Cumulative distribution function estimation under interval censoring case 1.

Elodie Brunel, Fabienne Comte

CUMULATIVE DISTRIBUTION FUNCTION ESTIMATION UNDER INTERVAL CENSORING CASE 1.

E. BRUNEL[1] AND F. COMTE[2]

Abstract. We consider projection methods for the estimation of cumulative distribution function under interval censoring, case 1. Such censored data also known as current status data, arise when the only information available on the variable of interest is whether it is greater or less than an observed random time. Two types of adaptive estimators are investigated. The first one is a two-step estimator built as a quotient estimator. The second estimator results from a mean square regression contrast. Both estimators are proved to achieve automatically the standard optimal rate associated with the unknown regularity of the function, but with some restriction for the quotient estimator. Simulation experiments are presented to illustrate and compare the methods.

March 2008

AMS Classification (2001): 62G05; 62G20.

Keywords and phrases. Adaptive estimation. Current status data. Minimax rate. Interval censoring. Nonparametric estimator. Penalized contrast.

1. Introduction

Let X be a survival time with unknown cumulative distribution function (cdf) F. In the interval censoring case 1 model, we are not able to observe the survival time X. Instead, an observation consists of the pair (U, δ) where U is an examination time and δ is the indicator function of the event $(X \leq U)$. Roughly speaking, the only knowledge about the variable of interest X is wether it has occurred before U or not. Early examples of such interval censoring can be found in demography studies, see Diamond and McDonald (1991). In epidemiology, this censoring schemes also arise for instance in AIDS studies or more generally in the study of infectious diseases when the infection time is an unobservable event. We assume that U is independent of X, that F has density f and that the cdf G of U has density g. Such data also known as current status data may remind us right-censored data where the observed data is the pair $(\min(X, C), \mathbb{I}_{(X \leq C)})$ where C is a censoring variable. But, the estimation procedure in these two censoring models is substantially different. In the right-censoring model, the Kaplan and Meier (1958) estimator is well studied and is asymptotically normal at the rate $\sqrt{n}$. Nevertheless, current status data have been studied by many authors in the last two decades see Jewell and van der Laan (2004) for a state of the art. In the interval censoring model, the nonparametric maximum likelihood estimator (NPMLE) of the survival function is proved to be uniformly consistent, pointwise convergent to a nonnormal asymptotic distribution

[1] Institut de Mathématiques et de Modélisation de Montpellier, UMR CNRS 5149. Montpellier 2 University, Cc 051, Place E. Bataillon, 34095 Montpellier, FRANCE. email: ebrunel@math.univ-montp2.fr.

[2] MAP 5, UMR CNRS 8145. Paris Descartes University, 45 rue des Saints Peres, 75 270 Paris cedex 06, FRANCE. email: fabienne.comte@univ-paris5.fr.

at the rate $n^{1/3}$ in Groeneboom and Wellner (1992). In van de Geer (1993), it is also established that the NPMLE converges at rate $n^{-1/3}$ in $\mathbb{L}^2$-norm. The locally linear smoother proposed by Yang (2000), contrary to the NPMLE may be non monotone, but it has a better convergence rate than the NPMLE when the density f is smooth and the kernel function and bandwidth properly chosen. In the same spirit, Ma and Kosorok (2006) introduced an adaptive modified penalized least square estimator built with smoothing splines and proved that it achieved asymptotic optimal nonparametric rate if the function F belongs to some Sobolev space.

Here, we present two different penalized minimum contrast estimators built on trigonometric, polynomial or wavelet spaces whose associated penalty terms are really simple; the minimization of the penalized contrast function allows to choose a space that leads to both a non asymptotic automatic squared bias/variance compromise and to an asymptotic optimal convergence rate according to the regularity of the function F in term of Besov spaces. An interesting feature of the procedure is that the estimators and their study are quite simple and clean thanks to the most powerful Talagrand (1996) inequality for empirical centered processes. We also use technical properties proved in a regression framework by Baraud *et al.* (2001) and Baraud (2002) for the mean-square estimator.

The plan of the paper is as follows. Section 2 introduces the quotient and the regression estimators, after the description of the lifetimes model. We also give a detailed description of the projection spaces with their main properties. Then, we study one projection estimator of the density of the failure times which have occurred before the examination time in Section 3. Both convergence and adaptation results are given. This estimator is then applied to the estimation of the cumulative distribution function *via* a quotient construction. Section 4 describes a direct adaptive procedure to estimate the distribution function based on a mean square regression contrast. Simulations compare both approaches in Section 5. Lastly, most proofs and technical lemmas are deferred to Section 6.

2. Definition of the estimators

2.1. **Model and assumptions.** Let $(U_1, \delta_1), \cdots (U_n, \delta_n)$ be a sample of the pair (U, δ) where $\delta_i = \mathbb{I}_{(X_i \leq U_i)}$ and the pair (U_i, δ_i) is independent of X_i for all $i = 1, \cdots, n$. We are interested in the estimation of the distribution function F of the lifetime X on a compact set A only. We take $A = [0, 1]$ without loss of generality. Remember that we denote by f and F the density and the cumulative distribution function of the unobserved lifetime X and g and G those of the examination time U. A function of interest is the density ψ of the U_i restricted to the individuals for which $\delta_i = 1$ defined by:

$$\psi(x) = F(x)\, g(x) \tag{2.1}$$

It is clear that this equation provides a way to build an estimator of F. This approach is developed in Section 3. The censoring mechanism is such that the conditional law of $\delta = \mathbb{I}_{(X \leq U)}$ given $U = u$ is a Bernoulli law with parameter $F(u)$ and as a consequence we have:

$$\mathbb{E}(\delta | U = u) = F(u) \tag{2.2}$$

This relation will lead to define a direct mean-square estimator of F.

Both strategies require the following assumption:

[A1] The density g of the random time U is lower and upper bounded on A so that there exist real constants $g_0 > 0$ and $g_1 > 0$ such that for all $x \in A$, $g_0 \leq g(x) \leq g_1$.

2.2. **Definition of the estimators.** Assume that we have at our disposal a collection of finite dimensional spaces of functions, denoted by $(S_m)_{m\in\mathcal{M}_n}$, satisfying the following assumption:

$(\mathcal{H}_1)$ $(S_m)_{m\in\mathcal{M}_n}$ is a collection of finite-dimensional linear sub-spaces of $\mathbb{L}^2([0,1])$, with dimension $\dim(S_m) = D_m$ such that $D_m \leq n$, $\forall m \in \mathcal{M}_n$ and satisfying:

$$\exists \Phi_0 > 0, \forall m \in \mathcal{M}_n, \forall t \in S_m, \|t\|_\infty \leq \Phi_0 \sqrt{D_m}\|t\|. \tag{2.3}$$

where $\|t\|^2 = \int_0^1 t^2(x)dx$, for t in $\mathbb{L}^2([0,1])$.

2.2.1. *Quotient estimator.* As already mentioned, the first strategy requires to estimate ψ and g. We take as estimator of g, $\tilde{g}$ an adaptive density estimator defined in Massart (2007), Chapter 7, namely: $\tilde{g} = \hat{g}_{\hat{m}_g}$ where $\hat{g}_m = \arg\min_{t\in S_m} \gamma_n^g(t)$,

$$\gamma_n^g(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n t(U_i),$$

and

$$\hat{m}_g = \arg\min_{m\in\mathcal{M}_n} \gamma_n^g(\hat{g}_m) + \mathrm{pen}_g(m). \tag{2.4}$$

with $\mathrm{pen}_g(m) = \kappa\Phi_0^2 D_m/n$.

For the estimation of ψ, we consider the following contrast function

$$\gamma_n^\psi(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n \delta_i t(U_i). \tag{2.5}$$

Let then

$$\hat{\psi}_m = \arg\min_{t\in S_m} \gamma_n^\psi(t). \tag{2.6}$$

Then we define $\tilde{\psi} = \hat{\psi}_{\hat{m}}$ where

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} [\gamma_n^\psi(\hat{\psi}_m) + \mathrm{pen}^\psi(m)].$$

The penalty function will be motivated and defined later. The contrasts γ_n^g and γ_n^ψ are both found as empirical versions of the $\mathbb{L}^2$ distance between a function t in S_m and the function of interest (g or ψ). To see this, take the expectation of e.g. γ_n^ψ:

$$\mathbb{E}(\gamma_n^\psi(t)) = \|t\|^2 - 2\langle t, \psi\rangle = \|t-\psi\|^2 - \|\psi\|^2$$

with $\langle t, s\rangle = \int t(x)s(x)dx$. This illustrates that minimizing γ_n^ψ is likely to provide a function t that minimizes in mean $\|t-\psi\|^2$ and thus estimate ψ, on the space S_m.

