NONPARAMETRIC CONDITIONAL CUMULATIVE DISTRIBUTION FUNCTION ESTIMATION IN PRESENCE OF CENSORING

E. BRUNEL[(*),(1)], F. COMTE[(**),(2)] & C. LACOUR[(**),(3)]

Abstract. Consider an i.i.d. sample (X_i, Y_i), $i = 1, \dots, n$ of observations and denote by $F(x, y)$ the conditional cumulative distribution function of Y_i given $X_i = x$. We provide a data driven nonparametric strategy to estimate F. We prove that, in term of the integrated mean square risk on a compact set, our estimator performs a squared-bias variance compromise. We deduce from this an upper bound for the rate of convergence of the estimator, in a context of anisotropic function classes. A lower bound for this rate is also proved. Then our procedure can be adapted to positive censored random variables Y_i's, i.e. when only $Z_i = \inf(Y_i, C_i)$ and $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$ are observed, for an i.i.d. censoring sequence $(C_i)_{1\leq i\leq n}$ independent of $(X_i, Y_i)_{1\leq i\leq n}$. Simulation experiments illustrate the method.

April 2008

AMS (2000) subject classification. 62N02, 62G07.

Keywords. Adaptive estimation. Censored data. Conditional distribution function. Nonparametric methods.

1. Introduction

Consider an i.i.d. sample $(X_i, Y_i)_{1\leq i\leq n}$ of real-valued random variables with common probability density function (pdf) $f_{(X,Y)}$. In presence of a covariable, it is of interest to consider the conditional distribution function of Y given $X = x$ defined for all real y and x such that the marginal density $f_X(x)$ of X is strictly positive, by

$$F(x, y) = \frac{\int_{-\infty}^{y} f_{(X,Y)}(x, u)du}{f_X(x)} = \int_{-\infty}^{y} \pi(x, u)du, \tag{1}$$

where $\pi(x, y)$ denotes the conditional density of Y given $X = x$.

The conditional distribution function is often useful in reliability or in survival analysis, because of getting involved in many applications. For instance, the conditional survival function $S(x, y) = 1 - F(x, y)$ is of interest, either by itself, or as denominator for conditional hazard rate estimation (defined by $h(x, y) = \pi(x, y)/S(x, y)$). Conditional quantiles can also be deduced from $F(x, y)$ by (pseudo)-inversion given x of the function $F(x, y)$ and the same procedure may be applied to the estimator of F to find conditional quantile estimators. In brief, the aim of the paper is to provide a nonparametric strategy to estimate F from the observations.

(*) I3M Montpellier 2 University, France.
(**) MAP5, Paris Descartes University, France.
(1) email: ebrunel@math.univ-montp2.fr,
(2) email: fabienne.comte@univ-paris5.fr,
(3) email: claire.lacour@math-info.univ-paris5.fr.

It must be well understood that a function such as F behaves like a distribution function only with respect to (w.r.t.) the y-variable but not in the x-direction. It is thus an interesting nonparametric estimation problem to find an estimator that takes this asymmetry into account.

The central formula in (1) suggests that a Nadaraya-Watson strategy building an estimator as the ratio of an estimator of $\int_{\infty}^{y} f_{(X,Y)}(x,u)du$ divided by an estimator of $f_X(x)$ is conceivable: this strategy is studied in Ould-Saïd (2006). But estimators resulting of such methods have the drawback of precisely involving a ratio, with a denominator which can be small. This is the reason why we rather provide a regression-type strategy based on a mean square contrast. By doing so, we can take advantage of tools developed for standard regression by Baraud et al. (2001), or for conditional density estimation by Brunel et al. (2008) in the same setting. We propose a simple data-driven strategy which both builds a collection of projection estimators on finite dimensional spaces and selects the adequate space by penalization of the mean square contrast. Then we prove that the resulting estimator asymptotically reaches the adequate nonparametric rate, in a data driven way. It is worth mentioning that the projection spaces need not to be the same in both x and y-directions, and thus, it allows to be adaptive in x and on a pre-selected space in y. Thus we can derive rates that take this into account.

We consider moreover the case of positive censored variables Y_i's which corresponds to the context of lifetime studies. In many situations, such as medical trials or reliability systems, the lifetimes may not be completely observed and instead only censored lifetimes are available. Let $(C_i)_{1\leq i\leq n}$ be an i.i.d. sequence of positive censoring variables, with cumulative distribution function (cdf) G, independent of $(X_i, Y_i)_{1\leq i\leq n}$. In this context, the observations are

$$(X_i, Z_i, \delta_i)_{1\leq i\leq n}, \quad Z_i = \inf(Y_i, C_i), \ \delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}. \tag{2}$$

Then, we can provide both modified contrast and penalty following the transformation device proposed by Koul *et al.* (1981) for censoring correction (see also Fan and Gijbels (1994)). Bitouzé et al. (1999)'s results as already exploited in Brunel and Comte (2005) allow an adequate control when substituting to a cdf its Kaplan-Meier estimator in presence of censoring.

Our results can be compared to those obtained by Stute (1986) in the uncensored context and Dabrowska (1989) in the censoring case for the Beran's estimator which can be seen as a generalized Kaplan-Meier estimator in the presence of covariables. Both works are partly based on kernel strategies, and assume some particular regularity w.r.t. x (twice differentiability of F). Their strategies are not adaptive w.r.t. the bandwidth. In this sense, we substantially improve the statistical strategy (general regularity context and adaptive estimator). Stute (1986) proves a central limit theorem and Dabrowska (1989) gives a control of a uniform supremum-type norm. Note also that Gonzalez-Manteiga and Cadarso-Suarez (1994) give an almost sure representation as a sum of independant variables of the Beran's estimator. Our estimator is different from theirs in the sense that it is adaptive w.r.t. the unknown regularity. Moreover, we do not control the same type of risk since we use a global integrated $\mathbb{L}^2$-type norm on a compact set. Some other ideas were later developed by Li and Doss (1995), Ducharme and Mint El Mouvid (2001) and Hall et al. (1999). Li and Doss (1995) present nonparametric methods which are very powerful but lack of adaptivity w.r.t. the conditioning variable. On the other hand, Hall et al.

(1999) discuss local linear methods and methods for which the conditional cdf estimate is constrained to lie between 0 and 1 and to be monotone increasing, a property which is not necessarily verified by our estimator.

The plan of the paper is the following. First we describe in Section 2 the uncensored case: the estimation procedure (Section 2.1), the model collection and associated assumptions (Section 2.2) and the upper risk bound (Section 2.4). The lower bound is proved to be in adequation with our rate in Section 2.5. The censored case is described in Section 3. The method in both uncensored and censored cases is illustrated via simulations and examples in Section 4. The proofs are gathered in Section 5.

2. The uncensored case

In this section, we assume that (X_i, Y_i), $i = 1, \ldots n$ are independent and identically distributed pairs of variables and are completely observed. We estimate the conditional distribution function $F(x, y)$ of Y_i given $X_i = x$ on a given compact set $A = A_1 \times A_2 \subset \mathbb{R}^2$ only. Obviously, the conditional distribution function F belongs to the space of bounded and square integrable functions on the compact set $A = A_1 \times A_2$ denoted by $L^\infty \cap L^2(A)$. Moreover, we require that the following assumption is fulfilled:

[$\mathcal{A}1$] The density f_X verifies $\|f_X\|_\infty := \sup_{x\in A_1} |f_X(x)| < \infty$ and there exists a positive real f_0 such that, for all x in A_1, $f_X(x) \geq f_0$.

2.1. **Estimation procedure.** In order to estimate F, we need to introduce a collection $\{S_m, m \in \mathcal{M}_n\}$ of projection spaces: S_m is called a model and $\mathcal{M}_n$ is a set of multi-indexes (see the examples in Section 2.2). For each $m = (m_1, m_2)$, S_m is a space of functions with support in A defined by using two spaces: $S^{(1)}_{m_1}$ and $S^{(2)}_{m_2}$ which are subspaces of $(L^2 \cap L^\infty)(\mathbb{R})$ respectively spanned by two orthonormal bases $(\varphi^{m_1}_j)_{j\in J_{m_1}}$ with $|J_{m_1}| = D^{(1)}_{m_1}$ and $(\psi^{m_2}_k)_{k\in K_{m_2}}$ with $|K_{m_2}| = D^{(2)}_{m_2}$. For all j and all k, the supports of φ^m_j and ψ^m_k are respectively included in A_1 and A_2. Here j and k are not necessarily integers, they can be couples of integers as in the case of a piecewise polynomial space, see Section 2.2. Then, we define

$$S_m = S^{(1)}_{m_1} \otimes S^{(2)}_{m_2} = \{t, \quad t(x, y) = \sum_{j\in J_{m_1}} \sum_{k\in K_{m_2}} a^m_{j,k} \varphi^{m_1}_j(x) \psi^{m_2}_k(y),\ a^m_{j,k} \in \mathbb{R}\}.$$

Let, for $i = 1, 2$, $\mathcal{S}^{(i)}_n$ denote the largest space among the $S^{(i)}_{m_i}$'s. We denote by $\mathcal{D}^{(i)}_n$ its dimension and by $(\varphi^n_j)_{j\in\mathcal{J}_n}$, $(\psi^n_k)_{k\in\mathcal{K}_n}$ their respective bases.

These spaces correspond to the model collection used in Brunel et al. (2008) for conditional density estimation.

Now, we introduce the following contrast $\gamma^0_n(t)$, for any function $t \in S_m$:

$$\gamma^0_n(t) = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} (t^2(X_i, y) - 2t(X_i, y)\mathbb{1}_{\{Y_i \leq y\}})dy. \tag{3}$$

The contrast can be justified simply. Indeed, we have, for $\pi(x,y)$ denoting the conditional density of Y_1 given $X_1 = x$,

$$\begin{aligned}
\mathbb{E}(\gamma_n^0(t)) &= \mathbb{E}\int \left(t^2(X_1,y) - 2t(X_1,y)\mathbb{1}_{\{Y_1\le y\}}\right) dy \\
&= \iint t^2(x,y)f_X(x)dxdy - 2\iint \left(\int t(x,y)\mathbb{1}_{\{u\le y\}}dy\right)\pi(x,u)f_X(x)dxdu \\
&= \|t\|_f^2 - 2\iint t(x,y)\left(\int \pi(x,u)\mathbb{1}_{\{u\le y\}}du\right) f_X(x)dxdy \\
&= \|t\|_f^2 - 2\langle t,F\rangle_f = \|t-F\|_f^2 - \|F\|_f^2
\end{aligned}$$

where $\|t\|_f^2 = \iint_A t^2(x,y)f_X(x)dxdy$ and $\langle s,t\rangle_f = \iint s(x,y)t(x,y)f_X(x)dxdy$.

Therefore $\gamma_n^0(t)$ is the empirical counterpart of $\|t-F\|_f^2 - \|F\|_f^2$ and thus minimizing it leads to minimize $\|t-F\|_f$ in mean. Moreover, under assumption $[\mathcal{A}1]$, the norm $\|.\|_f$ is equivalent to the usual $\mathbb{L}^2$ norm on A ($\|t\|^2 = \iint_A t^2(x,y)dxdy$).

This contrast is new and its originality actually stands in the fact that it is a mean square contrast w.r.t. the X_i's and a projection contrast w.r.t. y. We emphasize the particular choice that can be made here: we take advantage of the fact that F has a distribution function behavior with respect to y. No model selection is required in the y-direction. Indeed, if $t(x,y) = t_2(y)$ does not depend of x, then $\gamma_n^0(t) = \gamma_n^0(t_2) = \int t_2^2(y)dy - 2\int t_2(y)\hat{F}_{Y,n}(y)dy$ where $\hat{F}_{Y,n}(y) = (1/n)\sum_{i=1}^n \mathbb{1}_{\{Y_i\le y\}}$ is the standard empirical distribution function of y. This explains why the dimension w.r.t. y is fixed *a priori* to be the greatest as possible. Thus, instead of minimizing the contrast over classes of functions belonging to $S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$, we can directly select the second space. Therefore, the conditional distribution function estimator is defined by:

$$\hat{F}_{m_1} = \arg\min_{t\in S_{m_1}^{(1)}\otimes\mathcal{S}_n^{(2)}} \gamma_n^0(t), \tag{4}$$

where the minimization holds in the sense explained in Section 2.3.

It is noteworthy that this is different of the conditional density estimator described in Brunel et al. (2008) where both x- and y-dimensions have to be relevantly selected (and are constrained in the censored case to be the same).

Now, the estimator $\hat{F}_{m_1}$ is likely to be close of F_{m_1}, the $\mathbb{L}^2$-orthogonal projection of F on $S_{m_1}^{(1)} \otimes \mathcal{S}_n^{(2)}$, so that the following decomposition of the risk holds:

$$\mathbb{E}(\|\hat{F}_{m_1} - F\|^2) = \|F_{m_1} - F\|^2 + \mathbb{E}(\|\hat{F}_{m_1} - F_{m_1}\|^2).$$

The term $\|F_{m_1} - F\|^2$ is a deterministic squared bias term, which decreases when $D_{m_1}^{(1)} = \dim(S_{m_1}^{(1)})$ increases. The term $\mathbb{E}(\|\hat{F}_{m_1} - F_{m_1}\|^2)$ is often called a variance term and generally increases with $D_{m_1}^{(1)}$. This explains why the model $S_{m_1}^{(1)}$ must be selected in order to perform an adequate squared bias/variance compromise. Moreover, this term does not depend on $\mathcal{S}_n^{(2)}$ and in particular on its dimension: this explains why we can choose directly the largest space of the collection in the second direction.

