LAGUERRE ESTIMATION FOR k-MONOTONE DENSITIES OBSERVED WITH NOISE

D. BELOMESTNY[(1)], F. COMTE[(2)] & V. GENON-CATALOT[(3)]

ABSTRACT. We study the models $Z_i = Y_i + V_i, Y_i = X_i U_i,\ i = 1, \dots, n$ where the V_i's are nonnegative, *i.i.d.* with known density f_V, the U_i's are *i.i.d.* with $\beta(1,k)$ density, $k \geq 1$, the X_i's are *i.i.d.*, nonnegative with unknown density f. The sequences $(X_i), (U_i), (V_i)$ are independent. We aim at estimating f on $\mathbb{R}^+$ in the three cases of direct observations $(X_1, \dots, X_n)$, observations $(Y_1, \dots, Y_n)$, observations $(Z_1, \dots, Z_n)$. We propose projection estimators using a Laguerre basis and give upper bounds on the $\mathbb{L}^2$-risks on specific Sobolev-Laguerre spaces. Lower bounds matching with the upper bounds are proved in the case of direct observation of X and in the case of observation of Y. A general data-driven procedure is described and proved to perform automatically the bias variance compromise. The method is illustrated on simulated data.

(1) Duisburg-Essen University, email: denis.belomestny@uni-due.de,
(2) Université Paris Descartes, MAP5, UMR CNRS 8145, email: fabienne.comte@parisdescartes.fr
(3) Université Paris Descartes, MAP5, UMR CNRS 8145, valentine.genon-catalot@parisdescartes.fr.

Keywords. Adaptive estimation. Lower bounds. Model selection. Multiplicative censoring. k-monotone densities. Projection estimator.
MSC2010. 62G07

1. Introduction

Consider observations $Z_1, \dots, Z_n$ such that

$$(1) \qquad Z_i = Y_i + V_i, \quad Y_i = X_i U_i, \quad i = 1, \dots, n.$$

where X_i, U_i, V_i are nonnegative random variables, (X_i) are *i.i.d.* with unknown density f, (U_i) are *i.i.d.*, (V_i) are *i.i.d.*, U_i, V_i have known densities and the sequences $(X_i), (U_i), V_i)$ are independent.
If $V_i = 0$ and U_i has uniform density on $[0,1]$, the model $Z_i = X_i U_i = Y_i$ is called multiplicative censoring model and has been widely investigated in the past decades. As detailed in Vardi (1989), this model covers several important statistical problems, in particular estimation under monotonicity constraints. In this context, numerous papers deal with the estimation of f whether by nonparametric maximum likelihood (Vardi (1989), Vardi and Zhang (1992), Asgharian *et al.* (2012), by projection methods (Andersen and Hansen (2001), Abbaszadeh *et al.* (2012,2013)) or kernel methods (Brunel *et al.*, (2015)). We mention that taking logarithm of the Y_i allows a deconvolution method used by some authors (see e.g. van Es *et al.* (2003)): the method leads to estimate a distortion of f and cannot be straightforwardly extended to general model (1). Another approach to the above multiplicative deconvolution problem can be based on the Mellin transform technique (see Belomestny and Schoenmakers (2015) for a related problem).
On the other hand, density estimation from noisy observations, *i.e.*, estimation of the density f_Y

Date: January 13, 2016.

of Y_i when observing $Z_i = Y_i + V_i$ is also the subject of a huge number of contributions. For real-valued random variables, this deconvolution problem is classically solved by Fourier methods. However, recently, the study of one-sided errors, *i.e.* $V_i \geq 0$, was motivated by applications in the field of finance (see Jirak *et al.* (2014)) or in survival models, (see van Es *et al.* (1998), Jongbloed (1998)). Specific new approaches have been introduced to deal with deconvolution of nonnegative random variables. In particular, Mabon (2015) considers the case of $Y_i, V_i \geq 0$, V_i with known density f_V on $\mathbb{R}^+$ and proposes a projection estimator of f_Y using a Laguerre basis of $\mathbb{L}^2(\mathbb{R}^+)$ whose properties allow deconvolution of densities on $\mathbb{R}^+$.
In this paper, we consider model (1) and extend the class of distributions for the U_i's. We assume that U_i has $\beta(1,k)$ distribution with a density $f_U(u) := \rho_k(u) = k(1-u)^{k-1}\mathbf{1}_{[0,1]}(u)$, for $k \geq 1$. Then the random variables $Y_i = X_i U_i$ have density

$$f_{k,Y}(y) = f \odot \rho_k(y) = k \int_y^{\infty} \left(1 - \frac{y}{u}\right)^{k-1} \frac{f(u)}{u}\, du, \quad y \geq 0, \tag{2}$$

where $f \odot h$ denotes the density of XU with X, U independent, X with density f, U with density h . For $k = 1$, $f'_{1,Y}(y) = -f(y)/y \leq 0$ for all $y \geq 0$ so that $f_{1,Y}$ is a nonincreasing density on $\mathbb{R}^+$.
More generally, a k-monotone density follows the definition:

Definition 1.1. *A density g on $[0, +\infty[$ is k-monotone if $(-1)^{\ell} g^{(\ell)}$ is nonincreasing and convex for $\ell = 0, \ldots, k-2$ if $k \geq 2$ and simply nonincreasing if $k = 1$.*

In particular, it has been proved (see Williamson (1956)) that a k-monotone density g admits the representation

$$g(y) = k \int_0^{\infty} \frac{(u-y)_+^{k-1}}{u^k}\, dF(u) = k \int_0^{\infty} \frac{(1-y/u)_+^{k-1}}{u}\, dF(u) \tag{3}$$

where F is a distribution function $\mathbb{R}_+$. Consequently, $f_{k,Y}$ given by (2) is k-monotone with $dF(u) = f(u)du$. Therefore, the multiplicative censoring model is a special case of the setting of k-monotone densities. The problem of estimating f in model (1) is identical to the problem of reconstructing f for a k-monotone density $f_{k,Y}$ from noisy observations $Z_i = Y_i + V_i$. Note that k-monotone densities have been considered by Balabdaoui and Wellner (2007, 2010) or Chee and Wang (2014), from the point of view of estimating $f_{k,Y}$ (not f) under the k-monotonicity constraint.

In this paper, we propose nonparametric estimators of f built as projection estimators on a Laguerre basis. This basis is related to specific Sobolev-Laguerre spaces which have been introduced by Shen (2000) and Bongioanni and Torrea (2007). The link between projection coefficients and regularity conditions in these spaces has been described in Comte and Genon-Catalot (2015). We consider the three cases: first, the observations are the X_i's, second the Y_i's and third the Z_i's. In each case, we propose estimators of f. In the second and third cases, the estimators are obtained thanks to explicit formulae linking the projection coefficients of $f_{k,Y}$ (second case) or f_Z (third case) in the Laguerre basis, to those of f . We provide risk bounds for the estimators, allowing to compute upper bounds for the rates of convergence. Then we study lower bounds: even in the case of direct observations of the X_i's, that is in the simple density model, the corresponding lower bound over Sobolev-Laguerre balls is very difficult to establish, and this is why we start with this supposedly elementary case. Upper and lower bounds match up to a logarithmic term. Next, we prove a lower bound for the 1-monotone case, almost matching the upper bound. Lastly, we provide a general adaptive procedure for estimation in model (1). In Section 2, we describe the basis and the Sobolev-Laguerre spaces together with their useful properties. In Section 3, we provide a collection of projection estimators of f from observations $X_1, \ldots, X_n$, and a straightforward upper bound on the rate of convergence of the estimator. The difficult part is to establish a lower bound. Section 4 follows the same steps from observations $Y_1, \ldots, Y_n$, with new investigations to relate Laguerre projection coefficients of $f_{k,Y}$ and f. The lower bound for the case $k = 1$ relies on a clever modification of the Hamming distance and the Varshamov-Gilbert Lemma (see Tsybakov (2009)). In Section 5, we define a collection of projection estimators of f from observations $Z_1, \ldots, Z_n$, together with a general adaptive procedure which can work in all three cases.

2. About Laguerre bases and spaces

2.1. Laguerre basis. Below we denote the scalar product and the $\mathbb{L}^2$-norm on $\mathbb{L}^2(\mathbb{R}^+)$ by:

$$\forall s, t \in \mathbb{L}^2(\mathbb{R}^+),\ \langle s, t\rangle = \int_0^{+\infty} s(x)t(x)dx,\ \|t\|^2 = \int_0^{+\infty} t^2(x)dx.$$

Consider the Laguerre polynomials (L_j) and the Laguerre functions (φ_j) given by

$$L_j(x) = \sum_{k=0}^{j}(-1)^k\binom{j}{k}\frac{x^k}{k!}, \qquad \varphi_j(x) = \sqrt{2}L_j(2x)e^{-x}\mathbf{1}_{x\geq 0}, \quad j \geq 0.$$

The collection $(\varphi_j)_{j\geq 0}$ constitutes a complete orthonormal system on $\mathbb{L}^2(\mathbb{R}^+)$, and is such that (see Abramowitz and Stegun (1964)):

$$\forall j \geq 1,\ \ \forall x \in \mathbb{R}^+,\ \ |\varphi_j(x)| \leq \sqrt{2}. \tag{4}$$

For $h \in \mathbb{L}^2(\mathbb{R}^+)$, we can develop h on the Laguerre basis with:

$$h = \sum_{j\geq 0} a_j(h)\varphi_j, \qquad a_j(h) = \langle h, \varphi_j\rangle.$$

Note that, when h is a density,

$$a_0(h) = \langle h, \varphi_0\rangle = \sqrt{2}\int_0^{+\infty} h(x)e^{-x}dx > 0.$$

By convention, we set $\varphi_j = 0$ if $j \leq -1$ and define the vector of coefficients of h on $(\varphi_0, \ldots, \varphi_{m-1})$:

$$\vec{a}_{m-1}(h) := (a_j(h))_{0\leq j\leq m-1}.$$

We define the m-dimensional space $S_m = \mathrm{span}(\varphi_0, \varphi_1, \ldots, \varphi_{m-1})$. The function

$$h_m = \sum_{j=0}^{m-1} a_j(h)\varphi_j$$

is the orthogonal projection of h on S_m.

2.2. Sobolev-Laguerre spaces. For $s \geq 0$, the Sobolev-Laguerre space with index s (see Bongioanni and Torrea (2007)) is defined by:

$$W^s = \{h : \mathbb{R}^+ \to \mathbb{R},\ h \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k\geq 0} k^s a_k^2(h) < +\infty\}. \tag{5}$$

The following results have been proved in Section 7 of Comte and Genon-Catalot (2015). For s integer, if $h : \mathbb{R}^+ \to \mathbb{R}$ belongs to $\mathbb{L}^2(\mathbb{R}^+)$, then

$$|h|_s^2 := \sum_{k\geq 0} k^s a_k^2(h) < +\infty. \tag{6}$$

is equivalent to the property that h admits derivatives up to order $s-1$, with $h^{(s-1)}$ being absolutely continuous and for $m = 0, \dots, s-1$, the functions

$$\xi_{m+1}(x) := x^{(m+1)/2}(h(x)e^x)^{(m+1)}e^{-x} = x^{(m+1)/2}\sum_{j=0}^{m+1}\binom{m+1}{j}h^{(j)}(x)$$

belong to $\mathbb{L}^2(\mathbb{R}^+)$. Moreover, for $m = 0, 1, \dots, s-1$,

$$\|\xi_{m+1}\|^2 = \sum_{k\geq m+1} k(k-1)\dots(k-m)a_k^2(h).$$

For $h \in W^s$ with s integer, we set $\|h\|_0^2 = \|h\|^2$ and for $s \geq 1$

$$\|h\|_s = \|\xi_s\| = [\sum_{k\geq s} k(k-1)\dots(k-s+1)a_k^2(h)]^{1/2}. \tag{7}$$

Now we set

$$|||h|||_s^2 := \sum_{j=0}^{s}\|h\|_j^2.$$

Then the following property holds.

Lemma 2.1. *When s is integer, the two norms $|||h|||_s$ and $|h|_s$ are equivalent.*

We define the ball $W^s(D)$ via (see (5)-(6)):

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

3. Projection estimator of f in the Laguerre basis when X_i's are observed

3.1. **Upper bound.** We assume that f belongs to $\mathbb{L}^2(\mathbb{R}^+)$ and provide for each $m \geq 1$, a projection estimator of f by estimating the coefficients $a_j(f), j = 0, \dots, m-1$. In the case where the X_i's are observed, we define

$$\hat{a}_j(X) = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i) \quad \text{and} \quad \hat{f}_m^X = \sum_{j=0}^{m-1}\hat{a}_j(X)\varphi_j.$$

Doing so, we obtain an estimator of $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$.

We have $\|\hat{f}_m^X - f\|^2 = \|f - f_m\|^2 + \|\hat{f}_m^X - f_m\|^2$ by the Pythagoras Theorem. As $(\varphi_j)_j$ is orthonormal, we get $\|\hat{f}_m^X - f_m\|^2 = \sum_{j=0}^{m-1}(\hat{a}_j(X) - a_j(f))^2$. Clearly, $\hat{a}_j(X)$ is an unbiased estimator of $a_j(f)$, so we have

$$\mathbb{E}[(\hat{a}_j(X) - a_j(f))^2] = \frac{1}{n}\mathrm{Var}(\varphi_j(X)) \leq \frac{1}{n}\mathbb{E}(\varphi_j^2(X)).$$

Therefore, with (4), we obtain the risk bound:

$$\mathbb{E}(\|\hat{f}_m^X - f\|^2) \leq \|f - f_m\|^2 + 2\frac{m}{n}. \tag{8}$$

Remark 3.1. The risk bound decomposition given in (8) classically involves a bias term $\|f - f_m\|^2 = \sum_{j\geq m} a_j^2(f)$ which is decreasing with m and a variance term of order m/n which is increasing with m. Therefore, to evaluate the rate of convergence, we have to perform a compromise to select relevantly m.

For $f \in W^s(D)$, we have $\|f - f_m\|^2 = \sum_{j\geq m} a_j^2(f) \leq Dm^{-s}$ and choosing $m = m_n = C_0(s, D)n^{1/(s+1)}$ in the upper bound (8), implies

$$\mathbb{E}(\|\hat{f}^X_{m_n} - f\|^2) \leq C_1(s, D)n^{-s/(s+1)}$$

where $C_i(s, D)$, $i = 0, 1$ are constants depending on s and D only. What is unexpectedly difficult is to prove that this rate is optimal on the Sobolev-Laguerre spaces.