Now from the adaptive estimators $\tilde{\psi}$ of ψ and $\tilde{g}$ of g, and by using the definition 2.1, we can build a quotient estimator of the distribution function F by setting

$$\tilde{F}(x) = \begin{cases} 0 & \text{if } \tilde{\psi}(x)/\tilde{g}(x) < 0 \\ \dfrac{\tilde{\psi}(x)}{\tilde{g}(x)} & \text{if } 0 \leq \tilde{\psi}(x)/\tilde{g}(x) \leq 1 \\ 1 & \text{if } \tilde{\psi}(x)/\tilde{g}(x) > 1 \end{cases} \tag{2.7}$$

2.2.2. *Regression estimator.* On the other hand, a direct estimator of the cdf F can be obtained by considering the following mean-square contrast:

$$\gamma_n^{\mathrm{MS}}(t) = \frac{1}{n}\sum_{i=1}^{n}[\delta_i - t(U_i)]^2 \tag{2.8}$$

In this case, we set

$$\hat{F}_m = \arg\min_{t\in S_m} \gamma_n^{MS}(t) \tag{2.9}$$

in the sense that we always can compute a vector $(\hat{F}_m(U_1),\dots,\hat{F}_m(U_n))$ as the orthogonal projection of the vector $(\delta_1,\dots,\delta_n)$ on the sub-space of $\mathbb{R}^n$ defined by $\{(t(U_1),\dots,t(U_n)), t\in S_m\}$. Then we define $\hat{F}_{\hat{m}_0}$ by:

$$\hat{m}_0 = \arg\min_{m\in\mathcal{M}_n}\{\gamma_n^{MS}(\hat{F}_m) + \mathrm{pen}^{MS}(m)\}, \tag{2.10}$$

with

$$\mathrm{pen}^{MS}(m) = \kappa_0\frac{D_m}{n}. \tag{2.11}$$

where κ_0 is a numerical constant.

Before studying both estimators, let us give some examples of collections $(S_m)_{m\in\mathcal{M}_n}$.

2.3. **Spaces of approximation.** The main assumption is described by $(\mathcal{H}_1)$. In this setting, an orthonormal basis of S_m is denoted by $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ where $|\Lambda_m| = D_m$. Let us mention that it follows from Birgé and Massart (1997) that Property (2.3) in the context of $(\mathcal{H}_1)$ is equivalent to

$$\exists\Phi_0>0, \|\sum_{\lambda\in\Lambda_m}\varphi_\lambda^2\|_\infty \le \Phi_0^2 D_m. \tag{2.12}$$

Moreover, for some results we need the following additional assumption:

- $(\mathcal{H}_2)$ $(S_m)_{m\in\mathcal{M}_n}$ is a collection of nested models, we denote by $\mathcal{S}_n$ the space belonging to the collection, such that $\forall m\in\mathcal{M}_n, S_m\subset\mathcal{S}_n$. We denote by N_n the dimension of this nesting space: $\dim(\mathcal{S}_n) = N_n$ $(\forall m\in\mathcal{M}_n, D_m\le N_n)$.

We consider more precisely the following examples:

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

[W] *Dyadic wavelet generated spaces* with regularity r and compact support, as described e.g. in Donoho and Johnstone (1998).

All those spaces satisfy $(\mathcal{H}_1)$, with for instance $\Phi_0 = \sqrt{2}$ for collection [T] and $\Phi_0 = \sqrt{2r+1}$ for collection [P]. Moreover, [T], [DP] and [W] satisfy $(\mathcal{H}_2)$ since they are nested with $\mathcal{S}_n$ being the space with the largest dimension in the collection.

3. Study of the quotient estimator

Our aim is to estimate the cdf F from the observations (δ_i, U_i), $i = 1, \cdots, n$.

3.1. **Convergence results for one estimator.** An explicit expression of the estimator follows from definition (2.5)-(2.6) by using the orthonormal basis $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ of (S_m) described in $(\mathcal{H}_1)$:

$$\hat{\psi}_m = \sum_{\lambda\in\Lambda_m} \hat{a}_\lambda \varphi_\lambda \text{ with } \hat{a}_\lambda = \frac{1}{n}\sum_{i=1}^{n} \delta_i \varphi_\lambda(U_i). \tag{3.1}$$

We define also ψ_m as the orthogonal projection of ψ on S_m. We can write

$$\psi_m = \sum_{\lambda\in\Lambda_m} a_\lambda \varphi_\lambda \text{ with } a_\lambda = \int_0^1 \varphi_\lambda(x)\psi(x)dx. \tag{3.2}$$

The rate of the estimator $\hat{\psi}_m$ of ψ is quite easy to derive. Indeed, it follows from (3.1), (3.2) and Pythagoras theorem that

$$\begin{aligned}
\|\psi - \hat{\psi}_m\|^2 &= \|\psi - \psi_m\|^2 + \|\psi_m - \hat{\psi}_m\|^2 = \|\psi - \psi_m\|^2 + \sum_{\lambda\in\Lambda_m} (a_\lambda - \hat{a}_\lambda)^2 \\
&= \|\psi - \psi_m\|^2 + \sum_{\lambda\in\Lambda_m} \left(\frac{1}{n}\sum_{i=1}^{n} \delta_i \varphi_\lambda(U_i) - \int_0^1 \psi(x)\varphi_\lambda(x)dx \right)^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{E}(\|\psi - \hat{\psi}_m\|^2) &= \|\psi - \psi_m\|^2 + \sum_{\lambda\in\Lambda_m} \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_i \varphi_\lambda(U_i) \right) \\
&= \|\psi - \psi_m\|^2 + \frac{1}{n}\sum_{\lambda\in\Lambda_m} \mathrm{Var}\left(\delta_1 \varphi_\lambda(U_1)\right) \\
&\leq \|\psi - \psi_m\|^2 + \frac{1}{n}\mathbb{E}\left[\left(\sum_{\lambda\in\Lambda_m} \varphi_\lambda^2(U_1) \right) \delta_1 \mathbb{I}_{(U_1\leq 1)} \right] \\
&\leq \|\psi - \psi_m\|^2 + \frac{\Phi_0^2 D_m}{n}\mathbb{E}(\delta_1 \mathbb{I}_{(U_1\leq 1)}).
\end{aligned}$$

This can be summarized by the following Proposition:

Proposition 3.1. *Consider the model described in Section 2.1 and the estimator* $\hat{\psi}_m = \arg\min_{t\in S_m} \gamma_n^\psi(t)$ *where* $\gamma_n^\psi(t)$ *is defined by (2.5) and* S_m *is a* D_m*-dimensional linear space*

in a collection satisfying $(\mathcal{H}_1)$*. Then*

$$\mathbb{E}(\|\psi - \hat{\psi}_m\|^2) \leq \|\psi - \psi_m\|^2 + \frac{\Phi_0^2 D_m}{n}\mathbb{E}(\delta_1 \mathbb{I}_{(U_1 \leq 1)}). \tag{3.3}$$

Inequality (3.3) gives the asymptotic rate for one estimator if we consider that ψ belongs to a Besov space $\mathcal{B}_{\alpha_\psi,p,\infty}([0,1])$ with finite Besov norm denoted by $|\psi|_{\alpha_\psi,p}$. For a precise definition of those notions we refer to DeVore and Lorentz (1993) Chapter 2, Section 7, where it is also proved that $\mathcal{B}_{\alpha_\psi,p,\infty}([0,1]) \subset \mathcal{B}_{\alpha_\psi,2,\infty}([0,1])$ for $p \geq 2$. This justifies that we now restrict our attention to $\mathcal{B}_{\alpha_\psi,2,\infty}([0,1])$.
Then the following (standard) rate is obtained:

Corollary 3.1. *Consider the model described in Section 2.1 and the estimator* $\hat{\psi}_m = \arg\min_{t \in S_m} \gamma_n^\psi(t)$ *where* $\gamma_n^\psi(t)$ *is defined by (2.5) and* S_m *is a* D_m*-dimensional linear space in collection* [T], [P], *or* [W]. *Assume moreover that* ψ *belongs to* $\mathcal{B}_{\alpha_\psi,2,\infty}([0,1])$ *with* $r > \alpha_\psi > 0$ *and choose a model with* $m = m_n$ *such that* $D_{m_n} = O(n^{1/(2\alpha_\psi+1)})$*, then*

$$\mathbb{E}(\|\psi - \hat{\psi}_{m_n}\|^2) = O\left(n^{-\frac{2\alpha_\psi}{2\alpha_\psi+1}}\right). \tag{3.4}$$

Remark 3.1. The bound r stands for the regularity of the basis functions for collections [P] and [W]. For the trigonometric collection [T], no upper bound for the unknown regularity α_ψ is required.

