Mathematical Geology, Vol. 25, No. 5, 1993

Euclidean Distance Matrix Analysis (EDMA): Estimation of Mean Form and Mean Form Difference[1]

Subhash Lele[2]

Euclidean Distance Matrix Analysis (EDMA) of form is a coordinate free approach to the analysis of form using landmark data. In this paper, the problem of estimation of mean form, variance-covariance matrix, and mean form difference under the Gaussian perturbation model is considered using EDMA. The suggested estimators are based on the method of moments. They are shown to be consistent, that is as the sample size increases these estimators approach the true parameters. They are also shown to be computationally very simple. A method to improve their efficiency is suggested. Estimation in the presence of missing data is studied. In addition, it is shown that the superimposition method of estimation leads to incorrect mean form and variance–covariance structure.

KEY WORDS: coordinate free approach, invariance principle, moment estimators, non-central chi-square, nuisance parameters, procrustes methods, superimposition, missing data.

INTRODUCTION

Morphometrics, or the quantitative analysis of biological forms is an important subject. Many different kinds of data are utilized to analyze biological forms. Traditionally scientists have used various linear distances across the form. The technological advances in the last two decades have enabled the scientists to collect data on the complete outline of the object or coordinates of certain biological loci called landmarks. This paper concerns itself with the statistical analysis of landmark coordinate data. In particular, we suggest a method to estimate the mean form and variance–covariance parameters given a sample of n individuals from a population. These estimators are shown to be consistent, that is they approach the true population values as the sample size grows. In paleontology it is common to obtain fossils which are incomplete, that is all the

[1]Received 31 March 1992; accepted 10 November 1992.

[2]Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Maryland 21205.

0882-8121/93/0700-0573$07.00/1

landmarks may not be present on all the individuals. It is shown that the method suggested in this paper handles such missing data problems easily and elegantly.

The last two sections of the paper study the superimposition methods of estimation of form. It is proved that these methods in general lead to inconsistent and asymptotically inefficient estimators of form and shape. Variance–covariance parameter estimators are also in substantial error. Thus the testing procedures based on these methods can lead to incorrect results.

SOME PRELIMINARIES

This section develops notation and states the statistical assumptions and models used throughout the paper. Most biological objects contain specific points referred to as biological landmarks. Landmarks are structurally consistent loci which can have evolutionary, ontogenetic, and/or functional significance (see Van Valen, 1982; Roth, 1988; Lele and Richtsmeier, 1991 for more discussion and examples). We assume that the biological objects under study have K landmarks and have dimension D which is either 2 or 3. We also assume that $K > D$.

Thus a biological object is represented by a $K \times D$ matrix of real variables with the jth row corresponding to D coordinates of the jth landmark. Let us call this matrix a "landmark coordinate matrix."

Let X_i be the landmark coordinate matrix for the ith individual in a sample of size n from a given population. Thus our data consists of n $K \times D$ matrices, namely $X_1, X_2, \ldots, X_n$.

Our statistical model is the perturbation model used by Goodall (1991) among others. Let M be a $K \times D$ matrix corresponding to the mean form. Let

$$X_i = (M + E_i)\Gamma_i + t_i$$

Here, E_i is a $(K \times D)$ matrix valued Gaussian random variable with mean 0 and variance $\Sigma_K \otimes \Sigma_D$ where $\otimes$ denotes Kronecker product. Σ_K is a $K \times K$ positive definite matrix representing the variance–covariance of the columns of E_i and Σ_D is a $D \times D$ positive definite matrix representing the variance–covariance of the rows of E_i. Thus, each column of E_i is a Gaussian $(K \times 1)$ vector with mean 0 and variance–covariance Σ_k and each row a Gaussian $(1 \times D)$ vector with mean 0 and variance–covariance Σ_D. Γ_i is a $(D \times D)$ orthogonal matrix representing rotation and/or reflection of $(M + E_i)$, and t_i is a $K \times D$ matrix with identical rows representing translation. Under these assumptions,

$$X_i \sim MN_{K \times D}(M\Gamma_i + t_i, \Sigma_K, \Gamma_i^T \Sigma_D \Gamma_i)$$

for $i = 1, 2, \ldots, n$. Here "MN" stands for "matrix normal." Parameters of interest are (M, Σ_K, Σ_D) and (Γ_i, t_i) $i = 1, 2, \ldots, n$ are the nuisance parameters.

In words the perturbation model may be explained as follows. To generate random "geometrical objects" or K point configuration in a D dimensional Euclidean space, nature first chooses the mean form M (a $K \times D$ matrix whose columns sum to zero), perturbs each point according to a Gaussian distribution (not necessarily independently of the other points). The K point configuration so obtained is then rotated and/or reflected by an *unknown* angle and translated by an *unknown* amount. Such perturbed, translated, and rotated/reflected K point configurations constitute our data. Since form is considered invariant under rotation/reflection and translation, unknown rotation/reflection angles and translation are not of interest (hence nuisance parameters) when studying "form" of an object. M, the mean form and the variance–covariance matrices Σ_K and Σ_D which dictate the amount of the correlatedness of perturbations are scientifically interesting parameters.

IDENTIFIABILITY AND ESTIMABILITY OF THE PARAMETERS OF INTEREST

In the following, we study the identifiability and estimability aspects for the parameters M, Σ_K, and Σ_D.

Identifiability

Note that even if there were no nuisance parameters, Σ_K (or symmetrically Σ_D) is identifiable only up to a constant, that is, distributions corresponding to parameter combinations (M, Σ_K, Σ_D) and $(M, c\Sigma_K, 1/c\Sigma_D)$ are not distinguishable for any $c > 0$. This means that Σ_D or Σ_K can be estimated only up to a constant multiple.

As a consequence, (in the case that Σ_D has no zero element on the diagonal) without loss of generality, we assume that the first entry in Σ_D is 1; that is

$$\Sigma_2 = \begin{bmatrix} 1 & \rho \\ \rho & a \end{bmatrix} \quad \text{or} \quad \Sigma_3 = \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & a_2 & \rho_3 \\ \rho_2 & \rho_3 & a_3 \end{bmatrix}$$

Here ρ's denote the covariances between the perturbations along the D axes, at a given landmark.

Estimability

From Neyman and Scott (1948), we know that if there is only one observation per stratum, variance is non-estimable. Thus, typically Σ_K is not estimable in this problem (Lele, 1991b; Lele and Richtsmeier, 1990).

Assume that (without affecting the biological problem) M is such that its columns sum to zero, that is, it is a centered matrix. Now suppose we transform the X_i's so that their columns sum to zero, that is we also center X_i's. Let us denote these centered X_i's by X_i^c. Then simple algebra shows that

$$X_i^c \sim MN(M\Gamma_i, \Sigma_K^*, \Gamma_i^T \Sigma_D \Gamma_i)$$

where Σ_K^* is a $K \times K$ non-negative definite matrix of rank $(K - 1)$ corresponding to the variance of the columns of X_i^c. Note that even in the absence of the nuisance parameters Γ_i corresponding to rotation Σ_K is non-estimable, but Σ_K^* is estimable.

Results in the sections on consistent estimation of Σ_K^* and Σ_D show that these parameters are estimable.

Invariance and Nuisance Parameters

Whenever there are nuisance parameters, the first order of business is to eliminate them. For the Gaussian model considered previously, it turns out that a very simple transformation attains this goal.

Let us begin with a simple model (M, Σ_K, I_D), that is,

$$X_i \sim MN_{K \times \mathrm{D}}(M\Gamma_i + t_i, \Sigma_k, I_D)$$

and

$$X_i^c \sim MN_{K \times D}(M\Gamma_i, \Sigma_K^*, I_D)$$

It follows from standard theory (Arnold, 1981, Chap. 17, Sect. 3) that

$$\mathbf{B}_i = X_i^c (X_i^c)^T \sim \text{Wishart}_K(D, \Sigma_K^*, MM^T)$$

that is, the random variables $\mathbf{B}_i$s are $(K \times K)$ matrices and have a Wishart distribution *independent* of the nuisance parameters. Moreover, using $B_1, B_2, \ldots, B_n$, it is possible to obtain consistent estimators of Σ_K^* and MM^T. A natural question then is: Is it enough to estimate MM^T instead of M? Does MM^T represent "form" of the object given by the landmark coordinate matrix M? These questions are treated in the next section.

MAXIMAL INVARIANTS

The form of an object is defined to be that characteristic which remains invariant under translation, rotation, and reflection. Thus if X is a $K \times D$ matrix of landmark coordinates and $X^* = X\Gamma + t$, where Γ is an orthogonal $D \times D$ matrix representing rotation/reflection and t is a $K \times D$ matrix (with identical rows) representing translation, then Form (X) = Form (X^*). The concept of maximal invariant comes to play an important role.

Definition. Let S denote the space of all $K \times D$ matrices or equivalently the space of all objects in dimension D represented by K landmarks. Let $F(\cdot)$ be a function defined on this space such that for X and X^* in S, $F(X) = F(X^*)$ if and only if $X^* = X\Gamma + t$ where Γ is a $D \times D$ orthogonal matrix and t is a $K \times D$ matrix with identical rows. In other words, $F(X) = F(X^*)$ if and only if X^* is just a rotation, reflection, and/or translation of X.

