NONPARAMETRIC SURVIVAL FUNCTION ESTIMATION FOR DATA SUBJECT TO INTERVAL CENSORING CASE 2

OLIVIER BOUAZIZ[(1)], ELODIE BRUNEL[(2)], FABIENNE COMTE[(1)]

Abstract. In this paper, we propose a new strategy of estimation for the survival function S, associated to a survival time subject to interval censoring case 2. Our method is based on a least squares contrast of regression type with parameters corresponding to the coefficients of the development of S on an orthonormal basis. We obtain a collection of projection estimators where the dimension of the projection space has to be adequately chosen via a model selection procedure. For compactly supported bases, we obtain adaptive results leading to general nonparametric rates. However, our results can be used for non compactly supported bases, a true novelty in regression setting, and we use specifically the Laguerre basis which is $\mathbb{R}^+$-supported and thus well suited when nonnegative random variables are involved in the model. Simulation results comparing our proposal with previous strategies show that it works well in a very general context. A real dataset is considered to illustrate the methodology.

MSC 2010 subject classification: 62N01–62N02–62G05
Keywords and phrases: Interval censoring, nonparametric estimation, regression contrast, survival function.

1. Introduction

Let X_1 be a survival time of interest (the time at which the event of interest occurs) with unknown survival function S, $S(x) = \mathbb{P}(X_1 > x)$. Our aim is to propose a nonparametric estimator of S in a setting where X_1 is not observed, but subject to interval censoring case 2. To be more precise, the observations are $(L_i, U_i, \delta_i)_{1 \le i \le n}$ with

$$\delta_i = \begin{cases} -1 & \text{if} \quad X_i \le L_i \\ 0 & \text{if} \quad L_i < X_i \le U_i \\ 1 & \text{if} \quad X_i > U_i \end{cases} \tag{1}$$

We assume that the triples $(L_i, U_i, \delta_i)_{1 \le i \le n}$ are i.i.d. and that the (L_i, U_i) are independent of the X_i. Note that interval censoring case 1, also called current status data, corresponds to $U_i = L_i$ (or $L_i = -\infty$), so that the δ_is have only two modalities.

We are aware of previous proposals on the topic. First, Turnbull (1976) introduced an iterative procedure in order to obtain a Non Parametric Maximum Likelihood Estimator (NPMLE) of the survival function under different censoring and truncation types. Later on, Groeneboom and Wellner (1992) introduced the iterative convex minorant algorithm based on isotonic regression theory. Groeneboom (1996) summarizes this as follows: "If one wants to estimate the distribution function by the nonparametric maximum likelihood estimator (NPMLE), one has to use methods from isotonic regression theory and convex optimization to even compute the estimator in an efficient way", see p.69 therein; see also the chapter by Wellner (1995). In Groeneboom et al. (2010) it was proved that the resulting estimator has the rate $(n\log(n))^{-1/3}$ for interval-censoring

[(1)]: MAP5, UMR 8145 CNRS, Université Paris Descartes, FRANCE, email: olivier.bouaziz@parisdescartes.fr, email: fabienne.comte@parisdescartes.fr
[(2)]: IMAG, Univ Montpellier, CNRS, Montpellier, France, email: ebrunel@math.univ-montp2.fr .

case 1. The authors also introduced two smooth estimators: the maximum smoothed likelihood estimator (MSLE), a general likelihood-based M-estimator that will turn out to be smooth automatically and the smoothed maximum likelihood estimator (SMLE), obtained by convolving the discrete NPMLE. For interval censoring case 1, they proved that these estimators reach the faster rate of convergence $n^{-2/5}$. Now, the factor 2/5 in these rates should be read as $\alpha/(2\alpha+1)$ for $\alpha = 2$, α being the regularity of the function under estimation. For interval-censoring case 2, the NPMLE and SMLE estimators were studied in Groeneboom and Ketelaars (2011) (see also Geskus and Groeneboom (1999)). For $\alpha = 1$, it was conjectured that they reach the rate $n^{-1/3}$ in the so-called *separated case* where L_i and U_i cannot become arbitrarily close. In the *non-separated case* (where the density of (L_i, U_i) is positive on the diagonal), it was conjectured that these estimators achieve the improved rate $(n \log(n))^{-1/3}$. In Birgé (1999) an explicit histogram estimator was built for interval censoring case 2 which was proved to reach the rates $(n \log(n))^{-1/3}$ and $n^{-1/3}$ in the *non-separated* ans *separated* cases in Birgé (1999) and in Groeneboom and Ketelaars (2011).

Other types of smooth estimators have been proposed for interval censored data. For interval censoring case 1, Brunel and Comte (2009) proposed two adaptive estimators, one of quotient type and another one of regression type, both using projection methods. They proved the rate $n^{-\alpha/(2\alpha+1)}$ for their estimators. Note that the regression property was also exploited in similar context by Yang (2000). For interval censored data case 2, spline methods were introduced in Kooperberg and Stone (1992) and a smooth alternative to the NPMLE was proposed by introducing a log-concave constraint in the estimation procedure in Anderson-Bergman and Yu (2016). However, no theoretical rates were provided for these last two estimators.

In the present work, we propose a least squares contrast minimization method, in the spirit of Brunel and Comte (2009). First, we derive two naive procedures, one which relies on the regression equation $\mathbb{E}[1 - \mathbb{1}_{\delta_i=-1}|L_i] = S(L_i)$, and another one based on its counterpart $\mathbb{E}[\mathbb{1}_{\delta_i=1}|U_i] = S(U_i)$. We briefly discuss their obvious drawbacks and explain how we elaborate a method taking both relations into account. We finally define a mixed contrast leading to the main estimator of the paper. We detail in what sense it improves the estimation.

We also emphasize that the regression method presented here is simple to implement since it only requires the inversion of an easy-to-compute matrix involved in the definition of the coefficients of the estimator; the matrix being symmetric positive-definite, the inversion algorithm is fast. Our results look different from most previous ones because our aim is different: we do not provide pointwise decomposition of the error nor limiting distribution, but global integrated risk on a support which can be compact or not. The density weighting our risk avoids the distinction between *separated* and *non separated cases*, and our simulations show that the method is robust to both. Note that, similarly to Brunel and Comte (2009), we obtain a final estimator of the survival function taking values between 1 and 0 and a posteriori modified to be decreasing thanks to the procedure described in Chernozhukov et al. (2009), which is conveniently associated with a R-package **Rearrangement**. The need for such final modification may seem a drawback, but it is known that the risk bound proved for the initial estimator still holds true for the final one. This is a quite deserving alternative to isotonic regression techniques. Lastly, we provide and study a model selection procedure that drives to an automatic squared-bias variance tradeoff. Our estimator does not win the supplementary logarithm in the non-separated case but its rate, under additional assumptions discussed in Section 2.6, and in particular if the function to estimate is regular, automatically adapts to its regularity index whatever the hypothesis on the joint distribution of (L_i, U_i).

The bases used in Brunel and Comte (2009) are compactly supported. This requires to define the domain of estimation at the very beginning of the procedure. This step is avoided by using the Laguerre basis, which is $\mathbb{R}^+$-supported. However, this non-compact feature is excluded from the theoretical framework of Brunel and Comte (2009), as well as from most other papers on nonparametric least squares regression (see e.g. Baraud, 2002). Therefore, we borrow elements from a recent work by Comte and Genon-Catalot (2019), to include this possibility in our results.

The plan of the paper is the following. Section 2 describes the bases and projection spaces and explains the way estimators are built. Non asymptotic risk bounds are then proved, which allow to discuss asymptotic rates in a rather general setting. Section 3 develops the model selection strategy and the associated risk bounds. Then we show, in thorough simulation experiments presented in Section 4 that our estimator works well, especially when using the Laguerre basis, in comparison with the NPMLE implemented in the **prodlim** R package, the log-concave estimator proposed by Anderson-Bergman and Yu (2016) in the **logconPH** R package and the smoothed maximum likelihood estimator (SMLE) obtained by convolving the NPMLE with a smoothing kernel as in Groeneboom and Ketelaars (2011). Real interval censored data on HIV infections are analyzed in Section 5 using our estimator. Most proofs are gathered in Section 6.

2. Definition and study of projection estimators

We first present the different bases associated with projection estimators defined in the sequel.

2.1. **Projection spaces.** Consider $\Sigma_m = \Sigma_m(I) = \text{span}(\varphi_0, \dots, \varphi_{m-1})$ where $(\varphi_j)_{0\le j\le m-1}$ constitutes an orthonormal basis $\langle \varphi_j, \varphi_k\rangle = \delta_{j,k}$ with respect to the scalar product $\langle u, v\rangle = \int_I u(x)v(x)dx$. The domain I is the support of the basis and can be an interval $[a, b]$ which shall be taken equal to $[0, 1]$ for simplicity in the examples below. We will also consider the case where $I = \mathbb{R}^+$ which can be very convenient in this type of problems.

The examples of bases we have in mind are the following.

- Histogram basis with $I = [0, 1]$, defined by $h_j(x) = \sqrt{m}\mathbb{1}_{[j/m,(j+1)/m[}$ for $j = 0, \dots, m-1$. They can be generalized to piecewise polynomials with given degree r, by rescaling $Q_0, \dots, Q_r$ the Legendre basis on each sub-interval $[j/m, (j + 1)/m[$, $j = 0, \dots, m - 1$.
- Trigonometric basis, $I = [0, 1]$, $t_0(x) = \mathbb{1}_{[0,1]}(x)$, $t_{2j-1}(x) = \sqrt{2}\cos(2\pi jx)\mathbb{1}_{[0,1]}(x)$, $t_{2j}(x) = \sqrt{2}\sin(2\pi jx)\mathbb{1}_{[0,1]}(x)$, for $2j \le m - 1$. Generally, m is chosen odd and in this case $j = 1, \dots, \frac{m-1}{2}$.
- For the Laguerre basis associated with $I = \mathbb{R}^+$, we define the Laguerre polynomials (P_j) and the Laguerre functions (ℓ_j):

$$P_j(x) = \sum_{k=0}^{j}(-1)^k\binom{j}{k}\frac{x^k}{k!}, \qquad \ell_j(x) = \sqrt{2}P_j(2x)e^{-x}\mathbb{1}_{x\ge 0}, \quad j \ge 0. \tag{2}$$

  The collection $(\ell_j)_{j\ge 0}$ is a complete orthonormal system on $\mathbb{L}^2(\mathbb{R}^+)$, such that (see Abramowitz and Stegun, 1964) $\forall j \ge 0, \forall x \in \mathbb{R}^+, |\ell_j(x)| \le \sqrt{2}$. A function $f \in \mathbb{L}^2(\mathbb{R}^+)$ can be developed f on the Laguerre basis, $f = \sum_{j\ge 0} a_j(f)\ell_j, a_j(f) = \langle f, \ell_j\rangle$.

The general notation for all these bases is $(\varphi_j)_j$. They all satisfy

$$\forall m \in \mathbb{N}\setminus\{0\}, \quad \sup_{x\in I}\sum_{j=0}^{m-1}\varphi_j^2(x) := \|\sum_{j=0}^{m-1}\varphi_j^2\|_\infty \le c_\varphi^2 m, \tag{3}$$

for some constant $c_\varphi > 0$ depending on the basis only. For the histogram basis, and the trigonometric basis with odd m, we have $c_\varphi = 1$ and for the Laguerre basis, $c_\varphi^2 = 2$.

The first two bases are compactly supported, and the last one is not. Most regression results hold with compactly supported bases, a case which is generally exclusively considered. In this work, we provide results in the setting of non compactly supported bases, and show empirically that the Laguerre basis is very relevant for survival function estimation. It has the advantage that we do not have to choose an estimation support for the basis and thus for the computation of the coefficients of the function in the basis.

Moreover, we mention that we estimate the survival function rather than the cumulative distribution function because we need the function under estimation to be possibly square-integrable on $\mathbb{R}^+$, in order to use the Laguerre basis. Note that survival functions in all classical models are square-integrable on $\mathbb{R}^+$. For instance, $S(x) = P_{\lambda,k}(x)e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)$ for a $\gamma(k,\lambda)$ density, $P_{\lambda,k}$ being a polynomial depending on λ with degree $k-1$, $S(x) = e^{-(x/\lambda)^k}\mathbb{1}_{\mathbb{R}^+}(x)$ for a Weibull density with parameters k, λ, $S(x) = (x_m/x)^k \mathbb{1}_{[x_m,+\infty[}(x)$ for $x_m > 0$, $k > 1/2$ for a Pareto density, the Gompertz-Makeham density $S(x) = \exp\{-\lambda x - \frac{\alpha}{\beta}(e^{\beta x} - 1)\}\mathbb{1}_{x\geq 0}$, for $\alpha, \beta, \lambda > 0$, are square integrable.

2.2. **Notation.** Let $(L_i, U_i, \delta_i)_{1\leq i\leq n}$ be a n-sample from model (1). We denote by f_U and f_L the densities of U_1 and L_1 and by $f_{(L,U)}$ the joint density of (L_1, U_1). We denote by $(\varphi_j)_{0\leq j\leq m-1}$ an orthonormal $\mathbb{L}^2(I, dx)$ basis as described in section 2.1.

We also use all along the paper the following notation. For any measurable I-supported functions $\psi, \tilde{\psi}$, we define the weighted $\mathbb{L}^2(I, f_Z(x)dx)$-norms and scalar products, for $Z = L, U$,

$$\|\psi\|_Z^2 = \int \psi^2(x) f_Z(x)dx, \qquad \text{and } \langle \psi, \tilde{\psi}\rangle_Z = \int \psi(x)\tilde{\psi}(x) f_Z(x)dx, \tag{4}$$

as soon as $\|\psi\|_Z^2 < +\infty$, $\|\tilde{\psi}\|_Z^2 < +\infty$, and their empirical counterparts:

$$\|\psi\|_{n,Z}^2 = \frac{1}{n}\sum_{i=1}^n \psi^2(Z_i), \quad \langle \psi, \tilde{\psi}\rangle_{n,Z} = \frac{1}{n}\sum_{i=1}^n \psi(Z_i)\tilde{\psi}(Z_i). \tag{5}$$

Clearly, $\mathbb{E}(\|\psi\|_{n,Z}^2) = \|\psi\|_Z^2$, and $\mathbb{E}(\langle \psi, \tilde{\psi}\rangle_{n,Z}) = \langle \psi, \tilde{\psi}\rangle_Z$ for $Z = L, U$.

As classical in regression setting, the following matrices and vectors are useful:

$$\begin{cases} \Phi_m^{(L)} = (\varphi_j(L_i))_{1\leq i\leq n, 0\leq j\leq m-1}, & \vec{\delta}^{(L)} = (1 - \mathbb{1}_{\delta_i=-1})_{1\leq i\leq n} = (1 - \mathbb{1}_{X_i\leq L_i})_{1\leq i\leq n}\,, \\ \Phi_m^{(U)} = (\varphi_j(U_i))_{1\leq i\leq n, 0\leq j\leq m-1}, & \vec{\delta}^{(U)} = (\mathbb{1}_{\delta_i=1})_{1\leq i\leq n} = (1 - \mathbb{1}_{X_i\leq U_i})_{1\leq i\leq n}\,, \end{cases} \tag{6}$$

and

$$\Psi_{m,Z} = (\langle \varphi_j, \varphi_k\rangle_Z)_{1\leq j,k\leq m}\,, \quad \widehat{\Psi}_{m,Z} = (\langle \varphi_j, \varphi_k\rangle_{n,Z}) \text{ for } \quad Z = U, L. \tag{7}$$

We have $\Psi_{m,Z} = \mathbb{E}(\widehat{\Psi}_{m,Z})$ for $Z = L, U$ and

$$\widehat{\Psi}_{m,L} = \frac{1}{n}\,{}^t\Phi_m^{(L)}\Phi_m^{(L)}, \quad \widehat{\Psi}_{m,U} = \frac{1}{n}\,{}^t\Phi_m^{(U)}\Phi_m^{(U)}.$$

In the sequel, the norm associated to matrices is the operator norm $\|A\|_{\rm op}$ defined as the square-root of the largest eigenvalue of the matrix ${}^tA\,A$ (or $A\,{}^tA$). If A is a square symmetric and nonnegative matrix (i.e. for all vector $\vec{x}$, ${}^t\vec{x}\,Ax \geq 0$), then $\|A\|_{\rm op}$ is simply the largest of the eigenvalues of A, which are all nonnegative.

In particular, $\Psi_{m,Z}$ and $\widehat{\Psi}_{m,Z}$ are symmetric nonnegative matrices. Indeed, for $Z = L, U$, we have ${}^t\vec{a}\;\Psi_{m,Z}\vec{a} = \|t\|_Z^2 \geq 0$ where $t = \sum_{j=0}^{m-1} a_j\varphi_j$, and ${}^t\vec{a} = (a_0, \ldots, a_{m-1})$; analogously, ${}^t\vec{a}\;\widehat{\Psi}_{m,Z}\vec{a} = \|t\|_{n,Z}^2 \geq 0$.

