

Kernel regression with functional response

Frédéric Ferraty

Institut de Mathématiques de Toulouse, France
e-mail: ferraty@math.univ-toulouse.fr

Ali Laksaci

Univ. Djillali Liabès, Sidi Bel Abbès, Algérie
e-mail: alilak@yahoo.fr

Amel Tadj*

Département de Mathématiques
Univ. Djillali Liabès
BP 89, 22000 Sidi Bel Abbès, Algérie
e-mail: ameltdz@yahoo.fr

and

Philippe Vieu

Institut de Mathématiques de Toulouse, France
e-mail: vieu@math.univ-toulouse.fr

Abstract: We consider kernel regression estimate when both the response variable and the explanatory one are functional. The rates of uniform almost complete convergence are stated as function of the small ball probability of the predictor and as function of the entropy of the set on which uniformity is obtained.

AMS 2000 subject classifications: Primary 62G08.

Keywords and phrases: Uniform almost complete convergence, kernel estimators, functional data, entropy, semi-metric space.

Received October 2010.

Contents

1	Introduction	160
2	Some probability tools for functional variables	161
3	The functional regression with functional response	161
3.1	The doubly functional nonparametric setting	161
3.2	The general hypotheses	162
3.3	Uniform rates of convergence	163

*Corresponding author.

A Proof of technical lemmas	165
Acknowledgements	170
References	170

1. Introduction

Regression model is a main tool to examine the relationship between a response variable and an explanatory one. We are interested in estimating the nonparametric regression when both variables (response and explanatory) are functional. The study of statistical models adapted to such kind of data received a lot of attention in recent literature (see, [14], [2] and [7] for recent monographies and [5] for an handbook on statistics in infinite dimensional spaces). Since the previous paper by [6] the literature on nonparametric regression (see the survey by [8]) starts to be rather important when the response variable is scalar, but there are very few advances in this direction when this response is functional (see however [3] and [11] for earlier references). Various advances on this field have been already provided under some linear assumption on the regression operator (see for instance [13], [9], in standard i.i.d. case; see also [2] in the specific time series context).

This work presents some asymptotic property for a kernel-type regression estimator when both response and explanatory variables are functional. Precisely we state the uniform almost complete convergence rate of this doubly functional kernel estimate. As far as we know, our result is the first one stating uniform asymptotic results in nonparametric doubly functional regression problems. As usually in functional statistics, the topological structures on the infinite-dimensional spaces play a prominent role, and we present the rates of convergence in such a way to highlight these topological effects. Firstly, the problems linked with the high (high because infinite) dimensionality of the explanatory variable are dealt with by means of small ball probability considerations (and this is directly linked with the topological structure). Secondly, the uniformity of the convergence is obtained by means of entropy notions (which are, once again, direct topological considerations). Finally, the type of the Banach space on which the response variable takes its values acts also directly on the rates of convergence.

Section 2 is dedicated to some probability tools for functional variable valued in a Banach space. The doubly functional model and its estimate are presented in Section 3 and the uniform rates of convergence are stated therein. Some technical proofs are deferred to the Appendix.

Before closing this introduction it is worth being stressed that, even if we have deliberately chosen to present a short theoretical paper, there exists a wide scope of applied scientific fields for which our approach could be of interest. In the next future, our works will be concentrated on the implementation of this doubly functional nonparametric method. To fix the ideas, one can for instance find examples in Biometrics, Genetics or Environmetrics in [9], [13] and [10] respectively.

2. Some probability tools for functional variables

We present two general tools for random variables valued in Banach spaces. The topological complexity of the Banach space \mathcal{B} will appear through the following following notion.

Definition 1. A Banach space \mathcal{B} is of type $p \in]1, 2]$ if there exists a strictly positive constant c such that for any finite sequence Z_1, \dots, Z_n of independent \mathcal{B} -random variables such that $E\|Z_i\|^p < \infty$ and $EZ_i = 0$, we have

$$E\left\|\sum_{i=1}^n Z_i\right\|^p \leq c \sum_{i=1}^n E\|Z_i\|^p.$$

Remark 1. Clearly, \mathbb{R}^k ($k \geq 1$) or more generally any Hilbert space is a Banach space of type 2.

