1
Introduction

1.1 Why consider functional data at all?

Functional data come in many forms, but their defining quality is that they consist of functions—often, but not always, smooth curves. In this book, we consider functional data arising in many different fields, ranging from the shapes of bones excavated by archaeologists, to economic data collected over many years, to the path traced out by a juggler's finger. The fundamental aims of the analysis of functional data are the same as those of more conventional statistics: to formulate the problem at hand in a way amenable to statistical thinking and analysis; to develop ways of presenting the data that highlight interesting and important features; to investigate variability as well as mean characteristics; to build models for the data observed, including those that allow for dependence of one observation or variable on another, and so on.

We have chosen case studies to cover a wide range of fields of application, and one of our aims is to demonstrate how large is the potential scope of functional data analysis. If you work through all the case studies you will have covered a broad sweep of existing methods in functional data analysis and, in some cases, you will study new methodology developed for the particular problem in hand. But more importantly, we hope that the readers will gain an insight into functional ways of thinking.

What sort of data come under the general umbrella of functional data? In some cases, the original observations are interpolated from longitudinal data, quantities observed as they evolve through time. However, there are many other ways that functional data can arise. For instance, in our study of children with attention deficit hyperactivity disorder, we take a large number of independent numerical observations for each child, and the functional datum for that child is the estimated probability density of these observations. Sometimes our data are curves traced out on a surface or in space. The juggler's finger directly traces out the data we analyze in that case, but in another example, on the characteristics of examination questions, the functional data arise as part of the modeling process. In the archaeological example, the shape of a two-dimensional image of each bone is the functional datum in question. And of course images as well as curves can appear as functional data or as functional parameters in models, as we show in our study of electromyography recordings and speech articulation.

The field of functional data analysis is still in its infancy, and the boundaries between functional data analysis and other aspects of statistics are definitely fuzzy. Part of our aim in writing this book is to encourage readers to develop further the insights—both statistically and in the various subject areas from which the data come—that can be gained by thinking about appropriate data from a functional point of view. Our own view about what is distinctive about functional data analysis should be gained primarily from the case studies we discuss, as summarized in Section 1.3, but some specific remarks are made in Section 1.4 below.

1.2 The Web site

Working through examples for oneself leads to deeper insight, and is an excellent way into applying and adapting methods to one's own data. To help this process, there is a Web site associated with the text. The Web site contains many of the data sets and analyses discussed in the book. These analyses are *not* intended as a package or as a "cookbook", but our hope is that they will help readers follow the steps that we went through in carrying out the analyses presented in the case studies. Some of the analyses were carried out in MATLAB and some in S-PLUS.

At the time of printing the Web site is linked to the Springer Web site at `www.springer-ny.com`.

1.3 The case studies

In this section, the case studies are briefly reviewed. Further details of the context of the data sets, and appropriate bibliographic references, are given in the individual chapters where the case studies are considered in full. In most of them, in addition to the topics explicitly mentioned below, there is some discussion of computational issues and other fine points of

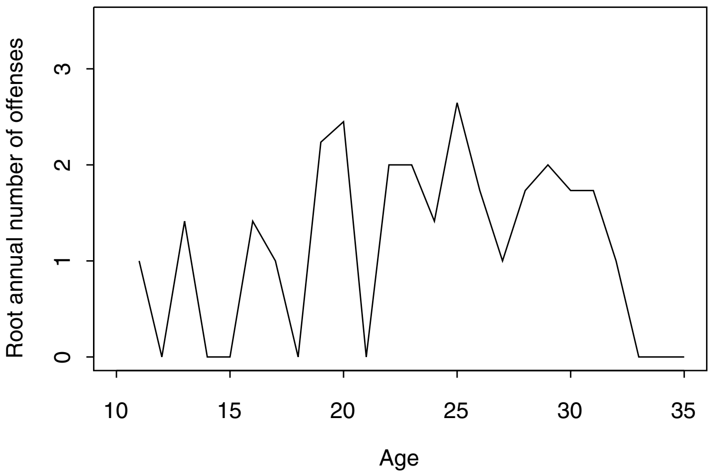

Figure 1.1. The functional datum corresponding to a particular individual in the criminology sample; it shows the way that the annual square root number of crimes varies over the life course.

methodology. In some chapters, we develop or explain some material that will be mainly of interest to statistical experts. These topics are set out in sections towards the end of the relevant chapter, and can be safely skipped by the more general reader.

