Ann. Inst. Statist. Math.
34 (1982), Part A, 531-541

ON A MEASURE OF MULTIVARIATE SKEWNESS AND A TEST FOR MULTIVARIATE NORMALITY

TAKAFUMI ISOGAI

(Received Nov. 25, 1980; revised Dec. 15, 1981)

Summary

We consider an extension of Pearson measure of skewness to a multivariate case and apply the proposed measure to a test of multivariate normality.

1. Introduction

For a test of multivariate normality several authors have introduced the measures of multivariate skewness, which are constructed by the 3rd order central moments or equivalently the 3rd order cumulants of the population (Malkovich and Afifi [4], Mardia [5]). In this paper we shall define a measure of multivariate skewness of a different kind by extension of Pearson's measure of skewness, say

$$\text{Skewness}=(\text{mode}-\text{mean})/\sigma$$

where σ is the standard deviation of the distribution under consideration (see Kendall and Stuart [3]).

When for a given sample we attempt to construct a measure of multivariate skewness, the inconvenience of determining a multivariate mode arises. The estimation of the sample mode is carried out by the use of a density estimator with respect to an appropriate kernel function.

In Sections 2 and 3 the distributional properties of the sample mode are investigated. In Section 4 asymptotic distributions of the proposed measure of multivariate skewness are derived under the null and non-null cases. In Section 5 we give simple examples for a test of normality.

Key words and phrases: density estimator, multivariate mode, multivariate skewness, test for multivariate normality.

2. Estimation and consistency of the multivariate mode

On estimation of the mode for a univariate case Parzen [6] has considered in detail. In order to obtain an estimator of the multivariate mode we have only to extend his procedure to a multivariate case.

Following Cacoullos [1], we define an estimator $f_n(x)$ of a p-variate density function $f(x)$ on the basis of a random sample $X_1, \cdots, X_n$ from $f(x)$. We consider $f_n(x)$ as the following form:

$$f_n(x) = \int \frac{1}{h^p(n)} K\left(\frac{x-y}{h(n)}\right) dF_n(y) = \frac{1}{nh^p(n)} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h(n)}\right) \tag{2.1}$$

where $F_n(y)$ is the empirical distribution function on a sample $X_1, \cdots, X_n$ and a sequence of positive numbers $\{h(n)\}$ satisfies

$$\lim_{n\to\infty} h(n) = 0 \; . \tag{2.2}$$

Furthermore, a weighting function $K(y)$ is a nonnegative Borel scalar function on R^p which is chosen to satisfy the following conditions, i.e.

$$\sup_y K(y) < \infty \; , \tag{2.3}$$

$$\int K(y) dy = 1 \; , \tag{2.4}$$

$$\lim_{|y|\to\infty} |y|^p K(y) = 0 \; , \tag{2.5}$$

$$K(y) = K(-y) \qquad \text{for all } y \tag{2.6}$$

where the notation $|y|$ denotes the norm of a vector y. We also assume that the Fourier transform of $K(y)$, i.e.

$$k(u) = \int e^{-iu'y} K(y) dy \tag{2.7}$$

is absolutely integrable. This fact indicates that we may think $K(y)$ to be uniformly continuous.

We can also express an estimator of the form of (2.1) as weighted averages over the sample characteristic function

$$\phi_n(u) = \int e^{iu'y} dF_n(y) = n^{-1} \sum_{j=1}^{n} e^{iu'X_j} \; . \tag{2.8}$$

As is easily checked, we may write

$$f_n(x) = \frac{1}{nh^p} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right) = \frac{1}{(2\pi)^p} \int e^{-iu'x} k(hu) \phi_n(u) du \; . \tag{2.9}$$

We can see that $f_n(x)$ is continuous and tends to 0 as $|x|\to\infty$. Thus there is a random variable θ_n such that

$$f_n(\theta_n)=\max_x f_n(x) . \tag{2.10}$$

We call θ_n a sample multivariate mode.

We suppose that the true probability density function $f(x)$ is uniformly continuous in x. It follows that $f(x)$ has a multivariate mode defined by

$$f(\theta)=\max_x f(x) . \tag{2.11}$$

We assume that θ is unique.

Under the above conditions the following Theorem 2.1 holds.

