ORIGINAL ARTICLE

Nonparametric estimation of the fragmentation kernel based on a partial differential equation stationary distribution approximation

Van Ha Hoang[1,2] | **Thanh Mai Pham Ngoc**[3] |
Vincent Rivoirard[4] | **Viet Chi Tran**[5]

[1]Faculty of Mathematics and Computer Science, University of Science, Ho Chi Minh City, Vietnam

[2]Vietnam National University, Ho Chi Minh City, Vietnam

[3]Laboratoire de Mathématiques, UMR 8628, Université Paris Sud, Orsay, France

[4]Ceremade, CNRS, UMR 7534, Université Paris-Dauphine, PSL Research University, Paris, France

[5]LAMA, Univ Gustave Eiffel, UPEM, Univ Paris Est Creteil, CNRS, Marne-la-Vallée, France

Correspondence
Van Ha Hoang, Faculty of Mathematics and Computer Science, University of Science, Ho Chi Minh City, Vietnam.
Email: hvha@hcmus.edu.vn

Funding information
ANR project Cadence, Grant/Award Number: ANR-16-CE32-0007; Chair "Modélisation Mathématique et Biodiversité"; Labex CEMPI, Grant/Award Number: ANR-11-LABX-0007-01

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 here address 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 techniques we derive a nonparametric estimator whose consistency is studied. The main difficulty comes from the nonstandard 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
cell division, deconvolution, growth-fragmentation, kernel rule, nonparametric estimation

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, that is, the half of their mother's size. The more h puts weight in the neighborhood 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.

Our biological motivation for studying this model comes from the understanding of aging phenomena associated with cell division. When a cell that contains toxic content divides asymmetrically, the daughter that contains less toxicity can be viewed as younger in the sense that it has a higher fitness. This toxic content could be detrimental cellular components, such as proteins, extrachromosomal rDNA circles or possibly damaged mitochondria, etc. The concentration of toxic content, that is an increasing function of time during the cell's life, can be seen as a "size." Asymmetry during the divisions impacts the distributions of toxicity among cells and the shapes of trees describing the successive generations of cells in continuous time. Statistical evidence of asymmetrical divisions and biological consequences are described in Stewart et al. (2005). See also Ackermann et al. (2003), Aguilaniu et al. (2003), Banks et al. (2011), Robert et al. (2014), Doumic et al. (2019), or Moseley (2013) for discussions on these topics.

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 $\mathbb{R}_+$ 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 (2012), Harris (1963), Jagers (1969) etc. until recent years (e.g., Bansaye & Tran, 2011; Bansaye et al., 2011; Cloez, 2011). Similar models in discrete time should also be mentioned (e.g., Bansaye, 2008; Bansaye et al., 2013; Bercu et al., 2009; Delmas & Marsalle, 2010; Guyon, 2007; Penda, 2015). For the individual-based model considered in this work, exact numerical simulations are possible. This model offers a convenient framework for statistics (e.g., Hoang, 2016, 2017; Hoffmann & Olivier, 2016). It also connects to the partial differential equations (PDEs) that are usually used in population dynamics (see Bansaye & Méléard, 2015).

We start from an initial population where the individuals are labeled 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 concatenating the mother's label with integers 0 or 1.

The population at time t is described by the point measure:

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

where δ_x is the Dirac measure at x, 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. In what follows, the parameter K will tend to $+\infty$. 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. Notice that if we follow a lineage starting from a cell at time 0 and choosing a daughter at random at each division, we recover an ergodic process with multiplicative jumps (see Hoang (2016), section 2.2.2) implying that the cell sizes are controlled over time, whatever the values of α and R.

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 $\Gamma_i = x_{i0}(T_i)/x_i(T_{i-})$, where $x_i(T_{i-}) = \lim_{t \to T_i, t < T_i} x_i(t)$. Estimating the function h from such a sample has been considered in Hoang (2016, 2017). Here, we focus on the situation when the division tree is not completely observed. Following ideas from Doumic et al. (2012, 2015, 2017) or Bourgeron et al. (2014) 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 $X_1, \ldots, X_n$. Along this paper, we do not take into account the approximation errors related to the asymptotic setting $K \to +\infty$ (the fluctuations associated to the convergence of Z^K could be established following Bansaye et al., 2011, Tran, 2014) nor the approximation by the stationary solution. The latter assumption is discussed in the next section. 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 & Lacour, 2011; Comte & Lacour, 2013), Comte et al. (2014), Neumann (1997) 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. In the setting of a large population close to its stationary state, we define an original estimator of h. 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 Appendix.

1.1 | 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., Billingsley, 1968).

The set of integrable (resp. bounded) nonnegative functions with respect to the Lebesgue measure on $\mathbb{R}_+$ is denoted by $\mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ (resp. $\mathbb{L}^\infty(\mathbb{R}_+, \mathbb{R}_+)$).

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) 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

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

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 $f : (x, s) \mapsto f(x, s) = f_s(x) \in C_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$, the process $(Z_t^K)_{t \geq 0}$ satisfies:*

$$\langle Z_t^K, f_t \rangle = \langle Z_0^K, f_0 \rangle + \int_0^t \int_{\mathbb{R}_+} (\partial_s f_s(x) + \alpha \partial_x f_s(x)$$
$$+ R \int_0^1 (f_s(\gamma x) + f_s((1 - \gamma)x) - f_s(x)) h(\gamma) d\gamma \, Z_s^K(dx) ds + M_t^{K,f}, \tag{3}$$

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

$$\langle M^{K,f} \rangle_t = \frac{1}{K} \int_0^t \int_{\mathbb{R}_+} \int_0^1 R(f_s(\gamma x) + f_s((1 - \gamma)x) - f_s(x))^2 h(\gamma) d\gamma \, Z_s^K(dx) ds. \tag{4}$$

The above equations in Proposition 1 show the evolution of a microscopic random system of particles. The drift coefficient (r.h.s. in the first line of (3)) indicates that each particle grows with speed α. When a particle of size x divides, it is replaced by two daughters of sizes γx and $(1 - \gamma)x$, where γ is drawn in the probability distribution with density h: this corresponds to the second line of (3). When the function h is piked at $1/2$, the daughter cells have almost equal sizes at division, whereas when h has large variance, it is likely to have an asymmetrical division.

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 variance of the martingale part $M^{K,f}$ is of order $1/K$ and we heuristically expect a deterministic limit when $K \to +\infty$. 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 the deterministic measure $\xi_0 \in \mathcal{M}_F(\mathbb{R}_+)$ 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 $\xi \in C(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}_+))$ of*

$$\langle \xi_t, f_t \rangle = \langle \xi_0, f_0 \rangle + \int_0^t \int_{\mathbb{R}_+}$$
$$\left(\partial_s f_s(x) + \alpha \partial_x f_s(x) + R \int_0^1 (f_s(\gamma x) + f_s((1 - \gamma)x) - f_s(x)) \, h(\gamma) d\gamma \right) \xi_s(dx) ds, \tag{5}$$

where $f_t(x) \in C_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ 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, for example, Perthame (2007) and Doumic et al. (2009).

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:*

$$\partial_t n(t,x) + \alpha \partial_x n(t,x) + R n(t,x) = 2R \int_0^\infty n(t,y) h\left(\frac{x}{y}\right) \frac{dy}{y}, \tag{6}$$

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

See Appendix A and Hoang (2016) (proposition 3.2.10) for the proof of this result. Besides the drift α associated with the continuous growth of individuals in time, the PDE (6) involves the death term $R n(t, x)$ and the birth term $2R \int_0^\infty n(t, y) h(x/y) dy/y$. These terms highlight that a particle disappearing at x is replaced by two particles whose sizes are fractions of x. The division is ruled by the density function h and as explained in the introduction, we are interested in the estimation of this density function.

The long time behavior of the solution of PDE (6) is well-known and presented in the following proposition. In this work, 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 (6) 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:

$$\int_0^1 h(u) \frac{du}{u} < +\infty. \tag{7}$$

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

Proposition 3. *Assume (7). Then, there exists a unique probability density $N \in \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ solving the following system:*

$$\begin{cases} \alpha \partial_x N(x) + 2R \, N(x) = 2R \int_0^\infty N(y) h\left(\frac{x}{y}\right) \frac{dy}{y}, & x \geq 0, \\ N(0) = 0, \quad \int N(x) dx = 1, \quad N(x) \geq 0. \end{cases} \tag{8}$$

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

$$\int_0^\infty |n(t,x) e^{-Rt} - \rho N(x)| dx \leq e^{-Rt} \left(\|g_0\|_1 + \frac{6R}{\alpha} \|G_0\|_1 \right), \tag{9}$$

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. Another remark is that the right-hand side of (8) is a multiplicative convolution between $y \mapsto N(y)/y$ and h. Multiplicative

convolutions appear naturally in problems where independent random variables are multiplied (here the size of the cell undergoing division and the random variable of density h ruling how the cell breaks into two daughters). Estimating h can thus be seen as performing a multiplicative deconvolution. 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 $X_1, X_2, \ldots, X_n$ with distribution $N(x)dx$. We estimate h from the data $X_1, \ldots, X_n$ and Equation (8). This experimental scheme has also been used in Doumic et al. (2012) and Bourgeron et al. (2014).

Starting with Equation (8), the multiplicative convolution $\int_0^\infty N(y)h\left(\frac{x}{y}\right)\frac{dy}{y}$ leads 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 (8) into an additive one. Then, we classically apply the Fourier transform and work with products of functions in the Fourier domain. We end up with a deconvolution problem which is more involved and quite different when compared with classical deconvolution problems (see Remark 2).

Let us now 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

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

Equation (8) becomes

$$\alpha D(u) + 2R\, M(u) = 2R\, (M \star g)\, (u), \tag{10}$$

where $\star$ denotes the standard convolution product, so

$$(M \star g)(u) = \int M(u - x)g(x)dx, \quad u \in \mathbb{R}.$$

We have $h(\gamma) = \gamma^{-1}g\left(\log(\gamma)\right)$ 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. Of course, since h is the density of a symmetric probability 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 C^β 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 on $\mathbb{R}_-$.

Remark 1. Assumption 1 implies (7). 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 proposition that will be essential to show consistency and derive rates of convergence (the proof is in Appendix B):

Proposition 4. *Under Assumption 1:*
(i) the first eigenvector N of the eigenproblem (8) satisfies

$$\int_0^{+\infty} x^{-\nu} N(x)dx < +\infty \quad \text{for} \quad \nu \in \{1, \dots, (\nu_0 + 2) \wedge ([\beta] + 1)\}. \tag{11}$$

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

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

(iii) The extension of $M^*(\xi)$ to the complex half-plane $\{\xi \in \mathbb{C} : \Im(\xi) < 1\}$, $\xi \mapsto M^*(\xi) = \int_{\mathbb{R}} e^{ix\xi} M(x)dx$, is holomorphic and thus, M^* admits only isolated zeros on this half-plane. 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 Equation (D3)). 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 (8). 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 (12). 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 (10). 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

$$g^*(\xi) = \frac{\alpha D^*(\xi)}{2RM^*(\xi)} + 1, \quad \xi \in \mathbb{R}. \tag{12}$$

Remark 2. Note that Equation (12) is not standard in classical inverse problems. Actually, classical deconvolution problems with independent noise (see (29) below) can be transformed so that, in the Fourier domain, they can be written similarly to (12) and assumptions are made on the asymptotic behavior of the Fourier transform of the noise density, which is the analog of M^* (see Comte and Lacour (2013) for instance). Such assumptions are not possible here since the smoothness of g is related to the smoothness of M via (10). Assumptions on $M^*(\xi)$ when $|\xi| \to +\infty$ would break these strong relationships between g and M. But these connections between g and M allow us to deduce the asymptotic behavior of $M^*(\xi)$ in Proposition 4 and the issues are circumvented.

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 $u \mapsto M(u) = e^u N(e^u)$. In view of (12), 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 $\widehat{M_n^*(\xi)}$ and $\widehat{D_n^*(\xi)}$, estimators of $M^*(\xi)$ and $D^*(\xi)$ respectively, and defined by

$$\widehat{M_n^*(\xi)} = \frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j}, \tag{13}$$

$$\widehat{D_n^*(\xi)} = (-\mathbf{i}\xi)\frac{1}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}. \tag{14}$$

Obviously, we have that $\widehat{M_n^*(\xi)}$ and $\widehat{D_n^*(\xi)}$ are unbiased estimators of $M^*(\xi) = \mathbb{E}\left[e^{\mathbf{i}\xi U_1}\right]$ and $D^*(\xi) = (-\mathbf{i}\xi)\mathbb{E}[e^{(\mathbf{i}\xi-1)U_1}]$, 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. Along the paper, we use the sinus cardinal kernel defined by $K(x) = \frac{\sin(x)}{\pi x}$ for which $K^*(t) = \mathbb{1}_{[-1,1]}(t)$. For $\ell > 0$, define

the rescaled kernel

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

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

$$\hat{g}_{n,\ell}^*(\xi) = K_\ell^*(\xi) \times \left(\frac{\alpha \widehat{D_n^*(\xi)}}{2R} \frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} + 1\right), \tag{15}$$

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

$$\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} = \begin{cases} \frac{1}{\widehat{M_n^*(\xi)}}, & \text{if } |\widehat{M_n^*(\xi)}| \geq n^{-1/2}, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

The technique used to obtain (15) is similar to inverse truncation filtering (see Bertero & Boccacci, 1998 or Byrne, 2014). Truncation is necessary to avoid explosion when $|\widehat{M_n^*(\xi)}|$ is close to 0. Finally, taking the inverse Fourier transform of $\hat{g}_{n,\ell}^*$, we obtain the estimator of g.

