SPLINE REGRESSION FOR HAZARD RATE ESTIMATION WHEN DATA ARE CENSORED AND MEASURED WITH ERROR

FABIENNE COMTE [(1)], GWENNAËLLE MABON [(1,2)], AND ADELINE SAMSON [(3)]

[(1)] MAP5, *UMR CNRS 8145, Université Paris Descartes, France*[1]
[(2)] CREST - ENSAE, *Malakoff, France*[2]
[(3)] *Laboratoire Jean Kuntzmann, UMR CNRS 5224, Univ Grenoble-Alpes, France*[3]

Abstract. In this paper, we study an estimation problem where the variables of interest are subject to both right censoring and measurement error. In this context, we propose a nonparametric estimation strategy of the hazard rate, based on a regression contrast minimized in a finite dimensional functional space generated by splines bases. We prove a risk bound of the estimator in term of integrated mean square error, discuss the rate of convergence when the dimension of the projection space is adequately chosen. Then, we define a data driven criterion of model selection and prove that the resulting estimator performs an adequate compromise. The method is illustrated via simulation experiments which prove that the strategy is successful. March 15, 2016

Keywords. Censored data; Measurement error; Nonparametric methods; Deconvolution; B-spline.

AMS Subject Classification 2010: Primary 62G05; secondary 62N01.

1. Introduction

In this paper, we consider a sample $(Y_j, \delta_j)_{1\leq j\leq n}$ of i.i.d. observations from the model

$$(1)\qquad \begin{aligned} Y_j &= (X_j \wedge C_j) + \varepsilon_j = (X_j + \varepsilon_j) \wedge (C_j + \varepsilon_j), \quad j = 1, \dots, n \\ \delta_j &= \mathbb{1}_{X_j \leq C_j}, \end{aligned}$$

where the X_js are the variables of interest. All variables X_j, C_j (right-censoring variable), ε_j (noise or measurement error variable) are independent and identically distributed (i.i.d.) and the three sequences are independent. The density f_ε of the noise is assumed to be known, the X_j and C_j are nonnegative random variables. In other words, the data are censored and measured with error. Right censoring corresponds to observations $U_j = X_j \wedge C_j$ and δ_j and measurement errors to observations $X_j + \varepsilon_j$. Obviously, $\mathbb{1}_{X_j\leq C_j} = \mathbb{1}_{X_j+\varepsilon_j\leq C_j+\varepsilon_j}$ so that the censoring indicator is unchanged by the measurement error.

The aim of the present work is to propose an estimation strategy for h_X, the hazard rate of X, defined by f_X/S_X where f_X is the density and S_X the survival function of X.

The literature has considered the two problems of measurement error and censoring mostly separately. On the one hand, deconvolution strategies have been developed for density estimation in presence of measurement error by Stefanski and Carroll (1990), Fan (1991) (kernels), Delaigle and Gijbels (2004) (bandwidth selection), Pensky and Vidakovic (1999); Fan and Koo (2002) (wavelet estimators), Comte et al. (2006) (projection methods with model selection). Cumulative distribution

[1] fabienne.comte@parisdescartes.fr
[2] gwennaelle.mabon@ensae.fr
[3] adeline.leclercq-samson@imag.fr

function estimation in presence of measurement error is considered by Dattner et al. (2011). On the other hand, nonparametric or hazard rate estimation for right censored data has been studied in Antoniadis et al. (1999), Li (2007, 2008) (wavelets), Dohler and Ruschendorf (2002) (penalized likelihood-based criterion), Brunel and Comte (2005, 2008), Reynaud-Bouret (2006), Akakpo and Durot (2010) (penalized contrast estimators).

The only reference treating both the censor and the measurement noise is Comte et al. (2015), where time between the onset of pregnancy and natural childbirth were to be estimated, from ultrasound data subject to both right censoring and measurement error. The method proposed therein is based on deconvolution strategies providing estimators of $f_X S_C$ and $S_X S_C$ where S_C denotes the survival function of C; then taking the ratio of these quantities delivers an estimate of h_X. The method is attractive, but theoretical rates are degraded by the survival step estimation. Moreover, the computation of an estimator built as a ratio is generally delicate and potentially unstable, as the denominator can get too small. This is why we explore here a one-step regression-type method. But the problem is difficult due to the double source of errors (censoring and measurement error). Here, we mainly propose an extension of the regression strategy developed in Comte et al. (2011) and Plancade (2011), relying on a projection contrast method, associated with spline spaces.

The paper is organized as follows. In Section 2, we state the assumptions. The estimator is then defined as a contrast minimizer associated with a contrast which is justified and to splines projection spaces which are described. Risk bounds in term of integrated mean squared error are given in Section 3, discussion about convergence rates are then provided. A model selection procedure is finally proposed and proved to perform an adequate compromise, at least in theory. For more practical aspects, simulation results are provided in Section 4, which allow to compare the performances of the present procedure with those of a previous quotient estimator. Proofs are gathered in Section 5.

2. Estimation procedure

2.1. **Notation and assumptions.**

We denote by f_U the density of a variable U, by $S_U(t) = \mathbb{P}(U \geq t)$ the survival function at point t of a random variable U and by $h_U(t) = f_U(t)/S_U(t)$ the hazard ratio at point t. The characteristic function of U is $f_U^*(t) = \mathbb{E}(e^{itU})$. We denote by $g^*(t) = \int e^{itx} g(x)\,\mathrm{d}x$ the Fourier transform of any integrable function g. For a function $g : \mathbb{R} \mapsto \mathbb{R}$, we denote $\|g\|^2 = \int_{\mathbb{R}} g^2(u)\,\mathrm{d}u$ the $\mathbb{L}^2$ norm. For two integrable and square-integrable functions g and h, we denote $g \star h$ the convolution product $g \star h(x) = \int g(x-u)h(u)\,\mathrm{d}u$. For two real numbers a and b, we denote $a \wedge b = \min(a, b)$.

Let us give the assumptions on the noise ε. We assume that the characteristic function of the noise is known and such that

$$\forall u \in \mathbb{R}, \quad f_\varepsilon^*(u) \neq 0.$$

Moreover, to be able to propose sets of functions satisfying several constraints, we restrict ourselves to the case of ordinary smooth errors, i.e. we assume that f_ε is such that

$$|f_\varepsilon^*(u)|^{-1} \sim \mathfrak{C}_\varepsilon (1 + u^2)^{\alpha/2}. \tag{2}$$

This condition allows to consider Laplace or Gamma distributions, but not Cauchy nor Gaussian.

On the other hand, for the variables X and C, we assume that the following assumption is fulfilled:
Assumption (A1). We assume that both X and C are nonnegative random variables and $\mathbb{E}(X) < +\infty$, $\mathbb{E}(C) < +\infty$.

Moreover, for A the compact set on which the function is estimated, we assume:
Assumption (A2). There exists a constant S_0 such that, $\forall x \in A, S_{X\wedge C}(x) \geq S_0 > 0$.

2.2. **Definition of the contrast.** From Model (1), we observe for $j = 1, \dots, n$ both Y_j and δ_j and we want to estimate the hazard rate h_X of X.

Under **(A1)**, $S_{X\wedge C}$ is integrable and square integrable on its support $\mathbb{R}^+$, so that $S^*_{X\wedge C}(u) = \int_0^{+\infty} e^{iuv} S_{X\wedge C}(v)dv$ is well defined. Then, the following Lemma, proved in Comte et al. (2015), holds and defines an estimator of $S^*_{X\wedge C}$, useful for the sequel.

Lemma 2.1. *Assume that* **(A1)** *holds and let* $(Y_j, \delta_j)_{1\le j\le n}$ *be observations drawn from model (1). Then the estimator* $\widehat{S}^*_{X\wedge C}$ *defined by*

$$\widehat{S}^*_{X\wedge C}(u) = \frac{1}{n}\frac{1}{iu}\sum_{j=1}^{n}\left(\frac{e^{iuY_j}}{f^*_\varepsilon(u)} - 1\right) \tag{3}$$

is an unbiased estimator of $S^*_{X\wedge C}(u)$.

Now, let t be a function such that t and t^2 are integrable, as well as t^*/f^*_ε and $(t^2)^*/f^*_\varepsilon$. We consider the following contrast

$$\gamma_n(t) = \frac{1}{2\pi}\left\{\int (t^2)^*(u)\widehat{S}^*_{X\wedge C}(-u)\,\mathrm{d}u - 2\int \frac{1}{n}\left(\sum_{j=1}^{n}\delta_j e^{iuY_j}\right)\frac{t^*(-u)}{f^*_\varepsilon(u)}\,\mathrm{d}u\right\}. \tag{4}$$

To understand why this proposal is relevant, let us compute the expectation of $\gamma_n(t)$, which, by the law of Large Numbers, coincides with its almost sure limit. First, notice that

$$\begin{aligned}\mathbb{E}\left[\delta_1 e^{iuY_1}\right] &= \mathbb{E}\left[\mathbb{1}_{X_1\le C_1} e^{iu(X_1\wedge C_1)} e^{iu\varepsilon_1}\right] = \mathbb{E}\left[\mathbb{1}_{X_1\le C_1} e^{iuX_1}\right] f^*_\varepsilon(u)\\ &= \mathbb{E}\left[S_C(X_1)e^{iuX_1}\right] f^*_\varepsilon(u) = (f_X S_C)^*(u) f^*_\varepsilon(u).\end{aligned} \tag{5}$$

Then, we can easily deduce that, by Parseval formula

$$\mathbb{E}\left[\frac{\delta_j}{2\pi}\int e^{iuY_j}\frac{t^*(-u)}{f^*_\varepsilon(u)}\,\mathrm{d}u\right] = \frac{1}{2\pi}\int (f_X S_C)^*(u)t^*(-u)\,\mathrm{d}u = \int t(x)S_C(x)f_X(x)\,\mathrm{d}x. \tag{6}$$

Moreover, using that by Lemma 2.1, $\mathbb{E}(\widehat{S}^*_{X\wedge C}(u)) = S^*_{X\wedge C}(u)$, we get

$$\mathbb{E}[\gamma_n(t)] = \frac{1}{2\pi}\int (t^2)^*(u)S^*_{X\wedge C}(-u)\,\mathrm{d}u - \frac{1}{\pi}\int (f_X S_C)^*(u)t^*(-u)\,\mathrm{d}u.$$

By Parseval formula, this yields

$$\mathbb{E}[\gamma_n(t)] = \langle t^2, S_{X\wedge C}\rangle - 2\langle t, f_X S_C\rangle.$$

Lastly, using $h_X = f_X/S_X$ and $S_{X\wedge C} = S_C S_X$, we obtain

$$\mathbb{E}[\gamma_n(t)] = \int (t^2(x) - 2t(x)h_X(x))S_{X\wedge C}(x)dx.$$

Therefore we get, if t is A-supported,

$$\mathbb{E}[\gamma_n(t)] = \int_A (t(x) - h_X(x))^2 S_{X\wedge C}(x)\,\mathrm{d}x - \int_A h^2_X(x)S_{X\wedge C}(x)\,\mathrm{d}x.$$

This is why, for large n, minimizing $\gamma_n(t)$ among an adequate collection of functions t amounts to minimize the term $\int_A (t(x) - h_X(x))^2 S_{X\wedge C}(x)\,\mathrm{d}x$, and brings an estimator of h_X. This contrast is new; it is of regression type and corresponds to a nontrivial generalization of the one studied in Comte et al. (2011) and Plancade (2011).

