Journal of Statistical Planning and Inference 140 (2010) 2809–2816

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On the distributions of multivariate sample skewness

Naoya Okamoto [a,*], Takashi Seo [b]

[a] Department of Food Sciences, Faculty of Health and Nutrition, Tokyo Seiei College, 1-4-6, Nishishinkoiwa, Katsushika-ku, Tokyo 124-8530, Japan
[b] Department of Mathematical Information Science, Faculty of Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

ARTICLE INFO

Article history:
Received 31 August 2008
Accepted 18 October 2009
Available online 6 March 2010

MSC:
62E20
62H10
65C05

Keywords:
Asymptotic distribution
Multivariate normal distribution
Approximate χ^2 test statistic

ABSTRACT

In this paper, we consider the multivariate normality test based on measure of multivariate sample skewness defined by Srivastava (1984). Srivastava derived asymptotic expectation up to the order N^{-1} for the multivariate sample skewness and approximate χ^2 test statistic, where N is sample size. Under normality, we derive another expectation and variance for Srivastava's multivariate sample skewness in order to obtain a better test statistic. From this result, improved approximate χ^2 test statistic using the multivariate sample skewness is also given for assessing multivariate normality. Finally, the numerical result by Monte Carlo simulation is shown in order to evaluate accuracy of the obtained expectation, variance and improved approximate χ^2 test statistic. Furthermore, upper and lower percentiles of χ^2 test statistic derived in this paper are compared with those of χ^2 test statistic derived by Mardia (1974) which is used multivariate sample skewness defined by Mardia (1970).

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In multivariate statistical analysis, the test for multivariate normality is an important problem and has been studied by many authors. Shapiro and Wilk (1965) derived test statistic using order statistic. Multivariate extensions of the Shapiro and Wilk (1965) test were proposed by Malkovich and Afifi (1973), Royston (1983), Srivastava and Hui (1987) and so on. Mardia (1970) and Srivastava (1984) gave different definitions of the multivariate sample skewness and kurtosis, and discussed test statistics using these measures for assessing multivariate normality. Mardia (1970) derived expectation of multivariate sample skewness and approximate χ^2 test statistic. Mardia (1974) derived more accurate approximate χ^2 test statistic using accessorial term than that of Mardia (1970). Tests based on multivariate skewness and kurtosis of Mardia, have been computationally implemented by Khattree and Naik (1999). Test statistic using the multivariate sample kurtosis of Srivastava was discussed by Seo and Ariga (2006). Srivastava (1984) derived asymptotic expectation up to the order N^{-1} for multivariate sample skewness and approximate χ^2 test statistic using its asymptotic expectation. Thus, for small N, it seems that multivariate normality test using approximate χ^2 test statistic cannot be carried out correctly.

In this paper, we consider the distribution of multivariate sample measure of skewness defined by Srivastava (1984). Expectation and variance of multivariate sample skewness which are different from Srivastava (1984) are derived. Further, improved approximate χ^2 test statistic is also given by using expectation of multivariate sample skewness derived in this paper. Finally, we investigate accuracy of the distributions of approximate χ^2 test statistics derived by Srivastava (1984) and this paper via a Monte Carlo simulation for selected values of parameters. Furthermore, approximate χ^2 test statistics derived by Srivastava (1984), Mardia (1974) and this paper, are compared with exact χ^2 test statistic.

* Corresponding author.
E-mail address: n_okamoto@auone.jp (N. Okamoto).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.03.003

2. Distributions of multivariate sample skewness

2.1. Mardia's measure of multivariate skewness

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be p-dimensional random vectors distributed identically and independently with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Mardia (1970) has defined the population measure of multivariate skewness as

$$\tilde{\beta}_{1,p} = \mathrm{E}[\{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\}^3].$$

We note that $\tilde{\beta}_{1,p} = 0$ under a multivariate normal population.

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be samples of size N from a multivariate population. Let $\overline{\boldsymbol{x}}$ and S be sample mean vector and sample covariance matrix as follows:

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

respectively. Mardia (1970) has defined the sample measure of multivariate skewness as

$$\tilde{b}_{1,p} = \frac{1}{N^2}\sum_{i,j=1}^{N}\{(\boldsymbol{x}_i-\overline{\boldsymbol{x}})'S^{-1}(\boldsymbol{x}_j-\overline{\boldsymbol{x}})\}^3.$$

In multivariate normal population, Mardia (1974) has given the mean of $\tilde{b}_{1,p}$ and approximate χ^2 statistic as follows:

$$\mathrm{E}[\tilde{b}_{1,p}] = \frac{p(p+2)}{(N+1)(N+3)}\{(N+1)(p+1)-6\},$$

$$\frac{N}{6}\tilde{b}_{1,p}\frac{(p+1)(N+1)(N+3)}{N\{(N+1)(p+1)-6\}} \sim \chi^2_{\frac{p}{6}(p+1)(p+2)}. \tag{1}$$

2.2. Srivastava's measure of multivariate skewness

Let $\boldsymbol{x}$ be a p-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma D_\lambda \Gamma'$, where $\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_p)$ is an orthogonal matrix and $D_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. Note that $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of Σ. Then Srivastava (1984) defined the population measure of multivariate skewness as

$$\beta^2_{1,p} = \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{\mathrm{E}[(y_i-\theta_i)^3]}{\lambda_i^{3/2}}\right\}^2,$$

where $y_i = \boldsymbol{\gamma}_i'\boldsymbol{x}$ and $\theta_i = \boldsymbol{\gamma}_i'\boldsymbol{\mu}$ $(i = 1, 2, \ldots, p)$. We note that $\beta^2_{1,p} = 0$ under a multivariate normal population.

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ be samples of size N from a multivariate population. Let $\overline{\boldsymbol{x}}$ and $S = HD_\omega H'$ be sample mean vector and sample covariance matrix as follows:

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

respectively, where $H = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$ is an orthogonal matrix and $D_\omega = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_p)$. We note that $\omega_1, \omega_2, \ldots, \omega_p$ are the eigenvalues of S. Then Srivastava (1984) defined the sample measure of multivariate skewness as

$$b^2_{1,p} = \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{1}{\omega_i^{3/2}}\sum_{j=1}^{N}\frac{(y_{ij}-\overline{y}_i)^3}{N}\right\}^2, \tag{2}$$

where $y_{ij} = \boldsymbol{h}_i'\boldsymbol{x}_j$ $(i=1,2,\ldots,p,\ j=1,2,\ldots,N)$, $\overline{y}_i = N^{-1}\sum_{j=1}^{N} y_{ij}$ $(i=1,2,\ldots,p)$.

For large N, Srivastava (1984) derived asymptotic expectation and approximate χ^2 test statistic for assessing multivariate normality as follows:

$$\mathrm{E}[b^2_{1,p}] = \frac{6}{N}, \tag{3}$$

$$\frac{Np}{6}b^2_{1,p} \sim \chi^2_p, \tag{4}$$

respectively.

Download English Version:

https://daneshyari.com/en/article/1147805

Download Persian Version:

https://daneshyari.com/article/1147805

Daneshyari.com