Nonparametric Laguerre estimation in the multiplicative censoring model

Denis Belomestny, Fabienne Comte, Valentine Genon-Catalot

NONPARAMETRIC LAGUERRE ESTIMATION IN THE MULTIPLICATIVE CENSORING MODEL

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

Abstract. We study the model $Y_i = X_i U_i,\ i = 1, \dots, n$ where 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 $(X_i), (U_i)$, are independent. We aim at estimating f on $\mathbb{R}^+$ from the observations $(Y_1, \dots, Y_n)$. We propose projection estimators using a Laguerre basis. A data-driven procedure is described in order to select the dimension of the projection space, which performs automatically the bias variance compromise. Then, we give upper bounds on the $\mathbb{L}^2$-risk on specific Sobolev-Laguerre spaces. Lower bounds matching with the upper bounds within a logarithmic factor are proved. The method is illustrated on simulated data. May 24, 2016

(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. Projection estimator.

MSC 2010. 62G07.

1. Introduction

Consider observations $Y_1, \dots, Y_n$ such that

$$Y_i = X_i U_i, \quad i = 1, \dots, n. \tag{1}$$

where (X_i) are *i.i.d.* nonnegative random variables with unknown density f, (U_i) are *i.i.d.* with $\beta(1,k)$ density given by $f_U(u) := \rho_k(u) = k(1-u)^{k-1}\mathbb{1}_{[0,1]}(u)$, with $k \geq 1$ and the sequences $(X_i), (U_i)$ are independent.

For $k = 1$, *i.e.* if U_i has uniform density on $[0,1]$, model (1) is refered to as the *multiplicative censoring model* and has been widely investigated in the past decades. It was first introduced in Vardi (1989) and covers several important statistical problems, such as estimation under monotonicity constraints or deconvolution of an exponential variable. Such a model is usually applied in survival analysis (see *e.g.* van Es *et al.* (2000)). Numerous papers deal with the estimation of f by various nonparametric methods. A nonparametric maximum likelihood approach is investigated in Vardi (1989), Vardi and Zhang (1992), Asgharian *et al.* (2012). However, in the latter papers, authors assume that a m-sample of direct observations $X_1, \dots, X_m$ is available in addition to the Y-sample and the method does not apply to the case $m = 0$. Using only the Y-sample, projection methods have been proposed. In Andersen and Hansen (2001), considering the estimation of f as an inverse problem, the authors apply singular value decomposition in different bases. Their procedure is not adaptive. Abbaszadeh *et al.* (2012, 2013) use projection estimators on wavelets bases to estimate the density f and its derivatives. They provide adaptive estimators, upper bounds of the $\mathbb{L}^p$-risks but no lower bounds. Kernel estimators of f and of the survival function $\bar{F}(x) = 1 - F(x)$, where F is the cumulative distribution function, are studied in Brunel *et al.*, (2015). Extensions of model (1) are considered in Chesneau (2013) who

Date: May 24, 2016.

assumes that the U_i's are a product of independent uniform variables and the sequence (X_i) is α-mixing.

In this paper, we consider the extension of the multiplicative sensoring model to the case where U_i has $\beta(1, k)$ distribution and propose nonparametric estimators of f built as projection estimators on a Laguerre basis under the assumption that $f \in \mathbb{L}^2(\mathbb{R}^+)$. Laguerre bases, which are orthonormal bases of $\mathbb{L}^2(\mathbb{R}^+)$, are well fitted for nonparametric estimation of $\mathbb{R}^+$-supported functions. Moreover, the support of the density under estimation being hidden by the noise, it is an advantage to have basis functions with non compact support. These bases have been recently used by several authors, for instance, in Comte *et al.* (2015), for Laplace deconvolution of a signal observed with noise, in Comte and Genon-Catalot (2015), for estimation of the mixing distribution of a Poisson mixture model, in Mabon (2015), for deconvolution of densities on $\mathbb{R}^+$. Laguerre bases are related to Sobolev-Laguerre spaces which were introduced in Shen (2000) and with more details in Bongioanni and Torrea (2007). The regularity properties of a function f belonging to a Sobolev-Laguerre space are characterized by the rate of decay of the coefficients of the development of f in the Laguerre basis. The link between projection coefficients and regularity conditions in these spaces has been described in Comte and Genon-Catalot (2015). In the present paper, we choose a Laguerre basis and first establish explicit formulae linking the projection coefficients of $f_{k,Y}$, the density of Y in model (1), in the Laguerre basis to those of f. This allows to define a collection $(\hat{f}_m)$ of estimators of f. We obtain a $\mathbb{L}^2$-risk bound for $\hat{f}_m$. Then, we propose a data-driven choice $\hat{m}_k$ of the dimension m leading to an adaptive estimator $\hat{f}_{\hat{m}_k}$. Using Sobolev-Laguerre regularity spaces, we determine upper bounds for the rate of convergence of the $\mathbb{L}^2$-risk. Then, we study lower bounds and prove that upper and lower bounds match up to a logarithmic term. The lower bound on Sobolev-Laguerre balls is difficult to obtain and follows several technical steps. We start proving it in the case of direct observations of the X_i's, that is in the simple density model and then we obtain it for model (1) when $k = 1$. To avoid more technical developments, we just indicate how to extend it for all integer k.

In Section 2, we describe the Laguerre basis, build the projection estimators of f and provide the upper bound of their $\mathbb{L}^2$-risk. This leads to the adaptive procedure. In Section 3, we introduce the Sobolev-Laguerre regularity spaces and obtain upper bounds on the rate of convergence of the projection estimators on Sobolev-Laguerre balls. To prove lower bounds, one can follow the general scheme described, *e.g.* in Tsybakov (2009). However, in the considered situation it is more natural to construct alternatives as finite combinations of Laguerre functions with coefficients taking values in $\{0, 1\}$. Such a construction makes the problem of attributing a hypothesis to Sobolev-Laguerre ball straightforward. Then the lower bounds are obtained via a modification of the Hamming distance and a corresponding extension of the Varshamov-Gilbert bound. In Section 4, we implement the adaptive estimators of f, based on direct observations $X_1, \dots, X_n$ and on multiplicative censored observations $Y_1, \dots, Y_n$ for $k = 1, 2$ and for various densities f. The method provides very good results for direct observations, which remain convincing for censored data. Extensions and concluding remarks are given in Section 5.

2. Projection estimators in the Laguerre basis

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)):

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

We assume that $f \in \mathbb{L}^2(\mathbb{R}^+)$, so that we can develop f on the Laguerre basis:

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

Let S_m be the m-dimensional subspace of $\mathbb{L}^2(\mathbb{R}^+)$ spanned by $(\varphi_0, \varphi_1, \dots, \varphi_{m-1})$. The function

$$f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j \tag{3}$$

is the orthogonal projection of f on S_m. Below, we define estimators $\hat{a}_j$ of $a_j(f)$ from the observations $Y_1, \dots, Y_n$. This leads to a collection of projection estimators $(\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j\varphi_j, m \geq 1)$.

2.2. **Preliminary properties and formulas.** Let $f_{k,Y}$ denote the density of Y_i given by (1). A straightforward computation leads to

$$f_{k,Y}(y) = k\int_y^{\infty} \left(1 - \frac{y}{u}\right)^{k-1} \frac{f(u)}{u}\, du\, \mathbb{I}_{y\geq 0}. \tag{4}$$

Moreover, another simple computation yields:

$$\|f_{k,Y}\| \leq \|f\|\mathbb{E}(\frac{1}{\sqrt{U_1}}) < +\infty. \tag{5}$$

Thus, $f_{k,Y}$ belongs to $\mathbb{L}^2(\mathbb{R}^+)$. In this paragraph, we prove that the coefficients $(a_j(f), j \geq 0)$ are linked with the coefficients $(a_j(f_{k,Y}) = \langle f_{k,Y}, \varphi_j\rangle, j \geq 0)$ of the density $f_{k,Y}$ on the Laguerre basis by a linear relation. This requires preliminary steps.

