Nonparametric estimation of the fragmentation kernel based on a PDE stationary distribution approximation

Van Hà Hoang[*], Thanh Mai Pham Ngoc[†], Vincent Rivoirard[‡], Viet Chi Tran[§]

Abstract

We consider a stochastic individual-based model in continuous time to describe a size-structured population for cell divisions. This model is motivated by the detection of cellular aging in biology. We address here the problem of nonparametric estimation of the kernel ruling the divisions based on the eigenvalue problem related to the asymptotic behavior in large population. This inverse problem involves a multiplicative deconvolution operator. Using Fourier technics we derive a nonparametric estimator whose consistency is studied. The main difficulty comes from the non-standard equations connecting the Fourier transforms of the kernel and the parameters of the model. A numerical study is carried out and we pay special attention to the derivation of bandwidths by using resampling.

Keywords: Growth-fragmentation; cell division; nonparametric estimation; Kernel rule; deconvolution;

MSC2010: 62G07; 92D25; 60J80; 45K05; 35B40

1 Introduction

We consider a population model with size structure in continuous time, where individuals are cells which grow continuously and undergo binary divisions after random exponential times at rate $R > 0$. When a cell of size x divides, it dies and is replaced by two daughter cells of sizes γx and $(1-\gamma)x$, where γ is assumed here to be a random variable drawn according to a distribution with a density with respect to the Lebesgue measure on $[0, 1]$: $\Gamma(d\gamma) = h(\gamma)d\gamma$. Between divisions, the sizes of the cells grow with speed $\alpha > 0$. Because the two daughter cells are exchangeable, we assume that h is a symmetric density with respect to $\gamma = 1/2$. When h is piked at $1/2$, then both daughters tend to have similar sizes, *i.e.* the half of their mother's size. The more h puts weight in the neighbourhood of 0 and 1, the more asymmetric the divisions are. They give birth to one small daughter and one big daughter with size close to its mother's. In this article, we are interested in the estimation of this function h in the case of large populations where the division tree is not observed. We stick to constant rate R and speed α for the sake of simplicity.

The population can be described by a stochastic individual-based (particle) model, where the population at time t is represented by a random measure that is the sum of Dirac masses on

[*]Laboratoire de Mathématiques, UMR 8628, Université Paris Sud, 91405 Orsay Cedex France; E-mail: vanha.hoang@math.u-psud.fr

[†]Laboratoire de Mathématiques, UMR 8628, Université Paris Sud, 91405 Orsay Cedex France; E-mail: thanh.pham_ngoc@math.u-psud.fr

[‡]Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL Research University, 75016 Paris, France; E-mail: Vincent.Rivoirard@dauphine.fr

[§]Univ. Lille, CNRS, UMR 8524 - Laboratoire P. Painlevé, UFR de Mathématiques, F-59000 Lille, France; E-mail: chi.tran@math.univ-lille1.fr

$\mathbb{R}_+$ weighting the cells' sizes. Stochastic continuous time individual-based models of dividing cell populations with size-structure have made the subject of an abundant literature starting from Athreya and Ney [1], Harris [22], Jagers [27] etc. until recent years (*e.g.* Bansaye *et al.* [6, 3], Cloez [10]). Similar models in discrete time should also be mentioned (*e.g.* [2, 5, 7, 35, 14, 21]). Individual-based models are easy to simulate and offer sometimes a convenient framework for statistics (see *e.g.* Hoffmann and Olivier [26], Hoang [25, 24]). They also connect to the partial differential equations (PDEs) that are usually used in population dynamics (see [4]).
We start from an initial population where the individuals are labelled in an exchangeable way by integers. The population of cells descending from these initial individuals can be seen as the forest of trees rooted in these initial individuals. We use the Ulam-Harris-Neveu notation to label the cells appearing in the population: if the mother has a label $i \in \mathcal{I} = \cup_{\ell \geq 1} \mathbb{N} \times \{0,1\}^{\ell-1}$, then the two daughters have labels $i0$ and $i1$ obtained by concatening the mother's label with integers 0 or 1.
The population at time t is described by the point measure:

$$Z_t^K = \frac{1}{K} \sum_{i \in V_t^K} \delta_{x_i(t)}, \tag{1.1}$$

where δ is the Dirac Delta function, V_t^K is the set of labels of living individuals at time t and K is a renormalizing parameter corresponding to the order of the initial population size. The parameter K will tend to $+\infty$ in the sequel. The individual with label $i \in V_t^K$ is represented by a Dirac mass weighting the size $x_i(t)$ of this individual at time t.

When the complete division forest is observed, we can associate to each division an independent random variable with distribution h: if T_i is the division time of the cell i, then, we define $\Gamma_i = x_{i0}(T_i)/x_i(T_{i-})$. Estimating the function h from such a sample has been considered in [24, 25]. Here, we focus on the situation when the division tree is not completely observed. Following ideas from Doumic *et al.* [17, 18, 15] or Bourgeron [9] whose aim was to recover the division rate R when the latter depends on the size, our strategy is to consider the PDE approximating the evolution of the measure-valued process $(Z_t^K)_{t\geq 0}$ when K is large. The long-time behavior of the solution of this PDE can be studied thanks to an eigenvalue problem. This yields a stationary distribution $N(x)dx$ from which we can assume that we have drawn a sample of n i.i.d. random variables $X_1, \dots, X_n$. The function h is then solution to an intricate inverse problem involving a multiplicative convolution operator. We use deconvolution techniques inspired by those used by Comte and Lacour [12, 11], Comte *et al.* [13], Neumann [34] to construct and study a kernel estimator of h. Changing variables and taking Fourier transforms lead us to an equation where the regularities of the different terms are strongly related to the regularity of the unknown function h to be estimated. The consistency of the estimator is studied, and simulations are performed. In particular, we discuss and illustrate numerically the bandwidth selection rules for the kernel estimator.

The paper is organized as follows. Section 2 describes the miscroscopic model. Section 3 tackles the problem of estimating the division kernel h. Section 4 presents the numerical performances of our estimation procedure. Eventually, all the proofs are gathered in the Appendix.

Notation: We denote by $\mathcal{M}_F(\mathbb{R}_+)$ the space of finite measures on $\mathbb{R}_+$ endowed with the weak convergence topology. For $\mu \in \mathcal{M}_F(\mathbb{R}_+)$ and for $f \in \mathcal{C}_b(\mathbb{R}_+, \mathbb{R})$ a bounded continuous real function on $\mathbb{R}_+$, $\langle \mu, f\rangle = \int_{\mathbb{R}_+} f d\mu$ is the integral of f with respect to μ. We denote by $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ the space of càdlàg functions from $\mathbb{R}_+$ to $\mathcal{M}_F(\mathbb{R}_+)$ embedded with the Skorokhod topology (e.g. [8]).

The Fourier transform of any integrable function f is defined by

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

2 Microscopic model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(Z_0^K)_{K\in\mathbb{N}^*}$ be a sequence of random point measures on $\mathbb{R}_+$ of the form (1.1) that converges to $\xi_0 \in \mathcal{M}_F(\mathbb{R}_+)$ in distribution and for the weak convergence topology on $\mathcal{M}_F(\mathbb{R}_+)$. We also assume that

$$\sup_{K\in\mathbb{N}^*} \mathbb{E}(\langle Z_0^K, 1\rangle^2) < +\infty. \tag{2.1}$$

For each $K \in \mathbb{N}^*$ and initial condition Z_0^K as above, we can represent the measure-valued process $(Z_t^K)_{t\geq 0}$ as the unique solution of a stochastic differential equation (SDE) driven by a Poisson point measure that satisfies the following martingale problem.

Proposition 1. *For a given $K \in \mathbb{N}^*$ and a test function $f(x,s) \in \mathcal{C}_b^1(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$, the process $(Z_t^K)_{t\geq 0}$ satisfies:*

$$\begin{aligned}\langle Z_t^K, f\rangle = \langle Z_0^K, f\rangle + \int_0^t \int_{\mathbb{R}_+} \Big(&\partial_s f_s(x) + \alpha\partial_x f_s(x) \\ &+ R\int_0^1 \big(f(\gamma x) + f((1-\gamma)x) - f(x)\big)h(\gamma)d\gamma\Big) Z_s^K(dx)ds + M_t^{K,f},\end{aligned} \tag{2.2}$$

where $(M_t^{K,f})_{t\geq 0}$ is a square integrable martingale started at 0 with bracket:

$$\langle M^{K,f}\rangle_t = \frac{1}{K}\int_0^t \int_{\mathbb{R}_+} R\big(f(\gamma x) + f((1-\gamma)x) - f(x)\big)^2 h(\gamma)d\gamma Z_s^K(dx)ds. \tag{2.3}$$

The detailed construction of the SDE satisfied by $(Z_t^K)_{t\geq 0}$ is given in Appendix A, as well as a sketch of proofs for the results of this section. The martingale property and quadratic variation are direct consequences of stochastic calculus with the SDE. The following theorem states the limit of $(Z^K)_{K\in\mathbb{N}^*}$ when $K \to +\infty$.

Theorem 1. *If $(Z_0^K)_{k\in\mathbb{N}^*}$ converges in distribution to $\xi_0 \in \mathcal{M}_F(\mathbb{R}_+)$ as $K \to +\infty$ then $(Z^K)_{K\in\mathbb{N}^*}$ converges in distribution in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ as $K \to +\infty$ to the unique solution $\xi \in \mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ of*

$$\begin{aligned}\langle \xi_t, f\rangle = \langle \xi_0, f\rangle + \int_0^t \int_{\mathbb{R}_+} \Big(&\partial_s f_s(x) + \alpha\partial_x f_s(x) \\ &+ R\int_0^1 \big(f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\big)\, h(\gamma)d\gamma\Big)\xi_s(dx)ds,\end{aligned} \tag{2.4}$$

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

When the limiting initial condition ξ_0 admits a smooth density with respect to the Lebesgue measure, the following proposition allows us to connect the measure-valued processes with the growth-fragmentation integro-differential equations usually introduced for cell divisions (*e.g.* [37, 19]).

Proposition 2. *If ξ_0 has a density $n_0 \in C_b^1(\mathbb{R}_+, \mathbb{R}_+)$ with respect to the Lebesgue measure on $\mathbb{R}_+$, then $\forall t \in \mathbb{R}_+$, $\xi_t(dx)$ admits a density $n(t,x)$ that is the unique solution of the PDE:*

$$\partial_t n(t,x) + \alpha \partial_x n(t,x) + Rn(t,x) = 2R \int_0^\infty n(t,y) h\left(\frac{x}{y}\right) \frac{dy}{y}, \tag{2.5}$$

where $h(x/y) = 0$ if $y < x$ (since h is supported by $[0,1]$).

The long time behaviour of the solution of PDE (2.5) is well-known and presented in the following proposition. In the sequel, we shall base our statistical estimation of h on the long time limit of the PDE. Notice that by change of variable in the integral, the right hand side of Equation (2.5) can also be rewritten as: $2R \int_0^1 n(t, x/u) h(u)\, du/u$. We observe that a convenient assumption on the density h is the following:

$$\int_0^1 h(u) \frac{du}{u} < +\infty. \tag{2.6}$$

In the sequel, a stronger assumption will be needed to obtain the consistency of our estimators.

Proposition 3. *Assume* (2.6). *Then, there exists a unique probability density $N \in \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ solving the following system:*

$$\begin{cases} \alpha \partial_x N(x) + 2R\, N(x) = 2R \int_0^\infty N(y) h\left(\frac{x}{y}\right) \frac{dy}{y}, \quad x \geq 0, \\ N(0) = 0, \quad \int N(x) dx = 1, \quad N(x) \geq 0. \end{cases} \tag{2.7}$$

With $\rho = \|n_0\|_1 = \int_0^\infty n_0(u) du$ (where n_0 has been introduced in Prop. 2), we have:

$$\int_0^\infty |n(t,x)e^{-Rt} - \rho N(x)| dx \leq e^{-Rt}\big(\|g_0\|_1 + \frac{6R}{\alpha}\|G_0\|_1\big), \tag{2.8}$$

where $g_0(x) = n_0(x) - \rho N(x)$, and $G_0(x) = \int_0^x g_0(y) dy$.

Proposition 3 shows that the renormalized population density $\rho^{-1} n(t,x) e^{-Rt}$ converges exponentially fast, when the time t tends to infinity, to a stationary density $N(x)$ that is obtained by solving an eigenvalue problem. The proof of Proposition 3 is given in Appendix A. Notice that we do not have such a strong result if the division rate is not a constant. We explain in the next section the building of our statistical estimation procedure based on the results of this proposition.

3 Estimation of the division kernel

3.1 Estimation procedure and assumptions

3.1.1 Principle

We consider the problem of estimating the density h in the case of incomplete data of divisions. As explained previously, we shall construct an estimator of h based on the stationary size distribution which results from the study of the large population limit $n(t,x)$. The long time behavior provides us an observation scheme for the estimation of the density h in the statistical approach: since $e^{-Rt} n(t,x)$ converges exponentially fast to $N(x)$ (up to a constant) as t increases by Proposition 3, when we pick n cells randomly in the population at a large time t, we can assume that we have n i.i.d observations $X_1, X_2, \dots, X_n$ with distribution $N(x)dx$. We estimate h from the data $X_1, \dots, X_n$ and Equation (2.7).

This corresponds to a deconvolution problem but it is more complicated and quite different when compared to classical deconvolution problems. The convolution in Equation (2.7) is a multiplicative convolution $\int_0^\infty N(y)h\left(\frac{x}{y}\right)\frac{dy}{y}$ leading to more intricate technical problems than for the classical additive convolution. So, we apply a logarithmic change of variables to transform the multiplicative convolution in the right hand side of (2.7) into an additive one. Then, we classically apply the Fourier transform and work with products of functions in the Fourier domain. Let us describe our estimation procedure in details.
By using the change of variable $x = e^u$ for $x > 0$ and $u \in \mathbb{R}$, we introduce the functions

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

and

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

Equation (2.7) becomes

$$\alpha D(u) + 2R\,M(u) = 2R\big(M \star g)(u). \tag{3.1}$$

We have $h(\gamma) = \gamma^{-1}g\big(\log(\gamma)\big)$ for $\gamma \in (0,1)$. Then, the estimator of h will be obtained from the estimator of g once we have obtained estimators for unknown functions M and D.

3.1.2 Assumptions on h

First, assumptions on the density h are needed in the sequel. Of course, since h is the density of a symmetric distribution on $[0,1]$, it satisfies $\int h(x)dx = 1$ and $\int xh(x)dx = 1/2$. For the proofs, we will also need the following condition.

