Nonparametric techniques for Functional Data Analysis

Ph.D. Thesis Proposal

Mattia Ciollaro
Department of Statistics
Carnegie Mellon University

Thesis Committee

Peter E. Freeman
Christopher R. Genovese
Jing Lei
Larry Wasserman

Department of Statistics

and

Ross O'Connell

Department of Physics

December 12th, 2014

Abstract

Functional Data Analysis (FDA) is a modern branch of Statistics dealing with infinite-dimensional data such as functions, curves, surfaces, or images. Functional data are more and more often encountered in diverse scientific fields such as Astronomy, Envirometrics, Chemometrics, Biometrics, Econometrics, and Medicine to name a few. This fact has led to an increasing demand for novel statistical techniques that can account for the infinite-dimensional nature of functional data and address the challenges arising in their analysis. My research focuses both on methodological and applied problems involving smooth functional data. In particular, I aim at studying pseudo-modes and gradient flows on pseudo-densities (where the word 'pseudo' hints at the difficulties related to defining proper probability density functions on infinite-dimensional spaces). My main goals are to extend the notion of local mode (defined as a local maximum of a probability density function) to smooth functional data belonging to an infinite-dimensional space, to provide a solid theoretical background for modal clustering of these data, and to develop effective algorithms for this task. The second part of my research project is devoted to the estimation of the smooth component of the so-called *Lyman-α forest* portion of the light spectrum of some astronomical objects called *quasars*. Here, the goal is to disentangle the smooth component of the light spectrum in the Lyman-α forest from a non-smooth absorption process caused by neutral hydrogen which strongly influences the observed shape of quasar light spectra. I suggest estimating the smooth component of the light spectrum in the Lyman-α forest by using a nonparametric functional regression model that predicts the entire smooth light spectrum in the forest (the response) on the basis of an estimate of the smooth light spectrum outside of the forest (the predictor). In order to quantify the uncertainty associated to this prediction procedure, I introduce a class of prediction bands that are specifically designed for functional responses and have finite sample coverage guarantees. An estimate of the smooth continuum in the Lyman-α forest is crucial to quantify the intensity of absorption at various locations in our Universe. The intensity of absorption and its spatial distribution are scientifically relevant in modern Cosmology as they convey precious information about the distribution of standard matter in portions of our Universe that cannot be easily probed otherwise, if not by means of the Lyman-α forest.

Contents

1 Introduction

Progress in mathematical statistics correspond, at least to some extent, to an increase in the level of sophistication of the structures that are assumed for the sample space, $\mathcal{X}$ (i.e. the space to which the data $X_1, \ldots, X_n$ are assumed to belong), and the space of the parameters characterizing the distribution P of the data, Θ. In this sense, the field of Statistics has witnessed a steady increase in the level of sophistication and, at least approximately, we can now distinguish among the following setups (Cuevas, 2014):

- Classical inference: $\mathcal{X} = \mathbb{R}$, $\Theta \subset \mathbb{R}$; dating back to 1920's
- Multivariate analysis: $\mathcal{X} = \mathbb{R}^d$ $(n >> d)$, $\Theta \subset \mathbb{R}^k$ $(n >> k)$; dating back to the 1940's
- Nonparametrics: $\mathcal{X} = \mathbb{R}^d$ $(n >> d)$, $\Theta =$ a function space; dating back to the 1960's
- **Functional data analysis**: $\mathcal{X} =$ a function space, $\Theta \subset \mathbb{R}^k$ or $\Theta =$ a function space; dating back to the 1990's
- High-dimensional problems: $\mathcal{X} = \mathbb{R}^d$ $(n < d)$, $\Theta \subset \mathbb{R}^k$; dating back to the 2000's.

In general, we can say that Functional Data Analysis (FDA in the following) is the branch of Statistics dealing with data that come in the form of a sample of (possibly vector-valued) functions $X_1(t), \ldots, X_n(t)$, where $t \in \mathcal{T}$ and $\mathcal{T}$ usually corresponds to a compact subset (typically a rectangle) of $\mathbb{R}^d$. The distinctive feature of FDA lies in the mathematical and methodological effort taken to account for the intrinsically infinite-dimensional nature of the sample units and, most frequently, of the sample space $\mathcal{X}$ as well.

Although from a conceptual point of view we can certainly conceive the idea of observing a sample of infinite-dimensional objects (such as a sample of curves), from a practical point of view we never actually observe infinite-dimensional sample units. Rather, in the case of curves for instance, we observe discretized versions of such objects, where the level of discretization and the discretization grids depend on the technological limits of the data-acquisition apparatus or on other constraints affecting the design of the experiment. Frequently, the observed discretized curves are also perturbed by random noise, systematic errors introduced by the data-acquisition apparatus or by other sources[1]. It is natural to ask why one cannot simply analyze functional data with the rich set of tools of multivariate statistics. There are at least two reasons why statistical methods of multivariate analysis tend to fail when applied to functional data: first, the ratio between the sample size (the number of curves in the sample) and the number of variables (the size of the grid at which the curves are sampled) is generally unbalanced, with the sample size being often too small compared to the number of variables; the other reason comes from the fact that if the curves are sufficiently smooth, then high correlations exist among the observed variables (i.e. the values of the curves at nearby

[1]Two different approaches arise in the FDA literature: some authors reconstruct the curves from the noisy observations on the grid by means of some smoothing procedure and then take these smoothed objects as the functional sample of observations. Other authors have been developing error-in-variables functional models in the attempt to keep account for the noise at the grid level even once the original discretized curves are transformed into smooth functional objects.

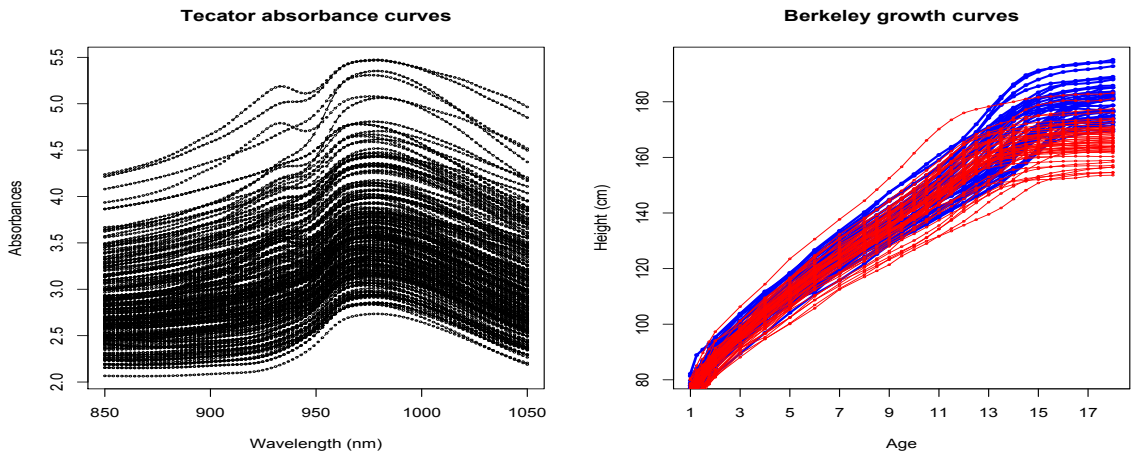

Figure 1: *Two examples of functional data. Left panel: 215 spectrometric curves corresponding to the absorbance level of finely chopped pieces of meat measured on a 100-channels spectrum. Right panel: growth curves of 39 boys (thicker blue lines) and 54 girls (thinner red lines) measured at 31 different ages from the age of 1 to 18 years.*

points in the grid). Strong correlations and collinearity are known to induce ill-conditioned problems in the context of multivariate statistical analyses. These facts motivate the need to develop ad hoc methods for the analysis of functional data.

