



**HAL**  
open science

# Nonparametric adaptive inference of birth and death models in a large population limit

Alexandre Boumezoued, Marc Hoffmann, Paulien Jeunesse

► **To cite this version:**

Alexandre Boumezoued, Marc Hoffmann, Paulien Jeunesse. Nonparametric adaptive inference of birth and death models in a large population limit. *Mathematical Statistics and Learning*, 2021, 3 (1), pp.1-69. 10.4171/MSL/18 . hal-03064772

**HAL Id: hal-03064772**

**<https://hal.science/hal-03064772>**

Submitted on 12 Jul 2024

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

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

# NONPARAMETRIC ADAPTIVE INFERENCE OF BIRTH AND DEATH MODELS IN A LARGE POPULATION LIMIT

ALEXANDRE BOUMEZOUED, MARC HOFFMANN AND PAULIEN JEUNESSE

ABSTRACT. Motivated by improving mortality tables from human demography databases, we investigate statistical inference of a stochastic age-evolving density of a population alimeted by time inhomogeneous mortality and fertility. Asymptotics are taken as the size of the population grows within a limited time horizon: the observation gets closer to the solution of the Von Foerster Mc Kendrick equation, and the difficulty lies in controlling simultaneously the stochastic approximation to the limiting PDE in a suitable sense together with an appropriate parametrisation of the anisotropic solution. In this setting, we prove new concentration inequalities that enable us to implement the Goldenshluger-Lepski algorithm and derive oracle inequalities. We obtain minimax optimality and adaptation over a wide range of anisotropic Hölder smoothness classes.

**Mathematics Subject Classification (2010):** 62G05, 62M05, 60J80, 60J20, 92D25.

**Keywords:** Age-structured models, large population limit, concentration inequalities, nonparametric adaptive estimation, anisotropic estimation, Goldenshluger-Lepski method.

## CONTENTS

1. Introduction	2
1.1. Setting	2
1.2. Motivation	3
1.3. Link with literature on death rate inference	3
1.4. Results and organisation of the paper	4
2. The microscopic model and its large population limit	6
2.1. Notation	6
2.2. Construction of the model	6
2.3. Stability of the model	7
3. Nonparametric estimation of $g$ and $\mu$	10
3.1. Kernel approximation	10
3.2. Construction of estimators of $g$ and $\mu$	10
3.3. Oracle inequalities	11
4. Adaptive estimation under anisotropic Hölder smoothness	14
4.1. The smoothness of the McKendrick Von Foerster equation	14
4.2. Minimax lower bounds	15
4.3. Adaptive estimation under anisotropic Hölder smoothness	16
5. Numerical illustration	17
6. Proof of Theorem 6	17
6.1. A first stability result	18
6.2. Stability of the stochastic term	22
6.3. Proof of Theorem 6	28

6.4. Remaining proofs of Section 2	31
7. Proofs of Section 3 and 4	32
7.1. Proof of Theorem 11	32
7.2. Proof of Theorem 13	34
7.3. Proof of Theorem 18	39
7.4. Proof of Theorem 19	43
7.5. Proof of Theorem 20	44
8. Appendix	45
8.1. Proof of Proposition 25	45
8.2. Proof of Proposition 7	46
8.3. Proof of Proposition 17	47
8.4. Further estimates on the McKendricks Von Foerster equation	49
References	49

## 1. INTRODUCTION

1.1. **Setting.** Suppose one wishes to recover a probability density  $g$  over the nonnegative real line  $\mathbb{R}_+ = [0, \infty)$  from a  $N$ -sample  $a_1, \dots, a_N$ , where the  $a_i$  are not necessarily independent. If  $Z^N = N^{-1} \sum_{i=1}^N \delta_{a_i}$  denotes the empirical distribution of the  $N$ -sample, designing a good statistical estimator of  $g$  requires a fine quantitative control of the fluctuations in the convergence

$$(1) \quad \int_{\mathbb{R}_+} \psi(a) Z^N(da) \rightarrow \int_0^\infty \psi(a) g(a) da$$

(at least in probability) as  $N$  grows, for a large enough class of test functions  $\psi$ . Moreover, the performance of such a procedure depends on the smoothness properties of the function  $g$ , typically quantified by a smoothness parameter, like a (possibly fractional) number of derivatives in any reasonable sense and is usually unknown by the practitioner. For suitable  $\psi$  (possibly data-dependent), optimal estimators can be found provided good concentration inequalities are available for (1), following the broad guiding principle of Lepski's method [30, 17, 18] or other adaptive methods like model selection or wavelets, see for instance the comprehensive textbooks of Giné and Nickl [16] or Härdle *et al.* [19] or Tsybakov [44]. In this paper, we generalise the classical situation described above by adding a time variable. We investigate statistical inference of a time-evolving particle system governed by stochastic dynamics: for every  $t \in [0, T]$ , we observe the state of a population of (approximately)  $N$  particles, encoded by its empirical measure  $Z^N = (Z_t^N(da))_{0 \leq t \leq T}$ . Informally,  $Z^N$  is solution to a certain stochastic differential equation (SDE)

$$\mathcal{H}_{b,\mu}^N(Z^N) = 0,$$

constructed in (7) below;  $\mathcal{H}_{b,\mu}^N$  is parametrised by two functions  $b$  and  $\mu$  and  $Z_t^N(da)$  represents the state of a population structured in age  $a \in \mathbb{R}_+$ , alimeted by a time-inhomogeneous fertility rate  $b(t, a)$  and decimated by a mortality rate  $\mu(t, a)$ . Moreover, we are given an initial empirical age distribution  $Z_0^N$  at time  $t = 0$ . Under appropriate regularity conditions we have a convergence  $\mathcal{H}_{b,\mu}^N \rightarrow \mathcal{H}_{b,\mu}$  in a large population limit  $N \rightarrow \infty$ , where  $\mathcal{H}_{b,\mu}(g) = 0$  is an inhomogeneous version

of the classical McKendrick Von Foerster renewal equation [35, 46], given by

$$(2) \quad \begin{cases} \partial_t g(t, a) + \partial_a g(t, a) + \mu(t, a)g(t, a) = 0 \\ g(0, a) = g_0(a), \quad g(t, 0) = \int_0^\infty b(t, a)g(t, a)da. \end{cases}$$

and that reveals the interplay between the limiting solution  $g$  and the model parameter  $b$  and  $\mu$ . In particular, we have a convergence

$$(3) \quad \int_0^T \int_{\mathbb{R}_+} \psi(t, a) Z_t^N(da) dt \rightarrow \int_0^T \int_0^\infty \psi(t, a) g(t, a) da dt$$

(at least in probability) as  $N$  grows for a rich enough class of functions  $\psi$ , and this situation generalises (1) in a time-dependent framework.

Informally, our statistical problem takes the following form: estimate  $g$  or the parameters of the model  $b, \mu$  from data  $Z^N$  in the limit  $N \rightarrow \infty$ . In this setting, it is crucial to understand: (i) the quantitative properties of the convergence (3) and in particular, how concentration inequalities can be obtained (with a view towards an adaptive estimation scheme in the idea of Lepski's principle) and (ii) what is the structure of the equation  $\mathcal{H}_{b, \mu}(g) = 0$  in terms of identification and interplay between the parameters  $b, \mu, g$  and their smoothness properties. In particular, the anisotropic smoothness of  $g$  viewed as a graph-manifold can benefit from the structure  $\mathcal{H}_{b, \mu}(g) = 0$  and lead to better approximation properties in certain directions along the characteristics of the transport.

**1.2. Motivation.** Of primary interest for us is human demography through the recent efforts and contributions for improving mortality estimates, see [8, 4, 5] among others and the references therein. In particular, the recent development of large human datasets like the Human Mortality Database (HMD) and Human Fertility Database (HFD) [21, 20] – in open access – allows one to process fertility and mortality data simultaneously, and subsequently addresses demographical issues such as the anomalies of cohort effects that have long fascinated demographers and actuaries [41, 7]. In this rejuvenated context, it becomes reasonable to study the estimation of population density or mortality rate in the enriched dynamical framework provided by birth-death particle systems that converge to the classical McKendrick Von Foerster equation in a large population limit, and revisit classical studies like *e.g.* [24, 38] for statistical estimation of the death rate; see the detailed literature review in next section. In this setting, we consider the idealised model where we can observe the (renormalised) evolution of the state of the population  $Z_t^N$  continuously for  $t \in [0, T]$ , where  $t = 0$  is the starting date for the observation of the population and  $t = T$  a terminal time horizon, fixed once for all. We are interested in identifying or estimating the parameters of the model. Of major importance is the inhomogeneous death rate  $\mu(t, a)$ . In our framework, we cannot recover the birth rate since we are not given any genealogical input: mathematically, this simply expresses the lack of injectivity of the mapping  $b \mapsto g$ . Still, our observation enables us to identify the functions  $(t, a) \mapsto g(t, a)$  and  $(t, a) \mapsto \mu(t, a)$  in the limit  $N \rightarrow \infty$  and establish a thorough nonparametric estimation program, in the methodology of adaptive minimax estimation.

**1.3. Link with literature on death rate inference.** The main difficulty in establishing a consistent theory to estimate mortality rates comes from two key points: (i) incorporate the fact that the death rate depends on both age and time (non-homogeneous setting) and (ii) use as observables the outcome of a stochastic population dynamics (birth-death process). In the literature, we argue that each point is treated separately. The inference of a time-dependent death rate also related to a time-dependent covariate (possibly age), which relates to the first

point has been addressed from a nonparametric perspective by e.g. [1, 12, 24, 34, 39, 6, 11] and the references therein. From [24], *"One way of understanding the difficulties in establishing an Aalen theory in the Lexis diagram is that although the diagram is two-dimensional, all movements are in the same direction (slope 1) and in the fully non-parametric model the diagram disintegrates into a continuum of life lines of slope 1 with freely varying intensities across lines. The cumulation trick from Aalen's estimator (generalizing ordinary empirical distribution functions and Kaplan & Meier's (1958) nonparametric empirical distribution function from censored data) does not help us here."* On the other side, the inference of an age-dependent death rate in an homogeneous birth-death model (or similar) - oiuyr second point - has been addressed in [9, 13, 22] among others. To the best of our knowledge, no statistical method deals with the usual problem faced by demographers related to the inference of a time and age-dependent death rate table based on the observation of population dynamics. Note that in this paper, the observation of the population is assumed to be continuous over time, whereas in practice the information on population exposure is extracted from census (point observation); these practical considerations are discussed in a companion paper, see [5].

**1.4. Results and organisation of the paper.** In a first part of the paper, Section 2, we construct the SDE that describes the state of the population  $Z^N$  by means of a birth-death process characterised via a stochastic differential equation – given in (7) – driven by a random Poisson measure. We recall its convergence in a large population limit to the solution of the McKendrick Von Foerster equation  $g$  based on classical results of [43, 36]. Our next step consists in quantifying the stability of the convergence  $Z^N \rightarrow g$ . To that end and anticipating the subsequent statistical analysis, we introduce two pseudo-distances:

$$\mathcal{W}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathbb{R}_+} w_2(t-a) f(t,a) (Z_t^N(da) - g(t,a) da)$$

and its integrated version

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathbb{R}_+} w_2(s-a) f(s,a) (Z_s^N(da) - g(s,a) da) ds,$$

where  $w_1$  and  $w_2$  are two bounded weight functions and  $\mathcal{F}$  a rich enough class of function with complexity measured in terms of entropy conditions. Note that formally  $\mathcal{W}_{w_2}^N(\mathcal{F})_t$  is a degenerate version of  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t$ . Taking  $w_1 = w_2 = 1$  is reminiscent of the 1-Wasserstein distance if  $\mathcal{F}$  consists of 1-Lipschitz functions. However, for the statistical analysis, we must be able to handle approximating kernels that do not have bounded Lipschitz norms, hence the presence of the weights  $w_1$  and  $w_2$  that can accomodate such kernels. The main result of this section, Theorem 6 states that under appropriate regularity conditions on  $b$  and  $\mu$ , if  $|w_2|_{1,\infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_0$  is of (small) order  $r_N$ , so are

$$|w_2|_{1,\infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_T \quad \text{and} \quad (|w_1|_{1,\infty} |w_2|_{1,\infty})^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T.$$

The rate of decay  $r_N$  possibly inflates by an order  $N^{-1/2}$  and the result holds in terms of exponential decay of the fluctuation probabilities. The functional control  $|\cdot|_{1,\infty} = (\|\cdot\|_{L^1} \cdot \|\cdot\|_{L^\infty})^{1/2}$  interpolates between  $L^1$  and  $L^\infty$ -norms, and is sufficient to handle the behaviour of statistical kernels in an optimal way, since it can therefore be compared to the usual  $L^2$ -norm that appears in variance terms. The concentration of  $\mathcal{W}_{w_2}^N(\mathcal{F})_T$  expresses a kind of stability of the particle system from  $t = 0$  to  $t = T$ , while the more intricate control of  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$  is crucial to control variance terms in bi-variate kernel estimators for the nonparametric estimation of  $g(t,a)$  and  $\mu(t,a)$ . The proof relies on a combination of martingales techniques in the spirit of Tran [43], a central reference for the paper, combined with classical tools from concentration of processes indexed by functions

under entropy controls, following for instance Ledoux-Talagrand [28].

In a second part, Section 3, we construct nonparametric estimators of  $g(t, a)$  and  $\mu(t, a)$  by means of kernel approximation: we consider estimators of the form

$$\widehat{g}_h^N(t, a) = K_h \star Z_t^N(a)$$

for  $g(t, a)$ , where  $\star$  denotes convolution and  $K_h = h^{-1}K(h^{-1}\cdot)$ , with  $|K|_1 = 1$ , is a kernel normalised in  $L^1$  with bandwidth  $h > 0$ . It is noteworthy that for estimating the population density  $g(t, a)$  at time  $t$ , the information  $Z_t^N$  is sufficient and we do not need the data  $(Z_s^N, s \neq t)$ . The situation is very different for estimating  $\mu(t, a)$  the main parameter of interest. We construct a quotient estimator, inspired from a Nadaraya-Watson type procedure, and use

$$(4) \quad \widehat{\mu}_{h_1, h_2, h_3}^N(t, a) = \frac{(H_{h_1} \otimes K_{h_2} \circ \varphi) \star \Gamma^N(du, ds)}{\widehat{g}_{h_3}^N(t, a)}$$

where  $\Gamma^N(du, ds)$  is the point process of the death occurrences in the population lifetime that can be extracted from  $Z^N$  and that converges to  $\pi = \mu g$ , see (17) in Section 3.2 for the details. In (4), we consider a bivariate kernel  $H \otimes K$  with bandwidth  $(h_1, h_2)$  and  $\varphi(t, a) = (t, t - a)$  is a certain change of coordinates that enables one to benefit from the smoothness along the characteristics of the transport. The choice of the bandwidths  $h_1, h_2, h_3$  is chosen according to the data  $Z^N$  itself, in the spirit of Lepski's principle [17, 18]. In Theorems 11 and 13, we derive oracle inequalities that control the pointwise risk of  $\widehat{g}_h^N(t, a)$  and  $\widehat{\mu}_{h_1, h_2, h_3}^N(t, a)$  in terms of optimal balance between the error propagation of Theorem 6 and the linear approximation kernels.

Section 4 is devoted the adaptive estimation of  $g$  and  $\mu$  for the pointwise risk under smoothness constraints. In a first part, we study the smoothness of  $g$  when  $b$  and  $\mu$  belong to anisotropic Hölder spaces (and for simplicity, we assume that the initial condition  $g_0$  is sufficiently smooth). Thanks to the relatively explicit form of the solution of the McKendrick Von Foester equation, we establish in Proposition 16 that when parametrised via  $\varphi$ , the function  $\tilde{g}$  in the representation  $g = \tilde{g} \circ \varphi$  has explicitly quantifiable improved smoothness over  $g$ , suggesting to consider the approximation kernel  $H_{h_1} \otimes K_{h_2} \circ \varphi$  for estimating  $\pi$  via the quotient estimator (4) that implicitly uses the representation of  $\mu = \pi/g$ . We establish in Theorem 18 minimax lower bounds for estimating  $g(t, a)$  and  $\mu(t, a)$  and prove in Theorems 19 and 20 that these bounds are optimal in some cases, thanks to the oracle inequalities established Theorems 11 and 13. In particular, we achieve minimax adaptation over anisotropic Hölder smoothness constraints, up to poly-logarithmic terms.

The techniques developed in this paper have at least two possible lines of extensions for considering more general models than (2): (i) first, when we replace the constant transport by an arbitrary aging function solution to  $dX_t = v(X_t)dt$  if  $X_t$  denotes the age evolution of an individual, and (ii) if we allow for interacting particle system in the following sense: we replace  $\mu(t, a)$  by a population dependent mortality rate  $\tilde{\mu}(t, a) + \int_{\mathbb{R}_+} U(a, a') Z_s^N(da')$ , as already studied for instance by Tran [9] for some baseline mortality rate  $\tilde{\mu}(t, a)$  affected in a mean-field sense by a kernel  $U(a, a')$ . Under appropriate regularity assumptions, the limiting model takes the form

$$\begin{cases} \partial_t g(t, a) + \partial_a(v(a)g(t, a)) + (\tilde{\mu}(t, a) + \int_{\mathbb{R}_+} U(a, a')g(t, a')da')g(t, a) = 0 \\ g(0, a) = g_0(a), \quad g(t, 0) = \int_0^\infty b(t, a)g(t, a)da. \end{cases}$$

We intend to describe the extension to this situation in a forthcoming work. Sections 6 is devoted to the proof of the main concentration result of Theorem 6 and auxiliary stability results of Section 2. In Section 7, we give the proofs of the statistical results of 3 and 4. The Appendix Section 8 contains some useful technical and auxiliary results.

## 2. THE MICROSCOPIC MODEL AND ITS LARGE POPULATION LIMIT

### 2.1. Notation.

*The function spaces.* We fix once for all a terminal time  $T > 0$  and  $\mathcal{D} = [0, T] \times \mathbb{R}_+$ . We work with the set of (measurable) functions

$$\mathcal{L}_{\mathcal{D}}^{\infty} = \{f : \mathcal{D} \rightarrow \mathbb{R}, \sup_{t,a} |f(t, a)| < \infty\},$$

implicitly continued on  $\mathbb{R} \times \mathbb{R}$  by setting  $f(t, a) = 0$  for  $(t, a) \notin \mathcal{D}$  and also introduce

$$\mathcal{L}_{\mathcal{D}}^{\text{time}} = \{f : [0, T] \rightarrow \mathbb{R}, \sup_t |f(t)| < \infty\}, \quad \mathcal{L}_{\mathcal{D}}^{\text{age}} = \{f : \mathbb{R}_+ \rightarrow \mathbb{R}, \sup_a |f(a)| < \infty\},$$

with natural embeddings  $\mathcal{L}_{\mathcal{D}}^{\text{time}} \subset \mathcal{L}_{\mathcal{D}}^{\infty}$  and also  $\mathcal{L}_{\mathcal{D}}^{\text{age}} \subset \mathcal{L}_{\mathcal{D}}^{\infty}$  for appropriate arguments. For  $p = 1, 2$ , we set

$$(5) \quad |f|_p = \left( \int_{\mathcal{D}} |f(t, a)|^p dt da \right)^{1/p}, \quad |f|_{\infty} = \sup_{(t,a) \in \mathcal{D}} |f(t, a)|, \quad |f|_{1,\infty} = (|f|_1 |f|_{\infty})^{1/2}.$$

We obviously have  $|f|_2 \leq |f|_{1,\infty}$ , but also the following interesting stability property under dilation: for every  $\tau > 0$ ,

$$|f|_2 |\tau^{1/2} f(\tau \cdot)|_{1,\infty} = |\tau^{1/2} f(\tau \cdot)|_2 |f|_{1,\infty}.$$

For  $0 \leq s \leq 1$ , we denote by  $\mathcal{C}_{\mathcal{D}}^s$  the set of  $s$ -Hölder continuous functions  $f$  on  $\mathcal{D}$  that satisfy

$$(6) \quad |f(t, a) - f(t', a')| \leq c(|t - t'|^s + |a - a'|^s)$$

for every  $(t, a), (t', a') \in \mathcal{D}$  and some  $c > 0$ .

*The random measures.*  $\mathcal{M}_F$  denotes the set of finite point measures on  $\mathbb{R}_+ = [0, \infty)$  and  $\mathcal{M}_{F+}$  the set of positive finite measures on  $\mathbb{R}_+$ . Any  $Z \in \mathcal{M}_F$  admits the representation  $Z = \sum_{i=1}^n \delta_{a_i}$  for some ordered set  $\{a_1, \dots, a_n\} \subset \mathbb{R}_+$ . For a real-valued function  $f$  defined on  $\mathbb{R}_+$ , we write

$$\langle Z, f \rangle = \int_{\mathbb{R}_+} f(a) Z(da) = \sum_{i=1}^n f(a_i).$$

In particular  $n = \langle Z, \mathbf{1} \rangle$ . For  $Z = \sum_{i=1}^n \delta_{a_i} \in \mathcal{M}_F$ , abusing notation slightly, we define the evaluation maps  $a_i(Z) = a_i$  and for  $t \geq 0$ , the shift  $\tau_t Z = \sum_{i=1}^n \delta_{a_i+t}$ .

**2.2. Construction of the model.** The basic assumptions on the model are the following:

**Assumption 1.** *We have*

- (i)  $b \in \mathcal{L}_{\mathcal{D}}^{\infty}$  and  $\mu \in \mathcal{L}_{\mathcal{D}}^{\infty}$ ,
- (ii)  $NZ_0^N \in \mathcal{M}_F$  is random and satisfies<sup>1</sup>  $\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$  almost-surely; moreover  $Z_0^N \rightarrow \xi_0$  narrowly, for some deterministic  $\xi_0 \in \mathcal{M}_+$ ,
- (iii)  $\xi_0(da) = g_0(a)da$  for some  $g_0 \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$  such that  $\int_0^{\infty} g_0(a)da < \infty$ .

<sup>1</sup>where  $a_N \lesssim b_N$  means  $\sup_{N \geq 1} a_N b_N^{-1} < \infty$ .

For  $t \in [0, T]$ , consider the equation

$$(7) \quad \begin{aligned} Z_t^N &= \tau_t Z_0^N + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq b(s, a_i(Z_{s-}^N)), i \leq \langle N Z_{s-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta) \\ &\quad - N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{a_i(Z_{s-}^N) + t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_{s-}^N)), i \leq \langle N Z_{s-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(ds, di, d\vartheta), \end{aligned}$$

where  $\mathcal{Q}_i$ ,  $i = 1, 2$  are independent Poisson random measures on  $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$  with intensity measure  $ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$ . In this setting, the distribution  $Z_0^N$  describes the renormalised state of the population at time  $t = 0$  and  $N \langle Z_0^N, \mathbf{1} \rangle$  its size.

Under Assumption 1 (i), we have existence and strong uniqueness of a solution to (7) in  $\mathbb{D}([0, T], \mathcal{M}_+)$ , the Skorokhod space of càdlàg processes with values in  $\mathcal{M}_+$ . Under Assumption 1 (i) and (ii)<sup>2</sup>, we even have the narrow convergence of  $Z^N$  in  $\mathbb{D}([0, T], \mathcal{M}_+)$  to a deterministic limit  $\xi \in \mathcal{C}([0, T], \mathcal{M}_+)$ , see *e.g.* [43, 15]. Under Assumption 1 (iii), the limit  $\xi = (\xi_t(da))_{0 \leq t \leq T}$  is

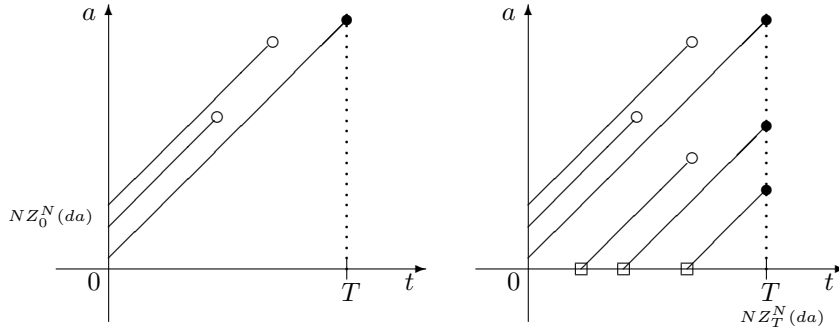


FIGURE 1. Sample path of  $N Z_0^N(da)$  and its evolution without births (left), sample path of  $(N Z_t^N(da))_{0 \leq t \leq T}$  (right).

smooth in the following sense: we have that  $\xi_t(da) = g(t, a)da$ , where  $g$  is a weak solution to the McKendrick Von Foerster equation (2) defined in Section 1.1 above (see [35, 46] and the comprehensive textbook of Perthame [40]). With the notation of Section 1.1, the equation  $\mathcal{H}_{b, \mu}^N(Z^N) = 0$  is given by (7) while  $\mathcal{H}_{b, \mu}(g) = 0$  is given by (2).

### 2.3. Stability of the model.

*Preliminaries.* The stability of  $Z_t^N(da)$  relative to its limit  $g(t, a)$  will be expressed in terms of weighted quantities of the form

$$\mathcal{W}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_{\mathbb{R}_+} w_2(t-a) f_t(a) (Z_t^N(da) - g(t, a)da)$$

and also

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \int_0^t w_1(s) \int_{\mathbb{R}_+} w_2(s-a) f_s(a) (Z_s^N(da) - g(s, a)da) ds,$$

<sup>2</sup>Actually, the condition of the almost-sure bound  $\sup_N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$  can be relaxed to the significant weaker moment condition  $\sup_{N \geq 1} \mathbb{E}[\langle Z_0^N, \mathbf{1} \rangle^{1+\epsilon}] < \infty$  for some  $\epsilon > 0$ .



where  $w_i$ ,  $i = 1, 2$  are two bounded weight functions (possibly taking negative values). For notational simplicity, we write  $f_t(a) = f(t, a)$  for  $f \in \mathcal{L}_D^\infty$  when no confusion is possible. Implicitly, we assume that  $\mathcal{F}$  is well-behaved in the sense that  $\mathcal{W}_{w_2}^N(\mathcal{F})_t$  and  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t$  are measurable, as random variables on the ambient probability space over which  $Z^N$  is defined.