Thus, to perform model selection, we set

$$\hat{m}^0 = \arg\min_{m_1\in\mathcal{M}_n} \{\gamma_n^0(\hat{F}_{m_1}) + \mathrm{pen}^0(m_1)\} \tag{5}$$

where pen^0 is a penalty function to be specified later. Then we can define

$$\tilde{F} = \hat{F}_{\hat{m}^0}.$$

2.2. **Assumptions on the models and examples.** The models we have in mind all fulfill the following key properties:

[M1] $\mathcal{D}_n^{(1)} \le n$.
[M2] There exist positive a real ϕ_1 such that, for all u in $S_{m_1}^{(1)}$, $\|u\|_\infty^2 \le \phi_1 D_{m_1}^{(1)} \int u^2$.
[M3] $D_{m_1}^{(1)} \le D_{m_1'}^{(1)} \Rightarrow S_{m_1}^{(1)} \subset S_{m_1'}^{(1)}$.

Note that assumption [M2] is equivalent to

$$\exists \phi_1 > 0, \| \sum_{j\in J_{m_1}} (\varphi_j^{m_1})^2 \|_\infty \le \phi_1 D_{m_1}^{(1)}.$$

Therefore, the spaces on which the estimators are built are $S_m = S_{m_1}^{(1)} \otimes \mathcal{S}_n^{(2)} = \{t, \quad t(x,y) = \sum_{j\in J_{m_1}} \sum_{k\in\mathcal{K}_n} a_{j,k}^m \varphi_j^{m_1}(x)\psi_k^n(y), \ \ a_{j,k}^m \in \mathbb{R}\}$. The third assumption [M3] ensures that, for m and m' in $\mathcal{M}_n$, $S_m + S_{m'}$ is included in a model (since $S_m + S_{m'} \subset S_{m''}$ with $D_{m_1''}^{(1)} = \max(D_{m_1}^{(1)}, D_{m_1'}^{(1)})$ and $D_{m_2''}^{(2)} = \mathcal{D}_n^{(2)}$). We denote by $\mathcal{S}$ the space with maximal dimension among the $(S_m)_{m\in\mathcal{M}_n}$. Thus for all m in $\mathcal{M}_n$, $S_m \subset \mathcal{S}$. Lastly, for $t \in S_m$, we denote by $\|t\|^2 = \iint t^2(x,y)dxdy$ and $\|t\|_\infty = \sup_{(x,y)\in A_1\times A_2} |t(x,y)|$.

Now, we show that Assumptions [M1]–[M3] are not too restrictive. Indeed, they are verified for the spaces $S_{m_i}^{(i)}$ $(i = 1,2)$ spanned by the following bases (see Barron et al. (1999)):

- Trigonometric basis: for $A_1 = [0,1]$, $\mathrm{span}(\varphi_0, \ldots, \varphi_{m_1-1})$ with $\varphi_0 = \mathbb{1}_{[0,1]}$, $\varphi_{2j}(x) = \sqrt{2}\cos(2\pi jx)\ \mathbb{1}_{[0,1]}(x)$, $\varphi_{2j-1}(x) = \sqrt{2}\sin(2\pi jx)\mathbb{1}_{[0,1]}(x)$ for $j \ge 1$. For this model $D_{m_1}^{(1)} = m_1$ and $\phi_1 = 2$ hold.
- Histogram basis: for $A_1 = [0,1]$, $\mathrm{span}(\varphi_1, \ldots, \varphi_{2^{m_1}})$ with $\varphi_j = 2^{m_1/2}\mathbb{1}_{[(j-1)/2^{m_1}, j/2^{m_1}[}$ for $j = 1, \ldots, 2^{m_1}$. Here $D_{m_1}^{(1)} = 2^{m_1}$, $\phi_1 = 1$.
- Regular piecewise polynomial basis: for $A_1 = [0,1]$, polynomials of degree $0, \ldots, r$ (where r is fixed) on each interval $[(l-1)/2^D, l/2^D[, l = 1, \ldots, 2^D$. In this case, $m_1 = (D,r)$, $J_m = \{j = (l,d),\ 1 \le l \le 2^D, 0 \le d \le r\}$, $D_{m_1}^{(1)} = (r+1)2^D$. We can put $\phi_1 = \sqrt{r+1}$.
- Regular wavelet basis: $\mathrm{span}(\Psi_{lk}, l = -1, \ldots, m_1, k \in \Lambda(l))$ where $\Psi_{-1,k}$ points out the translates of the father wavelet and $\Psi_{lk}(x) = 2^{l/2}\Psi(2^l x - k)$ where Ψ is the mother wavelet. We assume that the support of the wavelets is included in A_1 and that Ψ_{-1} belongs to the Sobolev space W_2^r.

We repeatedly use in the proofs the following property. Let $(\psi_k^{(m_2)}, 1 \le k \le D_{m_2}^{(2)})$ be an orthonormal basis of $S_{m_2}^{(2)}$, then, for any bounded function $\xi \in \mathbb{L}^2(A_2)$,

$$\sum_{k=1}^{D_{m_2}^{(2)}} \left(\int_{A_2} \psi_k^{(m_2)}(y)\xi(y)dy \right)^2 \le \int_{A_2} \xi^2(y)dy \le \ell(A_2) \sup_{y\in A_2} \xi^2(y), \tag{6}$$

where $\ell(A_2)$ denotes the length of A_2. Indeed, the left-hand-side of (6) is simply the $\mathbb{L}^2(A_2)$ norm of the orthogonal projection of ξ on $S_{m_2}^{(2)}$, and this is less than the $\mathbb{L}^2(A_2)$ norm of ξ.

2.3. **About the definition of the estimator.** We discuss now the definition of $\hat{F}_{m_1}$ given by (4). Let $t(x,y) = \sum_{j\in J_{m_1}} \sum_{k\in\mathcal{K}_n} a_{j,k}\varphi_j^{m_1}(x)\psi_k^n(y)$ be a function in $S_{m_1}^{(1)} \otimes \mathcal{S}_n^{(2)}$. Then,

(7)
$$\frac{\partial \gamma_n^0(t)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow \sum_{j\in J_{m_1}} a_{j,k_0}\frac{1}{n}\sum_{i=1}^n \varphi_j^{m_1}(X_i)\varphi_{j_0}^{m_1}(X_i) = \frac{1}{n}\sum_{i=1}^n \varphi_{j_0}^{m_1}(X_i)\int \psi_{k_0}^n(y)\mathbb{1}_{\{Y_i\le y\}}dy,$$

which implies that

$$\forall j_0 \forall k_0 \quad \frac{\partial \gamma_n^0(t)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow G_{m_1}A_{m_1} = \Upsilon_{m_1},$$

where A_{m_1} denotes the matrix $(a_{j,k})_{j\in J_{m_1}, k\in\mathcal{K}_n}$,

$$G_{m_1} = \left(\frac{1}{n}\sum_{i=1}^n \varphi_j^{m_1}(X_i)\varphi_l^{m_1}(X_i)\right)_{j,l\in J_{m_1}}$$

and

$$\Upsilon_{m_1} = \left(\frac{1}{n}\sum_{i=1}^n \varphi_j^{m_1}(X_i)\int \psi_k^n(y)\mathbb{1}_{\{Y_i\le y\}}dy\right)_{j\in J_{m_1}, k\in\mathcal{K}_n}.$$

In fact, we cannot define a unique minimizer of the contrast $\gamma_n^0(t)$, since G_{m_1} is not necessarily invertible. However, the following proposition still enables us to define an estimator:

Proposition 1.

$$\forall j_0 \forall k_0 \quad \frac{\partial \gamma_n^0(t)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow \forall u, (t(X_i,u))_{1\le i\le n} = P_{\mathcal{W}}\left(\left(\sum_k \psi_k^n(u)\int \psi_k^n(y)\mathbb{1}_{\{Y_i\le y\}}dy\right)_{1\le i\le n}\right)$$

where $P_{\mathcal{W}}$ denotes the orthogonal projection on $\mathcal{W} = \{(t(X_i,u))_{1\le i\le n},\ t \in S_{m_1}^{(1)} \otimes \mathcal{S}_n^{(2)}\}$ with the euclidian scalar product $\langle\ .\ ,\ .\ \rangle_{\mathbb{R}^n}$ in $\mathbb{R}^n$.

Remark 1. It follows from Proposition 1 that the minimization of $\gamma_n^0(t)$ leads to a unique vector $(\hat{F}_{m_1}(X_i,u))_{1\le i\le n}$ defined as the projection of the vector $\left(\sum_k \psi_k^n(u)\int \psi_k^n(y)\mathbb{1}_{\{Y_i\le y\}}dy\right)_{1\le i\le n}$ on $\mathcal{W}$. The associated function $\hat{F}_{m_1}(.,.)$ is not defined uniquely but we can choose a function $\hat{F}_{m_1}$ in $S_{m_1}^{(1)}\otimes\mathcal{S}_n^{(2)}$ whose values at (X_i,u) are fixed according to Proposition 1. For the sake of simplicity, we use the notation (4) but one must keep in mind that the underlying function is a theoretical tool: the estimator is actually the vector $(\hat{F}_{m_1}(X_i,u))_{1\le i\le n}$.

Remark 2. As a consequence of remark 1, we do not compute a standard $\mathbb{L}^2$-risk but instead the following empirical mean integrated squared error $\mathbb{E}\|F - \tilde{F}\|_n^2$ where $\|.\|_n$ is the empirical norm defined by

$$\|t\|_n = \left(\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} t^2(X_i,y)dy\right)^{1/2}. \tag{8}$$

This norm is the natural distance in this problem and we can notice that if t is deterministic with support included in $A_1 \times \mathbb{R}$

$$f_0\|t\|^2 \leq \mathbb{E}\|t\|_n^2 = \|t\|_f^2 \leq \|f_X\|_\infty\|t\|^2$$

and then the mean of this empirical norm is equivalent to the $\mathbb{L}^2(A)$ norm $\|.\|$.

2.4. **Main result.** For a function h and a subspace S, let

$$d(h,S) = \inf_{g\in S}\|h-g\| = \inf_{g\in S}\left(\iint |h(x,y)-g(x,y)|^2dxdy\right)^{1/2}.$$

With an inequality of Talagrand (1996), we can prove the following result in the uncensored case.

Theorem 1. *We consider the uncensored model satisfying Assumption* [$\mathcal{A}1$]. *We consider* $\tilde{F}$ *the estimator on* A *of the conditional distribution function* F *described in Section 2.1 with model collections verifying Assumptions* [M1]-[M2]-[M3], *and the following penalty:*

$$\mathrm{pen}^0(m_1) = K_0\ell(A_2)\frac{D^{(1)}_{m_1}}{n}, \tag{9}$$

where K_0 *is a numerical constant. Then*

$$\mathbb{E}\|F\mathbb{1}_A - \tilde{F}\|_n^2 \leq C\inf_{m_1\in\mathcal{M}_n}\{d^2(F\mathbb{1}_A, F_{m_1}) + \mathrm{pen}^0(m_1)\} + \frac{C'}{n} \tag{10}$$

where $C = \max(5\|f_X\|_\infty, 6)$ *and* C' *is a constant depending on* $\phi_1, \ell(A_2), f_0$.

The penalty (9) deserves some comments. First, the constant K_0 in the penalty is purely numerical and calibrated via simulations. Moreover, note that inequality (10) holds for any penalty $\mathrm{pen}^0(.)$ such that $\mathrm{pen}^0(m_1) \geq K_0\ell(A_2)D^{(1)}_{m_1}/n$.

We can deduce from Theorem 1 the rate of convergence of the risk. For that purpose, assume that F restricted to A belongs to the anisotropic Besov space on A with regularity $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. Let us recall the definition of $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$. Let e_1 and e_2 be the canonical basis vectors in $\mathbb{R}^2$ and for $i = 1, 2$, $A^r_{i,h} = \{x \in \mathbb{R}^2; x, x + he_i, \ldots, x + rhe_i \in A\}$. Next, for x in $A^r_{i,h}$, let

$$\Delta^r_{i,h}g(x) = \sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}g(x+khe_i) \tag{11}$$

the rth difference operator with step h. For $t > 0$, the directional moduli of smoothness are given by

$$\omega_{r_i,i}(g,t) = \sup_{|h|\leq t}\left(\int_{A^{r_i}_{i,h}} |\Delta^{r_i}_{h,i}g(x)|^2dx\right)^{1/2}.$$

We say that g is in the Besov space $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ if $\sup_{t>0}\sum_{i=1}^2 t^{-\alpha_i}\omega_{r_i,i}(g,t) < \infty$ for r_i integers larger than α_i.

The estimation procedure may allow an adaptation of the approximation space to each directional regularity. It happens that in the y-direction, the greatest space is directly chosen. Thus, we just have to select a relevant $S^{(1)}_{m_1}$.

Corollary 1. *Assume that F restricted to A belongs to the anisotropic Besov space $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ with regularity $\boldsymbol{\alpha}=(\alpha_1,\alpha_2)$ such that $\alpha_1>1/2$ and $\alpha_2>1$. We consider the spaces described in Subsection 2.2 (with the regularity r of the polynomials and the wavelets larger than α_i-1). Then, for $\mathcal{D}^{(2)}_n\geq\sqrt{n}$ and under the assumptions of Theorem 1,*

$$\mathbb{E}\|F\mathbb{1}_A-\tilde{F}\|_n^2=O(n^{-\frac{2\alpha_1}{2\alpha_1+1}}).$$

Thus we obtain a rate of convergence which would be standard for the estimation of a function of one variable with regularity α_1. We need to check the associated lower bound.