Definition 2. The estimator of g is

$$\hat{g}_{n,\ell}(u) = \frac{1}{2\pi}\int_\mathbb{R} \hat{g}_{n,\ell}^*(\xi)e^{-iu\xi}d\xi, \quad u \in \mathbb{R}_-. \tag{17}$$

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

$$\hat{h}_{n,\ell}(\gamma) = \gamma^{-1}\hat{g}_{n,\ell}\left(\log(\gamma)\right), \quad \gamma \in (0,1). \tag{18}$$

The main difficulty lies in the choice of ℓ. This problem is dealt with subsequently. Deconvolution estimators have been studied in Comte and Lacour (2011, 2013), Comte et al. (2014), Neumann (1997). 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 (8), in particular through Equation (12). This complicates the study of the rates of convergence. The next section studies the quadratic risk of $\hat{g}_{n,\ell}$ and $\hat{h}_{n,\ell}$.

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

Proposition 5. *For a Bernoulli random variable τ with parameter $1/2$ independent of $X_1, \ldots, X_n$, let us define the randomized estimator*

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

We have

$$\mathbb{E}\left[\|\hat{h}_{n,\ell} - h\|_2^2\right] = 2\mathbb{E}\left[\|\check{g}_{n,\ell} - g\|_2^2\right] = \mathbb{E}\left[\int_{\mathbb{R}_-} e^{-u}\left(\hat{g}_{n,\ell}(u) - g(u)\right)^2 du\right]. \tag{19}$$

The last equality in (19) shows that if we want to control the quadratic risk of $\hat{h}_{n,\ell}$ with respect to the Lebesgue measure, tight controls on the loss of $\hat{g}_{n,\ell}$ 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

$$\hat{h}_{n,\ell}^{sym}(x) = \frac{1}{2}\left(\hat{h}_{n,\ell}(x) + \hat{h}_{n,\ell}(1-x)\right), \tag{20}$$

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

Proposition 6. *Setting $m(x) = x(1-x)$, we have that*

$$\int_0^1 \left(\hat{h}_{n,\ell}^{sym}(x) - h(x)\right)^2 m(x)dx \leq \|\hat{g}_{n,\ell} - g\|_2^2. \tag{21}$$

Propositions 5 and 6 are proved in Appendix C. The previous result does not provide any control on boundaries of the interval $[0, 1]$ but the consistency of $\hat{g}_{n,\ell}$ will establish the consistency of $\hat{h}_{n,\ell}^{sym}$ on every compact set of $(0, 1)$. The study of the consistency of $\hat{g}_{n,\ell}$ 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 $\hat{g}_{n,\ell}$. More precisely, we establish the $\mathbb{L}^2$-consistency of $\hat{g}_{n,\ell}$ under a suitable choice of the bandwidth ℓ.

We first study the bias-variance decomposition of the $\mathbb{L}^2$-risk of $\hat{g}_{n,\ell}$. Recall that from Proposition 4(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*

$$\mathbb{E}\left[\|\hat{g}_{n,\ell} - g\|_2^2\right] \leq \|K_\ell \star g - g\|_2^2 + \frac{C}{n}S(\ell), \tag{22}$$

where

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

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

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*

$$\lim_{n\to+\infty} \frac{1}{n}\left(\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2 + \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2\right) = 0, \tag{23}$$

we have

$$\lim_{n \to +\infty} \mathbb{E}[\|\hat{g}_{n,\ell} - g\|_2^2] = 0. \tag{24}$$

The proof of Corollary 1 is straightforward. Indeed, due to the well-known results on kernel density, we have $\lim_{n \to +\infty} \|K_\ell \star g - g\|_2^2 = 0$ and under the assumptions of the corollary we have for the variance term $\lim_{n \to +\infty} n^{-1} S(\ell) = 0$. Thus we get the result (24). The proof of Theorem 2 is given in Appendix D. Note that under Assumption 1, we have by Proposition 4 that $|M^*(\xi)| = O(|\xi|^{-([\beta] \wedge (v_0+3))})$ when $|\xi| \to +\infty$. If we have $|M^*(\xi)| \sim C|\xi|^{-([\beta] \wedge (v_0+3))}$, for a constant $C > 0$, a bandwidth ℓ can be easily derived. Indeed,

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

and

$$\left\| \frac{K_\ell^*(\xi)\xi}{M^*(\xi)} \right\|_2^2 = \int_{-\ell^{-1}}^{\ell^{-1}} \frac{\xi^2}{|M^*(\xi)|^2} d\xi = O(\ell^{-(3+2([\beta] \wedge (v_0+3)))}),$$

and then, Assumption (23) is satisfied if

$$\ell^{-1} = o\left(n^{\frac{1}{3+2([\beta] \wedge (v_0+3))}} \right).$$

We obtain convergence rates for the quadratic risk of $\hat{g}_{n,\ell}$ under additional smoothness properties for the density g. For this purpose, we introduce Sobolev spaces defined as follows.

Definition 3. We consider Sobolev spaces $S(\beta, L)$ defined as the class of integrable functions $f : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int |f^*(t)|^2 (1 + t^2)^\beta dt \leq L^2.$$

We then obtain the following result.

Proposition 7. *If $g \in S(\beta, L)$ and $|M^*(\xi)| \sim C|\xi|^{-([\beta] \wedge (v_0+3))}$, for a constant $C > 0$, then we have*

$$\mathbb{E}[\|\hat{g}_{n,\ell} - g\|_2^2] = O\left(n^{-\frac{2\beta}{2\beta + 2([\beta] \wedge (v_0+3)) + 3}} \right).$$

The rate of convergence of Proposition 7 is the usual rate of convergence for ill-posed inverse problems involving a derivative and an ordinary smooth noise with a polynomial decay of order $[\beta] \wedge (v_0 + 3)$. This result shows good theoretical performances of our procedure.

4 | NUMERICAL SIMULATIONS

4.1 | Influence of the preliminary estimators $\widehat{M_n^*}$ and $\widehat{D_n^*}$ on the performances of $\hat{h}_{n,\ell}$ and $\hat{g}_{n,\ell}$

In this section, we study the numerical performances of our estimation procedure.

In the literature (e.g., Stewart et al., 2005; Wang et al., 2010), it is possible to obtain real datasets of sample size $n = 30,000$ or even a larger: in Stewart et al. (2005), the authors followed

divisions of *E. coli* and obtained a complete record of measurements of 35,049 cells, in Wang et al. (2010), the authors introduced their experimental procedures and techniques that allow to obtain a real dataset of 10^7 cells. Therefore, the simulations presented here are performed on simulated samples of sizes n varying from 1000 to 30,000.

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 (6) using the method presented in Doumic et al. (2009).

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 (12) and (15), the construction of the estimator $\hat{g}_{n,\ell}$ is based on the estimation of M^* and D^*. Remember that $D^*(\xi) = (-\mathbf{i}\xi)\mathbb{E}\left[e^{(\mathbf{i}\xi-1)U_1}\right]$ 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,
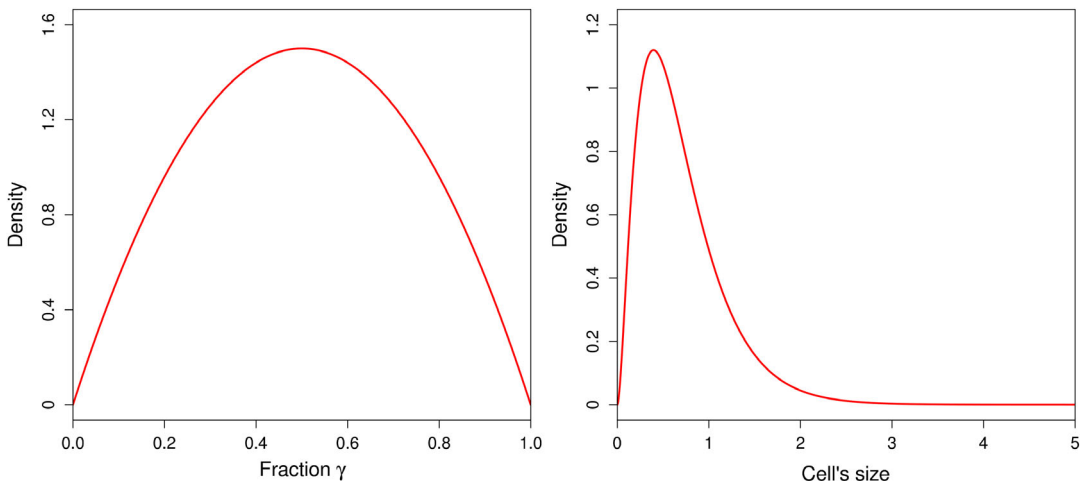

FIGURE 1 The Beta$(2, 2)$ density h_1 (left) and its corresponding stationary density N_1 (right)
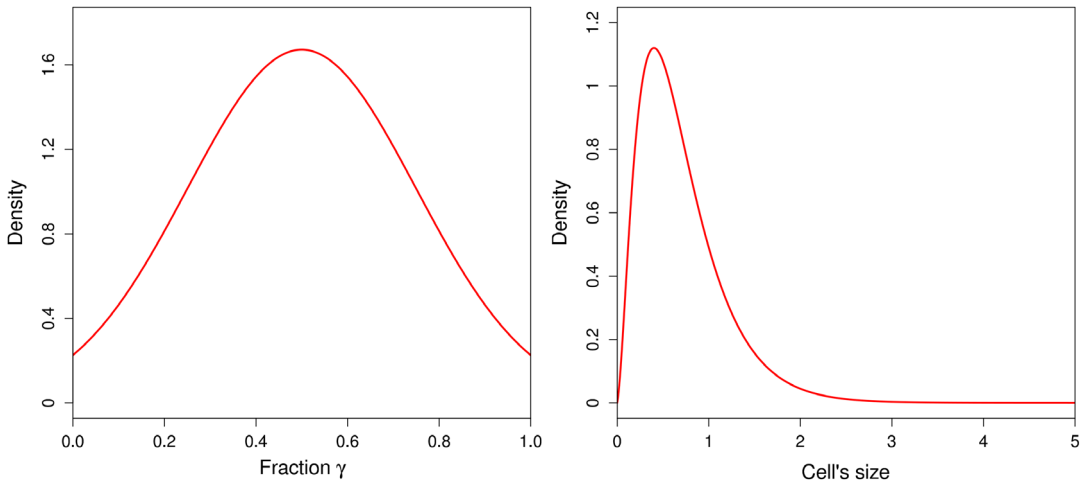
FIGURE 2 The truncated normal density h_2 (left) and its corresponding stationary density N_2 (right)
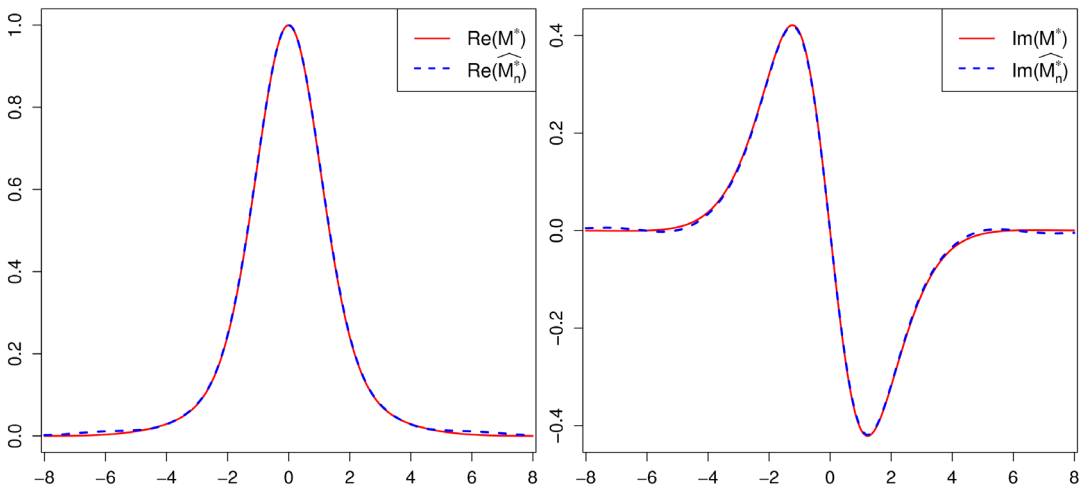