2.3. **B-Splines spaces of estimation.** As noted above, the contrast $\gamma_n(t)$ is well defined for functions t such that t and t^2 are integrable, as well as t^*/f_ε^* and $(t^2)^*/f_\varepsilon^*$. This is why we have to define our projection estimator on finite dimensional functional spaces generated by functions having very specific properties. Moreover, for technical reasons, the regression type of the contrast requires these functions to be compactly supported. This is the reason why we restrict to ordinary smooth error distributions given by (2). For simplicity, we set $A = [0, 1]$.

Now, we specify the spaces $\mathcal{S}_m$ of functions t among which we minimize the contrast. We choose the splines projection spaces, which verify the main properties needed for the estimation procedure. We consider dyadic B-splines on the unit interval $[0, 1]$. Let N_r be the B-spline of order r that corresponds to the r-times iterated convolution of $\mathbb{1}_{[0,1]}(x)$ and has knots at the points $0, 1, \ldots, r$. Using difference notation from de Boor (2001) or DeVore and Lorentz (1993), N_r is also defined as $N_r(x) = r[0, 1, \ldots, r](\,.\, - x)_+^{r-1}$. Let m be a positive integer and define,

$$\varphi_{m,k}(x) = 2^{m/2} N_r(2^m x - k), \quad k \in \mathbb{Z}.$$

Note that $\varphi_{m,k}$ has only non zero values on $]k/2^m, (k + r)/2^m]$.

For approximation on $[0, 1]$, one usually considers the B-splines $\varphi_{m,k}$ which do not vanish identically on $[0, 1]$. Let $\bar{\mathbb{K}}_m$ denote the set of integers k for which this holds, $\bar{\mathbb{K}}_m = \{-(r-1), \cdots -1, 0, 1, \ldots, 2^m - 1\}$ as N_r has support $[0; r]$. Let $\mathbb{K}_m$ denote the subset of $\bar{\mathbb{K}}_m$ of integers k such that the support of $\varphi_{m,k}$ is included in $[0, 1]$, i.e. $\mathbb{K}_m = \{0, 1, \ldots, 2^m - (r - 1)\}$.

We now define $\mathcal{S}_m$ as the linear span of the B-splines $\varphi_{m,k}$ for $k \in \mathbb{K}_m$. The linear space $\mathcal{S}_m$ is referred to as the space of dyadic splines, its dimension is $2^m - (r - 1)$ and any element t of $\mathcal{S}_m$ can be represented as

$$t(x) = \sum_{k \in \mathbb{K}_m} a_{m,k} \varphi_{m,k}$$

for a vector $\vec{\mathbf{a}}_m = (a_{m,k})_{k \in \mathbb{K}_m}$ with $2^m - (r - 1)$ coordinates. Note that the usual dyadic splines space is generated by the $2^m + (r - 1)$ functions corresponding to $k \in \bar{\mathbb{K}}_m$, and our restriction to $\mathbb{K}_m$ has consequences on the bias order.

The following properties of the splines are useful. For any $t \in \mathcal{S}_m$, $\|t\|_\infty \le \Phi_0 2^{m/2} \|t\|$. Moreover, there exists some constant Φ_0 such that

$$\Phi_0^{-2} \sum_{k \in \mathbb{K}_m} a_{m,k}^2 \le \left\| \sum_{k \in \mathbb{K}_m} a_{m,k} \varphi_{m,k} \right\|^2 \le \Phi_0^2 \sum_{k \in \mathbb{K}_m} a_{m,k}^2. \tag{7}$$

In the sequel we will use the following property of φ derived from their r-regularity:

$$\forall u \in \mathbb{R} \quad |N_r^*(u)| = \left| \frac{\sin(u/2)}{(u/2)} \right|^r \text{ and } |\varphi_{m,k}^*(u)| = 2^{-m/2} \left| \frac{\sin(u/2^{m+1})}{(u/2^{m+1})} \right|^r. \tag{8}$$

We denote by $\mathcal{S}$ the space S_{m_n} with the greatest dimension smaller than $n^{1/(2\alpha+1)}$. It is the maximal space that we will consider. In the following, we set $D_m = 2^m$.

2.4. **Minimizer of γ_n for the B-spline basis.** We can rewrite the contrast $\gamma_n(t)$ for $t = \sum_{k \in \mathbb{K}_m} a_{m,k} \varphi_{m,k}$ and $\vec{\mathbf{a}}_m$ denoting the vector of the coefficients of a function t in the basis, as

$$2\pi \gamma_n(t) = {}^t\vec{\mathbf{a}}_m \mathbf{G}_m \vec{\mathbf{a}}_m - 2\, {}^t\vec{\mathbf{a}}_m \Upsilon_m,$$

where

$$\mathbf{G}_m = \left(\int (\varphi_{m,j} \varphi_{m,k})^*(u) \widehat{S}_{X \wedge C}^*(-u) \, \mathrm{d}u \right)_{j,k \in \mathbb{K}_m}, \; \Upsilon_m = \left(\int \varphi_{m,k}^*(-u) \frac{\frac{1}{n} \sum_{j=1}^n \delta_j e^{iuY_j}}{f_\varepsilon^*(u)} \, \mathrm{d}u \right)_{k \in \mathbb{K}_m}$$

with $\widehat{S}_{X \wedge C}^*$ given by (3). Note that the matrix $\mathbf{G}_m$ is band-diagonal with bands of length $r - 1$ on each side of the diagonal (thus a symmetric diagonal band with global length $2r - 1$).

Then we define $\hat{h}_m$ the estimator of the hazard rate as the minimizer of $\gamma_n(t)$ over all functions $t \in \mathcal{S}_m$ whenever it is meaningful. For that, we derive $\gamma_n(t)$ to find $t = \sum_{k\in\mathbb{K}_m} a_{m,k}\varphi_{m,k}$ such that

$$\forall k_0 \in \mathbb{K}_m, \quad \frac{\partial \gamma_n}{\partial a_{m,k_0}}(t) = 0.$$

This yields the following system for the estimator $\hat{\vec{\mathbf{a}}}_m = (\hat{a}_{m,k})_{k\in\mathbb{K}_m}$ of the vector of coefficients $\vec{\mathbf{a}}_m$

$$\mathbf{G}_m \hat{\vec{\mathbf{a}}}_m = \Upsilon_m$$

Then, under **(A2)**, we set $\Gamma_m = \{\min \mathrm{sp}(\mathbf{G}_m) \geq (2/3)S_0\}$ where $\mathrm{sp}(B)$ denotes the set of the eigenvalues of a square matrix B. Clearly, $\mathbf{G}_m^{-1}$ exists on Γ_m. Finally, we consider the estimator

$$\hat{h}_m = \begin{cases} \arg\min_{t\in\mathcal{S}_m} \gamma_n(t) & \text{on} \quad \Gamma_m \\ 0 \quad \text{otherwise.} \end{cases} \tag{9}$$

Note that, as usual for mean square estimators, we have, on Γ_m, $\hat{\vec{\mathbf{a}}}_m = \mathbf{G}_m^{-1}\Upsilon_m$ and

$$\gamma_n(\hat{h}_m) = -\frac{1}{2\pi}\, {}^t\Upsilon_m \mathbf{G}_m^{-1}\Upsilon_m. \tag{10}$$

Formula (10) allows an easy computation of the contrast, and this is useful for the sequel.

3. Risk bound and model selection

3.1. **Risk bound and discussion about the rate.** First let us study the $\mathbb{L}_2$-risk of our estimator $\hat{h}_m$ defined by Equation (9). For this purpose, we introduce the norm $\|f\|_A^2 = \int_A f^2(x)\,\mathrm{d}x$. Let h_m be the projection of h_X on the space generated by the $\varphi_{m,k}$ for $k \in \mathbb{K}_m$. We can prove the following result.

Proposition 3.1. *Assume that f_ε satisfies (2) with regularity $\alpha > 1/2$ and the Y_j's admit moments of order 2. Assume also that* **(A1)** *and* **(A2)** *hold, that $r > \alpha + 1$and that f_Y, $S_C f_X$ and h_X are bounded on A. Let $\mathcal{M}_n = \{m, D_m^{2\alpha+1} \leq n\}$ and $\hat{h}_m$ be defined by (9). Then there exist constants $\mathfrak{C}_1, \mathfrak{C}_2, \mathfrak{C}_3$ such that, $\forall m \in \mathcal{M}_n$,*

$$\mathbb{E}\|h_X - \hat{h}_m\|_A^2 \leq \mathfrak{C}_1 \left(\|h_X - h_m\|_A^2 + \mathfrak{C}_{2,1}\frac{D_m^{2\alpha+1}}{n} + \mathfrak{C}_{2,2}\frac{D_m^{(2\alpha)\vee 1}}{n} + \mathfrak{C}_{2,3}\frac{D_m}{n} \right) + \frac{\mathfrak{C}_3}{n} \tag{11}$$

$$\leq \mathfrak{C}_1 \left(\|h_X - h_m\|_A^2 + \mathfrak{C}_2 \frac{D_m^{2\alpha+1}}{n} \right) + \frac{\mathfrak{C}_3}{n} \tag{12}$$

where $\mathfrak{C}_1 = \mathfrak{C}_1(S_0)$, $\mathfrak{C}_3 = \mathfrak{C}_3(\|h_X\|_A)$ are two constants and

$$\mathfrak{C}_2 = \mathfrak{C}_{2,1} + \mathfrak{C}_{2,2} + \mathfrak{C}_{2,2}. \tag{13}$$

with

$$2\pi\mathfrak{C}_{2,1} = \Phi_0^2 \mathfrak{C}_\varepsilon^2 \|S_C f_X\|_\infty I_1 + (2r-1)\Phi_0^4 \mathfrak{C}_\varepsilon^2 \|h_X\|_A^2 I_2/(2\pi)^2$$

$$2\pi\mathfrak{C}_{2,2} = 2(2r-1)\Phi_0^4\mathfrak{C}_\varepsilon^2 \; \|h_X\|_A^2 \|f_Y\|_\infty/(2\pi), \quad 2\pi\mathfrak{C}_{2,3} = 2^\alpha(2r-1)\Phi_0^4\mathfrak{C}_\varepsilon^2 \|h_X\|_A^2 \mathbb{E}[Y_1^2]/(2\pi),$$

where $I_1 = \int |N_r^*(z)|^2(1+z^2)^\alpha\,\mathrm{d}z$, $I_2 = \left(\int_{|v|>1} |v|^{-r}(1+v^2)^{\alpha/2}\,\mathrm{d}v\right)^2$.

The first term $\|h - h_m\|_A^2$ is the squared bias. The second term $\mathfrak{C}_2 D_m^{2\alpha+1}/n$ is the variance term. Inequality (12) gives the bias-variance decomposition, up to the negligible term $\mathfrak{C}_3/n$.

Comment about $\mathfrak{C}_2$. We summarize the variance as $\mathfrak{C}_2 D_m^{2\alpha+1}/n$ in (12), but it results in fact of three contributions detailed in (11). The first term is really of order $D_m^{2\alpha+1}$, the second one has order $D_m^{2\alpha\vee 1}$ and the last one is of order D_m.