Proof. The result is a straightforward consequence of the results of DeVore and Lorentz (1993) and of Lemma 12 of Barron *et al.* (1999), which imply that $\|\psi - \psi_m\|$ is of order $D_m^{-\alpha_\psi}$ in the three collections [T], [P] and [W], for any positive α_ψ. Thus the minimum order in (3.3) is reached for a model S_{m_n} with $D_{m_n} = O([n^{1/(1+2\alpha_\psi)}])$, which is less than n for $\alpha_\psi > 0$. Then, if $\psi \in \mathcal{B}_{\alpha_\psi,2,\infty}([0,1])$ for some $\alpha_\psi > 0$, we find the standard nonparametric rate of convergence $n^{-2\alpha_\psi/(1+2\alpha_\psi)}$. □

3.2. **Adaptive estimator of the density** ψ**.** The penalized estimator is defined in order to ensure an automatic choice of the dimension. Indeed, it follows from Corollary 3.1 that the optimal dimension depends on the unknown regularity α_ψ of the function to be estimated in the asymptotic setting and more generally on the unknown constants involved in the squared-bias/variance terms. Then we define

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n^\psi(\hat{\psi}_m) + \text{pen}^\psi(m)]$$

where the penalty function pen^ψ is determined in order to lead to the choice of a "good" model. First, we apply some Talagrand (1996) type inequality to the linear empirical process defined by

$$\nu_n(t) := \frac{1}{n}\sum_{i=1}^n (\delta_i t(U_i) - \langle t, \psi \rangle). \tag{3.5}$$

Then, by using the decomposition of the contrast given by

$$\gamma_n^\psi(t) - \gamma_n^\psi(s) = \|t - \psi\|^2 - \|s - \psi\|^2 - 2\nu_n(t - s), \tag{3.6}$$

we easily derive the following result:

Theorem 3.1. *Consider the model described in Section 2.1 and the estimator* $\hat{\psi}_m = \arg\min_{t\in S_m} \gamma_n^{\psi}(t)$ *where* $\gamma_n^{\psi}(t)$ *is defined by (2.5) and* S_m *is a* D_m*-dimensional linear space in a collection satisfying* $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$*. Then the estimator* $\hat{\psi}_{\hat{m}}$ *with* $\hat{m}$ *defined by*

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} [\gamma_n^{\psi}(\hat{\psi}_m) + \mathrm{pen}^{\psi}(m)]$$

and

$$\mathrm{pen}^{\psi}(m) = \kappa\Phi_0^2 \left(\int_0^1 \psi(x)dx\right) \frac{D_m}{n}$$

where κ *is a universal constant, satisfies*

$$\mathbb{E}(\|\hat{\psi}_{\hat{m}} - \psi\|^2) \leq \inf_{m\in\mathcal{M}_n} \left(3\|\psi - \psi_m\|^2 + 4\mathrm{pen}^{\psi}(m)\right) + \frac{C}{n}, \tag{3.7}$$

where C *is a constant depending on* Φ_0 *and on* $\int_0^1 \psi(x)dx$*.*

Therefore, the adaptive estimator automatically makes the squared-bias/variance compromise and from an asymptotic point of view, reaches the optimal rate, provided that the constant in the penalty is known. Note that Inequality (3.7) is nevertheless non-asymptotic.

Remark 3.2. In practice, the constant in the penalty, denoted above by κ, is found by simulation experiments taking into account very different types of functions ψ. See some examples of such a work in Birgé and Rozenholc (2006) or Comte and Rozenholc (2004).

The penalty given in Theorem 3.1 cannot be used in practice since it depends on the unknown quantity

$$\int_0^1 \psi(x)dx = \mathbb{E}(\delta_1 \mathbb{I}_{(U_1\leq 1)}).$$

A simple solution is to use that $\int_0^1 \psi(x)dx \leq 1$; it follows that Inequality (3.7) would hold for a penalty defined by $\mathrm{pen}^{\psi}(m) = \kappa\Phi_0^2 D_m/n$. This possibly works with a resulting overestimation of the penalty, in a way depending on the unknown function ψ. The alternative solution is to replace the unknown quantity by an estimator (rather than a bound), and to prove that the estimator of ψ built with this random penalty keeps the adaptation property of the theoretical penalized estimator. This is described in the following theorem whose proof is omitted since it is quite the same as the proof of Theorem 3.4 in Brunel and Comte (2005).

Theorem 3.2. *Assume that the assumptions of Theorem 3.1 are satisfied. Consider the estimator* $\hat{\psi}_{\hat{m}}$ *with* $\hat{m}$ *defined by*

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} [\gamma_n^{\psi}(\hat{\psi}_m) + \widehat{\mathrm{pen}}^{\psi}(m)]$$

and

$$\widehat{\mathrm{pen}}^{\psi}(m) = \kappa\Phi_0^2 \left(\frac{1}{n}\sum_{i=1}^n \delta_i\right) \frac{D_m}{n}$$

where κ *is a universal constant, then* $\hat{\psi}_{\hat{m}}$ *satisfies*

$$\mathbb{E}(\|\hat{\psi}_{\hat{m}} - \psi\|^2) \leq \inf_{m\in\mathcal{M}_n} K_0 \left[\|\psi - \psi_m\|^2 + \Phi_0^2 \left(\int_0^1 \psi(x)dx\right) \frac{D_m}{n}\right] + \frac{K}{n}, \tag{3.8}$$

where K_0 is a universal constant and K depends on ψ, Φ_0.

In particular, we can derive quite straightforwardly from results as Theorem 3.2 some adaptation results to unknown smoothness:

Proposition 3.2. *Consider the collection of models* [T], [DP] *or* [W], *with $r > \alpha_\psi > 0$. Assume that an estimator $\tilde{\psi}$ of ψ satisfies inequality (3.8) in Theorem 3.2 (respectively inequality (3.7) in Theorem 3.1). Let $L > 0$. Then*

$$\left(\sup_{\psi \in \mathbb{B}_{\alpha_\psi,2,\infty}(L)} \mathbb{E}\|\psi - \tilde{\psi}\|^2 \right)^{\frac{1}{2}} \leq C(\alpha_\psi, L) n^{-\frac{\alpha_\psi}{2\alpha_\psi+1}} \tag{3.9}$$

where $\mathbb{B}_{\alpha_\psi,2,\infty}(L) = \{t \in \mathcal{B}_{\alpha_\psi,2,\infty}([0,1]), |t|_{\alpha_\psi,2} \leq L\}$ where $C(\alpha_\psi, L)$ is a constant depending on α_ψ, L and also on ψ, Φ_0.

3.3. **Application to the estimation of the distribution function F.** Consider now the first estimator of F, given by (2.7).

