Nonparametric estimation of the division rate of a size-structured population

Marie Doumic-Jauffret, Marc Hoffmann, Patricia Reynaud-Bouret, Vincent Rivoirard

▶ **To cite this version:**

HAL Id: hal-00578694

https://hal.archives-ouvertes.fr/hal-00578694

Submitted on 21 Mar 2011

Nonparametric estimation of the division rate of a size-structured population

M. Doumic[*], M. Hoffmann[†], P. Reynaud-Bouret[‡] and V. Rivoirard[§]

March 21, 2011

Abstract

We consider the problem of estimating the division rate of a size-structured population in a nonparametric setting. The size of the system evolves according to a transport-fragmentation equation: each individual grows with a given transport rate, and splits into two offsprings of the same size, following a binary fragmentation process with unknown division rate that depends on its size. In contrast to a deterministic inverse problem approach, as in [23, 4], we take in this paper the perspective of statistical inference: our data consists in a large sample of the size of individuals, when the evolution of the system is close to its time-asymptotic behavior, so that it can be related to the eigenproblem of the considered transport-fragmentation equation (see [22] for instance). By estimating statistically each term of the eigenvalue problem and by suitably inverting a certain linear operator (see [4]), we are able to construct a more realistic estimator of the division rate that achieves the same optimal error bound as in related deterministic inverse problems. Our procedure relies on kernel methods with automatic bandwidth selection. It is inspired by model selection and recent results of Goldenschluger and Lepski [13, 14].

Keywords: Lepski method, Oracle inequalities, Adaptation, Aggregation-fragmentation equations, Statistical inverse problems, Nonparametric density estimation, Cell-division equation.

Mathematical Subject Classification: 35A05, 35B40, 45C05, 45K05, 82D60, 92D25,

[*]INRIA Rocquencourt, projet BANG, Domaine de Volceau, BP 105, 781153 Rocquencourt, France. *email*: marie.doumic-jauffret@inria.fr

[†]ENSAE-CREST and CNRS-UMR 8050, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France. *email*: marc.hoffmann@ensae.fr

[‡]CNRS-UMR 6621 and Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice cedex 02, France. *email*: patricia.reynaud-bouret@unice.fr

[§]CEREMADE, CNRS-UMR 7534, Université Paris Dauphine, Place Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France. INRIA Paris-Rocquencourt, projet Classic. *email*: Vincent.Rivoirard@dauphine.fr

1

62G05, 62G20

Contents

1 Introduction

1.1 Motivation

Structured models have long served as a representative deterministic model used to describe the evolution of biological systems, see for instance [19] or [20] and references therein. In their simplest form, structured models describe the temporal evolution of a population structured by a biological parameter such as size, age or any significant *trait*, by means of an evolution law, which is a mass balance at the macroscopic scale. A paradigmatic example is given by the transport-fragmentation equation in cell division, that reads

$$
\begin{cases}
\dfrac{\partial}{\partial t}n(t,x) + \dfrac{\partial}{\partial x}\big(g_0(x)n(t,x)\big) + B(x)n(t,x) = 4B(2x)n(t,2x), \;\; t \geq 0, \;\; x \geq 0, \\[2mm]
gn(t,x=0) = 0, \;\; t > 0, \\[2mm]
n(t=0,x) = n^0(x), \;\; x \geq 0.
\end{cases}
\tag{1.1}
$$

The mechanism captured by Equation (1.1) can be described as a mass balance equation (see [1, 20]): the quantity of cells $n(t,x)$ of size x at time t is fed by a transport term $g_0(x)$ that accounts for growth by nutrient uptake, and each cell can split into two offsprings of the same size according to a division rate $B(x)$. Supposing $g_0(x) = \kappa g(x)$, where we suppose a given model for the growth rate $g(x)$ known up to a multiplicative constant $\kappa > 0$, and experimental data for $n(t,x)$, the problem we consider here is to recover the division rate $B(x)$ and the constant κ.

In [23], Perthame and Zubelli proposed a deterministic method based on the asymptotic behavior of the cell amount $n(t,x)$: indeed, it is known (see *e.g.* [21, 22]) that under suitable assumptions on g and B, by the use of the *general relative entropy principle* (see [19]), one has

$$
\int_0^\infty \big|n(t,x)e^{-\lambda t} - \langle n^0, \phi\rangle N(x)\big|\phi(x)\,dx \underset{t\to\infty}{\longrightarrow} 0
\tag{1.2}
$$

where $\langle n^0, \phi\rangle = \int n^0(y)\phi(y)dy$ and ϕ is the adjoint eigenvector (see [21]). The density N is the first eigenvector, and (λ, N) the unique solution of the following eigenvalue problem

$$
\begin{cases}
\kappa\dfrac{\partial}{\partial x}\big(g(x)N(x)\big) + \lambda N(x) = 4BN(2x) - BN(x), \;\; x > 0, \\[2mm]
B(0)N(0) = 0, \qquad \int N(x)dx = 1, \qquad N(x) \geq 0, \qquad \lambda > 0.
\end{cases}
\tag{1.3}
$$

Moreover, under some supplementary conditions, this convergence occurs exponentially fast (see [22]). Hence, in the rest of this article, we work under the following analytical assumptions.

Assumption 1 (Analytical assumptions).

1. *For the considered nonnegative functions g and B and for $\kappa > 0$, there exists a unique eigenpair (λ, N) solution of Problem (1.3).*

2. *This solution satisfies, for all $p \geq 0$, $\int x^p N(x) dx < \infty$ and $0 < \int g(x) N(x) dx < \infty$.*

3. *The functions N and gN belong to $\mathcal{W}^{s+1}$ with $s \geq 1$, and in particular $\left\| N \right\|_\infty < \infty$ and $\left\| (gN)' \right\|_2 < \infty$. ($\mathcal{W}^{s+1}$ denotes the Sobolev space of regularity $s+1$ measured in $\mathbb{L}^2$-norm.)*

4. *We have $g \in \mathbb{L}^\infty(\mathbb{R}_+)$ with $\mathbb{R}_+ = [0, \infty)$.*

Hereafter $\|\bullet\|_2$ and $\|\bullet\|_\infty$ denote the usual $\mathbb{L}^2$ and $\mathbb{L}^\infty$ norms on $\mathbb{R}_+$. Assertions 1 and 2 are true under the assumptions on g and B stated in Theorem 1.1 of [3], under which we also have $N \in \mathbb{L}^\infty$. Assertion 3 is a (presumably reasonable) regularity assumption, necessary to obtain rates of convergence together with the convergence of the numerical scheme. Assertion 4 is restrictive, but mandatory in order to apply our statistical approach.

Thanks to this asymptotic behavior provided by the entropy principle (1.2), instead of requiring time-dependent data $n(t, x)$, which is experimentally less precise and more difficult to obtain, the inverse problem becomes: How to recover (κ, B) from observations on (λ, N) ? In [23, 4], as generally done in deterministic inverse problems (see [5]), it was supposed that experimental data were pre-processed into an approximation N_ε of N with an *a priori* estimate of the form $\|N - N_\varepsilon\| \leq \varepsilon$ for a suitable norm $\|\bullet\|$. Then, recovering B from N_ε becomes an inverse problem with a certain degree of ill-posedness. From a modelling point of view, this approach suffers from the limitation that knowledge on N is postulated in an abstract and somewhat arbitrary sense, that is not genuinely related to experimental measurements.

1.2 The statistical approach

In this paper, we propose to overcome the limitation of the deterministic inverse problems approach by assuming that we have n data, each data being obtained from the measurement of an individual cell picked at random, after the system has evolved for a long time so that the approximation $n(t, x) \approx N(x)e^{\lambda t}$ is valid. This is actually what happens if one observes cell cultures in laboratory after a few hours, a typical situation for *E. Coli* cultures for instance, provided, of course, that the underlying aggregation-fragmentation equation is valid.

Each data is viewed as the outcome of a random variable X_i, each X_i having probability

distribution $N(x)dx$. We thus observe $(X_1, \ldots, X_n)$, with

$$\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) = \prod_{i=1}^n N(x_i)dx_i,$$

and where $\mathbb{P}(\bullet)$ hereafter denotes probability[1]. We assume for simplicity that the random variables X_i are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that they are stochastically independent. Our aim is to build an estimator of $B(x)$, that is a function $x \rightsquigarrow \hat{B}_n(x, X_1, \ldots, X_n)$ that approximates the true $B(x)$ with optimal accuracy and nonasymptotic estimates. To that end, consider the operator

$$(\lambda, N) \rightsquigarrow \mathfrak{T}(\lambda, N)(x) := \kappa \frac{\partial}{\partial x}\big(g(x)N(x)\big) + \lambda N(x), \quad x \geq 0. \tag{1.4}$$

From representation (1.3), we wish to find B, solution to $\mathfrak{T}(\lambda, N) = \mathcal{L}(BN)$, where

$$\mathcal{L}\big(\varphi\big)(x) := 4\varphi(2x) - \varphi(x), \tag{1.5}$$

based on statistical knowledge of (λ, N) only. Suppose that we have preliminary estimators $\hat{L}$ and $\hat{N}$ of respectively $\mathfrak{T}(\lambda, N)$ and N, and an approximation $\mathcal{L}_k^{-1}$ of $\mathcal{L}^{-1}$. Then we can reconstruct B in principle by setting formally

$$\widehat{B} := \frac{\mathcal{L}_k^{-1}(\hat{L})}{\widehat{N}}.$$

This leads us to distinguish three steps that we briefly describe here. The whole method is fully detailed in Section 2.

The first and principal step is to find an optimal estimator $\hat{L}$ for $\mathfrak{T}(\lambda, N)$. To do so, the main part consists in applying twice the Goldenschluger and Lepski's method [14] (GL for short). This method is a new version of the classical Lepski method [9, 10, 11, 12]. Both methods are adaptive to the regularity of the unknown signal and the GL method furthermore provides with an oracle inequality. For the unfamiliar reader, we discuss *adaptive properties* later on, and explain in details the *GL method* and *the oracle point of view* in Section 2.

1. First, we estimate the density N by a kernel method, based on a kernel function K. We define $\hat{N} = \hat{N}_{\hat{h}}$ where $\hat{N}_h$ is defined by (2.1) and the bandwidth $\hat{h}$ is selected automatically by (2.3) from a properly-chosen set $\mathcal{H}$. (see Section 2.1 for more details). A so-called oracle inequality is obtained in Proposition 2 measuring the quality of estimation of N by $\hat{N}$. Notice that this result, which is just a simplified version of [14], is valid for estimating any density, since we have only assumed to observe an $n-$sample of N, so that this result can be considered *per se*.

---

[1]In the sequel, we denote by $\mathbb{E}(\bullet)$ the expectation operator with respect to $\mathbb{P}(\bullet)$ likewise.

2. Second, we estimate the density derivative (up to g) $D = \frac{\partial}{\partial x}(gN)$, again by a kernel method with the same kernel K as before, and select an optimal bandwidth $\tilde{h}$ given by Formula (2.6) similarly. This defines an estimator $\hat{D} := \hat{D}_{\tilde{h}}$ where $\hat{D}_h$ is specified by (2.4), and yields an oracle inequality for $\hat{D}$ stated in Proposition 3. In the saemway as for N, this result has an interest *per se* and is not a direct consequence of [14].

From there, it only remains to find estimators of λ and κ. To that end, we make the following *a priori* (but presumably reasonable) Assumption 2 on the existence of an estimator $\hat{\lambda}_n$ of λ.

Assumption 2 (Assumption on $\hat{\lambda}_n$). *There exists some $q > 1$ such that*

$$\varepsilon_{\lambda,n} = \left(\mathbb{E}[|\hat{\lambda}_n - \lambda|^q] \right)^{1/q} < \infty, \qquad R_{\lambda,n} = \mathbb{E}[\hat{\lambda}_n^{2q}] < \infty.$$

Indeed, in practical cell culture experiments, one can track n individual cells that have been picked at random through time. By looking at their evolution, it is possible to infer λ in a classical parametric way, via an estimator $\hat{\lambda}_n$ that we shall assume to possess from now on[2]. Based on the following simple equality

$$\kappa = \lambda \rho_g(N) \text{ where } \rho_g(N) = \frac{\int_{\mathbb{R}_+} x N(x) dx}{\int_{\mathbb{R}_+} g(x) N(x) dx}, \tag{1.6}$$

obtained by multiplying (1.3) by x and integrating by part, we then define an estimator $\hat{\kappa}_n$ by (2.8). Finally, defining $\hat{L} = \hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}$ ends this first step. The second step consists in the formal inversion of $\mathcal{L}$ and its numerical approximation: For this purpose, we follow the method proposed in [4] and recalled in Section 2.4. To estimate $H := BN$, we state

$$\hat{H} := \mathcal{L}_k^{-1}(\hat{L}) \tag{1.7}$$

where $\mathcal{L}_k^{-1}$ is defined by (2.10) on a given interval $[0, T]$. A new approximation result between $\mathcal{L}^{-1}$ and $\mathcal{L}_k^{-1}$ is given by Proposition 4. The third and final step consists in setting $\hat{B} := \frac{\hat{H}}{\hat{N}}$, clipping this estimator in order to avoid explosion when N becomes too small, finally obtaining

$$\tilde{B}(x) := \max(\min(\hat{B}(x), \sqrt{n}), -\sqrt{n}). \tag{1.8}$$

---

[2]Mathematically sepaking, this only amounts to enlarge the probability space to a rich enough structure that captures this estimator. We do not pursue that here.

1.3 Rates of convergence

Because of the approximated inversion of $\mathcal{L}$ on $[0, T]$, we will have access to error bounds only on $[0, T]$. We set $\|f\|_{2,T}^2 = \int_0^T f^2(x)dx$. for the $\mathbb{L}^2$-norm restricted to the interval $[0, T]$. If the fundamental (yet technical) statistical result is the oracle inequality for $\hat{H}$ stated in Theorem 1 (see Section 2.5), the relevant part with respect to existing works in the non-stochastic setting [4, 23] is its consequence in terms of rates of convergence. For presenting them, we need to assume that the kernel K has regularity and vanishing moments properties.

Assumption 3 (Assumptions on K). *The kernel K is differentiable with derivative K'. Furthermore, $\int K(x)dx = 1$ and $\|K\|_2$ and $\|K'\|_2$ are finite. Finally, there exists a positive integer m_0 such that $\int K(x)x^p dx = 0$ for $p = 1, \ldots, m_0 - 1$ and $I(m_0) := \int |x|^{m_0} K(x)dx$ is finite.*

Then our proposed estimators satisfy the following properties.

