Adaptive deconvolution on the nonnegative real line

Gwennaëlle Mabon

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 domains 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 works significantly better than Fourier methods for estimating $\mathbb{R}^+$-supported 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

In this paper we consider the following model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \tag{1}$$

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 the X_i's when the Z_i's are observed. We are going to show that the assumption of nonnegativity of the random variables is of huge importance for the estimation strategy.

The assumptions imply that, in Model (1), $h(x) = (f \star g)(x)$ where $(\varphi \star \psi)(x) = \int \varphi(x-u)\psi(u)\,\mathrm{d}u$ denotes the convolution product. This setting matches the setting of convolution models which is classical in nonparametric statistics. Indeed the problem of recovering the signal distribution f when it is observed with an additive noise with known error distribution, has been extensively

E-mail address: gwennaelle.mabon@ensae.fr.
Date: April 7, 2015.

studied. 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 bandwith selection, Comte et al. (2006) for projection strategies with penalization, or Meister (2009) and references therein. More recently deconvolution problems in additive models in the case of unknown error distribution have been addressed. For that some information on the error distribution is required. For instance, in a physical context, a preliminary sample of the noise can be derived. This led to the works of Neumann (1997) who proposed an estimation strategy still based on Fourier inversion, or Johannes (2009), Comte and Lacour (2011) and Kappus and Mabon (2014) who extended it to the adaptive strategy. Concerning the estimation of the cumulative distribution function (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. (2014). 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 the survival function 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 $\mathbb{R}$-supported. Therefore they are still valid when the variables are $\mathbb{R}^+$-supported. Nonetheless in the present paper, our goal is to propose a specific solution for nonnegative variables. We can already note that with this model we can allow Y to be an uniform distribution which is not the case in the Fourier approach. Moreover we shall illustrate that in practice our procedure works significantly better than Fourier methods for estimating $\mathbb{R}^+$-supported functions.

Model (1) is also related to the field of mixture models. We can cite the works of Roueff and Rydén (2005) and Rebafka and Roueff (2010) who study in particular mixtures of Exponential and Gamma. These models are justified by some applications in natural sciences such as radioactive decays for instance. In Rebafka and Roueff (2010) the authors use Legendre polynomials to derive their estimators but with this strategy they have to consider compact interval for the estimation. We do not have this constraint in our method thanks to the Laguerre basis.

More broadly 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. It is motivated from fields where the information provided about the error distribution is its support rather than its mean properties. Such matters arise in economics: for example in auction fields the underlying distribution of bidders' private values is identified from observed bids, see Guerre et al. (2000). This led Jirak et al. (2014) to a different approach based on local extreme values. We can also cite Bibinger et al. (2014) who are motivated by modelizing the dynamics of intra-day financial data from limit order books. In all these works one-sided errors models require new and different tools. More generally, the field of survival data analysis and reliability is widespread in many domains such as econometrics or biology, and they also involve nonnegative variables. Duration models are used as soon as the phenomenon of interest is modelized by nonnegative variables which generally corresponds to a waiting time until the occurrence of a certain event such as a failure time. Model (1) can thus be seen as a superposition of two such processes.

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

(A1) $f \in \mathbb{L}^2(\mathbb{R}^+)$

when the estimation of f is under study, or

(B1) $S_X \in \mathbb{L}^2(\mathbb{R}^+)$,

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 decompose the density function f on an appropriate orthonormal basis on $\mathbb{R}^+$, $(\varphi_k)_{k\geq 0}$,

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

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

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

This basis has already been used to estimate a nonnegative function f in Comte et al. (2013). 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, ε_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

$$h(x) = \int_0^x f(u)g(x-u)\,\mathrm{d}u \tag{3}$$

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

$$\int_0^x \varphi_k(u)\varphi_j(x-u)\,\mathrm{d}u = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right). \tag{4}$$

(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 $b_k(h)$, $k = 0, 1, \ldots, m-1$ respectively, we prove

$$\vec{h}_m = \mathbf{G}_m\vec{f}_m \tag{5}$$

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 as $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$. We then derive the same reasoning for the survival function estimation. Let us point out that we do not integrate the estimators 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 decomposition 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. These estimators are precisely defined and illustrated in Section 2.

We develop in Section 3 a study of the integrated risk 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 lead a 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 $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the $\mathbb{L}^2$ norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ 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 $\mathbf{A}$: $\varrho^2(\mathbf{A}) = \lambda_{\max}\left({}^t\mathbf{A}\mathbf{A}\right)$ where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of A in absolute value.

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, $\int_{\mathbb{R}^+} L_k(x)L_{k'}(x)e^{-x}\,\mathrm{d}x = \delta_{k,k'}$ 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

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

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

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

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

Thus under Assumption **(A1)**, f and g admit a development on the Laguerre basis. Since X and Y are independent and nonnegative variables, we have a convolution relation between h, f and g. Starting from Equation (3) and using the decomposition on the Laguerre basis, the following equality holds

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

So 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

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

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

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

(see Comte et al. (2013)). And as

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

$\mathbf{G}_m$ 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

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

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 $\mathbf{G}_m = \mathrm{I}_m$, with 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

$$\mathbf{G}_m = \mathrm{I}_m \quad \text{and} \quad \hat{\vec{f}}_m = \hat{\vec{h}}_m. \tag{10}$$

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 section. Let us consider for example $b_k(S_Z)$ the k-th coefficient of S_Z

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

with Φ_k a primitive of φ_k defined as $\Phi_k(x) = \int_0^x \varphi_k(u)\,\mathrm{d}u$. We can notice that

