Ann. Inst. Statist. Math.
37 (1985), Part A, 289-301

SOME EXTENSION OF HALDANE'S MULTIVARIATE MEDIAN AND ITS APPLICATION

TAKAFUMI ISOGAI

(Received Jan. 9, 1984; revised Jan. 11, 1985)

Summary

Some extension of Haldane's multivariate median is carried out by minimization principle of a specified distance function. Then, making use of the median, three types of measures of multivariate skewness are introduced and their asymptotic null distributions are obtained.

1. Introduction

Haldane's [2] multivariate median is defined as a minimizer θ_0 of the function $\mathrm{E}\|X-\theta\|$, where $X'=(X_1,\cdots,X_p)$ is a p-variate random vector distributed according to a distribution F with mean vector μ and covariance matrix Σ and E denotes the operation of expectation in X and $\|X-\theta\|$ denotes the standardized Euclidean norm, i.e. $\{(X-\theta)'\Sigma^{-1}(X-\theta)\}^{1/2}$.

Since a median is defined, in general, as the value of θ which is closest to X, it heavily depends on a particular choice of the distance between X and θ. In this paper, we introduce a median by way of a real function ϕ satisfying the following conditions and examine its statistical properties. Let ϕ be a strictly increasing function defined on the half interval $[0,\infty)$ with $\phi(0)\geqq 0$ and such that has the continuous 2nd order derivative on $(0,\infty)$ and that $\phi(u_1+u_2)\leqq\phi(u_1)+\phi(u_2)$ for all $u_1, u_2\geqq 0$. Note that the properties imply the inequality $\phi(u)\leqq\phi(1)\cdot(u+1)$. Now we define the distance between X and θ by $\Psi(X,\theta)=\phi(\|X-\theta\|)$.

In Section 2 under several assumptions we prove the consistency and asymptotic normality of a sample median. In Section 3 we apply the sample median to a testing problem of symmetry of multivariate data and propose three types of measures of multivariate skewness. Finally in Section 4 we give a simple example in which measures of multi-

Key words and phrases: Multivariate median, multivariate skewness, test for multivariate normality.

variate skewness are used, in particular, to test multivariate normality.

2. Generalized median

Let $X_1, \cdots, X_n$ be a random sample of size n from a p-variate population F with mean vector $\mu=(\mu_1, \cdots, \mu_p)'$ and covariance matrix $\Sigma=(\sigma_{ij})$, $i, j=1, \cdots, p$. Put $\bar{\Psi}(\theta)=n^{-1}\sum_{i=1}^{n}\Psi(X_i, \theta)$ and $\bar{\Psi}_n(\theta)=n^{-1}\sum_{i=1}^{n}\Psi_n(X_i, \theta)$ where $\Psi_n(X_i, \theta)$'s are defined by replacing Σ in $\Psi(X_i, \theta)$ by its consistent estimator Σ_n calculated from the sample. Set $\Lambda(\theta)=\mathrm{E}\,\Psi(X, \theta)$. We call a minimizer θ_0 of $\Lambda(\theta)$ the (generalized) median and a minimizer θ_n of $\bar{\Psi}_n(\theta)$ the sample median. Here note that if we transform $X\rightarrow AX+b$ with a non-singular constant $p\times p$ matrix A and a constant $p\times 1$ vector b, then $\theta_0\rightarrow A\theta_0+b$.

2.1. *Weak consistency*

We put the following two assumptions:

(A1) The domain of θ, Θ_0 say, is a compact set in R^p. (If the support of F, K say, is compact, we take as Θ_0 a bounded and closed sphere B including K and if K is not compact, we choose as Θ_0 a sufficiently large, bounded and closed sphere B such that the probability of the set $\{X \in B\}$ approaches one.)

(A2) $\Lambda(\theta)$ has an unique minimizer θ_0.

First note that the property of ψ: $\psi(u)\leqq\psi(1)(u+1)$ ensures the strong consistency of $\bar{\Psi}(\theta)$ to $\Lambda(\theta)$ with each $\theta \in \Theta_0$. Then, by (A1) and from the fact that $|\Psi(X, \theta_1)-\Psi(X, \theta_2)|\leqq\Psi(\theta_1, \theta_2)$ and $|\Lambda(\theta_1)-\Lambda(\theta_2)|\leqq\Psi(\theta_1, \theta_2)$, we have

LEMMA 2.1. *For any* $\varepsilon>0$,

$$\lim_{m\to\infty} \mathrm{P}\,\{\sup_{\theta\in\Theta_0}|\bar{\Psi}(\theta)-\Lambda(\theta)|<\varepsilon, \text{ all } n\geqq m\}=1\ .$$

Next, by the consistency of Σ_n and the continuity of $\Psi(X, \theta)$ in Σ, we have

LEMMA 2.2. *For any* $\varepsilon>0$,

$$\lim_{n\to\infty} \mathrm{P}\,\{\sup_{\theta\in\Theta_0}|\bar{\Psi}_n(\theta)-\bar{\Psi}(\theta)|>\varepsilon\}=0\ .$$

