



Global correction of projection estimators under local constraint

Fabienne Comte, Charlotte Dion

► **To cite this version:**

Fabienne Comte, Charlotte Dion. Global correction of projection estimators under local constraint. Journal of the Korean Statistical Society, Elsevier, 2020, 49 (4), pp.1255 - 1284. 10.1007/s42952-020-00055-8 . hal-01447605v2

HAL Id: hal-01447605

<https://hal.archives-ouvertes.fr/hal-01447605v2>

Submitted on 20 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Projection estimators under constraint at a single point

F. Comte⁽¹⁾ and C. Dion⁽²⁾

(1) *MAP5, UMR CNRS 8145, Université Paris Descartes,*
fabienne.comte@parisdescartes.fr

(2) *LPSM, UMR CNRS 8001, Sorbonne Université,*
charlotte.dion@upmc.fr

Abstract

This paper presents a general methodology for nonparametric estimation of a function s related to a nonnegative real random variable X , under a constraint of type $s(0) = c$. When a projection estimator of the target function is available, we explain how to modify it in order to obtain an estimator which satisfies the constraint. We extend risk bounds from the initial to the new estimator, and propose and study adaptive procedures for both estimators. The example of cumulative distribution function estimation illustrates the method for two different models: the multiplicative noise model ($Y = XU$ is observed, with U following a uniform distribution) and the additive noise model ($Y = X + V$ is observed where V is a nonnegative nuisance variable with known density).

Keywords. Adaptive estimation, constrained estimator, Laguerre basis, nonparametric projection estimator

AMS Classification. 62G05-62G07

1 Introduction

In this work we focus on a function s related to a nonnegative real random variable X which satisfies $s(0) = c$. The idea is to take into account this additional knowledge in a new nonparametric procedure of estimation. Indeed, if s is a survival function supported on \mathbb{R}^+ , it must satisfy the constraint with $c = 1$; there are also examples where densities should be constrained by $c = 0$. This starting value is subject to side effects and often badly estimated. Thus, including the constraint in the procedure is likely to improve the estimator.

1.1 Main motivation

In the statistical literature, different types of global constraints have been studied from nonparametric functional estimation point of view, such as convexity or monotonicity constraints. Specific procedures have been proposed to obtain for instance decreasing density estimators, see [17], [4]. We may also mention the proposal of [10] where an estimator of a weakly increasing function is modified to get a weakly increasing estimator, with no influence on the risk value.

We are interested in a different question, namely: given an estimator built in an orthonormal basis (denoted by \hat{s}_m), can we coherently modify it in order to fix its value in one specific point? By coherently, we mean that we look for a global correction of the estimator, and not a single

discontinuous change at the point. Indeed, correcting the value at 0 only automatically induces a discontinuity of the new estimator.

More precisely, consider a square integrable function s with support \mathbb{R}^+ , as this currently occurs for lifetimes densities or survival functions in survival analysis, reliability and actuarial sciences. When s is square integrable, a natural idea is to consider its development in an orthonormal basis $(\varphi_j)_{j \geq 0}$ with support $I = \mathbb{R}^+$ or $I = [0, \mathbf{a}]$. In other words, we write that $s = \sum_{j \geq 0} a_j(s) \varphi_j$ with $a_j(s) = \langle s, \varphi_j \rangle$. Next we consider that observations Y_1, \dots, Y_n related to s are available and allow us to build a projection estimator \widehat{s}_m of s : $\widehat{s}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j$ where \widehat{a}_j for $j = 0, \dots, m-1$ are known functions of the observations. Moreover we assume that $\mathbb{E}[\widehat{a}_j] = a_j(s)$ and call \widehat{s}_m a projection estimator of s as it is an unbiased estimator of $s_m = \sum_{j=0}^{m-1} a_j(s) \varphi_j$, the orthogonal projection of s on the m -dimensional space $\mathcal{S}_m = \text{span}(\varphi_0, \dots, \varphi_{m-1})$.

Now, instead of correcting the value in zero only, one may consider a new estimator defined by $\widehat{s}_m - \widehat{s}_m(0) + s(0)$. This time the estimator is entirely modified and keeps its regularity: it is simply translated. The resulting squared integrated bias of this estimator for $(\varphi_j)_j$ the trigonometric basis on $[0, \mathbf{a}]$ (see Section 2.1 for its definition) is the one of \widehat{s}_m plus $(1/\mathbf{a})(\sum_{j \geq m+1} a_j)^2$: this is made obvious by Figure 1. This simple fact motivates the work hereafter and shows the need for new ideas.

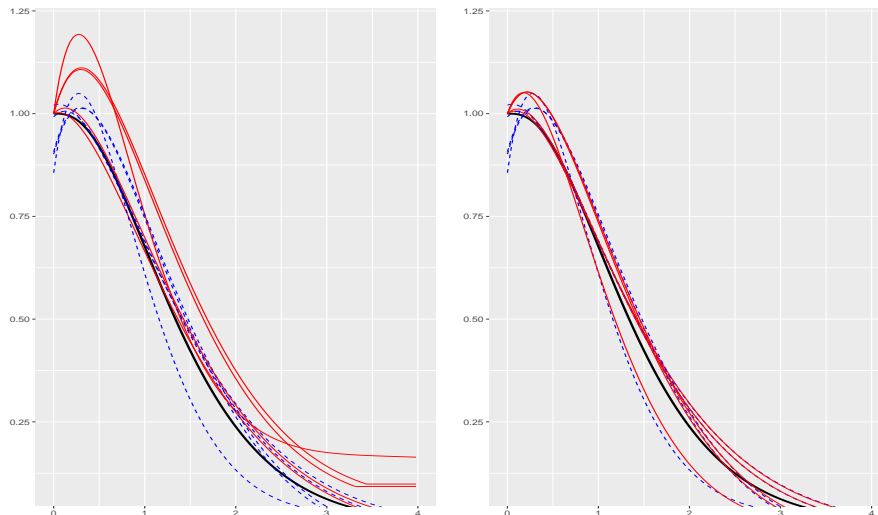


Figure 1: Example of survival function estimation for $n = 100$, zoom on $[0, 4]$. Plain bold black line: true survival function for a Gamma distribution. Dotted blue lines: 4 projection estimators of the survival function. Left: plain red lines: translated projection estimators. Right: plain red lines: proposed estimators.

1.2 Main contribution

To coherently modify the projection estimator \widehat{s}_m into, say \widetilde{s}_m , such that $\widetilde{s}_m(0) = c$ (when $s(0) = c$) we propose a general definition of the constrained estimator and prove that its global mean-square integrated risk on \mathbb{R}^+ is comparable to the risk of \widehat{s}_m . The idea is to write the projection estimator \widehat{s}_m as a minimum contrast estimator and then to modify this contrast by standard Lagrange multiplier strategy. Then, if a model selection procedure of a relevant choice of m is available for \widehat{s}_m , a modification is proposed, ensuring also a relevant choice of m for \widetilde{s}_m , under simple conditions.

Compared to the rearrangement procedure of [10], this procedure only preserves the speed of convergence of the original estimator (obtained without constraint). It does not improve it in the general case because the additional information is only in one point, and not global, contrary to monotonicity.

The procedure is described in full generality, but also illustrated through two examples. First, the multiplicative censoring model is considered, introduced by [22] and related to the field of survival analysis for example, see [21]. In this model, the observations are the $Y_i = X_i U_i$ with X_i and U_i independent and U_i following a uniform distribution on $[0, 1]$. The random variables $(X_i)_{1 \leq i \leq n}$ are independent and identically distributed (*i.i.d.*) and so are the $(U_i)_{1 \leq i \leq n}$. We observe an *i.i.d.* sample of Y_i 's while we are interested in the survival function of X .

This problem is studied as an inverse problem in [3]; projection wavelet estimators of the density of X are studied by [1]; kernel estimators of the density and the survival function are proposed in [8] and Laguerre projection estimators of the density are considered in [5]. Here, we propose and study a new projection estimator of the survival function in two different bases: the Fourier basis and the Laguerre basis. We obtain bounds for both estimators, without and with constraint correction, which are also new.

A second example is the convolution model, $Y_i = X_i + V_i$, where all variables are nonnegative, *i.i.d.*, and V is a nuisance process with known density; the function of interest is the survival function of X_i while only the Y_i 's are observed. The convolution model has been widely investigated, mainly with a Fourier approach [see 9, 16, 19, for example]. Recently, projection estimators of the density and the survival function supported by \mathbb{R}^+ are proposed in [18]. This approach relies on a Laguerre projection estimator and can be used in the present work to deduce a constrained estimator.

1.3 Organization of the paper

The paper is organized as follows. In Section 2 the general method is presented. The projection estimator in the Fourier or Laguerre basis is constructed, and its constrained version is deduced, their risks are compared. The conditions to obtain a model selection result for \hat{s}_m and \tilde{s}_m are given. Section 3.1 is dedicated to the multiplicative noise model case. The procedure is applied to the additive model in Section 3.2. A short simulation study is then provided. A last, Section 5 shows that the procedure may be extended to other frameworks, but is then also more intricate: the example of least-squares estimation of the survival function for interval censored data is shortly presented.

2 A general strategy

2.1 Notations

Let $I \subset \mathbb{R}^+$. The space $\mathbb{L}^2(I)$ is the space of square integrable functions on I . The associated \mathbb{L}^2 -norm is denoted $\|t\|^2 = \int_I |t(x)|^2 dx$ and $\langle t, u \rangle = \int_I t(x)u(x)dx$ for $t, u \in \mathbb{L}^2(I)$. Finally, the supremum norm of a bounded function t is denoted by $\|t\|_\infty = \sup_{x \in I} |t(x)|$.

In the following we focus on two orthogonal bases. First, the Laguerre basis is defined by

$$I = \mathbb{R}^+, \quad \varphi_0(x) = \sqrt{2}e^{-x}, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \text{ for } k \geq 1, \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}.$$

where L_k are the Laguerre polynomials. Indeed, the Laguerre basis is orthonormal for the integral scalar product on \mathbb{R}^+ , $\langle s, t \rangle = \int_0^{+\infty} s(x)t(x)dx$. It satisfies the orthonormality property $\langle \varphi_j, \varphi_k \rangle = \delta_{j,k}$ where $\delta_{j,k}$ is the Kronecker symbol, equal to 1 if $j = k$ and to zero otherwise. The following properties are used in the sequel:

$$\forall j \geq 0, \varphi_j(0) = \sqrt{2}, \text{ and } \|\varphi_j\|_\infty \leq \sqrt{2} \quad (2.1)$$

[see 2]. Any function of $\mathbb{L}^2(\mathbb{R}^+)$ can be decomposed on this basis.

Secondly, the Fourier basis (or trigonometric basis) on $I = [0, \mathbf{a}]$ is defined by

$$\varphi_0(x) = \mathbf{a}^{-1/2}, \quad j \geq 1, \varphi_{2j-1}(x) = \sqrt{2/\mathbf{a}} \cos(2\pi jx/\mathbf{a}), \quad \varphi_{2j}(x) = \sqrt{2/\mathbf{a}} \sin(2\pi jx/\mathbf{a}).$$

Note that

$$\varphi_0(0) = \mathbf{a}^{-1/2}, \quad \forall j \geq 1, \varphi_{2j-1}(0) = \sqrt{2/\mathbf{a}}, \quad \varphi_{2j}(0) = 0, \quad \forall j \geq 0, \|\varphi_j\|_\infty \leq \sqrt{2/\mathbf{a}}. \quad (2.2)$$

For these two bases, we set

$$\mathcal{S}_m = \text{span}\{\varphi_0, \varphi_1, \dots, \varphi_{m-1}\}.$$

2.2 Estimation method and assumptions

Let us denote the sample of observations: $(Y_i)_{1 \leq i \leq n}$ related to the variables of interest $(X_i)_{1 \leq i \leq n}$. All along the paper our strategy is a \mathbb{L}^2 -projection strategy requiring that the following condition holds:

$$\mathbf{(A1)}(s) \quad s \in \mathbb{L}^2(I).$$

Under $\mathbf{(A1)}(s)$, the development $s = \sum_{j \geq 0} a_j(s) \varphi_j$ with $a_j(s) = \langle s, \varphi_j \rangle$, holds in $\mathbb{L}^2(I)$. As $\varphi_j(0)$ is constant for the two bases described above, the following condition ensures that $s(0)$ exists and defines the constraint:

$$\mathbf{(A2)}(s) \quad \sum_{\ell \geq 0} |a_\ell(s)| < +\infty \text{ and } s(0) = c.$$

Note that, by (2.1)-(2.2), $\mathbf{(A2)}(s)$ implies that s is continuous, bounded, and satisfies $\|s\|_\infty \leq \mathbf{c} \sum_{\ell \geq 0} |a_\ell(s)| < +\infty$ with $\mathbf{c} = \sqrt{2}$ (Laguerre) or $\mathbf{c} = \sqrt{2/\mathbf{a}}$ (Fourier).