FIGURE 3 For the Beta$(2, 2)$ density, the real part (left) and the imaginary part (right) of $\hat{M}_n^*$ (blue line) compared with those of M^* (red line)

we introduce the modified formulas of D^* and $\widehat{D_n^*}$, denoted, respectively, by $\mathfrak{D}^*$ and $\widehat{\mathfrak{D}_n^*}$, obtained by removing $-\mathbf{i}\xi$ from the original formulas:

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

Figures 3, 4, and 5 provide a reconstruction of $\widehat{M_n^*}$, $\widehat{D_n^*}$ and $\widehat{\mathfrak{D}_n^*}$ based on a random sample $U_1, \ldots, U_n$ of size $n = 30,000$ for h_1. For each figure, we represent both the real part and the imaginary part of $\widehat{M_n^*}$ (resp. $\widehat{D_n^*}$, $\widehat{\mathfrak{D}_n^*}$) 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.
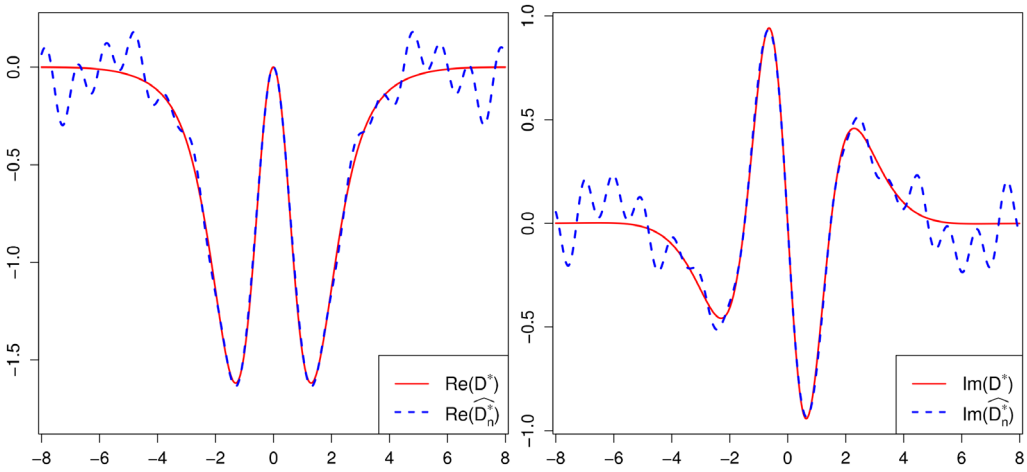
FIGURE 4 For the Beta$(2, 2)$ density, the real part (left) and the imaginary part (right) of $\hat{D}_n^*$ (blue line) compared with those of D^* (red line)
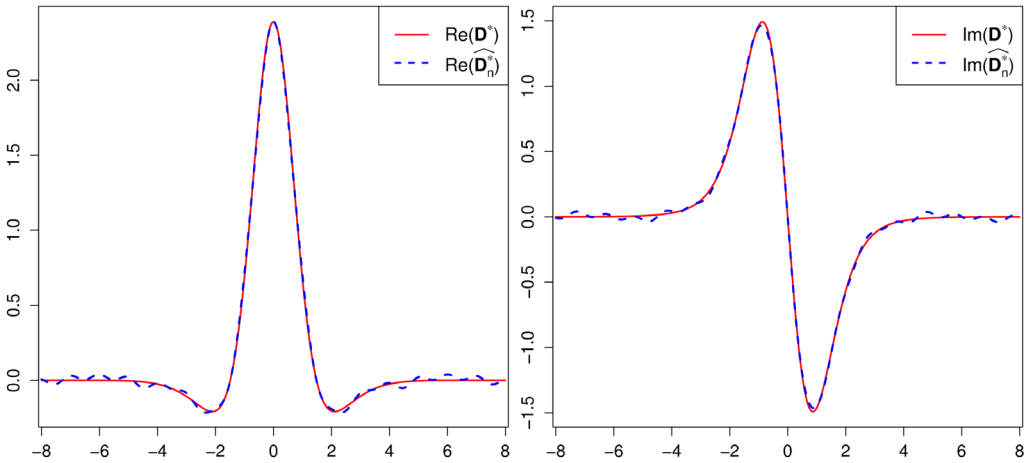

FIGURE 5 For the Beta$(2, 2)$ density, the real part (left) and the imaginary part (right) of $\hat{\mathfrak{D}}_n^*$ (blue line) compared with those of $\mathfrak{D}^*$ (red line)

Figures 3 and 5 show that the reconstructions of M^* and $\mathfrak{D}^*$ are very satisfying, whereas large oscillations in the reconstruction of $D^*(\xi)$ appear when ξ is large (see Figure 4), due to a large variance term. This confirms what we mentioned: the estimation of the derivative D^* has a strong influence for our statistical problem.

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

4.2 | Bandwidth selection rules

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

estimator $\hat{g}_{n,\ell}$ in the Fourier domain:

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

Define

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

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

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

We aim at constructing an estimator of $J(\ell)$, which is equivalent to providing an estimate of the scalar product $\langle\hat{g}_{n,\ell}^*, g\rangle$ since $\|\hat{g}_{n,\ell}^*\|_2^2$ 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 $\hat{g}_{n,\ell}^{*(t)}$ (resp. $\hat{g}_{n,\ell}^{*(v)}$) be the estimator of g^* constructed on the training set (resp. on the validation set). The heuristics is that if $\hat{g}_{n,\ell'}^{*(v)}$ is an estimator constructed on the validation set, then $\langle\hat{g}_{n,\ell}^{*(t)}, \hat{g}_{n,\ell'}^{*(v)}\rangle$ gives us an estimate of $\langle\hat{g}_{n,\ell}^{*(t)}, g^*\rangle$ 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 subsamples:

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

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 4. Let $(E_j, E_j^C)_{1 \leq j \leq V}$, $V \leq V_{\max}$ be the sequence of subsets selected from $\{1, \ldots, n\}$ and the corresponding sub-samples $(\mathbf{X}^{E_j}, \mathbf{X}^{E_j^C})_{1 \leq j \leq V}$. Let $\hat{g}_{n,\ell}^{*(E_j)}$ and $\hat{g}_{n,\ell}^{*(E_j^C)}$ be the estimators of g_ℓ^*

respectively constructed on the subsamples $\mathbf{X}^{E_j}$ and $\mathbf{X}^{E_j^C}$. Define

$$\hat{J}_{Crit1}(\ell) := \frac{1}{V} \sum_{j=1}^{V} \left[\|\hat{g}_{n,\ell}^{*(E_j)}\|_2^2 - 2 \left\langle \hat{g}_{n,\ell}^{*(E_j)}, \hat{g}_{n,\ell}^{*(E_j^C)} \right\rangle \right]. \tag{25}$$

Then the selected bandwidth is given by

$$\hat{\ell}_{Crit1} := \operatorname*{argmin}_{\ell \in \mathcal{L}} \hat{J}_{Crit1}(\ell). \tag{26}$$

Definition 5. Let $\hat{g}_{n,\ell}^{*(E_j)}$ and $\hat{g}_{n,\ell'}^{*(E_j^c)}$ be the estimators of $g_{n,\ell}^*$ as in Definition 4. Define,

$$\hat{J}_{Crit2}(\ell, \ell') := \frac{1}{V} \sum_{j=1}^{V} \left[\|\hat{g}_{n,\ell}^{*(E_j)}\|_2^2 - 2 \left\langle \hat{g}_{n,\ell}^{*(E_j)}, \hat{g}_{n,\ell'}^{*(E_j^C)} \right\rangle \right]. \tag{27}$$

Then an alternative bandwidth selection rule is given as follows:

$$\hat{\ell}_{Crit2} := \operatorname*{argmin}_{\ell \in \mathcal{L}} \left\{ \min_{\ell' \in \mathcal{L}} \hat{J}_{Crit2}(\ell, \ell') \right\}. \tag{28}$$

Note that the second criterion is more computationally intensive.

4.3 | Numerical results

Remember that we aim at reconstructing the densities h_1 and h_2, that is, 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 (15), (17) and (18) 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:

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

Of course, ℓ_{oracle} and $\hat{g}_{n,\ell_{\text{oracle}}}$ cannot be used in practice (since they depend on the true function to estimate) but they can be viewed as benchmark quantities. For $n = 30,000$ 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 4).

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 (20)) provides significant 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.

Table 1 shows the $\mathbb{L}^2$-risk of $\hat{g}_{n,\hat{\ell}_{Crit1}}$ and $\hat{g}_{n,\hat{\ell}_{Crit2}}$ where $\hat{\ell}_{Crit1}$ and $\hat{\ell}_{Crit2}$ are the bandwidths selected by our selection rules (see Definitions 4 and 5), 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 $n = 30,000$. We also provide associated Boxplots in Figures 9 and 10.
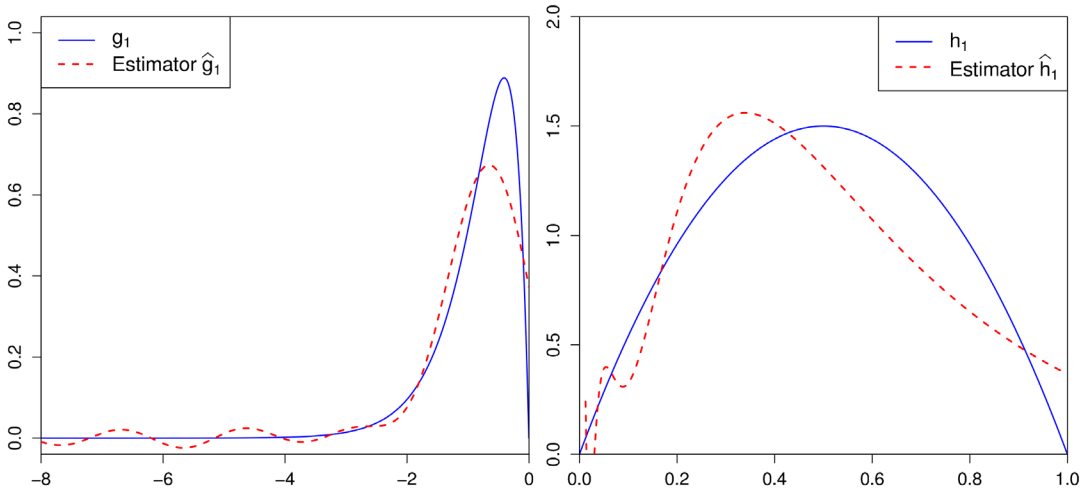
FIGURE 6 Estimation of $g_1(x) = e^x h_1(e^x)$ (left) and h_1 (right)
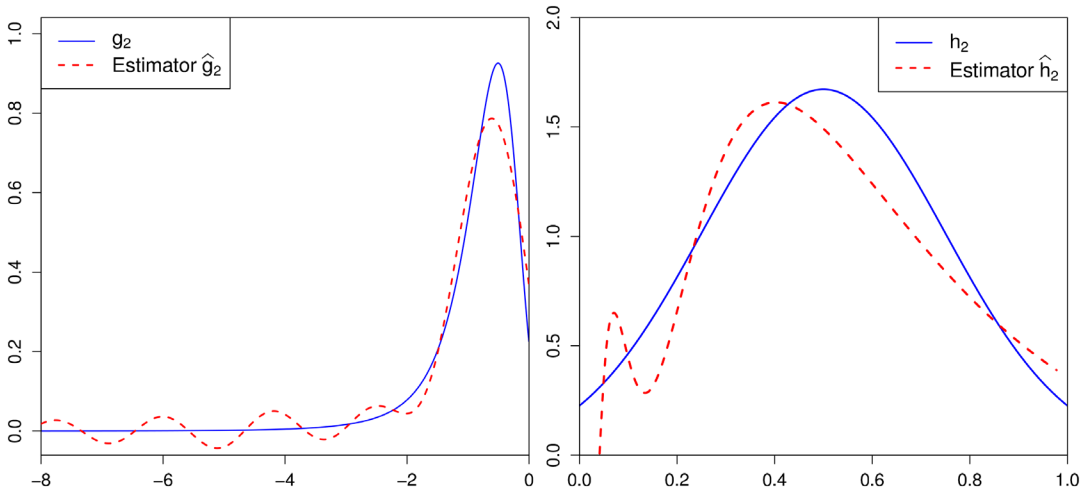

FIGURE 7 Estimation of $g_2(x) = e^x h_2(e^x)$ (left) and h_2 (right)

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 $V = 10$. Finally, for both estimation of g_1 and g_2 according to the $Beta(2, 2)$ distribution and the truncated normal distribution respectively, we illustrate in Figure 11 the regression lines of the logarithm of the mean squared error of $\hat{g}_{n,\hat{\ell}_{Crit1}}$ versus the logarithm of the sample size, with $n \in \{1000; 2000; 5000; 10,000; 20,000; 30,000\}$. One can observe that the MSE's decrease
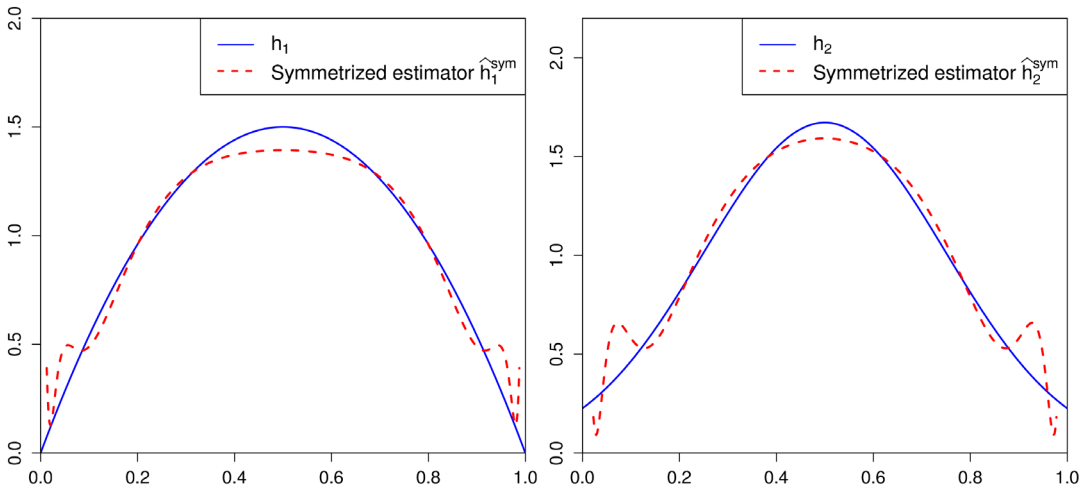
FIGURE 8 Reconstructions of h_1 (left) and h_2 (right) after symmetrization

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

		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
	$\bar{\tilde{\ell}}$	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
	$\bar{\tilde{\ell}}$	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
	$\bar{\tilde{\ell}}$	0.29837	0.29985	0.27583	0.30396	0.30303	0.27858

as the sample size n increases. This justifies the convergence of our estimators from the practical point of view.