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

where γ is the lower incomplete gamma function defined by formula 6.5.2. in Abramowitz and Stegun (1964). In order to apply the same method as for the density estimation, let us see how convolution is modified for survival functions. Let $z \geq 0$, by definition $S_Z(z) = \mathbb{P}(Z > z)$, and

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

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

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

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

$$S_{X,m}(x) = \sum_{k=0}^{m-1} b_k(S_X)\varphi_k(x). \tag{12}$$

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

$$\widehat{S}_{X,m}(x) = \sum_{k=0}^{m-1}\hat{b}_k(S_X)\varphi_k(x)$$
$$\text{with}\quad \widehat{\vec{S}}_{X,m} = \mathbf{G}_m^{-1}\left(\widehat{\vec{S}}_{Z,m} - \vec{S}_{Y,m}\right)\quad\text{and}\quad \hat{b}_k(S_Z) = \frac{1}{n}\sum_{i=1}^{n}\Phi_k(Z_i), \tag{13}$$

where $\vec{S}_{Y,m}$ is known since $b_k(S_Y) = \mathbb{E}\left[\Phi_k(Y_1)\right]$ 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. Indeed decomposing primitives of (φ_k) in the Laguerre basis is complicated.

3. Bounds on the $\mathbb{L}^2$ risk

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

3.1. Upper bounds.

Proposition 3.1. *Under **(A1)**, for $\mathbf{G}_m$ defined by (8) and $\hat{f}_m$ defined by (9), the following result holds*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n}\varrho^2(\mathbf{G}_m^{-1}). \tag{14}$$

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

Corollary 3.2. *Under **(A1)**, in the model without noise defined by (10) we get*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n}. \tag{15}$$

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

Proposition 3.3. *Under **(B1)** and if $\mathbb{E}[Z_1] < \infty$, for $\mathbf{G}_m$ defined by (8) and $\widehat{S}_{X,m}$ defined by (13), the following result holds*

$$\mathbb{E}\|S_X - \widehat{S}_{X,m}\|^2 \leq \|S_X - S_{X,m}\|^2 + \frac{\mathbb{E}[Z_1]}{n}\varrho^2(\mathbf{G}_m^{-1}). \tag{16}$$

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

Remark 2. The terms of the right-hand side of Equations (14), (15) and (16) 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 $\widehat{S}_{X,m}$ respectively defined by (9) and (13), we need to evaluate the smoothness of the signal along with the order of $\varrho^2\left(\mathbf{G}_m^{-1}\right)$. In the first place, we assume that f belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{ f : \mathbb{R}^+ \to \mathbb{R},\, f \in \mathbb{L}^2(\mathbb{R}^+),\, \sum_{k\geq 0} k^s b_k^2(f) \leq L < +\infty \right\} \quad \text{with} \quad s \geq 0 \tag{17}$$

where $b_k(f) = \langle f, \varphi_k \rangle$. Bongioanni and Torrea (2009) have introduced Laguerre-Sobolev space but the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2013). Indeed, let s be an integer, for $f : \mathbb{R}^+ \to \mathbb{R}$ 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 (2013). Now for $f \in W^s(\mathbb{R}^+, L)$ defined by (17),

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

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.

Proposition 3.5. *In the model without noise defined by (10), suppose that f belongs to $W^s(\mathbb{R}^+, L)$ defined by (17) 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 (14) and (16) which means assess the order of $\varrho^2\left(\mathbf{G}_m^{-1}\right)$. Comte et al. (2013) show that under the following conditions on the density g, we can recover the order of the spectral norm of $\mathbf{G}_m^{-1}$.

First we define an integer $r \geq 1$ such that

$$\frac{\mathrm{d}^j}{\mathrm{d}x^j} g(x) \mid_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \ldots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

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 $\infty + ib$.

Lemma 3.6 (Comte et al. (2013))**.** *If Assumptions **(C1)-(C2)** are true, then there exists a positive constants C_ϱ and C'_ϱ such that*

$$C'_\varrho m^{2r} \leq \varrho^2\left(\mathbf{G}_m^{-1}\right) \leq C_\varrho m^{2r}.$$

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

Proposition 3.7. *Assume that f belongs to $W^s(\mathbb{R}^+, L)$ defined by (17), that Assumptions **(C1)-(C2)** are fulfilled and let $m_{opt} \propto n^{1/(s+2r+1)}$, then*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E}\|f - \hat{f}_{m_{opt}}\|^2 \leq C_2(s, L, C_\varrho) n^{-s/(s+2r+1)}.$$

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 (17), that Assumptions **(C1)-(C2)** are fulfilled and let $m_{opt} \propto n^{1/(s+2r+1)}$, then*

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

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.

3.3. **In what context does the Laguerre procedure improve the Fourier approach ?** In this section we want to emphasize that for at least certain classes of functions the Laguerre method achieves better rates of convergence than Fourier method which are optimal if f belongs to a Sobolev class. Thus we consider Gamma distributions and mixed Gamma distributions which belong to a Sobolev class and are said to be ordinary smooth. We choose two differents distributions for Y: $\Gamma(q, \mu)$ and $\beta(a, b)$ for $a, b \in \mathbb{N}^*$ such that $b > a$. The results are reported in Table 1.

f	g	$\Gamma(q,\mu)$	$\beta(a,b)$ $b > a \geq 1$
$\Gamma(p,\theta)$	Laguerre	$\dfrac{(\log n)^{2q+1}}{n}$	$\dfrac{(\log n)^{2a+1}}{n}$
	Fourier	$n^{-(2p-1)/(2q+2p)}$	$n^{-(2p-1)/(2a+1)}$
$\mathcal{M}\Gamma$	Laguerre	$\dfrac{(\log n)^{2q+1}}{n}$	$\dfrac{(\log n)^{2a+1}}{n}$
	Fourier	$n^{-(2p^*-1)/(2q+2p^*)}$ $p^* = \min_i p_i$	$n^{-(2p^*-1)/(2a+1)}$ $p^* = \min_i p_i$

