



HAL
open science

Uniform Deconvolution for Poisson Point Processes

Anna Bonnet, Claire Lacour, Franck Picard, Vincent Rivoirard

► **To cite this version:**

Anna Bonnet, Claire Lacour, Franck Picard, Vincent Rivoirard. Uniform Deconvolution for Poisson Point Processes. 2020. hal-02964117v1

HAL Id: hal-02964117

<https://hal.science/hal-02964117v1>

Preprint submitted on 12 Oct 2020 (v1), last revised 1 Jul 2022 (v2)

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

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

Uniform Deconvolution for Poisson Point Processes

October 12, 2020

ANNA BONNET

Sorbonne Université, UMR CNRS 8001, LPSM, 75005 Paris, France
Univ. Lyon, Université Lyon 1, CNRS, LBBE UMR5558, F-69622 Villeurbanne, France

CLAIRE LACOUR

LAMA, Univ Gustave Eiffel, Univ Paris Est Créteil, CNRS, F-77454 Marne-la-Vallée, France

FRANCK PICARD

Univ. Lyon, Université Lyon 1, CNRS, LBBE UMR5558, F-69622 Villeurbanne, France

VINCENT RIVOIRARD

CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University, 75016 Paris,
France

Abstract

We focus on the estimation of the intensity of a Poisson process in the presence of a uniform noise. We propose a kernel-based procedure fully calibrated in theory and practice. We show that our adaptive estimator is optimal from the oracle and minimax points of view, and provide new lower bounds when the intensity belongs to a Sobolev ball. By developing the Goldenshluger-Lepski methodology in the case of deconvolution for Poisson processes, we propose an optimal data-driven selection of the kernel's bandwidth, and we provide a heuristic framework to calibrate the estimator in practice. Our method is illustrated on the spatial repartition of replication origins along the human genome.

Keywords: Convolution, Poisson process, Adaptive estimation

1 Introduction

Inverse problems for Poisson point processes have focused much attention in the statistical literature over the last years, mainly because the estimation of a Poisson process intensity in the presence of additive noise is encountered in many practical situations like tomography, microscopy, high energy physics. Our work is motivated by an original application field in high throughput biology, that has been revolutionized by the development of high throughput sequencing. The applications of such technologies are many, and we focus on the particular cases where sequencing allows the fine mapping of genomic features along the genome, like transcription factors of copy-number variations. The spatial repartition of these features can be modeled by a Poisson process with unknown intensity. Unfortunately, detections are prone to some errors, which produces data in the form of genomic intervals whose width is linked

to the precision of detection. Since the exact position of the peak is unknown within the interval, the appropriate error distribution is uniform, the level of noise being given by the width of the intervals.

In the 2000s, several wavelet methods have been proposed for Poisson intensity estimation from indirect data (Antoniadis and Bigot, 2006), as well as B-splines and empirical Bayes estimation (Kuusela et al., 2015). Other authors turned to variational regularization: see the survey of Hohage and Werner (2016), which also contains examples of applications and reconstruction algorithms. From a more theoretical perspective, Kroll (2019) studied the estimation of the intensity function of a Poisson process from noisy observations in a circular model. His estimator is based on Fourier series and is not appropriate for uniform noise (whose Fourier coefficients are zero except the first).

The specificity of uniform noise has rather been studied in the context of density deconvolution. In this case also, classical methods based on the Fourier transform do not work either in the case of a noise with vanishing characteristic function (Meister, 2009). Nevertheless several corrected Fourier approaches were introduced (Hall et al., 2001, 2007; Meister, 2008; Feuerverger et al., 2008). In this line, the work of Delaigle and Meister (2011) is particularly interesting, even if it is limited to a density to estimate with finite left endpoint. In a recent work, Belomestny and Goldenshluger (2019) have shown that the Laplace transform can perform deconvolution for general measurement errors. Another approach consists in using Tikhonov regularization for the convolution operator (Carrasco and Florens, 2011; Trong et al., 2014). In the specific case of uniform noise (also called boxcar deconvolution), it is possible to use *ad hoc* kernel methods (Groeneboom and Jongbloed, 2003; van Es, 2011). In this context of non-parametric estimation, each method depends on a regularization parameter (such as a resolution level, a regularization parameter or a bandwidth), and only a good choice of this parameter allows to achieve an optimal reconstruction. This parameter selection is often named adaption since the point is to adapt the parameter to the features of the target density. The above cited works (except Delaigle and Meister (2011)) do not address this adaption issue or only from a practical point of view, although this is central both from the practical and theoretical points of views.

We propose a kernel estimator to estimate the intensity of a Poisson process in the presence of a uniform noise. We provide theoretical guarantees of its performance by deriving the minimax rates of convergence of the integrated squared risk for an intensity belonging to a Sobolev ball. To ensure the optimality of our procedure, we establish new lower bounds on this smoothness space. Then we provide an adaptive procedure for bandwidth selection using the Goldenshluger-Lepski methodology, and we show its optimality in the oracle and minimax frameworks. From the practical point of view we propose a framework to tune the hyperparameters of the method, which is notoriously difficult in the context of Goldenshluger-Lepski methods. We propose two strategies, the first one aiming at tuning a universal hyperparameter to use the selection procedure in any configuration. Then we combine this procedure with an additional step based on the data. The empirical performance of our estimator is then studied by simulations. Finally we provide an illustration of our procedure on experimental data in Genomics, where the purpose is to study the spatial repartition of replication starting points along chromosomes in humans (Picard et al., 2014). The code is available at <https://github.com/AnnaBonnet/PoissonDeconvolution>.

2 Uniform deconvolution model

We consider $(X_i)_i$, the realization of a Poisson Process on \mathbb{R} , denoted by N^X , with $N_+ := N^X(\mathbb{R})$ the number of occurrences. The uniform convolution model consists in observing $(Y_i)_i$, occurrences of a Poisson process N^Y , a noisy version of N^X corrupted by a uniform noise, such that:

$$\forall i \in \{1, \dots, N_+\}, \quad Y_i = X_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{U}[-a; a], \quad (2.1)$$

where a , assumed to be known, is fixed. The errors $(\varepsilon_i)_i$ are supposed mutually independent, and independent of $(X_i)_i$. Then, we denote by λ_X the intensity function of N^X , and μ_X its mean measure assumed to satisfy $\mu_X(\mathbb{R}) < \infty$, so that

$$d\mu_X(x) = \lambda_X(x)dx, \quad x \in \mathbb{R}.$$

Note that $N_+ \sim \mathcal{P}(\mu_X(\mathbb{R}))$, where $\mathcal{P}(\theta)$ denotes the Poisson distribution with parameter θ . Then we further consider that observing N^X with intensity λ_X is equivalent to observing n i.i.d. Poisson processes with common intensity f_X , with $\lambda_X = n \times f_X$. This specification will be convenient to adopt an asymptotic perspective. As for λ_Y , the intensity of N^Y , it can easily be shown that

$$n^{-1} \times \lambda_Y = n^{-1} \times (\lambda_X \star f_\varepsilon) = f_X \star f_\varepsilon =: f_Y,$$

where f_ε stands for the density of the uniform distribution. The goal of the deconvolution method is to estimate f_X , based on the observation of N^Y on a compact interval $[0, T]$ for some fixed positive real number T . In the following, we provide an optimal estimator of f_X in the oracle and minimax settings. Minimax rates of convergence will be studied in the asymptotic perspective $n \rightarrow +\infty$ and parameters a , T and $\|f_X\|_1$ will be viewed as constant.

Notations. We denote by $\|\cdot\|_{2,T}$, $\|\cdot\|_{1,T}$ and $\|\cdot\|_{\infty,T}$ the \mathbb{L}_2 , \mathbb{L}_1 and sup-norm on $[0; T]$, and $\|\cdot\|_2$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ their analog on \mathbb{R} . Notation \lesssim means that the inequality is satisfied up to a constant and $a_n = o(b_n)$ means that the ratio a_n/b_n goes to 0 when n goes to $+\infty$. Finally, $\llbracket a; b \rrbracket$ denotes the set of integers larger or equal to a and smaller or equal to b .

3 Estimation procedure

3.1 Deconvolution with kernel estimator

To estimate f_X based on observations of N^Y , we introduce a kernel estimator which is based on the following heuristic arguments inspired from van Es (2011) who considered the setting of uniform deconvolution for density estimation. We observe that f_Y can be expressed by using the cumulative distribution of the X_i 's:

$$F_X(x) := \int_{-\infty}^x f_X(u)du \leq \|f_X\|_1, \quad x \in \mathbb{R}.$$

Indeed, for $x \in \mathbb{R}$,

$$\begin{aligned}
f_Y(x) &= \int_{\mathbb{R}} f_X(x-u)f_\varepsilon(u)du \\
&= \frac{1}{2a} \int_{-a}^a f_X(x-u)du \\
&= \frac{1}{2a} \left(F_X(x+a) - F_X(x-a) \right),
\end{aligned} \tag{3.1}$$

from which we deduce:

$$F_X(x) = 2a \sum_{k=0}^{+\infty} f_Y\left(x - (2k+1)a\right), \quad x \in \mathbb{R}.$$

Then, from heuristic arguments, we get

$$f_X(x) = 2a \sum_{k=0}^{+\infty} f'_Y\left(x - (2k+1)a\right), \tag{3.2}$$

which provides a natural form of our kernel estimate \widehat{f}_Y . Note that differentiability of f_Y is not assumed in the following. Indeed, we consider the kernel estimator of f_Y such that

$$\widehat{f}_Y(x) = \frac{1}{nh} \int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) dN_u^Y,$$

with dN^Y the point measure associated to N^Y , and K a kernel with bandwidth $h > 0$. Setting

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right),$$

we can write

$$\widehat{f}_Y(x) = \frac{1}{n} \sum_{i=1}^{N_+} K_h(x - Y_i).$$

Then, if K is differentiable, we propose the following kernel-based estimator of f_X

$$\widehat{f}_h(x) = \frac{2a}{nh^2} \sum_{k=0}^{+\infty} \sum_{i=1}^{N_+} K'\left(\frac{x - (2k+1)a - Y_i}{h}\right).$$

The proof of subsequent Lemma 1 in Appendix shows that the expectation of \widehat{f}_h is a regularization of f_X , as typically desired for kernel estimates since we have

$$\mathbb{E}[\widehat{f}_h] = K_h \star f_X. \tag{3.3}$$

Then, our objective is to provide an optimal selection procedure for the parameter h .

3.2 Symmetrization of the estimator

Our estimator is based on the inversion and differentiation of Equation (3.1), which can also be performed as follows:

$$2a \sum_{k=0}^{+\infty} f_Y(x + (2k+1)a) = \sum_{k=0}^{+\infty} \left(F_X(x + 2(k+1)a) - F_X(x + 2ka) \right) = \|f_X\|_1 - F_X(x)$$

and differentiated to obtain:

$$f_X(x) = -2a \sum_{k=0}^{+\infty} f'_Y \left(x + (2k+1)a \right),$$

which leads to another estimator

$$\check{f}_h(x) = -\frac{2a}{nh^2} \sum_{k=0}^{+\infty} \sum_{i=1}^{N_+} K' \left(\frac{x + (2k+1)a - Y_i}{h} \right).$$

In the framework of uniform deconvolution for densities, van Es (2011) proposes to use $\alpha \hat{f}_h(x) + (1 - \alpha) \check{f}_h(x)$, a convex combination of \hat{f}_h and \check{f}_h , as a combined estimator, to benefit from the small variance of $\check{f}_h(x)$ and $\hat{f}_h(x)$ for large and small values of x respectively. Unfortunately the combination that minimizes the asymptotic variance of the combined estimator is achieved for $\alpha = 1 - F_X(x)$, and thus depends on an unknown quantity. van Es (2011) suggested to use a plug-in estimator, but to avoid supplementary technicalities, we finally consider the following symmetric kernel-based estimator:

$$\tilde{f}_h(x) := \frac{1}{2} \left(\hat{f}_h(x) + \check{f}_h(x) \right) = \frac{a}{nh^2} \sum_{k=-\infty}^{+\infty} s_k \sum_{i=1}^{N_+} K' \left(\frac{x - (2k+1)a - Y_i}{h} \right), \quad (3.4)$$

with $s_k = 1$ if $k \geq 0$ and $s_k = -1$ if $k < 0$. Then it is shown in Lemma 1 that

$$\mathbb{E}[\tilde{f}_h] = (K_h \star f_X).$$

3.3 Risk of the kernel-based estimator

Our objective is to provide a selection procedure to select a bandwidth \hat{h} , that only depends on the data, so that the \mathbb{L}_2 -risk of $\tilde{f}_{\hat{h}}$ is smaller than the risk of the best kernel estimate (up to a constant), namely

$$\mathbb{E} \left[\|\tilde{f}_{\hat{h}} - f_X\|_{2,T}^2 \right] \lesssim \inf_{h \in \mathcal{H}} \mathbb{E} \left[\|\tilde{f}_h - f_X\|_{2,T}^2 \right].$$

