ESTIMATION OF CONDITIONAL CUMULATIVE DISTRIBUTION FUNCTION FROM CURRENT STATUS DATA

SANDRA PLANCADÉ

ABSTRACT. Consider a positive random variable of interest $Y$ depending on a covariate $X$, and a random observation time $T$ independent of $Y$ given $X$. Assume that the only knowledge available about $Y$ is its current status at time $T$: $\delta = \mathbb{I}_{\{Y \leq T\}}$. This paper presents a procedure to estimate the conditional cumulative distribution function $F$ of $Y$ given $X$ from an independent identically distributed sample of $(X,T,\delta)$.

A collection of finite-dimensional linear subsets of $L^2(\mathbb{R})$ called models are built as tensor products of classical approximation spaces of $L^2(\mathbb{R})$. Then a collection of estimators of $F$ is constructed by minimization of a regression-type contrast on each model and a data driven procedure allows to choose an estimator among the collection. We show that the selected estimator converges as fast as the best estimator in the collection up to a multiplicative constant and is minimax over anisotropic Besov balls. Finally simulation results illustrate the performance of the estimation and underline parameters that impact the estimation accuracy.

MSC 2000 Subject Classifications. 62G05-62N02

Keywords . Distribution function, Model selection, Anisotropic Besov spaces, Adaptivity.

1. INTRODUCTION

In some survival analysis studies, the observation of a positive variable of interest $Y$ called lifetime, is restricted to the knowledge of whether or not $Y$ exceeds a random measure time $T$. We only observe the time $T$ and the “current status” of the system at time $T$, namely $\delta = \mathbb{I}_{\{Y \leq T\}}$. Such data arise naturally for example in infectious disease studies, when the time $Y$ of infection is unobserved, and a test is carried out at time $T$. This framework is also called interval censoring (case 1) since the observation $(T,\delta)$ indicates whether $Y$ lies in $[0,T]$ or $(T, +\infty)$. The lifetime $Y$ and the observation time $T$ may depend on observed covariates $X$, and $Y$ and $T$ are usually assumed to be independent given $X$.

Current status data have been widely studied for the last two decades. Most results about nonparametric estimation of the survival function are based on NPMLE (Nonparametric Maximum Likelihood Estimator). Groeneboom and Wellner [1992] prove that the NPMLE is pointwise convergent at rate $n^{-1/3}$ which is the optimal rate, and van de Geer
[1993] establishes a similar result for the $L^2$-risk. This unusual rate of convergence differs from the uncensored and right-censored cases, in which the distribution function can be estimated with the parametric rate of convergence $n^{-1/2}$. Besides, as far as the author knows, no minimax rate of convergence has been computed on classical regularity spaces. More recently, estimators developed from the NPMLE allow to take into account the known regularity of the function. Hudgens et al. [2007] build three estimators derived from the NPMLE, and compare their performances on simulated and real data. van der Vaart and van der Laan [2006] apply smoothing methods to the NPLME to estimate the survival function from current status data in presence of high dimensional covariates. Birgé [1999] proposes an easily computable histogram estimator which reaches the minimax rate of convergence. Nevertheless the procedures proposed in these papers are not adaptive on classical regularity spaces. Few results about adaptivity are available, and they do not include covariates: Ma and Kosorok [2006] introduce a NPMLE and a least square estimator on Sobolev classes, and select the regularity parameter with a penalized criterion. Brunel and Comte [2009] consider a least-square estimator on classical bases and introduce a model selection procedure with a more easily computable penalty function.

We consider an i.i.d. sample $(X_i, Y_i)_{i=1,\ldots,n}$, where the $(X_i)$’s are i.i.d. random variables with common density $f_X$, and the $(Y_i)$’s are positive variables called survival times. For every $i$, $Y_i$ depends on $X_i$, and we denote by $F(x, y)$ the cumulative distribution function (c.d.f.) of $Y_i$ given $X_i$, namely

$$F(x, y) = P[Y \leq y | X = x]$$

where $P[E_1 | E_2]$ denotes the conditional probability of $E_1$ given $E_2$. We consider an i.i.d. sample $(T_i)_{i=1,\ldots,n}$ of positive random variables such that for every $i \in \{1, \ldots, n\}$, $T_i$ and $Y_i$ are independent given $X_i$, and we observe the sample

$$(1) \quad (X_i, T_i, \delta_i = \mathbb{1}_{Y_i \leq T_i})_{i=1,\ldots,n}.$$ 

This paper presents an estimator of the conditional cumulative distribution function $F$ from the sample (1). The estimation procedure, inspired from Brunel and Comte [2009], is based on the following heuristic. For every $(x, u)$,

$$\mathbb{E}[\delta|(X, T) = (x, u)] = \mathbb{E}[\mathbb{1}_{Y \leq u} | (X, T) = (x, u)].$$

Given $X = x$, $Y$ and $T$ are independent, thus

$$\mathbb{E}[\delta|(X, T) = (x, u)] = \mathbb{E}[\mathbb{1}_{Y \leq u} | X = x] = P[Y \leq u | X = x] = F(x, u).$$

Thus $F$ is the regression function of $\delta$ over $(X, T)$, and the interval censoring issue turns into a regression function estimation problem where all the variables involved $(X, T, \delta)$ are observed. Therefore we can apply methods developed for regression function estimation to our issue.

More precisely we consider a collection of linear subset of $L^2(\mathbb{R}^2)$, and build an estimator by minimization of a least-square contrast on each subset. Then a model selection criterion provides an estimator which converges as well as the best estimator among the collection, up to a multiplicative constant: the estimator is said to be adaptive. We first state adaptivity conditionally to the observed $\{(X_i, T_i)\}$’s under weak assumptions, for the
CDF ESTIMATION FROM CURRENT STATUS DATA

empirical norm:

$$\|\hat{F} - F\|_n^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{F}_i - F_i)^2 (X_i, T_i),$$

via a nonclassical use of Talagrand Inequality. The empirical risk $$\|\hat{F} - F\|_n^2$$ naturally arises since it indicates the quality of estimation on the set of observations $$\{(X_i, T_i)\}$$. Besides considering the empirical norm allows a direct transposition of the result for non random observation times ($$T_i$$)'s and covariates ($$X_i$$)'s. Then the integrated risk is controlled under additional assumptions about the collection of models as well as minimum regularity conditions for $$F$$. Nevertheless, considering the integrated risk enables us to conduct a minimax study and prove that our estimator is optimal over anisotropic Besov balls $$B^{\beta}_{2,\infty}(L)$$.

The paper is organised as follows. Section 2 introduces the tools involved in the estimation procedure. The definition of the estimator and the main result are presented in Section 3. In Section 4, we study the rate of convergence of the estimator over anisotropic Besov balls and prove that it is minimax. A numerical study is conducted on simulated data in Section 5. Section 6 is devoted to the proofs. Section 7 gathers the Talagrand deviation inequality used in the proofs and a linear algebra technical lemma.

2. Tools

2.1. Notations. For every i.i.d. random variables $$\{V_i, W_i\}$$, we denote by $$f_V$$ the density of $$V_i$$, by $$f_{V,W}(v, w)$$ the density of the couple $$(V_i, W_i)$$ and by $$f_{V|W}(v, w)$$ the conditional density of $$V_i$$ at $$v$$ given $$W_i = w$$ for every $$i = 1, \ldots, n$$.

The conditional cumulative distribution function $$F(x, y)$$ is estimated on a compact set $$A = A_1 \times A_2$$ where $$A_1$$ is a compact interval of $$\mathbb{R}$$, and $$A_2 = [0, a_2]$$ for some positive $$a_2$$.

For every $$t, s \in L^2(A)$$, let

$$\langle s, t \rangle_n = \frac{1}{n} \sum_{i=1}^{n} s(X_i, T_i) t(X_i, T_i) \quad \text{and} \quad \|s\|_n^2 = \langle s, s \rangle_n.$$ 

The expectations of the former are denoted by:

$$\langle s, t \rangle_{f_{(X,T)}} = \int_{x \in A_1} \int_{u \in A_2} s(x, u) t(x, u) f_{(X,T)}(x, u) dxdw \quad \text{and} \quad \|s\|^2_{f_{(X,T)}} = \langle s, s \rangle_{f_{(X,T)}}.$$ 

Let $$M$$ be a symmetric matrix of dimension $$d \times d$$ with non-negative coefficient, we denote by $$\rho$$ the spectral radium of $$M$$ :

$$\rho(M) = \sup \left( \frac{d}{\sum_{i,j=1}^{d} M_{i,j} a_i a_j} \right) = \sup \left( \frac{\sum_{i,j=1}^{d} M_{i,j} |a_i| |a_j|}{\sum_{i=1}^{d} a_i^2} \right).$$
2.2. Collection of models. We construct a collection of finite-dimensional linear subsets of \( L^2(A) \) called models as tensor products of models on \( A_1 \) and \( A_2 \). For \( j = 1 \) or \( 2 \), consider a collection of linear subsets of \( L^2(A_j) \)

\[
M_n^{(j)} = \{ S_{m_j}^{(j)}, m_j \in I_n^{(j)} \} \quad \text{where} \quad \text{Dim}(S_{m_j}^{(j)}) = D_{m_j}^{(j)} < +\infty.
\]

Then for every \( m = (m_1, m_2) \in I_n = I_n^{(1)} \times I_n^{(2)} \), we define

\[
S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)} = \left\{ t : A \to \mathbb{R}, \quad t(x, y) = \sum_{(k,l) \in J_m} a_{k,l} \phi_k^{m_1}(x) \psi_l^{m_2}(y) \right\}
\]

where \((\phi_k^{m_1})_{k=1,...,D_{m_1}^{(1)}}\) is an orthonormal basis of \( S_{m_1}^{(1)} \), \((\psi_l^{m_2})_{l=1,...,D_{m_2}^{(2)}}\) an orthonormal basis of \( S_{m_2}^{(2)} \) and

\[
(2) \quad J_m = \left( (1,1), \ldots, (1,D_{m_2}^{(2)}), (2,1), \ldots, (2,D_{m_2}^{(2)}), \ldots, (D_{m_1}^{(1)},1), \ldots, (D_{m_1}^{(1)},D_{m_2}^{(2)}) \right)
\]

We consider the following assumption that restricts the number of models in collections \( M_n^{(j)} \).

(H) Let \( j = 1 \) or \( 2 \). For every \( b > 0 \), there exists a constant \( B_j \) such that

\[
\sum_{m_j \in I_n^{(j)}} \exp \left( -b \sqrt{\frac{D_{m_j}^{(j)}}{m_j}} \right) \leq B_j, \quad \forall n \in \mathbb{N}^*.
\]

2.3. Regression-type contrast. The contrast is based on the following result which generalizes the heuristic presented in Section 1. It is proved in Section 6.1.