A simple case study allows to see that if $\tilde{\psi}(x)/\tilde{g}(x) < 0$ or $\tilde{\psi}(x)/\tilde{g}(x) > 1$, then $|\tilde{F}(x) - F(x)| \leq |\tilde{\psi}(x)/\tilde{g}(x) - F(x)|$, and thus the inequality $|\tilde{F}(x) - F(x)| \leq |\tilde{\psi}(x)/\tilde{g}(x) - F(x)|$ holds for any x. Also, our definition implies that $|\tilde{F}(x) - F(x)| \leq 1$, for any x. Moreover, to exploit [A1], we define

$$\Omega_g = \{\omega : \tilde{g}(x) - g(x) > -g_0/2, \forall x \in [0,1]\}.$$

Then, the following bounds are obtained:

$$\begin{aligned} \|\tilde{F} - F\|^2 = \int_0^1 (\tilde{F}(x) - F(x))^2 dx &= \int_0^1 (\tilde{F}(x) - F(x))^2 dx \mathbb{I}_{\Omega_g} + \int_0^1 (\tilde{F}(x) - F(x))^2 dx \mathbb{I}_{\Omega_g^c} \\ &\leq \int_0^1 (\frac{\tilde{\psi}(x)}{\tilde{g}(x)} - \frac{\psi(x)}{g(x)})^2 dx \mathbb{I}_{\Omega_g} + \int_0^1 dx \mathbb{I}_{\Omega_g^c}. \end{aligned}$$

Thus the first term can de decomposed as follows

$$\tilde{F} - F = \frac{\tilde{\psi} - \psi}{\tilde{g}} + F\left(\frac{g - \tilde{g}}{\tilde{g}}\right)$$

and thus, since $\tilde{g}(x) \geq g_0/2$ on Ω_g,

$$\int_0^1 (\frac{\tilde{\psi}(x)}{\tilde{g}(x)} - \frac{\psi(x)}{g(x)})^2 dx \mathbb{I}_{\Omega_g} \leq (\frac{2}{g_0})^2 \left(\|\hat{\psi}_{\hat{m}} - \psi\|^2 + \|\tilde{g} - g\|^2 \right).$$

For the second, taking the expectation, we use the following Lemma:

Lemma 3.1. *Assume that $g \in \mathcal{B}_{\alpha_g,2,\infty}([0,1])$ for some $\alpha_g > 0$ and consider a collection of spaces S_m such that $\log(n) \leq D_m \leq \sqrt{n}$. Then, under Assumptions* [$\mathbf{A_1}$] *and* ($\mathcal{H}_2$), *there exists a constant C such that*

$$\mathbb{P}(\Omega_g^c) \leq \mathbb{P}\left(\|\tilde{g} - g\|_\infty > g_0/2\right) \leq \frac{C}{n}. \tag{3.10}$$

Finally, by gathering the bounds, we obtain the following proposition:

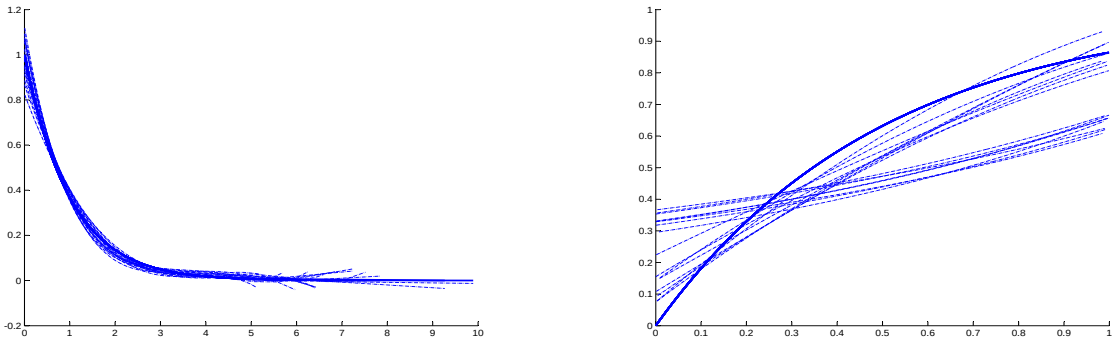

FIGURE 1. Plot of 15 Quotient estimators (left: density estimators $\tilde{g}$ of g, right: Quotient estimators $\tilde{F}$) for Model 4 with $n = 500$.

Proposition 3.3. *Under the assumptions of Lemma 3.1,*

$$\mathbb{E}\|\tilde{F} - F\|^2 \leq \frac{2^4}{g_0^2}\left(\mathbb{E}\|\tilde{\psi} - \psi\|^2 + \mathbb{E}\|\tilde{g} - g\|^2\right) + \frac{C(g_0, \|\psi\|)}{n}, \tag{3.11}$$

where $C(g_0, \|\psi\|)$ is a constant depending on g_0 and $\|\psi\|$.

From Inequality (3.11), we easily deduce by using results (3.7) or (3.8) that $\tilde{F}$ is an adaptive estimator of F if the functions g and ψ have the same regularity $\alpha = \alpha_g = \alpha_\psi$. Here we can state the following result:

Proposition 3.4. *Consider the collection of models* [T], [DP] *or* [W], *with $r > \alpha_F = \alpha_\psi = \alpha_g > 0$ and the estimator $\tilde{F}$ defined by 2.7. Let $L > 0$. Then*

$$\left(\sup_{F \in \mathbb{B}_{\alpha_F,2,\infty}(L)} \mathbb{E}\|F - \tilde{F}\|^2\right)^{\frac{1}{2}} \leq C(\alpha_F, L) n^{-\frac{\alpha_F}{2\alpha_F+1}} \tag{3.12}$$

where $\mathbb{B}_{\alpha_F,2,\infty}(L) = \{t \in \mathcal{B}_{\alpha_F,2,\infty}([0,1]), |t|_{\alpha_F,2} \leq L\}$ where $C(\alpha_F, L)$ is a constant depending on α_F, L and also on ψ, Φ_0 and g_0.

Note that Theorem 2 in Yang (2000) shows that the rate in the sup-norm over a compact is of order $O((\ln n/n)^{(1+\alpha)/(3+2\alpha)})$ *a.s.* where α stands for the regularity of the density function f.

If the index of regularity of F, α_F, is greater than the index of regularity of $\psi = Fg$, α_ψ, then the asymptotic rate of the estimator $\tilde{F}$ is given by $n^{-\alpha_\psi/(1+2\alpha_\psi)}$ instead of the optimal one $n^{-\alpha_F/(1+2\alpha_F)}$. This is the reason why we propose another contrast to estimate directly F.

4. Study of the mean square estimator

In this section, we study the mean square estimator of F from (2.2) and its adaptive version. In this context, we define the empirical norm $\|\cdot\|_n$ as follows: for $t \in S_m$,

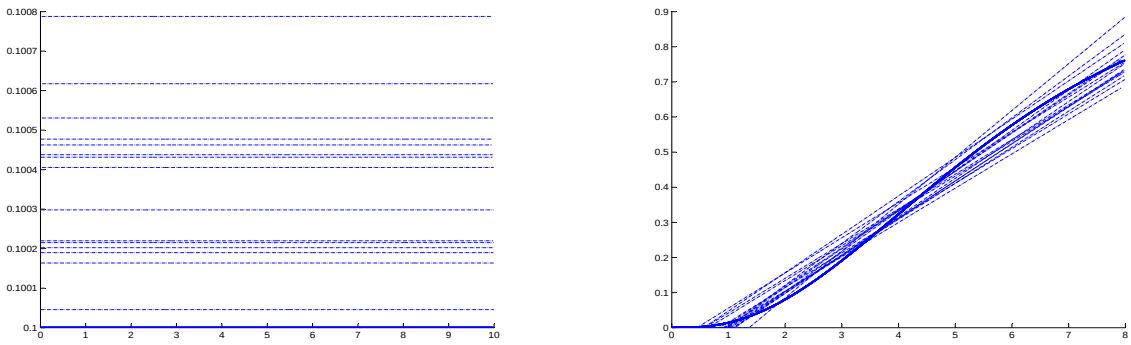

FIGURE 2. Plot of 15 Quotient estimators (left: density estimators $\tilde{g}$ of g, right: Quotient estimators $\tilde{F}$) for Model 5 with $n = 500$.