Discussion about the convergence rate. First, let us recall that a function f belongs to the Besov space $\mathcal{B}_{\beta,\ell,\infty}([0,1])$ if it satisfies

$$|f|_{\beta,\ell} = \sup_{y>0} y^{-\beta} w_d(f,y)_\ell < +\infty, \quad d = [\beta]+1, \tag{14}$$

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 $\mathcal{B}_{\beta,p,\infty}([0,1]) \subset \mathcal{B}_{\beta,2,\infty}([0,1])$ for $p \geq 2$. This justifies that we now restrict our attention to $\mathcal{B}_{\beta,2,\infty}([0,1])$. It follows from Theorem 3.3 in Chapter 12 of DeVore and Lorentz (1993) that $\|h_X - \bar{h}_m\|^2 = \mathrm{O}(D_m^{-2\beta})$, if h belongs to some Besov space $B_{\beta,2,\infty}([0,1])$ with $|h|_{\beta,2} \leq L$ for some fixed L and $\bar{h}_m$ is the projection of h_X on the space generated by the $\varphi_{m,k}$ for $k \in \bar{\mathbb{K}}_m$.

Consequently, under Besov regularity assumptions with regularity index β, we know that $\|h_X - \bar{h}_m\|^2 = \mathrm{O}(D_m^{-2\beta})$, which implies, for a choice $D_{m_{opt}} = \mathrm{O}(n^{1/(2\alpha+2\beta+1)})$, and if $\|h_X - \bar{h}_m\|^2 \asymp \|h_X - h_m\|^2$, a rate

$$\mathbb{E}\|h_X - \hat{h}_{m_{opt}}\|^2 = \mathrm{O}\left(n^{-2\beta/(2\alpha+2\beta+1)}\right)$$

Note that this rate is the optimal rate for density estimation in presence of noise satisfying (2) without censoring, therefore, it is likely to be optimal here as well [REF ??].

However, it is not straightforward to evaluate the distance between $\|h_X - \bar{h}_m\|^2$ and $\|h_X - h_m\|^2$. But this difference is likely to be small in practice. This also justify why we use $\mathbb{K}_m$ in the implementation of the estimator.

3.2. **Model selection.**

The aim of this section is to provide a data-driven estimator of the hazard rate h_X with a $\mathbb{L}_2$-risk as close as possible to the oracle risk defined by $\inf_m \|h_X - \hat{h}_m\|^2$. Thus we follow the model selection paradigm introduced by Birgé and Massart (1997), Birgé (1999), Massart (2003) which yields to a choice of the dimension of projection space m according to a penalized criterion.

To obtain the data driven selection of m, we propose the following additional step. We select

$$\widehat{m} = \arg\min_{m\in\mathcal{M}_n} \left\{\gamma_n(\hat{h}_m) + \mathrm{pen}(m)\right\} \quad \text{with} \quad \mathrm{pen}(m) = \kappa \frac{\mathfrak{C}_2 D_m^{2\alpha+1}}{n} \tag{15}$$

where $\mathfrak{C}_2$ is given by (13) and κ is a numerical constant. Then we consider the estimator $\tilde{h} = \hat{h}_{\widehat{m}}$. We recall that $\gamma_n(\hat{h}_m)$ is given by (10) and thus easy to compute. Then we can prove the following result.

Theorem 3.2. *Assume that the assumptions of Proposition 3.1 hold and the Y_j's admit moments of order 10 and consider the estimator $\hat{h}_{\widehat{m}}$ with $\widehat{m}$ given by (15). Then there exists κ_0 such that for all $\kappa \geq \kappa_0$,*

$$\mathbb{E}\|h_X - \tilde{h}\|_A^2 \leq \mathfrak{C}_1 \inf_{m\in\mathcal{M}_n} \left\{\|h_X - h_m\|_A^2 + \mathrm{pen}(m)\right\} + \frac{\mathfrak{C}_3}{n}, \tag{16}$$

where $\mathfrak{C}_1 = \mathfrak{C}_1(S_0)$ and $\mathfrak{C}_3 = \mathfrak{C}_3(\|h_X\|_A)$ are positive constants.

The first term is the squared bias. The second term $\mathrm{pen}(m)$ is of order of the variance term. The oracle inequality is achieved up to the negligible term $\mathfrak{C}_3/n$.

Many terms are known in $\mathfrak{C}_2$, but there are also unknown terms namely $\|S_C f_X\|_\infty$, $\|h_X\|_A^2$, $\|f_Y\|_\infty$ and $\mathbb{E}[Y_1^2]$. In practice, they are replaced by estimators.

4. Illustrations

The whole implementation is conducted using R software. The integrated squared errors $\|h_X - \tilde{h}\|_A^2$ are computed via a standard approximation and discretization (over 300 points) of the integral on an interval A. Then the mean integrated squared errors (MISE) $\mathbb{E}\|h_X - \tilde{h}\|_A^2$ are computed as the empirical mean of the approximated ISE over 500 simulation samples.

4.1. **Simulation setting.** The performance of the procedure is studied for the four following distributions for X. All the densities are normalized with unit variance.

- ▷ Mixed Gamma distribution: $X = W/\sqrt{5.48}$ with $W \sim 0.4\Gamma(5,1) + 0.6\Gamma(13,1)$, $A = [0, 5.5]$.
- ▷ Beta distribution: $X \sim \mathcal{B}(2,5)/\sqrt{0.025}$, $A = [0, 2.5]$.
- ▷ Gaussian distribution: $X \sim \mathcal{N}(5,1)$, $A = [0, 5.5]$.
- ▷ Gamma distribution: $X \sim \Gamma(5,1)/\sqrt{5}$, $A = [0, 2.5]$.

Data are simulated with a Laplace noise with variance $\sigma^2 = 2b^2$ as follows:

$$f_\varepsilon(x) = \frac{1}{2b} 2e^{-|x|/b} \quad \text{and} \quad f_\varepsilon^*(x) = \frac{1}{1 + b^2x^2}$$

Since the four target densities X are normalized with unit variance, it allows the ratio $1/\sigma^2$ to represent the signal-to-noise ratio, denoted $s2n$. We consider signal to noise ratios of $s2n = 2.5$ and $s2n = 10$ in the simulations which means that $b = 1/(2\sqrt{5})$ and $b = 1/(\sqrt{5})$.

The censoring variable C is simulated with an exponential distribution, with parameter λ chosen to ensure 20% or 40% of censored variables. We consider samples of size $n = 400$ and 1000.

We choose the same design of simulation as in Comte et al. (2015) in order to compare our procedure with theirs.

4.2. **Practical estimation procedure.** The adaptive procedure is implemented as follows:

- ▷ For $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$, compute $\gamma_n(\hat{h}_m) + \mathrm{pen}(m)$.
- ▷ Choose $\widehat{m}$ such that $\widehat{m} = \arg\min_{m \in \mathcal{M}_n} \left\{ \gamma_n(\hat{h}_m) + \mathrm{pen}(m) \right\}$.
- ▷ And compute $\tilde{h}(x) = \sum_{k \in \mathbb{K}_{\widehat{m}}} \hat{a}_{\widehat{m},k} \varphi_{\widehat{m},k}(x)$.

Gathering (10) and (15), our procedure consists in computing

$$\widehat{m} = \arg \min_{m \in \mathcal{M}_n} -{}^t\Upsilon_m \mathbf{G}_m^{-1} \Upsilon_m + \kappa \mathfrak{C}_3 \frac{D_m^{2\alpha+1}}{n}$$

with $\kappa = 0.5$ and $\mathcal{M}_n = \left\{ m \in \mathbb{N}, m \le n^{1/5}/\log 2 \right\}$.

When there is no noise (NN), our procedure reduces to Plancade (2011)'s, and $\mathbf{G}_m$ and Υ_m become more simply $\mathbf{G}_m^{\mathrm{NN}}$ and Υ_m^{NN} defined by

$$\mathbf{G}_m^{\mathrm{NN}} = \left(\frac{1}{n} \sum_{j=1}^n \int_A \varphi_{m,k}(u) \varphi_{m,k'}(u) \mathbb{1}_{\{u \le Y_j\}} \, \mathrm{d}u \right)_{k,k' \in \mathbb{K}_m}, \quad \Upsilon_m^{\mathrm{NN}} = \left(\frac{1}{n} \sum_{j=1}^n \delta_j \varphi_{m,k}(Y_j) \right)_{k \in \mathbb{K}_m}$$

where here $Y_j = X_j \wedge C_j$. The selection is performed via

$$\widehat{m}^{\mathrm{NN}} = \arg \min_{m \in \mathcal{M}_n^{\mathrm{NN}}} -{}^t\Upsilon_m^{\mathrm{NN}} (\mathbf{G}_m^{\mathrm{NN}})^{-1} \Upsilon_m^{\mathrm{NN}} + \kappa \|h_X\|_\infty \frac{D_m}{n},$$

with $\mathcal{M}_n^{\mathrm{NN}} = \{m \in \mathbb{N}, m \le n/\log 2\}$. Plancade (2011) shows that $\kappa > 1$ is a theoretical suitable value for the adaptive procedure. Since we do not use the same basis as the author we take $\kappa = 5$. Moreover, as in Plancade (2011) for sake of simplicity, we do not estimate $\|h_X\|_\infty$ and suppose that this quantity is known. Some preliminary numerical studies show that replacing it by its estimator hardly affects the result. Thus in our procedure we do not estimate $\|h_X\|_\infty$ either.

To evaluate the quality of the estimators, we approximate via Monte Carlo repetitions the minimal risk of the collection $(\hat{h}_m)$ (risk of the oracle), the quadratic risk of $\hat{h}_{\hat{m}}$, and the risk of the quotient estimator. More precisely, we denote $\hat{r}_{or}$ and $\hat{r}_{ad}$:

$$\hat{r}_{or} = \min_{m\in\mathcal{M}_n} \widehat{\mathbb{E}}\|h_X - \hat{h}_m\|_A^2 \quad \text{and} \quad \hat{r}_{ad} = \widehat{\mathbb{E}}\|h_X - \hat{h}_{\widehat{m}}\|_A^2 \tag{17}$$

where $\widehat{\mathbb{E}}$ is the approximation of theoretical expectation computed via Monte Carlo repetitions. We denote $\hat{r}_{quot}$ the estimated risk of the quotient estimator obtained by Comte et al. (2015).

4.3. **Simulation results.** The results of the simulations are given in Table 1. In the Table we report estimations of MISE when the data are not censored, or not noisy, or neither censored nor noisy. First we see that the risk decreases when the sample size increases. Likewise the risk increases when the variance and the censoring increase.

We can see that for the mixed Gamma distribution our results are less good than those of Comte et al. (2015). But we can notice that for a mixed distribution, our results are similar to those of Plancade (2011). However when the censoring level is high and the sample size small we obtain a better result. For the Gaussian distribution, the results of the two methods are equivalent. Our method improves the estimation for the Beta and Gamma distributions. Globally, we get better results than Comte et al. (2015) in 65% of the cases, which is good performance.

Note that our procedure requires only one constant calibration whereas Comte et al. (2015) use an estimator which is a quotient and choose independently the two dimensions and thus two constant calibrations.

