ESAIM: PS 21 (2017) 275–302
DOI: 10.1051/ps/2017011

ESAIM: Probability and Statistics
www.esaim-ps.org

ESTIMATING THE DIVISION KERNEL OF A SIZE-STRUCTURED POPULATION

Van Ha Hoang[1]

Abstract. We consider a size-structured model describing a population of cells proliferating by division. Each cell contain a quantity of toxicity which grows linearly according to a constant growth rate α. At division, the cells divide at a constant rate R and share their content between the two daughter cells into fractions Γ and $1-\Gamma$ where Γ has a symmetric density h on $[0,1]$, since the daughter cells are exchangeable. We describe the cell population by a random measure and observe the cells on the time interval $[0,T]$ with fixed T. We address here the problem of estimating the division kernel h (or fragmentation kernel) when the division tree is completely observed. An adaptive estimator of h is constructed based on a kernel function K with a fully data-driven bandwidth selection method. We obtain an oracle inequality and an exponential convergence rate, for which optimality is considered.

Mathematics Subject Classification. 60J80, 60J85, 62G05, 62G07, 92D25.

Received April 3, 2017. Revised June 14, 2017. Accepted June 14, 2017.

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 [3], Harris [18], Jagers [29]). 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.* [34], 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 to the concept of aging for cell lineage. This concept has been tackled in many papers (*e.g.* Ackermann *et al.* [2], Aguilaniu *et al.* [1], Lai *et al.* [23], Evans and Steinsaltz [15], Moseley [28]).

Here we consider a stochastic individual-based model of size-structured population in continuous time, where individuals are cells undergoing 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 linearly 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. Figure 1 illustrates trajectories of two daughter cells after a division. If $\Gamma = 1/2$, the two daughters

Keywords and phrases. Random size-structured population, division kernel, nonparametric estimation, Goldenshluger-Lepski's method, adaptive estimator, penalization.

[1] Laboratoire Paul Painlevé UMR CNRS 8524, Université de Lille 1, France. van-ha.hoang@math.univ-lille1.fr

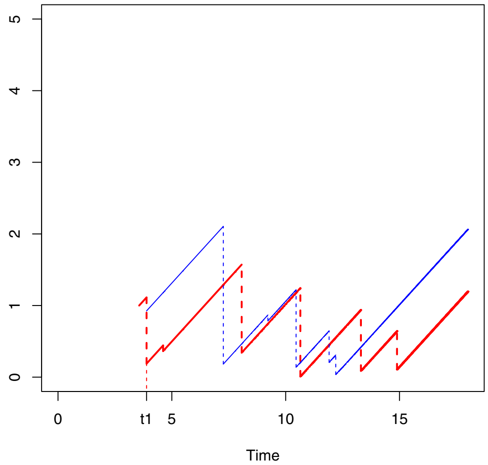


FIGURE 1. Trajectories of two daughter cells after a division, separating after the first division at time t_1.

both inherit toxicity $x/2$. If the distribution of Γ does not concentrate on the singleton $1/2$, divisions might be unequal. Here, we assume that the distribution of Γ admits a density h with respect to the Lebesgue measure. Because the daughter cells are exchangeable, we can assume that h is a symmetric function with respect to $1/2$ on $[0,1]$. When the density of h is concentrated around $1/2$, both daughters contain similar toxicities, close to the half of their mother's toxicity. The more h puts weight in the neighbourhood of 0 and 1, the more unequal the divisions are, with one daughter having little toxicity and the other a toxicity close to its mother's one. Modifications of this model to account for more complex phenomena have been considered in other papers. Bansaye and Tran [6], Cloez [11] or Tran [36] consider non-constant division and growth rates. Doumic *et al.* [13] studies a size-structured model in which the division rate depends on size of individuals growing exponential in time. Robert *et al.* [30] 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 questions as here, see for instance Bansaye *et al.* [4,7], Bercu *et al.* [8], Bitseki Penda [10], Delmas and Marsalle [12] or Guyon [17].

Individual-based models provide a natural framework for statistical estimation. Estimation of the division rate is, for instance, the subject of Doumic *et al.* [13,14] and Hoffmann and Olivier [20]. Here, the density h is the kernel division that we want to estimate. The interest of estimating h is to see whether the sharing is rather equal or unequal since the kurtosis of h provides indication on the age of lineage. If we consider that having a lot of toxicity is a kind of senescence, then, the knowledge of h allows to detect aging phenomena (see Lindner *et al.* [25]).

Assuming that we observe all the divisions of cells occurring 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 2.8. The construction of $\hat{h}$ is detailed in the sequel. From the oracle inequality we can infer adaptive exponential rates of convergence $\exp\left(-\frac{2\beta}{2\beta+1}RT\right)$ 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 [37]) 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 Sect. 2.2). By comparison, in [20] Hoffmann and Olivier obtain a similar rate of convergence for the kernel estimator of the division rate when the latter is not constant. However, their estimator is not adaptive since the choice of their optimal bandwidth still depends on the smoothness of the estimated function. The main difficulty for adaptation in their setting comes from the fact that the number of observations as well as the probability of inclusion of an individual both depend on the unknown division rate. In this paper, we do not consider these issues, but we aim at adaptivity and the kernel estimator $\hat{h}$ still depends on a random number of observations. Our estimator is adaptive with an "optimal" bandwidth chosen from a data-driven method. We derive upper bounds and lower bounds for the MISE (Mean Integrated Squared Error) of $\hat{h}$. The rate of convergence of our estimator $\hat{h}$ proves to be optimal in the minimax sense on the Hölder classes.

This paper 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. Our main results are stated in Theorems 2.11 and 2.12. Numerical results 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 $\|\cdot\|_2$ respectively denote the $\mathbb{L}^1$ and $\mathbb{L}^2$ norms on $\mathbb{R}$ with respect to the Lebesgue measure:

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

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

$$\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|.$$

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

$$f \star g(x) = \int_{\mathbb{R}} f(y) g(x-y) dy, \quad x \in \mathbb{R}.$$

2. Microscopic model and kernel estimator of h

2.1. The model

We recall the Ulam−Harris−Neveu notation used to describe a genealogical tree. The first cell is labelled by $\emptyset$ and when the cell j divides, the two descendants are labelled by $j0$ and $j1$. 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 living 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 let X_t^j be the quantity of toxicity in the cell j living at time t, we describe the population of cells at time t by a random point measure in $\mathcal{M}_F(\mathbb{R}_+)$:

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

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 $\langle \mu, f \rangle = \int_{\mathbb{R}_+} f \mathrm{d}\mu$.

We assume that between the birth time of cell j and its splitting time, the toxicity X_t^j grows linearly with the constant rate α,

$$\mathrm{d}X_t^j = \alpha \mathrm{d}t. \tag{2.3}$$

When the cells divide, the toxicity is shared between the daughter cells. This is described by the following stochastic differential equation (SDE) (2.5).

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 $Q(\mathrm{d}s, \mathrm{d}j, \mathrm{d}\gamma)$ be a Poisson point measure on $\mathbb{R}_+ \times \mathcal{E} := \mathbb{R}_+ \times J \times [0,1]$ with intensity $q(\mathrm{d}s, \mathrm{d}j, \mathrm{d}\gamma) = Rh(\gamma)\,\mathrm{d}s\, n(\mathrm{d}j)\mathrm{d}\gamma$. Here $n(\mathrm{d}j)$ is the counting measure on J and $\mathrm{d}s$ is the 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 (see Bansaye *et al.* [5]).

