A general method for constructing a test of multivariate normality

Tejas A. Desai*

The Indian Institute of Management, Vastrapur, Ahmedabad--380015, Gujarat, India

Abstract

We present a general method of constructing a test of multivariate normality using any given test of univariate normality of complete or randomly incomplete data. A simulation study considers multivariate tests constructed using the univariate versions of the Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von-Mises, and Anderson-Darling tests.

MSC: 62H15; 62H99

Keywords: Anderson-Darling test; Cramer-Von-Mises test; Kolmogorov-Smirnov test; Missingness at Random (MAR); Multivariate normality; Shapiro-Wilk test.

1. Introduction

Many tests of multivariate normality exist. Reviews by Gnanadesikan (1997) and Mardia (1980) outline many of these procedures. However, there is yet no single test of multivariate normality which is sensitive to all types of violation of multivariate normality. The aim of this paper is to demonstrate that given a test for *univariate* normaility which is sensitive to all kinds of violation of *univariate* normality, one can construct a test of multivariate normality that is sensitive to all kinds of *multivariate* normality. Section 2 demonstrates the general procedure of constructing such a multivariate test and gives a general result concerning the Type I error and the power of the multivariate test thus constructed and applied to complete multivariate data.

**E-mail address:* tdesai@iimahd.ernet.in

Section 3 extends the methodlogy of Section 2 to the case of randomly incomplete (MAR) multivariate data. Section 4 presents a simulation study wherein the Type 1 error and power of multivariate tests constructed using four widely used tests of univariate normality: the Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-Von-Mises, and Anderson-Darling tests. Simulations are performed for both complete and randomly incomplete multivariate data.

2. General methodology for complete data

Suppose X is a one-dimensional random variable whose distribution is unknown. Consider the following null and alternative hypotheses:

$$H_o : X \text{ is normally distributed and } H_a : \; X \text{ is not normally distributed} \tag{1}$$

Suppose that W is a statistic for testing the above hypotheses. Let $\alpha \in (0,1)$ denote any fixed significance level of the test; $n,$ a given sample size; $S(n)$, the function of sample size giving the true Type 1 error of the test for sample size n when H_o is true; and $P(n) \in [0,1]$, also a function of n giving the true power of the test for sample size n when H_o is false. Furthermore, suppose that

$$\underset{n \to \infty}{lim} S(n) = \alpha \; \text{ if } H_o \text{ is true and } \underset{n \to \infty}{lim} P(n) = 1 \text{ if } H_a \text{ is true} \tag{2}$$

Now suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ is a p-dimensional $(p \geq 1)$ random vector whose distribution is unknown and such that the distribution is continuous and defined over all of $\mathbb{R}^p$. Consider the following hypotheses:

$$H_o : \boldsymbol{X} \text{ is normally distributed and } H_a : \; \boldsymbol{X} \text{ is not normally distributed} \tag{3}$$

Suppose that we have a sample of random realizations of $\boldsymbol{X}$, namely, $\boldsymbol{X_n}$ where

$$\boldsymbol{X_n} = \begin{pmatrix} x_{11} & \ldots & x_{i1} & \ldots & x_{p1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1j} & \ldots & x_{ij} & \ldots & x_{pj} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1n} & \ldots & x_{in} & \ldots & x_{pn} \end{pmatrix}$$

We wish to test the null hypothesis in (3) using the significance level α stated above. Before we proceed to construct a test of the null hypothesis in (3), we establish some useful notation. Note that under the null hypothesis, the density of the (normal) distribution of $\boldsymbol{X}$ with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_p^2 \end{pmatrix}$$

may be written as

$$f(\boldsymbol{X}|\boldsymbol{\mu}, \mathbf{\Sigma}) = g_1(X_1) \times \prod_{i=2}^{n} g_i(X_i|X_1, \ldots, X_{i-1})$$

where

$$g_1(X_1) \text{ is the density of } N(\mu_1, \sigma_1^2)$$

and

$$g_i(X_i|X_1, \ldots, X_{i-1}) \text{ is the density of } \boldsymbol{N}(h_i(X_i|X_1, \ldots, X_{i-1}), \boldsymbol{\Phi_i}),\ i = 2, \ldots, p.$$