$\|t\|_n^2 = (1/n)\sum_{i=1}^n t^2(U_i)$. It is a natural norm in regression problems, and under [A1], it is equivalent in mean to the standard Lebesgue integrated $\mathbb{L}^2$-norm, i.e., under [A1]:

$$\forall t \in S_m,\ g_0 \leq \|t\|^2 \leq \mathbb{E}(\|t\|_n^2) = \int t^2(x)g(x)dx \leq g_1\|t\|^2.$$

Then, the mean-square contrast defined by (2.8) can be decomposed as follows:

$$\gamma_n^{MS}(t) - \gamma_n^{MS}(s) = \|t - F\|_n^2 - \|s - F\|_n^2 - 2\nu_n^{MS}(t - s) \tag{4.1}$$

where $\nu_n^{MS}(.)$ is defined by:

$$\nu_n^{MS}(t) = \frac{1}{n}\sum_{i=1}^n (\delta_i - F(U_i))t(U_i) \tag{4.2}$$

which is a centered process since $\mathbb{E}(\delta|U = u) = F(u)$.

In this case, we obtain the following result for the penalized estimator:

Theorem 4.1. *Consider the collections of models* [T] *with* $N_n \leq \sqrt{n}/\ln(n)$ *or* [DP] *or* [W] *with* $N_n \leq n/\ln^2(n)$. *Let* $\hat{F}_{\hat{m}_0}$ *be defined by (2.10), with* $\mathrm{pen}^{MS}(m)$ *defined by (2.11). Then,*

$$\mathbb{E}(\|\hat{F}_{\hat{m}_0} - F\|_n^2) \leq C \inf_{m\in\mathcal{M}_n} (\|F_m - F\|_n^2 + \mathrm{pen}^{MS}(m)) + C'\frac{1}{n} \tag{4.3}$$

where F_m *stands for the orthogonal projection of* F *on* S_m *and* C *and* C' *are constants depending on* Φ_0 *and* g.

Note that the computation of the estimator may be more tedious in practice than the quotient one, but the result is obtained directly for the estimator of F, without any regularity condition on ψ. As a consequence, we obtain here a rate only depending on the regularity of F, and we can state the following result:

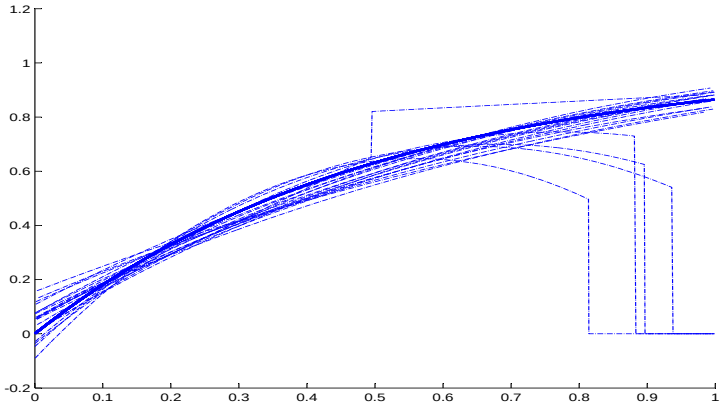

FIGURE 3. Plot of 15 Regression estimators for Model 4 with $n = 500$.

Proposition 4.1. *Consider the collection of models* [T], [DP] *or* [W], *with* $r > \alpha_F > 0$ *and the estimator* $\hat{F}_{\hat{m}_0}$ *defined by (2.10)-(2.11). Let* $L > 0$. *Then*

$$\left(\sup_{F \in \mathbb{B}_{\alpha_F,2,\infty}(L)} \mathbb{E}\|F - \tilde{F}\|_n^2 \right)^{\frac{1}{2}} \leq C(\alpha_F, L) n^{-\frac{\alpha_F}{2\alpha_F+1}} \tag{4.4}$$

where $\mathbb{B}_{\alpha_F,2,\infty}(L) = \{t \in \mathcal{B}_{\alpha_F,2,\infty}([0,1]), |t|_{\alpha_F,2} \leq L\}$ *where* $C(\alpha_F, L)$ *is a constant depending on* α_F, L *and also on* ψ, Φ_0 *and* g_0.

5. SIMULATIONS

Remember that the distribution of δ given $U = u$ is a Bernoulli variable with parameter $F(u)$. We consider the following models for generating data:

Model 1. *Uniform distribution* F: $U \sim \mathcal{U}(0,1)$ and $\delta \sim \mathcal{B}(1, U)$
Model 2. χ^2*-distribution* F: $U \sim \mathcal{U}(0,1)$ and $\delta \sim \mathcal{B}(1, F_{\chi_1^2}(U))$
Model 3. *Quadratic distribution* F: $U \sim \mathcal{U}(0,1)$ and $\delta \sim \mathcal{B}(1, U^2)$
Model 4. *Exponential distribution* F: $U \sim \gamma(1, \lambda)$ and $\delta \sim \mathcal{B}(1, 1-e^{-\mu U})$ with $\lambda = 1, \mu = 0.5$.
Model 5. *Gamma distribution* F: $U \sim \mathcal{U}(0, a)$ with $a = 10$ and $\delta \sim \mathcal{B}(1, F_{\gamma(3,2)}(U))$, where $F_{\gamma(p,\theta)}$ is the cdf of a Gamma distribution of parameter (p, θ).
Model 6. *Beta distribution (S-shape)* F: $U \sim \beta(4,6)$ and $\delta \sim \mathcal{B}(1, F_{\beta(4,8)}(U))$ where $F_{\beta(\alpha,\beta)}$ is the cdf of a Beta distribution of parameter (α, β).

Now, to study the quality of each estimation procedure and to compare them, we compute over J sample replications of size $n = 60, 200, 500$ and 1000 the mean squared errors (MSE) over the sample points $u_1, \ldots, u_K$ falling in $[a, b]$:

$$\text{MSE}_j = \frac{(b-a)}{K} \sum_{k=1}^{K} [F(u_k) - \hat{F}_j(u_k)]^2$$

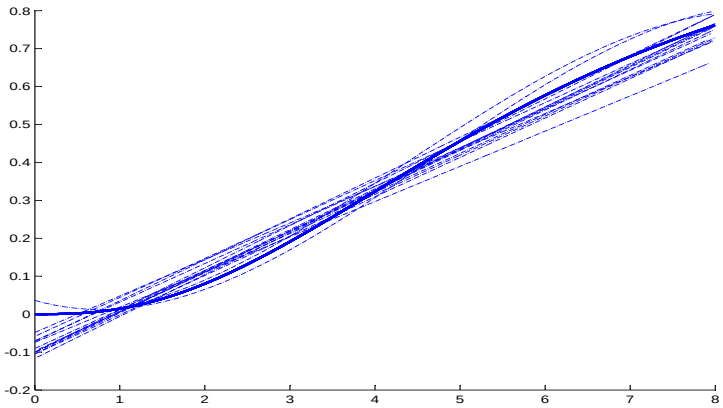

FIGURE 4. Plot of 15 Regression estimators for Model 5 with $n = 500$.

TABLE 1. Monte-Carlo results for the MSE ($\times 10^{-2}$) of the quotient and regression estimators of the cdf F, for $J = 100$ sample replications.

	Quotient est.				Regression est.			
n	60	200	500	1000	60	200	500	1000
model 1	1.56	0.36	0.20	0.066	0.77	0.14	0.055	0.028
model 2	2.23	0.82	0.28	0.15	10.2	0.87	0.18	0.033
model 3	1.64	0.57	0.25	0.074	1.61	0.34	0.069	0.025
model 4	5.65	3.78	1.60	0.40	6.8	1.85	0.46	0.10
model 5	61.2	32.7	1.15	0.95	5.74	3.31	1.22	0.93
model 6	1.92	0.83	0.40	0.20	0.89	0.23	0.12	0.07

where $\hat{F}_j$ stands for the (adaptive) quotient estimator $\tilde{F}$ or for the penalized regression estimator $\hat{F}_{\hat{m}_0}$ computed over the jth sample replication for $j = 1, \ldots, J$. To avoid boundary effects due to the sparsity of the observations at the end of the interval, the MSE_j's are truncated for each replication in the sense that we include in the mean only the u_k less than a given quantile value: $\mathbb{P}(X \leq 0.9) = 0.9$ for model 1, $\mathbb{P}(X \leq 1.8) = 0.82$ for model 2, $\mathbb{P}(X \leq 0.9) = 0.81$ for model 3, $\mathbb{P}(X \leq 1) = 0.86$ for model 4, $\mathbb{P}(X \leq 8) = 0.76$ for model 5 and $\mathbb{P}(X \leq 0.5) = 0.89$ for model 6; thus, the MSE_j are computed over $[a, b]$ with $a = 0$ and $b = 0.9$ for models 1 and 3, $b = 1.8, 1, 8$ and 0.5 for model 2, 4, 5 and 6 respectively. Therefore, the MSE's given in Table 1 stand effectively for the truncated arithmetic means of the MSE_j's. As we can see from results in Table 1, the regression estimator always makes better than the quotient estimators for large samples. However, for small sample size, the quotient estimator can behave as well as and even better than the regression one, see models 2 and 4 for $n = 60, 200$. Note also that, the density estimator $\tilde{g}$ of g is a very attractive estimator by itself as shown in Figures 1 and 2. In some cases

and particularly for model 4, see Figure 1, the quotient mechanism works wrong even if the density estimator is very performant. Figure 1 (right) shows that near than half of the curves don't give the good shape. This is a drawback of quotient strategies which do not have good robustness properties whereas this does not happen with the regression estimators (see Figures 3 and 4).

