Ann. Inst. Statist. Math.
Vol. 41, No. 1, 169-186 (1989)

# ON USING INFLUENCE FUNCTIONS FOR TESTING MULTIVARIATE NORMALITY

TAKAFUMI ISOGAI

*College of General Education, Osaka University, Toyonaka, Osaka 560, Japan*

(Received August 22, 1987; revised August 11, 1988)

**Abstract.** Changes in the joint distribution of influence functions for the mean vector and the covariance matrix are examined when the true probability distribution is contaminated. In particular, the formulas for influence functions of the first and second moments with respect to the above joint distribution are obtained and used to derive reasonable test statistics for multivariate normality. The formulas are extended by using the joint distribution of score functions for population parameters. An application of the extended formulas to the usual linear regression analysis leads to a measure of multivariate skewness which can be used to reduce the effect of non-normality of the response variable. Also, some relationship between the extended formulas and goodness-of-fit statistics is discussed and used to derive test statistics for multivariate normality.

*Key words and phrases*: Influence function, multivariate normality, measure of dependence, measures of multivariate skewness and kurtosis, score function, linear regression.

## 1. Introduction

Many of the standard multivariate statistical methods depend on the assumption of multivariate normality. Thus, in analyzing multivariate data, we often face the problem of detecting influential observations which may affect the estimation of the mean vector and the covariance matrix. Let $F$ be an underlying $p$-variate distribution function with mean vector $\mu' = (\mu_1,\dots,\mu_p)$ and covariance matrix $\Sigma = (\sigma_{ij})$, $i, j = 1,\dots,p$. An influential observation $X' = (X_1,\dots,X_p)$, which may have a possible effect on the estimation of $\mu$ and $\Sigma$, is detected by the influence functions for $\mu$ and $\Sigma$, that is,

$$\text{(1.1)} \qquad \mathbf{IF}(X;\mu) = \lim_{\varepsilon\to 0} (\mu(\tilde{F}) - \mu(F))/\varepsilon = X - \mu ,$$

and

$$\text{(1.2)} \qquad \mathbf{IF}(X;\Sigma) = \lim_{\varepsilon\to 0}(\Sigma(\tilde{F}) - \Sigma(F))/\varepsilon = (X-\mu)(X-\mu)' - \Sigma\,,$$

respectively, where $\tilde{F} = (1-\varepsilon)F + \varepsilon\delta_X$ $(0 \le \varepsilon \le 1)$ is a mixture distribution function constructed in terms of $F$ and a discrete distribution with a unit mass at $X$ (see Radhakrishnan and Kshirsagar (1981)).

If we take $F$ to be multivariate normal, then a well-known diagnostic tool for checking the distribution of $\mathbf{IF}(X;\mu)$ (or some function of $\mathbf{IF}(X;\Sigma)$) is Healy's (1968) $\chi^2$ probability plot based on $(X-\mu)'\Sigma^{-1}(X-\mu)$. This method is closely related to Wilks' (1963) outlier detection procedure, which examines the distribution of the influence function about the determinant of $\Sigma$. The influence function is defined by

$$\text{(1.3)} \qquad \mathbf{IF}(X;|\Sigma|) = [(X-\mu)'\Sigma^{-1}(X-\mu) - p]|\Sigma|$$

(again see Radhakrishnan and Kshirsagar (1981)). Wilks (1963) examined the ratio statistic $\mathbf{IF}(X;|\Sigma|)/|\Sigma|$, that is,

$$\text{(1.4)} \qquad R(X;|\Sigma|) = (X-\mu)'\Sigma^{-1}(X-\mu) - p\,.$$

The second moment of $R(X;|\Sigma|)$ was used by Mardia (1970) to derive his measure of multivariate kurtosis $\beta_{2,p}$ for testing multivariate normality.

On the other hand, a necessary and sufficient condition for multivariate normality is independence of the sample mean vector and the sample covariance matrix (see Section 15.24 in Stuart and Ord (1987)). This means that $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$ are independent, whenever $F$ is multivariate normal. Therefore, in order to investigate possible departures from multivariate normality for influential observations, it is natural to examine the structure of the joint distribution of $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$ when the distribution of an influential observation $X$, say $H$, is different from $F$.

In the following section, we first evaluate the first and second moments of the joint distribution of $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$ under the contaminated distribution $H = (1-\eta)F + \eta G$ $(0 \le \eta \le 1)$, where $G$ is an unknown $p$-variate distribution. Then, we derive influence functions for the first and second moments in the joint distribution of $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$. By using these influence functions, we give some interpretations of various non-null configurations of the $\chi^2$ probability plot, which are due to outliers, heteroscedasticity of variance, or non-normality. This consideration of the $\chi^2$ probability plot for non-normality gives us an idea about the basic relations of $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$, which enables us to introduce two types of measures of dependence for checking multivariate normality. A simulation study in 2-dimensional non-normal models shows that the two measures have good powers in comparison to other related test statistics.

In Section 3, the above basic relations are extended by using score functions for population parameters. An application of the extended relations to the usual linear regression analysis leads to a measure of multivariate skewness $\operatorname{tr}(S_2)$ (see Isogai (1983*a*)) which enables us to choose an experimental design so as to reduce the effect of non-normality of the response variable. Finally, some relationship between the extended relations and goodness-of-fit statistics is discussed and used to derive typical test statistics for multivariate normality.

## 2. Use of influence functions $\mathrm{IF}(X;\mu)$ and $\mathrm{IF}(X;\Sigma)$

Let us first examine the first two moments of $\mathbf{IF}(X;\mu)$ and $\mathbf{IF}(X;\Sigma)$. Here we assume that the distribution $H$ of an influential observation $X$ is expressed as $H=(1-\eta)F+\eta G$ with $0\le\eta\le 1$ and some $p$-variate non-normal or normal distribution $G$ with mean vector $\nu'=(\nu_1,\dots,\nu_p)$ and covariance matrix $\Psi=(\psi_{ij})$, $i,j=1,\dots,p$. We hereafter assume in this section that the distribution $F$ is $p$-variate normal. Put

$$(2.1)\qquad u_1=\mathbf{IF}(X;\mu)=X-\mu\,,$$

and

$$(2.2)\qquad u_2=\operatorname{vec}(\mathbf{IF}(X;\Sigma))=\operatorname{vec}\{(X-\mu)(X-\mu)'-\Sigma\}\,;$$

here, the vec operator on a given $p\times q$ matrix $A$ constructs a $pq\times 1$ column vector by stacking the $q$ column vectors of $A$ consecutively. In other words, if the matrix $A$ is given by $A=(a_1\colon a_2\colon\cdots\colon a_q)$, where each $a_i$ is a $p\times 1$ column vector, then $\operatorname{vec}(A)$ is defined by

$$\operatorname{vec}(A)'=(a_1',a_2',\dots,a_q')\,.$$

Also put

$$(2.3)\qquad w=\begin{pmatrix}u_1\\ u_2\end{pmatrix}.$$

The expectation of the $(p+p^2)$-dimensional random vector $w$ with respect to the distribution $H$ is

$$(2.4)\qquad E_H(w)=(1-\eta)E_F(w)+\eta E_G(w)$$
$$=\eta E_G(w)=\eta\begin{pmatrix}E_G(u_1)\\ E_G(u_2)\end{pmatrix}=\eta\begin{pmatrix}\nu-\mu\\ \operatorname{vec}\{\Psi-\Sigma+(\nu-\mu)(\nu-\mu)'\}\end{pmatrix},$$