In addition, and as it is usual in nonparametric statistics, one needs some kind of exponential inequality for getting rates of convergence. We will make use later in this paper of the following Bernstein's type inequality for sums of Banach-valued random elements:

Lemma 2.1. Let Z_1, \dots, Z_n be independent \mathcal{B} -random variables. If we have

$$\sum_{i=1}^n E\|Z_i\|^k \leq \frac{k!}{2} l^2 b^{k-2}, \forall k \geq 2,$$

for some positive constants $b = b(n)$ and $l = l(n)$, then:

$$\forall t > 0, P\left(\left|\|S_n\| - E\|S_n\|\right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2l^2 + 2bt}\right),$$

where $S_n = \sum_{i=1}^n Z_i$.

This inequality can be found at page 49 in [2], and a deeper discussion on \mathcal{B} -random variables can also be found in this book.

3. The functional regression with functional response

3.1. The doubly functional nonparametric setting

Let us consider a sample of independent pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathcal{B}$, where (\mathcal{F}, d) is a semi-metric space and $(\mathcal{B}, \|\cdot\|)$ is a Banach space of type $p \in]1, 2]$. Recall that a semi-metric (sometimes called pseudo-metric) is just a metric violating the property: $[d(x, y) = 0] \Rightarrow [x = y]$. The functional regression operator is defined by

$$m(x) = E[Y | X = x], \quad \forall x \in \mathcal{F}. \quad (3.1)$$

Nonparametric estimates of the operator m are constructed by local weighting ideas, as for instance the following doubly functional kernel estimate:

$$\widehat{m}(x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))Y_i}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}, \quad \forall x \in \mathcal{F},$$

where K is a kernel function and $h = h_n$ is a sequence of positive real numbers which goes to zero as n goes to infinity. We stay here with this simple nonparametric smoother, but alternative ones could be introduced such as the local functional one previously studied by [1] for scalar Y .

3.2. The general hypotheses

Now $S_{\mathcal{F}}$ is a fixed subset of \mathcal{F} , and for $\eta > 0$ we consider the following η -neighborhood of $S_{\mathcal{F}}$:

$$S_{\mathcal{F}}^{\eta} = \{x \in \mathcal{F}, \exists x' \in S_{\mathcal{F}}, d(x, x') \leq \eta\}.$$

We will use the notation $B(x, h) = \{x' \in \mathcal{F}, d(x', x) \leq h\}$ for the closed ball in the space \mathcal{F} . The model consists in assuming that the probability distribution of X is such that there exists a non-decreasing function ϕ such that:

$$(H1) \quad \exists(C_1, C_2), \forall x \in S_{\mathcal{F}}, \forall \epsilon > 0, 0 < C_1\phi(\epsilon) \leq P(X \in B(x, \epsilon)) \leq C_2\phi(\epsilon) < \infty,$$

while the joint distribution of (X, Y) has to satisfy:

$$(H2) \quad \exists C_3 < \infty, \exists b > 0, \exists \eta > 0, \forall x, x' \in S_{\mathcal{F}}^{\eta}, \|m(x) - m(x')\| \leq C_3 d^b(x, x'),$$

$$(H3) \quad \exists C_4, \forall r \geq 1, E(\|Y\|^r | X) < C_4 r! < \infty, \text{ a.s.},$$

where $r! = r(r-1)\dots(r-[r]+1)$, $[r]$ being the largest integer smaller than r . We also need the following technical conditions on the kernel K :

(H4) The kernel function has to be such that:

(i) K is a bounded and Lipschitz continuous function with support $[0, 1)$, and if $K(1) = 0$ it has also to fulfill, jointly with $\phi(\cdot)$, the conditions:

$$(ii) \quad \exists(C_5, C_6), \quad -\infty < C_5 < K'(t) < C_6 < 0;$$

$$(iii) \quad \exists C_7 > 0, \exists \eta_0 > 0, \forall \eta < \eta_0, \int_0^{\eta} \phi(u) du > C_7 \eta \phi(\eta).$$