5 | CONCLUSION

Many statistical papers interested in aging phenomena for population of dividing cells concentrate on the estimation of the division rate (the constant R in the present work) (Bourgeron et al., 2014; Doumic et al., 2012, 2015, 2017; Hoffmann & Olivier, 2016). In these papers, the division rate is assumed to be a function that depends on certain quantities growing with time and that can be seen as ages of the cells (bad chemical contents, size...). The decrease of the division rate with respect to these quantities can be understood as senescence at the individual level. At the population level, lineages along which the distribution of these quantities tend to shift to higher values can be seen as aging or senescent lineages. In the present work, we investigate another aspect that is the kernel $h(.)$ ruling the division of the mother cell. In most of the previous works, for example, (Doumic et al., 2012, 2015), the mother cell divides into two identical cells. From the microscopic
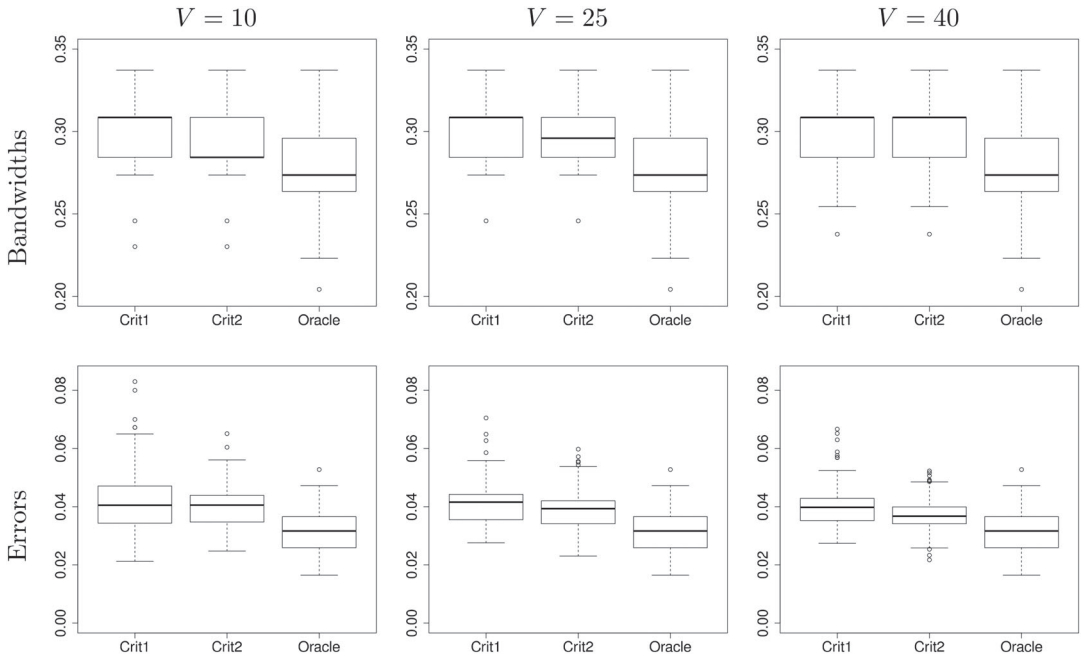
FIGURE 9   Bandwidths and errors for the estimation of h_1 (Beta$(2, 2)$ distribution)
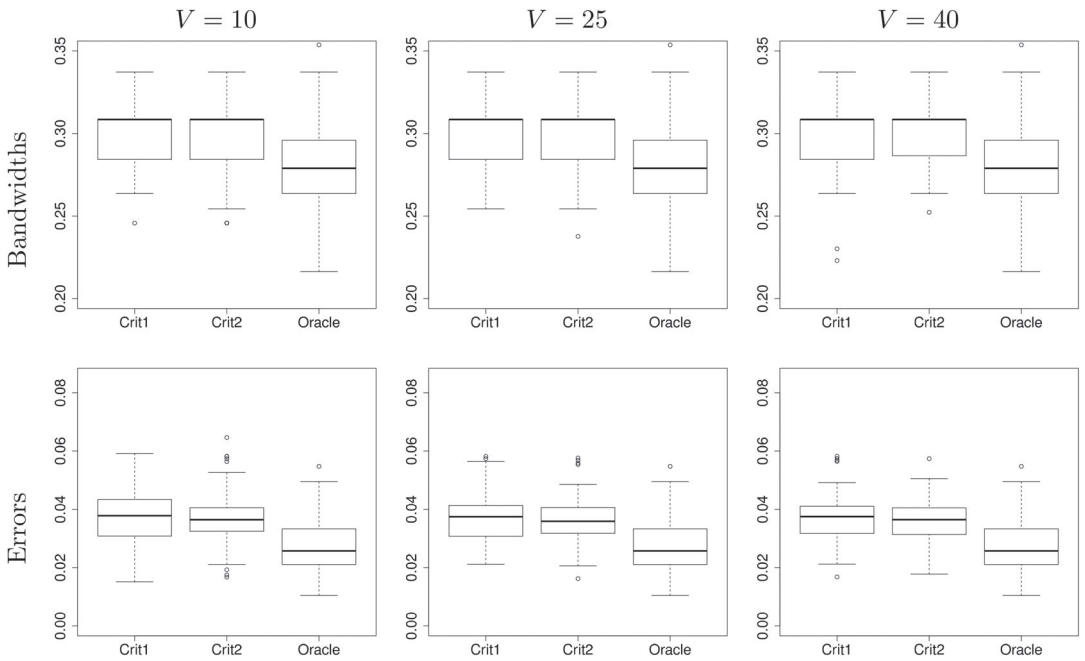

FIGURE 10   Bandwidths and errors for the estimation of h_2 (Truncated normal)
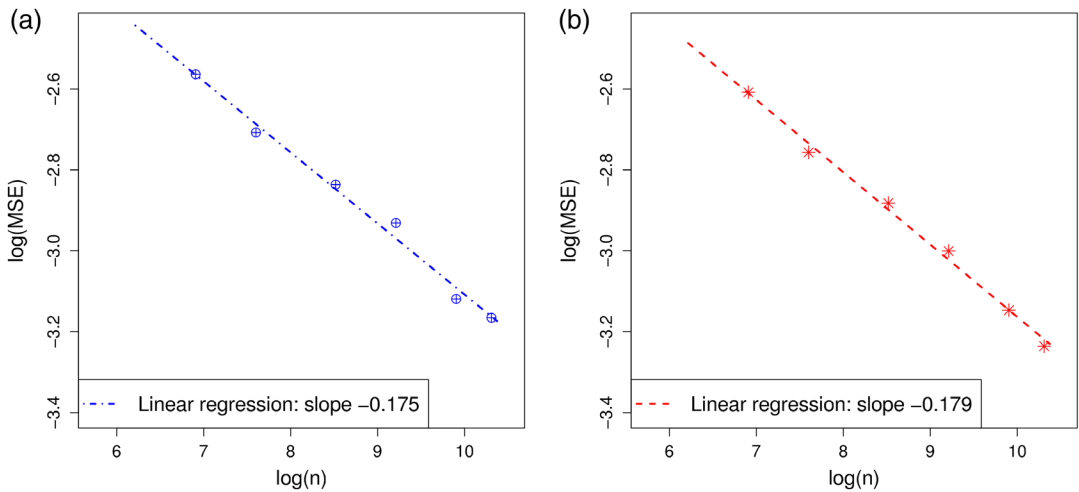
FIGURE 11 The regression lines for log-mean square error for the reconstruction of g_1 (left, Eq: $y = -1.351 - 0.175x$) and the reconstruction of g_2 (right, Eq: $y = -1.375 - 0.179x$) versus $\log(n)$, with the sample size n varying from 1000 to 30,000

individual-based point of view, the senescence in lineages then arises from the sole randomness in the times of division. The asymmetry between daughter cells is, however, an important feature that has to be taken into account. Some recent developments have been made in this direction: by the authors (Hoang, 2016, 2017) and also by Bourgeron et al. (2014), Doumic et al. (2017) but from a deterministic point of view.

The inverse problem arising from the estimation of h, Equation (12), is not a standard deconvolution problem. The function M^* is closely linked to the function of interest g (i.e., the function h in a logarithmic scale) through Equation (10). Compared with classical deconvolution problems (e.g., (29) below or Meister, 2009), M^* cannot be handled as an independent known noise. Firstly, the regularity and positivity of M^* have to be studied and give way to involved and technical proofs. Secondly, one has to estimate M^* in order to devise an estimator of g^*. This actually complicates the study of the proposed estimation procedure since we have to control the fluctuations of the empirical process $\hat{M}_n^*$. However, the theoretical study led in Section 3 allows us to circumvent these issues and to show consistency of our estimates. These nice performances are also illustrated from a numerical point of view by the use of artificial data whose size is consistent with real ones.

In line with the previous comment, a natural extension would consist in deriving the bandwidth by using an alternative theoretical approach to the cross-validation type approach described in Section 4.3. It would be natural to use, for instance, the Goldenshluger–Lepski methodology Goldenshluger and Lepski (2013) in the same spirit as Comte and Lacour (2013) or the PCO methodology Lacour et al. (2017), with the aim of deriving oracle inequalities. But such technical approaches require sharp controls of the variance of estimates and powerful concentration inequalities. Obtaining such results is beyond the scope of this paper but constitutes interesting challenges for future research.

Of course, our probabilistic model could be enhanced by taking into account some observational noise and instead of the X_i's we would observe,

$$Y_i = X_i + \varepsilon_i, \tag{29}$$

with X_i and the noise ϵ_i being independent, which corresponds to a classical deconvolution problem if we are interested in recovering the density of X_i's. The estimation of h would then require to combine classical deconvolution technics and the approach of this paper.

Furthermore, our study is in line with the references mentioned at the beginning of the conclusion: large populations close to their stationary states. This is justified by the exponentially fast convergence rate given in (9). A possible direction for further research would be to focus directly on the evolution problem (6). Following ideas of Comte and Genon-Catalot (2020), it could be possible to study a projection estimator computed from the finite particle system on the compact time interval $[0, T]$. These challenging inverse problems provide nice motivations for further work.

ORCID

Van Ha Hoang https://orcid.org/0000-0002-6020-7981

REFERENCES

Ackermann, M., Stearns, S. C., & Jenal, U. (2003). Senescence in a bacterium with asymmetric division. *Science*, *300*(5627), 1920–1920.

Aguilaniu, H., Gustafsson, L., Rigoulet, M., & Nyström, T. (2003). Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science*, *299*(5613), 1751–1753.

Athreya, K. B., & Ney, P. E. (2012). *Branching processes* (Vol. *196*). Springer Science & Business Media.

Banks, H. T., Sutton, K. L., Thompson, W. C., Bocharov, G., Roose, D., Schenkel, T., & Meyerhans, A. (2011). Estimation of cell proliferation dynamics using cfse data. *Bulletin of Mathematical Biology*, *73*(1), 116–150.

Bansaye, V. (2008). Proliferating parasites in dividing cells: Kimmel's branching model revisited. *The Annals of Applied Probability*, *18*(3), 967–996.