Then $F(\cdot)$ is called a *maximal invariant* defined on the space S under the group of rotation, reflection, and translation transformations.

It is obvious from the definition that any one–one function of a maximal invariant is also a maximal invariant. We consider the following maximal invariant used in Lele (1991a, b) and Lele and Richtsmeier (1991, 1992).

Euclidean Distance Matrix Representation

Note that an object X is just some configuration of K points in a D dimensional space. Consider the following square symmetric matrix, known as Euclidean Distance Matrix in the multidimensional scaling literature (Mardia et al., 1979, Chap. 14):

$$F(X) = \begin{bmatrix} 0 & d(1, 2) & d(1, 3) & \dots & d(1, K) \\ d(2, 1) & 0 & d(2, 3) & \dots & d(2, K) \\ d(3, 1) & d(3, 2) & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & 0 \end{bmatrix}$$

where $d(\ell, m)$ denotes the Euclidean distance between landmarks ℓ and m. For the sake of brevity we write $F(X) = [F_{\ell m}]$ and call it a form matrix.

Theorem 1 of Lele (1991a) proves that $F(\cdot)$ is a maximal invariant under the group of transformations consisting of translation, rotation, and reflection. Thus, $F(\cdot)$ retains all the relevant information about the form of an object.

In this formulation, any configuration of K points is represented by a point in a $K(K - 1)/2$ dimensional Euclidean space. For example, a configuration of three points (a triangle) corresponds to a point in R^3 with coordinates corresponding to the lengths of three sides. We know from elementary geometry that these three lengths have to satisfy the constraint, that the sum of any two sides is at least as great as the third. Thus the form space of three point configurations is given by:

$$M = \{(x, y, z): x \geq 0, y \geq 0, z \geq 0 \quad \text{and}$$
$$x + y \geq z, x + z \geq y, y + z \geq x\}$$

Form space for $K \geq 4$ is difficult to express. However, the following theorem establishes the relationship between number of landmarks K, dimension of the object D and symmetric positive semi-definite matrices.

Theorem 1. Let $Eu(X) = [F^2_{\ell m}]$,

$$A(X) = -\frac{1}{2} Eu(X), \qquad H = H - \frac{1}{K}(1^T 1)$$

where $1 = (1, 1, \ldots, 1)$ a $1 \times K$ vector, and

$$B(X) = HA(X)H = X^c (X^c)^T$$

where X^c is the mean centered X. $B(X)$ is known as centered inner product matrix.

If the configuration of K points lies in a D dimensional Euclidean space and F is its form matrix, then B is a symmetric positive semi-definite matrix B of rank D. Conversely given any $K \times K$ symmetric positive semidefinite matrix of rank D, there necessarily exists a configuration of K points in R^D such that its centered inner product matrix is exactly B.

In other words, the space of K landmark objects in D dimensions corresponds exactly to the space of $K \times K$ symmetric positive semi-definite matrices of rank D.

Proof. Follows from Theorem 14.1 of Mardia et al. (1979; see also Gower, 1966).

Note that MM^T in the previous section is the centered inner product matrix corresponding to the mean form M. The above theorem thus establishes that estimation of MM^T or $F(M)$ or $Eu(M)$ are equivalent to estimating the mean form. In other words, given MM^T one can construct M (up to translation, rotation, and reflection) and *vice versa*.

Lele and Richtsmeier (1991, 1992) developed methodology for studying form difference based on Euclidean distance matrix representation of form. In this paper some estimation procedures are developed and asymptotic properties of these estimators are studied.

CONSISTENT ESTIMATION UNDER THE MODEL (M, Σ_K, I_D)

We will first consider a model where the perturbation of landmarks along the D axes are independent and identical to each other (i.e., Σ_D is an identity matrix) but correlations between landmarks are allowed (i.e., Σ_K is not an identity matrix).

The main feature of this model is that there exists a non-iterative, closed form, consistent estimator for M and Σ_K^*. To be precise, one estimates $F(M)$ consistently, from which M can be obtained up to a similarity transform that is, up to differences attributable only to translation, rotation, and reflection.

We use the following notation:

(i) $F(X) = [F_{\ell m}]_{\substack{\ell=1,2,\ldots,K \\ m=1,2,\ldots,K}}$ where $F_{\ell m}$ is the Euclidean distance between landmarks ℓ and m.

(ii) $Eu(X) = [F_{\ell m}^2] = [e_{\ell m}]$ denotes the matrix of squared distances.

(iii) $B(X) = X^c(X^c)^T$ denotes the centered inner product matrix.

(iv) Let $\Sigma_K = [\sigma_{\ell m}]_{\substack{\ell=1,2,\ldots,K \\ m=1,2\ldots,K}}$ be the variance–covariance matrix and, $Eu(M) = [\delta_{\ell m}]_{\substack{\ell=1,2,\ldots,K \\ m=1,2,\ldots,K}}$ be the square Euclidean distance matrix for the mean form M.

Following theorems lead to the moment estimator for $\delta_{\ell m}$'s and prove its consistency.

Theorem 2. $e_{\ell,m} \sim \phi_{\ell m}\chi_D^2(\delta_{\ell m}/\phi_{\ell m})$ that is, squared Euclidean distances between pairs of landmarks have a non-central χ^2 distribution with D degrees of freedom, noncentrality parameter $\delta_{\ell m}$ and scaling parameter $\phi_{\ell m}$, where $\phi_{\ell m} = \sigma_{\ell\ell} + \sigma_{mm} - 2\sigma_{\ell m}$.

Proof. Follows from the following result regarding a sum of independent non-central χ^2 random variables.

If $W_i \sim \tau^2\chi_1^2(\eta_j/\tau^2)$ $\quad i = 1, 2, \ldots, k$ are mutually independent, then

$$\sum_{i=1}^{k} W_i \sim \tau^2\chi_k^2\left(\frac{\sum_1^k \eta_i}{\tau^2}\right)$$

See Johnson and Kotz (1970, Chap. 28). That is, sum of independent noncentral χ^2 random variables with the same scale parameter, is again a noncentral χ^2 random variable.

The following theorem gives the moments of the random variables $e_{\ell m}$.

Theorem 3. For a two-dimensional object,

$$E(e_{\ell,m}) = 2\phi_{\ell,m} + \delta_{\ell,m} = \alpha_1$$

$$\mathrm{Var}\,(e_{\ell,m}) = 4\phi_{\ell,m}^2 + 4\delta_{\ell,m}\phi_{\ell,m} = \alpha_2$$

and

$$\alpha_1^2 - \alpha_2 = (\delta_{\ell,m})^2 \tag{1}$$

Proof. See Johnson and Kotz (1970, Chap. 28) for moment formulae and then simple algebra proves the theorem.

We equate the sample moments to the population moments to obtain a moment estimator for $\delta_{\ell m}$. Note that (2) is the sample version of (1). The following theorem proves the consistency of the moment estimator for $\delta_{\ell m}$.

Theorem 4. Let $e_{\ell m}^i$ denote the squared Euclidean distance between landmarks ℓ and m in the ith object.

Let

$$\bar{e}_{\ell,m} = \frac{1}{n}\sum_{i=1}^{n} e^{i}_{\ell m}$$

$$S^2(e_{\ell,m}) = \frac{1}{n}\sum_{i=1}^{n} (e^{i}_{\ell,m} - \bar{e}_{\ell,m})^2$$

and

$$\hat{\delta}_{\ell,m} = [(\bar{e}_{\ell,m})^2 - S^2(e_{\ell,m})]^{1/2} \tag{2}$$

Then as $n \to \infty$,

$$\hat{\delta}_{\ell,m} \to \delta_{\ell,m} \quad \text{in probability}$$

Proof. This follows from the consistency of the sample moments and the continuity of the function $\hat{\delta}_{\ell,m}$ (Chung, 1974).

Corollary.

(a) $\widehat{Eu(M)} = [\hat{\delta}_{\ell,m}] \to Eu(M)$ in probability

(b) $\widehat{B(M)} \to B(M)$ in probability

(c) $\widehat{F(M)} \to F(M)$ in probability

Following two theorems generalize the moment estimator of $\delta_{\ell m}$ for two dimensional objects to three-dimensional objects.

Theorem 5. For a three-dimensional object,

$$E(e_{\ell,m}) = 3\phi_{\ell,m} + \delta_{\ell,m} = \beta_1$$

$$\text{Var}\,(e_{\ell,m}) = 6\phi^2_{\ell,m} + 4\delta_{\ell,m}\phi_{\ell,m} = \beta_2$$

and

$$\delta^2\ell,m = \beta_1^2 - \tfrac{3}{2}\beta_2$$

Proof. Similar to Theorem 3.

Theorem 6. Using the same notation as in Theorem 4, and

$$\hat{\delta}_{\ell,m} = [(\bar{e}_{\ell,m})^2 - \tfrac{3}{2}S^2(e_{\ell,m})]^{1/2} \tag{3}$$

It follows that

$$\hat{\delta}_{\ell,m} \to \delta_{\ell,m} \text{ in probability}$$

Proof. Similar to Theorem 4.

