ADAPTIVE ESTIMATION OF HAZARD RATE WITH CENSORED DATA.

E. BRUNEL¹ AND F. COMTE²

Abstract. This paper studies hazard rate estimation for right-censored data by projection estimator methods. We consider projection spaces generated by trigonometric or piecewise polynomials, splines or wavelet bases. We prove that the estimator reaches the standard optimal rate associated with the regularity of the hazard function, provided that the dimension of the projection space is relevantly chosen. Then we provide an adaptive procedure leading to an automatic choice of this dimension via a penalized minimum contrast estimator and automatically reaching the optimal rate. Our procedure is based on a random penalty function and is completely data driven. A small simulation experiment is provided, that compares our result with some others.

September 2004


1. INTRODUCTION

In medical trials, reliability and many other fields, the occurrence of the event of interest called lifetime (or time to failure) forms the modelling basis, although often these times are not completely observed. The best known example of such incomplete observation is right-censoring. Not all of a set of lifetimes are observed, so that for some individual under study, it is only known that his lifetime is larger than some given value. In these settings, hazard rate estimation for the lifetime event is a basic tool for processing survival analysis.

Many methods for hazard estimation have been considered in the literature, and in particular nonparametric ones have known an important recent development. Let us be more precise with this nonparametric setting. Patil (1997) considers a wavelet estimator for uncensored data but his approach can not be easily extended to censored data. Dyadic linear wavelets for incomplete data are used by Antoniadis et al. (1999). They propose a two-step procedure to estimate the density function of the lifetime and then define the hazard estimator by taking the ratio with some estimator of the survival function. The estimator is proved to achieve the best possible

¹ IUT de Paris V et Laboratoire MAP5, UMR CNRS 8145. email: brunel@math-info.univ-paris5.fr.
² Université Paris V, MAP5, UMR CNRS 8145. email: fabienne.comte@univ-paris5.fr.
convergence rate in the MISE sense; nevertheless their estimator is not adaptive and
the optimal resolution of the wavelet basis depends on the regularity of the function
to be estimated. Most recently, Wu and Wells (2003) consider another wavelet-type
estimator for left-truncated and right-censored data. Their estimator is directly
based on the wavelet transform of the Nelson-Aalen cumulative hazard estimator
rather than a two-step estimator. They provide an asymptotic formula for the MISE
and also study asymptotic normality of the estimator. Their results are not adaptive
with respect to the smoothness of the hazard function to be estimated either. Let
us mention also Kooperberg et al. (1995) who study the $L^2$ convergence rate of an
hazard rate estimator in a context of tensor product splines. An adaptive sieved
maximum-likelihood method is proposed by Döhler and Rüschendorf (2002), who
consider an original $\Lambda$-distance (which is equivalent to the $L^2$-norm): the resulting
estimator reaches the optimal rates, up to logarithmic factors only. The consistency
results are derived from Vapnik-Cervonenkis techniques. Moreover, as in Kooper-
berg et al. (1995), their work takes place in a multivariate framework conditionally
to a vector of covariates. A very general study is done by Reynaud-Bouret (2002)
in a general context embedding adaptive hazard rate estimation. Indeed, adaptive
results are obtained for penalized projection estimators of the Aalen multiplicative
intensity for counting processes. However, the only projection basis considered is a
basis of piecewise constant functions. An analogous method was used by Castellan
and Letué (2001) for the Cox model with right censorship.

Our method is related to model selection methods introduced by Barron and
Cover (1991), Birgé and Massart (1997) and Barron et al. (1999). Those technics
aim to an automatic (and data-driven) squared bias/variance compromise via some
relevant penalization of a contrast function. They generally involve the study of
some supremum of an empirical process and an important tool is provided by Tal-
lagrand’s (1996) inequality. Indeed, this inequality gives precisely some bounds for
the deviations of the supremum of some empirical centered process with respect to
its expectation. Another technical key in the present work is the strong representa-
tion of the Kaplan-Meier estimator via the original influence curves decomposition
initiated first by Reid (1981), and further used by Mielniczuk (1985), Delecroix and
Yazourh (1991) and independently, Lo et al. (1989). Once the decomposition is
available for a large variety of projection spaces, all proofs become very simple (con-
trary to Döhler and Rüschendorf’s (2002)). Consequently, our approach provides a
very general method allowing a wide range of models.

The plan of the paper is as follows. After the description in Section 2 of the
notations and assumptions, we present in Section 3 the study of one estimator
which is based on a projection contrast function. We describe the projection spaces
(namely the spaces generated by trigonometric polynomials, piecewise polynomials
or wavelets; the splines are considered apart from the others). We give the key
decomposition and some non asymptotic and asymptotic results, illustrating the
squared-bias-variance decomposition and the minimax rates that can be reached
2. NOTATIONS AND ASSUMPTIONS

We consider nonnegative i.i.d. random variables $X^0_i$, for $i = 1, \ldots, n$ (lifetimes for the $n$ subjects under study) with common continuous distribution function $F^0$, and $C_1, \ldots, C_n$ i.i.d. nonnegative random variables (“censoring sequence”) with common distribution function $G$, both sequences being independent. One classical problem when processing with lifetime data is the estimation of the hazard rate function or failure rate function $h$ defined, if $F^0$ has a density $f^0$ by

$$h(x) = \frac{d}{dx} H(x) = \frac{f^0(x)}{F^0(x)}, \quad \text{for } F^0(x) < 1.$$  

where $H = -\log(F^0)$ is called the cumulative hazard rate and $\bar{F}^0 = 1 - F^0$ is the survival function. In the setting of survival analysis data with random right censorship, one observes the bivariate sample $(X_1, \delta_1), \ldots, (X_n, \delta_n)$, where

$$X_i = X^0_i \wedge C_i, \quad \delta_i = I_{\{X^0_i \leq C_i\}}.$$  

In other words, $\delta_i = 1$ indicates that the $i^{th}$ subject’s time is uncensored. We denote by $F$ the common distribution function of the $X_i$’s and note that $\bar{F} = 1 - F = (1 - F^0)(1 - G)$.

Since $h$ is not square integrable on $\mathbb{R}^+$ for standard survival laws (like exponentials), the estimator of $h$ is standardly built on some interval $[0, b]$ (see e.g. Dölher and Rüschendorf (2002)). As mentioned in Delecroix and Yazourh (1992), this does not imply any practical restriction since, for estimation purpose, we can choose $b$ greater than the largest of the uncensored $X^0_i$’s. For sake of simplicity and without loss of generality, we set $b = 1$ in all the following.

A standard estimate of the survival function $\bar{F}^0$ is due to Kaplan and Meier (1958) and is defined in function of the $X_{(j)}$’s, where $X_{(j)}$ is the $j^{th}$ order statistic for the sample $(X_1, \ldots, X_n)$, by:

$$KM_n(x) = \left\{ \begin{array}{ll}
\prod_{i=1}^{n} \left( \frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{if } x \leq X_{(n)} \\
0 & \text{if } x > X_{(n)}. \end{array} \right.$$  

Here, $\delta_{(i)}$ is the induced order statistic corresponding to $X_{(i)}$. Note that this definition induces that the largest observation is supposed to be uncensored, whether or not it is. Now, in order to avoid the possibility that $KM_n(x) = 0$, since our purpose is to build an empirical estimator of the cumulative hazard rate function,
the Kaplan-Meier estimator is modified as in Lo et al. (1989). The survival function $\bar{F}^0$ is estimated by $\tilde{F}_n^0$ defined by

\[
\tilde{F}_n^0(x) = \begin{cases} 
\prod_{i=1,X(i) \leq x}^{n-i+1} \left( \frac{n-i+1}{n-i+2} \right)^{\delta(i)} & \text{if } x \leq X(n) \\
\frac{1}{n-i+2} \delta(i) & \text{if } x > X(n).
\end{cases}
\]

Useful properties are the following: $\tilde{F}_n^0(x) \geq \left( \frac{n+1}{n+2} \right)^{-1}$ for all $x$ and $\sup_{0 \leq x \leq T} |\bar{K}_n(x) - \tilde{F}_n^0(x)| = O(n^{-1})$ a.s. for any $0 < T < \inf \{ t \geq 0 : F(t) = 1 \}$.

3. Study of one estimator

3.1. Definition of the estimator. We consider the following contrast function

\[
\gamma_n(t) = \|t\|^2 - 2 \int_0^1 t(x) dH_n(x)
\]

where $t$ is a function of $L^2([0, 1])$, $\|t\|^2 = \int_0^1 t^2(x) dx$ and $H_n = -\ln(\tilde{F}_n^0)$, with $\tilde{F}_n^0$ defined by (2.1) and therefore

\[
\int_0^1 t(x) dH_n(x) = - \sum_{i/X(i) < 1} \delta(i) t(X(i)) \ln \left( 1 - \frac{1}{n-i+2} \right).
\]

We consider a collection of models $(S_m)_{m \in M}$ where each $S_m$ is a finite dimensional sub-space of $L^2([0, 1])$ with dimension $D_m$, generated by a basis $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ of functions where $\text{card}(\Lambda_m) = D_m$. Let then

\[
\hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t).
\]

An explicit expression of the estimator follows from this definition if the functions $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ are a collection of orthonormal functions:

\[
\hat{h}_m = \sum_{\lambda \in \Lambda_m} \hat{a}_\lambda \varphi_\lambda \text{ with } \hat{a}_\lambda = \int_0^1 \varphi_\lambda(x) dH_n(x).
\]

Note that one can easily write that

\[
\hat{h}_m(x) = - \sum_{i/X(i) < 1} \delta(i) K_m(X(i), x) \ln \left( 1 - \frac{1}{n-i+2} \right)
\]

with $K_m(X(i), x) = \sum_{\lambda \in \Lambda_m} \varphi_\lambda(X(i)) \varphi_\lambda(x)$. This expression shows that our estimator presents some analogy with a generalized kernel estimator by replacing $K_m(\cdot, x)$ with a kernel function.

Let also $h_m$ denote the orthogonal projection of $h$ on $S_m$. We can write
ADAPTIVE ESTIMATION OF HAZARD RATE

(3.3) \[ h_m = \sum_{\lambda \in \Lambda_m} a_{\lambda} \varphi_{\lambda} \] with \( a_{\lambda} = \int_0^1 \varphi_{\lambda}(x) dH(x). \)

It follows from (3.2), (3.3) and Pythagoras theorem that

\[
\|h - \hat{h}_m\|^2 = \|h - h_m\|^2 + \|h_m - \hat{h}_m\|^2
\]

\[
= \|h - h_m\|^2 + \sum_{\lambda \in \Lambda_m} (a_{\lambda} - \hat{a}_{\lambda})^2
\]

\[
= \|h - h_m\|^2 + \sum_{\lambda \in \Lambda_m} \left( \int_0^1 \varphi_{\lambda}(x) dH_n(x) - \int_0^1 \varphi_{\lambda}(x) dH(x) \right)^2.
\]

Let us define for a general \( t \) function and a real number \( a \in [0, 1] \),

(3.4) \[ C(t, a) = \int_0^a t(x) dH_n(x) - \int_0^a t(x) dH(x). \]

Then we have

(3.5) \[ \|h - \hat{h}_m\|^2 = \|h - h_m\|^2 + \sum_{\lambda \in \Lambda_m} C^2(\varphi_{\lambda}, 1). \]

The main difficulty related to a non-orthogonal basis (like splines that will be described in the next section) is that the explicit formula given for the coefficients \( \hat{a}_{\lambda} \) in formula (3.2) is no longer true. As a consequence, formula (3.5) which can be seen as a kind of bias-variance decomposition does not hold anymore. In the case of a non-orthogonal basis, the definition (3.1) of the contrast remains valid and is used directly

\[ \gamma_n(t) - \gamma_n(s) = \|t - h\|^2 - \|s - h\|^2 - 2 \int_0^1 (t - s) d(H_n - H)(x). \]

Then if \( h_m \) still denotes the projection of \( h \) on \( S_m \), \( \gamma_n(\hat{h}_m) \leq \gamma_n(h_m) \) implies

\[
\|\hat{h}_m - h\|^2 \leq \|h_m - h\|^2 + 2 C(\hat{h}_m - h_m, 1) \leq \|h_m - h\|^2 + 2 \|\hat{h}_m - h_m\| \sup_{t \in S_m, \|t\| = 1} |C(t, 1)|
\]

\[
\leq \|h_m - h\|^2 + \frac{1}{4} \|\hat{h}_m - h_m\|^2 + 4 \sup_{t \in S_m, \|t\| = 1} |C(t, 1)|^2,
\]

and since \( \|\hat{h}_m - h_m\|^2 \leq 2 \|h_m - h\|^2 + 2 \|\hat{h}_m - h\|^2 \), we have

(3.6) \[ \|\hat{h}_m - h\|^2 \leq 3 \|h_m - h\|^2 + 8 \left( \sup_{t \in S_m, \|t\| = 1} |C(t, 1)|^2 \right). \]
It appears that (3.6) can replace (3.5) when non orthogonal bases are considered.

