

AN INVERSE PROBLEM IN CELL DYNAMICS: RECOVERING AN INITIAL DISTRIBUTION OF TELOMERE LENGTHS FROM MEASUREMENTS OF SENESENCE TIMES

JULES OLAYÉ* 

Abstract. Telomeres are repetitive sequences situated at both ends of the chromosomes of eukaryotic cells. At each cell division, they are eroded until they reach a critical length that triggers a state in which the cell stops to divide: the senescent state. In this work, we are interested in the link between the initial distribution of telomere lengths and the distribution of senescence times. We propose a method to retrieve the initial distribution of telomere lengths, using only measurements of senescence times. Our approach relies on approximating our models with transport equations, which provide natural estimators for the initial telomere lengths distribution. We investigate this method from a theoretical point of view by providing bounds on the errors of our estimators, pointwise and in all Lebesgue spaces. We also illustrate it with estimations on simulations, and discuss its limitations related to the curse of dimensionality.

Mathematics Subject Classification. 45K05, 45Q05, 35Q49, 35R30, 62G07, 92-10.

Received February 21, 2025. Accepted February 11, 2026.

1. INTRODUCTION

Biological motivation. Linear chromosomes of eukaryotic cells have repeated sequences of nucleotides called telomeres at each of their two ends. These regions are non-coding, and prevent fusion between chromosomes. At each cell division, for each chromosome, one of its two telomeres is slightly shortened because the enzyme responsible for the DNA replication is unable to copy the last nucleotides of the DNA. This phenomenon is called the *end-replication problem* [39]. As telomeres progressively shorten, a cell may eventually reach a critical threshold in which telomeres are not long enough to protect the coding regions of DNA from degradation. To prevent this issue, when the shortest telomere of a cell attains a certain length, which is around 27 base pairs for the yeast [33], the cell enters a state called *senescence* [1, 8, 24, 40]. This state is characterised by the fact that the cell stops dividing, thus avoiding the loss of coding DNA. Understanding this phenomenon may help to understand how cancer cells can emerge. The main reason is that cancer cells have mutated to be able to lengthen their telomeres, preventing them from senescence and allowing them to proliferate [34].

Presentation of the problem. Since the criterion for a cell to become senescent is given by the lengths of its telomeres, initial lengths distribution and fate of cell lineages/population are deeply connected. In [33],

Keywords and phrases. Inverse problem, integro-differential equation, transport equation, telomere, cell division.

Institut de Mathématiques de Toulouse, CNRS, Université de Toulouse, 31062 Toulouse, France.

*Corresponding author: jules.olaye@math.univ-toulouse.fr

this link was studied with the following approach: how do the initial telomere lengths influence the fate of cell lineages/populations? Specifically, the authors investigated how the parameters of the initial distribution affect the population half-life time. In the current paper, we aim at understanding this link with the opposite approach. To do so, we propose to solve the following inverse problem: recovering the initial distribution of telomere lengths when only the senescence times distribution of cell lineages is observed.

Informal description of the models. To address the issue presented above, we use two deterministic telomere shortening models, whose dynamics are inspired from the stochastic models developed in recent years [2, 5, 6, 8, 24, 28, 33]. These models represent the mean value of telomere lengths concentration in cell lineages. They can be derived from stochastic telomere shortening models by proceeding as in the proof of Theorem 2.2.7 in [27], in which a deterministic telomere shortening model is derived from a stochastic population model by computing the mean value of telomere lengths concentration. The first model corresponds to a toy model for mathematical investigations, representing the evolution of cell lineages with only one telomere, see (2.1). In this model, cells are structured according to the length of their telomere $x \in \mathbb{R}_+$. The second model is a more biologically relevant one, and represents cell lineages with several telomeres, see (2.5). Each chromosome in a cell has two telomeres, one at each end. Thus, denoting by $k \in \mathbb{N}^*$ the number of chromosomes of the species we study, cells are structured by the lengths of their telomeres $x \in \mathbb{R}_+^{2k}$. In each model, the quantities modelled by our equations are the telomere lengths density over time, and the senescence times distribution. We assume that cells divide at a constant rate, and that at each division, the telomere (in the model on \mathbb{R}_+) or one telomere by chromosome (in the model on \mathbb{R}_+^{2k}) is shortened. We model senescence by a cemetery state, and assume that cells enter this state when one of their telomeres has a length below a threshold. This threshold is set to 0 for simplicity, as any other threshold value is equivalent to 0 by translation of the trait space. These dynamics give us a system of two integro-differential equations, including the deterministic equation of a jump process as a first equation, and the rate at which one of its coordinates reaches 0 as a second equation.

Difficulty. The main difficulty in solving this problem is that we work with integro-differential equations, which are non-local equations. This non-locality implies that we lose information about the telomere lengths distribution when only the senescence times are observed. Indeed, the models involve an integral over possible telomere lengths at the entry into senescence, see the second lines of (2.1) and (2.5). Thus, the value of the telomere length that triggers senescence is not observable in both models. This loss of information implies that the inverse of the operator linking the initial telomere lengths distribution and the senescence times distribution, if it exists, cannot be continuous, see Remark 2.2. The inverse problem is thus ill-posed in the sense of Hadamard or Tikhonov, see Definitions 2.1.1 and 2.1.3 in [20]. This also implies that it is difficult to prove that the operator mentioned above is invertible.

Estimation strategy. The estimation strategy we use consists in adapting the approach developed in [3, 13] to our setting. This approach allows us to retrieve a well-posed inverse problem, at the cost of an approximation. First, we assume that the telomere shortening values are small and that the cell division rate is high. This is biologically relevant because the average telomere shortening value is much smaller than the initial telomere length, see Section 2.3. Then, we obtain approximations of our models by transport equations with an absorption into a cemetery state at the boundary of \mathbb{R}_+ or \mathbb{R}_+^{2k} . Finally, we use the method of characteristics to construct estimators of the initial distribution on these approximated models.

Review of the literature. Inverse problems in cell biology have been widely studied over the last few years [3, 7, 12, 13, 29]. Among all these works, our article is in line with the following works in protein depolymerisation [3, 13]. Precisely, as in these articles, we have a trait that progressively degrades over time, and our aim is to retrieve the initial distribution of the trait. In [13], the approximation used is a transport-diffusion equation (second-order approximation) instead of a transport equation (first-order approximation). The present work is thus more in line with [3] in which a first-order approximation is done, justified later in [13]. In our case, a second-order approximation can also be done to construct an estimator for the model in \mathbb{R}_+ . This study has been conducted by the author, and will be presented in an upcoming article.

Integro-differential equations such as those described by our models, see (2.1) and (2.5), have been well-studied in the literature [4, 15, 19, 23], and are often referred to as *Volterra* or *Lotka–Volterra equations*. However, the theoretical study of mathematical models representing telomere shortening is very recent, and still quite unexplored. The main articles we can cite are [5, 6] or [28]. In these works, the question of the existence of a stationary profile was addressed. Previously, studies were conducted from a biomathematical or numerical perspective [8, 16, 24, 32, 33, 36, 40]. In some of these works, model approximations have been studied and justified from a modelling perspective [32, 36]. However, to the best of our knowledge, theoretical results on the approximation of telomere shortening models have not yet been obtained.

Main contributions. Our first contribution is to provide qualitative bounds for the approximation errors of our models and the errors of our estimators, pointwise and in all Lebesgue spaces. Specifically, we obtain bounds on the approximation and estimation errors that decrease exponentially fast after a certain length and a certain time. This completes the approximation result obtained in [13], in a discrete model similar to ours. In their work, a result in L^2 -norm for the approximation error was obtained with a bound increasing linearly with the time. Our main idea to obtain these new bounds is to rewrite the approximation errors of the telomere lengths densities as the integral of several sub-errors, see (3.19). This integral representation allows us to highlight for all $0 \leq s \leq t$ how the error generated at time s still influences the error at time t . Then, thanks to a maximum principle, we show that this error dissipates exponentially fast.

Our second contribution concerns the extension of the estimation to a multidimensional setting, *i.e.*, to the case where cells have several telomeres. From a theoretical point of view, we show that the estimator is less straightforward to obtain, as it requires a careful control over the rate at which cells remain outside the cemetery. From a practical point of view, we highlight the difficulties related to the curse of dimensionality and extreme value theory.

Organisation of the paper. First, in Section 2, we present the models and the estimators we use, and state the main result of the article. Then, we prove the main result for the model with one telomere and the model with several telomeres in Sections 3 and 4 respectively. Thereafter, we illustrate in Section 5 the quality of our estimators on simulations and data. Finally, we discuss the limits and the prospects of our work in Section 6. Auxiliary statements are presented in the Appendix A, as well as the proof of certain results given during the paper.

2. PRESENTATION OF THE MODELS AND THE MAIN RESULT

This section is devoted to the presentation of the notations and the results of the paper. First, in Section 2.1, we define the two models that we study. Then, in Section 2.2, we introduce our estimators, and give the assumptions and the main result of this work. Finally, in Section 2.3, we explain our choice of studying a model with only one telomere, and discuss our assumptions.

2.1. The models

Our goal is to have models representing the microfluidic experiments presented in [41], as our data are derived from them, see Section 5.3. In these experiments, cell lineages are tracked over time in microcavities until their last cell becomes senescent. Biologists then observe the senescence times of each lineage. We therefore need models which describe the telomere lengths distribution over time in lineages, and the time at which their last cell becomes senescent.

In both of our models, we assume that cell division times are distributed according to an exponential law with parameter $\tilde{b} > 0$. We consider $\tilde{\delta} > 0$ the maximum shortening value, and $\tilde{g} : [0, \tilde{\delta}] \rightarrow \mathbb{R}_+$ a probability density function representing the distribution of telomere shortening values at each division. We assume that \tilde{g} has finite first and second moments, and introduce the following constants and functions

$$\forall i \in \{1, 2\} : \tilde{m}_i := \int_0^{\tilde{\delta}} u^i \tilde{g}(u) \, du, \quad \text{and} \quad \forall x \in \mathbb{R}_+ : \tilde{G}(x) := \int_0^{\min(x, \tilde{\delta})} \tilde{g}(s) \, ds.$$

We also consider a non-negative function $n_0 \in L^1(\mathbb{R}_+)$ verifying $\|n_0\|_{L^1(\mathbb{R}_+)} = 1$. This function represents the initial telomere length distribution in the model with one telomere, and the initial length distribution for each individual telomere in the model with several telomeres. The goal of this paper is, thus, to infer this initial distribution knowing the distribution of senescence times.

Model with one telomere. In the first model, we assume that each cell has only one telomere. This is not biologically realistic because a cell has always several telomeres, since each chromosome of a cell has two of them. However, this model helps us to gain mathematical intuition, and it is discussed in Section 2.3 how this model can be used in practice.

Cells are structured according to their telomere length that belongs to \mathbb{R}_+ . We consider the density of telomere with length $x \geq 0$ at time $t \geq 0$ when several lineages are tracked, denoted $n^{(1)}(t, x)$. We also consider the rate at which cell lineages enter senescence at time $t \geq 0$, denoted $n_{\partial}^{(1)}(t)$. In fact, $n_{\partial}^{(1)}$ can also be seen as the senescence times distribution by Remark 2.1 below. In this model, at each cell division, the telomere of the dividing cell is shortened by a random value distributed according to \tilde{g} . If after this shortening, its telomere length is still greater than 0, then the cell remains in the dynamics. Otherwise, the cell goes to the cemetery, *i.e.* becomes senescent. Thus, the probability that the length $y \geq 0$ of the telomere of the dividing cell becomes negative after shortening is $\int_{\min(y, \tilde{\delta})}^{\tilde{\delta}} \tilde{g}(v) dv = 1 - \tilde{G}(y)$. Recalling that n_0 is the initial telomere length distribution, we then consider the following system of integro-differential equations

$$\begin{cases} \partial_t n^{(1)}(t, x) = \tilde{b} \int_0^{\tilde{\delta}} n^{(1)}(t, x + v) \tilde{g}(v) dv - \tilde{b} n^{(1)}(t, x), & \forall t \geq 0, x \geq 0, \\ n_{\partial}^{(1)}(t) = \tilde{b} \int_0^{\tilde{\delta}} n^{(1)}(t, y) (1 - \tilde{G}(y)) dy, & \forall t \geq 0, \\ n^{(1)}(0, x) = n_0(x), & \forall x \geq 0. \end{cases} \tag{2.1}$$

This system corresponds to the above description, and is our first model for telomere shortening. When we study this model, our aim is to estimate n_0 when $n_{\partial}^{(1)}$ is observed.

Remark 2.1. By integrating both lines of (2.1), we have conservation of the number of individuals, *i.e.* that for all $t \geq 0$

$$\int_0^{+\infty} n^{(1)}(t, x) dx + \int_0^t n_{\partial}^{(1)}(s) ds = \int_0^{+\infty} n^{(1)}(0, x) dx = 1.$$

The above yields that $\int_0^{+\infty} n_{\partial}^{(1)}(s) ds = 1$ by letting t tend to infinity, and that $n_{\partial}^{(1)}(t) \leq \tilde{b}$ for all $t \geq 0$ by the second line of (2.1). The fact that $n_{\partial}^{(1)}$ is bounded by \tilde{b} comes from the fact that the senescence rate cannot be higher than the division rate.

Remark 2.2. Let us denote the following sets, endowed respectively with the L^1 -norm, and the sum of the L^1 -norm and the L^∞ -norm,

$$\begin{aligned} \mathcal{F} &= \{f \in L^1(\mathbb{R}_+) \mid f \geq 0, \|f\|_{L^1(\mathbb{R}_+)} = 1\}, \\ \mathcal{F}_{\partial} &= \{f_{\partial} \in L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+) \mid f_{\partial} \geq 0, \|f_{\partial}\|_{L^1(\mathbb{R}_+)} = 1, \|f_{\partial}\|_{L^\infty(\mathbb{R}_+)} \leq \tilde{b}\}. \end{aligned}$$

We also consider $\Psi_{\partial} : \mathcal{F} \rightarrow \mathcal{F}_{\partial}$ the operator such that for all $f \in L^1(\mathbb{R}_+)$, $\Psi_{\partial}(f)$ corresponds to the cemetery $n_{\partial}^{(1)}$ of (2.1) when $n_0 = f$. This operator is well-posed by Proposition A.1 and Remark 2.1. In view of Proposition A.5 with $\xi = \tilde{b}g(y) dy$, if $(\Psi_{\partial})^{-1}$ exists, then it is not continuous (even if its domain is restricted on $\Psi_{\partial}(\mathcal{F})$). As a consequence, the inverse problem of estimating n_0 from $n_{\partial}^{(1)}$ is ill-posed in the sense of Hadamard and Tikhonov, see Definitions 2.1.1 and 2.1.3 in [20]. This difficulty is what motivates us to obtain a model approximation.

Model with $2k$ telomeres. We now create a model which takes into account the fact that a cell has several telomeres. We denote by $k \in \mathbb{N}^*$ the number of chromosomes of the species we study. Each chromosome

has 2 ends, so each cell has $2k$ telomeres. Hence, we use the space \mathbb{R}_+^{2k} to represent telomere lengths in our second model. For all $i \in \llbracket 1, k \rrbracket$, the i th and the $(i+k)$ th coordinates of a vector in \mathbb{R}_+^{2k} represent telomeres on the same chromosome (the chromosome i). At each cell division, for all $i \in \llbracket 1, k \rrbracket$, we have the following biological constraint:

- Either the telomere linked to the coordinate i is shortened and the telomere linked to the coordinate $i+k$ is unchanged, with probability $1/2$,
- Or the telomere linked to the coordinate $i+k$ is shortened and the telomere linked to the coordinate i is unchanged, with probability $1/2$.

To take into account this, we introduce the following set

$$\mathcal{I}_k = \left\{ I \in \mathcal{P}(\llbracket 1, 2k \rrbracket) \mid \#I = k, \forall (i, j) \in I^2 : i \neq j \implies i \bmod k \neq j \bmod k \right\}, \tag{2.2}$$

where $\mathcal{P}(\llbracket 1, 2k \rrbracket)$ is the power set of $\llbracket 1, 2k \rrbracket$, *i.e.* the set that contains all subsets of $\llbracket 1, 2k \rrbracket$. The set \mathcal{I}_k contains all the possible combinations of telomeres that can be shortened at each division. In particular, the condition in its definition represents the fact that only one end of the chromosome is shortened at each division. When a cell divides, we draw one set $I \in \mathcal{I}_k$ uniformly to know the indices where there is a shortening for the daughter cell we follow.

Example 2.3. When $k = 2$, we have

$$\mathcal{I}_2 = \{\{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}\}.$$

If at a cell division, we draw $\{1, 4\} \in \mathcal{I}_2$ (probability $1/4$) as a set where indices are shortened, then there is a shortening in the coordinates 1 and 4, and the coordinates 2 and 3 stay unchanged.

We also introduce the following measure for all $I \in \mathcal{I}_k$

$$d\tilde{\mu}^{(I)}(v) := \left[\prod_{i \in I} \tilde{g}(v_i) dv_i \right] \left[\prod_{i \in \llbracket 1, 2k \rrbracket \setminus I} \delta_0(dv_i) \right]. \tag{2.3}$$

This measure represents the shortening distribution at each division, knowing the fact that the shortened telomeres are those in I . Dirac measures represent the fact that the coordinates in $\llbracket 1, 2k \rrbracket \setminus I$ are unchanged, and the measures $(\tilde{g}(v_i) dv_i)_{i \in I}$ that the coordinates in I are shortened.

Finally, we consider

$$d\tilde{\mu}(v) := \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} d\tilde{\mu}^{(I)}(v). \tag{2.4}$$

This measure represents how telomeres are shortened at each division, taking into account all the possible combinations of telomeres that can be shortened at a division. The term 2^k comes from the fact that $\#\mathcal{I}_k = 2^k$, see Lemma A.6.

We denote by $n^{(2k)}(t, x)$ the density of telomere with lengths $x \in \mathbb{R}_+^{2k}$ at time $t \geq 0$. We also denote by $n_{\partial}^{(2k)}(t)$ the density of senescence times at time $t \geq 0$. We assume that at each division, the telomeres of the dividing cell are shortened by a random value distributed according to $\tilde{\mu}$, and that it becomes senescent if one of its telomere has a length below 0. Then, the probability that a cell with length $y \in \mathbb{R}_+^{2k}$ becomes senescent after division is $\tilde{\mu}(\{v \in \mathbb{R}_+^{2k} \mid y - v \notin \mathbb{R}_+^{2k}\})$. We also assume that the initial lengths distribution is $n^{(2k)}(0, x) = \prod_{i=1}^{2k} n_0(x_i)$, and refer to Remark 2.4 for more details about this choice. From the above description, we have that our second model is the following system

$$\begin{cases} \partial_t n^{(2k)}(t, x) = \tilde{b} \int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + v) d\tilde{\mu}(v) - \tilde{b} \cdot n^{(2k)}(t, x), & \forall t \geq 0, x \in \mathbb{R}_+^{2k}, \\ n_{\partial}^{(2k)}(t) = \tilde{b} \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \tilde{\mu}(\{v \in \mathbb{R}_+^{2k} \mid y - v \notin \mathbb{R}_+^{2k}\}) dy, & \forall t \geq 0, \\ n^{(2k)}(0, x) = \prod_{i=1}^{2k} n_0(x_i), & \forall x \in \mathbb{R}_+^{2k}. \end{cases} \tag{2.5}$$

Again, our aim is to estimate n_0 when $n_{\partial}^{(2k)}$ is observed.

Remark 2.4. Our assumption that $n^{(2k)}(0, x) = \prod_{i=1}^{2k} n_0(x_i)$ for all $x \in \mathbb{R}_+^{2k}$ is the same as in [24, 33, 40], and allows us to simplify computations. However, there are currently no theoretical or simulation-based results to support that this assumption is biologically relevant. Further studies will be conducted in this regard.

Remark 2.5. By adapting the computations done in Remark 2.1, we also have conservation of the number of individuals for this model, and that $n_{\partial}^{(2k)}$ is bounded by \tilde{b} . In particular, the following equalities hold, for all $t \geq 0$,

$$\int_{x \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x) dx + \int_0^t n_{\partial}^{(2k)}(s) ds = 1, \quad \text{and} \quad \int_{x \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x) dx = \int_t^{+\infty} n_{\partial}^{(2k)}(s) ds. \quad (2.6)$$

Remark 2.6. For the same reasons as those presented in Remark 2.2, retrieving n_0 from $n_{\partial}^{(2k)}$ is an ill-posed inverse problem in the sense of Hadamard and Tikhonov, see Definitions 2.1.1 and 2.1.3 in [20].

2.2. Assumptions and main result

The main result of this work is the construction of estimators for the initial distribution of telomere lengths in both models, with their respective error of estimation. We begin by presenting the main hypothesis of this work and some notations.

Assumptions. Our main result assumptions are the following. They are verified for example when n_0 is the density of an Erlang distribution. We study this case in detail in Section 5.

(H₁): There exist $N > 0$ (large), $\delta > 0$, $g : [0, \delta] \rightarrow \mathbb{R}_+$ a probability density function and $b > 0$ such that

$$\tilde{\delta} = \frac{\delta}{N}, \quad \forall x \in [0, \tilde{\delta}] : \tilde{g}(x) = Ng(Nx), \quad \text{and} \quad \tilde{b} = bN.$$

(H₂): It holds $n_0 \in W^{2,1}(\mathbb{R}_+)$, and there exist $\lambda > 0$, $C_{\lambda} > 0$ and $C'_{\lambda} > 0$ such that for all $x \in \mathbb{R}_+$

$$|n''_0(x)| \leq C_{\lambda} \exp(-\lambda x), \quad \text{and} \quad |n'_0(x)| \leq C'_{\lambda} \exp(-\lambda x).$$

(H₃): There exists $D_{\lambda} \geq 1$ such that for all $x \in \mathbb{R}_+$

$$n_0(x) \leq D_{\lambda} \lambda \exp(-\lambda x).$$

(H₄): There exist $\omega \geq \lambda$, $f_{\omega} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ non-decreasing, and $D_{\omega} \in (0, 1]$ such that for all $x \geq 0$

$$n_0(x) \geq \frac{D_{\omega} f_{\omega}(x) \exp(-\omega x)}{\int_0^{+\infty} f_{\omega}(y) \exp(-\omega y) dy}.$$

The above assumptions are further discussed in Section 2.3. Let us only emphasise one crucial point: in (H₁), N is assumed to be large. Then, when a result is stated, one needs to imagine that $N \rightarrow +\infty$ to understand it.

Notations. First, we denote the equivalents of $(\tilde{m}_i)_{i \in \{1,2\}}$ and \tilde{G} for g :

$$\forall i \in \{1, 2\} : m_i := \int_0^{\delta} u^i g(u) du, \quad \text{and} \quad \forall x \in \mathbb{R}_+ : G(x) := \int_0^{\min(x, \delta)} g(s) ds.$$

Then, we introduce the equivalents of $(\tilde{\mu}^{(I)})_{I \in \mathcal{I}_k}$ and $\tilde{\mu}$ for g :

$$\forall I \in \mathcal{I}_k : d\mu^{(I)}(v) := \left[\prod_{i \in I} g(v_i) dv_i \right] \left[\prod_{i \in [1, 2k] \setminus I} \delta_0(dv_i) \right], \quad \text{and} \quad d\mu(v) := \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} d\mu^{(I)}(v). \quad (2.7)$$

Thereafter, for all $d \in \mathbb{N}^*$ and ξ finite measure on \mathbb{R}_+^d , we consider $\mathcal{L}(\xi)$ the Laplace transform of ξ , defined for all $p \in \mathbb{C}$ such that $\text{Re}(p) > 0$ as:

$$\mathcal{L}(\xi)(p) := \int_{u \in \mathbb{R}_+^d} \exp\left(-p \sum_{j=1}^d u_j\right) \xi(du). \tag{2.8}$$

Finally, we define the following three constants:

$$\begin{aligned} \lambda_N &:= \frac{N}{m_1} \left[1 - \mathcal{L}(g)\left(\frac{\lambda}{N}\right) \right], & \lambda'_N &:= \frac{N}{km_1} \left[1 - \left(\mathcal{L}(g)\left(\frac{\lambda}{N}\right) \right)^k \right], \\ \omega'_N &:= \frac{N}{km_1} \left[1 - \left(\mathcal{L}(g)\left(\frac{\omega}{N}\right) \right)^k \right]. \end{aligned} \tag{2.9}$$

In view of the fact that $\mathcal{L}(g)(0) = \int_0^\delta g(u) du = 1$ and the definition of the derivative, we have that $\lim_{N \rightarrow +\infty} \lambda_N = -\frac{\lambda \mathcal{L}(g)'(0)}{m_1} = \lambda$. By the formula of the derivative of the power of a function, we also have that $\lim_{N \rightarrow +\infty} \lambda'_N = \lambda$ and $\lim_{N \rightarrow +\infty} \omega'_N = \omega$. Then, as N is supposed large, λ_N and λ'_N correspond to approximations of λ , and ω'_N to an approximation of ω .

Estimators and main result. To estimate n_0 from the observations of $n_\partial^{(1)}$ or $n_\partial^{(2k)}$, we use the estimators $\hat{n}_0^{(1)}$ and $\hat{n}_0^{(2k)}$ respectively, defined for all $x \geq 0$ as

$$\hat{n}_0^{(1)}(x) := \frac{1}{\tilde{b}\tilde{m}_1} n_\partial^{(1)}\left(\frac{x}{\tilde{b}\tilde{m}_1}\right), \quad \hat{n}_0^{(2k)}(x) := \frac{1}{k\tilde{b}\tilde{m}_1} \frac{n_\partial^{(2k)}\left(\frac{2x}{\tilde{b}\tilde{m}_1}\right)}{\left(\int_{\frac{2x}{\tilde{b}\tilde{m}_1}}^\infty n_\partial^{(2k)}(s) ds\right)^{1-\frac{1}{2k}}}. \tag{2.10}$$

The estimator on the left comes from the approximation of the pair $(n^{(1)}, n_\partial^{(1)})$ by a transport equation with drift $-bm_1$ and an absorbing state at $x = 0$, see Section 3. Thanks to this approximation, we have by using the method of characteristics that $n_\partial^{(1)}(t)$ is close to $bm_1 n_0(bm_1 t)$ for all $t \geq 0$ (see (3.6)). Therefore, by first taking $t = \frac{x}{bm_1}$, then dividing both terms by bm_1 , and finally using the following equality that comes from the change of variable $u' = \frac{u}{N}$ and (H₁):

$$bm_1 = b \int_0^\delta u g(u) du = bN^2 \int_0^{\frac{\delta}{N}} u' g(Nu') du' = \tilde{b}\tilde{m}_1, \tag{2.11}$$

we obtain that $\frac{1}{\tilde{b}\tilde{m}_1} n_\partial^{(1)}\left(\frac{x}{\tilde{b}\tilde{m}_1}\right)$ is close to $n_0(x)$ for all $x \geq 0$.

The estimator on the right of (2.10) also comes from an approximation of $(n^{(2k)}, n_\partial^{(2k)})$, see Section 4. This time, the approximated model is a transport equation with drift $-\frac{bm_1}{2}$ in each coordinate, and with an absorbing state at the boundary of \mathbb{R}_+^{2k} . Here are the steps to obtain the estimator. First, by (2.6), our model approximation, and the method of characteristics, we have that $\int_t^{+\infty} n_\partial^{(2k)}(s) ds$ is close to $\left[\int_{\frac{bm_1 t}{2}}^{+\infty} n_0(s) ds\right]^{2k}$ for all $t \geq 0$ (see Rem. 4.3). Then, by taking both terms to the power $\frac{1}{2k}$ and differentiating, we have that $\frac{1}{2k} n_\partial^{(2k)}(t) \left[\int_t^{+\infty} n_\partial^{(2k)}(s) ds\right]^{\frac{1}{2k}-1}$ is approximatively equal to $\frac{bm_1}{2} n_0\left(\frac{bm_1 t}{2}\right)$ for all $t \geq 0$. Finally, by taking $t = \frac{2x}{bm_1}$, then multiplying by $\frac{2}{bm_1}$, we have that $\frac{1}{k\tilde{b}\tilde{m}_1} n_\partial^{(2k)}\left(\frac{2x}{\tilde{b}\tilde{m}_1}\right) \left[\int_{\frac{2x}{\tilde{b}\tilde{m}_1}}^{+\infty} n_\partial^{(2k)}(s) ds\right]^{\frac{1}{2k}-1}$ is close to $n_0(x)$ for all $x \geq 0$. This last result justifies the expression of the estimator, because bm_1 and $\tilde{b}\tilde{m}_1$ are equals, see (2.11). We detail more rigorously these steps in Sections 3 and 4. This leads to the following result, which provides upper bounds on the errors between the estimators defined in (2.10) and n_0 .

Theorem 2.7 (Main result). *We recall the constants λ_N, λ'_N and ω'_N defined in (2.9). We also consider the constant $\beta'_N := (\lambda + 2k\lambda'_N) - (\omega + (2k - 1)\omega'_N)$. The following statements hold.*

(a) *Assume (H_1) – (H_2) . Then, there exists $c_1 > 0$ such that for all $x \geq 0$*

$$\left| \widehat{n}_0^{(1)}(x) - n_0(x) \right| \leq \frac{c_1}{N}(x + 1) \exp(-\lambda_N x),$$

where c_1 depends only on g, C_λ, C'_λ and D_λ .

