Nonparametric regression on functional data: inference and practical aspects

Frédéric FERRATY [*], André MAS [†‡] and Philippe VIEU [§]

March 3, 2006

Abstract

We consider the problem of predicting a real random variable from a functional explanatory variable. The problem is attacked by mean of nonparametric kernel approach which has been recently adapted to this functional context. We derive theoretical results by giving a deep asymptotic study of the behaviour of the estimate, including mean squared convergence (with rates and precise evaluation of the constant terms) as well as asymptotic distribution. Practical use of these results are relying on the ability to estimate these constants. Some perspectives in this direction are discussed. In particular a functional version of wild bootstrapping ideas is proposed and used both on simulated and real functional datasets.

Key Words: Asymptotic Normality, Functional Data, Nonparametric Model, Quadratic Error, Regression, Wild Functional Bootstrap.

[*] Université Paul Sabatier, Toulouse 3 and Université Toulouse le Mirail, Toulouse 2

[†] Université Montpellier 2

[‡] Corresponding author: Laboratoire de Probabilités et Statistique CC051, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier, France, mas@math.univ-montp2.fr

[§] Université Paul Sabatier, Toulouse 3

1 Introduction

Functional data are more and more frequently involved in statistical problems. Developping statistical methods in this special framework has been popularized during the last few years, particularly with the monograph by Ramsay & Silverman (2005). More recently, new developments have been carried out in order to propose nonparametric statistical methods for dealing with such functional data (see Ferraty & Vieu, 2006, for large discussion and references). These methods are also called doubly infinite dimensional (see Ferraty & Vieu, 2003). Indeed these methods deal with infinite-dimensional (i.e. functional) data and with a statistical model which depends on an infinite-dimensional unknown object (i.e. a nonparametric model). This double infinite framework motivates the appellation of Nonparametric Functional Statistics for such kind of methods. Our paper is centered on the functional regression model :

$$Y = r(\mathcal{X}) + error, \tag{1}$$

where Y is a real random variable, $\mathcal{X}$ is a functional random variable (that is, $\mathcal{X}$ takes values in some possibly infinite-dimensional space) and where the statistical model assumes only smoothness restriction on the functional operator r. At this point, it worth noting that the operator r is not constrained to be linear. This is a Functional Nonparametric Regression model (see Section 2 for deeper presentation).

The aim of this paper is to extend in several directions the current knowledges about functional nonparametric regression estimates presented in Section 2. In Section 3.1 we give asymptotic mean squared expansions, while in Section 3.2 the limiting distribution is derived. The main novelty/difficuly along the statement of these results relies on the exact calculation of the leading terms in the asymptotic expressions. Section 4 points out how such results can be used when the functional variable belongs to standard families of continuous time process. The accuracy of our asymptotic results leads to interesting perspectives from a practical point of view: minimizing mean squared errors can govern automatic bandwidth selection procedure while the limiting distribution of the error is a useful tool for building confidence bands. To this end, we propose in Section 5 a functional version of the wild bootstrap procedure, and we use it, both on simulated and on real functional datasets, to get some automatic rule for choosing the bandwidth. The concluding section 6 contains some important open questions which emerge naturally from the theoretical results given in this paper, such as the theoretical study of the accuracy of the functional wild bootstrap procedure used in our applications.

2 Kernel nonparametric functional regression

2.1 The model

The model is defined in the following way. Assume that $(Y_i, \mathcal{X}_i)$ is a sample of n i.i.d. pairs of random variables. The random variables Y_i are real and the $\mathcal{X}_i$'s are random elements with values in a functional space $\mathcal{E}$. In all the sequel we will take for $\mathcal{E}$ a separable Banach space endowed with a norm $\|\cdot\|$. This setting is quite general since it contains the space of continuous functions, L^p spaces as well as more complicated spaces like Sobolev or Besov spaces. Separability avoids measurability problems for the random variables $\mathcal{X}_i$'s. The model is classically written :

$$Y_i = r\left(\mathcal{X}_i\right) + \varepsilon_i \quad i = 1, ..., n,$$

where r is the regression function mapping $\mathcal{E}$ onto $\mathbb{R}$ and the ε_i's are such that for all i, $E(\varepsilon_i|\mathcal{X}_i) = 0$ and $E(\varepsilon_i^2|\mathcal{X}_i) = \sigma_\varepsilon^2(\mathcal{X}) < \infty$.

2.2 The estimate

Estimating r is a crucial issue in particular for predicting the value of the response given a new explanatory functional variable $\mathcal{X}_{n+1}$. However, it is also a very delicate task because r is a nonlinear operator (from $\mathcal{E}$ into $\mathbb{R}$) for which functional linear statistical methods were not planned. To provide a consistent procedure to estimate the nonlinear regression operator r, we propose to adapt the classical finite dimensional Nadaraya-Watson estimate to our functional model. We set

$$\widehat{r}(\chi) = \frac{\sum_{k=1}^{n} Y_k K\left(h^{-1}\left\|\mathcal{X}_k - \chi\right\|\right)}{\sum_{k=1}^{n} K\left(h^{-1}\left\|\mathcal{X}_k - \chi\right\|\right)}.$$

Several asymptotic properties of this estimate were obtained recently. It turns out that the existing literature adresses either the statement of upper bounds of the rates of convergence without specification of the exact constants (see Chapter 6 in Ferraty & Vieu, 2006), or abstract expressions of these constants which are unusable in practice (as for instance in the recent work by Masry, 2005, which has been published during the reviewing process of this paper). Our aim in this paper is to give bias, variance, means square errors and asymptotic distribution of the functional kernel regression estimate with exact computation of all the constants (see Section 3). We will focus on practical purposes in Section 5.

Several assumptions will be made later on the kernel K and on the bandwidth h. Remind that in a finite-dimensional setting pointwise mean squared error (at χ) of the estimate depends on the evaluation of the density (at χ) w.r.t. Lebesgue's measure and on the derivatives of this density. We refer to Schuster (1972) for an historical result about this topic. On infinite-dimensional spaces, there is no measure universally accepted (as the Lebesgue one in the finite-dimensional case) and there is need for developping a "free-density" approach. As discussed along Section 4 the problem of introducing a density for $\mathcal{X}$ is shifted to considerations on the measure of small balls with respect to the probability of X.

2.3 Assumptions and notations

Only pointwise convergence will be considered in the forthcoming theoretical results. In all the following, χ is a fixed element of the functional space $\mathcal{E}$. Let φ be the real valued function defined as

$$\varphi(s) = E\left[\left(r(\mathcal{X}) - r(\chi)\right) \mid \left\|\mathcal{X} - \chi\right\| = s\right],$$

and F be the c.d.f. of the random variable $\|\mathcal{X} - \chi\|$:

$$F(t) = P(\|\mathcal{X} - \chi\| \leq t).$$

Note that the crucial functions φ and F depends implicitely on χ. Consequently we should rather note them by φ_χ and F_χ but, as χ is fixed, we drop this index once and for all. Similarly, we will use in the remaining the notation σ_ε^2 instead of $\sigma_\varepsilon^2(\mathcal{X})$. Let us consider now the following assumptions.

H0 : r and σ_ε^2 are continuous in a neighborhood of χ, and $F(0) = 0$.

H1 : $\varphi'(0)$ exists.

H2 : The bandwidth h satisfies $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nF(h) = \infty$, while the kernel K is supported on $[0,1]$, has a continuous derivative on $[0,1)$, $K'(s) \leq 0$ and $K(1) > 0$.