We assume that we can build an estimator \widehat{s}_m of s on the subspace \mathcal{S}_m as follows:

$$\widehat{s}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \text{ where } \forall j \in \mathbb{N}, \quad \mathbb{E}[\widehat{a}_j] = a_j(s), \quad a_j(s) := \langle s, \varphi_j \rangle,$$

with \widehat{a}_j computed from a known transformation of the observations Y_1, \dots, Y_n . This clearly implies that $\mathbb{E}[\widehat{s}_m] = s_m = \sum_{j=0}^{m-1} a_j(s) \varphi_j$, where s_m is the orthogonal projection of s on \mathcal{S}_m . Thus \widehat{s}_m is an unbiased estimator of s_m and is called a projection estimator of s . As a consequence, the following decomposition of the MISE (Mean Integrated Squared Error) holds:

$$\mathbb{E} \left[\|\widehat{s}_m - s\|^2 \right] = \|s_m - s\|^2 + \mathbb{E}[\|\widehat{s}_m - s_m\|^2].$$

The estimator \widehat{s}_m can usefully be described as a minimum contrast estimator with respect to the contrast

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{s}_m \rangle, \quad \text{for } t \in \mathcal{S}_m. \quad (2.3)$$

This can be seen by noticing that $\gamma_n(t) = \|t - \widehat{s}_m\|^2 - \|\widehat{s}_m\|^2$.

Now in order to propose an estimator \widetilde{s}_m such that $\widetilde{s}_m(0) = c$, we consider the Lagrange multiplier method and the contrast

$$\widetilde{\gamma}_n(t, \lambda) = \gamma_n(t) - \lambda(t(0) - c) = \|t\|^2 - 2\langle t, \widehat{s}_m \rangle - \lambda(t(0) - c), \quad \text{for } t \in \mathcal{S}_m.$$

Minimizing $\widetilde{\gamma}_n(t, \lambda)$ with respect to $t \in \mathcal{S}_m$, and $\lambda \in \mathbb{R}$ leads to the estimator

$$\widetilde{s}_m = \sum_{j=0}^{m-1} \widetilde{a}_{j,m} \varphi_j, \quad \widetilde{a}_{j,m} = \widehat{a}_j - \widehat{K}_m \varphi_j(0), \quad (2.4)$$

with

$$\widehat{K}_m := \frac{\widehat{s}_m(0) - c}{\sum_{\ell=0}^{m-1} \varphi_\ell^2(0)} = \frac{\omega}{m} \left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell(0) - c \right), \quad (2.5)$$

and $\omega = 1/2$ for the Laguerre basis, $\omega = \mathbf{a}$ for the Fourier basis and odd m . Note that when $m = 1$, $\widetilde{a}_{0,1} = c\sqrt{\omega}$ and $\widetilde{s}_1(x) = c\sqrt{\omega}\varphi_0(x)$, that is $\widetilde{s}_1(x) = ce^{-x}\mathbf{1}_{\mathbb{R}^+}(x)$ in the Laguerre case and $\widetilde{s}_1(x) = c\mathbf{1}_{[0,\mathbf{a}]}(x)$ in the Fourier case. The estimator can also be written

$$\widetilde{s}_m := \widehat{s}_m - \widehat{K}_m \sum_{j=0}^{m-1} \varphi_j(0) \varphi_j, \quad (2.6)$$

with \widehat{K}_m given by (2.5). Clearly, if $\widehat{s}_m(0) = c$ i.e. directly satisfies the constraint, then $\widehat{K}_m = 0$ and $\widetilde{s}_m = \widehat{s}_m$. Otherwise, the new estimator has null risk at 0, i.e. $\widetilde{s}_m(0) = s(0)$, while \widehat{s}_m does not satisfy the constraint. Our aim is to compare the risk bound on \widetilde{s}_m to the one on \widehat{s}_m , globally.

2.3 Risk bound on the constrained estimator

As, under **(A1)**-**(A2)**(s), $s(0) = c = \sum_{\ell \geq 0} a_\ell(s) \varphi_\ell(0)$, we get $\mathbb{E}[\widehat{K}_m] = -(\omega/m) \sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0)$. Therefore, the new estimator is, in mean, a modification of s_m such that $\mathbb{E}[\widetilde{s}_m] = s_m + \mathbb{E}[\widehat{K}_m] \sum_{j=0}^{m-1} \varphi_j(0) \varphi_j$.

To evaluate the quality of this new estimator we prove the following result.

Proposition 2.1. *Under **(A1)**-**(A2)**(s), the MISE of the estimator \widetilde{s}_m of s , given by Equation (2.4) satisfies,*

$$\mathbb{E} \left[\|\widetilde{s}_m - s\|^2 \right] = \mathbb{E}[\|\widehat{s}_m - s\|^2] + B_m - V_{n,m}, \quad (2.7)$$

where

$$B_m := \frac{\omega}{m} \left(\sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right)^2, \quad V_{n,m} := \frac{\omega}{m} \text{Var} \left(\sum_{j=0}^{m-1} \widehat{a}_j \varphi_j(0) \right). \quad (2.8)$$

The proof of Proposition 2.1, as well as most other proofs, is relegated to Section 6.

Equation (2.7) implies that the MISE of \widetilde{s}_m has the same order as the risk of \widehat{s}_m up to two terms:

- B_m which depends only on m and increases the bias,
- $V_{n,m}$ which is of variance type and decreases the variance.

The two terms correspond to the pointwise bias and variance of \widehat{s}_m at 0, divided by m . For well-chosen m , they should have the same order and thus compensate each other, making the risks of \widetilde{s}_m and \widehat{s}_m essentially equal.

We provide hereafter a general study showing that, under mild assumptions, the additional bias term B_m has the same order as $\|s - s_m\|^2$, and thus does not change the order of the global bias (while correcting the bias at 0).

2.4 Bias order on Sobolev spaces

Sobolev regularity spaces are generally of type

$$W^\alpha(I, L) = \left\{ p : I \rightarrow \mathbb{R}, p \in \mathbb{L}^2(I), \sum_{k \geq 0} k^{c_\alpha} a_k^2(p) \leq L < +\infty \right\} \quad \text{with } \alpha \geq 0, c_\alpha \geq 0 \quad (2.9)$$

where $a_k(p) = \langle p, \varphi_k \rangle$. These spaces have been introduced for Laguerre basis, $I = \mathbb{R}^+$ and $c_\alpha = \alpha$ in [6] and the link with the coefficients of a function on a Laguerre basis was studied by [13]. For the Fourier developments, where $I = [0, \mathbf{a}]$ and $c_\alpha = 2\alpha$, they correspond to standard regularity spaces, see [20].

Now for $s \in W^\alpha(I, L)$ defined by (2.9), we have

$$\|s - s_m\|^2 = \sum_{k=m}^{\infty} a_k^2(s) = \sum_{k=m}^{\infty} a_k^2(s) k^{c_\alpha} k^{-c_\alpha} \leq L m^{-c_\alpha}.$$

If in addition, $c_\alpha > 1$, by Cauchy-Schwarz Inequality, under **(A2)**(s), it comes

$$B_m = \frac{1}{m} \left(\sum_{\ell \geq m} a_\ell(s) \right)^2 \leq \frac{1}{m} \sum_{\ell \geq m} \ell^{-c_\alpha} \sum_{\ell \geq m} \ell^{c_\alpha} a_\ell^2(s) \leq \frac{L}{c_\alpha - 1} m^{-c_\alpha},$$

using that $\sum_{\ell \geq m} \ell^{-c_\alpha} \leq m^{1-c_\alpha} / (c_\alpha - 1)$. Therefore, the additional bias term B_m has the same order in m as the standard bias term. The following Corollary summarizes this finding.

Corollary 2.2. *Assume that **(A1)**-**(A2)**(s) hold, and moreover that $s \in W^\alpha(I, L)$ for $c_\alpha > 1$, then the MISE of the estimator \widetilde{s}_m of s , given by Equation (2.4) satisfies,*

$$\mathbb{E} \left[\|\widetilde{s}_m - s\|^2 \right] \leq \frac{2c_\alpha}{c_\alpha - 1} L m^{-c_\alpha} + \mathbb{E}[\|\widehat{s}_m - s_m\|^2]. \quad (2.10)$$

The consequence is that the order of the upper risk bounds of \widetilde{s}_m and \widehat{s}_m are the same, and thus, both estimators have the same rates of convergence. Therefore, the procedure improves the estimator at 0 without degrading the rate of convergence.

2.5 Model selection

Let us describe now the empirical criterion used for selecting a relevant m , that is a value of m leading to a data-driven squared bias-variance tradeoff. We start with a standard procedure associated with \widehat{s}_m and explain how to extend it to \widetilde{s}_m .

With the bases we presented, the spaces \mathcal{S}_m are nested and, to select an adequate dimension, we look for m in a finite set

$$\mathcal{M}_n = \{1, \dots, m_{\max}\},$$

where, in any case, $m_{\max} \leq n$. In the examples, m_{\max} is specifically defined in each context and can depend on n . The contrast function defined by (2.3) can thus be written, for any $t \in \mathcal{S}_m$,

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{s}_{m_{\max}} \rangle. \quad (2.11)$$

This definition has the advantage that it no longer depends on m .

A selection procedure is generally defined for the estimator \widehat{s}_m , by replacing in its risk decomposition the squared bias term $\|s - s_m\|^2$ and the variance bound $\mathbb{E}[\|\widehat{s}_m - s_m\|^2] \leq V(m)$ by estimators. For the first term, according to Pythagoras Theorem, we have: $\|s - s_m\|^2 = \|s\|^2 - \|s_m\|^2$. The constant $\|s\|^2$ is dropped out, and $-\|s_m\|^2$ is estimated by $-\|\widehat{s}_m\|^2$. For the second term, it is estimated by $\widehat{V}(m)$ and a quantity $\widehat{\text{pen}}_1(m) = \kappa \widehat{V}(m)$ is plugged in the criterion, for a numerical constant κ . Note that $\widehat{V}(m)$ is taken equal to $V(m)$, when all terms in the variance bound are known. We get the following model choice:

$$\widehat{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \{-\|\widehat{s}_m\|^2 + \widehat{\text{pen}}_1(m)\},$$

For the new estimator, we introduce an additional term of penalization $\widehat{\text{pen}}_2$ which can also be computed from the observations, and takes the constraint into account as follows:

$$\widehat{\text{pen}}(m) := \widehat{\text{pen}}_1(m) + \widehat{\text{pen}}_2(m), \quad \widehat{\text{pen}}_2(m) := \frac{m}{2\omega} \widehat{K}_m^2 = \frac{\omega}{2m} \left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell(0) - c \right)^2. \quad (2.12)$$

Heuristically, $\widehat{\text{pen}}_2$ contains both B_m and $V_{n,m}$, as given in (2.7)-(2.8). Indeed, we have

$$\mathbb{E}[2\widehat{\text{pen}}_2(m)] = \frac{\omega}{m} \mathbb{E} \left[\left(\sum_{\ell=0}^{m-1} (\widehat{a}_\ell - a_\ell(s)) \varphi_\ell(0) - \sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right)^2 \right] = B_m + V_{n,m}. \quad (2.13)$$

Then we set

$$\widetilde{m} = \underset{m \in \mathcal{M}_n}{\operatorname{argmin}} \{-\|\widehat{s}_m\|^2 + \widetilde{\text{pen}}(m)\}. \quad (2.14)$$

Therefore, the estimate of the bias of \widehat{s}_m given by $-\|\widehat{s}_m\|^2$ is increased by the B_m term contained in $\text{pen}_2(m)$, and thus we get an estimate of the bias of \widetilde{s}_m . The $V_{n,m}$ contribution increases the variance part, but in a negligible way in most specific cases studied hereafter.

Now we can provide general conditions ensuring that both estimators are adaptive in the sense that they automatically realize the squared bias-variance tradeoff.

