the upper margin of the braincase with respect to the lower margin. It is not at all equivalent to the second weakest principal warp (dotted lines, frames (C) and (D)). The bending energy eigenanalysis has extracted these large scale patterns of shape covariance by explicitly weighting empirical covariance patterns inversely to geometric localizability. Other equally plausible geometric patterns, such as bending of the upper or lower structures, are not observed to bear any sample variance.

The example suggests the descriptive possibilities inherent in accommodating the metric geometry of Kendall's shape space to a biological subject matter. One can imagine other modifications of the metric in response to other contexts than the biometric. For instance, one can imagine the statistical study of the positions of a robot arm. When the state of the linkage is coded by the coordinates of its joints, then because certain parts of the robot are rigid, an appropriate measure of "distance" would be somewhat altered from the Procrustes. In another sort of constraint, certain "landmarks" might represent the loci of curves in the data—boundary arcs not otherwise labeled—and would thus be "deficient" by one coordinate; again the Procrustes metric needs to be modified. In a study of schools of fish, or flocks of birds, an appropriate shape metric might be the Cartesian product of a biological shape space by a hydro- or aerodynamic one (for the V of migrating geese, for instance). Yet other modifications would arise when the points of Kendall's space are "colored" in classes whose separate patterns cannot be usefully studied without reference to their interpenetration, as in problems of multispecies ecology. These and other possibilities represent an enrichment of the metric geometry of shape space within the global purview pursued so sparely and elegantly by David Kendall.

ACKNOWLEDGMENTS

Preparation of this comment was partially supported by Grant GM-37251 from the National Institutes of Health to F. L. Bookstein. Christopher Small commented usefully on an earlier draft. I and many of the rest of us in shape statistics are grateful to Colin Goodall for arranging a workshop around Kendall's Wilks Lectures, Princeton, May 1987. Part of this comment was presented at that workshop in even more preliminary form.

ADDITIONAL REFERENCES

BOOKSTEIN, F. L. (1988a). Toward a notion of feature extraction for plane mappings. In *Proc. Tenth International Conference on Information Processing in Medical Imaging* (C. de Graaf and M. Viergever, eds.) 23–43. Plenum, New York.

BOOKSTEIN, F. L. (1988b). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Machine Intelligence*. To appear.

BOOKSTEIN, F. L. (1989). *Morphometric Tools for Landmark Data*. In preparation.

RIOLO, M. L., MOYERS, R. E., MCNAMARA, J. S. and HUNTER, W. S. (1974). *An Atlas of Craniofacial Growth*. Center for Human Growth and Development, Univ. Michigan.

YAGLOM, I. M. (1979). *A Simple Non-Euclidean Geometry and its Physical Basis*. Springer, New York.

Comment

Christopher G. Small

With a high standard of rigor David Kendall has given us an interesting survey of the theory of shape analysis that he has pioneered with the help of others over the last decade. This work is now of sufficient volume that the many topics discussed in this survey can be only briefly touched upon. I certainly hope that this paper is a stimulus to additional consideration of this topic by statisticians. It may well be that on future occasions the topologists will have to introduce their theory of shape with preparatory remarks to the effect that it is not to be confused with the growing statistical theory of shape.

At first glance, this paper might seem to have much in common with the differential geometric techniques in statistics that are associated with Amari (1985) and others. However, despite the abstraction of some of the theory, the methods of Kendall are essentially data analytic rather than model theoretic: the differential geometry is on the sample space not the parameter space. So how much differential geometry must the data analyst know in order to implement the techniques that are described in this paper? The answer is largely dependent on the amount of software

Christopher G. Small is Associate Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

that can be developed to analyze shapes, because the differential geometry need not be duplicated by the data analyst if it is built into the software. Software that is analogous to the techniques available for projection pursuit would be especially useful here because it is impossible for most of us to visualize global properties of shape manifolds. So it would be helpful to be able to search a shape space interactively through various two-dimensional (or even three-dimensional) projections. The appropriate tools would then be able to analyze shapes in much the same way that multivariate data sets can be analyzed by "Grand Tours." (See Buja and Asimov, 1985 and Hurley, 1987.)

It is the job of a discussant to provide a different perspective on a paper. As I have been involved in the work with D. G. Kendall this is not an easy task. Nevertheless, I would like to indicate why there is room for flexibility in the choice of geometries available to the statistician. Each geometry can be examined in light of criteria such as naturalness, computational convenience or graphical convenience, etc. Although among such geometries the shape space geometry of D. G. Kendall is by far the most developed, and perhaps the most important, other geometries deserve consideration.