**Lemma 2.1.** Almost surely,

\[
\mathbb{E}[\delta_1(t(X_1, T_1)|(X_1, T_1)) = F(X_1, T_1)t(X_1, T_1).
\]

As already noticed in the introduction, considering \( t \equiv 1 \), this amounts to say that \( F \) is the regression function of \( \delta_1 \) over \((X_1, T_1)\). Thus we consider the classical least square contrast for regression function estimation:

\[
\gamma_n(t) = \frac{1}{n} \sum_{i=1}^{n} (t(X_i, T_i) - \delta_i)^2
\]

which measures the accuracy of the approximation of the \{\delta_i\}'s by the \{t(X_i, T_i)\}'s (see e.g. Baraud [2002]). Let us explain more precisely why it is relevant. For every \( t \in L^2(A) \),

\[
(3) \quad \gamma_n(t) = \|t\|_n^2 + \|\delta\|_n^2 - \frac{2}{n} \sum_{i=1}^{n} \delta_i t(X_i, T_i)
\]

where \( \|\delta\|_n = (1/n) \sum_{i=1}^{n} \delta_i^2 \). Thus

\[
\mathbb{E}[\gamma_n(t)] = \|t\|_n^2 + \frac{2}{n} \sum_{i=1}^{n} \mathbb{E}[\delta_i t(X_i, T_i)] + C
\]
with \(C = \mathbb{E}[\delta_1]\) is independent of \(t\). Therefore, by Lemma 2.1,
\[
\mathbb{E}[\gamma_n(t)] = \|t\|^2_{f(X,T)} - 2\langle F, t \rangle f(X,T) + C = \|F - t\|^2_{f(X,T)} + C'
\]
with \(C'\) independent of \(t\). This shows that minimizing \(\gamma_n(.)\) is equivalent to minimizing \(\|F - .\|_{f(X,T)}\) and should provide a function close to \(F\).

2.4. Minimum contrast estimators. For every model \(S_m \in \mathcal{M}_n\) we consider the estimator
\[
\hat{F}_m = \arg \min_{t \in S_m} \gamma_n(t).
\]
The definition (4) amounts to stating that \(\partial \gamma_n(\hat{F}_m)/\partial a_{k',l'} = 0\) for every \((k',l') \in J_m\).

Then, denote \(\hat{F}_m(x,u) = \sum_{(k,l) \in J_m} \hat{a}_{k,l}\phi_k^{m_1}(x)\psi_l^{m_2}(u)\), the coefficient column vector \(\hat{A}_m = [\hat{a}_{k,l}]_{(k,l) \in J_m}\) satisfies:
\[
\hat{G}_m \hat{A}_m = \hat{V}_m
\]
where
\[
\hat{G}_m = \left[ \frac{1}{n} \sum_{i=1}^n \phi_k^{m_1}(X_i)\psi_l^{m_2}(T_i) \phi_k^{m_1}(X_i)\psi_l^{m_2}(T_i) \right]_{(k,l),(k',l') \in J_m^2}
\]
is the \(D_m \times D_m\)-square Gram matrix related to \(\{\phi_j^{m_1}\psi_l^{m_2}\}_{(k,l) \in J_m}\) for the scalar product \(\langle \cdot, \cdot \rangle_n\) and
\[
\hat{V}_m = \left[ \frac{1}{n} \sum_{i=1}^n \phi_k^{m_1}(X_i)\psi_l^{m_2}(T_i) \delta_i \right]_{(k,l) \in J_m}
\]
is a column vector.

**Comment.** As the matrix \(\hat{G}_m\) is not always invertible, equation (5) does not provide a unique solution \(\hat{A}_m\). Nevertheless, consider an observed sample (1). Let \(\tilde{S}_m\) be the subset of \(\mathbb{R}^n\) defined by
\[
\tilde{S}_m = \{ (t(X_1,T_1), \ldots, t(X_n,T_n)) : t \in S_m \}
\]
and \(\tilde{Z}_m = \arg \min_{Z \in \tilde{S}_m} (1/n) \sum_{i=1}^n (Z_i - \delta_i)^2\). \(\tilde{Z}_m\) is the projection of \((\delta_1, \ldots, \delta_n)\) on \(\tilde{S}_m\) for the canonical norm on \(\mathbb{R}^n\), so \(\tilde{Z}_m\) is uniquely defined. Moreover, by definition of \(\tilde{S}_m\), there exists at least one function \(G \in S_m\) such that \(\tilde{Z}_m = (G(X_1,T_1), \ldots, G(X_n,T_n))\). Then \(G\) minimises \(\gamma_n(t)\) on \(S_m\). Moreover, if two such functions \(G\) exist, they are equal on the set \(\{(X_i,T_i)\}\), so \(\|\hat{F}_m - F\|^2_n\) remains the same. For that reason, the definition (4) of \(\hat{F}_m\) is sensible for the risk \(\mathbb{E}\left[\|\hat{F}_m - F\|^2_n\{((X_i,T_i))\}_{i=1,...,n}\right]\).

2.5. Bias-variance decomposition. The minimization of the contrast \(\gamma_n\) over the collection of models \(\mathcal{M}_n\) carried out in Section 2.4 provides a collection of estimators \(\{\hat{F}_m, m \in I_n\}\). Considering the empirical norm \(\|\cdot\|_n\), the best model, called the oracle, is the one which minimizes
\[
\mathbb{E}\left[\|\hat{F}_m - F\|^2_n\{((X_i,T_i))\}_{i=1,...,n}\right].
\]
This model is unknown, but the model selection procedure originally developed by Birgé and Massart allows to select a model which approaches the oracle. With Pythagoras’ Theorem, for every $m \in I_n$ the risk (6) splits in two terms called bias and variance:

\[
\mathbb{E} \left[ \left\| \hat{F}_m - F \right\|^2 \right] \{ (X_i, T_i) \}_{i=1,\ldots,n} = \| F - F_m \|_n^2 + \mathbb{E} \left[ \| \hat{F}_m - F_m \|_n^2 \right] \{ (X_i, T_i) \}_{i=1,\ldots,n},
\]

where $F_m = \text{arg min}_{t \in S_m} \| F - t \|_m^2$. We will build an estimator of this bias-variance sum and minimize it to select a model $\hat{m}$ (see Birgé and Massart [1998] for more details).

In order to clarify the calculations with conditional expectations, we adopt the following notations. Let $\{ (x_i, u_i) \}_{i=1,\ldots,n}$ be in $A^n$, we define the set

\[
A = \{ X_1 = x_1, \ldots, X_n = x_n, T_1 = u_1, \ldots, T_n = u_n \}
\]

and for every $s, t \in L^2(A)$ we set

\[
\langle s, t \rangle_0 = \frac{1}{n} \sum_{i=1}^n t(x_i, u_i)s(x_i, u_i) \quad \text{and} \quad \| t \|_0^2 = \frac{1}{n} \sum_{i=1}^n t^2(x_i, u_i).
\]

The norm and scalar product $\| . \|_0$ and $\langle ., . \rangle_0$ are equal to $\| . \|_n$ and $\langle ., . \rangle_n$ on the set $A$, thus

\[
\left\{ \begin{array}{l}
\mathbb{E} \left[ \| \hat{F}_m - F_m \|_n^2 | A \right] = \mathbb{E} \left[ \| \hat{F}_m - F_m \|_0^2 \right] , \\
\| F - F_m \|_n = \| F - F_m \|_0 \quad \text{on} \ A.
\end{array} \right.
\]

We consider a $\| . \|_0$-orthogonal basis of $S_m$: $(\varphi_\lambda)_{\lambda \in I_m}$ such that $\| \varphi_\lambda \|_0 = 0$ or 1 for every $\lambda \in I_m$. (Lemma 7.1 states the existence of such a basis.) Note that it is only a tool for variance upper bound and it is not involved in the estimation.

**Upper bound on the variance term** $\mathbb{E} \left[ \| \hat{F}_m - F_m \|_0^2 \right]$. By (10), $F_m = \text{arg min}_{t \in S_m} \| F - t \|_0$ on the set $A$, hence $F_m = \sum_{\lambda \in I_m} \langle \varphi_\lambda, F \rangle_0 \varphi_\lambda$. Let $\tilde{F}_m = \sum_{\lambda \in J_m} \tilde{b}_\lambda \varphi_\lambda$. Similarly to Section 2.4, the equality $\hat{F}_m = \text{arg min}_{t \in S_m} \gamma_n(t)$ is equivalent to

\[
\sum_{\lambda \in J_m} \tilde{b}_\lambda \left( \frac{1}{n} \sum_{i=1}^n \varphi_\lambda(X_i, U_i) \varphi_\lambda(X_i, U_i) \right) = \frac{1}{n} \sum_{i=1}^n \delta_0 \varphi_\lambda(X_i, U_i), \quad \forall \lambda' \in I_m.
\]

The family $(\varphi_\lambda)_{\lambda \in I_m}$ is $\| . \|_0$-orthogonal thus on the set $A$, (11) is equivalent to

\[
\tilde{b}_\lambda \| \varphi_\lambda \|_0^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq u_i} \varphi_\lambda(x_i, u_i), \quad \forall \lambda' \in I_m.
\]

For every $\lambda$ such that $\| \varphi_\lambda \|_0 \neq 0$, $\tilde{b}_\lambda = (1/n) \sum_{i=1}^n \mathbb{1}_{Y_i \leq u_i} \varphi_\lambda(x_i, u_i)$. If $\| \varphi_\lambda \|_0 = 0$ then $\varphi_\lambda(x_i, u_i) = 0$ for every $i$, therefore the arbitrary value of $\tilde{b}_\lambda$ does not affect the expression of $\hat{F}_m(x_i, u_i)$ and
where

$$\phi(E[\varphi(X, U)] - E[\varphi(X, U)]) = \left(\frac{1}{n} \sum_{i=1}^{n} (\varphi(X, U_i) - \delta_i) \right)^2$$

which is naturally estimated on $\mathcal{A}$ by

$$\min_{t \in S_m} \frac{1}{n} \sum_{i=1}^{n} (t(X_i, U_i) - \hat{\delta}_i)^2 = \gamma_n(\hat{F}_m).$$

3. Definition of the estimator and main result

3.1. Definition of the estimator of $F$. Consider the collection of model $\mathcal{M}_m$ defined in Section 2.2, the contrast $\gamma_n$ defined in (3) and the collection of estimators $\{\hat{F}_m, m \in I_n\}$ where $\hat{F}_m$ is defined in (4). Following the model selection procedure presented in Section 2.5 with the bias and variance estimations (12) and (13), we select the model:

$$\hat{m} = \arg \min_{m \in I_n} \left[ \gamma_n(\hat{F}_m) + \text{pen}(m) \right]$$

where $\text{pen}(m) = \theta D_m/n$ for some numerical constant $\theta > 1$.

Besides, the target function $F$ lies in $[0,1]$ by definition. We use this information to improve the estimation by constraining the values of our estimator to remain in the same interval. More precisely we consider the estimator $\hat{F}_{\hat{m}}$ where

$$\hat{F}_{\hat{m}}(x, u) = \begin{cases} 0 & \text{if } \hat{F}_{\hat{m}}(x, u) < 0 \\ 1 & \text{if } \hat{F}_{\hat{m}}(x, u) > 1 \\ \hat{F}_{\hat{m}}(x, u) & \text{otherwise.} \end{cases}$$
1) The restriction imposed in (14) is not only necessary in order to prove some convergence results but also improves the estimation. Indeed, for every \((x, u) \in A\),
\[
|\hat{F}_m(x, u) - F(x, u)| \leq |\hat{F}_m(x, u) - F(x, u)|, \quad \forall (x, u) \in A, \forall m \in \mathbb{N}^*
\]
almost surely. In particular, \(\|\hat{F}_m - F\|_n^2 \leq \|\hat{F}_m - F\|_n^2\). Thus, any upper bound on \(\mathbb{E}[\|\hat{F}_m - F\|_n^2 | \{(X_i, T_i)\}_{i=1,\ldots,n}]\) is an upper bound on \(\mathbb{E}[\|\hat{F}_m - F\|_n^2 | \{(X_i, T_i)\}_{i=1,\ldots,n}]\).

2) The condition on \(\theta\) could be weakened to \(\theta > 1/4\) under the same assumptions with slight technical changes in the proofs, but we assume that \(\theta > 1\) for sake of simplicity.