Assumption 1. *The function h is of class $\mathcal{C}^\beta$ on $[0,1]$, for some $\beta > 3$: the function h is $[\beta]$ times differentiable (where $[\beta]$ is the largest integer smaller than β) and the derivative of order $[\beta]$ is $\beta - [\beta]$ Hölder continuous.*
Moreover, we assume that there exists a positive integer $\nu_0 \geq 2$ such that for all $k \in \{0,\dots,\nu_0\}$, $h^{(k)}(0) = 0$.

Under Assumption 1, h can take positive values only on $(0,1)$, and the function g introduced previously is supported by $\mathbb{R}_-$.

Remark 1. *Assumption 1 implies (2.6). For $t \in (0,1)$, by Taylor's formula, there exists indeed $\theta \in (0,1)$ such that:*

$$0 \leq \frac{h(t)}{t} = \sum_{k=\nu_0+1}^{[\beta]-1} \frac{1}{k!}h^{(k)}(0)t^{k-1} + \frac{h^{[\beta]}(\theta t)}{[\beta]!}t^{[\beta]-1},$$

which is integrable in the neighborhood of 0 (the sum in the right hand side being 0 if $\nu_0 + 1 > [\beta] - 1$.

This remark shows that, under Assumption 1, the results of Proposition 3 are hence available to justify our approximation to start with a sample of i.i.d. random variables with density $N(x)$. We also have the following result that will be needed to show consistency (the proof is in Appendix B):

Lemma 1. *Under Assumption 1:*
(i) the first eigenvector N of the eigenproblem (2.7) satisfies

$$\int_0^{+\infty} x^{-\nu}N(x)dx < +\infty \quad \textit{for } \nu \in \{1,\dots,(\nu_0+2)\wedge([\beta]+1)\}. \tag{3.2}$$

(ii) M is of class $\mathcal{C}^{[\beta]}$ and its Fourier transform M^* satisfies:

$$\limsup_{|\xi|\to+\infty}\left\{|\xi|^{[\beta]\wedge(\nu_0+3)}\times|M^*(\xi)|\right\}<+\infty.$$

(iii) The extension of $M^*(\xi)$ to the complex plane, $\xi\in\mathbb{C}\mapsto M^*(\xi)=\int_{\mathbb{R}}e^{\mathbf{i}x\xi}M(x)dx$, is holomorphic and thus, M^* admits only isolated zeros. Moreover, M^* does not admit zeros on the real line.

The point (i) is crucial for proving the consistency. This proof relies on the use of the Rosenthal inequality (see Eq. (D.3)). This explains why we need $\nu\geq 4$ and hence $\nu_0\geq 2$ and $\beta>3$ in Assumption 1. The point (ii) establishes strong connections between the regularities of functions involved in (2.7). Paradoxically, the more regular h is, the faster M^* converges to 0 at infinity, which may lead to some difficulties in view of the subsequent (3.3). Fortunately, point (iii) shows that $M^*(\xi)$ does not vanish on the real line.

3.1.3 Fourier transformation

Notice that g is square integrable since we have

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

We can thus take the Fourier transform of both sides of equation (3.1). We obtain

$$\alpha D^*(\xi)+2R\,M^*(\xi)=2RM^*(\xi)\times g^*(\xi).$$

Therefore, under Assumption 1, the Fourier transform of g is obtained via the formula

$$g^*(\xi)=\frac{\alpha D^*(\xi)}{2RM^*(\xi)}+1,\quad \xi\in\mathbb{R}. \tag{3.3}$$

Note that Equation (3.3) is not standard in classical inverse problems. Indeed, for instance in the density deconvolution setting, the Fourier transform of the noise appears at the denominator in place of M^*. Here, M is connected to g which has to be estimated and thus cannot be handled as the usual noise.

3.1.4 Estimators of g and h

Given the sample of i.i.d random variables $X_1,\ldots,X_n$ with density function $x\mapsto N(x)$, we can consider the random variables $U_1,\ldots,U_n$ defined as $U_i=\log(X_i)$. These random variables are i.i.d of density function $u\mapsto M(u)=e^uN(e^u)$. In view of (3.3), the purpose is first to propose an estimator for g^* and then to apply the inverse Fourier transform to obtain an estimator of g. Our procedure will be naturally based on $\widehat{M^*(\xi)}$ and $\widehat{D^*(\xi)}$, estimators of $M^*(\xi)$ and $D^*(\xi)$ respectively, and defined by

$$\widehat{M^*(\xi)}=\frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j}, \tag{3.4}$$

$$\widehat{D^*(\xi)}=(-\mathbf{i}\xi)\frac{1}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}. \tag{3.5}$$

It's straightforward to show that $\widehat{M^*(\xi)}$ and $\widehat{D^*(\xi)}$ are unbiased estimators of $M^*(\xi) = \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big]$ and $D^*(\xi) = (-\mathbf{i}\xi)\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]$ respectively.

As usual in the nonparametric setting, the estimate of g will be obtained by regularization technics. For density estimation, convoluting by an appropriate rescaled kernel is a natural methodology. Convolution is expressed by products in the Fourier domain. So, let K a kernel function in $\mathbb{L}^2(\mathbb{R})$ such that its Fourier transform K^* exists and is compactly supported. A possible kernel is given by the sinus cardinal kernel $K(x) = \frac{\sin(x)}{\pi x}$ for which $K^*(t) = \mathbb{1}_{[-1,1]}(t)$. For $\ell > 0$, define

$$K_\ell(\cdot) := \frac{1}{\ell} K\left(\frac{\cdot}{\ell}\right).$$

Definition 1. *Given $\ell > 0$, the estimate $\hat{g}_\ell$ of g is defined through its Fourier transform:*

$$\widehat{g_\ell}^*(\xi) = K_\ell^*(\xi) \times \left(\frac{\alpha \widehat{D^*(\xi)}}{2R} \frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} + 1 \right), \tag{3.6}$$

where $\Omega = \big\{|\widehat{M^*(\xi)}| \geq n^{-1/2}\big\}$ and $\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}}$ is the truncated estimator of $\frac{1}{\widehat{M^*(\xi)}}$:

$$\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} = \begin{cases} \dfrac{1}{\widehat{M^*(\xi)}}, & \text{if } |\widehat{M^*(\xi)}| \geq n^{-1/2}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

Truncation is necessary to avoid explosion when $|\widehat{M^*(\xi)}|$ is close to 0. Finally, taking the inverse Fourier transform of $\widehat{g_\ell}^*$, we obtain the estimator of g.

Definition 2. *The estimator of g is*

$$\hat{g}_\ell(u) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{g}_\ell^*(\xi) e^{-\mathbf{i}u\xi} d\xi. \tag{3.8}$$

The estimator of the division kernel h is deduced from $\hat{g}_\ell$:

$$\hat{h}_\ell(\gamma) = \gamma^{-1} \hat{g}_\ell\big(\log(\gamma)\big), \quad \gamma \in (0,1). \tag{3.9}$$

The main difficulty lies in the choice of ℓ. This problem is dealt with subsequently. Deconvolution estimators have been studied in Comte and Lacour [12, 11], Comte *et al.* [13], Neumann [34]. However, the difference and the difficulty in our problem come from the fact that the regularities of g and h are closely related to the functions M and D that solve the eigenvalue problem (2.7), in particular through Equation (3.3). This complicates the study of the rates of convergence. The next section studies the quadratic risk of $\hat{g}_\ell$ and $\hat{h}_\ell$.

3.2 Study of the quadratic risk

3.2.1 Relations between the risks of the estimators of h and g

The first goal is to connect the $\mathbb{L}^2$-risk of $\hat{h}_\ell$ and the $\mathbb{L}^2$-risk of $\hat{g}_\ell$. Using a randomized estimator, we can show the following result.

Proposition 4. *For a Bernoulli random variable τ with parameter $1/2$ independent of $X_1, \dots, X_n$, let us define the randomized estimator*

$$\breve{g}_\ell(u) = \tau \hat{g}_\ell(u) + (1-\tau)\tilde{g}_\ell(u), \ \text{where } \tilde{g}_\ell(u) = e^u \hat{h}_\ell(1 - e^u).$$

We have

$$\mathbb{E}\big[\|\hat{h}_\ell - h\|_2^2\big] = 2\mathbb{E}\big[\|\breve{g}_\ell - g\|_2^2\big] = \mathbb{E}\Big[\int_{\mathbb{R}_-} e^{-u}\big(\hat{g}_\ell(u) - g(u)\big)^2 du\Big]. \tag{3.10}$$

The last equality in (3.10) shows that if we want to control the quadratic risk of $\hat{h}_\ell$ with respect to the Lebesgue measure, tight controls on the loss of $\hat{g}_\ell$ at $-\infty$ are needed. But, since h, as defined in our biological problem, is a symmetric function (as the daughter cells obtained after a division are exchangeable), it is natural to consider

$$\hat{h}_\ell^{sym}(x) = \frac{1}{2}\big(\hat{h}_\ell(x) + \hat{h}_\ell(1-x)\big), \tag{3.11}$$

whose quadratic risk is controlled by the quadratic risk of $\hat{g}_\ell$ except at boundaries of the interval $[0,1]$, as proved by the next proposition.

Proposition 5. *Setting $m(x) = x(1-x)$, we have that*

$$\int_0^1 \big(\hat{h}_\ell^{sym}(x) - h(x)\big)^2 m(x)dx \leq \|\hat{g}_\ell - g\|_2^2\,. \tag{3.12}$$

Propositions 4 and 5 are proved in Appendix C. The previous result does not provide any control on boundaries of the interval $[0,1]$ but the consistency of $\hat{g}_\ell$ will establish the consistency of $\hat{h}_\ell^{sym}$ on every compact set of $(0,1)$. The study of the consistency of $\hat{g}_\ell$ is the goal of the next section.

3.2.2 Consistency of the estimator of g for the quadratic-risk

This section is devoted to the theoretical study of the estimate $\hat{g}_\ell$. More precisely, we establish the $\mathbb{L}^2$-consistency of $\hat{g}_\ell$ under a suitable choice of the bandwidth ℓ.

We first study the bias-variance decomposition of the $\mathbb{L}^2$-risk of $\hat{g}_\ell$. Recall that from Lemma 1(iii), we have that under Assumption 1, $|M^*(\xi)|$ is strictly positive on every compact set of the real line $\xi \in [-A, A]$, $A > 0$, and thus lower bounded by a positive constant on each of these intervals (that depends on A).

Theorem 2. *Under Assumption 1, there exists a positive constant $C < +\infty$ such that*

$$\mathbb{E}\left[\|\hat{g}_\ell - g\|_2^2\right] \leq \|K_\ell \star g - g\|_2^2 + \frac{C}{n} S(\ell), \tag{3.13}$$

where

$$S(\ell) = \Big\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\Big\|_2^2 + \Big\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\Big\|_2^2.$$

Then the following corollary gives the $\mathbb{L}^2$-consistency of the estimator $\hat{g}_\ell$.

Corollary 1. *We suppose that Assumption 1 is satisfied and the kernel bandwidth $\ell = \ell(n)$ satisfies $\lim_{n\to+\infty} \ell = 0$. Provided that*

$$\lim_{n\to+\infty} \frac{1}{n}\left(\Big\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\Big\|_2^2 + \Big\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\Big\|_2^2\right) = 0, \tag{3.14}$$

we have

$$\lim_{n\to+\infty} \mathbb{E}\left[\|\hat{g}_\ell - g\|_2^2\right] = 0. \tag{3.15}$$

The proof of these results is given in Appendix D. Note that under Assumption 1, we have by Lemma 1 that $|M^*(\xi)| = O(|\xi|^{-([\beta]\wedge(\nu_0+3))})$ when $|\xi| \to +\infty$. If we have $|M^*(\xi)| \sim C|\xi|^{-([\beta]\wedge(\nu_0+3))}$, for a constant $C > 0$, then if we still take $K(x) = \frac{\sin(x)}{\pi x}$, we can derive a bandwidth ℓ. Indeed,

$$K_\ell^*(\xi) = K^*(\ell\xi) = \mathbb{1}_{[-\ell^{-1},\ell^{-1}]}(\xi)$$

and

$$\left\| \frac{K_\ell^*(\xi)\xi}{M^*(\xi)} \right\|_2^2 = \int_{-\ell^{-1}}^{\ell^{-1}} \frac{\xi^2}{|M^*(\xi)|^2} d\xi = O(\ell^{-(3+2([\beta]\wedge(\nu_0+3)))}),$$

and then, Assumption (3.14) is satisfied if

$$\ell^{-1} = o\left(n^{\frac{1}{3+2([\beta]\wedge(\nu_0+3))}} \right).$$

4 Numerical simulations

In this section, we study the numerical performances of our estimation procedure. We consider the density of the Beta(2, 2)-distribution and the density of the truncated normal distribution on $[0, 1]$ with mean $1/2$ and variance 0.25^2, respectively denoted h_1 and h_2. The density h_1 is proportional to $x(1-x)\mathbb{1}_{[0,1]}(x)$ and h_2 has the following form:

$$h_2(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)\right)}, \quad x \in [0, 1],$$

where $\mu = 0.5$, $\sigma = 0.25$ and $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the density and the cdf of the standard normal distribution. Furthermore, for all simulations we take $\alpha = 0.7$ and $R = 1$. Figures 1 and 2 show h_1, h_2 and their corresponding stationary densities N_1, N_2. The stationary densities are obtained by solving numerically the PDE (2.5) using the method presented in Doumic *et al.* [19].

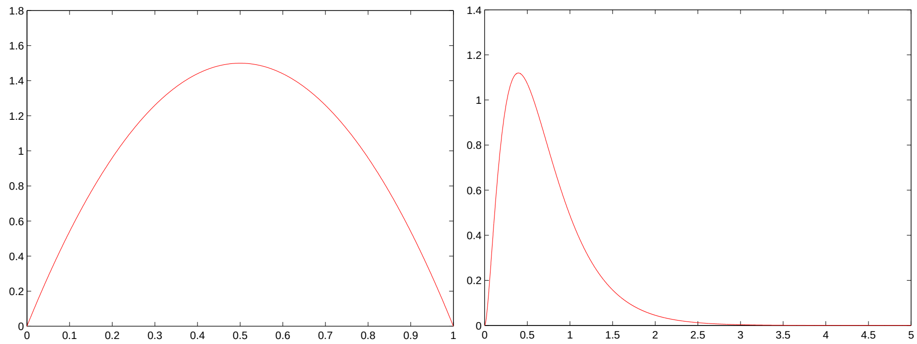

Figure 1: *The* Beta(2, 2) *density* h_1 *(left) and its corresponding stationary density* N_1 *(right).*

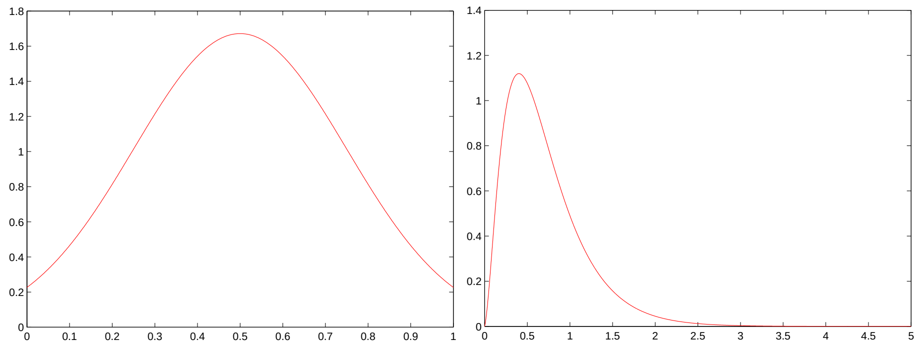

Figure 2: *The truncated normal density* h_2 *(left) and its corresponding stationary density* N_2 *(right).*

For the estimation of h_1 and h_2, even if theoretical boundary conditions stated in Assumption 1 are not satisfied, we shall observe that the procedure does a good job. Before presenting the numerical results, let us point out some difficulties that affect the quality of the estimation.

