

“Cytometry inference through adaptive atomic  
deconvolution”

Manon Costa, Sébastien Gadat, Pauline Gonnord, and Laurent Risser

# Cytometry inference through adaptive atomic deconvolution

Manon Costa <sup>1</sup>, Sébastien Gadat<sup>2</sup>, Pauline Gonnord<sup>3</sup>, and Laurent Risser<sup>1</sup>

<sup>1</sup> Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse III, France.

<sup>2</sup>Toulouse School of Economics, UMR 5604, Université de Toulouse 1, France.

<sup>3</sup>Centre de Physiopathologie Toulouse Purpan (CPTP), INSERM UMR1043, CNRS UMR 5282. Université Toulouse III, France.

August 11, 2017

## Abstract

In this paper we consider a statistical estimation problem known as atomic deconvolution. Introduced in reliability, this model has a direct application when considering biological data produced by flow cytometers. In these experiments, biologists measure the fluorescence emission of treated cells and compare them with their natural emission to study the presence of specific molecules on the cells' surface. They observe a signal which is composed of a noise (the natural fluorescence) plus some additional signal related to the quantity of molecule present on the surface if any. From a statistical point of view, we aim at inferring the percentage of cells expressing the selected molecule and the probability distribution function associated with its fluorescence emission. We propose here an adaptive estimation procedure based on a previous deconvolution procedure introduced by [vEGS08, GvES11]. For both estimating the mixing parameter and the mixing density automatically, we use the Lepskii method based on the optimal choice of a bandwidth using a bias-variance decomposition. We then derive some concentration inequalities for our estimators and obtain the convergence rates, that are shown to be minimax optimal (up to some log terms) in Sobolev classes. Finally, we apply our algorithm on simulated and real biological data.

**Keywords:** Mixture models, Atomic deconvolution, Adaptive kernel estimators, Inverse problems .

**MSC2010:** Primary: 62G07. Secondary: 62G20, 62P10

## 1 Introduction

### 1.1 Motivation

This paper deals with a statistical estimation problem close to the standard deconvolution problem in density estimation, and known as the problem of *atomic deconvolution*. This problem has been recently introduced in [vEGS08] and motivated by a reliability problem estimation. We consider in this work a natural application of this model to the biological datasets produced by flow cytometers. To motivate our theoretical and practical study, we have chosen to first introduce a common problem that should be addressed by biologist researchers when using flow cytometry measurements. A flow cytometer is an electronic machine that makes it possible to analyse a large number of cells and produce cell engineering results such as counting, sorting or bio-marker detection. This technology is commonly used for measuring the expression levels of proteins on the cells' surface. To this aim, the biologist use fluorescent antibodies which bind with specific proteins on the surface of cells. The flow cytometer then measures on each cells the fluorescent intensity, which is a reflect of the quantity of marker expressed by the cell. More precisely these measurements derive from a standard procedure in flow cytometry :

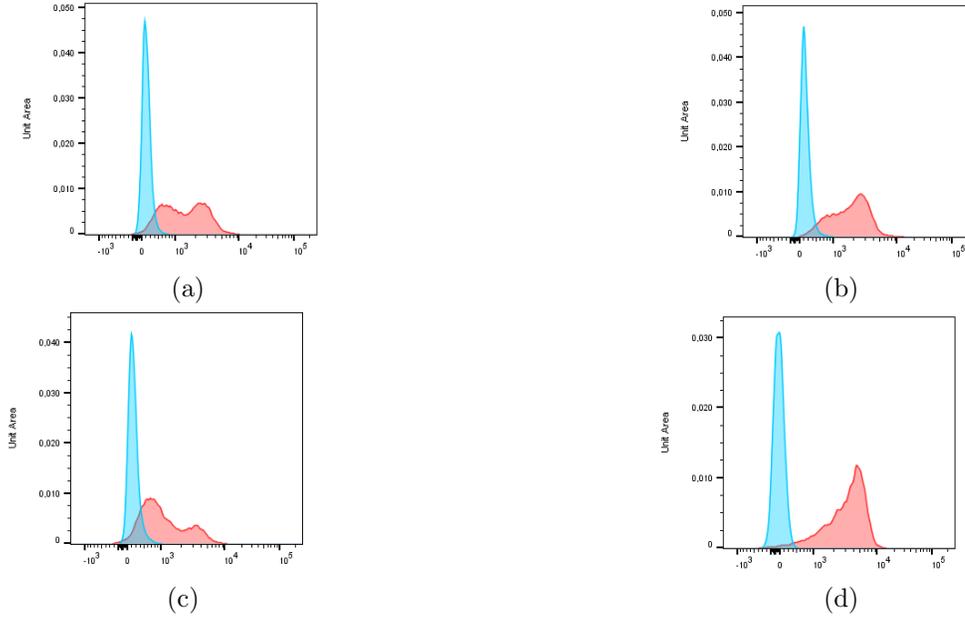


Figure 1: Effect of binding of an antibody specific for the CD27 protein expressed at the cell surface of lymphocytes extracted from the blood of an healthy individual (four different samples). The blue histograms represent the distribution of the baseline photo emission of untreated cells while the red histograms gives the fluorescent distribution of treated cells. We refer to Section 4.2 for more details on the samples.

- First the biologist performs a *calibration* that corresponds to the preliminary estimation of the baseline population of cells without any treatment: a large number of cells is placed in the cytometer and a fluorescent intensity is measured on each cell. This calibration ends with a baseline estimation of the the baseline photon emission by a population of cells.
- Then the biologist exposes the cells to a fluorescent antibody, which binds with the marker of interest on the surface of cells. The new fluorescence empirical distribution is built by the cytometer.
- The expertise of the biologist is at last used to calibrate a qualitative analysis to decide if there is an effect (or not) of the treatment and what is its mean effect. We should mention that this human expertise may be also assisted by some recent statistical computational analyses and have resulted in an open project called *FlowCAP* that makes it possible to produce standard statistical analysis (data clustering for example).

Figure 1 is an illustration of the empirical distribution of fluorescence measured by a cytometer before and after a treatment on different kind of cells with a logarithmic scale.

Our objective here is to enrich the possibilities of automatic estimation on flow cytometer datasets through a statistical procedure that permit to retrieve both the proportion of positive cells (*i.e.* the cells that express a given marker) and the distribution of the marker in a population of cells.

## 1.2 Model

The above phenomenon can be described as the following atomic deconvolution problem: we observe some i.i.d. realizations  $(Z_i)_{1 \leq i \leq n}$  of a model given by:

$$Z = U + AX, \tag{1}$$

where  $U$ ,  $A$  and  $X$  are jointly independent random variables. In Equation (1), the random variable  $Z$  is therefore the measurement of the flow cytometer of the fluorescent intensity on a given cell. The

random variable  $U$  represents the natural fluorescence of the cell: it is its baseline reaction regardless the impact of the drug administrated during the experiment. The random variable  $X$  stands for the quantitative effect of the reagent when the cell actively reacts to the treatment. At last,  $A$  represents the effectiveness of the treatment on the cell. We can therefore list below our definitions related to (1):

- $U$  represents the baseline “noise” on the observations, leading to the convolution model. We assume that the distribution of  $U$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$  and the density of  $U$  is assumed to be **known** and is denoted by  $g$ . This assumption for example translates a preliminary step of calibrating the cytometer before any treatment with a non-parametric estimation of the density  $g$ .
- $A$  follows a Bernoulli distribution  $A \sim \mathcal{B}(1 - p)$  and  $p$  is an **unknown** parameter. We use the convention  $\mathbb{P}[A = 0] = p$  (the cell does not react to the administrated reagent) while  $\mathbb{P}[A = 1] = 1 - p$  (the reagent induces an effect on the cell).
- $X$  is the effect of the treatment when the cell is reacting to the treatment and we assume that the distribution of  $X$  is also absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}$ , with an **unknown** density  $f$ . We also aim to estimate  $f$  to quantify the effect of the treatment.

For example, in each situation illustrated in Figure 1, we can see that the baseline density  $g$  is represented by the blue area while the “mixture” distribution obtained after the statistical contamination of  $U$  by  $AX$  is represented by the red area.

Even though relatively simple in appearance, we will see below that Equation (1) deserves both theoretical development and numerical efforts to obtain a practical and theoretically supported method of estimation for  $p$  and  $f$ . For example, we could think of the use of a moment estimation strategy to obtain informations on  $p$ . Unfortunately, we can rapidly compute that

$$\mathbb{E}(Z) = \mathbb{E}(U) + (1 - p)\mathbb{E}(X) \quad \text{and} \quad \mathbb{V}(Z) = \mathbb{V}(U) + p(1 - p)\mathbb{E}(X)^2 + (1 - p)\mathbb{V}(X),$$

leading to a non trivial relationship between the known moments of the distribution of  $U$ , the Bernoulli parameter  $p$ , and the moments of the distribution of  $X$ . This difficulty has been partially solved with the help of a non-parametric deconvolution strategy in [vEGS08, GvES11] adapted to the situation of atomic deconvolution. This estimator is inspired from the seminal contributions of [CH88] and [Fan91]. In particular, [vEGS08] builds an “ideal” consistent estimator and proves a CLT related to this estimator while [GvES11] studies the associated convergence rates and their minimax properties. Unfortunately, both works do not address the important question of adaptivity to the smoothness class  $\mathbf{H}_s(R)$  of the unknown density  $f$  (see below for details). The choice of the “good” smoothness parameter  $s$  may be a real issue from a practical point of view. In the general settings of non parametric regression, this adaptivity property may be attained by different strategies.

### 1.3 Adaptation

Hence, the method proposed in these works require an important improvement to produce a fully adaptive estimation strategy, and therefore to obtain a tractable algorithm for the estimation of  $p$  and  $f$ . Adaptivity of estimators is a very desirable property: in non-parametric statistics, estimators are importantly affected by a wrong choice of the parameters that are ideally smoothness-dependent. This is in particular the case for the bandwidth of a kernel deconvolution (see [Fan91]) or a frequency threshold in a Tychonov-type regularization (see [Tsy09]) for example. Different strategies have been designed these last decades to attain *ad-hoc* adaptivity of estimators. Among them, model selection theory introduced initially in [BM98] produced a lot of interesting derivations in non-parametric and high-dimensional statistics, while the cornerstone of a tractable use of such an approach relies on a suitable minimal penalty calibration. Nevertheless, this calibration is sometimes difficult both from a theoretical and from a practical point of view and other methods can also be used to produce efficient adaptive estimators. For example, resampling methods with an additional cross-validation procedure (see *e.g.* [PC84]) is a popular method but we should notice that deriving theoretical results with this simple method requires a significant amount of work (see the recent contributions of [AL16] for density estimation and of [Arl09] on general resampling methods). In statistical signal processing, adaptivity

of estimation may also be achieved with the help of a suitable decomposition basis and thresholding strategy. This is what is done with wavelets on Besov balls on specific situations: introduced for density estimation in [DJKP96] and used in different type of inverse problems in [BG10, BGKM13] for example. Nevertheless, wavelets are not so easy to use with mixture problems and we propose to follow another possible guideline. We will use the so-called Lepskii method to derive adaptive estimators. This method introduced in [Lep92] has been successfully applied in many non-parametric estimation problems and we refer to [GL11],[Lep15] for recent contributions using this method. We also refer to [Chi10] for an introduction of this method in different frameworks. The success of this method relies on a good bias-variance trade-off in the estimation procedure, and does not require a too much involved parameters tuning step.

## 1.4 Main assumptions

Below, we will use  $i$  to refer to the complex number such that  $i^2 = -1$  and the notation “:=” will refer to the definition of a mathematical object. We assume that the random variables  $U$  and  $X$  have a bounded second moment. The characteristic function of any random variable  $W$  will be denoted by  $\Phi_W$ . Therefore, we will frequently use the following notations:

$$\forall t \in \mathbb{R} \quad \Phi_Z(t) := \mathbb{E}[e^{itZ}] \quad \text{and} \quad \Phi_U(t) := \mathbb{E}[e^{itU}] \quad \text{and} \quad \Phi_X(t) := \mathbb{E}[e^{itX}],$$

where  $(Z, U, X)$  are the random variables involved in Equation (1). We assume that  $U$  (resp.  $X$ ) has a known density  $g$  (resp. unknown density  $f$ ) with respect to the Lebesgue measure. We will use for the sake of convenience the notation  $\Phi_h$  to refer to the Fourier transform of any density  $h$ :

$$\Phi_h(t) = \int_{\mathbb{R}} e^{itx} h(x) dx.$$

The notation  $\lesssim$  refers to an inequality up to a multiplicative constant independent of  $n$ .

It is well known that the Fourier transform of  $Z$  and its empirical counterpart may be used to obtain reliable estimations of  $\Phi_X$  when dealing with a standard convolution inverse problem (see [Fan91]) at the price of an assumption on the deconvolution operator translated in  $\Phi_U$ . Moreover, let us note that our model includes the situation where  $p = 0$ , which turns our atomic deconvolution problem into the standard deconvolution problem. Therefore, it is expected that our estimation problem can be solved at the minimal price of some standard smoothness and invertibility conditions on the Fourier transform of  $U$  and some smoothness assumptions on  $f$ . Indeed, these kind of assumptions are well known in the inverse problem literature and commonly used to derive convergence rates of estimators (see among many references the work *e.g.* [FK02] where this assumption is used in its great generality, or [CK02] where this assumption is specified in the Poisson inverse problem situation). We also refer to [BG10, BCG12] for other applications of this kind of assumptions with different “deconvolution + mixture problems”. We are therefore driven to introduce two sets of densities  $\mathbf{H}_s(R)$  and  $\mathbf{H}_\nu^+$ :

- The set  $\mathbf{H}_s(R)$  denotes the set of densities  $f$  that belong to the Sobolev space of regularity  $s$  (and radius  $R$ ) described with the help of the associated characteristic functions  $\Phi_f$  such that:

$$\mathbf{H}_s(R) := \left\{ h \in \mathbb{L}^2(\mathbb{R}) : \int_{\mathbb{R}} |\Phi_h(t)|^2 (1 + |t|^{2s}) dt \leq R^2 \right\}$$

Below, the density of  $X$  denoted by  $f$  is assumed to belong to  $\mathbf{H}_s(R)$  with an *unknown* parameter  $s$ , meaning that we assume that  $\int_{\mathbb{R}} |\Phi_X(t)|^2 (1 + |t|^{2s}) dt \leq R^2$ .

- The set  $\mathbf{H}_\nu^+$  denotes the set of densities such that the Fourier transform satisfies the smoothness and “invertibility” condition :  $\exists (d_0, d_1) : 0 < d_0 < d_1$  and  $d_0 |t|^{-\nu} \leq |\Phi(t)| \leq d_1 |t|^{-\nu}$  as  $|t| \rightarrow +\infty$ .

$$\mathbf{H}_\nu^+ := \{ h \in \mathbb{L}^2(\mathbb{R}) : \exists 0 < d_0 < d_1 \quad d_0 |t|^{-\nu} \leq |\Phi_h(t)| \leq d_1 |t|^{-\nu} \}.$$

Below, we will assume that the density  $g$  of the random variable  $U$  belongs to  $\mathbf{H}_\nu^+$ .

Let us briefly comment on this last assumption. In [Fan91],  $\mathbf{H}_\nu^+$  refers to the set of ordinary smooth densities of order  $\nu$ , which includes many distributions as gamma, double exponential distributions. Note that it would be possible to address the super-smooth case with an exponential decrease of the Fourier transform to handle Gaussian or Cauchy densities. Nevertheless, we have chosen in this work to restrict our study to the ordinary smooth case for the sake of brevity. As pointed in Section 1.2, we assume that  $g$  is known, meaning that  $\Phi_U$  is known on  $\mathbb{R}$ . Such an assumption is reasonable regarding the practical example we want to handle where we can repeat the experiments for the calibration of the cytometer many times.

The estimators introduced in [vEGS08] exploit the knowledge of  $\mathbf{H}_\nu^+$  and of  $\mathbf{H}_s(R)$  (which is more annoying for practical purpose) to obtain an optimal estimator with the important assumption of the knowledge of  $s$  (see Section 2.1).

## 1.5 Main results

In this article we introduce an adaptive procedure to estimate the unknown parameters  $p$  and  $f$ . We consider the estimators  $\hat{p}_n$  and  $\hat{f}_n$  proposed in [vEGS08] based on a kernel estimation using the relationship between the Fourier transform of each random variables because Equation (1) yields:

$$\forall t \in \mathbb{R} \quad \Phi_Z(t) = \Phi_U(t) [p + (1 - p)\Phi_X(t)]. \quad (2)$$

We detail in Section 2.1 and 3.1 the non-adaptive construction proposed by [vEGS08]. We then develop a well-designed strategy (see *e.g.* [Lep92]) to obtain an adaptive estimator of  $p$  and  $f$ . This method allows for choosing among a grid of regularity, the associated bandwidth parameter that realizes the bias-variance tradeoff. The precise construction of the adaptive estimator  $\hat{p}_{n, \hat{j}_n^p}$  and  $\hat{f}_{n, \hat{j}_n^f}$  are given in Section 2.2 and 3.2. The aim of this article is to establish the consistency rate of our adaptive strategy. More precisely, we will prove the following results. The first result concerns the estimation of  $p$ .