THEOREM 2.1. *If a sequence $h=h(n)$ satisfies*

$$\lim_{n\to\infty} nh^{2p}=\infty , \tag{2.12}$$

then for all $\varepsilon>0$

$$\mathrm{P}\,[\sup|f_n(x)-f(x)|<\varepsilon]\to 1 \qquad as\ n\to\infty . \tag{2.13}$$

If $\{\theta_n\}$ are sample multivariate modes, then for every $\varepsilon>0$

$$\mathrm{P}\,[|\theta_n-\theta|<\varepsilon]\to 1 \qquad as\ n\to\infty . \tag{2.14}$$

Remark. This theorem can be proved directly by the analogy of Theorem 3a in Parzen [6], so we shall omit the proof.

3. Asymptotic normality of the sample multivariate mode

In this section we shall investigate the conditions on a sequence $\{h(n)\}$ and a weighting function $K(y)$ such that the sample multivariate mode is asymptotically normal.

We assume that a density $f(x)$ has continuous second partial derivatives. We shall write the gradient vector of $f(x)$ as $\mathrm{grad}\,(f)=(\partial f/\partial x_1,\cdots,\partial f/\partial x_p)'$ and similarly the Hessian matrix of $f(x)$ as $H(f)=(\partial^2 f/\partial x_i\partial x_j)$ $(i, j=1,\cdots, p)$. Then $\mathrm{grad}\,(f)_\theta=0$ and $-H(f)_\theta$ is positive definite, where suffices denote values of these quantities at a unique mode $x=\theta$. Similarly, if the estimated density function is chosen to have continuous second partial derivatives (i.e. the corresponding weighting function has the same property), then $\mathrm{grad}\,(f_n)_{\theta_n}=0$ and $-H(f_n)_{\theta_n}$ is positive definite at the sample mode θ_n.

By use of Taylor's expansion we have

$$\mathrm{grad}\,(f_n)_{\theta_n}=\mathrm{grad}\,(f_n)_\theta+H(f_n)_{\hat\theta_n}(\theta_n-\theta)=0 \tag{3.1}$$

where some random variable $\hat{\theta}_n$ lies between θ_n and θ. If the matrix $H(f_n)$ dose not vanish at $\hat{\theta}_n$, we can write

$$\theta_n - \theta = -H^{-1}(f_n)_{\hat{\theta}_n} \operatorname{grad}(f_n)_\theta \tag{3.2}$$

where H^{-1} denotes the inverse of the Hessian matrix.

THEOREM 3.1. [*Consistency of the Hessian matrix* $H(f_n)$.] *If the Fourier transform* $k(u)$ *satisfies for all* i, j $(=1, \cdots, p)$

$$\int |u_i u_j k(u)| du < \infty \tag{3.3}$$

and a sequence $\{h(n)\}$ *satisfies*

$$\lim_{n\to\infty} nh^{2p+4} = \infty \tag{3.4}$$

and that the characteristic function $\phi(u)$ *satisfies*

$$\int |u_i u_j \phi(u)| du < \infty \qquad \text{for all } i, j = 1, \cdots, p\,, \tag{3.5}$$

then as n *becomes large, for every pair of* i *and* j $(=1, \cdots, p)$

$$\mathrm{E}\left[\sup\left|\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x) - \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right|\right] \to 0 \tag{3.6}$$

and

$$H(f_n)_{\hat{\theta}_n} \to H(f)_\theta \qquad \text{in probability}\,. \tag{3.7}$$

PROOF. Note first that

$$\begin{aligned} &\left|\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x) - \mathrm{E}\left[\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x)\right]\right| \\ &\quad \leqq (2\pi)^{-p} \int |u_i u_j k(hu)| |\phi_n(u) - \mathrm{E}[\phi_n(u)]| du\,. \end{aligned} \tag{3.8}$$

Thus

$$\begin{aligned} &\mathrm{E}\left[\sup\left|\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x) - \mathrm{E}\left[\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x)\right]\right|\right] \\ &\quad \leqq (2\pi)^{-p} \int |u_i u_j k(hu)| \sigma[\phi_n(u)] du \\ &\quad \leqq (2\pi)^{-p} (nh^{2p+4})^{-1/2} \int |u_i u_j k(u)| du \to 0 \qquad \text{as } n \to \infty \end{aligned} \tag{3.9}$$

where the first inequality holds from Schwarz's inequality. Similarly it can be easily shown that

$$\sup \left| \mathrm{E}\left[\frac{\partial^2 f_n}{\partial x_i \partial x_j}(x)\right] - \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right| \to 0 \qquad \text{as } n \to \infty . \tag{3.10}$$

From (3.9) and (3.10), (3.6) holds.