Corollary after Theorem 4 also holds. Next two theorems utilize the estimators of $\delta_{\ell m}$ to obtain a consistent estimator of the variance–covariance parameter Σ^*_K.

Theorem 7. $B(X_i) \sim \text{Wishart}_K(D, \Sigma^*_K, B(M))$.

Proof. Let X_i^c denote the mean centered observations. Then under the model and results of the section on estimability

$$X_i^c \sim MN_{K\times D}(M\Gamma_i, \Sigma_K^*, I)$$

and $B(X_i) = (X_i^c)(X_i^c)^T$. The result now follows from Section 17.3 of Arnold (1981).

Theorem 8. $E(B(X)) = D\Sigma_K^* + B(M)$ and

$$\hat{\Sigma}_K^* = \frac{1}{D}\left[\frac{1}{n}\sum_{i=1}^{n} B(X_i)\right] - B(M) \rightarrow \Sigma_K^* \text{ in probability.}$$

Proof. Follows from Arnold (1981, Theorem 17.6); consistency of moments and the corollary after Theorem 4.

Note. The following algorithm (Principal Coordinate Analysis, Gower, 1966) explains in detail how one can obtain $\hat{M}$, the estimated coordinates of the mean form (up to translation, rotation, and reflection transformations) using the estimators given in (2) and (3). A modified estimator for Σ_K^*, which may behave slightly better than the above estimator, is suggested. Although for medium to large sample sizes, they are almost identical.

$$Eu(M) = [\hat{\delta}_{\ell m}]_{\substack{\ell=1,2,\ldots,K\\ m=1,2,\ldots,K}}$$

be the symmetric $k \times k$ matrix of squared Euclidean distances where $\hat{\delta}_{\ell m}$ are the estimates obtained by Eqs. (2) or (3).

Step 1. Calculate $B(M) = H\{Eu(M)\}H$ where $H = I - 1/K(1^T1)$ is a $K \times K$ symmetric matrix (also used in Theorem 1) such that its diagonal entrees are $1 - 1/K$ and off diagonal entrees are $-1/K$.

Step 2. Calculate the eigenvalues and eigenvectors of $B(M)$. Let the eigenvalues be $\lambda_1 > \lambda_2 > \ldots > \lambda_K$ and the corresponding eigenvectors be h_1, $h_2, \ldots, h_K$.

Step 3. $\hat{M}$, the estimator of the coordinates of the mean form M (up to translation, rotation and reflection transformations) is given by:

For a two-dimensional object:

$$\hat{M} = [\sqrt{\lambda_1}h_1, \sqrt{\lambda_2}h_2]$$

Note that this is a $K \times 2$ matrix. For a three-dimensional object:

$$\hat{M} = [\sqrt{\lambda_1}h_1, \sqrt{\lambda_2} \cdot h_2, \sqrt{\lambda_3}h_3]$$

This is a $K \times 3$ matrix.

One may plot these coordinates to visualize the mean form pictorially.

Given $\hat{M}$, the modified estimator of Σ_K^* is given by:

$$\hat{\Sigma}_K^* = \frac{1}{n} \sum_{i=1}^{n} B(X_i) - \hat{M}\hat{M}^T$$

Again note that for large samples, $B(M)$ and $\hat{M}\hat{M}^T$ are almost identical. Hence the two estimators of Σ_K^* are equivalent as $n \to \infty$.

In summary, the above results imply the following: (1) Under the Gaussian perturbation model (M, Σ_K, I), it is possible to estimate the mean form M and variance–covariance matrix Σ_K^* consistently. (2) The estimators are extremely easy to calculate.

CONSISTENT ESTIMATION WHEN Σ_D IS A GENERAL POSITIVE DEFINITE SYMMETRIC-MATRIX

In this section, we consider estimation of M, Σ_K, and Σ_D when correlation between perturbations along the D axes is allowed, i.e., Σ_D is *not* an identity matrix.

When Σ_D is a general positive definite symmetric matrix, the problem is substantially more difficult. The number of parameters is larger, thus needing the higher order moments. In this section, we use the representation for quadratic forms involving Gaussian random variables (Dik and de Gunst, 1985) for calculating these moments. Given these moments, one can derive the moment equations and solve them numerically to get the moment estimators.

The following is a description of Dik and de Gunst's result (specialized to a particular case; see also de Gunst, 1987).

Notation. Let X be a Gaussian $(D \times 1)$ random vector with mean μ and variance Σ. Let Σ be positive definite.

Let $\lambda_1, \lambda_2, \ldots, \lambda_D$ be the eigenvalues of Σ. Let S be a square root of Σ, i.e., if the Jordan decomposition of $\Sigma = P\Lambda_p P^T$, then $S = P\Lambda_P^{1/2} P^T$. Let

$$\omega = \Lambda_P^{-1} P^T S^T \mu = \Lambda_P^{-1/2} P^T \mu$$

$$\xi = \mu^T \mu - \omega^T \Lambda_P \omega$$

A simple calculation shows that $\xi = 0$.

Theorem 9 (Dik and de Gunst, 1985):

$$X^T X \cong \sum_{i=1}^{D} \lambda_i (u_i + \omega_i)^2$$

where $u_1, u_2, \ldots, u_D$ are independent identically distributed standard normal random variables and ω_i, denotes the ith component of vector ω. $V \cong W$ denotes that V and W have the same distribution.

For the situation we are interested in, namely the distribution of the squared distances between any two landmarks, D is either 2 or 3. Using the above representation, moments for $e_{\ell m}$ with various Σ_Ds can be calculated as follows:

Let us consider any two landmarks on a two dimensional object. Without loss of generality, we label them 1 and 2 and assume that they have the following distribution:

$$\begin{pmatrix} X_1 & Y_1 \\ X_2 & Y_2 \end{pmatrix} \sim MN_{2\times 2}\left[\begin{pmatrix} 0 & 0 \\ \mu & 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & a \end{pmatrix}\right]$$

Hence,

$$(X_2 - X_1,\ Y_2 - Y_1) \sim N_{1\times 2}\left[(\mu, 0),\ \phi\begin{pmatrix} 1 & \rho \\ \rho & a \end{pmatrix}\right]$$

where

$$\phi = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$$

Thus:

$$\begin{aligned} e_{12} &= (X_2 - X_1)^2 + (Y_2 - Y_1)^2 \\ &\cong \sum_{i=1}^{2} \lambda_i (u_i + \omega_i)^2 \end{aligned}$$

where λ_is are the eigenvalues of the matrix $[\begin{smallmatrix} \phi & \rho\phi \\ \rho\phi & a\phi \end{smallmatrix}]$, P is the matrix of eigenvectors,

$$\omega_1 = \mu \frac{P_{11}}{\sqrt{\lambda_1}} \qquad \omega_2 = \mu \frac{P_{21}}{\sqrt{\lambda_2}}$$

and

$$P_{11}^2 + P_{21}^2 = 1$$

Thus for a two-dimensional object, the moments for the squared distance between any two landmarks are given by:

$$\begin{aligned} E(e_{ij}^k) &= E \sum_{i=1}^{2} \lambda_i (u_i + \omega_i)^{2K} \\ &= \sum_{\ell=0}^{K} \binom{K}{\ell} \lambda_1^{2\ell}\lambda_2^{2(k-\ell)} E(u_1 + \omega_1)^{2\ell} E(u_2 + \omega_2)^{2(k-\ell)} \end{aligned}$$

For a three-dimensional object, the corresponding expression is:

$$E(e_{ij}^k) = \sum_{\substack{0 \le \ell_i \le k \\ \sum_1^3 \ell_i = k}} \frac{K!}{\ell_1!\,\ell_2!\,\ell_3!} \lambda_1^{2\ell}\lambda_2^{2\ell_2}\lambda_3^{2\ell_3} \prod_{i=1}^{3} E(u_i + \omega_i)^{2\ell_i}$$

with obvious notation. Also it is easy to show that

$$E(u + \omega)^{2m} = E \sum_{j=1}^{2m} \binom{2m}{j} u^j \omega^{2m-j}$$
$$= \sum_{j=1}^{2m} \binom{2m}{j} E(u^j) \omega^{2m-j}$$

But $E(u^j) = 0$ for all odd values of j. Hence

$$E(u + \omega)^{2m} = \sum_{j=1}^{m} \binom{2m}{2j} \omega^{2m-2j} E(u^{2j})$$

and

$$E(u^{2j}) = \frac{(2j)!}{j!} \times \left(\frac{1}{2}\right)^j$$

For a two-dimensional object, since there are at most four parameters (μ, ϕ, ρ, a), one needs to solve four moment equations. Correspondingly, for a three-dimensional object, one needs to solve seven moment equations. Note that there is a one–one relationship between (λ, ω) and (μ, Σ_D, ϕ) parameterizations. The above information is sufficient to produce explicit equations. Obviously numerical methods are required to solve them. From the theory of estimating equations, we also know that, under typical conditions these estimators are consistent and asymptotically Gaussian.

There are some difficulties associated with these estimators:

1. It is not obvious that the given system of equations has a unique solution.
2. The estimators depend on moments of fairly high order. Thus although consistent, they conceivably are not very efficient.

