<Research Paper>

Journal of the Korean Statistical Society
Vol. 16, No. 1, 1987

On The Condition That Two Hyper-ellipsoids Have No Points in Common

Seong-Ju Kim*

ABSTRACT

The condition that two hyper-ellipsoids have no points in common is derived using the simultaneous diagonalization of the two hyper-ellipsoids. It is observed that the simultaneous diagonalization is composed of rotation and extension followed by another rotation. An approximation to this condition in terms of the generalized distance is discussed.

1. Introduction

Consider the distance from x to a in the p dimensional space. The Euclidean distance from x to a is defined, as usual, by

$$|x-a|=[(x-a)^T\ (x-a)]^{1/2}$$

As it happens, in the multivariate case where each variable has different variability, it is more reasonable to define the distance from x to a with respect to some $p\times p$ symmetric positive definite matrix A as

$$G(x,a,A)=(x-a)^T\ A^{-1}\ (x-a).$$

We will call $G(x,a,A)$ the generalized distance from x to a with respect to A. The Euclidean square distance from x to a is a special case of $G(x,a,A)$ when A is the identity matrix. Various kinds of the generalized distance have been used in the multivariate analysis.

For example, consider two independent p dimensional random samples as follows. $X_1, X_2, \cdots, X_{r_1}$, from the population with mean vector μ_1 and covariance matrix Σ_1;

* Department of Statistics, Sung Kyun Kwan University. Chongro-Ku, Seoul 110

$Y_1, Y_2, \cdots, Y_{n_2}$ from the population with mean vecter μ_2 and covariance matrix Σ_2. Define the sample mean vectors and sample covariance matrices as

$$\overline{X}=\sum_i X_i/n_1,\ S_1=\sum_i (X_i-\overline{X})\ (X_i-\overline{X})^T/(n_1-1),$$

$$\overline{Y}=\sum_j Y_j/n_2,\ S_2=\sum_j (Y_j-\overline{Y})\ (Y_j-\overline{Y})^T/(n_2-1)$$

and denote

$$\tilde{S}_1=S_1/n_1,\ \tilde{S}_2=S_2/n_2,$$

$$S=[(n_1-1)\ S_1+(n_2-1)S_2]/(n_1+n_2-2).$$

Suppose we want to make inferences for the multivariate normal population mean vectors μ_1, μ_2 and the difference $\mu_1-\mu_2$. Then Hotelling's T^2 for μ_1 and μ_2 are expressed as $G(\overline{X}, \mu_1, \tilde{S}_1)$ and $G(\overline{Y}, \mu_2, \tilde{S}_2)$ respectively. Under the assumption of $\Sigma_1=\Sigma_2$, the sample version of Mahalanobis distance between μ_1 and μ_2 is expressed as $G(\overline{X}, \overline{Y}, S)$ in our context. Suppose we want to classify a future observation x between the two populations. One way to handle this problem is to classify x according as

$$G(\overline{X}, x, S_1)>G(\overline{Y}, x, S_2).$$

We refer to Gnanadesikan(1977) for details concerning the distance measures.

From the geometrical point of view, the Euclidean square distance from x to a is the distance measured by the family of spheres centered at a. On the other hand, the generalized distance from x to a is the distance measured by the family of hyper-ellipsoids centered at a. For the two solid spheres, it is clear when they have no points in common. But it is not known yet for the two hyper-ellipsoids. Therefore it is quite interesting to inquire about the condition that two hyper-ellipsoids have no points in common.

We derive the exact condition using the simultaneous diagonalization of the two hyper-ellipsoids in Section 2. It is observed that the simultaneous diagonalization is composed of rotation and extension followed by another rotation.

An approximation to this condition in terms of the generalized distance is discussed in Section 3 followed by concluding remarks in Section 4.

2. Main results

Suppose A and B are $p\times p$ symmetric positive definite matrices. Without loss of generality, two hyper-ellipsoids centered at a and b in R^p are expressed by

$$\begin{cases} (x-a)^T A^{-1} (x-a) \leqq 1 \\ (x-b)^T B^{-1} (x-b) \leqq 1 \end{cases} \tag{1}$$

Consider the generalized eigenvalue problem of solving for d_k in

$$Ax = d_k(Bx)$$

for $k=1, 2, \cdots, p$. Using the simultaneous diagonalization theorem, Noble and Daniel (1977), a pair of symmetric positive definite matrices A and B are simultaneously diagonalized such that

$$Q^T AQ = D \text{ and } Q^T BQ = I \tag{2}$$

where Q is a nonsingular matrix, D is a diagonal matrix with diagonal elements

$$0 \leqq d_1 \leqq d_2 \leqq \cdots \leqq d_p$$

and I is the $p \times p$ identity matrix.

