ADAPTIVE AND MINIMAX ESTIMATION OF THE CUMULATIVE DISTRIBUTION FUNCTION GIVEN A FUNCTIONAL COVARIATE

GAËLLE CHAGNY$^{(A)}$ AND ANGELINA ROCHE$^{(B)}$

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 device in the spirit of the recent Godenshluger-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 criterion 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.

(A) LMRS, UMR CNRS 6085, Université de Rouen, France. gaelle.chagny@gmail.com

(B) I3M, UMR CNRS 5149, Université Montpellier II, France. angelina.roche@univ-montp2.fr

Keywords: Adaptive kernel estimator. Conditional cumulative distribution function. Functional random variable. Minimax estimation. Small ball probability.

AMS Subject Classification 2010: 62G05; 62H12.

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 (2005); Ferraty and Vieu (2006) for case studies and Ferraty and Romain (2011) for a recent overview. Here, we are interested in explaining the relationship between a functional random variable X and a scalar quantity Y. 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 (conditional c.d.f. in the sequel) of Y given X defined by

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

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

In the sequel, we consider kernel estimators similar to the ones defined by Ferraty et al. (2006, 2010), for which we provide a detailed non-asymptotic adaptive and minimax study.

Date: January 15, 2014.

The pioneering works on conditional distribution when the covariate is functional are the one of Ferraty and Vieu (2002); Ferraty et al. (2006), completed by Ferraty et al. (2010). Kernel estimators, which depend on a smoothness 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 (2007); Ferraty et al. (2007); Dabo-Niang and Rhomari (2009). The case of dependent data is the subject of the works of Masry (2005); Aspirot et al. (2009); Laib and Louani (2010); Dabo-Niang et al. (2012) under several assumptions (α-mixing, ergodic or non-stationary processes). Demongeot et al. (2010) consider local-linear estimators of the conditional density and conditional mode. Robust versions of the previous strategies are proposed by Crambes et al. (2008); Azzedine et al. (2008); Gheriballah et al. (2013). Gijbels et al. (2012) 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. However, only two papers tackle the problem of bandwidth selection: Ferraty and Vieu (2002) and Rachdi and Vieu (2007) suggest cross-validation procedures which are shown to be asymptotically optimal in regression contexts.

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 et al. (2010) and Plancade (2013) for c.d.f estimation with a real covariate and to Akakpo and Lacour (2011) 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 non-parametric setting.

In keeping with the studies of functional conditional distribution, we investigate the properties of the nonparametric Nadaraya-Watson-type estimators of Ferraty et al. (2006), 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

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

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 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 2011) 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 (2012) 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 2005; Ferraty et al. 2006; Geenens 2011) 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. Finally, the proofs are gathered in Section 6.

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:

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

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

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

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, 2005; Cardot et al., 1999) where either prediction error on random curves (Crambes et al., 2009) or prediction error over a fixed curve (Cai and Hall, 2006) are considered. Second, integrated risks have been relatively unexplored in non-parametric functional data analysis. Indeed, there is no measure universally accepted as the Lebesgue measure in finite-dimensional setting (see e.g. Delaigle and Hall 2010; Dabo-Niang and Yao 2013). The only measure at hand is the probability measure of X.

2.2. **Assumptions.** Hereafter, we denote by φ^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 that is to say the function $t \mapsto \mathbb{E}[X(t)]$ is supposed to be identically equal to 0. 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_K The kernel K is of type I (Ferraty and Vieu, 2006) i.e. its support is in $[0, 1]$ and there exist two constants $c_K, C_K > 0$ such that

$$c_K \mathbf{1}_{[0,1]} \leq K \leq C_K \mathbf{1}_{[0,1]}.$$

H_F There exists $\beta > 0$ 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_φ 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_K is quite classical in kernel methods for functional data (see Ferraty et al. 2006; Burba et al. 2009; Ferraty et al. 2010). We are aware that this is a strong assumption but alleviate it in a functional data context requires a lot of technical difficulties and it is still, to our knowledge, an open problem.

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 et al. (2006) or Assumption (H2') of Ferraty et al. (2010). Note, however, that, since both considered risks are integrated with respect to y, 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 et al. (2010, Corollary 1) only depends on the regularity of F with respect to x.

Assumption H_φ is very similar to assumptions made by Ferraty et al. 2006; Burba et al. 2009; Ferraty et al. 2010. This condition H_φ is reasonable, since the class of Gaussian processes fulfill 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 (Anderson, 1955) (see also Li and Shao 2001, Theorem 2.13 or Hoffmann-Jørgensen et al. 1979, Theorem 2.1, p.322) and from Hoffmann-Jørgensen et al. (1979, 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 non-asymptotic upper bound for the risk:

Theorem 1. *Suppose assumptions H_K and H_F are 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_φ 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_φ, C_φ, $|D|$ 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 et al. (2006, Theorem 3.1) and Ferraty et al. (2010, 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 β 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 κ 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 estimate 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 6). This motivates the following choice of the bandwidth:

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

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

This selection rule is inspired both on the recent version of the so-called Lepski method (see Goldenshluger and Lepski 2011) and model selection tools. The main idea is to estimate the bias term by looking at several estimators. Goldenshluger and Lepski (2011) propose to first define "intermediate" estimates $\widehat{F}^{X'}_{h,h'}$ $(h,h'\in\mathcal{H}_n)$, based on a convolution product of the kernel with the estimators 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 similar to the procedure proposed by Chagny (2013a) or Comte and Johannes (2012) for model selection purpose. Thus, $V(h)$ can also be seen as a penalty term. We also refer to the phD of Chagny (2013b, p.170) for technical details leading to this choice.

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 $C_0>12$ is a purely numerical constant (specified in the proofs).

Assumption H_{b1} fixes the size of the bandwidth collection: compared to the assumptions of Goldenshluger and Lepski (2011), we consider a discrete set and not an interval, which permits to use the classical tools of model selection theory in the proofs. We now state the following result.

Theorem 2. *Assume H_K, H_φ, H_F, H_b and that $n\geq 3$. There exist two constants $c,C>0$ depending on c_K, C_K, c_φ, C_φ, $|D|$, C_D such that*

$$\mathbb{E}\left[\left\|\widehat{F}^{X'}_{\widehat{h}}-F^{X'}\right\|_D^2\mathbf{1}_B(X')\right]\leq c\min_{h\in\mathcal{H}_n}\left\{h^{2\beta}+\frac{\ln(n)}{n\varphi(h)}\right\}+\frac{C}{n}. \tag{11}$$

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. 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, the penalty term $V(h)=\kappa\ln(n)/(n\varphi(h))$, which may have been classically defined cannot be used in practice. This explains why we define and plug in $\widehat{V}(h)$ the estimator (9). However, for the sake of clarity, we begin the proof by establishing the result with $\widehat{V}(h)$ replaced by its theoretical counterpart $V(h)=\kappa\ln(n)/(n\varphi(h))$.

