REVSTAT – Statistical Journal
Volume 18, Number 1, January 2020, 1–26

NONPARAMETRIC REGRESSION BASED ON DISCRETELY SAMPLED CURVES

Authors: Liliana Forzani
– Facultad de Ingeniería Química, UNL and researcher of CONICET, Argentina
liliana.forzani@gmail.com

Ricardo Fraiman
– Centro de Matemática, Facultad de Ciencias (UdelaR), Uruguay
fraimanricardo@gmail.com

Pamela Llop
– Facultad de Ingeniería Química, UNL and researcher of CONICET, Argentina
lloppamela@gmail.com

Received: January 2017 Revised: August 2017 Accepted: September 2017

Abstract:

- In the context of nonparametric regression, we study conditions under which the consistency (and rates of convergence) of estimators built from discretely sampled curves can be derived from the consistency of estimators based on the unobserved whole trajectories. As a consequence, we derive asymptotic results for most of the regularization techniques used in functional data analysis, including smoothing and basis representation.

Key-Words:

- *nonparametric regression; functional data; discrete curves.*

AMS Subject Classification:

- 62G08, 62M99.

1. INTRODUCTION

Technological progress in collecting and storing data provides datasets recorded at finite grids of points that become denser and denser over time. Although in practice data always comes in the form of finite dimensional vectors, from the theoretical point of view, the classic multivariate techniques are not well suited to deal with data which, essentially, is infinite dimensional and whose observations within the same curve are highly correlated.

From a practical point of view, a commonly used technique to treat this kind of data is to transform the (observed) discrete values into a function via smoothing or a series approximations (see [5], [21], [24, 25, 26], or chapter 9 of [13] and the references therein). For the analysis, we can use the intrinsic infinite dimensional nature of the data and assume the existence of continuous underlying stochastic processes which are observed ideally at every point. In this context, the theoretical analysis is performed on the functional space where they take values (see [15]). In what follows, we will refer to this last setting as the *full model*.

Nonparametric regression is an important tool in functional data analysis (FDA) which has received considerable attention from different authors in both settings. For the full model, consistency results have been obtained by, among others, [1], [3], [4], [7], [10], [15], [22], and [23]. In particular, [16] (see also the Corrigendum [17]) prove a consistency result close to universality for the kernel (with random bandwidth) estimator. The first contribution of the present paper will be to prove the consistency of the k-nearest neighbor with kernel regression estimator (Proposition 2.2) when the full trajectories are observed. This family, considered by [12], combines the smoothness properties of the kernel function with the locality properties of the k-nearest neighbors distances.

Regarding regression when discretized curves are available, [19] study the mean square consistency of the kernel estimator when the sample size as well as the grid size discretization go to infinity. More precisely, from independent realizations of a random process with continuous covariance structure, they estimate the regression function, assuming its smoothness. Under the same assumptions, but using interpolation of the data, [27], in a mainly practical approach, propose a method to estimate the regression function via smoothing splines (see also [20]). More recently, [8] establish minimax rates of convergence of estimators of the mean based on discretized sampled data while [9] establish the minimax rates of convergence for the covariance operator when data are observed on a lattice (see also [18] for the problem of principal components analysis for longitudinal data). In this context it is natural to assess the relation between the *ideal* nonparametric regression estimator constructed with the entire set of curves and the one computed with the discretized sample. In this direction, we are interested in addressing the following question:

- Under what conditions can the consistency (and rates of convergence) of the estimate computed with the discretized trajectories be derived from the consistency of the estimate based on the full curves?

Clearly, the asymptotic results for estimates computed with the discretized sample will not be a direct consequence of those for the full model. However, we provide reasonable conditions in order to still get the consistency and find rates of convergence of the estimator.

In this context we state the results for the well known kernel and k-nearest neighbor with kernel estimators. These results are a consequence of a more general result, which, besides discretization, also includes the cases of regularization via smoothing and basis representation.

This paper is organized as follows: In Section 2 we state the consistency of the k-nearest neighbor with kernel estimator in the infinite dimensional setting (for the full model). This result is not only interesting by itself but also, it will be used to prove consistency results when discretely sample data are available. In Section 3 we provide conditions for the consistency of the kernel and k-nearest neighbor with kernel estimators when we do not observe the whole trajectories but only a function of them (Theorems 3.1 and 3.2). In Section 4 the results for discretization, smoothing and basis representation are obtained as a consequence of Theorems 3.1 and 3.2. Finally, in Section 5 we perform a small simulation study where we compare the behaviour of the estimators computed with the discretized trajectories and with the full curves. Proofs are given in Appendices A and B.

2. CONSISTENCY RESULTS FOR FULLY OBSERVED CURVES

In this section we provide two L^2-consistency results for the full model, i.e., when ideally all trajectories are observed at every point of the interval $[0,1]$. The first one corresponds to kernel estimates, and was obtained in [16], while the second one for k-NN with kernel estimates is derived in the present paper. Both results will be used, in Section 3, to prove the consistency of that estimators when only discretely sampled curves in $[0,1]$ are observed.

We will use the notation $f \lesssim g$ when there exists a constant $C > 0$ such that $f \leq Cg$ and $f \approx g$ if there exists a constant $C > 0$ such that $f = Cg$.

Let $(\mathcal{H}, d)$ be a separable metric space and let $(\mathcal{X}_1, Y_1), ...(\mathcal{X}_n, Y_n)$ be independent identically distributed (i.i.d.) random elements in $\mathcal{H}\times\mathbb{R}$ with the same law as the pair $(\mathcal{X}, Y)$ fulfilling the model:

$$Y = \eta(\mathcal{X}) + e\,, \tag{2.1}$$

where the error e satisfies $\mathbb{E}_{e|\mathcal{X}}(e|\mathcal{X}) = 0$ and $\mathrm{var}_{e|\mathcal{X}}(e|\mathcal{X}) = \sigma^2 < \infty$. In this context, the regression function $E(Y|\mathcal{X}) = \eta(\mathcal{X})$ can be estimated by

$$\widehat{\eta}_n(\mathcal{X}) = \sum_{i=1}^{n} W_{ni}(\mathcal{X})\, Y_i\,, \tag{2.2}$$

where the weights $W_{ni}(\mathcal{X}) = W_{ni}(\mathcal{X}, \mathcal{X}_1, ..., \mathcal{X}_n) \geq 0$ and $\sum_{i=1}^{n} W_{ni}(\mathcal{X}) = 1$. In this paper, we first consider the weights corresponding to the family of kernel estimators given by

$$W_{ni}(\mathcal{X}) = \frac{K\left(\frac{d(\mathcal{X},\mathcal{X}_i)}{h_n(\mathcal{X})}\right)}{\sum_{j=1}^{n} K\left(\frac{d(\mathcal{X},\mathcal{X}_j)}{h_n(\mathcal{X})}\right)}\,, \tag{2.3}$$

where K is a regular kernel, i.e., there are constants $0 < c_1 < c_2 < \infty$ such that $c_1 \mathbb{I}_{[0,1]}(u) \leq K(u) \leq c_2\, \mathbb{I}_{[0,1]}(u)$. Here $0/0$ is assumed to be 0. In this general setting, [16] proved the following result.

Proposition 2.1 (Theorem 5.1 in [16]). *Assume that*

K1) *K is a regular and Lipschitz kernel;*

F1) *$(\mathcal{H}, d)$ is a separable metric space;*

F2) *$\{(\mathcal{X}_i, Y_i)\}_{i\geq 1}$ are i.i.d. random elements with the same law as the pair $(\mathcal{X}, Y) \in \mathcal{H}\times\mathbb{R}$ fulfilling model (2.1) with, for each $i = 1, ..., n$, joint distribution $\mathbb{P}_{\mathcal{X},\mathcal{X}_i}$;*

F3) *μ is a Borel probability measure of $\mathcal{X}$ and $\eta \in L^2(\mathcal{H}, \mu) = \big\{f : \mathcal{H} \to \mathbb{R} : \int_{\mathcal{H}} f^2(z)\, d\mu(z) < \infty\big\}$ is a bounded function which satisfies the Besicovitch condition:*

$$\lim_{\delta\to 0} \frac{1}{\mu\big(\mathcal{B}(\mathcal{X}, \delta)\big)} \int_{\mathcal{B}(\mathcal{X},\delta)} |\eta(z) - \eta(\mathcal{X})|\; d\mu(z) \;=\; 0\,, \tag{2.4}$$

in probability, where $\mathcal{B}(\mathcal{X}, \delta)$ is the closed ball of center $\mathcal{X}$ and radius δ with respect to d.

For any $x \in \mathrm{supp}\,(\mu)$ and any sequence $h_n(x) \to 0$ such that $\frac{n\mu(\mathcal{B}(x,h_n(x)))}{\log n} \to \infty$, the estimator given in (2.2) with weights given in (2.3) satisfies

$$\lim_{n\to\infty} \mathbb{E}\left((\widehat{\eta}_n(\mathcal{X}) - \eta(\mathcal{X}))^2\right) \;=\; 0\,.$$

Remark 2.1. The Besicovitch condition in F3 is a differentiation type condition which, as is well known, in finite dimensional spaces automatically holds for any integrable function η. Unfortunately, it is no longer true in infinite dimensional spaces and it can be proved, for instance, that it is necessary in order to get the L_1-consistency of uniform kernel estimates (see Proposition 5.1 in [16]). However, it holds in a general setting if, for instance, the function η is continuous. For a deeper reading on this topic see [10] or [16].

Remark 2.2. Note that for $x \in \mathrm{supp}\,(\mu)$ the consistency of this estimator holds for every sequence $\tilde{h}_n(x) \to 0$ such that $\tilde{h}_n(x) \geq h_n(x)$, where $h_n(x)$ is given in Proposition 2.1, since if $\tilde{h}_n(x) \geq h_n(x)$, then $\frac{n\mu(\mathcal{B}(x,\tilde{h}_n(x)))}{\log n} \geq \frac{n\mu(\mathcal{B}(x,h_n(x)))}{\log n} \to \infty$.

The existence of a sequence verifying $\frac{n\mu(\mathcal{B}(x,h_n(x)))}{\log n} \to \infty$ in Proposition 2.1 follows from the next lemma.

Lemma 2.1 (Lemma A.5 in [16]). *For any $x \in \mathrm{supp}\,(\mu)$, there exists a sequence of positive real numbers $h_n(x) \to 0$ such that $\frac{n\mu(\mathcal{B}(x,h_n(x)))}{\log n} \to \infty$.*

Let $H_n(x)$ be the distance from x to its k_n-nearest neighbor among $\{\mathcal{X}_1, ..., \mathcal{X}_n\}$. Recall that the k_n-nearest neighbor of x among $\{\mathcal{X}_1, ..., \mathcal{X}_n\}$ is the sample point $\mathcal{X}_i$ reaching the k_n-th smallest distance to x in the sample. Then, when the bandwidth in (2.3) is given by $H_n(x)$, we obtain the family of k_n-nearest neighbor (k-NN) with kernel estimates. For the uniform kernel, the consistency of the estimator was proven in [16], Theorem 4.1. For more general kernels, the consistency could be a consequence of Proposition 2.1 if we can prove that $H_n(x) \to 0$ and $\frac{n\mu(\mathcal{B}(x,H_n(x)))}{\log n} \to \infty$. Although it can be proved that $H_n(x) \to 0$ (see [16], Lemma A.4 stated below) the condition $\frac{n\mu(\mathcal{B}(x,H_n(x)))}{\log n} \to \infty$ is not necessary true for $H_n(x)$. However, as we will see in Proposition 2.2, we can still prove the mean square consistency of this estimator under the same weak conditions as in Proposition 2.1.