Remark 3. The empirical norm is the more natural in this problem, but if we were interested in a L^2 control of the risk, we may modify the estimation procedure as follows:

$$\tilde{F}^*=\begin{cases}\tilde{F} & \text{if } \|\tilde{F}\|\leq k_n\\ 0 & \text{else}\end{cases}\tag{12}$$

with $k_n=n^\lambda$ for some well-chosen $\lambda\in]0,1[$. We may prove in this framework a result similar to Theorem 1 but bounding $\mathbb{E}\|\tilde{F}^*-F\mathbb{1}_A\|^2$ instead of its empirical version, see Lacour (2007).

2.5. **Lower bound.** We denote by $\|.\|_A$ the norm in $L^2(A)$, i.e. $\|g\|_A=\left(\int_A|g|^2\right)^{1/2}$. The norm in the Besov space $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$ is

$$\|F\|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)}=\|F\|_A+|F|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)}$$

where

$$|F|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)}=\sup_{0\leq t\leq 1}\sum_{i=1}^2 t^{-\alpha_i}\sup_{|h|\leq t}\left\|\Delta^{r_i}_{i,h}F\right\|_{A^{r_i}_{h,i}}$$

with $\Delta^{r_i}_{h,i}F(x,y)$ defined as in (11) and r_i an integer strictly larger than α_i. We set

$$\mathcal{B}=\{F \text{ conditional distribution function on } \mathbb{R}^2 \text{ such that } \|F\|_{B^{\boldsymbol{\alpha}}_{2,\infty}(A)}\leq L\}$$

and $\mathbb{E}_{F,f}$ the expectation corresponding to the distribution of $(X_1,Y_1),\ldots,(X_n,Y_n)$ if the true conditional distribution function of (X,Y) is F and the true marginal density of X is f.

Theorem 2. *We assume that $\alpha_2>1$. Then, for all f uniformly bounded, there exists a positive constant C such that, if n is large enough,*

$$\inf_{\hat{F}_n}\sup_{F\in\mathcal{B}}\mathbb{E}_{F,f}\|\hat{F}_n-F\|_A^2\geq Cn^{-\frac{2\alpha_1}{2\alpha_1+1}}$$

where the infimum is taken over all estimators $\hat{F}_n$ of F based on the couples of observations $(X_1,Y_1),\ldots,(X_n,Y_n)$.

As a conclusion, we deduce from Theorem 2 that the estimate $\tilde{F}=\hat{F}_{\hat{m}_0}$ has the optimal rate of convergence and is thus adaptive and minimax.

3. The censored case

In this section, we consider the censoring framework (2). Remember that the censoring variables (C_i), $i = 1\ldots,n$ are independent and identically distributed positive variables, with cdf G; the C_i's are independent of the sequence (X_i, Y_i), $i = 1\ldots,n$. In this context, we observe (X_i, Z_i, δ_i) where $Z_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{Y_i \le C_i\}}$. As we have in mind lifetime studies, we only consider positive variables Y_i and C_i, $i = 1\ldots,n$.

Here, we have to mention that the constraints added to extend our results from uncensored to censored case are weak. In the conditional density estimation studied by Brunel et al. (2008), both x and y-dimensions need to be the same, which is a pure technical (and unnatural) constraint. With the cumulative conditional distribution, since model selection has only to be performed in the x-direction, this drawback is avoided. In any case, the theoretical penalty has to be replaced by a random one since it depends on the (unknown) lower bound of the survival function of the censoring variable on A_2 as explained in Section 4.

3.1. Estimation procedure.

In the censored case, it is classical to assume, in addition of $[\mathcal{A}1]$ the following:

$[\mathcal{A}2\ ]$ For all $y \in A_2$, $1 - G(y) \ge c_G > 0$.

Note that the lower bound condition $[\mathcal{A}2]$ is required only on the compact set A_2, which is a very mild assumption.

Now, let us take into account the fact that the Y_i's may be censored in the definition of the contrast. We use a standard transformation of the data (see Fan and Gijbels (1994)), based on the weights:

$$w_i = \frac{\delta_i}{\bar{G}(Z_i)} \tag{13}$$

where $\bar{G} = 1 - G$ is the survival function associated with the censoring variables. They are replaced by an empirical version and we choose the contrast function

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n} \int_{\mathbb{R}} (t^2(X_i, y) - 2\hat{w}_i t(X_i, y)\mathbb{1}_{\{Z_i \le y\}})dy, \quad \hat{w}_i = \frac{\delta_i}{\hat{\bar{G}}(Z_i)}. \tag{14}$$

Note that $\gamma_n^0(t)$ and $\gamma_n(t)$ coincide by defining the weights $w_i = 1$ and $\hat{w}_i = 1$ if no censoring occurs (and then $Z_i = Y_i$).

Here $\hat{\bar{G}}$ is the Kaplan Meier estimator of the survival function $\bar{G}$ of the C_i's, modified in the way suggested by Lo et al. (1989), and defined by

$$\hat{\bar{G}}(y) = \prod_{i\ /\ Z_{(i)} \le y} \left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_{(i)}}. \tag{15}$$

Note that $\hat{\bar{G}}$ is a slight modification of the standard Kaplan-Meier estimator in order to satisfy the following useful property: $\hat{\bar{G}}(y) \ge 1/(n+1)$, $\forall y$. Moreover, it is a very good estimator of $\bar{G}$ on the interval A_2, in the sense of Lemma 1 in Section 5, provided $A_2 \subsetneq [0, \tau]$ where $\tau = \sup\{y, G(y) < 1\}$; this condition is ensured by Assumption $[\mathcal{A}2]$.

Therefore, we define

$$\hat{F}_{m_1} = \arg \min_{t \in S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n} \gamma_n(t), \tag{16}$$

where the minimization problem holds in the sense explained in Section 2.3 (by replacing $\int \psi^n_k(y) \mathbb{1}_{\{Y_i \leq y\}} dy$ by $\hat{w}_i \int \psi^n_k(y) \mathbb{1}_{\{Y_i \leq y\}} dy$), and lastly, we select the adequate dimension $\hat{m}$ by:

$$\hat{m} = \arg \min_{m_1 \in \mathcal{M}_n} \{\gamma_n(\hat{F}_{m_1}) + \mathrm{pen}(m_1)\} \text{ and } \tilde{F} = \hat{F}_{\hat{m}}, \tag{17}$$

where pen is a penalty function to be specified later.

3.2. **Risk bound.** In the censored case, the penalty coming out contains terms that must be estimated. We obtain the following result:

Theorem 3. *We consider the censored model satisfying Assumptions* $[\mathcal{A}1]$–$[\mathcal{A}2]$. *We consider* $\tilde{F}$ *the estimator of the conditional distribution* $F\mathbb{1}_A$ *described in (16)-(17) with models verifying Assumptions* [M1]-[M2]-[M3] *with* $\mathcal{D}^{(1)}_n \leq \sqrt{n}$. *We choose the following penalty:*

$$\mathrm{pen}(m) = K_0 \ell(A_2) \Xi \frac{D^{(1)}_{m_1}}{n}, \text{ with } \Xi = \frac{\phi_1}{f_0} \mathbb{E}\left(\frac{\delta_1}{\bar{G}^2(Z_1)}\right) \text{ or } \Xi = \frac{1}{c_G}, \tag{18}$$

where K_0 *is a numerical constant. Then*

$$\mathbb{E}\|F\mathbb{1}_A - \tilde{F}\|^2_n \leq C \inf_{m_1 \in \mathcal{M}_n} \{d^2(F\mathbb{1}_A, S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n) + \mathrm{pen}(m_1)\} + \frac{C'}{n}$$

where $C = \max(5\|f_X\|_\infty, 6)$ *and* C' *is a constant depending on* $\phi_1, \ell(A_2), f_0, c_G$.

Here the penalty involves quantities that should be estimated: c_G in the last case (see Section 4); in the first case, f_0, the lower bound of f_X the density of the X_i's on the interval A_1 (which is replaced by an estimator built as the minimum value of an estimator of f_X) and the expectation $\mathbb{E}(\delta_1/\bar{G}^2(Z_1))$ (which is replaced by an empirical mean $(1/n)\sum_{i=1}^n \delta_i/\hat{\bar{G}}^2(Z_i)$). Finally, a similar result as Theorem 3 would hold with the unknown penalty replaced by its empirical version but is omitted for sake of brevity. Note that the dimension of the greatest space for the x-direction is constrained to be of order $\sqrt{n}$ however no heavy condition is required against the uncensored data case.

4. Simulations

We study the estimation procedure by generating samples (X_i, Y_i) following three models:

- Example 1. Let $Y_i = b(X_i) + \varepsilon_i$, with ε_i i.i.d. $\mathcal{N}(0,1)$, X_i i.i.d. uniform $\mathcal{U}([0,1])$, $b(x) = 2x + 5$. We take $A = [0,1] \times [3, 9.5]$.
- Example 2. Hall et al. (1999)'s example. Let $Y_i = b(X_i) + \varepsilon_i$, with ε_i i.i.d., X_i i.i.d. uniform, both with density $1 - |x|$ on $[-1,1]$, $b(x) = 2\sin(\pi x)$. We take $A = [-0.9, 0.9] \times [-2.8, 2.8]$.
- Example 3. Given $X_i = x$, let Y_i follow the distribution $0.5\mathcal{N}(8-4x,1) + 0.5\mathcal{N}(8+4x,1)$. The X_i's are i.i.d. $\mathcal{U}([0,1])$. We take $A = [0,1] \times [2,14]$.

	Censoring	Basis	$n = 200$	$n = 500$	$n = 1000$
Example 1	0%	H	0.0489	0.0424	0.0292
		P	0.0944	0.0368	0.0194
		T	0.0782	0.0705	0.0587
		T*	0.0274	0.0204	0.0111
Example 1	20%	H	0.0729	0.0638	0.0569
	($c = 8.7$)	P	0.1481	0.0670	0.0425
		T	0.1339	0.0986	0.0885
		T*	0.0790	0.0451	0.0353
Example 1	40%	H	0.1190	0.0989	0.0936
	($c = 6.97$)	P	0.2284	0.1226	0.0887
		T	0.2131	0.2035	0.1808
		T*	0.1492	0.1411	0.1184
Example 2	0%	H	0.3119	0.1266	0.1193
		P	1.2081	1.1200	1.1016
		T	0.2167	0.1079	0.1040
		T*	0.1161	0.0320	0.0288
Example 3	0%	H	0.0864	0.0768	0.0549
		P	0.1627	0.0638	0.0317
		T	0.1242	0.1150	0.0974
		T*	0.0402	0.0319	0.0180
Example 3	20%	H	0.1686	0.1466	0.1368
	($c = 11.65$)	P	0.3236	0.1632	0.1110
		T	0.2254	0.1694	0.1557
		T*	0.1358	0.0820	0.0690

TABLE 1. MISE values for the estimator $\tilde{F}$, for $K = 500$ replications, n observations and bases: H=histogram, P= Piecewise polynomials, T = Trigonometric polynomials, T* = Trigonometric polynomial on truncated interval.

The sets $A = A_1 \times A_2$ are fixed intervals, roughly calibrated with respect to each distribution. In the first two cases, the conditional distribution function F is given by

$$F(x, y) = F_\varepsilon(y - b(x)),$$

with $b(x) = 2x + 5$ (example 1) or $b(x) = 2\sin(\pi x)$ (example 2) and F_ε is the distribution function of ε. For Example 3, we have $F(x, y) = 0.5\Phi(-(y-8+4x)) + 0.5\Phi(-(y-8-4x))$, where Φ is the $\mathcal{N}(0, 1)$ distribution function.

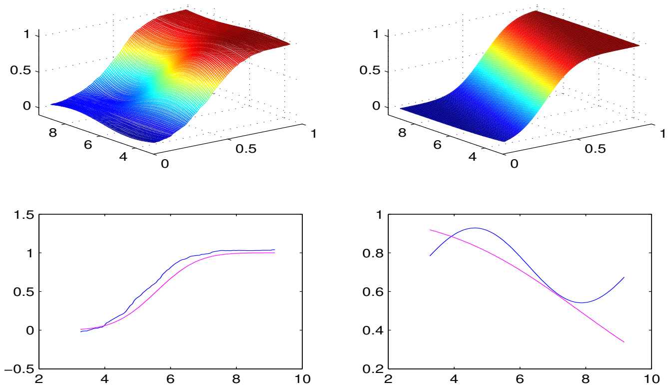

FIGURE 1. Plots of the estimated (top-left) with basis T⋆ (trigonometric polynomials with boundary truncation) and the true (top right) conditional distribution function and $y \mapsto F(x, y)$ (full line), $\tilde{F}(x, y)$ (dashed dotted line) for $x = 0.28$ (bottom-left) and $y = 6.48$ (bottom-right) with $n = 2000$ observations in Example 1 and with 20% of censoring.

The penalty is chosen as follows:

$$\ell(A_2) \max_{1\le i\le n} \left(\frac{1}{\hat{\hat{G}}(Z_i)} \right) \frac{D_{m_1}}{n}. \tag{19}$$

If the data are uncensored, the term $(\max(1/\hat{\hat{G}}(Z_i))$ is equal to one, and we recover the empirical version of (9) with constant K_0 calibrated as 1. In the censored case, this factor stands for $1/c_G$ and is quite great but it seems to be what is required for the estimation algorithm to give good results. The algorithm selects $D_m^{(1)}$ less than $\sqrt{n}$ and the other dimension $\mathcal{D}_n^{(2)}$ is fixed to the maximal value n. Three bases are considered, in accordance with the description of Section 2.2: the trigonometric basis denoted by T or T⋆ (see the explanation for $T^\star$ below, the histogram basis denoted by H, the piecewise polynomial basis with degree $r = 2$ denoted by P (and built by using the Legendre basis).