Let us remark that a density satisfying (4) is k-monotone, *i.e.* that $(-1)^\ell f_{k,Y}^{(\ell)}$ is nonincreasing and convex for $\ell = 0, \dots, k-2$ if $k \geq 2$ and simply nonincreasing if $k = 1$. This property is proved in Williamson (1956). Therefore, model (1) covers the case of observations with k-monotone densities. Note that k-monotone densities are considered in 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 Proposition 2.1 below, we state an inversion formula giving f from $f_{k,Y}$ defined by (4) proved in Williamson (1956). For convenience of the reader, we give a proof in the appendix.

Proposition 2.1. *Let $f_{k,Y}$ and f be linked by formula (4) 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 $k \geq 1$,*

$$f(y) = \frac{(-1)^k}{k!} y^k f_{k,Y}^{(k)}(y), \tag{6}$$

$$F(y) = F_{k,Y}(y) - yf_{k,Y}(y) + \cdots + \frac{(-1)^{k-1}}{(k-1)!} y^{k-1} f_{k,Y}^{(k-2)}(y) + \frac{(-1)^k}{k!} y^k f_{k,Y}^{(k-1)}(y). \tag{7}$$

Note that, setting $f_{0,Y} = f, F_{0,Y} = F$, these formulae contain the case where $Y_i = X_i$ ($U_i = 1$). So, below, we consider the case $k = 0$ in our results as the case of direct observations of the X_i's. With the two following propositions, we give the links between the coefficients of f and $f_{k,Y}$ on the Laguerre basis.

Proposition 2.2. *Assume that $\mathbb{E}X^{k-1} < +\infty$. Then, for all $j \geq 0$ and $k \geq 1$,*

$$a_j(f) = \langle f, \varphi_j \rangle = \frac{1}{k!} \langle f_{k,Y}, (y^k \varphi_j)^{(k)} \rangle \tag{8}$$

Proposition 2.2 provides a simple way of defining estimators of $a_j(f)$ by replacing the right-hand side of (8) by its empirical counterpart based on the observed Y-sample. Moreover, the proof relies explicitly on the fact that the φ_j's are not compactly supported. This is due to the integrations by parts used to obtain the result.

Proposition 2.3 hereafter gives another way of expressing the coefficients and is helpful for studying the rates of estimators. Define the matrices $H_m^{(k)}$ with size $m \times (m+k)$ by

$$H_m^{(0)} = Id_m,$$

and for $k \geq 1$,

$$[H_m^{(k)}]_{j,\ell} = h_\ell^{j,k} \;\text{ for } \ell = \sup((j-k), 1), \ldots, j+k, \qquad [H_m^{(k)}]_{j,\ell} = 0 \quad \text{otherwise},$$

where

$$h_\ell^{j,k} = \sum_{p=|\ell-j|}^{k} b_\ell^{j,p} \binom{k}{p} \frac{1}{p!}, \tag{9}$$

and the $(b_\ell^{j,p})$'s can be recursively computed by

$$b_\ell^{j,0} = \delta_{\ell,j}, \qquad b_\ell^{j,p+1} = -\frac{\ell+1}{2} b_{\ell+1}^{j,p} - (p+\frac{1}{2}) b_\ell^{j,p} + \frac{\ell}{2} b_{\ell-1}^{j,p} \quad \text{for } p \geq 0.$$

Proposition 2.3. *By convention, we set $\varphi_j = 0$ if $j \leq -1$ and define the column vectors of coefficients of f on $(\varphi_0, \ldots, \varphi_{m-1})$ and of $f_{k,Y}$ on $(\varphi_0, \ldots, \varphi_{m+k-1})$:*

$$\vec{a}_{m-1}(f) := (a_j(f))_{0 \leq j \leq m-1}, \quad \vec{a}_{m+k-1}(f_{k,Y}) = (a_j(f_{k,Y}))_{0 \leq j \leq m+k-1}.$$

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

$$\forall \ell \leq j+k, \quad |h_\ell^{j,k}| \leq C'_k (j+k)^k. \tag{10}$$

For each k, the coefficients have to be computed. In our simulations (Section 4), we use the two values $k = 1, 2$ and the coefficients are the following. For $k = 1$, $[H_m^{(1)}]_{j,\ell} = 0$ if $\ell \neq j, j-1, j+1$,

$$[H_m^{(1)}]_{j,j-1} = -\frac{j}{2}, \quad [H_m^{(1)}]_{j,j} = \frac{1}{2}, \quad [H_m^{(1)}]_{j,j+1} = \frac{j+1}{2}. \tag{11}$$

For $k = 2$, $[H_m^{(2)}]_{j,\ell} = 0$ if $\ell \neq j, j-1, j+1, j-2, j+2$ and

$$[H_m^{(2)}]_{j,j-2} = \frac{j(j-1)}{8}, \quad [H_m^{(2)}]_{j,j-1} = -\frac{1}{2} j, \quad [H_m^{(2)}]_{j,j} = -\frac{j^2+j-1}{4},$$

$$[H_m^{(2)}]_{j,j+1} = \frac{1}{2}(j+1), \quad [H_m^{(2)}]_{j,j+2} = \frac{(j+1)(j+2)}{8}.$$

For the study of the risk bounds, we need evaluate two norms of the matrix $H_m^{(k)}$. The first one is the spectral radius $\rho(H_m^{(k)})$ and the second one is the Frobenius norm $|H_m^{(k)}|_F$. We recall their definitions. The squared spectral radius of the matrix A, $\rho^2(A) = \lambda_{max}(A^t\, A)$, is equal to the largest eigenvalue of $A^t\, A$, where A^t denotes the transpose of A. The Frobenius squared norm of A is given by $|A|_F^2 = \mathrm{Tr}(A^t\, A)$ where $\mathrm{Tr}(M)$ is the trace of matrix M. The following result is deduced from Proposition 2.3.

Corollary 2.1. *For $m \geq 1$ and $k \geq 0$, there exist constants $c(k), C(k)$ depending on k only, such that*

$$c(k)m^{2k+1} \leq |H_m^{(k)}|_F^2 \leq m\rho^2(H_m^{(k)}) \leq C(k)m(m+k)^{2k}.$$

2.3. **Projection estimator and upper risk bound.** Proposition 2.3 leads us to define a collection of projection estimators of f by:

$$\hat{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \vec{\hat{a}}_{m-1} = (\hat{a}_j)_{0\leq j\leq m-1} = H_m^{(k)} \vec{\hat{a}}_{m+k-1}(Y), \qquad m \geq 1 \tag{12}$$

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

$$\hat{a}_j(Y) := \frac{1}{n}\sum_{i=1}^{n} \varphi_j(Y_i). \tag{13}$$

Note that, by Proposition 2.2, we have the other formula: $\hat{a}_j = (1/n)\sum_{i=1}^n \frac{1}{k!}(y^k\varphi_j)^{(k)}(Y_i)$. The following proposition gives the risk bound for the estimator $\hat{f}_m$.

Proposition 2.4. *Let $\hat{f}_m, f_m$ be given by (12) and (3). Then we have, for all $k, m \geq 0$,*

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \leq \|f - f_m\|^2 + 2\frac{[(m+k)\rho^2(H_m^{(k)})] \wedge [\|f_{k,Y}\|_\infty |H_m^{(k)}|_F^2]}{n},$$

where $x \wedge y = \inf(x, y)$. Moreover, it holds

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

with $\zeta_k = 2[(2k+1)C'_k]^2$ where C'_k is the constant in Proposition 2.2, formula (10).

Let us discuss the two terms in the infimum appearing in the first bound of Proposition 2.4. In light of Corollary 2.1, we have $|H_m^{(k)}|_F^2 \leq m\rho^2(H_m^{(k)})$ but $\|f_{k,Y}\|_\infty$ may be infinite. On the other hand, the two terms have the same order, given in second inequality of Proposition 2.4.

