Adaptive Laguerre density estimation for mixed Poisson models

Fabienne Comte, Valentine Genon-Catalot

ADAPTIVE LAGUERRE DENSITY ESTIMATION FOR MIXED POISSON MODELS.

F. COMTE[1] AND V. GENON-CATALOT[1]

Abstract. In this paper, we consider the observation of n i.i.d. mixed Poisson processes with random intensity having an unknown density f on $\mathbb{R}^+$. For fixed observation time T, we propose a nonparametric adaptive strategy to estimate f. We use an appropriate Laguerre basis to build adaptive projection estimators. Non-asymptotic upper bounds of the $\mathbb{L}^2$-integrated risk are obtained and a lower bound is provided, which proves the optimality of the estimator. For large T, the variance of the previous method increases, therefore we propose another adaptive strategy. The procedures are illustrated on simulated data. March 13, 2014

Keywords. Adaptive estimators. Inverse problem. Laguerre basis. Nonparametric estimation. Poisson mixture.
AMS Classification. 62G07 - 62C20.

1. Introduction

Consider n independent Poisson processes $(N_j(t), j = 1, \ldots, n)$ with unit intensity and n *i.i.d.* positive random variables $(C_j, j = 1, \ldots, n)$. Assume that the processes $(N_j(t), j = 1, \ldots, n)$ and the sequence $(C_j, j = 1, \ldots, n)$ are independent. Under these assumptions, the random time changed processes $(X_j(t) = N_j(C_j t), t \geq 0)$ are *i.i.d.* and such that the conditional distribution of X_j given $C_j = c$ is the distribution of a time-homogeneous Poisson process with intensity c. The process X_j is known as a mixed Poisson process (see *e.g.* Grandell (1997), Mikosch (2009)). Such processes are of common use in non-life insurance mathematics as well as in numerous other areas of applications (see Fabio *et al.* and references therein).

In this paper, we assume that the random variables C_j have an unknown density f on $(0, +\infty)$ and our concern is the nonparametric estimation of f from the observation of a n-sample $(X_j(T), j = 1, \ldots, n)$ for a given value T. We investigate this subject for large n and both for fixed T and large T with two different methods. The fixed T method performs well for small T (*e.g.* $T = 1$) and deteriorates as T increases while the large T method performs better and better as T increases. Thus, the two methods are complementary.

In Section 2, we consider the case $T = 1$. The distribution of $X_j(1) = N_j(C_j)$ is given by:

$$\mathbb{P}(N_j(C_j) = \ell) := \alpha_\ell(f) = \frac{1}{\ell!} \int_0^{+\infty} e^{-c} c^\ell f(c) dc, \quad \ell \geq 0, \tag{1}$$

which can be estimated by:

$$\hat{\alpha}_\ell = \frac{1}{n} \sum_{j=1}^{n} 1_{(N_j(C_j)=\ell)}, \quad \ell \geq 0. \tag{2}$$

The problem of estimating f from the discrete observations $(N_j(C_j), j = 1, \ldots, n)$ is thus an inverse problem, the problem of estimating a mixing density in a Poisson mixture. Several authors have considered this topic whether by kernel or projection methods, see Simar (1976),

[1] MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité.

Karr (1984), Zhang (1995), Loh and Zhang (1996, 1997), Hengartner (1997). These authors are mainly interested in estimating f on a compact subset of $(0,+\infty)$. We discuss with more details the links between the present results and the previous references in subsection 2.4.

In this paper, we assume that

(H) $f \in \mathbb{L}^2((0,+\infty))$

and propose a solution without any constraint on the support of the unknown function. We study the $\mathbb{L}^2((0,+\infty))$-risk and prove upper and lower bounds on an adequate function space. Our approach is a penalized projection method (see Massart (1997)) which provides a concrete adaptive estimator of f easily implementable. It is based on the following idea. By relations (1), $\alpha_\ell(f)$ is the $\mathbb{L}^2$ scalar product of f and the function $c \to e^{-c}c^\ell/\ell!$. Choosing an orthonormal basis (φ_k) of $\mathbb{L}^2((0,+\infty))$, (1) can be written as:

$$\alpha_\ell(f) = \sum_{k\geq 0} \theta_k(f)\Omega_k^{(\ell)}$$

where $\theta_k(f), \Omega_k^{(\ell)}$ are respectively the k-th component of f and $e^{-c}c^\ell/\ell!$ on the basis. The problem is to choose a basis such that the mapping $(\theta_k(f), k \geq 0) \to (\alpha_\ell(f), \ell \geq 0)$ can be simply and explicitly inverted. Then, by plugging the estimators $\hat{\alpha}_\ell$ in the inverse mapping, we get estimators of the coefficients $\theta_k(f)$ and deduce estimators of f. An appropriate choice of (φ_k) is thus a key tool: we consider the Laguerre bases defined by $((\sqrt{a}L_k(at)e^{-at/2}, k \geq 0))$ where $(L_k(t))$ are the Laguerre polynomials. Here, the choice $a = 2$ is especially relevant. Indeed, with

$$\varphi_k(t) = \sqrt{2}L_k(2t)e^{-t}, k \geq 0, t \geq 0 \tag{3}$$

$\Omega_k^{(\ell)} = 0$ for all $k > \ell$ and the matrix $\Omega_\ell = (\Omega_k^{(i)})_{0\leq i,k\leq \ell}$ is lower triangular and explicitly invertible (Propositions 2.1 and 2.2). Therefore, the inverse problem has a solution: the linear mapping on $\mathbb{R}^{\ell+1}$

$$\vec{\alpha}_\ell = (\alpha_k(f), k = 0, \ldots, \ell)' \to \vec{\theta}_\ell = (\theta_k(f), k = 0, \ldots, \ell)' = \Omega_\ell^{-1}\vec{\alpha}_\ell. \tag{4}$$

Moreover, a crucial consistency property holds: the first $\ell - 1$ coordinates of $\vec{\alpha}_\ell$ and $\vec{\theta}_\ell$ are equal to those of $\vec{\alpha}_{\ell-1}$ and $\vec{\theta}_{\ell-1}$. Note that, in Comte *et al.* (2013), another type of inverse problem involving functions of $\mathbb{L}^2((0,+\infty))$, has been solved also using a Laguerre basis.

So, we define a collection of estimators of f by $\hat{f}_\ell = \sum_{k=0}^{\ell} \hat{\theta}_k\varphi_k$, where $(\hat{\theta}_k)$ are defined using (2) and (4). We study their $\mathbb{L}^2$-risk (Proposition 2.3). For this, we introduce appropriate regularity subspaces of $\mathbb{L}^2((0,+\infty))$, the Sobolev-Laguerre spaces with index $s > 0$. These spaces are defined in Shen (2000) and Bongioanni and Torrea (2009). We precise (see Section 7) the rate of decay of the coefficients of a function f developed in a Laguerre basis when f belongs to a Sobolev-Laguerre space with index s. This allows to evaluate the order of the bias term $\|f - f_\ell\|^2$ where $\|.\|$ denotes the $\mathbb{L}^2((0,+\infty))$-norm. Using these regularity spaces, we discuss the possible rates of convergence of the $\mathbb{L}^2$-risk of $\hat{f}_\ell$. Functions belonging to a Sobolev-Laguerre ball with index s yield rates of order $O((\log n)^{-s})$. This rate is optimal, as we prove a lower-bound result. Afterwards, we propose a data-driven choice $\hat{\ell}$ of the dimension ℓ and study the $\mathbb{L}^2$-risk of the resulting adaptive estimator (Theorem 2.2). We interpret the results in the case where the observation is $(N_j(C_jT), j = 1, \ldots, n)$. This amounts to a change of scale which multiplies the variance term of the risk by a factor T and implies a deterioration of the estimator as T increases.

Section 3 is devoted to the estimation of f for large T. Our method relies on the property that for each j, $\widehat{C}_{j,T} = N_j(C_jT)/T$ is a consistent estimator of the random variable C_j as T tends to infinity. Then, we use the i.i.d. sample $(\widehat{C}_{j,T})_{1\leq j\leq n}$ to build estimators of f. We propose

projection estimators on the Laguerre basis (3) using other estimators of the coefficients $\theta_k(f)$ together with an adaptive choice of the space dimension (Proposition 3.1, Theorem 3.1). The criterion for the model selection is non standard: it involves a penalization which is the sum of two terms, one depending on n, ℓ and the other on T, ℓ.

Section 4 gives numerical simulation results and some concluding remarks are stated in Section 5. Proofs are gathered in Section 6. In Section 7, regularity spaces associated with Laguerre bases are discussed and useful inequality is recalled in Section 8 .

2. Estimation of the mixing density for $T = 1$.

2.1. **Projection estimator.** The Laguerre polynomials given by

$$L_k(t) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{t^j}{j!}, \quad k \geq 0 \tag{5}$$

are orthonormal polynomials with respect to the weight function $w(t) = e^{-t}$ on $(0, +\infty)$, *i.e.*, for all k, k', $\int_0^{+\infty} L_k(t) L_{k'}(t) e^{-t} dt = \delta_k^{k'}$ where $\delta_k^{k'}$ is the Kronecker symbol and the sequence (L_k) is an orthonormal basis of the space $\mathbb{L}^2((0, +\infty), w)$. Consequently, for all positive a, $(\sqrt{a} L_k(at), k \geq 0)$ is an orthonormal basis of $\mathbb{L}^2((0, +\infty), w(a.))$. Equivalently, $(\sqrt{a} L_k(at)\sqrt{w(a.)}, k \geq 0)$ is an orthonormal basis of $\mathbb{L}^2((0, +\infty))$. The choice $a = 2$ is especially well fitted to our problem. By (H), f admits a development on the basis (3)

$$f = \sum_{k \geq 0} \theta_k(f)\, \varphi_k, \quad \text{where} \quad \theta_k(f) = \int_0^{+\infty} f(c) \varphi_k(c) dc. \tag{6}$$

Developing the function $c \to c^\ell e^{-c}/\ell!$ on the same basis, we get

$$\frac{1}{\ell!} c^\ell e^{-c} = \sum_{k \geq 0} \Omega_k^{(\ell)} \varphi_k(c) \text{ where } \quad \Omega_k^{(\ell)} = \frac{1}{\ell!} \int_0^{+\infty} c^\ell \sqrt{2} L_k(2c) e^{-2c} dc. \tag{7}$$

As $(\sqrt{2} L_k(2c), k \geq 0)$ are orthogonal polynomials w.r.t. the weight function $w(2c) = e^{-2c}$, $\Omega_k^{(\ell)} = 0$ for $k > \ell$ (see Section 7 for more details). Thus,

$$\frac{1}{\ell!} c^\ell e^{-c} = \sum_{k=0}^{\ell} \Omega_k^{(\ell)} \varphi_k(c) \quad \text{and} \quad \alpha_\ell(f) = \sum_{k=0}^{\ell} \theta_k(f) \Omega_k^{(\ell)}$$

The coefficients $\Omega_k^{(\ell)}$ are given in the following proposition.

Proposition 2.1. *The coefficients $\Omega_k^{(\ell)}$ defined by (7) are equal to*

$$\Omega_k^{(\ell)} = \frac{(-1)^k}{\sqrt{2}\, 2^\ell} \binom{\ell}{k} 1_{(k \leq \ell)}. \tag{8}$$

Define the vectors $\vec{\theta}_\ell = (\theta_k(f), k = 0, \ldots, \ell)' \quad \vec{\alpha}_\ell = (\alpha_k(f), k = 0, \ldots, \ell)'$ and the triangular matrix $\Omega_\ell := (\Omega_k^{(i)})_{0 \leq i, k \leq \ell}$ where the diagonal terms are $\Omega_i^{(i)} = (-1)^i/(\sqrt{2}\, 2^i)$. The matrix Ω_ℓ is therefore invertible and its inverse is explicitly computed in the following proposition.

Proposition 2.2. *The following equality holds:*

$$\Omega_\ell^{-1} = \sqrt{2} \left((-1)^k \binom{j}{k} 2^k 1_{(k \leq j)} \right)_{0 \leq j, k \leq \ell}.$$

Therefore $\vec{\theta}_\ell = \Omega_\ell^{-1}\vec{\alpha}_\ell$. Note that since both Ω_ℓ and Ω_ℓ^{-1} are lower triangular, we have the consistency property: the first $\ell-1$ coordinates of $\vec{\alpha}_\ell$ and $\vec{\theta}_\ell$ are equal to those of $\vec{\alpha}_{\ell-1}$ and $\vec{\theta}_{\ell-1}$.

Now we have to define estimators of $(\theta_k(f))$. For this, consider the empirical estimators (2) of $\alpha_k := \alpha_k(f)$ and set

$$\vec{\hat{\alpha}}_\ell = {}^t(\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_\ell) \tag{9}$$

The vector $\vec{\theta}_\ell = (\theta_k(f), k = 0, \ldots, \ell)'$ of components of f is estimated by $\vec{\hat{\theta}}_\ell = \Omega_\ell^{-1}\vec{\hat{\alpha}}_\ell$. By the triangular form of Ω_ℓ, $\vec{\hat{\alpha}}_\ell$ and $\vec{\hat{\theta}}_\ell$ have their first $\ell-1$ coordinates equal to those of $\vec{\hat{\alpha}}_{\ell-1}$ and $\vec{\hat{\theta}}_{\ell-1}$.

Denote by $f_\ell = \sum_{k=0}^{\ell}\theta_k(f)\varphi_k$ the orthogonal projection of f on $S_\ell = \mathrm{span}(\varphi_0, \varphi_1, \ldots, \varphi_\ell)$.

We define the following collection of estimators of f by

$$\hat{f}_\ell = \sum_{k=0}^{\ell}\hat{\theta}_k\varphi_k, \quad \vec{\hat{\theta}}_\ell = \Omega_\ell^{-1}\vec{\hat{\alpha}}_\ell, \quad \ell \geq 0. \tag{10}$$

Recall that $\|.\|$ denotes the $\mathbb{L}^2$-norm of $\mathbb{L}^2((0,+\infty))$. The following risk decomposition holds.

Proposition 2.3. *The estimator $\hat{f}_\ell$ of f defined by (2)-(8)-(9)-(10) satisfies*

$$\mathbb{E}(\|\hat{f}_\ell - f\|^2) \leq \|f - f_\ell\|^2 + \frac{16}{15}\frac{2^{4\ell}}{n}.$$

Proposition 2.3 states a squared-bias/variance decomposition, and we need now to specify the bias order on adequate functional spaces, in order to evaluate optimal rates.

2.2. **Rates and rate optimality.** As it is always the case in nonparametric estimation[1], we must link the bias term $\|f - f_\ell\|^2$ with regularity properties of function f. In our context, these should be expressed in relation with the rate of decay of the coefficients $(\theta_k(f))_{k\geq 0}$. The Laguerre-Sobolev spaces described in Section 7 provide an adequate solution.

