



# Adaptive deconvolution on the nonnegative real line

Gwennaëlle Mabon

► **To cite this version:**

Gwennaëlle Mabon. Adaptive deconvolution on the nonnegative real line. MAP5 2014-33. 2016.  
<hal-01076927v3>

**HAL Id: hal-01076927**

**<https://hal.archives-ouvertes.fr/hal-01076927v3>**

Submitted on 16 Nov 2016

**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.

# ADAPTIVE DECONVOLUTION ON THE NONNEGATIVE REAL LINE

GWENNAËLLE MABON

CREST - ENSAE  
3 avenue Pierre Larousse  
92245 Malakoff, France

MAP5, Université Paris Descartes  
45 rue des Saints-Pères  
75006 Paris, France

**ABSTRACT.** In this paper we consider the problem of adaptive density or survival function estimation in an additive model defined by  $Z = X + Y$  with  $X$  independent of  $Y$ , when both random variables are nonnegative. This model is relevant, for instance, in reliability fields where we are interested in the failure time of a certain material which cannot be isolated from the system it belongs. Our goal is to recover the distribution of  $X$  (density or survival function) through  $n$  observations of  $Z$ , assuming that the distribution of  $Y$  is known. This issue can be seen as the classical statistical problem of deconvolution which has been tackled in many cases using Fourier-type approaches. Nonetheless, in the present case the random variables have the particularity to be  $\mathbb{R}^+$ -supported. Knowing that, we propose a new angle of attack by building a projection estimator with an appropriate Laguerre basis. We present upper bounds on the mean squared integrated risk of our density and survival function estimators. We then describe a nonparametric data driven strategy for selecting a relevant projection space. The procedures are illustrated with simulated data and compared to the performances of more classical deconvolution setting using a Fourier approach. Our procedure achieves faster convergence rates than Fourier methods for estimating these functions.

**Keywords.** Inverse problem. Adaptive estimation. Nonparametric density estimation. Survival function estimation. Laguerre basis. Deconvolution. Mean squared risk.

**AMS Subject Classification 2010:** 62G05, 62G07, 62G99, 62N99.

## 1. INTRODUCTION

**1.1. Model.** It occurs frequently that a variable of interest  $X$  is not directly observed ; instead we have at hand observations of  $Z$ , equal to the sum of  $X$  and another random variable  $Y$ . In many contexts,  $Y$  may be a measurement error, and as such, symmetric or centered. But we can also, in reliability fields, observe the sum of the lifetimes of two components, the second one being well known. In survival analysis,  $X$  can be the time of infection of a disease and  $Y$  the incubation time, and this happens in the so called back calculation problems in AIDS research. In these last two cases, distributions of  $X$  and  $Y$  are  $\mathbb{R}^+$ -supported.

More formally, we consider the following model

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

where the  $X_i$ 's are independent identically distributed (i.i.d.) nonnegative variables with unknown density  $f$  and unknown survival function  $S_X$  where  $S_X(x) = \mathbb{P}[X > x]$ . The  $Y_i$ 's are also i.i.d. nonnegative variables with known density  $g$  and survival function  $S_Y$ . We denote by  $h$  the density of the  $Z_i$ 's and  $S_Z$  its survival function. Moreover the  $X_i$ 's and the  $Y_i$ 's are assumed to be independent. Our target is the estimation of the density  $f$  along with the survival function  $S_X$  of

---

*E-mail address:* [gwennaelle.mabon@ensae.fr](mailto:gwennaelle.mabon@ensae.fr).

*Date:* November 15, 2016.

the  $X_i$ 's when the  $Z_i$ 's are observed. We want to emphasize that the estimation of the survival function  $S_X$  does not rely on the estimation of the density. A particular strategy is established. Thus we are going to show that the assumption of nonnegativity of the random variables is of huge importance for the estimation strategy.

**1.2. Bibliographical context.** The assumptions imply that, in Model (1),  $h(x) = (f \star g)(x)$  where  $(\varphi \star \psi)(x) = \int \varphi(x-u)\psi(u) du$  denotes the convolution product of two arbitrary functions  $\varphi$  and  $\psi$ . This setting matches the setting of convolution models which is classical in nonparametric statistics. [Groeneboom and Wellner \(1992\)](#) have introduced the problem of one-sided errors in the convolution model under monotonicity of the cumulative distribution function (c.d.f.). They derive nonparametric maximum likelihood estimators (NPMLE) of the c.d.f. Some particular cases have been tackled as uniform or exponential deconvolution by [Groeneboom and Jongbloed \(2003\)](#) and [Jongbloed \(1998\)](#) who propose NPMLE of the c.d.f. of the  $X_i$ 's, which have explicit expressions. For other cases [van Es et al. \(1998\)](#) circumvent the lack of explicit expression for the NPMLE by proposing an isotonic inverse estimator. More recently, [van Es \(2011\)](#), in the uniform deconvolution problem, proposes a density estimator and an estimator of the c.d.f. using kernel estimators and inversion formula. All these works are focused on deriving the asymptotic normality of monotonic estimators which are not always explicit. Here we adopt a very different point view since we are concerned by adaptive estimators of the survival and density functions. We do not focus on the monotonicity of our estimators but it can be noted that monotonic transformation of estimators defined in [Chernozhukov et al. \(2009\)](#) can be applied. Those techniques do not degrade theoretical results. In this paper, our method subsumes the existing ones and in this way unifies the approach to tackle the problem of nonnegative variables in the convolution model in a global setting.

Model (1) is also related to the field of mixture models. We cite [Roueff and Rydén \(2005\)](#) and [Rebafka and Roueff \(2015\)](#) who estimate mixtures of Exponential and Gamma, which are included in the present framework. [Rebafka and Roueff \(2015\)](#) use Legendre polynomials to derive their estimators and obtain a procedure relative to a basis of square-integrable functions on a compact interval  $A$ . The drawback of this method is that in practice this estimation interval depends on the data. In a direct problem, data bring information on supports, but this is more difficult in an inverse problem. We do not have this constraint in our method thanks to the Laguerre basis which is a basis of square-integrable functions on  $\mathbb{R}^+$ . Besides we will show that for the mixture gamma case, we can derive excellent convergence rates for our estimators.

More generally the problem of nonnegative variables appears in actuarial or insurance models. Recently, in a financial context, some papers as [Jirak et al. \(2014\)](#) or [Reiß and Selk \(2013\)](#) have addressed the problem of one-sided errors. The first authors are interested in the optimal adaptive estimation in nonparametric regression when the errors are not assumed to be centered anymore, and typically with Exponential density.

Concerning the convolution literature, the problem of recovering the signal distribution  $f$  when it is observed with an additive noise with known error distribution, has been extensively studied. In this context, the noise is more likely to be centered and thus not one-sided. Rates of convergence and their optimality for kernel estimators have been studied in [Carroll and Hall \(1988\)](#), [Stefanski \(1990\)](#), [Stefanski and Carroll \(1990\)](#), [Fan \(1991\)](#) and [Efromovich \(1997\)](#). For the study of sharp asymptotic optimality, we can cite [Butucea \(2004\)](#), [Butucea and Tsybakov \(2008a,b\)](#). For the most part, the adaptive bandwidth selection in deconvolution models has been addressed with a known error distribution, see for example [Pensky and Vidakovic \(1999\)](#) for wavelet strategy, [Delaigle and Gijbels \(2004\)](#) for bandwidth selection, [Comte et al. \(2006\)](#) for projection strategies with penalization, or [Meister \(2009\)](#) and references therein. Concerning the estimation of the c.d.f. in the convolution model, some papers can be found as [Zhang \(1990\)](#), [Fan \(1991\)](#), [Hall and Lahiri \(2008\)](#), [Dattner et al. \(2011\)](#), [Dattner and Reiser \(2013\)](#), [Dattner et al. \(2016\)](#). They all present pointwise estimation procedures since the distribution function is not square-integrable on  $\mathbb{R}$ . Note that the assumption is not so strong for survival functions on  $\mathbb{R}^+$ . The last two papers consider the pointwise estimation of the c.d.f. when the error distribution is unknown under the assumption that the tail of the characteristic function of the measurement error distribution has

a certain decay: polynomial or exponential. These estimators reach the optimal rates under the condition that the target function belongs to a Sobolev space.

All these works suppose that the variables  $X_i$ 's and  $Y_i$ 's are distributed on the real line. Therefore they are still valid when the variables are distributed on the nonnegative real line and we will provide some convergence rates of these estimators and compare them with our estimators to show that a more specific solution performs better.

**1.3. General strategy.** Let us describe now our specific method for the estimation of the density and survival functions when the random variables  $X$  and  $Y$  in Model (1) are  $\mathbb{R}^+$ -supported. We assume all along the paper that  $g$  belongs to  $\mathbb{L}^2(\mathbb{R}^+)$  and either  $f \in \mathbb{L}^2(\mathbb{R}^+)$  when the estimation of  $f$  is under study, or  $S_X \in \mathbb{L}^2(\mathbb{R}^+)$ , when we want to recover the survival function. In both cases, we use a penalized projection method (see [Birgé and Massart \(1997\)](#)). The idea is to expand the density function  $f$  on an appropriate orthonormal basis on  $\mathbb{L}^2(\mathbb{R}^+)$ ,  $(\varphi_k)_{k \geq 0}$ ,

$$f(x) = \sum_{k \geq 0} b_k(f) \varphi_k(x)$$

where  $b_k(f)$  represents the  $k$ -th component of  $f$  in the orthonormal basis and to estimate the  $m$  first ones  $b_0(f), \dots, b_{m-1}(f)$ . To deal with the particularity of nonnegative variables we introduce the Laguerre basis defined by

$$k \in \mathbb{N}, x \geq 0, \quad \varphi_k(x) = \sqrt{2} L_k(2x) e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}. \quad (2)$$

This basis has already been used to estimate a nonnegative function  $f$  in [Comte et al. \(2015\)](#). These authors consider a regression model defined by  $Y_i = f \star g(t_i) + \varepsilon_i$  where  $Y_i$  is observed,  $t_i$  are deterministic times of observation,  $\varepsilon_i$  is subgaussian and  $g$  is known. We can also cite [Vareschi \(2015\)](#) in a similar context with unknown  $g$ . For  $\mathbb{R}^+$ -supported functions, the convolution product writes

$$h(x) = \int_0^x f(u) g(x-u) du \quad (3)$$

and what makes the Laguerre basis relevant, in the previous works and in ours, is the relation

$$\int_0^x \varphi_k(u) \varphi_j(x-u) du = 2^{-1/2} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x)). \quad (4)$$

(see formula 22.13.14 in [Abramowitz and Stegun \(1964\)](#)). From this property, by decomposing  $f$ ,  $g$  and  $h$  on the Laguerre basis, we are able to define a linear transformation of the coefficients of the density function  $f$  to obtain those of  $h$ . More precisely, if we denote by  $\vec{h}_m$  and  $\vec{f}_m$   $m$ -dimensional vectors with coordinates  $b_k(f)$  and  $b_k(h)$ ,  $k = 0, 1, \dots, m-1$  respectively, we prove

$$\vec{h}_m = \mathbf{G}_m \vec{f}_m \quad (5)$$

where  $\mathbf{G}_m$  is a lower triangular invertible matrix depending on the coefficients of  $g$ . As  $g$  is known, so is  $\mathbf{G}_m$ . Thus we can recover the  $m$  first coefficients of  $f$ , from those of  $g$  which are known and those of  $h$  which can be estimated from the  $Z_i$ 's since  $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$ . We then derive the same reasoning for the survival function estimation. Let us point out that we do not integrate the estimator of  $f$  to obtain an estimator of  $S_X$ . Our idea is to directly project  $S_X$  on the Laguerre basis. This enables us to obtain directly the expansion of  $S_X$  on the Laguerre basis and thus its estimator. To our knowledge this a new strategy for the survival function estimation in a deconvolution setting.

**1.4. Outline of the paper.** The estimators are precisely defined and illustrated in Section 2.

We develop in Section 3 a study of the mean integrated squared error of the estimators of the density and survival function. We discuss the resulting rates of convergence of these two estimators. For that we introduce subspaces of  $\mathbb{L}^2(\mathbb{R}^+)$ , called Laguerre-Sobolev spaces with index  $s > 0$  which are defined in [Bongioanni and Torrea \(2009\)](#). This enables us to determine the order of the squared bias terms. This, together with variance order, provides rates of convergence of the estimators of

$f$  belonging to a Laguerre-Sobolev space. We also obtain rates of convergence for estimators of survival function.

In Section 4, we establish a data driven choice by penalization of the dimension  $m$  in our two models and oracle inequalities. For the estimation of the density and survival functions, the methods rely mostly on the fact that we are able to build nested models since the first  $m - 1$  coordinates  $\vec{h}_m$  and  $\vec{f}_m$  are the same as those of  $\vec{h}_{m-1}$  and  $\vec{f}_{m-1}$ . Finally we illustrate these procedures with some simulations and compare our results to those of Comte et al. (2006) in the case of the density estimation.

To sum up this paper is organized as follows. In Section 2, we give the notations, specify the statistical model and estimation procedures for  $f$  and  $S_X$ . In Section 3, we present upper bounds of the  $\mathbb{L}^2$  integrated risk and derive the corresponding rates of convergence. In Section 4, we propose a new adaptive procedure by penalization for the density and survival functions. Besides the theoretical properties of the adaptive estimators are studied. In Section 5, we study of the adaptive estimators through simulation experiments. Numerical results are then presented and compared to the performances in a more classical deconvolution setting using a Fourier approach. The results show that our procedure works significantly better than Fourier methods for estimating  $\mathbb{R}^+$ -supported functions. In the concluding Section 6 we give further possible developments or extensions of the method. All the proofs are postponed to Section 7.