First, one can observe in Figures 1 and 2 that shapes of functions N_1 and N_2 are very similar although functions h_1 and h_2 are very different. This illustrates a major difficulty of our inverse problem and leads to some difficulties for the estimation of the densities g and h.

Secondly, in view of (3.3) and (3.6), the construction of the estimator $\hat{g}_\ell$ is based on the estimation of M^* and D^*. Remember that $D^*(\xi) = (-\mathbf{i}\xi)\mathbb{E}\big[e^{(\mathbf{i}\xi-1)U_1}\big]$ and the leading term $-\mathbf{i}\xi$ of the last expression, coming from the computation of the Fourier transform of the derivation function D, gives large fluctuations for the estimation of D^* when ξ takes large values. To justify this point, we introduce the modified formulas of D^* and $\widehat{D^*}$, denoted respectively by $\mathfrak{D}^*$ and $\widehat{\mathfrak{D}^*}$, obtained by removing $-\mathbf{i}\xi$ from the original formulas:

$$\mathfrak{D}^*(\xi) = \mathbb{E}\big[e^{(\mathbf{i}\xi-1)U_1}\big] \quad \text{and} \quad \widehat{\mathfrak{D}^*}(\xi) = \frac{1}{n}\sum_{j=1}^{n} e^{(\mathbf{i}\xi-1)U_j}.$$

Figures 3, 4 and 5 provide a reconstruction of $\widehat{M^*}$, $\widehat{D^*}$ and $\widehat{\mathfrak{D}^*}$ based on a random sample $U_1, \dots, U_n$ of size $n = 30000$ for h_1. For each figure, we represent both the real part and the imaginary part of $\widehat{M^*}$ (resp. $\widehat{D^*}$, $\widehat{\mathfrak{D}^*}$) and we compare them with those of M^* (resp. D^*, $\mathfrak{D}^*$). The Fourier transforms M^*, D^* and $\mathfrak{D}^*$ are computed directly from the function N_1, indicating that one can consider M^*, D^* and $\mathfrak{D}^*$ as the "true" functions. Figure 3 shows that the reconstruction of $\widehat{M^*}$ is very satisfying, whereas many oscillations in the reconstruction of $\widehat{D^*}$ appear (see Figure 4). These oscillations vanish for $\widehat{\mathfrak{D}^*}$ (see Figure 5). This confirms what we mentioned: the estimation of the derivative D^* has a strong influence for our statistical problem.

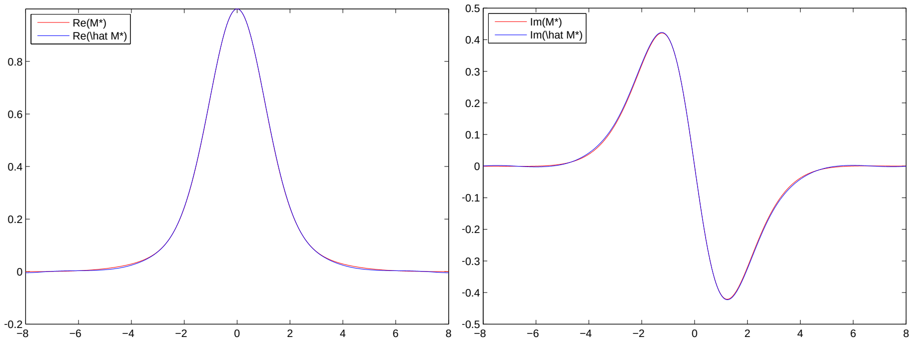

Figure 3: *For the* Beta(2, 2) *density, the real part (left) and the imaginary part (right) of* $\widehat{M^*}$ *(blue line) compared with those of* M^* *(red line).*

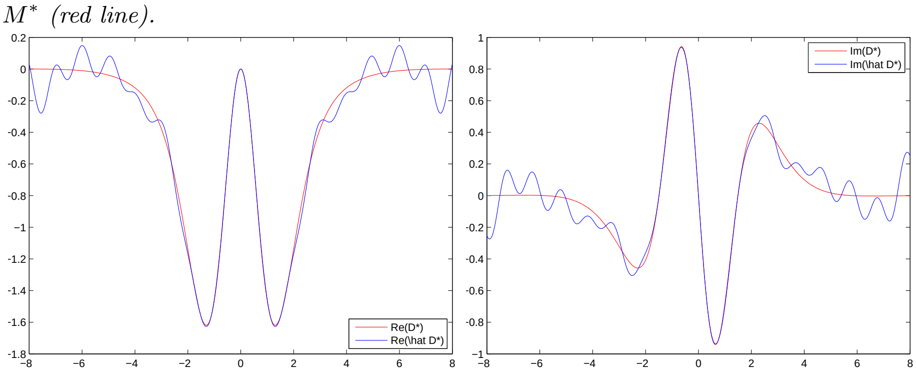

Figure 4: *For the* Beta(2, 2) *density, the real part (left) and the imaginary part (right) of* $\widehat{D^*}$ *(blue line) compared with those of* D^* *(red line).*

In the sequel, we introduce our bandwidth selection rules for the estimators $\hat{g}_\ell$ and $\hat{h}_\ell$, then we present some numerical results to illustrate the performances of our estimators.

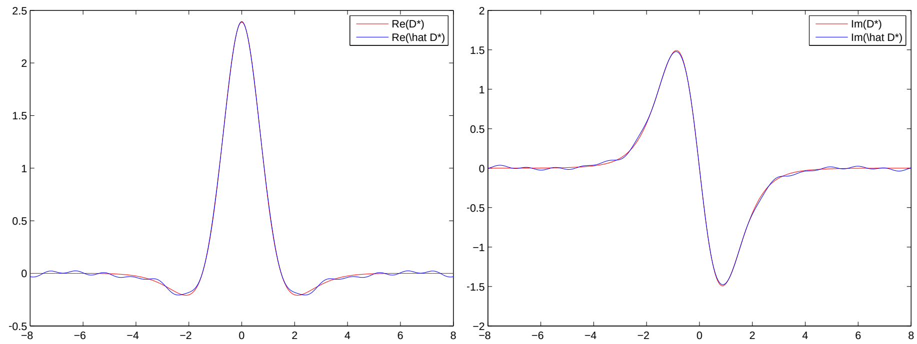

Figure 5: *For the* Beta(2, 2) *density, the real part (left) and the imaginary part (right) of* $\widehat{\mathfrak{D}}^*$ *(blue line) compared with those of* $\mathfrak{D}^*$ *(red line).*

4.1 Bandwidth selection rules

To establish a bandwidth section rule for the estimator $\hat{g}_\ell$ and $\hat{h}_\ell$, we use resampling techniques inspired from the principle of cross-validation. We first study the $\mathbb{L}^2$-risk of the estimator $\hat{g}_\ell$ in the Fourier domain:

$$\|\hat{g}_\ell - g\|_2^2 = \frac{1}{2\pi}\|\hat{g}_\ell^* - g^*\|_2^2 = \frac{1}{2\pi}\left(\|\hat{g}_\ell^*\|_2^2 - 2\langle \hat{g}_\ell^*, g^*\rangle\right) + \frac{1}{2\pi}\|g^*\|_2^2.$$

Define

$$J(\ell) := \|\hat{g}_\ell^*\|_2^2 - 2\langle \hat{g}_\ell^*, g^*\rangle$$

where the scalar product of two complex functions u and v is defined as

$$\langle u, v\rangle = \int_{\mathbb{R}} u(\xi)\overline{v(\xi)}d\xi.$$

Let $\mathcal{L}$ be a family of possible bandwidths, the optimal bandwidth is given by

$$\ell_{CV} := \underset{\ell\in\mathcal{L}}{\operatorname{argmin}}\, J(\ell) = \underset{\ell\in\mathcal{L}}{\operatorname{argmin}}\, \|\hat{g}_\ell - g\|_2^2.$$

We aim at constructing an estimator of $J(\ell)$, which is equivalent to providing an estimate of the scalar product $\langle \hat{g}_\ell^*, g\rangle$ since $\|\hat{g}_\ell^*\|_2^2$ is known. Instead of finding a closed formula for the estimator of the $\mathbb{L}^2$-risk which is intricate in our case, we use the following alternative approach: we start from a random sample and divide it into two disjoint sets, called the training set and the validation set. They are respectively used for computing the estimator and measuring its performance. For sake of simplicity, those sets have the same size. Let $\hat{g}_\ell^{*(t)}$ be the estimator of g^* constructed on the training set. The heuristics is that if $\hat{g}_{\ell'}^{*(v)}$ is an estimator constructed on the validation set, then $\langle \hat{g}_\ell^{*(t)}, \hat{g}_{\ell'}^{*(v)}\rangle$ gives us an estimate of $\langle \hat{g}_\ell^{*(t)}, g^*\rangle$ and subsequently an estimate of $J(\ell)$. The final bandwidth is the one which minimizes the average of all risk estimates computed over a number of couples of training-validation set selected from the same sample.

In detail, let $\{X_1, \ldots, X_n\}$ be a random sample. Let E and E^C be the subsets of $\{1, \ldots, n\}$ such that $|E| = n/2$ and $E^C = \{1, \ldots, n\} \setminus E$. We divide $\{X_1, \ldots, X_n\}$ into two sub-samples:

$$\mathbf{X}^E := (X_i)_{i\in E} \quad \text{and} \quad \mathbf{X}^{E^C} := (X_i)_{i\in E^C}.$$

There are $V_{\max}$ possibilities to select the subsets (E, E^c), where

$$V_{\max} := \binom{n}{n/2}.$$

If n is large then $V_{\max}$ will be huge. Hence we choose in practice a number V which is smaller than $V_{\max}$ to reduce computation time. We propose two criteria for the selection of bandwidths as follows.

Definition 3. *Let* $(E_j, E_j^C)_{1\le j\le V}$*,* $V \le V_{\max}$ *be the sequence of subsets selected from* $\{1,\dots,n\}$ *and the corresponding sub-samples* $(\mathbf{X}^{E_j}, \mathbf{X}^{E_j^C})_{1\le j\le V}$*. Let* $\hat{g}_\ell^{*(E_j)}$ *and* $\hat{g}_\ell^{*(E_j^c)}$ *be the estimators of* g_ℓ^* *respectively constructed on the sub-samples* $\mathbf{X}^{E_j}$ *and* $\mathbf{X}^{E_j^C}$*. Define*

$$\widehat{J}_{Crit1}(\ell) := \frac{1}{V}\sum_{j=1}^{V}\left[\left\|\hat{g}_\ell^{*(E_j)}\right\|_2^2 - 2\left\langle \hat{g}_\ell^{*(E_j)}, \hat{g}_\ell^{*(E_j^C)}\right\rangle\right]. \tag{4.1}$$

Then the selected bandwidth is given by

$$\hat{\ell}_{Crit1} := \operatorname*{argmin}_{\ell\in\mathcal{L}} \widehat{J}_{Crit1}(\ell). \tag{4.2}$$

Definition 4. *Let* $\hat{g}_\ell^{*(E_j)}$ *and* $\hat{g}_{\ell'}^{*(E_j^c)}$ *be the estimators of* g_ℓ^* *as in Definition 3. Define,*

$$\widehat{J}_{Crit2}(\ell,\ell') := \frac{1}{V}\sum_{j=1}^{V}\left[\left\|\hat{g}_\ell^{*(E_j)}\right\|_2^2 - 2\left\langle \hat{g}_\ell^{*(E_j)}, \hat{g}_{\ell'}^{*(E_j^C)}\right\rangle\right]. \tag{4.3}$$

Then an alternative bandwidth selection rule is given as follows:

$$\hat{\ell}_{Crit2} := \operatorname*{argmin}_{\ell\in\mathcal{L}} \left\{\min_{\ell'\in\mathcal{L}} \widehat{J}_{Crit2}(\ell,\ell')\right\}. \tag{4.4}$$

Note that the second criterion is more computationally intensive.

4.2 Numerical results

Remember that we aim at reconstructing the densities h_1 and h_2, *i.e.* the Beta$(2,2)$ density and the density of a truncated normal $\mathcal{N}(0.5, 0.25^2)$ on $[0,1]$. We apply formulas (3.6), (3.8) and (3.9) to construct the estimators for these densities. The bandwidth ℓ is chosen in the family $\mathcal{L} \subset \{1/(0.5\Delta),\ \Delta = 1,\dots,50\}$ according to two bandwidth selection rules. We compare the estimated densities when using our selection rules with those estimated with the oracle bandwidth. The oracle bandwidth is the optimal bandwidth obtained by assuming that we know the true density and defined as follows:

$$\ell_{\text{oracle}} := \operatorname*{argmin}_{\ell\in\mathcal{L}} \|\hat{g}_\ell - g\|_2^2.$$

Of course, ℓ_{oracle} and $\hat{g}_{\ell_{\text{oracle}}}$ cannot be used in practice (since they depend on the true function to estimate) but they can be viewed as benchmark quantities. For $n = 30000$ observations, we illustrate in Figures 6 and 7 the estimates of (g_1, h_1) and (g_2, h_2) using the first bandwidth selection rule (see Definition 3).