Lemmas 2.1 and 2.2 ensure that $\sup_{\theta\in\Theta_0}|\bar{\Psi}_n(\theta)-\Lambda(\theta)|$ tends to zero in probability. Thus, to prove the weak consistency of θ_n, it is enough to remark that

$$\begin{aligned}(2.1)\qquad |\Lambda(\theta_n)-\Lambda(\theta_0)| &\leqq |\Lambda(\theta_n)-\bar{\Psi}_n(\theta_n)|+|\bar{\Psi}_n(\theta_n)-\Lambda(\theta_0)| \\ &\leqq 2\sup_{\theta\in\Theta_0}|\bar{\Psi}_n(\theta)-\Lambda(\theta)|\end{aligned}$$

because by (A1), on putting $c_n=\max(\sup_{\theta\in\Theta_0}\bar{\Psi}_n(\theta), \sup_{\theta\in\Theta_0}\Lambda(\theta))$, we have

$$|\bar{\Psi}_n(\theta_n)-\Lambda(\theta_0)|\leqq|\sup_\theta[c_n-\bar{\Psi}_n(\theta)]-\sup_\theta[c_n-\Lambda(\theta)]|$$
$$\leqq\sup_\theta|\bar{\Psi}_n(\theta)-\Lambda(\theta)| .$$

Then, by (2.1) and (A2), we obtain

THEOREM 2.1. *For any* $\varepsilon>0$,

$$\lim_{n\to\infty} \mathrm{P}\{|\theta_n-\theta_0|>\varepsilon\}=0 ,$$

where $|\cdot|$ *means an usual Euclidean norm, i.e.* $|a|=\left(\sum_{i=1}^p a_i^2\right)^{1/2}$ *for* $a'=(a_1,\cdots,a_p)$.

2.2. *Asymptotic normality*

To prove asymptotic normality of θ_n, we utilize the results in Huber [3]. Following the formulation in Maronna [4], we write our parameter vector as $\nu'=(\theta', \tilde{\mu}', \mathrm{vec}(\tilde{\Sigma})')$ where $\tilde{\mu}$ is an arbitrary $p\times 1$ constant vector, $\tilde{\Sigma}=(\tilde{\sigma}_{ij})$ is an arbitrary $p\times p$ positive definite matrix and $\mathrm{vec}(\tilde{\Sigma})$ denotes the $p(p+1)/2\times 1$ vector formed from the elements on the right of the diagonal and including the diagonal of $\tilde{\Sigma}$, i.e.

$$\mathrm{vec}(\tilde{\Sigma})'=(\tilde{\sigma}_{11}, \tilde{\sigma}_{12},\cdots,\tilde{\sigma}_{1p}, \tilde{\sigma}_{22},\cdots,\tilde{\sigma}_{2p},\cdots,\tilde{\sigma}_{pp}) .$$

Let $\Theta_1=R^p$, $\Theta_2=R^{p(p+1)/2}$ and $\Theta=\Theta_0\times\Theta_1\times\Theta_2$. The product set Θ is normed by an usual Euclidean norm and Θ includes the true parameter $\nu_0'=(\theta_0', \mu', \mathrm{vec}(\Sigma)')$.

In the following we deal with the sample median θ_n calculated from $\bar{\Psi}_n(\theta)$ with $\Sigma_n=n^{-1}\sum_{i=1}^n(X_i-\bar{X})(X_i-\bar{X})'$ and $\bar{X}=n^{-1}\sum_{i=1}^n X_i$.

We first put

$$\Psi_0(X,\nu)=-\|X-\theta\|^{-1}\dot{\psi}(\|X-\theta\|)(X-\theta) , \tag{2.2}$$

$$\Psi_1(X,\nu)=-(X-\tilde{\mu}) , \tag{2.3}$$

$$\Psi_2(X,\nu)=\mathrm{vec}[-(X-\tilde{\mu})(X-\tilde{\mu})'+\tilde{\Sigma}] \tag{2.4}$$

where $\|X-\theta\|=\{(X-\theta)'\tilde{\Sigma}^{-1}(X-\theta)\}^{1/2}$ and $\dot{\psi}$ denotes the first derivative of ψ. Ψ_0 is derived from the gradient vector of $\Psi(X,\theta)$ with respect to θ. Note that the estimator $\nu_n'=(\theta_n', \bar{X}', \mathrm{vec}(\Sigma_n)')$ is a solution of the systems of equations $n^{-1}\sum_{i=1}^n\Psi_0(X_i,\nu)=0$, $n^{-1}\sum_{i=1}^n\Psi_1(X_i,\nu)=0$ and $n^{-1}\sum_{i=1}^n\Psi_2(X_i,\nu)=0$ where we put $\Psi_0(X,\nu)=0$ if $X=\theta$.

