**Electronic Journal of Statistics**
Vol. 8 (2014) 2352–2404
ISSN: 1935-7524
DOI: 10.1214/14-EJS956

# Adaptive and minimax estimation of the cumulative distribution function given a functional covariate

**Gaëlle Chagny**

*LMRS, UMR CNRS 6085, Université de Rouen, France*
*e-mail:* gaelle.chagny@gmail.com

**and**

**Angelina Roche**

*I3M, UMR CNRS 5149, Université Montpellier II, France*
*e-mail:* angelina.roche@univ-montp2.fr

**Abstract:** We consider the nonparametric kernel estimation of the conditional cumulative distribution function given a functional covariate. Given the bias-variance trade-off of the risk, we first propose a totally data-driven bandwidth selection mechanism in the spirit of the recent Goldenshluger-Lepski method and of model selection tools. The resulting estimator is shown to be adaptive and minimax optimal: we establish nonasymptotic risk bounds and compute rates of convergence under various assumptions on the decay of the small ball probability of the functional variable. We also prove lower bounds. Both pointwise and integrated criteria are considered. Finally, the choice of the norm or semi-norm involved in the definition of the estimator is also discussed, as well as the projection of the data on finite dimensional subspaces. Numerical results illustrate the method.

**MSC 2010 subject classifications:** Primary 62G05; secondary 62H12.
**Keywords and phrases:** Adaptive kernel estimator, conditional cumulative distribution function, minimax estimation, functional random variable, small ball probability.

Received February 2014.

## Contents

## 1. Introduction

The aim of Functional Data Analysis (FDA) is to analyse information on curves or functions. This field has attracted a lot of attention over the past decades, thanks to its numerous applications. We refer to Ramsay and Silverman [54], Ferraty and Vieu [33] for case studies and Ferraty and Romain [31] for a recent overview. In many situations, it is of interest to understand the link between a scalar quantity $Y$ and a functional random variable $X$. For instance, it has been pointed out (Glaser et al. [38, 39]) that the probability of occurrence of cerebral edema in children treated with diabetic ketoascidosis depends on the evolution over time of the drugs and fluids administred (and not only on their quantity). In another application, the focus is in explaining the link between the chemical composition of a sample of material and its spectrometric curve (such a problem related to food industry will be illustrated in the numerical study).

We suppose that the random variable $X$ takes values in a separable infinite-dimensional Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$. The latter can be $\mathbb{L}^2(I)$, the set of squared-integrable functions on a subset $I$ of $\mathbb{R}$, or a Sobolev space. The link between the predictor $X$ and the response $Y$ is classically described by regression analysis. However, this can also be achieved by estimating the entire conditional distribution of the variable $Y$ given $X$. The target function we want to recover is the conditional cumulative distribution function $F$ (denoted by conditional c.d.f. in the sequel) of $Y$ given $X$ defined by

$$F^x(y) := \mathbb{P}(Y \leq y | X = x), \quad (x, y) \in \mathbb{H} \times \mathbb{R}. \tag{1}$$

To estimate it, we have access to a data sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ distributed like the couple $(X, Y)$.

In the sequel, we consider kernel estimators similar to the ones defined by Ferraty, Laksaci and Vieu [29], Ferraty et al. [34], for which we provide a detailed nonasymptotic adaptive and minimax study.

The pioneering works on conditional distribution when the covariate is functional are the one of Ferraty and Vieu [32], Ferraty, Laksaci and Vieu [29], completed by Ferraty et al. [34]. Kernel estimators, which depend on a smoothing parameter, the so-called bandwidth, are built to address several estimation problems: regression function, conditional c.d.f., conditional density and its derivatives, conditional hazard rate, conditional mode and quantiles. A lot of research has then been carried out to extend or adapt the previous procedures to various statistical models. For instance, the estimation of the regression function is studied by Rachdi and Vieu [53], Ferraty, Mas and Vieu [30], Dabo-Niang and

Rhomari [24]. The case of dependent data is the subject of the works of Masry [51], Aspirot, Bertin and Perera [5], Laib and Louani [47], Dabo-Niang, Kaid and Laksaci [23] under several assumptions ($\alpha$-mixing, ergodic or non-stationary processes). Demongeot et al. [27] consider local-linear estimators of the conditional density and conditional mode. Robust versions of the previous strategies are proposed by Crambes, Delsol and Laksaci [21], Azzedine, Laksaci and Ould-Saïd [6], Gheriballah, Laksaci and Sekkal [36]. Gijbels, Omelka and Veraverbeke [37] investigate the estimation of the dependence between two variables conditionally to a functional covariate through copula modelling. Most of this literature focuses on asymptotic results (almost-complete convergence, asymptotic normality,...). Bias-variance decompositions are provided. Few papers tackle the problem of bandwidth selection: Rachdi and Vieu [53] and Benhenni et al. [7] have studied global or local cross-validation procedures which are shown to be asymptotically optimal in regression contexts. Recently, a Bayesian criterion has been investigated from a numerical point of view by Shang [55].

To our knowledge, adaptive estimation procedures in a nonasymptotic framework can only be found in conditional distribution estimation with real or multivariate covariates. We refer to Brunel, Comte and Lacour [10] and Plancade [52] for c.d.f estimation with a real covariate and to Akakpo and Lacour [2] and references therein for conditional density estimation with a multivariate covariate. Nevertheless, these works are based on projection estimators which cannot be extended directly to a functional framework in a nonparametric setting.

In keeping with the studies of functional conditional distribution, we investigate the properties of the nonparametric Nadaraya-Watson-type estimators of Ferraty, Laksaci and Vieu [29], but with a new perspective, only used so far for real and multivariate covariates. To estimate the c.d.f. defined by (1), we consider

$$\widehat{F}_h^x(y) := \sum_{i=1}^{n} W_h^{(i)}(x)\mathbf{1}_{\{Y_i \le y\}} \text{ where } W_h^{(i)}(x) := \frac{K_h(d(X_i,x))}{\sum_{j=1}^{n} K_h(d(X_j,x))}, \tag{2}$$

for any $(x,y) \in \mathbb{H} \times \mathbb{R}$, with $d$ a general semi-metric on the Hilbert space $\mathbb{H}$, $K_h : t \mapsto K(t/h)/h$, for $K$ a kernel function (that is $\int_{\mathbb{R}} K(t)dt = 1$) and $h$ a parameter to be chosen, the so-called bandwidth. We choose the kernel $K$ of type I (Ferraty and Vieu 33): its support is in $[0,1]$ and there exist two constants $c_K, C_K > 0$ such that $c_K \mathbf{1}_{[0,1]} \le K \le C_K \mathbf{1}_{[0,1]}$. We focus on the metric associated to the norm of the Hilbert space

$$d(x,x') := \|x - x'\|, \quad x, x' \in \mathbb{H}. \tag{3}$$

The main goal is to define a fully data-driven selection rule for the bandwidth $h$, which satisfies nonasymptotic adaptive results. The criterion we propose draws inspiration from both the so-called Lepski method (see the recent paper of Goldenshluger and Lepski 40) and model selection tools. We show that the bias-variance trade-off is realized and that the selected estimator automatically adapts to the unknown regularity of the target function. As usual, the variance

term of the risk depends on asymptotic properties of the small ball probability $\varphi(h) = \mathbb{P}(d(X,0) \leq h)$ when $h \to 0$. The behaviour of the small ball probability is a difficult problem which is still the subject of research studies. We compute precise rates for our estimator under several assumptions on the distribution of the process $X$, fulfilled e.g. by a large class of Gaussian processes. Consistently with the previous works, the rates we obtain are quite slow. However, we prove that they are minimax optimal. The results are also shown to be coherent with lower bounds computed by Mas [50] for the estimation of the regression function.

To bypass the difficulties inherent to the infinite dimensional nature of the data, some researchers (see e.g. Masry [51], Ferraty, Laksaci and Vieu [29], Geenens [35]) have suggested replacing the norm $\|\cdot\|$ in the definition of the estimator (2) by a semi-norm. The case of projection semi-norms has received particular attention. In that case the estimator can be redefined this way

$$\widehat{F}^x_{h,p}(y) := \sum_{i=1}^n W^{(i)}_{h,p}(x)\mathbf{1}_{\{Y_i \leq y\}} \text{ with } W^{(i)}_{h,p}(x) := \frac{K_h(d_p(X_i,x))}{\sum_{j=1}^n K_h(d_p(X_j,x))}, \tag{4}$$

where $d_p^2(x,x') := \sum_{j=1}^p \langle x - x', e_j\rangle^2$ and $(e_j)_{j\geq 1}$ is a basis of $\mathbb{H}$. Defining this estimator amounts to project the data into a $p$-dimensional space. We show that it does not improve the convergence rates of the Nadaraya-Watson estimator since the lower bounds are still valid. In order to understand what is going on, we briefly study a bias-variance decomposition of the risk of this estimator.

The paper is organized as follows: in Section 2, we provide a bias-variance decomposition of the estimator (2) in terms of two criteria, a pointwise and an integrated risk. The bandwidth $h$ is shown to influence significantly the quality of estimation. In Section 3, we define a bandwidth selection criterion achieving the best bias-variance trade-off. Rates of convergence of the resulting estimator are computed in Section 4. To ensure that these rates are optimal, we also prove lower bounds. Properties of the estimator defined with a projection semi-metric are investigated in Section 5. The results are illustrated via simulations and examples in Section 6. Finally, the proofs are presented in Section 7, and some details are postponed to the Appendix (Section A).

## 2. Integrated and pointwise risk of an estimator with fixed bandwidth

### 2.1. Considered risks

We consider two types of risks for the estimation of $(x,y) \mapsto F^x(y)$. Both are mean integrated squared error with respect to the response variable $y$.

The first criterion is a *pointwise risk* in $x$, integrated in $y$:

$$\mathbb{E}\left[\|\widehat{F}^{x_0}_h - F^{x_0}\|^2_D\right],$$

for a fixed $x_0 \in \mathbb{H}$, $D$ a compact subset of $\mathbb{R}$ and

$$\|f\|_D^2 := \int_D f(t)^2 dt,$$

keeping in mind that the Hilbert norm of $\mathbb{H}$ is $\|.\|$. We also denote by $|D| := \int_D dt$ the Lebesgue measure of the set $D$.

Next, we introduce a second criterion, which is an *integrated risk* with respect to the product of the Lebesgue measure on $\mathbb{R}$ and the probability measure $\mathbb{P}_X$ of $X$, defined by

$$\mathbb{E}\left[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')\right] = \mathbb{E}\left[\int_B \left(\int_D \left(\widehat{F}_h^x(y) - F^x(y)\right)^2 dy\right) d\mathbb{P}_X(x)\right], \tag{5}$$

where $X'$ is a copy of $X$ independent of the data sample and $B$ is a subset of $\mathbb{H}$.

The motivation for studying the two risks is twofold. First, in practice, we can either be interested in the estimation of $F^{X_{n+1}}$ where $X_{n+1}$ is a copy of $X$ independent of the sample or we can be interested in estimating the c.d.f conditionally to $X = x_0$ where $x_0$ is a point chosen in advance. Such an approach is rather classical in functional linear regression (Ramsay and Silverman [54], Cardot, Ferraty and Sarda [14]) where either prediction error on random curves (Crambes, Kneip and Sarda [22]) or prediction error over a fixed curve (Cai and Hall [13]) are considered. Second, integrated risks have been relatively unexplored in nonparametric functional data analysis. Indeed, there is no measure universally accepted as the Lebesgue measure in finite-dimensional setting (see e.g. Delaigle and Hall [26], Dabo-Niang and Yao [25]). The only measure at hand is the probability measure of $X$.

### *2.2. Assumptions*

Hereafter, we denote by $\varphi^x$ the shifted small ball probability:

$$\varphi^x(h) = \mathbb{P}(\|X - x\| \leq h), \quad h > 0, \quad x \in \mathbb{H}.$$

We write $\varphi(h)$ instead of $\varphi^0(h)$. If $X'$ is a random variable, $\varphi^{X'}$ is the conditional small ball probability: $\varphi^{X'}(h) = \mathbb{P}_{X'}(\|X - X'\| \leq h)$, where hereafter the notation $\mathbb{P}_{X'}$ (resp. $\mathbb{E}_{X'}$, $\mathrm{Var}_{X'}$) stands for the conditional probability (resp. expectation, variance) given $X'$. For simplicity, we assume that the curve $X$ is centred. We also consider the following assumptions. The first one is related to the choice of the kernel, the two following are regularity assumptions for the function to estimate and the process $X$.

$H_F$ There exists $\beta \in (0;1)$ such that $F$ belongs to the functional space $\mathcal{F}_\beta$, the class of the maps $(x,y) \in \mathbb{H} \times \mathbb{R} \mapsto F^x(y)$ such that:

- for all $x \in \mathbb{H}$, $F^x$ is a c.d.f;
- there exists a constant $C_D > 0$ such that, for all $x, x' \in \mathbb{H}$

$$\|F^x - F^{x'}\|_D \leq C_D \|x - x'\|^\beta.$$

$H_\varphi$ There exist two constants $c_\varphi, C_\varphi > 0$ such that for all $h \in \mathbb{R}$,

$$c_\varphi \varphi(h) \mathbf{1}_B(X') \leq \varphi^{X'}(h) \mathbf{1}_B(X') \leq C_\varphi \varphi(h) \mathbf{1}_B(X') \text{ a.s.},$$

where $X'$ is an independent copy of $X$.

Assumption $H_F$ is an Hölder-type regularity condition on the map $x \mapsto F^x$. This type of condition is natural in kernel estimation. It is very similar to Assumption (H2) of Ferraty, Laksaci and Vieu [29] or Assumption (H2') of Ferraty et al. [34]. Note, however, that no regularity condition on the map $y \mapsto F^x(y)$ is required here. A similar phenomenon appears for the estimation of the c.d.f when the covariate is real: for instance, the convergence rate given by Brunel, Comte and Lacour [10, Corollary 1] only depends on the regularity of $F$ with respect to $x$.

Assumption $H_\varphi$ is very similar to assumptions made by Ferraty, Laksaci and Vieu [29], Burba, Ferraty and Vieu [12], Ferraty et al. [34]. This condition $H_\varphi$ is reasonable, since the class of Gaussian processes fulfills it provided that $B$ is a bounded subset of $\mathbb{H}$. Indeed the upper bound is verified with $C_\varphi = 1$ thanks to Anderson's Inequality (see Anderson [3] and also Li and Shao [48, Theorem 2.13] or Hoffmann-Jørgensen, Shepp and Dudley [42, Theorem 2.1, p. 322]) and from Hoffmann-Jørgensen, Shepp and Dudley [42, Theorem 2.1, p. 322] we know that the lower bound is verified with $c_\varphi := e^{-R^2/2}$ where $R := \max\{\|x\|,\ x \in B\}$.

### 2.3. Upper bound

Under the assumptions above we are able to obtain a nonasymptotic upper bound for the risk, proved in Section 7.2.

**Theorem 1.** *Suppose that Assumption $H_F$ is fulfilled. Let $h > 0$ be fixed.*

*(i) For all $x_0 \in \mathbb{H}$ we have*

$$\mathbb{E}\left[\left\|\widehat{F}_h^{x_0} - F^{x_0}\right\|_D^2\right] \leq C\left(h^{2\beta} + \frac{1}{n\varphi^{x_0}(h)}\right), \tag{6}$$

*where $C > 0$ only depends on $c_K$, $C_K$, $|D|$ and $C_D$.*

*(ii) If, in addition, Assumption $H_\varphi$ is fulfilled,*

$$\mathbb{E}\left[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')\right] \leq C\left(h^{2\beta} + \frac{1}{n\varphi(h)}\right), \tag{7}$$

*where $C > 0$ only depends on $c_K$, $C_K$, $c_\varphi$, $C_\varphi$, $|D|$ and $C_D$.*

The first term of the right-hand-side of inequalities (6) and (7) corresponds to a bias term, and the second is a variance term, which increases when $h$ goes to 0 (since $\varphi^{x_0}(h)$ and $\varphi(h)$ decrease to 0 when $h \to 0$). Note that the upper bounds are very similar to the results of Ferraty, Laksaci and Vieu [29, Theorem 3.1] and Ferraty et al. [34, Corollary 3]. However, we do not have an extra-$\ln n$ factor in the variance term.

We deduce from Theorem 1 that the usual bias-variance trade-off must be done if one wants to choose $h$ in a family of possible bandwidths. The ideal compromise $h^*$ is called the oracle, and is defined by

$$h^* = \arg\min_h \mathbb{E}\left[\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right]. \tag{8}$$

It cannot be used as an estimator since it both depends on the unknown regularity index $\beta$ of $F$ and on the rate of decrease of the small ball probability $\varphi(h)$ of $X$ to 0. The challenge is to propose a fully data-driven method to perform the trade-off.

## 3. Adaptive estimation

In this section, we focus on the integrated risk. We refer to Remark 1 below for the extension of the results for the pointwise criterion.

### *3.1. Bandwidth selection*

We have at our disposal the estimators $\widehat{F}_h$ defined by (2) for any $h > 0$. Let $\mathcal{H}_n$ be a finite collection of bandwidths, with cardinality depending on $n$ and properties precised below. For any $h \in \mathcal{H}_n$, an empirical version for the small ball probability $\varphi(h) = \mathbb{P}(\|X\| \leq h)$ is

$$\widehat{\varphi}(h) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\|X_i\| \leq h\}}. \tag{9}$$

For any $h \in \mathcal{H}_n$, we define

$$\widehat{A}(h) = \max_{h' \in \mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 - \widehat{V}(h')\right)_+, \quad \widehat{V}(h) = \begin{cases} \kappa\dfrac{\ln(n)}{n\widehat{\varphi}(h)} \text{ if } \widehat{\varphi}(h) \neq 0 \\ +\infty \text{ otherwise,} \end{cases} \tag{10}$$

where $\kappa$ is a constant specified in the proofs which depends neither on $h$, nor on $n$, nor on $F^{X'}$. The quantity $\widehat{V}(h)$ is an estimator of the upper bound for the variance term (see (7)) and $\widehat{A}(h)$ is proved to be an approximation of the bias term (see Lemma 5). This motivates the following choice of the bandwidth:

$$\widehat{h} = \operatorname{argmin}_{h \in \mathcal{H}_n}\left\{\widehat{A}(h) + \widehat{V}(h)\right\}. \tag{11}$$

The selected estimator is $\widehat{F}_{\widehat{h}}$.

This selection rule is inspired both by the recent version of the so-called Lepski method (see Goldenshluger and Lepski [40]) and by model selection tools. The main idea is to estimate the bias term by looking at several estimators. Goldenshluger and Lepski [40] propose to first define "intermediate" estimators $\widehat{F}_{h,h'}^{X'}$ $(h, h' \in \mathcal{H}_n)$, based on a convolution product of the kernel with the esti-

mators with fixed bandwidths. However, this can only be done when the bias of the estimator is written as the convolution product of the kernel with the target function. Since it is not the case in our problem, we perform the bandwidth selection with $\widehat{F}^{X'}_{h,h'} = \widehat{F}^{X'}_{h\vee h'}$ in (10). This is analogous to the procedure proposed by Chagny [15] or Comte and Johannes [19] for model selection purpose. Thus, $V(h)$ can also be seen as a penalty term. We also refer to the doctoral thesis of Chagny [16, p. 170] for technical details leading to this choice. Finally, let us notice that a criterion based on the maximum $h \vee h'$ also appears in Kerkyacharian, Lepski and Picard [43], and more recently, similar ideas are used in Goldenshluger and Lepski [41].

### *3.2. Theoretical results*

To prove our main results, we consider the following hypothesis, in addition to the assumptions defined in Section 2.2.

$H_b$ The collection $\mathcal{H}_n$ of bandwidths is such that:

- $H_{b1}$ its cardinality is bounded by $n$,
- $H_{b2}$ for any $h \in \mathcal{H}_n$, $\varphi(h) \geq C_0 \ln(n)/n$, where $C_0 > 0$ is a purely numerical constant (specified in the proofs).

**Remark 1.**

- Assumption $H_{b1}$ determines the size of the bandwidth collection: compared to the assumptions of Goldenshluger and Lepski [40], we consider a discrete set and not an interval, which permits to use the classical tools of model selection theory in the proofs.
- In practice, it is impossible to verify Assumption $H_{b2}$ since the function $\varphi$ is unknown. However, this difficulty can be circumvented by introducing a random collection of bandwidths $\widehat{\mathcal{H}}_n$ verifying, for all $h \in \widehat{\mathcal{H}}_n$, $\widehat{\varphi}(h) \geq 2C_0 \ln(n)/n$ where $\widehat{\varphi}$ is an estimator of $\varphi$ (see Equation (9)). However, since it does not add significant difficulty (see Comte and Johannes [19], Brunel, Mas and Roche [11]) but would complicate the understandability of proofs, we choose to keep Assumption $H_{b2}$.