Assumptions **H0** and **H2** are clearly unrestrictive, since they are the same as those classically used in the finite-dimensional setting. Much more should be said on assumption **H1**. Note first that, obviously, $\varphi(0) = 0$. It is worth noting that, whereas we could expect assumptions on the local regularity of r (as in the finite-dimensional case), hypothesis **H1** skips over that point and avoids to go into formal considerations on differential calculus on Banach spaces. To fix the ideas, if we assumed differentiability of r, we would get by Taylor's expansion that

$$r(\mathcal{X}) - r(\chi) = \langle r'(\chi), \mathcal{X} - \chi \rangle + o(\|\mathcal{X} - \chi\|),$$

where $r'(\chi) \in \mathcal{E}^*$, $\mathcal{E}^*$ being the conjugate space of $\mathcal{E}$ and $\langle \cdot, \cdot \rangle$ being the duality bracket between $\mathcal{E}$ and $\mathcal{E}^*$. In this context, a non trivial link would appear between $\varphi'(0)$ and $r'(\chi)$ through the following relation:

$$\lim_{s\to 0} E\left[\left\langle r'(\chi), \frac{\mathcal{X} - \chi}{\|\mathcal{X} - \chi\|}\right\rangle \mid \|\mathcal{X} - \chi\| = s\right] = \varphi'(0).$$

Indeed, even if the link between the existency of $r'(\chi)$ and of $\varphi'(0)$ is strong, one can build counter-examples for their non equivalence (these counter-examples are available on request but they are out of the main scope of this paper). In the perspective of estimating the constants given in Theorem 1, it will be easier to estimate $\varphi'(0)$ (for instance by using $\widehat{r}$) than the operator $r'(\chi)$. Therefore, we prefer to express computations by mean of $\varphi'(0)$ instead of $r'(\chi)$. This has the additional advantage to produce more readable writings. Consequently, the differentiability of r is not needed.

Let us now introduce the function τ_h defined for all $s \in [0,1]$ as:

$$\tau_h(s) = \frac{F(hs)}{F(h)} = P\left(\|\mathcal{X} - \chi\| \leq hs \,|\, \|\mathcal{X} - \chi\| \leq h\right),$$

for which the following assumption is made:

H3 : For all $s \in [0,1]$, $\tau_h(s) \to \tau_0(s)$ as $h \to 0$.

Note that the function τ_h is increasing for all h. The measurable (as the pointwise limit of the sequence of measurable functions τ_h) mapping $\boldsymbol{\tau}_0$ is non decreasing. Let us finally mention that this function $\boldsymbol{\tau}_0$ will play a key role in our methodology, in particular when we will have to compute the exact constant terms involved in our asymptotic expansions. For the sake of clarity, the following proposition (whose a short proof will be given in the Appendix) will explicit the function τ_0 for various cases. By $1_{]0,1]}(\cdot)$ we denote the indicator function on the set $]0,1]$ and $\delta_1(\cdot)$ stands for the Dirac mass at 1.

Proposition 1

i) If $F(s) \sim Cs^\gamma$ for some $\gamma > 0$ then $\tau_0(s) = s^\gamma$.

ii) If $F(s) \sim Cs^\gamma |\ln s|^\kappa$ with $\gamma > 0$ and $\kappa > 0$ then $\tau_0(s) = s^\gamma$.

iii) If $F(s) \sim C_1 s^\gamma \exp(-C_2/s^p)$ for some $p > 0$ and some $\gamma > 0$ then $\tau_0(s) = \delta_1(s)$.

iv) If $F(s) \sim C/|\ln s|$ then $\tau_0(s) = 1_{]0,1]}(s)$.

A deeper discussion linking the above behavior of F with small ball probabilities notions will be given in Section 4.

3 Asymptotic study

In both following subsections we will state some asymptotic properties (respectively mean squared asymptotic evaluation and asymptotic normality) for the functional kernel regression estimate $\hat{r}$.

It is worth noting that all the results below can be seen as extensions to functional data of several ones already existing in the finite-dimensional case (the literature is quite extensive in this field and the reader will find in Sarda & Vieu (2000) deep results as well as a large scope of references). With other words, our technique for proving both Theorem 1 and Theorem 2 is also adapted to the scalar or vector regression model since the abstract space $\mathcal{E}$ can be of finite dimension (even, of course, if our main goal is to treat infinite-dimensional cases). Moreover, it turns out that the transposition to finite-dimensional situations of our key conditions (see discussion in Section 4 below) becomes (in some sense) less restrictive than what is usually assumed. With other words, the result of Theorem 1 and Theorem 2 can be directly applied to finite-dimensional settings, and will extend the results existing in this field (see again Sarda & Vieu, 2000) to situation when the density of the corresponding scalar or multivariate variable does not exist or has all its successive derivatives vanishing at point χ (see discussion in Section 4.3).

All along this section we assume that assumptions **H0-H3** hold. Let us first introduce the following notations:

$$M_0 = \left(K(1) - \int_0^1 (sK(s))' \tau_0(s)\, ds\right),$$
$$M_1 = \left(K(1) - \int_0^1 K'(s) \tau_0(s)\, ds\right),$$
$$M_2 = \left(K^2(1) - \int_0^1 \left(K^2\right)'(s) \tau_0(s)\, ds\right).$$

3.1 Mean Squared Convergence

The following result gives asymptotic evaluation of the mean squared errors of our estimate. The asymptotic mean squared errors have a standard convex shape, with large bias when the bandwidth h increases and large variance when h decays to zero. We refer to the Appendix for the proof of Theorem 1.

Theorem 1 *When **H0-H3** hold, we have the following asymptotic developments :*

$$E\widehat{r}(\chi) - r(\chi) = \varphi'(0)\frac{M_0}{M_1}h + O\left((nF(h))^{-1}\right) + o(h), \tag{2}$$

and

$$Var(\widehat{r}(\chi)) = \frac{1}{nF(h)}\frac{M_2}{M_1^2}\sigma_\varepsilon^2 + o\left(\frac{1}{nF(h)}\right). \tag{3}$$

3.2 Asymptotic Normality

Let us denote the leading bias term by:

$$B_n = \varphi'(0)\frac{M_0}{M_1}h.$$

Before giving the asymptotic normality, one has to be sure that the leading bias term does not vanish. This is the reason why we introduce the following additional assumption:

H4 : $\varphi'(0) \neq 0$ and $M_0 > 0$.

The first part of assumption **H4** is very close to what is assumed in standard finite-dimensional literature. It forces the nonlinear operator r not to be too smooth (for instance, if r is Lipschitz of order $\beta > 1$, then $\varphi'(0) = 0$). The second part of assumption **H4** is specific to the infinite-dimensional setting, and the next Proposition 2 will show that this condition is general enough to be satisfied in some standard situations. This proposition will be proved in the appendix.

Proposition 2

i) If $\tau_0(s) \neq 1_{]0,1]}(s)$ *and* τ_0 *is continuously differentiable on* $(0,1)$*, then* $M_0 > 0$ *for* *any kernel K satisfying **H2**.*

*ii) If $\tau_0(s) = \delta_1(s)$, then $M_0 > 0$ for any kernel K satisfying **H2**.*

To emphasize the interest of these results, they should be combined with those of Proposition 1. Note that the result *i*) includes the well-known family of processes for which $\tau_0(s) = s^\gamma$ (see Proposition 1-*i*), that is those whose distributions admit fractal dimensions (see Section 4.2). The second case when $\tau_0(s) = \delta_1(s)$ (for which a particular case is given in Proposition 1-*iii*) corresponds to nonsmooth processes (see Section 4.1). These two cases cover a large number of situations. However, if a more general function τ_0 has to be used, one can make additional hypotheses on the kernel K. In particular, if K is such that $\forall s \in [0,1]$, $(sK(s))' > 0$ and $\tau_0(s) \neq 1_{]0,1]}(s)$, then $M_0 > 0$, which covers the case of the uniform kernel. More complicated kernel functions K would lead to more technical assumptions linking K with τ_0. It is out of purpose to give these tedious details (available on request) but let us just note that the key restriction is the condition $\tau_0(s) \neq 1_{]0,1]}(s)$ (else we have $M_0 = 0$).