Theorem 2.3. *Assume that (A1)-(A2)(s) hold. Assume moreover that the centered empirical process defined, for $t \in \mathcal{S}_m$, by*

$$\nu_n(t) := \langle t, \widehat{s}_{m_{\max}} - s_{m_{\max}} \rangle. \quad (2.15)$$

satisfies, for some $a \in \{0, 1, 2\}$,

$$\mathbb{E} \left(\sup_{t \in B_{m, \widehat{m}}} \nu_n^2(t) - \frac{1}{4} \text{pen}_1(m \vee \widehat{m}) \right)_+ \leq C \frac{\log^a(n)}{n}, \quad (2.16)$$

where C is a constant, $B_{m, \widehat{m}} := \{t \in \mathcal{S}_{m \vee \widehat{m}}, \|t\| = 1\}$, and $\text{pen}_1(m)$ is such that

$$\mathbb{E}[\widehat{\text{pen}}_1(m)] \leq 2\text{pen}_1(m) \quad \text{and} \quad \mathbb{E}[(\text{pen}_1(\widehat{m}) - \widehat{\text{pen}}_1(\widehat{m}))_+] \leq C' \frac{\log^a(n)}{n}. \quad (2.17)$$

Then

$$\mathbb{E}[\|\widehat{s}_{\widehat{m}} - s\|^2] \leq 3 \inf_{m \in \mathcal{M}_n} \{\|s - s_m\|^2 + 2\text{pen}_1(m)\} + C'' \frac{\log^a(n)}{n}, \quad (2.18)$$

where C'' is a constant depending on s but not on m or n .

Moreover, the final estimator $\widetilde{s}_{\widetilde{m}}$ defined by (2.4) and (2.14) satisfies the oracle-type inequality:

$$\mathbb{E}[\|\widetilde{s}_{\widetilde{m}} - s\|^2] \leq 2 \inf_{m \in \mathcal{M}_n} \{3\|s - s_m\|^2 + 6\text{pen}_1(m) + \mathbb{E}[\widehat{\text{pen}}_2(m)]\} + C''' \frac{\log^a(n)}{n}, \quad (2.19)$$

where C''' is a constant which does not depend on n and $a \in \{0, 1, 2\}$.

Inequality (2.18) means that the classical projection estimator $\widehat{s}_{\widehat{m}}$ is adaptive and realizes the compromise between the bias term $\|s - s_m\|^2$ and the variance bound $\text{pen}_1(m)$.

The last inequality (2.19) in Theorem 2.3 states the same result for $\widetilde{s}_{\widetilde{m}}$, taking into account the additional terms associated with the risk bound of \widetilde{s}_m . To see this, we can provide a general rough bound on $\mathbb{E}[2\widehat{\text{pen}}_2(m)]$ following from the fact that, by Schwarz inequality, $V_{n,m} \leq \mathbb{E}[\|\widehat{s}_m - s_m\|^2]$, and from (2.13):

$$\mathbb{E}[2\widehat{\text{pen}}_2(m)] \leq \mathbb{E}[\|\widehat{s}_m - s_m\|^2] + \frac{\omega}{m} \left(\sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right)^2 = \mathbb{E}[\|\widehat{s}_m - s_m\|^2] + B_m. \quad (2.20)$$

Thus, the terms in the infimum of (2.19) are the ones involved in the risk of \widetilde{s}_m and the bound states the automatic trade-off resulting from the model selection strategy.

Both results are up to a negligible residual term $\log^a(n)/n$. Note that if $\widehat{\text{pen}}_1(m)$ is deterministic, then we take $\text{pen}_1(m) = \widehat{\text{pen}}_1(m)$ and condition (2.17) is automatically fulfilled.

3 Examples of applications

In this section, we show how the previous procedure applies to two different models. After a short presentation of the context, the definition of the projection estimator is given, together with its constrained version, with notation $(\widehat{a}_j, \widetilde{a}_j)$ for the coefficients of the estimator on the Laguerre or Fourier basis. Then a specific risk bound is provided in each example.

3.1 Survival function estimation in the multiplicative noise model

In this section a multiplicative noise model is considered. The common unknown density and survival function of *i.i.d.* X_i 's are denoted by f and S ($S(x) = \mathbb{P}(X > x)$) respectively, but we observe

$$Y_i = X_i U_i, \quad i = 1, \dots, n, \quad (3.1)$$

where the U_i 's follow a uniform distribution on $[0, 1]$: $U_i \sim \mathcal{U}([0, 1])$, the $(X_i)_{1 \leq i \leq n}$ and the $(U_i)_{1 \leq i \leq n}$ being independent. Our aim is to estimate the survival function S associated with f from observations $(Y_i)_{1 \leq i \leq n}$, when the X_i 's are assumed to be nonnegative (lifetimes data).

The common density f_Y of the *i.i.d.* observations $(Y_i)_{1 \leq i \leq n}$ is given by

$$f_Y(y) = \int_y^{+\infty} \frac{f(x)}{x} dx, \quad y \in]0, +\infty[. \quad (3.2)$$

Moreover, another useful property, implied by the model, see [8], is that for $t : I \rightarrow \mathbb{R}$ a bounded, derivable function, it holds

$$\mathbb{E}[t(Y_1) + Y_1 t'(Y_1)] = \mathbb{E}[t(X_1)]. \quad (3.3)$$

For any $t \in \mathbb{L}^2(\mathbb{R}^+)$, the following bound also holds [see 13]:

$$\mathbb{E}[(Y_1 t(Y_1))^2] \leq \|t\|^2 \mathbb{E}[X_1]. \quad (3.4)$$

Lastly, the key formula obtained from (3.2), is $S_Y(y) = S(y) - y f_Y(y)$, where S_Y is the survival function of Y . Thus, the coefficients of this function on the Laguerre basis are such that $a_j(S) = \langle S, \varphi_j \rangle = \mathbb{E}[Y \varphi_j(Y)] + \langle S_Y, \varphi_j \rangle$. As a consequence, we estimate the projection $S_m = \sum_{j=0}^{m-1} a_j(S) \varphi_j$ of S on \mathcal{S}_m by

$$\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \quad \widehat{a}_j = \frac{1}{n} \sum_{i=1}^n \left[\int_{\mathbb{R}^+} \varphi_j(x) \mathbb{1}_{Y_i \geq x} dx + Y_i \varphi_j(Y_i) \right], \quad (3.5)$$

which obviously satisfies $\mathbb{E}(\widehat{a}_j) = a_j$. Then we can prove the following result.

Proposition 3.1. *If $\mathbb{E}[X_1] < +\infty$, the estimator \widehat{S}_m (3.5) is an unbiased estimator of S_m and it satisfies*

$$\mathbb{E} \left[\left\| \widehat{S}_m - S \right\|^2 \right] \leq \|S_m - S\|^2 + \mathbb{E}[X_1] \frac{m+1}{n}. \quad (3.6)$$

In the case of the Laguerre basis, if $\mathbb{E}[X_1^{3/2}] < +\infty$, then, for any $m \geq 1$,

$$\mathbb{E} \left[\left\| \widehat{S}_m - S \right\|^2 \right] \leq \|S_m - S\|^2 + c^* \mathbb{E} \left[X_1^{3/2} \right] \frac{\sqrt{m}}{n}, \quad (3.7)$$

where c^ is a constant depending on the basis.*

Note that $\mathbb{E}[X_1] < +\infty$ implies that S is integrable and thus S^2 also so that **(A1)**(S) is fulfilled. Moreover, as $\mathbb{E}[Y_1] = \mathbb{E}[X_1] \mathbb{E}[U_1] = \mathbb{E}[X_1]/2$, the integrability condition can be set indifferently on X_1 or Y_1 .

Inequality (3.6) is new; it can be compared to the result obtained in [8], where a kernel estimator is studied. Inequality (3.7) is specific to the Laguerre basis and improves the bound on the variance with respect to (3.6): it has order \sqrt{m}/n instead of m/n under a slightly stronger moment condition

Then, using the steps described in Section 2, we can define a new estimator \widetilde{S}_m of S such that $\widetilde{S}_m(0) = 1$, with (2.4) or (2.6) and \widehat{K}_m given by (2.5) with $c = 1$:

$$\widetilde{S}_m = \widehat{S}_m - \widehat{K}_m \sum_{j=0}^{m-1} \varphi_j(0) \varphi_j, \quad \widehat{K}_m = \frac{\omega}{m} \left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell(0) - 1 \right).$$

Then \widetilde{S}_m satisfies Proposition 2.1 and we can moreover prove here:

Proposition 3.2. *Assume that $\mathbb{E}[Y_1] < +\infty$ and **(A2)**(S) holds. Then $V_{n,m} \leq 4\mathbb{E}[Y_1]/n$ and*

$$\mathbb{E}[\|\widehat{S}_m - S\|^2] + B_m - \frac{4\mathbb{E}[Y_1]}{n} \leq \mathbb{E} \left[\left\| \widetilde{S}_m - S \right\|^2 \right] \leq \mathbb{E}[\|\widehat{S}_m - S\|^2] + B_m,$$

where B_m and $V_{n,m}$ are defined in (2.8).

If $S \in W^\alpha(I, L)$, then it follows from Section 2.4 that, using bound (3.6) under $\mathbb{E}[X_1] < +\infty$ in the Fourier case and bound (3.7) under $\mathbb{E}[X_1^{3/2}] < +\infty$ in the Laguerre case, the adequate choice of the projection space is $m_n = O(n^{1/(2\alpha+1)})$ in the Fourier case and $m_n = O(n^{2/(2\alpha+1)})$

in the Laguerre case (see the values of c_α in Section 2.4). The rate of convergence of \tilde{S}_{m_n} to S is thus of order $n^{-2\alpha/(2\alpha+1)}$ in both cases.

For the selection procedure, we consider the Fourier basis case. We can prove that

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{S}_m\|^2 + \widehat{\text{pen}}_1(m)\}, \quad \widehat{\text{pen}}_1(m) = 2\kappa_1 \bar{Y}_n m/n, \quad \mathcal{M}_n = \{1, \dots, n\}, \quad (3.8)$$

defines an estimate $\hat{S}_{\hat{m}}$ which makes a data driven bias variance trade-off.

This implies that the procedure of Section 2.5 can be applied to select \tilde{m} and build the final estimator $\tilde{S}_{\tilde{m}}$ defined by (2.12)-(2.14).

Theorem 3.3. *Assume that (A1)-(A2)(S) hold and that $\mathbb{E}[X_1^4] < +\infty$. Let $\hat{S}_{\hat{m}}$ be defined by (3.5) and (3.8). Then there exists a constant κ_1^o such that for any $\kappa_1 \geq \kappa_1^o$, we have*

$$\mathbb{E}[\|\hat{S}_{\hat{m}} - S\|^2] \leq 3 \inf_{m \in \mathcal{M}_n} \{\|S - S_m\|^2 + 2\text{pen}_1(m)\} + C \frac{\log^2(n)}{n},$$

where $\text{pen}_1(m) = \kappa_1 \mathbb{E}[Y_1]m/n$ and C is a constant depending on $\mathbb{E}[X_1^4]$. Moreover

$$\mathbb{E}[\|\tilde{S}_{\tilde{m}} - S\|^2] \leq C_1 \inf_{m \in \mathcal{M}_n} \{\|S - S_m\|^2 + B_m + \text{pen}_1(m)\} + C_2 \frac{\log^2(n)}{n},$$

where B_m is defined in (2.8), C_1 is a numerical constant and C_2 is a constant depending on $\mathbb{E}[X_1^4]$.

The proof of Theorem 3.3 relies on the study of the empirical process

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^n \int t(x) \mathbb{1}_{Y_i \geq x} dx + Y_i t(Y_i) - \mathbb{E} \left[\int t(x) \mathbb{1}_{Y_i \geq x} dx + Y_i t(Y_i) \right]$$

which satisfies Assumption (2.16) with $\text{pen}_1(m) = 2\kappa_1 \mathbb{E}[Y_1]m/n$. Moreover, pen_1 and $\widehat{\text{pen}}_1$ satisfy Assumption (2.17) with $a = 2$.

The results can be adapted to the Laguerre with m replaced by \sqrt{m} and \bar{Y}_n by $n^{-1} \sum_{i=1}^n Y_i^{3/2}$ in the penalty. The numerical constant in the penalty is denoted κ_2 .

3.2 Laguerre survival function estimator in the convolution model

Now we explain briefly that the constrained strategy also applies to the convolution model

$$Y_i = X_i + V_i, \quad i = 1, \dots, n \quad (3.9)$$

where the $(X_i)_{1 \leq i \leq n}$ and the $(V_i)_{1 \leq i \leq n}$ are two independent sequences of *i.i.d.* nonnegative random variables. The X_i 's still have unknown density denoted by f and unknown survival function denoted by S , while the V_i 's have known density g . Here, the key property of the Laguerre basis [see 2] is:

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u) \varphi_j(x-u) du = \frac{1}{\sqrt{2}} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x)), \quad (3.10)$$

i.e., the convolution of two basis functions has a linear expression in function of two other basis functions.