2.4. **Adaptive estimation.** The risk bound decomposition of Proposition 2.4 classically involves a squared 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^{2k+1}/n which is increasing with m. Therefore, to select relevantly m, we have to perform a compromise. This can be done asymptotically by evaluating rates of convergence (see below), or, as we do now, on finite sample by a model section strategy. In view of the discussion on the risk bound, we define for $k \geq 0$,

$$\hat{m}_k = \arg\min_{m\in\mathcal{M}_n^{(k)}} \left(-\|\hat{f}_m\|^2 + \mathrm{pen}_k(m)\right), \quad \mathrm{pen}_k(m) = \kappa\frac{m\rho^2(H_m^{(k)})}{n}, \tag{14}$$

where

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

Note that the definition of $\hat{m}_k$ mimicks the squared-bias variance compromise as $-\|\hat{f}_m\|^2$ is an estimator of $-\|f_m\|^2$ which is, up to the constant $\|f\|^2$, equal to $\|f - f_m\|^2$ and $\mathrm{pen}_k(m)$ is proportional to the variance term. As $H_m^{(k)}$ is explicit, the computation of $\rho^2(H_m^{(k)})$ is obtained by a numerical algorithm (function `eigs` applied to $(H_m^{(k)})^t\ (H_m^{(k)})$ in Matlab).

Theorem 2.1. *Assume that $\mathbb{E}(1/X) < +\infty$. Let $\hat{f}_m$ be given by (12) and $\hat{m}_k$ by (14). There exists a constant κ_0 such that for any $\kappa \geq \kappa_0$, we have*

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

where C_1 is a numerical constant ($C_1 = 4$ suits) and C_2 depends on k and $\mathbb{E}(1/X)$.

It follows from Theorem 2.1 that the estimator $\hat{f}_{\hat{m}_k}$ is adaptive in the sense that its risk automatically realizes the squared bias-variance compromise.

The constant κ_0 provided by the proof is generally not optimal; finding the optimal theoretical value of κ in the penalty is far from easy (see for instance Birgé and Massart (2007) in a Gaussian regression model). This is why it is standard to calibrate the value κ in the penalty by preliminary simulations.

3. Rates of convergence on Sobolev-Laguerre balls

We now study the asymptotic point of view to find the dimension m_{opt} which realizes the bias variance compromise of the risk bound given in Proposition 2.4. We have already identified the rate of the variance term as m^{2k+1}/n. We now look at the bias term $\|f - f_m\|^2$. Classically, the bias rate is evaluated by choosing a regularity space for the function f. Sobolev-Laguerre spaces are well fitted to our framework.

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

$$W^s = \{h : (0, +\infty) \to \mathbb{R}, h \in \mathbb{L}^2((0, +\infty)), |h|_s^2 := \sum_{k \geq 0} k^s a_k^2(h) < +\infty\}. \tag{15}$$

where $a_k(h) = \int_0^{+\infty} h(u)\varphi_k(u)du$. For s integer, the property $|h|_s^2 < +\infty$ can be linked with regularity properties of the function h. We give details in the Appendix. We define the ball $W^s(D)$:

$$W^s(D) = \left\{h \in W^s, |h|_s^2 \leq D\right\}.$$

3.2. **Upper rates.** We can deduce from Proposition 2.4 the rates of convergence of the estimator on Sobolev-Laguerre balls. For $f \in W^s(D)$, we have $\|f - f_m\|^2 = \sum_{j \geq m} a_j^2(f) \leq Dm^{-s}$. This yields:

Corollary 3.1. *Assume that $f \in W^s(D)$. Let $\hat{f}_m$ be given by (12). Then choosing $m_{opt} = [n^{s+2k+1}]$ gives*

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

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

The rate may be interpreted as follows: we have an inverse problem, where s measures the smoothness and k the ill-posedeness.

For direct observations of $X_1, \ldots, X_n$ $(k = 0)$, this rate is the same as the one obtained by Juditsky and Lambert-Lacroix (2004) for estimation of a density on $\mathbb{R}$, over Hölder classes of densities.

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 $a_k(f) = \sqrt{2}[\theta/(\theta + 1)]\left((\theta - 1)/(\theta + 1)\right)^k$ 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 $[\log(n)]^{2k+1}/n$ for $m_{opt} = \log(n)/\rho$, $\rho = |\log[|(\theta - 1)/(\theta + 1)|]|$. The result can be extended to Gamma and mixed Gamma densities, see Comte and Genon-Catalot (2015), Mabon (2015). Thus, the Laguerre basis method provides excellent rates for the class of mixed gamma densities.

Nevertheless, we stress that the adaptive procedure does not require any knowledge on the rate of the bias and still automatically realizes the finite sample bias-variance compromise and also *automatically* reaches the best possible asymptotic rate.

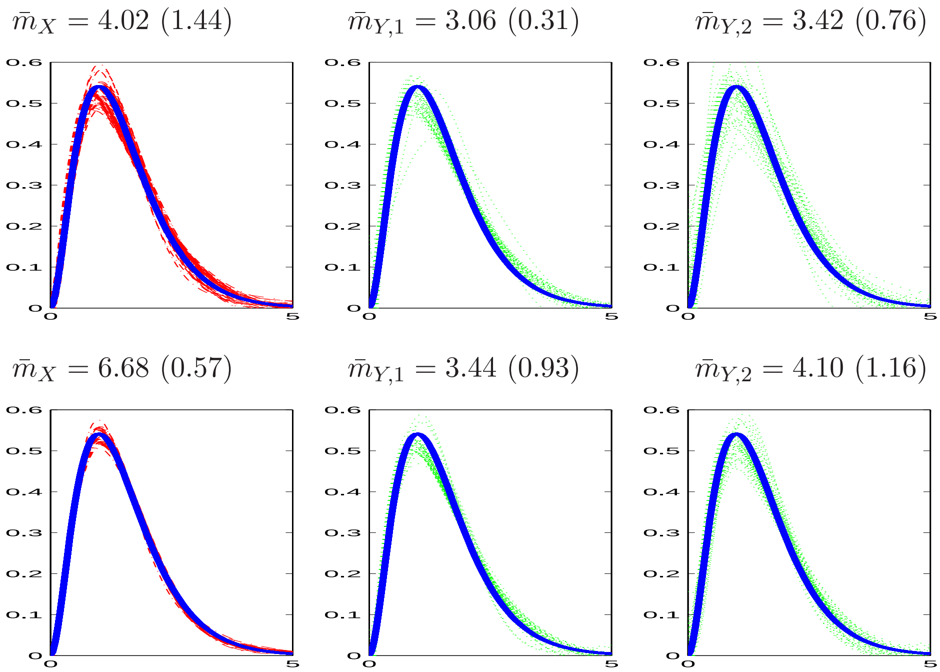

FIGURE 1. True density f of Model (i) (Gamma distribution) in bold (blue). 50 estimators of f, left: from direct observation of X in dotted (red); middle: from observation of $Y = XU$ with $U \sim \mathcal{U}([0,1])$, in dotted (green); right: from observation of $Y = XU$ with $U \sim \beta(1,2)$, in dotted (green). First line: $n = 400$. Second line: $n = 2000$. Above each plot, $\bar{m}_X$ (resp. $\bar{m}_{Y,1}$, resp $\bar{m}_{Y,2}$) is the mean of the selected dimensions from X (resp. from Y) with standard deviation in parenthesis.

So far, we have used that the φ_js are bounded. However, Szegö (1975) p.198 and p. 241. gives the following asymptotic bound: $\forall a > 0, \sup_{x>a} |\varphi_j(x)| \leq Cj^{-1/4}$. Therefore, for densities with support $[a, +\infty[$ with $a > 0$, we have $\sum_{j_0 \leq j \leq m} \mathbb{E}(\varphi_j^2(Y_1)) \leq C' m^{1/2}$ and the variance term of $\hat{f}_m$ has order $O(m^{2k+1/2}/n)$ instead of $O(m^{2k+1}/n)$. By choosing $m_{opt} = [n^{s+2k+(1/2)}]$, the upper rate becomes on this restricted class of densities, of order $O(n^{-s/(s+2k+(1/2))})$. Lower bounds for this class would require a completely different proof.