3.2. Description of the spaces of approximation. In this section, we describe the spaces \((S_m)_{m \in \mathcal{M}_n}\) considered in the sequel and we point out the key properties that they satisfy. They are all considered as subspaces of \(L^2([0,1])\).

3.2.1. Orthonormal basis.

[T] Trigonometric spaces: \(S_m\) is generated by \(\{1, \sqrt{2}\cos(2\pi j x), \sqrt{2}\sin(2\pi j x) \mid j = 1, \ldots, m\}\), \(D_m = 2m + 1\) and \(\mathcal{M}_n = \{1, \ldots, [n/2] - 1\}\).

[P] Regular piecewise polynomial spaces: \(S_m\) is generated by \(m(r+1)\) polynomials, \(r+1\) polynomials of degree 0, 1, \ldots, \(r\) on each subinterval \([\{(j-1)/m, j/m\} \mid j = 1, \ldots, m, D_m = (r+1)m, m \in \mathcal{M}_n = \{1, 2, \ldots, [n/(r+1)]\}\). For example, consider the orthogonal collection in \(L^2([-1,1])\) of Legendre polynomials \(Q_k\), where the degree of \(Q_k\) is equal to \(k\), \(|Q_k(x)| \leq 1, \forall x \in [-1,1], Q_k(1) = 1\) and \(\int_{-1}^{1} Q_k^2(u) du = 2/(2k + 1)\). Then the orthonormal basis is given by \(\varphi_{j,k}(x) = \sqrt{m(2k + 1)}Q_k(2mx - 2j + 1)I_{[(j-1)/m,j/m]}(x) \mid j = 1, \ldots, m\) and \(k = 0, \ldots, r\), with \(D_m = (r + 1)m\). In particular, the histogram basis corresponds to \(r = 0\) and is simply defined by \(\varphi_j(x) = \sqrt{D_m} I_{[(j-1)/D_m,j/D_m]}(x)\) and \(D_m = m\). We call dyadic collection of piecewise polynomials, and denote by [DP], the collection corresponding to dyadic subdivisions with \(m = 2^q\) and \(D_m = (r + 1)2^q\).

[W] Dyadic wavelet generated spaces with regularity \(r\) and compact support, as described e.g. in Donoho and Johnstone (1994): \(S_m\) is generated by \(\{\phi_{j_0,k}, \psi_{j,k} \mid k \in \mathbb{Z}, m \geq j \geq j_0\}\) for any fixed resolution \(j_0\) and with \(\phi_{j_0,k}(x) = \sqrt{2^{j_0}} \phi(2^{j_0} x - k)\) and \(\psi_{j,k}(x) = \sqrt{2^j} \psi(2^j x - k)\) where \(\phi\) and \(\psi\) denote respectively the scaling function and the mother wavelet on \([0,1]\) and are elements of the Hölder space \(C^r, r \geq 0\). In this case, the multi-resolution analysis is said to be \(r\) regular. Moreover, the wavelet \(\psi\) has vanishing moments up to order \(r\) (see for example Daubechies, 1992). Since \(\phi\) and \(\psi\) are compactly supported on \([0,1]\), for any fixed \(j\) the sum over \(k\) is finite in the wavelet serie, more precisely for a function \(t \in S_m\),

\[
t(x) = \sum_{k=0}^{2^{j_0}-1} a_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{2^r-1} \sum_{k=0}^{m} b_{j,k} \psi_{j,k}(x).
\]

Therefore, one can see that the generating basis is of cardinality \(D_m = 2^{m+1}\) and \(m \in \mathcal{M}_n = \{1, 2, \ldots, [\ln(n)/2] - 1\}\).

Note that all those collections of models satisfy the following properties: in addition of being finite dimensional linear spaces of dimension \(D_m\), they are such that:

\[
(3.7) \quad \exists \Phi_0 > 0, \forall t \in S_m, \|t\|_\infty \leq \Phi_0 \sqrt{D_m} \|t\|
\]
and

\[ \exists \Phi_0 > 0, \| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \|_\infty \leq \Phi_0 D_m. \]

In fact, it follows from Birgé and Massart (1997) that both properties above are equivalent. For instance \( \Phi_0 = \sqrt{2} \) for collection \([T]\) and \( \Phi_0 = \sqrt{2r + 1} \) for collection \([P]\).

Our description of the bases enhances their similarity, but they must be distinguished in the proofs: collection \([T]\) is different from collections \([P]\) and \([W]\) because the former is built with functions defined globally on the whole interval \([0, 1]\) whereas the latter are built with functions localized on small sub-intervals of \([0, 1]\).

3.2.2. B-Splines. The splines projection space leads to the same kind of results, but we choose to describe it separately because its generating basis is not orthonormal, even if it has also some very similar properties as compared to the bases already described. More specifically, we consider dyadic B-splines on the unit interval \([0, 1]\). Let \( N_r \) be the B-spline of order \( r \) which has knots at the points \( 0, 1, \ldots, r \), i.e. \( N_r(x) = r[0, 1, \ldots, r](1 - x)^{r-1} \) with the usual difference notation (see de Boor (1978), or DeVore and Lorentz (1993)). Let \( m \) be a positive integer and define, for \( D_m = 2^m + r - 1 \),

\[ B_{m,k}(x) = N_r(2^m x - k), \quad \tilde{B}_{m,k} = \sqrt{D_m} B_{m,k}, \quad k \in \mathbb{Z}. \]

For approximation on \([0, 1]\), we only consider the B-splines \( B_{m,k} \) which do not vanish identically on \([0, 1]\). Let \( \mathcal{K}_m \) denote the set of integers \( k \) for which this holds and let \( S_m \) be the linear span of the B-splines \( (B_{m,k}) \) (or \( \tilde{B}_{m,k} \)) for \( k \in \mathcal{K}_m \). The linear space \( S_m \) is referred to as the space of dyadic splines, its dimension is \( D_m \) and any element of \( S_m \) can be represented as \( \sum_{k \in \mathcal{K}_m} a_{m,k} B_{m,k} \) for a \( D_m \)-dimensional vector \( a_m = (a_{m,k})_{k \in \mathcal{K}_m} \). The following properties of the splines are useful and illustrate their similarity with the previously described bases:

(S1) \( \forall x \in \mathbb{R}, \ 0 \leq B_{m,k} \leq \sqrt{D_m} \),

(S2) \( \forall x \in \mathbb{R}, \ \sum_{k \in \mathbb{Z}} \tilde{B}_{m,k}(x) = \sqrt{D_m} \).

(S3) \( \tilde{B}_{m,k} \) has only non zero values on \([k/2^m, (k + r)/2^m]\),

(S4) \( \int \tilde{B}_{m,k} = \sqrt{D_m} 2^{-m} \),

(S5) There exists some constant \( \Phi_0 \) such that

\[ \Phi_0^{-2} \sum_{k \in \mathcal{K}_m} a_{m,k}^2 \leq \left( \sum_{k \in \mathcal{K}_m} a_{m,k} \tilde{B}_{m,k} \right)^2 \leq \Phi_0^2 \sum_{k \in \mathcal{K}_m} a_{m,k}^2, \]

(S6) For any \( t \in S_m \), \( \| t \|_{\infty} \leq \Phi_0 \sqrt{D_m} \| t \| \).

Let us denote \( \hat{a}_m = (\hat{a}_{m,k})_{k \in \mathcal{K}_m} \) the vector of the coefficients of \( \hat{h}_m \) in the spline basis. As already mentioned, the explicit formula given for the coefficients in formula (3.2) is no longer true with spline estimation. Indeed the inversion of a matrix (which is nevertheless block-diagonal with blocks of size \( r \) from (S3)) is
required to derive the expression of the coefficients $\hat{a}_{m,k}$'s. Consider $G_m(B) = (\langle \tilde{B}_{m,k}, \tilde{B}_{m,\ell} \rangle)_{k,\ell \in \mathbb{K}_m}$ and $\Delta_m(B) = \text{diag}(\|\tilde{B}_{m,k}\|^2)_{k \in \mathbb{K}_m}$ two $D_m \times D_m$ matrices and $\Xi_{n,m}(B, H) = (2 \int_0^1 \tilde{B}_{m,k}(x)dH_n(x))_{k \in \mathbb{K}_m}$ a $D_m \times 1$ vector, it is straightforward from (3.1) that:

$$(\Delta_m(B) + G_m(B))\hat{a}_m = \Xi_{n,m}(B, H).$$

In the following, we denote by $[B]$ the collection of models associated with the splines spaces $S_m$ for $m \geq 0$.

3.3. Decomposition of the variance term. By using a result of Lo and al. (1989) and also Reid (1981) and Delecroix and Yazourh (1992), we prove a decomposition result which is the main technical key of the proofs. This decomposition is described by the following Lemma:

**Lemma 3.1.** Consider collections $[T]$, $[P]$, $[W]$ and $[B]$. Then $\forall m \in \mathcal{M}_n$, $\forall t \in S_m$, $\forall a \in [0, 1]$,

$$(3.8) \quad C(t, a) = \frac{1}{n} \sum_{i=1}^n Z_i(t, a) + R_n^*(t, a)$$

where

$$(3.9) \quad Z_i(t, a) = \int_0^{X_i \wedge a} t(u) \frac{h(u)}{1 - F(u)} du - 1_{\{\delta_i = 1\} \cap \{X_i \leq a\}} \frac{t(X_i)}{1 - F(X_i)},$$

and

$$(3.10) \quad \mathbb{E} \sup_{t \in S_m, \|t\| = 1} [R_n^*(t, a)]^{2k} \leq \kappa(k, \Phi_0, a) \frac{D_m^{2k} \ln 2^{k}(n)}{n^{2k}} \text{ for any integer } k,$$

where $\kappa(k, \Phi_0, a)$ denotes a constant depending on $k, a$ and $\Phi_0$.

**Remark 3.1.** More generally, Lemma 3.1 holds true for any finite dimensional space generated by an orthonormal basis satisfying (3.7) and $\sup_{t \in S_m} N(t) \leq K_0 D_m$ if $t$ is continuous and differentiable on $[0, 1]$ or $\sum_{\lambda \in \Lambda_m} N^2(\varphi_\lambda) \leq K_0^2 D_m^2$ if the $\varphi_\lambda$'s are continuous and differentiable except on the boundary of their support with $N(t) = 2\|t\|_{\infty} + \int_0^1 |t'(x)| \, dx$.

Note that it follows from the definition of $R_n^*(t, a) = C(t, a) - n^{-1} \sum_{i=1}^n Z_i(t, a)$ that since $t \mapsto C(t, a)$ and $t \mapsto Z_i(t, a)$ are linear with respect to $t$, then $R_n^*(t, a)$ has the same linearity property. Then

$$\sup_{t \in S_m, \|t\| = 1} [R_n^*(t, a)]^2 \leq \sup_{\sum a^2 \leq 1} \sum_{\lambda \in \Lambda_m} a_{\lambda} R_n^*(\varphi_\lambda, a) \leq \sum_{\lambda \in \Lambda_m} R_n^{*2}(\varphi_\lambda, a)$$

and that the supremum is attained for $a_{\lambda} = R_n^*(\varphi_\lambda, a) / \sqrt{\sum_{\lambda} R_n^{*2}(\varphi_\lambda, a)}$. Therefore

$$\sup_{t \in S_m, \|t\| = 1} [R_n^*(t, a)]^2 = \sum_{\lambda \in \Lambda_m} R_n^{*2}(\varphi_\lambda, a).$$
In the sequel, the abbreviated notation $Z_i(t, 1) = Z_i(t)$ will be used (respectively $R_n^*(t) = R_n^*(t, 1)$).

3.4. Convergence results. The key point is that Lemma 3.1 gives a decomposition of $C(\varphi_\lambda, 1)$ into an empirical mean of i.i.d. centered random variables and a negligible term. Then we can easily prove:

**Proposition 3.1.** Consider the model described in section 2 and the estimator $\hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t)$ where $\gamma_n(t)$ is defined by (3.1) and $S_m$ is a $D_m$-dimensional linear space in collection [T], [P], [W] or [B]. Then

\[
E(\|h - \hat{h}_m\|^2) \leq C_1 \|h - h_m\|^2 + C_2 \frac{2\Phi_0^2 D_m}{n} \int_0^1 \frac{h(x)}{1 - F(x)} dx + \frac{D_m^2 \ln^2(n)}{n^2},
\]

where $\kappa$ is a constant depending on the basis, $C_1 = C_2 = 1$ for collection [T], [P] or [W] and $C_1 = 3$, $C_2 = 8$ for collection [B].

