Université de Lille 1 - Sciences et Technologies
Laboratoire Paul Painlevé - UMR CNRS 8524
École doctorale Sciences Pour l'Ingénieur - 072

THÈSE

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ de LILLE 1

Discipline: MATHÉMATIQUES APPLIQUÉES

Présentée par
Van Hà HOANG

Estimation adaptative pour des problèmes inverses avec des applications à la division cellulaire

---

Thèse soutenue publiquement le 28 Novembre 2016 devant le jury composé de

M.	Antoine	AYACHE	EXAMINATEUR
MME.	Gwénaëlle	CASTELLAN	EXAMINATRICE
MME.	Benoîte	DE SAPORTA	EXAMINATRICE
M.	Marc	HOFFMANN	RAPPORTEUR
MME.	Thanh Mai	PHAM NGOC	DIRECTRICE
M.	Vincent	RIVOIRARD	DIRECTEUR
MME	Adeline	SAMSON LECLERCQ	RAPPORTEUR
M.	Viet Chí	TRAN	DIRECTEUR

Remerciements

En tout premier lieu, je tiens à exprimer mes remerciements et ma profonde gratitude envers mes directeurs de thèse, Thanh Mai PHAM NGOC, Viet Chi TRAN et Vincent RIVOIRARD, pour m'avoir proposé un sujet très passionnant et pour m'avoir encadré durant ces trois années. Vous avez su m'apporter un soutien constant, une disponibilité, une confiance et des conseils précieux dès le début de ma thèse. Sans votre nombreux conseils, suggestions et orientations, ce manuscrit n'aurait jamais pu aboutir. Enfin, je vous remercie pour votre patience ainsi que vos encouragements tout au long de cette thèse.

Je remercie chaleureusement Adeline LECLERCQ SAMSON et Marc HOFFMANN pour avoir accepté de rapporter ma thèse. Je leur en suis très reconnaissant.

Je remercie vivement Gwénaëlle CASTELLAN, Benoîte DE SAPORTA et Antoine AYACHE de me faire l'honneur de participer à mon jury de soutenance.

Je tiens à remercier l'ensemble des membres du Laboratoire Paul Painlevé, en particulier l'équipe de Probabilités et Statistique, de m'avoir accueilli très chaleureusement pendant ces trois années de thèse. Je remercie également Aurore SMETS, Sabine HERTSOEN et Jean-Jaques DERYCKE pour leur aide et leur gentillesse.

I would like to thank my professors and colleagues of Faculty of Mathematics and Computer Science, University of Science in Ho Chi Minh city for providing a good working environment with a lot of kind assistance during the time when I have worked there.

Je remercie tous les organismes qui ont permis de financer mes travaux de recherche : Vietnam Ministère Éducation et Formation, le projet ANR *Calibration statistique* (ANR 2011 BS01 010 01) et la chaire MMB (Modélisation Mathématique et Biodiversité).

J'adresse également tous mes remerciements à cô Trâm pour m'avoir aidé à traverser sans encombres les méandres administratives qui accompagnent chaque inscription à l'école doctorale.

Je tiens à remercie tous les doctorants, post-doctorants et ATER au laboratoire pour leur accueil. En particulier, j'adresse un remerciement chaleureux à mes collègues au bureau M3-216+ : Céline Esser, Ahmed Ait Hammou Oulhaj, Geoffrey Boutard et Florent Dewez ainsi que Sara Mazzonetto, Huu Kien Nguyen, Pierre Houdebert et Simon Lestum, merci à vous pour tous les bons moments que nous avons passés ensemble.

Je souhaiterais également remercier Huy Cuong Vu Do, Hoang Anh Nguyen Le, Quang Huy Nguyen, Thi Mong Ngoc Nguyen, Nhat Nguyen Nguyen, Thi Le Thu Nguyen, Thi Trang Nguyen pour votre accompagnement et votre soutien tout au long de ma thèse.

Mes derniers remerciements vont à ma famille, pour leurs encouragements et leur soutien constant.

Contents

1 Introduction

Contents

1.1 Inverse problem models

This thesis is divided into two independent parts, both devoted to solve inverse problems in the statistical framework. In the first two chapters, a model motivated by the detection of the cellular aging in biology is studied. From the observation of a dividing cell population with size structure, we estimate the kernel determining how the daughters' sizes are related to their mother's one, with the purpose to see whether the sharing is rather symmetric or asymmetric. This leads us to both direct and indirect density estimation problems. In the latter, we deal with a deconvolution problem where the convolution operator is of multiplicative type. In the last chapter, we consider the estimation of a nonparametric regression function with errors in variables, using wavelets in a multivariate setting. To the best of our knowledge, this problem has seldom been studied. To take into account the noise in the covariates, deconvolution is involved in the construction of wavelet estimators.

Inverse problems arise in many scientific domains, such as biology, medicine, physics, engineering, finance, *etc.* Solving an inverse problem consists in reconstructing unknown quantities of interest from observed measurements. Loosely speaking, inverse problems are concerned with determining causes from known effects: for example, recovering the original image from a given blurred version of the image (see Bertero and Boccacci [13]) or determining the density of rock within the Earth from measurements of seismic waves produced by earthquakes (see for instance Mosegaard and Tarantola [85], Tarantola [102]), etc. Other applications of inverse problems in biology and medicine can be found in Comte et al [28], Stirnemann et al [100] for instance.

In mathematical terms, a general inverse problem is modelled as follows: let $\mathbb{H}$ and $\mathbb{K}$ two Hilbert spaces and let $\mathbf{A} : \mathbb{H} \to \mathbb{K}$ an operator. Given a function $g \in \mathbb{K}$, we want to find a good approximation of the solution $f \in \mathbb{H}$ of the equation:

$$g = \mathbf{A}f. \tag{1.1}$$

We may refer to Bertero and Boccacci [13], Bissantz [15], Cavalier [19], Cavalier and Hengartner [20], Tarantola [102], Tikhonov and Arsenin [104] for comprehensive studies and discussions of standard and statistical inverse problems.

Depending on the properties of the operator $\mathbf{A}$, the inverse problems are separated into two categories: well-posed and ill-posed problems. Following Hadamard [55], an inverse problem is called ill-posed if its solution does not exist ($\mathbf{A}$ is not surjective), if the solution is not unique ($\mathbf{A}$ is not injective) or if the solution is not continuous with respect to the data ($\mathbf{A}^{-1}$ is not continuous), *i.e.* a small change in the data g can cause a large change in the reconstruction of f. Otherwise a problem is said to be well-posed if it satisfies all three conditions above (existence, uniqueness and stability). One example of a well-posed problem is the forward problem of the heat equation. However, in practice inverse problems are often ill-posed problems which arise naturally in many domains, for instance, in computerized tomography, in image processing or in geophysics. The main issue is usually on the stability since it is very difficult to observe the function g without errors. To account for this, Equation (1.1) can be considered with an additive term ϵ modeled according to the considered approaches: ϵ can be either a deterministic perturbation or a random noise. In the deterministic approach, one supposes that the noisy observations of g is $g_\epsilon = \mathbf{A}f + \sigma\epsilon$ for some $\sigma > 0$ and ϵ belongs to a ball of the Hilbert space $\mathbb{G}$, *i.e* $\|\epsilon\| \leq 1$. This approach is the standard framework for inverse problems. We refer to Tikhonov [103], Tikhonov and Arsenin [104] for first studies of deterministic perturbations. In contrast, the statistical point of view assumes that we observe

$$Y = \mathbf{A}f + \sigma\epsilon, \tag{1.2}$$

where the noise ϵ is a random variable and σ corresponds to the noise level with $\sigma > 0$. In the case where $\mathbb{H}$ is a functional space, many nonparametric estimation problems are related to this second approach, such as probability density estimation, or estimation in the regression model and in the white noise model. Initial studies of inverse problems with random noise were proposed by Bakushinskii [6], Sudakov and Khalfin [101].

A classical example of the model (1.2) is the Gaussian white noise model with convolution (see for instance De Canditiis and Pensky [30], Johnstone [69], Kalifa and Mallat [70], Kerkyacharian et al. [71] and Pensky and Sapatinas [90]):

$$Y(x) = f \star G(x) + \sigma W(x), \quad x \in [0, 1], \tag{1.3}$$

where $W(x)$ is assumed to be a Gaussian white noise. Here the operator $\mathbf{A}$ is the convolution operator defined by:

$$\mathbf{A}f(x) = f \star G(x) = \int_0^1 f(z)G(x-z)dz.$$

and $f \in \mathbb{L}^2([0,1])$ is the unknown function we want to recover from noisy observations $Y(x)$. The function G is assumed to be known (the Green kernel for the heat equation for example). To obtain an estimate of f, we deal with a deconvolution problem to find an inverse of $\mathbf{A}$. In this thesis, deconvolution procedures arise in **Chapters 3 and 4**. The difficulties that are encountered are due to the fact that the convolutions are either multiplicative or in a multivariate setting.

This thesis consists of two statistical inverse problems: density estimation for a pure birth process with size structure or a growth-fragmentation partial differential equation (PDE), and multivariate nonparametric regression with errors in the variables. After giving a detailed presentation of the two problems that are considered in this thesis (Section 1.2), we introduce the Goldenshluger and Lepski's method (Section 1.3) which is applied for our adaptive estimation procedures and the present contributions of this thesis are summed up (Section 1.4).

1.2 Presentation of the models considered in this thesis

1.2.1 Size-structured population model

In the first part of the thesis, which consists of **Chapter 2** and **Chapter 3**, we study the problem of estimating a density function associated with an application for cell population with size structure.

Size-structured population models have recently attracted a lot of attention and interest in literature. We may cite the works of Calsina and Manuel [17], Diekmann et al. [35], Doumic and Gabriel [37], Doumic et al. [40], Doumic and Tine [41], Farkas and Hagen [48], Michel [83] and Perthame [91]. These authors study PDE models for size-structured population. We refer to Bansaye et al. [8], Bansaye and Méléard [9], Cloez [24], Doumic et al. [38] for size-structured population modelled by branching processes. Furthermore, related studies for populations with age structure can be found in Athreya and Ney [5], Harris [57], Jagers [67], Tran [107] and references therein.

We consider a stochastic individual-based model to describe a discrete population of cells in continuous time where the individuals are cells characterized by their varying sizes and undergoing binary divisions. The term 'size' covers variables such as volume, length, level of certain proteins, DNA content, *etc.* that grow deterministically with time. Here, we have in mind that each cell contains some toxicities which play the role of the size, in the spirit of the study of Stewart et al. [99]. For the sake of simplicity, we assume here that the toxicity X_t inside a cell at time t grows linearly with a constant rate $\alpha > 0$.

$$dX_t = \alpha dt. \tag{1.4}$$

When a division occurs, the mother cell dies and is replaced by two daughters. The division rate is $R > 0$, meaning that the lifelength of each cell is an exponential random variable with parameter $R > 0$. Upon division, a random fraction Γ of the cell's size goes into the first daughter cell and a fraction $1 - \Gamma$ into the second one. So if the division occurs at time t, and if the mother is of size X_t, the two daughter cells are of sizes ΓX_{t-} and $(1-\Gamma)X_{t-}$. The random variable Γ is on $[0, 1]$ is assumed to have a density $h(\gamma)$ with respect to Lebesgue measure. Our purpose here is to estimate the density h ruling the divisions. The originality of estimating h is to detect aging phenomena such as the one put into light by Stewart et al. [99]. More precisely, if h is piked at $\frac{1}{2}$ (*i.e.* $\Gamma \simeq \frac{1}{2}$), both daughters contain the same toxicity, *i.e.* the half of their mother's toxicity. The more h puts weight in the neighbourhood of 0 and 1, the more asymmetric the divisions are, with one daughter having little toxicity and the other's toxicity close to its mother's one. If we consider that having a lot of toxicity is a kind of senescene, then, the kurtosis of h provides indication on aging phenomena (see Lindner et al. [75]).

We stick to constant rates R and α for the sake of simplicity. Modifications of this model to account for more complex phenomena have been considered in other papers. Bansaye and Tran [11], Cloez [24] or Tran [106] consider non-constant division and growth rates. Robert et al. [95] studies whether divisions can occur only when a size threshold is reached. Notice that several similar models for binary cell division in discrete time also exist in the literature, see for instance Bansaye et al. [7, 10], Bercu et al. [12], Delmas and Marsalle [34], Guyon [54] or Bitseki Penda [89].

In our model, the genealogy of these dividing cells can be represented by a binary tree, namely a Yule tree, the nodes of which can be labelled with the Ulam-Harris-Neveu notation. Let $J = \bigcup_{m=0}^{+\infty} \mathbb{N} \times \{0, 1\}^m$ be the set of labels with the convention that $\{0, 1\}^0 = \emptyset$; J is the set of words starting with an integer followed by a sequence of 0s and 1s. The roots of the tree (the cells present at time 0) are given integer labels in $\mathbb{N}$. When a cell with label j splits, her daughters are given the labels $j0$ and $j1$ obtained by concatenating 0 or 1 to the word j. The population at time t consists of the collection of sizes $(X_t^i, i \in V_t)$ where $V_t \subset J$ is the set of labels of cells alive at time t. $N_t = |V_t|$ is the size of the population. Because the size of the population is random and increasing, it is more convenient to describe the population by a random point measure, rather than a vector of varying size. Let

$$Z_t(dx) = \sum_{i \in V_t} \delta_{X_t^i}(dx), \tag{1.5}$$

be the sum of Dirac masses weighting the sizes of cells alive at time t. It is a empirical measure on $\mathcal{M}(\mathbb{R}_+)$, the space of finite measures embedded with the topology of weak convergence.

The dynamics described above defines $(Z_t, t \geq 0)$ as a piecewise deterministic Markov process (PDMP) that can be associated to a stochastic differential equation (SDE) driven by a Poisson point measure (see e.g. [50, 106]): divisions correspond to random times generated by the Poisson measure and between two divisions, sizes grow following (1.4).

Let $Z_0 \in \mathcal{M}(\mathbb{R}_+)$ such that $\mathbb{E}\Big(\langle Z_0, 1\rangle\Big) < +\infty$. Let $Q(ds, di, d\gamma)$ be a Poisson point measure on $\mathbb{R}_+ \times \mathcal{E} := \mathbb{R}_+ \times J \times [0,1]$ with intensity $q(ds, di, d\gamma) = R ds n(di) H(d\gamma)$ where $n(di)$ is the counting measure on J and ds is Lebesgue measure on $\mathbb{R}_+$. Because Z_t, for $t \in \mathbb{R}_+$, is a random variable in the space of finite measures $\mathcal{M}_F(\mathbb{R}_+)$, we characerize its evolution by considering test functions f and giving SDEs for $\langle Z_t, f\rangle = \int_{\mathbb{R}_+} f(x) Z_t(dx)$.

For every test function $f : (x,t) \mapsto f(x,t) = f_t(x) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ (bounded of class $\mathcal{C}^1$ in t and x with bounded derivatives):

$$\begin{aligned}\langle Z_t, f\rangle =& \langle Z_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} \Big(\partial_s f_s(x) + \alpha \partial_x f_s(x)\Big) Z_s(dx) ds \\ &+ \int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i \le N_{s-}\}} \Big[f_s\Big(\gamma X^i_{s-}\Big) + f_s\Big((1-\gamma) X^i_{s-}\Big) - f_s(X^i_{s-})\Big] Q(ds, di, d\gamma).\end{aligned} \tag{1.6}$$

The second term in the right hand side of (1.6) corresponds to the growth of sizes in the cells and the third term gives a description of cell divisions where the sharing of mother cell's size into two daughter cells depends on the random fraction Γ.

The SDE (1.6) can easily be simulated as follows (see Bansaye and Méléard [9], Ferrière and Tran [49] or Fournier and Méléard [50] for more details):

1. We start with N_0 cells of sizes $X^0_0, \dots X^{N_0-1}_0$. The set of living cell labels at time 0 is $V_0 = \{0, \dots N_0 - 1\}$.
2. To each living cell at time $t = 0$, we attach an independent random exponential clock with parameter R.
3. The smallest clock value τ determines the cell u that divides next. The division happens at time $t + \tau$ and the cell sizes just before the division are $X^v_{t+\tau_-} = X^v_t + \alpha\tau$, $v \in V_{t+\tau_-}$. We draw an independent random variable Γ following the distribution $h(x)dx$ and replace the cell u by her daughters $u0$ of size $(1-\Gamma)X^u_{t+\tau_-}$ and $u1$ of size $\Gamma X^u_{t+\tau_-}$.
4. We iterate again from step 2, with the last division time instead of $t = 0$.

Since all the cells have the same division rate, we can of course simplify the algorithm: when the population is of size N, the next division occurs after an independent exponential time with parameter NR and we can select uniformly the cell that undergoes division.

In **Chapter 2**, we assume that we observe the evolution of cells up to a fixed time T, i.e. the whole division tree and the entire path $(Z_t, t \in [0,T])$ of the measure-valued process. From the algorithm presented in the previous paragraph, we can guess that the observation of the whole tree allows us to construct a sequence of independent random variables distributed as Γ, which we can use to construct an adaptive estimator of the kernel division, with a fully data-driven bandwidth selection method. The main difficulty comes from the fact that the number of observations of the sequence is random, and we have to study the SDE to obtain necessary estimates to establish upper and lower bounds for the convergence rate of the estimator. This is detailed in Section 1.4.

Hoffmann and Olivier in [59] describe the growth-fragmentation model by a PDMP $X = ((X_t^1, X_t^2, \ldots), \quad t \geq 0)$, with value in $\bigcup_{k=0}^{\infty} [0, \infty)^k$, where X_t^i denotes the age of the living individual at time t. The division rate $B(x)$ considered by these authors is a function of the size x. Kernel estimators of $B(x)$ are constructed by assuming that the whole genealogical tree on the time interval $[0, T]$ is observed. The main difficulty comes from the fact that the number of observations as well as the probability of inclusion of an individual both depend on B. In this thesis, we do not consider these issues, but other difficulties appear.

In **Chapter 3**, we observe the cell population at a fixed time T. The whole division tree is not completely observed and the estimation techniques developed in Chapter 2 cannot be considered anymore. Models with partial observations arise frequently in biology and have been investigated in many studies of both biostatistics and probability, see for instance Donnet and Samson [36], Samson and Thieullen [97], Wu [110]. In this case, the idea is to approximate the evolution of the population when the initial size is large by a growth-fragmentation PDE. Assuming that we have n independent observations $X_1, \ldots, X_n$ of distribution $n(T, x)dx$, where $n(T, x)$ defines the solution of the PDE, we can estimate h from an inverse problem involving a multiplicative convolution.

More precisely, assume that $N_0 = n$ and consider the renormalized random point measure:

$$Z_t^n(dx) = \frac{1}{n} \sum_{i=1}^{N_t^n} \delta_{X_t^i}(dx), \tag{1.7}$$

where N_t^n is the number of cells alive at time t. The large population limit corresponds to $n \to +\infty$. Assume that Z_0^n converges in distribution to the measure $n_0(x)dx \in \mathcal{M}_F(\mathbb{R}_+)$ as $n \to +\infty$. Following the work of Fournier and Méléard [50] (see also in Ethier and Kurtz [43] and Tran [107]), we prove that the renormalized random process converges to the weak solution of a growth-fragmentation equation:

$$\partial_t n(t,x) + \alpha \partial_x n(t,x) + Rn(t,x) = 2R \int_0^1 n\left(t, \frac{x}{\gamma}\right) \frac{1}{\gamma} h(\gamma) d\gamma, \qquad n(0,x) = n_0(x). \tag{1.8}$$

The PDE (1.8) is very close to the one considered in Bourgeron et al. [16] and Doumic and Tine [41]:

$$\begin{cases} \partial_t n(t,x) + \alpha \partial_x \Big(g(x) n(t,x)\Big) + B(x) n(t,x) = 2 \int_0^{+\infty} n(t,y) B(y) \kappa(x,y) dy, \\ t \geq 0, x \geq 0, \\ g(0) n(t,0) = 0, \quad t > 0, \\ n(0,x) = n^0(x), \quad x \geq 0, \end{cases} \tag{1.9}$$

where $n(t,x)$ is the density of the cells structured by the size x at time t, $\kappa(x,y) = \frac{1}{y} h\Big(\frac{x}{y}\Big)$ is the division kernel, $\alpha g(x)$ is the non-constant growth rate and $B(x)$ is the non-constant division rate (see also Doumic et al. [39] in the case of equal mitosis). Setting $y = x/\gamma$ in (1.8) gives the PDE (1.9) in the case where the division rate and the growth rate are the constants R and α respectively:

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) + Rn(t,x) = 2R\int_0^{+\infty} n(t,y)h\Big(\frac{x}{y}\Big)\frac{dy}{y}.$$

The study of the asymptotic behaviour of the solutions $n(t,x)$ is related to the following eigenvalue problem

$$\begin{cases} \alpha\partial_x N(x) + (\lambda+R)N(x) = 2R\int_0^{+\infty} N(y)h\Big(\frac{x}{y}\Big)\frac{dy}{y}, \; x \geq 0, \\ N(0)=0, \quad \int N(x)dx = 1, \quad N(x)\geq 0, \quad \lambda > 0, \end{cases} \tag{1.10}$$

where λ is the first eigenvalue and N is the first eigenvector. When $t \to +\infty$, it is proved (see *e.g* [84], [92]) that the approximation $n(t,x) \approx N(x)e^{\lambda t}$ is valid. We assume that we have n i.i.d observations $X_1, X_2, \ldots, X_n$ with distribution $N(x)dx$. For instance, each observation is drawn by measuring a cell selected randomly. Then, the estimation of h leads us to an inverse problem with a multiplicative convolution defined by the operator

$$\Big(f \vee g\Big)(x) = \int_0^{+\infty} f(y)g\Big(\frac{x}{y}\Big)\frac{dy}{y},$$

where f and g are two integrable functions defined on $\mathbb{R}_+$.

Deconvolution problem with multiplicative operator has seldom been studied in statistical inverse problems. Thus, our problem is more complicated and quite different, compared to the classical deconvolution problems. In a standard inverse problem, Bourgeron et al. [16] consider an equation similar to Equation (1.10) but the division rate R is replaced by the function $B(x)$ and they aim to estimate B. When dealing with the convolution $\int_0^\infty H(y)h\left(\frac{\cdot}{y}\right)\frac{dy}{y}$ with $H(y) = B(y)N(y)$, they apply the Mellin transform to replace the multiplicative convolution by a product. In our problem, we apply a logarithmic change of variables in the right hand side (r.h.s) of Equation (1.10) giving a term of the form $\int_{\mathbb{R}} M(v)g(u-v)dv = (M\star g)(u)$ where $M(u) = N(e^u)$ and $g(u) = h(e^u)$. Then we resort to Fourier techniques to recover the Fourier transform of g for which we can propose a kernel-based estimator. The estimator of the Fourier transform of g involves a quotient of two estimators and we use techniques inspired by Comte and Lacour [26, 27], Comte et al. [28], Neumann [87] for its construction. The estimator of h will be obtained from the estimator of g. We prove only consistency of our estimator. Indeed, it appears that regularities of functions to be estimated h or g are closely linked with functions $\partial_x N$ and N in Equation (1.10). This makes the problems intricate and thus the study of rates of convergence is a work in progress.

1.2.2 Nonparametric regression with errors-in-variables

In **Chapter 4**, we consider the multivariate regression with errors-in-variables. Suppose that we observe n i.i.d random vectors $(W_1, Y_1), \ldots, (W_n, Y_n)$ of the following model:

$$\begin{cases} Y_l &= m(X_l) + \varepsilon_l, \\ W_l &= X_l + \delta_l, \quad l = 1,\ldots,n, \end{cases} \tag{1.11}$$

where $Y_l \in \mathbb{R}$, $(\delta_1, \ldots, \delta_n)$ and $(X_1, \ldots, X_n)$ are i.i.d $\mathbb{R}^d$-valued vector. We denote by f_X the density of the X_l's assumed to be bounded from below by a positive constant, and by f_W the density of the W_l's. The covariates errors δ_l are i.i.d unobservable random variables having known density g. The ε_l's are i.i.d standard normal random variables with known variance s^2. The δ_l's are independent of the X_l's and Y_l's. Our aim is to estimate the regression function $m(x), x \in [0,1]^d$ based on the observations Y_l's and W_l's where direct observations of X_l's are not available. We aim at estimating the multidimensional regression function $m(x)$ for the pointwise risk. Model (1.11) has been mainly studied in the univariate case. Let us cite the work of Fan and Truong [47], Comte and Taupin [29] and Meister [79] among others and the goal is to extend these results to the multivariate setting. Introduction of **Chapter 4** describes the framework in which our contribution takes place more precisely.

One of the difficulties of the problem is due to the design points which are corrupted by additive errors leading to a necessary deconvolution step. To face this issue we devise a kernel projection procedure based on wavelets. The second issue we focus on is the choice of the multiresolution analysis on which projection is performed. Note that this problem is intricate and barely considered. Here, our approach based on an application of the Goldenshluger-Lepski rule leads to an automatic and adaptive choice of the multiresolution analysis, which varies locally with x. Note that our estimation procedure does not resort to thresholding. Considering the ordinary smooth case on the component densities of the errors covariates, we establish optimal rates of convergence for the estimator $\hat{m}(x)$ for the pointwise risk and over anisotropic Hölder classes. The summary of the theoretical and numerical results are presented in Section 1.4.

1.3 The Goldenshluger-Lepski method

This section is devoted to the presentation the Goldenshluger-Lepski (GL) method which is the main adaptation method used in this thesis. In **Chapter 2**, we apply the GL method for the selection of the bandwidth of the kernel estimator $\hat{h}$. In **Chapter 4**, we inspire from the GL methodology to propose a data-driven selection rule of the wavelet level resolution for the estimation of the multivariate regression function m. As a consequence of using the GL method, we obtain oracle inequalities which provide a good tool to measure the performance of our estimation procedures in both problems of estimating the division kernel h and the estimation of the regression function m.

The GL method is proposed in Goldenshluger and Lepski [52] and is widely used in recent studies of nonparametric estimation, see for instance, Comte et al. [25], Comte and Lacour [27], Doumic et al. [39], Reynaud-Bouret et al. [94] who apply the GL method for selecting a fully-data driven bandwidth of kernel estimators, and Bertin et al. [14], Chagny [21] who extent the GL method to model selection for selecting dimension of estimator by projection.

Here, for the sake of simplicity, we present the GL method for the problem of density estimation in the univariate case for the $\mathbb{L}_2$-loss: let X be a random variable in $\mathbb{R}$ having density f, we want to estimate f from the i.i.d sample $X_1, X_2, \ldots, X_n$ drawn from f. Let K be a kernel satisfying $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$. We introduce the kernel estimator of f as follows:

$$\hat{f}_\ell(x) := \frac{1}{n}\sum_{i=1}^{n} K_\ell(x - X_i),$$

where $K_\ell(\cdot) = (1/\ell)K(\cdot/\ell)$ and $\ell > 0$ is the bandwidth to be selected.

We first study the bias-variance decomposition under the $\mathbb{L}_2$-risk:

$$\mathbb{E}\left[\|f - \hat{f}_\ell\|_2^2\right] = \|f - \mathbb{E}[\hat{f}_\ell]\|_2^2 + \mathbb{E}\left[\|\mathbb{E}[\hat{f}_\ell] - \hat{f}_\ell\|_2^2\right].$$

For the variance term, we obtain easily

$$\mathbb{E}\left[\|\mathbb{E}[\hat{f}_\ell] - \hat{f}_\ell\|_2^2\right] \leq \frac{1}{n\ell}\int K^2(x)dx = \frac{\|K\|_2^2}{n\ell}.$$

Moreover, one can easily check that $\mathbb{E}[\hat{f}_\ell] = \int K_\ell(\cdot - y)f(y)dy = K_\ell \star f$, where $\star$ denotes the convolution operator. Then we obtain

$$\mathbb{E}\left[\|f - \hat{f}_\ell\|_2^2\right] \leq \|f - K_\ell \star f\|_2^2 + \frac{1}{n\ell}\|K\|_2^2.$$

Let $\mathcal{H}_n$ be a family of possible bandwidths, then the best choice for the bandwidth is the one which minimizes the bias-variance decomposition. This is the oracle bandwidth $\bar{\ell}$ defined as

$$\bar{\ell} := \underset{\ell \in \mathcal{H}_n}{\operatorname{argmin}} \left\{\|f - K_\ell \star f\|_2^2 + \frac{1}{n\ell}\|K\|_2^2\right\}.$$

However, it is impossible to obtain $\bar{\ell}$ in practice since we cannot compute the bias term $\|f - K_\ell \star f\|_2$ due to the unknown function f. The idea of the GL method is to estimate the bias term $\|f - K_\ell \star f\|_2$ by considering several estimators. To derive the selection rule, we define for any $\ell, \ell' \in \mathcal{H}_n$

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

Heuristically, if $\hat{f}_{\ell'}$ is an estimator of f then $\hat{f}_{\ell,\ell'} = K_\ell \star \hat{f}_{\ell'}$ is an estimator of $K_\ell \star f$.

Then, for any $\ell, \ell' > 0$, the estimator of the bias term is defined as

$$A(\ell) = \sup_{\ell' \in \mathcal{H}_n}\left\{\|\hat{f}_{\ell'} - \hat{f}_{\ell,\ell'}\|_2^2 - V(\ell')\right\}_+,$$

where $V(\ell') = \chi\frac{\|K\|_2^2}{n\ell'}$ with χ a constant to be tuned. We subtract the variance term $V(\ell')$ in the setting of $A(\ell)$ to take into account the fluctuations of the estimator (see [39] for more details).

Finally, the bandwidth is selected as

$$\hat{\ell} := \underset{\ell \in \mathcal{H}_n}{\operatorname{argmin}} \left\{A(\ell) + V(\ell)\right\},$$

and we define $\hat{f} := \hat{f}_{\hat{\ell}}$ the corresponding kernel estimator.

Under appropriate conditions, we can prove that this estimate satisfies an oracle inequality for the $\mathbb{L}_2$-risk:

$$\mathbb{E}\left[\|f-\hat{f}\|_2^2\right] \leq C \inf_{\ell \in \mathcal{H}} \left\{\|f - K_\ell \star f\|_2^2 + \frac{1}{n\ell}\right\} + \Delta_n,$$

where $C > 0$ is a constant independent of f and n and Δ_n is a negligible term.

1.4 Contributions

1.4.1 Estimation of the division kernel

This section presents the main results for the estimation of the division kernel in the case of complete data (**Chapter 2**) and in the case of incomplete data (**Chapter 3**).

Case of complete data:

Assume that we observe the whole division tree in the time interval $[0, T]$ with fixed T. At the i^{th} division time t_i, let us denote j_i the individual who splits into two daughter cells $X_{t_i}^{j_i 0}$ and $X_{t_i}^{j_i 1}$ and define

$$\Gamma_i^0 = \frac{X_{t_i}^{j_i 0}}{X_{t_i -}^{j_i}} \quad \text{and} \quad \Gamma_i^1 = \frac{X_{t_i}^{j_i 1}}{X_{t_i -}^{j_i}}$$

the random fractions that go into the daughter cells, with the convention $\frac{0}{0} = 0$. Γ_i^0 and Γ_i^1 are exchangeable with $\Gamma_i^1 = 1 - \Gamma_i^0$. Γ_i^0 and Γ_i^1 are thus not independent but the couples $(\Gamma_i^0, \Gamma_i^1)_{i \in \mathbb{N}^*}$ are independent and identically distributed with distribution (Γ^0, Γ^1) where Γ^1 has the density h and $\Gamma^0 = 1 - \Gamma^1$. Based on the observations of the Γ_i^1's, we construct an adaptive estimator of h by using a kernel method with a kernel function K.

Assume that we observe $(\Gamma_1^1, \dots, \Gamma_{M_T}^1)$ where $M_T > 0$ is the random number of random divisions in $[0, T]$, we define the kernel estimator of h as follows:

Definition 1.4.1. Let $K : \mathbb{R} \longrightarrow \mathbb{R}$ is an integrable function such that

$$\int_{\mathbb{R}} K(x)dx = 1 \ \text{ and } \ \int_{\mathbb{R}} K^2(x)dx < \infty.$$

For all $\gamma \in (0, 1)$, define

$$\hat{h}_\ell(\gamma) = \frac{1}{M_T} \sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1), \tag{1.12}$$

where $K_\ell = \frac{1}{\ell} K(\cdot/\ell)$ and $\ell > 0$ is the bandwidth to be chosen.

When using a kernel method, the choice of bandwidth is crucial. We apply here the GL method to propose a fully data-driven selection rule for which we can select an adaptive bandwidth and provide a non asymptotic oracle inequality. Details of our bandwidth selection rule is presented in Section 2.2.3 of **Chapter 2**.

We state an oracle inequality which highlights the bias-variance decomposition of the mean squared integrated error (MISE) of $\hat{h}$. We recall that the MISE of $\hat{h}$ is the quantity $\mathbb{E}\Big[\|\hat{h}-h\|_2^2\Big]$.

Theorem 1.4.2. *Let N_0 be the number of mother cells at the beginning of divisions and M_T is the random number of divisions in $[0,T]$. Consider H a countable subset of $\{\triangle^{-1} : \triangle = 1,\dots,\triangle_{\max}\}$ in which we choose the bandwidths and $\triangle_{\max} = \lfloor \delta M_T \rfloor$ for some $\delta > 0$. Assume $h \in L^\infty([0,1])$ and let $\hat{h}$ be a kernel estimator defined with the kernel $K_{\hat{\ell}}$ where $\hat{\ell}$ is chosen by the GL method. Define*

$$\varrho(T)^{-1} = \begin{cases} \dfrac{e^{-RT+\log(RT)}}{1-e^{-RT}}, & \text{if } N_0 = 1, \\ e^{-RT}, & \text{if } N_0 > 1. \end{cases}$$

Then,

$$\mathbb{E}\Big[\|\hat{h}-h\|_2^2\Big] \le C_1 \inf_{\ell \in H}\Big\{\|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\varrho(T)^{-1}\Big\} + C_2\varrho(T)^{-1},$$

where C_1 is a constant depending on N_0, $\|K\|_1$ and C_2 is a constant depending on N_0, δ, $\|K\|_1$, $\|K\|_2$ and $\|h\|_\infty$.

The constants C_1 and C_2 in the oracle inequality above depend also on a positive constant ϵ to be tuned. However we do not mention ϵ for the ease of exposition.

The term $\|K_\ell \star h - h\|_2^2$ is an approximation term, $\frac{\|K\|_2^2}{\ell}\varrho(T)^{-1}$ is a variance term and the last term $\varrho(T)^{-1}$ is asymptotically negligible. Hence the right hand side of the oracle inequality corresponds to a bias variance trade-off.

Adaptive minimax rates of convergence are derived when the density h belongs to a Hölder class $\mathcal{H}(\beta, L)$ of smoothness β and radius L and the kernel function K satisfies some regularity conditions. These assumptions are detailed in Section 2.2.3 of **Chapter 2**. The upper bound in Theorem 1.4.3 is deduced from the oracle inequality of Theorem 1.4.2. The lower bound is obtained by perturbation methods (Theorem 1.4.4) and is valid for any estimator $\widehat{h}_T$ of h, thus indicating the optimal convergence rate.

Theorem 1.4.3. *Let $\beta^* > 0$ and K be a kernel of order β^*. Let $\beta \in (0, \beta^*)$. Let $\hat{\ell}$ be the adaptive bandwidth. Then, for any $T > 0$, the kernel estimator $\hat{h}$ satisfies*

$$\sup_{h \in \mathcal{H}(\beta,L)} \mathbb{E}\|\hat{h}-h\|_2^2 \le C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}},$$

where $\varrho(T)^{-1}$ is defined as above and C_3 is a constant depending on N_0, δ, ϵ, $\|K\|_1$, $\|K\|_2$, $\|h\|_\infty$, β and L.

Theorem 1.4.3 illustrates adaptive properties of our procedure: it achieves the rate $\varrho(T)^{-\frac{2\beta}{2\beta+1}}$ over the Hölder class $\mathcal{H}(\beta, L)$ as soon as β is smaller than β^*. So, it automatically adapts to the unknown smoothness of the signal to estimate.

Theorem 1.4.4. *For any $T > 0$, $\beta > 0$ and $L > 0$. Assume that $h \in \mathcal{H}(\beta, L)$, then there exists a constant $C_4 > 0$ such that for any estimator $\hat{h}_T$ of h*

$$\sup_{h \in \mathcal{H}(\beta,L)} \mathbb{E}\|\hat{h}_T - h\|_2^2 \geq C_4 e^{-\frac{2\beta}{2\beta+1}RT}.$$

Remark 1.4.5. The difficulty for adaptation is that the number of observations is random. It is worth noting the difference between N_T, the number of cells living at time T, and the number of random divisions M_T. Indeed, we have $M_T = N_T - N_0$. Moreover, the optimality of the estimator $\hat{h}$ depends on N_0. When $N_0 = 1$, N_T is distributed according to a geometric distribution with parameter e^{-RT}. The upper bounds are $e^{\frac{2\beta}{2\beta+1}(-RT+\log(RT))}$ and differ with a logarithmic term from the lower bound. When $N_0 > 1$, N_T has a negative binomial distribution $\mathcal{NB}(N_0, e^{-RT})$ and both the upper bounds and lower bounds are in $e^{-\frac{2\beta}{2\beta+1}RT}$. Thus the rate of convergence is optimal when $N_0 > 1$.

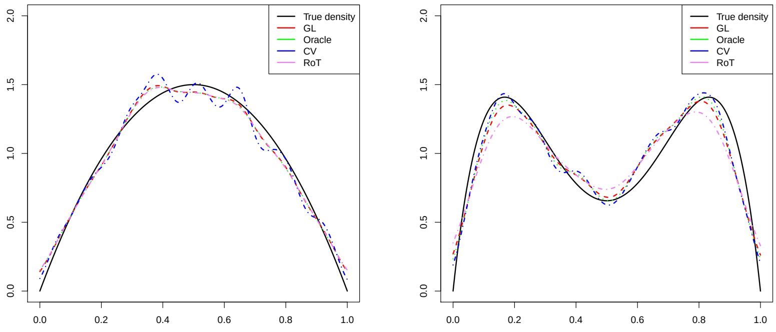
Lastly, we present here some numerical results to illustrate the performance of our estimator. Full details of the simulation study can be found in Section 2.3.

We first recall that the Beta density with parameters (a,b), denote here by $\text{Beta}(a, b)$, is proportional to $x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$. We implement simulations where h is the density of the $\text{Beta}(2, 2)$ distribution or a mixture distribution as $\frac{1}{2}\,\text{Beta}(2, 6) + \frac{1}{2}\,\text{Beta}(6, 2)$. The $\text{Beta}(2, 2)$ distribution represents symmetric divisions with kernel concentrated around $1/2$ while the choice of the Beta mixture gives us a bimodal density corresponding to very asymmetric divisions. We aim in both cases to estimate non-parametrically these densities with our nonparametric estimator.

We take the classical Gaussian kernel $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and reconstruct $\hat{h}$ by (1.12) where the adaptive bandwidth is selected by the GL method. We compare the estimated densities by the GL bandwidth with those estimated by the oracle bandwidth defined as

$$\ell_{\text{oracle}} = \operatorname*{argmin}_{\ell \in H} \mathbb{E}\Big[\|\hat{h}_\ell - h\|_2^2\Big].$$

Moreover, we are interested in comparing $\hat{h}$ with estimators obtained by Cross-Validation (CV) bandwidth, the rule-of-thumb (RoT) bandwidth. Computation of the CV and RoT bandwidths is described in Section 2.3. More details of these methods can be found in Silverman [98] or Tsybakov [108]. We shall call GL (resp. oracle, CV, RoT) estimator for the one estimated by using the GL (resp. oracale , CV, RoT) bandwidth. For a further comparison, in the reconstruction of $\text{Beta}(2, 2)$ density, we compare our nonparametric estimators with the parametric one obtained by assuming that the distribution is a $\text{Beta}(a, a)$ distribution and by using Maximum Likelihood (ML) method to estimate the parameter a.



(a) (b)

Figure 1.1: Reconstruction of division kernels with $T = 13$. (a): $Beta(2, 2)$. (b): Beta mixture.

To estimate the MISE, we implement Monte-Carlo simulations with $\mathcal{M} = 100$ repetitions:

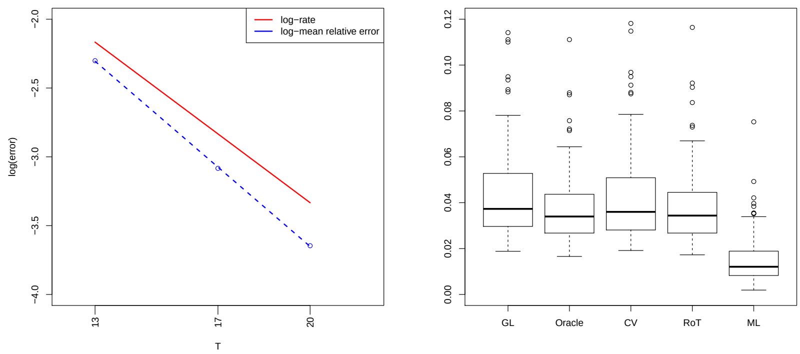
$$\bar{e} = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} e_i \text{ and } \sigma_e = \sqrt{\frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} (e_i - \bar{e})^2}$$

where

$$e_i = \frac{\|\hat{h}^{(i)} - h\|_2}{\|h\|_2}, \quad i = 1, \ldots, \mathcal{M},$$

and $\hat{h}^{(i)}$ denotes the estimator of h corresponding to i^{th} repetition.

Figure 1.1 illustrates a reconstruction for the density of Beta$(2, 2)$ and Beta mixture with $T = 13$, $R = 0.5$ and $\alpha = 0.35$.



(a) (b)

Figure 1.2: (a): The log-mean relative error for the reconstruction of Beta$(2, 2)$ compared to the log-rate (solide line) computed with $\beta = 1$. (b): Errors of estimated densities of Beta$(2, 2)$ when $T = 17$.

In Figure 1.2a, we illustrate on a log-log scale the mean relative error and the rate of convergence with respect to the time T. This shows that the error is close to the exponential rate predicted by the theory. The boxplots in Figure 1.2b represent the mean of relative error for estimated densities using the GL bandwidth, the oracle bandwidth, the CV bandwidth and the RoT bandwidth. One can observe that the GL estimator is closed to the oracle one which is the best estimator. Overall, we conclude that the GL method has a good behavior when compared to the CV method and rule-of-thumb. As usual, we also see that the ML errors are quite smaller than those of the nonparametric approach but the magnitude of the mean $\bar{e}$ remains similar.

Case of incomplete data:

As stated in Section 1.2.1, when the division tree is not fully observed, we can rely on the PDE approximation of the stochastic process $(Z_t^n, t \in \mathbb{R}_+)$, when $n \to +\infty$.

We denote by $\mathbb{D}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$ (resp. $\mathcal{C}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$) the set of càdlàg functions (resp. continuous functions) from $[0,T]$ to $\mathcal{M}_F(\mathbb{R}_+)$ embedded with the Skorohod distance (resp. embedded with the uniform convergence norm).

Theorem 1.4.6. *Consider the sequence $(Z^n)_{n\in\mathbb{N}^*}$ defined in (1.7). We assume that $Z_0^n \in \mathcal{M}_F(\mathbb{R}_+)$ satisfies*

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\Big(\langle Z_0^n, 1\rangle^2\Big) < +\infty.$$

If Z_0^n converges in distribution to $\mu_0 \in \mathcal{M}_F(\mathbb{R}_+)$ as $n \to +\infty$ then for every $T > 0$, $(Z^n)_{n\in\mathbb{N}^*}$ converges in distribution in $\mathbb{D}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$ as $n \to +\infty$ to the unique solution $\mu \in \mathcal{C}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$ of

$$\begin{aligned}\langle \mu_t, f\rangle = \langle \mu_0, f\rangle &+ \int_0^t \int_{\mathbb{R}_+} \Big(\partial_s f_s(x) + \alpha g(x)\partial_x f_s(x)\Big)\mu_s(dx)ds \\ &+ \int_0^t \int_{\mathbb{R}_+} \int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)\mu_s(dx)ds\end{aligned} \tag{1.13}$$

where $f_t(x) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ is a test function.

We are interested in the problem of estimation the division kernel h in the case where both the growth rate and the division rate are constants. Hence, we stick to $g(x) = 1$ and $B(x) = R$ for all $x \in \mathbb{R}_+$. If the measure μ_0 has a density, we can connect (1.13) to a growth-fragmentation PDE written under a more classical form:

Proposition 1.4.7. *We have the following results:*

i. If $\mu_0(dx) = n_0(dx)$ then $\forall t \in \mathbb{R}_+$, $\mu_t(dx)$ has a density $n(t,x)$.

ii. If $n(t,x) \in \mathcal{C}^{1,1}(\mathbb{R}_+, \mathbb{R}_+)$ then it satisfies the PDE:

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) + Rn(t,x) = 2R\int_0^1 n\left(t, \frac{x}{\gamma}\right)\frac{1}{\gamma}h(\gamma)d\gamma. \tag{1.14}$$

Proceeding as explained in Section 1.2.1, this leads us to the PDE (1.10), with a multiplicative convolution in the second term:

$$\alpha\partial_x N(x) + (\lambda + R)N(x) = 2R\int_0^{+\infty} N(y)h\Big(\frac{x}{y}\Big)\frac{dy}{y}.$$

We first change the variables in the PDE above to transform the multiplicative convolution into an additive one. Let us introduce the functions

$$g(u) = e^u h(e^u),$$

and

$$M(u) = e^u N(e^u), D(u) = \partial_u\Big(u \mapsto N(e^u)\Big) = \partial_u N(e^u).$$

Then the PDE (1.10) becomes

$$\alpha D(u) + (\lambda + R)M(u) = 2R\Big(M \star g)(u).$$

The estimator of h will be obtained from the estimator of g.

In the sequel, we denote by f^* the Fourier transform of an integrable function f defined by $f^*(\xi) = \int f(x)e^{\mathbf{i}x\xi}dx$.

Assume that the the density h and the Fourier transform M^* of M satisfy the Assumptions 3.3.1, 3.3.2 and 3.3.3 in Section 3.3.1 of **Chapter 3**, then the Fourier transform of g is given by:

$$g^*(\xi) = \frac{\alpha D^*(\xi)}{2RM^*(\xi)} + \frac{\lambda + R}{2R}, \quad \xi \in \mathbb{R}. \tag{1.15}$$

In view of (1.15), the purpose is first to propose an estimator for g^* and eventually to apply Fourier inversion to obtain an estimator of g.

Let K a kernel function in $\mathbb{L}^2(\mathbb{R})$ such that its Fourier transform K^* exists and is compactly supported. Define $K_\ell(\cdot) := \ell^{-1}K(\cdot/\ell)$ for $\ell > 0$. We set

$$g_\ell = K_\ell \star g.$$

Since $g_\ell^* = K_\ell^* \times g^*$, let $\widehat{M^*(\xi)}$ and $\widehat{D^*(\xi)}$ be unbiased estimators of M^* and D^* given in Proposition 3.3.6 (see Section 3.3.3), a natural estimator $\hat{g}_\ell$ of g is such that its Fourier transform takes the following form:

$$\widehat{g_\ell}^*(\xi) = K_\ell^*(\xi) \times \left(\frac{\alpha\widehat{D^*(\xi)}}{2R}\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} + \frac{\lambda + R}{2R}\right),$$

where $\Omega = \Big\{|\widehat{M^*(\xi)}| \geq n^{-1/2}\Big\}$ and $\mathbb{1}_\Omega/\widehat{M^*(\xi)}$ is the truncated estimator of $1/M^*(\xi)$ which is added to avoid an explosion when $\widehat{M^*(\xi)}$ is closed to 0.

Finally, taking inverse Fourier transform of $\widehat{g_\ell}^*$, we obtain

$$\hat{g}_\ell(u) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{g}_\ell^*(\xi)e^{-\mathbf{i}u\xi}d\xi.$$

Then the estimator of the division kernel h is given by

$$\hat{h}_\ell(\gamma) = \gamma^{-1}\hat{g}_\ell\Big(\log(\gamma)\Big), \quad \gamma \in (0,1).$$

We establish the $\mathbb{L}_2$-consistency of $\hat{g}_\ell$ under a suitable choice of the bandwidth ℓ in the following theorem. This result is proved in **Chapter 3**. Moreover, we consider only the consistency of the estimate $\hat{g}_\ell$ because of the difficulties explained in Section 3.4.

Theorem 1.4.8. *We suppose that Assumptions 3.3.1, 3.3.2 and 3.3.3 are satisfied and the kernel bandwidth ℓ which depends on n satisfies $\lim_{n\to+\infty} \ell = 0$. Provided that*

$$\lim_{n\to+\infty} \frac{1}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2 + \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right) = 0,$$

we have

$$\lim_{n\to+\infty} \mathbb{E}\left[\|\hat{g}_\ell - g\|_2^2\right] = 0.$$

1.4.2 Adaptive wavelet estimator for multivariate regression function in the errors-in-variables model

In this section, we present the statistical results for the estimation of the multivariate regression function $m(\cdot)$ in **Chapter 4**.

Notation. We denote by $\mathcal{F}$ the Fourier transform of any Lebesgue integrable function $f \in \mathbb{L}_1(\mathbb{R}^d)$ by

$$\mathcal{F}(f)(t) = \int_{\mathbb{R}^d} e^{-i<t,y>} f(y)dy, \quad t \in \mathbb{R}^d,$$

where $< .,. >$ denotes the usual scalar product.