Let Δ be the function defined by

$$\Delta(X, \nu)' = (\Psi_0(X, \nu)', \Psi_1(X, \nu)', \Psi_2(X, \nu)') .$$

For simplicity of the discussion, we denote the derivative of $\Delta(X, \nu)$ with respect to ν by $(D\Delta)(X, \nu)$ and write down the differential of Δ as follows:

$$\text{(2.5)} \qquad d\Psi_0(X, \nu) = [\|X-\theta\|^{-1}\dot{\psi}(\|X-\theta\|)I_p + h(\nu)\tilde{\Sigma}G(\nu)]d\theta + 2^{-1}h(\nu)\,\mathrm{tr}\,(G(\nu)d\tilde{\Sigma})(X-\theta) ,$$

$$\text{(2.6)} \qquad d\Psi_1(X, \nu) = d\tilde{\mu} ,$$

$$\text{(2.7)} \qquad d\Psi_2(X, \nu) = \mathrm{vec}\,[d\tilde{\Sigma} + d\tilde{\mu}(X-\tilde{\mu})' + (X-\tilde{\mu})d\tilde{\mu}']$$

where I_p denotes the identity matrix of order p and

$$G(\nu) = \|X-\theta\|^{-2}\tilde{\Sigma}^{-1}(X-\theta)(X-\theta)'\tilde{\Sigma}^{-1} ,$$

$$h(\nu) = \ddot{\psi}(\|X-\theta\|) - \|X-\theta\|^{-1}\dot{\psi}(\|X-\theta\|)$$

and also $\ddot{\psi}$ denotes the second derivative of ψ.

Set

$$\lambda(\nu)' = (\lambda_0(\nu)', \lambda_1(\nu)', \lambda_2(\nu)') = \mathrm{E}\,[\Delta(X, \nu)'] ,$$

$$C(\nu) = \mathrm{E}\,[\Delta(X, \nu)\Delta(X, \nu)'] = [C_{ij}(\nu)] \qquad (i, j = 0, 1, 2)$$

where $\lambda_i(\nu) = \mathrm{E}\,[\Psi_i(X, \nu)]$ and $C_{ij}(\nu) = \mathrm{E}\,[\Psi_i(X, \nu)\Psi_j(X, \nu)']$, Also we denote the derivative of $\lambda(\nu)$ with respect to ν by $(D\lambda)(\nu)$ and put $(D\lambda)(\nu) = [D_{ij}(\nu)]$ $(i, j = 0, 1, 2)$ where the partition of $D\lambda$ corresponds to its derivatives with θ, $\tilde{\mu}$ and $\mathrm{vec}\,(\tilde{\Sigma})$ in ν, for example, $D_{00}(\nu) = ((\partial/\theta_1)\lambda_0(\nu), \cdots, (\partial/\theta_p)\lambda_0(\nu))$ for $\theta' = (\theta_1, \cdots, \theta_p)$, $D_{01}(\nu) = ((\partial/\tilde{\mu}_1)\lambda_0(\nu), \cdots, (\partial/\tilde{\mu}_p)\lambda_0(\nu))$ and $D_{02}(\nu) = ((\partial/\tilde{\sigma}_{11})\lambda_0(\nu), \cdots, (\partial/\tilde{\sigma}_{1p})\lambda_0(\nu), (\partial/\tilde{\sigma}_{22})\lambda_0(\nu), \cdots, (\partial/\tilde{\sigma}_{pp})\lambda_0(\nu))$, etc.

Here we put further assumptions:

(A3) The underlying distribution F has a density with respect to Lebesgue measure and $\mathrm{E}\,|X|^4 < \infty$;

(A4) Operations of differentiation with respect to ν and expectation with respect to X are interchangeable for $\Delta(\theta)$ and $\lambda(\nu)$;

(A5) The following integrals exist, i.e.

$$\mathrm{E}\,|\dot{\psi}(\|X-\theta\|)|^2 < \infty , \qquad \mathrm{E}\,|\ddot{\psi}(\|X-\theta\|)|^2 < \infty ,$$

$$\mathrm{E}\,|\dot{\psi}(\|X-\theta\|)/\|X-\theta\||^2 < \infty$$

in some neighbourhood $B(\nu_0)$ $(\subset \Theta)$ of $\nu_0' = (\theta_0', \mu', \mathrm{vec}\,(\Sigma)')$;

(A6) $C(\nu)$ and $D_{00}(\nu)$ are nonsingular matrices for $\nu \in B(\nu_0)$.

Remark that (A4) implies $\lambda(\nu_0) = 0$ and with $(D\lambda)(\nu_0)$, D_{01}, D_{10}, D_{12}, D_{20} and D_{21} are all zero matrices and $D_{11} = I_p$ and also $D_{22} = I_{p(p+1)/2}$ where we omit the suffix ν_0 for simplicity when we evaluate the values of

$C(\nu)$ and $(D\lambda)(\nu)$ at the true parameter ν_0. The moment conditions (A3) and (A5) ensure the existence of $C(\nu)$ and $(D\lambda)(\nu)$ for $\nu \in B(\nu_0)$.

Now we check the conditions of Lemma 3 in Huber [3].

Define

$$u(X, \nu, d)=\sup_{|\nu_1-\nu|\leqq d} |\Delta(X, \nu_1)-\Delta(X, \nu)| \,.$$

and put $N(X, \nu)=\{(X-\theta)'\tilde{\Sigma}^{-1}(X-\theta)\}^{1/2}$ for $\nu'=(\theta', \tilde{\mu}', \mathrm{vec}\,(\tilde{\Sigma})') \in B(\nu_0)$.