Definition 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:

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

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. Simulations and results on the stabilization of the mean toxicity using this model are presented in the author's Ph.D. thesis (see [19], Sects. 2.2.2 and 2.3.2).

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

Proposition 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_i be the i^{th} jump time. Then*

$$\lim_{i\to+\infty} T_i = +\infty \text{ and } \lim_{T\to+\infty} N_T = +\infty \quad (a.s).$$

(ii) *When $N_0 > 1$, N_T is distributed according to a negative binomial distribution, denoted as $\mathcal{NB}(N_0, \mathrm{e}^{-RT})$. Its probability mass function is then*

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

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

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

Consequently, we have $\mathbb{E}\big[N_T\big] = N_0 \mathrm{e}^{RT}$.

(iii) *When $N_0 = 1$:*

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

When $N_0 > 1$, we have

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

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

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

The proof of Proposition 2.2 is presented in Appendix.

2.2. 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 $X_{t_i}^{j_i0}$ and $X_{t_i}^{j_i1}$ and define

$$\Gamma_i^0 = \frac{X_{t_i}^{j_i0}}{X_{t_i-}^{j_i}} \quad \text{and} \quad \Gamma_i^1 = \frac{X_{t_i}^{j_i1}}{X_{t_i-}^{j_i}},$$

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

Observe that 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 Γ^1 has the density $h(\gamma)$ and $\Gamma^0=1-\Gamma^1$.

The goal is to estimate h. For this purpose, we naturally use kernel estimates. Assume that we observe $(\Gamma_1^1,\dots,\Gamma_{M_T}^1)$ where $M_T>0$ is the random number of divisions in $[0,T]$, we define the kernel estimator of h as follows.

Definition 2.3. Let $K:\mathbb{R}\longrightarrow\mathbb{R}$ be an integrable function such that

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

Define

$$\hat{h}_\ell(\gamma) = \frac{1}{M_T}\sum_{i=1}^{M_T} K_\ell(\gamma-\Gamma_i^1), \quad \gamma\in(0,1), \tag{2.6}$$

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

Remark 2.4. Since $N_0\neq 0$, the number of random divisions M_T is different from the number of individuals living at time T. More precisely, we have $M_T=N_T-N_0$.

In (2.6), $\hat{h}_\ell$ depends also on T. However, we omit T for the sake of notation. The estimator $\hat{h}_\ell$ satisfies the following properties.

Proposition 2.5. *For all $\gamma\in(0,1)$, let $\mathbb{E}\big[\hat{h}_\ell(\gamma)|M_T\big]$ and $\mathbb{V}\mathrm{ar}\big[\hat{h}_\ell(\gamma)\big|M_T\big]$ respectively be the conditional expectation and conditional variance given M_T of $\hat{h}_\ell(\gamma)$. We have*

(i)
$$\mathbb{E}\big[\hat{h}_\ell(\gamma)|M_T\big] = K_\ell\star h(\gamma) \ \textit{ and } \ \mathbb{E}\big[\hat{h}_\ell(\gamma)\big] = K_\ell\star h(\gamma).$$

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

(ii)
$$\mathbb{V}\mathrm{ar}\big[\hat{h}_\ell(\gamma)\big|M_T\big] = \frac{1}{M_T}\mathbb{V}\mathrm{ar}\left[K_\ell(\gamma-\Gamma_1^1)\right].$$

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

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

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

Let $\hat{h}_\ell$ be the kernel estimator of h as in Definition 2.3. 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 2.6. *The $\mathbb{L}^2$-loss of $\hat{h}_\ell$ given M_T satisfies:*

$$\mathbb{E}\Big[\|\hat{h}_\ell - h\|_2 \big| M_T\Big] \leq \|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T \ell}}. \tag{2.7}$$

In the right hand side of the risk decomposition (2.7) 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 possible bandwidths H which will be defined later, the best bandwidth $\bar{\ell}$ would be

$$\bar{\ell} := \underset{\ell \in H}{\operatorname{argmin}} \left\{ \|h - K_\ell \star h\|_2 + \frac{\|K\|_2}{\sqrt{M_T \ell}} \right\}. \tag{2.8}$$

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 [16] 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. In a similar fashion, Doumic *et al.* [14] and Reynaud−Bouret *et al.* [32] applied this methodology. To apply the GL method, we set for any ℓ, $\ell' \in H$:

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

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

Definition 2.7. Given $\epsilon > 0$ and setting $\chi := (1+\epsilon)(1+\|K\|_1)$, we define

$$\hat{\ell} := \underset{\ell \in H}{\operatorname{argmin}} \left\{ A(\ell) + \frac{\chi \|K\|_2}{\sqrt{M_T \ell}} \right\}, \tag{2.9}$$

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

$$A(\ell) := \sup_{\ell' \in H} \left\{ \|\hat{h}_{\ell,\ell'} - \hat{h}_{\ell'}\|_2 - \frac{\chi \|K\|_2}{\sqrt{M_T \ell'}} \right\}_+, \tag{2.10}$$

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

$$\hat{h} := \hat{h}_{\hat{\ell}}. \tag{2.11}$$

An inspection of the proof of Theorem 2.8 which follows the term $A(\ell)$ provides a control for the bias $\|h - K_\ell \star h\|_2$ up to the term $\|K\|_1$ (see (4.5) and (4.6) in the proof of Thm. 2.8, Sect. 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 2.8. *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 $[0, T]$. Define*

$$\varrho(T)^{-1} = \begin{cases} \dfrac{\mathrm{e}^{-RT + \log(RT)}}{1 - \mathrm{e}^{-RT}}, & \textit{if} \quad N_0 = 1, \\ \mathrm{e}^{-RT}, & \textit{if} \quad N_0 > 1. \end{cases} \tag{2.12}$$

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

Assume $h \in \mathbb{L}^\infty([0,1])$. Consider H a countable subset of $\{\triangle^{-1} : \triangle = 1, \dots, \triangle_{\max}\}$ in which we choose the bandwidths and $\triangle_{\max} = \lfloor \delta \varrho(T) \rfloor$ for some $\delta > 0$. Let $\hat{h}$ be the kernel estimator defined with the kernel $K_{\hat{\ell}}$ where $\hat{\ell}$ is chosen by the GL method. Then, given $\epsilon > 0$

$$\mathbb{E}\Big[\|\hat{h} - h\|_2^2\Big] \leq C_1 \inf_{\ell \in H} \left\{ \|K_\ell \star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell} \varrho(T)^{-1} \right\} + C_2 \varrho(T)^{-1}, \tag{2.13}$$

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 $\|h\|_\infty$.

In view of iv) of Proposition 2.2 and Remark 2.4, recall that $\mathbb{E}\big[1/N_T\big] \leq \mathrm{e}^{-RT}/(N_0 - 1)$ and $M_T = N_T - N_0$. Since $\varrho(T)^{-1}$ is of the order of $\mathbb{E}\big[1/N_T\big]$, the last term of (2.13) is negligible. Hence the estimate $\hat{h}$ mimics the performances of the oracle estimator.