While (H1)–(H4) are standard conditions to get pointwise rates of convergence, the next assumptions are directly linked with our wishes to have uniform rates over the set $S_{\mathcal{F}}$. These conditions will make appear the topological complexity of the set $S_{\mathcal{F}}$ which will act through the Kolmogorov's entropy of $S_{\mathcal{F}}$ defined for any $\epsilon > 0$ by:

$$\psi_{S_{\mathcal{F}}}(\epsilon) = \log(N_{\epsilon}(S_{\mathcal{F}}))$$

where $N_{\epsilon}(S_{\mathcal{F}})$ is the minimal number of open balls in \mathcal{F} of radius ϵ which are necessary to cover $S_{\mathcal{F}}$. Of course these conditions will also cross restrictions on the small ball probability function ϕ introduced in (H1). Assume that:

$$(H5) \exists C_8 > 0, \exists \eta_0 > 0, \forall \eta < \eta_0, 0 \leq \phi'(\eta) < C_8,$$

$$(H6) \exists n_0, \forall n > n_0, \frac{(\log n)^2}{n\phi(h)} < \psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right) < \frac{n\phi(h)}{\log n},$$

$$(H7) \exists \beta > 1, \sum_{n=1}^{\infty} \exp\left\{(1-\beta)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty.$$

It is important to stress that, despite of its rather intricate form, this set of assumptions is not too much restrictive. Excepted (H3), all other conditions are related with the explanatory variable X and they have been discussed in various previous papers. The reader can look for instance at Chapter 13 in [7] to see how all these conditions can be shown to be true pending to suitable topological structure on the space \mathcal{F} (that is, pending to suitable choice of the semi-metric d). It is out of purpose to provide detailed discussion in this paper, because this is definitively not linked with the functional nature of the response (which is the main point that we wish to address) and also because such a discussion appears already in various other papers. Finally, the only condition which is specific to the functional response Y is the rather unrestrictive conditional moment existency assumed in (H3).

3.3. Uniform rates of convergence

The following theorem states the rate of convergence of \widehat{m} , uniformly over the set $S_{\mathcal{F}}$. The asymptotics are stated in terms of almost complete convergence (denoted by *a.co.*) which is known to imply both weak and strong convergences (see, among other, Section A-1 in [7]). The topological structure on the space \mathcal{F} acts directly on these rates through the functions ϕ and $\psi_{S_{\mathcal{F}}}$, while the topological complexity of the space \mathcal{B} will act through its type p . As discussed before, this is the first result of this kind in regression setting when both X and Y are functional. In the special simpler case when Y is real this result was given in [4]. To fix the ideas and to highlight the wide generality of the apparently highly technical assumptions (H1)–(H7) a few special cases will be considered later in Remark 2, while Remark 3 will present an interesting direct consequence of this general result.

Theorem 3.1. *Under the hypotheses (H1)–(H7), we have*

$$\begin{aligned} \sup_{x \in S_{\mathcal{F}}} \|\widehat{m}(x) - m(x)\| &= O(h^b) + O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}} \right) \\ &+ O_{a.co.} \left(\frac{1}{n\phi(h)} \right)^{1-\frac{1}{p}}. \end{aligned} \quad (3.2)$$

Proof. We consider the decomposition

$$\widehat{m}(x) - m(x) = \frac{(\widehat{g}(x) - E\widehat{g}(x))}{\widehat{f}(x)} + \frac{(E\widehat{g}(x) - m(x))}{\widehat{f}(x)} + \frac{(1 - \widehat{f}(x))m(x)}{\widehat{f}(x)} \quad (3.3)$$

where

$$\widehat{f}(x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))}{nE[K(h^{-1}d(x, X_1))]} \quad \text{and} \quad \widehat{g}(x) = \frac{\sum_{i=1}^n K(h^{-1}d(x, X_i))Y_i}{nE[K(h^{-1}d(x, X_1))]}.$$

The denominator $\widehat{f}(x)$ does not involve the functional response, and therefore the following results stated in [4] remain true:

$$\sup_{x \in S_{\mathcal{F}}} |\widehat{f}(x) - 1| = O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi(h)}} \right) \quad \text{and} \quad \sum_{n=1}^{\infty} P \left(\inf_{x \in S_{\mathcal{F}}} \widehat{f}(x) < \frac{1}{2} \right) < \infty.$$

So finally, Theorem 3.1 will be true as long as both following lemmas will be proved. The proofs of these lemmas are reported to the Appendix. \square

Lemma 3.2. *Under the hypotheses (H1), (H2) and (H4)–(H7), we have*

$$\sup_{x \in S_{\mathcal{F}}} \|E\widehat{g}(x) - m(x)\| = O(h^b).$$

Lemma 3.3. *Under the assumptions of Theorem 3.1, we have*

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{g}(x) - E\widehat{g}(x)\| = O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi(h)}} \right) + O_{a.co.} \left(\frac{1}{n\phi(h)} \right)^{1-1/p}.$$