Remark 3.2. Faster rates of convergence may be obtained if the bias is smaller. Exponential distributions provide examples of such a case. If X has exponential distribution $\mathcal{E}(\theta)$, $\theta > 0$, then the coefficients are given by $a_k(f) = \sqrt{2}[\theta/(\theta + 1)]\,((\theta - 1)/(\theta + 1))^k$ and the bias can be explicitly computed,

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} a_k^2(f) = \frac{\theta}{2}\left|\frac{\theta - 1}{\theta + 1}\right|^{2m}.$$

Then the bias is exponentially decreasing and the rate of convergence is of order $[\log(n)]/n$ for $m_n = \log(n)/\rho$, $\rho = |\log[|(\theta - 1)/(\theta + 1)|]|$. The result can be extended to Gamma and mixed Gamma densities, see Comte and Genon-Catalot (2015), Mabon (2015).

3.2. **Lower bound.** In the section, we prove the following result

Theorem 3.1. *Assume that s is an integer, $s > 1$.*
Then for any estimator $\hat{f}_n$ built as a measurable function of $X_1, \ldots, X_n$, for any $\epsilon > 0$ and for n large enough,

$$\sup_{f\in W^s(D)} \mathbb{E}_f\left[\|\hat{f}_n - f\|^2\right] \gtrsim \psi_n, \quad \psi_n = n^{-s/(s+1)}/\log^{(1+\epsilon)/(s+1)}(n).$$

The proof is based on Theorem 2.7 in Tsybakov (2009), and induces several steps. The main difficulty of the construction is to ensure that the density alternative proposal is really a density on $\mathbb{R}^+$, and in particular nonnegative.

Proof of Theorem 3.1. Define f_0 as the density

$$f_0(x) = \frac{c_{\alpha,\beta}}{(e + x)^\alpha \log^\beta(e + x)}\mathbf{1}_{\mathbb{R}^+}(x)$$

where $\alpha > 1$, and $\beta = (1 + \epsilon)/2 > 1/2$ with $\epsilon < 1$, and $c_{\alpha,\beta}$ is such that $\int f_0 = 1$. Note that as $\forall x \geq 0$, $1 \leq \log(e + x) \leq e + x$, we have, as $\beta < 1$,

$$\frac{c_{\alpha,\beta}}{(e + x)^{\alpha+1}} \leq f_0(x) \leq \frac{c_{\alpha,\beta}}{(e + x)^\alpha}. \tag{9}$$

Next we consider the functions

$$f_{\boldsymbol{\theta}}(x) := f_0(x) + \delta \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x)$$

for some $\delta > 0$, $K \in \mathbb{N}$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \{0, 1\}^K$.

Lemma 3.1. *Let $s > 1$ be integer. Then f_0 and $f_{\boldsymbol{\theta}}$ belong to $W^s(D)$ provided that $\alpha \geq (s + 1)/2(> 1)$ and $\delta^2K^{s+1} \leq D/C$ for some constant $C = C(s) > 0$.*

Lemma 3.2. *Suppose that $\sum_{k=1}^K \theta_k(-1)^k = 0$ and all partial sums $\sum_{k=1}^p \theta_k(-1)^k$, $p = 1, \ldots, K$, are uniformly bounded by 1, then under the choice $\delta = \delta' K^{-\alpha} \log^{-\beta}(K)$ for small enough constant $\delta' > 0$ not depending on K, we have that $f_{\boldsymbol{\theta}}$ is a probability density on $\mathbb{R}_+$.*

Next we have for any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \{0,1\}^K$,

$$\int_0^\infty \left(f_{\boldsymbol{\theta}}(x) - f_{\boldsymbol{\theta}'}(x)\right)^2 \, dx = \delta^2 \sum_{k=1}^K \left(\theta_k - \theta_k'\right)^2 = \delta^2 \rho(\boldsymbol{\theta}, \boldsymbol{\theta}'), \tag{10}$$

where $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{k=1}^K \mathbf{1}_{\theta_k \neq \theta_k'}$ is the so-called Hamming distance. Now to apply Theorem 2.7 p.101 in Tsybakov (2009), we need to extend the Varshamov-Gilbert bound (see Lemma 2.9 p. 104 in Tsybakov (2009)) as follows.

Lemma 3.3. *Fix some even natural number $K > 0$. There exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ and a constant $A_1 > 0$, such that $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$, all partial sums $\sum_{k=1}^N \theta_k^{(j)}(-1)^k$, $N = 1, \dots, K$, are uniformly bounded by 1,*

$$\sum_{k=1}^K \theta_k^{(j)}(-1)^k = 0 \quad \text{and} \quad \rho(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(l)}) \geq A_1 K,$$

for all $0 \leq j < l \leq M$. Moreover it holds that

$$M \geq 2^{A_2 K} \tag{11}$$

for some constant $A_2 > 0$.

Then we have the following Lemma.

Lemma 3.4.

$$\frac{1}{M} \sum_{j=1}^M \chi^2 \left((f_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (f_0)^{\otimes n}\right) \lesssim n\delta^2 K^{\alpha+4} \quad \text{and for } 0 \leq j \neq l \leq M, \ \|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2 \gtrsim \delta^2 K.$$

Now we are in position to end the proof of Theorem 3.2. Under the choices

$$\delta^2 = (\delta')^2 K^{-2\alpha} (\log K)^{-(1+\epsilon)} \text{ and } K \asymp (n/\log^{1+\epsilon}(n))^{1/(2\alpha)}$$

using inequality (11), $K \leq \log M/(A_2 \log 2)$, we get

$$\frac{1}{M} \sum_{j=1}^M \chi^2 \left((f_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (f_0)^{\otimes n}\right) \lesssim \log^{\alpha+4}(M)$$

and

$$\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2 \gtrsim (n/\log^{1+\epsilon})^{(1-2\alpha)/2\alpha}$$

for all $0 \leq j \neq l \leq M$. Finally, by taking $\alpha = (s+1)/2$ (recall that $\alpha \geq (s+1)/2$) and arbitrary small $\epsilon > 0$, we derive

$$\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2 \gtrsim (n/\log^{1+\epsilon}(n))^{-s/(s+1)} \log^{-(1+\epsilon))}(n) = n^{-s/(s+1)} [\log(n)]^{-(1+\epsilon)/(s+1)}.$$

This ends the proof of Theorem 3.2. □

4. Projection estimator of f in the Laguerre basis when Y_i's are observed

Now, our aim is to build an estimator of f from the observations $Y_1, \dots, Y_n$, still taking into account that all variables are nonnegative.

4.1. **Preliminary properties and formulas.** The construction in this case relies on the following steps.

The inversion formula giving f from $f_{k,Y}$ defined by (2) stated in Proposition 4.1 is given in Williamson (1956). For convenience of the reader, we give a proof in the appendix.

Proposition 4.1. *Let $f_{k,Y}$ and f be linked by Formula (3) and set $F(x) = \int_0^x f(t)\,dt$ (resp. $F_{k,Y}(y) = \int_0^y f_{k,Y}(t)\,dt$). Then we have, for any $y \geq 0$, for $k \geq 1$,*

$$f(y) = \frac{(-1)^k}{k!} y^k f_{k,Y}^{(k)}(y), \tag{12}$$

$$F(y) = F_{k,Y}(y) - y f_{k,Y}(y) + \cdots + \frac{(-1)^{k-1}}{(k-1)!} y^{k-1} f_{k,Y}^{(k-2)}(y) + \frac{(-1)^k}{k!} y^k f_{k,Y}^{(k-1)}(y). \tag{13}$$

With the two following Propositions, we give the links between the coefficients of f and $f_{k,Y}$ on the Laguerre basis which are used for the estimation procedure.

Proposition 4.2. *Assume that $\mathbb{E}X^{k-1} < +\infty$. Then, for all $j \geq 0$,*

$$a_j(f) = \langle f, \varphi_j \rangle = \frac{1}{k!} \langle f_{k,Y}, (y^k \varphi_j)^{(k)} \rangle \tag{14}$$

Proposition 4.3. *Define $b_\ell^{j,0} = \delta_{\ell,j}$, for $p \geq 0$,*

$$b_\ell^{j,p+1} = -\frac{\ell+1}{2} b_{\ell+1}^{j,p} - (p + \frac{1}{2}) b_\ell^{j,p} + \frac{\ell}{2} b_{\ell-1}^{j,p}, \quad \text{and} \quad h_\ell^{j,k} = \sum_{p=|\ell-j|}^{k} b_\ell^{j,p} \binom{k}{p} \frac{1}{p!}. \tag{15}$$

Define the matrices $H_m^{(k)}$ with size $m \times (m+k)$ by $[H_m^{(k)}]_{j,\ell} = h_\ell^{j,k}$ for $\ell = \sup((j-k), 1), \ldots, j+k$, otherwise $[H_m^{(k)}]_{j,\ell} = 0$. Then,

$$\vec{a}_{m-1}(f) = H_m^{(k)} \vec{a}_{m+k-1}(f_{k,Y}).$$

Moreover, the coefficients $h_{j,k}^\ell$ satisfy

$$\forall \ell \leq j+k, \quad |h_\ell^{j,k}| \leq C'_k (j+k)^k. \tag{16}$$

Example. For instance, for $k = 1$ (multiplicative censoring model for Y_i), $h_\ell^{j,1} = 0$ if $\ell \neq j, j-1, j+1$ and

$$h_{j-1}^{j,1} = -\frac{j}{2}, \quad h_j^{j,1} = \frac{1}{2}, \quad h_{j+1}^{j,1} = \frac{j+1}{2}, \tag{17}$$

meaning that $a_0(f) = (1/2)a_0(f_{1,Y}) + (1/2)a_1(f_{1,Y})$ and for $j \geq 1$,

$$a_j(f) = -\frac{j}{2} a_{j-1}(f_{1,Y}) + \frac{1}{2} a_j(f_{1,Y}) + \frac{j+1}{2} a_{j+1}(f_{1,Y}).$$

For $k = 2$,

$$h_{j-2}^{j,2} = \frac{j(j-1)}{8}, \quad h_{j-1}^{j,2} = -\frac{1}{2} j, \quad h_j^{j,2} = -\frac{j^2+j-1}{4}, \quad h_{j+1}^{j,2} = \frac{1}{2}(j+1), \quad h_{j+2}^{j,2} = \frac{(j+1)(j+2)}{8}.$$

4.2. **Projection estimator and upper risk bound.** Proposition 4.3 leads us to define a collection of projection estimators of f by:

$$\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \vec{\hat{a}}_{m-1} = (\hat{a}_j)_{0 \le j \le m-1} = H_m^{(k)} \vec{\hat{a}}_{m+k-1}(Y), \qquad m \ge 1 \tag{18}$$

where $\vec{\hat{a}}_{m+k-1}(Y) = [(\hat{a}_j(Y))_{0 \le j \le m+k-1}]$ and $\hat{a}_j(Y)$ is defined by

$$\hat{a}_j(Y) := \frac{1}{n} \sum_{i=1}^{n} \varphi_j(Y_i). \tag{19}$$

Note that, by formula (14), we also have the useful formula:

$$\hat{a}_j = \hat{a}_j(k) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{k!} (y^k \varphi_j)^{(k)}(Y_i).$$

For instance, for $k = 1$, we get $\hat{a}_j(1) = n^{-1} \sum_{i=1}^{n} [Y_i \varphi_j'(Y_i) + \varphi_j(Y_i)]$ and for $k = 2$, we obtain $\hat{a}_j(2) = (2n)^{-1} \sum_{i=1}^{n} [Y_i^2 \varphi_j"(Y_i) + 4\varphi_j'(Y_i) + 2\varphi_j(Y_i)]$.

Let $\rho^2(A) = \lambda_{max}(A^t\, A)$ denote the squared spectral radius of the matrix A, i.e. the largest eigenvalue of $A^t\, A$, where A^t denotes the transpose of A. We can prove the following risk bound for the estimator.

Proposition 4.4. *Let $\hat{f}_m$ be given by (18). Then we have*

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \le \|f - f_m\|^2 + 2\frac{(m+k)\rho^2(H_m^{(k)})}{n}.$$

Moreover, there exists a constant ζ_k such that

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \le \|f - f_m\|^2 + \zeta_k \frac{(m+k)^{2k+1}}{n}$$

with $\zeta_k = 2[(2k+1)C_k']^2$ where C_k' is the constant in Proposition 4.2, formula (16).

Remark 3.1 applies to the bound given in Proposition 4.4: the term term is unchanged and still decreasing with m and the variance term is now of order $(m+k)^{2k+1}/n$ which is increasing with m. Therefore, we have to perform a compromise to select relevantly m.

We can deduce from Proposition 4.4 rates of convergence of the estimator on Sobolev-Laguerre spaces described in Section 2.2.

Corollary 4.1. *Assume that $f \in W^s(D)$. Let $\hat{f}_m$ be given by (18). Then choosing $m_{opt} = [n^{s+2k+1}]$ gives*

$$\mathbb{E}(\|\hat{f}_{m_{opt}} - f\|^2) \le C_2(D, s, k) n^{-s/(s+2k+1)}$$

where $C(D, s, k)$ is a constant depending on D, s and k.

Remark 3.2 applies here. For exponential, Gamma or mixed Gamma densities f, the bias is unchanged and exponentially decreasing. Thus, the same choice m_n yields a rate of order $[\log(n)]^{2k+1}/n$.

4.3. **Lower bound in the 1-monotone case.** In this section, we prove that the upper bound obtained in Corollary 4.1 is almost optimal. We consider the case $k=1$, but the step from $k=0$ (case of direct observation of X) to $k=1$ suggests how to get a general result. However, given technicalities of the proof, we decided to remain with $k=1$.

Theorem 4.1. *Assume that s is an integer, $s>1$ and consider the model $Y=XU$, for X and U independent, $U\sim\mathcal{U}([0,1])$ with only Y observed.*
Then for any estimator $\hat{f}_n$ of f the density of X, for any $\epsilon>0$ and for n large enough,

$$\sup_{f\in W^s(D)} \mathbb{E}_f\left[\|\hat{f}_n-f\|^2\right]\gtrsim\tilde{\psi}_n,\quad \tilde{\psi}_n=n^{-(s+\epsilon)/(s+\epsilon+3)}.$$

Remark. We may take $\epsilon=0$ and have additional log terms, but this would add some technicalities again.