(b) *Assume (H_1) – (H_4) . Then, there exists $d_1 > 0$ such that for all $x \geq 0$*

$$\left| \widehat{n}_0^{(2k)}(x) - n_0(x) \right| \leq \frac{d_1}{N} \left(\frac{D_\lambda}{D_\omega} \right)^{2k} (k^2 x + k + 1) \exp[-\beta'_N x],$$

where d_1 depends only on $g, \delta, \lambda, C_\lambda, C'_\lambda, D_\lambda$, and D_ω .

Remark 2.8. From the above, we have that the pointwise errors tend to zero when $N \rightarrow +\infty$.

Let us comment on the dependence on x of the bounds in Theorem 2.7. First, assume that $\lambda + 2k\lambda'_N > \omega + (2k - 1)\omega'_N$. Then, for each of the estimators, the bound contains a term corresponding to a linear growth in x , and a term corresponding to an exponential decay in x . The linear growth is related to the accumulation of errors during the model approximation. The exponential decay is related to the fact that the influence of past errors dissipates over time at an exponential rate. In the case where $\lambda + 2k\lambda'_N \leq \omega + (2k - 1)\omega'_N$, there is no exponential decay for $\widehat{n}_0^{(2k)}$. This is because the term in the denominator of $\widehat{n}_0^{(2k)}$, which tends to 0 exponentially fast, compensates for the dissipation and leads to an exponential growth of the error. The fact that there is a condition for having an exponential decay is consistent with the fact that the hazard rate function associated to a given density, which is similar to $\widehat{n}_0^{(2k)}$, may explode when $x \rightarrow +\infty$, see Sections 7.VI and 7.IX in [21]. However, the condition $\lambda + 2k\lambda'_N > \omega + (2k - 1)\omega'_N$ is not optimal, as discussed in the penultimate paragraph of Section 6. This opens new perspectives, notably to improve our bounds.

Let us now comment on the dependence on k of the bound obtained in Theorem 2.7 (b). We observe that two terms contribute to the k -dependence of the error. The first one is the term $\left(\frac{D_\lambda}{D_\omega}\right)^{2k}$. It implies that the error bound grows exponentially when the number of chromosomes increase. In fact, this growth is mainly related to the fact that our assumptions are not optimal, which leads to a non-optimal bound. This non-optimality is due to the fact that comparing n_0 with the densities presented in (H_3) and (H_4) results in a loss of information about n_0 , and that this loss is multiplied when the dimension increases. We would like to manage this loss of information in a future work in order to obtain a bound without the term $\left(\frac{D_\lambda}{D_\omega}\right)^{2k}$. The second term that contributes to the k -dependence of the bound in Theorem 2.7 (b) is the term $k^2 x + k + 1$. It is related to the fact that the size of the space \mathbb{R}_+^{2k} becomes larger when k increases, making the model more difficult to approximate. In fact, the dependence on k of the error causes problems when one is interested in species for which $2k$ is large, such as yeast cells ($2k = 32$) or human cells ($2k = 92$). This issue is further studied in Section 5.1.3.

We now need to ensure that the accumulation of the pointwise errors does not lead to an explosion of the global error. To verify this, we obtain from Theorem 2.7 and the fact that $\int_0^{+\infty} x^p e^{-\beta x} dx = \frac{\Gamma(p+1)}{\beta^{p+1}}$ for all $\beta \geq 0$ and $p > 0$ the following corollary. It provides bounds on the errors of our estimators in all Lebesgue spaces.

Corollary 2.9 (Estimation errors in Lebesgue spaces). *We recall the constants λ_N, λ'_N and ω'_N defined in (2.9). We also consider the constant $\beta'_N := (\lambda + 2k\lambda'_N) - (\omega + (2k - 1)\omega'_N)$. The following statements hold.*

(a) *Assume (H_1) – (H_2) . Then, for all $p > 0$ we have*

$$\left\| \widehat{n}_0^{(1)} - n_0 \right\|_{L^p(\mathbb{R}_+)} \leq \frac{c_1}{N} \left(\frac{\Gamma(p+1)^{\frac{1}{p}}}{(\lambda_N p)^{1+\frac{1}{p}}} + \frac{1}{(\lambda_N p)^{\frac{1}{p}}} \right),$$

where c_1 is the same constant as in Theorem 2.7 (a).

(b) Assume (H_1) – (H_4) , and that $\beta'_N > 0$. Then, for all $p > 0$, we have

$$\left\| \widehat{n}_0^{(2k)} - n_0 \right\|_{L^p(\mathbb{R}_+)} \leq \frac{d_1}{N} \left(\frac{D_\lambda}{D_\omega} \right)^{2k} \left(\frac{k^2 \Gamma(p+1)^{\frac{1}{p}}}{(\beta'_N p)^{1+\frac{1}{p}}} + \frac{k+1}{(\beta'_N p)^{\frac{1}{p}}} \right),$$

where d_1 is the same constant as in Theorem 2.7 (a).

From the above, we have that if $\lambda + 2k\lambda'_N \leq \omega + (2k - 1)\omega'_N$, then the accumulation of the pointwise errors for $\widehat{n}_0^{(2k)}$ is too large to get a result in norm, due to the exponential growth of the pointwise errors (linear when $\lambda + 2k\lambda'_N = \omega + (2k - 1)\omega'_N$).

2.3. Discussion about the single-telomere model and the assumptions

We conclude this section by explaining why we work with a model with one telomere, and discussing our assumptions.

Relevance of using a model with one telomere. As said in the presentation of the first model, up to our knowledge, there is no species with only one telomere. However, there are at least two good reasons to work with such a model.

- It is possible with experimental methods to place ourselves in a setting very similar to the study of cells with one telomere, see p. 112 in [42].
- It has been deduced numerically in Figure 5.a of [8], which studies a discrete probabilistic model with the same dynamics as those presented in (2.5), that 60% of senescence times are signalled by the telomere that was the shortest at the beginning of the dynamics, so that this telomere quite often signals senescence. Therefore, even if the approximation is rough, we can assume that the shortest telomere at the beginning always signals senescence to gain first insight. Under this approximation, it is sufficient to use a single-telomere model.

Discussion about the assumptions. Let us present the consequences of our assumptions, how they are useful in our proofs, and why we made them.

- (H_1) : This is the key assumption of this paper, as it allows us to justify that we can approximate our models by transport equations, see Sections 3.1 and 4.1. Since N is assumed to be large, this assumption means that the shortening values are small compared to the scale where telomere lengths are initially distributed. This assumption that the shortening values are small is supported by the following biological reality:
 - The average telomere length of the budding yeast is of the order of 300 base pairs [31], and the average shortening value is of the order of 7.5 base pairs [16]. The ratio between these is $\frac{7.5}{300} = \frac{1}{40} = 2.5\%$, which is small.
 - The average telomere length of the human is of the order of 12.5 kilobase pairs [31], and the average shortening value is of the order of 0.125 kilobase pairs [17]. The ratio between these is $\frac{0.125}{12.5} = \frac{1}{100} = 1\%$, which is small.

By the above explanation, we have that $N = 40$ when we study the budding yeast, and $N = 100$ when we study the human.

The assumption on \tilde{b} means that we work on a time scale where division times occur very frequently. This allows us to compensate the fact that the shortening values are small, and to avoid to have senescence times that tend to infinity when $N \rightarrow +\infty$. This assumption also implies that the constraints

$$\left\| n_\partial^{(1)} \right\|_{L^\infty(\mathbb{R}_+)} \leq \tilde{b} = bN \quad \text{and} \quad \left\| n_\partial^{(2k)} \right\|_{L^\infty(\mathbb{R}_+)} \leq \tilde{b} = bN \tag{2.12}$$

presented in Remarks 2.1 and 2.5 become by letting N go to infinity:

$$\|n_{\partial}^{(1)}\|_{L^{\infty}(\mathbb{R}_+)} < +\infty \quad \text{and} \quad \|n_{\partial}^{(2k)}\|_{L^{\infty}(\mathbb{R}_+)} < +\infty,$$

which is no more restrictive. Hence, in the rest of the article, when it is required to approximate or estimate n_{∂} , we do not try to satisfy (2.12).

- (H₂): Due to a second-order Taylor expansion, the errors between the approximated models and the original models are mainly given by the second derivative of n_0 , see Section 3.3. The inequality on the left-hand side of (H₂) allows us to have a control on it. The inequality on the right-hand side, for its part, allows us to have a control on the variation of telomere length density close to 0. Controlling this is important because cells susceptible to become senescent have telomere lengths close to 0. For more information, we refer to Sections 3.4, 3.5 and 4.5.
- (H₃): The approximation error for the model with several telomeres does not only depend on the derivatives of n_0 , but also on n_0 itself. This assumption allows us to control it.
- (H₄): This assumption allows us to obtain a lower bound for $\int_t^{+\infty} n_{\partial}^{(2k)}(s) ds$, for all $t \geq 0$. It is important to have such lower bound because the inverse of $\int_t^{+\infty} n_{\partial}^{(2k)}(s) ds$ appears in the expression of $\hat{n}_0^{(2k)}$, see (2.10), and tends to 0 when $t \rightarrow +\infty$. Therefore, if the decay of the function $t \mapsto \int_t^{+\infty} n_{\partial}^{(2k)}(s) ds$ is too fast, then $\hat{n}_0^{(2k)}(x)$ explodes when $x \rightarrow +\infty$.

3. THE SINGLE-TELOMERE MODEL

We begin by bounding the error done by $\hat{n}_0^{(1)}$, *i.e.*, we prove Theorem 2.7 (a). This statement follows almost directly from the model approximation. First, in Section 3.1, we successively rewrite (2.1) using the parameters introduced in (H₁), explain how an approximation can be obtained from this rewriting, and prove Theorem 2.7 (a) assuming the approximation is true. Then, in Section 3.2, we present the auxiliary results necessary to obtain this approximation. Thereafter, in Sections 3.3 and 3.4, we prove these auxiliary statements. Finally, in Section 3.5, we prove the model approximation. Throughout this section, we assume that (H₁) holds.

3.1. Model approximation and proof of Theorem 2.7 (a)

The approximation of (2.1) is obtained by letting the scaling parameter N , introduced in (H₁), tend to infinity. Thus, we need to rewrite (2.1) to make N appears. Let us start with the first line of (2.1). In this equation, we replace \tilde{b} with bN , then $\tilde{g}(v)$ with $Ng(Nv)$ for all $v \geq 0$, and finally $\tilde{\delta}$ with $\frac{\delta}{N}$. Thereafter, we do the change of variable $v' = Nv$, and place $n^{(1)}(t, x)$ inside the integral by using that $\int_0^{\delta} g(v) dv = 1$. We obtain that for all $t \geq 0, x \geq 0$,

$$\begin{aligned} \partial_t n^{(1)}(t, x) &= bN^2 \int_0^{\frac{\delta}{N}} n^{(1)}(t, x + v)g(Nv) dv - bN \cdot n^{(1)}(t, x) \\ &= bN \int_0^{\delta} \left[n^{(1)}\left(t, x + \frac{v'}{N}\right) - n^{(1)}(t, x) \right] g(v') dv'. \end{aligned} \tag{3.1}$$

Now, we rewrite the second line of (2.1). First observe that by the change of variable $w' = \frac{w}{N}$, we have for all $v \in [0, \frac{\delta}{N}]$

$$G(Nv) = \int_0^{Nv} g(w) dw = N \int_0^v g(Nw') dw' = \int_0^v \tilde{g}(w') dw' = \tilde{G}(v). \tag{3.2}$$

In view of the above equality, in the second line of (2.1), we successively replace \tilde{b} with bN , $\tilde{G}(v)$ with $G(Nv)$ for all $v \geq 0$, and finally $\tilde{\delta}$ with $\frac{\delta}{N}$. Then, we do the change of variable $v' = Nv$. It comes the following rewriting,

for all $t \geq 0$,

$$n_{\partial}^{(1)}(t) = bN \int_0^{\frac{\delta}{N}} n^{(1)}(t, v)(1 - G(Nv)) \, dv = b \int_0^{\delta} n^{(1)}\left(t, \frac{v'}{N}\right)(1 - G(v')) \, dv'. \tag{3.3}$$

By combining (3.1) and (3.3), we now have below a new expression for (2.1) in which N appears

$$\begin{cases} \partial_t n^{(1)}(t, x) = bN \int_0^{\delta} [n^{(1)}(t, x + \frac{v}{N}) - n^{(1)}(t, x)]g(v) \, dv, & \forall t \geq 0, x \geq 0, \\ n_{\partial}^{(1)}(t) = b \int_0^{\delta} n^{(1)}(t, \frac{v}{N})(1 - G(v)) \, dv, & \forall t \geq 0, \\ n^{(1)}(0, x) = n_0(x), & \forall x \geq 0. \end{cases} \tag{3.4}$$

We aim to derive a system corresponding to the limit version of (3.4) as $N \rightarrow +\infty$. In fact, this system can be intuitively obtained. To do so, one has to observe that by the definition of the derivative and the equality $\int_0^{\delta} (1 - G(v)) \, dv = \int_0^{\delta} vg(v) \, dv = m_1$ (integration by part), we have the following two results, for all $t \geq 0$,

$$\begin{aligned} N \int_0^{\delta} [n^{(1)}(t, x + \frac{v}{N}) - n^{(1)}(t, x)]g(v) \, dv &\underset{N \rightarrow +\infty}{\approx} \int_0^{\delta} [v\partial_x n^{(1)}(t, x)]g(v) \, dv = m_1 \partial_x n^{(1)}(t, x), \\ \int_0^{\delta} n^{(1)}(t, \frac{v}{N})(1 - G(v)) \, dv &\underset{N \rightarrow +\infty}{\approx} m_1 n^{(1)}(t, 0). \end{aligned}$$

Then, by plugging the above in (3.4), we can conjecture that the following approximates (3.4)

$$\begin{cases} \partial_t u^{(1)}(t, x) = bm_1 \partial_x u^{(1)}(t, x), & \forall t \geq 0, x \geq 0, \\ u_{\partial}^{(1)}(t) = bm_1 u^{(1)}(t, 0), & \forall t \geq 0, \\ u^{(1)}(0, x) = n_0(x), & \forall x \geq 0. \end{cases} \tag{3.5}$$

Remark 3.1. By using the method of characteristics to solve the first line of (3.5), and then plugging the solution in the second line, we have for all $t \geq 0, x \geq 0$,

$$u^{(1)}(t, x) = n_0(bm_1 t + x), \quad \text{and} \quad u_{\partial}^{(1)}(t) = bm_1 n_0(bm_1 t). \tag{3.6}$$

In fact, our conjecture can be rigorously proven. Specifically, the following result, proved in Section 3.5, provides bounds on the pointwise errors between (3.4) and (3.5).

Proposition 3.2 (Pointwise approximation errors, one telomere). *We recall the constant λ_N defined in (2.9). The following statements hold.*

(a) Assume (H_1) – (H_2) . Then, there exists $c'_0 > 0$ such that for all $t \geq 0, x \geq 0$, we have

$$\left| n^{(1)}(t, x) - u^{(1)}(t, x) \right| \leq \frac{c'_0 bt}{N} \exp(-bm_1 \lambda_N t) \exp(-\lambda x),$$

where c'_0 depends only on g and C_{λ} .

(b) Assume (H_1) – (H_2) . Then, there exists $c'_1 > 0$ such that for all $t \geq 0$, we have

$$\left| n_{\partial}^{(1)}(t) - u_{\partial}^{(1)}(t) \right| \leq \frac{c'_1 b(bm_1 t + 1)}{N} \exp(-bm_1 \lambda_N t),$$

where c'_1 depends only on g, C_{λ} and C'_{λ} .

Remark 3.3. As done in Corollary 2.9, the errors in all Lebesgue spaces between $n^{(1)}$ and $u^{(1)}$, and $n_{\partial}^{(1)}$ and $u_{\partial}^{(1)}$ can be bounded thanks to the above inequalities. We do not state these bounds as they are not necessary to prove our main theorem.

The main interest of Proposition 3.2 is that it allows us to get a bound on the error done by $\widehat{n}_0^{(1)}$. To obtain it, first observe that by the right-hand side of (3.6), it holds $n_0(x) = \frac{1}{bm_1}u_{\partial}^{(1)}\left(\frac{x}{bm_1}\right)$ for all $x \geq 0$. Then, in view of the definition of $\widehat{n}_0^{(1)}$ given in (2.10) and equation (2.11), we have by applying Proposition 3.2 (b) the following corollary, which directly implies Theorem 2.7 (a).

Corollary 3.4 (Pointwise estimation error, one telomere). *We recall the constant λ_N defined in (2.9). Assume that (H_1) – (H_2) hold. Then, for all $x \geq 0$, we have*

$$\left| \widehat{n}_0^{(1)}(x) - n_0(x) \right| = \frac{1}{bm_1} \left| n_{\partial}^{(1)}\left(\frac{x}{bm_1}\right) - u_{\partial}^{(1)}\left(\frac{x}{bm_1}\right) \right| \leq \frac{c'_1(x+1)}{m_1N} \exp(-\lambda_N x),$$

where c'_1 is the same constant as in Proposition 3.2 (b).

From the above, if we manage to justify that Proposition 3.2 is true, then Theorem 2.7 (a) will be proved. Explaining why this proposition holds is thus what we do in the next subsections. In particular, we now present the main arguments and the auxiliary statements used to obtain it.

3.2. Plan of the proof of Proposition 3.2

The proof of the first statement of Proposition 3.2 consists in controlling the absolute value of $\bar{u}^{(1)} := n^{(1)} - u^{(1)}$. To do so, we first obtain an equation verified by $\bar{u}^{(1)}$. By taking the difference between the first lines of (3.4) and (3.5), then decomposing $n^{(1)}$ with the equality $n^{(1)} = \bar{u}^{(1)} + u^{(1)}$, and finally using that $bm_1 = bN \int_0^{\delta} \frac{v}{N}g(v) dv$, we have for all $t \geq 0, x \geq 0$,

$$\begin{aligned} \partial_t \bar{u}^{(1)}(t, x) &= bN \int_0^{\delta} \left[n^{(1)}\left(t, x + \frac{v}{N}\right) - n^{(1)}(t, x) \right] g(v) dv - bm_1 \partial_x u^{(1)}(t, x) \\ &= bN \int_0^{\delta} \left[\bar{u}^{(1)}\left(t, x + \frac{v}{N}\right) - \bar{u}^{(1)}(t, x) \right] g(v) dv \\ &\quad + bN \int_0^{\delta} \left[u^{(1)}\left(t, x + \frac{v}{N}\right) - u^{(1)}(t, x) - \frac{v}{N} \partial_x u^{(1)}(t, x) \right] g(v) dv. \end{aligned} \tag{3.7}$$

Therefore, if we are able to bound a solution of the above equation, then Proposition 3.2 (a) will be proved. The following lemma, proved in Section 3.3, allows us to do this. This lemma is stated for equations in a multidimensional trait space instead of only \mathbb{R}_+ , as we use it later in Section 4.4 for the approximation of the model with several telomeres.

Lemma 3.5 (Key lemma for approximating lengths densities). *Let $d \in \mathbb{N}^*$, ξ a probability measure on \mathbb{R}^d with finite first and second moments, and $F : \mathbb{R}_+ \times \mathbb{R}_+^d \rightarrow \mathbb{R}$. We also consider u_{ξ} the solution of the following integro-differential equation, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$,*

$$\begin{aligned} \partial_t u_{\xi}(t, x) &= bN \int_{v \in \mathbb{R}_+^d} \left[u_{\xi}\left(t, x + \frac{v}{N}\right) - u_{\xi}(t, x) \right] \xi(dv) \\ &\quad + bN \int_{v \in \mathbb{R}_+^d} \left[F\left(t, x + \frac{v}{N}\right) - F(t, x) - \sum_{i=1}^d \frac{v_i}{N} \partial_{x_i} F(t, x) \right] \xi(dv), \end{aligned} \tag{3.8}$$

with initial condition $u_\xi(0, \cdot) \equiv 0$. Assume that there exist $\alpha, \beta > 0$ and $C > 0$ such that

$$\forall (t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d : \sup_{(\ell, \ell') \in \llbracket 1, d \rrbracket^2} (|\partial_{x_\ell x_{\ell'}} F(t, x)|) \leq C \exp\left(-\alpha t - \beta \sum_{i=1}^d x_i\right), \tag{3.9}$$

$$bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N}\right)\right) \leq \alpha. \tag{3.10}$$

Then, denoting the constant $\sigma_\xi = \sum_{1 \leq \ell, \ell' \leq d} \int_{v \in \mathbb{R}_+^d} v_\ell v_{\ell'} \xi(dv)$, we have for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$

$$|u_\xi(t, x)| \leq \frac{btC\sigma_\xi}{2N} \exp\left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N}\right)\right)t\right] \exp\left(-\beta \sum_{i=1}^d x_i\right). \tag{3.11}$$

One can easily see that (3.7) is an equation of the form given in (3.8) with $d = 1$, $u_\xi = \bar{u}^{(1)}$, $\xi = g$ and $F = u^{(1)}$. The proof of the first statement of Proposition 3.2 is thus to check the assumptions of Lemma 3.5, and then to apply it. We do this in Section 3.5.

To prove Proposition 3.2 (b), we need this time to control $|\bar{u}_\partial^{(1)}| := |n_\partial^{(1)} - u_\partial^{(1)}|$. Again, we do this by obtaining the equation verified by $\bar{u}_\partial^{(1)}$, and then applying a general lemma. By taking the difference between the second lines of (3.4) and (3.5), then decomposing $n^{(1)}$ with the equality $n^{(1)} = \bar{u}^{(1)} + u^{(1)}$, and finally using that $\int_0^\delta (1 - G(v)) dv = \int_0^\delta vg(v) dv = m_1$ (integration by part), we have for all $t \geq 0$

$$\begin{aligned} \bar{u}_\partial^{(1)}(t) &= b \int_0^\delta n^{(1)}\left(t, \frac{v}{N}\right) (1 - G(v)) dv - bm_1 u^{(1)}(t, 0) \\ &= b \int_0^\delta \bar{u}^{(1)}\left(t, \frac{v}{N}\right) (1 - G(v)) dv + b \int_0^\delta \left[u^{(1)}\left(t, \frac{v}{N}\right) - u^{(1)}(t, 0)\right] (1 - G(v)) dv. \end{aligned} \tag{3.12}$$

We thus need to control an equation with the same form as (3.12), and the second statement of Proposition 3.2 will be proved. This control is done by using the following lemma, that is proved in Section 3.4.

Lemma 3.6 (Key lemma for approximating cemeteries). *Assume that the assumptions of Lemma 3.5 hold. We consider a set of functions $(h_i)_{i \in \llbracket 1, d \rrbracket}$ from $\mathbb{R}_+ \times \mathbb{R}^d$ to \mathbb{R} and $C' > 0$ such that for all $t \geq 0$, $x \in \mathbb{R}_+^d$ and $i \in \llbracket 1, d \rrbracket$, it holds*

$$|h_i(t, x)| \leq \frac{C' x_i}{N} \exp\left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N}\right)\right)t - \beta \sum_{j=1, j \neq i}^d x_j\right]. \tag{3.13}$$

We denote the function $v_\xi : \mathbb{R}_+ \rightarrow \mathbb{R}$, defined for all $t \geq 0$ as

$$\begin{aligned} v_\xi(t) &:= bN \int_{y \in \mathbb{R}_+^d} u_\xi(t, y) \xi(\{v \in \mathbb{R}_+^d \mid Ny - v \notin \mathbb{R}_+^d\}) dy \\ &\quad + b \sum_{i=1}^d \int_{y \in \mathbb{R}_+^d} h_i(t, y) (1 - G(y_i)) 1_{\{y_i \leq \delta\}} dy. \end{aligned} \tag{3.14}$$

Then, recalling the constant σ_ξ introduced in Lemma 3.5, we have for all $t \geq 0$

$$|v_\xi(t)| \leq \left(\frac{btC\sigma_\xi\alpha}{2\beta^d N} + \frac{bC'm_2d}{2\beta^{d-1}N}\right) \exp\left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N}\right)\right)t\right]. \tag{3.15}$$

This is less evident to see it, but equation (3.12) and a change of variable imply that $\bar{u}_\partial^{(1)}$ is of the form presented in (3.14) with $\xi = g$. More details about this are given in Section 3.5, as well as the proof of Proposition 3.2 (b) from Lemma 3.6.

We now prove the auxiliary statements given in this section, and then obtain Proposition 3.2 from them.

3.3. Proof of Lemma 3.5

We begin with preliminaries. To simplify notations, we denote $H_N : \mathbb{R}_+ \times \mathbb{R}_+^d \rightarrow \mathbb{R}$, the function defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$ as

$$\begin{aligned} H_N(t, x) &:= bN \int_{v \in \mathbb{R}_+^d} \left[F\left(t, x + \frac{v}{N}\right) - F(t, x) - \sum_{i=1}^d \frac{v_i}{N} \partial_{x_i} F(t, x) \right] \xi(dv) \\ &= \frac{b}{N} \sum_{1 \leq \ell, \ell' \leq d} \int_{v \in \mathbb{R}_+^d} \left[\int_0^1 (1-w) \left(\partial_{x_\ell x_{\ell'}} F\left(t, x + w \frac{v}{N}\right) v_\ell v_{\ell'} \right) dw \right] \xi(dv). \end{aligned} \tag{3.16}$$

The last equality in the above comes from expanding $F(t, x + \frac{v}{N}) - F(t, x) - \sum_{i=1}^d \frac{v_i}{N} \partial_{x_i} F(t, x)$ using a Taylor expansion with remainder in integral form. We also introduce the linear operator $\Phi : L^1(\mathbb{R}^d) \rightarrow C(\mathbb{R}_+, L^1(\mathbb{R}^d))$, defined such that for all $f_0 \in L^1(\mathbb{R}^d)$, $m = \Phi(f_0)$ is the solution in $C(\mathbb{R}_+, L^1(\mathbb{R}^d))$ of the following integro-differential equation

$$\begin{cases} \partial_t m(t, x) = bN \int_{v \in \mathbb{R}_+^d} m(t, x + \frac{v}{N}) \xi(dv) - bNm(t, x), & \forall t \geq 0, x \in \mathbb{R}_+^d, \\ m(0, x) = f_0(x), & \forall x \in \mathbb{R}_+^d. \end{cases} \tag{3.17}$$

This operator is well-posed by Proposition A.1.

The proof of Lemma 3.5 is done in two steps. First, in Step 1, we prove that for all $(t, s, x) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+^d$ such that $t \geq s$, it holds

$$|\Phi(H_N(s, \cdot))(t - s, x)| \leq \frac{bC\sigma_\xi}{2N} \exp\left[-bN\left(1 - \mathcal{L}(\xi)\left(\frac{\beta}{N}\right)\right)t\right] \exp\left(-\beta \sum_{i=1}^d x_i\right). \tag{3.18}$$

Then, in Step 2, we prove that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$, we have

$$u_\xi(t, x) = \int_0^t \Phi(H_N(s, \cdot))(t - s, x) ds. \tag{3.19}$$

In view of the fact that for all $s \geq 0$, $H_N(s, \cdot)$ is the source term in (3.8) at time s , equation (3.19) means that we have rewritten u_ξ as the sum of the evolutions of the source terms over time. This equation (3.18) imply that the lemma is proved, as we only have to plug (3.18) in (3.19) and then integrate to obtain (3.11).

Step 1. Assume first that for all $(s, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$ it holds

$$|H_N(s, x)| \leq \frac{bC\sigma_\xi}{2N} \exp(-\alpha s) f_\beta(x), \tag{3.20}$$

where $f_\beta(x) := \exp(-\beta \sum_{i=1}^d x_i)$. By Corollary A.2 and the definition of Φ , we have for all $(h, \varphi) \in (L^1(\mathbb{R}_+^d))^2$ verifying $|h| \leq \varphi$ that $|\Phi(h)| \leq \Phi(\varphi)$. To bound from above the left-hand side of (3.18), we first apply this inequality for $h = H_N(s, \cdot)$ and $\varphi = \frac{bC\sigma_\xi}{2N} \exp(-\alpha s) f_\beta$. Then, we use that it holds $\Phi(\varphi) = \frac{bC\sigma_\xi}{2N} \exp(-\alpha s) \Phi(f_\beta)$, as Φ is a linear operator. Finally, we apply equation (A.3) with $\omega = 0$ to compute $\Phi(f_\beta)$. We obtain that for all $(t, s, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$ such that $s \leq t$

$$|\Phi(H_N(s, \cdot))(t - s, x)| \leq \frac{bC\sigma_\xi}{2N} \exp(-\alpha s) \exp\left[-bN\left(1 - \mathcal{L}(\xi)\left(\frac{\beta}{N}\right)\right)(t - s)\right] \exp\left[-\beta \sum_{i=1}^d x_i\right].$$

Then, by using (3.10) to bound from above the first exponential, we get that (3.18) is true assuming that (3.20) holds.