For the survival function estimation, it is noticed in [18] that S_Y is the survival function for Y and S_V for V . Using relation (3.10), the convolution equation can be used to write, on the one hand:

$$\begin{aligned} S_Y(x) - S_V(x) &= S \star g(y) = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} a_j(S) a_k(g) \varphi_j \star \varphi_k(x) \\ &= \sum_{k=0}^{\infty} \varphi_k(x) \sum_{\ell=0}^k 2^{-1/2} (a_{k-\ell}(g) - a_{k-\ell-1}(g)) a_\ell(S). \end{aligned} \quad (3.11)$$

On the other hand,

$$S_Y(x) - S_V(x) = \sum_{k=0}^{\infty} (a_k(S_Y) - a_k(S_V)) \varphi_k(x).$$

Therefore, if we define the matrix $\mathbf{G}_m = ([\mathbf{G}_m]_{i,j})_{1 \leq i,j \leq m}$, by

$$[\mathbf{G}_m]_{i,j} = 2^{-1/2} (a_{i-j}(g) - a_{i-j-1}(g)) \text{ if } j \geq i, \text{ 0 otherwise,}$$

and convention $\varphi_j \equiv 0$ if $j < 0$, we have

$$\mathbf{G}_m \overrightarrow{(S_X)}_m = \overrightarrow{(S_Y)}_m - \overrightarrow{(S_V)}_m,$$

where $\overrightarrow{(S_Z)}_m = {}^t(a_0(S_Z), \dots, a_{m-1}(S_Z))$ for $Z = Y, V, U$ and $S_X = S$. The matrix \mathbf{G}_m is known as g is known. An important feature of \mathbf{G}_m is to be lower triangular and Toeplitz. As the diagonal elements $a_0(g) = \sqrt{2}\mathbb{E}[e^{-Y}] > 0$, the matrix \mathbf{G}_m has nonzero determinant and can be inverted. Therefore

$$\overrightarrow{(S_X)}_m = \mathbf{G}_m^{-1} \left(\overrightarrow{(S_Y)}_m - \overrightarrow{(S_V)}_m \right).$$

In view of this and as $a_j(S_Y) = \langle S_Y, \varphi_j \rangle = \mathbb{E}[(I\varphi)_j(Y_1)]$ where $(I\varphi)_j(y) := \int_0^y \varphi_j(x) dx$, the proposed estimator is

$$\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j(S) \varphi_j, \text{ with } \widehat{\overrightarrow{S}}_m = \mathbf{G}_m^{-1} \left(\widehat{\overrightarrow{(S_Y)}}_m - \overrightarrow{(S_V)}_m \right),$$

where $\widehat{\overrightarrow{S}}_m = {}^t(\widehat{a}_0, \dots, \widehat{a}_{m-1})$ and $\widehat{\overrightarrow{(S_Y)}}_m = {}^t(\widehat{a}_0(Y), \dots, \widehat{a}_{m-1}(Y))$,

$$[\widehat{\overrightarrow{(S_Y)}}_m]_j = \widehat{a}_{j-1}(Y) = \frac{1}{n} \sum_{i=1}^n (I\varphi)_{j-1}(Y_i), \text{ } j = 1, \dots, m.$$

The estimator \widehat{S}_m is an unbiased estimator of S_m satisfying the following MISE bound. If S satisfies **(A1)**(S) and $\mathbb{E}[Y_1] < +\infty$, then [see Proposition 3.3 in 18]

$$\mathbb{E}[\|\widehat{S}_m - S\|^2] \leq \|S - S_m\|^2 + \frac{\mathbb{E}[Y_1]}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2.$$

Clearly,

$$\widetilde{S}_m = \widehat{S}_m - \widehat{K}_m \sqrt{2} \sum_{j=0}^{m-1} \varphi_j \text{ with } \widehat{K}_m = (2m)^{-1} \left(\sum_{\ell=0}^{m-1} \sqrt{2} \widehat{a}_\ell - 1 \right)$$

satisfies the bound in Proposition 2.1. Here we can also prove

$$V_{n,m} = \frac{1}{m} \text{Var} \left(\sum_{\ell=0}^{m-1} \hat{a}_\ell \right) \leq \frac{\|\mathbf{G}_m^{-1}\|_F^2 \mathbb{E}[Y_1]}{nm}. \quad (3.12)$$

In the case of a Gamma noise, i.e. if $g \sim \gamma(p, \theta)$, it is known [see 11] that $\|\mathbf{G}_m^{-1}\|_F^2/m = O(m^{2p-1})$ while $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 = O(m^{2p})$ so that the bound (3.12) is such that $V_{n,m}$ is negligible with respect to the variance.

Survival function estimation in the additive convolution model [see 18] relies on the study of the empirical process $\nu_n(t) = \langle t, \hat{S}_{m_{\max}} - S_{m_{\max}} \rangle$ with $\hat{S}_{\hat{m}}$ defined by

$$\widehat{\text{pen}}_1(m) = \frac{2\kappa_3 \bar{Y}_n}{n} \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \log(n)$$

and $\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{S}_m\|^2 + \widehat{\text{pen}}_1(m)\}$ for $\mathcal{M}_n = \{m, \|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \log(n)/n \leq 1\}$. The empirical process fulfills (2.16) and the penalties $\widehat{\text{pen}}_1$ and $\widehat{\text{pen}}_1(m)$ satisfy condition (2.17) with $a = 0$. Thus, here again the procedure of Section 2.5 can be applied to \hat{S}_m .

4 Numerical illustrations

4.1 Description of the practical procedure

In this section, we compare the new constrained estimator to the standard adaptive projection estimator in different cases. We illustrate the two cases of survival function reconstruction presented in Section 3. The samples size is $n = 100$ or $n = 1000$. The variable of interest X is simulated from two different distributions:

- $X \sim \chi^2(10)/\sqrt{20}$ (denoted $X \sim \chi^2$),
- $X \sim 0.5\Gamma(2, 0.4) + 0.5\Gamma(11, 0.5)$ (denoted $X \sim M\Gamma$)

and in the additive model (3.9), we choose to illustrate

- $V \sim \Gamma(2, 1/\sqrt{8})$ (denoted $V \sim \Gamma$),
- $V \sim \exp(2)$.

First, a preliminary calibration step is conducted. The universal constants κ_i , $i = 1, 2, 3$ appearing in each of the procedures are calibrated with a large choice of setups, different from the ones of the simulation study. Empirical MISE are computed via 1000 Monte-Carlo experiments. In the multiplicative noise model, the first penalty ($\widehat{\text{pen}}_1$) is $2\kappa_2(\sum_{i=1}^n Y_i^{3/2})\sqrt{m}/n^2$ with $\kappa_2 = 1$. For the additive model, the penalty is $2\kappa_3(\sum_{i=1}^n Y_i)\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \log(n)/n$ with $\kappa_3 = 0.001$.

Then, we also illustrate the estimation of the density function in both cases with our procedure. Indeed, the two densities have the constraint to be null in 0. We also illustrate the example of the exponential density $\mathcal{E}(2)$ which is 2 in 0. In the densities examples, we assume that the value at 0 is known, a case which may occur in physical experiments or in economics. For the multiplicative case the density estimator is derived from [5] and from [18] for the additive case. In the additive case we illustrate the case of an exponential noise density $V \sim \mathcal{E}(2)$.

4.2 Illustrations

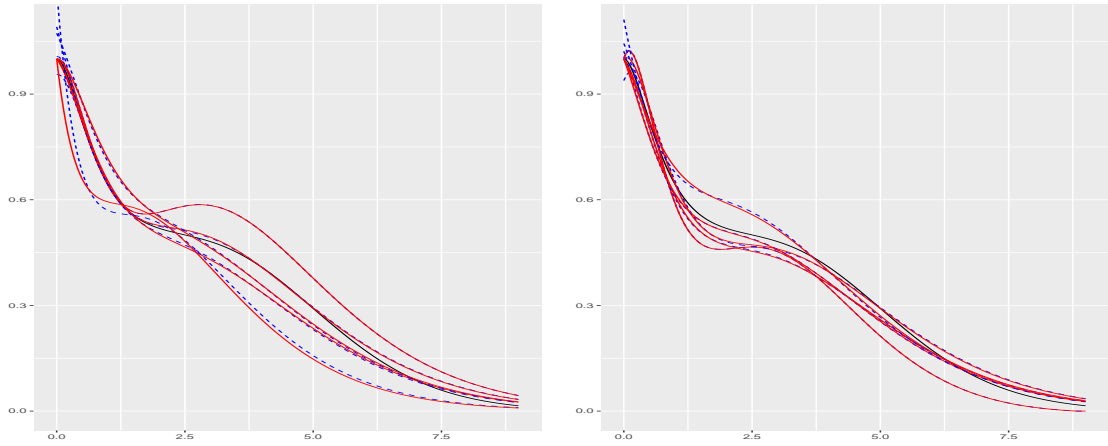


Figure 2: In black bold the true survival function S_X , in dotted blue 5 estimators $\widehat{S}_{\widehat{m}}$, in red 5 estimators $\widetilde{S}_{\widehat{m}}$ with Laguerre basis and $X \sim M\Gamma$ with $n = 100$. Left: multiplicative model (3.1). Right: additive model (3.9) with $V \sim \Gamma$.

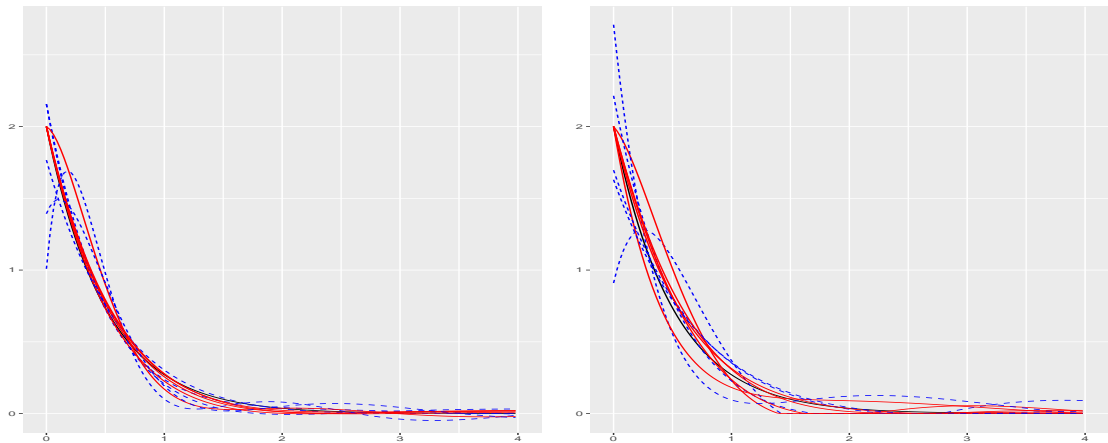


Figure 3: In black bold the true density f_X , in dotted blue 5 estimators $\widehat{f}_{\widehat{m}}$, in red 5 estimators $\widetilde{f}_{\widehat{m}}$ with Laguerre basis, $X \sim \mathcal{E}(2)$, $n = 100$. Left: multiplicative model (3.1). Right: additive model with $V \sim \Gamma$.

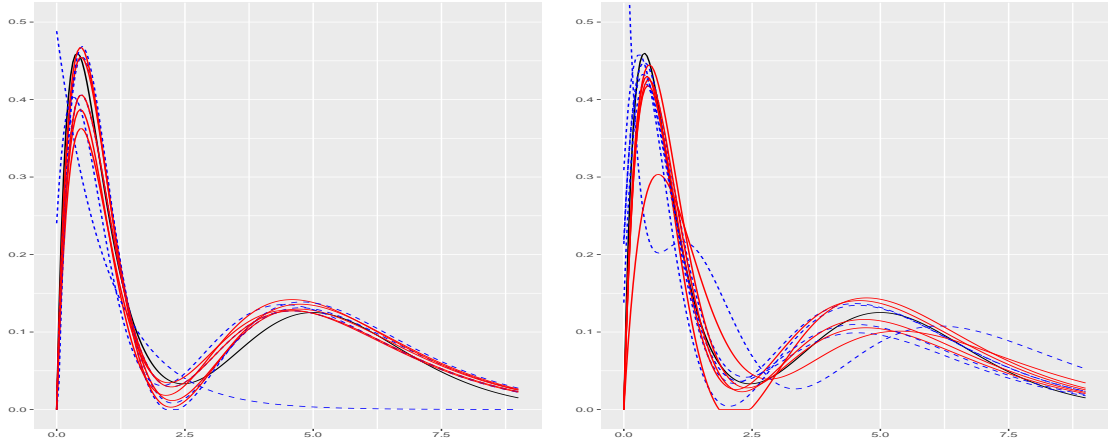


Figure 4: In black bold, the true density f_X , in dotted blue 5 estimators $\hat{f}_{\tilde{m}}$, in red 5 estimators $\tilde{f}_{\tilde{m}}$ with Laguerre basis, for $X \sim M\Gamma$, $n = 100$. Left: multiplicative model (3.1). Right: additive model (3.9) with $V \sim \mathcal{E}(2)$.