6. Proofs

6.1. **Talagrand's Inequality.** The following version of Talagrand's Inequality (see Talagrand (1996)) is very useful in most of the proofs:

Lemma 6.1. *Let $Z_1, \ldots, Z_n$ be i.i.d. random variables and $\nu_n(g)$ be defined by $\nu_n(g) = (1/n)\sum_{i=1}^n [g(Z_i) - \mathbb{E}(g(Z_i))]$ for g belonging to a countable class $\mathcal{G}$ of uniformly bounded measurable functions. Then for $\epsilon > 0$*

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}} |\nu_n(g)|^2 - 2(1+2\epsilon)H^2\right]_+ \leq \frac{6}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon\frac{nH^2}{v}} + \frac{8M_1^2}{K_1 n^2 C^2(\epsilon)}e^{-\frac{K_1 C(\epsilon)\sqrt{\epsilon}}{\sqrt{2}}\frac{nH}{M_1}}\right), \tag{6.1}$$

with $C(\epsilon) = (\sqrt{1+\epsilon} - 1) \wedge 1$, K_1 is a universal constant, and where

$$\sup_{g\in\mathcal{G}} \|g\|_\infty \leq M_1, \quad \mathbb{E}\left(\sup_{g\in\mathcal{G}} |\nu_n(g)|\right) \leq H, \quad \sup_{g\in\mathcal{G}} \mathrm{Var}(g(X_1)) \leq v.$$

6.2. **Proof of Lemma 3.1.** Let us write

$$\|\tilde{g} - g\|_\infty \leq \|g - g_{\hat{m}_g}\|_\infty + \|g_{\hat{m}_g} - \hat{g}_{\hat{m}_g}\|_\infty$$

with $\tilde{g} = \hat{g}_{\hat{m}_g}$ defined by (2.4). If g belongs to some Besov space $\mathcal{B}_{\alpha_g,2,\infty}([0,1]$ with $\alpha_g > 0$ then, Lemma 12 in Barron *et al.* (1999) gives (with the rectriction $D_m \geq \log(n), \forall m$):

$$\|g - g_{\hat{m}_g}\|_\infty \leq C D_{\hat{m}_g}^{-\alpha_g} \leq C(\log n)^{-\alpha_g}.$$

Thus, $\|g - g_{\hat{m}_g}\|_\infty$ decreases to 0 as n goes to ∞ and for some integer n_0 large enough, we have for $n \geq n_0$,

$$\mathbb{P}(\|\tilde{g} - g\|_\infty > g_0/2) \leq \mathbb{P}(\|g_{\hat{m}_g} - \hat{g}_{\hat{m}_g}\|_\infty > g_0/4)$$

Now, $\|g_{\hat{m}_g} - \hat{g}_{\hat{m}_g}\|_\infty \leq \Phi_0\sqrt{D_{\hat{m}_g}}\|g_{\hat{m}_g} - \hat{g}_{\hat{m}_g}\|$ and $\|g_{\hat{m}_g} - \hat{g}_{\hat{m}_g}\|^2 = \sum_{\lambda\in\Lambda_{\hat{m}_g}} \nu_{n,g}^2(\varphi_\lambda) = \sup_{t\in B_{\hat{m}_g}} |\nu_{n,g}^2(t)|$. This implies

$$\begin{aligned}
\mathbb{P}(\|\tilde{g} - g\|_\infty > g_0/2) &\leq \mathbb{P}\Big(\sup_{t\in B_{\hat{m}_g}} |\nu_{n,g}(t)| > \frac{g_0}{4\Phi_0\sqrt{D_{\hat{m}_g}}}\Big) \\
&\leq \sum_{m\in\mathcal{M}_n} \mathbb{P}\Big(\sup_{t\in B_m} |\nu_{n,g}(t)| > \frac{g_0}{4\Phi_0\sqrt{D_m}}\Big) \qquad (6.2)
\end{aligned}$$

We apply the version of Talagrand (1996) inequality given in Corollary 2 in Birgé and Massart (1998) which states that there exists a universal constant K_1 such that, for any positive η and λ,

$$\mathbb{P}(\sup_{f\in\mathcal{F}} |\nu_{n,g}(f)| \geq (1+\eta)E + \lambda) \leq 3\exp\left[-K_1 n\left(\frac{\lambda^2}{v} \wedge \frac{(\eta\wedge 1)\lambda}{b}\right)\right].$$

This result can be applied to the class of functions $\mathcal{F} = B_m(0,1)$ by taking $f = t - \mathbb{E}(t(U_1))$, with

$$\sup_{t\in B_m(0,1)} \mathrm{Var}(t(U_1)) \leq \sup_{t\in B_m(0,1)} \int_0^1 t^2(u)g(u)du \leq g_1 := v, \sup_{t\in B_m(0,1)} \|t\|_\infty \leq \Phi_0\sqrt{D_m} := b$$

and $\mathbb{E}(\sup_{t\in B_m(0,1)} \nu_{n,g}^2(t)) = (1/n)\sum_{\lambda\in\Lambda_m} \mathrm{Var}(\varphi_\lambda(U_1)) \leq \Phi_0^2 D_m/n := E^2$. By choosing $\eta = 1$ and $\lambda = g_0/(8\Phi_0\sqrt{D_m})$ and if $2E + \lambda \leq g_0/(4\Phi_0\sqrt{D_m})$, we obtain from (6.2):

$$\begin{aligned}\mathbb{P}(\|\tilde{g} - g\|_\infty > g_0/2) &\leq \sum_{m\in\mathcal{M}_n} 3\exp\left[-K_1 n\left(\frac{\lambda^2}{g_1} \wedge \frac{\lambda}{\Phi_0\sqrt{D_m}}\right)\right] \\ &\leq \sum_{m\in\mathcal{M}_n} 3\exp\left[-K_1C_1 n/D_m\right]\end{aligned}$$

with $C_1 = \left(\frac{g_0^2}{64g_1\Phi_0^2} \wedge \frac{g_0}{8\Phi_0^2}\right)$, if we ensure that $2E + \lambda \leq g_0/(4\Phi_0\sqrt{D_m})$. But with $E = \Phi_0\sqrt{D_m/n}$, this is verified if $D_m \leq [g_0/(16\Phi_0^2)]\sqrt{n}$. Thus, we can deduce that

$$\mathbb{P}(\|\tilde{g} - g\|_\infty > g_0/2) \leq 3|\mathcal{M}_n|\exp\left[-K_1C_1'\sqrt{n}\right]$$

with $C_1' = \left(\frac{g_0^2}{64g_1\Phi_0^2} \wedge \frac{g_0}{8\Phi_0^2}\right)/[g_0/(16\Phi_0^2)]$. Finally, since $|\mathcal{M}_n| \leq n$, if $D_m \leq (K_1C_1)n/(2\ln(n))$ then $\mathbb{P}(\|\tilde{g} - g\|_\infty > g_0/2) \leq 3/n$ and this concludes the proof. □

6.3. **Proof of Theorem 3.1.**

6.3.1. *Proof of a preliminary Lemma.* First, we prove the following lemma:

Lemma 6.2. *Assume that* $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$ *are fulfilled and denote by* $B_{m,m'}(0,1) = \{t \in S_m + S_{m'},\ \|t\| = 1\}$. *Let* $\nu_n(g_t)$ *be defined by (3.5) and*

$$g_t(u,\delta) = \delta t(u), \tag{6.3}$$

then for $\epsilon > 0$

$$\mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)} \nu_n^2(g_t) - p^\psi(m,m')\right)_+ \leq \frac{\kappa_1}{n}\left(e^{-\kappa_2\epsilon(D_m+D_{m'})} + \frac{e^{-\kappa_3\epsilon^{3/2}\sqrt{n}}}{C(\epsilon)^2}\right), \tag{6.4}$$

with $p^\psi(m,m') = 2(1+2\epsilon)\Phi_0^2\int_0^1\psi(x)dx\,(D_m + D_{m'})/n$ *and* $C(\epsilon) = (\sqrt{1+\epsilon} - 1) \wedge 1$. *The constants* κ_i *for* $i = 1,2,3$ *depend on* Φ_0, ψ *and* F.