Now if we behave as if $\Sigma_D = I$ (when estimation is easy) when in fact $\Sigma_D \neq I$, then we obviously will get inconsistent estimators and related confidence intervals. A question of interest would be: How far off are these estimators? This will be explored in the next section.

ORDER OF INCONSISTENCY FOR THE ESTIMATORS OF M When $\Sigma_D \neq I_D$

Let us consider a two dimensional object. Suppose we use the estimator (2) or (3) for estimating M. It is consistent when $\Sigma_D = I$. However, it is inconsistent when $\Sigma_D \neq I$. We will calculate the order of inconsistency when Σ_D is a correlation matrix. Similar calculations can be conducted for other Σ_Ds.

Let $\Sigma_D = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. Then the eigenvalues of Σ_D are given by $\lambda_1 = 1 - \rho$ and $\lambda_2 = 1 + \rho$. Let P denote the matrix of eigenvectors. Using the results of the previous section, it is easy to show that:

$$E(e_{\ell m}) = \delta_{\ell m} + 2\phi_{\ell m}$$

$$\begin{aligned}
\text{Var}\,(e_{\ell m}) &= E(e_{\ell m}^2) - [E(e_{\ell m})]^2 \\
&= 2\sum_{i=1}^{2}\lambda_i^2 + 4\sum_{1}^{2}\lambda_i^2\omega_i^2 \\
&= 2\phi_{\ell m}^2[(1-\rho)^2 + (1+\rho)^2] + 4\phi_{\ell m}\delta_{\ell m}(\lambda_1 p_{11}^2 + \lambda_2 p_{12}^2) \\
&= 4\phi_{\ell m}^2 + 4\rho^2\phi_{\ell m}^2 + 4\phi_{\ell m}\delta_{\ell m}(-2\rho p_{11}^2 + (1+\rho)) \\
&= 4\phi_{\ell m}^2 + 4\phi_\ell\delta_{\ell m} + 4\rho^2\phi_{\ell m}^2 + 4\rho\,\phi_{\ell m}\delta_{\ell m}(1 - 2p_{11}^2)
\end{aligned}$$

But note that for this Σ_D, $1 - 2P_{11}^2 = 0$. Hence

$$\text{Var}\,(e_{\ell m}) = 4\phi_{\ell m}^2 + 4\delta_{\ell m}\phi_{\ell m} + 4\rho^2\phi_{\ell m}^2$$

Thus:

$$\begin{aligned}
\hat{\delta}_{\ell m}^2[E(e_{\ell m})]^2 &- \text{Var}\,(e_{\ell m}) \\
&= \delta_{\ell m}^2 - 4\rho^2\phi_{\ell m}^2
\end{aligned}$$

Since in practice, $4\rho^2\phi_{\ell m}^2$ are fairly small compared to $\delta_{\ell m}^2$, estimators given by (2) are very good even when $\Sigma_D \neq I$. Model misspecification has very little effect.

ESTIMATION OF THE FORM DIFFERENCE

In practical situations, the quantity of interest is either form or shape difference between two populations.

Let $X_1, X_2, \ldots, X_n$ be n independent observations from population I and $Y_1, Y_2, \ldots, Y_m$ be m independent observations from population II. Let the mean form of population I be M^X with the corresponding form matrix $F(M^X)$ and corresponding quantities for population II be M^Y and $F(M^Y)$. There are several different ways to define the difference between mean forms M^X and M^Y. Goodall and Bose (1987, Eq. 3), Rohlf and Slice (1990) among others define it as the coordinatewise difference between M^X and M^Y taken after a proper superimposition. In Lele (1991a), this approach is critically evaluated on the scientific basis and it is argued that one should use form difference defined in terms of $F(M^X)$ and $F(M^Y)$. Following is a definition of form difference which considers relative changes in the forms. It is a vector of ratios of like distances in two forms. This is a scientifically interesting way of studying form difference.

Definition. Form difference between populations I and II is defined as:

$$FDM(M^X, M^Y) = \frac{F(M^X)}{F(M^Y)}$$

where the division is conducted elementwise, with the convention that 0/0 = 0.

We will now explore the consistent estimation of $FDM(M^X, M^Y)$. The following result is an immediate consequence of Theorem 4.

Theorem 10. Let the parameters for the two populations be $(M^X, \Sigma^*_{KX}, \Sigma_{DX})$ and $(M^Y, \Sigma^*_{KY}, \Sigma_{DY})$.

If $\Sigma_{DX} = \Sigma_{DY} = I$, then

$$FD\hat{M}(M^X, M^Y) = \frac{\hat{F}(M^X)}{\hat{F}(M^Y)} \to FDM(M^X, M^Y)$$

in probability.

This theorem shows that the form difference between two populations can be estimated consistently when landmarks are perturbed dependently along each axis but independently between the axes.

We will now explore the situations where $\Sigma_{DX} \neq I$ and $\Sigma_{DY} \neq I$, i.e., when between axes correlation is present. As in the previous section, we will calculate the rate of inconsistency. For further exploration, we will assume that $\Sigma^*_{KX} = \Sigma^*_{KY}$ and $\Sigma_{DX} = \Sigma_{DY}$, i.e., the two populations have the same variance–covariance but possibly different means. As usual we will consider the squared Euclidean distances matrices $Eu(M^X)$ and $Eu(M^Y)$. Define

$$\psi(M^X, M^Y) = \frac{Eu(M^X)}{Eu(M^Y)} = \left[\frac{\delta^X_{\ell m}}{\delta^Y_{\ell m}}\right]_{\substack{\ell=1,2,\ldots,K \\ m=1,2,\ldots,K}}$$

Thus, ψ is an elementwise squared matrix $FDM(M^X, M^Y)$. Let $\psi_{\ell m}$ denote the (ℓ, m)th element of ψ. It is just the ratio of squared Euclidean distance between landmarks ℓ and m in population I and II.

Let

$$\hat{\psi}^2_{\ell m} = \left(\frac{\hat{\delta}^X_{\ell m}}{\hat{\delta}^Y_{\ell m}}\right)^2$$

We are interested in calculating the order of the difference $(\hat{\psi}^2_{\ell m} - \psi^2_{\ell m})$ when $\Sigma_D \neq I$. Let $\Sigma_D = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. From the previous section, it follows that

$$\hat{\psi}^2_{\ell m} \to \frac{(\delta^X_{\ell m})^2 - 4\rho^2\phi^2_{\ell m}}{(\delta^2_{\ell m})^2 - 4\rho^2\phi^2_{\ell m}}$$

in probability, Consider,

$$\hat{\psi}^2_{\ell m} - \psi^2_{\ell m} = \frac{4(\rho\phi_{\ell m})^2[(\delta^X_{\ell m})^2 - (\delta^Y_{\ell m})^2]}{(\delta^Y_{\ell m})^2[(\delta^Y_{\ell m})^2 - 4\rho^2\phi^2_{\ell m}]} \tag{4}$$

The following results are now obvious:

Result 1. Note that when $M^X = M^Y$, $\delta^X_{\ell m} = \delta^Y_{\ell m}$. Hence, $\hat{\psi}^2_{\ell m} - \psi^2_{\ell m} \to 0$ in probability. That is, when the usual null hypothesis of no form difference holds, one can estimate the form difference consistently, provided the two populations have the same variance–covariance structure.

Result 2. If the two mean forms M^X and M^Y are not identical, then the form difference $FDM(M^X, M^Y)$ can be estimated consistently when the landmarks are perturbed independently between the axes, although correlation between landmarks along the axes is allowed.

Result 3. In the most general case where correlation along and between axes is allowed, simple estimators of $FDM(M^X, M^Y)$ obtained from (2) or (3) viz. $(\hat{\delta}^X_{\ell m}/\hat{\delta}^Y_{\ell m})^{1/2}$ are very accurate. The amount of error can be calculated from Equation (4). These estimators are thus fairly robust against model misspecifications.

ESTIMATION IN THE PRESENCE OF MISSING LANDMARKS

It is common in paleontological and anthropological studies to have individuals in the sample on which some of the landmarks are missing. For example, see Leakey et al. (1991) where form and shape differences between two extinct primate species are studied. Missing landmarks were constructed by anatomical knowledge, scientific experience, and intuition. In Lele (1992) a more quantitative approach is suggested to construct such missing landmarks when a few individuals in the sample have all the landmarks present. However consider the following situation where *none* of the individuals in the sample have all the landmarks present. Even in such an extreme situation it is shown that the methodology discussed in the previous section can be applied successfully in order to obtain *complete* geometrical information about the biological structure from such a *partial* set of observations.

Consider the biological structure defined by the landmarks Frontal Zygomatic Intersection (FZI), Nasion (NAS), and Zygomaxillare Superior (ZMS). Figure 1 illustrates this structure. It is not unusual to find fossils with one of these landmarks missing because the bony structure on which they lie tend to break easily (Leakey et al., 1991). Suppose now that we have n_1 individuals with only FZI and NAS present, n_2 individuals with only NAS and ZMS present, and n_3 individuals with only FZI and ZMS present. All of these individuals are from the same species and are of the same age group. Note that to estimate the form of the biological structure defined by the three landmarks FZI, NAS, and ZMS, one only needs to estimate the three distances FZI-NAS, NAS-ZMS, and ZMS-FZI. These are estimated consistently by the following estimators.