Figures 2, 3 and 4 illustrate the estimation procedure detailed in the paper for the multiplicative model (left plots), see (3.1) and the additive noise model (right plots), see (3.9). Figure 2 illustrates the reconstruction of the survival function of X and Figures 3 and 4 show density estimations. We compare estimated survival functions $\hat{S}_{\tilde{m}}$ in dotted blue and corrected estimators $\tilde{S}_{\tilde{m}}$ in red on Figure 2, when X has a mixed Gamma distribution. These graphs show the typical behavior of the procedures. For the additive noise model, all the estimators seem close to the true function S in bold black, even if the sample size is not large ($n = 100$). Nevertheless, the constrained estimator fits better the curve than the classical projection one, near zero. For the multiplicative case (left graphs) the estimation seems harder. The correction clearly improves the behavior of the estimators. These remarks are confirmed by Table 1 below.

Figure 3 presents estimated density functions $\hat{f}_{\tilde{m}}$ in dotted blue and corrected estimators $\tilde{f}_{\tilde{m}}$ in red, when X follows an exponential distribution ($f(0) = 2$); the case of X following a mixed-gamma distribution ($f(0) = 0$) is given in Figure 4. On these graphs, the improvement brought by the estimator $\tilde{f}_{\tilde{m}}$ for two different densities with two different noise distributions, appears clearly.

4.3 Detailed study

n	$X \sim \chi^2$		$X \sim M\Gamma$	
	100	1000	100	1000
$\hat{S}_{\tilde{m}}^{\text{mult}}$	1.993	0.194	3.721	0.455
$\tilde{S}_{\tilde{m}}^{\text{mult}}$	1.966	0.185	3.556	0.423
$\hat{S}_{\tilde{m}}^{\text{add}}$	0.915	0.075	1.539	0.247
$\tilde{S}_{\tilde{m}}^{\text{add}}$	0.955	0.064	1.494	0.244

Table 1: Survival function estimation with Laguerre basis. MISE $\times 100$ with 1000 repetitions. $\hat{S}_{\tilde{m}}^{\text{mult}}$, $\tilde{S}_{\tilde{m}}^{\text{mult}}$: survival function estimators in the multiplicative model. $\hat{S}_{\tilde{m}}^{\text{add}}$, $\tilde{S}_{\tilde{m}}^{\text{add}}$: survival function estimators in the additive case, with $V \sim \Gamma$.

n	$X \sim \chi^2$		$X \sim M\Gamma$	
	100	1000	100	1000
$\hat{f}_{\tilde{m}}^{\text{mult}}$	3.705	0.515	3.719	0.442
$\tilde{f}_{\tilde{m}}^{\text{mult}}$	3.315	0.442	2.678	0.384
$\hat{f}_{\tilde{m}}^{\text{add}}$	0.635	0.185	1.438	0.260
$\tilde{f}_{\tilde{m}}^{\text{add}}$	0.532	0.189	0.821	0.248

Table 2: Density function estimation with Laguerre basis. $\text{MISE} \times 100$ with 1000 repetitions. $\hat{f}_{\tilde{m}}^{\text{mult}}$, $\tilde{f}_{\tilde{m}}^{\text{mult}}$: density estimators in the multiplicative case. $\hat{f}_{\tilde{m}}^{\text{add}}$, $\tilde{f}_{\tilde{m}}^{\text{add}}$: density estimators in the additive case, with noise $V \sim \mathcal{E}(2)$.

Table 1 gives the empirical MISE multiplied by 100, for the survival function estimators computed in the Laguerre basis. As expected, increasing the sample size improves the results: the larger n , the smaller the MISE. Nevertheless, we can see that the new procedure gives better results in most cases: indeed, the estimator $\tilde{S}_{\tilde{m}}$ has smaller risks than the uncorrected projection estimator $\hat{S}_{\tilde{m}}$. The improvement is not large, but it is systematic. Table 2 gives the empirical MISE multiplied by 100, for the density function estimators computed in the Laguerre basis. The same conclusion can be drawn here, with a more pronounced improvement, for example in the bi-modal mixed gamma distribution. This confirms the theoretical result which states that the rate of convergence for the constrained and the non-constrained strategy should have the same order (see Corollary 2.2). But this numerical study goes further and claims that the new estimator performs better and gives a better fit of the estimated function than the unconstrained one.

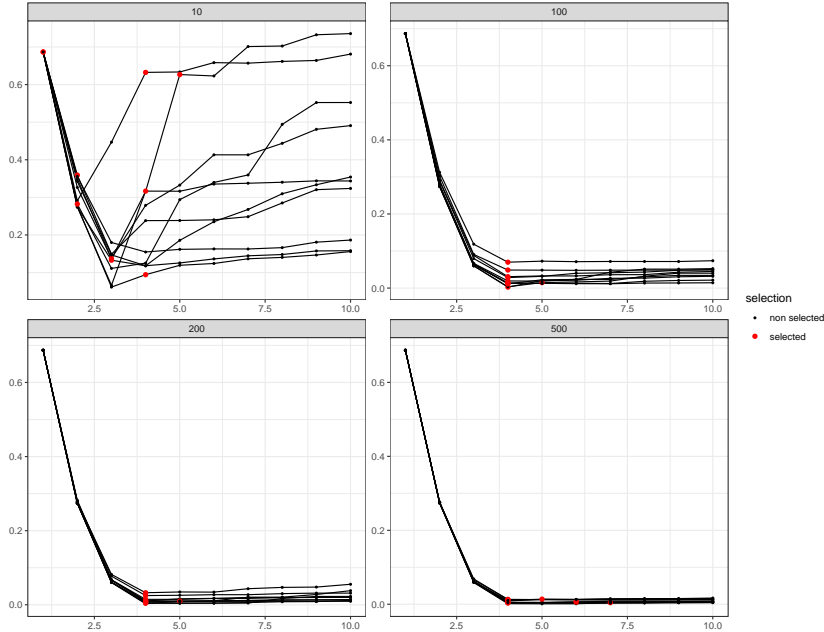


Figure 5: Curves $m \mapsto \|\tilde{S}_m - S\|^2$ for ten simulated samples, with $n = 10$ (top-left), $n = 100$ (top right), $n = 200$ (bottom left), $n = 500$ (bottom right). The selected \tilde{m} is in red. S is the survival function for $X \sim M\Gamma$ in the multiplicative noise case.

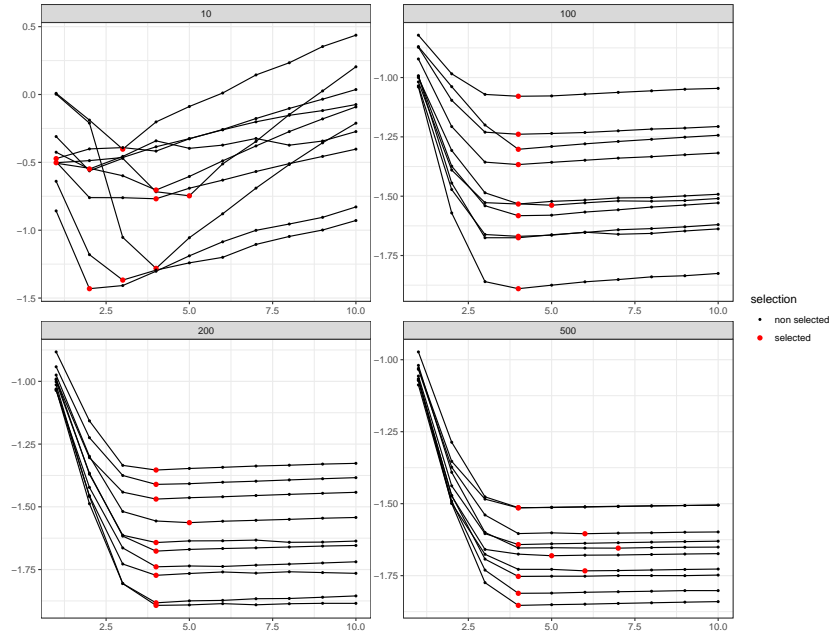


Figure 6: Curves $m \mapsto -\|\tilde{S}_m\|^2 + \widehat{\text{pen}}(m)$ for ten simulated samples, with $n = 10$ (top-left), $n = 100$ (top right), $n = 200$ (bottom left), $n = 500$ (bottom right). The selected \tilde{m} is in red. S is the survival function estimation for $X \sim MT$ in the multiplicative noise case.

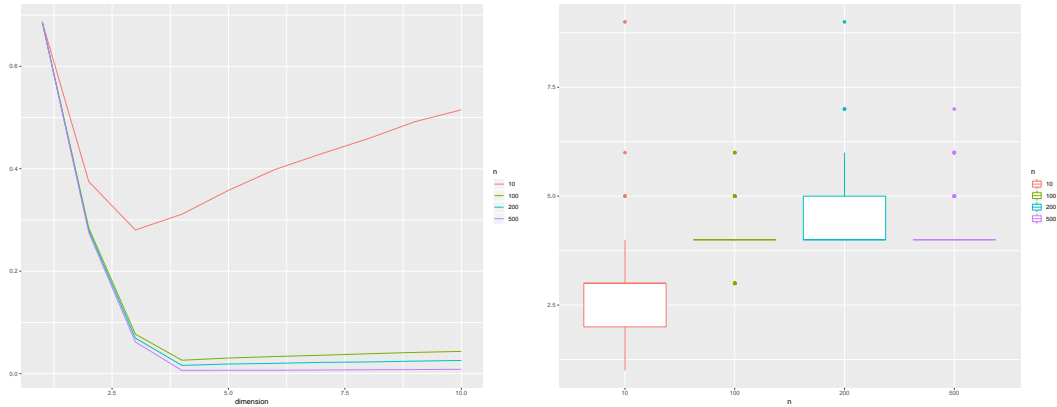


Figure 7: Left: Empirical MISE for the survival function estimation in the multiplicative noise case, for $X \sim MT$, with different values of n . Right: boxplots of the \tilde{m} for the survival function estimation in the multiplicative noise case, for $X \sim MT$, with different values of n and 1000 repetitions.

Let us look more precisely at the selection procedure. In Figure 5, $m \mapsto \|\tilde{S}_m - S\|^2$ is plotted, for 10 samples from the multiplicative model (3.1), with $X \sim MT$, the four picture correspond to different values of $n \in \{10, 100, 200, 500\}$. The red points correspond to the selected dimension for each of the 10 simulations. In comparison, Figure 6 shows the penalized criterion behavior. We see that the criterion has the same behavior as $\|\tilde{S} - S_m\|^2$ and the minimizer of the penalized criterion is often the oracle. This is what was expected. Figure 7-left shows $m \mapsto \mathbb{E}[\|\tilde{S}_m - S\|^2]$, where the expectation is computed as mean over 1000 Monte-Carlo

repetitions (mean behavior of the individual one, shown on Figure 5), each curve corresponds to a value in $n \in \{10, 100, 200, 500\}$. Figure 7-right shows the boxplots for the selected values of m , for the same 4 values of n . The illustrated case is the multiplicative model (3.1), where X is distributed according to the mixed gamma distribution. The oracle of the L^2 -risk is $m^* = 3$ according to the left graph of Figure 7 for $n = 10$ the trajectory of the MISE is easy to minimize and the selection procedure chooses most of the time $\tilde{m} = 2$ or 3. Note that for $n = 100$ and $n = 500$, flat boxes mean that the selected dimension is almost always equal to 4, the oracle value.

5 Extension to regression setting: the example of interval censored data

We mention in this section that the question may be studied for projection estimators in regression framework. In this context, the estimator is built in a different way and is not an unbiased estimator of the projection. To be more precise, we present a specific example: the problem of survival function estimation for interval censored data.