For $s \geq 0$, let

$$W_2^s((0,+\infty), K) = \{h : (0,+\infty) \to \mathbb{R}, h \in \mathbb{L}^2((0,+\infty)), \sum_{k\geq 0} k^s\theta_k^2(h) \leq K < +\infty\} \tag{11}$$

where $\theta_k(h) = \int_0^{+\infty} h(u)\varphi_k(u)du$. The subscript 2 corresponds to the scale parameter $a = 2$ of the basis. In particular, for s integer, if $h : (0,+\infty) \to \mathbb{R}$ belongs to $L^2((0,+\infty))$,

$$\sum_{k\geq 0} k^s(\theta_k(h))^2 < +\infty. \tag{12}$$

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

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

[1]Kernel methods use Hölder spaces for pointwise estimation, Nikol'ski classes for global estimation; projection methods use, on Fourier basis, Sobolev spaces, on wavelet bases, Besov spaces.

For any $h \in W_2^s((0,+\infty),K)$, we have $\|h-h_\ell\|^2 = \sum_{k=\ell+1}^{\infty} \theta_k^2(h) \leq K/\ell^s$ where h_ℓ is the orthogonal projection of h on S_ℓ.

Proposition 2.4. *Let for* $0<\epsilon<1$,

$$\ell_\epsilon = \frac{(1-\epsilon)\log(n)}{4\log(2)} \quad \text{and} \quad \ell^\star = \left[\frac{1}{4\log(2)}(\log(n) - s\log\log(n))\right] \vee 1.$$

We have

$$\sup_{f\in W_2^s((0,+\infty),K)} \mathbb{E}\left[\|\hat{f}_{\ell_\epsilon} - f\|^2\right] \leq K\left(\frac{4\log(2)}{1-\epsilon}\right)^s (\log(n))^{-s} + \frac{16}{15}\frac{1}{n^\epsilon},$$

and

$$\sup_{f\in W_2^s((0,+\infty),K)} \mathbb{E}\left[\|\hat{f}_{\ell^\star} - f\|^2\right] \leq \left(K(4\log(2))^s + \frac{16}{15}\right)(\log(n))^{-s}(1+o(1)).$$

Note that $\tilde{\ell}_\epsilon$ does not depend on s and is thus adaptive. With $\ell^\star$, the bias and variance terms have the same order $(\log(n))^{-s}$, which is better. In addition, the constant is improved. Nevertheless, this choice depends on s.

Proof. For $f \in W_2^s(0,+\infty),K)$, the risk bound in Proposition 2.3 writes

$$\mathbb{E}(\|\hat{f}_\ell - f\|^2) \leq \frac{K}{\ell^s} + \frac{16}{15}\frac{2^{4\ell}}{n}.$$

The variance term has exponential order $2^{4\ell}$ with respect to ℓ. Thus, we can not make the classical bias variance compromise. First we can choose ℓ such that the bias term dominates: this is obtained by choosing $\ell = \ell_\epsilon$. Second, a more precise tuning of both terms is obtained with $\ell = \ell^\star$. In both cases, the rate is of order $O([\log(n)]^{-s})$. □

We now prove that, for densities lying in Laguerre-Sobolev balls $W_2^s((0,+\infty),K)$, the rate $(\log n)^{-s}$ is optimal.[2]

Theorem 2.1. *Assume that* s *is a positive integer and let* $K \geq 1$. *There exists a constant* $c>0$ *such that*

$$\liminf_{n\to+\infty} (\log(n))^s \inf_{\hat{f}_n} \sup_{f\in W^s((0,+\infty),K)} \mathbb{E}_f\left[\|\hat{f}_n - f\|^2\right] \geq c$$

where $\inf_{\hat{f}_n}$ *denotes the infimum over all estimators of* f *based on* $(N_j(C_j))_{1\leq j\leq n}$.

The proof uses several lemmas established in Zhang (1995) and Loh and Zhang (1996).

2.3. **Model selection.** Model selection is justified as the bias may have much smaller order. For instance, it can be null if f admits a finite development in the Laguerre basis. Exponential distributions also provide examples of smaller bias. Indeed, consider f an exponential density $\mathcal{E}(\theta)$. Then

$$\theta_k(f) = \int_0^{+\infty} \varphi_k(c)\theta e^{-\theta c}dc = \sqrt{2}\theta\sum_{j=0}^{k}(-1)^j\binom{k}{j}\frac{2^j}{j!}\int_0^{+\infty} c^j e^{-(\theta+1)c}dc = \sqrt{2}\frac{\theta}{\theta+1}\left(\frac{\theta-1}{\theta+1}\right)^k.$$

[2]Note that analogous rates occur in the context of deconvolution for ordinary smooth function and supersmooth noise (severely ill-posed problem). Nevertheless, the logarithmic rate is proved to be optimal, see Fan (1991).

As a consequence

$$\|f - f_\ell\|^2 = \sum_{k=\ell+1}^{\infty} \theta_k^2(f) = \frac{\theta}{2}\left(\frac{\theta-1}{\theta+1}\right)^{2(\ell+1)}.$$

Choosing

$$\ell = \ell_{opt} = \lambda \log(n) \quad \text{with} \quad \lambda = \frac{1}{2(\log(2) + \log(|(\theta+1)/(\theta-1)|)}$$

yields the rate

$$O(n^{-1/(1+\mu)}) \quad \text{with} \quad \mu = \frac{2\log(2)}{\log(|(\theta+1)/(\theta-1)|)}.$$

The rate depends on θ and can be $O(n^{-\beta})$ for any $\beta < 1$. For instance if $\theta = 5/3$ the rate is $O(n^{-1/2})$, for $\theta = 1/2$, the rate is $O(n^{-0.44})$ (see Section 4) and it tends to $O(n^{-1})$ (the parametric rate) when θ tends to 1, which is coherent with the fact that the bias is null for $\theta = 1$.

This kind of result can be generalized to the case of a distribution f defined as a mixture of exponential distributions and to Gamma distributions $\Gamma(p, \theta)$, with p an integer. More precisely, if f_p is the density $\Gamma(p, \theta)$,

$$\theta_k(f_p) = \frac{\sqrt{2}}{\Gamma(p)}\left(\frac{\theta}{\theta+1}\right)^p S_{p,k}\left(\frac{2}{\theta+1}\right), \quad \text{with} \quad S_{p,k}(x) = \frac{d^{p-1}}{dx^{p-1}}\left[x^{p-1}(1-x)^k\right].$$

This term can be computed explicitly and we get, for $\ell \geq p-1$,

$$\sum_{k\geq\ell}[\theta_k(f_p)]^2 \leq \left(\frac{\theta-1}{\theta+1}\right)^{2(\ell-(p-1))} C(p,\theta), \quad \text{with} \quad 0 < C(p,\theta) < +\infty.$$

Note that the bias is null for $\theta = 1$ and $\ell > p-1$, which is expected since $f_p \in S_{p-1}$. Moreover, the bias order depends on θ, which can be seen in simulations.

Now we have to define an automatic selection rule of the adequate dimension ℓ. We make the selection among the following set:

$$\mathcal{M}_n = \left\{\ell \in \{0, 1, \ldots, L_n\}, L_n = \left[\frac{\log(n)}{\log(2)}\right] + 1\right\}$$

where $[x]$ denotes the integer part of the real number x. For κ a numerical constant, we define

$$\hat{\ell} = \arg\min_{\ell\in\mathcal{M}_n}\left\{-\|\hat{f}_\ell\|^2 + \text{pen}(\ell)\right\}, \quad \text{with} \quad \text{pen}(\ell) = \kappa\frac{\ell 2^{4\ell}}{n}. \tag{13}$$

We can prove the following result

Theorem 2.2. *Consider the estimator $\hat{f}_{\hat{\ell}}$ defined by (10) and (13). For any $\kappa \geq 8$, we have*

$$\mathbb{E}(\|\hat{f}_{\hat{\ell}} - f\|^2) \leq \inf_{\ell\in\mathcal{M}_n}\left(3\|f_\ell - f\|^2 + 4\text{pen}(\ell)\right) + \frac{C}{n}.$$

The infimum in the right-hand-side of the inequality above shows that the estimator is indeed adaptive. Note that the penalty is, up to a constant, equal to the variance multiplied by ℓ. This implies a possible negligible loss in the rate of the adaptive estimator w.r.t. the expected optimal rate.

Remark. Let us now assume that the observation is $(N_j(C_jT), j = 1, \ldots, n)$. The previous method applies directly to estimate the density f_T of C_jT *i.e.* $f_T(t) = (1/T)f(t/T)$. We can deduce the results for $f(c) = Tf_T(Tc)$. The function f is developed on the basis $(\varphi_k^{(T)} :=$

$\sqrt{T}\varphi_k(T.), k \geq 0)$ and the following relation holds $\theta_k^{(T)}(f) = \sqrt{T}\theta_k(f_T) =< f, \varphi_k^{(T)} >$. Denote by $f_\ell^{(T)}$ the orthogonal projection of f on the space $S_\ell^{(T)}$ spanned by $(\varphi_k^{(T)}, k \leq \ell)$. To estimate $f(c) = Tf_T(Tc)$, we set for all ℓ,

$$\hat{f}_\ell^{(T)}(c) := T\hat{f}_{T,\ell}(Tc)$$

where $\hat{f}_{T,\ell}$ is the estimator built for f_T using $(N_j(C_jT), j = 1, \ldots, n)$. The estimator $\hat{f}_\ell^{(T)}$ of f satisfies

$$\mathbb{E}(\|\hat{f}_\ell^{(T)} - f\|^2) \leq \|f - f_\ell^{(T)}\|^2 + T\frac{16}{15}\frac{2^{4\ell}}{n}.$$

Moreover, with $\hat{\ell}$ defined in (13), there exists $\kappa > 0$ such that

$$\mathbb{E}(\|\hat{f}_{\hat{\ell}}^{(T)} - f\|^2) \leq \inf_{\ell \in \mathcal{M}_n} \left(3\|f_\ell^{(T)} - f\|^2 + 4T\mathrm{pen}(\ell)\right) + \frac{CT}{n}.$$

The variance term in the $\mathbb{L}^2$-risk is multiplied by a factor T. This explains that the method may be worse when T increases. Actually, this was clear on simulated data.

2.4. **Related works.** In Simar (1976), it is proved that the cumulative distribution function $F(x)$ of C_j can be consistently estimated using $(\hat{\alpha}_\ell)$. The method is theoretical and concrete implementation is not easy. Noting that $\alpha_0(f)$ is simply the Laplace transform of f, Karr (1984) studies the properties of $\hat{\alpha}_0$ to estimate $\alpha_0(f)$ in the more general context of mixed point Poisson processes.

For comparison purposes, we detail some of the results of Zhang (1995), Hengartner (1995) and Loh and Zhang (1996, 1997) in the case of Poisson mixtures. In the case where f has compact support $[0, \theta^\star]$, Zhang (1995) gives a kernel estimator of $f(a)$ and studies pointwise quadratic risk on Hölder classes with index r (i.e. functions f admitting $\lfloor r \rfloor$ derivatives such that $f^{(\lfloor r \rfloor)3}$ is $r - \lfloor r \rfloor$-Hölder). The estimator has a MSE of order $[\log(n)/\log\log(n)]^{-2r}$ which does not correspond to his lower bound which is $[\log(n)]^{-2r}$. In the case of non compact support for f, the kernel estimator MSE has order $(\log(n))^{-r/2}$, with no associated lower bound. Loh and Zhang (1996) generalize the results of Zhang (1995) by studying a weighted-$\mathbb{L}_p$-risk.

Hengartner (1997) considers the case where f has a compact support. He builds projection estimators using orthogonal polynomials on the support. The upper bound of MISE has order $[\log(n)/\log\log(n)]^{-2r}$ on the same class as above and on Sobolev classes with index r. On the latter classes, he proves a lower bound of order $[\log(n)/\log\log(n)]^{-2r}$.

Loh and Zhang (1997), in the case of non compact support for f, use Laguerre polynomials and build projection estimators. Thus, the function is estimated by a polynomial; they study a weighted $\mathbb{L}^2$-risk. The upper bound is $O([\log(m)]^{-m/2})$ on the class of functions such that $\sum_{j \geq m} j^m \tau_j^2(f) < M$ where $\tau_j(f)$ is the coefficient of f on the development with respect to the Laguerre polynomials. Their lower bound is $O([\log(n)]^{-m})$, which does not correspond to the upper bound.

In all cases, the number of coefficients in the projection estimators does not depend on the regularity space. In this sense, the above methods are adaptive.

Let us now clarify our contribution. First, we use a $\mathbb{L}^2((0, +\infty))$-basis and a usual MISE, which is more fitted to the problem. Second, we clarify the functional spaces associated to the context of Laguerre bases on $(0, +\infty)$ and provide explicit links between regularity and coefficients of a development on these spaces. Upper and lower bounds match globally and without weights.

[3] where $\lfloor r \rfloor$ is the largest integer previous r

Here, the proof of our lower bound is inspired of Loh and Zhang's constructions. Therefore, our results synthesize and improves all these previous works.

Lastly, when the function under estimation has stronger regularity properties than considered in lower bounds, we show that the rate can be improved (polynomial instead of logarithmic). This justifies the proposal of an adaptive procedure, see Theorem 2.2, which is moreover non asymptotic.

3. Estimation for large T

Let us set

$$\widehat{C}_{j,T} := \frac{1}{T} N_j(C_j T).$$

Conditionally to $C_j = c$, we know that $\widehat{C}_{j,T}$ converges almost surely to c as T tends to infinity. Consequently, $\widehat{C}_{j,T}$ converges almost surely to C_j. We now use the i.i.d. sample $(\widehat{C}_{j,T})_{1\le j\le n}$ to build projection estimators of f, where the coefficients $\theta_k(f)$ are now estimated as follows.

$$\tilde{f}_\ell^{(T)} = \sum_{k=0}^{\ell} \tilde{\theta}_k \varphi_k, \quad \tilde{\theta}_k = \frac{1}{n}\sum_{j=1}^{n} \varphi_k(\widehat{C}_{j,T}). \tag{14}$$

Note that S_ℓ has the norm-connection property:

$$\forall t \in S_\ell, \quad \|t\|_\infty := \sup_{x\in\mathbb{R}^+} |t(x)| \le \sqrt{2(\ell+1)}\|t\|, \tag{15}$$

as can be seen from Lemma 6.1. We obtain the following risk bound.