Proof of Theorem 4.1. The proof follows the same steps as the proof of Theorem 3.1. First we define proposals $\tilde{f}_0$ and $\tilde{f}_{\boldsymbol{\theta}}$ for the densities of $X_1,\dots,X_n$ and compute the corresponding densities $f_{Y,0}$ and $f_{Y,\boldsymbol{\theta}}$ of $Y_1,\dots,Y_n$. Let us choose $\tilde{f}_0$ such that

$$f_{Y,0}(x)=\int_x^{+\infty}\frac{\tilde{f}_0(u)}{u}du=f_0(x)=\frac{c_{\alpha,\beta}}{(e+x)^\alpha\log^\beta(e+x)}\mathbf{1}_{\mathbb{R}^+}(x),$$

where $\beta=(1+\epsilon)/2$, with $0<\epsilon<1$ and $\alpha>1$. By derivation, we get

$$\tilde{f}_0(x)=-xf'_{Y,0}(x)=c_{\alpha,\beta}\frac{x}{(e+x)^{\alpha+1}\log^{\beta+1}(e+x)}[\alpha\log(e+x)+\beta]\mathbf{1}_{\mathbb{R}^+}(x),$$

Then we can compute by formula (2) for $k=1$, Next, let

$$\tilde{f}_{\boldsymbol{\theta}}(x)=\tilde{f}_0(x)+\delta\sum_{k=K+1}^{2K}\theta_{k-K}x\varphi'_k(x).$$

We have, as $\int\varphi_k(x)dx=\sqrt{2}(-1)^k$ that

$$\int x\varphi'_k(x)dx=[x\varphi_k(x)]_0^{+\infty}-\int_0^{+\infty}\varphi_k(x)dx=\sqrt{2}(-1)^{k+1}.$$

Therefore $\int\tilde{f}_{\boldsymbol{\theta}}(x)dx=1$ under the condition $\sum\theta_k(-1)^k=0$, as previously. Thanks to formula (40), we have

$$x\varphi'_k(x)=-\frac{k}{2}\varphi_{k-1}(x)-\frac{1}{2}\varphi_k(x)+\frac{k+1}{2}\varphi_{k+1}(x)$$

and we can write $\tilde{f}_{\boldsymbol{\theta}}$ as follows in the $(\varphi_k)_k$ basis:

$$\tilde{f}_{\boldsymbol{\theta}}(x)=\tilde{f}_0(x)+\delta\sum_{k=K}^{2K+1}\mu_k(\boldsymbol{\theta})\varphi_k(x)$$

with for $k=K,K+1,\dots,2K+1$,

$$\mu_k(\boldsymbol{\theta})=-\frac{k+1}{2}\theta_{k-K+1}-\frac{\theta_{k-K}}{2}+\frac{k}{2}\theta_{k-K-1}$$

under initial and final conditions $\theta_{-1}=\theta_0=\theta_{K+1}=\theta_{K+2}=0$.

Computing $\int_x^{+\infty}\tilde{f}_{\boldsymbol{\theta}}(u)/u\,du$ yields

$$f_{Y,\boldsymbol{\theta}}=f_{Y,0}(x)+\delta\sum_{k=K+1}^{2K}\theta_{k-K}\varphi_k(x).$$

We stress that by our construction, $f_{Y,0} = f_0$ and $f_{Y,\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}$, so that $\chi^2(f_{Y,\boldsymbol{\theta}}, f_{Y,0}) = \chi^2(f_{\boldsymbol{\theta}}, f_0)$ is already computed in the previous section (proof of Theorem 3.1).

Lemma 4.1. *Let s integer, $s > 1$. Then $\tilde{f}_0$ and $\tilde{f}_{\boldsymbol{\theta}}$ belong to $W^s(D)$, provided that $\alpha \geq (s+1)/2 \geq 1$ and $\delta^2 K^{s+3} \leq D/C$ for some constant $C = C(s) > 0$.*

Next, we have to see under which condition $\tilde{f}_{\boldsymbol{\theta}} \geq 0$.

Lemma 4.2. *Suppose that $\sum_{k=1}^{K} \theta_k(-1)^k = 0$ and all partial sums $\sum_{k=1}^{p} \theta_k(-1)^k$, $p = 1, \ldots, K$, are uniformly bounded by 1, then under the choice $\delta = \delta' K^{-(\alpha+1)} \log^{-\beta}$ for small enough constant $\delta' > 0$ not depending on K, we have that $\tilde{f}_{\boldsymbol{\theta}}$ is a probability density on $\mathbb{R}_+$.*

Next, we have

$$\|\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_{\boldsymbol{\theta}'}\|^2 = \delta^2 \sum_{k=K}^{2K+1} (\mu_k(\boldsymbol{\theta}) - \mu_k(\boldsymbol{\theta}'))^2 \tag{20}$$

Write that for $k = K, K+1, \ldots, 2K+1$, we have

$$\mu_k(\boldsymbol{\theta}) = -\frac{k}{2}(\theta_{k-K+1} - \theta_{k-K-1}) - \frac{\theta_{k-K} + \theta_{k-K+1}}{2}.$$

We notice that for $j = 0, 1, \ldots, K+1$, we have

$$|\mu_{K+j}(\boldsymbol{\theta}) - \mu_{K+j}(\boldsymbol{\theta}')| \geq [\frac{K+j}{2} - 1] \text{ if } \theta_{j+1} - \theta_{j-1} \neq \theta'_{j+1} - \theta'_{j-1}$$

since $|\theta_j - \theta'_j + \theta_{j+1} - \theta'_{j+1}|/2 \leq 1$. Therefore, we get

$$\sum_{k=K}^{2K+1} (\mu_k(\boldsymbol{\theta}) - \mu_k(\boldsymbol{\theta}'))^2 \geq \delta^2 (K/2 - 1)^2 \rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}'), \tag{21}$$

where

$$\rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sum_{k=K+1}^{2K} \mathbf{1}_{\theta_{k+1}-\theta_{k-1} \neq \theta'_{k+1}-\theta'_{k-1}}.$$

Therefore, we need to check that $\rho_1(.,.)$ is a distance and that the Varshamov-Gilbert Lemma holds with the Hamming distance replaced by $\rho_1(.,.)$.

Lemma 4.3. *Fix some even natural number $K > 0$. There exists a subset $\{\boldsymbol{\theta}^{(0)}, \ldots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ and a constant $A_1 > 0$, such that $\boldsymbol{\theta}^{(0)} = (0, \ldots, 0)$, all partial sums $\sum_{j=1}^{k} \theta_j^{(m)}(-1)^j$, $k = 1, \ldots, K$, are uniformly bounded by 1,*

$$\sum_{k=1}^{K} \theta_k^{(m)}(-1)^k = 0 \quad \text{and} \quad \rho_1(\boldsymbol{\theta}^{(m)}, \boldsymbol{\theta}^{(l)}) \geq \tilde{A}_1 K,$$

for all $0 \leq m < l \leq M$. Moreover it holds that

$$M \geq 2^{\tilde{A}_2 K} \tag{22}$$

for some constant $\tilde{A}_2 > 0$.

Next we prove

Lemma 4.4.

$$\frac{1}{M} \sum_{j=1}^{M} \chi^2\left((f_{Y,\boldsymbol{\theta}^{(j)}})^{\otimes n}, (f_{Y,0})^{\otimes n}\right) \lesssim n\delta^2 K^{\alpha+4} \text{ and for } 0 \leq j \neq l \leq M,\ \|\tilde{f}_{\boldsymbol{\theta}^{(j)}} - \tilde{f}_{\boldsymbol{\theta}^{(l)}}\|^2 \gtrsim \delta^2 K^3.$$

Now we end the proof of Theorem 4.1. We choose $\alpha = (s{+}1)/2$, $\delta^2 = (\delta')^2 K^{-2(\alpha+1)} \log^{-(1+\epsilon)}(K)$, $K = [n/\log^{1+\epsilon}(n)]^{1/[2(\alpha+1)]}$ and we obtain

$$\frac{1}{M}\sum_{j=1}^{M} \chi^2\left((f_{Y,\boldsymbol{\theta}^{(j)}})^{\otimes n}, (f_{Y,0})^{\otimes n}\right) \lesssim \log^{\alpha+4}(M).$$

and

$$\|\tilde{f}_{\boldsymbol{\theta}^{(j)}} - \tilde{f}_{\boldsymbol{\theta}^{(l)}}\|^2 \gtrsim n^{-s/(s+3)}[\log(n)]^{(1+\epsilon)/(1+s/3)}.$$

Note that $\delta^2 K^{s+3} = [\log(n)]^{-(1+\epsilon)}$ is bounded (constraint from Lemma 4.1). This ends the proof of Theorem 4.1. □

5. Projection estimator of f in the Laguerre basis when Z_i's are observed and adaptive procedure

5.1. **Projection estimator and risk bound.** We use the following result proved in Mabon (2015). Define the $m \times m$ triangular matrix $\mathbf{V}_m = (v_{i,j})_{0\le i,j\le m-1}$ where

$$v_{i,j} = 2^{-1/2}(\langle f_V, \varphi_{i-j}\rangle \mathbf{1}_{i-j\ge 0} - \langle f_V, \varphi_{i-j-1}\rangle \mathbf{1}_{i-j-1\ge 0}), \tag{23}$$

and the diagonal elements are $v_{i,i} = 2^{-1/2}\langle f_V, \varphi_0\rangle > 0$. We have for all $m \ge 1$,

$$\vec{a}_{m-1}(f_Y) = (a_j(f_Y))_{0\le j\le m-1} = \mathbf{V}_m^{-1}[(a_j(f_Z))_{0\le j\le m-1}] = \mathbf{V}_m^{-1}\vec{a}_{m-1}(f_Z), \tag{24}$$

If $V_i = 0$, $\mathbf{V}_m = I_m$. Formula (23) relies on a specific property of the Laguerre basis (φ_j) which can be used in $\mathbb{R}^+$-deconvolution setting:

$$\int_0^x \varphi_k(u)\varphi_j(x-u)du = \frac{1}{\sqrt{2}}\left(\varphi_{j+k}(x) - \varphi_{j+k+1}(x)\right), \tag{25}$$

(see 22.13.14 in Abramowitz and Stegun (1964)).

The following result is the basement of our estimation procedure.

Proposition 5.1. *Define the matrices $K_m^{(k)}$ with size $m \times (m+k)$ by*

$$K_m^{(k)} := H_m^{(k)}\mathbf{V}_{m+k}^{-1}.$$

Then,

$$\vec{a}_{m-1}(f) = K_m^{(k)}\vec{a}_{m+k-1}(f_Z).$$

Proposition 5.1 leads us to define a collection of projection estimators estimator of f by:

$$\tilde{f}_m = \sum_{j=0}^{m-1} \tilde{a}_j\varphi_j, \quad \vec{\tilde{a}}_{m-1} = (\tilde{a}_j)_{0\le j\le m-1} = K_m^{(k)}\vec{\tilde{a}}_{m+k-1}(Z), \qquad m \ge 1 \tag{26}$$

where $\vec{\tilde{a}}_{m+k-1}(Z) = [(\tilde{a}_j(Z))_{0\le j\le m+k-1}]$ and $\tilde{a}_j(Z)$ is defined by

$$\tilde{a}_j(Z) := \frac{1}{n}\sum_{i=1}^{n} \varphi_j(Z_i). \tag{27}$$

We can prove the following risk bound for the estimator.

Proposition 5.2. *Let $\tilde{f}_m$ be given by (26). Then we have*

$$\mathbb{E}(\|\tilde{f}_m - f\|^2) \le \|f - f_m\|^2 + 2\frac{(m+k)\rho^2(K_m^{(k)})}{n}.$$

Remark 3.1 still applies here. The bias term is unchanged. The variance term has now order $(m+k)\rho^2(K_m^{(k)}) \leq (m+k)^{2k+1}\rho^2(\mathbf{V}_{m+k}^{-1})/n$, and is increasing with m. Indeed, it is proved that $m \mapsto \rho^2(\mathbf{V}_{m+k}^{-1})$ is increasing (see Mabon (2015)). Therefore, we have to perform a compromise to select relevantly m.

5.2. **Asymptotic rate of convergence of the estimator.** We can deduce from Proposition 5.2 rates of convergence of the estimator on Sobolev-Laguerre spaces.

Let us define $r \geq 1$ as an integer such that

$$\frac{d^j f_V(t)}{dt^j}\bigg|_{t=0} = \begin{cases} 0, & \text{if } j = 0, ..., r-2, \\ B_r \neq 0, & \text{if } j = r-1. \end{cases} \tag{28}$$

with $r = 1$ if $f_V(0) = B_1 \neq 0$.

Consider the two assumptions

- **(A1)** f_V is r times differentiable with $f_V^{(r)} \in \mathbb{L}_1[0, \infty)$.
- **(A2)** Laplace transform $Lf_V(s)$ of f_V has no zeros with nonnegative real parts except for zeros of the form $s = \infty + ib$.

It is proved in Comte et *al.* that, under **(A1)-(A2)**, we have $\rho^2(\mathbf{V}_m^{-1}) \asymp Cm^{2r}$.

Therefore the following Corollary holds

Corollary 5.1. *Assume that $f \in W^s(D)$, and* **(A1)-(A2)** *are fulfilled. Let $\tilde{f}_m$ be given by (26). Then choosing $m_{opt} = [n^{2r+s+2k+1}]$ gives*

$$\mathbb{E}(\|\tilde{f}_{m_{opt}} - f\|^2) \leq C(K)n^{-s/(2r+s+2k+1)}.$$

Remark 5.1. If V follows an exponential distribution, then it satisfies assumptions **(A1)-(A2)** with $r = 1$ and more generally, a Gamma(p, θ) density satisfies assumptions **(A1)-(A2)** with $r = p$.

5.3. **Adaptive estimation.** We can propose a model selection method to select m automatically, aiming at an automatic bias variance compromise. We define

$$\tilde{m} = \arg\min_{m \in \mathcal{M}_n} \left(-\|\tilde{f}_m\|^2 + \kappa \frac{m\rho^2(K_m^{(k)})}{n} \right) \tag{29}$$

where

$$\mathcal{M}_n = \{m \in \mathbb{N}^*,\ m\rho^2(K_m^{(k)}) \leq n\}.$$

If $V = 0$, the procedure and the result are valid with matrix $K_m^{(k)}$ replaced by $H_m^{(k)}$ (i.e. $\mathbf{V}_m^{(k)} = \mathbf{Id}_{m+k}$, the identity matrix with size $(m+k) \times (m+k)$. If $V = 0$ and $U = 1$ (direct observation of X), then the procedure and result is valid with $k = 0$ and $H_m^{(0)} == \mathbf{Id}_m$.