2.3. **Two naive regression estimators.** The first idea is to extend the strategy developed in Brunel and Comte (2009) in presence of case 1 interval censoring. Noticing that

$$\mathbb{E}(1 - \mathbb{1}_{\delta_i=-1}|L_i) = S(L_i) \tag{8}$$

we can define

$$\widehat{S}_m^{(L)} = \arg\min_{t\in\Sigma_m} \gamma_n^{(L)}(t), \quad \gamma_n^{(L)}(t) = \frac{1}{n}\sum_{i=1}^n t^2(L_i) - \frac{2}{n}\sum_{i=1}^n (1 - \mathbb{1}_{\delta_i=-1})t(L_i).$$

This corresponds to the least squares estimator associated with the regression model (8), where S would be replaced by S_m, the projection of S on Σ_m and the m explanatory variables would be $(\varphi_j(L_i))_{1\le i\le n}$ for $j = 0, \ldots, m-1$. To understand why the estimator may be suitable, just compute the expectation of the criterion (which is also its almost sure limit when n tends to infinity). We have

$$\begin{aligned}
\mathbb{E}(\gamma_n^{(L)}(t)) &= \mathbb{E}[t^2(L_1)] - 2\mathbb{E}[\mathbb{E}((1 - \mathbb{1}_{X_1\le L_1})|L_1)t(L_1)] = \int_I t^2(x)f_L(x)dx - 2\mathbb{E}[S(L_1)t(L_1)] \\
&= \int_I (t(x) - S(x))^2 f_L(x)dx - \int_I S^2(x)f_L(x)dx.
\end{aligned}$$

Clearly, the resulting term is minimal for $t = S$ and thus, the minimizer of $\gamma_n^{(L)}$ is likely to asymptotically minimize $\|t - S\|_L^2$ and to be near of S.
Similarly, relying on the equality $\mathbb{E}(\mathbb{1}_{\delta_i=1}|U_i) = S(U_i)$, we can set

$$\widehat{S}_m^{(U)} = \arg\min_{t\in\Sigma_m} \gamma_n^{(U)}(t), \quad \gamma_n^{(U)}(t) = \frac{1}{n}\sum_{i=1}^n t^2(U_i) - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=1}t(U_i).$$

Standard computations analogous to those in linear regression models yield to the following formula, for $\vec{\hat{a}}_m^{(Z)} = {}^t(\hat{a}_0^{(Z)}, \ldots, \hat{a}_{m-1}^{(Z)})$, $Z = L, U$, the coordinates of $\widehat{S}_m^{(Z)}$ in the basis $(\varphi_j)_{0\le j\le m-1}$,

$$\widehat{S}_m^{(Z)}(x) = \sum_{j=0}^{m-1} \hat{a}_j^{(Z)}\varphi_j(x) \text{ with } \vec{\hat{a}}_m^{(Z)} = \left({}^t\Phi_m^{(Z)}\Phi_m^{(Z)}\right)^{-1} {}^t\Phi_m^{(Z)}\vec{\delta}^{(Z)} \tag{9}$$

where $\Phi_m^{(Z)}$, $\vec{\delta}^{(Z)}$ are defined in (6), provided that ${}^t\Phi_m^{(Z)}\Phi_m^{(Z)}$ is invertible. Note that, if the bases are compactly supported, their supports I_Z for $Z = L, U$ depend on the support of the L_i's denoted by supp(L) or the one of the U_i's denoted by supp(U): they are chosen such that $I_Z \subset \text{supp}(Z)$ for $Z = L, U$. The estimation spaces are thus $\Sigma_m(I_Z)$, and the basis should inherit from the same index, but it is omitted for the sake of readability. For the Laguerre basis, the support of the basis is fixed, $I_Z = \mathbb{R}^+$.

We can prove the following results, for the two estimators $\widehat{S}_m^{(L)}$ and $\widehat{S}_m^{(U)}$.

Proposition 1. *For $Z = L, U$, assume that ${}^t\Phi_m^{(Z)}\Phi_m^{(Z)}$ is invertible almost surely. Let $\widehat{S}_m^{(Z)}$ be the estimator of S on I_Z defined by coefficients $\vec{\hat{a}}_m^{(Z)}$ in the basis $\varphi_0, \ldots, \varphi_{m-1}$ as given by (9). Then denoting by $S_I = S\mathbb{1}_I$, we have*

$$\mathbb{E}(\|\widehat{S}_m^{(Z)} - S_{I_Z}\|_{n,Z}^2) \le \inf_{t\in\Sigma_m(I_Z)} \|t - S_{I_Z}\|_Z^2 + \frac{1}{4}\frac{m}{n}.$$

Remark 1. Note that $\int_{I_Z} S^2(x)f_Z(x)dx < +\infty$, and that $\inf_{t\in\Sigma_m(I_Z)} \|t - S_{I_Z}\|_Z^2 = \|S_m^{(Z)} - S_{I_Z}\|_Z^2$ where $S_m^{(Z)}$ is the orthogonal projection of S on $\Sigma_m(I_Z)$ with respect to the scalar product $\langle \cdot, \cdot\rangle_Z$ where $Z = L, U$. If moreover S is square-integrable on $I_Z \cap \text{supp}(Z)$ and f_L and f_U are upper bounded, by $f_{\max}^{(L)}$ and $f_{\max}^{(U)}$ respectively, we can recover a standard (non-weighted) $\mathbb{L}^2$-norm on $I_Z \cap \text{supp}(Z)$ and get, for the bias term

$$\inf_{t\in\Sigma_m} \|t - S_{I_Z}\|_Z^2 \le f_{\max}^{(Z)} \inf_{t\in\Sigma_m} \|(t - S_{I_Z})\mathbf{1}_{\text{supp}(Z)}\|^2 \le f_{\max}^{(Z)} \|S_m^{(Z)} - S_{I_Z}\|^2, \quad Z = L, U,$$

where $S_m^{(Z)}$ is the standard orthogonal projection of S_{I_Z} on $\Sigma_m(I_Z)$, $S_m^{(Z)} = \sum_{j=0}^{m-1}\langle S, \varphi_j\rangle\varphi_j$.

Let us also briefly discuss about the invertibility assumption. First, in the case of the histogram basis, the matrix ${}^t\Phi_m^{(Z)}\Phi_m^{(Z)}$ is diagonal (indeed in that case, $\varphi_j\varphi_k \equiv 0$ for $j \neq k$). For $I_Z = [0,1]$, it is thus invertible if no bin $[j/m,(j+1)/m[$ is empty, and then explicit formula for the coefficients is available (see Section 2.7). Now, if a bin is empty, the estimator still can be defined by choosing any value on this interval: this will not have any impact on the empirical norm of the estimator, which relies on observed values only. In the case of the Laguerre basis, it is easy to see that the matrix ${}^t\Phi_m^{(Z)}\Phi_m^{(Z)}$ is a.s. invertible, as soon as $m \ge n$.

Moreover, for all bases, $\widehat{\Psi}_{m,Z}$ tends to $\Psi_{m,Z}$ almost surely, for $Z = L, U$, when n tends to infinity. We noticed that ${}^t\vec{a}\,\Psi_{m,Z}\vec{a} = \|t\|_Z^2$ where $t = \sum_{j=0}^{m-1} a_j\varphi_j$, for $Z = L, U$. Assume that I_Z is compact and f_Z is lower bounded on I_Z by $f_0^{(Z)}$. Then, for $t \neq 0$, $\|t\|_Z^2 \ge f_0^{(Z)}\|t\|^2 > 0$, $Z = L, U$. Therefore, $\Psi_{m,Z}$ is invertible, which heuristically means that ${}^t\Phi_m^{(Z)}\Phi_m^{(Z)}$ is "asymptotically" invertible.

Now, by using this strategy, we can see that we take separately two parts of the available information while we would like to take it completely. Moreover, the estimators will clearly perform well, but only either on the support of L or on the one of U, and not on both, see Figure 1.

2.4. **Improved estimator.** Here, we explain our further investigations in order to obtain an estimator on a larger interval, in better accordance with all available data.

2.4.1. *First step: estimator of differences.* For $T(x,y) = \sum_{1\le j,k\le m} a_{j,k}\varphi_j(x)\varphi_k(y)$ belonging to $\Sigma_m \otimes \Sigma_m$, we may also consider, as $\mathbb{E}(\mathbb{1}_{\delta_i=0}|U_i, L_i) = S(L_i) - S(U_i)$, the contrast

$$\frac{1}{n}\sum_{i=1}^n T^2(L_i, U_i) - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=0} T(L_i, U_i).$$

In that way, we would take all the observations into account. However, the resulting estimator would provide an estimator of the bi-variate function $G(x,y) = S(x) - S(y), x < y$, without taking its specific form into account: the underlying function is $S(.)$ and it is univariate. However, due to the curse of dimensionality, the rate associated to the bidimensional problem would be bad, or at least worse than what we can expect for a univariate function. Now inserting in addition the specific form of G , we obtain

$$\tilde{\gamma}_n(t) = \frac{1}{n}\sum_{i=1}^n [t(L_i) - t(U_i)]^2 - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=0}[t(L_i) - t(U_i)].$$

This contrast has expectation

$$\mathbb{E}[\tilde{\gamma}_n(t)] = \iint [t(x) - t(y) - (S(x) - S(y))]^2 f_{(L,U)}(x,y)dxdy - \iint (S(x) - S(y))^2 f_{(L,U)}(x,y)dxdy.$$

Here, we estimate m coefficients, which may be relevant to recover S, except that the function is determined up to, at least, an additive constant. Now, the expectation can be re-written:

$$\begin{aligned}\mathbb{E}[\tilde{\gamma}_n(t)] &= \|t-S\|_L^2 + \|t-S\|_U^2 - 2\iint (t-S)(x)(t-S)(y) f_{(L,U)}(x,y)dxdy \\ &\quad - \iint (S(x)-S(y))^2 f_{(L,U)}(x,y)dxdy.\end{aligned}$$

The first two right-hand-side terms ($\|t-S_I\|_L^2 + \|t-S_I\|_U^2$) correspond to norms that we intend to simultaneously minimize, with $I \supseteq I_L \cup I_U$. The last term does not depend on the function t and can be omitted. This is why we tried to kill the third term, of cross-product type. Noticing that

$$\iint (t-S)(x)(t-S)(y) f_{(L,U)}(x,y)dxdy = \mathbb{E}[(t-S)(L_1)(t-S)(U_1)]$$

and that, by conditioning by (U_i, L_i), we have

$$\mathbb{E}[(t(U_i) - \mathbb{1}_{\delta_i=1})(t(L_i) - \mathbb{1}_{\delta_i\neq -1})] = \mathbb{E}[(t(U_i) - S(U_i))(t(L_i) - S(L_i))] + \underbrace{\mathbb{E}[S(U_i)(1-S(L_i))]}_{\text{independent of function } t}$$

we obtain an adequate term to add to the previous contrast.

2.4.2. *New estimator.* Thus, we corrected the contrast by replacing $\tilde{\gamma}_n(t)$ by

$$\frac{1}{n}\sum_{i=1}^n [t(L_i) - t(U_i)]^2 - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=0}[t(L_i) - t(U_i)] + \frac{2}{n}\sum_{i=1}^n (t(U_i) - \mathbb{1}_{\delta_i=1})(t(L_i) - \mathbb{1}_{\delta_i\neq -1}).$$

This formula can be rewritten

$$\|t\|_{n,U}^2 + \|t\|_{n,L}^2 - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=1} t(U_i) - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i\neq -1} t(L_i) + \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=1}.$$

This is how we obtained our main contrast:

$$\gamma_n(t) = \|t\|_{n,U}^2 + \|t\|_{n,L}^2 - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i=1} t(U_i) - \frac{2}{n}\sum_{i=1}^n \mathbb{1}_{\delta_i\neq -1} t(L_i). \tag{10}$$

where $\|t\|_{n,U}^2$ and $\|t\|_{n,L}^2$ are defined by (5). We note that this contrast appears as the sum of the two previous ones and we straightforwardly obtain the following result.

Proposition 2. *Using the norms defined in (4) and (5), we have*

$$\mathbb{E}(\gamma_n(t)) = \|t-S\|_U^2 + \|t-S\|_L^2 - \|S\|_U^2 - \|S\|_L^2.$$

We set

$$\|t-S\|_U^2 + \|t-S\|_L^2 = \int (t(x)-S(x))^2 (f_L(x) + f_U(x))dx := \|t-S\|_{L+U}^2.$$

Thus we define our final estimator by

$$\widehat{S}_m = \arg\min_{t\in\Sigma_m(I)} \gamma_n(t). \tag{11}$$

Assuming that ${}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}$ is invertible, then the estimator can be computed as

$$\widehat{S}_m = \sum_{j=1}^m \hat{a}_j\varphi_j \quad \vec{\hat{a}}_m = \begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_{m-1}\end{pmatrix} = \left[{}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}\right]^{-1}\left({}^t\Phi_m^{(L)}\vec{\delta}^{(L)} + {}^t\Phi_m^{(U)}\vec{\delta}^{(U)}\right),$$

where $\Phi_m^{(Z)}$, $\vec{\delta}^{(Z)}$ are defined in (6). This formula shows that the estimator uses all the data and is not the sum of the first two estimators. Note that ${}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}$ is in particular invertible when both ${}^t\Phi_m^{(L)}\Phi_m^{(L)}$ and ${}^t\Phi_m^{(U)}\Phi_m^{(U)}$ are, and this has already been discussed (and holds true for Laguerre basis and $m \leq n$).

The study of the estimator is more tedious, but it is interesting to see that we can prove the following result.

Proposition 3. *Assume that* ${}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}$ *is invertible almost surely. Then, for any* $m \in \{1, \dots, n\}$, *we have*

$$\mathbb{E}\left(\|\widehat{S}_m - S_I\|_{n,U}^2 + \|\widehat{S}_m - S_I\|_{n,L}^2\right) \leq \inf_{t \in \Sigma_m(I)}\left(\|t - S_I\|_{L+U}^2\right) + \frac{5}{2}\frac{m}{n}.$$

We already noticed that $\int_I S^2(x)(f_L(x) + f_U(x))dx < +\infty$. Note that the bias is now

$$\inf_{t \in \Sigma_m(I)}\left(\|t - S_I\|_U^2 + \|t - S_I\|_L^2\right) = \int (S_m(x) - S_I(x))^2(f_L(x) + f_U(x))dx$$

where S_m is the orthogonal projection of S on $\Sigma_m(I)$ with respect to the scalar product $\langle .,. \rangle_L + \langle .,. \rangle_U$. Following Remark 1, if f_L and f_U are bounded by $f_{\max}$, we get

$$\inf_{t \in \Sigma_m(I)}\left(\|t - S_I\|_U^2 + \|t - S_I\|_L^2\right) \leq 2f_{\max}\|S_m - S_I\|^2.$$

Proposition 3 shows that the risk of the estimator $\widehat{S}_m$ is bounded by a squared bias term $\inf_{t \in \Sigma_m(I)}\left(\|t - S_I\|_{L+U}^2\right)$, which decreases when m increases, plus a variance term $(5/2)(m/n)$ increasing with m. A data-driven procedure leading to an adequate compromise is proposed in Section 3.

2.5. **Discussion about separated and non separated cases.** Due to our specific empirical norm a.s. converging to $\|.\|_{L+U}$, when n tends to infinity, our estimator $\widehat{S}_m$ is expected to perform well on $(\text{supp}(L) \cup \text{supp}(U))$: this is due to the weight function $f_L + f_U$.

This means that if the basis is compactly supported, its support $I \supseteq I_L \cup I_U$ is in fact chosen by the user, and must be taken in accordance. Of course, it is larger than in the two naive strategies.

For non compactly supported bases such as Laguerre, no support has to be chosen *a priori* for the computation of the coefficients of the estimator. However, in the simulations, the errors of the estimator are computed on an interval $[\min(L_i), \max(U_i)]$ according to the range of the observations.

Here we want to underline that the separated/non separated cases are behind these support considerations. In the separated case, there is a hole between the supports of L and U, and thus a compact set of estimation containing both supports will lead, in the histogram case, to empty bins (see Section 2.7). Therefore the diagonal matrix ${}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}$ fails to be invertible: only the non-separated case should be considered for such a basis. Analogously, Birgé (1999) works also in the non-separated case, as he assumes that the joint distribution of (L_1, U_1) is lower bounded on the diagonal. This problem does not appear for the compactly supported trigonometric basis, nor for the global Laguerre basis (in these two cases, the matrix can be proved to be a.s. invertible).

Moreover, for Laguerre basis, even if the estimation interval is $I = \mathbb{R}^+$, the risk is controlled on $\text{supp}(L) \cup \text{supp}(U)$ thanks to the weight function $f_L + f_U$. In practice, when considering this basis, a hole between the supports of L and U does not imply any practical problem in the procedure (see Figure 1). Heuristically, this means that in the separated case, the procedure automatically extrapolates the estimator on the hole, from what has been built on each support of U and L. This is coherent if the function has indeed a global regularity on the whole nonnegative real line.

2.6. **Discussion about rates.** Inequalities provided in Proposition 1 and Proposition 3 can allow to compute convergence rates of the estimators.

• Consider the compactly supported bases described in Section 2.1 (such as histograms or piecewise polynomials). Assume that f_L and f_U are bounded on the support of the basis. The results stated in Brunel and Comte (2009), Corollary 3.1 p.8, apply here. They imply that the method provides convergent estimators $\widehat{S}^{(Z)}_{m^{(Z)}_{\rm opt}}$, with asymptotic rate $n^{-2\alpha/(2\alpha+1)}$ for α the Besov regularity of S_{I_Z} when S_{I_Z} belongs to a Besov ball, and $m^{(Z)}_{\rm opt} = O(n^{1/(2\alpha+1)})$ for $Z = L, U$, and the same holds for $\widehat{S}_m$ on I. Those rates constitute a generalization of the rate $n^{-1/3}$ corresponding to $\alpha = 1$ (rates obtained under Lipschitz type assumptions, e.g. in Birgé (1999)) to rates of order $n^{-\alpha/(2\alpha+1)}$ for a general regularity α which can be larger than one for trigonometric bases or piecewise polynomials with degree $r \geq \alpha$. As already mentioned this rate can be improved rate within a logarithmic factor, under a set of assumptions including a condition of non separation on the joint distribution $f_{(L,U)}$: we do not obtain this improvement.

