Intensity estimation of non-homogeneous Poisson processes from shifted trajectories

Jérémie Bigot, Sébastien Gadat, Thierry Klein and Clément Marteau

Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)
31062 Toulouse, Cedex 9, France
{Jeremie.Bigot, Sebastien.Gadat, Thierry.Klein, Clement.Marteau}@math.univ-toulouse.fr

June 14, 2012

Abstract

In this paper, we consider the problem of estimating nonparametrically a mean pattern intensity λ from the observation of n independent and non-homogeneous Poisson processes $N^1, \dots, N^n$ on the interval $[0,1]$. This problem arises when data (counts) are collected independently from n individuals according to similar Poisson processes. We show that estimating this intensity is a deconvolution problem for which the density of the random shifts plays the role of the convolution operator. In an asymptotic setting where the number n of observed trajectories tends to infinity, we derive upper and lower bounds for the minimax quadratic risk over Besov balls. Non-linear thresholding in a Meyer wavelet basis is used to derive an adaptive estimator of the intensity. The proposed estimator is shown to achieve a near-minimax rate of convergence. This rate depends both on the smoothness of the intensity function and the density of the random shifts, which makes a connection between the classical deconvolution problem in nonparametric statistics and the estimation of a mean intensity from the observations of independent Poisson processes.

Keywords: Poisson processes, Random shifts, Intensity estimation, Deconvolution, Meyer wavelets, Adaptive estimation, Besov space, Minimax rate.
AMS classifications: Primary 62G08; secondary 42C40

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS.

1 Introduction

Poisson processes became intensively studied in the statistical theory during the last decades. Such processes are well suited to model a large amount of phenomena. In particular, they are used in various applied fields including genomics, biology and imaging.

In the statistical literature, the estimation of the intensity of non-homogeneous Poisson process has recently attracted a lot of attention. In particular the problem of estimating a Poisson intensity from a single trajectory has been studied using model selection techniques [27] and non-linear wavelet thresholding [12], [21], [28], [33]. Poisson noise removal has also been considered by [13], [35] for image processing applications. Deriving optimal estimators of a Poisson intensity using a minimax point of view has been considered in [9], [27], [28] [33]. In all these papers, the intensity λ of the observed process is expressed as $\lambda(t) = \kappa\lambda_0(t)$ where the function to estimate is the scaled intensity λ_0 and κ is a positive real, representing an "observation time", that is let going to infinity to study asymptotic properties.

In this paper, we consider a slightly different framework. In many applications, data can be modeled as independent Poisson processes with *different* non-homogeneous intensities having nevertheless a similar shape. The simplest model which describes such situations is to assume that the intensities $\lambda_1, \ldots, \lambda_n$ of the Poisson processes $N^1, \ldots, N^n$ are randomly shifted versions $\lambda_i(\cdot) = \lambda(\cdot - \boldsymbol{\tau}_i)$ of an unknown intensity λ, where $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ are i.i.d. random variables. The intensity λ that we want to estimate is thus the same for all the observed processes up to random translations. Basically, such a model corresponds to the assumption that the recording of counts does not start at the same time (or location) from one individual to another. Such situation appears in biology, in particular when reading DNA sequences from different subjects in genomics [32].

Let us now describe more precisely our model. Let $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ be i.i.d. random variables with known density g with respect to the Lebesgue measure on $\mathbb{R}$. Let $\lambda : [0,1] \to \mathbb{R}_+$ a real-valued function. Throughout the paper, it is assumed that λ can be extended outside $[0,1]$ by 1-periodization. We suppose that, conditionally to $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$, the point processes $N^1, \ldots, N^n$ are independent Poisson processes on the measure space $([0,1], \mathcal{B}([0,1]), dt)$ with intensities $\lambda_i(t) = \lambda(t - \boldsymbol{\tau}_i)$ for $t \in [0,1]$, where dt is the Lebesgue measure. Hence, conditionally to $\boldsymbol{\tau}_i$, N^i is a random countable set of points in $[0,1]$, and we denote by $dN_t^i = dN^i(t)$ the discrete random measure $\sum_{T \in N^i} \delta_T(t)$ for $t \in [0,1]$, where δ_T is the Dirac measure at point T. For further details on non-homogeneous Poisson processes, we refer to [20]. The objective of this paper is to study the estimation of λ from a minimax point of view as the number n of observed Poisson processes tends to infinity. Since λ is 1-periodic, one may argue that the random shifts $\boldsymbol{\tau}_i$ are only defined modulo one, and therefore, without loss of generality, we also assume that g is restricted to have support in the interval $[0,1]$.

In this framework, our main result is that estimating λ corresponds to a deconvolution problem where the density g of the random shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ is a convolution operator that has to be inverted. Hence, estimating λ falls into the category of Poisson inverse problems. The presence of the random shifts significantly complicates the construction of upper and lower bounds for the minimax risk. In particular, to derive a lower bound, standard methods such as the Assouad's cube technique that is widely used for standard deconvolution problems in a white noise model (see e.g. [26] and references therein) have to be carefully adapted to take into account the effect of the random shifts. In this paper, our main tool is a likelihood ratio formula specific to Poisson processes (see Lemma 9.1 below) that yields to major differences in the proof with respect to the Gaussian case. In order to obtain an upper bound, we use Meyer wavelets which are well suited to deconvolution problems [19]. We construct a non-linear wavelet-based estimator with level-dependent and random thresholds that require the use of concentration inequalities for Poisson processes and an accurate estimation of the L^1-norm of the intensity λ. Note that estimating the intensity function of an indirectly observed non-homogeneous Poisson process from a single trajectory has been considered by [3], [9], [25], but adopting an inverse problem point of view to estimate a mean pattern intensity from the observation of n Poisson processes has not been proposed so far.

We point out that we assume throughout this paper that the density g of the random shifts is known. This assumption relies on an a priori knowledge of the random phenomenon generating the shifts. This hypothesis is realistic when dealing with Chip-Seq data for which the biologists are able to describe and to quantify the law of small random deformations leading to a shifted D.N.A. transcription. Note that a similar assumptions appears in [30] in the setting where the shifts τ_i are given, but when one only observes the sum $\sum_{i=1}^{n} N^i$ of n Poisson processes N^i with randomly shifted intensities.

The rest of the paper is organized as follows. In Section 2, we discuss some limitations of existing approaches (e.g. in genomics and bioinformatics) to estimate the mean pattern λ via an alignment step which consists in computing "estimators" $\hat{\tau}_1, \ldots, \hat{\tau}_n$ of the unobserved shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$. In Section 3, we describe the connection between estimating λ and standard approaches in statistical deconvolution problems. We also discuss the construction of a linear but nonadaptive estimator of the intensity λ. Section 4 is devoted to the computation of a lower bound for the minimax risk over Besov balls which is the main contribution of the paper. In Section 5, we construct an adaptive estimator using non-linear Meyer wavelet thresholding, that is used to obtain an upper bound of the minimax risk over Besov balls. Section 6 contains a conclusion and a discussion on some perspectives. The proofs of the main statements and of some technical lemmas are gathered in Section 7, Section 8, Section 9 and Section 10.

2 The standard approach to estimate a mean intensity via an alignment step

The motivation of our study comes from a practical problem encountered in DNA Chip-Seq data processing which can be described as follows. Chip-Sequencing is a fast biological analysis pipeline used to find and map genetic information along the genome. For any protein (transcription factor) which can read and interpret information in the genome, Chip-Seq provides a long sequence of tags (called reads) associated with specific genome locations where this transcription factor binds specific DNA sequences. Moreover, the Chip-Seq data provides a higher concentration of tags near transcription factor binding sites. We can number several goals for such analysis. Biologists are interested in the identification of true binding sites (where the rate is significantly high), as well as the estimation of the mean binding rate along the genome for such protein or the clustering of two populations of experiments which behaves very differently during such tag procedure.

From a statistical point of view, Chip-Seq data may be considered as repetitions of some Poisson counting processes (see [31]) which is not of homogeneous intensity as pointed in [11]. The unknown intensity of the underlying Poisson process quantifies the rate of expected reads for a specific choice of transcription factor. To obtain an estimator of this unknown intensity, a simple procedure is to average all the observed experiments. However, there is an additional difficulty in the analysis of such data which mainly relies on the acquisition method. The sequencing procedure puts some tags when reads occur along a very long DNA sequence. Then, it splits this counting process in a large number of sequences with smaller sizes which (roughly speaking) correspond to several chromosomes. But the demarcation of the beginning and ending locations of the chromosomes depends on some *a priori* knowledge which may be inaccurate. For each observed counting process, this generates some additional unknown random shifts of the underlying intensity.

In order to overcome this additional source of randomness, various strategies have been proposed to compute estimators $\hat{\tau}_1, \ldots, \hat{\tau}_n$ of the unobserved shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ (see e.g. [34, 1, 2]). An estimator of the intensity λ can then be computed by aligning and then averaging the observed processes. More precisely, if $\hat{\lambda}_i(\cdot)$ denotes an estimator of the shifted intensity $\lambda(\cdot - \boldsymbol{\tau}_i)$, obtained by some smoothing procedure applied to the process N^i, then an estimator of λ via an alignment step is defined by

$$\hat{\lambda}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda}_i(t + \hat{\tau}_i), \ t \in [0,1].$$

However, as pointed by Theorem 2.1 and Theorem 2.2 below, we show that an estimation of λ through an alignment step yields to non-consistent estimators. Indeed, a first result is that, under mild assumptions on the intensity λ and the density g of the random shifts, it is not possible to build consistent estimators of the random shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ in the sense that $\liminf_{n \to +\infty} \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\tau}_i - \boldsymbol{\tau}_i\right)^2\right) \neq 0$ for any estimators $(\hat{\tau}_1, \ldots, \hat{\tau}_n) \in [0,1]^n$.

Theorem 2.1 *Suppose that $\lambda \in L^2([0,1])$ is continuously differentiable and satisfies*

$$\lambda_0 := \inf_{t\in[0,1]} \{\lambda(t)\} > 0.$$

Assume that the density g of the random shifts has a compact support $[\tau_{\min}, \tau_{\max}] \subset [0,1]$ such that $\lim_{\tau\to\tau_{\min}} g(\tau) = \lim_{\tau\to\tau_{\max}} g(\tau) = 0$. Suppose that g is absolutely continuous and such that

$$\int_0^1 \left(\frac{\partial}{\partial \tau} \log g(\tau) \right)^2 g(\tau) d\tau < +\infty.$$

Let $(\hat{\tau}_1, \ldots, \hat{\tau}_n) \in [0,1]^n$ denote any estimators of the true random shifts $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ (i.e. a measurable mapping of the random processes N^i, $i = 1, \ldots, n$ taking its value in $[0,1]^n$). Then,

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} (\hat{\tau}_i - \boldsymbol{\tau}_i)^2 \right) \geq \frac{1}{\int_0^1 \left| \frac{\partial}{\partial t} \lambda(t) \right|^2 dt + \int_0^1 \left(\frac{\partial}{\partial \tau} \log g(\tau) \right)^2 g(\tau) d\tau} > 0. \tag{2.1}$$

Inequality (2.1) shows that building consistent estimators of the random shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ in the asymptotic setting $n \to +\infty$ is not feasible. This inconsistency result on the estimation of the shifts implies that a consistent estimation of λ via an alignment step is not possible. Indeed, consider the case of an ideal smoothing of the data with $\hat{\lambda}_i(t) = \lambda(t - \boldsymbol{\tau}_i)$, $t \in [0,1]$ which would lead to the ideal estimator

$$\bar{\lambda}_n(t) := \frac{1}{n} \sum_{i=1}^{n} \lambda(t - \boldsymbol{\tau}_i + \hat{\tau}_i), \; t \in [0,1],$$

where $(\tau_1, \ldots, \tau_n) \in [0,1]^n$ are estimators computed from the data $N^1, \ldots, N^n$. Then, the following theorem shows that $\bar{\lambda}_n$ is not a consistent estimator of λ as $n \to +\infty$.

Theorem 2.2 *Suppose that the assumptions of Theorem 2.1 still hold. Assume that $\lambda \in L^2([0,1])$ is such that*

$$\theta_1 := \int_0^1 \lambda(t) e^{-\mathbf{i}2\pi t} dt \neq 0.$$

Let $(\hat{\tau}_1, \ldots, \hat{\tau}_n) \in [0,1]^n$ denote any estimators of the true random shifts $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ satisfying the constraints $\sum_{i=1}^n \hat{\tau}_i = 0$ and $\tau_{\min} \leq \hat{\tau}_i \leq \tau_{\max}$ for all $i = 1, \ldots, n$. Suppose that the density g has zero expectation and finite variance i.e. $\int_0^1 \tau g(\tau) d\tau = 0$ and $\int_0^1 \tau^2 g(\tau) d\tau < +\infty$. Assume that $\tau_{\max} - \tau_{\min} := \frac{\delta}{4\pi}$ for some $0 < \delta < 3$, and consider the ideal estimator

$$\bar{\lambda}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \lambda(t - \boldsymbol{\tau}_i + \hat{\tau}_i), \; t \in [0,1].$$

Then,

$$\liminf_{n\to+\infty} \mathbb{E}\left(\int_0^1 \left| \bar{\lambda}_n(t) - \lambda(t) \right|^2 dt \right) \geq \left(\frac{\frac{2}{3}\pi^2 (3-\delta) |\theta_1|}{\int_0^1 \left| \frac{\partial}{\partial t} \lambda(t) \right|^2 dt + \int_0^1 \left(\frac{\partial}{\partial \tau} \log g(\tau) \right)^2 g(\tau) d\tau} \right)^2 > 0. \tag{2.2}$$

In Theorem (2.2), we have added the assumption that the estimators of the random shifts satisfy the constraint $\sum_{i=1}^n \hat{\tau}_i = 0$ and that the density g has zero expectation. Such assumptions are necessary when using an alignment procedure. Indeed, without such constraints, our model is not identifiable since for any $\tilde{\tau} \in \mathbb{R}$ one may replace the unknown intensity $\lambda(\cdot)$ by $\tilde{\lambda}(\cdot) = \lambda(\cdot - \tilde{\tau})$ and the random shifts by $\tilde{\boldsymbol{\tau}}_i = \boldsymbol{\tau}_i - \tilde{\tau}$ without changing the formulation of the problem. Under such assumptions, inequality (2.2) shows that

$$\liminf_{n\to+\infty} \mathbb{E}\left(\int_0^1 \left| \bar{\lambda}_n(t) - \lambda(t) \right|^2 dt \right) \neq 0,$$

and thus $\bar{\lambda}_n$ does not converge to λ as $n \to +\infty$ for the quadratic risk. Therefore, such a result illustrates the fact that standard procedures based on an alignment step do not yield consistent estimators of λ. In this paper, we therefore suggest an alternative method based on a deconvolution step that yields a consistent and adaptive estimator that converges with a (near-)optimal rate in the minimax sense.

3 A deconvolution problem formulation

3.1 A Fourier transformation of the data

For each observed counting process, the presence of a random shift complicates the estimation of the intensity λ. Indeed, for all $i \in \{1, \dots, n\}$ and any $f \in L^2([0,1])$ we have

$$\mathbb{E}\left[\int_0^1 f(t) dN_t^i | \boldsymbol{\tau}_i\right] = \int_0^1 f(t)\lambda(t - \boldsymbol{\tau}_i)dt, \tag{3.1}$$

where $\mathbb{E}[.|\boldsymbol{\tau}_i]$ denotes the conditional expectation with respect to the variable $\boldsymbol{\tau}_i$. Thus

$$\mathbb{E}\int_0^1 f(t)dN_t^i = \int_0^1 f(t)\int_{\mathbb{R}} \lambda(t-\tau)g(\tau)d\tau dt = \int_0^1 f(t)(\lambda \star g)(t)dt.$$

Hence, the mean intensity of each randomly shifted process is the convolution $\lambda \star g$ between λ and the density of the shifts g. This shows that a parallel can be made with the classical statistical deconvolution problem which is known to be an inverse problem. This parallel is highlighted by taking a Fourier expansion of the data. Let $(e_\ell)_{\ell\in\mathbb{Z}}$ the complex Fourier basis on $[0,1]$, *i.e.* $e_\ell(t) = e^{\mathbf{i}2\pi\ell t}$ for all $\ell \in \mathbb{Z}$ and $t \in [0,1]$. For $\ell \in \mathbb{Z}$, define

$$\theta_\ell = \int_0^1 \lambda(t)e_\ell(t)dt \text{ and } \gamma_\ell := \int_0^1 g(t)e_\ell(t)dt,$$

as the Fourier coefficients of the intensity λ and the density g of the shifts. Then, for $\ell \in \mathbb{Z}$, define y_ℓ as

$$y_\ell := \frac{1}{n}\sum_{i=1}^n \int_0^1 e_\ell(t)dN_t^i. \tag{3.2}$$

Using (3.1) with $f = e_\ell$, we obtain that

$$\mathbb{E}\left[y_\ell | \boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_n\right] = \frac{1}{n}\sum_{i=1}^n \int_0^1 e_\ell(t)\lambda(t-\boldsymbol{\tau}_i)dt = \frac{1}{n}\sum_{i=1}^n e^{-\mathbf{i}2\pi\ell\boldsymbol{\tau}_i}\theta_\ell = \tilde{\gamma}_\ell\theta_\ell,$$

where we have introduced the notation

$$\tilde{\gamma}_\ell := \frac{1}{n}\sum_{i=1}^n e^{\mathbf{i}2\pi\ell\boldsymbol{\tau}_i}, \ \forall \ell \in \mathbb{Z}. \tag{3.3}$$

Hence, the estimation of the intensity $\lambda \in L^2([0,1])$ can be formulated as follows: we want to estimate the sequence $(\theta_\ell)_{\ell\in\mathbb{Z}}$ of Fourier coefficients of λ from the sequence space model

$$y_\ell = \tilde{\gamma}_\ell\theta_\ell + \xi_{\ell,n}, \tag{3.4}$$

where the $\xi_{\ell,n}$ are centered random variables defined as

$$\xi_{\ell,n} := \frac{1}{n}\sum_{i=1}^n \left[\int_0^1 e_\ell(t)dN_t^i - \int_0^1 e_\ell(t)\lambda(t-\boldsymbol{\tau}_i)dt\right] \text{ for all } \ell \in \mathbb{Z}.$$

The model (3.4) is very close to the standard formulation of statistical linear inverse problems. Indeed, using the singular value decomposition of the considered operator, the standard sequence space model of an ill-posed statistical inverse problem is (see [8] and the references therein)

$$c_\ell = \theta_\ell \gamma_\ell + z_\ell, \tag{3.5}$$

where the γ_ℓ's are eigenvalues of a known linear operator, and the z_ℓ's represent an additive random noise. The issue in model (3.5) is to recover the coefficients θ_ℓ from the observations c_ℓ. A large class of estimators in model (3.5) can be written as

$$\hat{\theta}_\ell = \delta_\ell \frac{c_\ell}{\gamma_\ell},$$

where $\delta = (\delta_\ell)_{\ell \in \mathbb{Z}}$ is a sequence of reals with values in $[0,1]$ called filter (see [8] for further details).

Equation (3.4) can be viewed as a linear inverse problem with a Poisson noise for which the operator to invert is stochastic with eigenvalues $\tilde{\gamma}_\ell$ (3.3) that are unobserved random variables. Nevertheless, since the density g of the shifts is assumed to be known and $\mathbb{E}\tilde{\gamma}_\ell = \gamma_\ell$ with $\tilde{\gamma}_\ell \approx \gamma_\ell$ for n sufficiently large (in a sense which will be made precise later on), an estimation of the Fourier coefficients of f could be obtained by a deconvolution step of the form

$$\hat{\theta}_\ell = \delta_\ell \frac{y_\ell}{\gamma_\ell}, \tag{3.6}$$

where $\delta = (\delta_\ell)_{\ell \in \mathbb{Z}}$ is a filter whose choice has to be discussed.