The following lemma holds:

LEMMA 2.3. *There are some positive constants* c_i, $i=1,\cdots,6$, *which are independent of any* $\nu \in B(\nu_0)$, *such that for any* ν, ν_1 *in* $B(\nu_0)$ *and for any* $d>0$,

(i) $c_1|X-\theta|\leqq N(X, \nu)\leqq c_2|X-\theta|$;

(ii) $\sup_{|\nu_1-\nu|\leqq d} |N(X, \nu_1)-N(X, \nu)|\leqq(c_3|X-\theta|+c_4)d$;

(iii) $\sup_{|\nu_1-\nu|\leqq d} \left|\dfrac{X-\theta_1}{N(X, \nu_1)}-\dfrac{X-\theta}{N(X, \nu)}\right|\leqq\left(\dfrac{c_5}{N(X, \nu)}+c_6\right)d.$

PROOF. (i) is obvious. (ii) is obtained from (i) and the differentiability of $N(X, \nu)$ with respect to ν. (iii) is easily shown by use of (i) and (ii).

We need the further assumption:

(A7) For any ν, ν_1 in $B(\nu_0)$ and for any $d>0$,

$$\sup_{|\nu_1-\nu|\leqq d} |\psi(N(X, \nu_1))-\psi(N(X, \nu))|\leqq H(X)d \,,$$

where $H(X)$ is independent of $\nu \in B(\nu_0)$ and $\mathrm{E}|H(X)|^2<\infty$. (Note that one of the sufficient conditions for (A7) is Lipshitz' continuity of ψ.) Then

LEMMA 2.4. *Under assumptions* (A3)-(A7), *there are some positive constants* a, b, c *such that for every* ν *and* ν_1 *in* $B(\nu_0)$

(i) $|\lambda(\nu)|\geqq a|\nu-\nu_0|$;

(ii) $\mathrm{E}\,u(X, \nu, d)\leqq bd$;

(iii) $\mathrm{E}\,u(X, \nu, d)^2\leqq cd.$

PROOF. For (i), we note that

$$|\lambda(\nu)|^2=(\nu-\nu_0)'(D\lambda)(\nu^*)'(D\lambda)(\nu^*)(\nu-\nu_0)$$

where ν^* lies on the segment between ν and ν_0. Put

$$a=\{\inf_{\nu \in B(\nu_0)} r_{\min}[(D\lambda)(\nu)'(D\lambda)(\nu)]\}^{1/2}$$

where $r_{\min}[\cdot]$ means the minimum latent root of the argument matrix. For (ii) and (iii), using the results of Lemma 2.3 and (A7), we have

$$\begin{aligned} u(X, \nu, d) &\leqq \sup_{|\nu_1-\nu|\leqq d} |\dot\psi(N(X, \nu_1))-\dot\psi(N(X, \nu))| \left|\frac{X-\theta_1}{N(X, \nu_1)}\right| \\ &+ \sup_{|\nu_1-\nu|\leqq d} \left|\frac{X-\theta_1}{N(X, \nu_1)}-\frac{X-\theta}{N(X, \nu)}\right| |\dot\psi(N(X, \nu))| \\ &+ \sup_{|\nu_1-\nu|\leqq d} |\tilde\mu_1-\tilde\mu| + \sup_{|\nu_1-\nu|\leqq d} |\Psi_2(X, \nu_1)-\Psi_2(X, \nu)| \\ &\leqq [c_1^{-1}H(X)+(c_5N(X, \nu)^{-1}+c_6)|\dot\psi(N(X, \nu))|+c_7|X|+c_8]d \end{aligned}$$

where c_7 and c_8 are some positive constants independent of $\nu \in B(\nu_0)$. (ii) and (iii) are immediately verified by use of the last term in the above inequality.

Thus, by Theorem 3 in Huber [3], we have

THEOREM 2.2. *Under assumptions* (A1)-(A7), $n^{1/2}(\nu_n-\nu_0)$ *is asymptotically distributed as a* $(p+p+p(p+1)/2)$*-variate normal with mean vector* 0 *and covariance matrix* $[(D\lambda)(\nu_0)]^{-1}C(\nu_0)[(D\lambda)(\nu_0)]^{-1}$ $(=V=[V_{ij}]$ $i, j=0, 1, 2$, *say*). *Submatrices of* V *are defined by*

$$V_{00}=D_{00}^{-1}[C_{00}-D_{02}C_{20}-C_{02}D'_{02}+D_{02}C_{22}D'_{02}]D_{00}^{-1} ,$$

$$V_{01}=V'_{10}=D_{00}^{-1}[C_{01}-D_{02}C_{21}] ,$$

$$V_{02}=V'_{20}=D_{00}^{-1}[C_{02}-D_{02}C_{22}] ,$$

and

$$V_{i,j}=C_{i,j} \qquad i, j=1, 2 .$$

3. Application

First we consider some statistical interpretation of the median θ_0. Noting that $\lambda(\nu_0)=0$ under the assumptions (A1)-(A4), we have

$$\mathrm{E}\left[\frac{\Sigma^{-1/2}(X-\theta_0)}{\|X-\theta_0\|}\dot\psi(\|X-\theta_0\|)\right]=0 . \tag{3.1}$$

Since $\dot\psi$ plays a role of weighting the orientation statistic $\Sigma^{-1/2}(X-\theta_0)/\|X-\theta_0\|$, a natural choice for $\dot\psi$ is thought to be a monotonous function, i.e. $\dot\psi(u_1)\leqq\dot\psi(u_2)$ or $\dot\psi(u_1)\geqq\dot\psi(u_2)$ for $u_1\leqq u_2$. If $\dot\psi$ is nondecreasing and $\psi(0)=0$, the property of ψ: $\psi(u_1+u_2)\leqq\psi(u_1)+\psi(u_2)$ implies that $\psi(u)=\psi(1)u$, which gives us the equivalent criterion the Haldane's case. Thus we have only to consider the case that $\dot\psi$ is nonincreasing, i.e. $\ddot\psi\leqq 0$. Hence θ_0 will point a location at which the population has a high density.