*The structure of  $\mathcal{F}$ .* We describe the minimal structure we need to put on  $\mathcal{F}$  so that the subsequent concentration properties hold for  $\mathcal{W}_{w_2}^N(\mathcal{F})_t$  and  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_t$ . In particular, we must be able to control the complexity of  $\mathcal{F}$  measured in terms of entropy. Let  $\mathfrak{s}_t$ ,  $\mathfrak{t}_t$  and  $\mathfrak{u}_t$  be the operators on  $\mathcal{L}_D^\infty$  defined by

$$\mathfrak{s}_t(f) = ((s, a) \mapsto f(t, a + t)), \quad \mathfrak{t}_t(f) = ((s, a) \mapsto f(t, t - s)), \quad \mathfrak{u}_t(f) = ((s, a) \mapsto f(t, t + a - s)).$$

**Assumption 2.** *We have  $0, c_0, c_0 b, c_0 \mu \in \mathcal{F}$  for some constant  $c_0 > 0$ . Moreover, for every  $t \in [0, T]$ , the class  $\mathcal{F}$  is stable under the following operations:*

$$(8) \quad f \mapsto -f, \quad (f, g) \mapsto fg, \quad f \mapsto \mathfrak{s}_t(f), \quad f \mapsto \mathfrak{t}_t(f), \quad f \mapsto \mathfrak{u}_t(f).$$

Let  $\text{diam}_{|\cdot|_\infty}(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} |f - g|_\infty$  and write  $\mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)$  for the minimal number of  $\epsilon$ -balls for the  $|\cdot|_\infty$ -metric that are necessary to cover  $\mathcal{F}$ .

**Proposition 3.** *Let  $\mathcal{F}$  be the minimal set satisfying Assumption 2 for some  $c_0 > 0$  such that  $c_1 = c_0 \max(|b|_\infty, |\mu|_\infty) < 1$ . If moreover  $b, \mu \in \mathcal{C}_D^s$  for some  $s > 0$  ( $\mathcal{C}^s$  is the set of Hölder continuous functions defined in (6)), then*

$$(9) \quad e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

*Concentration properties.*

**Definition 4** (mild concentration). *A sequence of nonnegative random variables  $(X^N)_{N \geq 1}$  has a mild concentration property of order  $0 \leq r_N \rightarrow 0$  if for large enough  $N$ , we have*

$$\mathbb{P}(X^N \geq (1 + u)r_N) \leq \frac{1}{e^u - 1} \quad \text{for every } u \geq 0.$$

**Assumption 5.** *The sequence*

$$|w_2|_{1, \infty}^{-1} \max_{h=1, w_2} \mathcal{W}_h^N(\mathcal{F})_0$$

*has a mild concentration property of order  $r_N$  for some  $0 \leq r_N \rightarrow 0$ .*

**Theorem 6.** *Work under Assumptions 1, 2 and 5. Assume moreover  $\text{diam}_{|\cdot|_\infty}(\mathcal{F}) \leq 1$  and*

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

*If  $w_2$  has compact support with length support bounded in  $N$  by some  $u > 0$  and satisfies an estimate of the form*

$$(10) \quad |w_2|_\infty \lesssim \max(N^{1/2}, r_N^{-1}) |w_2|_1,$$

*then*

$$(|w_1|_{1, \infty} |w_2|_{1, \infty})^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_{1, \infty}^{-1} \mathcal{W}_{w_2}^N(\mathcal{F})_T$$

*share both a mild concentration property of order  $C \max(r_N, N^{-1/2})$ , for an explicitly computable  $C = C(u, e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty, g_0, |w_1|_1, |w_2|_1) > 0$  continuous in its arguments. In particular, if  $|w_i|_1$ ,  $i = 1, 2$  is uniformly bounded in  $N$ , then  $C$  can be chosen independently of  $N$ .*

Several remarks are in order: **1)** If the initial condition  $Z_0^N$  is close to its limit  $g_0$  in  $\mathcal{W}_{w_2}(\mathcal{F})_0$ -norm of order  $r_N$ , Theorem 6 states that the error inflates in  $\mathcal{W}_{w_2}(\mathcal{F})_t$ -norm by a factor no worse than  $N^{-1/2}$  for  $t \in [0, T]$ . In particular, whenever  $r_N \lesssim N^{-1/2}$ , the error propagation is stable. **2)** The order of magnitude of the error propagation is  $\max(N^{-1/2}, r_N)$ , as one could expect. As for the order in terms of  $w_1$  or  $w_2$ , the ideal order would be the integrated squared-error norm  $|w_i|_2$  as a variance term in a central limit theorem for instance. Here, we obtain the slightly worse interpolation quantity  $|w_i|_{1,\infty}$  which is always bigger than  $|w_i|_2$ . However, for statistical purposes, when  $w_i$  is replaced by a kernel  $w_i = h_N^{-1}K(h_N^{-1}\cdot)$  for some kernel  $K$  such that  $|K|_1 = 1$ , the order is sharp, since in that case

$$|w_i|_{1,\infty} \approx h_N^{-1/2} \approx |w_i|_2$$

and moreover  $|w_i|_1$  is uniformly bounded in  $N$ . The fact that we have here the correct order for dilating kernels is crucial for nonparametric estimation and is the main purpose (and difficulty) of Theorem 6. This seems to be a standard situation for nonparametric estimation in structured populations, where such effects are also met, see [13, 22, 3]. **3)** If  $w_2$  is not compactly supported or if (10) does not hold, we still have that

$$(|w_1|_{1,\infty}|w_2|_\infty)^{-1}\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_\infty^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T$$

share both a mild concentration property of order  $C \max(r_N, N^{-1/2})$ , as explicitly obtained in the proof. However, such a result is not sufficient for nonparametric estimation: picking  $w_2 = h_N^{-1}K(h_N^{-1}\cdot)$  yields  $|w_2|_\infty \approx h_N^{-1}$  which is dramatically worse than the expected  $h_N^{-1/2}$  in kernel estimation. **4)** The constant  $C$  also depends on the length of the support of  $w_2$ , but that may be considered as fixed once for all for later statistical purposes. **5)** Assumption 5 implies the moment estimate

$$(11) \quad \mathbb{E}\left[\max_{h=1,w_2} \mathcal{W}_h^N(\mathcal{F})_0^p\right] \lesssim |w_2|_{1,\infty}^p r_N^p \quad \text{for every } p > 0.$$

**6)** We finally give a reasonable and sufficient condition for Assumption 5 to hold.

**Proposition 7.** *If  $\mathcal{F}$  is uniformly bounded (in particular if  $\mathcal{F}$  is the minimal set of Proposition 3) and if  $NZ_0^N(da)$  consists of a  $N$ -drawn of independent random variables with common distribution  $g_0(a)da$  (with the normalisation assumption  $\int_{\mathbb{R}_+} g_0(a)da = 1$ ), we have Assumption 5.*

The proof is based on a concentration inequality of Klein and Rio [25] and is developed in a statistical setting in Comte *et al.* [10] and is delayed until Appendix 8.2

We end this section by giving a global stability result for the propagation of the error  $Z_t^N(da) - g(t, a)da$ , given a preliminary control on  $Z_0^N(da) - g(0, a)da$ , which relies on the techniques developed in Theorem 6, but with a weaker moment condition for the initial control of the particle system.

**Proposition 8.** *Work under Assumptions 1 and 2. If*

$$(12) \quad \mathbb{E}\left[\max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0^p\right] \leq |w_2|_{1,\infty}^p r_N^p$$

for some  $r_N \geq 0$  and  $p \geq 1$ , and if  $w_2$  is compactly supported and satisfies an estimate of the form  $|w_2|_\infty \lesssim \max(N^{1/2}, r_N^{-1})|w_2|_1$ , then

$$(13) \quad \mathbb{E}\left[\mathcal{W}_{w_2}^N(\mathcal{F})_T^p\right] \lesssim |w_2|_{1,\infty}^p \max(N^{-p/2}, r_N^p)$$

and

$$(14) \quad \mathbb{E}\left[\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T^p\right] \lesssim (|w_1|_{1,\infty}|w_2|_{1,\infty})^p \max(N^{-p/2}, r_N^p).$$

3. NONPARAMETRIC ESTIMATION OF  $g$  AND  $\mu$ 

## 3.1. Kernel approximation.

**Definition 9.** A kernel  $K$  of (integer) order  $\ell_0 \geq 1$  is a bounded function with compact support in  $\mathbb{R}_+$  such that

$$\int_0^\infty \kappa^{\ell-1} K(\kappa) d\kappa = \mathbf{1}_{\{\ell=1\}}, \quad \text{for } \ell = 1, \dots, \ell_0 - 1.$$

For a bandwidth  $h > 0$ , we set  $K_h(\kappa) = h^{-1}K(h^{-1}\kappa)$  so that  $|K_h|_1 = |K|_1$ . In order to approximate functions of  $\mathcal{L}_{\mathcal{D}}^\infty$ , we use bivariate kernels defined by

$$H \otimes K(t, a) = H(t)K(a) \quad \text{for } (t, a) \in \mathcal{D},$$

with  $H \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$ ,  $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ . For a bivariate bandwidth  $\mathbf{h} = (h_1, h_2)$  with  $h_i > 0$ , let

$$(H \otimes K)_{\mathbf{h}}(t, a) = H_{h_1}(t)K_{h_2}(a)$$

and define the linear approximation

$$(15) \quad (H \otimes K)_{\mathbf{h}} \star f(t, a) = \int_0^T \int_0^\infty f(s, u) (H \otimes K)_{\mathbf{h}}(t-s, a-u) ds du.$$

We may also approximate  $f$  in another system of coordinates: if  $\varphi : \mathcal{D} \rightarrow \mathcal{D}$  is invertible, reparametrise  $f$  via

$$f(t, a) = \tilde{f} \circ \varphi(t, a)$$

and define the  $\varphi$ -skewed linear approximation

$$(H \otimes K)_{\mathbf{h} \circ \varphi} \star f(t, a) = \int_0^T \int_0^\infty f(s, u) ((H \otimes K)_{\mathbf{h} \circ \varphi})(s-t, u-a) ds du$$

so that  $((H \otimes K)_{\mathbf{h} \circ \varphi} \star f)(t, a) = (H \otimes K)_{\mathbf{h}} \star \tilde{f}(\varphi(t, a))$ . The  $\varphi$ -skewed approximation potentially has better approximation properties for  $\tilde{f}$  in the vicinity of  $\varphi(t, a)$  than  $f$  in the vicinity of  $(t, a)$ , as will become transparent in Section 4 below.

3.2. Construction of estimators of  $g$  and  $\mu$ .

*Construction of an estimator of  $g$ .* Let  $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$  be a kernel of order  $\ell_0 \geq 0$ . For  $(t, a) \in \mathcal{D}$ , we consider the family of estimators

$$(16) \quad \hat{g}_h^N(t, a) = K_h \star Z_t^N(a) = \int_{\mathbb{R}_+} K_h(u-a) Z_t^N(du), \quad h > 0.$$

**Remark 10.** At first glance, it may seem slightly suprising to build an estimator of the bivariate function  $g(t, a)$  by means of (16) that uses data  $Z_t^N$  only and discards the observation  $(Z_s^N, s \neq t)$ . For instance, one may consider estimators of the form

$$((H \otimes K)_{\mathbf{h} \circ \varphi} \star Z^N)(t, a) = \int_0^T \int_{\mathbb{R}_+} ((H \otimes K)_{\mathbf{h} \circ \varphi})(s-t, u-a) Z_s^N(du)$$

Formally  $\hat{g}_h^N(t, a) = (H_{h_1=0} \otimes K_h) \star Z^N(t, a)$  without any specific change of coordinates and we will see that such a simple procedure already achieves minimax optimality, see Section 4.3 below.

*Construction of the process of death occurrences.* We first extract from the data  $(Z_t^N(da))_{0 \leq t \leq T}$  the random measure

$$\Gamma^N(dt, da) = \sum_{i \geq 1} \delta_{(T_i, A_i)}(dt, da) \text{ on } [0, T] \times \mathbb{R}_+$$

associated with the successive times  $T_i$  of the death occurrences of the population during the observation period  $[0, T]$ , together with the corresponding ages  $A_i$  of the individuals that die at time  $T_i$ .

Remember that the evaluation mappings  $a_i(Z_t^N)$  in the representation  $Z_t^N = N^{-1} \sum_{i \geq 1} \delta_{a_i(Z_t^N)}$  are ordered:

$$a_1(Z_t^N) < a_2(Z_t^N) < \dots$$

and that  $t \mapsto a_i(Z_t)$  is increasing with slope one unless a birth or a death occurs, in which case we have a non-negative or a negative jump. It follows that

$$(17) \quad \Gamma^N(dt, da) = \sum_{s > 0} \mathbf{1}_{\{i^* = \inf\{i \geq 1, \Delta a_i(Z_s^N) > 0\} < \infty\}} \delta_{(s, a_{i^*}(Z_s^N))}(dt, da)$$

on  $[0, T] \times \mathbb{R}_+$ , where we set  $\Delta a_i(Z_s^N) = a_i(Z_s^N) - a_i(Z_{s-}^N)$  and with the usual convention  $\inf \emptyset = \infty$ . This second representation in terms of the jump measure of the processes  $a_i(Z_t^N)$  gives an explicit construction of  $\Gamma^N(dt, da)$  as a function of  $(Z_t^N(da), t \in [0, T])$ .

*Construction of an estimator of  $\mu$ .* Let  $H \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$  and  $K \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$  be two kernels. For  $(t, a) \in \mathcal{D}$  and  $\varphi(t, a) = (t, t - a)$ , consider the family

$$(18) \quad \hat{\pi}_{\mathbf{h}}^N(t, a) = \int_0^T \int_{\mathbb{R}_+} ((H \otimes K)_{\mathbf{h}} \circ \varphi)(s - t, u - a) \Gamma^N(ds, du), \quad \mathbf{h} = (h_1, h_2) \text{ with } h_i > 0,$$

that estimate the function  $\pi = \mu g$ . An estimator of  $\mu(t, a)$  is obtained by considering the ratio

$$(19) \quad \hat{\mu}_{\mathbf{h}, \varpi}^N(t, a) = \frac{\hat{\pi}_{\mathbf{h}}^N(t, a)}{\hat{g}_{\mathbf{h}}^N(t, a) \vee \varpi}$$

for some threshold  $\varpi > 0$ , and is thus specified by the bandwidths  $h > 0$ ,  $\mathbf{h} = (h_1, h_2)$  with  $h_i > 0$  and  $\varpi > 0$ .

### 3.3. Oracle inequalities.

*Estimation of  $g$ , data-driven bandwidth.* Pick a lattice  $\mathcal{G}_1^N$  included in  $[N^{-1/2}, (\log N)^{-1}]$  and such that  $\text{Card}(\mathcal{G}_1^N) \lesssim N$ . The algorithm, based on the Lepski's principle as defined in the Goldenshluger-Lepski's method [17, 18] requires the family of linear estimators

$$\left( \hat{g}_{\mathbf{h}}^N(t, a), \mathbf{h} \in \mathcal{G}_1^N \right)$$

defined in (16) and selects an appropriate bandwidth  $h = \hat{h}^N(t, a)$  from the data  $(Z_t^N(da))_{0 \leq t \leq T}$ . For  $(t, a) \in \mathcal{D}$ , writing  $\{x\}_+ = \max(x, 0)$ , define

$$\mathbf{A}_{\mathbf{h}}^N(t, a) = \max_{h' \leq h, h' \in \mathcal{G}_1^N} \left\{ \left( \hat{g}_{\mathbf{h}}^N(t, a) - \hat{g}_{\mathbf{h}'}^N(t, a) \right)^2 - (\mathbf{V}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}'}^N) \right\}_+,$$

where

$$(20) \quad \mathbf{V}_{\mathbf{h}}^N = (4(\log N)C^*N^{-1/2}|K_{\mathbf{h}}|_{1, \infty})^2$$

and  $C^*$  is a (known) upper bound of the constant  $C$  of Theorem 6. (Remember that the constant  $C$  depends on the parameters of the model via  $|b|_{\infty}$ ,  $|\mu|_{\infty}$  and  $g_0$ .) Let

$$\widehat{h}^N(t, a) \in \operatorname{argmin}_{h \in \mathcal{G}_1^N} (\mathbf{A}_h^N(t, a) + \mathbf{V}_h^N).$$

The data-driven Goldenshluger-Lepski estimator of  $g(t, a)$  is defined as

$$(21) \quad \widehat{g}_\star^N(t, a) = \widehat{g}_{\widehat{h}^N(t, a)}^N(t, a).$$

*Oracle estimate.* We need some notation. Given a kernel  $K_h$ , the bias at scale  $h$  of  $g$  at point  $(t, a)$  is defined as

$$(22) \quad \mathcal{B}_h^N(g)(t, a) = \sup_{h' \leq h, h' \in \mathcal{G}_1^N} \left| \int_0^\infty K_{h'}(u - a)g(t, u)du - g(t, a) \right|.$$

We are ready to give our first estimation result for every  $(t, a) \in \mathcal{D}_- = \mathcal{D} \setminus \{t = a\}$ .

**Theorem 11.** *Work under Assumptions 1, 2 and 5 with  $r_N \leq N^{-1/2}$  and some  $\mathcal{F}$  that satisfies  $e(\mathcal{F}) < \infty$ . For  $(t, a) \in \mathcal{D}_-$ , specify  $\widehat{g}_\star^N(t, a)$  with a bounded and compactly supported kernel  $K$ . The following oracle inequality holds true*

$$\mathbb{E}[(\widehat{g}_\star^N(t, a) - g(t, a))^2] \lesssim \inf_{h \in \mathcal{G}_1^N} (\mathcal{B}_h^N(g)(t, a)^2 + \mathbf{V}_h^N) + \delta_N$$

for large enough  $N$ , with  $\delta_N = N^{-1}$  and up to a constant that depends on  $C^\star$  and  $K$ .

Some remarks: **1)** The fact that we measure the performance of  $\widehat{g}_\star^N$  at point  $(t, a)$  in pointwise squared-error loss is inessential here. Other integrated norms like  $|\cdot|_p$  would work as well, following the general proof of Lepski's principle [29, 17, 18]. However, if we need a fine control of the bias in terms of smoothness space, this is no longer true and is linked to the anisotropic and spatial inhomogeneous smoothness structure of the solution  $g$ . This will become transparent in Theorems 19 and 33 below. **2)** In (20), the choice of  $C^\star$  has to be set in principle prior to the data analysis and is of course difficult to calibrate. It depends on upper bounds on many quantities like  $e(\mathcal{F})$  that appear in the constant of Theorem 6 or supremum of norms of the unknown parameters  $b$  and  $\mu$ . Moreover, the explicit value  $C^\star$  obtained by tracking the constants in the computations of Section 6 is certainly too large. In practice, we need to inject some further prior knowledge and calibrate the threshold by some other method, possibly using data. Such approaches in the context of Lepski's principle have been developed lately in [26]. **3)** The proof relies on Theorem 6 which requires  $e(\mathcal{F})$  to be finite. However, this requirement is not heavy, as soon as  $b$  and  $\mu$  have a minimal global Hölder smoothness, as stems from Proposition 3.

*Estimation of  $\mu$ , data-driven bandwidth.* Analogously to the bandwidth-selection method for estimation of  $g$  following Lepski's principle, we pick a discrete set  $\mathcal{G}_2^N \subset [N^{-1/2}, (\log N)^{-1}]^2$  with cardinality  $\operatorname{Card}\mathcal{G}_2^N \lesssim N$ . The construction is similar to that of  $\widehat{g}_\star^N(t, a)$ , given in addition the family of estimators

$$(\widehat{\pi}_\mathbf{h}^N(t, a), \mathbf{h} \in \mathcal{G}_2^N)$$

defined in (18). For  $(t, a) \in \mathcal{D}$ , let

$$\mathbf{A}_\mathbf{h}^N(t, a) = \max_{h' \in \mathcal{G}_2^N} \{ (\widehat{\pi}_\mathbf{h}^N(t, a) - \widehat{\pi}_{\mathbf{h}'}^N(t, a))^2 - (\mathbf{V}_\mathbf{h}^N + \mathbf{V}_{\mathbf{h}'}^N) \}_+,$$

where

$$(23) \quad \mathbf{V}_\mathbf{h}^N = (4(\log N)C^\star N^{-1/2} |H_{h_1}|_{1, \infty} |K_{h_2}|_{1, \infty})^2$$

and  $C^\star$  is a (known) upper bound of the constant  $C$  of Theorem 6. Let

$$\widehat{\mathbf{h}}^N(t, a) \in \operatorname{argmin}_{\mathbf{h} \in \mathcal{G}_2^N} (\mathbf{A}_\mathbf{h}^N(t, a) + \mathbf{V}_\mathbf{h}^N(t, a)).$$

The data-driven Goldenshluger-Lepski estimator of  $\mu(t, a)$  is defined as

$$(24) \quad \widehat{\mu}_*^N(t, a)_\varpi = \mu_{\widehat{h}^N(t, a), \widehat{h}^N(t, a)}^N(t, a)_\varpi.$$

*Oracle estimates.* In order to estimate  $\mu$  in squared-error loss consistently with the quotient estimator (24), we need a (local) lower bound assumption on  $g(t, a)$ . Let

$$\begin{aligned} \mathcal{D}_U &= \{(t, a) \in \mathcal{D}, a > t\}, \\ \mathcal{D}_L &= \{(t, a) \in \mathcal{D}, a < t\}, \end{aligned}$$

and  $\mathcal{D}^- = \mathcal{D} \setminus \{t = a\}$  so that  $\mathcal{D}^- = \mathcal{D}_L \cup \mathcal{D}_U$ . A sufficient condition is given by the following

**Assumption 12.** *For every  $(t, a) \in \mathcal{D}^-$  there exists an open set  $\mathcal{U}_{(t, a)}$  such that*

$$(25) \quad \inf_{u \in \mathcal{U}_{(t, a)}} b(t - a, t - a + u)g_0(u) \geq \delta \text{ if } (t, a) \in \mathcal{D}_L$$

and

$$(26) \quad g_0(t - a) \geq \delta \text{ if } (t, a) \in \mathcal{D}_U,$$

for some  $\delta > 0$ .

We need some notation. For  $\mathbf{h} = (h_1, h_2)$  and  $\mathbf{h}' = (h'_1, h'_2)$  in  $\mathcal{G}_2^N$ , we say that  $\mathbf{h} \leq \mathbf{h}'$  if  $h_1 \leq h'_1$  and  $h_2 \leq h'_2$  hold simultaneously. Given a bivariate kernel  $H \otimes K$ , the bias at scale  $\mathbf{h}$  of  $\pi = \mu g$  at point  $(t, a)$  in the direction  $\varphi$  is defined as

$$(27) \quad \mathcal{B}_{\mathbf{h}}^N(\mu g)(t, a) = \sup_{\mathbf{h}' \leq \mathbf{h}, \mathbf{h}' \in \mathcal{G}_2^N} \left| \int_{\mathcal{D}} ((H \otimes K)_{\mathbf{h}'} \circ \varphi)(s - t, u - a) \pi(s, u) duds - \pi(t, a) \right|.$$

**Theorem 13.** *Work under Assumptions 1, 2, 5 with  $r_N \leq N^{-1/2}$  and some  $\mathcal{F}$  that satisfies  $\epsilon(\mathcal{F}) < \infty$  together with Assumption 12. For  $(t, a) \in \mathcal{D}^-$  specify  $\widehat{\mu}_*^N(t, a)_\varpi$  with kernels  $H, K$ . The following oracle inequality holds true*

$$\mathbb{E}[(\mu_*^N(t, a)_\varpi - \mu(t, a))^2] \lesssim \inf_{\mathbf{h} \in \mathcal{G}_1^N} (\mathcal{B}_{\mathbf{h}}^N(g)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) + \inf_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\mu g)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) + \delta_N$$

for large enough  $N$  and small enough  $\varpi > 0$ , with  $\delta_N = N^{-1}$  and up to a constant that depends on  $C^*$  and the kernels  $H, K$ .

Some remarks: **1)** Similar to the case of Theorem 11, other loss functions can be chosen. **2)** We see that the performance of  $\widehat{\mu}_*^N(t, a)_\varpi$  is similar to the worst performance of the estimation of the product  $\pi = \mu g$  and the estimation of  $g$ , as is standard in the study of quotient estimator in the classical Nadaraya-Watson (NW) sense [2, 37]. However, the situation is quite different here than what is customary in standard nonparametric regression with NW: the estimation of  $g(t, a)$  is actually equivalent to the estimation of a univariate function, while  $\pi(t, a)$  is related to a genuinely bi-variate estimation problem that suffers from a dimensional effect. Therefore, there is good hope to obtain here an optimal procedure, as will become transparent under Hölder anisotropic smoothness scales in the subsequent minimax theorems 18 and 20 below. **3)** The same remark about the choice of  $C^*$  (and also the threshold  $\varpi$ ) as in Theorem 11 above are valid in the context of the estimation of  $\mu(t, a)$ .

## 4. ADAPTIVE ESTIMATION UNDER ANISOTROPIC HÖLDER SMOOTHNESS

## 4.1. The smoothness of the McKendrick Von Foerster equation.

**Definition 14.** Let  $\alpha > 0$ ,  $x_0 \in \mathbb{R}$  and  $\mathcal{U}_{x_0}$  be a neighbourhood of  $x_0$ . We say that  $f : \mathcal{U}_{x_0} \rightarrow \mathbb{R}$  belongs to  $\mathcal{H}^\alpha(x_0)$  if<sup>3</sup> for every  $x, y \in \mathcal{U}_{x_0}$

$$(28) \quad |f^{(n)}(y) - f^{(n)}(x)| \leq C|y - x|^{\{\alpha\}}$$

having  $\alpha = n + \{\alpha\}$  for a non-negative integer  $n$  and  $0 < \{\alpha\} \leq 1$ .

We obtain a semi-norm by setting  $|f|_{\mathcal{H}^\alpha(x_0)} = \sup_{x \in \mathcal{U}_{x_0}} |f(x)| + C_{\mathcal{U}_{x_0}}(f)$ , where  $C_{\mathcal{U}_{x_0}}(f)$  is the smallest constant  $C$  for which (28) holds. The extension to multivariate functions is straightforward:

**Definition 15.** The bivariate function  $f$  belongs to the anisotropic Hölder class  $\mathcal{H}^{\alpha_1, \alpha_2}(x_0, y_0)$  if

$$|f|_{\mathcal{H}^{\alpha_1, \alpha_2}(x_0, y_0)} = |f(\cdot, y_0)|_{\mathcal{H}^{\alpha_1}(x_0)} + |f(x_0, \cdot)|_{\mathcal{H}^{\alpha_2}(y_0)} < \infty.$$

We write  $f \in \mathcal{H}$  if for every  $(t, a) \in \mathcal{D}$ , we have  $f \in \mathcal{H}^{\sigma, \tau}(t, a)$ .