Let $\tau : R^p \rightarrow R^p$ be the transformation defined by

$$y = \tau(x) = Q^T(x-a)$$

and denote

$$m = \tau(b)$$

for the given value of b. By the transformation $y=\tau(x)$, the two hyper-ellipsoids in (1) can be expressed as

$$\begin{cases} (y-m)^T (y-m) \leqq 1 \\ y^T D^{-1} y \leqq 1. \end{cases} \tag{3}$$

We then notice that (3) are the unit solid sphere centered at m and the diagonalized hyper-ellipsoid at the origin. We will denote them S (m, 1) and E respectively. Since τ is a nonsingular transformation, we will inquire about the condition using (3) in the transformed space. This is simpler than using (1) in the original space.

Here we will introduce a notion of parallel body and refer to Valentine(1964) for details concerning the geometry of convex sets. The parallel body E_1 of the diagonalized hyper-ellipsoid E is defined as

$$E_1 = \bigcup_{x \in E} S(x, 1)$$

where $S(x, 1)$ denote the unit solid sphere centered at x. The following theorem presents the required condition using (3) in the transformed space.

Theorem 1. The following (i), (ii) and (iii) are all equivalent.

(i) $S(m, 1)$ and E in (3) have no points in common.

(ii) $m \notin E_1$ (4)

(iii) $m^T[I-(I+kD^{-1})^{-1}]^2 m > 1$ (5)

whers k is a positive root of

$$m^T[D(I+kD^{-1})^2]^{-1}m=1$$

Proof: We will show the following implications.

(i)$\leftrightarrow$(ii). Imagine the unit solid sphere whose center is rolling on the surface of E. From the definition of the parallel body of E_1, $m\not\in E_1$ if and only if

$$S(m,1)\cap E=\phi$$

(ii)$\leftrightarrow$(iii). It is observed that $m\not\in E_1$ if and only if

$$(m-y^*)^T(m-y^*)>1 \text{ and } m\not\in E$$

where y^* denote the foot of m on E. (See Figure 1 in the Appendix.) Since the normal direction at y^* is $D^{-1}y^*$, $m\not\in E$ is equivalent to the fact that there exist some positive constant k such that

$$m=y^*+k(D^{-1}y^*).$$

If we eliminate y^* using the fact that y^* is on E, then the result follows.

Essentially the equivalence of (ii) and (iii) can be proved using Lagrange's method. But our approach is more appealing to the geometrical intuition. ▩

3. An approximation to the condition

While the Theorem 1 provides the exact condition that two hyper-ellipsoids have no points in common in the transformed space, neither (4) nor (5) is not convenient to use. Therefore we need to approximate the condition in the previous theorem. What does the parallel body E_1 look like? Since the parallel body E_1 was constructed by expanding the diagonalized hyper-ellipsoid E in its unit normal direction, it is convex and has intercept $\sqrt{d_i}+1$ for $i=1,2,\cdots,p$. Thus it is quite natural to approximate E_1 by the hyper-ellipsoid E^* which has the largest volume as a subset of E_1.

The hyper-ellipsoid E^* we are looking for is

$$y^T(D^{1/2}+I)^{-2}y\leq 1$$

because it has the same intercepts as E_1 and because the spherical parameterization shows $E^*\subseteq E_1$. Note that E^* and E_1 are identical if E is a solid sphere. Figure 2 in the Appendix was drawn from 3000 points esch on E, E_1, E^* and $S(m,1)$ when $d_1=1$, $d_1=16$ and $m=[2.8, 0.7]$. We can see that E^* and E_1 are quite close enough even if there is a big difference in the two eigenvalues. If we approximate E_1 by E^* the following corollary gives an approximate condition in the original space.

Corollary 1. $m \notin E^*$ if and only if

$$(b-a)^T (A^*)^{-1} (b-a) > 1 \tag{6}$$

in the original space where

$$A^* = A + B + 2B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2}. \tag{7}$$

Proof. First we observe that $m \notin E^*$ if and only if

$$m^T(D^{1/2}+I)^{-2}\ m = (b-a)^T\ [(Q^{-1})^T(D+I+2D^{1/2})Q^{-1}]^{-1}(b-a) > 1.$$

From (2), it follows that

$$(Q^{-1})^T DQ^{-1} = A \text{ and } (Q^{-1})^T Q^{-1} = B.$$

Therefore it suffices to show that

$$(Q^{-1})^T D^{1/2} Q^{-1} = B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2}.$$

Consider the spectral decomposition of

$$B = PLP^T$$

followed by another spectral decomposition of

$$(L^{-1/2}P^T)A(PL^{-1/2}) = OD^*O^T$$

where O, P are $p \times p$ orthogonal matrices and D^*, L are $p \times p$ diagonal metrices. Then D and Q in (2) are equivalent to

$$D = D^* \text{ and } Q = PL^{-1/2}O.$$

It is observed that Q is the simultaneous operation consisting of rotation and extension followed by another rotation. From the fact that

$$B^{-1/2}AB^{-1/2} = PODO^TP^T,$$

we have

$$\begin{aligned}(Q^{-1})^T D^{1/2} Q^{-1} &= (PL^{1/2}P^T)(PODO^TP^T)^{1/2}(PL^{1/2}P^T)\\ &= B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2}.\end{aligned}$$

Thus the proof is completed. ▨

Since A^* in (7) is symmetric and positive definite, the condition (6) in the corollary can be expressed in terms of the generalized distance from b to a with respect to A^*. That is approximately the two hyper-ellipsoids in (1) have no points in common if $G(b, a, A^*) > 1$.