(because $E_F(w)=0$), where by subscripts $F$, $G$ and $H$ we denote distributions under which their moments are calculated. Also the covariance matrix of $w$ under $H$ may be expressed as

$$(2.5)\qquad \begin{aligned} \mathrm{var}_H(w) &= E_H\{(w - E_H(w))(w - E_H(w))'\} \\ &= \mathrm{var}_F(w) + \eta[\,\mathrm{var}_G(w) - \mathrm{var}_F(w)] \\ &\quad + \eta(1-\eta)[E_G(w) - E_F(w)][E_G(w) - E_F(w)]' . \end{aligned}$$

We need the following lemma to write the formula of $\mathrm{var}_H(w)$ explicitly.

LEMMA 2.1. *For matrices $A$ of order $q \times r$, $B$ of order $r \times s$ and $C$ of order $s \times t$, we have*

$$(2.6)\qquad \mathrm{vec}\,[ABC] = (C' \otimes A)\,\mathrm{vec}\,[B] ,$$

*where the symbol $\otimes$ denotes the Kronecker product of matrices (for details, see Rao* (1973), *Section* 1b.8).

First we give the formula for $\mathrm{var}_G(w)$. Partition

$$(2.7)\qquad \mathrm{var}_G(w) = \begin{pmatrix} \mathrm{var}_G(u_1) & \mathrm{cov}_G(u_1, u_2) \\ \mathrm{cov}_G(u_2, u_1) & \mathrm{var}_G(u_2) \end{pmatrix},$$

where

$$(2.8)\qquad \mathrm{var}_G(u_1) = \Psi ,$$

$$(2.9)\qquad \begin{aligned} \mathrm{cov}_G(u_1, u_2) &= \mathrm{cov}_G(u_2, u_1)' \\ &= \Delta + \Psi \otimes (\nu - \mu)' + (\nu - \mu)' \otimes \Psi , \end{aligned}$$

and

$$(2.10)\qquad \begin{aligned} \mathrm{var}_G(u_2) = \Gamma &+ \Delta \otimes (\nu - \mu) + (\nu - \mu) \otimes \Delta + \Delta' \otimes (\nu - \mu)' \\ &+ (\nu - \mu)' \otimes \Delta' + \Psi \otimes (\nu - \mu)(\nu - \mu)' \\ &+ (\nu - \mu)(\nu - \mu)' \otimes \Psi + (\nu - \mu) \otimes \Psi \otimes (\nu - \mu)' \\ &+ (\nu - \mu)' \otimes \Psi \otimes (\nu - \mu) ; \end{aligned}$$

we now describe the matrices $\Delta$ and $\Gamma$. In particular, we put $E_G[(X - \nu) \cdot \mathrm{vec}\,\{(X - \nu)(X - \nu)' - \Psi\}'] \equiv \Delta = (\delta_{st})$, which is a $p \times p^2$ matrix given by

$$
(2.11) \qquad \Delta' = \begin{pmatrix}
\kappa_{111} & \kappa_{211} & \cdots & \kappa_{p11} \\
\kappa_{121} & \kappa_{221} & \cdots & \kappa_{p21} \\
\kappa_{131} & \kappa_{231} & \cdots & \kappa_{p31} \\
\vdots & \vdots & & \vdots \\
\kappa_{1p1} & \kappa_{2p1} & \cdots & \kappa_{pp1} \\
\kappa_{112} & \kappa_{212} & \cdots & \kappa_{p12} \\
\kappa_{122} & \kappa_{222} & \cdots & \kappa_{p22} \\
\vdots & \vdots & & \vdots \\
\kappa_{1p2} & \kappa_{2p2} & \cdots & \kappa_{pp2} \\
\vdots & \vdots & & \vdots \\
\kappa_{11p} & \kappa_{21p} & \cdots & \kappa_{p1p} \\
\kappa_{12p} & \kappa_{22p} & \cdots & \kappa_{p2p} \\
\vdots & \vdots & & \vdots \\
\kappa_{1pp} & \kappa_{2pp} & \cdots & \kappa_{ppp}
\end{pmatrix}.
$$

Next, we put $E_G[\text{vec}\,\{(X-\nu)(X-\nu)' - \Psi\}\ \text{vec}\,\{(X-\nu)(X-\nu)' - \Psi\}'] \equiv \Gamma = (\gamma_{st})$, which is a $p^2 \times p^2$ matrix given by

$$
(2.12) \qquad \Gamma = \left(\begin{matrix}
\sigma_{11,11} & \sigma_{11,21} & \cdots & \sigma_{11,p1} & \vdots & \sigma_{11,12} & \sigma_{11,22} & \cdots \\
\sigma_{21,11} & \sigma_{21,21} & \cdots & \sigma_{21,p1} & \vdots & \sigma_{21,12} & \sigma_{21,22} & \cdots \\
\vdots & \vdots & & \vdots & & \vdots & \vdots & \\
\sigma_{p1,11} & \sigma_{p1,21} & \cdots & \sigma_{p1,p1} & \vdots & \sigma_{p1,12} & \sigma_{p1,22} & \cdots \\
\vdots & \vdots & & \vdots & & \vdots & \vdots & \\
\sigma_{1p,11} & \sigma_{1p,21} & \cdots & \sigma_{1p,p1} & \vdots & \sigma_{1p,12} & \sigma_{1p,22} & \cdots \\
\vdots & \vdots & & \vdots & & \vdots & \vdots & \\
\sigma_{pp,11} & \sigma_{pp,21} & \cdots & \sigma_{pp,p1} & \vdots & \sigma_{pp,12} & \sigma_{pp,22} & \cdots
\end{matrix}\right.
$$

$$
\left.\begin{matrix}
\sigma_{11,p2} & \vdots & \cdots & \vdots & \sigma_{11,1p} & \sigma_{11,2p} & \cdots & \sigma_{11,pp} \\
\sigma_{21,p2} & \vdots & \cdots & \vdots & \sigma_{21,1p} & \sigma_{21,2p} & \cdots & \sigma_{21,1p} \\
\vdots & & & & \vdots & \vdots & & \vdots \\
\sigma_{p1,p2} & \vdots & \cdots & \vdots & \sigma_{p1,1p} & \sigma_{p1,2p} & \cdots & \sigma_{p1,pp} \\
\vdots & & & & \vdots & \vdots & & \vdots \\
\sigma_{1p,p2} & \vdots & \cdots & \vdots & \sigma_{1p,1p} & \sigma_{1p,2p} & \cdots & \sigma_{1p,pp} \\
\vdots & & & & \vdots & \vdots & & \vdots \\
\sigma_{pp,p2} & \vdots & \cdots & \vdots & \sigma_{pp,1p} & \sigma_{pp,2p} & \cdots & \sigma_{pp,pp}
\end{matrix}\right).
$$