Moreover, since the rate of convergence depends on the function $F(h)$ and for producing a reasonably usable asymptotic distribution it is worth having some estimate of this function. The most natural is its empirical counterpart:

$$\widehat{F}(h) = \frac{\#\left(i : \|\mathcal{X}_i - \chi\| \leq h\right)}{n}.$$

The pointwise asymptotic gaussian distribution for the functional nonparametric regression estimate is given in Theorem 2 below which will be proved in the appendix. Note that the symbol $\hookrightarrow$ stands for "convergence in distribution".

Theorem 2 *When **H0-H4** hold, we have*

$$\sqrt{n\widehat{F}(h)}\left(\widehat{r}(\chi) - r(\chi) - B_n\right)\frac{M_1}{\sqrt{M_2\sigma_\varepsilon^2}} \hookrightarrow \mathcal{N}(0,1).$$

A simpler version of this result is stated in Corollary 1 below whose proof is obvious. The key-idea relies in introducing the following additional assumption:

H5 : $\lim_{n\to\infty} h\sqrt{n\,F(h)} \;=\; 0$

which allows to cancel the bias term.

Corollary 1 *When **H0-H5** hold, we have*

$$\sqrt{n\widehat{F}\,(h)}\,(\widehat{r}\,(\chi) - r\,(\chi))\,\frac{M_1}{\sigma_\varepsilon\sqrt{M_2}} \hookrightarrow \mathcal{N}\,(0,1)\,.$$

In practice, the constants involved in Corollary 1 need to be estimated. In order to compute explicitely both constants M_1 and M_2, one may consider the simple uniform kernel and get easily the following result:

Corollary 2 *Under assumptions of Corrollary 1, if* $K(.) = 1_{[0,1]}(.)$ *and if* $\widehat{\sigma}_\varepsilon^2$ *is a consistent estimator of* σ_ε^2*, then we have:*

$$\sqrt{\frac{n\widehat{F}\,(h)}{\widehat{\sigma}_\varepsilon^2}}\,(\widehat{r}\,(\chi) - r\,(\chi)) \hookrightarrow \mathcal{N}\,(0,1)\,.$$

There are many possibilities for constructing a consistent conditional variance estimate. One among all the possibilities consists in writing that

$$\begin{aligned}\sigma_\varepsilon^2(\chi) &= E((Y - r(\mathcal{X}))^2|\mathcal{X} = \chi),\\ &= E(Y^2|\mathcal{X} = \chi) - (E(Y|\mathcal{X} = \chi))^2,\end{aligned}$$

and, by estimating each conditional expectation with the functional kernel regression technique.

4 Some Examples of small ball probabilities

The distribution function F plays a prominent role in our methodology. This appears clearly in our conditions (through the function τ_0) and in the rates of convergence of our estimate (through the asymptotic behavior of the quantity $n\,F(h)$). More precisely, the behaviour of F around 0 turns out to be of first importance. In other words, the small ball probabilities of the underlying functional variable $\mathcal{X}$ will be determining. In order to illustrate our ideas and to connect with existing probabilistic knowledges in this field, let us now just discuss how F (and hence τ_0) behave for different usual examples of processes $\mathcal{X}$ valued in an infinite-dimensional space.

4.1 Nonsmooth processes

Calculation of the quantity $P\left(\|\mathcal{X}-\chi\|<s\right)$ for "small" s (i.e. for s tending to zero) and for a fixed χ is known as a "small ball problem" in probability theory. This problem is unfortunately solved for very few random variables (or processes) $\mathcal{X}$, even when $\chi = 0$. In certain functional spaces, taking $\chi \neq 0$ yield considerable difficulties that may not be overcome. Authors usually focus on gaussian random elements. We refer to Li & Shao (2001) for a survey on the main results on small ball probability. If $\mathcal{X}$ is a gaussian random element on the separable Banach space $\mathcal{E}$ and if χ belongs to the reproducing kernel Hilbert space associated with $\mathcal{X}$, then the following well-known result holds:

$$P\left(\|\mathcal{X}-\chi\|<s\right) \sim C_\chi P\left(\|\mathcal{X}\|<s\right), \quad \text{as } s \to 0. \tag{4}$$

So, the small ball problem at any point χ may be shifted to a small ball problem at 0. Moreover, (4) can be precised in a few situations. For instance, Mayer-Wolf & Zeitouni (1993) investigate the case when $\mathcal{X}$ is a one-dimensional diffusion process and χ satisfies some conditions (see Mayer-Wolf & Zeitouni, 1993, p15). They also briefly mention the non gaussian case (see Mayer-Wolf & Zeitouni, 1993, Remark 3, p19) but many other authors have considered different settings (see Ferraty et *al.*, 2005, for a large discussion and references therein). As far as we know, the results which are available in the literature are basically all of the form:

$$P\left(\|\mathcal{X}-\chi\|<s\right) \sim c_\chi\, s^{-\alpha} \exp\left(-C/s^{\beta}\right), \tag{5}$$

where α, β, c_χ and C are positive constants and $\|\cdot\|$ may be a sup, a L^p, a Besov norm ... The next remark is a direct consequence of Proposition 1. It proves that non-smooth processes may satisfy the assumptions needed to get the asymptotic expansions of previous sections.

Remark 1 *In the case of "non-smooth" processes defined by (5) we have* $\tau_0(s) = \delta_1(s)$*. In addition, condition* $n\,F(h) \to +\infty$ *(in **H2**) is checked as soon as* $h = A/(\log n)^{1/\beta}$ *for* A *large enough.*

4.2 Fractal (or geometric) processes

Another family of infinite dimensional processes is the class of fractal processes for which the small ball probabilities are of the form

$$P(\|\mathcal{X} - \chi\| < s) \sim c'_\chi\, s^\gamma, \tag{6}$$

where c'_χ and γ are once again positive constants. Like above, it is elementary to get the following result from Proposition 1.

Remark 2 *Under (6), we have* $\tau_0(s) = s^\gamma$ *while the condition* $n\,F(h) \to +\infty$ *(in **H2**) is satisfied as soon as* $h = An^{-B}$ *for* B *small enough.*

4.3 Back to the finite dimensional setting

Finally, it is important to note that a special case of fractal processs is given by the usual multivariate case (that is, by the case when $\mathcal{E} = \mathbb{R}^p$). The following result is obvious for the uniform norm on $\mathbb{R}^p$ and extends directly to any norm, since all of them are equivalent in finite dimension.

Remark 3 *If* $\mathcal{E} = \mathbb{R}^p$*, then any random variable* $\mathcal{X}$ *on* $\mathbb{R}^p$ *which has a finite and non zero density function at point* χ *satisfies (6) with* $\gamma = p$*.*

From Remarks 2 and 3, it is clear that all the results of Section 3 apply in a finite dimensional setting. Besides, the assumptions needed for Theorems 1 and

2 are weaker than those described in Remark 3 since there is no need to assume the existence of a density for $\mathcal{X}$. In this sense, our results extend the standard multivariate literature (see the discussion at the beginning of Section 3).