**Assumption 16.** For some  $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$  and for every  $(t, a) \in \mathcal{D}$ , we have

$$b \in \mathcal{H}^{\alpha, \beta}(t, a), \quad \mu \in \mathcal{H}^{\gamma, \delta}(t, a), \quad g_0 \in \mathcal{H}^\nu(a).$$

We give two results about the pointwise smoothness of the solution of the McKendrick Von Foerster equation on  $\mathcal{D}^- = \mathcal{D} \setminus \{t = a\}$ , depending on the choice of coordinates. The smoothness of  $g$  differs on  $\mathcal{D}_U$  where only mortality affects the population and  $\mathcal{D}_L$ , where both mortality and birth come into play. Introduce also the change of coordinates  $\varphi(t, a) = (t, t - a)$  that maps

$$\mathcal{D}_U \rightarrow \varphi(\mathcal{D}_U) = \tilde{\mathcal{D}}_U = \{(t, a') \in \mathcal{D}, 0 \leq t \leq T, a' < 0\}$$

$$\mathcal{D}_L \rightarrow \tilde{\mathcal{D}}_L = \varphi(\mathcal{D}_L) = \{(t, a') \in \mathcal{D}, 0 \leq t \leq T, 0 < a' < t\}$$

onto smoothly. This defines in turn

$$\tilde{g} : \tilde{\mathcal{D}}_U \cup \tilde{\mathcal{D}}_L \rightarrow \mathbb{R}_+ \quad \text{via} \quad g(t, a) = \tilde{g} \circ \varphi(t, a).$$

**Proposition 17.** Work under Assumptions 1, and 16.

(i) We have

$$g \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta), \min(\alpha, \beta, \gamma+1, \delta)} \quad \text{on} \quad \mathcal{D}_L \quad \text{and} \quad g \in \mathcal{H}^{\min(\gamma+1, \delta), \max(\gamma \wedge (\delta+1), \delta)} \quad \text{on} \quad \mathcal{D}_U.$$

(ii) We have the following improvement of the anisotropic smoothness when the parametrisation is given by  $\tilde{g}$ :

$$\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\alpha, \beta, \gamma+1, \delta)} \quad \text{on} \quad \tilde{\mathcal{D}}_L \quad \text{and} \quad \tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \max(\gamma \wedge (\delta+1), \delta)} \quad \text{on} \quad \tilde{\mathcal{D}}_U.$$

The proof of Proposition 17 is relatively straightforward, given explicit representations of the solution  $g$  in terms of  $b$ ,  $\mu$  and  $g_0$ , and is given in Appendix 8.3.

<sup>3</sup>The definition depends on  $\mathcal{U}_{x_0}$ , further omitted in the notation.

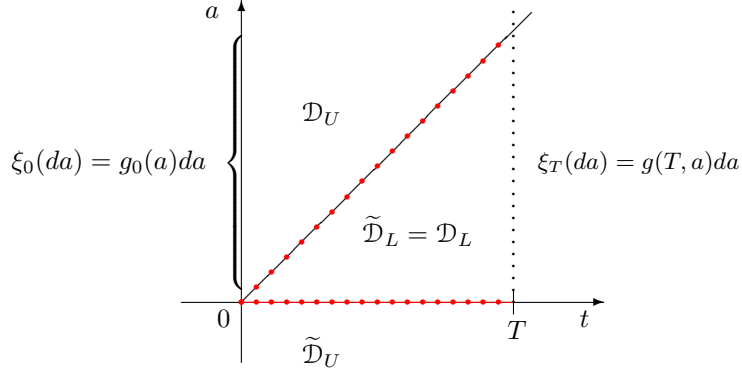


FIGURE 2.  $\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\alpha, \beta, \gamma+1, \delta)}$  on  $\tilde{\mathcal{D}}_L$  and  $\tilde{g} \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\gamma, \delta+1)}$  on  $\tilde{\mathcal{D}}_U$ .

**4.2. Minimax lower bounds.** For  $\alpha, \beta > 0$  and  $L > 0$ , we set

$$\mathcal{H}_L^{\alpha, \beta}(t, a) = \{f \in \mathcal{L}_{\mathcal{D}}^{\infty}, |f|_{\infty} + |f|_{\mathcal{H}^{\alpha, \beta}(t, a)} \leq L\},$$

where the semi-norm  $|\cdot|_{\mathcal{H}^{\alpha, \beta}(t, a)}$  is defined after Definition 15. We also set, for  $\epsilon > 0$ ,

$$\mathcal{L}_{\mathcal{D}, \epsilon}^{\infty} = \{f \in \mathcal{L}_{\mathcal{D}}^{\infty}, \inf_{(t, a) \in \mathcal{D}} f(t, a) \geq \epsilon\}$$

Remember that under Assumption 1, any point  $(b, \mu, g_0)$  with  $b, \mu, g_0 \in \mathcal{L}_{\mathcal{D}}^{\infty}$  defines a unique solution  $g$  to the McKendrick Von Foester equation (2). Let

$$s_{\text{dens}}^- = \max(\gamma, \delta) \text{ and } s_{\text{death}}^- = (\gamma^{-1} + \delta^{-1})^{-1}.$$

Under a non-degeneracy condition of the form  $\mu \in \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty}$ , we obtain the following minimax lower bound:

**Theorem 18.** *Work under Assumptions 1 and 12. Let  $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$  and  $L > 0$ . For every  $(t, a) \in \mathcal{D}^-$ , we have*

$$(29) \quad \inf_F \sup_{b, \mu, g_0} \mathbb{E}[|F - g(t, a)|] \gtrsim N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)}$$

and

$$(30) \quad \inf_F \sup_{b, \mu, g_0} \mathbb{E}[|F - \mu(t, a)|] \gtrsim N^{-s_{\text{death}}^- / (2s_{\text{death}}^- + 1)},$$

where the infimum is taken over all estimators and the supremum over

$$b \in \mathcal{H}_L^{\alpha, \beta}(t, a), \quad \mu \in \mathcal{H}_L^{\gamma, \delta}(t, a) \cap \mathcal{L}_{\mathcal{D}, \epsilon}^{\infty} \text{ and } g_0 \in \mathcal{H}_L^{\nu}(t, a).$$

Some remarks: **1)** As for the previous estimation results in Theorems 11 and 13, a glance at the proof shows that the lower bound actually holds for a wider class of loss functions, including loss in probability. We keep up to the statements (29) and (30) in expected pointwise absolute value for simplicity. **2)** If we take  $\gamma = \delta$  for simplicity, we see that  $s_{\text{dens}}^- = \gamma$  while  $s_{\text{death}}^- = \gamma/2$ . Therefore, although we are estimating bi-variate functions, the estimation difficulty for  $g(t, a)$  is really that of a 1-dimensional function while the estimation of  $\mu(t, a)$  remains that of a genuinely bivariate function. Heuristically, there is no information about the population density  $g(t, a)$  captured by  $(Z_s^N, s \neq t)$  while the estimation of the death rate  $\mu(t, a)$  requires dynamical knowledge from the process  $\Gamma^N(ds, du)$  for which a truly 2-dimensional information domain around  $(t, a)$  is required in order to identify  $\mu(t, a)$ .



**4.3. Adaptive estimation under anisotropic Hölder smoothness.** Our next result shows the performance of  $g_\star^N(t, a)$  defined in (21) and gives optimal up to inessential logarithmic factors in some cases. Moreover,  $g_\star^N(t, a)$  is nearly smoothness adaptive. More precisely, let

$$(31) \quad s_{\text{dens}}^+(t, a) = \max(\gamma \wedge (\delta + 1), \delta) \mathbf{1}_{\mathcal{D}_U}(t, a) + \min(\alpha, \beta, \gamma + 1, \delta) \mathbf{1}_{\mathcal{D}_L}(t, a),$$

and note that  $s_{\text{dens}}^+(t, a) \leq s_{\text{dens}}^-(t, a)$  always.

**Theorem 19.** *Work under Assumptions 1, 2, 5 with  $r_N \leq N^{-1/2}$  and some  $\mathcal{F}$  that satisfies  $e(\mathcal{F}) < \infty$ , and Assumption 12. Specify  $\widehat{g}_\star^N(t, a)$  with a compactly supported kernel of order  $\ell_0 \geq 0$  and pick*

$$\mathcal{G}_1^N = (x_1^N < x_2^N < \dots < x_N^N)$$

a subdivision of  $[N^{-1/2}, (\log N)^{-1}]$  with  $\max_{1 \leq i \leq N-1} (x_{i+1}^N - x_i^N) \lesssim N^{-1}$  so that  $\text{Card } \mathcal{G}_1^N \lesssim N$ . For every  $(t, a) \in \mathcal{D}_-$  and large enough  $N$ , we have

$$(32) \quad \sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{g}_\star^N(t, a) - g(t, a))^2])^{1/2} \lesssim \left( \frac{(\log N)^2}{N} \right)^{s_{\text{dens}}^+(t, a) \wedge \ell_0 / (2s_{\text{dens}}^+(t, a) \wedge \ell_0 + 1)},$$

where the supremum is taken over  $b \in \mathcal{H}_L^{\alpha, \beta}(t, a)$ ,  $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$ ,  $g_0 \in \mathcal{H}_L^\nu(t, a)$  with  $\alpha, \beta, \gamma, \delta > 0, \nu \geq \max(\gamma, \delta) + 1$  and  $L > 0$ .

Some remarks: **1)** Comparing with the minimax lower bound of Theorem 18, we see that both upper and lower bounds (29) and (32) agree on  $\mathcal{D}_U$  if  $\delta \leq \gamma \leq \delta + 1$  and on  $\mathcal{D}_L$  if  $\delta - 1 \leq \gamma \leq \delta$  (and if  $\alpha$  and  $\beta$  are sufficiently large too), provided the order  $\ell_0$  of the kernel  $K$  is sufficiently large. The rates are tight up to an inessential logarithmic factor. We do not know about the optimality in  $g$  beyond this domain, but we see that the difficulty of the estimation of  $g(t, a)$  is equivalent to the difficulty of the univariate function  $a \mapsto g(t, a)$  for which the time variable  $t$  is simply a parameter: it suffices to piece together the estimators  $\widehat{g}_\star^N(t, a)$  for every  $t$  in order to estimate the graph  $(t, a) \mapsto g(t, a)$ . **2)** While we already know that a logarithmic payment is unavoidable for a smoothness adaptive estimator (see the classical Lepski-Low phenomenon, [29, 33]) we do not know whether the order we find in the log term is correct (*i.e.*  $(\log N)^2$  versus the classical  $\log N$  payment). This stems from Theorem 6 and the mild concentration property as we define it, where exponential tail are obtained versus subgaussian tails, but this order seems genuinely linked to the Poissonian behaviour of the noise and it is not clear that we can extend our statistical result in order to remove the extra  $\log N$  error-term in (32).

Similarly,  $\mu_\star^N(t, a)$  defined in (24) also shares near optimality in some cases. Define

$$s_L(\alpha, \beta, \gamma, \delta) = (\min(\gamma, \delta)^{-1} + \min(\alpha, \beta, \gamma + 1, \delta)^{-1})^{-1},$$

$$s_U(\gamma, \delta) = (\min(\gamma, \delta)^{-1} + \delta^{-1})^{-1},$$

and

$$(33) \quad s_{\text{death}}^+(t, a) = s_U(\gamma, \delta) \mathbf{1}_{\mathcal{D}_U}(t, a) + s_L(\alpha, \beta, \gamma, \delta) \mathbf{1}_{\mathcal{D}_L}(t, a).$$

Note that  $s_{\text{death}}^+(t, a) \leq s_{\text{death}}^-(t, a)$  always.

**Theorem 20.** *Work under Assumptions 1, 2, 5 with  $r_N \leq N^{-1/2}$  and some  $\mathcal{F}$  that satisfies  $e(\mathcal{F}) < \infty$ , and Assumption 12. Specify  $\widehat{\mu}_\star^N(t, a)$  with kernels  $H, K$  of order  $\ell_0 \geq 0$  and pick  $\mathcal{G}_2^N = \mathcal{G}_1^N \times \mathcal{G}_1^N$  so that  $\text{Card } \mathcal{G}_2^N \lesssim N^2$ . For every  $(t, a) \in \mathcal{D}^-$  and large enough  $N$ , we have*

$$(34) \quad \sup_{b, \mu, g_0} (\mathbb{E}[(\widehat{\mu}_\star^N(t, a) - \mu(t, a))^2])^{1/2} \lesssim \left( \frac{(\log N)^2}{N} \right)^{s_{\text{death}}^+(t, a) \wedge \ell_0 / (2s_{\text{death}}^+(t, a) \wedge \ell_0 + 1)},$$

where the supremum is taken over  $b \in \mathcal{H}_L^{\alpha, \beta}(t, a)$ ,  $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$ ,  $g_0 \in \mathcal{H}_L^\nu(t, a)$ , with  $\alpha, \beta, \gamma, \delta > 0$ ,  $\nu \geq \max(\gamma, \delta) + 1$  and  $L > 0$ .

Some remarks: **1)** The same remark as **2)** after the statement of Theorem 19 holds here. **2)** The minimax optimality situation is somewhat clearer for estimating  $\mu$ : we see that we have near optimality on  $\mathcal{D}_U$  as soon as  $\gamma \leq \delta$ , while the upper and lower bounds only agree if  $\gamma \leq \delta \leq \gamma + 1$  on  $\mathcal{D}_L$  (and if  $\alpha$  and  $\beta$  are sufficiently large too), provided the order  $\ell_0$  of the kernel  $K$  is sufficiently large. Thus situation is somewhat similar to the estimation of  $g$  on  $\mathcal{D}_U$ , see Theorem 19 above. **3)** The rate of estimation is triggered by the smoothness of  $\pi = \mu g$  since the estimation of the quotient  $g$  will always be better, for

$$s_{\text{death}}^+(t, a) \leq s_{\text{dens}}^+(t, a) \text{ for every } (t, a) \in \mathcal{D}_-$$

always. However, in order to achieve optimality, we need to optimise the approximation property of  $\pi$  by looking at the smoothness of  $\tilde{\pi} = \tilde{\mu} \tilde{g}$ , with  $\mu = \tilde{\mu} \circ \varphi$ . This benefit is obtained thanks to Proposition 17 and is given in details in the proof. We would lose by a polynomial order in the rate of convergence given in (34) if we used a kernel of the form  $(H \otimes K)_{\mathbf{h}}$  instead of  $(H \otimes K)_{\mathbf{h}} \circ \varphi$  for the estimation of the numerator  $\pi$  in the representation  $\mu = \pi/g$ .

## 5. NUMERICAL ILLUSTRATION

We briefly explore the performance of our estimators on simulated data. We use the following parameters:

- (i) The initial condition  $g_0$  is taken as the density of Gaussian random variable centred in 40 with variance of 152 (*i.e.* a standard deviation of approximately 12 years), conditioned on living between 0 and 120.
- (ii) We pick  $b(t, a) = \mathbf{1}_{\{120 \leq a \leq 40\}}$ . Although  $b$  is not globally Hölder continuous, we still have (and can prove) similar results for such simple piecewise constant functions.
- (iii) We pick  $\mu(t, a) = 4 \cdot 10^{-2} \exp(7.4 \cdot 10^{-3} a) \exp(-5 \cdot 10^{-3} t)$ . We pick a relatively high death rate in order to guarantee sufficiently many events of death for the estimation of  $\mu(t, a)$  and avoid artefacts.

We consider the domain  $\mathcal{D} = [0, 20] \times [0, 120]$  which means  $T = 20$  and a maximal possible age of 120. We estimate  $g$  on the grid  $\mathcal{G}^g = \mathcal{T}^g \times \mathcal{A}^g$ , with  $\mathcal{T}^g = \{k \times 1.005, 0 \leq k < 20\}$  and  $\mathcal{A}^g = \{k \times 0.2002, 0 \leq k < 600\}$ . We estimate the functions  $\mu$  and  $\pi = \mu \cdot g$  on the grids  $\mathcal{G}^\mu = \mathcal{T}^\mu \times \mathcal{A}^\mu$ , with  $\mathcal{T}^\mu = \mathcal{T}^g$  and  $\mathcal{A}^\mu = \{k \times 1.0008, 0 \leq k < 120\}$ .

We first estimate  $g$  and  $\pi = \mu \cdot g$  and obtain consistent results in the regime  $N = 4000$ .

We end this section by exploring the estimation of  $\mu$  via our quotient estimator.

## 6. PROOF OF THEOREM 6

This section is devoted to the proof of the concentration properties of the model stated in Theorem 6. Recall that  $w_1 \in \mathcal{L}_{\mathcal{D}}^{\text{time}}$  and  $w_2 \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$  are two continuous weight functions. We introduce two fundamental processes for which we will establish concentration properties:

$$\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| \int_0^t w_1(s) M_s^N(w_2(s - \cdot) f_s) ds \right|,$$

where  $M_t^N(f)$  is defined in (36) below and

$$\mathcal{M}_{w_2}^N(\mathcal{F})_t = \sup_{f \in \mathcal{F}} \left| M_t^N(w_2(t - \cdot) f_t) \right|.$$

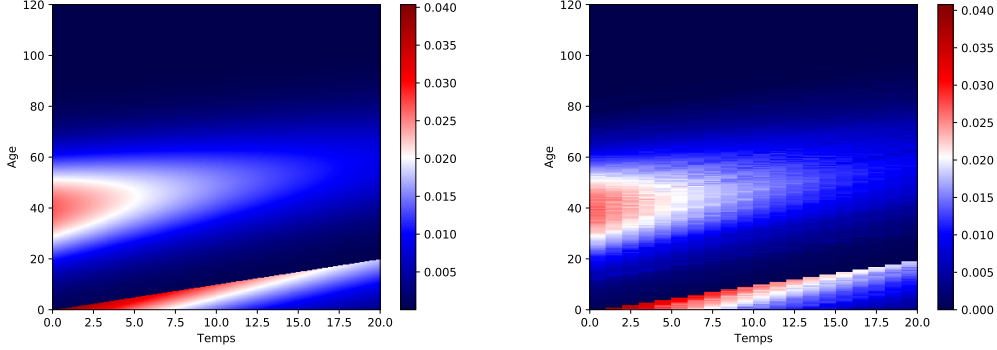


FIGURE 3. Left: true population density  $g$ . Right:  $\widehat{g}_{\widehat{h}^N}^N$  with  $N = 4000$  over a single simulation of  $Z^N$ .  $X$ -axis,  $Y$ -axis: units in years.

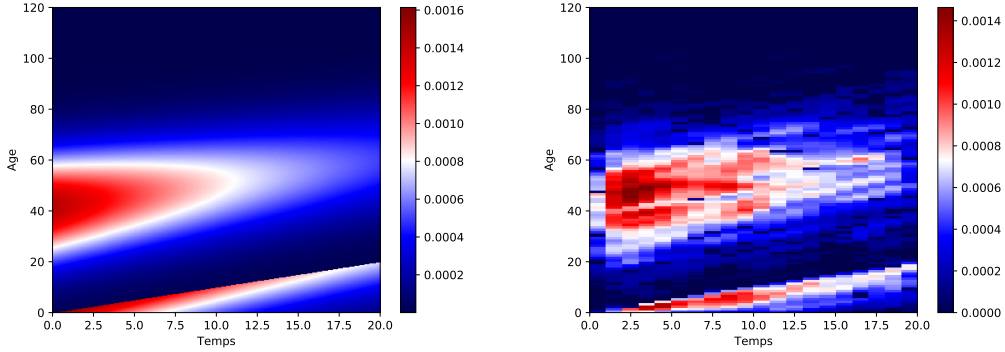


FIGURE 4. Left: true  $\pi = \mu \cdot g$ . Right: implementation of  $\widehat{\pi}_{\widehat{h}^N}^N$  with  $N = 4000$ .  $X$ -axis,  $Y$ -axis: units in years.

### 6.1. A first stability result.

**Proposition 21.** *Work under Assumptions 1 and 2. Then  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$  is bounded above by*

$$|w_1|_1 \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} \int_0^T |w_1(t)| (\mathcal{W}_{w_2, 1}^N(\mathcal{F})_t + \mathcal{W}_{1, w_2}^N(\mathcal{F})_t) dt + \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T,$$

where  $c_0$  is defined in Assumption 2.

*Proof.* By (7), the action  $\langle Z_t^N, f_t \rangle$  of  $Z_t^N(da)$  for  $f \in \mathcal{L}_{\mathcal{D}}^\infty$  can be written as

$$(35) \quad \begin{aligned} \langle Z_t^N, f_t \rangle &= \int_0^\infty f_t(t+a) Z_0^N(da) \\ &+ \int_0^t \int_0^\infty (b(s, a) f_t(t-s) - \mu(s, a) f_t(a+t-s)) Z_s^N(da) ds + M_t^N(f_t), \end{aligned}$$

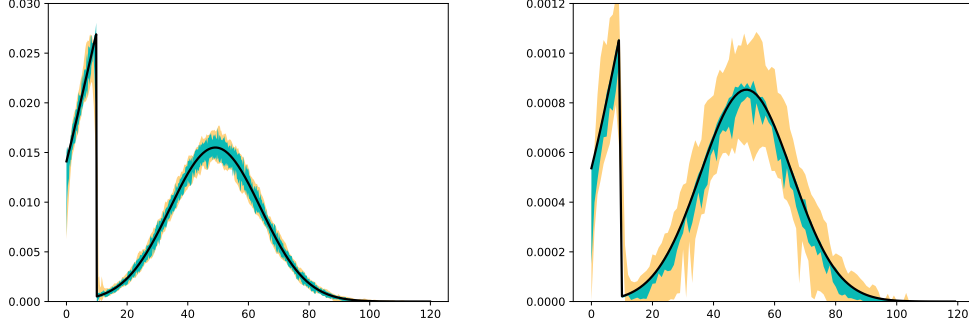


FIGURE 5. For fixed  $t = 10$ , and  $N = 4000$ , comparison between the true function (solid black) and a (pointwise) 95% confidence interval based on 50 Monte-Carlo simulations. Oracle estimator in green and our adaptative estimator in yellow. Left: Estimation of  $a \mapsto g(10, a)$ . Right: Estimation of  $a \mapsto \pi(10, a) = \mu(10, a)g(10, a)$ . X-axis: units in years, Y-axis: rate per unit of time.

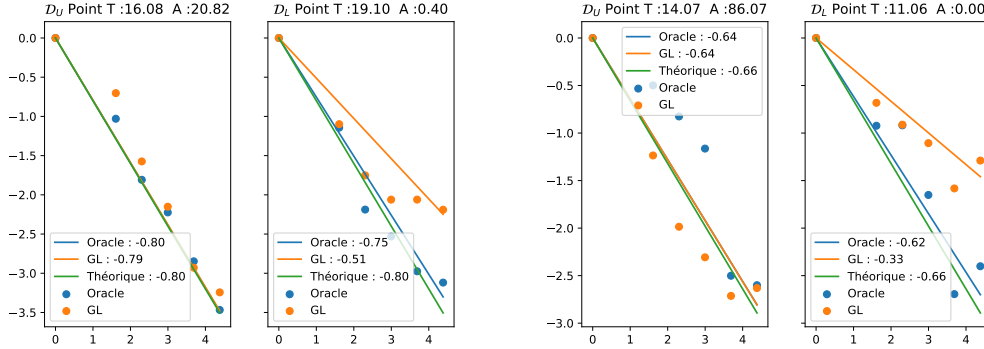


FIGURE 6.  $\log_{10}$ -integrated error of pointwise estimation against approximately  $\log_{10} N$  based on 50 Monte-carlo simulations. We compare the theoretical line given by the minimax theory (green) with a linear regression given by the oracle estimation (blue) and our estimation method (orange) for different values of  $\log_{10} N$  at two given points in  $\mathcal{D}_L$  and  $\mathcal{D}_U$  respectively. Left: estimation of  $g(t, a)$  at  $(t, a) = (16.08, 20.82)$  and  $(t, a) = (19.10, 0.40)$ . Right: estimation for  $\pi(t, a) = \mu(t, a)g(t, a)$  at  $(t, a) = (14.07, 86.07)$  and  $(t, a) = (11.06, 0.00)$ . X-axis: integers 0 to 5 correspond to  $N = 10^2, 5 \cdot 10^2, 10^3, 2 \cdot 10^3, 4 \cdot 10^3, 8 \cdot 10^3$ .

with

$$(36) \quad M_t^N(f_t) = N^{-1} \int_0^t \int_{\{i \leq n_s^N\} \times \mathbb{R}_+} (f_t(t-s) \mathbf{1}_{\{\theta \leq b\}} - f_t(a_i(Z_s^N) + t-s) \mathbf{1}_{\{b \leq \theta \leq b+\mu\}}) \tilde{\mathcal{Q}}(ds, di, d\theta).$$

In the above formula,  $n_t^N = N \langle Z_t^N, \mathbf{1} \rangle$  is the size of the population at time  $t$ , the functions  $b$  and  $\mu$  in the indicators are evaluated at points  $(s, a_i(Z_s^N))$  and  $\tilde{\mathcal{Q}}(ds, di, d\theta) = \mathcal{Q}(ds, di, d\theta) -$

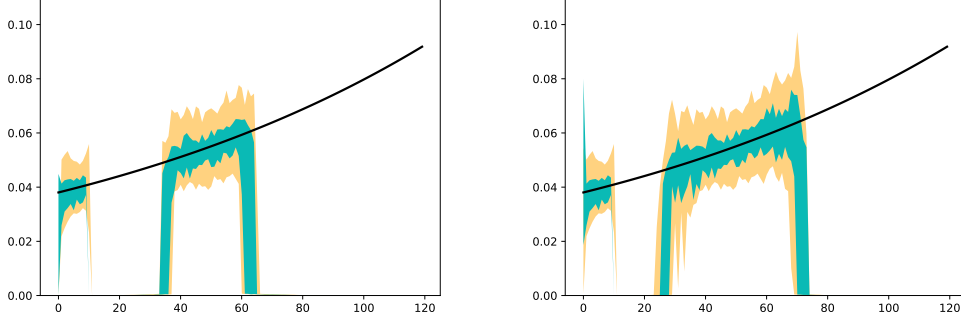


FIGURE 7. For fixed  $t = 10$ , and  $N = 4000$ , comparison between the true function (solid black) and a (pointwise) 95% confidence interval based on 50 Monte-Carlo simulations. Oracle estimator in green and our adaptative estimator in yellow. Estimation of  $a \mapsto \mu(10, a)$ . Left:  $\varpi = 10^{-2}$ . Right:  $\varpi = 5 \cdot 10^{-3}$ . X-axis: units in years, Y-axis: rate per unit of time. In order to improve on these results, one must either lower down  $\varpi$  or expect to be in a more favourable regime  $N \gg 4000$ .

$ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$  is the compensated measure of the Poisson measure  $\mathcal{Q}$ .

