Journal of Statistics: Advances in Theory and Applications
Volume 1, Number 2, 2009, Pages 207-220

ON JARQUE-BERA TESTS FOR ASSESSING MULTIVARIATE NORMALITY

KAZUYUKI KOIZUMI[1], NAOYA OKAMOTO[2] and TAKASHI SEO[1]

[1]Department of Mathematical Information Science
Faculty of Science
Tokyo University of Science
Japan
e-mail: koizu702@yahoo.co.jp

[2]Department of Food Sciences
Faculty of Health and Nutrition
Tokyo Seiei College
Japan

Abstract

In this paper, we consider some tests for the multivariate normality based on the sample measures of multivariate skewness and kurtosis. Sample measures of multivariate skewness and kurtosis were defined by Mardia [3], Srivastava [9] and so on. We derive new multivariate normality tests by using Mardia's and Srivastava's moments. For univariate case, Jarque and Bera [1] proposed bivariate test using skewness and kurtosis. We propose some new bivariate tests for assessing multivariate normality which are natural extensions of Jarque-Bera test. Finally, the numerical results by Monte Carlo simulation are shown in order to evaluate accuracy of expectations, variances and upper percentage points for new test statistics proposed in this paper.

2000 Mathematics Subject Classification: 62E20, 62H10.

Keywords and phrases: Jarque-Bera test, multivariate skewness, multivariate kurtosis, normality test.

Received June 31, 2009

1. Introduction

In statistical analysis, the test for normality is an important problem. This problem has been considered by many authors. Shapiro and Wilk [8]'s *W*-statistic is well known as the univariate normality test. For the multivariate case, some tests based on *W*-statistic were proposed by Malkovich and Afifi [2], Royston [6], Srivastava and Hui [10] and so on. Mardia [3] and Srivastava [9] gave different definitions of the multivariate measures of skewness and kurtosis, and discussed the test statistics using these measures for assessing multivariate normality, respectively. Mardia [4] derived exact expectations and variances of multivariate sample skewness and kurtosis, and discussed their asymptotic distributions. Srivastava [9]'s sample measures of multivariate skewness and kurtosis have been discussed by many authors. Seo and Ariga [7] derived normalizing transformations of test statistic using Srivastava's kurtosis by the asymptotic expansion. Okamoto and Seo [5] derived the exact expectation and variance of Srivastava's skewness and improved test statistic for assessing multivariate normality.

On the other hand, for univariate sample case, Jarque and Bera [1] proposed the bivariate test using skewness and kurtosis. Improved Jarque-Bera tests have been considered by many authors (see, e.g., Urzúa [11]). It seems that Jarque-Bera test for multivariate case has not been considered by any author. Our purpose is to propose new Jarque-Bera tests for assessing multivariate normality by using Mardia's and Srivastava's skewness and kurtosis, respectively. New test statistics are asymptotically distributed as χ^2-distribution. We investigate accuracy of expectations, variances and upper percentage points for multivariate Jarque-Bera tests by Monte Carlo simulation.

2. Multivariate Measures of Skewness and Kurtosis

2.1. Mardia's skewness and kurtosis

Let $\boldsymbol{x} = (x_1, x_2, \dots, x_p)'$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_p)'$ be random p-vectors distributed identically and independently with mean vector $\boldsymbol{\mu} =$

$(\mu_1, \mu_2, \ldots, \mu_p)'$ and covariance matrix Σ. Mardia [3] has defined the population measures of multivariate skewness and kurtosis as

$$\beta_{M,1} = E[\{(\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\}^3],$$

$$\beta_{M,2} = E[\{(\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}^2],$$

respectively. When $p = 1$, $\beta_{M,1}$ and $\beta_{M,2}$ are reduced to the ordinary univariate measures. It is obvious that for any symmetric distribution about $\boldsymbol{\mu}$, $\beta_{M,1} = 0$. For a normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$,

$$\beta_{M,1} = 0, \quad \beta_{M,2} = p(p+2).$$

To give the sample counterparts of $\beta_{M,1}$ and $\beta_{M,2}$, let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be samples of size N from a multivariate p-dimensional population. And let $\overline{\boldsymbol{x}}$ and S be the sample mean vector and the sample covariance matrix as follows:

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{x}_j,$$

$$S = \frac{1}{N}\sum_{j=1}^{N} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})',$$

respectively.