In this paper, the following type of assumption on g is considered:

Assumption 3.1 *The Fourier coefficients of g have a polynomial decay i.e. for some real $\nu > 0$, there exist two constants $C \geq C' > 0$ such that $C'|\ell|^{-\nu} \leq |\gamma_\ell| \leq C|\ell|^{-\nu}$ for all $\ell \in \mathbb{Z}$.*

In standard inverse problems such as deconvolution, the expected optimal rate of convergence from an arbitrary estimator typically depends on such smoothness assumptions for g. The parameter ν is usually referred to as the degree of ill-posedness of the inverse problem, which quantifies the difficult of inverting the convolution operator.

3.2 A linear estimator by spectral cut-off

This part allows us to shed some light on the connexion that may exist between our model and a deconvolution problem. For a given filter $(\delta_\ell)_{\ell \in \mathbb{Z}}$ and using (3.6), a linear estimator of λ is given by

$$\hat{\lambda}_\delta(t) = \sum_{\ell \in \mathbb{Z}} \hat{\theta}_\ell e_\ell(t) = \sum_{\ell \in \mathbb{Z}} \delta_\ell \gamma_\ell^{-1} y_\ell e_\ell(t), \; t \in [0,1], \tag{3.7}$$

whose quadratic risk can be written in the Fourier domain as

$$\mathcal{R}(\hat{\lambda}_\delta, \lambda) := \mathbb{E}\left(\sum_{\ell \in \mathbb{Z}} |\hat{\theta}_\ell - \theta_\ell|^2 \right).$$

The following proposition (whose proof can be found in Section 8) illustrates how the quality of the estimator $\hat{\lambda}_\delta$ (in term of quadratic risk) is related to the choice of the filter δ.

Proposition 3.1 *For any given non-random filter δ, the risk of $\hat{\lambda}^\delta$ can be decomposed as*

$$\mathcal{R}(\hat{\lambda}^\delta, \lambda) = \sum_{\ell \in \mathbb{Z}} |\theta_\ell|^2 (\delta_\ell - 1)^2 + \sum_{\ell \in \mathbb{Z}} \frac{\delta_\ell^2}{n} |\gamma_\ell|^{-2} \|\lambda\|_1 + \sum_{\ell \in \mathbb{Z}} \frac{\delta_\ell^2}{n} |\theta_\ell|^2 \left(|\gamma_\ell|^{-2} - 1 \right). \tag{3.8}$$

where $\|\lambda\|_1 = \int_0^1 \lambda(t) dt$.

Note that the quadratic risk of any linear estimator in model (3.4) is composed of three terms. The two first terms in the risk decomposition (3.8) correspond to the classical bias and variance in statistical inverse problems. The third term corresponds to the error related to the fact that the inversion of the operator is performed using $(\gamma_l)_{l\in\mathbb{Z}}$ instead of the (unobserved) random eigenvalues $(\tilde{\gamma}_l)_{l\in\mathbb{Z}}$. Consider now the following smoothness class of functions (a Sobolev ball)

$$H_s(A) = \left\{ \lambda \in L^2([0,1]) \; ; \sum_{\ell\in\mathbb{Z}} (1+|\ell|^{2s})|\theta_\ell|^2 \leq A \text{ and } \lambda(t) \geq 0 \text{ for all } t \in [0,1] \right\},$$

for some smoothness parameter $s > 0$, where $\theta_\ell = \int_0^1 e^{-2\mathbf{i}\ell\pi t}\lambda(t)dt$.

For the sake of simplicity we only consider the family of projection (or spectral cut-off) filters $\delta^M = (\delta_\ell)_{\ell\in\mathbb{Z}} = \left(\mathbb{1}_{\{|\ell|\leq M\}}\right)_{\ell\in\mathbb{Z}}$ for some $M \in \mathbb{N}$. Using Proposition 3.1, it follows that

$$\mathcal{R}(\hat{\lambda}^{\delta^M}, \lambda) = \sum_{\ell > M} |\theta_\ell|^2 + \frac{1}{n} \sum_{|\ell|<M} \left(|\gamma_\ell|^{-2} \|\lambda\|_1 + |\theta_\ell|^2 \left(|\gamma_\ell|^{-2} - 1 \right) \right). \tag{3.9}$$

For an appropriate choice of the spectral cut-off parameter M, the following proposition gives the asymptotic behavior of the risk of $\hat{\lambda}^{\delta^M}$, see equation (3.7).

Proposition 3.2 *Assume that f belongs to $H_s(A)$ with $s > 1/2$ and $A > 0$, and that g satisfies Assumption (3.1). If $M = M_n$ is chosen as the largest integer such $M_n \leq n^{\frac{1}{2s+2\nu+1}}$, then as $n \to +\infty$*

$$\sup_{\lambda \in H_s(A)} \mathcal{R}(\hat{\lambda}^{\delta^M}, \lambda) = \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}} \right).$$

The proof follows immediately from the decomposition (3.9), the definition of $H_s(A)$ and Assumption (3.1). Remark that Proposition 3.2 shows that under Assumption 3.1 the quadratic risk $\mathcal{R}(\hat{\lambda}^{\delta^M}, \lambda)$ is of polynomial order of the sample size n, and that this rate deteriorates as the degree of ill-posedness ν increases. Such a behavior is a well known fact for standard deconvolution problems, see e.g. [26], [19] and references therein. Proposition 3.2 shows that a similar phenomenon holds for the linear estimator $\hat{\lambda}^{\delta^M}$. Hence, there may exist a connection between estimating a mean pattern intensity from a set of non-homogeneous Poisson processes and the statistical analysis of deconvolution problems.

However, the choice of $M = M_n$ in Proposition 3.2 depends on the *a priori* unknown smoothness s of the intensity λ. Such a spectral cut-off estimator is thus non-adaptive, of limited interest for applications. Moreover, the result of Proposition 3.2 is only suited for smooth functions since Sobolev balls $H_s(A)$ for $s > 1/2$ are not well adapted to model intensities λ which may have singularities. This corresponds to a classical limitation of deconvolution using the Fourier basis which is not well suited to estimate an intensity λ with spikes for instance. In Section 5, we will thus consider the problem of constructing an adaptive estimator using non-linear wavelet decompositions, and we will derive an upper bound of the quadratic risk of such estimators over Besov balls.

4 Lower bound of the minimax risk over Besov balls

Denote by $\|\lambda\|_2^2 = \int_0^1 |\lambda(t)|^2 dt$ the squared norm of a function λ belonging to the space $L^2([0,1])$ of squared integrable functions on $[0,1]$ with respect to the Lebesgue measure dt. Let $\Lambda \subset L^2([0,1])$ be some smoothness class of functions, and let $\hat{\lambda}_n \in L^2([0,1])$ denote an estimator of the intensity function $\lambda \in \Lambda$, *i.e.* a measurable mapping of the random processes N^i, $i = 1, \ldots, n$ taking its value in $L^2([0,1])$. Define the quadratic risk of any estimator $\hat{\lambda}_n$ as

$$\mathcal{R}(\hat{\lambda}_n, \lambda) := \mathbb{E}\|\hat{\lambda}_n - \lambda\|_2^2,$$

and introduce the following minimax risk

$$\mathcal{R}_n(\Lambda) = \inf_{\hat{\lambda}_n} \sup_{\lambda \in \Lambda} \mathcal{R}(\hat{\lambda}_n, \lambda),$$

where the above infimum is taken over the set of all possible estimators constructed from $N^1, \ldots, N^n$. In order to investigate the optimality of an estimator, the main contributions of this paper are deriving upper and lower bounds for $\mathcal{R}_n(\Lambda)$ when Λ is a Besov ball, and constructing an adaptive estimator that achieves a near-minimax rate of convergence.

4.1 Meyer wavelets and Besov balls

Let us denote by ψ (resp. ϕ) the periodic mother Meyer wavelet (resp. scaling function) on the interval $[0, 1]$ (see e.g. [26, 19] for a precise definition). Any intensity $\lambda \in L^2([0, 1])$ can then be decomposed as follows

$$\lambda(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{+\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(t),$$

where $\phi_{j_0,k}(t) = 2^{j_0/2}\phi(2^{j_0}t - k)$, $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$, $j_0 \geq 0$ denotes the usual coarse level of resolution, and

$$c_{j_0,k} = \int_0^1 \lambda(t)\phi_{j_0,k}(t)dt, \; \beta_{j,k} = \int_0^1 \lambda(t)\psi_{j,k}(t)dt,$$

are the scaling and wavelet coefficients of λ. It is well known that Besov spaces can be characterized in terms of wavelet coefficients (see e.g [24]). Let $s > 0$ denote the usual smoothness parameter, then for the Meyer wavelet basis and for a Besov ball $B^s_{p,q}(A)$ of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has that

$$B^s_{p,q}(A) = \left\{ f \in L^2([0,1]) : \left(\sum_{k=0}^{2^{j_0}-1} |c_{j_0,k}|^p \right)^{\frac{1}{p}} + \left(\sum_{j=j_0}^{+\infty} 2^{j(s+1/2-1/p)q} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \leq A \right\}$$

with the respective above sums replaced by maximum if $p = \infty$ or $q = \infty$. The parameter s is related to the smoothness of the function f. Note that if $p = q = 2$, then a Besov ball is equivalent to a Sobolev ball if s is not an integer. For $1 \leq p < 2$, the space $B^s_{p,q}(A)$ contains functions with local irregularities.

4.2 A lower bound of the minimax risk

The following result provides a lower bound of reconstruction in $B^s_{p,q}(A)$ over a large range of values for s, p, q.

Theorem 4.1 *Suppose that g satisfies Assumption 3.1. Introduce the class of functions*

$$\Lambda_0 = \left\{ \lambda \in L^2([0,1]); \; \lambda(t) \geq 0 \text{ for all } t \in [0,1] \right\}.$$

Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $A > 0$ and assume that $s > 2\nu + 1$. Then, there exists a constant $C_0 > 0$ (independent of n) such that for all sufficiently large n

$$\mathcal{R}_n(B^s_{p,q}(A) \cap \Lambda_0) = \inf_{\hat{\lambda}_n} \sup_{\lambda \in B^s_{p,q}(A) \bigcap \Lambda_0} \mathcal{R}(\hat{\lambda}_n, \lambda) \geq C_0 n^{-\frac{2s}{2s+2\nu+1}},$$

where the above infimum is taken over the set of all possible estimators $\hat{\lambda}_n \in L^2([0,1])$ of the intensity λ (i.e the set of all measurable mapping of the random processes N^i, $i = 1, \ldots, n$ taking their value in $L^2([0,1])$).

Hence, Theorem 4.1 shows that under Assumption 3.1 the minimax risk $\mathcal{R}_n\left(\Lambda_0 \cap B^s_{p,q}(A)\right)$ is lower bounded by the sequence $n^{-\frac{2s}{2s+2\nu+1}}$ which goes to zero at a polynomial rate as the sample size n goes to infinity, and that this rate deteriorates as the degree of ill-posedness ν increases. Such a behavior is a well known fact for standard deconvolution problems, see e.g. [26], [19] and references therein. The proof of this result is postponed to Section 9. The arguments to derive this lower bound rely on a non-standard use of Assouad's cube technique that is classically used in statistical deconvolution problems to obtain minimax properties of an estimator (see e.g. [26] and references therein).

5 Adaptive estimation in Besov spaces

In this section, we describe a statistical procedure to build an adaptive (to the unknown smoothness s of λ) estimator using Meyer wavelets.

5.1 A deconvolution step to estimate scaling and wavelet coefficients

We use Meyer wavelets to build a non-linear and adaptive estimator as follows. Meyer wavelets satisfy the fundamental property of being band-limited function in the Fourier domain which make them well suited for deconvolution problems. More precisely, each $\phi_{j,k}$ and $\psi_{j,k}$ has a compact support in the Fourier domain in the sense that

$$\phi_{j_0,k} = \sum_{\ell \in D_{j_0}} c_\ell(\phi_{j_0,k}) e_\ell, \; \psi_{j,k} = \sum_{\ell \in \Omega_j} c_\ell(\psi_{j,k}) e_\ell,$$

with

$$c_\ell(\phi_{j_0,k}) := \int_0^1 e^{-2\mathbf{i}\ell\pi t} \phi_{j_0,k}(t)dt, \; c_\ell(\psi_{j,k}) := \int_0^1 e^{-2\mathbf{i}\ell\pi t} \psi_{j,k}(t)dt,$$

and where D_{j_0} and Ω_j are finite subsets of integers such that $\#D_{j_0} \leq C2^{j_0}$, $\#\Omega_j \leq C2^j$ for some constant $C > 0$ independent of j and

$$\Omega_j \subset [-2^{j+2}c_0, -2^j c_0] \cup [2^j c_0, 2^{j+2}c_0] \tag{5.1}$$

with $c_0 = 2\pi/3$. Then, thanks to Dirichlet theorem, the scaling and wavelets coefficients of λ satisfy

$$c_{j_0,k} = \sum_{\ell \in D_{j_0}} c_\ell(\phi_{j_0,k}) \theta_\ell, \; \beta_{j,k} = \sum_{\ell \in \Omega_j} c_\ell(\psi_{j,k}) \theta_\ell. \tag{5.2}$$

Therefore, one can plug the estimator $\hat{\theta}_\ell = \gamma_\ell^{-1} y_\ell$ of each θ_ℓ, see equation (3.4), in (5.2) to build estimators of the scaling and wavelet coefficients by defining

$$\hat{c}_{j_0,k} = \sum_{\ell \in \Omega_{j_0}} c_\ell(\phi_{j_0,k}) \hat{\theta}_\ell \quad \text{and} \quad \hat{\beta}_{j,k} = \sum_{\ell \in \Omega_j} c_\ell(\psi_{j,k}) \hat{\theta}_\ell. \tag{5.3}$$

5.2 Hard thresholding estimation

We propose to use a non-linear hard thresholding estimator defined by

$$\hat{\lambda}^h_n = \sum_{k=0}^{2^{j_0(n)}-1} \hat{c}_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0(n)}^{j_1(n)} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbb{1}_{\{|\hat{\beta}_{j,k}| \geqslant \hat{s}_j(n)\}} \psi_{j,k}. \tag{5.4}$$

In the above formula, $\hat{s}_j(n)$ refers to possibly random thresholds that depend on the resolution j, while $j_0 = j_0(n)$ and $j_1 = j_1(n)$ are the usual coarsest and highest resolution levels whose dependency on n will be specified later on. Then, let us introduce some notations. For all $j \in \mathbb{N}$, define

$$\sigma_j^2 = 2^{-j} \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2} \text{ and } \epsilon_j = 2^{-j/2} \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-1}, \tag{5.5}$$

and for any $\gamma > 0$, let

$$\tilde{K}_n(\gamma) = \frac{1}{n} \sum_{i=1}^{n} K_i + \frac{4\gamma \log n}{3n} + \sqrt{\frac{2\gamma \log n}{n^2} \sum_{i=1}^{n} K_i + \frac{5\gamma^2 (\log n)^2}{3n^2}}, \tag{5.6}$$

where $K_i = \int_0^1 dN_t^i$ is the number of points of the counting process N^i for $i = 1, \ldots, n$. Introduce also the class of bounded intensity functions

$$\Lambda_\infty = \left\{ \lambda \in L^2([0,1]);\ \|\lambda\|_\infty < +\infty \text{ and } \lambda(t) \geq 0 \text{ for all } t \in [0,1] \right\},$$

where $\|\lambda\|_\infty = \sup_{t \in [0,1]} \{|\lambda(t)|\}$.

Theorem 5.1 *Suppose that g satisfies Assumption 3.1. Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $A > 0$. Let $p' = \min(2, p)$, and assume that $s > 1/p'$ and $(s + 1/2 - 1/p')p > \nu(2 - p)$. Let $\delta > 0$ and suppose that the non-linear estimator $\hat{\lambda}_n^h$ (5.4) is computed using the random thresholds*

$$\hat{s}_j(n) = 4 \left(\sqrt{\sigma_j^2 \frac{2\gamma \log n}{n} \left(\|g\|_\infty \tilde{K}_n(\gamma) + \delta \right)} + \frac{\gamma \log n}{3n} \epsilon_j \right), \quad \textit{for } j_0(n) \leq j \leq j_1(n),$$

with $\gamma \geq 2$, and where σ_j^2 and ϵ_j are defined in (5.5). Define $j_0(n)$ as the largest integer such that $2^{j_0(n)} \leq \log n$ and $j_1(n)$ as the largest integer such that $2^{j_1(n)} \leq \left(\frac{n}{\log n} \right)^{\frac{1}{2\nu+1}}$. Then, as $n \to +\infty$,

$$\sup_{\lambda \in B_{p,q}^s(A) \bigcap \Lambda_\infty} \mathcal{R}(\hat{\lambda}_n^h, \lambda) = \mathcal{O} \left(\left(\frac{\log n}{n} \right)^{\frac{2s}{2s+2\nu+1}} \right).$$

The proof of Theorem 5.1 is postponed to Section 10. Hence, Theorem 5.1 shows that under Assumption 3.1 the quadratic risk of the non-linear estimator $\hat{\lambda}_n^h$ is of polynomial order of the sample size n, and that this rate deteriorates as ν increases. Again, this result illustrates the connection between estimating a mean intensity from the observation of Poisson processes and the analysis of inverse problems in nonparametric statistics. Note that the choices of the random thresholds $\hat{s}_j(n)$ and the highest resolution level j_1 do not depend on the smoothness parameter s. Hence, contrary to the linear estimator $\hat{\lambda}^{\delta^M}$ studied in Proposition 3.2, the non-linear estimator $\hat{\lambda}_n^h$ is said to be adaptive with respect to the unknown smoothness s. Moreover, the Besov spaces $B_{p,q}^s(A)$ may contain functions with local irregularities. The above described non-linear estimator is thus suitable for the estimation of non-globally smooth functions.

In Section 4, we have shown that the rate $n^{-\frac{2s}{2s+2\nu+1}}$ is a lower bound for the asymptotic decay of the minimax risk over a large scale of Besov balls. Hence, the wavelet estimator that we propose is almost optimal up to a logarithmic term. The presence of such a term is classical in wavelet-based denoising. It corresponds to the price to pay for adaptation when using estimators based on nonlinear thresholding in a wavelet basis.

6 Conclusion and perspectives

In this paper, we have considered the problem of adaptive estimation of a non-homogeneous intensity function from the observation of n independent Poisson processes having a similar intensity λ that is randomly shifted for each observed trajectory. It has been shown that this model turns out to be an inverse problem in which the density g of the random shifts plays the role of a convolution operator. These results have been derived under the assumption that the density g is known. It is a well-known fact (see e.g. [23]) that, in standard deconvolution problems, if the convolution kernel g is unknown no satisfying rate of convergence can be obtained after a regularization process. As explained in Section 2, instead of assuming the knowledge of g, one could try to preliminary construct "estimators" of the unobserved random shifts $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ and then to average the observed processes after an alignment step using these estimated shifts. However, as shown by the results of Theorems 2.1 and 2.2, to obtain a consistent estimator of a mean pattern intensity using estimated values of the shifts, it would be necessary to consider a double asymptotic setting where both the number n of observed trajectories and an "observation time" κ (such that $\lambda(t) = \kappa\lambda_0(t)$ where λ_0 is an unknown scaled intensity to be estimated) are let going to infinity. Nevertheless, this double asymptotic setting is far beyond the scope of this paper in which we have only considered the case where n tends to infinity. Another possibility to treat the case of an unknown g would be to adopt the point of view of inverse problems with an additive Gaussian noise in the setting of partially known (or noisy) operators as in [16] and [10]. However, the assumptions made in [16] and [10] to consistently estimate an unknown operator cannot be easily adapted to our framework.