Apply now (35) to the test function  $a \mapsto w_2(t-a)f_t(a)$  with  $f \in \mathcal{F}$ , subtract  $g(t, a)da$  in the equation above, noting that  $g(t, a)$  solves (2), set  $\eta_t^N(da) = Z_t^N(da) - g(t, a)da$  and obtain

$$\begin{aligned} \int_{\mathbb{R}_+} w_2(t-a)f_t(a)\eta_t^N(da) &= \int_{\mathbb{R}_+} w_2(-a)f_t(t+a)\eta_0^N(da) \\ &+ \int_0^t \int_{\mathbb{R}_+} (w_2(s)f_t(t-s)b(s, a) - w_2(s-a)f_t(a+t-s)\mu(s, a))\eta_s^N(da)ds + M_t^N(w_2(t-\cdot)f_t). \end{aligned}$$

Multiplying each term by  $\omega_1(t)$ , integrating from 0 to  $T$  and taking absolute values, we also have

$$\left| \int_0^T \omega_1(t) \int_{\mathbb{R}_+} w_2(t-a)f_t(a)\eta_t^N(da)dt \right| \leq I + II + III + IV,$$

with

$$\begin{aligned} I &= \left| \int_0^T \omega_1(t) \int_{\mathbb{R}_+} w_2(-a)f_t(t+a)\eta_0^N(da)dt \right|, \\ II &= \left| \int_0^T \omega_1(t) \int_0^t \int_{\mathbb{R}_+} w_2(s)f_t(t-s)b(s, a)\eta_s^N(da)dsdt \right|, \\ III &= \left| \int_0^T \omega_1(t) \int_0^t \int_{\mathbb{R}_+} w_2(s-a)f_t(a+t-s)\mu(s, a)\eta_s^N(da)dsdt \right|, \\ IV &= \left| \int_0^T \omega_1(t)M_t^N(w_2(t-\cdot)f_t)dt \right|. \end{aligned}$$

By Assumption 2, we have  $f_t(t+a) \in \mathcal{F}$  therefore  $I \leq \mathcal{W}_{w_2}^N(\mathcal{F})_0$ . Using that  $c_0 f_t(t-s)b(s,a) \in \mathcal{F}$ , we also have

$$\begin{aligned} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s) f_t(t-s) b(s,a) \eta_s^N(da) ds \right| &\leq c_0^{-1} \sup_{f \in \mathcal{F}} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s) \mathbf{1}(s-a) f_t(a) \eta_s^N(da) ds \right| \\ &= c_0^{-1} \mathcal{W}_{w_2,1}(\mathcal{F})_t, \end{aligned}$$

Therefore  $II \leq c_0^{-1} \int_0^T |w_1(t)| \mathcal{W}_{w_2,1}(\mathcal{F})_t dt$ . In the same way,

$$\begin{aligned} \left| \int_0^t \int_{\mathbb{R}_+} w_2(s-a) f_t(a+t-s) \mu(s,a) \eta_s^N(da) ds \right| &\leq c_0^{-1} \sup_{f \in \mathcal{F}} \left| \int_0^t \int_{\mathbb{R}_+} \mathbf{1}(s) w_2(s-a) f_t(a) \eta_s^N(da) ds \right| \\ &= c_0^{-1} \mathcal{W}_{1,w_2}(\mathcal{F})_t \end{aligned}$$

and  $III \leq c_0^{-1} \int_0^T |w_1(t)| \mathcal{W}_{1,w_2}(\mathcal{F})_t dt$  follows likewise. Finally,

$$|IV| \leq \sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) M_t^N(w_2(t-\cdot) f_t) dt \right| = \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T.$$

Summing up the estimates, we obtain the conclusion noting that

$$\sup_{f \in \mathcal{F}} \left| \int_0^T w_1(t) \int_{\mathbb{R}_+} w_2(t-a) f_t(a) \eta_s^N(da) ds \right| = \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$$

since  $\mathcal{F}$  is stable under  $f \mapsto -f$  by Assumption 2.  $\square$

**Proposition 22.** *Work under Assumptions 1 and 2. We have*

$$\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \lesssim |w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0 + \max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T,$$

where  $(k_1, k_2)$  and  $(l_1, \dots, l_4)$  range over permutations of  $(1, w_2)$  and  $(1, 1, w_1, w_2)$  respectively. The symbol  $\lesssim$  means inequality up to an explicitly computable constant depending on  $T$  and  $c_0$  from Assumption 2.

*Proof.* Apply first Proposition 21 with  $w_1 = 1$  and  $w_2 = 1$  to obtain

$$\begin{aligned} \mathcal{W}_{1,1}^N(\mathcal{F})_T &\leq T \mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \int_0^T \mathcal{W}_{1,1}^N(\mathcal{F})_t dt + \mathcal{M}_{1,1}^N(\mathcal{F})_T \\ &\leq (T \mathcal{W}_1^N(\mathcal{F})_0 + \mathcal{M}_{1,1}^N(\mathcal{F})_T) e^{2c_0^{-1} T} \\ &= \mathcal{G}^{(1), N}(\mathcal{F})_T, \end{aligned}$$

say, by Grönwall lemma. Next, by Proposition 21 applied to  $(w_2, 1)$ , we obtain

$$\begin{aligned} \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T &\leq |w_2|_{L^1([0, T])} \mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \int_0^T |w_2(t)| \mathcal{W}_{1,1}^N(\mathcal{F})_t dt + \mathcal{M}_{w_2, 1}^N(\mathcal{F})_T \\ &\leq |w_2|_{L^1([0, T])} (\mathcal{W}_1^N(\mathcal{F})_0 + 2c_0^{-1} \mathcal{G}^{(1), N}(\mathcal{F})_T) + \mathcal{M}_{w_2, 1}^N(\mathcal{F})_T \\ &= \mathcal{G}_{w_2}^{(2), N}(\mathcal{F})_T, \end{aligned}$$

say. Apply now Proposition 21 with  $(1, w_2)$  so that

$$\begin{aligned} \mathcal{W}_{1,w_2}^N(\mathcal{F})_T &\leq T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} \int_0^T (\mathcal{W}_{w_2,1}^N(\mathcal{F})_t + \mathcal{W}_{1,w_2}^N(\mathcal{F})_t) dt + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T \\ &\leq T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} T\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \int_0^T \mathcal{W}_{1,w_2}^N(\mathcal{F})_t dt + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T \\ &\leq (T\mathcal{W}_{w_2}^N(\mathcal{F})_0 + CT\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \mathcal{M}_{1,w_2}^N(\mathcal{F})_T) e^{c_0^{-1}T} \\ &= \mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T \end{aligned}$$

say, by the previous estimate and Grönwall lemma again. By Proposition 21 and the two previous bounds, we infer that  $\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$  is less than

$$|w_1|_{L^1([0,T])} \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1} |w_1|_{L^1([0,T])} (\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T + \mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T) + \mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T.$$

Expanding the estimates  $\mathcal{G}_{w_2}^{(2),N}(\mathcal{F})_T$  and  $\mathcal{G}_{w_2}^{(3),N}(\mathcal{F})_T$  in terms of their appropriate arguments concludes the proof.  $\square$

By Proposition 22, we see that the stability of the system is controlled by the initial approximation  $\mathcal{W}_{w_2}^N(\mathcal{F})_0$  (including  $w_2 = 1$ ) and the propagation of the stochastic term  $\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$ . We now turn to that latter term.

**6.2. Stability of the stochastic term.** For  $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , let

$$\tilde{\mathcal{M}}_{w_1,w_2}^N(f)_t = \int_0^t w_1(s) M_s^N(w_2(s - \cdot) f) ds$$

and

$$\tilde{\mathcal{M}}_{w_2}^N(f)_t = M_t^N(w_2(t - \cdot) f).$$

In particular, since  $\mathcal{F}$  is stable under  $f \mapsto -f$ , we have

$$(37) \quad \sup_{f \in \mathcal{F}} \tilde{\mathcal{M}}_{w_1,w_2}^N(f)_T = \sup_{f \in \mathcal{F}} |\tilde{\mathcal{M}}_{w_1,w_2}^N(f)_T| = \mathcal{M}_{w_2}^N(\mathcal{F})_T$$

and

$$\sup_{f \in \mathcal{F}} \tilde{\mathcal{M}}_{w_2}^N(f)_T = \sup_{f \in \mathcal{F}} |\tilde{\mathcal{M}}_{w_2}^N(f)_T| = \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

For  $\kappa \geq 0$ , consider the event

$$(38) \quad \mathcal{A}_\kappa^N = \left\{ \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \exp(|b|_\infty T) (1 + \kappa) \right\},$$

and for  $\lambda \geq 0$ , set

$$\vartheta_{w_1,w_2}^N(f)_\lambda = 2NT |w_1|_\infty^{-1} \exp(|b|_\infty T) (|b|_\infty + |\mu|_\infty) \rho(N^{-1} \lambda |w_1 w_2|_\infty |f|_\infty),$$

where  $\rho(x) = e^x - x - 1$ .

**Proposition 23.** *Work under Assumptions 1. For large enough  $N$ , we have*

$$(39) \quad \int_0^\infty \mathbb{P}((\mathcal{A}_\kappa^N)^c) e^\kappa d\kappa \leq \frac{1}{2}$$

and for  $\lambda \geq 0$ ,

$$(40) \quad \mathbb{E}[\exp(\lambda |\tilde{\mathcal{M}}_{w_1,w_2}^N(f)_T - \tilde{\mathcal{M}}_{w_1,w_2}^N(g)_T|) \mathbf{1}_{\mathcal{A}_\kappa^N}] \leq 2 \exp(|w_1|_1 (1 + \kappa) \vartheta_{w_1,w_2}^N(f - g)_\lambda).$$

Moreover, (40) remains true with  $\tilde{\mathcal{M}}_{w_2}^N(f)_T - \tilde{\mathcal{M}}_{w_2}^N(g)_T$ , replacing formally  $w_1$  by 1 in the right-hand side of the inequality.

*Proof.* We first prove (39), namely

$$\int_0^\infty e^\kappa \mathbb{P}\left(\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle > \exp(|b|_\infty T)(1 + \kappa)\right) d\kappa \leq \frac{1}{2}.$$

**Step 1)** Consider the equation

$$\tilde{Z}_t^N = \tau_t Z_0^N + N^{-1} \int_0^t \int_{\mathbb{N} \times \mathbb{R}_+} \delta_{t-s}(da) \mathbf{1}_{\{0 \leq \vartheta \leq |b|_\infty, i \leq N \langle \tilde{Z}_{s-}^N, \mathbf{1} \rangle\}} \mathcal{Q}_1(ds, di, d\vartheta)$$

defined on the same probability space as  $(Z_t(da))_{0 \leq t \leq T}$ . Applying (35) with  $b = |b|_\infty$ ,  $\mu = 0$  and  $f_t = 1$ , we obtain

$$\langle \tilde{Z}_t^N, \mathbf{1} \rangle = \langle \tilde{Z}_0^N, \mathbf{1} \rangle + |b|_\infty \int_0^t \langle \tilde{Z}_s^N, \mathbf{1} \rangle ds + M_t^N(\mathbf{1}),$$

and for every  $\lambda \geq 0$ , by Itô's formula:

$$\exp(\lambda \langle \tilde{Z}_t^N, \mathbf{1} \rangle) = \exp(\lambda \langle \tilde{Z}_0^N, \mathbf{1} \rangle) + N|b|_\infty (e^{\lambda/N} - 1) \int_0^t \langle \tilde{Z}_s^N, \mathbf{1} \rangle \exp(\lambda \langle \tilde{Z}_s^N, \mathbf{1} \rangle) ds + \xi_t,$$

where  $(\xi_t)_{0 \leq t \leq T}$  is a local martingale. By localisation, one can prove that  $\mathbb{E}[\xi_t] = 0$ . Writing  $f(t, \lambda) = \mathbb{E}[\exp(\lambda \langle \tilde{Z}_t^N, \mathbf{1} \rangle)]$ , it follows that

$$(41) \quad f(t, \lambda) = f(0, \lambda) + N|b|_\infty (e^{\lambda/N} - 1) \int_0^t \partial_\lambda f(s, \lambda) ds.$$

The solution of the transport equation (41) at time  $t = T$  with initial condition  $f(0, \lambda) = f_0(\lambda)$  is given by

$$f(T, \lambda) = f_0\left(N \log \frac{e^{\lambda/N - |b|_\infty T}}{1 - (1 - e^{-|b|_\infty T})e^{\lambda/N}}\right) \leq \exp\left(qN \log \frac{e^{\lambda/N - |b|_\infty T}}{1 - (1 - e^{-|b|_\infty T})e^{\lambda/N}}\right),$$

where the last inequality stems from  $f_0(\lambda) = \mathbb{E}[\exp(\lambda \langle \tilde{Z}_0^N, \mathbf{1} \rangle)] = \mathbb{E}[\exp(\lambda \langle Z_0^N, \mathbf{1} \rangle)] \leq e^{q\lambda}$  for some  $q$  by Assumption 1 (ii).

**Step 2)** With the notation  $r = \exp(-|b|_\infty T)$ , the usual Chernoff bound argument yields

$$\begin{aligned} \log \mathbb{P}(\langle \tilde{Z}_T^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa)) &\leq -\lambda r^{-1}(1 + \kappa) + qN \log \frac{r e^{\lambda/N}}{1 - (1-r)e^{\lambda/N}} \\ &\leq -Nr^{-1}(1 + \kappa) \log\left(\left(1 - \frac{rq}{\kappa+1}\right) \frac{1}{1-r}\right) + qN \log \frac{\kappa+1-rq}{1-r} \\ &\leq \log C_1 - C_2 N \kappa \end{aligned}$$

for the choice  $\lambda = N \log\left(\left(1 - \frac{rq}{\kappa+1}\right) \frac{1}{1-r}\right)$  and for two constants  $C_i = C_i(q, r) > 0$  that do not depend on  $N$ . Noting that by construction,  $\sup_{t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \langle \tilde{Z}_T^N, \mathbf{1} \rangle$ , we finally obtain

$$\begin{aligned} \int_0^\infty e^\kappa \mathbb{P}\left(\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa)\right) d\kappa &\leq \int_0^\infty e^\kappa \mathbb{P}(\langle \tilde{Z}_T^N, \mathbf{1} \rangle > r^{-1}(1 + \kappa)) d\kappa \\ &\leq C_1 \int_0^\infty e^{(1-C_2N)\kappa} d\kappa = \frac{C_1}{C_2N - 1} \leq \frac{1}{2} \end{aligned}$$

for  $N \geq (1 + 2C_1)/C_2$ , and (39) is proved.

**Step 3)** We now turn to (40). For  $t_0 \in [0, T]$  and  $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , define

$$(42) \quad B_{t, t_0}^N(f) = N \int_0^{t \wedge t_0} \int_{\mathbb{R}_+} \left(b(s, a) \rho(N^{-1} f(t_0 - s)) + \mu(s, a) \rho(N^{-1} f(a + t_0 - s))\right) Z_s^N(da) ds.$$



**Lemma 24.** *For every  $t_0 \in [0, T]$  and  $f, g \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , there exists a nonnegative random variable  $\Lambda_{t_0, t_0}^N(f - g)$  with  $\mathbb{E}[\Lambda_{t_0, t_0}^N(f - g)] = 1$  such that*

$$\mathbb{E}[\exp(M_{t_0}^N(f) - M_{t_0}^N(g))] = \mathbb{E}[\Lambda_{t_0, t_0}^N(f - g) \exp B_{t_0, t_0}^N(f - g)].$$

*Proof.* Fix  $t_0 \in [0, T]$  and for  $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , define the random process

$$\widetilde{M}_{t_0, t_0}^N(f) = N^{-1} \int_0^{t \wedge t_0} \int_{\{i \leq n_s^N\} \times \mathbb{R}_+} (f(t_0 - s) \mathbf{1}_{\{b \leq \theta\}} - f(a_i(Z_{s^-}^N) + t_0 - s) \mathbf{1}_{\{b \leq \theta \leq b + \mu\}}) \widetilde{\mathcal{Q}}(ds, di, d\theta),$$

obtained by keeping  $t = t_0$  fixed in the integrand of  $M_{t \wedge t_0}^N(f)$  defined in (36). By construction,  $(\widetilde{M}_{t_0, t_0}^N(f))_{0 \leq t \leq T}$  is a martingale. In turn, a simple consequence of Itô's formula, see *e.g.* Tran [43] shows that the random process

$$t \mapsto \Lambda_{t, t_0}^N(f) = \exp(\widetilde{M}_{t, t_0}^N(f) - B_{t, t_0}^N(f))$$

is a martingale such that  $\mathbb{E}[\Lambda_{t, t_0}^N(f)] = 1$ . Noting that  $M_{t_0}^N(f) = \widetilde{M}_{t_0, t_0}^N(f)$  at  $t = t_0$ , we also have

$$\begin{aligned} \mathbb{E}[\exp(M_{t_0}^N(f) - M_{t_0}^N(g))] &= \mathbb{E}[\exp(M_{t_0}^N(f - g))] \\ &= \mathbb{E}[\exp(\widetilde{M}_{t_0, t_0}^N(f - g))] \\ &= \mathbb{E}[\Lambda_{t_0, t_0}^N(f - g) \exp B_{t_0, t_0}^N(f - g)]. \end{aligned}$$

□

Let  $\lambda \geq 0$ . We substitute  $f - g$  by  $a \mapsto \lambda w_1(t_0) w_2(t_0 - a)(f(a) - g(a))$  and look for an upper bound for

$$B_{t_0, t_0}^N(\lambda w_1(t_0) w_2(t_0 - \cdot)(f - g)).$$

**Step 4)** Observe first that  $\rho(x) = e^x - x - 1$  implies that for any nonnegative function  $\psi \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , we have

$$\begin{aligned} \rho(N^{-1} \lambda \psi(a')(f(a) - g(a))) &\leq N^{-1} \psi(a') |f - g|_{\infty} \int_0^{\lambda} (\exp(\kappa N^{-1} |\psi|_{\infty} |f - g|_{\infty}) - 1) d\kappa \\ &= \frac{\psi(a')}{|\psi|_{\infty}} \rho(N^{-1} \lambda |\psi|_{\infty} |f - g|_{\infty}). \end{aligned}$$

Therefore, with  $\psi(a') = w_1(t_0) w_2(t_0 - a')$  and  $a' = t_0 - s$ , we derive

$$\rho(N^{-1} \lambda w_1(t_0) w_2(s)(f(t_0 - s) - g(t_0 - s))) \leq \frac{w_1(t_0) w_2(s)}{|w_1 w_2|_{\infty}} \rho(N^{-1} \lambda |w_1 w_2|_{\infty} |f - g|_{\infty})$$

and

$$\rho(N^{-1} \lambda w_1(t_0) w_2(s - a)(f(a + t_0 - s) - g(a + t_0 - s))) \leq \frac{w_1(t_0) w_2(s - a)}{|w_1 w_2|_{\infty}} \rho(N^{-1} \lambda |w_1 w_2|_{\infty} |f - g|_{\infty})$$

with  $a' = a + t_0 - s$  follows likewise. Plugging these two estimates in the definition (42) of  $B_{t_0, t_0}^N$ , we infer on  $\mathcal{A}_\kappa^N = \{\sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \leq \exp(|b|_\infty T)(1 + \kappa)\}$  the chain of inequalities

$$\begin{aligned}
& B_{t_0, t_0}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \\
& \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) \int_0^{t_0} \int_{\mathbb{R}_+} (w_2(s) + w_2(s - a)) Z_s^N(da) ds \\
& \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) 2T \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle \\
& \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) \exp(|b|_\infty T)(1 + \kappa) 2T \\
& = w_1(t_0)(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda.
\end{aligned}$$

We derive

$$\begin{aligned}
(43) \quad & \exp(\lambda w_1(t_0) M_{t_0}^N(w_2(t_0 - \cdot)(f - g))) \mathbf{1}_{\mathcal{A}_\kappa^N} \\
& \leq \exp(w_1(t_0)(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda) \Lambda_{t_0, t_0}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \mathbf{1}_{\mathcal{A}_\kappa^N}.
\end{aligned}$$

**Step 5)** For every integer  $n \geq 1$  and  $\lambda \geq 0$ ,  $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , define

$$\Delta_{w_1, w_2}^{N, n}(f - g) \lambda = \exp\left(\lambda T n^{-1} \sum_{i=1}^n w_1(iT n^{-1}) M_{iT n^{-1}}^N(w_2(iT n^{-1} - \cdot)(f - g))\right).$$

Applying repeatedly (43) with  $t_0 = iT n^{-1}$  and integrating with respect to  $\mathbb{E}_{\mathcal{A}_\kappa^N}[\cdot] = \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \mathbb{E}[\cdot \mathbf{1}_{\mathcal{A}_\kappa^N}]$ , we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}_\kappa^N}[\Delta_{w_1, w_2}^{N, n}(f - g) \lambda] & \leq \exp\left(T n^{-1} \sum_{i=1}^n w_1(iT n^{-1})(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda\right) \times \\
& \quad \times \mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \prod_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda T w_1(t_0)w_2(t_0 - \cdot)(f - g))^{1/n} \right] \\
& \leq \exp\left(T n^{-1} \sum_{i=1}^n w_1(iT n^{-1})(1 + \kappa) \vartheta_{w_1, w_2}^N(f - g) \lambda\right) \mathbb{P}(\mathcal{A}_\kappa^N)^{-1},
\end{aligned}$$

where we used the fact that the geometric mean is controlled by the arithmetic mean:

$$\prod_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))^{1/n} \leq n^{-1} \sum_{i=1}^n \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))$$

and the fact that

$$\begin{aligned}
& \mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \right] \\
& \leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \mathbb{E} \left[ \Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g)) \right] = \mathbb{P}(\mathcal{A}_\kappa^N)^{-1}
\end{aligned}$$

since  $\Lambda_{iT n^{-1}, iT n^{-1}}^N(\lambda w_1(t_0)w_2(t_0 - \cdot)(f - g))$  has expectation 1 by Lemma 24. Using

$$\liminf_{n \rightarrow \infty} \Delta_{w_1, w_2}^{N, n}(f - g) \lambda = \exp\left(\lambda \int_0^T w_1(s) M_s^N(w_2(s - \cdot)(f - g)) ds\right)$$

by convergence of Riemann sums, letting  $n \rightarrow \infty$ , we obtain by Fatou lemma

$$(44) \quad \mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \exp \left( \lambda \int_0^T w_1(s) M_s^N(w_2(s-\cdot))(f-g) ds \right) \right] \leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(|w_1|_1(1+\kappa)\vartheta_{w_1, w_2}^N(f-g)\lambda).$$

Noting that Lemma 24 also holds for  $-M_s^N(f)$  and applying (44) to  $-M_s^N(w_2(s-\cdot))(f-g)$ , we infer

$$\mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \exp \left( \lambda \left| \int_0^T w_1(s) M_s^N(w_2(s-\cdot))(f-g) ds \right| \right) \right] \leq 2\mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(|w_1|_1(1+\kappa)\vartheta_{w_1, w_2}^N(f-g)\lambda),$$

but since  $\int_0^T w_1(s) M_s^N(w_2(s-\cdot))(f-g) ds = \tilde{\mathcal{M}}_{w_1, w_2}^N(f)_T - \tilde{\mathcal{M}}_{w_1, w_2}^N(g)_T$ , the estimate (40) is established.

**Step 6** It remains to prove (40) for  $\mathcal{M}_{w_2}^N(f-g)_T$ . We first integrate (43) for  $w_1 = 1$  at  $t_0 = T$  so that  $|w_1|_1 = T$  and proceed exactly as in Step 5) to obtain

$$\mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \exp \left( \lambda M_T^N(w_2(T-\cdot))(f-g) \right) \right] \leq \mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(T(1+\kappa)\vartheta_{1, w_2}^N(f-g)\lambda).$$

Applying the same argument for  $-M_T^N$ , we also have

$$\mathbb{E}_{\mathcal{A}_\kappa^N} \left[ \exp \left( \lambda \left| M_T^N(w_2(T-\cdot))(f-g) \right| \right) \right] \leq 2\mathbb{P}(\mathcal{A}_\kappa^N)^{-1} \exp(T(1+\kappa)\vartheta_{1, w_2}^N(f-g)\lambda)$$

which is the desired result.  $\square$

Proposition 23 is the main ingredient to obtain a concentration inequality for the processes  $(\tilde{\mathcal{M}}_{w_2}^N(f)_T)_{f \in \mathcal{F}}$  and  $(\tilde{\mathcal{M}}_{w_1, w_2}^N(f)_T)_{f \in \mathcal{F}}$ , and in turn, a deviation bound for  $\mathcal{M}_{w_1, w_2}(\mathcal{F})_T$  and  $\mathcal{M}_{w_2}(\mathcal{F})_T$  thanks to (37). The proof is given in Section 6.4 below.

More precisely, consider the apparently more general situation where we have a real-valued process  $\xi(f)_{f \in \mathcal{F}}$  indexed by some metric set  $(\mathcal{F}, d)$  and a family of events  $\mathcal{A}(\kappa)_{\kappa > 0}$  satisfying the following properties:

$$(45) \quad \int_0^\infty \mathbb{P}(\mathcal{A}(\kappa)^c) e^\kappa d\kappa \leq \frac{1}{2},$$

and

$$(46) \quad \mathbb{E} \left[ \exp \left( \lambda |\xi(f) - \xi(g)| \right) \mathbf{1}_{\mathcal{A}(\kappa)} \right] \leq 2 \exp \left( c_1(1+\kappa)\rho(c_2 d(f, g)\lambda) \right),$$

for every  $\lambda \geq 0$  and some  $c_1, c_2 > 0$ .

**Proposition 25.** *Assume that  $\xi(f)_{f \in \mathcal{F}}$  and  $\mathcal{A}(\kappa)_{\kappa > 0}$  satisfy (45) and (46) and that  $\xi(f_0) = 0$  for some  $f_0 \in \mathcal{F}$ . Then there exists a choice  $\varpi = \varpi(c_1, c_2) > 0$  such that for every  $u \geq 0$ :*

$$\mathbb{P} \left( \sup_{f \in \mathcal{F}} |\xi(f)| \geq 8 \left( u + \int_0^{\text{diam}_{\tilde{d}}(\mathcal{F})} \log(1 + \mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)) d\epsilon \right) \right) \leq (e^{u/\text{diam}_{\tilde{d}}(\mathcal{F})} - 1)^{-1},$$

where  $\tilde{d} = \varpi d$ ,  $\text{diam}_{\tilde{d}}(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} \tilde{d}(f, g)$  and  $\mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)$  is the minimal number of balls of  $\tilde{d}$ -size  $\epsilon > 0$  that are necessary to cover  $\mathcal{F}$ .