3.1. *Measures of multivariate skewness*

Now we consider how to use the median θ_0 in testing symmetry of the population. The difference between μ and θ_0 gives us some information about asymmetry of data. Then, we define a Pearson-type measure of skewness by

$$(\mu-\theta_0)'\Sigma^{-1}(\mu-\theta_0) \qquad (=S_P, \text{ say}).$$

Alternatively, it is often useful to examine the configuration of data about the specified locations. Thus we check the distributions of the scaled residuals

$$e=\Sigma^{-1/2}(X-\mu) \qquad \text{and} \qquad \varepsilon=\Sigma^{-1/2}(X-\theta_0),$$

which can be converted to radius-and-angles representations

$$e=\|X-\mu\|\frac{\Sigma^{-1/2}(X-\mu)}{\|X-\mu\|} \qquad \text{and} \qquad \varepsilon=\|X-\theta_0\|\frac{\Sigma^{-1/2}(X-\theta_0)}{\|X-\theta_0\|}.$$

As formal test statistics Geary's test concerns $|e|$ or $|\varepsilon|$ and Rayleigh's test deals with $e/|e|$ or $\varepsilon/|\varepsilon|$. The difference between μ and θ_0 will reflect the basic elements of the above tests. Thus we define another measures of skewness by

$$\mathrm{E}\,[\|X-\mu\|-\|X-\theta_0\|]^2 \qquad (=S_G, \text{ say})$$

and

$$\mathrm{E}\left|\frac{\Sigma^{-1/2}(X-\mu)}{\|X-\mu\|}-\frac{\Sigma^{-1/2}(X-\theta_0)}{\|X-\theta_0\|}\right|^2 \qquad (=S_R, \text{ say}).$$

Remarks. (i) The measure S_P can be derived from the ratio of determinants of two dispersion matrices $\mathrm{E}\,(X-\mu)(X-\mu)'$ $(=\Sigma)$ and $\mathrm{E}\,(X-\theta_0)(X-\theta_0)'$. Noting that $\mathrm{E}\,(X-\theta_0)(X-\theta_0)'=\Sigma+(\mu-\theta_0)(\mu-\theta_0)'$, the ratio is written as

$$\frac{\det(\Sigma)}{\det(\Sigma+(\mu-\theta_0)(\mu-\theta_0)')}=(1+S_P)^{-1}.$$

Thus S_P measures the discrepancy between generalized variances about two locations.
(ii) For the definition of S_G, we can take another type of measure

$$\mathrm{E}\left[\frac{\|X-\theta_0\|}{\|X-\mu\|}-1\right]^2 \qquad (=S_G', \text{ say}).$$

3.2. *Asymptotic null distributions of measures of skewness*

Using a random sample $X_1,\cdots,X_n$ from the underlying p-variate distribution F having mean vector μ and covariance matrix Σ and satisfying the conditions (A1)–(A7), estimators for S_P, S_G, S_G' and S_R are represented as

$$\hat{S}_P=(\bar{X}-\theta_n)'\Sigma^{-1}(\bar{X}-\theta_n)\ ,$$

$$\hat{S}_G=(1/n)\sum_i\,[\|X_i-\theta_n\|-\|X_i-\bar{X}\|]^2\ ,$$

$$\hat{S}'_G=(1/n)\sum_i\left|\frac{\|X_i-\theta_n\|}{\|X_i-\bar{X}\|}-1\right|^2\ ,$$

$$\hat{S}_R=(1/n)\sum_i\left|\frac{\Sigma^{-1/2}(X_i-\bar{X})}{\|X_i-\bar{X}\|}-\frac{\Sigma^{-1/2}(X_i-\theta_n)}{\|X_i-\theta_n\|}\right|^2$$

where the summation is taken over all values of the index set $\{1,\cdots,n\}$, $\bar{X}=n^{-1}\sum_{i=1}^n X_i$, θ_n is the sample median calculated from $\bar{\Psi}_n(\theta)$ with $\Sigma_n=n^{-1}\sum_{i=1}^n(X_i-\bar{X})(X_i-\bar{X})'$ and for simplicity of the discussion we use Σ instead of Σ_n in the above and following expressions for $\hat{S}_P$, $\hat{S}_G$, $\hat{S}'_G$ and $\hat{S}_R$.