## 2. STATISTICAL MODEL AND ESTIMATION PROCEDURE

**2.1. Notations.** For two real numbers  $a$  and  $b$ , we denote  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ . For two functions  $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$  belonging to  $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ , we denote  $\|\varphi\|$  the  $\mathbb{L}^2$  norm of  $\varphi$  defined by  $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$ ,  $\langle \varphi, \psi \rangle$  the scalar product between  $\varphi$  and  $\psi$  defined by  $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx$ . Let  $d$  be an integer, for two vectors  $\vec{u}$  and  $\vec{v}$  belonging to  $\mathbb{R}^d$ , we denote  $\|\vec{u}\|_{2,d}$  the Euclidean norm defined by  $\|\vec{u}\|_{2,d}^2 = {}^t\vec{u}\vec{u}$  where  ${}^t\vec{u}$  is the transpose of  $\vec{u}$ . The scalar product between  $\vec{u}$  and  $\vec{v}$  is  $\langle \vec{u}, \vec{v} \rangle_{2,d} = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$ . We introduce the spectral norm of a matrix  $\mathbf{A}$ :  $\varrho^2(\mathbf{A}) = \lambda_{\max}({}^t\mathbf{A}\mathbf{A})$  where  $\lambda_{\max}(\mathbf{A})$  is the largest eigenvalue of  $A$  in absolute value. We denote by  $\|\mathbf{A}\|_F = \sum_{i,j} a_{ij}^2$  the Frobenius norm of a matrix  $\mathbf{A}$ .

**2.2. Laguerre basis.** The Laguerre polynomials  $L_k$  defined by (2) are orthonormal with respect to the weight function  $x \mapsto e^{-x}$  on  $\mathbb{R}^+$ . In other words,  $\int_{\mathbb{R}^+} L_k(x)L_{k'}(x)e^{-x} dx = \delta_{k,k'}$  where  $\delta_{k,k'}$  is the Kronecker symbol. Hence  $(\varphi_k)_{k \geq 0}$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R}^+)$ . We remind that the Laguerre basis verifies the following inequality for all integer  $k$

$$\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_{\infty} \leq \sqrt{2}. \quad (6)$$

We also introduce the space  $\mathcal{S}_m = \text{Span}\{\varphi_0, \dots, \varphi_{m-1}\}$ .

**2.3. Projection estimator of the density function.** For a function  $p$  in  $\mathbb{L}^2(\mathbb{R}^+)$ , we denote

$$p(x) = \sum_{k \geq 0} b_k(p)\varphi_k(x) \quad \text{where} \quad b_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u) du.$$

Thus if  $f, g \in \mathbb{L}^2(\mathbb{R}^+)$ ,  $f$  and  $g$  admit an expansion on the Laguerre basis. Since  $X$  and  $Y$  are independent and nonnegative variables, we have a convolution relation between  $h$ ,  $f$  and  $g$ . Plugging into (3) the decomposition on the Laguerre basis of  $f$  and  $g$ , the following equality holds

$$h(x) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} b_k(f)b_j(g) \int_0^x \varphi_k(u)\varphi_j(x-u) du. \quad (7)$$

Now we decompose  $h$  on the Laguerre basis as  $\sum_{k=0}^{\infty} b_k(h)\varphi_k(x)$  and apply Equation (4) to (7). We get for all  $x \in \mathbb{R}^+$

$$\sum_{k=0}^{\infty} b_k(h)\varphi_k(x) = \sum_{k=0}^{\infty} \varphi_k(x) \left( 2^{-1/2}b_k(f)b_0(g) + \sum_{l=0}^{k-1} 2^{-1/2}(b_{k-l}(g) - b_{k-l-1}(g))b_l(f) \right).$$

We finally obtain an infinite triangular system of linear equations. We can write for any  $m$  that  $\vec{h}_m = \mathbf{G}_m \vec{f}_m$  where  $\mathbf{G}_m$  is the lower triangular Toeplitz matrix with elements

$$[\mathbf{G}_m]_{i,j} = \begin{cases} 2^{-1/2} b_0(g) & \text{if } i = j, \\ 2^{-1/2} (b_{i-j}(g) - b_{i-j-1}(g)) & \text{if } j < i, \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

(see [Comte et al. \(2015\)](#)). As

$$b_0(g) = \int_{\mathbb{R}^+} g(u) \varphi_0(u) du = \sqrt{2} \int_{\mathbb{R}^+} g(u) e^{-u} du = \sqrt{2} \mathbb{E}[e^{-Y}] > 0.$$

$\mathbf{G}_m$  is invertible. The principle of a projection method for estimation is to reduce the question of estimating  $f$  to the one of estimating  $f_m$  the projection of  $f$  on  $\mathcal{S}_m$ . Clearly

$$f_m(x) = \sum_{k=0}^{m-1} b_k(f) \varphi_k(x)$$

and  $b_k(f)$  for  $k = 0, \dots, m-1$  can be estimated by Equation (5) since  $\mathbf{G}_m^{-1} \vec{h}_m = \vec{f}_m$ . So, as  $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$ , the projection of  $f$  on  $\mathcal{S}_m$  can be estimated by

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{b}_k(f) \varphi_k(x) \quad \text{with} \quad \hat{f}_m = \mathbf{G}_m^{-1} \hat{h}_m \quad \text{and} \quad \hat{b}_k(h) = \frac{1}{n} \sum_{i=1}^n \varphi_k(Z_i). \quad (9)$$

Let us notice that if  $Y = 0$  a.s. then  $g = \delta_0$ , and we have for any integer  $k$   $b_k(g) = \varphi_k(0) = \sqrt{2}$ . This implies  $\mathbf{G}_m = \mathbf{I}_m$ , with  $\mathbf{I}_m$  the identity matrix. Therefore if there is no additional noise, we are able to estimate  $f_m$  directly from the observations. It means that in Equation (9) we have  $Z_i = X_i$  and

$$\mathbf{G}_m = \mathbf{I}_m \quad \text{and} \quad \hat{f}_m = \hat{h}_m. \quad (10)$$

This projection estimator, in the direct problem, is not new but rather specific to the Laguerre basis. For more details on projection estimators in the direct case for the estimation of density, see Chap. 7 in [Massart \(2003\)](#).

**2.4. Projection estimator of the survival function.** We have to point out that in the case of the projection estimation of the survival function, the estimation of the coefficients in the Laguerre basis is slightly different from the previous estimators. Let us consider for example  $b_k(S_Z)$  the  $k$ -th coefficient of  $S_Z$

$$\begin{aligned} b_k(S_Z) &= \int_{\mathbb{R}^+} S_Z(u) \varphi_k(u) du = \int_{\mathbb{R}^+} \varphi_k(u) \left( \int_u^{+\infty} h(v) dv \right) du \\ &= \int_{\mathbb{R}^+} \left( \int_0^v \varphi_k(u) du \right) h(v) dv = \mathbb{E}[\Phi_k(Z_1)] \end{aligned}$$

with  $\Phi_k$  a primitive of  $\varphi_k$  defined as  $\Phi_k(x) = \int_0^x \varphi_k(u) du$ . We can notice that

$$\Phi_k(x) = \sqrt{2} \int_0^x \sum_{j=0}^k (-2)^j \binom{k}{j} \frac{u^j}{j!} e^{-u} du = \sqrt{2} \sum_{j=0}^k \frac{(-2)^j}{j!} \binom{k}{j} \gamma(j+1, x). \quad (11)$$

where  $\gamma$  is the lower incomplete gamma function defined by formula 6.5.2. in [Abramowitz and Stegun \(1964\)](#). In order to apply a similar method as for the density estimation, let us see how convolution is modified for survival functions with the following Lemma

**Lemma 2.1.** *If  $Z$  is drawn from Model (1), it holds that*

$$S_Z(z) = S_X \star g(z) + S_Y(z), \quad \forall z > 0.$$

*Proof.* Let  $z \geq 0$ , by definition  $S_Z(z) = \mathbb{P}(Z > z)$ , and

$$\begin{aligned} S_Z(z) &= \mathbb{P}(X + Y > z) = \iint \mathbf{1}_{x+y>z} f(x) \mathbf{1}_{x \geq 0} g(y) \mathbf{1}_{y \geq 0} dx dy \\ &= \int \left( \int_{z-y}^{+\infty} f(x) dx \right) g(y) \mathbf{1}_{y \geq 0} \mathbf{1}_{z-y \geq 0} dy + \int \left( \int_0^{+\infty} f(x) dx \right) g(y) \mathbf{1}_{y \geq 0} \mathbf{1}_{z-y \leq 0} dy \\ &= \int_0^z S_X(z-y) g(y) dy + \int_z^{+\infty} g(y) dy \\ &= \int_0^z S_X(z-y) g(y) dy + S_Y(z) = S_X \star g(z) + S_Y(z). \end{aligned}$$

□

We can notice that we have one more term: the survival function of  $Y$ . Nevertheless similarly to the density estimation the coefficients of  $S_X$ ,  $b_k(S_X)$  can also be represented as a solution of an infinite triangular system of linear equations as follows

$$\begin{aligned} S_Z(z) - S_Y(z) &= \sum_{k \geq 0} (b_k(S_Z) - b_k(S_Y)) \varphi_k(z) \\ &= 2^{-1/2} \sum_{k=0}^{\infty} \varphi_k(x) \left( b_k(S_X) b_0(g) + \sum_{l=0}^k (b_{(k-l)}(g) - b_{(k-l-1)}(g)) b_l(S_X) \right). \end{aligned}$$

Now let us define,  $S_{X,m}$  the projection of  $S_X$  on the space  $\mathcal{S}_m$

$$S_{X,m}(x) = \sum_{k=0}^{m-1} b_k(S_X) \varphi_k(x).$$

Thus, with  $\mathbf{G}_m$  defined by Equation (8) and  $\Phi_k$  defined by (11), the projection estimator of  $S_{X,m}$  on the Laguerre basis is given by

$$\begin{aligned} \hat{S}_{X,m}(x) &= \sum_{k=0}^{m-1} \hat{b}_k(S_X) \varphi_k(x) \\ \text{with } \hat{S}_{X,m} &= \mathbf{G}_m^{-1} \left( \hat{S}_{Z,m} - \vec{S}_{Y,m} \right) \quad \text{and} \quad \hat{b}_k(S_Z) = \frac{1}{n} \sum_{i=1}^n \Phi_k(Z_i), \end{aligned} \quad (12)$$

where  $\vec{S}_{Y,m}$  is known since  $b_k(S_Y) = \mathbb{E}[\Phi_k(Y_1)]$  and  $g$  is known.

**Remark 1.** It is worth mentioning that here we do not integrate the estimator of the density  $\hat{f}_m$  to estimate the survival function.

### 3. BOUNDS ON THE $\mathbb{L}^2$ RISK

In this section, we study the integrated risk of our estimators.

**3.1. Upper bounds.** Before stating any results, let us remind that

$$h(x) = \int f(u) g(x-u) \mathbf{1}_{u \geq 0} \mathbf{1}_{x-u \geq 0} du = \int_0^x f(u) g(x-u) du.$$

By Cauchy-Schwarz inequality, we have  $\forall x \in \mathbb{R}^+$ ,  $h(x) \leq \|f\| \|g\|$ . Thus for  $f$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$ , it yields that  $\|h\|_{\infty} \leq \|f\| \|g\| < \infty$ . Moreover, if  $\|g\|_{\infty} < \infty$  then we can bound  $\|h\|_{\infty}$  by  $\|g\|_{\infty}$ . Since  $g$  is known in our model, it can be interesting to switch the two quantities in a simulation setting for instance.

**Proposition 3.1.** *If  $f$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$ , with  $\mathbf{G}_m$  defined by (8) and  $\hat{f}_m$  defined by (9), the following result holds*

$$\mathbb{E} \|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n} \varrho^2(\mathbf{G}_m^{-1}) \wedge \frac{\|h\|_{\infty}}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2. \quad (13)$$



This result can easily be applied to the estimation of the density when  $Y = 0$ :

**Corollary 3.2.** *For  $f \in \mathbb{L}^2(\mathbb{R}^+)$ , in the model without noise defined by (10) we get*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + (2 \wedge \|h\|_\infty) \frac{m}{n}. \quad (14)$$

Finally, we derive the following upper bound for the projection estimator of the survival function.

**Proposition 3.3.** *If  $S_X$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$  and  $\mathbb{E}[Z_1] < \infty$ , for  $\mathbf{G}_m$  defined by (8) and  $\hat{S}_{X,m}$  defined by (12), the following result holds*

$$\mathbb{E}\|S_X - \hat{S}_{X,m}\|^2 \leq \|S_X - S_{X,m}\|^2 + \frac{\mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1}). \quad (15)$$

**Lemma 3.4.**  *$m \mapsto \varrho^2(\mathbf{G}_m^{-1})$  is nondecreasing.*

**Remark 2.** The terms of the right-hand side of Equations (13), (14) and (15) correspond to a squared bias and variance term. Indeed the first one gets smaller when  $m$  gets larger and vice versa for the other one thanks to Lemma 3.4.

**3.2. Rates of convergence.** In order to derive the corresponding rates of convergence of the estimators  $\hat{f}_m$  and  $\hat{S}_{X,m}$  respectively defined by (9) and (12), we need to evaluate the smoothness of the signal along with the order of  $\varrho^2(\mathbf{G}_m^{-1})$ . In the first place, we assume that  $f$  belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{ f : \mathbb{R}^+ \rightarrow \mathbb{R}, f \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s b_k^2(f) \leq L < +\infty \right\} \quad \text{with } s \geq 0 \quad (16)$$

where  $b_k(f) = \langle f, \varphi_k \rangle$ . Bongioanni and Torrea (2009) have introduced Laguerre-Sobolev spaces but the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2015). Indeed, let  $s$  be an integer, for  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  and  $f \in \mathbb{L}^2(\mathbb{R}^+)$ , we have that

$$\sum_{k \geq 0} k^s b_k^2(f) < +\infty$$

is equivalent to the fact that  $f$  admits derivatives up to order  $s-1$  with  $f^{(s-1)}$  absolutely continuous and for  $0 \leq k \leq s-1$ ,  $x^{(k+1)/2} \sum_{j=0}^{k+1} \binom{k+1}{j} f^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$ . For more details we refer to section 7 of Comte and Genon-Catalot (2015). Now for  $f \in W^s(\mathbb{R}^+, L)$  defined by (16),

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} b_k^2(f) = \sum_{k=m}^{\infty} b_k^2(f) k^s k^{-s} \leq L m^{-s}.$$

Before deriving the order of the spectral norm of  $\mathbf{G}_m^{-1}$ , we can already give the rate of convergence in the forward problem when we have no noise.