TABLE 1. Rates of convergence for the MISE

To derive the rates of convergence of Table 1 we have to do the bias variance tradeoff in the Laguerre and Fourier settings. First let us compute the squared bias where the signal X is a Gamma distribution $\Gamma(p, \theta)$ with $p \in \mathbb{N}^*$ and $\theta > 0$ in the Laguerre procedure and in the Fourier setting. In the Laguerre procedure we first need to compute the coefficients in the Laguerre basis of the signal

$$b_k(f) = \int_{\mathbb{R}^+} \theta^p u^{p-1} e^{-\theta u} \sqrt{2} e^{-u} L_k(2u)\,\mathrm{d}u = \sqrt{2}\theta^p \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}^+} u^{p+j-1} e^{-(1+\theta)u}\,\mathrm{d}u$$

$$= \frac{\sqrt{2}\theta^p}{(1+\theta)^p} \sum_{j=0}^{k} \binom{k}{j} \frac{(-2)^j}{(1+\theta)^j} \frac{(p+j-1)!}{j!} = \frac{\sqrt{2}\theta^p}{(1+\theta)^p} \frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}} \left[x^{p-1}(1-x)^k\right]\Big|_{x=2/(1+\theta)}. \quad (18)$$

It leads to

$$\|f - f_m\|^2 = \sum_{k \geq m} \left(\frac{\sqrt{2}\theta^p}{(1+\theta)^p} \frac{\mathrm{d}^{p-1}}{\mathrm{d}x^{p-1}} \left[x^{p-1}(1-x)^k\right]\Big|_{x=2/(1+\theta)} \right)^2 \leq C(p,\theta) \left(\frac{\theta-1}{\theta+1}\right)^{2m} m^{2(p-1)}$$

with $0 < C(p,\theta) < \infty$. For $X \sim \Gamma(p,\theta)$ with $\theta > 0$, we can note that the squared bias decays exponentially. It can also be noted that if $\theta = 1$ then the bias is null. The rate of convergence, in this case, reaches the parametric rate in the Laguerre setting. Moreover if $f \in \mathcal{M}\Gamma$ defined by

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

the squared bias decays exponentially

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

In the Fourier procedure, according to Comte et al. (2006), the squared bias is computed as follows

$$\|f - f_m\|^2 = \frac{1}{2\pi} \|f^* - f_m^*\|^2 = \frac{1}{2\pi} \int_{|t| \geq \pi m} |f^*(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{|t| \geq \pi m} (1 + \theta^2 t^2)^{-p} \, \mathrm{d}t \leq C m^{-2p+1}.$$

It yields that if $f \in \mathcal{M}\Gamma$ the order of the squared bias is $\max_{1 \leq i \leq p} m^{-2p_i+1}$. We notice in this setting, contrary to the Laguerre setting, that the bias is polynomial.

Secondly let us compute the order of the variance. If $g \sim \Gamma(q,\mu)$, for the Laguerre procedure the variance term is $2m\varrho^2(\mathbf{G}_m^{-1})/n$ and is upper bounded $2C_\varrho m^{2q+1}/n$ and if $g \sim \beta(a,b)$ it is upper bounded by $2C_\varrho m^{2a+1}/n$ for $b > a$, from Remark 3. We find the same orders of variance for the Fourier setting after computing $1/(2\pi n) \int_{-\pi m}^{\pi m} |g^*(t)|^{-2} \, \mathrm{d}t$. Thus we can derive the rates of convergence summarized in Table 1.

For the Laguerre procedure we see that the rates of convergence in both cases have order $(\log n)^\alpha/n$ with α depending on the model. While for the Fourier procedure we find classical rates of convergence of the deconvolution setting which are slower. For instance in the double Gamma case we have a rate of convergence $n^{-p/(2q+p+1)}$. Thus in the context of nonnegative variables of Gamma type, we recover faster rates of convergence with our Laguerre method than with a Fourier procedure. We can extend those results to the case of Exponential and Gamma mixtures. This context fits fields of survival analysis and duration models.

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.

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 $\inf_m \mathbb{E}\|S_X - \widehat{S}_{X,m}\|^2$. 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:

(A2) $\mathcal{M}_n^{(1)} = \left\{ 1 \leq m \leq d_1, \dfrac{m\varrho^2\left(\mathbf{G}_m^{-1}\right)}{n} \leq 1 \right\}$, where $d_1 < n$ may depend on n.

(A3) $\forall b > 0, \quad \displaystyle\sum_{m \in \mathcal{M}_n^{(1)}} \varrho^2\left(\mathbf{G}_m^{-1}\right) e^{-bm} < C(b) < \infty.$

We define the penalty as

$$\mathrm{pen}_1(m) = \frac{2\kappa_1 m}{n} \varrho^2\left(\mathbf{G}_m^{-1}\right) \tag{20}$$

where κ_1 is a numerical constant see our comment below.

Theorem 4.1. *Let us suppose that* ***(A1)-(A3)*** *are true. Let* $\hat{f}_{\widehat{m}_1}$ *be defined by (9) and*

$$\widehat{m}_1 = \underset{m\in\mathcal{M}_n^{(1)}}{\operatorname{argmin}} \left\{ -\|\hat{f}_m\|^2 + \operatorname{pen}_1(m) \right\}$$

with pen_1 *defined by (20), then there exists a positive numerical constant* κ_1 *such that*

$$\mathbb{E}\|f - \hat{f}_{\widehat{m}_1}\|^2 \leq 4 \inf_{m\in\mathcal{M}_n^{(1)}} \left\{ \|f - f_m\|^2 + \operatorname{pen}_1(m) \right\} + \frac{C}{n}, \tag{21}$$

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

It follows from the proof that $\kappa_1 = 32$ would suit. But in practice, values obtained from the theory are generally too large and the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments.