• For $s \geq 0$, Bongioanni and Torrea (2009) defined Sobolev-Laguerre spaces with index s by:

$$W^s = \{\theta : \mathbb{R}^+ \to \mathbb{R},\ \theta \in \mathbb{L}^2(\mathbb{R}^+), |\theta|_s^2 := \sum_{k\geq 0} k^s a_k^2(\theta) < +\infty\}.$$

where $a_k(\theta) = \int_{\mathbb{R}^+} \theta(x)\ell_k(x)dx$. We define the ball $W^s(D)$ by

$$W^s(D) = \left\{\theta \in W^s_L, |\theta|_s^2 = \sum_{k=0}^{\infty} k^s a_k^2(\theta) \leq D\right\}.$$

For details on these spaces, and especially for regularity properties of functions in these spaces, see Comte et al. (2015), Section 7.2.

Heuristically, the index α for Besov spaces and s for Laguerre Sobolev spaces correspond to the regularity order of the function on the domain (a bounded interval in the first case and $\mathbb{R}^+$ in the second). For instance, the survival function associated with a $\beta(a, b)$ distribution has infinite regularity on any interval included in $]0, 1[$ and regularity b on e.g. $[0, 2]$ or $\mathbb{R}^+$.

Now, if f_L and f_U are upper bounded on $I = \mathbb{R}^+$ and S_I belongs to $W^s(D)$, then the risks of $\widehat{S}^{(Z)}_m$ for $Z = L, U$ and of $\widehat{S}_m$ can be bounded by $Dm^{-s} + cst\, m/n$, see Proposition 1 ($cst = 1/4$) and Proposition 3 ($cst = 5/2$). Thus, choosing m of order $n^{1/(s+1)}$ yields a risk less than $n^{-s/(s+1)}$ and the estimators are therefore convergent.

Note that the choices of m proposed in this section are theoretical and not possible in practice as they depend on unknown regularity parameters.

2.7. **Histogram case.** In the specific case of histogram basis, the matrices $\Psi_{m,Z}$, $Z = L, U$, are diagonal. Thus, invertibility conditions are easy to study and explicit formulas for the estimators can be given.

We take in this section $\varphi_j = h_j$ for $j = 0, \ldots, m-1$, see Section 2.1. We define, the following cardinalities:

$$N_j := \mathrm{Card}\{i \in \{1, \ldots, n\}, L_i \in I_j\}, M_j := \mathrm{Card}\{i \in \{1, \ldots, n\}, U_i \in I_j\},$$

and

$N'_j = \text{Card}\{i \in \{1,\dots,n\}, L_i \in I_j \text{ and } \delta_i = -1\}, M'_j = \text{Card}\{i \in \{1,\dots,n\}, U_i \in I_j \text{ and } \delta_i = 1\}.$

Then $\langle \varphi_j, \varphi_k \rangle_{n,U} = 0$ if $j \neq k$ and $(m/n)M_j$ if $j = k$. So $\widehat{\Psi}_{m,U} = (m/n)\text{diag}(M_1,\dots,M_m)$. Analogously $\widehat{\Psi}_{m,L} = (m/n)\text{diag}(N_1,\dots,N_m)$. They are invertible if no M_j is null for the first one, no N_j is null for the second one. The estimator $\widehat{S}_m$ relies on the inversion of $\widehat{\Psi}_{m,L} + \widehat{\Psi}_{m,U}$ and the matrix is therefore invertible if M_j and N_j are never simultaneously null. We obtain

$$\hat{a}_j^{(L)} = \frac{1}{\sqrt{m}} \frac{N_j - N'_j}{N_j} = \frac{1}{\sqrt{m}} \left(1 - \frac{N'_j}{N_j}\right), \quad \hat{a}_j^{(U)} = \frac{1}{\sqrt{m}} \frac{M'_j}{M_j}, \quad \hat{a}_j = \frac{1}{\sqrt{m}} \frac{N_j - N'_j + M'_j}{M_j + N_j}.$$

It is worth underlining that the last estimator in not the sum of the estimators $\widehat{S}_m^{(L)}$ and $\widehat{S}_m^{(U)}$ and is different from Birgé's proposal.

3. Model selection

The choice of m as discussed in Section 2.6 is not possible as it is asymptotic, and depends on the unknown regularity order (α or s). Instead, we propose a finite sample strategy leading to a bias-variance compromise.

We proceed with a model selection strategy for the estimator $\widehat{S}_m$, that is a data driven way of selecting m from the data in a coherent way. Part of the tools we use here are inspired from the work on standard regression function estimation developed in Comte and Genon-Catalot (2019). They allow us to provide a generalization of the method presented in Brunel and Comte (2009) for interval censoring case 1, and dedicated to compactly supported bases. Note that a similar procedure would be possible for $\widehat{S}_m^{(Z)}$, $Z = L, U$, we experiment it numerically in section 4, but do not give theoretical details.

To take into account both compactly and non compactly supported bases, we define the random collection of models as follows:

$$\widehat{\mathcal{M}}_n = \left\{ m \in \mathbb{N} \setminus \{0\}, m(\|(\widehat{\Psi}_{m,L} + \widehat{\Psi}_{m,U})^{-1}\|_{\text{op}}^2 \vee 1) \leq 4\mathfrak{c} \frac{n}{\log(n)} \right\}, \tag{12}$$

where $\widehat{\Psi}_{m,Z}$ for $Z = L, U$ is defined by (7) and

$$\mathfrak{c} = \left(6 \wedge \frac{1}{\|f_L + f_U\|_\infty}\right) \frac{1}{48 c_\varphi^2},$$

with c_φ defined in (3). The theoretical (deterministic) counterpart of this random set is denoted by $\mathcal{M}_n$ and defined by

$$\mathcal{M}_n = \left\{ m \in \mathbb{N} \setminus \{0\}, m(\|(\Psi_{m,L} + \Psi_{m,U})^{-1}\|_{\text{op}}^2 \vee 1) \leq \mathfrak{c} \frac{n}{\log(n)} \right\}. \tag{13}$$

We propose to select the model following the rule:

$$\hat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_n} [\gamma_n(\widehat{S}_m) + \text{pen}(m)] \quad \text{with} \quad \text{pen}(m) = \kappa \frac{m}{n}, \tag{14}$$

where κ is a numerical constant. Relying on results stated in Comte and Genon-Catalot (2019), we can obtain the following result.

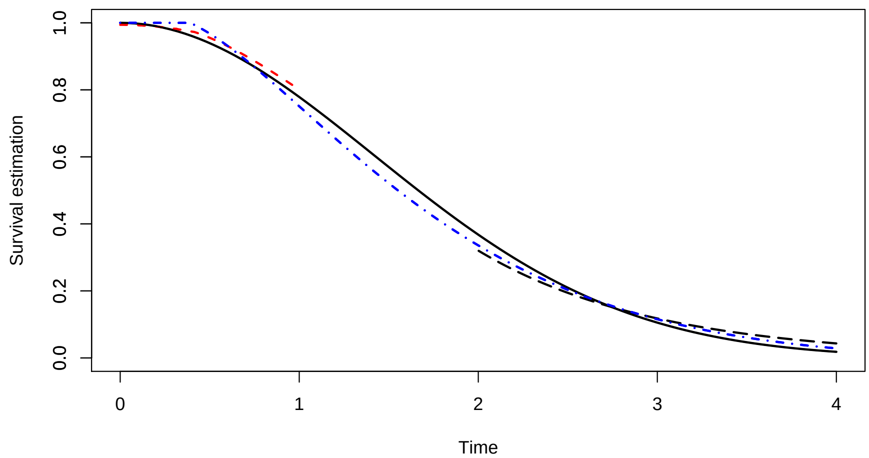


FIGURE 1. True survival curve (black solid line) S_I of a $\mathcal{W}eibull(2,2)$ distribution and Laguerre basis estimators with sample size $n = 1000$: $\widehat{S}^{(L)}_{\hat{m}_L}$ (red dashed line) built on supp(L), $\widehat{S}^{(U)}_{\hat{m}_U}$ (black longdashed line) built on supp(U) and $\widehat{S}_{\hat{m}}$ (blue dotdashed line) built on $I = [0, 4]$ for scenario 4.

Theorem 1. *Assume that $f_U + f_L$ is bounded. Consider a nested collection of models $(\Sigma_m)_{m\in\mathcal{M}_n}$ with models satisfying (3), $\mathcal{M}_n$ defined by (13), and the estimator defined by (10)-(11) and (14). Then there exists a value $\kappa_0 > 0$, such that $\forall\kappa \geq \kappa_0$,*

$$\mathbb{E}\big[\|\widehat{S}_{\hat{m}} - S_I\|^2_{L+U}\big] \leq C \inf_{m\in\mathcal{M}_n} \left(\inf_{t\in\Sigma_m} \|S_I - t\|^2_{L+U} + \kappa\frac{m}{n} \right) + \frac{C'}{n}$$

where C is a numerical constant and C' is a constant depending on f_L, f_U, $\mathfrak{c}$.

The procedure is well defined as soon as κ is fixed. Theorem 1 states that there exists a suitable value κ_0. Larger values may also be used, but the larger κ, the larger the risk bound. Thus, we should choose the smallest admissible value. Moreover, Theorem 1 also implies that the value κ_0 should be the same for the general problem, but the theoretical value of κ_0 obtained in the proofs is too large in practice. This is why we have to find a more relevant all-in-one value: it is standard to calibrate the value of κ through preliminary simulation experiments and we explain in the next section how we proceed. Then, it is definitely fixed to this value.

The inequality stated in Theorem 1 shows that the estimator makes an automatic bias-variance tradeoff, with a data driven selection criterion. The performance of $\widehat{S}_{\hat{m}}$ is valid on an interval which is larger than if $\widehat{S}^{(Z)}_m$ had been considered, for $Z = L$ or U. The loss of the procedure lies in the multiplicative constants C (the nearer of 1, the better), and in the restriction on the collection given in (12) and (13), which must let the optimal choice reachable. We discuss this in the next remark.

Remark 2. If the basis has compact support I and f_U and f_L are lower bounded on I by f_0, then we can prove

$$\max(\|\Psi^{-1}_{m,U}\|^2_{\text{op}}, \|\Psi^{-1}_{m,L}\|^2_{\text{op}}) \leq 1/f_0^2 \quad \text{and} \quad \|(\Psi_{m,L} + \Psi_{m,U})^{-1}\|^2_{\text{op}} \leq 4/f_0^2. \tag{15}$$

Thus, it turns out that the collection $\mathcal{M}_n$ defined by (13) is simply the set of models m such that $m \leq Cn/\log(n)$, which is a very weak constraint. This implies that the adaptive estimator automatically reaches the best possible rate on Besov-type regularity spaces (see the end of Sections 2.3 and 2.4) on the domain determined by the support of the basis.

In the case of non compactly supported bases, such as the Laguerre basis which works very well for survival function estimation, condition in (13) imposes a real restriction on the collection of models. For optimality issues, theoretical examples and illustrations, we refer to Comte and Genon-Catalot (2019).

4. Simulation study

Our aim is to compare our new penalized estimator, built using the Laguerre basis defined by (2), with other available competitors. We consider the log-concave Nonparametric Maximum Likelihood Estimator (NPMLE) of Anderson-Bergman and Yu (2016) implemented using the **logconPH** R package and the unconstrained NPMLE implemented using the **prodlim** R package. We also consider the smoothed maximum likelihood estimator (SMLE) obtained by convolving the NPMLE with a smoothing kernel as in Groeneboom and Ketelaars (2011).

4.1. **Models and censoring schemes.** We simulated $K = 100$ samples of size $n = 100$, 300 and 1000 from the following event time distributions :

- Model 1 : $X \sim \text{Log-}\mathcal{N}(0, 1)$ the survival function is $S(x) = 0.5(1 - \text{erf}(\log(x)/\sqrt{2}))$ with $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$.
- Model 2 : $X \sim \mathcal{W}eibull(a, b)$ with shape parameter $a = 0.5$ and scale parameter $b = 2$ corresponding to a non log-concave distribution.
- Model 3 : X is distributed as a $\mathcal{B}eta'(\alpha, \beta)$ a beta prime distribution or a beta of type II with survival function $S(x) = \int_x^{+\infty} u^{\alpha-1}(1 + u)^{-\alpha-\beta}/(B(\alpha, \beta))du$ for $x \geq 0$ where $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1 - t)^{\beta-1} dt$ is the beta function with $\alpha = 5$ and $\beta = 2$ two shape parameters.
- Model 4 : $X = 6Z$ with $Z \sim \mathcal{B}eta(2, 5)$ a standard beta distribution admitting the density function $f(x) = \Gamma(a + b)/(\Gamma(a)\Gamma(b))x^{a-1}(1 - x)^{b-1}$ for $0 \leq x \leq 1$ with shape parameters $a = 2$ and $b = 5$.

Note that Model 1 and 4 correspond to log-concave distributions while Model 2 and 3 do not. We also investigate different schemes for the distribution of the inspection times L and U :

- scenario 1 : $L \sim \mathcal{U}([0, 2.5])$ and $U = \mathcal{U}([3, 4])$.
- scenario 2 : $L \sim \mathcal{U}([0, 1])$ and $U = L + \mathcal{U}([0, 3])$.
- scenario 3 : $L, U \sim \mathcal{U}([0, 4])$ with the constraint $0 \leq U - L \leq 0.1$ so that the times L and U can be very close to each other.
- scenario 4 $L \sim \mathcal{U}([0, 1])$ and $U \sim \mathcal{U}([2, 4])$.

We took $K = 100$ to save computing time, but we checked the consistency of our results: we experimented on part of our models the larger value $K = 500$, and observed that the numerical results were not drastically different (see Table 3 in appendix B) and the conclusions about the performances of the different competitors remained the same. In scenarios 1 and 4, there is a hole between the supports of L and U. These scenarios make sense in the context of diseases with a long-distance follow-up care. On the opposite, the supports of L and U overlap in scenarios 2 and 3.

We illustrate how model selection performs for histogram and Laguerre bases on Figure 2. We can see the selected estimator among 10 estimators of the collection: on Figure 2 (left/right respectively) the selected dimension is $\hat{m} = 6$ for histogram basis while it is $\hat{m} = 2$ for Laguerre

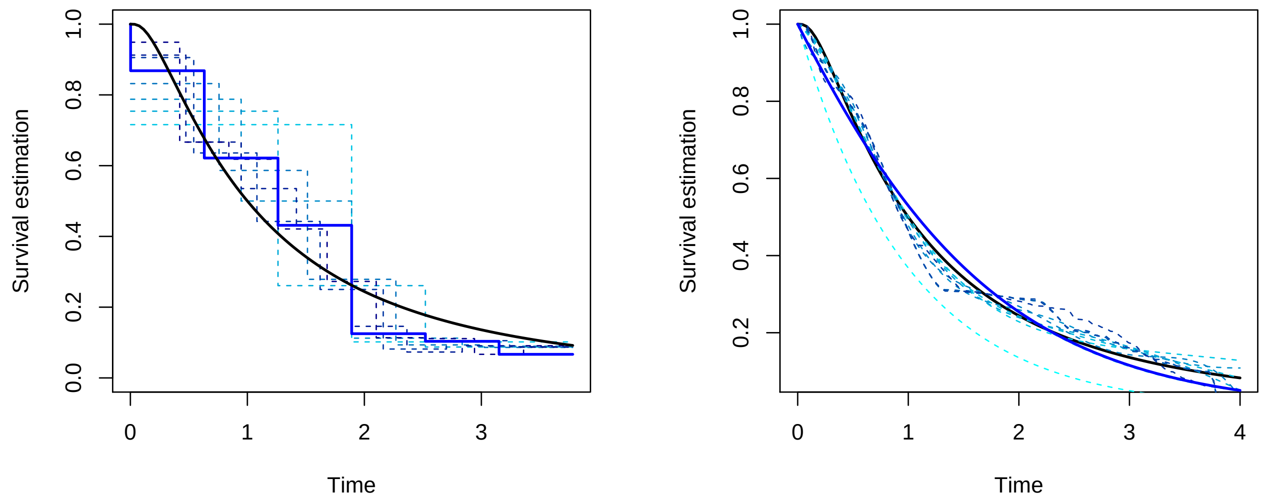


FIGURE 2. Model selection for mean squares estimators : Collection of estimators $\widehat{S}_m^{(Z)}$ for dimension $m = 1, \cdots, 10$ (from cyan to dark blue dotted line) and selected estimator (blue solid line) with $\hat{m} = 6$ for histogram basis (left) and with $\hat{m} = 2$ for Laguerre basis (right) both for Model 1, scenario 2 and sample size $n = 300$. True survival curve S_I (black solid line).

basis. But, in the sequel we choose to compare with competitors the projection estimator in the Laguerre basis only. Indeed, as mentioned in Section 2.7, histogram estimators are not well-defined on empty bins and additional conventions should be especially needed to build them in scenario 1 and 4. Nevertheless, histogram estimators behave well on an estimation interval without empty bins as shown in Figure 2 (left). Note that we do not apply the *a posteriori* monotonization procedure for histogram estimators but the selected estimator for dimension $\hat{m} = 6$ has the adequate monotony property. No correction seems to us neither relevant, nor obvious in that case.