Need the statistician consider geometry at all? The answer must be yes, because there is a natural interplay between geometrical and statistical considerations. Suppose for example we are to find a measure of location for a multivariate data set $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k$. A family of such measures based on minimum distance methods is given by using that point $\mathbf{x}$ which minimizes $\sum \rho(\mathbf{y}_i, \mathbf{x})$ where ρ is a metric that represents the geometry in use. If $\rho(\mathbf{y}, \mathbf{x}) = \| \mathbf{y} - \mathbf{x} \|^2$ then $\mathbf{x}$ is the centroid of the data set. However, the centroid is sensitive to outliers. So if a more robust measure of location is desired the metric $\rho(\mathbf{y}, \mathbf{x}) = \| \mathbf{y} - \mathbf{x} \|$ can be used. In this case $\mathbf{x}$ becomes the spatial median whose properties were investigated by Brown (1983). This example shows how a departure from the L_2 metric of "least squares" used in multivariate analysis leads to differing statistical properties in the resulting estimator.

To a statistician, the centroid of a data set is a sensible measure of location based upon the traditions of multivariate theory. Similarly, sums of squares are natural measures of scale. However, other geometries are worthy of consideration in contexts in which their properties are more useful than the standard geometry. Moreover, the assumption that there exists a unique natural geometry on the space of shapes presupposes that there exists a unique natural geometry on the space in which the observations lie. Yet different contexts seem to dictate different geometries, and differential geometric considerations only tell us how to transfer the geometry across from the space in which the observations lie to the shape space. A distinction must be made between the Euclidean space in which points lie (Kendall's $\mathbf{R}^m$) and the Euclidean space in which a data set is represented (Kendall's $\mathbf{R}^{m \times k}$). When $m \leq 3$ the geometry of the former usually has a physical interpretation. The geometry of the latter is typically a mathematical construction.

The following example illustrates the point. Suppose X_1, X_2 and X_3 are three independent random variables that are uniformly distributed over some common interval, say [0, 1]. The shape σ of the three points is an element of a unit circle $S^1(1)$. If we compute the shape density on the unit circle, we find it to be analytic on six arcs that correspond to the $3! = 6$ possible rankings of the three variables. Alternatively, however, we could introduce the L_∞ norm on $\mathbf{R}^3$ by setting $\| \mathbf{x} \|_\infty = \max(|x_1|, |x_2|, |x_3|)$. The resulting induced metric on $\mathbf{R}^3$ can in turn be related to a metric space on equivalence classes of shapes. Let $x_{(i)}$ be the ith order statistic. Define $m = [x_{(1)} + x_{(3)}]/2$ and $r = [x_{(3)} - x_{(1)}]/2$. Then the point

$$\left[\frac{x_1 - m}{r}, \frac{x_2 - m}{r}, \frac{x_3 - m}{r}\right]$$

is a representation of the shape of the triplet (x_1, x_2, x_3) and lives on a closed path made of six line segments (i.e., a bent hexagon) that is the shape space in this case. The corners of the hexagon correspond to the places in the former shape space $S^1(1)$ where the shape density failed to be analytic. The bent hexagon can be described by removing two diametrically opposite vertices of a cube and joining the six remaining vertices with those edges connecting neighboring vertices among the six. The induced distribution on the hexagon can be seen to be uniform and therefore very easy to work with. Calculations are further simplified if the labels of the points are ignored in which case all shapes are represented on a single line segment that is the fundamental region on the hexagon. For a larger number k of points in dimension 1, the resulting shape space will be a continuous image of a sphere of dimension $k - 2$ lying within the boundary of a k dimensional hypercube. It will be made of the union of $k(k-1)$ hypercubes of dimension $k - 2$ attached at their boundaries. Once again, the induced distribution on the shape space will be uniform. The smoothness of the shape manifolds Σ_2^k is deceptive for the induced distributions of shapes of points uniformly distributed in a planar region. In such cases, the sets on which the shape density fails to be analytic are quite complex. Therefore a geometry in which the singularities are inherent in the shape space itself becomes plausible in light of the computational difficulty of such models on smooth manifolds.