The oracle inequality (21) 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. Bias and variance orders in Equation (14) as shown in Section 3 yield an optimal value of m in function of n which m depends on unknown quantities, and thus cannot be implemented. That is why Equation (21) is of high interest: rates of convergence are reached without requiring to be specified in the framework.

Remark 6. 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 $\varrho^2\left(\mathbf{G}_m^{-1}\right) = \mathrm{O}(m^\alpha)$ then for any α, Assumption **(A3)** is true. It is satisfied for instance for Gamma distributions. Similarly if $\varrho^2\left(\mathbf{G}_m^{-1}\right) = \mathrm{O}(m^\alpha e^{\lambda m^\beta})$ then it is enough that $\beta < 1$ to ensure **(A3)**. If Assumption **(A3)** is released then an adaptive procedure can still be obtained with the associated penalty $\operatorname{pen}(m) = 2\kappa m\varrho^2\left(\mathbf{G}_m^{-1}\right)\log(n)/n$.

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

(B2) $\mathcal{M}_n^{(2)} = \left\{ 1 \leq m \leq d_2, \dfrac{\varrho^2\left(\mathbf{G}_m^{-1}\right)\log n}{n} \leq C \right\}$, where $d_2 < n$ may depend on n and $C > 0$.

(B3) $0 < \mathbb{E}[Z_1^3] < \infty$.

We define the penalty as

$$\operatorname{pen}_2(m) = \frac{\kappa_2 \mathbb{E}[Z_1]}{n} \varrho^2\left(\mathbf{G}_m^{-1}\right)\log n \tag{22}$$

Theorem 4.2. *Let us suppose that* ***(B1)-(B3)*** *are true. Let* $\widehat{S}_{X,\widehat{m}_2}$ *be defined by (13) and*

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

with pen_2 *defined by (22), then there exists a positive numerical constant* κ_2 *such that*

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

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 (16). 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 (22) which is made in the next corollary.

Corollary 4.3. *Let us suppose that* ***(B1)-(B3)*** *are true. Let $\widehat{S}_{X,\widetilde{m}_2}$ be defined by (13) and*

$$\widetilde{m}_2 = \underset{m\in\mathcal{M}_n^{(2)}}{\operatorname{argmin}} \left\{ -\|\widehat{S}_{X,m}\|^2 + \widehat{\operatorname{pen}}_2(m) \right\} \tag{23}$$

$$\widehat{\operatorname{pen}}_2(m) = \frac{2\kappa_2 \bar{Z}_n}{n} \varrho^2\left(\mathbf{G}_m^{-1}\right) \log n \quad \text{where} \quad \bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i, \tag{24}$$

then there exists a positive numerical constant κ_2 such that

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

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

5. Illustrations

The whole implementation is conducted using R software. The integrated squared errors $\|f - \hat{f}_{\widehat{m}}\|^2$ 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) $\mathbb{E}\|f - \hat{f}_{\widehat{m}_1}\|^2$ and $\mathbb{E}\|S_X - \widehat{S}_{X,\widetilde{m}_2}\|^2$ are computed as the empirical mean of the approximated ISE over 200 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 $\sigma^2 = 2/(4-\pi)$, $f(x) = \dfrac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$, $I_f = [0,5]$, $I_S = [0,25]$.
- ▷ Weibull, $X/\sqrt{\Gamma(4) - \Gamma(1+2/3)^2}$, $f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}\mathbb{1}_{x\ge 0}$, with $k = \frac{3}{2}$ and $\lambda = 1$, $I_f = [0,5]$, $I_S = [0,5]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5,1) + 0.6\Gamma(13,1)$, $I_f = [0,26]$, $I_S = [0,15]$.
- ▷ 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 σ^2 of the error distribution g takes the values 0, 1/10 and 1/4. The case where the variance σ^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 (18) for the computation of the matrix $\mathbf{G}_m$.

5.2. **Practical estimation procedure.** The adaptive procedure is then implemented as follows:

▷ For $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$, compute $-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)$.

▷ Choose $\hat{m}$ such that $\hat{m} = \underset{m\in\mathcal{M}_n}{\text{argmin}} \left\{-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\right\}$.

▷ 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 $\kappa_2 = 0.065$. We consider the two following model collections $\mathcal{M}_n^{(1)} = \left\{m \in \{1, \dots, n-1\}, \quad 1 \le m \le \lfloor n^{1/5} \rfloor\right\}$ and $\mathcal{M}_n^{(2)} = \left\{m \in \{1, \dots, n-1\}, \quad 1 \le m \le \lfloor (n \log n)^{1/4} \rfloor\right\}$ for the density and survival function estimation.