4.2. **About the implementation.** First, we have to fix the value of the constant κ in the penalty term (see (14)). To that aim, we perform preliminary simulation experiments for the calibration of the value of κ over some models, see Appendix A for details. This leads us to choose $\kappa = 0.5$, a value which is then fixed for all the sequel. This is a small value among the range of values we explored. But if large values of κ ensure stability of the estimators (over-penalization), they do not allow to achieve large dimension choices.

To assess the numerical performance of our penalized Least Squares estimator and its competitors, we compute the Mean Integrated Squared Error over a grid. We define a grid $t_1, \cdots, t_J$ of $J = 100$ equispaced points on $I = [\min(L_i), \max(U_i)]$. It is not always possible to evaluate the value of the NPMLE on the right of the interval, as for any product-limit estimator, it is biased and does not go to zero if the greatest observed value of U_i corresponds to $\delta_i = 1$. So we made the choice to shorten the grid at the upper bound of the last step of the NPMLE. Roughly speaking, the grid is shrunken at $\max(U_i, \delta_i = 1)$ instead of $\max(U_i)$. This choice is rather in favour of the NPMLE and SMLE but does not degrade significantly the results for the other

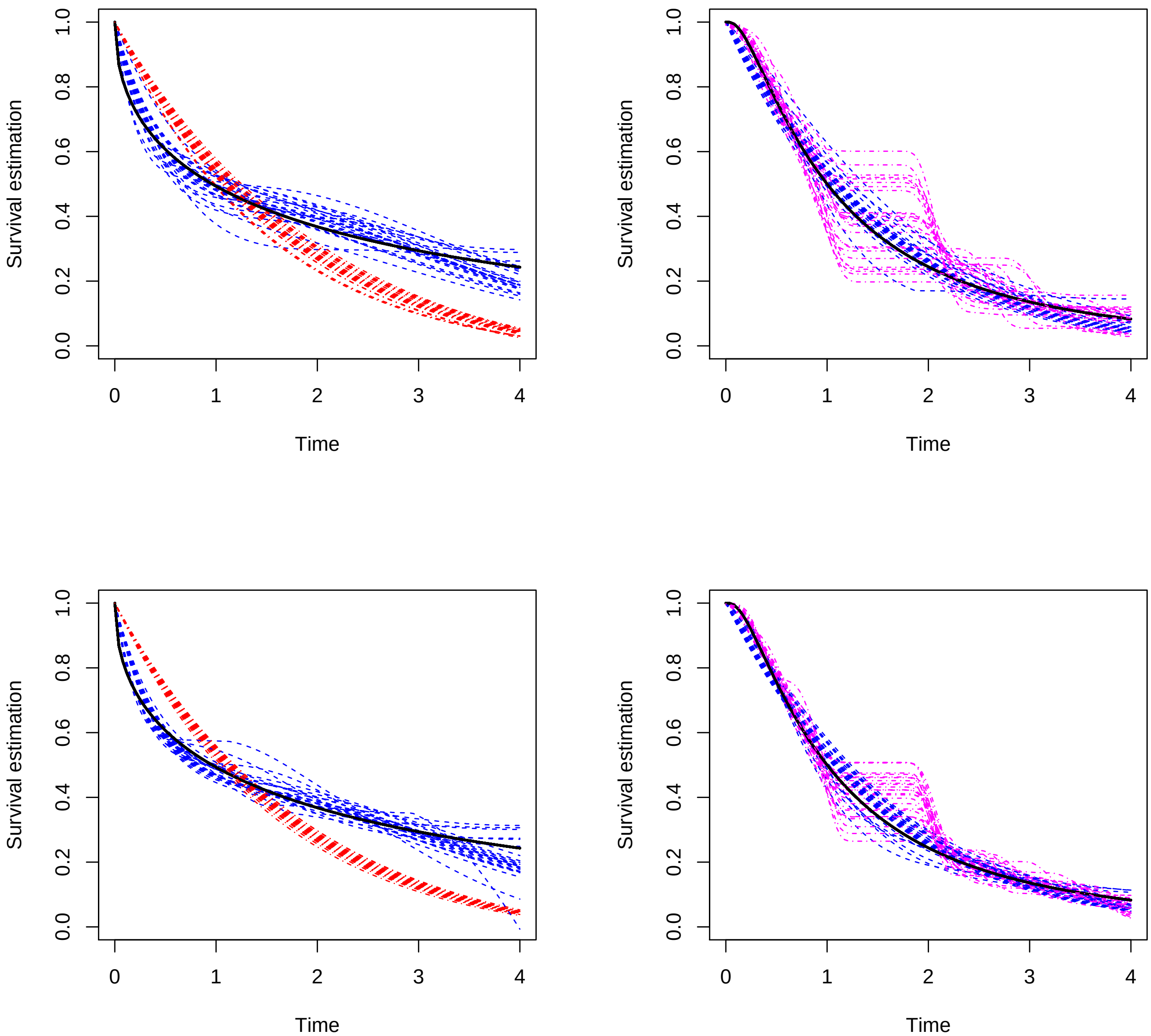
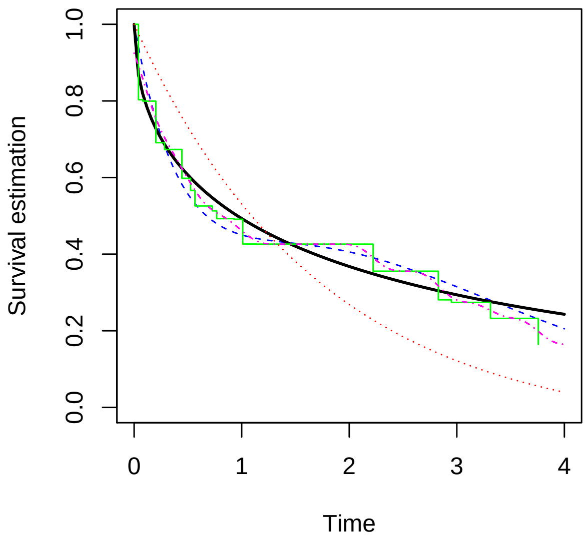


FIGURE 3. True Survival function S_I (black solid line) and bundles of 25 estimators : Anderson-Bergman and Yu estimators (red dot dashed) for Model 2, scenario 2 on the left and Smoothed Maximum Likelihood estimators (magenta dot dashed) for Model 1, scenario 4 on the right and Laguerre basis estimators (blue dashed), for $n = 300$ at the top and $n = 1000$ at the bottom.

estimators as far as we see in preliminary trials. For the SMLE, we take the symmetric triweight kernel and a bandwidth $h_n = n^{-1/5}$ as in Groeneboom and Ketelaars (2011).



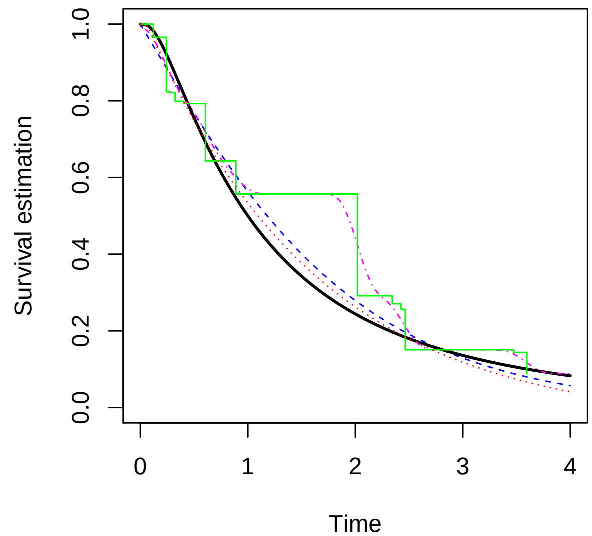


Figure 4. True survival function S_I on $I = [0, 4]$ (black solid line), Anderson-Bergman and Yu estimator (red dotted), our least squares estimators with Laguerre basis (blue dashed), NPMLE estimator (green step line) and SMLE (magenta dot dashed) for Model 2 in scenario 2 on the left and Model 1 in scenario 4 on the right, both for sample size $n = 300$.

We generate $K = 100$ samples for each Model and scenario. The error we computed is defined as follows:

$$\frac{1}{K}\sum_{k=1}^{K}\frac{1}{J}\sum_{j=1}^{J}\left(\widehat{S}^{[k]}(t_j) - S(t_j)\right)^2 \tag{16}$$

where $\widehat{S}^{[k]}$ stands for the estimator of S based on the kth generated sample. This quantity is the MISE we presented in the tables and the boxplots. We might have normalized by the length of the interval but this way of computing is standard and allows comparison between the estimators.

4.3. **Results.** The values of the Mean Integrated Squared Error MISE$\times 10^3$ are presented in Table 1 and 2 in Appendix B, with standard deviation in parenthesis and sample sizes 100, 300 and 1000. The results can be visualized through boxplots given in Figures 5 ($n = 300$) and 6 ($n = 1000$). Model 1 and 4 match with log-concave distributions and as expected the log-concave estimator of Anderson-Bergman and Yu gives the best results whatever the scenario for the inspection times is, but it doesn't work at all for Model 2 and 3. However, our Least Squares estimator challenges the NPMLE and SMLE especially for scenario 3 and 4. Even if it performs a little worse in mean for few models, the Least Squares estimator built with Laguerre basis has no important failure, contrary the other ones. In fact, on Figure 4, we illustrate typical bad behaviours of both constrained/unconstrained NPMLEs: the log-concave estimator is very bad for non log-concave distribution (Figure 4, left) while the NPMLE and SMLE perform badly for

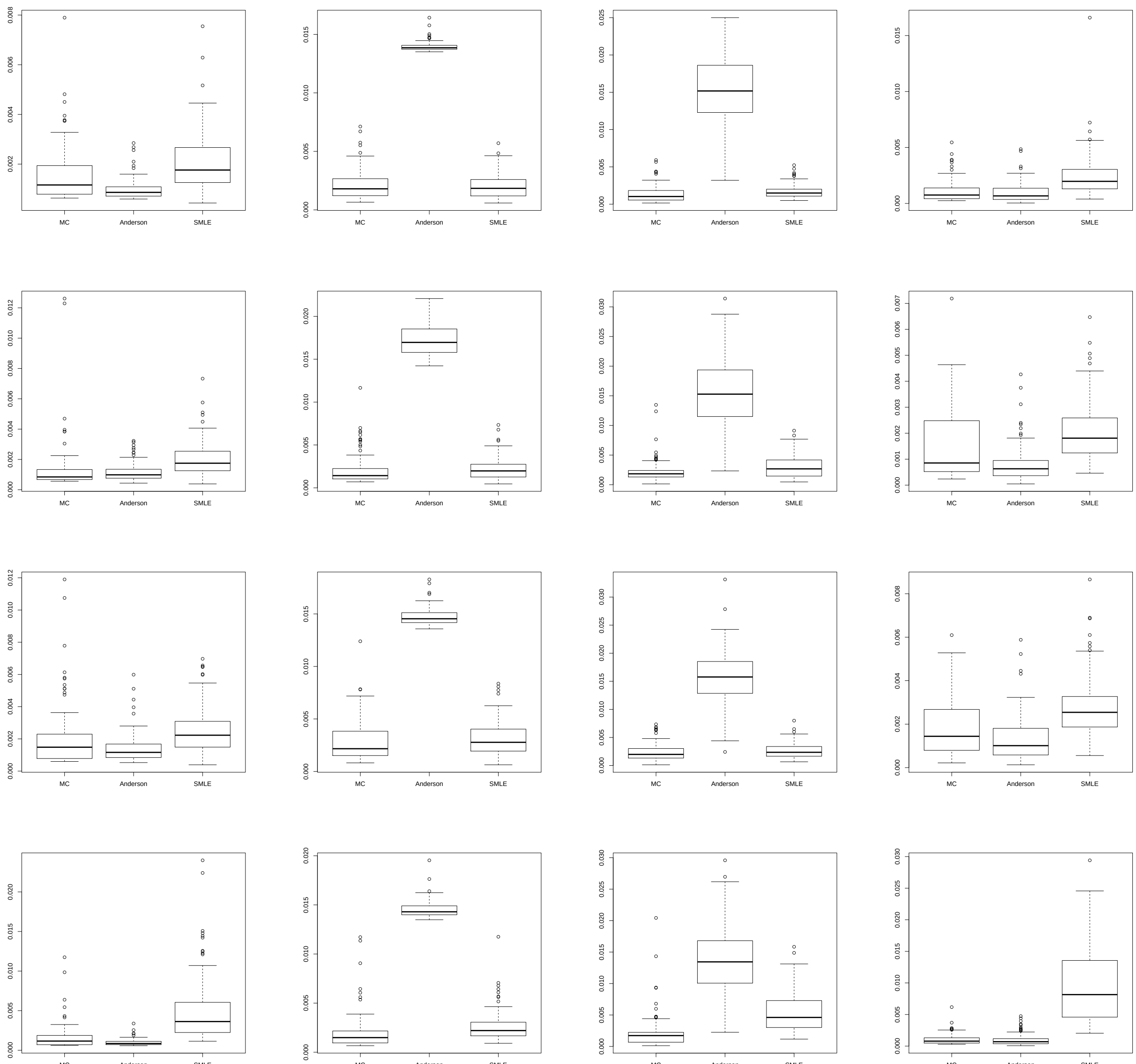


FIGURE 5. MISE for sample size $n = 300$ and $K = 100$ Monte-Carlo replications. From top to bottom scenario 1 to 4, from left to right Model 1 to 4. On each plot, Left: our estimator (MC), Middle: Anderson and Yu estimator (Anderson), Right: SMLE.

scenario 4 (Figure 4, right) when there is a hole between the supports of L and U. So the Least Squares estimator seems to be overall the most reliable. This fact is also illustrated on Figure 3 with bundles of estimators. From Figure 5 and 6, we can see that, except for a small number of extreme error values (which means that model selection failed), the Least Squares estimator appears to be quite a good compromise for any distribution type of the event time and for any support of the inspection times.

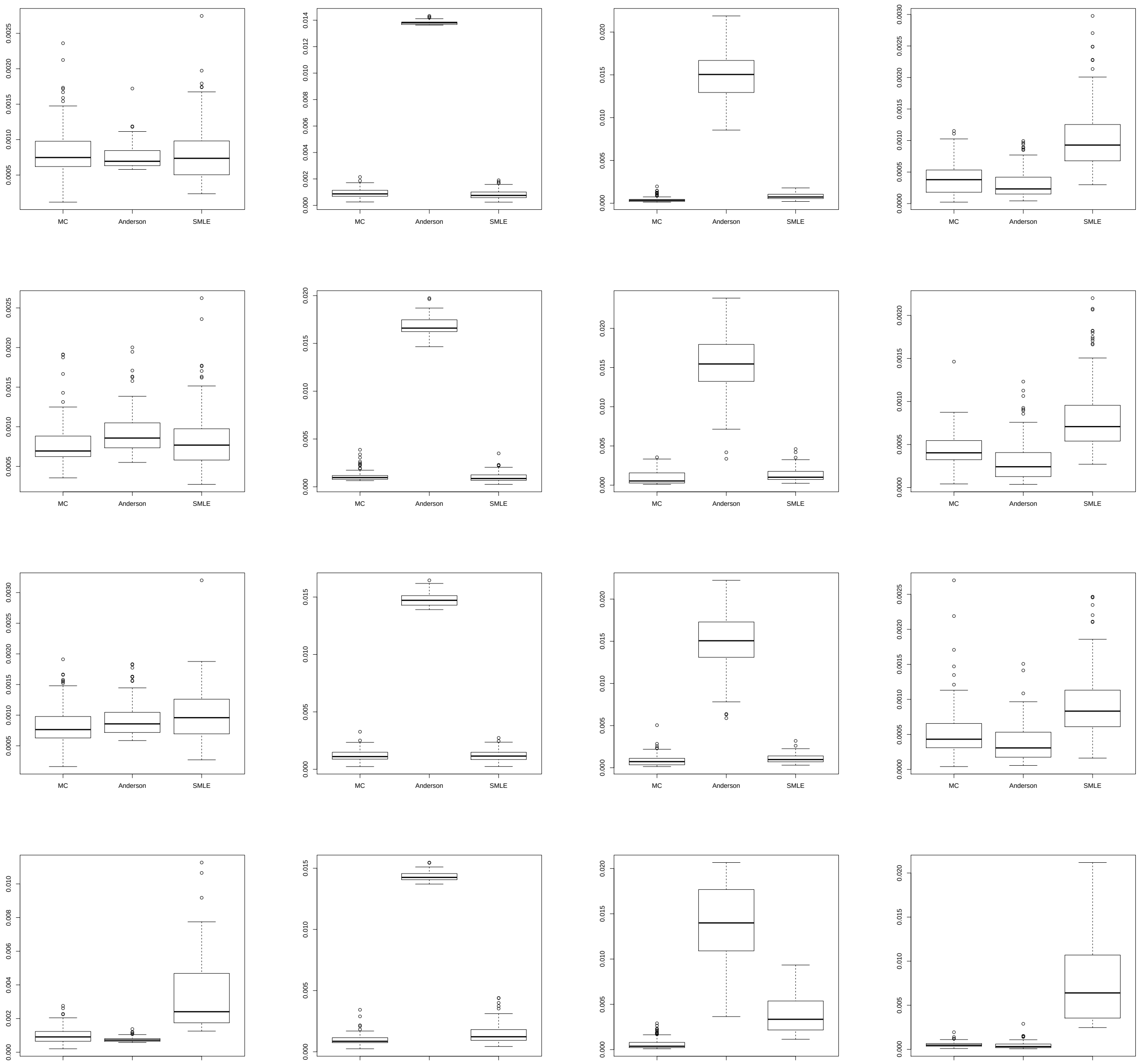


Figure 6. MISE for sample size $n = 1000$ and $K = 100$ Monte-Carlo replications. From top to bottom scenario 1 to 4, from left to right Model 1 to 4. On each plot, Left: our estimator (MC), Middle: Anderson and Yu estimator (Anderson), Right: SMLE.