Remark 1.
- In practice, it is impossible to verify Assumption H_{b2} since the function φ and the constant C_0 are unknown. However, this difficulty can be circumvented by introducing a random collection of bandwidths $\widehat{\mathcal{H}}_n$ verifying, for all $h \in \widehat{\mathcal{H}}_n$, $\widehat{\varphi}(h) \geq 2\widehat{C}_0 \ln(n)/n$ where $\widehat{\varphi}$ and $\widehat{C}_0$ are some estimators of φ (see Equation (9)) and C_0. However, since it does not add significant difficulty (see Comte and Johannes 2012; Brunel et al. 2013) but would complicate the understandability of proofs, we choose to keep Assumption H_{b_2}.
- 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 (11) 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 to zero of our estimators. Small ball problems have aroused considerable interest and attention in the past decades, and 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 2001), systematic studies have also been undertaken to handle the general case of (infinite) sum of independent random variables (Lifshits, 1997; Dunker et al., 1998; Mas, 2012). We consider in the sequel one of the three following hypothesis which allow to understand how the small ball probability decay influences the rates (see Section 4.2) and which are frequently used in the literature. We describe below large class 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 heavily 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 $\Gamma f(s) = \langle f, \mathrm{Cov}(X_., X_s)\rangle$. Recall the Karhunen-Loève decomposition of the process X, which can be written

$$X = \sum_{j\geq 1} \sqrt{\lambda_j}\eta_j\psi_j, \tag{12}$$

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 Γ) 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 2012, 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 $cj^{-2a} \leq \lambda_j \leq Cj^{-2a}$, $a \geq 1/2$ $(c, C > 0)$, 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 et al. 1979, 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 1975). More generally, if X is defined by a random series $X = \sum_{j\geq 1} j^{-2a} Z_j$, for variables Z_i with a c.d.f. regularly varying at 0 with positive index, one can also define γ_1, γ_2, and α such that $H_{X,L}$ is fulfilled (see Dunker et al. 1998, Proposition 4.1 p.11 and also Mas 2012, (19) p.9). The second case $H_{X,M}$ typically happens when the eigenvalues of the covariance operator exponentially decrease (see Dunker et al. 1998, 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 $H_{X,L}$ (Hoffmann-Jørgensen et al., 1979, Theorem 4.4 and example 4.7, pp. 333 and 336).

Finally, it also results of 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 Γ 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.** We now compute the upper bounds for the pointwise and integrated risks of the estimators, under the previous regularity assumptions.

Theorem 3. (a) *Under the assumptions of Theorem 1, 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 the results. The faster the small ball probability decreases (that is the more 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 et al. (2006) (section 5.3) in the same framework, and by Mas (2012) 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.

		$H_{X,L}$ (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)	$(\ln(n))^{-2\beta/\alpha}$	$\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha}(n)\right)$	$n^{-\frac{2\beta}{2\beta+\gamma}}$

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

Remark 2. We thus obtain various rates, depending on the regularity assumptions on X. This phenomenom also occurs in a deconvolution model: the rates for the kernel estimators are logarithmic if the noise is "supersmooth" and the signal to recover "ordinarysmooth", but can be improved by considering the case of a "supersmooth" signal, to recover at least rates which are intermediate between logarithmic and polynomial (see e.g. Lacour 2006; Comte and Lacour 2010).

We have already noticed that our adaptive procedure leads to the loss of a logarithm 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 4. *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 $x \in B$,*

$$\varphi(h) > 0 \text{ and } c_2\varphi(h) \leq \varphi^x(h) \leq C_2\varphi(h). \tag{13}$$

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 4 proves that the upper bounds of Theorem 3 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 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 et al. (2010) or Proposition 4.1 of Plancade (2013) 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 (2012) (Theorem 3) for regression estimation.

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 $\varphi_p(h) := \mathbb{P}\left(d_p(x,0) \le h\right) \sim_{h\to 0} C(p)h^p$ and then the variance is of order $1/(nh^p)$. Notice that, even if the variances order 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 are 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 4 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 the following proposition, 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'_φ There exist two constants $c_\varphi, C_\varphi > 0$ such that for all $h \in \mathbb{R}$, for all $p \in \mathbb{N}^*$,

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

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

H_ξ 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_ξ^{ind} the sequence of random variables $(\xi_j)_{j\ge 1}$ is independent and there exists a constant C_ξ such that, for all $j \ge 1$

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

H_ξ^b there exists a constant C_ξ such that, for all $j \ge 1$,

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

Remark that Assumption H'_φ is the equivalent of Assumption H_φ replacing d by d_p. If X is a Gaussian process, the vector $(\langle X, e_1\rangle, \dots, \langle X, e_p\rangle)$ is a Gaussian vector and

Assumption H_φ is also verified provided that B is bounded. Assumption H_ξ^{ind} is true if X is a Gaussian process and $(e_j)_{j\geq 1}$ is the Karhunen-Loève basis of X (see (12) above), and also Ash and Gardner 1975) and Assumption H_ξ^b is equivalent to suppose that X is bounded a.s. We are aware that both assumptions H_ξ^{ind} and H_ξ^b are strong since in most cases the Karhunen-Loève basis is unknown. We give here Proposition 1 below in the only aim of better understanding 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 1. *Suppose assumptions H_K, H_F and H_ξ are fulfilled. Let $h>0$ and $p\in\mathbb{N}^*$ be fixed.*

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

$$\mathbb{E}\left[\left\|\widehat{F}_{h,p}^{x_0}-F^{x_0}\right\|_D^2\right]\leq C\left(h^{2\beta}+\left(\sum_{j>p}\sigma_j^2\right)^\beta+\left(\sum_{j>p}\langle x_0,e_j\rangle^2\right)^\beta+\frac{1}{n\varphi_p^{x_0}(h)}\right), \tag{14}$$

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

(ii) *If, in addition, Assumption H_φ is fulfilled,*

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

where $C>0$ only depends on C_ξ, β, c_K, C_K, c_φ, C_φ, $|D|$ and C_D.

We have additional bias terms compared to Ferraty et al. (2006, 2010). This is due to the fact that our regularity assumption H_F (see Section 2.2) is here different from Assumption (H2) of Ferraty et al. (2006) or Assumption (H2') of Ferraty et al. (2010). 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 et al. (2006, 2010) 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 $j=1,\ldots,p$ (but $\langle x,e_j\rangle\neq\langle x',e_j\rangle$ for a $j>p$), 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 (2012) (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 $(\mathrm{span}\{e_1,\ldots,e_p\})^\perp$, 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 $K_h\left(\sqrt{\sum_{j=1}^p\langle x,e_j\rangle^2}\right)\neq 0$ 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 $\lim_{n\to+\infty}p=+\infty$.

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 rates obtained under Assumption $H_{X,F}$.* We start from the Karhunen-Loève decomposition of X defined in (12). For a Gaussian process, the variables η_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 fulfilled. 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 (15), 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 (14), 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 $x_0 \in \text{span}\{e_1, \ldots, e_p\}$.
- 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 rates obtained under Assumption $H_{X,M}$.* Thanks to Proposition 1, we are able to obtain the rate of convergence for the estimator.

Corollary 1. *Suppose that the assumptions of Proposition 1 are fulfilled 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 fulfilled 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.

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 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 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 (2002).Another approach may be to reduce the structural complexity of the model considering e.g. single or multiple-index models (Chen et al. 2011; Ait-Saïdi et al. 2008) .

6. Proofs

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.

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 of integrable function $L^q(\mathbb{R})$ (the notation will be used with $q=2$ and $q=\infty$).