In (2.11) and (2.12),

$$
\sigma_{ij,lm} = \kappa_{ijlm} + \psi_{il}\psi_{jm} + \psi_{im}\psi_{jl}\,,
$$

and $\kappa_{ijl}$ and $\kappa_{ijlm}$ denote the 3rd and 4th order multivariate cumulants of $G$, respectively. We note in addition that

$$\delta_{st} = \kappa_{ijl} \quad \text{for} \quad s = i \quad \text{and} \quad t = j + (l-1)p ,$$
$$\gamma_{st} = \sigma_{ij,lm} \quad \text{for} \quad s = i + (j-1)p \quad \text{and} \quad t = l + (m-1)p .$$

In terms of Kronecker products and a $p^2 \times p^2$ permutation matrix $T$, $\Gamma$ can be rewritten as follows:

$$\Gamma = K^{(4)} + \Psi \otimes \Psi + T(\Psi \otimes \Psi) ; \tag{2.13}$$

here $K^{(4)}$ is a $p^2 \times p^2$ matrix whose elements are the 4th order cumulants $\kappa_{ijlm}$, and the permutation matrix $T$ is given by

$$T = \begin{pmatrix} E_{11} & E_{21} & E_{31} & \cdots & E_{p1} \\ E_{12} & E_{22} & E_{32} & \cdots & E_{p2} \\ E_{1p} & E_{2p} & E_{3p} & \cdots & E_{pp} \end{pmatrix} ,$$

with $p \times p$ matrices $E_{ij}$'s whose $(i, j)$ element is 1 and the other elements are zero. Note that the permutation matrix $T$ has the following properties: (1) $T = T'$ and $T^2 = I$, the identity matrix of order $p^2$ and (2) $T \operatorname{vec}(A) = \operatorname{vec}(A')$ for an arbitrary $p \times p$ matrix $A$.

For the formula of $\operatorname{var}_F(w)$, we need only replace $\nu$ and $\Psi$ in the formula of $\operatorname{var}_G(w)$ by $\mu$ and $\Sigma$, respectively, and set the 3rd and 4th order multivariate cumulants equal to zero. Then

$$\operatorname{var}_F(u_1) = \Sigma, \qquad \operatorname{cov}_F(u_1, u_2) = \operatorname{cov}_F(u_1, u_2)' = 0 , \tag{2.14}$$

and

$$\operatorname{var}_F(u_2) = \Sigma \otimes \Sigma + T(\Sigma \otimes \Sigma) , \tag{2.15}$$

where $T$ denotes the $p^2 \times p^2$ permutation matrix defined above.

### 2.1 *Influence functions of the moments of* IF($X$; $\mu$) *and* IF($X$; $\Sigma$)

From the above results we can easily evaluate the influence functions of the moments $E_F(w)$ and $\operatorname{var}_F(w)$ of the influence functions **IF**($X$; $\mu$) and **IF**($X$; $\Sigma$). They are given by

$$\mathbf{IF}(G; E_F(w)) = E_G(w) , \tag{2.16}$$

and

(2.17) $\quad \mathbf{IF}(G; \mathrm{var}_F(w)) = \mathrm{var}_G(w) - \mathrm{var}_F(w) + E_G(w)E_G(w)'$ .

By utilizing these formulas we may investigate the first two moments of $R(X; |\Sigma|)$ in (1.4). Recall that

$$R(X; |\Sigma|) = \mathrm{tr}\, \Sigma^{-1}\{(X-\mu)(X-\mu)' - \Sigma\}$$
$$= \mathrm{vec}\,(\Sigma^{-1})'\, \mathrm{vec}\,\{(X-\mu)(X-\mu)' - \Sigma\} = \mathrm{vec}\,(\Sigma^{-1})' u_2 \,.$$

So, we have

$$E_H[R(X; |\Sigma|)] = \mathrm{vec}\,(\Sigma^{-1})' E_H(u_2) \,,$$
$$\mathrm{var}_H(R(X; |\Sigma|)) = \mathrm{vec}\,(\Sigma^{-1})'\, \mathrm{var}_H(u_2)\, \mathrm{vec}\,(\Sigma^{-1}) \,.$$

The influence functions of the first two moments of $R(X; |\Sigma|)$ are thus

(2.18) $\quad \mathbf{IF}(G; E_F[R(X; |\Sigma|)]) = \mathrm{vec}\,(\Sigma^{-1})' E_G(u_2)$ ,

and

(2.19) $\quad \mathbf{IF}(G; \mathrm{var}_F(R(X; |\Sigma|))) = \mathrm{vec}\,(\Sigma^{-1})'\{\mathrm{var}_G(u_2) - \mathrm{var}_F(u_2) + E_G(u_2)E_G(u_2)'\}\, \mathrm{vec}\,(\Sigma^{-1})$ .

We next evaluate (2.18) and (2.19) in some special cases:

*Case* 1. (location shift problem) $\Psi = \Sigma$, $F$ and $G$ are normal.

$$\mathbf{IF}[G; E_F(R(X; |\Sigma|))] = (\nu-\mu)'\Sigma^{-1}(\nu-\mu) \,,$$
$$\mathbf{IF}[G; \mathrm{var}_F(R(X; |\Sigma|))] = 4(\nu-\mu)'\Sigma^{-1}(\nu-\mu) + [(\nu-\mu)'\Sigma^{-1}(\nu-\mu)]^2 \,.$$

*Case* 2. (variance discrepancy problem) $\nu = \mu$, $F$ and $G$ are normal.

$$\mathbf{IF}[G; E_F(R(X; |\Sigma|))] = \mathrm{tr}\,(\Psi\Sigma^{-1}) - p \,,$$
$$\mathbf{IF}[G; \mathrm{var}_F(R(X; |\Sigma|))] = 2[\,\mathrm{tr}\,(\Sigma^{-1}\Psi\Sigma^{-1}\Psi) - p] + [\,\mathrm{tr}\,(\Psi\Sigma^{-1}) - p]^2 \,.$$

*Case* 3. (non-normal case) $\nu = \mu$, $\Psi = \Sigma$, $F$ is normal and $G$ is non-normal.