3) The convergence results presented in this paper are valid for any \(\theta > 1/4\), but in practical implementation a value of \(\theta\) has to be fixed. It can be either calibrated on simulated data from a large number of examples, or chosen a priori independently of the framework (a constant equal to 2 in the penalty is often considered as a reasonable value, see for example Massart [2008]).

4) Note that the constant involved in the penalty is a numerical constant whereas in many other frameworks it depends on unknown parameters of the problem and has to be estimated. This makes our model selection procedure especially simple to implement.

3.2. Risk for the empirical norm. The estimator \(\hat{F}_m\) satisfies the following oracle inequality.

**Theorem 3.1.** Assume that Assumption (H) holds, there exist numerical constants \(C_1\) and \(C_2\) such that almost surely,
\[
\mathbb{E}[\|\hat{F}_m - F\|_n^2 | \{(X_i, T_i)\}_{i=1,\ldots,n}] \leq C_1 \inf_{m \in I_n} \left\{ \inf_{t \in S_m} \|F - t\|_n^2 + \text{pen}(m) \right\} + \frac{C_2}{n}.
\]

**Comments**

1) For every model \(m \in I_n\), \(\left\{ \inf_{t \in S_m} \|F - t\|_n^2 + \text{pen}(m) \right\}\) has the same order as \(\|\hat{F}_m - F\|_n^2\) (see Section 2.5). Thus Theorem 3.1 indicates that up to a multiplicative constant, the model selection estimator \(\hat{F}_m\) converges as fast as the best estimator in the collection.

2) It is clear that the same result holds with non random observation times \((T_1, \ldots, T_n)\) and non random covariates \((X_1, \ldots, X_n)\).

3) According to Comment 2 in Section 3.1,
\[
\mathbb{E}[\|\hat{F}_m - F\|_n^2 | \{(X_i, T_i)\}_{i=1,\ldots,n}] \leq C_1 \inf_{m \in I_n} \left\{ \inf_{t \in S_m} \|F - t\|_n^2 + \text{pen}(m) \right\} + \frac{C_2}{n}.
\]

4. Convergence of the estimator on anisotropic Besov balls \(B_{2,\infty}^{\beta}\)

In this section, we prove that our estimator reaches the minimax rate of convergence over anisotropic besov balls. As the definition of Besov spaces refers to \(L^2\)-norms, it appears natural to consider the risk of our estimator for the integrated norm \(\|\cdot\|_{f(X,T)}\).

Convergence results are then derivated from Theorem 3.1 under additional assumptions.
4.1. Definition of anisotropic Besov spaces. We recall the definition of two-dimensional anisotropic Besov spaces stated for example in Hochmuth [2002]. Let \( \Omega \subset \mathbb{R}^2 \). For \( j = 1 \) or \( 2 \), \( r \in \mathbb{N}^* \) and \( h > 0 \), let

\[
\Delta_h^{r,j}(f)(x,y) = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f((x,y) + khe_j)
\]

be the directional partial difference operator for every \( (x,y) \in \Omega_{h,j}^r \) where \( \Omega_{h,j}^r = \{(x,y) \in \Omega, (x,y) + rhe_j \in \Omega \} \) and \( (e_1, e_2) \) is the canonical basis of \( \mathbb{R}^2 \). For \( t > 0 \), let \( \omega_{r,j}(f, t, \Omega) = \sup_{|h| \leq t} \| \Delta_h^{r,j}(f)(x,y) \|_{L^2(\Omega_{h,j})} \)

be the directional modulus of smoothness for the \( L^2 \)-norm. Let \( \beta = (\beta_1, \beta_2) \in (\mathbb{R}_+^*)^2 \) and \( r_j = \lceil \beta_j \rceil + 1 \). We define the anisotropic Besov space of parameters \((\beta, 2, \infty)\) as

\[
B_{2,\infty}^{\beta}(\Omega) = \left\{ f \in L^2(\Omega), |f|_{B_{2,\infty}^{\beta}(\Omega)} < +\infty \right\}
\]

where

\[
|f|_{B_{2,\infty}^{\beta}(\Omega)} = \sup_{t>0} \left[ t^{-\beta_1} \omega_{r_1,1}(f, t, \Omega) + t^{-\beta_2} \omega_{r_2,2}(f, t, \Omega) \right].
\]

We consider the following norm on \( B_{2,\infty}^{\beta}(\Omega) \): \( \| f \|_{B_{2,\infty}^{\beta}(\Omega)} = |f|_{B_{2,\infty}^{\beta}(\Omega)} + \| f \| \) and for \( L > 0 \),

\[
B_{2,\infty}^{\beta}(A, L) = \left\{ f \in B_{2,\infty}^{\beta}(A), \| f \|_{B_{2,\infty}^{\beta}(A)} \leq L \right\}.
\]

4.2. Additional assumptions. We consider bound conditions for \( f(x,t) \) as well as additional assumptions about the collections \( M_n^{(1)} \) and \( M_n^{(2)} \).

\( (A_1) \) There exist \( h_0 > 0 \), \( h_1 < +\infty \) such that \( h_0 \leq f(x,t)(x,u) \leq h_1 \), \( \forall (x,u) \in A \).

\( (A_2) \) For \( j = 1 \) and \( 2 \),

\[(16) \quad |M_n^{(j)}| \leq P^{(j)}(n), \quad \forall n \in \mathbb{N}
\]

for some polynomial \( P^{(j)} \). Moreover, there exists a model \( S_n^{(j)} \in M_n^{(j)} \) of dimension \( N_n^{(j)} \) such that, for every \( m_j \in I_n^{(j)} \), \( S_n^{(j)} \subset S_n^{(j)} \). Besides \( N_n^{(1)}N_n^{(2)} \leq \sqrt{n} \log n \).

\( (A_3) \) There exists a positive constant \( K_1 \) (resp. \( K_2 \)) such that, for every \( m_1 \in I_n^{(1)} \) (resp. \( m_2 \in I_n^{(2)} \)),

\[
\sup_{x \in A_1} \sum_{k=1}^{D_{m_1}^{(1)}} (\phi_k^{m_1}(x))^2 \leq K_1 D_{m_1}^{(1)} \quad \left( \text{resp.} \sup_{u \in A_2} \sum_{l=1}^{D_{m_2}^{(2)}} (\psi_l^{m_2}(u))^2 \leq K_2 D_{m_2}^{(2)} \right).
\]

Assumption \( (A_2) \) refers to the number of models in the collection, whereas \( (A_3) \) depends on the nature of the models. Assumption \( (A_3) \) holds in particular if the collections \( M_n^{(1)} \) and \( M_n^{(2)} \) consist of the following models.

\( (i) \) \( S_n^{(j)} \) is the set of piecewise polynomials with maximum degree \( s_j \) and step \( \text{lg}(A_j)/D_{m_j}^{(j)} \),

where \( \text{lg}(A_j) \) denotes the length of \( A_j \).
(ii) \( S_{m_j}^{(j)} = \text{vect}\{\chi_{l,k}, l \leq m_j, k \in \mathbb{Z}\} \) where \( \chi \) is a mother wavelet with regularity \( s_j \), \( \chi_{l,k}(x) = 2^{l/2}\chi(2^l x - k) \) and \( D_{m_j}^{(j)} = 2^{m_j} \).

(iii) \( S_{m_j}^{(j)} \) is the set of trigonometric polynomials with maximum degree \( D_{m_j}^{(j)} \).

### 4.3. Upper bound of \( \|\tilde{F}_m - F\|_{f(X,T)} \)

Under the additional assumptions from Section 4.2, Theorem 3.1 leads to the following result.

**Corollary 4.1.** Assume that (H), (A_1), (A_2) and (A_3) hold then

\[
\mathbb{E}\left[\|\tilde{F}_m - F\|^2_{f(X,T)}\right] \leq C_3 \inf_{m \in I_n} \left\{ \inf_{t \in S_m} \|F - t\|^2_{f(X,T)} + \text{pen}(m) \right\} + \frac{C_4}{n}
\]

where \( C_3 \) is a numerical constant and \( C_4 \) depends on \( h_0 \) and \( K \).

**Comment** Corollary 4.1 indicates that the rate of convergence of \( \tilde{F}_m \) for the \( \|\cdot\|_{f(X,T)} \)-risk is the one of the best estimator among the collection \( \{\tilde{F}_m, m \in I_n\} \) (see Comment after Theorem 3.1).

For the models (i) – (iii) described in Section 4.2, the bias term \( \inf_{t \in S_m} \|F - t\|^2_{f(X,T)} \) in the right-hand member of equation (17) is upper bounded on anisotropic Besov spaces.

**Lemma 4.1.** Assume that \( F \in \mathcal{B}_{2,\infty}^\beta(A,L) \) for some \( L > 0 \) and \( \beta = (\beta_1, \beta_2) \in (\mathbb{R}^+)^2 \), and the collection \( \mathcal{M}_n \) is set up from linear models (i), (ii) with \( s_j > \beta_j - 1 \), or (iii). There exists a positive constant \( C_0 \) such that

\[
\inf_{t \in S_m} \|F - t\| \leq C_0 \left( (D_{m_1}^{(1)})^{-\beta_1} + (D_{m_2}^{(2)})^{-\beta_2} \right).
\]

Lemma 4.1 is proved in Lacour [2007] based on papers from Hochmuth [2002] and Nikol’skii [1975]. Inserting this result into Corollary 4.1 provides the rate of convergence of our estimator on anisotropic Besov spaces.

**Corollary 4.2.** Assume that \( F \in \mathcal{B}_{2,\infty}^{\beta_1,\beta_2}(A,L) \) with \( \beta_1, \beta_2 > 1 \) and \( \mathcal{M}_n \) is set up from models (i), (ii) with \( s_j > \beta_j - 1 \), or (iii). Moreover assume that (A_1) and (A_2) hold, and

\[
N_n^{(j)} \leq \left( \frac{n}{\log^2 n} \right)^{1/4} \quad \text{for} \quad j = 1, 2.
\]

Then

\[
\mathbb{E}\left[\|\tilde{F}_m - F\|^2\right] \leq C_5 n^{-\overline{\beta}/(\overline{\beta} + 1)}
\]

for some positive constant \( C_5 \), where \( \overline{\beta} = 2\beta_1\beta_2/(\beta_1 + \beta_2) \) is the harmonic mean of \( (\beta_1, \beta_2) \).
Indeed, for every \( m = (m_1, m_2) \),

\[
\mathbb{E} \left[ \| \hat{F}_n - F \|_{(X,T)}^2 \right] 
\leq C_3 \left\{ \inf_{t \in \mathcal{S}_m} \| F - t \|_{(X,T)}^2 + \text{pen}(m) \right\} + \frac{C_4}{n}
\leq C_3 \left\{ h_1 \inf_{t \in \mathcal{S}_m} \| F - t \|^2 + \text{pen}(m) \right\} + \frac{C_4}{n}
\leq C_3 \left\{ 2h_1 C_0 \left( (D_{m_1}^{(1)})^{-2\beta_1} + (D_{m_2}^{(2)})^{-2\beta_2} \right) + \theta \frac{D_{m_1}^{(1)} D_{m_2}^{(2)}}{n} \right\} + \frac{C_4}{n}
\]

where \( C_0 \) depends on \((\beta, L)\). Let \( \bar{m}_1 \) and \( \bar{m}_2 \) be such that

\[
1 \leq D_{\bar{m}_1}^{(1)} n^{-\beta_2/(\beta_1 + \beta_2 + 2\beta_1 \beta_2)} \leq 2 \quad \text{and} \quad 1 \leq D_{\bar{m}_2}^{(2)} n^{-\beta_1/(\beta_1 + \beta_2 + 2\beta_1 \beta_2)} \leq 2,
\]