**Proof.** Applying (3.8) in decomposition (3.5) for an orthonormal basis leads to

\[
\|h - \hat{h}_m\|^2 = \|h - h_m\|^2 + \sum_{\lambda \in \Lambda_m} \left( \frac{1}{n} \sum_{i=1}^n Z_i(\varphi_\lambda) + R_n^*(\varphi_\lambda) \right)^2
\]

\[
\leq \|h - h_m\|^2 + 2 \sum_{\lambda \in \Lambda_m} \left( \frac{1}{n} \sum_{i=1}^n Z_i(\varphi_\lambda) \right)^2 + 2 \sum_{\lambda \in \Lambda_m} R_n^{*2}(\varphi_\lambda).
\]

Then, by using that the variables $Z_i(\varphi_\lambda)$ are i.i.d. centered with variance

\[
\int_0^1 \varphi_\lambda^2(x) \frac{h(x)}{1 - F(x)} dx
\]

and the bound (3.10) for [T], [P] and [W], we have

\[
E(\|h - \hat{h}_m\|^2) \leq \|h - h_m\|^2 + \frac{2}{n} \sum_{\lambda \in \Lambda_m} \int_0^1 \varphi_\lambda^2(x) \frac{h(x)}{1 - F(x)} dx
\]

\[
+ 2E \left( \sup_{t \in S_m, \|t\|=1} [R_n^*(t)] \right)
\]

\[
\leq \|h - h_m\|^2 + \frac{2\Phi_0^2 D_m}{n} \int_0^1 \frac{h(x)}{1 - F(x)} dx + 2K(1, \Phi_0, 1)D_m^2 \ln^2(n) n^2.
\]

For collection [B], the proof requires specific but straightforward properties of the spline basis and is deferred to the last section. \qed

Inequality (3.11) gives the asymptotic rate for one estimator, provided that we know that for a function $h$ with regularity $\alpha$, $\|h - h_m\|^2 \leq KD_m^{-2\alpha}$ and provided that $D_m$ is well chosen in function of $n$. More precisely, if $D_m$ is of order $n^{1/(2\alpha+1)}$, the resulting rate is of order $n^{-2\alpha/(2\alpha+1)}$. 
More formally, let us recall that a function \( f \) belongs to the Besov space \( B_{\alpha,\ell}\infty([0,1]) \) if it satisfies
\[
|f|_{\alpha,\ell} = \sup_{y>0} y^{-\alpha} w_d(f,y)\ell < +\infty, \quad d = [\alpha] + 1,
\]
where \( w_d(f,y)\ell \) denotes the modulus of smoothness. For a precise definition of those notions we refer to DeVore and Lorentz (1993) Chapter 2, Section 7, where it is also proved that \( B_{\alpha,p,\infty}([0,1]) \subset B_{\alpha,2,\ell}\infty([0,1]) \) for \( p \geq 2 \). This justifies that we now restrict our attention to \( B_{\alpha,2,\ell}\infty([0,1]) \).

Then the following (standard) rate is obtained:

**Corollary 3.1.** Consider the model described in section 2 and the estimator \( \hat{h}_m = \arg\min_{t \in S_m} \gamma_n(t) \) where \( \gamma_n(t) \) is defined by (3.1) and \( S_m \) is a \( D_m \)-dimensional linear space in collection [T], [P], [W] or [B]. Assume moreover that \( h \in B_{\alpha,2,\ell}\infty([0,1]) \) with \( \alpha > 0 \) and choose a model with \( m = m_n \) such that \( D_{m_n} = O(n^{1/(2\alpha+1)}) \), then
\[
E(\|h - \hat{h}_{m_n}\|^2) = O\left(n^{-\frac{2\alpha}{2\alpha+1}}\right).
\]

**Remark 3.2.** The bound \( \alpha \) stands for the regularity of the basis functions for collections [P], [W] and [B]. For the trigonometric collection [T], no upper bound for the unknown regularity \( \alpha \) is required.

**Proof.** The result is a straightforward consequence of the results of DeVore and Lorentz (1993) and of Lemma 12 of Barron, Birgé and Massart (1999), which imply that \( \|h - h_{m}\| \) is of order \( D_{m}^{-\alpha} \) in the three collections [T], [P] and [W], for any positive \( \alpha \). In the same way, it follows from Theorem 3.3 in Chapter 12 of DeVore and Lorentz (1993) that \( \|h - h_{m}\|^2 = O(D_{m}^{-2\alpha}) \) in collection [B], if \( h \) belongs to some Besov space \( B_{\alpha,2,\ell}\infty([0,1]) \) with \( |h|_{\alpha,2} \leq L \) for some fixed \( L \). Thus the minimum order in (3.11) is reached for a model \( S_{m_n} \) with \( D_{m_n} = O\left([n^{1/(1+2\alpha)}]\right) \), which is less than \( n \) for \( \alpha > 0 \). Then, if \( h \in B_{\alpha,2,\ell}\infty([0,1]) \) for some \( \alpha > 0 \), we find the standard nonparametric rate of convergence \( n^{-2\alpha/(1+2\alpha)} \). \( \square \)

Lastly, let us mention that we can obtain a punctual central limit theorem, but for the histogram basis only. This result is analogous to those given for wavelet bases in Antoniadis et al. (1999) or Wu and Wells (2003) who need to choose some “dyadic points” in a very particular way for providing their central limit theorem to hold. The particular feature of the histogram basis avoids this problem. More precisely we have the following result:

**Theorem 3.1.** Consider \( \hat{h}_m = \arg\min_{t \in S_m} \gamma_n(t) \) where \( \gamma_n(t) \) is defined by (3.1) and \( S_m \) is a \( D_m \)-dimensional linear space in collection [P] with \( r = 0 \). Assume moreover that \( h^{(1)} \) is bounded, then for \( D_m = D_{m_n} \) such that \( D_{m_n} = o(\sqrt{n}) \) and \( n/D_{m_n}^3 = o(1) \),
we have
\[
\sqrt{\frac{n}{D_m}} \left( \hat{h}_{m_n}(x) - h(x) \right) \overset{\mathcal{L}}{\rightarrow} n \to \infty \mathcal{N} \left( 0, \frac{h(x)}{1 - F(x)} \right).
\]

4. The adaptive estimator

4.1. Adaptation with a theoretical penalty. The penalized estimator is defined in order to ensure an automatic choice of the dimension. Indeed, it follows from Corollary 3.1 that the optimal dimension depends on the unknown regularity \( \alpha \) of the function to be estimated. Then we define
\[
\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left[ \gamma_n(\hat{h}_m) + \text{pen}(m) \right]
\]
where the penalty function \( \text{pen} \) is determined in order to lead to the choice of a “good” model. First, by applying some Talagrand’s (1996) type inequality to the empirical process defined by
\[
\nu_n(g_t) = \frac{1}{n} \sum_{i=1}^{n} Z_i(t) := \frac{1}{n} \sum_{i=1}^{n} g_t(X_i)
\]
where
\[
g_t(x) = \int_{0}^{x \wedge 1} \frac{h(u)}{1 - F(u)} du - I_{\{x = 1\}} \wedge \{x \leq 1\} \frac{t(x)}{1 - F(x)},
\]
so that \( g_t(X_i) = Z_i(t) \), we can prove the following lemma:

**Lemma 4.1.** Let \( \nu_n(g_t) \) be defined by (4.1) and (4.2), then for \( \epsilon > 0 \)
\[
\mathbb{E} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) - p(m,m') \right) \leq \frac{\kappa_1}{n} \left( e^{-\kappa_2 \epsilon (D_m + D_{m'})} + e^{-\kappa_3 \epsilon^3/\sqrt{n}} \right),
\]
with \( p(m,m') = 2(1 + 2\epsilon) C_3(D_m + D_{m'})/n \), \( C_3 = (\sqrt{1 + \epsilon - 1}) \wedge 1 \) and where \( B_{m,m'}(0,1) = \{ t \in S_m + S_{m'} / \| t \| \leq 1 \} \). The constants \( \kappa_i \) for \( i = 1,2,3 \) and \( C_3 \) depend on \( \Phi_0, h \) and \( F \).

Then, by using the decomposition of the contrast given by
\[
\gamma_n(t) - \gamma_n(s) = \| t - h \|^2 - \| s - h \|^2 - \frac{2}{n} \sum_{i=1}^{n} Z_i(t - s) - 2 R_n^*(t - s),
\]
we easily derive the following result:

**Theorem 4.1.** Consider the model described in section 2 and the estimator \( \hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t) \) where \( \gamma_n(t) \) is defined by (3.1) and \( S_m \) is a \( D_m \)-dimensional linear space in collection \( \mathcal{T}, \mathcal{DP}, \mathcal{B} \) or \( \mathcal{W} \), \( |\mathcal{M}_n| \leq n \) and \( D_m \leq \sqrt{n}, \forall m \in \mathcal{M}_n \). Then the estimator \( \hat{h}_{\hat{m}} \) with \( \hat{m} \) defined by
\[
\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left[ \gamma_n(\hat{h}_m) + \text{pen}(m) \right]
\]
and

$$\text{pen}(m) = \kappa \Phi_0^2 \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m}{n}$$

where $\kappa$ is a universal constant satisfies

$$\mathbb{E}(\|\hat{h}_m - h\|^2) \leq \inf_{m \in \mathcal{M}_n} (3\|h - h_m\|^2 + 4\text{pen}(m)) + \frac{K \ln^2(n)}{n},$$

where $K$ is a constant depending on $h$, $F$ and $\Phi_0$.

Therefore, the adaptive estimator automatically makes the squared-bias/variance compromise and from an asymptotic point of view, reaches the optimal rate, provided that the constant in the penalty is known. Note that Inequality (4.5) is nevertheless non-asymptotic.

**Remark 4.1.** Note that again, the result holds true for any collection of models $(S_m)_{m \in \mathcal{M}_n}$ such that $S_m$ is a linear subspace of $L_2([0, 1])$ with dimension $D_m$ satisfying condition (3.7), for which (3.10) holds and $|\mathcal{M}_n| \leq n$, $D_m \leq \sqrt{n}$, $\forall m \in \mathcal{M}_n$.

**Proof.** It follows from the definition of $\hat{h}_m$ that:

$$\gamma_n(\hat{h}_m + \text{pen}(\hat{m})) \leq \gamma_n(h_m) + \text{pen}(m).$$

Then by using decomposition (4.4), it follows from (4.6) and from the definition of the process $\nu_n(g_t)$ given in lemma 4.1 that:

$$\|\hat{h}_m - h\|^2 \leq \|h_m - h\|^2 + \frac{2}{n} \sum_{i=1}^{n} Z_i(\hat{h}_m - h_m) + 2R^*_n(\hat{h}_m - h_m) + \text{pen}(m) - \text{pen}(\hat{m})$$

$$\leq \|h_m - h\|^2 + \frac{1}{4}\|\hat{h}_m - h_m\|^2 + 8 \sup_{t \in B_{m,n}(0,1)} \nu^2_n(g_t) + 8 \sup_{t \in B_{m,n}(0,1)} R^2_n(t) + \text{pen}(m) - \text{pen}(\hat{m}).$$

(4.7)

Note that the norm connection as described by (3.7) still holds for elements $t$ of $S_{m} + S_{m'}$ as follows $\|t\|_{\infty} \leq \Phi_0 \sup(D_m, D_{m'})\|t\|$ since we restricted our attention to nested collections of models. We denote by $D(m')$ the dimension of $S_{m} + S_{m'}$ for the fixed $m$ considered in the following. Note that $D(m') = \sup(D_m, D_{m'}) \leq D_m + D_{m'}$.

Let $p(m, m')$ such that

$$8p(m, m') \leq \text{pen}(m) + \text{pen}(m')$$

for all $m, m'$ in $\mathcal{M}_n$.

Then $\forall m \in \mathcal{M}_n$, 

(4.8)
\[
\frac{1}{2} \| \hat{h}_m - h \|^2 \leq \frac{3}{2} \| h - h_m \|^2 + 2 \text{pen}(m) + 8 \left( \sup_{t \in B_{m,m}(0,1)} \nu_n^2(g_t) - p(m, \hat{m}) \right) + 8 \sup_{t \in B_{m,m}(0,1)} R_n^2(t).
\]

Then if we prove
\[
(4.9) \quad \mathbb{E} \left( \sup_{t \in B_{m,\hat{m}}(0,1)} \nu_n^2(g_t) - p(m, \hat{m}) \right) \leq C \frac{\ln^2(n)}{n},
\]
and
\[
(4.10) \quad \mathbb{E} \left( \sup_{t \in B_{m,\hat{m}}(0,1)} R_n^2(t) \right) \leq C' \frac{\ln^2(n)}{n},
\]
we have the following result: \( \forall m \in \mathcal{M}_n, \)
\[
\mathbb{E}(\| \hat{h}_m - h \|^2) \leq 3\| h - h_m \|^2 + 4\text{pen}(m) + \frac{K\ln^2(n)}{n},
\]
and this proves the theorem.