In this model, the variable of interest is still denoted by X , with unknown survival function S and density f . Now, the observations are $(U_i, \delta_i)_{1 \leq i \leq n}$ where $\delta_i = \mathbf{1}_{X_i \leq U_i}$ and both $(X_i)_{1 \leq i \leq n}$ and $(U_i)_{1 \leq i \leq n}$ are i.i.d., the two sequences being independent. Relying on the equality $\mathbb{E}[1 - \delta_i | U_i] = S(U_i)$, a least-squares regression contrast has been proposed in [7] to define a projection estimator of S :

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

We find that $\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j$ with

$$\widehat{\vec{S}}_m = \begin{pmatrix} \widehat{a}_0 \\ \vdots \\ \widehat{a}_{m-1} \end{pmatrix} = \left(\widehat{\Phi}_m \widehat{\Phi}_m \right)^{-1} \widehat{\Phi}_m \overrightarrow{1 - \delta}, \quad \widehat{\Phi}_m = (\varphi_j(U_i))_{1 \leq i \leq n, 0 \leq j \leq m-1}$$

and $\overrightarrow{1 - \delta} = \imath(1 - \delta_1, \dots, 1 - \delta_n)$, provided that $\widehat{\Phi}_m \widehat{\Phi}_m$ is invertible a.s. Let us define two norms specific to this problem:

$$\|t\|_n^2 = \frac{1}{n} \sum_{i=1}^n t^2(U_i) \quad \text{and} \quad \|t\|_U^2 = \int t^2(u) f_U(u) du$$

where f_U is the common density of the U_i 's. Then, it is easy to prove [see 7] that

$$\mathbb{E}[\|\widehat{S}_m - S\|_n^2] \leq \inf_{t \in \mathcal{S}_m} \|t - S\|_U^2 + \frac{1}{4} \frac{m}{n}. \quad (5.1)$$

Note that the risk is computed with respect to the empirical norm $\|\cdot\|_n$, and the bias refers to the weighted norm $\|\cdot\|_U$.

As previously, to obtain an estimator satisfying $\widetilde{S}_m(0) = 1$ (i.e. $c = 1$), we can define

$$\widetilde{S}_m = \underset{t \in \mathcal{S}_m, \lambda \in \mathbb{R}}{\operatorname{argmin}} \{ \gamma_n(t) - \lambda(t(0) - 1) \}.$$

Then we obtain, denoting by $\overrightarrow{\varphi(0)} = {}^t(\varphi_0(0), \dots, \varphi_{m-1}(0))$,

$$\tilde{a}_m = \hat{a}_m - \hat{K}_m \left({}^t\hat{\Phi}_m \hat{\Phi}_m \right)^{-1} \overrightarrow{\varphi(0)}, \quad \hat{K}_m = \frac{\overrightarrow{{}^t\varphi(0)} \widehat{S}_m - 1}{\overrightarrow{{}^t\varphi(0)} \left({}^t\hat{\Phi}_m \hat{\Phi}_m \right)^{-1} \overrightarrow{\varphi(0)}}. \quad (5.2)$$

Note that all additional terms in \tilde{S}_m involve the matrix $\left({}^t\hat{\Phi}_m \hat{\Phi}_m \right)^{-1}$, and this makes a clear difference with respect to the setting of Sections 2 and 3. We can prove a similar result to Proposition 2.1, but with more technicalities:

Proposition 5.1. *Assume that (A1)-(A2)(S) hold, that ${}^t\hat{\Phi}_m \hat{\Phi}_m$ is a.s. invertible, and that f_U is bounded. Then for any $m \leq n/\log^2(n)$,*

$$\mathbb{E}[\|\tilde{S}_m - S\|_n^2] \leq 2\mathbb{E}[\|\hat{S}_m - S\|_n^2] + 2\frac{\mathbb{E}[(\hat{S}_m(0) - S(0))^2]}{m} + \frac{c}{n}$$

where c is a positive constant depending on $\|f_U\|_\infty$.

Proposition 5.1 shows that we recover a bound of the same type as before, but in a less accurate way: the risk of the constrained estimator \tilde{S}_m is bounded by the risk of \hat{S}_m plus the attenuated risk at point 0, $\mathbb{E}[(\hat{S}_m(0) - S(0))^2]/m = B_m + V_{n,m}$, see inequality (2.7)-(2.8) in Proposition 2.1. Therefore, the global correction of \hat{S}_m should not deteriorate the rate.

However, the norms involved in the result are empirical norms, leading to weighted integrated norms when taking expectations. This is what makes the context different from the one presented in Section 2. Moreover, the point-wise risk is also more difficult to handle in this framework.

Model selection is left for further research but would be much more intricate than in the previous class of examples [see 15, for a description of the steps.]

6 Proofs

6.1 Proof of Proposition 2.1

We have the general equality

$$\mathbb{E}[\|\tilde{s}_m - s\|^2] = \|\mathbb{E}[\tilde{s}_m] - s\|^2 + \mathbb{E}[\|\tilde{s}_m - \mathbb{E}[\tilde{s}_m]\|^2]. \quad (6.1)$$

For the bias term, we have $\mathbb{E}[\tilde{s}_m] = s_m + (\omega/m) \left(\sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right) \sum_{j=0}^{m-1} \varphi_j(0) \varphi_j$. As a consequence

$$\begin{aligned} \|\mathbb{E}[\tilde{s}_m] - s\|^2 &= \sum_{j \geq m} a_j^2(s) + \left(\frac{\omega}{m} \right)^2 \left(\sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right)^2 \left(\sum_{k=0}^{m-1} \varphi_k^2(0) \right) \\ &= \sum_{j \geq m} a_j^2(s) + \frac{\omega}{m} \left(\sum_{\ell \geq m} a_\ell(s) \varphi_\ell(0) \right)^2. \end{aligned} \quad (6.2)$$

For the variance term, we have

$$\tilde{s}_m - \mathbb{E}[\tilde{s}_m] = \sum_{j=0}^{m-1} \left[\hat{a}_j - a_j(s) - \frac{\omega \varphi_j(0)}{m} \sum_{\ell=0}^{m-1} (\hat{a}_\ell \varphi_\ell(0) - a_\ell(s) \varphi_\ell(0)) \right] \varphi_j$$

and thus

$$\begin{aligned}
\|\tilde{s}_m - \mathbb{E}[\tilde{s}_m]\|^2 &= \sum_{j=0}^{m-1} \left[\hat{a}_j - a_j(s) - \frac{\omega \varphi_j(0)}{m} \sum_{\ell=0}^{m-1} (\hat{a}_\ell - a_\ell(s)) \varphi_\ell(0) \right]^2 \\
&= \sum_{j=0}^{m-1} (\hat{a}_j - a_j(s))^2 + \sum_{j=0}^{m-1} \frac{\omega^2 \varphi_j^2(0)}{m^2} \left(\sum_{\ell=0}^{m-1} (\hat{a}_\ell - a_\ell(s)) \varphi_\ell(0) \right)^2 - \frac{2\omega}{m} \left(\sum_{\ell=0}^{m-1} (\hat{a}_j - a_j(s)) \varphi_j(0) \right)^2 \\
&= \sum_{j=0}^{m-1} (\hat{a}_j - a_j(s))^2 - \frac{\omega}{m} \left(\sum_{\ell=0}^{m-1} (\hat{a}_\ell - a_\ell(s)) \varphi_\ell(0) \right)^2.
\end{aligned}$$

This yields $\mathbb{E} [\|\tilde{s}_m - \mathbb{E}[\tilde{s}_m]\|^2] = \sum_{j=0}^{m-1} \text{Var}(\hat{a}_j) - \frac{\omega}{m} \text{Var} \left(\sum_{\ell=0}^{m-1} \hat{a}_\ell \varphi_\ell(0) \right)$, and plugging this and (6.2) into (6.1) gives equality (2.7). \square

6.2 Proof of Theorem 2.3

We consider the estimator \hat{s}_m as a minimum contrast estimator associated with γ_n defined by (2.11). It satisfies $\gamma_n(\hat{s}_m) = -\|\hat{s}_m\|^2$, $\hat{s}_m = \underset{t \in \mathcal{S}_m}{\text{argmin}} \gamma_n(t)$. For $\nu_n(\cdot)$ defined by (2.15), it yields the relation for any $m, m' \in \mathcal{M}_n$ and $t \in \mathcal{S}_m$, $u \in \mathcal{S}_{m'}$,

$$\begin{aligned}
\gamma_n(t) - \gamma_n(u) &= \|t - s\|^2 - \|u - s\|^2 + 2\langle t - u, s \rangle - 2\langle t - u, \hat{s}_{m_{\max}} \rangle \\
&= \|t - s\|^2 - \|u - s\|^2 - 2\nu_n(t - u).
\end{aligned} \tag{6.3}$$

According to the relation: $\tilde{s}_m = \hat{s}_m - \hat{K}_m \sum_{j=0}^{m-1} \varphi_j(0) \varphi_j$ our strategy is to prove a result on the estimator $\hat{s}_{\tilde{m}}$ and to deduce one for $\tilde{s}_{\tilde{m}}$. Notice that the definitions of \tilde{m} and \hat{s}_m give that, for all $m \in \mathcal{M}_n$,

$$\gamma_n(\hat{s}_{\tilde{m}}) + \widehat{\text{pen}}(\tilde{m}) \leq \gamma_n(\hat{s}_m) + \widehat{\text{pen}}(m) \leq \gamma_n(s_m) + \widehat{\text{pen}}(m).$$

Therefore, $\gamma_n(\hat{s}_{\tilde{m}}) - \gamma_n(s_m) \leq \text{pen}(m) - \text{pen}(\tilde{m})$ and with (6.3) we get

$$\|\hat{s}_{\tilde{m}} - s\|^2 - \|s - s_m\|^2 - 2\nu_n(\tilde{s}_{\tilde{m}} - s_m) \leq \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\tilde{m}).$$

Therefore, denoting by $\mathcal{B}_{m,m'} = \{t \in \mathcal{S}_{m \vee m'}, \|t\| = 1\}$, we have

$$\begin{aligned}
\|\hat{s}_{\tilde{m}} - s\|^2 &\leq \|s - s_m\|^2 + \widehat{\text{pen}}(m) + 2\nu_n(\hat{s}_{\tilde{m}} - s_m) - \widehat{\text{pen}}_1(\tilde{m}) - \widehat{\text{pen}}_2(\tilde{m}) \\
&\leq \|s - s_m\|^2 + \widehat{\text{pen}}(m) + \frac{1}{4} \|\hat{s}_{\tilde{m}} - s_m\|^2 + 4 \sup_{t \in \mathcal{B}_{m,\tilde{m}}} \nu_n^2(t) - \widehat{\text{pen}}_1(\tilde{m}) - \widehat{\text{pen}}_2(\tilde{m}).
\end{aligned}$$

Writing that $\|\hat{s}_{\tilde{m}} - s_m\|^2 \leq 2\|\hat{s}_{\tilde{m}} - s\|^2 + 2\|s - s_m\|^2$ and gathering the terms implies

$$\begin{aligned}
\frac{1}{2} \|\hat{s}_{\tilde{m}} - s\|^2 &\leq \frac{3}{2} \|s - s_m\|^2 + \widehat{\text{pen}}(m) + 4 \sup_{t \in \mathcal{B}_{m,\tilde{m}}} \left(\nu_n^2(t) - \frac{1}{4} \text{pen}_1(\tilde{m} \vee m) \right)_+ \\
&\quad + \text{pen}_1(\tilde{m} \vee m) - \widehat{\text{pen}}_1(\tilde{m}) - \widehat{\text{pen}}_2(\tilde{m}) \\
&\leq \frac{3}{2} \|s - s_m\|^2 + \widehat{\text{pen}}(m) + 4 \sup_{t \in \mathcal{B}_{m,\tilde{m}}} \left(\nu_n^2(t) - \frac{1}{4} \text{pen}_1(\tilde{m} \vee m) \right)_+ + \text{pen}_1(\tilde{m}) \\
&\quad + \text{pen}_1(m) - \widehat{\text{pen}}_1(\tilde{m}) - \widehat{\text{pen}}_2(\tilde{m})
\end{aligned}$$

as $\text{pen}_1(\tilde{m} \vee m) \leq \text{pen}_1(\tilde{m}) + \text{pen}_1(m)$. Now taking expectation and using Assumption (2.16) and $\mathbb{E}[\widehat{\text{pen}}_1(m)] \leq 2\text{pen}_1(m)$ we get, $\forall m \in \mathcal{M}_n$,

$$\begin{aligned} \mathbb{E}[\|\widehat{s}_{\tilde{m}} - s\|^2] &\leq 3\|s - s_m\|^2 + 6\text{pen}_1(m) + 2\mathbb{E}[\widehat{\text{pen}}_2(m)] + \frac{8C}{n} \\ &\quad + 2\mathbb{E}[\text{pen}_1(\tilde{m}) - \widehat{\text{pen}}_1(\tilde{m})] - 2\mathbb{E}[\widehat{\text{pen}}_2(\tilde{m})]. \end{aligned}$$

This result implies Inequality (2.18): indeed, for $\widehat{\text{pen}}_2(m) = 0$, $\tilde{m} = \widehat{m}$.