We now establish upper and lower bounds for the MISE. The lower bound is obtained by perturbation methods (Thm. 2.12) and is valid for any estimator $\widehat{h}_T$ of h, thus indicating the optimal convergence rate. The upper bound is obtained in Theorem 2.11 thanks to the key oracle inequality of Theorem 2.8.

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

Definition 2.9. Let $\beta > 0$ and $L > 0$. The Hölder class of smoothness β and radius L is defined by

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

Definition 2.10. Let $\beta^* > 0$. An integrable function $K : \mathbb{R} \to \mathbb{R}$ is a kernel of order β^* if

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

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

Theorem 2.11. *Let $\beta \in (0, \beta^*)$ and let $\mathfrak{h} > 1$. Let K be a kernel of order β^* and let $\hat{\ell}$ be the adaptive bandwidth defined in (2.9). Then, for any $T > 0$, the kernel estimator $\hat{h}$ satisfies*

$$\sup_{\substack{h \in \mathcal{H}(\beta, L) \\ \|h\|_\infty \leq \mathfrak{h}}} \mathbb{E}\|\hat{h} - h\|_2^2 \leq C_3 \varrho(T)^{-\frac{2\beta}{2\beta+1}}, \tag{2.14}$$

where $\varrho(T)^{-1}$ is defined in (2.12) and C_3 is a constant depending on N_0, δ, ϵ, $\|K\|_1$, $\|K\|_2$, β, L and $\mathfrak{h}$.

We now establish a lower bound in Theorem 2.12.

Theorem 2.12. *For any $T > 0$, $\beta > 0$, $L > 0$ and $\mathfrak{h} > 1$, there exists a constant $C_4 > 0$ such that for any estimator $\hat{h}_T$ of h*

$$\sup_{\substack{h \in \mathcal{H}(\beta, L) \\ \|h\|_\infty \leq \mathfrak{h}}} \mathbb{E}\|\hat{h}_T - h\|_2^2 \geq C_4 \exp\left(-\frac{2\beta}{2\beta + 1} RT \right) \cdot \tag{2.15}$$

Contrary to the classical cases in nonparametric estimation (see for instance Tsybakov [37]), 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. For functions h in the Hölder class such that $\|h\|_\infty \leq \mathfrak{h}$, $\mathfrak{h} > 1$, Theorem 2.11 shows that the rate of convergence of $\hat{h}$ is $\varrho(T)^{-\frac{2\beta}{2\beta+1}} = \exp\left(-\frac{2\beta}{2\beta+1} RT\right)$, when $N_0 > 1$. In this case the upper bound matches the lower bound in Theorem 2.12. When $N_0 = 1$, the upper bound differs from the lower bound by an additional logarithmic term. 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 2.11 illustrates adaptive properties of our procedure: it achieves the rate $\varrho(T)^{-\frac{2\beta}{2\beta+1}}$ over the class $\mathcal{H}(\beta, L) \cap \{\|h\|_\infty \leq \mathfrak{h}\}$ as soon as β is smaller than β^*. So, our estimator automatically adapts to the unknown smoothness of the signal to estimate.

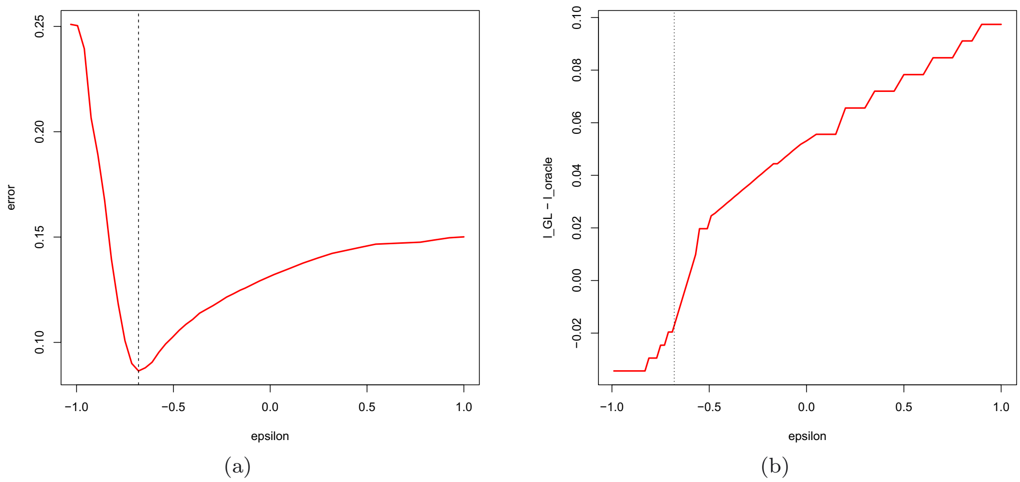


FIGURE 2. (a) MISE's as a function of ϵ. (b) $\hat{\ell} - \ell_{\text{oracle}}$ as a function of ϵ. The dotted lines indicate the optimal value of ϵ which is used in all simulations.

3. Numerical results

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 $\text{Beta}(2, 2)$. $\text{Beta}(a, b)$ distributions on $[0, 1]$ are characterized by their densities

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

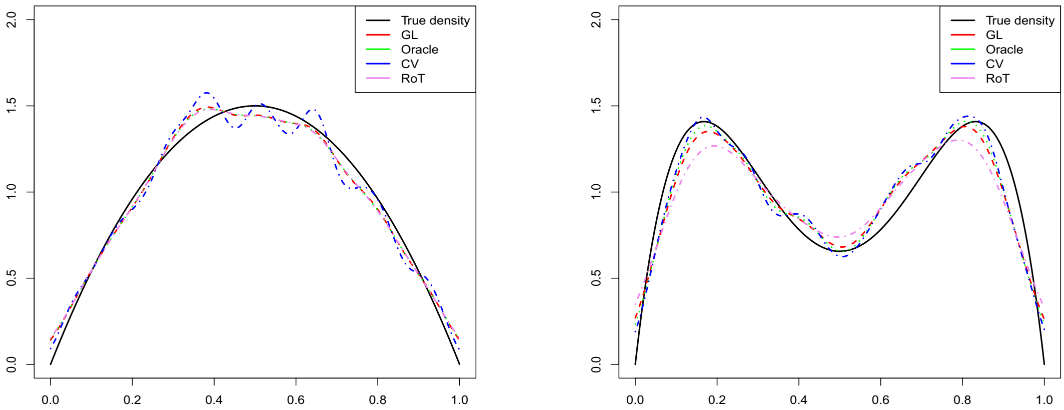
where $\mathcal{B}(a, b)$ is the renormalizing constant.

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.

We estimate $\hat{h}$ by using (2.6) and 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 $H \subset \{1, 2^{-1}, \ldots, \lfloor \delta \varrho(T) \rfloor^{-1}\}$ for some $\delta > 0$ small enough when $\varrho(T)$ 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 $\ell_{oracle} = \text{argmin}_{\,\ell \in H}\, \mathbb{E}\big[\|\hat{h}_\ell - h\|_2^2\big]$ and h is the density of $\text{Beta}(2, 2)$. In Figure 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 Thm. 2.8) 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.* [9] for more details about the GL methodology). Moreover, following the discussion in Lacour and Massart [24] we investigate (see Fig. 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.