$$\mathbf{IF}[G; E_F(R(X; |\Sigma|))] = 0 \,,$$

$$\mathbf{IF}[G;\ \mathrm{var}_F(R(X;\ |\Sigma|))] = \mathrm{vec}\ (\Sigma^{-1})' K^{(4)}\ \mathrm{vec}\ (\Sigma^{-1})$$
$$= \sum_i \sum_j \sum_l \sum_m \sigma^{ij} \sigma^{lm} \kappa_{ijlm}\ .$$

We focus now on Case 3, in which the value of $\mathbf{IF}[G;\ \mathrm{var}_F(R(X;\ |\Sigma|))]$ is equivalent to Mardia's (1970) measure of multivariate kurtosis $\beta_{2,p}$. Under the conditions of Case 3, (2.17) can be rewritten as

$$\mathrm{var}_G[R(X;\ |\Sigma|)] = \mathrm{var}_F[R(X;\ |\Sigma|)] + \mathbf{IF}[G;\ \mathrm{var}_F(R(X;\ |\Sigma|))]\ ,$$

or equivalently,

$$\mathrm{var}_G[\mathbf{IF}(X;\ |\Sigma|)] = \mathrm{var}_F[\mathbf{IF}(X;\ |\Sigma|)] + \mathbf{IF}[G;\ \mathrm{var}_F(\mathbf{IF}(X;\ |\Sigma|))]\ .$$

This relation well explains the character of $\beta_{2,p}$ as a diagnostic tool in using Healy's (1968) $\chi^2$ plot under normality, and also suggests some ways to define measures of testing multivariate normality. That is, under the assumption that $\mu = \nu$, $\Psi = \Sigma$, $F$ is normal and $G$ is non-normal, from (2.17) we have the basic relation

$$\mathrm{var}_G(w) = \mathrm{var}_F(w) + \mathbf{IF}(G;\ \mathrm{var}_F(w))\ . \tag{2.20}$$

By using $\mathrm{var}_G(w)$ in (2.20) we shall define measures for assessing multivariate normality.

## 2.2 *Measures of dependence*

Under the assumption that $G$ is a $p$-variate non-normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$, we shall consider the problem of evaluating the magnitude of the correlation between the random vectors $u_1$ and $u_2$ in $w$; recall that zero-correlation between $u_1$ and $u_2$ is a necessary condition for assuring multivariate normality.

Two types of measures of dependence $D_{\mathrm{sum}}$ and $D_{\mathrm{max}}$ are introduced. $D_{\mathrm{sum}}$ is defined as the sum of canonical correlations between $u_1$ and $u_2$, namely,

$$D_{\mathrm{sum}} = \mathrm{tr}\ (\Sigma^{-1} \Delta \Gamma^{-} \Delta')\ , \tag{2.21}$$

where $\Gamma^{-}$ denotes a generalized inverse of Moore-Penrose type (see Rao (1973), Section 1b.5). In the case of the 4th order multivariate cumulants $\kappa_{ijlm}$ being all zero, we have $D_{\mathrm{sum}} = (1/2)\beta_{1,p}$, where $\Gamma^{-} = (1/4)(I + T)$ and $\beta_{1,p}$ is a measure of multivariate skewness introduced by Mardia (1970).

$D_{\mathrm{sum}}$ is invariant under the affine transformation of the random vector $X \to QX + r$, where $Q$ is an arbitrary $p \times p$ nonsingular constant matrix and $r$ is an arbitrary $p \times 1$ constant vector. This invariance property can be easily shown by noting that under the transformation $X \to QX + r$, the random vector $w = (u_1', u_2')$ is transformed to

$$\begin{pmatrix} Q & 0 \\ 0 & Q' \otimes Q \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

For a given random sample $X_1, X_2, \ldots, X_n$ of size $n$, a sample version $\hat{D}_{\text{sum}}$ of $D_{\text{sum}}$ is defined by replacing the population cumulants by the corresponding $k$-statistics. Under normality, $(n/3)\hat{D}_{\text{sum}}$ has the same asymptotic distribution of $\chi^2$ with $p(p+1)(p+2)/6$ degrees of freedom as that of Mardia's $(n/6)b_{1,p}$.

The object of the other measure of dependence $D_{\max}$ is to examine the correlation between $u_1$ and $u_2$ by using $a'X$, a linear combination of the random vector $X$, where $a' = (a_1, \ldots, a_p)$ is a $p \times 1$ scalar vector. $D_{\max}$ is defined by

$$D_{\max} = \max_a \frac{[\operatorname{cov}_G(a'u_1, (a' \otimes a')u_2)]^2}{\operatorname{var}_G(a'u_1)\operatorname{var}_G((a' \otimes a')u_2)}. \tag{2.22}$$

Here we carry out the maximization in $a$ under the condition that $\operatorname{var}_G(a'u_1) = 1$. Then $D_{\max}$ may be rewritten as

$$\begin{aligned} D_{\max} &= \max_a \frac{[a'\Delta(a \otimes a)]^2}{(a' \otimes a')\Gamma(a \otimes a)} \\ &= \max_a \frac{\left[\sum_i \sum_j \sum_l a_i a_j a_l \kappa_{ijl}\right]^2}{\sum_i \sum_j \sum_l \sum_m a_i a_j a_l a_m \kappa_{ijlm} + 2}, \end{aligned} \tag{2.23}$$

under $a'\Sigma a = 1$.

In the case of the 4th order multivariate cumulants $\kappa_{ijlm}$ being all zero, we have $D_{\max} = (1/2)b_1^*$, where $b_1^*$ is a measure of multivariate skewness as defined by Malkovich and Afifi (1973).

Clearly, $D_{\max}$ is also affine invariant, and for a random sample of size $n$, a sample version $\hat{D}_{\max}$ of $D_{\max}$ is defined similarly as in the case of $\hat{D}_{\text{sum}}$. It is difficult to calculate the sampling distribution of $\hat{D}_{\max}$ exactly, but a simulation study is feasible.

## 2.3 *Monte Carlo study*

In this section we shall examine some finite sample properties of $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$ for various non-normal models with dimension $p = 2$. For the non-normal models, we consider the following two component normal mixture models, which were studied earlier in Isogai (1983*b*):

$$\text{(M1)} \qquad (1-\pi)N_2(0, I_2) + \pi N_2\left[\begin{pmatrix} m \\ m \end{pmatrix}, s^2\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}\right]$$

$$\text{with } 1 > \pi > 0,\ 1 > r > -1,\ m \geq 0\,,$$

$$\text{(M2)} \qquad (1-\pi)N_2(0, I_2) + \pi N_2\left[\begin{pmatrix} m \\ m \end{pmatrix}, s^2\begin{pmatrix} 1-r & 0 \\ 0 & 1+r \end{pmatrix}\right]$$

$$\text{with } 1 > \pi > 0,\ 1 > r \geq 0,\ m \geq 0\,,$$

where $N_2(\mu, \Sigma)$ in (M1) and (M2) denotes the density function of a 2-dimensional normal distribution $N_2(\mu, \Sigma)$.