Recall that we consider the following regression with errors-in-variables model in a multidimensional setting

$$Y_l = m(X_l) + \epsilon_l, \quad W_l = X_l + \delta_l, \quad l = 1,\ldots,n,$$

and f_X, f_W and g are the densities of the X_l's, the W_l's and the covariate errors δ_l's respectively. Our aim is to estimate the nonparametric multivariate regression function $m(x)$, $x \in [0,1]^d$. We provide an adaptive procedure in the multidimensional setting and study the pointwise risk over the anisotropic Hölder classes. For the ease of exposition, we refer to Section 4.2.1 of **Chapter 4**, pages 88 - 89 for the complete presentation of the assumptions on the regression function m, the design density f_X and the density of the errors covariates g.

To estimate the regression function m, the idea is that we try to estimate the conditional expectation

$$m(x) = \mathbb{E}[Y|X=x] = \frac{\int yf(x,y)dy}{f_X(x)} = \frac{p(x)}{f_X(x)},$$

where $f(x,y)$ denotes the joint density of (X,Y). Here we do not consider the problem of estimating the density f_X. We use an estimate $\hat{f}_X$ of f_X introduced in Comte and Lacour [27] (Section 3.4). This estimate is constructed from a deconvolution kernel and the bandwidth is selected by the GL method [53]. Hence the main task is to construct an estimate $\hat{p}$ of p. Then the estimator $\hat{m}$ of m at point x is given by

$$\hat{m}(x) = \frac{\hat{p}(x)}{\hat{f}_X(x)}.$$

For the reconstruction of $\hat{p}$ in the univariate case, using a kernel function $K(\cdot)$ with a bandwidth h_n, Fan and Truong [47] proposed an estimate of $\hat{p}$ with

$$\hat{p}_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n} Y_j L_n\left(\frac{x-W_j}{h_n}\right),$$

where $L(\cdot)$ is the deconvoluting kernel given by

$$L_n(x) = \frac{1}{2\pi}\int e^{-itx}\frac{\overline{\mathcal{F}(K)(t)}}{\mathcal{F}(g)(t/h_n)}dt.$$

Here we aim to construct an adaptive estimator of $p(x)$ using projection kernels on wavelets bases combined with a deconvolution operator which is a multidimensional wavelet analogous of $L_n(\cdot)$. Let φ a father wavelet satisfying the conditions (A1), (A2) and (A3) in Section 4.2.1 of **Chapter 4** and let $\left\{\varphi_{jk}, j\in\mathbb{Z}^d\right\}$, $j\in\mathbb{N}^d$ an orthonormal basis where for any x,

$$\varphi_{jk}(x) = \prod_{l=1}^{d} 2^{\frac{j_l}{2}}\varphi(2^{j_l}x_l - k_l), \quad j\in\mathbb{N}^d,\ k\in\mathbb{Z}^d.$$

Given a resolution level $j\in\mathbb{N}^d$, we define the estimator $\hat{p}_j(x)$ of $p(x)$ as follows

$$\hat{p}_j(x) = \frac{1}{n}\sum_{k}\sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u)\varphi_{jk}(x),$$

where the deconvolution operator $\mathcal{D}_j$ is defined as follows for a function f defined on $\mathbb{R}$

$$(\mathcal{D}_j f)(w) = \frac{1}{(2\pi)^d}\int e^{-i<t,w>}\prod_{l=1}^{d}\frac{\overline{\mathcal{F}(f)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt, w\in\mathbb{R}^d.$$

Furthermore, one can observe that the Fourier transform of the kernel K in the operator L_n has been replaced in our procedure by the Fourier transform of the wavelet φ_{jk} and the bandwidth h_n by 2^{-j}.

It remains to propose a method of selecting the wavelet resolution level j. We propose here a fully data-driven selection rule of the resolution level inspired from Goldenshluger and Lepski [53]. Then our estimation procedure automatically adapts to the unknown smoothness of the regression function to estimate. Details of the selection rule are described in Section 4.2.3 of **Chapter 4**.

Let $\hat{p}_{\hat{j}}$ be the final estimator of p where $\hat{j}$ is the adaptive index selected by our data-driven selection rule (see Section 4.2.3). We establish an oracle inequality (see Theorem 4.3.1 in Section 4.3.1) which highlights the bias-variance decomposition of the pointwise risk.

Let $\mathbb{H}_d(\vec{\beta}, L)$ be the anisotropic Hölder class with $\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_d) \in (\mathbb{R}_+^*)^d$ and $L > 0$ as in Definition 4.3.2, the following theorem give us the rates of convergence of the estimators $\hat{p}_{\hat{j}}$ over $\mathbb{H}_d(\vec{\beta}, L)$.

Theorem 1.4.9. *Let $q \geq 1$ be fixed and let $\hat{j}$ be the adaptive index. Let N be the number of vanishing moments of the father wavelet φ. Then, if for any l, $\lfloor \beta_l \rfloor \leq N$ and $L > 0$, it holds*

$$\sup_{p \in \mathbb{H}_d(\vec{\beta}, L)} \mathbb{E}\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_2 \left(\frac{\log n}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with $\bar{\beta} = \frac{1}{\frac{1}{\beta_1} + \cdots + \frac{1}{\beta_d}}$ and R_2 a constant depending on $\gamma, q, \varepsilon, \tilde{\gamma}, \mathfrak{m}, \mathfrak{d}, s, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

The constants $\mathfrak{m}$, $\mathfrak{d}$, c_g, $\mathcal{C}_g$ are related to the assumptions on m, f_X and g (see Assumptions 4.1 - 4.4) and the constants γ, $\tilde{\gamma}$ and ϵ are tuned constants constituted in the selection rule of the adaptive index $\hat{j}$ (see Section 4.2.3).

For the estimation of m, we define the estimator $\hat{m}$ for all x in $[0, 1]^d$:

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \tag{1.16}$$

where $\hat{f}_X$ is an estimate introduced in Comte and Lacour [27]. The term $n^{-1/2}$ is added to avoid the drawback when $\hat{f}_X$ is closed to 0.

The following theorems give the upper bounds and lower bounds for the estimator $\hat{m}$ under the pointwise risk. This shows that the estimate $\hat{m}$ achieves the optimal rate of convergence up to a logarithmic term.

Theorem 1.4.10. *Let $q \geq 1$ be fixed and let $\hat{m}$ defined as above. Then, if for any l, $\lfloor \beta_l \rfloor \leq N$ and $L > 0$, it holds*

$$\sup_{(m, f_X) \in \mathbb{H}_d(\vec{\beta}, L) \times \mathbb{H}_d(\vec{\beta}, L)} \mathbb{E}\left|\hat{m}(x) - m(x)\right|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_3 \left(\frac{\log n}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_3 a constant depending on $\gamma, q, \varepsilon, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

Theorem 1.4.11. *Let $q \geq 1$, $L > 0$ and for any l, $\lfloor \beta_l \rfloor \leq N$. Then for any estimator $\tilde{m}$ of m and for n large enough we have*

$$\sup_{(m, f_X) \in \mathbb{H}_d(\vec{\beta}, L) \times \mathbb{H}_d(\vec{\beta}, L)} \mathbb{E}\left|\tilde{m}(x) - m(x)\right|^q \geq R_4 n^{-q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_4 a positive constant depending on $\vec{\beta}$, L, s, C_g and $\mathcal{C}_g$.

We now describe briefly some numerical simulations to illustrate the theoretical results. We select the Doppler function for the regression function m and aim to estimate $\hat{m}$ at point $x_0 = 0.25$. The regression errors ϵ_l's are taken to be a standard normal variable with variance $s^2 = 0.15$. For the design density f_X, we consider Beta densities and the uniform density on $[0, 1]$. The uniform distribution is quite classical in regression with random design. The Beta$(2, 2)$ distribution give us symmetric distribution on $[0, 1]$ while the Beta$(0.5, 2)$ provides an asymmetric distribution with an accumulation of points near 0 and few points near 1. The choice of the Beta$(0.5, 2)$ density allows us to determine the influence of the design density at the boundaries of the interval $[0, 1]$. Moreover, despite the fact that Beta densities vanish in 0 and 1 and the design density f_X is assumed to be bounded from below (see Section 4.2.1), the choice of Beta distributions is still reasonable for simulations on any compact of $[0, 1]$ since the performances of the estimator are very bad at points very closed to 0 and 1. This is justified in Table 4.3 of Section 4.4.

For the choice of the density g of the covariate errors, we focus on the centered Laplace density with scale parameter $\sigma_{g_L} > 0$ that we denote g_L. This distribution has the explicit formula of the Fourier of transform of the g_L which implies the regularity of the noise density g_L. Moreover, σ_{g_L} is chosen according to the so-called reliability ratio (see Section 4.4 for details). We choose here $\sigma_L = 0.075$ and 0.1.

We estimate $m(x)$ by using Formula (1.16). First, we compute $\hat{p}_{\hat{j}}(x)$ an estimator of $p(x) = m(x) \times f_X(x)$ which is denoted "GL" in the graphics below. Then we divide $\hat{p}_{\hat{j}}(x)$ by the adaptive deconvolution density estimator $\hat{f}_X(x)$ of Comte and Lacour [27].

We compare the pointwise risk error of $\hat{p}_{\hat{j}}(x)$ (computed with 100 Monte Carlo repetitions) with the oracle risk one. The oracle is $\hat{p}_{j_{oracle}}$ with the index j_{oracle} defined as follows:

$$j_{oracle} := \arg\min_{j \in J} |\hat{p}_j(x) - p(x)|.$$

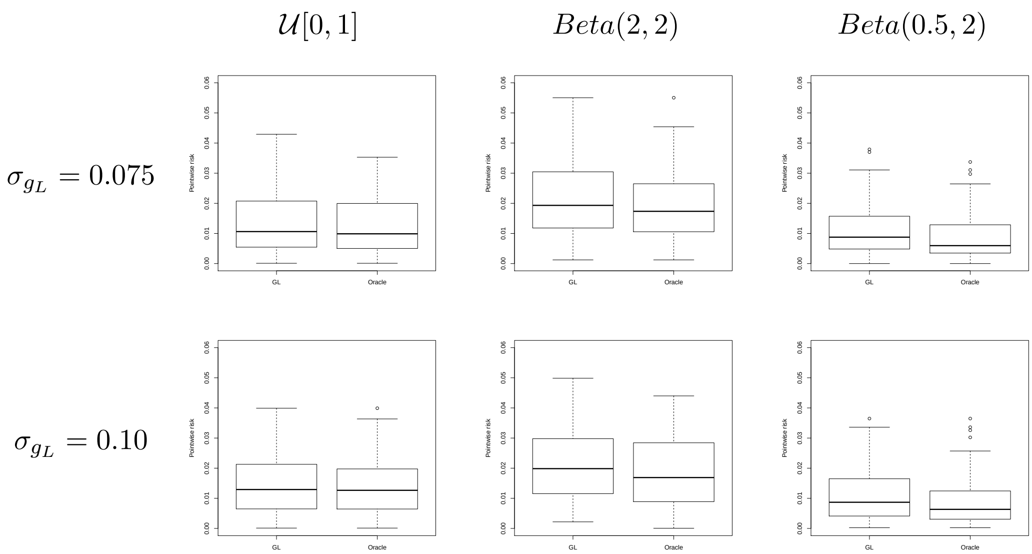


Figure 1.3: Estimation of $p(x)$ at $x_0 = 0.25$

Boxplots in Figure 1.3 illustrate the performances of our adaptive estimator at $x_0 = 0.25$: the risks of $\hat{p}_{\hat{j}}$ are close to those of the oracle. This shows that our performances are quite satisfying at $x_0 = 0.25$. Furthermore, one can observe that increasing the Laplace noise parameter σ_{g_L} deteriorates slightly the performances. Hence it seems that our procedure is robust to the noise in the covariates and accordingly to the deconvolution step. Finally, we refer to Section 4.4 for full details and discussions on the numerical results.

2

Estimating the division kernel of a size-structured population

Contents

This chapter is a version of the paper *Estimating the division kernel of a size-structured population* (Hoang [58]) submitted for publication.

2.1 Introduction

Models for populations of dividing cells possibly differentiated by covariates such as size have made the subject of an abundant literature in recent years (starting from Athreya and Ney [5], Harris [57], Jagers [67]...) Covariates termed as 'size' are variables that grow deterministically with time (such as volume, length, level of certain proteins, DNA content, *etc.*) Such models of structured populations provide descriptions for the evolution of the size distribution, which can be interesting for applications. For instance, in the spirit of Stewart et al. [99], we can imagine that each cell contains some toxicities whose quantity plays the role of the size. The asymmetric divisions of the cells, where one daughter contains more toxicity than the other, can lead under some conditions to the purge of the toxicity in the population by concentrating it into few lineages. These results are linked with the concept of aging for cell lineage. This concept has been tackled in many papers (e.g. Ackermann et al. [1], Aguilaniu et al. [2], Evans and Steinsaltz [44], C-Y. Lai et al. [74], Moseley [86]...).

Here we consider a stochastic individual-based model of size-structured population in continuous time, where individuals are cells undergoing asymmetric binary divisions and whose size is the quantity of toxicity they contain. A cell containing a toxicity $x \in \mathbb{R}_+$ divides at a rate $R > 0$. The toxicity grows inside the cell with rate $\alpha > 0$. When a cell divides, a random fraction $\Gamma \in [0,1]$ of the toxicity goes in the first daughter cell and $1-\Gamma$ in the second one. If $\Gamma = \frac{1}{2}$, the daughters are the same with toxicity $\frac{x}{2}$. We assume that Γ has a symmetric distribution on $[0,1]$ with a density h with respect to Lebesgue measure such that $\mathbb{P}(\Gamma = 0) = \mathbb{P}(\Gamma = 1) = 0$. If h is piked at $1/2$ (*i.e.* $\Gamma \simeq 1/2$), then both daughters contain the same toxicity, *i.e.* the half of their mother's toxicity. The more h puts weight in the neighbourhood of 0 and 1, the more asymmetric the divisions are, with one daughter having little toxicity and the other a toxicity close to its mother's one. If we consider that having a lot of toxicity is a kind of senescence, then, the kurtosis of h provides indication on aging phenomena (see [75]).

Modifications of this model to account for more complex phenomena have been considered in other papers. Bansaye and Tran [11], Cloez [24] or Tran [106] consider non-constant division and growth rates. Robert et al. [95] studies whether divisions can occur only when a size threshold is reached. Our purpose here is to estimate the density h ruling the divisions, and we stick to constant rates R and α for the sake of simplicity. Notice that several similar models for binary cell division in discrete time also exist in the literature and have motivated statistical question as here, see for instance Bansaye et al. [7, 10], Bercu et al. [12], Bitseki Penda [89], Delmas and Marsalle [34] or Guyon [54].

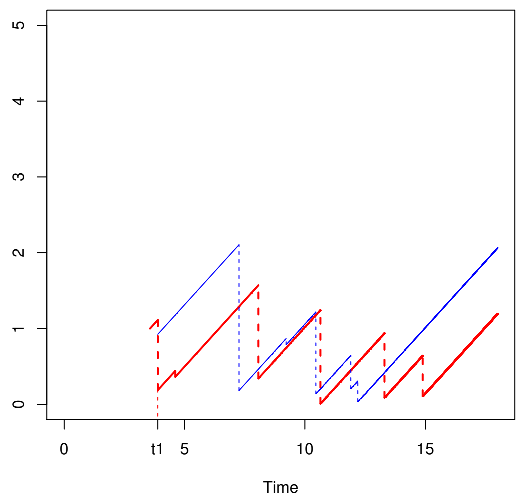


Figure 2.1: *Trajectories of two daughter cells after a division, separating after the first division at time t_1.*

Individual-based models provide a natural framework for statistical estimation. Estimation of the division rate is, for instance, the subject of Doumic et al. [38, 39] and Hoffmann and Olivier [59]. Here, the density h is the kernel division that we want to estimate. Assuming that we observe the divisions of cells in continuous time on the interval $[0, T]$, with $T > 0$, we propose an adaptive kernel estimator $\hat{h}$ of h for which we obtain an oracle inequality in Theorem 2.2.12. The construction of $\hat{h}$ is detailed in the sequel. From oracle inequality we can infer adaptive exponential rates of convergence with respect to T depending on β the smoothness of the density. Most of the time, nonparametric rates are of the form $n^{-\frac{2\beta}{2\beta+1}}$ (see for instance Tsybakov [108]) and exponential rates are not often encountered in the literature. The exponential rates are due to binary splitting, the number of cells $i.e$ the sample size increases exponentially in $\exp(RT)$ (see Section 2.3). By comparison, in [59] Hoffmann and Olivier obtain a similar rate of convergence $\exp\left(-\lambda_B \frac{\varsigma}{2\varsigma+1} T\right)$ of the kernel estimator of their division rate $B(x)$, where λ_B is the Malthus parameter and $\varsigma > 0$ is the smoothness of $B(x)$. However, their estimator $\hat{B}_T$ of B is not adaptive since the choice of their optimal bandwidth still depends on ς. Our estimator is adaptive with an "optimal" bandwidth chosen from a data-driven method. We derive upper bounds and lower bounds for asymptotic minimax risks on Hölder classes and show that they coincide. Hence, the rate of convergence of our estimator $\hat{h}$ proves to be optimal in the minimax sense on the Hölder classes.

This chapter is organized as follows. In Section 2, we introduce a stochastic differential equation driven by a Poisson point measure to describe the population of cells. Then, we construct the estimator of h and obtain upper and lower bounds for the MISE (Mean Integrated Squared Error). Our main results are stated in Theorems 2.2.15 and 2.2.16. Numerical results and discussions about aging effect are presented in Section 3. The main proofs are shown in Section 4.

Notation We introduce some notations used in the sequel.

Hereafter, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the $\mathbb{L}^1$ and $\mathbb{L}^2$ norms on $\mathbb{R}$ with respect to Lebesgue measure:

$$\|f\|_1 = \int_{\mathbb{R}} |f(\gamma)| d\gamma, \quad \|f\|_2 = \left(\int_{\mathbb{R}} |f(\gamma)|^2 d\gamma \right)^{1/2}.$$

The $\mathbb{L}^\infty$ norm is defined by

$$\|f\|_\infty = \sup_{\gamma \in (0,1)} |f(\gamma)|.$$

Finally, $f \star g$ denotes the convolution of two functions f and g defined by

$$f \star g(\gamma) = \int_{\mathbb{R}} f(u) g(\gamma - u) du.$$

2.2 Microscopic model and kernel estimator of h

2.2.1 The model

We recall the Ulam-Harris-Neveu notation used to describe the genealogical tree. The first cell is labelled by $\emptyset$ and when the cell i divides, the two descendants are labelled by $i0$ and $i1$. The set of labels is

$$J = \{\emptyset\} \cup \bigcup_{m=1}^{\infty} \{0,1\}^m . \tag{2.1}$$

We denote V_t the set of cells alive at time t, and $V_t \subset J$.

Let $\mathcal{M}_F(\mathbb{R}_+)$ be the space of finite measures on $\mathbb{R}_+$ embedded with the topology of weak convergence and X_t^i be the quantity of toxicity in the cell i at time t, we describe the population of cells at time t by a random point measure in $\mathcal{M}_F(\mathbb{R}_+)$:

$$Z_t(dx) = \sum_{i=1}^{N_t} \delta_{X_t^i}(dx), \qquad \text{where} \qquad N_t = \langle Z_t, 1\rangle = \int_{\mathbb{R}_+} Z_t(dx) \tag{2.2}$$

is the number of individuals living at time t. For a measure $\mu \in \mathcal{M}_F(\mathbb{R}_+)$ and a positive function f, we use the notation $\langle \mu, f\rangle = \int_{\mathbb{R}_+} f d\mu$.

Along branches of the genealogical tree, the toxicity $(X_t, t \geq 0)$ satisfies

$$dX_t = \alpha dt, \tag{2.3}$$

with $X_0 = x_0$. When the cells divide, the toxicity is shared between the daughter cells. This is described by the following stochastic differential equation (SDE).

Let $Z_0 \in \mathcal{M}_F(\mathbb{R}_+)$ be an initial condition such that

$$\mathbb{E}(\langle Z_0, 1\rangle) < +\infty, \tag{2.4}$$

and let $Q(ds, di, d\gamma)$ be a Poisson point measure on $\mathbb{R}_+ \times \mathcal{E} := \mathbb{R}_+ \times J \times [0,1]$ with intensity $q(ds, di, d\gamma) = R\, ds\, n(di) H(d\gamma)$. $n(di)$ is the counting measure on J and ds is Lebesgue measure on $\mathbb{R}_+$. We denote $\{\mathcal{F}_t\}_{t\geq 0}$ the canonical filtration associated with the Poisson point measure and the initial condition. The stochastic process $(Z_t)_{t\geq 0}$ can be described by a SDE as follows.

Definition 2.2.1. For every test function $f_t(x) = f(x,t) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ (bounded of class $\mathcal{C}^1$ in t and x with bounded derivatives), the population of cells is described by:

$$\begin{aligned}\langle Z_t, f_t\rangle &= \langle Z_0, f_0\rangle + \int_0^t \int_{\mathbb{R}_+} \Big(\partial_s f_s(x) + \alpha \partial_x f_s(x)\Big) Z_s(dx) ds \\ &+ \int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i\leq N_{s-}\}} \Big[f_s\left(\gamma X_{s-}^i\right) + f_s\left((1-\gamma) X_{s-}^i\right) - f_s\left(X_{s-}^i\right)\Big] Q(ds, di, d\gamma). \end{aligned}\tag{2.5}$$

The second term in the right hand side of (2.5) corresponds to the growth of toxicities in the cells and the third term gives a description of cell divisions where the sharing of toxicity into two daughter cells depends on the random fraction Γ.

We now state some properties of N_t that are useful in the sequel.

Proposition 2.2.2. *Let $T > 0$, and assume the initial condition N_0, the number of mother cells at time $t = 0$, is deterministic, for the sake of simplicity. We have*

i) Let T_j be the j^{th} jump time. Then:

$$\lim_{j\to+\infty} T_j = +\infty \text{ and } \lim_{T\to+\infty} N_T = +\infty \quad (a.s). \tag{2.6}$$

ii) N_T is distributed according to a negative binomial distribution, denoted as $\mathcal{NB}(N_0, e^{-RT})$. Its probability mass function is then

$$\mathbb{P}\left(N_T = n\right) = \binom{n-1}{n-N_0}\left(e^{-RT}\right)^{N_0}\left(1-e^{-RT}\right)^{n-N_0}, \tag{2.7}$$

for $n \geq N_0$. When $N_0 = 1$, N_T has a geometric distribution

$$\mathbb{P}\left(N_T = n\right) = e^{-RT}\left(1-e^{-RT}\right)^{n-1}. \tag{2.8}$$

Consequently, we have

$$\mathbb{E}\Big[N_T\Big] = N_0 e^{RT}. \tag{2.9}$$

iii) When $N_0 = 1$:

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \frac{RTe^{-RT}}{1-e^{-RT}}. \tag{2.10}$$

When $N_0 > 1$, we have:

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \left(\frac{e^{-RT}}{1-e^{-RT}}\right)^{N_0}(-1)^{N_0-1}\left(\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k e^{kRT}}{k} + RT\right). \tag{2.11}$$

iv) Furthermore, when $N_0 > 1$, we have

$$\frac{e^{-RT}}{N_0} \leq \mathbb{E}\left[\frac{1}{N_T}\right] \leq \frac{e^{-RT}}{N_0 - 1}. \tag{2.12}$$

The proof of Proposition 2.2.2 is presented in Section 4.

2.2.2 Influence of age

In this section, we study the aging effect via the mean age which is defined as follows.

Definition 2.2.3. The mean age of the cell population up to time $t \in \mathbb{R}_+$ is defined by:

$$\bar{X}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} X_t^i = \frac{\langle Z_t, f\rangle}{N_t}, \tag{2.13}$$

where $f(x) = x$.

Following the work of Bansaye et al. [8], we note that the long time behavior of the mean age is related to the law of an auxiliary process Y started at $Y_0 = \frac{X_0}{N_0}$ with infinitesimal generator characterized for all $f \in \mathcal{C}_b^{1,1}(\mathbb{R}_+, \mathbb{R})$ by

$$Af(x) = \alpha f'(x) + 2R \int_0^1 (f(\gamma x) - f(x))\, h(\gamma) d\gamma. \tag{2.14}$$

The empirical distribution $\frac{1}{N_t} \sum_{i=1}^{N_t} \delta_{X_t^i}$ gives the law of the path of a particle chosen at random at time t. Heuristically, the distribution of Y restricted to $[0, t]$ approximates this distribution. Hence, this explains the coefficient 2 which is a size-biased phenomenon, *i.e.* when one chooses a cell in the population at time t, a cell belonging to a branch with more descendants is more likely to be chosen.

Lemma 2.2.4. *Let Y be the auxiliary process with infinitesimal generator (2.14), for $t \in \mathbb{R}_+$,*

$$Y_t = \left(Y_0 - \frac{\alpha}{R}\right) e^{-Rt} + \frac{\alpha}{R} + \int_0^t e^{-R(t-s)} dU_s. \tag{2.15}$$

where U_t is a square-integrable martingale.

Consequently, we have

$$\mathbb{E}[Y_t] = \left(Y_0 - \frac{\alpha}{R}\right) e^{-Rt} + \frac{\alpha}{R}, \tag{2.16}$$

and

$$\lim_{t\to\infty} \mathbb{E}[Y_t] = \frac{\alpha}{R}. \tag{2.17}$$

We will show that the auxiliary process Y satisfies ergodic properties (see Section 2.4) which entails the following theorem.

Theorem 2.2.5. *Assume that there exists $\underline{h} > 0$ such that for all $\gamma \in (0, 1)$, $h(\gamma) \geq \underline{h}$. Then*

$$\lim_{t\to+\infty} \bar{X}_t = \lim_{t\to+\infty} \mathbb{E}(Y_t) = \frac{\alpha}{R}. \tag{2.18}$$

Theorem 2.2.5 is a consequence of the ergodic properties of Y, of Theorem 4.2 in Bansaye et al. [8] and of Lemma 2.2.4. It shows that the average of the mean age tends to the constant α/R when the time t is large. Simulations in Section 3 illustrate the results. The proofs of Lemma 2.2.4 and Theorem 2.2.5 are presented in Section 2.4 and Section 2.4.

Remark 2.2.6. When the population is large, we are interested in studying the asymptotic behavior of the random point measure. As in Doumic et al. [39], we can show that our stochastic model is approximated by a growth-fragmentation partial differential equation. This problem is a work in progress.

2.2.3 Estimation of the division kernel

Data and construction of the estimator

Suppose that we observe the evolution of the cell population in a given time interval $[0, T]$. At the i^{th} division time t_i, let us denote j_i the individual who splits into two daughters $X_{t_i}^{j_i 0}$ and $X_{t_i}^{j_i 1}$ and define

$$\Gamma_i^0 = \frac{X_{t_i}^{j_i 0}}{X_{t_i -}^{j_i}} \quad \text{and} \quad \Gamma_i^1 = \frac{X_{t_i}^{j_i 1}}{X_{t_i -}^{j_i}},$$

the random fractions that go into the daughter cells, with the convention $\frac{0}{0} = 0$.

Γ_i^0 and Γ_i^1 are exchangeable with $\Gamma_i^0 + \Gamma_i^1 = 1$, Γ_i^0 and Γ_i^1 are thus not independent but the couples $(\Gamma_i^0, \Gamma_i^1)_{i \in \mathbb{N}^*}$ are independent and identically distributed with distribution (Γ^0, Γ^1) where $\Gamma^1 \sim H(d\gamma)$ and $\Gamma^0 = 1 - \Gamma^1$.

Since h is a density function, it is natural to use a kernel method. We define an estimator $\hat{h}_\ell$ of h based on the data $(\Gamma_i^0, \Gamma_i^1)_{i \in \mathbb{N}^*}$ as follows.

Definition 2.2.7. Let $K : \mathbb{R} \longrightarrow \mathbb{R}$ is an integrable function such that

$$\int_{\mathbb{R}} K(x)dx = 1 \quad \text{and} \quad \int_{\mathbb{R}} K^2(x)dx < \infty.$$

Let M_T be the random number of divisions in the time interval $[0, T]$ and assume that $M_T > 0$. For all $\gamma \in (0, 1)$, define

$$\hat{h}_\ell(\gamma) = \frac{1}{M_T} \sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1), \tag{2.19}$$

where $K_\ell = \frac{1}{\ell} K(\cdot/\ell)$, $\ell > 0$ is the bandwidth to be chosen.

Remark 2.2.8. Since $N_0 \neq 0$, the number of random divisions M_T is not equal to the number of individuals living at time T. Indeed, we have $M_T = N_T - N_0$.

In (2.19), $\hat{h}_\ell$ depends also on T. However, we omit T for the sake of notation. The estimator $\hat{h}_\ell$ will satisfy the following properties.

Proposition 2.2.9.

i) The conditional expectation and conditional variance given M_T of $\hat{h}_\ell(\gamma)$ and variance $\hat{h}_\ell(\gamma)$ are:

$$\mathbb{E}\Big[\hat{h}_\ell(\gamma)|M_T\Big] = K_\ell \star h(\gamma) \;\; \textit{and} \;\; \mathbb{E}\Big[\hat{h}_\ell(\gamma)\Big] = K_\ell \star h(\gamma), \tag{2.20}$$

$$\mathbb{V}ar\Big[\hat{h}_\ell(\gamma)\Big|M_T] = \frac{1}{M_T} \mathbb{V}ar\left[K_\ell(\gamma - \Gamma_1^1)\right], \tag{2.21}$$

$$\mathbb{V}ar\Big[\hat{h}_\ell(\gamma)\Big] = \mathbb{E}\Big[\frac{1}{M_T}\Big] \mathbb{V}ar\left[K_\ell(\gamma - \Gamma_1^1)\right]. \tag{2.22}$$

Consequently, we have $\mathbb{E}\Big[\hat{h}_\ell(\gamma)|M_T\Big] = \mathbb{E}\Big[\hat{h}_\ell(\gamma)\Big]$.

ii) For all $\gamma \in (0, 1)$,

$$\lim_{T \to +\infty} \hat{h}_\ell(\gamma) = K_\ell \star h(\gamma) \quad \textit{(a.s).} \tag{2.23}$$

Adaptive estimation of h by the Goldenshluger-Lepski's (GL) method

Let $\hat{h}_\ell$ be the kernel estimator of h as in Definition 2.2.7. We measure the performance of $\hat{h}_\ell$ via its $\mathbb{L}^2$-loss *i.e* the average $\mathbb{L}^2$ distance between $\hat{h}_\ell$ and h. The objective is to find a bandwidth which minimizes this $\mathbb{L}^2$-loss. Since M_T is random, we first study the $\mathbb{L}^2$-loss conditionally to M_T.

Proposition 2.2.10. *The $\mathbb{L}^2$-loss of $\hat{h}_\ell$ given M_T satisfies :*

$$\mathbb{E}\Big[\|\hat{h}_\ell - h\|_2\Big|M_T\Big] \leq \|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T\ell}}. \tag{2.24}$$

In the right hand side of the risk decomposition (2.24) the first term is a bias term. Hence it decreases when $\ell \to 0$ whereas the second term which is a variance term increases when $\ell \to 0$. The best choice of ℓ should minimize this bias-variance trade-off. Thus, from a finite family of bandwidths H, the best bandwidth $\bar{\ell}$ would be

$$\bar{\ell} := \underset{\ell \in H}{\operatorname{argmin}} \Big\{\|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T\ell}}\Big\}. \tag{2.25}$$

The bandwidth $\bar{\ell}$ is called "the oracle bandwidth" since it depends on h which is unknown and then it cannot be used in practice. Since the oracle bandwidth minimizes a bias variance trade-off, we need to find an estimation for the bias-variance decomposition of $\hat{h}_\ell$. Goldenshluger and Lepski [53] developed a fully data-driven bandwidth selection method (GL method). The main idea of this method is based on an estimate of the bias term by looking at several estimators. In a similar fashion, Doumic et al. [39] and Reynaud-Bouret et al. [94] have used this method. To apply the GL method, we set for any ℓ, $\ell' \in H$:

$$\hat{h}_{\ell,\ell'} := \frac{1}{M_T}\sum_{i=1}^{M_T}\Big(K_\ell \star K_{\ell'}\Big)(\gamma - \Gamma_i^1) = \Big(K_\ell \star \hat{h}_{\ell'}\Big)(\gamma).$$

Finally, the adaptive bandwidth and the estimator of h are selected as follows:

Definition 2.2.11. Given $\epsilon > 0$ and setting $\chi := (1+\epsilon)(1+\|K\|_1)$, we define

$$\hat{\ell} := \underset{\ell \in H}{\operatorname{argmin}} \Big\{A(\ell) + \frac{\chi\|K\|_2}{\sqrt{M_T\ell}}\Big\}, \tag{2.26}$$

where, for any $\ell \in H$,

$$A(\ell) := \sup_{\ell' \in H}\Big\{\|\hat{h}_{\ell,\ell'} - \hat{h}_{\ell'}\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T\ell'}}\Big\}_+, \tag{2.27}$$

Then, the estimator $\hat{h}$ is given by

$$\hat{h} := \hat{h}_{\hat{\ell}}. \tag{2.28}$$

An inspection of the proof of Theorem 2.2.12 shows that the term $A(\ell)$ provides a control for the bias $\|h - K_\ell \star h\|_2$ up to the term $\|K\|_1$ (see (2.45) and (2.47) in the proof of Theorem 2.2.12, Section 4). Since $A(\ell)$ depends only on $\hat{h}_{\ell,\ell'}$ and $\hat{h}_{\ell'}$, the estimator $\hat{h}$ can be computed in practice.

We shall now state an oracle inequality which highlights the bias-variance decomposition of the MISE of $\hat{h}$. We recall that the MISE of $\hat{h}$ is the quantity $\mathbb{E}\Big[\|\hat{h}-h\|_2^2\Big]$.

Theorem 2.2.12. *Let $T > 0$ and assume that observations are taken on $[0,T]$. Let N_0 be the number of mother cells at the beginning of divisions and M_T is the random number of divisions in $[0,T]$. Consider H a countable subset of $\{\triangle^{-1} : \triangle = 1,\dots,\triangle_{\max}\}$ in which we choose the bandwidths and $\triangle_{\max} = \lfloor \delta M_T \rfloor$ for some $\delta > 0$. Assume $h \in L^\infty([0,1])$ and let $\hat{h}$ be a kernel estimator defined with the kernel $K_{\hat{\ell}}$ where $\hat{\ell}$ is chosen by the GL method. Define*

$$\varrho(T)^{-1} = \begin{cases} \dfrac{e^{-RT+\log(RT)}}{1-e^{-RT}}, & \text{if } N_0 = 1, \\ e^{-RT}, & \text{if } N_0 > 1. \end{cases} \tag{2.29}$$

For large T, the main term in $\varrho(T)$ is e^{-RT} in any case. It is exactly the order of $\varrho(T)$ for $N_0 > 1$. Then, given $\epsilon > 0$

$$\mathbb{E}\Big[\|\hat{h}-h\|_2^2\Big] \le C_1 \inf_{\ell \in H}\left\{\|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\varrho(T)^{-1}\right\} + C_2\varrho(T)^{-1}, \tag{2.30}$$

where C_1 is a constant depending on N_0, $\|K\|_1$ and ϵ and C_2 is a constant depending on N_0, δ, ϵ, $\|K\|_1$, $\|K\|_2$ and $\|h\|_\infty$.

The term $\|K_\ell \star h - h\|_2^2$ is an approximation term, $\frac{\|K\|_2^2}{\ell}\varrho(T)^{-1}$ is a variance term and the last term $\varrho(T)^{-1}$ is asymptotically negligible. Hence the right hand side of the oracle inequality corresponds to a bias variance trade-off.

We now establish upper and lower bounds for the MISE. The lower bound is obtained by perturbation methods (Theorem 2.2.16) and is valid for any estimator $\widehat{h}_T$ of h, thus indicating the optimal convergence rate. The upper bound is obtained in Theorem 2.2.15 thanks to the key oracle inequality of Theorem 2.2.12.

For the rate of convergence, it is necessary to assume that the density h and the kernel function K satisfy some regularity conditions introduced in the following definitions.

Definition 2.2.13. Let $\beta > 0$ and $L > 0$. The Hölder class of smoothness β and radius L is defined by

$$\mathcal{H}(\beta, L) = \Big\{ f : f \text{ has } k = \lfloor\beta\rfloor \text{ derivatives and } \forall x, y \in \mathbb{R}$$
$$\Big|f^{(k)}(y) - f^{(k)}(x)\Big| \le L|x-y|^{\beta-k}\Big\}.$$

Definition 2.2.14. Let $\beta^* > 0$. An integrable function $K : \mathbb{R} \to \mathbb{R}$ is a kernel of order β^* if

- $\int K(x)dx = 1$,
- $\int |x|^{\beta^*}|K(x)|dx < \infty$,
- For $k = \lfloor \beta^* \rfloor$, $\forall\, 1 \leq j \leq k$, $\int x^j K(x)dx = 0$.

Then, the following theorem gives the rate of convergence of the adaptive estimator $\hat{h}$.

Theorem 2.2.15. *Let $\beta^* > 0$ and K be a kernel of order β^*. Let $\beta \in (0, \beta^*)$. Let $\hat{\ell}$ be the adaptive bandwidth defined in (2.26). Then, for any $T > 0$, the kernel estimator $\hat{h}$ satisfies*

$$\sup_{h \in \mathcal{H}(\beta,L)} \mathbb{E}\|\hat{h} - h\|_2^2 \leq C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}}, \tag{2.31}$$

where $\varrho(T)^{-1}$ is defined in (2.29) and C_3 is a constant depending on N_0, δ, ϵ, $\|K\|_1$, $\|K\|_2$, $\|h\|_\infty$, β and L.

We now establish a lower bound in Theorem 2.2.16.

Theorem 2.2.16. *For any $T > 0$, $\beta > 0$ and $L > 0$. Assume that $h \in \mathcal{H}(\beta, L)$, then there exists a constant $C_4 > 0$ such that for any estimator $\hat{h}_T$ of h*

$$\sup_{h \in \mathcal{H}(\beta,L)} \mathbb{E}\|\hat{h}_T - h\|_2^2 \geq C_4 \exp\left(-\frac{2\beta}{2\beta+1} RT\right). \tag{2.32}$$

Contrary to the classical cases of nonparametric estimation (*e.g.* Tsybakov [108], ...), the number of observations M_T is a random variable that converges to $+\infty$ when $T \to +\infty$ which is one of the main difficulty here. From Theorem 2.2.15, when $N_0 > 1$ the upper bound is in $\exp\left(-\frac{2\beta}{2\beta+1}RT\right)$ which is the same rate as the lower bound. The rate of convergence $\hat{h}$ is thus optimal. When $N_0 = 1$, the upper bound is in $\exp\left(\frac{2\beta}{2\beta+1}\Big(-RT + \log(RT)\Big)\right)$ that differs with a logarithmic from the rate in the lower bound. The rate of convergence is thus slightly slower than in the case $N_0 > 1$ and our estimator is optimal up to a logarithmic factor. Furthermore, Theorem 2.2.15 illustrates adaptive properties of our procedure: it achieves the rate $\varrho(T)^{-\frac{2\beta}{2\beta+1}}$ over the Hölder class $\mathcal{H}(\beta, L)$ as soon as β is smaller than β^*. So, it automatically adapts to the unknown smoothness of the signal to estimate.

2.3 Numerical simulations

2.3.1 Numerical computation of $\hat{h}$

We use the **R** software to implement simulations with two original distributions of division kernel h and compare with their estimators. On the interval $[0, 1]$, the first distribution to test is Beta(2, 2). The Beta(a, b) distributions on $[0, 1]$ are characterized by their densities

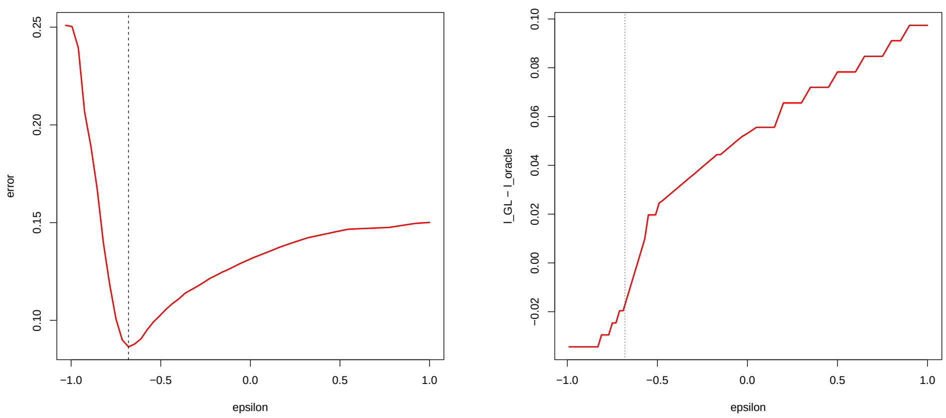
$$h_{\text{Beta}(a,b)}(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathcal{B}(a,b)}.$$

where $\mathcal{B}(a, b)$ is the renormalization constant.

Since h is symmetric, we only consider the distributions with $a = b$. Generally, asymmetric divisions correspond to $a < 1$ and symmetric divisions with kernels concentrated around $\frac{1}{2}$ correspond to $a > 1$. The smaller the parameter a, the more asymmetric the divisions. For the second density, we choose a Beta mixture distribution as

$$\frac{1}{2}\,\text{Beta}(2,6) + \frac{1}{2}\,\text{Beta}(6,2).$$

This choice gives us a bimodal density corresponding to very asymmetric divisions.



(a) (b)

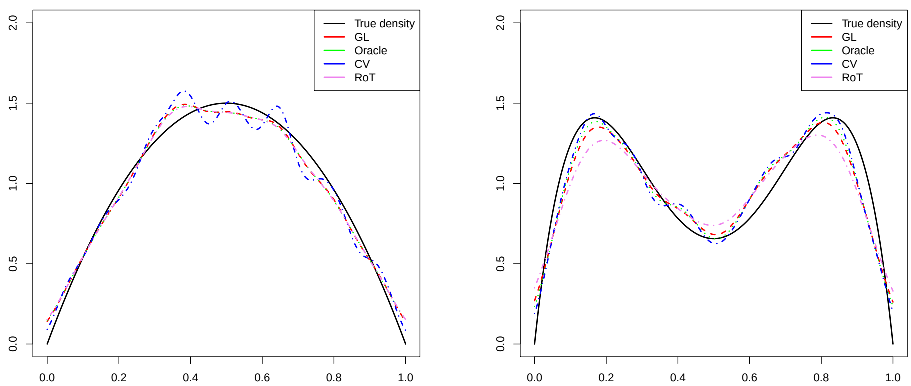
Figure 2.2: *(a): MISE's as a function of ϵ. (b): $\hat{\ell} - \ell_{oracle}$ as a function of ϵ. The dotted lines indicate the optimal value of ϵ which is used in all simulations.*

We estimate $\hat{h}$ by using (2.19) and we take the classical Gaussian kernel $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. For the choice of bandwidth, we apply the GL method with the family $H = \left\{1, 2^{-1}, \ldots, \lfloor \delta M_T \rfloor^{-1}\right\}$ for some $\delta > 0$ small enough when M_T is large to reduce the time of numerical simulation. We have $\|K\|_1 = 1$, $\|K\|_2 = 2^{-1/2}\pi^{-1/4}$ and $K_\ell \star K_{\ell'} = K_{\sqrt{\ell^2+\ell'^2}}$, hence it is not difficult to calculate in practice $\hat{h}_{\ell,\ell'}$ as well as $\hat{h}_{\ell'}$. Finally, the value of ϵ in $\chi = (1+\epsilon)(1+\|K\|_1)$ is chosen to find an optimal value of the MISE. To do this, we implement a preliminary simulation to calibrate ϵ in which we choose $\epsilon > -1$ to ensure that $1+\epsilon > 0$. We compute the MISE and $\hat{\ell} - \ell_{\text{oracle}}$ as functions of ϵ where $\ell_{oracle} = \text{argmin}_{\,\ell \in H}\,\mathbb{E}\left[\|\hat{h}_\ell - h\|_2^2\right]$ and h is the density of Beta(2, 2). In Figure 2.2a, simulation results show that the risk has minimum value at $\epsilon = -0.68$. This value is not justified from a theoretical point of view. The theoretical choice $\epsilon > 0$ (see Theorem 2.2.12) does not give bad results but this choice is too conservative for non-asymptotic practical purposes as often met in the literature (see Bertin et al. [14] for more details about the GL methodology). Moreover, following the discussion in Lacour and Massart [73] we investigate (see Figure 2.2b) the difference $\hat{\ell} - \ell_{\text{oracle}}$ and observe some explosions close to $\epsilon = -0.68$. Consequently, we choose $\epsilon = -0.68$ for all following simulations.

Figure 2.3 illustrates a reconstruction for the density of Beta(2, 2) and beta mixture $\frac{1}{2}\,\text{Beta}(2,6) + \frac{1}{2}\,\text{Beta}(6,2)$ when $T = 13$. We choose here the division rate and the growth rate $R = 0.5$ and $\alpha = 0.35$ respectively. We compare the estimated densities when using the GL bandwidth with those estimated with the oracle bandwidth. The oracle bandwidth is found by assuming that we know the true density. Moreover, the GL estimators are compared with estimators using the cross-validation (CV) method and the rule of thumb (RoT). The CV bandwidth is defined as follows:

$$\ell_{CV} = \underset{\ell \in H}{\text{argmin}} \left\{ \int \hat{h}_\ell^2(\gamma) d\gamma - \frac{2}{n} \sum_{i=1}^{n} \hat{h}_{\ell,-i}(\Gamma_i^1) \right\}$$

where $\hat{h}_{\ell,-i}(\gamma) = \frac{1}{n-1} \sum_{j \neq i} K_\ell\Big(\Gamma_j^1 - \gamma\Big)$. The RoT bandwidth can be calculated simply by using the formula $\ell_{RoT} = 1.06\hat{\sigma} n^{-1/5}$ where $\hat{\sigma}$ is the standard deviation of the sample $(\Gamma_1^1, \ldots, \Gamma_n^1)$. More details about these methods can be found in Section 3.4 of Silverman [98] or Tsybakov [108].



(a) *Reconstruction of* Beta(2, 2) (b) *Reconstruction of beta mixture*

Figure 2.3: *Reconstruction of division kernels with* $T = 13$.

To estimate the MISE, we implement Monte-Carlo simulations with respect to $T = 13, 17$ and 20. The number of repetitions for each simulation is $\mathcal{M} = 100$. Then, we compute the mean of relative error $\bar{e} = (1/\mathcal{M}) \sum_{i=1}^{\mathcal{M}} e_i$ and the standard deviation $\sigma_e = \sqrt{(1/\mathcal{M}) \sum_{i=1}^{\mathcal{M}} (e_i - \bar{e})^2}$ where

$$e_i = \frac{\|\hat{h}^{(i)} - h\|_2}{\|h\|_2}, \quad i = 1, \ldots, \mathcal{M}, \tag{2.33}$$

and $\hat{h}^{(i)}$ denotes the estimator of h corresponding to i^{th} repetition.

		GL	Oracle	CV	RoT	ML method
$T = 13$	$\bar{e}$	0.1001	0.0840	0.1009	0.0900	0.0610
	σ_e	0.0585	0.0481	0.0599	0.0577	0.0724
	$\bar{\hat{\ell}}$	0.0920	0.0845	0.0824	0.0727	
$T = 17$	$\bar{e}$	0.0458	0.0397	0.0459	0.0405	0.0166
	σ_e	0.0260	0.0230	0.0297	0.0237	0.0171
	$\bar{\hat{\ell}}$	0.0485	0.0497	0.0478	0.0470	
$T = 20$	$\bar{e}$	0.0261	0.0241	0.0262	0.0245	0.0088
	σ_e	0.0140	0.0114	0.0132	0.00121	0.0091
	$\bar{\hat{\ell}}$	0.0377	0.0359	0.0345	0.0354	

Table 2.1: *Mean of relative error and its standard deviation for the reconstruction of* Beta$(2,2)$. $\bar{\hat{\ell}}$ *is the average of bandwidths for* $M = 100$ *samples.*

		GL	Oracle	CV	RoT
$T = 13$	$\bar{e}$	0.1361	0.1245	0.1379	0.1686
	σ_e	0.0672	0.0562	0.0815	0.0537
	$\bar{\hat{\ell}}$	0.0618	0.0527	0.0522	0.0948
$T = 17$	$\bar{e}$	0.0539	0.0534	0.0550	0.0919
	σ_e	0.0180	0.0168	0.0168	0.00223
	$\bar{\hat{\ell}}$	0.0309	0.0272	0.0264	0.0590

Table 2.2: *Mean of relative error and its standard deviation for the reconstruction of beta mixture* $\frac{1}{2}\,\mathrm{Beta}(2,6) + \frac{1}{2}\,\mathrm{Beta}(6,2)$.

The MISE's are computed for estimated densities using the GL bandwidth, the oracle bandwidth, the CV bandwidth and the RoT bandwidth. For a further comparison, in the reconstruction of Beta$(2,2)$, we compute the relative error in a parametric setting by comparing the true density h with the density of Beta$(\hat{a},\hat{a})$ where $\hat{a}$ is a Maximum Likelihood (ML) estimator the parameter a. The simulation results are displayed in Table 2.1 and Table 2.2. For the density of Beta mixture, we only compute the error with $T = 13$ and $T = 17$. The boxplot in Figure 4 illustrates the MISE's in Table 2.1 when $T = 17$.