It thus remains to prove (3.20), and Step 1 will be done. To do so, we first plug (3.9) in the second line of (3.16). Then, we bound from above $\exp(-\alpha s - \sum_{i=1}^d \beta(x_i + w \frac{v_i}{N}))$ by $\exp(-\alpha s) f_\beta(x)$, and use the equality $\int_0^1 (1-w) dw = \frac{1}{2}$. It comes the following, which proves (3.20) and concludes the proof of the first step, for all $(s, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$,

$$\begin{aligned} |H_N(s, x)| &\leq \frac{bC}{N} \sum_{1 \leq \ell, \ell' \leq d} \int_{v \in \mathbb{R}_+^d} \left[\int_0^1 (1-w) \exp\left(-\alpha s - \sum_{i=1}^d \beta\left(x_i + w \frac{v_i}{N}\right)\right) v_\ell v_{\ell'} dw \right] \xi(dv) \\ &\leq \frac{bC}{2N} \left[\sum_{1 \leq \ell, \ell' \leq d} \int_{v \in \mathbb{R}_+^d} v_\ell v_{\ell'} \xi(dv) \right] \exp(-\alpha s) f_\beta(x) = \frac{bC \sigma_\xi}{2N} \exp(-\alpha s) f_\beta(x). \end{aligned}$$

Step 2. To obtain (3.19), we prove that the function \tilde{u}_ξ , defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$ as

$$\tilde{u}_\xi(t, x) = \int_0^t \Phi(H_N(s, \cdot))(t-s, x) ds \quad (3.21)$$

is a solution of (3.8), the equation verified by u_ξ . Then, by the uniqueness result stated in Proposition A.1 and the fact that u_ξ and \tilde{u}_ξ have the same initial conditions, we will have that $u_\xi = \tilde{u}_\xi$, so that (3.19) is true. To do the above, we first compute $\partial_t \tilde{u}_\xi(t, x)$, by using that for all $f \in W^{1,1}(\mathbb{R}_+, L^1(\mathbb{R}_+))$ such that $s \mapsto f(s, s) \in L^1(\mathbb{R}_+)$, it holds $\frac{d}{dt} (\int_0^t f(t, s) ds) = \int_0^t \partial_t f(t, s) ds + f(t, t)$. Thereafter, we simplify the terms $\partial_t \Phi(H_N(s, \cdot))(t-s, x)$ and $\Phi(H_N(t, \cdot))(0, x)$ that appear after the previous computation, in view of the fact that $\Phi(H_N(s, \cdot))$ is a solution of (3.17) with initial condition $f_0 = H_N(s, \cdot)$. Finally, we switch the integrals, and use (3.21) to replace the integrals in ds with the function \tilde{u}_ξ . We obtain that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$

$$\begin{aligned} \partial_t \tilde{u}_\xi(t, x) &= \int_0^t \partial_t \Phi(H_N(s, \cdot))(t-s, x) ds + \Phi(H_N(t, \cdot))(0, x) \\ &= bN \int_0^t \left[\int_{v \in \mathbb{R}_+^d} \Phi(H_N(s, \cdot))\left(t-s, x + \frac{v}{N}\right) \xi(dv) - \Phi(H_N(s, \cdot))(t-s, x) \right] ds + H_N(t, x) \\ &= bN \left[\int_{v \in \mathbb{R}_+^d} \tilde{u}_\xi\left(t, x + \frac{v}{N}\right) \xi(dv) - \tilde{u}_\xi(t, x) \right] + H_N(t, x). \end{aligned}$$

The above equation is in fact the same as (3.8) in view of the definition of H_N , see (3.16). Then, we have that $u_\xi = \tilde{u}_\xi$, which concludes the proof of the second step, and thus of the lemma. \square

3.4. Proof of Lemma 3.6

To simplify notations, we denote for all $y \in \mathbb{R}_+^d$ the set $A_{Ny} := \{v \in \mathbb{R}_+^d \mid Ny - v \notin \mathbb{R}_+^d\}$. We consider $v_{\xi,1}$ and $v_{\xi,2}$ the functions defined for all $t \geq 0$ as

$$v_{\xi,1}(t) = bN \int_{y \in \mathbb{R}_+^d} u_\xi(t, y) \xi(A_{Ny}) dy, \quad \text{and} \quad v_{\xi,2}(t) = b \sum_{i=1}^d \int_{y \in \mathbb{R}_+^d} h_i(t, y) (1 - G(y_i)) 1_{\{y_i \leq \delta\}} dy.$$

By (3.14), we have that $v_\xi = v_{\xi,1} + v_{\xi,2}$. Then, to prove this lemma, we bound $v_{\xi,1}$ and $v_{\xi,2}$, and conclude by summing their bounds, in view of the triangle inequality.

Let $t \geq 0$. We begin by bounding $v_{\xi,2}(t)$. First, for all $i \in \llbracket 1, d \rrbracket$, we apply (3.13) to bound the function h_i . Then, we compute the integrals with respect to $(dy_j)_{j \in \llbracket 1, d \rrbracket, j \neq i}$, by using that

$$\prod_{j \in \llbracket 1, d \rrbracket, j \neq i} \left(\int_{y_j \in \mathbb{R}_+} \exp(-\beta y_j) dy_j \right) = \left(\int_{s \in \mathbb{R}_+} \exp(-\beta s) ds \right)^{d-1} = \frac{1}{\beta^{d-1}}.$$

Finally, we use the equality $\int_{y_i \in [0, \delta]} y_i(1 - G(y_i)) dy_i = \int_{y_i \in [0, \delta]} \frac{(y_i)^2}{2} g(y_i) dy_i = \frac{m_2}{2}$ (integration by part) to compute the integral that remains. It comes

$$\begin{aligned} |v_{\xi,2}(t)| &\leq b \sum_{i=1}^d \frac{C'}{N} \frac{1}{\beta^{d-1}} \frac{m_2}{2} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right] \\ &= \frac{bC'm_2d}{2\beta^{d-1}N} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right]. \end{aligned} \tag{3.22}$$

Now, we focus on bounding $v_{\xi,1}(t)$. To do so, we first apply (3.11) to bound from above the term $u_{\xi}(t, y)$ in $|v_{\xi,1}(t)|$. Thereafter, we use the fact that for all $y \in \mathbb{R}^d$ it holds $\xi(A_N y) = 1 - \xi((A_N y)^c)$ to develop the bound in two different integrals. Finally, we compute the value of the first integral, which is

$$\int_{y \in \mathbb{R}_+^d} \exp \left(-\beta \sum_{i=1}^d y_i \right) dy = \frac{1}{\beta^d}.$$

We get the following

$$\begin{aligned} |v_{\xi,1}(t)| &\leq \frac{b^2 t C \sigma_{\xi}}{2} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right] \int_{y \in \mathbb{R}_+^d} \exp \left(-\beta \sum_{i=1}^d y_i \right) \xi(A_N y) dy \\ &= \frac{b^2 t C \sigma_{\xi}}{2} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right] \left[\frac{1}{\beta^d} - \int_{y \in \mathbb{R}_+^d} \exp \left(-\beta \sum_{i=1}^d y_i \right) \xi((A_N y)^c) dy \right]. \end{aligned} \tag{3.23}$$

To continue our computations, we need to find a better expression for the last term of (3.23). We notice that by the definition of $A_N y$, it holds $\xi((A_N y)^c) = \int_{v \in \mathbb{R}_+^d} 1_{\{v_i \in [1, d], y_i \geq \frac{v_i}{N}\}} \xi(dv)$, for all $y \geq 0$. Using this equality, then Fubini's theorem to switch the integrals, and finally the fact that $\int_{y_i \in [\frac{v_i}{N}, +\infty)} \exp(-\beta y_i) dy_i = \frac{\exp(-\frac{\beta v_i}{N})}{\beta}$ for all $i \in [1, d]$, we have

$$\begin{aligned} \int_{y \in \mathbb{R}_+^d} \exp \left[-\beta \sum_{i=1}^d y_i \right] \xi((A_N y)^c) dy &= \int_{v \in \mathbb{R}_+^d} \left[\int_{y \in \mathbb{R}_+^d} 1_{\{v_i \in [1, d], y_i \geq \frac{v_i}{N}\}} \exp \left(-\beta \sum_{i=1}^d y_i \right) dy \right] \xi(dv) \\ &= \frac{1}{\beta^d} \mathcal{L}(\xi) \left(\frac{\beta}{N} \right). \end{aligned}$$

The above expression is what we need to continue the computations interrupted at (3.23). First, we plug the above equation in (3.23), and put the terms $\frac{1}{\beta^d}$ in factors. Then, we use (3.10) to bound the term $1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right)$ that appears from the previous step by $\frac{\alpha}{bN}$. We obtain

$$\begin{aligned} |v_{\xi,1}(t)| &\leq \frac{b^2 t C \sigma_{\xi}}{2\beta^d} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right] \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) \\ &\leq \frac{btC\sigma_{\xi}\alpha}{2\beta^d N} \exp \left[-bN \left(1 - \mathcal{L}(\xi) \left(\frac{\beta}{N} \right) \right) t \right]. \end{aligned} \tag{3.24}$$

The lemma is thus proved by summing (3.22) and (3.24). □

3.5. Proof of Proposition 3.2

We prove this proposition statement by statement. We first deal with Proposition 3.2 (a). We need to verify the assumptions of Lemma 3.5 for $u_{\xi} = \bar{u}^{(1)}$ to obtain it. First, recall that (3.7), the equation verified by $\bar{u}^{(1)}$,

corresponds to (3.8) with $F = u^{(1)}$ and $\xi = g$. Then, notice that by the left-hand side of (3.6) and (H_2) , we have for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$

$$\left| \partial_{xx} u^{(1)}(t, x) \right| = |n_0''(bm_1 t + x)| \leq C_\lambda \exp(-\lambda bm_1 t - \lambda x),$$

so that equation (3.9) holds with $\alpha = \lambda bm_1$, $\beta = \lambda$ and $C = C_\lambda$. Finally, by the inequality $1 - e^{-x} \leq x$ for all $x \in \mathbb{R}$, one can easily obtain that

$$bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right) = bN \int_0^{+\infty} \left(1 - e^{-\frac{\lambda}{N} u} \right) g(u) \, du \leq \lambda b \int_0^{+\infty} u g(u) \, du = \lambda bm_1. \tag{3.25}$$

The above is exactly equation (3.10) with the same α and β as before. Then, we have that all the assumptions of Lemma 3.5 are verified. We therefore apply this lemma to bound $u_\xi = \bar{u}^{(1)}$, in view of the fact that

$$\sigma_\xi = \int_0^\delta u^2 g(u) \, du = m_2.$$

We obtain that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$

$$\left| \bar{u}^{(1)}(t, x) \right| \leq \frac{btC_\lambda m_2}{2N} \exp \left[-bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right) t \right] \exp(-\lambda x).$$

As it holds $bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right) = bm_1 \lambda_N$ by equation (2.9), the above inequality yields that Proposition 3.2 (a) is true with $c'_0 = C_\lambda \frac{m_2}{2}$.

Now, we deal with Proposition 3.2 (b). This time, we check the assumptions of Lemma 3.6 for the function $v_\xi = \bar{u}_\partial^{(1)}$. We denote h_1 the function defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$ as

$$h_1(t, x) = u^{(1)} \left(t, \frac{x}{N} \right) - u^{(1)}(t, 0). \tag{3.26}$$

In the first term of the last line of (3.12), we do the change of variable $y = \frac{v}{N}$, and replace $\bar{u}^{(1)}$ with u_ξ (same u_ξ as before). In the second term, we plug equation (3.26). We obtain that for all $t \geq 0$

$$\bar{u}_\partial^{(1)}(t) = bN \int_0^{\frac{\delta}{N}} u_\xi(t, y) (1 - G(Ny)) \, dy + b \int_0^\delta h_1(t, v) (1 - G(v)) \, dv,$$

so that (3.14) holds. It remains to prove (3.13). To do so, in view of (3.26), we first write h_1 as an integral of $\partial_x u^{(1)}$. Then, we use (3.6) to write $u^{(1)}$ in terms of n_0 . Finally, we apply (H_2) and (3.25) to bound n'_0 and integrate. We obtain that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$

$$|h_1(t, x)| = \left| \int_0^{\frac{x}{N}} \partial_x u^{(1)}(t, v) \, dv \right| = \left| \int_0^{\frac{x}{N}} n'_0(bm_1 t + v) \, dv \right| \leq \frac{C'_\lambda x}{N} \exp \left[-bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right) t \right].$$

This is exactly (3.13) with $C' = C'_\lambda$. Since all the assumptions of Lemma 3.6 are verified, we apply this lemma to bound $v_\xi = \bar{u}_\partial^{(1)}$ (the constants C , σ_ξ , α and β in (3.15) are the same as for the first statement). We then use that $bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right) = bm_1 \lambda_N$ to simplify the term $bN \left(1 - \mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right)$ in the bound, see (2.9). We obtain Proposition 3.2 (b). \square

4. THE MODEL WITH SEVERAL TELOMERES

Having solved the single-telomere case, let us now turn to the case with $2k$ telomeres. The difference with the previous section is that here, we have to control the term in the denominator of $\hat{n}_0^{(2k)}$, see (2.10). In addition, we need to prove an additional statement to Lemmas 3.5 and 3.6 to obtain the model approximation. First, in

Section 4.1, we rewrite (2.5), and explain how a model approximation can be conjectured from this rewriting. Then, in Section 4.2, we present the main arguments and the auxiliary statements required to approximate our model and obtain Theorem 2.7 (b). Thereafter, in Section 4.3, we prove these auxiliary statements. Finally, we prove the approximation of $n^{(2k)}$ in Section 4.4, the approximation of $n_{\partial}^{(2k)}$ in Section 4.5, and Theorem 2.7 (b) in Section 4.6. In all this section, we assume that (H_1) holds.

4.1. Rewriting of the integro-differential equation and model approximation

As for the single-telomere model, we obtain the model approximation by letting the scaling parameter N introduced in (H_1) tend to infinity. We split this subsection into two parts: first, we rewrite our model in order to make N appears. Then, we intuit and state a model approximation based on this rewriting.

Rewriting of the model. Our first goal is to substitute the term $\int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + v) \tilde{\mu}(dv)$ in the first line of (2.5) for the term $\int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + \frac{v}{N}) \mu(dv)$. To do so, we first denote for all $i \in \llbracket 1, 2k \rrbracket$ the i th vector of the canonical basis $e_i \in \mathbb{R}_+^{2k}$. Then, we use both equalities in (2.7) and integrate. We obtain that for all $t \geq 0$, $x \in \mathbb{R}_+^{2k}$,

$$\int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}\left(t, x + \frac{v}{N}\right) \mu(dv) = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \int_{(v_i, i \in I) \in \mathbb{R}_+^k} n^{(2k)}\left(t, x + \sum_{i \in I} \frac{v_i}{N} e_i\right) \left[\prod_{i \in I} g(v_i) dv_i \right]. \tag{4.1}$$

We therefore need to prove that $\int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + v) \tilde{\mu}(dv)$ is equal to the right-hand side of (4.1), and the substitution will be possible. To do this, we first develop $\tilde{\mu}$ by using (2.4) and (2.3). Then, for all $I \in \mathcal{I}_k$, $i \in I$, we replace $\tilde{g}(v_i)$ with $Ng(Nv_i)$ in view of (H_1) , and we integrate. Finally, we do the change of variable $v'_i = Nv_i$ in each coordinate. It comes for all $t \geq 0$, $x \in \mathbb{R}_+^{2k}$,

$$\begin{aligned} \int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + v) \tilde{\mu}(dv) &= \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \int_{(v_i, i \in I) \in \mathbb{R}_+^k} n^{(2k)}\left(t, x + \sum_{i \in I} v_i e_i\right) \left[\prod_{i \in I} Ng(Nv_i) dv_i \right] \\ &= \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \int_{(v'_i, i \in I) \in \mathbb{R}_+^k} n^{(2k)}\left(t, x + \sum_{i \in I} \frac{v'_i}{N} e_i\right) \left[\prod_{i \in I} g(v'_i) dv'_i \right]. \end{aligned}$$

Therefore, it holds $\int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + v) \tilde{\mu}(dv) = \int_{v \in \mathbb{R}_+^{2k}} n^{(2k)}(t, x + \frac{v}{N}) \mu(dv)$. We now rewrite the first line of (2.5) by using this last equality. First, we plug it in (2.5). Then, we apply (H_1) to replace \tilde{b} with bN . We obtain the following rewriting, for all $t \geq 0$, $x \in \mathbb{R}_+^{2k}$,

$$\partial_t n^{(2k)}(t, x) = bN \int_{v \in \mathbb{R}_+^{2k}} \left[n^{(2k)}\left(t, x + \frac{v}{N}\right) - n^{(2k)}(t, x) \right] d\mu(v). \tag{4.2}$$

Now, we aim at rewriting the second line of (2.5). In Section 3.1, the rewriting of the second line of (2.1) was obtained by proving an equality involving \tilde{G} and G , given in (3.2). Here, we need to obtain the equivalent equality, involving $\tilde{\mu}$ and μ , which is the following, for all $y \in \mathbb{R}_+^{2k}$,

$$\tilde{\mu}(\{v \in \mathbb{R}_+^{2k} \mid y - v \in \mathbb{R}_+^{2k}\}) = \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \in \mathbb{R}_+^{2k}\}). \tag{4.3}$$

By first using (2.7) to develop μ , then integrating, and finally applying (3.2), we have

$$\begin{aligned} \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \in \mathbb{R}_+^{2k}\}) &= \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \int_{v \in \mathbb{R}_+^{2k}} 1_{\{\forall i \in \llbracket 1, 2k \rrbracket: Ny_i \geq v_i\}} \left[\prod_{i \in I} g(v_i) dv_i \right] \left[\prod_{i \notin I} \delta_0(dv_i) \right] \\ &= \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \left(\prod_{i \in I} G(Ny_i) \right) = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \left(\prod_{i \in I} \tilde{G}(y_i) \right). \end{aligned} \tag{4.4}$$

In addition, by doing the first two steps presented above, replacing g with \tilde{g} , we have that the left-hand side of (4.3) is equal to the last term in (4.4). Hence, from this last result, we have that (4.3) is true. We now use this last equation to rewrite the second line of (2.5). First, we take the complement of both sets in (4.3). Then, we plug the equality obtained in the second line of (2.5). Finally, we replace \tilde{b} with bN , in view of (H₁). We obtain that for all $t \geq 0$

$$n_{\partial}^{(2k)}(t) = bN \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) dy. \quad (4.5)$$

We conclude by presenting the full rewriting of the model. By combining (4.2) and (4.5), we have the following new version of (2.5)

$$\begin{cases} \partial_t n^{(2k)}(t, x) = bN \int_{v \in \mathbb{R}_+^{2k}} [n^{(2k)}(t, x + \frac{v}{N}) - n^{(2k)}(t, x)] d\mu(v), & \forall t \geq 0, x \in \mathbb{R}_+^{2k}, \\ n_{\partial}^{(2k)}(t) = bN \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) dy, & \forall t \geq 0, \\ n^{(2k)}(0, x) = \prod_{i=1}^{2k} n_0(x_i), & \forall x \in \mathbb{R}_+^{2k}. \end{cases} \quad (4.6)$$

The above is the system of integro-differential equations that we study in the rest of the section.

Model approximation. As for the single-telomere model, we start by giving the intuition allowing us to obtain the approximant of (4.6). We begin by conjecturing the approximation of the first line of (4.6). In view of the Taylor expansion, we have that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$bN \int_{v \in \mathbb{R}_+^{2k}} \left[n^{(2k)}\left(t, x + \frac{v}{N}\right) - n^{(2k)}(t, x) \right] d\mu(v) \underset{N \rightarrow +\infty}{\approx} b \sum_{i=1}^{2k} \int_{v \in \mathbb{R}_+^{2k}} v_i d\mu(v) \partial_{x_i} n^{(2k)}(t, x).$$

In addition, the explicit value of $b \int_{v \in \mathbb{R}_+^{2k}} v_i d\mu(v)$ can be computed for all $i \in \llbracket 1, 2k \rrbracket$ thanks to the following statement.

Lemma 4.1 (First moment of μ). *Assume that (H₁) holds. Then, for all $i \in \llbracket 1, 2k \rrbracket$, we have*

$$\int_{v \in \mathbb{R}_+^{2k}} v_i d\mu(v) = \frac{m_1}{2}.$$

Proof. Let $i \in \llbracket 1, 2k \rrbracket$. To obtain the above expression, we first develop μ by using the right-hand side of (2.7). Then, we integrate both measures, in view of the fact that for all $I \in \mathcal{I}_k$, it holds by the left-hand side of (2.7): $\int_{v \in \mathbb{R}_+^{2k}} v_i \mu^{(I)}(dv) = m_1 1_{\{i \in I\}}$. We obtain

$$\int_{v \in \mathbb{R}_+^{2k}} v_i d\mu(v) = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} m_1 1_{\{i \in I\}} = \frac{m_1 \#(\{I \in \mathcal{I}_k \mid i \in I\})}{2^k}.$$

The lemma is then proved by using Lemma A.6 to compute the ratio. \square

Therefore, by denoting the vector $1_{2k} := (1, \dots, 1) \in \mathbb{R}^{2k}$, our conjecture is that (4.6) can be approximated by the function $u^{(2k)} \in C(\mathbb{R}_+, W^{2,1}(\mathbb{R}_+^{2k}))$, solution of the following equation, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$\partial_t u^{(2k)}(t, x) = \frac{bm_1}{2} \nabla \cdot (u^{(2k)} 1_{2k})(t, x). \quad (4.7)$$

We now intuit how the second line of (4.6) can be approximated. In comparison with what is done in Section 3.1, it is necessary to decompose $n_{\partial}^{(2k)}$ into two terms. Notice that by the second equality in (4.4), the fact that $G(x) = 1$ for all $x > \delta$, and Lemma A.6, we have

$$\forall y \in \left(\frac{\delta}{N}, +\infty\right)^{2k} : \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) = 1 - \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \left(\prod_{i \in I} G(Ny_i) \right) = 0. \quad (4.8)$$

Then, combining the above with the second line of (4.6) yields that our decomposition of $n_{\partial}^{(2k)}$ is the following, for all $t \geq 0$,

$$\begin{aligned} n_{\partial}^{(2k)}(t) &= bN \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} = 1\}} dy \\ &\quad + bN \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} \geq 2\}} dy \\ &=: n_{\partial,1}^{(2k)}(t) + n_{\partial,2}^{(2k)}(t). \end{aligned} \tag{4.9}$$

In fact, $n_{\partial,1}^{(2k)}$ and $n_{\partial,2}^{(2k)}$ have a different behaviour as $N \rightarrow +\infty$. The behaviour of $n_{\partial,2}^{(2k)}$ is easy to conjecture. As the domain of integration of the integral in its definition corresponds to the points with at least two coordinates smaller than $\frac{\delta}{N}$, we have that this integral vanishes at least at a rate $\frac{1}{N^2}$ as $N \rightarrow +\infty$. Then, by multiplying it by bN , we obtain that for all $t \geq 0$

$$n_{\partial,2}(t) = bN \times O\left(\frac{1}{N^2}\right)_{N \rightarrow +\infty} \approx 0. \tag{4.10}$$

To conjecture the behaviour of $n_{\partial,1}^{(2k)}$ as $N \rightarrow +\infty$, we need to obtain a better expression for it. To do so, in its definition given in the first line of (4.9), we first decompose the indicator, by using that for all $y \in \mathbb{R}_+^{2k}$, we have

$$1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} = 1\}} = \sum_{i=1}^{2k} 1_{\{y_i \leq \frac{\delta}{N}, \forall j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y_j > \frac{\delta}{N}\}}.$$

Then, we use the following equality to replace the term $\mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\})$ in the first line of (4.9). This equality is obtained by applying the second equality in (4.4), the fact that $G(x) = 1$ when $x > \delta$, and Lemma A.6. For all $i \in \llbracket 1, 2k \rrbracket$ and $y \in \mathbb{R}_+^{2k}$ verifying $y_i \leq \frac{\delta}{N}$ and $y_j > \frac{\delta}{N}$ when $j \neq i$, it holds

$$\begin{aligned} \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) &= 1 - \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \left(\prod_{j \in I} G(Ny_j) \right) \\ &= 1 - \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} (G(Ny_i) 1_{\{i \in I\}} + 1_{\{i \notin I\}}) = \frac{1}{2} (1 - G(Ny_i)). \end{aligned}$$

Finally, for each index $i \in \llbracket 1, 2k \rrbracket$ of the sum, we do the change of variables $y'_i = Ny_i$, and $y'_j = y_j$ for all $j \in \llbracket 1, 2k \rrbracket \setminus \{i\}$. We obtain that for all $t \geq 0$

$$\begin{aligned} n_{\partial,1}^{(2k)}(t) &= \frac{bN}{2} \sum_{i=1}^{2k} \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) (1 - G(Ny_i)) 1_{\{y_i \leq \frac{\delta}{N}, \forall j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y_j > \frac{\delta}{N}\}} dy \\ &= \frac{b}{2} \sum_{i=1}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} n^{(2k)}\left(t, \sum_{j=1, j \neq i}^{2k} y'_j e_j + \frac{y'_i}{N} e_i\right) (1 - G(y'_i)) 1_{\{y'_i \leq \delta, \forall j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y'_j > \frac{\delta}{N}\}} dy'. \end{aligned} \tag{4.11}$$

Then, we are now able to conjecture the behaviour of $n_{\partial,1}^{(2k)}$ as $N \rightarrow +\infty$. Indeed, by letting N tend to infinity, and by using that $\int_{s \in [0, \delta]} (1 - G(s)) ds = m_1$, we have for all $t \geq 0$

$$n_{\partial,1}^{(2k)}(t) \underset{N \rightarrow +\infty}{\approx} \frac{bm_1}{2} \sum_{i=1}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \delta_0(dy_i) \left(\prod_{j=1, j \neq i}^{2k} dy_j \right).$$

Combining the above with (4.10) and (4.7) finally yields that the following system seems a good approximant of (4.6)

$$\begin{cases} \partial_t u^{(2k)}(t, x) = \frac{bm_1}{2} \nabla \cdot (u^{(2k)} 1_{2k})(t, x), & \forall t \geq 0, x \in \mathbb{R}_+^{2k}, \\ u_{\partial}^{(2k)}(t) = \frac{bm_1}{2} \sum_{i=1}^{2k} \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \delta_0(dy_i) \left(\prod_{j=1, j \neq i}^{2k} dy_j \right), & \forall t \geq 0, \\ u^{(2k)}(0, x) = \prod_{i=1}^{2k} n_0(x_i), & \forall x \in \mathbb{R}_+^{2k}. \end{cases} \quad (4.12)$$

Remark 4.2. By using the method of characteristics to solve the first line of (4.12), we have for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$u^{(2k)}(t, x) = \prod_{i=1}^{2k} n_0\left(\frac{bm_1}{2}t + x_i\right). \quad (4.13)$$

In addition, by plugging (4.13) in the second line of (4.12), and then doing the change of variables $y' = y + \frac{bm_1 t}{2} 1_{2k}$, we have for all $t \geq 0$

$$u_{\partial}^{(2k)}(t) = \frac{bm_1}{2} \sum_{i=1}^{2k} n_0\left(\frac{bm_1 t}{2}\right) \prod_{\substack{j=1 \\ j \neq i}}^{2k} \left[\int_{\frac{bm_1 t}{2}}^{+\infty} n_0(y'_j) dy'_j \right] = kbm_1 n_0\left(\frac{bm_1 t}{2}\right) \left[\int_{\frac{bm_1 t}{2}}^{+\infty} n_0(s) ds \right]^{2k-1}. \quad (4.14)$$

Remark 4.3. The conservation of the number of individuals, presented in Remark 2.5, still holds for (4.12) when (H_2) – (H_3) are verified. To obtain it, first integrate (4.14), in view of the fact that for all $f \in W^{1,1}(\mathbb{R}_+) \cap W^{1,2k}(\mathbb{R}_+) \cap L^{2k-1}(\mathbb{R}_+)$ it holds $(f^{2k})' = 2k f' f^{2k-1}$. Then, do the change of variable $s' = s - \frac{bm_1 t}{2}$. Finally, apply (4.13). It comes for all $t \geq 0$

$$\int_t^{+\infty} u_{\partial}^{(2k)}(s) ds = \left[\int_{\frac{bm_1 t}{2}}^{+\infty} n_0(s) ds \right]^{2k} = \left[\int_0^{+\infty} n_0\left(\frac{bm_1}{2}t + s'\right) ds' \right]^{2k} = \int_{x \in \mathbb{R}_+^{2k}} u(t, x) dx. \quad (4.15)$$

Remark 4.4. Even in the case where the assumption presented in (2.4) does not hold, the computations done to obtain the first equality in (4.14) can be done, and yield that for all $t \geq 0$

$$u_{\partial}^{(2k)}(t) = \frac{bm_1}{2} \sum_{i=1}^{2k} \int_{(y'_j, j \in [1, 2k] \setminus \{i\}) \in \mathbb{R}_+^{2k-1}} n\left(0, \frac{bm_1 t}{2} e_i + \sum_{j=1, j \neq i}^{2k} y'_j e_j\right) \left[\prod_{j=1, j \neq i}^{2k} 1_{\{y'_j \geq \frac{bm_1 t}{2}\}} dy'_j \right].$$

We thus have by integrating, and then doing the change of variable $y'' = \frac{bm_1 s}{2} e_i + \sum_{j=1, j \neq i}^{2k} y'_j e_j$, that for all $t \geq 0$

$$\int_t^{+\infty} u_{\partial}^{(2k)}(s) ds = \sum_{i=1}^{2k} \int_{y'' \in \mathbb{R}_+^{2k}} n(0, y'') 1_{\{y''_i \geq \frac{bm_1 t}{2}\}} 1_{\{\forall j \in [1, 2k]: y''_j \geq y''_i\}} dy.$$

The latter yields, in view of the fact that $y''_j \geq y''_i$ for all $j \neq i$, that for all $t \geq 0$

$$\int_t^{+\infty} u_{\partial}^{(2k)}(s) ds = \int_{y'' \in \mathbb{R}_+^{2k}} n(0, y'') 1_{\{\min_{1 \leq i \leq 2k} (y''_i) \geq \frac{bm_1 t}{2}\}} dy''.$$

Hence, $u_{\partial}^{(2k)}$ is the distribution of the minimum initial telomere length, rescaled by $\frac{bm_1}{2}$.