By specifying the values of 4 parameters $\pi$, $r$, $m$ and $s^2$ in each of the above mixture models, we have the following non-normal cases:

(s1) $\pi = .25,\quad m = 3,\quad s^2 = 3,\quad 1 > r > -1$ with (M1) ,

(s2) $\pi = .25,\quad m = 3,\quad s^2 = 3,\quad 1 > r \geq 0$ with (M2) ,

(k1) $\pi = .25,\quad m = 0,\quad s^2 = 3,\quad 1 > r \geq 0$ with (M1) (or (M2)) ,

Table 1. 5% level power of $\hat{D}_{\text{sum}}$ and $\hat{D}_{\text{max}}$ and other related measures with sample size $n = 30$.

| Non-normal cases | | | Measures | | | | |
|---|---|---|---|---|---|---|---|
| | | | $\hat{D}_{\text{sum}}$ | $\hat{D}_{\text{max}}$ | $b_{1,2}$ | $\hat{\eta}^2_{\text{max}}$ | $b_1^*$ |
| (s1) | $\pi = .25$ | $r = -.9$ | 71% | 80% | 68% | 51% | 62% |
| | $s^2 = 3$ | $r = -.5$ | 66 | 80 | 66 | 40 | 61 |
| | $m = 3$ | $r = 0.0$ | 59 | 82 | 62 | 35 | 60 |
| | | $r = .5$ | 52 | 84 | 58 | 40 | 62 |
| | | $r = .9$ | 80 | 86 | 70 | 81 | 64 |
| (s2) | $\pi = .25$ | $r = 0.0$ | 59 | 82 | 61 | 35 | 59 |
| | $s^2 = 3$ | $r = .5$ | 59 | 81 | 61 | 40 | 63 |
| | $m = 3$ | $r = .9$ | 64 | 82 | 62 | 66 | 63 |
| (k1) | $\pi = .25$ | $r = 0.0$ | 14 | 16 | 23 | 12 | 20 |
| | $s^2 = 3$ | $r = .5$ | 14 | 16 | 22 | 12 | 21 |
| | $m = 0$ | $r = .9$ | 16 | 15 | 22 | 15 | 19 |
| (k2) | $s^2 = 1$ | $\pi = .25$ | 9 | 8 | 10 | 6 | 8 |
| | $m = 0$ | $\pi = .50$ | 15 | 15 | 19 | 13 | 19 |
| | $r = .9$ | $\pi = .75$ | 24 | 25 | 33 | 20 | 32 |
| (k3) | $\pi = .50$ | $r = -.9$ | 42 | 42 | 7 | 35 | 5 |
| | $s^2 = 1$ | $r = -.5$ | 15 | 19 | 4 | 13 | 3 |
| | $m = 3$ | $r = 0.0$ | 7 | 9 | 2 | 5 | 2 |
| | | $r = .5$ | 14 | 12 | 7 | 16 | 6 |
| | | $r = .9$ | 86 | 60 | 59 | 80 | 37 |
| (k4) | $\pi = .25$ | $m = 1$ | 36 | 25 | 36 | 32 | 30 |
| | $s^2 = 3$ | $m = 2$ | 46 | 49 | 54 | 41 | 48 |
| | $r = -.9$ | | | | | | |

(k2) $\pi = .25, .50, .75,\quad m = 0,\quad s^2 = 1,\quad r = .9$ with (M1) ,

(k3) $\pi = .50,\quad m = 3,\quad s^2 = 1,\quad 1 > r > -1$ with (M1) ,

(k4) $\pi = .25,\quad m = 1, 2,\quad s^2 = 3,\quad r = -.9$ with (M1) .

First, with respect to each of $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$, 5000 samples of size $n$ ( = 30, 50) were drawn from a 2-dimensional normal $N_2(0, I_2)$. The empirical 5% points were obtained from the corresponding order statistics of $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$.

Next, with $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$, 1000 samples of the same size as above were drawn from each of non-normal cases. The empirical power was calculated by the proportion of samples falling in the 5% empirical critical region.

The results for the non-normal cases are given in Tables 1 and 2, where for purposes of comparison we also reprint the simulation results of the measures $b_{1,2}$ and $b_1^*$ and $\hat{\eta}^2_{\max}$, taken from Isogai (1983*b*) (for details concerning the measure $\hat{\eta}^2_{\max}$, see Cox and Small (1978)). Obviously, $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$ seem to have good power compared to the other measures; in practice, we would recommend the use of $\hat{D}_{\text{sum}}$.

Table 2. 5% level power of $\hat{D}_{\text{sum}}$ and $\hat{D}_{\max}$ and other related measures with sample size $n = 50$.

| Non-normal cases | | | Measures | | | | |
|---|---|---|---|---|---|---|---|
| | | | $\hat{D}_{\text{sum}}$ | $\hat{D}_{\max}$ | $b_{1,2}$ | $\hat{\eta}^2_{\max}$ | $b_1^*$ |
| (s1) | $\pi = .25$ | $r = -.9$ | 95% | 96% | 92% | 76% | 88% |
| | $s^2 = 3$ | $r = -.5$ | 95 | 97 | 91 | 64 | 88 |
| | $m = 3$ | $r = 0.0$ | 95 | 97 | 89 | 54 | 88 |
| | | $r = .5$ | 93 | 98 | 86 | 62 | 90 |
| | | $r = .9$ | 97 | 98 | 96 | 97 | 93 |
| (s2) | $\pi = .25$ | $r = 0.0$ | 95 | 97 | 89 | 54 | 88 |
| | $s^2 = 3$ | $r = .5$ | 94 | 98 | 88 | 63 | 90 |
| | $m = 3$ | $r = .9$ | 94 | 97 | 89 | 90 | 91 |
| (k1) | $\pi = .25$ | $r = 0.0$ | 19 | 20 | 29 | 16 | 29 |
| | $s^2 = 3$ | $r = .5$ | 16 | 18 | 27 | 16 | 28 |
| | $m = 0$ | $r = .9$ | 17 | 17 | 26 | 20 | 27 |
| (k2) | $s^2 = 1$ | $\pi = .25$ | 9 | 8 | 9 | 8 | 10 |
| | $m = 0$ | $\pi = .50$ | 17 | 15 | 22 | 16 | 23 |
| | $r = .9$ | $\pi = .75$ | 26 | 24 | 40 | 27 | 42 |
| (k3) | $\pi = .50$ | $r = -.9$ | 53 | 44 | 12 | 46 | 7 |
| | $s^2 = 1$ | $r = -.5$ | 20 | 18 | 5 | 17 | 4 |
| | $m = 3$ | $r = 0.0$ | 7 | 8 | 1 | 5 | 2 |
| | | $r = .5$ | 23 | 17 | 11 | 26 | 9 |
| | | $r = .9$ | 98 | 81 | 90 | 98 | 68 |
| (k4) | $\pi = .25$ | $m = 1$ | 59 | 40 | 57 | 54 | 44 |
| | $s^2 = 3$ | $m = 2$ | 77 | 75 | 80 | 67 | 72 |
| | $r = -.9$ | | | | | | |