Lemma 2.2 (Lemma A.4 in [16]). *Let $\mathcal{H}$ be a separable metric space, μ a Borel probability measure, and $\{\mathcal{X}_i\}_{i=1}^n$ a random sample of $\mathcal{X}$. If $x \in \text{supp}\,(\mu)$ and k_n is a sequence of positive real numbers such that $k_n \to \infty$ and $k_n/n \to 0$, then $H_n(x) \to 0$.*

Proposition 2.2. *Assume K1, F1–F3 hold. Let k_n be a sequence of positive real numbers such that $k_n \to \infty$, $k_n/n \to 0$ and let $H_n(x)$ be the distance from x to its k_n-nearest neighbor among $\{\mathcal{X}_1, ..., \mathcal{X}_n\}$. Then, the estimator given by (2.2) with weights given in (2.3) is mean square consistent for any sequence $h_n(x) \to 0$ such that $h_n(x) \geq H_n(x)$, $x \in \text{supp}\,(\mu)$.*

Remark 2.3. Observe that, unlike [15] or [7], we ask d to be a metric not a semi-metric (which is a milder condition). Nevertheless, we do not ask for conditions neither on small ball probabilities nor on the smoothness of the regression function as in the cited papers. Further study is needed to extend ours results to the case of semi-metrics.

3. CONSISTENCY RESULTS FOR DISCRETELY SAMPLED CURVES

In this section we will assume that we are not able to observe the whole trajectories $\mathcal{X}_i$ in $\mathcal{H}$ given in F2, but only a function of them. As we will see in Section 4, different choices of that function will correspond to discretizations, eigenfunction expansions, or smoothing. In this context, the weights of the estimator given in (2.3) cannot be computed because we have not a distance d defined for the discretized sample curves (as a consequence, we do not have the validity of the Besicovitch condition (2.4) for the discretized data) or a bandwidth h_n.

We are interested in defining an estimator and proving its consistency in this setting. For that, let us consider the following assumptions:

- **H1**) *$(\mathcal{H}, d)$ is a separable (metric) Hilbert space and $F\colon \mathcal{H} \to \mathcal{H}$ is a function such that, for each $i = 1, ..., n$, $F(\mathcal{X}_i) = \mathcal{X}_i^p$;*
- **H2**) *$d_p\colon \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a semi-metric in $\mathcal{H}$ defined by $d_p(\mathcal{X}, \mathcal{Y}) = d(\mathcal{X}^p, \mathcal{Y}^p)$ such that there exists a sequence $c_{n,p} \to 0$ as $n, p \to \infty$ satisfying, for each $i = 1, ..., n$,*

$$n^2\, \mathbb{E}_{\mathcal{X}} \left(\mathbb{P}^2_{\mathcal{X}_i|\mathcal{X}} \left(|d(\mathcal{X}, \mathcal{X}_i) - d_p(\mathcal{X}, \mathcal{X}_i)| \geq c_{n,p} \,\middle|\, \mathcal{X} \in \text{supp}\,(\mu) \right) \right) \to 0\,. \tag{3.1}$$

Here, $\mathbb{P}^2_{\mathcal{Y}|\mathcal{X}}(\cdot)$ means the square of $\mathbb{P}_{\mathcal{Y}|\mathcal{X}}(\cdot)$.

Remark 3.1. Observe that in H1 neither $\mathcal{H}$ nor F change with the sample. This implies that in this case, the functional data falls into the category of sparsely and regularly sampled data.

The estimator of η based on $\{(\mathcal{X}_i^p, Y_i)\}_{i=1}^n$ will be defined as in (2.2) and (2.3) but with the semi-metric d_p instead of the metric d. More precisely, for $h_{n,p}(\mathcal{X}) > 0$, we define

$$\widehat{\eta}_{n,p}(\mathcal{X}) \;=\; \frac{\sum_{i=1}^n K\left(\frac{d_p(\mathcal{X},\mathcal{X}_i)}{h_{n,p}(\mathcal{X})}\right) Y_i}{\sum_{j=1}^n K\left(\frac{d_p(\mathcal{X},\mathcal{X}_j)}{h_{n,p}(\mathcal{X})}\right)}\,. \tag{3.2}$$

For this estimator, we state the following two asymptotic results.

Theorem 3.1. *Assume K1, F2, F3, H1 and H2 hold.*

(a) *(Kernel estimator) For any $x \in supp\,(\mu)$, let $h_n^*(x) \to 0$ be a sequence of positive real numbers such that $\frac{n\mu(\mathcal{B}(x,h_n^*(x)))}{\log n} \to \infty$. Then, for $c_{n,p}$ given in H2 and $h_{n,p}(x) \to 0$ such that there exists a sequence $h_n(x) \to 0$, $h_n(x) \geq h_n^*(x)$ satisfying:*

(H3.1) $\mathbb{E}_{\mathcal{X}}\left(c_{n,p}^2/h_n^2(\mathcal{X})\right) \to 0$ *as* $n,p \to \infty$;

(H3.2) $c_{n,p} \leq h_{n,p}(x) - h_n(x) \leq C_2\, c_{n,p}$ *for* $C_2 \geq 1$;

we have

$$\lim_{n,p\to\infty} \mathbb{E}\left((\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2\right) = 0\,. \tag{3.3}$$

(b) *(k_n-NN with kernel estimator) Let $c_{n,p}$ given in H2 and $H_n(x)$ the distance from x to its k_n-nearest neighbor among $\{\mathcal{X}_1,...,\mathcal{X}_n\}$. For any $x \in supp\,(\mu)$, let $h_{n,p}(x) \to 0$ be such that there exists a sequence $h_n(x) \to 0$, $h_n(x) \geq H_n(x)$ satisfying assumptions (H3.1) and (H3.2). Then, for $k_n \to \infty$ and $k_n/n \to 0$ we have (3.3).*

Remark 3.2. Observe that the sequence $h_n^*(x)$ in Theorem 3.1 always exists by Lemma 2.1. In addition, under H2, it is always possible to choose a sequence $h_{n,p}(x) \to 0$ fulfilling the conditions in Theorem 3.1. Indeed, taking $h_n(x) = \max\{h_n^*(x), \sqrt{c_{n,p}}\}$ and $h_{n,p}(x) = h_n(x) + Cc_{n,p}$, with $C \geq 1$, we have that $h_n(x) \to 0$, $h_{n,p}(x) \to 0$, $h_n(x) \geq h_n^*(x)$, (H3.1) holds since $h_n(x) \geq \sqrt{c_{n,p}}$ and (H3.2) holds by definition of $h_{n,p}(x)$. The same happens if instead of taking $h_n^*(x)$ we take $H_n(x)$.

Theorem 3.2. *Under the assumptions of Theorem 3.1, let $\gamma_n \to \infty$ as $n \to \infty$ be such that, as $n,p \to \infty$,*

(a) $\mathbb{E}_{\mathcal{X}}\left(\gamma_n\left(\frac{c_{n,p}}{h_n(\mathcal{X})}\right)^2\right) \to 0$;

(b) $\gamma_n\, n^2\, \mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_i|\mathcal{X}}\left(|d(\mathcal{X},\mathcal{X}_i) - d_p(\mathcal{X},\mathcal{X}_i)| \geq c_{n,p} \,\middle|\, \mathcal{X} \in supp\,(\mu)\right)\right) \to 0,$ *for each* $i = 1,...,n$.

Then

$$\lim_{n\to\infty} \mathbb{E}\left(\gamma_n(\widehat{\eta}_n(\mathcal{X}) - \eta(\mathcal{X}))^2\right) = 0$$

implies

$$\lim_{n,p\to\infty} \mathbb{E}\left(\gamma_n(\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2\right) = 0\,.$$

4. PARTICULAR CASES

In this section we provide definitions of $\mathcal{H}$ and d_p for discretization, smoothing, and eigenfunction expansions, which satisfy conditions H1 and H2. Then, for any sequence $h_{n,p}(x) \to 0$ satisfying (H3.1) and (H3.2) in Theorem 3.1, we get the consistency of $\hat{\eta}_{n,p}$ as a consequence of the consistency results for $\hat{\eta}_n$ in the full model.

Consider the case where the elements of the dataset are curves in $L^2([0,1])$ that are only observed at a discrete set of points in the interval $[0,1]$. More precisely, let us assume that $\{\mathcal{X}_i\}_{i=1}^n$ are observed only at some points: $(\mathcal{X}_i(t_1),...,\mathcal{X}_i(t_{p+1}))$ where $0 = t_1 < t_2 < \cdots < t_{p+1} = 1$, which for simplicity we will assume are equally spaced, i.e., $\Delta t = t_{i+1} - t_i = 1/p$. In this case, we will need to require the trajectories to satisfy some regularity condition. More precisely, we will assume that $\mathcal{X}$ is a random element of $\mathcal{H} \doteq H^1([0,1])$, the Sobolev space defined as

$$H^1([0,1]) \;=\; \Big\{f\colon [0,1] \to \mathbb{R}\colon f \text{ and } Df \in L^2([0,1])\Big\}\,,$$

where Df is the weak derivative of f, i.e., Df is a function in $L^2([0,1])$ which satisfies

$$\int_0^1 f(t)\, D\phi(t)\, dt \;=\; -\int_0^1 Df(t)\,\phi(t)\, dt\,, \qquad \forall\, \phi \in C_0^\infty.$$

In this space, the norm is defined by

$$\|f\|_{H^1([0,1])} \;=\; \|f\|_{L^2([0,1])} + \|Df\|_{L^2([0,1])}\,.$$

In this setting, we will prove consistency for the semi-metrics d_p given below.

4.1. Discretization

Consider the semi-metric

$$d_p(\mathcal{X},\mathcal{X}_1) \;=\; d(\mathcal{X}^p,\mathcal{X}_1^p) \;=\; \left(\frac{1}{p}\sum_{j=1}^p |\mathcal{X}(t_j) - \mathcal{X}_1(t_j)|^2\right)^{1/2}\,,$$

where $\mathcal{X}^p(t) = F(\mathcal{X})(t) = \sum_{j=1}^p \phi_j(t)\mathcal{X}(t_j)$ with $\phi_j(t) = \mathbb{I}_{[t_j,t_{j+1})}(t)$. In this case, consistency will hold for any sequence $c_{n,p} \to 0$ as $n,p \to \infty$ such that $n^2\,\mathbb{P}_{\mathcal{X},\mathcal{X}_1}(\|\mathcal{X}\|_{\mathcal{H}} + \|\mathcal{X}_1\|_{\mathcal{H}} \geq p\,c_{n,p}) \to 0$.

4.2. Kernel smoothing

Let us consider now the semi-metric

$$d_p(\mathcal{X},\mathcal{X}_1) \;=\; d(\mathcal{X}^p,\mathcal{X}_1^p) \;=\; \left(\int_0^1 |\mathcal{X}^p(t) - \mathcal{X}_1^p(t)|^2\, dt\right)^{1/2}\,,$$

where $\mathcal{X}^p(t) = F(\mathcal{X})(t) = \sum_{j=1}^p \phi_j(t)\mathcal{X}(t_j)$ with $\phi_j(t) = \frac{K(|t-t_j|/h)}{\sum_{i=1}^p K(|t-t_i|/h)}$ and K is a regular kernel supported in $[0,1]$. In this case, consistency will be true for any sequence $c_{n,p} \to 0$ as $n,p \to \infty$ satisfying $n^2\,\mathbb{P}_{\mathcal{X},\mathcal{X}_1}(\|\mathcal{X}\|_{\mathcal{H}} + \|\mathcal{X}_1\|_{\mathcal{H}} \geq p c_{n,p}) \to 0$.

Let us note that if $\mathbb{E}_{\mathcal{X}}(\|\mathcal{X}\|_{\mathcal{H}}^2) < \infty$, the consistency for the cases given in Sections 4.1 and 4.2 will hold for any sequence $c_{n,p}$ such that $\frac{n}{p\,c_{n,p}} \to 0$.