**Proposition 3.5.** *In the model without noise defined by (10), suppose that  $f$  belongs to  $W^s(\mathbb{R}^+, L)$  defined by (16) and let  $m_{opt} \propto n^{1/(s+1)}$ , then the following holds*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\|f - \hat{f}_{m_{opt}}\|^2 \leq C_1(s, L) n^{-s/(s+1)}$$

where  $C_1(s, L)$  is a positive constant.

Secondly in the deconvolution problem, we must evaluate the variance term of Equations (13) and (15) which means assess the order of  $\varrho^2(\mathbf{G}_m^{-1})$  and  $\|\mathbf{G}_m^{-1}\|_F^2$ . Let us notice that Equation (13) also admits the following upper bound

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{(2 \vee \|h\|_\infty)}{n} (m \varrho^2(\mathbf{G}_m^{-1}) \wedge \|\mathbf{G}_m^{-1}\|_F^2). \quad (17)$$

Now recall that we have the following equivalence between the spectral and Frobenius norms:

$$\frac{1}{\sqrt{m}} \|\mathbf{G}_m^{-1}\|_F \leq \varrho(\mathbf{G}_m^{-1}) \leq \|\mathbf{G}_m^{-1}\|_F \quad (18)$$



Now let us assess the order of these norms in function of  $m$ . [Comte et al. \(2015\)](#) show that under the following conditions on the density  $g$ , we can recover the order of the Frobenius norm and the spectral norm of  $\mathbf{G}_m^{-1}$ . First we define an integer  $r \geq 1$  such that

$$\frac{d^j}{dx^j}g(x) \Big|_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \dots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

And we make the two following assumptions:

**(C1)**  $g \in \mathbb{L}^1(\mathbb{R}^+)$  is  $r$  times differentiable and  $g^{(r)} \in \mathbb{L}^1(\mathbb{R}^+)$ .

**(C2)** The Laplace transform defined by  $G(z) = \mathbb{E}[e^{-zY}]$  of  $g$  has no zero with non negative real parts except for the zeros of the form  $\infty + ib$ .

**Remark 3.** A Gamma distribution of parameter  $p$  and  $\theta$  verifies these **(C1)**-**(C2)** for  $r = p$  ( $r = 1$  for an Exponential). On the contrary an Inverse Gamma distribution does not satisfy **(C1)** because there exists no  $r$  such that the derivative is different from 0 in 0.

**Lemma 3.6** ([Comte et al. \(2015\)](#)). *If Assumptions **(C1)**-**(C2)** are true, then there exists a positive constants  $C_\varrho$  and  $C'_\varrho$  such that*

$$\begin{aligned} C'_\varrho m^{2r} &\leq \varrho^2 (\mathbf{G}_m^{-1}) \leq C_\varrho m^{2r}, \\ C'_\varrho m^{2r} &\leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \leq C_\varrho m^{2r}. \end{aligned}$$

Thus under Assumptions **(C1)**-**(C2)** the spectral and Frobenius norms have the same order. Gathering Equations (17) and (18) the minimum of the two norms is the Frobenius norm.

**Proposition 3.7.** *Assume that  $f$  belongs to  $W^s(\mathbb{R}^+, L)$  defined by (16), that Assumptions **(C1)**-**(C2)** are fulfilled and let  $m_{opt} \propto n^{1/(s+2r)}$ , then*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \hat{f}_{m_{opt}}\|^2 \leq C_2(s, L, C_\varrho) n^{-s/(s+2r)}.$$

where  $C_2(s, L, C_\varrho)$  is a positive constant.

**Proposition 3.8.** *Assume that  $S_X$  belongs to  $W^{s+1}(\mathbb{R}^+, L)$  defined by (16), that Assumptions **(C1)**-**(C2)** are fulfilled and let  $m_{opt} \propto n^{1/(s+2r+1)}$ , then*

$$\sup_{S_X \in W^{s+1}(\mathbb{R}^+, L)} \mathbb{E} \|S_X - \hat{S}_{X,m}\|^2 \leq C_3(s, L, C_\varrho) n^{-(s+1)/(s+2r+1)}.$$

with  $C_3(s, L, C_\varrho)$  is a positive constant.

**Remark 4.** We clearly see that in Propositions 3.5, 3.7 and 3.8 the value of  $m$  that permits to compute the rate of convergence of the estimator depends on the regularity of the function under estimation. So the solution of the best compromise between the squared bias and the variance depends on unknown quantities  $L$  and  $s$ . That is why we consider the problem of data driven selection of  $m$ . Our goal is then to find a procedure that does not require prior information on  $f$  nor  $S_X$  and whose risk automatically reaches the optimal rate.

**Remark 5.** Lower bounds in deconvolution problems on the real line have been studied in [Fan \(1991\)](#) and [Butucea and Tsybakov \(2008a,b\)](#), yet those results cannot be extended to the setting of this paper since we do not consider the same spaces of regularity. Otherwise we can cite [Vareschi \(2015\)](#) who proves lower bounds in the context of a Laplace regression model. But this methodology cannot be applied in our context.

Recently, [Belomestny et al. \(2016\)](#) have studied the problem of the lower bound in Laguerre density estimation from direct observations. In their paper they explain that the difficulty lies in the fact that the density alternative proposal is a density on  $\mathbb{R}^+$ , and particularly that this density is indeed nonnegative. Due to this constraint, they are able to prove if  $s$  is an integer that the rate  $n^{-s/(s+1)}$  is nearly optimal up to a log factor.

**3.3. Some comparisons of convergence rates the Laguerre procedure and the Fourier approach.** In this section we want to emphasize that for at least certain classes of functions the Laguerre procedure achieves better rates of convergence than the estimators computed with Fourier method which are known to be optimal minimax if  $f$  belongs to a Sobolev class.

First let us introduce, the following space of Gamma mixtures :

$$\mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda}) = \left\{ f = \sum_{i=1}^p \lambda_i \gamma_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^p \lambda_i = 1 \quad \text{and} \quad \gamma_i \sim \Gamma(\alpha_i, \theta_i) \right\} \quad (19)$$

with  $\vec{\alpha} = (\alpha_1, \dots, \alpha_p)$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_p)$ , and  $\vec{\lambda} = (\lambda_1, \dots, \lambda_p)$ .

Now the important point of this subsection can be stated.

**Lemma 3.9.** *In Model (1), let  $f$  belongs to  $\mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})$  defined by (19) then*

a) *if  $g \sim \Gamma(q, \mu)$ , we get*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \sum_{i=1}^p \lambda_i C(\alpha_i, \theta_i) \left| \frac{\theta_i - 1}{\theta_i + 1} \right|^{2m} m^{2(\alpha_i - 1)} + (2 \vee \|h\|_\infty) C_\varrho \frac{m^{2q}}{n},$$

b) *if  $g \sim \beta(a, b)$  with  $b > a \geq 1$ , we get that*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \sum_{i=1}^p \lambda_i C(\alpha_i, \theta_i) \left| \frac{\theta_i - 1}{\theta_i + 1} \right|^{2m} m^{2(\alpha_i - 1)} + (2 \vee \|h\|_\infty) C_\varrho \frac{m^{2a}}{n},$$

For this family of densities, we obtain a more specific upper bound on the MISE than the previous section. Indeed if  $f$  belongs to (19) then the bias decays exponentially and not polynomially. For the Laguerre deconvolution, these distributions can be seen as *supersmooth* functions. The bias term is obtained by explicit computations of the coefficients of Gamma type density in the Laguerre basis (see proof in Section 7). Since the bias decays faster than when  $f \in W^s(\mathbb{R}^+, L)$  and the variance has the same order as in the previous section (Gamma density satisfies **(C1)**-**(C2)**), the upper bound for the asymptotic MISE will be smaller. We can deduce that for  $X \sim \text{Exp}(1)$  ( $f \in \mathcal{M}\Gamma(1, 1, 1, 1)$ ), the bias is null. The rate of convergence, in this particular case of inverse problem, reaches the parametric rate in the Laguerre setting.

**Corollary 3.10.** *Under the Assumption of Lemma 3.9, then*

a) *if  $g \sim \Gamma(q, \mu)$ , for  $m_{opt} = c \log n / \log \rho$  with  $c \geq 1$  and  $\rho = \max_i |\lambda_i - 1| / |\lambda_i + 1| \in (0, 1)$  we get*

$$\sup_{f \in \mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})} \mathbb{E}\|f - \hat{f}_{m_{opt}}\|^2 \leq C \frac{(\log n)^{2q}}{n}.$$

b) *if  $g \sim \beta(a, b)$  with  $b > a \geq 1$ , for  $m_{opt} = c \log n / \log \rho$  with  $c \geq 1$  and  $\rho = \max_i |\lambda_i - 1| / |\lambda_i + 1| \in (0, 1)$  we get*

$$\sup_{f \in \mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})} \mathbb{E}\|f - \hat{f}_{m_{opt}}\|^2 \leq C' \frac{(\log n)^{2a}}{n}.$$

where  $C$  and  $C'$  are positive constants.

**Remark 6.** In deconvolution problems, the minimax rates of convergence over Sobolev classes are well understood. The degree of ill-posedness depends on the regularity of the Fourier transforms of densities  $f$  and  $g$ . So two cases are distinguished : Fourier transform decays exponentially (supersmooth case) and Fourier transform decays polynomially (ordinary smooth case). In this context, the problem has been studied by Fan (1993) who proved lower bounds for  $\mathbb{L}^p$ -norms when  $f$  and  $g$  are ordinary smooth, Butucea (2004) when the signal  $f$  is supersmooth and the noise is ordinary smooth, Butucea and Tsybakov (2008a,b) when both the signal and the noise are supersmooth.

A Gamma distribution of parameter  $(p, \mu)$  belongs a Sobolev space of parameter  $(p - 1)/2$ . Thus

in Fourier deconvolution, the minimax rate of convergence is  $n^{-(2p-1)/(2q+2p-1)}$  when  $Y \sim \Gamma(q, \mu)$ . For comparison, Corollary 3.10 proves an upper bound on the risk with Laguerre procedure of order  $(\log n)^{2q}/n$ . Likewise when  $Y$  is a Beta distribution of parameter  $a$  and  $b$  with  $b > a \geq 1$ , the minimax rate is  $n^{-(2p-1)/(2a+2p-1)}$  to be compared to  $(\log n)^{2a}/n$  with Laguerre procedure. Formulas of the estimators in the Fourier setting are given in Section 5.

Thus for the Fourier procedure we find classical rates of convergence of the deconvolution setting which are slower. In the context of nonnegative variables of Gamma type, we recover faster upper bounds on the MISE with the Laguerre method than with a Fourier procedure. We can extend those results to the case of Exponential and Gamma mixtures. This context fits especially fields of survival analysis and duration models.

To conclude, we have illustrated that there exist some distributions that can be estimated faster with the Laguerre procedure than the Fourier procedure. Indeed we can distinguish, as in the Fourier setting, supersmooth (bias decays exponentially) and ordinary smooth (bias decays polynomially) densities. For the Gamma type functions, we can see that they are supersmooth for the Laguerre method but they are not for the Fourier method and the Sobolev spaces associated to the problem. This illustration also shows that Laguerre-Sobolev and Sobolev spaces are different. One is not included in the other. On the other hand we cannot deny that there are surely some distributions in the framework of Model (1) which can be better estimated with the Fourier procedure. Nonetheless we are not able to explicitly compute other coefficients than those of Gamma type densities. In any case the Fourier method is more general and then can be applied when the r.v. are distributed on  $\mathbb{R}$  or  $\mathbb{R}^+$ .

In the next section we provide a procedure to select the dimension which provides the squared bias/variance compromise whatever the underlying true density and noise distributions.

#### 4. MODEL SELECTION

The aim of this section is to provide an integer  $m$  that enables us to compute an estimator of the unknown density or survival function with the  $\mathbb{L}^2$  risk as close as possible to the oracle risk  $\inf_m \mathbb{E} \|f - \hat{f}_m\|^2$  or  $\inf_m \mathbb{E} \|S_X - \hat{S}_{X,m}\|^2$ . We follow the model selection paradigm (see Birgé and Massart (1997), Birgé (1999), Massart (2003)) and choose the dimension of projection spaces  $m$  as the minimizer of a penalized criterion.

**4.1. Adaptive density estimation.** We add the two following assumptions:

$$\text{(A1)} \quad \mathcal{M}_n^{(1)} = \left\{ 1 \leq m \leq d_1, m \varrho^2 (\mathbf{G}_m^{-1}) \leq \frac{n}{\log n} \right\}, \text{ where } d_1 < n \text{ may depend on } n.$$

$$\text{(A2)} \quad \forall b > 0, \sum_{m \in \mathcal{M}_n^{(1)}} \varrho^2 (\mathbf{G}_m^{-1}) e^{-bm} < C(b) < \infty \text{ with } C(b) \text{ uniformly independent of } n.$$

We define the penalty as

$$\text{pen}_1(m) = \frac{\kappa_1}{n} (2m \varrho^2 (\mathbf{G}_m^{-1}) \wedge \log(n) (\|g\|_\infty \vee 1)) \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$$

where  $\kappa_1$  is a numerical constant see our comment after Theorem 4.1.

**Theorem 4.1.** *If  $f$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$ ,  $\|g\|_\infty < \infty$ , let us suppose that (A1)-(A2) are true. Let  $\hat{f}_{\hat{m}_1}$  be defined by (9) and*

$$\hat{m}_1 = \arg \min_{m \in \mathcal{M}_n^{(1)}} \left\{ -\|\hat{f}_m\|^2 + \text{pen}_1(m) \right\}$$

with  $\text{pen}_1$  defined by (4.1), then there exists a positive numerical constant  $\kappa_0$  such that  $\kappa_1 \geq \kappa_0$  then

$$\mathbb{E} \|f - \hat{f}_{\hat{m}_1}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(1)}} \left\{ \|f - f_m\|^2 + \text{pen}_1(m) \right\} + \frac{C}{n}, \quad (20)$$

where  $C$  depends on  $\|f\|$  and  $\|g\|$ .