Figure 1 displays two examples of functional data. The left panel displays a sample of 215 absorbance curves of finely chopped pieces of meat. The data belong to the 'Tecator' dataset (Borggaard and Thodberg, 1992, `http://lib.stat.cmu.edu/datasets/tecator`; the dataset is also directly accessible from the R package 'fda.usc'). Each curve represents the absorbance of a single piece of meat measured by a spectrometer at 100 wavelengths between 850 nm and 1050 nm. Each sample unit is the discretized version of an underlying smooth curve and the level of discretization (in this case 100 points/curve) depends on the spectrograph. The right panel displays the growth curves of 39 boys and 54 girls of the 'Berkeley Growth Study' dataset (Tuddenham and Snyder, 1954; the dataset can be accessed through the R 'fda' package). Again, each sample unit is the discretized version of an intrinsically continuous curve. In this case, the height of each individual in the study is measured 31 times between the age of 1 and 18 years; however, because of the continuous nature of time, we could at least in principle imagine to sample such curves at arbitrarily high resolution.

Notable and interesting mathematical difficulties arise in FDA. For instance, in the high-dimensional regime, it is well-known that the *curse of dimensionality* (or *empty-space phenomenon*; see for example Section 1.2 of Lee and Verleysen, 2007) kicks in and negatively affects the quality of the statistical inferences. Furthermore, because functional data often belong to infinite-dimensional sample spaces, one expects the negative effects of the curse of dimensionality to be further amplified (leading to the *curse of infinite dimension*; Ferraty et al., 2006). Also, if the sample space $\mathcal{X}$ is infinite-dimensional, it is generally hard to meaningfully define a density function for the probability measure P associated to the data.

Thus, some of the challenges that arise in the analysis of functional data often come from the following sources:

- **Concentration of probability.** For a random element $X \sim P$ valued in a semi-metric[2] space $(\mathcal{X}, d)$, the small ball probability function at a point $x_0 \in \mathcal{X}$ is defined as

$$\varphi_{x_0}(h) = P\left(d(X - x_0) \leq h\right), \tag{1}$$

  i.e. $\varphi_{x_0}(h)$ describes the probability content of the ball of radius $h > 0$ centered at $x_0 \in \mathcal{X}$, and the balls are defined in terms of the topology induced by the semi-metric d. In the finite-dimensional setting where $\mathcal{X} = \mathbb{R}^d$, usually $\varphi_{x_0}(h) \sim C_{x_0} h^d$. In contrast, for general diffusion processes one typically has $\varphi_{x_0}(h) \sim C_{x_0}\, e^{-C/h^2}$ for some positive constants C_{x_0} and C when d is the distance induced by the supremum norm (see Ferraty et al., 2006 for details). It is clear that for diffusion processes φ vanishes at a faster rate as $h \to 0$ compared to the finite-dimensional case, hence probability is less concentrated. Li and Shao (2001) survey the state of the art about small ball probabilities for Gaussian processes. However, very little is known in general about small-ball probability functions in infinite-dimensional spaces aside from specific (and usually Gaussian) settings. Furthermore, it is in general hard to obtain explicit expressions for the shifted small ball probabilities of generic processes (i.e. expressions for φ_{x_0} when the center of the ball x_0 is arbitrary). Rates of convergence of nonparametric functional estimators depend explicitly on the asymptotic behavior of φ as $h \to 0$ (see, for instance, Ferraty and Vieu, 2006). The difficulties associated with small ball probabilities significantly complicate asymptotic analyses, and the infinite-dimensional nature of the sample space can lead to negative results in terms of minimax rates of convergence (see, for an example regarding regression, Mas, 2012).

- **Lack of a natural dominating measure.** Unlike the case $\mathcal{X} = \mathbb{R}^d$, an analog to the Lebesgue measure is not directly available when the sample space $\mathcal{X}$ is infinite-dimensional. This fact poses serious difficulties in terms of the definition of probability density functions for a functional random variable and, consequently, in terms of the definition of any density-related object (such as the notion of mode, for instance). It should be noted that simply assuming the existence of a dominating measure for a statistical model on a case by case basis is not a satisfying solution and neither is subjectively choosing a dominating measure for a particular model: this would introduce too much subjectivity in the statistical analysis of functional data. Frequently, a way to compensate for the lack of a proper density is to introduce some type of pseudo-density (see, for instance, Gasser et al., 1998, Hall and Heckman, 2002, Delaigle and Hall, 2010, or Ferraty et al., 2012a). For example, for a random element $X \sim P$ valued in the semi-metric space $(\mathcal{X}, d)$ and for some $h > 0$, one may take the functional

$$p_h(x_0) = E_P\, K\left(\frac{d(X, x_0)}{h}\right) = \int_{\mathbb{R}} K(t)\, dP_{d(X,x)/h}(t) \tag{2}$$

[2] In a semi-metric space $(\mathcal{X}, d)$ the distance function d satisfies all the usual properties of a proper distance function except for the coincidence axiom: while d must satisfy $x = y \implies d(x, y) = 0$ for all $x, y \in \mathcal{X}$, it is not required that $d(x, y) = 0 \implies x = y$ for all $x, y \in \mathcal{X}$.

to be a pseudo-density for P (see, for instance, Hall and Heckman, 2002), where K is a kernel function and $P_{d(X,x)/h}$ is the probability measure induced by P through the map $X \mapsto d(X,x)/h$. Notice that when K is the uniform kernel $K(t) = \mathbb{1}_{[0,1]}(t)$, then $p_h(x_0) = \varphi_{x_0}(h)$. However, an inconvenient feature of this population functional is that it explicitly depends on the resolution level corresponding to the bandwidth parameter h. An alternative is to assume that

$$\varphi_{x_0}(h) = P\left(d(X - x_0) \leq h\right) = p(x)\phi(h) + o(\phi(h)) \tag{3}$$

as $h \to 0$, where p is a non-negative functional only depending on the center of the ball and ϕ is a *concentration function* only depending on the size of the ball (see, for instance, Gasser et al., 1998 and Ferraty et al., 2012a). In this case, under appropriate identifiability conditions, the functional p is considered a pseudo-density for P. Generally, the factorization above corresponds to a moderately strong assumption as few processes can be shown to satisfy it. For example, in the case of Brownian motion, it can be shown (Li and Shao, 2001) that for all absolutely continuous functions $x : [0,1] \to \mathbb{R}$ such that $x(0) = 0$ and $\int_0^1 \left(x'(t)\right)^2 \, dt < \infty$ the small ball probability function satisfies the above factorization with

$$\varphi_x(h) \sim p(x)\phi(h) = \exp\left(-\frac{1}{2}\int_0^1 \left(x'(t)\right)^2 \, dt\right)\varphi_0(h) \tag{4}$$

as $h \to 0$. Notice, however, that the factorization is not necessarily satisfied when x is a trajectory of the Brownian motion.