(Assumption (18) guarantees the existence of such models for \( \beta_1, \beta_2 > 1 \)), then

\[
2h_1 C_0 \left( (D_{\bar{m}_1}^{(1)})^{-2\beta_1} + (D_{\bar{m}_2}^{(2)})^{-2\beta_2} \right) + \theta \frac{D_{m_1}^{(1)} D_{m_2}^{(2)}}{n} \leq C n^{-\bar{\beta} / (1+1)}.
\]

Moreover \( \mathbb{E} \left[ \| \hat{F}_n - F \|_{(X,T)}^2 \right] \leq (1/h_0) \mathbb{E} \left[ \| \hat{F}_n - F \|_{(X,T)}^2 \right] \) which proves Corollary 4.2. \( \Box \)

**Remark 4.1.** The condition \( \beta_1, \beta_2 > 1 \) in Corollary 4.2 can be generalised to

\[
(\beta_1, \beta_2) \in (\beta_1^*, +\infty) \times (\beta_2^*, +\infty)
\]

for a known couple \((\beta_1^*, \beta_2^*)\) with \( \overline{\beta}^* \geq 1 \), where \( \overline{\beta}^* \) is the harmonic mean of \( \beta_1^* \) and \( \beta_2^* \) by considering \( N_{n_1}^{(1)} \) and \( N_{n_2}^{(2)} \) such that

\[
N_{n_1}^{(1)} \leq (\log n)^{-1/2 n_{\beta_1^*}/(\beta_1^* + \beta_2^* + 2\beta_1^* \beta_2^*)} \quad \text{and} \quad N_{n_2}^{(2)} \leq (\log n)^{-1/2 n_{\beta_2^*}/(\beta_1^* + \beta_2^* + 2\beta_1^* \beta_2^*)}.
\]

This alternative assumption allows to take into account a priori knowledge on the regularity of \( F \) through an appropriate choice of \((N_{n_1}^{(1)}, N_{n_2}^{(2)})\). The estimation would be optimized by considering a smaller maximum size of models \((N_{n_1}^{(1)})\) in the direction where \( F \) is more regular.

### 4.4. Lower bound

Let \( \mathcal{F} \) be a set of conditional cumulative distribution functions on \( A \). A sequence \( (r_n)_{n \in \mathbb{N}} \) of positive numbers is called the minimax rate of convergence for \( F \) over \( \mathcal{F} \) if there exist two constants \( c \) and \( C \) such that

\[
c \leq \inf_{\hat{F}_n} \sup_{F \in \mathcal{F}} \left( r_n^{-1} \mathbb{E}[\| \hat{F}_n - F \|^2] \right) \leq C
\]

where the infimum is taken over all possible estimators \( \hat{F}_n \). Note that the minimax rate is defined up to a multiplicative constant.

According to Corollary 4.2, provided that \( \beta_1, \beta_2 > 1 \),

\[
\inf_{\hat{F}_n} \sup_{F \in \mathcal{B}_{2,\infty}(A,L)} \left( n^{-\overline{\beta}/(\overline{\beta}+1)} \mathbb{E}[\| \hat{F}_n - F \|^2] \right) \leq \sup_{F \in \mathcal{B}_{2,\infty}(A,L)} \left( n^{-\overline{\beta}/(\overline{\beta}+1)} \mathbb{E}[\| \hat{F}_n - F \|^2] \right) \leq C.
\]

Moreover, the following result holds.
Proposition 4.1. Let $\beta = (\beta_1, \beta_2) \in (1, +\infty) \times (1, +\infty)$. Assume that $h_1 = \|f_{(X,T)}\|_\infty < +\infty$, then there exists a constant $c$ which depends on $(\beta, L, h_1)$ such that

$$\inf_{\hat{F}_n} \sup_{F \in B_2^{\beta_2}(A,L)} \mathbb{E} \left[ n^{-\beta/(\beta +1)} \| \hat{F}_n - F \|^2 \right] \geq c.$$ 

Therefore, for every $\beta_1, \beta_2 > 1$, the minimax rate of convergence over $B_2^{\beta_2}(A,L)$ is $n^{-\beta/(\beta +1)}$, and $\hat{F}_n$ is minimax. This proves that our estimator adapts to the unknown regularity $\beta$ of the function $F$.

5. Graphical results on simulated data.

In this section, we present the performance of the estimator $\hat{F}_n$ on simulated data. In particular, we study the impact of the distance between the distributions of $Y$ and $T$ on the estimation accuracy.

5.1. About the numerical implementation. We have chosen to implement the procedure in an histogram basis. The basis of functions in which the estimator is computed are supposed to be fixed independently of the data, and the error of estimation is bounded on a set included in the support of the distribution of $(X,T)$. In practical cases this support is usually unknown and has to be estimated from the data. In our implementation, we consider histograms supported on the set

$$A = [\text{quantile}(0.01, X), \text{quantile}(0.99, X)] \times [\text{quantile}(0.01, T), \text{quantile}(0.99, T)].$$

We have chosen a constant $\theta = 2$ in the penalty but the estimation results seem quite robust when we change this value.

5.2. Results for several sample sizes. We consider the following distribution of $(X,Y,T)$.

$$\begin{cases} 
X \sim U([0,3]), \\
Y = X + \varepsilon \quad \text{with} \quad \varepsilon \sim \text{Exp}(1), \\
T = X + \varepsilon' \quad \text{with} \quad \varepsilon' \sim \text{Exp}(1).
\end{cases}$$

Figure 1 presents the conditional distribution function $F(x,y)$ as well as its estimators for several values of $n$. The same functions are plotted for a fixed $x$, $x = 2$, and a fixed $y$, $y = 3.3$, respectively on first and second rows of Figure 2. As expected the accuracy of the estimation increases with the size of the sample, but we notice that a quite large size of sample is required to get a correct estimation. This is not surprising given the nature of the current status framework in which the observed data give a very incomplete information about the variable of interest. Besides, we note that the estimation of the dependence of $F$ on $y$ when $x$ is fixed is substantially better than for the dependence on $x$. The same phenomenon is observed with other distributions of the variables.

5.3. Results for several distributions of the observation time. In a right-censoring framework, the rate of censoring (defined as the expected proportion of observations that are censored) is a parameter that impacts the accuracy of the estimation: the lower the rate of censoring, the better the estimation. Indeed as the rate of censoring decreases, the proportion of survival times actually observed increases and the estimation of the survival time distribution gets better.
In the interval censoring framework, the rate of censoring does not make sense. Nevertheless, as confirmed by simulations, the estimation accuracy is expected to increase while the distance between the distributions of $(X, T)$ and $(X, Y)$ decreases. Indeed consider a fixed $X = x$, the function $y \rightarrow F(x, y)$ varies more on a set where $f_{(X,Y)}$ is high and less on a set where $f_{(X,Y)}$ is small. Thus the estimation of $F$ improves if the observations $\{T_i\}$’s are concentrated on a set where $f_{(X,Y)}$ is high. On the opposite, if the main supports of $f_{(X,T)}$ and $f_{(X,Y)}$ are disconnected, the observations $\{T_i\}$’s provide no information about the distribution of $Y$ and the estimation will be impossible.
The link between the error of estimation and the distance between $f(X,T)$ and $f(X,Y)$ is implicit in the theoretical results since the risk upper bounded in Corollary 4.2 is weighted by $f(X,T)$.

In Section 5.3 we have considered a measurement time $T$ with same conditional distribution as $Y$. Now we add an offset $\alpha = 0.5, 1, 2, 3$ to the distribution of $T$.

Then the $L_1$-distance between $f(X,T)$ and $f(X,Y)$ is equal to $2(1 - \exp(-\alpha))$ and increases as $\alpha$ increases.

The true function $F$ is the same as Figure 1. Figure 3 presents $\hat{F}_m$ for several values of $\alpha$ and for $n = 3000$. The same plots are presented in Figure 4 for a fixed $x = 2$ (first row) and for a fixed $y = 3.3$ (second row). The product of intervals used to compute the estimators are the same as described in Section 5.1 but the plots are represented on a set which contains the main support of $(X,Y)$:

$$I = [\text{quantile}(0.01, X), \text{quantile}(0.99, X)] \times [0, \text{quantile}(0.99, Y)].$$

With the increase of $\alpha$ (corresponding to an increase of the distance between $f(X,T)$ and $f(X,Y)$) the estimation deteriorates. In particular for $\alpha = 2$, $\|f(X,T) - f(X,Y)\|_{L_1} = 1.7$ is close to 2 which indicates that the distributions of $(X,T)$ and $(X,Y)$ hardly overlap, and the estimation is very bad despite a large sample size.

6. Proofs

6.1. Proof of Lemma 2.1. Let $(x, u) \in \mathbb{R}^2$ be such that $f(X,T)(x, u) > 0$.

$$E[\delta_1|(X_1, T_1) = (x, u)] = E[\Pi_{(Y_1 \leq u)}|(X_1, T_1) = (x, u)] = \int_{A_2} \Pi_{(y \leq u)} \frac{f(Y,T)|X(y, x) f_X(x)}{f(X,T)(x, u)} dy.$$
Figure 4. Conditional distribution function $F(x, y)$ of $Y$ given $X$ (red line) and estimator $\tilde{F}_{\hat{m}}$ (black dotted line) for $n = 3000$, and for an offset $\alpha \in (0, 0.5, 1, 2)$ of the time of observation $T$. In the first column, $x = 2$ and in the second column $y = 3.3$.

$$\mathbb{E}[\delta_1 | (X_1, T_1) = (x, u)] = \int_{A_2} \mathbb{I}_{(y \leq u)} x \frac{f_{|X}(y, x)f_{|X,Y}(x, u)}{f_{|X}(x, u)} dy = \int_{A_2} \mathbb{I}_{(y \leq u)} f_{|X}(y, x) dy = F(x, u).$$

6.2. **Proof of Theorem 3.1.** Let $m = (m_1, m_2) \in I_n$ and $F_m \in S_m$. By definition of $\hat{m}$ and $\tilde{F}_{\hat{m}}$,

$$\gamma_n(\tilde{F}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(\tilde{F}_{m}) + \text{pen}(m) \leq \gamma_n(F_m) + \text{pen}(m).$$

Besides, for every $s, t \in S_n$,

$$\gamma_n(t) - \gamma_n(s) = \frac{1}{n} \sum_{i=1}^{n} \left[ (t(X_i, T_i) - \delta_i)^2 - (s(X_i, T_i) - \delta_i)^2 \right] = \| t - F \|^2_n - \| s - F \|^2_n - 2\nu_n(t - s)$$

where $\nu_n(t) = (1/n) \sum_{i=1}^{n} (\delta_i - F(X_i, T_i)) t(X_i, T_i)$. Thus (19) implies

$$\|\tilde{F}_{\hat{m}} - F\|^2_n \leq \|F_m - F\|^2_n + \text{pen}(m) - \text{pen}(\hat{m}) + 2\nu_n(\tilde{F}_{\hat{m}} - F_m) \leq \|F_m - F\|^2_n + \text{pen}(m) - \text{pen}(\hat{m}) + 2\|\tilde{F}_{\hat{m}} - F_m\|_n \sup_{t \in S_m + S_{\hat{m}}, \|t\|_n \leq 1} \nu_n(t).$$

This last inequality is the main distinction with the integrated risk framework (see e.g. Massart [2007]). In this context terms such as $\nu_n(\tilde{F}_{\hat{m}} - F_m)$ are upper bounded by $\|\tilde{F}_{\hat{m}} - F_m\|_g \sup_{t \in S_m + S_{\hat{m}}, \|t\|_g \leq 1} \nu_n(t)$ where $\cdot \|_g$ is the $L^2$-norm associated to a suitable
function \( g \). Technically, this change requires a non i.i.d. version of Talagrand inequality (Theorem 7.1) instead of the more classical i.i.d. version (see e.g. Lacour [2008], Section 6, Lemma 5). As a consequence weaker assumptions are required and smaller constants are obtained in the upper bounds.