The constant  $\kappa_0$  can be estimated from the proof, but in practice, values obtained from the theory are generally too large. Thus the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments.

The oracle inequality (20) establishes a non asymptotic oracle bound. It shows that the squared bias variance tradeoff is automatically made up to a multiplicative constant. We have shown in Section 3 that the rates of convergence in deconvolution problems are intricate and depend on the regularity types of the function  $f$  under estimation and the noise density  $g$ . As shown in Section 3, the best compromise between squared bias and variance orders in Equation (13) yields to a certain value of  $m$  function of  $n$  and depending on unknown quantities, and thus cannot be implemented. That is why Equation (20) is of high interest: rates of convergence are reached without requiring to be specified in the framework.

An oracle inequality can still be achieved by considering an estimator of  $\|h\|_\infty$  based on the data with a dimension of order  $\log n$ . Yet we do not state this result for sake of clarity.

**Remark 7.** Note it is common in the literature to assume that the distributions belong to a certain semi-parametric model which is not the case in this paper. In the deconvolution setting with a Fourier approach, papers as Comte et al. (2006) for instance, assume that the Fourier transform of the target and error densities have a particular decay behavior. Here this is replaced the spectral norm of the matrix  $\mathbf{G}_m^{-1}$ . We can notice that if  $\varrho^2(\mathbf{G}_m^{-1}) = O(m^\alpha)$  then for any  $\alpha$ , Assumption (A2) is true. It is satisfied for instance for Gamma distributions. Similarly if  $\varrho^2(\mathbf{G}_m^{-1}) = O(m^\alpha e^{\lambda m^\beta})$  then it is enough that  $\beta < 1$  to ensure (A2). If Assumption (A2) is relaxed then an adaptive procedure can still be obtained with the associated penalty  $\text{pen}(m) = \kappa \log(n)(2m\varrho^2(\mathbf{G}_m^{-1}) \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\text{F}}^2)/n$ .

**4.2. Adaptive survival function estimation.** In this particular framework, we make the two following assumptions:

$$\text{(B1)} \quad \mathcal{M}_n^{(2)} = \left\{ 1 \leq m \leq d_2, \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} \leq C \right\}, \text{ where } d_2 < n \text{ may depend on } n \text{ and } C > 0.$$

$$\text{(B2)} \quad 0 < \mathbb{E}[Z_1^3] < \infty.$$

We define the penalty as

$$\text{pen}_2(m) = \frac{\kappa_2 \mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1}) \log n \quad (21)$$

**Theorem 4.2.** *If  $S_X$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$ , let us suppose that (B1)-(B2) are true. Let  $\widehat{S}_{X, \widehat{m}_2}$  be defined by (12) and*

$$\widehat{m}_2 = \arg \min_{m \in \mathcal{M}_n^{(2)}} \left\{ -\|\widehat{S}_{X,m}\|^2 + \text{pen}_2(m) \right\}$$

*with  $\text{pen}_2$  defined by (21), then there exists a positive numerical constant  $\kappa'_0$  such that  $\kappa_2 \geq \kappa'_0$  then*

$$\mathbb{E}\|S_X - \widehat{S}_{X, \widehat{m}_2}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \left\{ \|S_X - S_{X,m}\|^2 + \text{pen}_2(m) \right\} + \frac{C}{n},$$

*where  $C$  is a constant depending on  $\mathbb{E}[Z_1^3]$ .*

We can also notice that in the penalty associated with this procedure a logarithmic term appears while it was not in the upper bound of Equation (15). Such logarithms often appear in adaptive procedures.

Comments after Theorem 4.1 still hold. This oracle inequality shows that the squared bias variance tradeoff is automatically made. Asymptotically, this ensures that the rates of convergence are reached up to a  $\log n$  factor. To our knowledge this the first time that a global adaptive procedure of the survival function is considered. This result rests upon the particularity of the Laguerre basis which enables to extend the adaptive estimation of the density function to the survival function.

Nevertheless this estimation cannot be computed directly since the penalty depends on the expectation of  $Z$ . A solution is to prove an oracle inequality for a random penalty associated to (21) which is made in the next corollary.

**Corollary 4.3.** *If  $S_X$  and  $g \in \mathbb{L}^2(\mathbb{R}^+)$ , let us suppose that **(B1)**-**(B2)** are true. Let  $\widehat{S}_{X, \widetilde{m}_2}$  be defined by (12) and*

$$\widetilde{m}_2 = \arg \min_{m \in \mathcal{M}_n^{(2)}} \left\{ -\|\widehat{S}_{X,m}\|^2 + \widehat{\text{pen}}_2(m) \right\} \quad (22)$$

$$\widehat{\text{pen}}_2(m) = \frac{2\kappa_2 \bar{Z}_n}{n} \varrho^2 (\mathbf{G}_m^{-1}) \log n \quad \text{where} \quad \bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad (23)$$

then there exists a positive numerical constant  $\kappa_2$  such that

$$\mathbb{E} \|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \left\{ \|S_X - S_{X,m}\|^2 + \text{pen}_2(m) \right\} + \frac{C}{n}$$

where  $C$  is a constant depending on  $\mathbb{E}[Z_1]$ ,  $\mathbb{E}[Z_1^3]$  and  $\text{Var}[Z_1]$ .

## 5. ILLUSTRATIONS

The whole implementation is conducted using R software. The integrated squared errors  $\|f - \widehat{f}_{\widehat{m}_1}\|^2$  and  $\|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2$  are computed via a standard approximation and discretization (over 300 points) of the integral on an interval of  $\mathbb{R}$  respectively denoted by  $I_f$  and  $I_S$ . Then the mean integrated squared errors (MISE)  $\mathbb{E}\|f - \widehat{f}_{\widehat{m}_1}\|^2$  and  $\mathbb{E}\|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2$  are computed as the empirical mean of the approximated ISE over 500 simulation samples.

**5.1. Simulation setting.** The performance of the procedure is studied for the seven following distributions for  $X$ . All the densities are normalized with unit variance except the Pareto distribution which has infinite variance.

- ▷ Exponential  $\mathcal{E}(1)$ ,  $I_f = [0, 5]$ ,  $I_S = [0, 10]$ .
- ▷ Gamma distribution :  $2 \cdot \Gamma(4, \frac{1}{4})$ ,  $I_f = [0, 10]$ ,  $I_S = [0, 5]$ .
- ▷ Gamma distribution :  $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$ ,  $I_f = [0, 13]$ ,  $I_S = [0, 5]$ .
- ▷ Rayleigh distribution with  $\sigma^2 = 2/(4 - \pi)$ ,  $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ ,  $I_f = [0, 5]$ ,  $I_S = [0, 25]$ .
- ▷ Weibull,  $X/\sqrt{\Gamma(1 + 4/3) - \Gamma(1 + 2/3)^2}$ ,  $f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \mathbb{1}_{x \geq 0}$ , with  $k = \frac{3}{2}$  and  $\lambda = 1$ ,  $I_f = [0, 5]$ ,  $I_S = [0, 5]$ .
- ▷ Mixed Gamma distribution :  $X = W/\sqrt{2.96}$ , with  $W \sim 0.4\Gamma(2, 1/2) + 0.6\Gamma(16, 1/4)$ ,  $I_f = [0, 5]$ ,  $I_S = [0, 10]$ .
- ▷ Chi-squared distribution with 10 degrees of freedom,  $\chi^2(10)/\sqrt{20}$ ,  $I_f = [0, 10]$ ,  $I_S = [0, 10]$ .
- ▷ Pareto distribution with shape parameter  $\alpha = 2$  and scale parameter  $x_m = 1$ ,  $I_f = [0, 5]$ ,  $I_S = [0, 10]$ .

Exponential and Weibull distributions are often used in survival and failure analysis. The Gamma distribution is also often used in insurance modelization. The Rayleigh distribution arises in wind velocity analysis for instance.

In the simulation, the variance  $\sigma^2$  of the error distribution  $g$  takes the values 0, 1/10 and 1/4. The case where the variance  $\sigma^2$  is null, which corresponds to the case  $Y = 0$ , is used as a benchmark for the quality of the estimation in the model with noise. We are not aware of any other specific global method of deconvolution on the nonnegative real line. In that case for the density function, we use our procedure with  $\mathbf{G}_m = \mathbf{I}_m$ . Concerning the survival function, we simply compute the empirical estimator  $S_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i > x\}$  (since  $Y = 0$ ) which reaches the parametric rate of convergence.

We then choose a Gamma distribution for the error distribution which verifies **(C1)**-**(C3)** for  $r = 2$ :

- ▷ Gamma noise:  $\Gamma(2, \frac{1}{\sqrt{20}})$  and  $\Gamma(2, \frac{1}{\sqrt{8}})$ .

Thus the first Gamma distribution has a variance 1/10 and the second 1/4. We refer to Equation (25) for the computation of the matrix  $\mathbf{G}_m$ .

**5.2. Practical implementation of the estimators.** The adaptive procedure is then implemented as follows:

- ▷ For  $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$ , compute  $-\|\hat{f}_m\|^2 + \text{pen}_1(m)$ .
- ▷ Choose  $\hat{m}$  such that  $\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{f}_m\|^2 + \text{pen}_1(m)\}$ .
- ▷ And compute  $\hat{f}_{\hat{m}}(x) = \sum_{k=0}^{\hat{m}-1} \hat{b}_k(f) \varphi_k(x)$ .

The procedure is given for the density estimation. For the survival case the three steps are the same with the right quantities associated to the problem and described in Section 4.2. Besides, the penalties are chosen according to Theorem 4.1 and Corollary 4.3. The constant calibrations were done with intensive preliminary simulations. We take  $\kappa_1 = 0.03$  and  $\kappa_2 = 0.065$ . It can be noted that the values of  $\kappa_1$  and  $\kappa_2$  are much smaller than what comes in theory. For the case of direct observations, we take the calibration constant equal to 0.1. We consider the two following model collections  $\mathcal{M}_n^{(1)} = \{m \in \llbracket 1, n-1 \rrbracket, 1 \leq m \leq \lfloor n^{1/5} \rfloor\}$  and  $\mathcal{M}_n^{(2)} = \{m \in \llbracket 1, n-1 \rrbracket, 1 \leq m \leq \lfloor (n \log n)^{1/4} \rfloor\}$  for the density and survival function estimation.

In order to measure the performances of our procedure (density estimation), we also compute the MISE obtained when using Fourier deconvolution approach. More precisely, we apply the procedure of Comte et al. (2006). It corresponds to a projection method with a  $\mathbb{R}$ -supported sinus cardinal basis or kernel. Besides this procedure is minimax optimal if  $f$  belongs to a Sobolev class in the case of a known ordinary smooth error distribution. We therefore compute the following estimator and penalty. Let  $g^*$  be the Fourier transform of  $g$  defined as  $g^*(x) = \int e^{iux} g(u) du$ . For a Gamma distribution of parameter  $p$  and  $\theta$ , its Fourier transform is  $g^*(u) = (1 - iu\theta)^{-p}$ . We compute

$$\hat{f}_{\text{Fo},m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{n^{-1} \sum_{j=1}^n e^{iuZ_j}}{g^*(u)} du \quad \text{pen}_{\text{Fo}}^{(1)}(m) = \frac{\kappa_{\text{Fo}}^{(1)}}{2\pi n} \int_{-\pi m}^{\pi m} \frac{du}{|g^*(u)|^2}$$

We select  $m$  by minimizing  $-\|\hat{f}_{\text{Fo},m}\|^2 + \text{pen}_{\text{Fo}}^{(1)}(m)$ . If  $Y = 0$ , we set  $g^* \equiv 1$  and  $\text{pen}_{\text{Fo}}^{(2)}(m) = \kappa_{\text{Fo}}^{(2)} m/n$ . The model collection is  $\{m/10 : m \in \mathbb{N}, 1 \leq m \leq 50\}$ . After calibration we find  $\kappa_{\text{Fo}}^{(1)} = 41$  and  $\kappa_{\text{Fo}}^{(2)} = 5$ . We consider two different penalties since in the case of the model without noise, the estimator  $\hat{f}_{\text{Fo},m}$  of  $f_m$  can be computed directly without approximating the integral.

Both procedures (Laguerre and Fourier) are fast.

**5.3. Simulation results.** The results are given in Tables 1 and 2. For both tables, the values of the MISE are multiplied by 100 for each case and computed from 200 simulated data. In Table 1 the abbreviations *Lag* and *Fou* correspond respectively to the Laguerre method and Fourier method of Comte et al. (2006). First we see that the risk decreases when the sample size increases. Likewise, the risk increases when the variance of the noise increases. If  $Y = 0$  i.e.  $\sigma^2 = 0$ , we see that the Laguerre procedure in the direct problem has better performances than the Fourier procedure. For instance, when  $n = 2000$  the MISE in the Fourier setting is at least twice larger than the Laguerre for the Gamma, Rayleigh, mixed Gamma and Chi-squared distributions between the Laguerre and Fourier methods. For the Exponential density estimation, the ratio of the MISE of Fourier divided by Laguerre is equal to 15 in average and for the Weibull distribution is equal to 60 in average. If  $\sigma^2$  equals 1/10 or 1/4, we can make the same kind of remarks in favor of the Laguerre procedure.

Let us concentrate on the Pareto distribution. This distribution contrarily to the others does not have a density close to the Laguerre basis. In the model without noise, we see that the Fourier procedure is better. For small sample size the risks are very close. When the sample size increases, the risk of the Laguerre estimator decreases very slowly while the risk of the Fourier estimator is divided by two. On the other hand, in the deconvolution setting the Laguerre procedure performs better than the Fourier method especially for  $\sigma^2 = 1/4$ .