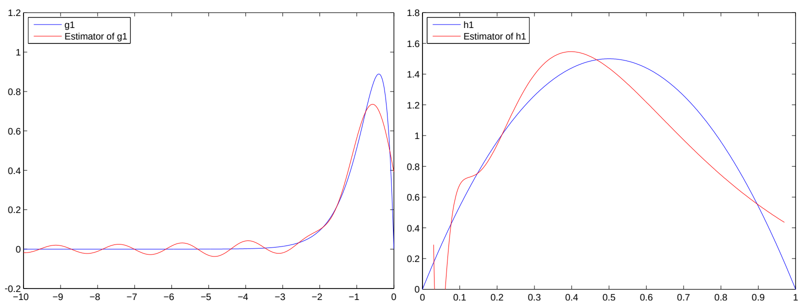

Figure 6: *Estimation of* $g_1(x) = e^x h_1(e^x)$ *and* h_1*.*

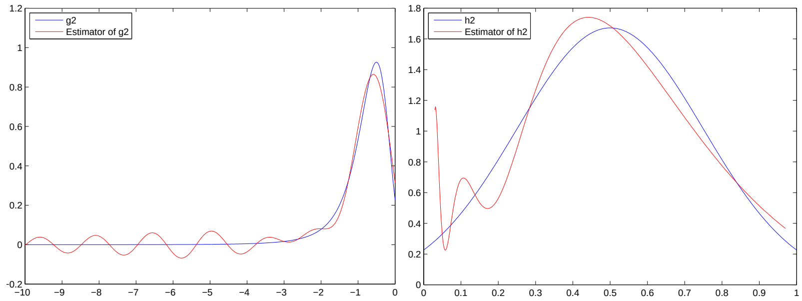

Figure 7: *Estimation of* $g_2(x) = e^x h_2(e^x)$ *and* h_2.

These graphs show bad behaviors when reconstructing h_1 and h_2 if we do not take into account the symmetry of theses densities. Considering symmetrization (see (3.11)) provides significative improvements (see Figure 8). Reconstructions of densities are quite satisfying except at boundaries of $[0, 1]$, which is expected in view of remarks of Section 3.2.1.

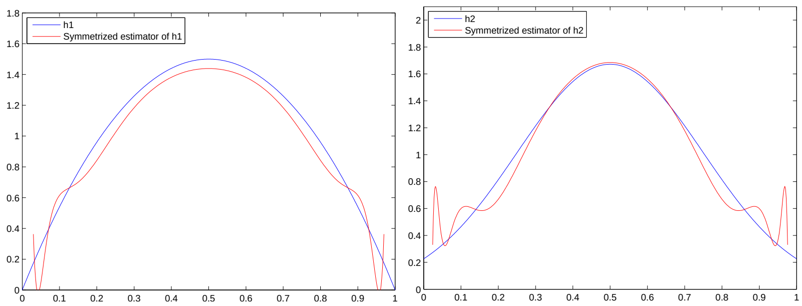

Figure 8: Reconstructions of h_1 (left) and h_2 (right) after symmetrization.

Table 1 shows the $\mathbb{L}^2$-risk of $\hat{g}_{\hat{\ell}_{Crit1}}$ and $\hat{g}_{\hat{\ell}_{Crit2}}$ where $\hat{\ell}_{Crit1}$ and $\hat{\ell}_{Crit2}$ are the bandwidths selected by our selection rules (see Definitions 3 and 4), over 100 Monte Carlo runs for estimating h_1 and h_2 with respect to $V = 10, 25$ and 40. The sample size for each repetition is $n = 30000$. We also provide associated Boxplots in Figure 9 and 10.

		h_1 - Beta$(2, 2)$			h_2 - Truncated normal		
		Crit1	Crit2	Oracle	Crit1	Crit2	Oracle
$V = 10$	$\bar{e}$	0.04155	0.04031	0.03056	0.03703	0.03669	0.02806
	$\bar{\hat{\ell}}$	0.29839	0.29606	0.27583	0.30255	0.30312	0.27858
$V = 25$	$\bar{e}$	0.04145	0.03898	0.03056	0.03679	0.03602	0.02806
	$\bar{\hat{\ell}}$	0.29732	0.29787	0.27583	0.30348	0.30155	0.27858
$V = 40$	$\bar{e}$	0.04039	0.03708	0.03056	0.03613	0.03440	0.02806
	$\bar{\hat{\ell}}$	0.29837	0.29985	0.27583	0.30396	0.30303	0.27858

Table 1: *Average of the* $\mathbb{L}^2$*-risk of* $\hat{g}_{\hat{\ell}_{Crit1}}$ *and* $\hat{g}_{\hat{\ell}_{Crit2}}$ *over* 100 *Monte Carlo repetitions for estimating* h_1 *and* h_2*, compared with those of the oracle.*

Table 1 and boxplots show that the performances of our estimators are close to those of the oracle. When comparing the first bandwidth selection rule Crit1 with the second one Crit2, one can observe that the performances of Crit2 are slightly better than those of Crit1 (see Table 1). However, Crit2 is more time-consuming than Crit1. For both selection rules, we observe that the performances are slightly better when we increase the number of selected subsamples V. Remember that the larger the value of V, the larger the computation time whereas the performances are improved marginally. Hence, in practice it is reasonable to choose the first bandwidth selection rule Crit1 with $V = 10$.

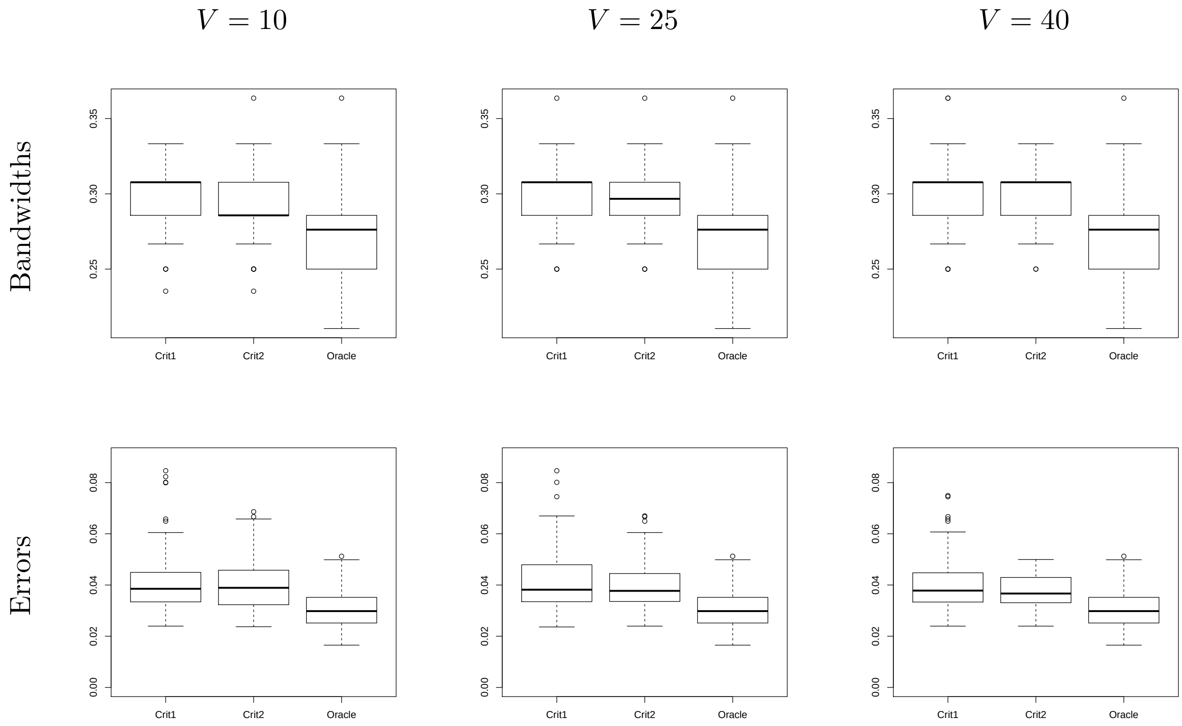

Figure 9: *Bandwidths and errors for the estimation of* h_1 *(*$\mathrm{Beta}(2,2)$ *distribution).*

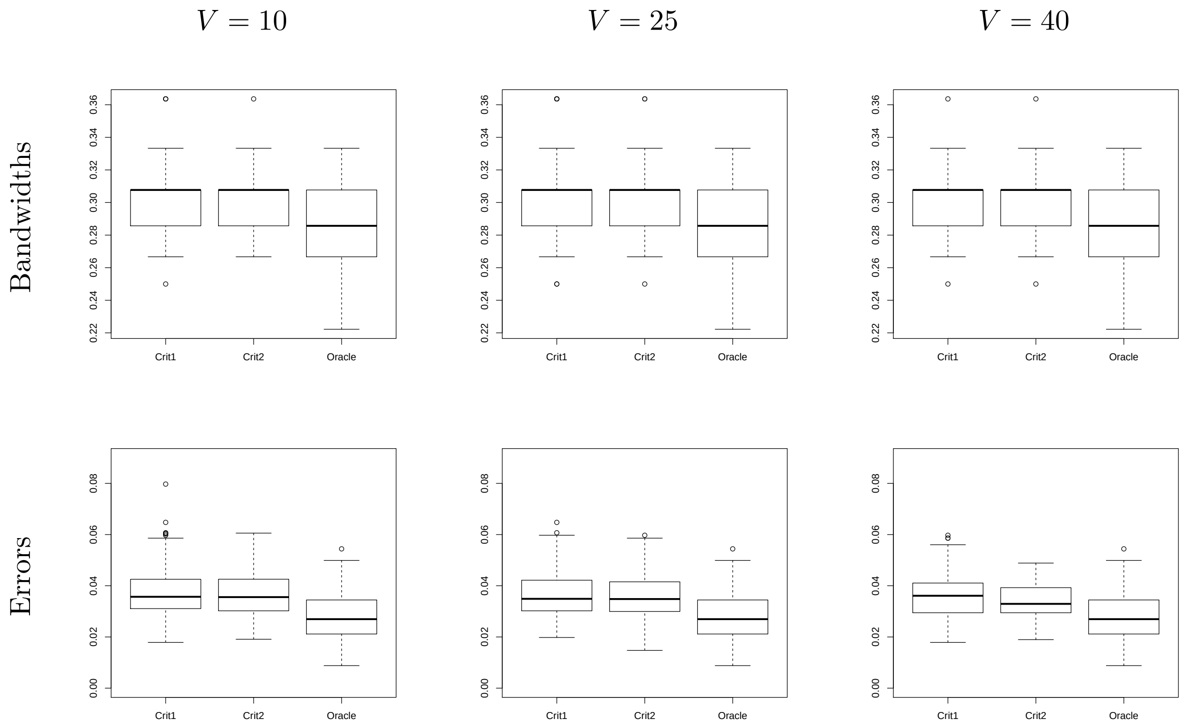

Figure 10: *Bandwidths and errors for the estimation of* h_2 *(Truncated normal).*

Appendix

This section is devoted to the proofs of the paper's results. C is a constant whose value may change from line to line.

A Large population renormalization

Before proving the results of Section 2, let us build the SDE satisfied by the process $(Z_t^K)_{t\geq 0}$. Consider

$$\widetilde{Z}_t^K = \frac{1}{K}\sum_{i\in V_t^K}\delta_{(i,x_i(t))}$$

the random point measure on $\mathcal{I}\times\mathbb{R}_+$ with marginal measure Z_t^K on $\mathbb{R}_+$, and that keeps track of the sizes and labels of the individuals in the population.

Let us consider as in Section 2 a sequence $(\widetilde{Z}_0^K)_{K\in\mathbb{N}^*}$ of random point measures on $\mathcal{I}\times\mathbb{R}_+$ such that the sequence of marginal measures $(Z_0^K)_{K\in\mathbb{N}^*}$ of the form (1.1) converges to $\xi_0\in\mathcal{M}_F(\mathbb{R}_+)$ in probability and for the weak convergence topology on $\mathcal{M}_F(\mathbb{R}_+)$ and satisfies (2.1). Let also $Q(ds,di,d\gamma)$ be a Poisson point measure on $\mathbb{R}_+\times\mathcal{E}:=\mathbb{R}_+\times\mathcal{I}\times[0,1]$ with intensity $q(ds,di,d\gamma)=Rds\,n(di)\,h(\gamma)d\gamma$ where $n(di)$ is the counting measure on $\mathcal{I}$ and ds and $d\gamma$ are Lebesgue measures on $\mathbb{R}_+$.
We denote $\{\mathcal{F}_t\}_{t\geq 0}$ the canonical filtration associated with the Poisson point measure and the sequence $(\widetilde{Z}_0^K)_{K\in\mathbb{N}^*}$.

For a given $K\in\mathbb{N}^*$, it is possible to describe the measure $\widetilde{Z}_t^K$ at time t by the following equation:

$$\begin{aligned}\widetilde{Z}_t^K = &\sum_{i\in V_0^K}\delta_{(i,x_i(0)+\alpha t)}\\ &+\int_0^t\int_E \mathbb{1}_{i\in V_{s_-}^K}\big(\delta_{(i0,\gamma x_i(s_-)+\alpha(t-s))}+\delta_{(i1,(1-\gamma)x_i(s_-)+\alpha(t-s))}-\delta_{(i,x_i(s_-))}\big)Q(ds,di,d\gamma),\end{aligned} \tag{A.1}$$

where the notation $x_i(s)$ stands for the size of the individual with label i in the population Z_s^K (we omit the dependence in K). This representation allows to take deterministic motions into account and the idea comes from [39, 33]: we build the population at time t by considering the contribution of the initial condition for this time t, and then the modifications due to all the divisions between times 0 and t. The first term in the r.h.s. of (A.1) corresponds to the individuals alive at time 0 with their sizes at time t if they don't die. In the integral with respect to the Poisson point process, an atom at (s,i,γ) of Q corresponds to a 'virtual' division event at time s of the individual i associated with the fraction γ. This event effectively takes place only if the individual with label i is alive at time s_-. In this case, the Dirac masses corresponding to the mother at t (at size $x_i(s_-)+t-s$) is replaced with the Dirac masses of the two daughters, at the size that they will have if they are still alive at time t ($\gamma x_i(s_-)+\alpha(t-s)$ and $(1-\gamma)x_i(s_-)+\alpha(t-s)$).

