



**HAL**  
open science

# Nonparametric density estimation for mixed Poisson processes

Fabienne Comte, Valentine Genon-Catalot

► **To cite this version:**

Fabienne Comte, Valentine Genon-Catalot. Nonparametric density estimation for mixed Poisson processes. 2013. hal-00848158v1

**HAL Id: hal-00848158**

**<https://hal.science/hal-00848158v1>**

Preprint submitted on 25 Jul 2013 (v1), last revised 13 Mar 2014 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# NONPARAMETRIC DENSITY ESTIMATION FOR MIXED POISSON PROCESSES.

F. COMTE<sup>1</sup> AND V. GENON-CATALOT<sup>1</sup>

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  $\mathbb{R}^+$ . Depending on the observation time, we propose two nonparametric adaptive strategies to estimate  $f$ . We use an appropriate Laguerre basis to build adaptive projection estimators and also propose kernel estimators with adaptive bandwidths. Non-asymptotic bounds of the  $\mathbb{L}^2$ -integrated risk are obtained in each case. The procedures are illustrated on simulated data. July 25, 2013

**Keywords.** Mixed Poisson process. Inverse problem. Nonparametric estimation. Adaptive estimators.  
**AMS Classification.** 62G07 - 62M09.

## 1. INTRODUCTION

Consider  $n$  independent Poisson processes  $(N_j(t), j = 1, \dots, n)$  with unit intensity and  $n$  i.i.d. positive random variables  $(C_j, j = 1, \dots, n)$ . Assume that the processes  $(N_j(t), j = 1, \dots, n)$  and the sequence  $(C_j, j = 1, \dots, n)$  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). The value  $X_j(t)$  represents for a subject  $j$  the number of occurrence of an event during the time interval  $[0, t]$  (e.g. the claim number in insurance mathematics). The randomness of the intensity  $C_j$  takes into account the heterogeneity among subjects which is more realistic. For instance, the distribution of  $C_j$  may be a mixture of distributions. The mixed Poisson process belongs to the more general class of mixed-effects models where parameters are assumed to be unobserved random variables.

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  $(X_j(T), j = 1, \dots, n)$  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:

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

which can be estimated by:

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

---

<sup>1</sup> MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité.

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.

Our approach is a penalized projection method (see Massart (1997)) which provides a concrete adaptive estimator of  $f$ . 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  $c \rightarrow e^{-c}c^\ell/l!$ . This leads us to assume:

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

Then choosing an orthonormal basis  $(\varphi_k)$  of this space, (1) can be written as:

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

where  $\theta_k, \Omega_k^{(\ell)}$  are respectively the  $k$ -th component of  $f$  and  $e^{-c}c^\ell/l!$  on the basis. The problem is to choose a basis such that the mapping  $(\theta_k, k \geq 0) \rightarrow (\alpha_\ell(f), \ell \geq 0)$  can be simply and explicitly inverted. Then, by plugging the estimators  $\hat{\alpha}_\ell$  in the inverse mapping, we get estimators of the coefficients  $\theta_k$  and deduce estimators of  $f$ . A relevant choice of the basis  $(\varphi_k)$  is thus a key tool. In what follows, we consider the basis

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

where  $(L_k(t))$  are the Laguerre polynomials. This choice is justified by the fact that  $\Omega_k^{(\ell)} = 0$  for all  $k > \ell$  and the matrix  $\Omega_\ell = (\Omega_j^{(i)})_{0 \leq i, j \leq \ell}$  is lower triangular and explicitly invertible (Propositions 2.1 and 2.2). This yields that the following linear mapping on  $\mathbb{R}^{\ell+1}$  is explicit for all  $\ell$ :

$$\vec{\alpha}_\ell = (\alpha_k(f), k = 0, \dots, \ell)' \rightarrow \vec{\theta}_\ell = (\theta_k, k = 0, \dots, \ell)' = \Omega_\ell^{-1} \vec{\alpha}_\ell$$

with a crucial 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  $\vec{\theta}_{\ell-1}$ .

Consequently, we define a collection of estimators of  $f$  by  $\hat{f}_\ell = \sum_{k=0}^{\ell} \hat{\theta}_k \varphi_k$  and study their  $\mathbb{L}^2$ -risk (Proposition 2.3). By introducing appropriate regularity subspaces of  $\mathbb{L}^2((0, +\infty))$ , the regularity of  $f$  is measured in terms of the convergence of series associated with the coefficients  $(\theta_k)$ . The justification is detailed in the Section 8. Using these regularity spaces, we discuss the possible rates of convergence of the  $\mathbb{L}^2$ -risk of  $\hat{f}_\ell$ . Smooth functions  $f$  yield rates of order  $O((\log n)^{-s})$ . "Super-smooth" functions yield rates of order  $O(n^{-b})$ ,  $b < 1$ . Afterwards, we propose a data-driven choice  $\hat{\ell}$  of the dimension  $\ell$  and study the  $\mathbb{L}^2$ -risk of the resulting adaptive estimator (Proposition 2.1).

In Section 3, we interpret the results of Section 2 in the case where the observation is  $(N_j(C_j T), j = 1, \dots, n)$ . 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 4 is devoted to the estimation of  $f$  for large  $T$ . Our methods rely on the property that for each  $j$ ,  $\hat{C}_{j,T} = N_j(C_j T)/T$  is a consistent estimator of the random variable  $C_j$  as  $T$  tends to infinity. Then, we use the i.i.d. sample  $(\hat{C}_{j,T})_{1 \leq j \leq n}$ , to build estimators of  $f$ . First we propose a kernel estimator with an adaptive bandwidth selection (Proposition 4.1, Theorem 4.1). Then we study projection estimators on the Laguerre basis (3) with an adaptive choice of the space dimension (Proposition 4.2, Theorem 4.2).

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

2. ESTIMATION FOR  $T = 1$ .

2.1. **Preliminaries.** The Laguerre polynomials given by

$$(4) \quad L_k(t) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{t^j}{j!}, \quad k \geq 0$$

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 Kroenecker 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)

$$(5) \quad f = \sum_{k \geq 0} \theta_k \varphi_k, \quad \text{where} \quad \theta_k = \int_0^{+\infty} f(c) \varphi_k(c) dc.$$

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

$$(6) \quad \frac{1}{\ell!} c^\ell e^{-c} = \sum_{k \geq 0} \Omega_k^{(\ell)} \varphi_k(c), \quad \text{where} \quad \Omega_k^{(\ell)} = \frac{1}{\ell!} \int_0^{+\infty} c^\ell \sqrt{2} L_k(2c) e^{-2c} dc.$$

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 8 for more details). Thus,

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

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

**Proposition 2.1.** *The coefficients  $\Omega_k^{(\ell)}$  defined by (6) are equal to*

$$(7) \quad \Omega_k^{(\ell)} = \frac{(-1)^k}{\sqrt{2} 2^\ell} \binom{\ell}{k} \mathbf{1}_{(k \leq \ell)}.$$

Define the vectors

$$\vec{\theta}_\ell = (\theta_k, k = 0, \dots, \ell)' \quad \vec{\alpha}_\ell = (\alpha_k(f), k = 0, \dots, \ell)'$$

and the triangular matrix

$$(8) \quad \Omega_\ell = \begin{pmatrix} \Omega_0^{(0)} & 0 & \dots & 0 \\ \Omega_0^{(1)} & \Omega_1^{(1)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \Omega_0^{(\ell)} & \Omega_1^{(\ell)} & \dots & \Omega_\ell^{(\ell)} \end{pmatrix},$$

where the diagonal terms are  $\Omega_i^{(i)} = (-1)^i / (\sqrt{2} 2^i)$ . The matrix  $\Omega_\ell$  is therefore invertible and its inverse is explicitly computed in the following proposition.

**Proposition 2.2.** *The following equality holds:*

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

Therefore  $\vec{\theta}_\ell = \Omega_\ell^{-1} \vec{\alpha}_\ell$ .

Note that since both  $\Omega_\ell$  and  $\Omega_\ell^{-1}$  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  $\vec{\theta}_{\ell-1}$ .

## 2.2. Estimation of the density of the random effect.

2.2.1. *Projection estimators.* We consider the empirical estimators (2) of  $\alpha_k := \alpha_k(f)$  and set

$$(9) \quad \vec{\tilde{\alpha}}_\ell = {}^t(\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_\ell)$$

The vector  $\vec{\theta}_\ell = (\theta_k, k = 0, \dots, \ell)'$  of components of  $f$  is estimated by  $\vec{\hat{\theta}}_\ell = \Omega_\ell^{-1} \vec{\tilde{\alpha}}_\ell$ .

By the triangular form of  $\Omega_\ell$ ,  $\vec{\tilde{\alpha}}_\ell$  and  $\vec{\hat{\theta}}_\ell$  have their first  $\ell - 1$  coordinates equal to those of  $\vec{\tilde{\alpha}}_{\ell-1}$  and  $\vec{\hat{\theta}}_{\ell-1}$ . Denote by  $f_\ell$  the orthogonal projection of  $f$  on

$$S_\ell = \text{span}(\varphi_0, \varphi_1, \dots, \varphi_\ell),$$

namely  $f_\ell = \sum_{k=0}^{\ell} \theta_k \varphi_k$ . We estimate  $f$  by  $\hat{f}_\ell = \sum_{k=0}^{\ell} \hat{\theta}_k \varphi_k$ . We denote by  $\|\cdot\|$  the  $\mathbb{L}^2$ -norm of  $\mathbb{L}^2((0, +\infty))$  and by  $|\cdot|$  the usual Euclidean norm in  $\mathbb{R}^{\ell+1}$ . The following risk decomposition holds.