From Tables 2.1 and 2.2, we can note that the accuracy of the estimation of Beta$(2,2)$ and Beta mixture by the GL bandwidth increases for larger T. In Figure 2.5, we illustrate on a log-log scale the mean relative error and the rate of convergence versus time T. This shows that the error is close to the exponential rate predicted by the theory. Moreover, we can observe that the errors of Beta mixture are larger than those of Beta$(2,2)$ with the same T due to the complexity of its density. In both cases, the error estimated by using oracle bandwidth is always smaller. The GL error is slightly smaller than the CV error. The RoT error can show very good behavior but lacks of stability. Overall, we conclude that the GL method has a good behavior when compared to the cross validation method and rule-of-thumb. As usual, we also see that the ML errors are quite smaller than those of nonparametric approach but the magnitude of the mean $\bar{e}$ remains similar.

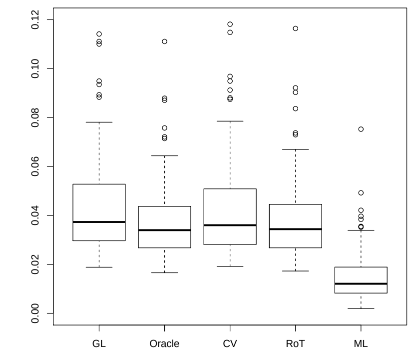


Figure 2.4: *Errors of estimated densities of* Beta(2, 2) *when* $T = 17$.

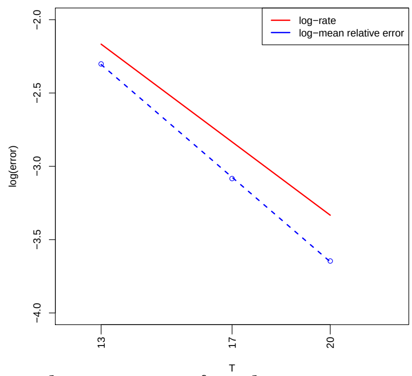


Figure 2.5: *The log-mean relative error for the reconstruction of* Beta(2, 2) *compared to the log-rate (solid line) computed with* $\beta = 1$.

Since h is symmetric on $[0, 1]$ with respect to $\frac{1}{2}$, the estimator $\hat{h}$ can be improved and we can introduce

$$\tilde{h}(x) = \frac{1}{2}\left(\hat{h}(x) + \hat{h}(1 - x)\right),$$

which is symmetric by construction and satisfies also (2.31). We compute the mean of relative error for the estimator $\tilde{h}$ with the estimation of Beta(2, 2) and Beta mixture. The results are displayed in Table 2.3. Compared with the error in Table 2.1 and 2.2, one can see as expected that the errors for the reconstruction of $\tilde{h}$ are smaller. However, these errors are of the same order, indicating that the estimator $\hat{h}$ had already good symmetric properties.

		GL	Oracle	CV	RoT
Beta(2, 2)	$T = 13$	0.0785	0.0634	0.0762	0.0644
	$T = 17$	0.0356	0.0309	0.0356	0.0309
Beta mixture	$T = 13$	0.1117	0.0953	0.1030	0.1584
	$T = 17$	0.0450	0.0414	0.0417	0.0893

Table 2.3: *Mean of relative error for the reconstruction of* $\tilde{h}$.

2.3.2 Influence of the distribution on the mean age

For $t \geq 0$, recall the mean age defined in (2.13). To study the influence of the distribution on the mean age, we simulate $n = 50$ trees with respect to $t = 6, 6 + \triangle t, \dots, 24$ with $\triangle t = 0.36$. For each sample $(\bar{x}_t^{(1)}, \dots, \bar{x}_t^{(n)})$, we compute the average mean, the 1st (Q_{25}) quartile and 3rd (Q_{75}) quartile. Figure 2.6a and 2.6b show the simulation results corresponding to the density of Beta(2, 2) with $\alpha = 0.45$ and $R = 0.4$. One can see that the average of mean age and the mean age converge to $\frac{\alpha}{R} = 1.125$ for larger t. This agrees with the theoretical result proved in Section 2.2.2.

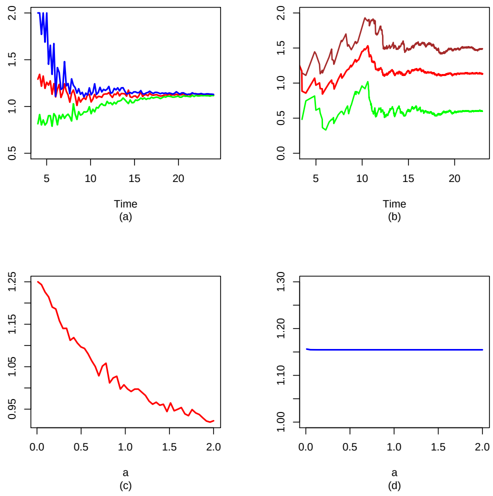


Figure 2.6: *(a) Average mean, 1st and 3rd quartiles for the sample of means for* 50 *trees. (b) Average mean, 1st and 3rd quartiles for one tree. (c) Average of* $Q_{75} - Q_{25}$ *with* $a \in [0, 2]$ *at* $t = 12$*. (d) Mean age with* $a \in [0, 2]$ *at* $t = 12$*.*

Moreover, Q_{25} and Q_{75} vary when the parameter a changes. In Figure 2.6c, we draw a fitted curve of the average of $(Q_{75} - Q_{25})$ when a varies from 0 to 2. As we mentioned in the introduction, if divisions are more asymmetric corresponding to small values of a, the toxicities concentrate on few cells, *i.e.* we have more older cells after the divisions. This explains the decreasing trend in the average of $(Q_{75} - Q_{25})$. Finally, Figure 2.6d displays the average of mean ages with respect to a. One can note that it does not change when we replace the kernel distribution, *e.g* Beta(0.6, 0.6) instead of Beta(2, 2).

2.4 Proofs

Proof of Proposition 2.2.2.

ii) The proof of ii) can be found easily in literature. Here we refer to [96], Section 5.3 for this proof.

i) Let us prove that $\lim_{T \to +\infty} N_T = \lim_{j \to +\infty} N_{T_j} = +\infty$. Since our model has only births and no death, $(N_t)_{t\in[0,T]}$ is a non-decreasing process: $N_{T_j} = N_0 + j$. All the T_j's are finite and $\lim_{j\to+\infty} N_{T_j} = +\infty$ a.s. From *ii)*, we have $\mathbb{E}[N_T] = N_0 e^{RT}$. Hence, we deduce from the estimate $\sup_{t\in[0,T]} \mathbb{E}[N_t] < +\infty$ for all $T > 0$ that $T_j \underset{j\to+\infty}{\longrightarrow} +\infty$ a.s. Then we also have $\lim_{T\to+\infty} N_T = +\infty$ a.s.

iii) Let $p = e^{-RT}$. When $N_0 = 1$, $N_T \sim \text{Geom}(p)$. Then we have

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \sum_{n=1}^{\infty} \frac{1}{n}\mathbb{P}\Big(N_T = n\Big) = \sum_{n=1}^{\infty} \frac{1}{n} p(1-p)^{n-1}$$
$$= \frac{p}{1-p}\sum_{n=1}^{+\infty} \frac{(1-p)^n}{n} = -\frac{p}{1-p}\log(p).$$

Replace p with e^{-RT}, we obtain (2.10).
When $N_0 > 1$, $N_T \sim \mathcal{NB}(N_0, p)$. Hence, we have

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \sum_{n=N_0}^{\infty} \frac{1}{n}\binom{n-1}{n-N_0} p^{N_0}(1-p)^{n-N_0}$$
$$= \left(\frac{p}{1-p}\right)^{N_0} \sum_{n=N_0}^{\infty} \frac{1}{n}\binom{n-1}{n-N_0}(1-p)^n$$
$$:= \left(\frac{p}{1-p}\right)^{N_0} f(1-p), \tag{2.34}$$

where $f(x) = \sum_{n=N_0}^{+\infty} \frac{1}{n}\binom{n-1}{n-N_0} x^n$. We can differentiate $f(x)$ by taking derivative under the sum. Then:

$$\frac{d}{dp} f(1-p) = -\sum_{n=N_0}^{+\infty} \binom{n-1}{n-N_0}(1-p)^{n-1}$$
$$= -\frac{(1-p)^{N_0-1}}{p^{N_0}} \sum_{n=N_0}^{+\infty} \binom{n-1}{n-N_0} p^{N_0}(1-p)^{n-N_0} = -\frac{1}{p}\left(\frac{1}{p}-1\right)^{N_0-1},$$

since the sum is 1 (we recognize the negative binomial).
Hence,

$$\frac{d}{dp} f(1-p) = -\frac{1}{p}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{1}{p^k}(-1)^{N_0-1-k} + (-1)^{N_0-1}\right]$$
$$=(-1)^{N_0}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{p^{k+1}} + \frac{1}{p}\right]. \tag{2.35}$$

Integrating equation (2.35) and notice that $f(0) = 0$, we get

$$f(1-p) =(-1)^{N_0}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{k}\Big(-\frac{1}{p^k}\Big) + \log(p)\right]$$
$$=(-1)^{N_0-1}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{k}\frac{1}{p^k} + \log\left(\frac{1}{p}\right)\right]. \tag{2.36}$$

Combine (2.34),(2.36) and replace p with e^{-RT}, we get (2.11).

iv) We first prove the lower bound of (2.12). From (2.5), taking $f_t(x) = 1$, we have

$$N_T = N_0 + \int_0^T \int_{\mathcal{E}} \mathbb{1}_{\{i \le N_{s-}\}} Q(ds, di, d\gamma). \tag{2.37}$$

Applying Itô formula for jump processes (see [63], Theorem 5.1 on p.67) to (2.37), we obtain

$$\begin{aligned} \frac{1}{N_T} &= \frac{1}{N_0} + \int_0^T \int_{\mathcal{E}} \left(\frac{1}{N_{s-}+1} - \frac{1}{N_{s-}} \right) \mathbb{1}_{\{i \le N_{s-}\}} Q(ds, di, d\gamma) \\ &= \frac{1}{N_0} - \int_0^T \int_{\mathcal{E}} \frac{1}{N_{s-}(N_{s-}+1)} \mathbb{1}_{\{i \le N_{s-}\}} Q(ds, di, d\gamma). \end{aligned}$$

Hence,

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \frac{1}{N_0} - \mathbb{E}\left[\int_0^T \frac{1}{N_s(N_s+1)} R N_s ds\right] = \frac{1}{N_0} - R\int_0^T \mathbb{E}\left[\frac{1}{N_s+1}\right] ds. \tag{2.38}$$

Since $N_s \ge N_0$, we have $\frac{1}{N_s+1} \le \frac{1}{N_s}$. Therefore, (2.38) implies that

$$\mathbb{E}\left[\frac{1}{N_T}\right] \ge \frac{1}{N_0} - R\int_0^T \mathbb{E}\left[\frac{1}{N_s}\right] ds. \tag{2.39}$$

By comparison of $\mathbb{E}\left[\frac{1}{N_T}\right]$ with the solutions of the ODE $\frac{d}{dT}u(T) = -Ru(T)$ with $u(0) = 1/N_0$, we finally obtain

$$\mathbb{E}\left[\frac{1}{N_T}\right] \ge \frac{1}{N_0} e^{-RT}.$$

For the upper bound, notice that $\mathbb{E}\left[\frac{1}{N_T}\right] \le \mathbb{E}\left[\frac{1}{N_T-1}\right]$ for $N_0 > 1$. Then we have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{N_T-1}\right] &= \sum_{n=N_0}^{+\infty} \frac{1}{n-1} \binom{n-1}{n-N_0} p^{N_0}(1-p)^{n-N_0} \\ &= \sum_{n=N_0}^{+\infty} \frac{(n-2)!}{(n-N_0)!(N_0-1)!} p^{N_0}(1-p)^{n-N_0} \\ &= \frac{p}{N_0-1} \sum_{n=N_0}^{+\infty} \frac{(n-2)!}{(n-N_0)!(N_0-2)!} p^{N_0-1}(1-p)^{n-N_0} \\ &= \frac{p}{N_0-1} \sum_{m=N_0-1}^{+\infty} \frac{(m-1)!}{(m-(N_0-1))!((N_0-1)-1)!} p^{N_0-1}(1-p)^{m-(N_0-1)} \\ &= \frac{p}{N_0-1} = \frac{e^{-RT}}{N_0-1}, \end{aligned}$$

by changing the index in the sum $(m = n-1)$ and by recognizing the negative binomial with parameter (N_0-1, p). Hence, we conclude that for $N_0 > 1$

$$\frac{e^{-RT}}{N_0} \le \mathbb{E}\left[\frac{1}{N_T}\right] \le \frac{e^{-RT}}{N_0-1}.$$

This ends the proof of Proposition 2.2.2.

□

Proof of Lemma 2.2.4.

By symmetry of h with respect to $1/2$, we have:

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \left(\alpha + 2R \int_0^1 (\gamma Y_s - Y_s)\, h(\gamma) d\gamma \right) ds + U_t \\ &= Y_0 + \int_0^t \left(\alpha - 2R Y_s \int_0^1 \gamma h(\gamma) d\gamma \right) ds + U_t \\ &= Y_0 + \int_0^t (\alpha - R Y_s)\, ds + U_t. \end{aligned}$$

where U_t is a square-integrable martingale.
Let $\tilde{Y}_t = Y_t e^{Rt}$, $\tilde{Y}_0 = Y_0$. By Itô formula, we get

$$\tilde{Y}_t = \tilde{Y}_0 + \frac{\alpha}{R}\left(e^{Rt} - 1\right) + \int_0^t e^{Rs} dU_s.$$

Replacing $\tilde{Y}_t$ by $Y_t e^{Rt}$, we obtain

$$Y_t = \left(Y_0 - \frac{\alpha}{R}\right) e^{-Rt} + \frac{\alpha}{R} + \int_0^t e^{-R(t-s)} dU_s.$$

We end the proof by taking the expectation and the limit as $t \to +\infty$ of Y_t to obtain (2.16) and (2.17).

□

Proof of Theorem 2.2.5.

We will show that the process Y satisfies ergodicity and integrability assumptions in Bansaye et al. [8] (see (H1) - (H4), Section 4). More precisely:

1. $\mathbb{E}\left[Y_t\right] < +\infty$ for all $t \geq 0$.
2. There exists $\varpi < R$ and $c > 0$ such that $\mathbb{E}\left[Y_t^2\right] < c e^{\varpi t}$ for all $t \geq 0$.

From (2.16) we note that $\mathbb{E}[Y_t] < +\infty$ for all $t \geq 0$. To prove the second point, from (2.14) we have

$$\begin{aligned} \mathbb{E}[Y_t^2] &= \mathbb{E}\left[Y_0^2 + \int_0^t \left(2\alpha Y_s + 2R \int_0^1 \left(\gamma^2 Y_s^2 - Y_s^2\right) h(\gamma) d\gamma \right) ds \right] \\ &= Y_0^2 + 2\alpha \int_0^t \mathbb{E}[Y_s] ds - 2\theta R \int_0^t \mathbb{E}[Y_s^2] ds, \end{aligned} \tag{2.40}$$

with $\theta = \int_0^1 (1-\gamma^2) h(\gamma) d\gamma$ and $0 < \theta < 1$.

Substituting $\mathbb{E}[Y_t] = (Y_0 - \alpha/R)e^{-Rt} + \alpha/R$ into (2.40), we see that $\mathbb{E}(Y_t^2)$ solves the following equation:

$$\frac{d\mathbb{E}[Y_t^2]}{dt} = -2\theta R \mathbb{E}[Y_t^2] + \left(2\alpha Y_0 - \frac{2\alpha^2}{R}\right) e^{-Rt} + \frac{2\alpha^2}{R}. \tag{2.41}$$

The solution of the equation (2.41) is:

$$\mathbb{E}[Y_t^2] = e^{-2\theta Rt}\left[Y_0^2 + \int_0^t e^{2\theta Rs}\left(\Big(2\alpha Y_0 - \frac{2\alpha^2}{R}\Big)e^{-Rs} + \frac{2\alpha^2}{R}\right)ds\right]. \tag{2.42}$$

Hence, if $\theta = \frac{1}{2}$, we have

$$\begin{aligned}\mathbb{E}[Y_t^2] &= Y_0^2 e^{-Rt} + \Big(2\alpha Y_0 - \frac{2\alpha^2}{R}\Big)te^{-Rt} + \frac{2\alpha^2}{R^2}\left(1 - e^{-Rt}\right)\\ &\le Y_0^2 e^{-Rt} + \Big(2\alpha Y_0 - \frac{2\alpha^2}{R}\Big)e^{-(R-\theta)t} + \frac{2\alpha^2}{R^2}\\ &\le \left(Y_0^2 + 2\alpha Y_0 + \frac{2\alpha^2}{R} + \frac{2\alpha^2}{R^2}\right)e^{(0\vee(\theta-R))t} = c_1 e^{\varpi t},\end{aligned}$$

with $\varpi = 0 \vee (\theta - R) := \max(0, \theta - R)$.

If $\theta \neq \frac{1}{2}$,

$$\begin{aligned}\mathbb{E}[Y_t^2] &= e^{-2\theta Rt}\left[Y_0^2 + \Big(2\alpha Y_0 - \frac{2\alpha^2}{R}\Big)\int_0^t e^{(2\theta-1)Rs}ds + \frac{2\alpha^2}{R}\int_0^t e^{2\theta Rs}ds\right]\\ &= Y_0^2 e^{-2\theta Rt} + \Big(2\alpha Y_0 - \frac{2\alpha^2}{R}\Big)\frac{1}{(2\theta-1)R}\left(e^{-Rt} - e^{-2\theta Rt}\right)\\ &\qquad + \frac{\alpha^2}{\theta R^2}\left(1 - e^{-2\theta Rt}\right)\\ &\le \left(Y_0^2 + \Big(2\alpha Y_0 + \frac{2\alpha^2}{R}\Big)\frac{1}{|2\theta-1|R} + \frac{\alpha^2}{\theta R^2}\right) = c_2.\end{aligned}$$

Thus, if we set $c = \max(c_1, c_2)$ then $\mathbb{E}\left[Y_t^2\right] < ce^{\varpi t}$ for all $t \geq 0$.

Now, we apply by Theorem 5.3 of Meyn and Tweedie [82] to show that there exists $\pi \in \mathcal{M}_F(\mathbb{R}_+)$ such that

$$\lim_{t\to+\infty} E[Y_t] = \langle \pi, f\rangle = \frac{\alpha}{R}.$$

The application of Meyn and Tweedie requires that the condition $(\mathcal{S})$ (cf. Meyn and Tweedie [82]) is satisfied, *i.e.* Y is a non-explosive right process, all compact sets are petite for some skeleton chain, and condition (CD2) (cf. Meyn and Tweedie [82]) holds for some compact set C and V bounded on C.

We first verify the condition (CD2). The infinitesimal generator A of Y is defined for $\mathcal{C}^1$ test functions as

$$Af(x) = \alpha f'(x) + 2R\int_0^1 \left(f(\gamma x) - f(x)\right)h(\gamma)d\gamma.$$

For $V(x) = x$ and $f(x) = x + 1$, we have

$$AV(x) = \alpha - Rx \le -\frac{R}{2}f(x) + \left(\alpha + \frac{R}{2}\right)\mathbb{1}_{\left\{x\le \frac{2\alpha}{R}+1\right\}}.$$

Hence, (CD2) is satisfied. To verify the property of petiteness, we have

$$dY_s = \alpha ds + 2R\int_0^1 (\gamma - 1)Y_{s-}Q(d\gamma, ds),$$

where $Q(ds, d\gamma)$ is a Poisson point measure with intensity $h(\gamma)d\gamma ds$. Let us consider the skeleton chain obtained by discretising the time in time steps $\triangle t$. Our purpose is to prove that there exist a compact set C and a measure ν such that the probability transition kernel of the skeleton chain satisfies

$$K(y, B) \geq \nu(B) \quad \forall y \in C, \forall B \text{ measurable set.}$$

We have

$$\begin{aligned} K(y, B) &= \mathbb{P}_y\Big(Y_{\triangle t} \in B\Big) \\ &\geq \mathbb{P}_y\Big(Y_{\triangle t} \in B, \quad 1 \text{ jump in } [0, \triangle t]\Big) \\ &= \mathbb{P}_y\Big(Y_{\triangle t} \in B \mid 1 \text{ jump in } [0, \triangle t]\Big)(2R\triangle t)e^{-2R\triangle t} \\ &= \mathbb{P}_y\Big((y + \alpha\triangle t)\Gamma \in B\Big)(2R\triangle t)e^{-2R\triangle t} \end{aligned}$$

where Γ is a random variable with density $h(\gamma)$. Thus:

$$K(y, B) \geq (2R\triangle t)e^{-2R\triangle t}\int_0^1 \mathbb{1}\Big\{\gamma \in \frac{B}{y + \alpha\triangle t}\Big\}h(\gamma)\ d\gamma.$$

Recall that $h(\gamma) \geq \underline{h} > 0$ for all $\gamma \in (0, 1)$. Since C is a compact set, then there exists $M < +\infty$ such that $\max_{y\in C}(y) \leq M$. Then we obtain:

$$\begin{aligned} K(y, B) \geq &(2R\triangle t)e^{-2R\triangle t}\underline{h}\,\text{Leb}\left(\frac{B}{y + \alpha\triangle t} \cap [0, 1]\right) \\ \geq &(2R\triangle t)e^{-2R\triangle t}\underline{h}\,\text{Leb}\left(\frac{B}{M + \alpha\triangle t} \cap [0, 1]\right) \end{aligned}$$

where Leb$(\cdot)$ denotes a Lebesgue measure. This implies that the property of petiteness is satisfied. Thus the condition $(\mathcal{S})$ is verified.

Finally, applying Theorem 4.2 of [8], we obtain the result

$$\lim_{t\to+\infty} \frac{\langle Z_t, f\rangle}{N_t} = \langle \pi, f\rangle = \frac{\alpha}{R}.$$

□

Proof of Proposition 2.2.9.

To prove (2.20), let us remark that the number of random divisions M_T is independent of $(\Gamma_i^1)_{i\in\mathbb{N}^*}$, because the division rate R is constant and because of the construction of our stochastic process. Therefore, we have

$$\begin{aligned} \mathbb{E}\Big[\hat{h}_\ell | M_T\Big] &= \mathbb{E}\Big[\frac{1}{M_T}\sum_{i=1}^{M_T} K_\ell(\gamma - \Gamma_i^1)\Big| M_T\Big] = \frac{M_T\mathbb{E}[K_\ell(\gamma - \Gamma_1^1)]}{M_T} \\ &= \mathbb{E}\Big[K_\ell(\gamma - \Gamma_1^1)\Big] = K_\ell \star h(\gamma), \end{aligned}$$

and $\mathbb{E}\Big[\hat{h}_\ell\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\hat{h}_\ell|M_T\Big]\Big] = K_\ell \star h(\gamma)$. By similar calculations as (2.20), we obtain (2.21) and (2.22).

To prove *ii)*, by the Strong Law of Large Numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n} K_\ell(\gamma - \Gamma_i^1) \xrightarrow{\text{a.s}} \mathbb{E}\Big[K_\ell(\gamma - \Gamma_1^1)\Big] \quad \text{as } n \to +\infty.$$

From (2.6), we have $\lim_{T\to+\infty} N_T = +\infty$ (a.s). Since $M_T = N_T - N_0$ and N_0 is deterministic, this yields

$$\frac{1}{M_T}\sum_{i=1}^{M_T} K_\ell\Big(\gamma - \Gamma_i^1\Big) \xrightarrow{\text{a.s}} \mathbb{E}\Big[K_\ell(\gamma - \Gamma_1^1)\Big] = K_\ell \star h(\gamma).$$

This ends the proof of Proposition 2.2.9.

□

Proof of Proposition 2.2.10.

We have

$$\mathbb{E}\Big[\|\hat{h}_\ell - h\|_2|M_T\Big] \le \|h - K_\ell \star h\|_2 + \mathbb{E}\Big[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2|M_T\Big].$$

For the variance term, using that $\mathbb{E}\Big[\hat{h}_\ell(\gamma)\Big] = \mathbb{E}\Big[\hat{h}_\ell(\gamma)|M_T\Big]$

$$\begin{aligned}
\mathbb{E}\Big[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T\Big] &= \mathbb{E}\Big[\int_{\mathbb{R}} \Big|\hat{h}_\ell(\gamma) - \mathbb{E}\Big[\hat{h}_\ell(\gamma)\Big]\Big|^2 d\gamma\Big|M_T\Big] \\
&= \int_{\mathbb{R}} \mathbb{V}ar\Big[\hat{h}_\ell(\gamma)\Big|M_T\Big] d\gamma \\
&= \frac{1}{M_T}\int_{\mathbb{R}} \mathbb{V}ar\Big[K_\ell(\gamma - \Gamma_1^1)\Big] d\gamma \\
&\le \frac{1}{M_T}\int_{\mathbb{R}} \mathbb{E}\Big[K_\ell^2(\gamma - \Gamma_1^1)\Big] d\gamma
\end{aligned}$$

By Fubini's theorem, we get

$$\begin{aligned}
\int_{\mathbb{R}} \mathbb{E}\Big[K_\ell^2(\gamma - \Gamma_1^1)\Big] d\gamma &= \int_{\mathbb{R}}\int_{\mathbb{R}} K_\ell^2(\gamma - u)h(u)du\, d\gamma \\
&= \int_{\mathbb{R}} h(u)\Big(\int_{\mathbb{R}} K_\ell^2(\gamma - u)d\gamma\Big) du \\
&= \|K_\ell\|_2^2 \int_{\mathbb{R}} h(u)du = \frac{\|K\|_2^2}{\ell}.
\end{aligned}$$

Then we have

$$\mathbb{E}\Big[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T\Big] \le \frac{\|K\|_2^2}{M_T\ell}. \tag{2.43}$$

Hence, applying Cauchy-Schwarz's inequality, we obtain (2.24). This ends the proof of Proposition 2.2.10.

□

Proof of Theorem 2.2.12.

This proof is inspired by the proof of Doumic et al. [39]. However, our problem here is that the number of observations M_T is random. To overcome this difficulty, we work conditionally to M_T to get concentration inequalities.

Hereafter, we refer $\int f$ to $\int_{\mathbb{R}} f$ and since the support of h is $(0,1)$, we can write $\int h(\gamma)d\gamma$ instead of $\int_0^1 h(\gamma)d\gamma$. Recall that

$$A(\ell) := \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell,\ell'} - \hat{h}_{\ell'}\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}} \right\}_+ .$$

Then, for any $\ell \in H$, we have

$$\|\hat{h} - h\|_2 \le A_1 + A_2 + A_3,$$

where

$$A_1 := \|\hat{h} - \hat{h}_{\hat{\ell},\ell}\|_2 \le A(\ell) + \frac{\chi\|K\|_2}{\sqrt{M_T \hat{\ell}}},$$

$$A_2 := \|\hat{h}_{\hat{\ell},\ell} - \hat{h}_\ell\|_2 \le A(\hat{\ell}) + \frac{\chi\|K\|_2}{\sqrt{M_T \ell}},$$

$$A_3 := \|\hat{h}_\ell - h\|_2.$$

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

$$A_1 + A_2 \le 2A(\ell) + 2\frac{\chi\|K\|_2}{\sqrt{M_T \ell}}, \tag{2.44}$$

and

$$\begin{aligned} A(\ell) &\le \sup_{\ell' \in H} \Big\{ \|\Big(\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\Big) - \Big(\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\Big)\|_2 \\ &\qquad + \|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}} \Big\}_+ \\ &\le \xi_T(\ell) + \sup_{\ell' \in H} \left\{ \|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \right\}, \end{aligned} \tag{2.45}$$

where

$$\xi_T(\ell) = \sup_{\ell' \in H} \left\{ \|\Big(\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\Big) - \Big(\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\Big)\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}} \right\}_+ . \tag{2.46}$$

For the term $\sup_{\ell' \in H} \left\{ \|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \right\}$, we have

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

Hence, we derive

$$\|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 = \|K_{\ell'} \star (K_\ell \star h - h)\|_2 \leq \|K\|_1 \|K_\ell \star h - h\|_2, \tag{2.47}$$

where the right hand side does not depend on ℓ' allowing us to take $\sup_{\ell' \in H}$ in the left hand side.

Thus, (2.44), (2.46) and (2.47) give

$$A_1 + A_2 \leq 2\xi_T(\ell) + 2\|K\|_1 \|K_\ell \star h - h\|_2 + 2\frac{\chi \|K\|_2}{\sqrt{M_T \ell}}.$$

Then,

$$\mathbb{E}\left[(A_1 + A_2)^2\right] \leq 12\mathbb{E}[\xi_T^2(\ell)] + 12\|K\|_1^2 \|K_\ell \star h - h\|_2^2 + 12\frac{\chi^2 \|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right]. \tag{2.48}$$

For the term A_3, we have from (2.43)

$$\begin{aligned}\mathbb{E}\left[A_3^2\right] &= \|\mathbb{E}[\hat{h}_\ell] - h\|_2^2 + \mathbb{E}\left[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2\right] \\ &\leq \|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right].\end{aligned}$$

Finally, replacing χ by $(1+\epsilon)(1+\|K\|_1)$, we have for any $\ell \in H$

$$\begin{aligned}\mathbb{E}\left[\|\hat{h} - h\|_2^2\right] &\leq 2\mathbb{E}\left[(A_1 + A_2)^2\right] + 2\mathbb{E}\left[A_3^2\right] \\ &\leq 24\mathbb{E}\left[\xi_T^2(\ell)\right] + 2\left(1 + 12\|K\|_1^2\right)\|K_\ell \star h - h\|_2^2 \\ &\quad + 2\Big(1 + 12(1+\epsilon)^2(1+\|K\|_1)^2\Big)\frac{\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right] \\ &\leq 24\mathbb{E}\left[\xi_T^2(\ell)\right] + C_1\left(\|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right]\right),\end{aligned} \tag{2.49}$$

with C_1 a constant depending on ϵ and $\|K\|_1$.

It remains to deal with the term $\mathbb{E}\left[\xi_T^2(\ell)\right]$ where $\xi_T(\ell)$ is defined in (2.46),

$$\begin{aligned}\xi_T(\ell) &\leq \sup_{\ell' \in H}\left\{\|\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\|_2 + \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}}\right\}_+ \\ &\leq \sup_{\ell' \in H}\left\{\|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \|K\|_1 + \|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}}\right\}_+ \\ &\leq \sup_{\ell' \in H}\left\{\Big(1 + \|K\|_1\Big)\|\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{(1+\epsilon)(1+\|K\|_1)\|K\|_2}{\sqrt{M_T \ell'}}\right\}_+ \\ &\leq (1 + \|K\|_1) S_T,\end{aligned}$$

where

$$S_T := \sup_{\ell \in H}\left\{\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{M_T \ell}}\right\}_+.$$

Hence,

$$\mathbb{E}\Big[\xi_T^2(\ell)\Big] \leq (1+\|K\|_1)^2 \mathbb{E}\Big[\mathbb{E}\Big[S_T^2|M_T\Big]\Big]. \tag{2.50}$$

If we show that

$$\mathbb{E}\Big[S_T^2|M_T=n\Big] \leq C_* \frac{1}{n}, \tag{2.51}$$

then

$$\mathbb{E}\Big[\xi_T^2(\ell)\Big] \leq C_*(1+\|K\|_1)^2 \mathbb{E}\left[\frac{1}{M_T}\right] \tag{2.52}$$

where C_* is a constant.

Let us establish (2.51). When $M_T = n$, $\forall n \in \mathbb{N}^*$, we set

$$\mathbb{E}\Big[\Sigma_n^2\Big] = \mathbb{E}\Big[S_T^2|M_T=n\Big]$$

where

$$\Sigma_n := \sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+,$$

with

$$Z_\ell = \hat{h}_\ell - \mathbb{E}[\hat{h}_\ell] = \frac{1}{n}\sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma-\Gamma_i^1)].$$

Then,

$$\begin{aligned}
\mathbb{E}\Big[\Sigma_n^2\Big] &= \mathbb{E}\left[\sup_{\ell\in H}\left\{\|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2\right] \\
&\leq \int_0^{+\infty} \mathbb{P}\left[\sup_{\ell\in H}\left\{\|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2 \geq x\right] dx \\
&\leq \sum_{\ell\in H}\int_0^{+\infty} \mathbb{P}\left[\left\{\|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2 \geq x\right] dx.
\end{aligned}$$

We bound this with Talagrand's inequality.

Let $\mathcal{A}$ be a countable dense subset of the unit ball of $\mathbb{L}_2([0,1])$. We express the norm $\|Z_\ell\|_2$ as

$$\begin{aligned}
\|Z_\ell\|_2 &= \sup_{a\in\mathcal{A}} \int a(\gamma) Z_\ell(\gamma) d\gamma \\
&= \sup_{a\in\mathcal{A}} \sum_{i=1}^n \int a(\gamma)\frac{1}{n}\left(K_\ell(\gamma-\Gamma_i^1) - \mathbb{E}[K_\ell(\gamma-\Gamma_i^1)]\right) d\gamma.
\end{aligned}$$

Let

$$V_{i,\Gamma} = \int a(\gamma)\frac{1}{n}\left(K_\ell(\gamma-\Gamma_i^1) - \mathbb{E}[K_\ell(\gamma-\Gamma_i^1)]\right) d\gamma.$$

Then $V_{i,\Gamma}$, $i=1,\dots,n$ is a sequence of i.i.d random variables with zero mean. Thus, we can apply Talagrand's inequality (see [78, p. 170]) to $\|Z_\ell\|_2 = \sup_{a\in\mathcal{A}} \sum_{i=1}^n V_{i,\Gamma}$. For all $\eta, x > 0$, one has

$$\mathbb{P}\left(\|Z_\ell\|_2 \geq (1+\eta)\mathbb{E}[\|Z_\ell\|_2] + \sqrt{2\nu x} + c(\eta) b x\right) \leq e^{-x},$$

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

$$\nu = \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E}\left[\left(\int a(\gamma) \left(K_\ell(\gamma - \Gamma_1^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] \right) d\gamma \right)^2 \right],$$

and,

$$b = \frac{1}{n} \sup_{y \in (0,1), a \in \mathcal{A}} \int a(\gamma) \left(K_\ell(\gamma - y) - \mathbb{E}[K_\ell(\gamma - \Gamma_1^1)] \right) d\gamma.$$

Next, we calculate the terms $\mathbb{E}[\|Z_\ell\|_2]$, ν and b. Applying Cauchy - Schwarz's inequality and using independence of variables, we get

$$\begin{aligned}
\mathbb{E}\Big[\|Z_\ell\|_2\Big] &\leq \left(\mathbb{E}\Big[\|Z_\ell\|_2^2\Big] \right)^{1/2} \\
&\leq \left(\mathbb{E}\left[\int \left(\frac{1}{n} \sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 d\gamma \right] \right)^{1/2} \\
&= \frac{1}{n} \left(\int \mathbb{E}\left[\left(\sum_{i=1}^n K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 \right] d\gamma \right)^{1/2} \\
&= \frac{1}{n} \left(\int \sum_{i=1}^n \mathbb{E}\left[\left(K_\ell(\gamma - \Gamma_i^1) - \mathbb{E}[K_\ell(\gamma - \Gamma_i^1)] \right)^2 \right] d\gamma \right)^{1/2} \\
&\leq \frac{1}{n} \left(n \int \mathbb{E}\left[K_\ell(\gamma - \Gamma_1^1)^2 \right] d\gamma \right)^{1/2} = \frac{\|K\|_2}{\sqrt{n\ell}}.
\end{aligned}$$

For the term ν, we have

$$\begin{aligned}
\nu &\leq \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E}\left[\left(\int a(\gamma) K_\ell(\gamma - \Gamma_1^1) d\gamma \right)^2 \right] \\
&\leq \frac{1}{n} \sup_{a \in \mathcal{A}} \mathbb{E}\left[\int |K_\ell(\gamma - \Gamma_1^1)| d\gamma \times \int a^2(\gamma) |K_\ell(\gamma - \Gamma_1^1)| d\gamma \right] \\
&\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \mathbb{E}\left[\int a^2(\gamma) |K_\ell(\gamma - \Gamma_1^1)| d\gamma \right] \\
&\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \int a^2(\gamma) \mathbb{E}\left[|K_\ell(\gamma - \Gamma_1^1)| \right] d\gamma \\
&\leq \frac{\|K\|_1}{n} \sup_{a \in \mathcal{A}} \int \int a^2(\gamma) |K_\ell(\gamma - u)| h(u) du d\gamma \\
&\leq \frac{\|h\|_\infty \|K\|_1^2}{n}.
\end{aligned}$$

For the term b, we have

$$\begin{aligned}
b &= \frac{1}{n} \sup_{y \in (0,1)} \|K_\ell(\cdot - y) - \mathbb{E}[K_\ell(\cdot - \Gamma_1^1)]\|_2 \\
&\leq \frac{1}{n} \left(\sup_{y \in (0,1)} \|K_\ell(\cdot - y)\|_2 + \left(\mathbb{E}\Big[\int K_\ell^2(\gamma - \Gamma_1^1) d\gamma \Big] \right)^{1/2} \right) \leq \frac{2\|K\|_2}{n\sqrt{\ell}}.
\end{aligned}$$

So, for all $\eta, x > 0$, we have

$$\mathbb{P}\left(\|Z_\ell\|_2 \geq (1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2x}{n}} + 2c(\eta)\frac{\|K\|_2 x}{n\sqrt{\ell}}\right) \leq e^{-x}.$$

Let W_ℓ be some strictly positive weights, we apply the previous inequality to $x = W_\ell + u$ for $u > 0$. We have

$$\mathbb{P}\Bigg(\|Z_\ell\|_2 \geq (1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{W_\ell}{n}} + 2c(\eta)\frac{\|K\|_2 W_\ell}{n\sqrt{\ell}} \\ + \|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{u}{n}} + 2c(\eta)\frac{\|K\|_2 u}{n\sqrt{\ell}}\Bigg) \leq e^{-W_\ell - u}.$$

If we set

$$\Psi_\ell = (1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}} + \|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{W_\ell}{n}} + 2c(\eta)\frac{\|K\|_2 W_\ell}{n\sqrt{\ell}},$$

then,

$$\mathbb{P}\left(\|Z_\ell\|_2 - \Psi_\ell \geq \|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{u}{n}} + 2c(\eta)\frac{\|K\|_2 u}{n\sqrt{\ell}}\right) \leq e^{-W_\ell - u}.$$

Let

$$\Lambda = \mathbb{E}\left[\sup_{\ell \in H}\left(\|Z_\ell\|_2 - \Psi_\ell\right)_+^2\right] = \int_0^{+\infty}\mathbb{P}\left[\sup_{\ell \in H}\left(\|Z_\ell\|_2 - \Psi_\ell\right)_+^2 \geq x\right]dx.$$

An upper bound of Λ is given by

$$\Lambda \leq \sum_{\ell \in H}\int_0^{+\infty}\mathbb{P}\left[\left(\|Z\|_2 - \Psi_\ell\right)_+^2 \geq x\right]dx.$$

Let us take u such that

$$x = f(u)^2 = \left(\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{u}{n}} + 2c(\eta)\frac{\|K\|_2 u}{n\sqrt{\ell}}\right)^2.$$

So,

$$dx = 2f(u)\left(\|h\|_\infty^{1/2}\|K\|_1\frac{1}{2\sqrt{nu}} + 2c(\eta)\frac{\|K\|_2}{n\sqrt{\ell}}\right)du.$$

Hence,

$$\begin{aligned}
\Lambda &\leq \sum_{\ell \in H}\int_0^{+\infty} e^{-W_\ell - u}2f(u)\left(\|h\|_\infty^{1/2}\|K\|_1\frac{1}{2\sqrt{nu}} + 2c(\eta)\frac{\|K\|_2}{n\sqrt{\ell}}\right)du \\
&\leq \sum_{\ell \in H}\int_0^{+\infty} e^{-W_\ell - u}2f(u)\left(\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{u}{n}} + 2c(\eta)\frac{\|K\|_2 u}{n\sqrt{\ell}}\right)u^{-1}du \\
&\leq 2\sum_{\ell \in H} e^{-W_\ell}\int_0^{+\infty} f^2(u)e^{-u}u^{-1}du \\
&\leq C_\eta\sum_{\ell \in H} e^{-W_\ell}\left(\|h\|_\infty\|K\|_1^2\int_0^{+\infty} e^{-u}du + \frac{\|K\|_2^2}{\ell^2}\int_0^{+\infty} ue^{-u}du\right) \times \frac{1}{n} \\
&\leq C_\eta\sum_{\ell \in H} e^{-W_\ell}\left(\|h\|_\infty\|K\|_1^2 + \frac{\|K\|_2^2}{\ell^2}\right) \times \frac{1}{n}. \qquad (2.53)
\end{aligned}$$

We need to choose W_ℓ and η such that

$$\mathbb{E}\Big[\Sigma_n^2\Big] = \mathbb{E}\left[\sup_{\ell\in H}\left\{\|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2\right] \le \Lambda. \tag{2.54}$$

Let $\theta > 0$, we choose

$$W_\ell = \frac{\theta^2\|K\|_2^2}{2\|h\|_\infty\|K\|_1^2\sqrt{\ell}},$$

the we have

$$\Psi_\ell = (1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}} + \frac{\theta\|K\|_2}{\sqrt{2n\sqrt{\ell}}} + \frac{c(\eta)\theta^2\|K\|_2^3}{\|h\|_\infty\|K\|_1^2}\frac{1}{n\ell}.$$

Obviously, the series in (2.53) is finite and for any $\ell \in H$, since $\ell \le 1$, we have

$$\begin{aligned}\Psi_\ell &\le (1+\eta+\theta)\frac{\|K\|_2}{\sqrt{n\ell}} + \frac{c(\eta)\theta^2\|K\|_2^3}{\|h\|_\infty\|K\|_1^2}\frac{1}{n\ell} \\ &\le \left(1+\eta+\theta+\frac{c(\eta)\theta^2\|K\|_2^2}{\|h\|_\infty\|K\|_1^2}\frac{1}{\sqrt{n\ell}}\right)\frac{\|K\|_2}{\sqrt{n\ell}}.\end{aligned}$$

Since $H \subset \left\{\triangle^{-1}, \triangle = 1,\dots,\triangle_{\max}\right\}$, if we choose $\triangle_{\max} = \lfloor \delta n\rfloor$ for some $\delta > 0$, then $\ell_{\min} = \triangle_{\max}^{-1}$ and we obtain

$$\Psi_\ell \le \left(1+\eta+\theta+\frac{c(\eta)\theta^2\|K\|_2^2\sqrt{\delta}}{\|h\|_\infty\|K\|_1^2}\right)\frac{\|K\|_2}{\sqrt{n\ell}}.$$

It remains to choose $\eta = \epsilon/2$ and θ small enough such that

$$\theta + \frac{c(\eta)\theta^2\|K\|_2^2\sqrt{\delta}}{\|h\|_\infty\|K\|_1^2} = \frac{\epsilon}{2},$$

then

$$\Psi_\ell \le (1+\epsilon)\frac{\|K\|_2}{\sqrt{n\ell}},$$

and we get

$$\mathbb{E}\left[\Sigma_n^2\right] \le C_* \times \frac{1}{n},$$

where C_* is a constant depending on δ,ϵ,$\|h\|_\infty$,$\|K\|_1$ and $\|K\|_2$. Hence, we get (2.52).

Combining (2.49) and (2.52), we obtain

$$\mathbb{E}\left[\|\hat{h}-h\|_2^2\right] \le C_1\left(\|K_\ell\star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right]\right) + C_*\mathbb{E}\left[\frac{1}{M_T}\right]$$

Moreover, since $N_T > N_0$, we have

$$\begin{aligned}\mathbb{E}\left[\frac{1}{M_T}\right] = \mathbb{E}\left[\frac{1}{N_T-N_0}\right] &= \mathbb{E}\left[\frac{N_T}{N_T-N_0}\frac{1}{N_T}\right] = \mathbb{E}\left[\frac{1}{1-\frac{N_0}{N_T}}\frac{1}{N_T}\right] \\ &\le \mathbb{E}\left[\frac{1}{1-\frac{N_0}{N_0+1}}\frac{1}{N_T}\right] \\ &\le (N_0+1)\mathbb{E}\left[\frac{1}{N_T}\right].\end{aligned} \tag{2.55}$$

Then, using (2.10), (2.12) and (2.55), recall the definition of $\varrho(T)^{-1}$ in (2.29), we obtain for any $\ell \in H$

$$\mathbb{E}\left[\|\hat{h} - h\|_2^2\right] \leq C_1 \left(\|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right) + C_2 \varrho(T)^{-1}.$$

This ends the proof of Theorem 2.2.12. □

Proof of Theorem 2.2.15.

We begin with the bias term $\|K_\ell \star h - h\|_2$ in the right hand side of the oracle inequality (2.30). For any $\ell \in H$ and $\gamma \in (0,1)$, let $k = \lfloor \beta \rfloor$ and $b(\gamma) = K_\ell \star h(\gamma) - h(\gamma)$, then we have

$$h(\gamma + u\ell) = h(\gamma) + h'(\gamma)u\ell + \cdots + \frac{(u\ell)^k}{(k-1)!} \int_0^1 (1-\theta)^{k-1} h^{(k)}(\gamma + \theta u\ell) d\theta.$$

Since K is a kernel of order β^* and $\beta \in (0, \beta^*)$, we get

$$b(\gamma) = \int K(u) \frac{(u\ell)^k}{(k-1)!} \left[\int_0^1 (1-\theta)^{k-1} \left(h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma) \right) d\theta \right] du.$$

Setting $E_{k,\ell}(u) = |K(u)| \frac{|u\ell|^k}{(k-1)!}$ for the sake of notation. Since $h \in \mathcal{H}(\beta, L)$ and applying twice the generalized Minskowki's inequality, we obtain

$$\begin{aligned}
&\|h - \mathbb{E}[\hat{h}]\|_2^2 = \int b^2(\gamma) d\gamma \\
&\leq \int \left(\int E_{k,\ell}(u) \left[\int_0^1 (1-\theta)^{k-1} \left| h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma) \right| d\theta \right] du \right)^2 d\gamma \\
&\leq \left(\int E_{k,\ell}(u) \left[\int \left(\int_0^1 (1-\theta)^{k-1} \left| h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma) \right| d\theta \right)^2 d\gamma \right]^{1/2} du \right)^2 \\
&\leq \left(\int E_{k,\ell}(u) \left[\int_0^1 (1-\theta)^{k-1} \left(\int \left| h^{(k)}(\gamma + \theta u\ell) - h^{(k)}(\gamma) \right|^2 d\gamma \right)^{1/2} d\theta \right] du \right)^2 \\
&\leq \left(\int E_{k,\ell}(u) \left[\int_0^1 (1-\theta)^{k-1} L(\theta u\ell)^{\beta-k} d\theta \right] du \right)^2 \\
&\leq \left(\int |K(u)| \frac{|u\ell|^k}{(k-1)!} \left[\int_0^1 (1-\theta)^{k-1} L(u\ell)^{\beta-k} d\theta \right] du \right)^2 \\
&\leq C_{K,L,\beta} \ell^{2\beta},
\end{aligned}$$

where $C_{K,L,\beta} = \left(\frac{L}{k!} \int |u|^\beta |K(u)| du \right)^2$.

Finally, we have

$$\mathbb{E}\left[\|\hat{h} - h\|_2^2\right] \leq C_1 \inf_{\ell \in H} \left\{ C_{K,L,\beta} \ell^{2\beta} + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right\} + C_2 \varrho(T)^{-1}. \tag{2.56}$$

Taking the derivative of the expression inside the $\inf_{\ell \in H}$ of (2.56) with respect to ℓ, we obtain the minimizer

$$\ell^* = \left(\frac{\|K\|_2^2}{2\beta C_{K,L,\beta}} \right)^{\frac{1}{2\beta+1}} \varrho(T)^{-\frac{1}{2\beta+1}}.$$

Since the optimal bandwidth $\hat{\ell}$ is proportional to ℓ^* up to a multiplicative constant. Therefore, by substituting ℓ by $\hat{\ell}$ in the right hand side of (2.56), we obtain

$$\mathbb{E}\left[\|\hat{h} - h\|_2^2 \right] \leq C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}},$$

with C_3 a constant depending on N_0, δ, ϵ, $\|K\|_1$, $\|K\|_2$, $\|h\|_\infty$, β and L. This ends the proof of Theorem 2.2.15. □

Proof of Theorem 2.2.16.

For $T > 0$, let us denote by $\hat{h}_T$ the estimator of h. To prove the Theorem 2.2.16, we apply the general reduction scheme proposed by Tsybakov [108] (Section 2.2, p.79). We will show the existence of a family $\mathcal{H}_{m,T} = \left\{ h_{j,T} : j = 0, 1, \ldots, m \right\}$ such that:

1) $h_{j,T} \in \mathcal{H}(\beta, L)$, $j = 0, \ldots, m$.

2) $\|h_{j,T} - h_{k,T}\|_2 \geq 2c\, e^{-\frac{\beta}{2\beta+1} RT}$, $0 \leq j < k \leq m$.