Proposition 3.1. *Recall that f_ℓ is the orthogonal projection of f on $S_\ell = \mathrm{span}(\varphi_0, \dots, \varphi_\ell)$. Then*

$$\mathbb{E}(\|\tilde{f}_\ell^{(T)} - f\|^2) \le \|f - f_\ell\|^2 + 2\frac{\ell+1}{n} + \frac{8(\ell+1)^5}{T^2} s_2, \quad s_2 := 3\mathbb{E}(C_1^2) + \frac{\mathbb{E}(C_1)}{T}.$$

The bound contains the usual decomposition into a squared-bias term $\|f - f_\ell\|^2$ and a variance term. The latter term is the sum of two components: the first one $2(\ell+1)/n$ is classical and no more exponential in ℓ, the second one is due to the approximation of the C_j's by the $\widehat{C}_{j,T}$'s and gets small when T increases. To define a penalization procedure, we must estimate s_2. Let

$$\widehat{s}_2 = \frac{1}{n}\sum_{j=1}^{n} [3(\widehat{C}_{j,T})^2 - 2\frac{\widehat{C}_{j,T}}{T}]. \tag{16}$$

As $3(\widehat{C}_{j,T})^2 - 2\widehat{C}_{j,T}/T = \widehat{C}_{j,T}(3N_j(C_jT) - 2)/T \ge 0$, $\hat{s}_2 \ge 0$. Elementary computations using conditioning on C_j show that $\mathbb{E}(\hat{s}_2) = s_2$. Now, set

$$\mathcal{M}_{n,T} = \left\{0, 1, \dots, n \wedge T^{2/5}\right\}$$

and

$$\tilde{\ell} = \arg\min_{\ell\in\mathcal{M}_{n,T}} \left\{-\|\tilde{f}_\ell^{(T)}\|^2 + \widetilde{\mathrm{pen}}(\ell)\right\} \quad \text{with} \quad \widetilde{\mathrm{pen}}(\ell) = \tilde{\kappa}_1 \frac{(\ell+1)}{n} + \tilde{\kappa}_2 \frac{(\ell+1)^5}{T^2}\hat{s}_2. \tag{17}$$

The following holds.

Theorem 3.1. *Assume that $\mathbb{E}(C_1^8) < +\infty$. Let $\tilde{f}_{\tilde{\ell}}^{(T)}$ the estimator defined by (14) and (17). Then there exist numerical constants $\tilde{\kappa}_1, \tilde{\kappa}_2$ such that*

$$\mathbb{E}(\|\tilde{f}_{\tilde{\ell}}^{(T)} - f\|^2) \le C \inf_{\ell\in\mathcal{M}_{n,T}} \left(\|f - f_\ell\|^2 + 2\tilde{\kappa}_1\frac{\ell+1}{n} + \frac{8\tilde{\kappa}_2(\ell+1)^5 s_2}{T^2}\right) + \frac{C'}{n}$$

where C is a numerical constant and C' a positive constant.

Thus, the estimator $\tilde{f}_{\tilde{\ell}}^{(T)}$ is adaptive and its risk automatically reaches the order of the bias-variance compromise.

4. Numerical simulations

In this paragraph, we illustrate on simulated data the two adaptive projection methods using the Laguerre basis: method 1 corresponds to Section 2 when $T = 1$, method 2 corresponds to section 3 for large T.

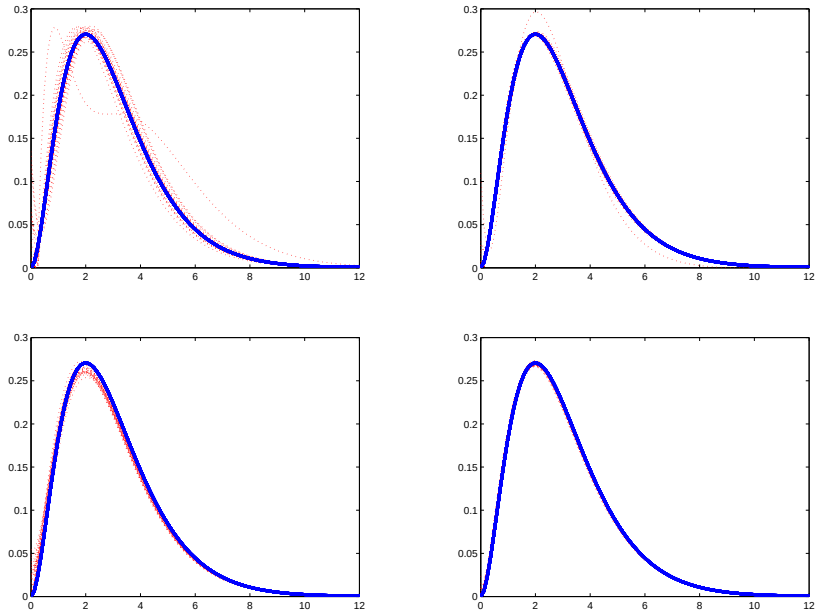

Figure 1. Estimation of the Gamma(3,1) density with method 1 (top left $n = 10000$ and top right $n = 100000$, for $T = 1$) and method 2 (bottom left, $n = 1000, T = 10$ and bottom right $n = 4000, T = 40$): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of the time $\hat{\ell} = 2$ for both methods.

We consider different distributions for the C_j's:

(1) a Gamma $\Gamma(p, \theta)$ for $p = 3, \theta = 1$,
(2) a mixed Gamma density $0.3\Gamma(3, 0.25) + 0.7\Gamma(10, 0.6)$.
(3) an exponential $\mathcal{E}(\theta)$, with $\theta = 1/2$, $f_\theta(x) = \theta e^{-\theta x} 1_{x>0}$,
(4) a Pareto density $f_{(p,\theta)}(x) = p(1 + p\theta x)^{-1-1/p} 1_{x>0}$, with $p = 5$ and $\theta = 1/2$,
(5) a Weibull density $f_{(p,\theta)}(x) = \theta p^{-\theta} x^{\theta-1} e^{-(x/p)^\theta} 1_{x>0}$ for $p = 3$ and $\theta = 2$.

Note that, as $\theta = 1$, the density (1) has only three nonzero coefficients θ_0, θ_1, θ_2 in its exact development in the Laguerre basis. For density (3), we know that the rate of the $\mathbb{L}^2$ risk depends on the value of θ ($n^{-0.44}$ for $\theta = 1/2$, see Section 2). In Figures 1-5, we illustrate the first method for $T = 1$ and $n = 10000$, $n = 100000$ and the second for sample sizes $n = 1000$ and $T = 10$, and $n = 4000$, $T = 40$, for the five densities defined above. We plot 25 consecutive estimates on the same picture together with the unknown density to recover, to show variability bands and

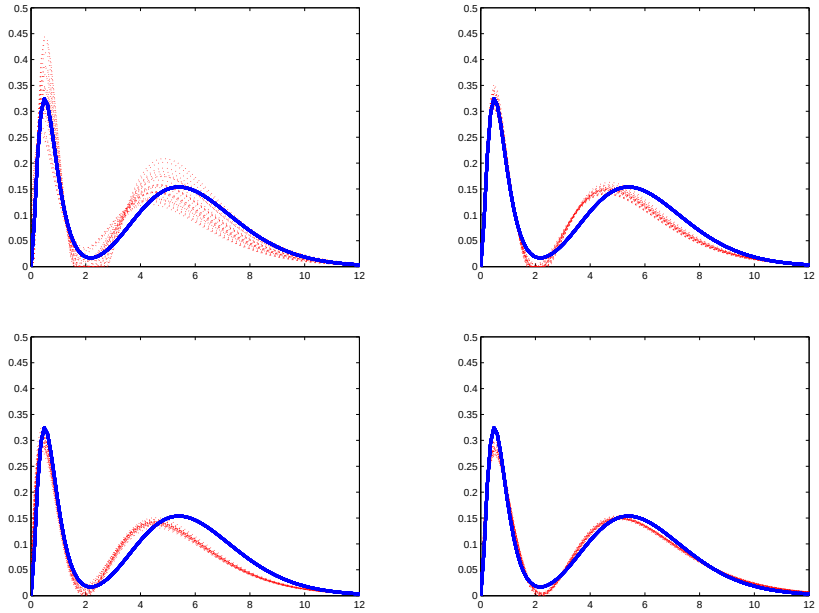

FIGURE 2. Estimation of the mixed Gamma density with method 1 (top left $n = 10000$ and top right $n = 100000$, for $T = 1$) and method 2 (bottom left, $n = 1000, T = 10$ and bottom right $n = 4000, T = 40$): true -thick (blue) line and 25 estimated (dashed (red) lines). The selected ℓ is 3 except for the bottom right plot where it is 4.

illustrate the stability of the procedures.

• Comments on method 1. The method is easy to implement. As it is standard for penalized methods, the theoretical constant is too large and in practice, is calibrated by preliminary simulations. We have selected the constant $\kappa = 0.001$ in the penalty. This prevents from possible explosion of the variance, which has exponential order. The adaptive estimator performs reasonably well for large values of n ($n \geq 10000$) but is very sensitive to the parameter values for distributions Gamma or exponential, as expected. The mixture density and the Pareto and Weibull densities, which do not admit finite developments in the basis, are correctly estimated. Increasing n improves significantly the estimation. We choose to select ℓ in $\{0, 1, \ldots, 2\lfloor \log(n) \rfloor - 1\}$. On the examples, the algorithm selects values of $\hat{\ell}$ belonging to $\{0, 1, \ldots, 4\}$.

• Comments on method 2. The method is also easy to implement. We have selected the constants $\tilde{\kappa}_1 = 1.5$, $\tilde{\kappa}_2 = 10^{-5}$. The very small value of $\tilde{\kappa}_2$ simply kills the effect of the second term in the penalty in order to allow not too large values of T. This second method gives better results than the first method, as soon as $T \geq 10$ (even $T \geq 5$ provides good estimators). The number of observations need not be very large. We kept the same set of possible values for ℓ in the selection algorithm; here again, the selected values $\tilde{\ell}$ are in $\{0, 1, \ldots, 4\}$.

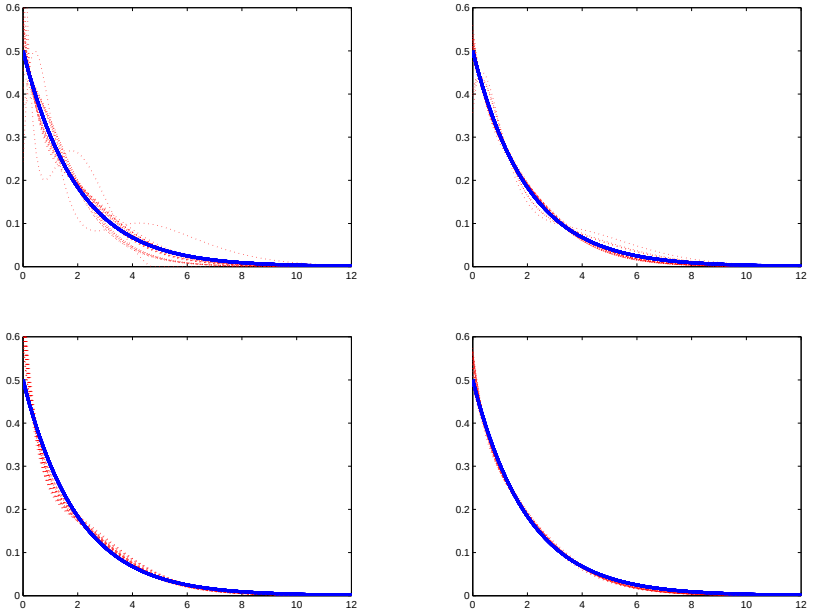

FIGURE 3. Estimation of the Exponential density with projection method 1 (top left $n = 10000$ and top right $n = 100000$, for $T = 1$) and method 2 (bottom left, $n = 1000, T = 10$ and bottom right $n = 4000, T = 40$): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of the time $\hat{\ell} = 2$.

5. Concluding remarks

In this paper, we study the nonparametric density estimation of a positive random variable C from the observation of $(N_j(C_jT), j = 1, \ldots, n)$, where (N_j) are i.i.d. Poisson processes with unit intensity, (C_j) are i.i.d. random variables distributed as C, and (N_j) and (C_j) are independent. Under the assumption that the unknown density f of the unobserved variables (C_j) is in $\mathbb{L}^2((0, +\infty))$ and for a fixed value T, we express the nonparametric problem as an inverse problem, which can be solved by using a Laguerre basis of $\mathbb{L}^2((0, +\infty))$. Explicit estimators of the coefficients of f on the basis are proposed and used to define a collection of projection estimators. The space dimension is then selected by a data driven criterion. For functions belonging to Sobolev-Laguerre spaces described in Section 2, f is estimated at a rate $O((\log(n))^{-s})$. So, an interesting question is to know whether there exist other functions than those of these spaces estimated at the same rate. This problem amounts to finding maximal functional classes for which a given rate of convergence of the estimators can be achieved.

For large T, estimators $\widehat{C}_{j,T}$ of the C_j's are used to build adaptive projection estimators in the Laguerre basis. In this approach, a moment condition on C_j is required.

The numerical simulation results show that the Laguerre basis is indeed appropriate, to obtain estimators with no boundary effects at 0.

Possible developments of this work are the following. We may use specific kernel estimators on $\mathbb{R}^+$, as in Comte and Genon-Catalot (2012), to compare them with projection Laguerre estimators. As in Fabio *et al.* (2012), we may enrich the data by considering several observation times. Another relevant extension is to study mixed compound Poisson processes, e.g. using the approach of Comte *et al.* (2014), or more general mixed Lévy processes.

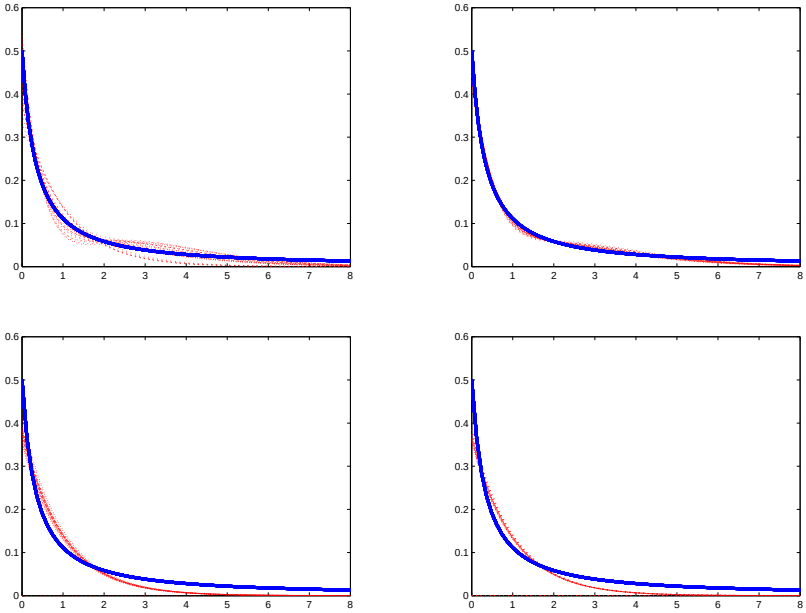

FIGURE 4. Estimation of the Pareto density with projection method 1 (top left $n = 10000$ and top right $n = 100000$, for $T = 1$) and method 2 (bottom left, $n = 1000, T = 10$ and bottom right $n = 4000, T = 40$): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of the time $\hat{\ell} = 2$ for the top pictures and 0 for the bottom ones.