We compute the empirical MISE (Mean Integrated Squared Error) over $N = 500$ replications of the samples, by averaging over the paths $j = 1, \dots, N$, the quantities

$$\frac{\ell(A_1)\ell(A_2)}{K^2} \sum_{k=1}^{K} (\tilde{F}^{(j)}(x_k, y_k) - F(x_k, y_k))^2,$$

where $\ell(A_i)$ is the length of the interval A_i, $i = 1, 2$, $(x_k)_{1\le k\le K}$, $(y_k)_{1\le k\le K}$ are subdivisions of A_1 and A_2 respectively, and $\tilde{F}^{(j)}$ is the estimate associated to the jth sample path. Note that we compute $\mathbb{L}^2$-type errors in both x- and y-directions instead of using the empirical norm in the x-direction.

When censoring occurs, the C_i's are generated as exponential random variables $\mathcal{E}(c)$ with parameter c empirically adjusted to reach a given censoring rate (20% or 40%), namely $c = 6.97$ for Example 1 and censoring level 40% and $c = 8.7$ for Example 1 and

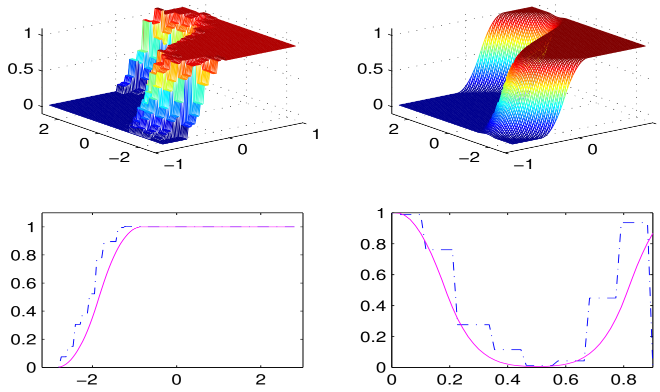

FIGURE 2. Plots of the estimated (top-left) with histogram basis and the true (top right) conditional distribution function and $y \mapsto F(x, y)$ (full line), $\tilde{F}(x, y)$ (dashed dotted line) for $x = -0.37$ (bottom-left) and $y = 1.10$ (bottom-right) with $n = 2000$ observations in Example 2 and without censoring.

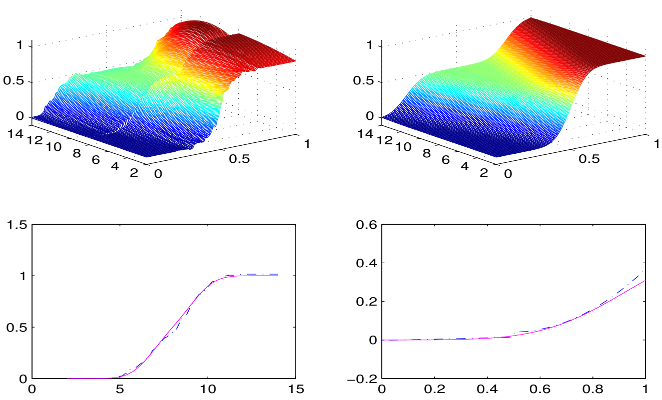

FIGURE 3. Plots of the estimated (top-left) with basis T (piecewise polynomials) and the true (top right) conditional distribution function and $y \mapsto F(x, y)$ (full line), $\tilde{F}(x, y)$ (dashed dotted line) for $x = 0.29$ (bottom-left) and $y = 7.94$ (bottom-right) with $n = 1500$ observations in Example 3 and without censoring.

censoring level 20%, $c = 11.65$ for Example 3 (censoring level 20%). The variables in Example 2 are not positive and therefore not included in the censoring study. Figures 1, 2 and 3 illustrate the appearance of our estimates. Note that it can be seen from Figure 2 that the histogram basis is very performant and that its localization properties are crucial. This is true for all examples.

Monte Carlo experiment results are reported in Table 1. As expected, censoring deteriorates the result, and a sample size increase improves them significantly. The best results are obtained for the basis $\mathrm{T}^\star$ that is the basis of trigonometric polynomials where only 90% of the points (x_k, y_k) are used to compute errors: we avoid like this boundary effect, and for $k = 1, \dots, 100$, this corresponds to MISE computed over the points $(x_k, y_k)_{6\le k\le 95}$, the plot in Figure 1 is given accordingly to the same points. This illustrates that the trigonometric basis is not localized but works globally on the domain of estimation; this implies important boundary effects. On the opposite, the histogram basis is always competitive probably because it is very localized, even for small sample size, whereas the piecewise polynomial basis requires larger sample sizes. Note that the piecewise polynomial basis fails in the case of example 2, probably because the true distribution is constant on subsets of the domain: thus, the degree of the polynomials should not be fixed but also selected.

5. Proofs

All proofs (except the proof of Theorem 2) are written in the general possibly censored setting. The uncensored case corresponds to weights $w_i = \hat{w}_i = 1$ and $Z_i = Y_i$ and we denote by $\mathrm{pen}(m)$ the general penalty function involved in Theorem 1 and 3.

5.1. **Proof of Proposition 1.** Here we prove

$$\forall j_0 \forall k_0, \frac{\partial \gamma_n^0(t)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow \forall u, (t(X_i,u))_{1\le i\le n} = P_{\mathcal{W}}\left(\left(\hat{w}_i \sum_k \psi_k^n(u) \int \psi_k^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\right)_{1\le i\le n}\right)$$

with notations of Proposition 1.
Equality (7) yields, by multiplying by $\psi_{k_0}^n(u)$,

$$\sum_{j\in J_{m_1}} a_{j,k_0} \sum_{i=1}^n \varphi_j^{m_1}(X_i)\psi_{k_0}^n(u)\varphi_{j_0}^{m_1}(X_i) = \sum_{i=1}^n \varphi_{j_0}^m(X_i)\hat{w}_i \int \psi_{k_0}^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\psi_{k_0}^n(u).$$

Then, we sum over k_0 in $\mathcal{K}_n$:

$$\sum_{i=1}^n t(X_i,u)\varphi_{j_0}^{m_1}(X_i) = \sum_{i=1}^n \sum_{k_0\in\mathcal{K}_n} \hat{w}_i \int \psi_{k_0}^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\psi_{k_0}^n(u)\varphi_{j_0}^{m_1}(X_i).$$

If we multiply this equality by $a'_{j_0,k}\psi_k^n(u)$ and if we sum over $k\in\mathcal{K}_n$ and $j_0\in J_{m_1}$, we obtain

$$\sum_{i=1}^n [t(X_i,u) - \sum_{k_0\in\mathcal{K}_n} \hat{w}_i \int \psi_{k_0}^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\psi_{k_0}^n(u)] \sum_{k\in\mathcal{K}_n}\sum_{j_0\in J_{m_1}} a'_{j_0,k}\varphi_{j_0}^{m_1}(X_i)\psi_k^n(u) = 0$$

i.e.

$$\sum_{i=1}^n [t(X_i,u) - \sum_{k_0\in\mathcal{K}_n} \hat{w}_i \int \psi_{k_0}^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\psi_{k_0}^n(u)] s(X_i,u) = 0$$

for all s in $S_{m_1}^{(1)} \otimes \mathcal{S}_n^{(2)}$. So the vector $(t(X_i,u) - \sum_{k\in\mathcal{K}_n} \hat{w}_i \int \psi_k^n(y)\mathbb{1}_{\{Z_i\le y\}}dy\psi_k^n(u))_{1\le i\le n}$ is orthogonal to each vector in $\mathcal{W}$. Since $t(X_i,y)$ belongs to $\mathcal{W}$, the proposition is proved.

5.2. **Proof of Theorem 1 and 3.** For ρ a real larger than 1, let

$$\Omega_\rho = \{\forall t \in \mathcal{S} \quad \|t\|_f^2 \le \rho \|t\|_n^2\}$$

We denote by F_{m_1} the orthogonal projection of F on $S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n$. Now,

$$\mathbb{E}\|\tilde{F} - F\mathbb{1}_A\|_n^2 = \mathbb{E}\left(\|\tilde{F} - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho}\right) + \mathbb{E}\left(\|\tilde{F} - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho^c}\right) \tag{20}$$

To bound the first term, we observe that for all s, t

$$\gamma_n(t) - \gamma_n(s) = \|t - F\|_n^2 - \|s - F\|_n^2 - 2\nu_n(t-s) - 2R_n(t-s)$$

where

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n \left\{ w_i \int t(X_i, y)\mathbb{1}_{\{Z_i \le y\}} dy - \int_{\mathbb{R}} t(X_i, y) F(X_i, y) dy \right\},$$

$$R_n(t) = \frac{1}{n}\sum_{i=1}^n \int t(X_i, y)\mathbb{1}_{\{Z_i \le y\}} dy \,[\hat{w}_i - w_i].$$

Since $\|t - F\|_n^2 = \|t - F\mathbb{1}_A\|_n^2 + \|F\mathbb{1}_{A^c}\|_n^2$, we can write

$$\gamma_n(t) - \gamma_n(s) = \|t - F\mathbb{1}_A\|_n^2 - \|s - F\mathbb{1}_A\|_n^2 - 2\nu_n(t-s) - 2R_n(t-s).$$

The definition of $\hat{m}$ gives, for some fixed $m_1 \in \mathcal{M}_n$,

$$\gamma_n(\tilde{F}) + \mathrm{pen}(\hat{m}) \le \gamma_n(F_{m_1}) + \mathrm{pen}(m_1).$$

And then

$$\begin{aligned} \|\tilde{F} - F\mathbb{1}_A\|_n^2 &\le \|F_{m_1} - F\mathbb{1}_A\|_n^2 + \mathrm{pen}(m_1) + 2R_n(\tilde{F} - F_{m_1}) \\ &\quad +2\nu_n(\tilde{F} - F_{m_1}) - \mathrm{pen}(\hat{m}) \\ &\le \|F_{m_1} - F\mathbb{1}_A\|_n^2 + \mathrm{pen}(m_1) + 2\|\tilde{F} - F_{m_1}\|_f \sup_{t \in B_f(\hat{m})} R_n(t) \\ &\quad +2\|\tilde{F} - F_{m_1}\|_f \sup_{t \in B_f(\hat{m})} \nu_n(t) - \mathrm{pen}(\hat{m}) \end{aligned}$$

where, for all m', $B_f(m') = \{t \in S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n + S^{(1)}_{m'} \otimes \mathcal{S}^{(2)}_n, \quad \|t\|_f = 1\}$. Let θ a real larger than 2ρ and $p(.,.)$ a function such that $2\theta p(m, m') \le \mathrm{pen}(m) + \mathrm{pen}(m')$. Then

$$\begin{aligned} \|\tilde{F} - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho} &\le \|F_{m_1} - F\mathbb{1}_A\|_n^2 + \frac{1}{\theta}\|\tilde{F} - F_{m_1}\|_f^2 \mathbb{1}_{\Omega_\rho} + 2\mathrm{pen}(m_1) \\ &\quad +2\theta \sum_{m' \in \mathcal{M}_n} \left[\sup_{t \in B_f(m')} \nu_n^2(t) - p(m_1, m')\right]_+ \mathbb{1}_{\Omega_\rho} \\ &\quad +2\theta \sup_{t \in B_f(\hat{m})} R_n^2(t) \mathbb{1}_{\Omega_\rho} \end{aligned} \tag{21}$$

But $\|\tilde{F} - F_{m_1}\|_f^2 \mathbb{1}_{\Omega_\rho} \le \rho\|\tilde{F} - F_{m_1}\|_n^2 \mathbb{1}_{\Omega_\rho} \le 2\rho\|\tilde{F} - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho} + 2\rho\|F\mathbb{1}_A - F_{m_1}\|_n^2$.

Then, inequality (21) becomes

$$\begin{aligned}
\|\tilde F - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho}\left(1-\frac{2\rho}{\theta}\right) &\le \left(1+\frac{2\rho}{\theta}\right)\|F_{m_1}-F\mathbb{1}_A\|_n^2 + 2\mathrm{pen}(m_1)\\
&\quad +2\theta\sum_{m'\in\mathcal{M}_n}\left[\sup_{t\in B_f(m')}\nu_n^2(t)-p(m_1,m')\right]_+\mathbb{1}_{\Omega_\rho}\\
&\quad +2\theta\sup_{t\in B_f(\hat m)} R_n^2(t)\mathbb{1}_{\Omega_\rho}
\end{aligned}$$

$$\begin{aligned}
\text{so}\quad \mathbb{E}\left(\|\tilde F - F\mathbb{1}_A\|_n^2 \mathbb{1}_{\Omega_\rho}\right) &\le \frac{\theta+2\rho}{\theta-2\rho}\mathbb{E}\|F\mathbb{1}_A - F_{m_1}\|_n^2 + \frac{2\theta}{\theta-2\rho}\mathrm{pen}(m_1)\\
&\quad +\frac{2\theta^2}{\theta-2\rho}\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\left[\sup_{t\in B_f(m')}\nu_n^2(t)-p(m_1,m')\right]_+\mathbb{1}_{\Omega_\rho}\right)\\
&\quad +\frac{2\theta^2}{\theta-2\rho}\mathbb{E}\left(\left[\sup_{t\in B_f(\hat m)}R_n^2(t)\right]_+\mathbb{1}_{\Omega_\rho}\right) \qquad (22)
\end{aligned}$$

We now use the following proposition:

Proposition 2. *Under the assumptions of Theorem 1, with* $p(m_1,m') = 2\ell(A_2)D^{(1)}(m_1,m')/n$ *where* $D^{(1)}(m_1,m') = \max(D^{(1)}_{m_1}, D^{(1)}_{m'})$*, or under the assumptions of Theorem 3, with either* $p(m_1,m') = 2\ell(A_2)D^{(1)}(m_1,m')/(nc_G)$ *or* $p(m_1,m') = 2(\ell(A_2)\phi_1/f_0)\mathbb{E}(\delta_1/\bar G^2(Z_1))D^{(1)}(m_1,m')/n$*, there exists a constant* C_1 *such that*

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\left[\sup_{t\in B_f(m')}\nu_n^2(t)-p(m_1,m')\right]_+\mathbb{1}_{\Omega_\rho}\right)\le\frac{C_1}{n}. \qquad (23)$$

Moreover, we can prove that

$$\mathbb{E}\left(\left[\sup_{t\in B_f(\hat m)}R_n^2(t))\right]_+\mathbb{1}_{\Omega_\rho}\right)\le \mathbb{E}\left(\left[\sup_{t\in B_f(n)}R_n^2(t))\right]_+\mathbb{1}_{\Omega_\rho}\right)\le\frac{C}{n} \qquad (24)$$

where $B_f(n)$ is the unit ball of the largest space of the (nested) collection.