3) $\dfrac{1}{m} \sum_{j=1}^m K(P_j, P_0) \leq \vartheta \log(m)$ for $0 < \vartheta < 1/8$. P_j and P_0 are the distribution of observations when the division kernels are $h_{j,T}$ and h_0, respectively. $K(P,Q)$ denotes the Kullback-Leibler divergence between two measures P and Q:

$$K(P,Q) = \begin{cases} \int \log \frac{dP}{dQ} dP, & \text{if } P \ll Q \\ +\infty, & \text{otherwise.} \end{cases}$$

Under the preceding conditions 1, 2, 3, Tsybakov [108] (Theorem 2.5, p.99) show that

$$\inf_{\hat{h}_T} \max_{h \in \mathcal{H}_{m,T}} \mathbb{P}\left(\|\hat{h}_T - h\|_2^2 \geq c^2 e^{-\frac{2\beta}{2\beta+1} RT} \right) \geq C', \tag{2.57}$$

where the infimum is taken over all estimators $\hat{h}_T$ and positive constant C' is independent of T. This will be sufficient to obtain Theorem 2.2.16 by [108, Theorem 2.7]. The proof ends with proposing a family $\mathcal{H}_{m,T}$ and checking the assumptions 1, 2, 3.

Construction of the family $\mathcal{H}_{m,T}$:

The idea is construct a family of perturbations around h_0 which is a symmetric density with respect to $\frac{1}{2}$ and belongs to $\mathcal{H}(\frac{L}{2}, \beta)$. For the simplification, we choose $h_0(\gamma) = \mathbb{1}_{(0,1)}(\gamma)$.

Let $c_0 > 0$ be a real number, and let $\gamma \in (0,1)$, $f(\gamma) = LD^{-\beta} g(D\gamma)$ where g is a regular function having support $(0,1)$ and $\int g(\gamma) d\gamma = 0$, $g \in \mathcal{H}(\frac{1}{2}, \beta)$, we define

$$D = \lceil c_0 e^{\frac{RT}{2\beta+1}} \rceil \quad \text{and} \quad f_k(\gamma) = f\left(\gamma - \frac{(k-1)}{D} \right),$$

By definition, the functions f_k's have disjoint support and one can check that the functions $f_k \in \mathcal{H}(\frac{L}{2}, \beta)$.

Then, the function $h_{j,T}$ will be chosen in

$$\mathcal{D} = \left\{ h_\delta(\gamma) = h_0(\gamma) + c_1 \sum_{k=1}^{D} \delta_k f_k(\gamma) : \delta = (\delta_1, \ldots, \delta_D) \in \{0,1\}^D \right\},$$

where

$$c_1 = \min\left(\frac{1}{LD^{-\beta}\|g\|_\infty}, 1 \right). \tag{2.58}$$

We now check that h_δ is a density, since $\int h_\delta(\gamma)d\gamma = \int h_0(\gamma)d\gamma = 1$, it remains to verify that $h_\delta(\gamma) \geq 0 \, \forall \, \gamma$. We have

$$\begin{aligned}
\inf_{(0,1)} h_\delta(\gamma) &\geq \inf_{(0,1)} h_0 - \|c_1 \sum_{k=1}^{D} \delta_k f_k\|_\infty \\
&\geq 1 - c_1 L D^{-\beta} \max_k \sup_\gamma |\delta_k| g\Big(D\gamma - (k-1)\Big) \\
&\geq 1 - c_1 L D^{-\beta} \|g\|_\infty \geq 0,
\end{aligned}$$

by the choice of c_1. Thus the family of densities $\mathcal{D}$ is well-defined.

1) The condition $h_{j,T} \in \mathcal{H}(\beta, L)$:

Let us denote $q = \lfloor \beta \rfloor$, then for all $\gamma, \gamma' \in (0,1)$ we have

$$\begin{aligned}
\left| h_\delta^{(q)}(\gamma) - h_\delta^{(q)}(\gamma') \right| &= \left| h_0^{(q)}(\gamma) - h_0^{(q)}(\gamma') + c_1 \sum_{k=1}^{D} \delta_k \Big(f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma') \Big) \right| \\
&\leq c_1 \sum_{k=1}^{D} |\delta_k| \left| f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma') \right| \\
&\leq c_1 \max_k \left| f_k^{(q)}(\gamma) - f_k^{(q)}(\gamma') \right| \\
&\leq c_1 L D^{-\beta} \max_k D^q \left| g^{(q)}(D\gamma - (k-1)) - g^{(q)}(D\gamma' - (k-1)) \right| \\
&\leq c_1 L D^{\lfloor \beta \rfloor - \beta} D^{\beta - \lfloor \beta \rfloor} |\gamma - \gamma'|^{\beta - \lfloor \beta \rfloor} \leq L |\gamma - \gamma'|^{\beta - \lfloor \beta \rfloor},
\end{aligned}$$

which is always satisfied with $c_1 = \min\left(\frac{1}{LD^{-\beta}\|g\|_\infty}, 1 \right)$, thus $h_\delta \in \mathcal{H}(L, \beta)$.

2) The condition $\|h_{j,T} - h_{k,T}\|_2 \geq 2c\, e^{-\frac{\beta}{2\beta+1} RT}$:

For all $\delta, \delta' \in \{0,1\}^D$, we have

$$\begin{aligned}\|h_\delta - h_{\delta'}\|_2 &= \left[\int_0^1 \Big(h_\delta(\gamma) - h_{\delta'}(\gamma)\Big)^2 d\gamma\right]^{1/2} = \left[\int_0^1 \left(c_1 \sum_{k=1}^D (\delta_k - \delta'_k) f_k(\gamma)\right)^2 d\gamma\right]^{1/2} \\ &= c_1 \left[\int_0^1 \sum_{k=1}^D (\delta_k - \delta'_k)^2 f_k^2(\gamma) d\gamma\right]^{1/2} = c_1 \left[\sum_{k=1}^D (\delta_k - \delta'_k)^2 \int_{\frac{k-1}{D}}^{\frac{k}{D}} f_k^2(\gamma) d\gamma\right]^{1/2} \\ &= c_1 \left[\sum_{k=1}^D (\delta_k - \delta'_k)^2 \int_{\frac{k-1}{D}}^{\frac{k}{D}} L^2 D^{-2\beta} g^2\left(D\gamma - (k-1)\right) d\gamma\right]^{1/2} \\ &= c_1 L D^{-\beta-1/2} \|g\|_2 \left[\sum_{k=1}^D (\delta_k - \delta'_k)^2\right]^{1/2} = c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{d_H(\delta, \delta')},\end{aligned}$$

where $d_H(\delta, \delta') = \sum_{k=1}^D \mathbb{1}\{\delta_k \neq \delta'_k\}$ is the Hamming distance between δ and δ'.

According to the Lemma of Varshamov-Gilbert (cf. Tsybakov [108], p.104), there exist a subset $\left\{\delta^{(0)}, \dots, \delta^{(m)}\right\}$ of $\{0,1\}^D$ with cardinal (2.59) such that $\delta^{(0)} = (0, \dots, 0)$,

$$m \geq 2^{D/8}, \tag{2.59}$$

and

$$d_H(\delta^{(j)}, \delta^{(k)}) \geq \frac{D}{8}, \ \forall\, 0 \leq j < k \leq m. \tag{2.60}$$

Then, by setting $h_{j,T}(x) = h_{\delta^{(j)}}(x)$, $j = 0, \dots, m$, we obtain

$$\begin{aligned}\|h_{j,T} - h_{k,T}\|_2 &= c_1 L D^{-\beta-\frac{1}{2}} \|g\|_2 \sqrt{d_H(\delta^{(j)}, \delta^{(k)})} \\ &\geq c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{\frac{D}{8}} \\ &\geq \frac{c_1 L}{4} \|g\|_2 D^{-\beta}\end{aligned}$$

whenever $D \geq 8$.

Suppose that $N_T \geq N_{T^*}$ where $T^* = \log\left(\frac{7}{c_0}\right) \frac{2\beta+1}{R}$. Then, $D \geq 8$ and $D^\beta \leq (2c_0)^\beta e^{\frac{\beta}{2\beta+1} RT}$. This implies:

$$\|h_{j,T} - h_{k,T}\|_2 \geq \frac{c_1 L}{4} \|g\|_2 (2c_0)^{-\beta} e^{-\frac{\beta}{2\beta+1} RT},$$

But,

$$\min\left(\frac{1}{L\|g\|_\infty}, 1\right) \leq c_1 \leq 1$$

Hence, we obtain

$$\|h_{j,T} - h_{k,T}\|_2 \geq 2c\, e^{-\frac{\beta}{2\beta+1} RT},$$

where

$$c = \frac{\min(1, L\|g\|_2)}{8} (2c_0)^{-\beta}.$$

3) The condition $\frac{1}{m}\sum_{j=1}^{m} K(P_j, P_0) \leq \vartheta \log(m)$ **for** $0 < \vartheta < 1/8$:

We need to show that for all $\delta \in \{0,1\}^D$,

$$K(P_\delta, P_0) \leq \vartheta \log(m),$$

where

$$K(P_\delta, P_0) = \mathbb{E}\left[\log \frac{dP_\delta}{dP_0}|_{\mathcal{F}_T}(Z)\right],$$

and where $(Z_t)_{t\in[0,T]}$ is defined in (2.5) with the random measure Q having intensity $q(ds, di, d\gamma) = Rh_\delta(\gamma)ds\, n(di)d\gamma$.

Here, the difficulty comes from the fact that N_T is variable because the observations result from a stochastic process Z_t. The law of these observations is not a probability distribution on a fixed $\mathbb{R}^n$ where n would be the sample size, but rather a probability distribution on a path space. P_δ is the probability distribution when the Poisson point measure Q has intensity $Rh_\delta(\gamma)ds\, n(di)d\gamma$. Thus a natural tool is to use Girsanov's theorem (see [65], Theorem 3.24, p. 159) saying that P_δ is absolutely continuous with respect to P_0 on $\mathcal{F}_T$ with

$$\frac{dP_\delta}{dP_0}|_{\mathcal{F}_T} = \mathfrak{D}_T^\delta,$$

where $(\mathfrak{D}_t^\delta)_{t\in[0,T]}$ is the unique solution of the following SDE (see Proposition 4.17 of [105] for a similar SDE):

$$\mathfrak{D}_T^\delta = 1 + \int_0^T \int_{\mathcal{E}} \mathfrak{D}_{s-}^\delta \mathbb{1}_{\{i\leq N_{s-}\}}\left(\frac{h_\delta(\gamma)}{h_0(\gamma)} - 1\right) Q(ds, di, d\gamma). \tag{2.61}$$

Apply Itô formula for jump processes to (2.61), we get

$$\begin{aligned}\log \mathfrak{D}_T^\delta &= \int_0^T \int_{\mathcal{E}} \mathbb{1}_{\{i\leq N_{s-}\}}\left[\log\left(\mathfrak{D}_{s-}^\delta - \left(\frac{h_\delta(\gamma)}{h_0(\gamma)} - 1\right)\mathfrak{D}_{s-}^\delta\right) - \log \mathfrak{D}_{s-}^\delta\right] Q(ds, di, d\gamma)\\ &= \int_0^T \int_{\mathcal{E}} \mathbb{1}_{\{i\leq N_{s-}\}} \log \frac{h_\delta(\gamma)}{h_0(\gamma)} Q(ds, di, d\gamma) = \sum_{i=1}^{N_T} \log \frac{h_\delta(\Gamma_i^1)}{h_0(\Gamma_i^1)}\end{aligned}$$

by definition of $(\Gamma_1^1, \ldots, \Gamma_{N_T}^1)$.
Then,

$$\begin{aligned}K(P_\delta, P_0) &= \mathbb{E}_\delta\left[\log \mathfrak{D}_T^\delta\right] = \mathbb{E}_\delta\left[\sum_{i=1}^{N_T} \log \frac{h_\delta(\Gamma_i^1)}{h_0(\Gamma_i^1)}\right]\\ &= \mathbb{E}\left[N_T\right] \mathbb{E}_\delta\left[\log \frac{h_\delta(\Gamma_1^1)}{h_0(\Gamma_1^1)}\right] = \mathbb{E}\left[N_T\right] \int_0^1 h_\delta(\gamma) \log \frac{h_\delta(\gamma)}{h_0(\gamma)} d\gamma.\end{aligned}$$

Here, $\mathbb{E}\left[N_T\right]$ does not depend on h_δ and we have $\mathbb{E}[N_T] = N_0 e^{RT}$. Thus, recall the definition of $h_\delta(\cdot)$ and note that $\log(1+x) \leq x$ for $x > -1$, we get

$$\begin{aligned}
K(P_\delta, P_0) &= N_0 e^{RT} \int_0^1 h_\delta(\gamma) \log(h_\delta(\gamma)) d\gamma \\
&= N_0 e^{RT} \int_0^1 \Big(1 + c_1 \sum_{k=1}^{D} \delta_k f_k(\gamma)\Big) \log\Big(1 + c_1 \sum_{k=1}^{D} \delta_k f_k(\gamma)\Big) d\gamma \\
&= N_0 e^{RT} \sum_{k=1}^{D} \int_{\frac{k-1}{D}}^{\frac{k}{D}} \Big(1 + c_1 \delta_k f_k(\gamma)\Big) \log\Big(1 + c_1 \delta_k f_k(\gamma)\Big) d\gamma \\
&= N_0 e^{RT} \sum_{k=1}^{D} \delta_k \int_0^{1/D} \Big(1 + c_1 f(\gamma)\Big) \log\Big(1 + c_1 f(\gamma)\Big) d\gamma \\
&\leq N_0 e^{RT} D \int_0^{1/D} \Big(1 + c_1 f(\gamma)\Big) c_1 f(\gamma) d\gamma \\
&\leq N_0 e^{RT} \Big[c_1 L D^{-\beta} \int_0^{1/D} g(D\gamma) D d\gamma + c_1^2 L^2 D^{-2\beta} \int_0^{1/D} g^2(D\gamma) D d\gamma \Big] \\
&\leq N_0 e^{RT} c_1^2 L^2 D^{-2\beta} \int_0^1 g^2(\gamma) d\gamma \\
&\leq N_0 c_1^2 L^2 \|g\|_2^2 e^{RT} c_0^{-2\beta} e^{-\frac{2\beta}{2\beta+1} RT} \\
&\leq N_0 L^2 \|g\|_2^2 c_0^{-2\beta-1} D \qquad \text{since } c_1 \leq 1.
\end{aligned}$$

From (2.59), we have $m \geq 2^{D/8}$ then

$$D \leq \frac{8 \log(m)}{\log(2)}.$$

Hence, if we set

$$c_0 = \left(\frac{8 N_0 L^2 \|g\|_2^2}{\vartheta \log(2)} \right)^{1/(2\beta+1)},$$

we obtain $K(P_\delta, P_0) \leq \vartheta \log(m)$. This ends the proof of Theorem 2.2.16.

□

2.5 Perspective

In this chapter, we construct an adaptive estimator for the division kernel h of a size-structured population model where the cell divisions occur under the assumption that both the division rate and the growth rate are constants. It would be interesting to extend the problem of estimating the division kernel to the case where the division kernel is a function of size of cells $B(x)$. The main difficulty arises in this case coming from the fact that the size of the population depends on the size distribution which itself depends on the division kernel h. In Hoffmann and Olivier [59], the authors also encounter this difficulty when they consider the problem of estimating the division rate $B(x)$. For this following work, we will need to study a stochastic model to describe the evolution of cells in the non-constant case of division rate, then propose an adaptive estimation procedure to construct an estimator of h by using a kernel-based estimator, if possible. Lastly, we expect to obtain an oracle type inequality and optimal rates of convergence for theoretical results.

3

Estimating of the division kernel of the size-structured population by stationary size distribution

Contents

This chapter is a work in progress and is the fruit of a collaboration with Viet Chi Tran (Université de Lille 1), Thanh Mai Pham Ngoc (Université Paris Sud) and Vincent Rivoirard (Université Paris Dauphine).

3.1 Introduction

In this chapter, we consider the size-structured population describing the binary cell divisions we studied in **Chapter 2** but in a general case where the division rate is a function $B(x)$ and the toxicity $x \in \mathbb{R}_+$ grows inside a cell with rate $\alpha g(x)$ where $g(x)$ is a continuous positive function and α is a positive constant. Recall that when a cell divides, a random fraction Γ of the toxicity goes in the first daughter cell and a fraction $(1 - \Gamma)$ in the second one. Here, Γ is assumed to be a random variable having a symmetric distribution $H(d\gamma)$ on $[0,1]$ such that $\mathbb{P}(\Gamma=0)=\mathbb{P}(\Gamma=1)=0$. Indeed, in Section 3.3 we shall assume that $H(d\gamma)$ has a density $h(\gamma)$ in the problem of estimating the division kernel. Furthermore, the division rate B and the function g are assumed to satisfy the following assumptions:

Assumption 3.1.1.

- *The division rate $B(x)$ is continuous and bounded by a positive constant $\bar{B}$.*
- *We have $g \in \mathbb{L}^\infty(\mathbb{R}_+)$.*

We assume that the toxicity $(X_t)_{t\geq 0}$ satisfies

$$dX_t = \alpha g(X_t)dt. \tag{3.1}$$

Then we describe the population of cells at time $t \geq 0$ by the following random point measure:

$$Z_t(dx) = \sum_{i=1}^{N_t} \delta_{X_t^i}(dx),$$

where N_t is the number of cells living at time t.

We are interested in studying a renormalization $(Z^n)_{n\in\mathbb{N}^*}$ of the microscopic process Z. We prove in Section 3.2.2 that the renormalized microscopic process converges in a large population limit (see for instance, Bansaye and Tran [11], Fournier and Méléard [50], Tran [106] for such studies on the large population limit) to a unique solution of the following growth-fragmentation equation:

$$\partial_t n(t,x)+\alpha\partial_x\Big(g(x)n(t,x)\Big)+B(x)n(t,x) = 2\int_0^{+\infty} n(t,y)B(y)h\left(\frac{x}{y}\right)\frac{dx}{y},\ t\geq 0, x\geq 0, \tag{3.2}$$

where $n(t,x)$ is the density of the cells structured by the toxicity x at time t and $H(d\gamma)$ is assumed to be having a density $h(\gamma)$. This growth-fragmentation equation is widely used for the problem of estimating the division rate $B(x)$ in both analysis and statistics. In Bourgeron et al. [16] and Doumic and Tine [41], inverse problem methods are used to estimate B. For the statistical approach, Doumic et al. [39] construct an estimator of B using a kernel method in the case of equal mitosis where $H(d\gamma)=\delta_{1/2}(d\gamma)$, *i.e.* the daughters of a cell x have size $x/2$. Here, we are interested in the estimation of the division kernel $h(\cdot)$ of the size-structured population introduced in **Chapter 2** but for the case where we do not observe completely the whole division tree, *i.e.* we only have the observations of the population at a fixed time T. In this context, a kernel estimator as in **Chapter 2** can not be used anymore since the observations $\Gamma_1^1,\ldots,\Gamma_{M_T}^1$ are not available. Thus we construct an estimate of h starting from the growth-fragmentation equation (3.2). Study of the asymptotic behavior of the solution $n(t,x)$ (see [37, 84, 92]) shows that $n(t,x)\approx N(x)e^{\lambda x}$ where (N,λ) is the unique solution of the following integro-differential equation:

$$\alpha\partial_x\Big(g(x)N(x)\Big)+(\lambda+B(x))N(x) = 2\int_0^{+\infty} N(y)B(y)h\Big(\frac{x}{y}\Big)\frac{dy}{y}, x\geq 0. \tag{3.3}$$

One can observe that there is a multiplicative convolution in the right hand side (r.h.s) of Equation (3.3). Thus we resort to Fourier techniques for a deconvolution procedure, then estimate h based on n observations having distribution $N(x)dx$.

This chapter is organized as follows: in Section 3.2, we describe the renormalized stochastic process $(Z^n)_{n\in\mathbb{N}^*}$ with its moment and martingale properties. Then we prove the convergence of $(Z^n)_{n\in\mathbb{N}^*}$ in the large population limit. Statistical estimation of the division rate $h(\cdot)$ and the consistency of the estimator are presented in Section 3.3.

Notation. We introduce some notations used in the sequel.

For two spaces metric E and F, we note that $\mathcal{C}_b(E,F)$ (resp. $\mathbb{D}(E,F)$, $C_b^{1,1}$, $\mathcal{B}_b(E,F)$) is the set of continuous bounded functions from E to F embedded with the uniform convergences norm (resp. of càdlàg functions from E to F embedded with the Skorohod distance, of bounded functions of class $\mathcal{C}^1$ in t and x with bounded derivatives, of bounded measurable functions).

For a measurable space $(E,\mathcal{E})$, we note $\mathcal{M}_F(E)$ the set of finite measure on E. For a measure $m\in\mathcal{M}_F(E)$ and $f\in\mathcal{B}_b(E)$, we note $\langle m,f\rangle=\int_E f dm$. The set of finite measure on $\mathbb{R}_+$ is denoted by $\mathcal{M}_F(\mathbb{R}_+)$. By default, we will consider the weak convergence topology. When necessary, we will write $\Big(\mathcal{M}_F(\mathbb{R}_+),w\Big)$ (resp. $\Big(\mathcal{M}_F(\mathbb{R}_+),v\Big)$) to precise the weak convergence topology (resp. vague convergence topology).

In this chapter, we denote by f^* the Fourier transform of an integrable function f defined by

$$f^*(\xi)=\int_{-\infty}^{+\infty} f(x)e^{\mathbf{i}x\xi}dx.$$

3.2 Renormalization of the microscopic process and large population limit

3.2.1 Renormalized microscopic processes

We first define the renormalized microscopic process, then introduce a SDE driven by a Poisson point measure which describes the evolution of cells.

Definition 3.2.1. For $n\in\mathbb{N}^*$, we define the renormalized process belonging to $\mathcal{M}_F(\mathbb{R}_+)$ by

$$Z_t^n(dx)=\frac{1}{n}\sum_{i=1}^{N_t^n}\delta_{X_t^i}(dx), \tag{3.4}$$

where $N_t^n=\langle nZ_t^n,1\rangle$ is the number of cells alive at time t. The parameter n is related to the large population limit which corresponds to $n\to+\infty$.

Definition 3.2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space:

1. Let $(Z_0^n)_{n\in\mathbb{N}^*}$ be a sequence of independent random variable such that

$$\forall n \in \mathbb{N}^*, \quad Z_0^n \in \mathcal{M}_F(\mathbb{R}_+) \quad \text{and} \quad \sup_{n\in\mathbb{N}^*} \mathbb{E}(\langle Z_0^n, 1\rangle^2) < +\infty. \tag{3.5}$$

2. Let $Q(ds, d\nu, di, d\gamma)$ be a Poisson point measure on $\mathbb{R}_+ \times \mathcal{E} := \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{N}^* \times [0,1]$ with intensity $q(ds, d\nu, di, d\gamma) = \bar{B}\, ds\, d\nu\, n(di)\, H(d\gamma)$ where $n(di)$ is the counting measure on $\mathbb{N}^*$ and $ds, d\nu$ are Lebesgue measures on $\mathbb{R}_+$. We denote $\{\mathcal{F}_t\}_{t\geq 0}$ the canonical filtration associated with the Poisson point measure.

Then, for every test function $f_t(x) = f(x,t) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ the cell population is described by a SDE for the renormalized process $Z^n \in \mathcal{M}_F(\mathbb{R}_+)$:

$$\begin{aligned}\langle Z_t^n, f\rangle &= \langle Z_0^n, f\rangle + \int_0^t \int_{\mathbb{R}_+} \big(\partial_s f_s(x) + \alpha g(x)\partial_x f_s(x)\big) Z_s^n(dx)ds \\ &+ \frac{1}{n}\int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i\leq N_{s-}^n\}} \mathbb{1}_{\left\{\nu \leq \frac{B(X_{s-}^i)}{\bar{B}}\right\}} \Big[f_s\big(\gamma X_{s-}^i\big) + f_s\big((1-\gamma)X_{s-}^i\big) - f_s\big(X_{s-}^i\big)\Big] Q(ds, d\nu, di, d\gamma).\end{aligned} \tag{3.6}$$

Remark 3.2.3. In fact, the X_t^i's depend also on n, since the division occurs differently when n varies. However, we omit the n for sake of notation.

In the sequel, we show some moment and martingale properties that will be useful in the proof of the convergence in large population limit.

Definition 3.2.4. Let $N > 0$ and $n \in \mathbb{N}^*$, we define the stopping-times as follows

$$\tau_N^n = \inf\left\{t \geq 0, \langle Z_t^n, 1\rangle \geq N\right\}. \tag{3.7}$$

Proposition 3.2.5. *Consider the sequence $(Z^n)_{n\in\mathbb{N}^*}$, if there exists $q \geq 1$ such that*

$$\sup_{n\in\mathbb{N}^*} \left(\mathbb{E}\langle Z_0^n, 1\rangle^q\right) < +\infty, \tag{3.8}$$

then we have, for all $T > 0$

$$\sup_{n\in\mathbb{N}^*} \mathbb{E}\left(\sup_{t\in[0,T]} \langle Z_t^n, 1\rangle^q\right) < +\infty. \tag{3.9}$$

Proof. We use ideas from [50, 105] to prove Proposition 3.2.5. Let $N > 0, n \in \mathbb{N}^*$ and τ_N^n is the stopping-times as in Definition 3.2.4. From equation (3.6), put $f \equiv 1$ we have

$$\langle Z_{t\wedge\tau_N^n}^n, 1\rangle = \langle Z_0^n, 1\rangle + \int_0^{t\wedge\tau_N^n} \int_{\mathcal{E}} \mathbb{1}_{\{i\leq N_{s-}^n\}} \mathbb{1}_{\left\{\nu \leq \frac{B(X_{s-}^i)}{\bar{B}}\right\}} Q(ds, d\nu, di, d\gamma).$$

Applying Ito's formula for jump process with finite variation (see [63], Theorem 5.1 on p.67), we get

$$\langle Z^n_{t\wedge\tau^n_N},1\rangle^q = \langle Z^n_0,1\rangle^q + \int_0^{t\wedge\tau^n_N}\int_{\mathcal{E}} \mathbb{1}_{\{i\le N^n_{s-}\}}\mathbb{1}_{\left\{\nu\le\frac{B(X^i_{s-})}{\bar{B}}\right\}}\left[\left(\langle Z^n_{s-},1\rangle+\frac{1}{n}\right)^q-\langle Z^n_{s-},1\rangle^q\right]Q(ds,d\nu,di,d\gamma).$$

Since we have increasing functions of t

$$\sup_{s\in[0,t\wedge\tau^n_N]}\langle Z^n_s,1\rangle^q = \langle Z^n_{t\wedge\tau^n_N},1\rangle^q.$$

Hence,

$$\begin{aligned}\mathbb{E}\left(\sup_{s\in[0,t\wedge\tau^n_N]}\langle Z^n_s,1\rangle^q\right) &= \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\\ &+\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\int_{\mathbb{R}_+}\left[\left(\langle Z^n_s,1\rangle+\frac{1}{n}\right)^q-\langle Z^n_s,1\rangle^q\right]nB(x)Z^n_s(dx)ds\right)\\ &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\bar{B}\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\int_{\mathbb{R}_+}\frac{1}{n^{q-1}}\left[\left(n\langle Z^n_s,1\rangle+1\right)^q-\left(n\langle Z^n_s,1\rangle\right)^q\right]Z^n_s(dx)ds\right)\end{aligned}$$

Since $(1+y)^q-y^q\le C(q)(1+y^{q-1})$, we get by Fubini's theorem and the choice of τ^n_N

$$\begin{aligned}\mathbb{E}\left(\sup_{s\in[0,t\wedge\tau^n_N]}\langle Z^n_s,1\rangle^q\right) &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)\\ &\quad+\bar{B}C(q)\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\int_{\mathbb{R}_+}\frac{1}{n^{q-1}}\left[1+\left(n\langle Z^n_s,1\rangle\right)^{q-1}\right]Z^n_s(dx)ds\right)\\ &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\bar{B}C(q)\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\int_{\mathbb{R}_+}\left[1+\langle Z^n_s,1\rangle^{q-1}\right]Z^n_s(dx)ds\right)\\ &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\bar{B}C(q)\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\left[1+\langle Z^n_s,1\rangle^{q-1}\right]\left(\int_{\mathbb{R}_+}Z^n_s(dx)\right)ds\right)\\ &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\bar{B}C(q)\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\left[\langle Z^n_s,1\rangle+\langle Z^n_s,1\rangle^q\right]ds\right)\\ &\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+\bar{B}C(q)\,\mathbb{E}\left(\int_0^{t\wedge\tau^n_N}\left[\langle Z^n_s,1\rangle+\langle Z^n_s,1\rangle^q\right]ds\right).\end{aligned}$$

Since

$$\mathbb{E}\left(\langle Z^n_s,1\rangle+\langle Z^n_s,1\rangle^q\right)\le 2\mathbb{E}\left(1+\langle Z^n_s,1\rangle^q\right)\le 2(1+\mathbb{E}(\sup_{u\in[0,s]}\langle Z^n_u,1\rangle^q)),$$

we have

$$\mathbb{E}\left(\sup_{s\in[0,t\wedge\tau^n_N]}\langle Z^n_s,1\rangle^q\right)\le \mathbb{E}\left(\langle Z^n_0,1\rangle^q\right)+2\bar{B}C(q)T+2\bar{B}C(q)T\int_0^t\mathbb{E}\Big(\sup_{u\in[0,s\wedge\tau^n_N]}\langle Z^n_u,1\rangle^q\Big)ds.$$

By using Gronwall's inequality, we obtain

$$\mathbb{E}\left(\sup_{t\in[0,T\wedge\tau_N^n]}\langle Z_t^n,1\rangle^q\right)\leq\Big(\mathbb{E}\left(\langle Z_0^n,1\rangle^q\right)+2\bar{B}C(q)T\Big)\,e^{2\bar{B}C(q)T}\leq C(q,T),\qquad(3.10)$$

with $C(q,T)$ a constant which is independent of n and N.

To end the proof, we need to show that

$$\tau_\infty^n=\lim_{N\to\infty}\tau_N^n=+\infty.\qquad(3.11)$$

If there exists $M<+\infty$ associated with $A_M\subset\Omega$ such that $\forall\omega\in A_M$, $\lim_{N\to\infty}\tau_N^n(\omega)<M$ and $\mathbb{P}(A_M)>0$, then

$$\forall T>M,\ \mathbb{E}\left(\sup_{t\in[0,T\wedge\tau_N^n]}\langle Z_t^n,1\rangle^q\right)\geq\mathbb{P}(A_M)\times N^q.$$

This contrasts with (3.10) since $\mathbb{P}(A_M)\times N^q$ depends on N. Therefore, $\tau_\infty^n=+\infty$.

Using condition (3.8) and Fatou's lemma, we obtain

$$\mathbb{E}\left(\liminf_{N\to\infty}\sup_{t\in[0,T\wedge\tau_N^n]}\langle Z_t^n,1\rangle^q\right)\leq\liminf_{N\to\infty}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tau_N^n]}\langle Z_t^n,1\rangle^q\right)\leq C(q,T)<+\infty,$$

and we conclude that $\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\langle Z_t^n,1\rangle^q\right)<+\infty.$

□

Corollary 3.2.6. *From Proposition 3.2.5, if*

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}(\langle Z_0^n,1\rangle)<+\infty,$$

then, for all $T>0$

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\langle Z_t^n,1\rangle\right)<+\infty$$

and we deduce that the SDE (3.6) has a solution $(Z_t^n)_{t\geq0}\in\mathbb{D}\left([0,T],\mathcal{M}_F(\mathbb{R}_+)\right)$.

Proposition 3.2.7 (Martingale property)**.** *Let $n\in\mathbb{N}^*$, we assume that $Z_0^n\in\mathcal{M}_F(\mathbb{R}_+)$ such that $\mathbb{E}(\langle Z_0^n,1\rangle^2)<+\infty$, then for all test function $f(x,t)\in\mathcal{C}_b^1(\mathbb{R}_+\times\mathbb{R}_+,\mathbb{R})$:*

$$\begin{aligned}M_t^{n,f}=\langle Z_t^n,f\rangle-\langle Z_0^n,f\rangle&-\int_0^t\int_{\mathbb{R}_+}\Big(\partial_sf_s(x)+\alpha g(x)\partial_xf_s(x)\Big)Z_s^n(dx)ds\\&-\int_0^t\int_{\mathbb{R}_+}\int_0^1\left[f_s(\gamma x)+f_s((1-\gamma)x)-f_s(x)\right]B(x)H(d\gamma)Z_s^n(dx)ds.\end{aligned}\qquad(3.12)$$

is a continuous square integrable martingale with quadratic variation

$$\langle M^{n,f}\rangle_t = \frac{1}{n}\int_0^t \int_{\mathbb{R}_+}\int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right]^2 B(x)H(d\gamma)Z_s^n(dx)ds. \tag{3.13}$$

The martingale property and quadratic variation are direct consequence of the stochastic process $(Z^n)_{n\in\mathbb{N}^*}$ w.r.t Poisson point measures.

Proof. Let τ_N^n be the stopping-times as in Definition 3.2.4. We have for $t \in \mathbb{R}_+$

$$\langle M^{n,f}\rangle_{t\wedge\tau_N^n} = \frac{1}{n}\int_0^{t\wedge\tau_N^n} \int_{\mathbb{R}_+}\int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right]^2 B(x)H(d\gamma)Z_s^n(dx)ds,$$

then

$$\langle M^{n,f}\rangle_{t\wedge\tau_N^n} \leq \frac{1}{n}\int_0^{t\wedge\tau_N^n} 3\bar{B}\|f\|_\infty^2 \left(\int_{\mathbb{R}_+} Z_s^n(dx)\right) ds.$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(\langle M^{n,f}\rangle_{t\wedge\tau_N^n}\right) &\leq \frac{1}{n}3\bar{B}\|f\|_\infty^2 \mathbb{E}\left(\int_0^{t\wedge\tau_N^n}\langle Z_s^n, 1\rangle ds\right)\\
&\leq \frac{1}{n}3\bar{B}\|f\|_\infty^2 \mathbb{E}\left(\int_0^{t}\sup_{s\in[0,t\wedge\tau_N^n]}\langle Z_s^n, 1\rangle ds\right)\\
&\leq \frac{t}{n}3\bar{B}\|f\|_\infty^2 \mathbb{E}\left(\sup_{s\in[0,t\wedge\tau_N^n]}\langle Z_s^n, 1\rangle\right).
\end{aligned}$$

Since $\lim_{N\to\infty}\tau_N^n = +\infty$ and using the Fatou's lemma, we obtain

$$\mathbb{E}\left(\langle M^{n,f}\rangle_t\right) \leq \liminf_{N\to\infty}\mathbb{E}\left(\langle M^{n,f}\rangle_{t\wedge\tau_N^n}\right) \leq \frac{t}{n}3\bar{B}\|f\|_\infty^2 \mathbb{E}\left(\sup_{s\in[0,t]}\langle Z_s^n, 1\rangle\right) < +\infty.$$

Hence, $M_t^{n,f}$ is a square integrable martingale.

□

3.2.2 Large population limit

The following theorem states the limit of $(Z^n)_{n\in\mathbb{N}^*}$ when $n \to +\infty$.

Theorem 3.2.8. *Consider the sequence $(Z^n)_{n\in\mathbb{N}^*}$ defined in Definition 3.2.1 and 3.2.2. If Z_0^n converges in distribution to $\mu_0 \in \mathcal{M}_F(\mathbb{R}_+)$ as $n \to +\infty$ then $(Z^n)_{n\in\mathbb{N}^*}$ converges in distribution in $\mathbb{D}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$ as $n \to +\infty$ to $\mu \in \mathcal{C}\left([0,T], \mathcal{M}_F(\mathbb{R}_+)\right)$, where μ is the unique solution of*

$$\begin{aligned}
\langle \mu_t, f\rangle = \langle \mu_0, f\rangle &+ \int_0^t\int_{\mathbb{R}_+}\Big(\partial_s f_s(x) + \alpha g(x)\partial_x f_s(x)\Big)\mu_s(dx)ds\\
&+ \int_0^t\int_{\mathbb{R}_+}\int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)\mu_s(dx)ds,
\end{aligned} \tag{3.14}$$

with $f_t(x) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+\times\mathbb{R}_+, \mathbb{R})$ a test function.

Proof. Following Fournier and Méléard [50], we divide the proof into two parts: first, we will prove that $(Z^n)_{n\in\mathbb{N}^*}$ is tight in $\mathbb{D}\left([0,T],\mathcal{M}_F(\mathbb{R}_+)\right)$, where $\mathcal{M}_F(\mathbb{R}_+)$ is embedded with the topology of weak convergence. Then, we identify the limit in the second part.

Step 1: Tightness of $(Z^n)_{n\in\mathbb{N}^*}$.

Firstly, we prove that the sequence $(Z^n)_{n\in\mathbb{N}^*}$ is uniformly tight in the space of probability measures on $\mathbb{D}\left([0,T],(\mathcal{M}_F(\mathbb{R}_+),v)\right)$ where $(\mathcal{M}_F(\mathbb{R}_+),v)$ is embedded with the vague topology.

For a test function $f_t(x) = f(x,t) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+\times\mathbb{R}_+,\mathbb{R})$, we define the finite variation part of $\langle Z_t^n, f\rangle$ given in (3.12) by

$$\begin{aligned} V_t^{n,f} &= \langle Z_t^n, f\rangle - \langle Z_0^n, f\rangle - M_t^{n,f} \\ &= \int_0^t\int_{\mathbb{R}_+}\int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)Z_s^n(dx)ds \\ &\qquad + \int_0^t\int_{\mathbb{R}_+}\Big(\partial_s f_s(x) + \alpha g(x)\partial_x f_s(x)\Big)Z_s^n(dx)ds. \end{aligned} \tag{3.15}$$

Using Aldous-Rebolledo and Roelly's criterion [3, 68], we need to show that

1. $\forall t\in\mathcal{T}$ dense in $\mathbb{R}_+$, $\left(\langle M^{n,f}\rangle_t\right)_{n\in\mathbb{N}^*}$ and $\left(V_t^{n,f}\right)_{n\in\mathbb{N}^*}$ are tight in $\mathbb{R}_+$.
2. $\forall T\geq 0$, $\forall\epsilon>0$, $\forall\eta>0$, $\exists\delta>0$, $n_0\in\mathbb{N}$ such that
$$\sup_{n\geq n_0}\mathbb{P}\left(\left|\langle M^{n,f}\rangle_{T_n} - \langle M^{n,f}\rangle_{S_n}\right|\geq\eta\right)\leq\epsilon, \tag{3.16}$$
and
$$\sup_{n\geq n_0}\mathbb{P}\left(\left|V_{T_n}^{n,f} - V_{S_n}^{n,f}\right|\geq\eta\right)\leq\epsilon, \tag{3.17}$$
for every couples of stopping-times $(S_n,T_n)_{n\in\mathbb{N}^*}$ such that $S_n\leq T_n\leq T$ and $T_n\leq S_n+\delta$.

To prove the first point, we need to show that, $\forall T>0$

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\left|\langle M^{n,f}\rangle_t\right|\right) < +\infty,$$

and

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\left|V_t^{n,f}\right|\right) < +\infty.$$

From the expression of $\langle M^{n,f}\rangle_t$ in (3.13), we have

$$\begin{aligned} \left|\langle M^{n,f}\rangle_t\right| &\leq \frac{1}{n}\int_0^t 3\bar{B}\|f\|_\infty^2\left(\int_{\mathbb{R}_+}Z_s^n(dx)\right)ds \leq \frac{1}{n}3\|f\|_\infty^2\bar{B}\int_0^t\sup_{u\in[0,T]}\langle Z_u^n,1\rangle ds \\ &\leq \frac{1}{n}3\bar{B}T\|f\|_\infty^2\sup_{t\in[0,T]}\langle Z_t^n,1\rangle. \end{aligned}$$

Hence,

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\left|\langle M^{n,f}\rangle_t\right|\right)\le 3\bar{B}T\|f\|_\infty^2\sup_{n\in\mathbb{N}^*}\left(\frac{1}{n}\sup_{t\in[0,T]}\langle Z_t^n,1\rangle\right)<+\infty.$$

Similarly, from (3.15) we have

$$\begin{aligned}\left|V_t^{n,f}\right| &\le \Big(\|\partial_t f\|_\infty+\alpha\|\partial_x f\|_\infty\|g\|_\infty\Big)T\sup_{t\in[0,T]}\langle Z_t^n,1\rangle+3\bar{B}T\|f\|_\infty\sup_{t\in[0,T]}\langle Z_t^n,1\rangle\\ &\le \Big(\|\partial_t f\|_\infty+\alpha\|\partial_x f\|_\infty\|g\|_\infty+3\bar{B}\|f\|_\infty\Big)T\sup_{t\in[0,T]}\langle Z_t^n,1\rangle.\end{aligned}$$

Hence,

$$\sup_{n\in\mathbb{N}^*}\mathbb{E}\left(\sup_{t\in[0,T]}\left|V_t^{n,f}\right|\right)<+\infty.$$

Next, $\forall T>0$, $\forall\epsilon>0$, $\forall\eta>0$, for $\delta>0$, we consider a sequence of couples of stopping-times $(S_n,T_n)_{n\in\mathbb{N}^*}$ such that $S_n\le T_n\le T$ and $T_n\le S_n+\delta$, we have

$$\begin{aligned}&\mathbb{E}\left(\left|\langle M^{n,f}\rangle_{T_n}-\langle M^{n,f}\rangle_{S_n}\right|\right)\\ &=\mathbb{E}\left(\left|\frac{1}{n}\int_{S_n}^{T_n}\int_{\mathbb{R}_+}\int_0^1\left[f_s(\gamma x)+f_s((1-\gamma)x)-f_s(x)\right]^2B(x)H(d\gamma)Z_s^n(dx)ds\right|\right)\\ &\le\mathbb{E}\left(\frac{1}{n}\int_{S_n}^{T_n}3\|f\|_\infty^2\bar{B}\int_{\mathbb{R}_+}Z_s^n(dx)ds\right)\\ &\le\frac{1}{n}3\bar{B}\delta\|f\|_\infty^2\mathbb{E}\left(\sup_{t\in[0,T]}\langle Z_t^n,1\rangle\right)\le\frac{\delta C_{f,\bar{B},T}}{n}.\end{aligned}$$

The upper bound can be as small as we wish with a proper choice of δ. Thus, we obtain (3.16). Similarly,

$$\begin{aligned}\mathbb{E}\left(\left|V_{T_n}^{n,f}-V_{S_n}^{n,f}\right|\right)&=\mathbb{E}\Big(\Big|\int_{S_n}^{T_n}\int_{\mathbb{R}_+}\Big(\partial_s f_s(x)+\alpha g(x)\partial_x f_s(x)\Big)Z_s^n(dx)ds\\ &\qquad+\int_{S_n}^{T_n}\int_{\mathbb{R}_+}\int_0^1\left[f_s(\gamma x)+f_s((1-\gamma)x)-f_s(x)\right]B(x)H(d\gamma)Z_s^n(dx)ds\Big|\Big)\\ &\le\Big(\|\partial_t f\|_\infty+\alpha\|\partial_x f\|_\infty\|g\|_\infty\Big)\delta\sup_{n\in\mathbb{N}^*}\mathbb{E}\Big(\sup_{t\in[0,T]}\langle Z_s^n,1\rangle\Big)+3\bar{B}\delta\|f\|_\infty\sup_{n\in\mathbb{N}^*}\mathbb{E}\Big(\sup_{t\in[0,T]}\langle Z_s^n,1\rangle\Big)\\ &\le\delta\left(\|\partial_t f\|_\infty+\alpha\|\partial_x f\|_\infty\|g\|_\infty+3\bar{B}\|f\|_\infty\right)\sup_{n\in\mathbb{N}^*}\mathbb{E}\Big(\sup_{t\in[0,T]}\langle Z_s^n,1\rangle\Big)\\ &\le\delta C_{f,g,\alpha,\bar{B},T}.\end{aligned}$$

By the choice of δ, we obtain (3.17) and we conclude that $(Z^n)_{n\in\mathbb{N}^*}$ is uniformly tight in $\mathbb{D}([0,T],(\mathcal{M}_F(\mathbb{R}_+),v))$. Next, by the same computation as in Méléard and Tran [81], we shall prove the tightness of $(Z^n)_{n\in\mathbb{N}^*}$ in $\mathbb{D}([0,T],(\mathcal{M}_F(\mathbb{R}_+),w))$.

Let us denote by μ a limiting process of $(Z^n)_{n\in\mathbb{N}^*}$. It is almost surely (a.s) continuous in $(\mathcal{M}_F(\mathbb{R}_+), v)$ since

$$\sup_{t\in\mathbb{R}_+} \sup_{f,\|f\|_\infty\le 1} \Big|\langle Z_t^n, f\rangle - \langle Z_{t-}^n, f\rangle\Big| \le \frac{3\|f\|_\infty}{n}. \tag{3.18}$$

The same computation as in Step 1 for $f_t(x) = 1$ implies that the sequence $(\langle Z^n, 1\rangle)_{n\in\mathbb{N}^*}$ is uniformly tight in $\mathbb{D}([0,T], \mathbb{R}_+)$. As a consequence, there exists an increasing sequence $(u_n)_{n\in\mathbb{N}^*}$ such that:

- $(Z^{u_n})_{n\in\mathbb{N}^*}$ converges in distribution to μ in $\mathbb{D}([0,T], (\mathcal{M}_F(\mathbb{R}_+), v))$.
- $(\langle Z^{u_n}, 1\rangle)_{n\in\mathbb{N}^*}$ converges in distribution in $\mathbb{D}([0,T], \mathbb{R}_+)$.

We can use the Méléard-Roelly's criterion [80] to prove that $(Z^{u_n})_{n\in\mathbb{N}^*}$ converges in distribution to $\mu \in \mathbb{D}([0,T], (\mathcal{M}_F(\mathbb{R}_+), w))$ provided $(\langle Z^n, 1\rangle)_{n\in\mathbb{N}^*}$ converges to $\langle \mu, 1\rangle$.

For the sake of simplicity, we will again denote u_n by n.

Now, we introduce of smooth function f_k defined on $\mathbb{R}_+$ and approximating $\mathbb{1}_{[k,+\infty)}(x)$. For $k \in \mathbb{N}^*$, let us define

$$f_k(x) = \psi(0 \vee (x-(k-1)) \wedge 1),$$

where $\psi(x) = 6x^5 - 15x^4 + 10x^3$ is a non-decreasing function such that

$$\psi(0) = \psi'(0) = \psi''(0) = 1 - \psi(1) = \psi'(1) = \psi''(1) = 0.$$

The sequence $(f_k)_{k\in\mathbb{N}^*}$ is non-increasing, and satisfies for $x \ge 0$ and $p \ge 1$ that

$$\begin{aligned} &\mathbb{1}_{[k,+\infty)}(x) \le f_k(x) \le \mathbb{1}_{[k-1,+\infty)}(x); \\ &f_k^{(p)}(x) \le \sup_{u\in[k-1,k]} |f_k^{(p)}(u)| \mathbb{1}_{[k-1,+\infty)}(x) \le \sup_{u\in[k-1,k]} |f_k^{(p)}(u)| f_{k-1}(x). \end{aligned} \tag{3.19}$$

To prove that $(\langle Z^n, 1\rangle)_{n\in\mathbb{N}^*}$ converges to $\langle \mu, 1\rangle$, we use the following lemma:

Lemma 3.2.9. *Under the assumption of Theorem 3.2.8,*

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\Big(\sup_{t\in[0,T]} \langle Z_t^n, f_k\rangle\Big) = 0, \tag{3.20}$$

where $(f_k)_{k\in\mathbb{N}}$ are defined as above.

The proof is postponed at the end of Theorem 3.2.8. From Lemma 3.2.9, we can deduce that

$$\lim_{k\to+\infty} \mathbb{E}\Big(\sup_{t\in[0,T]} \langle \mu_t, f_k\rangle\Big) = 0. \tag{3.21}$$

As a consequence, we can extract a subsequence of $\Big(\sup_{t\in[0,T]} \langle \mu_t, f_k\rangle\Big)_k$ that converges to 0 a.s, and since the process $(\mu_t)_{t\in[0,T]}$ is continuous from $[0,T]$ into $(\mathcal{M}_F(\mathbb{R}_+), v)$, one can deduce that it is also continuous from $[0,T]$ into $(\mathcal{M}_F(\mathbb{R}_+), w)$.

We now prove that $\langle Z^{u_n}, 1\rangle$ converges to $\langle \mu, 1\rangle$. Let G be a Lipschitz function on $\mathcal{C}\Big([0,T],(\mathcal{M}_F(\mathbb{R}_+),w)\Big)$, we have

$$\begin{aligned}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle Z^n,1\rangle)-G(\langle\mu,1\rangle))\Big| &\leq \lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle Z^n,1\rangle)-G(\langle Z^n,1-f_k\rangle))\Big| \\ &\quad+\lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle Z^n,1-f_k\rangle)-G(\langle\mu,1-f_k\rangle))\Big| \\ &\quad+\lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle\mu,1-f_k\rangle)-G(\langle\mu,1\rangle))\Big|.\end{aligned}$$

Since $\Big|G(\langle\nu,1-f_k\rangle)-G(\langle\nu,1\rangle))\Big| \leq C\sup_{t\in[0,T]}\langle\nu_t,f_k\rangle$ by Lipschitz property, according to Lemma 3.2.9 and Equation (3.21), we obtain

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle Z^n,1\rangle)-G(\langle Z^n,1-f_k\rangle))\Big|=0,$$

and

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle\mu,1-f_k\rangle)-G(\langle\mu,1\rangle))\Big|=0.$$

By the continuity of $\nu\mapsto\langle\nu,1-f_k\rangle$ in $\mathbb{D}\Big([0,T],(\mathcal{M}_F(\mathbb{R}_+),v)\Big)$ we have

$$\lim_{k\to+\infty}\limsup_{n\to+\infty}\Big|\mathbb{E}(G(\langle Z^n,1-f_k\rangle)-G(\langle\mu,1-f_k\rangle))\Big|=0.$$

Thus,

$$\lim_{n\to+\infty}F\Big(\langle Z^n,1\rangle\Big)=F\Big(\langle\mu,1\rangle\Big),$$

which completes the proof of tightness of $(Z^n)_{n\in\mathbb{N}^*}$ in $\mathbb{D}\Big([0,T],(\mathcal{M}_F(\mathbb{R}_+),w)\Big)$.