## 3. Use of a score function

The definitions of the influence functions $\mathbf{IF}(X; \mu)$ and $\mathbf{IF}(X; \Sigma)$ with respect to the population moments $\mu$ and $\Sigma$ can be easily extended by incorporation of the log likelihood of an influential observation $X$. Suppose that the true distribution $F$ has a density $f(x)$ belonging to the family of densities $\{f(x, \theta);\ \theta \in \Theta,\ \Theta$ is an appropriate subset in some Euclidean space$\}$ (say $\mathscr{F}$), with the densities $f(x, \theta)$ being sufficiently regular, and that the density $f(x)$ is indexed by $\theta = \theta_0 \in \Theta$. Also suppose that a contaminate component $G$ has a density $g(x)$ which may or may not belong to the family $\mathscr{F}$.

Then we can define a quasi-influence function for the population parameter $\theta = \theta_0$ by using its score function as

$$\text{(3.1)} \qquad \mathbf{QIF}(X; \theta_0) = \frac{\partial}{\partial \theta} \log f(X, \theta)|_{\theta=\theta_0}\,.$$

This definition (3.1) is obviously a generalization of (1.1) and (1.2). The formulas corresponding to (2.16) and (2.17) can be written as

$$\text{(3.2)} \qquad E_G[\mathbf{QIF}(X, \theta_0)] = \mathbf{IF}[G; E_{\theta_0}\{\mathbf{QIF}(X, \theta_0)\}]\,,$$

and

$$\text{(3.3)} \qquad \begin{aligned} &\mathrm{var}_G(\mathbf{QIF}(X, \theta_0)) + E_G[\mathbf{QIF}(X, \theta_0)]' E_G[\mathbf{QIF}(X, \theta_0)] \\ &\quad = \mathrm{var}_{\theta_0}(\mathbf{QIF}(X, \theta_0)) + \mathbf{IF}[G; \mathrm{var}_{\theta_0}(\mathbf{QIF}(X, \theta_0))]\,. \end{aligned}$$

In the following we shall apply formulas (3.2) and (3.3) to the usual regression analysis with some parametric families, and derive test statistics for multivariate normality.

### 3.1 *Application to regression analysis*

Using the concept of a quasi-influence function, we here consider an application to the usual linear regression analysis. The model is represented by

$$\text{(3.4)} \qquad y_i = z_i'\beta + e_i, \qquad i = 1, 2, \ldots, n\,,$$

where $z_i' = (z_{i1}, z_{i2}, \ldots, z_{iq})$ is a $q \times 1$ vector of known constants corresponding to the $i$-th level of a $q$-dimensional regression vector $z' = (z_1, z_2, \ldots, z_q)$, $y_i$ is the $i$-th observation of the response variable $y$, $\beta$ is a $q \times 1$ vector of unknown parameters, and $e_i$ is a random variable, typically called the error term, with $E(e_i) = 0$ and var $(e_i) = \sigma^2$ $(i = 1, \ldots, n)$.

In matrix notation the model becomes

$$Y = Z\beta + e\,, \tag{3.5}$$

where $Y' = (y_1,\ldots,y_n)$, $e' = (e_1,\ldots,e_n)$ and $Z$ is an $n\times q$ matrix of rank $q$ whose $i$-th row vector is $z_i'$. Usually, we assume that the distribution $F$ of the error vector $e$ is an $n$-variate normal $N_n(0, \sigma^2 I)$. Then, the least squares estimator of $\beta$, denoted by $\hat\beta$, is given by

$$\hat\beta = (Z'Z)^{-1}Z'Y\,, \tag{3.6}$$

and $\hat\beta$ is distributed as $N_q(\beta, \sigma^2(Z'Z)^{-1})$ under $F$.

From the density $N_q(\beta, \sigma^2(Z'Z)^{-1})$ of $\hat\beta$, quasi-influence functions of $\beta$ and $\sigma^2$ are obtained as

$$\mathbf{QIF}(Y;\beta) = (Z'Z)(\hat\beta-\beta) = Z'(Y-Z\beta) = Z'e\,, \tag{3.7}$$

and

$$\begin{aligned}\mathbf{QIF}(Y;\sigma^2) &= (\hat\beta-\beta)'(Z'Z)(\hat\beta-\beta) - q\sigma^2\\ &= e'Z(Z'Z)^{-1}Z'e - q\sigma^2\,.\end{aligned} \tag{3.8}$$

Under the normality assumption $\mathbf{QIF}(Y;\beta)$ and $\mathbf{QIF}(Y;\sigma^2)$ are independent. Thus, when the normality of the error vector $e$ is violated, we may use formulas (3.2) and (3.3) to evaluate the mutual dependence between $\mathbf{QIF}(Y;\beta)$ and $\mathbf{QIF}(Y;\sigma^2)$.

Suppose that under a contaminate component $G$ the error vector $e$ has $E_G(e) = 0$, $\mathrm{var}_G(e) = \sigma^2 I$ and the 3rd and 4th order multivariate cumulants $\kappa_{ijl}$ and $\kappa_{ijlm}$, $i, j, l, m = 1, 2,\ldots, n$. By applying formulas (2.8), (2.9) and (2.10) to $\mathbf{QIF}(Y;\beta)$ (say $\tilde u_1$) and $\mathbf{QIF}(Y;\sigma^2)$ (say $\tilde u_2$), we have

$$\mathrm{var}_G(\tilde u_1) = \sigma^2(Z'Z)\,, \tag{3.9}$$

$$\mathrm{cov}_G(\tilde u_1, \tilde u_2) = Z'\Delta\, \mathrm{vec}\,(V)\,, \tag{3.10}$$

$$\mathrm{var}_G(\tilde u_2) = \mathrm{vec}\,(V)'[K^{(4)} + (\sigma^2 I)\otimes(\sigma^2 I) + T(\sigma^2 I)\otimes(\sigma^2 I)]\,\mathrm{vec}\,(V)\,, \tag{3.11}$$

where we put

$$V = (v_{ij}) = Z(Z'Z)^{-1}Z'\,, \tag{3.12}$$

and $\Delta$ is an $n\times n^2$ matrix of $\kappa_{ijl}$'s, $K^{(4)}$ is an $n^2\times n^2$ matrix of $\kappa_{ijlm}$'s, $T$ is an $n^2\times n^2$ permutation matrix, and $\Delta$, $K^{(4)}$ and $T$ have the same structures as those defined in Section 2.