**Theorem 1 (Minimax adaptivity of  $p$ )** *Assume that  $f$  belongs to  $\mathbf{H}_s(R)$  and  $g$  to  $\mathbf{H}_\nu^+$  with  $\nu > 1$ , if  $p \in (0, 1)$ , then the estimator  $\hat{p}_{n, \hat{j}_n^p}$  defined in (10) and (13) satisfies:*

$$\mathbb{E} \left( \hat{p}_{n, \hat{j}_n^p} - p \right)^2 \lesssim (\log n) n^{-(2s+1)/(2s+2\nu)}.$$

The proof of Theorem 1 follows a multiple-testing strategy jointly used with a concentration inequality. The additional log term (regarding the minimax rate  $n^{-(2s+1)/(2s+2\nu)}$ ) involved in our upper bound is the price to pay for using a multiple testing strategy and identify the smoothness parameter.

Then we use a plug-in strategy to estimate the function  $f$  from its Fourier transform.

**Theorem 2 (Minimax adaptivity of  $f$ )** *For any  $p \in (0, 1)$ , assume that  $Z$  has a finite variance and assume that  $f \in \mathbf{H}_s(R)$  and  $g \in \mathbf{H}_\nu^+$  with  $\nu > 1$  and select  $\hat{j}_n^f$  as in Equation (29). Then, the estimator  $\hat{f}_{n, \hat{j}_n^f}$  defined in (30) satisfies:*

$$\mathbb{E} \|\hat{f}_{n, \hat{j}_n^f} - f\|_2 \lesssim (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)},$$

where  $a > 0$  is defined in Equation (26).

We emphasize that our estimators produced both for  $p$  and  $f$  induced by the Lepskii rule are non-asymptotic and fully-adaptive with respect to the smoothness parameter  $s$ . Therefore, our results produce an adaptive minimax upper bound up to some log term.

We should at last remark that the previous results require the important knowledge of the convolution operator brought by  $\Phi_U$ . Even if this assumption is legitimate for our intended applications, this may not be the case in other situations. When the operator described by  $\Phi_U$  is unknown and has to be estimated from the data, then the problem falls into the framework of deconvolution with noise in the operator (see [Cav11]). In such a case, it is highly suspected that the attainable rates may be damaged by the preliminary estimation of  $\Phi_U$ . Such a theoretical development is beyond the scope of this work and certainly deserves some careful derivations to understand how the noise on the eigenvalues of the “deconvolution operator” is propagated (see *e.g.* [CH05] and [CR07]).

The article is organized as follows: Section 2 details first the non-adaptive and then the adaptive estimation of the contamination parameter  $p$  while Section 3 proposes to solve similar estimation problems for  $f$ . We present numerical simulations in Section 4 on both simulated and real data-sets and the interest of our results for biological purposes.

## 2 Estimation of the contamination rate $1 - p$

We first consider the problem of estimating the contamination rate  $1 - p$ . We recall the non-adaptive results obtained by [vEGS08] and expose our adaptive strategy afterwards.

### 2.1 Non adaptive estimation of $p$

We describe below the estimators proposed by [vEGS08], that will be used to obtain our adaptive procedure. A key observation relies on Equation (2). This last equation makes it possible, from  $n$  independent observations  $(Z_i)_{1 \leq i \leq n}$ , to obtain an estimation of  $\Phi_X$  and  $p$ . Given this set of observations, the first step consists in the introduction of the empirical estimator of the Fourier transform of  $Z$ :

$$\forall t \in \mathbb{R} \quad \hat{\Phi}_{Z,n}(t) := \frac{1}{n} \sum_{i=1}^n e^{itZ_i}.$$

A key relationship between  $p$  and the several Fourier transforms is given by (2). Since  $f$  belongs to  $\mathbf{H}_s(\mathbb{R})$ , then  $\lim_{|t| \rightarrow +\infty} \Phi_X(t) = 0$ . Therefore, we expect to recover  $p$  by using the information brought by  $\Phi_Z$  and  $\Phi_U$  at large frequencies. In particular, Equation (2) implies the identifiability of the model and provides some insights for an estimation strategy. Second, we may use the knowledge of  $\Phi_U$  and in particular the fact that the density of  $U$  belongs to  $\mathbf{H}_\nu^+$ , which entails some lower bounds of  $\Phi_U$  for large frequencies.

We use the construction of [GvES11] and introduce a smooth real valued kernel  $\Phi_K$  in the Fourier domain that satisfies:

$$\Phi_K(t) \neq 0 \iff t \in [-1; 1] \quad \text{and} \quad \int_{-1}^1 \Phi_K(t) dt = 2, \quad (3)$$

and a flatness condition on the neighborhood of 0 ( $\mathcal{V}_k$  below denotes an open neighborhood of 0):

$$\forall k \in \mathbb{R}_+, \quad \exists C_k > 0, \quad \exists \mathcal{V}_k \quad \forall t \in \mathcal{V}_k : \quad |\Phi_K(t)| \leq C_k |t|^k.$$

This last condition is not restrictive and is satisfied when  $\Phi_K$  is chosen for example as:

$$\Phi_K(t) := \frac{e^{-a|t|^{-m}}}{C_{a,m}} \mathbf{1}_{|t| \leq 1},$$

for any  $a > 0$ ,  $m \geq 1$  ( $C_{a,m}$  is the normalizing constant associated to Equation (3)). Note that this local condition may be replaced by a global one because of  $\Phi_K$  is bounded on  $[-1; 1]$ . Therefore, we keep the notation  $C_k$  and assume that:

$$\forall k \in \mathbb{R}_+, \quad \exists C_k > 0 \quad \forall t \in [-1; 1] : \quad |\Phi_K(t)| \leq C_k |t|^k. \quad (4)$$

Following the works [Fan91, GvES11], we use the kernel  $\Phi_K$  on the Fourier transform  $p + (1 - p)\Phi_X$  and obtain:

$$\frac{h}{2} \int_{\mathbb{R}} \Phi_K(ht) \frac{\Phi_Z(t)}{\Phi_U(t)} dt = \frac{h}{2} \int_{\mathbb{R}} \Phi_K(ht) [p + (1 - p)\Phi_X(t)] dt = p + \frac{h(1 - p)}{2} \int_{-h^{-1}}^{h^{-1}} \Phi_K(ht) \Phi_X(t) dt.$$

The last term of the r.h.s. vanishes when  $h \rightarrow 0$ . Moreover, it is possible to obtain a tight upper bound in terms of  $h$  of this *bias* term:

$$\begin{aligned} \left| h \int_{-h^{-1}}^{h^{-1}} \Phi_K(ht) \Phi_X(t) dt \right| &= \left| h \int_{-h^{-1}}^{h^{-1}} \Phi_K(ht) |t|^{-s} \Phi_X(t) |t|^s dt \right| \\ &\leq h \sqrt{\int_{-h^{-1}}^{h^{-1}} |\Phi_K(ht)|^2 |t|^{-2s} dt} \int_{\mathbb{R}} |\Phi_X(t)|^2 |t|^{2s} dt \leq \sqrt{2} C_s R h^{s+1/2}, \quad (5) \end{aligned}$$

where we applied the Cauchy-Schwarz inequality and then Inequality (4) with  $k = s$  and the fact that  $f \in \mathbf{H}_s(R)$ . Therefore, we can write that:

$$\lim_{h \rightarrow 0} \frac{h}{2} \int_{\mathbb{R}} \Phi_K(ht) \frac{\Phi_Z(t)}{\Phi_U(t)} dt = p.$$

We can plug Equation (2) in the limit above and then define a natural estimator of  $p$  (that will depend on a small bandwidth parameter  $h_n$ ):

$$\hat{p}_n := \frac{h_n}{2} \int_{-h_n^{-1}}^{h_n^{-1}} \Phi_K(h_n t) \frac{\hat{\Phi}_{Z,n}(t)}{\Phi_U(t)} dt. \quad (6)$$

According to (5), we can compute an upper-bound of the bias of  $\hat{p}_n$  as:

$$[\mathbb{E}\hat{p}_n - p]^2 \leq 2C_s^2 R^2 h_n^{2s+1}.$$

The variance of  $\hat{p}_n$  is handled in a standard way following the arguments of [Fan91] with  $\nu > 1$ :

$$\mathbb{V}(\hat{p}_n) \lesssim \frac{h_n^2}{n} \int_{\mathbb{R}} \left| \int_{\mathbb{R}} e^{itz} \frac{\Phi_K(h_n t)}{\Phi_U(t)} dt \right|^2 dz \lesssim \frac{h_n}{n} \int_{\mathbb{R}} \left| \frac{\Phi_K(t)}{\Phi_U(th_n^{-1})} \right|^2 dt \lesssim \frac{1}{nh_n^{2\nu-1}}. \quad (7)$$

Now, a classical optimization of the bias-variance tradeoff yields the optimal (non-adaptive) choice for the bandwidth parameter  $h_n^* := n^{-\frac{1}{2s+2\nu}}$ . It can be shown that this choice leads to the minimax optimal consistency rate  $n^{-(2s+1)/(2s+2\nu)}$ . In other words, we have obtained:

**Theorem 3 (Theorem 1(i) and 5 in [GvES11])** *Assume that  $f \in \mathbf{H}_s(R)$  and  $g \in \mathbf{H}_\nu^+$  with  $\nu > 1$ , then the choice  $h_n = n^{-\frac{1}{2s+2\nu}}$  in (6) leads to an estimator  $\hat{p}_n$  that satisfies the consistency rate:*

$$\mathbb{E}[|\hat{p}_n - p|^2] \leq C(s, R) n^{-\frac{(2s+1)}{2s+2\nu}},$$

where  $C(s, R)$  is a positive constant that continuously depend of  $s$  and  $R$ . Moreover, this estimator is minimax optimal under the additional assumption that  $|\phi'(t)|(1 + |t|^\nu) \leq d_2$  for all  $t \in \mathbb{R}$  and  $d_2 > 0$ .

This result highly depends on the knowledge of  $s$  and a bad strategy for the choice of  $s$  and  $h_n$  will significantly arms the estimation procedure. Nevertheless it is our starting point to produce an adaptive estimator of  $p$ .

## 2.2 Adaptive estimation of $p$

We propose to adapt the Lepskii strategy (see *e.g.* [Lep92]) to obtain an adaptive estimator of  $p$ .

**Grid on  $s$**  We define a grid  $\mathcal{S}_n$  and will estimate  $s$  with an exploration of the possible values in  $\mathcal{S}_n$ . We consider an interval  $[0, s_0]$  where we assume that  $s \in [0, s_0]$  and define a regularly spaced sequence

$$s_{max} = s_0 > s_1 > \dots > s_{k_n} \geq 0,$$

such that for every  $0 \leq j \leq k_n$ ,

$$s_j := s_0 - j \frac{\varepsilon}{\log n}. \quad (8)$$

In the last formula,  $\varepsilon > 0$  is a fixed parameter. This parameter will be calibrated later on and will permit to obtain good estimation properties. For each  $0 \leq j \leq k_n$ , we associate the bandwidth parameter  $h_{n,j}$  that corresponds to the optimal bandwidth parameter chosen above when the smoothness of  $f$  is known and equal to  $s_j$ . This rule yields:

$$h_{n,j} := n^{-1/(2s_j+2\nu)}. \quad (9)$$

The corresponding estimator is then denoted by:

$$\hat{p}_{n,j} := 0 \vee \left( \frac{h_{n,j}}{2} \int_{-h_{n,j}^{-1}}^{h_{n,j}^{-1}} \Phi_K(h_n t) \frac{\hat{\Phi}_{Z,n}(t)}{\Phi_U(t)} dt \right) \wedge 1. \quad (10)$$

We choose to constraint the estimator to obtain a value in  $[0, 1]$ , which will not damage the properties of the estimator since  $p \in [0, 1]$ .

The associated minimax risk in the smoothness class  $\mathcal{H}_{s_j}(R)$  is of the order  $n^{-(s_j+1/2)/(2s_j+2\nu)}$  (see Theorem 3). Given a positive parameter  $\beta$ , we introduce a *penalty* term defined by this minimax risk up to a log term:

$$\kappa_{n,j} := \beta \sqrt{\log nn} n^{-(s_j+1/2)/(2s_j+2\nu)}, \quad (11)$$

We should note the important monotonic variations of the quantities defined above: when  $j > l$  then  $s_j < s_l$  and the bandwidth and penalty parameters satisfy  $h_{n,j} < h_{n,l}$  and  $\kappa_{n,j} > \kappa_{n,l}$ . We introduce the notation  $s^*$  for the closest element to the regularity  $s$  from below in the grid  $\mathcal{G}_n$ :

$$s^* := \sup \{s_j \in \mathcal{G}_n : s_j \leq s\}. \quad (12)$$

For the sake of convenience, we denote by  $j^*$  the integer in  $\{0, \dots, k_n\}$  such that  $s^* = s_{j^*}$ . This integer deterministically depends on  $n$  and  $\epsilon$ , but we omit this dependence to alleviate the notations. The calibration of the grid yields

$$0 \leq s - s^* \leq \frac{\epsilon}{\log n}.$$

**Model selection with the Lepskii rule** To construct the adaptive estimator, we use the following decision rule driven by the bias variance decomposition of each estimator of  $\hat{p}_{n,j}$ :

$$\hat{j}_n^p := \inf \{0 \leq j \leq k_n, \quad \forall l > j : |\hat{p}_{n,j} - \hat{p}_{n,l}| < \kappa_{n,l}\}. \quad (13)$$

The decision rule defined by (13) is costly from a computational point of view. However, alternative coding strategies inspired from [Kat99] make it possible to efficiently implement (13). These strategies rely on the construction of confidence intervals for the sequence  $(\hat{p}_{n,j})_j$ . We will see in Section 4 that if such a strategy works for the computation of a scalar paramter as  $p$ ; it is unfortunately no more possible for the estimation of a density.

The objective of the next paragraphs is to establish the consistency rate of this adaptive strategy, whose performance is described by Theorem 1. We first detail the Lepskii strategy, than give a concentration result and conclude.

### 2.2.1 Analysis of the Lepskii rule

We begin by the statement of two preliminary results. The first one links the average performance of  $\hat{p}_{n,\hat{j}_n^p}$  with some deviations inequalities.