Step2: Indentification the limit.

To prove that μ satisfies (3.14) a.s, we will show that, for every test function $f:(t,x)\mapsto f(t,x)\in\mathcal{C}_b^{1,1}(\mathbb{R}_+\times\mathbb{R}_+,\mathbb{R})$, the quantity

$$\begin{aligned}\Psi_t(\mu)=\langle\mu_t,f\rangle-\langle\mu_0,f\rangle-\int_0^t\int_{\mathbb{R}_+}\Big(\partial_s f_s(x)+\alpha g(x)\partial_x f_s(x)\Big)\mu_s(dx)ds \\ +\int_0^t\int_{\mathbb{R}_+}\int_0^1\left[f_s(\gamma x)+f_s((1-\gamma)x)-f_s(x)\right]B(x)H(d\gamma)\mu_s(dx)ds\end{aligned}$$

vanishes.

Proposition 3.2.7 shows that $\forall n\in\mathbb{N}^*$ the process

$$M_t^{f,\phi(n)}=\Psi_t\left(Z_t^{\phi(n)}\right)$$

is a square integrable martingale. Thus, for $t\in\mathbb{R}_+$ we have

$$\mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right)^2\leq\mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|^2\right)=\mathbb{E}\Big(\langle M^{f,\phi(n)}\rangle_t\Big)\leq\frac{tC}{\phi(n)}\mathbb{E}\Big(\sup_{s\in[0,t]}\langle Z_s^{\phi(n)},1\rangle\Big).$$

Then,

$$\forall t \in \mathbb{R}_+, \lim_{n\to+\infty} \mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right) = 0.$$

To show that, $\forall t \in \mathbb{R}_+$, $\Psi_t(\mu) = 0$ a.s, we need to prove that

$$\lim_{n\to+\infty} \mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right) = \mathbb{E}\left(\left|\Psi_t(\mu)\right|\right).$$

Indeed, we have that $(Z^{\phi(n)})_{n\in\mathbb{N}^*}$ converges in distribution to μ. Since μ is continuous almost surely, $f \in \mathcal{C}^1$ class with bounded derivatives and from the moment assumption of Proposition 3.2.5 then Ψ_t is continuous and

$$\lim_{n\to+\infty} \Psi_t(Z^{\phi(n)}) \xrightarrow{d} \Psi_t(\mu).$$

We have: $\forall t \in \mathbb{R}_+$, $\forall z \in \mathcal{M}_F(\mathbb{R}_+)$, $\Psi_t(z) \le C \sup_{s\in[0,t]} \langle z_s, 1\rangle$. Hence, by Proposition 3.2.5, the sequence $\left(|\Psi_t(Z^{\phi(n)})|\right)_{n\in\mathbb{N}^*}$ is uniform integrability. Then we get

$$\lim_{n\to+\infty} \mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right) = \lim_{n\to+\infty} \mathbb{E}\left(\left|\Psi_t(Z^{\phi(n)})\right|\right) = \mathbb{E}\left(\lim_{n\to+\infty}\left|\Psi_t(Z^{\phi(n)})\right|\right)$$

This ends the proof of Theorem 3.2.8.

□

Proof of Lemma 3.2.9. With the sequence $(f_k)_{k\in\mathbb{N}^*}$ defined in the proof of Theorem 3.2.8, we have

$$\begin{aligned}\langle Z_t^n, f_k\rangle = \langle Z_0^n, f_k\rangle &+ \int_0^t \int_{\mathbb{R}_+} f_k'(x)\alpha g(x) Z_s^n(dx)ds \\ &+ \int_0^t \int_{\mathbb{R}_+}\int_0^1 \Big[f_k(\gamma x) + f_k((1-\gamma)x) - f_k(x)\Big] B(x)H(d\gamma)Z_s^n(dx)ds + M_t^{n,k},\end{aligned}$$

where the martingale $M_t^{n,k}$ defined in (3.12) with f_k instead of f, and with quadratic variation given in (3.13). Similar arguments as above allow us to prove that

$$\langle M^{n,k}\rangle_t \le \frac{C_1}{n}\int_0^t \langle Z_s^n, 1\rangle ds, \tag{3.22}$$

where C_1 is a constant. By (3.19) and since the sequence $(f_k)_{k\in\mathbb{N}^*}$ is non-increasing, $\langle Z_s^n, f_k\rangle \le \langle Z_s^n, f_{k-1}\rangle$, we obtain

$$\begin{aligned}\langle Z_t^n, f_k\rangle &\le \langle Z_0^n, f_k\rangle + \alpha\|g\|_\infty C_2\int_0^t \langle Z_t^n, f_{k-1}\rangle ds + 3\bar{B}\int_0^t \langle Z_t^n, f_{k-1}\rangle ds + M_t^{n,k} \\ &\le \langle Z_0^n, f_k\rangle + C_3\int_0^t \langle Z_t^n, f_{k-1}\rangle ds + M_t^{n,k},\end{aligned}$$

where C_2, C_3 are the constants.

Let $\alpha_t^{n,k} = \mathbb{E}\Big(\sup_{s\le t}\langle Z_t^n, f_k\rangle\Big)$ and $\alpha_t^n = \mathbb{E}\Big(\sup_{s\le t}\langle Z_s^n, 1\rangle\Big)$ which bounded uniformly in $n\in\mathbb{N}^*$ and $t\in[0,T]$ according to (3.9). From (3.22), one can deduces that

$$\alpha_t^{n,k} \le \alpha_0^{n,k} + C_3\int_0^t \alpha_s^{n,k-1}ds + C_4\epsilon_n$$

where ϵ_n is a sequence that converges to 0 as $n\to+\infty$.

Iterating this inequality yields,

$$\begin{aligned}\alpha_t^{n,k} &\le \sum_{l=0}^{k-1}\alpha_0^{n,k-l}\frac{(C_3t)^l}{l!} + \frac{\Big(C_3\int_0^t\alpha_s^n ds\Big)^k}{k!} + \epsilon_n C_4\sum_{l=0}^{k-1}\frac{(C_3t)^l}{l!}\\ &\le \alpha_0^{n,\lfloor k/2\rfloor}e^{C_3t} + \alpha_0^n\sum_{l=\lfloor k/2\rfloor+1}^{+\infty}\frac{(C_3t)^l}{l!} + \frac{(C_3't)^k}{k!} + \epsilon_n C_4 e^{C_3t},\end{aligned}$$

where we use the monotonicity of $\alpha_0^{n,k}$ w.r.t k for the second inequality. Given the moment condition (3.9), the assumption of tightness in $(\mathcal{M}_F(\mathbb{R}_+), w)$ of the initial conditions $(Z_0^n)_{n\in\mathbb{N}^*}$ is equivalent to

$$\lim_{k\to+\infty}\sup_{n\to+\infty}\alpha_0^{n,k} = 0$$

Hence,

$$\lim_{k\to+\infty}\sup_{n\to+\infty}\alpha_t^{n,k} \le \sup_{n\in\mathbb{N}^*}\alpha_0^n\lim_{k\to+\infty}\sum_{\lfloor k/2\rfloor+1}^{+\infty}\frac{(C_3t)^l}{l!} + \lim_{k\to+\infty}\frac{(C_3't)^k}{k!}.$$

As a consequence, we deduce that

$$\lim_{k\to+\infty}\sup_{n\to+\infty}\mathbb{E}\Big(\sup_{s\le t}\langle Z_t^n, f_k\rangle\Big) = \lim_{k\to+\infty}\sup_{n\to+\infty}\alpha_t^{n,k} = 0,$$

which completes the proof of Lemma 3.2.9.

□

The following proposition give us a growth-fragmentation equation in the case where the growth rate is a constant.

Proposition 3.2.10. *We assume that $g(x) = 1$ for all $x\in\mathbb{R}_+$. This implies that $X_t = x_0 + \alpha(t-t_0)$ if the toxicity is x_0 at times t_0. Then, we have the following results:*

i. If $\mu_0(dx) = n_0(dx)$ then $\forall t\in\mathbb{R}_+$, $\mu_t(dx)$ has a density $n(t,x)$.

ii. If $n(t,x)\in\mathcal{C}^{1,1}(\mathbb{R}_+,\mathbb{R}_+)$ then it satisfies the PDE:

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) + B(x)n(t,x) = 2\int_0^1\frac{1}{\gamma}B\left(\frac{x}{\gamma}\right)n\left(t,\frac{x}{\gamma}\right)H(d\gamma) \quad (3.23)$$

Proof. *i.* We use an approach inspired by Tran [105, 106] but with more differences. In our problem, we consider the case where a cell divides with random fraction Γ and $(1-\Gamma)$ into the two daughter cells respectively and Γ has a distribution $H(d\gamma)$.

Let μ_t be a limit of $(Z_t^n)_{n\in\mathbb{N}^*}$ in $\mathcal{C}([0,T],\mathbb{R}_+)$, for a test function $f_t(x) = f(x,t)$ we have

$$\langle \mu_t, f_t\rangle = \langle \mu_0, f_0\rangle + \int_0^t \int_{\mathbb{R}_+} \left[\partial_s f_s(x) + \alpha \partial_x f_s(x)\right] \mu_s(dx)ds + \int_0^t \int_{\mathbb{R}_+} \int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)\mu_s(dx)ds.$$

To show that $\mu_t(dx)$ has a density, for fixed $t_0 \in \mathbb{R}_+$ and a positive function $\phi \in \mathcal{C}_b^1(\mathbb{R}_+,\mathbb{R}_+)$, we consider the function $f(x,t) = \phi(x-\alpha(t-t_0))$. Then, we get

$$\langle \mu_{t_0}, \phi\rangle = \int_{\mathbb{R}_+} \phi(x+\alpha t_0)\mu_0(dx) + \int_0^{t_0} \int_{\mathbb{R}_+} \int_0^1 \Big[\phi(\gamma x - \alpha(s-t_0)) + \phi((1-\gamma)x - \alpha(s-t_0)) - \phi(x-\alpha(s-t_0))\Big] B(x)H(d\gamma)\mu_s(dx)ds. \quad (3.24)$$

Consider the term

$$I_1 = \int_0^{t_0} \int_{\mathbb{R}_+} \int_0^1 \phi(\gamma x - \alpha(s-t_0))B(x)H(d\gamma)\mu_s(dx)ds.$$

By Fubini's theorem and setting $\nu(ds,dx) = \mu_s(dx)ds$, we get

$$I_1 = \int_0^1 H(d\gamma) \left(\int_0^{t_0} \int_{\mathbb{R}_+} \phi(\gamma x - \alpha(s-t_0))B(x)\nu(ds,dx) \right).$$

Let us show that

$$\nu(ds,dx) = \bar{\mu}(dx)q(x,ds) = \bar{\mu}(dx)q(x,s)ds,$$

with $\bar{\mu}(dx)$ the marginal measure of $\nu(ds,dx)$ such that

$$\int_{\mathbb{R}_+} g(x)\bar{\mu}(dx) = \int_0^{t_0} \int_{\mathbb{R}_+} g(x)\mu_s(dx)ds,$$

for every measurable function g.

Let A is a $ds-$ negligible set, we have $\forall g$

$$\int_0^{t_0} \int_{\mathbb{R}_+} \mathbb{1}_A(s)g(x)\nu(ds,dx) = \int_0^{t_0} \left[\int_{\mathbb{R}_+} \mathbb{1}_A(s)g(x)\mu_s(dx)\right] ds = 0.$$

Since $\nu(ds,dx) = \bar{\mu}(dx)q(x,ds)$, we have

$$\int_{\mathbb{R}_+} g(x)\bar{\mu}(dx) \int_0^{t_0} \mathbb{1}_A(s)q(x,ds) = 0,$$

so that

$$\int_0^{t_0} \mathbb{1}_A(s)q(x,ds) = 0 \qquad g(x)\bar{\mu}(dx) \text{ a.e.}$$

Hence, by Radon-Nikodym's theorem, there exists a density $q(x,s)$ such that

$$q(x,ds) = q(x,s)ds \qquad \bar{\mu}(dx) \text{ a.e.}$$

Thus, we get

$$I_1 = \int_0^1 H(d\gamma)\left(\int_0^{t_0}\int_{\mathbb{R}_+}\phi(\gamma x - \alpha(s-t_0))B(x)(dx)q(x,s)\bar{\mu}ds\right).$$

Let $y = \gamma x - \alpha(s-t_0)$, so that

$$\begin{aligned} I_1 &= \int_0^1 H(d\gamma)\left(\int_{\mathbb{R}_+}\int_{\gamma x}^{\gamma x+\alpha t_0}\phi(y)q\left(x,\frac{\gamma x - y + \alpha t_0}{\alpha}\right)B(x)\frac{dy}{\alpha}\bar{\mu}(dx)\right) \\ &= \int_0^1 H(d\gamma)\left(\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\mathbb{1}_{\gamma x\le y\le \gamma x+\alpha t_0}\phi(y)\tilde{q_1}(x,y)B(x)dy\bar{\mu}(dx)\right), \end{aligned}$$

where $\tilde{q_1}(x,y) = \dfrac{1}{\alpha}q\left(x,\frac{\gamma x-y+\alpha t_0}{\alpha}\right)$.

By Fubini's theorem, we obtain

$$I_1 = \int_{\mathbb{R}_+}\phi(y)\int_{\mathbb{R}_+}\int_0^1\mathbb{1}_{\gamma x\le y\le \gamma x+\alpha t_0}\phi(y)\tilde{q}(x,y)B(x)H(d\gamma)\bar{\mu}(dx)dy = \int_{\mathbb{R}_+}\phi(y)\Phi_1(y)dy,$$

where

$$\Phi_1(y) = \int_{\mathbb{R}_+}\int_0^1\mathbb{1}_{\gamma x\le y\le \gamma x+\alpha t_0}\phi(y)\tilde{q_1}(x,y)B(x)H(d\gamma)\bar{\mu}(dx). \tag{3.25}$$

Similarly, for the term

$$I_2 = \int_0^{t_0}\int_{\mathbb{R}_+}\int_0^1\phi\Big((1-\gamma)x - \alpha(s-t_0)\Big)B(x)H(d\gamma)\mu_s(dx)ds.$$

By putting $y = (1-\gamma)x - \alpha(s-t_0)$, and note that $\mu_s(dx)ds = \bar{\mu}(dx)q(x,s)ds$ we obtain

$$I_2 = \int_{\mathbb{R}_+}\phi(y)\Phi_2(y)dy,$$

where

$$\Phi_2(y) = \int_{\mathbb{R}_+}\int_0^1\mathbb{1}_{(1-\gamma)x\le y\le (1-\gamma)x+\alpha t_0}\phi(y)\tilde{q_2}(x,y)B(x)H(d\gamma)\bar{\mu}(dx). \tag{3.26}$$

For the last term

$$\begin{aligned} I_3 &= \int_0^{t_0}\int_{\mathbb{R}_+}\int_0^1\phi\Big(x-\alpha(s-t_0)\Big)B(x)H(d\gamma)\mu_s(dx)ds \\ &= \int_0^{t_0}\int_{\mathbb{R}_+}\phi\Big(x-\alpha(s-t_0)\Big)q(x,s)B(x)\bar{\mu}(dx)ds \\ &= \int_{\mathbb{R}_+}\phi(y)\Phi_3(y)dy. \end{aligned} \tag{3.27}$$

where

$$\Phi_3(y) = \int_{\mathbb{R}_+} \mathbb{1}_{x \le y \le x+\alpha t_0} \tilde{q}_3(x,y)\bar{\mu}(dx) \quad \text{and} \quad \tilde{q}_3(x,y) = \frac{1}{\alpha} q\left(x, \frac{x-y+\alpha t_0}{\alpha}\right).$$

Since

$$\int_{\mathbb{R}_+} \phi(x+\alpha t_0)\mu_0(dx) = \int_{\mathbb{R}_+} \phi(x+\alpha t_0) n_0(x)dx = \int_{\mathbb{R}_+} \phi(y) n_0(y-\alpha t_0)dy,$$

we set

$$\Phi(y) = \Phi_1(y) + \Phi_2(y) + \Phi_3(y) + n_0(y-\alpha t_0),$$

then

$$\int_{\mathbb{R}_+} \phi(x+\alpha t_0)\mu_0(dx) + \int_0^{t_0}\int_{\mathbb{R}_+}\int_0^1 \Big[\phi\Big(\gamma x - \alpha(s-t_0)\Big) + \phi\Big((1-\gamma)x - \alpha(s-t_0)\Big) \\ - \phi\Big(x - \alpha(s-t_0)\Big)\Big] H(d\gamma)B(x)\mu_s(dx)ds = \int_{\mathbb{R}_+} \phi(y)\Phi(y)dy.$$

We conclude $\mu_t(dx)$ has a density $n_t(x)dx$ w.r.t $\bar{\mu}_s(dx)ds$.

ii) From (3.14), we have

$$\langle \mu_t, f\rangle = \langle \mu_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} \alpha f'(x)\mu_s(dx) \\ + \int_0^t \int_{\mathbb{R}_+}\int_0^1 \Big[f(\gamma x) + f((1-\gamma)x) - f(x)\Big] B(x)H(d\gamma)\mu_s(dx)ds.$$

Replace $\mu_s(dx)$ by $n(t,x)dx$ and take the derivative by t, we obtain

$$\frac{\partial}{\partial t}\int_{\mathbb{R}_+} f(x)n(t,x)dx = \int_{\mathbb{R}_+} f'(x)\alpha n(t,x)dx - \int_{\mathbb{R}_+}\int_0^1 f(x)B(x)H(d\gamma)n(t,x)dx \\ + \int_{\mathbb{R}_+}\int_0^1 \Big[f(\gamma x) + f((1-\gamma)x)\Big] B(x)H(d\gamma)n(t,x)dx.$$

Integrate by part, we get

$$\int_{\mathbb{R}_+} f(x)\frac{\partial}{\partial t}n(t,x)dx = \alpha f(x)n(t,x)\Big|_0^{+\infty} - \int_{\mathbb{R}_+} \alpha f(x)\frac{\partial}{\partial x}n(t,x)dx \\ - \int_{\mathbb{R}_+} f(x)B(x)n(t,x)dx + \int_{\mathbb{R}_+} B(x)n(t,x)\left(\int_0^1 \left[f(\gamma x) + f((1-\gamma)x)\right] H(d\gamma)\right) dx. \tag{3.28}$$

By putting $\nu = 1-\gamma$ and $H(d\gamma)$ is a symmetric distribution, we get

$$\begin{aligned}
&\int_{\mathbb{R}_+} B(x)n(t,x)\left(\int_0^1 \left[f(\gamma x)+f((1-\gamma)x)\right]H(d\gamma)\right)dx \\
&= \int_{\mathbb{R}_+} B(x)n(t,x)\int_0^1 f(\gamma x)H(d\gamma)dx + \int_{\mathbb{R}_+} B(x)n(t,x)\int_0^1 f((1-\gamma)x)H(d\gamma)dx \\
&= \int_0^1 H(d\gamma)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{\gamma}\right)n\left(t,\frac{y}{\gamma}\right)\frac{dy}{\gamma} \\
&\qquad + \int_0^1 H(d\gamma)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{1-\gamma}\right)n\left(t,\frac{y}{1-\gamma}\right)\frac{dy}{1-\gamma} \\
&= \int_0^1 \frac{1}{\gamma}H(d\gamma)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{\gamma}\right)n\left(t,\frac{y}{\gamma}\right)dy + \int_0^1 \frac{1}{\nu}H(d\nu)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{\nu}\right)n\left(t,\frac{y}{\nu}\right)dy \\
&= 2\int_0^1 \frac{1}{\gamma}H(d\gamma)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{\gamma}\right)n\left(t,\frac{y}{\gamma}\right)dy. \qquad (3.29)
\end{aligned}$$

From (3.28) and (3.29) we otain:

$$\begin{aligned}
&\int_{\mathbb{R}_+} f(x)\frac{\partial}{\partial t}n(t,x)dx + \int_{\mathbb{R}_+} \alpha f(x)\frac{\partial}{\partial x}n(t,x)dx + \int_{\mathbb{R}_+} f(x)B(x)n(t,x)dx \\
&= 2\int_0^1 \frac{1}{\gamma}H(d\gamma)\int_{\mathbb{R}_+} f(y)B\left(\frac{y}{\gamma}\right)n\left(t,\frac{y}{\gamma}\right)dy,
\end{aligned}$$

or

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) + B(x)n(t,x) = 2\int_0^1 \frac{1}{\gamma}B\left(\frac{x}{\gamma}\right)n\left(t,\frac{x}{\gamma}\right)H(d\gamma).$$

This ends the proofs of Proposition 3.2.10.

□

In the following section, we assume the distribution $H(d\gamma)$ has a density $h(\gamma)$ that we aim to estimate.

3.3 Estimation of the division kernel $h(\cdot)$

3.3.1 Estimation procedure and assumptions

In this section, we consider the problem of estimating the density h in the case of incomplete data of divisions. As we introduced in Section 3.1, we shall construct an estimator of h based on the stationary size distribution which results from the study of the large population limit. Moreover, we consider the estimation of h in the case where the division rate and the growth rate are positive constants R and α respectively.

From Equation (3.23), with $H(d\gamma) = h(\gamma)d\gamma$ and $B(x) = R$, we obtain by setting $y = x/\gamma$

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) + Rn(t,x) = 2R\int_0^\infty n(t,y)h\left(\frac{x}{y}\right)\frac{dy}{y}, \tag{3.30}$$

where $h(x/y) = 0$ if $y < x$.

We assume that the division kernel h satisfies the following assumption:

Assumption 3.3.1. *There exists a positive constant C such that for any $t \in (0,1)$, for $\nu \in \{1,\ldots,4\}$*

$$\int_0^t h(x)dx \le \min\left(1, Ct^\nu\right). \tag{3.31}$$

Since h is the density of a symmetric distribution on $[0,1]$, it satisfies $\int h(x)dx = 1$ and $\int xh(x)dx = 1/2$. Moreover, under Assumption 3.3.1, it is proved in Doumic and Gabriel [37] (see Lemma 3 in the Appendix) that $\int x^2h(x)dx < 1/2$. Then, by general relative entropy principle (see [92]), the division kernel h satisfies the assumptions for the existence and uniqueness of the solution (λ, N) of the following eigenvalue problem (see [37] or [91] for more details):

$$\begin{cases} \alpha\partial_x N(x) + (\lambda + R)N(x) = 2R\int_0^\infty N(y)h\left(\frac{x}{y}\right)\frac{dy}{y}, & x \ge 0, \\ N(0) = 0, \quad \int N(x)dx = 1, \quad N(x) \ge 0, \quad \lambda > 0, \end{cases} \tag{3.32}$$

where N is the first eigenvector and λ is the first eigenvalue and

$$\int_0^\infty |n(t,x)e^{-\lambda t} - \rho N(x)|\phi(x)dx \to 0 \quad \text{as } t \to \infty,$$

with $\rho = \int_0^\infty n(t=0,u)\phi(u)du$.

The long-time asymptotic behavior of the solution $n(t,x)$ provides us an observation scheme for the estimation of the density h in the statistical approach: since $e^{-\lambda t}n(t,x) \approx N(x)$ as t is large, we assume that we have n i.i.d observations $X_1, X_2, \ldots, X_n$. Each observation is drawn by measuring an individual cell selected randomly. The X_i's have probability distribution $N(x)dx$ and we estimate h from the data $X_1, \ldots, X_n$ and Equation (3.32).

To estimate h, we face with a deconvolution problem but it is more complicated and quite different when compared to classical deconvolution problems. In particular, in Equation (3.32), one difficulty lies in the multiplicative convolution $\int_0^\infty N(y)h\left(\frac{x}{y}\right)\frac{dy}{y}$ leading to more intricate technical problems than for the classical additive convolution. So, we apply a logarithmic change of variables to transform the multiplicative convolution in the r.h.s of (3.32) into the additive one. Then we classically apply the Fourier transform and work with products of functions in the Fourier domain. Let us describe our estimation procedure in details. By using the change of variable $x = e^u$ for $x > 0$ and $u \in \mathbb{R}$, we introduce the functions

$$g(u) = e^u h(e^u),$$

and

$$M(u) = e^u N(e^u), \quad D(u) = \partial_u\Big(u \mapsto N(e^u)\Big) = \partial_u N(e^u).$$

Then Equation (3.32) becomes

$$\alpha D(u) + (\lambda + R)M(u) = 2R\Big(M \star g)(u). \tag{3.33}$$

Since $h(\gamma) = \gamma^{-1} g\Big(\log(\gamma)\Big)$ for $\gamma \in (0,1)$, the estimator of h will be obtained from the estimator of g. We need some assumptions on the density h and the Fourier transform of M.

Assumption 3.3.2. *The density h is continuous on $[0,1]$.*

Note that since h is continuous, g is square integrable since we have

$$\int_{\mathbb{R}} g^2(x)dx = \int_{\mathbb{R}} e^{2u} h^2(e^u)du = \int_0^\infty xh^2(x)dx = \int_0^1 xh^2(x)dx < +\infty.$$

Assumption 3.3.3. *For all $\xi \in \mathbb{R}$, $M^*(\xi) \neq 0$.*

The Fourier transform of g is given by the following proposition.

Proposition 3.3.4. *Under Assumption 3.3.3, we have*

$$g^*(\xi) = \frac{\alpha D^*(\xi)}{2RM^*(\xi)} + \frac{\lambda + R}{2R}, \quad \xi \in \mathbb{R}. \tag{3.34}$$

Proof. Recall Equation (3.32) and we set for $u, v \in \mathbb{R}$

$$x = e^u \text{ and } y = e^v.$$

Then we get

$$\alpha e^{-u}\partial_u N(e^u) + (\lambda + R)N(e^u) = 2R\int_{\mathbb{R}} N(e^v)h(e^{u-v})dv.$$

Multiply both sides of the equation above by e^u, we obtain

$$\alpha\partial_u N(e^u) + (\lambda + R)e^u N(e^u) = 2R\int_{\mathbb{R}} e^v N(e^v)e^{u-v}h(e^{u-v})dv.$$

Hence we get

$$\alpha D(u) + (\lambda + R)M(u) = 2R(M \star g)(u).$$

Taking the Fourier transform of both sides of equation (3.33), we obtain

$$\alpha D^*(\xi) + (\lambda + R)M^*(\xi) = 2RM^*(\xi) \times g^*(\xi).$$

Therefore, if $M^*(\xi) \neq 0$ for all $\xi \in \mathbb{R}$ then the Fourier transform of g is obtained by (3.34).

□

3.3.2 Moment condition

In this section, we prove that the moment condition required for the existence of the estimator of h is satisfied.

Proposition 3.3.5. *Under the Assumption 3.3.1, the first eigenvector N of the eigenproblem (3.32) satisfies*

$$\int_0^{+\infty} x^{-\nu} N(x)dx < +\infty \quad \textit{for } \nu \in \{1, \dots, 4\}. \tag{3.35}$$

Proof. Let $\epsilon > 0$ to be chosen small enough, we have

$$\begin{aligned}\int_0^{+\infty} x^{-\nu} N(x)dx &= \int_0^{\epsilon} x^{-\nu} N(x)dx + \int_{\epsilon}^{+\infty} x^{-\nu} N(x)dx \\ &\leq \int_0^{\epsilon} x^{-\nu} N(x)dx + \frac{1}{\epsilon^{\nu}} \int_{\epsilon}^{+\infty} N(x)dx \\ &\leq \int_0^{\epsilon} x^{-\nu} N(x)dx + \frac{1}{\epsilon^{\nu}}.\end{aligned}$$

Hence, it remains to prove

$$\int_0^{\epsilon} x^{-\nu} N(x)dx < +\infty.$$

We follow and adapt the steps of the proof of Theorem 1 in Doumic and Gabriel [37]. Integrating both side of equation (3.32) between 0 and $x_0 < x$, we get:

$$\alpha N(x_0) + (\lambda + R) \int_0^{x_0} N(y)dy = 2R \int_0^{x_0} \int_0^{+\infty} N(y)h\Big(\frac{z}{y}\Big)\frac{dy}{y}dz. \tag{3.36}$$

Thus,

$$\alpha N(x_0) \leq 2R \int_0^{x_0} \int_0^{+\infty} N(y)h\Big(\frac{z}{y}\Big)\frac{dy}{y}dz \leq 2R \int_0^{x} \int_0^{+\infty} N(y)h\Big(\frac{z}{y}\Big)\frac{dy}{y}dz.$$

Let us define:

$$f : x \mapsto \sup_{y \in (0,x)} N(y),$$

then we have for all x

$$f(x) \leq \frac{2R}{\alpha} \int_0^{x} \int_0^{+\infty} N(y)h\Big(\frac{z}{y}\Big)\frac{dy}{y}dz. \tag{3.37}$$

From Assumption 3.3.1 we have for all $x < \epsilon$:

$$\begin{aligned}f(x) &\leq \frac{2R}{\alpha} \int_0^{+\infty} dyN(y) \int_0^{x} h\Big(\frac{z}{y}\Big)\frac{dz}{y} \\ &\leq \frac{2R}{\alpha} \int_0^{+\infty} N(y) \min\Big(1, C\frac{x^{\nu}}{y^{\nu}}\Big)dy \\ &\leq \frac{2R}{\alpha} \bigg(\int_0^{x} N(y)dy + C\int_x^{\epsilon} N(y)\frac{x^{\nu}}{y^{\nu}}dy + C\int_{\epsilon}^{+\infty} N(y)\frac{x^{\nu}}{y^{\nu}}dy \bigg) \\ &\leq \frac{2R}{\alpha} \bigg(\int_0^{x} \sup_{z\in(0,y)} N(z)dy + Cx^{\nu} \int_x^{\epsilon} \sup_{z\in(0,y)} N(z)\frac{dy}{y^{\nu}} \bigg) + \bigg(\frac{2CR}{\alpha} \int_{\epsilon}^{+\infty} \frac{N(y)}{y^{\nu}}dy \bigg) x^{\nu} \\ &\leq \frac{2R\epsilon}{\alpha} f(x) + \frac{CRx^{\nu}}{\alpha} \int_x^{\epsilon} \frac{f(y)}{y^{\nu}}dy + Kx^{\nu},\end{aligned}$$

with $K = \dfrac{2CR}{\alpha\epsilon^{\nu}}$.

By setting $F(x) = x^{-\nu} f(x)$, we get

$$F(x) \leq \frac{K}{1 - \frac{2R\epsilon}{\alpha}} + \frac{2CR}{\alpha - 2R\epsilon} \int_x^{\epsilon} F(y)dy. \tag{3.38}$$

Hence, we have to choose ϵ such that

$$0 < \epsilon < \frac{\alpha}{2R}.$$

Then, applying Gronwall's inequality to (3.38), we find that

$$F(x) \leq \frac{K}{1 - \frac{2R\epsilon}{\alpha}} \exp\left(\frac{2CR\epsilon}{\alpha - 2R\epsilon}\right) := \tilde{C}, \quad \forall x \in [0, \epsilon].$$

Thus, we will have

$$x^{-\nu} N(x) \leq \tilde{C}, \quad \forall x \in [0, \epsilon].$$

We finally obtain

$$\int_0^{\epsilon} x^{-\nu} N(x) dx \leq \tilde{C}\epsilon < +\infty.$$

This ends the proof of Proposition 3.3.5.

□

3.3.3 Estimators of g and h

Assume that we observe the i.i.d random variables $X_1, \ldots, X_n$ having density function $x \mapsto N(x)$. We consider the random variables $U_1, \ldots, U_n$ where $U_i = \log(X_i)$ which are i.i.d of density function $u \mapsto M(u) = e^u N(e^u)$. In view of (3.34), the purpose is first to propose an estimator for g^* and then to apply the inverse Fourier transform to obtain an estimator of g.

Let K a kernel function in $\mathbb{L}^2(\mathbb{R})$ such that its Fourier transform K^* exists and is compactly supported. A possible kernel is given by the sinus cardinal kernel $K(x) = \frac{\sin(x)}{x}$. Define $K_\ell(\cdot) := \ell^{-1} K(\cdot/\ell)$ for $\ell > 0$. We set

$$g_\ell = K_\ell \star g.$$

From Equation (3.34) we have

$$g^*(\xi) = \frac{\alpha D^*(\xi)}{2RM^*(\xi)} + \frac{\lambda + R}{2R}.$$

Hence, a natural estimator of $g^*(\xi)$ is

$$\widehat{g^*(\xi)} = \frac{\alpha \widehat{D^*(\xi)}}{2R\widehat{M^*(\xi)}} + \frac{\lambda + R}{2R},$$

where $\widehat{M^*(\xi)}$ and $\widehat{D^*(\xi)}$ are unbiased estimators of M^* and D^* whose formulas are given in the following Proposition.

Proposition 3.3.6. *Unbiased estimators of the Fourier transform* $M^*(\xi)$ *and* $D^*(\xi)$ *are given by*

$$\widehat{M^*(\xi)} = \frac{1}{n}\sum_{j=1}^{n} e^{\mathbf{i}\xi U_j}, \tag{3.39}$$

$$\widehat{D^*(\xi)} = (-\mathbf{i}\xi)\frac{1}{n}\sum_{j=1}^{n} e^{(\mathbf{i}\xi-1)U_j}. \tag{3.40}$$

Proof. Since

$$M^*(\xi) = \int_{\mathbb{R}} e^u N(e^u) e^{\mathbf{i}u\xi} du = \mathbb{E}\Big[e^{\mathbf{i}\xi U_1}\Big],$$

and

$$\begin{aligned} D^*(\xi) &= \int_{\mathbb{R}} D(u) e^{\mathbf{i}u\xi} du = \int_{\mathbb{R}} \partial_u N(e^u) e^{\mathbf{i}u\xi} du \\ &= e^{\mathbf{i}\xi u} N(e^u)\Big|_{-\infty}^{+\infty} - \mathbf{i}\xi \int_{\mathbb{R}} N(e^u) e^{\mathbf{i}\xi u} du = (-\mathbf{i}\xi)\int_{\mathbb{R}} e^u N(e^u) e^{(\mathbf{i}\xi-1)u} du \\ &= (-\mathbf{i}\xi)\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right], \end{aligned}$$

then the unbiased estimators of the Fourier transform of M and D are given by (3.39) and (3.40).

□

Since $g^*_\ell = K^*_\ell \times g^*$, a natural estimator $\hat{g}_\ell$ of g is such that its Fourier transform takes the following form:

$$\widehat{g_\ell}^*(\xi) = K^*_\ell(\xi) \times \left(\frac{\alpha \widehat{D^*(\xi)}}{2R} \frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} + \frac{\lambda + R}{2R} \right), \tag{3.41}$$

where $\Omega = \Big\{|\widehat{M^*(\xi)}| \geq n^{-1/2}\Big\}$ and $\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}}$ is the truncated estimator of $\frac{1}{\widehat{M^*(\xi)}}$:

$$\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} = \begin{cases} \dfrac{1}{\widehat{M^*(\xi)}}, & \text{if } |\widehat{M^*(\xi)}| \geq n^{-1/2}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.42}$$

Truncation is necessary to avoid explosion when $|\widehat{M^*(\xi)}|$ is closed to 0.

Deconvolution estimators with close forms have been studied in Comte and Lacour [26, 27], Comte et al. [28], Neumann [87]. However, the difference and the difficulty in our problem come from the fact that the regularities of g and h are closely related to the functions M and D that solve the eigenvalue problem (3.32), in particular through the Equation (3.34). This complicates the study of the rates of convergence.

Finally, taking the inverse Fourier transform of $\widehat{g_\ell}^*$, we obtain

$$\hat{g}_\ell(u) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{g}^*_\ell(\xi) e^{-\mathbf{i}u\xi} d\xi. \tag{3.43}$$

Then the estimator of the division kernel h is given by

$$\hat{h}_\ell(\gamma) = \gamma^{-1}\hat{g}_\ell\Big(\log(\gamma)\Big), \quad \gamma \in (0,1). \tag{3.44}$$

3.3.4 Consistency for the $\mathbb{L}^2$-risk

This section is devoted to the theoretical study of the estimate $\hat{g}_\ell$. More precisely, we establish the $\mathbb{L}_2$-consistency of $\hat{g}_\ell$ under a suitable choice of the bandwidth ℓ.

Theorem 3.3.7. *We suppose that Assumptions 3.3.1, 3.3.2 and 3.3.3 are satisfied and the kernel bandwidth ℓ which depends on n satisfies $\lim_{n\to+\infty} \ell = 0$. Provided that*

$$\lim_{n\to+\infty} \frac{1}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2 + \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right) = 0, \tag{3.45}$$

we have

$$\lim_{n\to+\infty} \mathbb{E}\left[\|\hat{g}_\ell - g\|_2^2\right] = 0. \tag{3.46}$$

Note that if M^* is bounded from below by a positive constant on $[0,1]$ and if it satisfies

$$|M^*(\xi)| \overset{|\xi|\to+\infty}{\sim} |\xi|^{-\gamma}$$

then, Assumption (3.45) is satisfied if

$$\ell^{-1} = o\left(n^{\frac{1}{3+2\gamma}}\right).$$

Proof. To prove Theorem 3.3.7 we rely on the following proposition which gives a bias variance trade-off of the $\mathbb{L}^2$-risk of $\hat{g}_\ell$. In the sequel, the notation C denotes a positive constant which may change from line to line.

Proposition 3.3.8. *Under Assumptions 3.3.1, 3.3.2 and 3.3.3 , there exists a positive constant $C < +\infty$ such that*

$$\mathbb{E}\left[\|\hat{g}_\ell - g\|_2^2\right] \le \|K_\ell \star g - g\|_2^2 + \frac{C}{n}\square(\ell) \tag{3.47}$$

where

$$\square(\ell) = \left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2 + \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2.$$

Proof. We have

$$\|\hat{g}_\ell - g\|_2 \le \|g_\ell - g\|_2 + \|\hat{g}_\ell - g_\ell\|_2.$$

The first term of the above r.h.s inequality is a bias term whereas the second is a variance term. To control the variance term, we have by the Parseval's identity and by (3.41):

$$
\begin{aligned}
&\|\hat{g}_\ell - g_\ell\|_2^2 = \frac{1}{2\pi}\|\widehat{g_\ell}^* - g_\ell^*\|_2^2 \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| K_\ell^*(\xi)\Big[\Big(\frac{\alpha\widehat{D^*(\xi)}}{2R}\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} + \frac{\lambda+R}{2R}\Big) - g^*(\xi)\Big]\right|^2 d\xi \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| K_\ell^*(\xi)\Big[\Big(\frac{\alpha\widehat{D^*(\xi)}}{2R}\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\lambda+R}{2R}\Big) - g^*(\xi)\Big]\right|^2 d\xi \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| \frac{\alpha}{2R}K_\ell^*(\xi)\widehat{D^*(\xi)}\Big(\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}\Big) + K_\ell^*(\xi)\Big(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\lambda+R}{2R} - g^*(\xi)\Big)\right|^2 d\xi \\
&\le C\int_{\mathbb{R}} \left| K_\ell^*(\xi)\widehat{D^*(\xi)}\Big(\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}\Big)\right|^2 d\xi + C\int_{\mathbb{R}} |K_\ell^*(\xi)^2| \left|\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\lambda+R}{2R} - g^*(\xi)\right|^2 d\xi \\
&:= \mathrm{I} + \mathrm{II}\,.
\end{aligned}
$$

In the sequel, we deal with variance of complex variables. Note that for a complex variable say Z, by distinguishing real and imaginary parts one gets that

$$\mathbb{V}ar(Z) := \mathbb{E}[|Z-\mathbb{E}(Z)|^2] = \mathbb{E}[|Z|^2] - |\mathbb{E}[Z]|^2 \le \mathbb{E}[|Z|^2].$$

For the term II, because

$$
\begin{aligned}
\mathbb{E}\left(K_\ell^*(\xi)\left(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\lambda+R}{2R}\right)\right) &= K_\ell^*(\xi)\left(\frac{\alpha D^*(\xi)}{2RM^*(\xi)} + \frac{\lambda+R}{2R}\right) \\
&= K_\ell^*(\xi)g^*(\xi),
\end{aligned}
$$

we have

$$
\begin{aligned}
\mathbb{E}(\mathrm{II}) &= C\int_{\mathbb{R}} \mathbb{V}ar\left(K_\ell^*(\xi)\left(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\lambda+R}{2R}\right)\right) d\xi \\
&\le C\int_{\mathbb{R}} \mathbb{V}ar\left(K_\ell^*(\xi)\frac{\widehat{D^*(\xi)}}{M^*(\xi)}\right) d\xi \\
&\le C\int_{\mathbb{R}} \left|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right|^2 \mathbb{V}ar\left(\frac{(-\mathbf{i}\xi)}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}\right) d\xi \\
&\le \frac{C}{n}\int_{\mathbb{R}} \left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2 \mathbb{V}ar\left(e^{(\mathbf{i}\xi-1)U_1}\right) d\xi \\
&\le \frac{C}{n}\int_{\mathbb{R}} \left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2 \mathbb{E}\left[\left|e^{(\mathbf{i}\xi-1)U_1}\right|^2\right] d\xi \le \frac{C}{n}\int_{\mathbb{R}} \left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2 \mathbb{E}\left[e^{-2U_1}\right] d\xi \\
&\le \frac{C}{n}\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|^2,
\end{aligned}
$$

since $\mathbb{E}[e^{-2U_1}] < +\infty$ thanks to the Proposition 3.3.5.

We now set

$$\triangle(\xi) := \frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}.$$

Then we get

$$\begin{aligned}\mathbb{E}[\,\mathrm{I}\,] &\leq C\int_{\mathbb{R}} \mathbb{E}\left[\left|K_\ell^*(\xi)\widehat{D^*(\xi)}\triangle(\xi)\right|^2\right] d\xi \leq C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \mathbb{E}\left[\left|\widehat{D^*(\xi)}\right|^2\left|\triangle(\xi)\right|^2\right] d\xi \\ &\leq C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \mathbb{E}\left[\left|\widehat{D^*(\xi)} - \mathbb{E}\big[\widehat{D^*(\xi)}\big]\right|^2\left|\triangle(\xi)\right|^2\right] d\xi \\ &\qquad + C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \left|\mathbb{E}\big[\widehat{D^*(\xi)}\big]\right|^2 \mathbb{E}\big[|\triangle(\xi)|^2\big] d\xi \\ &:= \mathrm{III} + \mathrm{IV}.\end{aligned}$$

To control the term IV, we need the two following lemmas whose proofs are postponed in section 3.3.5.

Lemma 3.3.9. *Let*

$$\triangle(\xi) = \frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}.$$

Then there exists a positive constant C_p such that

$$\mathbb{E}\left[|\triangle(\xi)|^{2p}\right] \leq C_p \min\left\{\frac{1}{|M^*(\xi)|^{2p}}, \frac{n^{-p}}{|M^*(\xi)|^{4p}}\right\} \quad \textit{for } p = 1, 2. \tag{3.48}$$

Lemma 3.3.10. *The ratio $\left|\frac{D^*(\xi)}{M^*(\xi)}\right|$ is bounded:*

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{2R}{\alpha}\left(1 + \frac{\lambda + R}{2R}\right), \quad \forall \xi \in \mathbb{R}.$$

Since $\widehat{D^*}$ is an unbiased estimator of D^* using Lemma 3.3.9 we get

$$\mathrm{IV} \leq C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \left|D^*(\xi)\right|^2 \frac{n^{-1}}{|M^*(\xi)|^4} d\xi.$$

Then using Lemma 3.3.10 we get

$$\begin{aligned}\mathrm{IV} &\leq C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \frac{n^{-1}}{|M^*(\xi)|^2} d\xi \\ &\leq \frac{C}{n}\left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2.\end{aligned}$$

For the term III, we have by applying Cauchy-Schwarz's inequality and by Lemma 3.3.9:

$$\begin{aligned}
\mathrm{III} &\leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\left(\mathbb{E}\left[\left|\widehat{D^*(\xi)}-\mathbb{E}\left[\widehat{D^*(\xi)}\right]\right|^4\right]\right)^{1/2}\left(\mathbb{E}\left[\left|\triangle(\xi)\right|^4\right]\right)^{1/2}d\xi \\
&\leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\xi\right|^2\left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}-\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]\right|^2\right]\right)^{1/2} \\
&\qquad\qquad\times\min\left\{\frac{1}{|M^*(\xi)|^4},\frac{n^{-2}}{|M^*(\xi)|^8}\right\}^{1/2}d\xi \\
&\leq C\int_{\mathbb{R}}\frac{\left|K_\ell^*(\xi)\xi\right|^2}{|M^*(\xi)|^2}\left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j(\xi)\right|^4\right]\right)^{1/2}d\xi,
\end{aligned}$$

where $Z_j(\xi)=e^{(\mathbf{i}\xi-1)U_j}-\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]$. Since $Z_1(\xi),\dots,Z_n(\xi)$ are independent centered variables with

$$\mathbb{E}[|Z_1(\xi)|^4]\leq\mathbb{E}[|e^{(\mathbf{i}\xi-1)U_1}|^4]=\mathbb{E}[e^{-4U_1}]<+\infty,$$

by Proposition 3.3.5. Applying Rosenthal inequality to real and imaginary parts of complex variables Z_j's, we get

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j(\xi)\right|^4\right]\leq Cn^{-4}\left(n\mathbb{E}[|Z_1(\xi)|^4]+\left(n\mathbb{E}[|Z_1(\xi)|^2]\right)^2\right)\leq Cn^{-2}.$$

Hence

$$\mathrm{III}\leq\frac{C}{n}\int_{\mathbb{R}}\frac{\left|K_\ell^*(\xi)\xi\right|^2}{|M^*(\xi)|^2}d\xi=\frac{C}{n}\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|^2.$$

Finally, we obtain

$$\mathbb{E}\left[\|\hat{g}_\ell-g_\ell\|_2^2\right]\leq\|K_\ell\star g-g\|_2^2+\frac{C}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2+\left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right).$$

This ends the proof of Proposition 3.3.8.

□

Let us go back to the proof of Theorem 3.3.7. Using (3.47), due to well-known results on kernel density, the bias term converges to 0:

$$\lim_{n\to+\infty}\|K_\ell\star g-g\|_2^2=0,$$

and under the assumptions of the theorem we have for the variance term

$$\lim_{n\to+\infty}\frac{1}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2+\left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right)=0,$$

which completes the proof of Theorem 3.3.7.

□

3.3.5 Proof of technical lemmas

Proof of Lemma 3.3.9. This proof is inspired by the proof of Neumann [87]. We will prove the result with $p = 1$. For $p = 2$, the proof is similar.

We divide into two cases: $|M^*(\xi)| < 2n^{-1/2}$ and $|M^*(\xi)| \geq 2n^{-1/2}$. Recall that $\Omega = \left\{|\widehat{M^*(\xi)}| \geq n^{-1/2}\right\}$ and $\mathbb{E}\Big[\widehat{M^*(\xi)}\Big] = \mathbb{E}\Big[e^{\mathbf{i}\xi U_1}\Big] = M^*(\xi)$, we have:

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] = \mathbb{E}\left[\left|\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)]}\right|^2\right] = \mathbb{E}\left[\left|\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \left(\frac{\mathbb{1}_\Omega}{M^*(\xi)} + \frac{\mathbb{1}_{\Omega^c}}{M^*(\xi)}\right)\right|^2\right]$$
$$= \frac{\mathbb{P}(\Omega^c)}{|M^*(\xi)|^2} + \mathbb{E}\left[\mathbb{1}_\Omega \frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2}\right]. \tag{3.49}$$

<u>*i)* If $|M^*(\xi)| < 2n^{-1/2}$:</u>

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq \frac{1}{|M^*(\xi)|^2} + \frac{\mathbb{E}\left[|\widehat{M^*(\xi)} - M^*(\xi)|^2\right]n}{|M^*(\xi)|^2|}.$$

But

$$\mathbb{E}\left[\left|\widehat{M^*(\xi)} - M^*(\xi)\right|^2\right] = \mathbb{V}ar\left[\widehat{M^*(\xi)}\right] = \mathbb{V}ar\left[\frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j}\right]$$
$$\leq \frac{1}{n}\mathbb{V}ar\left(e^{\mathbf{i}\xi U_1}\right) \leq \frac{1}{n}\mathbb{E}\left[|e^{\mathbf{i}\xi U_1}|^2\right] = \frac{1}{n}.$$

Hence we obtain

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq \frac{C}{|M^*(\xi)|^2} \leq C\min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}, \tag{3.50}$$

since $|M^*(\xi)| < 2n^{-1/2}$.