Proposition 1. *Under Assumptions 1, 2 and 3, let us assume that $R_{\lambda,n}$ and $\sqrt{n}\epsilon_{\lambda,n}$ are bounded uniformly in n and specify $\mathcal{L}_k^{-1}$ with $k = n$ f. Assume further that the family of bandwidth $\mathcal{H} = \tilde{\mathcal{H}} = \{D^{-1} : D = D_{\min}, ..., D_{\max}\}$ depends on n is such that $1 \leq D_{\min} \leq n^{1/(2m_0+1)}$ and $n^{1/5} \leq D_{\max} \leq n^{1/2}$ for all n. Then $\hat{H}$ satisfies, for all $s \in [1; m_0 - 1]$*

$$\mathbb{E}\left[\|\hat{H} - H\|_{2,T}^q\right] = O\left(n^{-\frac{qs}{2s+3}}\right), \tag{1.9}$$

Furthermore, if the kernel K is Lipschitz-regular, if there exists an interval $[a, b]$ in $(0, T)$ such that

$$[m, M] := [\inf_{x \in [a,b]} N(x), \sup_{x \in [a,b]} N(x)] \subset (0, \infty), \qquad Q := \sup_{x \in [a,b]} |H(x)| < \infty,$$

and if $\ln(n) \leq D_{\min} \leq n^{1/(2m_0+1)}$ and $n^{1/5} \leq D_{\max} \leq (n/\ln(n))^{1/(4+\eta)}$ for some $\eta > 0$, then $\hat{B}$ satisfies, for all $s \in [1, m_0 - 1]$,

$$\mathbb{E}\left[\|(\tilde{B} - B)1_{[a,b]}\|_2^q\right] = O\left(n^{-\frac{qs}{2s+3}}\right). \tag{1.10}$$

1.4 Remarks and comparison to other works

1) Let us establish formal correspondences between the methodology and results when recovering B from (1.3) from the point of view of statistics or PDE analysis. ¡ After renormalization, we obtain the rate $n^{-s/(2s+3)}$ for estimating B, and this corresponds to ill-posed inverse problems of order 1 in nonparametric statistics. We can make a parallel with additive deterministic noise following Nussbaum and Pereverzev [18] (see also [15]

and the references therein). Suppose we have an approximate knowledge of N and λ up to deterministic errors $\zeta_1 \in \mathbb{L}^2$ and $\zeta_2 \in \mathbb{R}$ with noise level $\varepsilon > 0$: we observe

$$N_\varepsilon = N + \varepsilon\zeta_1, \quad \|\zeta_1\|_2 \le 1, \tag{1.11}$$

and $\lambda_\varepsilon = \lambda + \varepsilon\zeta_2, \ |\zeta_2| \le 1$. From the representation

$$B = \frac{\mathcal{L}^{-1}\mathfrak{T}(N,\lambda)}{N},$$

where $\mathfrak{T}(N,\lambda)$ is defined in (1.4), we have that the recovery of $\mathfrak{T}(N,\lambda)$ is ill-posed in the terminology of Wahba [25] for it involves the computation of the derivative of N. Since $\mathcal{L}$ is bounded with an inverse bounded in $\mathbb{L}^2$ and the dependence in λ is continuous, the overall inversion problem is ill-posed of degree $a = 1$. By classical inverse problem theory for linear cases[3], this means that if $N \in \mathcal{W}^s$, the optimal recovery rate in $\mathbb{L}^2$-error norm should be $\varepsilon^{s/(s+a)} = \varepsilon^{s/(s+1)}$ (see also the work of Doumic, Perthame and collaborators [23, 4]).

Suppose now that we replace the deterministic noise ζ_1 by a random Gaussian *white noise*: we observe

$$N_\varepsilon = N + \varepsilon\mathbb{B} \tag{1.12}$$

where $\mathbb{B}$ is a Gaussian white noise, *i.e.* a random distribution in $\mathcal{W}^{-1/2}$ that operates on test functions $\varphi \in \mathbb{L}^2$ and such that $\mathbb{B}(\varphi)$ is a centered Gaussian variable with variance $\|\varphi\|_2^2$. Model (1.12) serves as a representative toy model for most stochastic error models such as density estimation or signal recovery in the presence of noise. Let us formally introduce the α-fold integration operator $\mathcal{I}^\alpha$ and the derivation operator ∂. We can rewrite (1.12) as

$$N_\varepsilon = \mathcal{I}^1(\partial N) + \varepsilon\mathbb{B}$$

and applying $\mathcal{I}^{1/2}$ to both side, we (still formally) equivalently observe

$$Z_\varepsilon := \mathcal{I}^{1/2} N_\varepsilon = \mathcal{I}^{3/2}(\partial N) + \varepsilon\mathcal{I}^{1/2}\mathbb{B}.$$

We are back to a deterministic setting, since in this representation, we have that the noise $\varepsilon\mathcal{I}^{1/2}\mathbb{B}$ is in $\mathbb{L}^2$. In order to recover ∂N from Z_ε, we have to invert the operator $\mathcal{I}^{3/2}$, which has degree of ill-posedness $3/2$. We thus obtain the rate

$$\varepsilon^{s/(s+3/2)} = \varepsilon^{2s/(2s+3)} = n^{-s/(2s+3)}$$

for the calibration $\varepsilon = n^{-1/2}$ dictated by (1.12) when we compare our statistical model with the deterministic perturbation (see for instance [17] for establishing formally the correspondence $\varepsilon = n^{-1/2}$ is a general setting). This is exactly the rate we find in Proposition 1: the deterministic error model and the statistical error model coincide to that extent[4].

---

[3]although here the problem is nonlinear, but that will not affect the argument.

[4] The statistician reader willl note that the rate $n^{-s/(2s+3)}$ is also the minimax rate of convergence when estimating the derivative of a density, see [6].

2) The estimators $\hat{H}$ and $\hat{B}$ do not need the exact knowledge of s as an input to recover this optimal rate of convergence. We just need to know an upper bound $m_0 - 1$ to choose the regularity of the kernel K. This capacity to obtain the optimal rate without knowing the precise regularity is known in statistics as adaptivity in the minimax sense (see [24] for instance for more details). It is close in spirit to what the discrepancy principle can do in deterministic inverse problems [5]. However, in the deterministic framework, one needs to know the level of noise ε, which is not realistic in practice. In our statistical framework, this level of noise is linked to the size sample n through the correspondence $\varepsilon = n^{-1/2}$.

3) Finally, note that the rate is polynomial and no extra-logarithmic terms appear, as it is often the case when adaptive estimation is considered (see [9, 10, 11, 12]).

The next section explains in more details the GL approach and presents our estimators to a full extent, including the fundamental oracle inequalities. It also elaborates on the methodology related to oracle inequality. The main advantage of oracle inequalities is that they hold nonasymptotically (in n) and that they guarantee an optimal choice of bandwidth with respect to the selected risk. Section 3 is devoted to numerical simulations that illustrate the performance of our method. Proofs are delayed until Section 4.

2   Construction and properties of the estimators

2.1   Estimation of N by the GL method

We first construct an estimator of N. A natural approach is a kernel method, which is all the more appropriate for comparisons with analytical methods (see [4] for the deterministic analogue). The kernel function K should satisfy the following assumption, in force in the sequel.

Assumption 4 (Assumption on the kernel density estimator).   *$K : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$.*

For $h > 0$ and $x \in \mathbb{R}$, define

$$\hat{N}_h(x) := \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i), \tag{2.1}$$

where $K_h(x) = h^{-1}K(h^{-1}x)$. Note in particular that $\mathbb{E}(\hat{N}_h) = K_h \star N$, where $\star$ denotes convolution. We measure the performance of $\hat{N}_h$ via its squared integrated error, *i.e.* the average $\mathbb{L}^2$ distance between N and $\hat{N}_h$. It is easy to see that

$$\mathbb{E}[\|N - \hat{N}_h\|_2] \leq \left\|N - K_h \star N\right\|_2 + \mathbb{E}[\|K_h \star N - \hat{N}_h\|_2],$$

9

with

$$\mathbb{E}[\|K_h \star N - \hat{N}_h\|_2^2] = \frac{1}{n^2}\mathbb{E}\Big[\int\Big[\sum_{i=1}^{n}\Big(K_h(x-X_i) - \mathbb{E}\big(K_h(x-X_i)\big)\Big)\Big]^2 dx\Big]$$

$$= \frac{1}{n^2}\int\sum_{i=1}^{n}\mathbb{E}\Big[\Big(K_h(x-X_i) - \mathbb{E}\big(K_h(x-X_i)\big)\Big)^2\Big]dx$$

$$\leq \frac{1}{n}\mathbb{E}\Big[\int K_h^2(x-X_1)dx\Big] = \frac{\|K_h\|_2^2}{n} = \frac{\|K\|_2^2}{nh}.$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}[\|N - \hat{N}_h\|_2] \leq \|N - K_h \star N\|_2 + \frac{1}{\sqrt{nh}}\|K\|_2.$$

The first term corresponds to a bias term, it decreases when $h \to 0$. The second term corresponds to a variance term, which increases when $h \to 0$. If one has to choose h in a family $\mathcal{H}$ of possible bandwidths, the best choice is $\bar{h}$ where

$$\bar{h} := \operatorname{argmin}_{h \in \mathcal{H}}\big\{\|N - K_h \star N\|_2 + \frac{1}{\sqrt{nh}}\|K\|_2\big\}. \qquad (2.2)$$

This ideal compromise $\bar{h}$ is called the "oracle": it depends on N and then cannot be used in practice. Hence one wants to find an automatic (data-driven) method for selecting this bandwidth. The Lepski method [9, 10, 11, 12] is one of the various theoretical adaptive methods available for selecting a density estimator. In particular it is the only known method able to select a bandwidth for kernel estimators. However the method do not usually provide a non asymptotic oracle inequality. Recently, Goldenschluger and Lepski [13] developed powerful probabilistic tools that enable to overcome this weakness and that can provide with a fully data-driven bandwidth selection method. We give here a practical illustration of their work: how should one select the bandwidth for a given kernel in dimension 1?

The main idea is to estimate the bias term by looking at several estimators. The method consists in setting first, for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^{n}(K_h \star K_{h'})(x - X_i) = (K_h \star \hat{N}_{h'})(x).$$

Next, for any $h \in \mathcal{H}$, define

$$A(h) := \sup_{h' \in \mathcal{H}}\big\{\|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2\big\}_+$$

$$= \sup_{h' \in \mathcal{H}}\big\{\max\big\{0, \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2\big\}\big\},$$

10

where, given $\varepsilon > 0$, we set $\chi := (1 + \varepsilon)(1 + \|K\|_1)$. The quantity $A(h)$ is actually a good estimator of $\|N - K_h \star N\|_2$ up to the term $\|K\|_1$ (see (4.2) and (4.3) in Section 4). The next step consists then in setting

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \left\{ A(h) + \frac{\chi}{\sqrt{nh}} \|K\|_2 \right\}, \tag{2.3}$$

and our final estimator of N is obtained by putting $\hat{N} := \hat{N}_{\hat{h}}$. Let us specify what we are able to prove at this stage.

Proposition 2. *Assume $N \in \mathbb{L}^\infty$ and work under Assumption 4. If $\mathcal{H} \subset \{D^{-1}, D = 1, \ldots, D_{\max}\}$ with $D_{\max} = \delta n$ for $\delta > 0$, then, for any $q \geq 1$,*

$$\mathbb{E}\left[\|\hat{N} - N\|_2^{2q}\right] \leq C(q)\chi^{2q} \inf_{h \in \mathcal{H}} \left\{ \|K_h \star N - N\|_2^{2q} + \frac{\|K\|_2^{2q}}{(hn)^q} \right\} + C_1 n^{-q},$$

where $C(q)$ is a constant depending on q and C_1 is a constant depending on q, ε, δ, $\|K\|_2$, $\|K\|_1$ and $\|N\|_\infty$.

The previous inequality is called an oracle inequality, for we have $\mathbb{E}[\|\hat{N} - N\|_2] \leq (\mathbb{E}[\|\hat{N} - N\|_2^{2q}])^{1/(2q)}$ and $\hat{h}$ is performing as well as the oracle $\bar{h}$ up to some multiplicative constant. In that sense, we are able to select the *best bandwidth* within our family $\mathcal{H}$.

Remark 1. *As compared to the results of Goldenschluger and Lepski in [13], we do not consider the case where $\mathcal{H}$ is an interval and we do not specify K except for Assumption 4. This simpler method is more reasonable from a numerical point of view, since estimating N is only a preliminary step. The probabilistic tool we use here is classical in model selection theory (see Section 4 and [16]) and actually, we do not use directly [13]. In particular the main difference is that, in our specific case, we are able to get $\max(\mathcal{H})$ fixed whereas Goldenschluger and Lepski [13] require $\max(\mathcal{H})$ to tend to 0 with n. The price to pay is that we obtain a uniform bound (see Lemma 1 in Section 4.3) which is less tight, but that will be sufficient for our purpose.*

2.2 Estimation of $\frac{\partial}{\partial x}\big(g(x)N(x)\big)$ by the GL method

The previous method can of course be adapted to estimate

$$D(x) := \frac{\partial}{\partial x}\big(g(x)N(x)\big).$$

We adjust here the work of [14] to the setting of estimating a derivative. We again use kernel estimators with more stringent assumptions[5] on K.

---

[5]For sake of simplicity we use the same kernel to estimate N and D but this choice is not mandatory.

