Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 12132–12136, October 1996
Statistics

This contribution is part of a special series on Inaugural Articles by members of the National Academy of Sciences elected on April 25, 1995.

Statistical analysis of shape of objects based on landmark data

(shape/size/Mahalanobis distance/principal component analysis/canonical correlations)

C. RADHAKRISHNA RAO† AND SHAILAJA SURYAWANSHI‡

†Statistics Department, Pennsylvania State University, University Park, PA 16802; and ‡Merck Research Laboratories, Rahway, NJ 07065

Contributed by C. Radhakrishna Rao, September 5, 1996

ABSTRACT **Two objects with homologous landmarks are said to be of the same shape if the configurations of landmarks of one object can be exactly matched with that of the other by translation, rotation/reflection, and scaling. The observations on an object are coordinates of its landmarks with reference to a set of orthogonal coordinate axes in an appropriate dimensional space. The origin, choice of units, and orientation of the coordinate axes with respect to an object may be different from object to object. In such a case, how do we quantify the shape of an object, find the mean and variation of shape in a population of objects, compare the mean shapes in two or more different populations, and discriminate between objects belonging to two or more different shape distributions. We develop some methods that are invariant to translation, rotation, and scaling of the observations on each object and thereby provide generalizations of multivariate methods for shape analysis.**

1. Introduction

We consider objects that are characterized by the configuration of some recognizable landmarks, say k in number. The observations on a p-dimensional object are represented by a $p \times k$ matrix X, where the i-th column gives the coordinates of the i-th landmark with respect to a set of coordinate axes. The location and orientation of any particular object with reference to the coordinate axes may differ from object to object so that the observations made on different objects are not comparable. We shall also allow for variations in the units of the coordinate axes from object to object. In such a case, we can compare the objects only in shape, i.e., after filtering out the differences in location, scaling, and rotation/reflection. A convenient way of doing this is to consider the maximal number of functions of X, which are invariant under transformations of the type

$$\lambda R(X - \boldsymbol{\alpha}\mathbf{1}^T), \qquad [1.1]$$

$\forall\ \lambda > 0$, $\boldsymbol{\alpha} \in \mathfrak{R}^p$ and $R \in O(p)$, the set of orthogonal matrices. Such functions may be described as shape coordinates. When $p = 2$, there are $2k - 4$ such functions that can be represented as a point in $\mathfrak{R}^{2k-4}$, and when $p = 3$, the number of such functions is $3k - 7$. There is no unique way of choosing these functions. However, the inference based on a particular choice will be consistent with that based on any other choice provided the probability distribution of any chosen set of functions can be accurately specified.

There is considerable literature on the analysis of shape coordinates, starting with the seminal work of Kendall (1–3) and Bookstein (4–6). For recent work by Mardia, Dryden, Goodall, Kent, and others the reader is referred to the survey papers by Kendall (3) and Kent (7) and the references contained therein.

Another way of specifying an object through landmarks is to provide the $k(k - 1)/2$ Euclidean distances between all possible pairs of landmarks. The distances are invariant for change in location and rotation. They can be represented by a $k \times k$ symmetric matrix $D = (D_{ij})$, where D_{ij} is the distance between the landmarks i and j. In such a case, we need only make adjustments for scale to compare objects. Writing the entries above the diagonal in D in a vector form $\mathbf{d}_*$ the object can be represented by a ray

$$\mathbf{d}_*^{(s)} = \{\lambda \mathbf{d}_* : \lambda > 0\}, \qquad [1.2]$$

which represents the shape of the object. The appropriate statistical methodology based on $\mathbf{d}_*^{(s)}$ is developed by Lele in a series of papers (see refs. 8 and 9 and references therein).

In our approach we consider the vector $\mathbf{d}$ of the logarithms of the individual components of $\mathbf{d}_*$, in which case the shape of the object is characterized by the set

$$\mathbf{d}^{(s)} = \{\mathbf{d} + c\mathbf{1} : c \in \mathfrak{R}\}.$$

We develop the appropriate statistical methodology based on $\mathbf{d}^{(s)}$, which seems to have some advantages over the earlier approaches.