Bansaye, V., Delmas, J.-F., Marsalle, L., & Tran, V. C. (2011). Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *The Annals of Applied Probability*, *21*(6), 2263–2314.

Bansaye, V., & Méléard, S. (2015). *Some stochastic models for structured populations: Scaling limits and long time behavior*. arXiv:1506.04165.

Bansaye, V., Millan, J. C. P., & Smadi, C. (2013). On the extinction of continuous state branching processes with catastrophes. *Electronic Journal of Probability*, *18*(106), 1–31.

Bansaye, V., & Tran, V. C. (2011). Branching feller diffusion for cell division with parasite infection. *ALEA-Latin American Journal of Probability Mathematical Statistics*, *8*, 95–127.

Bercu, B., De Saporta, B., & Gegout-Petit, A. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability*, *14*(87), 2492–2526.

Bertero, M., & Boccacci, P. (1998). *Introduction to inverse problems in imaging*. CRC Press.

Billingsley, P. (1968). *Convergence of probability measures* (MR0233396). John Wiley & Sons.

Bourgeron, T., Doumic, M., & Escobedo, M. (2014). Estimating the division rate of the growth-fragmentation equaion with a self-similar kernel. *Inverse Problems*, *30*(2), 025007.

Byrne, C. L. (2014). *Signal processing: A mathematical approach*. CRC Press.

Cloez, B. (2011). *Limit theorems for some branching measure-valued processes*. arXiv preprint arXiv:1106.0660.

Comte, F., & Genon-Catalot, V. (2020). Nonparametric drift estimation for iid paths of stochastic differential equations. *Annals of Statistics to appear*.

Comte, F., & Lacour, C. (2011). Data driven density estimation in presence of unknown convolution operator. *Journal of the Royal Statistical Society: Series B*, *73*(4), 601–627.

Comte, F., & Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, *49*(2), 569–609.

Comte, F., Samson, A., & Stirnemann, J. J. (2014). Deconvolution estimation of onset of pregnancy with replicate observations. *Scandinavian Journal of Statistics*, *41*(2), 325–345.

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

Doumic, M., Escobedo, M., & Tournus, M. (2017). Estimating the division rate and kernel in the fragmentation equation. *submitted*. Retrieved from Https://hal.archives-ouvertes.fr/hal-01501811

Doumic, M., & Gabriel, P. (2010). Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, *20*(05), 757–783.

Doumic, M., Hoffmann, M., Krell, N., & Robert, L. (2015). Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, *21*(3), 1760–1799.

Doumic, M., Hoffmann, M., Reynaud-Bouret, P., & Rivoirard, V. (2012). Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal of Numerical Analyis*, *50*(2), 925–950.

Doumic, M., Olivier, A., & Robert, L. (2019). *Estimating the division rate from indirect measurements of single cells*. arXiv:1907.05108.

Doumic, M., Perthame, B., & Zubelli, J. P. (2009). Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, *25*(4), 045008.

Epstein, B. (1948). Some applications of the Mellin transform in statistics. *The Annals of Mathematical Statistics*, *19*(3), 370–379.

Ethier, S. N., & Kurtz, T. G. (2009). *Markov processes: Characterization and convergence* (Vol. *282*). John Wiley & Sons.

Goldenshluger, A., & Lepski, O. (2013). General selection rule from a family of linear estimators. *Theory of Probability & Its Applications*, *57*(2), 209–226.

Guyon, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *The Annals of Applied Probability*, *17*(5/6), 1538–1569.

Harris, T. E. (1963). *The theory of branching processes*. Springer.

Hoang, V. H. (2016). *Estimation adaptative pour des problés inverses avec des applications á la division cellulaire* (Ph.D. thesis). Université de Lille. (1).

Hoang, V. H. (2017). Estimating the division kernel of a size-structured population. *ESAIM: Probability and Statistics (ESAIM: P&S).21*, 275-302.

Hoffmann, M., & Olivier, A. (2016). Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Processes and Their Applications*, *126*(5), 1433–1471.

Jagers, P. (1969). A general stochastic model for population development. *Scandinavian Actuarial Journal*, *1969*(1-2), 84–103.

Joffe, A., & Métivier, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, *18*(1), 20–65.

Jourdain, B., Méléard, S., & Woyczynski, W. A. (2012). Lévy flights in evolutionary ecology. *Journal of Mathematical Biology*, *65*(4), 677–707.

Lacour, C., Massart, P., & Rivoirard, V. (2017). Estimator selection: A new method with applications to kernel density estimation. *Sankhya A*, *79*(2), 298–335.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics, vol. 193 of Lecture Notes in Statistics*. Springer-Verlag.

Méléard, S., & Roelly, S. (1993). Sur les convergences étroite ou vague de processus à valeur mesures. *Comptes Rendus de l'Academie des Sciences, Serie I*, *317*, 785–788.

Méléard, S., & Tran, V. C. (2012). Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and Their Applications*, *122*, 250–276.

Metz, J., & Tran, V. (2013). Daphnias: From the individual based model to the large population equation. *Journal of Mathematical Biology*, *66*(4-5), 915–933.

Moseley, J. B. (2013). Cellular aging: Symmetry evades senescence. *Current Biology*, *23*(19), R871–R873.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, *7*(4), 307–330.

Penda, S. V. B. (2015). Deviation inequalities for bifurcating Markov chains on Galton-Watson tree. *ESAIM: Probability and Statistics*, *19*, 689–724.

Perthame, B. (2007). *Transport equations arising in biology*. Frontiers in Mathematics.

Perthame, B., & Ryzhik, L. (2005). Exponential decay for the fragmentation or cell-division equation. *Journal of the Differential Equations*, *210*, 155–177.

Robert, L., Hoffmann, M., Krell, N., Aymerich, S., Robert, J., & Doumic, M. (2014). Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC Biology*, *12*(1), 1.

Stewart, E. J., Madden, R., Paul, G., & Taddei, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLOS Biology*, *3*(2), e45.

Titchmarsh, E. C. (1948). *Introduction to the theory of Fourier integrals* (2nd ed.). Oxford University Press.

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

Tran, V. C. (2008). Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics*, *12*, 345–386.

Tran, V. C. (2014). *Une ballade en forêts aléatoires* (HDR dissertation). Université Lille. (1). Retrieved from https://tel.archives-ouvertes.fr/tel-01087229v1.

Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., & Jun, S. (2010). Robust growth of escherichia coli. *Current Biology*, *20*(12), 1099–1103.

APPENDIX A. LARGE POPULATION RENORMALIZATION

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

Before proving the results of Section 2, let us build the SDE satisfied by the process $(Z_t^K)_{t\geq0}$. Consider

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

the random point measure on $\mathcal{I}\times\mathbb{R}_+ = \cup_{\ell\geq1}\mathbb{N}\times\{0,1\}^{\ell-1}\times\mathbb{R}_+$ 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 $(\tilde{Z}_0^K)_{K\in\mathbb{N}^*}$ 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) 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). 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 $(\tilde{Z}_0^K)_{K\in\mathbb{N}^*}$.

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

$$
\tilde{Z}_t^K = \sum_{i\in V_0^K} \delta_{(i,x_i(0)+\alpha t)}
$$
$$
+ \int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i\in V_{s_-}^K\}} \left(\delta_{(i0,\gamma x_i(s_-)+\alpha(t-s))} + \delta_{(i1,(1-\gamma)x_i(s_-)+\alpha(t-s))} - \delta_{(i,x_i(s_-)+\alpha(t-s))} \right) Q(ds,di,d\gamma), \quad \text{(A1)}
$$

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 Tran (2008) and Metz and Tran (2013): 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 right-hand side of (A1) corresponds to the individuals alive at time 0 with their sizes at time t if they do not 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 $x_i(s_-)+\alpha(t-s)$) 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) propagates to positive time and it is possible to show that for any $T>0$, (see Hoang, 2016, proposition 3.2.5)

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

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

$$
\langle Z_t^K, f_t \rangle = \langle Z_0^K, f_0 \rangle + \int_0^t \int_{\mathbb{R}_+} (\partial_s f_s(x) + \alpha \partial_x f_s(x)) Z_s^K(dx) ds
$$
$$
+ \frac{1}{K} \int_0^t \int_{\mathcal{E}} \mathbb{1}_{\{i\in V_{s_-}^K\}} (f_s(\gamma x_i(s-)) + f_s((1-\gamma)x_i(s_-)) - f_s(x_i(s_-))) Q(ds,di,d\gamma),
$$
$$
= \langle Z_0^K, f_0 \rangle + M_t^{K,f}
$$
$$
+ \int_0^t \int_{\mathbb{R}_+} \left(\partial_s f_s(x) + \alpha \partial_x f_s(x) + R \int_0^1 (f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)) h(\gamma) d\gamma \right) Z_s^K(dx) ds,
$$
$$
\text{(A2)}
$$

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

$$
\langle M^{K,f} \rangle_t = \frac{1}{K} \int_0^t \int_{\mathbb{R}_+} \int_0^1 R(f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x))^2 h(\gamma) d\gamma Z_s^K(dx) ds. \quad \text{(A3)}
$$

The proof of Proposition 1 then follows the ideas in (Tran, 2006, 2008) and are detailed in Hoang (2016). Equation (A2) corresponds to Equation (3) in the main body.

The proof of Theorem 1 uses the martingale problem established in Proposition 1 and standard arguments (e.g., Bansaye & Méléard, 2015; Ethier & Kurtz, 2009; Joffe & Métivier ,1986; Tran, 2014, theorem 1.1.8 and proof of theorem 1.1.11). Let us denote by $A^{K,f}$ the finite variation part of $\langle Z^K_t, f_t \rangle$:

$$A^{Kf}_t = \int_0^t \int_{\mathbb{R}_+} \left(\partial_s f_s(x) + \alpha \partial_x f_s(x) + R \int_0^1 (f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x))\, h(\gamma)d\gamma \right) Z^K_s(dx)ds. \quad \text{(A4)}$$

First, using the moment assumptions together with (A2) and (A3), we can show that the sequences of real-valued processes $(A^{K,f})_{K\in\mathbb{N}^*}$ 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 $(\langle Z^K_\cdot, f\rangle)_{K\in\mathbb{N}^*}$ for all test function $f \in C^1_b(\mathbb{R}_+, \mathbb{R})$. 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 $\overline{Z}$ to which subsequences of $(Z^K)_{K\in\mathbb{N}^*}$ converge vaguely, are continuous measure-valued processes of $C(\mathbb{R}_+, (\mathcal{M}_F(\mathbb{R}_+), w))$, where $\mathcal{M}_F(\mathbb{R}_+)$ is embedded with the weak convergence topology.

Thirdly, proceeding as in Tran (2014, proof of theorem 1.1.11) (see also Jourdain et al., 2012 and Méléard & Tran, 2012), we can prove that

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

where the functions φ_k are C^2 approximations of $1\!\!1_{\{x \geq k\}}$ 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 $\overline{Z}$, their masses converge in distribution to $\langle \overline{Z}, 1 \rangle$, which provides the tightness in $(\mathcal{M}_F(\mathbb{R}_+), w)$ by a criterion due to Méléard and Roelly (1993).

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 (5) (see Hoang, 2016). This integro-differential equation admits a unique solution. Indeed, let ξ^1 and ξ^2 be two solutions of (5) starting with the same initial condition ξ_0. For a test function $\varphi \in C^1_b(\mathbb{R}_+, \mathbb{R})$ and $t > 0$, setting

$$f(x,s) = f_s(x) = \varphi(x + \alpha(t-s)), \quad \text{(A5)}$$

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

$$\langle \xi^i_t, \varphi \rangle = \langle \xi_0, \varphi(. + \alpha t) \rangle + \int_0^t \int_{\mathbb{R}_+} \int_0^1 R\left(f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right) h(\gamma)d\gamma\, \xi^i_s(dx)\, ds.$$

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

$$\|\xi^1_t - \xi^2_t\|_{TV} \leq 3R\|\varphi\|_\infty \int_0^t \|\xi^1_s - \xi^2_s\|_{TV}ds,$$

where $\|.\|_{TV}$ stands for the total variation norm. Gronwall's inequality concludes the proof of uniqueness of the solution of (5). 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 Hoang (2016) and see also Tran (2008).