We now state the following nonasymptotic bound for the maximal risk over the class $\mathcal{F}_\beta$.

**Theorem 2.** *Assume $H_\varphi$, $H_b$ and that $n \geq 3$. There exist two constants $c, C > 0$ depending on $c_K$, $C_K$, $c_\varphi$, $C_\varphi$, $|D|$, $C_D$ and $\kappa$ such that*

$$\sup_{F\in\mathcal{F}_\beta} \mathbb{E}\left[\left\|\widehat{F}^{X'}_{\widehat{h}} - F^{X'}\right\|^2_D \mathbf{1}_B(X')\right] \leq c \min_{h\in\mathcal{H}_n}\left\{h^{2\beta} + \frac{\ln(n)}{n\varphi(h)}\right\} + \frac{C}{n}. \tag{12}$$

The optimal bias-variance compromise is reached by the estimator, which is thus adaptive with respect to the unknown smoothness of the target function $F$. The selected bandwidth $\widehat{h}$ is performing as well as the unknown oracle $h^*$ defined in (8), up to the multiplicative constant $c$, up to a remainding term of order $1/n$ which is negligible, and up to the $\ln(n)$ factor. This extra-quantity also appears in the term $V(h)$. The loss is due to adaptation. In Section 4, we

prove that it does not affect the convergence rates of the estimator which is nevertheless optimal in the minimax sense in most of the cases.

The proof of Theorem 2 is mainly based on model selection tools, specifically concentration inequalities (see Section 7.3). A specific difficulty comes from the fact that the variance term in (7) depends on the unknown distribution of $X$, through its small ball probability. Thus, we cannot estimate this variance term with $V(h) = \kappa \ln(n)/(n\varphi(h))$, since it cannot be used in practice. This explains why we plug (9), an estimator for $\varphi(h)$ in $\widehat{V}(h)$. However, for the sake of clarity, we begin the proof by establishing the result with $\widehat{V}(h)$ replaced by its theoretical counterpart $V(h) = \kappa \ln(n)/(n\varphi(h))$. Notice also that we could build an adaptive estimator for the pointwise risk. To do so, replace $\widehat{\varphi}(h)$ in (10) by $\widehat{\varphi}^{x_0}(h) = \sum_{i=1}^{n} \mathbf{1}_{\{\|X_i - x_0\|) \leq h\}}/n$ and $X'$ by $x_0$ in the definition of $\widehat{A}(h)$.

## 4. Minimax rates

In this section, we compute the convergence rate of the oracle $\widehat{F}_{h^*}$ with $h^*$ defined by (8), the rate of the selected estimator $\widehat{F}_{\widehat{h}}$, and prove lower bounds for the conditional c.d.f. estimation problem under various assumptions on the rate of decrease of the small ball probability of the covariate $X$.

### *4.1. Small ball probabilities*

The computation of the oracle $h^*$, as well as the computation of the minimum in the right-hand-side of (12) require to fix conditions on the rate of decrease of the small ball probability $\varphi(h)$. The choice of the assumptions is crucial and determines the rates of convergence of our estimators. Small ball problems have aroused considerable interest and attention in the past decades. Lots of studies propose to compute lower and upper bounds for $\varphi(h)$, in the case of particular types of process $X$. If much attention has been given to Gaussian processes (see for example the clear account provided by Li and Shao [48]), systematic studies have also been undertaken to handle the general case of (infinite) sum of independent random variables (Lifshits [49], Dunker, Lifshits and Linde [28], Mas [50]). We consider in the sequel one of the following three hypothesis which are frequently used in the literature. This allows to understand how the small ball probability decay influences the rates (see Section 4.2). We describe below large classes of processes for which they are fulfilled.

$H_X$ There exist some constants $c_1, C_1 > 0$ such that $\varphi^{x_0}(h)$ satisfies one of the following three assumptions, for any $h > 0$.

- $H_{X,L}$ There exist some constants $\gamma_1, \gamma_2 \in \mathbb{R}$, and $\alpha > 0$ such that $c_1 h^{\gamma_1} \exp(-c_2 h^{-\alpha}) \leq \varphi^{x_0}(h) \leq C_1 h^{\gamma_2} \exp(-c_2 h^{-\alpha})$;
- $H_{X,M}$ There exist some constants $\gamma_1, \gamma_2 \in \mathbb{R}$, and $\alpha > 1$, such that $c_1 h^{\gamma_1} \exp(-c_2 \ln^{\alpha}(1/h)) \leq \varphi^{x_0}(h) \leq C_1 h^{\gamma_2} \exp(-c_2 \ln^{\alpha}(1/h))$;
- $H_{X,F}$ There exists a constant $\gamma > 0$, such that $c_1 h^{\gamma} \leq \varphi^{x_0}(h) \leq C_1 h^{\gamma}$,

where we set $x_0 = 0$ if we consider the integrated risk.

Such inequalities are strongly connected with the rate of decrease of the eigenvalues of the covariance operator $\Gamma : f \in \mathbb{H} \mapsto \Gamma f \in \mathbb{H}$ with $\Gamma f(s) = \langle f(\cdot), \mathrm{Cov}(X(\cdot), X_s)\rangle$. Recall the Karhunen-Loève decomposition of the process $X$, which can be written

$$X = \sum_{j\geq 1} \sqrt{\lambda_j}\eta_j\psi_j, \tag{13}$$

where $(\eta_j)_{j\geq 1}$ are uncorrelated real-random variables, $(\lambda_j)_{j\geq 1}$ is a non-increasing sequence of positive numbers (the eigenvalues of $\Gamma$) and $(\psi_j)_{j\geq 1}$ an orthonormal basis of $\mathbb{H}$. When $X$ really lies in an infinite dimensional space, the set $\{j \geq 1,\ \lambda_j > 0\}$ is infinite, and under mild assumptions on the distribution of $X$, it is known that $\varphi(h)$ decreases faster than any polynomial of $h$ (see e.g. Mas [50], Corollary 1, p. 10). This is the case in Assumptions $H_{X,L}$ and $H_{X,M}$. Moreover, the faster the decay of the eigenvalues is, the more the data are concentrated close to a finite dimensional space, and the slower $\varphi(h)$ decreases.

For example, when $X$ is a Gaussian process with eigenvalues $(\lambda_j)_j$ such that $cj^{-2a} \leq \lambda_j \leq Cj^{-2a}$, $a > 1/2$ $(c, C > 0)$, Assumption $H_{X,L}$ is satisfied with $\gamma_1 = \gamma_2 = (3-a)/(2a-1)$, $c_2 = a(2a/(2a-1))^{1/(2a-1)}$ and $\alpha = 1/(a - 1/2)$ (Hoffmann-Jørgensen, Shepp and Dudley [42, Theorem 4.4 and example 4.5, p. 333–334]). This classical situation of such polynomial decay covers the example of the Brownian motion, with $a = 1$ (see Ash and Gardner [4]). More generally, if $X$ is defined by a random serie $X = \sum_{j\geq 1} j^{-2a} Z_j$, for variables $Z_i$ with a c.d.f. regularly varying at 0 with positive index, one can also define $\gamma_1, \gamma_2$, and $\alpha$ such that $H_{X,L}$ is fulfilled (see Dunker, Lifshits and Linde [28, Proposition 4.1 p. 11] and also Mas [50, (19) p. 9]). The second case $H_{X,M}$ typically happens when the eigenvalues of the covariance operator exponentially decrease (see Dunker, Lifshits and Linde [28], Proposition 4.3 p. 12). In the case of a Gaussian process with $c\exp(-2j)/j \leq \lambda_j \leq C\exp(-2j)/j$, we have $c_2 = 1/2$ and $\alpha = 2$ in $H_{X,M}$ (Hoffmann-Jørgensen, Shepp and Dudley [42, Theorem 4.4 and example 4.7, pp. 333 and 336]).

Finally, it also follows from the above considerations that $H_{X,F}$ only covers the case of finite dimensional processes (the set $\{j,\ \lambda_j > 0\}$ is finite, that is the operator $\Gamma$ has a finite rank). This is the extreme case of $H_{X,M}$ (with $\alpha = 1, \gamma_1 = \gamma_2 = 0$). Nevertheless, even if our main purpose is to study functional data, the motivation to keep this case is twofold. First, we show below that our estimation method allows to recover the classical rates (upper and lower bounds) obtained for c.d.f. estimation with multivariate covariates. Then, processes which fulfill $H_{X,F}$ can still be considered as functional data since the finite space to whom $X$ belongs is unknown for the statistician.

### *4.2. Convergence rates of kernel estimators*

Under the previous regularity assumptions, we compute the upper bounds for the pointwise and integrated risks of the estimators, the proofs are deferred to the Appendix-Section A.

TABLE 1
*Rates of convergence of the oracle estimator (line (a)) and the adaptive estimator (line (b)). Minimax lower bounds (line (c))*

| | | $H_{X,L}$ (lower rate) | $H_{X,M}$ (medium rate) | $H_{X,F}$ (fast rate) |
|---|---|---|---|---|
| (a) | Rates for $\widehat{F}_{h^*}$ (upper bounds) | $(\ln(n))^{-2\beta/\alpha}$ | $\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$ | $n^{-\frac{2\beta}{2\beta+\gamma}}$ |
| (b) | Rates for $\widehat{F}_{\widehat{h}}$ (upper bounds) | $(\ln(n))^{-2\beta/\alpha}$ | $\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$ | $\left(\frac{n}{\ln(n)}\right)^{-\frac{2\beta}{2\beta+\gamma}}$ |
| (c) | Minimax risk (lower bounds) | $(\ln(n))^{-2\beta/\alpha}$ | $\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$ | $n^{-\frac{2\beta}{2\beta+\gamma}}$ |

**Proposition 1.** *(a) Under the assumptions of Theorem 1, and if $H_F$ is fulfilled, the convergence rates of the pointwise risk $\mathbb{E}[\|\widehat{F}_{h^*}^{x_0} - F^{x_0}\|^2]$, and the integrated risk $\mathbb{E}[\|\widehat{F}_{h^*}^{X'} - F^{X'}\|_D^2]$ of the oracle $\widehat{F}_{h^*}$ are given in Table 1, line (a).*

*(b) Under the assumptions of Theorem 2, the convergence rates of the integrated risk $\mathbb{E}[\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')]$ of the estimator $\widehat{F}_{\widehat{h}}$ are given in Table 1, line (b).*

*For both cases, the upper bounds are given up to a multiplicative constant, and for the different cases $H_{X,L}$, $H_{X,M}$, and $H_{X,F}$.*

Let us comment on the results. The faster the small ball probability decreases (that is the less concentrated the measure of $X$ is), the slower the rate of convergence of the estimator is. In the generic case of a process $X$ which satisfies $H_{X,L}$, the rates are logarithmic, which is not surprising. It reflects the "curse of dimensionality" which affects the functional data. Similar rates are obtained by Ferraty, Laksaci and Vieu [29] (section 5.3) in the same framework, and by Mas [50] for regression estimation (section 2.3.1). However, we show that the results can be improved when the process $X$ is more regular, although still infinite dimensional. Under Assumption $H_{X,M}$, the rates we compute have the property to decrease faster than any logarithmic function. Assumption $H_{X,F}$ is the only one which yields to the faster rate, that is the polynomial one.

We have already noticed that our adaptive procedure leads to the loss of a logarithmic factor (see the comments following Theorem 2). Nevertheless, by comparing line (a) to line (b) in Table 1, we obtain that the adaptive estimator still achieves the oracle rate if $H_{X,L}$ or $H_{X,M}$ are fulfilled. The loss is actually negligible with respect to the rates.

### *4.3. Lower bounds*

We now establish lower bounds for the risks under mild additional assumptions, showing that the estimators suggested above attain the optimal rates of convergence in a minimax sense over the class of conditional c.d.f. $\mathcal{F}_\beta$ (defined in Section 2.2). The results for the integrated risk are obtained through non-straightforward extensions of the pointwise case.

**Theorem 3.** *Suppose that $H_X$ is fulfilled, and that $n \geq 3$.*

*(i) The minimax risk $\inf_{\widehat{F}} \sup_{F \in \mathcal{F}_\beta} \mathbb{E}_F[\|\widehat{F}^{x_0} - F^{x_0}\|^2]$ is lower bounded by a quantity proportional to the ones in line (c) in Table 1.*

*(ii) Assume moreover that $B$ contains the ball $\{x \in \mathbb{H},\ \|x\| \leq \rho\}$ where $\rho > 0$ is a constant to be specified in the proof, and that there exist two constants $c_2, C_2 > 0$ such that, for all $h > 0$, for all $x \in B$,*

$$\varphi(h) > 0 \text{ and } c_2\varphi(h) \leq \varphi^x(h) \leq C_2\varphi(h). \tag{14}$$

*Then the minimax risk $\inf_{\widehat{F}} \sup_{F \in \mathcal{F}_\beta} \mathbb{E}_F[\|\widehat{F}^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')]$ is also lower bounded by a quantity proportional to the ones in line (c) in Table 1.*

*For both cases, the infimum is taken over all possible estimators obtained with the data-sample $(X_i, Y_i)_{i=1,\dots,n}$. In (i), $\mathbb{E}_F$ is the expectation with respect to the law of $\{(X_i, Y_i), i = 1, \dots, n\}$ and in (ii), $\mathbb{E}_F$ is the expectation with respect to the law of $\{\{(X_i, Y_i), i = 1, \dots, n\}, X'\}$ when, for all $i = 1, \dots, n$, for all $x \in \mathbb{H}$, the conditional c.d.f. of $Y_i$ given $X_i = x$ is $F^x$.*

Theorem 3, which is proved in Section 7.4, shows that the upper bounds of Proposition 1 cannot be improved, not only among kernel estimators but also among all estimators, under assumptions $H_{X,L}$ and $H_{X,M}$. The estimator $\widehat{F}_{\widehat{h}}$ is thus both adaptive optimal in the oracle and in the minimax senses.

The computations are new for conditional c.d.f. estimation with a functional covariate. Under $H_{X,F}$, with $\gamma = 1$, the lower bounds we obtain are consistent with Theorem 2 of Brunel, Comte and Lacour [10] or Proposition 4.1 of Plancade [52] for c.d.f. estimation with a one-dimensional covariate, over Besov balls. In the functional framework, the results can only be brought close to those of Mas [50] (Theorem 3) for regression estimation.

**Remark 2.** The rates (both lower and upper-bounds) depend on the smoothness of the target function $F$ (assumption $H_F$, $F \in \mathcal{F}_\beta$, to control the bias term) and on the smoothness of the process $X$, through the rate of decay of the small ball probabilities (larger variance for smaller small ball probabilities). The broad range of rates we establish in Table 1 (from polynomial to logarithmic rates, with intermediate cases) can be compared to the rates which are computed for the density estimation of a variable $Z$ in a deconvolution model $W = Z + \varepsilon$, from a sample drawn as $W$, and with known noise $\varepsilon$ distribution. Actually, the standard rates in this setting depend on smoothness assumptions both on the density to recover and on the noise density. Similarly to our framework, the rates are known to be slow (logarithmic) under classical smoothness assumptions (a "supersmooth" noise, and an "ordinary" smooth signal) but they can be improved by considering differents assumptions. See *e.g.* Comte, Rozenholc and Taupin [20] and Lacour [45].

## 5. Impact of the projection of the data onto finite-dimensional spaces

We have seen in Section 4.2 that, when $X$ lies in an infinite dimensional space (assumptions $H_{X,M}$ and $H_{X,L}$), the rates of convergence are slow. This "curse of dimensionality" phenomenon is well known in kernel estimation for high or infinite dimensional datasets. The introduction of the projection semi-metrics $d_p$, leading to the estimators (4), has thus been proposed in order to circumvent this problem. Defining such estimators amounts to project the data into a $p$-dimensional space. Indeed, this permits to address the problem of variance reduction since $\varphi_p(h) := \mathbb{P}(d_p(x,0) \leq h) \sim_{h\to 0} C(p)h^p$ and then the variance is of order $1/(nh^p)$. Notice that, even if the variance orders of magnitude are the same, the situation here is different from Assumption $H_{X,F}$ with $\gamma = p$: $H_{X,F}$ amounts to suppose that the curve $X$ lies a.s. in an unknown finite-dimensional space (see Section 4.1) whereas, here, the data is projected into a finite-dimensional space but may lie in an infinite-dimensional space.

A first thing we can say is that, under our regularity assumption $H_F$, Theorem 3 remains true and the convergence rate of the risk of $\widehat{F}_{h,p}$ cannot be better than the lower bounds given in Table 1, line (c). This implies that, in our setting, the estimator $\widehat{F}_{h,p}$ cannot converge at significantly better rates than our adaptive estimator $\widehat{F}_{\widehat{h}}$ even if the couple of parameters $(p,h)$ is well-chosen. Precisely, as shown in Proposition 2 below, project data also adds an additional bias term which compensates for the decrease of the variance.

### *5.1. Assumptions*

In order to state the result, we need the following assumptions.

$H'_\varphi$ There exist two constants $c_\varphi, C_\varphi > 0$ such that for all $h \in \mathbb{R}$, for all $p \in \mathbb{N}^*$,

$$c_\varphi \varphi_p(h)\mathbf{1}_B(X') \leq \varphi_p^{X'}(h)\mathbf{1}_B(X') \leq C_\varphi \varphi_p(h)\mathbf{1}_B(X') \text{ a.s.},$$

where $X'$ is an independent copy of $X$ and $\varphi_p^{X'}(h) := \mathbb{P}_{X'}(d_p(X,X') \leq h)$.

$H_\xi$ Let $\xi_j := \langle X, e_j\rangle/\sigma_j$ where $\sigma_j := \mathrm{Var}(\langle X, e_j\rangle)$. One of the two following assumptions is verified:

$H_\xi^{ind}$ the sequence of random variables $(\xi_j)_{j\geq 1}$ is independent and there exists a constant $C_\xi$ such that, for all $j \geq 1$

$$\mathbb{E}\left[\xi_j^\beta\right] \leq C_\xi;$$

$H_\xi^b$ there exists a constant $C_\xi$ such that, for all $j \geq 1$,

$$|\xi_j| \leq C_\xi \text{ a.s.}$$

Note that Assumption $H'_\varphi$ is the equivalent of Assumption $H_\varphi$ replacing $d$ by $d_p$. If $X$ is a Gaussian process, the vector $(\langle X, e_1\rangle, \ldots, \langle X, e_p\rangle)$ is a Gaussian vector and Assumption $H_\varphi$ is also verified provided that $B$ is bounded. Assump-

tion $H_\xi^{ind}$ is true if $X$ is a Gaussian process and $(e_j)_{j\geq 1}$ is the Karhunen-Loève basis of $X$ (see (13) above, and also Ash and Gardner [4]). Assumption $H_\xi^b$ is equivalent to suppose that $X$ is bounded a.s. We are aware that both assumptions $H_\xi^{ind}$ and $H_\xi^b$ are strong since in most cases the Karhunen-Loève basis is unknown. The aim of Proposition 2 below is to permit a better understanding of the bias-variance decomposition of the risk when the data are projected. A further study would be needed to obtain weaker assumptions but this is beyond the scope of this paper.

### 5.2. Upper bound

**Proposition 2.** *Suppose assumptions $H_F$ and $H_\xi$ are satisfied. Let $h>0$ and $p\in\mathbb{N}^*$ be fixed.*

*(i) For all $x_0\in\mathbb{H}$ we have*

$$\mathbb{E}\big[\big\|\widehat{F}_{h,p}^{x_0}-F^{x_0}\big\|_D^2\big]\leq C\left(h^{2\beta}+\left(\sum_{j>p}\sigma_j^2\right)^{\beta}+\left(\sum_{j>p}\langle x_0,e_j\rangle^2\right)^{\beta}+\frac{1}{n\varphi_p^{x_0}(h)}\right), \tag{15}$$

*where $C>0$ only depends on $C_\xi$, $\beta$, $c_K$, $C_K$, $|D|$ and $C_D$.*

*(ii) If, in addition, Assumption $H_\varphi$ is satisfied,*

$$\mathbb{E}\big[\|\widehat{F}_{h,p}^{X'}-F^{X'}\|_D^2\mathbf{1}_B(X')\big]\leq C\left(h^{2\beta}+\left(\sum_{j>p}\sigma_j^2\right)^{\beta}+\frac{1}{n\varphi_p(h)}\right), \tag{16}$$

*where $C>0$ only depends on $C_\xi$, $\beta$, $c_K$, $C_K$, $c_\varphi$, $C_\varphi$, $|D|$ and $C_D$.*

We have additional bias terms compared to Ferraty, Laksaci and Vieu [29], Ferraty et al. [34]. This is due to the fact that our regularity assumption $H_F$ (see Section 2.2) is different from Assumption (H2) of Ferraty, Laksaci and Vieu [29] or Assumption (H2') of Ferraty et al. [34]. Our assumption is expressed with the norm of $\mathbb{H}$ whereas their assumptions are expressed with the semi-norm used in the definition of the estimator (here $d_p$). Remark that, with projection semi-norms, the assumptions of Ferraty, Laksaci and Vieu [29], Ferraty et al. [34] imply that the function $F^x$ only depends on $(\langle x,e_j\rangle)_{1\leq j\leq p}$. Indeed, if we take $x$ and $x'$ such that $\langle x,e_j\rangle=\langle x',e_j\rangle$ for $j=1,\dots,p$ (but $\langle x,e_j\rangle\neq\langle x',e_j\rangle$ for some $j>p$), both (H2) and (H2') imply that $F^x(y)=F^{x'}(y)$ for all $y$. Our assumption is then less restrictive.