Let FZI = landmark 1, NAS = landmark 2, and ZMS = landmark 3. From the previous sections, it is clear that it is enough to estimate δ_{12}, δ_{13}, and

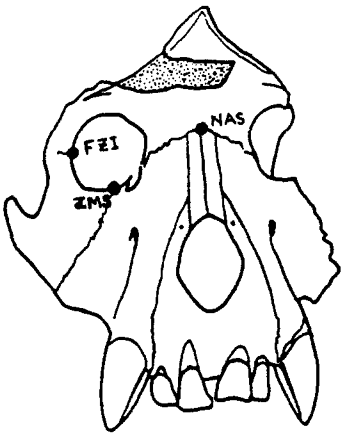

Fig. 1. Frontal view of Afropithecus turkanensis cranium showing the positions of landmarks Frontal Zygomatic Intersection (FZI), Nasion (NAS) and Zygomaxillare Superior (ZMS) (adapted from Leakey et al., 1991, figure 1).

δ_{23}. Given these three squared distances, one can construct the mean triangle defined by them uniquely. From (2), it follows that

$$\hat{\delta}_{12} = [\bar{e}_{12} - s^2(e_{12})]^{1/2}$$

$$\hat{\delta}_{13} = [\bar{e}_{13} - s^2(e_{13})]^{1/2}$$

$$\hat{\delta}_{23} = [\bar{e}_{23} - s^2(e_{23})]^{1/2}$$

are consistent estimators of δ_{12}, δ_{13}, δ_{23}, respectively. One gets $\bar{e}_{12}$ and $s^2(e_{12})$ from the n_1 individuals on which FZI and NAS are present and so on. Hence estimation of the complete geometrical structure using only partial observations is possible.

Some algebraic manipulations lead to a consistent estimator of the variance–covariance structure. Being able to handle such missing data is extremely important from the practical point of view. Note that the superimposition estimators that will be discussed later fail to handle this important practical situation.

ROBUSTNESS AGAINST MODEL SPECIFICATION

Referees of this paper have pointed out quite correctly that the nice properties of the estimators discussed in the earlier sections are derived under a Gaussian perturbation model. It is important to study robustness of these estimators under model misspecification.

First of all, we would like to point out that the variance–covariance structure considered previously is fairly general. The assumption of the Gaussianity of perturbations is difficult to test precisely. However note that our estimators depend only on the first two moments of the distribution of the squared distances between landmarks. If the perturbations are not extremely skewed, these squared distances have approximately a χ^2 distribution. Particularly the relationship between the first two moments is satisfied reasonably well. Thus the estimators should be reasonably robust.

Referees' querries also led us to the following tantalizing results. We will describe them briefly and only heuristically here. Suppose one does not want to assume Gaussian or any other particular perturbation distribution. However it might be reasonable to assume that two populations have identical variance covariance structure. Let us define difference between two forms as:

$$AFD(M^X, M^Y) = F(M^X) - F(M^Y)$$

where the differences are taken elementwise.

The definition used previously gives relative form difference whereas the above definition gives absolute form difference. It turns out that this form difference can be estimated consistently without any model assumptions other than the equality of variances as follows:

Note that $E(e_{\ell m}) = \delta_{\ell m} + 2\phi_{\ell m}$ irrespective of the Gaussianity assumption. Since sample moments converge to the population moments as $n \to \infty$,

$$\bar{e}^X_{\ell m} \to \delta^X_{\ell m} + 2\phi_{\ell m}$$

$$\bar{e}^Y_{\ell m} \to \delta^Y_{\ell m} + 2\phi_{\ell m}$$

When the two variance–covariance structures are identical,

$$\bar{e}^X_{\ell m} - \bar{e}^Y_{\ell m} \to \delta^X_{\ell m} - \delta^Y_{\ell m} \text{ in probability}$$

Thus one can estimate $AFD(M^X, M^Y)$ consistently without any model assumptions.

Moreover by applying Central Limit Theorem (Serfling, 1980), it follows that this estimator is also asymptotically Gaussian and hence can be used to obtain a model robust test procedure for form difference as well as model robust confidence intervals for the absolute form difference in the following fashion.

Consider the following quadratic form:

$$(\bar{e}^X_{\ell m} - \bar{e}^Y_{\ell m})^T S^+ (\bar{e}^X_{\ell m} - \bar{e}^Y_{\ell m})$$

where S^+ is a generalized inverse of the variance–covariance matrix of $(\bar{e}^X_{\ell m} - \bar{e}^Y_{\ell m})$. We need a generalization inverse because the variance–covariance matrix could be singular (Rao, 1973, Chap. 1). With proper standardization, this quadratic form has a χ^2 distribution, which can be used to obtain a testing procedure as well as confidence intervals for $AFD(M^X, M^Y)$.

Generalization to comparing shapes instead of forms is also possible in principle. We will not discuss the details here.

In conclusion, using the Euclidean Distance Methodology one can possibly obtain model robust testing procedures for form/shape differences.

SUPERIMPOSITION APPROACH FOR ESTIMATION OF (M, Σ_K, Σ_D)

Goodall (1991) and Bookstein (1986) suggest using superimposition methods to estimate mean form and shape as well as variance–covariance parameters Σ_K and Σ_D. Goodall (1991) also claims generalized procrustes estimators to be consistent, asymptotically efficient and maximum likelihood estimators. These claims are not supported by any mathematical proofs, neither is the likelihood function produced. In the following, it is shown that the superimposition approaches yield inconsistent estimators of mean form as well as shape. It is also shown that variance–covariance parameters are nonidentifiable under this scheme.

Following is a brief discussion of superimposition approach.

Given $X_1, X_2, \ldots, X_n$, one translates and rotates these matrices in such a manner that $\Sigma_{i \neq j}\phi(X_i, X_j)$ is minimized for some preselected, non-negative, real valued loss function $\phi(\cdot)$. The examples of $\phi(\cdot)$ are:

(a) Ordinary procrustes analysis (Goodall, 1991),

$$\phi(X_i, X_j) = tr\{(X_i - X_j)(X_i - X_j)^T\}$$

(b) Weighted procrustes analysis (Goodall, 1991),

$$\phi(X_i, X_j) = tr\{(X_i - X_j)W(X_i - X_j)^T\}$$

for some weight matrix W.

(c) Edge superimposition (Bookstein, 1986; Bookstein and Sampson, 1990); Fix an edge, (without loss of generality) say (1, 2), i.e., the line joining landmarks 1 and 2.

$$\phi(X_i, X_j) = \begin{cases} 0 & \text{if the edge (1, 2) in } X_i \text{ is aligned with edge (1, 2) in } X_j \\ \infty & \text{otherwise} \end{cases}$$

Let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ be the transformed variables obtained after using one of the superimposition metrics. Then under all superimposition schemes, mean form estimator corresponds to the coordinatewise average of the transformed variables.

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i = \overline{\tilde{X}}$$

Under Goodall's procrustes scheme (Goodall, 1991, Eq. 10.2, 10.3), the variance–covariance parameters are estimated by:

$$\hat{\Sigma}_K^* = \frac{1}{n}\sum_{i=1}^{n} (\tilde{X}_i - \hat{M})(\tilde{X}_i - \hat{M})^T/\text{tr}\,\hat{\Sigma}_D$$

$$\hat{\Sigma}_D = \frac{1}{n}\sum_{i=1}^{n} (\tilde{X}_i - \hat{M})^T(\tilde{X}_i - \hat{M})$$

Inconsistency of Edge Superimposition Estimator of the Mean Form M

For the sake of simplicity, consider an object with only three landmarks. Let the parameters be

$$M = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 10 \end{bmatrix}$$

$\Sigma_3 = \sigma^2 I_3$ and $\Sigma_2 = I_2$ where I_k denotes $k \times k$ identity matrix. Thus, each $X_i \sim MN_{3\times 2}(M, \sigma^2 I_3, I_2)$. Suppose that edge (1, 2) is used for superimposition. Then each X_i is transformed in such a manner its first row is (0, 0) and second row is $(a, 0)$ for some $a > 0$.