Our procedure is based on the bias-variance trade-off of the risk of any estimate \tilde{f}_h

$$\begin{aligned} \mathbb{E} \left[\|\tilde{f}_h - f_X\|_{2,T}^2 \right] &= \|\mathbb{E}[\tilde{f}_h] - f_X\|_{2,T}^2 + \mathbb{E} \left[\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T}^2 \right] \\ &=: B_h^2 + v_h. \end{aligned} \quad (3.5)$$

Then we use the following mild assumption:

Assumption 1. *The kernel K is supported on the compact interval $[-A, A]$, with $A \in \mathbb{R}_+^*$ and K is differentiable on $[-A, A]$.*

Then the variance of the estimator is such that:

Lemma 1. *For any $h \in \mathcal{H}$ and any $x \in [0, T]$, we have*

$$\mathbb{E}[\tilde{f}_h(x)] = (K_h \star f_X)(x).$$

Under Assumption 1 and if h is small enough so that $Ah \leq a$, then

$$v_h := \mathbb{E}\left[\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T}^2\right] = \frac{aT\|f_X\|_1\|K'\|_2^2}{2nh^3}.$$

The expectation of \tilde{f}_h has the expected expression (derived from (3.3)), but Lemma 1 also provides the exact expression of the variance term v_h of our very specific estimate. Since our framework is an inverse problem, this variance does not reach the classical $(nh)^{-1}$ bound. Moreover, v_h depends linearly on $\|f_X\|_1$ but also on T , which means that the estimation of f_X has to be performed on the compact interval $[0, T]$, for v_h to be finite. This requirement is due to Expression (3.4) of our estimate that shows that for any $x \in \mathbb{R}$, $\tilde{f}_h(x)$ is different from 0 almost surely. This dependence of v_h on T is a direct consequence of our strategy not to make any assumption on the support of f_X , that can be unknown or non-compact. Of course, if the support of f_X was known to be compact, like $[0, 1]$, then we would force \tilde{f}_h to be null outside $[0, 1]$ (for instance by removing large values of $|k|$ in the sum of (3.4)), and estimation would be performed on the set $[0, 1]$. Actually, estimating a non-compactly supported Poisson intensity on the whole real line leads to deterioration of classical non-parametric rates in general (Reynaud-Bouret and Rivoirard, 2010).

3.4 Bandwidth selection

The objective of our procedure is to choose the bandwidth h , based on the Goldenshluger-Lepski methodology (Goldenshluger and Lepski, 2013). First, we introduce a finite set \mathcal{H} of bandwidths such that for any $h \in \mathcal{H}$, $h \leq a/A$, which is in line with assumptions of Lemma 1. Then, for two bandwidths t and h , we also define

$$\tilde{f}_{h,t} := K_h \star \tilde{f}_t,$$

a twice regularized estimator, that satisfies the following property (see Lemma 2 in Appendix):

$$\tilde{f}_{h,t} = \tilde{f}_{t,h}.$$

Now we select the bandwidth as follows:

$$\hat{h} := \operatorname{argmin}_{h \in \mathcal{H}} \left\{ \mathcal{A}(h) + \frac{c\sqrt{N_+}}{nh^{3/2}} \right\}, \quad (3.6)$$

where

$$c = (1 + \eta)(1 + \|K\|_1)\|K'\|_2\sqrt{\frac{aT}{2}} \quad (3.7)$$

for some $\eta > 0$ and

$$\mathcal{A}(h) := \max_{t \in \mathcal{H}} \left\{ \|\tilde{f}_{h,t} - \tilde{f}_t\|_{2,T} - \frac{c\sqrt{N_+}}{nt^{3/2}} \right\}_+.$$

Finally, we estimate f_X with

$$\tilde{f} = \tilde{f}_{\hat{h}}. \quad (3.8)$$

Note that $\mathcal{A}(h)$ is an estimation of the bias term B_h of the estimator \tilde{f}_h . Indeed,

$$B_h := \|\mathbb{E}[\tilde{f}_h] - f_X\|_{2,T} = \|K_h \star f_X - f_X\|_{2,T}$$

and we replace the unknown function f_X with the kernel estimate \tilde{f}_t . The term $\frac{c\sqrt{N_+}}{nt^{3/2}}$ in $\mathcal{A}(h)$ controls the fluctuations of $\|\tilde{f}_{h,t} - \tilde{f}_t\|_{2,T}$. Finally, since $\mathbb{E}[N_+] = n\|f_X\|_1$, (3.6) mimics the bias-variance trade-off (3.5) (up to the squares). In order to fully define the estimation procedure, it remains to choose the set of bandwidths \mathcal{H} . This is specified in Section 4.

4 Theoretical results

4.1 Oracle approach

The oracle setting allows us to prove that the bandwidth selection procedure described in Section 3.4 is (nearly) optimal among all kernel estimates. Indeed, we obtain the following result.

Theorem 1. *Suppose that Assumption 1 is verified. We consider the estimate \tilde{f} such that the finite set of bandwidths \mathcal{H} satisfies*

$$\min \mathcal{H} = \frac{1}{\delta n} \quad \text{and} \quad \max \mathcal{H} = o(1)$$

for some constant $\delta > 0$. Then, for n large enough,

$$\mathbb{E} \left[\|\tilde{f}_{\hat{h}} - f_X\|_{2,T}^2 \right] \leq C_1 \inf_{h \in \mathcal{H}} \mathbb{E} \left[\|\tilde{f}_h - f_X\|_{2,T}^2 \right] + \frac{C_2}{n}, \quad (4.1)$$

where C_1 is a constant only depending on K and η and C_2 is a constant depending on $\|f_X\|_1$, T , a , δ , η and K .

Remark 1. *For the condition $\min \mathcal{H} = \frac{1}{\delta n}$, we can replace δn by any polynomial in n .*

Theorem 1 shows that our procedure achieves nice performance: Up to the constant C_1 and the negligible term C_2/n that goes to 0 quickly, our estimate has the smallest risk among all kernel rules under mild conditions on the set of bandwidths \mathcal{H} .

4.2 Minimax approach

The minimax approach is a framework that shows the optimality of an estimate among all possible estimates. For this purpose, we consider a class of functional spaces for f_X , then we derive the minimax risk associated with each functional space and show that our estimate achieves this rate. Here, we consider the class of Sobolev balls that can be defined, for instance, through the Fourier transform of \mathbb{L}_2 -functions: Given $\beta > 0$ and $L > 0$, the Sobolev ball of smoothness β and radius L is

$$\mathcal{S}^\beta(L) := \left\{ g \in \mathbb{L}_2 : \int_{-\infty}^{+\infty} |g^*(\xi)|^2 (\xi^2 + 1)^\beta d\xi \leq L^2 \right\},$$

where $g^*(\xi) := \int e^{ix\xi}g(x)dx$ is the Fourier transform of g . From a statistical perspective, the minimax rate associated with the space $\mathcal{S}^\beta(L)$ is

$$\mathcal{R}(\beta, L) := \inf_{Z_n} \sup_{f_X \in \mathcal{S}^\beta(L)} \mathbb{E} [\|Z_n - f_X\|_{2,T}^2],$$

where the infimum is taken over all estimators Z_n of f_X based on the observations $(Y_i)_{i=1, \dots, N_+}$. We first derive a lower bound for the minimax risk.

Theorem 2. *We assume that L is larger than an absolute constant. There exists a positive constant C_3 only depending on β , a and T such that, if n is large enough,*

$$\mathcal{R}(\beta, L) \geq C_3 L^{\frac{6}{2\beta+3}} n^{-\frac{2\beta}{2\beta+3}}. \quad (4.2)$$

The rate established in (4.2) is slower than the classical non-parametric rate $n^{-\frac{2\beta}{2\beta+1}}$. It is the expected rate since our deconvolution problem corresponds to an inverse problem of order 1, meaning that f_ε^* , the characteristic function of the noise, satisfies

$$|f_\varepsilon^*(\xi)| = O(\xi^{-1}) \text{ as } \xi \rightarrow \infty.$$

Note that the analog of the previous lower bound has been established in the density deconvolution context, first by Fan (1993), but with supplementary assumption $|f_\varepsilon^{*'}(\xi)| = O(\xi^{-2})$ which is not satisfied in our case of uniform noise (see Equation (7.6)). Our proof is rather inspired by the work of Meister (2009), but we face here a Poisson inverse problem, and we have to control the \mathbb{L}_2 -norm on $[0, T]$ instead of \mathbb{R} . Furthermore, Theorem 2.14 of Meister (2009) only holds for $\beta > 1/2$. Consequently, we use different techniques to establish Theorem 2, which are based on wavelet decompositions of the signal. Specifically, we use the Meyer wavelets of order 2.

We now show that the rate achieved by our estimate \tilde{f} corresponds to the lower bound (4.2), up to a constant. We have the following corollary, easily derived from Theorem 1, and based on the following assumption.

Assumption 2. *The kernel K is of order $\ell = \lfloor \beta \rfloor$, meaning that the functions $x \mapsto x^j K(x)$, $j = 0, 1, \dots, \ell$ are integrable and satisfy*

$$\int K(x)dx = 1, \quad \int x^j K(x)dx = 0, \quad j = 1, \dots, \ell.$$

Remark 2. *See Proposition 1.3 of Tsybakov (2008) or Section 3.2 of Goldenshluger and Lepski (2014) for the construction of kernels satisfying Assumptions 1 and 2.*

Corollary 1. *Suppose that Assumptions 1 and 2 are satisfied. We consider the estimate \tilde{f} such that the set of bandwidths \mathcal{H} is*

$$\mathcal{H} = \left\{ D^{-1} : D \in \llbracket \log n; \delta n^{\frac{1}{3}} \rrbracket \right\},$$

for some constant $\delta > 0$. Then, for n large enough,

$$\sup_{f_X \in \mathcal{S}^\beta(L)} \mathbb{E} \left[\|\tilde{f} - f_X\|_{2,T}^2 \right] \leq C_4 L^{\frac{6}{2\beta+3}} n^{-\frac{2\beta}{2\beta+3}},$$

where C_4 only depends on K , β , $\|f_X\|_1$, a and T .

To prove Corollary 1, we combine the upper bound (4.1) and the decomposition (3.5) to obtain for any f_X and any $h \in \mathcal{H}$,

$$\mathbb{E} \left[\|\tilde{f} - f_X\|_{2,T}^2 \right] \leq C_1 \mathbb{E} \left[\|\tilde{f}_h - f_X\|_{2,T}^2 \right] + \frac{C_2}{n} = C_1(B_h^2 + v_h) + \frac{C_2}{n}.$$

Under Assumption 2, we have for any $f_X \in \mathcal{S}^\beta(L)$

$$B_h = \|K_h * f_X - f_X\|_{2,T} \leq MLh^\beta,$$

for M a positive constant depending on K and β . Indeed, the space $\mathcal{S}^\beta(L)$ is included into the Nikol'ski ball $H(\beta, L')$ with L' equal to L up to a constant. We refer the reader to Proposition 1.5 of Tsybakov (2008) and Kerkycharian et al. (2001) for more details. Now, we plug $h \in \mathcal{H}$ of order $(L^{-2}n^{-1})^{\frac{1}{2\beta+3}}$ in the previous upper bound to obtain the desired bound of Corollary 1 thanks to Lemma 1.

Corollary 1 shows that our estimator is adaptive minimax, i.e. it achieves the best possible rate (up to a constant) and the bandwidth selection does not depend on the spaces parameters (β, L) on the whole range $\{0 < \beta < \ell + 1, L > 0\}$, where ℓ is the order of the chosen kernel. We have established the optimality of our procedure.