**Remark 3.** Notice that the estimator (4) is not consistent when $p$ is fixed. This is also noted by Mas [50] in a regression setting (see Remark 2, p. 4). It is coherent with the fact that we loose information when we project the data. Indeed, suppose that the signal $X$ lies a.s. in $(\mathrm{span}\{e_1,\dots,e_p\})^\perp$, then $d_p(X_i,x)=\sqrt{\sum_{j=1}^p\langle x,e_j\rangle^2}$ a.s. and $\widehat{F}_{h,p}^x(y)=\frac{1}{n}\sum_{i=1}^n\mathbf{1}_{\{Y_i\leq y\}}$ if $K_h(\sqrt{\sum_{j=1}^p\langle x,e_j\rangle^2})\neq 0$ and 0 otherwise. The bias of such an estimator is then constant and non null as

soon as there exists $F^x(y) \neq \mathbb{P}(Y \leq y)$ on a subset of $D$ of positive Lebesgue measure. Hence in order to obtain a consistent estimator in the case where $\sigma_j > 0$ for all $j$, we have to impose that $p$ is depending on $n$ and $\lim_{n\to+\infty} p = +\infty$.

### *5.3. Discussion*

The rates obtained can be compared to the lower bounds given in Table 1 in the Gaussian case under assumptions $H_{X,F}$ and $H_{X,M}$.

#### *5.3.1. Comparison with the rates obtained under Assumption $H_{X,F}$*

We start from the Karhunen-Loève decomposition of $X$ defined in (13). For a Gaussian process, the variables $\eta_j$ are independent standard normal, $(\lambda_j)_{j\geq 1}$ is a non-increasing sequence of positive numbers and $(\psi_j)_{j\geq 1}$ a basis of $\mathbb{H}$. If $\lambda_{\gamma+1} = 0$ and $\lambda_\gamma > 0$ and if the law of $(\eta_1, \ldots, \eta_\gamma)$ is non-degenerate then Assumption $H_{X,F}$ is satisfied. Two cases may then occur.

- If $e_j = \pm\psi_j$ for all $j$, then $\sigma_j^2 = \mathbb{E}[\langle X, e_j\rangle^2] = \mathbb{E}[\langle X, \psi_j\rangle^2] = \lambda_j$ and $\sigma_j = 0$ for $j > \gamma$. Then from Inequality (16), with a good choice of $(p, h)$, the integrated risk is upper bounded by $Cn^{-2\beta/(2\beta+\gamma)}$ which fits with the lower bound. According to Inequality (15), the pointwise risk is penalized by the term $\sum_{j>p}\langle x_0, e_j\rangle^2$ and the minimax rate is attained only if $x_0 \in \operatorname{span}\{e_1, \ldots, e_p\}$.
- However, if the basis $(e_j)_{j\geq 1}$ is not well-chosen for instance if $e_j = \psi_j$ for $j \notin \{\gamma, \gamma + l\}$ $(l > 0)$, $e_\gamma = \psi_{\gamma+l}$ and $e_{\gamma+l} = \psi_\gamma$, the integrated risk of the estimator is upper bounded by $Cn^{-2\beta/(2\beta+\gamma+l)}$ whereas the minimax rate is $n^{-2\beta/(2\beta+\gamma)}$.

#### *5.3.2. Comparison with the rates obtained under Assumption $H_{X,M}$*

Thanks to Proposition 2, we are able to obtain the rate of convergence for the estimator.

**Corollary 1.** *Suppose that the assumptions of Proposition 2 are satisfied and that $(\xi_1, \ldots, \xi_p)$ admits a density $f_p$ with respect to the Lebesgue measure on $\mathbb{R}^p$ such that there exists a constant $c_f$ verifying*

$$f_p(0) \geq c_f^p.$$

*Assume also that there exist $\delta > 1$, $c > 0$ such that $\sum_{j>p}\sigma_j^2 \leq cp^{-2\delta+1}$.*

*(i) Then, for all $x_0 \in \mathbb{H}$ such that there exist $\delta' > 1$, $c' > 0$ such that $\sum_{j>p}\langle x_0, e_j\rangle^2 \leq c'p^{-\delta'+1}$ we have*

$$\mathbb{E}\left[\|\widehat{F}_{h,p}^{x_0} - F^{x_0}\|_D^2\right] \leq C\left(\frac{\ln(n)}{\ln(\ln(n))}\right)^{\beta(1-2\min\{\delta,\delta'\})},$$

*for a well-chosen bandwidth $h$ and a good choice of $p$, and where $C > 0$ is a numerical constant.*

*(ii) We also have*

$$\mathbb{E}\left[\|\widehat{F}_{h,p}^{X'} - F^{X'}\|_D^2\right] \leq C\left(\frac{\ln(n)}{\ln(\ln(n))}\right)^{\beta(1-2\delta)},$$

*for a well-chosen bandwidth $h$ and a good choice of $p$, and where $C > 0$ is a numerical constant.*

If $cj^{-2a} \leq \lambda_j \leq Cj^{-2a}$, for two constants $c, C > 0$, then Assumption $H_{X,M}$ is satisfied with $\alpha = 1/(a-1/2)$, the estimator converges with the minimax rate if $\delta = a$ (adding the condition $\delta' \geq a$ for the pointwise risk). The conclusion is similar to Paragraph 5.3.1: if $e_j = \pm\psi_j$ for all $j \geq 1$ (recall that this condition is unrealistic since in most cases the basis $(\psi_j)_{j\geq 1}$ is unknown) then we can choose $p$ and $h$ such that the minimax rate is achieved, up to a logarithmic factor, for the integrated risk and the pointwise risk under an additional condition on $x_0$. Otherwise, we do not know if the minimax rate can be achieved.

## 6. Numerical study

We briefly present an illustration of the theoretical results. We first describe the parameters we use and the way the risks are computed. Then we illustrate the performances of our method with some figures and tables, for both simulated samples and a real dataset.

### *6.1. Implementation of the estimators*

We implement the two types of estimators (2) and (4) with the uniform kernel $K = \mathbf{1}_{(0;1)}$. We choose the bandwidth $h$ in a collection $\mathcal{H}_n$ with $\mathcal{H}_n = \{C/k,\ 1 \leq k \leq k_{\max}\}$. Thanks to the definition of the kernel $K$, it is useless to consider bandwidths $h$ which are larger than $\max\{\|X_i - x_0\|, i = 1, \ldots, n\}$. Indeed, for any $h > \max\{\|X_i - x_0\|, i = 1, \ldots, n\}$ we have $K_h(\|X_i - x_0\|) = 1/h$ and $\widehat{F}_h(y) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}}$. Then we set $C = \max\{\|X_i - x_0\|, i = 1, \ldots, n\}$. To choose the other parameter $k_{\max}$, keep in mind that $\widehat{\varphi}(h)$ should not be too small: the aim is to have an empirical counterpart for Assumption $H_{b2}$ and also to avoid instabilities in the calculation of $\widehat{V}(h)$. We propose to ensure that $\widehat{\varphi}(h) \geq \ln(n)/n$. This is the case when $C/k_{\max}$ is the quantile of order $\ln(n)/n$ of the sample $\{\|X_i\|,\ 1 \leq i \leq n\}$.

For the collection of estimators $(\widehat{F}_h)_{h\in\mathcal{H}_n}$ defined in (2), we obviously implement the selection algorithm which lead to $\widehat{F}_{\widehat{h}}$. The method to compute it is entirely described in Section 3.1. The $L^2$ norm involving in the definition of $\widehat{A}(h)$ (see (10)) is approximated by the trapezoidal rule over $D = (\min_{1\leq i\leq n} Y_i, \max_{1\leq i\leq n} Y_i)$. The only parameter to be adjusted is the constant $\kappa$ of the penalty term $\widehat{V}(h)$. The simulations permit to tune it (see Section 6.2 below).

This calibration, as well as the simulation study of the performances of the estimators requires risk computations. For an estimator $\widehat{F}$ of the conditional

c.d.f. $F$, the integrated risk $\mathbb{E}[\|\widetilde{F}^{X'} - F^{X'}\|_D^2]$ is approximated by the mean of $N = 50$ Monte-Carlo replications of the random variable $\|\widetilde{F}^{X'} - F^{X'}\|_D^2$. The interval $D$ is chosen as explained above. The set $B$ involved in (5) is necessary only for the theoretical study and can be chosen as large as possible. It is ignored here.

### *6.2. Simulation study*

#### *6.2.1. Simulation of the data-samples*

Our procedure is applied to different simulated samples $(X_i, Y_i)_{i\in\{1,\dots,n\}}$, with $n = 500$, drawn as $(X, Y)$. The functional covariate $X$ is simulated in the following way:

$$X(t) := \xi_0 + \sum_{j=1}^{J} \sqrt{\lambda_j}\xi_j\psi_j(t),$$

where $(\lambda_j)_{j\geq 1}$ is a sequence of positive real numbers such that $\sum_{j\geq 1}\lambda_j < +\infty$, $(\xi_j)_{j\geq 0}$ is a sequence of *i.i.d.* standard normal random variables and $\psi_j(t) := \sqrt{2}\sin(\pi(j-0.5)t)$, $t \in \mathbb{R}$. Remark that the simulated $X$ is a finite-dimensional Gaussian process. Hence, Assumption $H_{X,F}$ (see Section 4.1) is satisfied, for $\gamma = J + 1$. However, if $J$ is sufficiently large, we will consider that

$$X(t) = \sum_{j=1}^{J} \sqrt{\lambda_j}\xi_j\psi_j(t) \approx \sum_{j\geq 1} \sqrt{\lambda_j}\xi_j\psi_j(t)$$

and $X$ has thus a behaviour very similar to the one of an infinite-dimensional process. In the simulation study, only a few coefficients are estimated, since the sample size is finite, evidently.

The three following choices are considered, to illustrate the three regularity conditions studied above. The processes are plotted in Figure 1.

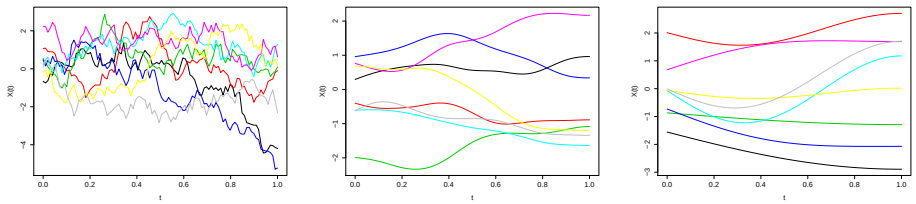
- **(a)** $\lambda_j = j^{-2}$, $J = 150$, corresponding to a Brownian motion, which verifies Assumption $H_{X,L}$.
- **(b)** $\lambda_j = e^{-2j}/j$, $J = 150$, which corresponds to a process $X$ verifying $H_{X,M}$.
- **(c)** $\lambda_j = j^{-2}$, $J = 2$, which leads to a process $X$ satisfying $H_{X,F}$ (with $\gamma = 3$).

When $X$ is simulated, we obtain $Y$ as follows.

**Example 1 (Regression model)** $Y = \langle\beta, X\rangle^2 + \varepsilon$ with $\beta(t) = \sin(4\pi t)$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, where $\varepsilon$ is independent of $X$ and $\sigma^2 = 0.1$. With this model, the cumulative distribution function of $Y$ given $X = x$ is

$$F_1^x(y) := \Phi((y - \langle\beta, x\rangle^2)/\sigma),$$

where $\Phi(z) = \mathbb{P}(Z \leq z)$ for a $Z \sim \mathcal{N}(0, 1)$.



(a) $-H_{X,L}$ (b) $-H_{X,M}$ (c) $-H_{X,F}$

FIG 1. *Realisations of the process $X$: 8 independent curves in Examples (a), (b), and (c).*

**Example 2 (Gaussian mixture model)** The conditional distribution of $Y$ given $X = x$ is $0.5\mathcal{N}(8-4\|x\|,1)+0.5\mathcal{N}(8+4\|x\|,1)$. Then, the cumulative distribution function of $Y$ given $X = x$ is

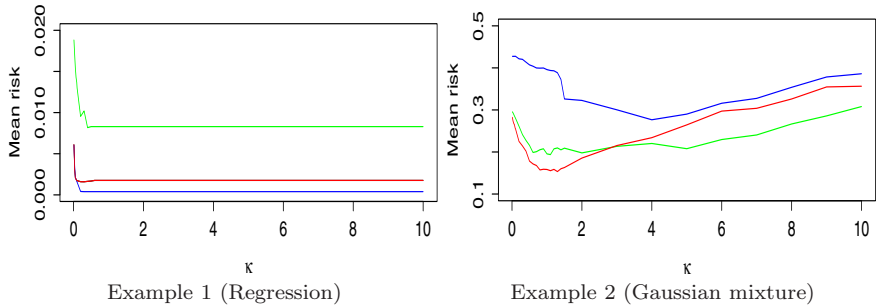
$$F_2^x(y) := 0.5\Phi(8-4\|x\|) + 0.5\Phi(8+4\|x\|).$$

### *6.2.2. Calibration of the penalty constant*

The question that needs to be answered now is: how to choose the constant $\kappa$ involved in the penalty term $\widehat{V}(h)$ (in (10))? Its choice is crucial for the quality of adaptive estimation. Keep in mind that $\widehat{h}$ is chosen as the bandwidth which realizes the best compromise of two terms, the estimator of the bias term, namely $\widehat{A}(h)$ and the variance term $\widehat{V}(h) = \kappa \ln(n)(n\widehat{\varphi}(h))^{-1}$. From (11), heuristically,

- if $\kappa$ is small, then $\widehat{A}(h)$ is the most influential term. Since this term tends to decrease with $h$, we select small bandwidths;
- if $\kappa$ is large, the reverse occurs: we select large bandwidths.

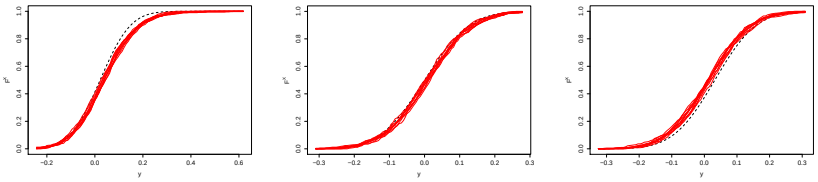
To illustrate this phenomenon, we plot the risk of the estimator $\widehat{F}_{\widehat{h}}$ in function of the parameter $\kappa$, see Figure 2. In the light of the results, as expected, $\kappa$ should not be chosen too small (otherwise the risk blows up), neither too large. However a choice for $\kappa$ larger than the minimizer does not make the risk blowing up (same




Example 1 (Regression) Example 2 (Gaussian mixture)

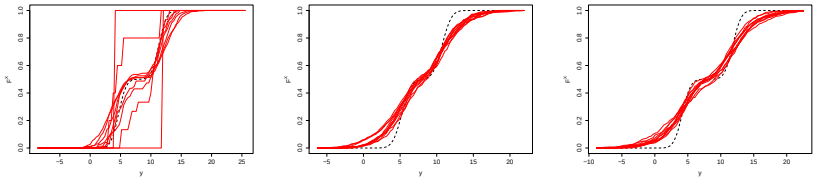
FIG 2. *Plot of approximated mean risk with respect to $\kappa$. Blue curves: Process (a) – $H_{X,L}$. Green curves: Process (b) – $H_{X,M}$. Red curves: Process (c) – $H_{X,F}$.*

**Example 1 (Regression model)**



(a) – $H_{X,L}$ (b) – $H_{X,M}$ (c) – $H_{X,F}$

**Example 2 (Gaussian mixture model)**



(a) – $H_{X,L}$ (b) – $H_{X,M}$ (c) – $H_{X,F}$

FIG 3. *Plots of true conditional c.d.f. (black dotted line) versus 10 estimators* $\widehat{F}^{x_0}_{\widehat{h}}$ *calculated from independent samples of size* $n = 500$ *(red thin lines), with* $x_0$ *a copy of* $X$*. First line: Example 1, true c.d.f.* $F_1^{x_0}$*. Second line: Example 2, true c.d.f.* $F_2^{x_0}$*.*

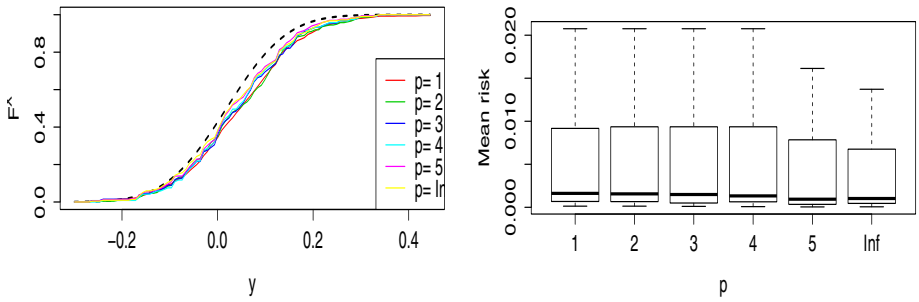
results are obtained in Bertin, Lacour and Rivoirard [8]). We thus fix $\kappa = 4$. Another value around 4 just improves the results for some c.d.f. and deteriorates them for some others. Data-driven calibration tools developed in model selection contexts, such as slope heuristic, could be very useful here. However this kind of tool does not exist in our bandwidth selection context, and this is beyond the scope of the paper.

### *6.2.3. Simulation results*

Beams of estimators are presented in Figure 3. The estimation is quite stable for the regression model (Example 1). However the estimation is harder for the Gaussian mixture model in the cases $H_{X,L}$: a more detailed study shows that our criterion selects models with too small $h$. The bandwidth selection criterion behaves better in the cases $H_{X,F}$ and $H_{X,M}$.

To study the impact of the projection of the data from the practical point of view, we also compare the estimator $\widehat{F}_{h,p}$ (see (2)) with $\widehat{F}_h$ (see (4)), which can be seen as $\widehat{F}_{h,p}$ for $p = \infty$. Since we have not developed adaptive estimation for $\widehat{F}_{h,p}$ with finite value of $p$, we choose to compare the corresponding oracle estimators $\widehat{F}_{h^*,p}$ with $h^*$ defined by (8) (just replace $\widehat{F}_h$ in the *r.h.s* of (8) by $\widehat{F}_{h,p}$). As expected from the theory (see Proposition 2), the smaller $p$, the more biased the estimators: as an example, the oracle estimates in Example 1 (b) (regression model with process $X$ satisfying $H_{X,M}$) are plotted on Figure 4. The boxplots for the corresponding risks (over 50 estimators) are also plotted

**Example 1 (Regression model), (a)**




**Example 2 (Gaussian mixture model), (a)**

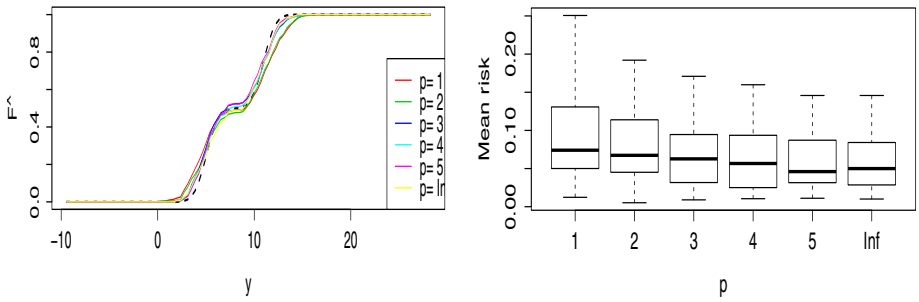



FIG 4. *First column: Plots of true conditional c.d.f. (black dotted line) versus estimators* $\widehat{F}^{x_0}_{h^*,p}$*, for* $p \in \{1, 2, 3, 4, 5, \infty\}$*, calculated from a sample of size* $n = 500$ *(thin lines), with* $x_0$ *a copy of* $X$*, in Example 1 (first line) and 2 (second line) with process (a). Second column: Boxplots of the risks of* 50 *previous estimators for the same values of* $p$*.*

(outliers are not drawn). The mean integrated risks for all the studied models are presented in Table 2. Both the boxplots of Figure 4 and the risks of Table 2 confirm that the risks generally become quite smaller when $p$ increases.