The existence of numerous geometries for shapes does not give carte blanche to the researcher to use any geometric representation whatsoever. As always in data analysis, the geometry used must serve the goals of the statistical analysis if it is to be anything more than a mathematical theory. The use of a measure of distance between shapes (the "geometry") generates clustering and classification algorithms and test statistics based upon minimum distance methods. If the measure of distance used is artificial the resulting statistical analysis will surely not be any less so. Geometries give rise to statistical methods which can, in turn, be evaluated in the context of criteria for good inference or data analysis. This is as true for shape analysis as it is for multivariate analysis.

In defining metrics within the space $\mathbf{R}^{m\times k}$ or some subset of that space the nature of shape "kinematics" should be considered. Shape changes or shape differences can arise through perturbations of the individual k points. Alternatively, shape changes can arise through global transformations of the space $\mathbf{R}^m$ which a fortiori induce a change on the shape of the individual points of the data set. The usual Euclidean metric seems inappropriate for the latter case. For example, suppose we assume that changes in the shape of a data set arise from a global transformation T of $\mathbf{R}^{m\times k}$. We could measure the distance from one data set to another by means of a measure of distance of T from the identity transformation. The induced metrics on the space of shapes from these methods have a very different character from Kendall's geometry. For example, suppose $k = 3$ and $m = 2$. Any three points which are not collinear can be transformed to any other such set by composition of an affine transformation T. A measure of distance (in fact a pseudometric) of this transformation from the identity defined by Bookstein (1986) induces the riemannian geometry of the hyperbolic plane on the shape space in which Kendall's great circle of collinear shapes becomes the circle at infinity of the hyperbolic plane.

In his survey, D. G. Kendall has provided some interesting inversion results, showing that for three points independently and uniformly distributed in a compact convex set the induced distribution of shape determines the convex set modulo a shape preserving transformation. The argument, based upon the set where the shape density fails to be analytic, is quite delicate. Unfortunately we cannot expect to detect discontinuities in the higher derivatives of the shape density through statistical means. So the result does not transfer immediately into a statistical test although I expect that this could be done. The delicacy of the inversion is rather different from that of the inversion theorems that I have been working on over the last few years (Small, 1983, 1984, 1985) which are based upon transform techniques for larger sets of points.

Attention has recently been turned to the representation of shapes of objects more complex than finite sets of points. Bookstein has proposed that the shapes of objects can be studied by choosing a representative set of points (called "landmarks") on the object and studying the shape of this set of points. Although this works for biological structures in which there is some differentiation in terms of function of the various sites on the structure, the problem of representing complex shapes using finite sets of points is complicated in general by the lack of obvious landmarks. In such cases, multidimensional generalizations of Bernstein polynomials may be of value in representing complex shapes through finite sets of points. Note that the problem here is not simply to find a finite dimensional parametrization of complex shapes, but to ensure that there is a geometrical relationship between the coordinates of the parametrization and the shape itself. Thus, for example, the coefficients of a polynomial parametrize the shape of the polynomial but do not do so in such a way that there is a clear geometrical interpretation of the shape to be found within them. Let Δ^n be the n-dimensional simplex of all points $(p_1, \cdots, p_{n+1})$ such that $p_1, p_2, \cdots, p_{n+1} \geq 0$ and $\sum p_i = 1$. Suppose $\mathbf{g}(\mathbf{p}) = [g_1(\mathbf{p}), \cdots, g_m(\mathbf{p})]$ defines a continuous m-dimensional image $\mathbf{g}(\Delta^n)$ of the simplex. The image $\mathbf{g}(\Delta^n)$ can be approximated by a polynomial image $\mathbf{g}_j(\Delta^n)$ defined by

$$\mathbf{g}_j(p_1, \cdots, p_{n+1}) = E\left[\mathbf{g}\left(\frac{X_1}{j}, \cdots, \frac{X_{n+1}}{j}\right)\right]$$

where $X_1, \cdots, X_{n+1}$ have a multinomial distribution with $\sum X_i = j$ and $E(X_i) = jp_i$. Then $\mathbf{g}_j$ is a jth degree polynomial that uniformly approximates $\mathbf{g}$. The image $\mathbf{g}_j(\Delta^n)$ is determined by the set of points $\mathbf{g}(x_1/j, \cdots, x_{n+1}/j)$ where $0 \leq x_i$ and $\sum x_i = j$. At this early stage I can only speculate as to the value of a theory of shape for random paths in applications such as dynamical systems.