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 $\varphi \in C^1(\mathbb{R}_+, \mathbb{R}_+)$ with nonnegative values, let us define the test function $f(x, s)$ as in (A5). Then, neglecting the negative terms in the second line of (5) and using the symmetry of h with respect to $1/2$:

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

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. Let us denote by $n(t, x)$ the density of ξ_t with respect to the Lebesgue measure dx on $\mathbb{R}_+$. Then, for a nonnegative test function $f \in C_b^1(\mathbb{R}_+, \mathbb{R}_+)$ depending only on x and using the symmetry of h, (5) becomes:

$$
\begin{aligned}
\int_{\mathbb{R}_+} f(x) n(t, x) dx &= \int_{\mathbb{R}_+} f(x) n_0(x) dx + \int_0^t \int_{\mathbb{R}_+} \alpha f'(x) n(s, x) dx \, ds \\
&\quad + \int_0^t \int_{\mathbb{R}_+} 2R \int_0^1 f(\gamma x) h(\gamma) d\gamma \, n(s, x) dx \, ds - R \int_0^t \int_{\mathbb{R}_+} f(x) n(s, x) dx \, ds \\
&= \int_{\mathbb{R}_+} f(x) n_0(x) dx + \int_0^t \int_{\mathbb{R}_+} \alpha f'(x) n(s, x) dx \, ds \\
&\quad + \int_0^t \int_{\mathbb{R}_+} f(x) \times 2R \int_x^{+\infty} h\left(\frac{x}{y}\right) n(s, y) \frac{dy}{y} \, dx \, ds - R \int_0^t \int_{\mathbb{R}_+} f(x) n(s, x) dx \, ds,
\end{aligned}
\tag{A6}
$$

where we used Fubini's Theorem for the third term in the right hand side. For the second term in the right-hand side, integrating by part gives:

$$
\begin{aligned}
\int_0^t \int_{\mathbb{R}_+} \alpha f'(x) n(s, x) dx \, ds &= \int_0^t \left\{ [\alpha f(x) n(s, x)]_0^{+\infty} - \int_{\mathbb{R}_+} \alpha f(x) \partial_x n(s, x) dx \right\} ds \\
&= -\alpha f(0) \int_0^t n(s, 0) ds - \int_0^t \int_{\mathbb{R}_+} \alpha f(x) \partial_x n(s, x) dx \, ds.
\end{aligned}
\tag{A7}
$$

Gathering (A6) and (A7) that are true for any test function f and time t, we can identify the equations satisfied by $(n(t, x), x \in \mathbb{R}_+, t > 0)$. We find that $n(t, 0) = 0$ for every $t \geq 0$ and that

$(n(t, x), x \in \mathbb{R}_+, t > 0)$ solves in distribution sense (6) for which uniqueness of the solution holds (e.g., Perthame, 2007, theorem 4.3, p. 90).

The proof of Proposition 3 is a particular case of Perthame (2007, theorem 4.6, p. 94) based on Krein-Rutman theorem (e.g., Perthame, 2007, theorem 6.5, p. 175; see also Doumic & Gabriel, 2010). In the case that we consider, the proof can be simplified compared with Perthame (2007). Let us consider the eigen elements (λ, N, ϕ) associated with (6), that is, the solution of:

$$\begin{cases} \alpha \partial_x N(x) + (\lambda + R)N(x) = 2R \int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma}, & x \geq 0, \\ N(0) = 0, \quad \int N(x) dx = 1, \quad N(x) \geq 0, \quad \lambda > 0, \\ \alpha \partial_x \phi(x) - (\lambda + R)\phi(x) = -2R \int_0^1 \phi(\gamma x) h(\gamma) d\gamma, & x \geq 0, \\ \phi(x) \geq 0, \quad \int_0^{+\infty} \phi(x) N(x) dx = 1. \end{cases} \tag{A8}$$

It is clear that $\lambda = R$ and $\phi \equiv 1$ solve the third equation of (A8). 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 $f \in C(\mathbb{R}_+, \mathbb{R}_+)$, and $M \in \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}) \cap C(\mathbb{R}_+, \mathbb{R}_+)$:

$$\alpha \partial_x N(x) + (\mu + R) N(x) - 2R \int_0^1 M\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = f(x), \quad x \geq 0 \quad ; \quad N(0) = 0. \tag{A9}$$

Equation (A9) is a first-order ordinary differential equation (ODE) that can be solved with the variation of constant method. It admits a unique solution, that we denote $T(M)$:

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

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

$$|T(M_1)(x) - T(M_2)(x)| \leq \frac{2R}{\alpha} \int_0^1 \int_0^{x/\gamma} e^{-\frac{\mu+R}{\alpha}(x-\gamma z)} |M_1(z) - M_2(z)| h(\gamma) dz \, d\gamma$$

$$\leq \frac{2R}{\mu + R} \left(1 - e^{-\frac{(\mu+R)x}{\alpha}}\right) \int_0^1 \frac{h(\gamma)}{\gamma} d\gamma \|M_1 - M_2\|_\infty. \tag{A10}$$

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

$$\alpha \partial_x N(x) + (\mu + R) N(x) - 2R \int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = f(x), \quad x \geq 0 \quad ; \quad N(0) = 0. \tag{A11}$$

Step 2: The map A that associates to $f \in C(\mathbb{R}_+, \mathbb{R}_+) \cap \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ the unique corresponding solution of (A11) is thus well defined. Following the path of Perthame (2007, section 6.6.2), we can show that this map is linear, continuous (with computation similar to (A10)) 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 (using similar truncations as in Perthame, 2007) to obtain that the spectral radius of A, $\rho(A)$, is a positive simple eigenvalue associated with a positive eigenvector satisfying:

$$\alpha \partial_x N(x) + \left(\mu + R - \frac{1}{\rho(A)} \right) N(x) - 2R \int_0^1 N\left(\frac{x}{\gamma}\right) h(\gamma) \frac{d\gamma}{\gamma} = 0, \quad x \geq 0 \quad ; \quad N(0) = 0. \quad \text{(A12)}$$

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 (9), are obtained by generalizing the proof of Perthame (2007, theorem 4.2, p. 88) (see also Perthame & Ryzhik, 2005). 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 $\partial_t \int_0^{+\infty} \left| G(t,x)e^{Rt} \right| dx \leq 0$. As a consequence,

$$\int_0^{+\infty} |G(t,x)|dx \leq e^{-Rt} \|G_0\|_1. \quad \text{(A13)}$$

From the PDE of g, $K(0,x) = \partial_t G(t,x)|_{t=0} = 2R \int_0^1 G_0(x/u)h(u)du - 2RG_0(x) - \alpha g_0(x)$. Proceeding similarly as for G, we show that

$$\int_0^{+\infty} |K(t,x)|dx \leq e^{-Rt} \int_0^{+\infty} |K(0,x)|dx \leq e^{-Rt} \left(3R\|G_0\|_1 + \alpha\|g_0\|_1 \right). \quad \text{(A14)}$$

Plugging (A13) and (A14) 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.

APPENDIX B. PROOF OF PROPOSITION 4

Proof of Proposition 4 (i). Let $\epsilon > 0$ to be chosen small enough. Since N is a probability density, we have for $v \leq (v_0 + 2) \wedge ([\beta] + 1)$:

$$\int_0^{+\infty} x^{-v} N(x)dx \leq \int_0^{\epsilon} x^{-v} N(x)dx + \frac{1}{\epsilon^v}.$$

Hence, it remains to prove

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

We follow and adapt the steps of the proof of theorem 1 in Doumic and Gabriel (2010). Integrating both side of Equation (8) between 0 and $x_0 \leq x$, we get:

$$\alpha N(x_0) + 2R \int_0^{x_0} N(y)dy = 2R \int_0^{x_0} \int_0^{+\infty} N(y)h\left(\frac{z}{y}\right) \frac{dy}{y} dz. \quad \text{(B1)}$$

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

$$f(x) \le \frac{2R}{\alpha} \int_0^{x} \int_0^{+\infty} N(y) h\left(\frac{z}{y}\right) \frac{dy}{y} dz. \tag{B2}$$

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

$$\int_0^{t} h(x) dx \le C \int_0^{t} x^{(\nu_0+1) \wedge [\beta]} dx \le C t^{(\nu_0+2) \wedge ([\beta]+1)} \le C t^{\nu}, \tag{B3}$$

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

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

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

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

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

$$F(x) \le \frac{K}{1 - \frac{2R\epsilon}{\alpha}} + \frac{2CR}{\alpha - 2R\epsilon} \int_x^{\epsilon} F(y) dy. \tag{B4}$$

Then, applying Gronwall's inequality to (B4), we obtain

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

and

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

We finally obtain

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

This ends the proof of Proposition 4(i). $\qquad\square$

Proof of Proposition 4(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

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

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

$$\begin{aligned}
\frac{\Phi(x+\epsilon) - \Phi(x)}{\epsilon} &= -\frac{1}{\epsilon}\int_x^{x+\epsilon} N(y) h\left(\frac{x}{y}\right)\frac{dy}{y} + \int_{x+\epsilon}^{+\infty} N(y)\frac{y}{\epsilon}\left[h\left(\frac{x+\epsilon}{y}\right) - h\left(\frac{x}{y}\right)\right]\frac{dy}{y^2} \\
&\to_{\epsilon\to 0} -\frac{N(x)h(1)}{x} + \int_x^{+\infty} N(y) h'\left(\frac{x}{y}\right)\frac{dy}{y^2} = \int_x^{+\infty} N(y) h'\left(\frac{x}{y}\right)\frac{dy}{y^2}, \quad (B5)
\end{aligned}$$

since $h(1) = h(0) = 0$ by Assumption 1. This shows that Φ is of class C^1. Plugging this information into (8), it follows that $\partial_x N$ is continuous, and hence N is of class C^1 and thus $\partial_x N$ also. This entails from the computation of Φ' that Φ is of class 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 C^{k-1} for $k \leq [\beta] \wedge \nu_0$. Then, since the successive derivatives of h at 0 vanish by Assumption 1,

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

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 C^k and that using this information in (8), $\partial_x N$ is of class C^{k-1} entailing that N is of class 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 $C^{[\beta]}$, and thus, $u \mapsto M(u)$ is also of class $C^{[\beta]}$.

Take $k \leq [\beta]$. That M is of class $C^{[\beta]}$ 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:

$$\int_{\mathbb{R}} e^{(\ell+1)u} \left|N^{(\ell)}(e^u)\right| du = \int_0^{+\infty} v^\ell \left|N^{(\ell)}(v)\right| dv. \quad (B6)$$

It is known (as a direct adaptation of Perthame, 2007, theorem 4.6, p. 95 for example) that as soon as $\mu < R/\alpha$,

$$N(x)e^{\mu x} \in \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+) \cap \mathbb{L}^\infty(\mathbb{R}_+, \mathbb{R}_+). \quad (B7)$$

Assume that for some $\ell < k$, we have proved that $\int_0^{+\infty} e^{\mu x} |N^{(\ell)}(x)| dx < +\infty$ for $\mu < R/\alpha$. Let us prove that this also holds for $\ell + 1$, which would entail (B6). Deriving (8) ℓ times, multiplying by $e^{\mu x}$ and integrating again in $x \in (0, +\infty)$, we obtain:

$$\alpha \int_0^{+\infty} e^{\mu x} |N^{(\ell+1)}(x)| dx \leq 2R \int_0^{+\infty} e^{\mu x} |N^{(\ell)}(x)| dx + 2R \int_0^{+\infty} e^{\mu x} \left| \int_x^{+\infty} h^{(\ell)} \left(\frac{x}{y} \right) \frac{N(y)}{y^{\ell+1}} dy \right| dx$$

$$\leq 2R \int_0^{+\infty} e^{\mu x} |N^{(\ell)}(x)| dx + 2R \int_0^{+\infty} \frac{N(y) e^{\mu y}}{y^\ell} \|h^{(\ell)}\|_1 dy. \tag{B8}$$

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. Using that $N(x) e^{\mu x} \in \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ implies that the second term is integrable at $+\infty$. Point (i) of Proposition 4 ensures the integrability at 0. Thus, the right-hand side of (B8) is finite. The use of point (i) of Proposition 4 explains why $[\beta] \wedge (v_0 + 3)$ appears in the announced result.

The finiteness of $\int_0^{+\infty} e^{\mu x} |N^{(\ell)}(x)| dx$, for $\ell \leq [\beta] \wedge (v_0 + 3)$, is thus proved by recursion, implying the finiteness of the terms in (B6) and concluding the proof. □

Proof of Proposition 4(iii). The proof is divided into several steps.

Step 1: First, notice that

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

We first prove that M^* has isolated zeros in $\{\xi \in \mathbb{C} : \mathfrak{I}(\xi) < 1\}$.

Because N is such that $e^{\mu x} N(x) \in \mathbb{L}^\infty(\mathbb{R}_+, \mathbb{R}_+) \cap \mathbb{L}^1(\mathbb{R}_+, \mathbb{R}_+)$ for $\mu < R/\alpha$ (see Perthame, 2007, p. 95), M^* is well defined on the half plane $\{\xi \in \mathbb{C} : \mathfrak{I}(\xi) \leq 0\}$. By Proposition 4(i), $N(y)/y$ is integrable on the neighborhood of zero. Hence, for $\xi_2 < 1$, so is $y^{-\xi_2} N(y)$. As a consequence, M^* is analytic on the lower half-plane $\{\xi \in \mathbb{C} : \mathfrak{I}(\xi) < 1\}$ which contains the real axis. The derivative of the integrand with respect to ξ has modulus $|\log(y)| y^{-\xi_2} N(y)$. The latter function is upper bounded, when $\xi_2 < 1$ and when $y < 1$, by $\frac{N(y)}{y}$ which is integrable on the neighborhood of zero by Proposition 4(i). It follows from the results on integrals with parameters that the extension of M^* to the complex plane is holomorphic on $\{\xi \in \mathbb{C} : \mathfrak{I}(\xi) < 1\}$. Because M^* is not the null function, its zeros in $\{\xi \in \mathbb{C} : \mathfrak{I}(\xi) < 1\}$ have to be isolated.