Assumption 5 (Assumption on the kernel of the derivative estimator). *The function K is differentiable, $\int K(x)dx = 1$ and $\int (K'(x))^2 dx < \infty$.*

For any bandwidth $h > 0$, we define the kernel estimator of D as

$$
\begin{aligned}
\hat{D}_h(x) &:= \frac{1}{n}\sum_{i=1}^n g(X_i)K_h'(x - X_i) \\
&= \frac{1}{nh^2}\sum_{i=1}^n g(X_i)K'\left(\frac{x-X_i}{h}\right).
\end{aligned}
\tag{2.4}
$$

Indeed

$$
\begin{aligned}
\mathbb{E}(\hat{D}_h(x)) &= \int K_h'(x - u)g(u)N(u)du \\
&= \big(K_h' \star (gN)\big)(x) = \big(K_h \star (gN)'\big)(x).
\end{aligned}
$$

Again we can look at the integrated squared error of $\hat{D}$. We obtain the following upper bound:

$$
\mathbb{E}[\|\hat{D}_h - D\|_2] \leq \|D - K_h \star D\|_2 + \mathbb{E}[\|K_h \star D - \hat{D}_h\|_2],
$$

with

$$
\begin{aligned}
\mathbb{E}[\|K_h \star D - \hat{D}_h\|_2^2] &= \frac{1}{n^2}\mathbb{E}\left[\int\left[\sum_{i=1}^n\Big(g(X_i)K_h'(x - X_i) - \mathbb{E}\big(g(X_i)K_h'(x - X_i)\big)\Big)\right]^2 dx\right] \\
&= \frac{1}{n^2}\int\sum_{i=1}^n\mathbb{E}\left[\Big(g(X_i)K_h'(x - X_i) - \mathbb{E}\big(g(X_i)K_h'(x - X_i)\big)\Big)^2\right]dx \\
&\leq \frac{1}{n}\mathbb{E}\left[\int g^2(X_1)K_h'^2(x - X_1)dx\right] \\
&\leq \frac{\|g\|_\infty^2\|K_h'\|_2^2}{n} = \frac{\|g\|_\infty^2\|K'\|_2^2}{nh^3}.
\end{aligned}
$$

Hence, by Cauchy-Schwarz inequality

$$
\mathbb{E}[\|D - \hat{D}_h\|_2] \leq \|D - K_h \star D\|_2 + \frac{1}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2.
$$

Once again, there is a bias-variance decomposition, but now the variance term is of order $\frac{1}{\sqrt{nh^3}}\|K'\|_2\|g\|_\infty$. We therefore define the oracle by

$$
\bar{\bar{h}} := \operatorname{argmin}_{h \in \tilde{\mathcal{H}}}\left\{\|D - K_h \star D\|_2 + \frac{1}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\right\}.
\tag{2.5}
$$

Now let us apply the GL method in this case. Let $\tilde{\mathcal{H}}$ be a family of bandwidths. We set for any $h, h' > 0$,

$$
\hat{D}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^n g(X_i)(K_h \star K_{h'})'(x - X_i)
$$

12

and

$$\tilde{A}(h) \ := \ \sup_{h' \in \tilde{\mathcal{H}}} \big\{ \|\hat{D}_{h,h'} - \hat{D}_{h'}\|_2 - \frac{\tilde{\chi}}{\sqrt{nh'^3}} \|g\|_\infty \|K'\|_2 \big\}_+,$$

where, given $\tilde{\varepsilon} > 0$, we put $\tilde{\chi} := (1 + \tilde{\varepsilon})(1 + \|K\|_1)$. Finally, we estimate D by using $\hat{D} := \hat{D}_{\tilde{h}}$ with

$$\tilde{h} := \mathrm{argmin}_{h \in \tilde{\mathcal{H}}} \big\{ \tilde{A}(h) + \frac{\tilde{\chi}}{\sqrt{nh^3}} \|g\|_\infty \|K'\|_2 \big\}. \tag{2.6}$$

As before, we are able to prove an oracle inequality for $\hat{D}$.

Proposition 3. *Assume $N \in \mathbb{L}^\infty$. Work under Assumption 5. If $\tilde{\mathcal{H}} = \{D^{-1}, D = 1 \ldots, \tilde{D}_{\max}\}$, with $\tilde{D}_{\max} = \sqrt{\tilde{\delta}n}$ for $\tilde{\delta} > 0$, then for any $q \geq 1$,*

$$\mathbb{E}\big[\|\hat{D} - D\|_2^{2q}\big] \leq \tilde{C}(q)\tilde{\chi}^{2q} \inf_{h \in \tilde{\mathcal{H}}} \big\{ \|K_h \star D - D\|_2^{2q} + \big(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\big)^{2q} \big\} + \tilde{C}_1 n^{-q},$$

where $\tilde{C}(q)$ is a constant depending on q and $\tilde{C}_1$ is a constant depending on q, $\tilde{\varepsilon}$, $\tilde{\delta}$, $\|K'\|_2$, $\|K'\|_1$, $\|g\|_\infty$ and $\|N\|_\infty$.

2.3   Estimation of κ (and λ)

As mentioned in the introduction, we will not consider the problem of estimating λ and we work under Assumption 2: an estimator $\hat{\lambda}_n$ of λ is furnished by the practitioner prior to the data processing for estimating B. It becomes subsequently straightforward to obtain an estimator of κ by estimating $\rho_g(N)$, see the form of (1.6). We estimate $\rho_g(N)$ by

$$\hat{\rho}_n := \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n g(X_i) + c}, \tag{2.7}$$

where $c > 0$ is a (small) tuning constant[6]. Next we simply put

$$\hat{\kappa}_n = \hat{\lambda}_n \hat{\rho}_n. \tag{2.8}$$

2.4   Approximated inversion of $\mathcal{L}$

From (2.7), the right-hand side of (1.3) is consequently estimated by

$$\hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}.$$

---

[6]In practice, one can take $c = 0$.

It remains to formally apply the inverse operator $\mathcal{L}^{-1}$. However Given φ, the dilation equation

$$\mathcal{L}(\psi)(x) = 4\psi(2x) - \psi(x) = \varphi(x), \quad x \in \mathbb{R}_+ \tag{2.9}$$

admits in general infinitely many solutions, see Doumic $et\ al.$ [4], Appendix A. Nevertheless, if $\varphi \in \mathbb{L}^2$, there is a unique solution $\psi \in \mathbb{L}^2$ to (2.9), see Proposition A.1. in [4], and moreover it defines a continuous operator $\mathcal{L}^{-1}$ from $\mathbb{L}^2$ to $\mathbb{L}^2$. Since K_h and gK'_h belong to $\mathbb{L}^2$, one can define a unique solution to (2.9) when $\varphi = \hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}$. This inverse is not analytically known but we can only approximate it via the fast algorithm described below.

Given $T > 0$ and an integer $k \geq 1$, we construct a linear operator $\mathcal{L}_k^{-1}$ that maps a function $\varphi \in \mathbb{L}^2$ into a function with compact support in $[0, T]$ as follows. Consider the regular grid on $[0, T]$ with mesh $k^{-1}T$ defined by

$$0 = x_{0,k} < x_{1,k} < \cdots < x_{i,k} := \tfrac{i}{k}T < \ldots < x_{k,k} = T.$$

We set

$$\varphi_{i,k} := \frac{k}{T} \int_{x_{i,k}}^{x_{i+1,k}} \varphi(x)dx \ \text{ for } \ i = 0, \ldots, k-1,$$

and define by induction the sequence [7]

$$H_{i,k}(\varphi) := \frac{1}{4}(H_{i/2,k}(\varphi) + \varphi_{i/2,k}), \quad i = 0, \ldots, k-1,$$

what gives, for $i = 0$ and $i = 1$

$$H_0(\varphi) := \tfrac{1}{3}\varphi_{0,k}, \quad H_1(\varphi) := \tfrac{4}{21}\varphi_{0,k} + \tfrac{1}{7}\varphi_{1,k}.$$

Finally, we define

$$\mathcal{L}_k^{-1}(\varphi)(x) := \sum_{i=0}^{k-1} H_{i,k}(\varphi) 1_{[x_{i,k}, x_{i+1,k})}(x). \tag{2.10}$$

As stated in the introduction, we eventually estimate $H = BN$ by

$$\hat{H} = \mathcal{L}_k^{-1}(\hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}).$$

The stability of the inversion is given by the fact that $\mathcal{L}_k^{-1} : \mathbb{L}^2 \to \mathbb{L}^2$ is continuous, see Lemma 4 in Section 4.3, and by the following approximation result between $\mathcal{L}^{-1}$ and $\mathcal{L}_k^{-1}$.

---

[7]for any sequence $u_i, i = 1, 2, \ldots$, we define

$$u_{i/2} := \left\{ \begin{array}{ll} u_{i/2} & \text{if } i \text{ is even} \\ \tfrac{1}{2}(u_{(i-1)/2} + u_{(i+1)/2}) & \text{otherwise.} \end{array} \right.$$

Proposition 4. *Let $T > 0$ and $\varphi \in \mathcal{W}^1$. Let $\mathcal{L}^{-1}(\varphi)$ denote the unique solution of (2.9) belonging to $\mathbb{L}^2$. We have for $k \geq 1$:*

$$\|\mathcal{L}_k^{-1}(\varphi) - \mathcal{L}^{-1}(\varphi)\|_{2,T} \leq C \frac{T}{\sqrt{k}} \|\varphi\|_{\mathcal{W}^1},$$

with $C < \frac{1}{\sqrt{6}}$.

Hence, $\mathcal{L}_k^{-1}$ behaves nicely over sufficiently smooth functions. Moreover the estimation of N and the estimators $\hat{\kappa}_n$ and $\hat{\lambda}_n$ are essentially regular. Finally we estimate B as stated in (1.8). The overall behaviour of the estimator is finally governed by the quality of estimation of the derivative D, which determines the accuracy of the whole inverse problem in all these successive steps.

2.5   Oracle inequalities for $\hat{H}$

We are ready to state our main result, namely the oracle inequality fulfilled by $\hat{H}$.

Theorem 1. *Work under Assumptions 1 and 2 and let K a kernel satisfying Assumptions 4 and 5. Define $\mathcal{H} \subset \{D^{-1}, D = 1 \ldots, D_{\max}\}$ with $D_{\max} = \delta n$ for $\delta > 0$ and $\tilde{\mathcal{H}} \subset \{D^{-1} : D = 1, ..., \tilde{D}_{\max}\}$ with $\tilde{D}_{\max} = \sqrt{\tilde{\delta}n}$ for $\tilde{\delta} > 0$. For $k \geq 1$ and $T > 0$, let us define $\mathcal{L}_k^{-1}$ by (2.10) on the interval $[0, T]$. Finally, define the estimator*

$$\hat{H} = \mathcal{L}_k^{-1}(\hat{\lambda}_n \hat{N}_{\hat{h}} + \hat{\kappa}_n \hat{D}_{\tilde{h}}),$$

where $\hat{N}_{\hat{h}}$ and $\hat{D}_{\tilde{h}}$ are kernel estimators defined respectively by (2.1) and (2.4), and where we have selected $\hat{h}$ and $\tilde{h}$ by (2.3) and (2.6). Moreover take $\hat{\kappa}_n$ as defined by (2.7) and (2.8) for some $c > 0$.

The following upper bound holds for any n:

$$\mathbb{E}\left[\|\hat{H} - H\|_{2,T}^q\right] \leq C_1 \Big\{ \sqrt{R_{\lambda,n}} \inf_{h \in \tilde{\mathcal{H}}} \big[\|K_h \star D - D\|_2^q + \big(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\big)^q\big]$$

$$+ \inf_{h \in \mathcal{H}} \big[\|K_h \star N - N\|_2^q + \big(\frac{\|K\|_2}{\sqrt{nh}}\big)^q\big] + \varepsilon_{\lambda,n}^q + \big((\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big)^q \Big\} + C_2 n^{-\frac{q}{2}},$$

where C_1 is a constant depending on q, g, N, ε, $\tilde{\varepsilon}$, $\|K\|_1$ and c; and C_2 is a constant depending on q, g, N, $\tilde{\varepsilon}$, ε, δ, $\tilde{\delta}$, $\|K\|_2$, $\|K\|_1$, $\|K'\|_2$, $\|K'\|_1$.

1) Note that the upper bound quantifies the additive cost of each step used in the estimation method. The first part is an oracle bound on the estimation of D times the size of the estimator of λ. The second part is the oracle bound for N. Of course, the

15

results are sharp only if $\hat{\lambda}$ is good, which can be seen through $\varepsilon_{\lambda,n}$. Finally, the bound is also governed by the approximated inversion through the only term where k appears. The last term is just a residual term, that will be in most of the cases negligible with respect to the other terms. In particular since all the previous errors are somehow unavoidable, this means that, as far as our method is concerned, our upper bound is the best possible that can be achieved in order to select the different bandwidths h, up to multiplicative constants. Moreover, one can see how the limitation in k influences the method and how large k needs to be chosen to guarantee that the main error comes from the fact that one estimates a derivative.

2) The result holds for any n. In particular, we expect the method to perform well for small sample size, see the numerical illustration below. This also shows that we are able to select a good bandwidth as far as the kernel K is fixed, even if there is no assumption on the moments of K and consequently on the approximation properties of K. In the next simulation section, we focus on a Gaussian kernel which has only one vanishing moment (hence one cannot really consider minimax adaptation for regular function with it) but for which the bandwidth choice is still important in practice. The previous result guarantees an optimal bandwidth choice even for this kernel, up to some multiplicative constant.

3) From this oracle result, we can easily deduce the rates of convergence of Proposition 1 at the price of further assumptions on the kernel K, *i.e.* Assumption 3 defined in the Introduction. Section 1.2.2 of [24] recalls how to build compactly supported kernels satisfying Assumption 3. If m_0 satisfies Assumption 3, then for any $s \leq m_0$, for any $f \in \mathcal{W}^s$,

$$\|K_h \star f - f\|_2 \leq C\|f\|_{\mathcal{W}^s} h^s,$$

where C is a constant that can be expressed by using K and m_0 (see Theorem 8.1 of [8]). Now, it is sufficient to choose $h = D^{-1}$ of order $n^{-1/(2s+3)}$ to obtain (1.9). The complete proof of Proposition 1 is delayed until the last section.

3   Numerical illustration

Let us illustrate our method through some simulations.

3.1   The numerical protocol

First, we need to build up simulated data: to do so, we depart from given g, κ and B on a regular grid $[0, dx, \ldots, X_M]$ and solve the direct problem by the use of the so-called *power algorithm* to find the corresponding density N and the principal eigenvalue λ (see for instance [4] for more details). We check that N almost vanishes outside the interval $[0, X_M]$; else, X_M has to be increased. The density N being given on this grid, we approximate it by a spline function, and build a n-sample by the rejection sampling algorithm. For the

sake of simplicity, we do not simulate an approximation on λ and keep the exact value, thus leading, in the estimate of Theorem 1, to $R_{\lambda,n} = \lambda^{2q}$ and $\hat{\lambda}_n = \lambda$.

We then follow step by step the method proposed here and detailed in Section 2.

1. The GL method for the choice of $\hat{N} = \hat{N}_{\bar{h}}$. We take the classical Gaussian kernel $K(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$, set $D_{\max} = n$ and limit ourselves to a logarithmic sampling $\mathcal{H} = \{1, 1/2, \ldots, 1/9, 1/10, 1/20, \ldots, 1/100, 1/200, \ldots, 1/n\}$ in order to reduce the cost of computations (The GL method is indeed the most time-consuming step in the numerical protocol).

2. The GL method for the choice of $\hat{D}_h$. The procedure is similar except that we choose here $D_{max} = \sqrt{n}$. The selected bandwidths $\bar{h}$ and $\tilde{h}$ can be different. We check that the GL method does not select an extremal point of $\mathcal{H}$.

3. The choice of $\hat{\kappa}_n$, as defined by (2.8).

4. The numerical scheme described in Section 2.4 and in [4] for the inversion of $\mathcal{L}$.

5. The division by $\hat{N}$ and definition of $\tilde{B}$ as described in (1.8).

At each step, we compare, in $\mathbb{L}^2$-norm, the reconstructed function and the original one: $\hat{N}$ vs N, $\frac{\partial}{\partial x}(g\hat{N})$ vs $\frac{\partial}{\partial x}(gN)$, $\hat{H}$ vs BN and finally $\hat{B}$ vs B.

3.2 Results on simulated data

We first test the three cases simulated in [4] in which the numerical analysis approach was dealt with. Namely on the interval $[0,4]$, we consider the cases where $g \equiv 1$ and first $B = B_1 \equiv 1$, second $B(x) = B_2(x) = 1$ for $x \leq 1.5$, then linear to $B_2(x) = 5$ for $x \geq 1.7$. This particular form is interesting because due to this fast increase on B, the solution N is not that regular and exhibits a 2-peaks distribution (see Figure 1). Finally, we test $B(x) = B_3(x) = \exp(-8(x-2)^2) + 1$.