### *6.3. Application to a spectrometric dataset*

We propose to study a classical dataset, widely studied in regression contexts (see Ferraty and Vieu [32, 33], Ferraty, Mas and Vieu [30]). The data is available on line[1], and have been recorded on Tecator Infratec Food and Feed Analyzer. For each unit $i$ ($i = 1, \ldots, n$, with $n = 215$), we observe one spectrometric curve $X_i$ which corresponds to the absorbances of a piece of chopped meat measured at 100 wavelengths ranging between 850 and 1050 nm. The aim is to describe the link between a spectrometric curve of a chopped meat and its fat content. For each curve $X_i$, we denote by $Y_i$ the corresponding fat content. The sample curves are plotted in Figure 5, first column.

In a first step, the data are centred to match our assumptions. Then we apply

[1] http://www.math.univ-toulouse.fr/~ferraty/SOFTWARES/NPFDA/

TABLE 2
*Values of the Mean integrated risks* $\times 10$ *averaged over* 50 *samples, for the estimators* $\widehat{F}^{x_0}_{h^*,p}$, *with* $x_0$ *a copy of* $X$, *and for* $p \in \{1, 2, 3, 4, 5, \infty\}$

Example 1

| p= | 1 | 2 | 3 | 4 | 5 | $+\infty$ |
|---|---|---|---|---|---|---|
| (a) | 0.0044 | 0.0043 | 0.0039 | 0.0039 | 0.0038 | 0.0037 |
| (b) | 0.1796 | 0.1788 | 0.1768 | 0.1773 | 0.1649 | 0.1483 |
| (c) | 0.0171 | 0.0166 | 0.0075 | 0.0074 | 0.0071 | 0.0061 |

Example 2

| p= | 1 | 2 | 3 | 4 | 5 | $+\infty$ |
|---|---|---|---|---|---|---|
| (a) | 0.7110 | 0.6603 | 0.6251 | 0.6545 | 0.6144 | 0.6184 |
| (b) | 1.2826 | 1.1817 | 1.0440 | 1.0212 | 0.9677 | 0.8481 |
| (c) | 1.1528 | 1.0509 | 0.7024 | 0.7034 | 0.6968 | 0.6596 |

our global adaptive estimation procedure. We choose ten curves $x_0^{(j)} := X(i_0^{(j)})$ $(j \in \{1, \ldots, 10\})$ randomly in the sample. For each curve $x_0^{(j)}$ we estimate the conditional distribution of $Y$ given $X = x_0^{(j)}$ using the information of $\{(X_i, Y_i), i \neq i_0^{(j)}\}$. The results are given in Figure 5. We see that when $Y_{i_0^{(j)}}$ is small (blue curves), the estimated conditional distribution function $y \mapsto \widehat{F}^{x_0^{(j)}}_{\widehat{h}}$ is strongly increasing in the interval $[0, 10]$, which indicates that our estimator detects that $Y$ must be small with large probability. Conversely, when $Y_{i_0^{(j)}}$ is large (red curves) the estimated conditional distribution function faster increases in the interval $[30, 50]$. These results prove that our estimators are able to capture the repartition of $Y$ given $X = x_0$ when $x_0$ is taken in the sample.

## Concluding remarks

- The estimation procedure we propose is not restricted to the case where the covariate $X$ is functional. Indeed the adaptive estimator $\widehat{F}_{\widehat{h}}$ can be calculated as soon as the covariate $X$ takes values in a general Hilbert

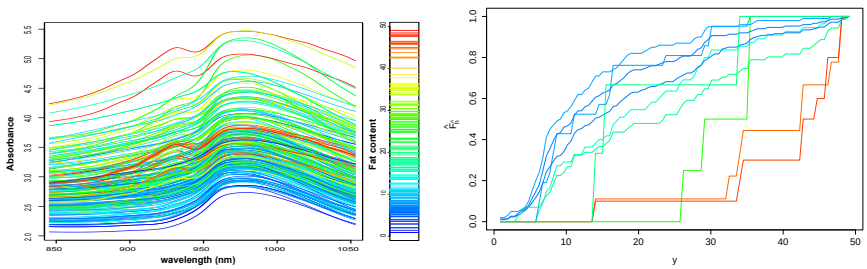



FIG 5. *First column: Curves of the sample (left). The colors of each curve is defined by its corresponding fat content (right). Second column: Plots of* $\widehat{F}^{x_0^{(j)}}_{\widehat{h}}$, *for* $j = 1, \ldots, 10$. *The color of each plot is defined by the fat content* $Y_{i_0^{(j)}}$ *as represented in the first column.*

space $(\mathbb{H}, \|\cdot\|)$. The results can be applied to a function space such as $\mathbb{L}^2(I)$ $(I \subset \mathbb{R})$, $\mathbb{L}^2(\mathbb{R}^d)$ or a Sobolev space but also $\mathbb{R}^d$, $\mathbb{C}^d$, $\ell^2(\mathbb{N})$,... The results given in sections 2 and 3 remain valid. For instance, in the case where $X \in \mathbb{R}^d$, an immediate consequence of Theorem 1, is that both the pointwise and integrated risks of $\widehat{F}_{\widehat{h}}$ converge to 0 at the rate $(n/\ln(n))^{-2\beta/(2\beta+d)}$.

- Is there a solution to the "curse of dimensionality"? We prove that, under our assumptions, the classical Nadaraya-Watson estimator (2) with $d(x, x') = ||x - x'||$ attains the minimax rate of convergence. Then, in our setting, even if these rates are slow, they cannot be significantly improved by changing the semi-norm $d$ in the kernel. A reflexion is under way on determining if it is possible to modify the rates considering more regular functions F than the ones of the class $\mathcal{F}_\beta$, for instance taking into account the derivatives of the covariate $X$ in the spirit of Ferraty and Vieu [32]. Another approach may be to reduce the structural complexity of the model considering e.g. single or multiple-index models (Chen, Hall and Müller [17], Ait-Saïdi et al. [1]).

## 7. Proofs

### 7.1. Preliminary notation and results

We will mainly focus on the proof of the results for the integrated risk (since it is the one for which adaptation results are provided), and only highlight the differences when choosing the pointwise criterion. Some technical proofs are postponed to the Appendix (Section A).

We denote by $\mathbb{E}_{X'}$ (resp. $\mathbb{P}_{X'}$, $\mathrm{Var}_{X'}$) the conditional expectation (resp. probability, variance) given $X'$. We also introduced the classical norm $\|.\|_{L^q(\mathbb{R})}$ of the space $L^q(\mathbb{R})$ of integrable functions (the notation will be used with $q = 2$ and $q = \infty$).

Recall that $K_h(x) := h^{-1}K(h^{-1}x)$. Assumptions on the kernel and $H_\varphi$ imply that, for all $l \geq 1$,

$$h^{-l} m_l \varphi(h) \mathbf{1}_B(X') \leq \mathbb{E}_{X'}\left[K_h^l(d(X, X'))\right] \mathbf{1}_B(X') \leq h^{-l} M_l \varphi(h) \mathbf{1}_B(X') \text{ a.s.} \tag{17}$$

where $m_l := c_K^l c_\varphi$ and $M_l := C_K^l C_\varphi$. These inequalities are useful in the sequel.

One of the key arguments in the proofs of Theorems 1, 2, and Proposition 2 is the control of the deviations (in probability and expectation) of the process $R_h^x$, for $x \in \mathbb{H}$, defined by

$$R_h^x = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^n \dfrac{K_h(d(X_i, X'))}{\mathbb{E}_{X'}\left[K_h(d(X, X'))\right]}, & \text{if } x = X', \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^n \dfrac{K_h(d(X_i, x))}{\mathbb{E}\left[K_h(d(X, x))\right]}, & \text{if } x \in \mathbb{H} \text{ is fixed.} \end{cases} \tag{18}$$

The following lemma, proved in Section A, establishes the result which is useful to control the integrated risk of the estimators. The proof can be found below.

**Lemma 1.** *Assume $H_\varphi$. For any $\eta > 0$, on the set $\{X' \in B\}$, the following inequality holds a.s.*

$$\mathbb{P}_{X'}\left(\left|R_h^{X'} - 1\right| > \eta\right) \leq 2\exp\left(-\frac{n\eta^2\varphi(h)}{2\left(\frac{M_2}{m_1^2} + \frac{C_K\eta}{m_1}\right)}\right). \tag{19}$$

*Moreover, assume also $H_{b2}$, and denote by $V_R(h) = \kappa_R \ln(n)/(n\varphi(h))$, we have a.s.*

$$\mathbb{E}_{X'}\left[\left(\left(R_h^{X'} - 1\right)^2 - V_R(h)\right)_+\right] \leq \min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n^\alpha}, \tag{20}$$

*for any $\alpha > 0$, as soon as $\kappa_R > \max(4M_2\alpha/m_1^2, 32C_K^2\alpha^2/m_1^2C_0)$.*

*Fix a point $x_0 \in \mathbb{H}$. Then Inequality (19) becomes*

$$\mathbb{P}\left(|R_h^{x_0} - 1| > \eta\right) \leq 2\exp\left(-\frac{n\eta^2\varphi^{x_0}(h)}{2\left(\frac{C_K^2}{c_K^2} + \frac{C_K\eta}{c_K}\right)}\right). \tag{21}$$

### *7.2. Proof of Theorem 1*

#### *7.2.1. Main part of the proof of the Inequality (7)*

Following Ferraty, Laksaci and Vieu [29], Ferraty et al. [34], we define

$$\widetilde{F}_h^{X'}(y) := \sum_{i=1}^n \widetilde{W}_h^{(i)}(X')\mathbf{1}_{\{Y_i \leq y\}}, \text{ where } \widetilde{W}_h^{(i)}(X') = \frac{K_h(d(X_i, X'))}{n\mathbb{E}_{X'}[K_h(d(X_1, X'))]}. \tag{22}$$

We also have $R_h^{X'} := \sum_{i=1}^n \widetilde{W}_h^{(i)}(X')$ (see Definition (18)). First, notice that since $\widehat{F}_h^{X'} \leq 1$ and $F^{X'} \leq 1$ a.s.,

$$\mathbb{E}\left[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_{\{R_h^{X'} < 1/2\}}\mathbf{1}_B(X')\right] \leq 4|D|\mathbb{P}\left(\left\{R_h^{X'} < 1/2\right\} \cap \{X' \in B\}\right).$$

Now, with $\mathbb{P}(\{R_h^{X'} < 1/2\} \cap \{X' \in B\}) \leq \mathbb{P}(\{|R_h^{X'} - 1| > 1/2\} \cap \{X' \in B\})$ and with Lemma 1 we get

$$\begin{aligned}\mathbb{E}\left[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_{\{R_h^{X'} < 1/2\}}\mathbf{1}_B(X')\right] &\leq 8|D|\exp\left(-\frac{m_1}{8\left(M_2/m_1 + C_K/2\right)}n\varphi(h)\right) \\ &\leq \frac{C}{n\varphi(h)},\end{aligned}$$

where $C = 64|D|e^{-1}\frac{M_2/m_1 + C_K/2}{m_1}$. The last inequality comes from the bound $xe^{-x} \leq e^{-1}$, $x > 0$.

We must now control $\mathbb{E}[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')\mathbf{1}_{\{R_h^{X'} \geq 1/2\}}]$. Recall that $\widehat{F}_h^{X'} = \widetilde{F}_h^{X'}/R_h^{X'}$. We thus have,

$$\begin{aligned}
&\mathbb{E}\left[\|\widehat{F}_h^{X'} - F^{X'}\|_D^2 \mathbf{1}_B(X')\mathbf{1}_{\{R_h^{X'} \geq 1/2\}}\right] \\
&\quad \leq 3\mathbb{E}\left[\left\|\frac{\widetilde{F}_h^{X'}}{R_h^{X'}} - \mathbb{E}_{X'}\left[\frac{\widetilde{F}_h^{X'}}{R_h^{X'}}\right]\right\|_D^2 \mathbf{1}_B(X')\mathbf{1}_{\{R_h^{X'} \geq 1/2\}}\right] \\
&\qquad + 3\mathbb{E}\left[\left\|\mathbb{E}_{X'}\left[\frac{\widetilde{F}_h^{X'}}{R_h^{X'}}\right] - \frac{F^{X'}}{R_h^{X'}}\right\|_D^2 \mathbf{1}_B(X')\mathbf{1}_{\{R_h^{X'} \geq 1/2\}}\right] \\
&\qquad + 3\mathbb{E}\left[\left\|\frac{F^{X'}}{R_h^{X'}} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\mathbf{1}_{\{R_h^{X'} \geq 1/2\}}\right] \\
&\quad \leq 12\mathbb{E}\left[\left\|\widetilde{F}_h^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2 \mathbf{1}_B(X')\right] + 12\mathbb{E}\left[\left\|\mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right] - F^{X'}\right\|_D^2\right] \\
&\qquad + 12\mathbb{E}\left[\left(1 - R_h^{X'}\right)^2 \left\|F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right]. \qquad (23)
\end{aligned}$$

The first and third terms are variance terms, bounded by Lemmas 2 and 3 proved below. The second one is a bias term, controlled by Lemma 4.

**Lemma 2.** *Under Assumption $H_\varphi$, on the set $\{X' \in B\}$,*

$$\mathbb{E}_{X'}\left[\left\|\widetilde{F}_h^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] \leq |D|\frac{M_2}{m_1^2}\frac{1}{n\varphi(h)}.$$

**Lemma 3.** *Under Assumption $H_\varphi$*

$$\mathbb{E}\left[\left(R_h^{X'} - 1\right)^2 \mathbf{1}_B(X')\right] \leq \frac{M_2}{m_1^2}\frac{1}{n\varphi(h)}.$$

**Lemma 4.** *Under Assumption $H_F$,*

$$\mathbb{E}\left[\left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] \leq C_D^2 h^{2\beta}.$$

This ends the proof of Inequality (7). The scheme can easily be adapted to prove (6). □

*7.2.2. Proof of Lemmas 2 and 3 (upper bounds for the variance terms)*

***Proof of Lemma 2*** By Fubini's Theorem

$$\mathbb{E}_{X'}\left[\left\|\widetilde{F}_h^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] = \int_D \mathbb{E}_{X'}\left[\left(\widetilde{F}_h^{X'}(y) - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}(y)\right]\right)^2\right] dy$$

$$= \int_D \mathrm{Var}_{X'}\left(\widetilde{F}_h^{X'}(y)\right) dy.$$

Since, for all $y \in D$, $\widetilde{F}_h^{X'}(y)$ is a mean of independent and identically distributed random variables (conditionally to $X'$), we have, on the set $\{X' \in B\}$,

$$\begin{aligned}\mathbb{E}_{X'}\left[\left\|\widetilde{F}_h^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] &= \frac{1}{n}\int_D \mathrm{Var}_{X'}\left(\frac{K_h\left(d(X_1,X')\right)\mathbf{1}_{\{Y_1\le y\}}}{\mathbb{E}_{X'}\left[K_h\left(d(X_1,X')\right)\right]}\right) dy \\ &\le \frac{|D|}{n}\mathbb{E}_{X'}\left[\frac{K_h^2\left(d(X_1,X')\right)}{\left(\mathbb{E}_{X'}\left[K_h\left(d(X_1,X')\right)\right]\right)^2}\right] \le \frac{|D|}{n}\frac{M_2}{m_1^2}\frac{1}{\varphi(h)},\end{aligned}$$

where the last inequality comes from Inequality (17). □

***Proof of Lemma 3*** Since $\mathbb{E}_{X'}[R_h^{X'}] = 1$, remark that,

$$\begin{aligned}\mathbb{E}\left[\left(R_h^{X'} - 1\right)^2 \mathbf{1}_B(X')\right] &= \mathbb{E}\left[\mathrm{Var}_{X'}\left(R_h^{X'}\right)\mathbf{1}_B(X')\right] \\ &= \frac{1}{n}\mathbb{E}\left[\mathrm{Var}_{X'}\left(\frac{K_h(d(X_1,X'))}{\mathbb{E}_{X'}\left[K_h(d(X_1,X'))\right]}\right)\mathbf{1}_B(X')\right],\end{aligned}$$

and the result comes also from Inequality (17).

*7.2.3. Proof of Lemma 4 (upper bound for the bias term)*

First remark that, for $y \in D$, a.s.

$$\mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}(y)\right] = n\mathbb{E}_{X'}\left[\mathbb{E}\left[\widetilde{W}_h^{(1)}(X')\mathbf{1}_{\{Y_1\le y\}}|X_1\right]\right] = n\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')F^{X_1}(y)\right]$$

and since $n\mathbb{E}_{X'}[\widetilde{W}_h^{(1)}(X')F^{X'}] = F^{X'}$,

$$F^{X'}(y) - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}(y)\right] = n\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left(F^{X'}(y) - F^{X_1}(y)\right)\right].$$

Then, thanks to the generalized Minkowski Inequality (see Lemma A.1 of Tsybakov [56]), which can be easily extended to the integral over Hilbert spaces

$$\begin{aligned}\mathbb{E}\left[\left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] &\le n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left\|F^{X_1} - F^{X'}\right\|_D\right]^2\right] \\ &\le C_D^2 n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left\|X_1 - X'\right\|^\beta\right]^2\right] \qquad (24)\end{aligned}$$

by $H_F$. Now, since $K$ is supported on $[0,1]$, if $d^2(X_1,X') = \|X_1 - X'\|^2 > h$ then $\widetilde{W}_h^{(1)}(X') = 0$,

$$\mathbb{E}\left[\left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] \le C_D^2 n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')h^\beta\right]^2\right].$$

But $\mathbb{E}_{X'}[\widetilde{W}_h^{(1)}(X')] = 1/n$, which ends the proof:

$$\mathbb{E}\left[\left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] \quad \leq \quad C_D^2 h^{2\beta}.$$

□

### 7.3. Proof of Theorem 2

#### 7.3.1. Main part of the proof of Theorem 2

Let $\Lambda$ be the set

$$\Lambda = \bigcap_{h\in\mathcal{H}_n} \left\{\left|\frac{\widehat{\varphi}(h)}{\varphi(h)} - 1\right| < \frac{1}{2}\right\}.$$

We split the loss function of the estimator

$$\left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 \leq \left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 (\mathbf{1}_\Lambda + \mathbf{1}_{\Lambda^c}).$$

We will argue as follows, with two steps. First, on the set $\Lambda$, $\widehat{\varphi}(h)$ is close to $\varphi(h)$, and the proof comes down to the control of the estimator of the bias approximation of the criterion (10). Second, the probability of the set $\Lambda^c$ is shown to be negligible. Let us prove these two claims.