Note that for each i $(=1,\cdots,n)$

$$\|X_i-\theta_n\|=\|X_i-\theta_0\|\left\{1-(\theta_n-\theta_0)'\frac{\Sigma^{-1}(X_i-\theta_0)}{\|X_i-\theta_0\|^2}+O_p(n^{-1})\right\}\ , \tag{3.2}$$

$$\frac{\Sigma^{-1/2}(X_i-\theta_n)}{\|X_i-\theta_n\|}=\frac{\Sigma^{-1/2}(X_i-\theta_0)}{\|X_i-\theta_0\|}-\left[I_p-\frac{\Sigma^{-1/2}(X_i-\theta_0)(X_i-\theta_0)'\Sigma^{-1/2}}{\|X_i-\theta_0\|^2}\right]\cdot\frac{\Sigma^{-1/2}(\theta_n-\theta_0)}{\|X_i-\theta_0\|}+O_p(n^{-1})\ . \tag{3.3}$$

Replacing θ_n by $\bar{X}$ and θ_0 by μ in (3.2) and (3.3), we obtain asymptotic expansions for $\bar{X}$.

Now we put the three null conditions as follows:

(N1) $\theta_0=\mu$;

(N2) $\|X-\theta_0\|$ and $\Sigma^{-1/2}(X-\theta_0)/\|X-\theta_0\|$ are independent;

(N3) The third cumulants of X are all zero.

By (3.2), (3.3) and (N1), $\hat{S}_G$, $\hat{S}'_G$ and $\hat{S}_R$ reduce to

$$\hat{S}_G=(\bar{X}-\theta_n)'\Sigma^{-1/2}[(1/n)\sum_i A_i]\Sigma^{-1/2}(\bar{X}-\theta_n)+O_p(n^{-3/2})\ ,$$

$$\hat{S}'_G=(\bar{X}-\theta_n)'\Sigma^{-1/2}\left[(1/n)\sum_i\frac{A_i}{U_i^2}\right]\Sigma^{-1/2}(\bar{X}-\theta_n)+O_p(n^{-3/2})\ ,$$

$$\hat{S}_R=(\bar{X}-\theta_n)'\Sigma^{-1/2}\left[(1/n)\sum_i\frac{(I_p-A_i)}{U_i^2}\right]\Sigma^{-1/2}(\bar{X}-\theta_n)+O_p(n^{-3/2})$$

where

$$A=\frac{\Sigma^{-1/2}(X-\theta_0)(X-\theta_0)'\Sigma^{-1/2}}{\|X-\theta_0\|^2}\ ,\qquad U=\|X-\theta_0\|$$

and A_i and U_i are given by replacing X of A and U by X_i, $i=1,\cdots,n$ respectively. Then, by (N1) we have $\mathrm{E}[U^2]=p$ and from this fact and by (N2) it follows that $\mathrm{E}[A]=(1/p)I_p$.

Thus under (A1)–(A7) and (N1)–(N3), we obtain

$$C_{00}=\Sigma^{1/2}\,\mathrm{E}[A]\Sigma^{1/2}\,\mathrm{E}[\psi(U)]^2=(1/p)\,\mathrm{E}[\dot{\psi}(U)]^2\Sigma=a_\psi\Sigma\ , \qquad \text{say}\ ,$$

$$C_{01}=(1/p)\,\mathrm{E}[U\dot{\psi}(U)]\Sigma\ , \qquad C_{20}=C'_{02}=0\ , \qquad C_{12}=C'_{21}=0\ ,$$

$$D_{00}=\{(p-1)p^{-1}\,\mathrm{E}[U^{-1}\psi(U)]+p^{-1}\,\mathrm{E}[\ddot{\psi}(U)]\}I_p=b_\psi I_p\ , \qquad \text{say}\ ,$$

$$D_{02}=0\ ,$$

and also

$$\mathrm{var}\,(\theta_n)=n^{-1}c_\psi\Sigma+O_p(n^{-3/2})\ , \tag{3.4}$$

$$\mathrm{cov}\,(\theta_n,\bar{X})=n^{-1}d_\psi\Sigma+O_p(n^{-3/2})\ , \tag{3.5}$$

$$\mathrm{var}\,(\theta_n-\bar{X})=n^{-1}e_\psi\Sigma+O_p(n^{-3/2}) \tag{3.6}$$

where $c_\psi=a_\psi/b_\psi^2$, $d_\psi=\mathrm{E}[U\dot{\psi}(U)]/(pb_\psi)$ and $e_\psi=c_\psi-2d_\psi+1=\mathrm{E}[\dot{\psi}(U)-b_\psi U]^2/(pb_\psi^2)$ and $\mathrm{var}(\cdot)$ and $\mathrm{cov}(\cdot,\cdot)$ denote a variance and a covariance with argument vectors respectively.

Here we remark that if, under the distribution F, the random variable $U\ (=\|X-\theta_0\|)$ has a density $g(u)$ of which first derivative is continuous and $\lim_{u\to 0}\dot{\psi}(u)g(u)=\lim_{u\to\infty}\dot{\psi}(u)g(u)=0$, then b_ψ can be rewritten as

$$b_\psi=p^{-1}\,\mathrm{E}[h(U)\dot{\psi}(U)] \tag{3.7}$$

where

$$h(u)=\frac{d}{du}\left(\log\frac{u^{p-1}}{g(u)}\right)\ . \tag{3.8}$$

Hence, from the Schwartz inequality we have

$$c_\psi\geqq p/\mathrm{E}[h(U)]^2 \tag{3.9}$$

which gives the lower bound of the variance of θ_n and the equality holds if and only if $\dot{\psi}(u)$ is proportional to $h(u)$.