5 Simulation study and numerical tuning

In the following we use numerical simulations to tune the hyperparameters our estimator and to assess the performance of our deconvolution procedure. We consider different shapes for the Poisson process intensity to challenge our estimator in different scenarii, by first generating Poisson processes on $[0, 1]$ based on the Beta probability distribution function, with $f_{\text{unisym}} = \text{Beta}(2, 2)$ (unimodal symmetric), $f_{\text{bisym}} = 0.5 \times \text{Beta}(2, 6) + 0.5 \times \text{Beta}(6, 2)$ (bimodal symmetric), $f_{\text{biasym}} = 0.5 \times \text{Beta}(2, 20) + 0.5 \times \text{Beta}(2, 2)$ (bimodal assymmetric). We also generate Poisson processes with Laplace distribution intensity (location 5, scale 0.5) to consider a sharp form and a different support. In this case, we consider that $T = 10$. We consider a uniform convolution model with increasing noise ($a \in \{0.05, 0.1\}$ for Beta, $a \in \{0.5, 1, 2, 3\}$ for Laplace) and Poisson processes with increasing number of occurrences ($n \in \{500, 1000\}$). For each set (f_X, n, a) , we present the median performance over 30 replicates. In order to keep a bounded variance of our estimators, we explore different values of h using a grid denoted by \mathcal{H} , with minimum value $(aT/n)^{1/3}$ (see Lemma 1 and Corollary 1). Finally, our procedure is computed with an Epanechnikov kernel, that is $K(u) = 0.75(1 - u^2)\mathbb{1}_{|u| \leq 1}$. Our estimator is challenged by the oracle estimator, that is the estimator \hat{f}_{h^*} , with h^* minimizing the mean squared error $\mathbb{E}\|f_X - \hat{f}_h\|_{2,T}^2$, with f_X the true intensity.

Then our selection procedure for the smoothing parameter h is based on the two-step method described in Section 3. Using this procedure in practice requires to tune the value of the hyper-parameter η that is part of the penalty $c(\eta)$:

$$c(\eta) = (1 + \eta)(1 + \|K\|_1) \sqrt{\frac{aT}{2}} \|K'\|_2,$$

with K the Epanechnikov kernel in our simulations. This penalty is at the core of the two-step method that consists in computing:

$$A_\eta(h) = \max_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \tilde{f}_{h,t}\|_{2,T} - c(\eta) \frac{\sqrt{N_+}}{nt^{3/2}} \right\}_+, \quad (5.1)$$

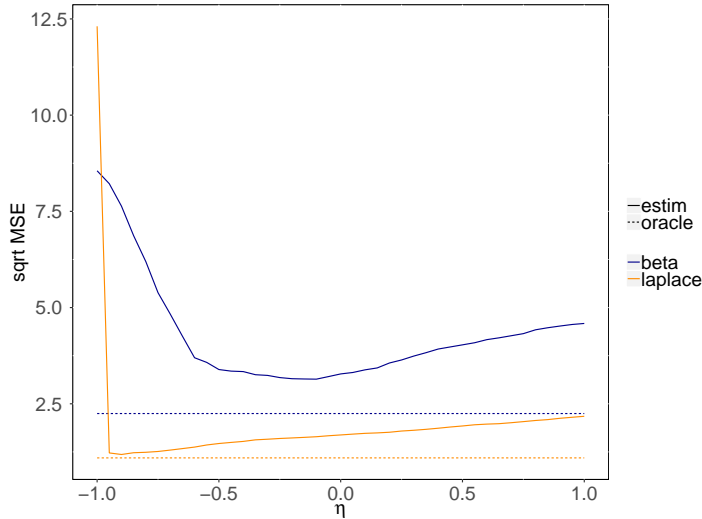


Figure 1: Mean squared error (square root) obtained on 30 simulations with our estimator for different values of η and with an oracle estimator (dotted line). The mean squared errors are computed as mean values on 30 simulations and over all different Beta scenarii (blue line) and Laplace scenarii (orange line)

followed by

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \left\{ A_\eta(h) + c(\eta) \frac{\sqrt{N_+}}{nh^{3/2}} \right\}. \quad (5.2)$$

5.1 Universal hyperparameter tuning

We first propose to investigate if we could find a "universal" value of parameter η that would be appropriate whatever the form of the intensity function. For a grid of η in $[-1; 1]$, we compare the mean squared errors (MSE) of estimators calibrated with different values of η to the MSE achieved by the oracle estimator that achieves the smallest MSE over the grid \mathcal{H} . Figure 1 shows that the optimal choice of η depends on the shape of the true intensity (~ -0.5 for Beta, ~ -0.9 for Laplace). However, choosing $\eta = -0.6$ appears to be a reasonable trade-off to obtain good performance in most settings. Reconstructions show (Figure 2) that the quality of the results strongly depends on the shape of the true intensity, which is over-smoothed, in particular for the bimodal symmetric Beta intensity with high noise ($a = 0.1$) and for the Laplace intensity.

5.2 Going further universal tuning

In order to improve the calibration of our procedure, we propose a tuning method that depends on the data. Our idea is to balance the estimation of the bias term ($A_\eta(h)$) with the penalty constant $c(\eta)$ by estimating a value for the hyperparameter η based on the data. Consequently, instead of calibrating η directly, we propose to restate the procedure such that:

$$\hat{h}_\gamma = \operatorname{argmin}_{h \in \mathcal{H}} \left(\min_{\eta} \left\{ A_\eta(h) + \gamma \times \frac{c(\eta)\sqrt{N_+}}{nh^{3/2}} \right\} \right). \quad (5.3)$$

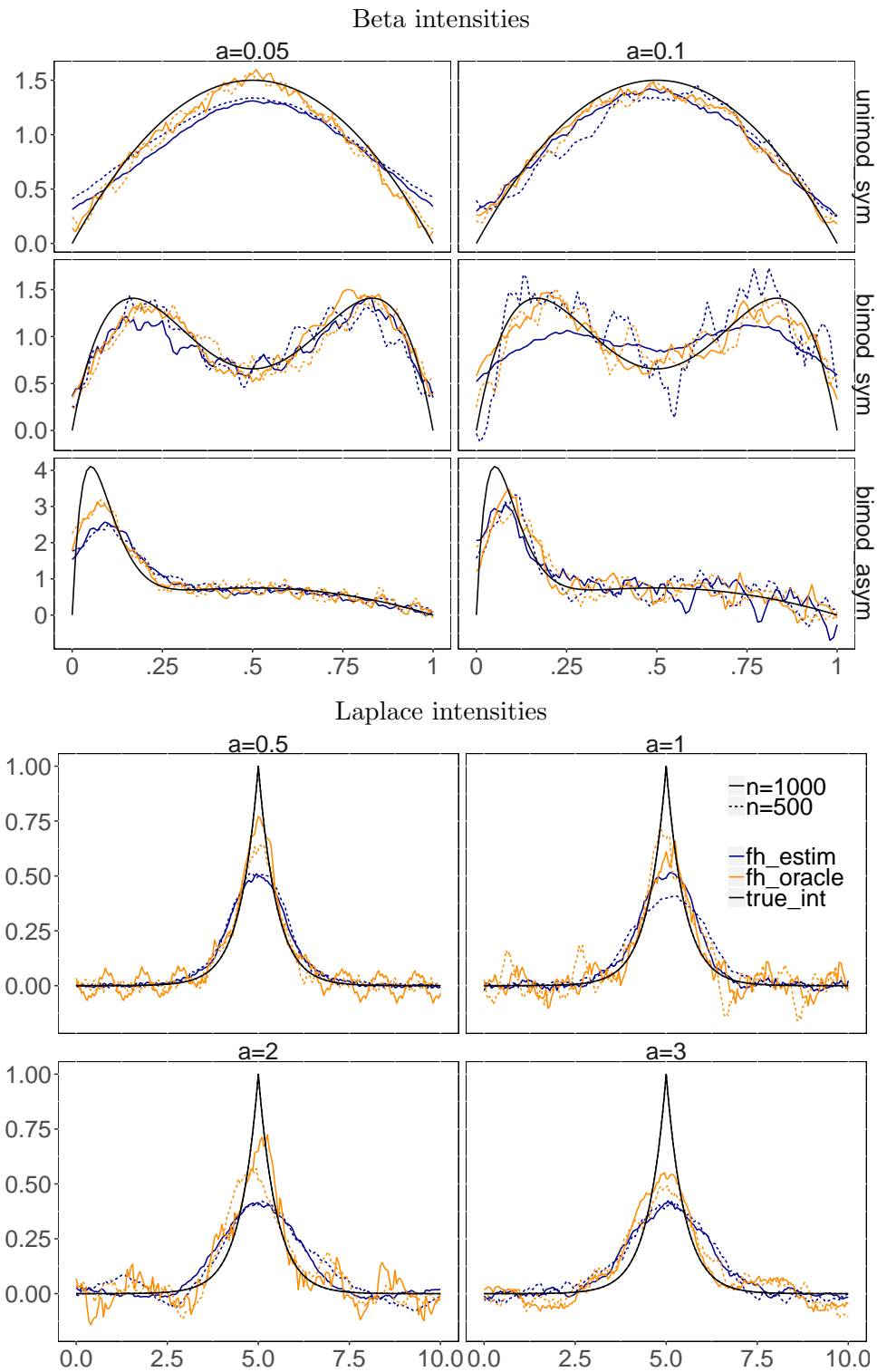


Figure 2: Estimated intensity with the tuning method described in Section 5.1 with $\eta = -0.6$, compared with the oracle estimator. The reconstruction is obtained with a generated dataset that verifies the median MSE over 30 simulations.

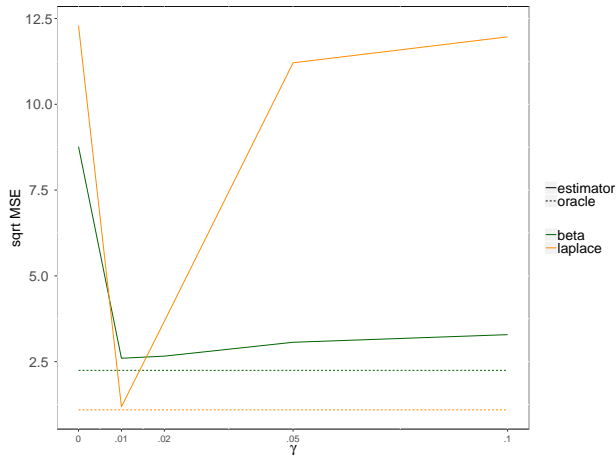


Figure 3: Mean squared error (square root) obtained with our estimator for different values of γ and with an oracle estimator, for different values of n . The mean squared are computed as mean values on 30 simulations and over all different Beta scenarii (green line) and Laplace scenarii (orange line).

Our proposition is to introduce a trade-off parameter γ that balances the two steps of the procedure, and to get rid of hyperparameter η by considering the minimum of the penalized bias $A_\eta(h)$ over a grid of η values ($\eta \in]-1, 1[$). This min-step corresponds to the objective of minimizing an empirical risk. Then we consider different values for γ and compare the estimation risk of \hat{f}_{h_γ} with the risk of the oracle. Figure 3 shows that there exists a value $\gamma = 0.01$ for which the risk of our estimator is close to the best risk, whatever the form of the true intensity, the noise (a values), and the number of observations (n values). Hence, in the following, we set $\gamma = 0.01$. Let us notice that this value of γ is very small: however, we see from Figure 3 that choosing $\gamma = 0$ substantially increases the mean squared error, which confirms the necessity of the penalization in (5.2). The estimated intensities with this second tuning method are displayed in Figure 4: we observe that the estimations are closer to the oracle estimator than with the first tuning approach, which validates the adaptive tuning procedure. Figure 5 confirms that for all scenarii, the adaptive tuning method provides the best results.