4.3. Eigenfunction expansions

Let $\mathcal{X}, \mathcal{X}_1$ be i.d. random elements on $\mathcal{H} = L^2[0,1]$. Let $v_1, v_2, \ldots$ be the orthonormal eigenfunctions of the covariance operator $\mathbb{E}_{\mathcal{X}}(\mathcal{X}(t)\mathcal{X}(s))$ (without loss of generality we have assumed that $\mathbb{E}(\mathcal{X}(t)) = 0$) associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ such that

$$\mathbb{E}_{\mathcal{X}}(\mathcal{X}(t)\mathcal{X}(s)) = \sum_{k=1}^{\infty} \lambda_k v_k(t) v_k(s) .$$

If $\mathbb{E}\left(\int \mathcal{X}^2(s)\, ds\right) < \infty$ is finite, using the Karhunen–Loève representation, we can write $\mathcal{X}$ as

$$\mathcal{X}(t) = \sum_{k=1}^{\infty} \left(\int \mathcal{X}(s) v_k(s)\, ds \right) v_k(t) \doteq \sum_{k=1}^{\infty} \xi_k v_k(t) , \tag{4.1}$$

with $\mathbb{E}(\xi_k) = 0$, $\mathbb{E}(\xi_k \xi_j) = 0$ (i.e., $\xi_1, \xi_2, \ldots$ uncorrelated) and $\operatorname{var}(\xi_k) = \mathbb{E}\left(\xi_k^2\right) = \lambda_k = \mathbb{E}\left(\left(\int \mathcal{X}(s) v_k(s)\, ds\right)^2\right)$. The classical L^2-norm in $\mathcal{H}$ can be written as

$$d(\mathcal{X}, \mathcal{X}_1) = \sqrt{\sum_{k=1}^{\infty} \left(\int (\mathcal{X}(t) - \mathcal{X}_1(t))\, v_k(t)\, dt \right)^2} . \tag{4.2}$$

If we consider the truncated expansion of $\mathcal{X}$ as given in [15],

$$\mathcal{X}^p(t) = \sum_{k=1}^{p} \left(\int \mathcal{X}(s) v_k(s)\, ds \right) v_k(t) , \tag{4.3}$$

we can define the parametrized class of seminorms from the classical L^2-norm given by

$$\|\mathcal{X}\|_p = \sqrt{\int \left(\mathcal{X}^p(t)\right)^2 dt} = \sqrt{\sum_{k=1}^{p} \left(\int \mathcal{X}(t) v_k(t)\, dt \right)^2} ,$$

which leads to the semi-metric

$$d_p(\mathcal{X}, \mathcal{X}_1) = d(\mathcal{X}^p, \mathcal{X}_1^p) = \sqrt{\sum_{k=1}^{p} \left(\int (\mathcal{X}(t) - \mathcal{X}_1(t))\, v_k(t)\, dt \right)^2} . \tag{4.4}$$

In this case, the consistency will hold for any sequence $c_{n,p} \to 0$ such that $\frac{n^2}{c_{n,p}^2} \sum_{k=p+1}^{\infty} \lambda_k \to 0$ as $n, p \to \infty$.

5. SIMULATION STUDY

In order to illustrate the results given in Theorems 3.1 and 3.2, we perform a small simulation study where we compare the behaviour of the estimators, $\widehat{\eta}_n$ and $\widehat{\eta}_{n,p}$ for finite sample sizes settings. Following [7], we simulate n pairs $\{(\mathcal{X}_i(t), Y_i)\}_{i=1}^n$ where, for $t \in [0, \pi]$, and for each $i = 1, \ldots, n$,

$$\mathcal{X}_i(t) = a_i \cos(2t) , \qquad a_i \sim N(0, \sigma = 0.1) .$$

The plot of $n = 100$ curves is shown in Figure 1.

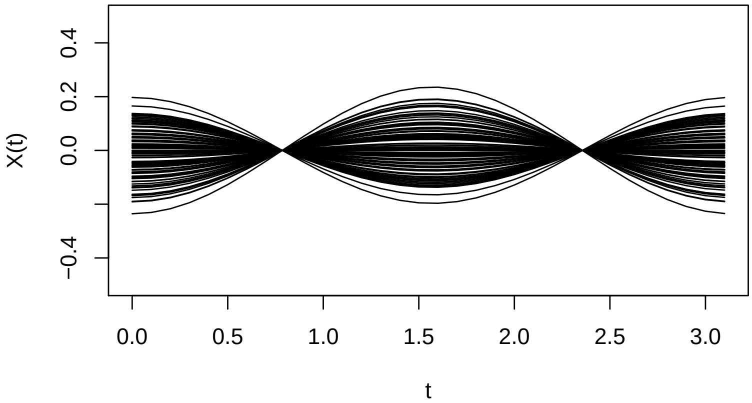

Figure 1: Simulated curves for $n = 100$.

The responses were generated following the model

$$Y_i = \eta(\mathcal{X}_i) + \epsilon_i \, , \qquad \epsilon_i \sim N(0, \sigma = 0.4) \, ,$$

for different regression functions η as listed below:

Setting 1: $\eta(\mathcal{X}_i) = a_i^2$ (see [7]);

Setting 2: $\eta(\mathcal{X}_i) = \left(\int_0^\pi \sin(4\pi t)\, \mathcal{X}_i(t)\, dt\right)^2$ (see [11]);

Setting 3: $\eta(\mathcal{X}_i) = \int_0^\pi |\mathcal{X}_i(t)| \log(|\mathcal{X}_i(t)|)\, dt$ (see [14]);

Setting 4: $\eta(\mathcal{X}_i) = \int_0^\pi \mathcal{X}_i^2(t)\, dt$ (see [2]).

For the *full model* we used the classical L^2-metric which in this case gives

$$\begin{aligned} d(\mathcal{X}_i, \mathcal{X}_j) \;&=\; \left(\int_0^\pi \big(\mathcal{X}_i(t) - \mathcal{X}_j(t)\big)^2\, dt\right)^{1/2} \;=\; \left(\int_0^\pi (a_i - a_j)^2 \cos^2(2t)\, dt\right)^{1/2} \\ &=\; \left(\int_0^\pi \cos^2(2t)\, dt\right)^{1/2} |a_i - a_j| \;=\; \sqrt{\frac{\pi}{2}}\; |a_i - a_j| \, . \end{aligned}$$

For the discretized model, we divided the interval of time $[0, \pi]$ in $p{+}1$ subintervals of length $\frac{\pi}{p}$. The semimetric in this case is given by

$$\begin{aligned} d_p(\mathcal{X}, \mathcal{X}_1) \;&=\; d(\mathcal{X}^p, \mathcal{X}_1^p) \;=\; \left(\int_0^\pi \left|\mathcal{X}^p(t) - \mathcal{X}_1^p(t)\right|^2\, dt\right)^{1/2} \\ &\approx\; \left(\frac{1}{p} \sum_{k=1}^p \big(\mathcal{X}_i(t_k) - \mathcal{X}_j(t_k)\big)^2\right)^{1/2} . \end{aligned}$$

For both estimators $\widehat{\eta}_n$ and $\widehat{\eta}_{n,p}$, we used the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)\, \mathbb{I}_{[0,1](u)}$ and the bandwidths h_n and $h_{n,p}$ were chosen via cross validation.

In both cases the sample of size n was divided in two samples of the same size, the learning sample, used to compute the optimal smoothing parameter and the testing sample, used to measure the power of both methods by the Mean Square Error (MSE). For different combination of n and p we repeated 250 times the procedure of building $n/2$ learning samples and $n/2$ testing samples and computing the MSE's for the full and discretized models.

The following tables show the mean over the 250 MSE's for all estimators. As we can see, the simulations confirm our theoretical results since, for the four different settings we can see the consistency as $n, p \to \infty$ stated in Theorem 3.1 and also the equal order or convergence stated in Theorem 3.2.

Table 1: MSE's for Setting 1.

n	Discretized model				Full model
	20	40	60	80	
50	0.1871725	0.1829381	0.1819154	0.1817674	0.1818614
100	0.1784129	0.1661579	0.1661309	0.1660854	0.1659922
150	0.1727869	0.1674195	0.1675846	0.1674071	0.1672996
200	0.1671014	0.1629972	0.1629855	0.1630360	0.1631458
250	0.1646048	0.1631582	0.1631817	0.1632266	0.1632193
300	0.1653583	0.1638297	0.1637960	0.1638118	0.1637993

Table 2: MSE's for Setting 2.

n	Discretized model				Full model
	20	40	60	80	
50	0.1919580	0.1796157	0.1795600	0.1789984	0.1789860
100	0.1787471	0.1684685	0.1684097	0.1684710	0.1685058
150	0.1731875	0.1661859	0.1661971	0.1663508	0.1663451
200	0.1695872	0.1646054	0.1646025	0.1646861	0.1646566
250	0.1658714	0.1622371	0.1621559	0.1621067	0.1621016
300	0.1655437	0.1633919	0.1634236	0.1634164	0.1634100

Table 3: MSE's for Setting 3.

n	Discretized model				Full model
	20	40	60	80	
50	0.1875816	0.1752962	0.1744660	0.1751941	0.1748388
100	0.1797477	0.1672346	0.1671503	0.1671671	0.1671481
150	0.1706658	0.1662048	0.1661369	0.1661024	0.1660888
200	0.1696802	0.1683357	0.1681568	0.1681344	0.1681435
250	0.1666817	0.1651802	0.1652298	0.1652369	0.1652162
300	0.1626991	0.1622967	0.1623146	0.1622935	0.1623169

Table 4: MSE's for Setting 4.

n	Discretized model				Full model
	20	40	60	80	
50	0.1951465	0.1867710	0.1872990	0.1870323	0.1869950
100	0.1824836	0.1694453	0.1694464	0.1695669	0.1695569
150	0.1717909	0.1655053	0.1656256	0.1657503	0.1657367
200	0.1692647	0.1657557	0.1655030	0.1655163	0.1655050
250	0.1651644	0.1630851	0.1631351	0.1630439	0.1630378
300	0.1665684	0.1655066	0.1655070	0.1654343	0.1654715

APPENDIX A – Proofs of auxiliary results

To prove the consistency of the examples given in sections 4.1 and 4.2 we need the following result.

Proposition A.1. *Let $\mathcal{X}^p(t) = \sum_{j=1}^p \phi_j(t)\mathcal{X}(t_j)$ with ϕ_j satisfying:*

(**a**) *for each $t \in [0,1]$, $\sum_{j=1}^p \phi_j(t) = 1$;*

(**b**) *for each $t \in [0,1]$, $\sum_{j=i}^p \phi_j^2(t) \leq C_3$ for some constant C_3;*

(**c**) *$\operatorname{supp}(\phi_j) \subset [t_{(j-m)}, t_{(j+m)}]$ with m independent of p.*

If $c_{n,p} \to 0$ as $n,p \to \infty$ is such that $n^2 \mathbb{P}_{\mathcal{X},\mathcal{X}_1}(\|\mathcal{X}\|_{\mathcal{H}} + \|\mathcal{X}_1\|_{\mathcal{H}} \geq p c_{n,p}) \to 0$, H2 is fulfilled.