2. Size and Shape Variables

Let $\mathbf{d}$ be the vector of logs of all possible, or a subset of, distances between landmarks and denote by m, the size of the vector $\mathbf{d}$. Any function of $\mathbf{d}$, which is invariant for translations of each component of $\mathbf{d}$ by a constant, is a function of the set

$$\mathbf{d}^{(s)} = \left\{ \mathbf{a}^T\mathbf{d} = \sum_{i=1}^{m} a_i d_i : \mathbf{a}^T\mathbf{1} = 0,\ \mathbf{a}^T\mathbf{a} = 1 \right\}. \qquad [2.1]$$

Note that $\exp(\mathbf{a}^T\mathbf{d})$ is a function of the ratios of the distances and, as such, the elements in Eq. **2.1** represent the shape of an object.

A basis of the set Eq. **2.1** is $H\mathbf{d}$, where H is an $m - 1 \times m$ matrix of rank $(m - 1)$, such that $H\mathbf{1} = \mathbf{0}$. We may consider the shape variables as

$$\mathbf{d}^{(s)} = H\mathbf{d}. \qquad [2.2]$$

Having defined shape variables unambiguously, we now look for a suitable characterization of the size of an object. Generally, size is defined as any function $f(\mathbf{d})$ such that $f(\mathbf{d} + c\mathbf{1}) = c + f(\mathbf{d})$, $\forall\ c \in \mathfrak{R}$, which in terms of the original distances can be written in the form $g(\lambda D) = \lambda g(D)\ \forall\ \lambda > 0$. There is no unique choice of f or g unless we impose some restrictions or require the function to have some desired properties. We suppose that $\mathbf{d}$ has a distribution with mean $\boldsymbol{\delta}$ and variance covariance matrix Σ.

(*i*) Consider a linear function $\mathbf{b}^T\mathbf{d}$ as a measure of size, which imposes the condition

$$\mathbf{b}^T(\mathbf{d} + c\mathbf{1}) = \mathbf{b}^T\mathbf{d} + c \Rightarrow \mathbf{b}^T\mathbf{1} = 1. \qquad [2.3]$$

If we require $\mathbf{b}^T\mathbf{d}$ to be uncorrelated with the shape variables, then

$$\mathrm{Cov}(\mathbf{b}^T\mathbf{d}, H\mathbf{d}) = \mathbf{b}^T\Sigma H^T = \mathbf{0} \Rightarrow \mathbf{b}^T\Sigma = \alpha\mathbf{1}^T \text{ or } \mathbf{b} = \alpha\Sigma^{-1}\mathbf{1},$$

where α is a constant. Using the condition Eq. **2.3**, $\alpha^{-1} = \mathbf{1}^T\Sigma^{-1}\mathbf{1}$ so that the required linear function is $\mathbf{1}^T\Sigma^{-1}\mathbf{d}/\mathbf{1}^T\Sigma^{-1}\mathbf{1}$.

(*ii*) Consider a linear function $\mathbf{b}^T\mathbf{d}$ as in Eq. **2.3**. The regression of $\mathbf{d}$ on $\mathbf{b}^T\mathbf{d}$ is

$$\frac{\mathrm{Cov}(\mathbf{d},\mathbf{b}^T\mathbf{d})}{\mathrm{Var}(\mathbf{b}^T\mathbf{d})} = \frac{\Sigma\mathbf{b}}{\mathbf{b}^T\Sigma\mathbf{b}}, \qquad [2.4]$$

where the vector on the right hand side of Eq. **2.4** represents the average increase (or decrease) in each of the variables for a unit increase in $\mathbf{b}^T\mathbf{d}$. We may characterize $\mathbf{b}^T\mathbf{d}$ as a measure of size if the vector in Eq. **2.4** has all positive elements. If we choose all the elements of Eq. **2.4** to be equal then

$$\frac{\Sigma\mathbf{b}}{\mathbf{b}^T\Sigma\mathbf{b}} \propto \mathbf{1} \text{ or } \mathbf{b} \propto \Sigma^{-1}\mathbf{1}. \qquad [2.5]$$

Using the condition $\mathbf{b}^T\mathbf{1} = 1$, we have $\mathbf{b} = \Sigma^{-1}\mathbf{1}/\mathbf{1}^T\Sigma^{-1}\mathbf{1}$, which is the same vector as that derived in (*i*).