3.3. **Lower bounds.** We prove that the upper rate obtained in Proposition 3.1 is optimal on Sobolev-Laguerre balls. This reveals unexpectedly difficult. We first treat the case $k = 0$ ($U_i = 1$, direct observations of X_i). Then, we deal with $k = 1$ and give indications on how to extend the result to all $k > 1$. The upper bound matches the lower bound up to a logarithmic term.

Theorem 3.1. *Assume that s is an integer, $s > 1$ and $X_1, \ldots, X_n$ are observed. Then for any estimator $\hat{f}_n$, 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 is in particular nonnegative.

Next 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, see Remark 6.1 in the proof. However, given technicalities of the proof, we detail only the case $k = 1$.

Theorem 3.2. *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,

$$\sup_{f \in W^s(D)} \mathbb{E}_f \left[\|\hat{f}_n - f\|^2 \right] \gtrsim \tilde{\psi}_n, \quad \tilde{\psi}_n = n^{-s/(s+3)} (\log n)^{(1+\epsilon)/(1+s/3)}.$$

4. Simulation results

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

- on direct observations $X_1, \dots, X_n$,
- on multiplicative censored observations $Y_1, \dots, Y_n$, $Y_i = X_i U_i$ with $U_i \sim \mathcal{U}([0,1])$,
- on multiplicative censored observations $Y_1, \dots, Y_n$, $Y_i = X_i U_i$ with $U_i \sim \beta(1,2)$.

We consider for f the densities

(i) Gamma$(3, 1/2)$,
(ii) a Gamma mixture: cX with $X \sim$ 0.4 Gamma(2,1/2)+0.6 Gamma(16,1/4) and $c = 5/8$.
(iii) Lognormal$(0.5, 0.5)$ (exponential of a Gaussian with mean 0.5 and variance 0.5^2).
(iv) $5X$ with $X \sim$ Beta$(4,5)$, a beta distribution.

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

After preliminary simulation experiments, direct estimation is penalized with $\kappa_1 = 0.75$. For U following a uniform distribution on $[0,1]$, we use $\kappa_2 = 0.25$ and $\kappa_3 = 0.025$ for U following a $\beta(1,2)$ distribution on $[0,1]$.

Beam of estimators are given in Figures 1-2 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 k (we have $k = 1$ when $U \sim \mathcal{U}([0,1])$ and $k = 2$ when $U \sim \beta(1,2)$) makes the problem more difficult. This is why Gamma mixtures are hard to reconstruct in presence of multiplicative censoring (see Figure 2). Selected dimensions can be of various orders (between 3 and 12 in our examples) and vary or be very stable (see the standard deviations).

	$n = 400$				$n = 2000$			
density	Kernel	Laguerre	$\mathcal{U}([0,1])$	$\beta(1,2)$	Kernel	Laguerre	$\mathcal{U}([0,1])$	$\beta(1,2)$
$\gamma(3,1/2)$	3.66	3.38	6.9	12.55	1.20	0.58	4.05	4.67
(i)	(2.19)	(1.35)	(7.77)	(14.95)	(0.47)	(0.51)	(1.28)	(3.21)
Mixed	22.25	6.17	51.07	41.54	12.00	1.82	9.47	12.62
Gamma (ii)	(2.69)	(1.98)	(16.84)	(19.50)	(1.24)	(0.69)	(4.29)	(5.95)
Lognormale	3.93	2.54	19.54	22.34	1.28	1.13	6.21	7.07
(iii)	(2.25)	(1.61)	(8.42)	(19.43)	(0.51)	(0.45)	(2.66)	(4.40)
$5\beta(4,5)$	2.51	2.06	11.87	18.52	0.71	0.67	8.21	8.40
(iv)	(1.31)	(1.64)	(8.36)	(14.49)	(0.36)	(0.46)	(2.18)	(4.56)

Table 1. MISE $\times$ 1000 with std $\times$ 1000 in parenthesis for 100 estimation of f with kernel or Laguerre projection estimators in the case of direct observation of X and with Laguerre projection in case $Y = XU$ is observed and U is $\mathcal{U}([0,1])$ or $\beta(1,2)$.

Table 1 gives the Mean Integrated Squared Error (MISE) for two sample sizes ($n = 400$ and $n = 2000$) and the three cases for the same X sample; ISE are computed on the interval of observation. The kernel estimator implemented for comparison is obtained via the function `ksdensity` of Matlab. The projection method is in general better than the kernel estimator,

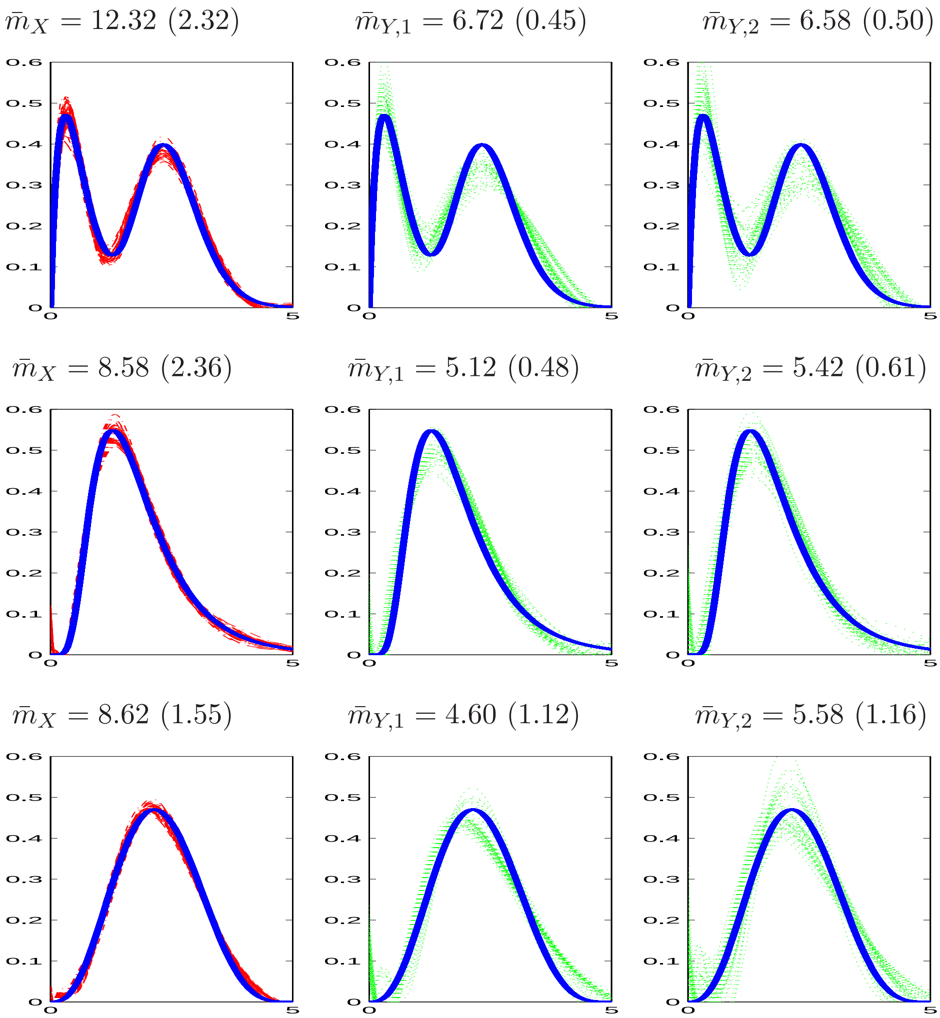

FIGURE 2. True density f of Model (ii) (Mixed Gamma) on the first line, Model (iii) (Lognormal) on the second line, and Model (iv) on the third line (Beta distribution) in bold (blue). Left: 50 estimators of f from direct observation of X in dotted (red). Middle: 50 estimators of f from observation of Y, in dotted (red) with $U \sim \mathcal{U}([0,1])$. Right: 50 estimators of f from observation of Y, in dotted (red) with $U \sim \beta(1,2)$, $n = 2000$ in all cases. Above each plot, $\bar{m}_X$ (resp. $\bar{m}_{Y,1}$, resp $\bar{m}_{Y,2}$) is the mean of the selected dimensions for X (resp. for Y) with standard deviation in parenthesis.

with slight improvement for models (iii) and (iv), and a much more important one in the Gamma and in the mixed Gamma case of models (i) and (ii). This was expected as theoretical rates are better for Gamma or mixed Gamma when using Laguerre projection method. Clearly, the inverse problem faced in the multiplicative censoring case makes the problem more difficult and the MISEs higher.

