 50$.

4. Real Data Based Example

Myocardial infarction (MI) is commonly caused by blood clots blocking the blood flow of the heart leading heart muscle injury. The heart disease is leading cause of death affecting about or higher than 20% of populations regardless of different ethnicities according to the Centers for Disease Control and Prevention (e.g., Schisterman et al. [29]).

We illustrate the application of the proposed approach based on a sample from a study that evaluates biomarkers associated with MI. The study was focused on the residents of Erie and Niagara counties, 35-79 years of age. The New York State department of Motor Vehicles drivers' license rolls was used as the sampling frame for adults between the age of 35 and 65 years, while the elderly sample (age 65-79) was randomly chosen from the Health Care Financing Administration database. The biomarkers called "thiobarbituric acid-reactive substances" (TBARS), "vitamin E" and "glucose" are often used as discriminant factors between individuals with (MI=1) and without (MI=0) myocardial infarction disease (e.g., Schisterman et al. [29]). The sample of 50 biomarkers values was collected on cases who survived on MI.

Oftentimes, measurements related to biological processes follow a log-normal distribution (e.g., Limpert et al. [19]). The aim of this study is to investigate the joint distribution of log-transformed TBARS measurements, say X_1 , log-transformed vitamin E measurements, say X_2 ,

and log-transformed glucose measurements, say X_3 , when MI=1. Figure 1 depicts the histograms based on values of X_1 , X_2 and X_3 .

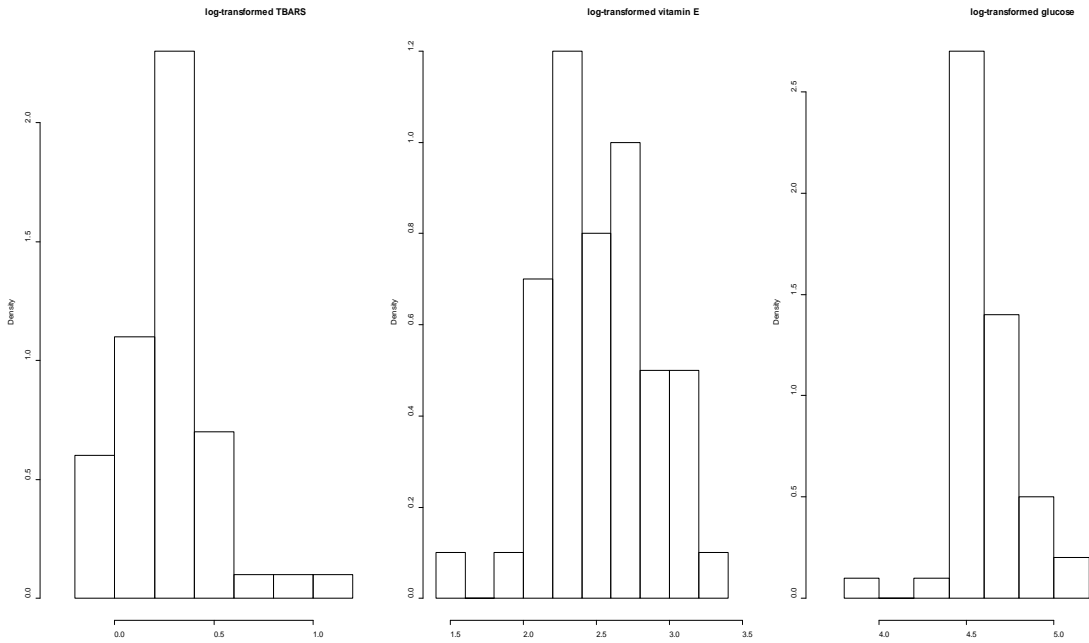


Fig. 1. Data based histograms related to observed values of log-transformed TBARS measurements, log-transformed vitamin E measurements and log-transformed glucose measurements.