**Remark 26.** *We show in Remark 36 at the end of the proof of Proposition 25 that if  $c_1 \geq 308$ , we may pick  $\varpi(c_1, c_2) = k\sqrt{c_1}c_2$ , with  $k = 2\sqrt{77}$ .*

The proof of Proposition 25 relies on standard concentration techniques and goes back to Dudley [14]. We use the classical textbook of Ledoux-Talagrand [28] and detail the computations in the Appendix section 8. Combining Proposition 23 and 25, we obtain the following

**Theorem 27.** *Work under Assumptions 1. Assume  $\text{diam}_{|\cdot|_\infty} \mathcal{F} \leq 1$  and*

$$e(\mathcal{F}) = \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon < \infty.$$

*For large enough  $N$ , there exists an explicit choice of  $C = C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) > 0$ , given in the proof below, such that for every  $u \geq 0$ :*

$$(47) \quad \mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq (1+u)CN^{-1/2}|w_1|_{1,\infty}|w_2|_\infty) \leq (e^{ue(\mathcal{F})} - 1)^{-1}$$

and

$$(48) \quad \mathbb{P}(\mathcal{M}_{w_2}^N(\mathcal{F})_T \geq (1+u)CN^{-1/2}|w_2|_\infty) \leq (e^{ue(\mathcal{F})} - 1)^{-1}.$$

*Proof.* We plan to apply Proposition 25 with  $\xi(f) = \tilde{\mathcal{M}}_{w_1, w_2}^N(f)_T$ , having  $\xi(f_0) = 0$  for  $f_0 = 0$ . We take  $\mathcal{A}(\kappa) = \mathcal{A}_\kappa^N$  defined in (38) and notice that (45) is satisfied by (39). Also, we have (46) by (40) with

$$(49) \quad c_1 = 2NT|w_1|_1|w_1|_\infty^{-1} \exp(|b|_\infty T)(|b|_\infty + |\mu|_\infty) \quad \text{and} \quad c_2 = |w_1 w_2|_\infty N^{-1},$$

for the metric  $d(f, g) = |f - g|_\infty$ . Setting  $\tilde{d} = \varpi d$  with  $\varpi$  taken from Proposition 25, we have  $\text{diam}_{\tilde{d}}(\mathcal{F}) = \varpi \text{diam}_d(\mathcal{F}) \leq \varpi$  by assumption and also  $\mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon) \leq \mathcal{N}(\mathcal{F}, d, \epsilon/\varpi)$ . It follows that

$$\int_0^{\text{diam}_{\tilde{d}}(\mathcal{F})} \log(1 + \mathcal{N}(\mathcal{F}, \tilde{d}, \epsilon)) d\epsilon \leq \int_0^\varpi \log(1 + \mathcal{N}(\mathcal{F}, d, \epsilon/\varpi)) d\epsilon = \varpi e(\mathcal{F}),$$

which is finite by assumption. Since  $\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T = \sup_{f \in \mathcal{F}} |\xi(f)| = \sup_{f \in \mathcal{F}} \xi(f)$ , remember (37), we may apply Proposition 25 and obtain, for every  $u \geq 0$ ,

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8(\varpi e(\mathcal{F}) + u)) \leq (e^{u/\varpi} - 1)^{-1},$$

or equivalently

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8\varpi e(\mathcal{F})(1+u)) \leq (e^{ue(\mathcal{F})} - 1)^{-1}.$$

By Remark 26 (see also Remark 36), we pick  $\varpi = \varpi(c_1, c_2) = k\sqrt{c_1}c_2$  with  $k = 2\sqrt{77}$ , assuming  $c_1 \geq 308$  which is satisfied for sufficiently large  $N$  by (49). Using (49) again, it follows that

$$\begin{aligned} 8k\sqrt{c_1}c_2e(\mathcal{F})(1+u) &= (1+u)8ke(\mathcal{F})e^{\frac{1}{2}|b|_\infty T} \sqrt{2T}(|b|_\infty + |\mu|_\infty)^{1/2} N^{-1/2} (|w_1|_1|w_1|_\infty)^{1/2} |w_2|_\infty \\ &= (1+u)CN^{-1/2}|w_1|_{1,\infty}|w_2|_\infty, \end{aligned}$$

say, with

$$(50) \quad C = C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) = 8ke(\mathcal{F})e^{\frac{1}{2}|b|_\infty T} \sqrt{2T}(|b|_\infty + |\mu|_\infty)^{1/2}$$

and (47) follows. The proof of (48) is obtained in the same way and is omitted.  $\square$

**Remark 28.** (i) *Up to inflating the constant  $C$  by a multiplicative factor  $\max(1, e(\mathcal{F}))$ , we see that Theorem 27 implies a mild concentration property for*

$$(|w_1|_{1,\infty}|w_2|_\infty)^{-1} \mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \quad \text{and} \quad |w_2|_\infty^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$$

*with rate  $CN^{-1/2}$ .*

(ii) *The initial bound  $|w_2|_{1,\infty}$  of Assumption 5 inflates to  $|w_2|_\infty$  in (48). This defect actually has dramatic consequences when applied to subsequent statistical estimation:  $w_2$  becomes a kernel depending on  $N$  that mimicks a Dirac mass which is not stable for the  $|\cdot|_\infty$  metric. Improving on this estimates is actually the key difficulty in the proof of Theorem 6.*

**6.3. Proof of Theorem 6.** The weakness of Theorem 27 lies in the use of Proposition 23, where the control (40) somehow needs to be improved. This improvement however uses the results of Theorem 27 that we are going to iterate.

**Step 1)** By Proposition 22, we have  $\mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T \lesssim I + II$ , with

$$I = |w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0$$

and

$$II = \max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T.$$

Since  $|w_1|_1 \max_{(k_1, k_2)} |k_1|_{L^1([0, T])} \mathcal{W}_{k_2}^N(\mathcal{F})_0 \lesssim \max_{k=1, w_2} \mathcal{W}_k^N(\mathcal{F})_0$  up to a constant that only depends on  $T$ ,  $|w_1|$  and  $|w_2|_{L^1([0, T])}$ , we have by Assumption 5 that  $(|w_1|_{1, \infty} |w_2|_\infty)^{-1} I$  has a mild concentration property (actually, we can even replace  $|w_2|_\infty$  by  $|w_2|_{1, \infty}$ ). Next, by Theorem 27, the mild concentration property also holds for

$$(|w_1|_{1, \infty} |w_2|_\infty)^{-1} \mathcal{W}_{k, l}^N(\mathcal{F})_T, \quad \text{with } (k, l) \in \{(1, 1), (w_2, 1), (1, w_2)\}$$

up to an appropriate change in the constants, and therefore it carries over to  $(|w_1|_{1, \infty} |w_2|_\infty)^{-1} II$  since  $\max_{(l_1, \dots, l_4)} |l_1|_{L^1([0, T])} |l_2|_{L^1([0, T])} \mathcal{M}_{l_3, l_4}^N(\mathcal{F})_T \lesssim \sum_{(k, l)} \mathcal{W}_{k, l}^N(\mathcal{F})_T$  where the summation holds over  $\{(1, 1), (w_2, 1), (1, w_2)\}$ . In turn,

$$(|w_1|_{1, \infty} |w_2|_\infty)^{-1} \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T$$

has a mild concentration property of order  $C' \max(r_N, N^{-1/2})$ , for some  $C' > 0$  that depends on  $c_0$  of Assumption 2,  $T$ ,  $|w_1|_1$ ,  $|w_2|_{L^1([0, T])}$  and the constant  $C(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty)$  of Theorem 27 defined in (50).

**Step 2)** We next carefully revisit Step 4) of the proof of Proposition 23. We have

$$\begin{aligned} & B_{t_0, t_0}^N(\lambda w_1(t_0) w_2(t_0 - \cdot)(f - g)) \\ & \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) \int_0^T \int_{\mathbb{R}_+} (w_2(s) + w_2(s - a)) Z_s^N(da) ds. \end{aligned}$$

Adding and subtracting the limit  $g(t, a) da$ , we also have

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}_+} (w_2(s) + w_2(s - a)) Z_s^N(da) ds \\ & \leq \int_0^T \int_0^\infty (w_2(s) + w_2(s - a)) g(s, a) da ds + \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T + \mathcal{W}_{1, w_2}^N(\mathcal{F})_T \\ & \leq |w_2|_g + \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T + \mathcal{W}_{1, w_2}^N(\mathcal{F})_T, \end{aligned}$$

where, for  $f \in \mathcal{F}_b^{\text{age}}$  we set

$$(51) \quad |f|_g = |f|_{L^1([0, T])} \sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da + T |f|_1 |g|_\infty \wedge |f|_\infty |g|_1.$$

This bound is tighter than the estimate  $2|w_2|_\infty T \sup_{0 \leq t \leq T} \langle Z_t^N, \mathbf{1} \rangle$  that we used in Step 4) of the proof of Proposition 23. Introduce now the family of events

$$\mathcal{B}_\kappa^N = \left\{ \mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \leq 5C' \max(r_N, N^{-1/2}) |w_2|_{1, \infty} (1 + \kappa) \right\}, \quad \kappa > 0,$$

and

$$\mathcal{C}_\kappa^N = \left\{ \mathcal{W}_{1,w_2}^N(\mathcal{F})_T \leq 5C' \max(r_N, N^{-1/2}) \sqrt{T} |w_2|_\infty (1 + \kappa) \right\}, \quad \kappa > 0,$$

where  $C'$  is the constant of Step 1). On  $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$ , we now have

$$\begin{aligned} & B_{t_0, t_0}^N (\lambda w_1(t_0) w_2(t_0 - \cdot) (f - g)) \\ & \leq N(|b|_\infty + |\mu|_\infty) \frac{w_1(t_0)}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty) (|w_2|_g + C_{w_2}^N) (1 + \kappa) \\ & = w_1(t_0) (1 + \kappa) \tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda, \end{aligned}$$

say, with

$$(52) \quad C_{w_2}^N = 5C' \max(r_N, N^{-1/2}) (|w_2|_{1, \infty} + |w_2|_\infty \sqrt{T})$$

and

$$\tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda = N(|b|_\infty + |\mu|_\infty) \frac{|w_2|_g + C_{w_2}^N}{|w_1 w_2|_\infty} \rho(N^{-1} \lambda |w_1 w_2|_\infty |f - g|_\infty).$$

We thus have established that (40) of Proposition 23 holds with  $\tilde{\vartheta}_{w_1, w_2}^N (f - g)_\lambda$  instead of  $\vartheta_{w_1, w_2}^N (f - g)_\lambda$  and  $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$  instead of  $\mathcal{A}_\kappa^N$ .

**Step 3)** We now prove an analogous bound as (39) replacing  $\mathcal{A}_\kappa^N$  by  $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$ . Applying Theorem 27 with  $(w_1, w_2) = (w_2, 1)$  up to an inflation of  $C$  by  $\max(e(\mathcal{F}), 1)$  with the substitution  $1 + u = 5(1 + \kappa)$ , we obtain

$$\begin{aligned} \mathbb{P}((\mathcal{B}_\kappa^N)^c) &= \mathbb{P}(\mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \geq 5C' \max\{r_N, N^{-1/2}\} |w_2|_{1, \infty} (1 + \kappa)) \\ &= \mathbb{P}(\mathcal{W}_{w_2, 1}^N(\mathcal{F})_T \geq (1 + u)C' \max(r_N, N^{-1/2}) |w_2|_{1, \infty}) \\ &\leq (\exp(u) - 1)^{-1} = (\exp(4 + \kappa + 4\kappa) - 1)^{-1} \leq e^{-5\kappa}. \end{aligned}$$

It follows that

$$\int_0^\infty \mathbb{P}((\mathcal{B}_\kappa^N)^c) e^\kappa d\kappa \leq \int_0^\infty e^{-4\kappa} d\kappa = \frac{1}{4}.$$

In the same way, applying Theorem 27 with  $(w_1, w_2) = (1, w_2)$  and up to an inflating the constant  $C$  again, we obtain

$$\mathbb{P}((\mathcal{C}_\kappa^N)^c) = \mathbb{P}(\mathcal{W}_{1, w_2}^N(\mathcal{F})_T \geq 5C' \max(r_N, N^{-1/2}) |w_2|_\infty \sqrt{T} (1 + \kappa)) \leq e^{-5\kappa}$$

Hence  $\int_0^\infty \mathbb{P}((\mathcal{C}_\kappa^N)^c) e^\kappa d\kappa \leq \frac{1}{4}$  follows likewise and (39) is proved with  $\mathcal{B}_\kappa^N \cap \mathcal{C}_\kappa^N$  in place of  $\mathcal{A}_\kappa^N$ .

**Step 4)** We may now reproduce the proof of Theorem 27 with our new estimates from Step 2) : the estimate (49) now becomes

$$c'_1 = N(|b|_\infty + |\mu|_\infty) \frac{|w_1|_1}{|w_1 w_2|_\infty} (|w_2|_g + C_{w_2}^N) \quad \text{and} \quad c'_2 = c_2 = N^{-1} |w_1 w_2|_\infty,$$

and thanks to Step 3), we may apply in this new setting Proposition 25 to obtain

$$\mathbb{P}(\mathcal{M}_{w_1, w_2}^N(\mathcal{F})_T \geq 8\varpi(c'_1, c'_2) e(\mathcal{F})(1 + u)) \leq (e^{ue(\mathcal{F})} - 1)^{-1}.$$

Again, we may pick  $\varpi = \varpi(c'_1, c'_2) = k\sqrt{c'_1}c'_2$  with  $k = 2\sqrt{77}$ , assuming  $c_1 \geq 308$  which is true for  $N$  is large enough, and it follows that

$$\begin{aligned} & 8k\sqrt{c'_1}c'_2e(\mathcal{F}) \\ &= 8k e(\mathcal{F})(|b|_\infty + |\mu|_\infty)^{1/2}N^{-1/2}|w_1|_{1,\infty}|w_2|_\infty^{1/2}(|w_2|_g + C_{w_2}^N)^{1/2} \\ &\leq C''N^{-1/2}|w_1|_{1,\infty}|w_2|_\infty^{1/2}(|w_2|_g + C_{w_2}^N)^{1/2} \end{aligned}$$

say, with

$$C'' = C''(e(\mathcal{F}), T, |b|_\infty, |\mu|_\infty) = 8k \max(1, e(\mathcal{F}))(|b|_\infty + |\mu|_\infty)^{1/2} \max(5C'\sqrt{T}, 1)^{1/2}.$$

For  $f \in \mathcal{L}_{\mathcal{D}}^{\text{age}}$ , define now

$$[f]_{1,\infty}^{\varepsilon_N} = |f|_\infty^{1/2}(|f|_g + \varepsilon_N(|f|_{1,\infty} + |f|_\infty))^{1/2}.$$

We have proved that for  $\varepsilon_N = \max(r_N, N^{-1/2})$ , the sequence

$$(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$$

has a mild concentration property with rate  $C''N^{-1/2}$ . Applying the same argument as for Step 1) above, the mild concentration property carries over to

$$(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$$

with rate  $C'' \max(r_N, N^{-1/2})$ , possibly up to inflating the constant  $C'' > 0$ .

**Step 5)** We finally show that  $[w_2]_{1,\infty}^{\varepsilon_N} \lesssim |w|_{1,\infty}$  up to a constant that only depends on  $|b|_\infty, |\mu|_\infty, |g_0|_\infty$  and  $T$ , under the additional assumption that  $w_2$  has compact support and  $|w_2|_\infty \lesssim \varepsilon_N^{-1}|w_2|_1$ . By definition of  $|w_2|_g$  in (51), we have

$$|w_2|_g \lesssim |w_2|_1 \left( \sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da + |g|_\infty \right) \lesssim |w_2|_1$$

by the estimates of Lemma 38 in Appendix 8.4. Moreover, the compact support of  $w_2$  implies  $|w_2|_{1,\infty} \leq |w_2|_\infty |\text{supp}(w_2)|^{1/2} \lesssim |w_2|_\infty$ . It follows that

$$[w_2]_{1,\infty}^{\varepsilon_N} \lesssim |w_2|^{1/2}(|w_2|_1 + \varepsilon_N|w_2|_\infty)^{1/2} \lesssim |w_2|_{1,\infty}.$$

Let us note that the constant may possibly depend on  $|\text{supp}(w_2)|$  which is bounded above by  $\mathbf{u}$  by assumption.

**Step 6)** It remains to prove a mild concentration property for  $([w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{W}_{w_2}^N(\mathcal{F})_T$  with rate  $C'' \max(r_N, N^{-1/2})$ . The property holds for

$$([w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_2}^N(\mathcal{F})_T$$

with the same proof as for  $(|w_1|_{1,\infty}[w_2]_{1,\infty}^{\varepsilon_N})^{-1}\mathcal{M}_{w_1,w_2}^N(\mathcal{F})_T$ . We omit the details. Next, reproducing the beginning of the proof of Proposition 21 and applying (35) to the test function  $a \mapsto w_2(t - a)f_t(a)$  with  $f \in \mathcal{F}$ , we obtain

$$\mathcal{W}_{w_2}^N(\mathcal{F})_T \leq \mathcal{W}_{w_2}^N(\mathcal{F})_0 + c_0^{-1}(\mathcal{W}_{w_2,1}^N(\mathcal{F})_T + \mathcal{W}_{1,w_2}^N(\mathcal{F})_T) + \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

By Proposition 22, we further have

$$\mathcal{W}_{w_2,1}^N(\mathcal{F})_T \lesssim \mathcal{W}_1^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T$$

and

$$\mathcal{W}_{1,w_2}^N(\mathcal{F})_T \lesssim \max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T,$$

up to a constant that only depends on  $T$ ,  $c_0$ ,  $|w_1|_1$  and  $|w_2|_{L^1([0,T])}$ , therefore  $\mathcal{W}_{w_2}^N(\mathcal{F})_T$  is of order

$$\max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T + \mathcal{M}_{w_2}^N(\mathcal{F})_T.$$

The mild concentration property of  $(|w_2|_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$  and  $(|w_1|_{1,\infty} [w_2]_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$  enables us to control the last two terms. The first term has the correct order by Assumption 5. The proof of Theorem 6 is complete.

#### 6.4. Remaining proofs of Section 2.

*Proof of Proposition 8.* We repeat the argument of Step 6) in the proof of Theorem 6 above. By Proposition 22, we have

$$\mathcal{W}_{w_2}^N(\mathcal{F})_T \lesssim \max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0 + \max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T$$

and thus

$$\mathbb{E}[\mathcal{W}_{w_2}^N(\mathcal{F})_T^p] \lesssim \mathbb{E}[\max_{k=1,w_2} \mathcal{W}_k^N(\mathcal{F})_0^p] + \mathbb{E}[\max_{h,k=1,w_2} \mathcal{M}_{h,k}^N(\mathcal{F})_T^p] + \mathbb{E}[\mathcal{M}_{w_2}^N(\mathcal{F})_T^p],$$

up to a constant that depends on  $p$ ,  $T$ ,  $c_0$ ,  $|w_1|_1$  and  $|w_2|_{L^1([0,T])}$ . The first term is of order  $|w_2|_{1,\infty}^p r_N$  by Assumption. For the two other terms we use the identity  $\mathbb{E}[Z^p] = p \int_0^\infty x^{p-1} \mathbb{P}(Z \geq x) dx$  for a nonnegative random variable  $Z$  and conclude with the mild concentration property of  $(|w_2|_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{M}_{w_2}^N(\mathcal{F})_T$  and  $(|w_1|_{1,\infty} [w_2]_{1,\infty}^{\varepsilon_N})^{-1} \mathcal{W}_{w_1,w_2}^N(\mathcal{F})_T$ .

*Proof of Proposition 3.* Let  $\mathcal{F}_0$  denote the minimal set that contains  $0$ ,  $c_0$ ,  $c_0\mu$ ,  $c_0b$  and that is stable under the operations defined in (8) except for the pointwise product  $(f, g) \mapsto f \cdot g$ . We also set, for  $f \in \mathcal{L}_{\mathcal{D}}^\infty$ :

$$\mathcal{A}(f)_{t_1,t_2}^{(k,l)} = ((s, a) \mapsto f(t_1, t_2 + ka - ls))$$

with  $t_1, t_2 \in [0, T]$  and  $k, l = 0, 1$ .

**Step 1)** We claim that

$$(53) \quad \mathcal{F}_0 \subseteq \left\{ 0, \pm c_0, \pm c_0\mu, \pm c_0b, \pm \mathcal{L}(c_0b)_{t_1,t_2}^{(k,l)}, \pm \mathcal{A}(c_0\mu)_{t_1,t_2}^{(k,l)}, \text{ for every } t_1, t_2 \in [0, T], k, l = 0, 1 \right\}.$$

Indeed, one can check the following stability properties:

$$\mathfrak{s}_t(\mathcal{A}(f)_{t_1,t_2}^{(k,l)})(s, a) = \mathcal{A}_{t_1,t_2}^{(k,l)}(t, t+a) = f(t_1, t_2 + kt + ka - lt) = \mathcal{A}_{t_1,t_2+kt-lt}^{(k,0)}(s, a),$$

$$\mathfrak{t}_t(\mathcal{A}(f)_{t_1,t_2}^{(k,l)})(s, a) = \mathcal{A}_{t_1,t_2}^{(k,l)}(t, t-s) = f(t_1, t_2 - lt + ka - ks) = \mathcal{A}_{t_1,t_2+kt-lt}^{(0,k)}(s, a),$$

$$\mathfrak{u}_t(\mathcal{A}(f)_{t_1,t_2}^{(k,l)})(s, a) = \mathcal{A}_{t_1,t_2}^{(k,l)}(t, t-s+a) = f(t_1, t_2 + kt - ks + ka - lt) = \mathcal{A}_{t_1,t_2+kt-lt}^{(k,k)}(s, a).$$

This proves (53).

**Step 2)** We now prove that if  $b, \mu \in \mathbb{C}^s$  for some  $0 < s \leq 1$  with Hölder constant  $L > 0$ , then

$$(54) \quad \mathcal{N}(\mathcal{F}_0, |\cdot|_\infty, \epsilon) \lesssim \epsilon^{-2/s},$$



up to a constant that only depends on  $s$ ,  $T$  and  $L$ . Indeed, if  $f \in \mathcal{C}^s$  with Hölder constant  $L > 0$ , we have

$$\begin{aligned} |\mathcal{L}(f)_{t_1, t_2}^{(k, l)} - \mathcal{L}(f)_{t'_1, t'_2}^{(k, l)}|_\infty &= \sup_{s, a} |f(t_1, t_2 + ka - ls) - f(t_1, t_2 + ka - ls)| \\ &\leq L(|t_1 - t'_1|^s + |t_2 - t'_2|^s), \end{aligned}$$

therefore, for fixed  $(k, l)$  and  $f \in \mathcal{C}^s$ , the  $\epsilon$ -covering number of  $\{\mathcal{L}(f)_{t_1, t_2}^{(k, l)}, t_1, t_2 \in [0, T]\}$  in  $|\cdot|_\infty$  is the same as that of  $[0, T]^2$  equipped with the metric  $d((t_1, t_2) - (t'_1, t'_2)) = L(|t_1 - t'_1|^s + |t_2 - t'_2|^s)$ . Since  $\mathcal{N}([0, T], \epsilon, L|\cdot|^\gamma) = T\mathcal{N}([0, T], (\epsilon/L)^{1/\gamma}, |\cdot|) = TL^{1/\gamma}\epsilon^{-1/\gamma}$ , we have that  $\mathcal{N}([0, T]^2, \epsilon, d) \lesssim \epsilon^{-2/s}$  and (54) is established.

**Step 3)** We now consider the class  $\mathcal{F}_0^{\text{prod}}$  that contains  $\mathcal{F}_0$  and that is stable under the operation  $(f, g) \mapsto fg$ . Since  $\mathfrak{s}_t(fg) = \mathfrak{s}_t(f)\mathfrak{s}_t(g)$ ,  $\mathfrak{t}_t(fg) = \mathfrak{t}_t(f)\mathfrak{s}_t(g)$ ,  $\mathfrak{u}_t(fg) = \mathfrak{u}_t(f)\mathfrak{s}_t(g)$ , the class  $\mathcal{F}_0^{\text{prod}}$  contains the minimal class  $\mathcal{F}$ .

Let  $f = \prod_{\ell=1}^m f_\ell \in \mathcal{F}_0^{\text{prod}}$ , with  $f_\ell \in \mathcal{F}_0$ . For every  $\ell$ , we have  $|f_\ell|_\infty \leq c_1 < 1$ , with  $c_1 = c_0 \max(b|_\infty, |\mu|_\infty) < 1$  by assumption. Therefore, if  $m \geq \log \epsilon / \log c_1 = m(\epsilon)$ , we have  $|f|_\infty = |f - 0|_\infty \leq \epsilon$ . Now, let  $g_i$  be  $\mathcal{N}(\mathcal{F}_0, \epsilon m(\epsilon)^{-1}, |\cdot|_\infty)$  functions in  $\mathcal{F}_0$  such that, for every  $f \in \mathcal{F}_0$ , there exists an index  $i(f)$  such that  $|f - g_{i(f)}|_\infty \leq \epsilon m(\epsilon)^{-1}$ . If  $m \leq \log \epsilon / \log c_0$ , we have

$$\left| f - \prod_{\ell=1}^m g_{i(f_\ell)} \right|_\infty = \left| \prod_{\ell=1}^m f_\ell - \prod_{\ell=1}^m g_{i(f_\ell)} \right|_\infty \leq c_1^{m-1} m \epsilon m(\epsilon)^{-1} \leq \epsilon.$$

As a result, the family  $\{0, \prod_{\ell=1}^k g_\ell, k = 1, \dots, m(\epsilon)\}$  is a family of centers of balls of radius at most  $\epsilon$  that are sufficient to cover  $\mathcal{F}_0^{\text{prod}}$ . It follows that

$$\mathcal{N}(\mathcal{F}_0^{\text{prod}}, \epsilon, |\cdot|_\infty) \leq \mathcal{N}(\mathcal{F}_0, m(\epsilon)\epsilon, |\cdot|_\infty)^{m(\epsilon)+1} \lesssim (\epsilon m(\epsilon))^{-2m(\epsilon)/s}.$$

**Step 4)** We have established  $\mathcal{F} \subseteq \mathcal{F}_0^{\text{prod}}$  and therefore

$$\begin{aligned} e(\mathcal{F}) &= \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}, |\cdot|_\infty, \epsilon)) d\epsilon \leq \int_0^1 \log(1 + \mathcal{N}(\mathcal{F}_0^{\text{prod}}, \epsilon, |\cdot|_\infty)) d\epsilon \\ &\lesssim \int_0^1 \log(\epsilon m(\epsilon))^{-2m(\epsilon)/s} d\epsilon \lesssim \int_0^1 (\log \epsilon)^2 d\epsilon < \infty. \end{aligned}$$

The proof of Proposition 3 is complete.

## 7. PROOFS OF SECTION 3 AND 4

**7.1. Proof of Theorem 11.** Remember that the condition  $r_N \leq N^{-1/2}$  is in force in this section.

*Preliminaries.* We first write a standard bias-variance decomposition in squared-error loss, based upon the stability result of Corollary 8.