Theorem 5.1. *Let $\tilde{f}_m$ be given by (26) and $\tilde{m}$ by (29) .There exists a constant κ_0 such that for any $\kappa \geq \kappa_0$, we have*

$$\mathbb{E}(\|\tilde{f}_{\tilde{m}} - f\|^2) \leq C_1 \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \text{pen}(m)\right) + \frac{C_2}{n}.$$

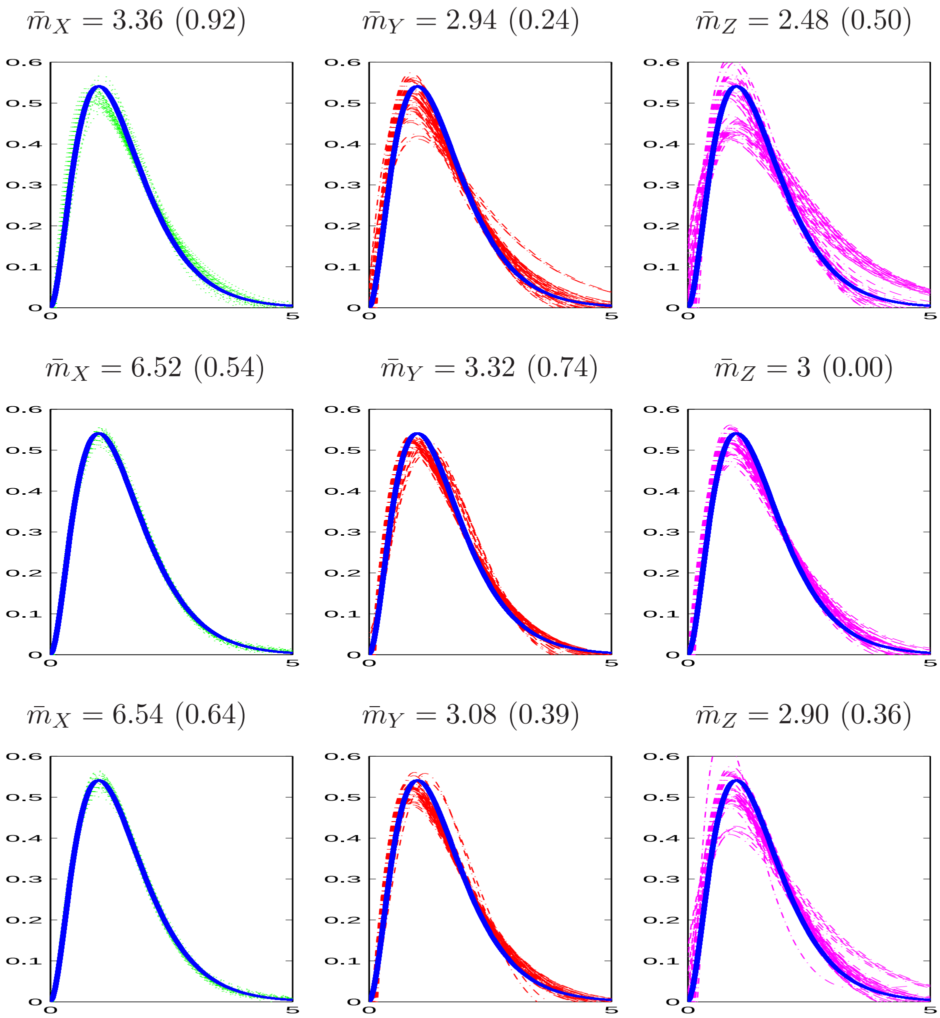

FIGURE 1. True density f of Model (i) (Gamma distribution) in bold (blue). Left: 50 estimators of f from direct observation of X in dotted (green). Middle: 50 estimators of f from observation of Y, in dotted (red). Right: 50 estimators of f from observation of Z, in dotted (magenta). First line: $n = 400$ and $U \sim \mathcal{U}([0,1])$. Second line: $n = 2000$ and $U \sim \mathcal{U}([0,1])$. Third line: $n = 2000$ and $U \sim \beta(1,2)$. Above each plot, $\bar{m}_X$ (resp. $\bar{m}_Y$, resp $\bar{m}_Z$) is the mean of the selected dimensions for X (resp. for Y, resp. for Z) with standard deviation in parenthesis.

6. Simulation results

We implement the adaptive estimators $\tilde{f}_{\tilde{m}}$ of f based

- on direct observations $X_1, \dots, X_n$,
- on multiplicative censored observations for $k = 1, 2, \quad Y_1, \dots, Y_n$,
- on $Z_1, \dots, Z_n$ from the complete model (1) with $V \sim \mathcal{E}(\lambda)$, an exponential variable with parameter λ (with mean $1/\lambda$, with $\lambda = 2$ in the simulations).

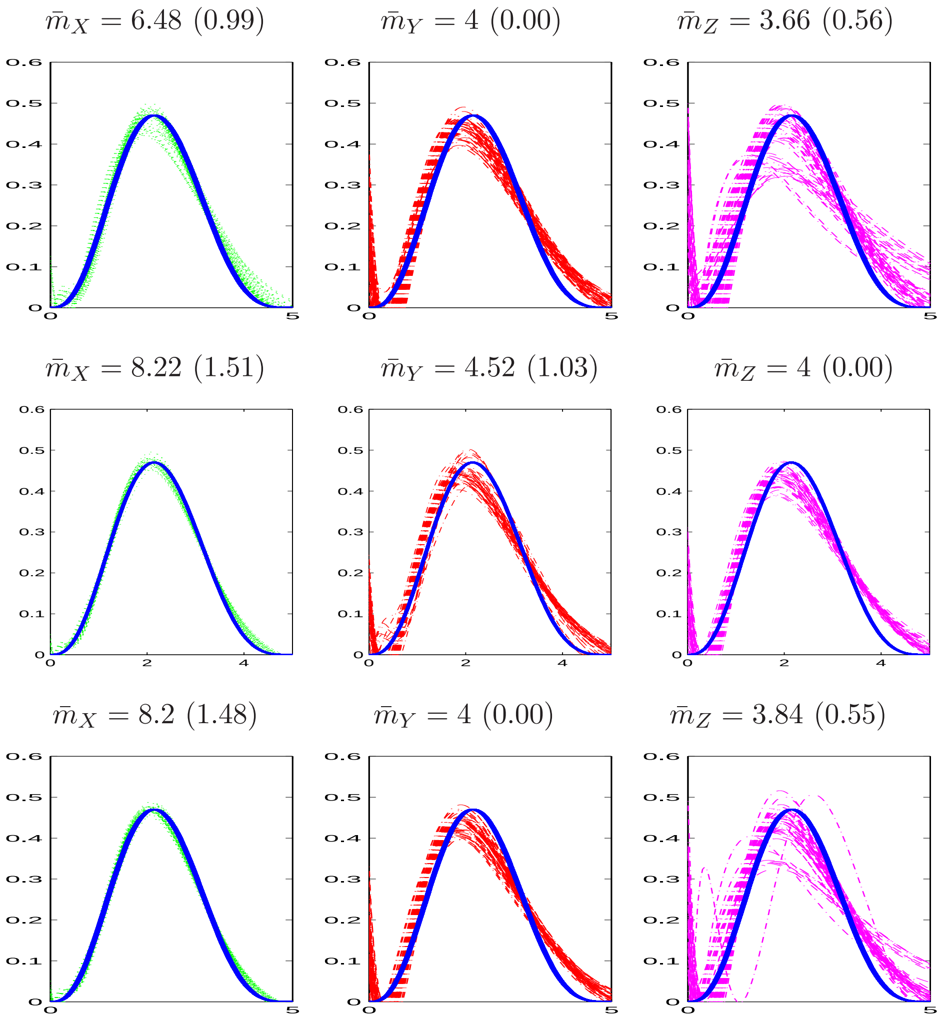

FIGURE 2. True density f of Model (ii) (Beta distribution) in bold (blue). Left: 50 estimators of f from direct observation of X in dotted (green). Middle: 50 estimators of f from observation of Y, in dotted (red). Right: 50 estimators of f from observation of Z, in dotted (magenta). First line: $n = 400$ and $U \sim \mathcal{U}([0,1])$. Second line: $n = 2000$ and $U \sim \mathcal{U}([0,1])$. Third line: $n = 2000$ and $U \sim \beta(1,2)$. Above each plot, $\bar{m}_X$ (resp. $\bar{m}_Y$, resp $\bar{m}_Z$) is the mean of the selected dimensions for X (resp. for Y, resp. for Z) with standard deviation in parenthesis.

For $V \sim \mathcal{E}(\lambda)$, we have $[\mathbf{V}_m]_{i,i} = \lambda/(1+\lambda)$ and

$$[\mathbf{V}_m]_{i,j} = -2\lambda \frac{(\lambda-1)^{i-j-1}}{(\lambda+1)^{(i-j+1)}} \text{ if } j < i$$

and $[\mathbf{V}_m]_{i,j} = 0$ otherwise.

We consider for f the densities

(i) Gamma$(3, 1/2)$,
(ii) 5 Beta$(4, 5)$,
(iii) a mixture of Gamma: c(0.4 Gamma(2,1/2)+0.6 Gamma(16,1/4)) with $c = 5/8$.

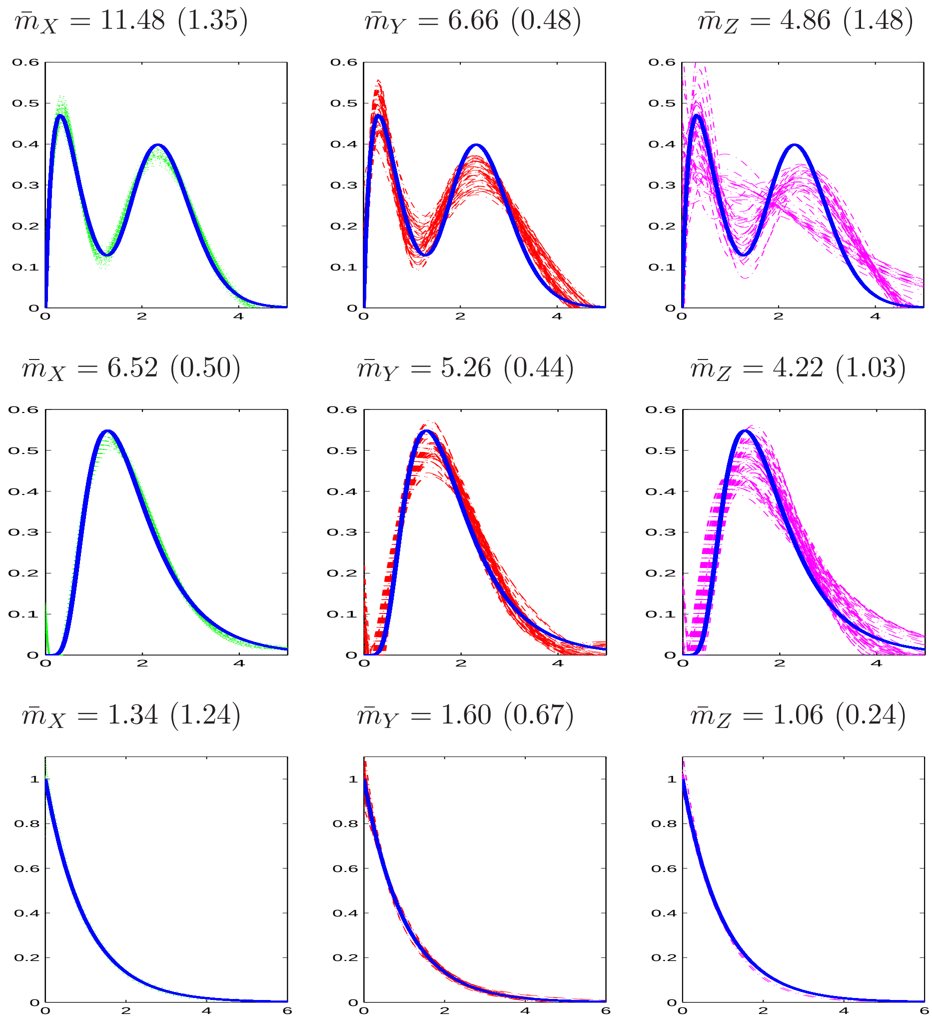

FIGURE 3. True density f in bold (blue) of Model (iii) (Mixed Gamma distribution) (first line), Model (iv) (Lognormal distribution) (second line) and $f \rightsquigarrow \mathcal{E}(1)$ (third line), with $n = 2000$. U uniform for lines 1 and 2, $U \rightsquigarrow \beta(1,2)$ for line 3. Left: 50 estimators of f from direct observation of X in dotted (green). Middle: 50 estimators of f from observation of Y, in dotted (red). Right: 50 estimators of f from observation of Z, in dotted (magenta). Above each plot, $\bar{m}_X$ (resp. $\bar{m}_Y$, resp $\bar{m}_Z$) is the mean of the selected dimensions for X (resp. for Y, resp. for Z) with standard deviation in parenthesis.

(iv) Lognormal$(0.5, 0.5)$ (exponential of a Gaussian with mean 0.5 and variance 0.5^2).

All factors and parameters are chosen to have the true densities with the same scales.

Direct estimation is penalized with $\kappa_1 = 1$ in all cases. For U following a uniform distribution on $[0,1]$, without additive noise, we use $\kappa_2 = 0.25$ and with additive noise, we use $\kappa_3 = 0.25$. For U following a $\beta(1,2)$ distribution on $[0,1]$, we take $\kappa_2 = 0.125$ without additive noise and $\kappa_3 = 0.25$ with additive noise.

Beam of estimators are given in Figures 1-3 and show clearly the performance of the method via variability bands. The Laguerre basis provides excellent estimation when using direct data, and the problem gets more difficult in presence of censoring. Increasing the order k (we took

$k = 1$ for $U \sim \mathcal{U}([0,1])$ and $k = 2$ for $U \sim \beta(1,2)$) makes the problem more difficult. Adding a nuisance process V creates also additional difficulty, this is why mixture of Gamma's can be correctly reconstructed only with multiplicative censoring, for instance (see Figure 3). Selected dimensions can be of various orders (between 2 and 12 in our examples) and vary or be very stable (see the null standard deviations).