Remark 2. *For any Hilbert space, it is clear that $p = 2$ (see, Remark 1) so the rate (3.2) becomes:*

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{m}(x) - m(x)\| = O(h^b) + O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi(h)}} \right).$$

In the special case when \mathcal{B} is the euclidian space \mathbb{R}^k , when $S_{\mathcal{F}}$ is compact and when X has a density with respect to the Lebesgue measure, then (3.2) becomes the usual multivariate nonparametric rate:

$$\sup_{x \in S_{\mathcal{F}}} \|\widehat{m}(x) - m(x)\| = O(h^b) + O_{a.co.} \left(\sqrt{\frac{\log n}{nh^k}} \right).$$

Remark 3. *Uniform consistency allows to replace a fixed x with a random element X . Indeed, as soon as $P(X \in S_{\mathcal{F}}) = 1$, one gets*

$$\|\widehat{m}(X) - m(X)\| \leq \sup_{x \in S_{\mathcal{F}}} \|\widehat{m}(x) - m(x)\|.$$

So, results on $\sup_{x \in S_{\mathcal{F}}} \|\widehat{m}(x) - m(x)\|$ remain valid for $\|\widehat{m}(X) - m(X)\|$. This can be of great practical interest for other problems (automatic bandwidth choice, semi-parametric modelling, ...) which are out of the scope here.

Appendix A: Proof of technical lemmas

In the following, we will denote $K_i(x) = K(h^{-1}d(x, X_i))$. First of all, one has

$$\exists 0 < C_9 \leq C_{10} < \infty, \forall x \in S_{\mathcal{F}}, C_9\phi(h) < E[K_1(x)] < C_{10}\phi(h). \quad (\text{A.1})$$

The result (A.1) is obvious when $K(1) > 0$ and can be extended to continuous kernel K satisfying (H4) as shown in Lemma 4.4, page 44, in [7]. From now on, we will denote by C is a generic nonnegative real constant, and we will take

$$\epsilon = \frac{\log n}{n}.$$

Note that condition (H5) implies that for n large enough:

$$0 \leq \phi(h) \leq Ch,$$

in such a way that (H6) implies both that

$$\lim_{n \rightarrow \infty} \frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)} = 0 \text{ and } \lim_{n \rightarrow \infty} \frac{\log n}{nh} = 0. \quad (\text{A.2})$$

Proof of Lemma 3.2. One has

$$\begin{aligned} \|E\hat{g}(x) - m(x)\| &= \frac{1}{E[K_1(x)]} \left\| E \frac{1}{n} \left[\sum_{i=1}^n K_i(x) Y_i \right] - m(x) E[K_1(x)] \right\| \\ &= \frac{1}{E[K_1(x)]} \|E[K_1(x) Y_1] - m(x) E[K_1(x)]\|. \end{aligned}$$

Hence, we get

$$\forall x \in S_{\mathcal{F}}, \quad \|E\hat{g}(x) - m(x)\| = \frac{1}{E[K_1(x)]} [EK_1(x)(\|m(X_1) - m(x)\|)].$$

Thus, with hypotheses (H1), (H2) and (A.1) we have, for n large enough:

$$\forall x \in S_{\mathcal{F}}, \quad \|E\hat{g}(x) - m(x)\| \leq C \frac{1}{E[K_1(x)]} [EK_1(x) \mathbb{1}_{B(x,h)}(X_1) d^b(X_1, x)] \leq Ch^b.$$

This last inequality yields the proof, since C does not depend on x . \square

Proof of Lemma 3.3. Let $x_1, \dots, x_{N_\epsilon(S_{\mathcal{F}})}$ be such that the following covering of $S_{\mathcal{F}}$ holds:

$$S_{\mathcal{F}} \subset \cup_{k=1}^{N_\epsilon(S_{\mathcal{F}})} B(x_k, \epsilon),$$

and define

$$\forall x \in S_{\mathcal{F}}, k(x) = \arg \min_{k \in \{1, 2, \dots, N_\epsilon(S_{\mathcal{F}})\}} d(x, x_k).$$