(A) *Reconstruction of* Beta(2, 2) (B) *Reconstruction of Beta mixture*

FIGURE 3. Reconstruction of division kernels with $T = 13$.

Figure 3 illustrates a reconstruction for the density of Beta(2, 2) and Beta mixture $\frac{1}{2}\,\mathrm{Beta}(2,6) + \frac{1}{2}\,\mathrm{Beta}(6,2)$ 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 CV bandwidth and the RoT bandwidth, *i.e.* the bandwidths obtained by using the cross-validation (CV) method and the rule of thumb (RoT) respectively. For a random sample of deterministic size n, $\Gamma_1^1, \ldots, \Gamma_n^1$, the CV bandwidth is defined as follows

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

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. More details about these methods can be found in Section 3.4 of Silverman [33] or Tsybakov [37].

To estimate the MISE, we implement Monte-Carlo simulations with respect to $T = 13, 17$ and 20. The mean numbers of observations for each T are respectively 664, 4914 and 22 025. The number of repetitions for each time is $M = 100$. Then, we compute the mean of relative error $\bar{e} = (1/M)\sum_{i=1}^{M} e_i$ and the standard deviation $\sigma_e = \sqrt{(1/M)\sum_{i=1}^{M}(e_i - \bar{e})^2}$ where

$$e_i = \frac{\|\hat{h}^{(i)} - h\|_2}{\|h\|_2}, \quad i = 1, \ldots, M, \tag{3.1}$$

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

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 of a. The simulation results are displayed in Tables 1 and 2. For the density of Beta mixture, we only compute the error with $T = 13$ and $T = 17$. Boxplots in Figure 4 illustrate the MISE's in Table 1 when $T = 17$.

From Tables 1 and 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 5, we illustrate on a log-log scale the mean relative error and

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

		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

the rate of convergence versus time T. This shows that the error is close to the exponential rate predicted by the theory. Note that here we simulate the division tree with $N_0 = 1$, so the mean relative error is compared with the rate $\varrho(T)^{-\frac{2\beta}{2\beta+1}} = \left(\frac{\exp(-RT+\log(RT))}{1-\exp(-RT)}\right)^{-\frac{2\beta}{2\beta+1}}$. 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.

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 Theorem 2.11. 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 3. Compared with the error in Tables 1 and 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.

4. Proofs

In the sequel, when writing $\int f$, we refer to $\int_{\mathbb{R}} f$. Moreover, for functions h having support $(0,1)$, we write $\int h(\gamma)\mathrm{d}\gamma$ instead of $\int_0^1 h(\gamma)\mathrm{d}\gamma$.

TABLE 3. Mean of relative error for the reconstruction of $\tilde{h}$.

		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

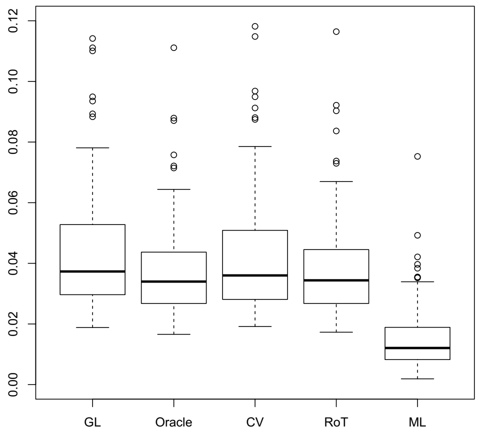


FIGURE 4. Errors of estimated densities of Beta(2, 2) when $T = 17$.

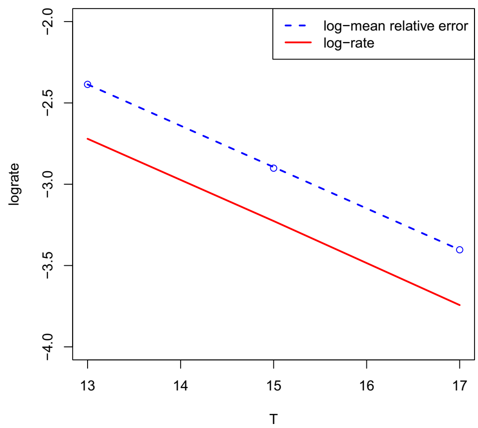


FIGURE 5. The log-mean relative error for the reconstruction of Beta(2, 2) compared to the log-rate (solid line) computed with $\beta = 1$.

4.1. Proof of Proposition 2.6

We have

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

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

$$\begin{aligned}\mathbb{E}\big[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T\big] &= \mathbb{E}\left[\int \left|\hat{h}_\ell(\gamma) - \mathbb{E}\big[\hat{h}_\ell(\gamma)\big]\right|^2 \mathrm{d}\gamma|M_T\right] \\ &= \int \mathbb{V}\mathrm{ar}\Big[\hat{\mathrm{h}}_\ell(\gamma)\big|\mathrm{M_T}\Big]\mathrm{d}\gamma \\ &= \frac{1}{M_T}\int \mathbb{V}\mathrm{ar}\Big[\mathrm{K}_\ell(\gamma - \Gamma_1^1)\Big]\mathrm{d}\gamma \\ &\le \frac{1}{M_T}\int \mathbb{E}\Big[K_\ell^2(\gamma - \mathit{\Gamma}_1^1)\Big]\mathrm{d}\gamma.\end{aligned}$$

By Fubini's theorem, we get

$$\begin{aligned}\int \mathbb{E}\Big[K_\ell^2(\gamma - \mathit{\Gamma}_1^1)\Big]\mathrm{d}\gamma &= \int\int K_\ell^2(\gamma - u)h(u)\mathrm{d}u\,\mathrm{d}\gamma \\ &= \int h(u)\left(\int K_\ell^2(\gamma-u)\mathrm{d}\gamma\right)\mathrm{d}u \\ &= \|K_\ell\|_2^2\int h(u)\mathrm{d}u = \frac{\|K\|_2^2}{\ell}\cdot\end{aligned}$$

Then we have

$$\mathbb{E}\big[\|\hat{h}_\ell - \mathbb{E}[\hat{h}_\ell]\|_2^2|M_T\big] \le \frac{\|K\|_2^2}{M_T\ell}\cdot \tag{4.1}$$

Hence, applying Cauchy−Schwarz's inequality, we obtain (2.7). This ends the proof of Proposition 2.6.

4.2. Proof of Theorem 2.8

This proof is inspired by the proof of Doumic *et al.* [14]. 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.