**Proposition 2.3.** *The estimator of  $f$  defined by*

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

where  $\vec{\tilde{\alpha}}_\ell$  is defined by (2) and (9) and  $\Omega_\ell$  by (7)-(8), satisfies

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

Let us discuss the possibles rates implied by Proposition 2.3. Consider the following regularity function spaces, for  $s, u, v \geq 0$ ,

$$(11) \quad W^{s,u,v}(\mathbb{R}^+, K) = \{h : \mathbb{R}^+ \rightarrow \mathbb{R}, h \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s \exp(uk^v) \theta_k^2(h) \leq K < +\infty\}$$

where  $\theta_k(h) = \int_0^{+\infty} h(u) \varphi_k(u) du$ . For  $u = 0$ , we simply note  $W^s(\mathbb{R}^+, K)$ . Regularity spaces linked with Laguerre bases are discussed in Section 8. In particular, for  $s$  integer, if  $h : (0, +\infty) \rightarrow \mathbb{R}$  belongs to  $L^2((0, +\infty))$ ,

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

is equivalent to the property that  $f$  admits derivatives up to order  $s - 1$ , with  $f^{(s-1)}$  absolutely continuous and for  $m = 0, \dots, s - 1$ , the functions  $\delta^m \circ \dots \circ \delta^1 \circ \delta^0 f$  belong to  $L^2((0, +\infty))$  where  $\delta^\alpha f = \sqrt{x} f' + \frac{1}{2}(\sqrt{x} - \frac{\alpha}{\sqrt{x}})f$ . Thus,  $W^s(\mathbb{R}^+, K)$  corresponds to a space of functions with regularity  $s$ . For any  $h \in W^s(\mathbb{R}^+, K)$ , we have  $\|h - h_\ell\|^2 = \sum_{k=\ell+1}^{\infty} \theta_k^2(h) \leq K/\ell^s$  where  $h_\ell$  is

the orthogonal projection of  $h$  on  $S_\ell$ .

Consequently, for  $f \in W^s(\mathbb{R}^+, K)$ , 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  $\ell$ . Thus, we can not make the classical bias variance compromise and must choose  $\ell$  so that the variance term does not explode. The bias becomes the dominating term. For instance, we can choose  $\ell$  such that  $2^{4\ell}/n = 1/\sqrt{n}$ , which yields  $\ell = \log(n)/(8 \log(2))$  and a rate of order  $O([\log(n)]^{-s})$ .

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

The logarithmic rate can be improved for  $f \in W^{s,u,v}(\mathbb{R}^+, K)$  with  $u > 0$ . For instance, consider for  $f$  an exponential density  $\mathcal{E}(\theta)$ . In this case,

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

As a consequence

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

Note that the bias is null for  $\theta = 1$ , since then,  $f = \varphi_0$ . 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 rates depend on  $\theta$  and can be  $O(n^{-b})$  for any  $b < 1$ . For instance if  $\theta = 5/3$  the rate is  $O(n^{-1/2})$ , and it tends to  $O(n^{-1})$  (the parametric rate) when  $\theta$  tends to 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  $\ell \geq p-1$ ,

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

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

2.2.2. *Model selection.* Now we have to define an automatic selection rule of the adequate dimension  $\ell$ . We make the selection among the following set, which guaranties that the variance remains bounded. Let

$$\mathcal{M}_n = \left\{ \ell \in \{0, 1, \dots, L_n\}, L_n = \left\lceil \frac{1 \log(n)}{4 \log(2)} \right\rceil + 1 \right\}$$

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

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

We can prove the following result

**Theorem 2.1.** *Consider the estimator  $\hat{f}_{\hat{\ell}}$  defined by (10) and (13). Then there exists  $\kappa > 0$  such that*

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

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  $\ell$ . This implies a possible negligible loss in the rate of the adaptive estimator w.r.t. the expected optimal rate.

### 3. ESTIMATION FOR GENERAL FIXED $T$

Let us now assume that the observation is  $(N_j(C_j T), j = 1, \dots, n)$ . The previous method applies directly to estimate the density  $f_T$  de  $C_j T$  i.e.  $f_T(t) = (1/T)f(t/T)$ . We can deduce the results for  $f(c) = T f_T(Tc)$ .

Developping  $f_T$  on the basis  $(\varphi_k, k \geq 0)$  yields coefficients  $\theta_k(f_T), k \geq 0$ . Now, developping  $f$  on the basis  $(\varphi_k^{(T)} := \sqrt{T}\varphi_k(T\cdot), k \geq 0)$  yields coefficients

$$\theta_k^{(T)}(f) = \sqrt{T}\theta_k(f_T).$$

Let us denote by

$$f_\ell^{(T)} = \sum_{k=0}^{\ell} \theta_k^{(T)}(f) \varphi_k^{(T)}$$

the orthogonal projection of  $f$  on the space  $S_\ell^{(T)}$  spanned by  $(\varphi_k^{(T)}, k \leq \ell)$ . The following relation holds for all  $\ell \geq 0$ :

$$f_\ell^{(T)}(c) = T f_{T,\ell}(cT)$$

where  $f_{T,\ell}$  denotes as above the orthogonal projection of  $f_T$  on the space  $S_\ell$  spanned by  $(\varphi_k, k \leq \ell)$ . This implies:

$$\|f - f_\ell^{(T)}\|^2 = T \|f_T - f_{T,\ell}\|^2$$

The coefficients

$$\alpha_k(f_T) = \alpha_k^{(T)}(f) = \int_0^{+\infty} f(c) e^{-cT} \frac{(cT)^k}{k!} dc = \mathbb{P}(N_j(C_j T) = k)$$

are estimated by

$$(14) \quad \hat{\alpha}_k = \frac{1}{n} \sum_{j=1}^n 1_{(N_j(C_j T)=k)}, \quad k \geq 0.$$

To estimate  $f(c) = Tf_T(Tc)$ , we set for all  $\ell$ ,

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

where  $\hat{f}_{T,\ell}$  is the estimator built for  $f_T$  using  $(N_j(C_jT), j = 1, \dots, n)$ . Therefore,

$$\|\hat{f}_\ell^{(T)} - f\|^2 = T\|\hat{f}_{T,\ell} - f\|^2.$$

Consequently, applying Proposition 2.3 for  $f_T$  yields the following result.

**Proposition 3.1.** *The estimator  $\hat{f}_\ell^{(T)}$  of  $f$  satisfies*

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

The construction of the estimator  $\hat{f}_\ell^{(T)}$  uses the basis  $(\varphi_k^{(T)})$ . The variance term in the  $\mathbb{L}^2$ -risk is multiplied by a factor  $T$ . This explains the deterioration of the method when  $T$  increases. With  $\hat{\ell}$  defined in (13), we set  $\hat{f}_{\hat{\ell}}^{(T)}(c) := T\hat{f}_{T,\hat{\ell}}(Tc)$ . The following result holds for the adaptive estimator.

**Theorem 3.1.** *There exists  $\kappa > 0$  such that*

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

#### 4. ESTIMATION FOR LARGE $T$

Let us set

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

Conditionally to  $C_j = c$ , we know that  $\hat{C}_{j,T}$  converges almost surely to  $c$  as  $T$  tends to infinity. Consequently,  $\hat{C}_{j,T}$  converges almost surely to  $C_j$ . We now use the i.i.d. sample  $(\hat{C}_{j,T})_{1 \leq j \leq n}$  to build estimators of  $f$ .

**4.1. Kernel estimator and adaptive bandwidth.** Consider  $K : \mathbb{R} \rightarrow \mathbb{R}$  a  $C^2$  function, integrable, with  $\int_{\mathbb{R}} K(u)du = 1$  and

$$(15) \quad \int_{\mathbb{R}} K^2(u)du < +\infty, \quad \int_{\mathbb{R}} (K''(u))^2(u)du < +\infty.$$

Let  $K_h(\cdot) = (1/h)K(\cdot/h)$  and set for  $h > 0$ ,

$$(16) \quad \hat{f}_h(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - \hat{C}_{j,T}}{h}\right), \quad x \in \mathbb{R}.$$

Obviously  $f = f1_{f \geq 0}$ . To distinguish the  $\mathbb{L}^2$ -norm on  $\mathbb{R}$  of the  $\mathbb{L}^2$ -norm on  $\mathbb{R}^+$ , we set  $\|h\|_2^2 = \int_{\mathbb{R}} f^2(x)dx$ .

The following Proposition gives the risk of the kernel estimator (16) with fixed bandwidth.

**Proposition 4.1.** *Consider the estimator  $\hat{f}_h$  given by (16) under (H) and (15). Assume moreover that  $\mathbb{E}(C_1^2) < +\infty$ . Then*

$$(17) \quad \mathbb{E}(\|\hat{f}_h - f\|_2^2) \leq 2\|f - f_h\|_2^2 + \frac{\|K\|_2^2}{nh} + \frac{2\|K''\|_2^2}{3h^5T^2} \left( 3\mathbb{E}(C_1^2) + \frac{\mathbb{E}(C_1)}{T} \right)$$

where  $f_h(x) = f * K_h(x)$  and  $*$  denotes the convolution product.



Under weak regularity assumptions on  $f$ , the bias term  $\|f - f_h\|_2^2$  tends to zero when  $h$  tends to zero. On the other hand, the bandwidth  $h$  must be such that  $1/(nh) + 1/(T^2 h^5)$  tends to zero. Since  $n$  is the number of *i.i.d.* observations, we express the rate of the  $\mathbb{L}^2$ -risk as a function of  $n$  and let  $T$  and  $h$  be also expressed as functions of  $n$ .

Recall that a kernel of order  $\ell$  satisfies  $\int_{\mathbb{R}} x^k K(x) dx = 0$  for  $k = 1, \dots, \ell$ . For constants  $\beta > 0$  and  $\Sigma > 0$ , the Nikol'ski class  $\mathcal{N}(\beta, \Sigma)$  is defined by:

$$\mathcal{N}(\beta, \Sigma) = \left\{ f : \mathbb{R} \mapsto \mathbb{R} : \left[ \int \left( f^{(\ell)}(x+t) - f^{(\ell)}(x) \right)^2 dx \right]^{1/2} \leq \Sigma |t|^{\beta-\ell}, \quad \forall t \in \mathbb{R} \right\},$$

where  $\ell = \lfloor \beta \rfloor$ .

The following corollary holds:

**Corollary 4.1.** *Assume that  $f$  belongs to  $\mathcal{N}(\beta, \Sigma)$  and that the kernel  $K$  has order  $\ell = \lfloor \beta \rfloor$  with  $\int_{\mathbb{R}} |x|^\beta |K(x)| dx < +\infty$ . Under the condition*

$$(18) \quad h \propto n^{-1/(2\beta+1)} \quad \text{and} \quad T^2 \geq n^{(2\beta+5)/(2\beta+1)},$$

we have

$$(19) \quad \mathbb{E}(\|\hat{f}_h - f\|_2^2) \leq C n^{-2\beta/(2\beta+1)}.$$

Corollary 4.1 standardly follows from Proposition 4.1. Indeed, if  $f$  belongs to  $\mathcal{N}(\beta, \Sigma)$  and if the kernel  $K$  has order  $\ell = \lfloor \beta \rfloor$ , then  $\|f - f_h\|_2^2 \leq C^2 h^{2\beta}$  with  $C = \Sigma \int_{\mathbb{R}} |u|^\beta |K(u)| du / \ell!$  (see Tsybakov, 2009). For the first two terms in the r.h.s. of (17), the classical rate-optimal compromise imposes that  $h \propto n^{-1/(2\beta+1)}$  thus implying  $h^{2\beta} + 1/(nh) \propto n^{-2\beta/(2\beta+1)}$ . Fitting the last term with this rate requires that  $1/(T^2 h^5) \leq n^{-2\beta/(2\beta+1)}$ . If  $T^2 \geq n^{(2\beta+5)/(2\beta+1)}$ , then  $n^{5/(2\beta+1)}/T^2 \leq n^{-2\beta/(2\beta+1)}$ . This constraint holds for any nonnegative  $\beta$  if  $T \geq n^{5/2}$ .