Chapter 2: Life course data in criminology

We study data on the criminal careers of over 400 individuals followed over several decades of their lifespan. For each individual a function is constructed over the interval $[11, 35]$, representing that person's level of criminal activity between ages 11 and 35. For reasons that are explained, it is appropriate to track the square root of the number of crimes committed each year, and a typical record is given in Figure 1.1. Altogether we consider 413 records like this one, and the records are all plotted in Figure 1.2. This figure demonstrates little more than the need for careful methods of summarizing and analyzing collections of functional data.

Data of this kind are the simplest kind of functional data: we have a number of independent individuals, for each of whom we observe a single function. In standard statistics, we are accustomed to the notion of a sequence of independent numerical observations. This is the functional equivalent: a sequence of independent *functional* observations.

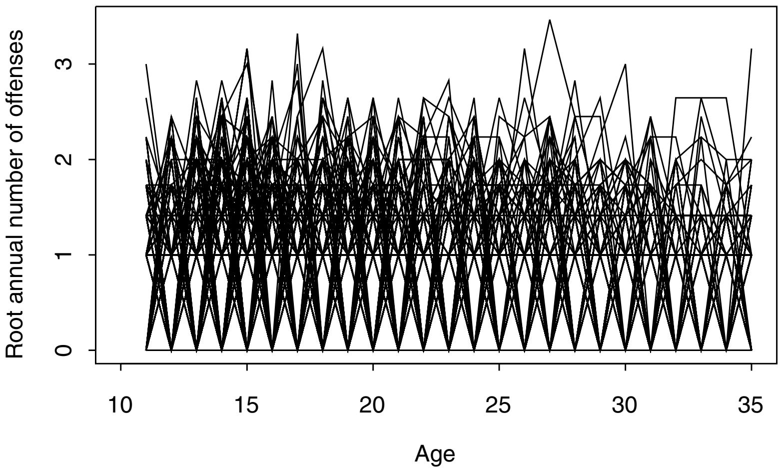

Figure 1.2. The functional data for all 413 subjects in the criminology study.

The questions we address in Chapter 2 include the following.

- What are the steps involved in making raw data on an individual's criminal record into a continuous functional observation?
- How should we estimate the mean of a population such as that in Figure 1.2, and how can we investigate its variability?
- Are there distinct groups of offenders, or do criminals reside on more of a continuum?
- How does our analysis point to salient features of particular data? Of particular interest to criminologists are those individuals who are juvenile offenders who subsequently mature into reasonably law-abiding citizens.

The answers to the third and fourth questions address controversial issues in criminology; it is of obvious importance if there is a "criminal fraternity" with a distinct pattern of offending, and it is also important to know whether reform of young offenders is possible. Quantifying reform is a key step towards this goal.

Chapter 3: The nondurable goods index

In Chapter 3 we turn to a single economic series observed over a long period of time, the U.S. index of nondurable goods production, as plotted