Recall that

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

Step 1. Let us prove that

$$\mathbb{E}\left[\|\hat{h} - h\|_2^2\right] \le 24\mathbb{E}\left[\xi_T^2(\ell)\right] + C_1\left(\|K_\ell\star h - h\|_2^2 + \frac{\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right]\right) \tag{4.2}$$

where

$$\xi_T(\ell) := \sup_{\ell'\in H}\left\{\|\big(\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\big) - \big(\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\big)\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T\ell'}}\right\}_+, \tag{4.3}$$

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

For any $\ell \in H$, we have

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

where

$$\begin{aligned}A_1 &:= \|\hat{h} - \hat{h}_{\hat{\ell},\ell}\|_2 \le A(\ell) + \frac{\chi\|K\|_2}{\sqrt{M_T\hat{\ell}}}, \\ A_2 &:= \|\hat{h}_{\hat{\ell},\ell} - \hat{h}_\ell\|_2 \le A(\hat{\ell}) + \frac{\chi\|K\|_2}{\sqrt{M_T\ell}}, \\ A_3 &:= \|\hat{h}_\ell - h\|_2.\end{aligned}$$

By definition (2.9) of $\hat{\ell}$, we have

$$A_1 + A_2 \leq 2A(\ell) + 2\frac{\chi\|K\|_2}{\sqrt{M_T \ell}}, \tag{4.4}$$

and

$$\begin{aligned} A(\ell) &\leq \sup_{\ell' \in H} \left\{ \left\| \left(\hat{h}_{\ell,\ell'} - \mathbb{E}[\hat{h}_{\ell,\ell'}]\right) - \left(\hat{h}_{\ell'} - \mathbb{E}[\hat{h}_{\ell'}]\right) \right\|_2 + \|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 - \frac{\chi\|K\|_2}{\sqrt{M_T \ell'}} \right\}_+ \\ &\leq \xi_T(\ell) + \sup_{\ell' \in H} \left\{ \|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 \right\}, \end{aligned} \tag{4.5}$$

with $\xi_T(\ell)$ defined in (4.3).

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

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

Hence, we derive

$$\|\mathbb{E}[\hat{h}_{\ell,\ell'}] - \mathbb{E}[\hat{h}_{\ell'}]\|_2 = \|K_{\ell'} \star (K_\ell \star h - h)\|_2 \leq \|K\|_1 \|K_\ell \star h - h\|_2, \tag{4.6}$$

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

Thus, (4.3)–(4.6) give

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

Then,

$$\mathbb{E}\left[(A_1 + A_2)^2\right] \leq 12\mathbb{E}[\xi_T^2(\ell)] + 12\|K\|_1^2 \|K_\ell \star h - h\|_2^2 + 12\frac{\chi^2\|K\|_2^2}{\ell}\mathbb{E}\left[\frac{1}{M_T}\right]\cdot \tag{4.7}$$

For the term A_3, we have from (4.1)

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

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

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

This proves (4.2).

Step 2. Let us deal with the term $\mathbb{E}\left[\xi_T^2(\ell)\right]$. We have

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

where

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

Hence,

$$\mathbb{E}\big[\xi_T^2(\ell)\big] \leq (1+\|K\|_1)^2\mathbb{E}\Big[\mathbb{E}\big[S_T^2|M_T\big]\Big]\cdot \tag{4.8}$$

If we show that

$$\mathbb{E}\big[S_T^2|M_T=n\big] \leq C_*\frac{1}{n}, \tag{4.9}$$

then

$$\mathbb{E}\big[\xi_T^2(\ell)\big] \leq C_*(1+\|K\|_1)^2\mathbb{E}\left[\frac{1}{M_T}\right] \tag{4.10}$$

where C_* is a constant.

Step 3. Let us establish (4.9). 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

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

Then,

$$\begin{aligned}\mathbb{E}\big[\Sigma_n^2\big] &= \mathbb{E}\left[\sup_{\ell\in H}\left\{\|Z_\ell\|_2-\frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2\right]\\ &\leq \int_0^{+\infty}\mathbb{P}\left[\sup_{\ell\in H}\left\{\|Z_\ell\|_2-\frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2\geq x\right]\mathrm{d}x\\ &\leq \sum_{\ell\in H}\int_0^{+\infty}\mathbb{P}\left[\left\{\|Z_\ell\|_2-\frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}}\right\}_+^2\geq x\right]\mathrm{d}x.\end{aligned} \tag{4.11}$$

We bound this with Talagrand's inequality.

Step 3.1. Talagrand's inequality provides deviation estimates close to the one needed in (4.11). 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

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

Let

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

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 Massart [26], p. 170) to $\|Z_\ell\|_2 = \sup\limits_{a\in\mathcal{A}} \sum_{i=1}^n V_{i,\Gamma}$. For all $\eta, x > 0$, one has

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

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

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

and,

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

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

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

For the term ν, we have

$$\begin{aligned}\nu &\leq \frac{1}{n}\sup_{a\in\mathcal{A}}\mathbb{E}\left[\left(\int a(\gamma)K_\ell(\gamma-\Gamma_1^1)\mathrm{d}\gamma\right)^2\right]\\ &\leq \frac{1}{n}\sup_{a\in\mathcal{A}}\mathbb{E}\left[\int |K_\ell(\gamma-\Gamma_1^1)|\mathrm{d}\gamma\times\int a^2(\gamma)|K_\ell(\gamma-\Gamma_1^1)|\mathrm{d}\gamma\right]\\ &\leq \frac{\|K\|_1}{n}\sup_{a\in\mathcal{A}}\mathbb{E}\left[\int a^2(\gamma)|K_\ell(\gamma-\Gamma_1^1)|\mathrm{d}\gamma\right]\\ &\leq \frac{\|K\|_1}{n}\sup_{a\in\mathcal{A}}\int a^2(\gamma)\mathbb{E}\left[|K_\ell(\gamma-\Gamma_1^1)|\right]\mathrm{d}\gamma\\ &\leq \frac{\|K\|_1}{n}\sup_{a\in\mathcal{A}}\int\int a^2(\gamma)|K_\ell(\gamma-u)|h(u)\mathrm{d}u\mathrm{d}\gamma\\ &\leq \frac{\|h\|_\infty\|K\|_1^2}{n}.\end{aligned} \tag{4.12}$$

For the term b, we have

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

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

$$\mathbb{P}\left(\|Z_\ell\|_2\geq(1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}}+\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2x}{n}}+2c(\eta)\frac{\|K\|_2x}{n\sqrt{\ell}}\right)\leq \mathrm{e}^{-x}.$$

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

$$\begin{aligned}\mathbb{P}\Bigg(\|Z_\ell\|_2\geq(1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}}&+\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2W_\ell}{n}}+2c(\eta)\frac{\|K\|_2W_\ell}{n\sqrt{\ell}}\\ &+\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2u}{n}}+2c(\eta)\frac{\|K\|_2u}{n\sqrt{\ell}}\Bigg)\leq \mathrm{e}^{-W_\ell-u}.\end{aligned}$$

If we set

$$\Psi_\ell=(1+\eta)\frac{\|K\|_2}{\sqrt{n\ell}}+\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2W_\ell}{n}}+2c(\eta)\frac{\|K\|_2W_\ell}{n\sqrt{\ell}},$$

then,

$$\mathbb{P}\Big(\|Z_\ell\|_2-\Psi_\ell\geq\|h\|_\infty^{1/2}\|K\|_1\sqrt{\frac{2u}{n}}+2c(\eta)\frac{\|K\|_2u}{n\sqrt{\ell}}\Big)\leq \mathrm{e}^{-W_\ell-u}. \tag{4.13}$$

Step 3.2. Let

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

An upper bound of Λ is given by

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

which we can make smaller than C^*/n as follows by using (4.13).