• **Step 1. Upper bound for $\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\|_D^2 \mathbf{1}_\Lambda$.** Let $h \in \mathcal{H}_n$ be fixed. We start with the following decomposition for the loss of the estimator $\widehat{F}_{\widehat{h}}^{X'}$:

$$\left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 \quad \leq \quad 3\left\|\widehat{F}_{\widehat{h}}^{X'} - \widehat{F}_{\widehat{h}\vee h}^{X'}\right\|_D^2 + 3\left\|\widehat{F}_{\widehat{h}\vee h}^{X'} - \widehat{F}_h^{X'}\right\|_D^2 + 3\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2.$$

The definitions of $\widehat{A}(h)$, $\widehat{A}(\widehat{h})$ and then the one of $\widehat{h}$ enable to write

$$\begin{aligned} 3\left\|\widehat{F}_{\widehat{h}}^{X'} - \widehat{F}_{\widehat{h}\vee h}^{X'}\right\|_D^2 + 3\left\|\widehat{F}_{\widehat{h}\vee h}^{X'} - \widehat{F}_h^{X'}\right\|_D^2 &\leq 3\big(\widehat{A}(h) + \widehat{V}(\widehat{h})\big) + 3\big(\widehat{A}(\widehat{h}) + \widehat{V}(h)\big), \\ &\leq 6\big(\widehat{A}(h) + \widehat{V}(h)\big). \end{aligned}$$

Thus,

$$\left\|\widehat{F}_{\widehat{h}}^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_\Lambda \leq \left\{6\widehat{A}(h) + 6\widehat{V}(h) + 3\left\|\widehat{F}_h^{X'} - F^{X'}\right\|_D^2\right\}\mathbf{1}_\Lambda.$$

We now split $\widehat{A}(h)$ to introduce the following intermediary

$$V(h) = \frac{2}{3}\kappa\frac{\ln(n)}{n\varphi(h)} \text{ and } A(h) = \max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 - V(h')\right)_+. \tag{25}$$

Compared to the data-driven criterion (10), the variance term $V(h)$ is deterministic here. Note that

$$\begin{aligned}\widehat{A}(h) &= \max_{h'\in\mathcal{H}_n,\,\widehat{V}(h')<\infty}\left\{\left\|\widehat{F}^{X'}_{h\vee h'}-\widehat{F}^{X'}_{h'}\right\|^2_D - V(h')+\left(V(h')-\widehat{V}(h')\right)\right\}_+,\\ &\le \max_{h'\in\mathcal{H}_n,\,\widehat{V}(h')<\infty}\left\{\left\|\widehat{F}^{X'}_{h\vee h'}-\widehat{F}^{X'}_{h'}\right\|^2_D - V(h')\right\}\\ &\quad+\max_{h'\in\mathcal{H}_n,\,\widehat{V}(h')<\infty}\left(V(h')-\widehat{V}(h')\right)_+,\\ &\le A(h)+\max_{h'\in\mathcal{H}_n}\left(V(h')-\widehat{V}(h')\right)_+.\end{aligned}$$

We obtain the decomposition

$$\begin{aligned}\left\|\widehat{F}^{X'}_{\widehat{h}}-F^{X'}\right\|^2_D\mathbf{1}_\Lambda &\le \left\{6A(h)+6V(h)+3\left\|\widehat{F}^{X'}_{h}-F^{X'}\right\|^2_D\right.\\ &\quad\left.+6\max_{h'\in\mathcal{H}_n}\left(V(h')-\widehat{V}(h')\right)_+ +6\left(\widehat{V}(h)-V(h)\right)\right\}\mathbf{1}_\Lambda.\end{aligned}\tag{26}$$

For $h'\in\mathcal{H}_n$, such that $\widehat{V}(h')<\infty$, we have

$$V(h')-\widehat{V}(h')=\frac{2}{3}\kappa\frac{\ln(n)}{n}\left(\frac{1}{\varphi(h')}-\frac{3}{2}\frac{1}{\widehat{\varphi}(h')}\right).$$

But on the set $\Lambda$, for any $h'\in\mathcal{H}_n$, $|\widehat{\varphi}(h')-\varphi(h')|<\varphi(h')/2$. In particular, we thus have $\widehat{\varphi}(h')-\varphi(h')<\varphi(h')/2$, that is $\widehat{\varphi}(h')<(3/2)\varphi(h')$. This proves that $V(h')-\widehat{V}(h')<0$, and hence

$$\max_{h'\in\mathcal{H}_n}\left(V(h')-\widehat{V}(h')\right)_+=0.$$

Moreover, on $\Lambda$, we also have, for $h\in\mathcal{H}_n$, $\varphi(h)-\widehat{\varphi}(h)<\varphi(h)/2$, that is $2/\varphi(h)>1/\widehat{\varphi}(h)$. Thus,

$$\widehat{V}(h)-V(h)=\frac{2}{3}\kappa\frac{\ln(n)}{n}\left(\frac{3}{2}\frac{1}{\widehat{\varphi}(h)}-\frac{1}{\varphi(h)}\right)\le\frac{2}{3}\kappa\frac{\ln(n)}{n}2\frac{1}{\varphi(h)}=2V(h).$$

Gathering the two bounds in (26) leads to

$$\left\|\widehat{F}^{X'}_{\widehat{h}}-F^{X'}\right\|^2_D\mathbf{1}_\Lambda\le 6A(h)+18V(h)+3\left\|\widehat{F}^{X'}_{h}-F^{X'}\right\|^2_D.$$

Besides, the quantity $\|\widehat{F}^{X'}_{h}-F^{X'}\|^2_D$ is the loss of an estimator with fixed bandwidth $h$ and has already been bounded (see Theorem 1 Inequality (7)). Hence we obtain

$$\mathbb{E}\left[\left\|\widehat{F}^{X'}_{\widehat{h}}-F^{X'}\right\|^2_D\mathbf{1}_\Lambda\mathbf{1}_B(X')\right]\le 6\mathbb{E}[A(h)\mathbf{1}_B(X')]+18V(h)+3C\left(h^{2\beta}+\frac{1}{n\varphi(h)}\right),\tag{27}$$

where $C$ is the constant of Theorem 1 (Inequality (7)). The remainding part of the step is the result of the lemma hereafter, the proof of which is postponed to the following section.

**Lemma 5.** *Let $h \in \mathcal{H}_n$ be fixed. Under the assumptions of Theorem 2, there exist two constants $C_1$ and $C_2$ such that,*

$$\mathbb{E}\left[A(h)\mathbf{1}_B(X')\right] \leq C_1 h^{2\beta} + \frac{C_2}{n}. \tag{28}$$

*The constant $C_2$ depends on $C_0$, $|D|$, $M_2$, $m_1$ and $C_K$ and the constant $C_1$ only depends on $C_D$.*

• **Step 2. Upper bound for $\|\widehat{F}^{X'}_{\widehat{h}} - F^{X'}\|^2_D \mathbf{1}_{\Lambda^c}$.** We roughly bound

$$\left\|\widehat{F}^{X'}_{\widehat{h}} - F^{X'}\right\|^2_D \mathbf{1}_{\Lambda^c} \leq 2\left(\left\|\widehat{F}^{X'}_{\widehat{h}}\right\|^2_D + \left\|F^{X'}\right\|^2_D\right)\mathbf{1}_{\Lambda^c} \leq 4|D|\mathbf{1}_{\Lambda^c}.$$

It remains to control $\mathbb{P}(\Lambda^c)$:

$$\begin{aligned}\mathbb{P}(\Lambda^c) &\leq \sum_{h\in\mathcal{H}_n} \mathbb{P}\left(|\widehat{\varphi}(h) - \varphi(h)| \geq \frac{\varphi(h)}{2}\right),\\ &= \sum_{h\in\mathcal{H}_n} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{d(0,X_i)\leq h\}} - \mathbb{E}\left[\mathbf{1}_{\{d(0,X_i)\leq h\}}\right]\right| \geq \frac{\varphi(h)}{2}\right).\end{aligned}$$

We apply Bernstein's Inequality (Lemma 9), with $T_i = \mathbf{1}_{\{d(0,X_i)\leq h\}}$ and $\eta = \varphi(h)/2$. Since $0 \leq T_i \leq 1$, we set $b_0 = 1$, and $v^2 = \mathrm{Var}(T_1) = \varphi(h)(1-\varphi(h))$. We derive

$$\begin{aligned}\mathbb{P}(\Lambda^c) &\leq 2\sum_{h\in\mathcal{H}_n} \exp\left(-\frac{n\varphi^2(h)/8}{\varphi(h)(1-\varphi(h)) + \varphi(h)/2}\right)\\ &= 2\sum_{h\in\mathcal{H}_n} \exp\left(-\frac{n\varphi(h)}{8(1-\varphi(h)) + 4}\right) \leq 2\sum_{h\in\mathcal{H}_n} \exp\left(-\frac{n\varphi(h)}{12}\right).\end{aligned}$$

We thus obtain

$$\mathbb{E}\left[\left\|\widehat{F}^{X'}_{\widehat{h}} - F^{X'}\right\|^2_D \mathbf{1}_{\Lambda^c}\mathbf{1}_B(X')\right] \leq 8|D| \sum_{h\in\mathcal{H}_n} \exp\left(-\frac{n\varphi(h)}{12}\right) \leq 8|D|n^{1-C_0/12}. \tag{29}$$

thanks to Assumptions $H_{b1}$ and $H_{b2}$. Taking $C_0 > 24$, we have $n^{1-C_0/12} < n^{-1}$. Thus (29) leads to $\mathbb{E}[\|\widehat{F}^{X'}_{\widehat{h}} - F^{X'}\|^2_D \mathbf{1}_{\Lambda^c}\mathbf{1}_B(X')] \leq 8|D|/n$. It remains to gather this bound with (27), to apply Lemma 5 and to take the infimum over $h \in \mathcal{H}_n$. This ends the proof of Theorem 2. □

*7.3.2. Proof of Lemma 5 (Upper bound for $A(h)$)*

Fix $h, h' \in \mathcal{H}_n$. We define the set $\Omega_{h,h'} = \{R^{X'}_{h'} \geq 1/2\} \cap \{R^{X'}_{h\vee h'} \geq 1/2\}$ and split

$$\left\|\widehat{F}^{X'}_{h'} - \widehat{F}^{X'}_{h\vee h'}\right\|^2_D \leq \left\|\widehat{F}^{X'}_{h'} - \widehat{F}^{X'}_{h\vee h'}\right\|^2_D \left(\mathbf{1}_{\Omega_{h,h'}} + \mathbf{1}_{\Omega^c_{h,h'}}\right).$$

Recall that we write the estimator $\widehat{F}_h^{X'}(y) = \widetilde{F}_h^{X'}(y)/R_h^{X'}$, with $\widetilde{F}_h^{X'}$ defined by (22) and $R_h^{X'}$ by (18). We split again

$$\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}} = \left\|\frac{\widetilde{F}_{h'}^{X'}}{R_{h'}^{X'}} - \frac{\widetilde{F}_{h\vee h'}^{X'}}{R_{h\vee h'}^{X'}}\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}} \leq 4\left(T_{h'}^a + B_{h,h'} + \widetilde{T}_{h\vee h'} + T_{h\vee h'}^b\right),$$

where

$$\begin{aligned} T_{h'}^a &= \left\|\frac{1}{R_{h'}^{X'}}\left(\widetilde{F}_{h'}^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right]\right)\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}}, \\ T_{h\vee h'}^b &= \left\|\tfrac{1}{R_{h\vee h'}^{X'}}\left(\widetilde{F}_{h\vee h'}^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right)\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}} \\ B_{h,h'} &= \frac{1}{\left(R_{h'}^{X'}\right)^2}\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}}, \\ \widetilde{T}_{h\vee h'} &= \left(\frac{1}{R_{h'}^{X'}} - \frac{1}{R_{h\vee h'}^{X'}}\right)^2 \left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}}. \end{aligned} \tag{30}$$

Thus, by subtracting $V(h')$ and taking the maximum over $h' \in \mathcal{H}_n$, we obtain

$$\begin{aligned} A(h) &= \max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 - V(h')\right)_+ \\ &\leq \max_{h'\in\mathcal{H}_n}\left(4T_{h'}^a - \frac{V(h')}{3}\right)_+ + \max_{h'\in\mathcal{H}_n}\left(4T_{h\vee h'}^b - \frac{V(h')}{3}\right)_+ \\ &\quad + \max_{h'\in\mathcal{H}_n}\left(4\widetilde{T}_{h\vee h'} - \frac{V(h')}{3}\right)_+ \\ &\quad + 4\max_{h'\in\mathcal{H}_n} B_{h,h'} + \max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 \mathbf{1}_{\Omega_{h,h'}^c}\right). \end{aligned} \tag{31}$$

We have not subtracted $V(h')$ to two of the above terms: we show below that we can bound these terms directly. We now deal with each of the terms involving in (31) on the set $\{X' \in B\}$.

• **Upper bound for the term depending on** $B_{h,h'}$. Since $B_{h,h'} = 0$ if $h' > h$, we first have $\max_{h'\in\mathcal{H}_n} B_{h,h'} = \max_{h'\in\mathcal{H}_n, h'\leq h} B_{h,h'}$. We then use the definition of the set $\Omega_{h,h'}$, and split the term to obtain the bias terms:

$$\begin{aligned} \max_{h'\in\mathcal{H}_n} B_{h,h'} &\leq 4\max_{\substack{h'\in\mathcal{H}_n \\ h'\leq h}}\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2 \\ &\leq 8\max_{\substack{h'\in\mathcal{H}_n \\ h'\leq h}}\left\{\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - F^{X'}\right\|_D^2 + \left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2\right\} \\ &\leq 8\left(\max_{\substack{h'\in\mathcal{H}_n \\ h'\leq h}} C_D^2(h')^{2\beta} + C_D^2 h^{2\beta}\right) \leq 16 C_D^2 h^{2\beta}, \end{aligned} \tag{32}$$

thanks to Lemma 4.

• **Upper bound for the term depending on $\mathbf{1}_{\Omega^c_{h,h'}}$.** It is the second term which does not depend on $V(h')$:

$$\max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}^{X'}_{h'}-\widehat{F}^{X'}_{h\vee h'}\right\|^2_D\mathbf{1}_{\Omega^c_{h,h'}}\right) \leq 2\max_{h'\in\mathcal{H}_n}\mathbf{1}_{\Omega^c_{h,h'}}\left(\left\|\widehat{F}^{X'}_{h'}\right\|^2_D+\left\|\widehat{F}^{X'}_{h\vee h'}\right\|^2_D\right).$$

Thus, since $|\widehat{F}^{X'}_{h'}(y)|\leq 1$ and $\left|\widehat{F}^{X'}_{h'\vee h}(y)\right|\leq 1$,

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}^{X'}_{h'}-\widehat{F}^{X'}_{h\vee h'}\right\|^2_D\mathbf{1}_{\Omega^c_{h,h'}}\mathbf{1}_B(X')\right)\right]\leq 4|D|\sum_{h'\in\mathcal{H}_n}\mathbb{P}\left(\Omega^c_{h,h'}\cap\{X'\in B\}\right).$$

Moreover, $\mathbb{P}(\Omega^c_{h,h'}\cap\{X'\in B\})\leq[\mathbb{P}(\{|R^{X'}_{h'}-1|>\frac{1}{2}\}\cap\{X'\in B\})+\mathbb{P}(\{|R^{X'}_{h'\vee h}-1|>\frac{1}{2}\}\cap\{X'\in B\})]$. Thus by applying Inequality (19) of Lemma 1, with $\eta=1/2$:

$$\begin{aligned}\sum_{h'\in\mathcal{H}_n}\mathbb{P}\left(\Omega^c_{h,h'}\cap\{X'\in B\}\right) &\leq \sum_{h'\in\mathcal{H}_n}2\exp\left(-\frac{n\varphi(h')}{8\left(\frac{M_2}{m_1^2}+\frac{C_K}{2m_1}\right)}\right)\\ &\quad+2\exp\left(-\frac{n\varphi(h\vee h')}{8\left(\frac{M_2}{m_1^2}+\frac{C_K}{2m_1}\right)}\right).\end{aligned}$$

Recall now that thanks to $H_{b2}$, $\varphi(h)\geq C_0\ln(n)/n$ for all $h\in\mathcal{H}_n$, with $C_0>16(M_2/m_1^2+C_K/2m_1)$. Use also $H_{b1}$ to deduce

$$\sum_{h'\in\mathcal{H}_n}\mathbb{P}\left(\Omega^c_{h,h'}\cap\{X'\in B\}\right) \leq 4\times n\times n^{-\frac{C_0}{8\left(\frac{M_2}{m_1^2}+\frac{C_K}{2m_1}\right)}}<\frac{4}{n}.$$

Thus, we have proved that

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}^{X'}_{h'}-\widehat{F}^{X'}_{h\vee h'}\right\|^2_D\mathbf{1}_{\Omega^c_{h,h'}}\right)\mathbf{1}_B(X')\right]\leq\frac{16|D|}{n}. \tag{33}$$

• **Upper bound for the term depending on $\widetilde{T}_{h,h'}$.** The definition of this term implies that

$$\begin{aligned}\widetilde{T}_{h,h'} &= \left(\frac{R^{X'}_{h\vee h'}-R^{X'}_{h'}}{R^{X'}_{h'}R^{X'}_{h\vee h'}}\right)^2\left\|\mathbb{E}_{X'}\left[\widetilde{F}^{X'}_{h\vee h'}\right]\right\|^2_D\mathbf{1}_{\Omega_{h,h'}},\\ &\leq 16|D|\left(R^{X'}_{h\vee h'}-R^{X'}_{h'}\right)^2,\\ &\leq 32|D|\left\{\left(R^{X'}_{h\vee h'}-1\right)^2+\left(R^{X'}_{h'}-1\right)^2\right\},\end{aligned}$$

using that $\mathbb{E}[\widetilde{F}^{X'}_{h\vee h'}] \leq 1$. We roughly bound the supremum over $h' \in \mathcal{H}_n$ by a sum over $h'$ and use the last inequality:

$$\begin{aligned}
&\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4\widetilde{T}_{h\vee h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \\
&\quad\leq 4\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(\widetilde{T}_{h\vee h'} - \frac{V(h')}{12}\right)_+ \mathbf{1}_B(X')\right], \\
&\quad\leq 4\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(32|D|\left\{\left(R^{X'}_{h\vee h'}-1\right)^2 + \left(R^{X'}_{h'}-1\right)^2\right\} - \frac{V(h')}{12}\right)_+ \mathbf{1}_B(X')\right], \\
&\quad\leq 4\left\{\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(32|D|\left(R^{X'}_{h\vee h'}-1\right)^2 - \frac{V(h')}{24}\right)_+ \mathbf{1}_B(X')\right]\right. \\
&\qquad\left. + \sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(32|D|\left(R^{X'}_{h'}-1\right)^2 - \frac{V(h')}{24}\right)_+ \mathbf{1}_B(X')\right]\right\}, \\
&\quad\leq 128|D|\left\{\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(\left(R^{X'}_{h\vee h'}-1\right)^2 - \frac{V(h')}{768|D|}\right)_+ \mathbf{1}_B(X')\right]\right. \\
&\qquad\left. + \sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(\left(R^{X'}_{h'}-1\right)^2 - \frac{V(h')}{768|D|}\right)_+ \mathbf{1}_B(X')\right]\right\}.
\end{aligned}$$

Then, Inequality (20) of Lemma 1 (with $\alpha = 2$) proves that, on the set $\{X' \in B\}$, a.s.,

$$\mathbb{E}_{X'}\left[\left(\left(R^{X'}_{h'}-1\right)^2 - V_R(h')\right)_+\right] \leq \min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n^2},$$

with $V_R(h') = \kappa_R \ln(n)/n\varphi(h')$ and $\kappa_R > \max(8M_2/m_1^2, 128C_K^2/m_1^2C_0)$. Choosing $2\kappa/3 > 768|D|\kappa_R$ in the definition of $V(h')$ (see (25)) leads to $V(h')/768|D| \geq V_R(h')$, and hence we also have

$$\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(\left(R^{X'}_{h'}-1\right)^2 - \frac{V(h')}{768|D|}\right)_+ \mathbf{1}_B(X')\right] \leq \min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n},$$

thanks to Assumption $H_{b1}$. Since $\varphi(h) \leq \varphi(h\vee h')$, $V(h') \geq V(h\vee h')$, the other term is bounded as follows

$$\begin{aligned}
&\mathbb{E}\left[\left(\left(R^{X'}_{h\vee h'}-1\right)^2 - \frac{V(h')}{768|D|}\right)_+ \mathbf{1}_B(X')\right] \\
&\qquad\leq \mathbb{E}\left[\left(\left(R^{X'}_{h\vee h'}-1\right)^2 - \frac{V(h\vee h')}{768|D|}\right)_+ \mathbf{1}_B(X')\right],
\end{aligned}$$

and same computations allow to deal with it. We thus deduce that

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4\widetilde{T}_{h\vee h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \leq 256|D|\min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n}. \quad (34)$$

• **Upper bound for the terms depending on $T^a_{h'}$ or $T^b_{h'}$.** First, by definition of $\Omega_{h,h'}$, $T^a_{h'} \leq 4\|\widetilde{F}^{X'}_{h'} - \mathbb{E}_{X'}[\widetilde{F}^{X'}_{h'}]\|^2_D$. Furthermore, noticing that $\widetilde{F}^{X'}_{h'}$ belongs to $L^2(D)$, the following equality is classical:

$$\left\|\widetilde{F}^{X'}_{h'} - \mathbb{E}_{X'}[\widetilde{F}^{X'}_{h'}]\right\|^2_D = \sup_{t\in\bar{S}_D(0,1)} \langle \widetilde{F}^{X'}_{h'} - \mathbb{E}_{X'}[\widetilde{F}^{X'}_{h'}], t\rangle^2_D, \tag{35}$$

where $\bar{S}_D(0,1)$ is a dense countable subset of the sphere $\mathcal{S}(0,1) = \{t \in L^2(D), \|t\|_D = 1\}$. Indeed, denoting $v = \widetilde{F}^{X'}_{h'} - \mathbb{E}_{X'}[\widetilde{F}^{X'}_{h'}]$, for any $t \in L^2(D)$, the Cauchy-Schwarz Inequality leads to $\sup_{t\in\mathcal{S}(0,1)}\langle v, t\rangle^2_D \leq \|v\|_D$, with equality when $t = v/\|v\|$. We can replace $\mathcal{S}(0,1)$ by $\bar{S}_D(0,1)$ thanks to the separability of $L^2(D)$, which gives (35). Moreover, we write the scalar product $\langle \widetilde{F}^{X'}_{h'} - \mathbb{E}_{X'}[\widetilde{F}^{X'}_{h'}], t\rangle_D \mathbf{1}_B(X') = \nu_{n,h}(t)$, for $t \in \bar{S}_D(0,1)$, where

$$\begin{aligned}\nu_{n,h}(t) &= \frac{1}{n}\sum_{i=1}^n \psi_{t,h}(X_i, Y_i) - \mathbb{E}_{X'}\left[\psi_{t,h}(X_i, Y_i)\right], \\ &\text{with } \psi_{t,h}(X_i, Y_i) = \frac{K_h(d(X_i, X'))}{\mathbb{E}_{X'}[K_h(d(X_i, X'))]}\langle \mathbf{1}_{[Y_i;\infty[}, t\rangle_D \mathbf{1}_B(X').\end{aligned} \tag{36}$$

Consequently,

$$\begin{aligned}&\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4T^a_{h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \\ &\leq 16\sum_{h'\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}_D(0,1)}\nu^2_{n,h'}(t) - \frac{V(h')}{48}\right)_+ \mathbf{1}_B(X')\right].\end{aligned}$$

We use the following lemma, which permits to control the empirical process defined by (36). Its proof can be found in Section A.