In Figures 1 and 2, we show the simulation results with $n = 5.10^4$ (a realistic value for *in vitro* experiments on *E. Coli* for instance) for the reconstruction of N, $\frac{\partial}{\partial x}(gN)$, BN and B.

One notes that the solution can well capture the global behavior of the division rate B, but, as expected, has more difficulties in recovering fine details (for instance, the difference between B_1 and B_3) and also gives much more error when B is less regular (case of B_2). One also notes that even if the reconstruction of N is very satisfactory, the critical point is the reconstruction of its derivative. Moreover, for large values of x, even if N and its derivative are correctly reconstructed, the method fails in finding a proper division rate B. This is due to two facts: first, N vanishes, so the division by N leads to error amplification.
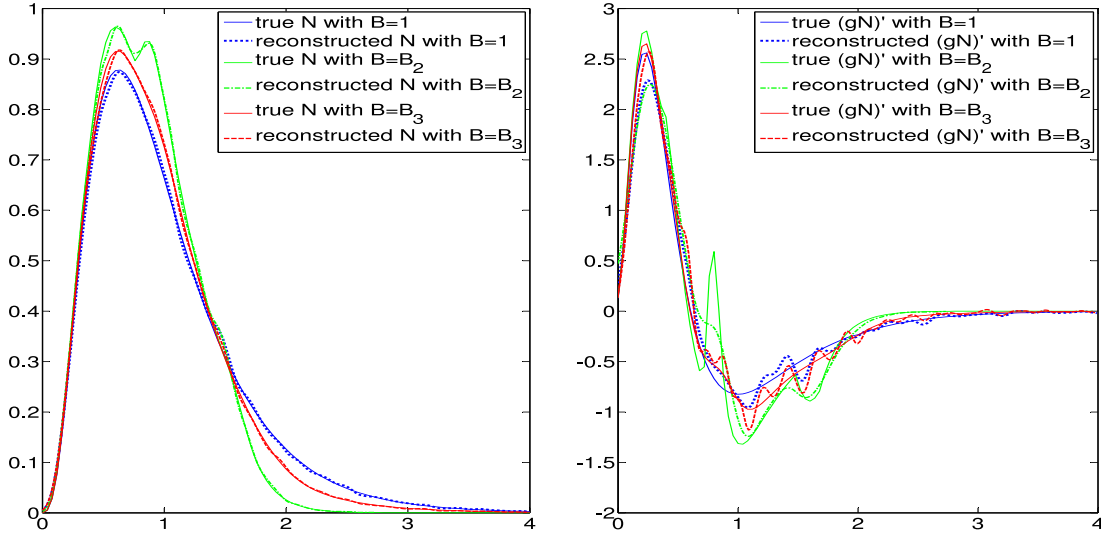
Figure 1: Reconstruction of N (left) and of $\frac{\partial}{\partial x}(gN)$ (right) obtained with a sample of $n = 5.10^4$ data, for three different cases of division rates B.

Second, the values taken by $B(x)$ for large x have little influence on the solutions N of the direct problem: whatever the values of B, the solutions N will not vary much, as shown by Figure 1 (left). A similar phenomenon occurred indeed when solving the deterministic problem in [4] (for instance, we refer to Fig. 10 of this article for a comparison of the results).

We also test a case closer to biological true data, namely the case $B(x) = x^2$ and $g(x) = x$. The results are shown on Figures 3 and 4 for n-samples of size 10^3, 5.10^3, 10^4 and 5.10^4.

One notes that reconstruction is already very good for N when $n = 10^3$, unlike the reconstruction of $\frac{\partial}{\partial x}(gN)$ that requires much more data.

Finally, in Table 3.2 we give average error results on 50 simulations, for $n = 1000$, $g \equiv B \equiv 1$. We display the relative errors in $\mathbb{L}^2$ norms, (defined by $||\phi - \hat{\phi}||_{\mathbb{L}^2}/||\phi||_{\mathbb{L}^2}$), and their empirical variances. In Table 3.2, for the case $g(x) = x$ and $B(x) = x^2$, we give some results on standard errors for various values of n, and compare them to $n^{-1/5}$, which is the order of magnitude of the expected final error on BN, since with a Gaussian kernel we have $s = 1$ in Proposition 1. We see that our numerical results are in line with the theoretical estimates: indeed, the error on H is roughly twice as large as $n^{-1/5}$.
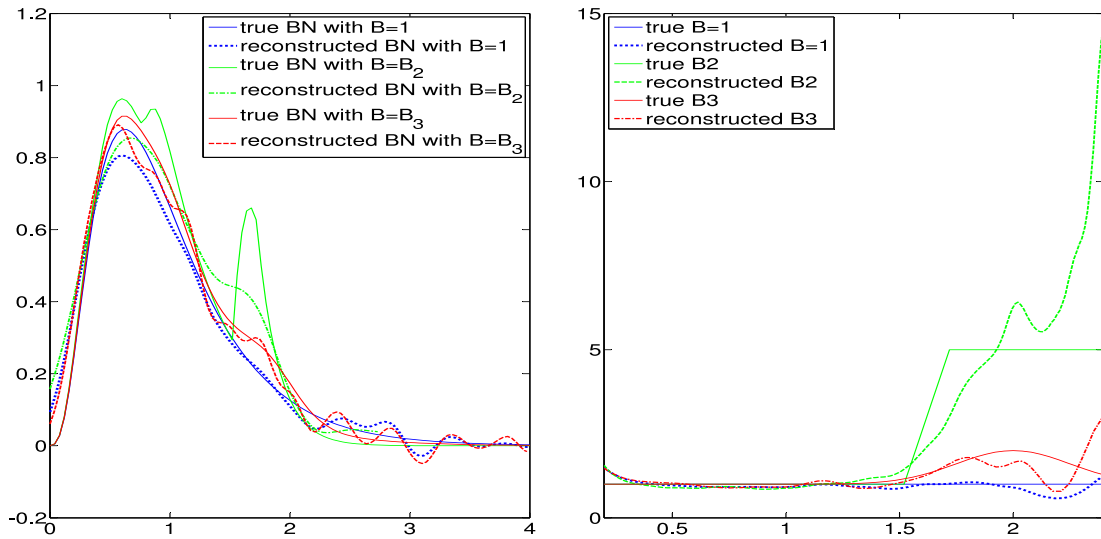
Figure 2: Reconstruction of BN (left) and of B (right) obtained with a sample of $n = 5.10^4$ data, for three different cases of division rates B.

Error on N: average	Variance	Error on $\frac{\partial}{\partial x}(gN)$: average	Variance
0.088	0.15	0.51	0.28
Error on BN: average	Variance	Average $\hat{h}$	Average $\tilde{h}$
0.39	0.29	0.12	0.40

n	$n^{-\frac{1}{5}}$	$\hat{h}$	$\tilde{h}$	error on N	error on D	error on H
10^3	0.25	0.1	0.5	0.06	0.68	0.42
5.10^3	0.18	0.07	0.3	0.03	0.45	0.28
10^4	0.16	0.08	0.3	0.035	0.46	0.29
5.10^4	0.11	0.04	0.2	0.014	0.31	0.19

4 Proofs

In Section 4.1, we first give the proofs of the main results of Section 2. This allows us, in Section 4.2, to prove the results of Section 2.5, which require the collection of all the results of Section 2, *i.e.* the oracle-type inequalities on the one hand and a numerical analysis result on the other hand. This illustrates the subject of our paper that lies at the frontier between these fields. Finally, we state and prove the technical lemmas used in Section 4.1. These technical tools are concerned with probabilistic results, namely concentration and Rosenthal-type inequalities that are often the main bricks to establish oracle inequalities,
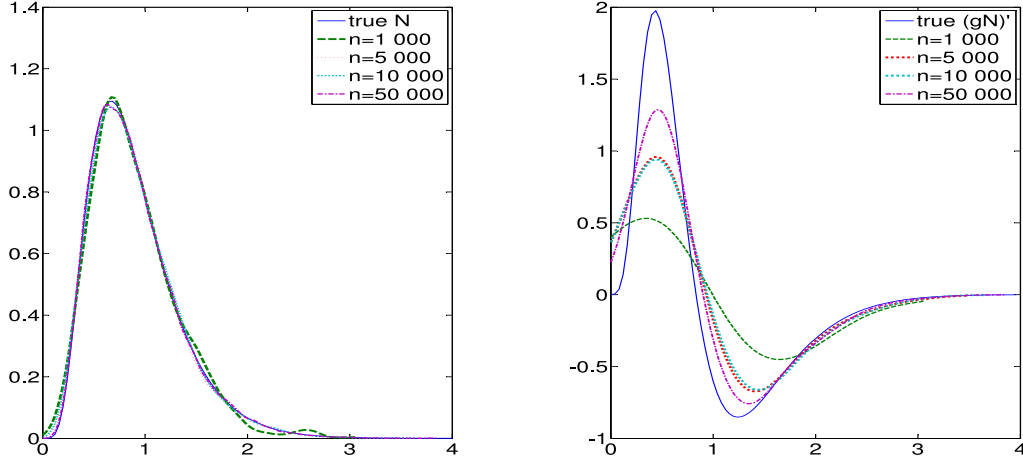
Figure 3: Reconstruction of N (left) and of $\frac{\partial}{\partial x}(gN)$ (right) obtained for $g(x) = x$ and $B(x) = x^2$, for various sample sizes.

and also the boundedness of $\mathcal{L}_k^{-1}$. In the sequel, the notation $\square_{\theta_1,\theta_2,\dots}$ denotes a generic positive constant depending on $\theta_1, \theta_2, \dots$ (the notation $\square$ simply denotes a generic positive absolute constant). It means that the values of $\square_{\theta_1,\theta_2,\dots}$ may change from line to line.

4.1 Proofs of the main results of Section 2

Proof of Proposition 2

For any $h^* \in \mathcal{H}$, we have:

$$
\begin{aligned}
\|\hat{N} - N\|_2 &\leq \|\hat{N}_{\hat{h}} - N_{\hat{h},h^*}\|_2 + \|N_{\hat{h},h^*} - \hat{N}_{h^*}\|_2 + \|\hat{N}_{h^*} - N\|_2 \\
&\leq A_1 + A_2 + A_3,
\end{aligned}
$$

with

$$
A_1 := \|\hat{N}_{\hat{h}} - N_{\hat{h},h^*}\|_2 \leq A(h^*) + \frac{\chi}{\sqrt{n\hat{h}}}\|K\|_2,
$$

$$
A_2 := \|N_{\hat{h},h^*} - \hat{N}_{h^*}\|_2 \leq A(\hat{h}) + \frac{\chi}{\sqrt{nh^*}}\|K\|_2
$$

and

$$
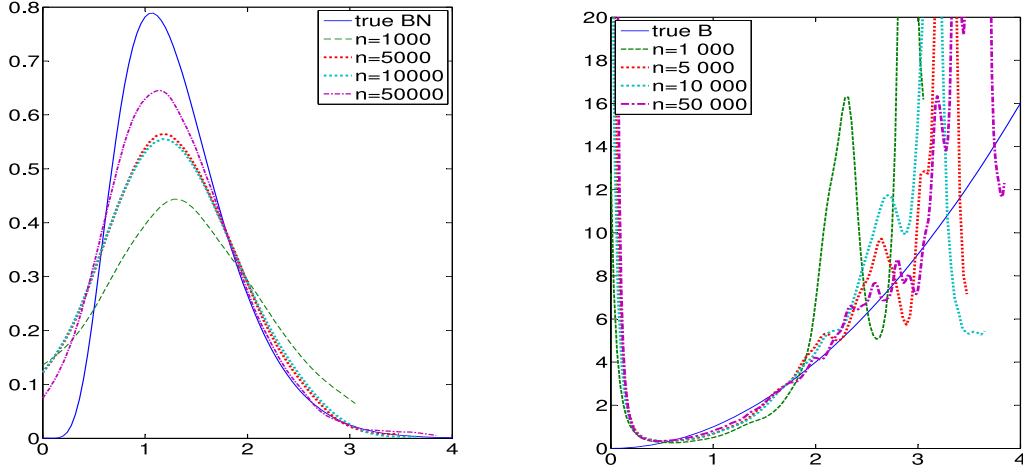A_3 := \|\hat{N}_{h^*} - N\|_2.
$$

20

Figure 4: Reconstruction of BN (left) and of B (right) obtained for $g(x) = x$ and $B(x) = x^2$, for various sample sizes.

We obtain

$$
\begin{aligned}
\|\hat{N} - N\|_2 &\leq A(h^*) + \frac{\chi}{\sqrt{n\hat{h}}}\|K\|_2 + A(\hat{h}) + \frac{\chi}{\sqrt{nh^*}}\|K\|_2 + \|\hat{N}_{h^*} - N\|_2 \\
&\leq 2A(h^*) + \frac{2\chi}{\sqrt{nh^*}}\|K\|_2 + \|\hat{N}_{h^*} - N\|_2.
\end{aligned}
\tag{4.1}
$$

Since we have

$$
\begin{aligned}
A(h^*) &= \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h^*,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2 \right\}_+ \\
&\leq \sup_{h' \in \mathcal{H}} \left\{ \left\{ \|\hat{N}_{h^*,h'} - \mathbb{E}[\hat{N}_{h^*,h'}] - (\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}])\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2 \right\}_+ \right. \\
&\qquad \left. + \|\mathbb{E}[\hat{N}_{h^*,h'}] - \mathbb{E}[\hat{N}_{h'}]\|_2 \right\}
\end{aligned}
\tag{4.2}
$$

and for any x and any h'

$$
\begin{aligned}
\mathbb{E}\left[\hat{N}_{h^*,h'}(x)\right) - \mathbb{E}(\hat{N}_{h'}(x)] &= \int (K_{h^*} \star K_{h'})(x - u)N(u)du - \int K_{h'}(x - v)N(v)dv \\
&= \int\int K_{h^*}(x - u - t)K_{h'}(t)N(u)dtdu - \int K_{h'}(x - v)N(v)dv \\
&= \int\int K_{h^*}(v - u)K_{h'}(x - v)N(u)dudv - \int K_{h'}(x - v)N(v)dv \\
&= \int K_{h'}(x - v)\Big(\int K_{h^*}(v - u)N(u)du - N(v)\Big)dv,
\end{aligned}
$$