<u>*ii)* If $|M^*(\xi)| \geq 2n^{-1/2}$:</u>

We first control the probability $\mathbb{P}(\Omega^c)$,

$$\mathbb{P}(\Omega^c) = \mathbb{P}\left(|\widehat{M^*(\xi)}| < n^{-1/2}\right) = \mathbb{P}\left(|\widehat{M^*(\xi)}| < |M^*(\xi)| - |M^*(\xi)| + n^{-1/2}\right)$$
$$\leq \mathbb{P}\left(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)| - n^{-1/2}\right)$$
$$\leq \mathbb{P}\left(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2\right). \tag{3.51}$$

Let $T_j(\xi) = e^{\mathbf{i}\xi U_j} - \mathbb{E}\Big[e^{\mathbf{i}\xi U_1}\Big]$, then

$$\widehat{M^*(\xi)} - M^*(\xi) = \frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j} - \mathbb{E}\Big[e^{\mathbf{i}\xi U_1}\Big] = \frac{1}{n}\sum_{j=1}^n T_j(\xi).$$

We have

$$|T_1(\xi)| = \left|e^{\mathbf{i}\xi U_j} - \mathbb{E}\left[e^{\mathbf{i}\xi U_1}\right]\right| \leq \left|e^{\mathbf{i}\xi U_j}\right| + \left|\mathbb{E}\left[e^{\mathbf{i}\xi U_1}\right]\right| \leq 2,$$

and

$$\mathbb{V}ar\Big(T_1(\xi)\Big) \leq \mathbb{E}\Big[|e^{\mathbf{i}\xi U_1}|^2\Big] = 1.$$

Thus, we get by Bernstein inequality (see Massart [78])

$$\begin{aligned}
\mathbb{P}\left(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2\right) &\leq 2\max\left\{\exp\left(-\frac{n|M^*(\xi)|^2}{16}\right), \exp\left(-\frac{n|M^*(\xi)|}{16}\right)\right\} \\
&\leq 2\exp\left(-\frac{n|M^*(\xi)|^2}{16}\right) \\
&\leq C\frac{n^{-1}}{|M^*(\xi)|^2}. \qquad (3.52)
\end{aligned}$$

We also have that

$$\begin{aligned}
\frac{1}{|\widehat{M^*(\xi)}|^2} = \frac{|M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2} &= \frac{|\widehat{M^*(\xi)} - \Big(\widehat{M^*(\xi)} - M^*(\xi)\Big)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2} \\
&\leq 2\left\{\frac{1}{|M^*(\xi)|^2} + \frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2}\right\}. \qquad (3.53)
\end{aligned}$$

Thus, from (3.49), (3.51) and (3.53) we have:

$$\begin{aligned}
\mathbb{E}\left[|\triangle(\xi)|^2\right] &\leq C\left\{\frac{n^{-1}}{|M^*(\xi)|^4} + \mathbb{E}\left[\mathbb{1}_\Omega\frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2||M^*(\xi)|^2}\right]\right\} \\
&\leq C\left\{\frac{n^{-1}}{|M^*(\xi)|^4} + \frac{\mathbb{E}\left[|\widehat{M^*(\xi)} - M^*(\xi)|^2\right]}{|M^*(\xi)|^4} + \frac{\mathbb{E}\left[|\widehat{M^*(\xi)} - M^*(\xi)|^4\right]n}{|M^*(\xi)|^4}\right\}. \qquad (3.54)
\end{aligned}$$

To find an upper bound for $\mathbb{E}\Big[|\widehat{M^*(\xi)} - M^*(\xi)|^4\Big]$, recall that $T_j(\xi) = e^{\mathbf{i}\xi U_j} - \mathbb{E}\Big[e^{\mathbf{i}\xi U_1}\Big]$. Since $T_1(\xi), \ldots, T_n(\xi)$ are independent centered variables with

$$\mathbb{E}[|T_1(\xi)|^4] \leq \mathbb{E}[|e^{\mathbf{i}\xi U_1}|^4] = 1.$$

Thus we get by Rosenthal's inequality applied to real and imaginary parts

$$\begin{aligned}
\mathbb{E}\Big[|\widehat{M^*(\xi)} - M^*(\xi)|^4\Big] &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n T_j(\xi)\right|^4\right] \\
&\leq Cn^{-4}\left(n\mathbb{E}[|T_1(\xi)|^4] + \Big(n\mathbb{E}[|T_1(\xi)|^2]\Big)^2\right) \leq Cn^{-2}
\end{aligned}$$

Thus, from (3.51) and (3.54) we get

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C\frac{n^{-1}}{|M^*(\xi)|^4}.$$

Furthermore

$$\frac{1}{|M^*(\xi)|^2} \geq \frac{n^{-1}}{|M^*(\xi)|^4},$$

since $|M^*(\xi)| > 2n^{-1/2}$. Hence

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C \min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}.$$

Combining the two cases, we obtain

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C \min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}.$$

This ends the proof of Lemma 3.3.9.

□

Proof of Lemma 3.3.10. From equation (3.34) we have

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{2R}{\alpha}\left(|g^*(\xi)| + \frac{\lambda + R}{2R}\right).$$

Using the change of variable $e^u = x$

$$\begin{aligned}|g^*(\xi)| &= \left|\int_{\mathbb{R}} e^{\mathbf{i}u\xi} g(u)du\right| = \left|\int_{\mathbb{R}} e^{\mathbf{i}u\xi} e^u h(e^u)du\right| = \left|\int_0^\infty e^{\mathbf{i}\xi \log x} h(x)dx\right| \\ &\leq \int_0^1 h(x)dx = 1,\end{aligned}$$

thus

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{2R}{\alpha}\left(1 + \frac{\lambda + R}{2R}\right),$$

which completes the proof.

□

3.4 Perspective

In this chapter, we study a microscopic process which describes evolution of cells and we study the large population limit that leads to obtain the PDE's approximation with a constant growth rate α. A study in the case of non-constant growth rate should be considered, then we expect to match the microscopic stochastic system with the following growth-fragmentation equation:

$$\partial_t n(t,x) + \alpha \partial_x\Big(g(x)n(t,x)\Big) + B(x)n(t,x) = 2\int_0^{+\infty} B(y)n(t,y)h\Big(\frac{x}{y}\Big)\frac{dy}{y}.$$

Assuming $B(x) = R$ and $g(x) = 1$, we have considered in Section 3.3 the problem of estimating the kernel h based on the eigenvalue problem corresponding to the PDE above and have only proved the consistency of the proposed estimator. Hence, we wish to establish rates of convergence and study optimality as well as adaptation for the estimator $\hat{g}$ then for the estimator $\hat{h}$. A natural approach would be the following: Starting from (3.33)

$$\alpha D(u) + (\lambda + R)M(u) = 2R\Big(M \star g)(u),$$

the first step would consist in estimating D with an estimate $\hat{D}$, then M with an estimate $\hat{M}$ and write (3.33) as

$$Y(u) = 2R(M \star g)(u) + W(u), \tag{3.55}$$

with observations

$$Y(u) := \alpha \hat{D}(u) + (\lambda + R)\hat{M}(u)$$

and "noise"

$$W(u) := \alpha(\hat{D}(u) - D(u)) + (\lambda + R)(\hat{M}(u) - M(u)).$$

By observing that (3.55) is quite similar to (1.3) studied by De Canditiis and Pensky [30], Johnstone [69], Kalifa and Mallat [70], Kerkyacharian et al. [71] and Pensky and Sapatinas [90], this problem could be dealt with by using tools proposed in these papers. However, two difficulties arise:

1. The convolution operator $g \mapsto M \star g$ is only known up to some errors (since M is estimated).

2. In the literature, minimax adaptation is studied under a smoothness condition on the unknown function and under an asymptotic condition on the noise with both conditions being independent. It's not the case here since the smoothness of g and the asymptotic behavior of W will both depend on smoothness of M (and D). See Proposition 3.3.4 for more details. Observe that assuming that M is smooth, say M belongs to a Sobolev space of parameter γ, implies some polynomial decay of its Fourier transform. In view of (3.34), since smoothnesses of M and D are closely related, this implies some intricate conditions on the function g, leading to a non-obvious control of the bias term $\|K_\ell \star g - g\|_2^2$ in inequality (3.47). In particular g cannot belong to a Sobolev class if asymptotically the ratio $\frac{D^*(\xi)}{M^*(\xi)}$ is not equivalent to $-(\lambda + R)/\alpha$. The study of adaptive minimax rates of convergence is a work in progress.

Finally, it is worth generalizing the problem of estimating h to the case where the division kernel and the growth rate are functions of size of cells.

4

Adaptive wavelet multivariate regression with errors in variables

Contents

This chapter is a version of the paper *Adaptive wavelet multivariate regression with errors in variables* (Chichignoud et al. [23]) written in collaboration with Michaël Chichignoud (Winton Capital Management), Thanh Mai Pham Ngoc (Université Paris Sud) and Vincent Rivoirard (Université Paris Dauphine), submitted for publication, in minor revision.

4.1 Introduction

We consider the problem of multivariate nonparametric regression with errors in variables. We observe the i.i.d dataset

$$(W_1, Y_1), \ldots, (W_n, Y_n)$$

where

$$Y_l = m(X_l) + \varepsilon_l$$

and

$$W_l = X_l + \delta_l,$$

with $Y_l \in \mathbb{R}$. The covariates errors δ_l are i.i.d unobservable random variables having error density g. We assume that g is known. The δ_l's are independent of the X_l's and Y_l's. The ε_l's are i.i.d standard normal random variables, independent of the X_l's with variance s^2 which is supposed to be known. We wish to estimate the regression function $m(x)$, $x \in [0,1]^d$, but direct observations of the covariates X_l are not available. Instead due to the measuring mechanism or the nature of the environment, the covariates X_l are measured with errors. Let us denote f_X the density of the X_l's assumed to be positive and f_W the density of the W_l's.

Use of errors-in-variables models appears in many areas of science such as medicine, econometry or astrostatistics and is appropriate in a lot of practical experimental problems. For instance, in epidemiologic studies where risk factors are partially observed (see Fan and Masry [46], Whittemore and Keller [109]) or in environmental science where air quality is measured with errors (Delaigle et al. [31]).

In the error-free case, that is $\delta_l = 0$, one retrieves the classical multivariate nonparametric regression problem. Estimating a function in a nonparametric way from data measured with error is not an easy problem. Indeed, constructing a consistent estimator in this context is challenging as we have to face to a deconvolution step in the estimation procedure. Deconvolution problems arise in many fields where data are obtained with measurement errors and has attracted a lot of attention in the statistical literature, see Meister [79] for an excellent source of references. The nonparametric regression with errors-in-variables model has been the object of a lot of attention as well, we may cite the works of Fan and Masry [46], Fan and Truong [47], Ioannides and Alevizos [64], Koo and Lee [72], Meister [79], Comte and Taupin [29], Chesneau [22], Du et al. [42], Caroll et al. [18], Delaigle et al. [31]. The literature has mainly to do with kernel-based approaches, based on the Fourier transform. All the works cited have tackled the univariate case except for Fan and Masry [46] where the authors explored the asymptotic normality for mixing processes. In the one dimensional setting, Chesneau [22] used Meyer wavelets in order to devise his statistical procedure but his assumptions on the model are strong since the corrupted observations W_l follow a uniform density on $[0,1]$. Comte and Taupin [29] investigated the mean integrated squared error with a penalized estimator based on projection methods upon Shannon basis. But the authors do not give any clue about how to choose the resolution level of the Shannon basis. Furthermore, the constants in the penalized term are calibrated via intense simulations.

In this chapter, our aim is to study the multidimensional setting and the pointwise risk. We would like to take into account the anisotropy for the function to estimate. Our approach relies on the use of projection kernels on wavelets bases combined with a deconvolution operator taking into account the noise in the covariates. When using wavelets, a crucial point lies in the choice of the resolution level. But it is well-known that theoretical results in adaptive estimation do not provide the way to choose the numerical constants in the resolution level and very often lead to conservative choices. We may cite the work of Gach et al. [51] which attempts to tackle this problem. For the density estimation problem and the sup-norm loss, the authors based their statistical procedure on Haar projection kernels and provide a way to choose locally the resolution level. Nonetheless, in practice, their procedure relies on heavy Monte Carlo simulations to calibrate the constants. In our paper the resolution level of our estimator is optimal and partially data-driven. It is automatically selected by a method inspired from Goldenshluger and Lepski [53] to tackle anisotropy problems. This method has been used recently in various contexts (see Doumic et al. [39], Comte and Lacour [27] and Bertin et al. [14]). Furthermore, we do not resort to thresholding which is very popular when using wavelets and our selection rule is adaptive to the unknown regularity of the regression function. We obtain oracle inequalities and provide optimal rates of convergence for anisotropic Hölder classes. The performances of our adaptive estimator, the negative impact of the errors in the covariates, the effects of the design density are assessed by examples based on simulations.

The chapter is organized as follows. In Section 4.2, we describe our estimation procedure. In Section 4.3, we provide an oracle inequality and rates of convergences of our estimator for the pointwise risk. Section 4.4 gives some numerical illustrations. Proofs of Theorems, propositions and technical lemmas are to be found in Section 4.5.

Notation Let $\mathbb{N} = \{0, 1, 2, \dots\}$ and $j = (j_1, \dots, j_d) \in \mathbb{N}^d$, we set $S_j = \sum_{i=1}^d j_i$ and for any $y \in \mathbb{R}^d$, we set, with a slight abuse of notation,

$$2^j y := (2^{j_1} y_1, \dots, 2^{j_d} y_d)$$

and for any $k = (k_1, \cdots, k_d) \in \mathbb{Z}^d$,

$$h_{j,k}(y) := 2^{\frac{S_j}{2}} h(2^j y - k) = 2^{\frac{S_j}{2}} h(2^{j_1} y_1 - k_1, \dots, 2^{j_d} y_d - k_d),$$

for any given function h. We denote by $\mathcal{F}$ the Fourier transform of any Lebesgue integrable function $f \in \mathbb{L}_1(\mathbb{R}^d)$ by

$$\mathcal{F}(f)(t) = \int_{\mathbb{R}^d} e^{-i<t,y>} f(y) dy, \quad t \in \mathbb{R}^d,$$

where $< ., . >$ denotes the usual scalar product.

For two integers a, b, we denote $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$. And $\lfloor y \rfloor$ denotes the largest integer smaller than y : $\lfloor y \rfloor \leq y < \lfloor y \rfloor + 1$.

4.2 The estimation procedure

For estimating the regression function m, the idea consists in writing m as the ratio

$$m(x) = \frac{m(x) f_X(x)}{f_X(x)}, \quad x \in [0, 1]^d.$$

In the sequel, we denote

$$p(x) := m(x) \times f_X(x).$$

First, we estimate p, then f_X. Since estimating f_X is a classical deconvolution problem, the main task consists in estimating p. We propose a wavelet-based procedure with an automatic choice of the maximal resolution level. Section 4.2.2 describes the construction of the projection kernel on wavelet bases depending on a maximal resolution level. Section 4.2.3 describes the Goldenshluger-Lepski procedure to select the resolution level adaptively.

4.2.1 Technical conditions

To facilitate the presentation, we collect in this subsection all the conditions that we need to ensure the existence of all quantities defined throughout the chapter.

First, some conditions are imposed on the regression function m and the design density f_X. We suppose that

$$m \in \mathcal{M}(\mathfrak{m}) = \{S : [0,1]^d \to \mathbb{R} : \|S\|_\infty \leq \mathfrak{m}\},\ \mathfrak{m} > 0, \tag{4.1}$$

and

$$f_X \in \mathcal{M}(\mathfrak{d}) = \{f \text{ density on } [0,1]^d \text{ and } \|f\|_\infty \leq \mathfrak{d}\},\ \mathfrak{d} > 0. \tag{4.2}$$

Futhermore, there exists $C_1 > 0$ such that for any $x \in [0,1]^d$, $f_X(x) \geq C_1$. We also suppose that $m \cdot f_X$ and $\mathcal{F}(m \cdot f_X) \in \mathbb{L}_1(\mathbb{R}^d)$.

To derive rates of convergence as we have to face a deconvolution step, we need some assumptions on the smoothness of the density of the errors covariates g. We suppose that

$$\mathcal{F}(g)(t) = \prod_{l=1}^{d} \mathcal{F}(g_l)(t_l),$$

and there exist positive constants c_g and C_g such that

$$c_g \prod_{l=1}^{d} (1+|t_l|)^{-\nu} \leq |\mathcal{F}(g)(t)| \leq C_g \prod_{l=1}^{d} (1+|t_l|)^{-\nu}, \quad \nu \geq 0, \quad t_l \in \mathbb{R}. \tag{4.3}$$

We require a supplementary condition on the derivative of the Fourier transform of g. There exists a positive constant $\mathcal{C}_g$ such that

$$|\mathcal{F}'(g)(t)| \leq \mathcal{C}_g \prod_{l=1}^{d} (1+|t_l|)^{-\nu-1}, \quad t_l \in \mathbb{R}. \tag{4.4}$$

Laplace and Gamma distributions satisfy the Assumptions (4.3) and (4.4) above. Assumptions (4.3) and (4.4) control the decay of the Fourier transform of each components of g at a polynomial rate controlled by the degree of ill-posedness ν. Hence we deal with a midly ill-posed inverse problem.

We consider a father wavelet φ on the real line satisfying the following conditions:

- (A1) The father wavelet φ is compactly supported on $[-A, A]$, where A is a positive integer.
- (A2) There exists a positive integer N, such that for any x

$$\int \sum_{k\in\mathbb{Z}} \varphi(x-k)\varphi(y-k)(y-x)^{\ell} dy = \delta_{0\ell}, \quad \ell = 0, \dots, N.$$

- (A3) φ is of class $\mathcal{C}^r$, where $r \geq \nu + 2$.

Conditions (A1), (A2) and (A3) are satisfied for instance by Coiflets wavelets (see Härdle et al. [56], chapter 8). Condition (A3) has already been encountered in the literature (see condition (A2) in Fan and Koo [45]). In particular, Condition (A3) is useful to prove Lemma 4.5.12.

4.2.2 Approximation kernels and family of estimators for p

The associated projection kernel on the space

$$V_j := \text{span}\{\varphi_{jk}, k \in \mathbb{Z}^d\}, \quad j \in \mathbb{N}^d,$$

is given for any x and y by

$$K_j(x,y) = \sum_k \varphi_{jk}(x)\varphi_{jk}(y),$$

where for any x,

$$\varphi_{jk}(x) = \prod_{l=1}^{d} 2^{\frac{j_l}{2}} \varphi(2^{j_l}x_l - k_l), \quad j \in \mathbb{N}^d,\ k \in \mathbb{Z}^d.$$

Therefore, the projection of p on V_j can be written for any z,

$$p_j(z) := K_j(p)(z) := \int K_j(z,y)p(y)dy = \sum_k p_{jk}\varphi_{jk}(z)$$

with

$$p_{jk} = \int p(y)\varphi_{jk}(y)dy.$$

First we estimate unbiasedly any projection p_j. Secondly to obtain the final estimate of p, it will remain to select a convenient value of j which will be done in Section 4.2.3. The natural approach is based on unbiased estimation of the projection coefficients p_{jk}. To do so, we adapt the kernel approach proposed by Fan and Truong [47] in our wavelets context. To this purpose, we set

$$\hat{p}_{jk} := \frac{1}{n}\sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u) = \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\frac{1}{n}\sum_{u=1}^{n} Y_u \int e^{-i<t,2^j W_u - k>} \prod_{l=1}^{d} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} dt_l,$$

$$\hat{p}_j(x) = \frac{1}{n}\sum_k \sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u)\varphi_{jk}(x),$$

where the deconvolution operator $\mathcal{D}_j$ is defined as follows for a function f defined on $\mathbb{R}$

$$(\mathcal{D}_j f)(w) = \frac{1}{(2\pi)^d}\int e^{-i<t,w>}\prod_{l=1}^{d}\frac{\overline{\mathcal{F}(f)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt, w \in \mathbb{R}^d. \tag{4.5}$$

Lemma 4.5.6, proved in Section 4.5.2 states that $\mathbb{E}[\hat{p}_j(x)] = p_j(x)$ which justifies our approach. Furthermore, the deconvolution operator $(\mathcal{D}_j f)(w)$ in (4.5) is the multidimensional wavelet analogous of the operator $K_n(x)$ defined in (2.4) in Fan and Truong [47]: the Fourier transform of their kernel K has been replaced in our procedure by the Fourier transform of the wavelet φ_{jk} and their bandwith h by 2^{-j}. Eventually, our estimator is well-defined. Indeed, using Lemma 4.5.11 and assumption (4.3) we have that

$$\prod_{l=1}^{d}\left|\frac{\mathcal{F}(\varphi)(t_l)}{\mathcal{F}(g_l)(2^{j_l}t_l)}\right| \le C\prod_{l=1}^{d}(1+|t_l|)^{-r}(1+|2^{j_l}t_l|)^{\nu} \le C2^{S_j\nu}\prod_{l=1}^{d}(1+|t_l|)^{\nu-r},$$

which is integrable using condition (A3).

Note that the definition of the estimator $\hat{p}_j(x)$ still makes sense when we do not have any noise on the variables X_l i.e $g(x) = \delta_0(x)$ because in this case $\mathcal{F}(g)(t) = 1$.

4.2.3 Selection rule by using the Goldenshluger-Lepski methodology

The second and final step consists in selecting the multidimensional resolution level j depending on x and based on a data-driven selection rule inspired from a method exposed in Goldenshluger and Lepski [53]. To define this latter we have to introduce some quantities. In the sequel we denote for any $w \in \mathbb{R}^d$,

$$T_j(w) := \sum_k (\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x)$$

and

$$U_j(y,w) := y\sum_k (\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x) = y \times T_j(w),$$

so we have

$$\hat{p}_j(x) = \frac{1}{n}\sum_{u=1}^{n} U_j(Y_u, W_u).$$

Proposition 4.5.1 in Section 4.5.2 shows that $\hat{p}_j(x)$ concentrates around $p_j(x)$. So the idea is to find a maximal resolution $\hat{j}$ that mimics the oracle index. The oracle index minimizes a bias variance trade-off. So we need an estimation of the bias-variance decomposition of $\hat{p}_j(x)$. We denote $\sigma_j^2 := \mathbb{V}ar(U_j(Y_1, W_1))$ and the variance of $\hat{p}_j$ is thus equal to $\frac{\sigma_j^2}{n}$. We set :

$$\hat{\sigma}_j^2 := \frac{1}{n(n-1)}\sum_{l=2}^{n}\sum_{v=1}^{l-1}(U_j(Y_l, W_l) - U_j(Y_v, W_v))^2, \tag{4.6}$$

and since $\mathbb{E}(\hat{\sigma}_j^2) = \sigma_j^2$, $\hat{\sigma}_j^2$ is a natural estimator of σ_j^2. To devise our procedure, we introduce a slightly overestimate of σ_j^2 given by:

$$\tilde{\sigma}_{j,\tilde{\gamma}}^2 := \hat{\sigma}_j^2 + 2C_j\sqrt{2\tilde{\gamma}\hat{\sigma}_j^2\frac{\log n}{n}} + 8\tilde{\gamma}C_j^2\frac{\log n}{n}, \tag{4.7}$$

where $\tilde{\gamma}$ is a positive constant and

$$C_j := \left(\mathfrak{m} + s\sqrt{2\tilde{\gamma}\log n}\right)\|T_j\|_\infty.$$

For any $\varepsilon > 0$, let $\gamma > 0$ and

$$\Gamma_\gamma(j) := \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2\log n}{n}} + \frac{c_j\gamma\log n}{n},$$

where

$$c_j := 16\,(2\mathfrak{m} + s)\,\|T_j\|_\infty.$$

Let

$$\Gamma_\gamma(j,j') := \Gamma_\gamma(j) + \Gamma_\gamma(j\wedge j'),$$

and

$$\Gamma_\gamma^*(j) := \sup_{j'}\Gamma_\gamma(j,j'). \tag{4.8}$$

We now define the selection rule for the resolution index. Let

$$\hat{R}_j := \sup_{j'}\left\{\left|\hat{p}_{j\wedge j'}(x) - \hat{p}_{j'}(x)\right| - \Gamma_\gamma(j',j)\right\}_+ + \Gamma_\gamma^*(j). \tag{4.9}$$

Then $\hat{p}_{\hat{j}}(x)$ is the final estimator of $p(x)$ with $\hat{j}$ such that

$$\hat{j} := \arg\min_{j\in J}\hat{R}_j, \tag{4.10}$$

where the set J is defined as

$$J := \left\{j \in \mathbb{N}^d : \quad 2^{S_j} \le \left\lfloor\frac{n}{\log^2 n}\right\rfloor\right\}. \tag{4.11}$$

Now, we shall highlight how the above quantities interplay in the estimation of the risk decomposition of $\hat{p}_j$. An inspection of the proof of Theorem 4.3.1 shows that a control of the bias of $\hat{p}_j$ is provided by :

$$\sup_{j'}\left\{\left|\hat{p}_{j\wedge j'}(x) - \hat{p}_{j'}(x)\right| - \Gamma_\gamma(j',j)\right\}_+.$$

The term $|\hat{p}_{j\wedge j'}(x) - \hat{p}_{j'}|$ is classical when using the Goldenshluger Lepski method (see Sections 2.1 and 5.2 in Bertin et al. [14]). Furthermore for technical reasons (see proof of Theorem 4.3.1), we do not estimate the variance of $\hat{p}_j(x)$ by $\frac{\hat{\sigma}_j^2}{n}$ but we replace it by $\Gamma_\gamma^2(j)$. Note that we have the straightforward control

$$\Gamma_\gamma(j) \le C\left(\hat{\sigma}_j\sqrt{\frac{\log n}{n}} + (C_j + c_j)\frac{\log n}{n}\right),$$

where C is a constant depending on ε, $\tilde{\gamma}$ and γ. Actually we prove that $\Gamma^2_\gamma(j)$ is of order $\frac{\log n}{n}\sigma_j^2$ (see Lemma 4.5.9 and 4.5.13). The dependence of $\tilde{\sigma}^2_{j,\tilde{\gamma}}$ (4.7) in $\mathfrak{m}$ appears only in smaller order terms. In conclusion, up to the knowledge of $\mathfrak{m}$ the procedure is completely data-driven. Next section explains how to choose the constants γ and $\tilde{\gamma}$. Our approach is non asymptotic and based on sharp concentration inequalities.

4.3 Rates of convergence

4.3.1 Oracle inequality and rates of convergence for $p(\cdot)$

First, we state an oracle inequality which highlights the bias-variance decomposition of the risk.

Theorem 4.3.1. *Let $q \geq 1$ be fixed and let $\hat{j}$ be the adaptive index defined as above. Then, it holds for any $\gamma > q(\nu+1)$ and $\tilde{\gamma} > 2q(\nu+2)$,*

$$\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q\right] \leq R_1\left(\inf_\eta \mathbb{E}\left[\left\{B(\eta) + \Gamma^*_\gamma(\eta)\right\}^q\right]\right) + R_1' n^{-q},$$

where

$$B(\eta) := \max\left(\sup_{j'}\left|\mathbb{E}\left[\hat{p}_{\eta\wedge j'}(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]\right|, |\mathbb{E}[\hat{p}_\eta(x)] - p(x)|\right)$$

R_1 a constant depending only on q and R_1' a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$ and φ.

The oracle inequality in Theorem 4.3.1 illustrates a bias-variance decomposition of the risk. The term $B(\eta)$ is a bias term. Indeed, one recognizes on the r.h.s the classical bias term

$$|\mathbb{E}[\hat{p}_\eta(x)] - p(x)| = |p_\eta(x) - p(x)|.$$

Concerning $\left|\mathbb{E}\left[\hat{p}_{\eta\wedge j'}(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]\right|$, for sake of clarity let us consider for instance the univariate case : if $j' \leq \eta$ this term is equal to zero. If $j' \geq \eta$, it turns to be

$$|\mathbb{E}\left[\hat{p}_\eta(x)\right] - \mathbb{E}\left[\hat{p}_{j'}(x)\right]| = |p_\eta(x) - p_{j'}(x)| \leq |p_\eta(x) - p(x)| + |p_{j'}(x) - p(x)|.$$

As we have the following inclusion for the projection spaces $V_\eta \subset V_{j'}$, the term $p_{j'}$ is closer to p than p_η for the $\mathbb{L}_2$-distance. Hence we expect a good control of $|p_{j'}(x) - p(x)|$ with respect to $|p_\eta(x) - p(x)|$.

We study the rates of convergence of the estimators over anisotropic Hölder Classes. Let us define them.

Definition 4.3.2 (Anisotropic Hölder Space)**.** Let $\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_d) \in (\mathbb{R}^*_+)^d$ and $L > 0$. We say that $f : [0,1]^d \to \mathbb{R}$ belongs to the anisotropic Hölder class $\mathbb{H}_d(\vec{\beta}, L)$ of functions if f is bounded and for any $l = 1, ..., d$ and for all $z \in \mathbb{R}$

$$\sup_{x\in[0,1]^d}\left|\frac{\partial^{\lfloor\beta_l\rfloor} f}{\partial x_l^{\lfloor\beta_l\rfloor}}(x_1, \ldots, x_l + z, \ldots, x_d) - \frac{\partial^{\lfloor\beta_l\rfloor} f}{\partial x_l^{\lfloor\beta_l\rfloor}}(x_1, \ldots, x_l, \ldots, x_d)\right| \leq L|z|^{\beta_l - \lfloor\beta_l\rfloor}.$$

The following theorem gives the rate of convergence of the estimator $\hat{p}_{\hat{j}}(x)$ over anisotropic Hölder space.

Theorem 4.3.3. *Let $q \geq 1$ be fixed and let $\hat{j}$ be the adaptive index defined in (4.10). Then, if for any l, $\lfloor \beta_l \rfloor \leq N$ and $L > 0$, it holds*

$$\sup_{p \in \mathbb{H}_d(\vec{\beta},L)} \mathbb{E}\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_2 \left(\frac{\log n}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with $\bar{\beta} = \frac{1}{\frac{1}{\beta_1}+\cdots+\frac{1}{\beta_d}}$ and R_2 a constant depending on $\gamma, q, \tilde{\gamma}, \mathfrak{m}, \mathfrak{d}, s, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

4.3.2 Rates of convergence for $m(\cdot)$

As mentioned above, the estimation of m requires an adaptive estimate of f_X. This is due to kernel estimators, e.g. projection estimators do not need an additional estimate (see Bertin et al. [14]). For this purpose, we use an estimate introduced by Comte and Lacour [27] (Section 3.4) denoted by $\hat{f}_X$. This estimate is constructed from a deconvolution kernel and the bandwidth is selected via a method described in Goldenshluger and Lepski [53]. We will not give the explicit expression of $\hat{f}_X$ for ease of exposition. Then, we define the estimate of m for all x in $[0,1]^d$:

$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \tag{4.12}$$

The term $n^{-1/2}$ is added to avoid the drawback when $\hat{f}_X$ is closed to 0.

Theorem 4.3.4. *Let $q \geq 1$ be fixed and let $\hat{m}$ defined as above. Then, if for any l, $\lfloor \beta_l \rfloor \leq N$ and $L > 0$, it holds*

$$\sup_{(m,f_X) \in \mathbb{H}_d(\vec{\beta},L)\times\mathbb{H}_d(\vec{\beta},L)} \mathbb{E}\left|\hat{m}(x) - m(x)\right|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_3 \left(\frac{\log n}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_3 a constant depending on $\gamma, q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$.

Next Theorem gives a lower bound for the pointwise risk:

Theorem 4.3.5. *Let $q \geq 1$, $L > 0$ and for any l, $\lfloor \beta_l \rfloor \leq N$. Then for any estimator $\tilde{m}$ of m and for n large enough we have*

$$\sup_{(m,f_X) \in \mathbb{H}_d(\vec{\beta},L)\times\mathbb{H}_d(\vec{\beta},L)} \mathbb{E}\left|\tilde{m}(x) - m(x)\right|^q \geq R_4 n^{-q\bar{\beta}/(2\bar{\beta}+2\nu+1)},$$

with R_4 a positive constant depending on $\vec{\beta}$, L, s, C_g and $\mathcal{C}_g$.

Consequently, the estimate $\hat{m}$ achieves the optimal rate of convergence up to a logarithmic term.

4.4 Numerical results

In this section, we implement some simulations to illustrate the theoretical results. We aim at estimating the Doppler regression function m at point $x_0 = 0.25$ (see Figure 4.1). We have $n = 1024$ observations and the regression errors ε_l's follow a standard normal density with variance $s^2 = 0.15^2$. As for the design density of the X_l's, we consider Beta densities and the uniform density on $[0, 1]$. The uniform distribution is quite classical in regression with random design. The $Beta(2, 2)$ and $Beta(0.5, 2)$ distributions reflect two different behaviors on $[0, 1]$. Indeed, we recall that the Beta density with parameters (a, b) (denoted here by $Beta(a, b)$) is proportional to $x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$. In Figure 4.2, we plot the noisy regression Doppler function according to the three design scenario. For the covariate errors δ_i's, we focus on the centered Laplace density with scale parameter $\sigma_{g_L} > 0$ that we denote g_L. This latter has the following expression :

$$g_L(x) = \frac{1}{2\sigma_{g_L}} e^{-\frac{|x|}{\sigma_{g_L}}}.$$

The choice of the centered Laplace noise is motivated by the fact that the Fourier transform of g_L is given by

$$\mathcal{F}(g_L)(t) = \frac{1}{1 + \sigma_{g_L}^2 t^2},$$

and according to Assumption (4.3), it gives an example of an ordinary smooth noise with degree of ill-posedness $\nu = 2$. Furthermore, when facing regression problems with errors in the design, it is common to compute the so-called reliability ratio (see Fan and Truong [47]) which is given by

$$R_r := \frac{\mathbb{V}ar(X)}{\mathbb{V}ar(X) + 2\sigma_{g_L}^2}.$$

R_r permits to assess the amount of noise in the covariates. The closer to 0 R_r is, the bigger the amount of noise in the covariates is and the more difficult the deconvolution step will be. For instance, Fan and Truong [47] chose $R_r = 0.70$. We computed the reliability ratio in Table 4.1 for the considered simulations.

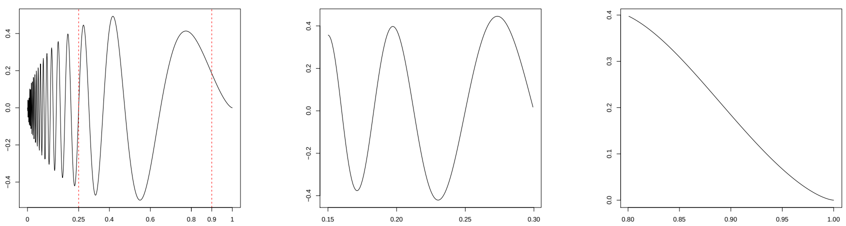
σ_{g_L}	design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.88	0.81	0.80
0.10	0.80	0.71	0.69

Table 4.1: Reliability ratio.

We recall that our estimator of $m(x)$ is given by the ratio of two estimators (see (4.12)) :

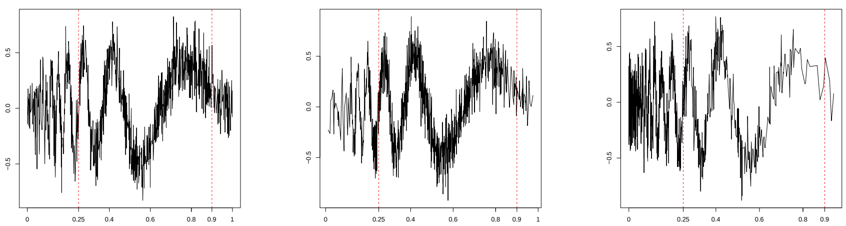
$$\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x) \vee n^{-1/2}}. \tag{4.13}$$

First, we compute $\hat{p}_{\hat{j}}(x)$ an estimator of $p(x) = m(x) \times f_X(x)$ which is denoted "GL" in the graphics below. We use coiflet wavelets of order 5. Then we divide $\hat{p}_{\hat{j}}(x)$ by the adaptive deconvolution density estimator $\hat{f}_X(x)$ of Comte and Lacour [27]. This latter is constructed with a deconvolution kernel and an adaptive bandwidth. For the selection of the coiflet level $\hat{j}$ in $\hat{p}_{\hat{j}}(x)$, we advise to use $\hat{\sigma}_j^2$ instead of $\tilde{\sigma}_{j,\tilde{\gamma}}^2$ and $\frac{2\max_i |Y_i| \|T_j\|_\infty}{3}$ instead of c_j. It remains to settle the value of the constant γ. To do so, we compute the pointwise risk of $\hat{p}_{\hat{j}}(x)$ in function of γ: Figure 4.3 shows a clear "dimension jump" and accordingly the value $\gamma = 0.5$ turns to be reasonable. Hence we fix $\gamma = 0.5$ for all simulations and our selection rule is completely data-driven.



(a) (b) (c)

Figure 4.1: a/ Representation of Doppler function. b/ A zoom of Doppler function on $[0.15, 0.30]$. c/ A zoom of Doppler function on $[0.80, 1]$.



(a) (b) (c)

Figure 4.2: a/ Noisy Doppler with $X_i \sim \mathcal{U}[0,1]$. b/ Noisy Doppler with $X_i \sim Beta(2,2)$. c/ Noisy Doppler function with $X_i \sim Beta(0.5, 2)$.

σ_{g_L}	design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.0144	0.0204	0.0071
0.10	0.0156	0.0206	0.0072

Table 4.2: MAE of $\hat{m}(x)$ at $x_0 = 0.25$

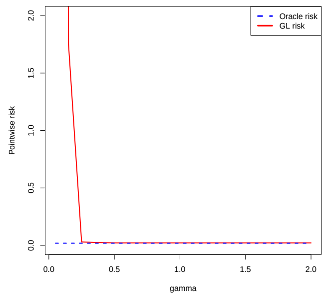


Figure 4.3: Pointwise risk of $\hat{p}_{\hat{j}}$ at $x_0 = 0.25$ in function of parameter γ for the $Beta(2,2)$ design and $\sigma_{g_L} = 0.075$.

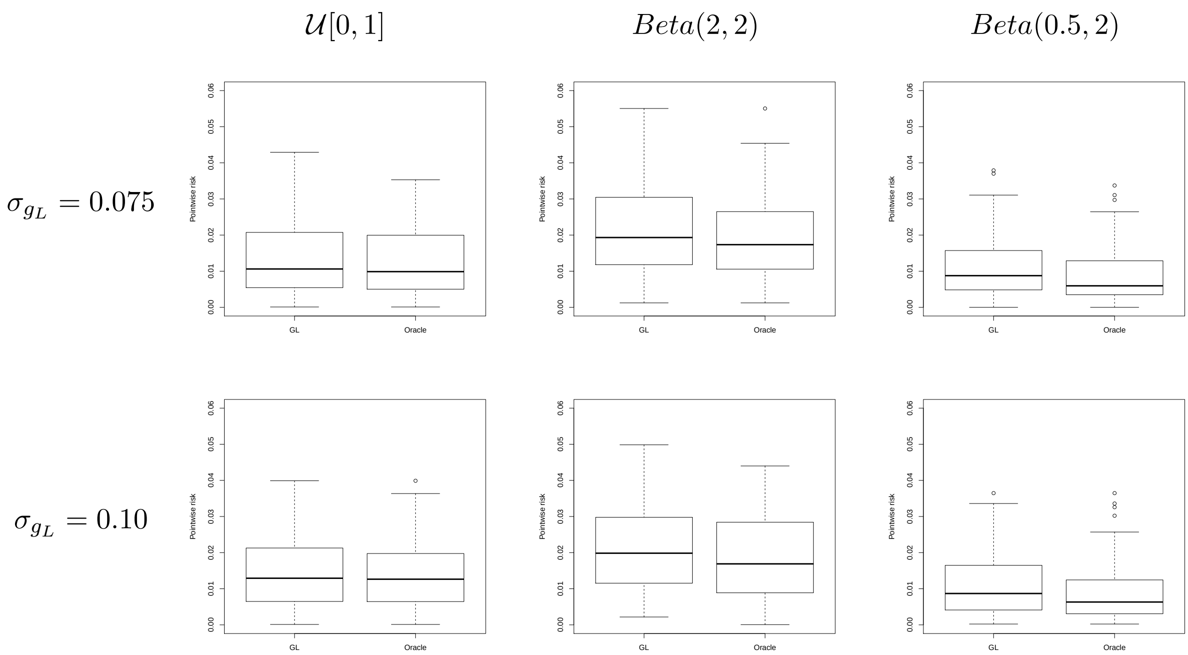


Figure 4.4: Estimation of $p(x)$ at $x_0 = 0.25$

Boxplots in Figure 4.4 summarize our numerical experiments. Theorem 4.3.1 gives an oracle inequality for the estimation of $p(x)$. We compare the pointwise risk error of $\hat{p}_{\hat{j}}(x)$ (computed with 100 Monte Carlo repetitions) with the oracle risk one. The oracle is $\hat{p}_{j_{oracle}}$ with the index j_{oracle} defined as follows:

$$j_{oracle} := \arg\min_{j \in J} |\hat{p}_j(x) - p(x)|.$$

In Table 4.2, we have computed the MAE (Mean Absolute Error) of $\hat{m}(x)$ over 100 Monte Carlo runs.

σ_{g_L}	design of the X_i			σ_{g_L}	design of the X_i		
	$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$		$\mathcal{U}[0,1]$	$Beta(2,2)$	$Beta(0.5,2)$
0.075	0.3461	0.5312	0.3445	0.075	0.2153	0.3429	0.5130
0.10	0.3668	0.5493	0.3589	0.10	0.2191	0.3453	0.5293

Table 4.3: MAE of $\hat{m}(x)$ at the points very closed to 0 and 1: on the left: $x_0 = 0.01$ and on the right: $x_0 = 0.98$.

We shall make a remark on the choice of Beta distributions. Beta densities are bounded from below on any compact included in $[0,1]$. One can see in Table 4.3 that the performances are very bad at points very closed to 0 and 1, in particular $x_0 = 0.01$ and $x_0 = 0.98$.

Our performances are close to those of the oracle (see Figure 4.4) and are quite satisfying at $x_0 = 0.25$. When going deeper into details, increasing the Laplace noise parameter σ_{g_L} deteriorates sligthly the performances. Hence it seems that our procedure is robust to the noise in the covariates and accordingly to the deconvolution step. Concerning the role of the design density, when considering the $Beta(0.5,2)$ distribution, we expect the performances to be better not too far from 0 as the observations tend to concentrate near 0. Indeed, this phenomenon is confirmed by Table 4.2.

4.5 Proofs

4.5.1 Proofs of theorems

This section is devoted to the proofs of theorems. These proofs use some propositions and technical lemmas which are respectively in Section 4.5.2 and 4.5.2. In the sequel, C is a constant which may vary from one line to another one.

Proof of Theorem 4.3.1.

We firstly recall the basic inequality $(a_1 + \cdots + a_p)^q \leq p^{q-1}(a_1^q + \cdots + a_p^q)$ for all $a_1, \ldots, a_p \in \mathbb{R}_+^p$, $p \in \mathbb{N}$ and $q \geq 1$. For ease of exposition, we denote $\hat{p}_{\hat{j}}(x) = \hat{p}_{\hat{j}}$. So, we can show for any $\eta \in \mathbb{N}^d$:

$$\begin{aligned}
&\left|\hat{p}_{\hat{j}} - p(x)\right| \leq \left|\hat{p}_{\hat{j}} - \hat{p}_{\hat{j}\wedge\eta}\right| + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma_\gamma(\hat{j},\eta) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| - \Gamma_\gamma(\eta,\hat{j}) + \Gamma_\gamma(\eta,\hat{j}) + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma_\gamma(\eta,\hat{j}) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| - \Gamma_\gamma(\eta,\hat{j}) + \Gamma_\gamma(\hat{j},\eta) + |\hat{p}_\eta - p(x)| \\
&\leq \left|\hat{p}_{\eta\wedge\hat{j}} - \hat{p}_{\hat{j}}\right| - \Gamma_\gamma(\hat{j},\eta) + \Gamma^*_\gamma(\eta) + \left|\hat{p}_{\hat{j}\wedge\eta} - \hat{p}_\eta\right| - \Gamma_\gamma(\eta,\hat{j}) + \Gamma^*_\gamma(\hat{j}) + |\hat{p}_\eta - p(x)| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\hat{p}_\eta - p(x)| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + |\hat{p}_\eta - \mathbb{E}[\hat{p}_\eta]| \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + |\hat{p}_\eta - \mathbb{E}[\hat{p}_\eta]| - \Gamma_\gamma(\eta) + \Gamma_\gamma(\eta) \\
&\leq \hat{R}_\eta + \hat{R}_{\hat{j}} + |\mathbb{E}[\hat{p}_\eta] - p(x)| + \sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]\right| - \Gamma_\gamma(j') \right\}_+ + \Gamma^*_\gamma(\eta).
\end{aligned}$$

By definition of $\hat{j}$, we recall that $\hat{R}_{\hat{j}} \leq \inf_\eta \hat{R}_\eta$ and

$$\hat{R}_\eta \leq \sup_{j,j'} \left\{ \left|\hat{p}_{j\wedge j'} - \mathbb{E}[\hat{p}_{j\wedge j'}]\right| - \Gamma_\gamma(j\wedge j') \right\}_+ + \sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]\right| - \Gamma_\gamma(j') \right\}_+ + \sup_{j'} \left|\mathbb{E}[\hat{p}_{\eta\wedge j'}] - \mathbb{E}[\hat{p}_{j'}]\right| + \Gamma^*_\gamma(\eta).$$

Hence

$$\begin{aligned}\left|\hat{p}_{\hat{j}} - p(x)\right| \leq & 2\Bigg[\sup_{j,j'} \left\{ \left|\hat{p}_{j\wedge j'} - \mathbb{E}[\hat{p}_{j\wedge j'}]\right| - \Gamma_\gamma(j\wedge j') \right\}_+ \\ & + \sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]\right| - \Gamma_\gamma(j') \right\}_+ + \sup_{j'} \left|\mathbb{E}[\hat{p}_{\eta\wedge j'}] - \mathbb{E}[\hat{p}_{j'}]\right| \Bigg] \\ & + 2\Gamma^*_\gamma(\eta) + \left|\mathbb{E}[\hat{p}_\eta] - p(x)\right| + \sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]\right| - \Gamma_\gamma(j') \right\}_+ + \Gamma^*_\gamma(\eta).\end{aligned}$$

By definition of $B(\eta) = \max\left(\sup_{j'}\left|\mathbb{E}\hat{p}_{\eta\wedge j'} - \mathbb{E}\hat{p}_{j'}\right|, \left|\mathbb{E}\hat{p}_\eta - p(x)\right|\right)$, we get

$$\begin{aligned}\left|\hat{p}_{\hat{j}} - p(x)\right| \leq & 2\sup_{j,j'} \left\{ \left|\hat{p}_{j\wedge j'} - \mathbb{E}[\hat{p}_{j\wedge j'}]\right| - \Gamma_\gamma(j\wedge j') \right\}_+ \\ & + 3\sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}[\hat{p}_{j'}]\right| - \Gamma_\gamma(j') \right\}_+ + 3B(\eta) + 3\Gamma^*_\gamma(\eta).\end{aligned}$$

Consequently

$$\begin{aligned}\left|\hat{p}_{\hat{j}} - p(x)\right|^q \leq & 3^{2q-1}\Bigg(\left[B(\eta) + \Gamma^*_\gamma(\eta)\right]^q + \sup_{j'} \left\{ \left|\hat{p}_{j'} - \mathbb{E}\hat{p}_{j'}\right| - \Gamma_\gamma(j') \right\}_+^q \\ & + \sup_{j,j'} \left\{ \left|\hat{p}_{j\wedge j'} - \mathbb{E}\hat{p}_{j\wedge j'}\right| - \Gamma_\gamma(j\wedge j') \right\}_+^q\Bigg).\end{aligned}$$

Using Proposition 4.5.4, we have

$$\mathbb{E}\left|\hat{p}_{\hat{j}} - p(x)\right|^q \leq C\left(\mathbb{E}\left[\left(B(\eta) + \Gamma^*_\gamma(\eta)\right)^q\right]\right) + R_1' n^{-q}.$$

Then, we get

$$\mathbb{E}\left|\hat{p}_{\hat{j}} - p(x)\right|^q \leq R_1\left(\inf_\eta \mathbb{E}\left[\left(B(\eta) + \Gamma^*_\gamma(\eta)\right)^q\right]\right) + R_1' n^{-q},$$

where R_1 is a constant only depending on q and R_1' a constant depending on $\mathfrak{m}$, $\mathfrak{d}$, s, φ, c_g, $\mathcal{C}_g$.

□

Proof of Theorem 4.3.3.

The proof is a direct application of Theorem 4.3.1 together with a standard bias-variance trade-off. We first recall the assertion of this theorem:

$$\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x)-p(x)\right|^q\right] \leq C\left(\inf_{\eta}\mathbb{E}\left[\left(B(\eta)+\Gamma^*_\gamma(\eta)\right)^q\right]\right)+R'_1 n^{-q}.$$

For the bias term, we use Proposition 4.5.5 to get:

$$B(\eta) \leq CL\sum_{l=1}^{d} 2^{-\eta_l\beta_l}, \text{for all } \eta \in J.$$

Now let us focus on $\mathbb{E}\left[\Gamma^*_\gamma(\eta)^q\right]$. We have

$$\begin{aligned}
\mathbb{E}\left[\Gamma_\gamma(\eta)^q\right] &= \mathbb{E}\left[\left(\sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{\eta,\tilde{\gamma}}\log n}{n}}+\frac{c_\eta\gamma\log n}{n}\right)^q\right]\\
&\leq 2^{q-1}\left(\left(\frac{2\gamma(1+\varepsilon)\log n}{n}\right)^{\frac{q}{2}}\mathbb{E}[\tilde{\sigma}^q_{\eta,\tilde{\gamma}}]+\left(\frac{c_\eta\gamma\log n}{n}\right)^q\right)\\
&\leq C\left(\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}}+\left(\frac{c_\eta\log n}{n}\right)^q\right),
\end{aligned}$$

using Lemma 4.5.9. But

$$c_\eta = 16\,(2\mathfrak{m}+s)\,\|T_\eta\|_\infty \leq C2^{S_\eta\nu+S_\eta},$$

using Lemma 4.5.13. Hence

$$\mathbb{E}\left[\Gamma_\gamma(\eta)^q\right] \leq C\left(\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}}+\left(\frac{\log n}{n}\right)^q 2^{(S_\eta\nu+S_\eta)q}\right).$$

We have

$$\left(\frac{\log n}{n}\right)^{\frac{q}{2}}2^{(2S_\eta\nu+S_\eta)\frac{q}{2}} \geq \left(\frac{\log n}{n}\right)^q 2^{(S_\eta\nu+S_\eta)q} \Longleftrightarrow 2^{S_\eta} \leq \frac{n}{\log n},$$

which is true since by (4.11), $2^{S_\eta} \leq \frac{n}{\log^2 n}$.