**Lemma 29.** *Let  $h \in \mathcal{G}_1^N$ . If  $\widehat{g}_h^N$  is specified with a bounded and compactly supported kernel  $K$ , we have*

$$\mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \lesssim \mathcal{B}_h^N(g)(t, a)^2 + \mathcal{V}_h^N,$$

where  $\mathcal{B}_h(g)(t, a)$  and  $\mathcal{V}_h^N$  are defined in (27) and (20) respectively.

*Proof.* Write  $\widehat{g}_h^N(t, a) - g(t, a) = I + II$ , with

$$I = \int_0^\infty K_h(u - a)g(t, u)du - g(t, a)$$

and

$$II = \int_{\mathbb{R}_+} K_h(u - a)(Z_t^N(du) - g(t, u))du.$$

We have  $I^2 \leq B_h(g)(t, a)^2$ . For the stochastic term, we have

$$(55) \quad |II| \leq \mathcal{W}_{K_h(t-a-\cdot)}^N(\mathcal{F})_t$$

Moreover

$$(56) \quad |K_h(t - a - \cdot)|_\infty \leq |K_h(t - a - \cdot)|_\infty = h^{-1}|K|_\infty \lesssim |K_h(t - a - \cdot)|_1 N^{1/2}$$

as soon as  $h^{-1} \lesssim N^{1/2}$  since  $|K_h(t - a - \cdot)|_1 = |K|_1 = 1$ . This condition is true for any  $h \in \mathcal{G}_1^N$  using the fact that  $K$  is bounded and compactly supported. We may then apply Corollary 8 and obtain

$$\mathbb{E}[II^2] \lesssim |K_h(t - a - \cdot)|_{1, \infty}^2 N^{-1} \lesssim (C^* N^{-1/2} |K_h|_{1, \infty})^2 = \mathbf{V}_h^N.$$

□

*Completion of proof of Theorem 11.* We essentially repeat the main argument of the Goldenshluger-Lepski method (see *e.g.* [17, 18] for the pointwise risk) in a setting that we need to adapt to our context.

**Step 1)** For any  $h \in \mathcal{G}_1^N$ , forcing  $\widehat{g}_h^N(t, a)$  in the risk decomposition and by definition of  $\mathbf{A}_h^N(t, a)$  and  $\widehat{h}^N(t, a)$ , we successively have

$$\begin{aligned} & \mathbb{E}[(\widehat{g}_*^N(t, a) - g(t, a))^2] \\ & \lesssim \mathbb{E}[(\widehat{g}_*^N(t, a) - \widehat{g}_h^N(t, a))^2] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\ & \lesssim \mathbb{E}[\{(\widehat{g}_{\widehat{h}^N(t, a)}^N(t, a) - \widehat{g}_h^N(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{\widehat{h}^N(t, a)}^N\}_+ + \mathbf{V}_h^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\ & \lesssim \mathbb{E}[\mathbf{A}_{\max(\widehat{h}^N(t, a), h)}^N(t, a) + \mathbf{V}_h^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\ & \lesssim \mathbb{E}[\mathbf{A}_h^N(t, a)] + \mathbf{V}_h^N + \mathbb{E}[\mathbf{A}_{\widehat{h}^N(t, a)}^N + \mathbf{V}_{\widehat{h}^N(t, a)}^N] + \mathbb{E}[(\widehat{g}_h^N(t, a) - g(t, a))^2] \\ & \lesssim \mathbb{E}[\mathbf{A}_h^N(t, a)] + \mathbf{V}_h^N + \mathcal{B}_h^N(g)(t, a)^2 \end{aligned}$$

where we applied Lemma 29 to obtain the last line.

**Step 2)** We first estimate  $\mathbf{A}_h^N(t, a)$ . Write  $g_h(t, a)$  for  $\int_{\mathbb{R}_+} K_h(u - a)g(t, u)du$ . For  $h, h' \in \mathcal{G}_1^N$  with  $h' \leq h$ , since

$$\begin{aligned} & (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}^N(t, a))^2 \\ & \leq 4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 + 4(g_h(t, a) - g(t, a))^2 + 4(g_{h'}(t, a) - g(t, a))^2 + 4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2, \end{aligned}$$

we have

$$\begin{aligned} & (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}^N(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N \\ & \leq 8\mathcal{B}_h^N(g)(t, a)^2 + (4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N) + (4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N). \end{aligned}$$

using  $h' \leq h$  in order to bound  $(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2$  by the bias at scale  $h$ . It follows that

$$\begin{aligned} & (\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N \\ & \leq 8\mathcal{B}_h^N(g)(t, a)^2 + 4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N + 4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N, \end{aligned}$$

and taking maximum over  $h' \leq h$ , we obtain

$$(57) \quad \max_{h' \leq h} \{(\widehat{g}_h^N(t, a) - \widehat{g}_{h'}(t, a))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N\}_+ \\ \leq 8\mathcal{B}_h^N(g)(t, a)^2 + \{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N\}_+ + \max_{h' \leq h} \{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N\}_+.$$

**Step 3)** We estimate the expectation of the first stochastic term in the right-hand side of (57). Since  $|\widehat{g}_h^N(t, a) - g_h(t, a)| \leq \mathcal{W}_{K_h(t-a-\cdot)}^N$ , we successively have

$$\begin{aligned} \mathbb{E}[\{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N\}_+] &= \int_0^\infty \mathbb{P}(4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N \geq \kappa) d\kappa \\ &= \int_0^\infty \mathbb{P}(|\widehat{g}_h^N(t, a) - g_h(t, a)| \geq \frac{1}{2}(\mathbf{V}_h^N + \kappa)^{1/2}) d\kappa \\ &\leq \int_0^\infty \mathbb{P}(\mathcal{W}_{K_h(t-a-\cdot)}^N \geq \frac{1}{2}(\mathbf{V}_h^N + \kappa)^{1/2}) d\kappa. \end{aligned}$$

We may apply Theorem 6 with  $w_2 = K_h(t-a-\cdot)$  since  $K$  is compactly supported and having (56) of Lemma 29 above. By the change of variable

$$\frac{1}{2}(\mathbf{V}_h^N + \kappa)^{1/2} = (1+u)C''|K_h|_{1,\infty}N^{-1/2},$$

we then obtain

$$\begin{aligned} & \mathbb{E}[\{4(\widehat{g}_h^N(t, a) - g_h(t, a))^2 - \mathbf{V}_h^N\}_+] \\ & \leq 8C''|K_h|_{1,\infty}N^{-1/2} \int_{\frac{1}{2C''}(\mathbf{V}_h^N)^{1/2}|K_h|_{1,\infty}^{-1}N^{1/2}-1}^\infty (1+u) \min((e^u-1)^{-1}, 1) du \\ & \lesssim \exp(-\frac{1}{2C''}(\mathbf{V}_h^N)^{1/2}|K_h|_{1,\infty}^{-1}N^{1/2}) \leq N^{-2} \end{aligned}$$

by definition of  $\mathbf{V}_h^N$ .

**Step 4)** For the second stochastic term, we use the rough estimate

$$\begin{aligned} \mathbb{E}[\max_{h' \leq h} \{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N\}_+] &\leq \sum_{h' \leq h} \mathbb{E}[\{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N\}_+] \\ &\lesssim \text{Card}(\mathcal{G}_1^N)N^{-2} \lesssim N^{-1} \end{aligned}$$

where we used Step 3) to bound each term  $\mathbb{E}[\{4(\widehat{g}_{h'}^N(t, a) - g_{h'}(t, a))^2 - \mathbf{V}_{h'}^N\}_+]$  independently of  $h$  together with  $\text{Card}(\mathcal{G}_1^N) \lesssim N$ . In conclusion, we have proved through Steps 2)-4) that  $\mathbb{E}[\mathbf{A}_h^N(t, a)] \lesssim \delta_N$ . Therefore, from Step 1), we conclude

$$\mathbb{E}[(\widehat{g}_*^N(t, a) - g(t, a))^2] \lesssim \mathcal{B}_h^N(g)(t, a)^2 + \mathbf{V}_h^N + \delta_N$$

for any  $h \in \mathcal{G}_1^N$ . The proof of Theorem 11 is complete.

## 7.2. Proof of Theorem 13.

*Preliminaries.* We first study the behaviour of the process  $\Gamma^N(dt, da)$  of death occurrences introduced in Section 3.2 and represented via (17).

**Lemma 30.** *With the notation of Section 2.2, we have*

$$(58) \quad \Gamma^N(dt, da) = N^{-1} \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} \delta_{a_i(Z_s^N)}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_s^N)), i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} \mathcal{Q}_2(dt, di, d\vartheta),$$

where  $\mathcal{Q}_2$  is a Poisson random measure on  $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$  with intensity  $dt(\sum_{k \geq 1} \delta_k(di))d\vartheta$ . Moreover, for nonnegative weights  $w_1 \in \mathcal{L}_D^{\text{time}}$  and  $w_2 \in \mathcal{L}_D^{\text{age}}$ , we have

$$(59) \quad \left| \int_0^T \int_{\mathbb{R}_+} w_1(s)w_2(s-u) (\Gamma^N(ds, du) - \mu(s, u)g(s, u)duds) \right| \leq \mathcal{W}_{w_1, w_2}^N(\mathcal{F})_T + |(\Delta_{w_1, w_2}^N)_T|,$$

where  $t \mapsto (\Delta_{w_1, w_2}^N)_t$  is a square integrable martingale with predictable compensator

$$(60) \quad \langle \Delta_{w_1, w_2}^N \rangle_t = N^{-1} \int_0^t \int_{\mathbb{R}_+} w_1(s)^2 w_2(s-u)^2 \mu(s, u) Z_s^N(du) ds.$$

*Proof.* The representation (58) is straightforward. We add and subtract in the left-hand side of (59) the term  $\int_0^T \int_0^\infty w_1(s)w_2(s-u)\mu(s, u)Z_s^N(du)ds$  and obtain the desired inequality with

$$\begin{aligned} (\Delta_{w_1, w_2}^N)_t &= \int_0^t \int_{\mathbb{R}_+} w_1(s)w_2(s-u) (\Gamma^N(ds, du) - \mu(s, u)Z_s^N(du)) \\ &= N^{-1} \int_0^t \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} w_1(s)w_2(s-u) \delta_{a_i(Z_s^N)}(da) \mathbf{1}_{\{0 \leq \vartheta \leq \mu(s, a_i(Z_s^N)), i \leq \langle NZ_s^N, \mathbf{1} \rangle\}} \tilde{\mathcal{Q}}_2(ds, di, d\vartheta), \end{aligned}$$

where  $\tilde{\mathcal{Q}}_2(ds, di, d\vartheta) = \mathcal{Q}_2(ds, di, d\vartheta) - ds(\sum_{k \geq 1} \delta_k(di))d\vartheta$  is the associated compensated measure. Thus  $(\Delta_{w_1, w_2}^N)_t$  is a martingale and (60) follows.  $\square$

We next study the deviation of  $(\Delta_{w_1, w_2}^N)_T$ . Define

$$(61) \quad V_{w_1, w_2}^N = (4C^*(\log N)N^{-1/2}|w_1|_{1, \infty}|w_2|_{1, \infty})^2.$$

where  $C^*$  is the constant defined in (23) in Section 3.3. Let also

$$\chi_{w_1, w_2}^N = N^{-1}|w_1|_\infty|w_2|_\infty|\mu|_\infty$$

and

$$(62) \quad \xi_{w_1, w_2}^N = 16N^{-1}|\mu|_\infty|g|_\infty|w_1|_2^2|w_2|_2^2(V_{w_1, w_2}^N)^{-1/2}(\log N)$$

**Lemma 31.** *For  $u > 2^{-6}V_{w_1, w_2}^N(\log N)^{-2}$ , we have*

$$\mathbb{P}(|(\Delta_{w_1, w_2}^N)_T| \geq u^{1/2}) \leq 2 \exp\left(-\frac{u^{1/2}}{2(\chi_{w_1, w_2}^N + \xi_{w_1, w_2}^N)}\right) + 2\mathbb{P}(N^{-1}|\mu|_\infty \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T \geq \frac{1}{2}\xi_{w_1, w_2}^N u^{1/2}).$$

*Proof.* We plan to apply a classical deviation inequality for martingales (see e.g. Lemma 2.1 in van de Geer [45] or the classical textbook by Shorak and Wellner [42]), namely:

$$(63) \quad \mathbb{P}((\Delta_{w_1, w_2}^N)_T \geq v, \langle \Delta_{w_1, w_2}^N \rangle_T \leq w) \leq \exp\left(-\frac{v^2}{2(v\chi_{w_1, w_2}^N + w)}\right)$$

for every  $v, w \geq 0$ , where  $\chi_{w_1, w_2}^N = N^{-1}|w_1|_\infty|w_2|_\infty|\mu|_\infty$  is an almost-sure bound of the size of the jumps of  $(\Delta_{w_1, w_2}^N)_T$ . With  $v = u^{1/2}$  and  $w = \xi_{w_1, w_2}^N u^{1/2}$ , inequality (63) gives

$$\mathbb{P}(|(\Delta_{w_1, w_2}^N)_T| \geq u^{1/2}) \leq 2 \exp\left(-\frac{u^{1/2}}{2(\chi_{w_1, w_2}^N + \xi_{w_1, w_2}^N)}\right) + 2\mathbb{P}(\langle \Delta_{w_1, w_2}^N \rangle_T \geq \xi_{w_1, w_2}^N u^{1/2}).$$

Inserting the term  $N^{-1} \int_0^t \int_{\mathbb{R}_+} w_1(s)^2 w_2(s-u)^2 g(s, u) duds$  in (60), we obtain

$$\langle \Delta_{w_1, w_2}^N \rangle_T \leq N^{-1} |\mu|_\infty (|w_1|_2^2 |w_2|_2^2 |g|_\infty + \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T),$$

therefore

$$\mathbb{P}(\langle \Delta_{w_1, w_2}^N \rangle_T \geq \xi_{w_1, w_2}^N u^{1/2}) \leq \mathbb{P}(N^{-1} |\mu|_\infty \mathcal{W}_{w_1^2, w_2^2}(\mathcal{F})_T \geq \frac{1}{2} \xi_{w_1, w_2}^N u^{1/2})$$

as soon as

$$(64) \quad N^{-1} |\mu|_\infty |g|_\infty |w_1|_2^2 |w_2|_2^2 < \frac{1}{2} \xi_{w_1, w_2}^N u^{1/2},$$

but by definition of  $\xi_{w_1, w_2}^N$  in (62), this condition is equivalent to  $u > 2^{-6} V_{w_1, w_2}^N (\log N)^{-2}$ .  $\square$

Under Assumption 12, we have a uniform lower bound on  $g(t, a)$ .

**Lemma 32.** *Work under Assumptions 1 and 12. Then, there exists  $\epsilon > 0$  depending on  $\delta(t, a)$  defined in (25) and (26) and  $|\mu|_\infty$  and  $T$  such that  $g(t, a) \geq \epsilon$ .*

The proof uses an explicit representation of  $g(t, a)$  established in Proposition 17 and is delayed until Appendix 8.4.

7.2.1. *Completion of proof of Theorem 13.* Let  $(h, \mathbf{h}) \in \mathcal{G}_1^N \times \mathcal{G}_2^N$  and set  $\pi(t, a) = \mu(t, a)g(t, a)$ .

**Step 1)** We plan to use the following decomposition

$$\widehat{\mu}_{h, \mathbf{h}}^N(t, a)_\varpi - \mu(t, a) = I + II,$$

with

$$I = \frac{\pi(t, a)(g(t, a) - \widehat{g}_h^N(t, a) \vee \varpi)}{g(t, a)\widehat{g}_h^N(t, a) \vee \varpi}$$

and

$$II = \frac{(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a))g(t, a)}{g(t, a)\widehat{g}_h^N(t, a) \vee \varpi}.$$

First, we have

$$|I| \leq (\epsilon\varpi)^{-1} |\mu|_\infty |g|_\infty |g(t, a) - \widehat{g}_h^N(t, a) \vee \varpi| \leq (\epsilon\varpi)^{-1} |\mu|_\infty |g|_\infty |g(t, a) - \widehat{g}_h^N(t, a)|$$

thanks to Lemma 32 as soon as  $\varpi \leq \epsilon \leq g(t, a)$ . In the same way,

$$|II| \leq (\epsilon\varpi)^{-1} |g|_\infty |\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a)|$$

follows. Picking  $h = \widehat{h}^N(t, a)$ ,  $\mathbf{h} = \widehat{\mathbf{h}}^N(t, a)$  and taking square and expectation, we have thus established

$$(65) \quad \mathbb{E}[(\widehat{\mu}_*^N(t, a)_\varpi - \mu(t, a))^2] \lesssim \mathbb{E}[(\widehat{g}_{\widehat{h}^N(t, a)}^N(t, a) - g(t, a))^2] + \mathbb{E}[(\widehat{\pi}_{\widehat{\mathbf{h}}^N(t, a)}^N(t, a) - \pi(t, a))^2]$$

as soon as  $\varpi \leq \epsilon$ . By Theorem 11, we already have the desired bound for the first term in the right-hand side of (65).

**Step 2)** We study the second term in the right-hand side of (65). For any  $\mathbf{h} \in \mathfrak{G}_2^N$ , repeating Step 1) of the proof of Theorem 11, we have

$$\mathbb{E}[(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi(t, a))^2] \lesssim \mathbb{E}[\mathbf{A}_{\mathbf{h}}^N(t, a)] + \mathbf{V}_{\mathbf{h}}^N + \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2.$$

In order to estimate  $\mathbb{E}[\mathbf{A}_{\mathbf{h}}^N(t, a)]$ , we repeat Step 2) of the proof of Theorem 11 and obtain

$$(66) \quad \max_{\mathbf{h}' \leq \mathbf{h}} \left\{ (\widehat{\pi}_{\mathbf{h}}^N(t, a) - \widehat{\pi}_{\mathbf{h}'}^N(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N - \mathbf{V}_{\mathbf{h}'}^N \right\}_+ \\ \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2 + \left\{ 4(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi_{\mathbf{h}}(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N \right\}_+ + \max_{\mathbf{h}' \leq \mathbf{h}} \left\{ 4(\widehat{\pi}_{\mathbf{h}'}^N(t, a) - \pi_{\mathbf{h}'}(t, a))^2 - \mathbf{V}_{\mathbf{h}'}^N \right\}_+.$$

**Step 3)** We estimate the expectation of the first stochastic term in the right-hand side of the last inequality. Using the same trick as in (55), we have by (59) that

$$\left\{ 4(\widehat{\pi}_{\mathbf{h}}^N(t, a) - \pi_{\mathbf{h}}(t, a))^2 - \mathbf{V}_{\mathbf{h}}^N \right\}_+ \lesssim I + II,$$

with

$$I = \left\{ 8\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N \right\}_+$$

and

$$II = \left\{ 8(\Delta_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N)_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N \right\}_+.$$

We bound each term separately. First, we have

$$\mathbb{E}[I] = \int_0^\infty \mathbb{P}(8\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T^2 - \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N \geq \kappa) d\kappa \\ = \int_0^\infty \mathbb{P}(\mathcal{W}_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N(\mathcal{F})_T \geq \frac{1}{2\sqrt{2}}(\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N + \kappa)^{1/2}) d\kappa \lesssim N^{-3}$$

applying Theorem 6 with  $w_1 = H_{h_1}(\cdot-t)$  and  $w_2 = K_{h_2}(\cdot-(t-a))$  in the same way as Step 3) in the proof of Theorem 11. As for  $II$ , we have

$$(67) \quad \mathbb{E}[II] = \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \mathbb{P}(|\Delta_{H_{h_1}(\cdot-t), K_{h_2}(\cdot-(t-a))}^N)_T| \geq \frac{1}{2\sqrt{2}}\kappa^{1/2}) d\kappa$$

and we plan to apply Lemma 31 with  $w_1 = H_{h_1}(\cdot-t)$  and  $w_2 = K_{h_2}(\cdot-(t-a))$ . Setting  $u = \frac{1}{8}\kappa$ , the condition of Lemma 31 is fulfilled as soon as  $\kappa > 8 \cdot 2^{-6} V_{H_{h_1}, K_{h_2}}^N (\log N)^{-2} = \frac{1}{8} V_{H_{h_1}, K_{h_2}}^N (\log N)^{-2}$  which is the case here since the integral in (67) above is taken for  $\kappa \geq \frac{1}{2}\mathbf{V}_{\mathbf{h}}^N = \frac{1}{2}V_{H_{h_1}, K_{h_2}}^N$ . It follows that

$$\mathbb{E}[II] \leq III + IV,$$

with

$$III = 2 \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \exp\left(-\frac{\kappa^{1/2}}{4\sqrt{2}(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)}\right) d\kappa$$

and

$$IV = 2 \int_{\frac{1}{2}\mathbf{V}_{\mathbf{h}}^N}^\infty \mathbb{P}(N^{-1}|\mu|_\infty \mathcal{W}_{(H_{h_1})^2, (K_{h_2})^2}(\mathcal{F})_T \geq \frac{1}{4\sqrt{2}}\xi_{H_{h_1}, K_{h_2}}^N \kappa^{1/2}) d\kappa.$$

First, we write

$$III = 4(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^2 \int_{v_N}^\infty \kappa e^{-\frac{\kappa}{4\sqrt{2}}} d\kappa$$

with

$$v_N = \frac{\sqrt{2}}{2}(\mathbf{V}_{\mathbf{h}}^N)^{1/2}(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^{-1}.$$

Note that

$$(V_{H_{h_1}, K_{h_2}}^N)^{1/2} = h_1^{-1/2} h_2^{-1/2} N^{-1/2} (\log N) 4C^* |H|_{1,\infty} |K|_{1,\infty}.$$

It follows that

$$\begin{aligned} & \chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N \\ &= N^{-1} |H_{h_1}|_\infty |K_{h_2}|_\infty |\mu|_\infty + 16N^{-1} |\mu|_\infty |g|_\infty |H_{h_1}|_2^2 |K_{h_2}|_2^2 (V_{|H_{h_1}|, |K_{h_2}|}^N)^{-1/2} (\log N) \\ &= N^{-1} h_1^{-1} h_2^{-1} |\mu|_\infty |H|_\infty |K|_\infty + N^{-1/2} h_1^{-1/2} h_2^{-1/2} 4C^* |\mu|_\infty |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}}. \end{aligned}$$

By definition of  $\mathcal{G}_2^N$  we have  $h_i \geq N^{-1/2}$  hence

$$(\chi_{H_{h_1}, K_{h_2}}^N + \xi_{H_{h_1}, K_{h_2}}^N)^2 \leq (|\mu|_\infty |H|_\infty |K|_\infty + 4C^* |\mu|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}})^2$$

follows and the term in front of the integral in *III* is bounded. Moreover,

$$\begin{aligned} v_N &= \frac{\sqrt{2}}{2} \frac{(\log N) 4C^* |H|_{1,\infty} |K|_{1,\infty}}{N^{-1/2} h_1^{-1/2} h_2^{-1/2} |\mu|_\infty |H|_\infty |K|_\infty + 4C^* |\mu|_\infty |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}}} \\ &\geq \frac{\sqrt{2}}{2} \frac{C^* |H|_{1,\infty} |K|_{1,\infty}}{|\mu|_\infty (|H|_\infty |K|_\infty + 4C^* |g|_\infty \frac{|H|_2^2 |K|_2^2}{|H|_{1,\infty} |K|_{1,\infty}})} (\log N) = C^{(3)} \log N \end{aligned}$$

say. Since

$$\int_{v_N}^\infty \kappa e^{-\frac{\kappa}{4\sqrt{2}}} d\kappa \lesssim v_N e^{-\frac{1}{4\sqrt{2}} v_N} \lesssim (\log N) N^{-C^{(3)}}$$

it suffices to check that  $C^{(3)} > 2$  in order to have that *III* is smaller in order than  $N^{-2}$  and thus asymptotically negligible. We finally bound the term *IV*. Applying Theorem 6 with  $(w_1, w_2) = (H_{h_1}^2, K_{h_2}^2)$ , by the change of variable

$$\frac{N}{4\sqrt{2}|\mu|_\infty} \xi_{H_{h_1}, K_{h_2}}^N \kappa^{1/2} = (1+u) C'' |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty} N^{-1/2}$$

we obtain that *IV* is of order

$$y_N \int_{z_N}^\infty (1+u) \min((\exp(u)-1)^{-1}, 1) du,$$

with

$$y_N = ((\xi_{H_{h_1}, K_{h_2}}^N)^{-1} |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty} N^{-3/2})^2 \quad \text{and} \quad z_N = \frac{N^{3/2} \xi_{H_{h_1}, K_{h_2}}^N (V_h^N)^{1/2}}{8C'' |H_{h_1}^2|_{1,\infty} |K_{h_2}^2|_{1,\infty}} - 1.$$

Straightforward computations show that  $y_N \lesssim h_1^{-2} h_2^{-2} N^{-2} \lesssim 1$  by construction of  $\mathcal{G}_2^N$ . Finally

$$z_N = \frac{8|g|_\infty |H|_2 |K|_2}{C'' |H|_\infty |K|_\infty} \frac{\log N}{h_1^{1/2} h_2^{1/2} N^{-1/2}} - 1 \geq C^{(4)} \log N - 1.$$

say. One can check that  $C^{(4)} > 2$  and we can therefore conclude that *IV* also has a negligible order.

**Step 4)** The control of the second term in the right-hand side of (66) is done in the same way as in Step 4) of the proof of Theorem 11 and only inflates the previous bound by a factor of order  $\text{Card}(\mathcal{G}_2^N) \lesssim N^2$ . In turn  $\mathbb{E}[A_h^N(t, a)] \lesssim N^{-1}$  and we have established by Step 2) that for any  $h \in \mathcal{G}_2^N$ ,

$$(68) \quad \mathbb{E}[(\widehat{\pi}_h^N(t, a) - \pi(t, a))^2] \lesssim \mathcal{B}_h^N(\pi)(t, a)^2 + V_h^N + \delta_N$$

holds true with  $\delta_N \lesssim N^{-1}$ . Putting together Step 1) and Theorem 11 completes the proof.

### 7.3. Proof of Theorem 18.

*Preliminaries.* We let  $(Z_t)_{0 \leq t \leq T}$  denote the canonical process on<sup>4</sup>  $\mathbb{D}([0, T], \mathcal{M}_{F+})$  endowed with the weak topology and equipped with its Borel sigma-field. If  $\Upsilon$  is a probability measure on  $\mathcal{M}_{F+}$  and if  $b, \mu \in \mathcal{L}_{\mathbb{D}}^{\infty}$ , we write  $\mathbb{P}_{b, \mu, \Upsilon}^N$  for the (necessarily unique) probability measure on  $\mathbb{D}([0, T], \mathcal{M}_{F+})$  under which  $(Z_t)_{0 \leq t \leq T}$  is a weak solution to (7) with  $\mathcal{L}(Z_0) = \Upsilon$ .