The conditions in (18) yield the rate (19).

We now define a data driven selection of the bandwidth  $h$ . The method developed by Golden-shluger and Lepski (2011) using iterated kernel estimators is applied. Here we have moreover to deal with the substitution of the unobserved  $C_j$ 's by the  $\hat{C}_{j,T}$ 's. As usual, bandwidth selection has to reproduce the bias-variance compromise. This requires to estimate the right-hand side of (17) for all  $h$ . Let

$$s_2 = 3\mathbb{E}(C_1^2) + \frac{\mathbb{E}(C_1)}{T} \quad \text{and} \quad \hat{s}_2 = \frac{1}{n} \sum_{j=1}^n [3(\hat{C}_{j,T})^2 - 2\frac{\hat{C}_{j,T}}{T}].$$

As

$$3(\hat{C}_{j,T})^2 - 2\frac{\hat{C}_{j,T}}{T} = \hat{C}_{j,T} \frac{3N_j(C_j T) - 2}{T} \geq 0,$$

$\hat{s}_2 \geq 0$ . Elementary computations using conditioning on  $C_j$  show that  $\mathbb{E}(\hat{s}_2) = s_2$ . We denote by

$$(20) \quad V(h) = V_1(h) + V_2(h) \quad \text{where} \quad V_1(h) = \kappa_1 \|K\|_1^2 \frac{\|K\|_2^2}{nh}, \quad V_2(h) = \kappa_2 \|K\|_1^2 \frac{\|K''\|_2^2}{h^5 T^2} \hat{s}_2,$$

and  $\kappa_1$  and  $\kappa_2$  are two numerical constants. Note that  $\|K\|_1 \geq 1$ . The term  $V(h)$  is, up to multiplicative constants, an unbiased estimator of the variance term in the bound (17). To estimate the bias term  $\|f - f_h\|_2^2$ , we introduce the iterated kernel estimators  $\hat{f}_{h,h'}(x) = K_{h'} * \hat{f}_h(x)$ . The idea is to estimate the bias term by  $\sup_{h'} \|\hat{f}_{h,h'} - \hat{f}_{h'}\|_2^2$ , where  $h'$  is in an appropriate

finite set denoted by  $\mathcal{H}_{n,T}$ . However this introduces an additional variance term which must be subtracted. Hence, to estimate the bias term, we set

$$(21) \quad A(h) = \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\hat{f}_{h,h'} - \hat{f}_{h'}\|_2^2 - V(h') \right)_+.$$

The adaptive bandwidth is now defined by

$$(22) \quad \hat{h} = \arg \min_{h \in \mathcal{H}_{n,T}} (A(h) + V(h)).$$

We consider the adaptive estimator  $\hat{f}_{\hat{h}}$ , for which a risk bound is given in Theorem below.

**Theorem 4.1.** *Assume that  $\int |K(u)|^{4/3} du < +\infty$  and  $\mathbb{E}(C_j^8) < +\infty$ . Assume moreover that the bandwidth collection  $\mathcal{H}_{n,T}$  is such that:  $\text{card}(\mathcal{H}_{n,T}) \leq n$ ,  $\forall h \in \mathcal{H}_{n,T}, 1/(nh) \leq 1$  and  $1/(h^5 T^2) \leq 1$ , and for any  $c > 0$ , there exists a finite constant  $\Sigma(c)$  (independent of  $n$  and  $T$ ) such that*

$$(23) \quad \sum_{h \in \mathcal{H}_{n,T}} h^{-1/2} e^{-c/h} \leq \Sigma(c).$$

Recall that  $f_h = K_h * f$ . Then there exist constants  $\kappa_1, \kappa_2$  such that

$$(24) \quad \mathbb{E} \left( \|\hat{f}_{\hat{h}} - f\|_2^2 \right) \leq C \inf_{h \in \mathcal{H}_{n,T}} \left( \|f - f_h\|_2^2 + \frac{\|K\|_1^2 \|K\|_2^2}{nh} + \frac{\|K\|_1^2 \|K^n\|_2^2 s_2}{h^5 T^2} \right) + C' \frac{1}{n},$$

where  $C$  is a numerical constant and  $C'$  is a constant depending on  $\|f\|_2, \|K\|_{4/3}, \|K\|_2$ .

The estimator  $\hat{f}_{\hat{h}}$  is indeed adaptive: the infimum in the right-hand-side of Inequality (24) shows that the bias variance compromise is automatically realized.

For instance, if  $T \geq n^{5/2}$ , we can choose  $\mathcal{H}_{n,T} = \{1/k, k = 1, \dots, \sqrt{n}\}$  or  $\mathcal{H}_{n,T} = \{1/2^k, k = 1, \dots, \lfloor \log(n)/\log(2) \rfloor\}$ .

**4.2. Projection estimator and model selection.** Kernel estimators are classical, but have the drawback to provide estimators on  $\mathbb{R}$ , which may show a boundary effect at 0, as the function to estimate is null on  $(-\infty, 0]$ . This is why we study the same projection estimators as in Section 2, using the Laguerre basis. The difference lies in the estimation of the coefficients  $\theta_k$ .

Let us set:

$$(25) \quad \tilde{f}_\ell^{(T)} = \sum_{k=0}^{\ell} \tilde{\theta}_k \varphi_k, \quad \tilde{\theta}_k = \frac{1}{n} \sum_{j=1}^n \varphi_k(\hat{C}_{j,T}).$$

We obtain the following risk bound.

**Proposition 4.2.** *Recall that  $f_\ell$  is the orthogonal projection of  $f$  on  $S_\ell = \text{span}(\varphi_0, \dots, \varphi_\ell)$ . Then*

$$\mathbb{E}(\|\tilde{f}_\ell^{(T)} - f\|^2) \leq \|f - f_\ell\|^2 + 2 \frac{\ell + 1}{n} + \frac{8(\ell + 1)^5 s_2}{T^2},$$

where we recall that the norm  $\|\cdot\|$  is the  $\mathbb{L}^2$ -norm on  $[0, +\infty)$ .

Now, we define the penalization procedure by setting

$$\mathcal{M}_{n,T} = \left\{ \ell \in \{0, 1, \dots, M_n\}, \quad M_n \leq n, \quad M_n \leq T^{5/2} \right\}$$

and

$$(26) \quad \tilde{\ell} = \arg \min_{\ell \in \mathcal{M}_{n,T}} \left\{ -\|\tilde{f}_\ell^{(T)}\|^2 + \widehat{\text{pen}}(\ell) \right\}$$

with

$$\widetilde{\text{pen}}(\ell) = \tilde{\kappa}_1 \frac{(\ell + 1)}{n} + \tilde{\kappa}_2 \frac{(\ell + 1)^5}{T^2} \hat{s}_2.$$

**Theorem 4.2.** *Let  $\tilde{f}_{\hat{\ell}}^{(T)}$  the estimator defined by (25) and (26). Then there exist numerical constants  $\tilde{\kappa}_1, \tilde{\kappa}_2$  such that*

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

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

## 5. 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 subsection 4.2 for large  $T$ .

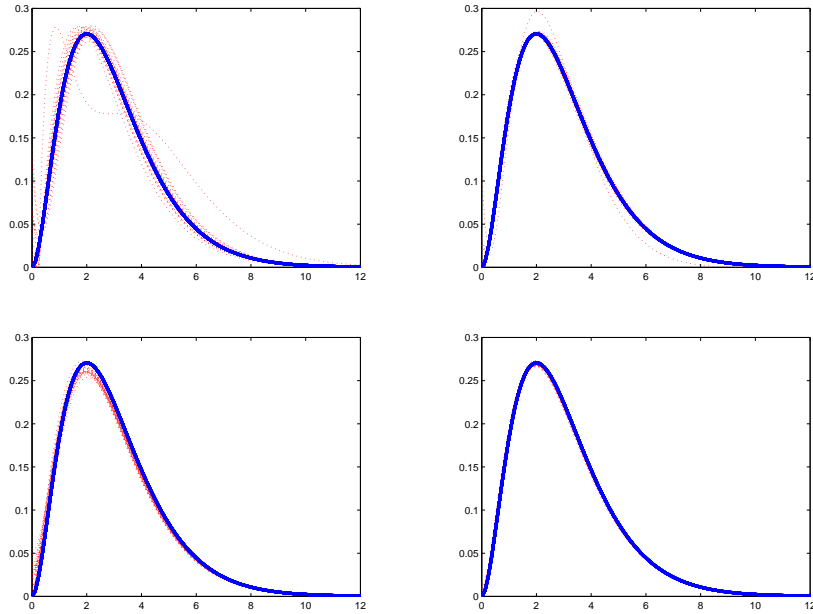


FIGURE 1. Estimation of the Gamma(3,1) density with method 1 (top left  $n = 10000$  and top right  $n = 100000$ , for  $T = 1$ ) and method 2 (bottom left,  $n = 1000, T = 10$  and bottom right  $n = 4000, T = 40$ ): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of the time  $\hat{\ell}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