In the expression above,

$$h_i(X_i|X_1, \ldots, X_{i-1}) = \mu_i + \boldsymbol{\Sigma_{i,i-1}}\boldsymbol{\Sigma^{-1}_{i-1,i-1}}((X_1, \ldots, X_{i-1}) - (\mu_1, \ldots, \mu_{i-1}))', i = 2, \ldots, p$$

and

$$\boldsymbol{\Phi_i} = \sigma_i^2 - \boldsymbol{\Sigma_{i,i-1}}\boldsymbol{\Sigma^{-1}_{i-1,i-1}}(\boldsymbol{\Sigma_{i,i-1}})', , i = 2, \ldots, p.$$

Note that

$$X_i = h_i(X_i|X_1, \ldots, X_{i-1}) + \epsilon$$

where

$$\epsilon \sim N(0, \boldsymbol{\Phi_i})$$

In the expressions above, $\boldsymbol{\Sigma_{i,i-1}}$ is the submatrix of $\boldsymbol{\Sigma}$ corresponding to the covariances of X_i with $X_1, \ldots X_{i-1}$, and $\boldsymbol{\Sigma_{i-1,i-1}}$ is the submatrix of $\boldsymbol{\Sigma}$ equal to the covariance matrix of $X_1, \ldots, X_{i-1}$. Let $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_p)'$ and

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \widehat{\sigma}_1^2 & \widehat{\sigma}_{12} & \ldots & \widehat{\sigma}_{1p} \\ \widehat{\sigma}_{12} & \widehat{\sigma}_2^2 & \ldots & \widehat{\sigma}_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\sigma}_{1p} & \widehat{\sigma}_{2p} & \ldots & \widehat{\sigma}_p^2 \end{pmatrix}$$

be the maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the null hypotheses in (3). For $i = 2, \ldots, p$, let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1}}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i-1,i-1}}$ be the maximum-likelihood estimates of $\boldsymbol{\Sigma_{i,i-1}}$ and $\boldsymbol{\Sigma_{i-1,i-1}}$, respectively. Finally, for $i = 2, \ldots, p$ and $j = 1, \ldots, n$, let

$$\widehat{h}_i(x_{ij}|x_{1j}, \ldots, x_{i-1,j}) = \widehat{\mu}_i + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i-1,i-1}}^{\boldsymbol{-1}} ((x_{1j}, \ldots, x_{i-1,j}) - (\widehat{\mu}_1, \ldots, \widehat{\mu}_{i-1}))'$$

and

$$\widehat{\boldsymbol{\Phi}}_{\boldsymbol{i}} = \widehat{\sigma}_i^2 - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1}} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i-1,i-1}}^{\boldsymbol{-1}} \left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1}} \right)'$$

Now we proceed to construct a stepwise test of the null hypothesis in (3). Let $\alpha_p = \alpha/p$. Consider the following stepwise procedure with a maximum of p possible steps:

Step 1: Using statistic W, test the null hypothesis in (1) using the univariate sample $(\, x_{11}, \quad \ldots \quad , x_{1j}, \quad \ldots \quad , x_{1n}\,)'$. Let k_1 equal to the p-value obtained as a result of the test. If $k_1 < \alpha_p$, then reject the null in (3), and stop the testing procedure at this step. Else, proceed to the next step.

Step i, $i > 1$: Consider the transformation

$$y_{ij} = x_{ij} - \widehat{h}_i(x_{ij}|x_{1j}, \ldots, x_{i-1,j}),\ j = 1, \ldots, n, \text{ and} \tag{4}$$

Using statistic W, test the univariate sample $(\, y_{i1}, \quad \ldots \quad , y_{ij}, \quad \ldots \quad , y_{in}\,)'$ for normality. Let k_i be the resulting p-value. If $k_i < \alpha_p/(1-\alpha_p)^{i-1}$, then stop the procedure and reject the null hypothesis in (3); else go to step $i+1$.