**Proposition 4** *Assume that  $f \in \mathbf{H}_s(R)$  and  $g$  is a known density in  $\mathbf{H}_\nu^+$ . Then the estimator  $\hat{p}_{n,\hat{j}_n^p}$  defined with (10) and (13) satisfies*

$$\mathbb{E} |\hat{p}_{n,\hat{j}_n^p} - p|^2 \leq (\beta^2 \log(n) + C) e^{2\epsilon/(2s+2\nu)} n^{-(2s+1)/(2s+2\nu)} + \sum_{l \geq j^*} \mathbb{P} \left( |\hat{p}_{n,l} - p| \geq \frac{\kappa_{n,l}}{2} \right).$$

#### Proof

We can decompose the estimation error as

$$\mathbb{E} \left( \left| \hat{p}_{n,\hat{j}_n^p} - p \right|^2 \right) = \mathbb{E} \left( \left| \hat{p}_{n,\hat{j}_n^p} - p \right|^2 \mathbf{1}_{\hat{j}_n^p \leq j^*} + \left| \hat{p}_{n,\hat{j}_n^p} - p \right|^2 \mathbf{1}_{\hat{j}_n^p > j^*} \right)$$

- We first consider the event  $\{\hat{j}_n^p \leq j^*\}$  and apply the inequality  $(a+b)^2 \leq 2a^2 + 2b^2$  to obtain:

$$\left| \hat{p}_{n,\hat{j}_n^p} - p \right|^2 \leq 2 \left| \hat{p}_{n,\hat{j}_n^p} - \hat{p}_{n,j^*} \right|^2 + 2 \left| \hat{p}_{n,j^*} - p \right|^2. \quad (14)$$

We study the first term of the r.h.s. of (14). On the event  $\{\hat{j}_n^p < j^*\}$ , the definition (13) of  $\hat{j}_n^p$  yields:

$$\left| \hat{p}_{n,\hat{j}_n^p} - \hat{p}_{n,j^*} \right| < \kappa_{n,j^*},$$

and obviously when  $\hat{j}_n^p = j^*$  the upper bound above also holds. Hence, we deduce that:

$$\mathbb{1}_{\hat{j}_n^p \leq j^*} |\hat{p}_{n, \hat{j}_n^p} - \hat{p}_{n, j^*}|^2 < \{\kappa_{n, j^*}\}^2.$$

Moreover, writing  $s^* = s + (s^* - s)$  in the definition of the penalty (11) leads to:

$$\begin{aligned} \kappa_{n, j^*} &= \beta \sqrt{\log(n)} n^{-(s^*+1/2)/(2s^*+2\nu)} \\ &= \beta \sqrt{\log(n)} n^{-(s+1/2)/(2s+2\nu)} \exp \left[ \left( \frac{s+1/2}{(2s+2\nu)} - \frac{s^*+1/2}{(2s^*+2\nu)} \right) \log n \right] \\ &\leq \beta \sqrt{\log(n)} n^{-(s+1/2)/(2s+2\nu)} \exp \left[ \frac{2\nu \epsilon \{\log n\}^{-1}}{(2s^*+2\nu)(2s+2\nu)} \log n \right] \\ &\leq \beta \exp \left[ \frac{\epsilon}{2s+2\nu} \right] \sqrt{\log(n)} n^{-(s+1/2)/(2s+2\nu)}. \end{aligned}$$

We can handle the second term of the r.h.s. of (14) easily: the nonadaptive estimator obtained with  $j^*$  satisfies the upper bound obtained in Theorem 3, so that:

$$\begin{aligned} \mathbb{E}(|\hat{p}_{n, j^*} - p|^2) &\leq C n^{-(2s^*+1)/(2s^*+2\nu)} \\ &\leq C \exp \left( \frac{2\epsilon}{2s+2\nu} \right) n^{-(2s+1)/(2s+2\nu)}, \end{aligned}$$

where  $C$  is a positive constant independent from  $s$ ,  $n$ , and  $\epsilon$  (the constant  $C$  only depends on  $R$  and  $s_0$ ). Then,

$$\mathbb{E}(\mathbb{1}_{\hat{j}_n^p \leq j^*} |\hat{p}_{n, \hat{j}_n^p} - p|^2) \leq (\beta^2 \log(n) + C) \exp \left( \frac{2\epsilon}{2s+2\nu} \right) n^{-(2s+1)/(2s+2\nu)}. \quad (15)$$

- We consider now the complementary event  $\{\hat{j}_n^p > j^*\}$  and first remark that

$$\hat{j}_n^p > j^* \iff \forall j \leq j^* \exists l > j : |\hat{p}_{n, j} - \hat{p}_{n, l}| \geq \kappa_{n, l}.$$

Therefore, we deduce that

$$\begin{aligned} \{\hat{j}_n^p > j^*\} &= \bigcap_{j \leq j^*} \bigcup_{l > j} \{|\hat{p}_{n, j} - \hat{p}_{n, l}| \geq \kappa_{n, l}\} \\ &\subset \bigcup_{l > j^*} \{|\hat{p}_{n, j^*} - \hat{p}_{n, l}| \geq \kappa_{n, l}\} \\ &\subset \bigcup_{l > j^*} \left\{ |\hat{p}_{n, l} - p| \geq \frac{\kappa_{n, l}}{2} \right\} \bigcup_{l > j^*} \left\{ |\hat{p}_{n, j^*} - p| \geq \frac{\kappa_{n, l}}{2} \right\}. \end{aligned} \quad (16)$$

where the second line comes from the inclusion  $\cup_{j \leq j^*} \{\cdot\} \subset \cup_{j=j^*} \{\cdot\}$  and the last line from the triangle inequality. We have seen that the map  $l \mapsto \kappa_{n, l}$  is increasing then when  $l > j^*$ , we have  $\kappa_{n, l} \geq \kappa_{n, j^*}$  so that

$$\bigcup_{l > j^*} \left\{ |\hat{p}_{n, j^*} - p| \geq \frac{\kappa_{n, l}}{2} \right\} \subset \left\{ |\hat{p}_{n, j^*} - p| \geq \frac{\kappa_{n, j^*}}{2} \right\}.$$

Using this last inclusion in the second union of the right hand side of (16), we finally deduce that

$$\{\hat{j}_n^p > j^*\} \subset \bigcup_{l \geq j^*} \left\{ |\hat{p}_{n, l} - p| \geq \frac{\kappa_{n, l}}{2} \right\}.$$

From the obvious upper-bound  $|\hat{p}_{n, \hat{j}_n^p} - p|^2 \leq 1$ , we can conclude that:

$$\mathbb{E}(|\hat{p}_{n, \hat{j}_n^p} - p|^2 \mathbb{1}_{\hat{j}_n^p > j^*}) \leq \mathbb{E}(\mathbb{1}_{\hat{j}_n^p > j^*}) \leq \sum_{l \geq j^*} \mathbb{P} \left( |\hat{p}_{n, l} - p| \geq \frac{\kappa_{n, l}}{2} \right). \quad (17)$$

Gathering (15) and (17) leads to the conclusion.  $\square$

## 2.2.2 Concentration inequality

The adaptivity property of the estimator  $\hat{p}_{n, \hat{j}_n^p}$  will be deduced from the multiple-testing strategy induced by Proposition 10. A baseline property to successively apply this approach will be a derivation of a concentration inequality on  $\hat{p}_{n,l} - p$  for each fixed  $l$ . Next proposition states that such a concentration inequality holds.

**Proposition 5** *Let  $f \in \mathbf{H}_s(R)$ ,  $g$  a known density in  $\mathbf{H}_\nu^+$  with  $\nu > 1/2$ , and  $l \geq j^*$ , then*

$$\mathbb{P}\left(|\hat{p}_{n,l} - p| \geq \frac{\kappa_{n,l}}{2}\right) \leq 2n^{-\beta^2/64}.$$

### Proof

The proof is divided into two steps.

*Step 1: reduction to a concentration inequality.* For any  $l \geq j^*$  and  $n \in \mathbb{N}$ , we write

$$\begin{aligned} \hat{p}_{n,l} - p &= \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\hat{\Phi}_{Z,n}(t)}{\Phi_U(t)} dt - p \\ &= \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\hat{\Phi}_{Z,n}(t) - \Phi_Z(t)}{\Phi_U(t)} dt + \frac{h_{n,l}^{-1}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\Phi_Z(t)}{\Phi_U(t)} dt - p \\ &= \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\hat{\Phi}_{Z,n}(t) - \Phi_Z(t)}{\Phi_U(t)} dt + \frac{(1-p)h_{n,l}^{-1}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \Phi_X(t) dt. \end{aligned}$$

The triangle inequality and the bias upper bound provided by Equation (5) yield:

$$|\hat{p}_{n,l} - p| \leq \left| \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\hat{\Phi}_{Z,n}(t) - \Phi_Z(t)}{\Phi_U(t)} dt \right| + C_s R h_{n,l}^{s+1/2}.$$

The r.h.s. of the upper bound is of the order  $h_{n,l}^{s+1/2} = n^{-(s+1/2)/(2s_l+2\nu)}$  while  $\kappa_{n,l} = \beta \sqrt{\log nn}^{-(s_l+1/2)/(2s_l+2\nu)}$ . From the definition of the grid on  $s$ ,  $l \geq j^*$  implies that  $s_l \leq s_{j^*}$  so that  $s_l \leq s$ . Therefore, we can write:

$$h_{n,l}^{s+1/2} \{\kappa_{n,l}\}^{-1} \lesssim \{\log n\}^{-1/2} n^{(s_l-s)/(2s_l+2\nu)} \lesssim \frac{1}{\sqrt{\log n}}.$$

We then deduce that a sufficiently large  $n$  exists such that for  $l \geq j^*$ , we have  $C_s R h_{n,l}^{s+1/2} < \kappa_{n,l}/4$  and

$$\begin{aligned} \mathbb{P}\left(|\hat{p}_{n,l} - p| \geq \frac{\kappa_{n,l}}{2}\right) &\leq \mathbb{P}\left(\left| \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{\hat{\Phi}_{Z,n}(t) - \Phi_Z(t)}{\Phi_U(t)} dt \right| \geq \frac{\kappa_{n,l}}{4}\right) \\ &\leq \mathbb{P}\left(\left| \underbrace{\frac{1}{n} \sum_{k=1}^n \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{e^{iZ_k t} - \Phi_Z(t)}{\Phi_U(t)} dt}_{:=\xi_{k,l}} \right| \geq \frac{\kappa_{n,l}}{4}\right). \end{aligned} \quad (18)$$

*Step 2: application of the Bernstein concentration inequality.* To handle the r.h.s. of Inequality (18), we will apply the Bernstein inequality (see Theorem 2.9 of [BLM13] and exercise 2.8 therein, a precise statement is given in Theorem 11 of our Appendix A) to the random variables

$$\xi_{k,l} = \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \Phi_K(h_{n,l}t) \frac{e^{iZ_k t} - \Phi_Z(t)}{\Phi_U(t)} dt.$$

We know that the i.i.d. random variables  $(\xi_{k,l})_{1 \leq k \leq n}$  are centered and satisfy an almost sure bound:

$$\forall k \in \{1 \dots n\} \quad |\xi_{k,l}| \leq \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}} \frac{|\Phi_K(h_{n,l}t)|}{|\Phi_U(t)|} dt.$$

We now use the assumption that  $g \in \mathbf{H}_\nu^+$  to deduce that a  $d_0$  exists such that  $\Phi_U(t)(1+|t|)^\nu \geq d_0$  for all  $t \in \mathbb{R}$ . Therefore, a large enough constant  $C$  exists such that for any  $k \in \{1 \dots n\}$ :

$$\begin{aligned} |\xi_{k,l}| &\leq \frac{h_{n,l}}{2} \int_{-h_{n,l}^{-1}}^{h_{n,l}^{-1}} \frac{|\Phi_K(h_{n,l}t)|(1+|t|)^\nu}{|\Phi_U(t)|(1+|t|)^\nu} dt \\ &\leq \frac{h_{n,l}}{2d_0} \int_{-h_{n,l}^{-1}}^{h_{n,l}^{-1}} |\Phi_K(h_{n,l}t)|(1+|t|)^\nu dt \\ &\leq (1+|h_{n,l}^{-1}|)^\nu \frac{h_{n,l}}{2d_0} \int_{-h_{n,l}^{-1}}^{h_{n,l}^{-1}} |\Phi_K(h_{n,l}t)| dt \leq Ch_{n,l}^{-\nu}. \end{aligned}$$

We can also bound the variance of each  $\xi_{k,l}$  and a computation similar to Equation (7) leads to  $\mathbb{V}(\xi_{k,l}) \leq h_{n,l}^{-2\nu+1}$ . We then apply Theorem 11 with  $b = Ch_{n,l}^{-\nu}$  and  $v = nh_{n,l}^{-2\nu+1}$  and deduce that:

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{k=1}^n \xi_{k,l} \right| \geq \frac{\kappa_{n,l}}{4} \right) \leq 2 \exp \left( - \frac{n\kappa_{n,l}^2}{32 \left( h_{n,l}^{-2\nu+1} + Ch_{n,l}^{-\nu} \kappa_{n,l} / 3 \right)} \right)$$

Now, remark that  $h_{n,l}^{-2\nu+1} = n^{(2\nu-1)/(2\nu+2s_l)}$  while  $h_{n,l}^{-\nu} \kappa_{n,l} = \beta n^{\nu/(2\nu+2s_l)} n^{-(s_l+1/2)/(2\nu+2s_l)} \sqrt{\log n}$ . Since we assumed that  $\nu > 1/2$ , the main contribution in the denominator of the exponential bound above is brought by  $h_{n,l}^{-2\nu+1}$ . We therefore conclude that a large enough  $n$  exists such that:

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{k=1}^n \xi_{k,l} \right| \geq \frac{\kappa_{n,l}}{4} \right) \leq 2 \exp \left( - \frac{n\kappa_{n,l}^2}{64h_{n,l}^{-2\nu+1}} \right) = 2n^{-\beta^2/64},$$

because the penalty term  $\kappa_{n,l}$  has been designed so that  $\frac{n\kappa_{n,l}^2}{h_{n,l}^{-2\nu+1}} = \beta^2 \log(n)$ .  $\square$

### 2.2.3 Proof of Theorem 1

We can achieve the proof of Theorem 1 and establish the adaptation of our estimator  $\hat{p}_{n,\hat{j}_n^p}$ : Propositions 4 and 5 yield

$$\mathbb{E} |\hat{p}_{n,\hat{j}_n^p} - p|^2 \leq (\beta^2 \log(n) + C) \exp \left( \frac{2\varepsilon}{2s+2\nu} \right) n^{-(2s+1)/(2s+2\nu)} + 2 \sum_{l \geq j^*} n^{-\beta^2/64}.$$

The size of the grid is bounded from above by  $\varepsilon^{-1} s_{max} \log n$ . Therefore, for  $n$  large enough

$$\mathbb{E} |\hat{p}_{n,\hat{j}_n^p} - p|^2 \leq (\beta^2 + 1) e^{\frac{2\varepsilon}{2s+2\nu}} (\log n) n^{-(2s+1)/(2s+2\nu)} + \varepsilon^{-1} s_{max} \log n n^{-\beta^2/64}.$$

It remains to choose the constant  $\varepsilon$  and  $\beta$ . We first pick  $\beta$  such that

$$n^{-\beta^2/64} \leq n^{-(2s+1)/(2s+2\nu)}, \quad (19)$$

then we choose  $\varepsilon$  in order to minimize the constant, regardless the value of  $s$ ,

$$(\beta^2 + 1) e^{\frac{2\varepsilon}{2s+2\nu}} + s_{max} \varepsilon^{-1}. \quad (20)$$

These choices yield the desired adaptive property (optimal up to a log term):

$$\mathbb{E} |\hat{p}_{n,\hat{j}_n^p} - p|^2 \lesssim (\log n) n^{-(2s+1)/(2s+2\nu)}.$$

$\square$

**Remark 6** For numerical experiments, the calibration of the parameters  $\varepsilon$  and  $\beta$  given by (19) and (20) is crucial. Since the value of  $s$  is unknown, these inequality have to be true for all  $s \in [0, s_{max}] = [0, s_0]$ . We are driven to an optimal choice of the form

$$\beta^* = 8 \sqrt{\frac{2s_{max} + 1}{2s_{max} + 2\nu}} \text{ and } \varepsilon^* = \operatorname{argmin} \left\{ ((\beta^*)^2 + 1) e^{\frac{\varepsilon}{2s+2\nu}} + s_{max} \varepsilon^{-1} \right\}.$$

### 3 Estimation of $f$

We keep the same presentation and first briefly describe the estimator proposed by [GvES11]. We describe our adaptive procedure.

#### 3.1 Non adaptive approach

The estimation of  $f$  is highly similar to the one of  $p$  and relies on the Fourier transform of  $U$  and of the empirical data. It exploits the relationship:

$$\Phi_X(t) = \frac{\Phi_Z(t) - p\Phi_U(t)}{(1-p)\Phi_U(t)},$$

and therefore uses a plug-in estimator of  $p$  and of  $\Phi_Z$  with another kernel smoothing strategy. Following [GvES11], we introduce a second kernel  $Q$  whose Fourier transform is denoted by  $\Phi_Q$  has its support in  $[-1, 1]$  and satisfies the following assumptions:

$$\Phi_Q(0) = 1, \quad \forall t \in \mathbb{R} \quad |\Phi_Q(t) - 1| \leq M|t|^s, \quad \int_{-1}^1 |\Phi_Q(t)|^2 dt < \infty, \quad (21)$$

where  $M > 0$ . It is worth saying that the choice of the kernel used for this deconvolution is important regarding the numerical results as well as the theoretical ones. For our purpose, we only handle ordinary smooth inverse problem in Sobolev spaces. In that case, it will be enough to handle a very simple kernel  $\Phi_Q$ , given by the sinc function:

$$\Phi_Q(t) := \mathbf{1}_{[-1;1]} \quad \text{so that} \quad \forall x \in \mathbb{R} \quad Q(x) = \frac{\sin x}{\pi x} \quad \text{with} \quad Q(0) = \frac{1}{\pi}.$$

It is immediate to check that this kernel automatically matches the requirement given in Equation (21) on the smoothness of  $\Phi_Q$  around 0. We refer to [DH06] for a detailed discussion on the influence of the kernel choice from a numerical point of view, and to [CL13] and [GL11] for deeper insights on functional spaces where the estimation is done (*e.g.* in anisotropic Nikol'skii classes). The plug-in strategy proposed in [GvES11] reads as follows:

$$\hat{\Phi}_{X,n} = \frac{\hat{\Phi}_{Z,n}(t) - \check{p}_n \Phi_U(t)}{(1 - \check{p}_n) \Phi_U(t)} \Phi_Q(\delta_n t), \quad (22)$$

where  $\check{p}_n$  is a preliminary truncated estimation of  $p$  that is plugged into Equation 22, with the following rule:

$$\check{p}_n := \hat{p}_n \wedge 1 - \tau_n. \quad (23)$$

The truncation step under  $1 - \tau_n$  makes it possible to avoid numerical instability when  $\hat{p}_n$  is estimated close to 1. It is important to remark that for a theoretical purpose, it is necessary to build a preliminary estimator of  $p$  with  $\check{p}_n$  independent from the estimated  $\Phi_Z$  in (22). This can be achieved using  $\lfloor n/2 \rfloor$  samples to compute  $\check{p}_n$  and the remaining ones to build  $\hat{\Phi}_{X,n}$  according to our plug-in strategy. We can now state the minimax consistency result of [GvES11], which is obtained with a similar bias-variance tradeoff.