7 Proof of the results in Section 2

7.1 Proof of Theorem 2.1

Some parts of the proof of are inspired by general results on Van Trees inequalities established in [14]. Consider first the case where the shifts $\tau_i, i = 1, \ldots, n$ are non-random parameters to be estimated, and let $\tau^n = (\tau_1, \ldots, \tau_n) \in [0,1]^n$. Let $\mathcal{N} = (\mathcal{N}^1, \ldots \mathcal{N}^n)$, where $\mathcal{N}^1, \ldots \mathcal{N}^n$ are n independent Poisson processes whose intensities are specified below. Thanks to Lemma 9.1, for any real-valued and bounded measurable function h of the random variable $\mathcal{N}$, one has that

$$\mathbb{E}_\tau \left(h(\mathcal{N}) \right) = \mathbb{E}_0 \left(h(\mathcal{N}) p(\mathcal{N}|\tau^n) \right)$$

with

$$p(\mathcal{N}|\tau^n) = \prod_{i=1}^{n} \exp\left[\lambda_0 - \int_0^1 \lambda(t-\tau_i)dt + \int_0^1 \log\left(\frac{\lambda(t-\tau_i)}{\lambda_0} \right) d\mathcal{N}_t^i \right],$$

where $\mathbb{E}_\tau$ denotes the expectation of the n Poisson counting processes in $\mathcal{N}$ under the assumption that each intensity is given by $\lambda_i(t) = \lambda(t-\tau_i),\ t \in [0,1], i = 1, \ldots, n$, (where $\tau_1, \ldots, \tau_n$ are fixed parameters) and $\mathbb{E}_0$ denotes the expectation of the n Poisson counting processes in $\mathcal{N}$ under the assumption that each intensity is given by $\lambda_i(t) = \lambda_0,\ t \in [0,1], i = 1, \ldots, n$. Since

$$\frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\tau^n) = \int_0^1 \frac{\partial}{\partial t}\lambda(t-\tau_i)dt - \int_0^1 \frac{\frac{\partial}{\partial t}\lambda(t-\tau_i)}{\lambda(t-\tau_i)} d\mathcal{N}_t^i$$

it follows from (3.1) that for $i = 1, \ldots, n$

$$\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\tau^n) \right) = 0. \tag{7.1}$$

Then, for $i_1 \neq i_2$ one has that $\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_{i_1}} \log p(x|\tau^n) \frac{\partial}{\partial \tau_{i_2}} \log p(x|\tau^n) \right) = 0$, and for $i_1 = i_2$, using Proposition 6 in [27] , one obtains that

$$\mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_{i_1}} \log p(\mathcal{N}|\tau^n) \right)^2 = \operatorname{Var}\left(\int_0^1 \frac{\frac{\partial}{\partial t}\lambda(t-\tau_i)}{\lambda(t-\tau_i)} d\mathcal{N}_t^i \right) = \int_0^1 \left| \frac{\partial}{\partial t}\lambda(t) \right|^2 dt \tag{7.2}$$

Suppose now that the shifts are i.i.d. random variables with density g satisfying the assumptions of Theorem 2.1. Let $\hat{\tau}^n = \hat{\tau}^n(\mathcal{N}) \in [0,1]^n$ denote any estimator of the true random shifts $\boldsymbol{\tau}^n = (\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$. Define the following random vectors U and $V = (V_1, \ldots, V_n)'$ in $\mathbb{R}^n$ as

$$U := \hat{\tau}^n - \boldsymbol{\tau}^n \text{ and } V_i := \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\boldsymbol{\tau}^n) g_n(\boldsymbol{\tau}^n)] \frac{1}{p(\mathcal{N}|\boldsymbol{\tau}^n) g_n(\boldsymbol{\tau}^n)} \text{ for } i = 1, \ldots, n,$$

where $g_n(\boldsymbol{\tau}^n) = \prod_{i=1}^n g(\boldsymbol{\tau}_i)$. Remark first that

$$\begin{aligned}
\mathbb{E}\left(U'V\right) &= \int_{[0,1]^n} \mathbb{E}_\tau \left(\sum_{i=1}^n (\hat{\tau}_i - \tau_i) \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] \frac{1}{p(\mathcal{N}|\tau^n) g_n(\tau^n)} \right) g_n(\tau^n) d\tau^n \\
&= \int_{[0,1]^n} \mathbb{E}_0 \left(\sum_{i=1}^n (\hat{\tau}_i - \tau_i) \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] \right) d\tau^n \\
&= \mathbb{E}_0 \left(\sum_{i=1}^n \hat{\tau}_i \int_{[0,1]^n} \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] d\tau^n \right) \\
&\quad - \mathbb{E}_0 \left(\sum_{i=1}^n \int_{[0,1]^n} \tau_i \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] d\tau^n \right)
\end{aligned}$$

Thanks to the assumption that g is absolutely continuous with a compact support $[\tau_{\min}, \tau_{\max}] \subset [0,1]$ such that $\lim_{\tau \to \tau_{\min}} g(\tau) = \lim_{\tau \to \tau_{\max}} g(\tau) = 0$, it follows that

$$\forall i \in \{1 \ldots n\} \qquad \int_{[0,1]^n} \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] d\tau^n = 0.$$

Moreover, an integration by part implies that

$$\int_{[0,1]^n} \tau_i \frac{\partial}{\partial \tau_i}[p(\mathcal{N}|\tau^n) g_n(\tau^n)] d\tau^n = -\int_{[0,1]^n} p(\mathcal{N}|\tau^n) g_n(\tau^n) d\tau^n.$$

Therefore $\mathbb{E}\left(U'V\right) = \mathbb{E}_0 \left(\sum_{i=1}^n \int_{[0,1]^n} p(\mathcal{N}|\tau^n) g_n(\tau^n) d\tau^n \right) = n$ and by Cauchy-Schwarz's inequality, it follows that $n^2 = \left(\mathbb{E}\left(U'V\right)\right)^2 \leq \mathbb{E}\left(U'U\right) \mathbb{E}\left(V'V\right)$. Then, note that

$$\mathbb{E}\left(U'U\right) = \mathbb{E}\left(\sum_{i=1}^n (\hat{\tau}_i - \boldsymbol{\tau}_i)^2 \right) = \mathbb{E}_0 \left(\int_{[0,1]^n} (\hat{\tau}_i - \tau_i)^2 p(\mathcal{N}|\tau^n) g_n(\tau^n) d\tau^n \right),$$

and

$$\begin{aligned}
\mathbb{E}\left(V'V\right) &= \mathbb{E}\left(\sum_{i=1}^n \left(\frac{\partial}{\partial \tau_i} \left[\log p(\mathcal{N}|\boldsymbol{\tau}^n) + \log g_n(\boldsymbol{\tau}^n)\right] \right)^2 \right) \\
&= \mathbb{E}\left(\sum_{i=1}^n \left(\frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\boldsymbol{\tau}^n) \right)^2 \right) + \mathbb{E}\left(\sum_{i=1}^n \left(\frac{\partial}{\partial \tau_i} \log g_n(\boldsymbol{\tau}^n) \right)^2 \right),
\end{aligned}$$

since by using (7.1) it follows that

$$\mathbb{E}\left(\sum_{i=1}^n \frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\boldsymbol{\tau}^n) \frac{\partial}{\partial \tau_i} \log g_n(\boldsymbol{\tau}^n) \right) = \sum_{i=1}^n \int_{[0,1]^n} \mathbb{E}_\tau \left(\frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\tau^n) \right) \frac{\partial}{\partial \tau_i} \log g_n(\tau^n) g_n(\tau^n) d\tau^n = 0.$$

Therefore by (7.2),

$$\begin{aligned}
\mathbb{E}\left(V'V\right) &= \sum_{i=1}^n \int_{[0,1]^n} \operatorname{Var}\left(\frac{\partial}{\partial \tau_i} \log p(\mathcal{N}|\tau^n) \right) g_n(\tau^n) d\tau^n + \mathbb{E} \sum_{i=1}^n \left(\frac{\partial}{\partial \tau_i} \log g(\tau_i) \right)^2 \\
&= n \int_0^1 \left| \frac{\partial}{\partial t} \lambda(t) \right|^2 dt + n \int_0^1 \left(\frac{\partial}{\partial \tau} \log g(\tau) \right)^2 g(\tau) d\tau.
\end{aligned}$$

Since we have shown that $n^2 \leq \mathbb{E}\left(U'U\right)\mathbb{E}\left(V'V\right)$, one finally obtains the following lower bound

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_i - \boldsymbol{\tau}_i)^2\right) = \frac{1}{n}\mathbb{E}\left(U'U\right) \geq \frac{1}{\int_0^1 \left|\frac{\partial}{\partial t}\lambda(t)\right|^2 dt + \int_0^1 \left(\frac{\partial}{\partial \tau}\log g(\tau)\right)^2 g(\tau)d\tau}, \tag{7.3}$$

which completes the proof. □

7.2 Proof of Theorem 2.2

A part of the proof is inspired by a similar result in [4]. Suppose that $(\hat{\tau}_1, \ldots, \hat{\tau}_n) \in [0,1]^n$ are estimators of the true random shifts $(\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n)$ satisfying the constraints $\sum_{i=1}^n \hat{\tau}_i = 0$ and $\tau_{\min} \leq \hat{\tau}_i \leq \tau_{\max}$ for all $i = 1, \ldots, n$. Let $\bar{\boldsymbol{\tau}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\tau}_i$ and define $\lambda^\star(t) = \lambda(t - \bar{\boldsymbol{\tau}})$ for $t \in [0,1]$. Note that applying Jensen's inequality and then Minkowski's inequality implies that

$$\left(\mathbb{E}\left(\int_0^1 \left|\bar{\lambda}_n(t) - \lambda(t)\right|^2 dt\right)\right)^{1/2} \geq \mathbb{E}\left(\int_0^1 \left|\bar{\lambda}_n(t) - \lambda(t)\right|^2 dt\right)^{1/2} \geq \left|\mathbb{E}I_1 - \mathbb{E}I_2\right|. \tag{7.4}$$

where $I_1 := \left(\int_0^1 \left|\bar{\lambda}_n(t) - \lambda^\star(t)\right|^2 dt\right)^{1/2}$ and $I_2 := \left(\int_0^1 \left|\lambda^\star(t) - \lambda(t)\right|^2 dt\right)^{1/2}$. Below, we derive an asymptotic lower bound (as $n \to +\infty$) of $|\mathbb{E}I_1 - \mathbb{E}I_2|$.

Control of the term $\mathbb{E}I_1$: remind that we note $\theta_1 = \int_0^1 \lambda(t)e^{-\mathbf{i}2\pi t}dt$. The Bessel's inequality restricted to the first Fourier term implies that

$$I_1 \geq |\theta_1| \left|\frac{1}{n}\sum_{i=1}^{n}\left(e^{\mathbf{i}2\pi(\hat{\tau}_i - \boldsymbol{\tau}_i + \bar{\boldsymbol{\tau}})} - 1\right)\right|. \tag{7.5}$$

Let $u_i = 2\pi(\hat{\tau}_i - \boldsymbol{\tau}_i + \bar{\boldsymbol{\tau}})$, $i = 1, \ldots, n$. Note that, given our assumptions, $|u_i| \leq 4\pi(\tau_{\max} - \tau_{\min}) = \delta < 3$. Let $F(u_1, \ldots, u_n) = \frac{1}{n}\sum_{i=1}^n e^{\mathbf{i}u_i}$. A Taylor expansion implies that for all $(a_1, \ldots, a_n) \in [-\delta, \delta]^n$, there exist some $(t_i)_{1 \leq i \leq n} \in [-\delta, \delta]^n$ such that

$$F(a_1, \ldots, a_n) = 1 + \frac{\mathbf{i}}{n}\sum_{i=1}^{n} a_i - \frac{1}{2n}\sum_{i=1}^{n} a_i^2 - \frac{\mathbf{i}}{6n}\sum_{i=1}^{n} a_i^3 e^{\mathbf{i}t_i}.$$

Given that $\sum_{i=1}^n \hat{\tau}_i = 0$, one has that $\sum_{i=1}^n u_i = 0$, and thus, using the above Taylor expansion with $a_1 = u_1, \ldots, a_n = u_n$, it follows that

$$\left|\frac{1}{n}\sum_{i=1}^{n} e^{\mathbf{i}u_i} - 1\right| = \left|-\frac{1}{2n}\sum_{i=1}^{n} u_i^2 - \frac{\mathbf{i}}{6n}\sum_{i=1}^{n} u_i^3 e^{\mathbf{i}t_i}\right| \geq \frac{1}{2n}\left|\sum_{i=1}^{n} u_i^2 - \left|\frac{\mathbf{i}}{3}\sum_{i=1}^{n} u_i^3 e^{\mathbf{i}t_i}\right|\right|.$$

Since $|u_i| \leq \delta$ for all $i = 1, \ldots, n$, we have that $\left|\frac{\mathbf{i}}{3}\sum_{i=1}^n u_i^3 e^{\mathbf{i}t_i}\right| \leq \frac{\delta}{3}\sum_{i=1}^n |u_i|^2$ which finally implies that $\left|\frac{1}{n}\sum_{i=1}^n e^{\mathbf{i}u_i} - 1\right| \geq \frac{3-\delta}{6}\frac{1}{n}\sum_{i=1}^n u_i^2$. Combined with (7.5), it proves that

$$I_1 \geq C(\delta)|\theta_1|\frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_i - \boldsymbol{\tau}_i + \bar{\boldsymbol{\tau}})^2 \geq C(\delta)|\theta_1|\left(I_{1,1} - I_{1,2}\right)$$

with $C(\delta) = 4\pi^2\frac{3-\delta}{6} > 0$, and where

$$I_{1,1} := \frac{1}{n}\sum_{i=1}^{n}(\hat{\tau}_i - \boldsymbol{\tau}_i)^2 \text{ and } I_{1,2} := 2|\bar{\boldsymbol{\tau}}|\left(\frac{1}{n}\sum_{i=1}^{n}|\hat{\tau}_i - \boldsymbol{\tau}_i|\right).$$

Given our assumptions on $\hat{\tau}_i$ and $\boldsymbol{\tau}_i$, it follows that $I_{1,2} \leq 4(\tau_{\max} - \tau_{\min})|\bar{\boldsymbol{\tau}}|$. Then, the assumption that $\int_0^1 \tau g(\tau)d\tau = 0$ implies that $\mathbb{E}|\bar{\boldsymbol{\tau}}|^2 = \frac{1}{n}\mathbb{E}|\boldsymbol{\tau}_1|^2$, and thus

$$\mathbb{E}I_{1,2} \leq 4(\tau_{\max} - \tau_{\min})\sqrt{\mathbb{E}|\bar{\boldsymbol{\tau}}|^2} \to 0 \text{ as } n \to +\infty.$$

Therefore, by Theorem 2.1

$$\liminf_{n\to+\infty} \mathbb{E} I_1 \geq \frac{C(\delta)|\theta_1|}{\int_0^1 \left|\frac{\partial}{\partial t}\lambda(t)\right|^2 dt + \int_0^1 \left(\frac{\partial}{\partial \tau}\log g(\tau)\right)^2 g(\tau)d\tau} \tag{7.6}$$

Control of the term $\mathbb{E}I_2$: using again the fact that $\mathbb{E}|\bar{\boldsymbol{\tau}}|^2 = \frac{1}{n}\mathbb{E}|\boldsymbol{\tau}_1|^2$, and the inequality $I_2 \leq |\bar{\boldsymbol{\tau}}| \times \sup_{t\in[0,1]} \left\{\left|\frac{\partial}{\partial t}\lambda(t)\right|\right\} |\bar{\boldsymbol{\tau}}|$, one obtains that

$$\mathbb{E}I_2 \leq \sup_{t\in[0,1]} \left\{\left|\frac{\partial}{\partial t}\lambda(t)\right|\right\} \sqrt{\mathbb{E}|\bar{\boldsymbol{\tau}}|^2} \to 0, \text{ as } n \to +\infty. \tag{7.7}$$

Therefore, by combining (7.6) and (7.7), it follows that

$$\liminf_{n\to+\infty} |\mathbb{E}I_1 - \mathbb{E}I_2| \geq \frac{4\pi^2\left(\frac{3-\delta}{6}\right)|\theta_1|}{\int_0^1 \left|\frac{\partial}{\partial t}\lambda(t)\right|^2 dt + \int_0^1 \left(\frac{\partial}{\partial \tau}\log g(\tau)\right)^2 g(\tau)d\tau}$$

which completes the proof. □

8 Proof of Proposition 3.1

Remark that for all $\ell \in \mathbb{Z}$

$$\hat{\theta}_\ell - \theta_\ell = \theta_\ell \left[\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right] + \frac{\delta_\ell}{n}\sum_{i=1}^n \epsilon_{\ell,i}, \tag{8.1}$$

where the $\epsilon_{\ell,i}$ are centered random variables defined as $\epsilon_{\ell,i} = \gamma_\ell^{-1}\int_0^1 e_\ell(t)\left(dN_t^i - \lambda(t-\boldsymbol{\tau}_i)dt\right)$. Now, to compute $\mathbb{E}|\hat{\theta}_\ell - \theta_\ell|^2$, remark first that

$$\begin{aligned} |\hat{\theta}_\ell - \theta_\ell|^2 &= \left[\theta_\ell \left[\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right] + \frac{\delta_\ell}{n}\sum_{i=1}^n \epsilon_{\ell,i}\right] \overline{\left[\theta_\ell \left[\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right] + \frac{\delta_\ell}{n}\sum_{i=1}^n \epsilon_{\ell,i}\right]} \\ &= \left[|\theta_\ell|^2 \left|\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right|^2 + 2\Re\text{e}\left(\theta_\ell \left[\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right] \overline{\frac{\delta_\ell}{n}\sum_{i=1}^n \epsilon_{\ell,i}}\right) + \frac{\delta_\ell^2}{n^2}\sum_{i,i'=1}^n \epsilon_{\ell,i}\overline{\epsilon_{\ell,i'}}\right]. \end{aligned}$$

Taking expectation in the above expression yields

$$\begin{aligned} \mathbb{E}|\hat{\theta}_\ell - \theta_\ell|^2 &= \mathbb{E}\left[\mathbb{E}|\hat{\theta}_\ell - \theta_\ell|^2 \middle| \boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n\right] \\ &= \mathbb{E}\left[|\theta_\ell|^2 \left|\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right|^2 + 2\Re\text{e}\left(\theta_\ell \left[\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right] \overline{\mathbb{E}\left[\frac{\delta_\ell}{n}\sum_{i=1}^n \epsilon_{\ell,i}\right]}\right) \middle| \boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n\right] \\ &\quad + \mathbb{E}\left[\frac{\delta_\ell^2}{n^2}\sum_{i,i'=1}^n \mathbb{E}\left[\epsilon_{\ell,i}\overline{\epsilon_{\ell,i'}} \middle| \boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n\right]\right]. \end{aligned}$$

Now, remark that given two integers $i \neq i'$ and the two shifts $\boldsymbol{\tau}_i, \boldsymbol{\tau}_{i'}$, $\epsilon_{\ell,i}$ and $\epsilon_{\ell,i'}$ are independent with zero mean. Therefore, using the equality

$$\mathbb{E}\left|\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right|^2 = \delta_\ell^2 |\gamma_\ell|^{-2}\mathbb{E}|\tilde{\gamma}_\ell - \gamma_\ell|^2 + (\delta_\ell - 1)^2 = (\delta_\ell - 1)^2 + \frac{\delta_\ell^2}{n}(|\gamma_\ell|^{-2} - 1),$$

one finally obtains

$$\begin{aligned}\mathbb{E}|\hat{\theta}_\ell - \theta_\ell|^2 &= |\theta_\ell|^2 \mathbb{E}\left|\delta_\ell \frac{\tilde{\gamma}_\ell}{\gamma_\ell} - 1\right|^2 + \mathbb{E}\left[\frac{\delta_\ell^2}{n^2}\sum_{i=1}^{n} \mathbb{E}\left[|\epsilon_{\ell,i}|^2 | \boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n\right]\right] \\ &= |\theta_\ell|^2(\delta_\ell - 1)^2 + \frac{\delta_\ell^2}{n}\left(|\theta_\ell|^2\left(|\gamma_\ell|^{-2} - 1\right) + \mathbb{E}|\epsilon_{\ell,1}|^2\right).\end{aligned}$$

Using in what follows the equality $\mathbb{E}|a+\mathbf{i}b|^2 = \mathbb{E}[|a|^2+|b|^2]$ with $a = \int_0^1 \cos(2\pi\ell t)\left(dN_t^1 - \lambda(t-\boldsymbol{\tau}_1)dt\right)$ and $b = \int_0^1 \sin(2\pi\ell t)\left(dN_t^1 - \lambda(t-\boldsymbol{\tau}_1)dt\right)$, we obtain

$$\begin{aligned}\mathbb{E}|\epsilon_{\ell,1}|^2 &= |\gamma_\ell|^{-2}\mathbb{E}\left[\mathbb{E}\left|\int_0^1 e_\ell(t)\left(dN_t^1 - \lambda(t-\boldsymbol{\tau}_1)dt\right)\right|^2 |\boldsymbol{\tau}_1\right] \\ &= |\gamma_\ell|^{-2}\mathbb{E}\int_0^1 \left(|\cos(2\pi\ell t)|^2 + |\sin(2\pi\ell t)|^2\right)\lambda(t-\boldsymbol{\tau}_1)dt = |\gamma_\ell|^{-2}\|\lambda\|_1,\end{aligned}$$

where the last equality follows from the fact that λ has been extended outside $[0,1]$ by periodization, which completes the proof of Proposition 3.1.