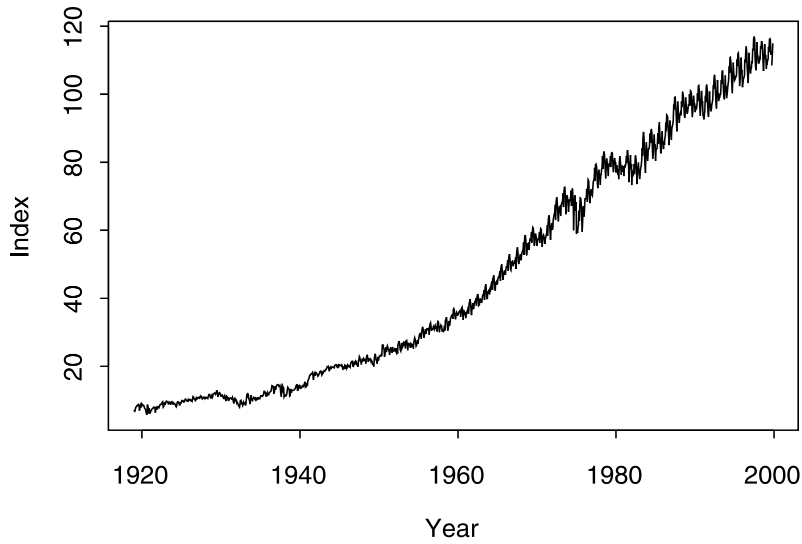

Figure 1.3. The nondurable goods index over the period 1919 to 2000.

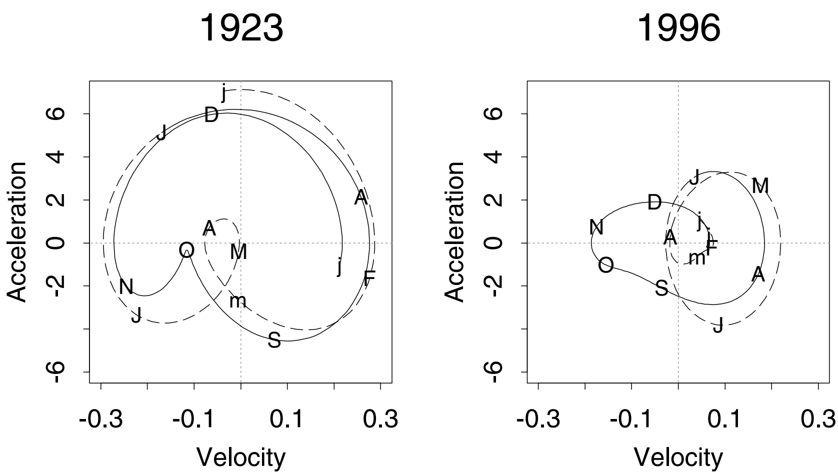

Figure 1.4. Phase-plane plots for two contrasting years: left 1923, right 1996.

in Figure 1.3. Although the index is only produced at monthly intervals, we can think of it as a continuously observed functional time series, with a numerical value at every point over a period of nearly a century. The record for each year may be thought of as an individual functional datum, although of course the point at which each such datum joins to the next is arbitrary; in our analysis, we take it to be the turn of the calendar year.

Our main concern is not the overall level of production, but an investigation of the dynamics of the index within individual years. It is obvious to everyone that goods production nowadays is higher than it was in the 1920s, but more interesting are structural changes in the economy that have affected the detailed behavior, as well as the overall level of activity, over the last century. We pay particular attention to a construct called the *phase-plane plot*, which plots the acceleration of the index against its rate of growth. Figure 1.4 shows phase-plane plots for 1923 and 1996, years near each end of the range of our data.

Our ability to construct phase-plane plots at all depends on the possibility of *differentiating* functional data. In Chapter 3, we use derivatives to construct useful presentations, but in later chapters we take the use of derivatives further, to build and estimate models for the observed functional phenomena.

Chapter 4: Bone shapes from a paleopathology study

Paleopathology is the study of disease in human history, especially taking account of information that can be gathered from human skeletal remains. The study described in Chapter 4 investigates the shapes of a large sample of bones from hundreds of years ago. The intention is to gain knowledge about osteoarthritis of the knee—not just in the past, but nowadays too, because features can be seen that are not easily accessible in living patients. There is evidence of a causal link between the shape of the joint and the incidence of arthritis, and there are plausible biomechanical mechanisms for this link.

We concentrate on images of the knee end of the femur (the upper leg bone); a typical observed shape is shown in Figure 1.5. The functional data considered in Chapter 4 are the outline shapes of bones like this one, and are cyclic curves, not just simple functions of one variable. It is appropriate to characterize these by the positions of *landmarks*. These are specific points picked out on the shapes, and may or may not be of direct interest in themselves.