Therefore by using equation (4.8) and the definition of \( p(m, m') \) in Lemma 4.1, we choose
\[
\text{pen}(m) = 16(1 + 2\epsilon) \int_0^1 \frac{h(x)}{1 - F(x)} \frac{D_m}{n} dx.
\]

Inequality (4.9) is a straightforward consequence of Lemma 4.1 since
\[
\sum_{m' \in \mathcal{M}_n} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) - p(m, m') \right) \leq \kappa_1 \left( \sum_{m' \in \mathcal{M}_n} e^{-\kappa_2 D_{m'}} \frac{D_m}{n} + \frac{\|\mathcal{M}_n\|}{n} e^{-\kappa_3 \sqrt{n}} \right).
\]

Then by taking \( \epsilon = 1/2 \) and assuming that \( |\mathcal{M}_n| \leq n \) and since for [T], [DP], [B] and [W], \( \sum_{m \in \mathcal{M}_n} e^{-a D_m} \leq \sum_{k=1}^n e^{-ka} \leq \Sigma(a) < +\infty, \forall a > 0, \) this leads to the bound
\[
\sum_{m' \in \mathcal{M}_n} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) - p(m, m') \right) \leq \frac{C}{n},
\]
and this ensures (4.9).

Lastly, let us verify that inequality (4.10) holds for all collections of model. For [T] and [B], we have
\[
\sup_{t \in B_{m,\hat{m}}(0,1)} (R_n^*(t))^2 \leq \sup_{t \in B_{m,\hat{m}}(0,1)} N(t) R_n^2 \leq C(D_m + D_{\hat{m}})^2 R_n^2.
\]
On the other hand, for [DP] and [W], we have already seen (in the proof of Lemma 3.1, see Inequality (6.6)) that
\[
\sup_{t \in B_m, \hat{m}} (R_n^*(t))^2 = \sum_{\lambda \in \Lambda_m, \hat{m}} R_n^2(\varphi_\lambda) \leq C'(D_m + \hat{D}_m)^2 R_n^2
\]
with the natural notation \(\Lambda_{m, \hat{m}} = \Lambda_m \cup \Lambda_{\hat{m}}\). In addition \(E(R_n^2) \leq \ln^2(n)/n^2\), then it is enough to conclude that for all collections [T], [B], [DP] and [W]
\[
E\left(\sup_{t \in B_m, \hat{m}} (R_n^*(t))^2\right) \leq K \frac{\ln(n)^2}{n} \quad \text{as soon as} \quad D_m \leq \sqrt{n}, \forall m \in \mathcal{M}_n.
\]

4.2. Random penalization. The penalty given in Theorem 4.1 can not be used in practice since it depends on the unknown quantity
\[
\int_0^1 h(x) \frac{1}{1 - F(x)} dx.
\]
Therefore we replace this quantity by an estimator and prove that the estimator of \(h\) built with this random penalty keeps the adaptation property of the theoretical estimator.

**Theorem 4.2.** Consider the model described in section 2 and the estimator \(\hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t)\) where \(\gamma_n(t)\) is defined by (3.1) and \(S_m\) is a \(D_m\)-dimensional linear space in collection [T], [DP], [B] or [W], \(|\mathcal{M}_n| \leq n\) and \(D_m \leq \sqrt{n}, \forall m \in \mathcal{M}_n\). Consider the estimator \(\hat{\hat{m}}\) with \(\hat{m}\) defined by
\[
\hat{\hat{m}} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{h}_m) + \tilde{\text{pen}}(m)]
\]
and
\[
\tilde{\text{pen}}(m) = \kappa \Phi_0^2 \left( \int_0^1 \frac{\hat{h}_n(x)}{1 - F_n(x)} dx \right) \frac{D_m}{n}
\]
where \(\kappa\) is a universal constant, \(\hat{h}_n = \hat{h}_{m_n}\) where \(m_n \in \mathcal{M}_n\) is one estimator in the collection and \(F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}\) stands for the standard empirical distribution function of the \(X_i\)’s. Then if the orthogonal projection \(h_{m_n}\) of \(h\) on \(S_{m_n}\), satisfies
\[
\|h_{m_n} - h\| \leq \frac{1}{16} (1 - F(1)) \int_0^1 \frac{h(x)}{1 - F(x)} dx
\]
then \(\hat{h}_{\hat{m}}\) satisfies
\[
\mathbb{E}(\|\hat{h}_{\hat{m}} - h\|^2) \leq \inf_{m \in \mathcal{M}_n} K_0 \left( \|h - h_m\|^2 + \Phi_0^2 \int_0^1 \frac{h(x)}{1 - F(x)} dx \frac{D_m}{n} \right) + K \ln^2(n)/n,
\]
where \(K_0\) is a universal constant and \(K\) depends on \(h, F, \Phi_0\).
We can see that the constraint given in (4.11) is not very strong since as already mentioned with Corollary 3.1, we know that for collection \([T],[P]\) or \([W]\) and \(h\) belonging to some Besov space \(B_{\alpha,2,\infty}([0,1])\), \(||h - h_{m_n}||^2 = O(D_{m_n}^{2\alpha})\) and therefore \(||h - h_{m_n}||^2\) tends to zero when \(D_{m_n}\) tends to infinity. This condition implies that \(D_{m_n}\) is not only bounded by above but also by below and must be great enough, as it is natural.

Note that another substitution can be done since \(\int_0^1 h(x)/(1 - F(x))\) can also be seen as the second order moment of the independent random variables \(I_{\{\delta_i=1,X_i\leq 1\}}/(1 - \hat{F}_n(X_i))\) and therefore can be estimated by

\[
\hat{s}_2 = \frac{1}{n} \sum_{i=1}^n \frac{I_{\{X_i\leq 1\}} I_{\{\delta_i=1\}}}{(1 - \hat{F}_n(X_i))^2}, \quad \hat{F}_n(x) = \frac{1}{n+1} \sum_{i=1}^n I_{\{X_i \leq x\}}.
\]

The following result holds:

**Corollary 4.1.** Under the assumptions of Theorem 4.2 but with

\[
\widehat{\text{pen}}(m) = \kappa \Phi_0^2 \hat{s}_2 \frac{D_m}{n}
\]

where \(\hat{s}_2\) is defined by (4.13) and \(\kappa\) is a universal constant, the estimator \(\hat{h}_{m_n}\) satisfies (4.12).

In particular, we can derive quite straightforwardly from results as Theorem 4.2 some adaptation results to unknown smoothness:

**Proposition 4.1.** Consider the collection of models \([T],[P]\) or \([W]\), with \(r > \alpha > 1/2\). Assume that an estimator \(\hat{h}\) of \(h\) satisfies inequality (4.12). Let \(L > 0\). Then

\[
\left( \sup_{h \in B_{\alpha,2,\infty}(L)} \mathbb{E}\|h - \hat{h}\|^2 \right)^{\frac{1}{2}} \leq C(\alpha, L) n^{-\frac{\alpha}{2\alpha + 1}}
\]

where \(B_{\alpha,2,\infty}(L) = \{t \in B_{\alpha,2,\infty}([0,1]), |t|_{\alpha,2} \leq L\}\) where \(C(\alpha, L)\) is a constant depending on \(\alpha, L\) and also on \(h, \Phi_0\).

### 4.3. Adaptive estimation with a general collection of models.

It is sometimes useful for estimation purpose to use non regular models, in order to take into account some variability in the regularity of the function to be estimated. This extension is possible under some restrictions. For instance Reynaud-Bouret (2003) proves some adaptation results for non regular histograms provided that knots are chosen in a set of cardinality less than \(n/\ln^2(n)\). Our restriction here is stronger since the largest cardinality we can consider is of order \(\sqrt{n}\). On the other hand, we consider the more general collection of piecewise polynomials, than histograms.

We assume in the following for the sake of simplicity (and without loss of generality), that \(\sqrt{n}\) is an integer. We consider the set of knots
\[ \Gamma = \{ \frac{\ell}{\sqrt{n}}, \ \ell = 1, \ldots, \sqrt{n} - 1 \}. \]

A general piecewise polynomial model, non necessarily regular, is then defined by the maximal degree \( r \) of the polynomials and a set of knots
\[ \{a_0 = 0, a_1, \ldots, a_\ell, a_{\ell+1} = 1\} \]
where \( \{a_1, \ldots a_\ell\} \) is any subset of \( \Gamma \) : its dimension is \( D_m = (\ell+1)(r+1) \). This means that with non regular collection for any fixed dimension \( D_m \), there is \( (\frac{\sqrt{n} - 1}{\ell}) \) associated models corresponding to the possible choices of the subset \( \{a_1, \ldots a_\ell\} \) with \( \ell = 1, \ldots, \sqrt{n} - 1 \). Therefore, the cardinality of the set \( \mathcal{M}_n \) of all possible \( m \) is:
\[
\sum_{\ell=1}^{\sqrt{n}-1} \left( \frac{\sqrt{n} - 1}{\ell} \right) = 2^{\sqrt{n} - 1} - 1 = \frac{1}{2} \exp(\sqrt{n} \ln(2)) - 1.
\]
It is exponentially great and in particular greater than the order \( n \) obtained in the regular case, when we consider only one model per dimension. The \( \varphi_\lambda \)'s for \( \lambda = (a_j, a_{j+1}; k) \in \Lambda_m \) are given by
\[
\sqrt{\frac{2k+1}{a_j - a_{j-1}}} Q_k \left( \frac{2}{a_j - a_{j-1}} x - \frac{a_j + a_{j-1}}{a_j - a_{j-1}} \right) 1_{[a_j, a_{j-1}]}(x)
\]
for \( k = 0, 1, \ldots, r \) and \( j = 0, \ldots, \ell + 1 \), where \( Q_k \) denotes the \( k \)th Legendre polynomial.

The main difficulty is that the connection between the supremum norm \( \|t\|_\infty \) and the \( L^2 \)-norm, \( \|t\| \) as given by (3.7) and its equivalent formulation is no longer true. We only have
\[
\forall m \in \mathcal{M}_n, \forall t \in S_m, \|t\|_\infty \leq \sqrt{(2r+1)\sqrt{n}}\|t\|
\]
and this inequality is clearly less powerful than (3.7) since we assumed that \( \forall m \in \mathcal{M}_n, D_m \leq \sqrt{n} \). In particular, the result in Lemma 3.1 must be re-examined in this light, as well as most bounds. We can prove the following result:

**Theorem 4.3.** Consider the model described in section 2 and the estimator \( \hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t) \) where \( \gamma_n(t) \) is defined by (3.1) and \( S_m \) is a \( D_m \)-dimensional linear space in the general collection of piecewise polynomials described above. Then the estimator \( \hat{h}_m \) with \( \hat{m} \) defined by
\[
\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{h}_m) + \text{pen}(m)]
\]
and
\[
\text{pen}(m) = \kappa (2r+1) \sup_{x \in [0,1]} \left( \frac{h(x)}{1 - F(x)} \right) \frac{D_m(1 + \ln^2(n))}{n}
\]
where $\kappa$ is a universal constant satisfies
\begin{equation}
E(\|\hat{h}_m - h\|^2) \leq \inf_{m \in M_n} (3\|h - h_m\|^2 + 4\text{pen}(m)) + \frac{K \ln^2(n)}{n},
\end{equation}
where $K$ is a constant depending on $h$, $F$ and $r$.

Two remarks about the penalty are in order. First, the $\ln^2(n)$ factor implies a $\ln^2(n)$ factor loss in the resulting rates. This is the price to pay for considering such a huge collection of models. Second, the unknown term in the penalty is now $\sup_{x \in [0,1]} [\hat{h}_n(x)/(1 - \hat{F}_n(x))]$ where $\hat{h}_n$ is a given estimator of $h$ on a well chosen regular space of piecewise polynomials and $\hat{F}_n$ is the empirical distribution of the data as defined above, would suit (but may lead to over penalization).

5. Simulations

In this section, we present the results of a short simulation experiment that aims to compare the performances of our estimator with two other ones that have been studied in the literature: Antoniadis et al. (1999)’s and Reynaud-Bouret (2003)’s. More precisely, Antoniadis et al. (1999) study a wavelet estimator with selection of the coefficients to keep by cross-validation. They present some simulation results in two cases:

- The first set of simulations is called in the following the “Gamma case”. In this case, the $X^0_i$’s are generated from a Gamma distribution with shape parameter 5 and scale 1 and the independent $C_i$’s from an exponential distribution with mean 6.
- The second set is called “the bimodal case”. In this case, the $X^0_i$’s have a bimodal density defined by
  \[ f^0 = 0.8u + 0.2v \]
where $u$ is the density of $\exp(Z/2)$ with $Z \sim \mathcal{N}(0,1)$ and $v = 0.17Z + 2$.

The $C_i$’s are generated from an exponential distribution with mean 2.5.

The authors give the mean squared errors of their estimator computed over $T = 200$ replications of samples of size $n = 200$ and $n = 500$. The error is computed over $K$ regularly spaced points $t_k$, $k = 1, \ldots, K$, of the interval in which the $X_i$’s fall ($[0, \max X_i]$), in the following standard way: the mean over the replications $t$ of
\begin{equation}
\text{MSE}_t = \frac{1}{K} \sum_{k=1}^{K} (h(t_k) - \hat{h}_t(t_k))^2
\end{equation}
where $\hat{h}_t$ is the estimate of $h$ for the sample number $t$, $t = 1, \ldots, T$.