In order to measure the performances of our procedure (density estimation) to the literature, 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 $g^*(x) = \int e^{iux} g(u)\,\mathrm{d}u$. For a Gamma distribution of parameter p and θ, its Fourier transform is $g^*(u) = (1 - iu\theta)^{-p}$. We compute

$$\hat{f}_{F,m}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-ixu} \frac{n^{-1}\sum_{j=1}^n e^{iuZ_j}}{g^*(u)}\,\mathrm{d}u \qquad \text{pen}_F^{(1)}(m) = \frac{\kappa_F^{(1)}}{2\pi n}\int_{-\pi m}^{\pi m} \frac{\mathrm{d}u}{|g^*(u)|^2}$$

We select m by minimizing $-\|\hat{f}_{F,m}\|^2 + \text{pen}_F^{(1)}(m)$. If $Y = 0$, we set $g^* \equiv 1$ and $\text{pen}_F^{(2)}(m) = \kappa_F^{(2)} m/n$ The model collection is $\{m/10 : \quad m \in \mathbb{N}, \quad 1 \le m \le 50\}$. After calibration we find $\kappa_F^{(1)} = 41$ and $\kappa_F^{(2)} = 5$. We consider two different penalties since in the case of the model without noise, the estimator of f_m $\hat{f}_{F,m}$ 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 2 and 3. For both tables, the values of the MISE are multiplied by 100 for each case and computed from 200 simulated data. In Table 2 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 deconvolution has better performances than the Fourier deconvolution. For instance, when $n = 2000$ the MISE in the Fourier setting is almost systematically 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 of 1 to 3 and for the Weibull distribution 1 to 33. If σ^2 equals $1/10$ or $1/4$, we can make the same kind of remarks noticing that the ratio is worse for the Fourier deconvolution: 1 to 18.

		$\sigma^2=0$		$\sigma^2=\frac{1}{10}$		$\sigma^2=\frac{1}{4}$	
		200	2000	200	2000	200	2000
Exponential	Lag	0.874	0.118	0.833	0.090	0.699	0.109
	Fou	3.950	3.234	6.002	3.359	11.228	3.889
Gamma	Lag	0.243	0.048	1.181	0.076	1.872	0.384
$\Gamma(4,1/4)$	Fou	0.585	0.076	2.027	0.250	8.497	0.860
Gamma	Lag	0.332	0.035	1.348	0.540	4.180	0.916
$\Gamma(20,1/2)$	Fou	0.521	0.059	1.917	0.245	2.728	1.092
Rayleigh	Lag	0.287	0.044	0.844	0.068	1.008	0.083
	Fou	0.498	0.073	1.546	0.248	7.523	0.797
Weibull	Lag	1.714	0.275	8.518	1.529	13.768	3.543
	Fou	7.004	6.611	8.751	6.839	14.237	7.421
Mixed Gamma	Lag	0.333	0.032	1.568	0.359	2.806	0.365
	Fou	0.488	0.062	1.038	0.204	8.317	0.829
Chi-squared	Lag	0.357	0.037	0.443	0.260	2.861	0.315
	Fou	0.542	0.069	1.887	0.250	8.135	0.844
Pareto	Lag	10.72	10.78	12.53	10.50	13.89	10.83
	Fou	12.42	6.54	18.70	9.38	30.06	26.28

TABLE 2. Results of simulation: MISE $\mathbb{E}\left(\|f-\hat{f}_{\widehat{m}}\|^2\right)\times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2=0$ corresponds to the model without noise ($Y=0$). The noise is $\Gamma(2,\frac{1}{\sqrt{20}})$ for $\sigma^2=\frac{1}{10}$ and $\Gamma(2,\frac{1}{\sqrt{8}})$ for $\sigma^2=\frac{1}{4}$.

	$\sigma^2=0$		$\sigma^2=\frac{1}{10}$		$\sigma^2=\frac{1}{4}$	
	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 3. Results of simulation: MISE $\mathbb{E}\left(\|S_X-\widehat{S}_{X,\widetilde{m}_2}\|^2\right)\times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2=0$ corresponds to the model without noise ($Y=0$). The noise is $\Gamma(2,\frac{1}{\sqrt{20}})$ for $\sigma^2=\frac{1}{10}$ and $\Gamma(2,\frac{1}{\sqrt{8}})$ for $\sigma^2=\frac{1}{4}$.

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 $\sigma^2=1/4$.

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

In Table 3, 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.