We apply Talagrand's inequality by taking $Z_i = (U_i, \delta_i)$ for $i = 1,\ldots,n$ and $g(u,\delta) = g_t(u,\delta)$. Usual density arguments show that this result can be applied to the class of functions $\mathcal{G} = \{g_t,\, t \in B_{m,m'}(0,1)\}$. Then we find for the present empirical process the following bounds:

$$\sup_{g\in\mathcal{G}} \|g\|_\infty = \sup_{t\in B_{m,m'}(0,1)} \|g_t\|_\infty \leq \Phi_0\sqrt{D(m,m')} := M_1$$

with $D(m,m')$ denoting the dimension of $S_m + S_{m'}$. Then

$$\begin{aligned}\sup_{g\in\mathcal{G}}\mathrm{Var}(g(U_1,\delta_1)) &= \sup_{t\in B_{m,m'}(0,1)}\mathrm{Var}(g_t(U_1,\delta_1)) = \sup_{t\in B_{m,m'}(0,1)}\mathbb{E}(\delta_1 t^2(U_1))\\ &= \sup_{t\in B_{m,m'}(0,1)}\int_0^1 t^2(u)\psi(u)du \le g_1 := v.\end{aligned}$$

Lastly,

$$\begin{aligned}\mathbb{E}\left(\sup_{g\in\mathcal{G}}\nu_n^2(g)\right) &= \mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}\nu_n^2(g_t)\right) \le \sum_{\lambda\in\Lambda_{m,m'}}\frac{1}{n}\mathrm{Var}(\delta_1\varphi_\lambda(U_1))\\ &\le \frac{\Phi_0^2 D(m,m')}{n}\int_0^1\psi(x)dx = C_1\frac{D(m,m')}{n} := H^2.\end{aligned}$$

with the natural notation $\Lambda_{m,m'} = \Lambda_m \cup \Lambda_{m'}$. Then it follows from (6.1) that

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

where κ_i for $i = 1,2,3$ are constant depending on K_1 and C_1 and $p^\psi(m,m') = 2(1+2\epsilon)C_1(D_m + D_{m'})/n$. □

6.3.2. *Proof of Theorem 3.1.* It follows from the definition of $\hat{\psi}_{\hat{m}}$ that: $\forall m \in \mathcal{M}_n$,

$$\gamma_n^\psi(\hat{\psi}_{\hat{m}}) + \mathrm{pen}^\psi(\hat{m}) \le \gamma_n^\psi(\psi_m) + \mathrm{pen}^\psi(m). \tag{6.5}$$

Then by using decomposition (3.6), it follows from (6.5) and from the definition of the process $\nu_n(g_t)$ given by (3.5) and (6.3) that:

$$\begin{aligned}\|\hat{\psi}_{\hat{m}} - \psi\|^2 &\le \|\psi_m - \psi\|^2 + 2\nu_n(g_{\hat{\psi}_{\hat{m}}-\psi}) + \mathrm{pen}^\psi(m) - \mathrm{pen}^\psi(\hat{m})\\ &\le \|\psi_m - \psi\|^2 + \frac{1}{4}\|\hat{\psi}_{\hat{m}} - \psi_m\|^2 + 4\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(g_t)\\ &\quad + \mathrm{pen}^\psi(m) - \mathrm{pen}^\psi(\hat{m}) \qquad (6.6)\end{aligned}$$

where we recall that $B_{m,\hat{m}}(0,1) = \{t \in S_m + S_{\hat{m}} \ / \ \|t\| \le 1\}$. Note that the norm connection as described by (2.3) still holds for any element t of $S_m + S_{m'}$ as follows: $\|t\|_\infty \le \Phi_0 \max(D_m, D_{m'})\|t\|$. Indeed, under $(\mathcal{H}_2)$, we restrict our attention to nested collections of models, so that $S_m + S_{\hat{m}}$ is equal to the larger of the two spaces. For a fixed integer m, we denote by $D(m')$ the dimension of $S_m + S_{m'}$, for all $m' \in \mathcal{M}_n$. Note that $D(m') = \max(D_m, D_{m'}) \le D_m + D_{m'}$.
Let $p^\psi(m,m')$ be such that

$$8p^\psi(m,m') \le \mathrm{pen}^\psi(m) + \mathrm{pen}^\psi(m') \quad \text{for all } m,\ m' \text{ in } \mathcal{M}_n. \tag{6.7}$$

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

$$\frac{1}{2}\|\hat{\psi}_{\hat{m}}-\psi\|^2 \;\leq\; \frac{3}{2}\|\psi-\psi_m\|^2+2\,\mathrm{pen}^{\psi}(m)+8\left(\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(g_t)-p^{\psi}(m,\hat{m})\right).$$

Then if we prove

(6.8)
$$\mathbb{E}\left(\sup_{t\in B_{m,\hat{m}}(0,1)}\nu_n^2(g_t)-p^{\psi}(m,\hat{m})\right)_+\leq\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}\nu_n^2(g_t)-p^{\psi}(m,m')\right)_+\leq\frac{C}{n}$$

we have the following result, which proves the theorem: $\forall m\in\mathcal{M}_n$,

$$\mathbb{E}(\|\hat{\psi}_{\hat{m}}-\psi\|^2)\leq 3\|\psi-\psi_m\|^2+4\mathrm{pen}^{\psi}(m)+\frac{C}{n}\,.$$

Therefore by using equation (6.7) and the definition of $p^{\psi}(m,m')$ in Lemma 6.2, we choose

$$\mathrm{pen}^{\psi}(m)=16(1+2\epsilon)\int_0^1\psi(x)dx\frac{D_m}{n}$$

Inequality (6.8) is a straightforward consequence of Lemma 6.2 since

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}\nu_n^2(g_t)-p^{\psi}(m,m')\right)_+\leq\kappa_1\left(\frac{\sum_{m'\in\mathcal{M}_n}e^{-\kappa_2\epsilon D(m')}}{n}+\frac{|\mathcal{M}_n|}{n}e^{-\kappa_3\epsilon^{3/2}\sqrt{n}}\right).$$

Then by taking $\epsilon=1/2$ and assuming that $|\mathcal{M}_n|\leq n$ and since, under $(\mathcal{H}_2)$, $\sum_{m\in\mathcal{M}_n}e^{-aD_m}\leq\sum_{k=1}^n e^{-ka}\leq\Sigma(a)<+\infty,\forall a>0$, this leads to the bound

$$\sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in B_{m,m'}(0,1)}\nu_n^2(g_t)-p^{\psi}(m,m')\right)_+\leq\frac{C}{n},$$

and this ensures (6.8). □

6.4. **Proof of Theorem 3.2.** We start by writing that, $\forall m\in\mathcal{M}_n$,

$$\gamma_n(\hat{F}_{\hat{m}_0})+\mathrm{pen}^{\mathrm{MS}}(\hat{m}_0)\leq\gamma_n(F_m)+\mathrm{pen}^{\mathrm{MS}}(m)$$

and by using the decomposition (4.1). It follows that

$$\|\hat{F}_{\hat{m}_0}-F\|_n^2\leq\|F_m-F\|_n^2+2\nu_n^{MS}(\hat{F}_{\hat{m}_0}-F_m)+\mathrm{pen}^{\mathrm{MS}}(m)-\mathrm{pen}^{\mathrm{MS}}(\hat{m}_0).$$