Recall that $K_h(x) := h^{-1}K(h^{-1}x)$. Assumptions H_K and H_φ imply that, for all $l \geq 1$,

$$h^{-l} m_l \varphi(h) \mathbf{1}_B(X') \leq \mathbb{E}_{X'}\left[K_h^l(d(X,X'))\right] \mathbf{1}_B(X') \leq h^{-l} M_l \varphi(h) \mathbf{1}_B(X') \text{ a.s.} \tag{16}$$

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

6.1. A preliminary result.

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

$$R_h^x = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^n \frac{K_h(d(X_i,X'))}{\mathbb{E}_{X'}\left[K_h(d(X,X'))\right]}, & \text{if } x = X', \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^n \frac{K_h(d(X_i,x))}{\mathbb{E}\left[K_h(d(X,x))\right]}, & \text{if } x \in \mathbb{H} \text{ is fixed.} \end{cases} \tag{17}$$

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

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

$$\mathbb{P}_{X'}\left(\left|R_h^{X'} - 1\right| > \eta\right) \leq 2\exp\left(-\frac{n\eta^2\varphi(h)}{2\left(\frac{M_2}{m_1^2} + \frac{C_K\eta}{m_1}\right)}\right). \tag{18}$$

Moreover, assume also H_{b_2}, and denote by $V_R(h) = \kappa_R \ln(n)/(n\varphi(h))$, we have a.s.

$$\mathbb{E}_{X'}\left[\left(\left(R_h^{X'} - 1\right)^2 - V_R(h)\right)_+\right] \leq \min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n^\alpha}, \tag{19}$$

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 (18) becomes*

$$\mathbb{P}\left(|R_h^{x_0} - 1| > \eta\right) \leq 2\exp\left(-\frac{n\eta^2\varphi^{x_0}(h)}{2\left(\frac{M_2}{m_1^2} + \frac{C_K\eta}{m_1}\right)}\right). \tag{20}$$

6.1.1. *Proof of Lemma 1.* To prove Inequality (18), the guideline is to apply Bernstein's Inequality (see Birgé and Massart 1998), for the conditional probability $\mathbb{P}_{X'}$.

Lemma 2. *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*

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

Then, for $\eta > 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}|S_n(T)| \geq \eta\right) &\leq 2\exp\left(-\frac{n\eta^2/2}{v^2 + b_0\eta}\right), \\ &\leq 2\min\left\{\exp\left(-\frac{n\eta^2}{4v^2}\right), \exp\left(-\frac{n\eta}{4b_0}\right)\right\}. \end{aligned} \tag{21}$$

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 (16) implies that

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

Similarly, for $l \geq 2$,

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

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 2, this complete the proof of Inequality (18). The proof may be adapted easily to demonstrate Inequality (20). For Inequality (19), we follow the same strategy as Comte and Genon-Catalot (2012), pages 20-21. First

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

thanks to Inequality (21). 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

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

Since, by Assumption H_{b_2}, $\varphi(h)\geq C_0\ln(n)/n$, 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 $(4M_2/C_0m_1^2)/n^\alpha$ as soon as $\kappa_R > 4M_2\alpha/m_1^2$. For the other integral, we begin with a lower bound for $n\sqrt{u+V_R(h)}/4b_0$,

$$\begin{aligned}\frac{n\sqrt{u+V_R(h)}}{4b_0} &\geq \frac{m_1}{4C_K}n\varphi(h)\frac{1}{\sqrt{2}}\left(\sqrt{V_R(h)}+\sqrt{u}\right),\\ &= \frac{m_1}{4\sqrt{2}C_K}\sqrt{\kappa_R}\sqrt{\ln(n)}\sqrt{n\varphi(h)} + \frac{m_1}{4\sqrt{2}C_K}n\varphi(h)\sqrt{u},\\ &\geq \frac{m_1\sqrt{C_0}}{4\sqrt{2}C_K}\sqrt{\kappa_R}\ln(n) + \frac{m_1C_0}{4\sqrt{2}C_K}\ln(n)\sqrt{u},\end{aligned}$$

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

$$\begin{aligned}\int_0^\infty \exp\left(-\frac{n\sqrt{u+V_R(h)}}{4b_0}\right)du &\leq n^{-\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}}\int_0^\infty \exp\left(-\frac{m_1C_0\ln(n)}{4\sqrt{2}C_K}\sqrt{u}\right)du\\ &= \frac{64C_K^2}{m_1^2C_0^2}\int_0^\infty s\exp(-s)ds\frac{1}{\ln^2(n)n^{\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}}}\\ &= \frac{64C_K^2}{m_1^2C_0^2}\frac{1}{\ln^2(n)n^{\frac{m_1\sqrt{C_0}\sqrt{\kappa_R}}{4\sqrt{2}C_K}}} \leq \frac{64C_K^2}{m_1^2C_0^2}\frac{1}{n^\alpha}.\end{aligned}$$

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

6.2. **Proof of Theorem 1.**

6.2.1. *Main part of the proof of the Inequality (7).* Following Ferraty et al. (2006, 2010), we define

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

We also have $R_h^{X'} := \sum_{i=1}^n \widetilde{W}_h^{(i)}(X')$ (see Definition (17)). First, notice that since $\widehat{F}_h^{X'}\leq 1$ and $F^{X'}\leq 1$ a.s.,

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

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

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

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

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

The first and third terms are variance terms, bounded by Lemmas 3 and 4 proved below. The second one is a bias term, controlled by Lemma 5.

Lemma 3. *Under Assumptions H_K and H_φ, 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 4. *Under Assumptions H_K and H_φ*

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

Lemma 5. *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).

□

6.2.2. *Proof of Lemmas 3 and 4 (upper bounds for the variance terms).*
Proof of Lemma 3. By Fubini's Theorem

$$\begin{aligned}
\mathbb{E}_{X'}\left[\left\|\widetilde{F}_h^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}\right]\right\|_D^2\right] &= \int_D \mathbb{E}_{X'}\left[\left(\widetilde{F}_h^{X'}(y) - \mathbb{E}_{X'}\left[\widetilde{F}_h^{X'}(y)\right]\right)^2\right] dy \\
&= \int_D \mathrm{Var}_{X'}\left(\widetilde{F}_h^{X'}(y)\right) dy.
\end{aligned}$$

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 $\{X' \in B\}$,

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

where the last inequality comes from Inequality (16).

□

Proof of Lemma 4. Since $\mathbb{E}_{X'}\left[R_h^{X'}\right] = 1$, 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 (16).

6.2.3. *Proof of Lemma 5 (upper bound for the bias term).* First remark that, for $y \in D$, a.s.

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

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

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

Then,

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

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

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

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

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

□

6.3. **Proof of an intermediate result for Theorem 2: the case of known small ball probability.** This section is an introduction to the proof of Theorem 2: we first deal with the toy case of known small ball probability. It is thus possible to define a selection rule by $\tilde{h} = \operatorname{argmin}_{h\in\mathcal{H}_n}\{A(h)+V(h)\}$, with

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

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

Assume that H_K, H_φ, H_F, H_{b1} and H_{b2} hold. The pseudo-estimator $\widehat{F}_{\tilde{h}}$ is such that

$$\mathbb{E}\left[\left\|\widehat{F}_{\tilde{h}}^{X'} - F^{X'}\right\|_D^2 \mathbf{1}_B(X')\right] \le c\min_{h\in\mathcal{H}_n}\left\{h^{2\beta} + \frac{\ln(n)}{n\varphi(h)}\right\} + \frac{C}{n}, \tag{26}$$

where c and C are constants which depend on c_K, C_K, c_φ, C_φ, $|D|$, and C_D.