6. PROOFS

6.1. **Proof of Proposition 2.1.** Using (5), we have

$$\Omega_k^{(\ell)} = \frac{1}{\ell!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\int_0^{+\infty}\sqrt{2}\frac{(2c)^j}{j!}c^\ell e^{-2c}dc = \frac{1}{\ell!}\frac{1}{\sqrt{2}\,2^\ell}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\frac{(\ell+j)!}{j!}.$$

Finally,

$$\Omega_k^{(\ell)} = \frac{1}{\sqrt{2}2^\ell}\sum_{j=0}^{k}(-1)^j\binom{k}{j}\frac{(\ell+j)(\ell+j-1)\dots(\ell+1)}{j!} \tag{18}$$

where we know that $\Omega_k^{(\ell)} = 0$ for $k > \ell$. Therefore $\ell \to 2^\ell\Omega_k^{(\ell)}$ is a polynomial of degree k which is equal to 0 for $\ell = 0, 1, \dots, k-1$. Hence, we have $2^\ell\Omega_k^{(\ell)} \propto \ell(\ell-1)(\ell-2)\dots(\ell-k+1)$. The proportionality coefficient is equal to the coefficient of ℓ^k is $(-1)^k/(\sqrt{2}\,k!)$. Hence the result. □

6.2. **Proof of proposition 2.2.** Denote by $\mathbb{R}_\ell[X]$ the space of polynomials with real coefficients and degree less than or equal to ℓ. The transpose of the matrix $\sqrt{2}\Omega_\ell$ represents the linear application of $\mathbb{R}_\ell[X]$, $P(X) \mapsto P\left(\frac{1-X}{2}\right)$, in the canonical basis $(1, X, \dots, X^\ell)$. The inverse linear mapping is $Q(X) \mapsto Q(1-2X)$. Hence the result. □

6.3. **Proof of Proposition 2.3.** We define by $|.|$ the usual Euclidean norm in $\mathbb{R}^{\ell+1}$.

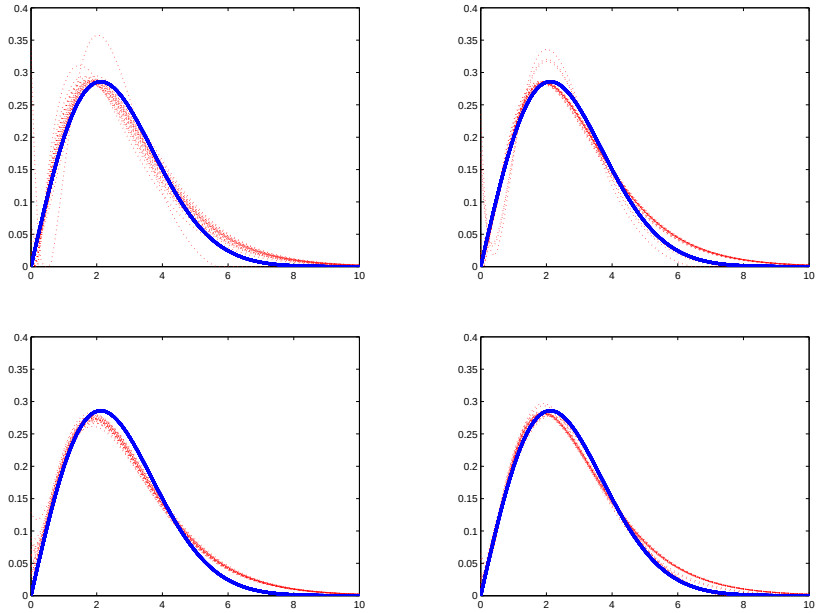

FIGURE 5. Estimation of the Weibull density with method 1 (top left $n = 10000$ and top right $n = 100000$, for $T = 1$) and method 2 (bottom left, $n = 1000, T = 10$ and bottom right $n = 4000, T = 40$): true -thick (blue) line and 25 estimated (dashed (red) lines). The selected ℓ's are 2, 3 or 4.

We have

$$\begin{aligned}\mathbb{E}(\|\hat{f}_\ell - f\|^2) &= \|f - f_\ell\|^2 + \mathbb{E}(\|\hat{f}_\ell - f_\ell\|^2) = \|f - f_\ell\|^2 + \mathbb{E}\left(\sum_{k=0}^{\ell}(\hat{\theta}_k - \theta_k)^2\right) \\ &= \|f - f_\ell\|^2 + \mathbb{E}(|\Omega_\ell^{-1}(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)|^2).\end{aligned}$$

Next, we write the variance term as follows:

$$\mathbb{E}(|\Omega_\ell^{-1}(\hat{\alpha}_\ell - \alpha_\ell)|^2) = \mathbb{E}\left({}^t(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)\,{}^t\Omega_\ell^{-1}\Omega_\ell^{-1}(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)\right). \tag{19}$$

Now, note that, if $M = (m_{i,j})_{0\le i,j\le \ell}$ is a $(\ell+1)\times(\ell+1)$ matrix,

$$\mathbb{E}({}^t(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)M(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)) = \sum_{0\le i,j\le \ell}\mathrm{cov}(\hat{\alpha}_i, \hat{\alpha}_j)m_{i,j}$$

where $\mathrm{cov}(\hat{\alpha}_i, \hat{\alpha}_j) = (\alpha_i\delta_i^j - \alpha_i\alpha_j)/n$ and δ_i^j is the Kronecker symbol. Thus, for M symmetric and nonnegative,

$$\mathbb{E}({}^t(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)M(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)) \le \mathrm{Tr}(MD_\alpha)/n$$

where $D_\alpha = diag(\alpha_0, \ldots, \alpha_\ell)$. Here, we get

$$\mathbb{E}(\|\hat{f}_\ell - f\|^2) \le \|f - f_\ell\|^2 + \frac{1}{n}\mathrm{Tr}({}^t\Omega_\ell^{-1}\Omega_\ell^{-1}D_\alpha). \tag{20}$$

Since $0 \leq \alpha_k \leq 1$ and $[\ {}^t\Omega_\ell^{-1}\Omega_\ell^{-1}]_{k,k} \geq 0$ for all k, we have

$$\mathbb{E}(\|\hat{f}_\ell - f\|^2) \leq \|f - f_\ell\|^2 + \frac{1}{n}\mathrm{Tr}(\ {}^t\Omega_\ell^{-1}\Omega_\ell^{-1}).$$

Note that $\mathrm{Tr}(\ {}^t\Omega_\ell^{-1}\Omega_\ell^{-1})$ is known as the squared Frobenius norm of the matrix Ω_ℓ^{-1}. It follows from Proposition 2.2 that

$$\mathrm{Tr}(\ {}^t\Omega_\ell^{-1}\Omega_\ell^{-1}) = 2\sum_{k=0}^{\ell}\sum_{j=0}^{k}\left(\binom{k}{j}\right)^2 2^{2j} \leq 2\sum_{k=0}^{\ell}2^{2k}\sum_{j=0}^{k}\left(\binom{k}{j}\right)^2. \tag{21}$$

Noting that

$$\sum_{j=0}^{k}\left(\binom{k}{j}\right)^2 = \binom{2k}{k} \leq 2^{2k-1},$$

we get

$$\mathrm{Tr}(\ {}^t\Omega_\ell^{-1}\Omega_\ell^{-1}) \leq \sum_{k=0}^{\ell}2^{4k} = \frac{2^{4(\ell+1)}-1}{2^4-1} \leq \frac{16}{15}2^{4\ell}. \tag{22}$$

As a consequence, we obtain the risk decomposition announced in Proposition 2.3.□

6.4. **Proof of Theorem 2.1.** From Tsybakov (2009) Chapter 2, we have to define two functions f_{0n}, f_{1n} such that

(1) f_{0n} and f_{1n} are densities,
(2) For some $K > 0$, f_{0n} and f_{1n} belong to $W_2^s((0,+\infty), K)$,
(3) For $j = 0, 1$, let $P_{jn} = (\alpha_x(f_{jn}), x \in \mathbb{N})$, then

$$V(P_{1n}, P_{0n}) = \sum_{x=0}^{+\infty}|\alpha_x(f_{1n}) - \alpha_x(f_{0n})| = O(1/n).$$

(4) $\|f_{0n} - f_{1n}\|^2 \geq C(\log(n))^{-s}$.

For the construction of the $f_{jn}, j = 0, 1$, we follow Loh and Zhang (1996,1997). Let $f_0(c) = e^{-c}$, $0 < c_0 < c_1 < b < c_2 < c_3$, and

$$f_{u,v}(c) = 1_{[c_0,c_1[}(c)\ell_{1,u,v}(c) + 1_{[c_1,c_2[}(c)\gamma_{u,v}(c) + 1_{[c_2,c_3[}(c)\ell_{2,u,v}(c),$$

where $\gamma_{u,v}(c) = (v^u/\Gamma(u))c^{u-1}e^{-vc}$ is the gamma density with parameter (u, v), $\ell_{i,u,v}, i = 1, 2$ are polynomials of degree $2s + 1$ such that $f_{u,v}$ is of class C^s. We set

$$u_n = \delta_0 \log n := u, \quad v_n = u_n/b := v.$$

Set $\chi_i(c) = 1_{[c_i,c_{i+1}[}(c), i = 0, 1, 2$. Then, for $\varepsilon > 0$, we set

$$f_{0n}(c) = f_0(c) + 3(\varepsilon/u^{1/4})\,(c_2/u)^{s/2}\,(f_{u,v}(c) - w_{on}f_0(c))\,.$$

We choose w_{0n} such that $\int f_{0n} = 1$. As $\int f_0 = 1$, we find

$$w_{0n} = \int_{c_0}^{c_3} f_{u,v}(c)dc.$$

Now, we define

$$f_{1n}(c) = f_{0n}(c) + (\varepsilon/u^{1/4})\,(c_2/u)^{s/2}\left(\cos\left(u\frac{c-b}{c_2}\right) - w_{1n}/w_{0n}\right)f_{u,v}(c).$$

Then, w_{1n} is chosen such that $\int f_{1n} = 1$ which yields:

$$w_{1n} = \int_{c_0}^{c_3} \cos\left(u\frac{c-b}{c_2}\right) f_{u,v}(c)dc.$$

Finally, δ_0 is chosen by

$$\delta_0 = \max\{\frac{c_2/(c_3-c_2)}{\log(c_3/c_2)}, \frac{2}{\log(1+b^2/c_2^2)}, \frac{1}{c_1/b-1-\log(c_1/b)}, \frac{1}{c_2/b-1-\log(c_2/b)}\} \tag{23}$$

- Step 1: $w_{0n} = 1 + o(1)$, $f_{jn} \geq 0, j = 0, 1$.

Proof. We first study f_{0n}. By construction,

$$\ell_{1,u,v}^{(j)}(c_0) = \ell_{2,u,v}^{(j)}(c_3) = 0, \;\; \ell_{1,u,v}^{(j)}(c_1) = \gamma_{u,v}^{(j)}(c_1), \;\; \ell_{2,u,v}^{(j)}(c_2) = \gamma_{u,v}^{(j)}(c_2), \;\; j = 0, \ldots, s.$$

On the space of polynomials of degree $2s+1$ on $[c_0, c_1]$,

$$\|Q\|_0 = \sum_{j=0}^{s} |Q^{(j)}(c_0)| + |Q^{(j)}(c_1)|$$

is a norm and all norms are equivalent. Therefore, there exists C such that

$$\|\ell_{1,u,v}\chi_0\|_\infty \leq C\|\ell_{1,u,v}\chi_0\|_0 = C\sum_{j=0}^{s} |\gamma_{u,v}^{(j)}(c_1)|, \;\; \|\ell_{2,u,v}\chi_2\|_\infty \leq C\sum_{j=0}^{s} |\gamma_{u,v}^{(j)}(c_2)|.$$

By Lemma 3 of Loh and Zhang (1996), $|\gamma_{u,v}^{(j)}(c_1)| + |\gamma_{u,v}^{(j)}(c_2)| = O(n^{-1}u^{j+(1/2)})$, and $\|\gamma_{u,v}(1 - 1_{[c_1,c_2]})\|_p = O(1)n^{-1}u^{(p-1)/(2p)}, 1 \leq p \leq \infty$. Thus,

$$\sum_{j=0}^{s} |\gamma_{u,v}^{(j)}(c_1)| = O(n^{-1}u^{s+(1/2)}), \quad \|\gamma_{u,v}(1 - 1_{[c_1,c_2]})\|_\infty = O(1)n^{-1}u^{1/2}.$$

We deduce

$$w_{0n} = \int_{c_0}^{c_1} \ell_{1,u,v} + \int_{c_1}^{c_2} \gamma_{u,v} + \int_{c_2}^{c_3} \ell_{2,u,v} = 1 + O(u^{1/2}/n) + (u^{s+(1/2)}/n)O(1) = 1 + o(1).$$

We have

$$\begin{aligned} f_{0n}(\chi_0 + \chi_2) &= f_0(\chi_0 + \chi_2)(1 - 3(\varepsilon/u^{1/4})(c_2/u)^{s/2} w_{0n}) \\ &\quad + 3(\varepsilon/u^{1/4})(c_2/u)^{s/2}(\ell_{1,u,v}\chi_0 + \ell_{2,u,v}\chi_2) \end{aligned}$$

and $\|\ell_{1,u,v}\chi_0 + \ell_{2,u,v}\chi_2\|_\infty = (u^{s+(1/2)}/n)O(1)$. Therefore, provided that ε is small enough, the first term of $f_{0n}(\chi_0 + \chi_2)$ is lower bounded as $c_0 > 0$ and the second term is $O(u^{(s/2)+(1/4)}/n) = o(1)$. Thus, we can choose ε small enough to have $f_{0n}((\chi_0 + \chi_2) \geq 0$.
Then, $f_{0n}\chi_1 > 0$ and $f_{0n}1_{[c_0,c_3]^c} > 0$. Therefore, $f_{0n} \geq 0$.
We have

$$|w_{1n}| \leq \int_{c_1}^{c_2} \gamma_{u,v} + (u^{s+(1/2)}/n)O(1) = w_{0n} + o(1).$$

We check that $f_{1n} \geq 0$ in the same way as for f_{0n}. □

- Step 2. For $j = 0, 1$, $f_{jn} \in W_2^s((0, +\infty), K)$ for all $K \geq 1$.