One considers now the following decomposition

$$\begin{aligned} \sup_{x \in \mathcal{S}_{\mathcal{F}}} \|\widehat{g}(x) - E\widehat{g}(x)\| &\leq \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \|\widehat{g}(x) - \widehat{g}(x_{k(x)})\|}_{G_1} + \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \|\widehat{g}(x_{k(x)}) - E\widehat{g}(x_{k(x)})\|}_{G_2} \\ &+ \underbrace{\sup_{x \in \mathcal{S}_{\mathcal{F}}} \|E\widehat{g}(x_{k(x)}) - E\widehat{g}(x)\|}_{G_3}. \end{aligned}$$

We will first deal with the terms G_1 and G_3 which are the simplest ones in the sense that they are not linked with the functional nature of Y and so one can make use of previous literature for scalar response Y to treat them. The term G_2 will need more specific attention.

i) *Study of the term G_1 .* We get directly from (H1) and (A.1):

$$\begin{aligned} G_1 &= \sup_{x \in \mathcal{S}_{\mathcal{F}}} \left\| \sum_{i=1}^n \left(\frac{1}{nE[K_1(x)]} K_i(x) Y_i - \frac{1}{nE[K_1(x_{k(x)})]} K_i(x_{k(x)}) Y_i \right) \right\| \\ &\leq \sup_{x \in \mathcal{S}_{\mathcal{F}}} \frac{1}{n\phi(h)} \sum_{i=1}^n \|Y_i\| \|K_i(x) - K_i(x_{k(x)})\| \mathbb{1}_{B(x,h) \cup B(x_{k(x)},h)}(X_i). \end{aligned}$$

In a first attempt, assume that $K(1) = 0$ (i.e. K is Lipschitz on $[0, 1]$) in order to get:

$$G_1 \leq \frac{C}{n} \sum_{i=1}^n Z_i \text{ with } Z_i = \frac{\epsilon \|Y_i\|}{h\phi(h)} \mathbb{1}_{B(x,h) \cup B(x_{k(x)},h)}(X_i).$$

Clearly, we get from (H3):

$$E[\|Y\|^m] = E[E[\|Y\|^m|X]] < Cm! < \infty,$$

which implies that

$$E(|Z_1|^m) \leq \frac{Cm!\epsilon^m}{h^m \phi(h)^{m-1}}. \quad (\text{A.3})$$

Moreover, by using the second result in (A.2) together with the definition of ϵ we have for n large enough:

$$\frac{\epsilon}{h} \leq C.$$

Both last results yield directly to

$$E(|Z_1|^m) \leq \frac{Cm!\epsilon^{m-1}}{h^{m-1} \phi(h)^{m-1}}.$$

So, by applying Corollary A.8 in [7] with $a^2 = \frac{\epsilon}{h\phi(h)}$, one gets

$$\frac{1}{n} \sum_{i=1}^n Z_i = EZ_1 + O_{a.co.} \left(\sqrt{\frac{\epsilon \log n}{nh\phi(h)}} \right).$$

Finally, applying again (A.3) for $m = 1$ one gets

$$G_1 = O_{a.co.} \left(\sqrt{\frac{\epsilon \log n}{n h \phi(h)}} \right) + O\left(\frac{\epsilon}{h}\right).$$

Now, using (H6) together with the second part of (A.2) and with the definition of ϵ , we get:

$$G_1 = O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n \phi(h)}} \right). \quad (\text{A.4})$$

The proof of (A.4) for the case $K(1) > C > 0$ (i.e. K Lipschitz on $[0, 1)$) is not presented here since one can proceed exactly as in Lemma 6 in [4] by splitting again G_1 into three terms.

ii) *Study of the term G_3 .* By definition of G_3 we have:

$$\begin{aligned} G_3 &\leq \sup_{x \in S_{\mathcal{F}}} E \left[\|\hat{g}(x_{k(x)}) - \hat{g}(x)\| \right] \\ &\leq E \left[\sup_{x \in S_{\mathcal{F}}} \|\hat{g}(x_{k(x)}) - \hat{g}(x)\| \right], \end{aligned}$$

the first inequality coming from the contractive property of the expectation operator (see [2], page 29). So we have finally $G_3 \leq EG_1$ which, combined with (A.4), leads directly to

$$G_3 = O \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n \phi(h)}} \right). \quad (\text{A.5})$$

iii) *Study of the term G_2 .* This part is the most technical because it involves directly the functional response Y . This is the main specificity of our work and so G_2 cannot be treated by the same techniques as if Y was real (as for G_1 and G_3). The proof will use the exponential inequality for Banach space valued random variables (see, Lemma 2.1). Let:

$$\Gamma_{ki} = \frac{1}{nE[K_1(x_{k(x)})]} [K_i(x_{k(x)})Y_i - E[K_i(x_{k(x)})Y_i]],$$

and

$$W_n = \left\| \sum_i^n \Gamma_{ki} \right\| - E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right).$$