The proof of such inequality is simpler than Theorem 2, and is a good illustration of the model selection tools required to deal with a data-driven selected bandwidth. To prove Theorem 2, we will then come down to Inequality (26).

6.3.1. *Main part of the proof of Inequality (26).* Let $h \in \mathcal{H}_n$ be fixed. We start with the following decomposition for the loss of the estimator$\widehat{F}_{\tilde{h}}^{X'}$:

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

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

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

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

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

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

Lemma 6. *Let $h \in \mathcal{H}_n$ be fixed. Under the assumptions of Theorem 2, there exist two constants C and C_1 such that,*

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

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

Applying Inequality (28) in (27) implies Inequality (26) by taking the infimum over $h \in \mathcal{H}_n$.

□

6.3.2. *Proof of Lemma 6 (Upper bound for $A(h)$).* Fix $h, h' \in \mathcal{H}_n$. We define the set $\Omega_{h,h'} = \{R_{h'}^{X'} \geq 1/2\} \cap \{R_{h\vee h'}^{X'} \geq 1/2\}$ and split

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

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 (17). 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

(29)

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

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

$$\begin{aligned}
A(h) &= \max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 - V(h')\right)_+ \qquad (30)\\
&\leq \max_{h'\in\mathcal{H}_n}\left(4T_{h'}^a - \frac{V(h')}{3}\right)_+ + \max_{h'\in\mathcal{H}_n}\left(4T_{h\vee h'}^b - \frac{V(h')}{3}\right)_+\\
&\quad + \max_{h'\in\mathcal{H}_n}\left(4\widetilde{T}_{h\vee h'} - \frac{V(h')}{3}\right)_+\\
&\quad +4\max_{h'\in\mathcal{H}_n} B_{h,h'} + \max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}_{h'}^{X'} - \widehat{F}_{h\vee h'}^{X'}\right\|_D^2 \mathbf{1}_{\Omega^c_{h,h'}}\right).
\end{aligned}$$

We have not subtracted $V(h')$ to two of the above terms: we show below that they are directly negligible. We now deal with each of the terms involving in (30) on the set $\{X' \in B\}$.

• **Upper bound for the term depending on** $B_{h,h'}$. We first use the definition of the set $\Omega_{h,h'}$, and split the term to obtain the bias terms:

$$\begin{aligned}
\max_{h'\in\mathcal{H}_n} B_{h,h'} &\leq 4\max_{h'\in\mathcal{H}_n}\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2\\
&= 4\max_{\substack{h'\in\mathcal{H}_n\\ h'\leq h}}\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2\\
&\leq 8\max_{\substack{h'\in\mathcal{H}_n\\ h'\leq h}}\left\{\left\|\mathbb{E}_{X'}\left[\widetilde{F}_{h'}^{X'}\right] - F^{X'}\right\|_D^2 + \left\|F^{X'} - \mathbb{E}_{X'}\left[\widetilde{F}_{h\vee h'}^{X'}\right]\right\|_D^2\right\}\\
&\leq 8\left(\max_{\substack{h'\in\mathcal{H}_n\\ h'\leq h}} C_D^2(h')^{2\beta} + C_D^2 h^{2\beta}\right) \leq 16 C_D^2 h^{2\beta}, \qquad (31)
\end{aligned}$$

thanks to Lemma 5.

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

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

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

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

Moreover,

$$\begin{aligned}\mathbb{P}\left(\Omega^c_{h,h'}\cap\{X'\in B\}\right) &\leq \mathbb{P}\left(\left\{R^{X'}_{h'}<\frac{1}{2}\right\}\cap\{X'\in B\}\right)+\mathbb{P}\left(\left\{R^{X'}_{h\vee h'}<\frac{1}{2}\right\}\cap\{X'\in B\}\right)\\ &\leq \mathbb{P}\left(\left\{\left|R^{X'}_{h'}-1\right|>\frac{1}{2}\right\}\cap\{X'\in B\}\right)+\mathbb{P}\left(\left\{\left|R^{X'}_{h'\vee h}-1\right|>\frac{1}{2}\right\}\cap\{X'\in B\}\right).\end{aligned}$$

Thus we apply Inequality (18) of Lemma 1, with $\eta=1/2$:

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

Recall now that thanks to H_{b_2}, $\varphi(h)\geq C_0\ln(n)/n$ for all $h\in\mathcal{H}_n$, with $C_0>16(M_2/m_1^2+C_K/2m_1)$. Use also H_{b_1} 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

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(\left\|\widehat{F}^{X'}_{h'}-\widehat{F}^{X'}_{h\vee h'}\right\|^2_D\mathbf{1}_{\Omega^c_{h,h'}}\right)\mathbf{1}_B(X')\right]\leq\frac{16|D|}{n}. \tag{32}$$

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

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

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

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

Then, Inequality (19) of Lemma 1 (with $\alpha = 2$) proves that, on the set $\{X' \in B\}$, a.s.,

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

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 $\kappa > 768|D|\kappa_R$ in the definition of $V(h')$ (see (25)) leads to $V(h')/768|D| \geq V_R(h')$, and hence we also have

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

thanks to Assumption H_{b_1}. Since $\varphi(h) \leq \varphi(h\vee h')$, $V(h') \geq V(h\vee h')$, the other term is bounded as follows

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

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

$$
\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4\widetilde{T}_{h\vee h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \leq 256|D|\min\left(\frac{4M_2}{m_1^2}, \frac{64C_K^2}{m_1^2}\right)\frac{1}{n}. \tag{33}
$$

• **Upper bound for the terms depending on $T_{h'}^a$ or $T_{h'}^b$.** First, by definition of $\Omega_{h,h'}$, $T_{h'}^a \leq 4\|\widetilde{F}_{h'}^{X'} - \mathbb{E}_{X'}[\widetilde{F}_{h'}^{X'}]\|_D^2$. Furthermore, noticing that $\widetilde{F}_{h'}^{X'}$ belongs to $L^1(D)\cap L^2(D)$, the following equality is classical:

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

where $\bar{S}_D(0,1)$ is a dense countable subset of the sphere $\{t \in L^1(D) \cap L^2(D), \|t\|_D = 1\}$ (such a set exists thanks to the separability of $L^2(D)$). Moreover, we write the scalar product $\langle \widetilde{F}_{h'}^{X'} - \mathbb{E}_{X'}[\widetilde{F}_{h'}^{X'}], t\rangle_D \mathbf{1}_B(X') = \nu_{n,h}(t)$, for $t \in \bar{S}_D(0,1)$, where

$$\nu_{n,h}(t) = \frac{1}{n}\sum_{i=1}^{n} \psi_{t,h}(X_i,Y_i) - \mathbb{E}_{X'}\left[\psi_{t,h}(X_i,Y_i)\right], \tag{34}$$

$$\text{with } \psi_{t,h}(X_i,Y_i) = \frac{K_h(d(X_i,X'))}{\mathbb{E}_{X'}[K_h(d(X_i,X'))]}\langle \mathbf{1}_{[Y_i;\infty[}, t\rangle_D \mathbf{1}_B(X').$$

Consequently,

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

We use the following lemma, which permits to control the empirical process defined by (34).