That (3.7) holds follows from (3.6) and the fact that $\hat{\theta}_n$ tends to θ, since it is between θ_n and θ and θ_n tends to θ.

THEOREM 3.2. [*Asymptotic normality of the sample multivariate mode.*] *If in addition to all the conditions* (3.3) *to* (3.5) $(\partial K/\partial y_i)(y)$ $(i=1,\cdots,p)$ *satisfies the following conditions*

$$\sup \left|\frac{\partial K}{\partial y_i}\right| < \infty , \quad \int \left|\frac{\partial K}{\partial y_i}\right| dy < \infty \quad \text{and} \quad \lim_{|y|\to\infty} |y|^p \left|\frac{\partial K}{\partial y_i}\right| = 0 , \tag{3.11}$$

then

$$\sqrt{nh^{p+2}}\, \mathrm{grad}\, (f_n)_\theta \sim N_p(0, f(\theta)J) \tag{3.12}$$

and

$$\sqrt{nh^{p+2}}\, (\theta_n - \theta) \sim N_p(0, f(\theta)H^{-1}(f)_\theta J H^{-1}(f)_\theta) \tag{3.13}$$

where the $p\times p$ *matrix* J *is defined by*

$$J = \begin{pmatrix} \int \left(\frac{\partial K}{\partial y_1}\right)^2 dy, \cdots, \int \frac{\partial K}{\partial y_1}\frac{\partial K}{\partial y_p} dy \\ \vdots \qquad\qquad \vdots \\ \int \frac{\partial K}{\partial y_p}\frac{\partial K}{\partial y_1} dy, \cdots, \int \left(\frac{\partial K}{\partial y_p}\right)^2 dy \end{pmatrix} . \tag{3.14}$$

PROOF. From (2.9) we can put for each i $(=1,\cdots,p)$

$$\frac{\partial f_n}{\partial x_i}(x) = n^{-1}\sum_{s=1}^n V_{i,n}^{(s)} \tag{3.15}$$

where we define

$$V_{i,n}^{(s)} = \frac{\partial}{\partial x_i}[h^{-p} K\{(x-X_s)/h\}] = h^{-(p+1)}\frac{\partial K}{\partial x_i}\{(x-X_s)/h\} \tag{3.16}$$

and $V_{i,n}^{(s)}$ $(s=1,\cdots,n)$ are independently and identically distributed as $V_{i,n} = h^{-(p+1)}(\partial K/\partial x_i)\{(x-X)/h\}$. We remark that for i $(=1,\cdots,p)$

$$\mathrm{E}\,[V_{i,n}] = -i(2\pi)^{-p}\int e^{-iu'x} k(hu) u_i \phi(u) du \tag{3.17}$$

and for i and j $(=1,\cdots,p)$

$$\mathrm{E}\,[V_{i,n} V_{j,n}] = h^{-(p+2)}\int \frac{\partial K}{\partial y_i}(y)\frac{\partial K}{\partial y_j}(y) f(x-hy) dy . \tag{3.18}$$

From (3.11), (3.17) and (3.18) it follows that for i and j $(=1,\cdots,p)$

$$\mathrm{E}\left[\frac{\partial f_n}{\partial x_i}(x)\right]\to\frac{\partial f}{\partial x_i}(x)\ ,$$

$$nh^{p+2}\operatorname{cov}\left(\frac{\partial f_n}{\partial x_i},\frac{\partial f_n}{\partial x_j}\right)\to f(x)\int\frac{\partial K}{\partial y_i}(y)\frac{\partial K}{\partial y_j}(y)dy$$

as $n\to\infty$ where $\operatorname{cov}(\cdot,\cdot)$ denotes the covariances of $\partial f_n/\partial x_i$'s.

To prove the asymptotic normality of the random vector $\operatorname{grad}(f_n)$ it suffices to show that for any real vector $a=(a_1,\cdots,a_p)'$ each linear combination $a'\operatorname{grad}(f_n)=\sum a_i(\partial f_n/\partial x_i)$ is asymptotically normal. Then to check the sufficient conditions for the asymptotic normality of $a'\operatorname{grad}(f_n)$ it is enough to show that

$$n^{-1/2}\,\mathrm{E}\left|h^{(p+2)/2}\sum_{i=1}^{p}a_iV_{i,n}\right|^3\to 0\qquad\text{as } n\to\infty\ . \tag{3.19}$$

For (3.19) it is sufficient to prove that

$$n^{-1/2}h^{3(p+2)/2}\,\mathrm{E}\,|V_{i,n}V_{j,n}V_{k,n}|\to 0\qquad\text{for all } i,\ j \text{ and } k$$

as n becomes large. This quantity is approximately equivalent to

$$(nh^p)^{-1/2}f(x)\int\left|\frac{\partial K}{\partial y_i}(y)\frac{\partial K}{\partial y_j}(y)\frac{\partial K}{\partial y_k}(y)\right|dy$$

and converges to 0 by the use of (3.4) and (3.11) as n is large.