Proof of Proposition A.1: Using the Fundamental Theorem of Calculus (FTC) (see Theorem 8.2 in [6]) for $H^1([0,1])$, we get

$$
\begin{aligned}
d^2(\mathcal{X}^p,\mathcal{X}) &= \int_0^1 \left| \sum_{j=1}^p \mathcal{X}(t_j)\phi_j(t) - \mathcal{X}(t) \right|^2 dt \\
&= \int_0^1 \left| \sum_{j=1}^p (\mathcal{X}(t_j) - \mathcal{X}(t))\phi_j(t) \right|^2 dt && \text{(by (a))} \\
&= \int_0^1 \left| \sum_{j=1}^p \left(\int_{t_j}^t D\mathcal{X}(s)\,ds \right) \phi_j(t) \right|^2 dt && \text{(from FTC)} \\
&\leq \int_0^1 \left(\sum_{j=1}^p \left(\int_{t_j}^t D\mathcal{X}(s)\,ds \right)^2 \mathbb{I}_{\{\operatorname{supp}(\phi_j)\}}(t) \right) \left(\sum_{j=1}^p \phi_j^2(t) \right) dt && \text{(by C-S Ineq.)} \\
&\lesssim \int_0^1 \sum_{j=1}^p \left(\int_{t_j}^t D\mathcal{X}(s)\,ds \right)^2 \mathbb{I}_{\{\operatorname{supp}(\phi_j)\}}(t)\,dt && \text{(by (b))} \\
&\lesssim \int_0^1 \sum_{j=1}^p \left(\int_{t_j}^t (D\mathcal{X}(s))^2\,ds \right) |t - t_j| \mathbb{I}_{\{\operatorname{supp}(\phi_j)\}}(t)\,dt && \text{(by C-S Ineq.)} \\
&= \sum_{i=1}^p \int_{t_i}^{t_{i+1}} \sum_{\substack{j=1 \\ j:|j-i|\leq m}}^p \left(\int_{t_j}^t (D\mathcal{X}(s))^2\,ds \right) |t - t_j|\,dt && \text{(by (c))} \\
&\lesssim \sum_{i=1}^p \sum_{\substack{j=1 \\ j:|j-i|\leq m}}^p \int_{t_{i-m}}^{t_{i+m}} (D\mathcal{X}(s))^2 \left(\int_{t_j}^{t_{j+1}} |t - t_j|\,dt \right) ds \\
&\lesssim \frac{m}{p^2} \sum_{i=1}^p \sum_{\substack{j=1 \\ j:|j-i|\leq m}}^p \int_{t_{i-m}}^{t_{i+m}} (D\mathcal{X}(s))^2\,ds
\end{aligned}
$$

$$
\begin{aligned}
&\lesssim \frac{m^2}{p^2}\sum_{i=1}^{p}\int_{t_{i-m}}^{t_{i+m}}(D\mathcal{X}(s))^2\,ds\\
&= \frac{m^2}{p^2}\int_0^1\sum_{i=1}^{p}\mathbb{I}_{[t_{i-m},t_{i+m}]}(s)(D\mathcal{X}(s))^2\,ds \;\lesssim\; \frac{1}{p^2}\left\|\mathcal{X}\right\|^2_{\mathcal{H}},
\end{aligned}
$$

from where we get $d(\mathcal{X}^p,\mathcal{X})\lesssim\frac{1}{p}\left\|\mathcal{X}\right\|_{\mathcal{H}}$. Analogously we can prove that $d(\mathcal{X}_1^p,\mathcal{X}_1)\lesssim\frac{1}{p}\left\|\mathcal{X}_1\right\|_{\mathcal{H}}$. By triangular inequality,

$$
\begin{aligned}
n^2\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_1|\mathcal{X}}\big(\,|d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1)|\geq c_{n,p}\Big|\mathcal{X}\in\mathrm{supp}\,(\mu)\,\big)\right)&\\
\leq\; n^2\mathbb{P}_{\mathcal{X},\mathcal{X}_1}\left(\left\|\mathcal{X}\right\|_{\mathcal{H}}+\left\|\mathcal{X}_1\right\|_{\mathcal{H}}\geq pc_{n,p}\right),&
\end{aligned}
$$

and therefore, for any $c_{n,p}\to 0$ such that $n^2\,\mathbb{P}_{\mathcal{X},\mathcal{X}_1}(\left\|\mathcal{X}\right\|_{\mathcal{H}}+\left\|\mathcal{X}_1\right\|_{\mathcal{H}}\geq p\,c_{n,p})\to 0$ H2 is fulfilled. □

A.1. Consistency for the example in Section 4.1

Since the functions $\phi_j(t)=\mathbb{I}_{[t_j,t_{j+1})}(t)$ satisfy trivially conditions (a)–(c) of Proposition A.1, H2 is fulfilled and therefore, for any sequence $h_{n,p}(x)\to 0$ satisfying (H3.1) and (H3.2) in Theorem 3.1, we get the consistency of $\hat{\eta}_{n,p}$.

A.2. Consistency for the example in Section 4.2

Observe that $\phi_j(t)=\frac{K(|t-t_j|/h)}{\sum_{i=1}^{p}K(|t-t_i|/h)}$ satisfies conditions (a)–(c) in Proposition A.1:

(**a**) for each $t\in[0,1]$, $\sum_{j=1}^{p}\phi_j(t)=\sum_{j=1}^{p}\frac{K(|t-t_j|/h)}{\sum_{i=1}^{p}K(|t-t_i|/h)}=1$;

(**b**) since K is nonnegative and $\frac{K(|t-t_j|/h)}{\sum_{i=1}^{p}K(|t-t_i|/h)}\leq 1$, for each $t\in[0,1]$, there exists $C_3=1$ such that

$$
\sum_{j=1}^{p}\phi_j^2(t)\;=\;\sum_{j=1}^{p}\left(\frac{K(|t-t_j|/h)}{\sum_{i=1}^{p}K(|t-t_i|/h)}\right)^2\leq\sum_{j=1}^{p}\frac{K(|t-t_j|/h)}{\sum_{i=1}^{p}K(|t-t_i|/h)}\;=1;
$$

(**c**) $\mathrm{supp}\,(\phi_j)=\mathrm{supp}\,(K(|t-t_j|/h))=[t_j-h,t_j+h]$, which implies that, for $h\leq m/p$, $\mathrm{supp}\,(\phi_j)\subset[t_{(j-m)},t_{(j+m)}]$.

This implies that H2 is fulfilled then, for any sequence $h_{n,p}(x)\to 0$ satisfying (H3.1) and (H3.2) in Theorem 3.1, we get the consistency of $\hat{\eta}_{n,p}$.

A.3. Consistency for the example in Section 4.3

Let us consider the truncated expansion of $\mathcal{X}$, $\mathcal{X}^p(t)$, given by (4.3) and the pseudo-metric $d_p(\mathcal{X},\mathcal{X}_1)=d(\mathcal{X}^p,\mathcal{X}_1^p)$ given by (4.4). In order to prove H2, let us consider $c_{n,p}$ such that $\frac{n^2}{c_{n,p}^2}\sum_{k=p+1}^{\infty}\lambda_k\to 0$. Using Chebyshev's Inequality in (3.1) followed by Cauchy Schwartz, we get

$$n^2\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_1|\mathcal{X}}\big(\,|d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1)|\geq c_{n,p}\Big|\mathcal{X}\in\operatorname{supp}(\mu)\,\big)\right) \leq \frac{n^2}{c_{n,p}^2}\mathbb{E}_{\mathcal{X},\mathcal{X}_1}\big(\,(d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1))^2\,\big). \tag{A.1}$$

Now, since $d(\mathcal{X},\mathcal{X}_1)\geq d_p(\mathcal{X},\mathcal{X}_1)$ we have that $0\leq d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1)=d(\mathcal{X},\mathcal{X}_1)-d(\mathcal{X}^p,\mathcal{X}_1^p)$ and, by triangular inequality $d(\mathcal{X},\mathcal{X}_1)\leq d(\mathcal{X},\mathcal{X}^p)+d(\mathcal{X}^p,\mathcal{X}_1^p)+d(\mathcal{X}_1^p,\mathcal{X}_1)$ which implies that

$$0\;\leq\;d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1)\;\leq\;d(\mathcal{X},\mathcal{X}^p)+d(\mathcal{X}_1^p,\mathcal{X}_1) \tag{A.2}$$

and, taking squares,

$$0\;\leq\;(d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1))^2\;\leq\;(d(\mathcal{X},\mathcal{X}^p)+d(\mathcal{X}_1^p,\mathcal{X}_1))^2\;\leq\;2\left(d^2(\mathcal{X},\mathcal{X}^p)+d^2(\mathcal{X}_1^p,\mathcal{X}_1)\right).$$

As a consequence, to proof this proposition it will sufficient to bound $\mathbb{E}_{\mathcal{X}}\left(d^2(\mathcal{X},\mathcal{X}^p)\right)$ (equivalently, $\mathbb{E}_{\mathcal{X}_1}\left(d^2(\mathcal{X}_1,\mathcal{X}_1^p)\right)$). Since v_k are orthonormal,

$$\begin{aligned} d^2(\mathcal{X},\mathcal{X}^p) &= \int\left(\mathcal{X}(s)-\sum_{k=1}^{p}\left(\int\mathcal{X}(t)v_k(t)\,dt\right)v_k(s)\right)^2 ds \\ &= \sum_{k=p+1}^{\infty}\left(\int\mathcal{X}(t)v_k(t)\,dt\right)^2. \end{aligned}$$

Then, we have

$$\begin{aligned} \mathbb{E}_{\mathcal{X}}\left(d^2(\mathcal{X},\mathcal{X}^p)\right) &= \mathbb{E}_{\mathcal{X}}\left(\sum_{k=p+1}^{\infty}\left(\int\mathcal{X}(t)v_k(t)\,dt\right)^2\right) \\ &= \sum_{k=p+1}^{\infty}\lambda_k \qquad\qquad \text{(from (4.1))}\,. \end{aligned}$$

Analogously we can prove that $\mathbb{E}_{\mathcal{X}_1}\left(d^2(\mathcal{X}_1,\mathcal{X}_1^p)\right)=\sum_{k=p+1}^{\infty}\lambda_k$. Therefore, in (A.1) we get

$$n^2\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_1|\mathcal{X}}\big(\,|d(\mathcal{X},\mathcal{X}_1)-d_p(\mathcal{X},\mathcal{X}_1)|\geq c_{n,p}\Big|\mathcal{X}\in\operatorname{supp}(\mu)\,\big)\right)\lesssim\frac{n^2}{c_{n,p}^2}\sum_{k=p+1}^{\infty}\lambda_k\;\to\;0\,.$$

This implies that H2 is fulfilled then, for any sequence $h_{n,p}(x)\to 0$ satisfying (H3.1) and (H3.2) in Theorem 3.1, we get the consistency of $\hat{\eta}_{n,p}$.

APPENDIX B – Proof of Proposition 2.2 and Theorems 3.1 and 3.2

To prove Proposition 2.2 we need some preliminary results whose proofs can be found in [16].

Theorem B.1 (Theorem 3.4). *If $\eta \in L^2(\mathcal{H}, \mu)$ and $\widehat{\eta}_n$ is the estimator given in (2.2) with weights $W_n(\mathcal{X}) = \{W_{ni}(\mathcal{X})\}_{i=1}^n$ satisfying the following conditions:*

(**i**) *there is a sequence of nonnegative random variables $a_n(\mathcal{X}) \to 0$ a.s. such that*
$$\lim_{n\to\infty} \mathbb{E}\left(\sum_{i=1}^n W_{ni}(\mathcal{X})\mathbb{I}_{\{d(\mathcal{X},\mathcal{X}_i)>a_n(\mathcal{X})\}}\right) = 0\,;$$

(**ii**)
$$\lim_{n\to\infty} \mathbb{E}\left(\max_{1\le i\le n} W_{ni}(\mathcal{X})\right) = 0\,;$$

(**iii**) *for all $\epsilon > 0$ there exists $\delta > 0$ such that for any η^* bounded and continuous function fulfilling $\mathbb{E}_{\mathcal{X}}((\eta(\mathcal{X}) - \eta^*(\mathcal{X}))^2) < \delta$ we have that*
$$\mathbb{E}\left(\sum_{i=1}^n W_{ni}(\mathcal{X})(\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2\right) < \epsilon\,;$$

then $\widehat{\eta}_n$ is mean square consistent.