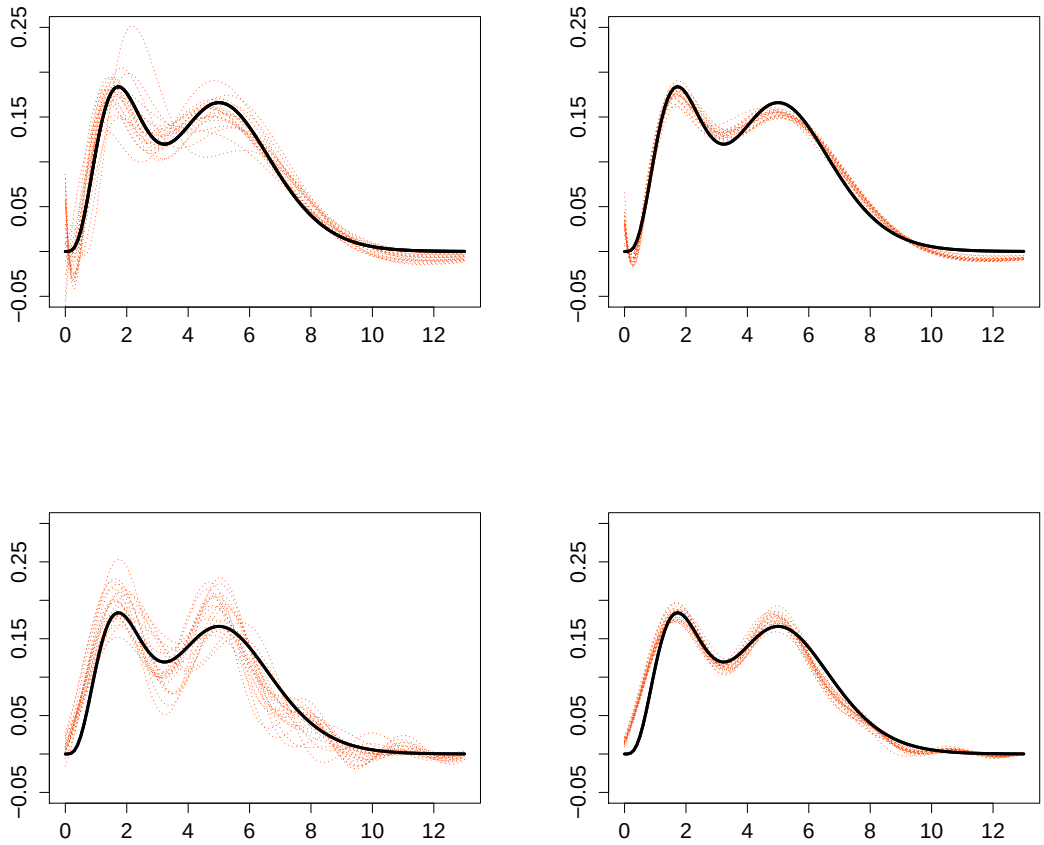

FIGURE 1. Estimation of the mixed Gamma density with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (top left for $n = 200$ and top right $n = 2000$), with $\sigma^2 = 1/10$.

6. CONCLUDING REMARKS

This paper deals with the estimation of nonnegative variables in a deconvolution setting with a known error distribution. First we have considered 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

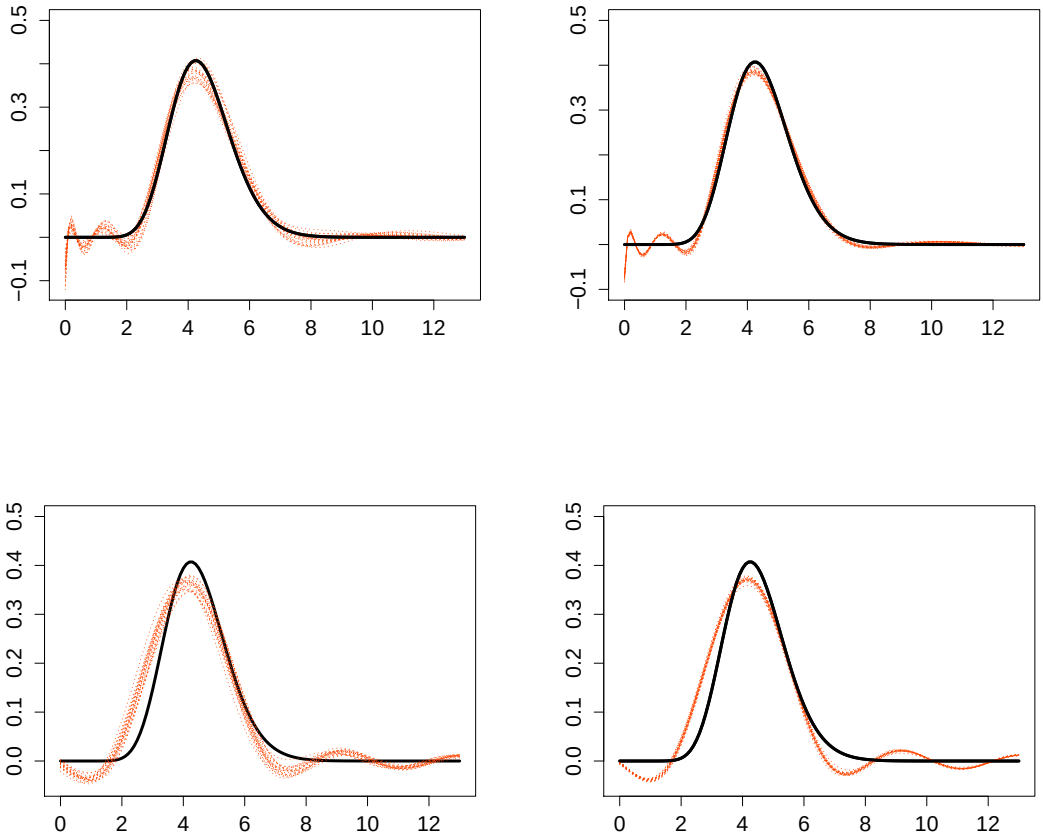

FIGURE 2. Estimation of the Gamma density $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$ with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (top left for $n = 200$ and top right $n = 2000$), with $\sigma^2 = 1/10$.

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

The assumption of the knowledge of the error distribution 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 distribution g on the Laguerre basis. Vareschi (2015), in a Laplace regression model, considers this problem ; he assumes that a pertubation 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 the true spectral norm of $\mathbf{G}_m^{-1}$.

7. PROOFS

7.1. Proof 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 (14). Let us study the second term: using the decomposition on the orthonormal Laguerre basis, we have

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

Finally we apply (6) and get

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

In the end we get: $\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \dfrac{2m}{n}\varrho^2(\mathbf{G}_m^{-1})$. □

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

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

We can notice that

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

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

$$\sum_{j=0}^{m-1}\Phi_j^2(Z_1) = \sum_{j=0}^{m-1}\left(\int \varphi_j(u)\mathbb{1}_{0\leq u\leq Z_1}\,\mathrm{d}u\right)^2 = \sum_{j=0}^{m-1}\langle\varphi_j, \mathbb{1}_{\cdot\leq Z_1}\rangle^2 \leq \|\mathbb{1}_{\cdot\leq Z_1}\|_{\mathbb{R}^+}^2 = Z_1 \tag{25}$$

