Journal of Multivariate Analysis 128 (2014) 19–32

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A test for multivariate skew-normality based on its canonical form

N. Balakrishnan [a,d], A. Capitanio [b], B. Scarpa [c,*]

[a] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
[b] Department of Statistical Sciences, University of Bologna, Bologna, Italy
[c] Department of Statistical Sciences, University of Padua, Padua, Italy
[d] Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 26 June 2013
Available online 12 March 2014

AMS subject classifications:
62H05
62E15
15A21
62G30

Keywords:
Multivariate skewness
Skew-normal distribution
Canonical form
Order statistics

ABSTRACT

A test to assess if a sample comes from a multivariate skew-normal distribution is proposed. The test statistic is obtained from the canonical form of the multivariate skew-normal distribution and its null distribution is derived. The power of the proposed test is evaluated through Monte Carlo simulations for different conveniently chosen alternatives. Finally, three numerical examples are presented for the purpose of illustration.

1. Introduction

Traditionally, many multivariate procedures have been developed in the literature based on multivariate normality; see, for example, the books by Anderson [1] and Muirhead [20]. However, in several practical situations, the assumptions of multivariate normality are violated as the data possess some level of skewness.

In this regard, Azzalini [3] introduced the univariate skew-normal (SN) distribution, extending the normal one, through the introduction of a shape (or skewness) parameter λ. Later, Azzalini and Dalla Valle [7] presented a multivariate extension of Azzalini's SN that has been further explored by Azzalini and Capitanio [5]. Specifically, a d-dimensional random variable $\mathbf{Z}$ is said to have a multivariate SN distribution, with zero location and scale parameters equal to one, if it is continuous and has its density function (pdf) as

$$2\phi_d(\mathbf{z};\bar{\mathbf{\Omega}})\Phi(\boldsymbol{\alpha}^T\mathbf{z}),\quad \mathbf{z}\in\Re^d, \tag{1}$$

where $\phi_d(\mathbf{z};\bar{\mathbf{\Omega}})$ is the d-dimensional normal density with zero mean and correlation matrix $\bar{\mathbf{\Omega}}$, $\Phi(\cdot)$ is the univariate $N(0,1)$ distribution function, and $\boldsymbol{\alpha}$ is a d-dimensional vector that regulates departure from symmetry. The multivariate normal

* Corresponding author.
E-mail address: scarpa@stat.unipd.it (B. Scarpa).

http://dx.doi.org/10.1016/j.jmva.2014.02.015

belongs to the family of SN distributions when $\boldsymbol{\alpha} = \mathbf{0}$. So, even though one may proceed by assuming a multivariate skew-normal distribution for data possessing some level of skewness, one still needs to formally test this key assumption to validate the results obtained subsequently in the analysis.

While a number of procedures are available to test the skewness of a distribution (see, e.g., [9]) there has been very little work on testing for multivariate SN against other distributions. The present paper focuses on developing tests for multivariate skew-normality; on the basis of a random sample $\mathbf{y}_1, \ldots, \mathbf{y}_n$ of n independent copies of a multivariate vector $\mathbf{Y}^T = [Y_1, \ldots, Y_d]$ with a generic distribution function $F_d(\mathbf{y})$, we are interested in a testing procedure for the hypothesis testing problem

$$\begin{cases} H_0 : \mathbf{Y} \text{ follows a multivariate skew-normal distribution with some location-scale parameters vs.} \\ H_1 : \mathbf{Y} \text{ follows a distribution other than the multivariate skew-normal.} \end{cases} \tag{2}$$

When considering tests for multivariate skew-normality, three issues are of interest: one is about the statistical performance of the test procedure, second is about the ease of use and computational efforts, and third is about the generalization to higher dimension.