**Lemma 6.** *Under the assumptions of Theorem 2, for $\delta_0 > \max(3528C_K^2|D|/M_2C_0, 12)$, there exists a constant $C_3 > 0$ (depending only on $m_1$, $M_2$, $\delta_0$, $C_0$ and $|D|$) such that*

$$\sum_{h\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}_D(0,1)}\nu^2_{n,h}(t) - 6\delta_0\frac{|D|M_2}{m_1^2}\frac{\ln(n)}{n\varphi(h)}\right)_+ \mathbf{1}_B(X')\right] \leq \frac{C_3}{n}.$$

Choosing $2\kappa/3 > 288\delta_0|D|M_2/m_1^2$ in the definition of $V(h')$ (see (25)) leads to $V(h')/48 \geq 6\delta_0|(D|M_2/m_1^2)\ln(n)/(n\varphi(h))$. This proves that

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4T^a_{h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \leq \frac{16C_3}{n}. \tag{37}$$

Recall finally that $V(h') \geq V(h\vee h')$, similar computations allow to also obtain the same bound for $\mathbb{E}[\max_{h'\in\mathcal{H}_n}(4T^b_{h\vee h'} - V(h')/3)_+]$.

Gathering Inequalities (32), (33), (34), and (37) in Inequality (31) completes the proof of Lemma 5. □

### *7.4. Proof of Theorem 3*

#### *7.4.1. Proof of (i), under Assumption $H_{X,L}$*

***Sketch of the proof*** The proof is based on the general reduction scheme described in Section 2.2 of Tsybakov [56]. We only describe in this section the main steps, whose proofs are given in Section A.2. Let $x_0 \in \mathbb{H}$ be fixed and $r_n = (\ln(n))^{-\beta/\alpha}$ the rate of convergence. We define two functions $F_0$ and $F_1$, called hypotheses, such that

(A) $F_l$ belongs to $\mathcal{F}_\beta$, for $l = 0, 1$,
(B) $\|F_0^{x_0} - F_1^{x_0}\|_D^2 \geq cr_n$ for a constant $c > 0$,
(C) $K(\mathbb{P}_1^{\otimes n}, \mathbb{P}_0^{\otimes n}) \leq \alpha$ for a real number $\alpha < \infty$ (which does not depend on $x_0$), where $\mathbb{P}_0^{\otimes n}$ (resp. $\mathbb{P}_1^{\otimes n}$) is the probability distribution of a sample $(X_{0,i}, Y_{0,i})_{i=1,\ldots,n}$ (resp. $(X_{1,i}, Y_{1,i})_{i=1,\ldots,n}$) for which the conditional c.d.f. of $Y_{0,i} \in \mathbb{R}$ given $X_{0,i} \in \mathbb{H}$ (resp. of $Y_{1,i}$ given $X_{1,i}$) is $F_0$ (resp. $F_1$). $K(P, Q)$ is the Kullback distance between two probability distributions $P$ and $Q$: $K(P, Q) = \int \ln(dP/dQ)dP$ if $P << Q$, and $K(P, Q) = +\infty$ otherwise.

Then, thanks to Theorem 2.2 in Tsybakov [56] (p. 90), the results hold for any $x_0$. Moreover, the multiplicative constant $c'$ involved in Theorem 2.2 is clearly independent on $x_0$. In the sequel, we define $F_0$ and $F_1$. The three conditions are checked in Section A.2.

***Construction of $F_0$ and $F_1$ and of the associated samples*** For $(x, y) \in \mathbb{H} \times \mathbb{R}$, let $F_0^x$ be the c.d.f. of the uniform distribution on $D$, that is $F_0^x(y) = \frac{y}{|D|}\mathbf{1}_{y\in D} + \mathbf{1}_{y>\sup D}$. Choose a real random variable $Y_0$ with a uniform distribution $\mathbb{P}_{\mathcal{U}_D}$ on the compact set $D$, and take any process $X_0$ on $\mathbb{H}$, independent on $Y_0$, with distribution $\mathbb{P}_X$ verifying $H_{X,L}$. For the second function, set

$$F_1^x(y) = F_0^x(y) + L\eta_n^\beta H\left(\frac{\|x - x_0\|}{\eta_n}\right)\int_{-\infty}^{y} \psi(t)dt,$$

where

- $\psi : \mathbb{R} \to \mathbb{R}$ is a non-zero continuous function with support $D$ with $\int_{\mathbb{R}} \psi(t)dt = 0$.
- $H : \mathbb{R}_+ \to \mathbb{R}_+$ is a function with support $[0; 1]$ such that $|H(u) - H(v)| \leq |u - v|^\beta$, for any $(u, v) \in \mathbb{R}_+^2$.
- $L$ is a real number such that $0<L<1/(\sup_{n\in\mathbb{N}^*}\{\eta_n^\beta\}|D|\|K\|L^\infty(\mathbb{R})\|\psi\|_{L^\infty(\mathbb{R})})$.
- $\eta_n$ is a non-negative real number such that

$$\eta_n^{2\beta} \geq c_{(B)}r_n \text{ and } \eta_n^{2\beta}\varphi^{x_0}(\eta_n) \leq \frac{c_{(C)}}{n}, \tag{38}$$

for two constants $c_{(B)} > 0$ and $c_{(C)} > 0$.

From $H_{X,L}$, a positive number $\eta_n$ for which the properties above hold is given by

$$\eta_n = \left( \frac{\ln n - ((2\beta+\gamma)/\alpha) \ln \ln n}{C_1} \right)^{-1/\alpha}. \tag{39}$$

We also choose a variable $Y_1$, such that, for any $x \in \mathbb{H}$, the conditional distribution of $Y_1$ given $X_0 = x$ is characterized by the c.d.f. $F_1^x$. The notation $\mathbb{P}_1$ is the distribution of $(X_0, Y_1)$.

### *7.4.2. Proof of (i), under Assumption $H_{X,M}$ or $H_{X,F}$*

The proofs exactly follow the same scheme as for (i) under $H_{X,L}$. The only difference is the choice of the sequence $(\eta_n)_n$ (see (39)).

- Under $H_{X,M}$, we set $r_n = \exp(-\frac{2\beta}{c_1^{1/\alpha}} \ln^{1/\alpha} n)$, and replace the previous choice (39) of $\eta_n$ by $\eta_n = \exp(-(c_1^{-1} \ln n - c_1^{-(\alpha+1)/\alpha}(2\beta+\gamma_2) \ln^{1/\alpha} n)^{1/\alpha})$. It verifies both of the required conditions (38).
- The case $H_{X,F}$ is the extreme case $\alpha = 1$ in $H_{X,M}$. We set $\eta_n = n^{1/(2\beta+\gamma)}$ and attain the lower bound $r_n = n^{-2\beta/(2\beta+\gamma)}$. □

### *7.4.3. Proof of (ii)*

The risk (5) is an integral w.r.t the measure $\mathbb{P}_X \otimes \mathbb{P}_{\mathcal{U}_D}$ where $\mathbb{P}_{\mathcal{U}_D}$ is the uniform distribution on the set $D$. The tools defined to prove (i) are useful and we refer to it. But it cannot be straightforwardly adapted, since for an integrated criterion, two hypotheses are not sufficient. We focus on the case of Assumption $H_{X,L}$ (the switch to Assumption $H_{X,M}$ and $H_{X,F}$ is the same as in (i)). Denote by $r_n = (\ln(n))^{-2\beta/\alpha}$ the rate of convergence again. We must build a set of functions $(F_\omega)_{\omega \in \Omega_n}$ where $\Omega_n$ is a non-empty subset of $\{0,1\}^{m_n}$ and $m_n$ is a positive integer which will be precised later, such that,

(A') $F_\omega$ belongs to $\mathcal{F}_\beta$, for all $\omega \in \Omega_n$,

(B') For all $\omega, \omega' \in \Omega_n$, $\omega \neq \omega'$, $\mathbb{E}[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')] \geq c r_n$ where $c > 0$ is a constant,

(C') For all $\omega \in \Omega_n$, $\mathbb{P}_\omega$ is absolutely continuous with respect to $\mathbb{P}_0$ and

$$\frac{1}{\mathrm{Card}(\Omega_n)} \sum_{\omega \in \Omega_n} K(\mathbb{P}_\omega^{\otimes n}, \mathbb{P}_0^{\otimes n}) \leq \zeta \ln(\mathrm{Card}(\Omega_n))$$

for a real number $\zeta \in ]0, 1/8[$, where $\mathbb{P}_\omega^{\otimes n}$ is the probability distribution of a sample $(X_{\omega,i}, Y_{\omega,i})_{i=1,\dots,n}$ for which the conditional c.d.f. of $Y_{\omega,i}$ given $X_{\omega,i}$ is given by $F_\omega$.

Then the result comes from Theorem 2.5 of Tsybakov [56] (p. 85–86). We follow the same steps as previously: below, we define the set of hypotheses $(F_\omega)_{\omega \in \Omega_n}$, and in Section A.2, we give some additional material to check conditions (A'), (B') and (C').

***Construction of the set of hypotheses $F_\omega$ and of the associated samples*** The first function $(x,y)\mapsto F_0^x(y)$ is defined as in the proof of (i). For all $\omega=(\omega_1,\dots,\omega_{m_n})\in\{0,1\}^{m_n}$, let

$$F_\omega^x(y)=F_0^x(y)+L\eta_n^\beta\int_{-\infty}^{y}\psi(t)dt\sum_{k=1}^{m_n}\omega_k H\left(\frac{\|x-x_k\|}{\eta_n}\right),$$

where $\psi$, $H$, $L$, and $(\eta_n)_n$ are introduced in the proof of (i) (a good choice of $\eta_n$ is (39)), and with $x_j=\sqrt{2}\sup_{n\in\mathbb{N}^*}\{\eta_n\}e_j$, for all $j\geq 1$, where $(e_j)_{j\geq 1}$ is an orthonormal basis of $\mathbb{L}^2([0,1])$. Moreover $\psi$ verifies the following condition: $\|\psi\|^2_{\mathbb{L}^2(\mathbb{R})}<\ln(2)/(64C_2|D|\|H\|_{\mathbb{L}^2(\mathbb{R})}c_{(C)})$ (where $C_2$ appears in Assumption (14), p. 2364 and $c_{(C)}$ in (39)).

We also choose a variable $Y_\omega$, such that, for any $x\in\mathbb{H}$, the conditional distribution of $Y_\omega$ given $X=x$ is characterized by the c.d.f. $F_\omega^x$. The notation $\mathbb{P}_\omega$ is the distribution of $(X,Y_\omega)$.

Remark that the definition of $(x_j)_{j=1,\dots,m_n}$ implies that,

$$H\left(\frac{\|x-x_k\|}{\eta_n}\right)H\left(\frac{\|x-x_j\|}{\eta_n}\right)=0 \text{ for all } x\in\mathbb{H}, \text{ as soon as } j\neq k. \tag{40}$$

Indeed, suppose that $H(\|x-x_k\|/\eta_n)\neq 0$, since $H$ has support $[0,1]$, we have $\|x-x_k\|\leq\eta_n$. Now remark that, as $(e_j)_{j\geq 1}$ is an orthonormal basis, for all $j\neq k$, $\|x_j-x_k\|^2=2\sup_{n\in\mathbb{N}^*}\{\eta_n^2\}(\|e_j\|^2-2\langle e_j,e_k\rangle+\|e_k\|^2)=4\sup_{n\in\mathbb{N}^*}\{\eta_n^2\}$. Then $\|x-x_j\|\geq\|x_j-x_k\|-\|x-x_k\|\geq 2\sup_{n\in\mathbb{N}^*}\{\eta_n\}-\eta_n>\eta_n$ and $H(\|x-x_k\|/\eta_n)=0$. □

### *7.5. Proof of Proposition 2*

#### *7.5.1. Main part of the proof*

The proof starts like the proof of Theorem 1. For Inequality (ii), we first bound $\mathbb{E}[\|\widehat{F}_h^{X'}-F^{X'}\|_D^2\mathbf{1}_{\{R_h^{X'}<1/2\}}\mathbf{1}_B(X')]$ by $C/(n\varphi_p(h))$, with $d$ replaced by $d_p$ in the definition of $R_h^{X'}$. For $\mathbb{E}[\|\widehat{F}_h^{X'}-F^{X'}\|_D^2\mathbf{1}_{\{R_h^{X'}\geq 1/2\}}\mathbf{1}_B(X')]$ we obtain the splitting (23). Lemmas 2 and Lemmas 3 remain valid (by replacing again $d$ by $d_p$ in every terms, and by using $H'_\varphi$ instead of $H_\varphi$). This first part is also easily adapted to the proof of Inequality (i).

The difference lies in the control of the bias term. We substitute to Lemma 4 the following result, the proof of which can be found below. This ends the proof.

**Lemma 7.** *Suppose that Assumptions $H_F$ and $H_\xi$ are fulfilled. Then*

$$\mathbb{E}\left[\left\|F^{X'}-\mathbb{E}_{X'}\left[\widetilde{F}_{h,p}^{X'}\right]\right\|_D^2\right]\leq C\left(h^{2\beta}+\left(\sum_{j>p}\sigma_j^2\right)^\beta\right)$$

*and*

$$\mathbb{E}\left[\left\|F^{x_0}-\mathbb{E}\left[\widetilde{F}^{x_0}_{h,p}\right]\right\|_D^2\right]\leq C\left(h^{2\beta}+\left(\sum_{j>p}\sigma_j^2\right)^{\beta}+\left(\sum_{j>p}\langle x_0,e_j\rangle^2\right)^{\beta}\right)$$

*where $C>0$ only depends on $C_D$, $\beta$, and $C_\xi$.* □

*7.5.2. Proof of Lemma 7*

Let us begin with the first inequality (integrated risk). Like in the proof of Lemma 4, we also obtain (24). Then,

$$\mathbb{E}\left[\left\|F^{X'}-\mathbb{E}_{X'}\left[\widetilde{F}^{X'}_{h,p}\right]\right\|_D^2\right]$$
$$\leq C_D^2n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left(d_p^2(X_1,X')+\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\right],$$

where $\xi_j^{(1)}:=(\langle X_1,e_j\rangle-\mu_j)/\sigma_j$ and $\xi_j':=(\langle X_1,e_j\rangle-\mu_j)/\sigma_j$ are the standardized versions of $\langle X_1,e_j\rangle$ and $\langle X',e_j\rangle$. The same arguments as in Lemma 4 lead to

$$\mathbb{E}\left[\left\|F^{X'}-\mathbb{E}_{X'}\left[\widetilde{F}^{X'}_{h,p}\right]\right\|_D^2\right]$$
$$\leq C_D^2n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left(h^2+\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\right].$$

Now, firstly, for all $a,b>0$, $(a+b)^{\beta/2}\leq(2\max\{a,b\})^{\beta/2}\leq 2^{\beta/2}(a^{\beta/2}+b^{\beta/2})$ and secondly $\mathbb{E}_{X'}[\widetilde{W}_h^{(1)}(X')]=1/n$. We thus obtain

$$\mathbb{E}\left[\left\|F^{X'}-\mathbb{E}_{X'}\left[\widetilde{F}^{X'}_{h,p}\right]\right\|_D^2\right]$$
$$\leq 2^{\beta+1}C_D^2\left(h^{2\beta}+n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(i)}(X')\left(\sum_{j>p}\sigma_j^2(\xi_j^{(i)}-\xi_j')^2\right)^{\beta/2}\right]^2\right]\right). \tag{41}$$

Under Assumption $H_\xi^b$, the results comes from the following bound

$$\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left(\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\right]\leq 4^\beta C_\xi^{2\beta}\left(\sum_{j>p}\sigma_j^2\right)^{\beta}\frac{1}{n^2}.$$

Under Assumption $H_\xi^{ind}$, remark that

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\left(\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\right] \\
&\quad=\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\right]^2\mathbb{E}_{X'}\left[\left(\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\right] \\
&\quad\le n^{-2}\mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2.
\end{aligned}$$

Now, we use the following lemma, proved in Section A.

**Lemma 8.** *Let $(\eta_j)_{j\ge1}$ a sequence of real random variables and $(\sigma_j)_{j\ge1}$ a sequence of real numbers verifying, for $\beta>0$,*

$$\sum_{j\ge1}\sigma_j^2<+\infty \text{ and } \forall j\ge1,\ \mathbb{E}\left[\eta_j^\beta\right]\le C_M,$$

*for a constant $C_M>1$, then, for all $p\in\mathbb{N}$*

$$\mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2\eta_j^2\right)^{\beta/2}\right]\le C_M\left(\sum_{j>p}\sigma_j^2\right)^{\beta/2}.$$

With $\eta_j=\xi_j^{(1)}-\xi_j'$ and $C_M=2^\beta C_\xi$, we get

$$\mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2(\xi_j^{(1)}-\xi_j')^2\right)^{\beta/2}\right]^2\le 2^{4\beta}C_\xi^2\left(\sum_{j>p}\sigma_j^2\right)^\beta,$$

and the result comes from Inequality (41). The proof of the first inequality of Lemma 7 is completed.

For the second inequality (pointwise risk), the only difference is that, from (24), we rather use

$$\begin{aligned}
\|X_1-x_0\|^\beta &= \left(d_p^2(X_1,x_0)+\sum_{j>p}(\sigma_j\xi_j-\langle x_0,e_j\rangle)^2\right)^{\beta/2} \\
&\le 3^{\beta/2}\left(d_p^\beta(X_1,x_0)+2^{\beta/2}\left(\sum_{j>p}\sigma_j^2\xi_j^2\right)^{\beta/2}+2^{\beta/2}\left(\sum_{j>p}\langle x_0,e_j\rangle^2\right)^{\beta/2}\right).
\end{aligned}$$

The final bound then follows similarly. □

**Acknowledgements**

The authors want to warmly thank Élodie Brunel, Fabienne Comte, Jean-Claude Fort, André Mas and both referees for their helpful comments and careful reading which have significantly improved the presentation of this work.

## Appendix A: Additional proofs

### *A.1. Proofs of the convergence rates of the estimators (upper-bounds)*

#### *A.1.1. Proof of Proposition 1*

***Proof of (a)*** We have to compute the convergence under three regularity assumptions ($H_{X,L}$, $H_{X,M}$ and $H_{X,F}$), and for the two criteria (pointwise and integrated). It follows from (7) of Theorem 1 that the risk of the estimator is bounded by

$$\widetilde{R}(h) = h^{2\beta} + 1/(n\varphi^{x_0}(h)),$$

up to a multiplicative constant, and with $x_0 = 0$ for the integrated risk. To obtain the convergence rates, it is thus sufficient to compute the bandwidth $h$ which minimizes the bound $\widetilde{R}(h)$ when assuming $H_X$. If the minimum cannot be easily computed, we can also choose a "good" value for $h$.

• *Convergence rate under Assumption $H_{X,L}$* With the lower bound on $\varphi^{x_0}$ of $H_{X,L}$, the quantity $\widetilde{R}(h)$ and thus also the risks are upper bounded by a quantity with order of magnitude

$$R(h) := h^{2\beta} + h^{-\gamma} \exp\left(c_2 h^{-\alpha}\right) n^{-1}.$$

Choosing the bandwidth $h_0$ such that

$$h_0 = \left(\frac{\ln n}{c_2} - \kappa \ln\left(\frac{\ln n}{c_2}\right)\right)^{-1/\alpha},$$

with $\kappa := c_2^{-1}(\gamma/\alpha + 2\beta/\alpha)$ ends the proof, since $R(h_0)$ has the announced order.