5 Perspectives on bandwidth choice

5.1 Introduction

The asymptotic results presented in the previous Section 3 are particularly appealing because, in addition to the specification of the rates of convergence, the exact constants involved in the leading terms of each result are precised. This is particularly interesting in practice. Let us focus now on the mean squared errors expansion given in Section 3.1. In fact, Theorem 1 could give clues for possible automatic bandwidth choice balancing the trade-off between variance and squared-bias effects. However, the constants are unknown in practice which could seem to be a serious drawback for practical purposes. This general problem is well-known in classical nonparametric statistics, but in our functional context this question gets even more intricate because of the rather complicated expression of M_0, M_1 and M_2. An appealing way to attack the problem is to use bootstrap ideas. In Section 5.2 we propose a track for building a functional version of the so-called wild bootstrap. We will show in Section 5.3, through some simulated examples, how this functional wild bootstrap procedure works on finite sample sizes for choosing automatically an optimal bandwidth. A case study, based on spectrometric functional data coming from the food industry, will be shortly presented in Section 5.4.

At this stage it is worth noting that we have no asymptotic support for this functional bootstrapping procedure. This open question will be one of the main point discussed in the concluding Section 6.

5.2 A Functional version of the wild bootstrap

Basically, when using bootstrapping techniques one expects to approximate directly the distribution of the error of estimation without having to estimate the leading terms involved in some asymptotic expansion of this error. In standard finite-dimensional problems (that is, when the variable $\mathcal{X}$ is valued in $\mathbb{R}^p$), a so-called wild bootstrap has been constructed for approximating the distribution of the error of estimation in kernel nonparametric regression. We refer to Härdle (1989) and Härdle & Marron (1991) for a previous presentation of the wild bootstrap in nonparameric regression. A selected set of additional references would include Mammen (2000) for the description of the state of art on nonparametric bootstrapping, Mammen (1993) for a large study of wild bootstrap, and Härdle, Huet & Jolivet (1995) for specific advances on wild bootstrap in (finite-dimensional) nonparametric regression setting.

The main interest of this kind of bootstrap relies on a resampling procedure of the residuals which makes it easily adaptable to our functional setting. Precisely, an adaptation to our functional setting could be the following functional wild bootstrap procedure:

i) Given the estimate $\hat{r}$ constructed with a bandwidth h, compute the residuals $\hat{\epsilon}_i = y_i - \hat{r}(\mathcal{X}_i)$, and construct a sequence of bootstrapped residuals such that each ϵ_i^* is drawn from a distribution G_i^* which is the sum of two Dirac distributions :

$$G_i^* = \frac{5+\sqrt{5}}{10}\delta_{\frac{\hat{\epsilon}_i(1-\sqrt{5})}{2}} + (\frac{5-\sqrt{5}}{10})\delta_{\frac{\hat{\epsilon}_i(1+\sqrt{5})}{2}}.$$

Such a distribution ensures that the first three moments of the bootstrapped residuals are respectively 0, $\hat{\epsilon}_i^2$ and $\hat{\epsilon}_i^3$ (see Härdle & Marron, 1991, for details).

ii) Given the bootstrapped residuals ϵ_i^*, and using a new kernel estimate $\tilde{r}$ which is defined as $\hat{r}$ but by using another bandwidth g, construct a bootstrapped sample $(\mathcal{X}_i^*, Y_i^*)$ by putting

$$\mathcal{X}_i^* = \mathcal{X}_i \ \text{ and } \ Y_i^* = \tilde{r}(\mathcal{X}_i) + \epsilon_i^*.$$

iii) Given the bootstrapped sample $(\mathcal{X}_i^*, Y_i^*)$, compute the kernel estimate $\hat{r}^*(\chi)$ which is defined as $\hat{r}$ (with the same bandwidth h) but using the bootstrapped sample $(\mathcal{X}_i^*, Y_i^*)$ instead of the previous sample $(\mathcal{X}_i, Y_i)$.

We suggest to repeat several times (let say N_B times) this bootstrap procedure, and to use the empirical distribution of $\hat{r}^*(\chi) - \tilde{r}(\chi)$ for bandwidth selection purpose. Precisely, the bootstrapped bandwidth is defined as follows:

Definition 1 *Given N_B replications of the above described bootstrapping scheme, and given a fixed set H of bandwidths, the bootstrapped bandwidth h^* is defined by:*

$$h^* = h^*(\chi) = \arg\min_{h \in H} \left(\frac{1}{N_B} \sum_{b=1}^{N_B} (\hat{r}^*(\chi) - \tilde{r}(\chi))^2 \right).$$

Of course, this procedure has still to be validated theoretically (see discussion in Section 6), but we will see in the next Sections 5.3 and 5.4 how it behaves both on simulated on and real data samples.

5.3 Some simulations

The aim of this section is to look at how the automatic bootstrapped bandwidth constructed in Definition 1 behaves on simulated samples. We construct random curves in the following way:

$$\mathcal{X}(t) = sin(\omega t) + (a + 2\pi)t + b, \ t \in (-1, +1),$$

where a and b (respectively ω) are r.r.v. drawn from a uniform distribution on $(0, 1)$ (respectively on $(0, 2\pi)$). Some of these curves are presented in Figure 1 below.

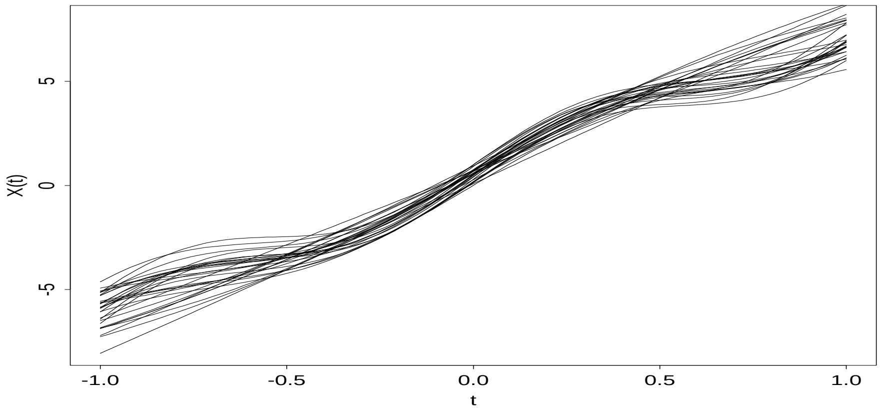

Figure 1: A sample of 30 simulated curves

The real response is simulated according to the following regression relation:

$$Y = r(\mathcal{X}) + \epsilon,$$

where

$$r(\mathcal{X}) = \int_{-1}^{+1} |\mathcal{X}'(t)|(1 - cos(\pi t))dt,$$

and where ϵ is drawn from a $\mathcal{N}(0, 2)$ distribution.

Our experience is based on the following lines. For each experience, we simulated two samples: a sample of size $n = 100$ on which all the estimates are computed and a testing sample of size $n = 50$ which is used to look at the behaviour of our method. Also, for each experience, the number of bootstrap replications was taken to be $N_B = 100$. Other values for J and N_B were also tried without changing the main conclusions. To improve the speed of our algorithm, the bandwidth h is assumed to belong to some grid in terms of nearest neighbours, that is

$$h = h(\chi) \in \{h_1, \ldots, h_{32}\} = H, \tag{7}$$

where h_k is the radius of the ball of center χ and containing exactly k among the curves data $\mathcal{X}_1, \dots \mathcal{X}_{100}$. Concerning the other parameters of our study, the kernel function K was chosen to be $K(u) = 1 - u^2,\ u \in (0, 1)$ and the norm $||.||$ was taken to be the L_2 one between the first order derivatives of the curves.