6 Deconvolution of Genomic data

Next generation sequencing technologies (NGS) has allowed the fine mapping of eukaryotes replication origins that constitute the starting points of chromosomes duplication. To maintain the stability and integrity of genomes, replication origins are under a very strict spatio-temporal control, and part of their positioning has been shown to be associated with cell differentiation (Picard et al., 2014). The spatial organization has become central to better understand genomes architecture and regulation. However, the positioning of replication origins is subject to errors, since any NGS-based high-throughput mapping consists of peak-calling based on the detection of an exceptional enrichment of short reads Picard et al. (2014). Consequently, the true position of the replication starting point is unknown, but rather in-

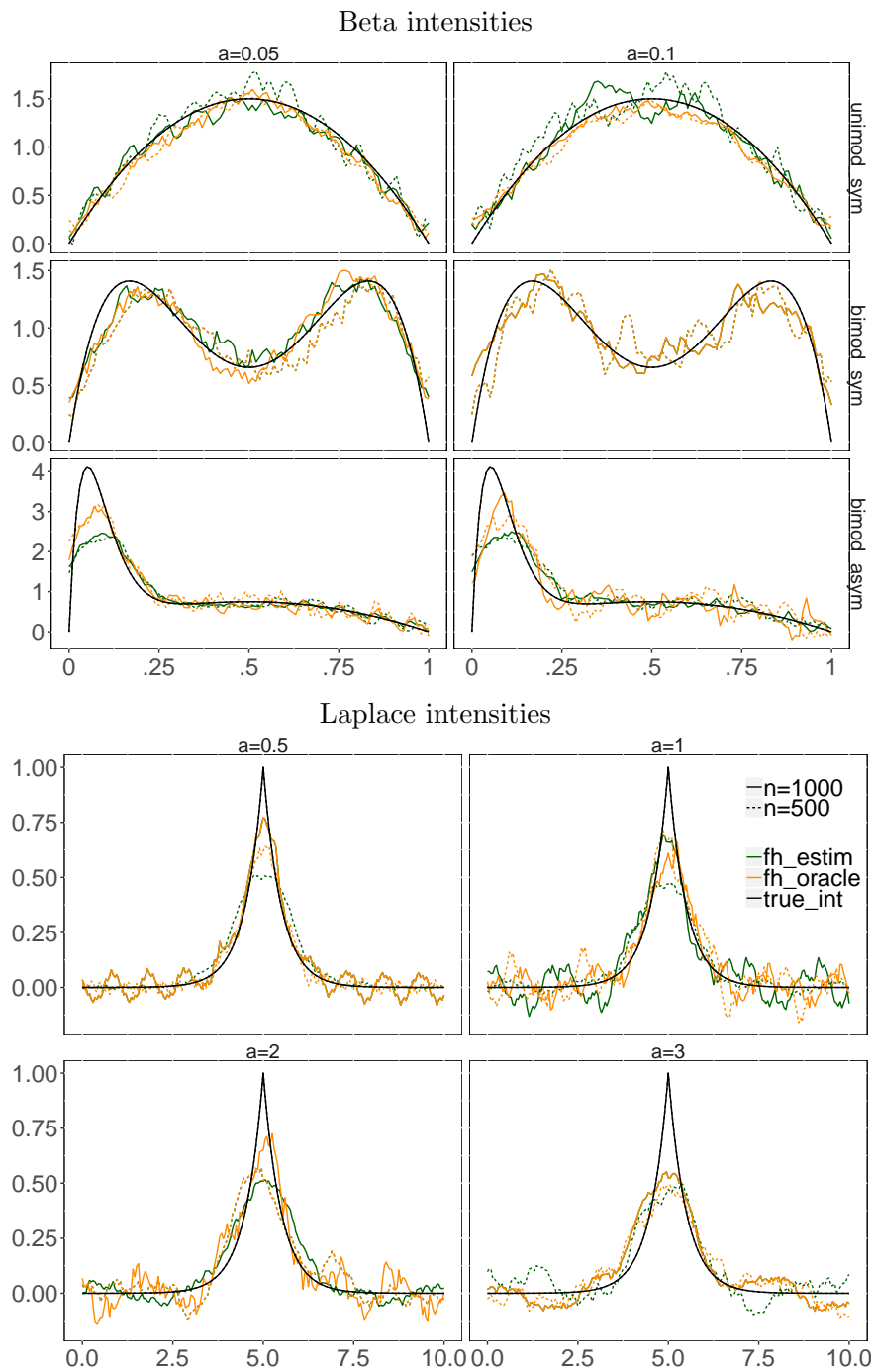


Figure 4: Estimated intensity with the adaptive tuning method described in Section 5.2, compared with the oracle estimator. The reconstruction is obtained with a generated dataset that verifies the median MSE over 30 simulations

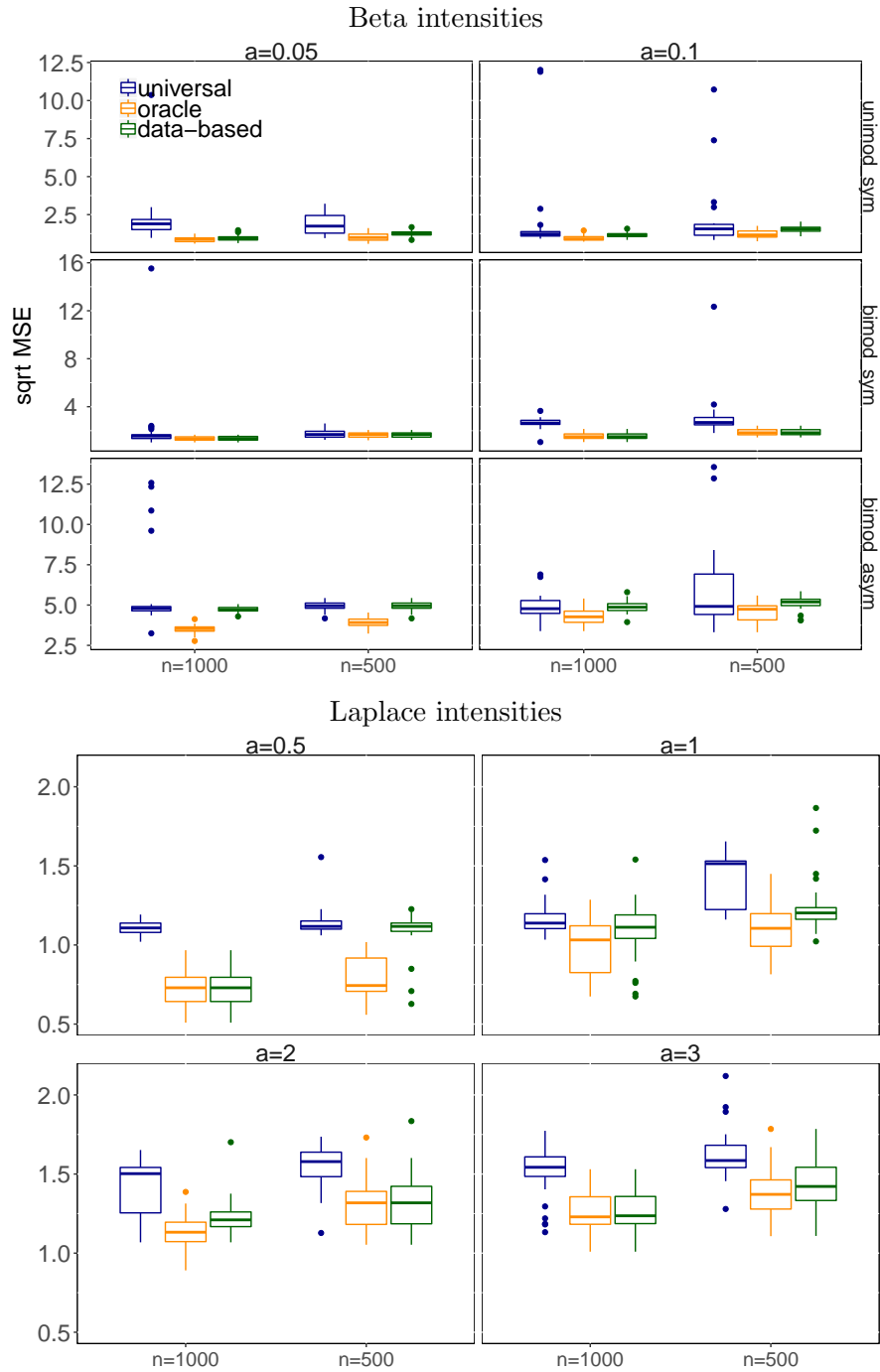


Figure 5: Mean squared errors (square root) obtained with both tuning methods ("fixed" for the first tuning method with $\eta = -0.6$ and "adaptive" for the second tuning method with $\gamma = 0.01$) and with an oracle estimator over 30 simulations.

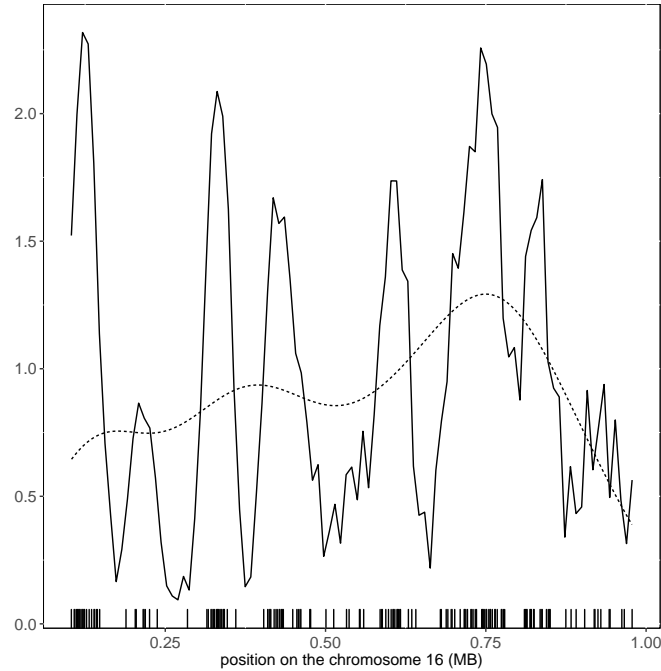


Figure 6: Estimation of the intensity of human replication origins along chromosome 16, with our procedure (plain line) that accounts for the uniform noise, and without correction (dashed line).

ferred from genomic intervals. The spatial control of replication being very strict, the precise quantification of the density of origins along chromosomes is central, but should account for this imprecision of the mapping step. If this density is estimated without deconvolution, Figure 6 shows that the naive estimator does not provide much information regarding the shape of the origins density along the genome, whereas our calibrated estimator (based on the second hybrid tuning procedure) perfectly illustrates the inhomogeneous nature of the replication starting points along the first human chromosome. In this illustration, parameter a has been estimated as the average length of the genomic intervals.

7 Proofs

If θ is a vector of constants (for instance $\theta = (T, a, K)$), we denote by \square_θ a positive constant that only depends on θ and that may change from line to line.

In the sequel, we use at several places the following property: Setting

$$S_k : x \mapsto \sum_{i=1}^{N_+} K'_h \left(x - (2k+1)a - Y_i \right),$$

since $Ah \leq a$, S_k and $S_{k'}$ have disjoint supports if $k \neq k'$.

7.1 Proof of Lemma 1

Proof. Considering first \widehat{f}_h , we have:

$$\begin{aligned} \mathbb{E}[\widehat{f}_h(x)] &= \frac{2a}{nh^2} \sum_{k=0}^{+\infty} \int_{\mathbb{R}} K' \left(\frac{x - (2k+1)a - u}{h} \right) n f_Y(u) du \\ &= \frac{1}{h^2} \sum_{k=0}^{+\infty} \int_{\mathbb{R}} K' \left(\frac{x - (2k+1)a - u}{h} \right) [F_X(u+a) - F_X(u-a)] du \\ &= \frac{1}{h^2} \sum_{k=0}^{+\infty} \int_{\mathbb{R}} K' \left(\frac{x - 2ka - v}{h} \right) F_X(v) dv - \int_{\mathbb{R}} K' \left(\frac{x - 2(k+1)a - v}{h} \right) F_X(v) dv \\ &= \frac{1}{h^2} \int_{\mathbb{R}} K' \left(\frac{x - v}{h} \right) F_X(v) dv \\ &= \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{x - v}{h} \right) f_X(v) dv = (K_h \star f_X)(x). \end{aligned}$$

The first point is then straightforward by using the definition of \widetilde{f}_h . For the second point, observe that

$$\begin{aligned} \widetilde{f}_h(t) - \mathbb{E}[\widetilde{f}_h(t)] &= \frac{a}{nh^2} \sum_{k=-\infty}^{+\infty} s_k \int K' \left(\frac{t - (2k+1)a - u}{h} \right) [dN_u^Y - n f_Y(u) du] \\ &= \frac{a}{nh} \int_{\mathbb{R}} L_h(t-u) [dN_u^Y - n f_Y(u) du], \end{aligned}$$

with

$$L_h(x) := \frac{1}{h} \sum_{k=-\infty}^{+\infty} s_k K' \left(\frac{x - (2k+1)a}{h} \right).$$

Using the support $[-A, A]$ of K , for each x

$$\begin{aligned} (L_h(x))^2 &= \frac{1}{h^2} \left(\sum_{k=-\infty}^{+\infty} s_k K' \left(\frac{x - (2k+1)a}{h} \right) \right)^2 \\ &= \frac{1}{h^2} \sum_{k=-\infty}^{+\infty} K'^2 \left(\frac{x - (2k+1)a}{h} \right) \end{aligned}$$

as soon as $Ah \leq a$. We have:

$$\begin{aligned}
\int_0^T \mathbb{E}[(\tilde{f}_h(t) - \mathbb{E}[\tilde{f}_h(t)])^2] dt &= \frac{a^2}{n^2 h^2} \int_0^T \text{Var} \left(\int_{\mathbb{R}} L_h(t-u) dN_u^Y \right) dt \\
&= \frac{a^2}{n^2 h^2} \int_0^T \int_{\mathbb{R}} L_h^2(t-u) n f_Y(u) du dt \\
&= \frac{a^2}{n h^4} \int_0^T \int_{\mathbb{R}} \sum_{k=-\infty}^{+\infty} \left(K' \left(\frac{t-u-(2k+1)a}{h} \right) \right)^2 f_Y(u) du dt \\
&= \frac{a^2}{n h^4} \int_0^T \int_{\mathbb{R}} \left(K' \left(\frac{t-v}{h} \right) \right)^2 \sum_{k=-\infty}^{+\infty} f_Y(v-(2k+1)a) dv dt \\
&= \frac{a}{2n h^4} \int_0^T \int_{\mathbb{R}} \left(K' \left(\frac{t-v}{h} \right) \right)^2 (\lim_{+\infty} F_X - \lim_{-\infty} F_X) dv dt,
\end{aligned}$$

which yields

$$\mathbb{E} \left[\|\tilde{f}_h(t) - \mathbb{E}[\tilde{f}_h(t)]\|_{2,T}^2 \right] = \frac{aT \|f_X\|_1 \|K'\|_2^2}{2nh^3}.$$

□

7.2 Auxiliary lemma

Our procedure needs the following result.