Lemma 7. *Under the assumptions of Theorem 2, for $\delta_0 > \max(3528C_K^2|D|/M_2C_0, 12)$, there exists a constant $C > 0$ (depending only on m_1, M_2, δ_0, C_0 and $|D|$) such that*

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

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

$$\mathbb{E}\left[\max_{h'\in\mathcal{H}_n}\left(4T^a_{h'} - \frac{V(h')}{3}\right)_+ \mathbf{1}_B(X')\right] \le \frac{16C}{n}. \tag{35}$$

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

Gathering Inequalities (31), (32), (33), and (35) in Inequality (30) completes the proof of Lemma 6.

□

6.3.3. *Proof of Lemma 7 (concentration of the empirical process).* The aim is to control the deviations of the supremum of the empirical process $\nu_{n,h}$ defined by (34). Since it is centred and bounded, the guiding idea is to apply the following concentration inequality.

Lemma 8. *[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*

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

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

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

Inequality (8) is a classical consequence of the Talagrand Inequality given in Klein and Rio (2005): see for example Lemma 5 (page 812) in Lacour (2008).

We first compute H^ν, M^ν and v^ν, involved in Lemma 8.

- For M^ν, 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 (16).
- For H^ν, 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 3.
- For v^ν, we also fix $t \in \bar{S}_D(0,1)$, and compute,

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

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 8 gives, for $\delta > 0$,

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

We choose $\delta = \delta_0\ln(n)$, for a δ_0 large enough, and given below. We compute the order of magnitude of the last upper bound, using Assumptions H_{b_1} and H_{b_2}. Recall that they imply $\sum_{h\in\mathcal{H}_n} 1/\varphi(h) \leq n^2/C_0\ln(n)$ and $\sum_{h\in\mathcal{H}_n} 1/\varphi^2(h) \leq n^3/C_0^2\ln^2(n)$. First,

$$\sum_{h\in\mathcal{H}_n}\frac{1}{n\varphi(h)}\exp\left(-\frac{\delta}{6}\right) = \frac{1}{n^{1+\delta_0/6}}\sum_{h\in\mathcal{H}_n}\frac{1}{\varphi(h)} \leq \frac{1}{C_0 n^{\delta_0/6-1}\ln(n)} \leq \frac{1}{C_0 n},$$

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

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

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, δ_0, C_0 and $|D|$. This shows that

$$
\sum_{h\in\mathcal{H}_n}\mathbb{E}\left[\left(\sup_{t\in\bar{S}_D(0,1)}\nu_{n,h}^2(t)-2(1+2\delta_0\ln(n))\left(H^\nu\right)^2\right)_+\mathbf{1}_B(X')\right]\leq\frac{C}{n}, \tag{36}
$$

for $(H^\nu)^2=|D|(M_2/m_1^2)/(n\varphi(h))$ and $C>0$ depending only on m_1, M_2, δ_0, C_0 and $|D|$. Since

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

Inequality (36) is also satisfied when we replace $2(1+2\delta_0\ln(n))\left(H^\nu\right)^2$ by this upper bound. Thus, the proof of Lemma 7 is completed.

□

6.4. **Proof of Theorem 2.** The idea is to come down to Inequality (26). Let Λ 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

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

We will argue as follows: first, on the set Λ, $\widehat{\varphi}(h)$ is close to $\varphi(h)$, and we use the same arguments as for (26). Second, the probability of the set Λ^c is negligible. Let us prove these two claims.

• **Upper bound for $\|\widehat{F}_{\widehat{h}}^{X'}-F^{X'}\|_D^2\mathbf{1}_\Lambda$.** It follows from the same arguments as in the beginning of the proof of Inequality (26) (see Section 6.3.1), that

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

Note that

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

We obtain the following decomposition:

$$\begin{aligned}
(37) \qquad \left\| \widehat{F}_{\widehat{h}}^{X'} - F^{X'} \right\|_D^2 \mathbf{1}_\Lambda \;\leq\; & \Bigg\{ 6A(h) + 6V(h) + 3 \left\| \widehat{F}_h^{X'} - F^{X'} \right\|_D^2 \\
& 6 \max_{h'\in\mathcal{H}_n} \left(V(h') - \widehat{V}(h') \right)_+ + 6 \left(\widehat{V}(h) - V(h) \right) \Bigg\} \mathbf{1}_\Lambda .
\end{aligned}$$

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

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

But on the set Λ, 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 Λ, 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,

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

Gathering the two bounds in (37) leads to

$$\left\| \widehat{F}_{\widehat{h}}^{X'} - F^{X'} \right\|_D^2 \mathbf{1}_\Lambda \leq 6A(h) + 8V(h) + 3 \left\| \widehat{F}_h^{X'} - F^{X'} \right\|_D^2 .$$

We thus obtain, like in the proof of Inequality (26),

$$\mathbb{E}\left[\left\| \widehat{F}_{\widehat{h}}^{X'} - F^{X'} \right\|_D^2 \mathbf{1}_\Lambda \mathbf{1}_B(X') \right] \leq 8V(h) + 3C \left(h^{2\beta} + \frac{1}{n\varphi(h)} \right).$$

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

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

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 2), with $T_i = \mathbf{1}_{\{d(0,X_i)\le h\}}$ and $\eta = \varphi(h)/2$. Since $0 \le T_i \le 1$, we set $b_0 = 1$, and $v^2 = \mathrm{Var}(T_1) = \varphi(h)(1-\varphi(h))$. We derive

$$\mathbb{P}(\Lambda^c) \le 2\exp\left(-\frac{n\varphi^2(h)/8}{\varphi(h)(1-\varphi(h))+\varphi(h)/2}\right) = 2\exp\left(-\frac{n\varphi(h)}{8(1-\varphi(h))+4}\right) \le 2\exp\left(-\frac{n\varphi(h)}{12}\right).$$

We thus obtain

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

thanks to Assumptions H_{b1} and H_{b2}. Since $C_0 > 12$, $n^{1-C_0/12} < n^{-1}$ which ends the proof.

□

6.5. **Proof of Theorem 3.**

6.5.1. *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. 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, or at least to choose a good value for it.

Convergence rate under Assumption $H_{X,F}$. With the lower bound on φ of $H_{X,F}$, 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

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

leads to what we want to prove, that is

$$R(h_0) \le C\exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha} n\right). \tag{38}$$

Indeed, if $n > \exp\left(c_2^{1/(1-\alpha)}(2\beta-\gamma_1)_+^{\alpha/(\alpha-1)}\right)$, we have

$$\frac{1}{c_2}\ln n - c_2^{-(\alpha+1)/\alpha}\left(2\beta-\gamma_1\right)\ln^{1/\alpha} n > 0$$

and

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

Now, let $s_n := \exp\left(-\frac{2\beta}{c_2^{1/\alpha}}\ln^{1/\alpha} n\right)$, we have

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

using that,

$$(1-c_2^{-1/\alpha}(2\beta-\gamma_1)\ln^{1/\alpha-1} n)^{1/\alpha} = 1-\frac{1}{\alpha}c_2^{-1/\alpha}(2\beta-\gamma_1)\ln^{1/\alpha-1} n + o(\ln^{1/\alpha-1} n).$$

Then $h_0^{2\beta} \leq Cs_n$. We deal similarly with the second term of $R(h)$,

$$\begin{aligned}&\frac{n^{-1}h_0^{-\gamma_1}\exp(c_2\ln^\alpha(1/h))}{s_n}\\ &\qquad = \exp\left(\gamma_1 c_2^{-1/\alpha}\ln^{1/\alpha} n\left(1-\left(1-c_2^{-1/\alpha}(2\beta-\gamma_1)\ln^{1/\alpha-1}\right)^{1/\alpha}\right)\right)\\ &\qquad\qquad \to_{n\to\infty} 1,\end{aligned}$$

which leads to (38), and ends the computation of the rate under $H_{X,M}$.