As for all $m \in \mathcal{M}_n$, $\|\widehat{s}_m - s\|^2 \leq 2\|\widehat{s}_m - s\|^2 + 2(m/\omega)\widehat{K}_m^2$, taking $m = \tilde{m}$ with $m\widehat{K}_m^2/(2\omega) = \widehat{\text{pen}}_2(m)$ leads to

$$\begin{aligned} \mathbb{E}[\|\widehat{s}_{\tilde{m}} - s\|^2] &\leq 2\mathbb{E}[\|\widehat{s}_{\tilde{m}} - s\|^2] + 4\mathbb{E}[\widehat{\text{pen}}_2(\tilde{m})] \\ &\leq 6\|s - s_m\|^2 + 12\text{pen}_1(m) + 4\mathbb{E}[\widehat{\text{pen}}_2(m)] + \frac{16C}{n} + 4\mathbb{E}[(\text{pen}_1(\tilde{m}) - \widehat{\text{pen}}_1(\tilde{m}))_+]. \end{aligned}$$

According to conditions (2.17) and (2.20), we get for any $m \in \mathcal{M}_n$ the result of Theorem 2.3. \square

6.3 Proof of Proposition 3.1

We study here the risk of the projection estimator of the survival function in the multiplicative model. We have $\mathbb{E}[\|\widehat{S}_m - S\|^2] = \|S_m - S\|^2 + \mathbb{E}[\|\widehat{S}_m - S_m\|^2]$, and we upper bound the term: $\mathbb{E}[\|\widehat{S}_m - S_m\|^2]$.

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{S}_m - S_m \right\|^2 \right] &= \sum_{j=0}^{m-1} \text{Var}(\widehat{a}_j) = \frac{1}{n} \sum_{j=0}^{m-1} \text{Var} \left(\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_1 \geq x} dx + Y_1 \varphi_j(Y_1) \right) \\ &\leq \frac{2}{n} \sum_{j=0}^{m-1} \mathbb{E} \left[(\Phi_j(Y_1) + Y_1 \varphi_j(Y_1))^2 \right] \end{aligned}$$

where $\Phi_j(x) = \int_0^x \varphi_j(u) du$. Now write that $\mathbb{E} \left[(\Phi_j(Y_1) + Y_1 \varphi_j(Y_1))^2 \right] = \mathbb{E}[\Phi_j^2(Y_1) + 2Y_1 \Phi_j(Y_1) \varphi_j(Y_1)] + \mathbb{E}[Y_1^2 \varphi_j^2(Y_1)]$. Now we note that, by Formula (3.3) applied to $t = \Phi_j^2$, the first rhs term is equal to $\mathbb{E}[\Phi_j^2(X_1)]$. Moreover, it follows from (3.4) that $\mathbb{E}[(Y_1 \varphi_j(Y_1))^2] \leq \mathbb{E}[X_1] \|\varphi_j\|^2 = \mathbb{E}[X_1]$. Therefore,

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{S}_m - S_m \right\|^2 \right] &\leq \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left(\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{X_1 \geq x}(x) dx \right)^2 \right] + \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E}[Y_1^2 \varphi_j^2(Y_1)] \\ &\leq \frac{1}{n} \mathbb{E} \left[\sum_{j=0}^{m-1} \langle \varphi_j, \mathbf{1}_{X_1 \geq \cdot} \rangle^2 \right] + \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E}[X_1] \leq \frac{1}{n} \mathbb{E}[\|\mathbf{1}_{X_1 \geq \cdot}\|^2] + \mathbb{E}[X_1] \frac{m}{n} = \mathbb{E}[X_1] \frac{m+1}{n}. \end{aligned}$$

This is the first result of Proposition 3.1.

From Lemma 8.2 in [14], specifically for the Laguerre basis, if $\mathbb{E}[Y_1^{3/2}] < +\infty$, then $\mathbb{E}[Y_1^2 \varphi_j^2(Y_1)] \leq c\mathbb{E}[Y_1^{3/2}]/\sqrt{j}$ for c a constant. This implies that $\sum_{j=0}^{m-1} \mathbb{E}[Y_1^2 \varphi_j^2(Y_1)] \leq c'\mathbb{E}[Y_1^{3/2}]\sqrt{m}$ and thus the second result of Proposition 3.1. \square

6.4 Proof of Proposition 3.2

We study here the additional variance term of the constrained estimator of the survival function in the multiplicative model. As previously, we apply Formula (3.3) applied to $t = (\sum_{j=0}^{m-1} \Phi_j)^2$

(recall that $\Phi_j(x) = \int_0^x \varphi_j(u)du$), and we find

$$\text{Var} \left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell(0) \right) \leq \frac{1}{n} \mathbb{E} \left[\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(0) \varphi_\ell(u) \mathbf{1}_{X_1 \geq u} du \right)^2 \right] + \frac{1}{n} \mathbb{E} \left[\left(\sum_{\ell=0}^{m-1} Y_1 \varphi_\ell(Y_1) \varphi_\ell(0) \right)^2 \right].$$

Now, we write by using Cauchy Schwarz and $\|\sum_{\ell=0}^{m-1} \varphi_\ell(0) \varphi_\ell\|^2 = m/\omega$,

$$\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(0) \varphi_\ell(u) \mathbf{1}_{x \geq u} du \right)^2 \leq \int_{\mathbb{R}^+} \left(\sum_{\ell=0}^{m-1} \varphi_\ell(0) \varphi_\ell(u) \right)^2 du \int_{\mathbb{R}^+} \mathbf{1}_{x \geq u} du = \frac{m}{\omega} x,$$

and we obtain $n^{-1} \mathbb{E}[(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(0) \varphi_\ell(u) \mathbf{1}_{X_1 \geq u} du)^2] \leq \mathbb{E}[X_1]m/(n\omega)$. Besides, we use the property (3.4). It comes,

$$\frac{1}{n} \mathbb{E} \left[\left(Y_1 \sum_{\ell=0}^{m-1} \varphi_\ell(0) \varphi_\ell(Y_1) \right)^2 \right] \leq \frac{\mathbb{E}[X_1]}{n} \left\| \sum_{\ell=0}^{m-1} \varphi_\ell(0) \varphi_\ell \right\|^2 = \mathbb{E}[X_1] \frac{m}{n\omega}.$$

We obtain finally: $\text{Var}(\sum_{\ell=0}^{m-1} \widehat{a}_\ell \varphi_\ell(0)) \leq 2\mathbb{E}[X_1]m/(n\omega) = 4\mathbb{E}[Y_1]m/(n\omega)$, and thus the bound given in Proposition 3.2. \square

6.5 Proof of Theorem 3.3

Let us now prove the oracle-type inequality for the survival function estimator in the multiplicative noise context. We prove that Assumptions (2.16)-(2.17) are fulfilled. The first part of Assumption (2.17) is obviously fulfilled as $\mathbb{E}\widehat{\text{pen}}_1(m) = 2\text{pen}_1(m)$. The second part follows from standard moment inequalities (see e.g. (iv) of Lemma 5.4 in [12]). Now we check (2.16). We set $p(m, m') = \text{pen}(m \vee m')$ and we write

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in B_{m, \widehat{m}}} \nu_n^2(t) - p(m, \widehat{m}) \right]_+ &\leq 3\mathbb{E} \left[\sup_{t \in B_{m, \widehat{m}}} \nu_{n,1}^2(t) - p(m, \widehat{m}) \right]_+ \\ &\quad + 3\mathbb{E} \left[\sup_{t \in B_{m, \widehat{m}}} \nu_{n,2}^2(t) \right] + 3\mathbb{E} \left[\sum_{t \in B_{m, \widehat{m}}} R_n^2(t) \right]_+ \end{aligned}$$

where $\nu_{n,1}(t) = \frac{1}{n} \sum_{i=1}^n (Y_i t(Y_i) \mathbf{1}_{Y_i \leq c_n} - \mathbb{E}[Y_i t(Y_i) \mathbf{1}_{Y_i \leq c_n}])$,

$$\nu_{n,2}(t) = \frac{1}{n} \sum_{i=1}^n (Y_i t(Y_i) \mathbf{1}_{Y_i > c_n} - \mathbb{E}[Y_i t(Y_i) \mathbf{1}_{Y_i > c_n}]),$$

and $R_n(t) = \frac{1}{n} \sum_{i=1}^n \left(\int t(x) \mathbf{1}_{Y_i \geq x} dx - \mathbb{E} \left[\int t(x) \mathbf{1}_{Y_i \geq x} dx \right] \right)$. First, as the models are nested,

$$\mathbb{E} \left[\left(\sup_{t \in B_{m, \widehat{m}}} R_n^2(t) \right)_+ \right] \leq \mathbb{E} \left[\sum_{j=1}^{m_{\max}-1} \left(\int \varphi_j(x) \frac{1}{n} \sum_{i=1}^n (\mathbf{1}_{Y_i \geq x} - S_Y(x)) dx \right)^2 \right] \leq \frac{\mathbb{E}[Y_1]}{n}. \quad (6.4)$$

Next we apply Talagrand's inequality (Theorem A.1 in supplementary material) to $\nu_{n,1}$ and to that aim, we compute H^2, v and M . Denoting $m^* = m \vee m'$, we have $\mathbb{E}[\sup_{t \in B_{m,m'}} \nu_{n,1}^2(t)] \leq \frac{\mathbb{E}[X_1]m^*}{n} = 2\frac{\mathbb{E}[Y_1]m^*}{n} := H^2$. Clearly here, H^2 can be obtained of order \sqrt{m}/n in the Laguerre case, and the proof would hold. Next using (3.4), we get

$$\sup_{\|t\|=1} \text{Var}(Y_1 t(Y_1) \mathbf{1}_{Y_1 \leq c_n}) \leq \sup_{\|t\|=1} \mathbb{E}[Y_1^2 t^2(Y_1)] \leq \sup_{\|t\|=1} \mathbb{E}[X_1] \|t\|^2 = 2\mathbb{E}[Y_1] := v.$$

Lastly, $\sup_{\|t\|=1} \sup_{y \in \mathbb{R}^+} |yt(y) \mathbf{1}_{y \leq c_n}| \leq c_n \sqrt{2m^*} = M$. Then we obtain, by taking $\alpha = 1/4$ in Theorem A.1, that, for $p(m, m') = 3\mathbb{E}[Y_1]m^*/n$

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in B_{m, \hat{m}}} \nu_{n,1}^2(t) - p(m, \hat{m}) \right)_+ \right] &\leq \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu_{n,1}^2(t) - p(m, m) \right)_+ \right] \\ &\leq \frac{C}{n} \sum_{m' \in \mathcal{M}_n} \left(\mathbb{E}[Y_1] e^{-bm^*/4} + c_n^2 \frac{m^*}{n} e^{-c' \sqrt{\mathbb{E}[Y_1] \sqrt{n}/c_n}} \right) \end{aligned}$$

where C and c' are two numerical constants. Therefore, choosing $c_n = (c'/\sqrt{2})\sqrt{\mathbb{E}[Y_1]n}/\log(n)$ and using that $\text{card}(\mathcal{M}_n) \leq n$ and $m^* \leq n$, yields

$$\mathbb{E} \left[\left(\sup_{t \in B_{m, \hat{m}}} \nu_{n,1}^2(t) - p(m, \hat{m}) \right)_+ \right] \leq \frac{K_2}{n}. \quad (6.5)$$

$$\begin{aligned} \text{Now, } \mathbb{E} \left[\sup_{t \in B_{m, \hat{m}}} \nu_{n,2}^2(t) \right] &\leq \mathbb{E} \left[\sum_{j=1}^{m\hat{m}} \nu_{n,2}^2(\varphi_j) \right] \leq \mathbb{E} \left[\sum_{j=1}^{m\max} \nu_{n,2}^2(\varphi_j) \right] \leq \frac{1}{n} \sum_{j=1}^{m\max} \text{Var}(Y_1 \varphi_j(Y_1) \mathbf{1}_{Y_1 > c_n}) \\ &\leq 2\mathbb{E}[Y_1^2 \mathbf{1}_{Y_1 > c_n}] \leq 2\frac{\mathbb{E}[Y_1^4]}{c_n^2} = \frac{\mathbb{E}[Y_1^4] \log^2(n)}{\mathbb{E}[Y_1] n}. \end{aligned} \quad (6.6)$$

Finally gathering (6.4), (6.5), (6.6) leads to (2.16) and, by Theorem 2.3), to the result of Theorem 3.3. \square

6.6 Proof of Inequality (3.12)

Let $\overrightarrow{(I\varphi)}_m = {}^t((I\varphi)_0(Y_1), \dots, (I\varphi)_{m-1}(Y_1))$. The result follows from:

$$\begin{aligned} n\text{Var} \left(\sum_{\ell=0}^{m-1} \hat{a}_\ell \right) &= \text{Var} \left(\sum_{\ell=0}^{m-1} [\mathbf{G}_m^{-1} \overrightarrow{(I\varphi)}_m]_\ell \right) \leq \mathbb{E} \left[\left(\sum_{\ell=0}^{m-1} [\mathbf{G}_m^{-1} \overrightarrow{(I\varphi)}_m]_\ell \right)^2 \right] \\ &\leq \mathbb{E} \left[\left(\sum_{\ell=0}^{m-1} \sum_{j=0}^{m-1} [\mathbf{G}_m^{-1}]_{\ell,j}^2 \sum_{\ell=0}^{m-1} (I\varphi)_\ell^2(Y_1) \right) \right] \leq \|\mathbf{G}_m^{-1}\|_F^2 \mathbb{E} \left[\sum_{\ell=0}^{m-1} (I\varphi)_\ell^2(Y_1) \right] \\ &\leq \|\mathbf{G}_m^{-1}\|_F^2 \mathbb{E}[Y_1]. \quad \square \end{aligned}$$

6.7 Proof of Proposition 5.1

It follows from the definition of the coefficients given in (5.2) that

$$\|\tilde{S}_m - S\|_n^2 \leq 2\|\hat{S}_m - S\|_n^2 + 2\frac{\hat{K}_m^2}{n} \sum_{i=1}^n \left(\sum_{j=0}^{m-1} [({}^t \hat{\Phi}_m \hat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)}]_j \varphi_j(U_i) \right)^2. \quad (6.7)$$

In the following, we denote by $\|\vec{x}\|_{\mathbb{R}^p}$ the euclidean norm of the vector \vec{x} belonging to \mathbb{R}^p .