Proof. This part is specific to our context as we do not have the same function spaces as Loh and Zhang (1996,1997). Step 2 is equivalent to proving that

$$|||f_{jn}|||_s^2 := \|c^{s/2}\left(f_{jn}(c)e^c\right)^{(s)}e^{-c}\|^2 = \sum_{k\geq s} k(k-1)\dots(k-s+1)\theta_k^2(f_{jn}) \leq K.$$

Note that, for a function f,

$$c^{s/2}\left(f(c)e^c\right)^{(s)}e^{-c} = c^{s/2}\sum_{j=0}^{s}\binom{s}{j}f^{(j)}(c).$$

We apply Lemmas 3 and 4 of Loh and Zhang (1996). We have, for $j=0,\dots,s$,

$$\|l_{1,u,v}^{(j)}\chi_0\|_2 = O(n^{-1}u^{s+1/2}), \quad \|l_{2,u,v}^{(j)}\chi_2\|_2 = O(n^{-1}u^{s+1/2}).$$

Moreover, for $j=0,\dots,s$, $u^{-1/4}\|\gamma_{u,v}^{(j)}\|_2 = O(u^{j/2})$. Consequently,

$$\begin{aligned}\varepsilon u^{-1/4}(c_2/u)^{s/2}\|c^{s/2}\sum_{j=0}^{s}\binom{s}{j}f_{u,v}^{(j)}(c)\|_2 &\leq C\varepsilon u^{-1/4}(c_2/u)^{s/2}c_3^{s/2}\sum_{j=0}^{s}\binom{s}{j}u^{j/2}u^{1/4}\\ &= C\varepsilon c_3^{s/2}(1+u^{1/2})^s(c_2/u)^{s/2} = \varepsilon O(1).\end{aligned}$$

Recall that

$$f_{0n}(c) = f_0(c)(1-3(\varepsilon/u^{1/4})(c_2/u)^{s/2}w_{0n}) + 3(\varepsilon/u^{1/4})(c_2/u)^{s/2}f_{u,v}(c)$$

where the coefficient of $f_0(c)$ is $1+o(1)$. Therefore

$$|||f_{0n}|||_s \leq |||f_0|||_s(1+\varepsilon o(1)) + \varepsilon O(1).$$

We have $\theta_0(f_0) = \sqrt{2}/2$, $\quad \theta_k(f_0) = 0, k \geq 1$. Therefore, $|||f_0|||_s^2 = 1/2$ and $|||f_{0n}|||_s^2 = (1/2)(1+\varepsilon O(1))$. The same holds for f_{1n}. We can choose any $K \geq 1$. □

- Step 3. $w_{1n}/w_{0n} = O(1/n)$ and

$$V(P_{1n},P_{0n}) = \sum_{x=0}^{+\infty}|\alpha_x(f_{1n}) - \alpha_x(f_{0n})| = o(1/n).$$

Proof. The proof is identical to Step 2 of Theorem 3 in Loh and Zhang (1996), p.574-575 ($\alpha' = s$). The choice of δ_0 by formula (23) is used in particular here and comes from Zhang (1995). □

- Step 4. There exists $C > 0$ such that

$$\|f_{1n} - f_{0n}\|^2 \geq C(\log n)^{-s}.$$

Proof. This part is also specific to our study: we only use two functions instead of three and our bound is global and not local. We have

$$(f_{1n}-f_{0n})^2 = \frac{\varepsilon^2 c_2^s}{u^{1/2+s}}\left(\cos^2\left(u\frac{c-b}{c_2}\right) - 2w_{1n}/w_{0n}\cos\left(u\frac{c-b}{c_2}\right) + (w_{1n}/w_{0n})^2\right)f_{u,v}^2,$$

where $w_{1n}/w_{0n} = o(1/n)$ and $|\cos| \leq 1$. Therefore, it is enough to bound from below:

$$\int f_{u,v}^2(c)\cos^2\left(u\frac{c-b}{c_2}\right)dc.$$

And as

$$\|\ell_{1,u,v}\chi_0 + \ell_{2,u,v}\chi_2\|_2 = (u^{s+(1/2)}/n)O(1),$$

we only look at

$$\int \chi_1(c)\gamma_{u,v}^2(c)\cos^2\left(u\frac{c-b}{c_2}\right)dc = \frac{1}{2}\int \chi_1(c)\gamma_{u,v}^2(c)\left(1+\cos\left(2u\frac{c-b}{c_2}\right)\right)dc := T_1+T_2$$

First $T_1 = O(\sqrt{u})$ because, by Lemma 3 of Loh and Zhang (1996, p.573),

$$\|\gamma_{u,v}\|^2 \sim \frac{1}{2\sqrt{\pi b}}\sqrt{u}, \quad \|\gamma_{u,v}(1-\chi_1)\|^2 = \frac{u^{1/2}}{n^2}O(1).$$

Next, we write

$$T_2 = \frac{1}{2}\int_0^{+\infty}\gamma_{u,v}^2(c)\cos\left(2u\frac{c-b}{c_2}\right)dc - \frac{1}{2}\int_0^{+\infty}(1-\chi_1(c))\gamma_{u,v}^2(c)\cos\left(2u\frac{c-b}{c_2}\right)dc := T_2'+T_2''$$

and as above $T_2'' = u^{1/2}/n^2 O(1)$. Moreover

$$T_2' = \frac{1}{2}\frac{v^{2u}}{\Gamma^2(u)}\frac{\Gamma(2u-1)}{(2v)^{2u-1}}\mathrm{Re}\left(\int_0^{\infty}\gamma_{2u-1,2v}(c)e^{2iu\frac{c-b}{c_2}}dc\right)$$

Now, notice that

$$\frac{v^{2u}}{\Gamma^2(u)}\frac{\Gamma(2u-1)}{(2v)^{2u-1}} = \|\gamma_{u,v}\|^2 = O(\sqrt{u}).$$

Moreover

$$J := \int_0^{\infty}\gamma_{2u-1,2v}(c)e^{2iu\frac{c-b}{c_2}}dc = e^{i\frac{2ub}{c_2}}\left(\frac{2v}{2v-\frac{2iu}{c_2}}\right)^{2u-1} = e^{i\frac{2ub}{c_2}}\left(\frac{1}{1-ib/c_2}\right)^{2u-1},$$

so that

$$|J| = \left(\frac{1}{1+\frac{b^2}{c_2^2}}\right)^{u-\frac{1}{2}} = \left(1+\frac{b^2}{c_2^2}\right)^{1/2} n^{-\delta_0\log(1+\frac{b^2}{c_2^2})} = O(1/n^2)$$

by the choice of δ_0. Therefore $T_2' = O(\sqrt{u}/n^2)$. Consequently

$$T_1+T_2 = \frac{1}{2\sqrt{\pi b}}\sqrt{u}\left(1+O(\frac{1}{n^2})\right).$$

It follows that

$$\|f_{1n}-f_{0n}\|^2 = \frac{\varepsilon^2 c_2^s}{u^{1/2+s}}\frac{1}{2\sqrt{\pi b}}\sqrt{u}(1+O(1/n)) = \frac{\varepsilon^2 c_2^s}{2\sqrt{\pi b}}\frac{1}{u^s}(1+O(1/n)).$$

This concludes step 3 as $u = \delta_0\log(n)$. □

6.5. **Proof of Theorem 2.2.** For simplicity, we set $L_n = L$. We define $\mathbf{S}_\ell = \{\mathbf{t} = {}^t(t_0,t_1,\dots,t_\ell,0,\dots,0) \in \mathbb{R}^{L+1}\}$, which can also be associated with the function $t = \sum_{k=0}^{\ell} t_k\varphi_k$ in S_ℓ and $|\mathbf{t}| = \|t\|$.

Now define, for $\mathbf{t}$ in any of the $\mathbf{S}_\ell$'s with $\ell \le L$,

$$\gamma_n(\mathbf{t}) = |\mathbf{t}|^2 - 2\langle \mathbf{t}, \Omega_L^{-1}\vec{\hat{\alpha}}_L\rangle.$$

For $\mathbf{t} \in \mathbf{S}_\ell$, note that $\gamma_n(\mathbf{t}) = |t^{(\ell)}|^2 - 2\langle t^{(\ell)}, \Omega_\ell^{-1}\vec{\hat{\alpha}}_\ell\rangle$, where $t^{(\ell)} = {}^t(t_0,t_1,\dots,t_\ell)$. Moreover, the vector $\hat{\mathbf{f}}_\ell = {}^t(\hat{\theta}_0,\dots,\hat{\theta}_\ell,0,\dots,0)$ is such that $\hat{\mathbf{f}}_\ell = \arg\min_{\mathbf{t}\in\mathbf{S}_\ell}\gamma_n(\mathbf{t})$ and satisfies

$$\gamma_n(\hat{\mathbf{f}}_\ell) = -\|\hat{f}_\ell\|^2 = -|\hat{\mathbf{f}}_\ell|^2 = -|\Omega_\ell^{-1}\vec{\hat{\alpha}}_\ell|^2.$$

For $\mathbf{s} \in \mathbf{S}_{\ell'}$ and $t \in \mathbf{S}_\ell$, the following decomposition holds:

$$\gamma_n(\mathbf{t}) - \gamma_n(\mathbf{s}) = \|t-f\|^2 - \|s-f\|^2 - 2\langle \mathbf{t}-\mathbf{s}, \Omega_L^{-1}(\vec{\hat{\alpha}}_L - \vec{\alpha}_L)\rangle$$

where $\|t-f\|^2=\sum_{k=0}^{L}(t_k-\theta_k)^2+\sum_{k=L+1}^{\infty}\theta_k^2$, for all k, $\theta_k=\langle f,\varphi_k\rangle$ and $t_{\ell+1},\dots,t_L$ are null when $\mathbf{t}\in\mathbf{S}_\ell$.

The integer $\hat\ell$ is given by

$$\hat\ell=\arg\min_{\ell\in\mathcal{M}_n}(\gamma_n(\hat{\mathbf{f}}_\ell)+\mathrm{pen}(\ell)),\quad\text{where }\hat{\mathbf{f}}_\ell={}^t(\hat\theta_0,\dots,\hat\theta_\ell,0,\dots,0)\in\mathbb{R}^{L+1}.$$

By definition of $\hat\ell$, $\gamma_n(\hat{\mathbf{f}}_{\hat\ell})+\mathrm{pen}(\hat\ell)\le\gamma_n(\mathbf{f}_\ell)+\mathrm{pen}(\ell)$ which implies

$$\|\hat f_{\hat\ell}-f\|^2\le\|f_\ell-f\|^2+\mathrm{pen}(\ell)+2\langle\hat{\mathbf{f}}_{\hat\ell}-\mathbf{f}_\ell,\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle-\mathrm{pen}(\hat\ell).$$

Now we have

$$2\langle\hat{\mathbf{f}}_{\hat\ell}-\mathbf{f}_\ell,\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle\le\frac14|\hat{\mathbf{f}}_{\hat\ell}-\mathbf{f}_\ell|^2+4\sup_{\mathbf{t}\in\mathbf{S}_{\hat\ell\vee\ell},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle^2$$

and $|\hat{\mathbf{f}}_{\hat\ell}-\mathbf{f}_\ell|^2\le2\|\hat f_{\hat\ell}-f\|^2+2\|f_\ell-f\|^2$. Thus we get

$$\begin{aligned}\mathbb{E}(\|\hat f_{\hat\ell}-f\|^2)\ &\le\ 3\|f_\ell-f\|^2+2\mathrm{pen}(\ell)+8\mathbb{E}\left(\sup_{\mathbf{t}\in S_{\hat\ell\vee\ell},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle^2-p(\ell,\hat\ell)\right)\\&\quad+\mathbb{E}(8p(\ell,\hat\ell)-2\mathrm{pen}(\hat\ell)).\end{aligned}\tag{24}$$

The following Proposition gives the appropriate choice for $p(\ell,\ell')$.

Proposition 6.1. *Let $p(\ell,\ell')=2\ell^*2^{4\ell*}/n$ with $\ell^*=\ell\vee\ell'$. Then, we have*

$$\mathbb{E}\left(\sup_{\mathbf{t}\in\mathbf{S}_{\hat\ell\vee\ell},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle^2-p(\ell,\hat\ell)\right)_+\le\frac{C'}{n}$$

The result of Proposition 6.1 inserted in Inequality (24), shows that for $\kappa\ge8$, we obtain $4p(\ell,\hat\ell)\le\mathrm{pen}(\hat\ell)+\mathrm{pen}(\ell)$ and

$$\mathbb{E}(\|\hat f_{\hat\ell}-f\|^2)\le3\|f_\ell-f\|^2+4\mathrm{pen}(\ell)+\frac{8C'}{n}$$

which is the result of Proposition 2.2. □

Proof of Proposition 6.1. We apply the Talagrand Inequality recalled in Lemma 8.1 of Section 8. First note that

$$\begin{aligned}&\mathbb{E}\left(\sup_{\mathbf{t}\in\mathbf{S}_{\hat\ell\vee\ell},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle^2-p(\ell,\hat\ell)\right)_+\\\le&\sum_{\ell'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{\mathbf{t}\in\mathbf{S}_{\ell^*},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat\alpha}_L-\vec\alpha_L)\rangle^2-p(\ell,\ell')\right)_+.\end{aligned}$$

Let us define $\|M\|^2=\mathrm{Tr}({}^tMM)$ and $\rho^2(M)$ the largest eigenvalue of ${}^tMM$. We consider the centered empirical process given by

$$\nu_n(\mathbf{t})=\frac1n\sum_{i=1}^n\langle\mathbf{t},\Omega_L^{-1}(\vec\beta_{i,L}-\vec\alpha_L)\rangle=\frac1n\sum_{i=1}^n(\psi_{\mathbf{t}}(\vec\beta_{i,L})-\mathbb{E}\psi_{\mathbf{t}}(\vec\beta_{i,L}))$$

where ${}^t\vec\beta_{i,L}=(1_{N_i(C_i\Delta)=0},\dots,1_{N_i(C_i\Delta)=L})$ are $L+1$-dimensional i.i.d. vectors and $\psi_{\mathbf{t}}(\vec x)=\langle\mathbf{t},\Omega_L^{-1}\vec x\rangle$. If $\mathbf{t}$ is in $\mathbf{S}_\ell$, $\nu_n(\mathbf{t})=\langle\mathbf{t},\Omega_\ell^{-1}(\vec{\hat\alpha}_\ell-\vec\alpha_\ell)\rangle$.

Recall that $\ell^*=\ell\vee\ell'$ and define the unit ball for the maximization by $\mathbf{B}_{\ell^*}=\{\mathbf{t}\in S_{\ell^*},|\mathbf{t}|=1\}$.

To apply Lemma 8.1, we specify ϵ, H^2, M and v^2.

Clearly

$$\mathbb{E}\left(\sup_{\mathbf{t}\in\mathbf{B}_{\ell^*}}\nu_n^2(\mathbf{t})\right)\leq\mathbb{E}(|\Omega_{\ell^*}^{-1}(\vec{\hat{\alpha}}_{\ell^*}-\vec{\alpha}_{\ell^*})|^2)\leq\frac{16}{15}\,\frac{2^{4\ell^*}}{n}:=H^2.$$

This bound was obtained in the computation of (20) (see (19), (21), (22)).