Then Mardia [3] has defined the sample measures of skewness and kurtosis by

$$b_{M,1} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' S^{-1}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})\}^3,$$

$$b_{M,2} = \frac{1}{N}\sum_{i=1}^{N} \{(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' S^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\}^2,$$

respectively.

Mardia [3, 4] has given the following lemma.

Lemma 1 (Mardia [3, 4]). *The exact expectation of* $b_{M,1}$ *and expectation and variance of* $b_{M,2}$ *when the population is* $N_p(\boldsymbol{\mu}, \Sigma)$ *are given by*

$$E(b_{M,1}) = \frac{p(p+2)}{(N+1)(N+3)}\{(N+1)(p+1)-6\},$$

$$E(b_{M,2}) = \frac{p(p+2)(N-1)}{N+1},$$

$$Var(b_{M,2}) = \frac{8p(p+2)(N-3)}{(N+1)^2(N+3)(N+5)}(N-p-1)(N-p+1).$$

Furthermore, Mardia [3] has obtained asymptotic distributions of $b_{M,1}$ and $b_{M,2}$ and used them to test the multivariate normality.

Theorem 1 (Mardia [3]). *Let* $b_{M,1}$ *and* $b_{M,2}$ *be the sample measures of multivariate skewness and kurtosis, respectively, on the basis of a random sample of size N drawn from* $N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$. *Then*

$$z_{M,1} \equiv \frac{N}{6} b_{M,1}$$

is asymptotically distributed as χ^2*-distribution with* $f \equiv p(p+1)(p+2)/6$ *degrees of freedom, and*

$$z_{M,2} \equiv \sqrt{\frac{N}{8p(p+2)}}(b_{M,2} - p(p+1))$$

is asymptotically distributed as $N(0, 1)$.

By making reference to moments of $b_{M,1}$ and $b_{M,2}$, Mardia [4] considered the following approximate test statistics as competitors of $z_{M,1}$ and $z_{M,2}$:

$$z^*_{M,1} = \frac{N}{6} b_{M,1} \frac{(p+1)(N+1)(N+3)}{N\{(N+1)(p+1)-6\}} \sim \chi^2_f \qquad (2.1)$$

asymptotically, and

$$z^*_{M,2} = \frac{\sqrt{(N+3)(N+5)}\{(N+1)b_{M,2} - p(p+2)(N-1)\}}{\sqrt{8p(p+2)(N-3)(N-p-1)(N-p+1)}} \sim N(0, 1) \quad (2.2)$$

asymptotically. It is noted that $z^*_{M,1}$ is formed so that $E(z^*_{M,1}) = f$.

2.2. Srivastava's skewness and kurtosis

Let $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$ be an orthogonal matrix such that $\Gamma' \Sigma \Gamma = D_\lambda$, where $D_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Note that $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of Σ. Then, Srivastava [9] defined the population measures of multivariate skewness and kurtosis as follows:

$$\beta_{S,1} = \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{E[(v_i - \theta_i)^3]}{\lambda_i^{\frac{3}{2}}}\right\}^2,$$

$$\beta_{S,2} = \frac{1}{p}\sum_{i=1}^{p}\frac{E[(v_i - \theta_i)^4]}{\lambda_i^2},$$

respectively, where $v_i = \gamma_i' \boldsymbol{x}$ and $\theta_i = \gamma_i' \boldsymbol{\mu} (i = 1, 2, \ldots, p)$. We note that $\beta_{S,1} = 0, \beta_{S,2} = 3$ under a multivariate normal population. Let $H = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$ be an orthogonal matrix such that $H'SH = D_w$, where $D_w = \mathrm{diag}(w_1, w_2, \ldots, w_p)$ and $w_1, w_2, \ldots, w_p$ are the eigenvalues of S. Then, Srivastava [9] defined the sample measures of multivariate skewness and kurtosis as follows:

$$b_{S,1} = \frac{1}{N^2 p}\sum_{i=1}^{p}\left\{w_i^{-\frac{3}{2}}\sum_{j=1}^{N}(v_{ij} - \bar{v}_i)^3\right\}^2,$$

$$b_{S,2} = \frac{1}{Np}\sum_{i=1}^{p} w_i^{-2}\sum_{j=1}^{N}(v_{ij} - \bar{v}_i)^4,$$

respectively, where $v_{ij} = \boldsymbol{h}_i' \boldsymbol{x}_j$, $\bar{v}_i = (1/N)\sum_{j=1}^{N} v_{ij}$.