Finally we get the result that under the conditions (A1)–(A7) and (N1)–(N3), $(n/c_\psi)\hat{S}_P$, $p(n/c_\psi)\hat{S}_G$, $(p/\mathrm{E}[U^{-2}])(n/c_\psi)\hat{S}'_G$ and $(p-1)^{-1}(p/\mathrm{E}[U^{-2}])\cdot(n/c_\psi)\hat{S}_R$ are respectively distributed as χ^2 variables with p degrees of freedom. Remark that in the case of replacing Σ by its consistent estimator Σ_n in the preceding expressions for $\hat{S}_P$, $\hat{S}_G$, $\hat{S}'_G$ and $\hat{S}_R$ we have the same results as above because of the continuity of the relevant quantities in Σ.

4. Example

In this section we first examine the set of our distance functions and discuss a rough test procedure based on a distance function which does not satisfy (A7). Next, using the above results, we consider the use of a rough test procedure based on a typical distance function in testing multivariate normality and check the performance of it by use of well known data.

4.1. *On distance functions*

We define the set of distance functions by $G=\{\psi \mid \psi$ is nonnegative on $[0, \infty)$, $\psi(u_1+u_2)\leqq\psi(u_1)+\psi(u_2)$ for all $u_1, u_2\geqq 0$, $\dot\psi>0$ and $\ddot\psi$ is nonpositive, nondecreasing and continuous$\}$. The set G contains various functions, for example, (1) $\psi(u)=1-\exp(-u/r)$, $r>0$, (2) $\psi(u)=\log(1+u/s)$, $s>0$, (3) $\psi(u)=u^t$, $0<t\leqq 1$, etc. G is abundant in the sense that if ψ_1 and ψ_2 belong to G, then $\psi_3=\psi_1\circ\psi_2$ (defined by $\psi_3(u)=\psi_1(\psi_2(u))$) again belongs to G.

In the above examples functions (1) and (2) satisfy the assumption (A7), but (3) except for $t=1$ does not. The boundedness of $\ddot\psi(0)$ is essential to (A7). Thus we divide G into $G_B=\{\psi \mid \psi\in G, |\ddot\psi(0)|$ is bounded$\}$ and $\bar G_B=G-G_B$. Also note that for any $\psi\in\bar G_B$, there is some sequence $\{\psi_\varepsilon\}$, $\varepsilon\downarrow 0$ such that $\psi_\varepsilon\in G_B$ for all $\varepsilon>0$ and ψ_ε converges to ψ uniformly as $\varepsilon\downarrow 0$, because for given ψ we can put $\psi_\varepsilon=\psi\circ\phi_\varepsilon$ with $\phi_\varepsilon(u)=u+\varepsilon$, $\varepsilon>0$, $\phi_\varepsilon\in G_B$. For a test of symmetry of the population, this fact enables us to use test statistics, i.e. our measures of multivariate skewness corresponding to ψ ($\in\bar G_B$) in the following sense.

For any given $\psi\in\bar G_B$ we have some uniformly convergent sequence $\{\psi_\varepsilon\}$, $\varepsilon\downarrow 0$, $\psi_\varepsilon\in G_B$. Let $\theta_0(\varepsilon)$ and $\theta_n(\varepsilon)$ denote the population median and the sample median corresponding to ψ_ε respectively. Under (A1)–(A7) and (N1)–(N3) we have $\theta_0(\varepsilon)=\mu=\theta_0$ for any $\varepsilon>0$ and $a_{\psi_\varepsilon}=p^{-1}\mathrm{E}[\dot\psi_\varepsilon(U)]$, $b_{\psi_\varepsilon}=p^{-1}\mathrm{E}[h(U)\dot\psi_\varepsilon(U)]$, etc. Thus, if under (A1)–(A7) and (N1)–(N3) $a_\psi=\lim_{\varepsilon\downarrow 0} a_{\psi_\varepsilon}$, $b_\psi=\lim_{\varepsilon\downarrow 0} b_{\psi_\varepsilon}$, etc., then we can regard test statistics formally constructed in terms of ψ as the approximations to the corresponding test statistics derived from ψ_ε with sufficiently small $\varepsilon>0$.

4.2. *Test for multivariate normality*

Suppose that the underlying distribution F is a p-variate normal $N_p(\mu, \Sigma)$. In this case the density of the random variable $U=\|X-\mu\|$ is $g(u)=[2^{(p-2)/2}\Gamma(p/2)]^{-1}u^{p-1}\exp(-u^2/2)$. Thus, from (3.8), $h(u)=u$. Also, from (3.5) and (3.7) we have $d_\psi=1$, which implies asymptotic independence between $\bar X$ and $\theta_n-\bar X$ for any choice of $\psi\in G_B$.