The moment assumption (2.1) propagates to positive time and it is possible to show that for any $T>0$, (see [24, Prop.3.2.5])

$$\sup_{K\in\mathbb{N}^*}\mathbb{E}\big(\sup_{t\in[0,T]}\langle Z_t^K,1\rangle^2\big)<+\infty.$$

For every $K \in \mathbb{N}^*$ and every test function $f_s(x) = f(x,s) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$, the stochastic process $(Z_t^K)_{t\in\mathbb{R}_+}$ satisfies:

$$\begin{aligned}
\langle Z_t^K, f\rangle =& \langle Z_0^K, f\rangle + \int_0^t \int_{\mathbb{R}_+} \big(\partial_s f_s(x) + \alpha \partial_x f_s(x)\big) Z_s^K(dx)ds \\
&+ \frac{1}{K}\int_0^t \int_E \mathbb{1}_{\{i\in V_{s-}^K\}} \Big(f_s\big(\gamma x_i(s-)\big) + f_s\big((1-\gamma)x_i(s_-)\big) - f_s\big(x_i(s_-)\big)\Big) Q(ds,di,d\gamma), \\
=& \langle Z_0^K, f\rangle + M_t^{K,f} \\
&+ \int_0^t \int_{\mathbb{R}_+} \Big(\partial_s f_s(x) + \alpha\partial_x f_s(x) + R\int_0^1 \big(f(\gamma x) + f((1-\gamma)x) - f(x)\big) h(\gamma)d\gamma\Big) Z_s^K(dx)ds,
\end{aligned} \tag{A.2}$$

where $(M_t^{K,f})_{t\geq 0}$ is a square integrable martingale started at 0 with bracket:

$$\langle M^{K,f}\rangle_t = \frac{1}{K}\int_0^t \int_{\mathbb{R}_+} R\big(f(\gamma x) + f((1-\gamma)x) - f(x)\big)^2 h(\gamma)d\gamma Z_s^K(dx)ds. \tag{A.3}$$

The proof of Proposition 1 then follows the ideas in [39, 38] and are detailed in [24]. Equation (A.2) corresponds to Equation 2.2 in the main body.

The proof of Theorem 1 uses the martingale problem established in Prop. 1 and standard arguments (see *e.g.* [20, 28, 4] and [40, Th.1.1.8 and proof of Th.1.1.11]). Let us denote by $A^{K,f}$ the finite variation part of $Z^{K,f}$:

$$A_t^{K,f} = \int_0^t \int_{\mathbb{R}_+} \Big(\partial_s f_s(x) + \alpha\partial_x f_s(x) + R\int_0^1 \big(f(\gamma x) + f((1-\gamma)x) - f(x)\big)h(\gamma)d\gamma\Big) Z_s^K(dx)ds. \tag{A.4}$$

First, using the moment assumptions together with (A.2)-(A.3), we can show that the sequences of real valued processes $(\langle A^{K,f}\rangle)_{K\in\mathbb{N}^*}$ and $(\langle M^{K,f}\rangle)_{K\in\mathbb{N}^*}$ are tight in $\mathbb{D}(\mathbb{R}_+,\mathbb{R})$, which by the Aldous-Rebolledo condition imply the tightness of the sequence $(\langle Z_.^K, f\rangle)_{K\in\mathbb{N}^*}$ for all test function $f \in \mathcal{C}_b^1(\mathbb{R}_+,\mathbb{R})$. As a consequence, the sequence $(Z^K)_{K\in\mathbb{N}^*}$ is tight in $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(\mathbb{R}_+), v))$, where $(\mathcal{M}_F(\mathbb{R}_+), v)$ means that the space of finite positive measures $\mathcal{M}_F(\mathbb{R}_+)$ is embedded with the topology of vague convergence.
Secondly, the limiting values $\bar{Z}$ to which subsequences of $(Z^K)_{K\in\mathbb{N}^*}$ converge vaguely, are continuous measure-valued processes of $\mathcal{C}(\mathbb{R}_+, (\mathcal{M}_F(\mathbb{R}_+), w))$, where $\mathcal{M}_F(\mathbb{R}_+)$ is embedded with the weak convergence topology.
Thirdly, proceeding as in [40, proof of Th.1.1.11] (see also [29, 32]), we can prove that

$$\lim_{k\to+\infty} \lim_{K\to+\infty} \mathbb{E}\big(\sup_{t\leq T} \langle Z_t^K, \varphi_k\rangle\big) = 0,$$

where the functions φ_k are $\mathcal{C}^2$ approximations of $\mathbb{1}_{x\geq k}$ for $k \in \mathbb{N}$ and are defined by $\varphi_0(x) = 1$ and for all $k \in \mathbb{N}^*$, $\varphi_k(x) = \psi(0 \vee (x-k+1) \wedge 1)$ with $\psi(x) = 6x^5 - 15x^4 + 10x^3$. This ensures that for every subsequence of $(Z^K)_{K\in\mathbb{N}^*}$ that converges vaguely to a limiting process $\bar{Z}$, their masses converge in distribution to $\langle \bar{Z}, 1\rangle$, which provides the tightness in $(\mathcal{M}_F(\mathbb{R}_+), w)$ by a criterion due to Méléard and Roelly [31].
We can now establish that the limiting values to which subsequences of $(Z^K)_{K\in\mathbb{N}^*}$ converge in $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(\mathbb{R}_+), w))$ are solutions of (2.4) (see [24]). This integro-differential equation admits a unique solution. Indeed, let ξ^1 and ξ^2 be two solutions of (2.4) starting with the same initial condition ξ_0. For a test function $\varphi \in \mathcal{C}_b^1(\mathbb{R}_+,\mathbb{R})$ and $t > 0$, setting

$$f(x,s) = f_s(x) = \varphi(x + \alpha(t-s)), \tag{A.5}$$

we obtain that for $i \in \{1,2\}$,

$$\langle \xi_t^i, \varphi\rangle = \langle \xi_0, \varphi(.+\alpha t)\rangle + \int_0^t \int_{\mathbb{R}_+} \int_0^1 R\big(f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\big) h(\gamma) d\gamma\, \xi_s^i(dx)\, ds.$$

Substracting these two equations for $i=1$ and $i=2$, we obtain

$$\|\xi_t^1 - \xi_t^2\|_{TV} \leq 3R\|\varphi\|_\infty \int_0^t \|\xi_s^1 - \xi_s^2\|_{TV} ds$$

where $\|.\|_{TV}$ stands for the total variation norm. Gronwall's inequality concludes the proof of uniqueness of the solution of (2.4). Since the limiting value of $(Z^K)_{K\in\mathbb{N}^*}$ is unique, the sequence hence converges in $\mathbb{D}(\mathbb{R}_+, (\mathcal{M}_F(\mathbb{R}_+), w))$ to this unique solution. This concludes the proof of Theorem 1.

The proof of Proposition 2 is detailed in [24] (see also [39]). First, notice that if $\xi_0(dx)$ admits a density $n_0(x)$ with respect to the Lebesgue measure, then for any $t > 0$, ξ_t also admits a density. Indeed, for a function $\varphi \in \mathcal{C}^1(\mathbb{R}_+, \mathbb{R}_+)$ with non-negative values, let us define the test function $f(x,s)$ as in (A.5). Then, neglecting the negative terms in the second line of (2.4) and using the symmetry of h with respect to $1/2$:

$$\begin{aligned}
\langle \xi_t, \varphi\rangle \leq & \int_{\mathbb{R}_+} \varphi(x+\alpha t) n_0(x) dx + 2R\int_0^t \int_{\mathbb{R}_+}\int_0^1 \varphi(\gamma x + \alpha(t-s)) h(\gamma) d\gamma\, \xi_s(dx)\, ds \\
= & \int_{\alpha t}^{+\infty} \varphi(y) n_0(y-\alpha t) dy + 2R\int_0^t \varphi(\alpha(t-s)) \xi_s(\{0\})\, ds \\
& \qquad + 2R\int_0^t \int_{\mathbb{R}_+\setminus\{0\}} \int_{\mathbb{R}} \mathbb{1}_{(\alpha(t-s), x+\alpha(t-s))}(y) \varphi(y) h\big(\frac{y-\alpha(t-s)}{x}\big) \frac{dy}{x}\, \xi_s(dx)\, ds \\
= & \int_{\alpha t}^{+\infty} \varphi(y) n_0(y-\alpha t) dy + 2R\int_0^{\alpha t} \varphi(y) \xi_{t-\frac{y}{\alpha}}(\{0\})\, \frac{dy}{\alpha} \\
& \qquad + 2R\int_0^{+\infty} \Big\{ \int_0^t \int_{\mathbb{R}_+\setminus\{0\}} \mathbb{1}_{(\alpha(t-s), x+\alpha(t-s))}(y) \frac{1}{x} h\big(\frac{y-\alpha(t-s)}{x}\big) \xi_s(dx)\, ds \Big\} \varphi(y) dy.
\end{aligned}$$

Since ξ_t is dominated by a nonnegative measure absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$ it follows that ξ_t admits itself a density. Denoting by $n(x,t)$ the density of ξ_t with respect to the Lebesgue measure dx on $\mathbb{R}_+$, we see that $(n(x,t), x\in\mathbb{R}_+, t>0)$ solves in distribution sense (2.5) for which uniqueness of the solution holds (*e.g.* [37, Th.4.3 p.90]).

The proof of Proposition 3 is a particular case of [37, Th.4.6 p. 94] based on Krein-Rutman theorem (*e.g.* [37, Th.6.5 p.175]) (see also [16]). In the case that we consider, the proof can be simplified compared with [37].
Let us consider the eigenelements (λ, N, ϕ) associated with (2.5), *i.e.* the solution of:

$$\left\{ \begin{array}{l}
\alpha \partial_x N(x) + (\lambda + R) N(x) = 2R\int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma}, \quad x \geq 0, \\
N(0) = 0, \quad \int N(x) dx = 1, \quad N(x) \geq 0, \quad \lambda > 0, \\
\alpha \partial_x \phi(x) - (\lambda + R)\phi(x) = -2R\int_0^1 \phi(\gamma x) h(\gamma) d\gamma, \quad x \geq 0, \\
\phi(x) \geq 0, \quad \int_0^{+\infty} \phi(x) N(x) dx = 1.
\end{array}\right. \tag{A.6}$$

It is clear that $\lambda = R$ and $\phi \equiv 1$ solve the third equation of (A.6). Because the first line is linear in N, we can forget for the proof the condition $\int N(x)dx = 1$: if there exists a nonnegative integrable solution, we can renormalize it.

Step 1: Let us consider the following auxiliary PDE, for a constant $\mu > 0$ and two functions $f \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$, and $M \in L^1(\mathbb{R}_+, \mathbb{R}) \cap \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$:

$$\alpha \partial_x N(x) + (\mu + R)\, N(x) - 2R \int_0^1 M\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = f(x), \quad 0 \le x \quad ; \quad N(0) = 0. \tag{A.7}$$

Equation (A.7) is a first order ODE that can be solved with the variation of constant method. It admits a unique solution, that we denote $T(M)$:

$$T(M)(x) = \frac{1}{\alpha} \int_0^x e^{-\frac{\mu+R}{\alpha}(x-y)} \Big(2R \int_0^1 M\left(\frac{y}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} + f(y) \Big) dy.$$

Consider M_1 an $M_2 \in L^1(\mathbb{R}_+, \mathbb{R}) \cap \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$. Then, for $x \ge 0$:

$$\begin{aligned} |T(M_1)(x) - T(M_2)(x)| \le & \frac{2R}{\alpha} \int_0^1 \int_0^{x/\gamma} e^{-\frac{\mu+R}{\alpha}(x-\gamma z)} |M_1(z) - M_2(z)| h(\gamma) dz\, d\gamma \\ \le & \frac{2R}{\mu + R} \big(1 - e^{-\frac{(\mu+R)x}{\alpha}}\big) \int_0^1 \frac{h(\gamma)}{\gamma} d\gamma \|M_1 - M_2\|_\infty. \end{aligned} \tag{A.8}$$

Provided the integral in the term above is finite, then for $\mu > 2R \int_0^1 h(\gamma)/\gamma\, d\gamma - R$, the map $M \in L^1(\mathbb{R}_+, \mathbb{R}) \cap \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+) \mapsto T(M) \in L^1(\mathbb{R}_+, \mathbb{R}) \cap \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+)$ is a contraction. Thus it admits a unique fixed point that is the unique solution of

$$\alpha \partial_x N(x) + (\mu + R)\, N(x) - 2R \int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = f(x), \quad 0 \le x \quad ; \quad N(0) = 0. \tag{A.9}$$

Step 2: The map A that associates to $f \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}_+) \cap L^1(\mathbb{R}_+, \mathbb{R}_+)$ the unique corresponding solution of (A.9) is thus well defined. Following the path of [37, Section 6.6.2], we can show that this map is linear, continuous (with computation similar to (A.8)) and strongly positive. Finally, the boundedness of N implies the boundedness of $\partial_x N$, with norms controlled by $\|f\|_\infty$. This allows to use Arzela-Ascoli theorem to obtain the compactness of the map A. We can then use Krein-Rutman theorem to obtain that the spectral radius of A, $\rho(A)$, is a positive simple eigenvalue associated with a positive eigenvector satisfying:

$$\alpha \partial_x N(x) + \big(\mu + R - \frac{1}{\rho(A)}\big)\, N(x) - 2R \int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = 0, \quad 0 \le x \quad ; \quad N(0) = 0. \tag{A.10}$$

The fact that $\lambda := \mu + R - \frac{1}{\rho(A)}$ is equal to $2R$ is a consequence of integrating the direct equation against the the adjoint eigenvector (here $\phi \equiv 1$) and using that $\int N(x) dx = 1$.

Step 3: The computation to establish the speed of convergence of $n(t, x) e^{-Rt}$ to $\rho N(x)$ stated in (2.8), are obtained by generalizing the proof of [37, Th.4.2 p.88] (see also [36]). Define $g(t, x) = n(x, t) e^{-Rt} - \rho N(x)$, $G(t, x) = \int_0^x g(t, y) dy$ and $K(t, x) = \partial_t G(t, x)$. One can write the PDEs satisfied by g and G. The PDE for G implies that $\partial_t \int_0^{+\infty} \big|G(t, x) e^{Rt}\big| dx \le 0$. As a consequence,

$$\int_0^{+\infty} |G(t, x)| dx \le e^{-Rt} \|G_0\|_1. \tag{A.11}$$

From the PDE of g, $K(0, x) = \partial_t G(t, x)|_{t=0} = 2R \int_0^1 G_0(x/u) h(u) du - 2R G_0(x) - \alpha g_0(x)$. Proceeding similarly as for G, we show that

$$\int_0^{+\infty} |K(t, x)| dx \le e^{-Rt} \int_0^{+\infty} |K(0, x)| dx \le e^{-Rt} \big(3R \|G_0\|_1 + \alpha \|g_0\|_1\big). \tag{A.12}$$

Plugging (A.11) and (A.12) in the PDE of g (where we notice that $g(t, x) = \partial_x G(t, x)$), we obtain the result announced in the proposition.