Let

$$X_i = \begin{bmatrix} Z_1^i \\ Z_2^i \\ Z_3^i \end{bmatrix}$$

where Z_j denotes the jth row and $Z_j = (Z_{1j}, Z_{2j})$ $j = 1, 2, 3,$

$$\Gamma(\theta_i) = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix} = (\alpha_1^i, \alpha_2^i) \text{ where } \alpha_1^i \text{ is the first column}$$

and α_2^i is the second column of the matrix $\Gamma(\theta_i)$, and

$$\theta_i = \sin^{-1}\left[\frac{Z_{22}^i - Z_{21}^i}{\|Z_2^i\|}\right] = \cos^{-1}\left[\frac{Z_{12}^i - Z_{11}^i}{\|Z_2^i\|}\right]$$

Then, with this notation, the transformed X_i's viz. $\tilde{X}_i$'s are:

$$\tilde{X}_i = \begin{bmatrix} 0 & 0 \\ \|Z_2^i\| & 0 \\ Z_3^i\alpha_1^i & Z_3^i\alpha_2^i \end{bmatrix}$$

Hence, the coordinatewise average of these $\tilde{X}_i$'s yields,

$$\bar{\tilde{X}} = \begin{bmatrix} 0 & 0 \\ \frac{1}{n}\sum_1^n \|Z_2^i\| & 0 \\ \frac{1}{n}\sum_1^n Z_3^i\alpha_1^i & \frac{1}{n}\sum_1^n Z_3^i\alpha_2^i \end{bmatrix}$$

Since

$$[(Z^i_{12} - Z^i_{11})^2 + (Z^i_{22} - Z^i_{21})^2]^{1/2} > [(Z^i_{12} - Z^i_{11})^2]^{1/2}$$

it follows, by taking expectations on both sides, that

$$E\|Z_2\| > \mu_2 = 1$$

Thus $\bar{X} \nrightarrow M$ in probability as $n \to \infty$. $\bar{X}$ is an *inconsistent* estimator of M. This also implies that the sample variance–covariance matrix does not estimate the true variance–covariance matrix consistently (Campbell, 1986 notes a similar result but does not realize its full implications).

Generalized Procrustes Analysis (GPA)

This section proves the following results.

(a) GPA estimator of mean form is inconsistent and hence asymptotically inefficient.

(b) The orbits defined on the parameter space $(M, \Sigma_K^*, \Sigma_D)$ by the Procrustes analysis are such that Σ_D is non-identifiable. This will be explained in the following. However we would like to note that this parameter is important biologically as it tells us the perturbation covariance along the different axes. Inability of the superimposition schemes to estimate this biologically important parameter is unsatisfactory from the scientific point of view.

Inconsistency of the GPA Estimator of M

Consider the model where $\Sigma_K = \sigma^2 I$, $\Sigma_D = I$. This is the simplest perturbation model also considered by Langron and Collins (1985). Let $X_1, X_2, \ldots, X_n$ be n random $K \times D$ matrices generated under the above model. Let G denote the procrustes sum of squares, viz.

$$G = \sum_{i=1}^{n} tr(\tilde{X}_i - \bar{\tilde{X}})(X_i - \bar{\tilde{X}})^T$$

where $\tilde{X}_i$ are translated and rotated (without scaling) figures such that G is minimized. The following results are well known.

Result 1. (Gower, 1975, Eq. 15)

$$G = tr \sum_{i=1}^{n} \tilde{X}_i\tilde{X}_i^T - ntr\bar{\tilde{X}}\bar{\tilde{X}}^T$$

Result 2. (Langron and Collins, 1985, Theorem 6.1)

$$G \sim \sigma^2 \chi^2_L$$

where $L = (n - 1)(KD - \frac{1}{2}D(D + 1))$.

Result 3. (Arnold, 1981, Chap. 17)

(a) $\tilde{X}_i\tilde{X}_i^T \sim \text{Wishart}_K(D, MM^T, \Sigma_K^*)$ for all $i = 1, 2, \ldots, n$

(b) $\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i\tilde{X}_i^T \to MM^T + D\Sigma_K^*$ in probability as $n \to \infty$

(c) $tr\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i\tilde{X}_i^T\right) \to tr\, MM^T + D\, tr\, \Sigma_K^* = tr\, MM^T + D(K-1)\sigma^2$

in probability as $n \to \infty$, since $tr\, \Sigma_K^* = (K-1)\sigma^2$.

Result 4. It follows trivially from Result 2 that

$$\frac{1}{n} G \to \sigma^2\left[KD - \frac{1}{2}D(D+1)\right] \text{ in probability as } n \to \infty$$

Let us now suppose that the GPA estimator of mean form is consistent. We show that one reaches a contradiction and thus prove the result by the method of *reductio ad absurdum*.

Suppose $\bar{X} \to M$ in probability as $n \to \infty$. (Note that this convergence is up to a similarity transform.) Then it implies, by Slutsky's theorem (Chung, 1974) that

$$tr\, \bar{\tilde{X}}\bar{\tilde{X}}^T \to tr\, MM^T \text{ in probability}$$

Combining Result 1 and Result 3 with this result, it follows that

$$\frac{1}{n} G \to D(K-1)\sigma^2 \text{ in probability}$$

But Result 4 shows that $(1/n)G$ converges to $\sigma^2(KD - \frac{1}{2}D(D+1))$ which is *not* equal to $D(K-1)\sigma^2$. We reach a contradiction, thus proving inconsistency of the GPA estimators of mean form even under the simplest model. This also implies that these estimators are not asymptotically efficient.

Non-Identifiability of Σ_D Under Superimposition Schemes Let us consider the model $(M, \Sigma_K^*, \Sigma_D)$. Goodall (1991) suggests the following as an estimator of Σ_D

$$\hat{\Sigma}_D = \frac{1}{n}\sum_{i=1}^{n} (\tilde{X}_i - \hat{M})^T(\tilde{X}_i - \hat{M})$$

where $\tilde{X}_i$s are the translated and rotated figures.

Now consider the following practical situation. Let person A get the sample $X_1, X_2, \ldots, X_n$ and let $\hat{\Sigma}_D^A$ denote his estimate of Σ_D. Suppose person B received the same sample except that each of X_is is now rotated by multiplying by an orthogonal matrix C, i.e., person B receives the sample $Y_i = X_iC$, $i = 1, 2, \ldots, n$. Let his estimate of Σ_D be denoted by $\hat{\Sigma}_D^B$. These two estimates are not equal to each other, in fact $\hat{\Sigma}_D^B = C^T\hat{\Sigma}_D^AC$. So should we take $\hat{\Sigma}_D^A$ or $\hat{\Sigma}_D^B$ as the estimator of Σ_D? This situation is not an academic one. Two scientists in two different laboratories may collect same landmarks off of the same fossil specimens but set them on the digitizer in different orientations, thus getting rotated versions of each other's data as described above.

This is not just a mathematical artifact either. What is going on here is the following.

Let $\theta = \{(M, \Sigma_K^*, \Sigma_D)$: M in $K \times D$ matrix, Σ_K^* is symmetric, positive semidefinite matrix of rank $K - 1$, Σ_D is symmetric, positive semidefinite matrix of rank $D\}$ be the parameter space. Then under the procrustes scheme (in fact, with minor modifications under any superimposition scheme) the orbits of equivalent parameters are given by:

$$\{(MC, \Sigma_K^*, C^T \Sigma_D C): (M, \Sigma_K^*, \Sigma_D) \in \theta, C \text{ is a } D \times D \text{ orthogonal matrix}\}.$$

These are the wrong orbits since clearly

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & a & \rho_3 \\ \rho_2 & \rho_3 & b \end{pmatrix} \text{ is not equivalent to } \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

in a biological sense. Biologically sensible orbits are given by

$$\{(MC, \Sigma_K^*, \Sigma_D): (M, \Sigma_K^*, \Sigma_D) \in \theta \text{ and } C \text{ is an orthogonal } D \times D \text{ matrix}\}$$

It is the "form" of an object which is invariant under rotation, not the variance–covariance structure. Biologically important parameters are non-identifiable under the superimposition schemes.

NUMERICAL EXAMPLES AND DISCUSSION

In this section, we provide numerical examples which clearly illustrate the inappropriateness of the procrustes estimators. This is especially glaring for the estimation of the variance–covariance parameter. We also discuss the intuitive reasoning behind its failure. To make these features most obvious ("an intraocular traumatic experience" to borrow a phrase due to Professor Berkson), we consider a situation which is somewhat extreme and hence magnifies the effects. Although as illustrated in Example 2, even under ideal conditions the procrustes estimators fail.

Example 1. In this example, we show that the variance–covariance structure obtained by using the procrustes estimators is wrong.

We generated 130 random geometrical objects as described previously using the following parameters. The number 130 was based on the maximum number of objects allowed by the software used for procrustes analysis. Each object has four landmarks and is two-dimensional.

$$M = \begin{bmatrix} 0 & 5 \\ 40 & 0 \\ 0 & -5 \\ -40 & 0 \end{bmatrix}, \Sigma_D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\Sigma_k = \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 10 & 0 & -9.999 \\ 0 & 0 & 0.01 & 0 \\ 0 & -9.999 & 0 & 10 \end{bmatrix}$$

Note that landmarks 2 and 4 are perturbed substantially and in a correlated fashion, whereas landmarks 1 and 3 are almost left unperturbed. These data are shown in Fig. 2a, depicting the true variance–covariance structure. These data were then subjected to the least squares procrustes analysis (without affine components) using the GRF program supplied with the edited volume by Rohlf and Bookstein (1990). Figure 2b shows the coordinates of the scaled, rotated and translated figures such that they minimize the procrustes distance. It is obvious that the variance–covariance structure after these operations is vastly different than the original one from which the data were created. In fact the Σ_k^* is given by:

$$\Sigma_k^* = \begin{bmatrix} 0.0064 & 0 & 0 & 0 \\ 0 & 10 & 0 & -9.98 \\ 0 & 0 & 0.0064 & 0 \\ 0 & -9.98 & 0 & 10 \end{bmatrix}$$

showing large variability around landmarks 2 and 4 and no variability around landmarks 1 and 3. Note that this is slightly different than Σ_k because this is the singular version (described previously) which is estimable. For numerical comparison, the procrustes estimate of Σ_k^* (Goodall, 1991, Eq. 10.2) is given by:

$$\hat{\Sigma}_k^* = \begin{bmatrix} 0.4935 & -0.0068 & -0.4795 & -0.0069 \\ -0.0068 & 0.0124 & -0.0022 & -0.0034 \\ -0.4795 & -0.0022 & 0.4825 & -0.0012 \\ -0.0069 & -0.0034 & -0.0012 & 0.0115 \end{bmatrix}$$