Indeed, let us take u such that

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

So,

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

Hence,

$$\begin{aligned}
\Lambda &\leq \sum_{\ell \in H} \int_0^{+\infty} \mathrm{e}^{-W_\ell - u} 2f(u) \left(\|h\|_\infty^{1/2} \|K\|_1 \frac{1}{\sqrt{2nu}} + 2c(\eta) \frac{\|K\|_2}{n\sqrt{\ell}} \right) \mathrm{d}u \\
&\leq \sum_{\ell \in H} \int_0^{+\infty} \mathrm{e}^{-W_\ell - u} 2f(u) \left(\|h\|_\infty^{1/2} \|K\|_1 \sqrt{\frac{2u}{n}} + 2c(\eta) \frac{\|K\|_2 u}{n\sqrt{\ell}} \right) u^{-1} \mathrm{d}u \\
&= 2 \sum_{\ell \in H} \mathrm{e}^{-W_\ell} \int_0^{+\infty} f(u)^2 \mathrm{e}^{-u} u^{-1} \mathrm{d}u \\
&\leq C_\eta \sum_{\ell \in H} \mathrm{e}^{-W_\ell} \left(\|h\|_\infty \|K\|_1^2 \int_0^{+\infty} \mathrm{e}^{-u} \mathrm{d}u + \frac{\|K\|_2^2}{\ell} \int_0^{+\infty} u \mathrm{e}^{-u} \mathrm{d}u \right) \times \frac{1}{n} \\
&= C_\eta \sum_{\ell \in H} \mathrm{e}^{-W_\ell} \left(\|h\|_\infty \|K\|_1^2 + \frac{\|K\|_2^2}{\ell} \right) \times \frac{1}{n} =: \frac{C^*}{n},
\end{aligned} \tag{4.14}$$

where we see that C^* depends on δ, ϵ, $\|h\|_\infty$, $\|K\|_1$, $\|K\|_2$.

Step 3.3. We need to choose W_ℓ and η such that

$$\mathbb{E}\left[\Sigma_n^2\right] = \mathbb{E}\left[\sup_{\ell \in H} \left\{ \|Z_\ell\|_2 - \frac{(1+\epsilon)\|K\|_2}{\sqrt{n\ell}} \right\}_+^2 \right] \leq \Lambda. \tag{4.15}$$

For this it is sufficient to prove that $\Psi_\ell \leq (1+\epsilon)\frac{\|K\|_2}{\sqrt{n\ell}}$. Let $\theta > 0$, we choose

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

so that 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 $\sum_{\ell \in H} \mathrm{e}^{-W_\ell}$ is finite and for any $\ell \in H$, since $\ell \leq 1$, we have

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

Since $H \subset \{\triangle^{-1}, \triangle = 1, \ldots, \triangle_{\max}\}$, if we choose $\triangle_{\max} = \lfloor \delta n \rfloor$ for some $\delta > 0$, then $\ell_{\min} = \triangle_{\max}^{-1}$. Moreover, because $\|h\|_\infty \geq 1$ since h is a probability density on $[0,1]$, we obtain

$$\Psi_\ell \leq \left(1 + \eta + \theta + \frac{c(\eta)\theta^2 \|K\|_2^2 \sqrt{\delta}}{\|K\|_1^2}\right) \frac{\|K\|_2}{\sqrt{n\ell}} \cdot \tag{4.16}$$

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

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

then

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

and we get from (4.14) and (4.15)

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

Hence, we get (4.9) and (4.10).

Step 4. Combining (4.2) and (4.10), we obtain

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

Moreover, since $N_T > N_0$, we have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{M_T}\right] &= \mathbb{E}\left[\frac{1}{N_T - N_0}\right] = \mathbb{E}\left[\frac{N_T}{N_T - N_0} \frac{1}{N_T}\right] = \mathbb{E}\left[\frac{1}{1 - \frac{N_0}{N_T}} \frac{1}{N_T}\right] \\ &\leq \mathbb{E}\left[\frac{1}{1 - \frac{N_0}{N_0+1}} \frac{1}{N_T}\right] \\ &\leq (N_0 + 1) \mathbb{E}\left[\frac{1}{N_T}\right] \cdot \end{aligned} \tag{4.17}$$

Then, using (iii) and (iv) of Proposition 2.2 and (4.17), recalling the definition of $\varrho(T)^{-1}$ in (2.12), 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 2.8.

4.3. Proof of Theorem 2.11

We begin with the bias term $\|K_\ell \star h - h\|_2$ in the right hand side of the oracle inequality (2.13). 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

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

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

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

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

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

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

Finally, we have

$$\mathbb{E}\left[\|\hat{h}-h\|_2^2\right] \leq C_1 \inf_{\ell\in H}\left\{C_{K,L,\beta}\ell^{2\beta} + \frac{\|K\|_2^2}{\ell}\varrho(T)^{-1}\right\} + C_2\varrho(T)^{-1}. \tag{4.18}$$

Taking the derivative of the expression inside the $\inf\limits_{\ell\in H}$ of (4.18) with respect to ℓ, we obtain the minimizer

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

Note that 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 (4.18), 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, $\|K\|_1$, $\|K\|_2$, β, L and $\mathfrak{h}$. Here, we notice that the dependence of C_3 on $\mathfrak{h}$ is deduced from the fact that the constant C_2 depends on $\|h\|_\infty$ and by using the assumption $\|h\|_\infty \leq \mathfrak{h}$. This ends the proof of Theorem 2.11.

4.4. Proof of Theorem 2.12

For $T > 0$, let us denote by $\hat{h}_T$ the estimator of h. To prove Theorem 2.12, we apply the general reduction scheme proposed by Tsybakov [37] (Sect. 2.2, p. 79). We will show the existence of a family $\mathcal{H}_{m,T} = \{h_{i,T} : i = 0, 1, \ldots, m\}$ such that:

(1) $h_{i,T} \in \mathcal{H}(\beta, L)$ and $\|h_{i,T}\|_\infty \leq \mathfrak{h}$, $\mathfrak{h} > 1$.

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

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

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

Under the preceding conditions 1, 2, 3, Tsybakov [37] (Thm. 2.5, p. 99) show that

$$\inf_{\hat{h}_T} \max_{h\in\mathcal{H}_{m,T}} \mathbb{P}\left(\|\hat{h}_T - h\|_2^2 \geq c^2 \mathrm{e}^{-\frac{2\beta}{2\beta+1}RT}\right) \geq C', \tag{4.19}$$

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 2.12 by [[37], Thm. 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 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 $\int g(\gamma)\mathrm{d}\gamma = 0$, $g \in \mathcal{H}(\frac{1}{2},\beta)$, we define

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

where $\lceil x \rceil$ denotes a smallest integer which is strictly larger than the real number x.

By definition, the functions f_k's have disjoint support $\big((k-1)/D, k/D\big]$ and one can check that the functions $f_k \in \mathcal{H}(\frac{L}{2},\beta)$.