B Proof of Lemma 1

Proof of Lemma 1 (i). Let $\epsilon > 0$ to be chosen small enough, we have for $\nu \le (\nu_0+2)\wedge([\beta]+1)$:

$$\begin{aligned}\int_0^{+\infty} x^{-\nu}N(x)dx &= \int_0^{\epsilon} x^{-\nu}N(x)dx + \int_\epsilon^{+\infty} x^{-\nu}N(x)dx\\ &\le \int_0^{\epsilon} x^{-\nu}N(x)dx + \frac{1}{\epsilon^\nu}\int_\epsilon^{+\infty} N(x)dx\\ &\le \int_0^{\epsilon} x^{-\nu}N(x)dx + \frac{1}{\epsilon^\nu}.\end{aligned}$$

Hence, it remains to prove

$$\int_0^{\epsilon} x^{-\nu}N(x)dx < +\infty.$$

We follow and adapt the steps of the proof of Theorem 1 in Doumic and Gabriel [16]. Integrating both side of equation (2.7) between 0 and $x_0 \le x$, we get:

$$\alpha N(x_0) + 2R\int_0^{x_0} N(y)dy = 2R\int_0^{x_0}\int_0^{+\infty} N(y)h\left(\frac{z}{y}\right)\frac{dy}{y}dz. \tag{B.1}$$

Thus,

$$\alpha N(x_0) \le 2R\int_0^{x_0}\int_0^{+\infty} N(y)h\left(\frac{z}{y}\right)\frac{dy}{y}dz \le 2R\int_0^{x}\int_0^{+\infty} N(y)h\left(\frac{z}{y}\right)\frac{dy}{y}dz.$$

Let us define:

$$f : x \mapsto \sup_{x_0\in(0,x]} N(x_0),$$

then we have for all x

$$f(x) \le \frac{2R}{\alpha}\int_0^{x}\int_0^{+\infty} N(y)h\left(\frac{z}{y}\right)\frac{dy}{y}dz. \tag{B.2}$$

Recall Assumption 1. Using a Taylor expansion, it implies that for any $t \in (0,1)$,

$$\int_0^t h(x)dx \le C\int_0^t x^{(\nu_0+1)\wedge[\beta]}dx \le Ct^{(\nu_0+2)\wedge([\beta]+1)} \le Ct^\nu \tag{B.3}$$

by choice of $\nu \le (\nu_0+2)\wedge([\beta]+1)$. Then, we have for all $x < \epsilon$:

$$\begin{aligned}f(x) &\le \frac{2R}{\alpha}\int_0^{+\infty} N(y)dy\int_0^x h\left(\frac{z}{y}\right)\frac{dz}{y}\\ &\le \frac{2R}{\alpha}\int_0^{+\infty} N(y)\min\left(1, C\frac{x^\nu}{y^\nu}\right)dy\\ &\le \frac{2R}{\alpha}\left(\int_0^x N(y)dy + C\int_x^\epsilon N(y)\frac{x^\nu}{y^\nu}dy + C\int_\epsilon^{+\infty} N(y)\frac{x^\nu}{y^\nu}dy\right)\\ &\le \frac{2R}{\alpha}\left(\int_0^x \sup_{z\in(0,x]} N(z)dy + Cx^\nu\int_x^\epsilon \sup_{z\in(0,y]} N(z)\frac{dy}{y^\nu}\right) + \left(\frac{2CR}{\alpha}\int_\epsilon^{+\infty}\frac{N(y)}{y^\nu}dy\right)x^\nu\\ &\le \frac{2R\epsilon}{\alpha}f(x) + \frac{2CRx^\nu}{\alpha}\int_x^\epsilon \frac{f(y)}{y^\nu}dy + Kx^\nu,\end{aligned}$$

with $K = \dfrac{2CR}{\alpha\epsilon^\nu}$. We choose ϵ such that

$$0 < \epsilon < \frac{\alpha}{2R}.$$

and by setting $F(x) = x^{-\nu} f(x)$, we get

$$F(x) \leq \frac{K}{1 - \frac{2R\epsilon}{\alpha}} + \frac{2CR}{\alpha - 2R\epsilon} \int_x^\epsilon F(y) dy. \tag{B.4}$$

Then, applying Gronwall's inequality to (B.4), we obtain

$$F(x) \leq \frac{K}{1 - \frac{2R\epsilon}{\alpha}} \exp\left(\frac{2CR\epsilon}{\alpha - 2R\epsilon}\right) =: \tilde{C}, \quad \forall x \in [0, \epsilon]$$

and

$$x^{-\nu} N(x) \leq \tilde{C}, \quad \forall x \in [0, \epsilon].$$

We finally obtain

$$\int_0^\epsilon x^{-\nu} N(x) dx \leq \tilde{C}\epsilon < +\infty.$$

This ends the proof of Lemma 1(i). □

Proof of Lemma 1(ii). Let us first notice that by the fixed point theorem in the proof of Proposition 3, N is continuous as uniform limit of a sequence of continuous functions. Let us show that under Assumption 1 the map

$$\Phi \,:\, x \in (0, +\infty) \mapsto \int_x^{+\infty} N(y) h\Big(\frac{x}{y}\Big) \frac{dy}{y}$$

is of class $\mathcal{C}^{[\beta]}$ on $(0, +\infty)$. We proceed by induction, and start by computing the first derivative of Φ for $x > 0$.

$$\begin{aligned} \frac{\Phi(x+\varepsilon) - \Phi(x)}{\varepsilon} &= -\frac{1}{\varepsilon} \int_x^{x+\varepsilon} N(y) h\Big(\frac{x}{y}\Big) \frac{dy}{y} + \int_{x+\varepsilon}^{+\infty} N(y) \frac{y}{\varepsilon} \Big[h\Big(\frac{x+\varepsilon}{y}\Big) - h\Big(\frac{x}{y}\Big)\Big] \frac{dy}{y^2} \\ &\to_{\varepsilon \to 0} -\frac{N(x)h(1)}{x} + \int_x^{+\infty} N(y) h'\Big(\frac{x}{y}\Big) \frac{dy}{y^2} = \int_x^{+\infty} N(y) h'\Big(\frac{x}{y}\Big) \frac{dy}{y^2}, \end{aligned} \tag{B.5}$$

since $h(1) = h(0) = 0$ by Assumption 1. This shows that Φ is of class $\mathcal{C}^1$. Plugging this information into (2.7), it follows that $\partial_x N$ is continuous, and hence N is of class $\mathcal{C}^1$, which itself entails from the computation of Φ' that Φ is of class $\mathcal{C}^2$.
Suppose that we have computed the successive derivatives of Φ up to $k-1$ and that we have shown that N is of class $\mathcal{C}^{k-1}$ for $k \leq [\beta] \wedge \nu_0$. Then, since the successive derivatives of h at 0 vanish by Assumption 1,

$$\Phi^{(k)}(x) = \int_x^{+\infty} N(y) h^{(k)}\Big(\frac{x}{y}\Big) \frac{dy}{y^{k+1}}.$$

Since N, h and their derivatives are bounded functions, the latter integrals are always finite for $x > 0$. This implies that Φ is of class $\mathcal{C}^k$ and that using this information in (2.7), $\partial_x N$ is of class $\mathcal{C}^{k-1}$ entailing that N is of class $\mathcal{C}^k$. As the computation of the first derivative of Φ shows, we are limited by the regularity of h.
So we finally have that $x \mapsto N(x)$ is of class $\mathcal{C}^{[\beta]}$, and thus, $u \mapsto M(u)$ is also of class $\mathcal{C}^{[\beta]}$.

Take $k \leq [\beta]$. That M is of class $\mathcal{C}^{[\beta]}$ implies that $(\mathbf{i}\xi)^k M^*(\xi)$ is the Fourier transform of $M^{(k)}$ and bounded on $\mathbb{R}$ provided we additionally prove that the derivatives of M up to the order k are integrable. Since $M(u) = e^u N(e^u)$, $M^{(k)}$ is a linear combination of terms of the form $e^{(\ell+1)u} N^{(\ell)}(e^u)$ with $\ell \leq k$. We thus have to check the finiteness, for all $\ell \leq k$, of:

$$\int_{\mathbb{R}} e^{(\ell+1)u} \big|N^{(\ell)}(e^u)\big| du = \int_0^{+\infty} v^\ell \big|N^{(\ell)}(v)\big| dv. \tag{B.6}$$

It is known (as a direct adaptation of [37, Th.4.6 p.95] for example) that $N(x)e^{\mu x} \in L^1 \cap L^\infty(\mathbb{R}_+,\mathbb{R}_+)$ as soon as $\mu < R/\alpha$. Assume that for some $\ell < k$, we have proved that $\int_0^{+\infty} e^{\mu x}\big|N^{(\ell)}(x)\big|dx < +\infty$ for $\mu < R/\alpha$. Let us prove that this also holds for $\ell+1$, which would entail (B.6). Deriving (2.7) ℓ times, multiplying by $e^{\mu x}$ and integrating again in $x \in (0,+\infty)$, we obtain:

$$\begin{aligned}\alpha\int_0^{+\infty} e^{\mu x}\big|N^{(\ell+1)}(x)\big|dx \leq & 2R\int_0^{+\infty} e^{\mu x}\big|N^{(\ell)}(x)\big|dx + 2R\int_0^{+\infty} e^{\mu x}\Big|\int_x^{+\infty} h^{(\ell)}\Big(\frac{x}{y}\Big)\frac{N(y)}{y^{\ell+1}}dy\Big|dx\\ \leq & 2R\int_0^{+\infty} e^{\mu x}\big|N^{(\ell)}(x)\big|dx + 2R\int_0^{+\infty}\frac{N(y)e^{\mu y}}{y^\ell}\|h^{(\ell)}\|_1 dy. \qquad \text{(B.7)}\end{aligned}$$

By the induction assumption, the first term in the right hand side is finite. Because $h^{(\ell)}$ is a continuous function on $[0,1]$, $\|h^{(\ell)}\|_1$ is finite. That $N(x)e^{\mu x} \in L^1(\mathbb{R}_+,\mathbb{R}_+)$ implies that the second term is integrable at $+\infty$. Point (i) of Lemma 1 ensures the integrability at 0. Thus, the right hand side of (B.7) is finite. The use of point (i) of Lemma 1 explains why $[\beta]\wedge(\nu_0+3)$ appears in the announced result.
The finiteness of $\int_0^{+\infty} e^{\mu x}\big|N^{(\ell)}(x)\big|dx$, for $\ell \leq [\beta]\wedge(\nu_0+3)$, is thus proved by recursion, implying the finiteness of the terms in (B.6) and concluding the proof. $\square$

Proof of Lemma 1(iii). Let us consider the application

$$\Phi\ :\ \xi = \xi_1 + \mathbf{i}\xi_2 \in \mathbb{C} \mapsto \int_{-\infty}^{+\infty} e^{\mathbf{i}x\xi}M(x)dx = \int_0^{+\infty} e^{\mathbf{i}\xi\log(y)}N(y)dy.$$

Because N is such that $e^{\mu x}N(x) \in L^\infty(\mathbb{R}_+,\mathbb{R}_+)\cap L^1(\mathbb{R}_+,\mathbb{R}_+)$ for $\mu < R/\alpha$ (see [37, p.95]), Φ is well defined on $\mathbb{C}$. The derivative of the integrand with respect to ξ has modulus $|\log(y)|\,N(y)$ that is upper bounded when y is close to zero by $\frac{N(y)}{y}$, which is integrable on the neighborhood of zero by Lemma 1(i). It follows from the results on integrals with parameters that the extension of $\xi \mapsto M^*(\xi)$ to the complex plane is holomorphic on $\mathbb{C}$. Because M^* is not the null function, its zeros have to be isolated.

To show that Φ admits no zero on the real line, we use the argument principle (see [23, Section 4.10]). Let Γ be a positively oriented Jordan contour. If there are J zeros of Φ inside Γ, with multiplicities $m_1,\dots,m_J$, then

$$\sum_{j=1}^J m_j = \frac{1}{2\pi\mathbf{i}}\int_\Gamma \frac{\Phi'(\xi)}{\Phi(\xi)}d\xi.$$

Let $A > 0$ and $\varepsilon > 0$. We choose for Γ the rectangle with vertices $(\pm A,\pm\varepsilon)$. Because the zeros of Φ are isolated, it is possible without restriction to assume that the contour Γ does not go through any of them. Then:

$$\begin{aligned}\int_\Gamma \frac{\Phi'(\xi)}{\Phi(\xi)}d\xi = & -\int_{-A}^{A}\frac{\partial_1\Phi(\xi_1+\mathbf{i}\varepsilon)}{\Phi(\xi_1+\mathbf{i}\varepsilon)}d\xi_1 + \int_{-A}^{A}\frac{\partial_1\Phi(\xi_1-\mathbf{i}\varepsilon)}{\Phi(\xi_1-\mathbf{i}\varepsilon)}d\xi_1\\ & + \int_{-\varepsilon}^{\varepsilon}\frac{\partial_2\Phi(A+\mathbf{i}\xi_2)}{\Phi(A+\mathbf{i}\xi_2)}d\xi_2 - \int_{-\varepsilon}^{\varepsilon}\frac{\partial_2\Phi(-A+\mathbf{i}\xi_2)}{\Phi(-A+\mathbf{i}\xi_2)}d\xi_2 \qquad \text{(B.8)}\end{aligned}$$

where ∂_1 and ∂_2 denote the derivatives with respect to ξ_1 and ξ_2:

$$\frac{\partial_1\Phi(\xi_1\pm\mathbf{i}\varepsilon)}{\Phi(\xi_1\pm\mathbf{i}\varepsilon)} = \frac{\int_{-\infty}^{+\infty}\mathbf{i}xe^{\mathbf{i}\xi_1 x\mp\varepsilon x}M(x)dx}{\int_{-\infty}^{+\infty}e^{\mathbf{i}\xi_1 x\mp\varepsilon x}M(x)dx} \qquad \text{(B.9)}$$

$$\frac{\partial_2\Phi(\pm A+\mathbf{i}\xi_2)}{\Phi(\pm A+\mathbf{i}\xi_2)} = -\frac{\int_{-\infty}^{+\infty}xe^{\pm\mathbf{i}Ax-\xi_2 x}M(x)dx}{\int_{-\infty}^{+\infty}e^{\pm\mathbf{i}Ax-\xi_2 x}M(x)dx}. \qquad \text{(B.10)}$$

We see that when $\varepsilon \to 0$ (for A fixed), the first two terms in the r.h.s. of (B.8) cancel each other by (B.9). The two last terms vanish in the limit as we integrate between $-\varepsilon$ and ε terms that are bounded at the neighborhood of 0. This shows that M^* does not have any zero on the real interval $[-A, A]$, for any $A > 0$. The proof is finished. $\square$