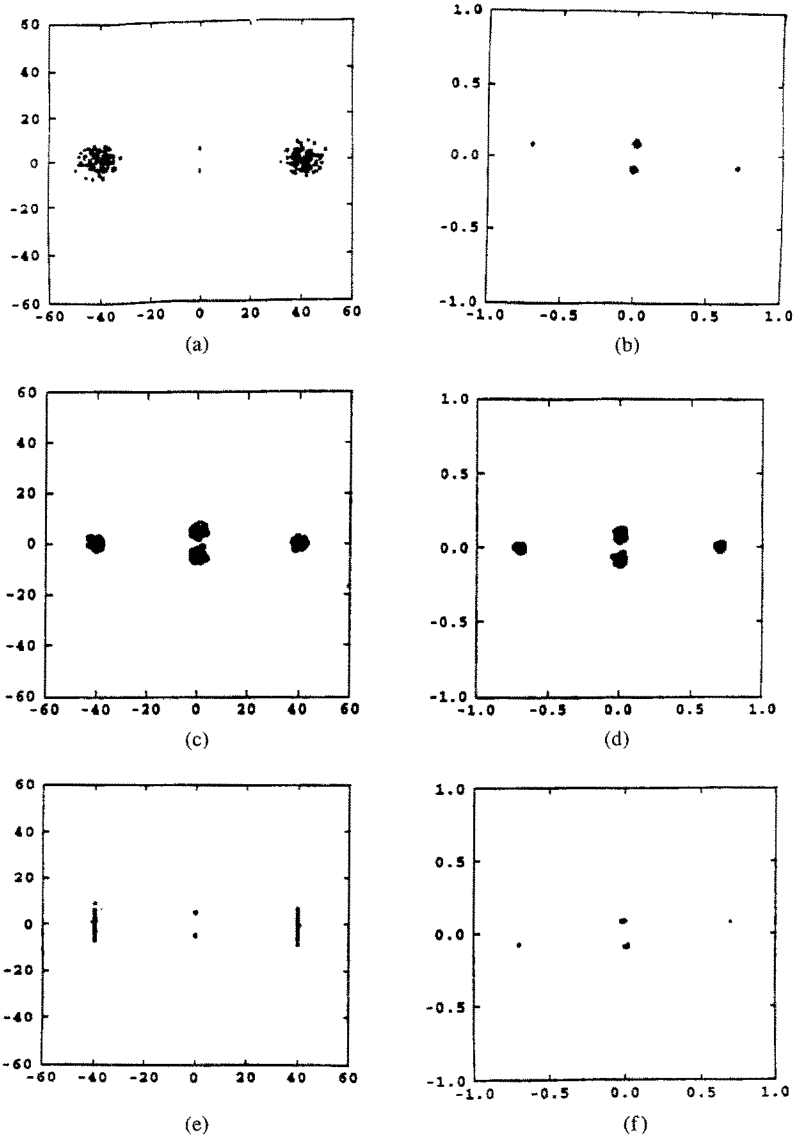

Fig. 2. The left-hand column of these graphs depict the true variance covariance structure in the data whereas the right-hand column depicts the variance–covariance structure estimated by the procrustes estimators using the data in the left-hand column. It is clear that procrustes estimators tend to reduce the variability around those landmarks that are far from the centroid and increase the variability around those landmarks that are closer to the centroid. The pair (e) and (f) shows that Σ_D is also incorrectly estimated by these estimators. Note that procrustes figures are scaled figures and hence the scales in two columns are different.

Since this is obtained after scaling this should have been a scaled version of the true Σ_k^* (or being an estimate, at least be close to it in its pattern). This clearly is not the case. Any testing procedures based on this estimator thus would lead to incorrect decisions.

The same data were analyzed using the distance based estimators described previously. The estimate of Σ_k^* is given by:

$$\hat{\Sigma}_k^* = \begin{bmatrix} 0.0038 & -0.0327 & -0.0035 & -0.0325 \\ -0.0327 & 9.2459 & -0.0200 & -11.4899 \\ -0.0035 & -0.0200 & 0.0095 & 0.2590 \\ -0.0325 & -11.4899 & 0.0259 & 8.7586 \end{bmatrix}$$

This obviously is closer to the true Σ_k^*.

Example 2. This example illustrates the same effect in the situation where $\Sigma_k = \sigma^2 I$. This is perhaps the most ideal situation for procrustes analysis. We used the same mean form and Σ_D as in Example 1 and 130 objects were created. Figures 2c and d illustrate the true variance covariance structure and the one obtained by the procrustes analysis. Clearly the variability around the landmarks that are away from the centroid is reduced whereas the variability around the landmarks closer to the centroid is increased. Numerical estimate supports this observation.

Example 3. This example illustrates the point that Σ_D may also be estimated incorrectly by these estimators. For this situation, we used the same values of M and Σ_k as in Example 1 but changed Σ_D to:

$$\Sigma_D = \begin{bmatrix} 0.001 & 0 \\ 0 & 1 \end{bmatrix}$$

Figures 2e and f illustrate the true variation and the procrustes estimate of it. The estimate not only decrease the variability around landmarks 2 and 4 and increases the variability around landmarks 1 and 3 but also claims that most of the variability around landmarks 1 and 3 is in the horizontal direction. This is completely different than the truth.

The above examples clearly show that almost any kind of statistical inference which uses the estimates of variance structure is going to go astray if procrustes estimators are used. This also casts doubt on the iteratively weighted least squares algorithm suggested in Goodall (1991, Section 10).

The intuitive reasoning behind the failure of procrustes estimators is the following. The procrustes fitting criterion which minimizes the sum of squared distances between the corresponding landmarks is such that the gain by fitting those landmarks that are farther away from the centroid tends to be much larger than the cost this type of rotation incurs by not fitting those landmarks which lie closer to the centroid. It thus tends to match outlying landmarks much more than the landmarks that are closer to the centroid. This is why the variability around the outlying landmarks is reduced drastically and the variability around the landmarks closer to the centroid is increased as illustrated in Fig. 2, irrespective of the true variability.

The inconsistency in the estimator of shape is difficult to illustrate numerically. Goodall (1991) claims consistency and efficiency of the procrustes shape estimator without any formal proof. Following intuitive example suggests that this claim may be unjustifiable. Considering the same mean form as in Example 1, let us assume that landmarks 1 and 3 are not perturbed at all and landmarks 2 and 4 are perturbed only in the vertical direction and in a perfectly correlated fashion such that if landmark 2 has a positive y-coordinate, landmark 4 has exactly the same negative y-coordinate. By doing this we make sure that the centroid for all the random figures is the origin. Following the intuitive reasoning given above, it is clear that procrustes rotations would rotate these figures such that landmarks 2 and 4 are closer to the x-axis at the cost of rotating landmarks 1 and 3 along an arc (see Fig. 3). Now if one takes the coordinatewise average of these rotated figures, the average for landmark 2 is necessarily larger than 40 whereas for landmark 4 it is smaller than -40. Similarly landmarks 1 and

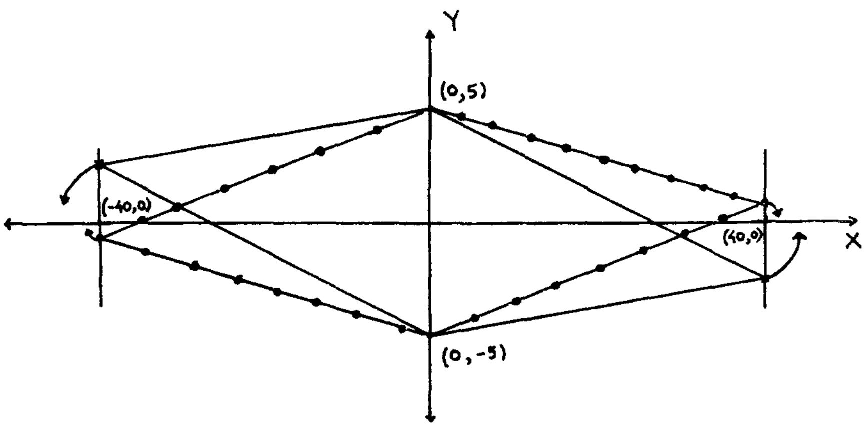

Fig. 3. An example showing the reasoning behind the failure of procrustes and edge superimposition estimators. This figure shows two random 4 landmark objects, one denoted by solid line and the other by line with solid dots. Note that after rotation all the points corresponding to landmarks 2 and 4 fall beyond 40 and -40 on the horizontal axis. Thus the coordinatewise average of these points falls beyond 40 and -40. Similarly points corresponding to landmarks 1 and 3 fall on an arc average of which falls within -5 and 5. Thus edge (2, 4) is estimated to have larger length than 80 whereas edge (1, 3) is estimated to have length smaller than 10. This thus estimates the shape of the quadrangle incorrectly. Note that the x and y scales are not equal for the sake of fitting the figure on the page.