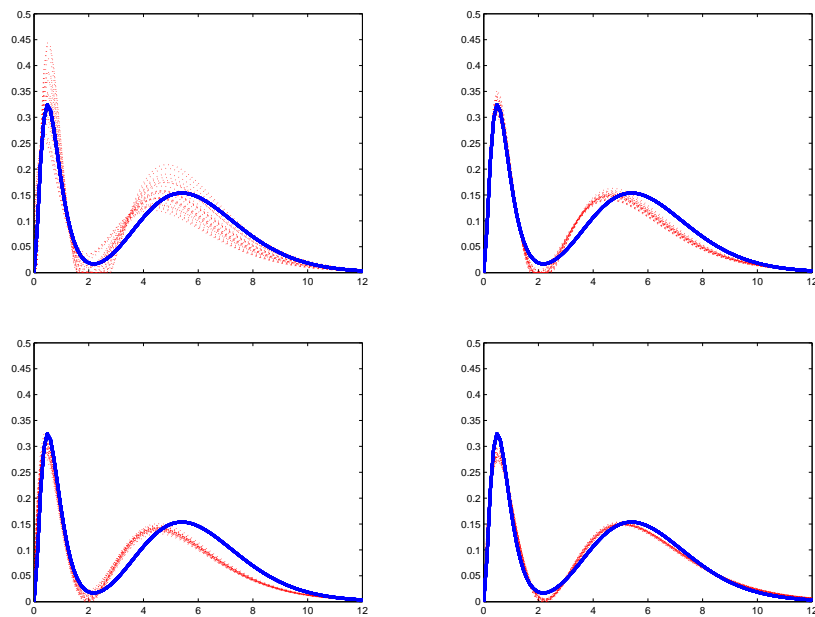


FIGURE 2. Estimation of the mixed Gamma density with method 1 (top left  $n = 10000$  and top right  $n = 100000$ , for  $T = 1$ ) and method 2 (bottom left,  $n = 1000, T = 10$  and bottom right  $n = 4000, T = 40$ ): true -thick (blue) line and 25 estimated (dashed (red) lines). The selected  $\ell$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

Note that, as  $\theta = 1$ , the density (1) has only three nonzero coefficients  $\theta_0, \theta_1, \theta_2$  in its exact development in the Laguerre basis. We thus expect the reconstruction to be perfect. On the contrary, for density (3), we know that the rate of the  $\mathbb{L}^2$  risk depends on the value of  $\theta$ , 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 illustrate the stability of the procedures.

- Comments on method 1. The method is easy to implement. We have selected the constant  $\kappa = 0.001$  in the penalty, after preliminary simulations. 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  $\ell$  in  $\{0, 1, \dots, 2\lfloor \log(n) \rfloor - 1\}$ . On the examples, the algorithm selects values of  $\hat{\ell}$  belonging to  $\{0, 1, \dots, 4\}$ .

- 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  $\ell$  in the selection algorithm; here again, the selected values  $\hat{\ell}$  are in  $\{0, 1, \dots, 4\}$ .

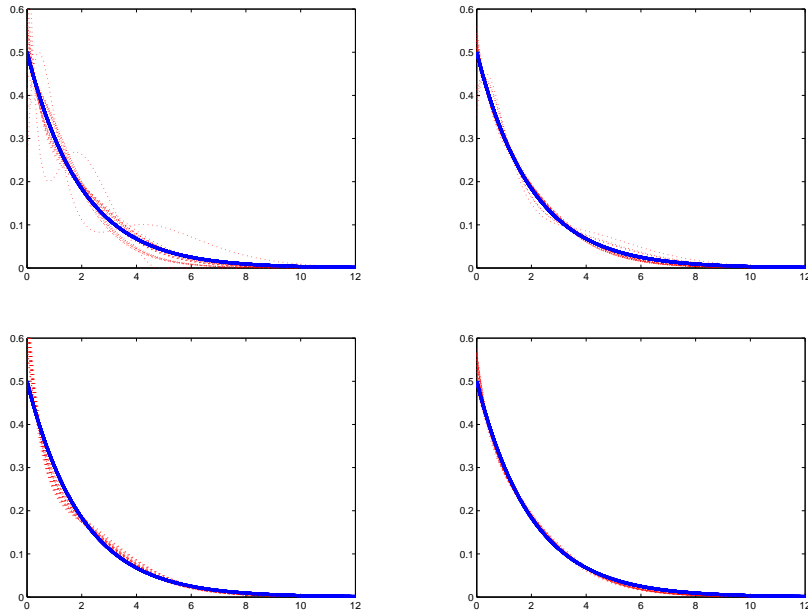


FIGURE 3. Estimation of the Exponential density with projection method 1 (top left  $n = 10000$  and top right  $n = 100000$ , for  $T = 1$ ) and method 2 (bottom left,  $n = 1000, T = 10$  and bottom right  $n = 4000, T = 40$ ): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of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

## 6. CONCLUDING REMARKS

In this paper, we study the nonparametric density estimation of a positive random variable  $C$  from the observation of  $(N_j(C_j T), j = 1, \dots, n)$ , 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  $\mathbb{L}^2(\mathbb{R}^+)$  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  $\mathbb{L}^2(\mathbb{R}^+)$ . 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 large  $T$ , estimators  $\hat{C}_{j,T}$  of the  $C_j$ 's are used to build adaptive kernel and adaptive projection estimators in the Laguerre basis. The interest of projection estimators on the Laguerre basis is that the unknown density is directly estimated on  $\mathbb{R}^+$ , contrary to standard kernel estimators. Moreover, the order of the bias term can be simply evaluated for functions in Sobolev Laguerre regularity spaces.

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.

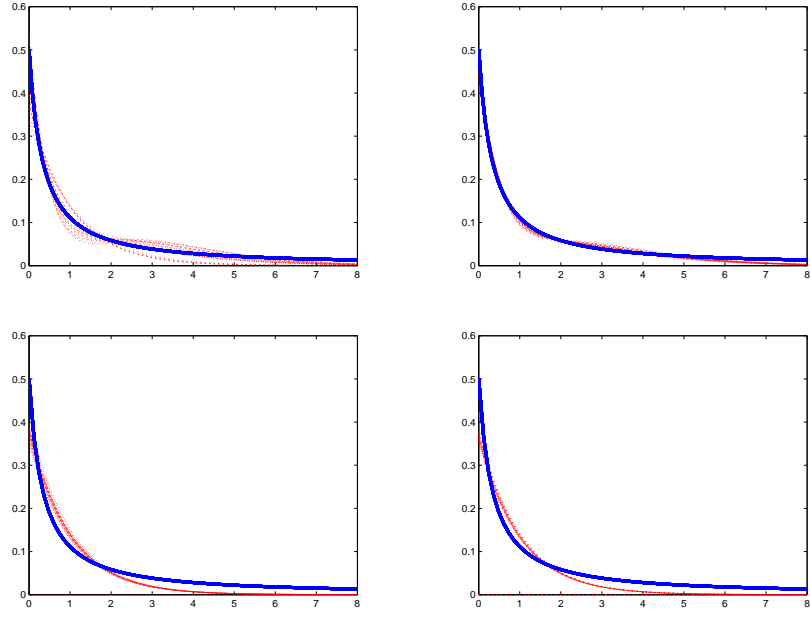


FIGURE 4. Estimation of the Pareto density with projection method 1 (top left  $n = 10000$  and top right  $n = 100000$ , for  $T = 1$ ) and method 2 (bottom left,  $n = 1000, T = 10$  and bottom right  $n = 4000, T = 40$ ): true -thick (blue) line and 25 estimated (dashed (red) lines). Most of the time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

- We may enrich the data by considering several observation times, see Fabio *et al.* (2012).
- We may extend the results to mixed compound Poisson processes by using the approach of Comte *et al.* (2013) based on observation of  $Y_j(t) = \sum_{i=1}^{N_j(C_j t)} \xi_i$  with  $\xi_i$  i.i.d. random variables independent of  $(N_j, C_j)_j$ .
- We may investigate the case of mixed Lévy processes.

## 7. PROOFS

7.1. **Proof of Proposition 2.1.** Using (4), 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,

$$(27) \quad \Omega_k^{(\ell)} = \frac{1}{\sqrt{2} 2^\ell} \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{(\ell+j)(\ell+j-1)\dots(\ell+1)}{j!}$$

where we know that  $\Omega_k^{(\ell)} = 0$  for  $k > \ell$ . Therefore  $\ell \rightarrow 2^\ell \Omega_k^{(\ell)}$  is a polynomial of degree  $k$  which is equal to 0 for  $\ell = 0, 1, \dots, k-1$ . Hence, we have:

$$2^\ell \Omega_k^{(\ell)} \propto \ell(\ell-1)(\ell-2)\dots(\ell-k+1).$$

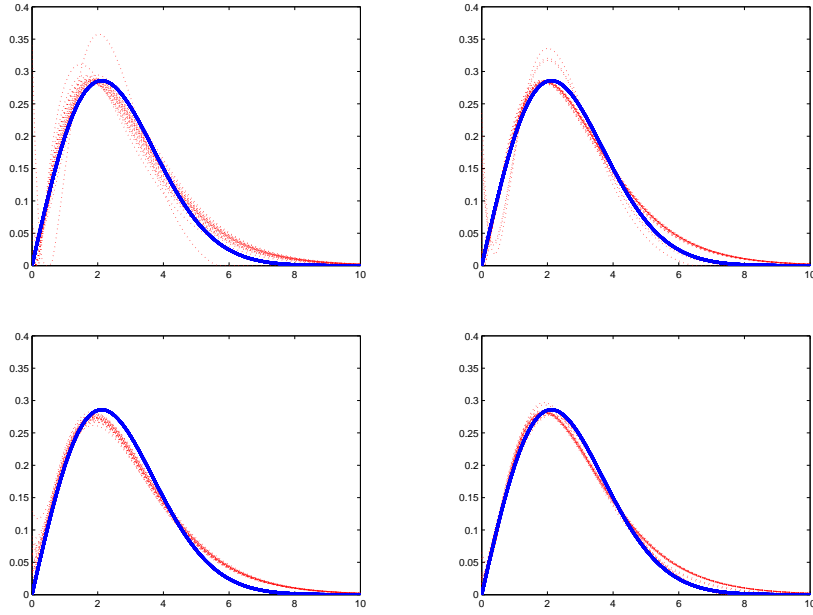


FIGURE 5. Estimation of the Weibull density with method 1 (top left  $n = 10000$  and top right  $n = 100000$ , for  $T = 1$ ) and method 2 (bottom left,  $n = 1000, T = 10$  and bottom right  $n = 4000, T = 40$ ): true -thick (blue) line and 25 estimated (dashed (red) lines). The selected  $\ell$ 's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

The proportionality coefficient is equal to the coefficient of  $\ell^k$  is  $(-1)^k / (\sqrt{2} k!)$ . Hence the result.  $\square$