It is clear that, $\forall \eta > 0$,

$$\begin{aligned} P(G_2 > \eta) &= P \left(\max_{k \in \{1, \dots, N_\epsilon(S_{\mathcal{F}})\}} \|\hat{g}(x_k) - E\hat{g}(x_k)\| > \eta \right) \\ &\leq N_\epsilon(S_{\mathcal{F}}) \max_{k \in \{1, \dots, N_\epsilon(S_{\mathcal{F}})\}} P(\|\hat{g}(x_k) - E\hat{g}(x_k)\| > \eta) \end{aligned}$$

$$\begin{aligned} &\leq N_\epsilon(S_{\mathcal{F}}) \max_{k \in \{1, \dots, N_\epsilon(S_{\mathcal{F}})\}} P \left(\left\| \sum_i^n \Gamma_{ki} \right\| > \eta \right) \\ &\leq N_\epsilon(S_{\mathcal{F}}) \max_{k \in \{1, \dots, N_\epsilon(S_{\mathcal{F}})\}} P \left(|W_n| > \eta - E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right) \right). \end{aligned}$$

Choosing now

$$\eta = \eta_0 \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} + E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right), \quad (\text{A.6})$$

we have

$$P(G_2 > \eta) \leq N_\epsilon(S_{\mathcal{F}}) \max_{k \in \{1, \dots, N_\epsilon(S_{\mathcal{F}})\}} P \left(|W_n| > \eta_0 \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} \right). \quad (\text{A.7})$$

To apply the inequality of Lemma 2.1, one must evaluate the quantities

$$\sum_{i=1}^n E \|\Gamma_{ki}\|^m \quad \text{and} \quad E \left(\left\| \sum_{i=1}^n \Gamma_{ki} \right\| \right). \quad (\text{A.8})$$

Using the condition (H3) we have for all $j \leq m$:

$$\begin{aligned} E \|Y_1 K_1(x)\|^j &= E \left[K_1^j(x) E \left[\|Y_1\|^j \mid X_1 \right] \right] \\ &\leq C j! E[K_1^j(x)], \end{aligned}$$

leading finally, by using the result (A.1) and the boundedness of K (see (H4)), to

$$E \|Y_1 K_1(x)\|^j \leq C j! \phi(h). \quad (\text{A.9})$$

Now we use the Newton's binomial expansion and we get:

$$\begin{aligned} \|\Gamma_{ki}\|^m &\leq \frac{1}{n^m (E[K_1(x)])^m} (\|K_i(x)Y_i\| + \|E[K_i(x)Y_i]\|)^m \\ &= \frac{1}{n^m (E[K_1(x)])^m} \sum_{k=0}^m C_{k,m} \|Y_1 K_1(x)\|^k \|E[Y_1 K_1(x)]\|^{m-k}, \end{aligned}$$

where $C_{k,m} = \frac{m!}{k!(m-k)!}$. We get by (A.9):

$$\begin{aligned} E \|\Gamma_{ki}\|^m &\leq \frac{1}{n^m (E[K_1(x)])^m} \sum_{k=0}^m C_{k,m} E \|Y_1 K_1(x)\|^k \|E[Y_1 K_1(x)]\|^{m-k} \\ &\leq C n^{-m} m! \phi(h)^{-m+1}. \end{aligned}$$

Note that we have used the fact that

$$\forall j \leq m, \quad C\phi(h) \leq E[K_1^j(x)].$$

which is again a consequence of the boundedness of K (see (H4)) and of the result (A.1). It follows that for all $m \geq 1$,

$$\sum_{i=1}^n E \|\Gamma_{ki}\|^m \leq Cm!(n\phi(h))^{-m+1}. \quad (\text{A.10})$$

On the other hand, we use the C_r inequality (see, [12], p.155) for the second quantity of (A.8) and we have

$$E\|Y_1 K_1(x) - E[Y_1 K_1(x)]\|^p \leq 2^{p-1} [E\|Y_1 K_1(x)\|^p + \|E[Y_1 K_1(x)]\|^p].$$