The literature and the open-source software on FDA are rapidly growing. If on the one hand it is not possible to attempt a literature review based on research papers here, we can certainly point out some of the landmarking monographs and some of the available open-source software for FDA. The books by Ramsay, Silverman, Hooker, and Graves (Ramsay et al., 2002, Ramsay and Silverman, 2005, whose first edition was published in 1997, and Ramsay et al., 2009) represent some of the monographs which helped popularizing (parametric) FDA both from a methodological and an applied perspective. They describe the functional extension of some of the most frequently used tools of multivariate analysis (such as principal component analysis, canonical correlation analysis, and discriminant analysis), and they also discuss descriptive statistics for functional data, parametric functional regression, curve registration, and other topics. The methods presented in these books are illustrated by means of various case studies. R and MATLAB implementations are available through the R 'fda' package. Another well-known R package for FDA is the 'fda.usc' package by Febrero-Bande and Oviedo de la Fuente. This R package features various functions to carry out exploratory and descriptive analysis of functional data; it also contains functions to perform functional regression with scalar response, supervised and unsupervised classification of functional data and functional analysis of variance. The book by Ferraty and Vieu (Ferraty and Vieu, 2006) complements the aforementioned monographs in that it focuses on the nonparametric (and doubly infinite-dimensional) setting in which both the data and the parameter space are infinite-dimensional. The book also summarizes some of the very interesting theoretical contribution of the French school about FDA on semi-metric spaces, the theoretical difficulties

associated with small ball probabilities, and how the curse of infinite dimension, the influence of small ball probabilities and the choice of the semi-metric are all intermingled. R code for the implementation of the proposed nonparametric functional procedures is available on the authors' research group website (`http://www.math.univ-toulouse.fr/staph/npfda/`). The general theory for autoregressive processes in function spaces is developed in Bosq (2000), which is another popular book within the FDA community. The monograph by Horváth and Kokoszka (Horváth and Kokoszka, 2012) offers a well-balanced mix of theory and applications (in particular, Chapter 2 provides a useful background reading on the theory of Hilbert spaces). Their book, which builds on part of the recent research of the two authors, also deals with some modern topics such as the analysis of dependent functional data, functional time series, change-point detection and spatial statistics for functional data. Finally, it is worth mentioning the monograph edited by Férraty and Romain (Ferraty and Romain, 2011), which contains invited discussions by leading international experts describing original ideas, current state of the art techniques, and avenues for future work in FDA. The book also contains comprehensive bibliographical information.

2 Two problems in Functional Data Analysis

My research project focuses mainly on two problems involving functional data. Part of my research is devoted to define and estimate functional modes by means of pseudo-density functions, use pseudo-densities to perform modal clustering of functional data, and perform statistical inference on the functional modes and on the associated clusters. Some steps in this direction are presented in Ciollaro et al. (2014b). Another part of my research project deals with a functional regression problem in modern Astronomy (Ciollaro et al., 2014a).

2.1 Modal clustering for functional data[3]

It is natural to define a functional mode as an element $x \in \mathcal{X}$ corresponding to a local maximum of a pseudo-density function p defined on the sample space $\mathcal{X}$ (such as in equations (2) or (3) for instance). Pseudo-densities can be used for clustering of functional data by means of the partition of the sample space induced by their modes: once a set of functional modes is estimated, the sample space is partitioned in the basins of attraction of the estimated sample modes (the basin of attraction of a mode is defined as the set of all points in the sample space whose gradient ascent lines converge to that mode). Figure 2 gives a visual representation of this idea in 2D. Unlike most clustering methodologies (which define and treat clusters exclusively as sample entities), modal clustering has a stronger inferential foundation and has gained an increasing appeal. According to the paradigm of modal clustering, one can clearly distinguish a population parameter to be estimated (the partition of $\mathcal{X}$ corresponding to the basins of attraction of the local modes of the true density function) and an estimator of this parameter (the partition of $\mathcal{X}$ corresponding to the basins of attraction of the local modes of an estimate of the true density function). Recent work on the inferential foundation of modal clustering includes Chacón (2012) and Chacón (2014).

[3] **Ongoing work with Christopher Genovese, Jing Lei, Daren Wang (Department of Mathematics), and Larry Wasserman.**

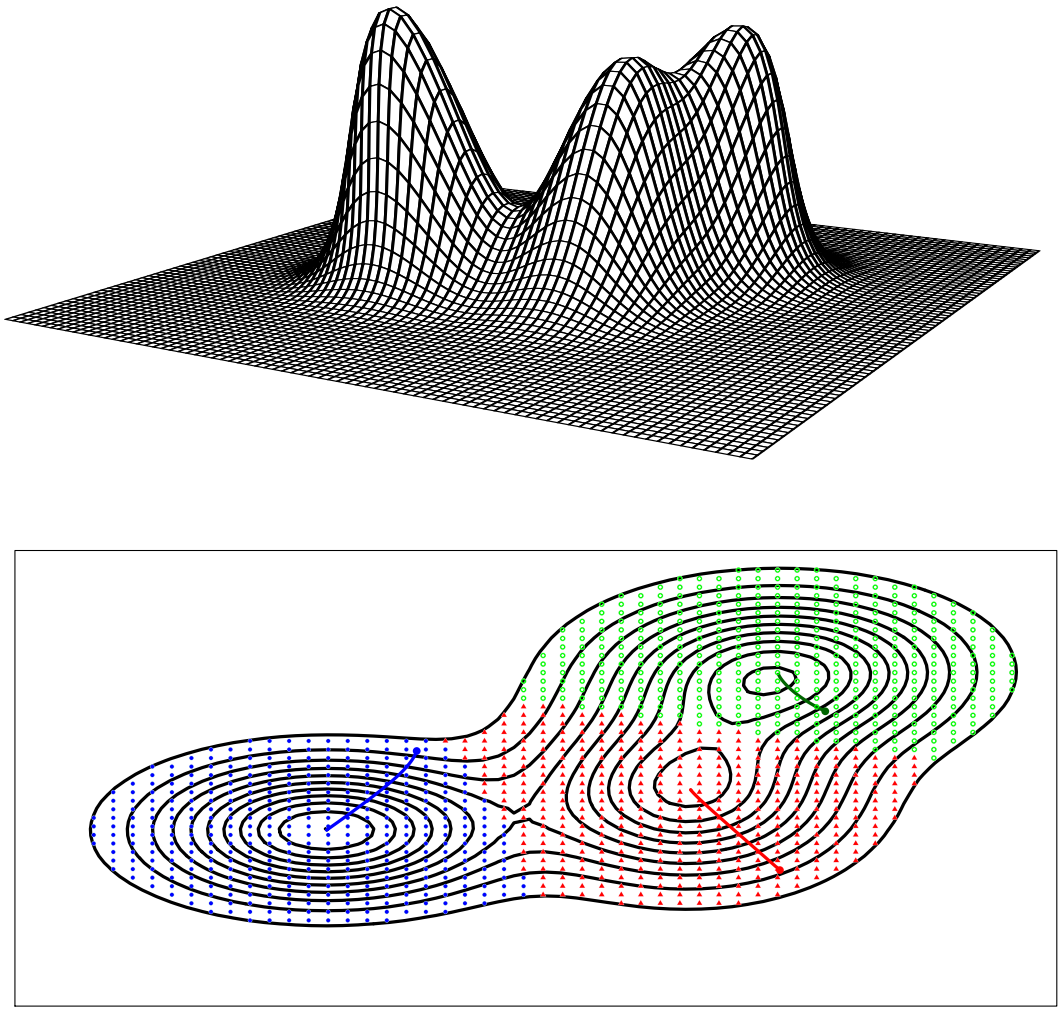

Figure 2: *Top panel: a bivariate density with compact support. The density has three local modes and therefore three well-defined clusters. Bottom panel: the three basins of attraction of the local modes. Within each basin of attraction, the gradient ascent path starting from a point belonging to that cluster is highlighted.*

When $\mathcal{X} = \mathbb{R}^d$ and p is a density function with Lipschitz derivative, the gradient ascent path $\pi_{x_0} : \mathbb{R}_+ \to \mathbb{R}^d$ joining an arbitrary starting point $x_0 \in \mathcal{X}$ to its closest local mode is the unique solution to the initial value problem

$$\begin{cases} \pi'_{x_0}(t) = \nabla p(\pi_{x_0}(t)) \\ \pi_{x_0}(0) = x_0. \end{cases} \tag{5}$$