**7.2. 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  $\ell$ . The transpose of the matrix  $\sqrt{2}\Omega_\ell$  represents the linear application

$$\begin{cases} \mathbb{R}_\ell[X] & \rightarrow \mathbb{R}_\ell[X] \\ P(X) & \mapsto P\left(\frac{1-X}{2}\right). \end{cases}$$

in the canonical basis  $(1, X, \dots, X^\ell)$ . The inverse linear mapping is

$$\begin{cases} \mathbb{R}_\ell[X] & \rightarrow \mathbb{R}_\ell[X] \\ Q(X) & \mapsto Q(1 - 2X). \end{cases}$$

hence the result.  $\square$

**7.3. Proof of Proposition 2.3.** 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:

$$(28) \quad \mathbb{E}(|\Omega_\ell^{-1}(\hat{\alpha}_\ell - \alpha_\ell)|^2) = \mathbb{E}({}^t(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell) {}^t\Omega_\ell^{-1}\Omega_\ell^{-1}(\vec{\hat{\alpha}}_\ell - \vec{\alpha}_\ell)).$$

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

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

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

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

where  $D_\alpha = \text{diag}(\alpha_0, \dots, \alpha_\ell)$ . Here, we get

$$(29) \quad \mathbb{E}(\|\hat{f}_\ell - f\|^2) \leq \|f - f_\ell\|^2 + \frac{1}{n}\text{Tr}({}^t\Omega_\ell^{-1}\Omega_\ell^{-1}D_\alpha).$$

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

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

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

$$(30) \quad \text{Tr}({}^t\Omega_\ell^{-1}\Omega_\ell^{-1}) = 2 \sum_{k=0}^{\ell} \sum_{j=0}^k \left( \binom{k}{j} \right)^2 2^{2j}$$

First write that

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

Now, notice that

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

We get

$$(31) \quad \text{Tr}({}^t\Omega_\ell^{-1}\Omega_\ell^{-1}) \leq \sum_{k=0}^{\ell} 2^{4k} = \frac{2^{4(\ell+1)} - 1}{2^4 - 1} \leq \frac{16}{15} 2^{4\ell}.$$

As a consequence, we obtain the risk decomposition announced in Proposition 2.3.  $\square$

**7.4. Proof of Proposition 2.1.** For simplicity, we set  $L_n = L$ . We define  $\mathbf{S}_\ell = \{t = {}^t(t_0, t_1, \dots, t_\ell, 0, \dots, 0) \in \mathbb{R}^{L+1}\}$ , which can also be associated with the function  $t = \sum_{k=0}^{\ell} t_k \varphi_k$  in  $S_\ell$  and  $|\mathbf{t}| = \|t\|$ .

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

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

For  $\mathbf{t} \in \mathbf{S}_\ell$ , note that  $\gamma_n(\mathbf{t}) = |t^{(\ell)}|^2 - 2\langle t^{(\ell)}, \Omega_\ell^{-1} \vec{\alpha}_\ell \rangle$ , 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

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



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

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

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

$$\hat{\ell} = \arg \min_{\ell \in \mathcal{M}_n} (\gamma_n(\hat{\mathbf{f}}_{\ell}) + \text{pen}(\ell)), \quad \text{where } \hat{\mathbf{f}}_{\ell} = {}^t(\hat{\theta}_0, \dots, \hat{\theta}_{\ell}, 0, \dots, 0) \in \mathbb{R}^{L+1}.$$

By definition of  $\hat{\ell}$ ,

$$\gamma_n(\hat{\mathbf{f}}_{\hat{\ell}}) + \text{pen}(\hat{\ell}) \leq \gamma_n(\mathbf{f}_{\ell}) + \text{pen}(\ell)$$

which implies

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

Now we have

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

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

$$\begin{aligned} \mathbb{E}(\|\hat{\mathbf{f}}_{\hat{\ell}} - f\|^2) &\leq 3\|\mathbf{f}_{\ell} - f\|^2 + 2\text{pen}(\ell) + 8\mathbb{E} \left( \sup_{\mathbf{t} \in \mathbf{S}_{\hat{\ell} \vee \ell}, |\mathbf{t}|=1} \langle \mathbf{t}, \Omega_L^{-1}(\vec{\alpha}_L - \vec{\alpha}_L) \rangle^2 - p(\ell, \hat{\ell}) \right) \\ (32) \quad &+ \mathbb{E}(8p(\ell, \hat{\ell}) - 2\text{pen}(\hat{\ell})). \end{aligned}$$

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

**Proposition 7.1.** *Let  $p(\ell, \ell') = 2\ell^* 2^{4\ell^*} / n$  with  $\ell^* = \ell \vee \ell'$ . Then, we have*

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

The result of Proposition 7.1 inserted in Inequality (32), shows that for  $\kappa \geq 8$ , we obtain  $4p(\ell, \hat{\ell}) \leq \text{pen}(\hat{\ell}) + \text{pen}(\ell)$  and

$$\mathbb{E}(\|\hat{\mathbf{f}}_{\hat{\ell}} - f\|^2) \leq 3\|\mathbf{f}_{\ell} - f\|^2 + 4\text{pen}(\ell) + \frac{8C'}{n}$$

which is the result of Proposition 2.1.  $\square$

**Proof of Proposition 7.1.** We apply the Talagrand Inequality recalled in Lemma 8.4 of Section 8.1. First note that

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

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

$$\nu_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \langle \mathbf{t}, \Omega_L^{-1}(\vec{\beta}_{i,L} - \vec{\alpha}_L) \rangle = \frac{1}{n} \sum_{i=1}^n (\psi_{\mathbf{t}}(\vec{\beta}_{i,L}) - \mathbb{E}\psi_{\mathbf{t}}(\vec{\beta}_{i,L}))$$

where  $\vec{\beta}_{i,L} = (1_{N_i(C_i\Delta)=0}, \dots, 1_{N_i(C_i\Delta)=L})$  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  $\mathbf{S}_\ell$ ,  $\nu_n(\mathbf{t}) = \langle \mathbf{t}, \Omega_\ell^{-1}(\vec{\alpha}_\ell - \bar{\alpha}_\ell) \rangle$ .

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 8.4, we specify  $\epsilon$ ,  $H^2$ ,  $M$  and  $v^2$ .

Clearly

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

This bound was obtained in the computation of (29) (see (28), (30), (31)).

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

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

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

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

$$\mathbb{E} \left( \sup_{\mathbf{t} \in \mathbf{S}_{\ell \vee \ell'}, |\mathbf{t}|=1} \langle \mathbf{t}, \Omega_L^{-1}(\vec{\alpha}_L - \bar{\alpha}_L) \rangle^2 - 2(1 + 2\delta\ell^*)H^2 \right)_+ \leq \frac{C_1}{n} \left( 2^{4\ell^*} e^{-C_2\delta\ell^*} + e^{-C_3\sqrt{\delta\ell^*n}} \right)$$

where we used that  $M^2/n \leq 2$  by definition of the set  $\mathcal{M}_n$  and where  $C_1, C_2, C_3$  are numerical constants. Then choosing  $\delta \geq \log(2)/C_2 + 1$  and  $\ell^* \geq 1$  gives the result.  $\square$

### 7.5. Proof of Proposition 4.1. Classically,

$$\begin{aligned} \mathbb{E}(\|f - \hat{f}_h\|_2^2) &= \|f - \mathbb{E}(\hat{f}_h)\|_2^2 + \mathbb{E}(\|\hat{f}_h - \mathbb{E}(\hat{f}_h)\|_2^2) \\ &\leq 2\|f - f_h\|_2^2 + 2\|f_h - \mathbb{E}(\hat{f}_h)\|_2^2 + \mathbb{E}(\|\hat{f}_h - \mathbb{E}(\hat{f}_h)\|_2^2). \end{aligned}$$

The last term is the usual variance term and is bounded by

$$(33) \quad \mathbb{E}(\|\hat{f}_h - \mathbb{E}(\hat{f}_h)\|_2^2) = \frac{1}{n} \int_{\mathbb{R}} \text{Var}(K_h(x - \hat{C}_{1,T})) dx \leq \frac{1}{n} \mathbb{E} \int_{\mathbb{R}} (K_h(x - \hat{C}_{1,T}))^2 dx \leq \frac{\int_{\mathbb{R}} K^2}{nh}.$$

The first term is a usual bias term. The specific term is the middle one. We note that  $f_h(x) = \mathbb{E}(K_h(x - C_1))$  and we apply the Taylor Formula with integral remainder:

$$\begin{aligned} K_h(x - \hat{C}_{1,T}) - K_h(x - C_1) &= \frac{(C_1 - \hat{C}_{1,T})}{h^2} K' \left( \frac{x - C_1}{h} \right) \\ &\quad + \frac{(C_1 - \hat{C}_{1,T})^2}{h^3} \int_0^1 (1-u) K'' \left( \frac{1}{h}(x - C_1 + u(C_1 - \hat{C}_{1,T})) \right) du. \end{aligned}$$

Noting that  $\mathbb{E}(\widehat{C}_{1,T}|C_1) = C_1$ , we get  $\mathbb{E}((C_1 - \widehat{C}_{1,T})K'(\frac{x-C_1}{h})) = 0$ . Using the Cauchy Schwarz inequality twice yields

$$\begin{aligned} \|f_h - \mathbb{E}(\widehat{f}_h)\|_2^2 &= \int \left( \mathbb{E}(K_h(x - \widehat{C}_{1,T}) - K_h(x - C_1)) \right)^2 dx \\ &= \int \left\{ \mathbb{E} \left[ \frac{(C_1 - \widehat{C}_{1,T})^2}{h^3} \int_0^1 (1-u)K'' \left( \frac{1}{h}(x - C_1 + u(C_1 - \widehat{C}_{1,T})) \right) du \right] \right\}^2 dx \\ &\leq \int \mathbb{E} \left[ \frac{(C_1 - \widehat{C}_{1,T})^4}{h^6} \int_0^1 (1-u)^2 (K'')^2 \left( \frac{1}{h}(x - C_1 + u(C_1 - \widehat{C}_{1,T})) \right) du \right] dx. \\ &\leq \frac{1}{3h^5} \mathbb{E} \left[ (C_1 - \widehat{C}_{1,T})^4 \right] \int (K'')^2(y) dy. \end{aligned}$$

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

$$(34) \quad \mathbb{E} \left[ (C_1 - \widehat{C}_{1,T})^4 \right] = \frac{1}{T^4} \mathbb{E}(C_1 T (1 + 3C_1 T)) = \frac{1}{T^2} \left( 3\mathbb{E}(C_1^2) + \frac{\mathbb{E}(C_1)}{T} \right).$$