5. Proofs

5.1. **About the spline basis.** Any function $t \in \mathcal{S}_m$ can be decomposed as $t(x) = \sum_{k\in\mathbb{K}_m} a_{m,k}\varphi_{m,k}(x)$. Moreover, when $\|t\| = 1$, we have $\sum_k a_{m,k}^2 \le \Phi_0$ from (7). The following relations will be used in the proofs.

Lemma 5.1.

1) For all $j = J, \dots, m$ and $k \in \Lambda_j$, we have

$$\varphi_{m,k}(\cdot) \quad = \quad 2^{m/2} N_r(2^m \cdot -k), \quad \varphi_{m,k}^*(u) = 2^{-m/2} e^{iuk/2^m} N_r^*\left(\frac{u}{2^m}\right).$$

2) For any m and any k, k', Lemma 4 from Lacour (2008) yields

$$(2\pi)|(\varphi_{m,k}\varphi_{m,k'})^*(u)| = |\varphi_{m,k}^* \star \varphi_{m,k'}^*(u)| \le \begin{cases} |2^{-m}u|^{1-r} & \text{if} \quad |u| > 2^m \\ 1 & \text{if} \quad |u| \le 2^m. \end{cases}$$

5.2. **Proof of Proposition 3.1.** Let us define

$$\Psi_n(t) = \frac{1}{2\pi}\int (t^2)^*(u)\widehat{S}_{X\wedge C}^*(u)\,\mathrm{d}u,$$

which is such that, for $t \in \mathcal{S}_m$,

$$\mathbb{E}\left[\Psi_n(t)\right] = \int_A t^2(x) S_{X\wedge C}(x)\,\mathrm{d}x := \|t\|_{S_{X\wedge C}}^2.$$

$s2n = \infty$		0% censoring		20% censoring		40% censoring	
f_X		$n=400$	$n=1000$	$n=400$	$n=1000$	$n=400$	$n=1000$
Mixed	$\hat{r}_{or}$	*0.49*	*0.19*	*0.58*	*0.27*	*0.71*	*0.37*
Gamma	$\hat{r}_{ad}$	1.21	0.49	1.24	0.58	**1.52**	0.72
	$\hat{r}_{quot}$	**0.82**	**0.29**	**1.15**	**0.38**	2.05	**0.67**
Beta	$\hat{r}_{or}$	*0.53*	*0.30*	*0.54*	*0.31*	*0.60*	*0.33*
	$\hat{r}_{ad}$	**0.78**	**0.54**	**0.77**	**0.53**	**0.85**	**0.54**
	$\hat{r}_{quot}$	1.37	0.66	1.83	0.79	2.40	1.05
Gaussian	$\hat{r}_{or}$	*0.28*	*0.13*	*0.42*	*0.24*	*1.06*	*0.52*
	$\hat{r}_{ad}$	0.63	0.27	**0.81**	**0.49**	**1.95**	**0.97**
	$\hat{r}_{quot}$	0.61	**0.19**	1.97	0.57	8.64	5.87
Gamma	$\hat{r}_{or}$	*0.38*	*0.18*	*0.48*	*0.24*	*0.53*	*0.26*
	$\hat{r}_{ad}$	**0.49**	**0.26**	**0.75**	0.31	**0.82**	0.35
	$\hat{r}_{quot}$	0.64	0.28	0.84	**0.29**	1.04	**0.32**
$s2n = 10$		0% censoring		20% censoring		40% censoring	
f_X		$n=400$	$n=1000$	$n=400$	$n=1000$	$n=400$	$n=1000$
Mixed	$\hat{r}_{or}$	*0.50*	*0.32*	*0.56*	*0.35*	*0.65*	*0.41*
Gamma	$\hat{r}_{ad}$	1.04	0.90	1.06	0.89	**1.13**	0.91
	$\hat{r}_{quot}$	**0.72**	**0.30**	**0.94**	**0.43**	1.34	**0.75**
Beta	$\hat{r}_{or}$	*0.99*	*0.44*	*1.09*	*0.47*	*1.22*	*0.54*
	$\hat{r}_{ad}$	**1.13**	**0.44**	**1.25**	**0.49**	**1.40**	**0.57**
	$\hat{r}_{quot}$	1.49	0.82	2.00	1.08	2.64	1.03
Gaussian	$\hat{r}_{or}$	*0.35*	*0.20*	*0.70*	*0.29*	*2.14*	*0.69*
	$\hat{r}_{ad}$	**0.48**	0.28	**0.90**	**0.41**	**2.47**	**0.96**
	$\hat{r}_{quot}$	0.56	**0.24**	1.34	0.71	7.39	5.87
Gamma	$\hat{r}_{or}$	*0.60*	*0.29*	*0.63*	*0.30*	*0.70*	*0.31*
	$\hat{r}_{ad}$	**0.65**	**0.33**	**0.66**	**0.34**	**0.72**	**0.35**
	$\hat{r}_{quot}$	0.78	0.37	0.89	0.39	0.98	1.02
$s2n = 2.5$		0% censoring		20% censoring		40% censoring	
f_X		$n=400$	$n=1000$	$n=400$	$n=1000$	$n=400$	$n=1000$
Mixed	$\hat{r}_{or}$	*0.59*	*0.37*	*0.67*	*0.41*	*0.80*	*0.47*
Gamma	$\hat{r}_{ad}$	**1.08**	0.92	**1.12**	0.91	**1.23**	**0.94**
	$\hat{r}_{quot}$	1.15	**0.48**	1.37	**0.68**	1.93	1.04
Beta	$\hat{r}_{or}$	*2.13*	*1.04*	*2.44*	*1.05*	*3.16*	*1.40*
	$\hat{r}_{ad}$	2.87	1.81	4.36	**1.80**	**4.66**	**2.14**
	$\hat{r}_{quot}$	**2.05**	**1.14**	**3.98**	1.92	5.72	2.57
Gaussian	$\hat{r}_{or}$	*0.61*	*0.27*	*1.87*	*0.50*	*6.04*	*1.69*
	$\hat{r}_{ad}$	0.91	**0.42**	2.54	**0.73**	**7.30**	**2.88**
	$\hat{r}_{quot}$	**0.86**	0.44	**2.15**	1.06	8.04	6.27
Gamma	$\hat{r}_{or}$	*1.17*	*0.46*	*1.23*	*0.51*	*1.44*	*0.53*
	$\hat{r}_{ad}$	1.67	0.62	**1.41**	**0.61**	**1.75**	**0.88**
	$\hat{r}_{quot}$	**1.32**	0.63	1.77	0.87	2.31	1.02

TABLE 1. MISE×100 of the estimation of h_X, compared with the MISE obtained when data are not censored, or not noisy, or neither censored nor noisy. MISE was averaged over 500 samples (1000 for r_{quot}). Data are simulated with a Laplace noise, and an exponential censoring variable.

Let us also consider the sets

$$\Delta_m = \left\{\omega \ / \ \forall t \in \mathcal{S}_m,\ \|t\|^2_{S_{X\wedge C}} \leq \frac{3}{2}\Psi_n(t)\right\}, \quad \Delta = \bigcap_{m\in\mathcal{M}_n} \Delta_m.$$

It is easy to see that $\forall m \in \mathcal{M}_n,\ \Delta \subset \Delta_m \subset \Gamma_m$, see Lacour (2008). Moreover, we can prove in the same way that $\mathbb{P}(\Delta^c) \leq c/n^3$.

Let h_m be the orthogonal-$\mathbb{L}^2(A,\,\mathrm{d}x)$ projection of h_X on $\mathcal{S}_m$. We have

$$\mathbb{E}\|h_X - \hat{h}_m\|^2_A \leq 2\|h_X - h_m\|^2_A + 2\mathbb{E}[\|\hat{h}_m - h_m\|^2_A]$$

and

$$\|\hat{h}_m - h_m\|^2_A = \|\hat{h}_m - h_m\|^2_A \mathbb{1}_\Delta + \|\hat{h}_m - h_m\|^2_A \mathbb{1}_{\Delta^c}.$$

Now we have the following Lemma:

Lemma 5.2. Under the Assumptions of Proposition 3.1, for any $m \in \mathcal{M}_n$, we have

$$\|\hat{h}_m\|^2 \leq n^2 \quad a.s.$$

This yields

$$\begin{aligned}\mathbb{E}\left[\|h_X - \hat{h}_m\|^2_A\right] &\leq 2\|h_X - h_m\|^2_A + 2\mathbb{E}\left[\|\hat{h}_m - h_m\|^2_A \mathbb{1}_\Delta\right] + 4\mathbb{E}\left[\left(\|\hat{h}_m\|^2_A + \|h_X\|^2_A\right)\mathbb{1}_{\Delta^c}\right] \\ &\leq 2\|h_X - h_m\|^2_A + \frac{2}{S_0}\mathbb{E}\left[\|\hat{h}_m - h_m\|^2_{S_{X\wedge C}} \mathbb{1}_\Delta\right] + 4\left(n^2 + \|h_X\|^2_A\right)\mathbb{P}\left[\Delta^c\right] \\ &\leq 2\|h_X - h_m\|^2_A + \frac{3}{S_0}\mathbb{E}\left[\Psi_n(\hat{h}_m - h_m)\mathbb{1}_\Delta\right] + \frac{4}{n}\left(1 + \frac{\|h_X\|^2_A}{n^2}\right). \qquad (18)\end{aligned}$$

Now let us define

$$\nu_{n,1}(t) = \frac{1}{2\pi}\int t^*(-u)\frac{(\hat{\theta}_Y(u) - \mathbb{E}[\hat{\theta}_Y(u)])}{f^*_\varepsilon(u)}\,\mathrm{d}u \quad \text{with} \quad \hat{\theta}_Y(u) = (1/n)\sum_{j=1}^n \delta_j e^{iuY_j},$$

$$\nu_{n,2}(t) = \frac{1}{2\pi}\int (th_m)(-u)\left(\widehat{S}^*_{X\wedge C}(u) - \mathbb{E}\left[\hat{S}^*_{X\wedge C}(u)\right]\right)\mathrm{d}u.$$

We have

$$\gamma_n(\hat{h}_m) - \gamma_n(h_m) = \Psi_n(\hat{h}_m - h_m) - 2\nu_{n,1}(\hat{h}_m - h_m) - 2\nu_{n,2}(h_m - \hat{h}_m) + 2\langle \hat{h}_m - h_m, h_m - h_X\rangle_{S_{X\wedge C}}.$$

Then writing that on $\Delta \subset \Gamma_m$,

$$\gamma_n(\hat{h}_m) \leq \gamma_n(h_m) \qquad (19)$$

and defining $\mathcal{B}(m) = \{t \in \mathcal{S}_m, \|t\| = 1\}$, we get, on Δ,

$$\begin{aligned}\Psi_n(\hat{h}_m - h_m) &\leq 2\nu_{n,1}(\hat{h}_m - h_m) + 2\nu_{n,2}(h_m - \hat{h}_m) + 2\langle \hat{h}_m - h_m, h_X\mathbb{1}_A - h_m\rangle_{S_{X\wedge C}} \\ &\leq 2\nu_{n,1}(\hat{h}_m - h_m) + 2\nu_{n,2}(h_m - \hat{h}_m) \\ &\quad + \frac{1}{4}\|\hat{h}_m - h_m\|^2_{S_{X\wedge C}} + 4\|h_X\mathbb{1}_A - h_m\|^2_{S_{X\wedge C}} \\ &\leq \frac{1}{4}\|\hat{h}_m - h_m\|^2_{S_{X\wedge C}} + \frac{8}{S_0}\sup_{t\in\mathcal{B}(m)}\nu^2_{n,1}(t) + \frac{8}{S_0}\sup_{t\in\mathcal{B}(m)}\nu^2_{n,2}(t) \\ &\quad + \frac{1}{4}\|\hat{h}_m - h_m\|^2_{S_{X\wedge C}} + 4\|h_X\mathbb{1}_A - h_m\|^2_{S_{X\wedge C}}\end{aligned}$$