All we have intuited above can in fact be justified rigorously. Specifically, the proposition below provides that the error between $(n^{(2k)}, n_{\partial}^{(2k)})$ and $(u^{(2k)}, u_{\partial}^{(2k)})$ tends to 0 when $N \rightarrow +\infty$. Its proof is separated into two parts, one part for each statement, done in Sections 4.4 and 4.5.

Proposition 4.5 (Pointwise approximation errors, several telomeres). *We recall the constant λ'_N defined in (2.9). The following statements hold.*

(a) *Assume (H₁)–(H₃). Then, there exists $d'_0 > 0$ such that for all $t \geq 0$, $x \in \mathbb{R}_+^{2k}$, we have*

$$\left| n^{(2k)}(t, x) - u^{(2k)}(t, x) \right| \leq d'_0 (\lambda D_\lambda)^{2k} \frac{k^2 bt}{2N} \exp(-k b m_1 \lambda'_N t) \exp\left(-\lambda \sum_{i=1}^{2k} x_i\right),$$

where d'_0 depends only on $g, \lambda, C_\lambda, C'_\lambda$ and D_λ .

(b) *Assume (H₁)–(H₃). Then, there exists $d'_1 > 0$ such that for all $t \geq 0$, we have*

$$\left| n_{\partial}^{(2k)}(t) - u_{\partial}^{(2k)}(t) \right| \leq \frac{b d'_1}{N} (D_\lambda)^{2k} \left(k^3 \frac{b m_1}{2} t + k^2 + k \right) \exp(-k b m_1 \lambda'_N t), \tag{4.16}$$

where d'_1 depends only on $g, \delta, \lambda, C_\lambda, C'_\lambda$ and D_λ .

Remark 4.6. As for Proposition 3.2, a result providing bounds on the approximation errors in all Lebesgue spaces can be obtained from the above, that we do not state.

We now present the main arguments and the auxiliary statements required to prove Proposition 4.5 and Theorem 2.7 (b).

4.2. Plan of the proofs of Proposition 4.5 and Theorem 2.7 (b)

We begin by presenting the plan to prove Proposition 4.5 (a). To get this statement, we need to control the absolute value of $\bar{u}^{(2k)} := n^{(2k)} - u^{(2k)}$. By proceeding as when we obtained (3.7), using this time Lemma 4.1 to write the term $\frac{b m_1}{2} \nabla \cdot (u^{(2k)} 1_{2k})(t, x)$ coming from (4.12) as an integral, we have that for all $t \geq 0, x \in \mathbb{R}_+^{2k}$

$$\begin{aligned} \partial_t \bar{u}^{(2k)}(t, x) &= bN \int_0^\delta \left[\bar{u}^{(2k)}\left(t, x + \frac{v}{N}\right) - \bar{u}^{(2k)}(t, x) \right] \mu(dv) \\ &\quad + bN \int_0^\delta \left[u^{(2k)}\left(t, x + \frac{v}{N}\right) - u^{(2k)}(t, x) - \sum_{i=1}^{2k} \frac{v_i}{N} \partial_{x_i} u^{(2k)}(t, x) \right] \mu(dv). \end{aligned} \tag{4.17}$$

Then, $\bar{u}^{(2k)}$ is a solution of an equation of the form presented in (3.8), with $d = 2k, u_\xi = \bar{u}^{(2k)}, \xi = \mu$ and $F = u^{(2k)}$. From this result, a natural plan to prove Proposition 4.5 is to proceed as in the proof of Proposition 3.2 (a), and check the assumptions of Lemma 3.5 to obtain Proposition 4.5 (a). Here, we do not do this directly because we must first obtain the value of the constant σ_ξ , defined in Lemma 3.5, for $\xi = \mu$. This value is given in the following statement, proved in Section 4.3.1.

Lemma 4.7 (Sum of second moments of μ). *Assume that (H₁) holds. Then, we have*

$$\sigma_\mu := \sum_{1 \leq \ell, \ell' \leq 2k} \int_{v \in \mathbb{R}_+^{2k}} v_\ell v_{\ell'} \mu(dv) = (m_1)^2 k^2 + (m_2 - (m_1)^2) k.$$

Now that we know the value of σ_μ , we prove Proposition 3.2 (a) in Section 4.4 by following the steps presented above.

To prove Proposition 4.5 (b), we control $|\bar{u}_{\partial}^{(2k)}| := |n_{\partial}^{(2k)} - u_{\partial}^{(2k)}|$. To do this, a decomposition of $\bar{u}_{\partial}^{(2k)}$ is required, inspired by the one given in (4.9). Let us introduce for all $t \geq 0$

$$\begin{aligned} u_{\partial,0}^{(2k)}(t) &:= bN \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) dy, \\ u_{\partial,1}^{(2k)}(t) &:= \frac{b}{2} \sum_{i=1}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} u^{(2k)}\left(t, \sum_{j=1, j \neq i}^{2k} y'_j e_j + \frac{y'_i}{N}\right) (1 - G(y'_i)) 1_{\{y'_i \leq \delta\}} dy', \\ u_{\partial,2}^{(2k)}(t) &:= u_{\partial,0}^{(2k)}(t) - u_{\partial,1}^{(2k)}(t). \end{aligned} \tag{4.18}$$

Then, as it holds $u_{\partial,0}^{(2k)} = u_{\partial,1}^{(2k)} + u_{\partial,2}^{(2k)}$, we have the following decomposition, for all $t \geq 0$,

$$\bar{u}_{\partial}^{(2k)}(t) = \left(n_{\partial}^{(2k)}(t) - u_{\partial,0}^{(2k)}(t) + u_{\partial,1}^{(2k)}(t) - u_{\partial}^{(2k)}(t) \right) + u_{\partial,2}^{(2k)}(t). \tag{4.19}$$

This decomposition allows us to control $|\bar{u}_{\partial}^{(2k)}|$, by bounding each of the terms that compose it. The term $n_{\partial}^{(2k)} - u_{\partial,0}^{(2k)} + u_{\partial,1}^{(2k)} - u_{\partial}^{(2k)}$, on the one hand, can be bounded by using Lemma 3.6, already proved in Section 3.4. The term $u_{\partial,2}^{(2k)}$, on the other hand, can be bounded with the following lemma, proved in Section 4.3.2.

Lemma 4.8 (Control of the probability of having several short telomeres). *Assume that (H₁) and (H₃) hold. Then, there exists $\tilde{d} > 0$ such that for all $t \geq 0$ we have*

$$\left| u_{\partial,2}^{(2k)}(t) \right| \leq \frac{b\tilde{d}}{N} (D_{\lambda})^{2k} k^2 \exp(-k b m_1 \lambda t), \tag{4.20}$$

where \tilde{d} depends only on δ, λ and g .

As a result, we have a control over $|\bar{u}_{\partial}^{(2k)}|$ by summing the bounds we have on the two terms presented above, in view of (4.19) and the triangle inequality. We detail how we apply these lemmas and sum the bounds to prove Proposition 4.5 (b) in Section 4.5.

The proof of Theorem 2.7 (b), finally, is based on the following statement. Its proof is relatively short, and consists in first taking $t = \frac{2x}{b m_1}$ in the first equality in (4.15), then raising both sides of the equality to the power $\frac{1}{2k}$, and finally taking the derivative.

Lemma 4.9 (Link between n_0 and u_{∂}). *Assume that (H₁)–(H₃) hold. Then, for all $x \geq 0$, we have*

$$n_0(x) = \frac{1}{k b m_1} \frac{u_{\partial}^{(2k)}\left(\frac{2x}{b m_1}\right)}{\left(\int_{\frac{2x}{b m_1}}^{\infty} u_{\partial}^{(2k)}(s) ds\right)^{1-\frac{1}{2k}}}. \tag{4.21}$$

From this lemma, if we replace $\hat{n}_0^{(2k)}$ with its definition (see (2.10)), thereafter use (2.11) to replace $\tilde{b} m_1$ with $b m_1$, and finally write n_0 with the right-hand side of (4.21), then we obtain that for all $x \geq 0$

$$\left| \hat{n}_0^{(2k)}(x) - n_0(x) \right| = \frac{1}{k b m_1} \left| \frac{n_{\partial}^{(2k)}\left(\frac{2x}{b m_1}\right)}{\left(\int_{\frac{2x}{b m_1}}^{\infty} n_{\partial}^{(2k)}(s) ds\right)^{1-\frac{1}{2k}}} - \frac{u_{\partial}^{(2k)}\left(\frac{2x}{b m_1}\right)}{\left(\int_{\frac{2x}{b m_1}}^{\infty} u_{\partial}^{(2k)}(s) ds\right)^{1-\frac{1}{2k}}} \right|. \tag{4.22}$$

We thus need to control this difference, and Theorem 2.7 (b) will be proved. Proposition 4.5 allows us to control the error between $n_{\partial}^{(2k)}$ and $u_{\partial}^{(2k)}$, or their integrals. However, there is still a difficulty related to the fact that the denominator of the two terms tends to 0 when $x \rightarrow +\infty$. We handle this by proving the following two lemmas, in Sections 4.3.3 and 4.3.4 respectively.

Lemma 4.10 (Power series expansion of the cemetery tail). *Assume that (H₁) and (H₃) hold. Let us consider the function $\tilde{n}_0 \in L^1(\mathbb{R}_+^{2k})$, defined for all $x \in \mathbb{R}_+^{2k}$ as*

$$\tilde{n}_0(x) := \prod_{i=1}^{2k} n_0(x_i). \tag{4.23}$$

Then, for all $t \geq 0$, we have

$$\int_t^{+\infty} n_{\partial}^{(2k)}(s) ds = e^{-b N t} \left[1 + \sum_{\ell \geq 1} \frac{(b N t)^{\ell}}{\ell!} \int_{(x,v) \in \mathbb{R}_+^{2k} \times (\mathbb{R}_+^{2k})^{\ell}} \tilde{n}_0 \left(x + \sum_{j=1}^{\ell} \frac{v_j}{N} \right) dx \mu(dv_1) \dots \mu(dv_{\ell}) \right]. \tag{4.24}$$

Lemma 4.11 (Lower bound for cemetery tails). *Assume that (H₁)–(H₄) hold, and recall the constant ω′_N defined in (2.9). Then, for all t ≥ 0, we have*

$$\int_t^{+\infty} n_{\partial}^{(2k)}(s) ds \geq (D_{\omega})^{2k} \exp(-k b m_1 \omega'_N t) \text{ and } \int_t^{+\infty} u_{\partial}^{(2k)}(s) ds \geq (D_{\omega})^{2k} \exp(-k b m_1 \omega t). \tag{4.25}$$

In fact, the statement that allows us to do this control is Lemma 4.11. Lemma 4.10 corresponds to an intermediate step to obtain Lemma 4.11. Thus, the proof of Theorem 2.7 (b) consists in controlling the right-hand side term of (4.22) by using Proposition 4.5 and Lemma 4.11. We do this in Section 4.6.

We now prove all the statements given in this section, and then obtain Proposition 4.5 and Theorem 2.7 (b) from them.

4.3. Proof of the auxiliary statements

This section is devoted to the proof of the auxiliary statements presented in Section 4.2. These statements are proved one by one, in the same order in which they were stated.

4.3.1. Proof of Lemma 4.7

We consider for all (ℓ, ℓ′) ∈ [1, 2k]² the integral I_{ℓ,ℓ′} := ∫_{y ∈ ℝ^{2k}₊} y_ℓ y_{ℓ′} μ(dv). One can easily see that

$$\sigma_{\mu} = \sum_{1 \leq \ell, \ell' \leq 2k} I_{\ell, \ell'}.$$

Thus, our aim is to compute the values of the integrals (I_{ℓ,ℓ′})_{(ℓ,ℓ′) ∈ [1,2k]²}, and then conclude by summing their values. To do this, we fix (ℓ, ℓ′) ∈ [1, 2k]² and do a distinction between cases.

Assume first that ℓ = ℓ′. In view of the left-hand side of (2.7), we have that for all I ∈ I_k

$$\int_{y \in \mathbb{R}_+^{2k}} y_{\ell} y_{\ell'} \mu^{(I)}(dy) = \begin{cases} \int_{y' \in [0, \delta]} (y')^2 g(y') dy' = m_2, & \text{if } \ell = \ell' \in I, \\ 0, & \text{otherwise.} \end{cases}$$

Then, by using the right-hand side of (2.7) and Lemma A.6, we obtain that

$$I_{\ell, \ell} = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k, \ell = \ell' \in I} m_2 = \frac{m_2}{2}. \tag{4.26}$$

Assume now that ℓ ≠ ℓ′. In this case, in view of the left-hand side of (2.7), we have that for all I ∈ I_k

$$\int_{y \in \mathbb{R}_+^{2k}} y_{\ell} y_{\ell'} \mu^{(I)}(dy) = \begin{cases} \left(\int_{y' \in [0, \delta]} y' g(y') dy' \right)^2 = (m_1)^2, & \text{if } \ell \in I \text{ and } \ell' \in I, \\ 0, & \text{otherwise.} \end{cases} \tag{4.27}$$

In addition, by the definition of I_k (see (2.2)), we have that when ℓ = ℓ′ mod k

$$\{I \in \mathcal{I}_k \mid \{\ell, \ell'\} \subset I\} = \emptyset.$$

Then, by combining these results, we obtain that when ℓ ≠ ℓ′ and ℓ = ℓ′ mod k

$$I_{\ell, \ell'} = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k, \ell \in I, \ell' \in I} (m_1)^2 = 0. \tag{4.28}$$

We also obtain by combining (4.27) with Lemma A.7 that when ℓ ≠ ℓ′ mod k

$$I_{\ell, \ell'} = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k, \ell \in I, \ell' \in I} (m_1)^2 = \begin{cases} 0, & \text{if } k = 1, \\ \frac{(m_1)^2}{4}, & \text{if } k \geq 2. \end{cases} \tag{4.29}$$

We now conclude. First, we combine (4.26), (4.28) and (4.29). Then, we use the fact that as the set

$$\{(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2 \mid \ell = \ell' \bmod k\} = \left(\bigcup_{i=1}^{2k} \{(i, i)\} \right) \cup \left(\bigcup_{i=1}^k \{(i, i+k)\} \right) \cup \left(\bigcup_{i=k+1}^{2k} \{(i, i-k)\} \right)$$

has a cardinality of $4k$, it holds $\#\{(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2 \mid \ell \neq \ell' \bmod k\} = (2k)^2 - 4k$. Finally, we use that as $k \in \mathbb{N}^*$, we have $((2k)^2 - 4k)1_{\{k \geq 2\}} = (2k)^2 - 4k$. We obtain the following, which ends the proof

$$\begin{aligned} \sigma_\mu &= \sum_{1 \leq \ell = \ell' \leq 2k} I_{\ell, \ell} + \sum_{\substack{1 \leq \ell \neq \ell' \leq 2k \\ \text{s.t. } \ell = \ell' \bmod k}} I_{\ell, \ell'} + \sum_{\substack{1 \leq \ell \neq \ell' \leq 2k \\ \text{s.t. } \ell \neq \ell' \bmod k}} I_{\ell, \ell'} \\ &= 2k \frac{m_2}{2} + 0 + ((2k)^2 - 4k) \frac{(m_1)^2}{4} 1_{\{k \geq 2\}} = (m_1)^2 k^2 + (m_2 - (m_1)^2) k. \end{aligned}$$

□

4.3.2. Proof of Lemma 4.8

We first need to obtain a better expression for $u_{\partial,2} = u_{\partial,0} - u_{\partial,1}$, where $u_{\partial,0}$ and $u_{\partial,1}$ are defined in (4.18). To do so, we develop the function $u_{\partial,0}$ in its definition. One can notice that by the definition of $u_{\partial,0}$ and (4.8), it holds for all $t \geq 0$

$$\begin{aligned} u_{\partial,0}(t) &= bN \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} = 1\}} dy \\ &\quad + bN \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} \geq 2\}} dy. \end{aligned}$$

The function on the first line of the above has the same definition as $n_{\partial,1}^{(2k)}$, see (4.9), with $u^{(2k)}$ instead of $n^{(2k)}$. Then, following exactly the same steps as those to obtain (4.11), replacing $n^{(2k)}$ with $u^{(2k)}$, yields that for all $t \geq 0$

$$\begin{aligned} u_{\partial,0}(t) &= \frac{b}{2} \sum_{i=1}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} u^{(2k)} \left(t, \sum_{j=1, j \neq i}^{2k} y'_j e_j + \frac{y'_i}{N} e_i \right) (1 - G(y'_i)) 1_{\{y'_i \leq \delta, \forall j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y'_j > \frac{\delta}{N}\}} dy' \\ &\quad + bN \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} \geq 2\}} dy. \end{aligned} \tag{4.30}$$

Now, we use the above equation to develop $u_{\partial,2}$. Specifically, we subtract $u_{\partial,1}$ from both sides of (4.30), in view of (4.18) and the following equality

$$1_{\{y'_i \leq \delta, \forall j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y'_j > \frac{\delta}{N}\}} - 1_{\{y'_i \leq \delta\}} = -1_{\{y'_i \leq \delta, \exists j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y'_j \leq \frac{\delta}{N}\}}.$$

We obtain that for all $t \geq 0$

$$\begin{aligned} u_{\partial,2}(t) &= -\frac{b}{2} \sum_{i=1}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} u^{(2k)} \left(t, \sum_{\ell=1, \ell \neq i}^{2k} y'_\ell e_\ell + \frac{y'_i}{N} e_i \right) (1 - G(y'_i)) 1_{\{y'_i \leq \delta, \exists j \in \llbracket 1, 2k \rrbracket \setminus \{i\}: y'_j \leq \frac{\delta}{N}\}} dy' \\ &\quad + bN \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) \mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\}) 1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} \geq 2\}} dy \\ &=: -u_{\partial,3}(t) + u_{\partial,4}(t). \end{aligned}$$

As $-u_{\partial,3}$ and $u_{\partial,4}$ have an opposite sign, the above implies that $|u_{\partial,2}(t)| \leq \max(u_{\partial,3}(t), u_{\partial,4}(t))$ for all $t \geq 0$. We thus now obtain an upper bound for both $u_{\partial,3}$ and $u_{\partial,4}$ in order to prove (4.20).

To bound $u_{\partial,3}$, we first bound $1_{\{y'_i \leq \delta, \exists j \in \llbracket 1, 2k \rrbracket \setminus \{i\} : y'_j \leq \frac{\delta}{N}\}}$ by $\sum_{j \in \llbracket 1, 2k \rrbracket \setminus \{i\}} 1_{\{y'_i \leq \delta, y'_j \leq \frac{\delta}{N}\}}$. Then, we use (4.13) to write $u^{(2k)}$ in terms of n_0 , and apply (H₃) to bound n_0 . Finally, we integrate in dy' , and use the equality $\int_0^\delta (1 - G(y'_i)) dy'_i = m_1$, the inequality $\int_0^{\frac{\delta}{N}} \exp(-\lambda y'_j) dy'_j \leq \frac{\delta}{N}$, and the equality $\int_0^{+\infty} \exp(-\lambda y'_\ell) dy'_\ell = \frac{1}{\lambda}$ for all $\ell \in \llbracket 1, 2k \rrbracket \setminus \{i, j\}$. We obtain that for all $t \geq 0$

$$\begin{aligned} |u_{\partial,3}(t)| &\leq \frac{b}{2} \sum_{i=1}^{2k} \sum_{\substack{j=1 \\ j \neq i}}^{2k} \int_{y' \in \mathbb{R}_+^{2k}} n_0\left(\frac{bm_1}{2}t + \frac{y'_i}{N}\right) \left[\prod_{\substack{\ell=1 \\ \ell \neq i}}^{2k} n_0\left(\frac{bm_1}{2}t + y_\ell\right) \right] (1 - G(y'_i)) 1_{\{y'_i \leq \delta, y_j \leq \frac{\delta}{N}\}} dy' \\ &\leq \frac{bm_1\delta}{2N} \sum_{i=1}^{2k} \sum_{\substack{j=1 \\ j \neq i}}^{2k} \frac{\lambda^{2k} (D_\lambda)^{2k}}{\lambda^{2k-2}} \exp(-kbm_1\lambda t) = 2k(2k-1) \frac{bm_1\delta}{2N} \lambda^2 (D_\lambda)^{2k} \exp(-kbm_1\lambda t). \end{aligned} \tag{4.31}$$

To bound $u_{\partial,4}$, we first bound the term $\mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\})$ in $u_{\partial,4}$ by 1, as μ is a probability measure. Then, we bound the term $1_{\{\#\{i \in \llbracket 1, 2k \rrbracket \mid y_i \leq \frac{\delta}{N}\} \geq 2\}}$ by the sum $\sum_{(i,j) \in \llbracket 1, 2k \rrbracket^2, i \neq j} 1_{\{y_i \leq \frac{\delta}{N}, y_j \leq \frac{\delta}{N}\}}$. Finally, as done before, we successively use equation (4.13) to write $u^{(2k)}$ in terms of n_0 , apply (H₃) to bound from above n_0 , and integrate by using the same inequalities/equalities. We obtain that for all $t \geq 0$

$$\begin{aligned} |u_{\partial,4}(t)| &\leq bN \sum_{i=1}^{2k} \sum_{\substack{j=1 \\ j \neq i}}^{2k} \int_{y \in \mathbb{R}_+^{2k}} u^{(2k)}(t, y) 1_{\{y_i \leq \frac{\delta}{N}, y_j \leq \frac{\delta}{N}\}} dy \\ &\leq 2k(2k-1) \frac{b\delta^2}{N} \lambda^2 (D_\lambda)^{2k} \exp(-kbm_1\lambda t). \end{aligned} \tag{4.32}$$

From (4.31), (4.32) and the fact that $|u_{\partial,2}| \leq \max(u_{\partial,3}, u_{\partial,4})$, the lemma is proved. □

4.3.3. Proof of Lemma 4.10

We consider two functions $F : \mathbb{R}_+ \times \mathbb{R}_+^{2k} \rightarrow \mathbb{R}$ and $\bar{F} : \mathbb{R}_+ \times \mathbb{R}_+^{2k} \rightarrow \mathbb{R}$, defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$ as

$$\begin{aligned} F(t, x) &:= \tilde{n}_0(x) + \sum_{\ell \geq 1} \frac{(bN)^\ell t^\ell}{\ell!} \int_{v_1 \in \mathbb{R}_+^{2k}} \cdots \int_{v_\ell \in \mathbb{R}_+^{2k}} \tilde{n}_0\left(x + \sum_{j=1}^{\ell} \frac{v_j}{N}\right) \mu(dv_\ell) \cdots \mu(dv_1), \\ \bar{F}(t, x) &:= e^{-bNt} F(t, x). \end{aligned} \tag{4.33}$$

We begin by proving that $n^{(2k)} = \bar{F}$. By deriving F , and then taking $m = \ell - 1$, we have that for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$\begin{aligned} \partial_t F(t, x) &= bN \sum_{m \geq 0} \frac{(bNt)^m}{m!} \int_{v_1 \in \mathbb{R}_+^{2k}} \cdots \int_{v_{m+1} \in \mathbb{R}_+^{2k}} \tilde{n}_0\left(x + \sum_{j=1}^{m+1} \frac{v_j}{N}\right) \mu(dv_{m+1}) \cdots \mu(dv_1) \\ &= bN \int_{v_1 \in \mathbb{R}_+^{2k}} F\left(t, x + \frac{v_1}{N}\right) \mu(dv_1). \end{aligned}$$

Then, by combining the above with the fact that $\partial_t \bar{F}(t, x) = e^{-bNt} \partial_t F(t, x) - bN e^{-bNt} F(t, x)$, we obtain that \bar{F} verifies the same integro-differential equation as $n^{(2k)}$ (see Eq. (4.6)), so that $n^{(2k)} = \bar{F}$ by Proposition A.1.

Now, we prove equation (4.24). By applying the equality on the right-hand side of (2.6), and then combining the fact that $n^{(2k)} = \bar{F}$ with the second line of (4.33), we have that for all $t \geq 0$

$$\int_t^{+\infty} n_{\partial}^{(2k)}(s) ds = e^{-bNt} \int_{x \in \mathbb{R}_+^{2k}} F(t, x) dx.$$

Therefore, by plugging the first line of (4.33) in the above equation, and then using that $\int_{x \in \mathbb{R}_+^{2k}} \tilde{n}_0(x) dx = 1$ to compute the term that is not in the sum, we obtain that equation (4.24) is true, which ends the proof. \square

4.3.4. Proof of Lemma 4.11

To prove this lemma, we proceed in two steps. In Step 1, we prove the left-hand side of (4.25), and in Step 2, we prove the right-hand side of (4.25).

Step 1. To simplify notations, we use in this step the function \tilde{n}_0 , defined in (4.23). Our aim here is to bound from below the right-hand side term of (4.24). To do so, we begin by bounding the coefficients in the sum. By applying (H4), and then using that f_ω is non-decreasing, we have that for all $\ell \in \mathbb{N}^*$, $x \in \mathbb{R}_+^{2k}$ and $(v_1, \dots, v_\ell) \in (\mathbb{R}_+^{2k})^\ell$

$$\begin{aligned} \tilde{n}_0\left(x + \sum_{j=1}^{\ell} \frac{v_j}{N}\right) &\geq (D_\omega)^{2k} \frac{\prod_{i=1}^{2k} \left(f_\omega\left(x_i + \sum_{j=1}^{\ell} \frac{(v_j)_i}{N}\right) \exp\left[-\omega\left(x_i + \sum_{j=1}^{\ell} \frac{(v_j)_i}{N}\right)\right]\right)}{\left(\int_0^{+\infty} f_\omega(y) \exp(-\omega y) dy\right)^{2k}} \\ &\geq (D_\omega)^{2k} \frac{\prod_{i=1}^{2k} (f_\omega(x_i) \exp[-\omega x_i])}{\left(\int_0^{+\infty} f_\omega(y) \exp(-\omega y) dy\right)^{2k}} \exp\left(-\omega \sum_{i=1}^{2k} \sum_{j=1}^{\ell} \frac{(v_j)_i}{N}\right). \end{aligned}$$

Then, by integrating both sides in dx and $(\mu(dv_i))_{i \in [1, \ell]}$, and simplifying the last term with a Laplace transform, we obtain that for all $\ell \in \mathbb{N}^*$

$$\begin{aligned} \int_{x \in \mathbb{R}_+^{2k}} \int_{v_1 \in \mathbb{R}_+^{2k}} \dots \int_{v_\ell \in \mathbb{R}_+^{2k}} \tilde{n}_0\left(x + \sum_{j=1}^{\ell} \frac{v_j}{N}\right) \mu(dv_\ell) \dots \mu(dv_1) dx \\ \geq (D_\omega)^{2k} \left[\prod_{j=1}^{\ell} \int_{v_j \in \mathbb{R}_+^{2k}} \exp\left(-\omega \sum_{i=1}^{2k} \frac{(v_j)_i}{N}\right) \mu(dv_j) \right] = (D_\omega)^{2k} \left(\mathcal{L}(\mu)\left(\frac{\omega}{N}\right)\right)^\ell. \end{aligned} \tag{4.34}$$

We now plug (4.34) in the right-hand side of (4.24). Then, we use that $1 \geq (D_\omega)^{2k}$ (as $D_\omega \leq 1$) to bound from below the term that is not in the sum. We obtain that for all $t \geq 0$

$$\begin{aligned} \int_t^{+\infty} n_\partial^{(2k)}(s) ds &\geq e^{-bNt} (D_\omega)^{2k} \left[1 + \sum_{\ell \geq 1} \frac{(bN)^\ell t^\ell}{\ell!} \left(\mathcal{L}(\mu)\left(\frac{\omega}{N}\right)\right)^\ell \right] \\ &= (D_\omega)^{2k} \exp\left[bN\left(\mathcal{L}(\mu)\left(\frac{\omega}{N}\right) - 1\right)t\right]. \end{aligned} \tag{4.35}$$

In addition, by first using the right-hand side of (2.7) to develop $\mathcal{L}(\mu)$, then the left-hand side of (2.7) to obtain that $\mathcal{L}(\mu^{(I)}) = (\mathcal{L}(g))^k$ for all $I \in \mathcal{I}_k$, and finally the first equality in Lemma A.6 to simplify the sum and the fraction, we have that for all $p \in \mathbb{C}$ such that $\text{Re}(p) > 0$

$$\mathcal{L}(\mu)(p) = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} \mathcal{L}(\mu^{(I)})(p) = \frac{1}{2^k} \sum_{I \in \mathcal{I}_k} (\mathcal{L}(g)(p))^k = (\mathcal{L}(g)(p))^k. \tag{4.36}$$

By plugging the above with $p = \frac{\omega}{N}$ in (4.35), in view of (2.9), we obtain that the left-hand side of (4.25) is proved.