Lemma 2. For any $h, t \in \mathcal{H}$,

$$\tilde{f}_{h,t} = \tilde{f}_{t,h}.$$

Proof. Since $K'_h(x) = (1/h^2)K'(x/h)$, we can write

$$\begin{aligned}
K_h \star \hat{f}_t &= K_h \star \left(\frac{2a}{n} \sum_{k=0}^{+\infty} \sum_{i=1}^{N_+} (K_t)'(x - (2k+1)a - Y_i) \right) \\
&= \frac{2a}{n} \sum_{k=0}^{+\infty} \sum_{i=1}^{N_+} K_h \star (K_t)'(x - (2k+1)a - Y_i)
\end{aligned}$$

Using that $K_h \star (K_t)' = (K_h \star K_t)' = (K_h)' \star K_t$, we obtain $K_h \star \hat{f}_t = K_t \star \hat{f}_h$.

In the same way, we can prove $K_h \star \check{f}_t = K_t \star \check{f}_h$ and then $K_h \star \tilde{f}_t = K_t \star \tilde{f}_h$. □

7.3 Proof of Theorem 1

Remember that

$$\hat{h} := \operatorname{argmin}_{h \in \mathcal{H}} \left\{ \mathcal{A}(h) + \frac{c\sqrt{N_+}}{nh^{3/2}} \right\},$$

with

$$\mathcal{A}(h) := \max_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \tilde{f}_{h,t}\|_{2,T} - \frac{c\sqrt{N_+}}{nt^{3/2}} \right\}_+.$$

For any $h \in \mathcal{H}$,

$$\|\tilde{f}_{\hat{h}} - f_X\|_{2,T} \leq A_1 + A_2 + A_3,$$

with

$$A_1 := \|\tilde{f}_{\hat{h}} - \tilde{f}_{\hat{h},h}\|_{2,T} \leq \mathcal{A}(h) + \frac{c\sqrt{N_+}}{n\hat{h}^{3/2}},$$

$$A_2 := \|\tilde{f}_h - \tilde{f}_{\hat{h},h}\|_{2,T} \leq \mathcal{A}(\hat{h}) + \frac{c\sqrt{N_+}}{nh^{3/2}},$$

and

$$A_3 := \|\tilde{f}_h - f_X\|_{2,T}.$$

By definition of \hat{h} , we have:

$$A_1 + A_2 \leq 2\mathcal{A}(h) + \frac{2c\sqrt{N_+}}{nh^{3/2}}.$$

Therefore, by setting

$$\zeta_n(h) := \sup_{t \in \mathcal{H}} \left\{ \|(\tilde{f}_{t,h} - \mathbb{E}[\tilde{f}_{t,h}]) - (\tilde{f}_t - \mathbb{E}[\tilde{f}_t])\|_{2,T} - \frac{c\sqrt{N_+}}{nt^{3/2}} \right\}_+,$$

we have:

$$\begin{aligned} A_1 + A_2 &\leq 2\zeta_n(h) + 2 \sup_{t \in \mathcal{H}} \|\mathbb{E}[\tilde{f}_{t,h}] - \mathbb{E}[\tilde{f}_t]\|_{2,T} + \frac{2c\sqrt{N_+}}{nh^{3/2}} \\ &\leq 2\zeta_n(h) + 2 \sup_{t \in \mathcal{H}} \|K_h \star K_t \star f_X - K_t \star f_X\|_{2,T} + \frac{2c\sqrt{N_+}}{nh^{3/2}} \\ &\leq 2\zeta_n(h) + 2\|K\|_1 \|K_h \star f_X - f_X\|_{2,T} + \frac{2c\sqrt{N_+}}{nh^{3/2}}. \end{aligned}$$

Finally, since $(\alpha + \beta + \gamma)^2 \leq 3\alpha^2 + 3\beta^2 + 3\gamma^2$,

$$\begin{aligned} \mathbb{E}[(A_1 + A_2)^2] &\leq 12\mathbb{E}[\zeta_n^2(h)] + 12\|K\|_1^2 \|K_h \star f_X - f_X\|_{2,T}^2 + \frac{12c^2\mathbb{E}[N_+]}{n^2h^3} \\ &\leq 12\mathbb{E}[\zeta_n^2(h)] + 12\|K\|_1^2 \|K_h \star f_X - f_X\|_{2,T}^2 + \frac{12c^2\|f_X\|_1}{nh^3}. \end{aligned}$$

For the last term, we obtain:

$$\begin{aligned} \mathbb{E}[A_3^2] &= \mathbb{E}[\|\tilde{f}_h - f_X\|_{2,T}^2] \\ &= \mathbb{E}[\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T}^2] + \|K_h \star f_X - f_X\|_{2,T}^2 \\ &= \frac{aT\|f_X\|_1\|K'\|_2^2}{2nh^3} + \|K_h \star f_X - f_X\|_{2,T}^2. \end{aligned}$$

Finally, replacing c with its definition, namely

$$c = (1 + \eta)(1 + \|K\|_1)\|K'\|_2\sqrt{\frac{aT}{2}},$$

we obtain: for any $h \in \mathcal{H}$,

$$\begin{aligned}
\mathbb{E}[\|\tilde{f}_h - f_X\|_{2,T}^2] &\leq 2\mathbb{E}[(A_1 + A_2)^2] + 2\mathbb{E}[A_3^2] \\
&\leq 2(1 + 12\|K\|_1^2)\|K_h \star f_X - f_X\|_{2,T}^2 + \square_{K,\eta} \frac{aT\|f_X\|_1}{nh^3} + 24\mathbb{E}[\zeta_n^2(h)] \\
&\leq \square_{K,\eta}\mathbb{E}[\|\tilde{f}_h - f_X\|_{2,T}^2] + 24\mathbb{E}[\zeta_n^2(h)], \tag{7.1}
\end{aligned}$$

by using Lemma 1. It remains to prove that $\mathbb{E}[\zeta_n^2(h)]$ is bounded by $\frac{1}{n}$ up to a constant. We have:

$$\begin{aligned}
\zeta_n(h) &\leq \sup_{t \in \mathcal{H}} \left\{ \|\tilde{f}_{t,h} - \mathbb{E}[\tilde{f}_{t,h}]\|_{2,T} + \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{c\sqrt{N_+}}{nt^{3/2}} \right\}_+ \\
&\leq \sup_{t \in \mathcal{H}} \left\{ (\|K\|_1 + 1)\|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{c\sqrt{N_+}}{nt^{3/2}} \right\}_+ \\
&\leq (\|K\|_1 + 1)S_n,
\end{aligned}$$

with

$$S_n := \sup_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{(1 + \eta)\|K'\|_2\sqrt{aTN_+}}{\sqrt{2}nt^{3/2}} \right\}_+.$$

For $\alpha \in (0, 1)$ chosen later, we compute:

$$A_n := \mathbb{E}[S_n^2 \mathbf{1}_{\{N_+ \leq (1-\alpha)^2 n \|f_X\|_1\}}].$$

Recall that (see the proof of Lemma 1)

$$\tilde{f}_t(x) = \frac{a}{nt} \int_{\mathbb{R}} L_t(x-u) dN_u^Y, \quad \text{with } L_t(x) = \frac{1}{t} \sum_{k=-\infty}^{\infty} s_k K' \left(\frac{x - (2k+1)a}{t} \right).$$

Since $At \leq a$,

$$\begin{aligned}
\left(\int_{\mathbb{R}} \sum_{k=-\infty}^{+\infty} s_k K' \left(\frac{x - (2k+1)a - u}{t} \right) dN_u^Y \right)^2 &= \left(\sum_{i=1}^{N_+} \sum_{k=-\infty}^{\infty} s_k K' \left(\frac{x - (2k+1)a - Y_i}{t} \right) \right)^2 \\
&\leq N_+ \sum_{i=1}^{N_+} \sum_{k=-\infty}^{\infty} \left(K' \left(\frac{x - (2k+1)a - Y_i}{h} \right) \right)^2 \\
&\leq N_+^2 \|K'\|_{\infty}^2,
\end{aligned}$$

which yields

$$\begin{aligned}
S_n^2 &\leq 2 \sup_{t \in \mathcal{H}} \|\tilde{f}_t\|_{2,T}^2 + 2 \sup_{t \in \mathcal{H}} \|\mathbb{E}[\tilde{f}_t]\|_{2,T}^2 \\
&\leq \sup_{t \in \mathcal{H}} \frac{2a^2}{n^2 t^4} \int_0^T \left(\int_{\mathbb{R}} \sum_{k=-\infty}^{+\infty} s_k K' \left(\frac{x - (2k+1)a - u}{t} \right) dN_u^Y \right)^2 dx + 2 \sup_{t \in \mathcal{H}} \|K_t \star f_X\|_{2,T}^2 \\
&\leq \sup_{t \in \mathcal{H}} \frac{2a^2 \|K'\|_{\infty}^2 T N_+^2}{n^2 t^4} + 2 \sup_{t \in \mathcal{H}} \frac{\|K\|_2^2 \|f_X\|_1^2}{t}.
\end{aligned}$$

Therefore, since $t \in \mathcal{H} \Rightarrow t^{-1} \leq \delta n$,

$$A_n \leq \left(2\delta^4 a^2 \|K'\|_\infty^2 T n^4 + 2\delta n \|K\|_2^2 \right) \|f_X\|_1^2 \times \mathbb{P}(N_+ \leq (1 - \alpha)^2 n \|f_X\|_1).$$

To bound the last term, we use, for instance, Inequality (5.2) of Reynaud-Bouret (2003) (with $\xi = (2\alpha - \alpha^2)n\|f\|_1$ and with the function $f \equiv -1$), which shows that there exists $\alpha' > 0$ only depending on α such that

$$\mathbb{P}(N_+ \leq (1 - \alpha)^2 n \|f_X\|_1) \leq \exp(-\alpha' n \|f_X\|_1).$$

This shows that

$$A_n \leq \frac{\square_{T,a,\delta,\alpha,K}}{n \|f_X\|_1}.$$

We now deal with

$$B_n := \mathbb{E}[S_n^2 \mathbf{1}_{\{N_+ > (1-\alpha)^2 n \|f_X\|_1\}}].$$

We take $\alpha = \min(\eta/2, 1/4)$. This implies

$$(1 + \eta)(1 - \alpha) \geq 1 + \frac{\eta}{4}$$

and

$$\begin{aligned} B_n &= \mathbb{E} \left[\sup_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{(1 + \eta) \|K'\|_2 \sqrt{aT N_+}}{\sqrt{2nt^{3/2}}} \right\}_+^2 \mathbf{1}_{\{N_+ > (1-\alpha)^2 n \|f_X\|_1\}} \right] \\ &\leq \mathbb{E} \left[\sup_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{(1 + \eta/4) \sqrt{aT} \|K'\|_2 \sqrt{\|f_X\|_1}}{\sqrt{2nt^{3/2}}} \right\}_+^2 \right] \\ &\leq \int_0^{+\infty} \mathbb{P} \left(\sup_{t \in \mathcal{H}} \left\{ \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{(1 + \eta/4) \sqrt{aT} \|K'\|_2 \sqrt{\|f_X\|_1}}{\sqrt{2nt^{3/2}}} \right\}_+ \geq x \right) dx \\ &\leq \sum_{t \in \mathcal{H}} \int_0^{+\infty} \mathbb{P} \left(\left\{ \|\tilde{f}_t - \mathbb{E}[\tilde{f}_t]\|_{2,T} - \frac{(1 + \eta/4) \sqrt{aT} \|K'\|_2 \sqrt{\|f_X\|_1}}{\sqrt{2nt^{3/2}}} \right\}_+ \geq x \right) dx. \end{aligned}$$

To conclude, it remains to control for any $x > 0$ the probability inside the integral. For this purpose, we use the following lemma.