This yields

$$\mathbb{E}[\Gamma^*_\gamma(\eta)^q] \leq C\left(\frac{2^{(2S_\eta\nu+S_\eta)}\log n}{n}\right)^{\frac{q}{2}}.$$

Eventually, we obtain the bound for the pointwise risk:

$$\mathbb{E}\left|\hat{p}_{\hat{j}}(x)-p(x)\right|^q \leq C\left(\inf_\eta\left\{L\sum_{l=1}^{d}2^{-\eta_l\beta_l}+\sqrt{\frac{2^{(2S_\eta\nu+S_\eta)}\log(n)}{n}}\right\}^q\right)+R'_1n^{-q}.$$

Setting the gradient of the right hand side of the inequality above with respect to η it turns out that the optimal η_l is proportional to $\frac{2}{\log 2}\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+2\nu+1)}(\log L + \frac{1}{2}\log(\frac{n}{\log(n)}))$, which leads for n large enough to

$$\mathbb{E}\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_2 \left(\frac{\log(n)}{n}\right)^{\frac{\bar{\beta}q}{2\bar{\beta}+2\nu+1}},$$

with R_2 a constant depending on $\gamma, q, \varepsilon, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$. The proof of Theorem 4.3.3 is completed.

□

Proof of Theorem 4.3.4.

We recall that $m(x) = \frac{p(x)}{f_X(x)}$ and $\hat{m}(x) = \frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x)\vee n^{-1/2}}$. We now state the main properties of the adaptive estimate $\hat{f}_X$ showed by Comte and Lacour [27] (Theorem 2): for all $q \geq 1$, all $\vec{\beta} \in (0,1]^d$, all $L > 0$ and n large enough, it holds

$$\mathbb{P}(E_1) := \mathbb{P}\left(|\hat{f}_X(x) - f_X(x)| \geq C\phi_n(\vec{\beta})\right) \leq n^{-2q}, \tag{4.14}$$

and

$$\mathbb{P}\left(|\hat{f}_X(x) - f_X(x)| \leq Cn\right) = 1, \tag{4.15}$$

where $\phi_n(\vec{\beta}) := (\log(n)/n)^{\bar{\beta}/(2\bar{\beta}+2\nu+1)}$. Although the construction of the estimate $\hat{f}_X(x)$ depends on q, we remove the dependency for ease of exposition (see Comte and Lacour [27] Section 3.4 for further details). From (4.14), we easily deduce, since $f_X(x) \geq C_1 > 0$, for n large enough that

$$\mathbb{P}(E_2) := \mathbb{P}\left(\hat{f}_X(x) < \frac{C_1}{2}\right) \leq n^{-2q}. \tag{4.16}$$

We now start the proof of the theorem. We have together with (4.15)

$$\begin{aligned}
|\hat{m}(x) - m(x)| &= \left|\frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x)\vee n^{-1/2}} - \frac{p(x)}{f_X(x)}\right| \\
&\leq \left|\frac{\hat{p}_{\hat{j}}(x)}{\hat{f}_X(x)\vee n^{-1/2}} - \frac{p(x)}{\hat{f}_X(x)\vee n^{-1/2}}\right| + \left|\frac{p(x)}{\hat{f}_X(x)\vee n^{-1/2}} - \frac{p(x)}{f_X(x)}\right| \\
&\leq \left|\frac{\hat{p}_{\hat{j}}(x) - p(x)}{\hat{f}_X(x)\vee n^{-1/2}}\right| + \|m\|_\infty \|f_X\|_\infty \left|\frac{(\hat{f}_X(x)\vee n^{-1/2}) - f_X(x)}{f_X(x)(\hat{f}_X(x)\vee n^{-1/2})}\right| \\
&:= \mathcal{A}_1 + \|m\|_\infty \|f_X\|_\infty \mathcal{A}_2.
\end{aligned}$$

Control of $\mathbb{E}[\mathcal{A}_1^q]$. Using Cauchy-Schwarz inequality and the inequality $\hat{f}_X(x)\vee n^{-1/2} \geq n^{-1/2}$, we obtain for n large enough

$$\begin{aligned}
\mathbb{E}[\mathcal{A}_1^q] &= \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2^c}] + \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2}] \\
&\leq \mathbb{E}[\mathcal{A}_1^q \mathbb{1}_{E_2^c}] + \sqrt{\mathbb{E}[\mathcal{A}_1^{2q}]}\sqrt{\mathbb{P}(E_2)} \\
&\leq C\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^q\right] + n^{q/2}\sqrt{\mathbb{E}\left[\left|\hat{p}_{\hat{j}}(x) - p(x)\right|^{2q}\right]}\sqrt{\mathbb{P}(E_2)}.
\end{aligned}$$

Then, using Theorem 4.3.3 and (4.16), we finally have $\mathbb{E}[\mathcal{A}_1^q] \leq C\phi_n^q(\vec{\beta})$.

Control of $\mathbb{E}[\mathcal{A}_2^q]$. Using (4.15) and the inequality $\hat{f}_X(x) \vee n^{-1/2} \geq n^{-1/2}$, it holds for n large enough

$$\mathbb{E}[\mathcal{A}_2^q] \leq \mathbb{E}[\mathcal{A}_2^q \mathbb{1}_{E_1^c \cap E_2^c}] + \mathbb{E}[\mathcal{A}_2^q(\mathbb{1}_{E_1} + \mathbb{1}_{E_2})] \leq \mathbb{E}[\mathcal{A}_2^q \mathbb{1}_{E_1^c \cap E_2^c}] + Cn^{3q/2}(\mathbb{P}(E_1) + \mathbb{P}(E_2)).$$

Then, using the definition of $\mathcal{A}_2$, (4.14) and (4.16), we obtain $\mathbb{E}[\mathcal{A}_2^q] \leq C\phi_n^q(\vec{\beta})$.

Eventually, by definitions of $\mathcal{A}_1$ and $\mathcal{A}_2$, the proof is completed and

$$\mathbb{E}[|\hat{m}(x) - m(x)|^q] \leq C(\mathbb{E}[\mathcal{A}_1^q] + \mathbb{E}[\mathcal{A}_2^q]) \leq L^{\frac{q(2\nu+1)}{2\bar{\beta}+2\nu+1}} R_3 \left(\frac{\log(n)}{n}\right)^{q\bar{\beta}/(2\bar{\beta}+2\nu+1)}$$

where R_3 is a constant depending on $\gamma, q, \varepsilon, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g, \vec{\beta}$. This completes the proof of Theorem 4.3.4.

□

Proof of Theorem 4.3.5.

Following Meister [79], the proof is straightforward. Indeed, for the regression problem with errors in variables in dimension 1, Theorem 3.5 in Meister [79] proves a lower bound in probability for the pointwise risk which claims that the minimax rate is $n^{-\frac{2\beta}{2\beta+2\nu+1}}$ for Hölder class of index β and degree-of-ill-posedness ν. Following step by step the proof of Theorem 3.5 in Meister [79] in dimension 2 (the extension to general case can be easily deduced), one obtains the lower bound of Theorem 4.3.5. Meister uses densities such as Cauchy distributions which admit multivariate counterparts.

□

4.5.2 Statements and proofs of auxiliary results

This section is devoted to statements and proofs of auxiliary results used in Section 4.5.1.

Statements and proofs of propositions

Let us start with Proposition 4.5.1 which states a concentration inequality of $\hat{p}_j$ around p_j.

Proposition 4.5.1. *Let j be fixed. For any $u > 0$,*

$$\mathbb{P}\left(|\hat{p}_j(x) - p_j(x)| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right) \leq 2e^{-u}, \tag{4.17}$$

where

$$\sigma_j^2 = \mathbb{V}ar(Y_1 T_j(W_1)).$$

For any $\tilde{\gamma} > 1$ we have for any $\tilde{\varepsilon} > 0$ that there exists R_4 only depending on $\tilde{\gamma}$ and $\tilde{\varepsilon}$ such that

$$\mathbb{P}(\sigma_j^2 \geq (1+\tilde{\varepsilon})\tilde{\sigma}_{j,\tilde{\gamma}}^2) \leq R_4 n^{-\tilde{\gamma}},$$

$\tilde{\sigma}_{j,\tilde{\gamma}}^2$ being defined in (4.7).

Proof.

First, note that

$$\hat{p}_j(x) = \sum_k \hat{p}_{jk}\varphi_{jk}(x) = \frac{1}{n}\sum_{l=1}^n Y_l \sum_k (\mathcal{D}_j\varphi)_{j,k}(W_l)\varphi_{jk}(x) = \frac{1}{n}\sum_{l=1}^n U_j(Y_l, W_l).$$

To prove Proposition 4.5.1, we apply the Bernstein inequality to the variables $U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]$ that are independent. Since,

$$U_j(Y_l, W_l) = Y_l T_j(W_l),$$

and

$$\mathbb{E}\left[\varepsilon_l T_j(W_l)\right] = 0,$$

we have for any $q \geq 2$,

$$\begin{aligned} A_q &:= \sum_{l=1}^n \mathbb{E}[|U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]|^q] \\ &= \sum_{l=1}^n \mathbb{E}\left[|m(X_l)T_j(W_l) + \varepsilon_l T_j(W_l) - \mathbb{E}[m(X_l)T_j(W_l)]|^q\right]. \end{aligned} \tag{4.18}$$

With $q = 2$,

$$\begin{aligned} A_2 &= \sum_{l=1}^n \mathbb{E}[|U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]|^2] \\ &= n\mathbb{V}ar(Y_1 T_j(W_1)) \\ &= n\mathbb{E}[(m(X_1)T_j(W_1) + \varepsilon_1 T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)])^2] \\ &= n\mathbb{E}[\varepsilon_1^2 T_j^2(W_1)] + n\mathbb{V}ar(m(X_1)T_j(W_1)) \\ &= n\left(s^2\mathbb{E}[T_j^2(W_1)] + \mathbb{V}ar(m(X_1)T_j(W_1))\right). \end{aligned}$$

Now, for any $q \geq 3$, with $Z \sim \mathcal{N}(0,1)$,

$$\begin{aligned} A_q &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] + \mathbb{E}[|\varepsilon_1 T_j(W_1)|^q]\right) \\ &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] + s^q\mathbb{E}[|Z|^q]\mathbb{E}[|T_j(W_1)|^q]\right) \\ &\leq n2^{q-1}\left(\mathbb{E}[|m(X_1)T_j W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] + s^q\mathbb{E}[|Z|^q]\mathbb{E}[T_j^2(W_1)]\|T_j\|_\infty^{q-2}\right). \end{aligned}$$

Furthermore,

$$\begin{aligned} &\mathbb{E}[|m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)]|^q] \\ &\qquad \leq \mathbb{E}[(m(X_1)T_j(W_1) - \mathbb{E}[m(X_1)T_j(W_1)])^2] \times (2\|m\|_\infty\|T_j\|_\infty)^{q-2} \\ &\qquad = \mathbb{V}ar(m(X_1)T_j(W_1)) \times (2\|m\|_\infty\|T_j\|_\infty)^{q-2}. \end{aligned}$$

Finally,

$$\begin{aligned}
A_q &\le n2^{q-1}\|T_j\|_\infty^{q-2}\left(\mathbb{V}ar(m(X_1)T_j(W_1)) \times (2\|m\|_\infty)^{q-2} + s^q\mathbb{E}[|Z|^q]\mathbb{E}[T_j^2(W_1)]\right)\\
&\le n2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q]\left(\mathbb{V}ar(m(X_1)T_j(W_1)) \times (2\|m\|_\infty)^{q-2} + s^q\mathbb{E}[T_j^2(W_1)]\right)\\
&\le n2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q]\left(\mathbb{V}ar(m(X_1)T_j(W_1)) + s^2\mathbb{E}[T_j^2(W_1)]\right)\\
&\quad \times\left((2\|m\|_\infty)^{q-2} + s^{q-2}\right)\\
&\le 2^{q-1}\|T_j\|_\infty^{q-2}\mathbb{E}[|Z|^q] \times A_2 \times (2\|m\|_\infty + s)^{q-2}.
\end{aligned}$$

Besides we have (see page 23 in [66]) denoting Γ the Gamma function

$$\mathbb{E}[|Z|^q] = \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right) \le 2^{q/2}2^{-1/2}q! \le 2^{(q-1)/2}q!, \tag{4.19}$$

as $\frac{1}{\sqrt{\pi}} \le \frac{1}{\sqrt{2}}$ and $\Gamma(\frac{q+1}{2}) \le \Gamma(q+1) = q!$. So, for $q \ge 3$,

$$\begin{aligned}
A_q &\le 2^{q-1}\|T_j\|_\infty^{q-2}2^{(q-1)/2}q! \times A_2 \times (2\|m\|_\infty + s)^{q-2}\\
&\le \frac{q!}{2} \times A_2 \times \left(2^{\frac{3q-1}{2(q-2)}}\|T_j\|_\infty\,(2\|m\|_\infty + s)\right)^{q-2}.
\end{aligned}$$

The function $\frac{3q-1}{2(q-2)}$ is decreasing in q. Hence for any $q \ge 3$, $2^{\frac{3q-1}{2(q-2)}} \le 16$. Thus

$$A_q \le \frac{q!}{2} \times A_2 \times {c_j}^{q-2}, \tag{4.20}$$

with

$$c_j := 16\|T_j\|_\infty\,(2\mathfrak{m} + s)\,.$$

We can now apply Proposition 2.9 of Massart [78]. We denote f_W the density of the W_l's. We have

$$\begin{aligned}
\mathbb{E}[T_j^2(W_1)] &= \int T_j^2(w)f_W(w)dw\\
&\le \|f_X\|_\infty\|T_j\|_2^2,
\end{aligned}$$

since the density f_W is the convolution of f_X and g, $\|f_W\|_\infty = \|f_X \star g\|_\infty \le \|f_X\|_\infty$. We have

$$\begin{aligned}
\mathbb{V}ar(m(X_1)T_j(W_1)) &\le \mathbb{E}[m^2(X_1)T_j^2(W_1)]\\
&\le \|m\|_\infty^2\int T_j^2(w)f_W(w)dw\\
&\le \|m\|_\infty^2\|f_X\|_\infty\|T_j\|_2^2.
\end{aligned}$$

Therefore, with

$$\sigma_j^2 = \frac{A_2}{n} = \mathbb{V}ar(Y_1T_j(W_1)), \tag{4.21}$$

$$\begin{aligned}
\sigma_j^2 &= \sigma_\varepsilon^2\mathbb{E}[T_j^2(W_1)] + \mathbb{V}ar(m(X_1)T_j(W_1)) \qquad (4.22)\\
&\le \sigma_\varepsilon^2\|f_X\|_\infty\|T_j\|_2^2 + \|m\|_\infty^2\|f_X\|_\infty\|T_j\|_2^2\\
&\le \|f_X\|_\infty\|T_j\|_2^2(s^2 + \|m\|_\infty^2).
\end{aligned}$$

We conclude that for any $u > 0$,

$$\mathbb{P}\left(|\hat{p}_j(x) - p_j(x)| \geq \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right) \leq 2e^{-u}. \tag{4.23}$$

Now, we can write

$$\begin{aligned}
\hat{\sigma}_j^2 &= \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - U_j(Y_v, W_v))^2 \\
&= \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)] - U_j(Y_v, W_v) + \mathbb{E}[U_j(Y_v, W_v)])^2 \\
&= s_j^2 - \frac{2}{n(n-1)} \xi_j,
\end{aligned}$$

with

$$\begin{aligned}
s_j^2 &:= \frac{1}{n(n-1)} \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2 + (U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_v, W_v)])^2 \\
&= \frac{1}{n} \sum_{l=1}^{n} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2
\end{aligned}$$

and

$$\xi_j := \sum_{l=2}^{n} \sum_{v=1}^{l-1} (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)]) \times (U_j(Y_v, W_v) - \mathbb{E}[U_j(Y_v, W_v)]).$$

In the sequel, we denote for any $\tilde{\gamma} > 0$,

$$\Omega_n(\tilde{\gamma}) = \left\{ \max_{1 \leq l \leq n} |\varepsilon_l| \leq s\sqrt{2\tilde{\gamma} \log n} \right\}.$$

We have that

$$\mathbb{P}(\Omega_n(\tilde{\gamma})^c) \leq n^{1-\tilde{\gamma}}. \tag{4.24}$$

Note that on $\Omega_n(\tilde{\gamma})$,

$$\|U_j(\cdot, \cdot)\|_\infty \leq C_j,$$

we recall that

$$C_j = (\mathfrak{m} + s\sqrt{2\tilde{\gamma} \log n}) \|T_j\|_\infty.$$

Lemma 4.5.2. *For any $\tilde{\gamma} > 1$ and any $u > 0$, there exists a sequence $e_{n,j} > 0$ such that $\limsup_j e_{n,j} = 0$ and*

$$\mathbb{P}\left(\sigma_j^2 \geq s_j^2 + 2C_j \sigma_j \sqrt{\frac{2u(1 + e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} \,\middle|\, \Omega_n(\tilde{\gamma})\right) \leq e^{-u}.$$

Proof.

We denote

$$\mathbb{P}_{\Omega_n(\tilde{\gamma})}(\cdot) = \mathbb{P}\left(\cdot|\Omega_n(\tilde{\gamma})\right), \quad \mathbb{E}_{\Omega_n(\tilde{\gamma})}(\cdot) = \mathbb{E}\left(\cdot|\Omega_n(\tilde{\gamma})\right).$$

Note that conditionally to $\Omega_n(\tilde{\gamma})$ the variables $U_j(Y_1, W_1), \ldots, U_j(Y_n, W_n)$ are independent. So, we can apply the classical Bernstein inequality to the variables

$$V_l := \frac{\sigma_j^2 - (U_j(Y_l, W_l) - \mathbb{E}[U_j(Y_l, W_l)])^2}{n} \leq \frac{\sigma_j^2}{n}.$$

Furthermore, as

$$\begin{aligned} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[U_j(Y_1, W_1)] &= \mathbb{E}[m(X_1)T_j(W_1)|\Omega_n(\tilde{\gamma})] + \mathbb{E}[\varepsilon_1 T_j(W_1)|\Omega_n(\tilde{\gamma})] \\ &= \mathbb{E}[m(X_1)T_j(W_1)] \\ &= \mathbb{E}[U_j(Y_1, W_1)] \end{aligned} \tag{4.25}$$

we get

$$\begin{aligned} \sum_{l=1}^{n} \mathbb{E}_{\Omega_n(\tilde{\gamma})}[V_l^2] &= \frac{\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[\left(\sigma_j^2 - (U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2\right)^2\right]}{n} \\ &= \frac{\sigma_j^4 + \mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^4\right]}{n} \\ &\qquad - \frac{2\sigma_j^2 \mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2\right]}{n} \\ &\leq \frac{\sigma_j^4 + (4C_j^2 - 2\sigma_j^2)\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2\right]}{n}. \end{aligned}$$

We shall find an upperbound for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2\right]$:

$$\begin{aligned} &\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1, W_1) - \mathbb{E}[U_j(Y_1, W_1)])^2\right] \\ &\qquad = \mathbb{V}ar(m(X_1)T_j(W_1)) + \mathbb{E}[\varepsilon_1^2 T_j^2(W_1)|\Omega_n(\tilde{\gamma})] \\ &\qquad = \mathbb{V}ar(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{\mathbb{E}[\varepsilon_1^2 \mathbb{1}_{\Omega_n(\tilde{\gamma})}]}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\qquad \leq \mathbb{V}ar(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{s^2}{\mathbb{P}(\Omega_n(\tilde{\gamma}))} \\ &\qquad \leq \mathbb{V}ar(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]\frac{s^2}{1 - n^{1-\tilde{\gamma}}} \\ &\qquad = \mathbb{V}ar(m(X_1)T_j(W_1)) + \mathbb{E}[T_j^2(W_1)]s^2(1 + \tilde{e}_n), \end{aligned}$$

where $\tilde{e}_n = n^{1-\tilde{\gamma}} + o(n^{1-\tilde{\gamma}})$.

Using (4.22) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]\leq(1+e_{n,j})\sigma_j^2, \tag{4.26}$$

where $(e_{n,j})$ is a sequence such that $\limsup_j e_{n,j}=0$.

Now let us find a lower bound for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]$:

$$\begin{aligned}
&\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]\\
&\qquad=\mathbb{V}ar(m(X_1)T_j(W_1))+\mathbb{E}[T_j^2(W_1)]\frac{\mathbb{E}[\varepsilon_1^2\mathbb{1}_{\Omega_n(\tilde{\gamma})}]}{\mathbb{P}(\Omega_n(\tilde{\gamma}))}\\
&\qquad\geq\mathbb{V}ar(m(X_1)T_j(W_1))+\mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2\mathbb{1}_{\Omega_n(\tilde{\gamma})}]\\
&\qquad=\mathbb{V}ar(m(X_1)T_j(W_1))+\mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2(1-\mathbb{1}_{\Omega_n^c(\tilde{\gamma})})]\\
&\qquad=\sigma_j^2-\mathbb{E}[T_j^2(W_1)]\mathbb{E}[\varepsilon_1^2\mathbb{1}_{\Omega_n^c(\tilde{\gamma})}].
\end{aligned}$$

Now using Cauchy Scharwz, (4.19) and (4.24) we have

$$\begin{aligned}
\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]&\geq\sigma_j^2-\mathbb{E}[T_j^2(W_1)](\mathbb{E}[\varepsilon_1^4])^{\frac{1}{2}}(\mathbb{P}(\Omega_n^c(\tilde{\gamma})))^{\frac{1}{2}}\\
&\geq\sigma_j^2-Cs^2\mathbb{E}[T_j^2(W_1)]n^{\frac{1-\tilde{\gamma}}{2}}\\
&=\sigma_j^2(1+\tilde{e}_{n,j}),
\end{aligned}\tag{4.27}$$

where $(\tilde{e}_{n,j})$ is a sequence such that $\limsup_j \tilde{e}_{n,j}=0$.

Finally, using the bounds we just got for $\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]$ yields

$$\begin{aligned}
\sum_{l=1}^n\mathbb{E}_{\Omega_n(\tilde{\gamma})}[V_l^2]&\leq\frac{\sigma_j^4+4C_j^2\sigma_j^2(1+e_{n,j})-2\sigma_j^4(1+\tilde{e}_{n,j})}{n}\\
&\leq\frac{4C_j^2\sigma_j^2(1+e_{n,j})-\sigma_j^4(1+2\tilde{e}_{n,j})}{n}\\
&\leq\frac{4C_j^2\sigma_j^2(1+e_{n,j})}{n}.
\end{aligned}$$

We obtain the claimed result.

□

Now, we deal with ξ_j.

Lemma 4.5.3. *There exists an absolute constant $c>0$ such that for any $u>1$,*

$$\mathbb{P}\left(\xi_j\geq c(n\sigma_j^2u+C_j^2u^2)\middle|\,\Omega_n(\tilde{\gamma})\right)\leq 3e^{-u}.$$

Proof. Note that conditionally to $\Omega_n(\tilde{\gamma})$, the vectors $(Y_l,W_l)_{1\leq l\leq n}$ are independent. We remind that by (4.25), (4.26) and (4.27) we have

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}[U_j(Y_1,W_1)]=\mathbb{E}[U_j(Y_1,W_1)] \tag{4.28}$$

and

$$\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_1,W_1)-\mathbb{E}[U_j(Y_1,W_1)])^2\right]=(1+e_{n,j})\sigma_j^2.$$

The ξ_j can be written as

$$\xi_j=\sum_{l=2}^{n}\sum_{v=1}^{l-1}g_j(Y_l,W_l,Y_v,W_v),$$

with

$$g_j(y,w,y',w')=(U_j(y,w)-\mathbb{E}[U_j(Y_1,W_1)])\times(U_j(y',w')-\mathbb{E}[U_j(Y_1,W_1)]).$$

Previous computations show that Conditions (2.3) and (2.4) of Houdré and Reynaud-Bouret [61] are satisfied. So that we are able to apply Theorem 3.1 of Houdré and Reynaud-Bouret [61]: there exist absolute constants c_1, c_2, c_3 and c_4 such that for any $u>0$,

$$\mathbb{P}_{\Omega_n(\tilde{\gamma})}\left(\xi_j\geq c_1C\sqrt{u}+c_2Du+c_3Bu^{3/2}+c_4Au^2\right)\leq 3e^{-u},$$

where A, B, C, and D are defined and controlled as follows. We have:

$$A=\|g_j\|_\infty\leq 4C_j^2.$$

$$C^2=\sum_{l=2}^{n}\sum_{v=1}^{l-1}\mathbb{E}_{\Omega_n(\tilde{\gamma})}[g_j^2(Y_l,W_l,Y_v,W_v)]=\frac{n(n-1)}{2}\sigma_j^4(1+e_{n,j})^2.$$

Let

$$\mathcal{A}=\left\{(a_l)_l,(b_v)_v:\quad\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[\sum_{l=2}^{n}a_l^2(Y_l,W_l)\right]\leq 1,\ \mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[\sum_{l=1}^{n-1}b_l^2(Y_l,W_l)\right]\leq 1\right\}.$$

We have:

$$\begin{aligned}
D &= \sup_{(a_l)_l,(b_v)_v\in\mathcal{A}}\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[\sum_{l=2}^{n}\sum_{v=1}^{l-1}g_j(Y_l,W_l,Y_v,W_v)a_l(Y_l,W_l)b_v(Y_v,W_v)\right]\\
&= \sup_{(a_l)_l,(b_v)_v\in\mathcal{A}}\left[\sum_{l=2}^{n}\sum_{v=1}^{l-1}\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_l,W_l)-[U_j(Y_l,W_l)]))a_l(Y_l,W_l)\right]\right.\\
&\qquad\left.\times\,\mathbb{E}_{\Omega_n(\tilde{\gamma})}\left[(U_j(Y_v,W_v)-\mathbb{E}[U_j(Y_v,W_v)]))b_v(Y_v,W_v)\right]\right]\\
&\leq \sup_{(a_l)_l,(b_v)_v\in\mathcal{A}}\sum_{l=2}^{n}\sum_{v=1}^{l-1}\sigma_j^2(1+e_{n,j})\sqrt{\mathbb{E}_{\Omega_n(\tilde{\gamma})}[a_l^2(Y_l,W_l)]\mathbb{E}_{\Omega_n(\tilde{\gamma})}[b_v^2(Y_v,W_v)]}\\
&\leq \sigma_j^2(1+e_{n,j})\sup_{(a_l)_l,(b_v)_v\in\mathcal{A}}\sum_{l=2}^{n}\sqrt{l-1}\sqrt{\mathbb{E}_{\Omega_n(\tilde{\gamma})}[a_l^2(Y_l,W_l)]\sum_{v=1}^{l-1}\mathbb{E}_{\Omega_n(\tilde{\gamma})}[b_v^2(Y_v,W_v)]}\\
&\leq \sigma_j^2(1+e_{n,j})\sqrt{\frac{n(n-1)}{2}}.
\end{aligned}$$

Finally,

$$\begin{aligned} B^2 &= \sup_{y,w} \sum_{v=1}^{n-1} \mathbb{E}_{\Omega_n(\tilde{\gamma})} \Big[(U_j(y,w) - \mathbb{E}[U_j(Y_1,W_1)]))^2 \times (U_j(Y_v,W_v) - \mathbb{E}[U_j(Y_1,W_1)])^2 \Big] \\ &\leq 4(n-1)C_j^2 \sigma_j^2 (1+e_{n,j}). \end{aligned}$$

Therefore, there exists an absolute constant $c > 0$ such that for any $u > 1$,

$$c_1 C\sqrt{u} + c_2 Du + c_3 Bu^{3/2} + c_4 Au^2 \leq c(n\sigma_j^2 u + C_j^2 u^2).$$

□

Let us go back to the proof of Proposition 4.5.1. We apply Lemmas 4.5.2 and 4.5.3 with $u > 1$ and we obtain, by setting

$$M_j(u) = \hat{\sigma}_j^2 + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} + \frac{2c(n\sigma_j^2 u + C_j^2 u^2)}{n(n-1)},$$

$$\begin{aligned} &\mathbb{P}\left(\sigma_j^2 \geq M_j(u)\right) \\ &\leq \mathbb{P}\left(\sigma_j^2 \geq s_j^2 - \frac{2}{n(n-1)}\xi_j + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} + \frac{2c(n\sigma_j^2 u + C_j^2 u^2)}{n(n-1)}\right) \\ &\leq \mathbb{P}\left(\sigma_j^2 \geq s_j^2 + 2C_j\sigma_j\sqrt{\frac{2u(1+e_{n,j})}{n}} + \frac{\sigma_j^2 u}{3n} \,\middle|\, \Omega_n(\tilde{\gamma})\right) \\ &\qquad + \mathbb{P}\left(\xi_j \geq c(n\sigma_j^2 u + C_j^2 u^2) \,\middle|\, \Omega_n(\tilde{\gamma})\right) + 1 - \mathbb{P}\Big(\Omega_n(\tilde{\gamma})\Big). \end{aligned}$$

Therefore, with $u = \tilde{\gamma}\log n$ and $\tilde{\gamma} > 1$, we obtain for n large enough:

$$\mathbb{P}\left(\sigma_j^2 \geq M_j(\tilde{\gamma}\log n)\right) \leq 5n^{-\tilde{\gamma}}.$$

And there exist a and b two absolute constants such that

$$\mathbb{P}\left(\sigma_j^2 \geq \hat{\sigma}_j^2 + 2C_j\sigma_j\sqrt{\frac{2\tilde{\gamma}\log n(1+e_{n,j})}{n}} + \frac{\sigma_j^2 a\tilde{\gamma}\log n}{n} + \frac{C_j^2 b^2\tilde{\gamma}^2\log^2 n}{n^2}\right) \leq 5n^{-\tilde{\gamma}}.$$

Now, we set

$$\theta_1 = \left(1 - \frac{a\tilde{\gamma}\log n}{n}\right), \quad \theta_2 = C_j\sqrt{\frac{2\tilde{\gamma}\log n(1+e_{n,j})}{n}}, \quad \theta_3 = \hat{\sigma}_j^2 + \frac{C_j^2 b^2\tilde{\gamma}^2\log^2 n}{n^2}$$

so

$$\mathbb{P}\left(\theta_1\sigma_j^2 - 2\theta_2\sigma_j - \theta_3 \geq 0\right) \leq 5n^{-\tilde{\gamma}}.$$

We study the polynomial

$$p(\sigma) = \theta_1\sigma^2 - 2\theta_2\sigma - \theta_3.$$

Since $\sigma \geq 0$, $p(\sigma) \geq 0$ means that

$$\sigma \geq \frac{1}{\theta_1}\left(\theta_2 + \sqrt{\theta_2^2 + \theta_1\theta_3}\right),$$

which is equivalent to

$$\sigma^2 \geq \frac{1}{\theta_1^2}\left(2\theta_2^2 + \theta_1\theta_3 + 2\theta_2\sqrt{\theta_2^2 + \theta_1\theta_3}\right).$$

Hence

$$\mathbb{P}\left(\sigma_j^2 \geq \frac{1}{\theta_1^2}\left(2\theta_2^2 + \theta_1\theta_3 + 2\theta_2\sqrt{\theta_2^2 + \theta_1\theta_3}\right)\right) \leq 5n^{-\tilde{\gamma}}.$$

So,

$$\mathbb{P}\left(\sigma_j^2 \geq \frac{\theta_3}{\theta_1} + \frac{2\theta_2\sqrt{\theta_3}}{\theta_1\sqrt{\theta_1}} + \frac{4\theta_2^2}{\theta_1^2}\right) \leq 5n^{-\tilde{\gamma}}.$$

So, there exist absolute constants δ, η, and τ' depending only on $\tilde{\gamma}$ so that for n large enough,

$$\mathbb{P}\Bigg(\sigma_j^2 \geq \hat{\sigma}_j^2\left(1 + \delta\frac{\log n}{n}\right) + \left(1 + \eta\frac{\log n}{n}\right)2C_j\sqrt{2\tilde{\gamma}\hat{\sigma}_j^2(1 + e_{n,j})\frac{\log n}{n}}$$
$$+ 8\tilde{\gamma}C_j^2\frac{\log n}{n}\left(1 + \tau'\left(\frac{\log n}{n}\right)^{1/2}\right)\Bigg) \leq 5n^{-\tilde{\gamma}}.$$

Finally, for all $\tilde{\varepsilon} > 0$ there exists R_4 depending on ε' and $\tilde{\gamma}$ such that for n large enough

$$\mathbb{P}(\sigma_j^2 \geq (1 + \varepsilon')\tilde{\sigma}_{j,\tilde{\gamma}}^2) \leq R_4 n^{-\tilde{\gamma}}.$$

Combining this inequality with (4.23), we obtain the desired result of Proposition 4.5.1.

□

Proposition 4.5.4 shows that the residual term in the oracle inequality is negligible.

Proposition 4.5.4. *We have for any $q \geq 1$,*

$$\mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x) - p_j(x)| - \Gamma_\gamma(j)\right)_+^q\right] \leq R_1' n^{-q}, \tag{4.29}$$

with R_1' a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ, c_g, $\mathcal{C}_g$ and φ.

Proof. We recall that $J = \left\{j \in \mathbb{N}^d : \quad 2^{S_j} \leq \lfloor \frac{n}{\log^2 n} \rfloor\right\}$.
Let $\tilde{\gamma} > 0$ and let us consider the event

$$\tilde{\Omega}_{\tilde{\gamma}} = \left\{\sigma_j^2 \leq (1 + \varepsilon)\tilde{\sigma}_{j,\tilde{\gamma}}^2, \ \forall j \in J\right\}.$$

Let $\gamma > 0$. We set in the sequel

$$E := \mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x) - p_j(x)| - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{j,\tilde{\gamma}}\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)^q_+ \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}}\right],$$

and $R_j := |\hat{p}_j(x) - p_j(x)|$. We have:

$$\begin{aligned}
E &= \int_0^\infty \mathbb{P}\left[\sup_{j\in J}\left(R_j - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{j,\tilde{\gamma}}\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)^q_+ \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}} > y\right] dy \\
&\leq \sum_{j\in J}\int_0^\infty \mathbb{P}\left[\left(R_j - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{j,\tilde{\gamma}}\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)^q_+ \mathbb{1}_{\tilde{\Omega}_{\tilde{\gamma}}} > y\right] dy \\
&\leq \sum_{j\in J}\int_0^\infty \mathbb{P}\left[\left(R_j - \sqrt{\frac{2\gamma\sigma^2_j\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)^q > y\right] dy.
\end{aligned}$$

Let us take u such that

$$y = h(u)^q,$$

where

$$h(u) = \sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}.$$

Note that for any $u > 0$,

$$h'(u) \leq \frac{h(u)}{u}.$$

Hence

$$\begin{aligned}
E &\leq C\sum_{j\in J}\int_0^\infty \mathbb{P}\left[R_j > \sqrt{\frac{2\gamma\sigma^2_j\log n}{n}} + \frac{c_j\gamma\log n}{n} + \sqrt{\frac{2u\sigma_j^2}{n}} + \frac{uc_j}{n}\right] h(u)^{q-1}h'(u)du \\
&\leq C\sum_{j\in J}\int_0^\infty \mathbb{P}\left[R_j > \sqrt{\frac{2\sigma^2_j(\gamma\log n + u)}{n}} + \frac{c_j(\gamma\log n + u)}{n}\right] h(u)^{q-1}h'(u)du.
\end{aligned}$$

Now using concentration inequality (4.17), we get

$$\begin{aligned}
E &\leq C\sum_{j\in J}\int_0^\infty e^{-(\gamma\log n + u)}h(u)^{q-1}h'(u)du \\
&\leq C\sum_{j\in J}\int_0^\infty e^{-(\gamma\log n + u)}h(u)^{q}\frac{1}{u}du \\
&\leq Ce^{-\gamma\log n}\sum_{j\in J}\int_0^\infty e^{-u}\left(\sqrt{\frac{2\sigma_j^2 u}{n}} + \frac{c_j u}{n}\right)^q \frac{1}{u}du \\
&\leq C\left(e^{-\gamma\log n}\sum_{j\in J}\left(\frac{\sigma_j^2}{n}\right)^{q/2}\int_0^\infty e^{-u}u^{\frac{q}{2}-1}du + \left(\frac{c_j}{n}\right)^q\int_0^\infty e^{-u}u^{q-1}du\right).
\end{aligned}$$

Now using Lemma 4.5.13, we have that $\sigma_j^2 \leq R_{10} 2^{(2S_j\nu+S_j)}$ and $c_j \leq C2^{S_j\nu+S_j}$. Hence,

$$\begin{aligned} E &\leq C\left(e^{-\gamma\log n}\sum_{j\in J}\left(\frac{2^{(2S_j\nu+S_j)}}{n}\right)^{q/2} + \left(\frac{2^{(S_j\nu+S_j)}}{n}\right)^{q}\right) \\ &\leq Cn^{-\gamma+q\nu}(\log n)^{-(2\nu+1)q} \leq Cn^{-q}, \end{aligned}$$

as soon as $\gamma > q(\nu+1)$.

It remains to find an upperbound for the following quantity:

$$E' := \mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x)-p_j(x)| - \sqrt{\frac{2\gamma(1+\varepsilon)\tilde{\sigma}^2_{j,\tilde{\gamma}}\log n}{n}} - \frac{c_j\gamma\log n}{n}\right)_+^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right].$$

We have

$$\begin{aligned} E' &\leq \mathbb{E}\left[\sup_{j\in J}\left(|\hat{p}_j(x)-p_j(x)|^q\,\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\right. \\ &\leq 2^{q-1}\left(\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] + \mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\right). \end{aligned}$$

First, let us deal with the term $\mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]$.

Following the lines of the proof of Lemma 4.5.10 we easily get that $\sum_k \varphi^2_{jk}(x) \leq C2^{S_j}$, hence

$$\begin{aligned} |p_j(x)| &= \left|\sum_k p_{jk}\varphi_{jk}(x)\right| \leq \left(\sum_k p^2_{jk}\right)^{\frac{1}{2}}\left(\sum_k \varphi^2_{jk}(x)\right)^{\frac{1}{2}} \\ &\leq C\|p\|_2 2^{\frac{S_j}{2}}. \end{aligned}$$

Now using Proposition 4.5.1 which states that $\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) \leq Cn^{-\tilde{\gamma}}$

$$\begin{aligned} \mathbb{E}\left[\sup_{j\in J}(|p_j(x)|)^q\mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] &\leq \sup_{j\in J}(\|p\|_2 2^{\frac{S_j}{2}})^q\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) && (4.30) \\ &\leq C\left(\frac{n}{\log^2 n}\right)^{\frac{q}{2}} n^{-\tilde{\gamma}}. && (4.31) \end{aligned}$$

It remains to find an upperbound for $\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]$. We have

$$\begin{aligned}
\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] &= \mathbb{E}\left[\sup_{j\in J}\left|\frac{1}{n}\sum_{l=1}^n Y_l T_j(W_l)\right|^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] \\
&\le \frac{1}{n^q}\mathbb{E}\left[\sup_{j\in J}\left(\sum_{l=1}^n |m(X_l)+\varepsilon_l|\,|T_j(W_l)|\right)^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] \\
&\le \frac{n^{q-1}}{n^q}\mathbb{E}\left[\sup_{j\in J}\sum_{l=1}^n |m(X_l)+\varepsilon_l|^q\,|T_j(W_l)|^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] \\
&\le \frac{C}{n}\mathbb{E}\left[\sup_{j\in J}\sum_{l=1}^n (\|m\|_\infty^q + |\varepsilon_l|^q)|T_j(W_l)|^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] \\
&\le C\left(\sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) + \sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{E}\left[|\varepsilon_1|^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right]\right) \\
&\le C\left(\sup_{j\in J}(\|T_j\|_\infty^q)\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}}) + s^q \sup_{j\in J}(\|T_j\|_\infty^q)\left(\mathbb{E}\left[|Z|^{2q}\right]\right)^{\frac{1}{2}}\left(\mathbb{P}(\tilde{\Omega}^c_{\tilde{\gamma}})\right)^{\frac{1}{2}}\right),
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$. Using (4.19) and $\|T_j\|_\infty \le T_4 2^{S_j(\nu+1)}$, we get

$$\mathbb{E}\left[\sup_{j\in J}(|\hat{p}_j(x)|)^q \mathbb{1}_{\tilde{\Omega}^c_{\tilde{\gamma}}}\right] \quad\le\quad C\left(\frac{n}{\log^2 n}\right)^{(\nu+1)q} n^{-\frac{\tilde{\gamma}}{2}}.$$

We have

$$\begin{aligned}
E' &\le Cn^{-\frac{\tilde{\gamma}}{2}}\left(\left(\frac{n}{\log^2 n}\right)^{\frac{q}{2}} + \left(\frac{n}{\log^2 n}\right)^{(\nu+1)q}\right) \\
&\le Cn^{-q},
\end{aligned}$$

as soon as $\tilde{\gamma} > 2q(\nu+2)$. This ends the proof of Proposition 4.5.4.

□

Proposition 4.5.5 controls the bias term in the oracle inequality.

Proposition 4.5.5. *For any $j = (j_1,\dots,j_d) \in \mathbb{Z}^d$ and $j' = (j'_1,\dots,j'_d) \in \mathbb{Z}^d$ and any x, if $p \in \mathbb{H}_d(\vec{\beta}, L)$*

$$|p_{j\wedge j'}(x) - p_{j'}(x)| \le R_{12} L \sum_{l=1}^d 2^{-j_l\beta_l},$$

where R_{12} is a constant only depending on φ and $\vec{\beta}$. We have denoted

$$j \wedge j' = (j_1 \wedge j'_1, \dots, j_d \wedge j'_d).$$

Proof. We first state three lemmas.

Lemma 4.5.6. *For any j and any k, we have:*

$$\mathbb{E}[\hat{p}_{jk}] = p_{jk}.$$

Proof. Recall that

$$\hat{p}_{jk} := \frac{1}{n}\sum_{u=1}^{n} Y_u \times (\mathcal{D}_j\varphi)_{j,k}(W_u) = \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\frac{1}{n}\sum_{u=1}^{n} Y_u \int e^{-i<t,2^jW_u-k>}\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^jt)}dt.$$

Let us prove now that $\mathbb{E}(\hat{p}_{jk}) = p_{jk}$.
We have

$$\mathbb{E}(\hat{p}_{jk}) = \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\left(\int \mathbb{E}(Y_1 e^{-i<t,2^jW_1-k>})\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^jt)}dt\right).$$

We shall develop the right member of the last equality. We have :

$$\begin{aligned}
\mathbb{E}\left[Y_1 e^{-i<t,2^jW_1-k>}\right] &= \mathbb{E}\left[(m(X_1)+\varepsilon_1)e^{-i<t,2^jW_1-k>}\right] \\
&= \mathbb{E}\left[m(X_1)e^{-i<t,2^jW_1-k>}\right] \\
&= \mathbb{E}\left[m(X_1)e^{-i<t,2^jX_1-k>}\right]\mathbb{E}\left[e^{-i<t,2^j\delta_1>}\right] \\
&= \int m(x)e^{-i<t,2^jx-k>}f_X(x)dx \times \mathcal{F}(g)(2^jt) \\
&= e^{i<t,k>}\mathcal{F}(p)(2^jt)\mathcal{F}(g)(2^jt).
\end{aligned}$$

Consequently

$$\begin{aligned}
\mathbb{E}\left[\hat{p}_{jk}\right] &= \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\int e^{i<t,k>}\mathcal{F}(p)(2^jt)\mathcal{F}(g)(2^jt)\frac{\overline{\mathcal{F}(\varphi)(t)}}{\mathcal{F}(g)(2^jt)}dt \\
&= \frac{2^{\frac{S_j}{2}}}{(2\pi)^d}\int e^{i<t,k>}\mathcal{F}(p)(2^jt)\overline{\mathcal{F}(\varphi)(t)}dt \\
&= \frac{1}{(2\pi)^d}\int \mathcal{F}(p)(t)\overline{\mathcal{F}(\varphi_{jk})(t)}dt.
\end{aligned}$$

Since by Parseval equality, we have

$$p_{jk} = \int p(t)\varphi_{jk}(t)dt = \frac{1}{(2\pi)^d}\int \mathcal{F}(p)(t)\overline{\mathcal{F}(\varphi_{jk})(t)}dt,$$

the result follows.

Note that in the case where we don't have any noise on the variable i.e $g(x) = \delta_0(x)$, since $\mathcal{F}(g)(t) = 1$, the proof above remains valid and we get $\mathbb{E}[\hat{p}_{jk}] = p_{jk}$.

□

Lemma 4.5.7. *If for any l, $\lfloor \beta_l \rfloor \leq N$, the following holds: for any $j \in \mathbb{Z}^d$ and any $p \in \mathbb{H}_d(\vec{\beta}, L)$,*

$$|\mathbb{E}[\hat{p}_j(x)] - p(x)| \leq L(\|\varphi\|_\infty \|\varphi\|_1)^d (2A+1)^d \sum_{l=1}^{d} \frac{(2A \times 2^{-j_l})^{\beta_l}}{\lfloor \beta_l \rfloor !}.$$

Proof. Let x be fixed and $j = (j_1, \ldots, j_d) \in \mathbb{Z}^d$. We have:

$$\int K_j(x,y)dy = \int \sum_{k_1} \cdots \sum_{k_d} \prod_{l=1}^{d} [2^{j_l} \varphi(2^{j_l} x_l - k_l) \varphi(2^{j_l} y_l - k_l) dy_l] = 1.$$

Therefore, using lemma 4.5.6

$$\begin{aligned}
\mathbb{E}[\hat{p}_j(x)] - p(x) &= p_j(x) - p(x) \\
&= \int K_j(x,y)(p(y) - p(x))dy \\
&= \sum_k \varphi_{jk}(x) \int \varphi_{jk}(y)(p(y) - p(x))dy \\
&= \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \varphi_{jk}(x) \int \prod_{l=1}^{d} 2^{\frac{j_l}{2}} \varphi(2^{j_l} y_l - k_l)(p(y) - p(x))dy.
\end{aligned}$$

Now, we use that

$$p(y) - p(x) = \sum_{l=1}^{d} p(x_1, \ldots, x_{l-1}, y_l, y_{l+1}, \ldots, y_d) - p(x_1, \ldots x_{l-1}, x_l, y_{l+1}, \ldots, y_d),$$

with $p(x_1, \ldots, x_l, y_{l+1}, \ldots, y_d) = p(x_1, \ldots, x_d)$ if $l = d$ and $p(x_1, \ldots, x_{l-1}, y_l, \ldots, y_d) = p(y_1, \ldots, y_d)$ if $l = 1$. Furthermore, the Taylor expansion gives: for any $l \in \{1, \ldots, d\}$, for some $u_l \in [0; 1]$,

$$\begin{aligned}
&p(x_1, \ldots, x_{l-1}, y_l, y_{l+1}, \ldots, y_d) - p(x_1, \ldots x_{l-1}, x_l, y_{l+1}, \ldots, y_d) = \\
&\sum_{k=1}^{\lfloor \beta_l \rfloor} \frac{\partial^k p}{\partial x_l^k}(x_1, \ldots x_{l-1}, x_l, y_{l+1}, \ldots, y_d) \times \frac{(y_l - x_l)^k}{k!} + \\
&\frac{\partial^{\lfloor \beta_l \rfloor} p}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1, \ldots x_{l-1}, x_l + (y_l - x_l)u_l, y_{l+1}, \ldots, y_d) \times \frac{(y_l - x_l)^{\lfloor \beta_l \rfloor}}{\lfloor \beta_l \rfloor !} \\
&- \frac{\partial^{\lfloor \beta_l \rfloor} p}{\partial x_l^{\lfloor \beta_l \rfloor}}(x_1, \ldots x_{l-1}, x_l, y_{l+1}, \ldots, y_d) \times \frac{(y_l - x_l)^{\lfloor \beta_l \rfloor}}{\lfloor \beta_l \rfloor !}.
\end{aligned}$$

Using vanishing moments of K_j and $p \in \mathbb{H}_d(\vec{\beta}, L)$, we obtain:

$$\begin{aligned}
&|p_j(x) - p(x)| \\
&\leq \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} |\varphi_{jk}(x)| \int \prod_{l=1}^{d} 2^{\frac{j_l}{2}} |\varphi(2^{j_l} y_l - k_l)| \sum_{l=1}^{d} L \frac{|y_l - x_\ell|^{\beta_l}}{\lfloor \beta_l \rfloor !} dy \\
&\leq \|\varphi\|_\infty^d \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \int_{[-A;A]^d} \prod_{l=1}^{d} |\varphi(u_l)| \sum_{l=1}^{d} L \frac{|2^{-j_l}(u_l + k_l) - x_l|^{\beta_l}}{\lfloor \beta_l \rfloor !} du.
\end{aligned}$$

Since for any l, $k_l \in \mathcal{Z}_{j,l}(x)$, we finally obtain

$$\begin{aligned}
&|p_j(x) - p(x)| \\
&\leq \|\varphi\|_\infty^d \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \int_{[-A;A]^d} \prod_{l=1}^d |\varphi(u_l)| \sum_{l=1}^d L \frac{(2A \times 2^{-j_l})^{\beta_l}}{\lfloor \beta_l \rfloor !} du \\
&\leq L(\|\varphi\|_\infty \|\varphi\|_1)^d (2A+1)^d \sum_{l=1}^d \frac{(2A \times 2^{-j_l})^{\beta_l}}{\lfloor \beta_l \rfloor !}.
\end{aligned}$$

□

Lemma 4.5.8. *We have for any $j = (j_1, \ldots, j_d) \in \mathbb{Z}^d$ and $j' = (j'_1, \ldots, j'_d) \in \mathbb{Z}^d$ and any x,*

$$K_{j'}(p_j)(x) = p_{j \wedge j'}(x).$$

Proof.