We computed, for the 32 different values of h, the average (over the χ's belonging to the second testing data sample) of the true error $(\hat{r}(\chi) - r(\chi))^2$ and of its bootstrap approximation $(\hat{r}^*(\chi) - \tilde{r}(\chi))^2$. Finally, this Monte Carlo scheme was repeated $J = 100$ times and the results are reported in Figure 2 (only 25 among the 100 curves are presented to make the plot clearer).

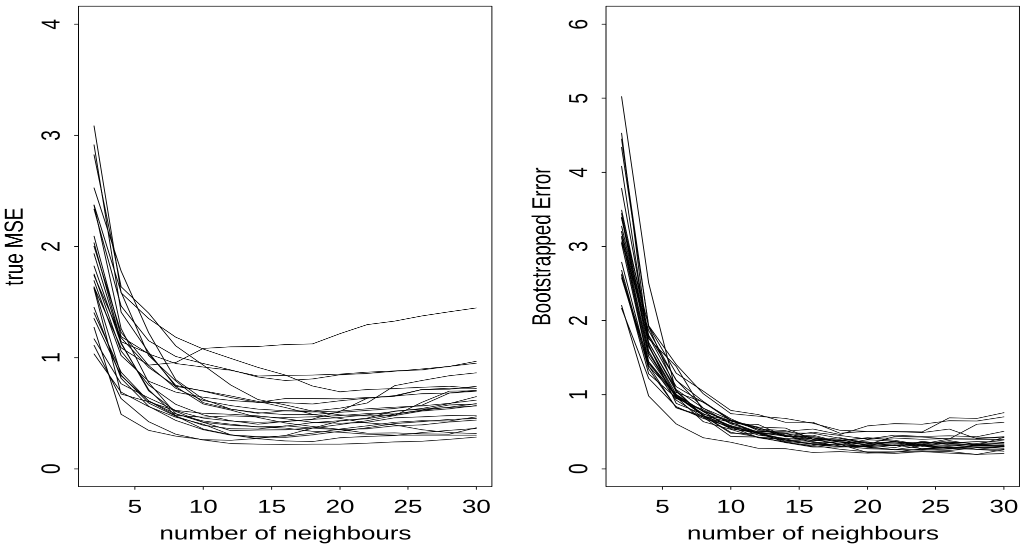

Figure 2: Simulations: True (left) and Bootstrap (right) Errors as functions of h

It appears clearly that both the theoretical quadratic loss and its data-driven bootstrapped version have the same convex shape. This convex shape is directly linked with the asymptotic expansion given in Theorem 1 before: large values of h give high bias, while small values of h lead to high variance. These results are quite promising in the sense that the similarity of the shapes of both sets of curves presented in Figure 2 let us expect that the bootstrapped bandwidths will be closed from the optimal ones. To check that point , we computed the theroretical minimal quadratic loss (that is, the error obtained by using the best bandwidth) and we compared it with the error obtained by using the boostrapped bandwidth h^*. This was done for each among the $J = 100$ experiences, and the results are reported in Figure 3 which gives mean, variance and density estimates of these two errors. Undoubtedly, these results show the good behaviour (at least on this example) of the bootstrapping method as an automatic bandwidth selection procedure. Of course, as discussed in Section 6, theoretical support for this functional bootstrap bandwidth selection rule is still an open question.

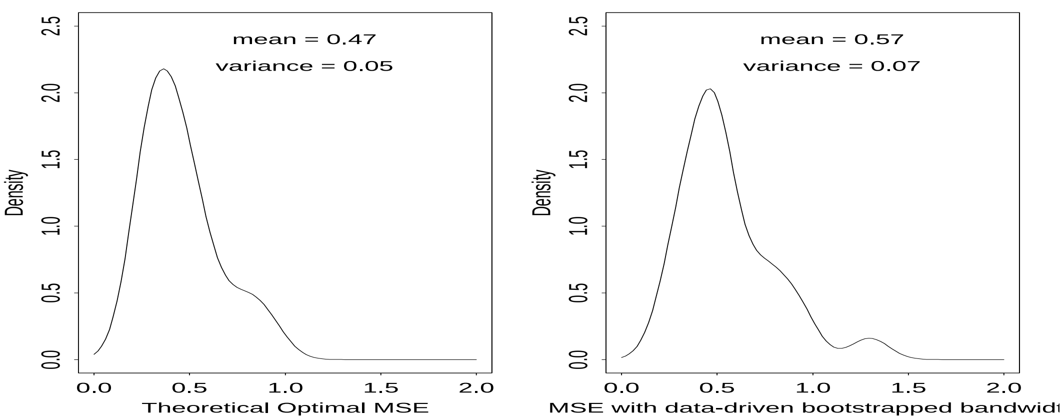

Figure 3: Simulations: MSE with optimal (left) and bootstrapped (right) bandwidth

5.4 A real data chemometric application

Let us now quickly show how our procedure is working on real data. These data contain of 215 spectra of light absorbance (Z_i, $i = 1, \ldots 215$) as functions of the wavelength, and observed on finely chopped pieces of meat. We present in Figure 4 the plots of the 215 spectra.

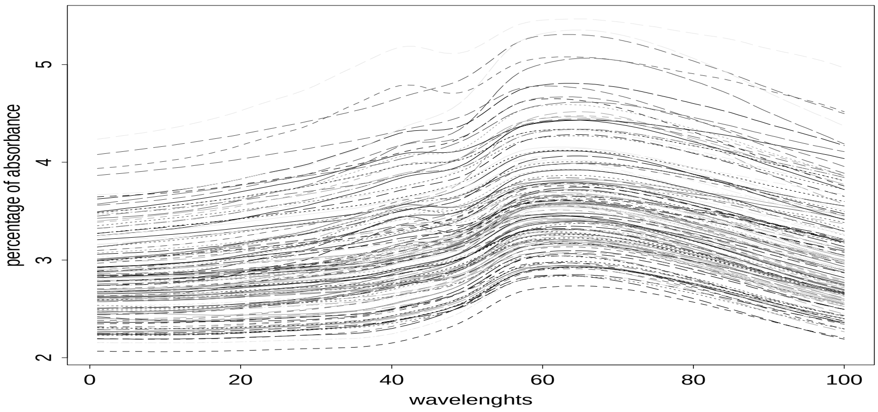

Figure 4: Spectrometric curves data

For each spectral cruve corresponds some real response Y_i which is the percentage of fatness, and our aim is to study the regression relation existing between the real variable Y and the functional one Z. These data have been widely studied and, inspired by previous studies (see Ferraty & Vieu, 2006) we decide to apply the functional kernel methodology to the curves $\mathcal{X} = Z''$, and by taking as norm $||.||$ between curves the usual L_2 norm between the second derivatives of the spectra. The kernel function K was chosen to be $K(u) = 1 - u^2,\ u \in (0, 1)$. Along our study we splitted the data into two subsamples. A first subsample of size $n = 165$ from which our estimates are computed, and a testing sample of size 50 on which they are applied.

In a first attempt, we used the automatic bootstrapping bandwidth selection rule, where H was defined as in (7). We present in Figure 5 the shape of the Bootsrapped Mean Square Error as a function of the number of neighbours (and thus, as function of the bandwidth).