Corollary B.1 (Corollary 3.3). *Let U_n be a sequence of probability weights satisfying conditions (i), (ii) and (iii) of Theorem B.1. If W_n is a sequence of weights such that $\sum_{i=1}^n W_{ni}(\mathcal{X}) = 1$ and, for each $n \ge 1$, $|W_n| \le MU_n$ for some constant $M \ge 1$, then the estimator given in (2.2) with weights $W_n(\mathcal{X})$ is mean square consistent.*

Lemma B.1 (Lemma A.1). *Let $\mathcal{H}$ be a separable metric space. If $A = supp\,(\mu) = \{x \in \mathcal{H}\colon \mu\big(\mathcal{B}(x,\epsilon)\big) > 0, \forall\, \epsilon > 0)\}$ then $\mu\big(A\big) = 1$.*

Proof of Proposition 2.2: Let $x \in \operatorname{supp}(\mu)$ be fixed. Let us observe that, since K is regular, there exist constants $0 < c_1 < c_2 < \infty$ such that, for each i,

$$W_{ni}(x) = \frac{K\left(\frac{d(\mathcal{X}_i,x)}{h_n(x)}\right)}{\sum_{j=1}^n K\left(\frac{d(\mathcal{X}_j,x)}{h_n(x)}\right)} \le \frac{c_2}{c_1}\frac{\mathbb{I}_{\{d(\mathcal{X}_i,x)\le h_n(x)\}}}{\sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}_j,x)\le h_n(x)\}}} \doteq \frac{c_2}{c_1}U_{ni}(x)\,. \tag{B.1}$$

Let $h_n(x) \to 0$ such that $h_n(x) \ge H_n(x)$ ($H_n(x) \to 0$ by Lemma 2.2, for $x \in \operatorname{supp}(\mu)$). From (B.1) and Corollary B.1, it suffices to prove that the weights U_{ni} satisfy conditions (i), (ii) and (iii) of Theorem B.1. To prove (i) let us take $a_n(x) = h_n^{1/2}(x) \to 0$. Then, by Lemma B.1,

$$\begin{aligned}\mathbb{E}&\left(\sum_{i=1}^n U_{ni}(\mathcal{X})\mathbb{I}_{\{d(\mathcal{X}_i,\mathcal{X})>h_n(\mathcal{X})^{1/2}\}}\right)\\ &= \mathbb{E}_{\mathcal{X}}\left(\mathbb{E}_{\mathcal{D}_n|\mathcal{X}}\left(\mathbb{I}_{\{\mathcal{X}\in\operatorname{supp}(\mu)\}}\sum_{i=1}^n U_{ni}(\mathcal{X})\mathbb{I}_{\{d(\mathcal{X}_i,\mathcal{X})>h_n(\mathcal{X})^{1/2}\}}\Big|\mathcal{X}\in\operatorname{supp}(\mu)\right)\right).\end{aligned}$$

Given $\epsilon > 0$, let $x \in \operatorname{supp}(\mu)$ be fixed. Since $h_n(x) \to 0$, there exists $N_1 = N_1(x)$ such that if $n \geq N_1$, $\mathbb{I}_{\left\{h_n(x)^{1/2} < d(x_i,x) \leq h_n(x)\right\}} = 0$ for all i and, consequently,

$$\mathbb{E}_{\mathcal{D}_n}\left(\frac{1}{\sum_{j=1}^n \mathbb{I}_{\{d(x_j,x)\leq h_n(x)\}}} \sum_{i=1}^n \mathbb{I}_{\left\{h_n(x)^{1/2} < d(x_i,x) \leq h_n(x)\right\}}\right) < \epsilon\,.$$

In addition, $\frac{\sum_{i=1}^n \mathbb{I}_{\left\{h_n(x)^{1/2} < d(x_i,x) \leq h_n(x)\right\}}}{\sum_{j=1}^n \mathbb{I}_{\left\{d(x_j,x)\leq h_n(x)\right\}}} \leq 1$, from what follows that

$$\mathbb{E}_{\mathcal{D}_n}\left(\frac{1}{\sum_{j=1}^n \mathbb{I}_{\{d(x_j,x)\leq h_n(x)\}}} \sum_{i=1}^n \mathbb{I}_{\left\{h_n(x)^{1/2} < d(x_i,x) \leq h_n(x)\right\}}\right) \leq 1\,.$$

Therefore, by the dominated convergence theorem we have that condition (i) is satisfied. Now, since $h_n(x) \geq H_n(x)$,

$$\sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}_j,x)\,\leq\, h_n(x)\}} \geq \sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}_j,x)\leq H_n(x)\}} \;=\; k_n \;\to\; \infty\,.$$

Therefore,

$$\max_{1\leq i\leq n} U_{ni}(x) \;\leq\; \max_{1\leq i\leq n} \frac{1}{\sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}_j,x)\leq h_n(x)\}}} \;\leq\; \frac{1}{k_n} \;\to\; 0\,,$$

from what we derive (ii) using the dominated convergence theorem. It remains to verify that condition (iii) holds. Since $\eta \in L^2(\mathcal{H},\mu)$ which is separable and complete, there exists η^* continuous and bounded such that, for all $\delta > 0$, $\mathbb{E}_{\mathcal{X}}((\eta(\mathcal{X}) - \eta^*(\mathcal{X}))^2) < \delta$. Then,

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^n U_{ni}(\mathcal{X})(\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2\right) \\
&\qquad = \;\mathbb{E}_{\mathcal{X}}\left(\mathbb{E}_{\mathcal{D}_n|\mathcal{X}}\left(\mathbb{I}_{\{\mathcal{X}\in\operatorname{supp}(\mu)\}} \sum_{i=1}^n U_{ni}(\mathcal{X})(\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2 | \mathcal{X} \in \operatorname{supp}(\mu)\right)\right).
\end{aligned}$$

Let $x \in \operatorname{supp}(\mu)$ be fixed. From [16], Lemma A.7, for any nonnegative bounded measurable function f, we have

$$\mathbb{E}_{\mathcal{D}_n}\left(\sum_{i=1}^n U_{ni}(x) f(\mathcal{X}_i)\right) \;\leq\; 12\,\frac{1}{\mu\big(\mathcal{B}(x,h_n(x))\big)} \int_{\mathcal{B}(x,h_n(x))} f(y)\,d\mu(y)\,.$$

Then, applying the inequality to $f(\mathcal{X}_i) = (\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2$, we get

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_n}\left(\sum_{i=1}^n U_{ni}(x)(\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2\right) \\
&\qquad\lesssim \frac{1}{\mu\big(\mathcal{B}(x,h_n(x))\big)} \int_{\mathcal{B}(x,h_n(x))} (\eta^*(y) - \eta(y))^2\,d\mu(y) \\
&\qquad\leq \frac{1}{\mu\big(\mathcal{B}(x,h_n(x))\big)} \int_{\mathcal{B}(x,h_n(x))} (\eta^*(y) - \eta^*(x))^2\,d\mu(y) \\
&\qquad\quad + \frac{1}{\mu\big(\mathcal{B}(x,h_n(x))\big)} \int_{\mathcal{B}(x,h_n(x))} (\eta^*(x) - \eta(x))^2\,d\mu(y) \\
&\qquad\quad + \frac{1}{\mu\big(\mathcal{B}(x,h_n(x))\big)} \int_{\mathcal{B}(x,h_n(x))} (\eta(x) - \eta(y))^2\,d\mu(y) \\
&\qquad\doteq f_{1,n}(x) + f_{2,n}(x) + f_{3,n}(x))\,.
\end{aligned}$$

This part will be complete if we show that the expectation with respect to $\mathcal{X}$ of these three functions converges to zero. For this, let $\epsilon > 0$ and $\delta \leq \epsilon$. Since η^* is continuous, there exists $r = r(x, \epsilon) > 0$ such that if $d(x, y) < r$ then $|\eta^*(x) - \eta^*(y)| < \epsilon$. On the other hand, since $h_n(x) \to 0$, for that $r(x, \epsilon) > 0$, there exists $N_2 = N_2(x, r(x, \epsilon))$ such that if $n \geq N_2$, $h_n(x) < r$. Then, $f_{1,n}(x) = \frac{1}{\mu\left(\mathcal{B}(x,h_n(x))\right)} \int_{\mathcal{B}(x,h_n(x))} (\eta^*(y) - \eta^*(x))^2 \, d\mu(y) < \epsilon$ for $n \geq N_2$ and in addition it is bounded so, by the dominated convergence theorem we have that

$$\mathbb{E}_{\mathcal{X}}(f_{1,n}(\mathcal{X})) \to 0 \, .$$

For the second term, since $\delta \leq \epsilon$, we have that

$$\mathbb{E}_{\mathcal{X}}(f_{2,n}(\mathcal{X})) = \mathbb{E}_{\mathcal{X}}((\eta(\mathcal{X}) - \eta^*(\mathcal{X}))^2) < \epsilon \, .$$

Finally, since η is bounded,

$$\mathbb{E}_{\mathcal{X}}(f_{3,n}(\mathcal{X})) \lesssim \mathbb{E}_{\mathcal{X}} \left(\frac{1}{\mu(\mathcal{B}(\mathcal{X}, h_n(\mathcal{X})))} \int_{\mathcal{B}(\mathcal{X},h_n(\mathcal{X}))} |\eta(\mathcal{X}) - \eta(y)| \, d\mu(y) \right),$$

which converge to zero if the bounded random variables

$$\frac{1}{\mu(\mathcal{B}(\mathcal{X}, h_n(\mathcal{X})))} \int_{\mathcal{B}(\mathcal{X},h_n(\mathcal{X}))} |\eta(\mathcal{X}) - \eta(y)| \, d\mu(y)$$

converge to zero in probability. To see this, let $\lambda > 0$ be fixed. For every $\delta_0 > 0$,

$$\begin{aligned} &\mathbb{P}_{\mathcal{X}} \left(\frac{1}{\mu(\mathcal{B}(\mathcal{X}, h_n(\mathcal{X})))} \int_{\mathcal{B}(\mathcal{X},h_n(\mathcal{X}))} |\eta(\mathcal{X}) - \eta(y)| \, d\mu(y) > \lambda \right) \\ &\qquad \leq \mathbb{P}_{\mathcal{X}} (h_n(\mathcal{X}) > \delta_0) + \sup_{\delta \leq \delta_0} \mathbb{P}_{\mathcal{X}} \left(\frac{1}{\mu(\mathcal{B}(\mathcal{X}, \delta))} \int_{\mathcal{B}(\mathcal{X},\delta)} |\eta(\mathcal{X}) - \eta(y)| \, d\mu(y) > \lambda \right). \end{aligned}$$

Since $h_n(\mathcal{X}) \to 0$ a.s. the first term converges to zero while the second term does thanks to the truth of the Besicovitch condition (2.4). $\square$

Proof of Theorem 3.1:

Proof of (a): Let us define $\mathcal{D}_n = \{\mathcal{X}_1, ..., \mathcal{X}_n\}$ and $\mathcal{C}_n = \{Y_1, ..., Y_n\}$. In order to prove the mean square consistency, we consider

$$\mathbb{E}\left((\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2)\right) = \mathbb{E}_{\mathcal{X}}\left(\mathbb{E}_{\mathcal{D}_n,\mathcal{C}_n|\mathcal{X}}\left((\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2\right) \big| \mathcal{X}\right)\right).$$