To prove (24), let us define

$$\Omega_G = \{\omega, \forall y\in A_2, \hat{\hat G}(y) - \bar G(y) > -c_G/2\}. \qquad (25)$$

On Ω_G, $\hat{\hat G}(y) > c_G/2$ and $\hat{\hat G}(y)\ge 1/(n+1)$, for all $y\in A_2$. Now we use the following key lemma, useful to control the probability of the uniform deviation of the estimator of the survival distribution function $\hat{\hat G}$:

Lemma 1. *For all* $k\in\mathbb{N}^*$*, there exists a constant* C_k *depending on* k *and* c_G *such that*

$$\mathbb{E}\left(\sup_{y\in A_2}|\hat{\hat G}(y)-\bar G(y)|^{2k}\right)\le\frac{C_k}{n^k}.$$

This lemma is proved in Brunel and Comte (2005), see Lemma 6.1. Now write, for sake of simplicity (φ_j,ψ_k) instead of $(\varphi_j^m,\psi_k^m)_{j\in\mathcal{I}_n,k\in\mathcal{K}_n}$, an orthonormal basis of the largest space of the collection,

$$\mathbb{E}\left(\left[\sup_{t\in B_f(n)} R_n^2(t))\right]_+ \mathbb{1}_{\Omega_\rho}\mathbb{1}_{\Omega_G^c}\right) \le \frac{1}{f_0}\sum_{j,k}\mathbb{E}(R_n^2(\varphi_j\psi_k)\mathbb{1}_{\Omega_G^c})$$

$$\le \frac{\ell(A_2)\phi_1\mathcal{D}_n^{(1)}}{c_G^2 f_0}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \frac{(\hat{\hat{G}}(Y_i)-\bar{G}(Y_i))^2\mathbb{1}_{A_2}(Y_i)}{\hat{\hat{G}}^2(Y_i)}\mathbb{1}_{\Omega_G^c}\right)$$

$$\le \frac{\ell(A_2)\phi_1\mathcal{D}_n^{(1)}}{c_G^2 f_0}(n+1)^2\mathbb{E}(\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|^2\mathbb{1}_{\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|>c_G/2})$$

$$\le \frac{2\ell(A_2)\phi_1 n^3}{c_G^2 f_0}(2/c_G)^6\mathbb{E}(\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|^8) \le \frac{C}{n}.$$

Next, we need to study R_n on Ω_G. To this end, we write that:

$$\mathbb{E}\left(\left[\sup_{t\in B_f(n)} R_n^2(t)\right]_+ \mathbb{1}_{\Omega_\rho}\mathbb{1}_{\Omega_G}\right)$$

$$\le \frac{4}{c_G^4}\mathbb{E}\left[\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|^2 \sup_{t\in B_f(n)}\left(\frac{1}{n}\sum_{i=1}^n (\int t(X_i,y)\mathbb{1}_{\{Y_i\le y\}}dy)^2\right)\right]$$

$$\le \frac{4}{c_G^4}\mathbb{E}\left(\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|^2\right)\sup_{t\in B_f(n)}\mathbb{E}[(\int t(X_1,y)\mathbb{1}_{\{Y_1\le y\}}dy)^2] \tag{26}$$

$$+\frac{4}{c_G^4}\mathbb{E}^{1/2}\left(\sup_{y\in A_2}|\hat{\hat{G}}(y)-\bar{G}(y)|^4\right)\mathbb{E}^{1/2}\left(\sup_{t\in B_f(n)}(\nu_n\text{"}(t))^2\right) \tag{27}$$

where

$$\nu_n\text{"}(t) = \frac{1}{n}\sum_{i=1}^n\left[\left(\int t(X_i,y)\mathbb{1}_{\{Y_i\le y\}}dy\right)^2 - \mathbb{E}[(\int t(X_1,y)\mathbb{1}_{\{Y_i\le y\}}dy)^2]\right].$$

It is easy to see that $\sup_{t\in B_f(n)}\mathbb{E}[(\int t(X_1,y)\mathbb{1}_{\{Y_1\le y\}}dy)^2] \le \ell(A_2)$, and thus Lemma 1 gives that (26) is of order $1/n$. Next, with Schwarz inequalities,

$$\mathbb{E}\left(\sup_{t\in B_f(n)}(\nu_n\text{"}(t))^2\right)$$

$$\le \frac{1}{nf_0^2}\sum_{j,j',k,k'}\mathbb{E}\left[\varphi_j^2(X_1)\varphi_{j'}^2(X_1)(\int\psi_k(y)\mathbb{1}_{\{Y_1\le y\}}dy)^2(\int\psi_{k'}(y)\mathbb{1}_{\{Y_1\le y\}}dy)^2\right]$$

$$\le \frac{\ell(A_2)^2\phi_1^2(\mathcal{D}_n^{(1)})^2}{nf_0^2}.$$

It follows that this term is bounded if $\mathcal{D}_n^{(1)} \le n^{1/2}$. This implies that (27) is also of order $1/n$. Gathering both terms (on Ω_G and Ω_G^c) gives Inequality (24).

Then, with $\theta = 3\rho$, inequalities (22) and (23) yield

$$\mathbb{E}\left(\|\tilde{F}-F\mathbb{1}_A\|_n^2\mathbb{1}_{\Omega_\rho}\right) \le 5\|f_X\|_\infty\|F_{m_1}-F\mathbb{1}_A\|^2 + 6\mathrm{pen}(m_1) + \frac{18\rho(C+C_1)}{n} \tag{28}$$

The penalty term $\text{pen}(m_1)$ has to verify $\text{pen}(m_1)+\text{pen}(m')\geq 12\rho M D^{(1)}(m_1,m')/n$ i.e. $12\rho M\max(D^{(1)}_{m_1};D^{(1)}_{m'}))/n\leq \text{pen}(m_1)+\text{pen}(m')$, where $M=\ell(A_2)$ in the uncensored case and $M=\ell(A_2)/c_G$ or $M=(\ell(A_2)\phi_1/f_0)\mathbb{E}(\delta_1/\bar{G}^2(Z_1))$ in the censored case. We choose $\rho=3/2$ and so $\text{pen}(m)=18MD^{(1)}_m/n$.

To bound the second term in (20), we recall that $(\hat{F}_{\hat{m}}(X_i,u))_{1\leq i\leq n}$ is the orthogonal projection of $(\sum_k \hat{w}_i(\int \psi_k^n(y)\mathbb{1}_{\{Y_i\leq y\}}dy)\,\psi_k^n(u))_{1\leq i\leq n}$ on

$$\mathcal{W}=\{(t(X_i,u))_{1\leq i\leq n},\quad t\in S^{(1)}_{\hat{m}}\otimes\mathcal{S}^{(2)}_n\},$$

and that $P_{\mathcal{W}}$ denote the orthogonal projection on $\mathcal{W}$. Let moreover $\|.\|_{\mathbb{R}^n}$ denote the Euclidean norm in $\mathbb{R}^n$, X be the vector $(X_i)_{1\leq i\leq n}$. We define

$$W_{i,k}=w_i\int\psi_k^n(y)\mathbb{1}_{\{Z_i\leq y\}}dy,\ \ \hat{W}_{i,k}=\hat{w}_i\int\psi_k^n(y)\mathbb{1}_{\{Z_i\leq y\}}dy\qquad \text{for } i\in\{1,\dots,n\}.$$

We need to study $(\hat{F}_{\hat{m}}(X,u))=P_{\mathcal{W}}((\sum_k\hat{W}_{i,k}\psi_k^n(u))_{1\leq i\leq n})$. Let $F_k(x)=\mathbb{E}[W_{i,k}|X_i=x]=\int F(x,y)\psi_k^n(y)dy$, then the following regression-type equations are useful:

$$W_{i,k}=F_k(X_i)+\varepsilon_{i,k},\ \text{where } \varepsilon_{i,k}=W_{i,k}-\mathbb{E}[W_{i,k}|X_i],\tag{29}$$

and

$$\hat{W}_{i,k}=W_{i,k}+R_{i,k}=F_k(X_i)+\varepsilon_{i,k}+R_{i,k},\tag{30}$$

where obviously $R_{i,k}=(\int\psi_k^m(y)\mathbb{1}_{\{Z_i\leq y\}}dy)(\hat{w}_i-w_i)$ is a negligible residual term. Note that this residual is null if there is no censoring. Then, denoting for simplicity ψ_k^n by ψ_k hereafter, we have the decomposition

$$\begin{aligned}&(\hat{F}_{\hat{m}}(X_i,u))_{1\leq i\leq n}=P_{\mathcal{W}}((\sum_k\hat{W}_{i,k}\psi_k(u))_{1\leq i\leq n})\\ =\ & P_{\mathcal{W}}((\sum_k F_k(X_i)\psi_k(u))_{1\leq i\leq n})+P_{\mathcal{W}}((\sum_k\varepsilon_{i,k}\psi_k(u))_{1\leq i\leq n})+P_{\mathcal{W}}((\sum_k R_{i,k}\psi_k(u))_{1\leq i\leq n})\end{aligned}$$

We denote by $\sum_k F_k(X)\psi_k(u)$ and $(F\mathbb{1}_A)(X,u)$ the vectors of $\mathbb{R}^n$: $(\sum_k F_k(X_i)\psi_k(u))_{1\leq i\leq n}$ and $(F\mathbb{1}_A)(X_i,u)_{1\leq i\leq n}$. We denote also by ε_k the vector $(\varepsilon_{i,k})_{1\leq i\leq n}$ and by R_k the vector $(R_{i,k})_{1\leq i\leq n}$. Thus

$$\begin{aligned}&\|F\mathbb{1}_A-\hat{F}_{\hat{m}}\|_n^2\\ =\ &\frac{1}{n}\int\|(F\mathbb{1}_A)(X,u)-P_{\mathcal{W}}(\sum_k F_k(X)\psi_k(u))-P_{\mathcal{W}}(\sum_k\varepsilon_k\psi_k(u))-P_{\mathcal{W}}(\sum_k R_k\psi_k(u))\|^2_{\mathbb{R}^n}du\\ \leq\ &\frac{2}{n}\int\|(F\mathbb{1}_A)(X,u)-P_{\mathcal{W}}(\sum_k F_k(X)\psi_k(u))\|^2_{\mathbb{R}^n}du+\frac{2}{n}\int\|P_{\mathcal{W}}(\sum_k(\varepsilon_k+R_k)\psi_k(u))\|^2_{\mathbb{R}^n}du\end{aligned}$$

Now, we can prove the following Lemma:

Lemma 2. $(1/n)\int\|(F\mathbb{1}_A)(X,u)-P_{\mathcal{W}}(\sum_k F_k(X)\psi_k(u))\|^2_{\mathbb{R}^n}du\leq(1/n)\int\|F\mathbb{1}_A(X,u)\|^2_{\mathbb{R}^n}du.$

Therefore,

$$
\begin{aligned}
&\|F\mathbb{1}_A - \hat{F}_{\hat{m}}\|_n^2 \\
\leq\ & \frac{2}{n}\int \|F\mathbb{1}_A(X,u)\|_{\mathbb{R}^n}^2 du + \frac{4}{n}\int \|\sum_k \varepsilon_k\psi_k(u)\|_{\mathbb{R}^n}^2 du + \frac{4}{n}\int \|\sum_k R_k\psi_k(u)\|_{\mathbb{R}^n}^2 du \\
\leq\ & \frac{2}{n}\sum_{i=1}^n \int (F\mathbb{1}_A)^2(X_i,u)du + \frac{4}{n}\sum_{i=1}^n \int [\sum_k \varepsilon_{i,k}\psi_k(u)]^2 du + \frac{4}{n}\sum_{i=1}^n \int [\sum_k R_{i,k}\psi_k(u)]^2 du \\
\leq\ & 2 + \frac{4}{n}\sum_{i=1}^n\sum_k \varepsilon_{i,k}^2 + \frac{4}{n}\sum_{i=1}^n\sum_k R_{i,k}^2.
\end{aligned}
$$