We only deal with the case $d = 2$. The extension to the general case can be easily deduced. If for $i = 1, 2$, $j_i \leq j'_i$ the result is obvious. It is also the case if for $l = 1, 2$, $j'_l \leq j_l$. So, without loss of generality, we assume that $j_1 \leq j'_1$ and $j'_2 \leq j_2$.

We have:

$$\begin{aligned}
K_{j'}(p_j)(x) &= \int K_{j'}(x, y) p_j(y) dy \\
&= \int \sum_k \varphi_{j'k}(x) \varphi_{j'k}(y) p_j(y) dy \\
&= \int \sum_{k_1} \sum_{k_2} \varphi_{j'_1 k_1}(x_1) \varphi_{j'_2 k_2}(x_2) \varphi_{j'_1 k_1}(y_1) \varphi_{j'_2 k_2}(y_2) p_j(y) dy_1 dy_2 \\
&= \int \sum_{k_1} \sum_{k_2} \varphi_{j'_1 k_1}(x_1) \varphi_{j'_2 k_2}(x_2) \varphi_{j'_1 k_1}(y_1) \varphi_{j'_2 k_2}(y_2) \\
&\quad \times \sum_{\ell_1} \sum_{\ell_2} \varphi_{j_1 \ell_1}(y_1) \varphi_{j_2 \ell_2}(y_2) \varphi_{j_1 \ell_1}(u_1) \varphi_{j_2 \ell_2}(u_2) p(u_1, u_2) du_1 du_2 dy_1 dy_2.
\end{aligned}$$

Since $j_1 \leq j'_1$, we have in the one-dimensional case, by a slight abuse of notation, $V_{j_1} \subset V_{j'_1}$ and

$$\int \sum_{k_1} \varphi_{j'_1 k_1}(x_1) \varphi_{j'_1 k_1}(y_1) \varphi_{j_1 \ell_1}(y_1) dy_1 = \int K_{j'_1}(x_1, y_1) \varphi_{j_1 \ell_1}(y_1) dy_1 = \varphi_{j_1 \ell_1}(x_1).$$

Similarly, since $j'_2 \leq j_2$, we have $V_{j'_2} \subset V_{j_2}$ and

$$\int \sum_{\ell_2} \varphi_{j_2 \ell_2}(y_2) \varphi_{j_2 \ell_2}(u_2) \varphi_{j'_2 k_2}(y_2) dy_2 = \int K_{j_2}(u_2, y_2) \varphi_{j'_2 k_2}(y_2) dy_2 = \varphi_{j'_2 k_2}(u_2).$$

Therefore, with $\tilde{j} = j \wedge j'$,

$$\begin{aligned} K_{j'}(p_j)(x) &= \int \sum_{k_2} \sum_{\ell_1} \varphi_{j'_2 k_2}(x_2)\varphi_{j_1\ell_1}(u_1)\varphi_{j_1\ell_1}(x_1)\varphi_{j'_2 k_2}(u_2)p(u_1,u_2)du_1du_2 \\ &= \int \sum_{\ell_1} \sum_{\ell_2} \varphi_{\tilde{j}_2\ell_2}(x_2)\varphi_{\tilde{j}_1\ell_1}(u_1)\varphi_{\tilde{j}_1\ell_1}(x_1)\varphi_{\tilde{j}_2\ell_2}(u_2)p(u_1,u_2)du_1du_2 \\ &= \int \sum_{\ell} \varphi_{\tilde{j}\ell}(x)\varphi_{\tilde{j}\ell}(u)p(u)du \\ &= p_{\tilde{j}}(x), \end{aligned}$$

which ends the proof of the lemma.

□

Now, we shall go back to the proof of Proposition 4.5.5. We easily deduce the result:

$$\begin{aligned} p_{j\wedge j'}(x) - p_{j'}(x) &= K_{j'}(p_j)(x) - K_{j'}(p)(x) \\ &= \int K_{j'}(x,y)(p_j(y) - p(y))dy. \end{aligned}$$

Therefore,

$$\begin{aligned} |p_{j\wedge j'}(x) - p_{j'}(x)| &\le \int |K_{j'}(x,y)||p_j(y) - p(y)|dy \\ &\le R_{12} L \sum_{l=1}^{d} 2^{-j_l\beta_l} \times \int |K_{j'}(x,y)|dy, \end{aligned}$$

where R_{12} is a constant only depending on φ and $\vec{\beta}$. We conclude by observing that

$$\begin{aligned} \int |K_{j'}(x,y)|dy &= \int \sum_{k_1} \cdots \sum_{k_d} \prod_{l=1}^{d} [2^{j'_l}|\varphi(2^{j'_l}x_l - k_l)||\varphi(2^{j'_l}y_l - k_l)|dy_i] \\ &\le \|\varphi\|_\infty^d \sum_{k_1\in\mathcal{Z}_{j',1}(x)} \cdots \sum_{k_d\in\mathcal{Z}_{j',d}(x)} \left(\int |\varphi(v)|dv\right)^d \\ &\le \left(\|\varphi\|_\infty \|\varphi\|_1(2A+1)\right)^d. \end{aligned}$$

We thus obtain the claimed result of Proposition 4.5.5.

□

Appendix

Technical lemmas are stated and proved below.

Lemma 4.5.9. *We have*

$$\mathbb{E}[(\tilde{\sigma}_{j,\tilde{\gamma}})^q] \le R_5 2^{S_j(2\nu+1)\frac{q}{2}},$$

with R_5 a constant depending on $q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g$.

Proof. First, let us focus on the case $q \geq 2$. We recall the expression of $\tilde{\sigma}^2_{j,\tilde{\gamma}}$

$$\tilde{\sigma}^2_{j,\tilde{\gamma}} = \hat{\sigma}^2_j + 2C_j\sqrt{2\tilde{\gamma}\hat{\sigma}^2_j\frac{\log n}{n}} + 8\tilde{\gamma}C^2_j\frac{\log n}{n}.$$

We shall first prove that

$$\mathbb{E}[(\hat{\sigma}_j)^q] \leq C2^{S_j(2\nu+1)\frac{q}{2}}.$$

Let us remind that

$$\hat{\sigma}^2_j = \frac{1}{2n(n-1)}\sum_{l\neq v}(U_j(Y_l,W_l) - U_j(Y_v,W_v))^2.$$

We easily get

$$\hat{\sigma}^2_j \leq \frac{C}{n}\sum_l(U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)])^2.$$

First let us remark that

$$\left(\sum_l(U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)])^2\right)^{\frac{q}{2}} \\ \leq C\left(\left(\sum_l((U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)])^2 - \sigma^2_j)\right)^{\frac{q}{2}} + n^{\frac{q}{2}}\sigma^q_j\right).$$

We will use Rosenthal inequality (see [56]) to find an upper bound for

$$\mathbb{E}\left[\left(\sum_l((U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)])^2 - \sigma^2_j)\right)^{\frac{q}{2}}\right].$$

We set

$$B_l := (U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)])^2 - \sigma^2_j.$$

The variables B_l are i.i.d and centered. We have to check that $\mathbb{E}[|B_l|^{\frac{q}{2}}] < \infty$. We have

$$\mathbb{E}[|B_l|^{\frac{q}{2}}] \leq C(\mathbb{E}[|(U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)]|^q] + \sigma^q_j),$$

but

$$\mathbb{E}[|(U_j(Y_l,W_l) - \mathbb{E}[U_j(Y_1,W_1)]|^q] = \frac{A_q}{n},$$

with A_q defined in (4.18). Hence

$$\mathbb{E}[|B_l|^{\frac{q}{2}}] \leq C\left(\frac{A_q}{n} + \sigma^q_j\right). \tag{4.32}$$

Using the control of A_q in (4.20), equation (4.21) and Lemma 4.5.13 we have

$$\begin{aligned} A_q &\leq Cn\sigma^2_j\|T_j\|^{q-2}_\infty \\ &\leq Cn2^{S_j(q\nu+q-1)}. \end{aligned} \tag{4.33}$$

Now, we are able to apply the Rosenthal inequality to the variables B_l which yields

$$\mathbb{E}\left[\left(\sum_l B_l\right)^{\frac{q}{2}}\right] \leq C\left(\sum_l \mathbb{E}[|B_l|^{\frac{q}{2}}] + \left(\sum_l \mathbb{E}[B_l^2]\right)^{\frac{q}{4}}\right),$$

and using (4.32) and (4.33) we get

$$\begin{aligned}\mathbb{E}\left[\left(\sum_l B_l\right)^{\frac{q}{2}}\right] &\leq C\left(\sum_l \left(\frac{A_q}{n} + \sigma_j^q\right) + \left(\sum_l \left(\frac{A_4}{n} + \sigma_j^4\right)\right)^{\frac{q}{4}}\right)\\ &\leq C\left(A_q + n\sigma_j^q + (A_4)^{\frac{q}{4}} + n^{\frac{q}{4}}\sigma_j^q\right)\\ &\leq C\left(n2^{S_j(q\nu+q-1)} + n2^{S_j(2\nu+1)\frac{q}{2}} + (n2^{S_j(4\nu+3)})^{\frac{q}{4}}\right).\end{aligned}$$

Consequently

$$\begin{aligned}\mathbb{E}[\hat{\sigma}_j^q] &\leq Cn^{-\frac{q}{2}}\left(n2^{S_j(q\nu+q-1)} + n2^{S_j(2\nu+1)\frac{q}{2}} + (n2^{S_j(4\nu+3)})^{\frac{q}{4}} + n^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}}\right)\\ &\leq C(n^{1-\frac{q}{2}}2^{S_j(q\nu+q-1)} + n^{1-\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}} + n^{-\frac{q}{4}}2^{S_j(4\nu+3)\frac{q}{4}} + 2^{S_j(2\nu+1)\frac{q}{2}}).\end{aligned}$$

Let us compare each term of the r.h.s of the last inequality. We have

$$n^{1-\frac{q}{2}}2^{S_j(q\nu+q-1)} \leq 2^{S_j(2\nu+1)\frac{q}{2}} \iff 2^{S_j} \leq n,$$

which is true by (4.11). Similarly we have

$$n^{-\frac{q}{4}}2^{S_j(4\nu+3)\frac{q}{4}} \leq 2^{S_j(2\nu+1)\frac{q}{2}} \iff 2^{S_j} \leq n,$$

and obviously

$$n^{1-\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{2}} \leq 2^{S_j(2\nu+1)\frac{q}{2}}.$$

Thus we get that the dominant term in r.h.s is $2^{S_j(2\nu+1)\frac{q}{2}}$. Hence

$$\mathbb{E}[\hat{\sigma}_j^q] \leq C2^{S_j(2\nu+1)\frac{q}{2}}.$$

Now using that

$$\mathbb{E}[\tilde{\sigma}_{j,\tilde{\gamma}}^q] \leq C\left(\mathbb{E}[\hat{\sigma}_j^q] + \left(2C_j\sqrt{2\tilde{\gamma}\frac{\log n}{n}}\right)^{\frac{q}{2}}\mathbb{E}[\hat{\sigma}_j^{\frac{q}{2}}] + \left(8\tilde{\gamma}C_j^2\frac{\log n}{n}\right)^{\frac{q}{2}}\right),$$

and since $C_j \leq C\sqrt{\log n}2^{S_j(\nu+1)}$, we have

$$\mathbb{E}[\tilde{\sigma}_{j,\tilde{\gamma}}^q] \leq C\left(2^{S_j(2\nu+1)\frac{q}{2}} + ((\log n)n^{-\frac{1}{2}}2^{S_j(\nu+1)})^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{4}} + \left(\frac{\log^2 n}{n}2^{2S_j(\nu+1)}\right)^{\frac{q}{2}}\right).$$

Let us compare the three terms of the right hand side. We have

$$2^{S_j\frac{q(2\nu+1)}{2}} \geq ((\log n)n^{-\frac{1}{2}}2^{S_j(\nu+1)})^{\frac{q}{2}}2^{S_j(2\nu+1)\frac{q}{4}} \iff 2^{S_j(q\nu+\frac{q}{2})} \geq (\log n)^{\frac{q}{2}}n^{-\frac{q}{4}}2^{S_j(q\nu+\frac{3q}{4})}$$
$$\iff 2^{S_j} \leq \frac{n}{\log^2 n},$$

which is true by (4.11). Furthermore we have

$$2^{S_j\frac{q(2\nu+1)}{2}} \geq \left(\frac{\log^2 n}{n}2^{2S_j(\nu+1)}\right)^{\frac{q}{2}} \iff 2^{S_j(q\nu+\frac{q}{2})} \geq \left(\frac{\log^2 n}{n}\right)^{\frac{q}{2}} 2^{S_j(q\nu+q)}$$
$$\iff 2^{S_j} \leq \frac{n}{\log^2 n},$$

which is true again by (4.11). Consequently

$$\mathbb{E}[\tilde{\sigma}^q_{j,\tilde{\gamma}}] \leq R_5 2^{S_j(2\nu+1)\frac{q}{2}},$$

with R_5 a constant depending on $q, \tilde{\gamma}, \mathfrak{m}, s, \mathfrak{d}, \varphi, c_g, \mathcal{C}_g$ and the lemma is proved for $q \geq 2$.

For the case $q \leq 2$ the result follows from Jensen inequality.

□

Lemma 4.5.10. *Under assumption (A1) on the father wavelet φ, we have for any $j = (j_1, \dots, j_d)$ and any $x \in \mathbb{R}^d$,*

$$\sum_k |\varphi_{jk}(x)| \leq (2A+1)^d \|\varphi\|_\infty^d 2^{\frac{S_j}{2}}.$$

Proof. Let $x \in \mathbb{R}^d$ be fixed. We set for any j and any $l \in \{1, \dots, d\}$,

$$\mathcal{Z}_{j,l}(x) = \left\{k_l : \quad |2^{j_l}x_l - k_l| \leq A\right\},$$

whose cardinal is smaller or equal to $(2A+1)$. Since

$$\varphi_{jk}(x) = \prod_{l=1}^d 2^{\frac{j_l}{2}}\varphi(2^{j_l}x_l - k_l),$$

then

$$\varphi_{jk}(x) \neq 0 \Rightarrow \forall\, l \in \{1, \dots, d\},\ k_l \in \mathcal{Z}_{j,l}(x).$$

Now,

$$\begin{aligned}
\sum_k |\varphi_{jk}(x)| &= \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \prod_{l=1}^d 2^{\frac{j_l}{2}} |\varphi(2^{j_l}x_l - k_l)| \\
&\leq \sum_{k_1 \in \mathcal{Z}_{j,1}(x)} \cdots \sum_{k_d \in \mathcal{Z}_{j,d}(x)} \|\varphi\|_\infty^d 2^{\frac{S_j}{2}} \\
&\leq (2A+1)^d \|\varphi\|_\infty^d 2^{\frac{S_j}{2}}.
\end{aligned}$$

□

Lemma 4.5.11. *Under condition (A1) and φ is $\mathcal{C}^r$, there exist constants R_6 and R_7 depending on φ such that*

$$\big|\mathcal{F}(\varphi)(t)\big| \leq R_6(1+|t|)^{-r}, \quad \textit{for any } \; t \in \mathbb{R}. \tag{4.34}$$

and

$$\Big|\overline{\mathcal{F}(\varphi)(t)}'\Big| \leq R_7(1+|t|)^{-r}, \quad \textit{for any } \; t \in \mathbb{R}. \tag{4.35}$$

Proof. First, let us focus on the case $|t| \geq 1$.

We have by integration by parts that

$$\mathcal{F}(\varphi)(t) = \int e^{-itx}\varphi(x)dx = \Big[-\frac{1}{it}e^{-itx}\varphi(x)\Big]_{-\infty}^{\infty} + \frac{1}{it}\int e^{-itx}\varphi'(x)dx.$$

Using that the father wavelet φ is compactly supported on $[-A, A]$, we get

$$\mathcal{F}(\varphi)(t) = \frac{1}{it}\int e^{-itx}\varphi'(x)dx.$$

By successive integration by parts and using that $|t| \geq 1$ one gets

$$\big|\mathcal{F}(\varphi)(t)\big| = \left|\frac{1}{(it)^r}\int e^{-itx}\varphi^{(r)}(x)dx\right| \leq \frac{2^r}{(1+|t|)^r}\int |\varphi^{(r)}(x)|dx,$$

the integral $\int_{-A}^{A} |\varphi^{(r)}(x)|dx$ being finite.

For the derivative we have

$$\overline{\mathcal{F}(\varphi)(t)}' = i\int e^{itx}x\varphi(x)dx.$$

Following the same scheme as for $\mathcal{F}(\varphi)(t)$, one gets by integration by parts and using the Leibniz formula that

$$\begin{aligned}\Big|\overline{\mathcal{F}(\varphi)(t)}'\Big| &= \left|\frac{1}{(it)^r}\int e^{itx}\frac{d^r}{dx^r}(x\varphi(x))dx\right| = \left|\frac{1}{(it)^r}\int e^{itx}\sum_{k=0}^{r}\binom{r}{k}x^{(k)}\varphi(x)^{(r-k)}dx\right| \\ &\leq \frac{2^r}{(1+|t|)^r}\sum_{k=0}^{r}\binom{r}{k}\int |x^{(k)}\varphi(x)^{(r-k)}|dx,\end{aligned}$$

the quantity $\sum_{k=0}^{r}\binom{r}{k}\int_{-A}^{A}|x^{(k)}\varphi(x)^{(r-k)}|dx$ being finite.

Hence the lemma is proved for $|t| \geq 1$.

The result for $|t| \leq 1$ is obvious since

$$\big|\mathcal{F}(\varphi)(t)\big| = \left|\int e^{-itx}\varphi(x)dx\right| \leq \int |\varphi(x)|dx < \infty,$$

and

$$\Big|\overline{\mathcal{F}(\varphi)(t)}'\Big| = \left|i\int e^{itx}x\varphi(x)dx\right| \leq \int |x\varphi(x)|dx < \infty.$$

Then the lemma is proved for any t.

□

Lemma 4.5.12. *Under conditions (A1) and (A3), for $\nu \geq 0$, we have*

$$\left|(\mathcal{D}_j\varphi)(w)\right| \leq R_8 2^{S_j\nu} \prod_{l=1}^{d}(1+|w_l|)^{-1},\ w \in \mathbb{R}^d$$

where R_8 is a constant depending on φ, $\mathcal{C}_g$ and c_g.

Proof. If all the $|w_l| < 1$ then using (4.3), Lemma 4.5.11 and $r \geq \nu + 2$ with $\nu \geq 0$ we have

$$\begin{aligned}
\left|(\mathcal{D}_j\varphi)(w)\right| &\leq C\int \prod_{l=1}^{d} \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)} \\
&\leq C\prod_{l=1}^{d}\int \left|\mathcal{F}(\varphi)(t_l)(1+2^{j_l}|t_l|)^{\nu}\right| dt_l \\
&\leq C2^{S_j\nu}\prod_{l=1}^{d}\int (1+|t_l|)^{\nu-r} dt_l \\
&\leq C2^{S_j\nu} \leq C2^{S_j\nu}\prod_{l=1}^{d}(1+|w_l|)^{-1}. \qquad (4.36)
\end{aligned}$$

Now we consider the case where there exists at least one w_l such that $|w_l| \geq 1$. We have

$$(\mathcal{D}_j\varphi)(w) = \frac{1}{(2\pi)^d}\prod_{l=1,|w_l|\leq 1}^{d}\int e^{-it_lw_l}\frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt_l \times \prod_{l=1,|w_l|\geq 1}^{d}\int e^{-it_lw_l}\frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt_l.$$

For the left-hand product on $|w_l| \leq 1$ we use the result (4.36). Now let us consider the right-hand product with $|w_l| \geq 1$. We set in the sequel

$$\eta_l(t_l) := \frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}.$$

We have

$$\prod_{l=1,|w_l|\geq 1}^{d}\int e^{-it_lw_l}\frac{\overline{\mathcal{F}(\varphi)(t_l)}}{\mathcal{F}(g_l)(2^{j_l}t_l)}dt_l = \prod_{l=1,|w_l|\geq 1}^{d}\int e^{-it_lw_l}\eta_l(t_l)dt_l.$$

Since $|\eta_l(t_l)| \to 0$ when $t_l \to \pm\infty$, an integration by part yields

$$\int e^{-it_lw_l}\eta_l(t_l)dt_l = iw_l^{-1}\int e^{-it_lw_l}\eta_l'(t_l)dt_l.$$

Let us compute the derivative of $\eta_l(t_l)$

$$\eta_l'(t_l) = \frac{\overline{\mathcal{F}(\varphi)(t_l)}'\mathcal{F}(g)(2^{j_l}t_l) - 2^{j_l}\mathcal{F}'(g)(2^{j_l}t_l)\overline{\mathcal{F}(\varphi)(t_l)}}{(\mathcal{F}(g)(2^{j_l}t_l))^2}.$$

Using Lemma 4.5.11, (4.3) and (4.4)

$$\begin{aligned}
|\eta_l'(t_l)| &\leq \left|\frac{\overline{\mathcal{F}(\varphi)(t_l)}'}{\mathcal{F}(g)(2^{j_l}t_l)}\right| + 2^{j_l}\left|\frac{\mathcal{F}'(g)(2^{j_l}t_l)\mathcal{F}(\varphi)(t_l)}{(\mathcal{F}(g)(2^{j_l}t_l))^2}\right| \\
&\leq C\left((1+|t_l|)^{-r}(1+2^{j_l}|t_l|)^{\nu} + 2^{j_l}(1+2^{j_l}|t_l|)^{-\nu-1}(1+|t_l|)^{-r}(1+2^{j_l}|t_l|)^{2\nu}\right) \\
&\leq C\left(2^{j_l\nu}(1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + 2^{j_l}(1+2^{j_l}|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right) \\
&\leq C\left(2^{j_l\nu}(1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + 2^{j_l\nu}(2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right) \\
&\leq C2^{j_l\nu}\left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\int e^{-it_lw_l}\eta_l(t_l)dt_l\right| \\
&\quad\leq |w_l|^{-1}\int|\eta_l'(t_l)|dt_l \\
&\quad\leq C|w_l|^{-1}2^{j_l\nu}\int\left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right)dt_l \\
&\quad\leq C|w_l|^{-1}2^{j_l\nu}(D_1+D_2+D_3),
\end{aligned}$$

with D_1, D_2 and D_3 defined below.

$$\begin{aligned}
D_1 &:= \int_{|t_l|\leq 2^{-j_l}}\left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right)dt_l \\
&\leq C\int_{|t_l|\leq 2^{-j_l}}\left((2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}\right)dt_l \\
&\leq C2^{-j_l}(2^{-j_l\nu}+2^{-j_l(\nu-1)}) \\
&\leq C.
\end{aligned}$$

$$\begin{aligned}
D_2 &:= \int_{2^{-j_l}\leq|t_l|\leq 1}\left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right)dt_l \\
&\leq C\int_{2^{-j_l}\leq|t_l|\leq 1}\left((2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}\right)dt_l \\
&\leq C\int_1^{2^{j_l}}((2^{-j_l}+2^{-j_l}s)^{\nu} + (2^{-j_l}+2^{-j_l}s)^{\nu-1})2^{-j_l}ds \\
&\leq C2^{-j_l(\nu+1)}\int_1^{2^{j_l}}s^{\nu}ds + C2^{-j_l\nu}\int_1^{2^{j_l}}s^{\nu-1}ds \\
&\leq C,
\end{aligned}$$

as soon as $\nu > 0$.

$$\begin{aligned}
D_3 &:= \int_{|t_l|\geq 1}\left((1+|t_l|)^{-r}(2^{-j_l}+|t_l|)^{\nu} + (2^{-j_l}+|t_l|)^{\nu-1}(1+|t_l|)^{-r}\right)dt_i \\
&\leq C\int_{|t_l|\geq 1}\left(|t_l|^{\nu-r} + |t_l|^{\nu-1-r}\right)dt_l \\
&\leq C,
\end{aligned}$$

since $\nu - r \leq -2$.

When $\nu = 0$ we still have

$$\left|\int e^{-it_l w_l}\eta_l(t_l)dt_l\right| \le C|w_l|^{-1}2^{j_l\nu} = C|w_l|^{-1}.$$

Indeed when $\nu = 0$

$$\eta_l(t_l) = \overline{\mathcal{F}(\varphi)(t_l)},$$

and

$$\begin{aligned}\left|iw_l^{-1}\int e^{-it_l w_l}\eta_l'(t_l)dt_l\right| &= \left|iw_l^{-1}\int e^{-it_l w_l}\overline{\mathcal{F}(\varphi)(t_l)}'dt_l\right| \\ &\le |w_l|^{-1}\int\left|\overline{\mathcal{F}(\varphi)(t_l)}'\right|dt_l \\ &\le C|w_l|^{-1}\int(1+|t|)^{-r}dt < C|w_l|^{-1},\end{aligned}$$

using Lemma 4.5.11 and $r \ge 2$.

□

Lemma 4.5.13. *There exist constants R_{10} depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ ,c_g, $\mathcal{C}_g$ and R_{11} depending on φ, c_g, $\mathcal{C}_g$ such that*

$$\sigma_j^2 \le R_{10}2^{S_j(2\nu+1)}, \quad \|T_j\|_\infty \le R_{11}2^{S_j(\nu+1)}.$$

Proof. We have

$$\begin{aligned}\sigma_j^2 &= \mathbb{V}ar(U_j(Y_1,W_1)) \\ &\le \mathbb{E}\left[|U_j(Y_1,W_1)|^2\right] \\ &= \mathbb{E}\left[\left|Y_1\sum_k(\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\ &= \mathbb{E}\left[\left|(m(X_1)+\varepsilon_1)\sum_k(\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\ &\le 2(\|m\|_\infty^2+\sigma_\varepsilon^2)\mathbb{E}\left[\left|\sum_k(\mathcal{D}_j\varphi)_{j,k}(W_1)\varphi_{jk}(x)\right|^2\right] \\ &\le 2(\|m\|_\infty^2+\sigma_\varepsilon^2)\int\left|\sum_k(\mathcal{D}_j\varphi)_{j,k}(w)\varphi_{jk}(x)\right|^2 f_W(w)dw \\ &\le 2(\|m\|_\infty^2+\sigma_\varepsilon^2)\|f_X\|_\infty\int 2^{S_j}\left|\sum_k(\mathcal{D}_j\varphi)\,(2^jw-k)\varphi_{jk}(x)\right|^2 dw.\end{aligned}$$

Now making the change of variable $z = 2^j w - k$, we get by using Lemma 4.5.10 and Lemma 4.5.12 to bound $(\mathcal{D}_j\varphi)(z)$

$$\begin{aligned}\sigma_j^2 &\leq 2(\|m\|_\infty^2 + \sigma_\varepsilon^2)\|f_X\|_\infty \int \left|\sum_k (\mathcal{D}_j\varphi)(z)\varphi_{jk}(x)\right|^2 dz \\ &\leq C\int 2^{2S_j\nu}\prod_{i=l}^{d}\frac{1}{(1+|z_l|)^2}\left(\sum_k |\varphi_{jk}(x)|\right)^2 dz \\ &\leq R_{10}2^{S_j(2\nu+1)},\end{aligned}$$

where R_{10} is a constant depending on s, $\mathfrak{m}$, $\mathfrak{d}$, φ ,c_g, $\mathcal{C}_g$. This gives the bound for σ_j^2.

For $\|T_j\|_\infty$, using again Lemma 4.5.10 and Lemma 4.5.12, we have

$$\begin{aligned}\|T_j\|_\infty &\leq \max_k \|(\mathcal{D}_j\varphi)_{j,k}\|_\infty \sum_k |\varphi_{jk}(x)| \leq 2^{\frac{S_j}{2}}\|(\mathcal{D}_j\varphi)\|_\infty \sum_k |\varphi_{jk}(x)| \\ &\leq R_{11}2^{S_j(\nu+1)},\end{aligned}$$

where R_{11} is a constant depending on φ, c_g, $\mathcal{C}_g$.

□

4.6 Perspective

In this chapter, we study the problem of adaptive estimation of multivariate regression function in the errors-in-variable model:

$$Y_l = m(X_l) + \epsilon_l, \quad W_l = X_l + \delta_l, \quad l = 1, \dots, n.$$

We propose a wavelet-based kernel estimator and obtain the optimal rates of convergence over anisotropic Hölder classes where the distribution of the errors δ_l's is ordinary smooth. It would be interesting if we complete the study with the supersmooth case on the density of the errors covariates g, $e.g$ we suppose that there exist $c_g, C_g > 0$, $\nu \in \mathbb{R}^d$, $\varpi, \varrho \in (\mathbb{R}_+^*)^d$ such that $\forall t \in \mathbb{R}^d$

$$c_g \prod_{l=1}^{d}(1+|t_l|)^{-\nu_l}\exp(-\varpi_l|t_l|^{\varrho_l}) \leq |\mathcal{F}(g)(t)| \leq C_g \prod_{l=1}^{d}(1+|t_l|)^{-\nu_l}\exp(-\varpi_l|t_l|^{\varrho_l}).$$

Standard examples of supersmooth densities are Gaussian and Cauchy distributions. Furthermore, most studies of errors-in-variables model assume that the distribution of the errors is known. However, in practice the noise density may be unknown and there are just a few studies in literature which have been investigated the errors-in-variables model with unknown error densities. We refer to the works of Delaigle et al. [32], Delaigle and Meister [33] and Linto and Whang [76] for related studies of univariate nonparametric regression in the case of unknown error distribution where these authors use replicated observations to construct an estimator for the unknown error density. Thus an extension to the case of unknown distribution of the errors in the multidimensional setting should be considered for future research.

A second perspective is that we consider the errors-in-variables model where the noise in the covariates is multiplicative. We aim to investigate the following model in the multidimensional setting:

$$\begin{cases} Y_i = m(X_i) + \epsilon_i, \\ W_i = X_i \odot \delta_i, \quad i = 1, \ldots, n, \end{cases} \tag{4.37}$$

where $(X_1, \ldots, X_n)$, $(\epsilon_1, \ldots, \epsilon_n)$ and $(\delta_1, \ldots, \delta_n)$ are i.i.d $\mathbb{R}^d$-valued vectors and $\odot$ is the Hadamard product (see [60, 93]) defined by $A \odot B = \Big(a_{ij} b_{ij}\Big)_{1 \le i,j \le n}$. Multiplicative errors-in-variables model is sparsely studied in statistics, but only in the parametric setting. We may cite here the works of Hwang [62] who considered a linear model with multiplicative errors in variables to analyse the data collected by the Department of Energy concerning energy consumption and housing characteristics of households in the United States where some of the predicting variables in the original data have been multiplied by random number to preserve confidentiality, and Nguyen et al. [88] who consider a linear mixed model where the data are contaminated by multiplicative errors. Other studies on multiplicative high-dimensional linear regression model can be found in Arellano-Valle et al. [4] and Loh and Wainwright [77]. Moreover, to the best of our knowledge, there is no study about nonparametric multiplicative errors-in-variables model. Hence, it would be interesting if we explore the errors-in-variables model with multiplicative errors together with a practical application.

References

[1] M. Ackermann, S. C. Stearns, and U. Jenal. Senescence in a bacterium with asymmetric division. *Science*, 300(5627):1920–1920, 2003.

[2] H. Aguilaniu, L. Gustafsson, M. Rigoulet, and T. Nyström. Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science*, 299(5613):1751–1753, 2003.

[3] D. Aldous. Stopping times and tightness. *The Annals of Probability*, 6(2):335–340, 1978.

[4] R. B. Arellano-Valle, H. Bolfarine, and L. Gasco. Measurement error models with nonconstant covariance matrices. *Journal of Multivariate Analysis*, 82(2):395 – 415, 2002.

[5] K. B. Athreya and P. E. Ney. *Branching processes*, volume 196. Springer Science & Business Media, 2012.

[6] A. Bakushinskii. On the construction of regularizing algorithms under random noise. In *Soviet Math. Doklady*, volume 189, pages 231–233, 1969.

[7] V. Bansaye. Proliferating parasites in dividing cells: Kimmel's branching model revisited. *The Annals of Applied Probability*, pages 967–996, 2008.

[8] V. Bansaye, J.-F. Delmas, L. Marsalle, and V. C. Tran. Limit theorems for markov processes indexed by continuous time galton-watson trees. *The Annals of Applied Probability*, 21, 2011.

[9] V. Bansaye and S. Méléard. *Some stochastic models for structured populations : scaling limits and long time behavior.* arXiv:1506.04165, 2015.

[10] V. Bansaye, J. C. P. Millan, C. Smadi, et al. On the extinction of continuous state branching processes with catastrophes. *Electron. J. Probab*, 18(106):1–31, 2013.

[11] V. Bansaye and V. C. Tran. Branching feller diffusion for cell division with parasite infection. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8, 2011.

[12] B. Bercu, B. De Saporta, and A. Gegout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability*, 14(87):2492–2526, 2009.

[13] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging.* CRC press, 1998.

[14] K. Bertin, C. Lacour, and V. Rivoirard. Adaptive pointwise estimation of conditional density function. *Ann. Inst. H. PoincarÃĽ Probab. Statist.*, 52(2):939–980, 05 2016.

[15] N. Bissantz and H. Holzmann. Statistical inference for inverse problems. *Inverse Problems*, 24(3):034009, 2008.

[16] T. Bourgeron, M. Doumic, and M. Escobedo. Estimating the division rate of the growth-fragmentation equaion with a self-similar kernel. *Inverse Problems*, 30, 2014.

[17] A. Calsina and M. Sanchón. Stability and instability of equilibria of an equation of size structured population dynamics. *Journal of Mathematical Analysis and Applications*, 286(2):435 – 452, 2003.

[18] R. J. Carroll, A. Delaigle, and P. Hall. Nonparametric prediction in measurement error models. *J. Amer. Statist. Assoc.*, 104(487):993–1003, 2009.

[19] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

[20] L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21(4):1345, 2005.

[21] G. Chagny. Penalization versus goldenshluger- lepski strategies in warped bases regression. *ESAIM: Probability and Statistics*, 17:328–358, 2013.

[22] C. Chesneau. On adaptive wavelet estimation of the regression function and its derivatives in an errors-in-variables model. *Curr. Dev. Theory Appl. Wavelets*, 4(2):185–208, 2010.

[23] M. Chichignoud, V. H. Hoang, T. M. Pham Ngoc, and V. Rivoirard. Adaptive wavelet multivariate regression with errors in variables. *hal-01253508*, 2016.

[24] B. Cloez. Limit theorems for some branching measure-valued processes. *arXiv preprint arXiv:1106.0660*, 2011.

[25] F. Comte, V. Genon-Catalot, and A. Samson. Nonparametric estimation for stochastic differential equations with random effects. *Stochastic Processes and their Applications*, 123(7):2522–2551, 2013.

[26] F. Comte and C. Lacour. Data driven density estimation in presence of unknown convolution operator. *Journal of the Royal Statistical Society: Series B*, 73(4):601–627, 2011.

[27] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):569–609, 2013.

[28] F. Comte, A. Samson, and J. J. Stirnemann. Deconvolution estimation of onset of pregnancy with replicate observations. *Scandinavian Journal of Statistics*, 41(2):325–345, 2014.

[29] F. Comte and M.-L. Taupin. Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statist. Sinica*, 17(3):1065–1090, 2007.

[30] D. De Canditiis and M. Pensky. Simultaneous wavelet deconvolution in periodic setting. *Scandinavian Journal of Statistics*, 33(2):293–306, 2006.

[31] A. Delaigle, P. Hall, and F. Jamshidi. Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):149–169, 2015.

[32] A. Delaigle, P. Hall, H.-G. Müller, et al. Accelerated convergence for nonparametric regression with coarsened predictors. *The Annals of Statistics*, 35(6):2639–2653, 2007.

[33] A. Delaigle and A. Meister. Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association*, 102(480):1416–1426, 2007.

[34] J.-F. Delmas and L. Marsalle. Detection of cellular aging in a galton–watson process. *Stochastic Processes and their Applications*, 120(12):2495–2519, 2010.

[35] O. Diekmann, M. Gyllenberg, J. J. A. Metz, and R. H. Thieme. On the formulation and analysis of general deterministic structured population modelsi. linear theory. *Journal of Mathematical Biology*, 36(4):349–388, 1998.

[36] S. Donnet and A. Samson. Estimation of parameters in incomplete data models defined by dynamical systems. *Journal of Statistical Planning and Inference*, 137(9):2815–2831, 2007.

[37] M. Doumic and P. Gabriel. Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, 20(05):757–783, 2010.

[38] M. Doumic, M. Hoffmann, N. Krell, L. Robert, et al. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3):1760–1799, 2015.

[39] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis*, 50(2):925–950, 2012.

[40] M. Doumic, B. Perthame, and J. P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25(4):045008, 2009.

[41] M. Doumic and L. M. Tine. Estimating the division rate for the growth-fragmentation equation. *Journal of mathematical biology*, 67(1):69–103, 2013.

[42] L. Du, C. Zou, and Z. Wang. Nonparametric regression function estimation for errors-in-variables models with validation data. *Statist. Sinica*, 21(3):1093–1113, 2011.

[43] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[44] S. N. Evans and D. Steinsaltz. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoretical population biology*, 71(4):473–490, 2007.

[45] J. Fan and J.-Y. Koo. Wavelet deconvolution. *IEEE Trans. Inform. Theory*, 48(3):734–747, 2002.

[46] J. Fan and E. Masry. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *J. Multivariate Anal.*, 43(2):237–271, 1992.

[47] J. Fan and Y. K. Truong. Nonparametric regression with errors in variables. *Ann. Statist.*, 21(4):1900–1925, 1993.

[48] J. Z. Farkas and T. Hagen. Stability and regularity results for a size-structured population model. *Journal of Mathematical Analysis and Applications*, 328(1):119 – 136, 2007.

[49] R. Ferrière and V. C. Tran. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM: Proc.*, 27:289–310, 2009.

[50] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Annals of applied probability*, 14:1880–1919, 2004.

[51] F. Gach, R. Nickl, and V. Spokoiny. Spatially adaptive density estimation by localised Haar projections. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):900–914, 2013.

[52] A. Goldenshluger and O. Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 11 2008.

[53] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: orcale inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.

[54] J. Guyon. Limit theorems for bifurcating markov chains. application to the detection of cellular aging. *The Annals of Applied Probability*, 17(5/6):1538–1569, 2007.

[55] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. JSTOR, 1925.

[56] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.

[57] T. E. Harris. *The theory of branching processes*. Springer, Berlin, 1963.

[58] V. H. Hoang. Estimating the division kernel of a size-structured population. 2015.

[59] M. Hoffmann and A. Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and their Applications*, 126(5):1433–1471, 2016.

[60] R. A. Horn. The hadamard product. In *Proc. Symp. Appl. Math*, volume 40, pages 87–169, 1990.

[61] C. Houdre and P. Reynaud-Bouret. Exponential inequalities, with constants, for v-statistics of order two. *Progress in Probability*, page 55, 05.

[62] J. T. Hwang. Multiplicative errors-in-variables models with applications to recent data released by the us department of energy. *Journal of the American Statistical Association*, 81(395):680–688, 1986.

[63] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. volume 24 of North-Holland Mathematical Library.

[64] D. A. Ioannides and P. D. Alevizos. Nonparametric regression with errors in variables and applications. *Statist. Probab. Lett.*, 32(1):35–43, 1997.

[65] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.

[66] K. P. Jagdish and B. Campbell. *Handbook of the normal distribution*. Marcel Dekker Inc., New York, NY, 1996.

[67] P. Jagers. A general stochastic model for population development. *Scandinavian Actuarial Journal*, 1969(1-2):84–103, 1969.

[68] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, pages 20–65, 1986.

[69] I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):547–573, 2004.

[70] J. Kalifa and S. Mallat. Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, 31(1):58–109, 2003.

[71] G. Kerkyacharian, D. Picard, and M. Raimondo. Adaptive boxcar deconvolution on full lebesgue measure sets. *Statistica Sinica*, 17(1):317–340, 2007.

[72] J.-Y. Koo and K.-W. Lee. B-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.*, 40(1):57–66, 1998.

[73] C. Lacour and P. Massart. Minimal penalty for goldenshluger–lepski method. *Stochastic Processes and their Applications*, 2016.

[74] C.-Y. Lai, E. Jaruga, C. Borghouts, and S. M. Jazwinski. A mutation in the atp2 gene abrogates the age asymmetry between mother and daughter cells of the yeast saccharomyces cerevisiae. *Genetics*, 162(1):73–87, 2002.

[75] A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart, and F. Taddei. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences*, 105(8):3076–3081, 2008.

[76] O. Linton and Y.-J. Whang. Nonparametric estimation with aggregated data. *Econometric Theory*, 18(02):420–468, 2002.

[77] B. P.-L. Loh and M. J. Wainwright. High-dimensional regression with noisy and missing data: provable guarantees wth nonconvexity. *The Annals of Statistics*, 40(3):1637–1664, 2012.

[78] P. Massart. *Concentration Inequalities and Model Selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, 6-23, 2003. Springer, 2007.

[79] A. Meister. *Deconvolution problems in nonparametric statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2009.

[80] S. Méléard and S. Roelly. Sur les convergences étroite ou vague de processus à valeur mesures. *Comptes Rendus de l'Academie des Sciences*, Serie I 317, 1993.

[81] S. Méléard and V. C. Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications*, 122, 2012.

[82] S. P. Meyn and R. L. Tweedie. Stability of markovian processes iii: Foster-lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, pages 518–548, 1993.

[83] P. Michel. Existence of a solution to the cell division eigenproblem. *Mathematical Models and Methods in Applied Sciences*, 16(supp01):1125–1153, 2006.

[84] P. Michel, S. Mischler, and B. Perthame. General relative entropy inequality: an illustration on growth models. *Journal de mathématiques pures et appliquées*, 84(9):1235–1260, 2005.

[85] K. Mosegaard and A. Tarantola. Probabilistic approach to inverse problems. In *International Handbook of Earthquake and Engineering Seismology*. Academic Press, Incorporated, 2002.

[86] J. B. Moseley. Cellular aging: symmetry evades senescence. *Current Biology*, 23(19):R871–R873, 2013.

[87] M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.

[88] D. V. Nguyen, D. Şentürk, and R. J. Carroll. Covariate-adjusted linear mixed effects model with an application to longitudinal data. *Journal of nonparametric statistics*, 20(6):459–481, 2008.

[89] S. V. B. Penda. Deviation inequalities for bifurcating markov chains on galton-watson tree. *ESAIM: Probability and Statistics*, 19:689–724, 2015.

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

[91] B. Perthame. *Transport equations arising in biology*. Frontiers in Mathematics. Birchkhauser, 2007.

[92] B. Perthame and L. Ryzhik. Exponential decay for the fragmentation or cell-division equation. *Journal of Differential equations*, 210(1):155–177, 2005.

[93] R. Reams. Hadamard inverses, square roots and products of almost semidefinite matrices. *Linear Algebra and its Applications*, 288:35 – 43, 1999.

[94] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience*, 4(1):1, 2014.

[95] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, and M. Doumic. Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC biology*, 12(1):1, 2014.

[96] S. M. Ross. *Stochastic processes*, volume 2. John Wiley & Sons New York, 1996.

[97] A. Samson and M. Thieullen. A contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122(7):2521–2552, 2012.

[98] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[99] E. J. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and death in an organism that reproduces by morphologically symmetric division. *PLOS Biology*, 3, 2005.

[100] J. Stirnemann, F. Comte, and A. Samson. Density estimation of a biomedical variable subject to measurement error using an auxiliary set of replicate observations. *Statistics in medicine*, 31(30):4154–4163, 2012.

[101] V. Sudakov and L. Khalfin. Statistical approach to correctness of problems in mathematical physics. *DOKLADY AKADEMII NAUK SSSR*, 157(5):1058, 1964.

[102] A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* siam, 2005.

[103] A. N. Tikhonov. Regularization of incorrectly posed problems. SOVIET MATHEMATICS DOKLADY, 1963.

[104] A. N. Tikhonov and V. Y. Arsenin. Solutions of ill-posed problems. 1977.

[105] V. C. Tran. *Modèles particulaires stochastiques pour des problèmes d'évolution adaptive et pour l'approximation de solutions statisques.* PhD dissertation, Université Paris X - Nanterre, http://tel.archives-ouvertes.fr/tel-00125100, 2006.

[106] V. C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics*, 12, 2008.

[107] V. C. Tran. *Une ballade en forêts aléatoires : Théorèmes limites pour des populations structurées et leurs généalogies, étude probabiliste et statistique de modèles SIR en épidémiologie, contributions à la géométrie aléatoire.* HDR dissertation, Université Lille 1, https://tel.archives-ouvertes.fr/tel-01087229v1, 2014.

[108] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer series in Statistics. Springer, New York, 2009.

[109] A. S. Whittemore and J. B. Keller. Approximations for regression with covariate measurement error. *Journal of the American Statistical Association*, 83(404):1057–1066, 1988.

[110] L. Wu. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *Journal of the American Statistical Association*, 99(467):700–709, 2004.

Titre: Estimation adaptative pour des problèmes inverses avec des applications à la division cellulaire

Résumé : Cette thèse se divise en deux parties indépendantes. Dans la première, nous considérons un modèle stochastique individu-centré en temps continu décrivant une population structurée par la taille. Ce modèle est motivé par la modélisation des divisions cellulaires et par la détection du vieillissement cellulaire en biologie. La population est représentée par une mesure ponctuelle évoluant suivant un processus aléatoire déterministe par morceaux. Nous étudions ici l'estimation non paramétrique du noyau régissant les divisions, sous deux schémas d'observation différents. Premièrement, nous observons l'évolution des cellules jusqu'au temps T et nous obtenons l'arbre entier des divisions. Nous construisons un estimateur à noyau avec une sélection adaptative de fenêtre dépendante des données. Nous obtenons une inégalité oracle et des vitesses de convergence exponentielles optimales. Deuxièmement, dans le cas où l'arbre de division n'est pas complètement observé, nous montrons que le processus microscopique renormalisé décrivant l'évolution de la population converge vers la solution faible d'une équation aux dérivés partielles (EDP). En considérant un probléme de valeurs propres lié à l'étude du comportement asymptotique des solutions de cette EDP, nous proposons un estimateur du noyau de division en utilisant des techniques de Fourier. Nous montrons la consistance de l'estimateur. L'étude de la vitesse de convergence est un travail en cours.

Dans la seconde partie de la thèse, nous considérons le modèle de régression non paramétrique avec erreurs sur les variables dans le contexte multidimensionnel. Notre objectif est d'estimer la fonction de régression multivariée inconnue. Nous proposons un estimateur adaptatif basé sur des noyaux de projection fondés sur une base d'ondelettes multi-index et sur un opérateur de déconvolution. Le niveau de résolution des ondelettes est obtenu par la méthode de Goldenshluger-Lepski. Nous obtenons une inégalité oracle et des vitesses de convergence optimales sur les espaces de Hölder anisotropes.

Mots clés : Population structurée par la taille, noyau de division, estimation non-paramétrique, méthode de Goldenshluger et Lepski, adaptation, ondelettes, déconvolution, régression anisotrope, erreurs de mesure.

Title: Adaptive estimation for inverse problems with applications to cell divisions

Abstract: This thesis is divided into two independent parts. In the first one, we consider a stochastic individual-based model in continuous time to describe a size-structured population for cell divisions. This model is motivated by the detection of cellular aging in biology. The random point measure describing the cell population evolves as a piecewise deterministic Markov process. We address here the problem of nonparametric estimation of the kernel ruling the divisions, under two observation schemes. First, we observe the evolution of cells up to a fixed time T and we obtain the whole division tree. We construct an adaptive kernel estimator of the division kernel with a fully data-driven bandwidth selection. We obtain an oracle inequality and optimal exponential rates of convergence. Second, when the whole division tree is not completely observed, we show that, in a large population limit, the renormalized microscopic process describing the evolution of cells converges to the weak solution of a partial differential equation (PDE). Considering an eigenvalue problem related to the asymptotic behavior of the PDE's solutions, we propose an estimator of the division kernel by using Fourier techniques. We prove the consistency of the estimator. The study of rates of convergence is a work in progress.

In the second part of this thesis, we consider the nonparametric regression with errors-in-variables model in the multidimensional setting. We estimate the multivariate regression function by an adaptive estimator based on projection kernels defined with multi-indexed wavelets and a deconvolution operator. The wavelet level resolution is selected by the method of Goldenshluger-Lepski. We obtain an oracle inequality and optimal rates of convergence over anisotropic Hölder classes.

Keywords: Random size-structured population, division kernel, nonparametric estimation, Goldenshluger-Lepski's method, adaptation, wavelets, deconvolution, anisotropic regression, measurement errors.