The same convex form as for the simulated data appears. This form matches the theoretical results obtained in Section 3.1, with high bias for large values of h and high variance for small bandwidths. These bootstrapped errors are, in this example, minimal for the value $k = 8$. That means that, for each new curve χ to be predicted,

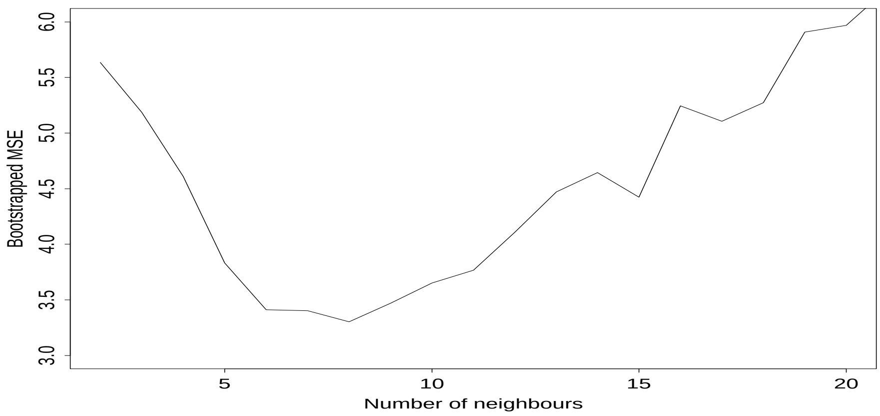

Figure 5: Spectrometric data: Bootstrapped Errors as function of the bandwidth

the data-driven bootstrapped bandwidth $h^*(\chi)$ is such that there are exactly 8 curves-data which are falling inside of the ball of radius $h^*(\chi)$.

These bandwidths lead to completely automatic data-driven fat contents prediction. For instance, we present in Figure 6 the fat content predictions for the 50 spectra in our testing sample. In order to highlight the nice behaviour of our prediction algorithm, Figure 6 plots the predicted values as functions of the true ones.

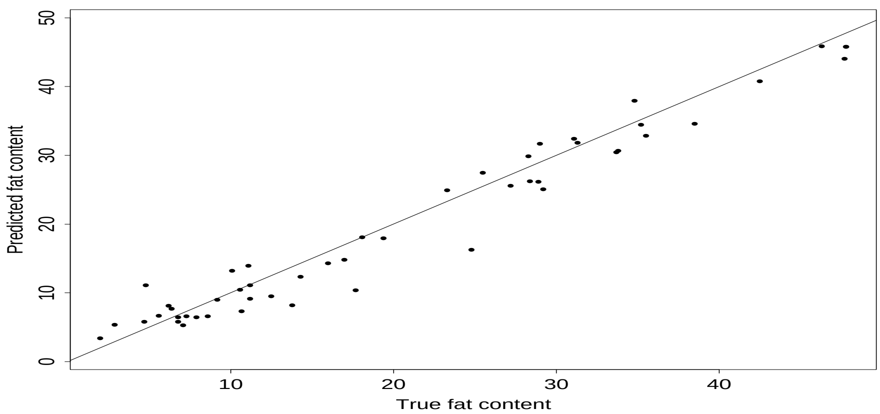

Figure 6: Spectrometric data: Predicted values on the testing sample

5.5 About implementation of the method

The implementation of the method was performed by using the *Splus* routine *funopare.kernel* which is included in the package *npfda*. This package will go with the monograph by Ferraty & Vieu (2006). This *Splus* package (as well as a similar *R* package) and the spectrometric dataset (as well as several other curves datasets) will be put in free access on line in the next future. By that time, programs and data are available on request.

6 Conclusions and open problems

This paper completes the recent advances existing in kernel nonparametric regression for functional data, by giving not only the rates of convergence of the estimates but also the exact expressions of the constant terms involved in these rates. These theoretical results deal with mean squared errors evaluations and asymptotic normality results. As explained in Section 5, these new results open interesting perspectives for applications, such as for instance data-driven automatic bandwidth selection and confidence band construction.

We support the idea that bootstrap methods offers interesting perspectives for the functional context. One of them is illustrated by our functional version of the wild bootstrap for selecting the smoothing parameter. We have observed nice results for bandwidth selection on some simulated and real datasets. It should be pointed out that theoretical support for this functional wild bootstrap bandwidth selection rule remains an open problem. Our guess is that it should be possible to extend to functional variables some results stated in finite dimension (for instance those in Härdle & Bowman, 1987), but this has still to be proved.

Another direct application of our result concerns the construction of confidence bands. Once again, the problem of estimating the constants involved in the asymptotic normal distribution can be attacked by the wild bootstrap track described before. One possible way for that would be to try to extend standard finite-dimensional knowledge (see for instance Härdle & Marron, 1991, or Härdle, Huet & Jolivet, 1995) to infinite-dimensional variables.

7 Appendix: Proofs

In the following, given some $\mathbb{R}$-valued random variable U, P^U will stand for the probability measure induced by U. To make its treatment easier, the kernel estimate $\hat{r}$ will be decomposed as follows:

$$\widehat{r}(\chi) = \frac{\widehat{g}(\chi)}{\widehat{f}(\chi)},$$

where

$$\widehat{g}(\chi) = \frac{1}{nF(h)} \sum_{k=1}^{n} Y_k K\left(\frac{\|\mathcal{X}_k - \chi\|}{h}\right)$$

and

$$\widehat{f}(\chi) = \frac{1}{nF(h)} \sum_{k=1}^{n} K\left(\frac{\|\mathcal{X}_k - \chi\|}{h}\right).$$

7.1 Proof of Theorem 1

The proof is split into two parts: computations of the bias and of the variance of the estimate. Each part is decomposed in technical lemmas that will be proved in Section 7.3.

- **Bias term: proof of (2)**. Let us write the following decomposition.

$$E\widehat{r}(\chi) = \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} + \frac{A_1}{\left(E\widehat{f}(\chi)\right)^2} + \frac{A_2}{\left(E\widehat{f}(\chi)\right)^2}, \tag{8}$$

with

$$A_1 = E\left[\widehat{g}(\chi)\left(\widehat{f}(\chi) - E\widehat{f}(\chi)\right)\right] \tag{9}$$

and

$$A_2 = E\left[\left(\widehat{f}(\chi) - E\widehat{f}(\chi)\right)^2 \widehat{r}(\chi)\right]. \tag{10}$$

The first step of the proof consists in rewritting the first term in right hand side of the decomposition (8) in the following way:

Lemma 1 *We have:*

$$\frac{E\widehat{g}\left(\chi\right)}{E\widehat{f}\left(\chi\right)} - r\left(\chi\right) = h\varphi'\left(0\right) I + o\left(h\right),$$

where

$$I = \frac{\int_0^1 tK\left(t\right) dP^{\|\mathcal{X}-\chi\|/h}\left(t\right)}{\int_0^1 K\left(t\right) dP^{\|\mathcal{X}-\chi\|/h}\left(t\right)}.$$

In a second attempt the next lemma will provide the constant term involved in this bias expression and its limit.

Lemma 2 *We have :*

$$I = \frac{K\left(1\right) - \int_0^1 \left(sK\left(s\right)\right)' \tau_h\left(s\right) ds}{K\left(1\right) - \int_0^1 K'\left(s\right) \tau_h\left(s\right) ds} \underset{n\to\infty}{\longrightarrow} \frac{M_0}{M_1}.$$

Finally, to finish this proof it suffices to prove that both last terms at right hand side of (8) are neglectible. This is done in next lemma.