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 $R(h) \leq Cn^{-2\beta/(2\beta+\gamma)}$.

□

6.5.2. *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.

□

6.6. **Proof of Theorem 4.**

6.6.1. *Proof of (i), under Assumption $H_{X,L}$.* The proof is based on the general reduction scheme described in Section 2.2 of Tsybakov (2009). 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 $(X_{0,i}, Y_{0,i})_{i=1,\dots,n}$ (resp. $(X_{1,i}, Y_{1,i})_{i=1,\dots,n}$) 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 (2009) (p.90), the results hold with c' independent on x_0. In the sequel, we define F_0 and F_1 and check the three conditions.

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 supported by $[0;1]$ 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 $0<L<1/(\sup_{n\in\mathbb{N}^*}\{\eta_n^\beta\}|D|\|K\|L^\infty(\mathbb{R})\|\psi\|_{L^\infty(\mathbb{R})})$.
- η_n is a non-negative real number such that

$$\eta_n^{2\beta}\geq c_{(B)}r_n \text{ and } \eta_n^{2\beta}\varphi^{x_0}(\eta_n)\leq\frac{c_{(C)}}{n}, \tag{39}$$

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

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

$$\eta_n=\left(\frac{\ln n-((2\beta+\gamma)/\alpha)\ln\ln n}{C_1}\right)^{-1/\alpha}. \tag{40}$$

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).

6.6.2. *Checks of the conditions (A) to (C).*
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 ψ is included in D) and if $y\in\mathring{D}$,

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

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 $I_\psi=\int_D(\int_{-\infty}^y\psi(t)dt)^2dy$,

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

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 η_n,

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

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). For any test function Φ on $\mathbb{H}\times\mathbb{R}$,

$$\begin{aligned}
\int_{\mathbb{H}\times\mathbb{R}} \Phi(x,y)d\mathbb{P}_1(x,y) &= \mathbb{E}\left[\Phi(X_0,Y_1)\right] = \mathbb{E}\left[\mathbb{E}\left[\Phi(X_0,Y_1)|X_0\right]\right],\\
&= \int_{\mathbb{H}} \mathbb{E}\left[\Phi(x,Y_1)|X_0 = x\right] d\mathbb{P}_{X_0}(x),\\
&= \int_{\mathbb{H}} \left(\int_{\mathbb{R}} \Phi(x,y) f_1^x(y)dy\right) d\mathbb{P}_{X_0}(x),\\
&= \int_{\mathbb{H}\times\mathbb{R}} \Phi(x,y)|D| f_1^x(y)\left(\frac{1}{|D|}\mathbf{1}_D(y)\right) dyd\mathbb{P}_{X_0}(x),\\
&= \int_{\mathbb{H}\times\mathbb{R}} \Phi(x,y)|D| f_1^x(y)d\mathbb{P}_0(x,y).
\end{aligned}$$

Consequently, $\mathbb{P}_1 << \mathbb{P}_0$, and $d\mathbb{P}_1/d\mathbb{P}_0(x,y) = |D|f_1^x(y)$. This enables to compute the Kullback distance

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

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

$$\begin{aligned}
K(\mathbb{P}_1,\mathbb{P}_0) &\le \mathbb{E}\left[\int_D |D|L\eta_n^\beta H\left(\frac{\|X_0-x_0\|}{\eta_n}\right)\psi(y)dy\right]\\
&\quad + \mathbb{E}\left[\int_D |D|\left(L\eta_n^\beta H\left(\frac{\|X_0-x_0\|}{\eta_n}\right)\psi(y)\right)^2 dy\right],\\
&= 0 + |D|L^2\eta_n^{2\beta}\int_D \psi^2(y)dy\mathbb{E}\left[H^2\left(\frac{\|X_0-x_0\|}{\eta_n}\right)\right],\\
&\le |D|L^2\eta_n^{2\beta}\|\psi\|^2_{L^2(\mathbb{R})}\|H\|^2_{L^\infty(\mathbb{R})}\mathbb{P}\left(\|X_0-x_0\| \le \eta_n\right)\\
&= |D|L^2\eta_n^{2\beta}\|\psi\|^2_{L^2(\mathbb{R})}\|H\|^2_{L^\infty(\mathbb{R})}\varphi^{x_0}(\eta_n),
\end{aligned}$$

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 η_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).

□

6.6.3. *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 (40)).

- Under $H_{X,M}$, we set $r_n = \exp\left(-\frac{2\beta}{c_1^{1/\alpha}} \ln^{1/\alpha} n\right)$, and replace the previous choice (40) of η_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 (39).
- 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 $r_n = n^{-2\beta/(2\beta+\gamma)}$.

□

6.6.4. *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 Ω_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_ω belongs to $\mathcal{F}_\beta$, for all $\omega \in \Omega_n$,

(B') For all $\omega, \omega' \in \Omega_n$, $\omega \neq \omega'$, $\mathbb{E}\left[\|F_\omega^{X'} - F_{\omega'}^{X'}\|_D^2 \mathbf{1}_B(X')\right] \geq c r_n$ 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 ζ, where $\mathbb{P}_\omega^{\otimes n}$ is the probability distribution of a sample $(X_{\omega,i}, Y_{\omega,i})_{i=1,\ldots,n}$ for which the conditional c.d.f. of $Y_{\omega,i}$ given $X_{\omega,i}$ is given by F_ω.

Then the result comes from Theorem 2.5 of Tsybakov (2009) (p.85-86). We follow the same steps as previously.

6.6.5. *Construction of the set of hypotheses F_ω 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 $\omega = (\omega_1, \ldots, \omega_{m_n}) \in \{0,1\}^{m_n}$, 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 ψ, H, L, and $(\eta_n)_n$ are introduced in the proof of (i) (a good choice of η_n is (40)), 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 $\mathbb{L}^2([0,1])$.

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

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

$$H\left(\frac{\|x-x_k\|}{\eta_n}\right)H\left(\frac{\|x-x_j\|}{\eta_n}\right)=0 \text{ for all } x\in\mathbb{H},\ \text{as soon as } j\neq k. \tag{41}$$

Indeed, suppose that $H(\|x-x_k\|/\eta_n)\neq 0$, since H is supported on $[0,1]$, we have $\|x-x_k\|\leq\eta_n$. Now remark that, as $(e_j)_{j\geq1}$ is an orthonormal basis, for all $j\neq k$, $\|x_j-x_k\|^2=2\sup_{n\in\mathbb{N}^*}\{\eta_n^2\}\left(\|e_j\|^2-2\langle e_j,e_k\rangle+\|e_k\|^2\right)=4\sup_{n\in\mathbb{N}^*}\{\eta_n^2\}$. 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 $H(\|x-x_k\|/\eta_n)=0$.