Specifying landmarks allows a sensible definition of an average bone shape. It also facilitates the investigation of variability in the population, via methods drawn from conventional statistics but with some original twists. Our functional motivation leads to appropriate ways of displaying this variability, and we are able to draw out differences between the bones that show symptoms of arthritis and those that do not.

Figure 1.5. A typical raw digital image of a femur from the paleopathology study.

Chapter 5: Modeling reaction time distributions

Attention deficit hyperactive disorder (ADHD) is a troubling condition, especially in children, but is in reality not easily characterized or diagnosed. One important factor may be the reaction time after a visual stimulus. Children that have difficulty in holding attention have slower reaction times than those that can concentrate more easily on a task in hand.

Reaction times are not fixed, but can be thought of as following a distribution specific to each individual. For each child in a study, a sample of about 70 reaction times was collected, and hence an estimate obtained of that child's density function of reaction time. Figure 1.6 shows typical estimated densities, one for an ADHD child and one for a control.

By estimating these densities we have constructed a set of functional data, one curve for each child in the sample. To avoid the difficulties caused by the constraints that probability densities have to obey, and to highlight features of particular relevance, we actually work with the functions obtained by taking logarithms of the densities and differentiating; one aspect of this transformation is that it makes a normal density into a straight line.

Investigating these functional data demonstrates that the difference between the ADHD and control children is not simply an increase in the mean reaction time, but is a more subtle change in the shape of the reaction time distribution.

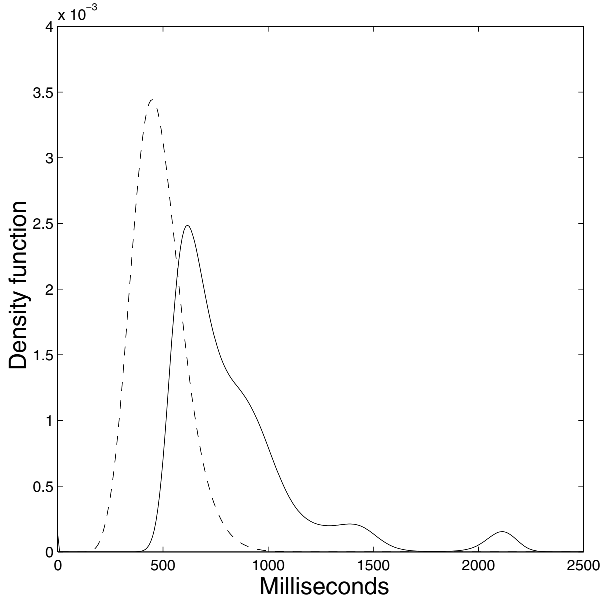

Figure 1.6. Estimated densities of reaction times for two children in the sample. The solid curve corresponds to a child with ADHD, and the dashed curve is one of the controls.

Chapter 6: Zooming in on human growth

Human growth is not at all the simple process that one might imagine at first sight—or even from one's own personal experience of growing up! Studies observing carefully the pattern of growth through childhood and adolescence have been carried out for many decades. A typical data record is shown in Figure 1.7. Collecting records like these is time-consuming and expensive, because children have to be measured accurately and tracked for a long period of their lives.

We consider how to make this sort of record into a useful functional datum to incorporate into further analyses. A smooth curve drawn through the points in Figure 1.7 is commonly called a growth curve, but growth is actually the *rate of increase* of the height of the child. In children this is necessarily positive because it is only much later in life that people begin to lose stature. We develop a monotone smoothing method that takes this sort of consideration into account and yields a functional datum that picks out important stages in a child's growth.