7. Proofs

7.1. **Proof of lemma 2.1.** Obviously $|h|_0 = |||h|||_0$ and $\|h\|_j^2 \le |h|_j^2$ for all j. Moreover $j \mapsto |h|_j$ is increasing. Therefore $|||h|||_s^2 \le (s+1)|h|_s^2$. On the other hand, let $b_{j,s}$ the coefficients such that

$$X^s = \sum_{j=1}^{s} b_{j,s} X(X-1)\dots(X-j+1).$$

Then

$$|h|_s^2 = \sum_{j=1}^{s} b_{j,s}\|h\|_j^2 \le A(s)|||h|||_s^2,$$

with

$$A(s) = \max(|b_{j,s}|, j = 1, \dots, s). \quad \square \tag{30}$$

7.2. **Proofs of Lemmas of Section 3.2 (Theorem 3.1).**

Proof of Lemma 3.1. Indeed we have

$$\|f_0\|_s^2 = \int_0^{+\infty} \left(x^{s/2} \sum_{j=0}^{s} \binom{s}{j} f_0^{(j)}(x) \right)^2 dx \le 2^s \sum_{j=0}^{s} \binom{s}{j} \int_0^{+\infty} \left(x^{s/2} f_0^{(j)}(x) \right)^2 dx.$$

The "worst" term in the above sum is $x^{s/2}(e+x)^{-\alpha}\log^{-\beta}(e+x)$. Thus, as $\alpha \ge (s+1)/2$ and $\beta > 1/2$,

$$x^{s/2} f_0^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$$

for $j = 0, \dots, s$ and there exists a constant $B(s,\alpha)$ such tat

$$\|f_0\|_s^2 \le B(s,\alpha).$$

It follows that

$$|f_0|_s^2 \le \tilde{B}(s,\alpha), \quad \tilde{B}(s,\alpha) := (s+1)B(s,\alpha)A(s)$$

where $A(s)$ is defined by (30). We take $D/4 \ge \tilde{B}(s,\alpha)$. Next

$$|f_{\boldsymbol{\theta}}|_s \le |f_0|_s + \delta \left| \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k \right|_s.$$

Let us define for $f, g \in W^s$, $\langle f, g\rangle_s = (1/2)(|f+g|_s^2 - |f|_s^2 - |g|_s^2)$ so that $|\varphi_k|_s^2 = k^s$ and $\langle \varphi_k, \varphi_\ell\rangle_s = 0$ for $k \ne \ell$. Therefore

$$\begin{aligned}
\left| \sum_{k=1+K}^{2K} \theta_{k-K}\varphi_k \right|_s^2 &= \sum_{k=1+K}^{2K} k^s \theta_{k-K}^2 \le \sum_{k=1+K}^{2K} k^s \\
&\le \sum_{k=1+K}^{2K} \int_k^{k+1} x^s\, ds = \frac{(2K+1)^{s+1} - (1+K)^{s+1}}{s+1},
\end{aligned}$$

and

$$|f_{\boldsymbol{\theta}}|_s^2 \leq 2|f_0|_s^2 + C\delta^2 K^{s+1}/(s+1)$$

for some constant $C = C(s) > 0$. Hence $|f_{\boldsymbol{\theta}}|_s^2 \leq D$ if $\delta^2 K^{s+1}/(s+1) \leq D/(2C)$. □

Proof of Lemma 3.2. First, noting that $\int \varphi_k(x)dx = \sqrt{2}(-1)^k$, we have

$$\begin{aligned}\int_0^\infty f_{\boldsymbol{\theta}}(x)\,dx &= 1 + \delta \sum_{k=1+K}^{2K} \theta_{k-K} \int_0^\infty \varphi_k(x)\,dx \\ &= 1 + \sqrt{2}\delta \sum_{k=1+K}^{2K} \theta_{k-K}(-1)^k = 1,\end{aligned}$$

so that our conditions ensure that $\int_0^\infty f_{\boldsymbol{\theta}}(x)\,dx = 1$.

Next we prove that $f_{\boldsymbol{\theta}}$ is nonnegative, which is surprisingly tricky. We have

$$f_{\boldsymbol{\theta}}(x)/f_0(x) = 1 + \delta \frac{(e+x)^\alpha \log^\beta(e+x)}{c_{\alpha,\beta}} \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x).$$

For any fixed $\mathtt{a} > 0$, for any $x \in [0, \mathtt{a}]$, we have $|f_{\boldsymbol{\theta}}(x)/f_0(x) - 1| \leq \delta K\sqrt{2}(e+\mathtt{a})^\alpha \log^\beta(e+\mathtt{a})/c_{\alpha,\beta} \lesssim \delta K = \delta' K^{1-\alpha}\log^{-\beta}(K)$ which is small as $\alpha \geq (s+1)/2 > 1$. Without loss of generality, we assume that $\mathtt{a} > 1$.

Thus, in order to prove that $f_{\boldsymbol{\theta}}$ is a nonnegative function, it is enough to show that

(31) $$\sup_{x>\mathtt{a}} \left| x^\lambda \log^\mu(x) \cdot \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \lesssim K^\lambda \log^\mu(K), \quad K \to \infty$$

for any fixed $\lambda > 0$, $\mu > 0$ and for sufficiently large $\mathtt{a} > 0$. Then by taking $\lambda = \alpha$, $\mu = \beta$ and $\delta = \delta' K^{-\alpha}\log^{-\beta}(K)$ for small enough constant $\delta' > 0$ not depending on K, we get $f_{\boldsymbol{\theta}}(x) \geq 0$, $x \in \mathbb{R}_+$.

We proceed in two steps for the proof of (31). First we study the supremum for large values of x, $2x \geq c\nu$, $c > 0$ and then for intermediate values of x ($2\mathtt{a} < 2x \leq b\nu$ with $b < 1$ and $\nu = 4K + 2$).

Step 1. Suppose that the sequence $\theta = (\theta_1, \ldots, \theta_K) \in \{0,1\}^K$ satisfies

$$\left| \sum_{k=1}^m \theta_k(-1)^k \right| \leq A$$

for all $m = 1, \ldots, K$, and some constant $A > 0$. Fix some real numbers λ, μ with $0 < \lambda < K$, and $\mu > 0$, then it holds for any $2x > 4K + 2\lambda + 1$,

(32) $$\left| x^\lambda \log^\mu(x) \cdot \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \leq AC_{\lambda,\mu} K^\lambda \log^\mu(K), \quad K \to \infty,$$

where $\varphi_k(x) = \sqrt{2}e^{-x}L_k(2x)$ and the constant $C_{\lambda,\mu}$ depends only on λ, μ.

To prove (32), we first study the case $\mu = 0$ and λ integer.

Lemma 7.1. *It holds for any integers n and $\lambda \leq n$,*

(33) $$x^\lambda L_n(x) = \sum_{k=-\lambda}^{\lambda} c_{k,n}^{(\lambda)} L_{n+k}(x),$$

where the coefficients $c_{k,n}^{(\lambda)}$ *can be computed via the relation*

$$c_{k,n}^{(\lambda)} = c_{k,n}^{(\lambda-1)}(2(n+k)+1) - c_{k-1,n}^{(\lambda-1)}(n+k) - c_{k+1,n}^{(\lambda-1)}(n+k+1)$$

for $|k| < \lambda$ *with* $c_{k,n}^{(0)} = \delta_{0,k}$ *and*

$$c_{\lambda,n}^{(\lambda)} = -c_{\lambda-1,n}^{\lambda-1}(n+\lambda), \quad c_{-\lambda,n}^{(\lambda)} = -c_{-\lambda+1,n}^{\lambda-1}(n-\lambda+1).$$

Proof. For $\lambda = 0$ the relation (33) obviously holds. Suppose that it holds for some $\lambda = K$, then due to a well known formula

$$xL_n(x) = (2n+1)L_n(x) - (n+1)L_{n+1} - nL_{n-1}(x),$$

we have

$$\begin{aligned}
x^{K+1}L_n(x) &= \sum_{k=-K}^{K} c_{k,n}^{(K)} xL_{n+k}(x) \\
&= \sum_{k=-K}^{K} c_{k,n}^{(K)} \left[(2(n+k)+1)L_{n+k}(x) - (n+k+1)L_{n+k+1} - (n+k)L_{n+k-1}(x)\right] \\
&= \sum_{k=-K}^{K} c_{k,n}^{(K)}(2(n+k)+1)L_{n+k}(x) - \sum_{k=-K+1}^{K+1} c_{k-1,n}^{(K)}(n+k)L_{n+k} \\
&\quad - \sum_{k=-K-1}^{K-1} c_{k+1,n}^{(K)}(n+k+1)L_{n+k}(x) \\
&= \sum_{k=-K-1}^{K+1} c_{k,n}^{(K+1)} L_{n+k}(x).
\end{aligned}$$

This ends the proof of Lemma 7.1. □

We deduce by induction from Lemma 7.1 the following Corollary.

Corollary 7.1. *Each coefficient* $c_{k,n}^{(\lambda)}$ *in (33) can be represented in the form*

(34) $$c_{k,n}^{(\lambda)} = \sum_{r=(r_1,\ldots,r_\lambda),\, r_i \in S_{\lambda,1/2}^{\lambda}} b_{k,r}^{(\lambda)} \prod_{i=1}^{\lambda}(n+r_i)$$

with $n \geq \lambda$, $S_{\lambda,1/2} = \{-\lambda,\ldots,\lambda\} \cup \{-\lambda+1/2,\ldots,\lambda+1/2\}$ *and some coefficients* $b_{k,r}^{(\lambda)}$ *not depending on* n.

The following property is given e.g. in Muckenhoupt (1970).

Lemma 7.2. *Set* $\nu = 4N = 4n+2$, $t = x/\nu$, *then it holds for all* $x \geq d\nu$ *for any* $d > 0$

$$e^{-x/2}L_n(x) = (-1)^n \frac{N^{N+1/6}e^{-N}}{n!(-x\phi'(t))^{1/2}} \left[Ai(-\nu^{2/3}\phi(t)) + O\left(\frac{Ai(-\nu^{2/3}\phi(t))}{x}\right)\right],$$

where

$$\phi(t) = -\left[3\gamma(t)/2\right]^{2/3}, \quad \gamma(t) = \frac{1}{2}(t^2-t)^{1/2} - \frac{1}{2}\cosh^{-1}(t^{1/2})$$

and $Ai(t)$ *is the Airy function (see Abramowitz and Stegun (1964)).*

Corollary 7.2. *Under conditions of the previous lemma, we have a representation*

$$e^{-x/2}L_n(x) = (-1)^n a_n(x), \tag{35}$$

where for any $x > c\nu$ with $c > 1$ the sequence a_n is bounded (uniformly in x), positive and increasing in n for $4n+2 \leq x$.

Proof. The function $Ai\big(-\nu^{2/3}\phi(x/\nu)\big)/(-x\phi'(x/\nu))^{1/2}$ is monotone increasing in ν for any $x \geq \nu = 4n+2$. Moreover, the function $N^{N+1/6}e^{-N}/n!$ is monotone increasing in n. The uniform boundedness of $a_n(x)$ follows from the boundedness of $\left|e^{-x/2}L_n(x)\right|$. □

Proof of Step 1. First we prove (32) for $\mu = 0$ and λ integer. From (33), (34) and (35), we have

$$x^\lambda \sum_{k=K+1}^{2K} \theta_{k-K} e^{-x/2} L_k(x) = \sum_{r=(r_1,\dots,r_\lambda), r_i \in S^\lambda_{\lambda,1/2}} \sum_{\ell=-\lambda}^{\lambda} (-1)^\ell b^{(\lambda)}_{\ell,r} \Sigma_K(\ell, r)$$

with

$$\Sigma_K(\ell, r) = \sum_{k=K+1}^{2K} \theta_{k-K}(-1)^k a_{\ell+k}(x) \rho_k^{(\lambda)}(r), \quad \rho_k^{(\lambda)}(r) = \prod_{i=1}^{\lambda}(k + r_i).$$

Note that $k \mapsto \rho_k^{(\lambda)}(r) a_{\ell+k}(x)$ is nonnegative and nondecreasing and $a_{\ell+k}(x)$ is bounded. Inequality (32) for $\mu = 0$ and λ an integer follows from the next Lemma.

Lemma 7.3. *Let $K_1 < K_2$ be two natural numbers and let ρ_n be an increasing sequence of nonnegative numbers, then for any $x > 4K_2 + 2$, we have*

$$\left| \sum_{n=K_1+1}^{K_2} e^{-x/2} \theta_n \rho_n L_n(x) \right| \leq \rho_{K_2} a_{K_2}(x) \max_{K_1+1 \leq n \leq K_2} \left| \sum_{n=K_1+1}^{n} \theta_n (-1)^n \right|.$$

Proof. Due to the Abel summation theorem, we get

$$\begin{aligned}
\sum_{n=K_1+1}^{K_2} e^{-x/2} \theta_n \rho_n L_n(x) &= \sum_{n=K_1+1}^{K_2} \theta_n \rho_n (-1)^n a_n(x) \\
&= S_{K_2} \rho_{K_2} a_{K_2}(x) + \sum_{n=K_1+1}^{K_2-1} S_n(\rho_{n+1} a_{n+1}(x) - \rho_n a_n(x)),
\end{aligned}$$

where $S_n \doteq \sum_{j=K_1+1}^{n} (-1)^j \theta_j$ for $n > K_1$. Since the sequence $\rho_n a_n(x)$ is non-decreasing and non-negative, we get the desired estimate. □

Now consider the case of λ a real number and write that $\lambda = [\lambda] + \{\lambda\}$ where $\{\lambda\}$ is the fractional part of λ and belongs to $(0,1)$. For any $2x > 4K + 2[\lambda] + 3$,

$$\begin{aligned}
\left| x^\lambda \sum_{k=K+1}^{2K} \theta_{k-K} \varphi_k(x) \right| &= |x^{\{\lambda\}-1}| \left| x^{[\lambda]+1} \sum_{k=K+1}^{2K} \theta_{k-K} \varphi_k(x) \right| \\
&\leq (4K + 2\lambda + 3)^{\{\lambda\}-1} A C_{[\lambda]+1} K^{[\lambda]+1},
\end{aligned}$$

and the result follows.

Now we study the case $\mu > 0$ and we want to prove that, for $2x > 4K + 2[\lambda] + 3$,

$$\left| x^\lambda \log^\mu(x) \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \lesssim \log^\mu(K) K^\lambda.$$

If λ is an integer, we write

$$\left| [x^{-1}\log^\mu(x)] x^{\lambda+1} \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \lesssim \frac{\log^\mu(K)}{K} K^{\lambda+1} = \log^\mu(K) K^\lambda,$$

since $x \mapsto \log^\mu(x)/x$ is decreasing for x large enough ($x > e^\mu$).