Let $S^*(n)$ denote the function of n which is equal to the Type I error of the above step-wise test for sample size n when H_o is true. Let $P^*(n)$ denote the function of n which is equal to the power of the above step-wise test for sample size n when H_o is false. Then we have the following fundamental result

Theorem 1. *Suppose* $(X_1, \ldots, X_p)$ *is a random vector having a continuous distribution defined over all of* $\mathbb{R}^P$. *Then under assumption* (2), *it follows for the above step-wise test that*

$$\underset{n \to \infty}{lim} S^*(n) \leq \alpha \ \text{ if } H_o \text{ is true and } \underset{n \to \infty}{lim} P^*(n) = 1 \text{ if } H_a \text{ is true}$$

Proof. Suppose that H_o is true. Then let $A_i,\ i = 1, \ldots, p,$ be indicator variables whose values are set to 0 prior to beginning the step-wise testing procedure. Moreover, each A_i is updated if and only if the testing procedure reaches step i. The updating is as follows:

$$A_i = \begin{cases} 1 \text{ if the null is rejected in Step } i \\ 0 \text{ otherwise} \end{cases}$$

If the null hypothesis is true, then $(\ x_{11}, \quad \ldots \quad, x_{1j}, \quad \ldots \quad, x_{1n}\ )'$ is a random sample from $N(\mu_1, \sigma_1^2)$. (5)

Moreover, when H_o is true, then for i, $1 < i \leq p$, $(\ y_{i1}, \quad \ldots \quad, y_{ij}, \quad \ldots \quad, y_{in}\ )'$ represents a sample of size n from the distribution of a random variable $Y^{(n)}$ such that, because of (4),

$$Y^{(n)} \xrightarrow{p} N\big(0, \sigma_i^2 - \boldsymbol{\Sigma_{i,i-1}} \boldsymbol{\Sigma^{-1}_{i-1,i-1}} (\boldsymbol{\Sigma_{i,i-1}})'\big) \tag{6}$$

Thus, it follows that when H_o is true

$$S^*(n) = Pr(A_1 = 1 \text{ or } A_2 = 1 \text{ or} \ldots \text{or } A_p = 1) \xrightarrow{p} \sum_{i=2}^{p} Pr(A_i = 1) \leq \alpha^*$$

where

$$\alpha^* = = \sum_{i=1}^{p} \left(1 - \frac{\alpha}{p}\right)^{i-1} \cdot \left[\alpha_p / \left(1 - \frac{\alpha}{p}\right)^{i-1}\right] = \alpha$$

Now suppose that H_a is true. Then (5) and/or (6) are violated for some j,$1 \leq j \leq p$.

Let j^* be the smallest such j. Let $P_{j^*}(n)$ denote the function of n which is equal to the power of the test conditional on the event that H_o is rejected at step j^* of the above step-wise test for sample size n. Then, since H_a is true, (2) implies that

$$\underset{n \longrightarrow \infty}{lim} P_{j^*}(n) = 1.$$

Note that

$$P^*(n) \geq P_{j^*}(n) \tag{7}$$

Therefore, (7) implies that

$$\underset{n \longrightarrow \infty}{lim} P^*(n) = 1 \text{ if } H_a \text{ is true} \qquad \square$$

3. General methodology for randomly incomplete data (MAR)

When $\boldsymbol{X_n}$ is randomly incomplete, the first step is to find the estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$. Unlike in the case of complete data, this estimate may not have an analytical form. However,$\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ may be obtained by using an iterative algorithm such as the EM or Newton-Raphson algorithms to maximize the likelihood of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ given the observed data. To avoid problems of non-estimability of parameters, we assume that the probability of *all* the p components being observed is greater than 0, but may be less than 1. The methodology of section 2 is then easily extended to the MAR case:

Step 1: Same as Step 1 in section 2, except that we use only the observed values among $x_{11}, \ldots, x_{1n}$. Let k_1 equal to the p-value obtained as a result of the test. If $k_1 < \alpha_p$, then reject the null in (3), and stop the testing procedure at this step. Else, proceed to the next step.

Step i, $i > 1$: For $j = 1, \ldots, n$, repeat the following process. If x_{ij} is missing then set y_{ij} to a missing value. If x_{ij} is observed and $x_{1j}, \ \ldots, x_{i-1,j}$ are all missing then let