Not all children go through events such as puberty at the same age. Once the functional data have been obtained, an important issue is time-warping or *registration*. Here the aim is to refer all the children to a common biological clock. Only then is it really meaningful to talk about a mean growth pattern or to investigate variability in the sample. Also, the relationship of

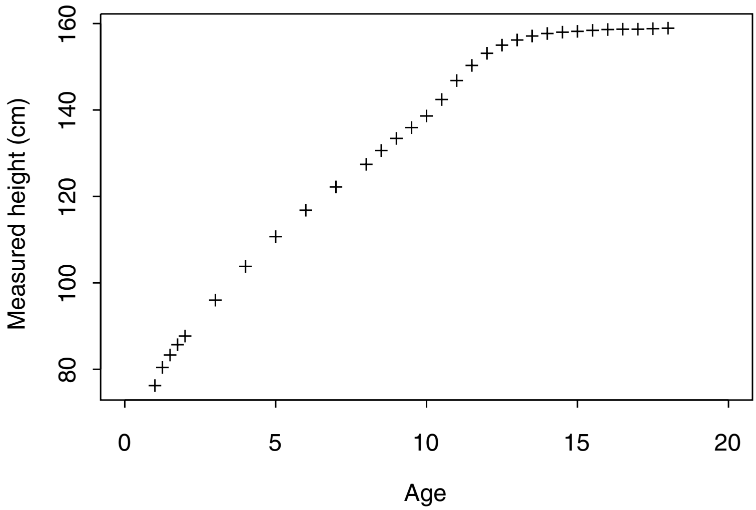

Figure 1.7. The raw data for a particular individual in a classical growth study.

biological to chronological age is itself important, and can also be seen as an interesting functional datum for each child.

The monotone smoothing method also allows the consideration of data observed on much shorter time scales than those in Figure 1.7. The results are fascinating, demonstrating that growth does not occur smoothly, but consists of short bursts of rapid growth interspersed by periods of relative stability. The length and spacing of these *saltations* can be very short, especially in babies, where our results suggest growth cycles of length just a few days.

Chapter 7: Time warping handwriting and weather records

In much biomechanical research nowadays, electronic tracking equipment is used to track body movements in real time as certain tasks are performed. One of us wrote the characters “fda” 20 times, and the resulting pen traces are shown in Figure 1.8. But the data we are actually able to work with are the full trace in time of all three coordinates of the pen position.

To study the important features of these curves, time registration is essential. We use this case study to develop more fully the ideas of registration introduced in Chapter 6, and we discover that there are dynamic patterns that become much more apparent once we refer to an appropriate time scale.

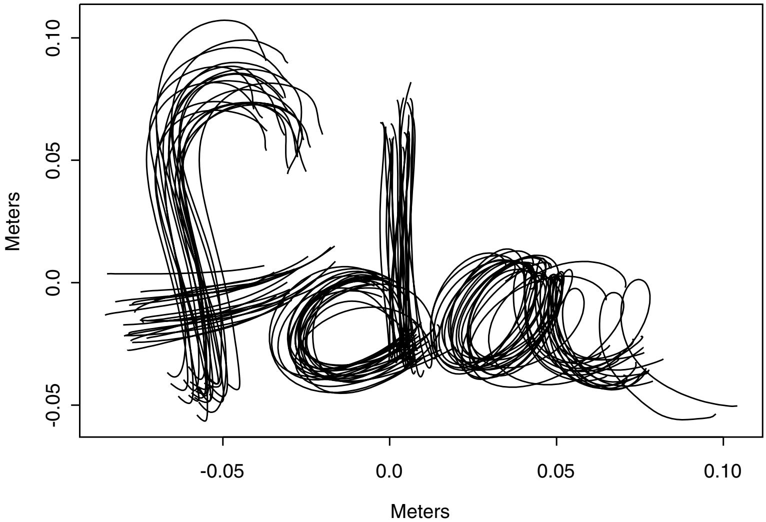

Figure 1.8. The characters “fda” written by hand 20 times.

Weather records are a rich source of functional data, as variables such as temperature and pressure are recorded through time. We know from our own experience that the seasons do not always fall at exactly the same calendar date, and one of the effects of global climate change may be disruption in the annual cycle as much as in the actual temperatures achieved. Both *phase variation*, the variability in the time warping function, and *amplitude variation*, the variability in the actual curve values, are important. This study provides an opportunity to explain how these aspects of variability can be separated, and to explore some consequences for the analysis of weather data.

Chapter 8: How do bone shapes indicate arthritis?