We implemented the new test \hat{T}_n (see Remark 5) and the classical HZ, M, SW tests (see Section 3 for their descriptions). In this study, the considered four tests significantly reject the hypotheses that $(X_1, X_2, X_3)^T$ is trivariate normally distributed (H_0). Then, we organized a Bootstrap/Jackknife type study to examine the power performances of the test-statistics. The conducted strategy was that a sample with size $n = 35$ was randomly selected from the data to be tested for trivariate normality at 5% level of significance. We repeated this strategy 25,000 times calculating the frequencies of the events $\{\hat{T}_n \text{ rejects } H_0\}$, $\{\text{HZ rejects } H_0\}$, $\{\text{M rejects } H_0\}$ and $\{\text{SW rejects } H_0\}$. The obtained experimental powers of the four tests are 0.55, 0.29,

0.28, 0.40, respectively. In this study, the presented test outperforms the classical procedures in terms of the power properties when detecting that the log-transformed biomarkers' values are not jointly distributed as normal random variables. That is, the proposed test can be expected to be more sensitive as compared with the known methods to rejecting the null hypothesis of MN regarding the joint distribution of the log-transformed values of the biomarkers.

5. Concluding Remarks

This paper established new univariate likelihood based projections of the MN distribution. It can be attractive to release the conditions used in the presented theorems as well as extend and methodize the likelihood based concept to characterize different multivariate distributions (e.g., Costa and Hero [5]).

Using the likelihood based characterization of MN, we demonstrated an example of developing a new approach for testing of MN. The displayed procedure is simple and can be easily applied in practice, since reliable software products for performing modules of the likelihood projections based tests for MN are available. Through extensive Monte Carlo simulation studies and a real data based example, we showed that, employing the well-known tests based on univariate observations, we developed the strategy to assess MN that is superior to the classical procedures across a variety of settings when non-MN distributed vectors consist of normal variables. In future studies, many types of corresponding univariate-based plots can be constructed to be both easy to make and simple to use for detecting departures from assumed multivariate distributions. It is hoped that the present paper will convince the readers of the usefulness of multivariate distributions' characterizations via relevant likelihood functions.

Acknowledgments

The author thanks Professor Lev B. Klebanov, Charles University, Czech Republic, and Professor Abram Kagan, University of Maryland College Park, USA, for many insightful and

helpful discussions regarding the presented results. I am grateful to the Editor, the AE and reviewers for helpful comments.

Reference

- [1] Baldessari, B. (1967), The Distribution of a Quadratic Form of Normal Random Variables, *Ann. Math. Statist*, 38, 6, 1700-1704.
- [2] Baringhaus, L. and Henze, N. (1988), A Consistent Test for Multivariate Normality Based on the Empirical Characteristic Function, *Metrika*, 35, 339-348.
- [3] Bogdan, M. (1999), Data driven smooth tests for bivariate normality. *Journal of Multivariate Analysis*, 68, 26-53.
- [4] Bose, A., Dasgupta, A. and Rubin, H. (2002), A Contemporary Review and Bibliography of Infinitely Divisible Distributions and Processes, *Sankhyā: The Indian Journal of Statistics, Series A*, 64. 763-819.
- [5] Costa, J. A, and Hero, A. O. (2002), *A Characterization of the Multivariate Distributions Maximizing Renyi Entropy*. Proceedings IEEE International Symposium on Information Theory. Lausanne, 2002, p. 263. doi: 10.1109/ISIT.2002.1023535
- [6] Eaton, M. R., and Perlman, M. D. (1973), The non-singularity of generalized sample covariance matrices, *Ann Statist*, 1, 710-717
- [7] Golikova, N. N. and Kruglov, V. M. (2015), A Characterisation of the Gaussian Distribution through the Sample Variance, *Sankhyā: The Indian Journal of Statistics, Series A*, 77, 330-336.
- [8] Cuesta-Albertos, J .A., del Barrio, E., Fraiman, R. and Matrán, C. (2012), Letter to the editor. *Journal of Multivariate Analysis*, 103, 161-162.
- [9] Ducharme G. R. and Micheaux P. L. D., (2019), A goodness-of-fit test for elliptical distributions with diagnostic capabilities, <http://arxiv.org/abs/1902.03622v1>