• *Convergence rate under Assumption $H_{X,M}$ or $H_{X,F}$* First assume $H_{X,M}$. The risk is bounded (up to a multiplicative constant) by the quantity

$$R(h) = h^{2\beta} + n^{-1} h^{-\gamma_1} \exp\left(c_2 \ln^{\alpha}(1/h)\right).$$

Choosing

$$h_0 = \exp\left(-\left(\frac{1}{c_2}\ln n - c_2^{-(\alpha+1)/\alpha}\,(2\beta - \gamma_1)\ln^{1/\alpha} n\right)_+^{1/\alpha}\right),$$

leads to what we want to prove, that is

$$R(h_0) \leq C \exp\left( -\frac{2\beta}{c_2^{1/\alpha}} \ln^{1/\alpha} n \right). \tag{42}$$

When assuming $H_{X,F}$, the optimal $h$ can be computed: the one which minimizes $R(h) = h^{2\beta} + h^{-\gamma}$ has the order $n^{1/(2\beta+\gamma)}$ and immediatly gives $R(h) \leq Cn^{-2\beta/(2\beta+\gamma)}$. □

***Proof of (b)*** The proof comes down to the proof of (a) since Theorem 1 gives a bound of the risks of $\widehat{F}_{\widehat{h}}$ with the form $\min_h \widetilde{R}(h)$ ($\widetilde{R}(h)$ defined in the proof of (a)). The computation of the (a) bound for this minimum has thus been done in the previous section. □

### *A.1.2. Proof of Corollary 1*

The proof is based on the same ideas as the ones used to prove Proposition 1 in Section A.1.1. We begin with the result (ii) (integrated risk).

***Proof of*** (ii) Thanks to Proposition 2 (ii), the risk of the estimator is bounded by $h^{2\beta} + (\sum_{j>p} \sigma_j^2)^\beta + n^{-1}\varphi_p^{-1}(h)$, up to a multiplicative constant. Remark that

$$\varphi_p(h) = \int_{\left\{\mathbf{x}\in\mathbb{R}^p,\ \sum_{j=1}^p \sigma_j^2 x_j^2 \leq h^2\right\}} f_p(x_1, \ldots, x_p) d\mathbf{x},$$

where $f_p$ is the density of $(\xi_1, \ldots, \xi_p)$. By noticing that

$$\left\{ \mathbf{x} \in \mathbb{R}^p,\ \sum_{j=1}^p \sigma_j^2 x_j^2 \leq h^2 \right\} \supset \left\{ \mathbf{x} \in \mathbb{R}^p,\ |x_j| \leq \frac{h}{\sqrt{\sum_{j=1}^p \sigma_j^2}} \right\},$$

we get

$$\varphi_p(h) \quad \geq \quad 2^p h^p \int_{\left[0, \left(\sum_{j=1}^p \sigma_j^2\right)^{-1/2}\right]^p} f_p(hx_1, \ldots, hx_p) d\mathbf{x} \geq c^p h^p,$$

where $c$ only depends on $\sum_{j\geq 1} \sigma_j^2$ and $c_f$. With the assumption on $\sigma_j$, we thus obtain the following upper bound for the risk, up to a constant $R(h,p) := h^{2\beta} + p^{\beta(1-2\delta)} + c^{-p} n^{-1} h^{-p}$. Choosing $h = (\frac{pc^{-p}}{2\beta})^{1/(2\beta+p)} n^{-1/(2\beta+p)}$ and $p = [\ln(n)/(\delta - 1/2)\ln\ln(n) - 2\beta]$ gives the result.

***Proof of*** (i) We deduce from Proposition 2 (ii), from the assumption $\sum_{j>p} \langle x_0, e_j \rangle \leq C \sum_{j>p} \sigma_j^2$, and from the left-hand-side inequality of $H_\varphi$ that the risk is upper bounded by $h^{2\beta} + (\sum_{j>p} \sigma_j^2)^\beta + n^{-1}\varphi_p^{-1}(h)$, up to a multiplicative constant. Thus, the reasoning is the same as for (ii). □

### A.2. Supplementary details for the proof of Theorem 3

#### A.2.1. Additional material for the proof of (i), under Assumption $H_{X,L}$

All that remains to be done is to check that the hypotheses $F_0$ and $F_1$ defined in Section 7.4.1 satisfy the three conditions (A), (B) and (C) (also defined in Section 7.4.1).

***Check (A). Belonging to the space $\mathcal{F}_\beta$*** For any $x \in \mathbb{H}$, the function $F_0^x$ is a c.d.f. by construction (it does not depend on $x$ and is simply the c.d.f. of the uniform distribution on $D$), and $\|F_0^x - F_0^{x'}\|^2 = 0$ $(x, x' \in \mathbb{H})$. Thus, $F_0$ belongs to $\mathcal{F}_\beta$.

Let $x \in \mathbb{H}$ be fixed. The function $y \mapsto F_1^x(y)$ is continuous, with limit 0 when $y$ goes to $-\infty$ (recall that $D$ is a bounded set), and 1 when $y$ goes to $+\infty$ (since $\int_{\mathbb{R}} \psi(t)dt = 0$). If $y \notin \bar{D}$, $(F_1^x)'(y) = 0$ (the support of $\psi$ is included in $D$) and if $y \in \mathring{D}$,

$$\begin{aligned}(F_1^x)'(y) &= \frac{1}{|D|} + L\eta_n^\beta H\left(\frac{\|x - x_0\|}{\eta_n}\right)\psi(y)\\ &\geq \frac{1}{|D|} - L\eta_n^\beta \|H\|_{L^\infty(\mathbb{R})}\|\psi\|_{L^\infty(\mathbb{R})} > 0,\end{aligned}$$

thanks to the definition of $L$ above. Thus $F_1^x$ is increasing, and $F_1^x$ is a conditional distribution function. Moreover, for any $x, x' \in \mathbb{H}$, denoting by $I_\psi = \int_D (\int_{-\infty}^y \psi(t)dt)^2 dy$,

$$\begin{aligned}\|F_1^x - F_1^{x'}\|_D^2 &= L^2\eta_n^{2\beta} I_\psi \left(H\left(\frac{\|x - x_0\|}{\eta_n}\right) - H\left(\frac{\|x' - x_0\|}{\eta_n}\right)\right)^2,\\ &\leq L^2\eta_n^{2\beta} I_\psi \left(\frac{\|x - x_0\|}{\eta_n} - \frac{\|x' - x_0\|}{\eta_n}\right)^{2\beta} \leq L^2 I_\psi \|x - x'\|^{2\beta},\end{aligned}$$

thanks to the regularity property of the function $H$. Therefore, $F_1$ also belongs to $\mathcal{F}_\beta$.

***Check (B). Condition on the loss $\|F_0^{x_0} - F_1^{x_0}\|_D^2$*** We have, thanks to the lower bound for $\eta_n$,

$$\|F_1^{x_0} - F_0^{x_0}\|_D^2 = L^2\eta_n^{2\beta} H^2(0) I_\psi \geq L^2 H^2(0) I_\psi c_{(C)} r_n.$$

***Check (C). Upper bound for the Kullback divergence $K(P_1^{\otimes n}, P_0^{\otimes n})$*** In a first step, we prove that the measure $\mathbb{P}_1$ is absolutely continuous with respect to $\mathbb{P}_0$, and compute the Radon-Nikodym derivative. First, notice that

$$F_1^x(y) = \int_{-\infty}^y \frac{1}{|D|}\mathbf{1}_D(t) + L\eta_n^\beta H\left(\frac{\|x - x_0\|}{\eta_n}\right)\psi(t)dt.$$

Therefore, keeping in mind that $\int_{\mathbb{R}} \psi(t)dt = \int_D \psi(t)dt = 0$, the conditional distribution of $Y_1$ given $X_0 = x$ admits a density with respect to the Lebesgue measure on $D$ given by

$$f_1^x(y) = \left( \frac{1}{|D|} + L\eta_n^\beta H\left( \frac{\|x - x_0\|}{\eta_n} \right) \psi(y) \right) \mathbf{1}_D(y).$$

We can thus compute the distribution $\mathbb{P}_1$ of the random couple $(X_0, Y_1)$. By definition, we see that $\mathbb{P}_1 << \mathbb{P}_0$ and $d\mathbb{P}_1/d\mathbb{P}_0(x,y) = |D| f_1^x(y)$. This enables to compute the Kullback distance

$$\begin{aligned} K(\mathbb{P}_1, \mathbb{P}_0) &= \int \ln\left( \frac{d\mathbb{P}_1}{d\mathbb{P}_0} \right) d\mathbb{P}_1 = \int_{\mathbb{H}\times\mathbb{R}} \ln\left( |D| f_1^x(y) \right) f_1^x(y) dy d\mathbb{P}_{X_0}(x), \\ &= \mathbb{E}\left[ \int_D \ln\left( |D| f_1^{X_0}(y) \right) f_1^{X_0}(y) dy \right], \\ &= \mathbb{E}\left[ \int_D \ln\left( 1 + |D| L\eta_n^\beta H\left( \frac{\|X_0 - x_0\|}{\eta_n} \right) \psi(y) \right) \right. \\ &\quad \left. \times \left( \frac{1}{|D|} + L\eta_n^\beta H\left( \frac{\|X_0 - x_0\|}{\eta_n} \right) \psi(y) \right) dy \right]. \end{aligned}$$

Noting that $\ln(1+u) \leq u$ for every $u > -1$, we obtain

$$K(\mathbb{P}_1, \mathbb{P}_0) \leq |D| L^2 \eta_n^{2\beta} \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} \varphi^{x_0}(\eta_n),$$

by using successively that $\int_{\mathbb{R}} \psi(y)dy = 0$ and that the support of $H$ is $[0;1]$.

Thus, thanks to the definition of $\eta_n$, we get

$$K(\mathbb{P}_1, \mathbb{P}_0) \leq |D| L^2 \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} c_{(C)}/n,$$

and finally,

$$K(\mathbb{P}_1^{\otimes n}, \mathbb{P}_0^{\otimes n}) = nK(\mathbb{P}_1, \mathbb{P}_0) \leq |D| L^2 \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} c_{(C)},$$

which completes the proof of (C). □

### *A.2.2. Additional material for the proof of (ii)*

Similarly, it remains to check that the set of hypotheses $(F_\omega)_{\omega\in\Omega_n}$ satisfy the three conditions (A'), (B') and (C') defined in Section 7.4.3.

***Check (A')*** We have already checked that $F_0$ belongs to $\mathcal{F}_\beta$. Let $\omega \in \{0,1\}^{m_n}$ be fixed. To prove that $F_\omega^x$ is non increasing ($x \in \mathbb{H}$ fixed), as for $F_1^x$, we bound,

$$(F_\omega^x)'(y) \geq \frac{1}{|D|} - L \sup_{n\in\mathbb{N}^*} \{\eta_n^\beta\} \|H\|_{L^\infty(\mathbb{R})} \|\psi\|_{L^\infty(\mathbb{R})} > 0,$$

for $y \in D$, thanks to Property (40) and the definition of $L$ above. Thus, as $F_1$ in the proof of (i), $F_\omega$ is a conditional distribution function, and we also similarly obtain $F_\omega \in \mathcal{F}_\beta$.

***Check (B')*** For all $\omega, \omega' \in \{0,1\}^{m_n}$,

$$\mathbb{E}\left[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')\right] = L^2\eta_n^{2\beta} I_\psi \mathbb{E}\left[\left(\sum_{k=1}^{m_n}(\omega_k - \omega_k')H\left(\frac{\|X' - x_k\|}{\eta_n}\right)\right)^2 \mathbf{1}_B(X')\right],$$

with $I_\psi$ defined in the proof of (i). From Property (40), we get:

$$\mathbb{E}\left[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')\right] = L^2\eta_n^{2\beta} I_\psi \sum_{k=1}^{m_n}(\omega_k - \omega_k')^2 \mathbb{E}\left[H^2\left(\frac{\|X' - x_k\|}{\eta_n}\right)\mathbf{1}_B(X')\right].$$

Now set $c_H := \min_{x,\|x\|\le 1/2} H(x)$, since $H$ is continuous and $H(x) > 0$ for all $x \in \mathbb{H}$, such that $\|x\| \le 1$, we have $c_H > 0$ and

$$\begin{aligned}\mathbb{E}\left[H^2\left(\frac{\|X' - x_k\|}{\eta_n}\right)\mathbf{1}_B(X')\right] &\ge \mathbb{E}\left[H^2\left(\frac{\|X' - x_k\|}{\eta_n}\right)\mathbf{1}_{\left\{\frac{\|X'-x_k\|}{\eta_n}\le 1/2\right\}}\mathbf{1}_B(X')\right] \\ &\ge c_H^2 \mathbb{P}\left(\{\|X' - x_k\| \le \eta_n/2\} \cap \{X' \in B\}\right).\end{aligned}$$

Now recall that, by definition, $\|x_k\| = \sqrt{2}\sup_{n\in\mathbb{N}}\{\eta_n\}$, and that $B$ contains the ball of $\mathbb{H}$ centred at 0 and of radius $\tilde{\rho}$. Then, as soon as, $\tilde{\rho} > (1/2 + \sqrt{2})\sup_{n\in\mathbb{N}}\{\eta_n\}$, we have $\{\|X' - x_k\| \le \eta_n/2\} \subset \{\|X'\| \le \rho\} \subset \{X' \in B\}$. Then, since $\|x_k\| < \rho$, we also have $x_k \in B$ and we can apply Condition (14) to get a lower bound on the shifted small ball probability $\mathbb{P}(\|X' - x_k\| \le \eta_n/2) = \varphi^{x_k}(\eta_n/2)$. We get

$$\mathbb{E}\left[H^2\left(\frac{\|X' - x_k\|}{\eta_n}\right)\mathbf{1}_B(X')\right] \ge c_H^2 c_2 \varphi(\eta_n/2),$$

and

$$\mathbb{E}\left[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')\right] \ge L^2 c_H^2 c_2 \eta_n^{2\beta} I_\psi \varphi(\eta_n/2)\rho(\omega,\omega'),$$

where $\rho$ is the Hamming distance on $\{0,1\}^{m_n}$ defined by $\rho(\omega,\omega') = \sum_{j=1}^{m_n}\mathbf{1}_{\{\omega_k \ne \omega_k\}}$. Now, from Varshamov-Gilbert bound (Lemma 2.7 of Tsybakov [56]), there exists a subset $\Omega_n$ of $\{0,1\}^{m_n}$ such that

$$\rho(\omega,\omega') \ge \frac{m_n}{8}, \text{ for all } \omega,\omega' \in \Omega_n, \omega \ne \omega', \text{ and } \operatorname{Card}(\Omega_n) \ge 2^{m_n/8}. \tag{43}$$

Then fix $m_n := \lfloor \varphi(\eta_n/2)^{-1} \rfloor$ where $\lfloor\cdot\rfloor$ is the integer part. For all $\omega \ne \omega'$, by definition of $\eta_n$

$$\mathbb{E}\left[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')\right] \ge \frac{1}{8}L^2 c_H^2 c_2 \eta_n^{2\beta} I_\psi m_n \varphi(\eta_n/2) \ge \frac{1}{8}L^2 c_H^2 c_2 r_n.$$

***Check (C')*** We also prove that the measure $\mathbb{P}_\omega$ is absolutely continuous with respect to $\mathbb{P}_0$, with derivative $d\mathbb{P}_\omega/d\mathbb{P}_0(x,y) = |D| f_\omega^x(y)$ and $d\mathbb{P}_\omega(x,y) = f_\omega^x(y)dyd\mathbb{P}_X(x)$, like in the proof of (i). Arguing again as in (i), we get

$$\begin{aligned} K(\mathbb{P}_\omega, \mathbb{P}_0) &\leq |D| L^2 \eta_n^{2\beta} \int_D \psi^2(y)dy \sum_{k=1}^{m_n} \omega_k \mathbb{E}\left[H^2\left(\frac{\|X - x_k\|}{\eta_n}\right)\right], \\ &\leq m_n |D| L^2 \eta_n^{2\beta} \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} \mathbb{P}\left(\|X - x_k\| \leq \eta_n\right). \end{aligned}$$

Now, arguing again as in Check (A'), we can apply Assumption (14) and get that $\mathbb{P}(\|X - x_k\| \leq \eta_n) \leq C_2 \mathbb{P}(\|X\| \leq \eta_n) = C_2 \varphi(\eta_n)$. Thanks to the definition of $\eta_n$, we now obtain (as in (i))

$$K(\mathbb{P}_\omega^{\otimes n}, \mathbb{P}_0^{\otimes n}) \leq C_2 m_n |D| L^2 \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} c_{(C)}.$$

Finally, condition (43) on the cardinal of $\Omega_n$ leads to $m_n \leq (8/\ln 2)\ln(\mathrm{Card}(\Omega_n))$, which completes the proof of (C'), taking $\zeta = (8/\ln 2) C_2 |D| L^2 \|\psi\|^2_{L^2(\mathbb{R})} \|H\|^2_{L^\infty(\mathbb{R})} c_{(C)}$ (we verify that the condition on $\|\psi\|_{\mathbb{L}^2(\mathbb{R})}$ implies that $\zeta < 1/8$). This completes at the same time the proof of all the lower bounds. □

### *A.3. Proofs of some technical lemmas*

Here we prove the technical results which are used in Section 7.

#### *A.3.1. Proof of Lemma 1*

To prove Inequality (19), the guideline is to apply Bernstein's Inequality (see Birgé and Massart [9]), for the conditional probability $\mathbb{P}_{X'}$.

**Lemma 9.** *Let $T_1, T_2, \ldots, T_n$ be independent random variables and $S_n(T) = \sum_{i=1}^n (T_i - \mathbb{E}[T_i])$. Assume that*

$$Var(T_1) \leq v^2 \text{ and } \forall l \geq 2, \ \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[|T_i|^l\right] \leq \frac{l!}{2} v^2 b_0^{l-2}.$$

*Then, for $\eta > 0$,*

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}|S_n(T)| \geq \eta\right) &\leq 2\exp\left(-\frac{n\eta^2/2}{v^2 + b_0\eta}\right), \\ &\leq 2\max\left\{\exp\left(-\frac{n\eta^2}{4v^2}\right), \exp\left(-\frac{n\eta}{4b_0}\right)\right\}. \end{aligned} \tag{44}$$

Here, $T_i = K_h(d(X_i, X'))/\mathbb{E}_{X'}[K_h(d(X_i, X'))]$, and $R_h^{X'} - 1 = S_n(T)/n$ (recall that we consider here conditional expectation and probability with respect

to $X'$). Let us compute the quantities $v$ and $b_0$ involved in the inequality. First, on the set $\{X' \in B\}$, Inequality (17) implies that

$$\begin{aligned}\mathrm{Var}_{X'}(T_1) &\leq \mathbb{E}_{X'}\left[T_1^2\right] = \frac{\mathbb{E}_{X'}\left[K_h^2(d(X_1,X'))\right]}{\left(\mathbb{E}_{X'}\left[K_h(d(X_1,X'))\right]\right)^2} \\ &\leq \frac{h^{-2}M_2\varphi(h)}{\left(h^{-1}m_1\varphi(h)\right)^2} = \frac{M_2}{m_1^2}\frac{1}{\varphi(h)} =: v^2.\end{aligned}$$

Similarly, for $l \geq 2$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{X'}\left[|T_i|^l\right] &= \mathbb{E}_{X'}\left[|T_1|^l\right] = \frac{\mathbb{E}_{X'}\left[K_h^l(d(X_1,X'))\right]}{\left(\mathbb{E}_{X'}\left[K_h(d(X_1,X'))\right]\right)^l} \\ &\leq \frac{h^l M_l\varphi(h)}{\left(h\varphi(h)m_1\right)^l} = \frac{M_l}{m_1^l}\frac{1}{\varphi^{l-1}(h)}.\end{aligned}$$

By splitting $M_l = C_K^l C_\varphi = M_2 C_K^{l-2}$, the last upper bound can be written

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{X'}\left[|T_i|^l\right] \leq \frac{M_2}{m_1^2}\frac{1}{\varphi(h)}\frac{C_K^{l-2}}{m_1^{l-2}}\frac{1}{(\varphi(h))^{l-2}} = v^2 b_0^{l-2},$$

with $b_0 = C_K/(m_1\varphi(h))$. We now apply the first inequality of Lemma 9, this complete the proof of Inequality (19). The proof may be adapted easily to demonstrate Inequality (21). For Inequality (20), we follow the same strategy as Comte and Genon-Catalot [18], pages 20–21. First

$$\begin{aligned}&\mathbb{E}_{X'}\left[\left(\left(R_h^{X'}-1\right)^2 - V_R(h)\right)_+\right] \\ &= \int_0^\infty \mathbb{P}_{X'}\left(\left(\left(R_h^{X'}-1\right)^2 - V_R(h)\right)_+ \geq u\right) du, \\ &\leq \int_0^\infty \mathbb{P}_{X'}\left(\left|R_h^{X'}-1\right| \geq \sqrt{V_R(h)+u}\right) du, \\ &\leq 2\max\left\{\int_0^\infty \exp\left(-\frac{n(u+V_R(h))}{4v^2}\right) du, \right. \\ &\qquad \left. \int_0^\infty \exp\left(-\frac{n\sqrt{u+V_R(h)}}{4b_0}\right) du\right\},\end{aligned}$$

thanks to Inequality (44). Now,

$$\frac{n(u+V_R(h))}{4v^2} = n\varphi(h)u\frac{m_1^2}{4M_2} + \ln(n)\frac{m_1^2\kappa_R}{4M_2},$$

which leads to

$$\int_0^\infty \exp\left(-\frac{n(u+V_R(h))}{4v^2}\right) du \leq \frac{4M_2}{m_1^2}\frac{1}{n^{1+\kappa_R m_1^2/4M_2}}\frac{1}{\varphi(h)}.$$

Since, by Assumption $H_{b2}$, $\varphi(h) \geq C_0 \ln(n)/n$, we obtain

$$\int_0^\infty \exp\left(-\frac{n(u+V_R(h))}{4v^2}\right) du \leq \frac{4M_2}{C_0 m_1^2} \frac{1}{\ln(n) n^{\kappa_R m_1^2/4M_2}},$$

and the last upper bound is smaller than $(4M_2/C_0 m_1^2)/n^\alpha$ as soon as $\kappa_R > 4M_2\alpha/m_1^2$. For the other integral, we begin with a lower bound for $n\sqrt{u+V_R(h)}/4b_0$,

$$\frac{n\sqrt{u+V_R(h)}}{4b_0} \quad \geq \quad \frac{m_1\sqrt{C_0}}{4\sqrt{2}C_K}\sqrt{\kappa_R}\ln(n) + \frac{m_1 C_0}{4\sqrt{2}C_K}\ln(n)\sqrt{u},$$

by using $\varphi(h) \geq C_0 \ln(n)/n$ another time. Thus,

$$\begin{aligned}\int_0^\infty \exp\left(-\frac{n\sqrt{u+V_R(h)}}{4b_0}\right) du &\leq n^{-\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}} \int_0^\infty \exp\left(-\frac{m_1 C_0 \ln(n)}{4\sqrt{2}C_K}\sqrt{u}\right) du \\ &= \frac{64C_K^2}{m_1^2 C_0^2}\int_0^\infty s\exp(-s)ds \frac{1}{\ln^2(n) n^{\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}}} \\ &= \frac{64C_K^2}{m_1^2 C_0^2}\frac{1}{\ln^2(n) n^{\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}}} \leq \frac{64C_K^2}{m_1^2 C_0^2}\frac{1}{n^\alpha}.\end{aligned}$$

as soon as $\kappa_R > 32C_K^2\alpha^2/m_1^2 C_0$. This ends the proof of Lemma 1. □

*A.3.2. Proof of Lemma 6 (concentration of the empirical process)*

The aim is to control the deviations of the supremum of the empirical process $\nu_{n,h}$ defined by (36). Since it is centred and bounded, the guiding idea is to apply the following concentration inequality.