21

we derive

$$\| \mathbb{E}(\hat{N}_{h^*,h'}) - \mathbb{E}(\hat{N}_{h'}) \|_2 \leq \|K\|_1 \|E_{h^*}\|_2, \tag{4.3}$$

where

$$E_{h^*}(x) := (K_{h^*} \star N)(x) - N(x), \quad x \in \mathbb{R}_+$$

represents the approximation term. Combining (4.1), (4.2) and (4.3) entails

$$\|\hat{N} - N\|_2 \leq \|\hat{N}_{h^*} - N\|_2 + 2\|K\|_1 \|E_{h^*}\|_2 + \frac{2\chi}{\sqrt{nh^*}} \|K\|_2 + 2\zeta_n,$$

with

$$\begin{aligned}
\zeta_n &:= \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h^*,h'} - \mathbb{E}[\hat{N}_{h^*,h'}] - (\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}])\|_2 - \frac{\chi}{\sqrt{nh'}} \|K\|_2 \right\}_+ \\
&= \sup_{h' \in \mathcal{H}} \left\{ \|K_{h^*} \star (\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]) - (\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}])\|_2 - \frac{(1+\varepsilon)(1+\|K\|_1)}{\sqrt{nh'}} \|K\|_2 \right\}_+ \\
&\leq (1+\|K\|_1) \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]\|_2 - \frac{(1+\varepsilon)}{\sqrt{nh'}} \|K\|_2 \right\}_+ .
\end{aligned}$$

Hence

$$\mathbb{E}\left[\|\hat{N} - N\|_2^{2q} \right] \leq \square_q \left(\mathbb{E}\left[\|\hat{N}_{h^*} - N\|_2^{2q} \right] + \|K\|_1^{2q} \|E_{h^*}\|_2^{2q} + \chi^{2q} \frac{\|K\|_2^{2q}}{(nh^*)^q} + (1+\|K\|_1)^{2q} \mathbb{E}[\xi_n^{2q}] \right),$$

where

$$\xi_n = \sup_{h' \in \mathcal{H}} \left\{ \|\hat{N}_{h'} - \mathbb{E}(\hat{N}_{h'})\|_2 - \frac{(1+\varepsilon)}{\sqrt{nh'}} \|K\|_2 \right\}_+ .$$

Now, we have:

$$\begin{aligned}
\mathbb{E}\left[\|\hat{N}_{h^*} - N\|_2^{2q} \right] &\leq 2^{2q-1} \left(\mathbb{E}\left[\|\hat{N}_{h^*} - \mathbb{E}[\hat{N}_{h^*}]\|_2^{2q} \right] + \|\mathbb{E}[\hat{N}_{h^*}) - N\|_2^{2q} \right) \\
&\leq 2^{2q-1} \left(\mathbb{E}\left[\|\hat{N}_{h^*} - \mathbb{E}[\hat{N}_{h^*}]\|_2^{2q} \right] + \|E_{h^*}\|_2^{2q} \right).
\end{aligned}$$

Then, by setting

$$Kc_{h^*}(X_i, x) := K_{h^*}(x - X_i) - \mathbb{E}(K_{h^*}(x - X_1)),$$

we obtain

$$\begin{aligned}
\mathbb{E}\left[\|\hat{N}_{h^*} - \mathbb{E}[\hat{N}_{h^*}]\|_2^{2q} \right] &= \mathbb{E}\left[\left(\int \left(\frac{1}{n} \sum_{i=1}^n Kc_{h^*}(X_i, x) \right)^2 dx \right)^q \right] \\
&\leq \frac{2^{q-1}}{n^{2q}} \left(\mathbb{E}\left[\left(\sum_{i=1}^n \int Kc_{h^*}^2(X_i, x) dx \right)^q \right] \right. \\
&\qquad \left. + \mathbb{E}\left[\left| \sum_{1 \leq i,j \leq n \ i \neq j} \int Kc_{h^*}(X_i, x) Kc_{h^*}(X_j, x) dx \right|^q \right] \right).
\end{aligned}$$

22

Since

$$\int Kc_{h^*}^2(X_i, x)dx = \int \left(K_{h^*}(x - X_i) - \mathbb{E}\left[K_{h^*}(x - X_1) \right] \right)^2 dx$$

$$\leq 2\Big(\int K_{h^*}^2(x - X_i)dx + \int \left(\mathbb{E}\left[K_{h^*}(x - X_1) \right] \right)^2 dx \Big)$$

$$\leq 2\big(\|K_{h^*}\|_2^2 + \int \mathbb{E}\left[K_{h^*}^2(x - X_1) \right] dx \big)$$

$$\leq 4\|K_{h^*}\|_2^2 = \frac{4}{h^*}\|K\|_2^2,$$

the first term can be bounded as follows

$$\mathbb{E}\left[\big(\int \sum_{i=1}^n Kc_{h^*}^2(X_i, x)dx \big)^q \right] \leq \big(\frac{4n}{h^*}\|K\|_2^2 \big)^q.$$

For the second term, we apply Theorem 8.1.6 of de la Peña and Giné (1999) (with $2q \geq 2$) combined with the Cauchy-Schwarz inequality:

$$\mathbb{E}\left[\Big| \sum_{1 \leq i,j \leq n \ \ i \neq j} \int Kc_{h^*}(X_i, x)Kc_{h^*}(X_j, x)dx \Big|^q \right]$$

$$\leq \Big(\mathbb{E}\left[\Big| \sum_{1 \leq i,j \leq n \ \ i \neq j} \int Kc_{h^*}(X_i, x)Kc_{h^*}(X_j, x)dx \Big|^{2q} \right] \Big)^{\frac{1}{2}}$$

$$\leq \Box_q n^q \Big(\mathbb{E}\left[\big| \int Kc_{h^*}(X_1, x)Kc_{h^*}(X_2, x)dx \big|^{2q} \right] \Big)^{\frac{1}{2}}$$

$$\leq \Box_q n^q \Big(\mathbb{E}\left[\big| \int Kc_{h^*}^2(X_1, x)dx \big|^{2q} \right] \Big)^{\frac{1}{2}} \leq \Box_q \big(\frac{4n}{h^*}\|K\|_2^2 \big)^q.$$

It remains to deal with the term $\mathbb{E}(\xi_n^{2q})$. By Lemma 1 below, we obtain

$$\mathbb{E}[\xi_n^{2q}] \leq \Box_{q,\eta,\delta,\|K\|_2,\|K\|_1,\|N\|_\infty} n^{-q}$$

and the conclusion follows.

Proof of Proposition 3

The proof is similar to the previous one and we avoid most of the computations for simplicity. For any $h_0 \in \tilde{\mathcal{H}}$,

$$\|\hat{D}_{\tilde{h}} - D\|_2 \leq \|\hat{D}_{\tilde{h}} - \hat{D}_{\tilde{h},h_0}\|_2 + \|\hat{D}_{\tilde{h},h_0} - \hat{D}_{h_0}\|_2 + \|\hat{D}_{h_0} - D\|_2$$

$$\leq \tilde{A}_1 + \tilde{A}_2 + \tilde{A}_3,$$

23

with

$$\tilde{A}_1 := \|\hat{D}_{\tilde{h}} - \hat{D}_{\tilde{h},h_0}\|_2 \le \tilde{A}(h_0) + \frac{\tilde{\chi}}{\sqrt{n\tilde{h}^3}}\|g\|_\infty\|K'\|_2,$$

$$\tilde{A}_2 := \|\hat{D}_{\tilde{h},h_0} - \hat{D}_{h_0}\|_2 \le \tilde{A}(\tilde{h}) + \frac{\tilde{\chi}}{\sqrt{nh_0^3}}\|g\|_\infty\|K'\|_2$$

and

$$\tilde{A}_3 := \|\hat{D}_{h_0} - D\|_2.$$

Then,

$$\|\hat{D}_{\tilde{h}} - D\|_2 \le 2\tilde{A}(h_0) + \frac{2\tilde{\chi}}{\sqrt{nh_0^3}}\|g\|_\infty\|K'\|_2 + \|\hat{D}_{h_0} - D\|_2.$$

To study $\tilde{A}(h_0)$, we first evaluate

$$
\begin{aligned}
\mathbb{E}[\hat{D}_{h_1,h_2}(x)] - \mathbb{E}[\hat{D}_{h_2}(x)]. &= (K_{h_1} \star K_{h_2} \star (gN)')(x) - (K_{h_2} \star (gN)')(x) \\
&= \int D(t)(K_{h_1} \star K_{h_2})(x-t)dt - \int D(t)K_{h_2}(x-t)dt \\
&= \int D(t)\int K_{h_1}(x-t-u)K_{h_2}(u)du\,dt - \int D(t)K_{h_2}(x-t)dt \\
&= \int D(t)\int K_{h_1}(v-t)K_{h_2}(x-v)dv\,dt - \int D(v)K_{h_2}(x-v)dv \\
&= \int K_{h_2}(x-v)\Big(\int D(t)K_{h_1}(v-t)dt - D(v)\Big)dv \\
&= (K_{h_2} \star \tilde{E}_{h_1})(x),
\end{aligned}
$$

where we set, for any real number x

$$
\begin{aligned}
\tilde{E}_{h_1}(x) &:= \int D(t)K_{h_1}(x-t)dt - D(x) \\
&= (K_{h_1} \star D)(x) - D(x). \tag{4.4}
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\tilde{A}(h_0) &= \sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_{h_0,h} - \hat{D}_h\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ \\
&\le \sup_{h\in\tilde{\mathcal{H}}} \big\{\big\{\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}] - (\hat{D}_h - \mathbb{E}[\hat{D}_h])\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ \\
&\qquad + \|\mathbb{E}[\hat{D}_{h_0,h}] - \mathbb{E}[\hat{D}_h]\|_2\big\} \\
&\le \sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}] - (\hat{D}_h - \mathbb{E}[\hat{D}_h])\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ + \|K\|_1\|\tilde{E}_{h_0}\|_2 \\
&\le (1 + \|K\|_1)\sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2 - \frac{(1+\tilde{\varepsilon})}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ + \|K\|_1\|\tilde{E}_{h_0}\|_2, \tag{4.5}
\end{aligned}
$$

In order to obtain the last line, we use the following chain of arguments:

$$
\begin{aligned}
\hat{D}_{h_0,h}(x) &= \frac{1}{n}\sum_{i=1}^{n}g(X_i)\int K_h'(x - X_i - t)K_{h_0}(t)dt \\
&= \int K_{h_0}(t)\Big(\frac{1}{n}\sum_{i=1}^{n}g(X_i)K_h'(x - X_i - t)\Big)dt
\end{aligned}
$$

and

$$
\mathbb{E}\big[\hat{D}_{h_0,h}(x)\big] = \int K_{h_0}(t)\Big(\int g(u)K_h'(x - u - t)N(u)du\Big)dt,
$$

therefore

$$
\hat{D}_{h_0,h}(x) - \mathbb{E}\big[\hat{D}_{h_0,h}(x)\big] = \int K_{h_0}(t)G(x - t)dt = K_{h_0}\star G(x),
$$

with

$$
\begin{aligned}
G(x) &= \frac{1}{n}\sum_{i=1}^{n}g(X_i)K_h'(x - X_i) - \int g(u)K_h'(x - u)N(u)du \\
&= \hat{D}_h(x) - \mathbb{E}\big[\hat{D}_h(x)\big].
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}]\|_2 &\leq \|K_{h_0}\|_1\|G\|_2 \\
&\leq \|K\|_1\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2,
\end{aligned}
$$

which justifies (4.5). In the same way as in the proof of Proposition 2, we can establish the following:

$$
\mathbb{E}\big[\|\tilde{E}_{h_0}\|_2^{2q}\big] = \mathbb{E}\big[\|\hat{D}_{h_0} - D\|_2^{2q}\big] \leq \square_q\Big(\|\tilde{E}_{h_0}\|_2^{2q} + \Big(\frac{\|g\|_\infty\|K'\|_2}{\sqrt{nh_0^3}}\Big)^{2q}\Big).
$$

Finally, we successively apply (4.4), (4.5) and Lemma 1 in order to conclude the proof.

Proof of Proposition 4

We use the notation and definitions of Section 2.4. We have

$$
\|\mathcal{L}_k^{-1}(\varphi) - \mathcal{L}^{-1}(\varphi)\|_{2,T}^2 = \sum_{i=0}^{k-1}\int_{x_{i,k}}^{x_{i+1,k}}\big(H_{i,k} - \mathcal{L}^{-1}(\varphi)(x)\big)^2 dx := \sum_{i=0}^{k-1}L_{i,k}.
$$

We prove by induction that for all i, one has $L_{i,k} \leq C^2 \frac{T^2}{k^2} \|\varphi\|_{\mathcal{W}^1}^2$. The result follows by summation over i. We first prove the two following estimates:

$$\int_{x_{i,k}}^{x_{i+1,k}} \left(\varphi_{i,k} - \varphi(x)\right)^2 dx \leq \frac{T^2}{4\pi^2 k^2} \|\varphi\|_{\mathcal{W}^1}^2, \tag{4.6}$$

$$|\varphi_{i+1,k} - \varphi_{i,k}|^2 \leq \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2. \tag{4.7}$$

By definition, $\varphi_{i,k}$ is the average of the function φ on the interval $[x_{i,k}, \ x_{i+1,k}]$ of size $\frac{T}{k}$. Thus (4.7) is simply Wirtinger inequality applied to $\varphi \in \mathcal{W}^1$ on the interval $[x_{i,k}, \ x_{i+1,k}]. For (4.7)$, we use the Cauchy-Schwarz inequality:

$$|\varphi_{i+1,k} - \varphi_{i,k}|^2 = \frac{k^2}{T^2} \left(\int_{x_{i,k}}^{x_{i+1,k}} \left(\varphi(x + \frac{T}{k}) - \varphi(x)\right) dx\right)^2 = \frac{k^2}{T^2} \left(\int_{x_{i,k}}^{x_{i+1,k}} \int_{x}^{x+\frac{T}{k}} \varphi'(z) dz \, dx\right)^2$$

$$\leq \frac{k^2}{T^2} \left(\int_{x_{i,k}}^{x_{i+1,k}} \sqrt{\frac{T}{k}} \|\varphi\|_{\mathcal{W}^1} dx\right)^2 = \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2.$$

We are ready to prove by induction the two following inequalities:

$$L_{i,k} \leq C_1^2 \frac{T^2}{k^2} \|\varphi\|_{\mathcal{W}^1}^2, \tag{4.8}$$

$$|H_{i+1,k} - H_{i,k}|^2 \leq C_2^2 \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2. \tag{4.9}$$

for two constants C_1 and C_2 specified later on. First, for $i = 0$, we have

$$L_{0,k} = \int_0^{\frac{T}{k}} |H_{0,k}(\varphi) - \mathcal{L}^{-1}(\varphi)(x)|^2 dx = \int_0^{\frac{T}{k}} \left|\frac{1}{3}\varphi_{0,k} - \mathcal{L}^{-1}(\varphi)(x)\right|^2 dx.$$