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

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

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 $\vec{u}_m = \arg\max_{\|\vec{u}\|_{2,m}^2=1}\|\mathbf{T}_m\vec{u}\|_{2,m}$ with $\mathbf{T}_m$ a lower triangular matrix and $\mathbf{T}_m$ a submatrix of $\mathbf{T}_{m+1}$. We put $\vec{v}_{m+1}$ with m first coordinates equal to $\vec{u}_m$. It yields that if we note $(a_i)_{1\leq i\leq m+1}$ the coefficients of the vector $\mathbf{T}_{m+1}v_{m+1}$ we get

$$\|\mathbf{T}_{m+1}\vec{v}_{m+1}\|_{2,m+1}^2 = \sum_{i=1}^{m+1}a_i^2 = \sum_{i=1}^{m}a_i^2 + a_{m+1}^2 = \|\mathbf{T}_m\vec{u}_m\|^2 + a_{m+1}^2 \geq \varrho^2(\mathbf{T}_m).$$

□

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

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

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

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

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

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

So we clearly have that

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

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

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

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,

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

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

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

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

It yields that

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

which implies

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

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

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

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

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

we then have

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

which implies

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

Using Equation (26), we have

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

which implies

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

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

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

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 $p_1(m,m') = \dfrac{8(m\vee m')}{n}\varrho^2\left(\mathbf{G}_{m\vee m'}^{-1}\right)$*

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

In the end: $\mathbb{E}\|f - \hat{f}_{\widehat{m}}\|^2 \leq 4\inf_{m\in\mathcal{M}_n^{(1)}}\left\{\|f - f_m\|^2 + \mathrm{pen}_1(m)\right\} + \dfrac{C_1}{n}$, as soon as $\kappa_1 \geq 32$. □

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

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

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

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

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

We now apply Cauchy-Schwarz inequality and get

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

We then set $H:=\sqrt{\frac{2m^*}{n}\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}$.

Before deriving a bound for the term of variance, let us remind that for any $x\in\mathbb{R}^+$, we have

$$h(x)=\int f(u)g(x-u)\mathbb{1}_{u\ge0}\mathbb{1}_{x-u\ge0}\,\mathrm{d}u=\int_0^x f(u)g(x-u)\,\mathrm{d}u.$$

Then we apply Cauchy-Schwarz inequality and get $\forall x\in\mathbb{R}^+$, $h(x)\le\|f\|\|g\|$. If Assumption **(A1)** is fulfilled, we clearly have $\|h\|_\infty\le\|f\|\|g\|<\infty$.

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

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

which implies

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

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

• Now applying Cauchy-Schwarz inequality

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

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

• We can now apply Talagrand's inequality for $\xi^2=1/2$

$$\mathbb{E}\left[\left\{\sup_{\vec{t}\in\mathcal{B}(m,m')}|\nu_n(\vec{t})|^2-8m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)\right\}_+\right]\leq C_1\frac{\|h\|_\infty\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n}e^{-C_2am^*/\|h\|_\infty}+C_3\frac{m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n^2}e^{-C_4\sqrt{n}}$$

which implies that

$$\sum_{m'\in\mathcal{M}_n^{(1)}}\mathbb{E}\left[\left\{\sup_{\vec{t}\in\mathcal{B}(m,m')}|\nu_n(\vec{t})|^2-8m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)\right\}_+\right]$$
$$\leq C\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{\|h\|_\infty\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n}e^{-C_2\frac{m^*}{\|h\|_\infty}}+\frac{m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n^2}e^{-C_4\sqrt{n}}$$

Yet under Assumption **(A3)**, we have

$$\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{\|h\|_\infty\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n}e^{-C_2\frac{m^*}{\|h\|_\infty}}\leq\frac{C}{n}.$$

Moreover according to Assumption **(A2)**, we also have

$$\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n^2}e^{-C_4n}\leq\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{e^{-C_4\sqrt{n}}}{n}\leq\frac{C}{n}.$$

In the end we have the desired result. □

Proof of Remark 6. We have to prove that Proposition 7.1 is still valid although Assumption **(A3)** is no longer true. We set $\xi^2=2\log n/K_1$, $H^2=2m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}/\,n\right.$, $v=2m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)$, $M_1=\sqrt{2m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}$. Under **(A2)**, we have

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

For $C(\xi)$ defined in Lemma A.1, we get

$$\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{M_1^2}{K_1C^2(\xi^2)n^2}\exp\left(-K_1\frac{C(\xi)\xi nH}{\sqrt{2}M_1}\right)$$
$$=\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{K_1C^2(\xi^2)n^2}\exp\left(-K_1C(\xi)\xi\sqrt{n}\right)\leq\sum_{m'\in\mathcal{M}_n^{(1)}}\frac{C_2m^*\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{(\log n)^2n^2}\exp\left(-C_1\log n\sqrt{n}\right)$$
$$\leq\frac{C_3|\mathcal{M}_n^{(1)}|}{(\log n)^2n}\exp\left(-C_1\log n\sqrt{n}\right)\leq\frac{C_3}{n}.$$

□

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

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

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,

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

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

So we have the following inequality

$$\|\widehat{S}_{X,\widehat{m}} - S_X\|^2 \le 3\|S_X - S_{X,m}\|^2 + 2\mathrm{pen}_2(m) + 16\left(\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2\right) - 2\mathrm{pen}_2(\widehat{m}). \tag{28}$$