For the proof of (3.13) we have only to recall the relation (3.2).

4. Measure of multivariate skewness

In this section we shall generalize Pearson's measure of skewness to a multivariate case. First we define a measure of multivariate skewness (which we shall call Skew in brief) by

$$\text{Skew}=(\theta-\mu)'\omega^{-1}(\Sigma)(\theta-\mu) \tag{4.1}$$

where θ is the population mode, μ is the mean vector and $\omega(\Sigma)$ is an appropriate function of the covariance matrix $\Sigma=(\sigma_{ij})$. As for choices of $\omega(\Sigma)$ we can take, for example, (1) $\omega_1(\Sigma)=\Sigma$, (2) $\omega_2(\Sigma)=(2\pi)^{p/2}|\Sigma|^{1/2}\cdot\Sigma J\Sigma$ (which is derived under the null condition that the population is $N_p(\mu,\Sigma)$), or (3) $\omega_3(\Sigma)=\operatorname{diag}(\sigma_{11},\cdots,\sigma_{pp})$.

For a given random sample $X_1,\cdots,X_n$ we shall take

$$(\theta-\bar{X})'\omega^{-1}(\hat{\Sigma})(\theta_n-\bar{X})\qquad(=\hat{\text{S}}\text{kew, say}) \tag{4.2}$$

as an estimator of Skew where θ_n is the sample mode defined in Section 2, $\bar{X}$ is the sample mean vector and $\hat{\Sigma}$ is the sample covariance matrix.

Now we consider the asymptotic distribution of (4.2). We notice that the asymptotic distribution of

$$nh^{p+2}(\theta_n-\bar{X})'\omega^{-1}(\hat{\Sigma})(\theta_n-\bar{X}) \tag{4.3}$$

is the same as that of

$$nh^{p+2}(\theta_n-\mu)'\omega^{-1}(\Sigma)(\theta_n-\mu) , \tag{4.4}$$

because both $\bar{X}$ and $\hat{\Sigma}$ have higher order of consistency than θ_n. Then the following theorem holds.

THEOREM 4.1. [*Asymptotic null distribution of* $\hat{S}kew$.] *In addition to conditions* (3.3) *to* (3.5) *and* (3.12) *to* (3.14), *if the true population is a p-variate normal* $N_p(\mu, \Sigma)$, *then* $nh^{p+2}\ \hat{S}kew$ *is distributed as*

$$\sum_{i=1}^{p} \lambda_i \chi_i^2(1) \tag{4.5}$$

where λ_i $(i=1,\cdots,p)$ *are the latent roots of the following determinantal equation*

$$|(2\pi)^{p/2}|\Sigma|^{1/2}\Sigma J\Sigma-\lambda\omega(\Sigma)|=0 \tag{4.6}$$

and $\chi_i^2(1)$ $(i=1,\cdots,p)$ *are independent* χ^2 *variables with one degree of freedom.*

PROOF. Since the population is $N_p(\mu, \Sigma)$, the population mode θ coincides with its mean μ and hence the value at $x=\mu$ of the Hessian $H(f)$ is

$$H(f)_\mu=-f(\mu)\Sigma^{-1}=-(2\pi)^{-p/2}|\Sigma|^{-1/2}\Sigma^{-1} . \tag{4.7}$$

From (3.14) of Theorem 3.2 the asymptotic distribution of (4.4) is the same as (4.5) and so the rest of the proof is straightforward.

We shall define some notations to deal with the non-null case. We put $\Sigma_\theta=f(\theta)H^{-1}(f)_\theta JH^{-1}(f)_\theta$. Let λ_i $(i=1,\cdots,p)$ be the latent roots of the determinantal equation

$$|\Sigma_\theta-\lambda\omega(\Sigma)|=0 \tag{4.8}$$

and also let P be some $p\times p$ matrix constructed by the corresponding latent vectors, which satisfies

$$P\Sigma_\theta^{1/2}\omega^{-1}(\Sigma)\Sigma_\theta^{1/2}P'=\mathrm{diag}\,(\lambda_1,\cdots,\lambda_p) . \tag{4.9}$$

Then, using the same argument as Theorem 4.1 with above notations, the following theorem holds.