We recall (see Proposition A.1. of [4]) that $\mathcal{L}^{-1}(\varphi)(x) = \sum_{n=1}^{\infty} 2^{-2n} \varphi(2^{-n}x)$, and we use the fact that $\frac{1}{3} = \sum_{n=1}^{\infty} 2^{-2n}$ and for $a, b > 0$, $ab \leq \frac{1}{2}(a^2 + b^2)$ in order to write

$$L_{0,k} = \int_0^{\frac{T}{k}} \left|\sum_{n=1}^{\infty} 2^{-2n} \left(\varphi_{0,k} - \varphi(2^{-n}x)\right)\right|^2 dx \leq \sum_{n,n'=1}^{\infty} 2^{-2n-2n'} \int_0^{\frac{T}{k}} |\varphi_{0,k} - \varphi(2^{-n}x)|^2 dx$$

$$\leq \frac{1}{3} \sum_{n=1}^{\infty} 2^{-n} \int_0^{2^{-n}\frac{T}{k}} |\varphi_{0,k} - \varphi(y)|^2 dy \leq \frac{1}{3} \frac{T^2}{4\pi^2 k^2} \|\varphi\|_{\mathcal{W}^1}^2.$$

26

This proves the first induction assumption for $i = 0$, and

$$|H_{1,k} - H_{0,k}|^2 = \left|\frac{1}{7}(\varphi_{1,k} - \varphi_{0,k})\right|^2 \leq \frac{1}{7^2}\frac{T}{k}\|\varphi\|_{\mathcal{W}^1}^2,$$

proves the second one. Let us now suppose that the two induction assumptions are true for all $j \leq i - 1$, and take $i \geq 1$. Let us first evaluate

$$L_{i,k} = \int_{x_{i,k}}^{x_{i+1,k}} \left(H_{i,k} - \mathcal{L}^{-1}(\varphi)(x)\right)^2 dx = \frac{1}{16}\int_{x_{i,k}}^{x_{i+1,k}} \left(H_{\frac{i}{2},k} + \varphi_{\frac{i}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2}) - \varphi(\frac{x}{2})\right)^2 dx.$$

We distinguish the case when i is even and when i is odd. Let i be even: then, by definition

$$L_{i,k} \leq \frac{1}{8}\int_{x_{i,k}}^{x_{i+1,k}} \left(H_{\frac{i}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2})\right)^2 dx + \frac{1}{8}\int_{x_{i,k}}^{x_{i+1,k}} \left(\varphi_{\frac{i}{2},k} - \varphi(\frac{x}{2})\right)^2 dx \leq \frac{1}{4}(C_1^2 + \frac{1}{4\pi^2})\frac{T^2}{k^2}\|\varphi\|_{\mathcal{W}^1}^2$$

by the induction assumption and Assertion 4.7 on φ for $j = \frac{i}{2}$. If i is odd, we write by definition

$$L_{i,k} = \frac{1}{16}\int_{x_{i,k}}^{x_{i+1,k}} \left(H_{\frac{i-1}{2},k} + \varphi_{\frac{i-1}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2}) - \varphi(\frac{x}{2}) + \frac{1}{2}(H_{\frac{i+1}{2},k} - H_{\frac{i-1}{2},k}) + \frac{1}{2}(\varphi_{\frac{i+1}{2},k} - \varphi_{\frac{i-1}{2},k})\right)^2 dx.$$

Hence, re-organizing terms, we can write

$$L_{i,k} \leq \frac{1}{2}\int_{x_{\frac{i-1}{2},k}}^{x_{\frac{i-1}{2}+1,k}} \left(H_{\frac{i-1}{2},k} - \mathcal{L}^{-1}(\varphi)(y)\right)^2 dy + \frac{1}{2}\int_{x_{\frac{i-1}{2},k}}^{x_{\frac{i-1}{2}+1,k}} \left(\varphi_{\frac{i-1}{2},k} - \varphi(y)\right)^2 dy$$

$$+ \frac{1}{16}\frac{T}{k}(H_{\frac{i+1}{2},k} - H_{\frac{i-1}{2},k})^2 + \frac{1}{16}\frac{T}{k}(\varphi_{\frac{i+1}{2},k} - \varphi_{\frac{i-1}{2},k})^2$$

Putting together the four inequalities above (the estimates for φ and the induction assumptions), we obtain

$$L_{i,k} \leq \frac{T^2}{k^2}\|\varphi\|_{\mathcal{W}^1}^2 \left(\frac{C_1^2}{2} + \frac{1}{8\pi^2} + \frac{C_2^2}{16} + \frac{1}{16}\right)$$

and (4.8) is proved. It remains to establishe (4.9). Let us write it for i even (the case of an odd i is similar):

$$|H_{i+1,k} - H_{i,k}|^2 = \frac{1}{16}|H_{\frac{i+1}{2}} - H_{\frac{i}{2}} + \varphi_{\frac{i+1}{2}} - \varphi_{\frac{i}{2}}|^2 = \frac{1}{32}|H_{\frac{i}{2}+1} - H_{\frac{i}{2}} + \varphi_{\frac{i}{2}+1} - \varphi_{\frac{i}{2}}|^2.$$

Hence, as previously, we obtain

$$|H_{i+1,k} - H_{i,k}|^2 \leq \frac{1}{16}\frac{T}{k}\|\varphi\|_{\mathcal{W}^1}^2(C_2^2 + 1).$$

To complete the proof, we remark that $C_2^2 = \frac{1}{15}$ and $C_1^2 = \frac{1}{4\pi^2} + \frac{1}{8}(1 + \frac{1}{15}) < \frac{1}{6}$ are suitable. It is consequently sufficient to take $C = C_1$.

4.2 Proof of Theorem 1 and Proposition 1

Proof of Theorem 1

It is easy to see that

$$
\begin{aligned}
\|\widehat{H} - H\|_{2,T} &= \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D} + \hat{\lambda}_n\hat{N}) - \mathcal{L}^{-1}(\mathcal{L}(BN))\|_{2,T} \\
&\leq \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D} + \hat{\lambda}_n\hat{N}) - \mathcal{L}_k^{-1}(\mathcal{L}(BN))\|_{2,T} \\
&\qquad + \|\mathcal{L}_k^{-1}(\mathcal{L}(BN)) - \mathcal{L}^{-1}(\mathcal{L}(BN))\|_{2,T} \\
&\leq \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D} + \hat{\lambda}_n\hat{N} - (\kappa D + \lambda N))\|_{2,T} \\
&\qquad + \frac{1}{3}\frac{T}{\sqrt{k}}\|\mathcal{L}(BN)\|_{\mathcal{W}^1},
\end{aligned}
$$

thanks to Proposition 4. Note that $\mathcal{L}(BN) = \kappa(gN)' + \lambda N$ so that we can write

$$\|\mathcal{L}(BN)\|_{\mathcal{W}^1} \leq C(\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2}).$$

We obtain, thanks to Lemma 4 that gives the boundedness of the operator $\mathcal{L}_k^{-1}$:

$$
\begin{aligned}
\|\widehat{H} - H\|_{2,T} &\leq \square\big(\big\|\hat{\kappa}_n\hat{D} - \kappa D\big\|_{2,T} + \big\|\hat{\lambda}_n\hat{N} - \lambda N\big\|_{2,T} + (\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big) \\
&\leq \square\big(|\hat{\kappa}_n|\|\hat{D} - D\|_2 + |\hat{\lambda}_n|\|\hat{N} - N\|_2 + |\hat{\kappa}_n - \kappa|\|D\|_2 + |\hat{\lambda}_n - \lambda|\|N\|_2 \\
&\qquad + (\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big) \\
&\leq \square\big(|\hat{\lambda}_n||\hat{\rho}_n|\|\hat{D} - D\|_2 + |\hat{\lambda}_n|(\|\hat{N} - N\|_2 + |\hat{\rho}_n - \rho_g(N)|\|D\|_2) \\
&\qquad + (\|N\|_2 + |\rho_g(N)|\|D\|_2)|\hat{\lambda}_n - \lambda| + (\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big).
\end{aligned}
$$

Taking expectation and using Cauchy-Schwarz inequality, we obtain for any $q \geq 1$,

$$
\begin{aligned}
\mathbb{E}[\|\widehat{H} - H\|_{2,T}^q] &\leq \square_q\Big[\big(\mathbb{E}[\hat{\lambda}_n^{2q}]\big)^{1/2}\big\{\big(\mathbb{E}[\hat{\rho}_n^{4q}]\big)^{1/4}\big(\mathbb{E}[\|\hat{D} - D\|_2^{4q}]\big)^{1/4} + \big(\mathbb{E}[\|\hat{N} - N\|_2^{2q}]\big)^{1/2} \\
&\qquad + \|D\|_2^q\big(\mathbb{E}[|\hat{\rho}_n - \rho_g(N)|^{2q}]\big)^{1/2}\big\} \\
&\qquad + \big(\|N\|_2 + \rho_g(N)\|D\|_2\big)^q\mathbb{E}[|\hat{\lambda}_n - \lambda|^q] + \big((\|N\|_{\mathcal{W}^1} + \|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big)^q\Big].
\end{aligned}
$$

Now, Lemma 3 gives the behaviour of $\mathbb{E}[|\hat{\rho}_n - \rho_g(N)|^{2q}]$. In particular, we obtain

$$\mathbb{E}[\hat{\rho}_n^{4q}] \leq \square_{q,g,N,c}.$$

We finally apply successively Propositions 2 and 3 to obtain the proof of Theorem 1.

Proof of Proposition 1

We have already proved (1.9). It remains to prove (1.10). We introduce the event

$$\Omega_n = \{2\hat{N}(x) \geq m \text{ for any } x \in [a,b]\}.$$

Then, for n larger that Q^2,

$$\mathbb{E}\left[\left(\int_a^b \left(\tilde{B}(x) - B(x)\right)^2 dx\right)^{\frac{q}{2}}\right]$$

$$= \mathbb{E}\left[\left(\int_a^b \left(\tilde{B}(x) - B(x)\right)^2 dx \times \mathbf{1}_{\Omega_n}\right)^{\frac{q}{2}}\right] + \mathbb{E}\left[\left(\int_a^b \left(\tilde{B}(x) - B(x)\right)^2 dx \times \mathbf{1}_{\Omega_n^c}\right)^{\frac{q}{2}}\right]$$

$$\leq \mathbb{E}\left[\left(\int_a^b \left(\hat{B}(x) - B(x)\right)^2 dx \times \mathbf{1}_{\Omega_n}\right)^{\frac{q}{2}}\right] + \left(2(b-a)(n+Q^2)\right)^{\frac{q}{2}} \mathbb{P}(\Omega_n^c)$$

$$\leq \mathbb{E}\left[\left(\int_a^b \left(\frac{\hat{H}}{\hat{N}} - \frac{H}{N}\right)^2 \times \mathbf{1}_{\Omega_n}\right)^{\frac{q}{2}}\right] + \left(4n(b-a)\right)^{\frac{q}{2}} \mathbb{P}(\Omega_n^c)$$

$$\leq \mathbb{E}\left[\left(\int_a^b \left(\frac{\hat{H}N - \hat{N}H}{\hat{N}N}\right)^2 \times \mathbf{1}_{\Omega_n}\right)^{\frac{q}{2}}\right] + \left(4n(b-a)\right)^{\frac{q}{2}} \mathbb{P}(\Omega_n^c)$$

$$\leq \square_{q,m,M,Q}\left(\mathbb{E}\left[\left\|\hat{H} - H\right\|_2^q\right] + \mathbb{E}\left[\left\|\hat{N} - N\right\|_2^q\right]\right) + \left(4n(b-a)\right)^{\frac{q}{2}} \mathbb{P}(\Omega_n^c).$$

The first term of the right hand side is handled by (1.9) and Proposition 2. The second term is handled by Lemma 2 that establishes that $\mathbb{P}(\Omega_n^c) = O(n^{-q})$.

4.3 Technical lemmas

Concentration inequalities

We first state the following concentration result. Note that a more general version of this result can be found in [13]. We nevertheless give a proof for the sake of completeness.

Lemma 1. *We have the following estimates*

- *Assume that $\|K\|_2, \|K\|_1, \|g\|_\infty$ and $\|N\|_\infty$ are finite. For every $q > 0$, introduce the grid $\mathcal{H} \subset \{1, 1/2, ..., 1/D_{\max}\}$ and $D_{\max} = \delta n$ for some $\delta > 0$. Then, for every $\eta > 0$*

$$\mathbb{E}\left[\sup_{h \in \mathcal{H}} \left\{\|\hat{N}_h - \mathbb{E}[\hat{N}_h]\|_2 - \frac{(1+\eta)}{\sqrt{nh}}\|K\|_2\right\}_+^{2q}\right] \leq \square_{q,\eta,\delta\|K\|_2,\|K\|_1,\|N\|_\infty} n^{-q}.$$

- *Assume that $\|K'\|_2, \|K'\|_1, \|g\|_\infty$ and $\|N\|_\infty$ are finite. For every $q > 0$, introduce the grid $\tilde{\mathcal{H}} \subset \{1, 1/2, ..., 1/\tilde{D}_{\max}\}$ and $\tilde{D}_{\max} = \sqrt{\tilde{\delta}n}$ for some $\tilde{\delta} > 0$. Then for every $\eta > 0$*

$$\mathbb{E}\Big[\sup_{h\in\tilde{\mathcal{H}}}\big\{\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2 - \frac{(1+\eta)}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+^{2q}\Big] \leq \Box_{q,\eta,\tilde{\delta},\|K'\|_2,\|K'\|_1,\|g\|_\infty,\|N\|_\infty} n^{-q}.$$

Proof. Let X be a real random variable and let us consider the random process

$$\forall t \in \mathbb{R}, \quad w(t,X) = \varphi(X)\Psi(t-X),$$

where φ and Ψ are measurable real-valued functions. Let $X_1, ..., X_n$ be n independent and identically distributed random variables with the same distribution as X and let us consider the process

$$\forall t \in \mathbb{R}, \quad \xi_{\varphi,\Psi}(t) = \sum_{i=1}^{n}\big(w(t,X_i) - \mathbb{E}\big[w(t,X)\big]\big).$$

First, let us study the behavior of

$$\|\xi_{\varphi,\Psi}\|_2 = \Big(\int \xi_{\varphi,\Psi}^2(t)dt\Big)^{1/2}.$$

If $\mathcal{B}$ denotes the unit ball in $\mathbb{L}^2$ and $\mathcal{A}$ is a countable subset of $\mathcal{B}$, we have

$$
\begin{aligned}
\|\xi_{\varphi,\Psi}\|_2 &= \sup_{a\in\mathcal{B}}\int a(t)\xi_{\varphi,\Psi}(t)dt \\
&= \sup_{a\in\mathcal{A}}\int a(t)\xi_{\varphi,\Psi}(t)dt \\
&= \sup_{a\in\mathcal{A}}\sum_{i=1}^{n}\int a(t)\big(w(t,X_i) - \mathbb{E}\big[w(t,X)\big]\big)dt.
\end{aligned}
$$

Hence one can apply Talagrand's inequality (see the version of Bousquet in the independent and identically distributed case [16, p 170]). For all $\varepsilon, x > 0$, one has:

$$\mathbb{P}\big(\|\xi_{\varphi,\Psi}\|_2 \geq (1+\varepsilon)\,\mathbb{E}[\|\xi_{\varphi,\Psi}\|_2] + \sqrt{2vx} + c(\varepsilon)bx\big) \leq e^{-x},$$

where $c(\varepsilon) = 1/3 + \varepsilon^{-1}$,

$$v = n\sup_{a\in\mathcal{A}}\mathbb{E}\Big[\big(\int a(t)\big[w(t,X) - \mathbb{E}\big(w(t,X)\big)\big]dt\big)^2\Big],$$

and

$$b = \sup_{y\in\mathbb{R},a\in\mathcal{A}}\int a(t)\big[w(t,y) - \mathbb{E}\big(w(t,X)\big)\big]dt.$$

We study each term of the right hand term within the expectation.