Then, the function $h_{i,T}$ 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

$$c_1 = \min\left(1, \frac{1}{LD^{-\beta}\|g\|_\infty}, \frac{\mathfrak{h}-1}{LD^{-\beta}\|g\|_\infty}\right). \tag{4.20}$$

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

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

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

(1) The condition $h_{i,T} \in \mathcal{H}(\beta, L)$ and $\|h_{i,T}\|_\infty \leq \mathfrak{h}$:

Let us denote $q = \lfloor \beta \rfloor$, using the function f_k's have disjoint support, then for all $\gamma, \gamma' \in (0,1)$ we have

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

which is always satisfied with $c_1 = \min\left(1, \frac{1}{LD^{-\beta}\|g\|_\infty}, \frac{\mathfrak{h}-1}{LD^{-\beta}\|g\|_\infty}\right)$.

Moreover, we also have by the choice of c_1

$$
\begin{aligned}
\|h\|_\infty = \Big\| h_0 + c_1 \sum_{k=1}^{D} \delta_k f_k \Big\|_\infty &\leq \|h_0\|_\infty + \Big\| c_1 \sum_{k=1}^{D} \delta_k f_k \Big\|_\infty \\
&\leq 1 + c_1 L D^{-\beta} \max_k \sup_\gamma |\delta_k| g\big(D\gamma - (k-1)\big) \\
&\leq 1 + c_1 L D^{-\beta} \|g\|_\infty \\
&\leq \mathfrak{h}.
\end{aligned}
$$

Thus $h_\delta \in \mathcal{H}(\beta, L) \cap \{\|h\|_\infty \leq \mathfrak{h}\}$, $\mathfrak{h} > 1$.

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

For all $\delta, \delta' \in \{0,1\}^D$, recalling the function f_k's have disjoint support, we have

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

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 [37], p. 104), if $D \geq 8$ then there exists a subset $\{\delta^{(0)}, \ldots, \delta^{(m)}\}$ of $\{0,1\}^D$ with cardinal (4.21) such that $\delta^{(0)} = (0, \ldots, 0)$,

$$m \geq 2^{D/8}, \tag{4.21}$$

and

$$d_H(\delta^{(i)}, \delta^{(k)}) \geq \frac{D}{8},\ \forall\, 0 \leq i < k \leq m. \tag{4.22}$$

Recall that $D = \lceil c_0 \mathrm{e}^{\frac{RT}{2\beta+1}} \rceil$. If we choose $T^* = \frac{2\beta+1}{R} \log\left(\frac{7}{c_0}\right)$ then we have $D \geq 8$ for all $T \geq T^*$. Furthermore, we also have by that choice of T^*

$$D^\beta \leq \left(1 + \frac{1}{7}\right)^\beta c_0^\beta \mathrm{e}^{\frac{\beta}{2\beta+1}RT} \leq (2c_0)^\beta \mathrm{e}^{\frac{\beta}{2\beta+1}RT}, \quad T \geq T^*.$$

Then, by setting $h_{i,T}(x) = h_{\delta^{(i)}}(x)$, $i = 0, \ldots, m$, we obtain by the Lemma of Varshamov−Gilbert

$$\begin{aligned}\|h_{i,T} - h_{k,T}\|_2 &= c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{d_H(\delta^{(i)}, \delta^{(k)})} \\ &\geq c_1 L D^{-\beta-1/2} \|g\|_2 \sqrt{\frac{D}{8}} \\ &\geq \frac{c_1 L}{\sqrt{8}} \|g\|_2 D^{-\beta} \\ &\geq \frac{c_1 L}{\sqrt{8}} \|g\|_2 (2c_0)^{-\beta} \mathrm{e}^{-\frac{\beta}{2\beta+1}RT}.\end{aligned}$$

But,

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

Hence, we obtain

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

where

$$c = \min\left(L\|g\|_2, \frac{\|g\|_2}{\|g\|_\infty}, \frac{(\mathfrak{h}-1)\|g\|_2}{\|g\|_\infty}\right) \frac{(2c_0)^{-\beta}}{2\sqrt{8}}.$$

(3) The condition $\frac{1}{m}\sum_{i=1}^m K(P_i, P_0) \leq \vartheta \log(m)$ **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

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

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

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 $Rh_\delta(\gamma)\mathrm{d}s\, n(\mathrm{d}j)\mathrm{d}\gamma$. Thus a natural tool is to use Girsanov's theorem (see Jacob and Shiryaev [22], Thm. 3.24, p. 159) saying that P_δ is absolutely continuous with respect to P_0 on $\mathcal{F}_T$ with

$$\frac{\mathrm{d}P_\delta}{\mathrm{d}P_0}|_{\mathcal{F}_T} = \mathfrak{D}_T^\delta,$$

where $(\mathfrak{D}_t^\delta)_{t\in[0,T]}$ is the unique solution of the following SDE (see Prop. 4.17 of Tran [35] for a similar SDE):

$$\mathfrak{D}_T^\delta = 1 + \int_0^T \int_{\mathcal{E}} \mathfrak{D}_{s-}^\delta \mathbb{1}_{\{j\leq N_{s-}\}} \left(\frac{h_\delta(\gamma)}{h_0(\gamma)} - 1\right) Q(\mathrm{d}s, \mathrm{d}j, \mathrm{d}\gamma). \tag{4.23}$$

Apply Itô formula for jump processes to (4.23), we get

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

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

Then,

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

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

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

From (4.21), we have $m \geq 2^{D/8}$ then

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

Hence, if we set for some $\vartheta \in (0, 1/8)$

$$c_0 := \left(\frac{8N_0L^2\|g\|_2^2}{\vartheta\log(2)}\right)^{1/(2\beta+1)},$$

we obtain $K(P_\delta, P_0) \leq \vartheta\log(m)$. This ends the proof of Theorem 2.12.

Appendix

The proofs of technical results are presented below.

Proof of Proposition 2.2.

(ii) The proof of (ii) can be found easily in literature. Here we refer to [31], Section 5.3 for this proof.

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

(iii) Let $p = \mathrm{e}^{-RT}$. When $N_0 = 1$, $N_T \sim \mathrm{Geom}(p)$. 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 $\mathbb{E}\Big[\frac{1}{N_T}\Big] = \frac{RT\mathrm{e}^{-RT}}{1-\mathrm{e}^{-RT}}$.
When $N_0 > 1$, $N_T \sim \mathcal{NB}(N_0, p)$. Hence, we have

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

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

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

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

Hence,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}f(1-p) =& -\frac{1}{p}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{1}{p^k}(-1)^{N_0-1-k}+(-1)^{N_0-1}\right]\\ =&(-1)^{N_0}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{p^{k+1}}+\frac{1}{p}\right].\end{aligned} \tag{A.2}$$

Integrating equation (A.2) and notice that $f(0)=0$, we get

$$\begin{aligned}f(1-p) =&(-1)^{N_0}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{k}\Big(-\frac{1}{p^k}\Big)+\log(p)\right]\\ =&(-1)^{N_0-1}\left[\sum_{k=1}^{N_0-1}\binom{N_0-1}{k}\frac{(-1)^k}{k}\frac{1}{p^k}+\log\left(\frac{1}{p}\right)\right].\end{aligned} \tag{A.3}$$

Combine (A.1),(A.3) and replace p with e^{-RT}, this proves (iii).