		$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
		200	2000	200	2000	200	2000
Exponential	Lag	0.874	0.118	2.669	0.247	2.436	0.799
	Fou	3.950	3.234	6.002	3.359	11.23	3.889
Gamma $\Gamma(4, 1/4)$	Lag	0.243	0.048	1.277	0.125	2.921	0.292
	Fou	0.585	0.076	2.027	0.250	8.497	0.860
Gamma $\Gamma(20, 1/2)$	Lag	0.332	0.035	0.576	0.054	0.912	0.088
	Fou	0.521	0.059	1.917	0.245	2.728	1.092
Rayleigh	Lag	0.287	0.044	1.297	0.128	2.887	0.304
	Fou	0.498	0.073	1.546	0.248	7.523	0.797
Weibull	Lag	0.467	0.062	1.505	0.151	3.924	0.427
	Fou	7.004	6.611	8.751	6.839	14.23	7.421
Mixed Gamma	Lag	0.705	0.083	1.235	0.126	3.237	0.306
	Fou	1.548	0.817	7.392	0.817	1.499	0.814
Chi-squared	Lag	0.357	0.037	1.102	0.122	2.506	0.263
	Fou	0.542	0.069	1.887	0.250	8.135	0.844
Pareto	Lag	10.72	10.78	11.94	10.66	13.72	10.82
	Fou	12.42	6.542	18.70	9.387	30.06	26.28

TABLE 1. Results of simulation: MISE  $\mathbb{E} \left( \|f - \hat{f}_{\hat{m}}\|^2 \right) \times 100$  averaged over 500 samples.  $\sigma^2$  denotes the level of variance of the noise.  $\sigma^2 = 0$  corresponds to the model without noise ( $Y = 0$ ). The noise is  $\Gamma(2, \frac{1}{\sqrt{20}})$  with  $\sigma^2 = \frac{1}{10}$  and  $\Gamma(2, \frac{1}{\sqrt{8}})$  with  $\sigma^2 = \frac{1}{4}$  respectively.

Thus the results point out the relevance of a specific method for nonnegative variables in a deconvolution problem.

		$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
		200	2000	200	2000	200	2000
Exponential		0.262	0.022	0.122	0.014	0.134	0.012
Gamma		0.263	0.024	0.688	0.203	1.363	0.233
Rayleigh		0.813	0.115	0.878	0.199	1.336	0.297
Chi-squared		0.310	0.027	1.313	0.117	1.445	0.679

TABLE 2. Results of simulation: MISE  $\mathbb{E} \left( \|S_X - \hat{S}_{X, \tilde{m}_2}\|^2 \right) \times 100$  averaged over 500 samples.  $\sigma^2$  denotes the level of variance of the noise.  $\sigma^2 = 0$  corresponds to the model without noise ( $Y = 0$ ). The noise is  $\Gamma(2, \frac{1}{\sqrt{20}})$  with  $\sigma^2 = \frac{1}{10}$  and  $\Gamma(2, \frac{1}{\sqrt{8}})$  with  $\sigma^2 = \frac{1}{4}$  respectively.

In Table 2, the first two columns correspond to the estimation with the empirical estimator of the survival function if we observe directly the data. The estimation is very good: this was expected since the estimator converges to the true function with rate  $\sqrt{n}$ . Yet for the estimation of the Exponential distribution we note that the penalization procedure always beats the empirical estimator  $S_n$ . It is also the case for the density estimation. It is explained by the fact that the Exponential density with parameter 1 corresponds to the first function of the basis. We notice that the risk decreases when the sample size increases. For the Exponential distribution, it is divided by 10, by 3.5 for the Gamma distribution, by 4.5 for the Rayleigh distribution, by 13 for



the Chi-squared distribution. And risk increases when the variance of the noise increases.

We also illustrate the results with some figures. Figure 1 and 2 display the results of the data driven estimation respectively for the mixed Gamma and the Gamma  $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$  for the Laguerre and Fourier methods. We can observe some oscillations near the origin for the Laguerre procedure, while for the Fourier method we can see that the estimators are a little bit shifted from the true density. For both methods the sample size  $n$  needs to be large enough to estimate the two modes of the mixed Gamma.

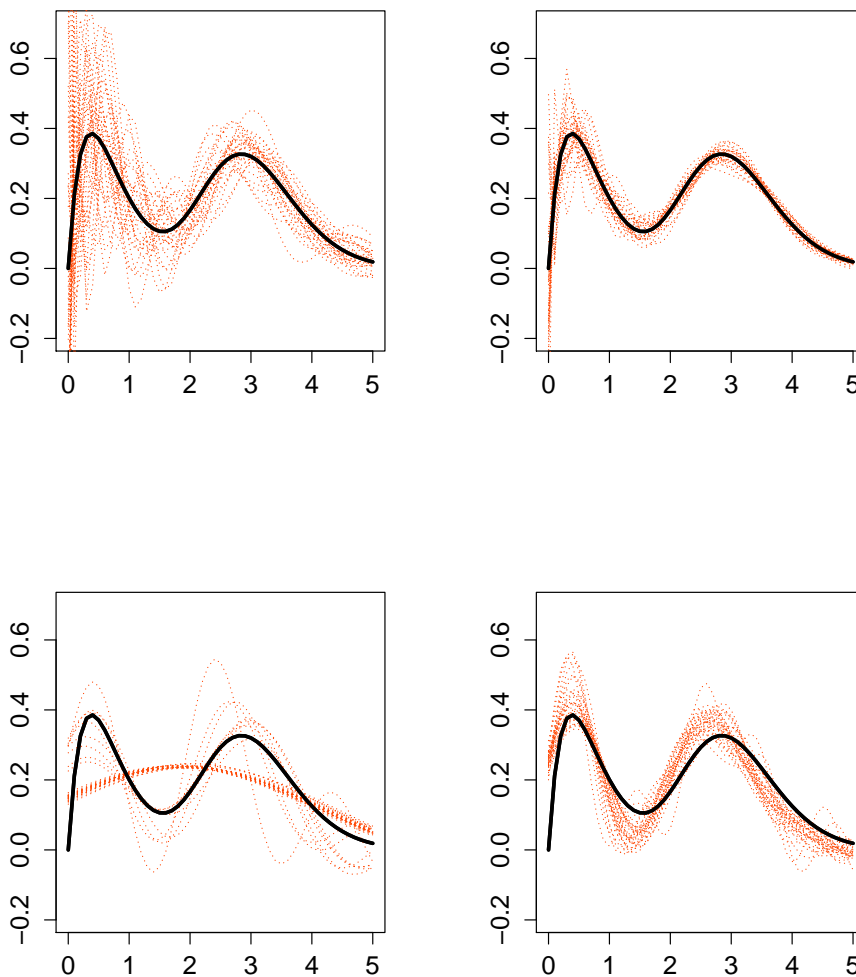


FIGURE 1. Estimation of the mixed Gamma density with Laguerre method (top left for  $n = 200$  and top right  $n = 2000$ ) and with Fourier method (bottom left for  $n = 200$  and bottom right  $n = 2000$ ), with  $\sigma^2 = 1/10$ .

## 6. CONCLUDING REMARKS

This paper deals with the estimation of densities and survival functions on  $\mathbb{R}^+$  in a deconvolution setting with a known error distribution and, as a particular case, to that of direct estimation. First we have considered a projection estimator and then the adaptive estimation of the density  $f$  of the  $X_i$ 's in a deconvolution setting and deduced a procedure when there is no additional noise. Secondly we have tackled the problem of the adaptive estimation of the survival function which is new to our knowledge, in a global estimation setting on  $\mathbb{R}^+$ . Moreover we have illustrated the

performances of our Laguerre procedure and compared it, when it is possible to the performances of the Fourier procedure described in [Comte et al. \(2006\)](#). The Laguerre procedure outperforms the previous one in the simulations. These results show that the Laguerre procedure is worthy of interest when the variables are nonnegative.

Assuming that error distribution to be known is often not realistic in applications. Nevertheless this would require additional information on the error distribution. In the deconvolution literature with unknown error distribution it is assumed that we have access to a preliminary sample of the noise, see for instance [Neumann \(1997\)](#). Thanks to this preliminary observation we could estimate the coefficients of the matrix  $\mathbf{G}_m$  since we could provide unbiased estimators of the coefficients of the matrix which are the coefficients of the density  $g$  on the Laguerre basis. [Vareschi \(2015\)](#), in a Laplace regression model, considers this problem ; he assumes that a perturbation of the coefficients of the matrix  $\mathbf{G}_m$  are observed instead of preliminary sample drawn from  $g$ . At last, in our model we would need to control the deviation of the spectral norm of  $\widehat{\mathbf{G}}_m^{-1}$  around  $\mathbf{G}_m^{-1}$ .

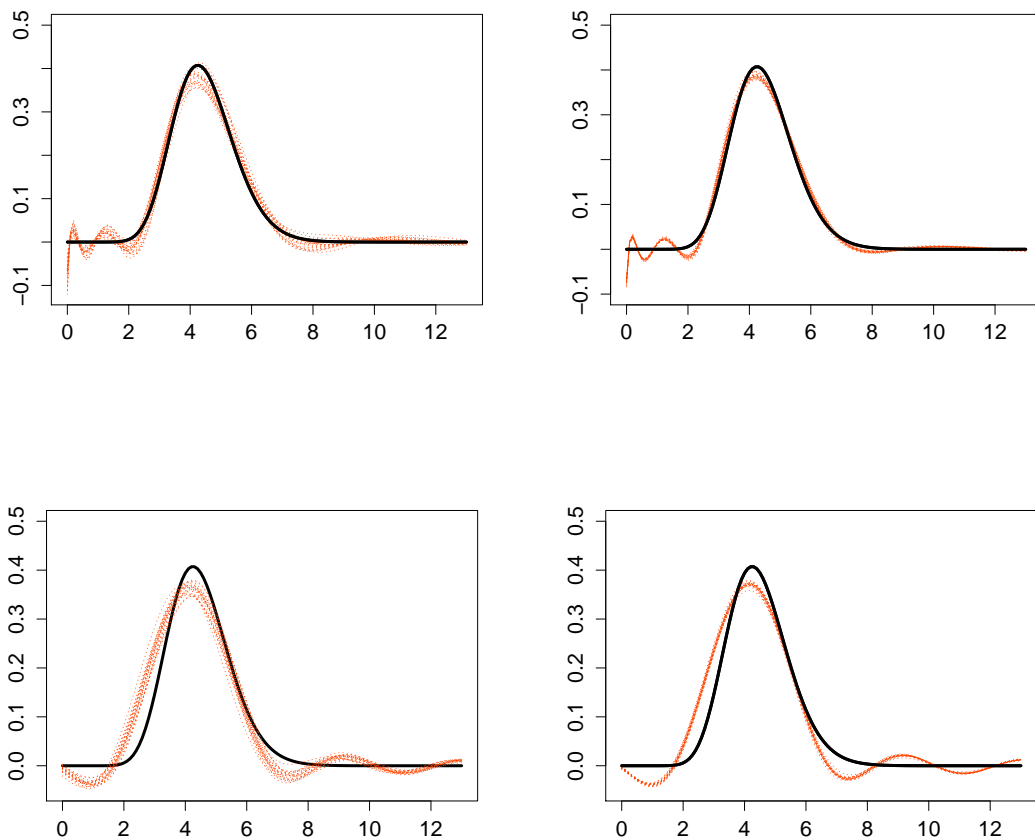


FIGURE 2. Estimation of the Gamma density  $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$  with Laguerre method (top left for  $n = 200$  and top right  $n = 2000$ ) and with Fourier method (bottom left for  $n = 200$  and bottom right  $n = 2000$ ), with  $\sigma^2 = 1/10$ .

## 7. PROOFS

### 7.1. Proof of results of Section 3.

*Proof of Proposition 3.1.* According to the Pythagorean theorem, we have

$$\|f - \hat{f}_m\|^2 = \|f - f_m + f_m - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2.$$

The first term corresponds to the bias term of Equation (13). Let us study the second term: using the decomposition on the orthonormal Laguerre basis, we have

$$\|f_m - \hat{f}_m\|^2 = \sum_{k=0}^{m-1} \left( b_k(f) - \hat{b}_k(f) \right)^2.$$

First we apply (6) and get

$$\begin{aligned} \mathbb{E}\|f_m - \hat{f}_m\|^2 &= \mathbb{E}\|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \leq \varrho^2(\mathbf{G}_m^{-1})\mathbb{E}\|\hat{\vec{h}}_m - \vec{h}_m\|_{2,m}^2 \\ &\leq \varrho^2(\mathbf{G}_m^{-1})\mathbb{E}\left[ \sum_{j=1}^m \left( \frac{1}{n} \sum_{i=1}^n \varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_1)] \right)^2 \right] \\ &\leq \frac{\varrho^2(\mathbf{G}_m^{-1})}{n} \sum_{j=1}^m \mathbb{E}[\varphi_j^2(Z_1)] \leq \varrho^2(\mathbf{G}_m^{-1}) \frac{2m}{n}. \end{aligned}$$

Secondly, we can notice that

$$\begin{aligned} \mathbb{E}\|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 &= \mathbb{E} \sum_{k=1}^m \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} (b_{j-1}(h) - \hat{b}_{j-1}(h)) \right)^2 \\ &= \mathbb{E} \sum_{k=1}^m \sum_{j'=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} (b_{j-1}(h) - \hat{b}_{j-1}(h)) [\mathbf{G}_m^{-1}]_{k,j'} (b_{j'-1}(h) - \hat{b}_{j'-1}(h)) \\ &= \sum_{k=1}^m \sum_{j'=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} [\mathbf{G}_m^{-1}]_{k,j'} \mathbb{E} \left[ (b_{j-1}(h) - \hat{b}_{j-1}(h))(b_{j'-1}(h) - \hat{b}_{j'-1}(h)) \right], \end{aligned}$$

and since  $b_{j-1}(h) - \hat{b}_{j-1}(h) = (1/n) \sum_{i=1}^n (\varphi_{j-1}(Z_i) - \mathbb{E}[\varphi_{j-1}(Z_i)])$ , it yields that

$$\begin{aligned} &\mathbb{E}\|\mathbf{G}_m^{-1}(\hat{\vec{h}}_m - \vec{h}_m)\|_{2,m}^2 \\ &= \frac{1}{n} \sum_{k=1}^m \sum_{j=1}^m \text{Var} \left[ [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1) \right] \leq \frac{1}{n} \mathbb{E} \left[ \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1) \right)^2 \right] \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \int_{\mathbb{R}^+} \left( \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(u) \right)^2 du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{1 \leq j, j' \leq m} [\mathbf{G}_m^{-1}]_{k,j} [\mathbf{G}_m^{-1}]_{k,j'} \int_{\mathbb{R}^+} \varphi_{j-1}(u) \varphi_{j'-1}(u) du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j}^2 = \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2. \end{aligned}$$

In the end we get:  $\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n} \varrho^2(\mathbf{G}_m^{-1}) \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$ .  $\square$

*Proof of Proposition 3.3.* As in the previous proof, we can write that

$$\|S_X - \hat{S}_{X,m}\|^2 = \|S_X - S_{X,m}\|^2 + \|S_{X,m} - \hat{S}_{X,m}\|^2.$$

We can notice that

$$\|S_{X,m} - \hat{S}_{X,m}\|^2 = \left\| \hat{S}_{X,m} - \vec{S}_{X,m} \right\|_{2,m}^2 = \left\| \mathbf{G}_m^{-1} \left( \hat{S}_{Z,m} - \vec{S}_{Z,m} \right) \right\|_{2,m}^2$$

Then we repeat the same scheme as in the proof of Proposition 3.1 and we get

$$\mathbb{E}\|S_{X,m} - \widehat{S}_{X,m}\|^2 \leq \frac{1}{n} \varrho^2(\mathbf{G}_m^{-1}) \sum_{j=0}^{m-1} \mathbb{E}[\Phi_j^2(Z_1)].$$

Yet

$$\sum_{j=0}^{m-1} \Phi_j^2(Z_1) = \sum_{j=0}^{m-1} \left( \int \varphi_j(u) \mathbf{1}_{0 \leq u \leq Z_1} du \right)^2 = \sum_{j=0}^{m-1} \langle \varphi_j, \mathbf{1}_{\cdot \leq Z_1} \rangle^2 \leq \|\mathbf{1}_{\cdot \leq Z_1}\|_{\mathbb{R}^+}^2 = Z_1 \quad (24)$$

which implies  $\mathbb{E} \left[ \sum_{j=0}^{m-1} \Phi_j^2(Z_1) \right] \leq \mathbb{E}[Z_1]$ .