The gradient ascent curve π_{x_0} is most often referred to as a *integral curve* and the set of all the integral curves associated to p is usually referred to as the *gradient flow* on p. In the finite dimensional case, one can show that as long as p is a smooth function and all its critical points are non-degenerate[4] (i.e. the determinant of the Hessian at each critical point of p is not null), the gradient flow on p satisfies some interesting properties. In particular, for an arbitrary initial point x_0 the integral curve π_{x_0} exists and is unique, and any two gradient ascent lines starting at different initial values x_0 and y_0 can only intersect at a critical point of p. Moreover, the equivalence class of points of $\mathcal{X}$ whose gradient ascent lines culminate at the same local mode of p corresponds to the basin of attraction of that local mode. Finally, the collection of these basins of attraction forms an ('essential', i.e. up to a null probability set) partition of the sample space $\mathcal{X}$ (Chacón, 2012, Chacón, 2014 and references therein) and therefore a well-defined clustering of the sample space. As soon as an i.i.d. sample $\mathcal{S} = \{X_1, \ldots, X_n\}$ is available, p and its gradient flows can be consistently estimated by means of a kernel density estimator[5]

$$\hat{p}(x_0) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\|X_i - x_0\|}{h}\right) \tag{6}$$

and by applying a suitable gradient ascent algorithm (Arias-Castro et al., 2013). A very well-known adaptive gradient ascent algorithm which can be used to locate the modes of $\hat{p}$ is the *mean-shift algorithm* (Fukunaga and Hostetler, 1975, Cheng, 1995). Briefly, the mean-shift algorithm is an iterative procedure that repeatedly shifts a point $x \in \mathbb{R}^d$ towards its closest local sample mean by means of the update equation

$$x \leftarrow \frac{\sum_{X \in \mathcal{S}} K\left(\frac{\|X-x\|}{h}\right) X}{\sum_{X \in \mathcal{S}} K\left(\frac{\|X-x\|)}{h}\right)}. \tag{7}$$

The mean-shift trajectory starting from x eventually culminates at a local mode of $\hat{p}$. More details about the connection between the mean-shift algorithm and gradient ascent can be found in Ciollaro et al. (2014b).

In the case of functional data valued in a semi-metric space, Ferraty et al. (2012a) show that the pseudo-density p of equation (3) can be consistently estimated by

$$\hat{p}(x_0) = \frac{\frac{1}{n} \sum_{i=1}^{n} K\left(\frac{d(X_i,x_0)}{h}\right)}{\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{d(X_i,X_j)}{h}\right)}. \tag{8}$$

[4]Such functions are usually referred to as *Morse functions*.

[5]Of course, the kernel density estimator can be generalized to $\hat{p}(x_0) = \frac{1}{n}|H|^{-1/2} \sum_{i=1}^{n} K\left(\|H^{-1/2}(X_i - x_0)\|\right)$ for a positive-definite bandwidth matrix H.

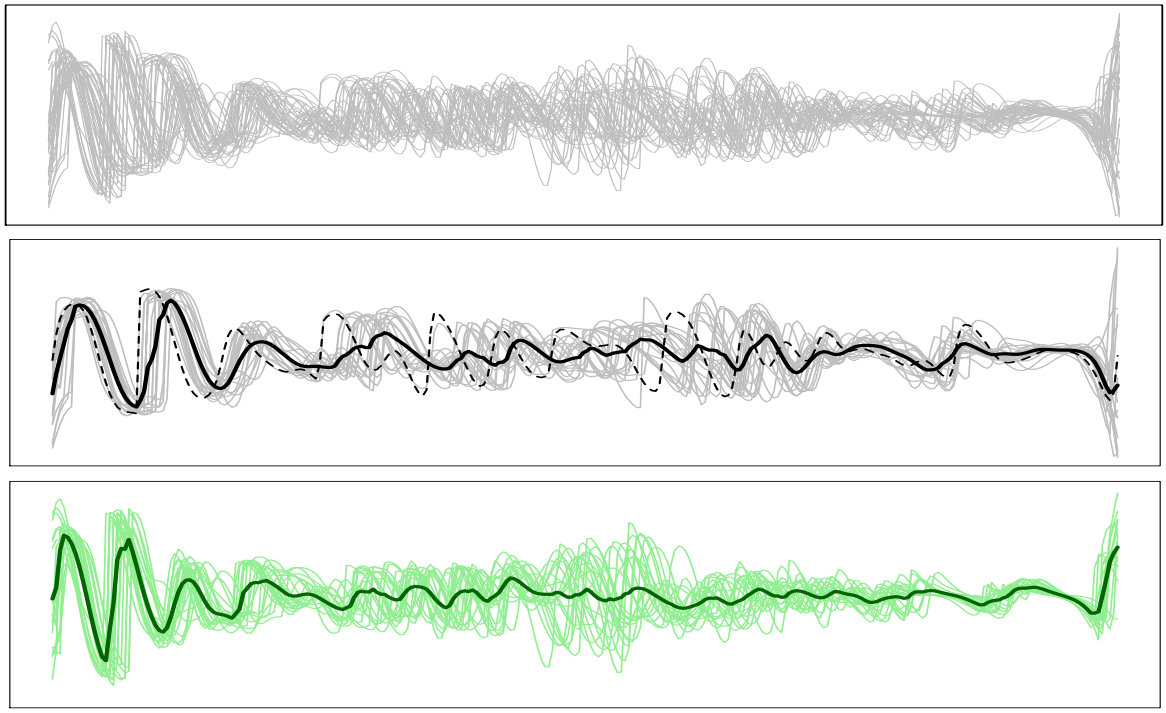

Figure 3: *Top panel: tangential accelerations corresponding to 40 signatures (20 original and 20 forged). Mid panel: one of the two clusters of curves obtained with the FMSA. 19 of the 20 original signatures are clustered together in this cluster. The FMSA assigns the dashed curve (corresponding to the remaining original signature) to a separate atomic cluster. Bottom panel: all of the 20 forged signatures are correctly clustered together in a second distinct cluster by the FMSA. The bold curves in the mid and in the bottom panels correspond to the two estimated functional modes.*

Ciollaro et al. (2014b) introduce a version of the mean-shift algorithm based on (8) which is tailored for functional data that belong to a potentially infinite-dimensional Hilbert space (the *functional mean-shift algorithm*, abbreviated FMSA in the following). An approximate bootstrap test for the significance of the functional modes estimated by the FMSA, inspired by Genovese et al., 2013, is also proposed in the same paper.

Figure 3 shows an example of modal clustering of a set of smooth curves corresponding to the tangential accelerations of 40 hand-written signatures[6] (20 original and 20 forged) obtained applying the FMSA with $d(x, y) = \|x - y\|_{L_2}$. Related work on mode estimation for functional data includes the work of Hall and Heckman (2002), who describe a non-adaptive gradient ascent method to locate the modes of $p_h(\cdot) = E_P\, K\left(\frac{d(X,\cdot)}{h}\right)$, and Dabo-Niang et al. (2004).

A careful extension of the framework described above to the case where p is a pseudo-density

[6]The dataset is part of the 2004 Signature Verification Competition, `http://www.cse.ust.hk/svc2004/index.html`. See also Geenens (2011).

as in equations (2) or (3) and $\mathcal{X}$ is not a finite dimensional space, but rather an infinite dimensional separable Hilbert space, poses some interesting challenges as I explain in the next section.

2.1.1 Future work

If one considers the problem of equation (5) when $\mathcal{X}$ is an infinite-dimensional Hilbert space, some complications arise and strong assumptions may be needed on both the pseudo-density p and on the sample space $\mathcal{X}$ in order to develop a sound theory of modal clustering for functional data. One is interested in the solution $\pi_{x_0} : \mathbb{R}_+ \to \mathcal{X}$ of the problem

$$\begin{cases} \pi'_{x_0}(t) = \nabla p(\pi_{x_0}(t)) \\ \pi_{x_0}(0) = x_0. \end{cases} \tag{9}$$

where now $\nabla p(x)$ is the *functional gradient* of p at $x \in \mathcal{X}$, i.e. the element of $\mathcal{X}$ such that the Fréchet derivative[7] of p at x, $Dp(x)$, can be written as

$$Dp(x)(\cdot) = \langle \cdot, \nabla p(x) \rangle \tag{10}$$

by means of the Riesz representation theorem. Our goals towards the development of a theory of modal clustering for functional data include:

1. clarify the conditions under which the gradient flow on p exists and the gradient ascent path starting from any arbitrary point $x_0 \in \mathcal{X}$ is unique and converges; establishing these facts is a key step to ensure that the population clustering induced by the population pseudo-density p is a well-defined object as it is in the finite dimensional case considered in Chacón (2012) and Chacón (2014)

2. developing confidence regions for the functional modes of p and developing statistical tests for the significance of the estimated functional modes

3. extending the clustering consistency theorem of Chacón (2014) to the functional case

4. obtaining uniform rates of consistency of the pseudo-density estimator $\hat{p}$ of equation (8) and of its derivatives; this would be a natural extension of the results of Ferraty et al. (2010) and Ferraty et al. (2012a).