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

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

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 $q(m,m') = \kappa_2 \varrho^2 \left(\mathbf{G}_{m\vee m'}^{-1}\right) \mathbb{E}[Z_1] \dfrac{\log n}{n}$*

$$(i) \quad \mathbb{E}\left[\left\{\sup_{\vec{t}\in\mathcal{B}(m,m')} (\zeta_n^{(1)}(\vec{t}))^2 - q(m,m')\right\}_+\right] \le \frac{C}{n}$$

$$(ii) \quad \mathbb{E}\left[\sup_{\vec{t}\in\mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2\right] \le \frac{\mathbb{E}[Z_1^3]}{n}.$$

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

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

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

We now apply Cauchy-Schwarz inequality and get

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

It follows that, with Equation (25)

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

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

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

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

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

• First notice again

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

Now applying Cauchy-Schwarz inequality and using Equation (25) again

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

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

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

$$\begin{aligned}&\mathbb{E}\left[\left\{\sup_{\vec{t}\in\mathcal{B}(m,m')}|\zeta_n^{(1)}(\vec{t})|^2-\kappa_2\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)\mathbb{E}[Z_1]\frac{\log n}{n}\right\}_+\right]\\&\quad\leq\frac{4}{K_1}\frac{\mathbb{E}[Z_1]\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n}e^{-K_1\xi^2}+\frac{98n\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{K_1n^2C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\sqrt{\mathbb{E}[Z_1]}n^{1/4}}\end{aligned}$$

which implies that

$$\begin{aligned}&\sum_{m'\in\mathcal{M}_n^{(2)}}\mathbb{E}\left[\left\{\sup_{\vec{t}\in\mathcal{B}(m,m')}|\zeta_n^{(1)}(\vec{t})|^2-\kappa_2\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)\mathbb{E}[Z_1]\frac{\log n}{n}\right\}_+\right]\\&\quad\leq\sum_{m'\in\mathcal{M}_n^{(2)}}\frac{4}{K_1}\frac{\mathbb{E}[Z_1]\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{n}e^{-K_1\xi^2}+\frac{98\sqrt{n}\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{K_1n^2C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\sqrt{\mathbb{E}[Z_1]}n^{1/4}}.\end{aligned}$$

Thus we have under Assumption **(B2)**

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

And

$$\sum_{m'\in\mathcal{M}_n^{(2)}}\frac{\sqrt{n}\varrho^2\left(\mathbf{G}_{m^*}^{-1}\right)}{K_1n^2C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\sqrt{\mathbb{E}[Z_1]}n^{1/4}}\leq\sum_{m'\in\mathcal{M}_n^{(2)}}\frac{C\sqrt{n}}{(\log n)^2n}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\sqrt{\mathbb{E}[Z_1]}n^{1/4}}\leq\frac{C}{n}.$$

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

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

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

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

In the end we have the desired result. □

7.4. **Proof of Corollary 4.3].** The beginning of the proof is the same as the proof of Theorem 4.2 except that we consider $\widetilde{m}_2$ (defined by Equation (23)) instead of $\widehat{m}_2$ and $\widehat{\mathrm{pen}}_2$ (defined by Equation (24)) instead of pen_2. Starting from Equation (28) we have

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

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*

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

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

Proof of Proposition 7.3. First let us notice the following

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

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

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

which yields

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

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

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

Now we apply Cauchy-Schwarz

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

We apply Markov inequality then Rosenthal inequality

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

Moreover under Assumption **(B2)**

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

which ends the proof. □

Appendix A.

Lemma A.1. *(Talagrand's inequality) Let $Y_1, \ldots, Y_n$ be i.i.d. variables and*

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

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

$$\mathbb{E}\left[\left\{\sup_{f\in\mathcal{F}}|r_n(f)|^2 - 2(1+2\xi^2)H^2\right\}_+\right] \leq \frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\xi^2\frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)}e^{-\frac{2K_1C(\xi)\xi}{7\sqrt{2}}\frac{nH}{M_1}}\right)$$

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

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

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.

Bibinger, M., Jirak, M., and Reiß, M. (2014). Improved volatility estimation based on limit order books. Preprint, arXiv:1408.3768.

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. *The Canadian Journal of Statistics*, 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. *Journal of the American Statistical Association*, 83(404):1184–1186.

Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2013). Laplace deconvolution and its application to dynamic contrast enhanced imaging. Prépublication MAP5 2012-17.

Comte, F. and Genon-Catalot, V. (2013). Adaptive Laguerre density estimation for mixed Poisson models. Preprint, hal-00848158.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *The Canadian Journal of Statistics*, 3(34):431–452.

Dattner, I., Goldenshluger, A., and Juditsky, A. (2011). On deconvolution of distribution functions. *The Annals of Statistics*, 39(5):2477–2501.

Dattner, I. and Reiser, B. (2013). Estimation of distribution functions in measurement error models. *Journal of Statistical Planning and Inference*, 143(3):479 – 493.

Dattner, I., Reiß, M., and Trabs, M. (2014). Adaptive quantile estimation in deconvolution with unknown error distribution. *ArXiv e-prints 1303.1698, To appear in Bernoulli.*

Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *Journal of the American Statistical Association*, 92:526–535.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574.

Hall, P. and Lahiri, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *The Annals of Statistics*, 36(5):2110–2134.

Jirak, M., Meister, A., and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *The Annals of Statistics*, 42(5):1970–2002.

Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323.

Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution.

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 bheld in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

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. *Journal of Nonparametric Statistics*, 7(4):307–330.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6):2033–2053.

Rebafka, T. and Roueff, F. (2010). Nonparametric estimation of the mixing density using polynomials. *ArXiv e-prints 1002.4516*.

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. *Statistics and Probability Letters*, 9:229–235.

Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21:169–184.

Vareschi, T. (2015). Noisy Laplace deconvolution with error in the operator. *Journal of Statistical Planning and Inference*, 157–158:16–35.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, 18(2):806–831.