(*iii*) Let $\mathbf{d}_1$ and $\mathbf{d}_2$ be two vectors associated with two individuals. Then the square of the Mahalanobis distance between the individuals admits the decomposition

$$\begin{aligned}(\mathbf{d}_1 - \mathbf{d}_2)^T\Sigma^{-1}(\mathbf{d}_1 - \mathbf{d}_2) &= (\mathbf{d}_1 - \mathbf{d}_2)^T H^T \\ &(H\Sigma H^T)^{-1}H(\mathbf{d}_1 - \mathbf{d}_2) \\ &+ \frac{[\mathbf{1}^T\Sigma^{-1}(\mathbf{d}_1 - \mathbf{d}_2)]^2}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}},\end{aligned} \qquad [2.6]$$

using the identity in ref. 10, p. 77,

$$\Sigma^{-1} = H^T(H\Sigma H^T)^{-1}H + \Sigma^{-1}\mathbf{1}(\mathbf{1}^T\Sigma^{-1}\mathbf{1})^{-1}\mathbf{1}^T\Sigma^{-1}. \qquad [2.7]$$

The first part of Eq. **2.6** is the square of the Mahalanobis distance in shape and the second may be interpreted as the square of the Mahalanobis distance in size. This again leads to the function $\mathbf{1}^T\Sigma^{-1}\mathbf{d}/\mathbf{1}^T\Sigma^{-1}\mathbf{1}$ as an indicator of size.

(*iv*) One requirement which biometricians seem to prefer is that size should be stochastically independent of shape. The problem then is to determine a function $f(\mathbf{d})$, such that $f(\mathbf{d} + c\mathbf{1}) = c + f(\mathbf{d})$ and is distributed independently of the shape variables $H\mathbf{d}$. It has been shown by Sampson and Siegel (11) that among linear functions of size, $\mathbf{1}^T\Sigma^{-1}\mathbf{d}/\mathbf{1}^T\Sigma^{-1}\mathbf{1}$ is the unique function that is independent of $H^T\mathbf{d}$ when $\mathbf{d}$ has the multivariate normal distribution with the variance covariance matrix Σ. In the appendix to this paper, we show that this result is true without the assumption that f is linear.

3. Statistical Analysis of Shape

Consider the full vector $\mathbf{d}$ of the logs of all possible $m = k(k-1)/2$ Euclidean distances, and let $\mathbf{d}_1, \ldots, \mathbf{d}_n$ be the observations (values of $\mathbf{d}$) on n objects of a population. We denote the sample mean and covariance matrix by

$$\bar{\mathbf{d}} = n^{-1}(\mathbf{d}_1 + \ldots + \mathbf{d}_n) \qquad [3.1]$$

and

$$S = (n-1)^{-1}(\mathbf{d}_1\mathbf{d}_1^T + \ldots + \mathbf{d}_n\mathbf{d}_n^T - n\bar{\mathbf{d}}\bar{\mathbf{d}}^T), \qquad [3.2]$$

which constitute the summary statistics of a sample on which further calculations are based. If we are working with a selected subset of distances the relevant components in Eqs. **3.1** and **3.2** are chosen.