Using (6) yields,

$$
\begin{aligned}
\varepsilon_{i,k}^2 &\leq 2(w_i\int \psi_k(y)\mathbb{1}_{\{Z_i\leq y\}}dy)^2 + 2\mathbb{E}[w_i\int \psi_k(y)\mathbb{1}_{\{Z_i\leq y\}}dy|X_i]^2 \\
\text{and}\quad \sum_k \varepsilon_{i,k}^2 &\leq 2\sum_k \frac{(\int \psi_k(y)\mathbb{1}_{\{Z_i\leq y\}}dy)^2}{\bar{G}^2(Z_i)} + 2\mathbb{E}[\sum_k \frac{(\int \psi_k(y)\mathbb{1}_{\{Z_i\leq y\}}dy)^2}{\bar{G}^2(Z_i)}|X_i] \leq 4\frac{\ell(A_2)}{c_G^2}
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\sum_k R_{i,k}^2 &\leq \sum_k (\int \psi_k(y)\mathbb{1}_{\{Z_i\leq y\}}dy)^2 \frac{|\hat{\bar{G}}(Y_i)-\bar{G}(Y_i)|^2\mathbb{1}_{A_2}(Y_i)}{c_G^2\hat{\bar{G}}^2(Y_i)} \\
&\leq \frac{\ell(A_2)}{c_G^2}\frac{|\hat{\bar{G}}(Y_i)-\bar{G}(Y_i)|^2\mathbb{1}_{A_2}(Y_i)}{\hat{\bar{G}}^2(Y_i)}.
\end{aligned}
$$

Thus, it follows from the previous study (and in particular from Lemma 1), that the following inequality holds

$$
\mathbb{E}^{1/2}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{|\hat{\bar{G}}(Y_i)-\bar{G}(Y_i)|^2\mathbb{1}_{A_2}(Y_i)}{\hat{\bar{G}}^2(Y_i)}\right)^2\right] \leq \frac{\kappa}{n}. \tag{31}
$$

Using (31), we obtain

$$
\mathbb{E}\left(\|F\mathbb{1}_A - \hat{F}_{\hat{m}}\|_n^2\mathbb{1}_{\Omega_\rho^c}\right) \leq 2(1+8\frac{\ell(A_2)}{c_G^2})\mathbb{P}(\Omega_\rho^c) + \frac{4\kappa\ell(A_2)}{c_G^2 n}\mathbb{P}^{1/2}(\Omega_\rho^c). \tag{32}
$$

Now we use the following proposition proved in Brunel et al. (2008) (Proposition 3):

Proposition 3. *Let $\rho > 1$. Then, under the assumptions of Theorem 1, there exists $C_2 > 0$ such that $P(\Omega_\rho^c) \leq \frac{C_2}{n^{3/2}}$.*

This proposition implies that $\mathbb{E}\left(\|F\mathbb{1}_A - \hat{F}_{\hat{m}}\|_n^2\mathbb{1}_{\Omega_\rho^c}\right) \leq \frac{C_3}{n}$.

Now we use (28) and we observe that this inequality holds for all m_1 in $\mathcal{M}_n$, so

$$
\mathbb{E}\|\tilde{F} - F\mathbb{1}_A\|_n^2 \leq C\inf_{m\in\mathcal{M}_n}(\|F\mathbb{1}_A - F_m\|^2 + \text{pen}(m)) + \frac{C_4}{n}
$$

with $C = \max(5\|f_X\|_\infty, 6)$.

5.3. **Proof of Proposition 2.** Let $\Gamma_i(t)=\delta_i/\bar{G}(Y_i)\int t(X_i,y)\mathbb{1}_{\{Y_i\le y\}}dy-\int t(X_i,y)F(X_i,y)dy$. Then $\nu_n(t)=(1/n)\sum_{i=1}^n\Gamma_i(t)$. The proof of Proposition 2 relies on (6) and on the following result:

Lemma 3. *(Talagrand (1996))*
Let $U_1,\dots,U_n$ i.i.d. variables and $(\zeta_t)_{t\in B}$ a set of functions and B is a unit ball of a finite dimensional subspace of $\mathbb{L}^2(A)$. Let $\nu_n(t)=(1/n)\sum_{i=1}^n[\zeta_t(U_i)-\mathbb{E}(\zeta_t(U_i))]$. We suppose that
(i) $\sup_{t\in B}\|\zeta_t\|_\infty\le M_1$, *(ii)* $\mathbb{E}(\sup_{t\in B}|\nu_n(t)|)\le H$, *(iii)* $\sup_{t\in B}\mathrm{Var}[\zeta_t(U_1)]\le v$.
Then, there exists $K>0$, $K_1>0$, $K_2>0$ such that

$$\mathbb{E}\left[\sup_{t\in B}\nu_n^2(t)-2H^2\right]_+\le K\left[\frac{v}{n}e^{-K_1\frac{nH^2}{v}}+\frac{M_1^2}{n^2}e^{-K_2\frac{nH}{M_1}}\right]$$

Here $\zeta_t(u,v,\delta)=\delta\int t(x,y)\mathbb{1}_{\{v\le y\}}dy/\bar{G}(v)-\int t(u,y)F(u,y)dy$ and $B=B_f(m')$. We now compute the constants M_1, H and v involved in *(i), (ii), (iii)*.

(i) We recall that in general, $S_m+S_{m'}$ is included in the model $S_{m''}$ with dimension $\max(D_{m_1}^{(1)},D_{m_1'}^{(1)})\max(D_{m_2}^{(2)},D_{m_2'}^{(2)})$. Here, the space $S_m+S_{m'}=S_{m_1}^{(1)}\otimes\mathcal{S}_n^{(2)}+S_{m'}^{(1)}\otimes\mathcal{S}_n^{(2)}$ is a space of the collection included in $S_{m_1,m'}^*\otimes\mathcal{S}_n^{(2)}$ (where $S_{m_1,m'}^*$ is the largest of $S_{m_1}^{(1)}$ and $S_{m'}^{(1)}$) and has dimension $\max(D_{m_1}^{(1)},D_{m'}^{(1)})\mathcal{D}_n^{(2)}$. Let $(\bar\varphi_j\otimes\psi_k^n)_{(j,k)}$ be an orthonomal basis of this space w.r.t. $\|.\|_f$ and let $t\in B_f(m')$.

$$\begin{aligned}|\delta\int t(u,y)\mathbb{1}_{\{v\le y\}}dy/\bar{G}(v)| &\le \frac{1}{c_G}|\sum_{j,k}a_{j,k}\bar\varphi_j(u)\int\psi_k^n(y)\mathbb{1}_{\{v\le y\}}dy|\\ &\le \frac{1}{c_G}\left(\sum_{j,k}a_{j,k}^2\sum_j\bar\varphi_j^2(u)\sum_k\left(\int_{A_2}\psi_k^n(y)\mathbb{1}_{\{v\le y\}}dy\right)^2\right)^{1/2}\\ &\le \frac{1}{c_G}\|t\|_f\left(\ell(A_2)\sum_j\bar\varphi_j^2(u)\right)^{1/2}\quad\text{with (6),}\\ &\le \frac{\sqrt{\phi_1\ell(A_2)\max(D_{m_1}^{(1)},D_{m'}^{(1)})}}{\sqrt{f_0}c_G}.\end{aligned}$$

Analogously,

$$|\int t(u,y)F(u,y)dy|=|\sum_{j,k}a_{j,k}\bar\varphi_j(u)\int\psi_k^n(y)F(u,y)dy|$$

$$\le\left(\sum_{j,k}a_{j,k}^2\sum_j\bar\varphi_j^2(u)\sum_k\left(\int\psi_k^n(y)F(u,y)dy\right)^2\right)^{1/2}\le\|t\|_f\left(\ell(A_2)\sum_j\bar\varphi_j^2(u)\right)^{1/2}\quad\text{with (6),}$$

$$\le\sqrt{\frac{\ell(A_2)\phi_1\max(D_{m_1}^{(1)},D_{m'}^{(1)})}{f_0}}.$$

Therefore

$$\sup_{t\in B}\|\zeta_t\|_\infty \leq \frac{(1/c_G+1)\sqrt{\phi_1\ell(A_2)}}{\sqrt{f_0}}\sqrt{\max(D_{m_1}^{(1)},D_{m'}^{(1)})}.$$

Then we set $M_1=\dfrac{(1/c_G+1)\sqrt{\phi_1\ell(A_2)}}{\sqrt{f_0}}\sqrt{\max(D_{m_1}^{(1)},D_{m'}^{(1)})}$.

$$\begin{aligned}
\textit{(ii)}\ \mathrm{Var}[\zeta_t(U_1)]=\mathbb{E}\left([\Gamma_1(t)]^2\right) &\leq \mathbb{E}\left[\left(\frac{\int t(X_1,y)\mathbb{1}_{\{Y_1\leq y\}}dy}{\bar{G}(Y_1)}\right)^2\right]\\
&\leq \frac{1}{c_G}\mathbb{E}\left[\int t^2(X_1,y)\mathbb{1}_{\{Y_1\leq y\}}dy\right]\\
&= \frac{1}{c_G}\iiint t^2(x,y)\mathbb{1}_{\{u\leq y\}}\pi(x,u)f_X(x)dxdydu\\
&= \frac{1}{c_G}\iint t^2(x,y)F(x,y)f_X(x)dxdy\leq \frac{\|t\|_f^2}{c_G}.
\end{aligned}$$

Then $v=\dfrac{1}{c_G}$.

(iii) Let $(\bar{\varphi}_j\otimes\psi_k)_{(j,k)}$ be as previously.

$$\begin{aligned}
\mathbb{E}(\sup_{t\in B}|\nu_n^2(t)|) &\leq \sum_{j,k}\mathbb{E}(\nu_n^2(\bar{\varphi}_j\otimes\psi_k))=\sum_{j,k}\frac{1}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^n\Gamma_i(\bar{\varphi}_j\otimes\psi_k)\right)^2\right]\\
&= \sum_{j,k}\frac{1}{n}\mathbb{E}\left(\Gamma_1^2(\bar{\varphi}_j\otimes\psi_k)\right).
\end{aligned}$$

In the uncensored case, we find

$$\begin{aligned}
\mathbb{E}(\sup_{t\in B}|\nu_n^2(t)|) &\leq \sum_{j,k}\frac{1}{n}\mathbb{E}\left(\bar{\varphi}_j^2(X_1)\left(\int\psi_k(y)\mathbb{1}_{\{Y_1\leq y\}}dy\right)^2\right)\\
&= \sum_{j,k}\frac{1}{n}\iint\left(\bar{\varphi}_j^2(x)\left(\int\psi_k(y)\mathbb{1}_{\{v\leq y\}}dy\right)^2\right)\pi(x,v)f_X(x)dxdv\\
&\leq \frac{1}{n}\sum_j\iint\bar{\varphi}_j^2(x)\left(\int_{A_2}\mathbb{1}_{\{v\leq y\}}dy\right)\pi(x,v)f_X(x)dxdv\\
&\leq \frac{1}{n}\sum_j\int\bar{\varphi}_j^2(x)\int_{A_2}\left(\int\mathbb{1}_{\{v\leq y\}}\pi(x,v)dv\right)f_X(x)dxdy\\
&\leq \frac{1}{n}\sum_j\int\bar{\varphi}_j^2(x)\left(\int_{A_2}F(x,y)dy\right)f_X(x)dx\\
&\leq \frac{\ell(A_2)}{n}\sum_j\int\bar{\varphi}_j^2(x)f_X(x)dx=\frac{\ell(A_2)\max(D_{m_1}^{(1)},D_{m'}^{(1)})}{n}.
\end{aligned}$$

In the censored case, the bound becomes $\ell(A_2)\max(D^{(1)}_{m_1}, D^{(1)}_{m'})/(nc_G)$. But, in order to avoid the constant c_G which would be uneasy to estimate, it is also possible to bound the term differently. Take $(\varphi_j\otimes\psi^n_k)_{j,k}$ an orthonormal w.r.t. the standard $\mathbb{L}^2$ scalar product.

$$\begin{aligned}\mathbb{E}(\sup_{t\in B}|\nu_n^2(t)|) &\leq \sum_{j,k}\frac{1}{nf_0}\mathbb{E}\left(\frac{\delta_1\varphi_j^2(X_1)\left(\int\psi_k^n(y)\mathbb{1}_{\{Z_1\leq y\}}dy\right)^2}{\bar{G}^2(Z_1)}\right)\leq\frac{\ell(A_2)}{nf_0}\mathbb{E}\left(\frac{\delta_1\|\sum_j\varphi_j^2\|_\infty}{\bar{G}^2(Z_1)}\right)\\ &\leq \mathbb{E}\left(\frac{\delta_1}{\bar{G}^2(Z_1)}\right)\frac{\ell(A_2)\phi_1}{f_0}\frac{\max(D^{(1)}_{m_1}, D^{(1)}_{m'})}{n}.\end{aligned}$$

Then $\mathbb{E}(\sup_{t\in B}\nu_n^2(t))\leq M\max(D^{(1)}_{m_1}, D^{(1)}_{m'})/n$ and $H^2 = M\max(D^{(1)}_{m_1}, D^{(1)}_{m'})/n$ with $M = \ell(A_2)$ in the uncensored case or $M=\ell(A_2)/c_G$ or $M=(\ell(A_2)\phi_1/f_0)\mathbb{E}(\delta_1/\bar{G}^2(Z_1))$ in the censored case. We set $D^{(1)}(m_1,m')=\max(D^{(1)}_{m_1}, D^{(1)}_{m'})$

According to Lemma 3, there exists $K'>0$, $K_1>0$, $K_2'>0$ such that

$$\mathbb{E}\left[\sup_{t\in B_f(m')}\nu_n^2(t)-2H^2\right]_+\leq K'\left[\frac{1}{n}e^{-K_1D^{(1)}(m_1,m')}+\frac{1}{n}e^{-K_2'\sqrt{n}}\right].$$