**Theorem 7 (Theorem 2(i) and 3(i) in [GvES11])** *Assume  $f \in \mathbf{H}_s(\mathbb{R})$  and  $g \in \mathbf{H}_\nu^+$  with  $\nu > 1$  and set  $\tau_n = \log(3n)^{-1}$ . Then the choice  $h_n = n^{-\frac{1}{2s+2\nu}}$  in (6) and  $\delta_n = n^{-1/(2s+2\nu+1)}$  in (22) leads to an estimator  $\hat{f}_n$  that satisfies the consistency rate:*

$$\mathbb{E}[\|\hat{f}_n - f\|^2] \lesssim n^{-\frac{(2s)}{2s+2\nu+1}}.$$

It is shown in [GvES11] that such estimator is minimax optimal under the additional assumptions that  $s \geq 1/2$  and  $|\phi'(t)|(1 + |t|^\nu) \leq d_2$  for all  $t \in \mathbb{R}$  and  $d_2 > 0$ . In particular, we note that the consistency rate corresponds to the standard rate of ordinary smooth deconvolution inverse problem (see *e.g.* [Fan91]). Again, we stress the fact that the proposed estimator highly depends on  $s$  through  $\delta_n = n^{-1/(2s+2\nu+1)}$ , which is unknown in practice and should be estimated. The next paragraph describes our proposed strategy.

### 3.2 Adaptive estimation of $f$

**Grid on  $s$**  The adaptation follows the same strategy as the one for  $p$ , even though significantly harder from a theoretical point of view. We still use the same grid  $\mathcal{S}_n$ :  $s_0 > s_1 > \dots > s_{k_n} \geq 0$ , such that for every  $0 \leq j \leq k_n$ ,  $s_j = s_0 - j \frac{\varepsilon}{\log n}$ , and the notation  $h_{n,j}$  for the bandwidth parameter associated to the optimal estimation of  $p$  in  $\mathbf{H}_s(R)$  when  $s = s_j$ , *i.e.* we use  $h_{n,j} := n^{-1/(2s_j+2\nu)}$ . A second bandwidth parameter associated to the estimation of  $f$  is defined by  $\delta_{n,j}, \forall 0 \leq j \leq k_n$ :

$$\delta_{n,j} := n^{-1/(2s_j+2\nu+1)}. \quad (24)$$

With these two bandwidth parameters derived from a regularity  $s_j$ , we build a preliminary estimator of  $p$  with the observations  $(Z_i)_{1 \leq i \leq \lfloor n/2 \rfloor}$ :

$$\hat{p}_{n,j}^{(1)} = 0 \vee \left\{ \left\{ \frac{h_{n,j}}{2} \int_{-h_{n,j}^{-1}}^{h_{n,j}^{-1}} \Phi_K(h_{n,j}t) \frac{\hat{\Phi}_{Z,n}^{(1)}(t)}{\Phi_U(t)} dt \right\} \wedge 1 - \tau_n \right\} \quad (25)$$

where  $\hat{\Phi}_{Z,n}^{(1)}(t)$  and  $\hat{\Phi}_{Z,n}^{(2)}(t)$  are two unbiased *independent* estimates of  $\Phi_Z(t)$ :

$$\forall t \in \mathbb{R} \quad \hat{\Phi}_{Z,n}^{(1)}(t) := \frac{1}{\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} e^{itZ_k} \quad \text{and} \quad \hat{\Phi}_{Z,n}^{(2)}(t) := \frac{1}{n-1-\lfloor n/2 \rfloor} \sum_{k=\lfloor n/2 \rfloor+1}^n e^{itZ_k}.$$

Theorem 7 entails for a good choice of the truncation parameter introduced in Equation (23) and we choose

$$\tau_n := \log(n)^{-a}, \quad (26)$$

for  $a > 0$ . Then, we use  $\hat{p}_{n,j}^{(1)}$  to obtain an estimator of the empirical Fourier transforms of  $X$  by:

$$\hat{\Phi}_{X_{n,j}}^{(2)} := \frac{\hat{\Phi}_{Z,n}^{(2)}(t) - \hat{p}_{n,j}^{(1)} \Phi_U(t)}{(1 - \hat{p}_{n,j}^{(1)}) \Phi_U(t)} \Phi_Q(\delta_{n,j}t) \quad (27)$$

The application of the Lepskii method relies on a penalty term, which is defined as

$$\rho_{n,l} := \beta(\log n)^{a+1/2} n^{-s_l/(2s_l+2\nu+1)}. \quad (28)$$

This penalty is the minimax risk of estimation of  $f$  in  $\mathbf{H}_s(R)$  when  $s_l = s$ , up to a log term (see Theorem 7). We recall that for the estimation of  $p$ , the penalty was associated with a supplementary  $(\log n)^{1/2}$  term. However, such a penalty is not strong enough to obtain a good concentration inequality for the estimator of  $f$  (see Proposition 10 below).

**Lepski's rule for the estimation of  $f$**  We use the penalization introduced in (28) to define:

$$\hat{j}_n^f := \inf\{0 \leq j \leq k_n, \quad \forall l > j : \quad \|\hat{f}_{n,j} - \hat{f}_{n,l}\|_2 < \rho_{n,l}\}.$$

Thanks to the Parseval identity, this can be also formulated in terms of the Fourier transforms:

$$\hat{j}_n^f := \inf\{0 \leq j \leq k_n, \quad \forall l > j : \quad \|\hat{\Phi}_{X_{n,j}}^{(2)} - \hat{\Phi}_{X_{n,l}}^{(2)}\|_2 < \rho_{n,l}\}. \quad (29)$$

The estimator derived from the penalization above is then written as  $\hat{\Phi}_{X_{n,\hat{j}_n^f}}^{(2)}$  where we used the selection of  $\hat{j}_n^f$  in (29) with the definition of  $\hat{\Phi}_{X_{n,\hat{j}_n^f}}^{(2)}$  given in (27). It leads to the estimation of  $f$  itself using the Fourier reconstruction formula:

$$\hat{f}_{n,\hat{j}_n^f}(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \hat{\Phi}_{X_{n,\hat{j}_n^f}}^{(2)}(t) dt. \quad (30)$$

Again, we start with the statement of a proposition that links the average performance of  $\hat{\Phi}_{X_{n,\hat{j}_n^f}}^{(2)}$  with a family of deviation inequalities.

**Proposition 8** Let  $f \in \mathbf{H}_s(R)$  and  $g$  a known density in  $\mathbf{H}_\nu^+$ , then the estimator  $\hat{f}_{n,\hat{j}_n^f}$  satisfies:

$$\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \lesssim (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} + \sqrt{\sum_{l \geq j^*} \mathbb{P}\left(\|\hat{f}_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2}\right)}.$$

**Proof**

The proof is close to the one of Proposition 4. We denote by  $s^* = s_{j^*}$  the closest element from  $s$  from below in the grid  $\mathcal{S}_n$  (see Equation (12)). Following the same guidelines, we obtain that:

$$\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 = \mathbb{E}\left(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f \leq j^*} + \|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f > j^*}\right).$$

On the event  $\{\hat{j}_n^f \leq l^*\}$ : we first apply the triangle inequality

$$\mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f \leq j^*}) \leq \mathbb{E}\left[\left(\|\hat{f}_{n,\hat{j}_n^f} - \hat{f}_{j^*}\|_2 + \|\hat{f}_{j^*} - f\|_2\right) \mathbf{1}_{\hat{j}_n^f \leq j^*}\right]$$

On  $\hat{j}_n^f \leq j^*$ , the first term is upper bounded by

$$\begin{aligned} \rho_{n,j^*} &= \beta(\log n)^{a+1/2} n^{-s_{j^*}/(2s_{j^*}+2\nu+1)} \\ &= \beta(\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} \exp\left[\left(\frac{s}{2s+2\nu+1} - \frac{s_{j^*}}{2s_{j^*}+2\nu+1}\right) \log n\right] \\ &\leq \beta(\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} \exp\left(\frac{\varepsilon}{2s+2\nu+1}\right). \end{aligned}$$

For the second term, we apply the results of the non adaptive estimator in Theorem 7:

$$\mathbb{E}\left(\|\hat{f}_{j^*} - f\|_2 \mathbf{1}_{\hat{j}_n^f \leq j^*}\right) \leq C n^{-s_{j^*}/(2s_{j^*}+2\nu+1)},$$

and deduce that:

$$\mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f \leq j^*}) \leq \left(\beta(\log n)^{a+1/2} + C\right) n^{-s/(2s+2\nu+1)} \exp\left(\frac{\varepsilon}{2s+2\nu+1}\right). \quad (31)$$

where  $C$  is a positive constant independent from  $\beta$ ,  $\varepsilon$  and  $n$ .

On the event  $\{\hat{j}_n^f > l^*\}$ : On this event, the triangle inequality yields  $\|\hat{f}_{n,l} - \hat{f}_{n,j}\|_2 \leq \|\hat{f}_{n,l} - f\|_2 + \|\hat{f}_{n,j} - f\|_2$ , so that

$$\{\hat{j}_n^f > j^*\} \subset \bigcup_{l \geq j^*} \left\{ \|\hat{f}_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2} \right\}.$$

Therefore, the Cauchy-Schwarz inequality and the inclusion above yields

$$\begin{aligned} \mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f > j^*}) &\leq \sqrt{\mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2^2)} \sqrt{\mathbb{E}(\mathbf{1}_{\hat{j}_n^f > j^*})} \\ &\leq \sqrt{\sum_{j=0}^{k_n} \mathbb{E}\left(\|\hat{f}_{n,j} - f\|_2^2\right)} \sqrt{\mathbb{E}\left(\mathbf{1}_{\bigcup_{l \geq j^*} \left\{ \|\hat{f}_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2} \right\}}\right)} \end{aligned}$$

We now use a refinement of Theorem 7 given by Proposition 12 in the appendix Section A to bound  $\mathbb{E}(\|\hat{f}_{n,j} - f\|_2^2)$ :

$$\mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f > j^*}) \leq \sqrt{\sum_{j=0}^{k_n} \Phi(s, R) n^{-2 \min(s_j, s)/(2s_j+2\nu+1)}} \sqrt{\mathbb{E}\left(\mathbf{1}_{\bigcup_{l \geq j^*} \left\{ \|\hat{f}_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2} \right\}}\right)}$$

Since  $k_n \leq s_0 \epsilon^{-1} \log(n)$  and  $\Phi$  is continuous in  $s$ , then  $C = \{\sup_{s \in [0, s_0]} \Phi(s, R)\}^{1/2} \sqrt{s_0 \epsilon^{-1}}$  is a finite constant. Setting now  $\alpha = s_0 \epsilon^{-1}$ , we shall now study the next sum:

$$\sum_{j=0}^{k_n} n^{-s_j/(s_j+\nu+1/2)} = \sum_{j=0}^{\alpha \log(n)} n^{-\frac{j\epsilon/\log(n)}{j\epsilon/\log(n)+\nu+1/2}}.$$

This sum may be divided into two parts: one with  $s_j \leq 1$  and the other part with smoothness  $s_j \geq 1$ . If  $j\epsilon/\log(n) \geq 1$ , then we use the rough bound  $s_j/(s_j + \nu + 1/2) \geq 1/(\nu + 3/2)$  and obtain that:

$$\sum_{j=\epsilon^{-1} \log(n)}^{k_n} n^{-s_j/(s_j+\nu+1/2)} \leq \alpha \log(n) n^{-1/(\nu+3/2)},$$

which is a bounded sequence. Now, if  $j\epsilon/\log(n) \leq 1$ , we have:

$$\sum_{j=0}^{\epsilon^{-1} \log(n)} n^{-\frac{j\epsilon/\log(n)}{j\epsilon/\log(n)+\nu+1/2}} \leq \sum_{j=0}^{\epsilon^{-1} \log(n)} e^{-\frac{j\epsilon}{j\epsilon/\log(n)+\nu+1/2}} \leq \sum_{j=0}^{\epsilon^{-1} \log(n)} e^{-\frac{\epsilon}{\nu+3/2}j} \leq \frac{1}{1 - e^{-\frac{\epsilon}{3/2+\nu}}}.$$

Therefore, a constant  $C(\epsilon, s_0, R)$  exists such that:

$$\mathbb{E}(\|\hat{f}_{n, \hat{j}_n^f} - f\|_2 \mathbf{1}_{\hat{j}_n^f > j^*}) \leq C(\epsilon, s_0, R) \sqrt{\sum_{l \geq j^*} \mathbb{P}\left(\|f_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2}\right)}. \quad (32)$$

The conclusion follows from (31) and (32).  $\square$

### 3.2.1 Bias upper bound

Once again, the main difficulty of the proof is to obtain a deviation inequality on each event:

$$\Omega_{n,l} := \left\{ \|\hat{f}_{n,l} - f\|_2 \geq \frac{\rho_{n,l}}{2} \right\} = \left\{ \|\hat{\Phi}_{X_{n,l}}^{(2)} - \Phi_X\|_2^2 \geq \frac{\rho_{n,l}^2}{4} \right\}.$$

where the last equality comes from the Parseval identity. We begin by the following statement:

**Proposition 9** *Assume  $l \geq j^*$ , for  $n$  large enough one has*

$$\Omega_{n,l} \subset \left\{ \|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2^2 \geq \frac{\rho_{n,l}^2}{8} \right\}.$$

#### Proof

The main ingredient of the proof uses the bias-variance decomposition of the squared  $\mathbb{L}^2$  norm. Hence, we define

$$T_1 := \|\mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)} - \Phi_X\|_2^2 \quad \text{and} \quad T_2 := \mathbb{V}(\hat{\Phi}_{X_{n,l}}^{(2)}) = \mathbb{E}\|\mathbb{E}(\hat{\Phi}_{X_{n,l}}^{(2)}) - \hat{\Phi}_{X_{n,l}}^{(2)}\|_2^2. \quad (33)$$

A key remark comes from the fact that  $\hat{\Phi}_{Z,n}^{(2)}(t)$  is built with a sample of  $n - \lfloor n/2 \rfloor$  observations, that are independent of the one used to estimate  $p$  with  $\hat{p}_{n,l}^{(1)}$ . We can write that:

$$\begin{aligned} \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}(t) &= \mathbb{E}\left[\frac{p\Phi_U(t) + (1-p)\Phi_U(t)\Phi_X(t) - \hat{p}_{n,l}^{(1)}\Phi_U(t)}{(1 - \hat{p}_{n,l}^{(1)})\Phi_U(t)}\Phi_Q(\delta_{n,l}t)\right] \\ &= \mathbb{E}\left[\frac{p - \hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}}\right]\Phi_Q(\delta_{n,l}t) + \left[\mathbb{E}\left[\frac{1-p}{1 - \hat{p}_{n,l}^{(1)}}\right]\Phi_Q(\delta_{n,l}t) - 1\right]\Phi_X(t) + \Phi_X(t) \\ &= \mathbb{E}\left[\frac{p - \hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}}\right]\Phi_Q(\delta_{n,l}t)(1 - \Phi_X(t)) + [\Phi_Q(\delta_{n,l}t) - 1]\Phi_X(t) + \Phi_X(t). \end{aligned}$$

Using  $(a+b)^2 \leq 2a^2 + 2b^2$ , the bias term  $T_1$  defined in (33) is upper bounded by

$$T_1 \leq 2 \underbrace{\left\{ \mathbb{E} \left[ \frac{p - \hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right]^2 \right\}}_{:=T_{1,1}} \int_{\mathbb{R}} |\Phi_Q(\delta_{n,l}t)(1 - \Phi_X(t))|^2 dt + 2 \underbrace{\int_{\mathbb{R}} |[\Phi_Q(\delta_{n,l}t) - 1]\Phi_X(t)|^2 dt}_{:=T_{1,2}},$$

We may apply the Cauchy-Schwarz inequality and Lemma 13 of the appendix, Section A and obtain:

$$\left\{ \mathbb{E} \left[ \frac{p - \hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right]^2 \right\} \leq \mathbb{E} \left[ \frac{p - \hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right]^2 \lesssim h_{n,l}^{2s_l+1}.$$

The function  $|\Phi_X|$  is bounded by 1 so that

$$T_{1,1} \lesssim h_{n,l}^{2s_l+1} \int_{\mathbb{R}} |\Phi_Q(\delta_{n,l}t)(1 - \Phi_X(t))|^2 dt \lesssim h_{n,l}^{2s_l+1} \int_{\mathbb{R}} |\Phi_Q(\delta_{n,l}t)|^2 dt \lesssim h_{n,l}^{2s_l+1} \delta_{n,l}^{-1} \|\Phi_Q\|_2^2.$$

Moreover, for a given smoothness parameter  $s_l$ , the estimation of  $p$  is “easier” (faster) than the estimation of  $f$ , which is translated by  $h_{n,l} = o(\delta_{n,l})$ . We conclude that:

$$T_{1,1} \lesssim h_{n,l}^{2s_l} \sim n^{-2s_l/(2s_l+2\nu)} = o(\rho_{n,l}^2).$$

The upper bound of the second term  $T_{1,2}$  uses the smoothness assumption on  $f \in \mathbf{H}_s(R)$  and the construction of the kernel  $Q$ : applying Equation (21) with  $\alpha = s$ , we obtain:

$$T_{1,2} \leq 2 \int_{\mathbb{R}} |\delta_{n,l}t|^{2s} |\Phi_X(t)|^2 dt \lesssim \delta_{n,l}^{2s} = \mathcal{O}\left(n^{-2s/(2s_l+2\nu+1)}\right).$$

Since  $l \geq j^*$ , we deduce that  $s_l \leq s$  and the additional log term in the definition of  $\rho_{n,l}$  in (28) permits to conclude that

$$T_{1,2} \leq \delta_{n,l}^{2s_l} = o(\rho_{n,l}^2).$$

Therefore,  $T_1$  is smaller than  $\rho_{n,l}^2/8$  for  $n$  large enough.