9 Proof of the lower bound (Theorem 4.1)

9.1 Some properties of Meyer wavelets

Meyer wavelet functions satisfies the following proposition which will be useful for the construction of a lower bound of the minimax risk.

Proposition 9.1 *There exists a universal constant $c(\psi)$ such that for any $j \in \mathbb{N}$ and for any $(\omega_k)_{0\leq k\leq 2^j-1} \in \{0,1\}^{2^j}$*

$$\sup_{x\in[0,1]}\left|\sum_{k=0}^{2^j-1}\omega_k \psi_{j,k}(x)\right| \leq c(\psi)2^{j/2}.$$

Proof : Recall that the periodic Meyer mother wavelet ψ (on the interval $[0,1]$) has been obtained from the periodization of a mother Meyer wavelet, say $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$, that generates a wavelet basis of $L^2(\mathbb{R})$ (see e.g. [26, 19]). The Meyer mother wavelet $\tilde{\psi}$ is not compactly supported, but it satisfies the following inequality $\sup_{x\in\mathbb{R}}\sum_{\ell\in\mathbb{Z}}|\tilde{\psi}(x-\ell)| < \infty$, which implies that there exists some universal constant $c = c(\tilde{\psi}) > 0$ such that $\sup_{x\in\mathbb{R}}\left\{\sum_{k\in\mathbb{Z}}\left|\tilde{\psi}_{j,k}(x)\right|\right\} \leq c2^{j/2}$, for any $j \geq 0$, where $\tilde{\psi}_{j,k}(x) = 2^{j/2}\tilde{\psi}(2^j x - k)$. Hence, the proof follows using the fact that the periodic Meyer wavelet $\psi_{j,k}(x) = \sum_{\ell\in\mathbb{Z}}\tilde{\psi}_{j,k}(x-\ell)$ for $x \in [0,1]$ is the periodization of the (classical) Meyer basis $\tilde{\psi}_{j,k}$ (with infinite support). □

9.2 Definitions and notations

Recall that $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$ are i.i.d. random variables with density g, and that for $\lambda \in \Lambda_0$ a given intensity, we denote by $N^1, \ldots, N^n$ the counting processes such that conditionally to $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_n$, $N^1, \ldots, N^n$ are independent Poisson processes with intensities $\lambda(\cdot - \boldsymbol{\tau}_1), \ldots, \lambda(\cdot - \boldsymbol{\tau}_n)$. Then, the notation $\mathbb{E}_\lambda$ will be used to denote the expectation with respect to the distribution $\mathbb{P}_\lambda$ (tensorized law) of the multivariate counting process $N = \left(N^1, \ldots, N^n\right)$ with the coupled randomness on the shifts and the counting processes. In the rest of the proof, we also assume that p, q denote two integers such that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, A is a positive constant, and that s is a positive real such that $s > 2\nu + 1$, where ν is the degree of ill-posedness defined in Assumption 3.1.

A key step in the proof is the use of the likelihood ratio $\Lambda(H_0, H_1)$ between two measures associated to two hypotheses H_0 and H_1 on the intensities of the Poisson processes we consider.

The following lemma, whose proof can be found in [6], is a Girsanov's like formula for Poisson processes when random shifts do not appear in the model (notation $\tilde{\mathbb{P}}$ instead of $\mathbb{P}$ below).

Lemma 9.1 (Girsanov's like formula) *Let $\mathcal{N}_0$ (hypothesis H_0) and $\mathcal{N}_1$ (hypothesis H_1) two Poisson processes having respectively intensity $\lambda_0(t) = \rho$ and $\lambda_1(t) = \rho + \mu(t)$ for all $t \in [0,1]$, where $\rho > 0$ is a positive constant and $\mu \in \Lambda_0$ is a positive function. Let $\tilde{\mathbb{P}}_{\lambda_1}$ (resp. $\tilde{\mathbb{P}}_{\lambda_0}$) be the distribution of $\mathcal{N}_1$ (resp. $\mathcal{N}_0$). Then, the likelihood ratio between H_0 and H_1 is*

$$\Lambda(H_0, H_1)(\mathcal{N}) := \frac{d\tilde{\mathbb{P}}_{\lambda_1}}{d\tilde{\mathbb{P}}_{\lambda_0}}(\mathcal{N}) = \exp\left[-\int_0^1 \mu(t)dt + \int_0^1 \log\left(1 + \frac{\mu(t)}{\rho}\right) d\mathcal{N}_t\right], \tag{9.1}$$

where $\mathcal{N}$ is a Poisson process with intensity belonging to Λ_0.

The above lemma means that if $F(\mathcal{N})$ is a real-valued and bounded measurable function of the counting process $\mathcal{N} = \mathcal{N}_1$ (hypothesis H_1), then

$$\mathbb{E}_{H_1}\left[F(\mathcal{N})\right] = \mathbb{E}_{H_0}\left[F(\mathcal{N})\Lambda(H_0, H_1)(\mathcal{N})\right]$$

where $\mathbb{E}_{H_1}$ denotes the expectation with respect to $\tilde{\mathbb{P}}_{\lambda_1}$ (hypothesis H_1), and $\mathbb{E}_{H_0}$ denotes the expectation with respect to $\tilde{\mathbb{P}}_{\lambda_0}$ (hypothesis H_0).

Obviously, one can adapt Lemma 9.1 to the case of n independent Poisson processes $\mathcal{N} = (\mathcal{N}^1, \ldots \mathcal{N}^n)$ with respective intensities $\lambda_i(t) = \rho + \mu_i(t),\ t \in [0,1], i = 1, \ldots, n$ under H_1 and $\lambda_i(t) = \rho,\ t \in [0,1], i = 1, \ldots, n$ under H_0, where $\mu_1, \ldots, \mu_n$ are positive intensities in Λ_0. In such a case, the Girsanov's like formula (9.1) becomes

$$\Lambda(H_0, H_1)(\mathcal{N}) = \prod_{i=1}^{n} \exp\left[-\int_0^1 \mu_i(t)dt + \int_0^1 \log\left(1 + \frac{\mu_i(t)}{\rho}\right) d\mathcal{N}_t^i\right]. \tag{9.2}$$

9.3 Minoration of the minimax risk using the Assouad's cube technique

Let us first describe the main idea of the proof which expoits the Assouad's cube approach (see e.g. [7]).

- In Section 9.3.1, we build a set of test functions which are appropriate linear combinations of Meyer wavelets. The construction of this set follows the idea of the Assouad's cube technique to derive lower bounds for minimax risks (see e.g. [15, 26]).

- In Section 9.3.2 we give a key result in Lemma 9.2 that relates a lower bound on the minimax risk to a problem of statistical testing of different hypotheses. The first main step in the proof of this lemma is the use of the likelihood ratio formula (9.2). The second main step exploits the fact that, under the hypothesis that the intensity $\lambda(t) = \lambda_0(t) = \rho > 0$ is a constant function then the distribution of the data is invariant through the action of the random shifts.

- In Section 9.3.3 we specify the size of the set of test functions used in the Assouad's cube approach.

- In Section 9.3.4 we give the proof of the technical Lemma 9.3 which controls the asymptotic behavior of the likelihood ratio (9.5) defined in Lemma 9.2 under well-chosen hypotheses H_1 and H_0.

The result of Theorem 4.1 then follows from the combination of the results of these four sections.

9.3.1 Assouad's cube

Given an integer $D \geq 1$, introduce

$$S_D(A) = \{f \in \Lambda_0 \cap B^s_{p,q}(A) \quad | \quad \langle f, \psi_{j,k} \rangle = 0 \; \forall j \neq D \; \forall k \in \{0 \ldots 2^j - 1\}\}.$$

For any $\omega = (\omega_k)_{k=0,\ldots,2^D-1} \in \{0,1\}^{2^D}$ and $\ell \in \{0,\ldots,2^D-1\}$, we define $\bar{\omega}^\ell \in \{0,1\}^{2^D}$ as $\bar{\omega}^\ell_k = \omega_k, \forall k \neq l$ and $\bar{\omega}^\ell_\ell = 1 - \omega_\ell$. In what follows, we will use the likelihood ratio formula (9.2) with the intensity

$$\lambda_0(t) = \rho(A) = \frac{A}{2}, \forall t \in [0,1], \tag{9.3}$$

which corresponds to the hypothesis H_0 under which all the intensities of the observed counting processes are constant and equal to $A/2$ where A is the radius of the Besov ball $B^s_{p,q}(A)$. Next, for any $\omega \in \{0,1\}^{2^D-1}$, we denote by $\lambda_{D,\omega}$ the intensity defined as

$$\lambda_{D,\omega} = \rho(A) + \xi_D \sum_{k=0}^{2^D-1} w_k \psi_{D,k} + \xi_D 2^{D/2} c(\psi), \text{ with } \xi_D = c 2^{-D(s+1/2)}, \tag{9.4}$$

for some constant $0 < c \leq A/(2+c(\psi))$, and where $c(\psi)$ is the constant introduced in Proposition 9.1. For the sake of convenience, we omit in what follows the subscript D and write λ_ω instead of $\lambda_{D,\omega}$. First, remark that each function λ_ω can be written as $\lambda_\omega = \rho(A) + \mu_\omega$ where

$$\mu_\omega = \xi_D \sum_{k=0}^{2^D-1} w_k \psi_{D,k} + \xi_D 2^{D/2} c(\psi),$$

is a positive intensity belonging to Λ_0 by Proposition 9.1. Moreover, it can be checked that the condition $c \leq A/(2 + c(\psi))$ implies that $\lambda_\omega \in B^s_{p,q}(A)$. Therefore, $\lambda_\omega \in S_D(A)$ for any $\omega \in \{0,1\}^{2^D}$. The following lemma provides a lower bound on S_D.

9.3.2 Lower bound on the minimax risk

Lemma 9.2 *Using the notations defined in the Assouad's cube paragraph, the following inequality holds*

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in S_D(A)} \mathbb{E}_\lambda \|\hat{\lambda}_n - \lambda\|_2^2 \geq \frac{\xi_D^2}{4} \frac{1}{2^{2D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \mathbb{E}_{\lambda_\omega} \left[1 \wedge \mathcal{Q}_{k,\omega}(N)\right],$$

with $N = (N^1, \ldots, N^n)$ *and*

$$\mathcal{Q}_{k,\omega}(N) = \frac{\int_{\mathbb{R}^n} \prod_{i=1}^n \exp\left[-\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dt + \int_0^1 \log\left(1 + \frac{\mu_{\bar{\omega}^k}(t-\alpha_i)}{\rho(A)}\right) dN^i_t\right] g(\alpha_i) d\alpha_i}{\int_{\mathbb{R}^n} \prod_{i=1}^n \exp\left[-\int_0^1 \mu_\omega(t-\alpha_i)dt + \int_0^1 \log\left(1 + \frac{\mu_\omega(t-\alpha_i)}{\rho(A)}\right) dN^i_t\right] g(\alpha_i) d\alpha_i}. \tag{9.5}$$

<u>*Proof :*</u> Let $\hat{\lambda}_n = \hat{\lambda}_n(N) \in L^2([0,1])$ denote any estimator of $\lambda \in S_D(A)$ (a measurable function of the process N). Note that, to simplify the notations, we will drop in the proof the dependency of $\hat{\lambda}_n(N)$ on N and n, and we write $\hat{\lambda}$ instead of $\hat{\lambda}_n(N)$. Then, define

$$R(\hat{\lambda}) = \sup_{\lambda \in S_D(A)} \mathbb{E}_\lambda \|\hat{\lambda} - \lambda\|_2^2.$$

Since $\lambda_\omega \in S_D(A)$ for any $\omega \in \{0,1\}^{2^D}$, it follows from Parseval's relation that

$$R(\hat{\lambda}) \geq \sup_{\omega \in \{0,1\}^{2^D}} \mathbb{E}_{\lambda_\omega} \|\hat{\lambda} - \lambda_\omega\|_2^2 \geq \sup_{\omega \in \{0,1\}^{2^D}} \mathbb{E}_{\lambda_\omega} \sum_{k=0}^{2^D-1} |\beta_{D,k}(\hat{\lambda}) - \omega_k \xi_D|^2,$$

where we have used the notation $\beta_{D,k}(\hat{\lambda}) = \langle \hat{\lambda}, \psi_{D,k}\rangle$. For all $k \in \{0,\ldots,2^D-1\}$ define

$$\hat{\omega}_k = \hat{\omega}_k(N) := \arg\min_{v\in\{0,1\}} |\beta_{D,k}(\hat{\lambda}(N)) - v\xi_D|.$$

Then, the triangular inequality and the definition of $\hat{\omega}_k$ imply that

$$\xi_D|\hat{\omega}_k - \omega_k| \leq |\hat{\omega}_k\xi_D - \beta_{D,k}(\hat{\lambda})| + |\beta_{D,k}(\hat{\lambda}) - \omega_k\xi_D| \leq 2|\beta_{D,k}(\hat{\lambda}) - \omega_k\xi_D|.$$

Thus,

$$\begin{aligned} R(\hat{\lambda}) &\geq \frac{\xi_D^2}{4} \sup_{\omega\in\{0,1\}^{2^D}} \mathbb{E}_{\lambda_\omega} \sum_{k=0}^{2^D-1} |\hat{\omega}_k(N) - \omega_k|^2, \\ &\geq \frac{\xi_D^2}{4}\frac{1}{2^{2^D}} \sum_{\omega\in\{0,1\}^{2^D}} \sum_{k=0}^{2^D-1} \mathbb{E}_{\lambda_\omega}|\hat{\omega}_k(N) - \omega_k|^2. \end{aligned} \tag{9.6}$$

Let $k \in \{0,\ldots,2^D-1\}$ and $\omega \in \{0,1\}^{2^D}$ be fixed parameters. Conditionally to the vector $\boldsymbol{\tau} = (\boldsymbol{\tau}_1,\ldots\boldsymbol{\tau}_n) \in \mathbb{R}^n$, we define the two hypothesis H_0 and $H_\omega^{\boldsymbol{\tau}}$ as

H_0: $N^1,\ldots,N^n$ are independent Poisson processes with intensities $(\lambda_0(\cdot - \boldsymbol{\tau}_1),\ldots,\lambda_0(\cdot - \boldsymbol{\tau}_n)) = (\lambda_0(\cdot),\ldots,\lambda_0(\cdot))$, where λ_0 is the constant intensity defined by (9.3),

$H_\omega^{\boldsymbol{\tau}}$: $N^1,\ldots,N^n$ are independent Poisson processes with intensities $(\lambda_\omega(\cdot - \boldsymbol{\tau}_1),\ldots,\lambda_\omega(\cdot - \boldsymbol{\tau}_n))$.

In what follows, we use the notation $\mathbb{E}_{H_0}$ (resp. $\mathbb{E}_{H_\omega^{\boldsymbol{\tau}}}$) to denote the expectation under the hypothesis H_0 (resp. $H_\omega^{\boldsymbol{\tau}}$) conditionally to the shifts $\boldsymbol{\tau} = (\boldsymbol{\tau}_1,\ldots\boldsymbol{\tau}_n)$. The Girsanov formula (9.2) yields

$$\begin{aligned} \mathbb{E}_{\lambda_\omega}|\hat{\omega}_k(N) - \omega_k|^2 &= \int_{\mathbb{R}^n} \mathbb{E}_{H_1^\tau}|\hat{\omega}_k(N) - \omega_k|^2 g(\tau_1)\ldots g(\tau_n)d\tau \\ &= \int_{\mathbb{R}^n} \mathbb{E}_{H_0}\left[|\hat{\omega}_k(N) - \omega_k|^2 \Lambda(H_0, H_\omega^\tau)(N)\right] g(\tau_1)\ldots g(\tau_n)d\tau, \end{aligned}$$

with $d\tau = d\tau_1,\ldots,d\tau_n$ and

$$\Lambda(H_0, H_\omega^\tau)(N) = \prod_{i=1}^n \exp\left[-\int_0^1 \mu_\omega(t-\tau_i)dt + \int_0^1 \log\left(1 + \frac{\mu_\omega(t-\tau_i)}{\rho(A)}\right) dN_t^i\right],$$

for $N = (N^1,\ldots,N^n)$. Now, remark that under the hypothesis H_0, the law of the random variable $\hat{\omega}_k(N)$ does not depend on the random shifts $\boldsymbol{\tau} = (\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n)$ since λ_0 is a constant intensity. Thus, we obtain the following equality

$$\mathbb{E}_{\lambda_\omega}|\hat{\omega}_k(N) - \omega_k|^2 = \mathbb{E}_{H_0}\left[|\hat{\omega}_k(N) - \omega_k|^2 \int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\tau)(N) g(\tau_1)\ldots g(\tau_n)d\tau\right]. \tag{9.7}$$

Using equality (9.7), we may re-write the lower bound (9.6) on $R(\hat{\lambda})$ as

$$\begin{aligned} R(\hat{\lambda}) &\geq \frac{\xi_D^2}{4}\frac{1}{2^{2^D}} \sum_{\omega\in\{0,1\}^{2^D}} \sum_{k=0}^{2^D-1} \mathbb{E}_{H_0}\left[|\hat{\omega}_k(N) - \omega_k|^2 \int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\tau)(N) g(\tau_1)\ldots g(\tau_n)d\tau\right] \\ &= \frac{\xi_D^2}{4}\frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega\in\{0,1\}^{2^D}|w_k=1} \left(\mathbb{E}_{H_0}\left[|\hat{\omega}_k(N) - \omega_k|^2 \int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\tau)(N) g(\tau_1)\ldots g(\tau_n)d\tau\right]\right. \\ &\quad \left. + \mathbb{E}_{H_0}\left[|\hat{\omega}_k(N) - \bar{\omega}_k^k|^2 \int_{\mathbb{R}^n} \Lambda(H_0, H_{\bar{\omega}^k}^\tau)(N) g(\tau_1)\ldots g(\tau_n)d\tau\right]\right). \end{aligned}$$

The key inequality $|1-v|^2 z + |v|^2 z' \geq z \wedge z'$ (true for all $v \in \{0,1\}$ and all reals $z, z' > 0$) yields

$$\begin{aligned}
R(\hat{\lambda}) \geq\; & \frac{\xi_D^2}{4} \frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \mathbb{E}_{H_0} \left\{ \int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\tau)(N) g(\tau_1) \dots g(\tau_n) d\tau \wedge , \right. \\
& \left. \int_{\mathbb{R}^n} \Lambda(H_0, H_{\bar{\omega}^k}^\tau)(N) g(\tau_1) \dots g(\tau_n) d\tau \right\} \\
\geq\; & \frac{\xi_D^2}{4} \frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \mathbb{E}_{H_0} \int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\tau)(N) g(\tau_1) \dots g(\tau_n) d\tau \, \Big(1 \wedge \\
& \left. \frac{\int_{\mathbb{R}^n} \Lambda(H_0, H_{\bar{\omega}^k}^\alpha)(N) g(\alpha_1) \dots g(\alpha_n) d\alpha}{\int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\alpha)(N) g(\alpha_1) \dots g(\alpha_n) d\alpha} \right), \\
\geq\; & \frac{\xi_D^2}{4} \frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \int_{\mathbb{R}^n} \mathbb{E}_{H_0} \left[\Lambda(H_0, H_\omega^\tau)(N) \left(1 \wedge \mathcal{Q}_{k,\omega}(N)\right) \right] g(\tau_1) \dots g(\tau_n) d\tau,
\end{aligned}$$

where

$$\mathcal{Q}_{k,\omega}(N) = \frac{\int_{\mathbb{R}^n} \Lambda(H_0, H_{\bar{\omega}^k}^\alpha)(N) g(\alpha_1) \dots g(\alpha_n) d\alpha}{\int_{\mathbb{R}^n} \Lambda(H_0, H_\omega^\alpha)(N) g(\alpha_1) \dots g(\alpha_n) d\alpha},$$

and $d\alpha = d\alpha_1 \dots d\alpha_n$. Then, using again the formula (9.2), we obtain the lower bound

$$R(\hat{\lambda}) \geq \frac{\xi_D^2}{4} \frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \mathbb{E}_{\lambda_\omega} \left[1 \wedge \mathcal{Q}_{k,\omega} \right],$$

which is independent of $\hat{\lambda}$. This ends the proof of the lemma. □

We detail in the next paragraph how to use Lemma 9.2 with a suitable value for the parameter D in order to obtain the desired lower bound on the minimax risk.