So, if $p(m_1,m')=2H^2=2MD^{(1)}(m_1,m')/n$,

$$\begin{aligned}&\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\sup_{t\in B_f(m')}\nu_n^2(t)-p(m_1,m')\right]_+\\ &\leq\frac{K'}{n}\left[\sum_{m'\in\mathcal{M}_n}(e^{-K_1D^{(1)}(m_1,m')}+e^{-K_2'\sqrt{D^{(1)}(m_1,m')}})\right]\leq\frac{A_1}{n},\end{aligned}$$

as $D^{(1)}(m_1,m')\leq\mathcal{D}^{(1)}_n\leq n$ and the result follows. $\square$

5.4. **Proof of Lemma 2.** First, remark that, for any $t\in S^{(1)}_{m_1}\otimes S^{(2)}_{m_2}$, we can write $t(x,y)=\sum_k t_k(x)\psi_k^{m_2}(y)$ where $t_k(x)=\int t(x,u)\psi_k^{m_2}(u)du$. Then, if we denote $P_{\mathcal{W}}((\sum_k F_k(X_i)\psi^n_k(u))_{1\leq i\leq n}) = (s_F(X_i,u))_{1\leq i\leq n}$, and $s_F(X_i,u)=\sum_k s_{F,k}(X_i)\psi_k^n(u)$, we know that, for any function $s\in S^{(1)}_{m_1}\otimes\mathcal{S}^{(2)}_n$ and any $u\in A_2$,

$$\sum_{i=1}^n(\sum_k s_k(s_k(X_i)\psi^n_k(u))(\sum_k F_k(X_i)\psi^n_k(u))=\sum_{i=1}^n(\sum_k s_k(X_i)\psi^n_k(u))(\sum_k s_{F,k}(X_i)\psi^n_k(u)).$$

Therefore, integrating w.r.t. u and setting $s=s_F$ yields:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\left(\sum_k s_{F,k}(X_i)F_k(X_i)\right) &= \frac{1}{n}\sum_{i=1}^n\sum_k s^2_{F,k}(X_i)\\ &= \frac{1}{n}\int\|P_{\mathcal{W}}((\sum_k F_k(X_i)\psi^n_k(u))_{1\leq i\leq n})\|^2_n du.\end{aligned}$$

Therefore

$$\begin{aligned}
\langle F\mathbb{1}_A(X,u), P_{\mathcal{W}}((\sum_k F_k(X_i)\psi_k(u)\rangle_n &= \frac{1}{n}\int\sum_{i=1}^n F(X_i,u)\sum_k s_{F,k}(X_i)\psi_k^n(u)du \\
&= \frac{1}{n}\sum_{i=1}^n\sum_k\left(\int F\mathbb{1}_A(X_i,u)\psi_k^n(u)du\right)s_{F,k}(X_i) \\
&= \frac{1}{n}\sum_k\sum_{i=1}^n F_k(X_i)s_{F,k}(X_i) \\
&= \frac{1}{n}\int\|P_{\mathcal{W}}((\sum_k F_k(X_i)\psi_k^n(u))_{1\le i\le n})\|_n^2du.
\end{aligned}$$

This in turn implies that

$$\begin{aligned}
&\frac{1}{n}\int\|F\mathbb{1}_A(X,u)-P_{\mathcal{W}}((\sum_k F_k(X_i)\psi_k^n(u))_{1\le i\le n})\|_{\mathbb{R}^n}^2du \\
=\ &\frac{1}{n}\int\|F\mathbb{1}_A(X,u)\|_{\mathbb{R}^n}^2du-\frac{1}{n}\int\|P_{\mathcal{W}}((\sum_k F_k(X_i)\psi_k^n(u))_{1\le i\le n})\|_n^2du.
\end{aligned}$$

This gives the result of Lemma 2. □

5.5. **Proof of Corollary 1.** To control the bias term, we state the following lemma proved in Lacour (2007) and following from Hochmuth (2002) and Nikol'skii (1975):

Lemma 4. *Let h be a function and $A = A_1 \times A_2$ be a product of compact sets, and assume that $h\mathbb{1}_A = h_A$ belongs to $B^{\boldsymbol{\alpha}}_{2,\infty}(A)$. We consider that S'_m is one of the following spaces on A:*

- *a space of piecewise polynomials of degrees bounded by $s_i > \alpha_i - 1$ ($i = 1, 2$) based on a partition with rectangles of sidelengthes $1/D_{m_1}$ and $1/D_{m_2}$,*
- *a linear span of $\{\phi_\lambda\psi_\mu, \lambda \in \cup_0^{m_1}\Lambda(j), \mu \in \cup_0^{m_2}M(k)\}$ where $\{\phi_\lambda\}$ and $\{\psi_\mu\}$ are orthonormal wavelet bases of respective regularities $s_1 > \alpha_1 - 1$ and $s_2 > \alpha_2 - 1$ (here $D_{m_i} = 2^{m_i}, i = 1, 2$),*
- *the space of trigonometric polynomials with degree smaller than D_{m_1} in the first direction and smaller than D_{m_2} in the second direction.*

Let h'_m be the orthogonal projection of $h_A = h\mathbb{1}_A$ on S'_m. Then, there exists a positive constant C_0 such that

$$\left(\int_A |h_A - h'_m|^2\right)^{1/2} \le C_0[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}].$$

If we choose for S'_m the set of the restrictions to A of the functions of $S^{(1)}_{m_1} \otimes \mathcal{S}^{(2)}_n$ and F_A the restriction of F to A, we can apply Lemma 4. But F'_m is also the restriction to A of F_m so that

$$\|F\mathbb{1}_A - F_m\| \le C_0[(D^{(1)}_{m_1})^{-\alpha_1} + (\mathcal{D}^{(2)}_n)^{-\alpha_2}].$$

According to Theorem 1

$$\mathbb{E}\|\tilde{F} - F\mathbb{1}_A\|_n^2 \le C'' \inf_{m_1\in\mathcal{M}_n}\left\{(D^{(1)}_{m_1})^{-2\alpha_1} + n^{-1} + \frac{D^{(1)}_{m_1}}{n}\right\},$$

provided that $\mathcal{D}_n^{(2)} \geq \sqrt{n}$ and $\alpha_2 > 1$.

In particular, if m^* is such that $D_{m_1^*} = \lfloor n^{\frac{1}{2\alpha_1+1}} \rfloor$ then

$$\mathbb{E}\|\tilde{\pi} - \pi \mathbb{1}_A\|_n^2 \leq C''' \left\{ D_{m_1^*}^{-2\alpha_1} + \frac{D_{m_1^*}}{n} \right\} = O\left(n^{-\frac{2\alpha_1}{2\alpha_1+1}} \right).$$

5.6. **Proof of Theorem 2.** The proof follows the one of Lacour (2007). Let ψ be a regular wavelet with compact support. For $J = (j_1, j_2) \in \mathbb{Z}^2$ to be chosen below and $K = (k_1, k_2) \in \mathbb{Z}^2$, we set $\psi_{JK}(x, y) = 2^{(j_1+j_2)/2}\psi(2^{j_1}x - k_1)\psi(2^{j_2}y - k_2)$ and

$$\Psi_{JK}(x, y) = \int_{-\infty}^{y} \psi_{JK}(x, u)du = 2^{(j_1-j_2)/2}\psi(2^{j_1}x - k_1)\Psi(2^{j_2}y - k_2),$$

where $\Psi(y) = \int_{-\infty}^{y} \psi(u)du$. Now we set R_J the maximal subset of $\mathbb{Z}^2$ such that

$$\mathrm{Supp}(\psi_{JK}) \subset A \quad \forall K \in R_J, \qquad \mathrm{Supp}(\psi_{JK}) \cap \mathrm{Supp}(\psi_{JK'}) = \emptyset \text{ if } K \neq K'.$$

Note that the Ψ_{JK} have the same support that the ψ_{JK}, since $\int \psi = 0$. The cardinal of R_J is $|R_J| = c2^{j_1+j_2}$, with c a positive constant which depends only on A and the support of ψ. Let, for all $\varepsilon = (\varepsilon_K) \in \{-1, 1\}^{|R_J|}$,

$$F_\varepsilon(x, y) = |B|^{-1} \int_{-\infty}^{y} \mathbb{1}_B(u)du + \frac{1}{\sqrt{n}} \sum_{K \in R_J} \varepsilon_K \Psi_{JK}(x, y),$$

with B a compact set such that $A \subset B \times B$, $\inf(B) = \inf(A_2)$ and $|B|$ large enough to be specified later. Let us denote by $\mathcal{G}$ the set of all such F_ε.

We want to show that $\mathcal{G} \subset \mathcal{B}$. Clearly, all F_ε are continuous with limit 0 in $-\infty$ and $+1$ in ∞. In addition,

$$\frac{\partial F_\varepsilon}{\partial y}(x, y) = |B|^{-1}\mathbb{1}_B(y) + \frac{1}{\sqrt{n}} \sum_{K \in R_J} \varepsilon_K \psi_{JK}(x, y).$$

Then, if $(x, y) \notin A$, $\frac{\partial F_\varepsilon}{\partial y} \geq 0$. If $(x, y) \in A$, $\frac{\partial F_\varepsilon}{\partial y} \geq |B|^{-1} - 2^{(j_1+j_2)/2}\|\psi\|_\infty^2/\sqrt{n}$ and then $\frac{\partial F_\varepsilon}{\partial y} \geq |B|^{-1}/2 > 0$ as soon as

$$\text{Condition C1:} \qquad \left(\frac{2^{j_1+j_2}}{n} \right)^{1/2} \leq \frac{|B|^{-1}}{2\|\psi\|_\infty^2}.$$

Thus, if the Condition C1 is verified, $\frac{\partial F_\varepsilon}{\partial y} \geq 0$ and F_ε is non decreasing. We now have to bound $\|F_\varepsilon\|_{B_{2,\infty}^{\alpha}(A)}$. We can write

$$\|F\|_{B_{2,\infty}^{\alpha}(A)} \leq |B|^{-1} \| \int_{-\infty}^{y} \mathbb{1}_B(u)du\|_{B_{2,\infty}^{\alpha}(A)} + \frac{1}{\sqrt{n}} \| \sum_{K \in R_J} \varepsilon_K \Psi_{JK}\|_{B_{2,\infty}^{\alpha}(A)}.$$

To compute these quantities, we use Lemma 5 proved below. First

$$\begin{aligned} \| \int_{-\infty}^{y} \mathbb{1}_B(u)du\|_{B_{2,\infty}^{\alpha}(A)} &= \| \int_{-\infty}^{y} \mathbb{1}_B(u)du\|_A + | \int_{-\infty}^{y} \mathbb{1}_B(u)du|_{B_{2,\infty}^{\alpha}(A)} \\ &\leq \|y - \inf(B)\|_A + C|\mathbb{1}_B(y)|_{B_{2,\infty}^{(\alpha_1, \alpha_2-1)}(A)} \\ &= \left(\frac{|A_1||A_2|^3}{3} \right)^{1/2} \end{aligned}$$

Moreover, using Lemma 5,

$$\begin{aligned}
\|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_{B^{\alpha}_{2,\infty}(A)} &= \|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_A + |\sum_{K\in R_J}\varepsilon_K\Psi_{JK}|_{B^{(\alpha_1,\alpha_2)}_{2,\infty}(A)} \\
\leq \|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_A &+ [1\vee \max_{K\in R_J}(\|g_K\|_{A_2}\|g'_K\|^{-1}_{A_2})]|\sum_{K\in R_J}\varepsilon_K\psi_{JK}|_{B^{(\alpha_1,\alpha_2-1)}_{2,\infty}(A)}
\end{aligned}$$

with $g_K(y) = 2^{-j_2/2}\Psi(2^{j_2}y - k_2)$. Hochmuth (2002) proves that for ψ smooth enough $|\sum_{K\in R_J}\varepsilon_K\psi_{JK}|_{B^{(\alpha_1,\alpha_2-1)}_{2,\infty}(A)} \leq (2^{j_1\alpha_1} + 2^{j_2(\alpha_2-1)})\|\sum_{K\in R_J}\varepsilon_K\psi_{JK}\|_A$. Then

$$\begin{aligned}
\|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_{B^{\alpha}_{2,\infty}(A)} &\leq \|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_A \\
&\quad +(1\vee 2^{-j_2}\|\Psi\|)(2^{j_1\alpha_1} + 2^{j_2(\alpha_2-1)})\|\sum_{K\in R_J}\varepsilon_K\psi_{JK}\|_A
\end{aligned}$$

For the latter term of the right-hand-side of the previous inequality, we have

$$\|\sum_{K\in R_J}\varepsilon_K\psi_{JK}\|^2_A = \sum_{K\in R_J}|\varepsilon_K|^2 = c2^{j_1+j_2}$$

and for the first one

$$\|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|^2_A = \sum_{K\in R_J}|\varepsilon_K|^2\|\Psi_{JK}\|^2_A = c2^{j_1+j_2}2^{-2j_2}\|\Psi\|^2,$$

as

$$\begin{aligned}
\|\Psi_{JK}\|^2_A &= \iint_A 2^{j_1}|\psi(2^{j_1}x-k_1)|^2 2^{-j_2}|\Psi(2^{j_2}y-k_2)|^2 dxdy \\
&= 2^{-2j_2}\iint |\psi(x)|^2|\Psi(y)|^2 dxdy = 2^{-2j_2}\|\Psi\|^2. \qquad (33)
\end{aligned}$$

It follows that:

$$\|\sum_{K\in R_J}\varepsilon_K\Psi_{JK}\|_{B^{\alpha}_{2,\infty}(A)} \leq (1\vee 2^{-j_2}\|\Psi\|)(1+2^{j_1\alpha_1}+2^{j_2(\alpha_2-1)})c^{1/2}2^{(j_1+j_2)/2}.$$

Finally

$$\|F\|_{B^{\alpha}_{2,\infty}(A)} \leq |B|^{-1}\left(\frac{|A_1||A_2|^3}{3}\right)^{1/2} + \frac{1\vee 2^{-j_2}\|\Psi\|}{\sqrt{n}}(1+2^{j_1\alpha_1}+2^{j_2(\alpha_2-1)})c^{1/2}2^{(j_1+j_2)/2}.$$

We take B such that $|B| \geq (2/L)\left(|A_1||A_2|^3/3\right)^{1/2}$ so that

$$|B|^{-1}\left(\frac{|A_1||A_2|^3}{3}\right)^{1/2} \leq \frac{L}{2}$$

and from now on, we suppose that Condition C2 is verified where

$$\text{Condition C2:} \qquad \frac{(1+2^{j_1\alpha_1}+2^{j_2(\alpha_2-1)})2^{(j_1+j_2)/2}}{\sqrt{n}} \leq \frac{L}{2\sqrt{c}(1\vee 2^{-j_2}\|\Psi\|)}.$$

Then for all ε, $F_\varepsilon \in \mathcal{B}$.