Lemma 3. *Let $\varepsilon > 0$ and $h \in \mathcal{H}$ be fixed. For any $x > 0$, with probability larger than $1 - \exp(-x)$,*

$$\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T} \leq (1 + \varepsilon) \|K'\|_2 \sqrt{\frac{aT \|f_X\|_1}{2nh^3}} + \sqrt{12x} \|K'\|_1 \sqrt{\frac{\|f_X\|_1 (T + 4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1})x \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T + 4a)}{2}}.$$

Proof. We set:

$$\begin{aligned} U(t) &= \tilde{f}_h(t) - \mathbb{E}[\tilde{f}_h(t)] \\ &= \frac{a}{nh^2} \sum_{k=-\infty}^{+\infty} s_k \int_{\mathbb{R}} K' \left(\frac{t - (2k + 1)a - u}{h} \right) [dN_u^Y - n f_Y(u) du] \\ &= \frac{a}{nh} \int_{\mathbb{R}} L_h(t - u) [dN_u^Y - n f_Y(u) du], \end{aligned}$$

with

$$L_h(x) := \frac{1}{h} \sum_{k=-\infty}^{+\infty} s_k K' \left(\frac{x - (2k+1)a}{h} \right).$$

Let \mathcal{D} a countable dense subset of the unit ball of $\mathbb{L}_2[0, T]$. We have:

$$\begin{aligned} \|U\|_{2,T} &= \sup_{g \in \mathcal{D}} \int_0^T g(t) U(t) dt \\ &= \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \Psi_g(u) [dN_u^Y - n f_Y(u) du], \end{aligned}$$

with

$$\Psi_g(u) := \frac{a}{nh} \int_0^T L_h(t-u) g(t) dt = \frac{a}{nh} (\tilde{L}_h \star g)(u)$$

and $\tilde{L}_h(x) = L_h(-x)$, where the convolution product is computed on $[0, T]$. We use Corollary 2 of Reynaud-Bouret (2003). So, we need to bound $\mathbb{E}[\|U\|_{2,T}]$ and

$$v_0 := \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \Psi_g^2(u) n f_Y(u) du.$$

We also have to determine b , a deterministic upper bound for all the Ψ_g 's. We have already proved in the proof of Lemma 1 that

$$\mathbb{E}[\|U\|_{2,T}^2] = \frac{aT \|f_X\|_1 \|K'\|_2^2}{2nh^3},$$

which implies

$$\mathbb{E}[\|U\|_{2,T}] \leq \|K'\|_2 \sqrt{\frac{aT \|f_X\|_1}{2nh^3}}. \quad (7.2)$$

If we denote $I(h, u) := \{k \in \mathbb{Z} : -u - Ah - a \leq 2ka \leq Ah + T - u - a\}$, then

$$\begin{aligned} \Psi_g^2(u) &= \frac{a^2}{n^2 h^2} \left(\int_0^T L_h(t-u) g(t) dt \right)^2 \\ &\leq \frac{a^2}{n^2 h^2} \int_0^T L_h^2(t-u) dt \times \int_0^T g^2(t) dt \\ &\leq \frac{a^2}{n^2 h^4} \int_0^T \sum_{k \in I(h, u)} \left(K' \left(\frac{t-u-(2k+1)a}{h} \right) \right)^2 dt \\ &\leq \frac{a^2}{n^2 h^3} \|K'\|_2^2 \times \text{card}(I(h, u)) \\ &\leq \frac{a^2}{n^2 h^3} \|K'\|_2^2 \times (T/(2a) + Ah/a + 1) \end{aligned}$$

and we can set, under the condition on \mathcal{H} ,

$$b := \frac{a}{nh^{3/2}} \|K'\|_2 \sqrt{\frac{T+4a}{2a}},$$

which is negligible with respect to the upper bound of $\mathbb{E}[\|U\|_{2,T}]$ given in (7.2). We now deal with

$$v_0 := n \times \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \Psi_g^2(u) f_Y(u) du.$$

We have:

$$\begin{aligned} v_0 &= \frac{a^2}{nh^2} \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \left(\int_0^T L_h(t-u) g(t) dt \right)^2 f_Y(u) du \\ &\leq \frac{a^2}{nh^2} \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \left(\int_0^T |L_h(t-u)| dt \int_0^T |L_h(t-u)| g^2(t) dt \right) f_Y(u) du. \end{aligned}$$

Since

$$\begin{aligned} \int_0^T |L_h(t-u)| dt &\leq \frac{1}{h} \int_0^T \sum_{k \in I(h,u)} \left| K' \left(\frac{t-u-(2k+1)a}{h} \right) \right| dt \\ &\leq \|K'\|_1 \text{card}(I(h,u)) \\ &\leq \|K'\|_1 \frac{(T+4a)}{2a}, \end{aligned}$$

we obtain

$$\begin{aligned} v_0 &\leq \frac{a^2}{nh^2} \|K'\|_1 \frac{(T+4a)}{2a} \sup_{g \in \mathcal{D}} \int_{\mathbb{R}} \int_0^T \sum_{k=-\infty}^{+\infty} \frac{1}{h} \left| K' \left(\frac{t-u-(2k+1)a}{h} \right) \right| g^2(t) dt f_Y(u) du \\ &\leq \frac{a}{2nh^2} \|K'\|_1 (T+4a) \sup_{g \in \mathcal{D}} \int_0^T \left(\int \frac{1}{h} \left| K' \left(\frac{t-v}{h} \right) \right| \sum_{k=-\infty}^{+\infty} f_Y(v-(2k+1)a) dv \right) g^2(t) dt \\ &\leq \frac{a}{2nh^2} \|K'\|_1 (T+4a) \sup_{g \in \mathcal{D}} \int_0^T \left(\int \frac{1}{h} \left| K' \left(\frac{t-v}{h} \right) \right| dv \right) \frac{\|f_X\|_1}{2a} g^2(t) dt \\ &\leq \frac{\|f_X\|_1}{4nh^2} \|K'\|_1^2 (T+4a). \end{aligned}$$

Inequality (5.7) of Reynaud-Bouret (2003) yields, for any $x > 0$,

$$\mathbb{P}(\|U\|_{2,T} \geq (1+\varepsilon)\mathbb{E}[\|U\|_{2,T}] + \sqrt{12v_0x} + (1.25 + 32\varepsilon^{-1})bx) \leq \exp(-x).$$

Since

$$\begin{aligned} RHS &:= (1+\varepsilon)\mathbb{E}[\|U\|_{2,T}] + \sqrt{12v_0x} + (1.25 + 32\varepsilon^{-1})bx \\ &\leq (1+\varepsilon)\|K'\|_2 \sqrt{\frac{aT\|f_X\|_1}{2nh^3}} + \sqrt{12x}\|K'\|_1 \sqrt{\frac{\|f_X\|_1(T+4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1})x \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T+4a)}{2}}. \end{aligned}$$

□

The previous lemma states that for any sequence of weights $(w_h)_{h \in \mathcal{H}}$, setting $x = w_h + u$, with $u > 0$, with probability larger than $1 - \exp(-u) \sum_{h \in \mathcal{H}} \exp(-w_h)$, for all $h \in \mathcal{H}$,

$$\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T} \leq M_h + \sqrt{12u}\|K'\|_1 \sqrt{\frac{\|f_X\|_1(T+4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1})u \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T+4a)}{2}}$$

with

$$\begin{aligned}
M_h &:= (1 + \varepsilon) \|K'\|_2 \sqrt{\frac{aT \|f_X\|_1}{2nh^3}} + \sqrt{12} w_h \|K'\|_1 \sqrt{\frac{\|f_X\|_1 (T + 4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1}) w_h \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T + 4a)}{2}} \\
&= \|K'\|_2 \sqrt{\frac{aT \|f_X\|_1}{2nh^3}} \left(1 + \varepsilon + \sqrt{12} w_h h \frac{\|K'\|_1}{\|K'\|_2} \frac{\sqrt{T + 4a}}{\sqrt{2aT}} + \frac{(1.25 + 32\varepsilon^{-1}) w_h}{\sqrt{n \|f_X\|_1}} \sqrt{\frac{T + 4a}{T}} \right) \\
&\leq \frac{(1 + \eta/4) \sqrt{aT} \|K'\|_2 \sqrt{\|f_X\|_1}}{\sqrt{2nh^{3/2}}},
\end{aligned}$$

for ε small enough and for n large enough, by taking $w_h = h^{-1/2} |\log h|^{-1}$ for instance, since in this case,

$$w_h h = o(1) \quad \text{and} \quad h^{-1} = O(n \|f_X\|_1).$$

Therefore,

$$B_n \leq \sum_{h \in \mathcal{H}} \int_0^{+\infty} \mathbb{P} \left(\left\{ \|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|_{2,T} - M_h \right\}_+^2 \geq x \right) dx.$$

By setting u such that

$$x := (g(u))^2 = \left(\sqrt{12} u \|K'\|_1 \sqrt{\frac{\|f_X\|_1 (T + 4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1}) u \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T + 4a)}{2}} \right)^2,$$

so

$$dx = 2g(u) \times \left(\frac{\sqrt{12}}{2\sqrt{u}} \|K'\|_1 \sqrt{\frac{\|f_X\|_1 (T + 4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1}) \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T + 4a)}{2}} \right) du$$

and

$$\begin{aligned}
B_n &\leq \sum_{h \in \mathcal{H}} \int_0^{+\infty} e^{-(w_h + u)} \times 2g(u) \times \left(\frac{\sqrt{12}}{2\sqrt{u}} \|K'\|_1 \sqrt{\frac{\|f_X\|_1 (T + 4a)}{4nh^2}} + (1.25 + 32\varepsilon^{-1}) \frac{\|K'\|_2}{nh^{3/2}} \sqrt{\frac{a(T + 4a)}{2}} \right) du \\
&\leq 2 \sum_{h \in \mathcal{H}} e^{-w_h} \int_0^{+\infty} e^{-u} (g(u))^2 u^{-1} du \\
&\leq \square_{a,K,T,\eta} \left(\frac{\|f_X\|_1}{n} + \frac{1}{n^2} \right).
\end{aligned}$$

Finally, we obtain

$$\mathbb{E}[\zeta_n^2(h)] \leq (\|K\|_1 + 1)^2 \mathbb{E}[S_n^2] \leq (\|K\|_1 + 1)^2 \left(\frac{\square_{T,a,\delta,\alpha,K}}{n \|f_X\|_1} + \square_{a,K,T,\eta} \left(\frac{\|f_X\|_1}{n} + \frac{1}{n^2} \right) \right),$$

which concludes the proof of the theorem.

7.4 Proof of Theorem 2

To prove Theorem 2, without loss of generality, we assume that T is a positive integer. We denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. The cardinal of a finite set m is denoted by $|m|$.

As usual in the proofs of lower bounds, we build a set of intensities $(f_m)_{m \in \mathcal{M}}$ quite distant from each other in terms of the L_2 -norm, but whose distance between the resulting models is small. This set of intensities is based on wavelet expansions. More precisely, let ψ be the Meyer wavelet built with with C^2 -conjugate mirror filters (see for instance Section 7.7.2 of Mallat (2009)). We shall use in particular that ψ is C^∞ and there exists a positive constant c_ψ such that

1. $|\psi(x)| \leq c_\psi(1 + |x|)^{-2}$ for any $x \in \mathbb{R}$,
2. ψ^* is C^2 and ψ^* has support included into $[-8\pi/3, -2\pi/3] \cup [2\pi/3, 8\pi/3]$,

where $\psi^*(\xi) = \int e^{it\xi} \psi(x) dx$ is the Fourier transform of ψ . Observe that this implies that the functions

$$\xi \mapsto \psi^*(\xi), \quad \xi \mapsto \psi^*(\xi)\xi^{-1}, \quad \xi \mapsto \psi^*(\xi)\xi^{-2}, \quad \xi \mapsto (\psi^*)'(\xi) \quad \text{and} \quad \xi \mapsto (\psi^*)'(\xi)\xi^{-1}$$

are bounded by a constant. Without loss of generality, we assume that this constant is c_ψ .

Let

$$f_1(x) = \frac{c_1}{1 + x^2},$$

where c_1 is a positive constant small enough, which is chosen such that f_1 belongs to $\mathcal{S}^\beta(L/2)$. Indeed, note that

$$c_\beta^2 := \int |c_1^{-1} f_1^*(\xi)|^2 (\xi^2 + 1)^\beta = \int \pi^2 \exp(-2|\xi|) (\xi^2 + 1)^\beta < \infty$$

so that it is sufficient to choose $c_1 = c_\beta^{-1} L/2$.

We recall a combinatorial lemma due to Birgé and Massart (see Lemma 8 in Reynaud-Bouret (2003), see also Lemma 2.9 in Tsybakov (2008)).