3.1. Mean Shape. A typical component of $\mathbf{d}_r$ is $\log D_{ij}^{(r)}$, the log of the distance between the landmarks i and j of object r, and the corresponding component in $\bar{\mathbf{d}}$ is

$$n^{-1}\sum_{r=1}^{n} \log D_{ij}^{(r)} = \log\hat{D}_{ij} \text{ say,}$$

so that $\hat{D}_{ij}$ is the geometric mean of $D_{ij}^{(r)}$, $r = 1, \ldots, n$. The geometric mean distance $\hat{D}_{ij}$, $i, j = 1, \ldots, k$, may not be embeddable in a Euclidean space of the same dimension as the objects are. For graphical representation of shape, we may do a metrical scaling in the required dimension using the method proposed by Torgerson (12). For details of this method and some modifications reference may be made to Rao (ref. 13, section 14). We call the distances calculated from the configuration obtained by metric scaling the regularized mean distances between landmarks, and denote them by $\tilde{D}_{ij}$. The ratios of $\hat{D}_{ij}$ are independent of the sizes of the individual objects, i.e., we get the same ratios if instead of $D_{ij}^{(r)}$ we consider $\lambda_r D_{ij}^{(r)}$ as the distances in the r-th object for any arbitrary λ_r. The same is true of $\tilde{D}_{ij}$s. In this sense, the distances $\hat{D}_{ij}$ or $\tilde{D}_{ij}$ provide a characterization of the mean shape of objects.

There are other ways of defining mean shape conforming to the dimensions of the objects. Let

$$\mathbf{d}_r = (d_{r1}, \ldots, d_{rm})^T, r = 1, \ldots, n, \qquad [3.3]$$

with $m = k(k-1)/2$, and consider a $p \times k$ matrix M (i.e., of the same order as the configuration of an individual object) with the associated m vector of log distances

$$\mathbf{d}_M = (d_{M1}, \ldots, d_{Mm})^T. \qquad [3.4]$$

Further, let $w_1, \ldots, w_m$ be non-negative weights adding up to unity, and denote

$$\bar{d}_r = \sum w_j d_{rj}, \ \bar{d}_M = \sum w_j d_{Mj}. \qquad [3.5]$$

We define the mean configuration as

$$M_* = \arg\min_M \sum_{j=1}^{m} w_j \sum_{r=1}^{n} [(d_{rj} - \bar{d}_r) - (d_{Mj} - \bar{d}_M)]^2. \qquad [3.6]$$

Another possibility is to work directly with distances instead of their logarithms. Representing the actual distances in Eqs. **3.3–3.6** with an asterisk, we may define the mean configuration as

$$M_* = \arg \min_{M,\lambda_1,\ldots,\lambda_n} \sum_{r=1}^{n}\sum_{j=1}^{m} w_j(d_{*Mj} - \lambda_r d_{*rj})^2, \qquad [3.7]$$

subject to the condition $\sum w_j d_{*Mj}^2 = 1$. The expression in Eq. **3.7** can be simplified to

$$M_* = \underset{M}{\operatorname{argmax}} \sum_{r=1}^{n} \frac{(\sum w_j d_{*rj} d_{*Mj})^2}{(\sum w_j d_{*rj}^2)}. \quad [3.8]$$

Suitable algorithms have to be developed to solve the optimization problems in Eqs. **3.6** and **3.8**.

3.2. Comparison of Two Populations in Size and Shape. Let $\bar{\mathbf{d}}_i$ and S_i be the statistics in Eqs. **3.1** and **3.2** based on a sample of size n_i from population i, $i = 1, 2$. We wish to test the hypotheses that the populations are the same in size and in shape with possible differences in size. Denote $S = [(n_1 - 1)S_1 + (n_2 - 1)S_2]/(n_1 + n_2 - 2)$. Then the overall Mahalanobis distance between the populations has the decomposition

$$\begin{aligned} D_0^2 &= (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2)^T S^{-1} (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2) \\ &= (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2)^T H^T (HSH^T)^{-1} H (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2) + \frac{[\mathbf{1}^T S^{-1} (\bar{\mathbf{d}}_1 - \bar{\mathbf{d}}_2)]^2}{\mathbf{1}^T S^{-1} \mathbf{1}} \\ &= D_{sh}^2 + D_{si}^2, \end{aligned} \quad [3.9]$$