Here we return to the bones considered in Chapter 4, and focus attention on the *intercondylar notch*, the inverted U-shape between the two ends of the bone as displayed in Figure 1.5. There are anatomical reasons why the shape of the intercondylar notch may be especially relevant to the incidence of arthritis. In addition, some of the bones are damaged in ways that exclude them from the analysis described in Chapter 4, but do not affect the intercondylar notch.

The landmark methods used when considering the entire cyclical shape are not easily applicable. Therefore we develop landmark-free approaches to the functional data analysis of curves, such as the notch outlines, traced out in two (or more) dimensions. Once these curves are represented in an appropriate way, it becomes possible to analyze different modes of variability in the data.

Of particular interest is a functional analogue of *linear discriminant analysis*. If we wanted to find out a way of distinguishing arthritic and nonarthritic intercondylar notch shapes, simply finding the mean shape within each group is not a very good way to go. On the other hand, blindly applying discriminant methods borrowed from standard multivariate analysis gives nonsensical results. By incorporating *regularization* in the right way, however, we can find a mode of variability that is good at separating the two kinds of bones. What seems to matter is the twist in the shape of the notch, which may well affect the way that an important ligament lies in the joint.

Chapter 9: Functional models for test items

Now we move from the way our ancestors walked to the way our children are tested in school. Perhaps surprisingly, functional data analysis ideas can bring important insights to the way that different test questions work in practice. Assume for the moment that we have a one-dimensional abstract measure θ of ability. For question i we can then define the *item response function* $P_i(\theta)$ to be the probability that a candidate of ability θ answers this question correctly.

The particular case study concentrates on the performance of 5000 candidates on 60 questions in a test constructed by the American College Testing Program. Some of the steps in our analysis are the following.

- There is no explicit definition of ability θ, but we construct a suitable θ from the data, and estimate the individual item response functions $P_i(\theta)$.
- By considering the estimated item response functions as functional data in their own right, we identify important aspects of the test questions, both as a sample and individually. Both graphical and more analytical methods are used.
- We investigate important questions raised by splitting the sample into female and male candidates. Can ability be assessed in a gender-neutral way? Are there questions on which men and women perform differently? There are only a few such test items in our data, but results for two of them are plotted in Figure 1.9. Which of these questions you would find easier would depend both on your gender and on your position on the overall ability range as quantified by the estimated score θ.

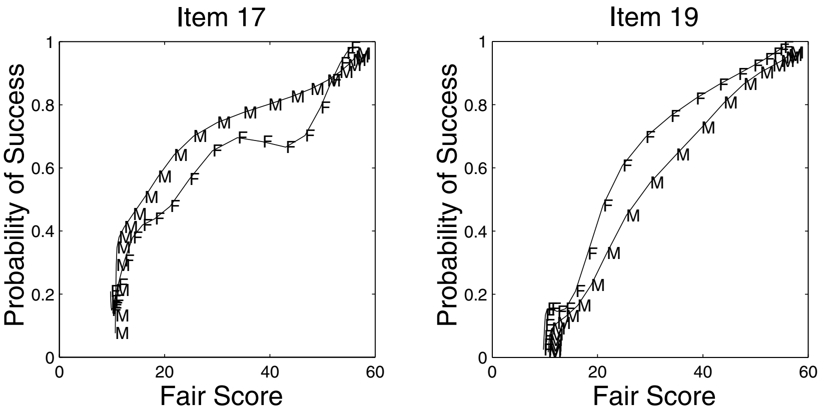

Figure 1.9. Probabilities of success on two test questions are displayed for both females and males, against a fair score that is a reasonable gender-neutral measure of ability.

Chapter 10: Predicting lip acceleration from electromyography

Over 100 muscles are involved in speech, and our ability to control and coordinate them is remarkable. The limitation on the rate of production of phonemes—perhaps 14 per second—is cognitive rather than physical. If we were designing a system for controlling speech movements, we would plan sequences of movements as a group, rather than simply executing each movement as it came along. Does the brain do this?

This big question can be approached by studying the movement of the lower lip during speech and taking electromyography (EMG) recordings to detect associated neural activity. The lower lip is an obvious subset of muscles to concentrate on because it is easily observed and the EMG recordings can be taken from skin surface electrodes. The larynx would offer neither advantage!