Remark 3. A drawback of our estimation procedure is that it doesn't build a strict estimator of a survival function. In fact, the penalized estimator may start at a value different than 1 and may fail to be monotone. As it is consistent, this does not happen for large enough sample sizes. However, we propose an *a posteriori* transformation to correct these two facts. We compute first the original penalized estimator $\widehat{S}_{\hat{m}}$. Then, we reevaluate the coefficients of our penalized estimator by adding a constraint in the least squares contrast to make the estimator be equal to 1 at the origin. The constrained least squares contrast can be expressed with the Lagrange multiplier $\gamma_n(t) - \lambda(t(0) - 1)$. From a computational point of view, the procedure is straightforward and leads to a smooth correction of the estimator. We do not investigate the theoretical properties of the resulting constrained estimator, but a study of its properties can be found in another context in Comte and Dion (2017). Finally, the procedure of Chernozhukov et al. (2009) available in the R-package **Rearrangement** allows to overcome the possible problem of monotony and can be applied without degrading the rate of the original estimator. These corrections are applied to the estimators plotted in Figures 1, 3, 4.

5. Application to a real dataset

In this Section we study a dataset from Melbye et al. (1984). In this dataset, a cohort of homosexual men from two cities in Denmark has been examined for HIV-antibody positivity on six different dates: December 1981, April 1982, February 1983, September 1984, April 1987, and May 1989. The dataset comprises a total of 297 people who have been tested at least once. Among all these people, 26 were diagnosed with infection at the first examination date (which corresponds to $\delta_i = -1$), 39 were diagnosed with infection at another examination date (which corresponds to $\delta_i = 0$) and 232 were examined without HIV infection (which corresponds to $\delta_i = 1$). See also Becker and Melbye (1991) and Carstensen (1996) for more informations on the dataset.

Our new estimator with Laguerre basis is applied to the dataset using calendar time as the time scale. In order to deal with the high time values of the dataset which may cause numerical difficulties, we rescale the observations for the estimator computations. The rescaled sample $(L'_i, U'_i)_{1\leq i\leq n}$ is obtained by applying the transformation $t \mapsto (t - \min(L_i))$ to the original data $(L_i, U_i)_{1\leq i\leq n}$. Then, the final curve is plotted in its original scale.

From the collection of models defined in (12), only four different models are allowed. Setting $\kappa = 0.5$ as in the simulation studies, our selection procedure chooses the model $m = 4$. The corresponding estimator is displayed in Figure 7 along with three competitors: the NPMLE implemented from the **prodlim** package, the Anderson-Bergman and Yu estimator implemented from the **logconPH** package and the SMLE from Groeneboom and Ketelaars (2011). As described in Remark 3, only the constrained and monotone version of our estimators are displayed in Figure 7.

We also decided to consider a bootstrap estimation strategy. The idea behind this bootstrap method is to construct an estimator that is robust to model selection. The data $(L_i, U_i, \delta_i)_{i=1,\ldots,n}$ were bootstrapped 5 000 times and for each bootstrap we select a model based on our rule (14) and we compute the corresponding estimator. A final estimator is then computed by taking the median over all 5 000 estimators at each time point. The bootstrap estimator is represented on the left panel of Figure 8 along with the three competitors. This resampling strategy allows us to provide a robust estimator that averages over all possible models instead of choosing only one model. In fact, with this bootstrap estimation method, the model $m = 2$ was chosen in 11% of cases, the model $m = 3$ was chosen in 79% of cases and the model $m = 4$ was chosen in 10% of cases. Finally, resampling the data allows us to construct 95% pointwise confidence intervals by taking the 0.975 and 0.025 empirical quantiles at each time point. Our bootstrap estimate along with its 95% confidence interval is displayed on the right panel of Figure 8.

Our estimators ($\hat{m} = 4$ and the bootstrap) are in accordance with the NPMLE. The bootstrap offers the advantage to provide a smooth and regular estimation of the survival curve. The SMLE is also smooth but contrary to our bootstrap estimator it exhibits a lot of fluctuations located at every step of the NPMLE function. Those variations are due to the construction of the SMLE which is a smooth version of the NPMLE and they probably do not reflect any real variation of the survival function of HIV infection. Finally, the Anderson-Bergman and Yu estimator seems to be biased on these data. For illustration, the chance of being HIV negative among Danish homosexual men in 1986 is estimated to 35.4% from the Anderson-Bergman and Yu estimator, to 78.1% from the NPMLE and SMLE and to 76.3% [73.5%, 84.7%] from our bootstrap estimator (with 95% confidence interval in brackets). The chance of being HIV negative among Danish homosexual men in 1990 is estimated to 2.7% from the Anderson-Bergman and Yu estimator, to 71.7% from the NPMLE and SMLE and to 73.1% [66.1%, 76.1%] from our bootstrap estimator (with 95% confidence interval in brackets). To conclude it seems that the Anderson-Bergman and Yu method is not adapted to this dataset because it implicitly assumes that the time distribution is log-concave while the NPMLE, SMLE and our estimator work for more general survival distributions. Our bootstrap estimator is smooth with less fluctuation than the SMLE. As a result our bootstrap estimator seems to provide the most realistic fit of the data. It also offers the possibility to construct confidence intervals with no additional cost. We emphasize that our estimator is fast to compute and the 5 000 bootstrap samples were computed within only a few seconds. Finally we want to stress that the bootstrapped estimator is not necessarily monotone even though this was the case on this dataset. In order to guarantee the final bootstrapped estimator to be monotone the method of Chernozhukov et al. (2009) would have to be applied.

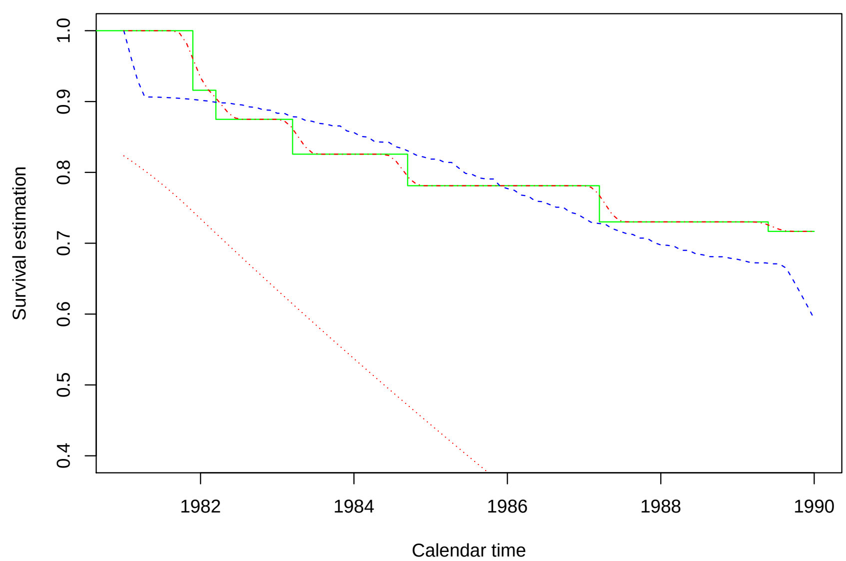


FIGURE 7. Survival estimates of HIV infection using the NPMLE (green solid line), our estimator with Laguerre basis after model selection with $m = 4$ (blue dashed line), the log-concave estimator from Anderson-Bergman and Yu (red dotted line) and the SMLE (red dot dashed line).

6. PROOFS

6.1. **Proof of Proposition 1.** Let $\Pi_m^{(L)}$ denote the orthogonal projection (for the scalar product in $\mathbb{R}^n$) on the subspace $\{\,{}^t(t(L_1), \ldots, t(L_n)), t \in \Sigma_m(I_L)\}$ of $\mathbb{R}^n$ and let $\Pi_m^{(L)} S$ be the projection

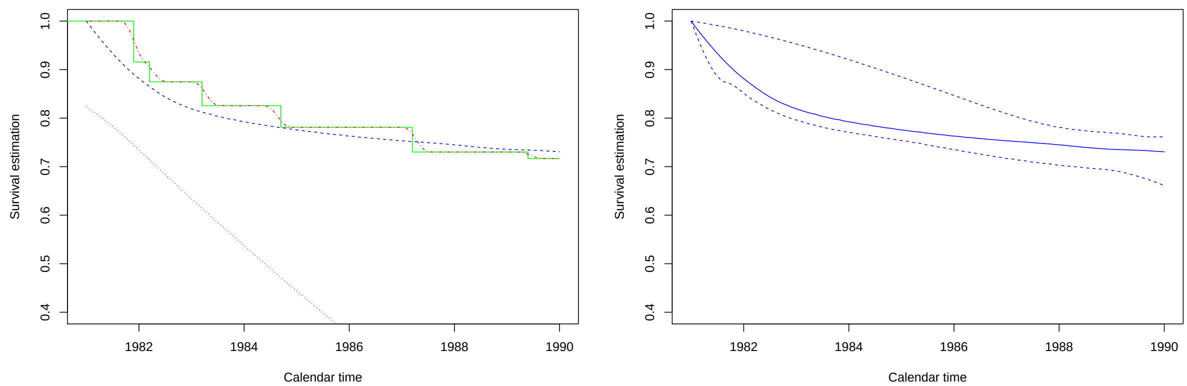


FIGURE 8. Survival estimates of HIV infection. Left panel: the NPMLE (green solid line), our bootstrap estimator with Laguerre basis (blue dashed line), the log-concave estimator from Anderson-Bergman and Yu (red dotted line) and the SMLE (red dot dashed line). Right panel: our bootstrap estimator with Laguerre basis along with its 95% confidence interval.

of ${}^t(S(L_1), \dots, S(L_n))$. Then by Pythagoras,

$$\|\hat{S}_m^{(L)} - S_{I_L}\|_{n,L}^2 = \|\Pi_m^{(L)} S - S_{I_L}\|_{n,L}^2 + \|\hat{S}_m^{(L)} - \Pi_m^{(L)} S\|_{n,L}^2 = \inf_{t \in \Sigma_m(I_L)} \|t - S_{I_L}\|_{n,L}^2 + \|\hat{S}_m^{(L)} - \Pi_m^{(L)} S\|_{n,L}^2.$$

By taking the expectation of the above formula, we have

$$\mathbb{E}[\|\hat{S}_m^{(L)} - S_{I_L}\|_{n,L}^2] \leq \int_{t \in \Sigma_m(I_L)} \|t - S_{I_L}\|_L^2 + \mathbb{E}\left[\|\hat{S}_m^{(L)} - \Pi_m^{(L)} S\|_{n,L}^2\right]. \tag{17}$$

Now, we compute and bound $\mathbb{E}\left[\|\hat{S}_m^{(L)} - \Pi_m^{(L)} S\|_{n,L}^2\right]$. We have

$$\vec{\hat{S}}_m^{(L)} := \begin{pmatrix} \widehat{S}_m^{(L)}(L_1) \\ \vdots \\ \widehat{S}_m^{(L)}(L_n) \end{pmatrix} = \Phi_m^{(L)} \vec{\hat{a}}_m^{(L)} = \Phi_m^{(L)} ({}^t\Phi_m^{(L)} \Phi_m^{(L)})^{-1} \, {}^t\Phi_m^{(L)} \vec{\delta}^{(L)}.$$

We set $\Xi_m^{(L)} = \Phi_m^{(L)} ({}^t\Phi_m^{(L)} \Phi_m^{(L)})^{-1} \, {}^t\Phi_m^{(L)}$ and note that is corresponds to the matrix of the orthogonal projection $\Pi_m^{(L)}$. Therefore

$$\Pi_m^{(L)} S = \Xi_m^{(L)} \mathbf{S}(L) \text{ where } \mathbf{S}(L) = {}^t(S(L_1), \dots, S(L_n)).$$

Therefore, denoting by $\vec{\varepsilon}^{(L)}(L) = {}^t(\varepsilon^{(L)}(L_1), \dots, \varepsilon^{(L)}(L_n))$, where $\varepsilon^{(L)}(L_i) = 1 - \mathbb{1}_{\delta_i = -1} - S(L_i)$, we get

$$\|\hat{S}_m^{(L)} - \Pi_m^{(L)} S\|_{n,L}^2 = \|\Xi_m^{(L)} \vec{\varepsilon}^{(L)}(L)\|_{n,L}^2 = \frac{1}{n} \, {}^t\vec{\varepsilon}^{(L)}(L) \, {}^t\Xi_m^{(L)} \Xi_m^{(L)} \vec{\varepsilon}^{(L)}(L) = \frac{1}{n} \, {}^t\vec{\varepsilon}^{(L)}(L) \Xi_m^{(L)} \vec{\varepsilon}^{(L)}(L).$$

Now,

$$\mathbb{E}\left[\vec{\varepsilon}^{(L)}(L)\Xi_m^{(L)}\vec{\varepsilon}^{(L)}(L)\right] = \sum_{1\le i,k\le n}\mathbb{E}\left(\varepsilon^{(L)}(L_i)\varepsilon^{(L)}(L_k)[\Xi_m^{(L)}]_{i,k}\right) = \sum_{i=1}^n \mathbb{E}(\varepsilon^{(L)}(L_i)^2[\Xi_m^{(L)}]_{i,i})$$
$$= \sum_{i=1}^n \mathbb{E}(S(L_i)(1-S(L_i))[\Xi_m^{(L)}]_{i,i})$$
$$\le \frac{1}{4}\sum_{i=1}^n \mathbb{E}([\Xi_m^{(L)}]_{i,i}) = \frac{1}{4}\mathbb{E}\left(\mathrm{Tr}(\Xi_m^{(L)})\right).$$

Indeed $\Xi_m^{(L)}$ is a symmetric positive matrix, so that ${}^t x\Xi_m^{(L)}x > 0$ for all vector x, and thus its diagonal coefficients are positive. Now $\mathrm{Tr}(\Xi_m^{(L)}) = \mathrm{Tr}(({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1}\,{}^t\Phi_m^{(L)}\Phi_m^{(L)}) = \mathrm{Tr}(I_m) = m$. Thus

$$\mathbb{E}\left[\|\hat{S}_m^{(L)} - \Pi_m^{(L)}S\|_{n,L}^2\right] \le \frac{1}{4}\frac{m}{n}$$

and plugging this in (17) gives the result of Proposition 1 for $\hat{S}_m^{(L)}$. The same ideas give the result for $\hat{S}_m^{(U)}$. $\square$

6.2. **Proof of Proposition 3.** We start by the contrast decomposition: let $t, t' \in \Sigma_m$, then

$$\gamma_n(t) - \gamma_n(t') = \|t - S_I\|_{n,U}^2 + \|t - S_I\|_{n,L}^2 - (\|t' - S_I\|_{n,U}^2 + \|t' - S_I\|_{n,L}^2)$$
$$-2\nu_{n,U}(t-t') - 2\nu_{n,L}(t-t'), \tag{18}$$

where

$$\nu_{n,U}(t) = \frac{1}{n}\sum_{i=1}^n t(U_i)(\mathbb{1}_{\delta_i=1} - S(U_i)), \quad \nu_{n,L}(t) = \frac{1}{n}\sum_{i=1}^n t(L_i)(\mathbb{1}_{\delta_i\ne -1} - S(L_i)).$$

Writing that $\gamma_n(\widehat{S}_m) \le \gamma_n(S_m)$ for any $S_m \in \Sigma_m$, we get
$$\|\widehat{S}_m - S_I\|_{n,U}^2 + \|\widehat{S}_m - S_I\|_{n,L}^2 \le \|S_m - S_I\|_{n,U}^2 + \|S_m - S_I\|_{n,L}^2 + 2\nu_{n,U}(\widehat{S}_m - S_m) + 2\nu_{n,L}(\widehat{S}_m - S_m).$$
Denoting by $\varepsilon^{(L)}(L_i) = \mathbb{1}_{\delta_i\ne -1} - S(L_i)$ and $\varepsilon^{(U)}(U_i) = \mathbb{1}_{\delta_i=1} - S(U_i)$, the inequality writes

$$\mathbb{E}\left[\|\widehat{S}_m - S_I\|_{n,U}^2 + \|\widehat{S}_m - S_I\|_{n,L}^2\right] \le \mathbb{E}\left[\|S_m - S_I\|_{n,U}^2 + \|S_m - S_I\|_{n,L}^2\right]$$
$$+\frac{2}{n}\mathbb{E}\left[\sum_{i=1}^n\left(\varepsilon^{(L)}(L_i)(\widehat{S}_m - S_m)(L_i) + \varepsilon^{(U)}(U_i)(\widehat{S}_m - S_m)(U_i)\right)\right]$$
$$\le \|S_m - S_I\|_U^2 + \|S_m - S_I\|_L^2$$
$$+\frac{2}{n}\underbrace{\mathbb{E}\left[\sum_{i=1}^n\left(\varepsilon^{(L)}(L_i)\widehat{S}_m(L_i) + \varepsilon^{(U)}(U_i)\widehat{S}_m(U_i)\right)\right]}_{:=\mathbb{T}} \tag{19}$$

Let us set

$$\Theta_m = {}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}. \tag{20}$$