30

- Obviously, one has:

$$\mathbb{E}\left[\|\xi_{\varphi,\Psi}\|_2\right] \leq \left(\mathbb{E}\left[\int \left(\sum_{i=1}^{n}\left(w(t,X_i) - \mathbb{E}\left[w(t,X)\right]\right)\right)^2 dt\right]\right)^{1/2}$$

$$= \left(\int \sum_{i=1}^{n} \mathbb{E}\left[\left(w(t,X_i) - \mathbb{E}\left[w(t,X)\right]\right)^2\right] dt\right)^{1/2}$$

$$\leq \left(n \int \mathbb{E}\left[w(t,X)^2\right] dt\right)^{1/2}.$$

But we easily see that for all y,

$$\int w^2(t,y)dt \leq \|\varphi\|_\infty^2 \|\Psi\|_2^2, \tag{4.10}$$

hence

$$\mathbb{E}[\|\xi_{\varphi,\Psi}\|_2] \leq \sqrt{n}\|\varphi\|_\infty\|\Psi\|_2.$$

- Since v is a supremum of variance terms,

$$\begin{aligned}
v &\leq n\sup_{a\in\mathcal{A}}\mathbb{E}\left[\left(\int a(t)w(t,X)dt\right)^2\right]\\
&\leq n\sup_{a\in\mathcal{A}}\mathbb{E}\left[\int |w(t,X)|dt \int a^2(t)|w(t,X)|dt\right]\\
&\leq n\sup_{y\in\mathbb{R}}\int |w(t,y)|dt \times \sup_{t\in\mathbb{R}}\mathbb{E}[|w(t,X)|]\\
&\leq n\|\varphi\|_\infty^2\|N\|_\infty\|\Psi\|_1^2.
\end{aligned}$$

- The Cauchy-Schwarz inequality and (4.10) give

$$\begin{aligned}
b &= \sup_{y\in\mathbb{R}}\|w(\bullet,y) - \mathbb{E}[w(.,X)]\|_2\\
&\leq \sup_{y\in\mathbb{R}}\|w(\bullet,y)\|_2 + \left(\mathbb{E}\left[\int w(t,X)^2 dt\right]\right)^{1/2}\\
&\leq 2\|\varphi\|_\infty\|\Psi\|_2.
\end{aligned}$$

The main point here is that $\sqrt{v}$ may be much smaller than $\mathbb{E}[\|\xi_{\varphi,\Psi}\|_2]$. So, for all $\varepsilon, x > 0$,

$$\mathbb{P}\left(\|\xi_{\varphi,\Psi}\|_2 \geq (1+\varepsilon)\sqrt{n}\|\varphi\|_\infty\|\Psi\|_2 + \|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\sqrt{2nx} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2 x\right) \leq e^{-x}.$$

Now we consider a family $\mathcal{M}$ of possible functions φ and Ψ. Let us introduce some strictly positive weights $L_{\varphi,\Psi}$ and let us apply the previous inequality to $x = L_{\varphi,\Psi} + u$ for $u > 0$. Hence with probability larger than $1 - \sum_{(\varphi,\Psi)\in\mathcal{M}} e^{-L_{\varphi,\Psi}} e^{-u}$, for all $(\varphi,\Psi) \in \mathcal{M}$, one has

$$
\begin{aligned}
\|\xi_{\varphi,\Psi}\|_2 \leq & \ (1+\varepsilon)\sqrt{n}\|\varphi\|_\infty\|\Psi\|_2 + \|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\sqrt{2nL_{\varphi,\Psi}} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2 L_{\varphi,\Psi} \\
& + \|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\sqrt{2nu} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2 u. \quad (4.11)
\end{aligned}
$$

Let

$$
M_{\varphi,\Psi} = (1+\varepsilon)\sqrt{n}\|\varphi\|_\infty\|\Psi\|_2 + \|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\sqrt{2nL_{\varphi,\Psi}} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2 L_{\varphi,\Psi}.
$$

It is also easy to obtain an upper bound of R_q for any $q \geq 1$ with

$$
R_q = \mathbb{E}\Big[\sup_{(\varphi,\Psi)\in\mathcal{M}}\big(\|\xi_{\varphi,\Psi}\|_2 - M_{\varphi,\Psi}\big)_+^{2q}\Big] = \int_0^{+\infty}\mathbb{P}\Big(\sup_{(\varphi,\Psi)\in\mathcal{M}}\big(\|\xi_{\varphi,\Psi}\|_2 - M_{\varphi,\Psi}\big)_+^{2q} \geq x\Big)dx.
$$

Indeed

$$
R_q \leq \sum_{(\varphi,\Psi)\in\mathcal{M}}\int_0^{+\infty}\mathbb{P}\Big(\big(\|\xi_{\varphi,\Psi}\|_2 - M_{\varphi,\Psi}\big)_+^{2q} \geq x\Big)dx.
$$

Then, let us take u such that

$$
x = f(u)^{2q} := \big(\|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\sqrt{2nu} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2 u\big)^{2q},
$$

so

$$
dx = 2q\big(f(u)\big)^{2q-1}\Big(\sqrt{2n}\|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\frac{1}{2\sqrt{u}} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2\Big)du.
$$

Hence

$$
\begin{aligned}
R_q \leq & \sum_{(\varphi,\Psi)\in\mathcal{M}}\int_0^{+\infty} e^{-(L_{\varphi,\Psi}+u)}2q\big(f(u)\big)^{2q-1}\Big(\sqrt{2n}\|\varphi\|_\infty\|N\|_\infty^{1/2}\|\Psi\|_1\frac{1}{2\sqrt{u}} + 2c(\varepsilon)\|\varphi\|_\infty\|\Psi\|_2\Big)du \\
\leq & \ 2q\sum_{(\varphi,\Psi)\in\mathcal{M}} e^{-L_{\varphi,\Psi}}\int_0^{+\infty} f(u)^{2q}e^{-u}u^{-1}du, \\
\leq & \ \Box_{q,\varepsilon}\sum_{(\varphi,\Psi)\in\mathcal{M}} e^{-L_{\varphi,\Psi}}\Big[n^q\|N\|_\infty^q\|\varphi\|_\infty^{2q}\|\Psi\|_1^{2q}\int_0^{+\infty} e^{-u}u^{q-1}du + \|\varphi\|_\infty^{2q}\|\Psi\|_2^{2q}\int_0^{+\infty} e^{-u}u^{2q-1}du\Big].
\end{aligned}
$$

Finally, we have proved that

$$
R_q \leq \Box_{q,\varepsilon}\sum_{(\varphi,\Psi)\in\mathcal{M}} e^{-L_{\varphi,\Psi}}\Big[n^q\|N\|_\infty^q\|\varphi\|_\infty^{2q}\|\Psi\|_1^{2q} + \|\varphi\|_\infty^{2q}\|\Psi\|_2^{2q}\Big]. \quad (4.12)
$$

Now let us evaluate what this inequality means for each set-up.

- First, when $\varphi = 1/n$ and $\Psi = 1/hK(\bullet/h)$, the family $\mathcal{M}$ corresponds to the family $\mathcal{H}$. In that case $M_{\varphi,\Psi}$ and $L_{\varphi,\Psi}$ will respectively be denoted by M_h and L_h. The upper bound given in (4.12) becomes

$$R_q \leq \Box_{q,\varepsilon} \sum_{h \in \mathcal{H}} e^{-L_h} \left[n^{-q} \|N\|_\infty^q \|K\|_1^{2q} + n^{-2q} h^{-q} \|K\|_2^{2q} \right]. \qquad (4.13)$$

Now it remains to choose L_h. But

$$M_h = (1+\varepsilon) \frac{1}{\sqrt{nh}} \|K\|_2 + \|N\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{2L_h}{n}} + 2c(\varepsilon) \frac{1}{n\sqrt{h}} \|K\|_2 L_h.$$

Let $\theta > 0$ and let

$$L_h = \frac{\theta^2 \|K\|_2^2}{2\|N\|_\infty \|K\|_1^2 \sqrt{h}}.$$

Obviously the series in (4.13) is finite and for any $h \in \mathcal{H}$, since $h \leq 1$, we have:

$$M_h \leq (1+\varepsilon+\theta) \frac{\|K\|_2}{\sqrt{nh}} + \frac{c(\varepsilon)\theta^2 \|K\|_2^3}{\|N\|_\infty \|K\|_1^2} \frac{1}{nh}.$$

Since $D_{\max} = \delta n$, one obtains that

$$M_h \leq \left(1 + \varepsilon + \theta + \frac{c(\varepsilon)\theta^2 \|K\|_2^3 \sqrt{\delta}}{\|N\|_\infty \|K\|_1^2} \right) \frac{\|K\|_2}{\sqrt{nh}}.$$

It remains to choose $\varepsilon = \eta/2$ and θ small enough such that $\theta + \frac{c(\varepsilon)\theta^2 \|K\|_2^3 \sqrt{\delta}}{\|N\|_\infty \|K\|_1^2} = \eta/2$ to obtain the desired inequality.

- Secondly, if $\varphi = g/n$ and $\Psi = 1/h^2 K'(\bullet/h)$ the family $\mathcal{M}$ corresponds to the family $\tilde{\mathcal{H}}$. So, $M_{\varphi,\Psi}$ and $L_{\varphi,\Psi}$ will be denoted by M'_h and L'_h respectively. The upper bound given by (4.12) now becomes

$$R_q \leq \Box_{q,\varepsilon} \sum_{h \in \tilde{\mathcal{H}}} e^{-L'_h} \left[n^{-q} h^{-2q} \|N\|_\infty^q \|g\|_\infty^{2q} \|K'\|_1^{2q} + n^{-2q} h^{-3q} \|g\|_\infty^{2q} \|K'\|_2^{2q} \right]. \quad (4.14)$$

But

$$M'_h = (1+\varepsilon) \frac{1}{\sqrt{nh^{3/2}}} \|g\|_\infty \|K'\|_2 + \|g\|_\infty \|N\|_\infty^{1/2} \|K'\|_1 \sqrt{\frac{2L'_h}{nh^2}} + 2c(\varepsilon) \|g\|_\infty \|K'\|_2 \frac{L'_h}{nh^{3/2}}.$$

Let $\theta > 0$ and let

$$L'_h = \frac{\theta^2 \|K'\|_2^2}{2\|N\|_\infty \|K'\|_1^2 h}.$$

33

Obviously the series in (4.14) is finite and we have:

$$M_h' \le (1 + \varepsilon + \theta)\|g\|_\infty \frac{\|K'\|_2}{\sqrt{nh^3}} + \frac{c(\varepsilon)\theta^2\|K'\|_2^3\|g\|_\infty}{\|N\|_\infty\|K'\|_1^2} \frac{1}{nh^{5/2}}.$$

But $h^2 \ge (\tilde{\delta}n)^{-1}$. Hence

$$M_h' \le \left(1 + \varepsilon + \theta + \frac{c(\varepsilon)\theta^2\|K'\|_2^3\|g\|_\infty\sqrt{\tilde{\delta}}}{\|N\|_\infty\|K'\|_1^2}\right)\frac{\|K'\|_2}{\sqrt{nh^3}}.$$

As previously, it remains to choose ε and θ accordingly to conclude.

∎

The second result is based on probabilistic arguments as well.

Lemma 2. *Under Assumptions and notations of Proposition 1, if there exists an interval $[a, b]$ in $(0, T)$ such that*

$$[m, M] := [\inf_{x\in[a,b]} N(x), \sup_{x\in[a,b]} N(x)] \subset (0, \infty), \qquad Q := \sup_{x\in[a,b]} |H(x)| < \infty,$$

and if $\ln(n) \le D_{\min} \le n^{1/(2m_0+1)}$ and $n^{1/5} \le D_{\max} \le (n/\ln(n))^{1/(4+\eta)}$ for some $\eta > 0$ fixed, then there exists C_η, a constant depending on η, such that for n large enough,

$$\mathbb{P}(\Omega_n^c) \le C_\eta n^{-q}.$$

Proof. Let x, y be two fixed point of $[a, b]$ and for $h \in \mathcal{H}'$, first let us look at

$$Z_h(x, y) = \hat{N}_h(x) - \hat{N}_h(y) - [K_h \star N(x) - K_h \star N(y)].$$

One can apply Bernstein inequality to $Z_h(x, y)$ (see *e.g.* (2.10) and (2.21) of [16]) to obtain:

$$\mathbb{E}\left[e^{\lambda Z_h(x,y)}\right] \le \exp\left[\frac{\lambda^2 v^2(x, y)}{2(1 - \lambda c(x, y))}\right], \quad \forall \lambda \in (0, 1/c(x, y)),$$

with

$$c(x, y) = \frac{1}{3n} \sup_{u\in\mathbb{R}} |K_h(x - u) - K_h(y - u)|$$

and

$$v^2(x, y) = \frac{1}{n} \int_{\mathbb{R}} |K_h(x - u) - K_h(y - u)|^2 N(u)du.$$

But, with $\mathcal{K}$ the Lipschitz constant of K,

$$c(x, y) \le \frac{\mathcal{K}}{3nh^2}|x - y|$$

34

and

$$v^2(x, y) \leq \frac{\mathcal{K}^2}{nh^4}|x - y|^2.$$

Let x_0 be a fixed point of $[a, b]$. One can apply Theorem 2.1 of [2] which gives that, if

$$Z_h = \sup_{x \in [a,b]} |Z_h(x, x_0)|$$

then for all positive x,

$$\mathbb{P}(Z_h \geq 18(\sqrt{v^2(x+1)} + c(x+1))) \leq 2e^{-x},$$

with

$$v^2 = \frac{\mathcal{K}^2}{nh^4}|b - a|^2$$

and

$$c = \frac{\mathcal{K}}{3nh^2}|b - a|.$$

But one can similarly prove, using Bernstein's inequality, that

$$\mathbb{P}\left(|\hat{N}_h(x_0) - K_h \star N(x_0)| \geq \sqrt{\frac{\square_{N,K} x}{nh}} + \square_K \frac{x}{nh}\right) \leq 2e^{-x}.$$

Hence

$$\mathbb{P}\left(\exists x \in [a, b], |\hat{N}_h(x) - K_h \star N(x)| \geq \sqrt{\frac{\square_{N,K} x}{nh^4}} + \square_K \frac{x}{nh^2}\right) \leq 4e^{-x}.$$

We apply this inequality with $x = q\log(n) + D^\eta$, $D = D'_{\min}, \ldots, D'_{\max}$. We obtain

$$\mathbb{P}\left(\exists h \in \mathcal{H}', \exists x \in [a, b], |\hat{N}_h(x) - K_h \star N(x)| \geq \sqrt{\frac{\square_{N,K,q} \log(n)}{nh^{4+\eta}}} + \square_{K,q} \frac{\log(n)}{nh^{2+\eta}}\right)$$

$$\leq 4 \sum_{D=1}^{D_{\max}} e^{-D^\eta} n^{-q} = \square_\eta n^{-q}.$$

Consequently, since $\frac{\log(n)}{nh^{4+\eta}} \to 0$ uniformly in $h \in \mathcal{H}'$, for n large enough,

$$\mathbb{P}(\exists h \in \mathcal{H}', \exists x \in [a, b], |\hat{N}_h(x) - K_h \star N(x)| \geq m/4) \leq \square_\eta n^{-q}.$$

But since $N \in \mathcal{W}^{s+1}$, $s \geq 1$, then $N' \in \mathcal{W}^s$ so N' is bounded on $\mathbb{R}$. Then

$$
\begin{aligned}
K_h \star N(x) - N(x) &= \int K_h(x - y)(N(y) - N(x))dy \\
&= \int K\left(\frac{x - y}{h}\right)\left(\frac{y - x}{h}\right) \int_0^1 N'(x + u(y - x))du\,dy \\
&= h \int tK(t) \int_0^1 N'(x - uht)du\,dt.
\end{aligned}
$$

35

So,
$$|N(x) - K_h \star N(x)| \leq \square_{N,K} h.$$

Because of the definition of $\mathcal{H}'$ this term tends to 0, uniformly in x and h. Hence for n large enough, for all $x \in [a,b]$ and all $h \in \mathcal{H}'$
$$K_h \star N(x) \geq N(x) - m/4 \geq 3m/4.$$

Consequently, for n large enough,
$$\mathbb{P}(\exists h \in \mathcal{H}', \exists x \in [a,b], 2\hat{N}_h(x) < m) \leq \square_\eta n^{-q}.$$

■

Rosenthal-type inequality

The following result studies the behavior of the moments of $\hat{\rho}_n$.