Let $x \in \operatorname{supp}(\mu)$ be fixed. To simplify the notation, we set $\mathbb{E}(\cdot) = \mathbb{E}_{\mathcal{D}_n,\mathcal{C}_n|\mathcal{X}}(\cdot)$. Then, for a particular $h_n(x) \geq h_n^*(x)$ to be defined later, let us define the *theoretical quantities*

$$K\left(\frac{d(x, \mathcal{X}_i)}{h_n(x)}\right) \doteq K_i(x) \doteq K_i \qquad \text{and} \qquad K\left(\frac{d_p(x, \mathcal{X}_i)}{h_{n,p}(x)}\right) \doteq K_{i,p}(x) \doteq K_{i,p} \, ,$$

and, as in (2.3),

$$\frac{K_i}{\sum_{j=1}^n K_j} \doteq W_i \qquad \text{and} \qquad \frac{K_{i,p}}{\sum_{j=1}^n K_{j,p}} \doteq W_{i,p} \, .$$

Let us consider the following auxiliary unobservable quantities:

$$\widehat{\eta}_n(x) = \sum_{i=1}^n W_i Y_i \, , \qquad \eta_n(x) = \sum_{i=1}^n W_i \eta(\mathcal{X}_i) \qquad \text{and} \qquad \eta_{n,p}(x) = \sum_{i=1}^n W_{i,p} \eta(\mathcal{X}_i) \, .$$

Then, we have

$$\begin{aligned}
\widehat{\eta}_{n,p}(x)-\eta(x) &= [\widehat{\eta}_{n,p}(x)-\eta_{n,p}(x)]+[\eta_{n,p}(x)-\eta_n(x)]+[\eta_n(x)-\widehat{\eta}_n(x)]+[\widehat{\eta}_n(x)-\eta(x)] \\
&= \sum_{i=1}^n W_{i,p}(Y_i-\eta(\mathcal{X}_i))+\sum_{i=1}^n (W_{i,p}-W_i)\eta(\mathcal{X}_i)+\sum_{i=1}^n W_i(\eta(\mathcal{X}_i)-Y_i) \\
&\quad + [\widehat{\eta}_n(x)-\eta(x)] \\
&= \sum_{i=1}^n (W_{i,p}-W_i)(Y_i-\eta(\mathcal{X}_i))+\sum_{i=1}^n (W_{i,p}-W_i)\eta(\mathcal{X}_i) \\
&\quad + [\widehat{\eta}_n(x)-\eta(x)]\,.
\end{aligned}$$

Taking squares and expectation in $\mathcal{D}_n, \mathcal{C}_n$, we have

$$\begin{aligned}
\mathbb{E}\left((\widehat{\eta}_{n,p}(x)-\eta(x))^2)\right) &\lesssim \mathbb{E}\left(\left(\sum_{i=1}^n (W_{i,p}-W_i)(Y_i-\eta(\mathcal{X}_i))\right)^2\right) \\
&\quad + \mathbb{E}\left(\left(\sum_{i=1}^n (W_{i,p}-W_i)\eta(\mathcal{X}_i)\right)^2\right) \\
&\quad + \mathbb{E}\left(([\widehat{\eta}_n(x)-\eta(x)])^2\right) \\
&\doteq I + II + III\,.
\end{aligned}$$

By Proposition 2.1 and Remark 2.2 (since $h_n(x)\to 0$ and $h_n(x)\geq h_n^*(x)$), taking expectation on $\mathcal{X}$ we have that term III converges to zero. For the first term, we have

$$\begin{aligned}
I &\approx \mathbb{E}\left(\left(\sum_{i=1}^n (W_{i,p}-W_i)(Y_i-\eta(\mathcal{X}_i))\right)^2\right) \\
&= \mathbb{E}\left(\sum_{i=1}^n\sum_{j=1}^n (W_{i,p}-W_i)(W_{j,p}-W_j)e_ie_j\right) && (Y_i-\eta(\mathcal{X}_i)=e_i) \\
&= \mathbb{E}\left(\sum_{i=1}^n\sum_{j=1}^n (W_{i,p}-W_i)(W_{j,p}-W_j)\mathbb{E}_{\mathcal{C}_n|\mathcal{D}_n}\left(\boldsymbol{e}_i\boldsymbol{e}_j|\mathcal{D}_n\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^n |W_{i,p}-W_i|^2\mathbb{E}_{\mathcal{C}_n|\mathcal{D}_n}\left(\boldsymbol{e}_i^2|\mathcal{D}_n\right)\right) && \text{(cond. ind.)} \\
&= \sigma^2\mathbb{E}\left(\sum_{i=1}^n |W_{i,p}-W_i|^2\right).
\end{aligned}$$

On the other hand, since η is bounded, in II we have

$$II = \mathbb{E}\left(\left(\sum_{i=1}^n (W_{i,p}-W_i)\eta(\mathcal{X}_i)\right)^2\right) \lesssim \mathbb{E}\left(\left(\sum_{i=1}^n |W_{i,p}-W_i|\right)^2\right).$$

We will see that terms I and II converge to zero by splitting the sum in different pieces:

(**1**) $A_1 \doteq \Big\{i\colon d_p(x,\mathcal{X}_i) > h_{n,p}(x), d(x,\mathcal{X}_i) > h_n(x)\Big\}$;

(**2**) $A_2 \doteq \Big\{i\colon d_p(x,\mathcal{X}_i) > h_{n,p}(x), d(x,\mathcal{X}_i) \le h_n(x)\Big\}$;

(**3**) $A_3 \doteq \Big\{i\colon d_p(x,\mathcal{X}_i) \le h_{n,p}(x), d(x,\mathcal{X}_i) > 3h_n(x)\Big\}$;

(**4**) $A_4 \doteq \Big\{i\colon d_p(x,\mathcal{X}_i) \le h_{n,p}(x), d(x,\mathcal{X}_i) \le 3h_n(x)\Big\}$.

Case (**1**) is trivial since in this case K is supported in $[0,1]$ which implies that $W_{i,p} = W_i = 0$. Let us start, therefore, with case (2).

(**2**) Let $\boldsymbol{A_2 \doteq \{i\colon d_p(x,\mathcal{X}_i) > h_{n,p}(x), d(x,\mathcal{X}_i) \le h_n(x)\}}$. Observe that in this case $W_{i,p} = 0$ since K is supported in $[0,1]$. Therefore, since $|W_i| \le 1$ we get

$$I_{A_2} \doteq \mathbb{E}\left(\sum_{i=1}^{n} |W_i|^2 \mathbb{I}_{\{i\in A_2\}}\right) \le \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{I}_{\{i\in A_2\}}\right)$$

and

$$II_{A_2} \doteq \mathbb{E}\left(\left(\sum_{i=1}^{n} |W_i| \mathbb{I}_{\{i\in A_2\}}\right)^2\right) \le \mathbb{E}\left(\left(\sum_{i=1}^{n} \mathbb{I}_{\{i\in A_2\}}\right)^2\right) \doteq C_{A_2}\,. \tag{B.2}$$

Observe that the i.i.d. random variables $\mathbb{I}_{\{i\in A_2\}}$ have a Bernoulli distribution with parameter

$$\begin{aligned} p &= \mathbb{P}_{\mathcal{X}_1}\left(d_p(x,\mathcal{X}_1) > h_{n,p}(x), d(x,\mathcal{X}_1) \le h_n(x)\right) \\ &\le \mathbb{P}_{\mathcal{X}_1}\left(d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1) \ge h_{n,p}(x) - h_n(x)\right) \\ &\le \mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right) \qquad \text{(by H3.2)}\,. \end{aligned}$$

As a consequence, the random variable $Z \doteq \sum_{i=1}^{n} \mathbb{I}_{\{i\in A_2\}}$ has Binomial distribution with parameters n and p and expectation $\mathbb{E}(Z) = np$. This implies that

$$I_{A_2} \lesssim \mathbb{E}(Z) \le n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right), \tag{B.3}$$

and, since $\mathbb{E}\left(Z^2\right) = np(1-p) + n^2p^2 \le np + (np)^2$,

$$\begin{aligned} II_{A_2} \le C_{A_2} \lesssim \mathbb{E}\left(Z^2\right) &\le n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right) \\ &\quad + \Big(n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right)\Big)^2. \end{aligned} \tag{B.4}$$

(**3**) Let $\boldsymbol{A_3 \doteq \{i\colon d_p(x,\mathcal{X}_i) \le h_{n,p}(x), d(x,\mathcal{X}_i) > 3h_n(x)\}}$. Observe that in this case $W_i = 0$ since K is supported in $[0,1]$. Then, since $\forall\, i$, $|W_{i,p}| \le 1$, we get

$$I_{A_3} \doteq \mathbb{E}\left(\sum_{i=1}^{n} |W_{i,p}|^2\, \mathbb{I}_{\{i\in A_3\}}\right) \le \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{I}_{\{i\in A_3\}}\right),$$

and

$$II_{A_3} \doteq \mathbb{E}\left(\left(\sum_{i=1}^{n} |W_{i,p}|\, \mathbb{I}_{\{i\in A_3\}}\right)^2\right) \le \mathbb{E}\left(\left(\sum_{i=1}^{n} \mathbb{I}_{\{i\in A_3\}}\right)^2\right). \tag{B.5}$$

Now, the i.i.d. random variables $\mathbb{I}_{\{i\in A_3\}}$ have Bernoulli distribution with parameter

$$\begin{aligned} p &= \mathbb{P}_{\mathcal{X}_1}\left(d_p(x,\mathcal{X}_1) \le h_{n,p}(x), d(x,\mathcal{X}_1) > 3h_n(x)\right) \\ &\le \mathbb{P}_{\mathcal{X}_1}\left(d(x,\mathcal{X}_1) - d_p(x,\mathcal{X}_1) \ge 3h_n(x) - h_{n,p}(x)\right) . \end{aligned}$$

As a consequence, the random variable $Z \doteq \sum_{i=1}^n \mathbb{I}_{\{i\in A_3\}}$ has Binomial distribution with parameters n and p. But from (H3.1), for n large enough, $h_n(x) \ge \left(\frac{1+C_2}{2}\right) c_{n,p}$ which, together with H3.2 implies that

$$3h_n(x) - h_{n,p}(x) \;\ge\; 2h_n(x) - C_2 c_{n,p} \;\ge\; c_{n,p}\,,$$

and then, for n large enough,

$$p \;\le\; \mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right).$$

Therefore, since $\mathbb{E}\left(Z\right) = np$ we have

$$I_{A_3} \;\lesssim\; \mathbb{E}\left(Z\right) \;\le\; n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p}\right), \tag{B.6}$$

and since $\mathbb{E}\left(Z^2\right) = np(1-p) + n^2p^2 \le np + (np)^2$,

$$\begin{aligned} II_{A_3} \;\lesssim\; \mathbb{E}\left(Z^2\right) \;&\le\; n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p})\right) \\ &\quad + \Big(n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| \ge c_{n,p})\right)\Big)^2 . \end{aligned} \tag{B.7}$$

(4) Let $\boldsymbol{A_4 \doteq \left\{i\colon d_p(x,\mathcal{X}_i) \le h_{n,p}(x), d(x,\mathcal{X}_i) \le 3h_n(x)\right\}}$. In this case we write,

$$\begin{aligned} W_{i,p} - W_i &= \frac{K_{i,p}}{\sum_{j=1}^n K_{j,p}} - \frac{K_i}{\sum_{j=1}^n K_j} \\ &= \frac{K_{i,p}}{\sum_{j=1}^n K_{j,p}} - \frac{K_i}{\sum_{j=1}^n K_{j,p}} + \frac{K_i}{\sum_{j=1}^n K_{j,p}} - \frac{K_i}{\sum_{j=1}^n K_j} \\ &= (K_{i,p} - K_i)\frac{1}{\sum_{j=1}^n K_{j,p}} + K_i \frac{\sum_{j=1}^n (K_j - K_{j,p})}{\sum_{j=1}^n K_j \sum_{j=1}^n K_{j,p}} \\ &= (K_{i,p} - K_i)\frac{1}{\sum_{j=1}^n K_{j,p}} + W_i \frac{\sum_{j=1}^n (K_j - K_{j,p})}{\sum_{j=1}^n K_{j,p}} . \end{aligned}$$