**Proposition 33.** *For  $i = 1, 2$ , let  $b_i, \mu_i \in \mathcal{L}_{\mathbb{D}}^{\infty}$  such that  $\text{supp}(b_2) \subset \text{supp}(b_1)$  and  $\text{supp}(\mu_2) \subset \text{supp}(\mu_1)$ . For any initial condition  $\mathcal{L}(Z_0) = \Upsilon$ , we have*

$$\|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV} \lesssim N^{1/2} (|b_1^{-1}b_2 - 1|_2 + |\mu_1^{-1}\mu_2 - 1|_2),$$

where  $\|\cdot\|_{TV}$  denotes total variation distance, up to an explicitly computable constant that only depends on  $\mu_1$  and  $b_1$ .

*Proof.* The proof is classical, and we only sketch it. Thanks to the Doléans-Dade exponential for semimartingales (see e.g. [23] or Löcherbach [32, 31] in the context of birth and death processes) and abbreviating  $f(s, a_i(Z_s^-))$  by  $f^i(s)$ , we have

$$\begin{aligned} \frac{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N}{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} &= N^{-1} \int_0^T \int_{\mathbb{R}_+} (b_2(s, a) - b_1(s, a) + \mu_2(s, a) - \mu_1(s, a)) Z_s^N(da) ds \\ &+ \int_0^T \int_{\mathbb{N} \setminus \{0\} \times \mathbb{R}_+} \mathbf{1}_{\{i \leq (NZ_s^N, \mathbf{1})\}} \left( \mathbf{1}_{\{0 \leq \vartheta \leq b_1^i(s)\}} \log \frac{b_2^i(s)}{b_1^i(s)} + \mathbf{1}_{\{b_1^i(s) \leq \vartheta \leq \mu_1^i(s)\}} \log \frac{\mu_2^i(s)}{\mu_1^i(s)} \right) \mathcal{Q}_1(ds, di, d\vartheta), \end{aligned}$$

where  $\mathcal{Q}_1$  is a Poisson random measures on  $\mathbb{R}_+ \times \mathbb{N} \setminus \{0\} \times \mathbb{R}_+$  with intensity  $ds(\sum_{k \geq 1} \delta_k(di)) d\vartheta$  under  $\mathbb{P}_{b_1, \mu_1, \Upsilon}^N$ . By Pinsker's inequality, it follows that

$$\begin{aligned} &\|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV}^2 \\ &\leq \frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[ \log \frac{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N}{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N} \right] \\ &= -\frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[ \log \frac{d\mathbb{P}_{b_2, \mu_2, \Upsilon}^N}{d\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \right] \\ &= \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[ \int_0^T \int_{\mathbb{R}_+} (b_2 - b_1 + \mu_2 - \mu_1 - b_1 \log \frac{b_2}{b_1} - \mu_1 \log \frac{\mu_2}{\mu_1})(s, a) Z_s^N(da) \right] \\ &= \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[ \int_0^T \int_{\mathbb{R}_+} (b_1 \varphi(b_1^{-1}b_2 - 1) + \mu_1 \varphi(\mu_1^{-1}\mu_2 - 1))(s, a) Z_s^N(da) \right], \end{aligned}$$

with  $\varphi(x) = x - \log(1+x) \leq x^2$  for  $x \geq 0$ . Therefore

$$\begin{aligned} \|\mathbb{P}_{b_1, \mu_1, \Upsilon}^N - \mathbb{P}_{b_2, \mu_2, \Upsilon}^N\|_{TV}^2 &\lesssim \frac{N}{2} \mathbb{E}_{\mathbb{P}_{b_1, \mu_1, \Upsilon}^N} \left[ \int_0^T \int_{\mathbb{R}_+} ((b_1^{-1}b_2 - 1)^2 + (\mu_1^{-1}\mu_2 - 1)^2)(s, a) Z_s^N(da) \right] \\ &\lesssim N (|b_1^{-1}b_2 - 1|_2^2 + |\mu_1^{-1}\mu_2 - 1|_2^2) \end{aligned}$$

and Proposition 33 is proved.  $\square$

<sup>4</sup>remember that  $\mathcal{M}_{F+}$  denotes the set of positive finite measures on  $\mathbb{R}_+$



*Representation of  $g$  in terms of  $(g_0, b, \mu)$ .* We need some notation. Let

$$L_{b,\mu}(t, a) = b(t, a) \exp\left(-\int_{t-a}^t \mu(s, s-t+a) ds\right) \text{ for } (t, a) \in \mathcal{D}_L,$$

$$M_{b,\mu,g_0}(t) = \int_0^\infty b(t, t+u) g_0(u) \exp\left(-\int_0^t \mu(s, u+s) ds\right) du, \text{ for } t \in [0, T],$$

and define  $B_{b,\mu,g_0} : [0, T] \rightarrow \mathbb{R}_+$  as the solution to the integral equation

$$(69) \quad B_{b,\mu,g_0}(t) = M_{b,\mu,g_0}(t) + \int_0^t B_{b,\mu,g_0}(a) L_{b,\mu}(t, t-a) da \text{ for every } t \in [0, T].$$

Note that Assumptions 1 and 16 ensure the existence and uniqueness of (69). Define next

$$(70) \quad g(t, a) = \begin{cases} g_0(a-t) \exp\left(-\int_0^t \mu(s, a-t+s) ds\right) & \text{on } \mathcal{D}_U \\ B_{b,\mu,g_0}(t-a) \exp\left(-\int_{t-a}^t \mu(s, a+s-t) ds\right) & \text{on } \mathcal{D}_L \end{cases}$$

and set for instance  $g(t, a) = 0$  on  $\{a = t\}$ . One can check that  $g$  defined in (70) is a weak solution to the McKendricks Von Voester equation (2).

*Completion of proof of Theorem 18.* We follow a classical two-point lower bound argument using Le Cam's lemma: if  $\mathbb{P}_i, i = 1, 2$  are two probability measures defined on the same probability space and  $\Psi(\mathbb{P}_i) \in \mathbb{R}$  is a functional of  $\mathbb{P}_i$ , we have

$$(71) \quad \inf_F \max_{i=1,2} \mathbb{E}_{\mathbb{P}_i} [|F - \Psi(\mathbb{P}_i)|] \geq \frac{1}{2} |\Psi(\mathbb{P}_1) - \Psi(\mathbb{P}_2)| (1 - \|\mathbb{P}_1 - \mathbb{P}_2\|_{TV}),$$

where the infimum is taken over all estimators of  $\Psi(\mathbb{P}_i)$ , see *e.g.* [27] among many other references.

**Step 1)** To prove (29), we pick

$$g_0 \in \mathcal{H}_L^\nu(a), \quad b_0 \in \mathcal{H}_L^{\alpha,\beta}(t, a), \quad \mu_1 \in \mathcal{H}_L^{\gamma,\delta}(t, a) \cap \mathcal{L}_{\mathcal{D},\epsilon}^\infty$$

arbitrarily, together with a sequence  $\Upsilon^N$  such that  $N \langle Z_0^N, \mathbf{1} \rangle \lesssim 1$  almost-surely under  $\Upsilon^N$  and  $\Upsilon^N(da) \rightarrow g_0(da)$  weakly as  $N \rightarrow \infty$ . Next, define

$$\mu_2^N(s, u) = \mu_1(s, u) (1 + \psi_{t-a}^N(s, u)),$$

where

$$\psi_{t-a}^N(s, u) = cN^{-1/2} \tau_N^{1/2} \psi(\tau_N(s-u-(t-a))),$$

with  $\tau_N = N^{1/(2s_{\text{death}}^- + 1)} = N^{1/(2 \max(\gamma, \delta) + 1)}$  and an infinitely many times differentiable nonnegative function  $\psi$  with compact support that satisfies  $\psi(0) = 1, |\psi|_2^2 = 1$ . Finally, pick  $c > 0$  small enough so that the property

$$\mu_2^N \in \mathcal{H}_L^{\gamma,\delta}(t, a) \cap \mathcal{L}_{\mathcal{D},\epsilon}^\infty$$

holds, uniformly in  $N$ . This is possible since

$$|\psi_{t-a}^N(\cdot, t-a)|_{\mathcal{H}^\gamma(t)} = cN^{-1/2} \tau_N^{1/2+\gamma} |\psi|_{\mathcal{H}^\gamma(t)} \leq c|\psi|_{\mathcal{H}^\gamma(t)} \leq c|\psi|_{\mathcal{H}^\gamma(t,a)}$$

and

$$|\psi_{t-a}^N(t, \cdot)|_{\mathcal{H}^\delta(t-a)} = cN^{-1/2} \tau_N^{1/2+\delta} |\psi|_{\mathcal{H}^\delta(t-a)} \leq c|\psi|_{\mathcal{H}^\delta(t-a)} \leq c|\psi|_{\mathcal{H}^\delta(a)}.$$

By Proposition 33, we have

$$(72) \quad \|\mathbb{P}_{b_0, \mu_1, \Upsilon^N} - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}\|_{TV} \lesssim N^{1/2} |\mu_1^{-1} \mu_2^N - 1|_2 = N^{1/2} |\psi_{t-a}^N|_2 = c^{1/2} |\psi|_2^2 \leq \frac{1}{2}$$

say, for large enough  $N$  and sufficiently small  $c$ .

**Step 2)** Let  $(t, a) \in \mathcal{D}_U$ . We let

$$\Psi(\mathbb{P}_{b,\mu,\Upsilon}^N) = g(t, a) = \tilde{g}(t', a') = g_0(-a') \exp\left(-\int_0^{t'} \mu(s, s-a') ds\right)$$

by (70) above, with  $(t', a') = (t, t-a) = \varphi(t, a)$ . It follows that

$$\begin{aligned} & \left| \Psi(\mathbb{P}_{b_0,\mu_2^N,\Upsilon}^N) - \Psi(\mathbb{P}_{b_0,\mu_1,\Upsilon}^N) \right| \\ &= g_0(-a') \exp\left(-\int_0^{t'} \mu_1(s, s-a') ds\right) \left| \exp\left(-\int_0^{t'} \psi_{a'}^N(s, s-a') ds\right) - 1 \right| \\ &\geq g_0(-a') \exp\left(-(|\mu_1|_\infty + |\psi_{a'}^N|_\infty)t'\right) \int_0^{t'} \psi_{a'}^N(s, s-a') ds \\ &\geq \frac{1}{2} g_0(-a') \exp(-|\mu_1|_\infty t') cN^{-1/2} \tau_N^{1/2} \psi(0) t' \\ (73) \quad &\gtrsim N^{-s_{\text{dens}}^-(2s_{\text{dens}}^-+1)} \end{aligned}$$

using  $|e^{-x} - 1| \geq xe^{-x}$  for  $x \geq 0$  and the fact that  $e^{-t'|\psi_{a'}^N|_\infty} \geq \frac{1}{2}$  say, for sufficiently large  $N$ .

**Step 3)** Let  $(t, a) \in \mathcal{D}_L$ . We now have

$$\Psi(\mathbb{P}_{b,\mu,\Upsilon}^N) = g(t, a) = \tilde{g}(t', a') = B_{b,\mu,g_0}(a') \exp\left(-\int_{a'}^{t'} \mu(s, s-a') ds\right),$$

by (70) and where  $B_{b,\mu,g_0}$  is defined in (69). It follows that

$$\begin{aligned} & \left| \Psi(\mathbb{P}_{b_0,\mu_2^N,\Upsilon}^N) - \Psi(\mathbb{P}_{b_0,\mu_1,\Upsilon}^N) \right| \\ &= \left| B_{b_0,\mu_2^N,g_0}(a') \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - B_{b_0,\mu_1,g_0}(a') \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right) \right| \\ &\geq |I| - |II|, \end{aligned}$$

with

$$I = B_{b_0,\mu_1,g_0}(a') \left( \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right) \right)$$

and

$$II = (B_{b_0,\mu_2^N,g_0}(a') - B_{b_0,\mu_1,g_0}(a')) \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right).$$

To bound  $I$  from below, we proceed as in Step 2). For simplicity, we assume moreover here that  $b_0(t, a) = b_0$  is constant. We have  $B_{b_0,\mu_1,g_0}(a') \geq M_{b_0,\mu_1,g_0}(t) \geq b_0|g_0|_1 e^{-|\mu_1|_\infty a'}$  and in the same way as for (73) one can check that

$$\begin{aligned} & \left| \exp\left(-\int_{a'}^{t'} \mu_2^N(s, s-a') ds\right) - \exp\left(-\int_{a'}^{t'} \mu_1(s, s-a') ds\right) \right| \\ &\geq \frac{1}{2} e^{-|\mu_1|_\infty(t'-a')} (t'-a') N^{-s_{\text{dens}}^-(2s_{\text{dens}}^-+1)} \end{aligned}$$

for large enough  $N$  hence

$$(74) \quad |I| \geq \frac{1}{2} b_0 |g_0|_1 e^{-|\mu_1|_\infty t'} (t'-a') cN^{-s_{\text{dens}}^-(2s_{\text{dens}}^-+1)}.$$

In order to bound  $II$  from above, we use the following technical facts that are checked in the same way as before: for every  $(t, a) \in \mathcal{D}_L$ , we have

$$\begin{aligned}
|M_{b_0, \mu_2^N, g_0}(a') - M_{b_0, \mu_1, g_0}(a')| &\leq b_0 |\mu_1|_\infty T \int_0^\infty g_0(u) \psi_{(t, a)}^N(-u) du \\
&\leq b_0 |\mu_1|_\infty T |g_0|_\infty |\psi_{t-a}^N|_1 \\
(75) \qquad \qquad \qquad &= b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 \ll N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)}
\end{aligned}$$

and

$$(76) \quad |L_{b_0, \mu_2^N, g_0}(a', a' - a) - L_{b_0, \mu_1, g_0}(a', a' - a)| \leq b_0 |\mu_1|_\infty c N^{-1/2} \tau_N^{1/2} \psi(\tau_N(u - a'))(a' - u)$$

and since  $B_{b_0, \mu, g_0}(t) \leq b_0 |g_0|_1 + b_0 \int_0^t B_{b_0, \mu, g_0}(s) ds$  for every  $t \in [0, T]$ , we infer

$$(77) \quad B_{b_0, \mu, g_0}(t) \leq b_0 |g_0|_1 e^{b_0 T}$$

by Grönwall lemma. It follows that

$$\begin{aligned}
&B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a') = M_{b_0, \mu_2^N, g_0}(a') - M_{b_0, \mu_1, g_0}(a') \\
&+ \int_0^{a'} B_{b_0, \mu_1, g_0}(a) (L_{b_0, \mu_2^N, g_0}(a', a' - a) - L_{b_0, \mu_1, g_0}(a', a' - a)) da \\
&+ \int_0^{a'} L_{b_0, \mu_2^N, g_0}(a', a' - a) (B_{b_0, \mu_2^N, g_0}(a) - B_{b_0, \mu_1, g_0}(a)) da.
\end{aligned}$$

Taking absolute values and using (75), (76) and (77), we derive

$$\begin{aligned}
|B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a')| &\leq b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 \\
&+ b_0^2 |g_0|_1 e^{b_0 T} |\mu_1|_\infty c N^{-1/2} \tau_N^{1/2} \int_0^{a'} \psi(\tau_N(u - a'))(a' - u) du \\
&+ b_0 \int_0^{a'} |B_{b_0, \mu_2^N, g_0}(s) - B_{b_0, \mu_1, g_0}(s)| ds.
\end{aligned}$$

Using  $\tau_N^{1/2} \int_0^{a'} \psi(\tau_N(u - a'))(a' - u) du \leq \tau_N^{-1/2} T |\psi|_1$ , we derive

$$|B_{b_0, \mu_2^N, g_0}(a') - B_{b_0, \mu_1, g_0}(a')| \leq b_0 |\mu_1|_\infty T |g_0|_\infty c N^{-1/2} \tau_N^{-1/2} |\psi|_1 (1 + b_0 e^{b_0 T}) e^{b_0 a'}$$

by Grönwall lemma again. We conclude

$$(78) \quad |II| \ll N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)}.$$

Comparing (74) and (78), we see that

$$(79) \quad |\Psi(\mathbb{P}_{b_0, \mu_2^N, \Upsilon}^N) - \Psi(\mathbb{P}_{b_0, \mu_1, \Upsilon}^N)| \gtrsim N^{-s_{\text{dens}}^- / (2s_{\text{dens}}^- + 1)}.$$

**Step 4).** Combining (73) or (79) with (71) and (72), we successively obtain

$$\begin{aligned}
\sup_{b, \mu, g_0} \mathbb{E}_{\mathbb{P}_{b, \mu, \Upsilon^N}^N} [|F - g(t, a)|] &\geq \frac{1}{2} \max_{i=1,2} \mathbb{E}_{\mathbb{P}_{b_0, \mu_i, \Upsilon^N}^N} [|F - \Psi(\mathbb{P}_{b_0, \mu_i, \Upsilon^N}^N)|] \\
&\geq \frac{1}{4} |\Psi(\mathbb{P}_{b_0, \mu_1, \Upsilon^N}^N) - \Psi(\mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}^N)| (1 - \|\mathbb{P}_{b_0, \mu_1, \Upsilon^N}^N - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}^N\|_{TV}) \\
&\gtrsim N^{-s_{\text{death}}^- / (2s_{\text{death}}^- + 1)}
\end{aligned}$$

and (29) follows.

**Step 5)** To prove (30), we proceed as in Step 1), considering now the perturbation

$$\mu_2^N(s, u) = \mu_1(s, u)(1 + \psi_{t,a}^N(s, u)),$$

with

$$\psi_{t,a}^N(u) = cN^{-1/2}\tau_N^{1/2}\psi(\tau_N(s-t))(\tilde{\tau}_N)^{1/2}\psi(\tilde{\tau}_N(u-a))$$

and  $\tau_N^\delta = (\tilde{\tau}_N)^\gamma = N^{s(\gamma,\delta)/(2s(\gamma,\delta)+1)}$  and an infinitely many times differentiable function  $\psi$  with compact support that satisfies  $\psi(0) = 1$ ,  $|\psi|_2^2 = 1$ . Finally, we pick  $c > 0$  small enough so that the property

$$\mu_2^N \in \mathcal{H}_L^{\gamma,\delta}(t, a) \cap \mathcal{L}_{\mathcal{D},\epsilon}^\infty$$

holds, uniformly in  $N$ . This is possible since

$$|\psi_{(t,a)}^N|_{\mathcal{H}^\gamma(t)} \leq cN^{-1/2}\tau_N^{1/2+\gamma}(\tilde{\tau}_N)^{1/2}|\psi|_{\mathcal{H}^\gamma(t)} \lesssim c$$

and

$$|\psi_{(t,a)}^N|_{\mathcal{H}^\delta(a)} \leq cN^{-1/2}\tau_N^{1/2}(\tilde{\tau}_N)^{1/2+\delta}|\psi|_{\mathcal{H}^\delta(a)} \lesssim c$$

likewise. Finally, we note that

$$(80) \quad |\mu_2^N(t, a) - \mu_1(t, a)| \geq |\mu_1(t, a)\psi_{t,a}^N(t, a)| \geq \epsilon cN^{-1/2}\tau_N^{1/2}(\tilde{\tau}_N)^{1/2} \gtrsim N^{-s_{\text{death}}^-/(2s_{\text{death}}^-+1)}$$

and

$$(81) \quad \|\mathbb{P}_{b_0, \mu_1, \Upsilon^N}^N - \mathbb{P}_{b_0, \mu_2^N, \Upsilon^N}^N\|_{TV} \lesssim N^{1/2}|\mu_1^{-1}\mu_2^N - 1|_2 = N^{1/2}|\psi_{(t,a)}^N|_2 = c^{1/2} \leq \frac{1}{2}$$

say, for sufficiently small  $c > 0$ , by Proposition 33, which conditions are satisfied since  $\mu_1$  and  $\mu_2^N$  are bounded below. The end of the proof is similar to that of Step 4) with  $\Psi(\mathbb{P}_{b, \mu, \Upsilon^N}^N) = \mu(t, a)$  together with the bounds (80) and (81). Therefore (30) is proved and Theorem 18 follows.

**7.4. Proof of Theorem 19.** By (i) of Proposition 17 the smoothness assumptions on  $(b, \mu, g)$  imply

$$(82) \quad u \mapsto g(t, u) \in \mathcal{H}_{L'}^{\min(\alpha, \beta, \gamma+1, \delta)}(a) \text{ for } (t, a) \in \mathcal{D}_L,$$

and

$$(83) \quad u \mapsto g(t, u) \in \mathcal{H}_{L'}^{\max(\gamma \wedge (\delta+1), \delta)}(a) \text{ for } (t, a) \in \mathcal{D}_U$$

for some  $L'$  that depends on  $L$  and the smoothness parameters only. For any  $h \in \mathcal{G}_1^N$ , by standard kernel approximation, see *e.g.* [44] the smoothness properties (82) and (83) together with the definition (31) of  $s_{\text{dens}}^+$  imply

$$\left| \int_0^\infty K_h(u-a)g(t, u)du - g(t, a) \right| \lesssim h^{s_{\text{dens}}^+ \wedge \ell_0},$$

up to a constant that depends on  $K$ ,  $s_{\text{dens}}^+$  and  $L'$  only. It follows that

$$\mathcal{B}_h^N(g)(t, a)^2 \lesssim h^{2s_{\text{dens}}^+ \wedge \ell_0}.$$

We also have

$$\mathbb{V}_h^N \lesssim (\log N)^2 N^{-1} h^{-1}$$

up to a constant that depends on  $C''$  of Theorem 6 and  $K$ . By Theorem 11, we conclude

$$\begin{aligned} \mathbb{E}[(\hat{g}_*^N(t, a) - g(t, a))^2] &\lesssim \min_{h \in \mathcal{G}_1^N} (h^{2s_{\text{dens}}^+ \wedge \ell_0} + (\log N)^2 N^{-1} h^{-1}) + \delta_N \\ &\lesssim \left( \frac{(\log N)^2}{N} \right)^{2s_{\text{dens}}^+ \wedge \ell_0 / (2s_{\text{dens}}^+ \wedge \ell_0 + 1)} \end{aligned}$$

using the definition of  $\mathcal{G}_1^N$ . Moreover, this estimate is uniform in  $(b, \mu, g_0)$ . The proof of Theorem 19 is complete.

**7.5. Proof of Theorem 20.** Define  $\tilde{\mu}$  via  $\mu = \tilde{\mu} \circ \varphi$  and set  $\tilde{\pi} = \tilde{\mu} \tilde{g}$ .

**Step 1)** Write  $\mu(t, a) = \tilde{\mu}(t', a') = \mu(t', t' - a')$  with  $(t', a') = \varphi(t, a) = (t, t - a)$ . The property  $\mu \in \mathcal{H}_L^{\gamma, \delta}(t, a)$  for every  $(t, a) \in \mathcal{D}$  implies  $\tilde{\mu} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \delta}(t', a')$  for every  $(t', a') \in \varphi(\mathcal{D}) = \mathcal{D}$ , for some other constant  $L'$  that depends on  $L$ . By (ii) of Proposition 17 it follows that

$$\tilde{\pi} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \min(\alpha, \beta, \gamma+1, \delta)}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_L$$

and

$$\tilde{\pi} \in \mathcal{H}_{L'}^{\min(\gamma, \delta), \delta}(t, a) \text{ for } (t, a) \in \tilde{\mathcal{D}}_U.$$

Let  $(t, a) \in \mathcal{D}_L$  so that  $\varphi(t, a) \in \tilde{\mathcal{D}}_L$ . By standard kernel approximation again, we infer

$$\begin{aligned} & \left| ((H \otimes K)_{\mathbf{h}} \circ \varphi) \star \pi(t, a) - \pi(t, a) \right| \\ &= \left| (H \otimes K)_{\mathbf{h}} \star \tilde{\pi}(\varphi(t, a)) - \tilde{\pi}(\varphi(t, a)) \right| \\ &= \left| \int_0^T \int_0^\infty H_{h_1}(\varphi_1(t, a) - s) K_{h_2}(\varphi_2(t, a) - u) \tilde{\pi}(s, u) ds du - \tilde{\pi}(\varphi(t, a)) \right| \\ &\lesssim h_1^{\min(\gamma, \delta) \wedge \ell_0} + h_2^{\min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} \end{aligned}$$

up to a constant that depends on  $H, K, L'$  and the smoothness parameters only and where we have set  $\varphi(t, a) = (\varphi_1(t, a), \varphi_2(t, a))$ . Similarly, if  $(t, a) \in \mathcal{D}_U$ , we have

$$\left| ((H \otimes K)_{\mathbf{h}} \circ \varphi) \star \pi(t, a) - \pi(t, a) \right| \lesssim h_1^{\min(\gamma, \delta) \wedge \ell_0} + h_2^{\delta \wedge \ell_0}.$$

It follows that

$$(84) \quad \mathcal{B}_{\mathbf{h}}^N(\pi)(t, a)^2 \lesssim \begin{cases} h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} & \text{if } (t, a) \in \mathcal{D}_L \\ h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \delta \wedge \ell_0} & \text{if } (t, a) \in \mathcal{D}_U. \end{cases}$$

We also have

$$(85) \quad \mathbf{V}_{\mathbf{h}}^N \lesssim (\log N)^2 N^{-1} h_1^{-1} h_2^{-1}$$

up to a constant that depends on  $C''$  of Theorem 6 and  $H, K$ .

**Step 2)** By Theorems 13 and 19, we have

$$(86) \quad \mathbb{E}[(\mu_{\star}^N(t, a)_{\varpi} - \mu(t, a))^2] \lesssim \left( \frac{(\log N)^2}{N} \right)^{2s_{\text{dens}}^+ \wedge \ell_0 / (2s_{\text{dens}}^+ \wedge \ell_0 + 1)} + \min_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\gamma)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) + \delta_N.$$

Moreover, by definition of  $s_L$  involved in (33), we have

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\alpha, \beta, \gamma+1, \delta) \wedge \ell_0} + (\log N)^2 N^{-1} h_1^{-1} h_2^{-1}) \lesssim \left( \frac{(\log N)^2}{N} \right)^{2s_L \wedge \ell_0 / (2s_L \wedge \ell_0 + 1)}$$

and likewise, by definition of  $s_U$  involved in (33), we have

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (h_1^{2 \min(\gamma, \delta) \wedge \ell_0} + h_2^{2 \min(\gamma, \delta) \wedge \ell_0} + (\log N)^2 N^{-1} h_1^{-1} h_2^{-1}) \lesssim \left( \frac{(\log N)^2}{N} \right)^{2s_U \wedge \ell_0 / (2s_U \wedge \ell_0 + 1)}.$$

Therefore, putting together (84) and (85) and using the definition of  $s_{\text{death}}$  in (33) we obtain

$$\min_{\mathbf{h} \in \mathcal{G}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\gamma)(t, a)^2 + \mathbf{V}_{\mathbf{h}}^N) \lesssim \left( \frac{(\log N)^2}{N} \right)^{s_{\text{death}}^+(t, a) \wedge \ell_0 / (2s_{\text{death}}^+(t, a) \wedge \ell_0 + 1)}.$$

Since  $s_{\text{dens}}^+ \geq s_{\text{death}}$ , inequality (86) becomes

$$\mathbb{E}[(\mu_{\star}^N(t, a)_{\varpi} - \mu(t, a))^2] \lesssim \left( \frac{(\log N)^2}{N} \right)^{s_{\text{death}}(t, a) \wedge \ell_0 / (2s_{\text{death}}(t, a) \wedge \ell_0 + 1)} + \delta_N.$$

Since the estimate is uniform in  $(b, \mu, g_0)$  and  $\delta_N \lesssim N^{-1}$ , this completes the proof of Theorem 20.