9.3.3 Quantitative settings

In the rest of the proof, we will suppose that $D = D_n$ satisfies the asymptotic equivalence

$$2^{D_n} \sim n^{\frac{1}{2s+2\nu+1}} \text{ as } n \to +\infty. \tag{9.8}$$

To simplify the notations we will drop the subscript n, and we write $D = D_n$. For two sequences of reals $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$ we use the notation $a_n \asymp b_n$ if there exists two positive constants $C, C' > 0$ such that $C \leq \frac{a_n}{b_n} \leq C'$ for all sufficiently large n. Then, define $m_{D_n} = 2^{D_n/2} \xi_{D_n}$. Since $\xi_{D_n} = c 2^{-D_n(s+1/2)}$, it follows that

$$m_{D_n} \asymp n^{-s/(2s+2\nu+1)} \to 0$$

as $n \to \infty$. Remark also that the condition $s > 2\nu + 1$ implies that

$$n m_{D_n}^3 \asymp n^{-(s-2\nu-1)/(2s+2\nu+1)} \to 0$$

as $n \to \infty$.

9.3.4 Lower bound of the likelihood ratio $\mathcal{Q}_{k,\omega}$

The above quantitative settings combined with Lemma 9.2 will allow us to obtain a lower bound of the minimax risk. For this purpose, let $0 < \delta < 1$, and remark that Lemma 9.2 and Markov inequality imply that

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in S_D(A)} \mathbb{E}_\lambda \|\hat{\lambda} - \lambda\|_2^2 \geq \frac{\delta \xi_D^2}{4} \frac{1}{2^{2^D}} \sum_{k=0}^{2^D-1} \sum_{\omega \in \{0,1\}^{2^D} | w_k = 1} \mathbb{P}_{\lambda_\omega} \left(\mathcal{Q}_{k,\omega}(N) \geq \delta \right). \tag{9.9}$$

The remainder of the proof is thus devoted to the construction of a lower bound in probability for the random variable $\mathcal{Q}_{k,\omega}(N) := \frac{I_1}{I_2}$ where

$$I_1 = I_1(N) = \int_{\mathbb{R}^n} \prod_{i=1}^n \exp \left[- \int_0^1 \mu_{\bar{\omega}^k}(t - \alpha_i) dt + \int_0^1 \log\left(1 + \mu_{\bar{\omega}^k}(t - \alpha_i)\right) dN_t^i \right] g(\alpha_i) d\alpha_i \tag{9.10}$$

and

$$I_2 = I_2(N) = \int_{\mathbb{R}^n} \prod_{i=1}^n \exp \left[- \int_0^1 \mu_{\omega}(t - \alpha_i) dt + \int_0^1 \log\left(1 + \mu_{\omega}(t - \alpha_i)\right) dN_t^i \right] g(\alpha_i) d\alpha_i, \tag{9.11}$$

where to simplify the presentation of the proof we have taken $\rho(A) = 1$ *i.e.* $A = 2$. Then, the following lemma holds (which is also valid for $\rho(A) \neq 1$).

Lemma 9.3 *There exists $0 < \delta < 1$ and a constant $p_0(\delta) > 0$ such that for any $k \in \{0 \ldots 2^{D_n} - 1\}$, any $\omega \in \{0,1\}^{2^{D_n}}$ and all sufficiently large n*

$$\mathbb{P}_{\lambda_\omega} \left(\mathcal{Q}_{k,\omega}(N) \geq \delta \right) \geq p_0(\delta) > 0.$$

Proof :
Sketch of proof: we first give a brief summary of the main ideas of the proof. The arguments that we use are not standard due to the structure of the likelihood ratio $\mathcal{Q}_{k,\omega}(N)$ which involves a kind of mixture structure with respect to the law of the random shifts (integration over $\mathbb{R}^n$ with respect to $g(\alpha_1) \ldots g(\alpha_n) d\alpha$).

In the first part of the proof, the main idea is to use several Taylor expansions to obtain a tractable asymptotic approximation of $\mathcal{Q}_{k,\omega}(N)$. Note that due to our quantitative settings stated in Section 9.3.3, we have to provide Taylor expansions up to the second or third order (since $nm_{D_n}^2$ does not converge to 0). In the second part of the proof, we use the minoration of the log-likelihood given in equations (9.15)- (9.19), and then classical concentration inequalities to obtain lower bound in probability of $\mathcal{Q}_{k,\omega}(N)$.

Note that, in the proof, we repeatedly use the following inequalities that hold for any $\omega \in \{0,1\}^{2^{D_n}}$

$$\begin{aligned} \|\mu_\omega\|_2 \leq \ & \|\mu_\omega\|_\infty && \leq 2c(\psi) m_{D_n} \to 0, \\ \|\lambda_\omega\|_2 \leq \ & \|\lambda_\omega\|_\infty && \leq \rho(A) + 2c(\psi) m_{D_n} \to \rho(A) = 1, \end{aligned} \tag{9.12}$$

as $n \to +\infty$. Recall that $\mathcal{Q}_{k,\omega}(N) := \frac{I_1}{I_2}$ where I_1 is given by (9.10) and I_2 is given by (9.11). Finally, and to be more precise, the proof is composed of the three following steps:

- Step 1: using a second order expansion of the logarithm in order to control I_1 and I_2, we will first show that

$$\begin{aligned} \mathcal{Q}_{k,\omega}(N) \quad &\geq \quad e^{\mathcal{O}_p(nm_{D_n}^3)} \frac{\prod_{i=1}^n \int_{\mathbb{R}} g(\alpha_i) \exp\left[\int_0^1 \left\{ \mu_{\bar{\omega}^k}(t - \alpha_i) - \frac{\mu_{\bar{\omega}^k}^2(t-\alpha_i)}{2} \right\} dN_t^i \right] d\alpha_i}{\prod_{i=1}^n \int_{\mathbb{R}} g(\alpha_i) \exp\left[\int_0^1 \left\{ \mu_{\omega}(t - \alpha_i) - \frac{\mu_{\omega}^2(t-\alpha_i)}{2} \right\} dN_t^i \right] d\alpha_i}, \\ &:= \quad e^{\mathcal{O}_p(nm_{D_n}^3)} \frac{J_1}{J_2}. \end{aligned} \tag{9.13}$$

- Step 2: using again a second order expansion of the exponential, we then show that

$$\ln(\mathcal{Q}_{k,\omega}(N)) \geq \ln(J_1) - \ln(J_2) + \mathcal{O}_p(nm_{D_n}^3) \tag{9.14}$$

$$= \sum_{i=1}^{n} \left\{ \mathbb{E}_{\lambda_\omega} \left(\int_0^1 g \star \{\mu_{\bar{\omega}^k}(t) - \mu_\omega(t)\} dN_t^i \right) + \frac{1}{2}\|g \star \lambda_\omega\|_2^2 - \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2 \right. \tag{9.15}$$

$$+ \int_0^1 g \star \{\mu_{\bar{\omega}^k}(t) - \mu_\omega(t)\} dN_t^i - \mathbb{E}_{\lambda_\omega} \left(\int_0^1 g \star \{\mu_{\bar{\omega}^k}(t) - \mu_\omega(t)\} dN_t^i \right) \tag{9.16}$$

$$+ \frac{1}{2} \int_0^1 g \star \mu_\omega^2(t) dN_t^i - \frac{1}{2} \int_{\mathbb{R}} g(\alpha_i) \left(\int_0^1 \mu_\omega(t - \alpha_i) dN_t^i \right)^2 d\alpha_i \tag{9.17}$$

$$+ \frac{1}{2} \int_{\mathbb{R}} g(\alpha_i) \left(\int_0^1 \mu_{\bar{\omega}^k}(t - \alpha_i) dN_t^i \right)^2 d\alpha_i - \frac{1}{2} \int_0^1 g \star \mu_{\bar{\omega}^k}^2(t) dN_t^i \tag{9.18}$$

$$\left. - \frac{1}{2} \left(\int_0^1 g \star \mu_{\bar{\omega}^k}(t) dN_t^i \right)^2 + \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2 + \frac{1}{2} \left(\int_0^1 g \star \mu_\omega(t) dN_t^i \right)^2 - \frac{1}{2}\|g \star \lambda_\omega\|_2^2 \right\} \tag{9.19}$$

$$+ \mathcal{O}_p(nm_{D_n}^3).$$

- Step 3: it consists in controlling (9.15), (9.16), (9.17), (9.18) and (9.19), more precisely:

  1. there exists a constant $0 < c_0 < +\infty$ such that for all sufficiently large n the deterministic term (9.15) satisfies

  $$(9.15) = \sum_{i=1}^{n} \left[\mathbb{E}_{\lambda_\omega} \left(\int_0^1 g \star \{\lambda_{\bar{\omega}^k}(t) - \lambda_\omega(t)\} dN_t^i + \frac{1}{2}\|g \star \lambda_\omega\|_2^2 - \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2 \right) \right] \geq -c_0.$$

  2. there exists a constant $c_1 > 0$ such that for all sufficiently large n

  $$\mathbb{P}\left(|(9.16)| \leq c_1\right) = \mathbb{P}\left(\left| \xi_D \sum_{i=1}^{n} \int_0^1 g \star \psi_{D,k}(t) d\tilde{N}_t^i \right| \leq c_1 \right) \geq 1/2. \tag{9.20}$$

  3. (9.17) + (9.18) converges to zero in probability as $n \to +\infty$.
  4. (9.19) converges to zero in probability as $n \to +\infty$.

Putting together all these results Lemma 9.3 is finally proved.

- Proof of Step 1: since for any k, one has $\int_0^1 \psi_{D,k}(t)dt = 0$, it follows that for any ω and α, $\int_0^1 \mu_\omega(t - \alpha)dt = c(\psi)\xi_{D_n} 2^{D_n/2} = c(\psi)m_{D_n}$. Therefore,

  $$I_1 = \int_{\mathbb{R}^n} g(\alpha_1) \ldots g(\alpha_n) e^{-c(\psi)nm_{D_n}} \prod_{i=1}^{n} \exp\left[\int_0^1 \log\left(1 + \mu_{\bar{\omega}^k}(t - \alpha_i)\right) dN_t^i \right] d\alpha,$$

  and

  $$I_2 = \int_{\mathbb{R}^n} g(\alpha_1) \ldots g(\alpha_n) e^{-c(\psi)nm_{D_n}} \prod_{i=1}^{n} \exp\left[\int_0^1 \log\left(1 + \mu_\omega(t - \alpha_i)\right) dN_t^i \right] d\alpha.$$

  Let $z > 0$ be a positive real, and consider the following second order expansion of the logarithm

  $$\log(1 + z) = z - \frac{z^2}{2} + \frac{z^3}{3} u^{-3} \text{ for some } 1 \leq u \leq 1 + z. \tag{9.21}$$

  From (9.21) , we obtain

  $$\int_0^1 \log\left(1 + \mu_\omega(t - \alpha_i)\right) dN_t^i \leq \int_0^1 \left\{ \mu_\omega(t - \alpha_i) - \frac{\mu_\omega^2(t - \alpha_i)}{2} \right\} dN_t^i + \int_0^1 \mu_\omega^3(t - \alpha_i) dN_t^i, \tag{9.22}$$

and that

$$\int_0^1 \log\left(1+\mu_{\bar{\omega}^k}(t-\alpha_i)\right) dN_t^i \geq \int_0^1 \left\{\mu_{\bar{\omega}^k}((t-\alpha_i) - \frac{\mu_{\bar{\omega}^k}^2(t-\alpha_i)}{2}\right\} dN_t^i. \tag{9.23}$$

Then, remark that inequalities (9.12) imply that

$$\begin{aligned}
\mathbb{E}_{\lambda_\omega} \int_0^1 \mu_\omega^3(t-\alpha_i) dN_t^i &= \int_0^1 \mu_\omega^3(t-\alpha_i) \int_{\mathbb{R}} \lambda_\omega(t-\tau_i) g(\tau_i) d\tau_i dt \\
&\leq \|\mu_\omega\|_\infty \|\mu_\omega\|_2^2 \|\lambda_\omega\|_\infty = \mathcal{O}\left(m_{D_n}^3\right).
\end{aligned}$$

Therefore, by Markov's inequality it follows that there exists a constant $K > 0$ such that

$$\forall \gamma > 0, \quad \mathbb{P}\left(\left|\sum_{i=1}^n \int_0^1 \mu_\omega^3(t-\alpha_i) dN_t^i\right| \geq \gamma\right) \leq K\gamma^{-1} \mathbb{E} \sum_{i=1}^n \left|\int_0^1 \mu_\omega^3(t-\alpha_i) dN_t^i\right| \leq K\gamma^{-1} n m_{D_n}^3,$$

and thus

$$\sum_{i=1}^n \int_0^1 \mu_\omega^3(t-\alpha_i) dN_t^i = \mathcal{O}_p\left(n m_{D_n}^3\right). \tag{9.24}$$

Hence, using inequality (9.22), one obtains that

$$I_2 \leq e^{-c(\psi) n m_{D_n} + \mathcal{O}_p\left(n m_{D_n}^3\right)} \int_{\mathbb{R}^n} g(\alpha_1) \ldots g(\alpha_n) \prod_{i=1}^n \exp\left[\int_0^1 \left\{\mu_\omega(t-\alpha_i) - \frac{\mu_\omega^2(t-\alpha_i)}{2}\right\} dN_t^i\right] d\alpha,$$

and by inequality (9.23) it follows that

$$I_1 \geq e^{-c(\psi) n m_{D_n}} \int_{\mathbb{R}^n} g(\alpha_1) \ldots g(\alpha_n) \prod_{i=1}^n \exp\left[\int_0^1 \left\{\mu_{\bar{\omega}^k}(t-\alpha_i) - \frac{\mu_{\bar{\omega}^k}^2(t-\alpha_i)}{2}\right\} dN_t^i\right] d\alpha.$$

Combining the above inequalities and the Fubini's relation we obtain that

$$\begin{aligned}
\mathcal{Q}_{k,\omega}(N) &\geq e^{\mathcal{O}_p\left(n m_{D_n}^3\right)} \frac{\prod_{i=1}^n \int_{\mathbb{R}} g(\alpha_i) \exp\left[\int_0^1 \left\{\mu_{\bar{\omega}^k}(t-\alpha_i) - \frac{\mu_{\bar{\omega}^k}^2(t-\alpha_i)}{2}\right\} dN_t^i\right] d\alpha_i}{\prod_{i=1}^n \int_{\mathbb{R}} g(\alpha_i) \exp\left[\int_0^1 \left\{\mu_\omega(t-\alpha_i) - \frac{\mu_\omega^2(t-\alpha_i)}{2}\right\} dN_t^i\right] d\alpha_i}, \\
&:= e^{\mathcal{O}_p\left(n m_{D_n}^3\right)} \frac{J_1}{J_2}. \qquad (9.25)
\end{aligned}$$

- <u>Proof of Step 2:</u> let $z \in \mathbb{R}$ and consider the following second order expansion of the exponential

$$\exp(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} \exp(u) \text{ for some } -|z| \leq u \leq |z|. \tag{9.26}$$

Let us now use (9.26) with $z_i = \int_0^1 \left\{\mu_{\bar{\omega}^k}(t-\alpha_i) - \frac{1}{2}\mu_{\bar{\omega}^k}^2(t-\alpha_i)\right\} dN_t^i$. By inequalities (9.12), one has that

$$\begin{aligned}
\mathbb{E}_{\lambda_\omega} |z_i| &\leq \int_0^1 \left(\mu_{\bar{\omega}^k}(t-\alpha_i) + \frac{1}{2}\mu_{\bar{\omega}^k}^2(t-\alpha_i)\right) \int_{\mathbb{R}} \lambda_\omega(t-\tau_i) g(\tau_i) d\tau_i dt, \\
&\leq \|\lambda_\omega\|_\infty \left(\|\mu_{\bar{\omega}^k}\|_2 + \frac{1}{2}\|\mu_{\bar{\omega}^k}\|_2^2\right) = \mathcal{O}\left(m_{D_n}\right).
\end{aligned}$$

Since $m_{D_n} \to 0$, we obtain by using (9.26) that for each $i \in \{1, \dots, n\}$,

$$
\begin{aligned}
&\exp(z_i) \\
&= 1 + \int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i - \frac{1}{2}\int_0^1 \mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i \\
&\quad + \frac{1}{2}\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i - \frac{1}{2}\int_0^1 \mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 + \frac{z_i^3}{6}e^{u_i}, \qquad \text{with } -|z_i| \le u_i \le |z_i| \\
&= 1 + \int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i - \frac{1}{2}\int_0^1 \mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i + \frac{1}{2}\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 + \frac{z_i^3}{6}e^{u_i} + R_i,
\end{aligned}
$$

where

$$
2R_i = \left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i - \frac{1}{2}\int_0^1 \mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 - \left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2.
$$

By inequalities (9.12) one obtains that for any shift $\alpha_i \in [0,1]$:

$$
2R_i = \left(-\frac{1}{2}\int_0^1 \mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right) \times \left(\int_0^1 2\mu_{\bar{\omega}^k}(t-\alpha_i) - \frac{1}{2}\mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right), \tag{9.27}
$$

and Cauchy-Schwarz's inequality yields a bound uniform in α_i:

$$
\mathbb{E}_{\lambda_\omega}(|R_i|) = \mathcal{O}\left(m^3_{D_n}\right). \tag{9.28}
$$

From the definition of J_1 in (9.25), we can use a stochastic version of the Fubini theorem (see [17], Theorem 5.44) to obtain

$$
\begin{aligned}
J_1 &= \prod_{i=1}^n \Bigg[1 + \int_{\mathbb{R}}\int_0^1 g(\alpha_i)\mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i d\alpha_i - \frac{1}{2}\int_{\mathbb{R}}\int_0^1 g(\alpha_i)\mu^2_{\bar{\omega}^k}(t-\alpha_i)dN_t^i d\alpha_i \\
&\quad + \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 d\alpha_i + \int_0^1 \left(\frac{z_i^3}{6}e^{u_i} + R_i\right) g(\alpha_i)d\alpha_i\Bigg], \\
&= \prod_{i=1}^n \Bigg[1 + \int_0^1 g \star \mu_{\bar{\omega}^k}(t)dN_t^i - \frac{1}{2}\int_0^1 g \star \mu^2_{\bar{\omega}^k}(t)dN_t^i \\
&\quad + \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 d\alpha_i + \int_0^1 \left(\frac{z_i^3}{6}e^{u_i} + R_i\right) g(\alpha_i)d\alpha_i\Bigg].
\end{aligned}
$$

At this step, it will be more convenient to work with the logarithm of the term J_1. We have

$$
\begin{aligned}
\ln(J_1) &= \sum_{i=1}^n \ln\Bigg[1 + \int_0^1 g \star \mu_{\bar{\omega}^k}(t)dN_t^i - \frac{1}{2}\int_0^1 g \star \mu^2_{\bar{\omega}^k}(t)dN_t^i \\
&\quad + \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 d\alpha_i + \int_0^1 \left(\frac{z_i^3}{6}e^{u_i} + R_i\right) g(\alpha_i)d\alpha_i\Bigg].
\end{aligned}
$$

Using again the second order expansion of the logarithm (9.21), we obtain that

$$
\begin{aligned}
\ln(J_1) &= \sum_{i=1}^n \Bigg[\int_0^1 g \star \mu_{\bar{\omega}^k}(t)dN_t^i - \frac{1}{2}\int_0^1 g \star \mu^2_{\bar{\omega}^k}(t)dN_t^i \\
&\quad + \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 d\alpha_i - \frac{1}{2}\left(\int_0^1 g \star \mu_{\bar{\omega}^k}(t)dN_t^i\right)^2 + \int_0^1 \left(\frac{z_i^3}{6}e^{u_i} + \tilde{R}_i\right) g(\alpha_i)d\alpha_i\Bigg],
\end{aligned}
$$

where $\tilde{R}_i$ is a remainder term that can be shown to satisfy $\mathbb{E}\left(|\tilde{R}_i|\right) = \mathcal{O}\left(m_{D_n}^3\right)$ by using the same arguments to derive (9.28). By a similar expansion of the term J_2 defined in (9.25), we obtain that

$$\begin{aligned}\ln(J_2) &= \sum_{i=1}^{n}\left[\int_0^1 g\star\mu_\omega(t)dN_t^i - \frac{1}{2}\int_0^1 g\star\mu_\omega^2(t)dN_t^i\right.\\ &+\frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_\omega(t-\alpha_i)dN_t^i\right)^2 d\alpha_i - \frac{1}{2}\left(\int_0^1 g\star\mu_\omega(t)dN_t^i\right)^2 + \left.\int_0^1\left(\frac{z_i^3}{6}e^{u_i}+\bar{R}_i\right)g(\alpha_i)d\alpha_i\right],\end{aligned}$$

for some remainder term $\bar{R}_i$ satisfying also $\mathbb{E}\left(|\bar{R}_i|\right) = \mathcal{O}\left(m_{D_n}^3\right)$. Then, by Markov's inequality and using the same arguments that those used to derive (9.24) we obtain that

$$\sum_{i=1}^{n}\bar{R}_i = \mathcal{O}_p(nm_{D_n}^3) \quad \text{and} \quad \sum_{i=1}^{n}\tilde{R}_i = \mathcal{O}_p(nm_{D_n}^3). \tag{9.29}$$

Now, let us study the term $\displaystyle\int_0^1\sum_{i=1}^{n}\frac{z_i^3}{6}e^{u_i}g(\alpha_i)d\alpha_i$.

