12
Functional linear models

12.1 Introduction

We have been exploring the variability of a functional variable without asking how much of its variation is explainable by other variables. It is now time to consider the use of covariates. In classical statistics, the analysis of variance, linear regression and the general linear model serve this purpose, and we now extend the notion of a linear model to the functional context.

Linear models can be functional in one or both of two ways:

1. The dependent or response variable x with argument t is functional.
2. One or more of the independent variables or covariates z is functional.

We will see in Chapter 13 that predicting a functional response with values $x(t)$ by a conventional design matrix (*functional analysis of variance*) or by a set of scalar variables (*functional multiple regression*) involves a fairly straightforward modification of ways of thinking and computational strategies already familiar in ordinary analysis of variance or multiple regression. The main change is that regression coefficients now become regression coefficient functions with values $\beta_j(t)$.

On the other hand, when one or more covariates are themselves functional, a wider range of ways of using a functional covariate to explain a response are available. Let us take a preliminary look at a few of these situations here in this chapter before considering them in detail in later chapters.

Consider the weather at our 35 Canadian sites. The precipitation $\texttt{Prec}(t)$ measured over times t will be the functional dependent variable, and either

- which of four climate zones the station falls in will be a categorical independent variable
- or the temperature $\texttt{Temp}(s)$ measured over times s will be the functional independent variable.

12.2 A functional response and a categorical independent variable

Does the shape of the mean annual precipitation profile depend on which climate zone the station is in? With the zones Atlantic, Arctic, Continental and Pacific, that answer is "almost certainly." Still, we may feel the need to see if the data reject the null hypothesis that there is no difference. And if this happens, then we will want to characterize the differences in functional terms.

In formal terms, we have a number N_g of weather stations in each climate zone $g = 1, \dots, 4$, and the model for the mth precipitation function in the gth group, indicated by $\texttt{Prec}_{mg}$, is

$$\texttt{Prec}_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t). \tag{12.1}$$

In this model, function μ is the grand mean across all 35 weather stations, and the *effect functions* α_g represent departures from the grand mean specific to climate zones. The residual variation left over after we have explained as much as we can using climate zones are captured in the residual functions $\epsilon_{mg}(t)$. Our task is to use the data $\texttt{Prec}_{mg}$ as well as the design matrix coding climate zone membership to estimate the functional parameters μ and α_g.

Moreover, we may want to test more localized hypotheses such as "there are no differences in mid-summer" or "the differences in mid-winter are essentially differences in amount of precipitation rather than in the shape of the precipitation profile." That is, we may have interesting *functional contrasts* specified in advance of looking at the data.

Finally, we can also have the familiar multiple comparisons problem, but this time in functional form. That is, we may simply ask, "Over which time intervals are there significant differences between climate zones?"

More generally, the model may involve a design matrix $\mathbf{Z}$ containing values of p scalar independent variables rather than just 0's and 1's coding category membership, or it may involve both types of predictors. As in the multivariate linear model, these two situations are essentially the same, and this applies here, too.

From an application perspective, a functional response with scalar covariates is a common situation, and our experience indicates that the majority of functional linear model analyses are of this form. We will take up this model in the next chapter, and there we will try to provide as many helpful suggestions for analysis and inference as we can at this point. On the whole, though, tools already familiar to us in working with multivariate data will need only relatively obvious modifications to be adapted to the functional response context.

12.3 A scalar response and a functional independent variable

The converse of the situation considered above may apply. Consider the question, "Does the total amount of precipitation depend on specific features of the temperature profile of a weather station?" Here we can take the response variable as being

$$\texttt{Prectot}_i = \int_0^{365} \texttt{Prec}_i(t)\, dt,$$

where i indexes the 35 weather stations.

Now the issue is how to weight information *within* the single covariate $\texttt{Temp}(s)$ *across* values of s. We do this using the linear model

$$\texttt{Prectot}_i = \alpha + \int_0^{365} \texttt{Temp}_i(s)\beta(s)\, ds + \epsilon_i\,. \tag{12.2}$$

Here the constant α is the usual intercept term that adjusts for the origin of the precipitation variable. The functional parameter of interest is again the regression coefficient function β.

This situation formally resembles conventional multiple regression if we think of each time s as indexing a separate scalar independent variable, namely $\texttt{Temp}(s)$. But then we realize that we now have a potentially unlimited number of independent variables at our disposal to predict 35 scalar values. This seems ridiculous; over-fitting the data now seems inevitable.

The way out of the problem is to force the weighting of information across s to be sufficiently smooth that we can know that a bad fit is in principle possible. This smoothing over s will involve the *regularization* process that we have already seen in action in the spline smoothing chapter 5. Chapter 15 is given over to this situation.

Note that we will always use a different letter s for the argument for a covariate function than we use for the dependent variable. Although in this example context both s and t index time in years over an annual cycle, more generally they could index entirely different continua, such as space for s and time for t.