5. Extensions and concluding remarks

In this paper, we propose a nonparametric adaptive estimator of the density f of X_i in the model $Y_i = X_iU_i$ where X_i are *i.i.d.* nonnegative random variables and the sequences $(U_i)_i$ and $(X_i)_i$ are independent. We develop the case of $U_i \sim \beta(1,k)$ for $k \in \mathbb{N}$, where $k = 0$ corresponds to the direct observation of the X_i's (i.e. $U_i \equiv 1$). Using a Laguerre basis a collection of projection estimators is built and a date driven procedure is proposed to select the dimension of the projection space. The risk bound of the adaptive estimator provides a an automatic bias variance compromise which is non asymptotic. From the asymptotic point of view, we prove upper rates over Sobolev Laguerre balls. We obtain lower bounds matching with the previous rates up to a logarithmic factor, the proof of which requires specific extensions of the classical scheme.

The method can be extended to other noise distributions. As in Chesneau (2013), we can consider that $U_i = U_i^{(1)} \dots U_i^{(\ell)}$ with $U_i^{(j)}$'s *i.i.d.* and uniform. Then denoting the density of Y_i by $f_Y^{(\ell)}$, Proposition 2.3 applies and yields

$$\vec{a}_{m-1}(f) = H_m^{(1)} H_{m+1}^{(1)} \dots H_{m+\ell-1}^{(1)} \vec{a}_{m+\ell-1}(f_Y^{(\ell)}).$$

Propositions 2.4 and 2.1 can be generalized without difficulty.

Another possible extension of the noise distribution is to consider that U_i follows a $\beta(r,k)$ distribution, $r \geq 1$. Indeed, an inversion formula extending Proposition 2.1 holds. Denoting by $f_{r,k,Y}$ the density of $Y_i = X_i U_i$ with $U_i \sim \beta(r,k)$, we can prove (see Section 6) that if $\mathbb{E}(1/X^{r-1}) < +\infty$

$$f(x) = (-1)^k \frac{x^{k+r-1}}{(r+k-1)(r+k-2)\dots r} \frac{d^k}{dx^k}\left(\frac{f_{r,k,Y}(x)}{x^{r-1}}\right). \tag{16}$$

Therefore, we can obtain an analogous of Proposition 2.3 and develop a complete study.

It is worth stressing that the model $Z_i = X_i U_i + V_i$ can be treated by our approach. Indeed Laguerre functions are a convenient tool for deconvolution on $\mathbb{R}^+$, as done in Mabon (2015). Moreover we provide a precise description of the strategy in the model $Z_i = X_i U_i + V_i$ in Belomestny *et al.* (2016).

Another way of treating the subject could be to take the logarithm of (1) and estimate the density of $\log(X)$ by deconvolution (mainly Fourier methods). This method can work for a large class of noise distributions. On the other hand, the function which is estimated is $f_{\log(X)}$, the density of $\log(X)$. The relation $f_X(x) = f_{\log(X)}(\log(x))/x$ implies that the estimator is not defined in 0 and the integrated risk has to be computed on $[a, +\infty[$, with $a > 0$. This is a significant drawback and justifies the use of the Laguerre strategy.

6. Proofs

6.1. **Proof of Propositions 2.2 and 2.3 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 (8) 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

$$L_j'(x) = -L_{j-1}^{(1)}(x). \tag{17}$$

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

$$xL_j^{(1)}(x) = (j+1)[L_j(x) - L_{j+1}(x)], \tag{18}$$

and Formula (22.7.12) therein

$$xL_j(x) = -(j+1)L_{j+1}(x) + (2j+1)L_j(x) - jL_{j-1}(x). \tag{19}$$

We have to compute $2yL_j'(2y) - yL_j(2y)$ or $tL_j'(t) - \frac{t}{2}L_j(t)$. Combining relations (17)-(19), 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

$$b_{j-1}^{j,1} = -\frac{j}{2}, \quad b_j^{j,1} = -\frac{1}{2}, \quad b_{j+1}^{j,1} = \frac{j+1}{2}. \tag{20}$$

Finally,

$$(y\varphi_j)' = \varphi_j(y) + y\varphi_j'(y) = -\frac{j}{2}\varphi_{j-1}(y) + \frac{1}{2}\varphi_j(y) + \frac{j+1}{2}\varphi_{j+1}(y). \tag{21}$$

This gives the result for $k = 1$. □

6.2. Proof of Proposition 2.2 for $k \geq 2$.

Let $f_k = f_{k,Y}$. Using (6), 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.*:

$$\left[\sum_{\ell=1}^{k} f_k^{(k-\ell)}(y)(y^k\varphi_j(y))^{(\ell-1)}(-1)^{\ell-1} \right]_0^{+\infty} = 0 \tag{22}$$

Therefore, we obtain Formula (8) after proving that (22) holds.

Proof of (22): 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 (7), $\Sigma_0(y) = (-1)^k k!(F(y) - F_{k,Y}(y))$. 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 $f_k^{(p)}(y)y^{p+2}, p = 0, \dots, k-2$ tends to 0 at both 0 and $+\infty$. Indeed,

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

$$|f_k^{(p)}(y)y^{p+2}| \lesssim \int_y^{+\infty} \frac{y^{p+2}}{u^{p+1}} f(u)du \leq y \int_y^{+\infty} f(u)du \tag{23}$$

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

$$|f_k^{(p)}(y)y^{p+2}| \lesssim \int_y^{+\infty} \frac{y^{p+2}}{u^{p+1}} f(u)du \leq \int_y^{+\infty} uf(u)du \tag{24}$$

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

$$|f_k^{(p)}(y)y^{p+t+1}| \lesssim \int_y^{+\infty} \frac{y^{p+t+1}}{u^{p+1}} f(u)du \leq \int_y^{+\infty} u^t f(u)du \tag{25}$$

which tends to 0 at $+\infty$ by the moment assumption $\mathbb{E}(X^t) < +\infty$. The proof of (22) is complete.□

6.3. **Proof of Proposition 2.3.** 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 $\ell = 0, 1, \dots, j+k$:

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

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

$$\frac{1}{k!}(y^k\varphi_j)^{(k)} = \sum_{p=0}^{k} \binom{k}{p} \frac{1}{p!} y^p \varphi_j^{(p)}. \tag{26}$$

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

Lemma 6.1. *We have*

$$y^p\varphi_j^{(p)}(y) = \sum_{\ell=0\vee(j-p)}^{j+p} b_\ell^{j,p} \varphi_\ell(y), \tag{27}$$

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

$$b_\ell^{j,p+1} = -\frac{\ell+1}{2} b_{\ell+1}^{j,p} - (p+\frac{1}{2}) b_\ell^{j,p} + \frac{\ell}{2} b_{\ell-1}^{j,p}. \tag{28}$$

Moreover

$$\forall \ell \le j+p,\ |b_\ell^{j,p}| \le C_p (j+p)^p. \tag{29}$$

Applying Lemma 6.1, and interchanging sums in (26) yields formula (9). Next, we use Formula (29) to get

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

This gives the bound (10).