Tests for univariate skew-normality have been discussed by Mateu-Figueras et al. [17] and Meintanis [18], while for testing multivariate skew-normality, to our knowledge, only the procedure given by Meintanis and Hlavka [19] is available, which is based on the moment generating function with distribution obtained by bootstrap resampling, and is computationally available only for the bivariate case. The test proposed here is based on the canonical form associated with the multivariate skew-normal variable. More specifically, we exploit the property that the components of the canonical form are mutually independent, with a single component 'absorbing' all the asymmetry in the multivariate skew-normal distribution [5].

The rest of the paper is organized as follows. In Section 2, we briefly introduce the 'canonical form' of the multivariate SN and then discuss how we can, in practice, obtain the appropriate linear transform. In Section 3, we propose the goodness-of-fit test statistic and discuss its distribution under the null hypothesis, based on a single observation. Section 4 generalizes the procedure for the case when more observations are available by also allowing for unknown location and scale parameters in the distribution. In Section 5, we carry out a Monte Carlo simulation study to examine the performance of the proposed test, and for this purpose we consider some specific alternative distributions for the evaluation of the power of the proposed test. In Section 6, we present three real examples to illustrate the proposed procedure, and finally some concluding remarks are made in Section 7.

2. Canonical form of the multivariate SN

Proposition 4 of Azzalini and Capitanio [5] states that for a d-dimensional SN random variable $\mathbf{Z}$ with density (1), there exists a linear transform $\mathbf{Z}^* = \mathbf{A}^*\mathbf{Z}$, where $\mathbf{A}^*$ is a $d \times d$ non-singular matrix, such that $\mathbf{Z}^*$ is a skew-normal variate with density

$$2\phi_d(\mathbf{z}; \mathbf{I}_d)\Phi(\alpha_* z_m), \quad \mathbf{z} \in \Re^d,\ m \in \{1, \ldots, d\},$$

where $\alpha_* = (\boldsymbol{\alpha}^T \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{1/2}$ is the only non-zero component of the shape parameter of $\mathbf{Z}^*$.

The transformed variable $\mathbf{Z}^*$ is said to be the *canonical form*, which is not unique; but, its density factorizes into a product of $d - 1$ scalar $N(0, 1)$ densities and a scalar skew-normal density with shape (or skewness) parameter α_*. From the factorization of the density, it is immediate to note that the d marginal univariate components of $\mathbf{Z}^*$ are independent.

It should be mentioned that this proposition states the existence of the canonical form, but it is not constructive. A method for obtaining the canonical form is described in the following proposition, given in [11], which considers the more general form of SN obtained when a vector $\boldsymbol{\xi}$ of location and a diagonal matrix $\boldsymbol{\omega}$ of scale parameters, such that $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$ is a covariance matrix, are introduced through the form

$$\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}.$$

Henceforth, we shall use the notation $\mathbf{Y} \sim SN_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ for this general form of SN. Then, the following proposition holds.

Proposition 1. *Let $\mathbf{Y} \sim SN_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. Consider the affine non-singular transform*

$$\mathbf{Y}^* = (\mathbf{C}^{-1}\mathbf{P})^T \boldsymbol{\omega}^{-1}(\mathbf{Y} - \boldsymbol{\xi}),$$

where $\mathbf{C}^T\mathbf{C} = \bar{\boldsymbol{\Omega}}$ and $\mathbf{P}$ is an orthogonal matrix having the first column to be proportional to $\mathbf{C}\boldsymbol{\alpha}$. Then,

$$\mathbf{Y}^* \sim SN(0, \mathbf{I}_d, \boldsymbol{\alpha}_{\mathbf{Y}^*}),$$

where $\boldsymbol{\alpha}_{\mathbf{Y}^*} = [\alpha_*, 0, \ldots, 0]^T$ and $\alpha_* = (\boldsymbol{\alpha}^T \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{1/2}$.

This result ensures the existence of the canonical form for the location-scale SN family. Note that the symmetric components of $\mathbf{Y}^*$ are standardized, having mean 0 and variance 1. In the following, without loss of generality, we always choose as a skew component the first one.

In [11], the following proposition is given, that gives a constructive approach for obtaining such a canonical form for the SN variable.