For every function \( p(m, m') \) of \( m \) and \( m' \),
\[
\| \hat{F}_m - F \|_n^2 \leq \| F_m - F \|_n^2 + \text{pen}(m) - \text{pen}(\hat{m}) + \frac{1}{4} \| \hat{F}_m - F_m \|_n^2 + 4 \sup_{t \in S_m + S_{\hat{m}}, \| t \|_n \leq 1} (\nu_n(t))^2 \\
= \| F_m - F \|_n^2 + \text{pen}(m) - \text{pen}(\hat{m}) + 4p(m, \hat{m}) + \frac{1}{4} \| \hat{F}_m - F_m \|_n^2 \\
+ 4 \sup_{t \in S_m + S_{\hat{m}}, \| t \|_n \leq 1} (\nu_n(t))^2 - p(m, \hat{m})
\]

Now, consider \( p(m, m') = (1/4)(\text{pen}(m) + \text{pen}(m')) \) then
\[
\| \hat{F}_m - F \|_n^2 \leq \| F_m - F \|_n^2 + 2\text{pen}(m) + \frac{1}{4} \left(2 \| \hat{F}_m - F \|_n^2 + 2\| F_m - F \|_n^2\right) \\
+ 4 \sup_{t \in S_m + S_{\hat{m}}, \| t \|_n \leq 1} (\nu_n(t))^2 - p(m, \hat{m})
\]

and
\[
\frac{1}{2} \| \hat{F}_m(21F) \|_n^2 \leq \frac{3}{2} \| F_m - F \|_n^2 + 2\text{pen}(m) + 4 \sum_{m' \in I_n} \left(\sup_{t \in S_m + S_{m'}, \| t \|_n \leq 1} [\nu_n(t)]^2 - p(m, m')\right)
\]

The following result is derived from Talagrand Inequality (Theorem 7.1).

**Lemma 6.1.** There exist numerical constants \( C_0 \) and \( \kappa_0 \) which only depend on the constant \( \theta \) in the penalty such that, for every \( m, m' \in I_n \), \( (x_1, \ldots, x_n) \in A_1^n \) and \( (u_1, \ldots, u_n) \in A_2^n \),
\[
\mathbb{E} \left[ \sup_{t \in S_m + S_{m'}, (1/n) \sum_{i=1}^n t^2(x_i, u_i) \leq 1} \left( \frac{1}{n} \sum_{i=1}^n (I_{Y_i \leq u_i} - F(x_i, u_i))t(x_i, u_i) \right)^2 - p(m, m') \right]_+ \leq \frac{C_0}{n} \exp(-\kappa_0 \sqrt{D_m + D_{m'}})
\]

where \( \mathcal{A} \) is defined in (8).

Therefore, after plugging the result of Lemma 6.1 in (21), Assumption (H) leads to
\[
\mathbb{E} \left[ \| \hat{F}_m - F \|_n^2 | \{(X_i, T_i)\}_{i=1}^n \right] \leq 2\| F - F_m \|_n^2 + 4\text{pen}(m) + \frac{4C_0 B}{n}
\]

for some numerical constant \( B \), which concludes the proof of Theorem 3.1. \( \square \)

**Proof of Lemma 6.1**
The proof relies on Talagrand Inequality (Theorem 7.1). Let \((x_1, \ldots, x_n) \in A_1^n, (u_1, \ldots, u_n) \in A_2^n\) and \(m, m' \in I_n\). Let

\[
\mu_n(t) = \frac{1}{n} \sum_{i=1}^{n} (\mathbb{I}_{Y_i \leq u_i} - F(x_i, u_i)) t(x_i, u_i).
\]

Then

\[
Z = \sup_{t \in S_m + S_{m'}, \|t\|_0 \leq 1} (\mu_n(t))^2 = \sup_{f \in \mathcal{F}_{m,m'}} \left( \frac{1}{n} \sum_{i=1}^{n} f_i(\mathbb{I}_{Y_i \leq u_i}) \right)^2
\]

where \(\|\cdot\|_0\) is defined in (9) and \(\mathcal{F}_{m,m'}\) is the following set of functions from \(\mathbb{R}\) to \(\mathbb{R}^n\):

\[
\mathcal{F}_{m,m'} = \left\{ f = (f^{(1)}, \ldots, f^{(n)}), f^{(i)}(x) = t(x_i, u_i)(x - F(x_i, u_i)), t \in S_m + S_{m'} \text{ and } \|t\|_0 \leq 1 \right\}.
\]

Let \((\varphi_\lambda)_{\lambda=1}^{D_{m+m'}}\) be a \(\|\cdot\|_0\)-orthogonal basis of \(S_m + S_{m'}\) such that \(\|\varphi_\lambda\|_0 = 0\) or 1, where \(D_{m+m'}\) denotes the dimension of \(S_m + S_{m'}\) (see Lemma 7.1). Let \(\Gamma = \{\lambda \in \{1, \ldots, D_{m+m'}\}, \|\varphi_\lambda\|_0 \neq 0\}\) then for every \(t = \sum_{\lambda=1}^{D_{m+m'}} a_\lambda \varphi_\lambda \in S_m + S_{m'}, \|t\|_0^2 = \sum_{\lambda=1}^{D_{m+m'}} a_\lambda^2\).

In order to apply Theorem 7.1 to \(Z\) we have to compute \(b, v\) and \(H\). First we compute the term \(H\).

\[
E[Z^2 | A] = E \left[ \sup_{t \in S_m + S_{m'}, \|t\|_0 \leq 1} (\mu_n(t))^2 \right] = E \left[ \sup_{\sum_{\lambda=1}^{D_{m+m'}} a_\lambda^2 \leq 1} \left( \sum_{\lambda=1}^{D_{m+m'}} a_\lambda \mu_n(\varphi_\lambda) \right)^2 \right] A.
\]

Besides, for every \(\lambda \notin \Gamma\), \(\mu_n(\varphi_\lambda) = 0\). Therefore

\[
E[Z^2 | A] = E \left[ \sup_{\sum_{\lambda=1}^{D_{m+m'}} a_\lambda^2 \leq 1} \left( \sum_{\lambda=1}^{D_{m+m'}} a_\lambda \mu_n(\varphi_\lambda) \right)^2 \right] A \leq E \left[ \sup_{\sum_{\lambda=1}^{D_{m+m'}} a_\lambda^2 \leq 1} \left( \sum_{\lambda=1}^{D_{m+m'}} a_\lambda^2 \left( \sum_{\lambda=1}^{D_{m+m'}} \mu_n(\varphi_\lambda)^2 \right) \right) \right] A.
\]

In the same way as the upper bound of the variance term in Section 2.5, we obtain:

\[
E[Z^2 | A] \leq \frac{1}{4n} \sum_{\lambda=1}^{D_{m+m'}} \left( \frac{1}{n} \sum_{i=1}^{n} \varphi_\lambda(x_i, u_i) \right)^2 = \frac{1}{4n} \sum_{\lambda=1}^{D_{m+m'}} \|\varphi_\lambda\|_0^2 = \frac{|\Gamma|}{4n} \leq \frac{D_{m} + D_{m'}}{4n} = H^2.
\]

Now we compute the terms \(b\) and \(v\). \(\mathbb{I}_{\{\leq u_i\}}\) and \(F(x_i, u_i)\) are in \([0, 1]\), so \(\|\mathbb{I}_{\{\leq u_i\}} - F(x_i, u_i)\|_\infty \leq 1\) a.s. Moreover, let \(t \in S_m + S_{m'}\) be such that \(\|t\|_0 \leq 1\), for every \(i \in \{1, \ldots, n\}\)

\[
t^2(x_i, u_i) \leq \sum_{l=1}^{n} t^2(x_l, u_l) = n\|t\|_0^2 \leq n.
\]
Thus
\[
\sup_{t \in S_m + S_m', \|t\|_0 \leq 1} \left( \sup_{i=1, \ldots, n} \left\| (\mathbb{I}_{\langle \cdot, u_i \rangle} - F(x_i, u_i)) t(x_i, u_i) \right\|_{\infty} \right) \leq \sqrt{n} = b.
\]

Moreover
\[
\sup_{t \in S_m + S_m', \|t\|_0 \leq 1} \left( \frac{1}{n} \sum_{i=1}^{n} \text{Var} \left( (\mathbb{I}_{\{Y_i \leq u_i\}} - F(x_i, u_i)) t(x_i, u_i) \right) | X_i = x_i, T_i = u_i \right)
\]
\[
= \sup_{t \in S_m + S_m', \|t\|_0 \leq 1} \left( \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ (\mathbb{I}_{\{Y_i \leq u_i\}} - F(x_i, u_i))^2 \right] | X_i = x_i, T_i = u_i \right) t^2(x_i, u_i)
\]
\[
\leq \frac{1}{4} = v.
\]

Finally recalling that \( p(m, m') = \theta \|m\|^2 \) with \( \theta > 1 \), Theorem 7.1 concludes the proof of Lemma 6.1. \( \square \)

6.3. Proof of Corollary 4.1. The proof of is divided in two propositions. Let
\[
\Omega = \left\{ \left\| \frac{t}{\|t\|_{f(X,T)}} \right\|_n^2 \leq \frac{1}{2}, \forall t \in S_n \right\}.
\]

**Proposition 6.1.** Under the assumptions of Corollary 4.1,
\[
\mathbb{E} \left[ \left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 \mathbb{I}_\Omega \right] \leq C'_1 \inf_{m \in I_n} \left\{ \inf_{t \in S_m} \left\| F - t \right\|_{f(X,T)}^2 + \text{pen}(m) \right\} + C'_2 \frac{1}{n}
\]
where \( C'_1 \) and \( C'_2 \) are numerical constants.

**Proposition 6.2.** Under the assumptions of Corollary 4.1,
\[
\mathbb{E} \left[ \left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 \mathbb{I}_{\Omega^c} \right] \leq C_6 \frac{1}{n}
\]
where \( C_6 \) depends on \( h_0 \) and \( K \).

6.3.1. Proof of Proposition 6.1. First of all, \( \tilde{G}_m \) is the Gram matrix related to the orthonormal basis \( \{ \phi_j^{m_1} \psi_k^{m_2} \}_{(k,l) \in J_m} \) for the scalar product \( \langle \cdot, \cdot \rangle_n \). Lemma 3.1 in Baraud [2000] indicates that \( \rho(\tilde{G}_m) = \sup_{t \in S_n} (\|t\|_n^2 / \|t\|_{f(X,T)}^2) \). Then on \( \Omega \), \( \rho(\tilde{G}_m) \geq 1/2 > 0 \) hence \( \tilde{G}_m \) is invertible for every \( m \).