Noting that (A.9) can be obtained by the same route for non necessary integer power (that is changing j into p) one gets

$$n^p \phi^{p-1}(h) E\|\Gamma_{k1}\|^p \leq C.$$

Now, we use that \mathcal{B} is a space of type $p \in]1, 2]$ and we have

$$\left(E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right) \right)^p \leq E \left(\left\| \sum_i^n \Gamma_{ki} \right\|^p \right) \leq C \sum_{i=1}^n E\|\Gamma_{ki}\|^p,$$

and hence

$$E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right) = O((n\phi(h))^{-1+1/p}). \quad (\text{A.11})$$

Because of (A.10), we are now in position for applying Lemma 2.1, by taking

$$l^2 = b = \frac{C}{n\phi(h)} \text{ and } t = \eta_0 \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}},$$

and we arrive at

$$\begin{aligned} P \left(\left| \left\| \sum_i^n \Gamma_{ki} \right\| - E \left(\left\| \sum_i^n \Gamma_{ki} \right\| \right) \right| > \eta_0 \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} \right) &\leq 2 \exp \left(-\frac{t^2}{2l^2 + 2bt} \right) \\ &\leq 2 \exp \left(-C \frac{t^2}{2l^2} \right) \\ &\leq 2N_{\epsilon}(S_{\mathcal{F}})^{-C\eta_0^2}, \end{aligned}$$

the second inequality coming from the first result in (A.2). Therefore, by using (A.11), (A.7) and (A.6), we have:

$$P \left(G_2 > \eta_0 \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} + \left(\frac{1}{n\phi(h)} \right)^{1-1/p} \right) \leq CN_{\epsilon}(S_{\mathcal{F}})^{1-\beta}.$$

Because of (H7) there exists some $\beta > 1$ such that

$$\sum_{n=1}^{\infty} N_{\epsilon}(S_{\mathcal{F}})^{1-\beta} < \infty,$$

and we obtain that

$$G_2 = O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\epsilon)}{n\phi(h)}} \right) + O_{a.co.} \left(\frac{1}{n\phi(h)} \right)^{1-1/p}. \quad (\text{A.12})$$

Now, Lemma 3.3 can be easily deduced from (A.4), (A.5) and (A.12). \square

Acknowledgements

The authors thank the participants in the working group STAPH in Toulouse (<http://www.math.univ-toulouse.fr/staph>) for continuous support. Special thanks go also to the Editor, an Associate Editor and two referees for the speed and the quality of the reviewing of a previous draft of our work. The strong relevance of their comments had a real impact on the final quality of this paper.

References

- [1] BARRIENTOS-MARIN, J., FERRATY, F. and VIEU, P. (2010). Locally modelled regression and functional data. *J. Nonparametr. Stat.*, **22** 617–632. [MR2682211](#)
- [2] BOSQ, D. (2000). *Linear processes in function spaces. Theory and Application. Lectures Notes in Statistics*, Vol 149. Springer-Verlag. [MR1783138](#)
- [3] BOSQ, D. and DELECROIX, M. (1985). Nonparametric prediction of a Hilbert-space valued random variable. *Stochastic Process. Appl.*, **19** 271–280. [MR0787585](#)
- [4] FERRATY, F., LAKSACI, A., TADJ, A. and VIEU, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140** 335–352. [MR2558367](#)
- [5] FERRATY, F. and ROMAIN, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.
- [6] FERRATY, F. and VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*, **17** 545–564. [MR1952697](#)
- [7] FERRATY, F. and VIEU, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer-Verlag. [MR2229687](#)
- [8] FERRATY, F. and VIEU, P. (2011). Kernel regression estimation for functional data. In *The Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain)*. Oxford University Press.
- [9] HE, G., MÜLLER, H.G., WANG, J.L. and YANG, W. (2009). Functional Linear Regression Via Canonical Analysis. *Bernoulli*, **16** 705–729. [MR2730645](#)
- [10] HLUBINKA, D. and PRCHAL, L. (2007). Changes in atmospheric radiation from the statistical point of view. *Comput. Statist. Data Anal.*, **51**(10) 4926–4941. [MR2364550](#)

- [11] LECOUTRE, J.P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.*, **10** 145–149. [MR1072501](#)
- [12] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand Princeton. [MR0203748](#)
- [13] MÜLLER, H.G., CHIOU, J.M. and LENG, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC Bioinformatics*, **9**, 60.
- [14] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New-York. [MR2168993](#)