6.6.6. *Checks of the conditions (A') to (C').*
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^x_ω is non increasing ($x\in\mathbb{H}$ fixed), as for F^x_1, we bound,

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

for $y\in D$, thanks to Property (41) and the definition of L above. Thus, as F_1 in the proof of (i), F_ω is a conditional distribution function, and we also similarly obtain $F_\omega\in\mathcal{F}_\beta$.
Check (B'). For all $\omega,\omega'\in\{0,1\}^{m_n}$,

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

with I_ψ defined in the proof of (i). From Property (41), we get:

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

Now set $c_H:=\min_{x,\|x\|\leq1/2}H(x)$, since H is continuous and $H(x)>0$ for all $x\in\mathbb{H}$, such that $\|x\|\leq1$, we have $c_H>0$ and

$$\begin{aligned}\mathbb{E}\left[H^2\left(\frac{\|X'-x_k\|}{\eta_n}\right)\mathbf{1}_B(X')\right] &\geq \mathbb{E}\left[H^2\left(\frac{\|X'-x_k\|}{\eta_n}\right)\mathbf{1}_{\left\{\frac{\|X'-x_k\|}{\eta_n}\leq1/2\right\}}\mathbf{1}_B(X')\right]\\ &\geq c_H^2\mathbb{P}\left(\{\|X'-x_k\|\leq\eta_n/2\}\cap\{X'\in B\}\right).\end{aligned}$$

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 ρ. Then, as soon as, $\rho>(1/2+\sqrt{2})\sup_{n\in\mathbb{N}}\{\eta_n\}$, we have $\{\|X'-x_k\|\leq\eta_n/2\}\subset\{\|X'\|\leq\rho\}\subset\{X'\in B\}$. Then, since $\|x_k\|<\rho$, we also have $x_k\in B$ and we can apply Condition (13) to get a lower bound on the shifted small ball probability $\mathbb{P}\left(\|X'-x_k\|\leq\eta_n/2\right)=\varphi^{x_k}(\eta_n/2)$. We get

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

and

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

where ρ is the Hamming distance on $\{0,1\}^{m_n}$ defined by $\rho(\omega,\omega')=\sum_{j=1}^{m_n}\mathbf{1}_{\{\omega_k\neq\omega_k\}}$. Now, from Varshamov-Gilbert bound (Lemma 2.7 of Tsybakov 2009), there exists a subset Ω_n

of $\{0,1\}^{m_n}$ such that

$$\rho(\omega,\omega') \geq \frac{m_n}{8}, \text{ for all } \omega,\omega' \in \Omega_n, \omega \neq \omega', \quad \text{and} \quad \mathrm{Card}(\Omega_n) \geq 2^{m_n/8}. \tag{42}$$

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

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

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

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

Now, arguing again as in Check (A'), we can apply Assumption (13) 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 η_n, we now obtain (as in (i))

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

Finally, condition (42) on the cardinal of Ω_n leads to $m_n \leq (8/\ln 2)\ln(\mathrm{Card}(\Omega_n))$, which completes the proof of (C'), and at the same time the proof of all the lower bounds.

$\square$

6.7. **Proof of Proposition 1.**

6.7.1. *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 3 and Lemmas 4 remain valid (by replacing again d by d_p in every terms, and by using H'_φ instead of H_φ). 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 5 the following result, the proof of which can be found below. This ends the proof.

Lemma 9. *Suppose that Assumptions H_F and H_ξ 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

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

where $C>0$ only depends on C_D, β, and C_ξ.

$\square$

6.7.2. *Proof of Lemma 9.* Let us begin with the first inequality (integrated risk). Like in the proof of Lemma 5, we also obtain (24). Then,

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

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 5 lead to

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

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 $\mathbb{E}_{X'}\left[\widetilde{W}_h^{(1)}(X')\right] = 1/n$. We thus obtain

$$\begin{aligned} \mathbb{E}\left[\left\|F^{X'}-\mathbb{E}_{X'}\left[\widetilde{F}_{h,p}^{X'}\right]\right\|_D^2\right] \;&\leq \\ &2^{\beta+1}C_D^2\left(h^{2\beta}+n^2\mathbb{E}\left[\mathbb{E}_{X'}\left[\widetilde{W}_h^{(i)}\left(\sum_{j>p}\sigma_j^2(\xi_j^{(i)}-\xi_j')^2\right)^{\beta/2}\right]^2\right]\right). \qquad (43) \end{aligned}$$

Under Assumption H_ξ^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_ξ^{ind}, remark that

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

Now applying Lemma 10 below with $\eta_j = \xi_j^{(1)} - \xi_j'$ and $C_M = 2^\beta C_\xi$, we get

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

and the result comes from Inequality (43). The proof of the first inequality of Lemma 9 is completed.

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

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

The final bound then follows similarly.

□

Lemma 10. *Let* $(\eta_j)_{j\geq 1}$ *a sequence of real random variables and* $(\sigma_j)_{j\geq 1}$ *a sequence a real numbers verifying, for* $\beta > 0$*,*

$$
\sum_{j\geq 1}\sigma_j^2 < +\infty \text{ and } \forall j \geq 1,\ \mathbb{E}\left[\eta_j^\beta\right] \leq C_M,
$$

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

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

Proof of Lemma 10. 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

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

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:

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

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

□

6.8. **Proof of Corollary 1.** The proof is based on the same ideas as the ones used to prove Theorem 3 in Section 6.5. We begin with the result (ii) (integrated risk).

6.8.1. *Proof of* (ii). Thanks to Proposition 1 (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

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

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

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

we get

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

where c only depends on $\sum_{j\ge1}\sigma_j^2$ and c_f. With the assumption on σ_j, we thus obtain the following upper bound for the risk, up to a constant $R(h,p) := h^{2\beta} + p^{\beta(1-2\delta)} + c^{-p}n^{-1}h^{-p}$. We then compute the partial derivatives

$$\begin{aligned}
\frac{\partial R}{\partial h}(h,p) &= 2\beta h^{2\beta-1} - pc^{-p}n^{-1}h^{-p-1},\\
\frac{\partial R}{\partial p}(h,p) &= \beta(1-2\delta)p^{\beta(1-2a)-1} - \ln(ch)c^{-p}n^{-1}h^{-p}.
\end{aligned}$$

If (h^*,p^*) is the minimizer of R we have $\frac{\partial R}{\partial h}(h^*,p^*) = 0$, which leads to

$$h^* = \left(\frac{p^* c^{-p^*}}{2\beta}\right)^{1/(2\beta+p^*)} n^{-1/(2\beta+p^*)}.$$

Moreover, for all $p\in\mathbb{N}^*$,

$$\begin{aligned}
R(h^*,p^*) &= \left(\frac{p^* c^{-p^*}}{2\beta}\right)^{2\beta/(2\beta+p^*)} n^{-2\beta/(2\beta+p^*)} + (p^*)^{\beta(1-2\delta)}\\
&\quad + c^{-p}n^{-1}\left(\frac{p^* c^{-p^*}}{2\beta}\right)^{-p^*/(2\beta+p^*)} n^{-p^*/(2\beta+p^*)}\\
&\le \left(\frac{pc^{-p}}{2\beta}\right)^{2\beta/(2\beta+p)} n^{-2\beta/(2\beta+p)} + p^{\beta(1-2\delta)}\\
&\quad + c^{-p}n^{-1}\left(\frac{pc^{-p}}{2\beta}\right)^{-p/(2\beta+p)} n^{-p/(2\beta+p)},
\end{aligned}$$

and this last bound has the order $n^{-2\beta/(2\beta+p)} + p^{\beta(1-2\delta)}$. Choosing $h = (pc^{-p}/2\beta)^{1/(2\beta+p)}\, n^{-1/(2\beta+p)}$ and $p = [\ln(n)/(\delta-1/2)\ln\ln(n) - 2\beta]$ gives the result.