As we have

$$\mathbb{T} = ({}^t\vec{\varepsilon}^{(L)}(L)\Phi_m^{(L)} + {}^t\vec{\varepsilon}^{(U)}(U)\Phi_m^{(U)})\Theta_m^{-1}({}^t\Phi_m^{(L)}\vec{\delta}^{(L)} + {}^t\Phi_m^{(U)}\vec{\delta}^{(U)})$$

we find

$$\mathbb{E}(\mathbb{T}) = \mathbb{E}\left(({}^t\vec{\varepsilon}^{(L)}(L)\Phi_m^{(L)} + {}^t\vec{\varepsilon}^{(U)}(U)\Phi_m^{(U)})\Theta_m^{-1}({}^t\Phi_m^{(L)}\vec{\varepsilon}^{(L)}(L) + {}^t\Phi_m^{(U)}\vec{\varepsilon}^{(U)}(U))\right).$$

We get

$$\mathbb{E}(\mathbb{T}) := \mathbb{E}(\mathbb{T}_L) + \mathbb{E}(\mathbb{T}_U) + \mathbb{E}(\mathbb{T}_{L,U}) \tag{21}$$

where, by using that $\mathbb{E}[(\varepsilon^{(L)}(L_i))^2|L_i] = S(L_i)(1-S(L_i))$ and $\mathbb{E}[(\varepsilon^{(U)}(U_i))^2|U_i] = S(U_i)(1-S(U_i))$,

$$\mathbb{T}_L = \sum_{i=1}^n S(L_i)(1-S(L_i))[\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}]_{i,i},\ \mathbb{T}_U = \sum_{i=1}^n S(U_i)(1-S(U_i))[\Phi_m^{(U)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}]_{i,i}$$

and

$$\mathbb{T}_{L,U} = \sum_{i=1}^n S(U_i)(1-S(L_i))[\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)} + \Phi_m^{(U)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}]_{i,i}.$$

Let us denote by $\|\vec{x}\|_{2,d}^2 = x_1^2 + \cdots + x_d^2$ the euclidean norm of a vector $\vec{x}$ of $\mathbb{R}^d$ and by $\vec{e}_{i,d}$ the i-th canonical basis vector in $\mathbb{R}^d$, that is the d-dimensional vector with all coordinates null except the i-th which is equal to 1. Then we have, for $Z = L, U$,

$$[\Phi_m^{(Z)}\Theta_m^{-1}\,{}^t\Phi_m^{(Z)}]_{i,i} = {}^t\vec{e}_{i,n}\Phi_m^{(Z)}\Theta_m^{-1}\,{}^t\Phi_m^{(Z)}\vec{e}_{i,n} = \|\Theta_m^{-1/2}\,{}^t\Phi_m^{(Z)}\vec{e}_{i,n}\|_{2,n}^2 \geq 0,$$

where $\Theta_m^{-1/2}$ is a matrix symmetric square root of Θ_m^{-1}. Thus for $Z = L, U$, we have

$$\mathbb{E}(\mathbb{T}_Z) \leq \frac{1}{4}\mathbb{E}\left(\sum_{i=1}^n [\Phi_m^{(Z)}\Theta_m^{-1}\,{}^t\Phi_m^{(Z)}]_{i,i}\right) = \frac{1}{4}\mathbb{E}(\mathrm{Tr}(\Phi_m^{(Z)}\Theta_m^{-1}\,{}^t\Phi_m^{(Z)})) = \frac{1}{4}\mathbb{E}(\mathrm{Tr}(\Theta_m^{-1}\,{}^t\Phi_m^{(Z)}\Phi_m^{(Z)})).$$

It follows that

$$\mathbb{E}(\mathbb{T}_L + \mathbb{T}_U) \leq \frac{1}{4}\mathbb{E}(\mathrm{Tr}(\Theta_m^{-1}(\,{}^t\Phi_m^{(L)}\Phi_m^{(L)} + {}^t\Phi_m^{(U)}\Phi_m^{(U)}))) = \frac{1}{4}\mathrm{Tr}(I_m) = \frac{m}{4}. \tag{22}$$

Now we prove that $\mathbb{T}_{L,U} \leq m$. Let us set $D^2 = \mathrm{diag}(d_1^2, \ldots, d_n^2)$ with $d_i^2 = S(U_i)(1-S(L_i))$. We have

$$\mathbb{T}_{L,U} = \mathrm{Tr}\left(D^2(\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)} + \Phi_m^{(U)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)})\right) = \mathrm{Tr}\left(\Theta_m^{-1}(\,{}^t\Phi_m^{(U)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(L)}D^2\Phi_m^{(U)})\right).$$

Let us denote by $\Theta_{m,D} := {}^t\Phi_m^{(L)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(U)}D^2\Phi_m^{(U)}$. We remark that, for any vector $\vec{x} \in \mathbb{R}^m$, we have

$${}^t\vec{x}\,{}^t(D(\Phi_m^{(L)} - \Phi_m^{(U)})\ (D(\Phi_m^{(L)} - \Phi_m^{(U)})\vec{x} = \|D(\Phi_m^{(L)} - \Phi_m^{(U)})\vec{x}\|_{2,n}^2 \geq 0$$

and the term is also equal to

$${}^t\vec{x}\,{}^t(D(\Phi_m^{(L)} - \Phi_m^{(U)}))\ (D(\Phi_m^{(L)} - \Phi_m^{(U)}))\vec{x} = {}^t\vec{x}\,{}^t\left(\Theta_{m,D} - (\,{}^t\Phi_m^{(U)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(L)}D^2\Phi_m^{(U)})\right)\vec{x}.$$

Setting $\vec{x} = \Theta_m^{-1/2}\vec{y}$, we get

$${}^t\vec{y}\,\Theta_m^{-1/2}(\,{}^t\Phi_m^{(U)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(L)}D^2\Phi_m^{(U)})\Theta_m^{-1/2}\vec{y} \leq {}^t\vec{y}\,\Theta_m^{-1/2}\Theta_{m,D}\Theta_m^{-1/2}\vec{y}.$$

Choosing $\vec{y} = \vec{e}_{i,m}$ and summing up the terms over i, we obtain that

$$\begin{aligned}\mathrm{Tr}\left(\Theta_m^{-1}(\,{}^t\Phi_m^{(U)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(L)}D^2\Phi_m^{(U)})\right) &= \mathrm{Tr}\left(\Theta_m^{-1/2}(\,{}^t\Phi_m^{(U)}D^2\Phi_m^{(L)} + {}^t\Phi_m^{(L)}D^2\Phi_m^{(U)})\Theta_m^{-1/2}\right)\\ &\leq \mathrm{Tr}\left(\Theta_m^{-1/2}\Theta_{m,D}\Theta_m^{-1/2}\right) = \mathrm{Tr}\left(\Theta_{m,D}\Theta_m^{-1}\right).\end{aligned}$$

Now, let λ be an eigenvalue of $\Theta_{m,D}\Theta_m^{-1}$, associated to a nonzero eigenvector $\vec{x}$, $\vec{x} \in \mathbb{R}^m$, we have

$$\Theta_m^{-1}\Theta_{m,D}\vec{x} = \lambda\vec{x} \ \Rightarrow \Theta_{m,D}\vec{x} = \lambda\Theta_m\vec{x} \Rightarrow {}^t\vec{x}\Theta_{m,D}\vec{x} = \lambda\,{}^t\vec{x}\Theta_m\vec{x}.$$

It is easy to see that ${}^t\vec{x}\Theta_{m,D}\vec{x} \geq 0$ and ${}^t\vec{x}\Theta_m\vec{x} > 0$ as Θ_m is assumed to be invertible, and thus

$$\lambda = \frac{{}^t\vec{x}\Theta_{m,D}\vec{x}}{{}^t\vec{x}\Theta_m\vec{x}} = \frac{{}^t\vec{z}_1 D^2 \vec{z}_1 + {}^t\vec{z}_2 D^2 \vec{z}_2}{{}^t\vec{z}_1\vec{z}_1 + {}^t\vec{z}_2\vec{z}_2}$$

where $\vec{z}_k = \Phi_m^{(Z)}\vec{x} \in \mathbb{R}^n$ where $k = 1$ for $Z = L$ and $k = 2$ for $Z = U$. It follows that

$$\lambda = \frac{\sum_{i=1}^n d_i^2([\vec{z}_1]_i^2 + [\vec{z}_2]_i^2)}{\sum_{i=1}^n([\vec{z}_1]_i^2 + [\vec{z}_2]_i^2)} \leq 1$$

since $\forall i, d_i^2 \leq 1$. Now the trace of a square $m \times m$ matrix which has all its eigenvalues less than 1 (and is diagonalizable), is less than m. This implies that

$$\mathbb{T}_{L,U} \leq m. \tag{23}$$

Gathering (21), (22) and (23), we get that

$$\frac{2}{n}\mathbb{E}(\mathbb{T}) \leq \frac{5}{2}\frac{m}{n}$$

and plugging this in (19), we obtain, for any $S_m \in \Sigma_m$,

$$\mathbb{E}\left[\|\widehat{S}_m - S_I\|_{n,U}^2 + \|\widehat{S}_m - S_I\|_{n,L}^2\right] \leq \mathbb{E}\left[\|S_m - S_I\|_{n,U}^2 + \|S_m - S_I\|_{n,L}^2\right] + \frac{5}{2}\frac{m}{n}.$$

Now, using that $\mathbb{E}\left[\|S_m - S_I\|_{n,Z}^2\right] = \|S_m - S_I\|_Z^2$ for $Z = L, U$, we obtain the result of Proposition 3. □

6.3. **Proof of Theorem 1.** The result is mainly a particular case of Theorem 2 of Comte and Genon-Catalot (2019), in a simpler case of bounded noise. This is why we only present here a sketch of proof.

The main tools in the proof of Comte and Genon-Catalot (2019) are the Talagrand Inequality and Tropp's (2015) matricial Bernstein Inequality. Both still apply here. For Talagrand, we lose the independence property of the noise with respect to the regressor, but get a simplified setting due to the boundedness property of $\varepsilon^{(L)}(L_i) = \mathbb{1}_{\delta_i \neq -1} - S(L_i)$ and $\varepsilon^{(U)}(U_i) = \mathbb{1}_{\delta_i = 1} - S(U_i)$. For Tropp's Inequality, it allows to have here the following fundamental Lemma:

Lemma 1. *Let $(L_1, U_1), \ldots, (L_n, U_n)$ be i.i.d. such that the densities f_U and f_L are bounded, $\sup_{x \in I} f_Z(x) := \|f_Z\|_\infty < +\infty$ for $Z = L, U$. Let the basis be such that $\|\sum_{j=0}^{m-1}\varphi_j^2\|_\infty \leq c_\varphi^2 m$. Then, for all $u > 0$,*

$$\mathbb{P}\left[\|\Psi_m - \widehat{\Psi}_m\|_{\text{op}} \geq u\right] \leq 2m\exp\left(-\frac{n\ u^2/2}{2c_\varphi^2\ [(\|f_L\|_\infty + \|f_U\|_\infty)\ + u/3]}\right).$$

The proof is the same as the proof of Proposition 4 in Comte and Genon-Catalot (2019) with here the bound $c_\varphi^2 m/n$ in (26) replaced by $2c_\varphi^2 m/n$ and the bound on $\nu_n(\mathbf{S}_m)$, $c_\varphi^2\|f\|_\infty m/n$ replaced by $2c_\varphi^2(\|f_L\|_\infty + \|f_U\|_\infty)m/n$.

This result is useful to study the set Ω_n defined by

$$\Omega_n = \cap_{m \in \mathcal{M}_n}\Omega_m \quad \text{with} \quad \Omega_m = \left\{\left|\frac{\|t\|_n^2}{\|t\|_{L+U}^2} - 1\right| \leq \frac{1}{2}, \forall t \in \Sigma_m(I) \setminus \{0\}\right\}. \tag{24}$$

where $\|t\|_n^2 = \|t\|_{n,L}^2 + \|t\|_{n,U}^2$. Indeed Lemma 5 in Comte and Genon-Catalot (2019) can be written here as follows:

Lemma 2. *Under the assumptions of Theorem 1, $\mathbb{P}(\Omega_n^c) \leq c/n^4$ where c is a positive constant.*

To understand the link between Lemma 1 and Lemma 2, we mention that the main point of the proof is the equality

$$\begin{aligned} & \mathbb{P}\left(\exists t \in \Sigma_m(I), \left|\frac{\|t\|_n^2}{\|t\|_{L+U}^2} - 1\right| \le \frac{1}{2}\right) \\ = \ & \mathbb{P}\left(\sup_{t\in\Sigma_m(I), \|t\|_{L+U}=1} \left|\frac{1}{n}\sum_{i=1}^n [t^2(L_i) + t^2(U_i) - \mathbb{E}(t^2(L_i) + t^2(U_i))]\right| > \frac{1}{2}\right), \end{aligned}$$

and the bound

$$\sup_{t\in\Sigma_m(I), \|t\|_{L+U}=1} \left|\frac{1}{n}\sum_{i=1}^n [t^2(L_i) + t^2(U_i) - \mathbb{E}(t^2(L_i) + t^2(U_i))]\right| \le \|\Psi_m^{-1}\|_{\rm op} \|\widehat{\Psi}_m - \Psi_m\|_{\rm op}.$$

Let us start the proof of Theorem 1 in a simplified context: we consider the estimator $\widehat{S}_m$ with $\hat{m}$ selected in the non random collection $\mathcal{M}_n$ and the empirical norm for the risk. The step from this to the effective random collection is given in the proof of Theorem 2 of Comte and Genon-Catalot (2019) as well as the last step to get a risk bound in term of integral norm weighted by $f_L + f_U$. The starting point is the contrast decomposition (18). We use this to write that, for all $m \in \mathcal{M}_n$, for all $S_m \in \Sigma_m(I)$:

$$\gamma_n(\widehat{S}_{\hat{m}}) + \text{pen}(\hat{m}) \le \gamma_n(S_m) + \text{pen}(m).$$

We get

$$\begin{aligned} \|\widehat{S}_{\hat{m}} - S_I\|_{n,U}^2 + \|\widehat{S}_{\hat{m}} - S_I\|_{n,L}^2 \ \le \ & \|S_m - S_I\|_{n,U}^2 + \|S_m - S_I\|_{n,L}^2 + \text{pen}(m) \\ & +2\nu_{n,U}(\widehat{S}_{\hat{m}} - S_m) + 2\nu_{n,L}(\widehat{S}_{\hat{m}} - S_m) - \text{pen}(\hat{m}). \end{aligned} \tag{25}$$

Define

$$\nu_n(t) = \nu_{n,L}(t) + \nu_{n,U}(t) = \frac{1}{n}\sum_{i=1}^n \left[\varepsilon^{(L)}(L_i)t(L_i) + \varepsilon^{(U)}(U_i)t(U_i)\right]$$

and recall that $\mathbb{E}(\|t\|_n^2) = \mathbb{E}(\|t\|_{n,L}^2) + \mathbb{E}(\|t\|_{n,U}^2) = \|t\|_{L+U}^2 = \int t^2(x)(f_L(x) + f_U(x))dx$.

In the following, we write Σ_m for $\Sigma_m(I)$, for sake of brevity. Taking expectation of (25) yields

$$\begin{aligned} \mathbb{E}\left(\|\widehat{S}_{\hat{m}} - S_I\|_n^2\right) \ \le \ & \|S_m - S_I\|_{L+U}^2 + \text{pen}(m) \\ & +2\mathbb{E}\left(\|\widehat{S}_{\hat{m}} - S_m\|_{L+U} \sup_{t\in\Sigma_m+\Sigma_{\hat{m}}, \|t\|_{L+U}=1} |\nu_n(t)|\right) - \mathbb{E}(\text{pen}(\hat{m})) \\ \le \ & \|S_m - S_I\|_{L+U}^2 + \text{pen}(m) + \frac{1}{4}\mathbb{E}(\|\widehat{S}_{\hat{m}} - S_m\|_{L+U}^2) \\ & +4\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{\hat{m}}, \|t\|_{L+U}=1} \nu_n^2(t)\right) - \mathbb{E}(\text{pen}(\hat{m})), \end{aligned}$$

where we use that $2|ab| \le (1/4)a^2 + 4b^2$ for all real numbers a, b. Now we bound separately the terms on Ω_n and Ω_n^c where Ω_n is defined by (24). We get

$$\begin{aligned} \mathbb{E}\left(\|\widehat{S}_{\hat{m}} - S_I\|_n^2 \mathbb{1}_{\Omega_n}\right) \ \le \ & \|S_m - S_I\|_{L+U}^2 + \text{pen}(m) + \frac{1}{4}\mathbb{E}\left(\|\widehat{S}_{\hat{m}} - S_m\|_n^2 \mathbb{1}_{\Omega_n}\right) \\ & +4\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{\hat{m}}, \|t\|_{L+U}=1} \nu_n^2(t)\right) - \mathbb{E}(\text{pen}(\hat{m})). \end{aligned}$$

Thus

$$\frac{1}{2}\mathbb{E}\left(\|\widehat{S}_{\hat{m}}-S_I\|_n^2\mathbb{1}_{\Omega_n}\right) \leq \frac{3}{2}\|S_m-S_I\|_{L+U}^2+\text{pen}(m) + 4\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{\hat{m}},\|t\|_{L+U}=1}\nu_n^2(t)-p(m,\hat{m})\right)_+ + \mathbb{E}(4p(m,\hat{m})-\text{pen}(\hat{m}))$$

Then we can apply Talagrand inequality to get the Lemma:

Lemma 3. *Under the assumptions of Theorem 1, we have*

$$\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{\hat{m}},\|t\|_{L+U}=1}\nu_n^2(t)-p(m,\hat{m})\right)_+ \leq \frac{c}{n}$$

where $p(m,m')=2(m+m')/n$.