Then, as $\|\hat{h}_m - h_m\|^2_{S_{X\wedge C}} 1_\Delta \leq \frac{3}{2}\Psi_n(\hat{h}_m - h_m)1_\Delta$, we obtain that, on Δ,

$$\frac{1}{4}\Psi_n(\hat{h}_m - h_m) \leq 4\|h_X\mathbb{1}_A - h_m\|^2_{S_{X\wedge C}} + \frac{8}{S_0}\sup_{t\in\mathcal{B}(m)}\nu^2_{n,1}(t) + \frac{8}{S_0}\sup_{t\in\mathcal{B}(m)}\nu^2_{n,2}(t). \qquad (20)$$

Now, plugging (20) in (18), we get

$$\begin{aligned}\mathbb{E}\left[\|h_X-\hat{h}_m\|_A^2\right] &\leq \mathfrak{C}_1\left(\|h_X-h_m\|_A^2+\mathbb{E}\left[\sup_{t\in\mathcal{B}(m)}\nu_{n,1}^2(t)\right]+\mathbb{E}\left[\sup_{t\in\mathcal{B}(m)}\nu_{n,2}^2(t)\right]\right)\\ &\quad +\frac{4}{n}\left(1+\frac{\|h_X\|_A^2}{n^2}\right).\end{aligned} \tag{21}$$

with $\mathfrak{C}_1=\max_{i=1,2}(\mathfrak{C}_{0,i})$ with $\mathfrak{C}_{0,1}=2+48/S_0$, $\mathfrak{C}_{0,2}=96/S_0^2$. To conclude, we use the following proposition proved below:

Proposition 5.3. Under the assumptions of Proposition 3.1, for $i=1,2$

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}(m)}\nu_{n,i}^2(t)\right]\leq K_i\frac{D_m^{2\alpha+1}}{n}, \tag{22}$$

where K_1,K_2 are constants which do not depend on n nor m.

Inserting (22) in (21) gives the result and ends the proof of Proposition 3.1. □

5.2.1. *Proof of Lemma 5.2.* If $\hat{h}_m$ is non zero, then we are on Γ_m. Therefore $\|\hat{h}_m\|^2\leq\Phi_0^2\|\hat{\vec{\mathbf{a}}}_m\|^2=\|\mathbf{G}_m^{-1}\Upsilon_m\|^2\leq[9/(4S_0^2)]\|\Upsilon_m\|^2$. Then, using that $|\hat{\theta}_Y(u)|\leq 1$ a.s., we get

$$\begin{aligned}\|\Upsilon_m\|^2 &= \sum_{k\in\mathbb{K}_m}\left|\int\varphi_{m,k}^*(u)\frac{\hat{\theta}_Y(u)}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\\ &\leq \sum_k\left(\int\frac{|\varphi_{m,k}^*(u)|}{|f_\varepsilon^*(u)|}\,\mathrm{d}u\right)^2\leq\sum_k\left(\int\frac{|2^{-m/2}N_r^*(u/2^m)|}{|f_\varepsilon^*(u)|}\,\mathrm{d}u\right)^2\\ &\leq \mathfrak{C}_\varepsilon^2\sum_k 2^m2^{2m\alpha}\left(\int|N_r^*(v)|(1+v^2)^{\alpha/2}\,\mathrm{d}v\right)^2\\ &\leq \mathfrak{C}_\varepsilon^2D_m^{2\alpha+2}\left(\int\left(\frac{\sin(v/2)}{v/2}\right)^r(1+v^2)^{\alpha/2}\,\mathrm{d}v\right)^2\\ &\leq \mathfrak{C}_1D_m^{2\alpha+2}\leq n^2\end{aligned}$$

where we use $D_m^{2\alpha+1}\leq n$ in $\mathcal{M}_n$. The constant is such that $\mathfrak{C}_1=\mathfrak{C}_\varepsilon^2\int\left(\frac{\sin(v/2)}{v/2}\right)^r(1+v^2)^{\alpha/2}\,\mathrm{d}v$ and is finite provided that $r>\alpha+1$. □

5.2.2. *Proof of Proposition 5.3.* We start with $\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,1}(t)|^2\right]$.

$$\begin{aligned}
\Phi_0^{-2}\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,1}(t)|^2\right] &\leq \sum_{k\in\mathbb{K}_m}\mathbb{E}\left[\nu_{n,1}^2(\varphi_{m,k})\right]\\
\leq\ & \frac{1}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^n\int\varphi_{m,k}^*(-u)\delta_j\frac{e^{iuY_j}-\mathbb{E}e^{iuY_j}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right)^2\right]\\
=\ & \frac{1}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\frac{1}{n}\mathbb{V}\mathrm{ar}\left(\int\varphi_{m,k}^*(-u)\delta_1\frac{e^{iuY_1}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right)\leq\frac{1}{n}\frac{1}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\mathbb{E}\left|\int\varphi_{m,k}^*(-u)\delta_1\frac{e^{iuY_1}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\\
=\ & \frac{1}{n}\frac{1}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\mathbb{E}\left|\int\varphi_{m,k}^*(-u)\delta_1\frac{e^{iu(X_1+\varepsilon_1)}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\\
=\ & \frac{1}{n}\frac{1}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\int\int\left|\int\varphi_{m,k}^*(-u)\frac{e^{iu(x+e)}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2 S_C(x)f_X(x)f_\varepsilon(e)\,\mathrm{d}x\,\mathrm{d}e\\
\leq\ & \frac{1}{n}\frac{\|S_Cf_X\|_\infty}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\int\left|\int\varphi_{m,k}^*(-u)\frac{e^{iuz}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\mathrm{d}z\leq\frac{1}{n}\frac{\|S_Cf_X\|_\infty}{2\pi}\sum_{k\in\mathbb{K}_m}\int\left|\frac{\varphi_{m,k}^*(-u)}{f_\varepsilon^*(u)}\right|^2\mathrm{d}u
\end{aligned}$$

by using Parseval equality. Using that the noise is ordinary-smooth with constant α, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,1}(t)|^2\right] &\leq \frac{\Phi_0^2}{n}\frac{\|S_Cf_X\|_\infty}{2\pi}\sum_{k\in\mathbb{K}_m}\int\frac{|N_r^*(z)|^2}{|f_\varepsilon^*(2^mz)|^2}\,\mathrm{d}z\\
&\leq \frac{\Phi_0^2\mathfrak{C}_\varepsilon^2}{n}\frac{\|S_Cf_X\|_\infty}{2\pi}\sum_{k\in\mathbb{K}_m}\int|N_r^*(z)|^2(1+(2^mz)^2)^\alpha\,\mathrm{d}z\\
&\leq \frac{\Phi_0^2\mathfrak{C}_\varepsilon^2}{n}\frac{\|S_Cf_X\|_\infty}{2\pi}\sum_{k\in\mathbb{K}_m}2^{2m\alpha}\int|N_r^*(z)|^2(1+z^2)^\alpha\,\mathrm{d}z
\end{aligned}$$

If $r>\alpha+\frac{1}{2}$, the integral $\int|N_r^*(z)|^2(1+z^2)^\alpha\,\mathrm{d}z$ is finite. We obtain

$$\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,1}(t)|^2\right]\leq K_1\frac{D_m^{2\alpha+1}}{n},$$

with

$$K_1=\frac{\Phi_0^2\mathfrak{C}_\varepsilon^2\|S_Cf_X\|_\infty}{2\pi}\int|N_r^*(z)|^2(1+z^2)^\alpha\,\mathrm{d}z$$

which is the announced result for $i=1$.

• Now we consider $\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,2}(t)|^2\right]$. The following inequality will be used:

$$\forall u\in\mathbb{R},\forall a\in\mathbb{R},\quad\left|\frac{e^{iua}-1}{u}\right|\leq|a|. \tag{23}$$

First, because of convergence problems near 0, we split this term in three parts:

$$\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,2}(t)|^2\right]\leq 3(T_1+T_2+T_3)$$

where $Z_j(u) = e^{iuY_j} - 1$ and for $\ell = 1, 2, 3$,

$$T_\ell = \frac{1}{(2\pi)^2}\mathbb{E}\left[\sup_{\|t\|=1}\left|\int_{D_\ell}(th_m)^*(u)\frac{1}{n}\frac{1}{iu}\sum_{j=1}^n \frac{Z_j(u) - \mathbb{E}[Z_j(u)]}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right],$$

with $D_1 = \{|u| \le 1\}$, $D_2 = \{1 < |u| \le 2^m\}$ and $D_3 = \{|u| > 2^m\}$. Indeed

$$e^{iuY_j} - f_\varepsilon^*(u) - \mathbb{E}[e^{iuY_j} - f_\varepsilon^*(u)] = Z_j(u) - \mathbb{E}[Z_j(u)].$$

In all cases we start by decomposing the functions t and h_m on the basis $(\varphi_{m,j})_j$:

$$T_\ell \le \frac{\Phi_0^4\|h_m\|_A^2}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\mathbb{E}\left[\left|\int_{D_\ell}(\varphi_{m,k}\varphi_{m,j})^*(u)\frac{1}{n}\frac{1}{iu}\sum_{j=1}^n\frac{Z_j(u) - \mathbb{E}[Z_j(u)]}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right].$$

For T_1 we write that, by Schwarz's inequality

$$\begin{aligned}
T_1 &\le \frac{\Phi_0^4\|h_m\|_A^2}{(2\pi)^2}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\int_{|u|\le 1}|(\varphi_{m,k}\varphi_{m,j})^*(u)|^2\mathbb{E}\left[\left|\frac{1}{n}\frac{1}{iu}\sum_{j=1}^n\frac{Z_j(u) - \mathbb{E}[Z_j(u)]}{f_\varepsilon^*(u)}\right|^2\right]\mathrm{d}u\\
&\le \frac{\Phi_0^4\|h_m\|_A^2}{(2\pi)^2 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\int_{|u|\le 1}\frac{|(\varphi_{m,k}\varphi_{m,j})^*(u)|^2}{|f_\varepsilon^*(u)|^2}\mathbb{E}\left[\left|\frac{Z_1(u) - \mathbb{E}(Z_1(u))}{iu}\right|^2\right]\mathrm{d}u.
\end{aligned}$$

Now we use Equation (23) and Lemma 5.1, 2) (as $(\varphi_{m,k}\varphi_{m,j})^* = \varphi_{m,k}^* \star \varphi_{m,j}^*/(2\pi)$) to get

$$\begin{aligned}
T_1 &\le \frac{\Phi_0^4\|h_m\|_A^2}{(2\pi)^3 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\int_{|u|\le 1}\frac{\mathrm{d}u}{|f_\varepsilon^*(u)|^2}\mathbb{E}[Y_1^2]\\
&= \frac{\Phi_0^4 2^\alpha\mathfrak{C}_\varepsilon^2\|h_X\|_A^2(2r-1)\mathbb{E}[Y_1^2]D_m}{(2\pi)^3 n}
\end{aligned}$$