A subject is observed repeatedly saying a particular phrase. After preprocessing, smoothing, and registration, this yields paired functional observations $(Y_i(t), Z_i(t))$, where Y_i is the lip acceleration and Z_i is the EMG level. If the brain just does things on the fly, then these data could be modeled by the pointwise model

$$Y_i(t) = \alpha(t) + Z_i(t)\beta(t) + \epsilon_i(t). \tag{1.1}$$

On the other hand, if there is feedforward information for a period of length δ in the neural control mechanism, then a model of the form

$$Y_i(t) = \alpha(t) + \int_{t-\delta}^{t} Z_i(s)\beta(s,t)\,ds + \epsilon_i(t) \tag{1.2}$$

may be more appropriate.

The study investigates aspects of these formulations of *functional linear regression*. The EMG functions play the role of the independent variable and the lip accelerations that of the dependent variable. Because of the functional nature of both, there is a choice of the structure of the model to fit. For the particular data studied, the indication is that there is indeed feedforward information, especially in certain parts of the articulated phrase.

Chapter 11: The dynamics of handwriting printed characters

The subject of this study is handwriting data as exemplified in Figure 1.8. Generally, we are used to identifying people we know well by their handwriting. Since in this case we have dynamic data about the way the pen actually moved during the writing, even including the periods it is off the paper, we might expect to be able to do better still.

It turns out that the X-, Y-, and Z-coordinates of data of this kind can all be modeled remarkably closely by a linear differential equation model of the form

$$u'''(t) = \alpha(t) + \beta_1(t)u'(t) + \beta_2(t)u''(t). \tag{1.3}$$

The coefficient functions $\alpha(t)$, $\beta_1(t)$, and $\beta_2(t)$ depend on which coordinate of the writing one is considering, and are specific to the writer. In this study, we investigate the ways that models of this kind can be fitted to data using a method called *principal differential analysis*.

The principal differential analysis of a particular person's handwriting gives some insight into the biomechanical processes underlying handwriting. In addition, we show that the fitted model is good at the classification problem of deciding who wrote what. You may well be able to forge the shape of someone else's signature, but you will have difficulty in producing a pen trace in real time that satisfies that person's differential equation model.

Chapter 12: A differential equation for juggling

Nearly all readers will be good at handwriting, but not many will be equally expert jugglers. An exception is statistician Michael Newton at Wisconsin, and data observed from Michael's juggling are the subject of our final case study. Certainly to less talented mortals, there is an obvious difference between handwriting and juggling: when we write, the paper remains still and we are always trying to do the same thing; a juggler seems to be catching and throwing balls that all follow different paths.

Various markers on Michael's body were tracked, but we concentrate on the tip of his forefinger. The juggling cycles are not of constant length, because if the ball is thrown higher it takes longer to come back down, and so there is some preprocessing to be done. After this has been achieved, the