Then,

$$\begin{aligned} I_{A_4} &\doteq \mathbb{E}\left(\sum_{i=1}^n |W_{i,p} - W_i|^2 \mathbb{I}_{\{i\in A_4\}}\right) \\ &\lesssim \mathbb{E}\left(\sum_{i=1}^n |K_{i,p} - K_i|^2 \frac{\mathbb{I}_{\{i\in A_4\}}}{(\sum_{j=1}^n K_{j,p})^2}\right) \\ &\quad + \mathbb{E}\left(\sum_{i=1}^n W_i^2 \mathbb{I}_{\{i\in A_4\}} \left(\frac{\sum_{j=1}^n (K_j - K_{j,p})}{\sum_{j=1}^n K_{j,p}}\right)^2\right) \\ &\lesssim \mathbb{E}\left(\sum_{i=1}^n |K_{i,p} - K_i|^2 \frac{\mathbb{I}_{\{i\in A_4\}}}{(\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j) \le h_{n,p}(x)\}})^2}\right) && (K \text{ regular}) \\ &\quad + \mathbb{E}\left(\left(\frac{\sum_{j=1}^n |K_j - K_{j,p}|}{\sum_{j=1}^n K_{j,p}}\right)^2\right) && \left(|W_i| \le 1,\; \sum_{i=1}^n W_i = 1\right) \\ &\doteq I_{A_4}^1 + I_{A_4}^2 \end{aligned} \tag{B.8}$$

and

$$
\begin{aligned}
II_{A_4} &\doteq \mathbb{E}\left(\left(\sum_{i=1}^{n}|W_{i,p}-W_i|\mathbb{I}_{\{i\in A_4\}}\right)^2\right)\\
&\lesssim \mathbb{E}\left(\left(\sum_{i=1}^{n}|K_{i,p}-K_i|\frac{\mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^{n}K_{j,p}}\right)^2\right)\\
&\quad+\mathbb{E}\left(\left(\sum_{i=1}^{n}W_i\mathbb{I}_{\{i\in A_4\}}\frac{\sum_{j=1}^{n}(K_j-K_{j,p})}{\sum_{j=1}^{n}K_{j,p}}\right)^2\right) \qquad\qquad \text{(B.9)}\\
&\lesssim \mathbb{E}\left(\left(\sum_{i=1}^{n}|K_{i,p}-K_i|\frac{\mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}}\right)^2\right) \qquad (K \text{ regular})\\
&\quad+\mathbb{E}\left(\left(\frac{\sum_{j=1}^{n}|K_j-K_{j,p}|}{\sum_{j=1}^{n}K_{j,p}}\right)^2\right) \qquad (|W_i|\le 1)\\
&\doteq II_{A_4}^1+II_{A_4}^2\,.
\end{aligned}
$$

Observe that if $\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}=0$ then $\forall\, j$, $\mathbb{I}_{\{j\in A_4\}}=0$ so in this case, $I_{A_4}^1$ and $II_{A_4}^1$ are zero. Then, in what follows we will assume that $\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}\neq 0$. Since K is Lipschitz and we are only considering the indexes i such that $d_p(x,\mathcal{X}_i)\le h_{n,p}(x)$, we get

$$
\begin{aligned}
|K_{i,p}-K_i| &= \left|K\left(\frac{d_p(x,\mathcal{X}_i)}{h_{n,p}(x)}\right)-K\left(\frac{d(x,\mathcal{X}_i)}{h_n(x)}\right)\right|\\
&\lesssim \left|\frac{d_p(x,\mathcal{X}_i)}{h_{n,p}(x)}-\frac{d(x,\mathcal{X}_i)}{h_n(x)}\right|\\
&= \frac{|d_p(x,\mathcal{X}_i)h_n(x)-d(x,\mathcal{X}_i)h_{n,p}(x)|}{h_{n,p}(x)h_n(x)}\\
&\le \frac{|d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|}{h_n(x)}+\frac{d_p(x,\mathcal{X}_i)|h_n(x)-h_{n,p}(x)|}{h_n(x)h_{n,p}(x)}\\
&\lesssim \frac{|d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|}{h_n(x)}+\frac{c_{n,p}}{h_n(x)} \qquad (\text{by H3.2})\,.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
I_{A_4}^1 &\lesssim \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\sum_{i=1}^{n}|d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|^2\frac{\mathbb{I}_{\{i\in A_4\}}}{\left(\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}\right)^2}\right)\\
&\quad+\left(\frac{c_{n,p}}{h_n(x)}\right)^2\mathbb{E}\left(\sum_{i=1}^{n}\frac{\mathbb{I}_{\{i\in A_4\}}}{\left(\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}\right)^2}\right) \qquad\qquad \text{(B.10)}\\
&\lesssim \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\sum_{i=1}^{n}|d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|^2\frac{\mathbb{I}_{\{j\in A_4\}}}{\left(\sum_{j=1}^{n}\mathbb{I}_{\{j:d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}\right)^2}\right)\\
&\quad+\left(\frac{c_{n,p}}{h_n(x)}\right)^2
\end{aligned}
$$

and

$$
\begin{aligned}
II^1_{A_4} &\lesssim \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\left(\sum_{i=1}^n |d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|\frac{\mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}}\right)^2\right)\\
&\quad + \left(\frac{c_{n,p}}{h_n(x)}\right)^2\mathbb{E}\left(\left(\sum_{i=1}^n \frac{\mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}}\right)^2\right)\\
&\lesssim \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\left(\sum_{i=1}^n |d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)|\frac{\mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}}\right)^2\right)\\
&\quad + \left(\frac{c_{n,p}}{h_n(x)}\right)^2.
\end{aligned}
\tag{B.11}
$$

(4.1) Let $\boldsymbol{A_{41} \doteq A_4 \cap \{i\colon |d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)| \le c_{n,p}\}}$. In this case, by (H3.1) we get

$$
I^1_{A_{41}} \doteq \frac{c^2_{n,p}}{h_n^2(x)}\,\mathbb{E}\left(\frac{\sum_{i=1}^n \mathbb{I}_{\{i\in A_4\}}}{\left(\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}\right)^2}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 \lesssim \left(\frac{c_{n,p}}{h_n(x)}\right)^2
\tag{B.12}
$$

and

$$
II^1_{A_{41}} \doteq \frac{c^2_{n,p}}{h_n^2(x)}\,\mathbb{E}\left(\left(\frac{\sum_{i=1}^n \mathbb{I}_{\{i\in A_4\}}}{\sum_{j=1}^n \mathbb{I}_{\{j: d_p(x,\mathcal{X}_j)\le h_{n,p}(x)\}}}\right)^2\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 \lesssim \left(\frac{c_{n,p}}{h_n(x)}\right)^2.
\tag{B.13}
$$

(4.2) Let $\boldsymbol{A_{42} \doteq A_4 \cap \{i\colon |d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)| > c_{n,p}\}}$. Let us define the i.i.d. random variables $Z_i \doteq d_p(x,\mathcal{X}_i)-d(x,\mathcal{X}_i)$, $i=1,\dots,n$. Since $d_p(x,\mathcal{X}_i)\le h_{n,p}(x)$ and $d(x,\mathcal{X}_i)\le 3h_n(x)$ we have that $|Z_i|\le h_{n,p}(x)+3h_n(x)$. Observe that, from (H3.2) and (H3.1), respectively, for n large enough we have

$$
h_{n,p} \le h_n(x) + C_2 c_{n,p} \le C h_n(x)\,.
$$

Which implies that, for n large enough, $|Z_i|\le C h_n(x)$. Therefore,

$$
\begin{aligned}
I^1_{A_{42}} &\doteq \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\sum_{i=1}^n |Z_i|^2 \mathbb{I}_{\{i: c_{n,p}\le |Z_i|\le C h_n(x)\}}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2\\
&\le \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\sum_{i=1}^n |Z_i|^2 \mathbb{I}_{\{i: c_{n,p}\le |Z_i|\le C h_n(x)\}}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2\\
&\le \frac{n}{h_n^2(x)}\,\mathbb{E}\left(|Z_1|^2 \mathbb{I}_{\{c_{n,p}\le |Z_1|\le C h_n(x)\}}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 && (\#A_{42}\le n)\\
&\lesssim \frac{n}{h_n(x)}\,\mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p}\le |Z_1|\le C h_n(x)\}}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 && (|Z_1|\lesssim h_n(x))\,.
\end{aligned}
\tag{B.14}
$$

On the other hand,

$$
\begin{aligned}
II^1_{A_{42}} &\doteq \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\left(\sum_{i=1}^n |Z_i| \mathbb{I}_{\{i: c_{n,p}\le |Z_i|\le C h_n(x)\}}\right)^2\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2\\
&\le \frac{1}{h_n^2(x)}\,\mathbb{E}\left(\left(\sum_{i=1}^n |Z_i| \mathbb{I}_{\{i: c_{n,p}\le |Z_i|\le C h_n(x)\}}\right)^2\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2.
\end{aligned}
\tag{B.15}
$$

Observe that, for $i \neq j$, Z_i is independent of Z_j, then

$$
\begin{aligned}
&\mathbb{E}\left(\left(\sum_{i=1}^{n} |Z_i| \mathbb{I}_{\{i c_{n,p} \le |Z_i| \le C h_n(x)\}}\right)^2\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} |Z_i||Z_j| \mathbb{I}_{\{i c_{n,p} \le |Z_i| \le C h_n(x)\}} \mathbb{I}_{\{j c_{n,p} \le |Z_j| \le C h_n(x)\}}\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{n} |Z_i|^2 \mathbb{I}_{\{i c_{n,p} \le |Z_i| \le C h_n(x)\}}\right) \\
&\quad + \mathbb{E}\left(\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} |Z_i||Z_j| \mathbb{I}_{\{i c_{n,p} \le |Z_i| \le C h_n(x)\}} \mathbb{I}_{\{j c_{n,p} \le |Z_j| \le C h_n(x)\}}\right) \\
&\le n\mathbb{E}\left(|Z_1|^2 \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) + n^2 \mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) \mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) \\
&\lesssim n h_n(x) \mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) + n^2 \left(\mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right)\right)^2 \qquad \left(|Z_1| \lesssim h_n(x)\right).
\end{aligned}
$$

Using this bound in (B.15), we get

$$
\begin{aligned}
II^1_{A42} \lesssim\ & \frac{n}{h_n(x)}\, \mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) \\
&+ \frac{n^2}{h_n^2(x)} \left(\mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right)\right)^2 + \left(\frac{c_{n,p}}{h_n(x)}\right)^2.
\end{aligned} \tag{B.16}
$$

We need to compute the expectation $\mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right)$, which is

$$
\begin{aligned}
\mathbb{E}\left(|Z_1| \mathbb{I}_{\{c_{n,p} \le |Z_1| \le C h_n(x)\}}\right) &= \int_{c_{n,p}}^{h_n(x)} \mathbb{P}\left(|Z_1| > t\right) dt \\
&\le \mathbb{P}\left(|Z_1| > c_{n,p}\right) \int_{c_{n,p}}^{h_n(x)} dt \\
&\le \mathbb{P}\left(|Z_1| > c_{n,p}\right) h_n(x).
\end{aligned}
$$