Srivastava [9] obtained the following lemma:

Lemma 2 (Srivastava [9]). *For large N, the expectations of* $\sqrt{b_{S,1}}$ *and* $b_{S,1}$ *and expectation and variance of* $b_{S,2}$ *when the population is* $N_p(\boldsymbol{\mu}, \Sigma)$ *are given by*

$$E(\sqrt{b_{S,1}}) = 0, \; E(b_{S,1}) = \frac{6}{N},$$

$$E(b_{S,2}) = 3, \; Var(b_{S,2}) = \frac{24}{Np}.$$

By using Lemma 2, Srivastava [9] derived the following theorem:

Theorem 2 (Srivastava [9]). *Let* $b_{S,1}$ *and* $b_{S,2}$ *be the sample measures of multivariate skewness and kurtosis, respectively, on the basis of a random sample of size N drawn from* $N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$. *Then*

$$z_{S,1} \equiv \frac{Np}{6} b_{S,1}$$

is asymptotically distributed as χ^2*-distribution with p degrees of freedom, and*

$$z_{S,2} \equiv \sqrt{\frac{Np}{24}}(b_{S,2} - 3)$$

is asymptotically distributed as $N(0, 1)$.

Further, Okamoto and Seo [5] gave the expectation of multivariate sample skewness $b_{S,1}$ without using Taylor expansion. By using the same way as Okamoto and Seo [5], we can obtain the expectation and variance of multivariate sample kurtosis $b_{S,2}$. Hence, we can get the following lemma:

Lemma 3. *For large N, we give mean of* $b_{S,1}$ *and mean and variance of* $b_{S,2}$ *when the population is* $N_p(\boldsymbol{\mu}, \Sigma)$.

$$E(b_{S,1}) = \frac{6(N-2)}{(N+1)(N+3)},$$

$$E(b_{S,2}) = \frac{3(N-1)}{N+1},$$

$$Var(b_{S,2}) = \frac{24}{p} \frac{N(N-2)(N-3)}{(N+1)^2(N+3)(N+5)},$$

respectively.

Also, Seo and Ariga [7] gave asymptotic expansions of expectation and variance of $b_{S,2}$. By making reference to moments of $b_{S,1}$ and $b_{S,2}$, we consider following approximate test statistics as competitors of $z_{S,1}$ and $z_{S,2}$:

$$z_{S,1}^{*} = \frac{(N+1)(N+3)}{6(N-2)} p b_{S,1} \sim \chi_p^2 \tag{2.3}$$

asymptotically, and

$$z_{S,2}^{*} = \frac{\sqrt{p(N+3)(N+5)}\{(N+1)b_{S,2} - 3(N-1)\}}{\sqrt{24N(N-2)(N-3)}} \sim N(0, 1) \tag{2.4}$$

asymptotically.

3. Multivariate Jarque-Bera Tests

In this section, we consider new tests for multivariate normality when the population is $N_p(\boldsymbol{\mu}, \Sigma)$. From Theorem 1, we propose a new test statistic using Mardia's measures as follows:

$$MJB_M = N\left\{\frac{b_{M,1}}{6} + \frac{(b_{M,2} - p(p+2))^2}{8p(p+2)}\right\}.$$

MJB_M statistic is asymptotically distributed as χ_{f+1}^2-distribution.

From Theorem 2, we propose a new test statistic using Srivastava's measures as follows:

$$MJB_S = Np\left\{\frac{b_{S,1}}{6} + \frac{(b_{S,2} - 3)^2}{24}\right\}.$$

MJB_S statistic is asymptotically distributed as χ^2_{p+1}-distribution.

Further, by using (2.1) and (2.2), modified MJB_M is given by

$$MJB_M^* = z_{M,1}^* + z_{M,2}^{*2}.$$

In the same as MJB_M, this statistic MJB_M^* is distributed as χ^2_{f+1}-distribution asymptotically.