Lemma 3. *For any $p \geq 2$,*
$$\mathbb{E}\left[|\hat{\rho}_n - \rho_g(N)|^p\right] \leq \square_{p,g,N,c} n^{-p/2}.$$

Proof. We have:
$$\mathbb{E}\left[|\hat{\rho}_n - \rho_g(N)|^p\right] \leq \square_p(A_n + B_n),$$

with
$$A_n := \mathbb{E}\left[\left|\frac{\sum_{i=1}^n X_i}{n \int g(x)N(x)dx + c} - \frac{\int_{\mathbb{R}_+} xN(x)dx}{\int_{\mathbb{R}_+} g(x)N(x)dx}\right|^p\right]$$

and
$$B_n := \mathbb{E}\left[\left|\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n g(X_i) + c} - \frac{\sum_{i=1}^n X_i}{n \int g(x)N(x)dx + c}\right|^p\right].$$

We use the Rosenthal inequality (see *e.g.* the textbook [7]): if $R_1, \ldots, R_n$ are independent centered variables such that $\mathbb{E}[|R_1|^p] < \infty$, with $p \geq 2$ then

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n R_i\right|^p\right] \leq \square_p n^{-p}\left(n\,\mathbb{E}\,|R_1|^p + \left(n\,\mathbb{E}\,R_1^2\right)^{p/2}\right). \tag{4.15}$$

36

We recall that Assumption 1 ensures that $\mathbb{E}[X_i^p] = \int x^p N(x)dx < \infty$, for any $p \geq 1$. Hence, for the first term A_n, using (4.15), we have:

$$
\begin{aligned}
A_n &= \mathbb{E}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^{n}X_i}{\int g(x)N(x)dx + \frac{c}{n}} - \frac{\int_{\mathbb{R}_+}xN(x)dx}{\int_{\mathbb{R}_+}g(x)N(x)dx}\right|^p\right] \\
&\leq \square_{g,N,c}\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i\int_{\mathbb{R}_+}g(x)N(x)dx - \int_{\mathbb{R}_+}xN(x)dx\left(\int g(x)N(x)dx + \frac{c}{n}\right)\right|^p\right] \\
&\leq \square_{p,g,N}\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i - \int_{\mathbb{R}_+}xN(x)dx\right|^p\right] + \square_{g,N,c}\,n^{-p} \\
&\leq \square_{p,g,N,c}\,n^{-p/2}.
\end{aligned}
$$

Let us turn to the term B_n:

$$
\begin{aligned}
B_n &= \mathbb{E}\left[\left|\frac{\sum_{i=1}^{n}X_i}{\sum_{i=1}^{n}g(X_i)+c} - \frac{\sum_{i=1}^{n}X_i}{n\int g(x)N(x)dx + c}\right|^p\right] \\
&\leq \left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|^{2p}\right]\right)^{1/2} \times \left(\mathbb{E}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^{n}g(X_i) - \int g(x)N(x)dx}{\left(\frac{1}{n}\sum_{i=1}^{n}g(X_i)+\frac{c}{n}\right)\left(\int g(x)N(x)dx+\frac{c}{n}\right)}\right|^{2p}\right]\right)^{1/2} \\
&\leq \|gN\|_1^{-p}\left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right|^{2p}\right]\right)^{1/2}\left(\mathbb{E}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^{n}g(X_i)-\int g(x)N(x)dx}{\frac{1}{n}\sum_{i=1}^{n}g(X_i)+\frac{c}{n}}\right|^{2p}\right]\right)^{1/2} \\
&\leq \square_{p,g,N}\left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}X_i - \mathbb{E}[X_1]\right|^{2p} + \left|\mathbb{E}[X_1]\right|^{2p}\right)^{1/2} \times \\
&\quad\quad \left(\mathbb{E}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^{n}g(X_i)-\int g(x)N(x)dx}{\frac{1}{n}\sum_{i=1}^{n}g(X_i)+\frac{c}{n}}\right|^{2p}\right]\right)^{1/2} \\
&\leq \square_{p,g,N}\left(\mathbb{E}\left[\left|\frac{\frac{1}{n}\sum_{i=1}^{n}g(X_i)-\int g(x)N(x)dx}{\frac{1}{n}\sum_{i=1}^{n}g(X_i)+\frac{c}{n}}\right|^{2p}\right]\right)^{1/2}.
\end{aligned}
$$

Now, we set for $\gamma > 3p$

$$
\Omega_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n}g(X_i) - \int g(x)N(x)dx\right| \leq \sqrt{\frac{2\gamma\,\mathrm{Var}(g(X_1))\log n}{n}} + \frac{\gamma\|g\|_\infty\log n}{3n}\right\}
$$

(recall that Assumption 1 states that $\|g\|_\infty < \infty$, which also implies $\mathbb{E}[g(X_1)^2] < \infty$). Since g is positive, the Bernstein inequality (see Section 2.2.3 of [16]) gives:

$$
\mathbb{P}(\Omega_n) \geq 1 - 2n^{-\gamma}. \tag{4.16}
$$

Therefore, we bound from above the term B_n by a constant $\square_{p,g,N}$ times

$$\big(\mathbb{E} \big[\big| \frac{\frac{1}{n}\sum_{i=1}^n g(X_i) - \int g(x)N(x)dx}{\frac{1}{n}\sum_{i=1}^n g(X_i) + \frac{c}{n}} \big|^{2p} \mathbf{1}_{\Omega_n} + \big| \frac{\frac{1}{n}\sum_{i=1}^n g(X_i) - \int g(x)N(x)dx}{\frac{1}{n}\sum_{i=1}^n g(X_i) + \frac{c}{n}} \big|^{2p} \mathbf{1}_{\Omega_n^c} \big] \big)^{1/2}$$

$$\leq \quad \square_{p,g,N,c} n^{-p/2} + \square_{p,g,N}\sqrt{2}(nc^{-1}\|g\|_\infty)^p n^{-\gamma/2}$$

$$\leq \quad \square_{p,g,N,c} n^{-p/2},$$

where we have used (4.15) for the first term and (4.16) for the second one. This concludes the proof of the lemma. ∎

The boundedness of $\mathcal{L}_k^{-1}$

Lemma 4. *For any function φ, we have:*

$$\big\| \mathcal{L}_k^{-1}(\varphi) \big\|_{2,T} \leq \sqrt{\frac{1}{3}} \|\varphi\|_{2,T}.$$

Proof. We have:

$$\int |\mathcal{L}_k^{-1}(\varphi)(x)|^2 dx \;=\; \frac{T}{k}\sum_{i=0}^{k-1} |H_{i,k}(\varphi)|^2 = \frac{T}{k}\sum_{i=0}^{k-1}\Big(\tfrac{1}{4}(H_{\frac{i}{2},k}(\varphi) + \varphi_{\frac{i}{2},k}) \Big)^2$$

$$=\; \frac{T}{k}\Big(\sum_{j=0}^{[\frac{k-1}{2}]} \tfrac{1}{16}\big(H_{j,k} + \varphi_{j,k}\big)^2 + \sum_{j=1}^{[\frac{k}{2}]} \tfrac{1}{64}(H_{j,k} + \varphi_{j,k} + H_{j-1,k} + \varphi_{j-1,k})^2 \Big)$$

$$\leq\; \frac{T}{k}\Big(\sum_{j=0}^{[\frac{k-1}{2}]} \tfrac{1}{8}\big(H_{j,k}^2 + \varphi_{j,k}^2\big) + \sum_{j=1}^{[\frac{k}{2}]} \tfrac{1}{16}(H_{j,k}^2 + \varphi_{j,k}^2 + H_{j-1,k}^2 + \varphi_{j-1,k}^2) \Big)$$

$$\leq\; \tfrac{1}{4}\frac{T}{k}\sum_{i=0}^{k-1}\Big(H_{j,k}^2 + \varphi_{j,k}^2 \Big) = \tfrac{1}{4}\Big(\int |\mathcal{L}_k^{-1}(\varphi)(x)|^2 dx + \frac{T}{k}\sum_{i=0}^{k-1}\varphi_{j,k}^2 \Big).$$

At the second line, we have distinguished the i's that are even and the i's that are odd. At the third line, we have used the inequalities $(a+b)^2 \leq 2a^2 + 2b^2$ and $(a+b+c+d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$. By substraction, we obtain

$$\int |\mathcal{L}_k^{-1}(\varphi)(x)|^2 dx \leq \frac{1}{3}\frac{T}{k}\sum_{i=0}^{k-1}\varphi_{j,k}^2.$$

The Cauchy-Schwarz inequality gives:

$$\Big(\int_{x_{i,k}}^{x_{i+1,k}} \varphi^2(x)dx \Big)^2 \leq \frac{T}{k}\int_{x_{i,k}}^{x_{i+1,k}} \varphi^2(x)dx,$$

38

so that

$$\frac{T}{k}\sum_{i=0}^{k-1}\varphi_{j,k}^2 = \frac{T}{k}\sum_{i=0}^{k-1}\frac{k^2}{T^2}\Big(\int_{x_{i,k}}^{x_{i+1,k}}\varphi(x)dx\Big)^2 \leq \int_0^T \varphi^2(x)dx$$

and finally we obtain the desired result:

$$\int |\mathcal{L}_k^{-1}(\varphi)(x)|^2 dx \leq \frac{1}{3}\int \varphi^2(x)dx.$$

■

References

[1] H. T. Banks, Karyn L. Sutton, W. Clayton Thompson, Gennady Bocharov, Dirk Roose, Tim Schenkel and Andreas Meyerhans, *Estimation of Cell Proliferation Dynamics Using CFSE Data*, Bull. of Math. Biol., doi: 10.1007/s11538-010-9524-5.

[2] Baraud, Y. *A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression.* to appear in Bernoulli.

[3] Doumic, M. and Gabriel, P. (2010) *Eigenelements of a General Aggregation-Fragmentation Model.* Math. Models Methods Appl. Sci. 20 (2010), no. 5, 757–783.

[4] Doumic, M., Perthame, B. and Zubelli, J. (2009) *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics.* Inverse Problems, **25**, 25pp.

[5] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Springer Verlag, 1996.

[6] Gasser, T. and Müller, H.G. (1979) *Optimal convergence properties of kernel estimates of derivatives of a density function.* In Lecture Notes in Mathematics **757**. Springer, Berlin, 144–154.

[7] Hall, P., Heyde, C.C., (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[8] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov A. (1998) *Wavelets, Approximation and Statistical Applications.* Springer-Verlag, Berlin.

[9] Lepski, O. V. (1990). *One problem of adaptive estimation in Gaussian white noise.* Theory Probab. Appl. **35** 459–470.

[10] Lepski, O. V. (1991). *Asymptotic minimax adaptive estimation. 1. Upper bounds.* Theory Probab. Appl. **36** 645–659.

[11] Lepski, O. V. (1992). *Asymptotic minimax adaptive estimation. 2. Statistical models without optimal adaptation. Adaptive estimators.* Theory Probab. Appl. **37** 468–481.

[12] Lepski, O. V. (1992). *On problems of adaptive estimation in white Gaussian noise.* In Advances in Soviet Mathematics (R. Z. Khasminskii, ed.) **12** 87–106. Amer. Math. Soc., Providence.

[13] Goldenshluger, A., Lepski, O. (2009) *Uniform bounds for norms of sums of independent random functions* arXiv:0904.1950.

[14] Goldenshluger, A., Lepski, O. (2010) *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality* arXiv:1009.1016.

[15] Mair, B. A. and Ruymgaart, F.H. (1996) *Statistical inverse estimation in Hilbert scales.* SIAM J. Appl. Math. **56**, 1424–1444.

[16] Massart, P. (2007) *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Springer, Berlin.

[17] Nussbaum, M. (1996). *Asymptotic equivalence of density estimation and white noise.* Ann. Statist. **24** 2399-2430.

[18] Nussbaum, M. and Pereverzev, S. (1999) *The degrees of ill-posedness in stochastic and deterministic noise models.* Preprint WIAS 509.

[19] Perthame, B. Transport equations arising in biology (2007). In *Frontiers in Mathematics*, Frontiers in Mathematics. Birckhauser.

[20] Metz, J.A.J. and Dieckmann O. (1986). *Formulating models for structured populations.* In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, Lecture Notes in Biomath., vol. 68, 78–135.

[21] P. Michel, S. Mischler, B. Perthame, *General relative entropy inequality: an illustration on growth models,* J. Math. Pures Appl. (9) 84, 1235–1260 (2005).

[22] B. Perthame and L. Ryzhik, *Exponential decay for the fragmentation or cell-division equation,* J. of Diff. Eqns, **210**, 155–177 (2005).

[23] B. Perthame and J. P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

[24] Tsybakov, A. (2004). *Introduction à l'estimation non-paramétrique.* Springer, Berlin.

[25] Wahba, G. (1977) *Practical approximate solutions to linear operator equations when the data are noisy.* SIAM J. Numer. Anal. **14**, 651–667.