5

Functional Nonparametric Prediction Methodologies

This chapter describes several approaches concerning the nonparametric prediction of some scalar response. The functional setting appears through the explanatory functional variable. We focus on three complementary prediction methods, namely the conditional expectation, the conditional median and the conditional mode. Conditional expectation refers to the well-known regression method whereas both the conditional median and conditional mode are strongly linked with the estimation of the conditional distribution. After introducing some bibliographical aspects (Section 5.1), we present in Section 5.2 the three functional nonparametric prediction methods. Section 5.3 presents the nonparametric models associated with these prediction problems, while Section 5.4 focuses on the construction of the estimators.

5.1 Introduction

There are many situations in which one may wish to study the link between two variables, with the main goal to be able to predict new values of one of them given the other one. This prediction problem has been widely studied in the literature when both variables are of finite dimensions. Of course the same problem can occur when some of the variables are functional. Our wish is to investigate this problem when the explanatory variable is functional and the response one is still real. Both to fix the ideas and to emphasize the great interest and usefulness of this problem in many fields of applied sciences, let us quickly come back to the chemometric data presented in Chapter 2.

As discussed in Section 2.1, recall that the statistical sample (of size $n = 215$) is composed of spectrometric curves $\chi_1, \ldots, \chi_n$ (these are the functional data) corresponding to the spectra observed for 215 pieces of finely chopped meat. In addition, by an analytic chemical process we have measured the fat content of each piece $y_1, \ldots, y_n$ (these are the scalar responses). Thus, we collect the observations of a scalar response (the fat content) and an explanatory functional variable (spectra). One question is: given an observed

spectrum of a piece of meat, can we predict its corresponding fat content? This is typically a functional prediction problem. To answer the question, we have to estimate the link between the fat content and the spectra. Unfortunately, there is neither a way to display this relationship nor is there structural information about it. Therefore, it becomes natural to introduce nonparametric models in order to make as few assumptions as possible on the shape of the link. The functional aspect of the problem is very important too, and we have to attack it in such a way as to use the whole spectrometric curve. In particular, continuity and other functional features of the spectra have to be taken into account. There is therefore real need for developing methods combining both nonparametric concepts and functional variable modelling.

Of course, there is a consistent literature both around nonparametric prediction and functional data. But, until now, functional variables have been studied essentially in a parametric setting. This has been popularized by [RS97] (mainly for practical points of view) and previous theoretical developments can be found in [B00] in the specific context of dependent functional variables. Recent practical advances can be found for instance in [CGS04] whereas some asymptotic studies are detailed in [CFS03] and [CFF02].

In another direction of statistical research, nonparametric prediction problems have been investigated intensively both in real and multivariate cases. It is impossible to give an exhaustive description of the related bibliography, but to fix the ideas the reader could look at the precursor works by [W64] and [N64], at the intermediary survey by [C85] and at [S00] or [AP03] for a description of the state of the art.

The aim of this book is to marry advantages of free-modelling together with a fully functional methodology in order to answer to functional prediction problem such as the spectrometric one. In this chapter, we present three functional nonparametric statistical approaches for the prediction problem. The reader has to keep in mind Defintion 1.4 and the fact that in the designation *functional nonparametric prediction method*, the word *functional* refers to the concept of functional variable (implicitly "we have to take into account the functional feature of the variable") while the word *nonparametric* means that we use a free-parameter modelling for the nonlinear operators to be estimated. In addition, it is important to note that our methodology is also based on free-distribution modelling since no parametric assumption is necessary for the distribution of the random variables.

5.2 Various Approaches to the Prediction Problem

Let us start by recalling some notation. Let $(\boldsymbol{\mathcal{X}}_i, Y_i)_{i=1,\dots,n}$ be n independent pairs, identically distributed as $(\boldsymbol{\mathcal{X}}, Y)$ and valued in $E \times \mathbb{R}$, where (E, d) is a semi-metric space (i.e. $\boldsymbol{\mathcal{X}}$ is a f.r.v. and d a semi-metric). Let χ (resp. y) be a fixed element of E (resp. $\mathbb{R}$), let $\mathcal{N}_\chi \subset E$) be a neighboorhood of χ and S