Thus

$$T_1 \le \mathfrak{C}_{2,3}D_m/n, \quad \text{with} \quad \mathfrak{C}_{2,3} = \Phi_0^4\mathfrak{C}_\varepsilon^2 2^\alpha\|h_X\|_A^2(2r-1)\mathbb{E}[Y_1^2]/(2\pi)^3. \tag{24}$$

For T_2, we use Parseval formula and Lemma 5.1, 2) and we obtain

$$\begin{aligned}
T_2 &\le \frac{\Phi_0^4\|h_m\|_A^2}{(2\pi)^2 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\mathbb{E}\left[\left|\int_{1<|u|\le 2^m}(\varphi_{m,k}\varphi_{m,j})^*(u)\frac{1}{iu}\frac{e^{iuY_1}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right]\\
&\le \frac{\Phi_0^4\|h_m\|_A^2\|f_Y\|_\infty}{2\pi n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\int_{1<|u|\le 2^m}\frac{|(\varphi_{m,k}\varphi_{m,j})^*(u)|^2}{u^2|f_\varepsilon^*(u)|^2}\,\mathrm{d}u\\
&\le \frac{\Phi_0^4\|h_m\|_A^2\|f_Y\|_\infty}{(2\pi)^2 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\int_{1<|u|\le 2^m}\frac{\mathrm{d}u}{u^2|f_\varepsilon^*(u)|^2}\\
&\le \frac{2(2r-1)D_m\Phi_0^4\|h_X\|_A^2\mathfrak{C}_\varepsilon^2\|f_Y\|_\infty}{(2\pi)^2 n}(2^m)^{(2\alpha-1)_+} \le \frac{2(2r-1)\Phi_0^4\mathfrak{C}_\varepsilon^2\|h_X\|_A^2\|f_Y\|_\infty}{(2\pi)^2 n}D_m^{2\alpha\vee 1}.
\end{aligned}$$

Thus

$$T_2 \le \mathfrak{C}_{2,2}D_m^{2\alpha}/n, \quad \text{with} \quad \mathfrak{C}_{2,2} = 2(2r-1)\mathfrak{C}_\varepsilon^2\Phi_0^4\|h_X\|_A^2\|f_Y\|_\infty/(2\pi)^2, \tag{25}$$

since we assume that $\alpha > 1$.

Lastly, for T_3, we use Lemma 5.1, 2).

$$\begin{aligned}
T_3 &\leq \frac{\Phi_0^4\|h_X\|_A^2}{(2\pi)^2 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\mathbb{E}\left[\left|\int_{|u|>2^m}(\varphi_{m,k},\varphi_{m,j})^*(u)\frac{1}{iu}\frac{e^{iuY_1}}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right]\\
&\leq \frac{\mathfrak{C}_\varepsilon^2\Phi_0^4\|h_X\|_A^2}{(2\pi)^3 n}\sum_{k\in\mathbb{K}_m}\sum_{j,|j-k|<r}\left(\int_{|u|>2^m}|2^{-m}u|^{1-r}\frac{(1+u^2)^{\alpha/2}}{|u|}\,\mathrm{d}u\right)^2\\
&\leq \frac{\mathfrak{C}_\varepsilon^2\Phi_0^4\|h_X\|_A^2}{(2\pi)^3 n}(2r-1)D_m\left(\int_{|v|>1}|v|^{1-r}\frac{(1+2^{2m}v^2)^{\alpha/2}}{2^m|v|}2^m\,\mathrm{d}v\right)^2\\
&\leq \frac{(2r-1)\mathfrak{C}_\varepsilon^2\Phi_0^4\|h_X\|_A^2 D_m^{2\alpha+1}}{(2\pi)^3 n}\left(\int_{|v|>1}|v|^{-r}(1+v^2)^{\alpha/2}\,\mathrm{d}v\right)^2\leq K_2\frac{D_m^{2\alpha+1}}{n}, \qquad (26)
\end{aligned}$$

with

$$K_2=\frac{(2r-1)\mathfrak{C}_\varepsilon^2\Phi_0^2\|h_X\|_A^2}{(2\pi)^3}\left(\int_{|v|>1}|v|^{-r}(1+v^2)^{\alpha/2}\,\mathrm{d}v\right)^2,$$

if $\alpha-r+1<0$ i.e. $r>\alpha+1$. Then gathering (24), (25) and (26) implies the result with $\mathfrak{C}_{2,1}=K_1+K_2$. □

5.3. **Proof of Theorem 3.2.** The proof starts similarly as proof of Proposition 3.1 with $\hat{h}_m$ replaced by $\hat{h}_{\widehat{m}}$ until inequality (19) which is now replaced by the fact that $\forall m\in\mathcal{M}_n$,

$$\gamma_n(\hat{h}_{\widehat{m}})+\mathrm{pen}(\widehat{m})\leq\gamma_n(h_m)+\mathrm{pen}(m).$$

Here, we define $\mathcal{B}(m,m')=\{t\in\mathcal{S}_m+\mathcal{S}_{m'},\|t\|=1\}$ and we get

$$\begin{aligned}
\Psi_n(\hat{h}_{\widehat{m}}-h_m) &\leq 2\nu_{n,1}(\hat{h}_{\widehat{m}}-h_m)+2\nu_{n,2}(\hat{h}_{\widehat{m}}-h_m)+2\langle\hat{h}_{\widehat{m}}-h_m,h_m-h_X\rangle_{S_{X\wedge C}}\\
&\quad+\mathrm{pen}(m)-\mathrm{pen}(\widehat{m})\\
&\leq 2\nu_{n,1}(\hat{h}_{\widehat{m}}-h_m)+2\nu_{n,2}(\hat{h}_{\widehat{m}}-h_m)\\
&\quad+\frac14\|\hat{h}_{\widehat{m}}-h_m\|_{S_{X\wedge C}}^2+4\|h_X-h_m\|_{S_{X\wedge C}}^2+\mathrm{pen}(m)-\mathrm{pen}(\widehat{m})\\
&\leq \frac14\|\hat{h}_{\widehat{m}}-h_m\|_{S_{X\wedge C}}^2+\frac{8}{S_0}\sup_{t\in\mathcal{B}(m,\widehat{m})}\nu_{n,1}^2(t)+\frac{8}{S_0}\sup_{t\in\mathcal{B}(m,\widehat{m})}\nu_{n,2}^2(t)\\
&\quad+\frac14\|\hat{h}_{\widehat{m}}-h_m\|_{S_{X\wedge C}}^2+4\|h_X-h_m\|_{S_{X\wedge C}}^2+\mathrm{pen}(m)-\mathrm{pen}(\widehat{m})\\
&\leq 4\|h_X-h_m\|_{S_{X\wedge C}}^2+\mathrm{pen}(m)+\frac12\|\hat{h}_{\widehat{m}}-h_m\|_{S_{X\wedge C}}^2\\
&\quad+\frac{8}{S_0}\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})}\nu_{n,1}^2(t)-p_1(m,\widehat{m})\right\}_++\frac{8}{S_0}\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})}\nu_{n,2}^2(t)-p_2(m,\widehat{m})\right\}_+\\
&\quad+\frac{8}{S_0}(p_1(m,\widehat{m})+p_2(m,\widehat{m}))-\mathrm{pen}(\widehat{m})
\end{aligned}$$

We shall choose the penalty so that $\forall m,m'\in\mathcal{M}_n$,

$$(8/S_0)(p_1(m,m')+p_2(m,m'))\leq\mathrm{pen}(m)+\mathrm{pen}(m').$$

Therefore we get, on Δ,

$$\begin{aligned}\Psi_n(\hat{h}_{\widehat{m}} - h_m) &\leq 4\|h_X - h_m\|^2_{S_{X\wedge C}} + 2\text{pen}(m) + \frac{1}{2}\|\hat{h}_{\widehat{m}} - h_m\|^2_{S_{X\wedge C}} \\ &+8\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,1}(t) - p_1(m,\widehat{m})\right\}_+ + 8\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,2}(t) - p_2(m,\widehat{m})\right\}_+\end{aligned}$$

Then, $\|\hat{h}_{\widehat{m}} - h_m\|^2_{S_{X\wedge C}} 1_\Delta \leq \frac{3}{2}\Psi_n(\hat{h}_{\widehat{m}} - h_m)$ implies that, on Δ,

$$\begin{aligned}\frac{1}{4}\Psi_n(\hat{h}_{\widehat{m}} - h_m) &\leq 4\|h_X - h_m\|^2_{S_{X\wedge C}} + 2\text{pen}(m) + 8\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,1}(t) - p_1(m,\widehat{m})\right\}_+ \\ &+8\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,2}(t) - p_2(m,\widehat{m})\right\}_+\end{aligned}$$

Now, inserting this in (18) with $\hat{h}_m$ replaced by $\tilde{h} = \hat{h}_{\widehat{m}}$, we get

$$\begin{aligned}\mathbb{E}\|h_X - \tilde{h}\|^2_A &\leq \mathfrak{C}_1\left(\|h_X - h_m\|^2_A + \text{pen}(m) + \mathbb{E}\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,1}(t) - p_1(m,\widehat{m})\right\}_+\right. \\ &\left.+\mathbb{E}\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,2}(t) - p_2(m,\widehat{m})\right\}_+\right) + \frac{4}{n}\left(1 + \frac{\|h_X\|^2_A}{n}\right).\end{aligned}$$

with $\mathfrak{C}_1 = \max_{i=1,2}(\mathfrak{C}_{0,i})$ with $\mathfrak{C}_{0,1} = 2 + 48/S_0$, $\mathfrak{C}_{0,2} = 296/S_0^2$. Now we use the following proposition proved below:

Proposition 5.4. Let, for $i = 1, 2$, $p_i(m, m') = \tau_i \log(n) D_m^{2\alpha+1}/n$ with τ_i proportional to K_i, then

$$\mathbb{E}\left\{\sup_{t\in\mathcal{B}(m,\widehat{m})} \nu^2_{n,i}(t) - p_i(m,\widehat{m})\right\}_+ \leq \frac{\mathfrak{C}}{n}.$$

This ends the proof of Theorem 3.2. □

5.3.1. *Proof of Proposition 5.4.* To prove Proposition 5.4, we apply to $\nu_{n,1}$ and $\nu_{n,2}$ the following version of Talagrand inequality.

Lemma 5.5. Let $T_1, \ldots, T_n$ be independent random variables and $\nu_n(r) = (1/n)\sum_{j=1}^n (r(T_j) - \mathbb{E}[r(T_j)])$, for r belonging to a countable class $\mathcal{R}$ of measurable functions. Then, for $\epsilon > 0$,

$$\mathbb{E}\left\{\sup_{r\in\mathcal{R}} |\nu_n(r)|^2 - (1+2\epsilon)H^2\right\}_+ \leq C\left(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{M^2}{n^2C^2(\epsilon)}e^{-K_2C(\epsilon)\sqrt{\epsilon}\frac{nH}{M}}\right) \tag{27}$$

with $K_1 = 1/6$, $K_2 = 1/(21\sqrt{2})$, $C(\epsilon) = \sqrt{1+\epsilon} - 1$ and C a universal constant and where

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \leq M, \quad \mathbb{E}\left[\sup_{r\in\mathcal{R}} |\nu_n(r)|\right] \leq H, \quad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{j=1}^n \mathbb{V}\text{ar}[r(T_j)] \leq v.$$

Inequality (27) is a straightforward consequence of the Talagrand inequality given in Klein and Rio (2005). Moreover, standard density arguments allow to apply it to the unit ball of spaces.