- [10] Hamedani, G. G. (1984), Nonnormality of Linear Combinations of Normal Random Variables, *The American Statistician*, 38, 295-296.
- [11] Henze, N. (2002), Invariant tests for multivariate normality: a critical review, *Statistical Papers*, 43, 467-506.
- [12] Horn, R. A. and Steutel, F. W. (1978), On multivariate infinitely divisible distributions, *Stochastic Processes and Their Applications*, 6, 139-151.
- [13] Inglot, T. and Ledwina, T. (2006), Towards data driven selection of a penalty function for data driven Neyman tests, *Linear Algebra and its Appl*, 417, 579-590.
- [14] Kallenberg, W. C. M. and Ledwina, T. (1999), Data-Driven Rank Tests for Independence, *Journal of the American Statistical Association*, 94, 285-301.
- [15] Kim, I. and Park, S. (2018). Likelihood ratio tests for multivariate normality. *Communication in statistics-Theory and Methods*, 47, 1923-1934.
- [16] Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000), *Continuous Multivariate Distributions*. John Wiley & Sons, Inc. New York
- [17] Kruglov, V. M. (2013), A Characterization of the Gaussian Distribution, *Stochastic Analysis and Applications*, 31:5, 872-875, DOI: 10.1080/07362994.2013.817250
- [18] Ledwina, T. (1994), Data driven version of Neyman's smooth test of fit, *Journal of the American Statistical Association*, 89, 1000-1005.
- [19] Limpert, E., Stahel, W. A. and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize lognormal distributions, which can provide deeper insight into variability and probability—Normal or log-normal: That is the question. *BioScience*, 51(5), 341–352.

- [20] Looney, S. W. (1995), How to Use Tests for Univariate Normality to Assess Multivariate Normality, *The American Statistician*, 49, 64-70
- [21] Miecznikowski, J.C., Hsu, E., Chen, Y., and Vexler, A. (2018), testforDEP: An R Package for Modern Distribution-free Tests and Visualization Tools for Independence, *The R Journal*, 10/2, 282-294. DOI: 10.32614/RJ-2018-057
- [22] Nagaraja, H. (2006), *Characterizations of Probability Distributions*. In Pham, H., editor, Springer Handbook of Engineering Statistics, (pp. 79–95). Springer, London, UK. DOI: 10.1007/978-1-84628-288-1_4
- [23] Pierre, P. A. (1971), Infinitely Divisible Distributions, Conditions for Independence, and Central Limit Theorems, *Journal of Mathematical Analysis*, 33, 341-354.
- [24] Prokhorov A. V. and Ushakov N. G. (2002), On the problem of reconstructing a summands distribution by the distribution of their sum, *Theory of Probability and its Applications*, 46, 420–430.
- [25] R Development Core Team (2012), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 <<http://www.R-project.org>>.
- [26] Rao, M. M. (2012), *Random and Vector Measures*. World Scientific, Singapore.
- [27] Rayner, J. C. W., Thas O. and Best, D. J. (2009), *Smooth Tests of Goodness of Fit: Using R*. John Wiley, New York.
- [28] Ruben, H. (1978), On Quadratic Forms and Normality, *Sankhyā: The Indian Journal of Statistics*, Series A, 40, 156-173
- [29] Schisterman, E. F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong D., Reiser, B. and Trevisan. M. (2001). TBARS and cardiovascular disease in a population-based sample. *Journal of Cardiovascular Risk*, 8, 1–7

- [30] Shao, Y. and Zhou, M. (2010), A characterization of multivariate normality through univariate projections, *Journal of Multivariate Analysis*, 101, 2637-2640.
- [31] Stoyanov, J. M. (2014), *Counterexamples in Probability: Third Edition*. Dover Publications. New York.
- [32] Styan, G. P. H. (1970), Notes on the distribution of quadratic forms in singular normal variables, *Biometrika*, 57, 3, 667.
- [33] Szkutnik, Z. (1987), On invariant tests for multidimensional normality, *Probab. Math. Statist*, **8**, 1-10.
- [34] Veeh, J. (1982), Infinitely divisible measures with independent marginals, *Z. Wahrsch. Verw. Gebiete*, 61, 303-308.
- [35] Vexler, A. and Hutson, A. (2018), *Statistics in the Health Sciences: Theory, Applications, and Computing*. CRC Press, New York.
- [36] Wesolowski, J. (1993), Multivariate infinitely divisible distributions with the Gaussian second order conditional structure. *In Stability Problems for Stochastic Models* (Suzdal, 1991), Lecture Notes in Mathematics 1546, V.V. Kalashnikov and V.M. Zolotarev, eds., Springer, Berlin, 180-183
- [37] Zhu, L-X., Wong, H. L. and Fang, K-T. (1995), A test for multivariate normality based on sample entropy and projection pursuit, *Journal of Statistical Planning and Inference*, 45, 373-385.