Step 2: Let us now show that M^* admits no zero on the real line. First we link the Fourier transform to the Mellin transform of N. Recall that M^* is related to the solution N of (8) where the right-hand term is a multiplicative convolution of $x \mapsto N(x)/x$ and h. A natural way of treating multiplicative deconvolution is by using Mellin transform (e.g., Epstein, 1948; Titchmarsh, 1948).

It is natural to set

$$\psi(x) = \frac{N(x)}{x}, \quad x \geq 0. \tag{B9}$$

The Mellin transforms of ψ and h are defined for $s \in \mathbb{C}$ as:

$$\Psi(s) = \int_0^{+\infty} x^{s-1} \psi(x) dx, \quad \text{and} \quad H(s) = \int_0^1 x^{s-1} h(x) dx. \tag{B10}$$

Notice that for $\xi \in \mathbb{R}$,

$$M^*(\xi) := \int_{-\infty}^{\infty} M(x)e^{ix\xi}dx = \int_{-\infty}^{\infty} e^x N(e^x)e^{ix\xi}dx = \int_0^{+\infty} N(u)u^{i\xi}du = \Psi(i\xi + 2). \quad \text{(B11)}$$

So the Point (iii) of Proposition 4 will be proved if Ψ does not vanish on the line $2 + i\mathbb{R}$. We first establish an equation satisfied by Ψ (Step 3) and then conclude with the proof by adapting the results of Doumic et al. (2017).

Recall (B7). Let us prove below that for any $s \in \mathbb{C}$ such that

$$\lim_{x \to 0} x^{s-1} N(x) = 0, \quad \text{(B12)}$$

the function Ψ satisfies:

$$(1 - s)\Psi(s) = \frac{2R}{\alpha}(H(s) - 1)\Psi(s + 1). \quad \text{(B13)}$$

Notice firstly that (B13) is in particular true for all s such that $\Re(s) > 1$ since (B12) is then satisfied. Secondly, let us add that (B13) is reminiscent of equation (2.4) in Doumic et al. (2017), with the difference that our equation is establish on the eigenvalue problem (8) with constant branching rate R, while Doumic and co-authors work with an evolution equation with no evolution of the cell sizes but power-law branching rate.

To prove (B13), observe that

$$N(x) = x\psi(x) \quad \text{and} \quad \partial_x N(x) = x\psi'(x) + \psi(x).$$

Therefore, (8) can be expressed by using ψ instead of N:

$$\alpha \partial_x N(x) + 2R\, N(x) = 2R \int_0^{\infty} N(y)h\left(\frac{x}{y}\right)\frac{dy}{y}, \quad x \geq 0$$

$$\Leftrightarrow x\psi'(x) + \left(1 + \frac{2R}{\alpha}x\right)\psi(x) = \frac{2R}{\alpha}\int_x^{+\infty} y\psi(y)h\left(\frac{x}{y}\right)\frac{dy}{y}, \quad x \geq 0.$$

Multiplying each side of this equation by x^{s-1} and integrating on $x \in [0, +\infty)$, we obtain:

$$\int_0^{+\infty} x^s \psi'(x)dx + \int_0^{+\infty}\left(1 + \frac{2R}{\alpha}x\right)x^{s-1}\psi(x)dx = \frac{2R}{\alpha}\int_0^{+\infty} x^{s-1}\int_x^{+\infty} y\psi(y)h\left(\frac{x}{y}\right)\frac{dy}{y}\,dx$$

$$\Leftrightarrow (1 - s)\Psi(s) = \frac{2R}{\alpha}(H(s) - 1)\Psi(s + 1),$$

by using an integration by parts for the left term and the Fubini theorem for the right term. This shows (B13).

Step 3: We now prove that for any $\xi \in \mathbb{R}$, and any $n \in \mathbb{N}$,

$$M^*(\xi) \neq 0 \quad \Leftrightarrow \quad \Psi(i\xi + n + 2) \neq 0. \quad \text{(B14)}$$

This is true for $n = 0$, by (B11). For $\xi \in \mathbb{R}$, applying (B13) with $s = 2 + i\xi$, for which (B12) is satisfied:

$$M^*(\xi) = \Psi(i\xi + 2) = -\frac{2R}{\alpha(1 + i\xi)} \left(H(i\xi + 2) - 1 \right) \Psi(i\xi + 3).$$

By induction, we have for any $n \in \mathbb{N}$ (n will be chosen subsequently),

$$M^*(\xi) = \left(-\frac{2R}{\alpha} \right)^n \prod_{k=1}^{n} \frac{1}{k + i\xi} \left(H(i\xi + k + 1) - 1 \right) \times \Psi(i\xi + n + 2).$$

Since for any $\xi = \xi_1 + i\xi_2 \in \mathbb{C}$ such that $\xi_1 \geq 1$,

$$|H(i\xi_2 + \xi_1 + 1)| = \left| \int_0^1 x^{i\xi_2 + \xi_1} h(x) dx \right| \leq \int_0^1 x h(x) dx = \frac{1}{2}, \tag{B15}$$

by symmetry of h, then the product in the right-hand side is nonzero and this leads to (B14).

As a consequence, to prove that M^* admits no zero on the real line, it is sufficient to prove that Ψ admits no zero on some line $n_0 + i\mathbb{R}$ where n_0 is an integer larger than 2.

Step 4: Following the computation in Doumic et al. (2017), we can prove that:

$$\forall s \in \mathbb{C} \text{ with } \Re(s) \geq 2, \quad |\Psi(s)| \neq 0. \tag{B16}$$

First, notice that (B15) implies that $H(s) - 1 \neq 0$ as soon as $\Re(s) \geq 2$. For such $s \in \mathbb{C}$, dividing by $H(s) - 1$, we can reformulate (B13) as:

$$\Psi(s + 1) = \Phi(s)\Psi(s), \quad \text{where} \quad \Phi(s) = \frac{\alpha}{2R} \times \frac{1 - s}{H(s) - 1}. \tag{B17}$$

Notice that our equation has much more regularities than the one studied in Doumic et al. since the application Φ is analytic on $\{s \in \mathbb{C}, \Re(s) \geq 2\}$ and does not vanish on this half-plane.

Let us fix $s_0 \in \mathbb{R}$ such that $s_0 \geq 2$. For $s \in \mathbb{C}$ such that $s_0 < \Re(s) < s_0 + 1$, we look for particular solutions of (B17) of the form

$$\Psi(s) = \exp\left(P(\zeta(s)) \right) \quad \text{where} \quad \zeta(s) = e^{i2\pi(s - s_0)}.$$

Then, following the steps of proposition 2 in Doumic et al. (2017), the function P solves a Carleman equation, from which a solution of (B17) can be obtain. There exist several solutions to (B17), and computing the inverse Mellin transforms of the latters, it appears that the solution corresponding to the Mellin transform of the solution of (8) is:

$$\Psi(s) = \exp\left(-\int_{\Re(\sigma) = s_0} \log\left(\Phi(\sigma) \right) \left[\frac{1}{1 - e^{2i\pi(s - \sigma)}} - \frac{1}{1 + e^{2i\pi(s_0 - \sigma)}} \right] d\sigma \right), \tag{B18}$$

for $s \in \{\Re(s) \in (s_0, s_0 + 1)\}$, where the chosen determination of the logarithm is $\log(z) = \log|z| + i \arg(z)$ with $\arg(z) \in [0, 2\pi)$. That the right-hand side of (B18) is well defined and does

not vanish on $\{\mathfrak{R}(s) \in (s_0, s_0 + 1)\}$ follows closely the proofs in Doumic et al. (2017, lemma 3 and section 6.3) by careful study of the behavior of $\Phi(s)$ when the imaginary part of s tend to $\pm\infty$.

Conclusion: choose for example $s_0 = \frac{5}{2}$ so that $s_0 \geq 2$ and $n_0 = 3 \in (s_0, s_0 + 1)$. The function given in (B18) is analytic, nonvanishing and coincide with the Mellin transform of ψ on $\{s \in \mathbb{C} : \mathfrak{R}(s) \in (s_0, s_0 + 1)\}$. This implies that Ψ admits no zero on $n_0 + i\mathbb{R}$ and finishes the proof. $\square$

APPENDIX C. PROOF OF PROPOSITIONS 5 AND 6

Proof of Proposition 5. We have

$$\|\hat{g}_{n,\ell} - g\|_2^2 = \int_{\mathbb{R}_-} \left(\hat{g}_{n,\ell}(u) - g(u)\right)^2 du = \int_{\mathbb{R}_-} \left(e^u \hat{h}_{n,\ell}(e^u) - e^u h(e^u)\right)^2 du$$

$$= \int_0^1 \left(\hat{h}_{n,\ell}(x) - h(x)\right)^2 x \, dx. \tag{C1}$$

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

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

Thus,

$$\mathbb{E}\left[\|\check{g}_{n,\ell} - g\|_2^2\right] = \mathbb{E}\left[\|\tau \hat{g}_{n,\ell} + (1 - \tau)\tilde{g}_{n,\ell} - g\|_2^2\right]$$

$$= \frac{1}{2}\mathbb{E}\left[\|\hat{g}_{n,\ell} - g\|_2^2\right] + \frac{1}{2}\mathbb{E}\left[\|\tilde{g}_{n,\ell} - g\|_2^2\right] = \frac{1}{2}\mathbb{E}\left[\|\hat{h}_{n,\ell} - h\|_2^2\right], \tag{C2}$$

since $\|\hat{g}_{n,\ell} - g\|_2^2 + \|\tilde{g}_{n,\ell} - g\|_2^2 = \|\hat{h}_{n,\ell} - h\|_2^2$. Let us now compute $\mathbb{E}\left[\|\tilde{g}_{n,\ell} - g\|_2^2\right]$. 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

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

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

$$\mathbb{E}[\|\tilde{g}_{n,\ell} - g\|_2^2] = \mathbb{E}\left[\int_{\mathbb{R}_-} (\tilde{g}_{n,\ell}(u) - g(u))^2 du\right]$$

$$= \mathbb{E}\left[\int_{\mathbb{R}_-} \left(\frac{e^u}{1 - e^u}\right)^2 (\hat{g}_{n,\ell}(\log(1 - e^u)) - g(\log(1 - e^u)))^2 du\right]$$

$$= \mathbb{E}\left[\int_{\mathbb{R}_-} \left(\frac{1 - e^v}{e^v}\right)(\hat{g}_{n,\ell}(v) - g(v))^2 dv\right].$$

As a consequence, the middle term in (C2) is

$$\frac{1}{2}\mathbb{E}\left[\|\hat{g}_{n,\ell} - g\|_2^2\right] + \frac{1}{2}\mathbb{E}\left[\|\tilde{g}_{n,\ell} - g\|_2^2\right] = \mathbb{E}\left[\int_{\mathbb{R}_-} \frac{1}{2}\left(1 + \frac{1 - e^v}{e^v}\right)\left(\hat{g}_{n,\ell}(v) - g(v)\right)^2 dv\right]$$

$$= \mathbb{E}\left[\int_{\mathbb{R}_-} \frac{e^{-v}}{2}\left(\hat{g}_{n,\ell}(v) - g(v)\right)^2 dv\right].$$

This concludes the proof. $\qquad\qquad\square$

Proof of Proposition 6. Remember (C1). Then, since $h(x) = h(1 - x)$,

$$\int_0^1 \left(\hat{h}_{n,\ell}^{sym}(x) - h(x)\right)^2 m(x)dx$$

$$= \frac{1}{4}\int_0^1 \left(\hat{h}_{n,\ell}(x) - h(x) + \hat{h}_{n,\ell}(1 - x) - h(1 - x)\right)^2 m(x)dx$$

$$\leq \frac{1}{2}\int_0^1 \left(\hat{h}_{n,\ell}(x) - h(x)\right)^2 m(x)dx + \frac{1}{2}\int_0^1 \left(\hat{h}_{n,\ell}(1 - x) - h(1 - x)\right)^2 m(1 - x)dx$$

$$= \int_0^1 \left(\hat{h}_{n,\ell}(x) - h(x)\right)^2 m(x)dx$$

$$\leq \int_0^1 \left(\hat{h}_{n,\ell}(x) - h(x)\right)^2 xdx = \|\hat{g}_{n,\ell} - g\|_2^2.$$

This concludes the proof. $\qquad\qquad\square$

APPENDIX D. PROOF OF THEOREM 2

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

$$\|\hat{g}_{n,\ell} - g\|_2 \leq \|g_\ell - g\|_2 + \|\hat{g}_{n,\ell} - g_\ell\|_2.$$