In order to take into account the sparsity of the observations at the end of the interval, ($\mathbb{P}(X^0 > 6) = 0.25$ in the Gamma case and $\mathbb{P}(X^0 > 2) = 0.16$ in the bimodal
case), they also compute an error $\text{MSE}_2$ defined by the same kind of mean squared error but only for the $t_k$’s less than 6 in the Gamma case and 2 in the bimodal case. For $K = 64$, they obtain the following result:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Gamma</th>
<th>Bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>MSE</td>
<td>0.112</td>
<td>0.0995</td>
</tr>
<tr>
<td>$\text{MSE}_2$</td>
<td>0.0025</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Table 1. Results of Antoniadis et al. (1999, Table 2), $T = 200$ replications.

Reynaud-Bouret (2003) studies an histogram adaptive estimator based on a contrast different from ours and for the same kind of data, when using what she calls a “regular histogram strategy”. She finds the results recalled in Table 2.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Gamma</th>
<th>Bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0333</td>
<td>0.0376</td>
</tr>
<tr>
<td>$\text{MSE}_2$</td>
<td>0.0086</td>
<td>0.0048</td>
</tr>
</tbody>
</table>

Table 2. Results of Reynaud-Bouret (2003, Table 10), $T = 200$ replications.

Note that Reynaud-Bouret (2003) studies also some other strategies that may be studied in our case also, but this is beyond the scope of the present paper. Analogously, some more ambitious simulation study may lead to a better and more complicated choice of the penalty function (see Comte and Rozenholc (2001) for a study of this type for penalized regression and volatility functions estimation). Here we considered a penalized estimator with penalty

$$\hat{\text{pen}}(m) = \hat{\text{pen}}(D, r) = \kappa \hat{s}_2 \frac{D(r + 1)}{n},$$

and $\kappa = 1, 1.5, 2, 2.5, 3$. A dimension $\hat{D}$ is selected by contrast penalization among the dyadic set of values $2^0, 2^1, 2^2, 2^3, 2^4$ and we compare the results for $r = 0, 1, 2$. Here again a more ambitious simulation study may select empirically the best degree (see some strategies in order to do so in Comte and Rozenholc (2001)).
Table 3. Results of our penalized estimator, $T = 200$ replications, $\kappa = 2.5$.

We can see that in both cases the best choice seems to be $r = 2$. Globally, our results are of the same order as the one obtained with the other methods, and it is most likely that a precise calibration of the penalty function may lead to globally better results.

Table 4. Best results when selecting the degree (among $r = 0, 1, 2$) and the constant $\kappa$ in the penalty (among $\kappa = 1, 1.5, 2, 2.5, 3$), $T = 200$ replications.

This is suggested by the results gathered in Table 4 that gives the best performance that is obtained when we select a posteriori both the degree and the constant in the penalty, for each set of simulations. The first line gives the error and the second in parenthesis the corresponding degree and value of $\kappa$. Under such unfair conditions, we obtain results that are almost always (except for the MSE2 corresponding to the bimodal sample) better results than our competitors.

6. PROOFS

6.1. Proof of Lemma 3.1. In all the sequel, we consider that $0 \leq a \leq 1$. For $t \equiv 1$, the decomposition (3.8) above follows from the decomposition of $H_n(x) - H(x)$ given in Lemma A.1. of Lo et al. (1989). It can be written

$$H_n(x) - H(x) = n^{-1} \sum_{i=1}^{n} \zeta_i(x) + r_n(x)$$
where

\[
\zeta_i(x) = \int_0^{x_i \land x} \frac{h(u)}{1 - F(u)} \, du - \mathbf{1}_{\{\delta_i = 1\} \cap \{X_i \leq x\}} \frac{1}{1 - F(X_i)} = Z_i(1, x)
\]

and

\[
\mathbb{P}\left( \sup_{0 \leq x \leq 1} |r_n(x)| > \theta \ln(n)/n \right) = O(n^{-\beta})
\]

for any \( \beta > 0 \) and some constant \( \theta \) depending on \( \beta \). Note that the \( \zeta_i \)'s are uniformly bounded on \([0, T]\), and as we have already mentioned \((n + 1)^{-1} \leq F_n(x) \leq 1\) for all \( x \) in \([0, 1]\), it follows that \( \sup_{0 \leq x \leq 1} |r_n(x)| = O(\ln(n)) \). From this, we easily deduce that, for any \( \beta \geq 1 \),

\[
\mathbb{E}\left( \sup_{0 \leq x \leq 1} |r_n(x)|^\beta \right) = \mathbb{E}\left( \sup_{0 \leq x \leq 1} |r_n(x)|^\beta \mathbf{1}_{\{\sup_{0 \leq x \leq 1} |r_n(x)| > \theta \ln(n)/n\}} \right) \\
+ \mathbb{E}\left( \sup_{0 \leq x \leq 1} |r_n(x)|^\beta \mathbf{1}_{\{\sup_{0 \leq x \leq 1} |r_n(x)| \leq \theta \ln(n)/n\}} \right) \\
\leq \kappa \ln^2(n) \mathbb{P}\left( \sup_{0 \leq x \leq 1} |r_n(x)| > \theta \ln(n)/n \right) + \left( \frac{\theta \ln(n)}{n} \right)^\beta \\
\leq \frac{\kappa \ln^2(n)}{n^\beta} + \left( \frac{\theta \ln(n)}{n} \right)^\beta
\]

so that, since \( \theta = \theta(\beta) \),

\[
\mathbb{E}\left( \sup_{0 \leq x \leq 1} |r_n(x)|^\beta \right) \leq \frac{C_\beta \ln^2(n)}{n^\beta} \text{ for any } \beta \geq 1.
\]

Next the global decomposition (3.8) follows by integration by part, if the function is continuous and differentiable on the considered interval. Indeed in that case, it is shown in Delecroix and Yazourh (1992) that for a function \( t \) continuous and differentiable on \([0, a]\), then

\[
\int_0^a t(x) \, d(H_n - H)(x) = t(a)(H_n - H)(a) + \int_0^a t'(x)(H_n - H)(x) \, dx.
\]

The above decomposition of \( H_n - H \) implies then

\[
\int_0^a t(x) \, d(H_n - H)(x) = t(a) \frac{1}{n} \sum_{i=1}^n \zeta_i(a) - \int_0^a t'(x) \frac{1}{n} \sum_{i=1}^n \zeta_i(x) \, dx + t(a)r_n(a) - \int_0^a t'(x) \, r_n(x) \, dx.
\]

(6.2)

From Delecroix and Yazourh (1992), we know that

\[
t(a) \frac{1}{n} \sum_{i=1}^n \zeta_i(a) - \int_0^a t'(x) \frac{1}{n} \sum_{i=1}^n \zeta_i(x) \, dx = \frac{1}{n} \sum_{i=1}^n Z_i(t, a).
\]

(6.3)
Next setting $R_n^*(t, a) = t(a)r_n(a) - \int_0^a r_n(x)t'(x)dx$ gives, for

$$R_n = \sup_{0 \leq x \leq 1} |r_n(x)|$$

the bound:

$$R_n^*(t, a) \leq R_n \left(|t(a)| + \int_0^a |t'(x)|dx\right).$$

Among the collections we described, the functions in $[T]$ are continuous and differentiable functions. Then consider $t(x) = a_0 + \sum_{j=1}^m a_j \sqrt{2} \cos(2\pi j x) + \sum_{j=1}^m b_j \sqrt{2} \sin(2\pi j x)$, $\|t\|^2 = \int_0^1 t^2(x)dx = a_0^2 + \sum_{j=1}^m a_j^2 + \sum_{j=1}^m b_j^2$, $D_m = 2m + 1$, $\|t\|_\infty \leq \Phi_0 \sqrt{D_m}$ for $\|t\| = 1$ and $\Phi_0 = \sqrt{2}$, and

$$\int_0^a |t'(x)|dx \leq \sqrt{a}\|t'\| = \sqrt{a} \left( \sum_{j=1}^m [(2\pi j)^2(a_j^2 + b_j^2)] \right)^{1/2} \leq \sqrt{a}[(2\pi D_m)\|t\|] = 2\pi \sqrt{a}D_m\|t\|.$$

It follows that Inequality (3.10) is fulfilled for the trigonometric basis with $\kappa(k, \Phi_0, a) = C_{2k} 2^{2k-1}(\Phi_0^{2k} + (2\pi a)^{2k})$.

Analogously, the collection $[B]$ for $r \geq 1$ allows to use the bound (6.4)-(6.5). Next $|t(a)| \leq \Phi_0 \sqrt{D_m}$ for $\|t\| = 1$ with (S1). On the other hand, for $t = \sum_{k \in \mathbb{K}_m} a_{m,k} \tilde{B}_{m,k}$,

$$\int_0^a |t'(x)|dx \leq \sum_{k \in \mathbb{K}_m} |a_{m,k}| \sqrt{D_m} \int_{k/2^m}^{(k+r)/2^m} 2^m |N_r'(2^m x - k)|dx, \text{ with (S3)}$$

$$= \sum_{k \in \mathbb{K}_m} |a_{m,k}| \sqrt{D_m} \int_0^r |N_r'(u)|du \leq C_r D_m \sqrt{\sum_{k \in \mathbb{K}_m} a_{m,k}^2} \leq C_r \Phi_0 D_m \text{ with (S5)},$$

where we set $C_r = \int_0^r |N_r'(u)|du$. Inequality (3.10) is therefore fulfilled for the dyadic splines with $\kappa(k, \Phi_0, a) = 2^{2k} \Phi_0^{2k} C_r^{2k}$.

Now we must consider the case where $t$ is only piecewise continuous and differentiable on the interval $[0, 1]$ with knots at the points $j/m$, $j = 1, \ldots, m$ as are the functions in collection $[P]$, and denote by $\tilde{t}_j$ a continuous and differentiable extension of $tI_{[j-1)/m, j/m]}$ to the whole interval $[0, 1]$. Then

$$\int_{(j-1)/m}^{j/m} t(x)d(H_n - H)(x) = \int_0^{j/m} \tilde{t}_j(x)d(H_n - H)(x) - \int_0^{(j-1)/m} \tilde{t}_j(x)d(H_n - H)(x).$$
Applying the decomposition resulting from (6.2) and (6.3) above to both terms gives

\[ \int_{(j-1)/m}^{j/m} t(x) d(H_n - H)(x) = \frac{1}{n} \sum_{i=1}^{n} [Z_i(t, j/m) - Z_i(t, (j-1)/m)] \]

\[ + t((\frac{j}{m})^{-}) r_n(\frac{j}{m}) - t((\frac{j-1}{m})^{+}) r_n(\frac{j-1}{m}) \]

\[ + \int_{(j-1)/m}^{j/m} r_n(u) t'(u) du. \]

Therefore, by taking \( a = j_0/m \) \((j_0 \leq m)\), for simplicity, gives

\[ \int_{0}^{a} t(x) d(H_n - H)(x) = \sum_{j=1}^{j_0} \int_{(j-1)/m}^{j/m} t(x) d(H_n - H)(x) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} Z_i(t, a) + \int_{0}^{a} r_n(u) t'(u) du \]

\[ + \sum_{j=1}^{j_0} \left[ t\left( \left[ \frac{j}{m} \right]^{-} \right) r_n\left( \frac{j}{m} \right) - t\left( \left[ \frac{j-1}{m} \right]^{+} \right) r_n\left( \frac{j-1}{m} \right) \right]. \]

Therefore the first part of the previous decomposition remains valid but the rest is now

\[ R_n^*(t, a) = \sum_{j=1}^{j_0} \left[ t\left( \left[ \frac{j}{m} \right]^{-} \right) r_n\left( \frac{j}{m} \right) - t\left( \left[ \frac{j-1}{m} \right]^{+} \right) r_n\left( \frac{j-1}{m} \right) \right] + \int_{0}^{a} r_n(u) t'(u) du. \]

Then a bound is valid for functions \( t \) chosen as a basis function with support in \([ (j - 1)/m, j/m ] \), that is if \( t = \varphi \lambda \) when considering collection [P]:

\[ |R_n^*(\varphi \lambda, a)| \leq R_n(2\|\varphi \lambda\|_{\infty} + \int_{0}^{1} |\varphi \lambda'(x)| dx) := R_nN(\varphi \lambda), \]

with \( R_n \) defined by (6.4).