where H is an $m - 1 \times m$ matrix of rank $(m - 1)$ such that $H\mathbf{1} = \mathbf{0}$. The statistic D_{si}^2 reflects the difference in size and

$$X_{si}^2 = \frac{n_1 n_2}{n_1 + n_2} D_{si}^2 \quad [3.10]$$

is asymptotically distributed as χ^2 on 1 degree of freedom under the null hypothesis of no difference in size. The statistic D_{sh}^2 reflects differences in shape, and

$$X_{sh}^2 = \frac{n_1 n_2}{n_1 + n_2} D_{sh}^2 \quad [3.11]$$

is asymptotically distributed as χ^2 on $(m - 1)$ degrees of freedom under the null hypothesis of no difference in shape, where m is the number of distances chosen.

As an example, we consider the differences in size and shape of the cranii of chimpanzee and gorilla based on a selection of 13 distances out of $8(8 - 1)/2$ possible distances. The data were collected by Paul Higgins (University College, London). The total Mahalanobis distance and that due to size and shape are $79.25 = 17.63 + 61.62$.

The sample sizes for chimpanzee and gorilla were 28 and 29, and the χ^2 for size is $(28 \times 29/28 + 29)17.63 = 251.15$, which is significant on 1 degree of freedom. The χ^2 for shape is $(28 \times 29/28 + 29)61.62 = 877.81$, which is significant on 12 degrees of freedom.

If we want to discriminate between the objects of two populations by shape alone, the appropriate linear discriminant function is

$$\begin{aligned} &(\mathbf{d}_1 - \mathbf{d}_2)^T H^T (HSH^T)^{-1} H \mathbf{d} \\ &\quad = (\mathbf{d}_1 - \mathbf{d}_2)^T \left(S^{-1} - \frac{S^{-1} \mathbf{1} \mathbf{1}^T S^{-1}}{\mathbf{1}^T S^{-1} \mathbf{1}} \right) \mathbf{d}, \end{aligned} \quad [3.12]$$

where $\mathbf{d}$ is the vector of log distances on an object. In terms of the original distances the discriminant function is of the form

$$\prod D_{ij}^{a_{ij}}, \ \sum \sum a_{ij} = 0. \quad [3.13]$$

3.3. Comparison of Many Populations. The test statistics in Eqs. **3.10** and **3.11** can be generalized for testing equality in size and shape of several populations. Let $\bar{\mathbf{d}}_i$ and S_i be the statistics in Eqs. **3.1** and **3.2** based on a sample of size n_i from population i, $i = 1, 2, \ldots, r$. Let

$$\begin{aligned} n &= \sum_i n_i, \\ S &= \sum_i (n_i - 1) S_i / \sum_i (n_i - 1), \\ B &= n_1 \bar{\mathbf{d}}_1 \bar{\mathbf{d}}_1^T + \ldots, n_r \bar{\mathbf{d}}_r \bar{\mathbf{d}}_r^T - n \bar{\mathbf{d}} \bar{\mathbf{d}}^T. \end{aligned} \quad [3.14]$$

Then the statistic for testing differences in size is

$$X_{si}^2 = \frac{\mathbf{1}^T S^{-1} B S^{-1} \mathbf{1}}{\mathbf{1}^T S^{-1} \mathbf{1}}, \quad [3.15]$$

which is distributed asymptotically as χ^2 on $(r - 1)$ degrees of freedom, and the statistic for differences in shape is

$$X_{sh}^2 = \operatorname{tr}(BS^{-1}) - \frac{\mathbf{1}^T S^{-1} B S^{-1} \mathbf{1}}{\mathbf{1}^T S^{-1} \mathbf{1}}, \quad [3.16]$$

which is distributed asymptotically as χ^2 on $(m - 1)(r - 1)$ degrees of freedom.

One question that arises in practice is about the choice of a subset of the distances, out of the $k(k - 1)/2$ possible distances, in carrying out the tests of differences in size and shape.