Also, by using (2.3) and (2.4), modified MJB_S is given by

$$MJB_S^* = z_{S,1}^* + z_{S,2}^{*2}.$$

In the same as MJB_S, this statistic MJB_S^* is distributed as χ^2_{p+1}-distribution asymptotically.

4. Simulation Studies

Accuracy of expectations, variances and upper percentage points of multivariate Jarque-Bera tests MJB_M, MJB_S, MJB_M^* and MJB_S^* is evaluated by Monte Carlo simulation study. Simulation parameters are as follows: $p = 3, 10, 20$, $N = 20, 50, 100, 200, 400, 800$. As a numerical experiment, we carry out 100,000 and 1,000,000 replications for the case of Mardia's measures and Srivastava's measures, respectively.

From Tables 1-2, expectations of approximate χ^2 statistic MJB_M^* and MJB_S^* are invariant for any sample sizes *N*. That is, MJB_M^* and MJB_S^* are almost close to exact expectations even for small *N*. However, accuracy of expectation of MJB_M and MJB_S is not good especially for small *N*. We note that expectations of MJB_M and MJB_S are convergence in those of exact χ^2-distribution for large *N*. Hence, it may be notice that both MJB_M^* and MJB_S^* are improvements of MJB_M and MJB_S, respectively.

On the other hand, from Tables 1-2, variances of MJB_M^* and MJB_S^* are larger than those of MJB_M and MJB_S. The tendency appears well when sample size N is small. But the coming off values of MJB_M^* and MJB_S^* are more than those of MJB_M and MJB_S. Therefore, there is a tendency for magnitude of variance to become large.

Finally, in Table 3, we give upper percentage points of MJB_M and MJB_M^*, by using Mardia's skewness and kurtosis. MJB_M tends to be conservative. But MJB_M^* is closer to the upper percentage points of χ^2_{f+1}-distribution even when the sample size N is small. In Table 4, we give upper percentage points of MJB_S and MJB_S^* by using Srivastava's skewness and kurtosis. We note that the tendency is similar to the case using Mardia's moments.

5. Concluding Remarks

For univariate case, Jarque-Bera test is well known to mount easily on practical use. In this paper, we have proposed four new test statistics for assessing multivariate normality. We recommend to use MJB_M and MJB_S from the view of practical use in the case of multivariate normality tests. But approximations of expectations, variances and upper percentage points of MJB_M and MJB_S are not good when the sample size N is small. Also, in this paper, we propose improved multivariate normality tests MJB_M^* and MJB_S^*. Hence, we have been improved expectations and upper percentage points of MJB_M and MJB_S. But variances of MJB_M and MJB_S are not improved. This is a future problem. We recommend to use MJB_M^* and MJB_S^* from the aspect of accuracy of approximations to upper percentage points of test statistics especially for small N.

References

[1] C. M. Jarque and A. K. Bera, A test for normality of observations and regression residuals, Int. Statist. Rev. 55 (1987), 163-172.

[2] J. R. Malkovich and A. A. Afifi, On tests for multivariate normality, J. Amer. Statist. Assoc. 68 (1973), 176-179.

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

[4] K. V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhyá B 36 (1974), 115-128.

[5] N. Okamoto and T. Seo, On the Distribution of Multivariate Sample Skewness for Assessing Multivariate Normality, Technical Report No. 08-01, Statistical Research Group, Hiroshima University, (2008).

[6] J. P. Royston, Some techniques for assessing multivariate normality based on the Shapiro-Wilk W, Appl. Statist. 32 (1983), 121-133.

[7] T. Seo and M. Ariga, On the Distribution of Kurtosis Test for Multivariate Normality, Technical Report No. 06-04, Statistical Research Group, Hiroshima University, (2006).

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

[9] M. S. Srivastava, A measure of skewness and kurtosis and a graphical method for assessing multivariate normality, Statist. & Prob. Lett. 2 (1984), 263-267.

[10] M. S. Srivastava and T. K. Hui, On assessing multivariate normality based on Shapiro-Wilk W statistic, Statist. & Prob. Lett. 5 (1987), 15-18.

[11] C. M. Urzúa, On the correct use of omnibus tests for normality, Econom. Lett. 90 (1996), 304-309.