From (2.21), a measure of dependence $D_{\text{REG}}$ is defined as

$$D_{\text{REG}} = \frac{\text{cov}_G(\tilde{u}_1, \tilde{u}_2)'\{\text{var}_G(\tilde{u}_1)\}^{-1}\text{cov}_G(\tilde{u}_1, \tilde{u}_2)}{\text{var}_G(\tilde{u}_2)} \tag{3.13}$$

$$= \frac{\text{vec}\,(V)'\Delta' V\Delta\,\text{vec}\,(V)}{\sigma^2[\,\text{vec}\,(V)'K^{(4)}\,\text{vec}\,(V) + 2q(\sigma^2)^2]}\,.$$

Note that $D_{\text{REG}}$ is invariant under the linear transformation $Z \to ZQ$ with a $q \times q$ nonsingular matrix $Q$.

To simplify the expression (3.13) of $D_{\text{REG}}$, we consider the case where the $e_i$'s, the elements of the error vector $e$, are independently and identically distributed under $G$. Then the 3rd and 4th order multivariate cumulants $\kappa_{ijl}$ and $\kappa_{ijlm}$ of $e$ are

$$\kappa_{111} = \kappa_{222} = \cdots = \kappa_{nnn} \qquad (\text{say } \kappa_3)\,,$$

$$\kappa_{ijl} = 0 \qquad (\text{otherwise})\,,$$

and

$$\kappa_{1111} = \kappa_{2222} = \cdots = \kappa_{nnnn} \qquad (\text{say } \kappa_4)\,,$$

$$\kappa_{ijlm} = 0 \qquad (\text{otherwise})\,.$$

Under this condition $D_{\text{REG}}$ is reduced to

$$D_{\text{REG}} = \frac{(\upsilon_S)(\gamma_1)^2}{[(\upsilon_K)\gamma_2 + 2q]}\,, \tag{3.14}$$

where we put $\gamma_1 = \kappa_3/(\sigma^2)^{3/2}$, $\gamma_2 = \kappa_4/(\sigma^2)^2$,

$$\upsilon_S = \sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ii}\upsilon_{ij}\upsilon_{jj} \tag{3.15}$$

and

$$\upsilon_K = \sum_{i=1}^{n} (\upsilon_{ii})^2\,. \tag{3.16}$$

We remark that $\upsilon_K$ corresponds to the quantity $m$ introduced by Box and Watson (1962); and, to use $D_{\text{REG}}$ in practice we must estimate $\gamma_1$ and $\gamma_2$ by appropriate residuals. Here we shall consider how to choose an experimental design which would reduce the dependence between **QIF**$(Y; \beta)$ and **QIF**$(Y; \sigma^2)$ under non-normality.

From the facts that $1 \geq \upsilon_{ii} \geq 0$, $\sum_{i=1}^{n} \upsilon_{ii} = q$ and $\upsilon_K \geq \upsilon_S \geq 0$, and that $D_{\text{REG}}$ is continuous in both $\upsilon_K$ and $\upsilon_S$, we may infer that $D_{\text{REG}}$ is bounded. If $\gamma_1$ and $\gamma_2$ are known, we may choose the design $Z$, which minimizes $D_{\text{REG}}$. But $\gamma_1$ and $\gamma_2$ are usually unknown, and so instead of using $D_{\text{REG}}$ directly, we shall deal with a quantity which dominates $D_{\text{REG}}$.

From the fact that $q^2/n \leq \upsilon_K \leq q$, we have

$$(3.17) \qquad D_{\text{REG}} \leq \max_{\upsilon_K} D_{\text{REG}} = \begin{cases} (\upsilon_S \gamma_1^2)/[(q^2/n)\gamma_2 + 2q]^{1/2} & \text{for} \quad \gamma_2 \geq 0\,, \\ (\upsilon_S \gamma_1^2)/[q\gamma_2 + 2q]^{1/2} & \text{for} \quad \gamma_2 \leq 0\,. \end{cases}$$

The factor $\upsilon_S$ is essential to reduce $\max_{\upsilon_K} D_{\text{REG}}$.

To examine the structure of the factor $\upsilon_S$, we further assume that the $q$-dimensional regression vector $z$ is divided into one constant term and the $q-1$ (say $p$)-dimensional regression vector $x$, that is, $z' = (1, x')$ with $q = 1 + p$. Then the $i$-th level of the regression vector $z$ is represented as $z_i' = (1,\ x_i')$, $i = 1, \ldots, n$ in terms of the $i$-th level of the $p$-dimensional regression vector $x$.

Following the method of Box and Watson (1962), in which they derived a measure of kurtosis $C_X$, we have

$$(3.18) \quad \upsilon_S = \frac{1}{n} + \frac{p^2 + 2p(n-1)}{n(n-1)^2} + \frac{(n-2)^2}{n^2(n-1)} \sum_i \sum_{i'} \sum_j \sum_{j'} \sum_l \sum_{l'} k^{ii'} k_{ii'j} k^{jj'} k_{j'l l'} k^{ll'}\,,$$

where $i, i', j, j', l$ and $l'$ range from 1 to $p$, $k_{ij}$ and $k_{ijl}$ are, respectively, the 2nd and 3rd order $k$-statistics based on $n$ observed regression vectors $x_1, x_2, \ldots, x_n$, and we set $(k^{ij}) = (k_{ij})^{-1}$.

Here we put

$$(3.19) \qquad S_X = \sum_i \sum_{i'} \sum_j \sum_{j'} \sum_l \sum_{l'} k^{ii'} k_{ii'j} k^{jj'} k_{j'l l'} k^{ll'}\,.$$

$S_X$ is a sample version for a measure of multivariate skewness $\operatorname{tr}(S_2)$ introduced by Isogai (1983*a*), which is nonnegative and invariant under the affine transformation of the regression vector $x \to Qx + r$ with a $p \times p$ nonsingular matrix $Q$ and a $p \times 1$ vector $r$. Thus, if we can choose an experimental design which makes $S_X = 0$, that is, if we select $n$ design vectors $x_1, \ldots, x_n$ to have zero multivariate skewness in the sense of $\operatorname{tr}(S_2)$, we would reduce $D_{\text{REG}}$, or at least we can ensure that $\max_{\upsilon_K} D_{\text{REG}}$ is of the same order $n^{-1}$ as $\upsilon_S$.

3.2 *Index of goodness-of-fit*

In this section we remark that under certain circumstances formulas (3.2) and (3.3) reduce to Fisher's information, which may be used to evaluate a discrepancy between different distributions. Also, by using Fisher's information measure we may derive some indices of goodness-of-fit and test statistics for multivariate normality.