The first term of the above right hand side 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 (15):

$$\|\hat{g}_{n,\ell} - g_\ell\|_2^2 = \frac{1}{2\pi}\|\hat{g}_{n,\ell}^* - g_\ell^*\|_2^2$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \left|K_\ell^*(\xi)\left[\left(\frac{\alpha\widehat{D_n^*(\xi)}}{2R}\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} + 1\right) - g^*(\xi)\right]\right|^2 d\xi$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \left|K_\ell^*(\xi)\left[\left(\frac{\alpha\widehat{D_n^*(\xi)}}{2R}\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \frac{\alpha\widehat{D_n^*(\xi)}}{2RM^*(\xi)} + \frac{\alpha\widehat{D_n^*(\xi)}}{2RM^*(\xi)} + 1\right) - g^*(\xi)\right]\right|^2 d\xi$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \left|\frac{\alpha}{2R}K_\ell^*(\xi)\widehat{D_n^*(\xi)}\left(\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \frac{1}{M^*(\xi)}\right) + K_\ell^*(\xi)\left(\frac{\alpha\widehat{D_n^*(\xi)}}{2RM^*(\xi)} + 1 - g^*(\xi)\right)\right|^2 d\xi$$

$$\leq C\int_{\mathbb{R}} \left|K_\ell^*(\xi)\widehat{D_n^*(\xi)}\left(\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \frac{1}{M^*(\xi)}\right)\right|^2 d\xi + C\int_{\mathbb{R}} |K_\ell^*(\xi)^2|\left|\frac{\alpha\widehat{D_n^*(\xi)}}{2RM^*(\xi)} + 1 - g^*(\xi)\right|^2 d\xi$$

$$:= I + II.$$

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

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

For the term II, because

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

we have

$$\begin{aligned}
\mathbb{E}[\text{II}] &= C\int_{\mathbb{R}}\mathbb{V}\text{ar}\left(K_\ell^*(\xi)\left(\frac{\alpha\widehat{D_n^*(\xi)}}{2RM^*(\xi)} + 1\right)\right)d\xi \\
&\leq C\int_{\mathbb{R}}\mathbb{V}\text{ar}\left(K_\ell^*(\xi)\frac{\widehat{D_n^*(\xi)}}{M^*(\xi)}\right)d\xi \\
&\leq C\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right|^2\mathbb{V}\text{ar}\left(\frac{(-\mathbf{i}\xi)}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi-1)U_j}\right)d\xi \\
&\leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{V}\text{ar}(e^{(\mathbf{i}\xi-1)U_1})d\xi \\
&\leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{E}\left[|e^{(\mathbf{i}\xi-1)U_1}|^2\right]d\xi \leq \frac{C}{n}\int_{\mathbb{R}}\left|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right|^2\mathbb{E}[e^{-2U_1}]d\xi \\
&\leq \frac{C}{n}\left\|\frac{K_\ell^*(\xi)\xi}{M^*(\xi)}\right\|_2^2,
\end{aligned}$$

since $\mathbb{E}[e^{-2U_1}] = \int_0^{+\infty}x^{-2}N(x)dx < +\infty$ thanks to Proposition 4.

We now set

$$\triangle(\xi) := \frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \frac{1}{M^*(\xi)}. \tag{D1}$$

Then we get

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

To control the term IV, we need the following lemma whose proof is postponed in Appendix E.

Lemma 1. *There exists a positive constant C_p such that*

$$\mathbb{E}[|\triangle(\xi)|^{2p}] \leq C_p \min\left\{\frac{1}{|M^*(\xi)|^{2p}}, \frac{n^{-p}}{|M^*(\xi)|^{4p}}\right\} \quad \text{for } p = 1, 2. \tag{D2}$$

Since $\hat{D}_n^*$ is an unbiased estimator of D^* using Lemma 1 with $p = 1$ we get

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

Moreover, from equation (12) and using that g^* is the Fourier transform of a density function, we get

$$\left|\frac{D^*(\xi)}{M^*(\xi)}\right| \leq \frac{2R}{\alpha}(|g^*(\xi)| + 1) \leq \frac{4R}{\alpha}.$$

Thus we obtain

$$\text{IV} \leq \frac{C}{n} \int_{\mathbb{R}} \left|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right|^2 d\xi = \frac{C}{n} \left\|\frac{K_\ell^*(\xi)}{M^*(\xi)}\right\|_2^2.$$

For the term III, we have by applying Cauchy–Schwarz's inequality and by Lemma 1:

$$\text{III} \leq C \int_{\mathbb{R}} \left|K_\ell^*(\xi)\right|^2 \left(\mathbb{E}\left[\left|\widehat{D_n^*(\xi)} - \mathbb{E}\left[\widehat{D_n^*(\xi)}\right]\right|^4\right]\right)^{1/2} \left(\mathbb{E}\left[|\triangle(\xi)|^4\right]\right)^{1/2} d\xi$$

$$\leq C \int_{\mathbb{R}} \left|K_\ell^*(\xi)\xi\right|^2 \left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n e^{(\mathbf{i}\xi - 1)U_j} - \mathbb{E}\left[e^{(\mathbf{i}\xi - 1)U_1}\right]\right|^4\right]\right)^{1/2}$$

$$\times \min\left\{\frac{1}{|M^*(\xi)|^4}, \frac{n^{-2}}{|M^*(\xi)|^8}\right\}^{1/2} d\xi$$

$$\leq C \int_{\mathbb{R}} \frac{\left|K_\ell^*(\xi)\xi\right|^2}{|M^*(\xi)|^2} \left(\mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j(\xi)\right|^4\right]\right)^{1/2} d\xi,$$

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

$$\mathbb{E}[|Z_1(\xi)|^4] = \mathbb{E}\left[\left|e^{(\mathbf{i}\xi - 1)U_1} - \mathbb{E}\left[e^{(\mathbf{i}\xi - 1)U_1}\right]\right|^4\right] \leq \mathbb{E}\left[\left(\left|e^{(\mathbf{i}\xi - 1)U_1}\right| + \left|\mathbb{E}\left[e^{(\mathbf{i}\xi - 1)U_1}\right]\right|\right)^4\right]$$

$$\leq 2^3 \left(\mathbb{E}\left[\left|e^{(\mathbf{i}\xi - 1)U_1}\right|^4\right] + \left|\mathbb{E}\left[e^{(\mathbf{i}\xi - 1)U_1}\right]\right|^4\right)$$

$$\leq 16\mathbb{E}\left[e^{-4U_1}\right] = 16 \int_0^{+\infty} x^{-4} N(x) dx < +\infty, \tag{D3}$$

by Proposition 4, applying Rosenthal inequality to real and imaginary parts of complex variables Z_j's, we get

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

Hence

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

Finally, we obtain

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

This ends the proof of Theorem 2. □

Proof of Proposition 7. First let us find an upper bound for the bias term $\|K_\ell \star g - g\|_2^2$. We assume that g^* is integrable and g belongs to the Sobolev class $S(\beta, L)$. Proposition 1 of Comte and Lacour (2013) yields

$$\|K_\ell \star g - g\|_2^2 \leq L^2 \ell^{2\beta}.$$

Now we shall consider the variance term. We have

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

Performing the usual trade-off between the bias and the variance terms, we get the following choice of the bandwith ℓ, for a fixed $\alpha > 0$:

$$\ell = \alpha n^{-\frac{1}{2\beta+2([\beta]\wedge(\nu_0+3))+3}},$$

which concludes the proof. □

APPENDIX E. PROOF OF TECHNICAL LEMMA

Proof of Lemma 1. This proof is inspired by the proof of Neumann (1997). 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 $\Omega_n(\xi) = \{|\widehat{M_n^*(\xi)}| \geq n^{-1/2}\}$ and $\mathbb{E}\left[\widehat{M_n^*(\xi)}\right] = \mathbb{E}\left[e^{\mathbf{i}\xi U_1}\right] = M^*(\xi)$, we have:

$$\mathbb{E}[|\triangle(\xi)|^2] = \mathbb{E}\left[\left|\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \frac{1}{M^*(\xi)}\right|^2\right] = \mathbb{E}\left[\left|\frac{\mathbb{1}_{\Omega_n(\xi)}}{\widehat{M_n^*(\xi)}} - \left(\frac{\mathbb{1}_{\Omega_n(\xi)}}{M^*(\xi)} + \frac{\mathbb{1}_{\Omega_n^c(\xi)}}{M^*(\xi)}\right)\right|^2\right]$$

$$= \frac{\mathbb{P}(\Omega_n^c(\xi))}{|M^*(\xi)|^2} + \mathbb{E}\left[\mathbb{1}_{\Omega_n(\xi)}\frac{|\widehat{M_n^*(\xi)} - M^*(\xi)|^2}{|\widehat{M_n^*(\xi)}|^2|M^*(\xi)|^2}\right]. \tag{E1}$$

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

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

But

$$\mathbb{E}[|\widehat{M_n^*(\xi)} - M^*(\xi)|^2] = \mathbb{V}\text{ar}[\widehat{M_n^*(\xi)}] = \mathbb{V}\text{ar}\left[\frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}\xi U_j}\right]$$

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

Hence we obtain

$$\mathbb{E}[|\bigtriangleup(\xi)|^2] \leq \frac{C}{|M^*(\xi)|^2} \leq C\min\left\{\frac{1}{|M^*(\xi)|^2}, \frac{n^{-1}}{|M^*(\xi)|^4}\right\}, \tag{E2}$$

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

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

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

$$\mathbb{P}(\Omega_n^c(\xi)) = \mathbb{P}(|\widehat{M_n^*(\xi)}| < n^{-1/2}) = \mathbb{P}(|\widehat{M_n^*(\xi)}| < |M^*(\xi)| - |M^*(\xi)| + n^{-1/2})$$

$$\leq \mathbb{P}(|\widehat{M_n^*(\xi)} - M^*(\xi)| > |M^*(\xi)| - n^{-1/2})$$

$$\leq \mathbb{P}(|\widehat{M_n^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2). \tag{E3}$$

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

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

We have

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

and

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

Since $|M^*(\xi)| \leq 1$ for all $\xi \in \mathbb{R}$ because of M is a density function, we get by Bernstein inequality (cf. for instance Comte & Lacour, 2013, lemma 2, p. 20)

$$\mathbb{P}(|\widehat{M_n^*(\xi)} - M^*(\xi)| > |M^*(\xi)|/2) \leq 2\max\left\{\exp\left(-\frac{n|M^*(\xi)|^2}{16}\right), \exp\left(-\frac{n|M^*(\xi)|}{16}\right)\right\}$$

$$\leq 2\exp\left(-\frac{n|M^*(\xi)|^2}{16}\right)$$

$$\leq C\frac{n^{-1}}{|M^*(\xi)|^2}. \tag{E4}$$

We also have that

$$\frac{1}{|\widehat{M_n^*(\xi)}|^2} = \frac{|M^*(\xi)|^2}{|\widehat{M_n^*(\xi)}|^2|M^*(\xi)|^2} = \frac{\left|\widehat{M_n^*(\xi)} - \left(\widehat{M_n^*(\xi)} - M^*(\xi)\right)\right|^2}{|\widehat{M_n^*(\xi)}|^2|M^*(\xi)|^2}$$

$$\leq 2\left\{\frac{1}{|M^*(\xi)|^2} + \frac{|\widehat{M_n^*(\xi)} - M^*(\xi)|^2}{|\widehat{M_n^*(\xi)}|^2|M^*(\xi)|^2}\right\}. \tag{E5}$$

Thus, from (E1), (E3), and (E5) we have:

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

$$\leq C \left\{ \frac{n^{-1}}{|M^*(\xi)|^4} + \frac{\mathbb{E}\left[|\widehat{M_n^*(\xi)} - M^*(\xi)|^2\right]}{|M^*(\xi)|^4} + \frac{\mathbb{E}\left[|\widehat{M_n^*(\xi)} - M^*(\xi)|^4\right] n}{|M^*(\xi)|^4} \right\}. \qquad \text{(E6)}$$

To find an upper bound for $\mathbb{E}\left[|\widehat{M_n^*(\xi)} - M^*(\xi)|^4\right]$, recall that $T_j(\xi) = e^{\mathbf{i}\xi U_j} - \mathbb{E}\left[e^{\mathbf{i}\xi U_1}\right]$. By similar calculations as obtained (D3), we have $\mathbb{E}[|T_1(\xi)|^4] < +\infty$ as $|T_1(\xi)| \leq 2$. Thus we get by Rosenthal's inequality applied to real and imaginary parts of the sequence of independent centered variables $T_1(\xi), \dots, T_n(\xi)$:

$$\mathbb{E}\left[|\widehat{M_n^*(\xi)} - M^*(\xi)|^4\right] = \mathbb{E}\left[\left| \frac{1}{n} \sum_{j=1}^n T_j(\xi) \right|^4 \right]$$

$$\leq Cn^{-4}\left(n\mathbb{E}[|T_1(\xi)|^4] + \left(n\mathbb{E}[|T_1(\xi)|^2]\right)^2 \right) \leq Cn^{-2}.$$

Thus, from (E3) and (E6) we get

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

Furthermore

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

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

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

Combining the two cases, we obtain

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

This ends the proof of Lemma 1. □