Let us be now more precise on the piecewise Legendre polynomial basis. Since \( \varphi_{j,k}(x) = \sqrt{m(2k+1)} Q_k(2mx - 2j + 1) I_{(j-1)/m, j/m}(x) \) for \( j = 1, \ldots, m \) and \( k = 0, \ldots, r \), with \( D_m = (r + 1)m \), we have \( \|\varphi_{j,k}\|_{\infty} = \sqrt{m(2k+1)} \leq \sqrt{2m(r + 1)} \leq \sqrt{2D_m} \). Moreover,

\[ \int_{0}^{1} |\varphi_{j,k}'(x)| dx = \sqrt{m(2k+1)} \int_{(j-1)/m}^{j/m} 2m |Q_k'(2mx - 2j + 1)| dx \]

\[ = \sqrt{m(2k+1)} \int_{-1}^{1} |Q_k'(u)| du \leq C_r \sqrt{2D_m} \]
where $C_r = \max_{k=0,\ldots,r} \int_{-1}^{1} |Q_k'(u)| du$. It follows that

$$N(\varphi_{j,k}) \leq (1 + C_r) \sqrt{2D_m}.$$  

The same type of proof obviously holds for wavelets with $\varphi_{j,k} = \phi_{j,k}$, it is clear that $\|\varphi_{j,k}\|_{\infty} \leq 2^{j/2} \|\phi\|_{\infty} \leq \|\phi\|_{\infty} \sqrt{D_m/2}$ and

$$\int_{0}^{1} |\varphi'_{j,k}(x)| \, dx = \sqrt{2^j} \int_{k/2^j}^{(k+1)/2^j} 2^j |\phi'(2^j x - k)| \, dx = \sqrt{D_m/2} \int_{0}^{1} |\phi'(u)| \, du.$$  

The same holds with $\phi$ replaced by $\psi$ and it follows that

$$N(\varphi_{j,k}) \leq C(\phi, \psi) \sqrt{D_m/2},$$  

with $C(\phi, \psi) = \left( \max(\|\phi\|_{\infty}; \|\psi\|_{\infty}) + \max(\int_{0}^{1} |\phi'(u)| \, du; \int_{0}^{1} |\psi'(u)| \, du) \right)$. 

As a consequence, for collections $[P]$ and $[W]$, we find

$$(6.6) \quad \sup_{t \in S_m, \|t\|=1} [R^*_n(t, a)]^2 = \sum_{\lambda \in \Lambda_m} [R^*_n(\varphi_{\lambda})]^2 \leq R^2_n \sum_{\lambda \in \Lambda_m} [N(\varphi_{\lambda})]^2 \leq CR^2_n D_m^2.$$  

And more generally, we have

$$\sup_{t \in S_m, \|t\|=1} [R^*_n(t, a)]^{2k} \leq \left( \sup_{t \in S_m, \|t\|=1} [R^*_n(t, a)]^2 \right)^k \leq C^k R^2_n D_m^{2k}.$$  

By using (6.1) for $\beta = 2k$, this implies (3.10).  

6.2. Proof of Proposition 3.1. It remains to prove the result for the spline basis. By starting from (3.6), we find that

$$\mathbb{E}(\|\hat{h}_m - h\|^2) \leq 3\|h_m - h\|^2 + 8 \mathbb{E} \left( \sup_{t \in S_m, \|t\|=1} |C(t, 1)|^2 \right) .$$  

By using the standard decomposition, we find

$$\mathbb{E} \left( \sup_{t \in S_m, \|t\|=1} C^2(t, 1) \right) \leq 2 \mathbb{E} \left[ \sup_{t \in S_m, \|t\|=1} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(t, 1) \right)^2 \right] + 2 \mathbb{E} \left( \sup_{t \in S_m, \|t\|=1} R^*_n(t, 1) \right) .$$  

Obviously from Lemma (3.1),

$$\mathbb{E} \left( \sup_{t \in S_m, \|t\|=1} R^2_n(t, 1) \right) \leq \Phi_0 \frac{D^2_m \ln^2(n)}{n^2} .$$
and

$$\mathbb{E} \left[ \sup_{t \in S_m, \|t\| = 1} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(t, 1) \right)^2 \right]$$

$$\leq \mathbb{E} \left\{ \sup_{t = \sum_{k \in \mathbb{K}_m} a_{m,k} \hat{B}_{m,k}, \|t\| = 1} \left[ \sum_{k \in \mathbb{K}_m} a_{m,k} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(\hat{B}_{m,k}, 1) \right) \right]^2 \right\}$$

$$\leq \mathbb{E} \left\{ \sup_{t = \sum_{k \in \mathbb{K}_m} a_{m,k} \hat{B}_{m,k}, \|t\| = 1} \left[ \sum_{k \in \mathbb{K}_m} a_{m,k}^2 \right] \left[ \sum_{k \in \mathbb{K}_m} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(\hat{B}_{m,k}, 1) \right)^2 \right] \right\}$$

$$\leq \Phi_0^2 \mathbb{E} \left[ \sum_{k \in \mathbb{K}_m} \left( \frac{1}{n} \sum_{i=1}^{n} Z_i(\hat{B}_{m,k}, 1) \right)^2 \right] \text{ with (S5)}$$

$$\leq \frac{\Phi_0^2}{n} \sum_{k \in \mathbb{K}_m} \text{Var}(Z_1(\hat{B}_{m,k}, 1)) = \frac{\Phi_0^2}{n} \sum_{k \in \mathbb{K}_m} \int_0^1 \frac{\hat{B}_{m,k}^2(x)h(x)}{1-F(x)} \, dx$$

$$\leq \frac{\Phi_0^2}{n} \sqrt{D_m} \int_0^1 \sum_{k \in \mathbb{K}_m} B_{m,k}(x) \frac{h(x)}{1-F(x)} \, dx \text{ with (S1)},$$

$$\leq \frac{\Phi_0^2}{n} \sqrt{D_m} \int_0^1 \sum_{k \in \mathbb{Z}} B_{m,k}(x) \frac{h(x)}{1-F(x)} \, dx$$

$$\leq \frac{\Phi_0^2 D_m}{n} \int_0^1 \frac{h(x)}{1-F(x)} \, dx \text{ with (S2)}.$$

6.3. Proof of Theorem 3.1. We have to prove the two following results

$$h_m(x) - h(x) = o \left( \sqrt{\frac{D_m}{n}} \right)$$

and that

$$\sqrt{\frac{n}{D_m}} \left( h_m(x) - h_m(x) \right) \rightarrow \mathcal{N}(0, h(x)/(1 - F(x))).$$

To prove (6.7), we prove that $h_m(x) - h(x) = O(1/D_m)$ and the result follows from the assumption $n/D_m^3 \rightarrow 0$ as $n \rightarrow \infty$. Since histograms are particular piecewise polynomials, we can see that $D_m = (r+1)m = m$ with $r = 0$. Let $x \in [0, 1]$, then for any fixed $n$ and associated $D_m = m_n$, there exists some integer $j_n$ such that
\[ x \in [(j_n - 1)/D_m, j_n/D_m]. \] Therefore
\[ h(x) - h_m(x) = h(x) - D_m \left( \int_{(j_n - 1)/D_m}^{j_n/D_m} h(u) du \right) 1_{[(j_n - 1)/D_m, j_n/D_m]}(x). \]

Then, by writing that \( h(u) = h(x) + (x - u)h'(u_x) \) for some \( u_x \in [u, x] \), we find
\[ h(x) - h_m(x) = D_m \left( \int_{(j_n - 1)/D_m}^{j_n/D_m} (x - u)h'(u_x) du \right) 1_{[(j_n - 1)/D_m, j_n/D_m]}(x) \]
and consequently,
\[ |h(x) - h_m(x)| \leq D_m \|h'(1)\|_{\infty} \left( \int_{(j_n - 1)/D_m}^{j_n/D_m} |x - u| du \right) \leq \frac{\|h'(1)\|_{\infty}}{D_m} = o(\sqrt{D_m/n}). \]

Let us turn to (6.8). First \( \hat{h}_m - h_m = \sum_{\lambda} (\hat{a}_\lambda - a_\lambda) \varphi_{\lambda} = \sum_{\lambda} C(\varphi_{\lambda}, 1) \varphi_{\lambda} \), and following decomposition (3.8) in Lemma 3.1, we get
\[ \hat{h}_m(x) - h_m(x) = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{\lambda} \varphi_{\lambda}(x) Z_i(\varphi_{\lambda}, 1) \right) + \sum_{\lambda} R^*_n(\varphi_{\lambda}, 1) \varphi_{\lambda}(x) \]
Now, if the remainder term \( \sum_{\lambda} R^*_n(\varphi_{\lambda}, 1) \varphi_{\lambda}(x) = o_P\left(\sqrt{\frac{n}{D_m}}\right) \), it suffices to prove that
\[ \sqrt{\frac{n}{D_m}} \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{\lambda} \varphi_{\lambda}(x) Z_i(\varphi_{\lambda}, 1) \right) \rightarrow N(0, \eta^2(x)) \]
Since the \( Y_{i,n}(x) = \sum_{\lambda} \varphi_{\lambda}(x) Z_i(\varphi_{\lambda}, 1) \) are centered i.i.d. variables, a Lyapounov central limit theorem applies.

So, let us compute the asymptotic variance of \( \frac{1}{\sqrt{n} D_m} \sum_{i=1}^{n} Y_{i,n}(x) \). Now remember that \( \mathbb{E}(Z_1(\varphi_{\lambda}, 1))^2 = \int_0^1 \varphi_{\lambda}^2(x) h(x)/(1 - F(x)) dx \). Let \( j_n \) and \( D_m \) be defined as previously and let \( \hat{h}(u) = h(u)/(1 - F(u)) \).
\[
\frac{1}{D_m} \mathbb{E}(Y^2_{1,n}(x)) = \frac{1}{D_m} \mathbb{E} \left( \sum_{\lambda} \varphi_{\lambda}(x) Z_1(\varphi_{\lambda}, 1) \right)^2
= \frac{1}{D_m} \int_0^1 \left( D_m \sum_{j=1}^{D_m} 1_{[(j_n - 1)/D_m, j_n/D_m]}(x) 1_{[(j_n - 1)/D_m, j_n/D_m]}(u) \right) \hat{h}(u) du
= D_m \int_{(j_n - 1)/D_m}^{j_n/D_m} \tilde{h}(u) du.
\]
By applying a Taylor expansion, we find
\[ \tilde{h}(u) = \tilde{h}(x) + (x - u)\tilde{h}'(u_x), \quad u_x \in \left[ \frac{j_n - 1}{D_m}, \frac{j_n}{D_m} \right] \]
which implies
\[ \frac{1}{D_m} \mathbb{E}(Y_{1,n}^2(x)) = \tilde{h}(x) + D_m \int_{(j_n-1)/D_m}^{j_n/D_m} \tilde{h}'(u_x)(x - u)du. \]
Since as previously
\[ D_m \left| \int_{(j_n-1)/D_m}^{j_n/D_m} \tilde{h}'(u_x)(x - u)du \right| \leq \frac{\|\tilde{h}'\|_{\infty}}{D_m}, \]
we get
\[ \lim_{n \to +\infty} \frac{1}{D_m} \mathbb{E}(Y_{1,n}^2(x)) = \tilde{h}(x). \]
Then Lyapounov’s Theorem applies if in addition
\[ \frac{1}{(\sqrt{n}D_m)^4} \sum_{i=1}^{n} \mathbb{E}(Y_{i,n}^4(x)) \to 0. \]
This follows from the following obvious bound, provided that \( D_m = o(\sqrt{n}) \):
\[ \frac{1}{(\sqrt{n}D_m)^4} \sum_{i=1}^{n} \mathbb{E}(Y_{i,n}^4(x)) = \frac{1}{n} \mathbb{E}(Z_1^4(\varphi_{j_n}, 1)) \leq \frac{D_m^2}{n} \left( \int_0^1 \tilde{h}(u)du + \sup_{x \in [0,1]} (1 - F(x))^{-1} \right)^4. \]
Lastly, to make the proof complete, we have to prove that the remainder term
\[ \sum_{\lambda} R_n^*(\varphi_{\lambda}, 1) \varphi_{\lambda}(x) \text{ is } o_P \left( \frac{\sqrt{n}}{D_m} \right). \]
\begin{align*}
\mathbb{E} \left( \sum_{\lambda} R_n^*(\varphi_{\lambda}) \varphi_{\lambda}(x) \right)^2 & \leq D_m \mathbb{E} \left( \sum_{\lambda} R_n^*(\varphi_{\lambda})^2 \right) \\
& \leq D_m \mathbb{E} \left( \sup_{t \in S_m, \|t\|=1} [R_n^*(t, 1)]^2 \right) \\
& \leq \frac{D_m^3 \ln^2(n)}{n^2} \quad \text{by applying Lemma 3.1.}
\end{align*}
This last inequality gives
\[ \mathbb{P} \left( \sqrt{\frac{D_m}{n}} \sum_{\lambda} R_n^*(\varphi_{\lambda}, 1) \varphi_{\lambda}(x) > \varepsilon \right) \leq \frac{D_m^3 \ln^2(n)}{n^3 \varepsilon^2} \]
which tends to 0 since \( D_m \leq \sqrt{n} \). \hfill \Box
6.4. Proof of Lemma 4.1. In order to control
\[ \mathbb{E} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) - p(m, m') \right), \]
we use the following version of Talagrand’s Inequality (see Talagrand (1996)):