• *Study of* $\nu_{n,1}$. We need to compute the three bounds involved in Lemma 5.5. We proved the bound for $\mathbb{E}\left[\sup_{\|t\|=1} |\nu_{n,1}(t)|^2\right]$ in Proposition 5.3, and the spaces are nested so that $\mathcal{S}_m + \mathcal{S}_{m'}$ is just equal to the largest of the two spaces. Therefore $H^2 = \frac{C}{n}(D_m \vee D_{m'})^{2\alpha+1}\|S_C f_X\|_\infty$.

Now we compute v such that $\sup_{t\in\mathcal{S}_m\vee\mathcal{S}_{m'},\|t\|=1}\mathbb{V}\mathrm{ar}[\delta_1\int\frac{t^*(u)}{f_\varepsilon^*(u)}e^{iuY_1}\,\mathrm{d}u]\le v$. For this we denote by $S_{m^*}=\mathcal{S}_m\vee\mathcal{S}_{m'}$ and write, using Parseval equality as previously,

$$\begin{aligned}
&\sup_{t\in S_{m^*},\|t\|=1}\mathbb{V}\mathrm{ar}\left[\frac{\delta_1}{2\pi}\int\frac{t^*(u)}{f_\varepsilon^*(u)}e^{iuY_1}\,\mathrm{d}u\right]\le\frac{\|f_XS_C\|_\infty}{2\pi}\sup_{t\in S_{m^*},\|t\|=1}\int\left|\frac{t^*(u)}{f_\varepsilon^*(u)}\right|^2\mathrm{d}u\\
\le\;&\frac{\|f_XS_C\|_\infty}{2\pi}\sup_{t\in S_{m^*},\|t\|=1}\|t^*\|\left(\int\frac{|t^*(u)|^2}{|f_\varepsilon^*(u)|^4}\,\mathrm{d}u\right)^{1/2}\\
\le\;&\frac{\|f_XS_C\|_\infty\Phi_0^2}{\sqrt{2\pi}}\left(\sum_{k\in\mathbb{K}_{m^*}}C_\varepsilon^4\int|N_r^*(v)|^2(1+(2^{m^*}v)^2))^{2\alpha}\,\mathrm{d}v\right)^{1/2}\\
\le\;&C_\varepsilon^2\|f_XS_C\|_\infty D_{m^*}^{2\alpha+1/2}\frac{\Phi_0^2}{\sqrt{2\pi}}\left(\int|N_r^*(w)|^2(1+w^2)^\alpha\,\mathrm{d}w\right)^{1/2}:=v.
\end{aligned}$$

The last constant to be computed in order to apply Lemma 5.5 is M.

$$\begin{aligned}
&\sup_{t\in\mathcal{S}_m,\|t\|=1}\sup_z\left|\delta_1\frac{1}{2\pi}\int e^{iuz}\frac{t^*(u)}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|\le\sup_{t\in\mathcal{S}_m,\|t\|=1}\sum_{k\in\mathbb{K}_m}\int a_{m,k}\left|\frac{\varphi_{m,k}^*(u)}{f_\varepsilon^*(u)}\right|\mathrm{d}u\\
\le\;&\left(\sum_{k\in\mathbb{K}_m}\left(\int\left|\frac{\varphi_{m,k}^*(u)}{f_\varepsilon^*(u)}\right|\mathrm{d}u\right)^2\right)^{1/2}\le\left(\sum_{k\in\mathbb{K}_m}2^{-m}\left(\int\left|\frac{N_r^*(u/2^m)}{f_\varepsilon^*(u)}\right|\mathrm{d}u\right)^2\right)^{1/2}\\
\le\;&\left(2^{2m}\left(\int\left|\frac{N_r^*(v)}{f_\varepsilon^*(2^mv)}\right|\mathrm{d}v\right)^2\right)^{1/2}\le\left(C\;2^{2m}2^{2m\alpha}\right)^{1/2}\le CD_m^{\alpha+1}=:M
\end{aligned}$$

Finally, Lemma 5.5 with $\epsilon=1/2$, $M=(D_m^*)^{\alpha+1}$, $v=C(D_m^*)^{2\alpha+1/2}$ and $H^2=C(D_m^*)^{2\alpha+1}/n$ where $D_m^*=D_m\vee D_{m'}$ yields

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in\mathcal{S}_m\vee S_{\widehat{m}}}\nu_{n,1}^2(t)-2H^2\right]\;&\le\;\sum_{m'/\;D_{m'}\in\mathcal{M}_n}\mathbb{E}\left[\sup_{t\in\mathcal{S}_m\vee\mathcal{S}_{m'}}\nu_{n,1}^2(t)-2H^2\right]\\
&\le\;C_0\left(\sum_{m'}\frac{(D_m\vee D_{m'})^{2\alpha+1}}{n}e^{-\mathfrak{C}_1\sqrt{D_m\vee D_{m'}}}\right.\\
&\qquad\left.+\frac{(D_m\vee D_{m'})^{2\alpha+2}}{n^2}e^{-\mathfrak{C}_2\sqrt{n}/\sqrt{D_m\vee D_{m'}}}\right)\le K/n
\end{aligned}$$

since the first series is convergent and $\forall m\in\mathcal{M}_n$, $D_m\le\sqrt{n}$ (indeed, $D_m^{2\alpha+1}\le n$ and $\alpha>1/2$); thus $\sqrt{n/D_m\vee D_{m'}}\ge n^{1/4}$ and $\exp(-\mathfrak{C}_2\sqrt{n}/\sqrt{D_m\vee D_{m'}})\le\exp(-\mathfrak{C}_2n^{1/4})$ for any m. □

• *Study of* $\nu_{n,2}$. First, because of boundedness problems, we split the expectation in two terms

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in\mathcal{S}_m\vee S_{\widehat{m}}}\nu_{n,2}^2(t)-p_2(m,\widehat{m})\right]\;&\le\;2\mathbb{E}\left[\sup_{t\in\mathcal{S}_m\vee S_{\widehat{m}}}|\nu_{n,2}^{(1)}(t)|^2\right]\\
&\quad+2\mathbb{E}\left[\sup_{t\in\mathcal{S}_m\vee S_{\widehat{m}}}|\nu_{n,2}^{(2)}(t)|^2-p_2(m,\widehat{m})/2\right]
\end{aligned}$$

where, for $\ell = 1, 2$,

$$\nu_{n,2}^{(\ell)}(t) = \int t^* \star h_m^*(u) \frac{1}{n} \frac{1}{iu} \sum_{j=1}^n \frac{Z_j^{(\ell)}(u) - \mathbb{E}(Z_j^{(\ell)}(u))}{f_\varepsilon^*(u)} \, \mathrm{d}u,$$

with $Z_j^{(1)}$ and $Z_j^{(2)}$ defined, introducing a cut-off c_n, as:

$$Z_j^{(1)}(u) = (e^{iuY_j} - 1)\mathbb{1}_{\{|Y_j| > c_n\}}, \;\; Z_j^{(2)}(u) = (e^{iuY_j} - 1)\mathbb{1}_{\{|Y_j| \le c_n\}}.$$

We choose

$$c_n = c_0 \sqrt{n} / \log(n)$$

where c_0 is detailed below.

$\triangleright$ We start with $\nu_{n,2}^{(1)}$. We want to bound $\mathbb{E}_1 := \mathbb{E}\left[\sup_{\|t\|=1, t\in\mathcal{S}_m + S_{\widehat{m}}} |\nu_{n,2}^{(1)}(t)|^2\right]$. Let $\mathcal{S}_n$ be a nesting space such that $\forall m \in \mathcal{M}_n, \; \mathcal{S}_m \subset \mathcal{S}_n$. Then

$$\begin{aligned}
\mathbb{E}_1 &\le \mathbb{E}\left[\sup_{\|t\|=1, t\in\mathcal{S}_n} |\nu_{n,2}^{(1)}(t)|^2\right] \\
&\le \mathbb{E}\left[\sum_{k\in\mathbb{K}_{m_n}} \left| \int \varphi_{m_n,k}^* \star h_m^*(u) \frac{1}{n}\frac{1}{iu} \sum_{j=1}^n \frac{Z_j^{(1)}(u) - \mathbb{E}(Z_j^{(1)}(u))}{f_\varepsilon^*(u)} \, \mathrm{d}u \right|^2\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{k\in\mathbb{K}_{m_n}} \left| \int \varphi_{m_n,k}^* \star h_m^*(u) \frac{1}{iu} \frac{Z_1^{(1)}(u) - \mathbb{E}(Z_1^{(1)}(u))}{f_\varepsilon^*(u)} \, \mathrm{d}u \right|^2\right] \\
&\le \frac{\mathbb{E}[Y_1^2 \mathbb{1}_{|Y_1|>c_n}]}{n} \sum_{k\in\mathbb{K}_{m_n}} \left(\int \left| \frac{\varphi_{m_n,k}^* \star h_m^*(u)}{f_\varepsilon^*(u)} \, \mathrm{d}u \right| \right)^2
\end{aligned}$$

Now, decomposing h_m along the $\varphi_{m_n,\ell}$ we get

$$\sum_{k\in\mathbb{K}_{m_n}} \left(\int \left| \frac{\varphi_{m_n,k}^* \star h_m^*(u)}{f_\varepsilon^*(u)} \, \mathrm{d}u \right| \right)^2 \le \Phi_0^2 \|h_m\|_A^2 \sum_{k\in\mathbb{K}_{m_n}} \sum_{\ell, |\ell-k|<r} \left(\int \left| \frac{\varphi_{m_n,k}^* \star \varphi_{m_n,\ell}^*(u)}{f_\varepsilon^*(u)} \, \mathrm{d}u \right| \right)^2.$$

With Lemma 5.1, 2), we get

$$\begin{aligned}
&\sum_{k\in\mathbb{K}_{m_n}} \left(\int \left| \frac{\varphi_{m_n,k}^* \star h_m^*(u)}{f_\varepsilon^*(u)} \, \mathrm{d}u \right| \right)^2 \\
&\le 4\Phi_0^2 \|h_m\|_A^2 \mathfrak{C}_\varepsilon^2 \sum_{k\in\mathbb{K}_{m_n}} \sum_{\ell, |\ell-l|<r} \left(\int_{|u|\le 2^{m_n}} (1+u^2)^{\alpha/2} \, \mathrm{d}u + \int_{|u|>2^{m_n}} |2^{-m_n} u|^{1-r} (1+u^2)^{\alpha/2} \, \mathrm{d}u \right)^2 \\
&\le 4\Phi_0^2 \|h_X\|_A^2 (2r-1) D_{m_n} \mathfrak{C}_\varepsilon^2 (2^{2m_n(\alpha+1)} + 2^{m_n(2\alpha+2)} (\int_{|v|>1} |v|^{1-r} (1+v^2)^{\alpha/2} \, \mathrm{d}v)^2) = \mathrm{O}(\mathcal{D}_n^{2\alpha+3}).
\end{aligned}$$

All dimensions considered are such that $D_m^{2\alpha+1}/n \le \mathfrak{C}$, so that $\mathcal{D}_n^{2\alpha+3} \le n^3$. Then for any p, we have

$$\mathbb{E}_1 \le \mathfrak{C} n^2 \mathbb{E}[Y_1^2 \mathbb{1}_{|Y_1|>c_n}] \le \mathfrak{C} n^2 \mathbb{E}[|Y_1|^{2+p}]/c_n^p$$

so that $p = 8$ implies that $\mathbb{E}_1 \le \mathfrak{C}\mathbb{E}[|Y_1|^{10}](\log(n))^8/n^2 \le \mathfrak{C}/n$.