Let us introduce, in the same way as Baraud *et al.* (2001), for $\|t\|_g^2=\int_{[0,1]}t^2(u)g(u)\,du$, the ball $B^g_{m,m'}(0,1)=\{t\in S_m+S_{m'},\|t\|_g=1\}$ and the set

$$\Omega_n=\left\{\omega,\left|\frac{\|t\|_n^2}{\|t\|_g^2}-1\right|\leq\frac{1}{2},\ \forall t\in\bigcup_{m,m'\in\mathcal{M}_n}(S_m+S_{m'})\setminus\{0\}\right\}.$$

On the complement of Ω_n, a separate study leads to the following lemma:

Lemma 6.3. *If $N_n\leq\sqrt{n}/\ln(n)$ for* [T] *or $N_n\leq n/\ln^2(n)$ for* [P] *or* [W]*, then $\mathbb{P}(\Omega_n^c)\leq c/n$ and, $\mathbb{E}\left(\|\hat{F}_{\hat{m}_0}-F\|_n^2\mathbb{1}_{\Omega_n^c}\right)\leq c'/n$, where c and c' are positive constants.*

Proof of Lemma 6.3. That $\mathbb{P}(\Omega_n^c) \leq c/n^2$ is in fact a pure property of the basis and is proved under our assumptions in Baraud (2002). Moreover, $\|\hat{F}_{\hat{m}_0} - F\|_n^2 \leq 2(\|\hat{F}_{\hat{m}_0}\|_n^2 + \|F\|_n^2)$. Now $\|F\|_n^2 \leq 1$ and $\|\hat{F}_{\hat{m}_0}\|_n^2 = (1/n)\|\Pi_{\hat{m}_0}\delta\|_{\mathbb{R}^n}^2$ where $\delta = (\delta_1, \ldots, \delta_n)$, $\Pi_{\hat{m}}$ is the orthogonal projection in $\mathbb{R}^n$ on $\{t(U_1), \ldots, t(U_n)), t \in S_m\}$ and $\|\cdot\|_{\mathbb{R}^n}$ is the Euclidean norm in $\mathbb{R}^n$. It follows that $\|\hat{F}_{\hat{m}_0}\|_n^2 \leq (1/n)\|\delta\|_{\mathbb{R}^n}^2 = (1/n)\sum_{i=1}^n \delta_i^2 \leq 1$. Therefore $\mathbb{E}\left(\|\hat{F}_{\hat{m}_0} - F\|_n^2 \mathbb{1}_{\Omega_n^c}\right) \leq 2\mathbb{P}(\Omega_n^c) \leq c'/n$. □

Therefore, we focus on the study of the bounds on Ω_n, where the inequality $\|t\|_g^2 \leq 2\|t\|_n^2$ is fulfilled. We obtain

$$\begin{aligned}
\|\hat{F}_{\hat{m}_0} - F\|_n^2 \mathbb{1}_{\Omega_n} &\leq \|F_m - F\|_n^2 + \frac{1}{8}\|\hat{F}_{\hat{m}_0} - F_m\|_f^2 \mathbb{1}_{\Omega_n} + 16 \sup_{t \in B^g_{\hat{m}_0,m}(0,1)} [\nu_n^{MS}]^2(t) \\
&\quad + \mathrm{pen}^{\mathrm{MS}}(m) - \mathrm{pen}^{\mathrm{MS}}(\hat{m}_0) \\
&\leq \left(1 + \frac{1}{2}\right)\|F_m - F\|_n^2 + \frac{1}{2}\|\hat{F}_{\hat{m}_0} - F\|_n^2 \mathbb{1}_{\Omega_n} \\
&\quad + 16\Big(\sup_{t \in B^g_{\hat{m}_0,m}(0,1)} [\nu_n^{MS}]^2(t) - \tilde{p}(m, \hat{m}_0)\Big)_+ \\
&\quad + \mathrm{pen}^{MS}(m) + 16\tilde{p}(m, \hat{m}_0) - \mathrm{pen}^{MS}(\hat{m}_0).
\end{aligned}$$

Let $(\bar{\varphi}_\lambda)_{\lambda \in \Lambda_{m,m'}}$ be an orthonormal basis of $S_m + S_{m'}$ for the scalar product $\langle \cdot, \cdot \rangle_g$ (built by Gramm-Schmidt orthonormalization). It is easy to see that:

$$\begin{aligned}
\mathbb{E}\Big(\sup_{t \in B^g_{m',m}(0,1)} [\nu_n^{MS}]^2(t)\Big) &\leq \sum_{\lambda \in \Lambda_{m,m'}} \frac{1}{n}\mathrm{Var}\Big([\delta_1 - F(U_1)]\bar{\varphi}_\lambda(U_1)\Big) \\
&\leq \sum_{\lambda \in \Lambda_{m,m'}} \frac{1}{n}\mathbb{E}_X\Big(\int_0^1 [\mathbb{1}_{X \leq u} - F(u)]^2 \bar{\varphi}_\lambda(u)^2 g(u) du\Big) \\
&\leq \frac{1}{n}\sum_{\lambda \in \Lambda_{m,m'}} \Big(\int_0^1 \mathbb{E}_X[\mathbb{1}_{X \leq u} - F(u)]^2 \bar{\varphi}_\lambda^2(u) g(u) du\Big) \\
&\leq \frac{1}{n}\sum_{\lambda \in \Lambda_{m,m'}} \Big(\int_0^1 F(u)(1 - F(u))\bar{\varphi}_\lambda^2(u) g(u) du\Big) \\
&\leq \frac{D_m \vee D_{m'}}{n}
\end{aligned}$$

as $F(u)(1 - F(u)) \leq 1$. Therefore, we obtain by applying Talagrand's Inequality

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\Big(\sup_{t \in B^g_{m',m}((0,1)} [\nu_n^{MS}]^2(t) - \tilde{p}(m, m')\Big)_+ \leq \frac{c}{n}$$

with

$$\tilde{p}(m, m') = 4\frac{D_m \vee D_{m'}}{n} := 4H^2,$$

$$\sup_{t\in B^g_{m',m}(0,1)} \mathrm{Var}[(\delta_1 - F(U_1))t(U_1)] \leq \sup_{t\in B^g_{m',m}(0,1)} \mathbb{E}(t^2(U_1)) = 1 := v,$$

$$\text{and } \sup_{t\in B^g_{m',m}(0,1)} \|(\delta_1 - F(U_1))t\|_\infty \leq \sup_{t\in B^g_{m',m}(0,1)} \|t\|_\infty \leq \Phi_0\sqrt{D_{m,m'}}/g_0 := M_1.$$

Acknowledgments

Here, we would like to express to Y. Rozenholc our sincere appreciation for providing us very performant Matlab programs. All the material is available on http://www.math-info.univ-paris5.fr/∼rozen.

References

Baraud, Y. (2002) Model selection for regression on a random design. *ESAIM P&S* **6**, 127-146.

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

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.

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

Birgé, L. and Rozenholc, Y. (2006). How many bins should be put in a regular histogram. *ESAIM Probab. Stat.* **10**, 24-45.

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

Comte, F. and Rozenholc, Y. (2004). A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.* **56**, 449-473.

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

Diamond, I.D. and McDonald, J.W. (1991). The analysis of current status data. *Demographic applications of event history analysis.*(eds J. Trussell, R. Hankinson & J. Tilton). Oxford University Press, Oxford.

Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation with wavelet shrinkage. *Ann. Statist.* **26**, 879-921.

Groeneboom, P. and Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. Boston, Birkhäuser Verlag.

Jewell, N. P. and van der Laan, M. (2004).*Current status data: review, recent developments and open problems.* Advances in survival analysis, 625–642, Handbook of Statist., 23, Elsevier, Amsterdam.

Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

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

Massart, P. (2007) Concentration inequalities and model selection. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin.

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

VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric likelihood estimators. *Ann. Statist.* **21**, 14-44.

YANG, S. (2000). Functional estimation under interval censoring case 1. *J. Statist. Plann. Inference* **89**, 135-144.

E. BRUNEL, I3M, UMR CNRS 5149, UNIVERSITÉ MONTPELLIER 2, CC 051, PLACE E. BATAILLON, 34095 MONTPELLIER, FRANCE