**Lemma 6.1.** Let \( X_1, \ldots, X_n \) be i.i.d. random variables and \( \nu_n(g) = (1/n) \sum_{i=1}^n [g(X_i) - \mathbb{E}(g(X_i))] \) for \( f \) belonging to a countable class \( G \) of uniformly bounded measurable functions. Then for \( \epsilon > 0 \)
\[ (6.9) \]
\[ \mathbb{E} \left[ \sup_{g \in G} |\nu_n(g)|^2 - 2(1 + 2\epsilon)H^2 \right] \leq \frac{6}{K_1} \left( \frac{1}{n} e^{-K_1 n^2 C(\epsilon)} e^{\frac{-K_1 C(\epsilon)}{\sqrt{2} M_1}} \right), \]
with \( C(\epsilon) = (\sqrt{1 + \epsilon} - 1) \wedge 1 \), \( K_1 \) is a universal constant, and where
\[ \sup_{g \in G} \|g\|_\infty \leq M_1, \quad \mathbb{E} \left( \sup_{g \in G} |\nu_n(g)| \right) \leq H, \quad \sup_{g \in G} \text{Var}(g(X_1)) \leq v. \]

We apply Talagrand’s inequality by taking
\[ g(x) = g_t(x) = \int_0^{x \wedge 1} t(u) \frac{h(u)}{1 - F(u)} du - I_{\{\delta_i = 1\} \cap \{x \leq 1\}} \frac{t(x)}{1 - F(x)}. \]

Usual density arguments show that this result can be applied to the class of functions \( G = \{g_t, t \in B_{m,m'}(0,1)\} \). Then we find for the present empirical process the following bounds:
\[ \sup_{g \in G} \|g\|_\infty = \sup_{t \in B_{m,m'}(0,1)} \|g_t\|_\infty \leq \Phi_0 C_1 \sqrt{D(m')} := M_1 \]
with \( D(m') \) denoting the dimension of \( S_m + S_{m'} \) and
\[ (6.10) \]
\[ C_1 = \int_0^1 \frac{h(u)}{1 - F(u)} du + \sup_{0 \leq x \leq 1} (1 - F(x))^{-1}. \]

Then
\[ \sup_{g \in G} \text{Var}(g(X_1)) = \sup_{t \in B_{m,m'}(0,1)} \text{Var}(g_t(X_1)) \leq \sup_{x \in [0,1]} \left( \frac{h(x)}{1 - F(x)} \right) = C_2 := v. \]
Lastly,

\[
\mathbb{E} \left( \sup_{g \in B} \nu_n^2(g) \right) = \mathbb{E} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) \right) \\
\leq \sum_{\lambda \in \Lambda_{m,m'}} \frac{1}{n} \text{Var}(Z_1(\varphi_{\lambda})) \\
\leq \frac{\Phi_0^2 D(m')}{n} \int_0^1 \frac{h(x)}{1 - F(x)} \, dx = C_3 \frac{D(m')}{n} := H^2.
\]

Then it follows from (6.9) that

\[
\mathbb{E} \left( \sup_{t \in B_{m,m'}(0,1)} \nu_n^2(g_t) - p(m, m') \right) \leq \kappa_1 \left( \frac{1}{n} e^{-\kappa_2 D(m')} + \frac{1}{nC^2(\epsilon)} e^{-\kappa_3 \epsilon^{3/2}/\sqrt{n}} \right),
\]

where \( \kappa_i \) for \( i = 1, 2, 3 \) are constant depending on \( K_1, C_1, C_2 \) and \( C_3 \) and \( p(m, m') = 2(1 + 2\epsilon)C_3(D_m + D_{m'}/n) \).

\[\square\]

6.5. Proof of Theorem 4.2. Let

\[
\Omega_b = \left\{ \left( \int_0^1 \frac{\hat{h}_n(x)}{1 - \hat{F}_n(x)} \, dx \right) / \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) - 1 \right\} < b, \quad 0 < b < 1.
\]

Then on \( \Omega_b \), the proof is quite similar to the proof of theorem 4.1. The following inequalities hold

\[
\int_0^1 \frac{\hat{h}_n(x)}{1 - \hat{F}_n(x)} \, dx < (b+1) \int_0^1 \frac{h(x)}{1 - F(x)} \, dx, \quad \int_0^1 \frac{h(x)}{1 - F(x)} \, dx < \frac{1}{1-b} \int_0^1 \frac{\hat{h}_n(x)}{1 - \hat{F}_n(x)} \, dx.
\]

Then we can mimick the proof of (4.7) with \( \text{pen}(m) \) replaced now by \( \text{pen}(m) = \kappa \Phi_0^2 \int_0^1 \hat{h}_n(x)/(1 - \hat{F}_n(x)) \, dx(D_m/n) \) and by defining

\[
p(m, m') = \Phi_0^2 \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m + D_{m'}}{n},
\]

we have

\[
\frac{1}{2} \left\| \hat{h}_m - h \right\|^2 \leq \frac{3}{2} \left\| h_m - h \right\|^2 + 8 \sup_{t \in B_{m,m'}(0,1)} (\nu_n^2(g_t) - p(m, \hat{m})) + 8 \sup_{t \in B_{m,m'}(0,1)} R_n^2(t) + 8 \text{pen}(m) + \text{pen}(\hat{m}).
\]

By taking \( \kappa = 8\Phi_0^2/(1 - b) \), we find that on \( \Omega_b \),
8p(m, \hat{m}) - \text{pen}(\hat{m}) + \text{pen}(m) \leq 8\Phi_0^2 \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m}{n} \\
+ 8\Phi_0^2 \left( \int_0^1 \frac{\hat{h}_n(x)}{1 - F_n(x)} \, dx \right) \frac{D_m}{n} \\
\leq \frac{16\Phi_0^2}{1 - b} \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m}{n}.

It follows that \forall m \in \mathcal{M}_n,

\begin{align*}
\mathbb{E} \left( \|\hat{h}_n - h\|^2 \mathbf{1}_{\Omega_b} \right) & \leq 3\|h_m - h\|^2 + 16 \sum_{m' \in \mathcal{M}_n} \left( \sup_{t \in B_{m, m'}} (\nu_n^2(g_t) - p(m, m')) \right) \\
& + 16 \mathbb{E} \left( \sup_{t \in B_{m, m'}(0, 1)} R^2_n(t) \right) + \frac{32\Phi_0^2}{1 - b} \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m}{n} \\
& \leq 3\|h_m - h\|^2 + \frac{32\Phi_0^2}{1 - b} \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) \frac{D_m}{n} + \frac{K}{n}.
\end{align*}

Next we need to prove that

(6.11) \quad \mathbb{E} \left( \|\hat{h}_n - h\|^2 \mathbf{1}_{\Omega_b} \right) \leq \frac{K'}{n}.

It follows from (3.5) and (3.6) that

\begin{align*}
\|\hat{h}_n - h\|^2 & \leq C_1\|h - h_m\|^2 + 2C_2 \sup_{t \in S_m, \|t\|=1} \nu_n^2(g_t) + 2C_2 \sup_{t \in S_m, \|t\|=1} [R_n^*(t)]^2 \\
& \leq C_1\|h\|^2 + 2C_2 \sup_{t \in S_m, \|t\|=1} \nu_n^2(g_t) + 2C_2 \sup_{t \in S_m, \|t\|=1} [R_n^*(t)]^2
\end{align*}

(6.12)

with (C_1, C_2) = (1, 1) for [T], [P] and [W] and (C_1, C_2) = (3, 8) for [B]. Then

\begin{align*}
\sup_{t \in S_m, \|t\|=1} \nu_n^2(g_t) & = \left( \sup_{t \in S_m, \|t\|=1} \nu_n^2(g_t) - \text{pen}(\hat{\hat{m}}) \right) + \text{pen}(\hat{\hat{m}}) \\
& \leq \sum_{m \in \mathcal{M}_n} \left( \sup_{t \in S_m, \|t\|=1} \nu_n^2(g_t) - \text{pen}(m) \right) + \text{pen}(\hat{\hat{m}}).
\end{align*}
Then we know by Lemma 4.1 that, for some well-known \( \kappa \) in \( \text{pen}(m) \),

\[
\mathbb{E} \left( \sum_{m \in M_n} \left( \sup_{t \in S_m, ||t||=1} (\nu_n^2(t) - \text{pen}(m))I_{\Omega_k} \right) \right) 
\leq \mathbb{E} \left( \sum_{m \in M_n} \left( \sup_{t \in S_m, ||t||=1} (\nu_n^2(t) - \text{pen}(m)) \right) \right) \leq K/n. 
\]

On the other hand, \( \forall m \in M_n, \text{pen}(m) \leq K' \), with \( K' = \kappa \Phi_0^2 \int_0^1 h(x)/(1 - F(x)) \, dx \), so that

\[
\mathbb{E} \left( \text{pen}(\hat{m}) I_{\Omega_k} \right) \leq K' \mathbb{P}(\Omega_k^c). 
\]

Therefore

(6.13) \[
\mathbb{E} \left( \sup_{t \in S_m, ||t||=1} \nu_n(g_t)^2 I_{\Omega_k} \right) \leq \frac{K}{n} + K' \mathbb{P}(\Omega_k^c). 
\]

In all cases of basis \([T], [P], [W], [B]\) we have as soon as \( D_m \leq \sqrt{n} \ \forall m \in M_n, \)

\[
\sup_{t \in S_m, ||t||=1} [R_n^*(t)]^2 \leq K D_m^2 R_n^2 \leq K n R_n^2, 
\]

so that we find,

(6.14) \[
\mathbb{E} \left( \sup_{t \in S_m, ||t||=1} [R_n^*(t)]^2 I_{\Omega_k} \right) \leq K_n \mathbb{E}^{1/2}(R_n^4) \mathbb{P}^{1/2}(\Omega_k^c) \leq K \frac{n^2(n)}{n} \mathbb{P}^{1/2}(\Omega_k^c), 
\]

so that, by gathering (6.12), (6.13) and (6.14), the result (6.11) holds provided that

(6.15) \[
\mathbb{P}(\Omega_k^c) \leq \frac{1}{n}. 
\]

We recall first that \( \hat{h}_n = \hat{h}_{m_n} \) is one particular estimator of our collection of models. Let \( B = b \int_0^1 h(x)/(1 - F(x)) \, dx \).

\[
\left| \int_0^1 \frac{\hat{h}_{m_n}(x)}{1 - \hat{F}_n(x)} \, dx - \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right|
\leq \int_0^1 \frac{|\hat{h}_{m_n}(x) - h(x)|}{1 - \hat{F}_n(x)} \, dx + \int_0^1 h(x) \left| \frac{1}{1 - \hat{F}_n(x)} - \frac{1}{1 - F(x)} \right| \, dx 
\leq \int_0^1 \frac{|\hat{h}_{m_n}(x) - h(x)|}{1 - \hat{F}_n(1)} \, dx + \int_0^1 \frac{h(x) |\hat{F}_n(x) - F(x)|}{(1 - \hat{F}_n(1))(1 - F(1))} \, dx 
\]

and since \( |\hat{F}_n(1) - F(1)| < (1 - F(1))/2 \) is equivalent to

\[
\frac{2}{3(1 - F(1))} < \frac{1}{1 - \hat{F}_n(1)} < \frac{2}{1 - F(1)}
\]

it follows that.
\[\mathbb{P}(\Omega^c) \leq \mathbb{P}\left( \int_0^1 \frac{1}{1 - \hat{F}_n(x)} |\hat{h}_{mn}(x) - h(x)|dx > B/2 \right)\]

\[+ \mathbb{P}\left( 1 - \hat{F}_n(1) \int_0^1 h(x)\hat{F}_n(x) - F(x)|dx > B/2 \right)\]

\[\leq \mathbb{P}\left( \frac{2}{1 - F(1)} \int_0^1 |\hat{h}_{mn}(x) - h(x)|dx > B/2 \right)\]

\[+ \mathbb{P}\left( \frac{2}{(1 - F(1))^2} \int_0^1 h(x)\hat{F}_n(x) - F(x)|dx > B/2 \right)\]

\[+ 2\mathbb{P}(|\hat{F}_n(1) - F(1)| > (1 - F(1))/2).\]

We bound successively the three above terms. First,

\[\mathbb{P}(|\hat{F}_n(1) - F(1)| > (1 - F(1))/2) \leq \frac{4}{(1 - F(1))^2} \mathbb{E}(\hat{F}_n(1) - F(1))^2\]

\[= \frac{4}{(1 - F(1))^2} \left[ \frac{F(1)(1 - F(1))}{n + 1} + \frac{(F(1))^2}{n + 1} \right]\]

\[\leq \frac{4 F(1)/(1 - F(1))^2}{n + 1}.\]