Lemma 4. *Let D an integer and Γ be a finite set with cardinal D . There exist absolute constants τ and σ such that there exists $\mathcal{M}_D \subset \mathcal{P}(\Gamma)$, satisfying $\log |\mathcal{M}_D| \geq \sigma D$ and such that for all distinct sets m and m' belonging to \mathcal{M}_D the symmetric difference of m and m' , denoted $m\Delta m'$, satisfies $|m\Delta m'| \geq \tau D$.*

Here we choose $\Gamma := \{0, \dots, D-1\}$ with $D := T2^{j-1}$ where j is an integer to be chosen later, and we denote $\mathcal{M} := \mathcal{M}_D$ given in the previous lemma. Thus $\log |\mathcal{M}| \geq \sigma T2^{j-1}$ and for all $m, m' \in \mathcal{M} : \tau T2^{j-1} \leq |m\Delta m'| \leq T2^{j-1}$.

Now, for $a_j > 0$ to be chosen, for $m \in \mathcal{M}$, for $x \in \mathbb{R}$, we set

$$f_m(x) := f_1(x) + a_j \sum_{k \in m} \psi_{jk}(x),$$

where we have denoted, as usual, $\psi_{jk}(x) := 2^{j/2} \psi(2^j x - k)$.

We compute $\psi_{jk}^*(\xi) = 2^{-j/2} e^{i\xi k 2^{-j}} \psi^*(\xi 2^{-j})$, which gives

$$\begin{aligned} \int |(f_m - f_1)^*(\xi)|^2 (1 + \xi^2)^\beta d\xi &= \int \left| a_j 2^{-j/2} \psi^*(\xi 2^{-j}) \sum_{k \in m} e^{i\xi k 2^{-j}} \right|^2 (1 + \xi^2)^\beta d\xi \\ &= a_j^2 \int \left| \psi^*(t) \sum_{k \in m} e^{ikt} \right|^2 (1 + t^2 2^{2j})^\beta dt \\ &\leq \square_{\psi, \beta} a_j^2 2^{2j\beta} \int_{-3\pi}^{3\pi} \left| \sum_{k \in m} e^{ikt} \right|^2 dt \\ &\leq \square_{\psi, \beta} 2^{2j\beta} |m| a_j^2 \leq \square_{\psi, \beta} T 2^{j(2\beta+1)} a_j^2, \end{aligned}$$

using Parseval's theorem and $|m| \leq D = T 2^{j-1}$. We assume from now on that

$$T 2^{j(2\beta+1)} a_j^2 \leq C(\psi, \beta) L^2 \quad (7.3)$$

for $C(\psi, \beta)$ a constant only depending on β and L small enough, so that $(f_m - f_1)$ belongs to $\mathcal{S}^\beta(L/2)$ and then $f_m \in \mathcal{S}^\beta(L)$.

Let us verify that f_m is non-negative, and then is an intensity of a Poisson process. For all real x ,

$$\frac{f_m(x) - f_1(x)}{f_1(x)} = c_1^{-1} (1 + x^2) a_j 2^{j/2} \sum_{k \in m} \psi(2^j x - k).$$

Recall that $\psi(x) \leq c_\psi (1 + |x|)^{-2}$. Let us now study 3 cases.

1. If $0 \leq |x| \leq T + 1$, we have:

$$(1+x^2) \left| \sum_{k \in m} \psi(2^j x - k) \right| \leq (T^2 + 2T + 2) c_\psi \sum_{k \in m} (1 + |2^j x - k|)^{-2} \leq 2c_\psi (T^2 + 2T + 2) \sum_{\ell=1}^{+\infty} \ell^{-2},$$

and the last upper bound is smaller than a finite constant only depending on T and c_ψ .

2. If $x \geq T + 1$, since $|m| \leq D = T 2^{j-1}$, we have:

$$(1+x^2) \left| \sum_{k \in m} \psi(2^j x - k) \right| \leq c_\psi T 2^{j-1} (1 + 2^j(x-T))^{-2} (1+x^2) \leq c_\psi T 2^{-j-1} \sup_{x \geq T+1} \frac{1+x^2}{(x-T)^2},$$

and the last expression is smaller than a finite constant only depending on T and c_ψ .

3. If $x \leq -T - 1$,

$$(1+x^2) \left| \sum_{k \in m} \psi(2^j x - k) \right| \leq c_\psi T 2^{j-1} (1 + 2^j(-x))^{-2} (1+x^2) \leq c_\psi T 2^{-j-1} \sup_{x \leq -T-1} \frac{1+x^2}{(-x)^2},$$

Finally we obtain that there exists $\bar{C}(T, c_\psi)$ a constant only depending on T and c_ψ such that

$$\frac{|f_m(x) - f_1(x)|}{f_1(x)} \leq c_1^{-1} a_j 2^{j/2} \bar{C}(T, c_\psi).$$

We take a_j such that

$$c_1^{-1} a_j 2^{j/2} \bar{C}(T, c_\psi) \leq \frac{1}{2}. \quad (7.4)$$

This ensures that $f_m \geq \frac{1}{2} f_1 \geq 0$. Another consequence is that $f_\varepsilon \star f_m \geq \frac{1}{2} f_\varepsilon \star f_1$. This provides

$$\begin{aligned} f_\varepsilon \star f_m(x) &\geq \frac{1}{2} \int_{-a}^a \frac{1}{2a} \frac{c_1}{1+(x-t)^2} dt \geq \frac{1}{2} \frac{c_1}{1+(|x|+a)^2} \\ &\geq \frac{1}{2} \frac{c_1}{1+2a^2+2x^2} \geq \frac{c_2^{-1}}{1+x^2}, \end{aligned}$$

denoting $c_2 = c_2(a, \beta, L) = \max(4, 2 + 4a^2)/c_1$.

Finally, we evaluate the distance between the distributions of the observations N^Y when N^X has intensity nf_m and $nf_{m'}$. We denote by \mathbb{P}_m the probability measure associated with N^Y , which has intensity $g_m := f_\varepsilon \star nf_m$, and we denote by $K(\mathbb{P}_m, \mathbb{P}_{m'})$ the Kullback-Leibler divergence between \mathbb{P}_m and $\mathbb{P}_{m'}$. Using Cavalier and Koo (2002), we have

$$K(\mathbb{P}_m, \mathbb{P}_{m'}) = \int g_m(x) \phi \left(\log \left(\frac{g_{m'}(x)}{g_m(x)} \right) \right) dx$$

where for any $x \in \mathbb{R}$, $\phi(x) = \exp(x) - x - 1$. Since for any $x > -1$, $\log(1+x) \geq x/(1+x)$, we have

$$K(\mathbb{P}_m, \mathbb{P}_{m'}) \leq \int \frac{(g_m(x) - g_{m'}(x))^2}{g_m(x)} dx = n \int \frac{((f_\varepsilon \star f_m)(x) - (f_\varepsilon \star f_{m'})(x))^2}{(f_\varepsilon \star f_m)(x)} dx.$$

For m and m' in \mathcal{M} , denote

$$\theta(x) = a_j^{-1} (f_\varepsilon \star (f_m - f_{m'}))(x) = \sum_{k \in m \Delta m'} b_k (f_\varepsilon \star \psi_{jk})(x)$$

with $b_k = 1$ if $k \in m$ and $b_k = -1$ if $k \in m'$. Denote also $\theta^*(\xi) = \int e^{i\xi x} \theta(x) dx$ its Fourier transform, and $(\theta^*)'(\xi) = \int i x e^{i\xi x} \theta(x) dx$ the derivative of θ^* . Parseval's theorem gives

$$\|\theta\|_2^2 = \frac{1}{2\pi} \|\theta^*\|_2^2, \quad \text{and} \quad \|x\theta(x)\|_2^2 = \frac{1}{2\pi} \|(\theta^*)'\|_2^2.$$

Thus

$$\begin{aligned} \frac{1}{n} K(\mathbb{P}_m, \mathbb{P}_{m'}) &\leq \int \frac{((f_\varepsilon \star f_m)(x) - (f_\varepsilon \star f_{m'})(x))^2}{(f_\varepsilon \star f_m)(x)} dx \leq c_2 \int (1+x^2) (f_\varepsilon \star (f_m - f_{m'})(x))^2 dx \\ &\leq c_2 a_j^2 \int (1+x^2) \theta(x)^2 dx \leq \frac{c_2}{2\pi} a_j^2 (\|\theta^*\|_2^2 + \|(\theta^*)'\|_2^2). \end{aligned}$$

We recall that $\psi_{jk}^*(\xi) = 2^{-j/2} e^{i\xi k 2^{-j}} \psi^*(\xi 2^{-j})$, which gives

$$\begin{aligned} \theta^*(\xi) &= \sum_{k \in m \Delta m'} b_k f_\varepsilon^*(\xi) \psi_{jk}^*(\xi) = \sum_{k \in m \Delta m'} b_k f_\varepsilon^*(\xi) 2^{-j/2} e^{i\xi k 2^{-j}} \psi^*(\xi 2^{-j}) \\ &= 2^{-j/2} f_\varepsilon^*(\xi) \psi^*(\xi 2^{-j}) \sum_{k \in m \Delta m'} b_k e^{i\xi k 2^{-j}}. \end{aligned}$$

Thus, remembering that for $\xi \in \mathbb{R}$,

$$|f_\varepsilon^*(\xi)| = \left| \frac{\sin(a\xi)}{a\xi} \right| \leq \min(1, |a\xi|^{-1}), \quad (7.5)$$

we have

$$\begin{aligned} \|\theta^*\|_2^2 &= \int \left| 2^{-j/2} f_\varepsilon^*(\xi) \psi^*(\xi 2^{-j}) \sum_{k \in m\Delta m'} b_k e^{i\xi k 2^{-j}} \right|^2 d\xi \\ &= \int \left| f_\varepsilon^*(u 2^j) \psi^*(u) \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq \int \min(1, |a 2^j u|^{-2}) |\psi^*(u)|^2 \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq \int_{-8\pi/3}^{8\pi/3} \min(1, |a 2^j|^{-2}) c_\psi^2 \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \end{aligned}$$

using the properties of ψ . Parseval's theorem gives

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du = \sum_{k \in m\Delta m'} b_k^2 = |m\Delta m'| \leq T 2^{j-1}.$$

Then

$$\|\theta^*\|_2^2 \leq 3\pi c_\psi^2 T 2^j (a^{-2} 2^{-2j} \wedge 1).$$

Let us now bound $\|(\theta^*)'\|_2^2$. First

$$(\psi_{jk}^*)'(\xi) = 2^{-3j/2} e^{i\xi k 2^{-j}} (\psi^*(\xi 2^{-j}) ik + (\psi^*)'(\xi 2^{-j})),$$

then

$$\begin{aligned} (\theta^*)'(\xi) &= \sum_{k \in m\Delta m'} b_k [(f_\varepsilon^*)'(\xi) \psi_{jk}^*(\xi) + f_\varepsilon^*(\xi) (\psi_{jk}^*)'(\xi)] \\ &= \sum_{k \in m\Delta m'} b_k (f_\varepsilon^*)'(\xi) 2^{-j/2} e^{i\xi k 2^{-j}} \psi^*(\xi 2^{-j}) + b_k f_\varepsilon^*(\xi) 2^{-3j/2} e^{i\xi k 2^{-j}} (\psi^*(\xi 2^{-j}) ik + (\psi^*)'(\xi 2^{-j})) \\ &= \alpha_1(\xi) + \alpha_2(\xi) + \alpha_3(\xi) \end{aligned}$$

where

$$\begin{aligned} \alpha_1(\xi) &= 2^{-j/2} (f_\varepsilon^*)'(\xi) \psi^*(\xi 2^{-j}) \sum_{k \in m\Delta m'} b_k e^{i\xi k 2^{-j}}, \\ \alpha_2(\xi) &= 2^{-3j/2} f_\varepsilon^*(\xi) \psi^*(\xi 2^{-j}) \sum_{k \in m\Delta m'} ik b_k e^{i\xi k 2^{-j}}, \\ \alpha_3(\xi) &= 2^{-3j/2} f_\varepsilon^*(\xi) (\psi^*)'(\xi 2^{-j}) \sum_{k \in m\Delta m'} b_k e^{i\xi k 2^{-j}}. \end{aligned}$$

Reasoning as before, and using that

$$|(f_\varepsilon^*)'(\xi)| = \left| \frac{\cos(a\xi)}{\xi} - \frac{1}{\xi} \times \frac{\sin(a\xi)}{a\xi} \right| \leq \frac{2}{|\xi|}, \quad (7.6)$$

we can write

$$\begin{aligned} \|\alpha_1\|_2^2 &= \int \left| (f_\varepsilon^*)'(u2^j)\psi^*(u) \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq \int \frac{4}{u^2 2^{2j}} |\psi^*(u)|^2 \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq 4c_\psi^2 2^{-2j} \times 6\pi |m\Delta m'| \\ &\leq 12\pi c_\psi^2 T 2^{-j}. \end{aligned}$$