We can conclude the proof using the triangle inequality

$$\|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2 \geq \|\hat{\Phi}_{X_{n,l}}^{(2)} - \Phi_X\|_2 - \|\mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)} - \Phi_X\|_2.$$

Hence

$$\|\hat{\Phi}_{X_{n,l}}^{(2)} - \Phi_X\|_2 \geq \frac{\rho_{n,l}}{2} \implies \|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2 \geq \frac{\rho_{n,l}}{2} - \sqrt{T_1} \geq \frac{\rho_{n,l}}{2\sqrt{2}}.$$

□

### 3.2.2 Concentration inequality

From Proposition 9, we can see that the Lepskii rule will perform well if we succeed in bounding the probability

$$\mathbb{P}\left(\|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2^2 \geq \frac{\rho_{n,l}^2}{8}\right).$$

For a given  $t \in \mathbb{R}$ , we can rewrite this expression as

$$\begin{aligned} \hat{\Phi}_{X_{n,l}}^{(2)}(t) - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}(t) &= \Phi_Q(\delta_{n,l}t) \left[ \frac{\hat{\Phi}_{Z,n}^{(2)}(t) - \hat{p}_{n,l}^{(1)}\Phi_U(t)}{(1 - \hat{p}_{n,l}^{(1)})\Phi_U(t)} \right] - \Phi_Q(\delta_{n,l}t)\mathbb{E} \left[ \frac{\hat{\Phi}_{Z,n}^{(2)}(t) - \hat{p}_{n,l}^{(1)}\Phi_U(t)}{(1 - \hat{p}_{n,l}^{(1)})\Phi_U(t)} \right] \\ &= \frac{\Phi_Q(\delta_{n,l}t)}{\Phi_U(t)} \left[ \frac{\hat{\Phi}_{Z,n}^{(2)}(t) - \hat{p}_{n,l}^{(1)}\Phi_U(t)}{1 - \hat{p}_{n,l}^{(1)}} \right] - \frac{\Phi_Q(\delta_{n,l}t)}{\Phi_U(t)} \Phi_Z(t)\mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] + \Phi_Q(\delta_{n,l}t)\mathbb{E} \left[ \frac{\hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right] \\ &= \frac{\Phi_Q(\delta_{n,l}t)}{\Phi_U(t)} \left[ \frac{\hat{\Phi}_{Z,n}^{(2)}(t)}{1 - \hat{p}_{n,l}^{(1)}} - \Phi_Z(t)\mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right] - \Phi_Q(\delta_{n,l}t) \left[ \frac{\hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{\hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right] \right] \\ &= T_{2,1} + T_{2,2} - T_{2,3}, \end{aligned}$$

where

$$T_{2,1} := \frac{\Phi_Q(\delta_{n,l}t) \hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t)}{\Phi_U(t) (1 - \hat{p}_{n,l}^{(1)})} \quad T_{2,2} := \frac{\Phi_Q(\delta_{n,l}t) \Phi_Z(t)}{\Phi_U(t)} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right],$$

and

$$T_{2,3} := \Phi_Q(\delta_{n,l}t) \left[ \frac{\hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{\hat{p}_{n,l}^{(1)}}{1 - \hat{p}_{n,l}^{(1)}} \right] \right].$$

The terms  $T_{2,2}$  and  $T_{2,3}$  can be upper bounded easily while  $T_{2,1}$  deserves a specific attention.

**Study of  $T_{2,2}$ :** We first remark that  $\mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right]$  is close to  $\frac{1}{1-p}$  using the upper bound:

$$\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \frac{1}{1-p} \right| \leq \frac{|p - \hat{p}_{n,l}^{(1)}|}{1-p} (\log n)^a, \quad (34)$$

where the inequality derives from  $\hat{p}_{n,l}^{(1)} \leq 1 - \tau_n$  and (26). Therefore, applying Theorem 3 with the smoothness parameter  $s_l$  we deduce that

$$\left| \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] - \frac{1}{1-p} \right| \lesssim (\log n)^a n^{-(s_l+1/2)/(2s_l+2\nu)}, \quad (35)$$

Consider a constant  $c > 0$ , the term  $T_{2,2}$  is then handled as follows:

$$\begin{aligned} \mathbb{P}(\|T_{2,2}\|_2 \geq c\rho_{n,l}) &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right| \geq \frac{c\rho_{n,l}}{\|\Phi_Q(\delta_{n,l}\cdot)\Phi_Z\Phi_U^{-1}\|_2}\right) \\ &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right| \geq \frac{c\rho_{n,l}}{\|\Phi_Q(\delta_{n,l}\cdot)(p + (1-p)\Phi_X)\|_2}\right) \\ &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right| \geq \frac{c}{2\|\Phi_Q\|_2} \rho_{n,l} \delta_{n,l}^{1/2}\right), \end{aligned}$$

where we use the upper bound of the  $L_2$  norm:  $\|\Phi_Q(\delta_{n,l}\cdot)(p + (1-p)\Phi_X)\|_2^2 \leq 4 \int_{\mathbb{R}} |\Phi_Q(\delta_{n,l}t)|^2 dt \leq 4\delta_{n,l}^{-1} \|\Phi_Q\|_2^2$ . We now use Equation (35) and obtain:

$$\begin{aligned} \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right| \geq \frac{c}{2\|\Phi_Q\|_2} \rho_{n,l} \delta_{n,l}^{1/2}\right) &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \frac{1}{1-p} \right| + \left| \frac{1}{1-p} - \mathbb{E} \left[ \frac{1}{1 - \hat{p}_{n,l}^{(1)}} \right] \right| \geq \frac{c}{2\|\Phi_Q\|_2} \rho_{n,l} \delta_{n,l}^{1/2}\right) \\ &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \frac{1}{1-p} \right| \geq \frac{c}{2\|\Phi_Q\|_2} \rho_{n,l} \delta_{n,l}^{1/2} - C(\log n)^a n^{-(s_l+1/2)/(2s_l+2\nu)}\right) \\ &\leq \mathbb{P}\left(\left| \frac{1}{1 - \hat{p}_{n,l}^{(1)}} - \frac{1}{1-p} \right| \geq \frac{c}{4\|\Phi_Q\|_2} \rho_{n,l} \delta_{n,l}^{1/2}\right), \end{aligned}$$

for  $n$  large enough because the approximation term involved in the upper bound of Equation (35) is negligible comparing to  $\rho_{n,l} \delta_{n,l}^{1/2}$ . Using again (34), we obtain that for a sufficiently small constant  $\eta$  (independent on  $n$  and  $l$ ):

$$\mathbb{P}(\|T_{2,2}\|_2 \geq c\rho_{n,l}) \leq \mathbb{P}\left(|p - \hat{p}_{n,l}^{(1)}| \geq \eta \rho_{n,l} \delta_{n,l}^{1/2}\right).$$

In order to apply the Bernstein inequality (see Theorem 11 in the appendix Section A) , let us remark first that the penalty  $\kappa_{n,l}$  defined by (11) satisfies

$$\kappa_{n,l} = o\left(\rho_{n,l}\delta_{n,l}^{1/2}\right).$$

Since for  $n$  large enough we have  $p \leq 1 - \tau_n$  , the truncated estimator satisfies

$$\|\hat{p}_{n,l}^{(1)} - p\| \leq \|\hat{p}_{n/2,l} - p\|$$

where  $\hat{p}_{n,l}^{(1)}$  is given by (25). Thus we can conclude that

$$\mathbb{P}(\|T_{2,2}\|_2 \geq c\rho_{n,l}) \lesssim n^{-\beta^2/64}. \quad (36)$$

**Study of  $T_{2,3}$ :** This study is of the same nature because  $\mathbb{E}\left[\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}}\right]$  is close to  $\frac{p}{1-p}$ : Inequality (35) reads:

$$\left|\mathbb{E}\left[\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}}\right] - \frac{p}{1-p}\right| = \left|\mathbb{E}\left[\frac{p - \hat{p}_{n,l}^{(1)}}{(1-p)(1-\hat{p}_{n,l}^{(1)})}\right]\right| \lesssim (\log n)^a n^{-(s_l+1/2)/(2s_l+2\nu)},$$

Hence

$$\begin{aligned} \mathbb{P}(\|T_{2,3}\|_2 \geq c\rho_{n,l}) &\leq \mathbb{P}\left(\left|\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}} - \mathbb{E}\left[\frac{p}{1-\hat{p}_{n,l}^{(1)}}\right]\right| \geq \frac{c\rho_{n,l}}{2\|\Phi_Q(\delta_{n,l}\cdot)\|_2}\right) \\ &\leq \mathbb{P}\left(\left|\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}} - \mathbb{E}\left[\frac{p}{1-\hat{p}_{n,l}^{(1)}}\right]\right| \geq \frac{c}{2\|\Phi_Q\|_2}\rho_{n,l}\delta_{n,l}^{1/2}\right), \end{aligned}$$

We use again Equation (35) and obtain that:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}} - \mathbb{E}\left[\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}}\right]\right| \geq \frac{c}{2\|\Phi_Q\|_2}\rho_{n,l}\delta_{n,l}^{1/2}\right) \\ \leq \mathbb{P}\left(\left|\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}} - \frac{p}{1-p}\right| + \left|\frac{p}{1-p} - \mathbb{E}\left[\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}}\right]\right| \geq \frac{c}{2\|\Phi_Q\|_2}\rho_{n,l}\delta_{n,l}^{1/2}\right) \\ \leq \mathbb{P}\left(\left|\frac{\hat{p}_{n,l}^{(1)}}{1-\hat{p}_{n,l}^{(1)}} - \frac{p}{1-p}\right| \geq \frac{c}{4\|\Phi_Q\|_2}\rho_{n,l}\delta_{n,l}^{1/2}\right), \end{aligned}$$

and similar arguments used for  $T_{2,2}$  yields:

$$\mathbb{P}(\|T_{2,3}\|_2 \geq c\rho_{n,l}) \lesssim n^{-\beta^2/64}. \quad (37)$$

**Study of  $T_{2,1}$ :** We recall that the support of  $\Phi_Q$  is  $[-1; 1]$ , which implies that

$$\|T_{2,1}\|_2^2 = \int_{-\delta_{n,l}^{-1}}^{\delta_{n,l}^{-1}} \frac{|\Phi_Q(\delta_{n,l}t)|^2}{|\Phi_U(t)|^2} \frac{|\hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t)|^2}{(1-\hat{p}_{n,l}^{(1)})^2} dt.$$

We now define  $M_{n,l}$  as the supremum of the empirical process  $\hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t)$  when  $t$  varies between  $-\delta_{n,l}^{-1}$  and  $\delta_{n,l}^{-1}$ :

$$M_{n,l} := \sup_{|t| \leq \delta_{n,l}^{-1}} \left| \hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t) \right|,$$

and write

$$\|T_{2,1}\|_2 \leq M_{n,l} \times \frac{1}{(1-\hat{p}_{n,l}^{(1)})} \times \left\| \frac{\Phi_Q(\delta_{n,l}t)}{\Phi_U(t)} \right\|_2.$$

We then deduce that

$$\begin{aligned} \mathbb{P}(\|T_{2,1}\|_2 \geq \rho_{n,l}) &\leq \mathbb{P}\left(\left\|\frac{\Phi_Q(\delta_{n,l}t)}{\Phi_U(t)}\right\|_2 \frac{M_{n,l}}{1 - \hat{p}_{n,l}^{(1)}} \geq \rho_{n,l}\right) \\ &\leq \mathbb{P}\left(\|\Phi_Q(\delta_{n,l}\cdot)\Phi_U(\cdot)^{-1}\|_2 M_{n,l} \geq \tau_n \rho_{n,l}\right). \end{aligned}$$

To upper bound the  $\mathbb{L}^2$  norm of  $\Phi_Q(\delta_{n,l}\cdot)\Phi_U^{-1}$ , we follow a standard argument of [Fan91] using that  $g \in \mathbf{H}_\nu^+$  and deduce that

$$\|\Phi_Q(\delta_{n,l}\cdot)\Phi_U(\cdot)^{-1}\|_2^2 = \int_{\mathbb{R}} \frac{|\Phi_Q(\delta_{n,l}t)|^2}{|\Phi_U(t)|^2} dt \leq \frac{1}{\delta_{n,l}} \int_{-1}^1 \frac{|\Phi_Q(\xi)|^2}{|\Phi_U(\xi\delta_{n,l}^{-1})|^2} d\xi \leq \frac{\|\Phi_Q\|_2^2}{d_2^{-2}} \delta_{n,l}^{-(2\nu+1)}.$$

This inequality yields

$$\mathbb{P}(\|T_{2,1}\|_2 \geq \rho_{n,l}) \leq \mathbb{P}\left(M_{n,l} \geq \frac{d_2}{\|\Phi_Q\|_2} \tau_n \rho_{n,l} \delta_{n,l}^{1/2+\nu}\right). \quad (38)$$

It remains to obtain a deviation inequality on  $M_{n,l}$ , which is the supremum norm of  $\hat{\Phi}_{Z,n}^{(2)} - \Phi_Z$  on the interval  $[-\delta_{n,l}^{-1}, \delta_{n,l}^{-1}]$ . For this purpose, we could try to use the roadmap of applying the Talagrand inequality associated to a chaining strategy (see [BLM13] for example) but these general theorems will not lead to a satisfactory deviation bound because of our assumption on  $U$ , which is not a sub-Gaussian random variable.