Lemma 3 *We have:*

$$A_1 = O\left(\left(nF\left(h\right)\right)^{-1}\right) \text{ and } A_2 = O\left(\left(nF\left(h\right)\right)^{-1}\right).$$

So the proof of (2) is complete.

- **Variance term: proof of (3)**. The starting point of the proof is the following decomposition. This decomposition has been obtained in earlier work by Collomb (1976) (see also Sarda & Vieu (2000)) in the finite-dimensional case, but since the proof is only using analytic arguments about Taylor expansion of the function $1/z$ around 0, it extends obviously to our functional setting:

$$Var\left(\widehat{r}\left(\chi\right)\right) = \frac{Var\widehat{g}\left(\chi\right)}{\left(E\widehat{f}\left(\chi\right)\right)^{2}} - 4\frac{E\widehat{g}\left(\chi\right)Cov\left(\widehat{g}\left(\chi\right),\widehat{f}\left(\chi\right)\right)}{\left(E\widehat{f}\left(\chi\right)\right)^{3}} + 3Var\widehat{f}\left(\chi\right)\frac{\left(E\widehat{g}\left(\chi\right)\right)^{2}}{\left(E\widehat{f}\left(\chi\right)\right)^{4}} + o\left(\frac{1}{nF\left(h\right)}\right). \tag{11}$$

Finally, the result (3) will follow directly from this decomposition together with both following lemmas.

Lemma 4 *We have successively:*

$$E\widehat{f}\left(\chi\right) \rightarrow K\left(1\right) - \int_{0}^{1} K'\left(s\right)\tau_{0}\left(s\right)ds = M_{1},$$

$$E\widehat{g}\left(\chi\right) \rightarrow r\left(\chi\right)\left(K\left(1\right) - \int_{0}^{1} K'\left(s\right)\tau_{0}\left(s\right)ds\right) = r\left(\chi\right)M_{1}.$$

Lemma 5 *We have successively:*

$$\left(Var\widehat{f}\left(\chi\right)\right) = \frac{M_{2}}{nF\left(h\right)}\left(1+o\left(1\right)\right),$$

$$\left(Var\widehat{g}\left(\chi\right)\right) = \left(\sigma_{\varepsilon}^{2} + r^{2}\left(\chi\right)\right)\frac{M_{2}}{nF\left(h\right)}\left(1+o\left(1\right)\right),$$

$$Cov\left(\widehat{g}\left(\chi\right),\widehat{f}\left(\chi\right)\right) = r\left(\chi\right)\frac{M_{2}}{nF\left(h\right)}\left(1+o\left(1\right)\right).$$

7.2 Proof of Theorem 2

The following lemma states a preliminary pointwise limiting distribution result. This lemma stems from the bias and variance expressions obtained along Theorem 1; it will be proved in the next subsection.

Lemma 6 *We have:*

$$\sqrt{nF\left(h\right)}\left(\widehat{r}\left(\chi\right) - r\left(\chi\right) - B_{n}\right)\frac{M_{1}}{\sqrt{\sigma_{\varepsilon}^{2}M_{2}}} \hookrightarrow \mathcal{N}\left(0,1\right). \tag{12}$$

Because of standard Glivenko-Cantelli type results, we have

$$\frac{\widehat{F}(h)}{F(h)} \xrightarrow{\mathbb{P}} 1,$$

and this is enough, combined with the result of Lemma 6, to get the conclusion of Theorem 2.

7.3 Proofs of technical lemmas

- **Proof of Lemma 1:** To calculate

$$\frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} - r(\chi) = \frac{E\left[(Y - r(\chi))\, K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right]}{EK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)}, \tag{13}$$

note first that

$$\begin{aligned} E\left[(Y - r(\chi))\, K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right] &= E\left[(r(\mathcal{X}) - r(\chi))\, K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right] \\ &= E\left[\varphi(\|\mathcal{X}-\chi\|)\, K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right]. \end{aligned}$$

Moreover, it comes:

$$\begin{aligned} & E\left[\varphi(\|\mathcal{X}-\chi\|)\, K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right] \\ &= \int \varphi(t)\, K\left(\frac{t}{h}\right) dP^{\|\mathcal{X}-\chi\|}(t) \\ &= \int \varphi(ht)\, K(t)\, dP^{\|\mathcal{X}-\chi\|/h}(t) \\ &= h\varphi'(0) \int tK(t)\, dP^{\|\mathcal{X}-\chi\|/h}(t) + o(h), \end{aligned}$$

the last line coming from the first order Taylor's expansion for φ around 0.

For the denumerator in (13) we have

$$E\left[K\left(\|\mathcal{X}-\chi\|/h\right)\right]=\int K\left(t\right)dP^{\|\mathcal{X}-\chi\|/h}\left(t\right).$$

Finally, it appears clearly that the first order bias term is $h\varphi'\left(0\right)I$.

- **Proof of Lemma 2:** We note that

$$tK\left(t\right)=K\left(1\right)-\int_{t}^{1}\left(sK\left(s\right)\right)'ds.$$

Applying Fubini's Theorem we get

$$\begin{aligned}\int_{0}^{1}tK\left(t\right)dP^{\|\mathcal{X}-\chi\|/h}\left(t\right)&=K\left(1\right)F\left(h\right)-\int_{0}^{1}\left(\int_{t}^{1}\left(sK\left(s\right)\right)'ds\right)dP^{\|\mathcal{X}-\chi\|/h}\left(t\right)\\&=K\left(1\right)F\left(h\right)-\int_{0}^{1}\left(sK\left(s\right)\right)'F\left(hs\right)ds.\end{aligned}$$

Similarly, we have

$$\int_{0}^{1}K\left(t\right)dP^{\|\mathcal{X}-\chi\|/h}\left(t\right)=K\left(1\right)F\left(h\right)-\int_{0}^{1}K'\left(s\right)F\left(hs\right)ds. \tag{14}$$

So the proof of this lemma is finished by applying the Lebesgue's dominated convergence theorem since the denumerator may be easily bounded above by $K\left(1\right)>0$ (K being decreasing).

- **Proof of Lemma 4**: The first assertion follows directly from (14), while the second one can be proved similarly according to the following lines:

$$\begin{aligned}EYK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)&=E\left[E\left(Y|\mathcal{X}\right)K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right]\\&=\left(r\left(\chi\right)+o(1)\right)E\left[K\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right].\end{aligned}$$

- **Proof of Lemma 5**: We write the variance of $\widehat{f}\left(\chi\right)$ as:

$$\left(Var\widehat{f}(\chi)\right) = \frac{1}{nF^2(h)}\left[EK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) - \left(EK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right)^2\right],$$

and note that, as for getting (14), it holds:

$$EK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) = F(h)\left(K^2(1) - \int_0^1 \left(K^2\right)'(s)\,\tau_h(s)\,ds\right). \qquad (15)$$

The first assertion of lemma 4 gives

$$\left(EK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right)^2 = O\left(F^2(h)\right).$$

At last

$$\left(Var\widehat{f}(\chi)\right) \sim (n\,F(h))^{-1}\left(K^2(1) - \int_0^1 \left(K^2\right)'(s)\,\tau_0(s)\,ds\right), \qquad (16)$$

which finishes the proof of the first assertion of our lemma.