THEOREM 4.2. [*Asymptotic non-null distribution of $\hat{Skew}$.*] *Under the conditions* (3.3) *to* (3.5) *and* (3.12) *to* (3.14), $nh^{p+2}\ \hat{Skew}$ *is distributed as*

$$\sum_{i=1}^{p} \lambda_i \chi^2_{c_i}(1) \tag{4.10}$$

where λ's are the latent roots of (4.8) *and* $\chi^2_{c_i}(1)$ $(i=1,\cdots,p)$ *are independent non-central* χ^2 *variables with one degree of freedom whose noncentrality parameters are given by*

$$c_i=\sqrt{nh^{p+2}}\,\psi_i \qquad \text{for } i=1,\cdots,p\,, \tag{4.11}$$

where ψ_i is the ith element of the vector Ψ defined by

$$\Psi=P\Sigma_\theta^{-1/2}(\mu-\theta)\ . \tag{4.12}$$

Remark. When we transform original observations X_i $(i=1,\cdots,n)$ to $Z_i=AX_i+b$ with an arbitrary nonsingular $p\times p$ matrix A and an arbitrary constant vector b, $\hat{\text{S}}$kew is transformed to

$$(\theta_n-\bar{X})'A'\omega^{-1}(A\hat{\Sigma}A')A(\theta_n-\bar{X})\ .$$

It can be easily shown that $\hat{\text{S}}$kew is not invariant under the affine transformation. Nevertheless, if we demand invariant property of $\hat{\text{S}}$kew in any limited situation, we have to restrict both a class of weighting functions $K(y)$ and that of transformation matrices A. Such a choice as satisfies our requirements is to choose a class of $K(y)$ which leads to $J=cI_p$ (where c is some constant number and I_p is the identity matrix, for instance, a class of $K(y)$ is composed of $N_p(0, \sigma^2 I_p)$ $(0<\sigma^2<\infty)$) and at the same time to restrict the affine transformation to the orthogonal transformation. In this limited situation, $\hat{\text{S}}$kew with $\omega_1(\hat{\Sigma})$ and $\omega_2(\hat{\Sigma})$ proves to be invariant.

5. Test for multivariate normality

First we note that from the remark of Section 4, $\hat{\text{S}}$kew is not invariant under the scale transformation $X'=(X_1,\cdots,X_p)\to Y'=(a_1X_1,\cdots,a_pX_p)$ $(a_1,\cdots,a_p$ are nonzero scalars) and so we consider a standardized random vector X with $\sigma[X_i]=1$ $(i=1,\cdots,p)$.

With an optimum choice of a number h for a fixed n, we evaluate

$$\mathrm{E}\,[\sup_x |f_n(x)-f(x)|] \tag{5.1}$$

which gives us the sufficient condition for the consistency of θ_n as well as the uniform consistency of $f_n(x)$ in Theorem 2.1.

Using Theorem 3a in Parzen [6] and Theorem 3.4 in Cacoullos [1], we have the following result that

$$(5.2)\qquad \mathrm{E}\,[\sup|f_n(x)-f(x)|] \leqq (2\pi)^{-p}(nh^{2p})^{-1/2}\int |k(u)|du+(I/2)h^2 \qquad (=L(h),\ \text{say})$$

where $I=\sup_x\left|\sum_{i=1}^p\sum_{j=1}^p \frac{\partial^2 f(x)}{\partial x_i\partial x_j}\int y_iy_jK(y)dy\right|$.

The value h which minimizes $L(h)$ is given by

$$(5.3)\qquad h_{\min}=\left\{p(2\pi)^{-p}\int |k(u)|du/I\right\}^{1/(p+2)}n^{-1/(2p+4)}.$$

If we choose h as above, then (5.1) tends to 0 as $n^{-1/(p+2)}$.

On the other hand, from the condition (3.4), it follows that if we put $h=O(n^{-\alpha})$, then $(2p+4)^{-1}>\alpha>0$. This range of α does not include the optimum $\alpha^*=(2p+4)^{-1}$ corresponding to $h_{\min}$ in (5.3). It suggests choosing α just smaller than α^*. However, from a practical point of view, we put $h=n^{-1/(2p+4)}$ at first and examine the behavior of $\widehat{\mathrm{S}}$kew.