$$y_{ij} = \left(\frac{x_{ij} - \widehat{\mu}_i}{\widehat{\sigma}_i} \right)$$

where $\widehat{\boldsymbol{\Phi}}_{\boldsymbol{i}}$ is defined as in Section 2. If If x_{ij} is observed and some but not all of $x_{1j}, \ldots, x_{i-1,j}$ are observed, then let $\boldsymbol{x_{i-1,obs}}$ denote the observed subvector of $x_{1j}, \ldots, x_{i-1,j}$, and $\widehat{\boldsymbol{\mu}}_{\boldsymbol{i-1,obs}}$ be the *row* vector of corresponding entries in $\widehat{\boldsymbol{\mu}}$**.** Let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1,obs}}$ be the submatrix of $\widehat{\boldsymbol{\Sigma}}$ corresponding to the covariance of x_{ij} with $\boldsymbol{x_{i-1,obs}}$**.** Let $\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{i-1,i-1,obs}}$ be the submatrix of $\widehat{\boldsymbol{\Sigma}}$ equal to the (estimated) covariance matrix of $\boldsymbol{x_{i-1,obs}}$. Then let

$$\widehat{h}_i(x_{ij}|\boldsymbol{x_{i-1,obs}}) = \widehat{\mu}_i + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1,obs}}\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{i-1,i-1,obs}}(\boldsymbol{x_{i-1,obs}} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{i-1,obs}})' \text{ and}$$

$$y_{ij} = (x_{ij} - \widehat{h}_i)\Big/\sqrt{\widehat{\sigma}_i^2 - \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1,obs}}\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{-1}}_{\boldsymbol{i-1,i-1,obs}}\left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{i,i-1,obs}}\right)^{\boldsymbol{\prime}}}$$

Using statistic W, test the observed values among $y_{i1}, \ldots, y_{in}$ for normality. Let k_i be the resulting p-value. If $k_i < \alpha_p/\left(1 - \frac{\alpha}{p}\right)^{i-1}$, then stop the procedure and reject the null hypothesis in (3); else go to step $i+1$.

Then we have the following general result. Let $S^*(n)$ denote the function of n which is equal to the Type I error of the above step-wise test for sample size n when H_o is true. Let $P^*(n)$ denote the function of n which is equal to the power of the above step-wise test for sample size n when H_o is false. Then we have the following fundamental result

Theorem 2. *Suppose $\boldsymbol{X} = (X_1, \ldots, X_p)$ is a random vector having a continuous distribution defined over all of $\mathbb{R}^P$. Suppose that $\boldsymbol{X}$ is subject to random missingness such that the probability of all the p components of $\boldsymbol{X}$ being observed is greater than 0, but may be less than 1. Then under assumption* (2)*, it follows for the above step-wise test that*

$$\underset{n \to \infty}{lim} S^*(n) \le \alpha \text{ if } H_o \text{ is true and } \underset{n \to \infty}{lim} P^*(n) = 1 \text{ if } H_a \text{ is true}$$

The proof of the above theorem is very similar to that of Theorem 1, and hence is not repeated here. The only difference in this case is that the asymptotic law (6) becomes

$$Y^{(n)} \xrightarrow{p} N(0,1)$$

4. Simulation

4.1 Complete Data Case

Complete data sets of different sizes were simulated from a tri-variate normal distribution with parameters

$$\boldsymbol{\mu} = (0,0,0)',\ \boldsymbol{\Sigma} = \begin{pmatrix} 1 & -0.3 & 0.5 \\ -0.3 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \text{ and } \alpha = 0.05$$

For testing the null hypothesis in (3), four multivariate tests were constructed by applying the methodology presented in Section 3 to four different statistics for assessing univariate normality, namely, the Shapiro-Wilk (SW), Kolmogorov-Smirnov (KS), Cramer-Von-Mises (CVM), and Anderson-Darling (AD) statistics. For each sample size considered, 1000 simulations were run. The rates of rejection of H_o in (3) when H_o is true are given in Table 1.

To investigate power, we considered a considered a bi-variate random variable (X, Y) such that X and Y are independent, X is distributed as standard normal, while Y is distributed as Student's t with 4 degrees of freedom. Thus the distribution of (X, Y) is symmetric but non-normal. Once again, for each sample size considered, 1000 simulations were run. The rates of rejection of H_o in (3) when H_o is false are given in Table 2.

Table 1 and Table2 suggest that of the 4 multivariate tests, the one based on Kolmogorov-Smirnoff statistic is the most conservative, while the one based on Shapiro-Wilk statistic is the most anti-conservative, though all four tests exhibit the asymptotic properties outlined in Theorem 1.