We study the second term of the right hand side of (6.7),

$$\sum_{i=1}^n \left(\sum_{j=0}^{m-1} [({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)}]_j \varphi_j(U_i) \right)^2 = \|\widehat{\Phi}_m ({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)}\|_{\mathbb{R}^n}^2 = \overrightarrow{\varphi(0)} ({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)},$$

so, using the definition of \widehat{K}_m in (5.2), we want to bound

$$\mathbf{E} := \frac{\widehat{K}_m^2}{n} \sum_{i=1}^n \left(\sum_{j=0}^{m-1} [({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)}]_j \varphi_j(U_i) \right)^2 = \mathbb{E} \left[\frac{1}{n} \frac{(\widehat{S}_m(0) - S(0))^2}{\overrightarrow{\varphi(0)} ({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)}} \right].$$

Let us define, $\Psi_m = (\int \varphi_j \varphi_k f_U)_{0 \leq j, k \leq m-1}$ and $\widehat{\Psi}_m = (1/n) {}^t\widehat{\Phi}_m \widehat{\Phi}_m$, so that $\mathbb{E}[\widehat{\Psi}_m] = \Psi_m$, and let $\|M\|_{\text{op}}$ denote the operator norm of the matrix M , that is $\|M\|_{\text{op}} = \sqrt{\lambda_{\max}({}^tMM)}$ the square root of the largest eigenvalue of tMM . Also, let

$$\Omega_m = \{\|\widehat{\Psi}_m - \Psi_m\|_{\text{op}} \leq \|\Psi_m\|_{\text{op}}/2\}.$$

Our assumptions imply that $\widehat{\Psi}_m$ is invertible, so that this matrix and its inverse are symmetric positive definite. We have

$$n \overrightarrow{\varphi(0)} ({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)} \geq \lambda_{\min}(\widehat{\Psi}_m^{-1}) \|\overrightarrow{\varphi(0)}\|_{\mathbb{R}^m}^2 = \frac{1}{\lambda_{\max}(\widehat{\Psi}_m)} \|\overrightarrow{\varphi(0)}\|_{\mathbb{R}^m}^2. \quad (6.8)$$

First, has the vector basis φ_j are bounded,

$$\begin{aligned} \lambda_{\max}(\widehat{\Psi}_m) &= \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} {}^t\vec{x} \widehat{\Psi}_m \vec{x} = \frac{1}{n} \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} \|\widehat{\Psi}_m \vec{x}\|_{\mathbb{R}^m} \\ &\leq \frac{1}{n} \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} \sum_{i=1}^n \left(\sum_{j=0}^{m-1} x_j \varphi_j(U_i) \right)^2 \leq \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^{m-1} \varphi_j^2(U_i) \leq \frac{m}{\omega}. \end{aligned}$$

On the other hand, as $\lambda_{\max}(\Psi_m) = \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} {}^t\vec{x} \Psi_m \vec{x}$, we get

$$\lambda_{\max}(\Psi_m) = \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} \int \left(\sum_{j=0}^{m-1} x_j \varphi_j(u) \right)^2 f_U(u) du \leq \|f_U\|_{\infty} \sup_{\|\vec{x}\|_{\mathbb{R}^m}=1} \int \left(\sum_{j=0}^{m-1} x_j \varphi_j(u) \right)^2 du = \|f_U\|_{\infty}.$$

Therefore

$$\begin{aligned} \lambda_{\max}(\widehat{\Psi}_m) &= \|\widehat{\Psi}_m\|_{\text{op}} = \|\widehat{\Psi}_m\|_{\text{op}} \mathbf{1}_{\Omega_m^c} + \|\widehat{\Psi}_m\|_{\text{op}} \mathbf{1}_{\Omega_m} \\ &\leq \frac{m}{\omega} \mathbf{1}_{\Omega_m^c} + (\|\widehat{\Psi}_m - \Psi_m\|_{\text{op}} + \|\Psi_m\|_{\text{op}}) \mathbf{1}_{\Omega_m} \\ &\leq \frac{m}{\omega} \mathbf{1}_{\Omega_m^c} + \frac{3}{2} \|\Psi_m\|_{\text{op}} \mathbf{1}_{\Omega_m} \leq \frac{m}{\omega} \mathbf{1}_{\Omega_m^c} + \frac{3}{2} \|f_U\|_{\infty} \mathbf{1}_{\Omega_m} \end{aligned} \quad (6.9)$$

Therefore, using that $\|\overrightarrow{\varphi(0)}\|_{\mathbb{R}^m}^2 = m/\omega$ for our two bases, we get by plugging (6.9) into (6.8) that

$$n \overrightarrow{\varphi(0)} ({}^t\widehat{\Phi}_m \widehat{\Phi}_m)^{-1} \overrightarrow{\varphi(0)} \geq \mathbf{1}_{\Omega_m^c} + \frac{2}{3 \|f_U\|_{\infty}} \frac{m}{\omega} \mathbf{1}_{\Omega_m}.$$

With $(\widehat{S}_m(0) - S(0))^2 \leq 2(\widehat{S}_m(0))^2 + 2$, this implies that

$$\mathbf{E} \leq 2\mathbb{E} \left[\frac{(\widehat{S}_m(0))^2}{\overrightarrow{\widehat{\Psi}_m^{-1}\varphi(0)}} \mathbf{1}_{\Omega_m^c} \right] + 2\mathbb{P}(\Omega_m^c) + \frac{3\omega\|f_U\|_\infty}{2} \frac{\mathbb{E} \left[(\widehat{S}_m(0) - S(0))^2 \right]}{m}. \quad (6.10)$$

Next, we have

$$\begin{aligned} (\widehat{S}_m(0))^2 &= [\overrightarrow{\widehat{\Psi}_m^{-1}\varphi(0)} (\overrightarrow{t\Phi_m\Phi_m})^{-1} \overrightarrow{t\Phi_m\mathbf{1} - \delta}]^2 \leq \|(\overrightarrow{t\Phi_m\Phi_m})^{-1/2} \overrightarrow{\varphi(0)}\|_{\mathbb{R}^m}^2 \|(\overrightarrow{t\Phi_m\Phi_m})^{-1/2} \overrightarrow{t\Phi_m\mathbf{1} - \delta}\|_{\mathbb{R}^n}^2 \\ &\leq \overrightarrow{\varphi(0)} (\overrightarrow{t\Phi_m\Phi_m})^{-1} \overrightarrow{\varphi(0)} \overrightarrow{t\mathbf{1} - \delta} \overrightarrow{t\Phi_m} (\overrightarrow{t\Phi_m\Phi_m})^{-1/2} \overrightarrow{t\Phi_m\mathbf{1} - \delta} \\ &\leq \overrightarrow{\varphi(0)} (\overrightarrow{t\Phi_m\Phi_m})^{-1} \overrightarrow{\varphi(0)} \|\overrightarrow{t\mathbf{1} - \delta}\|_{\mathbb{R}^n}^2 \leq n \overrightarrow{\varphi(0)} (\overrightarrow{t\Phi_m\Phi_m})^{-1} \overrightarrow{\varphi(0)} = \overrightarrow{\varphi(0)} \widehat{\Psi}_m^{-1} \overrightarrow{\varphi(0)}, \end{aligned}$$

where we used that $\|\Phi_m(\overrightarrow{t\Phi_m\Phi_m})^{-1/2} \overrightarrow{t\Phi_m}\|_{\text{op}} = 1$ as it is a projection matrix. Therefore, we get from (6.10),

$$\mathbf{E} \leq 4\mathbb{P}(\Omega_m^c) + \frac{3\omega\|f_U\|_\infty}{2} \frac{\mathbb{E} \left[(\widehat{S}_m(0) - S(0))^2 \right]}{m}. \quad (6.11)$$

Now, using Proposition 3.1 in [15], we get that

$$\mathbb{P}(\Omega_m^c) \leq \mathbb{P}(\|\widehat{\Psi}_m - \Psi_m\|_{\text{op}} > \|f_U\|_\infty/2) \leq \frac{c}{n}$$

for $m \leq n/\log^2(n)$. Plugging this in (6.11) and then in (6.7) implies the result. \square

A Integrated Talagrand's inequality

The following result follows from the Talagrand concentration inequality (see Klein and Rio (2005)).

Theorem A.1. Consider $n \in \mathbb{N}^*$, \mathcal{F} a class at most countable of measurable functions, and $(X_i)_{i \in \{1, \dots, n\}}$ 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 $\alpha > 0$, setting $C(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$, and $b = 1/6$,

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

References

- [1] M. Abbaszadeh, C. Chesneau, and H.I Doosti. Multiplicative censoring: estimation of a density and its derivatives under the l-p risk. *Revstat Statistical Journal*, 11:255–276, 2013.
- [2] M. Abramowitz and I. A Stegun. Handbook of mathematical functions. *Applied Mathematics Series*, 1966.
- [3] K.E. Andersen and M.E. Hansen. Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference*, 98(1–2):137 – 155, 2001.

- [4] Fadoua Balabdaoui and Jon A. Wellner. Estimation of a k -monotone density: limit distribution theory and the spline connection. *Ann. Statist.*, 35(6):2536–2564, 2007.
- [5] D Belomestny, F. Comte, and V Genon-Catalot. Nonparametric laguerre estimation in the multiplicative censoring model. *Electron. J. Statist.*, 10(2):3114–3152, 2016.
- [6] B. Bongioanni and J. L Torrea. What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2):147–172, 2009.
- [7] E. Brunel and F. Comte. Cumulative distribution function estimation under interval censoring case 1. *Electron. J. Stat.*, 3:1–24, 2009.
- [8] E. Brunel, F. Comte, and V. Genon-Catalot. Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST*, 25(3):570–590, 2016.
- [9] R. Carroll and P Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.
- [10] V. Chernozhukov, I. Fernández-Val, and A. Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009.
- [11] F. Comte, C.-A. Cuenod, M. Pensky, and Y. Rozenholc. Laplace deconvolution and its application to dynamic contrast enhanced imaging. *J.R.Stat.Soc.Ser.B.Stat.Methodol.*, 79(1):69–94, 2017.
- [12] F. Comte and C Dion. Nonparametric estimation in a multiplicative censoring model with symmetric noise. *Journal of Nonparametric Statistics*, 28:1–34, 2016.
- [13] F. Comte and V Genon-Catalot. Adaptive laguerre density estimation for mixed poisson models. *Electronic Journal of Statistics*, 9:1113–1149, 2015.
- [14] F. Comte and V Genon-Catalot. Laguerre and hermite bases for inverse problems. *Journal of the Korean Statistical Society*, 47:273–296, 2018.
- [15] F. Comte and V Genon-Catalot. Regression function estimation on non compact support as a partly inverse problem. *The Annals of the Institute of Mathematical Statistics*, page To appear, 2019.
- [16] J Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991.
- [17] Jian Huang and Jon A. Wellner. Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.*, 22(1):3–33, 1995.
- [18] G Mabon. Adaptive deconvolution on the nonnegative real line. *Scandinavian Journal of Statistics*, 44(3):707–740, 2017.
- [19] M. Pensky and B Vidakovic. Adaptative wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053, 1999.
- [20] A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats, 2009.

- [21] B. van Es, C. A. J. Klaassen, and K Oudshoorn. Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *J. Statist. Plann. Inference*, 91(2):295–312, 2000.
- [22] Y Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.