3 have the average smaller than 1 and larger than -1. Thus the shape estimate so obtained would be incorrect.

The same example also shows why the edge superimposition estimator gives a wrong estimate of shape. Suppose one matches edge joining landmarks 2 and 4. Since for each random object generated under this model this edge necessarily has length larger than 80 (hypotaneous of a right angle triangle is longer than the base), the coordinatewise average or equivalently the average of these lengths is larger than 80 and similarly average of the edge joining landmarks 1 and 3 is smaller than 10. Hence the proportions of the edges (i.e., shape of the object) are not the same as in the true mean shape. The result in the section on inconsistency of these estimators is a formalization of this logic for a general perturbation model.

In summary, the conclusions of this paper can be described as follows:

(a) Distance based methods provide statistically correct, computationally simple estimators of mean form and the variance covariance structure. The mean form estimators are robust against model specification as shown theoretically in previous sections. These methods can also deal with the practically important problem of missing data very easily (section on missing landmarks).

(b) Statistical properties such as consistency are easily derivable for these estimators. Small to medium sample behavior needs to be studied. Our own philosophy, however, is that: If the sample size is small, one should perhaps only do exploratory analysis because precision of most statistical procedures is not good enough to warrant precise scientific inferences. One may possibly rely on unrealistic statistical models (Lele and Richtsmeier, 1990) to gain a (mostly) false sense of security. If the sample size is medium to large, one can use distance based estimators. However for testing or generating confidence intervals for various quantities of scientific interest one is perhaps better off using nonparametric bootstrapping (e.g., Lele and Richtsmeier, 1991).

(c) The maximum likelihood estimators based on exact shape densities (e.g., Mardia and Dryden, 1989) are usually difficult to obtain computationally and are possibly not very robust against model specification. However, if one truly believes in the model, one can use the distance based estimators suggested in this paper as the starting values in the numerical maximization routines to obtain maximum likelihood estimators.

(d) If one truly believes in the Gaussian model described herein, one can improve the efficiency of the distance based estimators by finding the maximum likelihood estimators of MM^T and Σ_k^* under the Wishart density with distance based estimators as the starting values in the maximization routine. Of course, the Wishart distribution given is singular and hence does not possess a density. However, if the sample size n is such that it exceeds K/D, the number of landmarks divided by the dimension of the object, then $\Sigma_{i=1}^{n} B(X_i)$ has a nonsingular Wishart distribution with parameters (nD, nMM^T, Σ_k^*). Hence in principle, consistent, asymptotically efficient maximum likelihood estimation is feasible.

(e) One can also consider the following modification of the model given previously for generating random objects.

$$X_i = b_i(M + E_i)\Gamma_i + t_i \quad \text{for } b_i > 0$$

That is, each object is also scaled randomly before being translated and rotated/reflected. Then, if $b_i > 0$, the distribution of the centered inner product matrices is given by:

$$B(X_i) = X_i X_i^T \sim \text{Wishart}_k(D, b_i^2 MM^T, b_i^2 \Sigma_k^*)$$

Thus we still retain the nuisance parameters b_i. One probably can estimate the parameters of interest by utilizing the methodology described in Kiefer and Wolfowitz (1956) or Lindsay (1983). Although the mathematics and computations are difficult.

(f) The superimposition methods for analysis of form or shape are in general scientifically unsatisfactory (Lele, 1991a). This paper shows that they are also statistically unsatisfactory. These are also highly computationally intensive as against the methods suggested in this paper. The estimation of the variance–covariance structure is particularly unsatisfactory. In our opinion, superimposition methods should not be used for any statistical analysis.

ACKNOWLEDGMENT

This paper is an expanded version of the paper presented at the Statistics and Computer Science Interface Meetings, April 1991. Kind encouragement and many useful comments by Professor Robert J. Serfling are gratefully acknowledged. Dr. Joan T. Richtsmeier helped conduct the numerical analysis as well as provided valuable advice on the missing landmarks section. Mr. Tao Wang wrote the computer programs for distance based estimators. The author is also grateful to Professors Robert Ehrlich, William Full, and the anonymous referees for many important and useful comments. This research was partially supported by National Science Foundation (DBS9209083).

REFERENCES

Arnold, S. F., 1981, Theory of Linear Models and Multivariate Analysis: John Wiley and Sons, New York.

Bookstein, F., 1986, Size and Shape Spaces for Landmark Data in Two Dimensions: Stat. Sci., v. 1, p. 181-242.

Bookstein, F., and Sampson, P., 1990, Statistical Methods for the Geometric Components of Shape Change: Comm. Stat. Theory Methods, v. 19, p. 1939-1972.

Campbell, G., 1986, Comments on "Size and Shape Spaces for Landmark Data in Two Dimensions" by F. L. Bookstein: Stat. Sci., v. 1, p. 227-228.

Chung, K. L., 1974, A Course in Probability Theory: Academic Press, New York.

deGunst, M. C. M., 1987, On the Distribution of General Quadratic Functions in Normal Vectors: Statistica Neerlandica, p. 245-251.

Dik, J. J., and deGunst, M. C. M., 1985, The Distribution of General Quadratic Forms in Normal Variables: Statistica Neerlandica, p. 14-26.

Goodall, C., 1991, Procrustes Methods in the Statistics Analysis of Shape: J. Roy. Stat. Soc. Ser. B, v. 53, p. 285-339.

Goodall, C., and Bose, A., 1987, Models and Procrustes Methods for the Analysis of Shape Difference: Proc. 19th Symp. Interface Between Computer Science and Statistics, Philadelphia, PA, pp. 86-92.

Gower, J. C., 1966, Some Distance Properties of Latent Root and Vector Methods in Multivariate Analysis: Biometrika, v. 53, p. 315-328.

Gower, J., 1975, Generalized Procrustes Analysis: Psychometrika, v. 40, p. 33-50.

Johnson, N., and Kotz, S., 1970, Distributions in Statistical—Continuous Univariate Distributions 2: Houghton Mifflin Company, Boston.

Kiefer, J., and Wolfowitz, J., 1956, Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters: Ann. Math. Statist., v. 27, p. 887-906.

Langron, S. P., and Collins, A. J., 1985, Perturbation Theory for Generalized Procrustes Analysis: J. Roy. Statist. Soc. Ser. B, v. 47, p. 277-284.

Leakey, M. G., Leakey, R. E., Richtsmeier, J. T., Simons, E. L., and Walker, A. C., 1991, Similarities in Aegyptopithecus and Afropithecus Facial Morphology: Folia Primatol., v. 56, p. 65-85.

Lele, S., 1991a, Some Comments on Coordinate Free and Scale Invariant Method in Morphometrics: Am. J. Phys. Anthropol., v. 85, p. 407-418.

Lele, S., 1991b, Comments on Goodall's Paper: J. Roy. Stat. Soc. B, v. 53, p. 334.

Lele, S., 1992, A Quantitative Method for the Reconstruction of Missing Landmarks on Biological Objects: Unpublished manuscript.

Lele, S., and Richtsmeier, J. T., 1990, Statistical Models in Morphometrics: Are They Realistic: Syst. Zool., v. 39, n. 1, p. 60-69.

Lele, S., and Richtsmeier, J. T., 1991, Euclidean Distance Matrix Analysis: A Coordinate Free Approach for Comparing Biological Shapes: Am. J. Phys. Anthropol., v. 86, p. 415-427.

Lele, S., and Richtsmeier, J. T., 1992, On Comparing Biological Shapes: Detection of Influential Landmarks: Am. J. Phys. Anthropol., v. 87, p. 49-65.

Lindsay, B. G., 1983, The Geometry of Mixture Likelihoods, Part II: The Exponential Family: Ann. Stat., v. 11, p. 783-792.

Mardia, K. V., and Dryden, I., 1989, The Statistical Analysis of Shape Data: Biomtk., v. 76, p. 271-282.

Mardia, K. V., Kent, T., and Bibby, J. M., 1979, Multivariate Analysis: Academic Press, New York.

Neyman, J., and Scott, E. J., 1948, Consistent Estimates Based on Partially Consistent Observations: Econometrika v. 16, p. 1–32.

Rao, C. R., 1973, Linear Statistical Inference and Its Applications: John Wiley, New York.

Rohlf, J., and Bookstein, F., ed., 1990, Proceedings of the Michigan Morphometrics Workshop: Special Publication No. 2, Museum of Zoology, University of Michigan, Ann Arbor, Michigan.

Rohlf, F. J., and Slice, D., 1990, Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks: Syst. Zool., v. 39, p. 40–59.

Roth, V. L., 1988, The Biological Basis of Homology, *in* C. J. Humphries (Ed.), Ontogeny and Systematics: Columbia University Press.

Serfling, R. J., 1980, Approximation Theorems of Mathematical Statistics: John Wiley and Sons, New York.

Van Valen, L. M., 1982, Homology and Causes: J. Morphol., v. 173, p. 305–312.