4.2 The Case of Randomly Incomplete (MAR) Data

To investigate Type I error when the null hypothesis H_o in (3) is true and when data is missing at random, random missingness was created in the trivariate normal data simulated in section 4.1. Note that since $p = 3$, there are 8 possible missingness patterns. Each pattern may be described using a 3-dimensional indicator vector $\boldsymbol{Y}$ such that a given component of $\boldsymbol{Y}$ is 1 if the corresponding component of $\boldsymbol{X}$ is observed, and is 0 otherwise. The missingness patterns were generated using the probabilities in $Table\,3$. For each sample size considered, 1000 simulations were run with $\alpha = 0.05$. The Type 1 errors under H_o and random missingness are given in Table 4. Note that none of the rejection rates in Table 4 is significantly different from 0.05 at the 5% significance level. Table 4 illustrates that, just as in the complete-data case, the Type 1 error of each of the 4 tests approaches $\alpha = 0.05$ as the sample size increases, albeit somewhat faster than in the complete-data case.

To investigate power, the same bivariate random vector (X, Y) was considered as in the investigation of power in Section 4.1. Missing data were simulated using the missingness probabilities in Table 5. For each sample size considered, 1000 simulations were run. The rates of rejection of H_o in (3) when H_o is false are given in Table 6. It is clear from Table 6 that, just as in the complete-data case, the power of each of the 4 tests approaches 1 as the sample size increases, albeit more slowly than in the complete-data case. Furthermore, Tables 4 and 6 underscore the afore-mentioned observation that the test based on the Kolmogorov-Smirnoff statistic is the most conservative, whereas the test based the Shapiro-Wilk statistic is the most anti-conservative.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Hoboken: Wiley.

Gantz, B. J. et al. (1988). Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance. *Laryngoscope* **98**, 1100-1106.

Gnanadesikan, R. (1997). *Methods for statistical data analysis of multivariate observations*. New York: Wiley

Johnson, R. A., Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*. Upper Saddle River: Prentice Hall

Mardia, K. V., 1980. Tests for univariate and multivariate normality. In *Handbook of Statistics, Vol 1: Analysis of Variance*, P. R. Krishnaiah (ed.), 279-320, Amsterdam: North Holland.

$Table\,1$: Rates of rejection of H_o when H_o is true and data is complete

Sample size	Underlying statistic			
	SW	KS	CVM	AD
$n = 5$	0.048	0.050	0.041	0.042
$n = 15$	0.049	0.058	0.052	0.051
$n = 30$	0.060	0.059	0.059	0.056

$Table\,2$: Rates of rejection of H_o when H_o is false and data is complete

Sample size	Underlying statistic			
	SW	KS	CVM	AD
$n = 100$	0.635	0.349	0.634	0.632
$n = 250$	0.939	0.716	0.942	0.941
$n = 500$	0.999	0.966	0.995	0.993

Table 3: Probabilities of missingness patterns under H_o

Pattern $\boldsymbol{Y}$	$Pr(\boldsymbol{Y})$
$(1, 1, 1)$	0.70
$(1, 1, 0)$	0.08
$(1, 0, 1)$	0.08
$(1, 0, 0)$	0.02
$(0, 1, 1)$	0.08
$(0, 1, 0)$	0.02
$(0, 0, 1)$	0.02
$(0, 0, 0)$	0.00

$Table\,4$: Rates of rejection of H_o when H_o is true and data is randomly incomplete

Sample size	Underlying statistic			
	SW	KS	CVM	AD
$n = 10$	0.054	0.044	0.049	0.050
$n = 15$	0.049	0.048	0.048	0.049
$n = 30$	0.056	0.053	0.057	0.053

$Table\, 5$: Probabilities of missingness patterns under H_o

Pattern $\boldsymbol{Y}$	$Pr(\boldsymbol{Y})$
$(1, 1)$	0.70
$(1, 0)$	0.15
$(0, 1)$	0.15
$(0, 0)$	0.00

$Table\, 6$: Rates of rejection of H_o when H_o is false and data is randomly incomplete

Sample size	Underlying statistic			
	SW	KS	CVM	AD
$n = 100$	0.483	0.250	0.482	0.480
$n = 250$	0.838	0.555	0.841	0.840
$n = 500$	0.984	0.869	0.980	0.978