Now we take $\psi(u)=u^t$, $0<t\leqq 1$ as a distance function because of its easy feasibility in data analysis and examine the performance of our measures of skewness derived from ψ. This distance function is the limit of a uniformly convergent sequence $\{\psi_\varepsilon\}$, $\varepsilon\downarrow 0$, $\psi_\varepsilon(u)=(u+\varepsilon)^t$, $\psi_\varepsilon \in G_B$. Under the null hypothesis $a_\psi=\lim_{\varepsilon\downarrow 0} a_{\psi_\varepsilon}$, $b_\psi=\lim_{\varepsilon\downarrow 0} b_{\psi_\varepsilon}$, etc., and so

$$c_\psi=\Gamma\left(\frac{1}{2}p+1\right)\Gamma\left(\frac{1}{2}p+t-1\right)\Big/\left\{\Gamma\left[\frac{1}{2}(p+t)\right]\right\}^2=\lim_{\varepsilon\downarrow 0} c_{\psi_\varepsilon}\ .$$

Note that c_ψ is strictly decreasing in t $(0<t\leqq 1)$ and from (3.9) it is seen that $c_\psi>1$. By the assumption (A5), the range of dimension p is restricted as $p>2(2-t)$ for $0<t<1$. Therefore, with any t, $0<t<1$, the sufficient condition for p is $p\geqq 4$.

Let us consider the well known iris data in Fisher [1] and examine multivariate skewness of the data on iris setosa, which comprise 4 measurements on each of 50 plants. We take 4 measurements as variables, which are x_1=sepal length, x_2=sepal width, x_3=petal length and x_4=petal width.

For each t, $0<t\leqq 1$, Table 1 shows the estimated median θ_n and estimated measures $\hat{S}_P$, $\hat{S}_G$, $\hat{S}_R$ under which we also put the values of the corresponding test statistics.

Table 1. Analysis of the data on Iris setosa.
Upper values, medians and measures $\hat{S}_P$, $\hat{S}_G$ and $\hat{S}_R$.
Lower values, the corresponding test statistics.

The value of t	Median				Measures of skewness		
	$x_1\times 10^{-1}$	$x_2\times 10^{-1}$	$x_3\times 10^{-1}$	x_4	$\hat{S}_P$	$\hat{S}_G\times 10$	$\hat{S}_R\times 10$
1.0	.5000	.3406	.1458	.2290	.0311	.0858	.1208
					11.82	13.02	12.23
.8	.5004	.3403	.1458	.2243	.0537	.1478	.2140
					12.95	14.26	13.76
.5	.5013	.3399	.1460	.2184	.0973	.2706	.4027
					12.78	14.21	14.10
.3	.5030	.3406	.1466	.2139	.1506	.4215	.6445
					13.47	15.08	15.37
.1	.5034	.3405	.1470	.2089	.2027	.5585	.9076
					12.35	13.60	14.74

From Table 1 we know that there exists a strong evidence of non-normality for each t, because 5% point of a χ^2 variable with 4 degrees of freedom is 9.48773. We also note that as t decreases, the values of $\hat{S}_P$, $\hat{S}_G$, $\hat{S}_R$ increase. This fact indicates the feature that in the presence of non-normality the median is inclined to depart from the sample mean for small values of t.

Table 2. Analysis of the transformed data.
Upper values, measures $\hat{S}_P$, $\hat{S}_G$ and $\hat{S}_R$.
Lower values, the corresponding test statistics.

The value of t	Measures on					
	x_1, x_2, x_3 and $\log x_4$			$\log x_i$, $i=1, 2, 3, 4$		
	$\hat{S}_P$	$\hat{S}_G\times 10$	$\hat{S}_R\times 10$	$\hat{S}_P$	$\hat{S}_G\times 10$	$\hat{S}_R\times 10$
1.0	.0100	.0291	.0425	.0108	.0320	.0430
	3.81	4.42	4.30	4.10	4.86	4.35
.8	.0199	.0589	.0856	.0221	.0660	.0907
	4.82	5.68	5.51	5.33	6.37	5.83
.5	.0492	.1441	.2234	.0507	.1515	.2232
	6.46	7.57	7.82	6.66	7.95	7.81
.3	.0787	.2307	.3704	.0887	.2598	.4084
	7.04	8.25	8.83	7.93	9.29	9.74
.1	.1132	.3282	.5524	.1245	.3639	.6010
	6.89	7.99	8.97	7.58	8.86	9.76

The marginal analysis for each variable (see Small [6]) leads us to transform $x_4 \rightarrow \log x_4$. The result for the data on x_1, x_2, x_3, $\log x_4$ is shown in Table 2. There still exists some evidence for the deviation from normality. In Table 2 we also present the result of the data on all transformed variables $\log x_i$, $i=1, \cdots, 4$. Comparing these cases, we see that the unnecessary data transformation may make the data dirty.

Finally we remark that in the above analyses the sample medians were calculated, using the sample mean as the starting point, by Algorithm AS 47 provided by O'Neill [5].

Acknowledgement

The author would like to express his thanks to the referee, the editor and Mr. Y. Kano for useful comments on an earlier draft of the present paper.

OSAKA UNIVERSITY

REFERENCES

[1] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems, *Ann. Eugenics*, **VII**, 179-188.

[2] Haldane, J. B. S. (1948). Note on the median of a multivariate distribution, *Biometrika*, **35**, 414-415.

[3] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, **1**, 221-233, Univ. of California Press.

[4] Maronna, R. A. (1976). Robust *M*-estimators of multivariate location and scatter, *Ann. Statist.*, **4**, 51-67.

[5] O'Neill, R. (1971). Function Minimization using a Simplex Procedure, *Appl. Statist.*, **20**, 338-345.

[6] Small, N. J. H. (1980). Marginal Skewness and Kurtosis in Testing Multivariate Normality, *Appl. Statist.*, **29**, 85-87.