For $\varepsilon \in \{-1,1\}^{|R_J|}$, put $\varepsilon_{*K} = (\varepsilon'_I)_{I\in R_J}$ such that $\varepsilon'_I = \varepsilon_I$ if $I \neq K$ and $\varepsilon'_I = -\varepsilon_I$ if $I = K$. Let also the likelihood ratio

$$\Lambda_n(\varepsilon_{*K}, \varepsilon) = \prod_{i=1}^{n} \frac{\pi_{\varepsilon_{*K}}(X_i, Y_i)}{\pi_\varepsilon(X_i, Y_i)}$$

denoting, for all ε, $\pi_\varepsilon(x,y) = \frac{\partial F_\varepsilon}{\partial y}(x,y)$ the conditional density associated to F_ε. With the same proof as for Lemma 10.2 p.160 in Härdle *et al.* (1998), we can prove the following result: if there exist $\lambda > 0$ and p_0 such that $P_{F_\varepsilon,f}(\Lambda_n(\varepsilon_{*K},\varepsilon) > e^{-\lambda}) \geq p_0$, then, for any estimator $\hat{F}_n$,

$$\max_{F_\varepsilon \in \mathcal{G}} \mathbb{E}_{F_\varepsilon,f} \|\hat{F}_n - F_\varepsilon\|_A^2 \geq \frac{|R_J|}{2} \delta^2 e^{-\lambda} p_0$$

where $\delta = \inf_{\varepsilon\neq\varepsilon'} \|F_\varepsilon - F_{\varepsilon'}\|_A/2$. It is proved in Lacour (2007) section 7.5 that, if $2^{j_1+j_2} = o(n)$, $P_{\pi_\varepsilon}(\Lambda_n(\varepsilon_{*K},\varepsilon) > e^{-\lambda}) \geq p_0$ (the context is slightly different but it is sufficient to replace X_{i+1} by Y_i and to bound f by $\|f\|_\infty$).

Let us now compute $\delta = \inf_{\varepsilon\neq\varepsilon'} \|F_\varepsilon - F_{\varepsilon'}\|_A/2 = \|\varepsilon_K \Psi_{JK}/\sqrt{n}\|_A = \|\Psi_{JK}\|_A/\sqrt{n}$. Since it follows from (33 that $\|\Psi_{JK}\|_A^2 = 2^{-2j_2}\|\Psi\|^2$, we set $\delta = 2^{-j_2}\|\Psi\|/\sqrt{n}$.

Now for all n we choose $J = J(n) = (j_1(n), j_2(n))$ such that

$$c_1/2 \leq 2^{j_1} n^{-\frac{1}{2\alpha_1+1}} \leq c_1 \quad \text{and} \quad c_2/2 \leq 2^{j_2} \leq c_2$$

with c_1 and c_2 such that $\sqrt{c_1c_2} \leq |B|^{-1}/(2\|\psi\|_\infty^2)$ so that Condition C1 is satisfied, and such that $(1+c_1^{\alpha_1}+c_2^{\alpha_2-1})\sqrt{c_1c_2} \leq L/(2c^{1/2}(1\vee 2^{-j_2}\|\Psi\|))$ so that Condition C2 is satisfied. Then we have

$$|R_J|\delta^2 = c\|\Psi\|^2 \frac{2^{j_1-j_2}}{n} \geq \frac{cc_1c_2}{2}\|\Psi\|^2 n^{-\frac{2\alpha_1}{2\alpha_1+1}}.$$

Thus

$$\max_{F_\varepsilon \in \mathcal{G}} \mathbb{E}_{F_\varepsilon,f} \|\hat{F}_n - F_\varepsilon\|_A^2 \geq \frac{ce^{-\lambda}p_0c_1c_2}{4}\|\Psi\|^2 n^{\frac{-2\alpha_1}{2\alpha_1+1}}.$$

And then for all estimator

$$\sup_{F\in\mathcal{B}} \mathbb{E}_{F,f}\|\hat{F}_n - F\|_A^2 \geq Cn^{-\frac{2\alpha_1}{2\alpha_1+1}}$$

with $C = ce^{-\lambda}\|\Psi\|^2 p_0c_1c_2/4$.

□

Lemma 5. *Let $\alpha_1 > 0, \alpha_2 > 1$. For any functions $(f_i)_{i\in I}$ and any differentiable functions $(g_i)_{i\in I}$ such that $(g_i)_{i\in I}$ and $(g'_i)_{i\in I}$ are orthogonal families,*

$$|\sum_{i\in I} f_i \otimes g_i|_{B_{2,\infty}^{(\alpha_1,\alpha_2)}(A)} \leq 1 \vee \max_{i\in I}(\|g_i\|_{A_2}\|g'_i\|_{A_2}^{-1}) \left|\sum_{i\in I} f_i\otimes g'_i\right|_{B_{2,\infty}^{(\alpha_1,\alpha_2-1)}(A)}$$

Proof of Lemma 5: Let r_2 an integer strictly larger than α_2 and r_1 an integer strictly larger than α_1. Using the definition

$$|\sum_{i\in I} f_i\otimes g_i|_{B_{2,\infty}^{(\alpha_1,\alpha_2)}(A)} = t^{-\alpha_1}\sup_{|h|\leq t}\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i)\right\|_{A_{h,1}^{r_1}} + t^{-\alpha_2}\sup_{|h|\leq t}\left\|\Delta_{2,h}^{r_2}(\sum_{i\in I} f_i\otimes g_i)\right\|_{A_{h,2}^{r_2}}$$

We now bound these two terms.

• First $\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i) = \sum_{i\in I}(\Delta_h^{r_1} f_i)\otimes g_i$ and $\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i') = \sum_{i\in I}(\Delta_h^{r_1} f_i)\otimes g_i'$ where $\Delta_h^r(f)(x) = \sum_{k=0}^r (-1)^{r-k}\binom{r}{k} f(x+kh)$

Then

$$\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i)\right\|^2_{A_{h,1}^{r_1}} = \iint_{A_{h,1}^{r_1}}\left(\sum_{i\in I}(\Delta_h^{r_1} f_i)\otimes g_i\right)^2 = \sum_{i\in I}\int(\Delta_h^{r_1} f_i)^2\int g_i^2$$

using the orthogonality of $(g_i)_{i\in I}$. Hence,

$$\begin{aligned}\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i)\right\|^2_{A_{h,1}^{r_1}} &\leq \max_{i\in I}\left(\|g_i\|^2_{A_2}\|g_i'\|^{-2}_{A_2}\right)\sum_{i\in I}\int(\Delta_h^{r_1} f_i)^2\int(g_i')^2\\ &\leq \max_{i\in I}\left(\|g_i\|^2_{A_2}\|g_i'\|^{-2}_{A_2}\right)\iint_{A_{h,1}^{r_1}}\left(\sum_{i\in I}(\Delta_h^{r_1} f_i)\otimes g_i'\right)^2\end{aligned}$$

using this time the orthogonality of $(g_i')_{i\in I}$. So

$$t^{-\alpha_1}\sup_{|h|\leq t}\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i)\right\|_{A_{h,1}^{r_1}} \leq \max_{i\in I}\left(\|g_i\|_{A_2}\|g_i'\|^{-1}_{A_2}\right)t^{-\alpha_1}\sup_{|h|\leq t}\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i')\right\|_{A_{h,1}^{r_1}}$$

• For the direction y, we can write

$$\Delta_{2,h}^{r_2}(\sum_{i\in I} f_i\otimes g_i) = \sum_{i\in I} f_i\otimes\Delta_h^{r_2}(g_i)$$

But

$$\begin{aligned}\Delta_h^{r_2}(g_i)(y) &= \sum_{k=0}^{r_2}(-1)^{r_2-k}\binom{r_2}{k}g_i(y+kh)\\ &= \sum_{k=0}^{r_2-1}(-1)^{r_2-1-k}\binom{r_2-1}{k}[g_i(y+(k+1)h)-g_i(y+kh)]\\ &= \sum_{k=0}^{r_2-1}(-1)^{r_2-1-k}\binom{r_2-1}{k}h\int_0^1 g_i'(y+kh+th)dt\\ &= h\int_0^1\Delta_h^{r_2-1}(g_i')(y+th)dt\end{aligned}$$

Then we obtain

$$\begin{aligned}\left\|\Delta_{2,h}^{r_2}(\sum_{i\in I} f_i\otimes g_i)\right\|^2_{A_{h,2}^{r_2}} &= \iint_{A_{h,2}^{r_2}}\left(h\int_0^1\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')(x,y+th)dt\right)^2 dxdy\\ &\leq h^2\iint_{A_{h,2}^{r_2}}\int_0^1(\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')(x,y+th))^2 dtdxdy\end{aligned}$$

using the Schwartz inequality. By inverting the integrals, we get

$$\begin{aligned}
\left\|\Delta_{2,h}^{r_2}(\sum_{i\in I} f_i\otimes g_i)\right\|^2_{A_{h,2}^{r_2}} &\leq h^2\int_0^1\iint_{A_{h,2}^{r_2}}(\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')(x,y+th))^2 dxdydt\\
&\leq h^2\int_0^1\iint_{A_{h,2}^{r_2-1}}(\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')(x,z))^2 dxdzdt\\
&\leq h^2\left\|\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')\right\|^2_{A_{h,2}^{r_2-1}}
\end{aligned}$$

It follows that

$$t^{-\alpha_2}\sup_{|h|\leq t}\left\|\Delta_{2,h}^{r_2}(\sum_{i\in I} f_i\otimes g_i)\right\|_{A_{h,2}^{r_2}}\leq t^{1-\alpha_2}\sup_{|h|\leq t}\left\|\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')\right\|_{A_{h,2}^{r_2-1}}$$

Finally

$$\begin{aligned}
|\sum_{i\in I} f_i\otimes g_i|_{B_{2,\infty}^{(\alpha_1,\alpha_2)}(A)} &\leq \max_{i\in I}\left(\|g_i\|_{A_2}\|g_i'\|_{A_2}^{-1}\right)t^{-\alpha_1}\sup_{|h|\leq t}\left\|\Delta_{1,h}^{r_1}(\sum_{i\in I} f_i\otimes g_i')\right\|_{A_{h,1}^{r_1}}\\
&\quad + t^{1-\alpha_2}\sup_{|h|\leq t}\left\|\Delta_{2,h}^{r_2-1}(\sum_{i\in I} f_i\otimes g_i')\right\|_{A_{h,2}^{r_2-1}}\\
&\leq \max(1,\max_{i\in I}\left(\|g_i\|_{A_2}\|g_i'\|_{A_2}^{-1}\right))|\sum_{i\in I} f_i\otimes g_i'|_{B_{2,\infty}^{(\alpha_1,\alpha_2-1)}(A)}
\end{aligned}$$

which yields the result. □

References

Baraud, Y., Comte, F., and Viennet, G. (2001). Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.* **29**, 839-875.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301-413.

Bitouzé, D., Laurent, B., and Massart, P. (1999). A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. H. Poincaré Probab. Statist.* **35**, 735-763.

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

Brunel, E., Comte, F. and C. Lacour (2008). Adaptive estimation of the conditional density in presence of censoring. *Sankhya*, to appear.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* **17**, 1157-1167.

Ducharme, G. R. and Mint El Mouvid, M. (2001). Almost sure convergence of the local linear estimator of the conditional cumulative distribution function. *C. R. Acad. Sci. Paris Sér. I Math.* **333**, 873-876.

Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560-570.

Gonzalez-Manteiga, W., Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonparametr. Statist.* **4**, 65–78.

Hall, P., Wolff, R.C.L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94**, 154-163.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Hochmuth, R. (2002). Wavelet characterizations for anisotropic Besov spaces. Appl. *Comput. Harmon. Anal.* **12**, 179-208.

Koul, Susarla, H., V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

Lacour, C. (2007). Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist.* **43**, 571-597.

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.* **23**, 787-823.

Lo, S. H., Mack, Y. P., and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Related Fields* **80**, 461-473.

Nikol'skii, S. M. (1975). *Approximation of functions of several variables and imbedding theorems.* Springer-Verlag, New York. Translated from the Russian by John M. Danskin, Jr., Die Grundlehren der Mathematischen Wissenschaften, Band 205.

Ould-Saïd, E. (2006). A strong uniform convergence rate of kernel conditional quantile estimator under random censorship. *Statist. Probab. Lett.* **76**, 579-586.

Stute, W. (1986). Conditional empirical processes. *Ann. Statist.* **14**, 638-647.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.