If λ is not an integer,

$$\left| [x^{\{\lambda\}-1}\log^\mu(x)] x^{[\lambda]+1} \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \lesssim \frac{\log^\mu(K)}{K^{1-\{\lambda\}}} K^{[\lambda]+1} = \log^\mu(K) K^\lambda,$$

since for any $\omega > 0$, $x \mapsto \log^\mu(x)/x^\omega$ is decreasing for x large enough ($x > e^{\mu/\omega}$). □

Step 2. Now we want to prove (32) for $x \le b\nu$, $b < 1$, $\nu = 4K + 2$. It holds (see Muckenhoupt (1970) p.288)

$$e^{-x/2} L_n(x) \asymp \left[\frac{1}{2} \frac{\psi(x/\nu)}{\frac{x}{\nu}\psi'(x/\nu)} \right]^{1/2} \left[J_0(\nu\psi(x/\nu)) + O\left(\frac{x^{1/2}}{\nu^{3/2}} \tilde{J}_0(\nu\psi(\frac{x}{\nu})\right) \right]$$

for $x \le b\nu$ for some $b < 1$ and $\nu = 4n + 2$, where

$$\psi(t) = \frac{1}{2}(t - t^2)^{1/2} + \frac{1}{2}\arcsin(\sqrt{t}),$$

J_0 is the Bessel function and $\tilde{J}_0(u) = \mathbf{1}_{]0,1]}(u) + u^{-1/2}\mathbf{1}_{u>1}$. Since

$$\frac{\psi(t)}{t\psi'(t)} = 2 + \frac{2}{3}t + O(t^{3/2}), \quad t \to 0$$

and

$$\psi(t) = \sqrt{t} + O\left(t^{3/2}\right), \quad t \to 0,$$

it follows from the asymptotic behavior of the Bessel function J_0

$$\begin{aligned} e^{-x/2} L_n(x) &= J_0\left(\sqrt{x\nu}\right)(1 + o(1)) \\ &= \sqrt{\frac{2}{\pi}} (x\nu)^{-1/4} \cos(\frac{\pi}{4} - \sqrt{x\nu}) - \frac{1}{4}\sqrt{\frac{1}{2\pi}} (x\nu)^{-3/4} \sin(\frac{\pi}{4} - \sqrt{x\nu}) + O((x\nu)^{-5/4}), \end{aligned}$$

provided $x\nu$ is large. Suppose that $x > 1$ and $\lambda \ge 1$, then

$$x^\lambda \sum_{n=K+1}^{2K} e^{-x/2}\theta_n L_n(x) = \sum_{n=K+1}^{2K} \theta_n x^\lambda \frac{\cos\left[\frac{\pi}{4} - \sqrt{x(4n+2)}\right]}{(x(4n+2))^{1/4}} + R_n(x).$$

Since

$$\sum_{n=K+1}^{2K} \frac{1}{(4n+2)^{3/4}} \lesssim \int_K^{2K} \frac{1}{(1+s)^{3/4}}\, ds \lesssim K^{1/4},$$

we have $|R_n(x)| \lesssim x^{\lambda-3/4}K^{1/4} \leq K^\lambda$ for $x \leq K$. So we need to investigate the series

$$S_K(x) \doteq \sum_{n=K+1}^{2K} \theta_n x^\lambda \frac{\cos\left[\frac{\pi}{4} - \sqrt{x(4n+2)}\right]}{(x(4n+2))^{1/4}}.$$

It is clear that we can restrict our attention to the case $x > K^{\frac{\lambda-3/4}{\lambda-1/4}}$, because if $x^{\lambda-1/4} \leq K^{\lambda-3/4}$, we have

$$|S_K(x)| \leq x^{\lambda-1/4} \sum_{n=K+1}^{2K} \frac{1}{(4n+2)^{1/4}} \lesssim x^{\lambda-1/4}K^{3/4} \leq K^\lambda.$$

Since

$$\cos[\frac{\pi}{4} - \sqrt{x(4n+2)}] = 2^{-1/2}(\cos[\sqrt{x(4n+2)}] + \sin[\sqrt{x(4n+2)}]),$$

it is enough to study the asymptotic behavior of the series

$$\Sigma_{1,K}(x) = \sum_{n=K+1}^{2K} \theta_n \frac{\cos[\sqrt{xn}]}{n^{1/4}}, \quad \Sigma_{2,K}(x) = \sum_{n=K+1}^{2K} \theta_n \frac{\sin[\sqrt{xn}]}{n^{1/4}}$$

as $x \to \infty$. Now, as $\Sigma_{1,K}(x)$ and $\Sigma_{2,K}(x)$ are harmonic sums, their asymptotic behaviour $(x, K \to \infty)$ can be analysed using the Mellin transform approach, which yields that $|S_K(x)| \lesssim K^\lambda$ for $x > K^{\frac{\lambda-3/4}{\lambda-1/4}}$. This yields (31) for $\mu = 0$ and $2x < 4K+2$. The case $\mu \neq 0$ is here straightforward. This ends the proof of Step 2. □

Therefore (31) is proved so the proof of Lemma 3.2 is complete. □

Proof of Lemma 3.3. Set for any $j \in \mathbb{N}_0 = \{1, 2, \dots, \}$,

$$\Theta_{2j} := \left\{ (\theta_1, \dots, \theta_{2j}) \in \{0,1\}^{2j} : \sum_{k=1}^{2l} \theta_k(-1)^k = 0, \quad l = 1, \dots, j \right\},$$

then it obviously holds

$$\left|\Theta_{2(j+1)}\right| = 2\left|\Theta_{2j}\right|, \quad |\Theta_2| = 1.$$

Indeed

$$\Theta_{2(j+1)} = \{(\theta_1, \dots, \theta_{2j}, 0, 0); (\theta_1, \dots, \theta_{2j}, 1, 1), (\theta_1, \dots, \theta_{2j}) \in \Theta_{2j}\}.$$

Thus

$$|\Theta_{2j}| = 2^j.$$

And, for any sequence $\boldsymbol{\theta} \in \Theta_{2j}$, it holds $\left|\sum_{k=1}^{l} \theta_k(-1)^k\right| \leq 1$ for any $l = 1, \dots, 2k$. Hence the set

$$\Omega_K \doteq \left\{ (\theta_1, \dots, \theta_K) \in \{0,1\}^K : \left|\sum_{k=1}^{l} \theta_k(-1)^k\right| \leq 1, \quad l = 1, \dots, K, \quad \sum_{k=1}^{K} \theta_k(-1)^k = 0 \right\}$$

satisfies $|\Omega_K| \geq 2^{K/2}$ for all even K. Next we follow the proof of the Varshamov-Gilbert bound (see Tsybakov (2009)) applied to the set Ω_K and get that for any even $K \geq 16$ there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of Ω_K such that

$$\rho(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^{(l)}) \geq K/16, \quad 0 \leq j < l \leq M,$$

and $M \geq 2^{K/16}$. □

Proof of Lemma 3.4. Equality (10) and Lemma 3.3 imply $\|f_{\boldsymbol{\theta}^{(j)}} - f_{\boldsymbol{\theta}^{(l)}}\|^2 \geq A_1\delta^2 K$, for $0 \leq j \neq l \leq M$.

From (9), we have

$$\begin{aligned}\chi^2(f_{\boldsymbol{\theta}}, f_0) &= \int_0^\infty \frac{(f_{\boldsymbol{\theta}}(x) - f_0(x))^2}{f_0(x)}\, dx \\ &\leq C_1 \int_0^\infty (f_{\boldsymbol{\theta}}(x) - f_0(x))^2\, dx + C_2 \int_0^\infty \left(x^{(\alpha+1)/2} f_{\boldsymbol{\theta}}(x) - x^{(\alpha+1)/2} f_0(x)\right)^2 dx\end{aligned}$$

for some constants $C_1, C_2 > 0$. First

$$\int_0^\infty (f_{\boldsymbol{\theta}}(x) - f_0(x))^2\, dx = \delta^2 \sum_{k=1}^K \theta_k^2 \leq \delta^2 K$$

Next, using the relation (see formula (22.7.12) in Abramowitz and Stegun (1964)),

$$xL_k(x) = (2k+1)L_k(x) - kL_{k-1}(x) - (k+1)L_{k+1}(x),$$

we derive that for $(\alpha+1)/2$ integer,

$$x^{(\alpha+1)/2}(f_{\boldsymbol{\theta}}(x) - f_0(x)) = \delta \sum_{k=K+1-(\alpha+1)/2}^{2K+(\alpha+1)/2} \psi(k, K, \alpha, \boldsymbol{\theta})\varphi_k(x)$$

where $|\psi(k, K, \alpha, \boldsymbol{\theta})| \lesssim K^{(\alpha+1)/2}$. Now, with the orthonormality of the system $\{\varphi_k\}$, we get

$$\chi^2(f_{\boldsymbol{\theta}}, f_0) \lesssim \delta^2 K^{\alpha+2}, \quad K \to \infty$$

uniformly in $\boldsymbol{\theta} \in \{0,1\}^K$.

If $(\alpha+1)/2$ is not an integer, splitting the last integral between 0 and 1 and 1 and ∞, we get a bound $\delta^2 K^{\alpha_0+1}$ where α_0 is the smallest even integer larger than $\alpha+1$. Therefore,

$$\chi^2(f_{\boldsymbol{\theta}}, f_0) \lesssim \delta^2 K^{\alpha+4}, \quad K \to \infty$$

uniformly in $\boldsymbol{\theta} \in \{0,1\}^K$ and we get Lemma 3.4. □

7.3. **Proof of Propositions 4.2 and 4.3 for $k = 1$.** We first look at the case $k = 1$ before the general k-monotone case.

Set $f_1 = f_{1,Y}$. We have

$$\langle f_1, (y\varphi_j)' \rangle = [f_1(y) y\varphi_j(y)]_{y=0}^{y=+\infty} + \int_0^{+\infty} \frac{f(y)}{y} \times y\varphi_j(y)dy = \langle f, \varphi_j \rangle.$$

This yields (14) for $k = 1$.

As $y\varphi_j'(y)e^y = \sqrt{2}y[2L_j'(2y) - L_j(2y)]$ is a polynomial with degree $j+1$, it can be decomposed in the Laguerre polynomial basis of degree $j+1$. There exist coefficients $b_\ell^{j,1}$ such that

$$y\varphi_j'(y) = \sum_{\ell=0}^{j+1} b_\ell^{j,1} \varphi_\ell(y)$$

and using the specific properties of Laguerre polynomials we can compute the coefficient $b_\ell^{j,1}$. Let $L_j^{(\alpha)}$ be the generalized Laguerre polynomials given by Formula (22.3.9) in Abramowitz and

Stegun (1964) and $L_j = L_j^{(0)}$. By (22.5.17) for $m = 1$ in Abramowitz and Stegun (1964), we have

$$L_j'(x) = -L_{j-1}^{(1)}(x). \tag{36}$$

Moreover, Formula (22.7.31) in Abramowitz and Stegun (1964) gives

$$xL_j^{(1)}(x) = (j+1)[L_j(x) - L_{j+1}(x)], \tag{37}$$

and Formula (22.7.12) therein

$$xL_j(x) = -(j+1)L_{j+1}(x) + (2j+1)L_j(x) - jL_{j-1}(x). \tag{38}$$

We have to compute $2yL_j'(2y) - yL_j(2y)$ or $tL_j'(t) - \frac{t}{2}L_j(t)$. Combining relations (36),-(38), we get

$$tL_j'(t) - \frac{t}{1}L_j(t) = \frac{j+1}{2}L_{j+1}(t) - \frac{1}{2}L_j(t) - \frac{j}{2}L_{j-1}(t).$$

Thus, $b_\ell^{j,1} = 0$ for $\ell \neq j-1, j, j+1$ and

$$b_{j-1}^{j,1} = -\frac{j}{2}, \quad b_j^{j,1} = -\frac{1}{2}, \quad b_{j+1}^{j,1} = \frac{j+1}{2}. \tag{39}$$

Finally,

$$(y\varphi_j)' = \varphi_j(y) + y\varphi_j'(y) = -\frac{j}{2}\varphi_{j-1}(y) + \frac{1}{2}\varphi_j(y) + \frac{j+1}{2}\varphi_{j+1}(y). \tag{40}$$

This gives the result for $k = 1$. □

7.4. **Proof of Proposition 4.2 for $k \geq 2$.**
Let $f_k = f_{k,Y}$. Using (12), we write

$$\langle f, \varphi_j \rangle = \frac{(-1)^k}{k!} \int_0^{+\infty} f_k^{(k)}(y)(y^k\varphi_j(y))dy$$

and by integration by part we have

$$\langle f, \varphi_j \rangle = -\frac{(-1)^k}{k!} \int_0^{+\infty} f_k^{(k-1)}(y)(y^k\varphi_j(y))^{(1)}dy = \cdots = (-1)^k \frac{(-1)^k}{k!} \int_0^{+\infty} f_k(y)(y^k\varphi_j(y))^{(k)}dy$$

provided that all terms appearing in the integration by parts are null, *i.e.*:

$$\left[\sum_{\ell=1}^k f_k^{(k-\ell)}(y)(y^k\varphi_j(y))^{(\ell-1)}(-1)^{\ell-1} \right]_0^{+\infty} = 0 \tag{41}$$

Therefore, we obtain Formula (14) after proving that (41) holds.