Remark that since $|u_i|\leq|z_i|$, we have $\displaystyle\left|\int_0^1\sum_{i=1}^{n}\frac{z_i^3}{6}e^{u_i}g(\alpha_i)d\alpha_i\right| \leq \sum_{i=1}^{n}\sum_{k\geq0}\int_0^1\frac{|z_i|^{3+k}}{6k!}g(\alpha_i)d\alpha_i$.

Hence, for any $\gamma>0$ and by Markov's inequality, one has

$$\begin{aligned}\mathbb{P}\left(\left|\int_0^1\sum_{i=1}^{n}\frac{z_i^3}{6}e^{u_i}g(\alpha_i)d\alpha_i\right|\geq\gamma\right) &\leq \mathbb{P}\left(\sum_{i=1}^{n}\sum_{k\geq0}\int_0^1\frac{|z_i|^{3+k}}{6k!}g(\alpha_i)d\alpha_i\geq\gamma\right),\\ &\leq \frac{n}{6\gamma}\sum_{k\geq0}\int_0^1\mathbb{E}\left(\frac{|z_i|^{3+k}}{k!}\right)g(\alpha_i)d\alpha_i.\end{aligned}$$

Moreover, by inequality (9.12) it follows that for any $\alpha_i\in[0,1]$:

$$\begin{aligned}\mathbb{E}\left(\frac{|z_i|^{3+k}}{k!}\right) &= \frac{1}{k!}\mathbb{E}\left(\left|\int_0^1\left\{\mu_{\bar{\omega}^k}(t-\alpha_i)-\frac{1}{2}\mu_{\bar{\omega}^k}^2(t-\alpha_i)\right\}dN_t^i\right|\right)^{3+k}\\ &\leq \frac{1}{k!}\mathbb{E}\left(\int_0^1\left(2c(\psi)m_{D_n}+\frac{1}{2}(2c(\psi))^2m_{D_n}^2\right)dN_t^i\right)^{3+k}\end{aligned}$$

For n large enough, we have $2c(\psi)m_{D_n}+\frac{1}{2}(2c(\psi))^2m_{D_n}^2\leq4c(\psi)m_{D_n}$. Then, using the fact that if X is a Poisson random variable with intensity μ_0, the p-th moment of X is bounded by $(p+\mu_0)\mu_0^{p-1}$ one obtains that

$$\mathbb{E}\left(\frac{|z_i|^{3+k}}{k!}\right)\leq\frac{1}{k!}(4c(\psi)m_{D_n}\|\lambda\|_1)^{3+k}+\frac{4}{(k-1)!}(4c(\psi)m_{D_n})^4\|\lambda\|_1^3(4c(\psi)m_{D_n}\|\lambda\|_1)^{3+k-1},$$

which implies that uniformly in α_i:

$$\sum_{k\geq0}\mathbb{E}\left(\frac{|z_i|^{3+k}}{k!}\right)\leq(4c(\psi))^3m_{D_n}^3\|\lambda\|_1^3\left(1+1024c(\psi)m_{D_n}\right)e^{4c(\psi)m_{D_n}\|\lambda\|_1}.$$

Hence,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\int_0^1\frac{z_i^3}{6}e^{u_i}g(\alpha_i)d\alpha_i\right|\geq\gamma\right)\leq\frac{(4c(\psi))^3nm_{D_n}^3\|\lambda\|_1^3\left(1+1024c(\psi)m_{D_n}\right)e^{4c(\psi)m_{D_n}\|\lambda\|_1}}{6\gamma}, \tag{9.30}$$

which proves that $\sum_{i=1}^n\frac{z_i^3}{6}e^{u_i}=\mathcal{O}_p(nm_{D_n}^3)$. Therefore, combing the above equalities for $\ln J_1$ and $\ln J_2$ and (9.29), (9.30), we finally obtain the lower bound (9.14) for $\ln(\mathcal{Q}_{k,\omega}(N))$.

- Proof of Step 3: in what follows, we will show that, for all sufficiently large n, the terms (9.15)-(9.19) are bounded from below (in probability). Since $nm_{D_n}^3 \to 0$, this will imply that there exists $c > 0$ (not depending on λ_ω) and a constant $p(c) > 0$ such that for all sufficiently large n

$$\mathbb{P}_{\lambda_\omega}\left(\ln\left(\mathcal{Q}_{k,\omega}(N)\right) \geq -c\right) = \mathbb{P}_{\lambda_\omega}\left(\mathcal{Q}_{k,\omega}(N) \geq \exp(-c)\right) \geq p(c) > 0$$

which is the result stated in Lemma 9.3.

Lower bound for (9.15): since for any $1 \leq i \leq n$

$$\mathbb{E}_{\lambda_\omega}\left(\int_0^1 g \star \{\mu_{\bar{\omega}^k}(t) - \mu_\omega(t)\} dN_t^i\right) \quad = \quad \int_0^1 g \star \{\lambda_{\bar{\omega}^k}(t) - \lambda_\omega(t)\}\{g \star \lambda_\omega(t)\} dt.$$

We obtain that

$$\sum_{i=1}^n \left[\mathbb{E}_{\lambda_\omega}\left(\int_0^1 g \star \{\lambda_{\bar{\omega}^k}(t) - \lambda_\omega(t)\} dN_t^i + \frac{1}{2}\|g \star \lambda_\omega\|_2^2 - \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2\right)\right] \quad = \quad -\frac{n}{2}\|g \star \{\mu_\omega - \mu_{\bar{\omega}^k}\}\|_2^2.$$

Remark that $\mu_\omega - \mu_{\bar{\omega}^k} = \pm \xi_D \psi_{Dk}$. In what follows we will repeatidely use the following relation

$$\|\psi_{Dk} \star g\|_2^2 = \int_0^1 \left(\psi_{Dk} \star g(t)\right)^2 dt = \sum_{\ell \in \Omega_D} |c_\ell(\psi_{Dk})|^2 |\gamma_\ell|^2 \asymp 2^{-2D\nu} \qquad (9.31)$$

which follows from Parseval's relation, from the fact that $\#\Omega_D \asymp 2^D$ and that under Assumption 3.1 $|\gamma_\ell| \asymp 2^{-D\nu}$ for all $\ell \in \Omega_D$. Therefore

$$\|g \star \{\mu_\omega - \mu_{\bar{\omega}^k}\}\|_2^2 = \xi_D^2 \int_0^1 \left(\psi_{Dk} \star g(t)\right)^2 dt \asymp \xi_D^2 2^{-2D\nu} \asymp n^{-1},$$

and

$$-\sum_{i=1}^n \left[\mathbb{E}_{\lambda_\omega}\left(\int_0^1 g \star \{\lambda_{\bar{\omega}^k}(t) - \lambda_\omega(t)\} dN_t^i + \frac{1}{2}\|g \star \lambda_\omega\|_2^2 - \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2\right)\right] \asymp 1,$$

which implies that there exists a constant $0 < c_0 < +\infty$ such that for all sufficiently large n the deterministic term (9.15) satisfies

$$(9.15) = \sum_{i=1}^n \left[\mathbb{E}_{\lambda_\omega}\left(\int_0^1 g \star \{\lambda_{\bar{\omega}^k}(t) - \lambda_\omega(t)\} dN_t^i + \frac{1}{2}\|g \star \lambda_\omega\|_2^2 - \frac{1}{2}\|g \star \lambda_{\bar{\omega}^k}\|_2^2\right)\right] \geq -c_0.$$

In the rest of the proof, we show that, for all sufficiently large n, the terms (9.16)-(9.19) are bounded from below in probability. Without loss of generality, we consider only the case $\mu_\omega - \mu_{\bar{\omega}^k} = \xi_D \psi_{Dk}$.

Lower bound for (9.16): rewrite first (9.16) as

$$(9.16) = -\xi_D \sum_{i=1}^n \int_0^1 g \star \psi_{D,k}(t) d\tilde{N}_t^i,$$

where $d\tilde{N}_t^i = d\tilde{N}_t^i - \lambda(t - \tau_i)dt$. Then, using the fact that, conditonnaly to $\tau_1, \ldots, \tau_n$, the counting process $\sum_{i=1}^n N^i$ is a Poisson process with intensity $\sum_{i=1}^n \lambda_\omega(t - \tau_i)$, it follows from an analogue of Bennett's inequality for Poisson processes (see e.g. Proposition 7 in [27]) that for any $y > 0$

$$\mathbb{P}\left(\left|\xi_D \sum_{i=1}^n \int_0^1 g\star\psi_{D,k}(t)d\tilde{N}_t^i\right| \leq \sqrt{2y\xi_D^2\int_0^1\sum_{i=1}^n |g\star\psi_{D,k}(t)|^2\lambda_\omega(t-\tau_i)dt} + \frac{1}{3}y\xi_D\|g\star\psi_{D,k}\|_\infty \middle| \tau_1,\dots,\tau_n\right) \geq 1-\exp(-y)$$

Since $\int_0^1\sum_{i=1}^n |g\star\psi_{D,k}(t)|^2\lambda_\omega(t-\tau_i)dt \leq n\|g\star\psi_{D,k}\|_2^2\|\lambda_\omega\|_\infty$ for any $\tau_1,\dots,\tau_n$, letting $y=\log(2)$

$$\mathbb{P}\left(\left|\xi_D \sum_{i=1}^n \int_0^1 g\star\psi_{D,k}(t)d\tilde{N}_t^i\right| \leq \sqrt{2\log(2)\xi_D^2 n\|g\star\psi_{D,k}(t)\|_2^2\|\lambda_\omega\|_\infty} + \frac{1}{3}\log(2)\xi_D\|g\star\psi_{D,k}\|_\infty\right) \geq 1/2.$$

Now, using that $\xi_D^2 n\|g\star\psi_{D,k}(t)\|_2^2\|\lambda_\omega\|_\infty \asymp 1$ and $\xi_D\|g\star\psi_{D,k}\|_\infty \leq \|\psi\|_\infty 2^{D/2}\xi_D \to 0$, we can deduce that there exists a constant $c_1>0$ such that for all sufficiently large n

$$\mathbb{P}\left(|(9.16)|\leq c_1\right) = \mathbb{P}\left(\left|\xi_D \sum_{i=1}^n \int_0^1 g\star\psi_{D,k}(t)d\tilde{N}_t^i\right| \leq c_1\right) \geq 1/2. \qquad (9.32)$$

Lower bound for (9.17) and (9.18): define

$$\begin{aligned} X_i &= \frac{1}{2}\int_0^1 g\star\mu_\omega^2(t)dN_t^i - \frac{1}{2}\int_0^1 g\star\mu_{\bar{\omega}^k}^2(t)dN_t^i \\ &+\frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)dN_t^i\right)^2 d\alpha_i - \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\int_0^1 \mu_\omega(t-\alpha_i)dN_t^i\right)^2 d\alpha_i, \end{aligned}$$

and note that $(9.17)+(9.18) = \sum_{i=1}^n X_i$. For any $1\leq i\leq n$

$$\begin{aligned} \mathbb{E}_{\lambda_\omega}X_i &= \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\left(\int_0^1 \mu_{\bar{\omega}^k}(t-\alpha_i)g\star\lambda_\omega(t)dt\right)^2 - \left(\int_0^1 \mu_\omega(t-\alpha_i)g\star\lambda_\omega(t)dt\right)^2\right)d\alpha_i \\ &= \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\left(\int_0^1 -\xi_D\psi_{D,k}(t-\alpha_i)g\star\lambda_\omega(t)dt\right)\left(\int_0^1 \left(\mu_\omega(t-\alpha_i)-\mu_{\bar{\omega}^k}(t-\alpha_i)\right)g\star\lambda_\omega(t)dt\right)\right) \\ &= \frac{1}{2}\int_{\mathbb{R}} g(\alpha_i)\left(\left(\int_0^1 -\xi_D\psi_{D,k}(t-\alpha_i)g\star\mu_\omega(t)dt\right)\left(\int_0^1 \left(\mu_\omega(t-\alpha_i)-\mu_{\bar{\omega}^k}(t-\alpha_i)\right)g\star\mu_\omega(t)dt\right)\right) \end{aligned}$$

which implies that

$$|\mathbb{E}_{\lambda_\omega}X_i| \leq \frac{1}{2}\xi_D 2^{D/2}\|\psi\|_\infty\|\mu_\omega\|_\infty^2\left(\|\mu_\omega\|_\infty+\|\mu_{\bar{\omega}^k}\|_\infty\right) \asymp m_{D_n}^4.$$

Therefore $\sum_{i=1}^n \mathbb{E}_{\lambda_\omega}X_i \to 0$ as $n\to+\infty$, since $nm_{D_n}^4\to 0$. Now, remark that $X_1,\dots,X_n$ are i.i.d variables satisfying for all $1\leq i\leq n$

$$|X_i| \leq \frac{1}{2}(\|\mu_\omega\|_\infty^2+\|\mu_{\bar{\omega}^k}\|_\infty^2)(K_i+K_i^2) \leq 2c^2(\psi)m_{D_n}^2(K_i+K_i^2) \qquad (9.33)$$

where $K_i=\int_0^1 dN_t^i$. Conditionally to τ_i, K_i is a Poisson variable with intensity $\int_0^1\lambda_\omega(t-\tau_i)dt = \int_0^1\lambda_\omega(t)dt = \|\lambda_\omega\|_1$. Hence, the bound (9.12) for $\|\lambda_\omega\|_\infty$ and inequality (9.33) implies that there exists a constant $C>0$ (not depending on λ_ω) such that

$$\mathbb{E}X_1^2 \leq Cm_{D_n}^4,$$

which implies that $\mathrm{Var}(\sum_{i=1}^n X_i) = n\mathrm{Var}(X_1) \leq n\mathbb{E}X_1^2 \to 0$ as $n\to+\infty$ since $nm_{D_n}^4\to 0$. Therefore, $(9.17)+(9.18)=\sum_{i=1}^n X_i$ converges to zero in probability as $n\to+\infty$ using

Chebyshev's inequality.

Lower bound for (9.19): we denote by S_i the difference

$$S_i := 2\left(-\frac{1}{2}\left(\int_0^1 g\star\mu_{\bar{\omega}^k}(t)dN_t^i\right)^2+\frac{1}{2}\|g\star\lambda_{\bar{\omega}^k}\|_2^2+\frac{1}{2}\left(\int_0^1 g\star\mu_{\omega}(t)dN_t^i\right)^2-\frac{1}{2}\|g\star\lambda_{\omega}\|_2^2\right),$$

and remark that (9.19) $=\frac{1}{2}\sum_{i=1}^n S_i$. First, we have

$$\begin{aligned}\mathbb{E}_{\lambda_\omega}S_i &= \|g\star\lambda_{\bar{\omega}^k}\|_2^2-\|g\star\lambda_\omega\|_2^2+\int_0^1(g\star\mu_\omega)^2(t)g\star\lambda_\omega(t)dt-\int_0^1(g\star\mu_{\bar{\omega}^k})^2(t)g\star\lambda_\omega(t)dt\\ &\quad+\int_{\mathbb{R}}g(\tau_i)\left(\left\{\int_0^1(g\star\mu_\omega)(t)\lambda_\omega(t-\tau_i)dt\right\}^2-\left\{\int_0^1(g\star\mu_{\bar{\omega}^k})(t)\lambda_\omega(t-\tau_i)dt\right\}^2\right)d\tau_i.\end{aligned}$$

Since $\|g\star\mu_{\bar{\omega}^k}\|_2^2-\|g\star\mu_\omega\|_2^2=\|g\star\lambda_{\bar{\omega}^k}\|_2^2-\|g\star\lambda_\omega\|_2^2$ and $g\star\lambda_\omega=1+g\star\mu_\omega$ it follows that

$$\begin{aligned}\mathbb{E}_{\lambda_\omega}S_i &= \underbrace{\int_0^1(g\star\mu_\omega)^2(t)g\star\mu_\omega(t)dt-\int_0^1(g\star\mu_{\bar{\omega}^k})^2(t)g\star\mu_\omega(t)dt}_{S_{i,1}}\\ &\quad+\underbrace{\int_{\mathbb{R}}g(\tau_i)\left(\left\{\int_0^1(g\star\mu_\omega)(t)\lambda_\omega(t-\tau_i)dt\right\}^2-\left\{\int_0^1(g\star\mu_{\bar{\omega}^k})(t)\lambda_\omega(t-\tau_i)dt\right\}^2\right)d\tau_i}_{S_{i,2}}.\end{aligned}$$

One has that

$$|S_{i,1}|\le\|\mu_\omega\|_\infty^3+\|\mu_{\bar{\omega}^k}\|_\infty^2\|\mu_\omega\|_\infty\le 16c^3(\psi)m_{D_n}^3,$$

and that

$$S_{i,2} = \xi_D^2\int_{\mathbb{R}}g(\tau_i)\left(\left(\int_0^1 g\star\psi_{D,k}(t)\lambda_\omega(t-\tau_i)dt\right)\left(\int_0^1 g\star(\mu_\omega+\mu_{\bar{\omega}^k})(t)\lambda_\omega(t-\tau_i)dt\right)\right)d\tau_i$$

Hence using (9.12) and (9.31) it follows that there exists a constant $C>0$ such that for all sufficiently large n

$$|S_{i,2}|\le\xi_D^2\|g\star\psi_{D,k}\|_2\left(\|\mu_\omega\|_\infty+\|\mu_{\bar{\omega}^k}\|_\infty\right)\le Cn^{-\frac{3s+\nu+1}{2s+2\nu+1}}$$

Then, since $s>2\nu+1>\nu$ it follows that

$$\sum_{i=1}^n\mathbb{E}_{\lambda_\omega}S_i=\mathcal{O}\left(n^{-\frac{(s-2\nu-1)}{2s+2\nu+1}}+n^{-\frac{(s-\nu)}{2s+2\nu+1}}\right)\to 0.$$