Step 2. First, notice that by applying (H₄), then doing the change of variable $s' = s - \frac{bm_1 t}{2}$, and finally using that f_ω is non-decreasing, we have for all $t \geq 0$

$$\begin{aligned} \int_{\frac{bm_1}{2}t}^{+\infty} n_0(s) \, ds &\geq \frac{D_\omega \int_{\frac{bm_1}{2}t}^{+\infty} f_\omega(s) \exp(-\omega s) \, ds}{\int_0^{+\infty} f_\omega(z) \exp(-\omega z) \, dz} \\ &= \frac{D_\omega \int_0^{+\infty} f_\omega\left(s' + \frac{bm_1}{2}t\right) \exp(-\omega s' - \frac{bm_1 \omega}{2}t) \, ds'}{\int_0^{+\infty} f_\omega(z) \exp(-\omega z) \, dz} \geq D_\omega \exp\left(-\frac{bm_1 \omega}{2}t\right). \end{aligned}$$

Then, by combining the above equation with the first equality in (4.15), we obtain that the right-hand side of (4.25) is true, which ends the proof. □

4.4. Proof of Proposition 4.5 (a)

To recall, in view of (4.17), our aim is to check the assumptions of Lemma 3.5 for $u_\xi = \bar{u}^{(2k)}$, $\xi = \mu$ and $F = u^{(2k)}$, and then to apply it to obtain the statement. First, notice that by (4.13), we have for all $(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2$, $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$,

$$\partial_{x_\ell x_{\ell'}} u^{(2k)}(t, x) = \begin{cases} n'_0\left(\frac{bm_1}{2}t + x_\ell\right) n'_0\left(\frac{bm_1}{2}t + x_{\ell'}\right) \prod_{i=1, i \notin \{\ell, \ell'\}}^{2k} n_0\left(\frac{bm_1}{2}t + x_i\right), & \text{if } \ell \neq \ell', \\ n''_0\left(\frac{bm_1}{2}t + x_\ell\right) \prod_{i=1, i \neq \ell}^{2k} n_0\left(\frac{bm_1}{2}t + x_i\right), & \text{if } \ell = \ell'. \end{cases}$$

Then, by applying (H₂) and (H₃) to the above, we obtain that for all $(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2$ and $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$\left| \partial_{x_\ell x_{\ell'}} u^{(2k)}(t, x) \right| \leq \begin{cases} (C'_\lambda)^2 (\lambda D_\lambda)^{2k-2} \exp\left(-k b m_1 \lambda t - \lambda \sum_{i=1}^{2k} x_i\right), & \text{if } \ell \neq \ell', \\ C_\lambda (\lambda D_\lambda)^{2k-1} \exp\left(-k b m_1 \lambda t - \lambda \sum_{i=1}^{2k} x_i\right), & \text{if } \ell = \ell'. \end{cases}$$

This implies that (3.9) holds with $C = \max((C'_\lambda)^2, C_\lambda \lambda D_\lambda) (\lambda D_\lambda)^{2k-2}$, $\alpha = k b m_1 \lambda$, and $\beta = \lambda$. In addition, in view of the inequality $1 - e^{-x} \leq x$ for all $x \in \mathbb{R}$ and Lemma 4.1, we have that

$$\begin{aligned} bN \left[1 - \mathcal{L}(\mu) \left(\frac{\lambda}{N} \right) \right] &= bN \int_{u \in \mathbb{R}_+^{2k}} \left(1 - e^{-\frac{\lambda}{N} \sum_{i=1}^{2k} u_i} \right) \mu(du) \\ &\leq b\lambda \sum_{i=1}^{2k} \int_{u \in \mathbb{R}_+^{2k}} u_i \mu(du) = k b m_1 \lambda, \end{aligned} \tag{4.37}$$

so that (3.10) holds with the same α and β as before. From these two results, all the assumptions of Lemma 3.5 are verified. We thus apply this lemma, and it comes for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$

$$\begin{aligned} \left| \bar{u}^{(2k)}(t, x) \right| &\leq \max\left((C'_\lambda)^2, C_\lambda \lambda D_\lambda\right) (\lambda D_\lambda)^{2k-2} \frac{b t \sigma_\mu}{2N} \\ &\quad \times \exp\left[-bN \left(1 - \mathcal{L}(\mu) \left(\frac{\lambda}{N} \right) \right) t - \lambda \sum_{i=1}^{2k} x_i\right]. \end{aligned} \tag{4.38}$$

Now, by Lemma 4.7, we have that there exists $d'_0 > 0$, independent of b and k , such that

$$\max\left((C'_\lambda)^2, C_\lambda \lambda D_\lambda\right) \sigma_\mu \leq d'_0 (\lambda D_\lambda)^2 k^2. \tag{4.39}$$

Moreover, in view of (4.36) and (2.9), we have that

$$bN \left[1 - \mathcal{L}(\mu) \left(\frac{\lambda}{N} \right) \right] = bN \left[1 - \left(\mathcal{L}(g) \left(\frac{\lambda}{N} \right) \right)^k \right] = k b m_1 \lambda'_N. \tag{4.40}$$

Then, by plugging (4.39) and (4.40) in (4.38), we obtain that Proposition 4.5 (a) is true. □

4.5. Proof of Proposition 4.5 (b)

In view of (4.19), we consider the function $v_\mu := n_\partial^{(2k)} - u_{\partial,0}^{(2k)} + u_{\partial,1}^{(2k)} - u_\partial^{(2k)}$. Assume that there exists a set of functions $(h_i)_{i \in \llbracket 1, 2k \rrbracket}$ from $\mathbb{R}_+ \times \mathbb{R}_+^{2k}$ to \mathbb{R} verifying (3.13) with $C' = \frac{1}{2}C'_\lambda(\lambda D_\lambda)^{2k-1}$, $\xi = \mu$ and $\beta = \lambda$, such that for all $t \geq 0$

$$u_{\partial,1}^{(2k)}(t) - u_\partial^{(2k)}(t) = b \sum_{i=1}^{2k} \int_{y \in \mathbb{R}_+^{2k}} h_i(t, y)(1 - G(y_i))1_{\{y_i \leq \delta\}} dy. \tag{4.41}$$

By developing $n_\partial^{(2k)}$ and $u_{\partial,0}^{(2k)}$ with respectively (4.6) and (4.18), we have that for all $t \geq 0$

$$n_\partial^{(2k)}(t) - u_{\partial,0}^{(2k)}(t) = bN \int_{y \in \mathbb{R}_+^{2k}} \left[n^{(2k)}(t, y) - u^{(2k)}(t, y) \right] \mu(\{w \in \mathbb{R}_+^{2k} \mid Ny - w \notin \mathbb{R}_+^{2k}\}) dy. \tag{4.42}$$

Then, by plugging (4.41) and (4.42) in the definition of v_μ , we have that v_μ verifies (3.14) with $u_\xi = n^{(2k)} - u^{(2k)}$ and $\xi = \mu$. As u_ξ verifies the assumptions of Lemma 3.5 by the proof of Proposition 4.5 (a), see Section 4.4, this means that all the assumptions of Lemma 3.6 are verified for $v_\xi = v_\mu$. Then, by applying this lemma, and using (4.39) and (4.40) to simplify the bound (as done in Section 4.4), we obtain that for all $t \geq 0$

$$|v_\mu(t)| \leq \frac{b}{N} (D_\lambda)^{2k-1} \left(d'_0 \lambda D_\lambda \frac{k^3 b m_1 t}{2} + \frac{C'_\lambda m_2 k}{2} \right) \exp(-k b m_1 \lambda'_N t).$$

Recalling equation (4.19), combining the above with (4.20) through a triangle inequality, and then using that $-k b m_1 \lambda \leq -k b m_1 \lambda'_N$ (as a consequence of (4.37) and (4.40)), yield that (4.16) is true. Then, Proposition 4.5 (b) is proved, assuming that the set of functions $(h_i)_{i \in \llbracket 1, 2k \rrbracket}$ presented at the beginning of the proof exists.

It thus remains to prove that such a sequence exists. To do so, we consider for all $i \in \llbracket 1, 2k \rrbracket$, $(t, y) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$,

$$h_i(t, y) := \frac{1}{2} \left[u^{(2k)} \left(t, \sum_{j=1, j \neq i}^{2k} y_j e_j + \frac{y_i}{N} \right) - u^{(2k)} \left(t, \sum_{j=1, j \neq i}^{2k} y_j e_j \right) \right]. \tag{4.43}$$

By taking the difference between the second lines of equations (4.18) and (4.12), and using that $m_1 = \int_{y_i \in [0, \delta]} (1 - G(y_i)) dy_i$ for each index $i \in \llbracket 1, 2k \rrbracket$ of the sum in the second line of (4.12), we have that (4.41) holds with the set of functions defined in (4.43). In addition, by using (4.13), then writing n_0 as an integral of n'_0 , and finally applying (H2) and (H3), we have that for all $i \in \llbracket 1, 2k \rrbracket$, $(t, y) \in \mathbb{R}_+ \times \mathbb{R}_+^{2k}$,

$$\begin{aligned} h_i(t, y) &= \frac{1}{2} \left[n_0 \left(\frac{b m_1}{2} t + \frac{y_i}{N} \right) - n_0 \left(\frac{b m_1}{2} t \right) \right] \left[\prod_{j=1, j \neq i}^{2k} n_0 \left(\frac{b m_1}{2} t + y_j \right) \right] \\ &= \frac{1}{2} \left[\int_0^{\frac{y_i}{N}} n'_0 \left(\frac{b m_1}{2} t + u \right) du \right] \left[\prod_{j=1, j \neq i}^{2k} n_0 \left(\frac{b m_1}{2} t + y_j \right) \right] \\ &\leq \frac{1}{2} C'_\lambda (\lambda D_\lambda)^{2k-1} \frac{y_i}{N} \exp \left[-k b m_1 \lambda t - \lambda \sum_{j=1, j \neq i}^{2k} y_j \right]. \end{aligned}$$

This implies, by using (4.37) to bound the coefficient $-k b m_1 \lambda$ in the exponential, that $(h_i)_{i \in \llbracket 1, 2k \rrbracket}$ verifies (3.13) with $C' = \frac{1}{2}C'_\lambda(\lambda D_\lambda)^{2k-1}$, $\xi = \mu$ and $\beta = \lambda$. Then, from these points, we have that the set of functions we need exists, which concludes the proof of Proposition 4.5 (b). □

4.6. Proof of Theorem 2.7 (b)

Let $x \geq 0$. To simplify notations, we denote

$$\widehat{N}_0(x) := \int_{\frac{2x}{bm_1}}^{+\infty} n_{\partial}^{(2k)}(s) ds, \quad \text{and} \quad \widehat{U}_0(x) := \int_{\frac{2x}{bm_1}}^{+\infty} u_{\partial}^{(2k)}(s) ds.$$

In view of (4.22) and the triangle inequality, the following holds

$$\begin{aligned} \left| \widehat{n}_0^{(2k)}(x) - n_0(x) \right| &\leq \frac{\left| n_{\partial}^{(2k)}\left(\frac{2x}{bm_1}\right) - u_{\partial}^{(2k)}\left(\frac{2x}{bm_1}\right) \right|}{kbm_1 \widehat{N}_0(x)^{1-\frac{1}{2k}}} + \frac{u_{\partial}^{(2k)}\left(\frac{2x}{bm_1}\right)}{kbm_1} \left| \frac{1}{\widehat{N}_0(x)^{1-\frac{1}{2k}}} - \frac{1}{\widehat{U}_0(x)^{1-\frac{1}{2k}}} \right| \\ &=: \Delta_1(x) + \Delta_2(x). \end{aligned} \tag{4.44}$$

Thus, our aim is to obtain an upper bound for both $\Delta_1(x)$ and $\Delta_2(x)$. Theorem 2.7 (b) then comes by summing these bounds.

By applying Proposition 4.5 (b) to bound the numerator, and Lemma 4.11 to bound the denominator, we have the following

$$\begin{aligned} \Delta_1(x) &\leq \frac{bd'_1(D_\lambda)^{2k} (k^3x + k^2 + k) \exp(-2k\lambda'_N x)}{kbm_1 \left[(D_\omega)^{2k} \exp(-2k\omega'_N x) \right]^{1-\frac{1}{2k}}} \\ &= \frac{d'_1 D_\omega}{Nm_1} \left(\frac{D_\lambda}{D_\omega} \right)^{2k} (k^2x + k + 1) \exp(-2k\lambda'_N x + (2k - 1)\omega'_N x). \end{aligned} \tag{4.45}$$

We thus now focus on finding an upper bound for $\Delta_2(x)$, which requires more computations. For this purpose, we begin by obtaining an intermediate inequality. In the equation below, first apply the equality $|\frac{1}{c} - \frac{1}{d}| = \frac{|c-d|}{cd}$ for $c = \widehat{N}_0(x)^{1-\frac{1}{2k}}$ and $d = \widehat{U}_0(x)^{1-\frac{1}{2k}}$ to develop $\Delta_2(x)$. Then, use that it holds $\frac{1}{kbm_1} u_{\partial}^{(2k)}\left(\frac{2x}{bm_1}\right) = n_0(x) \widehat{U}_0(x)^{1-\frac{1}{2k}}$ by Lemma 4.9 to simplify the term $\widehat{U}_0(x)^{1-\frac{1}{2k}}$ in the denominator. Finally, apply (H3) to bound from above the term $n_0(x)$ coming from the previous computation, and use Lemma 4.11 to bound the term $\widehat{N}_0(x)^{1-\frac{1}{2k}}$ in the denominator. It comes the following inequality

$$\Delta_2(x) = n_0(x) \frac{\left| \widehat{N}_0(x)^{1-\frac{1}{2k}} - \widehat{U}_0(x)^{1-\frac{1}{2k}} \right|}{\widehat{N}_0(x)^{1-\frac{1}{2k}}} \leq \lambda D_\lambda \exp(-\lambda x) \frac{\left| \widehat{N}_0(x)^{1-\frac{1}{2k}} - \widehat{U}_0(x)^{1-\frac{1}{2k}} \right|}{(D_\omega)^{2k-1} \exp[-(2k-1)\omega'_N x]}. \tag{4.46}$$

To continue our computations, we need to bound the numerator of (4.46). To do so, we first develop it by using the inequality $|c^\ell - d^\ell| \leq \ell \frac{|c-d|}{\min(c,d)^{1-\ell}}$ for $c = \widehat{N}_0(x)$, $d = \widehat{U}_0(x)$ and $\ell = 1 - \frac{1}{2k}$. This inequality comes from the Taylor's inequality applied to the function $y \mapsto y^\ell$, and is true when $\ell \in (0, 1)$. Thereafter, we apply Lemma 4.11 to bound the term $\frac{1}{\min(c,d)^{1-\ell}}$ coming from the previous step, in view of the fact that $\omega'_N \leq \omega$ by (4.37) and (4.40), that also holds when λ_N and λ'_N are replaced with ω and ω'_N respectively. Finally, to bound the term $|c - d|$ coming from the first step, we use the following equality, which comes from (2.6)–(4.15), Proposition 4.5 (a), and the fact that $\int_{y \in \mathbb{R}_+^{2k}} \exp(-\lambda \sum_{i=1}^{2k} y_i) dy = \frac{1}{\lambda^{2k}}$,

$$\left| \widehat{N}_0(x) - \widehat{U}_0(x) \right| = \left| \int_{y \in \mathbb{R}_+^{2k}} \left(n^{(2k)} - u^{(2k)} \right) \left(\frac{2x}{bm_1}, y \right) dy \right| \leq d'_0(D_\lambda)^{2k} \frac{k^2 x}{Nm_1} \exp(-2k\lambda'_N x).$$

We obtain

$$\left| \widehat{N}_0(x)^{1-\frac{1}{2k}} - \widehat{U}_0(x)^{1-\frac{1}{2k}} \right| \leq \left(1 - \frac{1}{2k} \right) \frac{d'_0(D_\lambda)^{2k} \frac{k^2 x}{Nm_1} \exp[-2k\lambda'_N x]}{D_\omega \exp(-\omega x)}. \tag{4.47}$$

Then, by plugging (4.47) in (4.46) and bounding the term $1 - \frac{1}{2k}$ by 1, we get the following upper bound for $\Delta_2(x)$

$$\begin{aligned} \Delta_2(x) &\leq \lambda D_\lambda \exp(-\lambda x) \frac{d'_0(D_\lambda)^{2k} \frac{k^2 x}{Nm_1} \exp(-2k\lambda'_N x)}{(D_\omega)^{2k} \exp[-(2k-1)\omega'_N x - \omega x]} \\ &= \frac{d'_0 \lambda D_\lambda}{Nm_1} \left(\frac{D_\lambda}{D_\omega}\right)^{2k} k^2 x \exp[-(\lambda + 2k\lambda'_N)x + (\omega + (2k-1)\omega'_N)x]. \end{aligned} \tag{4.48}$$

We now conclude. As it holds $\omega \geq \lambda$, we have that

$$-2k\lambda'_N + (2k-1)\omega'_N \leq -(\lambda + 2k\lambda'_N) + (\omega + (2k-1)\omega'_N). \tag{4.49}$$

Therefore, by plugging (4.45) and (4.48) in (4.44), and then using (4.49) to bound the term coming from (4.45), we obtain Theorem 2.7 (b). \square

5. ESTIMATION ON SIMULATIONS AND EXPERIMENTAL DATA

Now that we have studied the quality of our estimators from a theoretical point of view, we verify if they work in practice. First, in Section 5.1, we present estimation results in the case where we observe $n_\partial^{(1)}$ and $n_\partial^{(2k)}$ without noise. Then, in Section 5.2, we study how our inference method can be adapted in a more realistic framework where we observe noisy values of $n_\partial^{(1)}$ and $n_\partial^{(2k)}$, focusing in particular on the noise related to sampling. Finally, in Section 5.3, we test our inference method on experimental data.

5.1. Estimation results on noise-free data

We study here the case where we observe the exact values of $\widehat{n}_0^{(1)}$ and $\widehat{n}_0^{(2k)}$. We first present in Section 5.1.1 two examples in which the estimation is satisfactory in the single-telomere model. Then, in Section 5.1.2, we present the issues that arise in this model when n_0 has a small variability. Finally, in Section 5.1.3, we present estimation results for the model in \mathbb{R}_+^{2k} .

5.1.1. Estimation results in the single-telomere model: large variability

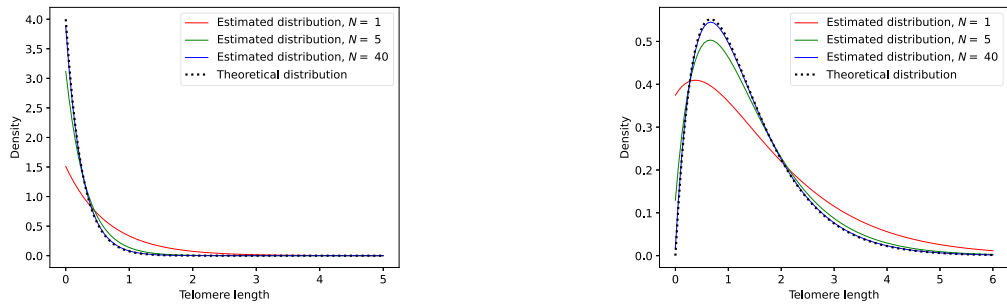
Let us start by presenting the framework we use here. We introduce for all $(\ell, \beta) \in \mathbb{N}^* \times \mathbb{R}_+^*$ the following functions, for all $x \geq 0$,

$$h_{\ell,\beta}(x) := \frac{\beta^\ell}{(\ell-1)!} x^{\ell-1} \exp(-\beta x), \quad \text{and} \quad H_{\ell,\beta}(x) := \int_0^x \frac{\beta^\ell}{(\ell-1)!} y^{\ell-1} \exp(-\beta y) dy. \tag{5.1}$$

These functions correspond respectively to the probability density and the cumulative density function of an Erlang distribution with parameter (ℓ, β) . In the estimations presented here, we choose n_0 belonging to the set $\{h_{\ell,\beta} \mid (\ell, \beta) \in \mathbb{N}^* \times \mathbb{R}_+^*\}$. The first reason is that in this case, either $\widehat{n}_0^{(1)}$ or $n^{(1)}$ can be computed thanks to an explicit formula, see Propositions A.8 and A.9. The second reason is that our main result assumptions are verified for these functions. Finally, the last reason is that when $\ell \geq 2$, these distributions are biologically relevant since they correspond to unimodal distributions, see [40].

We assume that (H_1) is verified for $b = 1$, $g = 1_{[0,1]}$, and N not yet fixed. Our aim is to check numerically that the curve of $\widehat{n}_0^{(1)}$ is close to the one of n_0 . To do so, we plot in Figures 1a and 1b the curve of $\widehat{n}_0^{(1)}$ as a function of telomere lengths when $n_0 \in \{h_{1,4}, h_{2,1}\}$, and for $N \in \{1, 5, 40\}$. In each of these figures, the curve of n_0 is also plotted in black for comparison. We observe that when N is large, the estimation performs very well. Indeed, in Figures 1a and 1b, the blue curves, which correspond to the estimations with $N = 40$ are almost superposed with the dotted curves. The estimation results are thus very satisfactory.

Remark 5.1. We have chosen $N = 40$ for the maximum scaling parameter because this is the most realistic value for budding yeast, see the discussion about (H_1) in Section 2.3.



(a) Estimation results when $n_0 = h_{1,4}$ and $N \in \{1, 5, 40\}$.

(b) Estimation results when $n_0 = h_{2,1.5}$ and $N \in \{1, 5, 40\}$.

FIGURE 1. Estimation results in the single-telomere model for different values of N , when $b = 1$, $g = 1_{[0,1]}$ and $n_0 \in \{h_{1,4}, h_{2,1.5}\}$.

5.1.2. Estimation results in the single-telomere model: small variability

As mentioned in the previous section, it is required to have an initial distribution with a sufficiently large coefficient of variation to ensure a good estimation. We present here the problems that occur when this is not the case and the reasons behind it. We assume that (H_1) is verified with $b = 1$, $g = 1_{[0,1]}$, and $N = 40$. We proceed to estimations of initial distributions with the same mean, but different coefficients of variation (cv). To do so, in view of equation (5.1) and Proposition A.10, we first introduce for all $cv > 0$ the function $\mathcal{H}_{cv} = h_{1/cv^2, 1/cv^2}$, which is the density of an Erlang distribution with mean $m = 1$ and coefficient of variation cv . Then, we plot in Figure 2 the curve of $\hat{n}_0^{(1)}$ (blue curves), that we compare with the curve of n_0 (dotted curves), for $n_0 \in \{\mathcal{H}_{1/2}, \mathcal{H}_{1/3}, \mathcal{H}_{1/5}, \mathcal{H}_{1/7}\}$. We observe that the smaller the coefficient of variation is, the worse the estimation becomes. In particular, the spread of the initial laws is not well-captured, although the position of the mode is correctly estimated.

A qualitative reason for this observation is that when the initial distribution has a small coefficient of variation, the randomness of shortening values and cell division times plays a greater role than the variability of the initial distribution. This leads to a poor estimation, as the information on n_0 obtained from senescence times is blurred by the information on g and b . A quantitative reason is that the second derivative of the initial distribution is large when its coefficient of variation is small. Thus, as the error between $\hat{n}_0^{(1)}$ and n_0 is mainly given by the second derivative of n_0 (see Sects. 3.3 and 3.5), the estimation is not reliable. To illustrate this, by using the formula in the left-hand side of (5.1), we compute the L^2 -norm of the second derivatives of $\mathcal{H}_{1/2}$, $\mathcal{H}_{1/3}$, $\mathcal{H}_{1/5}$, and $\mathcal{H}_{1/7}$. We obtain that $\|\mathcal{H}_{1/2}''\|_{L^2(\mathbb{R}_+)} \approx 5.657$, $\|\mathcal{H}_{1/3}''\|_{L^2(\mathbb{R}_+)} \approx 9.445$, $\|\mathcal{H}_{1/5}''\|_{L^2(\mathbb{R}_+)} \approx 28.17$ and $\|\mathcal{H}_{1/7}''\|_{L^2(\mathbb{R}_+)} \approx 62.41$, so that the values of the second derivative of n_0 increase when the coefficient of variation decreases.

5.1.3. Estimation results in the model with several telomeres

We now study how the inference method works on the model with several telomeres. To do so, we proceed in the same way as in Section 5.1. First, we assume that (H_1) is verified with $b = 1$, $g = 1_{[0,1]}$, and $N = 40$. Then, we plot in Figures 3a and 3b the curve of the estimator $n_0^{(2k)}$ for different values of k , by using the formula given in Proposition A.8. In each case, this curve is compared with the one of n_0 , plotted in dotted lines. The estimations presented in Figure 3a correspond to estimation when $k \in \{1, 3, 5\}$, so to estimations in which k is small. Conversely, the estimations presented in Figure 3b correspond to estimations when $k \in \{15, 30, 50\}$, so to estimations in which k is large. This case is important to consider because biologists mostly study species with a large number of telomeres: budding yeast cells have 32 telomeres ($k = 16$), and human cells have 92 telomeres ($k = 46$). We observe in Figure 3a that the estimated initial distributions almost overlap with the theoretical

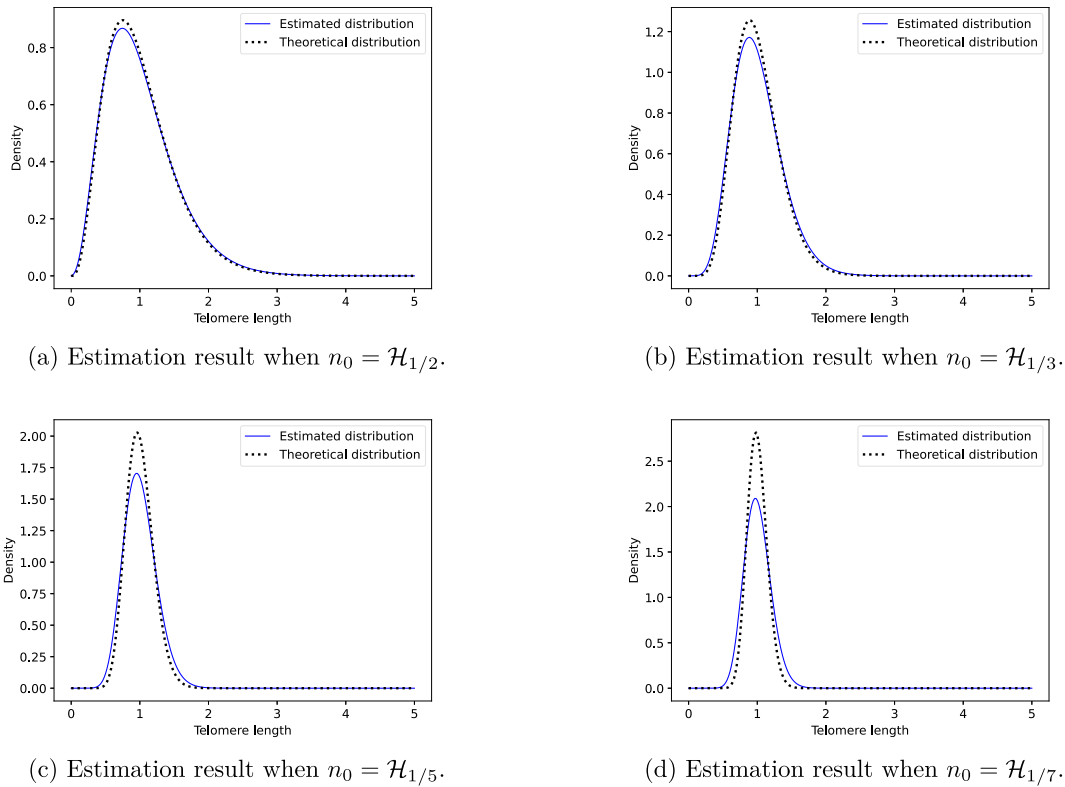


FIGURE 2. Estimation results in the single-telomere model when $b = 1$, $g = 1_{[0,1]}$, $N = 40$, and $n_0 \in \{\mathcal{H}_{1/2}, \mathcal{H}_{1/3}, \mathcal{H}_{1/5}, \mathcal{H}_{1/7}\}$.