Regarding 1., the standard theory of partial differential equations guarantees that a unique solution π_{x_0} to the initial value problem of equation (9) exists if the functional gradient of p is a Lipschitz map[8]. On the other hand, while the conditions required to have convergent (as $t \to \infty$) gradient ascent paths starting at any arbitrary point $x_0 \in \mathcal{X}$ are essentially always

[7] The Fréchet derivative of a functional p on a Banach space $\mathcal{X}$ at a point $x \in \mathcal{X}$ is the continuous linear operator (thus an operator that belongs to $\mathcal{X}^*$, the dual of $\mathcal{X}$) $Dp(x)(\cdot) : \mathcal{X} \to \mathbb{R}$ satisfying $\frac{\|p(x+h)-p(x)-Dp(x)(h)\|}{\|h\|} \to 0$ as $\|h\| \to 0$.

[8] The functional gradient of p, ∇p, is a Lipschitz map if and only if there exists $L > 0$ such that $\|\nabla p(x) - \nabla p(y)\| \leq \|x - y\|$ for any pair $x, y \in \mathcal{X}$.

trivially satisfied in the finite-dimensional case, when $\mathcal{X}$ is infinite-dimensional it is hard to guarantee convergence in general. Thus, either one is willing to deal with a possibly large subset of points of $\mathcal{X}$ for which the gradient ascent lines starting do not necessarily converge, or some compactness condition on $\mathcal{X}$ has to be imposed together with appropriate conditions on the first and the second derivatives of p to rule out divergence. Quite possibly, one has to rely on compact embedding theorems in the latter case. For example, for the case of smooth curves, if one decides to measure the distance between two curves by means of the L_2 distance (thus $d(x, y) = \|x - y\|_{L_2}$ in equation (3)), one may have to introduce an assumption of the type

$$P\left(X \in \overline{B_{H_1}(0, M)}^{L_2}\right) = 1 \text{ for some } M > 0, \tag{11}$$

where $\overline{B_{H_1}(0, M)}^{L_2}$ indicates the L_2-closure of the H_1-ball[9] of radius M. In fact, $\overline{B_{H_1}(0, M)}^{L_2}$ can be compactly embedded in L_2. An assumption of this type appears somewhat strong and would suggest that modal clustering of functional data could be feasible only when both the pseudo-density and its effective domain are sufficiently regular. Furthermore, another problem likely arise at this point if one assumes (11). How can one constrain the gradient ascent paths on p to stay within $\overline{B_{H_1}(0, M)}^{L_2}$ or even H_1? One may naively require p to be null outside of $\overline{B_{H_1}(0, M)}^{L_2}$, or $p(x) \to 0$ as $\|x\|_{H_1} \to \infty$, but then one can show that any of the previous two conditions together with the L_2 continuity of p imply that $p = 0$! A different route is imposing appropriate conditions on the Fréchet derivative of p in such a way constrain the gradient flow; otherwise, one may consider the gradient flows on p and $\widehat{p}$ with respect to the H_1 norm rather than the L_2 norm as in Jung et al. (2009) for instance (although the change of norm alone is not sufficient to guarantee the convergence of the gradient ascent paths to elements of H_1). Mathematically, this last hypothesis makes the problem rather interesting as one is then dealing with two different topologies at the same time: the L_2 topology can be thought as the 'main' topology for the problem, since the gradient ascent paths would converge with respect to the L_2 norm; on the other hand, the H_1 topology can be seen as an 'auxiliary' topology used to efficiently climb the pseudo-density and locate the functional modes.

Regarding 2., a natural starting point to perform statistical inference on the estimated clustering structure associated to a pseudo-density is the development of significance tests and/or confidence regions for the estimated functional modes and the curvature of p at the estimated local modes. An approximate test for the significance of the estimated functional modes based on the bootstrap is proposed in Ciollaro et al. (2014b), but a thorough theoretical justification and the consistency of the bootstrap still need to be clearly established for that procedure.

Regarding 3., once the theory of modal clustering is rigorously adapted to functional data, it would be of great interest to generalize the clustering consistency theorem of Chacón (2014). Chacón (2014) shows that for a univariate density p with compact support and a sequence of density estimators $(\widehat{p}_n)_{n=1}^{\infty}$ such that the corresponding sequence of derivatives $\left(\widehat{p}_n^{(j)}\right)_{n=1}^{\infty}$ converge uniformly almost surely to $p^{(j)}$ for $j = 0, 1, 2$, the sequence of clusterings induced

[9] $H_1([0, 1])$ is the Sobolev space of weakly differentiable functions on $[0, 1]$ with finite H_1 norm: $\|x\|_{H_1} = \|x\|_{L_2} + \|x'\|_{L_2}$.

by the estimated density, $\left(\widehat{\mathcal{C}}_n\right)_{n=1}^{\infty}$, is such that $d_H\left(\widehat{\mathcal{C}}_n, \mathcal{C}\right) \to 0$ and $d_P\left(\widehat{\mathcal{C}}_n, \mathcal{C}\right) \to 0$ where $\mathcal{C}$ is the unknown population clustering induced by p, and d_H and d_P indicate distances between partitions based on the Hausdorff distance and on the distance in measure respectively. An extension of the clustering consistency theorem to the high-dimensional or to the infinite-dimensional setting appears a challenging goal in light of the difficulties outlined above. Moreover, as noted by the author,

"The proof of this result [the clustering consistency theorem] is shown in the appendix and, as stated, it covers only the univariate case. The analysis of the proposed distances between clusterings is greatly simplified in the univariate case, since the cluster boundaries are solely determined by the points of local minima of the density. The extension of this result for dimension $d \geq 2$ seems quite challenging and far beyond the scope and length of the present paper, since the cluster boundaries in dimension d are $(d-1)$-dimensional manifolds which may have very intricate forms.".

Finally, regarding 4., on the basis of the previous work of Ferraty et al. (2010), we expect that whenever $\mathcal{X}_0$ is a subset of $\mathcal{X}$ with finite Kolmogorov entropy and the pseudo-density p satisfies

$$|p(x) - p(y)| \leq d^b(x, y) \tag{12}$$

for any $x, y \in \mathcal{X}_0$ and for some $b > 0$ (and if some additional technical conditions hold), then

$$\sup_{x \in \mathcal{X}_0} |\hat{p}(x) - p(x)| = O\left(h^b\right) + O_{a.co.}\left(\sqrt{\frac{\psi_{\mathcal{X}_0}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right) \tag{13}$$

as $n \to \infty$. The notation $O_{a.co.}$ is used to mean boundedness in the sense of almost complete convergence[10], ϕ is the concentration function of equation (3), and $\psi_{\mathcal{X}_0}(\epsilon)$ is the Kolmogorov ϵ-entropy of $\mathcal{X}_0$, i.e. the logarithm of the minimal number of balls of radius ϵ needed to cover $\mathcal{X}_0$. Similar rates of consistency are expected for the derivatives of $\hat{p}$ under appropriate smoothness assumptions, although to the best of our knowledge no results exist at the moment in this direction.