We shall consider the physiological data given by Tabe [7], which comprise 8 measurements on each of 107 patients of liver disease. We take 2 measurements in the order: GOT, GPT. These variables are related to the function of a liver. We apply $\widehat{\mathrm{S}}$kew with $\omega_2(\hat{\Sigma})$ to the data for a test of multivariate normality with respect to multivariate skewness Skew.

To test the null hypothesis that Skew$=0$, we can use the result from Theorem 4.1, that

$$(5.4)\qquad A=nh^{p+2}\widehat{\mathrm{S}}\mathrm{kew}\qquad \text{with } \omega_2(\hat{\Sigma})$$

has a χ^2 distribution with p degrees of freedom. As for a choice of $K(y)$ we take $K(y)=N_p(0, I_p)$.

Table 1. Physiological data on patients. First four moments. Upper values, the original. Lower values, the transformed.

	Mean	St. dev.	g_1	g_2
GOT	61.850	51.173	2.3460	7.7952
	3.8583	0.7282	0.1120	−.2492
GPT	72.626	69.383	1.7942	2.9318
	3.9038	0.8747	0.1983	−.7152

The upper estimates in Table 1 give the first four moments of the marginal distributions. The estimate of the correlation coefficient was

0.860. The large values of g_1 and g_2 suggest that the data have a strong deviation from normality. The vector of the estimated mode was (34.00, 37.50) and hence $\hat{\text{Skew}}=1.5414$. From (5.4) we have $A=15.9$ which is highly significant, because the 0.5% value of χ_2^2 is 10.60.

The positive values of g_1 led us to apply a log transformation to the data. The marginal moments for the transformed data with log (GOT) and log (GPT) are recorded in the lower values of Table 1. The estimate of the correlation coefficient was 0.881. These values indicate the improvement of the data from non-normality. The estimated mode was (3.6771, 3.6798) and hence $\hat{\text{Skew}}=.3197$ which gives $A=3.29$. The 10% value of χ_2^2 is 4.61. Therefore $\hat{\text{Skew}}$ is not significant.

As a final example we shall consider the well known iris data given by Fisher [2] and examine the data on Iris virginica, which comprise 4 measurements on each of 50 plants. We take 4 measurements in the order: sepal length, sepal width, petal length, petal width.

Table 2. Data on Iris virginica. First four moments.

	Mean	St. dev.	g_1	g_2
Sepal length	6.588	0.6359	0.1144	−.0879
Sepal width	2.974	0.3225	0.3549	0.5198
Petal length	5.552	0.5519	0.5328	−.2565
Petal width	2.026	0.2747	−.1256	−.6613

Table 2 gives the estimated marginal moments. The values of g_1 and g_2 are not significant. The estimated mode was (6.4798, 2.9770, 5.4305, 2.0362) and hence $\hat{\text{Skew}}=2.512$, which gives $A=17.70$. This is highly significant, because the 0.5% value of χ_4^2 is 14.86. Note that the estimated modes of the data on Iris setosa and Iris versicolor were (4.9971, 3.4144, 1.4710, .2447) and (5.9569, 2.8040, 4.3135, 1.3388) respectively which give $\hat{\text{Skew}}=.0883$ and $A=.624$ for Iris setosa and $\hat{\text{Skew}}=4.291$ and $A=30.34$ for Iris versicolor. The last value of A is also highly significant.

These results suggest that there is a strong evidence of multivariate skewness on the data with Iris virginica and Iris versicolor respectively, though the analysis of the marginal moments did not detect non-normality on the data.

Acknowledgements

I wish to express my thanks to the referee for his useful comments and to Professor M. Okamoto for his valuable advices and comments.

OSAKA CITY UNIVERSITY

REFERENCES

[1] Cacoullos, T. (1966). Estimation of a multivariate density, *Ann. Inst. Statist. Math.*, **18**, 179-189.
[2] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, **VII**, 179-188.
[3] Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics*, vol. 1, 4th edition, Griffin, London.
[4] Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality, *J. Amer. Statist. Ass.*, **68**, 176-179.
[5] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519-530.
[6] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.
[7] Tabe, K. (1967). Clinical and experimental studies on Rheumatoid Like Factor in Chronic Liver Disease, *J. Osaka City Med. C.*, **16**, 157-176 (in Japanese).