Nevertheless, it is possible to exploit the feature of the process  $(\hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t))_{-\delta_{n,l}^{-1} \leq t \leq \delta_{n,l}^{-1}}$ , which is an average of centered bounded random processes. We handle a discretization of the interval  $[-\delta_{n,l}^{-1}, \delta_{n,l}^{-1}]$  and define  $t \mapsto W_n(t)$  by:

$$\forall t \in \mathbb{R} : \quad W_n(t) = \hat{\Phi}_{Z,n}^{(2)}(t) - \Phi_Z(t).$$

We introduce a parameter  $\alpha > 0$  that will be chosen small enough below. The control of  $M_{n,l}$  relies on the simple remark that:

$$W_n(t+u) - W_n(t) = \frac{1}{n - n' + 1} \sum_{j=n'}^n [e^{itZ_j} - e^{i(t+u)Z_j}] - \mathbb{E}[e^{itZ} - e^{i(t+u)Z}],$$

where we set  $n' = \lceil n/2 \rceil + 1$ . The elementary remark  $|e^{ia} - e^{ib}| \leq |a - b|$  for all  $(a, b) \in \mathbb{R}^2$  yields

$$|W_n(t+u) - W_n(t)| \leq |u| \left( \frac{1}{n - n' + 1} \sum_{j=n'}^n |Z_j| + \mathbb{E}|Z| \right). \quad (39)$$

In particular, if we define the event

$$\Omega_n := \left\{ \frac{1}{n - n' + 1} \sum_{j=n'}^n |Z_j| < 2\mathbb{E}|Z| \right\}.$$

The Tchebychev inequality applied to the  $\mathbb{L}^2$ -random variables  $(Z_j)_{n' \leq j \leq n}$  shows that:

$$\mathbb{P}(\Omega_n^c) \leq \frac{\text{Var}(|Z|)}{n\mathbb{E}[|Z|]^2} \leq \frac{c}{n},$$

with  $c = \text{Var}(|Z|) \{\mathbb{E}[|Z|]\}^{-2}$ . Obviously,  $\mathbb{E}[|Z|] > 0$  because  $Z = U + AX$  and  $Z$  cannot be a.s. 0 because  $U$  is assumed to satisfy  $\mathbf{H}_\nu^+$ . The important point on  $\Omega_n$  is that Inequality (39) does not depend on  $t$ , so that  $\forall s > 0$ :

$$\begin{aligned} \mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}, |u| \leq \alpha} |W_n(t+u) - W_n(t)| > s\right) &\leq \mathbb{P}\left(\alpha \left(\frac{1}{n-n'} \sum_{j=n'}^n |Z_j| + \mathbb{E}|Z|\right) > s \cap \Omega_n\right) + \mathbb{P}(\Omega_n^c) \\ &\leq \mathbf{1}_{3\alpha\mathbb{E}[|Z|] > s} + \frac{c}{n} \end{aligned} \quad (40)$$

We now define the threshold  $s_{n,l}$  and the window size  $\alpha_{n,l}$  by:

$$s_{n,l} := \frac{1}{2} \frac{d_2}{\|\Phi_Q\|_2} \tau_n \rho_{n,l} \delta_{n,l}^{1/2+\nu} \quad \text{and} \quad \alpha_{n,l} := \frac{s_{n,l}}{3\mathbb{E}[|Z|]}. \quad (41)$$

We are naturally driven to handle a discrete grid  $\mathcal{T}_{n,l}$  regularly spaced from  $-\delta_{n,l}^{-1}$  to  $\delta_{n,l}^{-1}$  whose step-size is  $\alpha_{n,l}$ . We have

$$|\mathcal{T}_{n,l}| \leq \frac{2\delta_{n,l}^{-1}}{\alpha_{n,l}} = \frac{6\mathbb{E}[|Z|]\delta_{n,l}^{-1}}{s_{n,l}} \quad \text{and} \quad \forall t \in [-\delta_{n,l}^{-1}, \delta_{n,l}^{-1}], \quad \exists \tau_t \in \mathcal{T}_{n,l} : \quad |t - \tau_t| \leq \alpha_{n,l}.$$

These settings permit to deduce from (40) that:

$$\mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}, |u| \leq \alpha_{n,l}} |W_n(t+u) - W_n(t)| > s_{n,l}\right) \leq \frac{c}{n}. \quad (42)$$

Now, for each element of the grid  $\tau \in \mathcal{T}_{n,l}$ ,  $W_n(\tau)$  is a mean of  $n$  random variables that whose modulus are bounded by 1. The Hoeffding inequality implies that  $\forall \tau \in \mathcal{T}_{n,l}$ :

$$\begin{aligned} \mathbb{P}(|W_n(\tau)| > s_{n,l}) &\leq \mathbb{P}\left(\left|\frac{1}{n-n'} \sum_{j=n'}^n \cos(\tau Z_j) - \mathbb{E}(\cos(\tau Z))\right| < \frac{s_{n,l}}{2}\right) \\ &\quad + \mathbb{P}\left(\left|\frac{1}{n-n'} \sum_{j=n'}^n \sin(\tau Z_j) - \mathbb{E}(\sin(\tau Z))\right| < \frac{s_{n,l}}{2}\right) \\ &\leq 4 \exp\left(-\frac{(n-n')s_{n,l}^2}{8}\right). \end{aligned} \quad (43)$$

We now produce an upper bound of  $\mathbb{P}(\|T_{2,1}\|_2 \geq \rho_{n,l})$ . Equation (38) yields:

$$\begin{aligned} \mathbb{P}(\|T_{2,1}\|_2 \geq \rho_{n,l}) &= \mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}} |W_n(t)| \geq 2s_{n,l}\right) \\ &\leq \mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}, |u| \leq h_{n,l}} |W_n(t+u) - W_n(t)| \geq s_{n,l} \quad \text{and} \quad \sup_{|t| \leq \delta_{n,l}^{-1}} |W_n(t)| \geq 2s_{n,l}\right) \\ &\quad + \mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}, |u| \leq h_{n,l}} |W_n(t+u) - W_n(t)| \leq s_{n,l} \quad \text{and} \quad \sup_{|t| \leq \delta_{n,l}^{-1}} |W_n(t)| \geq 2s_{n,l}\right) \\ &\leq \mathbb{P}\left(\sup_{|t| \leq \delta_{n,l}^{-1}, |u| \leq h_{n,l}} |W_n(t+u) - W_n(t)| \geq s_{n,l}\right) + \mathbb{P}\left(\sup_{\tau \in \mathcal{T}_{n,l}} |W_n(\tau)| \geq s_{n,l}\right) \\ &\leq \frac{c}{n} + \sum_{\tau \in \mathcal{T}_{n,l}} \mathbb{P}(|W_n(\tau)| \geq s_{n,l}) \\ &\leq \frac{c}{n} + \frac{2\delta_{n,l}^{-1}}{\alpha_{n,l}} \times 4 \exp\left(-\frac{(n-n')s_{n,l}^2}{8}\right). \end{aligned}$$

where we successively applied the triangle inequality, a union bound over  $\mathcal{T}_{n,l}$ , and then inequalities (42) and (43). An immediate computation from (41) (26), (28) and (24) shows that:

$$\begin{aligned}
ns_{n,l}^2 &= n \frac{d_2^2}{4\|\Phi_Q\|_2^2} \tau_n^2 \rho_{n,l}^2 \delta_{n,l}^{1+2\nu} \\
&= n \frac{d_2^2}{4\|\Phi_Q\|_2^2} \times (\log n)^{-2a} \times \beta^2 (\log n)^{2a+1} n^{-2s_l/(2s_l+2\nu+1)} \times n^{-(1+2\nu)/(2s_l+2\nu+1)} \\
&= \frac{\beta^2 d_2^2}{4\|\Phi_Q\|_2^2} \log(n),
\end{aligned}$$

and

$$\delta_{n,l}^{-1} \alpha_{n,l}^{-1} = \frac{6\mathbb{E}(|Z|)\|\Phi_Q\|_2}{d_2} \beta \log(n)^{-1/2} n^{\frac{3/2+\nu+s_l}{2\nu+2s_l+1}}.$$

We therefore deduce that

$$\mathbb{P}(\|T_{2,1}\|_2 \geq \rho_{n,l}) \leq \frac{c}{n} + \beta \log(n)^{-1/2} \frac{48\mathbb{E}(|Z|)\|\Phi_Q\|_2}{d_2} n^{-\frac{\beta^2 d_2^2}{24\|\Phi_Q\|_2^2} + \frac{3/2+\nu+s_l}{2\nu+2s_l+1}}. \quad (44)$$

**Conclusion** To conclude with a concentration estimate of  $\hat{\Phi}_{X_{n,l}}^{(2)}$  it remains to combine (36), (37) and (44) to obtain the following result.

**Proposition 10** *Let  $f \in \mathbf{H}_s(R)$ ,  $g$  a known density in  $\mathbf{H}_\nu^+$  with  $\nu > 1$ , and  $l \geq j^*$ , then*

$$\mathbb{P}\left(\|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2^2 \geq \frac{\rho_{n,l}^2}{8}\right) \lesssim n^{-\frac{\beta^2}{64}} + n^{-1} + \log(n)^{-1/2} n^{-\frac{\beta^2 d_2^2}{24\|\Phi_Q\|_2^2} + \frac{3/2+\nu+s_l}{2\nu+2s_l+1}}.$$

### 3.2.3 Proof of Theorem 2

Let us recall that Proposition 8 and 9 yield:

$$\begin{aligned}
\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 &\lesssim (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} + \sqrt{\sum_{m \geq j^*} \mathbb{P}\left(\|\hat{f}_{n,m} - f\|_2 \geq \frac{\rho_{n,m}}{2}\right)} \\
&\lesssim (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} + \sqrt{\sum_{m \geq j^*} \mathbb{P}\left(\|\hat{\Phi}_{X_{n,l}}^{(2)} - \mathbb{E}\hat{\Phi}_{X_{n,l}}^{(2)}\|_2^2 \geq \frac{\rho_{n,l}^2}{8}\right)}
\end{aligned}$$

From the definition of the grid  $\mathcal{S}_n$  (see the beginning of Section 3.2), it contains a number of order  $s_0 \varepsilon^{-1} \log(n)$  points and for all  $l$ ,  $n^{-\frac{3/2+\nu+s_l}{2s_l+2\nu+1}} \leq n^{-1/2}$ . Then, we conclude from Proposition 10 that:

$$\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \lesssim \beta (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)} + \sqrt{s_0 \varepsilon^{-1} \log n \left( n^{-\frac{\beta^2}{64}} + n^{-1} + (\log n)^{-1/2} n^{-\frac{\beta^2 d_2^2}{24\|\Phi_Q\|_2^2} + \frac{3/2+\nu+s_l}{2\nu+2s_l+1}} \right)}$$

Therefore, for  $\beta$  large enough:

$$\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2 \lesssim (\log n)^{a+1/2} n^{-s/(2s+2\nu+1)},$$

which ends the proof of the almost optimal (up to some log terms) adaptivity of our Lepski's rule.  $\square$

## 4 Numerical experiments

### 4.1 Simulated data

Numerical simulation where implemented in Python. For the simulated random variables  $U, A, X$  such that  $Z = U + AX$ , we have considered both Gamma random variables and we recall below their property. A Gamma distribution function with parameters  $k$  and  $\theta$  (denoted  $\text{Gamma}(k, \theta)$ ) admits a density  $f(x)$  ( $x \in \mathbb{R}$ ) and characteristic function  $\phi(t)$  ( $t \in \mathbb{R}$ ) given by:

$$f(x) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k)\theta^k}, \quad \Phi(t) = (1 - \theta it)^{-k}.$$

We considered two settings :

## First data set

$$U \sim \text{Gamma}(2, 1), \quad A \sim \text{Ber}(0.4), \quad X \sim \text{Gamma}(\gamma, 1),$$

for  $\gamma \in \{4, 6, 8\}$  (see Figure 2). Therefore the smoothness parameter  $s$  satisfies  $s = \gamma$ , this parameter varies while  $\nu$  is kept fixed  $\nu = 2$ .

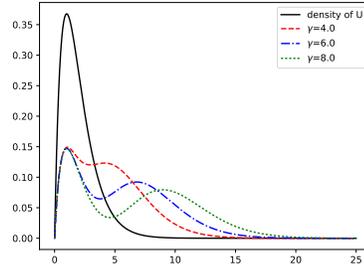


Figure 2: Density of the baseline perturbation  $U$  (in plain black) and of the observed signal  $Z$  for different values of  $\gamma$  (dotted curves).

## Second data set

$$U \sim \text{Gamma}(\gamma, 1), \quad A \sim \text{Ber}(0.4), \quad X \sim \text{Gamma}(2, 1),$$

for  $\gamma \in \{2, 4, 6, 8\}$  Therefore the smoothness parameter  $\nu$  satisfies  $\nu = \gamma$ , this parameter varies while  $s$  is kept fixed  $s = 2$ .

**Results - Estimation of  $p$**  We first focus on the estimation of  $p$ . For each data sets we performed Monte Carlo replications (with 100 repetitions) in order to approach  $\mathbb{E}\left(\|\hat{p}_{n, \hat{j}_n^p} - p\|_2^2\right)$  for different values of the sample size  $n \in \{100, 200, 500, 700, 1000, 2000\}$ . Let us remark that the Lepskii procedure is costly in terms of numerical simulations since the criterion for selection requires to compute all the estimators on a grid. An alternative procedure has thus been proposed by [Kat99] based on the construction of confidence intervals. However, in the setting of the estimation of a density, this construction is no longer possible. Therefore, our simulations are based on the exact Lepskii decision rule given in (13) for  $p$  and (29) for  $f$ . Figure 3 gives in a log – log scale the estimators of  $\mathbb{E}\left(\|\hat{p}_{n, \hat{j}_n^p} - p\|_2^2\right)$  obtained by the Monte Carlo replications as a function of the numbers of observations  $n$ . The left and right sides of Figure 3 correspond respectively to the first and second data sets.

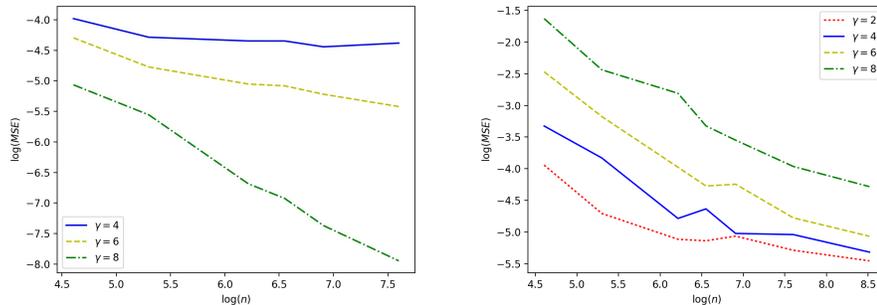


Figure 3: Log-Log plot of  $n \mapsto \mathbb{E}\left(\|\hat{p}_{n, \hat{j}_n^p} - p\|_2^2\right)$  obtained by Monte Carlo replications. Left: first data set for several values of  $s = \gamma$ . Right: second data set for several values of  $\nu = \gamma$ . For these simulations, we choose  $\varepsilon = 0.5$  and  $\beta = 9.0$ .

In both cases, logarithm of the mean squared error decreases linearly with  $\log(n)$ , which is consistent with Theorem 1 and 3. We observe in Figure 3 that the estimation of  $p$  is easier for large values of  $s$  (large  $\gamma$ ) on the left graph, while the estimation gets more difficult as  $\nu$  increases (large values of  $\gamma$ ) on the right graph. This phenomenon is completely consistent with the rates derived in Theorem 1.

**Results - Estimation of  $f$  and  $\mathbb{E}(\|\hat{f}_{n,\hat{j}_n^f} - f\|_2)$**  For the non parametric estimation of  $f$ , we compared the estimation obtained with the Lepskii procedure proposed in this article with a similar one assuming that the mixing parameter  $p$  is known. In Figure 4, we draw the expected Fourier transform  $\phi_X$  and estimators obtained with the Lepskii procedure for different sample sizes  $n \in \{100, 1000, 10000\}$ . The left side corresponds to the complete Lepskii procedure (29) while in the right side the parameter  $p$  is given. In this case, the simulated data correspond to the first data set with  $\gamma = 4$ . On these graphs, we can first see that when the sample size increases, the bandwidth selected in the algorithm is improved since the support of  $\hat{\Phi}_{X_n}$  widens.

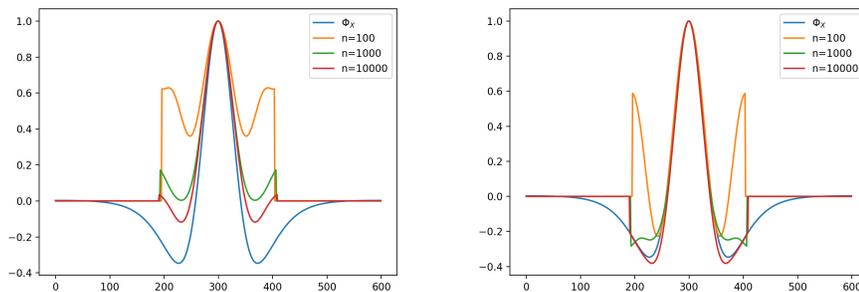


Figure 4: True characteristic function  $\Phi_X$  as well as its estimations counterpart for different sample sizes  $n \in \{100, 1000, 10000\}$ . Left: Lepskii procedure described in Section 3.2. Right: the true value of the mixing parameter  $p$  is assumed to be known. We used data according to the first data set with  $\gamma = 4$  and we set  $\varepsilon = 0.5$  and  $\beta = 9$ .

We shall observe first that the algorithm leads to better results when the mixing parameter is known, which is not surprising since the estimation of  $p$  provides an additional noise in our statistical estimator. As for the estimation of  $p$ , we computed with Monte Carlo replications the mean squared estimation error for different sample sizes (see Figure 5). Once again we compared our algorithm (left side) with the case where  $p$  is known (right side). We find that the estimators are more accurate when  $p$  is known and that the regularity  $s = \gamma$  of the function  $f$  improves the convergence speed.

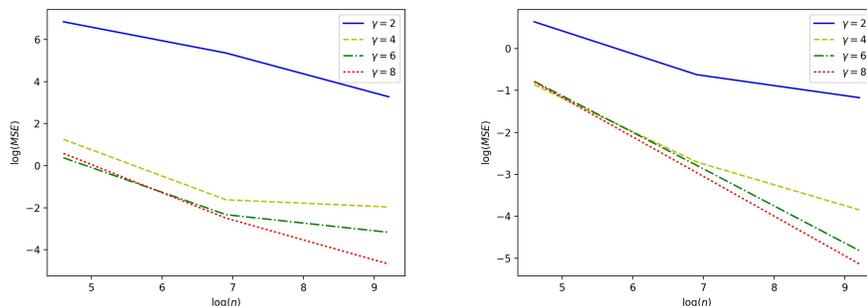


Figure 5: On these graphs, we draw in a log – log scale the estimators of  $\mathbb{E}\|\hat{f}_{n,\hat{j}_n^f} - f\|_2$  obtained by Monte Carlo simulations (for  $n \in \{100, 1000, 10000\}$ ). The left panel corresponds to the case where  $p$  is unknown while on the right panel, the value of  $p$  was given. For these simulations, we used the first data set with different values of  $\gamma$  and chose  $\beta = 9.0$  and  $\varepsilon = 0.5$ .