Now, note that $\mathrm{Var}(\sum_{i=1}^n S_i)=n\mathrm{Var}(Y_1)$ where

$$Y_1=\left(\int_0^1 g\star\mu_\omega(t)dN_t^1\right)^2-\left(\int_0^1 g\star\mu_{\bar{\omega}^k}(t)dN_t^1\right)^2.$$

Since $|Y_1|\le\left(\|\mu_\omega\|_\infty^2+\|\mu_{\bar{\omega}^k}\|_\infty^2\right)K_1^2$ with $K_1=\int_0^1 dN_t^1$ being, conditionally to τ_1, a Poisson variable with intensity $\int_0^1\lambda_\omega(t-\tau_1)dt=\int_0^1\lambda_\omega(t)dt=\|\lambda_\omega\|_1$. Therefore, (9.12) again implies that there exists a constant $C>0$ (not depending on λ_ω) such that

$$\mathrm{Var}(\sum_{i=1}^n S_i)=n\mathrm{Var}(Y_1)\le n\mathbb{E}Y_1^2\le Cnm_{D_n}^4\to 0.$$

Therefore, using Chebyshev's inequality, we obtain that (9.19) $=\frac{1}{2}\sum_{i=1}^n S_i$ converges to zero in probability as $n\to+\infty$. □

9.4 Lower bound on $B_s^{p,q}(A)$

By applying inequality (9.9) and Lemma 9.3, we obtain that there exists $0 < \delta < 1$ such that for all sufficiently large n

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in S_D(A)} \mathbb{E}_\lambda \|\hat{\lambda}_n - \lambda\|_2^2 \geq C \xi_{D_n}^2 2^{D_n},$$

for some constant $C > 0$ that is independent of D_n. From the definition (9.4) of ξ_{D_n} and using the choice (9.8) for D_n, we obtain that

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in S_D(A)} \mathbb{E}_\lambda \|\hat{\lambda}_n - \lambda\|_2^2 \geq C \xi_{D_n}^2 2^{D_n} \asymp 2^{-2sD_n} \asymp n^{-\frac{2s}{2s+2\nu+1}}.$$

Now, since $S_D(A) \subset B_{p,q}^s(A)$ for any $D \geq 1$ we obtain from the above inequalities that there exists a constant $C_0 > 0$ such that for all sufficiently large n

$$\begin{aligned} \inf_{\hat{\lambda}_n} \sup_{\lambda \in B_{p,q}^s(A) \bigcap \Lambda_0} n^{\frac{2s}{2s+2\nu+1}} \mathbb{E}_\lambda \|\hat{\lambda}_n - \lambda\|_2^2 &\geq \inf_{\hat{\lambda}_n} \sup_{\lambda \in S_{D_n}(A)} \mathbb{E}_\lambda \|\hat{\lambda} - \lambda\|_2^2, \\ &\geq C_0 n^{-\frac{2s}{2s+2\nu+1}}, \end{aligned}$$

which concludes the proof of Theorem 4.1. □

10 Proof of the upper bound (Theorem 5.1)

Following standard arguments in wavelet thresholding (see e.g. [26]), one needs to bound the centered moment of order 2 and 4 of $\hat{c}_{j_0,k}$ and $\hat{\beta}_{j,k}$ (see Proposition 10.1), as well as the deviation in probability between $\hat{\beta}_{j,k}$ and $\beta_{j,k}$ (see Proposition 10.2). In the proof, C, C', C_1, C_2 denote positive constants that are independent of λ and n, and whose value may change from line to line. The proof requires technical results that are postponed and proved in Section 10.2. We will use the following quantities

$$\tilde{\psi}_{j,k}(t) = \sum_{\ell \in \Omega_j} \gamma_\ell^{-1} c_\ell(\psi_{j,k}) e_\ell(t), \quad V_j^2 = \|g\|_\infty 2^{-j} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|^2}{|\gamma_\ell|^2}, \; \delta_j = 2^{-j/2} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|}{|\gamma_\ell|},$$

and

$$\Delta_{jk}^n(\gamma) = \sqrt{\|\tilde{\psi}_{j,k}\|_2^2 \left(\|g\|_\infty \tilde{K}_n(\gamma) \frac{2\gamma \log n}{n} + u_n(\gamma) \right)} + \frac{\gamma \log n}{3n} \|\tilde{\psi}_{j,k}\|_\infty, \tag{10.1}$$

where $\tilde{K}_n(\gamma)$ is introduced in (5.6), $u_n(\gamma)$ is a real sequence such that $u_n(\gamma) = o\left(\frac{\gamma \log n}{n}\right)$ as $n \to +\infty$.

10.1 Proof of Theorem 5.1

As classically done in wavelet thresholding, use the following risk decomposition

$$\mathbb{E}\|\hat{\lambda}_n^h - \lambda\|_2^2 = R_1 + R_2 + R_3 + R_4,$$

where

$$\begin{aligned} R_1 &= \sum_{k=0}^{2^{j_0}-1} \mathbb{E}(\hat{c}_{j_0,k} - c_{j_0,k})^2, \quad R_2 = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\beta}_{j,k} - \beta_{j,k})^2 \mathbb{1}_{\{|\hat{\beta}_{j,k}| \geq \hat{s}_j(n)\}} \right], \\ R_3 &= \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[\beta_{j,k}^2 \mathbb{1}_{\{|\hat{\beta}_{j,k}| < \hat{s}_j(n)\}} \right], \quad R_4 = \sum_{j=j_1+1}^{+\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2. \end{aligned}$$

Bound on R_4: first, recall that following our assumptions, Lemma 19.1 of [18] implies that

$$\sum_{k=0}^{2^j-1} \beta_{jk}^2 \leq C 2^{-2js^*}, \text{ with } s^* = s + 1/2 - 1/p', \tag{10.2}$$

where C is a constant depending only on p, q, s, A. Since by definition $2^{-j_1} \leq 2(\frac{\log n}{n})^{-\frac{1}{2\nu+1}}$, equation (10.2) implies that $R_4 = \mathcal{O}\left(2^{-2j_1 s^*}\right) = \mathcal{O}\left((\frac{\log n}{n})^{-\frac{2s^*}{2\nu+1}}\right)$, as $n \to +\infty$. Note that in the case $p \geq 2$, then $s^* = s$ and thus $\frac{2s}{2\nu+1} > \frac{2s}{2s+2\nu+1}$. In the case $1 \leq p < 2$, then $s^* = s + 1/2 - 1/p$, and one can check that the conditions $s > 1/p$ and $s^* p > \nu(2-p)$ imply that $\frac{2s^*}{2\nu+1} > \frac{2s}{2s+2\nu+1}$. Hence in both cases one has that

$$R_4 = \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}}\right), \text{ as } n \to +\infty. \tag{10.3}$$

Bound on R_1: using Proposition 10.1 and the inequality $2^{j_0} \leq \log n$ it follows that

$$R_1 \leq C \frac{2^{j_0(2\nu+1)}}{n} \leq C \frac{(\log n)^{2\nu+1}}{n} = \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}}\right). \tag{10.4}$$

Bound on R_2 and R_3. remark that $R_2 \leq R_{21} + R_{22}$ and $R_3 \leq R_{31} + R_{32}$ with

$$R_{21} = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\beta}_{j,k} - \beta_{j,k})^2 \mathbb{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \hat{s}_j(n)/2\}}\right], R_{22} = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\hat{\beta}_{j,k} - \beta_{j,k})^2 \mathbb{1}_{\{|\beta_{j,k}| \geq \hat{s}_j(n)/2\}}\right],$$

$$R_{31} = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[\beta_{j,k}^2 \mathbb{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \hat{s}_j(n)/2\}}\right] \text{ and } R_{32} = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[\beta_{j,k}^2 \mathbb{1}_{\{|\beta_{j,k}| < \frac{3}{2}\hat{s}_j(n)\}}\right].$$

Now, applying Cauchy-Schwarz's inequality, we get that

$$\begin{aligned} R_{21} + R_{31} &= \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[\left((\hat{\beta}_{j,k} - \beta_{j,k})^2 + \beta_{j,k}^2\right) \mathbb{1}_{\{|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \hat{s}_j(n)/2\}}\right] \\ &\leq \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\left(\mathbb{E}(\hat{\beta}_{j,k} - \beta_{j,k})^4\right)^{1/2} + \beta_{j,k}^2\right) \left(\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \hat{s}_j(n)/2)\right)^{1/2} \end{aligned}$$

Bound on $\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \hat{s}_j(n)/2)$: using that $|c_\ell(\psi_{j,k})| \leq 2^{-j/2}$ one has that $\|\tilde{\psi}_{j,k}\|_2^2 \leq \sigma_j^2$ and $\|\tilde{\psi}_{j,k}\|_\infty \leq \epsilon_j$. Thus, by definition of $\hat{s}_j(n)$ it follows that

$$2\Delta_{jk}^n(\gamma) \leq \hat{s}_j(n)/2 \tag{10.5}$$

for all sufficiently large n where $\Delta_{jk}^n(\gamma)$ is defined in (10.1). Moreover, by (5.1) there exists two constants C_1, C_2 such that for all $\ell \in \Omega_j$, $C_1 2^j \leq |\ell| \leq C_2 2^j$. Since $\lim_{|\ell| \to +\infty} \theta_\ell = 0$ uniformly for $f \in B_{p,q}^s(A)$ it follows that as $j \to +\infty$

$$V_j^2 = \|g\|_\infty 2^{-j} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|^2}{|\gamma_\ell|^2} = o\left(2^{-j} \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2}\right) = o\left(\sigma_j^2\right) \text{ and } \delta_j = 2^{-j/2} \sum_{\ell \in \Omega_j} \frac{|\theta_\ell|}{|\gamma_\ell|} = o\left(\epsilon_j\right).$$

Now, define the non-random threshold

$$s_j(n) = 4\left(\sqrt{\sigma_j^2 \frac{2\gamma \log n}{n} (\|g\|_\infty \|\lambda\|_1 + \delta)} + \frac{\gamma \log n}{3n} \epsilon_j\right), \text{ for } j_0(n) \leq j \leq j_1(n). \tag{10.6}$$

Using that $V_j^2 = o(\sigma_j^2)$ and $\delta_j = o(\epsilon_j)$ as $j \to +\infty$, and that $j_0(n) \to +\infty$ as $n \to +\infty$ it follows that for all sufficiently large n and $j_0(n) \le j \le j_1(n)$

$$2\left(\sqrt{\frac{2V_j^2\gamma\log n}{n}} + \delta_j\frac{\gamma\log n}{3n}\right) \le s_j(n)/2 \tag{10.7}$$

From equation (10.27) (see below), one has that $\mathbb{P}\left(\|\lambda\|_1 \ge \tilde{K}_n\right) \le 2n^{-\gamma}$, which implies that $s_j(n) \le \hat{s}_j(n)$ with probability larger than $1 - 2n^{-\gamma}$. Hence, by inequalities (10.5) and (10.7), it follows that for all sufficiently large n

$$2\max\left(\Delta_{jk}^n(\gamma), \sqrt{\frac{2V_j^2\gamma\log n}{n}} + \delta_j\frac{\gamma\log n}{3n}\right) \le \hat{s}_j(n)/2 \tag{10.8}$$

with probability larger than $1 - 2n^{-\gamma}$. Therefore, for all sufficiently large n, Proposition 10.2 and inequality (10.8) imply that

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| > \hat{s}_j(n)/2\right) \le Cn^{-\gamma}, \tag{10.9}$$

for all $j_0(n) \le j \le j_1(n)$.

<u>Bound on $R_{21} + R_{31}$:</u> Using the assumption that $\gamma \ge 2$, inequality (10.2) and Proposition 10.1, one has that for all sufficiently large n

$$R_{21} + R_{31} \le C\frac{1}{n}\left[\sum_{j=j_0}^{j_1} 2^j\left(\frac{2^{4j\nu}}{n^2}\left(1 + \frac{2^j}{n}\right)\right)^{1/2} + \sum_{j=j_0}^{j_1} 2^{-2js^*}\right].$$

By definition of j_1 one has that $\frac{2^j}{n} \le C$ for all $j \le j_1$, which implies that (since $s^* > 0$)

$$R_{21} + R_{31} \le C\frac{1}{n}\left[\sum_{j=j_0}^{j_1}\frac{2^{j(2\nu+1)}}{n} + \sum_{j=j_0}^{j_1} 2^{-2js^*}\right] = \mathcal{O}(n^{-\frac{2s}{2s+2\nu+1}}),\ \text{as } n \to +\infty, \tag{10.10}$$

using the fact that $\frac{2^{j(2\nu+1)}}{n} \le C$ for all $j \le j_1(n) \le \frac{1}{2\nu+1}\log_2 n$.

Finally, it remains to bound the term $T_2 = R_{22} + R_{32}$. For this purpose, let j_2 be the largest integer such that $2^{j_2} \le n^{\frac{1}{2s+2\nu+1}}(\log n)^\beta$ with $\beta = -\frac{1}{2s+2\nu+1}$, and partition T_2 as $T_2 = T_{21} + T_{22}$ where the first component T_{21} is calculated over the resolution levels $j_0 \le j \le j_2$ and the second component T_{22} is calculated over the resolution levels $j_2 + 1 \le j \le j_1$ (note that given our assumptions then $j_2 \le j_1$ for all sufficiently large n). Using the definition of the threshold $\hat{s}_j(n)$ it follows that

$$\hat{s}_j(n)^2 \le C\left(\sigma_j^2(\|g\|_\infty\tilde{K}_n + \delta)\frac{\log(n)}{n} + \frac{(\log n)^2}{n^2}\epsilon_j^2\right). \tag{10.11}$$

From Assumption 3.1 on the γ_ℓ's and equation (5.1) for Ω_j it follows that

$$\sigma_j^2 \le C2^{2j\nu} \text{ and } \epsilon_j \le C2^{j(\nu+1/2)}.$$

Since, for $2^j\frac{\log n}{n} \le \left(\frac{\log n}{n}\right)^{-\frac{2\nu}{2\nu+1}}$ all $j \le j_1$, it follows that $\frac{(\log n)^2}{n^2}\epsilon_j^2 \le C2^{2j\nu}\frac{\log(n)}{n}$ and thus

$$\hat{s}_j(n)^2 \le C2^{2j\nu}(\|g\|_\infty\tilde{K}_n + \delta + 1)\frac{\log(n)}{n}. \tag{10.12}$$

Using Proposition 10.1, the bound (10.12), the fact that

$$\mathbb{E}\tilde{K}_n \leq \|\lambda_1\|_1 + \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \tag{10.13}$$

and the definition of j_2 one obtains that

$$\begin{aligned} T_{21} \leq \sum_{j=j_0}^{j_2}\sum_{k=0}^{2^j-1}\left(\mathbb{E}(\hat{\beta}_{j,k}-\beta_{j,k})^2 + \frac{9}{4}\mathbb{E}\hat{s}_j(n)^2\right) &= \mathcal{O}\left(\frac{2^{j_2(2\nu+1)}}{n}\log(n)\right) \\ &= \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}}(\log n)^{\frac{2s}{2s+2\nu+1}}\right), \end{aligned} \tag{10.14}$$

as $n \to +\infty$. Then, it remains to obtain a bound for T_{22}. Recall that $\hat{s}_j(n) \geq s_j(n)$ with probability larger that $1-2n^{-\gamma}$, where $s_j(n)$ is defined in (10.6). Therefore, using Cauchy-Schwarz's inequality

$$\begin{aligned} &\mathbb{E}\left[(\hat{\beta}_{j,k}-\beta_{j,k})^2 1\!\!1_{\{|\beta_{j,k}|\geq \hat{s}_j(n)/2\}}\right] \\ &\leq \mathbb{E}(\hat{\beta}_{j,k}-\beta_{j,k})^2 1\!\!1_{\{|\beta_{j,k}|\geq s_j(n)/2\}} + \left(\mathbb{E}(\hat{\beta}_{j,k}-\beta_{j,k})^4\right)^{1/2}\left(\mathbb{P}(\hat{s}_j(n)\leq s_j(n))\right)^{1/2}. \end{aligned}$$

Then, by Assumption 3.1 one has that $\sigma_j^2 \geq C2^{2j\nu}$. Therefore, using Proposition 10.1 it follows that $\mathbb{E}(\hat{\beta}_{j,k}-\beta_{j,k})^2 \leq Cs_j^2(n)$ and that $\mathbb{E}(\hat{\beta}_{j,k}-\beta_{j,k})^4 \leq C\frac{2^{4j\nu}}{n^2}$ for all $j \leq j_1$. Finally, using that $\gamma \geq 2$ and the fact that $\mathbb{P}(\hat{s}_j(n) \leq s_j(n)) \leq 2n^{-\gamma}$, one finally obtains that for any $j \leq j_1$

$$\mathbb{E}\left[(\hat{\beta}_{j,k}-\beta_{j,k})^2 1\!\!1_{\{|\beta_{j,k}|\geq \hat{s}_j(n)/2\}}\right] \leq C\left(\frac{s_j^2(n)}{4}1\!\!1_{\{|\beta_{j,k}|\geq s_j(n)/2\}} + \frac{2^{2j\nu}}{n^2}\right) \tag{10.15}$$

Let us first consider the case $p \geq 2$. Using inequality (10.15) one has that

$$\begin{aligned} T_{22} &\leq C\left(\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}\frac{s_j^2(n)}{4}1\!\!1_{\{|\beta_{j,k}|\geq s_j(n)/2\}} + \frac{2^{2j\nu}}{n^2} + |\beta_{j,k}|^2\right) \\ &\leq C\left(\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^2 + \frac{1}{n}\sum_{j=j_2+1}^{j_1}\frac{2^{j(2\nu+1)}}{n}\right). \end{aligned}$$

Then (10.2), the definition of j_2, j_1 and the fact that $s^* = s$ imply that

$$T_{22} = \mathcal{O}\left(2^{-2j_2 s} + \frac{1}{n}\sum_{j=j_2+1}^{j_1}\frac{2^{j(2\nu+1)}}{n}\right) = \mathcal{O}\left(n^{-\frac{2s}{2s+2\nu+1}}(\log n)^{\frac{2s}{2s+2\nu+1}}\right) \tag{10.16}$$

Now, consider the case $1 \leq p < 2$. Using again inequality (10.15) one obtains that

$$\begin{aligned} T_{22} &\leq C\left(\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}\frac{s_j^2(n)}{4}1\!\!1_{\{|\beta_{j,k}|\geq s_j(n)/2\}} + \frac{2^{2j\nu}}{n^2} + \mathbb{E}|\beta_{j,k}|^2 1\!\!1_{\{|\beta_{j,k}|<\frac{3}{2}\hat{s}_j(n)\}}\right) \\ &\leq C\left(\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}s_j(n)^{2-p}|\beta_{j,k}|^p + |\beta_{j,k}|^p\mathbb{E}\hat{s}_j(n)^{2-p} + \frac{1}{n}\sum_{j=j_2+1}^{j_1}\frac{2^{j(2\nu+1)}}{n}\right) \end{aligned} \tag{10.17}$$