Next since $\vec{\beta}_{i,L}$ has only one nonzero coordinate, equal to 1, we have to bound $\psi_{\mathbf{t}}(\vec{x})=\langle\mathbf{t},\Omega_L^{-1}\vec{x}\rangle$ for $\vec{x}=e_j$ vector of the canonical basis of $\mathbb{R}^{L+1}$, with $j\leq\ell^*$ and $\mathbf{t}\in\mathbf{B}_{\ell^*}$. For such vectors $\vec{x}$,

$$|\psi_{\mathbf{t}}(\vec{x})|\leq\rho(\Omega_{\ell^*}^{-1})\leq\|\Omega_{\ell^*}^{-1}\|\leq\sqrt{16/15}\;2^{2\ell^*}:=M.$$

Lastly

$$\sup_{\mathbf{t}\in\mathbf{B}_{\ell^*}}\mathrm{Var}(\psi_t(\vec{\beta}_{i,L}))\leq\rho^2(\Omega_{\ell^*}^{-1})\mathbb{E}(\|\vec{\beta}_{i,L}\|^2)\leq\rho^2(\Omega_{\ell^*}^{-1})\leq\|\Omega_{\ell^*}^{-1}\|^2\leq\frac{15}{16}2^{4\ell^*}:=v^2$$

as $\mathbb{E}(\|\vec{\beta}_{i,L}\|^2)=\mathbb{E}(\sum_{k=0}^L 1^2_{N_i(C_i)=k})=\mathbb{P}(N_i(C_i)\in\{0,1,\dots,L\})\leq 1$.

We have $nH/M=\sqrt{n}$ and $nH^2/v^2=1$. We take $\epsilon^2=\delta\ell^*$ and for δ to be chosen afterwards, we get

$$\begin{aligned}\mathbb{E}\left(\sup_{\mathbf{t}\in\mathbf{S}_{\ell^*},|\mathbf{t}|=1}\langle\mathbf{t},\Omega_L^{-1}(\vec{\hat{\alpha}}_L-\vec{\alpha}_L)\rangle^2-2(1+2\delta\ell^*)H^2\right)_+ &\leq \frac{C_1}{n}\left(2^{4\ell^*}e^{-C_2\delta\ell^*}+e^{-C_3\delta\ell^*\sqrt{n}+4\ell^*\log(2)}\right)\\ &\leq \frac{C_1}{n}\left(2^{4\ell^*}e^{-C_2\delta\ell^*}+e^{-C_3\delta\ell^*\sqrt{n}/2}\right),\end{aligned}$$

where C_1,C_2,C_3 are numerical constants, provided that $\delta\geq 8\log(2)/C_3$. Then choosing $\delta\geq\max(\log(2)/C_2+1,8\log(2)/C_3)$ and $\ell^*\geq 1$ gives the result. □

6.6. **Proof of Proposition 3.1.**

Lemma 6.1. *$\forall x\geq 0$, $|\varphi_k(x)|\leq\sqrt{2}$, $|\varphi_k'(x)|\leq\sqrt{2}(2k+1)\leq 2\sqrt{2}(k+1)$ and $|\varphi_k''(x)|\leq 2\sqrt{2}(k+1)^2$.*
As a consequence, $\sum_{k=0}^\ell\varphi_k^2(x)\leq 2(\ell+1)$, $\sum_{k=0}^\ell[\varphi_k'(x)]^2\leq 8(\ell+1)^3$, $\sum_{k=0}^\ell[\varphi_k''(x)]^2\leq 8(\ell+1)^5$.

Proof of Lemma 6.1. The proof uses the Laguerre polynomials L_k^α, see Section 7, and relies on the relations $[L_k^\alpha(x)]'=-L_{k-1}^{\alpha+1}(x)$ and the bound (34). Recall that $\varphi_k(x)=\sqrt{2}L_k(2x)e^{-x}=\sqrt{2}L_k^0(2x)e^{-x}$. Bound (34) implies straightforwardly that $|\varphi_k(x)|\leq\sqrt{2}$, $\forall x\geq 0$. The second bound is obtained by writing that $\varphi_k'(x)=\sqrt{2}(2L_k'(2x)-L_k(2x))e^{-x}$and $|L_k'(x)|=|-L_{k-1}^1(x)|\leq ke^{x/2}$ and the third one by computing $\varphi_k''(x)=\sqrt{2}(4L_k''(2x)-4L_k'(x)+L_k(2x))e^{x/2}$ and $|L_k''(x)|=|-[L_{k-1}^1(x)]'|=|L_{k-2}^2(x)|\leq k(k-1)e^{x/2}/2$. □

First, by Pythagoras, $\|\tilde{f}_\ell^{(T)}-f\|^2=\|\tilde{f}_\ell^{(T)}-\mathbb{E}(\tilde{f}_\ell^{(T)})+\mathbb{E}(\tilde{f}_\ell^{(T)})-f_\ell\|^2+\|f_\ell-f\|^2$ and next,

$$\mathbb{E}(\|\tilde{f}_\ell^{(T)}-f\|^2)=\mathbb{E}(\|\tilde{f}_\ell^{(T)}-\mathbb{E}(\tilde{f}_\ell^{(T)})\|^2)+\|\mathbb{E}(\tilde{f}_\ell^{(T)})-f_\ell\|^2+\|f_\ell-f\|^2. \tag{25}$$

We have

$$\mathbb{E}(\|\tilde{f}_\ell^{(T)}-\mathbb{E}(\tilde{f}_\ell^{(T)})\|^2)=\mathbb{E}\left(\sum_{j=0}^\ell(\tilde{\theta}_j-\mathbb{E}(\tilde{\theta}_j))^2\right)=\frac{1}{n}\sum_{j=0}^\ell\mathrm{Var}\left(\varphi_j(\widehat{C}_{1,T})\right)$$

which yields with Lemma 6.1,

$$\mathbb{E}(\|\tilde{f}_\ell^{(T)}-\mathbb{E}(\tilde{f}_\ell^{(T)})\|^2)\leq\frac{1}{n}\mathbb{E}(\sum_{j=0}^\ell\varphi_j^2(\widehat{C}_{1,T}))\leq\frac{2(\ell+1)}{n}. \tag{26}$$

Note that, for N a Poisson variable with parameter λ, $\mathbb{E}((N-\lambda)^4) = \lambda(1+3\lambda)$. This implies

$$\mathbb{E}\left[(C_1 - \widehat{C}_{1,T})^4\right] = \frac{1}{T^4}\mathbb{E}\left(C_1 T(1+3C_1T)\right) = \frac{s_2}{T^2}. \tag{27}$$

Now, for some $\xi_T \in (C_1, \widehat{C}_{1,T})$, using Lemma 6.1 and (27), we get

$$\begin{aligned}
\|\mathbb{E}(\tilde{f}_\ell^{(T)}) - f_\ell\|^2 &= \sum_{j=0}^{\ell}\left[\mathbb{E}(\varphi_j(\widehat{C}_{1,T}) - \varphi_j(C_1))\right]^2 = \sum_{j=0}^{\ell}[\mathbb{E}((\widehat{C}_{1,T}-C_1)^2\varphi_j''(\xi_T))]^2 \\
&\leq \mathbb{E}\left\{(\widehat{C}_{1,T}-C_1)^4\sum_{j=0}^{\ell}[\varphi_j''(\xi_T)]^2\right\} \leq 8\frac{(\ell+1)^5}{T^2}s_2. \qquad (28)
\end{aligned}$$

Gathering (25), (26) and (28) yields the result. □

6.7. **Proof of Theorem 3.1.** Let

$$\tau_n(t) = \frac{1}{n}\sum_{j=1}^{n}[t(\widehat{C}_{j,T}) - \langle t, f\rangle] := \tilde{\nu}_n(t) + R(t),$$

$$\tilde{\nu}_n(t) = \frac{1}{n}\sum_{j=1}^{n}[t(\widehat{C}_{j,T}) - \mathbb{E}(t(\widehat{C}_{j,T}))], \qquad R(t) = \mathbb{E}[t(\widehat{C}_{1,T})] - \langle t, f\rangle.$$

Let $\tilde{\gamma}_n(t) = \|t\|^2 - 2n^{-1}\sum_{j=1}^{n} t(\widehat{C}_{j,T})$. Remark that $\tilde{f}_\ell^{(T)} = \arg\min_{t\in S_\ell}\tilde{\gamma}_n(t)$ and $\tilde{\gamma}_n(\tilde{f}_\ell^{(T)}) = -\|\tilde{f}_\ell^{(T)}\|^2$. Moreover we have

$$\tilde{\gamma}_n(t) - \tilde{\gamma}_n(s) = \|t-f\|^2 - \|s-f\|^2 - 2\tau_n(t-s)$$

and by definition of the penalty, $\forall \ell \in \mathcal{M}_{n,T}$, $\tilde{\gamma}_n(\tilde{f}_{\tilde{\ell}}^{(T)}) + \widetilde{\mathrm{pen}}(\tilde{\ell}) \leq \tilde{\gamma}_n(f_\ell) + \widetilde{\mathrm{pen}}(\ell)$. Therefore

$$\|\tilde{f}_{\tilde{\ell}}^{(T)} - f\|^2 \leq \|f_\ell - f\|^2 + \widetilde{\mathrm{pen}}(\ell) + 2\tau_n(\tilde{f}_{\tilde{\ell}}^{(T)} - f_\ell) - \widetilde{\mathrm{pen}}(\tilde{\ell}). \tag{29}$$

Using that $t \mapsto \tau_n(t)$ is linear and $2xy \leq x^2/4 + 4y^2$, we get

$$2\tau_n(\tilde{f}_{\tilde{\ell}}^{(T)} - f_\ell) \leq 2\|\tilde{f}_{\tilde{\ell}}^{(T)} - f_\ell\| \sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tau_n(t)| \leq \frac{1}{4}\|\tilde{f}_{\tilde{\ell}}^{(T)} - f_\ell\|^2 + 4\sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tau_n(t)|^2,$$

where $B_\ell = \{t \in S_\ell, \|t\| = 1\}$. Plugging this in (29) and using that $\|\tilde{f}_{\tilde{\ell}}^{(T)} - f_\ell\|^2 \leq 2\|\tilde{f}_{\tilde{\ell}}^{(T)} - f\|^2 + 2\|f - f_\ell\|^2$, we get

$$\begin{aligned}
\|\tilde{f}_{\tilde{\ell}}^{(T)} - f\|^2 &\leq 3\|f_\ell - f\|^2 + 2\widetilde{\mathrm{pen}}(\ell) + 8\sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tau_n(t)|^2 - 2\widetilde{\mathrm{pen}}(\tilde{\ell}) \\
&\leq 3\|f_\ell - f\|^2 + 2\widetilde{\mathrm{pen}}(\ell) + 16\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tilde{\nu}_n(t)|^2 - p_1(\ell,\tilde{\ell})\right)_+ \\
&\quad +16\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|R(t)|^2 - p_2(\ell,\tilde{\ell})\right)_+ + 16p_1(\ell,\tilde{\ell}) + 16p_2(\ell,\tilde{\ell}) - 2\widetilde{\mathrm{pen}}(\tilde{\ell}).
\end{aligned}$$

We define $p_i(\ell,\ell')$ in the following results:

Proposition 6.2. *Assume that* $\mathbb{E}(C_1^8) < +\infty$. *Define* $p_1(\ell,\ell') = 4(\ell\vee\ell'+1)/n$, $p_2(\ell,\ell') = 8\hat{s}_2(\ell\vee\ell'+1)^5/T^2$. *Then*

$$\mathbb{E}\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tilde{\nu}_n(t)|^2 - p_1(\ell,\tilde{\ell})\right)_+ \leq \frac{c}{n}, \qquad \mathbb{E}\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|R(t)|^2 - p_2(\ell,\tilde{\ell})\right)_+ \leq \frac{c'}{n}$$

where c, c' *are positive constants.*

The proof of Proposition 6.2 is given in Section 6.8. Now, the definitions of p_1, p_2 and $\widetilde{\text{pen}}(.)$ imply that

$$8p_1(\ell,\ell') + 8p_2(\ell,\ell') \leq \widetilde{\text{pen}}(\ell) + \widetilde{\text{pen}}(\ell')$$

for $\tilde{\kappa}_1 \geq 32$ and $\tilde{\kappa}_2 \geq 64$, $\forall \ell,\ell' \in \mathcal{M}_{n,T}$. Therefore, we obtain

$$\|\tilde{f}_{\tilde{\ell}}^{(T)} - f\|^2 \leq 3\|f_\ell - f\|^2 + 4\widetilde{\text{pen}}(\ell) + \frac{c"}{n}$$

which ends the proof of Theorem 3.1. □

6.8. **Proof of Proposition 6.2.** First we study $\tilde{\nu}_n(t)$ and apply the Talagrand Inequality. To do this, we evaluate the bounds H^2, M, v as defined in Lemma 8.1. Clearly

$$\mathbb{E}\left(\sup_{t\in B_{\ell'\vee\ell}}|\tilde{\nu}_n(t)|^2\right) \leq \sum_{k=0}^{\ell\vee\ell'}\text{Var}(\tilde{\nu}_n(\varphi_k)) = \frac{1}{n}\sum_{k=0}^{\ell\vee\ell'}\text{Var}(\varphi(\widehat{C}_{1,T})) \leq \frac{2(1+\ell\vee\ell')}{n} := H^2$$

by Lemma 6.1. Moreover, using (15), on $B_{\ell\vee\ell'}$, $\|t\|_\infty \leq \sqrt{2(\ell\vee\ell'+1)} := M$. Next, to find v, we split in two parts:

$$\sup_{t\in B_{\ell'\vee\ell}}\text{Var}(t(\widehat{C}_{1,T})) \leq 2(T_1+T_2)$$

where

$$T_1 := \sup_{t\in B_{\ell'\vee\ell}}\mathbb{E}(t^2(C_{1,T})) \leq \sup_{t\in B_{\ell'\vee\ell}}\|t\|_\infty\left(\int t^2\int f^2\right)^{1/2} \leq \sqrt{2(1+\ell\vee\ell')}\|f\|$$

and

$$T_2 := \sup_{t\in B_{\ell'\vee\ell}}\mathbb{E}[(t(\widehat{C}_{1,T}) - t(C_{1,T}))^2].$$

We write that

$$(t(\widehat{C}_{1,T}) - t(C_{1,T}))^2 = (\widehat{C}_{1,T} - C_{1,T})^2[t'(\xi_T)]^2 \leq (\widehat{C}_{1,T} - C_{1,T})^2\sum_{k=0}^{\ell}(\varphi_k'(\xi_T))^2$$

where we apply the Taylor Formula and $\xi_T \in (C_1, \widehat{C}_{1,T})$. Using Lemma 6.1 again, we get

$$T_2 \leq \mathbb{E}[(\widehat{C}_{1,T} - C_{1,T})^2]8(1+\ell\vee\ell')^3.$$

To conclude we use that $\mathbb{E}[(\widehat{C}_{1,T} - C_{1,T})^2] = \mathbb{E}(C_1)/T$ and that by definition of $\mathcal{M}_{n,T}$, $(1+\ell\vee\ell')^2/T \leq \sqrt{1+\ell\vee\ell'}$. Therefore, we obtain $v = C\sqrt{1+\ell\vee\ell'}$. Now the Talagrand Inequality implies that there exist constants A_i, $i = 1,2,3$ such that

$$\mathbb{E}\left(\sup_{t\in B_{\ell'\vee\ell}}|\tilde{\nu}_n(t)|^2 - \frac{8(1+\ell\vee\ell')}{n}\right)_+ \leq \frac{A_1}{n}\left((\ell\vee\ell')e^{-A_2\sqrt{\ell\vee\ell'}} + \frac{\ell\vee\ell'}{n}e^{-A_3\sqrt{n}}\right)$$

so that as

$$\mathbb{E}\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|\tilde{\nu}_n(t)|^2-p_1(\ell,\tilde{\ell})\right)_+\leq\sum_{\ell'\in\mathcal{M}_{n,T}}\mathbb{E}\left(\sup_{t\in B_{\ell'\vee\ell}}|\tilde{\nu}_n(t)|^2-\frac{8(1+\ell\vee\ell')}{n}\right)_+\leq c/n$$

which is the announced bound.