4. Concluding remarks

As an application of the results to the multivariate analysis, consider the multivariate

Behrens-Fisher problem testing the hypothesis $H; \mu_1 = \mu_2$ where μ_1 and μ_2 donote the mean vectors of two multivariate normal populations. Here any assumption about the population covariance matrices is not made. There are huge literature on this historic problem. However an appropriate solution is not available yet as discussed by Subrahmaniam and Subrahmaniam (1973). From the geometrical point of view of distance, we will think of this problem as follows.

Suppose two families of confidence ellipsoids for μ_1 and μ_2 are constructed from Hotelling's T^2, varying the confidence coefficients α_1 and α_2 respectively. It is intuitively conceivable that the two confidence ellipsoids have no points in common for reasonable α_1 and α_2 if μ_1 is really different from μ_2. Suppose the new test procedure is to reject the hypothesis H if two confidence ellipsoids have no points in common for given α_1 and α_2. If α_1 and α_2 are chosen such that the rejection probability is α under the hypothesis H, the new test procedure is a level α test. Since the power of the new test procedure depends on α_1 and α_2 as well, α_1 and α_2 are to be chosen carefully. A more rigorous development can be found in a technical report by Kim (1987).

At the beginning we started to think about the condition just for fun. We thought that the exact condition could be expressed in terms of the generalized distance before the concept of parallel body was introduced. The hyper-ellipsoid must be an ideal shape next to the sphere, but there is a big difference between the best and the second best. That might be the reason why the Euclidean distance is so popular.

Appendix

Figure 1

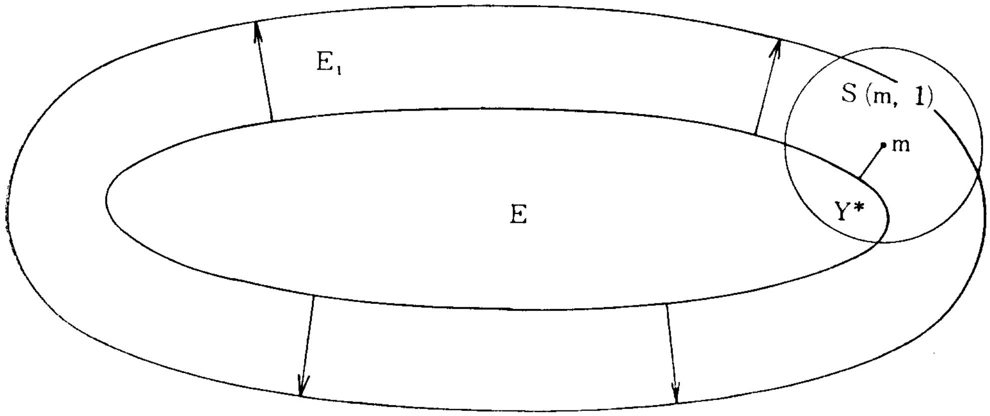
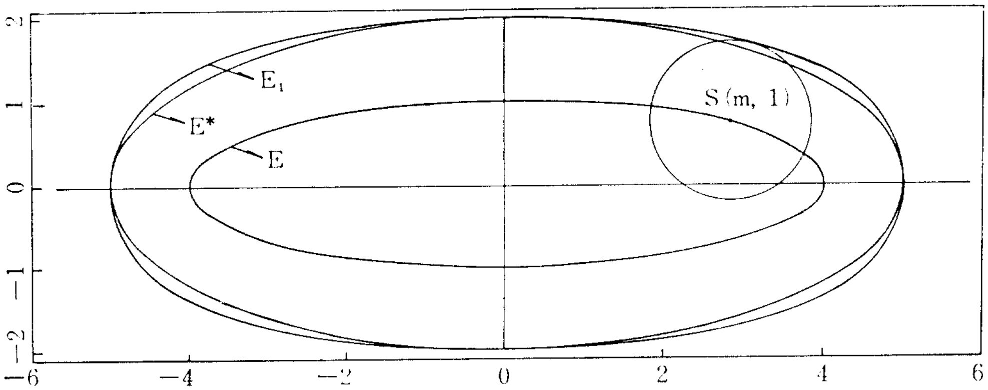


Figure 2



References

(1) Gnanadesikan, R.(1977). *Methods for Statistical Data Aanalysis of Multivariate Observations.* John Wiley and Sons.

(2) Kim, S-J. (1987). A Practical solution to the Multivariate Behrens-Fisher Problem. *Technical Report.* Sung Kyun Kwan University.

(3) Noble, B. and Daniel, J.W. (1977). *Applied Linear Algebra (Second Ed.).* Prentice Hall.

(4) Subrahmaniam, K. and Subrahmaniam, K. (1973). On the Multivariate Behrens-Fisher Problem. *Biometrika,* Vol. 60, 107~111.

(5) Valentine, F.A.(1964). *Convex Sets.* McGraw-Hill.