In the end:  $\mathbb{E}\|S_X - \widehat{S}_{X,m}\|^2 \leq \|S_X - S_{X,m}\|^2 + \frac{\mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1})$ .  $\square$

*Proof of Lemma 3.4.* To see that the spectral norm grows with the dimension  $m$ , recall that for a matrix  $\mathbf{A}$  of dimension  $m$  the spectral norm can be written as  $\varrho^2(\mathbf{A}) = \max_{\|\vec{u}\|_{2,m}=1} \|\mathbf{A}\vec{u}\|_{2,m}^2$ . Now consider  $\vec{u}_m = \arg \max_{\|\vec{u}\|_{2,m}=1} \|\mathbf{T}_m \vec{u}\|_{2,m}$  with  $\mathbf{T}_m$  a lower triangular matrix and  $\mathbf{T}_m$  a submatrix of  $\mathbf{T}_{m+1}$ . We put  $\vec{v}_{m+1} = (\vec{u}_m \ 0) = (u_1, \dots, u_m, 0)$  thus  $\|\vec{v}_{m+1}\|_{2,m+1}^2 = 1$ . Then we get

$$\|\mathbf{T}_{m+1} \vec{v}_{m+1}\|_{2,m+1}^2 = \|\mathbf{T}_m \vec{u}_m\|_{2,m}^2 + \sum_{i=1}^m \left( [\mathbf{T}_{m+1}]_{m+1,i} u_i \right)^2 \geq \varrho^2(\mathbf{T}_m).$$

It yields

$$\varrho^2(\mathbf{T}_{m+1}) = \max_{\|\vec{v}\|_{2,m+1}=1} \|\mathbf{T}_{m+1} \vec{v}\|_{2,m+1}^2 \geq \varrho^2(\mathbf{T}_m). \quad \square$$

*Proof of Lemma 3.9.* According to Proposition 3.1, we have for squared integrable density  $f$  and gon  $\mathbb{R}^+$

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n} \varrho^2(\mathbf{G}_m^{-1}) \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2.$$

First we want to give an upper bound for the squared bias :

$$\|f - f_m\|^2 = \|\vec{f} - \vec{f}_m\|_m^2 = \sum_{k \geq m} b_k^2(f)$$

Let us compute the coefficients when  $f$  belongs to  $\mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})$  defined by (19)

$$\begin{aligned} b_k(f) &= \int_{\mathbb{R}^+} \theta^p u^{p-1} e^{-\theta u} \sqrt{2} e^{-u} L_k(2u) du = \sqrt{2} \theta^p \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}^+} u^{p+j-1} e^{-(1+\theta)u} du \\ &= \frac{\sqrt{2} \theta^p}{(1+\theta)^p} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{(1+\theta)^j} \frac{(p+j-1)!}{j!} = \frac{\sqrt{2} \theta^p}{(1+\theta)^p} \frac{d^{p-1}}{dx^{p-1}} \left[ x^{p-1} (1-x)^k \right] \Big|_{x=2/(1+\theta)}. \end{aligned} \quad (25)$$

It leads to

$$\|f - f_m\|^2 = \sum_{k \geq m} \left( \sum_{i=1}^p \lambda_i b_k(\gamma_i) \right)^2 \leq \sum_{k \geq m} \sum_{i=1}^p \lambda_i (b_k(\gamma_i))^2 \leq \sum_{i=1}^p \lambda_i C(\alpha_i, \theta_i) \left| \frac{\theta_i - 1}{\theta_i + 1} \right|^{2m} m^{2(\alpha_i - 1)}.$$

At least for the variance if  $g \sim \Gamma(q, \mu)$  then it satisfies Assumptions (C1)-(C2) with  $r = q$  which implies that  $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \asymp \varrho^2(\mathbf{G}_m^{-1}) \asymp m^{2q}$ . So we have

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \sum_{i=1}^p \lambda_i C(\alpha_i, \theta_i) \left| \frac{\theta_i - 1}{\theta_i + 1} \right|^{2m} m^{2(\alpha_i - 1)} + (2 \vee \|h\|_\infty) C_\varrho \frac{m^{2q}}{n},$$

To obtain b), let us notice that Beta distributions with parameter  $a$  and  $b$  with  $b > a$  also satisfy Assumptions (C1)-(C2) but with  $r = a$ .  $\square$

7.2. **Proof of Theorem 4.1.** First for  $m \in \mathcal{M}_n^{(1)}$ , let us define the associated subspaces  $\mathcal{S}_{d_1}^m \subseteq \mathbb{R}^{d_1}$

$$\mathcal{S}_{d_1}^m = \left\{ \vec{t}_m \in \mathbb{R}^{d_1} / \vec{t}_m = {}^t(b_0(t), b_1(t), \dots, b_{m-1}(t), 0, \dots, 0) \right\}.$$

This space is defined to give nested models. When we increase the dimension from  $m$  to  $m+1$  we only compute one more coefficient. Then for any  $\vec{t} \in \mathbb{R}^{d_1}$ , we define the following contrast for the density estimation

$$\gamma_n(\vec{t}) = \|\vec{t}\|_{2,d_1}^2 - 2\langle \vec{t}, \mathbf{G}_{d_1}^{-1} \hat{h}_{d_1} \rangle_{2,d_1}.$$

Let us notice that for  $\vec{t}_m \in \mathcal{S}_{d_1}^m$ , thanks to the null coordinates of  $\vec{t}_m$  and the lower triangular form of  $\mathbf{G}_{d_1}$  and  $\mathbf{G}_m$ , we have

$$\langle \vec{t}_m, \mathbf{G}_{d_1}^{-1} \hat{h}_{d_1} \rangle_{2,d_1} = \langle \vec{t}_m, \mathbf{G}_m^{-1} \hat{h}_m \rangle_{2,m} = \langle \vec{t}_m, \hat{f}_m \rangle_{2,m}.$$

So we clearly have that

$$\hat{f}_m = \arg \min_{\vec{t}_m \in \mathcal{S}_{d_1}^m} \gamma_n(\vec{t}_m).$$

Now let  $m, m' \in \mathcal{M}_n^{(1)}$ ,  $\vec{t}_m \in \mathcal{S}_{d_1}^m$  and  $\vec{s}_{m'} \in \mathcal{S}_{d_1}^{m'}$ . Denote  $m^* = m \vee m'$ . Notice that

$$\gamma_n(\vec{t}_m) - \gamma_n(\vec{s}_{m'}) = \|\vec{t}_m - \vec{f}\|_{2,d_1}^2 - \|\vec{s}_{m'} - \vec{f}\|_{2,d_1}^2 - 2\langle \vec{t}_m - \vec{s}_{m'}, \mathbf{G}_{d_1}^{-1}(\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$$

and due to orthonormality of Laguerre basis, for any  $m$  we have the following relations between the  $\mathbb{L}^2$  norm and the Euclidean norms,

$$\|\hat{f}_m - f\|^2 = \|\hat{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (b_j(f))^2 \quad \text{and} \quad \|f_m - f\|^2 = \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (b_j(f))^2 \quad (26)$$

We set  $\nu_n(\vec{t}) = \langle \vec{t}, \mathbf{G}_{d_1}^{-1}(\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$  for  $\vec{t} \in \mathbb{R}^{d_1}$ .

According to the definition of  $\hat{m}$ , for any  $m$  in the model collection  $\mathcal{M}_n^{(1)}$ , we have the following inequality

$$\gamma_n(\hat{f}_{\hat{m}}) + \text{pen}_1(\hat{m}) \leq \gamma_n(\vec{f}_m) + \text{pen}_1(m).$$

It yields that

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 - \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 - 2\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) \leq \text{pen}_1(m) - \text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) + \text{pen}_1(m) - \text{pen}_1(\hat{m}).$$

Let us notice that  $\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) = \|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \nu_n \left( \frac{\hat{f}_{\hat{m}} - \vec{f}_m}{\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}} \right)$  and due to the relation  $2ab \leq a^2/4 + 4b^2$ , we have the following inequalities

$$\begin{aligned} \|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m}) \\ &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{4}\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m}) \end{aligned}$$

where  $\mathcal{B}(m, \hat{m}) = \{\vec{t}_{m \vee \hat{m}} \in \mathcal{S}_{d_1}^{m \vee \hat{m}}, \|\vec{t}_{m \vee \hat{m}}\|_{2,d_1} = 1\}$ . Now notice that

$$\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 \leq 2\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + 2\|\vec{f}_m - \vec{f}\|_{2,d_1}^2$$

we then have

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq 3\|\vec{f} - \vec{f}_m\|_{2,d_1}^2 + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m}).$$

Using Equation (26), we have

$$\|\hat{f}_{\hat{m}} - f\|^2 - \sum_{j=d_1}^{\infty} (b_j(f))^2 \leq 3 \left( \|f - f_m\|^2 - \sum_{j=d_1}^{\infty} (b_j(f))^2 \right) + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f - f_m\|^2 + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m}) \quad (27)$$

Now let  $p_1$  be a function such that for any  $m, m'$ , we have :  $4p_1(m, m') \leq \text{pen}_1(m) + \text{pen}_1(m')$ .

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq 3\|f - f_m\|^2 + 4\text{pen}_1(m) + 8 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - p_1(m, \hat{m}) \right]_+ \\ &\leq 3\|f - f_m\|^2 + 4\text{pen}_1(m) + 8 \sum_{m' \in \mathcal{M}_n^{(1)}} \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} \nu_n^2(\vec{t}) - p_1(m, m') \right\}_+ \end{aligned}$$

We now use the following result which ensures the validity of Theorem 4.1.

**Proposition 7.1.** *Under the assumptions of Theorem 4.1, there exists a constant  $C_1 > 0$  depending on  $\|h\|_{\infty}$  such that for*

$$p_1(m, m') = \frac{\kappa_1}{n} \min(2(m \vee m') \varrho^2(\mathbf{G}_{m \vee m'}^{-1}), \log n (\|g\|_{\infty} \vee 1) \|\mathbf{G}_{m \vee m'}^{-1}\|_{\mathbb{F}}^2),$$

we have

$$\mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} \nu_n^2(\vec{t}) - p(m, m') \right\}_+ \right] \leq \frac{C_1}{n}.$$

In the end:

$$\mathbb{E}\|f - \hat{f}_{\hat{m}}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(1)}} \{\|f - f_m\|^2 + \text{pen}_1(m)\} + \frac{C_1}{n},$$

as soon as  $\kappa_1 \geq 294$ . □

*Proof of Proposition 7.1.* To prove Proposition 7.1, we apply a Talagrand inequality. So we need to determine  $H$ ,  $M_1$  and  $v$  defined as

$$\begin{aligned} \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(\cdot) \rangle_{2,d_1}\|_{\infty} &\leq M_1, \quad \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})| \right] \leq H, \\ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \text{Var} \left[ \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2,d_1} \right] &\leq v. \end{aligned}$$

where  $m^* = m \vee m'$ .

- Let us start with the empirical process, first let us notice that

$$\begin{aligned}
& \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})|^2 \right] \\
&= \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \left( \frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{d_1}(Z_i) - \mathbb{E}[\vec{\varphi}_{d_1}(Z_i)]) \right) \rangle_{2, d_1} \right|^2 \right] \\
&= \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \rangle_{2, m^*} \right|^2 \right]
\end{aligned}$$

We now apply Cauchy-Schwarz inequality and get

$$\begin{aligned}
& \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})|^2 \right] \\
&\leq \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\vec{t}_{m^*}\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \right\|_{2, m^*}^2 \right] \\
&\leq \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E} \left[ \sum_{j=0}^{m^*-1} \left( \frac{1}{n} \sum_{i=1}^n (\varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_i)]) \right)^2 \right] \\
&\leq \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \sum_{j=0}^{m^*-1} \text{Var}(\varphi_j(Z_1)) \leq \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \sum_{j=0}^{m^*-1} \mathbb{E}[\varphi_j^2(Z_1)] \leq \frac{2m^*}{n} \varrho^2 (\mathbf{G}_{m^*}^{-1}).
\end{aligned}$$

Moreover as in proof of Proposition 3.1, we can notice that

$$\begin{aligned}
\mathbb{E} \left[ \left\| \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \right\|_{2, m^*}^2 \right] &\leq \frac{(\|h\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\text{F}}^2}{n} \\
&\leq \frac{(\|g\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\text{F}}^2}{n}.
\end{aligned}$$

We then set  $H := \sqrt{\frac{2m^*}{n} \varrho^2 (\mathbf{G}_{m^*}^{-1})} \wedge \frac{(\|g\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\text{F}}^2}{n}$ .