## 4.2 Real data

**Description of the dataset** The dataset used in this study is coming from fluorescence distribution measured using a flow cytometer instrument (BD Biosciences, LSRII FORTRESSA) on cells obtained from human blood. Lymphocytes are cells of the immune system useful for cancer cell destruction in immunotherapy (among other), they were extracted from the fresh blood of a healthy individual by the biologist and this cell suspension was then used for the experiment. Technically, the cell suspension was split into two parts. One half was left untouched and the baseline photo emission of untreated cells was recorded by the cytometer. The second half was mixed with a fluorescently labelled antibody (reagent) that specifically binds to the CD27 protein. Then, the photo emission by treated cells was again recorded by the cytometer. The amount of reagent binding is reflected by the fluorescence emitted by the cell that is coming only from the antibody treatment (see Figure 1). The biological experiment’s goal was to assess the expression of the molecule CD27 by human lymphocytes. This protein is expressed at the surface of lymphocytes and reflect their activation status. It is therefore used to estimate the functionality of a lymphocyte population. Methods that would help to decipher the mixture of functionalities among a lymphocyte population is thus of major biological interest.

In this setting, we use the estimation procedure developed in our paper to infer the percentage of lymphocytes expressing CD27 on their surface and the conditional probability for a cell to express CD27 conditionally to its fluorescence intensity.

**Preliminary estimation of the distribution of the noise  $U$**  From the recorded fluorescence intensity of untreated cells, we infer the density  $g$  of  $U$  with a preliminary kernel density estimation using a Gaussian kernel through the `scipy.stats.gaussian_kde` Python software. We then use the recorded fluorescence intensity of treated cells as data for the analysis (see Figure 6(a)). In this case, the size of the recorded data is  $n = 12645$ .

Of course, our work only deals with the situation where the density  $g$  is known, which is unfortunately not possible in our biological situation. Therefore, we admit as a reasonable approximation the estimation of  $g$  provided by a preliminary kernel density estimation. We have not treated in our theoretical study the consequences of such a preliminary non-parametric estimation, and we leave this subject open as a future subject of investigation. Such a work would then fall into the field of statistical inverse problems with noise in the operator (see *e.g* [CH05] and the references therein).

**Estimation of  $p$**  For the estimation of the proportion of cells expressing CD27 on their cells’ surface, which corresponds to  $1 - p$ , we use the Lepskii procedure proposed in Section 2.2. To verify the convergence of our algorithm, we use a subsampling strategy of our data set and repeat it on permuted versions of the data.

We observe on Figure 6(b) the good behaviour of our algorithm: it produces a sharp estimation even using the half of the data and it converges to the value  $\hat{p}_{n, \hat{j}_n^p} = 0.31$ . We aim at comparing this value with the measurement classically used in cytometric analysis. Usually biologist use a quantity called *percentage of positive cells* with reflects the percentage of marked cells whose fluorescent intensity is higher than a given threshold. This threshold is computed as the 95th percentile of the density  $g$  (see Figure 6(a)). On this data, the percentage of positive cells is 0.85, which is much larger than the quantity we estimate  $1 - \hat{p}_{n, \hat{j}_n^p} = 0.69$ .

**Estimation of  $f$**  Next, we focus on the estimation of the density  $f$ , which represents here the distribution of fluorescence intensity due to the presence of CD27 on the cells’ surface. We use two different estimation strategies:

- First, we apply the global Lepskii estimation procedure introduced in Section 3.2.
- Second, we use a similar approach but we consider that the value of  $p$  is known and given by the previously estimated  $\hat{p}_{n, \hat{j}_n, p}$  above. We then use the Lepskii procedure with a fixed value of  $p$  as described in Section 3.2.

In both cases we then compute an approximation of the Fourier transform  $\Phi_Z$  of the mixed signal and compared it with the empirical estimator  $\hat{\Phi}_{Z, n}$  in Figure 6(c). In particular, in Figure 6(c), we show several colored curves:

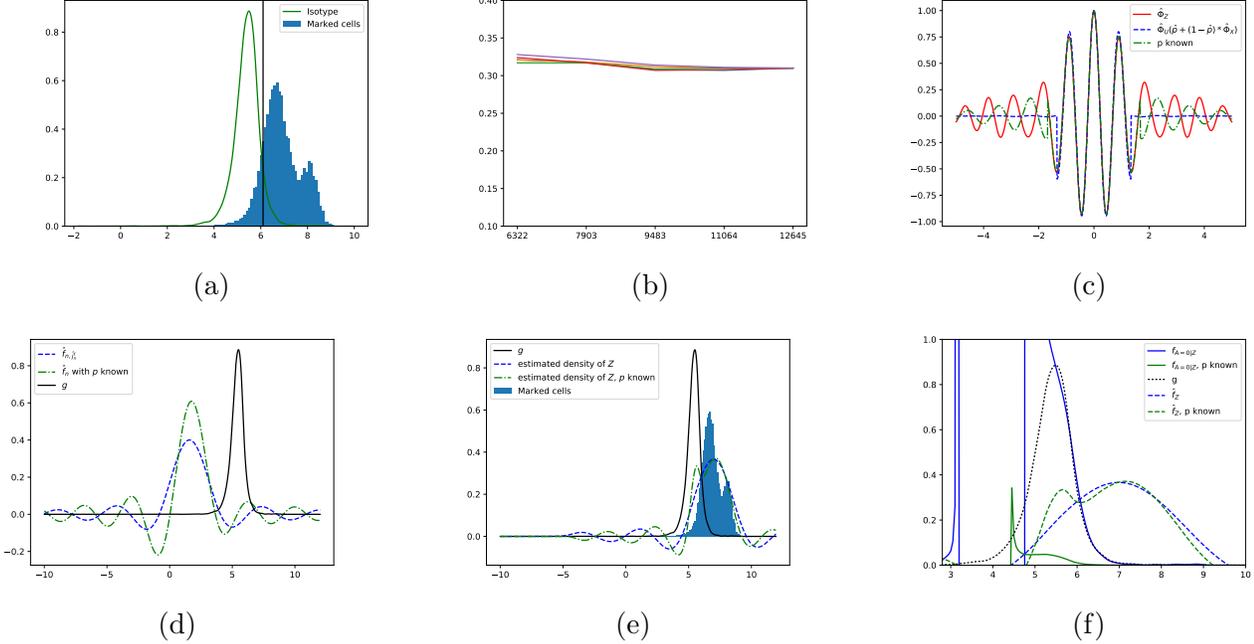


Figure 6: Results of the Lepskii procedure on real data. (a) Histogram of the fluorescence of treated cells and estimated probability distribution for the fluorescence of un-treated cells. (b) Estimation of  $p$  on an increasing sub-sample of the data (the data being permuted 5 times). (c) In red, empirical Fourier transform  $\hat{\Phi}_Z$  and two estimators computed respectively with the values  $\hat{f}_{n, \hat{j}_n^f}$  and  $\hat{p}_{n, \hat{j}_n^f}^{(1)}$  obtained with the Lepskii procedure for  $f$  (in blue) and with the values obtained when  $p$  is first estimated and then used in the estimation of  $f$  (in green). (d) Comparison of the obtained estimators of  $f$ . (e) Estimated density of  $Z$  computed using the two different strategies for estimating  $f$  and comparison with the histogram of data. (f) Estimators of the conditional density  $f_{A=0|Z}(z)$ .

- the red curve is the empirical estimator  $\hat{\Phi}_{Z,n}$
- the blue dotted curve is the estimator derived from the global Lepskii procedure
- the green dotted curve corresponds to the modified Lepskii procedure for  $f$  when  $p$  is previously estimated).

We observe that the estimator of  $f$  computed with  $p = 0.31$  seems to behave better, which is not surprising since the Lepskii procedure for  $f$  does not provide any theoretical guaranty for estimating  $p$  (we obtained  $\hat{p}_{n, \hat{j}_n^f}^{(1)} = 0.081$ , which seems very low and not realistic for our biological framework). Then we calculate the inverse transform of both estimators (see Figure 6(d) ), compute the estimated density of  $Z$  and compare it with the observed histogram (Figure 6(e)). We observe first that our estimation can take negative values (due to the bandwidth parameter in the estimation) and secondly that it fails to capture the exact shape of the density, however it reproduces reasonably well the mean and the variance.

Finally, we study the probability that a cell expresses CD27 on its surface conditionally to the value of the observed fluorescence. We compute estimators of this conditional density with the Bayes formula:

$$f_{A=0|Z}(z) = \frac{pg(z)}{pg(z) + (1-p)f * g(z)}$$

The results are drawn on Figure 6(f) for both estimators of  $f$ . Since our estimation of  $f$  can take negative values and vanish, we observe artificial singularities that should be omitted for the biological interpretation. We have not pushed further our investigations from a biological point of view since the goal of our paper is mainly theoretical.

## A Technical results

### A.1 Bernstein inequality

We first state a classical concentration inequality that may be found for example in [BLM13].

**Theorem 11 (Bernstein inequality)** *Let  $X_1, \dots, X_n$  be independent random variables with finite variance such that  $|X_i| \leq b$  for some  $b > 0$  almost surely for all  $i \leq n$ . Let  $S = \sum_{i=1}^n X_i - \mathbb{E}[X_i]$  and  $v = \sum_{i=1}^n \mathbb{E}[X_i^2]$ . Then*

$$\mathbb{P}(|S| \geq t) \leq 2 \exp\left(-\frac{t^2}{2(v + bt/3)}\right).$$

### A.2 Non adaptive estimation

In this section, we sketch the proof of Theorem [GvES11] with  $\tau_n = \log(n)^{-a}$  and  $a > 0$  instead of the choice  $\tau_n = \log(3n)^{-1}$  given in [GvES11], in order to control  $\mathbb{E}(\|\hat{f}_{n,j_n} - f\|_2)$ .

**Proposition 12** *If  $\mathbf{H}_s(R)$  and  $\mathbf{H}_\nu^+$  hold with  $\nu > 1$ , then the choices  $h_{n,j} = n^{-\frac{1}{2s_j+2\nu}}$  in (6),  $\tau_n = \log(n)^{-a}$  in (26) and  $\delta_{n,j} = n^{-1/(2s_j+2\nu+1)}$  in (24) lead to an estimator  $\hat{f}_{n,j}$  that satisfies the consistency rate:*

$$\mathbb{E}[\|\hat{f}_{n,j} - f\|_2^2] \leq C(s, R) n^{-\frac{2 \min(s, s_j)}{2s_j+2\nu+1}},$$

where  $C(s, R)$  is a positive constant depending continuously in  $s$  and  $R$ .

#### Proof

The bias-variance decomposition and Fubini's Theorem yield

$$\mathbb{E}[\|\hat{f}_{n,j} - f\|_2^2] \leq \int_{-\infty}^{\infty} \left(\mathbb{E}(\hat{f}_{n,j}(t) - f(t))\right)^2 dt + \int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,j}(t)) dt$$

**Integrated bias** In this section, we introduce more explicit notations, we set  $\hat{f}_{n,h_{n,j},\delta_{n,j}} = \hat{f}_{n,j}$  defined by (27) and set  $\hat{f}_{n,\delta_{n,j}}$  the estimator of  $f$  when  $p$  is known :

$$\hat{f}_{n,\delta_{n,j}} = \frac{1}{2\pi} \int e^{-itx} \frac{\hat{\Phi}_{Z,n}^{(2)}(t) - p\Phi_U(t)}{(1-p)\Phi_U(t)} \Phi_Q(\delta_{n,j}t) dt.$$

We also define

$$\tilde{f}_{n,\delta_{n,j}} = \frac{1}{2\pi} \int e^{-itx} \frac{\hat{\Phi}_{Z,n}^{(2)}(t)}{\Phi_U(t)} \Phi_Q(\delta_{n,j}t) dt.$$

Then

$$\begin{aligned} \int \left(\mathbb{E}(\hat{f}_{n,j} - f)\right)^2 &= \int \left(\mathbb{E}(\hat{f}_{n,h_{n,j},\delta_{n,j}} - f)\right)^2 \\ &\leq 2 \int \left(\mathbb{E}(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})\right)^2 + 2 \int \left(\mathbb{E}(\hat{f}_{n,\delta_{n,j}} - f)\right)^2 \end{aligned}$$

Using the Parseval inequality, then assumptions (21) and  $f \in \mathbf{H}_s(R)$  we deduce that the second term behaves like:

$$\begin{aligned} \int \left(\mathbb{E}(\hat{f}_{n,\delta_{n,j}} - f)\right)^2 &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\Phi_f(t)|^2 \frac{|\Phi_Q(\delta_{n,j}t) - 1|^2}{(\delta_{n,j}t)^{2s}} (\delta_{n,j}t)^{2s} dt, \\ &\leq MR \delta_{n,j}^{2s}. \end{aligned}$$

Then, for the first term we obtain that:

$$\begin{aligned}
\int \left( \mathbb{E}(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}}) \right)^2 &\leq 2 \int \mathbb{E} \left[ \tilde{f}_{n,\delta_{n,j}} \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \\
&\quad + 2 \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \int \left( \frac{1}{\delta_{n,j}} Q(t\delta_{n,j}^{-1}) \right)^2 dt \\
&\leq 2 \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \int \mathbb{E} [\tilde{f}_{n,\delta_{n,j}}]^2 \\
&\quad + 2 \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \int \left( \frac{1}{\delta_{n,j}} Q(t\delta_{n,j}^{-1}) \right)^2 dt
\end{aligned}$$

Assumptions (21) on the kernel  $Q$  and the Parseval inequality lead to:

$$\int_{-\infty}^{\infty} \left( \frac{1}{\delta_{n,j}} Q(t\delta_{n,j}^{-1}) \right)^2 dt = \delta_{n,j}^{-1} \int_{-1}^1 Q(t)^2 dt < \infty.$$

Moreover, we use a concentration inequality similar to Proposition 5 stated in Lemma 13. It remains to compute

$$\int \mathbb{E}[\tilde{f}_{n\delta_{n,j}}]^2 = \int \mathbb{V}(\tilde{f}_{n\delta_{n,j}}) + \int \mathbb{E}[\tilde{f}_{n\delta_{n,j}}^2].$$

From the definition of  $\tilde{f}_{n\delta_{n,j}}$ , we remark that

$$\tilde{f}_{n\delta_{n,j}}(x) = \frac{1}{\delta_{n,j}(n-n')} \sum_{k=n'}^n W_{n,j} \left( \frac{x - Z_k}{\delta_{n,j}} \right)$$

where  $W_{n,j}(x) = \int_{-1}^1 e^{-itx} \Phi_Q(t) \Phi_U(t\delta_{n,j}^{-1})^{-1} dt$ . Then

$$\int \mathbb{V}(\tilde{f}_{n\delta_{n,j}}) \leq \frac{1}{\delta_{n,j}^2(n-n')} \int_{-\infty}^{\infty} \mathbb{E} \left( W_{n,j} \left( \frac{x - Z}{\delta_{n,j}} \right)^2 \right) dx$$

A straightforward computation using that  $g \in \mathbf{H}_\nu^+$  leads to:

$$\int_{-\infty}^{\infty} \mathbb{E} \left( W_{n,j} \left( \frac{x - Z}{\delta_{n,j}} \right)^2 \right) dx \leq \frac{4\delta_{n,j}^{1-2\nu}}{d_2} \int_{-1}^1 |\Phi_Q|^2 (\delta_{n,j}^\nu + t^\nu)^2 dt$$

Thus for some constant  $C$ :

$$\int \mathbb{V}(\tilde{f}_{n\delta_{n,j}}) \leq C\delta_{n,j}^{-1-2\nu}(n-n')^{-1},$$

which entails as  $s_j > 0$  that  $\mathbb{V}(\tilde{f}_{n\delta_{n,j}}) \rightarrow 0$ .