Proof of Lemma 6.1 Initialization of (27) for $p=0$ is obvious. Formula (20) shows that the induction formula (28) holds for $p=0$ ($p=0$ to $p=1$).

Next, we differentiate (27) 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 (20), we get

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

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 (28). Inequality (29) is obtained by straightforward induction. The proof of Proposition 2.2 is now complete. □

6.4. **Proof of Corollary 2.1.** The general inequality $|H_m^{(k)}|_F^2 \le m\rho^2(H_m^{(k)})$ holds for all $m\times(m+k)$ matrices. For $k=0$, $H_m^{(0)} = I_m$ the $m\times m$ identity matrix, and the two above terms are equal to m. First we prove the upper bound for $\rho^2(H_m^{(k)})$, $k\ge 1$.

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

We consider first $m\ge k$ and use (10) to get

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

Next write that

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

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

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

If $m < k$, the bound obviously holds.

Now we prove the lower bound on $|H_m^{(k)}|_F^2$. First

$$|H_m^{(k)}|_F^2 = \sum_{j=1}^{m} \sum_{\ell=(j-k)^+\vee 1}^{j+k} [H_m^{(k)}]_{j,\ell}^2 \geq \sum_{j=1}^{m} [H_m^{(k)}]_{j,j+k}^2.$$

Now, Proposition 2.2 yields $[H_m^{(k)}]_{j,j+k} = h_{j+k}^{j,k} = b_{j+k}^{j,k}/k!$ and $b_{j+k}^{j,k} = ((j+k)/2)b_{j+k-1}^{j,k}$. Indeed, coefficients $b_\ell^{j,p}$ are zero if $\ell > j+p$ (see formula (27)). Therefore, as $h_{j+1}^{j,1} = (j+1)/2$, we get, by elementary induction that

$$h_{j+k}^{j,k} = \frac{1}{k!}\frac{(j+1)(j+2)\dots(j+k)}{2^k}.$$

We obtain

$$\begin{aligned} |H_m^{(k)}|_F^2 &\geq \sum_{j=1}^{m}\left(\frac{1}{k!}\frac{(j+1)(j+2)\dots(j+k)}{2^k}\right)^2 \geq \frac{1}{(k!2^k)^2}\sum_{j=1}^{m}(j+1)^{2k} \\ &\geq \frac{1}{(k!2^k)^2}\int_1^m x^{2k}dx = \frac{m^{2k+1}}{(2k+1)(k!2^k)^2}, \end{aligned}$$

which ends the proof. □

6.5. **Proof of Proposition 2.4.** The risk bound of the estimator can be written as follows

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

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, see (12),

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

So,

$$\begin{aligned} \mathbb{E}(\|\hat{f}_m - f_m\|^2) &\leq \rho^2(H_m^{(k)})\mathbb{E}(\|\vec{a}(Y)_{m+k-1} - \mathbb{E}(\vec{a}(Y)_{m+k-1})\|^2) \\ &\leq \rho^2(H_m^{(k)}) \sum_{j=0}^{m+k-1} \text{Var}(\hat{a}_j(Y)) = \frac{1}{n}\rho^2(H_m^{(k)}) \sum_{j=0}^{m+k-1} \text{Var}(\varphi_j(Y_1)) \\ &\leq \frac{1}{n}\rho^2(H_m^{(k)}) \sum_{j=0}^{m+k-1} \mathbb{E}(\varphi_j^2(Y_1)) \\ &\leq \frac{2(m+k)\rho^2(H_m^{(k)})}{n}, \end{aligned}$$

as $\sum_{j=0}^{m+k-1}\varphi_j^2(x)\le 2(m+k)$, $\forall x\in\mathbb{R}^+$. This gives a first bound. For the second one, we can write, if $\|f_Y\|_\infty<+\infty$,

$$\mathbb{E}(\|\hat f_m-f_m\|^2)=\mathbb{E}\left(\sum_\ell\left[H_m^{(k)}(\vec{\hat a}(Y)_{m+k-1}-\mathbb{E}(\vec{\hat a}(Y)_{m+k-1}))\right]_\ell^2\right)$$

$$\begin{aligned}
&= \mathbb{E}\left[\sum_\ell\left(\sum_j[H_m^{(k)}]_{\ell,j}[\vec{\hat a}(Y)_{m+k-1}-\mathbb{E}(\vec{\hat a}(Y)_{m+k-1}))]_j\right)^2\right]\\
&= \sum_\ell\sum_{j,j'}[H_m^{(k)}]_{\ell,j}[H_m^{(k)}]_{\ell,j'}\mathrm{cov}(\hat a_j(Y),\hat a_{j'}(Y))=\frac1n\sum_\ell\sum_{j,j'}[H_m^{(k)}]_{\ell,j}[H_m^{(k)}]_{\ell,j'}\mathrm{cov}(\varphi_j(Y_1),\varphi_{j'}(Y_1))\\
&= \frac1n\sum_\ell\mathrm{Var}\left(\sum_j[H_m^{(k)}]_{\ell,j}\varphi_j(H_1)\right)\le\frac1n\sum_\ell\mathbb{E}\left[\left(\sum_j[H_m^{(k)}]_{\ell,j}\varphi_j(Y_1)\right)^2\right]\\
&\le \frac{\|f_Y\|_\infty}{n}\sum_\ell\int\left(\sum_j[H_m^{(k)}]_{\ell,j}\varphi_j(y)\right)^2dy=\frac{\|f_Y\|_\infty}{n}\sum_\ell\sum_j[H_m^{(k)}]_{\ell,j}^2=\frac{\|f_Y\|_\infty}{n}|H_m^{(k)}|_F^2,
\end{aligned}$$

which gives the second part of the bound.

It follows from Corollary 2.1 that $m\rho^2(H_m^{(k)})$ and $|H_m^{(k)}|_F^2$ are both of orders m^{2k+1}, but the second bound involves $\|f_Y\|_\infty$. This term is unknown, difficult to estimate and additional assumption is required to ensure its finiteness, for instance $\mathbb{E}(1/X)<+\infty$ for $k=1$. □

6.6. **Proof of Theorem 2.1.** In the proof, we omit the index k in $\mathcal{M}_n^{(k)}$, $\mathrm{pen}_k(m)$ and $\hat m_k$.

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

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

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 form of the matrices $H_m^{(k)}$ (band), we have, for $\vec t\in S_m$,

$$\langle\vec t,H_M^{(k)}\vec{\hat a}_{M+k-1}(Y)\rangle_M=\langle\vec t_m,H_m^{(k)}\vec{\hat a}_{m+k-1}(Y)\rangle_m=\langle\vec t_m,\vec{\hat a}_{m-1}\rangle_m.$$

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

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

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

$$\begin{aligned}
\gamma_n(\vec t)-\gamma_n(\vec s) &= \|\vec t-\vec f_M\|_M^2-\|\vec s-\vec f_M\|_M^2-2\langle\vec t-\vec s,H_M^{(k)}\vec{\hat a}_{M+k-1}(Y)-\vec f_M\rangle_M\\
&= \|\vec t-\vec f_M\|_M^2-\|\vec s-\vec f_M\|_M^2-2\langle\vec t-\vec s,H_M^{(k)}(\vec{\hat a}_{M+k-1}(Y)-\vec a_{M+k-1}(f_Y)\rangle_M
\end{aligned}$$

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

$$\nu_n(\vec t)=\langle\vec t,H_M^{(k)}(\vec{\hat a}_{M+k-1}(Y)-\vec a_{M+k-1}(f_Y)\rangle_M,$$

and note that

$$\|\hat f_m-f\|^2=\|\vec{\hat f}_{(m)}-\vec f_M\|_M^2+\sum_{j=M}^\infty a_j^2(f),\quad \|f_m-f\|^2=\|\vec f_m-\vec f_M\|_M^2+\sum_{j=M}^\infty a_j^2(f).\tag{30}$$

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

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

which writes

$$\|\vec{\hat{f}}_{(\hat{m})} - \vec{f}_M\|_M^2 \leq \|\vec{f}_m - \vec{f}_M\|_M^2 + \mathrm{pen}(m) + 2\nu_n(\vec{\hat{f}}_{(\hat{m})} - \vec{f}_m) - \mathrm{pen}(\hat{m}). \tag{31}$$

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

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

We get by plugging this in (31),

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

Let $p(m, m')$ be such that $4p(m, m') \leq \mathrm{pen}(m) + \mathrm{pen}(m')$ and use (30), to get

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

Now, we have

$$\mathbb{E}\left[\left(\sup_{\vec{t}\in B(\hat{m},m)} \nu_n^2(\vec{t}) - p(m, \hat{m}) \right)_+ \right] \leq \frac{c}{n}, \tag{32}$$

where c depends on k and $\|f_Y\|_\infty = \mathbb{E}(1/X)$. The proof of (32) follows the line of the proof of Proposition 7.1 in Mabon (2015) and delivers the value of κ_0. Thus, we obtain the result announced in Theorem 2.1.