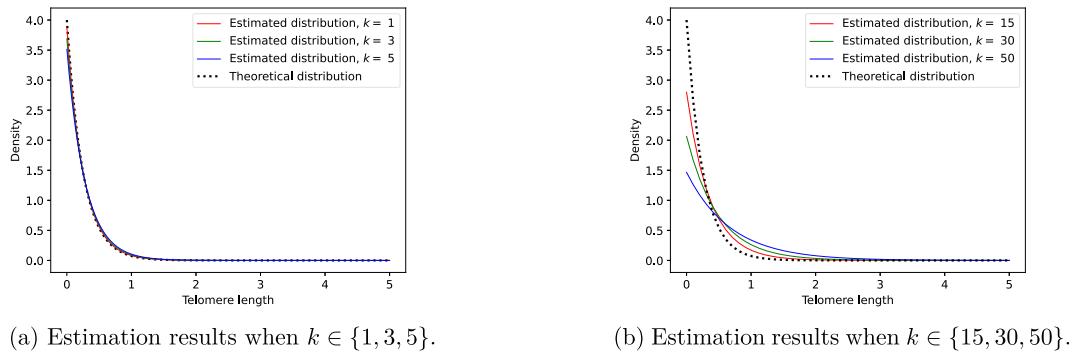


FIGURE 3. Estimation results in the model with several telomeres when $b = 1$, $g = 1_{[0,1]}$, $N = 40$, $n_0 = h_{1,4}$, and $k \in \{1, 3, 5, 15, 30, 50\}$.

initial distribution. We also observe in Figure 3b that the estimated curves are far from the curve of n_0 (dotted curve). We thus have that the estimation is satisfactory when k is small, but poor when k is large.

A quantitative reason for these observations is that the model approximation used to construct our estimator is no longer valid when k is large. Indeed, the bounds obtained in Proposition 4.5 and Theorem 2.7 increase with the value of k . It follows that using $\hat{n}_0^{(2k)}$ to estimate n_0 does not really make sense, because we now

have no control over its estimation error. A qualitative reason is that the senescence times distribution becomes more and more determined by the distribution of cell division times when $k \rightarrow +\infty$. The estimation is thus less accurate as the information on n_0 is blurred by the influence of the division times distribution. The increasing influence of the division times distribution is related to the fact that the probability of having at least one short telomere initially grows with k . As a result, the initial minimum length tends to concentrate near zero, and fewer shortenings are required before having a telomere with length below 0. This yields that the randomness of the initial distribution and the shortening values lose their impact on the senescence times distribution, and that only the randomness of the division times remains to influence the latter.

5.2. Estimation with random variables

In practice, we never have the full density $n_\partial^{(1)}$ or $n_\partial^{(2k)}$. Instead, we observe senescence times $(T_i)_{1 \leq i \leq m}$, where $m \in \mathbb{N}^*$, that are noisy measurements of random variables distributed according to $n_\partial^{(1)}$ or $n_\partial^{(2k)}$. In this subsection, we investigate whether our inference method can be adapted to this setting, by considering only the noise related to sampling, and not to measurement error. We first focus in Section 5.2.1 on the case where we observe senescence times distributed according to $n_\partial^{(1)}$. We then do the same in the case where senescence times are distributed according to $n_\partial^{(2k)}$ in two different parts, as it requires more work. More precisely, we present the estimator we use in Section 5.2.2, and check its quality on simulations in Section 5.2.3. We show in this last case that issues related to the curse of the dimensionality arise. From now on, we assume that (H_1) is verified with $b = 1$, $g = 1_{[0,1]}$, and $N = 40$.

5.2.1. Estimator and estimation with random variables distributed according to $n_\partial^{(1)}$

Let us consider $n_s \in \mathbb{N}^*$, and $(T_{1,i})_{1 \leq i \leq n_s}$ a sequence of random variables independent and identically distributed according to $n_\partial^{(1)}$. Our first objective in this section is to construct an estimator of n_0 based on these random variables. In fact, this can be easily done. What we only have to do is to first construct an estimator of $n_\partial^{(1)}$, and then to adapt the expression of $\hat{n}_0^{(1)}$ given in (2.10) to this estimator. Let us detail these two steps.

To obtain an estimator of $n_\partial^{(1)}$ with the times $(T_{1,i})_{1 \leq i \leq n_s}$, we do a log-transform kernel density estimation [9, 26]. Denoting for all $x \geq 0$ the Gaussian kernel $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, and $\alpha > 0$ a smoothing parameter, the log-transform kernel density estimator is defined for all $t \geq 0$ as

$$\bar{n}_\partial^{(1,\alpha)}(t) := \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{1}{t\alpha} \rho\left(\frac{1}{\alpha} \log\left(\frac{t}{T_{1,i}}\right)\right). \tag{5.2}$$

This is a classical method to estimate a density on \mathbb{R}_+ . The main idea behind this estimator is to smooth the Dirac measures in the empirical estimator of the density, defined as $\frac{1}{n_s} \sum_{i=1}^{n_s} \delta_{T_{1,i}}$. This smoothing is due to that for all $f \in \mathcal{C}_c^\infty(\mathbb{R}_+^*)$ and $X \in \mathbb{R}_+^*$, by the change of variable $t' = \log(\frac{t}{X})$ and the fact that $\frac{1}{\alpha} \rho(\frac{\cdot}{\alpha}) \xrightarrow{\alpha \rightarrow 0} \delta_0$ (as it is a classical mollifier), we have

$$\lim_{\alpha \rightarrow 0} \int_{t \in \mathbb{R}_+^*} \frac{1}{t\alpha} \rho\left(\frac{1}{\alpha} \log\left(\frac{t}{X}\right)\right) f(t) dt = \lim_{\alpha \rightarrow 0} \int_{t' \in \mathbb{R}} \frac{1}{\alpha} \rho\left(\frac{t'}{\alpha}\right) f(X \exp(t')) dt' = \delta_X(f). \tag{5.3}$$

Thus, by the above and the fact that $\rho \in \mathcal{C}^\infty(\mathbb{R})$, the terms that are summed in (5.2) are smoothed approximations of the measures $(\delta_{T_{1,i}})_{i \in \llbracket 1, n_s \rrbracket}$. In particular, when α is small, $\bar{n}_\partial^{(1,\alpha)}$ is close to the empirical estimator n_∂ but the smoothing is weak. When α is large, it is the opposite. We do not use a classical kernel density estimation [38] because this method works for densities with support on \mathbb{R} , and is much less satisfactory on \mathbb{R}_+ , see Section 1 in [26].

Adapting now the definition of \widehat{n}_0 , see (2.10), by replacing $n_{\partial}^{(1)}$ with $\overline{n}_{\partial}^{(1,\alpha)}$, we obtain the following estimator for n_0 that depends on the smoothing parameter $\alpha > 0$, for all $x \geq 0$,

$$\overline{n}_0^{(1,\alpha)}(x) := \frac{1}{\widehat{b}\widehat{m}_1} \overline{n}_{\partial}^{(1,\alpha)}\left(\frac{x}{\widehat{b}\widehat{m}_1}\right) = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{1}{x\alpha} \rho\left(\frac{1}{\alpha} \log\left(\frac{x}{\widehat{b}\widehat{m}_1 T_{1,i}}\right)\right). \tag{5.4}$$

This estimator is what we use to estimate n_0 from the variables $(T_{1,i})_{1 \leq i \leq n_s}$. A theoretical result concerning its quality, whose proof is sketched in Appendix B, is provided below. It states that the error of this estimator tends to 0 with probability 1 when both $n_s \rightarrow +\infty$ and $N \rightarrow +\infty$.

Proposition 5.2 (Confidence intervals, single-telomere model). *Assume that (H1)–(H2) hold. We denote the constant*

$$C_{\widehat{n},1} := \left\| Id^2 \cdot \left(\widehat{n}_0^{(1)}\right)' + Id \cdot \widehat{n}_0^{(1)} \right\|_{L^\infty(\mathbb{R}_+)}. \tag{5.5}$$

We also fix $p \in (0, 1]$, and the smoothing parameter $\alpha_p := \frac{1}{(C_{\widehat{n},1})^{\frac{1}{2}}} \left(\frac{\log(\frac{2}{p})}{2n_s}\right)^{\frac{1}{4}}$. Then, there exists a sequence $(L_{1,n})_{n \in \mathbb{N}}$ of positive real numbers independent of p such that $\limsup_{n \rightarrow +\infty} L_{1,n} < +\infty$, and such that with probability at least $1 - p$ it holds

$$\left\| Id \left[\overline{n}_0^{(1,\alpha_p)} - n_0 \right] \right\|_{L^\infty(\mathbb{R}_+)} \leq 2\sqrt{\frac{2}{\pi}} (C_{\widehat{n},1})^{\frac{1}{2}} \left(\frac{\log(\frac{2}{p})}{2n_s}\right)^{\frac{1}{4}} + \frac{L_{1,N}}{N}.$$

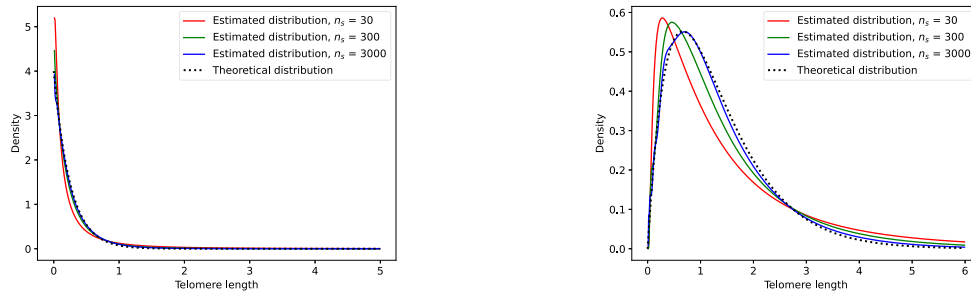
Remark 5.3. We observe that we need to weight the L^∞ -norm by the identity function to obtain a qualitative bound on the error. This weighting implies that the guarantees we have on the estimation of $n_0(x)$, where $x \geq 0$, decrease when x is small. This is due to the fact the logarithm explodes near 0, which generates instability.

We now proceed to estimations on simulations. We first choose $n_s \in \{30, 300, 3000\}$, and simulate a sequence of random variables $(T_{1,i})_{1 \leq i \leq n_s}$ distributed according to $n_{\partial}^{(1)}$, when $n_0 \in \{h_{1,4}, h_{2,1.5}\}$. These random variables are obtained by simulating a probabilistic telomere shortening model, see [5, 28]. Then, in Figure 4, we compare the curve of $\overline{n}_0^{(1,\alpha)}$ with the curve of n_0 , for $\alpha = \alpha_{0.1}$ defined in Proposition 5.2. We observe that when $n_s = 3000$, the curve of $\overline{n}_0^{(1,\alpha_{0.1})}$ follows perfectly the one of n_0 , so the estimation is very satisfactory. We also observe that when $n_s = 300$, the quality of the estimation decreases, but remains correct. When $n_s = 30$, the curve of $\overline{n}_0^{(1,\alpha_{0.1})}$ does not follow the one of n_0 , so the estimation is not completely satisfactory. However, this is expected because we have a very small number of data points in this case. We therefore have from these observations that the estimation stays reliable when we use $\overline{n}_0^{(1,\alpha_{0.1})}$ to estimate instead of $\widehat{n}_0^{(1)}$, provided that n_s is not too small.

Remark 5.4. Choosing for smoothing parameter $\alpha_{0.1}$ allows us to have an estimation with a confidence level of 0.9 by Proposition 5.2. However, in practice, choosing for smoothing parameter one of the constants in $(\alpha_p)_{p \in (0,1]}$ is impossible in most cases, as it requires to have *a priori* information on the initial distribution (in particular, on $C_{\widehat{n},1}$). Developing a method for choosing a good smoothing parameter without *a priori* information on n_0 is quite challenging. We thus postpone this study to another work.

5.2.2. Estimator with random variables distributed according to $n_{\partial}^{(2k)}$

We now focus on the case in which we observe a sequence of random variables $(T_{2k,i})_{1 \leq i \leq n_s}$ identically distributed according to $n_{\partial}^{(2k)}$, where $n_s \in \mathbb{N}^*$. In particular, in this section, we construct the estimator of n_0 we use in this context. We assume that $T_{2k,i} \leq T_{2k,i+1}$ for all $i \in \llbracket 1, n_s - 1 \rrbracket$, and that the random variables



(a) Comparison between the curves of $\widehat{n}_0^{(1, \alpha_{0.1})}$ and $n_0 = h_{1,4}$, for $n_s \in \{30, 300, 3000\}$.

(b) Comparison between the curves of $\widehat{n}_0^{(1, \alpha_{0.1})}$ and $n_0 = h_{2,1.5}$, for $n_s \in \{30, 300, 3000\}$.

FIGURE 4. Estimation results for the estimator $\widehat{n}_0^{(1, \alpha_{0.1})}$ defined in (5.4), when $b = 1, g = 1_{[0,1]}$, $N = 40, n_0 \in \{h_{1,4}, h_{2,1.5}\}$, and for different values of n_s .

$(T_{2k,i})_{1 \leq i \leq n_s}$ are independent up to a permutation. We consider the survival function associated to n_∂ , defined for all $t \geq 0$ as $N_\partial(t) := \int_t^{+\infty} n_\partial^{(2k)}(s) ds$, and its empirical estimator

$$\forall t \geq 0 : \widehat{N}_\partial(t) := \frac{1}{n_s} \sum_{i=1}^{n_s} 1_{\{T_{2k,i} > t\}} = 1 - \frac{1}{n_s} \sum_{i=1}^{n_s} 1_{\{T_{2k,i} \leq t\}}. \tag{5.6}$$

The strategy we follow here for constructing our estimator of n_0 is slightly different from the one followed in the previous section. It is based on the fact that $\widehat{n}_0^{(2k)}$ is the weak derivative of the function $x \mapsto -[N_\partial(\frac{2x}{b\widehat{m}_1})]^{\frac{1}{2k}}$, see (2.10). From this last property and the fact that N_∂ can be estimated by \widehat{N}_∂ , a natural estimator for n_0 would be to use the weak derivative of $M(x) := -[\widehat{N}_\partial(\frac{2x}{b\widehat{m}_1})]^{\frac{1}{2k}}$. The problem is that the latter cannot be represented by a function, as we can see in the following statement.

Proposition 5.5 (Derivative of M). *We denote $T_{2k,0} = 0$ and $T_{2k,n_s+1} = +\infty$. Then, it holds*

$$M' = \sum_{j=0}^{n_s} \left(1 - \frac{j}{n_s}\right)^{\frac{1}{2k}} \left[\delta_{\frac{b\widehat{m}_1}{2} T_{2k,j+1}} - \delta_{\frac{b\widehat{m}_1}{2} T_{2k,j}} \right].$$

Proof. As the times $(T_{2k,i})_{0 \leq i \leq n_s+1}$ are ordered in increasing order, we have for all $t \geq 0$

$$\begin{aligned} \sum_{i=1}^{n_s} 1_{\{T_{2k,i} \leq t\}} &= \sum_{i=1}^{n_s} \sum_{j=i}^{n_s} 1_{\{T_{2k,j} \leq t < T_{2k,j+1}\}} \\ &= \sum_{j=1}^{n_s} \sum_{i=1}^j 1_{\{T_{2k,j} \leq t < T_{2k,j+1}\}} = \sum_{j=0}^{n_s} j \cdot 1_{\{T_{2k,j} \leq t < T_{2k,j+1}\}}. \end{aligned} \tag{5.7}$$

Therefore, by first replacing M with its definition, and then combining (5.7) with (5.6), in view of the fact that the intervals in the indicators in (5.7) are disjoint, we obtain the following equality, which ends the proof, for

all $f \in C_c^\infty(\mathbb{R}_+^*)$,

$$\begin{aligned} - \int_{x \in \mathbb{R}_+^*} M(x) f'(x) dx &= \sum_{j=0}^{n_s} \left(1 - \frac{j}{n_s}\right)^{\frac{1}{2k}} \int_{x \in \mathbb{R}_+^*} 1_{\{T_{2k,j} \leq \frac{2x}{\tilde{b}\tilde{m}_1} < T_{2k,j+1}\}} f'(x) dx \\ &= \sum_{j=0}^{n_s} \left(1 - \frac{j}{n_s}\right)^{\frac{1}{2k}} \left[f\left(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,j+1}\right) - f\left(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,j}\right) \right]. \end{aligned}$$

□

To solve the issue presented above, we use a smoothed version of M' to estimate n_0 . This version is obtained by using that in view of (5.3), we can approximate $\delta_{\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,j}}$ for all $j \in \llbracket 0, n_s + 1 \rrbracket$ by the measure $\frac{1}{\alpha x} \rho\left(\frac{1}{\alpha} \log\left(\frac{2x}{\tilde{b}\tilde{m}_1 T_{2k,j}}\right)\right) dx$, where $\alpha > 0$ is a smoothing parameter. This gives us the following estimator, depending on the parameter α , for all $x \geq 0$,

$$\bar{n}_0^{(2k,\alpha)}(x) := \sum_{j=0}^{n_s} \left(1 - \frac{j}{n_s}\right)^{\frac{1}{2k}} \frac{1}{\alpha x} \left[\rho\left(\frac{1}{\alpha} \log\left(\frac{2x}{\tilde{b}\tilde{m}_1 T_{2k,j+1}}\right)\right) - \rho\left(\frac{1}{\alpha} \log\left(\frac{2x}{\tilde{b}\tilde{m}_1 T_{2k,j}}\right)\right) \right]. \tag{5.8}$$

The following proposition provides us a bound on the error of this estimator. It is proved in Appendix B, and further discussed in Section 5.2.3.

Proposition 5.6 (Confidence intervals, model with several telomeres). *Assume that (H₁)–(H₄) hold. We denote the constant*

$$C_{\hat{n},2k} := \left\| Id^2 \cdot \left(\hat{n}_0^{(2k)}\right)' + Id \cdot \hat{n}_0^{(2k)} \right\|_{L^\infty(\mathbb{R}_+)}. \tag{5.9}$$

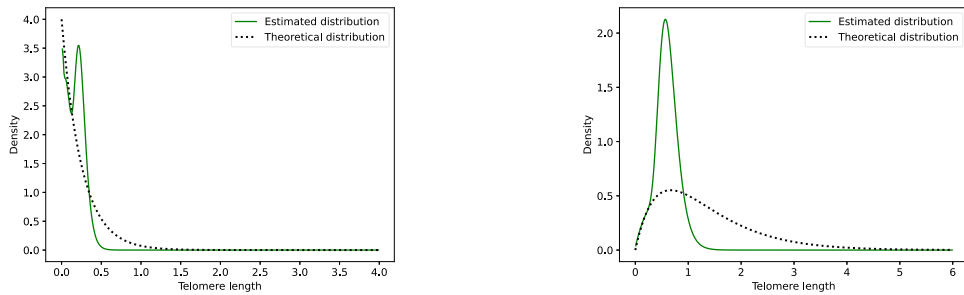
We also fix $p \in (0, 1]$, and the smoothing parameter $\tilde{\alpha}_p := \frac{1}{(C_{\hat{n},2k})^{\frac{1}{2}}} \left(\frac{\log\left(\frac{2}{p}\right)}{2n_s}\right)^{\frac{1}{8k}}$. Then, there exists a sequence $(L_{2k,n})_{n \in \mathbb{N}^-}$ of positive real numbers independent of p , such that $\limsup_{n \rightarrow +\infty} L_{2k,n} < +\infty$, and such that with probability at least $1 - p$ it holds

$$\left\| Id \left[\bar{n}_0^{(2k,\tilde{\alpha}_p)} - n_0 \right] \right\|_{L^\infty(\mathbb{R}_+)} \leq 2\sqrt{\frac{2}{\pi}} (C_{\hat{n},2k})^{\frac{1}{2}} \left(\frac{\log\left(\frac{2}{p}\right)}{2n_s}\right)^{\frac{1}{8k}} + \frac{L_{2k,N}}{N}. \tag{5.10}$$

At first sight, other estimators than the one given in (5.8) seem more natural to use. In view of the definition of $\hat{n}_0^{(2k)}$, see (2.10), we can for example mention estimators corresponding to the ratio between an estimator of the numerator of $\hat{n}_0^{(2k)}$ and an estimator of the denominator of $\hat{n}_0^{(2k)}$. We prefer here to use the estimator presented in (5.8) for two reasons. The first one is that it is easier to obtain qualitative results about the quality of the estimators $(\bar{n}_0^{(2k,\alpha)})_{\alpha>0}$, see Proposition 5.6. The second one is that this prevents us from having an estimator with a denominator tending to 0 when $x \rightarrow +\infty$, which can generate instability.

5.2.3. Estimation with random variables distributed according to $n_\partial^{(2k)}$

We consider two examples: first, the case where the model with several telomeres, defined in (2.5), has for parameters $(n_0, k) = (h_{1,4}, 5)$. Second, the case where this model has for parameters $(n_0, k) = (h_{2,1.5}, 16)$. For each of these examples, we fix $n_s = 3000$ and simulate a sequence of random variables $(T_{2k,i})_{1 \leq i \leq n_s}$ distributed according to $n_\partial^{(2k)}$ by using a probabilistic telomere shortening model [5, 28]. Then, we compute an estimation of n_0 from these times by using $\bar{n}_0^{(2k,\alpha)}$ for $\alpha = 0.275$. The smoothing parameter has been chosen manually,



(a) Comparison between the curves of $\hat{n}_0^{(2k,\alpha)}$ and $n_0 = h_{1,4}$, for $k = 5$.

(b) Comparison between the curves of $\hat{n}_0^{(2k,\alpha)}$ and $n_0 = h_{1,4}$, for $k = 16$.

FIGURE 5. Estimation results for the estimator $\hat{n}_0^{(2k,\alpha)}$ defined in (5.8), when $b = 1$, $g = 1_{[0,1]}$, $N = 40$, $n_s = 3000$, $\alpha = 0.275$ and $(n_0, k) \in \{(h_{1,4}, 5), (h_{2,1.5}, 16)\}$.

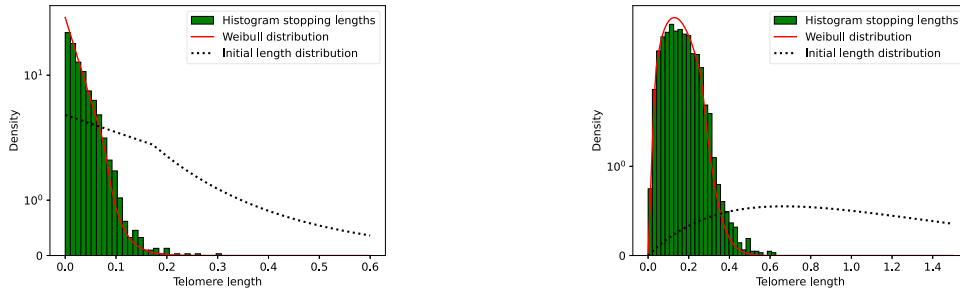
and does not need to be optimal. We do not choose one of the smoothing parameters $(\tilde{\alpha}_p)_{p \in (0,1]}$ introduced in Proposition 5.6 for the reason given in Remark 5.7. We plot in Figure 5 the estimated initial distributions (green curves), and compare them with their theoretical values (dotted curves). What we observe is surprising. The left tails of the initial distributions are first correctly estimated. However, at a certain telomere length, a peak appears on each estimated curve and the estimated density is 0 immediately afterwards. We thus have a very poor estimation.

To understand this, we investigate the n_s simulations of the probabilistic model used to generate our sequence of senescence times $(T_{2k,i})_{1 \leq i \leq n_s}$. For each of these simulations, we successively collect the initial length of the telomere signalling senescence, plot in Figure 6 the histogram of these lengths (green bars), and superpose the curve of n_0 on this histogram (black curve). In Figure 6a, which corresponds to the case where $(n_0, k) = (h_{1,4}, 5)$, we observe that even if n_0 has a large density after the length 0.35, no telomere with an initial length greater than 0.35 has signalled senescence. In Figure 6b, which corresponds to the case where $(n_0, k) = (h_{1,4}, 5)$, we have a similar observation. No telomere with an initial length greater than 0.65 has signalled senescence, whereas the density of n_0 is still large at this length. The random variables $(T_{2k,i})_{1 \leq i \leq n_s}$ therefore do not contain a significant part of the information about n_0 in these two cases. This lack of information explains the poor estimation in Figure 5. It also explains why we observe peaks: a smoothed Dirac measure appears in each estimation due to the abrupt loss of information.

The only way to manage this issue is to do the estimation with a larger number of data points. The reason is that it allows us to have better coverage of the telomere length that signals senescence at time $t = 0$. The problem is that the number of data points required to obtain a sufficiently large coverage of n_0 is far too high. The latter is illustrated by the bound on the errors of the estimators $(\bar{n}_0^{(2k,\tilde{\alpha}_p)})_{p \in (0,1]}$ we have obtained in Proposition 5.6. Indeed, this bound is of order $\frac{1}{(n_s)^{8k}}$ as $n_s \rightarrow +\infty$. Hence, the number of data points required to have a qualitative bound increases exponentially fast with the dimension.

Remark 5.7. The fact that we have a bound of order $\frac{1}{(n_s)^{8k}}$ as $n_s \rightarrow +\infty$ explains why we do not use one of the smoothing parameter $(\tilde{\alpha}_p)_{p \in (0,1]}$ in the estimations presented in Figure 5. As the bounds provided in Proposition 5.6 are not qualitative when $n_s = 3000$ and $k \in \{5, 16\}$, there is not benefit in choosing one of these smoothing parameters.

This problem of lack of data is in fact very common when working in high dimension, and is often referred to as *curse of the dimensionality*. In this context, we can go further, and link this problem to extreme value theory [22], which studies the distribution of the maximum and minimum of i.i.d. random variables. This link is shown in the following statement.



(a) Histogram when $(n_0, k) = (h_{1,4}, 5)$, and comparison with a Weibull distribution with parameters $(1, (H_{1,4})^{-1}(\frac{1}{10}))$.

(b) Histogram when $(n_0, k) = (h_{2,1.5}, 16)$, and comparison with a Weibull distribution with parameters $(2, (H_{2,1.5})^{-1}(\frac{1}{32}))$.

FIGURE 6. Histograms of the initial lengths of telomeres signalling senescence in the simulations done to generate $(T_{2k,i})_{1 \leq i \leq n_s}$, where $n_s = 3000$, at the symlog scale [37]. Comparison with the curve of n_0 (black curves) and with the density of a Weibull distribution (red curves). The symlog scale uses a logarithmic scale for large values and a linear scale for small ones. It allows us to make appear very small bars, while preserving the shape of the histogram close to that of the linear scale.

Proposition 5.8 (Minimum initial length when $k \rightarrow +\infty$). *Let $(\ell, \beta) \in \mathbb{N}^* \times \mathbb{R}_+^*$ and $(X_i)_{1 \leq i \leq 2k}$ a sequence of i.i.d. random variables with distribution given by $h_{\ell, \beta}$. Then, we have for all $x \geq 0$*

$$\lim_{k \rightarrow +\infty} \mathbb{P} \left[\frac{\min(X_1, \dots, X_{2k})}{(H_{\ell, \beta})^{-1}(\frac{1}{2k})} \leq x \right] = 1 - \exp(-x^\ell), \tag{5.11}$$

where $(H_{\ell, \beta})^{-1}$ is the reciprocal of the function $H_{\ell, \beta}$ defined in (5.1).

Remark 5.9. The right-hand side of (5.11) corresponds to the cumulative distribution function of a Weibull distribution with parameters $(\ell, 1)$. Hence, equation (5.11) means that $\min(X_1, \dots, X_{2k})$ can be approximated by a Weibull distribution with parameters $(\ell, (H_{\ell, \beta})^{-1}(\frac{1}{2k}))$ when $k \rightarrow +\infty$.

Proof. First, notice that as X_1 has for cumulative distribution function $H_{\ell, \beta}$, by the L'Hospital's rule and (5.1), we have for all $s \geq 0$

$$\lim_{h \rightarrow 0+} \frac{\mathbb{P}[-X_1 > -hs]}{\mathbb{P}[-X_1 > -h]} = \lim_{h \rightarrow 0+} \frac{H_{\ell, \beta}(hs)}{H_{\ell, \beta}(h)} = \lim_{h \rightarrow 0+} \frac{s \times h_{\ell, \beta}(hs)}{h_{\ell, \beta}(h)} = s^\ell. \tag{5.12}$$

Therefore, by using (5.12) to verify the assumptions of the Fisher–Tippett–Gnedenko's theorem in Corollary 1.6.3, Type III of [22], we obtain that for all $x \geq 0$

$$\lim_{k \rightarrow +\infty} \mathbb{P} \left[\frac{\max(-X_1, \dots, -X_{2k})}{(H_{\ell, \beta})^{-1}(\frac{1}{2k})} \leq -x \right] = \exp(-x^\ell).$$

We thus have that the proposition is proved, by using the above and the fact that for all $y \geq 0$ it holds $\mathbb{P}[\min(X_1, \dots, X_{2k}) \leq y] = 1 - \mathbb{P}[\max(-X_1, \dots, -X_{2k}) \leq -y]$. \square

From this result, we have that if n_0 is the density of an Erlang distribution and k is large, then the distribution of the shortest telomere at $t = 0$ depends only on the left tail of n_0 . In addition, we recall that in the approximated model, the distribution of senescence times is a scaled version of the distribution of the shortest telomere at

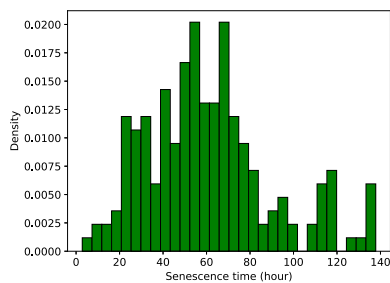
$t = 0$ (see Rem. 4.4). Then, we only have a good estimation of the left tail in Figure 5 because the cemetery only provides information about it in the original model, when $k \rightarrow +\infty$. This explanation is all the more valid since each of the histograms presented in Figure 6 has a shape similar to that of the density of a Weibull distribution with the parameters given in Remark 5.9.