6.8.2. *Proof of* (i). We deduce from Proposition 1 (ii), from the assumption $\sum_{j>p}\langle x_0, e_j\rangle \le C\sum_{j>p}\sigma_j^2$, and from the left-hand-side inequality of H_φ 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).

□

References

A. Ait-Saïdi, F. Ferraty, R. Kassa, and P. Vieu. Cross-validated estimations in the single-functional index model. *Statistics*, 42(6):475–494, 2008.

N. Akakpo and C. Lacour. Inhomogeneous and anisotropic conditional density estimation from dependent data. *Electron. J. Stat.*, 5:1618–1653, 2011.

T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.*, 6:170–176, 1955.

R. B. Ash and M. F. Gardner. *Topics in stochastic processes*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. Probability and Mathematical Statistics, Vol. 27.

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

N. Azzedine, A. Laksaci, and E. Ould-Saïd. On robust nonparametric regression estimation for a functional regressor. *Statist. Probab. Lett.*, 78(18):3216–3221, 2008.

L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

E. Brunel, F. Comte, and C. Lacour. Minimax estimation of the conditional cumulative distribution function. *Sankhya A*, 72(2):293–330, 2010.

E. Brunel, A. Mas, and A. Roche. Non-asymptotic Adaptive Prediction in Functional Linear Models. Submitted, hal-00763924, 2013.

F. Burba, F. Ferraty, and P. Vieu. k-nearest neighbour method in functional nonparametric regression. *J. Nonparametr. Stat.*, 21(4):453–469, 2009.

T. T. Cai and P. Hall. Prediction in functional linear regression. *Ann. Statist.*, 34(5): 2159–2179, 2006.

H. Cardot, F. Ferraty, and P. Sarda. Functional linear model. *Statist. Probab. Lett.*, 45 (1):11–22, 1999.

G. Chagny. Penalization versus goldenshluger-lepski strategies in warped bases regression. *ESAIM Probab. Statist.*, 17:328–358 (electronic), 2013a.

G. Chagny. *Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie.* PhD thesis, Univ. Paris Descartes, July 2013b. URL `http://tel.archives-ouvertes.fr/tel-00863141`.

D. Chen, P. Hall, and H.-G. Müller. Single and multiple index functional regression models with nonparametric link. *Ann. Statist.*, 39(3):1720–1747, 2011.

F. Comte and V. Genon-Catalot. Convolution power kernels for density estimation. *J. Statist. Plann. Inference*, 142(7):1698–1715, 2012.

F. Comte and J. Johannes. Adaptive functional linear regression. *Ann. Statist.*, 40(6): 2765–2797, 2012.

F. Comte and C. Lacour. Pointwise deconvolution with unknown error distribution. *C. R. Math. Acad. Sci. Paris*, 348(5-6):323–326, 2010.

C. Crambes, L. Delsol, and A. Laksaci. Robust nonparametric estimation for functional data. *J. Nonparametr. Stat.*, 20(7):573–598, 2008.

C. Crambes, A. Kneip, and P. Sarda. Smoothing splines estimators for functional linear regression. *Ann. Statist.*, 37(1):35–72, 2009.

S. Dabo-Niang and N. Rhomari. Kernel regression estimation in a Banach space. *J. Statist. Plann. Inference*, 139(4):1421–1434, 2009.

S. Dabo-Niang and A-F. Yao. Kernel spatial density estimation in infinite dimension space. *Metrika*, 76(1):19–52, 2013.

S. Dabo-Niang, Z. Kaid, and A. Laksaci. On spatial conditional mode estimation for a functional regressor. *Statist. Probab. Lett.*, 82(7):1413–1421, 2012.

A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *Ann. Statist.*, 38(2):1171–1193, 2010.

J. Demongeot, A. Laksaci, F. Madani, and M. Rachdi. Estimation locale linéaire de la densité conditionnelle pour des données fonctionnelles. *C. R. Math. Acad. Sci. Paris*, 348(15-16):931–934, 2010.

T. Dunker, M. A. Lifshits, and W. Linde. Small deviation probabilities of sums of independent random variables. In *High dimensional probability (Oberwolfach, 1996)*, volume 43 of *Progr. Probab.*, pages 59–74. Birkhäuser, Basel, 1998.

F. Ferraty and Y. Romain. *The Oxford Handbook of Functional Data Analysis*. Oxford Handbooks in Mathematics. OUP Oxford, 2011.

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

F. Ferraty and P. Vieu. *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York, 2006. Theory and practice.

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

F. Ferraty, A. Mas, and P. Vieu. Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.*, 49(3):267–286, 2007.

F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu. Rate of uniform consistency for nonparametric estimates with functional variables. *J. Stat. Plan. Inference*, 140(2):335–352, February 2010.

G. Geenens. Curse of dimensionality and related issues in nonparametric functional regression. *Stat. Surv.*, 5:30–43, 2011.

A. Gheriballah, A. Laksaci, and S. Sekkal. Nonparametric M-regression for functional ergodic data. *Statist. Probab. Lett.*, 83(3):902–908, 2013.

I. Gijbels, M. Omelka, and N. Veraverbeke. Multivariate and functional covariates and conditional copulas. *Electron. J. Stat.*, 6:1273–1306, 2012.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011.

J. Hoffmann-Jørgensen, L. A. Shepp, and R. M. Dudley. On the lower tail of Gaussian seminorms. *Ann. Probab.*, 7(2):319–342, 1979.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.

C. Lacour. Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris*, 342(11):877–882, 2006.

C. Lacour. Adaptive estimation of the transition density of a particular hidden Markov chain. *J. Multivariate Anal.*, 99(5):787–814, 2008.

N. Laib and D. Louani. Nonparametric kernel regression estimation for functional stationary ergodic data: asymptotic properties. *J. Multivariate Anal.*, 101(10):2266–2281, 2010.

W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland, Amsterdam, 2001.

M. A. Lifshits. On the lower tail probabilities of some random series. *Ann. Probab.*, 25 (1):424–442, 1997.

A. Mas. Lower bound in regression for functional data by representation of small ball probabilities. *Electron. J. Stat.*, 6:1745–1778, 2012.

E. Masry. Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, 115(1):155–177, 2005.

S. Plancade. Adaptive estimation of the conditional cumulative distribution function from current status data. *J. Statist. Plann. Inference*, 143(9):1466–1485, 2013.

M. Rachdi and P. Vieu. Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference*, 137(9):2784–2801, 2007.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.

A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by V. Zaiats.