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

$$N_T = N_0 + \int_0^T\int_{\mathcal{E}} \mathbb{1}_{\{j\le N_{s-}\}}Q(\mathrm{d}s,\mathrm{d}j,\mathrm{d}\gamma). \tag{A.4}$$

Applying Itô formula for jump processes (see [21], Thm. 5.1 on p. 67) to (A.4), we obtain

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

Hence,

$$\mathbb{E}\left[\frac{1}{N_T}\right] = \frac{1}{N_0} - \mathbb{E}\left[\int_0^T \frac{1}{N_s\left(N_s+1\right)}RN_s\mathrm{d}s\right] = \frac{1}{N_0} - R\int_0^T\mathbb{E}\left[\frac{1}{N_s+1}\right]\mathrm{d}s. \tag{A.5}$$

Since $N_s \ge N_0$, we have $\frac{1}{N_s+1}\le\frac{1}{N_s}$. Therefore, (A.5) implies that

$$\mathbb{E}\left[\frac{1}{N_T}\right] \ge \frac{1}{N_0} - R\int_0^T\mathbb{E}\left[\frac{1}{N_s}\right]\mathrm{d}s. \tag{A.6}$$

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

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

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

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

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 $N_0 > 1$

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

This ends the proof of Proposition 2.2. □

Proof of Proposition 2.5.

To prove i), let us remark that the number of random divisions M_T is independent of Γ_i^1, $i = 1, \ldots, M_T$, 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 $\mathbb{E}\big[\hat{h}_\ell\big] = \mathbb{E}\left[\mathbb{E}\big[\hat{h}_\ell|M_T\big]\right] = K_\ell \star h(\gamma)$. By similar calculations as to obtain (i), we obtain (ii) and (iii).

To prove (iv), 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 i) of Proposition 2.2, 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 2.5. □

Acknowledgements. The author is deeply grateful to Tran, Rivoirard and Pham Ngoc for the guidance and useful suggestions. The author would like to express his sincere thanks to P. Massart for constructive comments that improved the results. The author wish to thank the anonymous referee for a thorough reading and many helpful suggestions, which led to improvements of the paper. The author would like to thank Centre de Resources Informatiques de Lille 1 (CRI) for the computational cluster to implement the numerical simulations. This work is supported by Program 911 of Vietnam Ministry of Education and Training and is partly supported by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

References

[1] H. Aguilaniu, L. Gustafsson, M. Rigoulet and T. Nyström, Asymmetric Inheritance of Oxidatively Damaged Proteins During Cytokinesis. *Sci.* **299** (2003) 1751.

[2] M. Ackermann, S.C. Stearns and U. Jenal, Senescence in a Bacterium with Asymmetric Division. *Sci.* **300** (2003) 1920.

[3] K.B. Athreya and P.E. Ney, Branching Processes. Springer edition (1970).

[4] V. Bansaye, Proliferating parasites in dividing cells: Kimmels branching model revisited. *Ann. Appl. Probab.* (2008) 967–996.

[5] V. Bansaye, J.-F. Delmas, L. Marsalle and V.C. Tran, Limit theorems for Markov processes indexed by continuous time Galton - Watson trees. *Ann. Appl. Probab.* **21** (2011).

[6] V. Bansaye and V.C. Tran, Branching Feller diffusion for cell division with parasite infection. *ALEA, Lat. Am. J. Probab. Math. Stat.* **8** (2011) 95–127.

[7] V. Bansaye, J.C. Pardo and C. Smadi, On the extinction of continuous state branching processes with catastrophes. *Electron. J. Probab.* **18** (2013) 1–31.

[8] B. Bercu, B. De Saporta and A. Gégout-Petit, Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* **14** (2009) 2492–2526.

[9] K. Bertin, C. Lacour and V. Rivoirard, Adaptive pointwise estimation of conditional density function. *Ann. Inst. Henri Poincaré Probab. Statist.* (2015).

[10] S. Valère Bitseki Penda, Deviation inequalities for bifurcating Markov chains on Galton – Watson tree. *ESAIM: PS* **19** (2015) 689–724.

[11] B. Cloez, Limit theorems for some branching measure-valued processes. *Adv. Appl. Probab.* **49** (2017) 549–580.

[12] J.-F. Delmas and L. Marsalle, Detection of cellular aging in a Galton-Watson process. *Stochastic Processes and their Application* **120** (2010) 2495–2519.

[13] M. Doumic, M. Hoffmann, N. Krell and L. Robert, Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli* **21** 1760–1799.

[14] M. Doumic, M. Hoffmann, P. Reynaud−Bouret and V. Rivoirard, Nonparametric estimation of the division rate of a size-structured population. *SIAM J. Numer. Anal.* **50** (2012).

[15] S.N. Evans and D. Steinsaltz, Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoret. Popul. Biol.* **71** (2007) 473–490.

[16] A. Goldenshluger and O. Lepski, Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** (2011) 1608–1632.

[17] J. Guyon, Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17** (2007) 1538–1569.

[18] T.E. Harris, The Theory of Branching Processes. Springer, Berlin (1963).

[19] V.H. Hoang, *Adaptive estimation for inverse problems with applications to cell divisions*. PhD dissertation, Université de Lille 1 Sciences et Technologies, https://tel.archives-ouvertes.fr/tel-01417780 (2016).

[20] M. Hoffmann and A. Olivier, Nonparametric estimation of the division rate of an age dependent branching process. *Stoch. Process. Appl.* **126** (2016) 1433–1471.

[21] N. Ikeda and S. Wanatabe, Stochastic differential equations and diffusion processes. Vol. 24 of North-Holland Mathematical Library.

[22] J. Jacob and A.N. Shiryaev, Limit Theorems for Stochastic Processes. Springer Verlag, Berlin (1987).

[23] C.-Y. Lai, E. Jaruga, C. Borghouts and S.M. Jazwinski1, A Mutation in the ATP2 Gene Abrogates the Age Asymmetry Between Mother and Daughter Cells of the Yeast Saccharomyces cerevisiae. *Genetics* **162** (2002) 73–87.

[24] C. Lacour and P. Massart. Minimal penalty for Goldenshluger-Lepski method. *Stoch. Process. Appl.* **126** (2016) 3774–3789.

[25] 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. *PNAS* **105** (2015) 3076–3081.

[26] 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).

[27] S.P. Meyn and R.L. Tweedie, Stability of Markovian processes iii: Foster - Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.* **25** (1993).

[28] J.B. Moseley, Cellular Aging: Symmetry Evades Senescence. *Current Biology*, Vol. 23, Issue 19, R871 − R873 (2013).

[29] J. Peter, A general stochastic model for population development. *Scandinavian Actuarial J.* **1969** (1969) 84–103.

[30] 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** (2014).

[31] Sh.M. Ross, *Stochastic Processes*. Wiley, 2nd edition (1995).

[32] P. Reynaud−Bouret, V. Rivoirard, F. Grammont and C. Tuleau−Malot, Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neuroscience* **4** (2014).

[33] B.W. Silverman, *Density estimation for statistics and data analysis*, volume 26. CRC press (1986).

[34] 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).

[35] V.C. Tran, *Modèles particulaires stochastiques pour des problèmes d'évolution adaptive et pour l'approximation de solutions statisques*. Ph.D. dissertation, Université Paris X - Nanterre, http://tel.archives-ouvertes.fr/tel-00125100 (2006).

[36] V.C. Tran, Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probab. Statis.* **12** (2008).

[37] A.B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer series in Statistics. Springer (2004).