From these observations, we conclude that even if our inference method is theoretically adaptable when we observe random variables instead of $n_{\partial}^{(2k)}$, it is in fact not usable in practice. In most cases, due to the lack of data, only information about the left tail can be inferred, and not about the whole curve. This suggests that the senescence times distribution is in fact mainly influenced by the left tail of n_0 , and much less by the overall distribution.

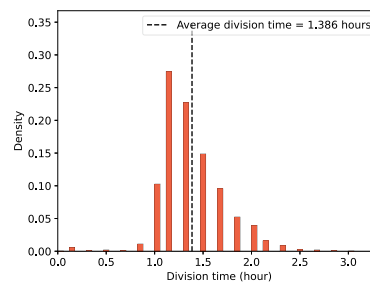
5.3. Estimation on experimental data

To conclude this study, we test our inference method on experimental data. The dataset we use here comes from [24]. It contains the senescence times of 187 different lineages, and the cell cycle duration of 1430 cells, both measured in hours. The histograms corresponding to these data are presented in Figures 7a and 7b. Our aim is to use the senescence times to obtain an estimation of n_0 . Then, we will compare this estimation with a previous estimation of the initial length distribution, obtained in [40] on the basis of an elongation-shortening model, see also [5, 28].

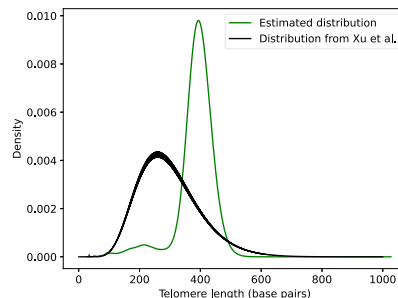
The budding yeast is a species with 16 chromosomes. We can estimate its division rate as $1/1.386 = 0.7216 \text{ h}^{-1}$ (hereafter denoted as h^{-1}) by computing the inverse of the average division times presented in Figure 7b. In addition, the shortening distribution of this species can be modelled by a uniform distribution from 5 to 10 base



(a) Histogram of senescence times coming from experiments.



(b) Histogram of division times coming from experiments.



(c) Estimation results on experimental data using $\bar{n}_0^{(32, \alpha)}$ with $\alpha = 0.1$. Comparison with the distribution obtained in Xu et al. [40].

FIGURE 7. Inference on experimental data: dataset and estimation results.

pairs (bp), see [35], and the threshold for senescence has been recently be estimated as $L_{\min} = 27$ bp in [33]. We thus proceed to the estimation by using the estimator $\bar{n}_0^{(2k,\alpha)}$ with the parameters $k = 16$, $b = 0.7216 \text{ h}^{-1}$, $m_1 = 7.5$ bp, $N = 40$ (see Rem. 5.1 and Sect. 2.3), and the smoothing parameter $\alpha = 0.1$. We apply this estimator to the senescence times presented in the previous paragraph, and plot in Figure 7c the density we have inferred shifted by L_{\min} (green curve). We also plot the estimation of the initial distribution coming from [40] on the same figure (black curve). As for the estimations presented in Section 5.2.3, we observe in Figure 7c that we first have an estimation of the left tail, then a peak, and finally an estimate of 0 everywhere. Let us comment on the estimated left tail, that is the only reliable estimation, see Section 5.2.3.

We observe in Figure 7c that our estimated tail does not grow as quickly as the tail of the distribution obtained in [40]. What is the most surprising is that even at lengths close to the mode of the black curve (around 200–240 base pairs), the tail of the estimated distribution has not grown yet. We thus have an estimation that differs from the one of [40]. The reason for this difference is surely linked to the fact that the model described by (2.5) is too simple. For example, using an exponential distribution for division times is not realistic, as illustrated by the shape of the histogram in Figure 7b. In addition, we do not know yet if the assumption discussed in Remark 2.4 is problematic or not. Solving this inverse problem in more complex models will allow us to identify which model assumptions are too restrictive. This work therefore opens up new modelling perspectives.

6. DISCUSSION

Our work began with a theoretical study, during which we study the connection between the senescence times distribution and the initial length density. We showed that there is a strong link between the two, provided that the ratio between the average telomere length and the average shortening value, denoted N , is large. We then have done a numerical study to understand how our method works in practice. Even if this study is encouraging in many cases, it has revealed several limitations. The first limitation is the poor estimation of initial distributions with a small coefficient of variation, see Section 5.1.2. In this case, the information on the initial distribution is blurred by the other sources of randomness (shortening values, cell cycle duration). The second limitation is the poor estimation of initial distributions in high dimension. In this case, the number of data points required for a good estimation is considerable, see Section 5.2.3. In addition, the error between (2.5) and (4.12) is large, see Section 5.1.3.

This work opens up new perspectives for telomere shortening models, including applications beyond those presented in this study. We have provided a rigorous justification that these models can be approximated by transport equations, and have used these approximations to solve our inverse problem. However, this justification has limitations, since for the species that motivate our study, N is not very large (for example, $N = 40$ for the yeast). The first perspective is thus to improve our approximations by adding a diffusion term to our transport equations, see [13, 32, 36]. Indeed, if we manage to show that telomere shortening models can be approximated by a transport-diffusion equations with an error of order $\frac{1}{N^2}$ instead of $\frac{1}{N}$, then it will be possible to construct estimators with estimation errors of order $\frac{1}{N^2}$ and improve the results presented in Section 5.1.2. In addition, on this more complex approximation, we strongly believe that thanks to the diffusion, the distribution of senescence times distribution is not only influenced by the distribution of the initial shortest telomere, see Remark 4.4, but also by the second shortest, third shortest etc. Hence, by obtaining information on the distribution of the shortest and second shortest initial telomeres thanks to this approximation, and then studying if these two distributions are close or not, it might be possible to obtain information on the correlations of the initial telomere lengths when the assumption presented in Remark 2.4 does not hold. Obtaining such a model approximation corresponds to a work in progress.

The second perspective is to find a model approximation when $k \rightarrow +\infty$. As mentioned above, problems arise when k is large, that may be related to extreme value theory or to the fact that the approximation error depends on k (see Thm. 2.7). It is therefore necessary to understand how the dynamics evolve when k is large. The main approach to achieve this is to use extreme value theory to our advantage, by constructing a typical particle representing the telomere with the smallest length among the $2k$ present in a cell. In view of the

Fisher–Tippett–Gnedenko’s Theorem 1.6.2 in [22], it will thus be possible to obtain a parametric law for the evolution of the shortest telomere over time, and thus to simplify the study of the phenomenon.

The third perspective, more directly linked to the inverse problem, is to improve our bounds on the approximation and estimation errors for the model with several telomeres. As explained in Section 2.2, the condition to have an exponential decay of the error in Theorem 2.7 (b) is that $\lambda + 2k\lambda'_N > \omega + (2k - 1)\omega'_N$. However, this condition is not optimal. The reason we think this is that from (A.22), a maximum principle in view of (H₃), and (4.36), one has that for all $t \geq 0$

$$n_{\partial}^{(2k)}(t) \leq (D_{\lambda})^{2k} k b m_1 \lambda'_N \exp[-k b m_1 \lambda'_N x].$$

Using the latter and Lemma 4.11 to bound $n_0^{(2k)}$, in view of the right-hand side of (2.10) and (2.11), yields that for all $x \geq 0$

$$\widehat{n}_0^{(2k)}(x) \leq (D_{\lambda})^{2k} \frac{\lambda'_N \exp[-2k\lambda'_N x]}{(D_{\omega} \exp[-2k\omega'_N x])^{1 - \frac{1}{2k}}} = \frac{(D_{\lambda})^{2k}}{(D_{\omega})^{2k-1}} \lambda'_N \exp[-2k\lambda'_N x + (2k - 1)\omega'_N x],$$

for which the right-hand side vanishes exponentially when $2k\lambda'_N > (2k - 1)\omega'_N$. In addition, the function n_0 vanishes exponentially when $x \rightarrow +\infty$ in view of (H₃). We thus have that the condition under which $|n_0^{(2k)}(x) - n_0(x)|$ vanishes exponentially is when $2k\lambda'_N > (2k - 1)\omega'_N$, which is less restrictive than the condition $\lambda + 2k\lambda'_N > \omega + (2k - 1)\omega'_N$ since $\omega \geq \lambda$. We thus would like to know if this is possible to obtain a better coefficient in the exponential in the bound of Theorem 2.7 (b), with a coefficient before the exponential at least of the same order. This can maybe be done by better controlling the term $\Delta_2(x)$ introduced in Section 4.6, where $x \in \mathbb{R}_+$. In addition, the constants $(D_{\lambda})^{2k}$ and $(\frac{D_{\lambda}}{D_{\omega}})^{2k}$ in respectively Proposition 4.5 and Theorem 2.7 (b) are not optimal. There are related to the loss of information on n_0 that occurs when we apply (H₃) or (H₄) to bound n_0 , and this loss increases exponentially with the dimension. We would like to obtain bounds on the estimation and approximation errors in which these two constants do not appear. This will allow us to prove that the approximation and estimation errors grow at most polynomially when the number of chromosomes increases.

The other perspectives are not related to model approximations, and correspond to questions that remain open. First, we would like to prove that the operator Ψ_{∂} defined in Remark 2.2, and its equivalent on (2.5), are invertible. Second, we would like to adapt our method for age-dependent models, see [28]. These two perspectives are the subject of future work.

ACKNOWLEDGMENTS

The author warmly thanks Marie Doumic for her guidance during the realisation of this work, and her careful proofreading of the article. He also thanks Milica Tomašević for her careful proofreading of the introduction. Finally, he thanks the anonymous referee for their comments and questions that allow him to improve this manuscript.

FUNDING

This work was partially funded by the Fondation Mathématique Jacques Hadamard and by the European Union ERC-2024-COG MUSEUM-101170884. Additional support was provided by the France 2030 program, administered by the French National Research Agency (ANR), under grant ANR-23-EXMA-0005. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (ERCEA). Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

- [1] P. Abdallah, P. Luciano, K.W. Runge, M. Lisby, V. Géli, E. Gilson and M.T. Teixeira, A two step model for senescence triggered by a single critically short telomere. *Nat. Cell Biol.* **11** (2009) 988–993.
- [2] O. Arino, M. Kimmel and G.F. Webb, Mathematical modeling of the loss of telomere sequences. *J. Theor. Biol.* **177** (1995) 45–57.

- [3] A. Armiento, M. Doumic, P. Moireau and H. Rezaei, Estimation from moments measurements for amyloid depolymerisation. *J. Theor. Biol.* **397** (2016) 68–88.
- [4] G. Barles, S. Mirrahimi and B. Perthame, Concentration in Lotka–Volterra parabolic or integral equations: a general convergence result. *MAA* **16** (2009) 321–340.
- [5] A. Benetos, C. Fritsch, E. Horton, L. Lenotre, S. Toupance and D. Villemonais, Stochastic branching models for the telomeres dynamics in a model including telomerase activity. *J. Math. Biol.* **91** (2025). DOI: [10.1007/s00285-025-02230-z](https://doi.org/10.1007/s00285-025-02230-z).
- [6] A. Benetos, O. Coudray, A. Gégout-Petit, L. Lenôtre, S. Toupance and D. Villemonais, A branching model for intergenerational telomere length dynamics. *J. Math. Biol.* **90** (2025). DOI: [10.1007/s00285-025-02185-1](https://doi.org/10.1007/s00285-025-02185-1).
- [7] T. Bourgeron, M. Doumic and M. Escobedo, Estimating the division rate of the growth-fragmentation equation with a self-similar kernel. *Inverse Probl.* **30** (2014). DOI: [10.1088/0266-5611/30/2/025007](https://doi.org/10.1088/0266-5611/30/2/025007).
- [8] T. Bourgeron, Z. Xu, M. Doumic and M.T. Teixeira, The asymmetry of telomere replication contributes to replicative senescence heterogeneity. *Sci. Rep.* **5** (2015). <https://pubmed.ncbi.nlm.nih.gov/26468778/>.
- [9] A. Charpentier and E. Flachaire, Log-transform kernel density estimation of income distribution. *L'Actualité économique* **91** (2015) 141–159. <https://ideas.repec.org/a/ris/actuec/0116.html>.
- [10] D. Cioranescu and P. Donato, An Introduction to Homogenization. Vol. 17 of *Oxford Lecture Series in Mathematics and Its Applications*. Oxford University Press (1999). DOI: [10.1093/oso/9780198565543.001.0001](https://doi.org/10.1093/oso/9780198565543.001.0001).
- [11] L. Comtet, *Advanced Combinatorics*, 1st edition. Springer, Dordrecht (1974). DOI: [10.1007/978-94-010-2196-8](https://doi.org/10.1007/978-94-010-2196-8).
- [12] M. Doumic and L.M. Tine, Estimating the division rate for the growth-fragmentation equation. *J. Math. Biol.* **67** (2013) 69–103. DOI: [10.1007/s00285-012-0553-6](https://doi.org/10.1007/s00285-012-0553-6).
- [13] M. Doumic and P. Moireau, Asymptotic approaches in inverse problems for depolymerization estimation. *Inverse Probl. Imaging* **20** (2025) 105–155. <https://www.aims sciences.org/article/id/6827171e4940ac5bf2c88f2e>.
- [14] A. Dvoretzky, J. Kiefer and J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** (1956) 642–669.
- [15] E. Ellefsen and N. Rodríguez, On some theory of monostable and bistable pure birth-jump integro-differential equations. *Ecol. Complex.* **45** (2021). <https://www.sciencedirect.com/science/article/pii/S1476945X20301720>.
- [16] S. Eugène, T. Bourgeron and Z. Xu, Effects of initial telomere length distribution on senescence onset and heterogeneity. *J. Theor. Biol.* **413** (2017) 58–65.
- [17] H. Hwang, A. Kreig, J. Calvert, J. Lormand, Y. Kwon, J.M. Daley, P. Sung, P.L. Opresko and S. Myong, Telomeric overhang length determines structural dynamics and accessibility to telomerase and ALT associated proteins. *Structure* **22** (2014) 842–853.
- [18] O.C. Ibe, *Fundamentals of Applied Probability and Random Processes*, 2nd edition. Academic Press (2014).
- [19] A. Jeddi, Convergence of a discrete selection-mutation model with exponentially decaying mutation kernel to a Hamilton–Jacobi equation. Preprint [arXiv:2412.06657](https://arxiv.org/abs/2412.06657) (2026).
- [20] S.I. Kabanikhin, *Inverse and Ill-posed Problems: Theory and Applications*. De Gruyter (2011). DOI: [10.1515/9783110224016](https://doi.org/10.1515/9783110224016).
- [21] D.G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text*, 3rd edition. *Statistics for Biology and Health*. Springer, New York (2012).
- [22] M.R. Leadbetter, G. Lindgren and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, 1st edition. *Springer Series in Statistics*. Springer, New York (1983). DOI: [10.1007/978-1-4612-5449-2](https://doi.org/10.1007/978-1-4612-5449-2).
- [23] A. Makroglou, Integral equations and actuarial risk management: some models and numerics. *Math. Model. Anal.* **8** (2003) 143–154.
- [24] H. Martin, M. Doumic, M.T. Teixeira and Z. Xu, Telomere shortening causes distinct cell division regimes during replicative senescence in *Saccharomyces cerevisiae*. *Cell Biosci.* **11** (2021). DOI: [10.1186/s13578-021-00693-3](https://doi.org/10.1186/s13578-021-00693-3).
- [25] P. Massart, The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** (1990) 1269–1283.
- [26] H.D. Nguyen, A.T. Jones and G.J. McLachlan, Positive data kernel density estimation via the log-KDE package for R. *Data Mining* (2019) 269–280. DOI: [10.1007/978-981-13-6661-1_21](https://doi.org/10.1007/978-981-13-6661-1_21).
- [27] J. Olayé, *Stochastic and deterministic approaches for studying telomere shortening and other cell dynamics*. Ph.D. thesis, École Polytechnique (2025).
- [28] J. Olayé and M. Tomasevic, Long-time behaviour of a multidimensional age-dependent branching process with a singular jump kernel modelling telomere shortening. *Electron. J. Probab.* **31** (2026). DOI: [10.1214/25-EJP1469](https://doi.org/10.1214/25-EJP1469).
- [29] J. Olayé, H. Bouzidi, A. Aristov, A. Barizien, S. Gutiérrez Ramos, C. Baroud and V. Bansaye, Estimation of the lifetime distribution from fluctuations in Bellman–Harris processes. *J. Math. Biol.* **90** (2025). DOI: [10.1007/s00285-025-02219-8](https://doi.org/10.1007/s00285-025-02219-8).

- [30] B. Perthame, Transport Equations in Biology, 1st edition. *Frontiers in Mathematics*. Birkhäuser, Basel (2007). DOI: [10.1007/978-3-7643-7842-4](https://doi.org/10.1007/978-3-7643-7842-4).
- [31] V. Pfeiffer and J. Lingner, Replication of telomeres and the regulation of telomerase. *Cold Spring Harb. Perspect. Biol.* **5** (2013). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3632062/>.
- [32] A.M. Portillo, E. Varela and J.A. García-Velasco, Influence of telomerase activity and initial distribution on human follicular aging: moving from a discrete to a continuum model. *Math. Biosci.* **358** (2023). <https://www.sciencedirect.com/science/article/pii/S0025556423000263>.
- [33] A. Rat, V. Martinez Fernandez, M. Doumic, M.T. Teixeira and Z. Xu, Mathematical model linking telomeres to senescence in *Saccharomyces cerevisiae* reveals cell lineage versus population dynamics. *Nat. Commun.* **16** (2025). <https://www.nature.com/articles/s41467-025-56196-z>.
- [34] N.J. Robinson and W.P. Schieman, Telomerase in cancer: function, regulation, and clinical translation. *Cancers* **14** (2022). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8834434/>.
- [35] J. Soudet, P. Jolivet and M. Teixeira, Elucidation of the DNA end-replication problem in *saccharomyces cerevisiae*. *Mol. Cell* **53** (2014) 954–964.
- [36] J.A.D. Wattis, Q. Qi and H.M. Byrne, Mathematical modelling of telomere length dynamics. *J. Math. Biol.* **80** (2020) 1039–1076.
- [37] J.B.W. Webber, A bi-symmetric log transformation for wide-range data. *Meas. Sci. Technol.* **24** (2012). <https://dx.doi.org/10.1088/0957-0233/24/2/027001>.
- [38] S. Węglarczyk, Kernel density estimation and its application. *ITM Web Conf.* **23** (2018).
- [39] D. Wynford-Thomas and D. Kipling, The end-replication problem. *Nature* **389** (1997) 551–552.
- [40] Z. Xu, K.D. Duc, D. Holcman and M.T. Teixeira, The length of the shortest telomere as the major determinant of the onset of replicative senescence. *Genetics* **194** 847–857 (2013).
- [41] Z. Xu, E. Fallet, C. Paoletti, S. Fehrmann, G. Charvin, and M.T. Teixeira, Two routes to senescence revealed by real-time analysis of telomerase-negative single lineages. *Nat. Commun.* **6** (2015). <http://www.nature.com/articles/ncomms8680>.
- [42] B. Zeinoun, *Replicative senescence-associated oxidative stress and Hog1 activation in telomerase negative Saccharomyces cerevisiae cells*. Ph.D. thesis, Sorbonne Université (2023). <https://theses.hal.science/tel-04480254v1>.

Please help to maintain this journal in open access!



This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.

APPENDIX A. AUXILIARY STATEMENTS

We present here auxiliary statements which are quite standard to obtain, and used frequently during all the paper. First, we provide in Appendix A.1 results related to the integro-differential equations we use in this work. Then, in Appendix A.2, we present statements dealing with the cardinalities of subsets of \mathcal{I}_k . Finally, in Appendix A.3, we give results about Erlang distributions that we use in our numerical study, see Section 5.

A.1. Auxiliary statements related to our integro-differential equations

We begin by presenting the statement that justifies the well-posedness of the integro-differential equations used in this work, as well as their uniqueness. The proof of this statement is based on a fixed-point argument, and is inspired by the ideas presented in Section 3.3 of [30]. We do not detail this proof, as it is relatively classical. We however mention, for the interested readers, that the detailed proof of this result is presented in p. 194 of [27].

Proposition A.1 (Well-posedness of the equations). *Let $d \in \mathbb{N}^*$, $\tilde{\xi}$ a finite measure on \mathbb{R}_+^d , $F \in C(\mathbb{R}_+, L^1(\mathbb{R}_+^d))$, and $w_0 \in L^1(\mathbb{R}_+^d)$. Then, there exists a unique solution in $C(\mathbb{R}_+; L^1(\mathbb{R}_+^d))$ to the following integro-differential equation*

$$\begin{cases} \partial_t w(t, x) = \int_{v \in \mathbb{R}_+^d} [w(t, x + v) - w(t, x)] \tilde{\xi}(dv) + F(t, x), & \forall (t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d, \\ w(0, x) = w_0(x), & \forall x \in \mathbb{R}_+^d. \end{cases} \tag{A.1}$$

In addition, if F and w_0 are non-negative, then w is non-negative.

We now present the second statement of this section, which corresponds to a maximum principle for solutions of (A.1) when the source term F is identically equal to 0. Again, we do not detail its proof, and refer to p. 195 in [27] for the full proof of this result.

Corollary A.2 (Maximum principle). *We work under the setting of Theorem 2.7, with $F = 0$ and w_0 non-negative. We consider $v_0 \in L^1(\mathbb{R}_+^d)$ verifying $|v_0| \leq w_0$, and $(v, w) \in (C(\mathbb{R}_+; L^1(\mathbb{R}_+^d)))^2$ two solutions of (A.1) with respective initial conditions v_0 and w_0 . Then, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$, we have*

$$|v(t, x)| \leq w(t, x).$$

The third statement we present is an intermediate statement to compute an explicit solution to (A.1) when $F = 0$. It can be proved by simply computing the derivative in time of the function given in (A.2).

Lemma A.3 (Complex explicit solutions to (A.1)). *Let $d \in \mathbb{N}^*$, $\tilde{\xi}$ a finite measure on \mathbb{R}_+^d , and two constants $C > 0$ and $p \in \mathbb{C}$. Then, the function $\zeta_{d,C,p} : \mathbb{R}_+ \times \mathbb{R}_+^d \rightarrow \mathbb{C}$ defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$ as*

$$\zeta_{d,C,p}(t, x) = C^d \exp \left[- \left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(p) \right) t - p \sum_{j=1}^d x_j \right] \tag{A.2}$$

verifies the first line of (A.1) with $F = 0$.

Now, the fourth statement we give is a consequence of Lemma A.3. It provides the explicit solution to (A.1) when w_0 is a product of an exponential and a cosine or sine function. It can be proved by simply computing $\frac{1}{2}(\zeta_{d,C,\beta+i\omega} + \zeta_{d,C,\beta-i\omega})$ and $\frac{1}{2i}(\zeta_{d,C,\beta+i\omega} - \zeta_{d,C,\beta-i\omega})$ for all $d \in \mathbb{N}^*$, $C > 0$ and $(\beta, \omega) \in \mathbb{R}_+^* \times \mathbb{R}$ by using (A.2), and then using the uniqueness result stated in Proposition A.1.

Proposition A.4 (Real explicit solutions to (A.1)). *We work under the setting of Theorem 2.7 with $F = 0$. We consider $C > 0$ and $(\beta, \omega) \in \mathbb{R}_+^* \times \mathbb{R}$, and denote $\tilde{\xi}(1) := \int_{v \in \mathbb{R}_+^d} 1 \tilde{\xi}(dv)$. The following statements hold.*

(1) *Assume that $w_0(x) = C^d \cos(\omega \sum_{j=1}^d x_j) \exp(-\beta \sum_{j=1}^d x_j)$ for all $x \in \mathbb{R}_+^d$. Then, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$, we have*

$$\begin{aligned} w(t, x) &= C^d \cos \left[\operatorname{Im} \left(\mathcal{L}(\tilde{\xi})(\beta + i\omega) \right) t - \omega \sum_{j=1}^d x_j \right] \\ &\times \exp \left[- \left(\tilde{\xi}(1) - \operatorname{Re} \left(\mathcal{L}(\tilde{\xi})(\beta + i\omega) \right) \right) t - \beta \sum_{j=1}^d x_j \right]. \end{aligned} \tag{A.3}$$

(2) *Assume that $w_0(x) = C^d \sin(\omega \sum_{j=1}^d x_j) \exp(-\beta \sum_{j=1}^d x_j)$ for all $x \in \mathbb{R}_+^d$. Then, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+^d$, we have*

$$\begin{aligned} w(t, x) &= C^d \sin \left[\operatorname{Im} \left(\mathcal{L}(\tilde{\xi})(\beta + i\omega) \right) t - \omega \sum_{j=1}^d x_j \right] \\ &\times \exp \left[- \left(\tilde{\xi}(1) - \operatorname{Re} \left(\mathcal{L}(\tilde{\xi})(\beta + i\omega) \right) \right) t - \beta \sum_{j=1}^d x_j \right]. \end{aligned} \tag{A.4}$$

Finally, the last statement that we present is useful to show that the cemeteries in (2.1) and (2.5) can be very close, even for two different initial conditions. This statement is essential to justify rigorously that the inverse problem we study in this work is ill-posed.

Proposition A.5 (Close cemeteries for different initial conditions). *We work under the setting of Theorem 2.7 with $\tilde{\xi}(\mathbb{R}_+^d \setminus \{0\}) > 0$, $F = 0$, and $w_0(x) = \exp(-\sum_{j=1}^d x_j)$ for all $x \in \mathbb{R}_+^d$. We denote $\tilde{\xi}(1) := \int_{v \in \mathbb{R}_+^d} 1 \tilde{\xi}(dv)$, the following function:*

$$\forall t \geq 0 : w_\partial(t) = \int_{y \in \mathbb{R}_+^d} w(t, y) \tilde{\xi}\left(\left\{v \in \mathbb{R}_+^d \mid y - v \notin \mathbb{R}_+^d\right\}\right) dy, \tag{A.5}$$

and the following sequences of functions, for all $n \in \mathbb{N}$:

- $w_0^{(n)}(x) = \frac{(1 + \sin(n \sum_{j=1}^d x_j)) \exp(-\sum_{j=1}^d x_j)}{\int_{y \in \mathbb{R}_+^d} (1 + \sin(n \sum_{j=1}^d y_j)) \exp(-\sum_{j=1}^d y_j) dy}$ for all $x \in \mathbb{R}_+^d$,
- $w^{(n)} \in C(\mathbb{R}_+, L^1(\mathbb{R}_+^d))$ corresponding to the solution of (A.1) with initial condition $w_0^{(n)}$,
- $w_\partial^{(n)}(t) = \int_{y \in \mathbb{R}_+^d} w^{(n)}(t, y) \tilde{\xi}(\{v \in \mathbb{R}_+^d \mid y - v \notin \mathbb{R}_+^d\}) dy$ for all $t \geq 0$.