- Now for the term of variance, let  $\vec{t}_{m^*} \in \mathcal{B}(m, m')$ . By definition we have the following equalities

$$\begin{aligned}
& \mathbb{E} \left[ \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2, d_1} \right|^2 \right] = \mathbb{E} \left[ \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(Z_1) \rangle_{2, m^*} \right|^2 \right] \\
&= \mathbb{E} \left[ \left| \sum_{j=0}^{m^*-1} b_j(t) \sum_{k=0}^{m^*-1} [\mathbf{G}_{m^*}^{-1}]_{jk} \varphi_k(Z_1) \right|^2 \right] = \int_{\mathbb{R}^+} \left| \sum_{0 \leq k, j \leq m^*-1} b_j(t) [\mathbf{G}_{m^*}^{-1}]_{jk} \varphi_k(u) \right|^2 h(u) du.
\end{aligned}$$



which implies

$$\begin{aligned}
& \mathbb{E} \left[ \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2, d_1} \right|^2 \right] \\
& \leq \|h\|_\infty \int_{\mathbb{R}^+} \sum_{0 \leq j, j', k, k' \leq m^* - 1} b_j(t) b_{j'}(t) [\mathbf{G}_{m^*}^{-1}]_{jk} [\mathbf{G}_{m^*}^{-1}]_{j'k'} \varphi_k(u) \varphi_{k'}(u) du \\
& \leq \|g\|_\infty \sum_{0 \leq j, j', k, k' \leq m^* - 1} b_j(t) b_{j'}(t) [\mathbf{G}_{m^*}^{-1}]_{jk} [\mathbf{G}_{m^*}^{-1}]_{j'k'} \delta_{k, k'} \\
& \leq \|g\|_\infty \sum_{0 \leq j, j', k \leq m^* - 1} b_j(t) b_{j'}(t) [\mathbf{G}_{m^*}^{-1}]_{jk} [\mathbf{G}_{m^*}^{-1}]_{j'k} \\
& \leq \|g\|_\infty \vec{t}_{m^*}^* \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{t}_{m^*} \leq \|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1}) \|\vec{t}_{m^*}\|_{2, m^*}^2 \leq \|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1}).
\end{aligned}$$

So we set  $v := \|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})$ .

• Now applying Cauchy-Schwarz inequality

$$\begin{aligned}
& \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(x) \rangle_{2, d_1} \right| = \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x) \rangle_{2, m^*} \right| \\
& \leq \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \|\vec{t}_{m^*}\|_{2, m^*} \|\mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x)\|_{2, m^*} \leq \sup_{x \in \mathbb{R}^+} \|\mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x)\|_{2, m^*} \\
& \leq \sup_{x \in \mathbb{R}^+} \sqrt{\vec{t}_{m^*}^*(x) \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x)} \leq \sqrt{\varrho^2(\mathbf{G}_{m^*}^{-1}) \sup_{x \in \mathbb{R}^+} \sum_{j=0}^{m^*-1} \varphi_j^2(x)} \leq \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}.
\end{aligned}$$

We take  $M_1 = \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}$ .

• We can now apply Talagrand's inequality (see Appendix 8.1).

i) First case :  $2m^* \varrho^2(\mathbf{G}_{m^*}^{-1}) \wedge \|g\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2 = 2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})$ .

For  $\xi^2 = 1/2$  with  $K_1 = 1/6$ , Talagrand's inequality implies

$$\begin{aligned}
& \sum_{m' \in \mathcal{M}_n^{(1)}} \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\nu_n(\vec{t})|^2 - 4 \frac{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \right\}_+ \right] \\
& \leq C \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \frac{m^*}{\|g\|_\infty}} + \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{n^2} e^{-c_1 \sqrt{n}}.
\end{aligned}$$

Yet under Assumption **(A2)**, we have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \frac{m^*}{\|g\|_\infty}} \leq \frac{C}{n}.$$

Moreover according to Assumptions **(A1)** and **(A2)**, we also have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{n^2} e^{-c_1 \sqrt{n}} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{e^{-c_1 \sqrt{n}}}{n} \leq \frac{C}{n}.$$

ii) Second case :  $2m^* \varrho^2(\mathbf{G}_{m^*}^{-1}) \wedge \|g\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2 = \|g\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2$ .

For  $\xi^2 = \mathbf{a}/K_1 \log n$  with  $\mathbf{a} = 7^2/K_1$ , Talagrand's inequality implies

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\nu_n(\vec{t})|^2 - 2(1 + \frac{\mathbf{a}}{K_1} \log n) \frac{(\|g\|_\infty \vee 1) \mathbf{G}_{m^*}^{-1} \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{n} \right\}_+ \right] \\ & \leq C \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \exp \left( -\mathbf{a} \log(n) \frac{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{\varrho^2(\mathbf{G}_{m^*}^{-1})} \right) \\ & \quad + \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{C^2(\xi^2) n^2} \exp \left( -\frac{2K_1}{7\sqrt{2}} C(\xi^2) \sqrt{\frac{\mathbf{a}}{K_1}} \sqrt{\log(n)} n \frac{(\|g\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}} \right). \end{aligned} \quad (28)$$

The first summand of Equation (28) implies that

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \exp \left( -\mathbf{a} \log(n) \frac{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{\varrho^2(\mathbf{G}_{m^*}^{-1})} \right) \\ & \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \exp(-\mathbf{a} \log(n)) \quad (\text{Equivalence between norms}) \\ & \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|g\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^{\mathbf{a}}} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \|g\|_\infty \frac{1}{n^{\mathbf{a}}} \quad (\text{Assumption (A1)}) \\ & = |\mathcal{M}_n^{(1)}| \frac{\|g\|_\infty}{n^{\mathbf{a}}} = \frac{\|g\|_\infty}{n^{\mathbf{a}-1}} \leq \frac{\|g\|_\infty}{n} \end{aligned}$$

since  $\mathbf{a} > 2$ .

Now let us bound the second summand of Equation (28). Assumption (A1) gives that  $m^* \varrho^2(\mathbf{G}_{m^*}^{-1}) \leq n/\log n$ , yet

$$\varrho(\mathbf{G}_{m^*}^{-1}) \geq \sqrt{2}/a_0(g) = 1/\mathbb{E}[e^{-Y}] \geq 1,$$

where we use that the operator norm is greater than the spectral radius. Then it implies that  $m^* \leq m^* \varrho^2(\mathbf{G}_{m^*}^{-1})$ . This remark combined with the equivalence between norms (see Eq. (18)) yields

$$-n \frac{(\|g\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}} \leq -\frac{n}{\sqrt{2n}} \frac{(\|g\|_\infty \vee 1)}{\sqrt{m^*}} \leq -\frac{1}{\sqrt{2}} \frac{n \sqrt{\log n}}{\sqrt{n} \sqrt{n}} \leq -\frac{\sqrt{\log n}}{\sqrt{2}}.$$

Secondly we have  $-C(\xi^2) = -\left(\sqrt{1 + \frac{\mathbf{a}}{K_1} \log n} - 1 \wedge 1\right) \leq -1$ , which implies that  $C(\xi^2) = 1$  since  $C(\xi^2) \leq 1$  by definition. This condition is verified as soon as  $n \geq e^{K_1/\mathbf{a}}$ , in other words for  $n \geq 2$ . Then we get that

$$\begin{aligned} & \exp \left( -\frac{2K_1}{7\sqrt{2}} C(\xi^2) \sqrt{\frac{\mathbf{a}}{K_1}} \sqrt{\log(n)} n \frac{(\|g\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}} \right) \leq \exp \left( -\frac{2K_1}{7\sqrt{2}} \sqrt{\frac{\mathbf{a}}{K_1}} \log n \frac{1}{\sqrt{2}} \right) \\ & \leq \exp \left( -\frac{\sqrt{\mathbf{a}K_1}}{7} \log n \right) \leq \frac{1}{n} \end{aligned}$$

since  $\mathbf{a} = 7^2/K_1$ . Thus, invoking again Assumption **(A1)**, the second summand of Equation (28) can be bounded as follows

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{C(\xi^2)n^2} \exp\left(-\frac{2K_1}{7\sqrt{2}}C(\xi^2)\sqrt{\log(n)}n \frac{(\|g\|_\infty \vee 1)\sqrt{\|\mathbf{G}_{m^*}^{-1}\|_F^2}}{\sqrt{n}\sqrt{2m^*\varrho^2(\mathbf{G}_{m^*}^{-1})}}\right) \\ & \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{C(\xi^2)n^2} \frac{1}{n} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{1}{C(\xi^2)n^2} = \frac{|\mathcal{M}_n^{(1)}|}{C(\xi^2)n^2} = \frac{1}{C(\xi^2)n} \leq \frac{1}{n}. \end{aligned}$$

In the end we have the desired result.  $\square$

*Proof of Remark 7.* We have to prove that Proposition 7.1 is still valid although Assumption **(A2)** is no longer true. We set  $\xi^2 = 2\log n/K_1$ ,  $H^2 = 2m^*\varrho^2(\mathbf{G}_{m^*}^{-1})/n$ ,  $v = 2m^*\varrho^2(\mathbf{G}_{m^*}^{-1})$ ,  $M_1 = \sqrt{2m^*\varrho^2(\mathbf{G}_{m^*}^{-1})}$ . Under **(A2)**, we have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{v}{n} \exp\left(-K_1\xi^2 n \frac{H^2}{v}\right) = \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{2m^*\varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^2} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{2}{n^2} \leq \frac{2|\mathcal{M}_n^{(1)}|}{n^2} \leq \frac{2}{n}.$$

For  $C(\xi^2)$  defined in Lemma 8.1, we get

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{M_1^2}{K_1 C^2(\xi^2) n^2} \exp\left(-\frac{2K_1}{7\sqrt{2}} \frac{C(\xi^2)\xi n H}{M_1}\right) \\ & = \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{K_1 C^2(\xi^2) n^2} \exp\left(-\frac{2K_1}{7\sqrt{2}} C(\xi^2)\xi\sqrt{n}\right) \\ & \leq C_2 \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{n^2} \exp\left(-C_1\sqrt{n\log n}\right) \\ & \leq \frac{C_3|\mathcal{M}_n^{(1)}|}{n} \exp\left(-C_1\sqrt{n\log n}\right) \leq \frac{C_3}{n}. \end{aligned}$$

$\square$

**7.3. Proof of Theorem 4.2 and Corollary 4.3.** For any  $\vec{t} \in \mathbb{R}^{d_2}$ , we define the following contrast for the survival function estimation

$$\delta_n(\vec{t}) = \|\vec{t}\|_{2,d_2}^2 - 2\langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left( \widehat{\vec{S}}_{Z,d_2} - \vec{S}_{Y,d_2} \right) \rangle_{2,d_2}$$

and we also have

$$\langle \vec{t}_m, \mathbf{G}_{d_2}^{-1} \left( \widehat{\vec{S}}_{Z,d_2} - \vec{S}_{Y,d_2} \right) \rangle_{2,d_2} = \langle \vec{t}_m, \mathbf{G}_m^{-1} \left( \widehat{\vec{S}}_{Z,m} - \vec{S}_{Y,m} \right) \rangle_{2,m} = \langle \vec{t}_m, \widehat{\vec{S}}_{X,m} \rangle_{2,m}$$

which yields that

$$\widehat{\vec{S}}_{X,m} = \arg \min_{\vec{t}_m \in \mathcal{S}_{d_2}^m} \delta_n(\vec{t}_m).$$

7.3.1. *Proof of Theorem 4.2.* The beginning of the proof is the same as the proof of Theorem 4.1 with the quantities associated to the survival function estimation. Then we start from Equation (27) with  $\nu_n(\vec{t})$  replacing the following empirical process  $\zeta_n(\vec{t}) := \zeta_n^{(1)}(\vec{t}) + \zeta_n^{(2)}(\vec{t})$  where,

$$\begin{aligned}\zeta_n^{(1)}(\vec{t}) &:= \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \rangle_{2, d_2} \\ \zeta_n^{(2)}(\vec{t}) &:= \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} \right] \right) \right) \rangle_{2, d_2}.\end{aligned}$$

So we have the following inequality

$$\begin{aligned}\|\widehat{S}_{X, \widehat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X, m}\|^2 + 2\text{pen}_2(m) \\ &\quad + 16 \left( \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) - 2\text{pen}_2(\widehat{m}).\end{aligned}\quad (29)$$

Now let  $q$  be a function such that for any  $m, m'$ , we have :  $4q(m, m') \leq \text{pen}_2(m) + \text{pen}_2(m')$ .

$$\begin{aligned}\|\widehat{S}_{X, \widehat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X, m}\|^2 + 4\text{pen}_2(m) + 16 \left[ \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, \widehat{m}) \right]_+ \\ &\quad + 16 \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \\ &\leq 3\|S_X - S_{X, m}\|^2 + 4\text{pen}_2(m) + 16 \sum_{m' \in \mathcal{M}_n^{(2)}} \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, m') \right\}_+ \\ &\quad + 16 \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2.\end{aligned}$$

We now use the following result which ensures the validity of Theorem 4.2.

**Proposition 7.2.** *Under the assumptions of Theorem 4.2, then there exists a universal constant  $C > 0$  such that for  $q(m, m') = \kappa_2 \varrho^2 (\mathbf{G}_{m \vee m'}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n}$*

$$\begin{aligned}(i) \quad &\mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, m') \right\}_+ \right] \leq \frac{C}{n} \\ (ii) \quad &\mathbb{E} \left[ \sup_{\vec{t} \in \mathcal{B}(m, \widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right] \leq \frac{\mathbb{E}[Z_1^3]}{n}.\end{aligned}$$

Finally,  $\mathbb{E}\|\widehat{S}_{X, \widehat{m}} - S_X\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \{ \|S_X - S_{X, m}\|^2 + \text{pen}_2(m) \} + \frac{C}{n}$ . □

*Proof of Proposition 7.2.* To prove (i), we apply a Talagrand inequality. So we need to determine  $H$ ,  $M_1$  and  $v$ .