2.2 Functional regression for quasar spectra[11]

We observe noisy light curves f_i such as that depicted in Figure 4. These curves can be thought as being generated according to the model

$$f_i(t) = (1 - \alpha_i(t))f_i^*(t) + \epsilon_i(t), \tag{14}$$

[10] Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. X_n converges almost completely to 0 if and only if $\sum_{n=1}^{\infty} P(|X_n| > \epsilon) < \infty$ for any $\epsilon > 0$. Almost complete convergence implies almost sure convergence by the first Borel-Cantelli lemma. If $(a_n)_{n=1}^{\infty}$ is a sequence of positive numbers, we say that $X_n = O_{a.co.}(a_n)$ if and only if $\exists \epsilon_0 > 0$ such that $\sum_{n=1}^{\infty} P(|X_n| > \epsilon_0 a_n) < \infty$.

[11] **Ongoing work with Jessi Cisewski, Peter Freeman, Christopher Genovese, Jing Lei, Ross O'Connell, and Larry Wasserman.**

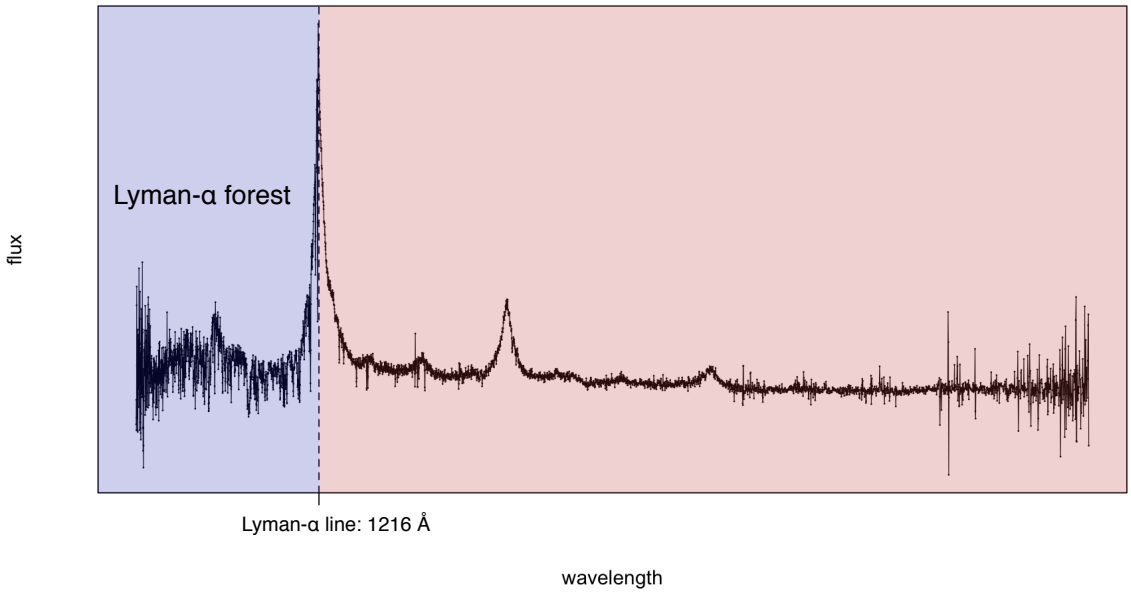

Figure 4: *A quasar light curve from the BOSS catalog. The Lyman-α line, corresponding to the sharp peak in the light curve at 1216 Å, marks the boundary between the Lyman-α forest ($\mathcal{T}_0$, blue) and the absorption-free region ($\mathcal{T}_1$, red). The goal is to predict the continuous component of the curve on the blue side (before the absorption process) using the information contained in the smooth component of the curve on the red side).*

where f_i^* is an underlying smooth function, ϵ_i are i.i.d. mean zero error processes and $\alpha_i \in [0,1]$ are absorption processes[12]. The absorption processes α_i are relevant because they convey important physical information regarding the distribution of matter in regions of our Universe that cannot be easily probed otherwise, if not by means of the Lyman-α forest. The scientific goal is therefore drawing inferences on some features of these processes. In this setting, the variable $t \in \mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$ is a wavelength in the ultraviolet portion of the light spectrum, $\mathcal{T}_0 \cap \mathcal{T}_1 = \emptyset$, and $\alpha_i = 0$ on $\mathcal{T}_1$. Our goal is to estimate f_i^* (which can be thought as a nuisance parameter from a scientific perspective) on $\mathcal{T}_0$ so that this estimate can be subtracted from the corresponding observed light curve f_i to finally get an estimate of the absorption level α_i on $\mathcal{T}_0$ (which, from a scientific standpoint, is the parameter of interest). To do that, one can use a smooth estimate of f_i^* on $\mathcal{T}_1$ which can be obtained by directly smoothing f_i on the absorption-free domain $\mathcal{T}_1$ where $\alpha_i = 0$ (using, say, a local polynomial estimator) as the predictor function in the regression model

$$E\left(f^*_{i,\mathcal{T}_0}(t) \mid \widehat{f}^*_{i,\mathcal{T}_1}\right) = r_{\widehat{f}^*_{i,\mathcal{T}_1}}(t). \tag{15}$$

In equation (15), $r \in \mathcal{R}$ is an operator mapping functions with domain $\mathcal{T}_1$ into functions with domain $\mathcal{T}_0$, and $\mathcal{R}$ is a suitable class of such operators. Once an estimate $\widehat{r}$ of r is available,

[12]More precisely, equation (14) can be reparametrized as $f_i(t) = e^{-\tau_i(t)} f_i^*(t) + \epsilon_i(t)$, where $\tau_i(t)$ represents the *optical depth* (essentially, the local density) of the intervening hydrogen absorbers at wavelength t.

the smooth component f^* on $\mathcal{T}_0$ can be estimated as

$$\widehat{f}^*_{i,\mathcal{T}_0}(t) = \widehat{r}_{\widehat{f}^*_{i,\mathcal{T}_1}}(t) \tag{16}$$

and the intensity of the absorption can be quantified by means of

$$\widehat{\alpha}_i(t) = \left(\widehat{f}^*_{i,\mathcal{T}_0}(t) - f^*_{i,\mathcal{T}_0}(t)\right) / \widehat{f}^*_{i,\mathcal{T}_0}(t). \tag{17}$$

Suppose that we possess a catalog of quasar spectra for which the absorption processes α_i are approximately null both on $\mathcal{T}_0$ and $\mathcal{T}_1$. Then the regression operator r can be consistently estimated by means of the Nadaraya-Watson type kernel smoother (Ferraty et al., 2012b)

$$\widehat{r}_{\widehat{f}^*_{\text{target},\mathcal{T}_1}}(t) = \frac{\sum_{i\in\text{catalog}} K\left(\frac{d\left(\widehat{f}^*_{i,\mathcal{T}_1}, \widehat{f}^*_{\text{target},\mathcal{T}_1}\right)}{h}\right) \widehat{f}^*_{i,\mathcal{T}_0}(t)}{\sum_{i\in\text{catalog}} K\left(\frac{d\left(\widehat{f}^*_{i,\mathcal{T}_1}, \widehat{f}^*_{\text{target},\mathcal{T}_1}\right)}{h}\right)}, \tag{18}$$

where $\widehat{f}^*_{\text{target},\mathcal{T}_1}$ is the estimated smooth component of the target quasar spectrum on the absorption-free region $\mathcal{T}_1$, $\widehat{f}^*_{i,\mathcal{T}_0}$ are estimates of the smooth component on $\mathcal{T}_0$ of the quasar light curves in the aforementioned absorption-free catalog, $h > 0$ is a bandwidth parameter, d is a semi-metric, and K is a kernel function. To get an estimate of the smooth component $f^*_{\text{target},\mathcal{T}_0}$ for the given target spectrum, we finally set

$$\widehat{f}^*_{\text{target},\mathcal{T}_0}(t) = \widehat{r}_{\widehat{f}^*_{\text{target},\mathcal{T}_1}}(t) \tag{19}$$

for $t \in \mathcal{T}_0$.