So far as the estimation of the mean shape is concerned, it is desirable to use all the $k(k - 1)/2$ distances. In tests of significance and discriminant analysis, all the distances can be used if the sample sizes are sufficiently large. The fact that the distances are functionally related does not invalidate the asymptotic tests since the relationships are not linear. However, the information content may not increase with increase in the number of distances chosen. This together with small sample sizes ordinarily met with in practice suggests that there is some advantage in choosing a subset of the distances. (see ref. 14 for loss in efficiency in using a large number of variables when sample sizes are small). It is known that a choice of $3k - 6$ (or $2k - 3$ when the relative positions of the landmarks are known) distances is sufficient to specify the configuration of landmarks on a two-dimensional object. There may be different choices of such distances, but in practice any one of these choices would be sufficient to indicate differences in the populations. If we are using nonparametric density estimation for classification purposes it is essential that we should choose a minimal set of distances, which can specify the configuration of landmarks uniquely. Once a set of distances is chosen, the basic statistics needed for statistical analysis can be obtained from Eqs. **3.1** and **3.2** by omitting some elements. In practice, a few different alternative choices may be tried to see if they lead to different conclusions. An example where $2k - 3$ distances are sufficient to specify the configuration of landmarks is that of the profile of the human face as shown in Fig. 1.

Once the differences in shape between the populations are revealed through appropriate tests, it will be of interest to examine the exact nature of differences. For this purpose, we consider the mean shape through regularized mean distances as discussed in Section 3.1. If $\hat{D}_{ij}^{(1)}$ and $\hat{D}_{ij}^{(2)}$ are the distances between the landmarks i and j, we consider all possible ratios

$$\delta_{ij} = \hat{D}_{ij}^{(1)} / \hat{D}_{ij}^{(2)}, \quad i,j = 1, \ldots, k. \quad [3.17]$$

An overall measure of difference in the mean shapes of the two populations is the Hilbert's distance

$$\log \frac{\max \delta_{ij}}{\min \delta_{ij}}. \quad [3.18]$$

A study of the pattern of the δ_{ij} values would indicate regions of the object where shapes in the two populations are more different than in the others.

4. Other Types of Exploratory Analysis

4.1. Principal Component Analysis of Shape. Let S be the estimated covariance matrix of the vector $\mathbf{d}$ of logs of chosen distances. The basic set of shape variables as defined in Eq. **2.1** is

$$\{\mathbf{a}^T\mathbf{d} : \mathbf{a}^T\mathbf{1} = 0, \mathbf{a}^T\mathbf{a} = 1\}. \quad \textbf{[4.1]}$$

We wish to find a member in the set of Eq. **4.1**, which has the maximum variance. The algebraic problem is that of determining **a**, which maximizes $\mathbf{a}^TS\mathbf{a}$ subject to the conditions $\mathbf{a}^T\mathbf{1} = 0$ and $\mathbf{a}^T\mathbf{a} = 1$. The equations leading to the optimum **a** are

$$S\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{1}, \; \mathbf{1}^T\mathbf{a} = 0. \quad \textbf{[4.2]}$$

It is easy to show that the optimum **a** is the eigenvector of QSQ corresponding to the maximum eigenvalue, where $Q = (I - m^{-1}\mathbf{1}\mathbf{1}^T)$; m being the number of components in the vector **d**. The other principal components, $(m - 2)$ in number, correspond to the $(m - 2)$ eigenvectors associated with the other non-zero eigenvalues of QSQ.

4.2. Canonical Correlations. It may be of interest in some situations to know whether the shapes in different regions of an object are related. For instance, we may wish to know whether there is a high correlation between the shapes of the upper and lower parts of a cranium. Let $\mathbf{d}_1$ and $\mathbf{d}_2$ be m_1 and m_2 vectors of log distances arising out of landmarks in the upper and lower parts of the cranium. Corresponding to the choice of $\mathbf{d}_1$ and $\mathbf{d}_2$, we have the covariance matrix,