Consequently,

$$(35) \quad \|f_h - \mathbb{E}(\widehat{f}_h)\|_2^2 \leq \frac{1}{3h^5 T^2} \left( 3\mathbb{E}(C_1^2) + \frac{\mathbb{E}(C_1)}{T} \right) \int (K'')^2(y) dy.$$

Then (33) and (35) imply the announced result.  $\square$

**7.6. Proof of Theorem 4.1.** For all  $h \in \mathcal{H}_{n,T}$ ,

$$\begin{aligned} &\|\widehat{f}_{\hat{h}} - f\|_2^2 \\ &\leq 3\|\widehat{f}_{\hat{h}} - \widehat{f}_{h,\hat{h}}\|_2^2 + 3\|\widehat{f}_{h,\hat{h}} - \widehat{f}_h\|_2^2 + 3\|\widehat{f}_h - f\|_2^2 \\ &\leq 3 \left( A(h) + V(\hat{h}) \right) + 3 \left( A(\hat{h}) + V(h) \right) + 3\|\widehat{f}_h - f\|_2^2 \\ &\leq 6A(h) + 6V(h) + 3\|\widehat{f}_h - f\|_2^2. \end{aligned}$$

Indeed, the second inequality holds by the definition of  $A(h)$ , i.e. for all  $h, h' \in \mathcal{H}_{n,T}$  we have  $A(h) + V(h') \geq \|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|_2^2$ . The last inequality holds by the definition of  $\hat{h}$ , that is  $A(\hat{h}) + V(\hat{h}) \leq A(h) + V(h)$  for all  $h \in \mathcal{H}_{n,T}$ . The bound for  $\mathbb{E} \left[ \|\widehat{f}_h - f\|_2^2 \right]$  is given by Proposition 4.1.

Hence, it is sufficient to study the term  $\mathbb{E}[A(h)]$ .

We state that

$$(36) \quad A(h) = \sup_{h' \in \mathcal{H}_{n,T}} \left[ \|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|_2^2 - V(h') \right]_+ \leq 5(D_1 + D_2 + D_3 + D_4 + D_5),$$

where  $D_1 = \sup_{h' \in \mathcal{H}_{n,T}} \|f_{h,h'} - f_{h'}\|_2^2$ ,

$$\begin{aligned} D_2 &= \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\widehat{f}_{h'} - \mathbb{E}\widehat{f}_{h'}\|_2^2 - \frac{V_1(h')}{10} \right)_+, \quad D_3 = \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\widehat{f}_{h,h'} - \mathbb{E}\widehat{f}_{h,h'}\|_2^2 - \frac{V_1(h')}{10} \right)_+, \\ D_4 &= \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\mathbb{E}(\widehat{f}_{h'}) - f_{h'}\|_2^2 - \frac{V_2(h')}{10} \right)_+, \quad D_5 = \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\mathbb{E}(\widehat{f}_{h,h'}) - f_{h,h'}\|_2^2 - \frac{V_2(h')}{10} \right)_+. \end{aligned}$$

Applying Young's Inequality  $\|f * g\|_p \leq \|f\|_1 \|g\|_p$  for  $p = 2$ , we get

$$\|f_{h,h'} - f_{h'}\|_2^2 = \|K_{h'} * (f_h - f)\|_2^2 \leq \|K_{h'}\|_1^2 \|f_h - f\|_2^2 = \|K\|_1^2 \|f_h - f\|_2^2,$$

and  $D_1 \leq \|K\|_1^2 \|f_h - f\|_2^2$ .

Next, to study  $D_2$ , we write

$$\|\hat{f}_{h'} - \mathbb{E}\hat{f}_{h'}\|_2^2 = \sup_{t \in \mathbb{L}^2, \|t\|_2=1} [\nu_{n,h'}(t)]^2,$$

where

$$\nu_{n,h}(t) = \langle t, \hat{f}_{h'} - \mathbb{E}\hat{f}_{h'} \rangle = \frac{1}{n} \sum_{j=1}^n \left[ t * K_h^-(\hat{C}_{j,T}) - \mathbb{E}(t * K_h^-(\hat{C}_{j,T})) \right]$$

and  $K_h^-(x) = K_h(-x)$ .

We have

$$\mathbb{E}(D_2) \leq \sum_{h' \in \mathcal{H}_{n,T}} \mathbb{E} \left( \sup_{t \in \mathcal{B}(1)} [\nu_{n,h'}(t)]^2 - \frac{V_1(h')}{10} \right)_+$$

where  $\mathcal{B}(1)$ , is the set of functions  $t$  of  $\mathbb{L}^2$  with  $\|t\|_2 = 1$ . We apply the Talagrand Inequality 8.4. To that aim, we need to evaluate the quantities  $M$ ,  $H$  and  $v^2$ . First,

$$\sup_{t \in \mathcal{B}(1)} \sup_x |\langle t, K_{h'}(\cdot - x) \rangle| \leq \sup_{t \in \mathcal{B}(1)} \|t\|_2 \|K_{h'}\|_2 = \|K\|_2 / \sqrt{h'} := M.$$

Next,

$$\mathbb{E} \left( \sup_{t \in \mathcal{B}(1)} [\nu_{n,h'}(t)]^2 \right) = \mathbb{E}(\|\hat{f}_{h'} - \mathbb{E}\hat{f}_{h'}\|_2^2) \leq \frac{\|K\|_2^2}{nh'} := H^2.$$

The most difficult is to obtain  $v^2$ . We split the term

$$\sup_{t \in \mathcal{B}(1)} \mathbb{E}(t * K_{h'}^-(\hat{C}_{1,T}))^2 \leq 2 \sup_{t \in \mathcal{B}(1)} \mathbb{E}(t * K_{h'}^-(C_1))^2 + 2 \sup_{t \in \mathcal{B}(1)} \mathbb{E}[t * (K_{h'}^-(\hat{C}_{1,T}) - K_{h'}^-(C_1))]^2 := T_1^2 + T_2^2.$$

First

$$\mathbb{E}(t * K_{h'}^-(C_1))^2 = \int [t * K_{h'}^-(c)]^2 f(c) dc \leq \|f\|_2 \|t * K_{h'}^-\|_4^2$$

By the Young Inequality (see Section 8.1) for  $p = 2$  and  $r = 4$  thus  $q = 4/3$ , we obtain  $\|t * K_{h'}^-\|_4^2 \leq \|t\|_2^2 \|K_{h'}^-\|_{4/3}^2$ . Consequently  $T_1^2 \leq \|f\|_2 \|K\|_{4/3}^2 / \sqrt{h'} := v_1^2$ . For the other term, we write with the mean value Theorem that

$$K_{h'}(x - \hat{C}_{1,T}) - K_{h'}(x - C_1) = (C_1 - \hat{C}_{1,T}) \int_0^1 (K_{h'})'(x - C_1 + u(C_1 - \hat{C}_{1,T})) du$$

so that

$$\begin{aligned} \left[ t * K_{h'}^-(\hat{C}_{1,T}) - t * K_{h'}^-(C_1) \right]^2 &= (C_1 - \hat{C}_{1,T})^2 \left( \int t(x) \int_0^1 (K_{h'})'(x - C_1 + u(C_1 - \hat{C}_{1,T})) du dx \right)^2 \\ &\leq (C_1 - \hat{C}_{1,T})^2 \|t\|_2^2 \int \int_0^1 [(K_{h'})'(x - C_1 + u(C_1 - \hat{C}_{1,T}))]^2 du dx \\ &\leq (C_1 - \hat{C}_{1,T})^2 \int (K_{h'}'(y))^2 dy. \end{aligned}$$

Thus, using  $\mathbb{E}[(\hat{C}_{1,T} - C_1)^2] = \mathbb{E}(C_1)/T$  and  $T \leq (h')^{5/2}$  by definition of  $\mathcal{H}_{n,T}$ , we obtain

$$\mathbb{E} \left( \left[ t * K_{h'}^-(\hat{C}_{1,T}) - t * K_{h'}^-(C_1) \right]^2 \right) \leq \frac{\|K'\|_2^2 \mathbb{E}(C_1)}{T(h')^3} \leq \frac{\|K'\|_2^2 \mathbb{E}(C_1)}{\sqrt{h'}}.$$

We obtain  $T_2^2 \leq \|K'\|_2^2 \mathbb{E}(C_1)/\sqrt{h'} := v_2^2$ . Thus  $v^2 = v_1^2 + v_2^2$ . The Talagrand Inequality yields, with  $v^2 = A_0/\sqrt{h'}$  and  $A_0 = \|f\|_2 \|K\|_{4/3}^2 + \|K'\|_2^2 \mathbb{E}(C_1)$ ,

$$\mathbb{E} \left( \sup_{t \in \mathcal{B}(1)} [\nu_{n,h'}(t)]^2 - \frac{V_1(h')}{10} \right)_+ \leq \frac{A_1}{n} \left( \frac{1}{\sqrt{h'}} e^{-A_2/\sqrt{h'}} + \frac{1}{nh'} e^{-A_3\sqrt{n}} \right),$$

for constants  $A_i$ ,  $i = 1, 2, 3$  depending on  $K$  and  $f$ . Using that  $1/(nh) \leq 1$  for  $h \in \mathcal{H}_{n,T}$ , we get, for  $\kappa_1/10 \geq 4$  (we take  $\epsilon^2 = 1/2$ ),

$$\mathbb{E}(D_2) \leq \frac{A_4}{n} \sum_{h \in \mathcal{H}_{n,T}} \frac{1}{\sqrt{h}} e^{-C'/\sqrt{h}} \leq \frac{C}{n}$$

under  $\sum_{h \in \mathcal{H}_{n,T}} h^{-1/2} e^{-C'/\sqrt{h}} \leq \Sigma(C) < +\infty$ ,  $\forall C > 0$ .

Next,  $D_3$  is analogous to  $D_2$  except that  $K_{h'}$  is replaced by  $K_h * K_{h'}$ , so that, by using Young's Inequality, all the bounds are simply multiplied by  $\|K\|_1$  for  $M$ ,  $\|K\|_1^2$  for  $H^2$  and  $v^2$ . For  $v^2$ , we can also remark that  $(K_h * K_{h'})' = K_h * (K_{h'})'$ .