The same steps can be followed to prove the second assertion. We write

$$\left(Var\widehat{g}(\chi)\right) = \frac{1}{nF^2(h)}\left[EY^2K^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) - \left(EYK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right)^2\right].$$

The second term at right hand side of this expression is treated directly by using the second assertion of Lemma 4, while the first one is treated as follows by conditioning on $\mathcal{X}$:

$$EY^2K^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) \;=\; Er^2(\mathcal{X})\,K^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) + E\sigma_\varepsilon^2(\mathcal{X})K^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right).$$

The continuity of r^2 and of $\sigma_\varepsilon^2(.)$ insure that

$$Var\widehat{g}(\chi) = \frac{1}{nF^2(h)}\left(\sigma_\varepsilon^2 + r^2(\chi)\right)EK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)(1+o(1)).$$

Combining this result with (15) allows to finish the proof of the second assertion of our lemma. Let us deal now with the covariance term :

$$
\begin{aligned}
&Cov\left(\widehat{g}\,(\chi)\,,\widehat{f}\,(\chi)\right)\\
&= \frac{1}{nF^2\,(h)}\left[EYK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) - EK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)EYK\left(\frac{\|\mathcal{X}-\chi\|}{h}\right)\right].
\end{aligned}
$$

The last two terms were computed before, while the first one is treated by conditioning on $\mathcal{X}$ and using continuity of r:

$$
EYK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right) = (r\,(\chi)+o(1))\,EK^2\left(\frac{\|\mathcal{X}-\chi\|}{h}\right).
$$

The proof of this lemma is now finished.

- **Proof of Lemma 3**: Both assertions of this lemma are direct consequences of Lemmas 4 and 5.

- **Proof of Lemma 6:** On one hand, (8) and Lemma 3 allows us to get

$$
\widehat{r}(\chi) - E\widehat{r}(\chi) \;=\; \frac{\widehat{g}(\chi)}{\widehat{f}(\chi)} - \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} + o\left(\frac{1}{\sqrt{n\,F(h)}}\right).
$$

On the other hand, the following decomposition holds:

$$
\frac{\widehat{g}(\chi)}{\widehat{f}(\chi)} - \frac{E\widehat{g}(\chi)}{E\widehat{f}(\chi)} = \frac{(\widehat{g}(\chi)-E\widehat{g}(\chi))\,E\widehat{f}(\chi) + \left(E\widehat{f}(\chi)-\widehat{f}(\chi)\right)E\widehat{g}(\chi)}{\widehat{f}(\chi)\,E\widehat{f}(\chi)}.
$$

Using Slutsky's theorem and Theorem 1, we get

$$
\sqrt{\frac{n\,F(h)\,M_1^2}{\sigma_\varepsilon^2 M_2}}\,(\widehat{r}(\chi)-E\widehat{r}(\chi)) \hookrightarrow \mathcal{N}\,(0,1)\,,
$$

noting that $\widehat{r}(\chi) - E\widehat{r}(\chi)$ can be expressed as an array of independent centered random variables (and the Central Limit Theorem applies). Let us remark that

$$\widehat{r}(\chi) - E\widehat{r}(\chi) \;=\; \widehat{r}\,(\chi) - r\,(\chi) - B_n,$$

which achieves the proof of this lemma.

7.4 Proof of Propositions 1 and 2

- **Proof of Proposition 1-*i* and *ii***: Obvious.

- **Proof of Proposition 1-*iii***: We have:

$$\tau_h\,(s) = \exp\left(-\frac{C_2}{h^p}\left(\frac{1}{s^p} - 1\right)\right) s\gamma.$$

If $s = 1$ we have $\tau_h\,(1) = 1, \forall h$, while if $s = 0$ we have $\tau_h\,(1) = \tau_0\,(1) = F(0) = 0, \forall h$. To complete this proof it suffices to note that, for $s \in (0,1)$, we have: $1/s^p - 1 > 0$, and so we have $\tau_h\,(s) \to 0$ as $h \to 0$.

- **Proof of Propostion 1-*iv***: For any $s > 0$ and any $h < 1$, we have:

$$\tau_h\,(s) = \frac{|\ln h|}{|\ln h + \ln s|} = \frac{\ln h}{\ln h + \ln s} = \frac{1}{1 + \frac{\ln s}{\ln h}}$$

and so we have $\tau_0(s) = 1, \forall s > 0$. To complete this proof it suffices to note that $\tau_h\,(1) = \tau_0\,(1) = F(0) = 0, \forall h$.

- **Proof of Propostion 2-*i***: By simple integration by parts we arrive at

$$M_0 = \int_0^1 sK(s)T_0(s)ds,$$

where $T_0' = \tau_0$. Because of **H2** and because τ_0 is non decreasing, there exists some nonempty interval $[a,b] \subset (0,1)$ such that both K and T_0 do not vanish on $[a, b]$, and therefore we arrive at:

$$M_0 \geq \int_a^b sK(s)T_0(s)ds > 0.$$

- **Proof of Propostion 2-i:** Obvious.

Acknowledgements. The authors are grateful to all the participants of the working group STAPH (http://www.lsp.ups-tlse.fr/staph) on Functional Statistics at the University Paul Sabatier of Toulouse for their stimulating and helpful comments, with special thanks to Laurent Delsol for his cosnructie reading of a earlier version of this paper. This work has been significantly improved along the refereing procedure, and we would like to thank two anonymous referees, an Associate Editor and the Editor for their comments and suggestions.

References

[1] Collomb, G. (1976). *Estimation nonparametrique de la regression par la methode du noyau (in french).* PhD Univ. Paul Sabatier, Toulouse, France.

[2] Ferraty, F., Laksaci, A. & Vieu, P. (2005). Estimation of some characteristics of the conditional distribution in nonparametric functional models. *Statist.Inf. for Stoch. Process.* In print.

[3] Ferraty, F. & Vieu, P. (2003). Functional nonparametric statistics: a double infinite dimensional framework. In *Recent Advances and Trends in Nonparametric Statistics.* Ed. M. Akritas & D. Politis, 61-78. Elsevier.

[4] Ferraty, F. & Vieu, P. (2006). *Nonparametric functional data analysis.* Springer New-York, in print.

[5] Härdle, W. (1989). Resampling for inference from curves. Proceedings of the 47th Session of the ISI. *Bull. Inst. Internat. Statist.*, **53**, 3, 53-64.

[6] Härdle, W. & Bowman, A. (1987). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.*, **83**, 127-141.

[7] Härdle, W., Huet, S. & Jolivet, E. (1995). Better bootstrap confidence intervals for regression curve estimation. *Statistics*, **26**, 4, 287-306.

[8] Härdle, W. & Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. of Statist.*, **16**, 1696-1708.

[9] Li, W.V. & Shao, Q.M. (2001). Gaussian processes : inequalities, small ball probabilities and applications. In Handbook of Statistics **19**, 533-597, Elsevier Science.

[10] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. of Statist.*, **21**, 255-285.

[11] Mammen, E. (2000). Resampling methods for nonparametric regression. In *Smoothing and regression; approaches, computation and application. Ed. M. Schimek*, Wiley, 425-251.

[12] Masry, E. (2005). Nonparametric regression estimation for dependent fucnctional data: asymptotic normality. *Stochastic Process. Appl.*, **115**(1), 155-177.

[13] Mayer-Wolf, E. & Zeitouni, O. (1993). The probability of small gaussian ellipsoids and associated conditional moments. *Ann. Probab.* **21**, 1, 14-24.

[14] Ramsay, J.O. & Silverman, B.W. (2005). *Functional data analysis. 2nd Edition.* Springer Series in Statistics.

[15] Sarda, P. & Vieu, P. (2000). Kernel regression. In *Smoothing and regression; approaches, computation and application. Ed. M. Schimek*, Wiley, 43-70.

[16] Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math Statist.*, **43**, 139-149.