22
Some perspectives on FDA

22.1 The context of functional data analysis

We conclude this volume with some historical remarks and pointers to bibliographic references which have not been included in the main course of our development. We are, of course, acutely aware that many branches of statistical science consider functional models and the data that go with them. FDA has a long historical shadow, extending at least back to the attempts of Gauss and Legendre to estimate a comet's trajectory (Gauss, 1809; Legendre, 1805). So what we offer here is perhaps little more than a list of personal inspirations. In addition we suggest some directions for further research.

22.1.1 Replication and regularity

While we want to leave the question of exactly what constitutes FDA soft around the edges, functional data problems as we have described them have two general features: replication and regularity. These are intimately related. Both permit the use of information from multiple data values to identify patterns; replication implies summaries taken across different observations, while regularity allows us to exploit information across argument values.

Replication is closely bound up with the key concept of a functional observation as a single entity, rather than a set of individual numbers or values. The availability of a sample of N related functional observations then leads to an interest in structure and variability in the data that requires more than one observation to detect. This is in contrast with much of the literature on nonparametric regression or curve estimation, where the focus is on estimating a single curve.

Functional principal components analysis, regression analysis, and canonical correlation, like their multivariate counterparts, characterize variation in terms of features that have stability across replicates. Likewise, principal differential analysis and the use of an estimated linear differential operator for smoothing presume a model structure that belongs to the entire sample. Even curve registration aims to remove one important source of inter-curve variation so as to render more obvious the structure that remains.

Regularity implies that we exploit the smoothness of the processes generating the data, even though these data usually come to us in discrete form. The assumption that a certain number of derivatives exist has been used in most of the analyses that we have considered. The roughness penalty approach used throughout the book controls the size of derivatives and mixtures of derivatives of the functional parameters that we have estimated. In this way we stabilize estimated principal components, regression functions, monotone transformations, canonical weight functions, and linear differential operators.

Are there more general concepts of regularity that would aid FDA? For example, wavelet approaches to smoothing briefly discussed in Section 3.6.1 are probably relevant, because of their ability to accommodate notions of regularity that, nevertheless, allow certain kinds of local misbehavior.

Independent identically distributed observations are only one type of regularity. For example, can we use the replication principle implicit in stationary time series and where the values of the process are functions, to define useful FDAs? Besse and Cardot (1997) offer an interesting first step in this direction.

22.1.2 Some functional aspects elsewhere in statistics

Analysis of variance is often concerned with within-replication treatments. While an ANOVA design does not assume as a rule that these treatments correspond to variation over time or some other continuum, in practice this is often the case. Consequently texts on ANOVA such as Maxwell and Delaney (2003) pay much attention to topics that arise naturally when treatments correspond to events such as days, related spatial positions, and so on. Modifications taking account of a more complex correlational structure for the residuals and the use of contrasts to make inferences about linear, quadratic, and other types of trend across treatments are examples.

As we indicated at the end of Chapter 5, fields such as longitudinal data analysis (Diggle et al., 1994), analysis of repeated measurements (Keselman and Keselman, 1993 and Lindsey, 1993) and growth curve analysis are cognate to functional data analysis. Two classic papers that use principal components analysis to describe the modes of variation among replicated curves are Rao (1958) and Tucker (1958); Rao (1987) offers a summary of his and others' work on growth curves. Two more recent applications are Castro, Lawton and Sylvestre (1986) and Grambsch et al. (1995).

But these and the many other studies of curve structure do not give the regularity of the phenomena a primary role, placing more emphasis instead on replication. Likewise, empirical Bayes, hierarchical linear model, or multilevel linear model approaches do treat functional data in principle, with the added feature of using prior information, but the nature of the prior structure tends to be multivariate rather than functional. In particular, as we noted in Chapter 5, the estimation of a between–curve variance-covariance matrix whose order is equal to the number of basis functions used to represent the curves places severe limits on the the complexity of the functional variation.

Nevertheless, we expect that further research will show that the experience gained and tools developed in these collateral disciplines can be put to good use in FDA.

22.1.3 Functional analytic treatments of statistical methods

One topic clearly within the scope of FDA is the description of statistical methods using functional analysis. For example, principal components analysis is a technique that lends itself naturally to many types of generalization. The notion of the eigenanalysis of a symmetric matrix was extended to integral operators with symmetric kernels in the last century, and the Karhunen-Loève decomposition of more general linear operators (Karhunen, 1947; Loève, 1945) is essentially the singular value decomposition in a wider context.

Parzen's papers (1961, 1963) are classics, and have had a great influence on the spline smoothing literature by calling attention to the important role played by the reproducing kernel. Grenander (1981) contributed further development, Eaton (1983) provided a systematic coverage of multivariate analysis using inner product space notation, and Stone (1987) also proposed a coordinate-free treatment.

Applied mathematicians and statisticians in France have been particularly active in recasting procedures originally developed in a conventional discrete or multivariate setting into a functional analytic notational framework. Deville (1974) considered the PCA of functional observations with values in $\mathcal{L}^2$. Cailliez and Pagès (1976) wrote an influential textbook on multivariate statistics that was both functional analytic in notation and coordinate-free in a geometrical sense. This was a courageous attempt to present advanced concepts to a mathematically unsophisticated readership, and it deserves to be better known. Dauxois and Pousse (1976) produced a comprehensive and sophisticated functional analytic exposition of PCA and CCA that unhappily remains in unpublished form.