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \quad \textbf{[4.3]}$$

in the partitioned form. We now consider the shape measurements $\mathbf{y}_1 = H_1\mathbf{d}_1$ and $\mathbf{y}_2 = H_2\mathbf{d}_2$, where H_1 is of the order $m_1 - 1 \times m_1$ and $H_1\mathbf{1} = \mathbf{0}$ and H_2 is of order $m_2 - 1 \times m_2$ and $H_2\mathbf{1} = \mathbf{0}$, with the associated covariance matrix

$$\begin{pmatrix} H_1S_{11}H_1^T & H_1S_{12}H_2^T \\ H_2S_{21}H_1^T & H_2S_{22}H_2^T \end{pmatrix} = \begin{pmatrix} S_{11}^* & S_{12}^* \\ S_{21}^* & S_{22}^* \end{pmatrix} \quad \textbf{[4.4]}$$

and find the canonical correlations between $\mathbf{y}_1$ and $\mathbf{y}_2$. These correlations are independent of the choice of H_1 and H_2 and in any problem they can be chosen in a convenient way.

5. Conclusions

It is demonstrated in this paper that by considering log distances between landmarks the traditional methods of multivariate analysis can be used to study differences in terms of subsets of landmarks in a consistent way. The approaches based on the coordinates as in Kendall (3), Mardia and Dryden (15), and others, the mean shape of a subset of landmarks may depend on other landmarks included in the study.

Appendix

THEOREM 1. *Let* $X \sim N_p(\boldsymbol{\mu}, \Sigma)$ *and* $f(X)$ *be a function such that* $f(X + c\mathbf{1}) = c + f(X)$ *and is distributed independently of* $\mathbf{z} = HX$, *where* H *is a* $p - 1 \times p$ *matrix of rank* $p - 1$ *and* $H\mathbf{1} = \mathbf{0}$. *Then* f *is the unique function, apart from a constant,*

$$f(X) = \frac{\mathbf{1}^T\Sigma^{-1}X}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}. \quad \textbf{[A.1]}$$

Proof: Write $f(X) = g(y, \mathbf{z})$, where $y = \mathbf{1}^T\Sigma^{-1}X/\mathbf{1}^T\Sigma^{-1}\mathbf{1}$. Then,

$$g(y + c,\mathbf{z}) = c + g(y,\mathbf{z}) \Rightarrow \frac{\partial}{\partial y}g(y,\mathbf{z}) = 1 \Rightarrow g(y,\mathbf{z}) = y + h(\mathbf{z}), \quad \textbf{[A.2]}$$