The above methodology is described in detail in Ciollaro et al. (2014a) and applied both on mock quasar spectra and spectra from the Hubble Space Telescope Faint Object Spectrograph (HST-FOS) and the Baryon Oscillation Spectroscopic Survey (BOSS) catalogs. The same paper also describes a type of prediction bands for the functional response variable that are based on the conformal prediction principle (Vovk et al., 2005, Vovk et al., 2009, Lei et al., 2013) and have finite sample coverage guarantees. Figure 5 depicts the prediction of the smooth component of the Lyman-α forest on two BOSS quasar spectra when the nonparametric functional regression model described above is fitted on 89 HST-FOS quasar spectra for which the absorption component in the Lyman-α forest is essentially null. The bands around the predicted curves on the left hand side of the pictures correspond to conformal prediction bands with finite sample coverage of at least 90%.

2.2.1 Future work

The problem of predicting the smooth component of a quasar light curve in the Lyman-α forest presents a series of additional methodological challenges and also raises some interesting theoretical questions. These include:

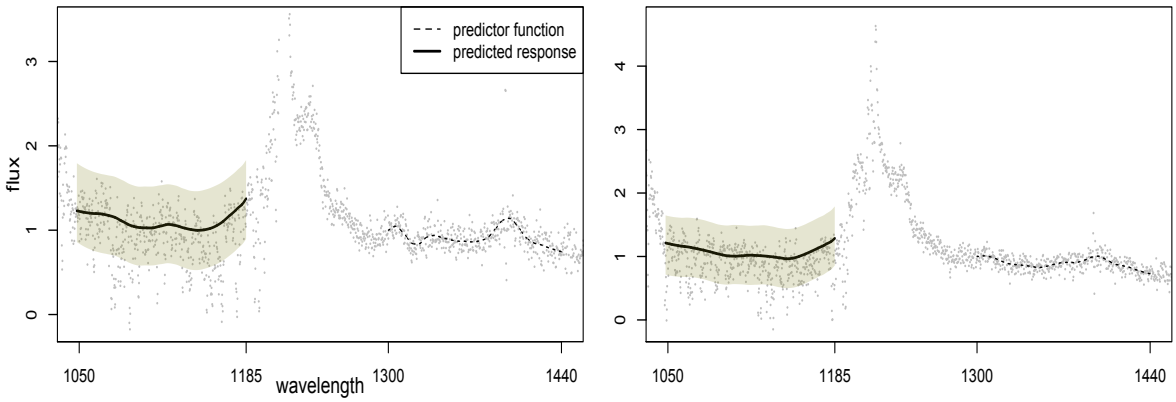

Figure 5: *Prediction of the smooth component of the Lyman-α forest on two BOSS quasar spectra. The nonparametric functional regression model is fitted on a sample of 89 HST-FOS spectra for which the absorption component in the Lyman-α forest is essentially null. The yellow bands correspond to conformal prediction bands with finite sample coverage of at least 90%.*

1. improving the current model and its ability to accurately predict the amplitude of the smooth component of the Lyman-α forest

2. analyzing from a theoretical perspective a new class of regression models in which the 'noise' term can be separated into two distinct components having different degrees of smoothness

3. developing fully-functional prediction bands and confidence bands that avoid the need to project the curves onto some finite-dimensional subspace

4. producing prediction of the smooth component in the Lyman-α forest for quasars catalogs of current large scale sky surveys such as the SDSS III/BOSS survey (`https://www.sdss3.org/surveys/boss.php`), and making a thorough comparison with already available predictions obtained with other methods.

Regarding 1., in Ciollaro et al. (2014a) it is noted that while the predictions obtained by means of the above nonparametric functional regression model tend to be accurate in terms of the predicted shape, the amplitude of the prediction is often incorrect even on mock spectra (see Figure 6). The same difficulty has been noted by other authors when other methods are used to perform the prediction, and some ad hoc post-prediction adjustments have been proposed (see for instance the mean-flux regulation procedure of Lee et al., 2012). It is known that the smooth component of quasar spectra tends to follow, at least approximately, a power law; however, some authors observed that a break in this behavior occurs somewhere between 1200 Å and 1300 Å (Zheng et al., 1997, Desjacques et al., 2007, Pâris et al., 2011, Lee et al., 2012) and thus the amplitude on the right-hand side of a quasar spectrum may carry little information about the amplitude of the smooth component of the Lyman-α forest. It is clear that an improvement of the prediction model in this direction is both scientifically and methodologically desirable. In light of the argument above, it appears that some information

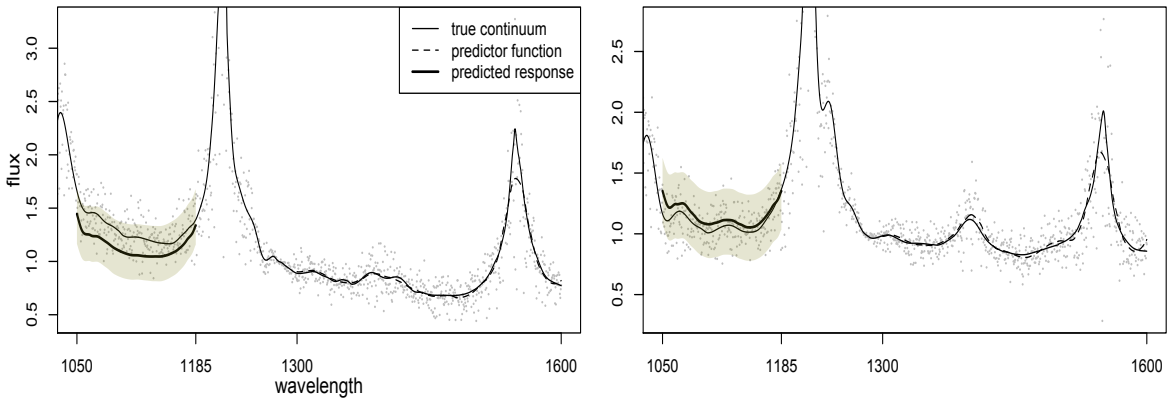

Figure 6: *Prediction of the smooth component of the Lyman-α forest on two mock quasar spectra. The shape of the predicted continuum matches that of the true continuum, but the amplitude of the continuum is not correctly recovered. The yellow bands correspond to conformal prediction bands with finite sample coverage of at least 90%.*

contained in the Lyman-α forest side of the quasar spectra (the domain of the response function) must necessarily be used to improve the accuracy of predicted amplitude.

Regarding 2., from the perspective of mathematical statistics, this application motivates the theoretical study of a class of nonparametric regression models of the type

$$y = f(x) + \epsilon = f^*(x) + \alpha + \epsilon \tag{20}$$

or

$$y = f(x) + \epsilon = \alpha f^*(x) + \epsilon, \tag{21}$$

where ϵ may be a simple white noise process and f^* is a smooth function in an appropriate class. The above regression models differ from the more familiar setting

$$y = f^*(x) + \epsilon \tag{22}$$

because of the presence of the additional (additive or multiplicative) stochastic component α, which is assumed to have smoothness features that should make it separable both from f^* and ϵ. The goal in the context of the regression models considered above would be to understand under which conditions one is able to disentangle the contributions from the three distinct components f^*, α and ϵ, and then make inferences both on f^* and α.