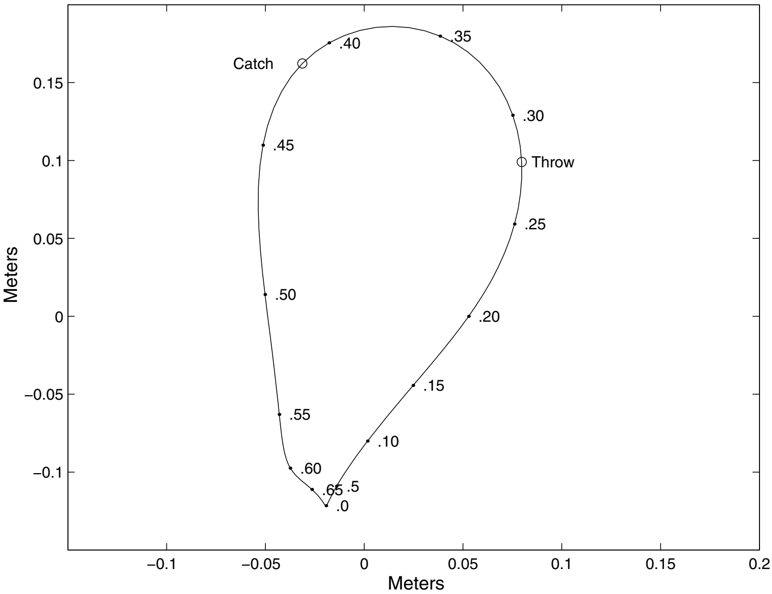

Figure 1.10. The average juggling cycle as seen from the juggler's perspective facing forward. The points on the curve indicate times in seconds, and the total cycle takes 0.711 seconds. The time when the ball leaves the hand and the time of the catch are shown as circles.

average juggling cycle is shown from one view in Figure 1.10. More details are given in Chapter 12.

Although individual cycles vary, they can all be modeled closely by a differential equation approach building on that of Chapter 11. There is a key difference, however; for the handwriting data the model (1.3) was used to model each coordinate separately. In juggling, there is crosstalk between the coordinates, with the derivatives and second derivatives of some affecting the third derivatives of others. However, there is no need for the terms corresponding to $\alpha(t)$ in the model.

Various aspects of the coordinate functions $\beta(t)$ are discussed. Most interestingly, the resulting system of differential equations controls all the individual juggling cycles almost perfectly, despite the outward differences among the cycles. Learning to juggle almost corresponds to wiring the system of differential equations into one's brain and motor system.

1.4 How is functional data analysis distinctive?

The actual term *functional data analysis* was coined by Ramsay and Dalzell (1991), although many of the ideas have of course been around for much longer in some form. What has been more distinctive about recent research is the notion of functional data analysis as a unified way of thinking, rather than a disparate set of methods and techniques.

We have quite deliberately refrained from attempting an exhaustive definition of functional data analysis, because we do not wish to set hard boundaries around the field. Nevertheless, it may be worth noting some common aspects of functional data that arise frequently in this book and elsewhere.

- *Conceptually, functional data are continuously defined.* Of course, in practice they are usually observed at discrete points and also have to be stored in some finite-dimensional way within the computer, but this does not alter our underlying way of thinking.
- *The individual datum is the whole function*, rather than its value at any particular point. The various functional data will often be independent of one another, but there are no particular assumptions about the independence of different values within the same functional datum.
- In some cases the data are functions of time, but *there is nothing special about time as a variable.* In the case studies we have been involved in, the data are functions of a one-dimensional variable, but most of the insights carry over straightforwardly to functions of higher-dimensional variables.
- There is no general requirement that the data be smooth, but *often smoothness or other regularity will be a key aspect of the analysis.* In some cases, derivatives of the observed functions will be important. On other occasions, even though the data themselves need not be smooth, smoothness assumptions will be appropriate for mean functions or other functions involved in modeling the observed data.

1.5 Conclusion and bibliography

Those wishing to read further are referred initially to the book by Ramsay and Silverman (1997), which gives a thematic treatment of many of the topics introduced by case studies in the present volume. That book also contains many additional bibliographic references and technical details. Of particular relevance to this introduction are Chapters 1 and 16 of Ramsay and Silverman (1997). These both stand aside somewhat from specific methods but discuss the general philosophy of functional data analysis. Chapter 16, in particular, considers the historical context of the subject as well as raising some issues for further investigation. Many of the case studies presented in this book are the fruits of our own continuing research in response to this challenge. Although our present book approaches functional data analysis from a different direction, the remark (Ramsay and Silverman, 1997, page 21) made in our previous book remains equally true:

> In broad terms, we have a grander aim: to encourage readers to think about and understand functional data in a new way. The methods we set out are hardly the last word in approaching the particular problems, and we believe that readers will gain more benefit by using the principles we have laid down than by following our suggestions to the letter.

Even more than a thematic treatment, case studies will always lead the alert reader to suggest and investigate approaches that are different, and perhaps better, than those originally presented. If a reader is prompted by one of our chapters to find a better way of dealing with a functional data set, then our aim of encouraging further functional data analysis research and development will certainly have been fulfilled.