Now, we turn to  $D_4$ . We have, by definition of  $\mathcal{H}_{n,T}$ ,

$$\begin{aligned} D_4 &= \sup_{h' \in \mathcal{H}_{n,T}} \left( \|\mathbb{E}(\hat{f}_{h'}) - f_{h'}\|_2^2 - \frac{V_2(h')}{10} \right)_+ \\ &\leq \sup_{h' \in \mathcal{H}} \left( \|\mathbb{E}(\hat{f}_{h'}) - f_{h'}\|_2^2 - \frac{\kappa_2}{10} \frac{1}{2} \|K\|_1^2 \frac{\|K''\|_2^2}{h^5 T^2} s_2 \right)_+ \\ &\quad + \frac{\kappa_2}{10} \|K\|_1^2 \|K''\|_2^2 \left( \frac{1}{2} s_2 - \hat{s}_2 \right)_+ \end{aligned}$$

The bound proved in (35) implies that, for  $\kappa_2/20 \geq 1/3$  i.e.  $\kappa_2 \geq 20/3$ , then

$$\left( \|\mathbb{E}(\hat{f}_{h'}) - f_{h'}\|_2^2 - \frac{\kappa_2}{10} \frac{1}{2} \|K\|_1^2 \frac{\|K''\|_2^2}{h^5 T^2} s_2 \right)_+ = 0.$$

Let

$$\Omega = \left\{ |\hat{s}_2 - s_2| \leq \frac{s_2}{2} \right\}.$$

We get

$$\mathbb{E}(D_4) \leq \frac{\kappa_2}{10} \|K\|_1^2 \|K''\|_2^2 \mathbb{E} \left( \left( \frac{1}{2} s_2 - \hat{s}_2 \right)_+ 1_{\Omega^c} \right)$$

since  $(\frac{1}{2} s_2 - \hat{s}_2)_+ 1_{\Omega} = 0$ . Therefore,

$$\mathbb{E}(D_4) \leq \frac{\kappa_2}{10} \|K\|_1^2 \|K''\|_2^2 \mathbb{E}^{1/2} \left[ \left( \frac{1}{2} s_2 - \hat{s}_2 \right)_+^2 \right] \mathbb{P}^{1/2}(\Omega^c).$$

By the Markov inequality, we have  $\mathbb{P}(\Omega^c) \leq \frac{2^4}{s_2^4} \mathbb{E}(|\hat{s}_2 - s_2|^4)$  and we use the Rosenthal Inequality to get

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

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

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

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  $m_2^2 \leq c_2$ .

Thus we get that  $\mathbb{E}(D_4) \leq C/n$ . Lastly,  $D_5$  is similar to  $D_4$  with additional factor  $\|K\|_1^2$  which explains the  $\|K\|_1$  added.  $\square$

### 7.7. Proof of Proposition 4.2.

**Lemma 7.1.**  $\forall x \geq 0$ ,  $|\varphi_k(x)| \leq \sqrt{2}$ ,  $|\varphi'_k(x)| \leq \sqrt{2}(2k+1) \leq 2\sqrt{2}(k+1)$  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 7.1. The proof uses the Laguerre polynomials  $L_k^\alpha$ , see the Appendix, Section 8, and relies on the relations  $[L_k^\alpha(x)]' = -L_{k-1}^{\alpha+1}(x)$  and the bound (47). Recall that  $\varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} = \sqrt{2}L_k^0(2x)e^{-x}$ . Bound (47) 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  $|L''_k(x)| = |- [L_{k-1}^1(x)]'| = |L_{k-2}^2(x)| \leq k(k-1)e^{x/2}/2$ .  $\square$

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,

$$(37) \quad \mathbb{E}(\|\tilde{f}_\ell^{(T)} - f\|^2) = \mathbb{E}(\|\tilde{f}_\ell^{(T)} - \mathbb{E}(\tilde{f}_\ell^{(T)})\|^2) + \|\mathbb{E}(\tilde{f}_\ell^{(T)}) - f_\ell\|^2 + \|f_\ell - f\|^2.$$

We have

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

which yields with Lemma 7.1,

$$(38) \quad \mathbb{E}(\|\tilde{f}_\ell^{(T)} - \mathbb{E}(\tilde{f}_\ell^{(T)})\|^2) \leq \frac{1}{n} \mathbb{E} \left( \sum_{j=0}^{\ell} \varphi_j^2(\hat{C}_{1,T}) \right) \leq \frac{2(\ell+1)}{n}.$$

On the other hand, for some  $\xi_T \in (C_1, \hat{C}_{1,T})$ ,

$$\begin{aligned} \|\mathbb{E}(\tilde{f}_\ell^{(T)}) - f_\ell\|^2 &= \sum_{j=0}^{\ell} \left[ \mathbb{E}(\varphi_j(\hat{C}_{1,T}) - \varphi_j(C_1)) \right]^2 = \sum_{j=0}^{\ell} [\mathbb{E}((\hat{C}_{1,T} - C_1)^2 \varphi_j''(\xi_T))]^2 \\ &\leq \mathbb{E} \left\{ (\hat{C}_{1,T} - C_1)^4 \sum_{j=0}^{\ell} [\varphi_j''(\xi_T)]^2 \right\} \\ (39) \quad &\leq 8\mathbb{E}[(\hat{C}_{1,T} - C_1)^4](\ell+1)^5 = 8\frac{(\ell+1)^5}{T^2}s_2 \text{ (see (34)).} \end{aligned}$$

Gathering (37), (38) and (39) yields the result.  $\square$

### 7.8. Proof of Theorem 4.2. Let

$$\begin{aligned} \tau_n(t) &= \frac{1}{n} \sum_{j=1}^n [t(\hat{C}_{j,T}) - \langle t, f \rangle] := \nu_n(t) + R(t), \\ \nu_n(t) &= \frac{1}{n} \sum_{j=1}^n [t(\hat{C}_{j,T}) - \mathbb{E}(t(\hat{C}_{j,T}))], \quad R(t) = \mathbb{E}[t(\hat{C}_{1,T})] - \langle t, f \rangle. \end{aligned}$$

Let  $\tilde{\gamma}_n(t) = \|t\|^2 - 2n^{-1} \sum_{j=1}^n t(\widehat{C}_{j,T})$ , and 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,  $\forall \ell \in \mathcal{M}_{n,T}$ ,

$$\tilde{\gamma}_n(\tilde{f}_\ell^{(T)}) + \widetilde{\text{pen}}(\tilde{\ell}) \leq \tilde{\gamma}_n(f_\ell) + \widetilde{\text{pen}}(\ell).$$

Therefore

$$(40) \quad \|\tilde{f}_\ell^{(T)} - f\|^2 \leq \|f_\ell - f\|^2 + \widetilde{\text{pen}}(\ell) + 2\tau_n(\tilde{f}_\ell^{(T)} - f_\ell) - \widetilde{\text{pen}}(\tilde{\ell}).$$

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

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

Plugging this in (40) and using that  $\|\tilde{f}_\ell^{(T)} - f_\ell\|^2 \leq 2\|\tilde{f}_\ell^{(T)} - f\|^2 + 2\|f - f_\ell\|^2$ , we get

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

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

**Proposition 7.2.** *Define  $p_1(\ell, \ell') = 4(\ell \vee \ell' + 1)/n$ , then*

$$(41) \quad \mathbb{E} \left( \sup_{t \in S_{\tilde{\ell} \vee \ell}} |\nu_n(t)|^2 - p_1(\ell, \tilde{\ell}) \right)_+ \leq \frac{c}{n},$$

where  $c$  is a positive constant.

Define  $p_2(\ell, \ell') = 8\hat{s}_2(\ell \vee \ell' + 1)^5/T^2$ , then

$$(42) \quad \mathbb{E} \left( \sup_{t \in S_{\tilde{\ell} \vee \ell}} |R(t)|^2 - p_2(\ell, \tilde{\ell}) \right)_+ \leq \frac{c'}{n}$$

where  $c'$  is a positive constant.

The proof of Proposition 7.2 follows the same line as the bounds for  $D_2$  and  $D_4$  in the proof of Theorem 4.1 and is therefore omitted. Now, the definitions of  $p_1$ ,  $p_2$  and  $\widetilde{\text{pen}}(\cdot)$  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

$$\|\tilde{f}_\ell^{(T)} - f\|^2 \leq 3\|f_\ell - f\|^2 + 4\widetilde{\text{pen}}(\ell) + \frac{c''}{n}$$

which ends the proof of Theorem 4.2.  $\square$

## 8. APPENDIX.

**Laguerre polynomials and associated regularity spaces.**

For  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  a Borel function, let

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

When  $\rho \equiv 1$ , 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  $(\phi_k^\rho)$  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

$$(43) \quad \rho(x) = x^\alpha e^{-x} = w_\alpha(x), \quad \alpha \geq 0$$

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

Consider the second order differential equation:

$$(44) \quad \sigma(x)y'' + \tau(x)y' + \lambda y = 0 \text{ with } \lambda = \lambda_k = -k\tau' - k\frac{k-1}{2}\sigma'',$$

and the weight function  $\rho$  given by  $[\sigma(x)\rho(x)]' = \tau(x)\rho(x)$ . For  $\sigma(x) = x, \tau(x) = \alpha + 1 - x, \rho(x) = w_\alpha(x)$ , the solution of (44) is  $y_k(x) = L_k^\alpha(x)$  the Laguerre polynomial with index  $\alpha$  and order  $k$ . The function  $L_k^\alpha$  is a polynomial of degree  $k$ , and the sequence  $(L_k^\alpha)$  is orthogonal with respect to the weight function  $w_\alpha$ . The orthogonality relations are equivalent to:

$$(45) \quad \int_0^{+\infty} x^\ell L_k^\alpha(x)w_\alpha(x)dx = 0 \quad \text{for } k > \ell.$$

We have

$$L_k^\alpha(x) = \frac{1}{k!} e^x x^{-\alpha} \frac{d^k}{dx^k} (x^{k+\alpha} e^{-x}).$$

The following holds:

$$(46) \quad (L_k^\alpha(x))' = -L_{k-1}^{\alpha+1}(x), \quad \int_0^{+\infty} (L_k^\alpha(x))^2 w_\alpha(x)dx = \frac{\Gamma(k+\alpha+1)}{k!}.$$

Moreover, we have the bound (see Abramowitz and Stegun 22.14.13)

$$(47) \quad \forall \alpha \geq 0, \forall k \geq 0, |L_k^\alpha(x)| \leq \frac{\Gamma(k+\alpha+1)}{k!\Gamma(\alpha+1)} e^{x/2}.$$

Setting  $\phi_k^\alpha(x) = L_k^\alpha(x) \left(\frac{k!}{\Gamma(k+\alpha+1)}\right)^{1/2}$ , 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  $L^2((0, +\infty), w)$ , with  $w = w_0$ . Noting that

$$(x^{\alpha+1} e^{-x})' = x^\alpha e^{-x}(\alpha + 1 - x),$$

we obtain, using (44) and (46),

$$(48) \quad \frac{d}{dx} (x^{\alpha+1} e^{-x} L_{k-1}^{\alpha+1}(x)) = x^\alpha e^{-x} k L_k^\alpha(x).$$

We can now prove the following result.