Table 1. Expectations and variances of MJB_M and MJB_M^*

p	N	$E(MJB_M)$	$E(MJB_M^*)$	$f+1$	$Var(MJB_M)$	$Var(MJB_M^*)$	$2(f+1)$
3	20	8.79	10.98	11	14.40	35.78	22
	50	10.00	11.01	11	23.27	36.01	22
	100	10.47	11.01	11	24.95	31.67	22
	200	10.73	11.00	11	24.27	27.55	22
	400	10.83	10.96	11	23.18	24.77	22
	800	10.91	10.98	11	22.51	23.27	22
10	20	189.20	221.02	221	176.65	304.25	442
	50	206.90	220.91	221	427.59	558.73	442
	100	213.73	220.99	221	482.40	562.74	442
	200	217.27	220.96	221	475.28	518.03	442
	400	219.12	220.98	221	468.08	490.30	442
	800	219.93	220.87	221	454.81	465.84	442
20	50	1449.36	1541.18	1541	2447.04	3031.77	3082
	100	1493.80	1541.17	1541	3467.95	3934.50	3082
	200	1516.99	1541.06	1541	3549.40	3818.71	3082
	400	1529.23	1541.37	1541	3406.70	3548.59	3082
	800	1534.60	1540.69	1541	3227.97	3298.83	3082

Table 2. Expectations and variances of MJB_S and MJB_S^*

p	N	$E(\mathrm{MJB_S})$	$E(MJB_S^*)$	$p+1$	$Var(MJB_S)$	$Var(MJB_S^*)$	$2(p+1)$
3	20	2.93	4.02	4	5.46	18.25	8
	50	3.50	4.01	4	9.06	15.67	8
	100	3.73	4.00	4	9.65	13.03	8
	200	3.86	4.00	4	9.24	10.87	8
	400	3.93	4.00	4	8.74	9.52	8
	800	3.96	4.00	4	8.37	8.74	8
10	20	8.66	11.08	11	12.14	36.89	22
	50	9.91	11.00	11	19.28	31.65	22
	100	10.43	11.00	11	21.42	27.98	22
	200	10.71	11.01	11	22.08	25.44	22
	400	10.86	11.01	11	22.18	23.86	22
	800	10.92	10.99	11	22.06	22.90	22
20	50	19.09	21.01	21	34.82	56.24	42
	100	20.01	21.01	21	39.31	50.83	42
	200	20.49	21.01	21	40.99	46.92	42
	400	20.75	21.01	21	41.72	44.73	42
	800	20.87	21.00	21	41.63	43.13	42

Table 3. The upper percentage points of MJB_M and MJB_M^*

p	N	MJB_M	MJB_M^*	$\chi^2_{f+1}(0.05)$
3	20	15.80	22.07	19.68
	50	18.67	21.76	19.68
	100	19.43	21.17	19.68
	200	19.71	20.61	19.68
	400	19.64	20.09	19.68
	800	19.63	19.85	19.68
10	20	212.42	252.35	256.68
	50	243.03	262.64	256.68
	100	252.04	262.59	256.68
	200	254.98	260.40	256.68
	400	256.16	258.97	256.68
	800	256.33	257.74	256.68
20	50	1535.31	1637.63	1633.44
	100	1594.99	1649.32	1633.44
	200	1618.15	1646.28	1633.44
	400	1627.42	1641.72	1633.44
	800	1629.50	1636.56	1633.44

Table 4. The upper percentage points of MJB_S and MJB_S^*

p	N	MJB_S	MJB_S^*	$\chi^2_{p+1}(0.05)$
3	20	6.81	11.24	9.49
	50	8.42	10.58	9.49
	100	8.98	10.16	9.49
	200	9.28	9.90	9.49
	400	9.39	9.71	9.49
	800	9.45	9.60	9.49
10	20	15.03	22.50	19.68
	50	17.86	21.37	19.68
	100	18.87	20.76	19.68
	200	19.34	20.33	19.68
	400	19.54	20.05	19.68
	800	19.60	19.86	19.68
20	50	29.79	34.84	32.67
	100	31.35	34.06	32.67
	200	32.08	33.48	32.67
	400	32.41	33.13	32.67
	800	32.49	32.86	32.67