6.7. **Proof of Theorem 3.1.**

6.7.1. *Main steps.* 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 $\beta < 1$,

$$\frac{c_{\alpha,\beta}}{(e+x)^{\alpha+1}} \leq f_0(x) \leq \frac{c_{\alpha,\beta}}{(e+x)^\alpha}. \tag{33}$$

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

Lemma 6.3. *Suppose that* $\sum_{k=1}^{K}\theta_k(-1)^k = 0$ *and all partial sums* $\sum_{k=1}^{p}\theta_k(-1)^k$, $p = 1,\dots,K$, *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 $\boldsymbol\theta, \boldsymbol\theta' \in \{0,1\}^K$,

$$\int_0^\infty \left(f_{\boldsymbol\theta}(x) - f_{\boldsymbol\theta'}(x)\right)^2 dx = \delta^2 \sum_{k=1}^{K}\left(\theta_k - \theta'_k\right)^2 = \delta^2\rho(\boldsymbol\theta,\boldsymbol\theta'), \tag{34}$$

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 6.4. *Fix some even integer* $K > 0$. *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,

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

for all $0 \leq j < l \leq M$. *Moreover it holds that, for some constant* $A_2 > 0$,

$$M \geq 2^{A_2 K}. \tag{35}$$

Then we have the following Lemma.

Lemma 6.5.

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

Now we are in position to end the proof of Theorem 3.3. Under the choices

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

using inequality (35), $K \leq \log M/(A_2\log 2)$, 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

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

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 3.1. □

Proof of Lemma 6.2. We have

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

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

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

It follows that

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

where $A(s)$ is defined by (49). We take $D/4 \geq \tilde{B}(s, \alpha)$. Next

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

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 $k \neq \ell$. Therefore

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

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 6.3. 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 difficult. 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

$$\sup_{x>\mathtt{a}} \left| x^\lambda \log^\mu(x) \cdot \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \lesssim K^\lambda \log^\mu(K), \quad K \to \infty \tag{36}$$

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

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

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

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

$$\left| x^\lambda \log^\mu(x) \cdot \sum_{k=K+1}^{2K} \theta_{k-K}\varphi_k(x) \right| \le AC_{\lambda,\mu} K^\lambda \log^\mu(K), \quad K \to \infty, \tag{37}$$

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

To prove (37), we first study the case $\mu = 0$ and λ integer.

Lemma 6.6. *It holds for any integers n and $\lambda \le n$,*

$$x^\lambda L_n(x) = \sum_{k=-\lambda}^{\lambda} c_{k,n}^{(\lambda)} L_{n+k}(x), \tag{38}$$

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 $\lambda = 0$, (38) obviously holds. Suppose that it holds for some $\lambda = K$, then due to formula (19), 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 6.6. □

We deduce by induction from Lemma 6.6 the following Corollary.

Corollary 6.1. *Each coefficient $c_{k,n}^{(\lambda)}$ in (38) can be represented in the form*

$$c_{k,n}^{(\lambda)} = \sum_{r=(r_1,\dots,r_\lambda), r_i \in S_{\lambda,1/2}^{\lambda}} b_{k,r}^{(\lambda)} \prod_{i=1}^{\lambda}(n+r_i) \tag{39}$$

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

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

Lemma 6.7. *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 6.2. *Under conditions of the previous lemma, we have a representation*

$$e^{-x/2}L_n(x) = (-1)^n a_n(x), \tag{40}$$

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

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 (37) for $\mu = 0$ and λ integer. From (38), (39) and (40), we have

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

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 (37) for $\mu = 0$ and λ an integer follows from the next Lemma.

Lemma 6.8. *Let* $K_1 < K_2$ *be two integers 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 ($x > e^{\mu/\omega}$). $\square$

Step 2. Now we want to prove (37) for $x \leq b\nu$, $b < 1$, $\nu = 4K + 2$. It holds that (see Muckenhoupt (1970) p.288)

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

for $x \leq b\nu$ 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

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

it follows from the asymptotic behavior of the Bessel function J_0, that

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

provided $x\nu$ is large. Suppose that $x > 1$ and $\lambda \geq 1$, 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 $x^{\lambda-1/4} \leq K^{\lambda-3/4}$, then

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

Now, as $S_K(x)$ is a harmonic sum, its asymptotic behaviour can be analysed using the Mellin transform approach, which yields that $|S_K(x)| \leq K^{\lambda}$ for $x > K^{\frac{\lambda-3/4}{\lambda-1/4}}$. This yields (37) for $\mu = 0$ and $2x < 4K + 2$. The case $\mu \neq 0$ is here straightforward. This ends the proof of Step 2. □

Therefore (36) is proved so the proof of Lemma 6.3 is complete. □

Proof of Lemma 6.4. 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 6.5. Equality (34) and Lemma 6.4 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 (33), 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

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

Next, using the relation (19), 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 6.5. □

6.8. **Proof of Theorem 3.2.** 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 $X_1, \ldots, X_n$ and compute the corresponding densities $f_{Y,0}$ and $f_{Y,\boldsymbol{\theta}}$ of $Y_1, \ldots, Y_n$. 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

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

Then we can compute by formula (4) 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 (21), 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 $k = K, K+1, \ldots, 2K+1$,

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

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 6.9. *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 6.10. *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 $\delta = \delta' K^{-(\alpha+1)} \log^{-\beta}(K)$ 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

$$\|\tilde{f}_{\boldsymbol{\theta}} - \tilde{f}_{\boldsymbol{\theta}'}\|^2 = \delta^2 \sum_{k=K}^{2K+1} (\mu_k(\boldsymbol{\theta}) - \mu_k(\boldsymbol{\theta}'))^2 \tag{41}$$

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

$$\sum_{k=K}^{2K+1} (\mu_k(\boldsymbol{\theta}) - \mu_k(\boldsymbol{\theta}'))^2 \geq (K/2-1)^2 \rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}'), \tag{42}$$

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

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

for all $0 \leq m < l \leq M$. Moreover it holds that, for some constant $\tilde{A}_2 > 0$,

$$M \geq 2^{\tilde{A}_2 K}. \tag{43}$$

Next we prove

Lemma 6.12.