In the same way, using (7.5),

$$\begin{aligned} \|\alpha_2\|_2^2 &= 2^{-2j} \int \left| f_\varepsilon^*(u2^j)\psi^*(u) \sum_{k \in m\Delta m'} ikb_k e^{iku} \right|^2 du \\ &\leq 2^{-2j} \int (1 \wedge a^2 2^{-2j} u^{-2}) |\psi^*(u)|^2 \left| \sum_{k \in m\Delta m'} ikb_k e^{iku} \right|^2 du \\ &\leq 3c_\psi^2 2^{-2j} (a^{-2} 2^{-2j} \wedge 1) \int_{-\pi}^{\pi} \left| \sum_{k \in m\Delta m'} ikb_k e^{iku} \right|^2 du \\ &\leq 6\pi c_\psi^2 2^{-2j} (a^{-2} 2^{-2j} \wedge 1) \sum_{k \in m\Delta m'} k^2 \end{aligned}$$

and we obtain that

$$\|\alpha_2\|_2^2 \leq cc_\psi^2 T^3 2^j (a^{-2} 2^{-2j} \wedge 1),$$

for c an absolute constant. Similarly,

$$\begin{aligned} \|\alpha_3\|_2^2 &= 2^{-2j} \int \left| f_\varepsilon^*(u2^j)(\psi^*)'(u) \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq 2^{-2j} \int (1 \wedge a^2 2^{-2j} u^{-2}) |(\psi^*)'(u)|^2 \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq 3c_\psi^2 2^{-2j} (a^{-2} 2^{-2j} \wedge 1) \int_{-\pi}^{\pi} \left| \sum_{k \in m\Delta m'} b_k e^{iku} \right|^2 du \\ &\leq 3\pi c_\psi^2 T 2^{-j} (a^{-2} 2^{-2j} \wedge 1). \end{aligned}$$

Finally, since a is smaller than an absolute constant and T is larger than an absolute constant, we have that

$$K(\mathbb{P}_m, \mathbb{P}_{m'}) \leq Cc_2 c_\psi^2 n a_j^2 T^3 2^j (a^{-2} 2^{-2j} \wedge 1),$$

for C an absolute constant.

Now, let us give the following version of Fano's lemma, derived from Birgé (2001).

Lemma 5. Let $(\mathbb{P}_i)_{i \in \{0, \dots, I\}}$ be a finite family of probability measures defined on the same measurable space Ω . One sets

$$\bar{K}_I = \frac{1}{I} \sum_{i=1}^I K(\mathbb{P}_i, \mathbb{P}_0).$$

Then, there exists an absolute constant B ($B = 0.71$ works) such that if Z is a random variable on Ω with values in $\{0, \dots, I\}$, one has

$$\inf_{0 \leq i \leq I} \mathbb{P}_i(Z = i) \leq \max \left(B, \frac{\bar{K}_I}{\log(I+1)} \right).$$

We apply this lemma with \mathcal{M} instead of $\{0, \dots, I\}$, whose log-cardinal is larger than $T2^{j-1}$ up to an absolute constant. We take a_j such that

$$\frac{C c_2 c_\psi^2 n a_j^2 T^3 2^j (a^{-2} 2^{-2j} \wedge 1)}{\log |\mathcal{M}|} \leq B,$$

which is satisfied if

$$a_j^2 \leq \frac{C(\psi)}{n T^2 c_2} (a^2 2^{2j} \vee 1), \quad (7.7)$$

with $C(\psi)$ a constant only depending on ψ small enough. Thus if Z is a random variable with values in m , $\inf_{m \in \mathcal{M}} \mathbb{P}_m(Z = m) \leq B$. Now,

$$\begin{aligned} \inf_{Z_n} \sup_{f_X \in \mathcal{S}^\beta(L)} \mathbb{E}_{f_X} [\|Z_n - f_X\|_{2,T}^2] &\geq \inf_{Z_n} \sup_{m \in \mathcal{M}} \mathbb{E}_{f_m} [\|Z_n - f_m\|_{2,T}^2] \\ &\geq \frac{1}{4} \inf_{m' \in \mathcal{M}} \sup_{m \in \mathcal{M}} \mathbb{E}_{f_m} [\|f_{m'} - f_m\|_{2,T}^2]. \end{aligned} \quad (7.8)$$

For the last inequality, we have used that if Z_n is an estimate, we define

$$m' \in \arg \min_{m \in \mathcal{M}} \mathbb{E}_{f_m} [\|Z_n - f_m\|_{2,T}^2]$$

and for $m \in \mathcal{M}$,

$$\|f_{m'} - f_m\|_{2,T} \leq \|f_{m'} - Z_n\|_{2,T} + \|f_m - Z_n\|_{2,T} \leq 2\|f_m - Z_n\|_{2,T}.$$

Since $f_m - f_{m'} = a_j \sum_{k \in m \Delta m'} b_k \psi_{jk}$ and (ψ_{jk}) is an orthonormal family, we have for $m \neq m'$,

$$\|f_m - f_{m'}\|_2^2 = a_j^2 |m \Delta m'| \geq \tau a_j^2 T 2^{j-1}, \quad (7.9)$$

for τ the absolute constant defined in Lemma 4. Furthermore,

$$\begin{aligned} 0 \leq \|f_m - f_{m'}\|_2^2 - \|f_m - f_{m'}\|_{2,T}^2 &= \int (f_m(x) - f_{m'}(x))^2 1_{\{|x| > T\}} dx \\ &= a_j^2 \int \left(\sum_{k \in m \Delta m'} b_k \psi_{jk}(x) \right)^2 1_{\{|x| > T\}} dx \\ &\leq a_j^2 \sum_{k \in m \Delta m'} b_k^2 \times \sum_{k \in m \Delta m'} \int \psi_{jk}^2(x) 1_{\{|x| > T\}} dx \\ &\leq a_j^2 |m \Delta m'| \times \sum_{k \in m \Delta m'} \int \psi_{jk}^2(x) 1_{\{|x| > T\}} dx. \end{aligned}$$

Then, since $0 \leq k \leq T2^{j-1}$,

$$\begin{aligned}
\int \psi_{jk}^2(x) 1_{\{|x|>T\}} dx &= \int \psi^2(u) 1_{\{|2^{-j}(u+k)|>T\}} du \\
&\leq c_\psi^2 \int (1+|u|)^{-4} 1_{\{|2^{-j}(u+k)|>T\}} du \\
&\leq c_\psi^2 \int_{2^{jT-k}}^{+\infty} (1+|u|)^{-4} du + c_\psi^2 \int_{-\infty}^{-2^{jT-k}} (1+|u|)^{-4} du \\
&\leq 2c_\psi^2 \int_{2^{j-1}T}^{+\infty} (1+u)^{-4} du \\
&\leq \frac{2c_\psi^2}{3} (T2^{j-1})^{-3}.
\end{aligned}$$

We finally obtain

$$0 \leq \|f_m - f_{m'}\|_2^2 - \|f_m - f_{m'}\|_{2,T}^2 \leq \frac{2c_\psi^2}{3} (T2^{j-1})^{-2} \times a_j^2$$

and using (7.9), for j large enough depending on T and $m \neq m'$,

$$\|f_m - f_{m'}\|_{2,T}^2 \geq C'(\psi) a_j^2 T 2^j,$$

for $C'(\psi)$ a constant only depending on ψ . Finally, applying (7.8) and Lemma 5, we obtain:

$$\begin{aligned}
\inf_{Z_n} \sup_{f_X \in \mathcal{S}^\beta(L)} \mathbb{E} [\|Z_n - f_X\|_{2,T}^2] &\geq \frac{1}{4} \inf_{m' \in \mathcal{M}} \sup_{m \in \mathcal{M}} \mathbb{E}_{f_m} [\|f_{m'} - f_m\|_{2,T}^2] \\
&\geq \frac{C'(\psi) a_j^2 T 2^j}{4} \inf_{m' \in \mathcal{M}} \sup_{m \in \mathcal{M}} \mathbb{P}_m(m' \neq m) \geq \frac{C'(\psi) a_j^2 T 2^j}{4} (1 - B).
\end{aligned}$$

Now, we choose $a_j > 0$ as large as possible such that (7.3), (7.4) and (7.7) are satisfied, meaning that

$$a_j^2 T 2^j = \left(C(\psi, \beta) L^2 2^{-2j\beta} \right) \wedge \left(\frac{c_1^2 T \bar{C}(T, c_\psi)^{-2}}{4} \right) \wedge \left(\frac{C(\psi) 2^j}{n T c_2} (a^2 2^{2j} \vee 1) \right).$$

Since $c_2 = \max(4, 2 + 4a^2)/c_1$ and $c_1 = c_\beta^{-1} L/2$, it simplifies in

$$a_j^2 T 2^j = \square_{T, \psi, \beta, a} \left(L^2 2^{-2j\beta} \wedge L^2 \wedge \frac{L 2^j}{n} (2^{2j} \vee 1) \right).$$

We assume that L is larger than a constant and we take

$$2^j \leq (L^2 n)^{1/(2\beta+3)} < 2^{j+1},$$

which yields that for n large enough,

$$\inf_{Z_n} \sup_{f_X \in \mathcal{S}^\beta(L)} \mathbb{E} [\|Z_n - f_X\|_2^2] \geq \square_{T, \psi, \beta, a} L^{6/(2\beta+3)} n^{-2\beta/(2\beta+3)}.$$

References

- Antoniadis, A. and Bigot, J. (2006). Poisson inverse problems. *Ann. Statist.*, 34(5):2132–2158.
- Belomestny, D. and Goldenshluger, A. (2019). Density deconvolution under general assumptions on the distribution of measurement errors. *arXiv preprint arXiv:1907.11024*.
- Birgé, L. (2001). A new look at an old result: Fano’s lemma. Technical report, Université Pierre et Marie Curie.
- Carrasco, M. and Florens, J.-P. (2011). A spectral method for deconvolving a density. *Econometric Theory*, pages 546–581.
- Cavalier, L. and Koo, J.-Y. (2002). Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Transactions on Information Theory*, 48(10):2794–2802.
- Delaigle, A. and Meister, A. (2011). Nonparametric function estimation under fourier-oscillating noise. *Statistica Sinica*, pages 1065–1092.
- Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *The Annals of Statistics*, pages 600–610.
- Feuerverger, A., Kim, P. T., and Sun, J. (2008). On optimal uniform deconvolution. *Journal of Statistical Theory and Practice*, 2(3):433–451.
- Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on R^d . *Probability Theory and Related Fields*, 159(3-4):479–543.
- Goldenshluger, A. V. and Lepski, O. V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl.*, 57(2):209–226.
- Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Statistica Neerlandica*, 57(1):136–157.
- Hall, P., Meister, A., et al. (2007). A ridge-parameter approach to deconvolution. *The Annals of Statistics*, 35(4):1535–1558.
- Hall, P., Ruymgaart, F., van Gaans, O., and van Rooij, A. (2001). Inverting noisy integral equations using wavelet expansions: a class of irregular convolutions. In *State of the art in probability and statistics (Leiden, 1999)*, volume 36 of *IMS Lecture Notes Monogr. Ser.*, pages 533–546. Inst. Math. Statist., Beachwood, OH.
- Hohage, T. and Werner, F. (2016). Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, 32(9):093001.
- Kerkycharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2):137–170.
- Kroll, M. (2019). Nonparametric intensity estimation from noisy observations of a Poisson process under unknown error distribution. *Metrika*, 82(8):961–990.

- Kuusela, M., Panaretos, V. M., et al. (2015). Statistical unfolding of elementary particle spectra: Empirical bayes estimation and bias-corrected uncertainty quantification. *The Annals of Applied Statistics*, 9(3):1671–1705.
- Mallat, S. (2009). *A wavelet tour of signal processing*. Elsevier/Academic Press, Amsterdam, third edition.
- Meister, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems*, 24(1):015003, 14.
- Meister, A. (2009). Density deconvolution. In *Deconvolution Problems in Nonparametric Statistics*, pages 5–105. Springer.
- Picard, F., Cadoret, J.-C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.-N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet*, 10(5):e1004282.
- Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields*, 126(1):103–153.
- Reynaud-Bouret, P. and Rivoirard, V. (2010). Near optimal thresholding estimation of a Poisson intensity on the real line. *Electron. J. Stat.*, 4:172–238.
- Trong, D. D., Phuong, C. X., Tuyen, T. T., and Thanh, D. N. (2014). Tikhonov’s regularization to the deconvolution problem. *Communications in Statistics-Theory and Methods*, 43(20):4384–4400.
- Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.
- van Es, B. (2011). Combining kernel estimators in the uniform deconvolution problem. *Stat. Neerl.*, 65(3):275–296.