Secondly,

\[\mathbb{P}\left( \int_0^1 h(x)\hat{F}_n(x) - F(x)|dx > \frac{B(1 - F(1))^2}{4} \right) \leq \left( \frac{4}{B(1 - F(1))^2} \right)^2 \mathbb{E}\left( \int_0^1 h^2(x)(\hat{F}_n(x) - F(x))^2dx \right)\]

\[\leq \left( \frac{4}{B(1 - F(1))^2} \right)^2 \int_0^1 h^2(x) \left[ \frac{F(1)(1 - F(x))}{n + 1} + \frac{(F(1))^2}{n + 1} \right] dx\]

\[\leq \frac{24 \int_0^1 h^2(x)dx / (B^2(1 - F(1)))^4}{n}.\]

Lastly as

\[\int_0^1 |\hat{h}_{mn}(x) - h(x)|dx \leq ||\hat{h}_{mn} - h|| \leq ||\hat{h}_{mn} - h_{mn}|| + ||h_{mn} - h||\]

\[\leq ||\hat{h}_{mn} - h_{mn}|| + B(1 - F(1))/8\]
under condition (4.11),
\[ P\left( \int_0^1 |\hat{h}_{mn}(x) - h(x)| \, dx > B(1 - F(1))/4 \right) \leq \frac{8}{B(1 - F(1))} \left( \frac{\|\hat{h}_{mn} - h_{mn}\|}{1 - F(1)} \right)^4 \]
\[ \leq \left( \frac{8}{B(1 - F(1))} \right)^4 \mathbb{E}(\|\hat{h}_{mn} - h_{mn}\|^4). \]

Then we need to study the condition on \( D_{mn} \) that ensures \( \mathbb{E}(\|\hat{h}_{mn} - h_{mn}\|^4) \leq 1/n. \)

\[ \mathbb{E} \left[ \left( \|\hat{h}_{mn} - h_{mn}\|^2 \right)^2 \right] \leq 4 \mathbb{E} \left\{ \left[ \sum_{\lambda \in \Lambda_{mn}} \left( \frac{1}{n} \sum_{i=1}^n Z_i(\varphi_\lambda) \right)^2 \right]^2 \right\} + 4 \mathbb{E} \left[ \left( \sum_{\lambda \in \Lambda_{mn}} R_n^2(\varphi_\lambda) \right)^2 \right] (6.16) \leq 4 D_{mn} \sum_{\lambda \in \Lambda_{mn}} \mathbb{E} \left( \left( \frac{1}{n} \sum_{i=1}^n Z_i(\varphi_\lambda) \right)^4 \right) + 4 \mathbb{E} \left( \sup_{t \in S_{mn}, \|t\|=1} |R_n^*(t)|^4 \right). \]

For the last term, the bound in (3.10) gives
\[ \mathbb{E} \left( \sup_{t \in S_{mn}, \|t\|=1} |R_n^*(t)|^4 \right) \leq K(2, \Phi_0, 1) \frac{D_{mn}^4 \ln^4(n)}{n^4} \leq K(2, \Phi_0, 1) \frac{\ln^4(n)}{n^2}. \]

For the first term, we apply Rosenthal’s Inequality (see Petrov (1995)). Let \( U_1, \ldots, U_n \) be independent centered random variables with values in \( \mathbb{R} \). For any \( p \geq 2 \), we have
\[ \mathbb{E} \left[ \left( \sum_{i=1}^n U_i \right)^p \right] \leq C(p) \left( \sum_{i=1}^n \mathbb{E}[|U_i|^p] + \left( \sum_{i=1}^n \mathbb{E}[U_i^2] \right)^{p/2} \right). \]

This yields
\[ \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n Z_i(\varphi_\lambda) \right)^4 \leq \frac{C(4)}{n^4} \left( n\mathbb{E}(Z_1^4(\varphi_\lambda)) + n^2[\mathbb{E}(Z_1^2(\varphi_\lambda))]^2 \right). \]

Since
\[ \sum_{\lambda \in \Lambda_{mn}} [\mathbb{E}(Z_1^2(\varphi_\lambda))]^2 = \sum_{\lambda \in \Lambda_{mn}} \left( \int_0^1 \varphi_\lambda^2(x) \frac{h(x)}{(1 - F(x))} \, dx \right)^2 \leq \sum_{\lambda \in \Lambda_{mn}} \int_0^1 \varphi_\lambda^2(x) \, dx \times \int_0^1 \varphi_\lambda^2(x) \frac{h^2(x)}{(1 - F(x))^2} \, dx \leq \| \sum_{\lambda \in \Lambda_{mn}} \varphi_\lambda^2 \|_{\infty} \int_0^1 \frac{h^2(x)}{(1 - F(x))^2} \, dx \leq \Phi_0^2 D_{mn} \int_0^1 \frac{h^2(x)}{(1 - F(x))^2} \, dx \]

On the other hand

\[ \sum_{\lambda \in \Lambda_{mn}} \mathbb{E}(Z_1^4(\varphi_\lambda)) \leq \sum_{\lambda \in \Lambda_{mn}} \int_0^1 \varphi_\lambda^4(x) \left( 1 + \frac{h^3(x)}{1 - F(x)} \right) \frac{h(x)}{(1 - F(x))^3} \, dx \]

we find, by using that \( \sum_{\lambda \in \Lambda_{mn}} \varphi_\lambda^2(x) \leq \| \varphi_\lambda \|_{\infty} \| \sum_{\lambda \in \Lambda_{mn}} \varphi_\lambda^2 \|_{\infty} \leq \Phi_0^4 D_{mn}^2 \), and from (6.16)

\[ \mathbb{E} \left( \left( \| \hat{h}_{mn} - h_{mn} \|^2 \right)^2 \right) \leq K'' \left( \frac{D_{mn}^3}{n^3} + \frac{D_{mn}^2}{n^2} \right) \leq \frac{K''}{n} \]

which gives the announced order 1/n as soon as \( D_{mn} \leq \sqrt{n} \). \( \square \)

6.6. Proof of Corollary 4.1. The first part of the proof is the same as the proof of theorem 4.2 with now

\[ \widetilde{\Omega}_b = \left\{ \left( \frac{1}{n} \sum_{i=1}^n I(X_i \leq 1) I(\delta_i = 1) \right) / \left( \int_0^1 \frac{h(x)}{1 - F(x)} \, dx \right) - 1 > b \right\} \]

The result holds if we can prove that \( \mathbb{P}(\widetilde{\Omega}_b) \leq 1/n \). Let \( B = b \int_0^1 h(x)/(1 - F(x)) \, dx \), then by writing that
where

\[
\sup_{\lambda > 0} \mathbb{P}(\sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \geq \lambda) \leq 2e^{-2\lambda^2},
\]

we find

\[
\mathbb{P}(\tilde{\Omega}_k) \leq \mathbb{P}\left( |\hat{F}_n(1) - F(1)| > \frac{1 - F(1)}{2} \right)
+ \mathbb{P}\left( \frac{4}{(1 - F(1))^4} \left( \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq 1) I(\delta_i = 1) |\hat{F}_n(X_i) - F(X_i)| \right) > B/2 \right)
+ \mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} \left( \frac{I(X_i \leq 1) I(\delta_i = 1)}{(1 - F(X_i))^2} - \int_0^1 \frac{h(x)}{1 - F(x)} dx \right) \right| > B/2 \right)
\leq \frac{4}{(1 - F(1))^2} \left[ \frac{F(1)(1 - F(1))}{n + 1} + \left( \frac{F(1)}{n + 1} \right)^2 \right]
+ K_1 \mathbb{E}\left( \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq 1) I(\delta_i = 1) |\hat{F}_n(X_i) - F(X_i)| \right)^2
+ K_2 \mathbb{E}\left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{I(X_i \leq 1) I(\delta_i = 1)}{(1 - F(X_i))^2} - \int_0^1 \frac{h(x)}{1 - F(x)} dx \right) \right)^2
\leq \frac{6}{n(1 - F(1))^2} + K_1 \mathbb{E}(\sup_{0 \leq x \leq 1} |\hat{F}_n(x) - F(x)|^2) + K_2 \frac{\text{Var} \left( \frac{I(X_i \leq 1) I(\delta_i = 1)}{(1 - F(X_i))^2} \right)}{n},
\]

where $K_1 = 64(1 - F(1))^4/B^2$ and $K_2 = 4/B^2$. Therefore, since it follows from Massart (1990) that $\forall \lambda > 0, \mathbb{P}(\sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \geq \lambda) \leq 2e^{-2\lambda^2}$, where
\[ F_n(x) = (1/n) \sum_{i=1}^{n} I(x_i \leq x) \] we have

\[
\mathbb{E}[ \sup_{0 \leq x \leq 1} (F_n(x) - F(x))^2 ] \leq 2 \mathbb{E}[ \sup_{0 \leq x \leq 1} (\hat{F}_n(x) - F_n(x))^2 ] + 2 \mathbb{E}[ \sup_{0 \leq x \leq 1} (F_n(x) - F(x))^2 ]
\]

\[
\leq \frac{2}{(n+1)^2} + \frac{4}{n} \leq \frac{5}{n}.
\]

and finally

\[
\mathbb{P}(\omega_c) \leq K'/n.
\]

\[ 6.7. \textbf{Proof of Theorem 4.3.} \]

Lemma 3.1 still holds but with since now \( N, N' \leq \sqrt{(2r+1)n} \leq (2r+1)n \)

so that we simply replace \( \Phi \) by \( \sqrt{(2r+1)n} \). Therefore Inequality (4.7) still holds with

\[
\mathbb{E}(\sup_{t \in B_{m',(0,1)}} R_n^2(t)) \leq \frac{(2r+1)\ln^2(n)}{n}.
\]

Moreover, we keep applying Lemma 6.1 but now

\[
\sup_{t \in B_{m',(0,1)}} \|g_t\|_\infty \leq C_1 \quad \sup_{t \in B_{m,(0,1)}} \|g_t\|_\infty \leq C_1 \sqrt{(2r+1)n^{1/4}} := M_1
\]

where \( C_1 \) is given by (6.10). The bound giving \( v = C_2 \) is unchanged. Lastly

\[
\mathbb{E}\left( \sup_{t \in B_{m',(0,1)}} \nu_n^2(g_t) \right) \leq \frac{1}{n} \sum_{\lambda \in \Lambda_{m',m}} \int \varphi_{\lambda}(x) h(x) \frac{1}{1-F(x)} dx
\]

\[
\leq \frac{1}{n} \sup_{x \in [0,1]} \frac{h(x)}{1-F(x)} \sum_{\lambda \in \Lambda_{m',m}} \int \varphi_{\lambda}(x) dx
\]

\[
\leq \frac{D_m + D_{m'}}{n} \sup_{x \in [0,1]} \frac{h(x)}{1-F(x)} := H^2.
\]

We denote by

\[
C_3 = \sup_{x \in [0,1]} \frac{h(x)}{1-F(x)}.
\]

Then it follows from (6.9) that

\[
\mathbb{E}\left( \sup_{t \in B_{m',(0,1)}} \nu_n^2(g_t) - p(m,m') \right) \leq \kappa_1 \left( \frac{1}{n} e^{-\kappa_2 D(m')} + \frac{1}{n} e^{-\kappa_3 \sqrt{n^{1/4} D(m')}} \right),
\]
where $\kappa_i$ for $i = 1, 2, 3$ are constant depending on $K_1, C_1, C_2$ and $C_3$ and $p(m, m') = 2(1 + 2\epsilon)C_3\frac{D_m + D_{m'}}{n}$ with $\epsilon > 1$ since we need $\epsilon = \alpha \ln^2(n)$. Therefore we choose

$$\text{pen}(m) = 16(1 + 2\alpha \ln^2(n)) \sup_{x \in [0,1]} \frac{h(x)}{1 - F(x)} \frac{D_m}{n}$$

and we find

$$\sum_{m' \in \mathcal{M}_n} \left( \sup_{t \in B_{m,m'}(0,1)} \nu^2_n(q_t) - p(m, m') \right) \leq \frac{\kappa_1}{n} \sum_{m' \in \mathcal{M}_n} \left( e^{-\kappa_2 \alpha \ln(n) D_{m'}} + e^{-\kappa_3 \sqrt{n} \ln(n) n^{1/4} \sqrt{D_{m'}}} \right)$$

by using that $\sqrt{D_{m'}} \leq n^{1/4}$ and $\ln^2(n) \geq \ln(n)$, for $n \geq 3$. Now we note that

$$\sum_{m' \in \mathcal{M}_n} e^{-\beta \ln(n) D_{m'}} = \sum_{\ell = 1}^{\sqrt{n} - 1} \left( \sqrt{n} - 1 \right) e^{-\beta \ln(n) \ell} = (1 + n^{-\beta})\sqrt{n} - 1$$

by using that there are $\left( \sqrt{n} - 1 \right)$ models of dimension $\ell$. The resulting term is bounded as soon as $\beta \geq 1/2$ since it has the order of $\exp(\sqrt{n} \ln(1 + n^{-\beta})) \sim \exp(\sqrt{n} n^{-\beta})$. The choice $\alpha = 1/(2\kappa_3^2) + 1/(2\kappa_2)$ suits. \hfill \Box

**References**