C Proof of Propositions 4 and 5

Proof of Proposition 4. We have

$$\begin{aligned}
\|\hat{g}_\ell - g\|_2^2 &= \int_{\mathbb{R}_-} \big(\hat{g}_\ell(u) - g(u)\big)^2 du = \int_{\mathbb{R}_-} \Big(e^u \hat{h}_\ell(e^u) - e^u h(e^u)\Big)^2 du \\
&= \int_0^1 \big(\hat{h}_\ell(x) - h(x)\big)^2 x dx. \qquad \text{(C.1)}
\end{aligned}$$

Since $g(u) = e^u h(e^u) = e^u h(1 - e^u)$ by the symmetry of h, we can show that

$$\|\tilde{g}_\ell - g\|_2^2 = \int_0^1 \big(\hat{h}_\ell(x) - h(x)\big)^2 (1 - x) dx.$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[\|\breve{g}_\ell - g\|_2^2\right] &= \mathbb{E}\left[\|\tau \hat{g}_\ell + (1-\tau)\tilde{g}_\ell - g\|_2^2\right] \\
&= \frac{1}{2}\mathbb{E}\big[\, \|\hat{g}_\ell - g\|_2^2 \,\big] + \frac{1}{2}\mathbb{E}\big[\, \|\tilde{g}_\ell - g\|_2^2 \,\big] = \frac{1}{2}\mathbb{E}\big[\|\hat{h}_\ell - h\|_2^2\big], \qquad \text{(C.2)}
\end{aligned}$$

since $\|\hat{g}_\ell - g\|_2^2 + \|\tilde{g}_\ell - g\|_2^2 = \|\hat{h}_\ell - h\|_2^2$. Let us now compute $\mathbb{E}\big[\, \|\tilde{g}_\ell - g\|_2^2 \,\big]$. Recall that $h = 0$ on $\mathbb{R} \setminus (0, 1)$, so $g = 0$ on $\mathbb{R}_+$. For $u < 0$, we define the new variable $v \in \mathbb{R}_-^*$ such that $e^v = 1 - e^u$. We have

$$\tilde{g}_\ell(u) = e^u \hat{h}_\ell(1 - e^u) = e^u \hat{h}_\ell(e^v) = e^{u-v} \hat{g}_\ell(v) = \frac{e^u}{1 - e^u} \hat{g}_\ell\big(\log(1 - e^u)\big).$$

Similarly, we have that $g(u) = \frac{e^u}{1-e^u} g(\log(1 - e^u))$ and thus

$$\begin{aligned}
\mathbb{E}\big[\, \|\tilde{g}_\ell - g\|_2^2 \,\big] &= \mathbb{E}\Big[\int_{\mathbb{R}_-} \big(\tilde{g}_\ell(u) - g(u)\big)^2 du\Big] \\
&= \mathbb{E}\Big[\int_{\mathbb{R}_-} \Big(\frac{e^u}{1 - e^u}\Big)^2 \Big(\hat{g}_\ell\big(\log(1 - e^u)\big) - g\big(\log(1 - e^u)\big)\Big)^2 du\Big] \\
&= \mathbb{E}\Big[\int_{\mathbb{R}_-} \Big(\frac{1 - e^v}{e^v}\Big) \big(\hat{g}_\ell(v) - g(v)\big)^2 dv\Big].
\end{aligned}$$

As a consequence, the middle term in (C.2) is

$$\begin{aligned}
\frac{1}{2}\mathbb{E}\big[\, \|\hat{g}_\ell - g\|_2^2 \,\big] + \frac{1}{2}\mathbb{E}\big[\, \|\tilde{g}_\ell - g\|_2^2 \,\big] &= \mathbb{E}\Big[\int_{\mathbb{R}_-} \frac{1}{2}\Big(1 + \frac{1 - e^v}{e^v}\Big) \big(\hat{g}_\ell(v) - g(v)\big)^2 dv\Big] \\
&= \mathbb{E}\Big[\int_{\mathbb{R}_-} \frac{e^{-v}}{2} \big(\hat{g}_\ell(v) - g(v)\big)^2 dv\Big].
\end{aligned}$$

This concludes the proof. $\square$

Proof of Proposition 5. Remember (C.1). Then, since $h(x) = h(1-x)$,

$$\begin{aligned}
\int_0^1 & \left(\hat{h}_\ell^{sym}(x) - h(x)\right)^2 m(x)dx \\
&= \frac{1}{4}\int_0^1 \left(\hat{h}_\ell(x) - h(x) + \hat{h}_\ell(1-x) - h(1-x)\right)^2 m(x)dx \\
&\le \frac{1}{2}\int_0^1 \left(\hat{h}_\ell(x) - h(x)\right)^2 m(x)dx + \frac{1}{2}\int_0^1 \left(\hat{h}_\ell(1-x) - h(1-x)\right)^2 m(1-x)dx \\
&= \int_0^1 \left(\hat{h}_\ell(x) - h(x)\right)^2 m(x)dx \\
&\le \int_0^1 \left(\hat{h}_\ell(x) - h(x)\right)^2 x dx = \|\hat{g}_\ell - g\|_2^2 .
\end{aligned}$$

This concludes the proof. $\square$

D Proof of Theorem 2 and Corollary 1

Proof of Theorem 2. Let $g_\ell = K_\ell \star g$. We have

$$\|\hat{g}_\ell - g\|_2 \le \|g_\ell - g\|_2 + \|\hat{g}_\ell - g_\ell\|_2.$$

The first term of the above r.h.s inequality is a bias term whereas the second is a variance term. To control the variance term, we have by the Parseval's identity and by (3.6):

$$\begin{aligned}
&\|\hat{g}_\ell - g_\ell\|_2^2 = \frac{1}{2\pi}\|\widehat{g_\ell}^* - g_\ell^*\|_2^2 \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| K_\ell^*(\xi)\Big[\Big(\frac{\alpha\widehat{D^*(\xi)}}{2R}\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} + 1\Big) - g^*(\xi)\Big]\right|^2 d\xi \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| K_\ell^*(\xi)\Big[\Big(\frac{\alpha\widehat{D^*(\xi)}}{2R}\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + \frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + 1\Big) - g^*(\xi)\Big]\right|^2 d\xi \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \left| \frac{\alpha}{2R}K_\ell^*(\xi)\widehat{D^*(\xi)}\Big(\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}\Big) + K_\ell^*(\xi)\Big(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + 1 - g^*(\xi)\Big)\right|^2 d\xi \\
&\le C\int_{\mathbb{R}} \left| K_\ell^*(\xi)\widehat{D^*(\xi)}\Big(\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}\Big)\right|^2 d\xi + C\int_{\mathbb{R}} |K_\ell^*(\xi)^2| \left|\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + 1 - g^*(\xi)\right|^2 d\xi \\
&:= \mathrm{I} + \mathrm{II} .
\end{aligned}$$

In the sequel, we deal with variance of complex variables. Note that for a complex variable, say Z, by distinguishing real and imaginary parts one gets that

$$\mathbb{V}\mathrm{ar}(Z) := \mathbb{E}[|Z - \mathbb{E}(Z)|^2] = \mathbb{E}[|Z|^2] - |\mathbb{E}[Z]|^2 \le \mathbb{E}[|Z|^2].$$

For the term II, because

$$\begin{aligned}
\mathbb{E}\left(K_\ell^*(\xi)\left(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + 1\right)\right) &= K_\ell^*(\xi)\left(\frac{\alpha D^*(\xi)}{2RM^*(\xi)} + 1\right) \\
&= K_\ell^*(\xi)g^*(\xi),
\end{aligned}$$

we have

$$\mathbb{E}[\mathrm{II}] = C\int_{\mathbb{R}} \mathbb{V}\mathrm{ar}\left(K_\ell^*(\xi)\left(\frac{\alpha\widehat{D^*(\xi)}}{2RM^*(\xi)} + 1\right)\right) d\xi$$

$$
\begin{aligned}
&\leq C\int_{\mathbb{R}} \mathbb{V}\mathrm{ar}\left(K_\ell^*(\xi)\frac{\widehat{D^*(\xi)}}{M^*(\xi)}\right)d\xi\\
&\leq C\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right|^2\mathbb{V}\mathrm{ar}\left(\frac{(-\mathbf{i}\xi)}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}\right)d\xi\\
&\leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{V}\mathrm{ar}\left(e^{(\mathbf{i}\xi-1)U_1}\right)d\xi\\
&\leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{E}\left[\left|e^{(\mathbf{i}\xi-1)U_1}\right|^2\right]d\xi \leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{E}\left[e^{-2U_1}\right]d\xi\\
&\leq \frac{C}{n}\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|^2,
\end{aligned}
$$

since $\mathbb{E}[e^{-2U_1}] = \int_0^{+\infty} x^{-2}N(x)dx < +\infty$ thanks to Lemma 1.
We now set

$$
\triangle(\xi) := \frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)}. \tag{D.1}
$$

Then we get

$$
\begin{aligned}
\mathbb{E}[\mathrm{I}] &\leq C\int_{\mathbb{R}}\mathbb{E}\left[\left|K_\ell^*(\xi)\widehat{D^*(\xi)}\triangle(\xi)\right|^2\right]d\xi \leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\mathbb{E}\left[\left|\widehat{D^*(\xi)}\right|^2\left|\triangle(\xi)\right|^2\right]d\xi\\
&\leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\mathbb{E}\left[\left|\widehat{D^*(\xi)} - \mathbb{E}\big[\widehat{D^*(\xi)}\big]\right|^2\left|\triangle(\xi)\right|^2\right]d\xi\\
&\qquad\qquad + C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\left|\mathbb{E}\big[\widehat{D^*(\xi)}\big]\right|^2\mathbb{E}\big[|\triangle(\xi)|^2\big]d\xi\\
&:= \mathrm{III} + \mathrm{IV}.
\end{aligned}
$$

To control the term IV, we need the two following lemmas whose proofs are postponed in Appendix E.

Lemma 2. *There exists a positive constant C_p such that*

$$
\mathbb{E}\left[|\triangle(\xi)|^{2p}\right] \leq C_p \min\left\{\frac{1}{|M^*(\xi)|^{2p}}, \frac{n^{-p}}{|M^*(\xi)|^{4p}}\right\} \quad \text{for } p = 1, 2. \tag{D.2}
$$

Lemma 3. *The ratio $\left|\frac{D^*(\xi)}{M^*(\xi)}\right|$ is bounded:*

$$
\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{4R}{\alpha}, \quad \forall \xi \in \mathbb{R}.
$$

Since $\widehat{D^*}$ is an unbiased estimator of D^* using Lemma 2 we get

$$
\mathrm{IV} \leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\left|D^*(\xi)\right|^2\frac{n^{-1}}{|M^*(\xi)|^4}d\xi.
$$

Then using Lemma 3 we get

$$
\begin{aligned}
\mathrm{IV} &\leq C\int_{\mathbb{R}}\left|K_\ell^*(\xi)\right|^2\frac{n^{-1}}{|M^*(\xi)|^2}d\xi\\
&\leq \frac{C}{n}\left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2.
\end{aligned}
$$

For the term III, we have by applying Cauchy-Schwarz's inequality and by Lemma 2:

$$
\begin{aligned}
\mathrm{III} &\le C \int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \left(\mathbb{E}\left[\left|\widehat{D^*(\xi)} - \mathbb{E}\left[\widehat{D^*(\xi)}\right]\right|^4\right]\right)^{1/2} \left(\mathbb{E}\left[\left|\triangle(\xi)\right|^4\right]\right)^{1/2} d\xi \\
&\le C \int_{\mathbb{R}} \left|K_\ell^*(\xi)\xi\right|^2 \left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j} - \mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]\right|^2\right]\right)^{1/2} \\
&\qquad\qquad\qquad\qquad \times \min\left\{\frac{1}{|M^*(\xi)|^4}, \frac{n^{-2}}{|M^*(\xi)|^8}\right\}^{1/2} d\xi \\
&\le C \int_{\mathbb{R}} \frac{\left|K_\ell^*(\xi)\xi\right|^2}{|M^*(\xi)|^2} \left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j(\xi)\right|^4\right]\right)^{1/2} d\xi,
\end{aligned}
$$

where $Z_j(\xi) = e^{(\mathbf{i}\xi-1)U_j} - \mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]$. Since $Z_1(\xi), \dots, Z_n(\xi)$ are independent centered variables with

$$
\begin{aligned}
\mathbb{E}[|Z_1(\xi)|^4] &= \mathbb{E}\left[\left|e^{(\mathbf{i}\xi-1)U_1} - \mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]\right|^4\right] \le \mathbb{E}\left[\left(\left|e^{(\mathbf{i}\xi-1)U_1}\right| + \left|\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]\right|\right)^4\right] \\
&\le 2^3 \left(\mathbb{E}\left[\left|e^{(\mathbf{i}\xi-1)U_1}\right|^4\right] + \left|\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]\right|^4\right) \\
&\le 8 \left(\mathbb{E}\left[e^{-4U_1}\right] + \left|\mathbb{E}\left[e^{-U_1}\right]\right|^4\right) \\
&\le 8 \left(\int_0^{+\infty} x^{-4} N(x)dx + \left(\int_0^{+\infty} x^{-1} N(x)dx\right)^4\right) < +\infty
\end{aligned} \tag{D.3}
$$

by Lemma 1, applying Rosenthal inequality to real and imaginary parts of complex variables Z_j's, we get

$$
\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j(\xi)\right|^4\right] \le Cn^{-4}\left(n\mathbb{E}[|Z_1(\xi)|^4] + \left(n\mathbb{E}[|Z_1(\xi)|^2]\right)^2\right) \le Cn^{-2}.
$$

Hence

$$
\mathrm{III} \le \frac{C}{n} \int_{\mathbb{R}} \frac{\left|K_\ell^*(\xi)\xi\right|^2}{|M^*(\xi)|^2} d\xi = \frac{C}{n} \left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|^2.
$$

Finally, we obtain

$$
\mathbb{E}\left[\,\|\hat{g}_\ell - g_\ell\|_2^2\,\right] \le \|K_\ell \star g - g\|_2^2 + \frac{C}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2 + \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right).
$$

This ends the proof of Proposition 2. □

Proof of Corollary 1. Using (3.13), due to well-known results on kernel density, the bias term converges to 0:

$$
\lim_{n\to+\infty} \|K_\ell \star g - g\|_2^2 = 0,
$$

and under the assumptions of the theorem we have for the variance term

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

which completes the proof of Theorem 1.

□

E Proofs of technical lemmas

Proof of Lemma 2. This proof is inspired by the proof of Neumann [34]. We will prove the result with $p = 1$. For $p = 2$, the proof is similar.