12.4 A functional response and a functional independent variable

Now we throw open the gates. How does a precipitation profile depend on the associated temperature profile? Now we consider how the functional covariate value $\texttt{Temp}(s)$ influences precipitation $\texttt{Prec}(t)$ specifically at time t. Here are some possibilities.

12.4.1 Concurrent

We might only use the temperature at the same time $s = t$ because we imagine that precipitation now depends only on the temperature now. Our model is

$$\texttt{Prec}_i(t) = \alpha(t) + \texttt{Temp}_i(t)\beta(t) + \epsilon_i(t). \tag{12.3}$$

We might call this model *concurrent* or *point-wise.* Should we use regularization to force β to be smooth in t?

This model has already been discussed in some detail prior to the first edition of this book by Hastie and Tibshirani (1993) under the name of the *varying coefficient model.* It deserves here a chapter of its own, 14, in part because we will show that all functional linear models can be reduced to this form.

12.4.2 Annual or total

We may prefer to allow for temperature influence on $\texttt{Prec}(t)$ to extend over the whole year. The model expands to become

$$\texttt{Prec}_i(t) = \alpha(t) + \int_0^{365} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t). \tag{12.4}$$

We face the additional complexity of the regression coefficient function β being bivariate; the value $\beta(s,t)$ determines the impact of temperature at time s on precipitation at time t.

We suspect from the discussion of the scalar response and functional covariate that it may be essential to smooth β as a function of s. But what is the difference between s-smoothing and smoothing with respect to t?

12.4.3 Short-term feed-forward

We may choose for reasons of parsimony to use only the temperature now and over an interval back in time in order to allow for some cumulative effects. For example, it may be that what counts is whether the temperature has been falling rapidly up to time t. The model expands to

$$\texttt{Prec}_i(t) = \alpha(t) + \int_{t-\delta}^{t} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t). \tag{12.5}$$

Here δ is the time lag over which we use temperature information. In addition to being bivariate, now β is only defined over the somewhat complicated trapezoidal domain: $t \in [0, 365], t - \delta \le s \le t$.

Since in this situation the data are periodic, so we won't have particular problems with s being negative at $t = 0$ since we can borrow information from the previous year. But for non-periodic data, we would want to remove the triangle implied by $s < 0$ from the domain.

12.4.4 Local influence

Finally, after some reflection, we may open up the model to allow integration over s within a t-dependent set Ω_t. Why? Well, for example, if the temperature first falls rapidly, and then rises rapidly immediately after, and if the time t in question is in the middle of the summer, this may be a thunderstorm, and will therefore have the potential for a very large amount of rainfall within a short time period. The model may therefore be

$$\texttt{Prec}_i(t) = \alpha(t) + \int_{\Omega_t} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t). \tag{12.6}$$

Here there is the potential complexity of the domain over which β is defined that will challenge our computational resources.

These examples indicate that the functional linear model has the potential to be rather complex. Indeed, there is no reason why the covariate z might not be a function of both s and t. For example, we may predict rainfall at a station by integrating information over both space and time if we are on the Canadian prairies where precipitation in the summer tends to be *convective*, meaning thunder storms, hail storms and tornadoes that tend to be spatially limited and to follow curvilinear tracks.

12.5 What about predicting derivatives?

We may choose to model the rate of change in precipitation, $D\texttt{Prec}$ instead of precipitation itself. When a model is designed to explain a derivative of some order, we call it a *dynamic model.* In this case, the model is a *differential equation,* meaning simply that a derivative is involved.

When the response is a derivative, then there is the potential for the function itself to be a useful covariate. For example, the concurrent linear model

$$D\texttt{Prec}_i(t) = \texttt{Prec}_i(t)\beta(t) + \epsilon_i(t) \tag{12.7}$$

is a called a *homogeneous first order linear differential equation* in precipitation, and if we also include an influence of temperature,

$$D\texttt{Prec}_i(t) = \texttt{Prec}_i(t)\beta_0(t) + \texttt{Temp}_i(t)\beta_1(t) + \epsilon_i(t), \tag{12.8}$$

the equation is said to be *nonhomogeneous* rather than *homogeneous.* Temperature in the equation is called a *forcing function.*

The final chapters in the book will take up the story of differential equations, and we will see that the power of functional data analysis is remarkably extended in this way.

12.6 Overview

Although we dedicate separate chapters to these situations for the good reason that each involves some specialized techniques and issues, at a broader level the differences between the various models outlined above are more apparent than real. For example, a scalar response can always be expressed as a functional response with a constant basis, and the same is true for a scalar covariate. Of course, specialized computational issues arise as we take advantage at an algorithmic level of the fact that scale variables are involved.

A central theme common to all functional linear models is that of smoothing regression coefficient functions. Functional linear models usually involve more predictive power than we want to use for a finite amount of noisy data. Deciding how much to smooth and how to define smoothness itself will be a central issue in most applications.

Probably the most fundamental issue is the nature of the potentially t-specific domain Ω_t in (12.6). Both the point-wise and total influence models are comparatively easy to deal with computationally, as we shall see. But localized feed-forward influence is often essential, and already well represented in statistics in the form of ARIMA and state-space models in time series analysis.