Therefore $\forall m,m',\quad 4p(m,m')-\text{pen}(m')\leq\text{pen}(m)$ provided that $\text{pen}(m)=\kappa m/n$ with $\kappa\geq 8$. Thus we get, $\forall m\in\mathcal{M}_n$, $\forall S_m\in\Sigma_m$

$$\mathbb{E}\left(\|\widehat{S}_{\hat{m}}-S_I\|_n^2\mathbb{1}_{\Omega_m}\right) \leq 3\|S_m-S_I\|_{L+U}^2+4\text{pen}(m)+\frac{2c}{n}. \tag{26}$$

On the other hand, we need to propose a rough bound for $\|\widehat{S}_{\hat{m}}-S_I\|_n^2$ in order to control this term on the set Ω_n^c. To that aim, we prove the Lemma

Lemma 4. *Under the Assumption of Theorem 1, for all* $m\in\mathcal{M}_n$, $\|\widehat{S}_m-S_I\|_n^2\leq 18$, *almost surely.*

It follows from Lemma 4 and Lemma 2, that

$$\mathbb{E}\left(\|\widehat{S}_{\hat{m}}-S_I\|_n^2\mathbb{1}_{(\Omega_n)^c}\right)\leq 3\sqrt{2\mathbb{P}[(\Omega_n)^c]}\leq\frac{c^*}{n}, \tag{27}$$

where c^* is a constant. Gathering (26) and (27) gives the first step of the result. □

Proof of Lemma 3. We obtain the result by applying the following Talagrand concentration inequality given in Klein and Rio (2005).

Theorem 2. *Consider* $n\in\mathbb{N}^*$, $\mathcal{F}$ *a class at most countable of measurable functions, and* $(X_i)_{i\in\{1,\ldots,n\}}$ *a family of real independent random variables. Define, for* $f\in\mathcal{F}$, $\nu_n(f)=(1/n)\sum_{i=1}^n(f(X_i)-\mathbb{E}[f(X_i)])$, *and assume that there are three positive constants* M, H *and* v *such that* $\sup_{f\in\mathcal{F}}\|f\|_\infty\leq M$, $\mathbb{E}[\sup_{f\in\mathcal{F}}|\nu_n(f)|]\leq H$, *and* $\sup_{f\in\mathcal{F}}(1/n)\sum_{i=1}^n\text{Var}(f(X_i))\leq v$. *Then for all* $\alpha>0$,

$$\mathbb{E}\left[\left(\sup_{f\in\mathcal{F}}|\nu_n(f)|^2-2(1+2\alpha)H^2\right)_+\right] \leq \frac{4}{b}\left(\frac{v}{n}e^{-b\alpha\frac{nH^2}{v}}+\frac{49M^2}{bC^2(\alpha)n^2}e^{-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7}\frac{nH}{M}}\right)$$

with $C(\alpha)=(\sqrt{1+\alpha}-1)\wedge 1$, *and* $b=\frac{1}{6}$.

Classically, by density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space (as it is needed in our case), after checking that $f\to\nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

For our process, we first note that, the collection of models being nested $\Sigma_m + \Sigma_{m'} = \Sigma_{m\vee m'}$. Let $\bar\varphi_j$ be a linear transformation of the basis $(\varphi_j)_j$ orthonormal with respect to the scalar product weighted by $f_U + f_L$ (by Gramm-Schmidt orthonormalisation), then

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\nu_n^2(t)\right) &\le \sum_{j=1}^{m\vee m'}\mathbb{E}(\nu_n^2(\bar\varphi_j)) = \frac{1}{n}\sum_{j=1}^{m\vee m'}\mathrm{Var}\left(\bar\varphi_j(L_1)\varepsilon^{(L)}(L_1)+\bar\varphi_j(U_1)\varepsilon^{(U)}(U_1)\right)\\ &\le \frac{2}{n}\sum_{j=1}^{m\vee m'}\mathbb{E}\left(\bar\varphi_j^2(L_1)(\mathbb{1}_{\delta_1\neq -1}-S(L_1))^2+\bar\varphi_j^2(U_1)(\mathbb{1}_{\delta_1=1}-S(U_1))^2\right)\\ &\le \frac{2}{n}\sum_{j=1}^{m\vee m'}\mathbb{E}(S(L_1)(1-S(L_1))\bar\varphi_j^2(L_1)+S(U_1)(1-S(U_1))\bar\varphi_j^2(U_1))\\ &\le \frac{m\vee m'}{2n}\le\frac{m+m'}{2n}:=H^2,\end{aligned}$$

using that $x(1-x)\le 1/4$ for any $x\in[0,1]$ and $\mathbb{E}(\bar\varphi_j^2(U_1)+\bar\varphi_j^2(L_1))=1$ by definition of $\bar\varphi_j$. Next,

$$\begin{aligned}\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\mathrm{Var}\left(t(L_1)\varepsilon^{(L)}(L_1)+t(U_1)\varepsilon^{(U)}(U_1)\right) &\le 2\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\mathbb{E}\left(t^2(L_1)+t^2(U_1)\right)\\ &= 2 := v.\end{aligned}$$

Lastly,

$$\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\ \sup_{(x,u)\in\mathbb{R}^+\times\mathbb{R}^+}|\varepsilon^{(L)}(x)t(x)+\varepsilon^{(U)}(u)t(u)|\le 2\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\ \sup_{x\in\mathbb{R}^+}|t(x)|.$$

For $t=\sum_{j=0}^{m-1}a_j\varphi_j$, we have $\|t\|_{U+L}^2 = {}^t\vec a\,\Psi_m\vec a = \|\sqrt{\Psi_m}\vec a\|_{2,m}^2$, where $\vec a = {}^t(a_0,a_1,\dots,a_n)$. Thus, for any m,

$$\begin{aligned}\sup_{t\in\Sigma_m,\|t\|_{U+L}=1}\sup_x|t(x)| &\le c_\varphi\sqrt{m}\sup_{\|\sqrt{\Psi_m}\vec a\|_{2,m}=1}\|\vec a\|_m\\ &\le c_\varphi\sqrt{m}\sup_{\|\vec u\|_{2,m}=1}|\|\sqrt{\Psi_m^{-1}}\vec u\|_{2,m} = c_\varphi\sqrt{m}\sqrt{\|\Psi_m^{-1}\|_{\mathrm{op}}}.\end{aligned}$$

Using the definition of $\mathcal{M}_n$, we have

$$\sqrt{m}\sqrt{\|\Psi_m^{-1}\|_{\mathrm{op}}}\le\left(m\|\Psi_m^{-1}\|_{\mathrm{op}}^2\right)^{1/4}m^{1/4}\le\left(\mathfrak{c}\frac{n}{\log(n)}\right)^{1/4}m^{1/4}.$$

Now, we get a bound similar to the one in Comte and Genon-Catalot (2019) (with $k_n=2$) and

$$M_1 = 2c_\varphi\left(\mathfrak{c}\frac{n}{\log(n)}\right)^{1/4}(m\vee m')^{1/4}.$$

Therefore, applying Talagrand inequality recalled in Theorem 2 gives

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{\hat m},\|t\|_{L+U}=1}\nu_n^2(t)-p(m,\hat m)\right)_+ &\le \sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{t\in\Sigma_m+\Sigma_{m'},\|t\|_{L+U}=1}\nu_n^2(t)-p(m,m')\right)_+\\ &\le\sum_{m'\in\mathcal{M}_n}\frac{C_1}{n}\left(e^{-C_2(m\vee m')}+\frac{(m\vee m')^{1/2}}{n^{1/2}}e^{-C_3(n(m\vee m'))^{1/4}}\right)\\ &\le \frac{C_4}{n},\end{aligned}$$

for C_i, $i = 1, \ldots, 4$ constants and $p(m, m') = 4H^2$ $(\alpha = 1/2)$. This ends the proof. □

Proof of Lemma 4. First recall that $\|\widehat{S}_{\hat{m}} - S_I\|_n^2 = \|\widehat{S}_{\hat{m}} - S_I\|_{n,L}^2 + \|\widehat{S}_{\hat{m}} - S_I\|_{n,U}^2$ and we prove that the first term is bounded by 3, the other term being similar. Now we consider the euclidean norm and recalling the definition of the estimator and of Θ_m (see (20)), we have, for any $m \leq n$:

$$\begin{aligned} n\,\|\widehat{S}_m - S_I\|_{n,L}^2 &= \|\Phi_m^{(L)}\Theta_m^{-1}({}^t\Phi_m^{(L)}\vec{\delta}^{(L)} + {}^t\Phi_m^{(U)}\vec{\delta}^{(U)}) - \vec{S}_I(L)\|_{2,n}^2 \\ &\leq 3[\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\vec{\delta}^{(L)}\|_{2,n}^2 + \|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\vec{\delta}^{(U)}\|_{2,n}^2 + \|\vec{S}_I(L)\|_{2,n}^2] \end{aligned}$$

with $\vec{S}_I(L) = {}^t(S_I(L_1), \ldots, S_I(L_n))$. Now we prove that each of the three terms is smaller than or equal to n. Clearly, this is true for $\|\vec{S}_I(L)\|_{2,n}^2 = S_I^2(L_1) + \cdots + S_I^2(L_n) \leq n$. Next, by definition of the operator norm, it follows that

$$\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\vec{\delta}^{(L)}\|_{2,n}^2 \leq \|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\|_{\text{op}}^2\|\vec{\delta}^{(L)}\|_{2,n}^2.$$

Since $\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}$ is a symmetric positive definite matrix, its operator norm corresponds to its largest eigenvalue. Let λ be any eigenvalue of $\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}$ associated with an eigenvector $\vec{x}$: $\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\vec{x} = \lambda\vec{x}$. Multiplying both sides by ${}^t\Phi_m^{(L)}$, we get, for $\vec{y} = {}^t\Phi_m^{(L)}\vec{x}$, ${}^t\Phi_m^{(L)}\Phi_m^{(L)}\Theta_m^{-1}\vec{y} = \lambda\vec{y}$, which means that λ is an eigenvalue of ${}^t\Phi_m^{(L)}\Phi_m^{(L)}\Theta_m^{-1}$. Now setting $\vec{z} = ({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1/2}\vec{y}$ where $S^{1/2}$ is a symmetric square root of a symmetric matrix S, we obtain that λ is also an eigenvalue of

$$({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{1/2}\Theta_m^{-1}({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{1/2} = \left[\text{Id}_m + ({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1/2}({}^t\Phi_m^{(U)}\Phi_m^{(U)})({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1/2}\right]^{-1},$$

where Id_m is the $m\times m$ identity matrix. Clearly $M := ({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1/2}({}^t\Phi_m^{(U)}\Phi_m^{(U)})({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1/2}$ is symmetric positive definite and is diagonalizable in an orthonormal basis as $\text{diag}(a_1, \ldots, a_m)$ with $a_i > 0$ for $i = 1, \ldots, m$. In this basis $(I + M)^{-1}$ is equal to $(1/(1 + a_1), \ldots, 1/(1 + a_m))$, and all these eigenvalues are in $(0, 1)$. Therefore $\lambda \leq 1$ and thus $\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\|_{\text{op}}^2 \leq 1$. Consequently, using that all the coordinates of $\vec{\delta}^{(L)}$ belong to $[-1, 1]$, we get the second bound

$$\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}\vec{\delta}^{(L)}\|_{2,n}^2 \leq \|\vec{\delta}^{(L)}\|_{2,n}^2 \leq n.$$

For the last term, we also start with

$$\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\vec{\delta}^{(U)}\|_{2,n}^2 \leq \|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\|_{\text{op}}^2\|\vec{\delta}^{(U)}\|_{2,n}^2.$$

Here the matrix $\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}$ is not symmetric and thus

$$\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\|_{\text{op}}^2 = \lambda_{\max}(\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\Phi_m^{(U)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}),$$

where $\lambda_{\max}(A)$ stands for the largest eigenvalue of a matrix A and $\|A\|_{\text{op}}^2 = \lambda_{\max}({}^tAA)$. As previously an eigenvalue of $\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\Phi_m^{(U)}\Theta_m^{-1}\,{}^t\Phi_m^{(L)}$ is also an eigenvalue of

$$\left(\text{Id}_m + {}^t\Phi_m^{(U)}\Phi_m^{(U)}({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1}\right)^{-1}\left(\text{Id}_m + {}^t\Phi_m^{(L)}\Phi_m^{(L)}({}^t\Phi_m^{(U)}\Phi_m^{(U)})^{-1}\right)^{-1},$$

as both matrices are equal (note that ${}^t\Phi_m^{(U)}\Phi_m^{(U)}({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1}$ is the inverse of ${}^t\Phi_m^{(L)}\Phi_m^{(L)}({}^t\Phi_m^{(U)}\Phi_m^{(U)})^{-1}$). Consider a basis in which the first one is diagonal and of the form $\text{Diag}(a_1, \ldots, a_m)$, then the whole matrix is of the form $\text{Diag}(1/[(1 + a_1)(1 + a_1^{-1})], \ldots, 1/[(1 + a_m)(1 + a_m^{-1})])$, that is the eigenvalues are less that 1 as soon as the a_i's are positive. Now let a be an eigenvalue with $\vec{x}$

associated eigenvector, that is ${}^t\Phi_m^{(U)}\Phi_m^{(U)}({}^t\Phi_m^{(L)}\Phi_m^{(L)})^{-1}\vec{x} = a\vec{x}$. Then ${}^t\Phi_m^{(U)}\Phi_m^{(U)}\vec{y} = a\,{}^t\Phi_m^{(L)}\Phi_m^{(L)}\vec{y}$ and taking the scalar product with $\vec{y}$ yields

$$ {}^t\vec{y}\,{}^t\Phi_m^{(U)}\Phi_m^{(U)}\vec{y} = a\,{}^t\vec{y}\,{}^t\Phi_m^{(L)}\Phi_m^{(L)}\vec{y} $$

that is $\|\Phi_m^{(U)}\vec{y}\|_{2,n}^2 = a\|\Phi_m^{(L)}\vec{y}\|_{2,n}^2$. Thus $a \geq 0$. Lastly $a \neq 0$ because of invertibility assumptions. We obtain that $\|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\|_{\rm op}^2 \leq 1$ and thus

$$ \|\Phi_m^{(L)}\Theta_m^{-1}\,{}^t\Phi_m^{(U)}\vec{\delta}^{(U)}\|_{2,n}^2 \leq n. $$

Therefore, gathering the three bounds n for the euclidean norms gives the bound 9 for the empirical norm and ends the proof.

6.4. **Proof of Inequality (15).** We already mentioned that ${}^t\vec{v}\Psi_{m,Z}\vec{v} = \int_I v^2(x)f_Z(x)dx$ for $Z = L, U$ and $v(x) = \sum_{j=0}^{m-1} v_j\varphi_j(x)$ where $\vec{v} = {}^t(v_0, \ldots, v_{m-1})$. Thus if $\forall x \in I$, $f_Z(x) \geq f_0$, we get for any vector $\vec{v} \in \mathbb{R}^d$,

$$ {}^t\vec{v}\Psi_{m,Z}\vec{v} \geq f_0 \int_I v^2(x)dx = f_0 \sum_{j=0}^{m-1} v_j^2. $$

As a consequence, for any vector $\vec{v} \in \mathbb{R}^d$,

$$ {}^t\vec{v}(\Psi_{m,L} + \Psi_{m,U})\vec{v} \geq 2f_0 \int_I v^2(x)dx = 2f_0 \sum_{j=0}^{m-1} v_j^2. $$

all eigenvalues of $\Psi_{m,L} + \Psi_{m,U}$ are larger than $2f_0$ and therefore, as they are all positive, the largest eigenvalue of $(\Psi_{m,L} + \Psi_{m,U})^{-1}$ is smaller that $1/(2f_0)$. □

References

Abramowitz, M. and Stegun, I. A. (1964), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Vol. 55, Courier Corporation.

Anderson-Bergman, C. and Yu, Y. (2016), 'Computing the log concave npmle for interval censored data', *Statistics and Computing* **26**(4), 813–826.

Baraud, Y. (2002), 'Model selection for regression on a random design', *ESAIM: Probability and Statistics* **6**, 127–146.

Becker, N. G. and Melbye, M. (1991), 'Use of a log-linear model to compute the empirical survival curve from interval-censored data, with application to data on tests for hiv positivity', *Australian Journal of Statistics* **33**(2), 125–133.

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

Bongioanni, B. and Torrea, J. L. (2009), 'What is a Sobolev space for the Laguerre function systems?', *Studia Mathematica* **192**, 147–172.

Brunel, E. and Comte, F. (2009), 'Cumulative distribution function estimation under interval censoring case 1', *Electronic Journal of Statistics* **3**, 1–24.

Carstensen, B. (1996), 'Regression models for interval censored survival data: application to HIV infection in danish homosexual men', *Statistics in Medicine* **15**(20), 2177–2189.

Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2009), 'Improving point and interval estimators of monotone functions by rearrangement', *Biometrika* **96**(3), 559–575.

Comte, F. and Dion, C. (2017), 'Laguerre estimation under constraint at a single point', *Preprint* MAP5 2017-04 and hal-01447605.

Comte, F. and Genon-Catalot, V. (2019), 'Regression function estimation on non compact support as a partly inverse problem', available online in *The Annals of the Institute of Statistical Mathematics*, https://doi.org/10.1007/s10463-019-00718-2

Comte, F. and Genon-Catalot, V. (2015), 'Adaptive Laguerre density estimation for mixed poisson models', *Electronic Journal of Statistics* **9**(1), 1113–1149.