By Holder inequality, it follows that for any $\alpha > 1$, $\mathbb{E}\hat{s}_j(n)^{2-p} \leq \left(\mathbb{E}\hat{s}_j(n)^{\alpha(2-p)}\right)^{1/\alpha}$. Hence, by taking $\alpha = 2/(2-p)$, we get that $\mathbb{E}\hat{s}_j(n)^{2-p} \leq \left(\mathbb{E}\hat{s}_j(n)^2\right)^{(2-p)/2}$. Then, using the following upper bounds (as a consequence of the definition of $s_j^2(n)$ and the arguments used to derive inequalities (10.12), (10.13))

$$s_j^2(n) \le C 2^{2j\nu} \frac{\log(n)}{n} \text{ and } \mathbb{E}\hat{s}_j(n)^2 \le C 2^{2j\nu} \mathbb{E}\tilde{K}_n \frac{\log(n)}{n} \le C 2^{2j\nu} \frac{\log(n)}{n},$$

it follows that inequality (10.17) and the fact that for $\lambda \in B_{p,q}^s(A)$, $\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \le C 2^{-jps^*}$ (with $ps^* = ps + p/2 - 1$) imply that

$$\begin{aligned} T_{22} &\le C \left(\sum_{j=j_2+1}^{j_1} 2^{2j\nu(1-p/2)} \left(\frac{\log(n)}{n} \right)^{1-p/2} 2^{-jps^*} + \frac{1}{n} \sum_{j=j_2+1}^{j_1} \frac{2^{j(2\nu+1)}}{n} \right) \\ &\le C \left(\left(\frac{\log n}{n} \right)^{1-p/2} \sum_{j=j_2+1}^{j_1} 2^{j(\nu(2-p)-ps^*)} + \frac{1}{n} \sum_{j=j_2+1}^{j_1} \frac{2^{j(2\nu+1)}}{n} \right) \\ &= \mathcal{O} \left(\left(\frac{\log n}{n} \right)^{1-p/2} 2^{j_2(\nu(2-p)-ps^*)} + \frac{1}{n} \sum_{j=j_2+1}^{j_1} \frac{2^{j(2\nu+1)}}{n} \right) \\ &= \mathcal{O} \left(n^{-\frac{2s}{2s+2\nu+1}} (\log n)^{\frac{2s}{2s+2\nu+1}} \right) \end{aligned} \tag{10.18}$$

where we have used the assumption $\nu(2-p) < ps^*$ and the definition of j_2, j_1 for the last inequalities. Finally, combining the bounds (10.3), (10.4), (10.10), (10.14), (10.16) and (10.18) completes the proof of Theorem 5.1. □

10.2 Technical results

Arguing as in the proof of Proposition 3 in [5], one has the following lemma:

Lemma 10.1 *Suppose that g satisfies Assumption 3.1. Then, there exists a constants $C > 0$ such that for any $j \ge 0$ and $0 \le k \le 2^j - 1$*

$$\|\tilde{\psi}_{j,k}\|_\infty \le C 2^{j(\nu+1/2)}, \ \|\tilde{\psi}_{j,k}\|_2^2 \le C 2^{2j\nu} \text{ and } \|\tilde{\psi}_{j,k}^2\|_2^2 \le C 2^{j(4\nu+1)}.$$

Proposition 10.1 *There exists $C > 0$ such that for any $j \ge 0$ and $0 \le k \le 2^j - 1$*

$$\mathbb{E}|\hat{c}_{j,k} - c_{j,k}|^2 \le C \frac{2^{2j\nu}}{n} \left(1 + \|\lambda\|_2 \|g\|_\infty\right), \quad \mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^2 \le C \frac{2^{2j\nu}}{n} \left(1 + \|\lambda\|_2 \|g\|_\infty\right), \tag{10.19}$$

and

$$\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^4 \le C \frac{2^{4j\nu}}{n^2} \left(1 + \frac{2^j}{n}\right) \left(1 + \|\lambda\|_2^2 \|g\|_\infty^2 + \|\lambda\|_2 \|g\|_\infty + \|\lambda\|_2^2 \|g\|_\infty\right). \tag{10.20}$$

Proof : We only prove the proposition for the wavelet coefficients $\hat{\beta}_{j,k}$ since the arguments are the same for the scaling coefficients $\hat{c}_{j,k}$. Remark first that $\hat{\beta}_{j,k} - \beta_{j,k} = \sum_{\ell \in \Omega_j} c_\ell(\psi_{j,k})(\hat{\theta}_\ell - \theta_\ell) = Z_1 + Z_2$, where Z_1 and Z_2 are the centered variables

$$Z_1 := \sum_{\ell \in \Omega_j} (\tilde{\gamma}_\ell \gamma_\ell^{-1} - 1) \theta_\ell c_\ell(\phi_{j,k}).$$

and

$$Z_2 := \frac{1}{n} \sum_{i=1}^n \int_0^1 \tilde{\psi}_{j,k}(t) d\tilde{N}_t^i.$$

where $d\tilde{N}_t^i = dN_t^i - \lambda(t - \boldsymbol{\tau}_i)dt$.

Control of the moments of Z_1: by arguing as in the proof of Proposition 3 in [5], one obtains that there exists a universal constant $C > 0$ such that

$$\mathbb{E}|Z_1|^2 \leq C\frac{2^{2j\nu}}{n} \text{ and } \mathbb{E}|Z_1|^4 \leq C\left(\frac{2^{4j\nu}}{n^2} + \frac{2^{j(4\nu+1)}}{n^3}\right). \tag{10.21}$$

The main arguments to obtain (10.21) rely on concentration inequalities on the variables $\boldsymbol{\tau}_i, i = 1, \ldots, n$.

Control of the moments of Z_2: using Lemma 10.1 remark that

$$\begin{aligned}\mathbb{E}|Z_2|^2 &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\int_0^1 \tilde{\psi}_{j,k}^2(t)\lambda(t-\boldsymbol{\tau}_i)dt = \frac{1}{n}\int_0^1 \tilde{\psi}_{j,k}^2(t)\lambda \star g(t)dt,\\ &\leq C\frac{2^{2j\nu}}{n}\|\lambda \star g\|_\infty \leq C\frac{2^{2j\nu}}{n}\|\lambda\|_2\|g\|_\infty.\end{aligned}$$

Let us now bound $\mathbb{E}|Z_2|^4$ by using Rosenthal's inequality [29]

$$\mathbb{E}\left|\sum_{i=1}^{n} Y_i\right|^{2p} \leq \left(\frac{16p}{\log(2p)}\right)^{2p}\max\left\{\left(\sum_{i=1}^{n}\mathbb{E}Y_i^2\right)^p;\sum_{i=1}^{n}\mathbb{E}|Y_i|^{2p}\right\},$$

which is valid for independent, centered and real-valued random variables $(Y_i)_{i=1\ldots,n}$. We apply this inequality to $Y_i = \int_0^1 \tilde{\psi}_{j,k}(t)d\tilde{N}_t^i$ with $p = 2$. Conditionnaly to $\boldsymbol{\tau}_i$, using Proposition 6 in [27] and the Jensen's inequality, it follows that

$$\begin{aligned}\mathbb{E}\left[Y_i^4|\boldsymbol{\tau}_i\right] &= \int_0^1 \tilde{\psi}_{j,k}^4(t)\lambda(t-\boldsymbol{\tau}_i)dt + 3\left(\int_0^1 \tilde{\psi}_{j,k}^2(t)\lambda(t-\boldsymbol{\tau}_i)dt\right)^2,\\ &\leq \int_0^1 \tilde{\psi}_{j,k}^4(t)\left(\lambda(t-\boldsymbol{\tau}_i) + 3\lambda^2(t-\boldsymbol{\tau}_i)\right)dt.\end{aligned}$$

Hence $\mathbb{E}\sum_{i=1}^n Y_i^4 \leq n\int_0^1 \tilde{\psi}_{j,k}^4(t)\left(\lambda \star g(t) + 3\lambda^2 \star g(t)\right)dt$. Then, using Lemma 10.1 $\mathbb{E}\sum_{i=1}^n Y_i^4 \leq Cn2^{j(4\nu+1)}\left(\|\lambda\|_2 + \|\lambda\|_2^2\right)\|g\|_\infty$. Using again Proposition 6 in [27] and Lemma 10.1 one obtains that $\mathbb{E}Y_i^2 = \int_0^1 \tilde{\psi}_{j,k}^2(t)\lambda \star g(t)dt \leq C2^{2j\nu}\|\lambda\|_2\|g\|_\infty$ which ends the proof of the proposition. □

Proposition 10.2 *Assume that* $\lambda \in \Lambda_\infty$ *and let* $\gamma > 0$. *Then, there exists a constant* $C > 0$ *such that for any* $j \geq 0$, $k \in \{0 \ldots 2^j - 1\}$ *and all sufficiently large* n

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| > 2\max\left(\Delta_{jk}^n(\gamma)), \sqrt{\frac{2V_j^2\gamma\log n}{n}} + \delta_j\frac{\gamma\log n}{3n}\right)\right) \leq Cn^{-\gamma}, \tag{10.22}$$

where $\Delta_{jk}^n(\gamma)$ *is defined in (10.1).*

<u>*Proof :*</u>

Using the notations introduced in the proof of Proposition 10.1, write $\hat{\beta}_{j,k} - \beta_{j,k} = Z_1 + Z_2$ and remark that for any $u > 0$

$$\mathbb{P}(|Z_1 + Z_2| > u) \leq \mathbb{P}(|Z_1| > u/2) + \mathbb{P}(|Z_2| > u/2) \tag{10.23}$$

Now, arguing as in Proposition 4 in [5] and using Bernstein's inequality, one has immediately that

$$\mathbb{P}\left(|Z_1| > \sqrt{\frac{2V_j^2\gamma\log n}{n}} + \delta_j\frac{\gamma\log n}{3n}\right) \leq 2n^{-\gamma}. \tag{10.24}$$

Let us now control the deviation of $Z_2 = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \tilde{\psi}_{j,k}(t)d\tilde{N}_t^i$. First, remark that conditionnaly to the shifts $\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n$, the process $\sum_{i=1}^{n} N^i$ is a Poisson process with intensity $\sum_{i=1}^{n}\lambda(. - \boldsymbol{\tau}_i)$. For the sake of convenience, we introduce some additionnal notations. For $n \geq 1$, $j \geq 0$ and $0 \leq k \leq 2^j - 1$, define

$$M_{jk}^n = \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \tilde{\psi}_{jk}^2(t)\lambda(t-\boldsymbol{\tau}_i)dt, \text{ and } M_{jk} = \mathbb{E}M_{jk}^n = \int_0^1 \tilde{\psi}_{jk}^2(t)\lambda \star g(t)dt.$$

Using an analogue of Bennett's inequality for Poisson processes (see e.g. Proposition 7 in [27]), we get that for any $s > 0$

$$\mathbb{P}\left(|Z_2| > \sqrt{\frac{2s}{n}M_{jk}^n} + \frac{s}{3n}\|\tilde{\psi}_{j,k}\|_\infty \Big| \boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n\right) \leq 2\exp(-s) \tag{10.25}$$

Remark that the quantity M_{jk}^n is not computable from the data as its depends on λ and the unobserved shifts $\boldsymbol{\tau}_1,\ldots,\boldsymbol{\tau}_n$. Nevertheless it is possible to compute a data-based upper bound for M_{jk}^n. Indeed, note that Bernstein's inequality (see e.g. Proposition 2.9 in [22]) implies that

$$\mathbb{P}\left(M_{jk}^n > M_{jk} + \tilde{M}_{jk}\left(\frac{\gamma\log n}{3n} + \sqrt{\frac{2\gamma\log n}{n}}\right)\right) \leq n^{-\gamma}.$$

with $\tilde{M}_{jk} = \|\lambda\|_\infty\|\tilde{\psi}_{j,k}\|_2^2$. Obviously, $\tilde{M}_{jk}$ is unknown but for all sufficiently large n, one has that

$$\tilde{M}_{jk} = \|\lambda\|_\infty\|\tilde{\psi}_{j,k}\|_2^2 \leq \log n\|\tilde{\psi}_{j,k}\|_2^2.$$

Moreover, remark that $M_{jk} = \|\psi_{jk}\sqrt{\lambda \star g}\|_2^2 \leq \|\psi_{jk}\|_2^2\|g\|_\infty\|\lambda\|_1$. Hence,

$$\mathbb{P}\left(M_{jk}^n > \|\tilde{\psi}_{j,k}\|_2^2\left(\|g\|_\infty\|\lambda\|_1 + \left(\frac{\gamma(\log n)^2}{3n} + \sqrt{\frac{2\gamma(\log n)^3}{n}}\right)\right)\right) \leq n^{-\gamma}. \tag{10.26}$$

In order to obtain a data-based upper bound for M_{jk}^n, it remains to derive an upper bound for $\|\lambda\|_1$. Recall that we have denoted by K_i the number of points of the process N^i. Conditionally to $\boldsymbol{\tau}_i$, K_i is real random variable that follows a Poisson distribution with intensity $\int_0^1 \lambda(t-\boldsymbol{\tau}_i)dt$. Since λ is assumed to be periodic with period 1, it follows that for any $i = 1,\ldots,n$, $\int_0^1 \lambda(t-\boldsymbol{\tau}_i)dt = \int_0^1 \lambda(t)dt$, and thus $(K_i)_{i=1,\ldots,n}$ are i.i.d. random variables following a Poisson distribution with intensity $\|\lambda\|_1 = \int_0^1 \lambda(t)dt$. Using standard arguments to derive concentration inequalities one has that for any $u > 0$

$$\mathbb{P}\left(\|\lambda\|_1 \geq \frac{1}{n}\sum_{i=1}^{n}K_i + \sqrt{\frac{2u\|\lambda\|_1}{n}} + \frac{u}{3n}\right) \leq 2\exp(-u).$$

Now, define the function $h(y) = y^2 - \sqrt{2a}y - a/3$ for $y \geq 0$ and with $a = u/n$. Then, the above inequality can be written as

$$\mathbb{P}\left(h\left(\sqrt{\|\lambda\|_1}\right) \geq \frac{1}{n}\sum_{i=1}^{n}K_i\right) \leq 2\exp(-u).$$

Since h restricted on $[\sqrt{a}(\sqrt{30}+3\sqrt{2})/6; +\infty[$ is invertible with $h^{-1}(y) = \sqrt{y + \frac{5a}{6}} + \sqrt{\frac{a}{2}}$ it follows that for $u = \gamma\log n$ and all sufficiently large n

$$\mathbb{P}\left(\|\lambda\|_1 \geq \bar{K}_n + \frac{4\gamma\log n}{3n} + \sqrt{\frac{2\gamma\log n}{n}\bar{K}_n + \frac{5\gamma^2(\log n)^2}{3n^2}}\right) \leq 2n^{-\gamma}, \tag{10.27}$$

where $\bar{K}_n = \frac{1}{n}\sum_{i=1}^{n} K_i$. Therefore, using (10.26) it follows that

$$\mathbb{P}\left(M_{jk}^n > \|\tilde{\psi}_{j,k}\|_2^2\left(\|g\|_\infty \tilde{K}_n(\gamma) + \left(\frac{\gamma(\log n)^2}{3n} + \sqrt{\frac{2\gamma(\log n)^3}{n}}\right)\right)\right) \leq 3n^{-\gamma}, \tag{10.28}$$

where $\tilde{K}_n(\gamma)$ is defined in (5.6). Hence, combining (10.25) with $s = \gamma \log n$ and (10.28) we obtain that

$$\mathbb{P}\left(|Z_2| > \sqrt{\frac{2\gamma \log n}{n}\|\tilde{\psi}_{j,k}\|_2^2\left(\|g\|_\infty \tilde{K}_n(\gamma) + \left(\frac{\gamma(\log n)^2}{3n} + \sqrt{\frac{2\gamma(\log n)^3}{n}}\right)\right)} + \frac{\gamma \log n}{3n}\|\tilde{\psi}_{j,k}\|_\infty\right) \leq 5n^{-\gamma} \tag{10.29}$$

Combining inequalities (10.23), (10.24) and (10.29) concludes the proof. □

References

[1] Schwartzman A., Jaffe A., Gavrilov Y., and Meyer C. Multiple testing of local maxima for detection of peaks in chip-seq data. *Submitted to Annals of Applied Statistics*, pages 1–22, 2012.

[2] Schwartzman A., Gavrilov Y., and Adler R. Multiple testing of local maxima for detection of peaks in 1d. *Annals of Statistics*, 39:3290–3319, 2011.

[3] A. Antoniadis and J. Bigot. Poisson inverse problems. *Annals of Statistics*, 34(5):2132–2158, 2006.

[4] J. Bigot and B. Charlier. On the consistency of frechet means in deformable models for curve and image analysis. *Electronic Journal of Statistics*, 5:1054–1089, 2011.

[5] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Annals of Statistics*, 38(4):2422–2464, 2010.

[6] P. Brémaud. *Point processes and Queues, Martingale Dynamics*. Springer series in Statistics, 1981.

[7] J. Bretagnolle and C. Huber. Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete*, 47(2):119–137, 1979.

[8] L. Cavalier, G. Golubev, D. Picard, and A. Tsybakov. Oracle inequalities for inverse problems. *Annals of Statistics*, 30(3):843–874, 2002. Dedicated to the memory of Lucien Le Cam.

[9] L. Cavalier and J.-Y. Koo. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Transactions on Information Theory*, 48(10):2794–2802, 2002.

[10] Laurent Cavalier and Nicolas Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.

[11] Johnson D., Mortazavi A., Myers R., and Wold B. Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316:1497–1502, 2007.

[12] D. L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Different perspectives on wavelets (San Antonio, TX, 1993)*, volume 47 of *Proc. Sympos. Appl. Math.*, pages 173–205. Amer. Math. Soc., Providence, RI, 1993.

[13] F.-X. Dupé, J. M. Fadili, and J.-L. Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, 18(2):310–321, 2009.

[14] Richard D. Gill and Boris Y. Levit. Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995.

[15] W. Hardle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation and. Statistical Applications.*, volume 129. Lecture Notes in Statistics, New York: Spriner-Verlag, 1998.

[16] Marc Hoffmann and Markus Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *Annals of Statistics*, 36(1):310–336, 2008.

[17] J. Jacod. *Calcul stochastique et problèmes de martingales.* Springer Verlag, 1979.

[18] I. M. Johnstone. Function estimation in gaussian noise: Sequence models. *Unpublished Monograph, http://www-stat.stanford.edu/~imj/*, 2002.

[19] I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society (B)*, 66:547–573, 2004.

[20] J. F. C. Kingman. *Poisson Processes*, volume 3 of *Oxford Studies in Probability.* Oxford University Press, Oxford, 1993.

[21] E. D. Kolaczyk. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statistica Sinica*, 9(1):119–135, 1999.

[22] P. Massart. *Concentration Inequalities and Model Selection: Ecole d'ÕtÕ de ProbabilitÕs de Saint-Flour XXXIII - 2003.* Lecture Notes in Mathematics Springer, 2006.

[23] C. Matias. Semiparametric deconvolution with unknown noise variance. *ESAIM Probability and Statistics*, 6:271–292, 2002.

[24] Y. Meyer. *Ondelettes et opérateurs, I.* Hermann, 1989.

[25] R. D. Nowak and E. D. Kolaczyk. A statistical multiscale framework for Poisson inverse problems. *IEEE Transactions on Information Theory*, 46(5):1811–1825, 2000. Information-theoretic imaging.

[26] M. Pensky and T. Sapatinas. Functional deconvolution in a periodic setting: uniform case. *Annals of Statistics*, 37(1):73–104, 2009.

[27] P. Reynaud-Bourret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126:103–153, 2003.

[28] P. Reynaud-Bourret and V. Rivoirard. Near optimal thresholding estimation of a poisson intensity on the real line. *Electronic Journal of Statistics*, 4:171–238, 2010.

[29] H. P. Rosenthal. On the span in L^p of sequences of independent random variables. II. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 149–167, Berkeley, Calif., 1972. Univ. California Press.

[30] L. Sansonnet. Wavelet thresholding estimation in a poissonian interactions model with application to genomic data. *available at http://arxiv.org/abs/1107.4219*, 2011.

[31] Mikkelsen T., Manching K., Jaffe D., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T., Koche R., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E., and Bernstein B. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448:553–560, 2007.

[32] J. Wang, A. Huda, V. Lunyak, and I. Jordan. A gibbs sampling strategy applied to the mapping of ambiguous short-sequence tags. *Bioinformatics*, 26:2501–2508, 2010.

[33] R. M. Willett and R. D. Nowak. Multiscale Poisson intensity and density estimation. *IEEE Transactions on Information Theory*, 53(9):3171–3187, 2007.

[34] Zhang Y., Liu T., Meyer C., Eeckhoute J., Johnson D., Bernstein B., Nusbaum C., Myers R., Brown M., and Li W. Model-based analysis of chip-seq (macs). *Genome Biology*, 9(9):R137, 2008.

[35] B. Zhang, J. M. Fadili, and J.-L. Starck. Wavelets, ridgelets, and curvelets for Poisson noise removal. *IEEE Transactions on Image Processing*, 17(7):1093–1108, 2008.