Proof of (41): Let

$$S(y) = \sum_{\ell=1}^k f_k^{(k-\ell)}(y)(y^k\varphi_j(y))^{(\ell-1)}(-1)^{\ell-1} = \sum_{p=0}^{k-1} f_k^{(p)}(y)(y^k\varphi_j(y))^{(k-p-1)}(-1)^{k-p-1}.$$

Using the Leibniz formula and interchanging sums yields

$$S(y) = \sum_{t=0}^{k-1} \varphi_j^{(t)}(y)\Sigma_t(y)$$

with

$$\Sigma_t(y) = \sum_{p=0}^{k-1-t} (-1)^{k-p-1} f_k^{(p)}(y) y^{p+1+t} \binom{k+p-1}{t} k \times (k-1) \ldots \times (p+t+2).$$

As $\varphi_j^{(t)}(y)$ is continuous at 0 and tends to 0 at $+\infty$, we only need to prove that $\Sigma_t(y)$ tends to 0 at 0 and $+\infty$. We look at the coefficient of $\varphi_j^{(0)} = \varphi_j$:

$$\Sigma_0(y) = (-1)^k k! \sum_{p=0}^{k-1} y^{p+1} (-1)^{p+1} f_k^{(p)}(y) \frac{1}{(p+1)!}.$$

By (13), $\Sigma_0(y) = (-1)^k k!(F(y) - F_{k,Y}(y))$. As F and $F_{k,Y}$ are continuous c.d.f. on $\mathbb{R}^+$, they are null at 0 and both tend to 1 at $+\infty$. Therefore, as y tends to 0 and $+\infty$,

$$\Sigma_0(y) \to 0.$$

For the term $\Sigma_1(y)$, we prove that each term $f_k^{(p)}(y) y^{p+2}, p = 0, \ldots, k-2$ tends to 0 at both 0 and $+\infty$. Indeed,

$$f_k^{(p)}(y) y^{p+2} \propto y^{p+2} \int_y^{+\infty} \frac{(u-y)^{k-1-p}}{u^k} f(u) du.$$

$$|f_k^{(p)}(y) y^{p+2}| \lesssim \int_y^{+\infty} \frac{y^{p+2}}{u^{p+1}} f(u) du \le y \int_y^{+\infty} f(u) du \tag{42}$$

which tends to 0 as y tends to 0. Also,

$$|f_k^{(p)}(y) y^{p+2}| \lesssim \int_y^{+\infty} \frac{y^{p+2}}{u^{p+1}} f(u) du \le \int_y^{+\infty} u f(u) du \tag{43}$$

which tends to 0 as y tends to $+\infty$ as $\mathbb{E}(X) < +\infty$. We proceed analogously for all terms $\Sigma_t(y), t \le k-1$. We prove that $f_k^{(p)}(y) y^{p+t+1}, p = 0, \ldots, k-t-1$ tends to 0 at both 0 and $+\infty$. The convergence at 0 is already done. For the convergence at $+\infty$, we use that

$$|f_k^{(p)}(y) y^{p+t+1}| \lesssim \int_y^{+\infty} \frac{y^{p+t+1}}{u^{p+1}} f(u) du \le \int_y^{+\infty} u^t f(u) du \tag{44}$$

which tends to 0 at $+\infty$ by the moment assumption $\mathbb{E}(X^t) < +\infty$. The proof of (41) is complete.□

7.5. **Proof of Proposition 4.3.** The function $(y^k \varphi_j)^{(k)}/k!$ belongs to S_{j+k}, and therefore admits a decomposition on the basis of the φ_ℓ, for $\ell = 0, 1, \ldots, j+k$:

$$\frac{1}{k!}(y^k \varphi_j)^{(k)} = \sum_{\ell=0}^{j+k} h_\ell^{j,k} \varphi_\ell(y).$$

This decomposition is obtained as follows. The Leibnitz formula yields:

$$\frac{1}{k!}(y^k \varphi_j)^{(k)} = \sum_{p=0}^{k} \binom{k}{p} \frac{1}{p!} y^p \varphi_j^{(p)}. \tag{45}$$

Next, the development of $y^p \varphi_j^{(p)}(y)$ is given in the following lemma.

Lemma 7.4. *We have*

$$y^p \varphi_j^{(p)}(y) = \sum_{\ell=0\vee(j-p)}^{j+p} b_\ell^{j,p} \varphi_\ell(y), \tag{46}$$

where $b_\ell^{j,0} = \delta_{\ell,j}$ *and for* $p \geq 0$,

$$b_\ell^{j,p+1} = -\frac{\ell+1}{2} b_{\ell+1}^{j,p} - (p+\frac{1}{2}) b_\ell^{j,p} + \frac{\ell}{2} b_{\ell-1}^{j,p}. \tag{47}$$

Moreover

$$\forall \ell \leq j+p, \ |b_\ell^{j,p}| \leq C_p (j+p)^p. \tag{48}$$

Applying Lemma 7.4, and interchanging sums in (45) yields formula (15). Next, we use Formula (48) to get

$$|h_\ell^{j,k}| \leq \sum_{p=|\ell-j|}^{k} C_p (j+p)^p \binom{k}{p} \frac{1}{p!} \leq \max_{p\leq k}(C_p)(j+k+1)^k \leq C'_k (j+k)^k.$$

This gives the bound (16).

Proof of Lemma 7.4 Initialization of (46) for $p=0$ is obvious. Formula (39) shows that the induction formula (47) holds for $p=0$ ($p=0$ to $p=1$).

Next, we differentiate (46) and multiply by y, to get

$$y\left(y^p \varphi_j^{(p+1)}(y) + p y^{p-1} \varphi_j^{(p)}(y)\right) = \sum_{\ell=0\vee j-p}^{p+j} b_\ell^{j,p} \ y\varphi'_\ell(y)$$

Now using (39), we get

$$y^{p+1}\varphi_j^{(p+1)}(y) = -p y^p \varphi_j^{(p)}(y) + \sum_{\ell=0\vee j-p}^{j+p} b_\ell^{j,p} \left(-\frac{\ell}{2}\varphi_{\ell-1}(y) - \frac{1}{2}\varphi_\ell(y) + \frac{\ell+1}{2}\varphi_{\ell+1}(y)\right).$$

Taking into account that $-py^p\varphi_j^{(p)}(y) = -\sum_{\ell=0\vee(j-p)}^{j+p} p\, b_\ell^{j,p}\varphi_\ell(y)$ gives formula (47). Inequality (48) is obtained by straightforward induction. The proof of Proposition 4.2 is now complete. □

7.6. **Proof of Proposition 4.4 and 5.2.** We give the proof of Proposition 5.2. The other result follows by setting $\mathbf{V}_{m+k} = I_{m+k}$ the identity matrix of size $(m+k)\times(m+k)$.

The risk bound of the estimator can be written as follows

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

where $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$ is the projection of f on $S_m = \text{span}(\varphi_0, \dots, \varphi_{m-1})$ and $\|f-f_m\|^2$ is the usual bias of a projection estimate. Next we have

$$\|\tilde{f}_m - f_m\|^2 = \sum_{j=0}^{m-1} (\tilde{a}_j - a_j(f))^2 = \|K_m^{(k)}(\vec{\tilde{a}}(Z)_{m+k-1} - \mathbb{E}(\vec{\tilde{a}}(Z)_{m+k-1}))\|^2$$

where $\vec{\tilde{a}}(Z)_{m+k-1} = (\tilde{a}_j(Z))_{0\le j\le m+k-1}$. So,

$$\begin{aligned}
\mathbb{E}(\|\tilde{f}_m - f_m\|^2) &\le \rho^2(K_m^{(k)})\mathbb{E}(\|\vec{\tilde{a}}(Z)_{m+k-1} - \mathbb{E}(\vec{\tilde{a}}(Z)_{m+k-1})\|^2)\\
&\le \rho^2(K_m^{(k)})\sum_{j=0}^{m+k-1}\mathrm{Var}(\tilde{a}_j(Z)) = \frac{1}{n}\rho^2(K_m^{(k)})\sum_{j=0}^{m+k-1}\mathrm{Var}(\varphi_j(Z_1))\\
&\le \frac{1}{n}\rho^2(K_m^{(k)})\sum_{j=0}^{m+k-1}\mathbb{E}(\varphi_j^2(Z_1))\\
&\le \frac{2(m+k)\rho^2(K_m^{(k)})}{n},
\end{aligned}$$

as $\sum_{j=0}^{m+k-1}\varphi_j^2(x) \le 2(m+k)$, $\forall x \in \mathbb{R}^+$. We have

$$\rho^2(K_m^{(k)}) \le \rho^2(H_m^{(k)})\rho^2(\mathbf{V}_{m+k}^{-1})$$

$$\rho^2(H_m^{(k)}) = \sup_{x\in\mathbb{R}^{m+k},\|x\|^2=1} x^t\,(H_m^{(k)})^t\,H_m^{(k)}x = \sup_{x\in\mathbb{R}^{m+k},\|x\|^2=1}\sum_{j=1}^{m}(\sum_{\ell=(j-k)^+}^{j+k}[H_m^{(k)}]_{j,\ell}x_\ell)^2$$

We consider first $m \ge k$ and use (16) to get

$$\begin{aligned}
\sum_{j=1}^{m}(\sum_{\ell=(j-k)^+}^{j+k}[H_m^{(k)}]_{j,\ell}x_\ell)^2 &\le (C_k')^2(2k+1)\sum_{j=1}^{m}(j+k)^{2k}\sum_{\ell=(j-k)^+}^{j+k}x_\ell^2\\
&\le (C_k')^2(2k+1)(m+k)^{2k}\sum_{j=1}^{m}\sum_{\ell=(j-k)^+}^{j+k}x_\ell^2.
\end{aligned}$$

Next write that

$$\sum_{j=1}^{m}\sum_{\ell=(j-k)^+}^{j+k}x_\ell^2 = \sum_{j=1}^{k}\sum_{\ell=0}^{j+k}x_\ell^2 + \sum_{j=k+1}^{m}\sum_{\ell=j-k}^{j+k}x_\ell^2$$

Interchanging sums yields

$$\sum_{j=k+1}^{m}\sum_{\ell=j-k}^{j+k}x_\ell^2 = \sum_{\ell=1}^{m+k}\sum_{j=(\ell-k)\vee 1}^{(\ell+k)\wedge m}x_\ell^2 \le (2k+1)\sum_{\ell=1}^{m+k}x_\ell^2$$

Therefore we get

$$\rho^2(H_m^{(k)}) \le C_k''(m+k)^{2k} \text{ with } C_k'' = (C_k')^2(2k+1)(3k+1).$$

If $m < k$, the bound obviously holds.
Therefore we get

$$\begin{aligned}
\mathbb{E}(\|\tilde{f}_m - f\|^2) &\le \|f-f_m\|^2 + \frac{2(m+k)\rho^2(K_m^{(k)})}{n}\\
&\le \|f-f_m\|^2 + C_k\frac{(m+k)^{2k+1}\rho^2(\mathbf{V}_{m+k}^{-1})}{n}
\end{aligned}$$

This ends the proof. □

7.7. **Proof of the Lemmas of Section 4.3 (Theorem 4.1).**

Proof of Lemma 4.1. For $\tilde{f}_0$ we follow the same line as in the proof of Lemma and omit the details. Next $\tilde{f}_{\boldsymbol{\theta}}$ belongs to $W^s(D)$ if $\delta^2 \sum_{k=K}^{2K+1} \mu_k^2(\boldsymbol{\theta}) k^s \le D/4$ i.e.

$$\delta^2 K^{s+3} \le D/C$$

where $C = C(s)$ is a constant. □

Proof of Lemma 4.2. First note that $\tilde{f}_0(0) = \tilde{f}_{\boldsymbol{\theta}}(0) = 0$ and $\tilde{f}_0'(0) = c_{\alpha,\beta}(\alpha+\beta)/e^{\alpha+1} > 0$ and

$$\tilde{f}_{\boldsymbol{\theta}}'(0) = \tilde{f}_0'(0) + \delta \sum_{k=K+1}^{2K} \theta_{k-K} \varphi_k'(0) = \tilde{c}_\alpha - \delta\sqrt{2} \sum_{k=K+1}^{2K} (2k+1)\theta_{k-K}.$$

Now $\tilde{f}_{\boldsymbol{\theta}}'(0) > 0$ if

$$\delta K^2 << 1.$$

Under this condition, $\tilde{f}_{\boldsymbol{\theta}}$ is nonnegative on an interval $[0, \mathtt{a}]$, $\mathtt{a} > 0$.

For $x > \mathtt{a}$, we follow the arguments in the proof of Lemma 3.2 for each of the three terms involved in the definition of $\mu_k(\boldsymbol{\theta})$. Thus we must prove that

$$\sup_{x>\mathtt{a}} \left| x^\lambda \log^\mu(e+x) \sum_{k=K}^{2K+1} \theta_{k-K} k \varphi_k(x) \right| \lesssim K^{\lambda+1} \log^\mu(K).$$

This is obtained as previously (just change $\rho_k^{(\lambda)}(r)$ into $k\rho_k^{(\lambda)}(r)$, see Step 1 of the proof of Lemma 3.2). Then by taking $\lambda = \alpha$, $\mu = \beta$ and $\delta = \delta' K^{-\alpha-1} \log^{-\beta}(K)$ for small enough constant $\delta' > 0$ not depending on K, we get $\tilde{f}_{\boldsymbol{\theta}}(x) \ge 0$, $x \in \mathbb{R}_+$. □

Proof of Lemma 4.3. Let

$$\Theta = \{(\theta_0, \dots, \theta_{2K+1}), \theta_0 = \theta_1 = 0, \theta_j \in \{0,1\}, \text{ for } j = 2, \dots, 2K+1\}.$$

We prove that $\rho_1(.,.)$ is a distance on Θ. Due to the initial conditions $\theta_0 = \theta_1 = 0$, $\rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}') = 0$ implies that $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

For $\boldsymbol{\theta} \in \Theta$, we separate $\boldsymbol{\theta} = (\theta_0, \dots, \theta_{2K+1})$ as $\boldsymbol{\theta}^{(even)} := (\theta_{2j}, 0 \le j \le K)$ and $\boldsymbol{\theta}^{(odd)}$ accordingly. Let $\rho_2(\omega, \omega') = \sum_{k=0}^K \mathbf{1}_{\omega_{k+1}-\omega_k \neq \omega'_{k+1}-\omega'_k}$, then

$$\rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}') = \rho_2((\boldsymbol{\theta})^{(even)}, (\boldsymbol{\theta}')^{(even)}) + \rho_2((\boldsymbol{\theta})^{(odd)}, (\boldsymbol{\theta}')^{(odd)}).$$

Now we can check that ρ_2 satisfies the triangular inequality on $\Omega = \{(\omega_0, \dots, \omega_K), \omega_0 = 0, \omega_j \in \{0,1\}, j = 1, \dots, K\}$. For $\epsilon, \epsilon' \in \{-1, 0, 1\}$, we note that

$$\mathbf{1}_{\epsilon \neq \epsilon'} = \frac{1}{2}\left(|\epsilon - \epsilon'| + ||\epsilon| - |\epsilon'||\right) = d(\epsilon, \epsilon')$$

where $d(.,.)$ satisfies the triangular inequality. Setting $\epsilon_k = \omega_{k+1} - \omega_k$, we get that $\rho_2(\omega, \omega') = \sum_{k=0}^K d(\epsilon_k, \epsilon'_k)$ satisfies the triangular inequality on Ω.

Thus, it is enough to prove the Lemma for the set Ω and the distance ρ_2.