Geskus, R. and Groeneboom, P. (1999)'Asymptotically optimal estimation of smooth functionals for interval censoring, case 2', *The Annals of Statistics* **27**, 627-674.

Groeneboom, P. (1996), Lectures on inverse problems. *Lectures on Probability Theory and Statistics (Saint-Flour, 1994)*, 67-164, Lecture Notes in Math., 1648, Springer, Berlin, 1996.

Groeneboom, P., Jongbloed, G. and Witte, B.I. (2010)'Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model', *The Annals of Statistics* **38**, 352-387.

Groeneboom, P. and Ketelaars, T. (2011), 'Estimators for the interval censoring problem', *Electronic Journal of Statistics* **5**, 1797–1845.

Groeneboom, P. and Wellner, J. A. (1992), *Information bounds and nonparametric maximum likelihood estimation*, Vol. 19, Springer Science; Business Media.

Klein, T. and Rio, E. (2005) Concentration around the mean for maxima of empirical processes. *The Annals of Probability* **33**, no. 3, 1060-1077.

Kooperberg, C. and Stone, C. J. (1992), 'Logspline density estimation for censored data', *Journal of Computational and Graphical Statistics* **1**(4), 301–328.

Melbye, M., Biggar, R. J., Ebbesen, P., Sarngadharan, M., Weiss, S. H., Gallo, R. C. and Blattner, W. A. (1984), 'Seroepidemiology of htlv-iii antibody in danish homosexual men: prevalence, transmission, and disease outcome.', *British Medical Journal (Clinical Research Edition)* **289**(6445), 573–575.

Tropp, J.A. (2015), 'An introduction to matrix concentration inequalities', *Found. Trends Mach. Learn.*, 8(1-2):1-230.

Turnbull, B. W. (1976), 'The empirical distribution function with arbitrarily grouped, censored and truncated data', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 290–295.

Wellner, J. A. (1995), 'Interval censoring, case 2: alternative hypotheses'. *Analysis of censored data* (Pune, 1994/1995), 271-291, IMS Lecture Notes Monogr. Ser., 27, Inst. Math. Statist., Hayward, CA.

Yang, S. (2000), 'Functional estimation under interval censoring case 1', *Journal of Statistical Planning and Inference* **89**(1), 135–144.

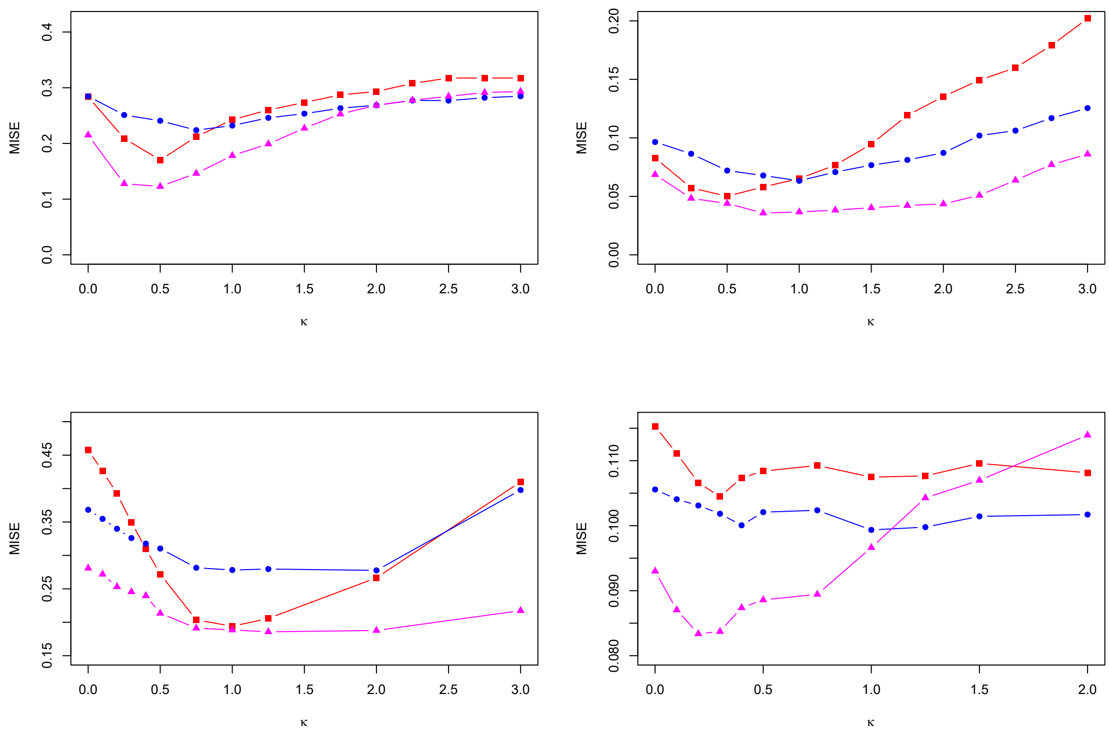


FIGURE 9. MISE$\times 10^2$ computed for $K = 100$ samples of the penalized estimators computed for different values of the constant κ. At the top : $X \sim \mathcal{E}xp(1/2)$, at the bottom $X \sim$ Log-$\mathcal{N}(0, 2)$ with sample size $n = 300$ (left) and $n = 1000$ (right) in scenario 2 (red circle), scenario 3 (blue square) and scenario 4 (magenta triangle).

APPENDIX A. SMALL CALIBRATION STUDY

In order to give some explanations on the choice of the value $\kappa = 0.5$ in Section 4, we propose a small study of the effect of the choice of κ on the proposed estimator. So, we compute the MISE error defined by (16) of our penalized estimator with κ taking different values in the penalty term pen(m) in (14). As the value of κ must be the same independently of the distribution of $(X_i)_{1\leq i\leq n}$ or the sample size n, we choose different distributions and sample sizes. For illustration, we present here the results for $n = 300$ and $n = 1000$, and densities:

- $X \sim \mathcal{E}xp(1/2)$ with density function $f(x) = \frac{1}{2}\exp(-\frac{x}{2})$, $x \geq 0$.
- $X = \exp(Y)$ with $Y \sim \mathcal{N}(0, 2)$ that is $X \sim$ Log-$\mathcal{N}(0, 2)$

Note that, to avoid overfitting, the densities here are different from the ones chosen in the simulation study of Section 4. We use scenario 2, 3 and 4 for the generation of the censoring pair $(L_i, U_i)_{1\leq i\leq n}$ as in the simulation study defined in Section 4.1.

On each picture of Figure 9, we can see that the MISEs have rather the same behaviour for both distributions of X and both sample sizes. We can identify a range of admissible values of

κ corresponding to values where the MISE is near its minimum. The value $\kappa = 0.5$ seems to be an adequate compromise.

Appendix B. Numerical results

We give in Table 1 and 2 the numerical results corresponding to the simulations and boxplots of Section 4. We also add the errors computed for the empirical survival function evaluated with the whole sample $(X_i)_{1 \le i \le n}$, as a benchmark or a kind of "oracle" estimator which represents the best we could obtain if we had observed directly the $X_i's$ without censoring, instead of the intervals $[L_i, U_i]$. We study in Table 3 the effect of the number of replications on the quality of estimations.

Inspection time Scenario	size n		Event time Models							
			Log-$\mathcal{N}(0,1)$		$\mathcal{W}eib(0.5,2)$		$\mathcal{B}eta'(5,2)$		$\mathcal{B}eta(5,2)$	
sc. 1 : $L\sim U[0,2.5]$, $U\sim U[3,4]$	100	MC	5.30	(4.83)	4.61	(4.00)	4.43	(4.27)	3.41	(4.27)
		AndYu	1.92	(1.53)	14.4	(0.97)	15.0	(1.66)	2.63	(2.44)
		NPMLE	7.00	(3.27)	6.76	(3.60)	6.78	(3.77)	8.09	(3.84)
		SMLE	4.72	(3.10)	5.08	(3.33)	5.46	(3.71)	5.50	(3.55)
		oracle	1.56	(1.36)	2.17	(2.24)	2.18	(2.25)	1.49	(1.20)
	300	MC	1.53	(1.12)	2.16	(1.31)	1.36	(1.16)	1.11	(1.00)
		AndYu	0.98	(0.44)	14.0	(0.42)	15.0	(4.56)	0.97	(0.91)
		NPMLE	3.10	(1.26)	2.77	(1.12)	2.53	(1.04)	3.70	(2.10)
		SMLE	2.05	(1.19)	2.06	(1.08)	1.73	(0.93)	2.45	(1.98)
		oracle	0.59	(0.64)	0.62	(0.63)	0.47	(0.48)	0.40	(0.35)
	1000	MC	0.80	(0.43)	0.92	(0.39)	0.42	(0.33)	0.40	(0.26)
		AndYu	0.75	(0.18)	13.8	(0.14)	14.9	(2.75)	0.31	(0.23)
		NPMLE	1.28	(0.48)	1.04	(0.31)	1.21	(0.40)	1.61	(0.56)
		SMLE	0.84	(0.45)	0.82	(0.33)	0.81	(0.34)	1.06	(0.55)
		oracle	0.20	(0.23)	0.18	(0.16)	0.15	(0.16)	0.12	(0.11)
sc. 2 : $L\sim U[0,1]$, $U=L+U[0,3]$	100	MC	5.46	(5.36)	6.87	(7.80)	5.99	(7.21)	2.92	(2.14)
		AndYu	2.24	(1.76)	18.0	(3.67)	16.35	(8.21)	2.41	(2.49)
		NPMLE	6.48	(3.44)	6.30	(3.52)	8.03	(4.90)	7.85	(4.37)
		SMLE	4.57	(3.37)	5.02	(3.37)	6.26	(4.43)	5.38	(4.16)
		oracle	1.66	(1.45)	2.20	(2.28)	1.81	(1.76)	1.63	(1.36)
	300	MC	1.36	(1.77)	2.05	(1.77)	2.17	(2.01)	1.48	(1.23)
		AndYu	1.17	(0.60)	17.3	(1.84)	14.8	(5.91)	0.81	(0.72)
		NPMLE	2.88	(1.22)	2.75	(1.47)	4.08	(2.20)	3.22	(1.23)
		SMLE	2.03	(1.22)	2.20	(1.32)	2.99	(1.88)	2.07	(1.15)
		oracle	0.61	(0.67)	0.62	(0.62)	0.48	(0.46)	0.43	(0.37)
	1000	MC	0.79	(0.30)	1.15	(0.59)	0.95	(0.82)	0.45	(0.22)
		AndYu	0.93	(0.30)	16.8	(1.02)	15.3	(3.60)	0.32	(0.26)
		NPMLE	1.19	(0.40)	1.19	(0.62)	1.78	(0.99)	1.34	(0.44)
		SMLE	0.84	(0.42)	1.00	(0.51)	1.29	(0.81)	0.84	(0.44)
		oracle	0.20	(0.23)	0.18	(0.16)	0.15	(0.16)	0.13	(0.12)

TABLE 1. MISE $\times 10^3$ and standard deviation in parenthesis for $K = 100$ samples, for scenario 1 and 2 : our penalized Least Squares estimator built with Laguerre basis (MC), the log-concave Anderson-Bergman and Yu's NPMLE (AndYu), the unconstrained NPMLE (NPMLE), the smoothed NPMLE (SMLE) and the "oracle" empirical survival function.

Inspection time Scenario	size n	Estimator	Log-$\mathcal{N}(0,1)$		$\mathcal{W}eib(0.5,2)$		$\mathcal{B}eta'(5,2)$		$\mathcal{B}eta(5,2)$	
sc. 3 : $L,U \sim U[0,4]$, $U \geq L$ and $U - L \leq 0.1$	100	MC	7.85	(5.70)	9.63	(7.57)	6.69	(5.85)	5.21	(7.26)
		AndYu	3.43	(3.34)	15.4	(2.26)	15.5	(7.24)	3.99	(3.89)
		NPMLE	9.81	(5.19)	10.3	(5.94)	8.76	(4.47)	10.5	(5.63)
		SMLE	7.09	(5.02)	8.24	(5.67)	6.75	(4.11)	7.52	(5.32)
		oracle	1.52	(1.38)	2.30	(2.46)	1.66	(1.43)	1.48	(1.14)
	300	MC	2.05	(1.97)	2.82	(1.98)	2.32	(1.64)	1.93	(1.42)
		AndYu	1.42	(0.91)	14.7	(0.84)	15.7	(5.28)	1.37	(1.10)
		NPMLE	3.82	(1.55)	3.90	(1.75)	3.58	(1.59)	4.17	(1.50)
		SMLE	2.59	(1.52)	3.13	(1.69)	2.62	(1.43)	2.82	(1.44)
		oracle	0.56	(0.51)	0.66	(0.58)	0.52	(0.50)	0.45	(0.38)
	1000	MC	0.81	(0.35)	1.22	(0.48)	0.88	(0.74)	0.54	(0.42)
		AndYu	0.94	(0.29)	14.8	(0.53)	14.9	(3.47)	0.39	(0.29)
		NPMLE	1.47	(0.44)	1.51	(0.52)	1.47	(0.58)	1.46	(0.56)
		SMLE	1.02	(0.44)	1.20	(0.48)	1.08	(0.53)	0.95	(0.51)
		oracle	0.16	(0.13)	0.17	(0.14)	0.20	(0.19)	0.15	(0.12)
sc. 4 : $L \sim U[0,1]$, $U \sim U[2,4]$	100	MC	5.31	(4.92)	4.43	(4.27)	4.18	(3.65)	2.93	(2.93)
		AndYu	1.97	(1.83)	15.0	(1.66)	14.5	(6.74)	2.91	(3.05)
		NPMLE	10.7	(6.78)	6.78	(3.78)	10.6	(5.30)	17.8	(10.6)
		SMLE	8.32	(6.68)	5.46	(3.72)	8.66	(5.04)	14.1	(10.0)
		oracle	1.58	(1.37)	2.18	(2.25)	1.84	(1.79)	1.54	(1.24)
	300	MC	1.62	(1.67)	2.04	(1.93)	2.16	(2.82)	1.06	(0.88)
		AndYu	1.00	(0.45)	14.6	(0.90)	13.6	(5.28)	1.01	(0.95)
		NPMLE	5.42	(4.37)	3.22	(1.63)	6.51	(3.19)	11.7	(6.40)
		SMLE	5.14	(4.25)	2.62	(1.60)	5.36	(3.04)	9.69	(5.96)
		oracle	0.59	(0.65)	0.62	(0.63)	0.47	(0.48)	0.41	(0.35)
	1000	MC	0.99	(0.47)	1.00	(0.49)	0.64	(0.64)	0.54	(0.29)
		AndYu	0.75	(0.15)	14.3	(0.37)	13.9	(4.00)	0.46	(0.42)
		NPMLE	4.05	(2.24)	1.71	(0.89)	4.48	(1.94)	8.95	(5.11)
		SMLE	3.33	(2.19)	1.49	(0.86)	3.84	(1.93)	7.60	(4.83)
		oracle	0.20	(0.23)	0.18	(0.16)	0.15	(0.16)	0.13	(0.11)

Event time Models (columns: Log-$\mathcal{N}(0,1)$, $\mathcal{W}eib(0.5,2)$, $\mathcal{B}eta'(5,2)$, $\mathcal{B}eta(5,2)$)

TABLE 2. MISE $\times 10^3$ and standard deviation in parenthesis for $K = 100$ samples, for scenario 3 and 4 : our penalized Least Squares estimator built with Laguerre basis (MC), the log-concave Anderson-Bergman and Yu's NPMLE (AndYu), the unconstrained NPMLE (NPMLE), the smoothed NPMLE (SMLE) and the "oracle" empirical survival function.

size n		Log-$\mathcal{N}(0,1)$ (sc. 3)				$\mathcal{W}eib(0.5,2)$ (sc. 2)			
		$K=100$		$K=500$		$K=100$		$K=500$	
300	MC	2.05	(1.97)	2.18	(2.27)	2.05	(1.77)	2.23	(2.11)
	AndYu	1.42	(0.91)	1.44	(0.89)	17.3	(1.84)	17.1	(1.97)
	NPMLE	3.82	(1.55)	3.88	(1.58)	2.75	(1.47)	2.82	(1.44)
	SMLE	2.59	(1.52)	2.69	(1.55)	2.20	(1.32)	2.28	(1.32)
	oracle	0.56	(0.51)	0.54	(0.47)	0.62	(0.62)	0.69	(0.68)
1000	MC	0.81	(0.35)	0.86	(0.44)	1.15	(0.59)	1.17	(0.58)
	AndYu	0.94	(0.29)	0.94	(0.29)	16.8	(1.02)	16.7	(1.05)
	NPMLE	1.47	(0.44)	1.50	(0.54)	1.19	(0.62)	1.28	(0.73)
	SMLE	1.02	(0.44)	1.03	(0.55)	1.00	(0.51)	1.07	(0.63)
	oracle	0.16	(0.13)	0.16	(0.14)	0.18	(0.16)	0.20	(0.21)

TABLE 3. Effect of the number K of replications in the Monte-Carlo study: MISE $\times 10^3$ and standard deviation in parenthesis for our penalized Least Squares estimator built with Laguerre basis (MC), the log-concave Anderson-Bergman and Yu's NPMLE (AndYu), the unconstrained NPMLE (NPMLE), the smoothed NPMLE (SMLE) and the "oracle" empirical survival function.