It remains to bound  $\int \mathbb{E}[\tilde{f}_{n,\delta_{n,j}}^2]$ . The Parseval inequality and the fact that  $|\Phi_X| \leq 1$  yields:

$$\int \mathbb{E}[\tilde{f}_{n,\delta_{n,j}}^2] \leq \delta_{n,j}^{-1} \int_{-1}^1 |\Phi_Q(t)|^2 dt.$$

Therefore, combining the previous inequality we deduce that:

$$\begin{aligned}
\int \left( \mathbb{E}(\hat{f}_{n,j} - f) \right)^2 &\leq 2MR\delta_{n,j}^{2s} + 4C(s,R)(n-n')^{-\frac{2\min(s,s_j)+1}{2s_j+2\nu}} \\
&\quad \times \left( \delta_{n,j}^{-1} \int Q(t)^2 dt + C\delta_{n,j}^{-1-2\nu}(n-n')^{-1} + \delta_{n,j}^{-1} \int_{-1}^1 |\Phi_Q(t)|^2 dt \right)
\end{aligned}$$

The second term of the r.h.s. is of order  $h_{n,j}^{2\min(s,s_j)+1}\delta_{n,j}^{-1}$  which is negligible before  $\delta_{n,j}^{2\min(s,s_j)}$ . This entails that:

$$\int \left( \mathbb{E}(\hat{f}_{n,j} - f) \right)^2 \leq C(s,R)(n)n^{-\frac{2\min(s,s_j)}{2s_j+2\nu+1}}. \quad (45)$$

**Integrated variance** Let us first remark that:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,j}) &= \int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,h_{n,j},\delta_{n,j}}) \\ &\leq 2 \int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,h_{n,j}}) + 2 \int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,h_{n,j}}).\end{aligned}$$

We have proved above that

$$2 \int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,h_{n,j}}) \leq C \delta_{n,j}^{-1-2\nu} (n - n')^{-1}.$$

It remains to consider the second term. We use once again an auxiliary sequence  $\Psi_n \rightarrow 0$ :

$$\begin{aligned}\int_{-\infty}^{\infty} \mathbb{V}(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}}) &\leq \int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2] \\ &\leq \int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2 \mathbf{1}_{|\hat{p}_{n,j} - p| > \Psi_n}] \\ &\quad + \int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2 \mathbf{1}_{|\hat{p}_{n,j} - p| \leq \Psi_n}].\end{aligned}$$

Similarly as for the integrated bias we obtain that:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2 \mathbf{1}_{|\hat{p}_{n,j} - p| > \Psi_n}] &\leq 2 \int \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2 (1 - \hat{p}_{n,j}^{(1)})^2} \tilde{f}_{n,\delta_{n,j}}^2 \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \right] \\ &\quad + 2 \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2 (1 - \hat{p}_{n,j}^{(1)})^2} \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \right] \int \left( \frac{1}{\delta_{n,j}} Q(t \delta_{n,j}^{-1}) \right)^2 dt.\end{aligned}$$

The second term of the r.h.s. is bounded by

$$2 \delta_{n,j}^{-1} \frac{2(1 - \tau_n)^2 + 2p^2}{\tau_n^2 (1 - p)} \left( \int_{-1}^1 Q(t) dt \right) \mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n).$$

We prove below in (46) that  $\mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n)$  decreases exponentially fast to 0 when  $\psi_n$  is chosen as  $\psi_n = n^{-1/(2(2s_{\max} + 2\nu))}$ . Thus, the previous term is  $o\left(n^{-\frac{2 \min(s, s_j)}{2s_j + 2\nu + 1}}\right)$ .

Concerning the other term, we apply the Fubini theorem and obtain that:

$$\begin{aligned}\int \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2 (1 - \hat{p}_{n,j}^{(1)})^2} \tilde{f}_{n,\delta_{n,j}}^2 \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \right] &= \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2 (1 - \hat{p}_{n,j}^{(1)})^2} \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \int \tilde{f}_{n,\delta_{n,j}}^2 \right] \\ &= \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2 (1 - \hat{p}_{n,j}^{(1)})^2} \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \int \frac{|\hat{\Phi}_{Z,n}(t) \Phi_Q(\delta_{n,j} t)|^2}{|\Phi_U(t)|^2} dt \right] \\ &\leq \frac{2 + 2p^2}{\tau_n^2 (1 - p)^2} \frac{4 \delta_{n,j}^{-1-2\nu}}{d^2} \int_{-1}^1 |\Phi_Q(t)|^2 (1 + t^\nu)^2 dt \mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n).\end{aligned}$$

Hence, we deduce that similarly

$$\int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2 \mathbf{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n}] \leq o\left(n^{-\frac{2 \min(s, s_j)}{2s_j + 2\nu + 1}}\right).$$

We finally consider the event where  $|\hat{p}_{n,j}^{(1)} - p| \leq \Psi_n$  and use the independence between  $\hat{p}_{n,j}^{(1)}$  and  $\tilde{f}_{n,\delta_{n,j}}$  to obtain

$$\begin{aligned} \int_{-\infty}^{\infty} \mathbb{E}[(\hat{f}_{n,h_{n,j},\delta_{n,j}} - \hat{f}_{n,\delta_{n,j}})^2 \mathbb{1}_{|\hat{p}_{n,j}^{(1)} - p| \leq \Psi_n}] &\leq 2 \int \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2(1-\hat{p}_{n,j}^{(1)})^2} \mathbb{1}_{|\hat{p}_{n,j}^{(1)} - p| \leq \Psi_n} \right] \mathbb{E} [\hat{f}_{n,\delta_{n,j}}^2] \\ &+ 2 \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2(1-\hat{p}_{n,j}^{(1)})^2} \mathbb{1}_{|\hat{p}_{n,j}^{(1)} - p| \leq \Psi_n} \right] \int \left( \frac{1}{\delta_{n,j}} Q(t\delta_{n,j}^{-1}) \right)^2 dt. \end{aligned}$$

From Lemma 13 and an argument similar to the one used for the integrated bias, we conclude that this last term is equal to  $O\left(n^{-\frac{2 \min(s,s_j)}{2s_j+2\nu+1}}\right)$ , which concludes the proof of Proposition 12.  $\square$

**Lemma 13** *Assume  $s_j \geq 1/2$  and that  $f \in \mathbf{H}_s(R)$ ,  $g \in \mathbf{H}_\nu^+$ , with the choices of  $h_{n,j} = n^{-1/(2s_j+2\nu)}$  and  $\tau_n = \log(n)^{-a}$  then*

$$\mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \leq C(s, R) n^{-\frac{2 \min(s,s_j)+1}{2s_j+2\nu}}.$$

### Proof

Let us first remark that using the Cauchy-Schwarz inequality and a truncation strategy, we have that:

$$\mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \leq \frac{1}{\tau_n^2(1-p)^2} \mathbb{E}[(\hat{p}_{n,j}^{(1)} - p)^2].$$

Moreover using (5) and (7)

$$\mathbb{E}[(\hat{p}_{n,j}^{(1)} - p)^2] \leq 2C_s R^2 h_{n,j}^{2s+1} + \frac{cte(s)}{n h_{n,j}^{2\nu-1}},$$

where  $cte(s)$  is a constant which depend continuously in  $s$ . Therefore, with our choice of  $h_{n,j}$  in (9) we deduce that there exist a constant  $\phi(s, R)$  depending continuously on  $s$  and  $R$  such that

$$\mathbb{E}[(\hat{p}_{n,j}^{(1)} - p)^2] \leq \phi(s, R) n^{-\frac{2 \min(s,s_j)+1}{2s_j+2\nu}},$$

and thus

$$\mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 \leq \frac{1}{\tau_n^2(1-p)^2} \phi(s, R) n^{-\frac{2 \min(s,s_j)+1}{2s_j+2\nu}}.$$

With this simple reasoning our upper bound depend on the truncation  $\tau_n$ . Therefore, we will refine the result by using an auxiliary sequence  $\Psi_n \rightarrow 0$  and split the events into two sub-cases:

$$\begin{aligned} \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)}{(1-p)(1-\hat{p}_{n,j}^{(1)})} \right]^2 &\leq \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2(1-\hat{p}_{n,j}^{(1)})^2} \right] \\ &\leq \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2(1-\hat{p}_{n,j}^{(1)})^2} \mathbb{1}_{|\hat{p}_{n,j}^{(1)} - p| \leq \Psi_n} \right] + \mathbb{E} \left[ \frac{(\hat{p}_{n,j}^{(1)} - p)^2}{(1-p)^2(1-\hat{p}_{n,j}^{(1)})^2} \mathbb{1}_{|\hat{p}_{n,j}^{(1)} - p| > \Psi_n} \right] \\ &\leq \frac{1}{(1-p)^2(1-p-\Psi_n)^2} \phi(s, R) n^{-\frac{2 \min(s,s_j)+1}{2s_j+2\nu}} + \frac{1}{(1-p)^2 \tau_n^2} \mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n). \end{aligned}$$

In the following, our aim is to calibrate the sequence  $\Psi_n$  such that the second term is negligible comparing to the first one. We have studied concentration inequalities in Section 2.2.2 for a non truncated version of  $\hat{p}_{n,j}^{(1)}$ , that we denote here  $\tilde{p}_{n,j}^{(1)}$ . Then

$$\begin{aligned}
\mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n) &\leq \mathbb{P}(|\hat{p}_{n,j}^{(1)} - \tilde{p}_{n,j}^{(1)}| > \Psi_n) + \mathbb{P}(|\tilde{p}_{n,j}^{(1)} - p| > \Psi_n) \\
&\leq \mathbb{E}(\mathbf{1}_{\tau_n > \Psi_n/2} \mathbf{1}_{\tilde{p}_{n,j}^{(1)} > 1 - \tau_n}) + \mathbb{P}(|\tilde{p}_{n,j}^{(1)} - p| > \Psi_n/2), \\
&\leq \mathbb{E}(\mathbf{1}_{\tau_n > \Psi_n/2} \mathbf{1}_{\tilde{p}_{n,j}^{(1)} > 1 - \tau_n}) + \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_{k,j}\right| > \Psi_n/4\right) + \mathbb{P}(C_s R h_{n,l}^{s+1/2} > \Psi_n/4).
\end{aligned}$$

where the random variables  $\xi_{k,j}$  are defined in (18) Therefore

$$\begin{aligned}
\mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n) &\leq \mathbb{P}(|\hat{p}_{n,j}^{(1)} - \tilde{p}_{n,j}^{(1)}| > \Psi_n) + \mathbb{P}(|\tilde{p}_{n,j}^{(1)} - p| > \Psi_n) \\
&\leq \mathbb{E}(\mathbf{1}_{\tau_n > \Psi_n/2} \mathbf{1}_{\tilde{p}_{n,j}^{(1)} > 1 - \tau_n}) + \mathbb{P}(|\tilde{p}_{n,j}^{(1)} - p| > \Psi_n/2), \\
&\leq \mathbb{E}(\mathbf{1}_{\tau_n > \Psi_n/2} \mathbf{1}_{\tilde{p}_{n,j}^{(1)} > 1 - \tau_n}) + \mathbb{P}\left(\left|\frac{1}{n - n'} \sum_{k=n'}^n \xi_{k,j}\right| > \Psi_n/4\right) \\
&\quad + \mathbb{P}(C_s R h_{n,l}^{s+1/2} > \Psi_n/4).
\end{aligned}$$

Using the Bernstein inequality, we then have:

$$\begin{aligned}
\mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n) &\leq \mathbf{1}_{\tau_n > \Psi_n/2} \mathbb{P}(\tilde{p}_{n,j}^{(1)} > 1 - \tau_n) + \exp\left(-\frac{(n - n')\Psi_n^2}{8(C' h_{n,l}^{-2\nu+1} + C h_{n,l}^{-\nu} \Psi_n/3)}\right) \\
&\quad + \mathbf{1}_{C_s R h_{n,l}^{s+1/2} > \Psi_n/4}.
\end{aligned}$$

We choose  $\Psi_n = n^{-1/2(2s_{max} + 2\nu)}$  such that the last term of the r.h.s. is always null. Thus, it remains to consider  $\mathbb{P}(\tilde{p}_{n,j}^{(1)} > 1 - \tau_n)$ . Applying similarly Bernstein inequality leads to:

$$\mathbb{P}(\tilde{p}_{n,j}^{(1)} > 1 - \tau_n) \leq \mathbb{P}(|\tilde{p}_{n,j}^{(1)} - p| > 1 - \tau_n - p) \leq \exp\left(-\frac{(n - n')(1 - \tau_n - p)^2}{8(C' h_{n,l}^{-2\nu+1} + C h_{n,l}^{-\nu}(1 - \tau_n - p)/3)}\right)$$

The two previous inequalities yield:

$$\begin{aligned}
\mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n) &\leq \exp\left(-\frac{(n - n')(1 - \tau_n - p)^2}{8(C' h_{n,l}^{-2\nu+1} + C h_{n,l}^{-\nu}(1 - \tau_n - p)/3)}\right) \\
&\quad + \exp\left(-\frac{(n - n')\Psi_n^2}{8(C' h_{n,l}^{-2\nu+1} + C h_{n,l}^{-\nu} \Psi_n/3)}\right).
\end{aligned} \tag{46}$$

It is clear that  $\tau_n^{-2} \mathbb{P}(|\hat{p}_{n,j}^{(1)} - p| > \Psi_n) = o(n^{-1/2(2s_{max} + 2\nu)})$ , which concludes the proof.  $\square$

## References

- [AL16] S. Arlot and M. Lerasle. Choice of  $v$  for  $v$ -fold cross-validation in least-squares density estimation. *Journal of Machine Learning Research*, 17(208):1–50, 2016.
- [Arl09] S. Arlot. Model selection by resampling penalization. *Electronic Journal of Statistics*, 3:557–624, 2009.
- [BCG12] J. Bigot, C. Christophe, and S. Gadat. Random action of compact Lie groups and minimax estimation of a mean pattern. *IEEE, Transactions on Information Theory*, 58(6):3509–3520, 2012.
- [BG10] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *The Annals of Statistics*, 38(4):2422–2464, 2010.

- [BGKM13] J. Bigot, S. Gadat, T. Klein, and C. Marteau. Intensity estimation of non-homogeneous poisson processes from shifted trajectories. *Electronic Journal of Statistics*, 7:317–372, 2013.
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities, A nonasymptotic theory of independence*. Oxford University Press, Oxford, 2013.
- [BM98] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4:329–375, 1998.
- [Cav11] L. Cavalier. Inverse problems in statistics. In P. Alquier, E. Gautier, and G. Stoltz, editors, *Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, August 31 - September 4, 2009*, pages 3–96. Springer Berlin Heidelberg, 2011.
- [CH88] Raymond J Carroll and Peter Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.
- [CH05] L. Cavalier and N.W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345–1361, 2005.
- [Chi10] M. Chichignoud. Statistical performances of nonlinear estimators. *PhD Thesis of Aix-Marseille University*, 2010.
- [CK02] L. Cavalier and J. Koo. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE, Transactions on Information Theory*, 48:2794–2802, 2002.
- [CL13] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Annales de l’Institut Henri Poincaré - Probabilités et Statistiques*, 49:569–609, 2013.
- [CR07] L. Cavalier and M. Raimondo. Wavelet deconvolution with noisy eigenvalues. *IEEE, Transactions on signal processing*, 55(6):2414–2424, 2007.
- [DH06] A. Delaigle and P. Hall. On optimal kernel choice for deconvolution. *Statistics and Probability Letters*, 76:1594–1602, 2006.
- [DJKP96] D. Donoho, I. Johnstone, G. Kerkycharian, and D. Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24:508–539, 1996.
- [Fan91] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19:1257–1272, 1991.
- [FK02] J. Fan and J.-Y. Koo. Wavelet deconvolution. *IEEE, Transactions on Information Theory*, 48(3):734–747, 2002.
- [GL11] A. Goldenshluger and O. Lepskii. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.
- [GvES11] S. Gugushvili, B. van Es, and P. Spreij. Deconvolution for an atomic distribution: rates of convergence. *Journal of Nonparametric Statistics*, 23:1003–1029, 2011.
- [Kat99] V. Katkovnik. A new method for varying adaptive bandwidth selection. *IEEE, Transactions on signal processing*, 47(9):2567–2571, 1999.
- [Lep92] O. Lepskii. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992.
- [Lep15] O. Lepskii. Adaptive estimation over anisotropic functional classes via oracle approach. *The Annals of Statistics*, 43(3):1178 – 1242, 2015.
- [PC84] R. Picard and R. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984.
- [Tsy09] A. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.
- [vEGS08] B. van Es, S. Gugushvili, and P. Spreij. Deconvolution for an atomic distribution. *Electronic Journal of Statistics*, 2:265–297, 2008.

† Institut de Mathématiques de Toulouse  
Université Toulouse 3 Paul Sabatier  
118, route de Narbonne  
31062 Toulouse Cedex 9, France

‡Toulouse School of Economics  
Université Toulouse 1 Capitole  
21, allées de Brienne  
31000 Toulouse, France

‡‡ Centre de Physiopathologie Toulouse Purpan (CPTP),  
INSERM UMR1043, CNRS UMR 5282,  
Université Toulouse III, France  
CHU Purpan, BP 3028, 31024 Toulouse CEDEX 03, France

{costa<sup>†</sup>,gadat<sup>‡</sup>,risser<sup>†</sup>}@math.univ-toulouse.fr  
pauline.gonnord<sup>‡‡</sup>@inserm.fr