**Lemma 10** (Talagrand's Inequality)**.** *Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables, and define $\nu_n(r) = \frac{1}{n}\sum_{i=1}^n r(\xi_i) - \mathbb{E}[r(\xi_i)]$, for $r$ belonging to a countable class $\mathcal{R}$ of real-valued measurable functions. Then, for $\delta > 0$, there exists a universal constant $C$ such that*

$$\begin{aligned}&\mathbb{E}\left[\left(\sup_{r\in\mathcal{R}}\left(\nu_n(r)\right)^2 - c(\delta)(H^\nu)^2\right)_+\right] \\ &\quad\leq C\left\{\frac{v^\nu}{n}\exp\left(-\frac{\delta}{6}\frac{n(H^\nu)^2}{v^\nu}\right) + \frac{(M_1^\nu)^2}{C^2(\delta)n^2}\exp\left(-\frac{1}{21\sqrt{2}}C(\delta)\sqrt{\delta}\frac{nH^\nu}{M_1^\nu}\right)\right\},\end{aligned}$$

*with, $C(\delta) = (\sqrt{1+\delta}-1)\wedge 1$, $c(\delta) = 2(1+2\delta)$ and*

$$\sup_{r\in\mathcal{R}}\|r\|_{L^\infty} \leq M_1^\nu,\ \mathbb{E}\left[\sup_{r\in\mathcal{R}}|\nu_n(r)|\right] \leq H^\nu,\ \textit{and}\ \sup_{r\in\mathcal{R}} Var\left(r\left(\xi_1\right)\right) \leq v^\nu.$$

Lemma 10 is a classical consequence of the Talagrand Inequality given in Klein and Rio [44]: see for example Lemma 5 (page 812) in Lacour [46].

We first compute $H^\nu$, $M^\nu$ and $v^\nu$, involved in Lemma 10.

- For $M^\nu$, let $t \in \bar{S}_D(0,1)$, $x \in \mathbb{H}$ and $y \in \mathbb{R}$ be fixed. By the Cauchy-Schwarz Inequality,

$$|\psi_{t,h}(x,y)| \leq |D|\|t\|_D \frac{\|K_h\|_{L^\infty(\mathbb{R})}}{m_1 h^{-1}\varphi(h)} \leq \frac{|D|C_K}{m_1\varphi(h)} := M_1^\nu,$$

  thanks to (17).
- For $H^\nu$, recall that

$$\mathbb{E}_{X'}\left[\sup_{t\in\bar{S}_D(0,1)} \nu_{n,h}^2(t)\right] = \mathbb{E}_{X'}\left[\left\|\widetilde{F}_{h'}^{X'} - \mathbb{E}_{X'}[\widetilde{F}_{h'}^{X'}]\right\|_D^2\right] \leq |D|\frac{M_2}{m_1^2}\frac{1}{n\varphi(h)} := (H^\nu)^2$$

  a.s. on the set $\{X' \in B\}$ with the same computation as for the variance term, see Lemma 2.
- For $v^\nu$, we also fix $t \in \bar{S}_D(0,1)$, and compute,

$$\mathrm{Var}_{X'}\left(\psi_{t,h}(X_1,Y_1)\right) \leq \mathbb{E}_{X'}\left[\psi_{t,h}^2(X_1,Y_1)\right], = \mathbb{E}_{X'}\left[\left(\int_D \mathbf{1}_{Y_1\leq y} t(y)dy\right)^2 \frac{K_h^2(d(X_1,X'))}{\left(\mathbb{E}_{X'}\left[K_h(d(X_1,X'))\right]\right)^2}\right]\mathbf{1}_B(X').$$

  The integral is controlled with the Cauchy-Schwarz Inequality: $(\int_D \mathbf{1}_{Y_1\leq y}t(y)dy)^2 \leq |D|\|t\|_D^2 = |D|$, and the other quantity has already been bounded: we obtain

$$\mathrm{Var}_{X'}\left(\psi_{t,h}(X_1,Y_1)\right) \leq v^\nu := \frac{|D|M_2}{m_1^2\varphi(h)}.$$

Then, Lemma 10 gives, for $\delta > 0$,

$$\mathbb{E}\left[\left(\sup_{t\in\bar{S}_D(0,1)} \nu_{n,h}^2(t) - 2(1+2\delta)\left(H^\nu\right)^2\right)_+ \mathbf{1}_B(X')\right] \leq C\left\{\frac{|D|M_2}{m_1^2}\frac{1}{n\varphi(h)}\exp\left(-\frac{\delta}{6}\right) + \frac{1}{C^2(\delta)}\frac{C_K^2|D|^2}{m_1^2}\frac{1}{n^2\varphi^2(h)}\exp\left(-\frac{1}{21\sqrt{2}}C(\delta)\sqrt{\delta}\frac{\sqrt{M_2}\sqrt{n\varphi(h)}}{\sqrt{|D|}C_K}\right)\right\}.$$

We choose $\delta = \delta_0 \ln(n)$, for a $\delta_0$ large enough, and given below. We compute the order of magnitude of the last upper bound, using Assumptions $H_{b1}$ and $H_{b2}$.

Recall that they imply $\sum_{h\in\mathcal{H}_n} 1/\varphi(h) \le n^2/C_0\ln(n)$ and $\sum_{h\in\mathcal{H}_n} 1/\varphi^2(h) \le n^3/C_0^2\ln^2(n)$. First,

$$\sum_{h\in\mathcal{H}_n}\frac{1}{n\varphi(h)}\exp\left(-\frac{\delta}{6}\right)=\frac{1}{n^{1+\delta_0/6}}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi(h)}\le\frac{1}{C_0n^{\delta_0/6-1}\ln(n)}\le\frac{1}{C_0n},$$

as soon as $\delta_0\ge 12$ since we can reasonably assume $n\ge 3$. Then, $C(\delta_0\ln(n))=\sqrt{1+\delta_0\ln(n)}-1\ge 1$, if $\delta_0\ln(n)\ge 3$, that is $\ln(n)\ge 3/\delta_0$. This is satisfied since $\delta_0>12$ and $n\ge 2$. Hence

$$\begin{aligned}
&\frac{1}{n^2C^2(\delta)}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi^2(h)}\exp\left(-\frac{1}{21\sqrt{2}}C(\delta_0\ln(n))\sqrt{\delta_0\ln(n)}\frac{\sqrt{M_2}}{\sqrt{|D|}C_K}\sqrt{n\varphi(h)}\right)\\
&\quad\le\ \frac{1}{n^2}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi^2(h)}\exp\left(-\frac{1}{21\sqrt{2}}\sqrt{\delta_0\ln(n)}\frac{\sqrt{M_2}}{\sqrt{|D|}C_K}\sqrt{n\varphi(h)}\right),\\
&\quad\le\ \frac{1}{n^2}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi^2(h)}\exp\left(-\frac{\sqrt{C_0}}{21\sqrt{2}}\sqrt{\delta_0}\frac{\sqrt{M_2}}{\sqrt{|D|}C_K}\ln(n)\right),\\
&\quad=\ n^{-2-\frac{\sqrt{C_0M_2\delta_0}}{21\sqrt{2|D|}C_K}}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi^2(h)}\le\frac{1}{C_0^2\ln^2(n)}n^{-\frac{\sqrt{C_0M_2\delta_0}}{21\sqrt{2|D|}C_K}+1}\le\frac{1}{C_0^2n},
\end{aligned}$$

as soon as $\sqrt{C_0M_2\delta_0}/(21\sqrt{2|D|}C_K)-1>1$ that is $\delta_0>3528C_K^2|D|/C_0M_2$ with $C>0$ depending only on $m_1$, $M_2$, $\delta_0$, $C_0$ and $|D|$. This shows that

$$\sum_{h\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}_D(0,1)}\nu_{n,h}^2(t)-2(1+2\delta_0\ln(n))\left(H^\nu\right)^2\right)_+\mathbf{1}_B(X')\right]\le\frac{C}{n},\qquad(45)$$

for $(H^\nu)^2=|D|(M_2/m_1^2)/(n\varphi(h))$ and $C>0$ depending only on $m_1$, $M_2$, $\delta_0$, $C_0$ and $|D|$. Since

$$2(1+2\delta_0\ln(n))\left(H^\nu\right)^2\le 6\delta_0\frac{|D|M_2}{m_1^2}\frac{\ln(n)}{n\varphi(h)},$$

Inequality (45) is also satisfied when we replace $2(1+2\delta_0\ln(n))(H^\nu)^2$ by this upper bound. Thus, the proof of Lemma 6 is completed. □

*A.3.3. Proof of Lemma 8*

First suppose that $\beta/2\in\mathbb{N}^*$, we have

$$\left(\sum_{j>p}\sigma_j^2\eta_j^2\right)^{\beta/2}=\sum_{j_1,\ldots,j_{\beta/2}>p}\prod_{l=1}^{\beta/2}\sigma_{j_l}^2\eta_{j_l}^2,$$

and, by a classical generalization of Hölder's Inequality

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2\eta_j^2\right)^{\beta/2}\right] &= \sum_{j_1,\dots,j_{\beta/2}>p}\prod_{l=1}^{\beta/2}\sigma_{j_l}^2\mathbb{E}\left[\prod_{l=1}^{\beta/2}\eta_{j_l}^2\right]\\ &\leq \sum_{j_1,\dots,j_{\beta/2}>p}\prod_{l=1}^{\beta/2}\sigma_{j_l}^2\prod_{l=1}^{\beta/2}\mathbb{E}\left[\eta_{j_l}^{\beta}\right]^{2/\beta}\\ &\leq C_M\sum_{j_1,\dots,j_{\beta/2}>p}\prod_{l=1}^{\beta/2}\sigma_{j_l}^2 \leq C_M\left(\sum_{j>p}\sigma_j^2\right)^{\beta/2}.\end{aligned}$$

Now suppose that $\beta\in\mathbb{Q}\cap]0,+\infty[$, we can write without loss of generality that $\beta/2=p/q$ with $p\in\mathbb{N}^*$ and $q>1$ (if $q=1$, $\beta/2\in\mathbb{N}^*$). Then the function $x\mapsto x^{1/q}$ is concave and by Jensen's Inequality:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2\eta_j^2\right)^{\beta/2}\right] &\leq \mathbb{E}\left[\left(\sum_{j>p}\sigma_j^2\eta_j^2\right)^{p}\right]^{1/q}\\ &\leq C_M^{1/q}\left(\sum_{j>p}\sigma_j^2\right)^{\beta/2} \leq C_M\left(\sum_{j>p}\sigma_j^2\right)^{\beta/2}.\end{aligned}$$

The case $\beta>0$ follows immediately from the density of $\mathbb{Q}$ into $\mathbb{R}$. □

## References


[1] Ait-Saïdi, A., Ferraty, F., Kassa, R. and Vieu, P. (2008). Cross-validated estimations in the single-functional index model. *Statistics* **42** 475–494. MR2465129 (2010a:62246)

[2] Akakpo, N. and Lacour, C. (2011). Inhomogeneous and anisotropic conditional density estimation from dependent data. *Electron. J. Stat.* **5** 1618–1653. MR2870146

[3] Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. MR0069229 (16,1005a)

[4] Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics, Vol. 27. MR0448463 (56 ##6769)

[5] Aspirot, L., Bertin, K. and Perera, G. (2009). Asymptotic normality of the Nadaraya-Watson estimator for nonstationary functional data and applications to telecommunications. *J. Nonparametr. Stat.* **21** 535–551. MR2543572 (2010i:62085)

[6] Azzedine, N., Laksaci, A. and Ould-Saïd, E. (2008). On robust nonparametric regression estimation for a functional regressor. *Statist. Probab. Lett.* **78** 3216–3221. MR2479480 (2010j:62117)
[7] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.* **22** 353–369. MR2336341
[8] Bertin, K., Lacour, C. and Rivoirard, V. (2014). Adaptive estimation of conditional density function. Submitted, hal-00922555.
[9] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4** 329–375. MR1653272 (2000g:62071)
[10] Brunel, E., Comte, F. and Lacour, C. (2010). Minimax estimation of the conditional cumulative distribution function. *Sankhya A* **72** 293–330. MR2746114 (2012c:62317)
[11] Brunel, E., Mas, A. and Roche, A. (2013). Non-asymptotic Adaptive Prediction in Functional Linear Models. Submitted, hal-00763924.
[12] Burba, F., Ferraty, F. and Vieu, P. (2009). $k$-nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.* **21** 453–469. MR2571722 (2011b:62105)
[13] Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 (2008k:62132)
[14] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45** 11–22. MR1718346 (2000g:62152)
[15] Chagny, G. (2013a). Penalization versus Goldenshluger-Lepski strategies in warped bases regression. *ESAIM Probab. Statist.* **17** 328–358 (electronic). MR3066383
[16] Chagny, G. (2013b). Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie PhD thesis, Univ. Paris Descartes.
[17] Chen, D., Hall, P. and Müller, H. G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. MR2850218 (2012g:62113)
[18] Comte, F. and Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *J. Statist. Plann. Inference* **142** 1698–1715. MR2903382
[19] Comte, F. and Johannes, J. (2012). Adaptive functional linear regression. *Ann. Statist.* **40** 2765–2797. MR3097959
[20] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34** 431–452. MR2328553
[21] Crambes, C., Delsol, L. and Laksaci, A. (2008). Robust nonparametric estimation for functional data. *J. Nonparametr. Stat.* **20** 573–598. MR2454613 (2009m:62111)
[22] Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* **37** 35–72. MR2488344 (2010i:62089)

[23] Dabo-Niang, S., Kaid, Z. and Laksaci, A. (2012). On spatial conditional mode estimation for a functional regressor. *Statist. Probab. Lett.* **82** 1413–1421. MR2929795

[24] Dabo-Niang, S. and Rhomari, N. (2009). Kernel regression estimation in a Banach space. *J. Statist. Plann. Inference* **139** 1421–1434. MR2485136 (2010d:62093)

[25] Dabo-Niang, S. and Yao, A.-F. (2013). Kernel spatial density estimation in infinite dimension space. *Metrika* **76** 19–52. MR3018356

[26] Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.* **38** 1171–1193. MR2604709 (2011c:62097)

[27] Demongeot, J., Laksaci, A., Madani, F. and Rachdi, M. (2010). Estimation locale linéaire de la densité conditionnelle pour des données fonctionnelles. *C. R. Math. Acad. Sci. Paris* **348** 931–934. MR2677995 (2011g:62099)

[28] Dunker, T., Lifshits, M. A. and Linde, W. (1998). Small deviation probabilities of sums of independent random variables. In *High Dimensional Probability (Oberwolfach, 1996). Progr. Probab.* **43** 59–74. Birkhäuser, Basel. MR1652320 (2000h:60035)

[29] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes.* **9** 47–76. MR2224840 (2007h:62034)

[30] Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* **49** 267–286. MR2396496 (2009c:62075)

[31] Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis. Oxford Handbooks in Mathematics.* OUP Oxford. MR2917982

[32] Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.* **17** 545–564. MR1952697 (2003m:62102)

[33] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Springer Series in Statistics.* Springer, New York. Theory and practice. MR2229687 (2007b:62008)

[34] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Stat. Plan. Inference* **140** 335–352. MR2558367 (2011e:62078)

[35] Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.* **5** 30–43. MR3018508

[36] Gheriballah, A., Laksaci, A. and Sekkal, S. (2013). Nonparametric $M$-regression for functional ergodic data. *Statist. Probab. Lett.* **83** 902–908. MR3040321

[37] Gijbels, I., Omelka, M. and Veraverbeke, N. (2012). Multivariate and functional covariates and conditional copulas. *Electron. J. Stat.* **6** 1273–1306. MR2988448

[38] Glaser, N. S., Barnett, P., McCaslin, I., Nelson, D., Trainor, J., Louie, J., Kaufman, F., Quayle, K., Roback, M., Malley, R. et al. (2001). Risk factors for cerebral edema in children with diabetic ketoacidosis. *New Engl. J. Med.* **344** 264–269.

[39] Glaser, N. S., Wootton-Gorges, S. L., Buonocore, M. H., Tancredi, D. J., Marcin, J. P., Caltagirone, R., Lee, Y., Murphy, C. and Kuppermann, N. (2013). Subclinical cerebral edema in children with diabetic ketoacidosis randomized to 2 different rehydration protocols. *Pediatrics* **131** e73–e80.

[40] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 (2012k:62091)

[41] Goldenshluger, A. and Lepski, O. (2013). On adaptive minimax density estimation on Rd. *Probab. Theory Related Fields* to appear. MR3230001

[42] Hoffmann-Jørgensen, J., Shepp, L. A. and Dudley, R. M. (1979). On the lower tail of Gaussian seminorms. *Ann. Probab.* **7** 319–342. MR0525057 (80j:60051)

[43] Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916 (2002j:62044)

[44] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312 (2006c:60022)

[45] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342** 877–882. MR2224640

[46] Lacour, C. (2008). Adaptive estimation of the transition density of a particular hidden Markov chain. *J. Multivariate Anal.* **99** 787–814. MR2405092 (2009i:62166)

[47] Laib, N. and Louani, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivariate Anal.* **101** 2266–2281. MR2719861 (2011h:62133)

[48] Li, W. V. and Shao, Q. M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods. Handbook of Statist.* **19** 533–597. North-Holland, Amsterdam. MR1861734

[49] Lifshits, M. A. (1997). On the lower tail probabilities of some random series. *Ann. Probab.* **25** 424–442. MR1428515 (98b:60100)

[50] Mas, A. (2012). Lower bound in regression for functional data by representation of small ball probabilities. *Electron. J. Stat.* **6** 1745–1778. MR2988463

[51] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.* **115** 155–177. MR2105373 (2005h:62117)

[52] Plancade, S. (2013). Adaptive estimation of the conditional cumulative distribution function from current status data. *J. Statist. Plann. Inference* **143** 1466–1485. MR3070252

[53] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference* **137** 2784–2801. MR2323791 (2009a:62190)

[54] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, second ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[55] Shang, H. L. (2013). Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Comput. Statist. Data Anal.* **67** 185–198. MR3079596

[56] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by V. Zaiats. MR2724359 (2011g:62006)