Therefore, with this inequality in (B.14), we have

$$
\begin{aligned}
I^1_{A_{42}} &\lesssim n\mathbb{P}\left(|Z_1| > c_{n,p}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 \\
&= n\mathbb{P}\left(|d_p(x, \mathcal{X}_1) - d(x, \mathcal{X}_1)| > c_{n,p}\right) + \left(\frac{c_{n,p}}{h_n(x)}\right)^2
\end{aligned} \tag{B.17}
$$

and, with the same inequality in (B.16),

$$
\begin{aligned}
II^1_{A_{42}} &\lesssim n\mathbb{P}\left(|Z_1| > c_{n,p}\right) + \left(n\mathbb{P}\left(|Z_1| > c_{n,p}\right)\right)^2 + \left(\frac{c_{n,p}}{h_n(x)}\right)^2 \\
&= n\mathbb{P}\left(|d_p(x, \mathcal{X}_1) - d(x, \mathcal{X}_1)| > c_{n,p}\right) \\
&\quad + \left(n\mathbb{P}\left(|d_p(x, \mathcal{X}_1) - d(x, \mathcal{X}_1)| > c_{n,p}\right)\right)^2 + \left(\frac{c_{n,p}}{h_n(x)}\right)^2.
\end{aligned} \tag{B.18}
$$

Then, with (B.12) and (B.17) in (B.10) we get

$$I^1_{A_4} \lesssim \left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) \tag{B.19}$$

and, with (B.13) and (B.18) in (B.11),

$$\begin{aligned} II^1_{A_4} \lesssim & \left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) \\ & + \Big(n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right)\Big)^2. \end{aligned} \tag{B.20}$$

On the other hand, observe that $I^2_{A_4} = \mathbb{E}\left(\left(\frac{\sum_{j=1}^n |K_j - K_{j,p}|}{\sum_{j=1}^n K_{j,p}}\right)^2\right)$. Since $A_4^c = \{j : d(x,\mathcal{X}_j) > 3h_n(x)\} \cup \{j : d_p(x,\mathcal{X}_j) > h_{n,p}(x)\}$, we can write

$$\begin{aligned} \frac{\sum_{j=1}^n |K_j - K_{j,p}|}{\sum_{j=1}^n K_{j,p}} \leq & \frac{\sum_{j=1}^n |K_j - K_{j,p}| \mathbb{I}_{\{j \in A_4\}}}{\sum_{j=1}^n K_{j,p}} \\ & + \frac{\sum_{j=1}^n |K_j - K_{j,p}| \mathbb{I}_{\{j : d(x,\mathcal{X}_j) > 3h_n(x)\}}}{\sum_{j=1}^n K_{j,p}} \\ & + \frac{\sum_{j=1}^n |K_j - K_{j,p}| \mathbb{I}_{\{j : d_p(x,\mathcal{X}_j) > h_{n,p}(x)\}}}{\sum_{j=1}^n K_{j,p}}. \end{aligned}$$

Using that K is regular and that $\sum_{j=1}^n K_{j,p} \geq 1$ (this is since $\{j : d_p(x,\mathcal{X}_j) \leq h_{n,p}(x)\} \neq \emptyset$), we get

$$\begin{aligned} I^2_{A_4} &= \mathbb{E}\left(\left(\frac{\sum_{j=1}^n |K_j - K_{j,p}|}{\sum_{j=1}^n K_{j,p}}\right)^2\right) \\ &\lesssim II^1_{A_4} + \mathbb{E}\left(\left(\sum_{j=1}^n |W_{j,p}| \mathbb{I}_{\{j : d_p(x,\mathcal{X}_j) \leq h_{n,p}(x), d(x,\mathcal{X}_j) > 3h_n(x)\}}\right)^2\right) \\ &\quad + \frac{\sum_{j=1}^n K_j \mathbb{I}_{\{j : d_p(x,\mathcal{X}_j) > h_{n,p}(x)\}}}{\sum_{j=1}^n K_{j,p}} \\ &\lesssim II^1_{A_4} + II_{A_3} + \mathbb{E}\left(\left(\sum_{j=1}^n \mathbb{I}_{\{j : d_p(x,\mathcal{X}_j) > h_{n,p}(x), d(x,\mathcal{X}_j) \leq h_n(x)\}}\right)^2\right) \\ &\leq II^1_{A_4} + II_{A_3} + C_{A_2}, \end{aligned}$$

where $II^1_{A_4}$ was defined in (B.9), II_{A_3} in (B.5), and C_{A_2} in (B.2). Then, from (B.20), (B.7) and (B.4), we have

$$\begin{aligned} I^2_{A_4} \lesssim & \left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) \\ & + \Big(n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right)\Big)^2. \end{aligned} \tag{B.21}$$

Therefore, with (B.19) and (B.21) in (B.8) we have

$$I_{A_4} \lesssim \left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) + \Big(n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right)\Big)^2, \tag{B.22}$$

and with (B.20) and (B.21) in (B.9),

$$II_{A_4} \lesssim \left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) + \Big(n\mathbb{P}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right)\Big)^2. \tag{B.23}$$

Finally, to complete the proof of this result (i.e. that I and II converge to zero) we need to show that the expectation on $\mathcal{X}$ of

$$\left(\frac{c_{n,p}}{h_n(x)}\right)^2 + n\mathbb{P}_{\mathcal{X}_1}\left(|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p}\right) + (n\mathbb{P}^2_{\mathcal{X}_1}|d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1)| > c_{n,p})$$

converges to zero. In order to show it, recall that from H2 we have

$$n^2\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_1|\mathcal{X}}\big(|d_p(\mathcal{X},\mathcal{X}_1) - d(\mathcal{X},\mathcal{X}_1)| \geq c_{n,p}\big)\big|\mathcal{X} \in \operatorname{supp}(\mu)\right) \to 0\,,$$

and consequently, by Cauchy Schwartz inequality,

$$n\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}_{\mathcal{X}_1|\mathcal{X}}\left(|d_p(\mathcal{X},\mathcal{X}_1) - d(\mathcal{X},\mathcal{X}_1)| \geq c_{n,p}\right)\big|\mathcal{X} \in \operatorname{supp}(\mu)\right) \to 0\,.$$

In addition, from (H3.1) we have

$$\mathbb{E}_{\mathcal{X}}\left(\left(\frac{c_{n,p}}{h_n(\mathcal{X})}\right)^2\right) \to 0\,.$$

Therefore, taking expectation with respect to $\mathcal{X}$ in (B.3), (B.4), (B.6), (B.7), (B.22) and (B.23), we prove Part (a) of the Theorem.

Proof of (b): The only difference with item (a) is the convergence of term III to zero which is ensured by Proposition 2.2. □

Proof of Theorem 3.2: Let $\gamma_n \to \infty$ as $n \to \infty$ a sequence such that, as $n, p \to \infty$, $\mathbb{E}_{\mathcal{X}}\left(\gamma_n\left(\frac{c_{n,p}}{h_n(\mathcal{X})}\right)^2\right) \to 0$ and, for each $i = 1, \ldots, n$,

$$\gamma_n n^2\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}^2_{\mathcal{X}_i|\mathcal{X}}\big(\,|d(\mathcal{X},\mathcal{X}_i) - d_p(\mathcal{X},\mathcal{X}_i)| \geq c_{n,p}\Big|\mathcal{X} \in \operatorname{supp}(\mu)\big)\right) \to 0\,.$$

From proof of Theorem 3.1 we get

$$\mathbb{E}\left(\gamma_n(\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2\right) \lesssim \gamma_n n\mathbb{E}_{\mathcal{X}}\left(\mathbb{P}_{\mathcal{X}_1}\left(d_p(x,\mathcal{X}_1) - d(x,\mathcal{X}_1) \geq c_{n,p}\right)\right) + \mathbb{E}_{\mathcal{X}}\left(\gamma_n\left(\frac{c_{n,p}}{h_n(\mathcal{X})}\right)^2\right) + \mathbb{E}\left(\gamma_n(\widehat{\eta}_n(\mathcal{X}) - \eta(\mathcal{X}))^2\right),$$

from what follows that

$$\lim_{n,p\to\infty} \mathbb{E}\left(\gamma_n(\widehat{\eta}_{n,p}(\mathcal{X}) - \eta(\mathcal{X}))^2\right) = 0\,.$$ □

ACKNOWLEDGMENTS

The authors would like to thank one of the anonymous referees and the Associate Editor for their constructive comments which improved the present version of the paper.

REFERENCES

[1] Abraham, C.; Biau, G. and Cadre, B. (2006). On the kernel rule for function classification, *Ann. Inst. Statist. Math.*, **58**, 619–633.

[2] Amiri, A,; Crambes, C. and Thiam, B. (2014). Recursive estimation of nonparametric regression with functional covariate, *Comput. Statist. Data Anal.*, **69**, 154–172.

[3] Biau, G.; Bunea, F. and Wegkamp, M.H. (2005). Functional classification in Hilbert spaces, *IEEE Trans. Inf. Theory*, **51**, 2163–2172.

[4] Biau, G.; Cérou, F. and Guyader, A. (2010). Rates of convergence of the functional k-nearest neighbor estimate, *IEEE Trans. Inform. Theory*, **56**, 2034–2040.

[5] Bigot, J. (2006). Landmark-based registration of curves via the continuous wavelet transform, *J. Comput. Graph. Statist.*, **15**, 542–564.

[6] Brezis, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer-Verlag, Berlin.

[7] Burba, F.; Ferraty, F. and Vieu, P. (2009). k-nearest neighbor method in functional nonparametric regression, *Journal of Nonparametric Statistics*, **21**, 453–469.

[8] Cai, T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition, *The Annals of Statistics*, **39**, 2330–2355.

[9] Cai, T. and Yuan, M. (2016). Minimax and Adaptive Estimation of Covariance Operator for Random Variables Observed on a Lattice Graph, *J. Amer. Statist. Assoc.*, **39**, 2330–2355.

[10] Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension, *ESAIM Probab. Stat.*, **10**, 340–355.

[11] Chagny, G. and Roche, A. (2014). Adaptive and minimax estimation of the cumulative distribution function given a functional covariate, *Electron. J. Stat.*, **8**, 2352–2404.

[12] Collomb, G. (1980). Estimation de la regression par la méthode des k points les plus proches avec noyau: Quelques propitétés de convergence ponctuelle, *Lectures Notes in Mathematics*, **821**, 159–175, Springer-Verlag, Berlin.

[13] Ferraty, F. and Romain, Y., Eds. (2011). *The Oxford Handbook of Functional Data Analysis*, Oxford University Press.

[14] Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data, *Comput. Statist.*, **17**(4), 545–564.

[15] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice.* Springer-Verlag, Berlin.

[16] Forzani, L.; Fraiman, R. and Llop, P. (2012). Consistent nonparametric regression for functional data under the Stone–Besicovitch conditions, *IEEE Trans. Inform. Theory*, **58**, 6697–6708.

[17] Forzani, L.; Fraiman, R. and Llop, P. (2014). Corrigendum to consistent nonparametric regression for functional data under the Stone–Besicovitch conditions, *IEEE Trans. Inform. Theory*, **60**, 3069.

[18] Hall, P.; Müller, H.G. and Wang, J.L. (2006). Properties of principal component methods for functional and longitudinal data analysis, *The Annals of Statistics*, **34**(3), 1493–1517.

[19] Hart, J.D. and Wherly, T.E. (1986). Kernel regression estimation using repeated measurements data, *J. Amer. Statist. Assoc.*, **81**, 1080–1088.

[20] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman and Hall, London.

[21] Kneip, A. and Ramsay, J.O. (2008). Combining registration and fitting for functional models, *J. Amer. Statist. Assoc.*, **103**, 1155–1165.

[22] Lian, H. (2011). Convergence of functional k-nearest neighbor regression estimate with functional responses, *Electronic Journal of Statistics*, **5**, 31–40.

[23] Müller, S. (2011). *Consistency and bandwidth selection for dependent data in non-parametric functional data analysis*, Ph.D. Thesis, Faculty of Mathematics and Physics, University of Stuttgart, Germany.

[24] Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*, McGraw-Hill, New York.

[25] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis. Methods and Case Studies*. Springer-Verlag, New York.

[26] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed., Springer-Verlag, New York.

[27] Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves, *J. Roy. Statist. Soc. Ser. B*, **53**, 233–243.