## 8. APPENDIX

### 8.1. Proof of Proposition 25.

*Preliminaries.* For  $x \geq 0$  and  $q \geq 1$ , define  $\psi_q(x) = \exp(x^q) - 1$ . Let also

$$\|\xi(f)\|_{\psi_q} = \inf \{c > 0, \mathbb{E}[\psi_q(c^{-1}\xi(f))]\} \leq 1\}$$

and

$$D = \text{diam}_d(\mathcal{F}) = \sup_{f, g \in \mathcal{F}} d(f, g).$$

**Proposition 34** (Theorem 11.2, Eq. (11.4) p. 302 in [28]). *In the setting of Proposition 25, if  $\|\xi(f) - \xi(g)\|_{\psi_q} \leq d(f, g)$  and  $E = \int_0^D \psi_q^{-1}(\mathcal{N}(\mathcal{F}, d, \epsilon)) d\epsilon < \infty$ , then*

$$\mathbb{P}(\sup_{f \in \mathcal{F}} |\xi(f)| \geq 8(E + u)) \leq \psi_q(u/D)^{-1},$$

provided  $\xi(f_0) = 0$  for some  $f_0 \in \mathcal{F}$ .

We also recall the following bound based on a classical Chernoff bound argument, proof of which we omit. For  $x \geq 0$ , let  $\tilde{\rho}(x) = (1 + x) \log(1 + x) - x$ .

**Lemma 35.** *Let  $X$  be a non-negative random variable on some probability space equipped with a probability measure  $\mathbb{Q}$ . If, for some  $k_1, k_2, k_3 > 0$ , we have*

$$\mathbb{E}_{\mathbb{Q}}[e^{\lambda X}] \leq k_1 \exp(k_2 \rho(k_3 \lambda)) \text{ for every } \lambda \geq 0,$$

then, for every  $u \geq 0$ ,

$$\mathbb{Q}(X \geq u) \leq k_1 \exp(-k_2 \tilde{\rho}(u/k_2 k_3)).$$

*Proof of Proposition 25.* Thanks to Proposition 34, all we need is an upper bound for  $\|\xi(f) - \xi(g)\|_{\psi_1}$ . Let  $\kappa > 0$ . We plan to apply Lemma 35 with  $\mathbb{Q} = \mathbb{P}(\cdot | \mathcal{A}(\kappa))$ ,  $X = |\xi(f) - \xi(g)|$ ,  $k_1 = 2\mathbb{P}(\mathcal{A}(\kappa))$ ,  $k_2 = c_1(1 + \kappa)$ ,  $k_3 = c_2 d(f, g)$  and using (46). It follows that for every  $u \geq 0$

$$(87) \quad \mathbb{P}(|\xi(f) - \xi(g)| \geq u) \leq 2 \exp(-c_1(1 + \kappa) \tilde{\rho}(u/c_1(1 + \kappa)c_2 d(f, g))) + \mathbb{P}(\mathcal{A}(\kappa)^c).$$

Now, let  $c > 0$ . We have

$$\begin{aligned} \mathbb{E}[\psi_1(c^{-1}|\xi(f) - \xi(g)|)] &= \mathbb{E}[\exp(c^{-1}|\xi(f) - \xi(g)|)] - 1 \\ &= \int_1^\infty \mathbb{P}(\exp(c^{-1}|\xi(f) - \xi(g)|) \geq \kappa) d\kappa \\ &= \int_0^\infty \mathbb{P}(|\xi(f) - \xi(g)| \geq c\kappa) e^{\kappa} d\kappa \\ &\leq 2 \int_0^\infty \exp\left(-c_1(1 + \kappa) \tilde{\rho}(c\kappa/c_1(1 + \kappa)c_2 d(f, g))\right) e^{\kappa} d\kappa + \frac{1}{2}, \end{aligned}$$

where we applied (87) with  $u = c\kappa$  and used (45) for bounding the second term. It suffices then to pick  $\varpi = \varpi(c_1, c_2) > 0$  such that

$$(88) \quad 2 \int_0^\infty \exp(-c_1(1+\kappa)\tilde{\rho}(\varpi(c_1, c_2)\kappa/c_1(1+\kappa)c_2))e^\kappa d\kappa \leq \frac{1}{2}.$$

Using (88) in the previous estimate with  $c = \varpi d(f, g)$ , we obtain

$$\mathbb{E}[\psi_1(\varpi d(f, g))^{-1}|\xi(f) - \xi(g)|] \leq 1$$

and therefore

$$\|\xi(f) - \xi(g)\|_{\psi_1} \leq \varpi d(f, g) = \tilde{d}(f, g),$$

say. We may then apply Proposition 34 with  $\tilde{d}$  instead of  $d$  and Proposition 25 follows.

**Remark 36.** In (88), we may choose  $\varpi(c_1, c_2) = k\sqrt{c_1}c_2$  for some  $k > 0$  that does not depend on  $c_1$  nor  $c_2$ . Indeed, since  $\tilde{\rho}(x) \geq \frac{1}{4}x^2$  for  $x \in [0, 1]$ , given the ansatz  $\varpi(c_1, c_2) = k\sqrt{c_1}c_2$  in (88), it suffices to show the existence of  $k$  satisfying  $k \leq \sqrt{c_1}$  and

$$(89) \quad \int_0^\infty \exp\left(-\frac{k^2}{4} \frac{\kappa^2}{1+\kappa} + \kappa\right) d\kappa \leq \frac{1}{4}.$$

One can check that (89) holds for large enough  $k$ . A rough bound is  $k = 2\sqrt{77}$ , and therefore  $c_1 \geq 308$  ensures the requirement  $k \leq \sqrt{c_1}$ .

**8.2. Proof of Proposition 7.** We have  $Z_0^N = N^{-1} \sum_{i=1}^N \delta_{A_i}$ , where the  $A_i$  are independent with common distribution  $g_0(a)da$ . Define  $\mathcal{F}_{w_2} = \{f = w_2(\cdot)g, g \in \mathcal{F}\}$ . We claim that

$$(90) \quad \sup_{f \in \mathcal{F}_{w_2}} |f|_\infty \lesssim |w_2|_\infty,$$

$$(91) \quad \mathbb{E}\left[\sup_{f \in \mathcal{F}_{w_2}} |\langle Z_0^N, f \rangle - \mathbb{E}[\langle Z_0^N, f \rangle]|\right] \lesssim N^{-1/2}|w_2|_2,$$

and

$$(92) \quad N^{-1} \sup_{f \in \mathcal{F}_{w_2}} \sum_{i=1}^N \text{Var}(f(A_i)) \lesssim |w_2|_2^2.$$

The estimates (90) and (92) are straightforward. We turn to (91). Write  $f = w_2(\cdot)g$  for  $f \in \mathcal{F}_{w_2}$ , with  $g \in \mathcal{F}$ . Adding and subtracting  $\int_0^\infty w_2(-a)g(A_i)g_0(a)da$ , we have

$$(93) \quad \langle Z_0^N, f \rangle - \mathbb{E}[\langle Z_0^N, f \rangle] = N^{-1} \sum_{i=1}^N (w_2(-A_i) - \mathbb{E}[w_2(-A_i)])g(A_i) + \left(\int_0^\infty w_2(-a)g_0(a)da\right)\nu_{w_2}^N(g),$$

with  $\nu_{w_2}^N(g) = N^{-1} \sum_{i=1}^N (g(A_i) - \mathbb{E}_{w_2}[g(A_i)])$  and where  $\mathbb{E}_{w_2}$  denotes expectation under a bias sampling proportional to  $w_2(\cdot)$ . Since  $\mathcal{F}$  is stable under  $g \mapsto -g$  and uniformly bounded, we have

$$\sup_{g \in \mathcal{F}} \left| N^{-1} \sum_{i=1}^N (w_2(-A_i) - \mathbb{E}[w_2(-A_i)])g(A_i) \right| \lesssim N^{-1} \sum_{i=1}^N (w_2(-A_i) - \mathbb{E}[w_2(-A_i)]).$$

By Cauchy-Schwarz inequality, it follows that

$$\mathbb{E}\left[\sup_{g \in \mathcal{F}} \left| N^{-1} \sum_{i=1}^N (w_2(-A_i) - \mathbb{E}[w_2(-A_i)])g(A_i) \right|\right] \leq N^{-1/2} \text{Var}(w_2(-A_i))^{1/2} \lesssim N^{-1/2}|w_2|_2.$$

In the same way as in the proof of Proposition 25, with  $\psi_2(x) = e^{x^2} - 1$ , we show using the tools in [28], p.322, that for  $g_1, g_2 \in \mathcal{F}$ , we have  $\|\nu_{w_2}^N(g_1) - \nu_{w_2}^N(g_2)\|_{\psi_2} \lesssim N^{-1/2}|g_1 - g_2|_\infty$  by Hoeffding inequality. It follows that  $\mathbb{E}[\sup_{g \in \mathcal{F}} \nu_{w_2}^N(g)] \lesssim N^{-1/2}$ . Noticing that the term  $\int_0^\infty w_2(-a)g_0(a)da$  in front of  $\nu_{w_2}^N(g)$  in (93) is of order  $|w_2|_2$  enables us to conclude the proof of (91). Noting that  $\mathcal{W}_{w_2}(\mathcal{F})_0 = \sup_{f \in \mathcal{F}_{w_2}} \langle Z_0^N, f \rangle$ , the proof of Proposition 7 is now a consequence of Lemma 6.1. in Comte *et al.* [10] on the concentration properties of  $\langle Z_0^N, f \rangle$ , based on the bounds (90), (91) and (92). We omit the details.

**8.3. Proof of Proposition 17.** The behaviour of the solution  $\xi_t(da) = g(t, a)da$  of the McKendricks Von Voester transport equation is studied in numerous textbooks, see *e.g.* [40]. The proof goes along a classical representation of  $g$  in terms of an auxiliary function solution to a certain renewal equation that enables one to study the pointwise smoothness of  $(t, a) \mapsto g(t, a)$ .

*Preliminaries.* We start with the following technical result, which is merely an observation:

**Lemma 37.** *If for some  $\sigma, \tau > 0$  and for every  $(t, a) \in \mathcal{D}$  we have  $f \in \mathcal{H}^{\sigma, \tau}(t, a)$ , then, for every  $(t', a') \in \mathcal{D}$ ,*

- (i)  $u \mapsto \int_0^u f(s, u)ds \in \mathcal{H}^{\min(\sigma+1, \tau)}(a')$ ,
- (ii)  $u \mapsto \int_0^{t'} f(s, u+s)ds \in \mathcal{H}^{\max(\sigma \wedge (\tau+1), \tau)}(a')$ .

*Proof.* Property (i) is straightforward. To obtain (ii), we first write

$$G_{t'}(u) = \int_0^{t'} f(s, u+s)ds = \int_u^{u+t'} f(s-u, s)ds = \int_u^{u+t'} \tilde{f}(s, u)ds$$

with  $\tilde{f}(s, u) = f(s-u, s)$ , so that  $\tilde{f} \in \mathcal{H}^{\min(\sigma, \tau), \sigma}(t' + a', a')$  for every  $(t', a') \in \mathcal{D}$ . Writing

$$\int_u^{u+t'} \tilde{f}(s, u)ds = \int_0^{u+t'} \tilde{f}(s, u)ds - \int_0^u \tilde{f}(s, u)ds$$

an applying (i), we obtain  $u \mapsto \int_0^u \tilde{f}(s, u)ds \in \mathcal{H}^{\min(\min(\sigma, \tau)+1, \sigma)}(a') = \mathcal{H}^{\min(\sigma, \tau+1)}(a')$  for every  $a' \in \mathbb{R}_+$ . Similarly,  $u \mapsto \int_u^{u+t'} \tilde{f}(s, u)ds \in \mathcal{H}^{\min(\sigma, \tau+1)}(a')$  therefore  $G_{t'} \in \mathcal{H}^\tau(a') \in \mathcal{H}^{\min(\sigma, \tau+1)}$ . But since  $G_{t'} \in \mathcal{H}^\tau(a')$  trivially holds, we have in fact  $G_{t'} \in \mathcal{H}^\tau(a') \cap \mathcal{H}^{\min(\sigma, \tau+1)}(a') = \mathcal{H}^{\max(\sigma \wedge (\tau+1), \tau)}(a')$ .  $\square$

*Completion of proof of Proposition 17.* For  $\sigma, \tau > 0$ , we write  $f \in \mathcal{H}^{\sigma, \tau}$  if  $f \in \mathcal{H}^{\sigma, \tau}(t, a)$  for every  $(t, a) \in \mathcal{D}$ .

**Step 1)** For fixed  $a$ , we have  $(s, t) \mapsto \mu(s, a-t+s) \in \mathcal{H}^{\gamma \wedge \delta, \delta}$  hence by (i) of Lemma 37 we have  $t \mapsto \int_0^t \mu(s, a-t+s)ds \in \mathcal{H}^{\min((\gamma \wedge \delta)+1, \delta)} = \mathcal{H}^{\min(\gamma+1, \delta)}$ . For fixed  $t$ ,  $(s, a) \mapsto \mu(s, a-t) \in \mathcal{H}^{\gamma, \delta}$  holds true, hence  $a \mapsto \int_0^t \mu(s, a-t+s)ds \in \mathcal{H}^{\min(\gamma, \delta+1) \vee \delta}$  by (ii) of Lemma 37. It follows that

$$(t, a) \mapsto \exp\left(-\int_0^t \mu(s, a-t+s)ds\right) \in \mathcal{H}^{\min(\gamma+1, \delta), \max(\gamma \wedge (\delta+1), \delta)}.$$

Also  $(t, a) \mapsto g_0(a-t) \in \mathcal{H}^{\nu, \nu} \subset \mathcal{H}^{\min(\gamma+1, \delta), \max(\gamma \wedge (\delta+1), \delta)}$  since  $\nu \geq \max(\gamma, \delta) + 1$  hence the result on  $\mathcal{D}_U$ . In the same way, on  $\mathcal{D}_L$ , we have  $t \mapsto \int_0^t \mu(s, a-t+s)ds \in \mathcal{H}^{\min(\gamma+1, \delta)}$  and  $t \mapsto \int_0^{t-a} \mu(s, a-t+s)ds \in \mathcal{H}^{\min(\gamma+1, \delta)}$  by (i) of Lemma 37 hence

$$(94) \quad t \mapsto \int_{t-a}^t \mu(s, a-t+s)ds \in \mathcal{H}^{\min(\gamma+1, \delta)}.$$



Moreover,  $\int_{t-a}^t \mu(s, a+s-t) ds = -\int_0^a \mu(s+t, a+s) ds$  and  $(s, a) \mapsto \mu(s, a+s-t) \in \mathcal{H}^{\gamma \wedge \delta, \delta}$  for fixed  $t$ , therefore

$$(95) \quad a \mapsto \int_{t-a}^t \mu(s, a-t+s) ds \in \mathcal{H}^{\min(\gamma+1, \delta)}$$

by (i) of Lemma 37 likewise. Putting together (94) and (95), we conclude

$$(96) \quad (t, a) \mapsto \exp\left(-\int_{t-a}^t \mu(s, a-t+s) ds\right) \in \mathcal{H}^{\min(\gamma+1, \delta), \min(\gamma+1, \delta)}.$$

The property  $b \in \mathcal{H}^{\alpha, \beta}$  together with (96) entail  $L_{b, \mu} \in \mathcal{H}^{\min(\alpha, \gamma+1, \delta), \min(\beta, \gamma+1, \delta)}$  hence

$$(t, a) \mapsto L_{b, \mu}(t, t-a) \in \mathcal{H}^{\min(\alpha, \gamma+1, \delta), \min(\alpha, \beta, \gamma+1, \delta)}$$

and

$$(97) \quad t \mapsto \int_0^t B_{b, \mu, g_0}(a) L_{b, \mu}(t, t-a) da \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta)},$$

follows by (i) of Lemma 37. Plainly,

$$(98) \quad t \mapsto M_{b, \mu, g_0}(t) \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta+1)}$$

and putting together (97) and (98), we conclude

$$(99) \quad t \mapsto B_{b, \mu, g_0}(t) \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta)}.$$

hence  $(t, a) \mapsto B_{b, \mu, g_0}(t-a) \in \mathcal{H}^{\min(\alpha, \beta, \gamma+1, \delta), \min(\alpha, \beta, \gamma+1, \delta)}$ . The result of Proposition 17 (i) follows.

**Step 2)** Writing  $(t', a') = \varphi(t, a) = (t, t-a)$ , the representation (70) now becomes

$$(100) \quad g(t, a) = \tilde{g}(t', a') = \begin{cases} g_0(-a') \exp\left(-\int_0^{t'} \mu(s, s-a') ds\right) & \text{on } \tilde{\mathcal{D}}_U \\ B_{b, \mu, g_0}(a') \exp\left(-\int_{a'}^{t'} \mu(s, s-a') ds\right) & \text{on } \tilde{\mathcal{D}}_L. \end{cases}$$

On  $\tilde{\mathcal{D}}_U$ , we have  $t' \mapsto \int_0^{t'} \mu(s, s-a') ds \in \mathcal{H}^{\min(\gamma, \delta)+1}$  and  $a' \mapsto \int_0^{t'} \mu(s, s-a') ds \in \mathcal{H}^{\max(\gamma \wedge (\delta+1), \delta)}$  by (ii) of Lemma 37 for the second case, hence

$$(t', a') \mapsto \exp\left(-\int_0^{t'} \mu(s, s-a') ds\right) \in \mathcal{H}^{\min(\gamma+1, \delta+1), \min(\gamma, \delta+1)}.$$

Since  $(t', a') \mapsto g_0(-a') \in \mathcal{H}^{\infty, \nu}$  hence the result since  $\nu \geq \delta$ . Similarly, on  $\tilde{\mathcal{D}}_L$ , by (99), we have  $(t', a') \mapsto B_{b, \mu, g_0}(a') \in \mathcal{H}^{\infty, \min(\alpha, \beta, \gamma+1, \delta)}$  and the same arguments as before yield

$$(t', a') \mapsto \int_{a'}^t \mu(s, s-a') ds \in \mathcal{H}^{\min(\gamma+1, \delta+1), \max(\gamma \wedge (\delta+1), \delta)}.$$

Combining these two properties gives the result on  $\tilde{\mathcal{D}}_L$  and completes (ii) of Proposition 17.

**8.4. Further estimates on the McKendricks Von Foerster equation.** The following result is a classical estimate of the renewal equation, see for instance [40].

**Lemma 38** ([40], Theorem 2.2. in Chapter 2). *Work under Assumptions 1. We have*

$$\sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da \leq \int_0^\infty g_0(a) da e^{|b-\mu|_\infty T}$$

and

$$|g|_\infty \leq \max(|g_0|_\infty, |b|_\infty \sup_{0 \leq t \leq T} \int_0^\infty g(t, a) da)$$

*Proof of Lemma 32.* On  $\mathcal{D}_U$ , by (70) in the proof of Proposition 17, we have

$$g(t, a) = g_0(a - t) \exp\left(-\int_0^t \mu(s, a - t + s) ds\right) \geq \delta(t, a) e^{-|\mu|_\infty T}$$

by (26) of Assumption 12. On  $\mathcal{D}_L$ , (70) yields the representation

$$g(t, a) = B_{b, \mu, g_0}(t - a) \exp\left(-\int_{t-a}^a \mu(s, a + s - t) ds\right) \geq B_{b, \mu, g_0}(t - a) e^{-|\mu|_\infty t}$$

and by (69), we further have

$$\begin{aligned} B_{b, \mu, g_0}(t - a) &\geq M_{b, \mu, g_0}(t - a) \\ &= \int_0^\infty b(t - a, t - a + u) g_0(u) \exp\left(-\int_0^{t-a} \mu(s, u + s) ds\right) du \\ &\geq \delta |\mathcal{U}_{(t, a)}| e^{-|\mu|_\infty (t-a)} \end{aligned}$$

by (25) of Assumption 12. The proof of Lemma 32 is complete.

**Acknowledgements** We gratefully acknowledge insightful comments and discussions with N. Champagnat, N. El Karoui, O. Lepski and V. C. Tran.

#### REFERENCES

- [1] Rudolf Beran. Nonparametric regression with randomly censored survival data. Technical report, Technical Report, Univ. California, Berkeley, 1981.
- [2] Herman J. Bierens. *Topics in advanced econometrics*. Cambridge University Press, Cambridge, 1994. Estimation, testing, and specification of cross-section and time series models.
- [3] S. Valère Bitseki Penda, Marc Hoffmann, and Adélaï de Olivier. Adaptive estimation for bifurcating Markov chains. *Bernoulli*, 23(4B):3598–3637, 2017.
- [4] Alexandre Boumezoued. Improving HMD mortality estimates with HFD fertility data. *To appear in the North American Actuarial Journal*, 2016.
- [5] Alexandre Boumezoued, Marc Hoffmann, and Paulien Jeunesse. A new inference strategy for general population mortality tables. Preprint hal-01773665, 2018.
- [6] Elodie Brunel, Fabienne Comte, and Agathe Guilloux. Estimation strategies for censored lifetimes with a lexis-diagram type model. *Scandinavian Journal of Statistics*, 35(3):557–576, 2008.
- [7] Andrew J. G. Cairns, David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, Alen Ong, and Igor Balevich. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *N. Am. Actuar. J.*, 13(1):1–35, 2009.
- [8] Andrew JG Cairns, David Blake, Kevin Dowd, and Amy R Kessler. Phantoms never die: living with unreliable population data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):975–1005, 2016.
- [9] Stéphane Cléménçon, Viet Chi Tran, and Hector De Arazoza. A stochastic SIR model with contact-tracing: large population limits and statistical inference. *Journal of Biological Dynamics*, 2(4):392–414, 2008.

- [10] Fabienne Comte, Jérôme Dedecker, and Marie-Luce Taupin. Adaptive density deconvolution with dependent inputs. *Math. Meth. Statist.*, 17:87–112, 2008.
- [11] Fabienne Comte, Stéphane Gaïffas, and Agathe Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 47, pages 1171–1196. Institut Henri Poincaré, 2011.
- [12] Dorota M Dabrowska. Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*, pages 181–197, 1987.
- [13] Marie Doumic, Marc Hoffmann, Nathalie Krell, and Lydia Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3):1760–1799, 2015.
- [14] R. M. Dudley. Universal Donsker classes and metric entropy. *Ann. Probab.*, 15(4):1306–1326, 1987.
- [15] Nicolas Fournier and Sylvie Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.
- [16] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016.
- [17] Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008.
- [18] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632, 2011.
- [19] Wolfgang Härdle, Gerard Kerkycharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.
- [20] HFD. The human fertility database. max planck institute for demographic research (germany) and vienna institute of demography (austria).
- [21] HMD. The human mortality database.
- [22] Marc Hoffmann and Adélaïde Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and their Applications*, 126(5):1433–1471, 2016.
- [23] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [24] Niels Keiding. Statistical inference in the lexis diagram. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 332(1627):487–509, 1990.
- [25] Thierry Klein and Emmanuel Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33:1060–1077, 2005.
- [26] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2):298–335, 2017.
- [27] Lucien Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
- [28] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [29] O. V. Lepskiĭ. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.
- [30] O. V. Lepskiĭ. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4):645–659, 1991.
- [31] E. Löcherbach. Likelihood ratio processes for Markovian particle systems with killing and jumps. *Stat. Inference Stoch. Process.*, 5(2):153–177, 2002.
- [32] Eva Löcherbach. LAN and LAMN for systems of interacting diffusions with branching and immigration. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(1):59–90, 2002.
- [33] Mark G. Low. Nonexistence of an adaptive estimator for the value of an unknown probability density. *Ann. Statist.*, 20(1):598–602, 1992.
- [34] Ian W McKeague and Klaus J Utikal. Inference for a nonlinear counting process regression model. *The Annals of Statistics*, pages 1172–1187, 1990.
- [35] A.G. McKendrick. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.*, 54:98–130, 1926.
- [36] Sylvie Méléard and Viet Chi Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Process. Appl.*, 122(1):250–276, 2012.
- [37] È. A. Nadaraja. On a regression estimate. *Teor. Veroyatnost. i Primenen.*, 9:157–159, 1964.

- [38] Jens P. Nielsen and Oliver B. Linton. Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.*, 23(5):1735–1748, 1995.
- [39] Jens P. Nielsen and Oliver B. Linton. Kernel estimation in a nonparametric marker dependent hazard model. *The Annals of Statistics*, pages 1735–1748, 1995.
- [40] Benoît Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.
- [41] SJ Richards. Detecting year-of-birth mortality patterns with limited data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(1):279–298, 2008.
- [42] Galen R. Shorack and Jon A. Wellner. *Empirical processes with applications to statistics*, volume 59 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. Reprint of the 1986 original [MR0838963].
- [43] Viet Chi Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Probab. Stat.*, 12:345–386, 2008.
- [44] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [45] Sara van de Geer. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.*, 23(5):1779–1801, 1995.
- [46] H. Von Foerster. *The Kinetics of Cellular Proliferation*. Grune & Stratton, 1959.

A. BOUMEZOUED, MILLIMAN R&D, 14 AVENUE DE LA GRANDE ARMÉE, 75017 PARIS, FRANCE.  
*E-mail address:* alexandre.boumezoued@milliman.com

MARC HOFFMANN, UNIVERSITÉ PARIS-DAUPHINE & PSL, CNRS, CEREMADE, 75016 PARIS, FRANCE  
*E-mail address:* hoffmann@ceremade.dauphine.fr

PAULIEN JEUNESSE, UNIVERSITÉ PARIS-DAUPHINE & PSL, CNRS, CEREMADE, 75016 PARIS, FRANCE  
*E-mail address:* jeunesse@ceremade.dauphine.fr