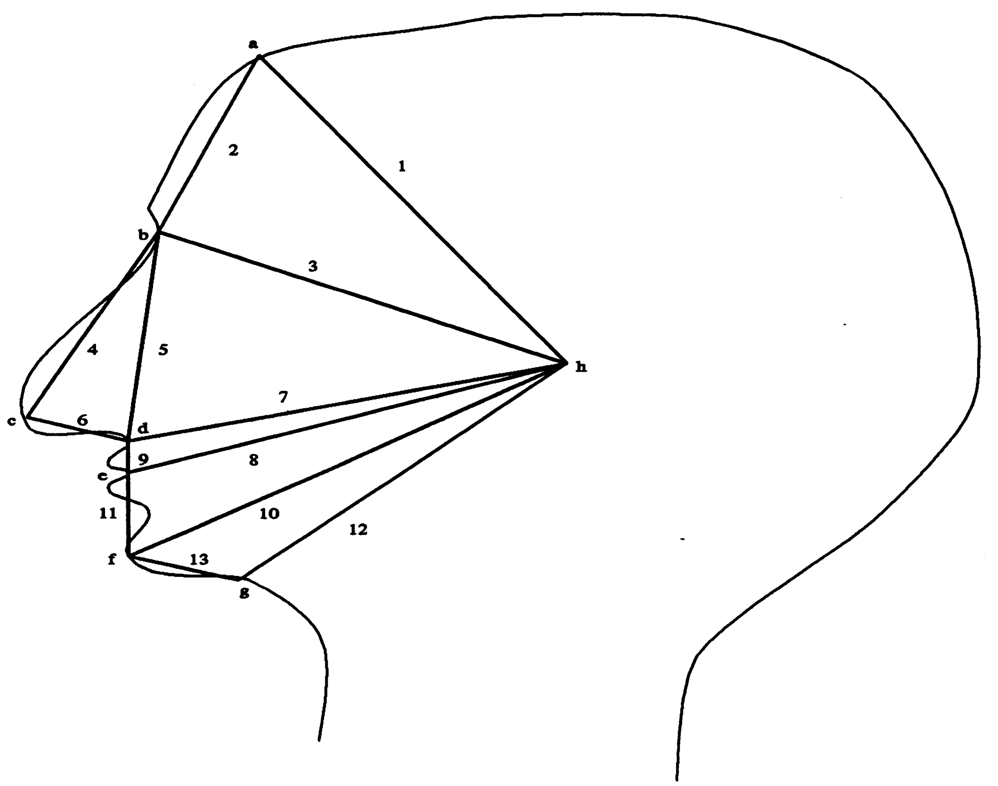

FIG. 1. Minimum number of distances required to fix a human facial profile.

where h is some function of $\mathbf{z}$. Because $X \sim N_p(\boldsymbol{\mu}, \Sigma)$, y, and $\mathbf{z}$, and hence y and $h(\mathbf{z})$ are independently distributed. It is given that $f(X)$ and $\mathbf{z}$ are independent and, therefore, $y + h(\mathbf{z})$ and $h(\mathbf{z})$ are independent.

Now, suppose that $h(\mathbf{z})$ in Eq. **A.2** is not degenerate and let h_1 and h_2 be two distinct support points of the distribution of h($\mathbf{z}$), and a be an arbitrary point such that $a - h_1$ and $a - h_2$ are continuity points of the distribution of y. We have,

$$P\{h(\mathbf{z}) + y \le a\} = P\{h(\mathbf{z}) + y \le a \mid h(\mathbf{z})\} \text{ a.s.}$$
$$= P\{y \le a - h_i\},\ i = 1,2,$$

which is impossible since the set of a's is dense in $\mathfrak{R}$. Hence $h(\mathbf{z})$ must be degenerate.

The research for this paper is supported by the Army Research Office under Grant DAAH04-96-1-0082.

1. Kendall, D. G. (1977) *Adv. Appl. Probab.* **9,** 428–430.
2. Kendall, D. G. (1984) *Bull. Lond. Math. Soc.* **16,** 81–121.
3. Kendall, D. G. (1989) *Stat. Sci.* **4,** 87–120.
4. Bookstein, F. L. (1986) *Stat. Sci.* **1,** 181–242.
5. Bookstein, F. L. (1990) *Commun. Stat. Part A* **19,** 1939–1972.
6. Bookstein, F. L. (1991) *Morphometric Tools for Landmark Data Geometry and Biology* (Cambridge Univ. Press, Cambridge, U.K.).
7. Kent, J. T. (1992) in *The Art of Statistical Science*, ed. Mardia, K. V. (Wiley, New York), pp. 115–127.
8. Lele, S. & Richtsmeier, J. (1991) *Am. J. Phys. Anthropol.* **87,** 49–65.
9. Lele, S. (1993) *Math. Geol.* **25,** 573–602.
10. Rao, C. R. (1973) *Linear Statistical Inference and Its Applications* (Wiley, New York).
11. Sampson, P. D. & Siegel, A. F. (1985) *J. Am. Stat. Assoc.* **80,** 910–914.
12. Torgerson, W. S. (1958) *Theory and Methods of Scaling* (Wiley, New York).
13. Rao, C. R. (1964) *Sankhyā Ser. A* **26,** 329–358.
14. Rao, C. R. (1952) *Advanced Statistical Methods in Biometric Research*, (Wiley, New York); reprinted (1974) by Haffner, New York.
15. Mardia, K. V. & Dryden, I. L. (1994) *Adv. Appl. Probab.* **26,** 334–340.