**Proposition 8.1.** *For  $s$  integer,  $w(x) = e^{-x}$  and  $g : (0, +\infty) \rightarrow \mathbb{R}$ , 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

$$(49) \quad \sum_{k \geq 0} k^s \tau_k^2(g) < +\infty,$$

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

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

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

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  $(\phi_{k'}^\alpha, k' \geq 0)$ , and for  $\alpha = 0$ ,  $\tau_k^0(h) = \tau_k(h)$ . The proof relies on the following Lemma:

**Lemma 8.1.** *Let  $\alpha \geq 0$ . If  $g : (0, +\infty) \rightarrow \mathbb{R}$  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 (48) yields

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

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  $[\dots]_0^{+\infty} = 0$ . Using that  $x^\alpha \leq x^{\alpha+1}$  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  $\lim_{x \rightarrow +\infty} g(x)x^{\alpha/2}e^{-x/2} = 0$ . Therefore,

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

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

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

□

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

$$(k(k-1)\dots(k-m+1))^{1/2} \tau_k(g) = (-1)^m \tau_{k-m}^m(g^{(m)}).$$

Therefore,

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

So we have (2). Moreover

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

Let us prove that (2)  $\Rightarrow$  (1). We have an analogous lemma.

**Lemma 8.2.** *Let  $\alpha \geq 0$ . Assume that  $g : (0, +\infty) \rightarrow \mathbb{R}$  belongs to  $L^2((0, +\infty), w_\alpha)$  and that  $\sum_{k \geq 0} k (\tau_k^\alpha(g))^2 < +\infty$ . Then,  $g$  is absolutely continuous,  $g'$  belongs to  $L^2((0, +\infty), w_{\alpha+1})$  and for all  $k \geq 1$ ,  $\tau_{k-1}^{\alpha+1}(g') = -\sqrt{k}\tau_k^\alpha(g)$ .*

*Proof.* We have  $g = \sum_{k \geq 0} \tau_k^\alpha(g) \phi_k^\alpha$  with  $\phi_0^\alpha$  a constant. Thus,

$$g(y) - g(x) = \sum_{k \geq 1} \tau_k^\alpha(g) \int_x^y (\phi_k^\alpha(t))' dt = - \sum_{k \geq 1} \sqrt{k} \tau_k^\alpha(g) \int_x^y \phi_{k-1}^{\alpha+1}(t) dt.$$

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

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

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

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

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

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

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

Let us turn now to bases of  $\mathbb{L}^2((0, +\infty))$  and to the definition of Laguerre-Sobolev spaces. For more details, we refer to in Bongioanni and Torrea (2009). The Laguerre functions are defined using the normalized Laguerre polynomials by:

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

The sequence  $(\mathcal{L}_k^\alpha, k \geq 0)$  constitute an orthonormal basis of  $\mathbb{L}^2((0, +\infty))$ . They are connected with the operator

$$\mathcal{L}_\alpha f = -x f'' - f' + \left(\frac{x}{4} + \frac{\alpha^2}{4x}\right) f$$

The following relations hold

$$\mathcal{L}_\alpha(\mathcal{L}_k^\alpha) = \left(k + \frac{\alpha + 1}{2}\right) \mathcal{L}_k^\alpha, \quad k \geq 0.$$

Now, let us introduce the operators

$$\delta^\alpha f = \sqrt{x} f' + \frac{1}{2} \left(\sqrt{x} - \frac{\alpha}{\sqrt{x}}\right) f, \quad (\delta^\alpha)^* f = -\sqrt{x} f' + \frac{1}{2} \left(\sqrt{x} - \frac{\alpha + 1}{\sqrt{x}}\right) f,$$

where  $(\delta^\alpha)^*$  denotes the (formal) adjoint of  $\delta^\alpha$  in  $\mathbb{L}^2((0, +\infty))$ . The following relations hold for  $f$  twice derivable :

$$\mathcal{L}_\alpha f = (\delta^\alpha)^* \delta^\alpha f + \frac{\alpha + 1}{2} f.$$

The operator  $\mathcal{L}_\alpha$  is linked with the operator  $L_\alpha$  given by  $L_\alpha g = x g'' + (\alpha + 1 - x) g'$  corresponding to the Laguerre polynomials by the relation:

$$x^{\alpha/2} e^{-x/2} L_\alpha(e^{x/2} x^{-\alpha/2} f) = -\mathcal{L}_\alpha f + \frac{\alpha + 1}{2} f.$$

We can now interpret the two Lemmas 8.1 and 8.2 in terms of Laguerre functions instead of Laguerre polynomials.

**Lemma 8.3.**  $\bullet$  *If  $f$  belongs to  $\mathbb{L}^2((0, +\infty))$ , is absolutely continuous and  $\delta^\alpha f$  belongs to  $\mathbb{L}^2((0, +\infty))$ , then,*

$$(50) \quad \sqrt{k} \theta_k^\alpha(f) = -\theta_{k-1}^{\alpha+1}(\delta^\alpha f)$$

where  $\theta_k^\alpha(f) = \int_0^{+\infty} f(x) \mathcal{L}_k^\alpha(x) dx$  is the  $k$ -th component of  $f$  on the basis  $(\mathcal{L}_k^\alpha)$  of  $\mathbb{L}^2((0, +\infty))$ .

- $\bullet$  *If  $f$  belongs to  $\mathbb{L}^2((0, +\infty))$  and  $\sum_{k \geq 0} k(\theta_k^\alpha(f))^2 < +\infty$ , then,  $f$  is absolutely continuous,  $\delta^\alpha f$  belongs to  $\mathbb{L}^2((0, +\infty))$  and the relation (50) holds.*

*Proof.* Note that  $g \in \mathbb{L}^2((0, +\infty), w_\alpha)$  is equivalent to  $f = g x^{\alpha/2} e^{-x/2} \in \mathbb{L}^2((0, +\infty))$  and  $f$  is absolutely continuous if and only if  $g$  is. A simple computation yields:

$$\delta^\alpha f = g' e^{-x/2} x^{(\alpha+1)/2}.$$

We can apply Lemma 8.1 to  $g$ . Observing that

$$\sqrt{k} \tau_k^\alpha(g) = \sqrt{k} \theta_k^\alpha(f), \quad \tau_{k-1}^{\alpha+1}(g') = \theta_{k-1}^{\alpha+1}(\delta^\alpha f)$$

yields (50).

The second item is obtained analogously using that  $g' \in \mathbb{L}^2((0, +\infty), w_{\alpha+1})$  is equivalent to  $\delta^\alpha f \in \mathbb{L}^2((0, +\infty))$ .  $\square$

Let us now introduce the Laguerre-Sobolev space  $W^s$  for  $s$  integer as the set of functions  $f \in \mathbb{L}^2((0, +\infty))$  admitting derivatives up to order  $s - 1$ , with  $f^{(s-1)}$  absolutely continuous and  $\delta^m \circ \dots \circ \delta^1 \circ \delta^0 f \in \mathbb{L}^2((0, +\infty))$ ,  $m = 0, 1, \dots, s - 1$ . Then,

$$f \in W^s \Leftrightarrow \sum_{k \geq 0} k^s (\theta_k^0(f))^2 < +\infty.$$

For all  $0 \leq m \leq s$  and  $k \geq m$ ,

$$(k(k-1) \dots (k-m+1))^{1/2} \theta_k^0(f) = (-1)^m \theta_{k-m}^m(\delta^{m-1} \circ \dots \circ \delta^0 f)$$

so that

$$\sum_{k \geq m} k(k-1) \dots (k-m+1) (\theta_k^0(f))^2 = \|\delta^{m-1} \circ \dots \circ \delta^0 f\|^2.$$

These properties can be deduced from Lemma 8.3 by iterating relation (50). The space  $W^s$  is a special case of the Laguerre-Sobolev spaces  $W_\alpha^{k,p}$  in Bongioanni and Torrea (2009) (see Definition 4 and Theorem 3, p.150).

### 8.1. Useful inequalities.

- First 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 8.4.** (*Talagrand Inequality*) *Let  $Y_1, \dots, Y_n$  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  $\epsilon^2 > 0$*

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

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

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

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.

- Next, we give the Young Inequality (see Hirsch and Lacombe (1999)). Let  $f$  be a function belonging to  $\mathbb{L}^p(\mathbb{R})$  and  $g$  belonging to  $\mathbb{L}^q(\mathbb{R})$ , let  $p, q, r$  be real numbers in  $[1, +\infty]$  and such that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1.$$

Then

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

where  $f * g$  is the convolution product and  $\|f\|_p^p = \int |f(x)|^p dx$ . In particular, for  $p = 1$ ,  $r = q = 2$ , we have  $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$ .

• The Rosenthal inequality is given in *e.g.* Hall and Heyde (1980, p.23). Let  $(X_i)_{1 \leq i \leq n}$  be  $n$  independent centered random variables, such that  $\mathbb{E}(|X_i|^p) < +\infty$  for an integer  $p \geq 1$ . Then there exists a constant  $C(p)$  such that

$$(51) \quad \mathbb{E} \left( \left| \sum_{i=1}^n X_i \right|^p \right) \leq C(p) \left( \sum_{i=1}^n \mathbb{E}(|X_i|^p) + \left( \sum_{i=1}^n \mathbb{E}(X_i^2) \right)^{p/2} \right).$$

## 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 For sale by the Superintendent of Documents, U.S. Government Printing Office, 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., Duval, C. and Genon-Catalot, V. (2013). Nonparametric density estimation in compound Poisson process using convolution power estimators. Preprint MAP5 2013-05.
- [5] Comte, F. and Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *J. Statist. Plann. Inference* **142**, 1698-1715.
- [6] 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.
- [7] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, no. 3, 1257-1272.
- [8] Goldenshluger, A., Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.
- [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] Hirsch, F. and Lacombe, G. (1999). *Elements of functional analysis*. Graduate Texts in Mathematics, 192. Springer-Verlag, New York.
- [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] 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
- [15] Mikosch, T. (2009). *Non-life insurance mathematics. An introduction with the Poisson process*. Second edition. Universitext. Springer-Verlag, Berlin.
- [16] Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4**, 1200-1209.
- [17] 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.