- Let us start with the empirical process, first let us notice that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ &= \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \rangle_{2, d_2} \right|^2 \right] \\ &= \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \rangle_{2, m^*} \right|^2 \right]. \end{aligned}$$

We now apply Cauchy-Schwarz inequality and get

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ & \leq \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\vec{t}_{m^*}\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \right\|_{2, m^*}^2 \right] \\ & \leq \mathbb{E} \left[ \left\| \mathbf{G}_{m^*}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \right\|_{2, m^*}^2 \right]. \end{aligned}$$

It follows that, with Equation (24)

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ & \leq \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E} \left[ \sum_{j=0}^{m^*-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \Phi_j(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[ \Phi_j(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right)^2 \right] \\ & \leq \varrho^2 (\mathbf{G}_{m^*}^{-1}) \sum_{j=0}^{m^*-1} \text{Var} \left( \Phi_j(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right) \\ & \leq \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \sum_{j=0}^{m^*-1} \mathbb{E} \left[ \Phi_j^2(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right] \leq \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \mathbb{E} \left[ Z_1 \mathbf{1}_{Z_1 \leq \sqrt{n}} \right]. \end{aligned}$$

We then set  $H := \sqrt{\frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \mathbb{E} [Z_1]}$ .

- Now for the term of variance, let  $\vec{t}_{m^*} \in \mathcal{B}(m, m')$

$$\begin{aligned} & \mathbb{E} \left[ \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \rangle_{2, d_2} \right|^2 \right] = \mathbb{E} \left[ \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \rangle_{2, m^*} \right|^2 \right] \\ & \leq \mathbb{E} \left[ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\vec{t}_{m^*}\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right\|_{2, m^*}^2 \right] \\ & \leq \mathbb{E} \left[ \left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right\|_{2, m^*}^2 \right] \leq \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E} \left[ Z_1 \mathbf{1}_{Z_1 \leq \sqrt{n}} \right]. \end{aligned}$$

So we set  $v := \mathbb{E} [Z_1] \varrho^2 (\mathbf{G}_{m^*}^{-1})$ .

- First notice again

$$\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, d_2} = \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, m^*}$$

Now applying Cauchy-Schwarz inequality and using Equation (24) again

$$\begin{aligned} & \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, d_2}\|_\infty \\ & \leq \sup_{x \in \mathbb{R}^+} \left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq n} \right\|_{2, m^*} \leq \sup_{x \in \mathbb{R}^+} \sqrt{t \vec{\Phi}_{m^*}(x) t \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq \sqrt{n}}} \\ & \leq \sqrt{\varrho^2 (\mathbf{G}_{m^*}^{-1}) \sup_{x \in \mathbb{R}^+} \sum_{j=0}^{m^*-1} \Phi_j^2(x) \mathbf{1}_{x \leq \sqrt{n}}} \leq \sqrt{\sqrt{n} \varrho^2 (\mathbf{G}_{m^*}^{-1})}. \end{aligned}$$

We take  $M_1 = \sqrt{\sqrt{n} \varrho^2 (\mathbf{G}_{m^*}^{-1})}$ .

We apply Talagrand's inequality for  $\xi^2 = \frac{2}{K_1} \log(n)$ . We get

$$\begin{aligned} & \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \\ & \leq \frac{4}{K_1} \frac{\mathbb{E}[Z_1] \varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} + \frac{98n \varrho^2 (\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \end{aligned}$$

which implies that

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(2)}} \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \\ & \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{4}{K_1} \frac{\mathbb{E}[Z_1] \varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} + \frac{98\sqrt{n} \varrho^2 (\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}}. \end{aligned}$$

Thus we have under Assumption **(B1)**

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^2} \leq \frac{|\mathcal{M}_n^{(2)}|}{n^2} \leq \frac{1}{n}.$$

And

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\sqrt{n} \varrho^2 (\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{C\sqrt{n}}{(\log n)^2 n} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \leq \frac{C}{n}.$$

So for  $q(m, m') = \kappa_2 \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n}$  we just showed that

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \mathbb{E} \left[ \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2 (\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \leq \frac{C}{n}.$$

Now we prove (ii). We have, using **(A2)**,

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} |\zeta_n^{(2)}(\vec{t}_{m^*})|^2 \right] \\ & = \mathbb{E} \left[ \sup_{\vec{t} \in \mathcal{B}(d_2, d_2)} \left| \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \left( \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} - \mathbb{E} \left[ \vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} \right] \right) \right) \rangle_{2, d_2} \right|^2 \right] \\ & \leq \frac{\varrho^2 (\mathbf{G}_{d_2}^{-1})}{n} \mathbb{E} \left[ Z_1 \mathbf{1}_{Z_1 > \sqrt{n}} \right] \leq \frac{\mathbb{E}[Z_1^3]}{n}. \end{aligned}$$

In the end we have the desired result.  $\square$

7.3.2. *Proof of Corollary 4.3.* The beginning of the proof is the same as the proof of Theorem 4.2 except that we consider  $\tilde{m}_2$  (defined by Equation (22)) instead of  $\hat{m}_2$  and  $\widehat{\text{pen}}_2$  (defined by Equation (23)) instead of  $\text{pen}_2$ . Starting from Equation (29) we have

$$\begin{aligned} \|\widehat{S}_{X,\hat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left( \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) \\ &\quad - 2\widehat{\text{pen}}_2(\tilde{m}_2) \\ &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left( \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) \\ &\quad - 2\text{pen}_2(\tilde{m}_2) + 2\text{pen}_2(\tilde{m}_2) - 2\widehat{\text{pen}}_2(\tilde{m}_2) \\ &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left\{ \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, \tilde{m}_2) \right\}_+ \\ &\quad + 2\text{pen}_2(m) + 16 \sup_{\vec{t} \in \mathcal{B}(m,\hat{m})} (\zeta_n^{(2)}(\vec{t}))^2 + 2 \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+. \end{aligned}$$

We now apply the following Proposition which ensures the validity of Corollary 4.3

**Proposition 7.3.** *Under the Assumptions of Corollary 4.3, the following holds*

$$\mathbb{E}[\widehat{\text{pen}}_2(m)] = 2\text{pen}_2(m) \quad \text{and} \quad \mathbb{E} \left[ \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] \leq \frac{C}{n}.$$

Finally,  $\mathbb{E}\|S_X - \widehat{S}_{X,\tilde{m}_2}\|^2 \leq \inf_{m \in \mathcal{M}_n^{(2)}} \{\|S_X - S_{X,m}\|^2 + \text{pen}_2(m)\} + \frac{C}{n}$ . □

*Proof of Proposition 7.3.* First let us notice the following

$$\mathbb{E}[\widehat{\text{pen}}_2(m)] = 2\kappa_2 \mathbb{E}[\bar{Z}_n] \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} = 2\kappa_2 \mathbb{E}[Z_1] \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} = 2\text{pen}_2(m).$$

For the second inequality, let us introduce the following favorable set:

$$\Lambda = \{|\mathbb{E}[Z_1] - \bar{Z}_n| \leq \mathbb{E}[Z_1]/2\},$$

which yields

$$\begin{aligned} \mathbb{E} \left[ \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] &= \mathbb{E} \left[ \left\{ 2\kappa_2 \left( \frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \frac{\log n}{n} \right\}_+ \mathbf{1}_\Lambda \right] \\ &\quad + \mathbb{E} \left[ \left\{ 2\kappa_2 \left( \frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \frac{\log n}{n} \right\}_+ \mathbf{1}_{\Lambda^c} \right]. \end{aligned}$$

Yet on the set  $\Lambda$ ,  $\mathbb{E}[Z_1]/2 - \bar{Z}_n \leq 0$  which yields

$$\begin{aligned} \mathbb{E} \left[ \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] &= \mathbb{E} \left[ 2\kappa_2 \left( \frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \mathbf{1}_{\Lambda^c} \right] \frac{\log n}{n} \\ &\leq \mathbb{E} \left[ 2\kappa_2 |\mathbb{E}[Z_1] - \bar{Z}_n| \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \mathbf{1}_{\Lambda^c} \right] \frac{\log n}{n}. \end{aligned}$$

Now we apply Cauchy-Schwarz

$$\mathbb{E} \left[ |\mathbb{E}[Z_1] - \bar{Z}_n| \mathbf{1}_{\Lambda^c} \right] \leq \sqrt{\mathbb{E} |\mathbb{E}[Z_1] - \bar{Z}_n|^2} \sqrt{\mathbb{P}[\Lambda^c]} = \sqrt{\text{Var}[\bar{Z}_n]} \sqrt{\mathbb{P} \left[ |\mathbb{E}[Z_1] - \bar{Z}_n| \geq \frac{\mathbb{E}[Z_1]}{2} \right]}.$$

We apply Markov inequality then Rosenthal inequality

$$\mathbb{E} \left[ |\mathbb{E}[Z_1] - \bar{Z}_n| \mathbf{1}_{\Lambda^c} \right] \leq \frac{\sqrt{\text{Var}[\bar{Z}_n]}}{\sqrt{n}} \sqrt{\frac{\mathbb{E} \left[ |\mathbb{E}[Z_1] - \bar{Z}_n|^2 \right]}{\mathbb{E}[Z_1]^2}} \leq \frac{\sqrt{\text{Var}[\bar{Z}_n]}}{\mathbb{E}[Z_1]n}.$$



Moreover under Assumption **(B1)**

$$\mathbb{E} \left[ \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] \leq \frac{C}{n} \mathbb{E} \left[ \varrho^2 \left( \mathbf{G}_{\tilde{m}_2}^{-1} \right) \frac{\log n}{n} \right] \leq \frac{C'}{n}$$

which ends the proof.  $\square$

## 8. APPENDIX

**Lemma 8.1.** (*Talagrand's inequality*) Let  $Y_1, \dots, Y_n$  be i.i.d. variables and

$$r_n(f) = \frac{1}{n} \sum_{k=1}^n \left( f(Y_k) - \mathbb{E}[f(Y_k)] \right)$$

for  $f$  belonging to some countable set  $\mathcal{F}$  of uniformly bounded measurable functions. Then for  $\xi^2 > 0$ ,

$$\mathbb{E} \left[ \left\{ \sup_{f \in \mathcal{F}} |r_n(f)|^2 - 2(1 + 2\xi^2)H^2 \right\}_+ \right] \leq \frac{4}{K_1} \left( \frac{v}{n} e^{-K_1 \xi^2 \frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \frac{nH}{M_1}} \right)$$

with constants  $C(\xi^2) = (\sqrt{1 + \xi^2} - 1) \wedge 1$  and  $K_1 = \frac{1}{6}$ ,  $M_1$ ,  $H$  and  $v$  are such that

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \quad \mathbb{E} \left[ \sup_{f \in \mathcal{F}} |r_n(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \text{Var}(f(Y_1)) \leq v.$$

## ACKNOWLEDGEMENTS

The author would like to thank Fabienne Comte for her fruitful advice and suggestions and Valentine Genon-Catalot for her comments which improved this work.

## REFERENCES

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.
- Belomestny, D., Comte, F., and Genon-Catalot, V. (2016). Nonparametric Laguerre estimation in the multiplicative censoring model. *To appear in Electronic Journal of Statistics*. preprint hal-01252143.
- Birgé, L. (1999). An alternative point of view on Lepski's method. In *State of the art in probability and statistics: Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series.
- Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer.
- Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2):147–172.
- Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Can. J. Stat.*, 32(2):181–192.
- Butucea, C. and Tsybakov, A. (2008a). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1):24–39.
- Butucea, C. and Tsybakov, A. (2008b). Sharp optimality in density deconvolution with dominating bias II. *Theory Proba. Appl.*, 52(2):237–249.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Am. Stat. Assoc.*, 83(404):1184–1186.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575.

- Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2015). Laplace deconvolution and its application to dynamic contrast enhanced imaging. *J. R. Stat. Soc., Ser. B.*
- Comte, F. and Genon-Catalot, V. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.*, 9:1112–1148.
- Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.*, 3(34):431–452.
- Dattner, I., Goldenshluger, A., and Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.*, 39(5):2477–2501.
- Dattner, I. and Reiser, B. (2013). Estimation of distribution functions in measurement error models. *J. Stat. Plann. Inference*, 143(3):479 – 493.
- Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, 22(1):143–192.
- Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Stat. Math.*, 56(1):19–47.
- Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *J. Am. Stat. Assoc.*, 92:526–535.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.
- Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, 21(2):600–610.
- Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Stat. Neerl.*, 57(1):136–157.
- Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.
- Hall, P. and Lahiri, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.*, 36(5):2110–2134.
- Jirak, M., Meister, A., and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.*, 42(5):1970–2002.
- Jongbloed, G. (1998). Exponential deconvolution: two asymptotically equivalent estimators. *Stat. Neerl.*, 52(1):6–17.
- Massart, P. (2003). *Concentration Inequalities and Model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.
- Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Stat.*, 7(4):307–330.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6):2033–2053.
- Rebafka, T. and Roueff, F. (2015). Nonparametric estimation of the mixing density using polynomials. *Math. Methods Stat.*, 24(3):200–224.
- Reiß, M. and Selk, L. (2013). Estimating nonparametric functionals efficiently under one-sided errors. Preprint, arXiv:1407.4229.
- Roueff, F. and Rydén, T. (2005). Nonparametric estimation of mixing densities for discrete distributions. *Ann. Statist.*, 33(5):2066–2108.
- Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Stat. Probab. Lett.*, 9:229–235.
- Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184.
- van Es, B. (2011). Combining kernel estimators in the uniform deconvolution problem. *Stat. Neerl.*, 65(3):275–296.
- van Es, B., Jongbloed, G., and van Zuijlen, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.*, 26(6):2395–2406.

- Vareschi, T. (2015). Noisy Laplace deconvolution with error in the operator. *J. Stat. Plann. Inference*, 157–158:16–35.
- Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18(2):806–831.