Moreover, let \( F_n = \arg \min_{t \in S_n} \| F - t \|_{f(X,T)}^2 \) be the projection of \( F \) on the global model \( S_n \). Then \( (\tilde{F}_m - F_n) \in S_n \) hence
\[
\left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 = \left\| \tilde{F}_m - F_n \right\|_{f(X,T)}^2 + \left\| F_n - F \right\|_{f(X,T)}^2.
\]

Thus, by definition of \( \Omega \),
\[
\mathbb{E} \left[ \left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 \mathbb{I}_\Omega \right] \leq 2 \mathbb{E} \left[ \left\| \tilde{F}_m - F_n \right\|_{f(X,T)}^2 \mathbb{I}_\Omega \right] + \mathbb{E} \left[ \left\| F_n - F \right\|_{f(X,T)}^2 \right]
\]
\[
\leq 4 \mathbb{E} \left[ \left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 \mathbb{I}_\Omega \right] + 4 \mathbb{E} \left[ \left\| F_n - F \right\|_{f(X,T)}^2 \mathbb{I}_\Omega \right] + \mathbb{E} \left[ \left\| F_n - F \right\|_{f(X,T)}^2 \right]
\]
\[
\leq 4 \mathbb{E} \left[ \left\| \tilde{F}_m - F \right\|_{f(X,T)}^2 \right] + 5 \mathbb{E} \left[ \left\| F_n - F \right\|_{f(X,T)}^2 \right].
\]
On the one hand $\mathbb{E}\left[\left\|\tilde{F}_m - F\right\|_2^2\right]$ is upper bounded with Comment 3 after Theorem 3.1. On the other hand, for every $m \in J_n$, $S_m \subset S_n$ so
\[ \|F_n - F\|_{f(x,t)}^2 = \inf_{t \in S_n} \|F - t\|_{f(x,t)}^2 \leq \inf_{t \in S_m} \|F - t\|_{f(x,t)}^2. \]

Besides, according to Comment 1 in Section 3.1, $\mathbb{E}\left[\left\|\tilde{F}_m - F\right\|_{f(x,t)}^2 I_{\Omega}\right] \leq \mathbb{E}\left[\left\|\tilde{F}_m - F\right\|_{f(x,t)}^2 I_{\Omega}\right]$, which ends the proof of Proposition 6.1. □

6.3.2. Proof of Proposition 6.2. The proof is based on the following Lemma.

**Lemma 6.2.** Under the assumptions of Theorem 3.1,
\begin{equation}
P[\Omega^c] \leq 2(N_n)^2 \exp\left(-\frac{3 - 2\sqrt{2}}{2} \frac{nh_0}{(N_n)^2K^2}\right).
\end{equation}

Assume that Lemma 6.2 holds. On the one hand for every $(x, u) \in A$, $\tilde{F}_m(x, u)$ and $F(x, u)$ lie in $[0,1]$ hence $\|\tilde{F}_m - F\|_{f(x,t)}^2 \leq 1$. On the other hand let $c_0 = h_0(3 - 2\sqrt{2})/(2K^2)$,
\[ \mathbb{E}\left[\left\|\tilde{F}_m - F\right\|_{n, f(x,t)}^2 I_{\Omega^c}\right] \leq 2(N_n)^2 \exp\left(-c_0 \frac{n}{(N_n)^2}\right) \leq 2n \exp(-c_0 \log^2 n) \leq \frac{C_6}{n} \quad \square \]

**Proof of Lemma 6.2.** Let $\{\chi_{\lambda, \lambda} \in J_n\}$ be an $\|\|_{f(x,t)}$-orthonormal basis of the global space $S_n$ where $J_n$ is the set of index defined in (2) for $D_{m_1}^{(1)} = N_n^{(1)}$ and $D_{m_2}^{(2)} = N_n^{(2)}$. Denote
\[ S_{\lambda, \lambda} = \frac{1}{n} \sum_{i=1}^n (\chi_{\lambda}(X_i, T_i) \chi_{\lambda}(X_i, T_i) - \mathbb{E}[\chi_{\lambda}(X_i, T_i) \chi_{\lambda}(X_i, T_i)]). \]

By definition of $\Omega$,
\[ P[\Omega^c] = P\left[\sup_{\lambda \in J_n} \left|\sum_{\lambda' \in J_n} a_{\lambda} a_{\lambda'} S_{\lambda, \lambda'}\right| > \frac{1}{2}\right] \leq P\left[\sup_{\lambda \in J_n} \left(\sum_{\lambda' \in J_n} a_{\lambda}^2 \right) \left|S_{\lambda, \lambda'}\right| > \frac{1}{2}\right]. \]

Let $C$ and $V$ be the following $N_n \times N_n$-square matrix:
\[ V = \sqrt{v_{\lambda, \lambda'}}(\lambda, \lambda') \in J_n^2 \quad \text{where} \quad v_{\lambda, \lambda'} = \mathbb{E}\left[\chi_{\lambda}(X_i, T_i) \chi_{\lambda'}(X_i, T_i)\right], \]
\[ C = (c_{\lambda, \lambda'})(\lambda, \lambda') \in J_n^2 \quad \text{where} \quad c_{\lambda, \lambda'} = \|\chi_{\lambda}(X_i, T_i) \chi_{\lambda'}(X_i, T_i)\|_\infty \]
and let
\begin{equation}
x = \frac{3 - 2\sqrt{2}}{2} \min\left(\frac{1}{\rho^2(V)}, \frac{1}{\rho(C)}\right)
\end{equation}
where $\rho$ is the spectral radius defined in Section 2.1. Then $\sqrt{2}x\rho(V) + x\rho(C) \leq 1/2$ and
\[ P[\Omega^c] \leq P \left[ \sup_{\sum_{\lambda \notin J_n} a_\lambda^2 = 1} \left( \sum_{(\lambda, \lambda') \in J_n^2} |a_\lambda| |a_{\lambda'}| |S_{\lambda, \lambda'}| \right) > \sqrt{2} x \rho(V) + x \rho(C) \right]. \]

Besides, let \((a_\lambda)_{\lambda \in J_n}\) be such that \(\sum_{\lambda \in J_n} a_\lambda^2 = 1\),

\[ \sqrt{2} x \rho(V) + x \rho(C) = \sup_{\lambda \in J_n^2} \left( \sum_{(\lambda, \lambda') \in J_n^2} |b_\lambda| |b_{\lambda'}| (\sqrt{2} x v_{\lambda, \lambda'} + x c_{\lambda, \lambda'}) \right) \geq \sum_{(\lambda, \lambda') \in J_n^2} |a_\lambda| |a_{\lambda'}| (\sqrt{2} x v_{\lambda, \lambda'} + x c_{\lambda, \lambda'}) \]

Hence,

\[ P[\Omega^c] \leq P \left[ \exists (\lambda, \lambda') \in J_n^2, |S_{\lambda, \lambda'}| > (\sqrt{2} x v_{\lambda, \lambda'} + x c_{\lambda, \lambda'}) \right] \leq \sum_{(\lambda, \lambda') \in J_n^2} P \left[ |S_{\lambda, \lambda'}| > \sqrt{2} x v_{\lambda, \lambda'} + x c_{\lambda, \lambda'} \right]. \]

Finally we use Bernstein Deviation Inequality presented in Birgé and Massart [1998], Lemma 8. Then

\[ (24) \quad P[\Omega^c] \leq \sum_{(\lambda, \lambda') \in J_n^2} 2 \exp(-nx) = 2(N_n)^2 \exp(-nx). \]

Besides, \(\max (\rho^2(V), \rho(C))\) is upper bounded similarly to Baraud et al. [2001]. According to equations (2.9) and (2.10) in Baraud et al. [2001], under Assumptions (A_2) and (A_3),

\[ \sup_{i \in \mathcal{S}_n} \|t\|_\infty \leq \frac{1}{h_0} \sup_{t \in \mathcal{S}_n} \|t\|_\infty = \frac{1}{h_0} \sup_{t \in (x, y) \in A} \sum_{k \in J_n} (\varphi_k(x) \psi_k(y))^2 \leq \frac{K_1 K_2}{h_0} \dim(S_n). \]

Then, with Lemma 2 in Baraud et al. [2001], \(\max (\rho^2(V), \rho(C)) \leq (K_1 K_2 / h_0) N_n^{(1)} N_n^{(2)}\) which concludes the proof of Lemma 6.2. □

6.4. Proof of Proposition 4.1. The proof is based on the following theorem (see Tsybakov [2004], Chapter 2, Theorem 2.5). Let \(B = B_{2, \infty}^2(A, L)\). Denote by \(K(P, Q)\) the Kullback distance between the distributions \(P\) and \(Q\):

\[ K(P, Q) = \begin{cases} \int \log(dP/dQ)dP & \text{if } P << Q \\ +\infty & \text{otherwise} \end{cases} \]

Theorem 6.1. Assume that there exist \(M \geq 2\) and \(F_0, \ldots, F_M\) such that

1. \(F_j \in B\) for every \(j \in \{0, \ldots, M\}\).
2. \(\|F_j - F_l\|^2 \geq 2r\) for every \(j \neq l \in \{0, \ldots, M\}\).
3. \(P_j^{(n)} << P_0^{(n)}\) for every \(j \in \{0, \ldots, M\}\), where \(P_j^{(n)}\) denotes the distribution of \((X_i, T_i, \delta_i)_{i=1,\ldots,n}\) if \(F = F_j\), and for some \(0 < \alpha < 1/8\)

\[ \frac{1}{M} \sum_{j=1}^M K(P_j^{(n)}, P_0^{(n)}) \leq \alpha \log M. \]
Then there exists a constant $c$ such that
\[
\inf \sup_{\hat{F}_n, F \in \mathbb{B}} \mathbb{E} \left[ r \| \hat{F}_n - F \|^2 \right] \geq c.
\]

We construct a set of distribution functions $\{F_0, \ldots, F_M\}$ which satisfies conditions 1, 2 and 3. Up to rescalings and translations, we assume that $A = [0, 1] \times [0, 1]$.

6.4.1. Construction of the $(F_i)$’s. Let
\[
F_0(x, u) = \mathbb{I}_{[0,1]}(x) (a \mathbb{I}_{[0, +\infty]}(u) + au \mathbb{I}_{[0,1]}(u) + (1 - a) \mathbb{I}_{(1, +\infty]}(u))
\]
with $a = \min(1/3, L/2)$. For every $x \in [0, 1]$,
- $F_0(x, u) = 0$, $\forall u < 0$,
- $F_0(x, u) = 1$, $\forall u > 1$,
- $F_0(x,.)$ is increasing on $[0, 1]$ and $F_0(x, u) \in [a, 2a] \subset (0, 1)$ for every $u \in [0, 1]$,
thus $F_0$ is a conditional distribution. Let $\psi$ be a one-dimensional wavelet supported on $[0, 1]$. Let $J = (j_1, j_2)$ be a couple of non-negative integers determined further. For every $S = (s_1, s_2) \in \mathbb{Z}^2$, let
\[
\psi_{J,S}(x, u) = 2^{(j_1 + j_2)/2} \psi(2^{j_1} x - s_1) \psi(2^{j_2} u - s_2)
\]
There exists a subset $R_J$ of $\mathbb{Z}^2$ such that
- $\text{Supp}(\psi_{J,S}) = I_{J,S} \subset [0, 1]^2$ for every $S \in R_J$,
- The applications $\{\psi_{J,S}, S \in R_J\}$ have disjoint supports,
- $|R_J| = 2^{j_1 + j_2}$.