Now we study $R(t)$. Let $D=\left(\sup_{t\in B_{\tilde{\ell}\vee\ell}}|R(t)|^2-p_2(\ell,\tilde{\ell})\right)_+$.

$$\begin{aligned}\mathbb{E}(D) &\leq \sum_{\ell'\in\mathcal{M}_{n,T}}\mathbb{E}\left(\sum_{j=0}^{\ell'\vee\ell}\left\{\mathbb{E}[\varphi_j(\widehat{C}_{1,T})-\varphi_j(C_1)]\right\}^2-4\frac{s_2(1+\ell'\vee\ell)^5}{T^2}\right)_+\\ &\quad+\mathbb{E}\left(8(\frac{s_2}{2}-\hat{s}_2)_+\frac{(1+\tilde{\ell}\vee\ell)^5}{T^2}\right)\end{aligned}$$

By Inequality (28), the first rhs term is zero. To deal with the second term, let $\Omega=\{|\widehat{s}_2-s_2|\leq s_2/2\}$. Using the definition of $\mathcal{M}_{n,T}$, we get

$$\mathbb{E}(D)\leq\mathbb{E}(8\mathbb{1}_{\Omega^c}(\frac{s_2}{2}-\hat{s}_2)_+).$$

since $(\frac{1}{2}s_2-\widehat{s}_2)_+1_\Omega=0$. By the Markov inequality, we have $\mathbb{P}(\Omega^c)\leq(2/s_2)^4\mathbb{E}(|\widehat{s}_2-s_2|^4)$ and we use the Rosenthal Inequality (see Hall and Heyde (1980, p.23)) to get

$$\mathbb{E}(|\widehat{s}_2-s_2|^4)\leq C_p(n^{-3}m_4^4+n^{-2}m_2^4)$$

where m_4 is the fourth centered moment of $X_j=3\widehat{C}_{j,T}^2-2\widehat{C}_{j,T}/T$ and m_2^2 the variance of X_j. We write

$$X_j-\mathbb{E}(X_j)=3(\widehat{C}_{j,T}-C_j)^2+3(C_j^2-\mathbb{E}(C_j^2))+6(C_j-\frac{2}{T})(\widehat{C}_{j,T}-C_j)-\frac{2}{T}(C_j-\mathbb{E}(C_j))+\frac{3}{T}\mathbb{E}(C_j).$$

After some elementary computations using the centered moments of a Poisson distribution, we obtain that, if $\mathbb{E}(C_j^8)<+\infty$, then there exist constants c_1,c_2 such that $m_4^4\leq c_1$ and $m_2^2\leq c_2$. Finally $\mathbb{E}(D)\leq c/n$.

7. Sobolev-Laguerre spaces

7.1. **Laguerre polynomials and associated regularity spaces: General properties.** For $\rho:\mathbb{R}^+\to\mathbb{R}^+$ a Borel function, let

$$\mathbb{L}^2(\mathbb{R}^+,\rho)=\{g:\mathbb{R}^+\to\mathbb{R},\int_0^{+\infty}g^2(x)\rho(x)dx:=\|g\|_\rho^2<+\infty\}.$$

When $\rho\equiv1$, we denote this space as usual by $\mathbb{L}^2(\mathbb{R}^+)$ with $\|g\|^2=\int_0^{+\infty}g^2(x)dx$. Obviously, $g\in\mathbb{L}^2(\mathbb{R}^+,\rho)$ is equivalent to $g\sqrt{\rho}\in\mathbb{L}^2(\mathbb{R}^+)$ and $\|g\|_\rho=\|g\sqrt{\rho}\|$. For any orthonormal basis (ϕ_k^ρ) of $\mathbb{L}^2(\mathbb{R}^+,\rho)$, $(\sqrt{\rho}\phi_k^\rho)$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. We are especially interested in the weight functions

$$\rho(x)=x^\alpha e^{-x}=w_\alpha(x),\quad\alpha\geq0 \tag{30}$$

and the associated orthonormal bases of $\mathbb{L}^2(\mathbb{R}^+,w_\alpha)$, namely the Laguerre polynomials. Consider the second order differential equation:

$$L_\alpha g=-kg,\quad\text{with}\quad L_\alpha g=xg''+(\alpha+1-x)g'. \tag{31}$$

The solution is $g(x) = L_k^\alpha(x)$ the Laguerre polynomial with index α and order k. The function L_k^α is a polynomial of degree k, and the sequence (L_k^α) is orthogonal with respect to the weight function w_α. The orthogonality relations are equivalent to:

$$\int_0^{+\infty} x^\ell L_k^\alpha(x) w_\alpha(x) dx = 0 \quad \text{for} \quad k > \ell. \tag{32}$$

We have

$$L_k^\alpha(x) = \frac{1}{k!} e^x x^{-\alpha} \frac{d^k}{dx^k}\left(x^{k+\alpha} e^{-x}\right), \ \ (L_k^\alpha(x))' = -L_{k-1}^{\alpha+1}(x). \tag{33}$$

The following holds, for all integer k and $\alpha \geq 0$:

$$\int_0^{+\infty} \left(L_k^\alpha(x)\right)^2 w_\alpha(x) dx = \frac{\Gamma(k+\alpha+1)}{k!}, \quad \forall x, \ \ |L_k^\alpha(x)| \leq \frac{\Gamma(k+\alpha+1)}{k!\Gamma(\alpha+1)} e^{x/2}. \tag{34}$$

Setting

$$\phi_k^\alpha(x) = L_k^\alpha(x) \left(\frac{k!}{\Gamma(k+\alpha+1)}\right)^{1/2}, \tag{35}$$

the sequence $(\phi_k^\alpha), k \geq 0)$ constitutes an orthonormal basis of the space $L^2((0,+\infty), w_\alpha)$. In particular, $\phi_k^0(x) = L_k^0(x) = L_k(x), k \geq 0$ constitute an orthonormal basis of $\mathbb{L}^2((0,+\infty), w)$, with $w(x) = w_0(x) = e^{-x}$. Noting that $\left(x^{\alpha+1} e^{-x}\right)' = x^\alpha e^{-x}(\alpha + 1 - x)$, we obtain, using (31) and (33),

$$\frac{d}{dx}\left(x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x)\right) = x^\alpha e^{-x} k L_k^\alpha(x). \tag{36}$$

For these formulas, see Abramowitz and Stegun (1964).

We can now prove the following result.

Proposition 7.1. *For s integer, $w(x) = e^{-x}$ and $g : (0,+\infty) \to \mathbb{R}$, the following two statements are equivalent:*

1. *g admits derivatives up to order $s-1$, $g^{(s-1)}$ is absolutely continuous and for $0 \leq m \leq s$, $x^{m/2} g^{(m)}$ belongs to $L^2((0,+\infty), w)$ ($g^{(s)}$ is the Radon-Nikodym derivative of $g^{(s-1)}$).*
2. *g belongs to $L^2((0,+\infty), w)$ and*

$$\sum_{k \geq 0} k^s \tau_k^2(g) < +\infty, \tag{37}$$

where $\tau_k(g) = \int_0^{+\infty} g(x) L_k(x) w(x) dx$ is the k-th component of g on the basis $(L_{k'}, k' \geq 0)$ of $\mathbb{L}^2((0,+\infty), w)$.

For all $m = 0, \ldots, s$, $\|x^{m/2} g^{(m)}\|_w^2 = \sum_{k \geq m} k(k-1)\ldots(k-m+1)\tau_k^2(g)$. If π_ℓ denotes the orthogonal projection of g on the space spanned in $L^2((0,+\infty), w)$ by $(L_k, k \leq \ell)$,

$$\|g - \pi_\ell g\|_w^2 \leq \frac{1}{\ell(\ell-1)\ldots(\ell-s+1)} \|x^{s/2} g^{(k)}\|_w^2.$$

We can now define for $s \geq 0$, the Sobolev-Laguerre space[4] with weight function w by:

$$W^s((0,+\infty), w) = \{g \in L^2((0,+\infty), w), \sum_{k \geq 0} k^s \tau_k^2(g) < +\infty\}. \tag{38}$$

[4]Bongioanni and Torrea (2009) introduce Sobolev-Laguerre spaces but do not establish the link with the coefficients of a function on a Laguerre basis.

Consider, for $a>0$, the space $L^2(\mathbb{R}^+,w(a.))$ corresponding to the weight function $w(ax)=e^{-ax}$. The sequence $(\sqrt{a}L_k(at))$ is an orthonormal basis of $L^2(\mathbb{R}^+,w(a.))$. Setting $g_a(t)=(t/\sqrt{a})g(t/a))$,

$$\tau_{k,a}(g):=\int_0^{+\infty}g(x)\sqrt{a}L_n(ax)w(ax)dx=\tau_k(g_a).$$

So we can define

$$W^s((0,+\infty),w(a.))=\{g\in L^2((0,+\infty),w(a.)),\sum_{k\geq 0}k^s\tau_{k,a}^2(g)<+\infty\}. \tag{39}$$

For s integer, $g\in W^s((0,+\infty),w(a.))$ is equivalent to $g\in L^2((0,+\infty),w(a.))$ and g admits derivatives up to order $s-1$, $g^{(s-1)}$ is absolutely continuous and for $0\leq m\leq s$, $x^{m/2}g^{(m)}$ belongs to $L^2((0,+\infty),w(a.))$.

Let us now interpret the result of Proposition 7.1 in terms of bases of $\mathbb{L}^2((0,+\infty))$. The Laguerre functions are defined using the normalized Laguerre polynomials by

$$\mathcal{L}_k^\alpha(x)=e^{-x/2}x^{\alpha/2}\phi_k^\alpha(x), \tag{40}$$

where ϕ_k^α is defined in (35). The sequence $(\mathcal{L}_k^\alpha,k\geq 0)$ is an orthonormal basis of $\mathbb{L}^2((0,+\infty))$. With L_α given in (31), we have

$$x^{\alpha/2}e^{-x/2}L_\alpha(e^{x/2}x^{-\alpha/2}f)=-\mathcal{L}_\alpha f+\frac{\alpha+1}{2}f,\quad\text{with}\quad\mathcal{L}_\alpha f=-xf''-f'+(\frac{x}{4}+\frac{\alpha^2}{4x})f.$$

Now,

$$f\in\mathbb{L}^2((0,+\infty))\Longleftrightarrow g=fe^{(x/2)}\in\mathbb{L}^2((0,+\infty),w)$$

and

$$\tau_k(g)=\theta_k^0(f):=\int_0^{+\infty}f(x)\mathcal{L}_k^0(x)dx.$$

We can thus set:

$$W^s((0,+\infty))=\{f\in\mathbb{L}^2((0,+\infty)),\sum_{k\geq 0}k^s(\theta_k^0(f))^2<+\infty\}. \tag{41}$$

We have to deduce the properties of $W^s((0,+\infty))$ from those of $W^s((0,+\infty),w)$. Using that

$$f\in\mathbb{L}^2((0,+\infty))\Longleftrightarrow g_\alpha\in\mathbb{L}^2((0,+\infty),w_\alpha),\quad\text{with}\quad g_\alpha=fx^{-\alpha/2}e^{x/2}$$

and the fact that f is abolutely continuous if and only if g_α is, a simple computation yields,

$$g_\alpha'e^{-x/2}x^{(\alpha+1)/2}=\delta^\alpha f\quad\text{where}\quad\delta^\alpha f=\sqrt{x}f'+\frac{1}{2}(\sqrt{x}-\frac{\alpha}{\sqrt{x}})f. \tag{42}$$

Observing that $\tau_k^\alpha(g_\alpha)=\int_0^{+\infty}g_\alpha(x)\phi_k^\alpha(x)dx=\theta_k^\alpha(f)=\int_0^{+\infty}f(x)\mathcal{L}_k^\alpha(x)dx$, we get $\tau_{k-1}^{\alpha+1}(g_\alpha')=\theta_{k-1}^{\alpha+1}(\delta^\alpha f)$ and

$$g_\alpha'\in\mathbb{L}^2((0,+\infty),w_{\alpha+1})\Longleftrightarrow x^{(\alpha+1)/2}g_\alpha'\in\mathbb{L}^2((0,+\infty),w)\Longleftrightarrow\delta^\alpha f\in\mathbb{L}^2((0,+\infty)).$$

We can state:

Proposition 7.2. *For s integer, the following properties are equivalent:*

(1) $f\in W^s((0,+\infty))\Longleftrightarrow g=fe^{x/2}\in W^s((0,+\infty),w)$,

(2) $f \in \mathbb{L}^2((0,+\infty))$, $f', f'', \ldots, f^{(s-1)}$ *exist,* $f^{(s-1)}$ *is absolutely continuous and for* $m = 0, \ldots, s-1$, $\delta^m \circ \ldots \circ \delta^1 \circ \delta^0 f \in \mathbb{L}^2((0,+\infty))$, *where, with* δ^α *given in (42), we have*

$$\delta^m \circ \ldots \circ \delta^1 \circ \delta^0 f = x^{(m+1)/2} g^{(m+1)} e^{-x/2}$$

$$\text{and} \quad \sum_{k \geq m} k(k-1)\ldots(k-m+1)(\theta_k^0(f))^2 = \|\delta^{m-1} \circ \ldots \circ \delta^0 f\|^2. \tag{43}$$

The proof of the above proposition is simply deduced from Proposition 7.1 that is proved below.