The spaces that Kendall has called "size and shape spaces" may well turn out to be more useful than shape spaces themselves. Commonly in statistical problems it is impossible to ignore size variables without losing information that is important to the understanding of the data. The geometry of an object possessing no intrinsic position or orientation falls naturally into such a space. For cases where a group acts upon a riemannian manifold, the resulting quotient space is termed a "shape space" by Kendall. The generalization is an important one, although the terminology tends to separate the subject from an area that is well established in the statistical literature: the theory of maximal invariance. Thus a shape in this generalized sense is also a maximal invariant under the action of the group $\mathscr{G}$. Such maximal invariants for data on riemannian manifolds are not uncommon in statistics. For example, data in directional statistics live on a one- or two-dimensional sphere for which the most useful group to generate transformation models is the rotation group. For models in which the rotation group generates a nuisance parameter, questions involving the testing of a concentration parameter in the absence of knowledge of the nuisance parameter require the reduction to the maximal rotation invariant. Fraser (1968) has emphasized the importance of transformation models and the fibers of data sets equivalent under the action of a group. Some relationships with the statistics of shape are developed in Small (1983).

I would like to close these comments with some remarks of a more specialized technical nature. The elegance of D. G. Kendall's theory of shape is especially clear for data sets in dimensions 1 and 2. In higher dimensions, singularities start to emerge. Although these singularities are not obviously detrimental to a theory of shape, they do detract from the elegance of the representation. Even in dimensions 1 and 2, the shape spaces are more easily constructed and represented than the corresponding size and shape spaces. The reason for the elegance of the shape space for the cases where $m = 1, 2$ is that in these cases a shape preserving transformation can be uniquely decomposed into two transformations that correspond to multiplication and addition in the real line and complex plane for the respective dimensions. For both $m = 1$ and $m = 2$ the group of shape preserving transformations is a solvable group with a dimension (i.e., number of degrees of freedom), which is an integral multiple of m. But in dimension $m = 3$ this fails to be the case. The group of shape preserving transformations is of dimension 7, which is not a multiple of m.

Let me conclude my remarks by congratulating David Kendall on some very interesting work. The new directions that are sketched in this paper seem to be promising for the analysis of geometric data of various kinds and from various sources. I hope that much more is forthcoming.

ADDITIONAL REFERENCES

AMARI, S. (1985). *Differential-Geometric Methods in Statistics.* Springer, New York.

BROWN, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30.

BUJA, A. and ASIMOV, D. (1985). Grand tour methods: An outline. *Computer Science and Statistics: Proceedings of the 17th Symposium on the Interface* (D. M. Allen, ed.) 63–67. North-Holland, Amsterdam.

FRASER, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.

HURLEY, C. (1987). The data viewer: A program for graphical data analysis. Ph.D. dissertation and Technical Report, Dept. Statistics, Univ. Washington.

SMALL, C. G. (1983). Characterization of distributions from maximal invariant statistics. *Z. Wahrsch. verw. Gebiete* **63** 517–527.

SMALL, C. G. (1984). A classification theorem for planar distributions based on the shape statistics of independent tetrads. *Math. Proc. Cambridge Philos. Soc.* **96** 543–547.

SMALL, C. G. (1985). Decomposition of models whose marginal distributions are mixtures. *Canad. J. Statist.* **13** 131–136.

Comment: Some Contributions to Shape Analysis

Kanti V. Mardia

There are no words to express the profound depth of Kendall's work. I have been working in this area intermittently since 1976 and I believe his fundamental work (as well as Bookstein, 1986) has opened up the field.

Kanti V. Mardia is Professor and Head, Department of Statistics, University of Leeds, Leeds LS2 9JT, England.

Bookstein (1986) has used the model for shape analysis assuming that the points are distributed independently as $N_2(\mu_i, \sigma^2\mathbf{I})$, $i = 1, 2, \cdots, p$. Consider $p = 3$. Let $\mathbf{x}$ be the point in Kendall's spherical shape space from these three points with ℓ representing the corresponding point in Kendall's space from μ_i's. Let $\bar{\mu}$ be their mean vector. Then using Mardia and Dryden (1989), it can be shown that the probability element of $\mathbf{x}$ is given by

$$\{1 + \varkappa(\ell'\mathbf{x} + 1)\}e^{\varkappa(\ell'\mathbf{x}-1)}dS, \quad \mathbf{x} \in S_2,$$