Let $b$ be a positive constant which will be determined later. For every $\varepsilon \in \{0, 1\}^{|R_J|}$, let
\[
G_{\varepsilon} = \sqrt{\frac{b}{n}} \sum_{S \in R_J} \varepsilon_S \psi_{J,S}
\]
and $F_{\varepsilon} = F_0 + G_{\varepsilon}$. For every $x \in [0, 1]$,
- $F_{\varepsilon}(x, u) = F_0(x, u) = 0$, $\forall u < 0$
- $F_{\varepsilon}(x, u) = F_0(x, u) = 1$, $\forall u > 1$

Moreover let $(x, u) \in [0, 1]^2$,
\[
F_{\varepsilon}(x, u) = a + \int_0^u \left( a + \sqrt{\frac{b}{n}} \sum_{S \in R_J} \varepsilon_S \frac{\partial \psi_{J,S}}{\partial y}(x, y) \right) dy.
\]
Assume that
\[
\sqrt{\frac{b}{n}} 2^{j_1/2} 2^{j_2/2} \| \psi \|_{\infty} \| \psi' \|_{\infty} \leq \frac{a}{2}
\]
Let $y \in [0, u]$ and $S_0$ be such that $(x, y) \in I_{J,S_0}$
\[
\left| \sqrt{\frac{b}{n}} \sum_{S \in R_J} \varepsilon_S \frac{\partial \psi_{J,S}}{\partial y}(x, y) \right| = \left| \sqrt{\frac{b}{n}} \varepsilon_{S_0} \frac{\partial \psi_{J,S_0}}{\partial y}(x, y) \right| < \frac{a}{2}.
\]
Therefore the term in the integral in (25) is positive and the application \( F_\varepsilon(x,.) \) is increasing on \([0,1]\). Moreover, as \( \psi_{J,S}(x,1) = 0 \) for every \( S \in R_J \), \( F_\varepsilon(x,1) = F_0(x,1) = 2a < 1 \). Thus \( F_\varepsilon \) is a conditional distribution function on \([0,1]^2\).

6.4.2. Condition which guarantees that \( F_\varepsilon \in B \) for every \( \varepsilon \). On the one hand, assume that \( \psi \) is regular enough, then according to Hochmuth [2002] (Theorem 3.5),

\[
|G_\varepsilon|_{B_{2,\infty}^\beta([0,1]^2)} \leq (2^{j_1\beta_1} + 2^{j_2\beta_2})\|G_\varepsilon\|.
\]

Moreover, as the \( \{\psi_{J,S}, S \in R_J\} \) have disjoint supports,

\[
\|G_\varepsilon\|^2 = \frac{n}{b} \left| \sum_{S \in R_J} \varepsilon_S \psi_{J,S} \right|^2 = \frac{n}{b} \sum_{S \in R_J} \varepsilon_S^2 \|\psi_{J,S}\|^2.
\]

By definition of the wavelets, \( \|\psi_{J,S}\| = \|\psi\| = 1 \), hence

\[
\|G_\varepsilon\| \leq \sqrt{\frac{n}{b}} |R_J| = \sqrt{\frac{b}{n}} 2^{(j_1+j_2)/2}.
\]

Thus

\[
\|G_\varepsilon\|_{B_{2,\infty}^\beta([0,1]^2)} = |G_\varepsilon|_{B_{2,\infty}^\beta([0,1]^2)} + \|G_\varepsilon\| \leq \sqrt{\frac{b}{n}} 2^{(j_1+j_2)/2}(2^{j_1\beta_1} + 2^{j_2\beta_2} + 1).
\]

On the other hand, \( |F_0|_{B_{2,\infty}^\beta([0,1]^2)} = 0 \). Indeed, let \( r_i = \lfloor \beta_i \rfloor + 1 \) for \( i = 1 \) and \( 2 \). Then \( r_1 \geq 1, r_2 \geq 2 \) and

\[
|F_0|_{B_{2,\infty}^\beta([0,1]^2)} = \sup_{t>0} \left[ t^{-\beta_1} \omega_{r_1,1}(F_0, t, [0,1]^2)_2 + t^{-\beta_2} \omega_{r_2,2}(F_0, t, [0,1]^2)_2 \right].
\]

Besides let \( h > 0 \) and

\[
\Omega_{h,1}^r = \{(x,u) \in [0,1]^2, (x+r_1h,u) \in [0,1]^2\}.
\]

For every \( (x,u) \in \Omega_{h,1}^r \), \( F_0(x+h,u) = F_0(x,u) \). So, as \( r_1 \geq 1, \Delta_{h,1}^{r_1} F_0(x,u) = 0 \). Hence

\[
\omega_{r_1,1}(F_0, t, [0,1]^2)_2 = \sup_{|h| \leq t} \|\Delta_{h,1}^{r_1} F_0\|_{L^2(\Omega_{h,1}^r)} = 0.
\]

Moreover on \([0,1]^2\), \( F_0(x,u) = a(1+u) \) if \( u < 1 \) and \( F_0(x,1) = 1 \). Thus, let \( \tilde{F}_0(x,u) = a(1+u) \) for every \((x,u) \in [0,1]^2\), \( F_0 \) and \( \tilde{F}_0 \) are equal on \([0,1]^2\) except on a set of measure 0, so \( \|\Delta_{h,2}^{r_2} F_0\| = \|\Delta_{h,2}^{r_2} \tilde{F}_0\| \). Besides, for all \((x,u) \in [0,1]^2\),

\[
\Delta_{h,2}^{r_2} \tilde{F}_0(x,u) = ah \quad \Rightarrow \quad \Delta_{h,2}^{r_2} \tilde{F}_0(x,u) = \Delta_{h,2}^{r_2-1} \Delta_{h,2}^{r_1} \tilde{F}_0(x,u) = 0
\]

as \( r_2 - 1 \geq 1 \). Then \( \omega_{r_2,2}(F_0, t, [0,1]^2)_2 = 0 \) and consequently \( |F_0|_{B_{2,\infty}^\beta([0,1]^2)} = 0 \). Moreover

\[
\|F_0\|_{B_{2,\infty}^\beta([0,1]^2)} = \sqrt{\int_0^1 \int_0^1 a^2(1+u)^2du dx = \sqrt{\frac{7}{3}} a}
\]

and

\[
\|F_\varepsilon\|_{B_{2,\infty}^\beta([0,1]^2)} \leq \|F_0\|_{B_{2,\infty}^\beta([0,1]^2)} + \|G_\varepsilon\|_{B_{2,\infty}^\beta([0,1]^2)} \leq \sqrt{\frac{7}{3}} a + \sqrt{\frac{b}{n}} 2^{(j_1+j_2)/2}(2^{j_1\beta_1} + 2^{j_2\beta_2} + 1).
\]
By definition \( a \leq L/2 \) so \( \|F_\varepsilon\|_{L^2([0,1]^2)} \leq L \) as soon as

\[
\sqrt{\frac{b}{n}} (j_1+j_2)/2 (2^{j_1} + 2^{j_2} + 1) \leq L \left( 1 - \sqrt{\frac{7}{12}} \right).
\]

6.4.3. Expression of \( \|F_\varepsilon - F_{\varepsilon'}\|^2 \).

\[
\|F_\varepsilon - F_{\varepsilon'}\|^2 = \frac{b}{n} \sum_{S \in \mathcal{R}_J, J,S} \int_{I_{J,S}} (\varepsilon_S - \varepsilon_S')^2 \psi_S^2 d(x,u)du = \frac{b}{n} \sum_{S \in \mathcal{R}_J} \mathbb{I}_{\{\varepsilon_S \neq \varepsilon_S'\}} = \frac{b}{n} p(\varepsilon, \varepsilon').
\]

6.4.4. Upper bound of \( K(F_\varepsilon^{(n)}, P_0^{(n)}) \). For every \( i \in \{1, \ldots, n\} \), under \( F_\varepsilon \), \( (X_i, T_i, \delta_i) \) has density

\[
p_\varepsilon(x, u, d) = \left[(F_\varepsilon(x, u))^d(1 - F_\varepsilon(x, u))^{1-d}\right] f_{(X,T)}(x, u)
\]
with respect to \( \mathcal{L} \otimes \mathcal{L} \otimes \mu \) where \( \mathcal{L} \) is the Lebesgue measure and \( \mu \) is the counting measure on \( \mathbb{N} \). Similarly, under \( F_0 \), \( (X_i, T_i, \delta_i) \) has density

\[
p_0(x, u, d) = \left[(F_0(x, u))^d(1 - F_0(x, u))^{1-d}\right] f_{(X,T)}(x, u)
\]
with respect to \( \mathcal{L} \otimes \mathcal{L} \otimes \mu \). For every \( \varepsilon \in \{0, 1\}^{\mathcal{R}_J} \), \( P_\varepsilon \) is absolutely continuous with respect to \( P_0 \). Indeed,

\[
F_0(x, u) = 0 \quad \Rightarrow \quad (x, u) \notin [0, 1] \times [0, +\infty[ \quad \Rightarrow \quad F_\varepsilon(x, u) = 0,
\]

\[
F_0(x, u) = 1 \quad \Rightarrow \quad (x, u) \in [0, 1] \times [1, +\infty[ \quad \Rightarrow \quad F_\varepsilon(x, u) = 1.
\]

Then

\[
K(P_\varepsilon, P_0) = \int_{\mathbb{R}^2} \left[ \log \left( \frac{F_\varepsilon(x, u)}{F_0(x, u)} \right) F_\varepsilon(x, u) + \log \left( \frac{1 - F_\varepsilon(x, u)}{1 - F_0(x, u)} \right) (1 - F_\varepsilon(x, u)) \right] f_{(X,T)}(x, u)dxdu
\]

Out of the intervals \( \{I_{J,S}, S \in \mathcal{R}_J\} \), \( F_\varepsilon \) and \( F_0 \) are equal. Hence

\[
K(P_\varepsilon, P_0) = \sum_{S \in \mathcal{R}_J} \int_{I_{J,S}} \left[ \log \left( 1 + \frac{\theta_S}{a(1+u)} \right) (a(1+u) + \theta_S) \right.
\]

\[
+ \log \left( 1 - \frac{\theta_S}{1 - a(1+u)} \right) (1 - a(1+u) - \theta_S) \right] f_{(X,T)}(x, u)dxdu
\]

where \( \theta_S = \varepsilon_S \sqrt{b/n} \psi_{I_{J,S}}(x, u) \). For every \( S \in \mathcal{R}_J \) and \( (x, u) \in I_{J,S} \)

\[
\frac{\theta_S}{a(1+u)} = \frac{F_\varepsilon(x, u)}{F_0(x, u)} - 1 > -1 \quad \text{and} \quad -\frac{\theta_S}{1 - a(1+u)} = \frac{1 - F_\varepsilon(x, u)}{1 - F_0(x, u)} - 1 > -1
\]

Noting that \( \log(1+v) \leq v \) for every \( v > -1 \) we obtain

\[
K(P_\varepsilon, P_0) \leq \sum_{S \in \mathcal{R}_J} \int_{I_{J,S}} \left[ \theta_S + \frac{\theta_S^2}{a(1+u)} - \theta_S + \frac{\theta_S^2}{1 - a(1+u)} \right] f_{(X,T)}(x, u)dxdu.
\]

For every \( u \in [0, 1] \),

\[
0 < \frac{1}{a(1+u)} \leq \frac{1}{a} \quad \text{and} \quad 0 < \frac{1}{1 - a(1+u)} \leq \frac{1}{1 - 2a}.
\]
Thus
\[ K(P_\varepsilon, P_0) \leq \left( \frac{1}{a} + \frac{1}{1-2a} \right) \sum_{S \in R_J} \int_{I_{j,S}} \theta_S^2 f(x,T)(x,u)dxdu \]
\[ \leq \left( \frac{1}{a} + \frac{1}{1-2a} \right) b n \|f(x,T)\|_\infty |R_J| = a'b \|f(x,T)\|_\infty \frac{2^{j_1+j_2}}{n}. \]
where \( a' = 1/a + 1/(1-2a) \). Finally,
\[ K(P_\varepsilon^{(n)}, P_0^{(n)}) \leq a'b \|f(x,T)\|_\infty 2^{j_1+j_2}. \]

6.4.5. Conclusion. According to Lemma 2.7, Chapter 2 in Tsybakov [2004], there exists a family \((\varepsilon^{(0)}, \ldots, \varepsilon^{(M)}) \subset \{0, 1\}^{|R_J|}\) with \( \varepsilon^{(0)} = (0, \ldots, 0) \) such that
\[ p(\varepsilon^{(i)}, \varepsilon^{(i')}) \geq \frac{|R_J|}{8} = \frac{2^{j_1+j_2}}{8}, \quad \forall i \neq i' \in \{0, \ldots, M\} \]
and \( \log(M) \geq (\log 2/8)2^{j_1+j_2} \) where the distance \( p \) is defined in (28).