Regarding 3., the conformal prediction bands C_n for the functional response proposed in Ciollaro et al. (2014a) exhibit the following property for any n and for any α:

$$P\left(Y \in C_n(X)\right) = P\left(\sup_t |Y(t) - \hat{r}(X)(t)| \leq q\right) \geq 1 - \alpha, \tag{23}$$

for some $q > 0$ which is in practice chosen by means of a sample splitting procedure. The adjective 'conformal' comes from the fact that the band is built on the basis of the score

$$c(x, y; X_1, \ldots, X_n, Y_1, \ldots, Y_n) = -\sup_t |y(t) - \hat{r}(x)(t)| \tag{24}$$

which represents a measure of the 'conformity' of the test pair (x, y) with the sample $\{(X_i, Y_i)\}_{i=1}^n$. The band guarantees coverage of at least $1 - \alpha$ for any finite sample of size n. The above probability statement concerns the joint distribution of the pair (X, Y) (here denoting the functional predictor and the functional response respectively), and not the conditional distribution of Y given X. This is why this type of band is said to be *marginally valid*. In regression, it is often useful to have prediction bands that are valid with respect to the conditional distribution of Y given X. While it is not possible to have non-trivial conformal prediction bands that are *conditionally valid* and have finite sample coverage guarantees (see Lemma 1 of Lei and Wasserman, 2014), one can in principle use the bootstrap to obtain conditional prediction bands with asymptotic coverage guarantees. Therefore, a theoretical future goal is the development of valid bootstrap prediction bands for the case nonparametric regression model with functional predictor and functional response. Using a similar approach, I aim at developing bootstrap-based confidence bands for the regression operator $\hat{r}(x)(\cdot) = E(Y \mid X = x)(\cdot)$ that avoid the need of projecting the curve $r(x)$ onto the finite-dimensional subspace spanned by the first k functions of an orthogonal basis (as proposed in Ferraty et al., 2012b, for example).

Regarding 4., we aim at complementing the set of continuum templates that are currently available for cosmological analyses based on the Lyman-α forest. In contrast to the currently available templates, these estimates shall include some quantitative evaluation of their predictive accuracy (e.g. the width of the associated prediction bands), and thus will improve subsequent analyses involving optimal weighting schemes that depend on the uncertainty associated with the prediction procedure. An example of such analyses is the estimation of the correlation function of the relative flux absorption α_i across different locations in our Universe (i.e. across different quasar lines of sight[13]), as discussed in Ciollaro et al. (2014a). Once an estimate of α_i is available for a given quasar line of sight, it is clear from equation (17) that its contribution in the computation of the correlation function should depend (also) on the uncertainty associated to the prediction of the smooth component of the Lyman-α forest. However, current estimates of the correlation function do not generally account for the uncertainty associated with the prediction of $f^*_{\mathcal{T}_0}$.

References

Ery Arias-Castro, David Mason, and Bruno Pelletier. On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *unpublished manuscript*, 2013.

Claus Borggaard and Hans H. Thodberg. Optimal minimal neural interpretation of spectra. *Analytical Chemistry*, 64(5):545–551, 1992.

Denis Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Springer, 2000.

José E. Chacón. Clusters and water flows: a novel approach to modal clustering through morse theory. *arXiv preprint arXiv:1212.1384*, 2012.

[13]Informally, a quasar line of sight is the line segment joining the quasar and the observer.

José E. Chacón. A population background for nonparametric density-based clustering. *arXiv preprint arXiv:1408.1381*, 2014.

Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

Mattia Ciollaro, Jessi Cisewski, Peter E. Freeman, Christopher R. Genovese, Jing Lei, Ross O'Connell, and Larry Wasserman. Functional regression for quasar spectra. *arXiv preprint arXiv:1404.3168*, 2014a.

Mattia Ciollaro, Christopher R. Genovese, Jing Lei, and Larry Wasserman. The mean-shift algorithm for mode hunting and clustering in infinite dimensions. *arXiv preprint arXiv:1408.1187*, 2014b.

Antonio Cuevas. A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23, 2014.

Sophie Dabo-Niang, Frédéric Ferraty, and Philippe Vieu. Estimation du mode dans un espace vectoriel semi-normé. *Comptes Rendus Mathematique*, 339(9):659–662, 2004.

Aurore Delaigle and Peter Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38(2):1171–1193, 2010.

Vincent Desjacques, Adi Nusser, and Ravi K. Sheth. The probability distribution function of the lyman-α transmitted flux from a sample of sloan digital sky survey quasars. *Monthly Notices of the Royal Astronomical Society*, 374(2):206–219, 2007.

Frédéric Ferraty and Yves Romain. *The Oxford Handbook of Functional Data Analaysis*. Oxford University Press, 2011.

Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, 2006.

Frédéric Ferraty, Ali Laksaci, and Philippe Vieu. Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9(1):47–76, 2006.

Frédéric Ferraty, Ali Laksaci, Amel Tadj, and Philippe Vieu. Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, 140(2):335–352, 2010.

Frédéric Ferraty, Nadia Kudraszow, and Philippe Vieu. Nonparametric estimation of a surrogate density function in infinite-dimensional spaces. *Journal of Nonparametric Statistics*, 24(2):447–464, 2012a.

Frédéric Ferraty, Ingrid Van Keilegom, and Philippe Vieu. Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109:10–28, 2012b.

Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975.

Theo Gasser, Peter Hall, and Brett Presnell. Nonparametric estimation of the mode of a distribution of random curves. *Journal of the Royal Statistical Society, Series B*, 60(4): 681–691, 1998.

Gery Geenens. A nonparametric functional method for signature recognition. In Frédéric Ferraty, editor, *Recent Advances in Functional Data Analysis and Related Topics*, Contributions to Statistics, pages 141–147. Physica-Verlag HD, 2011.

Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Nonparametric inference for density modes. *arXiv preprint arXiv:1312.7567*, 2013.

Peter Hall and Nancy E. Heckman. Estimating and depicting the structure of a distribution of random functions. *Biometrika*, 89(1):145–158, 2002.

Lajos Horváth and Piotr Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

Miyoun Jung, Ginmo Chung, Ganesh Sundaramoorthi, Luminita A. Vese, and Alan L. Yuille. Sobolev gradients and joint variational image segmentation, denoising, and deblurring. In *IS&T/SPIE Electronic Imaging*, pages 72460I–72460I. International Society for Optics and Photonics, 2009.

John A. Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer, 2007.

Khee-Gan Lee, Nao Suzuki, and David N. Spergel. Mean-flux-regulated principal component analysis continuum fitting of sloan digital sky survey lyα forest spectra. *The Astronomical Journal*, 143(2):51–65, 2012.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society, Series B*, 76(1):71–96, 2014.

Jing Lei, Alessandro Rinaldo, and Larry Wasserman. A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, pages 1–15, 2013.

Wenbo V. Li and Qi-Man Shao. Gaussian processes: inequalities, small ball probabilities and applications. *Stochastic processes: theory and methods*, 19:533–597, 2001.

André Mas. Lower bound in regression for functional data by representation of small ball probabilities. *Electronic Journal of Statistics*, 6:1745–1778, 2012.

Isabelle Pâris, Patrick Petitjean, Emmanuel Rollinde, et al. A principal component analysis of quasar uv spectra at z $\sim$ 3. *Astronomy and Astrophysics*, 530(11), 2011.

James O. Ramsay and Bernard W. Silverman. *Functional Data Analysis*. Springer, 2005.

James O. Ramsay, Giles Hooker, and Spencer Graves. *Applied Functional Data Analysis - Methods and Case Studies*. Springer, 2002.

James O. Ramsay, Giles Hooker, and Spencer Graves. *Functional Data Analysis with R and MATLAB*. Springer, 2009.

Read D. Tuddenham and Margaret M. Snyder. Physical growth of california boys and girls from birth to age 18. *Publications in Child Development. University of California, Berkeley*, 1(2):183–364, 1954.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. On-line predictive linear regression. *The Annals of Statistics*, 37(3):1566–1590, 2009.

Wei Zheng, Gerard A. Kriss, Randal C. Telfer, John P. Grimes, and Arthur F. Davidsen. A composite hst spectrum of quasars. *The Astrophysical Journal*, 475(2):469–478, 1997.