We split the proof in two cases: $|M^*(\xi)| < 2n^{-1/2}$ and $|M^*(\xi)| \geq 2n^{-1/2}$. Recall that $\Omega = \left\{|\widehat{M^*(\xi)}| \geq n^{-1/2}\right\}$ and $\mathbb{E}\big[\widehat{M^*(\xi)}\big] = \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big] = M^*(\xi)$, we have:

$$\begin{aligned}\mathbb{E}\left[|\triangle(\xi)|^2\right] &= \mathbb{E}\left[\Big|\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \frac{1}{M^*(\xi)]}\Big|^2\right] = \mathbb{E}\left[\left|\frac{\mathbb{1}_\Omega}{\widehat{M^*(\xi)}} - \left(\frac{\mathbb{1}_\Omega}{M^*(\xi)} + \frac{\mathbb{1}_{\Omega^c}}{M^*(\xi)}\right)\right|^2\right] \\ &= \frac{\mathbb{P}(\Omega^c)}{|M^*(\xi)|^2} + \mathbb{E}\left[\mathbb{1}_\Omega \frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2 |M^*(\xi)|^2}\right]. \end{aligned} \tag{E.1}$$

i) If $|M^*(\xi)| < 2n^{-1/2}$*:*

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq \frac{1}{|M^*(\xi)|^2} + \frac{\mathbb{E}\left[|\widehat{M^*(\xi)} - M^*(\xi)|^2\right] n}{|M^*(\xi)|^2|}.$$

But

$$\begin{aligned}\mathbb{E}\left[\left|\widehat{M^*(\xi)} - M^*(\xi)\right|^2\right] &= \mathbb{V}\mathrm{ar}\left[\widehat{M^*(\xi)}\right] = \mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j}\right] \\ &\leq \frac{1}{n}\mathbb{V}\mathrm{ar}\left(e^{\mathbf{i}\xi U_1}\right) \leq \frac{1}{n}\mathbb{E}\left[|e^{\mathbf{i}\xi U_1}|^2\right] = \frac{1}{n}.\end{aligned}$$

Hence we obtain

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq \frac{C}{|M^*(\xi)|^2} \leq C\min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}, \tag{E.2}$$

since $|M^*(\xi)| < 2n^{-1/2}$.

ii) If $|M^*(\xi)| \geq 2n^{-1/2}$*:*

We first control the probability $\mathbb{P}(\Omega^c)$,

$$\begin{aligned}\mathbb{P}\left(\Omega^c\right) &= \mathbb{P}\left(|\widehat{M^*(\xi)}| < n^{-1/2}\right) = \mathbb{P}\left(|\widehat{M^*(\xi)}| < |M^*(\xi)| - |M^*(\xi)| + n^{-1/2}\right) \\ &\leq \mathbb{P}\left(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)| - n^{-1/2}\right) \\ &\leq \mathbb{P}\left(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2\right).\end{aligned} \tag{E.3}$$

Let $T_j(\xi) = e^{\mathbf{i}\xi U_j} - \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big]$, then

$$\widehat{M^*(\xi)} - M^*(\xi) = \frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j} - \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big] = \frac{1}{n}\sum_{j=1}^n T_j(\xi).$$

We have

$$|T_1(\xi)| = \left|e^{\mathbf{i}\xi U_j} - \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big]\right| \leq \left|e^{\mathbf{i}\xi U_j}\right| + \left|\mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big]\right| \leq 2,$$

and

$$\mathbb{V}\mathrm{ar}\big(T_1(\xi)\big) \leq \mathbb{E}\big[|e^{\mathbf{i}\xi U_1}|^2\big] = 1.$$

Since $|M^*(\xi)| \leq 1$ for all $\xi \in \mathbb{R}$ because of M is a density function, we get by Bernstein inequality (see Massart [30])

$$\begin{aligned}
\mathbb{P}\Big(|\widehat{M^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2\Big) &\leq 2\max\Big\{\exp\Big(-\frac{n|M^*(\xi)|^2}{16}\Big), \exp\Big(-\frac{n|M^*(\xi)|}{16}\Big)\Big\} \\
&\leq 2\exp\Big(-\frac{n|M^*(\xi)|^2}{16}\Big) \\
&\leq C\frac{n^{-1}}{|M^*(\xi)|^2}. \qquad \text{(E.4)}
\end{aligned}$$

We also have that

$$\begin{aligned}
\frac{1}{|\widehat{M^*(\xi)}|^2} = \frac{|M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2} &= \frac{|\widehat{M^*(\xi)} - \big(\widehat{M^*(\xi)} - M^*(\xi)\big)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2} \\
&\leq 2\Bigg\{\frac{1}{|M^*(\xi)|^2} + \frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2|M^*(\xi)|^2}\Bigg\}. \qquad \text{(E.5)}
\end{aligned}$$

Thus, from (E.1), (E.3) and (E.5) we have:

$$\begin{aligned}
\mathbb{E}\left[|\triangle(\xi)|^2\right] &\leq C\Bigg\{\frac{n^{-1}}{|M^*(\xi)|^4} + \mathbb{E}\Bigg[\mathbb{1}_\Omega\frac{|\widehat{M^*(\xi)} - M^*(\xi)|^2}{|\widehat{M^*(\xi)}|^2||M^*(\xi)|^2}\Bigg]\Bigg\} \\
&\leq C\Bigg\{\frac{n^{-1}}{|M^*(\xi)|^4} + \frac{\mathbb{E}\big[|\widehat{M^*(\xi)} - M^*(\xi)|^2\big]}{|M^*(\xi)|^4} + \frac{\mathbb{E}\big[|\widehat{M^*(\xi)} - M^*(\xi)|^4\big]n}{|M^*(\xi)|^4}\Bigg\}. \qquad \text{(E.6)}
\end{aligned}$$

To find an upper bound for $\mathbb{E}\big[|\widehat{M^*(\xi)} - M^*(\xi)|^4\big]$, recall that $T_j(\xi) = e^{\mathbf{i}\xi U_j} - \mathbb{E}\big[e^{\mathbf{i}\xi U_1}\big]$. By similar calculations as obtained (D.3), we have $\mathbb{E}[|T_1(\xi)|^4] < +\infty$. Thus we get by Rosenthal's inequality applied to real and imaginary parts of the sequence of independent centered variables $T_1(\xi), \ldots, T_n(\xi)$:

$$\begin{aligned}
\mathbb{E}\big[|\widehat{M^*(\xi)} - M^*(\xi)|^4\big] &= \mathbb{E}\left[\Big|\frac{1}{n}\sum_{j=1}^n T_j(\xi)\Big|^4\right] \\
&\leq Cn^{-4}\Big(n\mathbb{E}[|T_1(\xi)|^4] + \big(n\mathbb{E}[|T_1(\xi)|^2]\big)^2\Big) \leq Cn^{-2}.
\end{aligned}$$

Thus, from (E.3) and (E.6) we get

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C\frac{n^{-1}}{|M^*(\xi)|^4}.$$

Furthermore

$$\frac{1}{|M^*(\xi)|^2} \geq \frac{n^{-1}}{|M^*(\xi)|^4},$$

since $|M^*(\xi)| > 2n^{-1/2}$. Hence

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C\min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}.$$

Combining the two cases, we obtain

$$\mathbb{E}\left[|\triangle(\xi)|^2\right] \leq C\min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}.$$

This ends the proof of Lemma 2. □

Proof of Lemma 3. From equation (3.3) we have

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{2R}{\alpha}\left(|g^*(\xi)|+1\right).$$

Using the change of variable $e^u = x$

$$\begin{aligned}|g^*(\xi)| &= \left|\int_{\mathbb{R}} e^{\mathbf{i}u\xi} g(u)du\right| = \left|\int_{\mathbb{R}} e^{\mathbf{i}u\xi} e^u h(e^u)du\right| = \left|\int_0^\infty e^{\mathbf{i}\xi\log x} h(x)dx\right| \\ &\leq \int_0^1 h(x)dx = 1,\end{aligned}$$

thus

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{4R}{\alpha},$$

which completes the proof.

□

Acknowledgements: The authors thank Sylvain Arlot, Thibault Bourgeron and Matthieu Lerasle for helpful discussions. Van Hà Hoang and Viet Chi Tran have been supported by the Chair "Modélisation Mathématique et Biodiversité" of Veolia Environnement-Ecole Polytechnique-Museum National d'Histoire Naturelle-Fondation X, and also acknowledge support from Labex CEMPI (ANR-11-LABX-0007-01) and ANR project Cadence (ANR-16-CE32-0007).

References

[1] Krishna B Athreya and Peter E Ney. Branching processes, volume 196. Springer Science & Business Media, 2012.

[2] Vincent Bansaye. Proliferating parasites in dividing cells: Kimmel's branching model revisited. The Annals of Applied Probability, pages 967–996, 2008.

[3] Vincent Bansaye, Jean-Francois Delmas, Laurance Marsalle, and Viet Chi Tran. Limit theorems for markov processes indexed by continuous time galton-watson trees. The Annals of Applied Probability, 21(6):2263–2314, 2011.

[4] Vincent Bansaye and Sylvie Méléard. Some stochastic models for structured populations : scaling limits and long time behavior. arXiv:1506.04165, 2015.

[5] Vincent Bansaye, Juan Carlos Pardo Millan, Charline Smadi, et al. On the extinction of continuous state branching processes with catastrophes. Electron. J. Probab, 18(106):1–31, 2013.

[6] Vincent Bansaye and Viet Chi Tran. Branching feller diffusion for cell division with parasite infection. ALEA, Lat. Am. J. Probab. Math. Stat., 8:95–127, 2011.

[7] Bernard Bercu, Benoîte De Saporta, and Anne Gegout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. Electronic Journal of Probability, 14(87):2492–2526, 2009.

[8] Patrick Billingsley. Convergence of probability measures. MR0233396. John Wiley & Sons, 1968.

[9] Thibault Bourgeron, Marie Doumic, and Miguel Escobedo. Estimating the division rate of the growth-fragmentation equaion with a self-similar kernel. Inverse Problems, 30:28p, 2014.

[10] Bertrand Cloez. Limit theorems for some branching measure-valued processes. arXiv preprint arXiv:1106.0660, 2011.

[11] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. Ann. Inst. Henri Poincaré Probab. Stat., 49(2):569–609, 2013.

[12] Fabienne Comte and Claire Lacour. Data driven density estimation in presence of unknown convolution operator. Journal of the Royal Statistical Society: Series B, 73(4):601–627, 2011.

[13] Fabienne Comte, Adeline Samson, and Julien J Stirnemann. Deconvolution estimation of onset of pregnancy with replicate observations. Scandinavian Journal of Statistics, 41(2):325–345, 2014.

[14] Jean-François Delmas and Laurence Marsalle. Detection of cellular aging in a galton–watson process. Stochastic Processes and their Applications, 120(12):2495–2519, 2010.

[15] Marie Doumic, Miguel Escobedo, and Magali Tournus. Estimating the division rate and kernel in the fragmentation equation. submitted, 2017. https://hal.archives-ouvertes.fr/hal-01501811.

[16] Marie Doumic and Pierre Gabriel. Eigenelements of a general aggregation-fragmentation model. Mathematical Models and Methods in Applied Sciences, 20(05):757–783, 2010.

[17] Marie Doumic, Marc Hoffmann, Nathalie Krell, Lydia Robert, et al. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. Bernoulli, 21(3):1760–1799, 2015.

[18] Marie Doumic, Marc Hoffmann, Patricia Reynaud-Bouret, and Vincent Rivoirard. Nonparametric estimation of the division rate of a size-structured population. SIAM Journal of Numerical Analyis, 50(2):925–950, 2012.

[19] Marie Doumic, Benoît Perthame, and Jorge P Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. Inverse Problems, 25(4):045008, 2009.

[20] Stewart N Ethier and Thomas G Kurtz. Markov processes: characterization and convergence, volume 282. John Wiley & Sons, 2009.

[21] Julien Guyon. Limit theorems for bifurcating markov chains. application to the detection of cellular aging. The Annals of Applied Probability, 17(5/6):1538–1569, 2007.

[22] Theodore E Harris. The theory of branching processes. Springer, Berlin, 1963.

[23] P. Henrici. Applied and computational complex analysis. John Wiley & sons, 1997.

[24] Van Ha Hoang. Estimation adaptative pour des problmes inverses avec des applications la division cellulaire. PhD thesis, Université de Lille 1, 2016.

[25] Van Ha Hoang. Estimating the division kernel of a size-structured population. to appear. ESAIM: Probability and Statistics (ESAIM: P&S), 2017.

[26] 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.

[27] Peter Jagers. A general stochastic model for population development. Scandinavian Actuarial Journal, 1969(1-2):84–103, 1969.

[28] Anatole Joffe and Michel Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. Advances in Applied Probability, pages 20–65, 1986.

[29] Benjamin Jourdain, Sylvie Méléard, and Wojbor A. Woyczynski. Lévy flights in evolutionary ecology. Journal of Mathematical Biology, 65(4):677–707, 2012.

[30] Pascal Massart. Concentration Inequalities and Model Selection. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, 6-23, 2003. Springer, 2007.

[31] Sylvie Méléard and Sylvie Roelly. Sur les convergences étroite ou vague de processus à valeur mesures. Comptes Rendus de l'Academie des Sciences, Serie I 317:785–788, 1993.

[32] Sylvie Méléard and Viet Chi Tran. Slow and fast scales for superprocess limits of age-structured populations. Stochastic Processes and their Applications, 122:250–276, 2012.

[33] J.A.J. Metz and V.C. Tran. Daphnias: from the individual based model to the large population equation. Journal of Mathematical Biology, 66(4-5):915–933, 2013. Special issue in honor of Odo Diekmann.

[34] Michael H Neumann. On the effect of estimating the error density in nonparametric deconvolution. Journal of Nonparametric Statistics, 7(4):307–330, 1997.

[35] S Valère Bitseki Penda. Deviation inequalities for bifurcating markov chains on galton- watson tree. ESAIM: Probability and Statistics, 19:689–724, 2015.

[36] B. Perthame and L. Ryzhik. Exponential decay for the fragmentation or cell-division equation. Journal of the Differential Equations, 210:155–177, 2005.

[37] Benoıt Perthame. Transport equations arising in biology. Frontiers in Mathematics. Birchkhauser, 2007.

[38] Viet Chi Tran. Modèles particulaires stochastiques pour des problèmes d'évolution adaptive et pour l'approximation de solutions statisques. PhD dissertation, Université Paris X - Nanterre, http://tel.archives-ouvertes.fr/tel-00125100, 2006.

[39] Viet Chi Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. ESAIM: Probability and Statistics, 12:345–386, 2008.

[40] Viet Chi Tran. Une ballade en forêts aléatoires. HDR dissertation, Université Lille 1, https://tel.archives-ouvertes.fr/tel-01087229v1, 2014.