Now parameters \( B_0, b, j_1 \) and \( j_2 \) are choosen so that the family \((F_{\varepsilon^{(0)}}, \ldots, F_{\varepsilon^{(M)}})\) satisfies the assumptions of Theorem 6.1 with
\[ r = B_0 n^{-\beta_3/(\beta+1)}. \]
Let
\[ b = \frac{\log 2}{72 \|f(x,T)\|_\infty a'}, \quad c_0 = \left[ \frac{L}{4 \sqrt{b}} \left( 1 - \sqrt{\frac{7}{12}} \right) \right]^{1/(1+\beta_1+\beta_2)} \quad \text{and} \quad B_0 = 32/bc_0^2. \]

Let \( j_1 \) and \( j_2 \) be in \( \mathbb{N}^* \) such that
\[ (c_0/2)n^{\beta_2/(\beta_1+\beta_2+2\beta_1\beta_2)} \leq 2^{j_1} \leq c_0 n^{\beta_2/(\beta_1+\beta_2+2\beta_1\beta_2)} \]
\[ (c_0/2)n^{\beta_1/(\beta_1+\beta_2+2\beta_1\beta_2)} \leq 2^{j_2} \leq c_0 n^{\beta_1/(\beta_1+\beta_2+2\beta_1\beta_2)}. \]
The existence of \( j_1 \) and \( j_2 \) is guaranteed for \( n \) larger than an integer \( n_0 \) depending on \((c_0, \beta)\). Then for every \( i, i' \in \{0, \ldots, M\} \)
\[ \|F_{\varepsilon^{(i)}} - F_{\varepsilon^{(i')}}\|^2 \geq \frac{b 2^{j_1+j_2}}{n} \geq \frac{bc_0^2}{32n} n^{(\beta_1+\beta_2)/(\beta_1+\beta_2+2\beta_1\beta_2)} = B_0 n^{-2\beta_1\beta_2/(\beta_1+\beta_2+2\beta_1\beta_2)} = B_0 n^{-\beta_3/(\beta+1)} \]
hence condition (2) in Theorem 6.1 is satisfied.

Moreover
\[ \frac{1}{M} \sum_{i=0}^{M} K(P^{(n)}_{\varepsilon^{(i)}}, P_0^{(n)}) \leq a' \|f(x,T)\|_\infty b 2^{j_1+j_2} = \frac{\log 2}{72} 2^{j_1+j_2} \leq \frac{\log M}{9} \]
hence condition (3) in Theorem 6.1 is satisfied with \( \alpha = 1/9 \).

Finally condition (1) in Theorem 6.1 is satisfied as soon as (26) and (27) hold. Besides, \( \beta_1 > 0 \) and \( \beta_2 > 1 \) and by (29) \( j_1 \) and \( j_2 \) are increasing with \( n \). Therefore \( 2^{(j_1+j_2)/2} (2^{j_1} \beta_1 + \cdots) \)
CDF ESTIMATION FROM CURRENT STATUS DATA

$2^j \beta_2 + 1$ increases faster than $2^{j/2} \beta_2^{j/2}$ and for $n$ larger than an integer $n_1$ depending on $\psi$ and $L$, (26) holds as soon as (27) holds. Moreover (27) holds as soon as

$$\sqrt{b_c} n^{-\beta_1 \beta_2 / (\beta_1 + \beta_2 + 2 \beta_1 \beta_2)} \left( (c_0^{\beta_1} + c_0^{\beta_2}) n^{-\beta_1 / (\beta_1 + \beta_2 + 2 \beta_1 \beta_2)} + 1 \right) \leq L \left( 1 - \sqrt{\frac{7}{12}} \right)$$

which is ensured if

(30)

$$\sqrt{bc} \left( c_0^{\beta_1} + c_0^{\beta_2} \right) \leq L \left( 1 - \sqrt{\frac{7}{12}} \right)$$

and

(31)

$$\sqrt{b_c} n^{-\beta_1 \beta_2 / (\beta_1 + \beta_2 + 2 \beta_1 \beta_2)} \leq L \left( 1 - \sqrt{\frac{7}{12}} \right).$$

On the one hand (30) holds as soon as

$$2 c_0^{\beta_1 + \beta_2 + 1} \leq \frac{L}{2 \sqrt{b}} \left( 1 - \sqrt{\frac{7}{12}} \right)$$

which is guaranteed by definition of $c_0$. On the other hand there exists an integer $n_2$ depending on $(\beta, c_0)$ such that (31) is satisfied for every $n \geq n_2$.

Thus for every $n \geq \max(n_0, n_1, n_2)$, conditions 1, 2 and 3 in Theorem 6.1 hold with $r = B_0 n^{-\beta / (\beta + 1)}$, which concludes the proof of Proposition 4.1. □

7. Appendix

7.1. Talagrand Inequality. We use the following version of Talagrand Inequality.

**Theorem 7.1.** Let $(V_1, \ldots, V_n)$ be independent random variables, and $\mathcal{F}$ be a countable set of applications from $\mathbb{R}$ to $\mathbb{R}^n$ such that $-\mathcal{F} = \mathcal{F}$. Let

$$Z = \sup_{f \in \mathcal{F}} \left( \frac{1}{n} \sum_{i=1}^{n} (f^{(i)}(V_i) - \mathbb{E}[f^{(i)}(V_i)]) \right)$$

and $b$, $v$ and $\mathbb{H}$ be such that

$$\sup_{f \in \mathcal{F}} \left( \sup_{i=1, \ldots, n} \|f^{(i)}\|_\infty \right) \leq b, \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \text{Var}(f^{(i)}(V_i)) \leq v \quad \text{and} \quad \mathbb{E}Z \leq \mathbb{H}.$$

Then, for every $\theta > 1$, there exists $\overline{C}$, $\overline{C}'$, $\overline{K}$, $\overline{K}'$ such that for every $n$,

$$\mathbb{E} \left[ (Z^2 - \theta \mathbb{H}^2)^+ \right] \leq \frac{\overline{C} v}{n} \exp \left( -\frac{\pi n \mathbb{H}^2}{v} \right) + \frac{\overline{C}' b^2}{n^2} \exp \left( -\frac{\pi n \mathbb{H}}{b} \right).$$

Theorem 7.1 is derived from Theorem 1.1 in Klein and Río [2005] by setting $s^{(i)}(t) = \frac{1}{t} (f^{(i)}(t) - \mathbb{E}[f^{(i)}(V_i)])$. Similarly to Birgé and Massart [1998], Corollary 2 we get

$$P[Z \geq (1 + \nu)\mathbb{H} + x] \leq \exp \left( -\frac{n}{3} \min \left( \frac{y^2}{2v}, \frac{2 \min(1, \nu) y}{7b} \right) \right).$$
Then we use the formula \(E[X_+] \leq \int_0^{+\infty} P[X \geq s] ds\) to get
\[
E \left[ (Z^2 - \theta H^2)_+ \right] \leq \int_0^{+\infty} P[Z \geq \sqrt{\theta H^2 + s}] ds \geq \int_0^{+\infty} P[Z \geq \alpha_1 H + \sqrt{\alpha_2 H + \alpha_3 s}] ds
\]
for some positive \((\alpha_1, \alpha_2, \alpha_3)\), and a simple integration provides the result of Theorem 7.1.

7.2. Linear algebra lemma.

**Lemma 7.1.** Let \(V = \text{Vect}(\xi_1, \ldots, \xi_D)\) be a linear subspace of a vector space \(E\). Let \(\langle s, t \rangle_0\) be a scalar product on \(E\), and \(\|t\|_0 = \sqrt{\langle t, t \rangle_0}\) the corresponding semi-norm. There exists a basis \((\varphi_1, \ldots, \varphi_D)\) of \(V\) which is orthogonal for the \(\|\cdot\|_0\)-norm.

**Proof of Lemma 7.1.** The proof follows exactly the Gram Schmidt orthogonalisation procedure, but with a possibly linearly dependent family.

- Let \(\varphi_1 = \xi_1\).
- For every \(k \in \{1, \ldots, D - 1\}\), we set \(\varphi_{k+1} = \xi_{k+1} + \sum_{j=1}^{k} a_j \varphi_j\) where

\[
a_j = \begin{cases} 
0 & \text{if } \|\varphi_j\|_0 = 0 \\
-\langle \xi_{k+1}, \varphi_j \rangle_0/\|\varphi_j\|_0^2 & \text{otherwise.}
\end{cases}
\]

Thus, for every \(k \in \{1, \ldots, D\}\), \(\text{Vect}(\xi_1, \ldots \xi_k) = \text{Vect}(\varphi_1, \ldots, \varphi_k)\) and the \((\varphi_j)\)’s are orthogonal for the \(\|\cdot\|_0\) semi-norm. \(\square\)

**Acknowledgements**

I am grateful to Fabienne Comte for her advices about this work and to Cristina Butucea for her help on the minimax study.

**References**

Y. Baraud. Model selection for regression on a fixed design. *Probab. Theory Related Fields*, 117(6):467–493, 2000.

Y. Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6:127–146, 2002.

Y. Baraud, F. Comte, and G. Viennot. Adaptive estimation in autoregression or \(\beta\)-mixing regression via model selection. *Ann. Statist.*, 29(3):839–875, 2001.

L. Birgé. Interval censoring: a nonasymptotic point of view. *Math. Methods Statist.*, 8(3):285–298, 1999.

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

E. Brunel and F. Comte. Cumulative distribution function estimation under interval censoring case 1. *Electron. J. Stat.*, 3:1–24, 2009.

P. Groeneboom and J.A. Wellner. *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1992. ISBN 3-7643-2794-4.

R. Hochmuth. Wavelet characterization for anisotropic Besov spaces. *Appl. Comput. Harmon. Anal.*, 12(2):179–208, 2002.
M.G. Hudgens, M.H. Maathuis, and P.B. Gilbert. Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable. *Biometrics*, 63(2):372–380, 2007.

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

C. Lacour. Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(5):571–597, 2007.

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

S. Ma and M.R. Kosorok. Adaptive penalized M-estimation with current status data. *Ann. Inst. Statist. Math.*, 58(3):511–526, 2006.

P. Massart. *Concentration inequalities and model selection*, Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23. Lecture Notes in Mathematics. Springer, Berlin, 2007.

P. Massart. *Selection de modele : de la theorie a la pratique*. *J. Soc. Fr. Stat. and Rev. Stat. Appl.*, 149(4):5–27, 2008.

S. M. Nikol’skii. Approximation of functions of several variables and imbedding theorems. *Jr. Die Grundlehren der Mathematischen Wissenschaften*, 1975.

A. B. Tsybakov. *Introduction à l’estimation non paramétrique*. Springer-Verlag, Berlin, 2004.

S. van de Geer. Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.*, 21(1):14–44, 1993.

A. van der Vaart and M.J. van der Laan. Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.*, 2:Art. 9, 42, 2006.