While the exercise of recasting the usual matrix treatments of multivariate methods into the more general language of functional analysis is intrinsically interesting to those with a taste for mathematical abstraction, it also defined directly the corresponding methods for infinite-dimensional or functional data. Some facility in functional analysis is a decided asset for certain aspects of research in FDA, as it already is in many other branches of applied mathematics.

22.2 Challenges for the future

We now turn to a few areas where there is clearly need for further research. These should be be seen as a small selection of the wide range of topics that a functional data analytic outlook opens up.

22.2.1 Probability and inference

The presence of replication inevitably invites some consideration of random functions and probability distributions on function spaces. Of course, there is a large literature on stochastic processes and random functions, but because of our emphasis on data analysis we have not emphasized these topics in the present volume.

We note, in passing, that functional observations can be random in a rather interesting variety of ways. We pointed out in Section 21.3 that the problem of spline smoothing is intimately related to the theory of stochastic processes defined by the nonhomogeneous linear differential equation $Lx = f$ where L is a deterministic linear differential operator and f is white noise. Should we allow for some stochastic behavior or nonlinearity in L? Is white noise always an appropriate model for f? Should we look more closely at the behavior of an estimate of f in defining smoothing criteria, FDAs, and diagnostic analyses and displays, exploiting this estimate in ways analogous to our use of residuals in regression analysis? There is a large literature on such *stochastic differential equations*; see, for example Øksendal (1995). Though stochastic differential equations are of great current interest in financial mathematics, they have had relatively little impact on statistics more generally. This seems like a way to go.

We discussed the extension of classical inferential tools such as the t-test or F-ratio to the functional domain. We often need simulation to assess the significance of statistics once we move beyond the context of inference for a fixed argument value t. For a rather different approach to inference that incorporates both theoretical arguments and simulation, see Fan and Lin (1998).

Because of the infinite-dimensional nature of functional variation, the whole matter of extending conventional methods of inference—whether parametric or nonparametric, Bayesian or frequentist—is one that will require considerable thought before being well understood. We consider that there is much to do before functional data analysis will have an inferential basis as developed as that of multivariate statistics.

22.2.2 Asymptotic results

There is an impressive literature on the asymptotic and other theoretical properties of smoothing methods. Although some would argue that theoretical developments have not always had immediate practical interest or relevance, there are many examples clearly directed to practical concerns. For a recent paper in the smoothing literature that addresses the issue of the positive interaction between theoretical and practical research, see, for example, Donoho et al. (1995).

Some investigations of various asymptotic distributional aspects of FDA are Dauxois et al. (1982), Besse (1991), Pousse (1992), Leurgans et al. (1993), Pezzulli and Silverman (1993), Silverman (1996) and Kneip and Engel (1995), for example. Nevertheless, theoretical aspects of FDA have not been researched in sufficient depth, and it is hoped that appropriate theoretical developments will feed back into advances in practical methodology.

22.2.3 Multidimensional arguments

Although we have touched multivariate functions of a single argument t, coping with more than one dimension in the domain of our functions has been mainly beyond our scope. But of course there is a rapidly growing number of fields where data are organized by space instead of or as well as time. Consider, for example, the great quantities of satellite and medical image data, where spatial dimensionality is two or three and temporal dependence is also of growing importance.

There is a large and growing literature on spatial data analysis; see, for example, Cressie (1991) and Ripley (1991). Likewise smoothing over two or more dimensions of variation is a subject of active research (Scott, 1992). In particular, Wahba (1990) has pioneered the extension of regularization techniques to multivariate arguments. In principle, there is no conceptual difficulty in extending our own work on FDA to the case of multivariate arguments by using the roughness penalties relevant to tensor or thin-plate splines. Indeed, the paper by Hastie et al. (1995) reviewed in Section 11.7 uses roughness penalty methods to address a functional data analysis problem with a spatial argument. However, there are questions about multivariate roughness penalty methods in FDA that require further research.

22.2.4 Practical methodology and applications

Clearly, much research is needed on numerical methods, as is evident when one considers the work on something as basic as the point-wise linear model underlying spline smoothing. We think that regularization techniques will play a strong role, in part because they are so intuitively appealing. But of course there are often simpler approaches that may work more or less as well.

It is our hope that this book will give impetus to the wider dissemination and use of FDA techniques. More importantly than any of the detailed methodological issues raised in this chapter, the pressing need is for the widespread use of functional data analytic techniques in practice.

22.2.5 Back to the data!

Finally, we say simply that it is the data that we have analyzed, and our colleagues who collected them, that are responsible for our understanding of functional data analysis. If what this book describes is found to deserve a name for itself, it will be because we have discovered, with each new set of functional data, challenges and invitations to develop new methods. Statistics shows its finest aspects when exciting data find existing statistical technology not entirely satisfactory. It is this process that informs this book, and ensures that unforeseen adventures in research await us all.