$\triangleright$ Now we consider $\nu_{n,2}^{(2)}$ and more precisely, $\mathbb{E}\left[\sup_{\|t\|=1} |\nu_{n,2}^{(2)}(t)|^2\right]$.

We apply Talagrand Inequality and compute the three bounds involved in Lemma 5.5. First, because of convergence problems near 0, similarly to proof of Proposition 5.3, we split this term in three parts:

$$\begin{aligned}\mathbb{E}\left[\sup_{\|t\|=1}|\nu_{n,2}^{(2)}(t)|^2\right] &\leq 3\mathbb{E}\left[\sup_{\|t\|=1}\left|\int_{|u|\leq 1} t^*\star h_m^*(u)\frac{1}{n}\sum_{j=1}^n\frac{Z_j^{(2)}(u)-\mathbb{E}[Z_j^{(2)}(u)]}{iuf_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right]\\ &+3\mathbb{E}\left[\sup_{\|t\|=1}\left|\int_{1<|u|\leq 2^{m\vee m'}} t^*\star h_m^*(u)\frac{1}{n}\sum_{j=1}^n\frac{Z_j^{(2)}(u)-\mathbb{E}[Z_j^{(2)}(u)]}{iuf_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right]\\ &+3\mathbb{E}\left[\sup_{\|t\|=1}\left|\int_{|u|>2^{m\vee m'}} t^*\star h_m^*(u)\frac{1}{n}\sum_{j=1}^n\frac{Z_j^{(2)}(u)-\mathbb{E}[Z_j^{(2)}(u)]}{iuf_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\right]\end{aligned}$$

Bounding these terms follows the same line as bounding $\mathbb{E}[\sup_{\|t\|=1}|\nu_{n,2}(t)|^2]$ in the proof of Proposition 5.3, so that we get $H^2=\mathfrak{C}(D_m\vee D_{m'})^{2\alpha+1}/n$.

Now we look for $v=3(v_1+v_2+v_3)$ with obvious notation. For the first two terms corresponding to previous T_1 and T_2, we can clearly obtain $v_1=K_{2,1}D_m\vee D_{m'}$ and $v_2=K_{2,2}D_m^{2\alpha}$ under the assumption $\alpha>1/2$. For T_3, we have

$$\begin{aligned}\sup_{t\in S_{m^*},\|t\|=1}\mathbb{V}\mathrm{ar}\left[\frac{1}{2\pi}\int_{|u|\geq 2^{m^*}}\frac{(th_m)^*(u)}{iuf_\varepsilon^*(u)}e^{iuY_1}\,\mathrm{d}u\right] &\leq \frac{\|f_Y\|_\infty}{2\pi}\sup_{t\in S_{m^*},\|t\|=1}\int_{|u|\geq 2^{m^*}}\left|\frac{(th_m)^*(u)}{uf_\varepsilon^*(u)}\right|^2\mathrm{d}u\\ \leq \frac{\|f_Y\|_\infty\Phi_0^4\|h_X\|_A^2}{2\pi}\sum_{j\in\mathbb{K}_{m^*}}\sum_{|j-k|<r}\int_{|u|\geq 2^{m^*}}\left|\frac{(\varphi_{m^*,k}\varphi_{m^*,j})^*(u)}{uf_\varepsilon^*(u)}\right|^2\mathrm{d}u&\\ \leq \frac{\|f_Y\|_\infty\Phi_0^4\|h_X\|_A^2}{2\pi}\sum_{j\in\mathbb{K}_{m^*}}\sum_{|j-k|<r}\int_{|u|\geq 2^{m^*}}\left|\frac{(2^{-m^*}u)^{1-r}}{uf_\varepsilon^*(u)}\right|^2\mathrm{d}u&\\ \leq \frac{\|f_Y\|_\infty\Phi_0^4\|h_X\|_A^2(2r-1)}{2\pi}D_{m^*}^{2\alpha}\int_{|z|>1}|z|^{-2r}(1+z^2)^\alpha\,\mathrm{d}z=v_3,&\end{aligned}$$

for $r>\alpha+1/2$. Therefore we obtain $v=\theta(D_m\vee D_{m'})^{2\alpha}$.

Next, we look for M_1. The term to be bounded is

$$\sup_{\|t\|=1}\sup_{y\in\mathbb{R}^+}\left|\int\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\frac{(e^{iuy}-1)\mathbb{1}_{|y|<c_n}}{u}\,\mathrm{d}u\right|$$

We split it into two terms as previously, depending on $|u|<2^m$ or $|u|\geq 2^m$.

$$\begin{aligned}\sup_{\|t\|=1}\sup_{y\in\mathbb{R}^+}\left|\int_{|u|\leq 2^m}\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\frac{(e^{iuy}-1)\mathbb{1}_{|y|<c_n}}{u}\,\mathrm{d}u\right| &\leq c_n\sup_{\|t\|=1}\int_{|u|\leq 2^m}\left|\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\right|\mathrm{d}u\\ &\leq c_n2^{m\alpha}\sup_{\|t\|=1}\left|\int_{|u|\leq 2^m}|t^*\star h_m^*(u)|\,\mathrm{d}u\right|\end{aligned}$$

As previously, we have

$$\begin{aligned}\sup_{\|t\|=1}\left|\int_{|u|\le 2^m}|t^*\star h_m^*(u)|\,\mathrm{d}u\right|^2 &\le 2^m\int |t^*\star h_m^*(u)|^2\,\mathrm{d}u = 2\pi 2^m\|th_m\|_A^2\\ &\le 2\pi\|h_m\|_{\infty,A}^2 D_m,\end{aligned}$$

so that, as $\|h_m\|_{\infty,A}\le 2\|h_X\|_{\infty,A}$, we obtain

$$\sup_{\|t\|=1}\sup_{y\in\mathbb{R}^+}\left|\int_{|u|\le 2^m}\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\frac{(e^{iuy}-1)\mathbb{1}_{|y|<c_n}}{u}\,\mathrm{d}u\right|\le Kc_nD_m^{\alpha+1/2}.$$

Now, we bound the second part with $|u|>2^m$

$$\begin{aligned}&\sup_{\|t\|=1}\sup_{y\in\mathbb{R}^+}\left|\int_{|u|>2^m}\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\frac{(e^{iuy}-1)\mathbb{1}_{|y|<c_n}}{u}\,\mathrm{d}u\right|\\
\le\;& \sup_{\|t\|=1}\int_{|u|>2^m}\left|\frac{t^*\star h_m^*(u)}{f_\varepsilon^*(u)}\frac{2}{u}\right|\mathrm{d}u\\
\le\;& \|h_m\|_A\left(\sum_{k\in\mathbb{K}_m}\sum_{\ell,|k-\ell|<r}\left(\int_{|u|>2^m}\left|\frac{\varphi_{m,k}^*\star\varphi_{m,\ell}^*(u)}{f_\varepsilon^*(u)}\frac{2}{u}\right|\mathrm{d}u\right)^2\right)^{1/2}\\
\le\;& 2\|h_m\|_A\left(\sum_{k\in\mathbb{K}_m}\sum_{|k-\ell|<r}\left(\int_{|u|>2^m}\left|\varphi_{m,k}^*\star\varphi_{m,\ell}^*(u)\frac{(1+u^2)^{\alpha/2}}{u}\right|\mathrm{d}u\right)^2\right)^{1/2}\\
\le\;& 2\|h_X\|_A\left(\sum_{k\in\mathbb{K}_m}\sum_{\ell,|k-\ell|<r}\left(\int_{|u|>2^m}|2^{-m}u|^{1-r}\frac{(1+u^2)^{\alpha/2}}{u}\,\mathrm{d}u\right)^2\right)^{1/2}\le 2\mathfrak{C}(2r-1)\|h_X\|_AD_m^{\alpha+1/2}.\end{aligned}$$

The two bounds yield $M_1=Kc_nD_m^{\alpha+1/2}$.

Applying Talagrand Inequality with $\epsilon=1/2$ and $c_n=c_0\sqrt{n}/\log(n)$ gives the result for $\nu_{n,2}^{(2)}$, following the same lines as for $\nu_{n,1}$. □

References

Akakpo, N. and Durot, C. (2010). Histogram selection for possibly censored data. *Mathematical Methods of Statistics*, 19:189–218.

Antoniadis, A., Gregoire, G., and Nason, G. (1999). Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Stat. Soc., Ser. B*, 61:63–84.

Birgé, L. (1999). An alternative point of view on Lepski's method. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer.

Brunel, E. and Comte, F. (2005). Penalized contrast estimation of density and hazard rate with censored data. *Sankhya*, 67:441–475.

Brunel, E. and Comte, F. (2008). Adaptive estimation of hazard rate with censored data. *Communications in Statistics, Theory and methods*, 37:1284–1305.

Comte, F., Gaiffas, S., and Guilloux, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47:1171–1196.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.*, 34(3):431–452.

Comte, F., Samson, A., and Stirnemann, J. J. (2015). Hazard estimation for censored data contaminated with additive measurement error: Application to length of pregnancy. *MAP5 2015-18*.

Dattner, I., Goldenshluger, A., and Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.*, 39:2477–250.

de Boor, C. (2001). *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, revised edition.

Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.*, 56(1):19–47.

DeVore, R. and Lorentz, G. (1993). *Constructive approximation.* Springer-Verlag.

Dohler, S. and Ruschendorf, L. (2002). Adaptive estimation of hazard functions. *Probab. Math. Statist.*, 22:355–379.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.

Fan, J. and Koo, J.-Y. (2002). Wavelet deconvolution. *Information Theory, IEEE Transactions on*, 48:734–747.

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

Lacour, C. (2008). Least squares type estimation of the transition density of a particular hidden markov chain. *Electronic Journal of Statistics*, 2:1–39.

Li, L. (2007). On the minimax optimality of wavelet estimators with censored data. *J. Statist. Plann. Inference*, 137:1138–1150.

Li, L. (2008). On the block thresholding wavelet estimators with censored data. *J. Multivariate Anal.*, 99:1518–1543.

Massart, P. (2003). *Concentration Inequalities and Model selection*, volume 1896 of *Lecture Notes in Mathematics.* Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053.

Plancade, S. (2011). Model selection for hazard rate estimation in presence of censoring. *Metrika*, 4:313–347.

Reynaud-Bouret, P. (2006). Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, 12:633–661.

Stefanski, L. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184.