Now let us consider the case where the contaminate component $G$ belongs to the family $\mathscr{F}$ and its density $g(x)$ is specified by some $\theta \in \Theta$, but we have little knowledge of its shape. As the first step, from (3.2) it is reasonable to examine the behavior of $E_\theta[\mathbf{QIF}(X, \theta_0)]$ as $\theta$ tends to $\theta_0$. Under appropriate regularity conditions, we have $E_{\theta_0}[\mathbf{QIF}(X, \theta_0)] = 0$, and so

$$(3.20) \qquad \frac{\partial}{\partial \theta} E_\theta[\mathbf{QIF}(X, \theta_0)]|_{\theta=\theta_0}$$

is meaningful. The quantity (3.20) is usually equal to $\mathrm{var}_{\theta_0}(\mathbf{QIF}(X, \theta_0))$ and denotes the Fisher's information matrix, which is an additive component of the right hand of (3.3). Therefore, we can use the quantity (3.20) as a measure to evaluate the discrepancy between $G$ and $F$.

A measure of the discrepancy $D(F, G)$ is defined by

$$(3.21) \qquad \begin{aligned} D(F, G) &= \frac{\partial}{\partial \theta} E_\theta[\mathbf{QIF}(X, \theta_0)]|_{\theta=\theta_0} \\ &= \mathrm{var}_{\theta_0}(\mathbf{QIF}(X, \theta_0)) . \end{aligned}$$

Here, we give some examples in the following.

*Example* 1. $\mathscr{F} = \{(1-\theta)f(x) + \theta g(x); 0 \le \theta \le 1\}$ .

$$(3.22) \qquad D(F, G) = \mathrm{var}_F\left(\frac{g(X)}{f(X)}\right),$$

which is called the mean square contingency.

*Example* 2. $\mathscr{F} = \{f(x)^{1-\theta} g(x)^\theta / C(\theta);\ C(\theta) = \int f(x)^{1-\theta} g(x)^\theta dx$ and $0 \le \theta \le 1\}$. Then

$$(3.23) \qquad D(F, G) = \mathrm{var}_F\left(\log \frac{g(X)}{f(X)}\right).$$

*Remark*. In Examples 1 and 2, let the density $f(x)$ be a $p$-variate normal with zero mean vector, and covariance matrix $I_p$ the identity matrix of order $p$: therefore, $f(x)$ can be expressed as $f(x)=\prod_{t=1}^{p}\alpha(x_t)$ with $\alpha(x)$ denoting the density of the univariate standard normal distribution. Also, let the density $g(x)$ be a $p$-variate non-normal, with zero mean vector and covariance matrix $I_p$, which has a Gram-Charlier Type A series expansion. That is, suppose that $g(x)$ may be formally expanded as

$$g(x)=\left\{\prod_{t=1}^{p}\alpha(x_t)\right\}\times\left\{1+\Sigma'\kappa^{(3)}_{r_1r_2\cdots r_p}\left(\prod_{t=1}^{p}H_{r_t}(x_t)\right)\Big/(r_1!r_2!\cdots r_p!)\right.$$
$$\left.+\Sigma''\kappa^{(4)}_{r_1r_2\cdots r_p}\left(\prod_{t=1}^{p}H_{r_t}(x_t)\right)\Big/(r_1!r_2!\cdots r_p!)\right\},$$

where $\Sigma'$ and $\Sigma''$ denote the summations with $r_i$, $i=1,\dots,p$ over the sets $\left\{(r_1, r_2,\dots,r_p);\ \sum_{t=1}^{p}r_t\geq 0,\ r_t\text{'s are integers}\right\}$ and $\left\{(r_1,\dots,r_p);\ \sum_{t=1}^{p}r_t=4,\ r_t\geq 0,\ r_t\text{'s are integers}\right\}$, respectively, $\kappa^{(3)}_{r_1r_2\cdots r_p}$ and $\kappa^{(4)}_{r_1\cdots r_p}$ are the 3rd and 4th order multivariate cumulants, respectively, and $r_i$ denotes the degree of the Chebyshev-Hermite polynomial $H_{r_i}(\cdot)$ with the corresponding random variables, and $H_r(x)$ is defined by $H_r(x)=\{(-d/dx)^r\alpha(x)\}/\alpha(x)$. Then we have

$$D(F,G)=(3!)^{-1}\beta_{1,p}+(4!)^{-1}\,\mathrm{tr}\,(K_2)\,, \tag{3.24}$$

where $\beta_{1,p}$ is Mardia's (1970) measure of multivariate skewness and tr $(K_2)$ is a measure of multivariate kurtosis introduced by Isogai (1983*a*).

*Example* 3. $\mathscr{F}=\{f(x,\theta);\ f(x)$ is a $p$-variate normal with mean vector $\mu$ and covariance matrix $\Sigma$ and $f(x,\theta)$ is defined by

$$f(x,\theta)=C(\mu,\Sigma,\theta)^{-1}\exp\left[-2^{-1}\|x-\mu\|^{2+\theta}\right]$$

with

$$\|x-\mu\|=\{(x-\mu)'\Sigma^{-1}(x-\mu)\}^{1/2}\,,$$
$$C(\mu,\Sigma,\theta)=\int_{R^p}\exp\left[-2^{-1}\|x-\mu\|^{2+\theta}\right]dx\,.$$

$\theta$ belongs to some open interval $\Theta$ such that $C(\mu,\Sigma,\theta)$ exists$\}$. Thus $f(x)$ is specified by $\theta=0$. Then we have

$$D(F,G)=16^{-1}\,\mathrm{var}_F(\|X-\mu\|^2\log\|X-\mu\|^2)\,. \tag{3.25}$$

The random variable $\|X-\mu\|^2\log\|X-\mu\|^2$ is a multivariate version of a test statistic $Z^2\log Z^2$ ($Z$ is a standard normal variate) introduced by

Spiegelhalter (1983) to examine normality against symmetric families.

## Acknowledgements

The author would like to thank the referees for useful comments on an earlier draft of this paper. Also, he greatly appreciates the detailed suggestions made by the associate editor.

## REFERENCES

Box, G. E. P. and Watson, G. S. (1962). Robustness to non-normality of regression test, *Biometrika*, **49**, 93-106.

Cox, D. R. and Small, N. J. H. (1978). Testing multivariate normality, *Biometrika*, **65**, 263-272.

Healy, M. J. R. (1968). Multivariate normal plotting, *Appl. Statist.*, **17**, 157-161.

Isogai, T. (1983*a*). On measures of multivariate skewness and kurtosis, *Math. Japon.*, **28**, 251-261.

Isogai, T. (1983*b*). Monte Carlo study on some measures for evaluating multinormality, *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.*, **30**, 1-10.

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality, *J. Amer. Statist. Assoc.*, **68**, 176-179.

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

Radhakrishnan, R. and Kshirsagar, A. M. (1981). Influence functions for certain parameters in multivariate analysis, *Comm. Statist. A—Theory Methods*, **10**, 515-529.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Spiegelhalter, D. J. (1983). Diagnostic tests of distributional shape, *Biometrika*, **70**, 401-409.

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*, Vol. 1, 5th ed., Griffin, London.

Wilks, S. S. (1963). Multivariate statistical outliers, *Sankhyā Ser. A*, **25**, 407-426.