Following the proof of the Varshamov-Gilbert Lemma as given in Tsybakov, this amounts to proving that for $\omega^{(0)} = (0, \dots, 0) \in \Omega$, $\text{Card}(\{(\omega_k) \in \Omega, \rho_2(\omega, \omega^{(0)}) = i\}) = \binom{K}{i}$. Let

$$A_{m,i} := \text{Card}\left(\left\{\omega \in \Omega, \sum_{k=0}^K \mathbf{1}_{\omega_{k+1}-\omega_k = 0} = i\right\}\right).$$

Note that

$$
\begin{aligned}
A_{K,i} &= \text{Card}\left(\left\{\omega\in\Omega, \omega_1-\omega_0=0, \sum_{k=1}^K \mathbf{1}_{\omega_{k+1}-\omega_k=0}=i-1\right\}\right)\\
&\quad +\text{Card}\left(\left\{\omega\in\Omega, \omega_1-\omega_0=1, \sum_{k=1}^K \mathbf{1}_{\omega_{k+1}-\omega_k=0}=i\right\}\right)\\
&= \text{Card}\left(\left\{\omega\in\Omega, \omega_0=0, \omega_1=0 \sum_{k=1}^K \mathbf{1}_{\omega_{k+1}-\omega_k=0}=i-1\right\}\right)\\
&\quad +\text{Card}\left(\left\{\omega\in\Omega, \omega_0=0, \omega_1=1, \sum_{k=1}^K \mathbf{1}_{\omega_{k+1}-\omega_k=0}=i\right\}\right)\\
&= A_{K-1,i-1}+A_{K-1,i}.
\end{aligned}
$$

As $A_{1,0}=A_{1,1}=1$, we deduce $A_{K,i}=\binom{K}{i}$ by the definition of the binomial coefficients. □

Proof of Lemma 4.4. The first inequality follows from Lemma 3.4 and $f_{Y,\boldsymbol{\theta}}=f_{\boldsymbol{\theta}}$ and $f_{Y,0}=f_0$. The second inequality follows from (20), (21) and Lemma 4.3. □

7.8. **Proof of Theorem 5.1.** Let $M=\max\mathcal{M}_n$ the maximal element of the collection. Let for $m\le M$, $S_m=\{\vec{t}\in\mathbb{R}^M\ \vec{t}=(t_1,\dots,t_m,0,\dots,0)\}$ and for any $\vec{t}\in\mathbb{R}^M$, let

$$\gamma_n(\vec{t})=\|\vec{t}\|_M^2-2\langle\vec{t},K_M^{(k)}\vec{\tilde{a}}_{M+k-1}(Z)\rangle_M,$$

where $\|\vec{x}\|_M^2$ is the Euclidean norm in $\mathbb{R}^M$ and $\langle\cdot,\cdot\rangle_M$ the associated scalar product. For $\vec{t}\in S_m$, we denote by $\vec{t_m}$ the vector of $\mathbb{R}^m$ with the m first coordinates of $\vec{t}$ (those which are not necessarily zero). Thanks to the particular forms of the matrices $H_m^{(k)}$ (band) and $\mathbf{V}_{m+k}^{-1}$ (lower triangular), we have, for $\vec{t}\in S_m$,

$$\langle\vec{t},K_M^{(k)}\vec{\tilde{a}}_{M+k-1}(Z)\rangle_M=\langle\vec{t_m},K_m^{(k)}\vec{\tilde{a}}_{m+k-1}(Z)\rangle_m=\langle\vec{t_m},\vec{\tilde{a}}_{m-1}\rangle_m.$$

Therefore the vector $\vec{\tilde{f}}_{(m)}$ in $\mathbb{R}^M$ with m first coordinates $\vec{\tilde{a}}_{m-1}$ and following coordinates null is such that $\vec{\tilde{f}}_{(m)}=\arg\min_{\vec{t}\in S_m}\gamma_n(\vec{t})$ and $\gamma_n(\vec{\tilde{f}}_{(m)})=-\|\tilde{f}_m\|^2$. Therefore

$$\tilde{m}=\arg\min_{m\in\mathcal{M}_n}\gamma_n(\vec{\tilde{f}}_{(m)})+\text{pen}(m)).$$

Now for $m,m'\in\mathcal{M}_n$, and $\vec{t}\in S_m$, $\vec{s}\in S_{m'}$, we have

$$
\begin{aligned}
\gamma_n(\vec{t})-\gamma_n(\vec{s}) &= \|\vec{t}-\vec{f}_M\|_M^2-\|\vec{s}-\vec{f}_M\|_M^2-2\langle\vec{t}-\vec{s},K_M^{(k)}\vec{\tilde{a}}_{M+k-1}(Z)-\vec{f}_M\rangle_M\\
&= \|\vec{t}-\vec{f}_M\|_M^2-\|\vec{s}-\vec{f}_M\|_M^2-2\langle\vec{t}-\vec{s},K_M^{(k)}(\vec{\tilde{a}}_{M+k-1}(Z)-\vec{a}_{M+k-1}(f_Z)\rangle_M
\end{aligned}
$$

where $\vec{f}_M=(a_j(f))_{0\le j\le M-1}$. Let us define

$$\nu_n(\vec{t})=\langle\vec{t},K_M^{(k)}(\vec{\tilde{a}}_{M+k-1}(Z)-\vec{a}_{M+k-1}(f_Z)\rangle_M,$$

and note that

$$\|\tilde{f}_m-f\|^2=\|\vec{\tilde{f}}_{(m)}-\vec{f}_M\|_M^2+\sum_{j=M}^{\infty}a_j^2(f),\quad \|f_m-f\|^2=\|\vec{f}_m-\vec{f}_M\|_M^2+\sum_{j=M}^{\infty}a_j^2(f). \tag{49}$$

By definition of $\tilde{m}$, we have

$$\gamma_n(\vec{\tilde{f}}_{(\tilde{m})}) + \text{pen}(\tilde{m}) \leq \gamma_n(\vec{f}_m) + \text{pen}(m),$$

which writes

$$\|\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_M\|_M^2 \leq \|\vec{f}_m - \vec{f}_M\|_M^2 + \text{pen}(m) + 2\nu_n(\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_m) - \text{pen}(\tilde{m}). \tag{50}$$

Let $B(\tilde{m}, m) = \{\vec{t} \in S_{m\vee\tilde{m}}, \|\vec{t}\|_M = 1\}$ and note that

$$\begin{aligned}
2\nu_n(\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_m) &\leq 2\|\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_m\|_M \sup_{\vec{t}\in B(\tilde{m},m)} |\nu_n(\vec{t})| \\
&\leq \frac{1}{4}\|\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_m\|_M^2 + 4 \sup_{\vec{t}\in B(\tilde{m},m)} \nu_n^2(\vec{t}) \\
&\leq \frac{1}{2}\|\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_M\|_M^2 + \frac{1}{2}\|\vec{f}_m - \vec{f}_M\|_M^2 + 4 \sup_{\vec{t}\in B(\tilde{m},m)} \nu_n^2(\vec{t}).
\end{aligned}$$

We get by plugging this in (50),

$$\frac{1}{2}\|\vec{\tilde{f}}_{(\tilde{m})} - \vec{f}_M\|_M^2 \leq \frac{3}{2}\|\vec{f}_m - \vec{f}_M\|_M^2 + \text{pen}(m) + 4 \sup_{\vec{t}\in B(\tilde{m},m)} \nu_n^2(\vec{t}) - \text{pen}(\tilde{m})$$

Let $p(m, m')$ be such that $4p(m, m') \leq \text{pen}(m) + \text{pen}(m')$ and use (49), to get

$$\frac{1}{2}\|\tilde{f}_{\tilde{m}} - f\|^2 \leq \frac{3}{2}\|f_m - f\|^2 + 2\text{pen}(m) + 4\left(\sup_{\vec{t}\in B(\tilde{m},m)} \nu_n^2(\vec{t}) - p(m,\tilde{m})\right)_+$$

Now, if we prove that

$$\mathbb{E}\left[\left(\sup_{\vec{t}\in B(\tilde{m},m)} \nu_n^2(\vec{t}) - p(m,\tilde{m})\right)_+\right] \leq \frac{c}{n} \tag{51}$$

the result follows.

The proof of (51) follows the line of the proof of Proposition 7.1 in Mabon (2015).

8. Appendix

For simplicity, set $f_{k,Y} = f_k$. For $k = 1$,

$$f_1(y) = \int_y^{+\infty} \frac{f(u)}{u} du 1_{y\geq 0}.$$

Derivating yields the first equality in (12). Integrating between 0 and y gives the second equality which implies:

$$\lim_{y\to 0} y\, f_1(y) = \lim_{y\to+\infty} y\, f_1(y) = 0. \tag{52}$$

To get (12), we proceed by induction and prove that, for any p such that $1 \leq p \leq k-1$,

$$\frac{d^p}{dy^p}[f_k(y)] = (-1)^p k \times \cdots \times (k-p) \sum_{j=0}^{k-1-p} \binom{k-1-p}{j} (-y)^j \int_y^{+\infty} \frac{f(u)}{u^{j+p+1}} du. \tag{53}$$

The formula is true for $p=0$ as (2) implies

$$f_k(y) = k\sum_{j=0}^{k-1}\binom{k-1}{j}(-y)^j\int_y^{+\infty}\frac{f(u)}{u^{j+1}}du.$$

Now if we admit the formula for order p, we can deduce that, derivating once more,

$$\begin{aligned}\frac{d^{p+1}}{dy^{p+1}}[f_k(y)] &= (-1)^p k\times\cdots\times(k-p)\left\{\sum_{j=1}^{k-1-p}\binom{k-1-p}{j}(-1)^j(jy^{j-1})\int_y^{+\infty}\frac{f(u)}{u^{j+p+1}}du\right.\\ &\left.+\sum_{j=0}^{k-1-p}\binom{k-1-p}{j}(-1)^{j+1}y^j\frac{f(y)}{y^{j+p+1}}\right\}.\end{aligned}$$

The last sum is equal to

$$-\frac{f(y)}{y^{p+1}}\sum_{j=0}^{k-1-p}\binom{k-1-p}{j}(-1)^j = -\frac{f(y)}{y^{p+1}}(1-1)^{k-1-p}=0$$

and for the first one, we note that

$$j\binom{k-1-p}{j} = (k-1-p)\binom{k-2-p}{j-1}$$

so that we get

$$\frac{d^{p+1}}{dy^{p+1}}[f_k(y)] = (-1)^p k\times\cdots\times(k-p)\times(k-p-1)\sum_{j=1}^{k-1-p}\binom{k-2-p}{j-1}(-1)^j y^{j-1}\int_y^{+\infty}\frac{f(u)}{u^{j+p+1}}du$$

and setting $j' = j-1$ in the sum gives the result at order $p+1$. Therefore Formula (53) is proved for all $p=0,\dots k-1$. Taking $p=k-1$ and derivating once more gives Formula (12).

To obtain (13), we integrate (12) between 0 and y. The successive integrations by part give the result provided that, for $\ell = 0,\dots,k$,

$$y^{k-\ell}f_k^{(k-\ell-1)}(y)\to 0, \quad \text{as} \quad y\to 0.$$

For this notice that, as for $u\geq y\geq 0$, $u-y\leq u$ and $y/u\leq 1$,

$$|y^{k-\ell}f_k^{(k-\ell-1)}(y)| \propto y^{k-\ell}\int_y^{+\infty}\frac{(u-y)^{k-1-(k-\ell-1))}}{u^k}f(u)du \ \leq y\int_y^{+\infty}\frac{f(u)}{u}du.$$

The r.h.s. above is equal to $yf_1(y)$ and tends to 0 as y tends to 0 by (52).□

References

[1] Abbaszadeh, M., Chesneau, C. and Doosti, H. (2012) Nonparametric estimation of density under bias and multiplicative censoring via wavelet methods. *Statist. Probab. Lett.* **82**, 932-941.

[2] Abbaszadeh, M, Chesneau, C. and Doosti, H. (2013) Multiplicative censoring: estimation of a density and its derivatives under the Lp-risk. *REVSTAT* **11**, 255-276.

[3] Abramowitz, M. and Stegun, I. A. (1964) Handbook of mathematical functions with formulas, graphs, and mathematical tables. National Bureau of Standards Applied Mathematics Series, 55 For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

[4] Asgharian, M., Carone, M., Fakoor, V. (2012) Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.* **40**, 159-187.

[5] Andersen, K. E. and Hansen, M. B. (2001) Multiplicative censoring: density estimation by a series expansion approach. *J. Statist. Plann. Inference* **98**, 137-155.

[6] Balabdaoui, F. and Wellner, J. A. (2007) Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.* **35**, 2536-2564.

[7] Balabdaoui, F. and Wellner, J. A. (2010) Estimation of a k-monotone density: characterizations, consistency and minimax lower bounds. *Stat. Neerl.* **64**, 45-70.

[8] Belomestny, D. and Schoenmakers, J. (2015). Statistical Skorohod embedding problem: optimality and asymptotic normality, *Statistics and Probability Letters*, **104**, 169–180.

[9] Bongioanni, B. and Torrea, J.L. (2009). What is a Sobolev space for the Laguerre function system? *Studia Mathematica* **192** (2), 147-172.

[10] Brunel, E., Comte, F. and Genon-Catalot, V. (2015) Nonparametric density and survival function estimation in the multiplicative censoring model. Preprint Hal and MAP5 2015-08. Accepted for publication in *Test*.

[11] Chee, C.-S. and Wang, Y. (2014) Least squares estimation of a k-monotone density function. *Comput. Statist. Data Anal.* **74**, 209-216.

[12] Comte, F., Cuenod, C.-A., Pensky, M. and Rozenholc, Y. (2013). Laplace deconvolution and its application to Dynamic Contrast Enhanced imaging. Preprint ArXiv, Preprint MAP5 2012-17. Accepted for publication in *J. R. Stat. Soc.* Ser. **B**.

[13] Comte, F. and Genon-Catalot, V. (2015) Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* **9**, 1113-1149.

[14] van Es, B., Spreij, P. and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli* **9**, 451-465.

[15] van Es, B., Jongbloed, G. and van Zuijlen, M. (1998) Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.* **26**, 2395-2406.

[16] Jirak, M., Meister, A. and Reiss, M. (2014) Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42**, 1970-2002.

[17] Jongbloed, G. (1998) Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica* **52**, 6-17.

[18] Mabon, G. (2015). Adaptive deconvolution on the nonnegative real line. Preprint HAL hal-01076927, version 2.

[19] Muckenhoupt, B. (1970). Asymptotic forms for Laguerre polynomials. *Proceedings of the American Mathematical Society*, 288-292.

[20] Shen, J. (2000). Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.* **38**, 1113-1133.

[21] Szegö, G. (1975) *Orthogonal polynomials.* Fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII. American Mathematical Society, Providence, R.I.

[22] Tsybakov, Alexandre B. (2009) Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009.

[23] Vardi, Y. (1989) Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751-761.

[24] Vardi, Y. and Zhang, C.-H. (1992) Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.

[25] Williamson R. E. (1956), Multiply monotone functions and their Laplace transforms, *Duke Mathematical Journal* **23**, 189-207.