It remains to interpret also the results for the scale changed bases $(\sqrt{a} L_k(ax) e^{(-ax/2)}, k \geq 0)$ of $\mathbb{L}^2((0,+\infty))$. For all $a > 0$ and $\alpha \geq 0$,

$$f \in \mathbb{L}^2((0,+\infty)) \Longleftrightarrow g_{\alpha,a} \in \mathbb{L}^2((0,+\infty), x^\alpha e^{-ax}), \quad \text{with} \quad g_{\alpha,a} = f x^{-\alpha/2} e^{ax/2},$$

and

$$x^{(\alpha+1)/2} e^{-ax/2} g'_{\alpha,a} = \delta_a^\alpha f, \quad \text{with} \quad \delta_a^\alpha f = \sqrt{x} f' + f\left(\frac{a\sqrt{x}}{2} - \frac{\alpha}{2\sqrt{x}}\right). \tag{44}$$

Noting that with $g = f e^{ax/2}$,

$$\tau_{k,a}^0(g) = \int_0^{+\infty} g(x)\sqrt{a} L_k(ax) e^{-ax} dx = \theta_{k,a}^0(f) = \int_0^{+\infty} f(x)\sqrt{a} L_k(ax) e^{(-ax/2)} dx,$$

we can set:

$$W_a^s((0,+\infty)) = \{f \in \mathbb{L}^2((0,+\infty)), \sum_{k \geq 0} k^s (\theta_{k,a}^0(f))^2\}. \tag{45}$$

We have

$$f \in W_a^s((0,+\infty)) \Longleftrightarrow g = f e^{ax/2} \in W^s((0,+\infty), e^{-ax})$$

and the statement analogous to Proposition 7.2 holds with δ_a^α instead of δ^α and $w(ax) = e^{-ax}$ instead of w.

Let us state the analogous of Proposition 7.2 with the scaled-changed basis corresponding to $a = 2$.

Proposition 7.3. *For* s *integer, the following properties are equivalent:*

(1) $f \in W_2^s((0,+\infty)) \Longleftrightarrow g = f e^x \in W^s((0,+\infty), w(2.))$,

(2) $f \in \mathbb{L}^2((0,+\infty))$, $f', f'', \ldots, f^{(s-1)}$ *exist,* $f^{(s-1)}$ *is absolutely continuous and for* $m = 0, \ldots, s-1$, $\delta_2^m \circ \ldots \circ \delta_2^1 \circ \delta_2^0 f \in \mathbb{L}^2((0,+\infty))$, *where, with* δ_a^α *given in (42), we have*

$$\delta_2^m \circ \ldots \circ \delta_2^1 \circ \delta_2^0 f = x^{(m+1)/2} (f e^x)^{(m+1)} e^{-x} = x^{(m+1)/2} \sum_{j=0}^{m+1} \binom{m+1}{j} f^{(j)}.$$

$$\text{and} \quad \sum_{k \geq m} k(k-1)\ldots(k-m+1)(\theta_{k,2}^0(f))^2 = \|\delta_2^{m-1} \circ \ldots \circ \delta_2^0 f\|^2. \tag{46}$$

In the text, we have set $\theta_{k,2}^0(f) = \theta_k(f)$ and $\varphi_k(t) = \sqrt{2} L_k(2t) e^{-t}$.

7.2. **Proof of Proposition 7.1.**

Proof. Recall that, for a function $g : (0, +\infty) \to \mathbb{R}$,

$$x^{m/2} g \in L^2((0,+\infty), w) \Longleftrightarrow g \in L^2((0,+\infty), w_m)$$

and $\|x^{m/2} g\|_w^2 = \int_0^{+\infty} x^m g^2(x) w(x) dx = \|g\|_{w_m}^2$.
We start by proving that (1) $\Rightarrow$ (2). For $h \in \mathbb{L}^2((0,+\infty), w_\alpha)$, let $\tau_k^\alpha(h) = \int_0^{+\infty} h(x)\phi_k^\alpha(x)dx$ denote the k-th component of h on the basis $(\phi_{k'}^\alpha = L_{k'}, k' \geq 0)$, and for $\alpha = 0$, $\tau_k^0(h) = \tau_k(h)$. The proof relies on the following Lemma:

Lemma 7.1. *Let $\alpha \geq 0$. If $g : (0,+\infty) \to \mathbb{R}$ is absolutely continuous with $x^{\alpha/2} g \in L^2((0,+\infty), w)$ and $x^{(\alpha+1)/2} g' \in L^2((0,+\infty), w)$, then for all $k \geq 1$, $\sqrt{k}\tau_k^\alpha(g) = -\tau_{k-1}^{\alpha+1}(g')$.*

Proof. By the assumption, g is continuous on $(0,+\infty)$. For $k \geq 1$, using (36) yields

$$\begin{aligned} k\int_0^{+\infty} g(x) L_k^\alpha(x) x^\alpha e^{-x} dx &= \int_0^{+\infty} g(x) \frac{d}{dx}\left(x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x)\right) dx \\ &= \left[g(x) x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x)\right]_0^{+\infty} - \int_0^{+\infty} g'(x) x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x) dx \end{aligned}$$

where the integrals are well-defined by assumption. We multiply both sides by $((k-1)!/\Gamma(k+\alpha+1))^{1/2}$. On the left-hand side, appears $\sqrt{k}\phi_k^\alpha$, on the right-hand side, $\phi_{k-1}^{\alpha+1}$. Hence, to get the result, it is enough to prove that $[\ldots]_0^{+\infty} = 0$. Using that $x^a \leq x^{a+1}$ for $x \geq 1$, we get $\int_1^{+\infty} e^{-x} g^2(x) x^{\alpha-1} dx < +\infty$, and

$$\left(\int_1^{+\infty} |g(x)g'(x)| x^\alpha e^{-x} dx\right)^2 \leq \int_1^{+\infty} g^2(x) x^\alpha e^{-x} dx \int_1^{+\infty} (g'(x))^2 x^\alpha e^{-x} dx < +\infty.$$

Thus,

$$\int_1^{+\infty} g^2(x) x^\alpha e^{-x} dx = -[g^2(x) x^\alpha e^{-x}]_1^{+\infty} + \int_1^{+\infty} e^{-x}(2g(x)g'(x)x^\alpha + \alpha g^2(x) x^{\alpha-1}) dx.$$

The integrals in the left-hand side and right-hand side above are finite. Therefore, the limit of $g^2(x)x^\alpha e^{-x}$ as x tends to infinity exists. As $\int_1^{+\infty} g^2(x) x^\alpha e^{-x} dx < +\infty$, this limit is necessarily equal to 0. This implies $\lim_{x\to+\infty} g(x) x^{\alpha/2} e^{-x/2} = 0$. Therefore,

$$\lim_{x\to+\infty} g(x) x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x) = 0.$$

The assumption on g implies $\int_0^{+\infty} |g(x)| x^\alpha e^{-x} dx < +\infty$ and $\int_0^{+\infty} |g'(x)| x^{\alpha+1} e^{-x} dx < +\infty$. Thus, $\int_0^1 |g(x)| x^\alpha dx < +\infty$ and $\int_0^1 |g'(x)| x^{\alpha+1} dx < +\infty$. We have:

$$\int_0^1 g(x) x^\alpha dx = \frac{1}{\alpha+1}[g(x)x^{\alpha+1}]_0^1 - \frac{1}{\alpha+1}\int_0^1 g'(x) x^{\alpha+1} dx.$$

Therefore, the limit of $g(x)x^{\alpha+1}$ as x tends to 0^+, exists and is finite. As $\int_0^1 x^\alpha |g(x)| dx < +\infty$, this limit is necessarily equal to 0. This implies

$$\lim_{x\to 0} g(x) x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x) = 0.$$

□

Now, let g satisfy (1). By the Lemma, $\sqrt{k}\tau_k(g) = -\tau^1_{k-1}(g')$, $\sqrt{k-1}\tau^1_{k-1}(g') = -\tau^2_{k-2}(g'')$ and so on. By elementary induction, we get for $m = 0, 1, \ldots, s$ and $k \geq m$,

$$(k(k-1)\ldots,(k-m+1))^{1/2}\tau_k(g) = (-1)^m\tau^m_{k-m}(g^{(m)}).$$

Therefore,

$$\sum_{k\geq 0} k(k-1)\ldots(k-s+1)\tau^2_k(g) = \sum_{k\geq 0}\left(\tau^s_k(g^{(s)})\right)^2 = \|g^{(s)}\|^2_{w_s} = \|x^{(s/2)}g^{(s)}\|_w < +\infty.$$

So we have (2). Moreover $\|g-\pi_\ell g\|^2_w \leq [\ell(\ell-1)\ldots(\ell-s+1)]^{-1}\|x^{(s/2)}g^{(s)}\|_w$. Let us prove that (2) $\Rightarrow$ (1). We have an analogous lemma.

Lemma 7.2. *Let $\alpha \geq 0$. Assume that $g : (0,+\infty) \to \mathbb{R}$ belongs to $L^2((0,+\infty), w_\alpha)$ and that $\sum_{k\geq 0} k\,(\tau^\alpha_k(g))^2 < +\infty$. Then, g is absolutely continuous, g' belongs to $L^2((0,+\infty), w_{\alpha+1})$ and for all $k \geq 1$, $\tau^{\alpha+1}_{k-1}(g') = -\sqrt{k}\tau^\alpha_k(g)$.*

Proof. We have $g = \sum_{k\geq 0}\tau^\alpha_k(g)\phi^\alpha_k$ with ϕ^α_0 a constant. Thus,

$$g(y) - g(x) = \sum_{k\geq 1}\tau^\alpha_k(g)\int_x^y(\phi^\alpha_k(t))'dt = -\sum_{k\geq 1}\sqrt{k}\tau^\alpha_k(g)\int_x^y\phi^{\alpha+1}_{k-1}(t)dt.$$

The function $h(t) = \sum_{k\geq 1}\sqrt{k}\tau^\alpha_k(g)\phi^{\alpha+1}_{k-1}(t)$ is well-defined and $h_N(t) = \sum_{k=1}^N\sqrt{k}\tau^\alpha_k(g)\phi^{\alpha+1}_{k-1}(t)$ converges to h in $\mathbb{L}^2((0,+\infty), w_{\alpha+1})$, thus in $L^1((0,+\infty), w_{\alpha+1})$ also. Consequently, for $0 < x \leq y$,

$$\inf_{u\in[x,y]}(u^{\alpha+1}e^{-u})\int_x^y|h_N(t)-h(t)|dt \leq \int_x^y|h_N(t)-h(t)|t^{\alpha+1}e^{-t}dt \to_{N\to+\infty} 0.$$

This implies $g(y) - g(x) = -\int_x^y h(t)dt$. Thus, g is absolutely continuous with $g' = h$ and $-\tau^{\alpha+1}_{k-1}(g') = \sqrt{k}\tau_k\alpha(g)$. As $\sum_{k\geq 0} k(\tau^\alpha_k(g))^2 < +\infty$, $g' \in \mathbb{L}^2((0,+\infty), w_{\alpha+1})$ which is equivalent to $t^{(\alpha+1)/2}g' \in \mathbb{L}^2((0,+\infty), w)$. □

Now, let g satisfy (2). Applying the lemma, we get that g is absolutely continuous and that $g' = -\sum_{k\geq 1}\sqrt{k}\tau_k(g)\phi^1_{k-1}(t)$ belongs to $\mathbb{L}^2((0,+\infty), w_1)$. Then, we have that g' is absolutely continuous with $g'' = (-1)^2\sum_{k\geq 2}\sqrt{k(k-1)}\tau_k(g)\phi^2_{k-2}(t)$ belonging to $\mathbb{L}^2((0,+\infty), w_2)$.

By induction, for $m = 0, \ldots, s$, $g^{(m)}$ belongs to $\mathbb{L}^2((0,+\infty), w_m)$ with

$$g^{(m)} = (-1)^m\sum_{k\geq m}(k(k-1)\ldots(k-m+1))^{1/2}\tau_k(g)\phi^m_{k-m}.$$

Thus, $t^{m/2}g^{(m)}$ belongs to $\mathbb{L}^2((0,+\infty), w)$ for $m = 0, \ldots, s$. So the proof of the proposition is complete. □

8. A useful inequality.

We recall the Talagrand inequality. The result below follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 8.1. *(Talagrand Inequality) Let $Y_1, \ldots, Y_n$ be independent random variables, let $\nu_{n,Y}(f) = (1/n)\sum_{i=1}^n[f(Y_i) - \mathbb{E}(f(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2 > 0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2 - 2(1+2\epsilon^2)H^2\Big]_+ \quad\leq\quad \frac{4}{K_1}\left(\frac{v^2}{n}e^{-K_1\epsilon^2\frac{nH^2}{v^2}} + \frac{98M^2}{K_1n^2C^2(\epsilon^2)}e^{-\frac{2K_1C(\epsilon^2)\epsilon}{7\sqrt{2}}\frac{nH}{M}}\right),$$

with $C(\epsilon^2) = \sqrt{1+\epsilon^2} - 1$, $K_1 = 1/6$, *and*

$$\sup_{f\in\mathcal{F}} \|f\|_\infty \le M, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_{n,Y}(f)|\Big] \le H, \ \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{k=1}^{n} \mathrm{Var}(f(Y_k)) \le v^2.$$

By standard density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \mapsto \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

[1] 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, Washington, D.C.

[2] Birgé, L., and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.

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

[4] Comte, F., Cuenod, C.-A., Pensky, M. and Rozenholc Y. (2013). Laplace deconvolution on the basis of time domain data and its application to Dynamic Contrast Enhanced imaging. Preprint MAP5 2012-17.

[5] Comte, F., Duval, C. and Genon-Catalot, V. (2014). Nonparametric density estimation in compound Poisson process using convolution power estimators. *Metrika* **77**, 163-183.

[6] Comte, F. and Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *J. Statist. Plann. Inference* **142**, 1698-1715.

[7] Fabio, L. C., Paula, G. A. and de Castro, M. (2012). A Poisson mixed model with nonnormal random effect distribution. *Comput. Statist. Data Anal.* **56**, 1499-1510.

[8] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, no. 3, 1257-1272.

[9] Grandell, J. (1997). *Mixed Poisson processes.* Monographs on Statistics and Applied Probability, 77. Chapman & Hall, London.

[10] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its applications.* Academic press, Inc. (London) LTD.

[11] Hengartner, N. W. (1997) Adaptive demixing in Poisson mixture models. *Ann. Statist.* **25**, 917-928.

[12] Karr, A. F. (1984). Combined nonparametric inference and state estimation for mixed Poisson processes. *Z. Wahrsch. Verw. Gebiete* **66**, 81-96.

[13] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.

[14] Loh, W.-L. and Zhang, C.-H. (1996). Global properties of kernel estimators for mixing densities in discrete exponential family models. *Statist. Sinica* **6**, 561-578.

[15] Loh, W.-L. and Zhang, C.-H. (1997). Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Statist.* **24**, 15-32.

[16] Massart, P. (2007). *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. With a foreword by Jean Picard. Lecture Notes in Mathematics, 1896. Springer, Berlin

[17] Mikosch, T. (2009). *Non-life insurance mathematics. An introduction with the Poisson process.* Second edition. Universitext. Springer-Verlag, Berlin.

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

[19] Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4**, 1200-1209.

[20] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.

[21] Zhang, C.-H. (1995). On estimating mixing densities in discrete exponential family models. *Ann. Statist.* **23**, 929-945.