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

Now we end the proof of Theorem 3.2. We choose $\alpha=(s+1)/2$, $\delta^2=(\delta')^2K^{-2(\alpha+1)}\log^{-(1+\epsilon)}(K)$, $K=[n/\log^{1+\epsilon}(n)]^{1/[2(\alpha+1)]}$ 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 $\delta^2K^{s+3}=[\log(n)]^{-(1+\epsilon)}$ is bounded (constraint from Lemma 6.9). This ends the proof of Theorem 3.2. □

Proof of Lemma 6.9. For $\tilde f_0$ we follow the same line as in the proof of Lemma 6.2 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. for $C=C(s)$ a constant,

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

Proof of Lemma 6.10. 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 6.3 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 6.3). 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 6.11. 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 $\rho_2(\omega,\omega')=\sum_{k=0}^K\mathbf{1}_{\omega_{k+1}-\omega_k\ne\omega_{k+1}'-\omega_k'}$, 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

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

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 $\omega^{(0)} = (0, \dots, 0) \in \Omega$, $\mathrm{Card}(\{(\omega_k) \in \Omega, \rho_2(\omega, \omega^{(0)}) = i\}) = \binom{K}{i}$. Let

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

Note that

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

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 6.12. The first inequality follows from Lemma 6.5 and $f_{Y,\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}$ and $f_{Y,0} = f_0$. The second inequality follows from (41), (42) and Lemma 6.11. □

Remark 6.1. *For $k > 1$, we choose $f_{k,Y,0}(x) = f_{Y,0}$ and deduce $f_{X,0}$, via formula (6). Similarly we set $f_{k,Y,\boldsymbol{\theta}} = f_{Y,0} + \delta \sum_{j=K+1}^{2K} \theta_{j-K}\varphi_j$. This leads to $f_{X,\boldsymbol{\theta}} = f_{X,0} + \delta \sum_{j=K+1-k}^{2K} \nu_j(\boldsymbol{\theta})\varphi_j$, with $\nu_j(\boldsymbol{\theta})$ to be computed. The proof can be completed analogously but with more tedious computations.*

6.9. **Proof of Formula (16).** We have

$$f_U(u) = \frac{1}{B(r,k)} u^{r-1}(1-u)^{k-1}\mathbb{1}_{[0,1]}(u) \text{ where } \frac{1}{B(r,k)} = \frac{(r+k-1)(r+k-2)\dots r}{(k-1)!},$$

and thus

$$f_{r,k,Y}(y) = \frac{y^{r-1}}{B(r,k)} \int_y^{+\infty} \left(1 - \frac{y}{v}\right)^{k-1} \frac{f(v)}{v^{r-1}} \frac{dv}{v}.$$

If we define

$$\theta_Y(y) = \frac{f_{r,k,Y}(y)}{y^{r-1}}, \quad \theta_X(x) = \frac{1}{kB(r,k)} \frac{f(x)}{x^{r-1}}$$

we have the analogous of relation (4)

$$\theta_Y(y) = k \int_y^{+\infty} \left(1 - \frac{y}{v}\right)^{k-1} \theta_X(v) \frac{dv}{v}. \tag{44}$$

Noting that $\mathbb{E}(1/U^{r-1}) = 1/(kB(r,k))$, under the assumption $\mathbb{E}(1/X^{r-1}) < +\infty$, we have $\int_0^{+\infty} \theta_Y(y)dy = \int_0^{+\infty} \theta_X(x)dx$. Therefore, relation (44) implies (see formula (6) in Proposition 2.1) $\theta_X(x) = ((-1)^k/k!)x^k\theta_Y^{(k)}(x)$. This gives Formula (16). □

7. Appendix

7.1. **Proof of (6) and (7).** For simplicity, set $f_{k,Y} = f_k$. For $k = 1$, $f_1(y) = \int_y^{+\infty} (f(u)/u) du 1_{y \geq 0}$. Derivating yields the first equality in (6). Integrating between 0 and y gives the second equality which implies:

$$\lim_{y \to 0} y\, f_1(y) = \lim_{y \to +\infty} y\, f_1(y) = 0. \tag{45}$$

To get (6), we proceed by induction and prove that, for any p such that $1 \leq p \leq k-1$,

$$\frac{d^p}{dy^p}[f_k(y)] = (-1)^p k \times \cdots \times (k-p) \sum_{j=0}^{k-1-p} \binom{k-1-p}{j} (-y)^j \int_y^{+\infty} \frac{f(u)}{u^{j+p+1}} du. \tag{46}$$

The formula is true for $p = 0$ as (4) 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,

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

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 (46) is proved for all $p = 0, \ldots k-1$. Taking $p = k-1$ and derivating once more gives Formula (6).

To obtain (7), we integrate (6) between 0 and y. The successive integrations by part give the result provided that, for $\ell = 0, \ldots, k$,

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

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 $y f_1(y)$ and tends to 0 as y tends to 0 by (45).□

7.2. **Norms in Sobolev-Laguerre spaces.** For $s \geq 0$, the Sobolev-Laguerre space with index s (see Bongioanni and Torrea (2007)) is defined in (15). The following results have been proved in Section 7 of Comte and Genon-Catalot (2015). For s integer, if $h : (0, +\infty) \to \mathbb{R}$ belongs to $\mathbb{L}^2((0, +\infty))$,

$$|h|_s^2 := \sum_{k \geq 0} k^s a_k^2(h) < +\infty. \tag{47}$$

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

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

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

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

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

$$\|h\|_s = \|x^{s/2} \sum_{j=0}^{s} \binom{s}{j} h^{(j)}\| = [\sum_{k \geq s} k(k-1)\dots(k-s+1)a_k^2(h)]^{1/2}, \quad |||h|||_s^2 := \sum_{j=0}^{s} \|h\|_j^2. \tag{48}$$

Then the following property holds.

Lemma 7.1. *When s is integer, the two norms $|||h|||_s$ and $|h|_s$ are equivalent.*

Proof of lemma 7.1. Obviously $|h|_0 = |||h|||_0$ and $\|h\|_j^2 \leq |h|_j^2$ for all j. Moreover $j \mapsto |h|_j$ is increasing. Therefore $|||h|||_s^2 \leq (s+1)|h|_s^2$. 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 $|h|_s^2 = \sum_{j=1}^{s} b_{j,s}\|h\|_j^2 \leq A(s)|||h|||_s^2$, with

$$A(s) = \max(|b_{j,s}|, j = 1, \dots, s). \quad \square \tag{49}$$

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., Comte, F. and Genon-Catalot, V. (2016) Laguerre estimation for k-monotone densities observed with noise. Preprint MAP5 2016-01, https://hal.archives-ouvertes.fr/hal-01252143/

[9] Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33-73.

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

[11] 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.

[12] Chee, C.-S. and Wang, Y. (2014) Least squares estimation of a k-monotone density function. *Comput. Statist. Data Anal.* **74**, 209-216.
[13] Chesneau, C. (2013) Wavelet estimation of a density in a GARCH-type model. *Comm. Statist. Theory Methods* **42**, 98-117.
[14] 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.
[15] Comte, F. and Genon-Catalot, V. (2015) Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.* **9**, 1113-1149.
[16] van Es, B., Klaassen, C. A. J. and Oudshoorn, K. Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). *J. Statist. Plann. Inference* **91**, 295-312, 2000.
[17] van Es, B., Spreij, P. and van Zanten, H. (2003). Nonparametric volatility density estimation. *Bernoulli* **9**, 451-465.
[18] van Es, B., Jongbloed, G. and van Zuijlen, M. (1998) Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.* **26**, 2395-2406.
[19] Jirak, M., Meister, A. and Reiss, M. (2014) Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42**, 1970-2002.
[20] Juditsky, A., Lambert-Lacroix, S. (2004) On minimax density estimation on R. *Bernoulli* **10**, 187-220.
[21] Jongbloed, G. (1998) Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica* **52**, 6-17.
[22] Mabon, G. (2015). Adaptive deconvolution on the nonnegative real line. Preprint HAL hal-01076927, version 2.
[23] Magnus, J. R. and Neudecker, H. (1999) *Matrix differential calculus with applications in statistics and econometrics.* Revised reprint of the 1988 original. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
[24] Shen, J. (2000). Stable and efficient spectral methods in unbounded domains using Laguerre functions. *SIAM J. Numer. Anal.* **38**, 1113-1133.
[25] Szegö, G. (1975) *Orthogonal polynomials.* Fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII. American Mathematical Society, Providence, R.I.
[26] Tsybakov, Alexandre B. (2009) Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009.
[27] Vardi, Y. (1989) Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76**, 751-761.
[28] Vardi, Y. and Zhang, C.-H. (1992) Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.
[29] Williamson R. E. (1956), Multiply monotone functions and their Laplace transforms, *Duke Mathematical Journal* **23**, 189-207.