Then, it holds:

$$\lim_{n \rightarrow +\infty} \left[\|w_\partial^{(n)} - w_\partial\|_{L^1(\mathbb{R}_+)} + \|w_\partial^{(n)} - w_\partial\|_{L^\infty(\mathbb{R}_+)} \right] = 0. \tag{A.6}$$

and

$$\lim_{n \rightarrow +\infty} \|w_0^{(n)} - w_0\|_{L^1(\mathbb{R}_+^d)} = \frac{2}{\pi}. \tag{A.7}$$

Proof. We begin by proving (A.6). For all $a > 0$, $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ verifying $\sup_{t \geq 0} (e^{at}|f(t)|) < +\infty$, one has that

$$\begin{aligned} \int_0^\infty |f(t)| dt + \sup_{t \geq 0} (|f(t)|) &\leq \left[\int_0^\infty e^{-at} dt \right] \left[\sup_{t \geq 0} (e^{at}|f(t)|) \right] + \sup_{t \geq 0} (e^{at}|f(t)|) \\ &= \left(\frac{1}{a} + 1 \right) \sup_{t \geq 0} (e^{at}|f(t)|). \end{aligned}$$

Therefore, in view of the above and the fact that $\mathcal{L}(\tilde{\xi})(1) < \mathcal{L}(\tilde{\xi})(0) = \tilde{\xi}(1)$ (because we have $\tilde{\xi}(\mathbb{R}_+^d \setminus \{0\}) > 0$, see (2.8)), we are going to prove the following in order to get (A.6):

$$\lim_{n \rightarrow +\infty} \left[\sup_{t \geq 0} \left[\exp\left[\left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1)\right)t\right] \left| w_\partial^{(n)}(t) - w_\partial(t) \right| \right] \right] = 0. \tag{A.8}$$

To do this, we first fix $n \in \mathbb{N}$, and denote the following to simplify notations:

$$\begin{aligned} C_n &:= \int_{y \in \mathbb{R}_+^d} \left(1 + \sin\left(n \sum_{j=1}^d y_j\right) \right) \exp\left(-\sum_{j=1}^d y_j\right) dy, \\ \forall y \in \mathbb{R}_+^d : A_y &:= \left\{ v \in \mathbb{R}_+^d \mid y - v \notin \mathbb{R}_+^d \right\}. \end{aligned} \tag{A.9}$$

By using the linearity of (A.1), and then (A.4), one has that for all $(t, y) \in \mathbb{R}_+ \times \mathbb{R}_+^d$

$$\begin{aligned} w^{(n)}(t, y) &= \frac{1}{C_n} \left(w(t, y) + \sin \left[\operatorname{Im} \left(\mathcal{L}(\tilde{\xi})(1 + in) \right) t - n \sum_{j=1}^d y_j \right] \right. \\ &\quad \left. \times \exp \left[- \left(\tilde{\xi}(1) - \operatorname{Re} \left(\mathcal{L}(\tilde{\xi})(1 + in) \right) \right) t - \sum_{j=1}^d y_j \right] \right). \end{aligned} \tag{A.10}$$

In addition, from (2.8) and the inequality $\cos(x) \leq 1$ for all $x \in \mathbb{R}$, one has that

$$\operatorname{Re} \left(\mathcal{L}(\tilde{\xi})(1 + in) \right) = \int_{y \in \mathbb{R}_+^d} \cos \left[-n \sum_{j=1}^d y_j \right] \exp \left[-\sum_{j=1}^d y_j \right] \tilde{\xi}(dy) \leq \mathcal{L}(\tilde{\xi})(1),$$

which implies that

$$\exp\left[\left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1)\right)t\right] \leq \exp\left[\left(\tilde{\xi}(1) - \operatorname{Re}\left(\mathcal{L}(\tilde{\xi})(1 + in)\right)\right)t\right]. \tag{A.11}$$

Therefore, by first subtracting both sides of (A.10) by $w(t, y)$, then integrating with respect to the measure $\exp[(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1))t]\tilde{\xi}(A_y) dy$ in view of (A.5), and finally applying the triangle inequality and (A.11) to simplify the second exponential in time, we have for all $t \geq 0$

$$\begin{aligned} \exp\left[\left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1)\right)t\right] \left|w_\partial^{(n)}(t) - w_\partial(t)\right| &\leq \left|\frac{1}{C_n} - 1\right| \exp\left[\left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1)\right)t\right] w_\partial(t) \\ &\quad + \frac{1}{C_n} \left| \int_{y \in \mathbb{R}_+^d} \sin\left[\operatorname{Im}\left(\mathcal{L}(\tilde{\xi})(1 + in)\right)t - n \sum_{j=1}^d y_j\right] \right. \\ &\quad \left. \times \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy \right|. \end{aligned} \tag{A.12}$$

Now, thanks to Theorem 2.6 in [10] and the following equalities (the third equality comes from the fact that $x \in \mathbb{R} \mapsto \cos(x)$ is 2π -periodic), for all $(u_1, \dots, u_{d-1}) \in \mathbb{R}_+^{d-1}$,

$$\begin{aligned} \int_{u_d \in (0, 2\pi)} \cos\left(\sum_{j=1}^d u_j\right) du_d &= \int_{u_d \in (0, 2\pi)} \cos\left(\sum_{j=1}^{d-1} u_j + u_d\right) du_d \\ &= \int_{u'_d \in (\sum_{j=1}^{d-1} u_j, \sum_{j=1}^{d-1} u_j + 2\pi)} \cos(u'_d) du'_d = \int_{u'_d \in (0, 2\pi)} \cos(u'_d) du'_d = 0, \end{aligned}$$

one can obtain that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \int_{y \in \mathbb{R}_+^d} \cos\left[n \sum_{j=1}^d y_j\right] \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy &= \left[\frac{1}{(2\pi)^d} \int_{u \in (0, 2\pi)^d} \cos\left[\sum_{j=1}^d u_j\right] du\right] \\ &\quad \times \left[\int_{y \in \mathbb{R}_+^d} \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy\right] = 0. \end{aligned} \tag{A.13}$$

By proceeding as above, we also have that

$$\lim_{n \rightarrow +\infty} \int_{y \in \mathbb{R}_+^d} \sin\left[n \sum_{j=1}^d y_j\right] \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy = 0. \tag{A.14}$$

Therefore, by first applying the equality $\sin(c-d) = \sin(c) \cos(d) + \cos(c) \sin(d)$ for the real numbers $c = \operatorname{Im}(\mathcal{L}(\tilde{\xi})(1 + in))t$ and $d = n \sum_{j=1}^d y_j$, then using the triangle inequality and the facts that $\sup_{s \geq 0} (|\sin(s)|) = 1$ and $\sup_{s \geq 0} (|\cos(s)|) = 1$, and finally applying (A.13) and (A.14) (the limit is equal to the limit superior in these equations), we obtain that

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \left[\sup_{t \geq 0} \left[\left| \int_{y \in \mathbb{R}_+^d} \sin\left[\operatorname{Im}\left(\mathcal{L}(\tilde{\xi})(1 + in)\right)t - n \sum_{j=1}^d y_j\right] \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy \right| \right] \right] \\ \leq \limsup_{n \rightarrow +\infty} \left[\sup_{t \geq 0} \left[\left| \sin[\operatorname{Im}(\mathcal{L}(\tilde{\xi})(1 + in))t] \right| \int_{y \in \mathbb{R}_+^d} \cos\left[n \sum_{j=1}^d y_j\right] \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy \right] \right] \\ + \limsup_{n \rightarrow +\infty} \left[\sup_{t \geq 0} \left[\left| \cos[\operatorname{Im}(\mathcal{L}(\tilde{\xi})(1 + in))t] \right| \int_{y \in \mathbb{R}_+^d} \sin\left[n \sum_{j=1}^d y_j\right] \exp\left[-\sum_{j=1}^d y_j\right] \tilde{\xi}(A_y) dy \right] \right] = 0. \end{aligned} \tag{A.15}$$

Moreover, in view of the first line of (A.9) and Theorem 2.6 in [10] (we proceed as in (A.13)), and then of (A.5) and (A.3) for $\beta = 1$ and $\omega = 0$ (recall that $w_0(x) = \exp(-\sum_{j=1}^d x_j)$ for all $x \in \mathbb{R}_+^d$), one has that

$$\begin{aligned} \lim_{n \rightarrow +\infty} C_n &= 1 + \lim_{n \rightarrow +\infty} \int_{y \in \mathbb{R}_+^d} \sin\left(n \sum_{j=1}^d y_j\right) \exp\left(-\sum_{j=1}^d y_j\right) dy = 1, \\ \sup_{t \geq 0} \left[\exp\left[\left(\tilde{\xi}(1) - \mathcal{L}(\tilde{\xi})(1)\right)t\right] w_\partial(t) \right] &< +\infty. \end{aligned} \tag{A.16}$$

Therefore, by taking the supremum in $t \geq 0$ and the limit when $n \rightarrow +\infty$ in (A.12) in view of (A.15) and (A.16), we obtain that (A.8) is true, which yields (A.6).

Now, it remains to prove (A.7). Let $n \in \mathbb{N}$. In view of the definitions of $w_0^{(n)}$, w_0 , and C_n , one has that

$$\|w_0^{(n)} - w_0\|_{L^1(\mathbb{R}_+^d)} = \int_{y \in \mathbb{R}_+^d} \left| \left(\frac{1}{C_n} - 1\right) + \frac{\sin\left(n \sum_{j=1}^d y_j\right)}{C_n} \right| \exp\left(-\sum_{j=1}^d y_j\right) dy.$$

This yields, in view of the reverse triangle inequality, and the following classical results:

$$\forall y \in \mathbb{R}_+^d : \left| 1 + \sin\left(n \sum_{j=1}^d y_j\right) \right| \leq 2 \quad \text{and} \quad \int_{y \in \mathbb{R}_+^d} \exp\left(-\sum_{j=1}^d y_j\right) dy = 1, \tag{A.17}$$

that

$$\begin{aligned} &\left\| \|w_0^{(n)} - w_0\|_{L^1(\mathbb{R}_+^d)} - \int_{y \in \mathbb{R}_+^d} \left| \sin\left(n \sum_{j=1}^d y_j\right) \right| \exp\left(-\sum_{j=1}^d y_j\right) dy \right\| \\ &\leq \int_{y \in \mathbb{R}_+^d} \left| \left(\frac{1}{C_n} - 1\right) \left(1 + \sin\left(n \sum_{j=1}^d y_j\right)\right) \right| \exp\left(-\sum_{j=1}^d y_j\right) dy \leq 2 \left| \frac{1}{C_n} - 1 \right|. \end{aligned}$$

Combining the latter with the first line of (A.16), and then using Theorem 2.6 in [10] in view of the fact that $x \in \mathbb{R}^d \mapsto |\sin(\sum_{j=1}^d x_j)|$ is $(0, \pi)^d$ -periodic (see [10], Def. 2.1), yields that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \|w_0^{(n)} - w_0\|_{L^1(\mathbb{R}_+^d)} &= \lim_{n \rightarrow +\infty} \int_{y \in \mathbb{R}_+^d} \left| \sin\left(n \sum_{j=1}^d y_j\right) \right| \exp\left(-\sum_{j=1}^d y_j\right) dy \\ &= \left[\frac{1}{\pi^d} \int_{u \in (0, \pi)^d} \left| \sin\left(\sum_{j=1}^d u_j\right) \right| du \right] \left[\int_{y \in \mathbb{R}_+^d} \exp\left(-\sum_{j=1}^d y_j\right) dy \right]. \end{aligned} \tag{A.18}$$

In addition, in view of the fact that $\int_{(0, \pi)} |\sin(x + \sigma)| dx = \int_0^\pi |\sin(x)| dx$ for all $\sigma \in \mathbb{R}$ because the function $x \in \mathbb{R} \mapsto |\sin(x)|$ is π -periodic, and then the inequality $\sin(x) \geq 0$ for $x \in (0, \pi)$, we have that

$$\begin{aligned} \frac{1}{\pi^d} \int_{u \in (0, \pi)^d} \left| \sin\left(\sum_{j=1}^d u_j\right) \right| du &= \frac{1}{\pi^d} \int_{(u_1, \dots, u_{d-1}) \in (0, \pi)^{d-1}} \left[\int_{u_d \in (0, \pi)} |\sin(u_d)| du_d \right] du_1 \dots du_{d-1} \\ &= \frac{1}{\pi^d} \int_{(u_1, \dots, u_{d-1}) \in (0, \pi)^{d-1}} \left[\int_{u_d \in (0, \pi)} \sin(u_d) du_d \right] du_1 \dots du_{d-1} = \frac{2}{\pi}. \end{aligned}$$

Then, by plugging the above and the right-hand side of (A.17) in (A.18), we obtain (A.7). □

A.2. Auxiliary statements related to \mathcal{I}_k

We now provide statements related to the cardinality of subsets of \mathcal{I}_k . The first statement we present is the following, and corresponds to Lemma 4.9 in [28]. It provides information about the number of sets in \mathcal{I}_k , and about the number of sets that contain/not contain a chosen index.

Lemma A.6 (Cardinality of subsets of \mathcal{I}_k , less than 1 index fixed). *It holds*

$$\#(\mathcal{I}_k) = 2^k, \quad \text{and} \quad \forall i \in \llbracket 1, 2k \rrbracket : \frac{\#\{I \in \mathcal{I}_k \mid i \in I\}}{2^k} = \frac{\#\{I \in \mathcal{I}_k \mid i \notin I\}}{2^k} = \frac{1}{2}.$$

The second lemma we present is not given in [28]. It provides the number of sets in \mathcal{I}_k that contain a chosen pair. We briefly sketch its proof but do not detail it, as it mainly consists in adapting the proof of Lemma 4.9 in [28].

Lemma A.7 (Cardinality of subsets of \mathcal{I}_k , 2 indexes fixed). *For all $(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2$ such that $\ell \neq \ell' \bmod k$, we have*

$$\#\{I \in \mathcal{I}_k \mid \ell \in I, \ell' \in I\} = \begin{cases} 0, & \text{if } k = 1, \\ 2^{k-2}, & \text{if } k \geq 2. \end{cases} \quad (\text{A.19})$$

Proof. When $k = 1$, we have by (2.2) that $\mathcal{I}_1 = \{\{1\}, \{2\}\}$, so that (A.19) is true for all $(\ell, \ell') \in \llbracket 1, 2 \rrbracket^2$ such that $\ell \neq \ell'$. When $k = 2$, the expression of \mathcal{I}_2 is given in Example 2.3, and we see that whatever the values of $(\ell, \ell') \in \llbracket 1, 4 \rrbracket^2$ verifying $\ell \neq \ell' \bmod 2$, (A.19) holds. We thus now focus in the case where $k \geq 3$. We consider $(\ell, \ell') \in \llbracket 1, 2k \rrbracket^2$ such that $\ell \neq \ell' \bmod k$. Following Section 4.4.3 in [28], one can prove that $f : \{0, 1\}^{k-2} \rightarrow \{I \in \mathcal{I}_k \mid \ell \in I, \ell' \in I\}$, defined for all $x \in \{0, 1\}^{k-2}$ as

$$f(x) = \{kx_j + j \mid j \in \llbracket 1, k \rrbracket, j \neq \ell \bmod k, j \neq \ell' \bmod k\} \cup \{\ell, \ell'\}$$

is bijective. Hence, as $\#\{0, 1\}^{k-2} = 2^{k-2}$, we obtain that (A.19) is true when $k \geq 3$, which concludes the proof of the lemma. \square

A.3. Auxiliary statements related to Erlang distributions

Throughout this section, we denote for all $\beta > 0$ the following constants:

$$\beta_{N,1} = \frac{N}{m_1} \left[1 - \mathcal{L}(g) \left(\frac{\beta}{N} \right) \right], \quad \beta_{N,2k} = \frac{N}{km_1} \left[1 - \mathcal{L}(\mu) \left(\frac{\beta}{N} \right) \right]. \quad (\text{A.20})$$

These constants are similar to the ones introduced in (2.9) ($\beta_{N,2k}$ is similar to λ'_N and ω'_N by (4.36)), and correspond to approximations of β when N is large.

The first statement we present is useful in the specific case where n_0 is an exponential distribution, *i.e.* an Erlang distribution with parameter $\ell = 1$. It gives the explicit formula of the estimator $\widehat{n}_0^{(d)}$, for all $d \in \{1, 2k\}$.

Proposition A.8 (Explicit estimators for exponential initial distributions). *Assume that there exists $\beta > 0$ such that $n_0 = h_{1,\beta}$, defined in (5.1). Then, for all $d \in \{1, 2k\}$ and $x \geq 0$, we have*

$$\widehat{n}_0^{(d)}(x) = \beta_{N,d} \exp[-\beta_{N,d}x]. \quad (\text{A.21})$$

Proof. The proof of this proposition when $d = 1$ follows the same steps as when $d = 2k$, but is easier. Hence, we only give the proof when $d = 2k$. Notice that by (2.5), (A.3) with $\omega = 0$ and (A.20), we have that

$$n^{(2k)}(t, x) = \beta^{2k} \exp[-kbm_1\beta_{N,2k}t - \beta \sum_{i=1}^{2k} x_i]$$

for all $t \geq 0$, $x \in \mathbb{R}_+^{2k}$. Our aim is to use this expression of $n^{(2k)}$ to obtain (A.21). To do so, we first rewrite $\mu(\{v \in \mathbb{R}_+^{2k} \mid Ny - v \notin \mathbb{R}_+^{2k}\})$ in the second line of (4.6) with an integral. Then, we switch the integrals and use the equality $1_{\{Ny - v \notin \mathbb{R}_+^{2k}\}} = 1 - 1_{\{v_i \in \llbracket 1, 2k \rrbracket : y_i \geq \frac{v_i}{N}\}}$. Finally, we apply the expression of $n^{(2k)}$ and equation (A.20). We obtain for all $t \geq 0$

$$\begin{aligned} n_{\partial}^{(2k)}(t) &= bN \int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \left[\int_{v \in \mathbb{R}_+^{2k}} 1_{\{Ny - v \notin \mathbb{R}_+^{2k}\}} \mu(dv) \right] dy \\ &= bN \int_{v \in \mathbb{R}_+^{2k}} \left[\int_{y \in \mathbb{R}_+^{2k}} n^{(2k)}(t, y) \left(1 - 1_{\{v_i \in \llbracket 1, 2k \rrbracket : y_i \geq \frac{v_i}{N}\}} \right) dy \right] \mu(dv) \\ &= bN \left[1 - \mathcal{L}(\mu) \left(\frac{\beta}{N} \right) \right] \exp[-kbm_1\beta_{N,2k}t] = kbm_1\beta_{N,2k} \exp[-kbm_1\beta_{N,2k}t]. \end{aligned} \quad (\text{A.22})$$

Therefore, by plugging the above in (2.10), and then using (2.11) to replace bm_1 with $\widetilde{b}\widetilde{m}_1$, we obtain (A.21) \square

The second proposition we present is devoted to the more general case in which n_0 is an Erlang distribution, but not necessarily an exponential distribution. In this case, $n^{(1)}$ is explicit. One can then compute $n_\delta^{(1)}$ by using (2.1), and then $\widehat{n}_0^{(1)}$ thanks to (2.10).

Proposition A.9 (Explicit solutions for Erlang initial distributions). *Assume that there exist $\ell \in \mathbb{N}^*$ and $\beta > 0$ such that $n_0 = h_{\ell,\beta}$, defined in (5.1). We consider for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$ the function $\psi_{x,t} \in L^1(\mathbb{R}_+^*)$, defined for all $\alpha > 0$ as $\psi_{x,t}(\alpha) := -bN[1 - \mathcal{L}(g)(\frac{\alpha}{N})]t - \alpha x$. Then, for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$, we have*

$$n^{(1)}(t, x) = (-1)^{\ell-1} \frac{\beta^\ell}{(\ell-1)!} B_{\ell-1} \left[\frac{d}{d\alpha} \psi_{x,t}(\beta), \dots, \frac{d^{\ell-1}}{d\alpha^{\ell-1}} \psi_{x,t}(\beta) \right] \exp(\psi_{x,t}(\beta)), \tag{A.23}$$

where $B_{\ell-1}$ is a complete Bell polynomial of order $\ell - 1$, see equation 3.c, p. 134 in [11].

Proof. We denote $f \in L^1(\mathbb{R}_+^* \times \mathbb{R}_+^2)$ the function defined for all $(\alpha, t, x) \in \mathbb{R}_+^* \times \mathbb{R}_+^2$ as $f(\alpha, t, x) := \exp[\psi_{x,t}(\alpha)]$. One can easily see that by integrating f and computing its derivative, we have for all $(\alpha, t, x) \in \mathbb{R}_+^* \times \mathbb{R}_+^2$

$$\partial_t f(\alpha, t, x) = bN \int_0^\delta [f(\alpha, t, x + u) - f(\alpha, t, x)] g(u) du. \tag{A.24}$$

We also know by the Faà di Bruno’s formula Theorem C, p. 139 in [11] and the definition of a complete Bell polynomial equation 3.c, p. 134 in [11], that it holds for all $(\alpha, t, x) \in \mathbb{R}_+^* \times \mathbb{R}_+^2$

$$\frac{d^{\ell-1}}{d\alpha^{\ell-1}} f(\alpha, t, x) = B_{\ell-1} \left[\frac{d}{d\alpha} \psi_{x,t}(\alpha), \dots, \frac{d^{\ell-1}}{d\alpha^{\ell-1}} \psi_{x,t}(\alpha) \right] \exp(\psi_{x,t}(\alpha)). \tag{A.25}$$

Then, by using equations (A.25) and (A.24), we have that the function $\varphi_1 \in L^1(\mathbb{R}_+^2)$ defined for all $(t, x) \in \mathbb{R}_+ \times \mathbb{R}_+$ as $\varphi_1(t, x) := (-1)^{\ell-1} \frac{\beta^\ell}{(\ell-1)!} \frac{d^{\ell-1}}{d\alpha^{\ell-1}} f(\beta, t, x)$ is equal to the right-hand side of (A.23), and verifies the same equation as $n^{(1)}$, see the first line of (3.4). In addition, we have that $\varphi_1(0, x) = \frac{\beta^\ell}{(\ell-1)!} x^{\ell-1} e^{-\beta x} = h_{\ell,\beta}(x)$ for all $x \geq 0$, in view of the fact that by equation (A.25), equation 3.n’, p. 136 in [11], and equation 3.c, p. 134 in [11], it holds

$$\frac{d^{\ell-1}}{d\alpha^{\ell-1}} f(\beta, 0, x) = B_{\ell-1}[-x, 0, \dots, 0] \exp(-\beta x) = (-x)^{\ell-1} \exp(-\beta x).$$

As the first line of (3.4) has a unique solution in $C(\mathbb{R}, L^1(\mathbb{R}_+^2))$ with initial condition $h_{\ell,\beta}$ by Proposition A.1, we obtain that $\varphi_1 = n^{(1)}$, so that equation (A.23) is true. □

The final proposition we present provides the value of the mean, the variance, and the coefficient of variation of an Erlang distribution. This result is quite standard, so we do not prove it. We however refer to p. 138 in [18] for a proof of the expression of the mean and the variance.

Proposition A.10 (Moments of Erlang distributions). *Let us fix $\ell \in \mathbb{N}^*$ and $\beta > 0$. We consider*

$$m = \int_0^{+\infty} x h_{\ell,\beta}(x) dx \quad \text{and} \quad \sigma^2 = \int_0^{+\infty} x^2 h_{\ell,\beta}(x) dx - m^2,$$

which correspond respectively to the mean and the variance of a random variable distributed according to $h_{\ell,\beta}$. We also consider $cv = \frac{\sigma}{\mu}$ the coefficient of variation of this distribution. Then, we have

$$m = \frac{\ell}{\beta}, \quad \sigma^2 = \frac{\ell}{\beta^2}, \quad \text{and} \quad cv = \frac{1}{\sqrt{\ell}}.$$

APPENDIX B. PROOF OF PROPOSITIONS 5.2 AND 5.6

This section is devoted to the proof of Propositions 5.2 and 5.6. First, in Appendix B.1, we introduce the notations and auxiliary results we need to obtain these propositions. Then, in Appendix B.2, we prove one of these auxiliary results. Finally, in Appendix B.3, we detail the proof of Proposition 5.6 from this statement. The proof of Proposition 5.2 follows exactly the same steps as those given in Appendix B.3, and is even slightly easier. Hence, we do not detail it.

B.1. Notations and auxiliary statements

We begin by introducing the notations we use in this section. We first consider the following set

$$\mathcal{X}^\infty(\mathbb{R}) = \{f \in \mathcal{D}'(\mathbb{R}) \mid \exists \varphi \in L^\infty(\mathbb{R}) \text{ s.t. } f = \varphi'\}.$$

This set contains distributions that are not necessarily representable by a function. For example, as for all $x \in \mathbb{R}$ and $x_0 \in \mathbb{R}$ it holds $\frac{d}{dx}(1_{x \leq x_0}) = \delta_{x_0}(\{x\})$, and as the shifted Heaviside functions are in $L^\infty(\mathbb{R})$, we have $\delta_{x_0} \in \mathcal{X}^\infty(\mathbb{R})$. We then introduce the functional $\mathcal{K} : \mathcal{X}^\infty(\mathbb{R}) \rightarrow \mathbb{R}_+$, defined for all $f \in \mathcal{X}^\infty(\mathbb{R})$ as

$$\mathcal{K}(f) = \inf_{\varphi \in L^\infty(\mathbb{R}) \text{ s.t. } f = \varphi'} (\|\varphi\|_{L^\infty(\mathbb{R})}). \quad (\text{B.1})$$

This functional allows us to control the distance between two distributions in $\mathcal{X}^\infty(\mathbb{R})$ by using two functions. For example, we can control the distance between two Dirac measures by two shifted Heaviside functions. We finally define the following functions, for all $d \in \{1, 2k\}$, $y \in \mathbb{R}$,

$$n_{\log}^{(d)}(y) = \exp(y) \widehat{n}_0^{(d)}(\exp(y)), \quad (\text{B.2})$$

and the following measures

$$\begin{aligned} \bar{n}_{\log}^{(1)}(dx) &= \frac{1}{n_s} \sum_{i=1}^{n_s} \delta_{\log(\tilde{b}\tilde{m}_1 T_{1,i})}(dx), \\ \bar{n}_{\log}^{(2k)}(dx) &= \sum_{i=0}^{n_s} \left(1 - \frac{i}{n_s}\right)^{\frac{1}{2k}} \left[\delta_{\log(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i+1})}(dx) - \delta_{\log(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i})}(dx) \right], \end{aligned} \quad (\text{B.3})$$

where $(T_{1,i})_{1 \leq i \leq n_s}$ and $(T_{2k,i})_{0 \leq i \leq n_s+1}$ are the sequences of random variables introduced in Sections 5.2.1 and 5.2.2. Since Dirac measures belong to $\mathcal{X}^\infty(\mathbb{R})$, we have that $\bar{n}_{\log}^{(d)} \in \mathcal{X}^\infty(\mathbb{R})$ for all $d \in \{1, 2k\}$. The function $n_{\log}^{(d)}$, defined in (B.2), is the distribution of the random variable $\log(Z^{(d)})$, where $Z^{(d)}$ is distributed according to $\widehat{n}_0^{(d)}$. The measure $\bar{n}_{\log}^{(1)}$, in the first line of (B.3), is the empirical estimator of $n_{\log}^{(1)}$. The measure $\bar{n}_{\log}^{(2k)}$, in the next line, is an estimator of $n_{\log}^{(2k)}$. The latter has been constructed by computing the weak derivative of the empirical cumulative distribution function associated to $n_{\log}^{(2k)}$, as done in the proof of Proposition 5.5.

We now present the two auxiliary statements that we need to prove Propositions 5.2 and 5.6. The first statement we provide is the following. It corresponds to an intermediate statement to obtain the second one. We briefly sketch its proof below but do not detail it, as the inequalities used in the proof are relatively classical, see [3, 7, 12].

Lemma B.1 (Bounds on the logarithm of the data). *Let us consider $d \in \{1, 2k\}$, and $\alpha > 0$. Assume that (H₁)–(H₄) hold, and that $\beta'_N = (\lambda + 2k\lambda'_N) - (\omega + (2k-1)\omega'_N) > 0$ when $d = 2k$. We denote for all $x \in \mathbb{R}$: $\rho_\alpha(x) = \frac{1}{\alpha} \rho(\frac{x}{\alpha})$. Then, it holds*

$$\left\| \rho_\alpha * \bar{n}_{\log}^{(d)} - n_{\log}^{(d)} \right\|_{L^\infty(\mathbb{R})} \leq \alpha^{-1} \sqrt{\frac{2}{\pi}} \mathcal{K}(\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) + \alpha \sqrt{\frac{2}{\pi}} \left\| (n_{\log}^{(d)})' \right\|_{L^\infty(\mathbb{R})}. \quad (\text{B.4})$$

Proof. First, notice that by the triangle inequality, it holds

$$\left\| \rho_\alpha * \bar{n}_{\log}^{(d)} - n_{\log}^{(d)} \right\|_{L^\infty(\mathbb{R})} \leq \left\| \rho_\alpha * (\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) \right\|_{L^\infty(\mathbb{R})} + \left\| \rho_\alpha * n_{\log}^{(d)} - n_{\log}^{(d)} \right\|_{L^\infty(\mathbb{R})}. \quad (\text{B.5})$$

Then, notice that by adapting the proof of Lemmas A.1.(1) and A.1.(4) in [7], we can obtain similar inequalities as in Lemmas 3.(iii) and 3.(i) of [3], for all $\varphi \in L^\infty(\mathbb{R})$ such that $\bar{n}_{\log}^{(d)} - n_{\log}^{(d)} = \varphi'$ and $f \in W^{1,\infty}(\mathbb{R})$,

$$\begin{aligned} \left\| \rho_\alpha * (\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) \right\|_{L^\infty(\mathbb{R})} &= \left\| \rho_\alpha * \varphi' \right\|_{L^\infty(\mathbb{R})} \leq \alpha^{-1} \|\rho'\|_{L^1(\mathbb{R})} \|\varphi\|_{L^\infty(\mathbb{R})}, \\ \left\| \rho_\alpha * f - f \right\|_{L^\infty(\mathbb{R})} &\leq \alpha \|\text{Id} \cdot \rho\|_{L^1(\mathbb{R})} \|f'\|_{L^\infty(\mathbb{R})}. \end{aligned} \quad (\text{B.6})$$

In view of (B.1), by taking the infimum in the first line of (B.6), we have that

$$\left\| \rho_\alpha * (\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) \right\|_{L^\infty(\mathbb{R})} \leq \alpha^{-1} \|\rho'\|_{L^1(\mathbb{R})} \mathcal{K}(\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}).$$

Therefore, by first plugging the above and the second line of (B.6) in (B.5), and then using that

$$\|\rho'\|_{L^1(\mathbb{R})} = \|\text{Id} \cdot \rho\|_{L^1(\mathbb{R})} = \sqrt{\frac{2}{\pi}}$$

as ρ is the density of a standard Gaussian distribution, we obtain (B.4). □

The second statement we give is the following. It provides for all $d \in \{1, 2k\}$ a bound on the supremum error between $\bar{n}_0^{(d,\alpha)}$ and n_0 , when the error between $\bar{n}_{\log}^{(d)}$ and $n_{\log}^{(d)}$ is known. It is proved in Section B.2.

Lemma B.2 (Error bounds for simulated data). *We recall the constants $C_{\hat{n},1}$ and $C_{\hat{n},2k}$ defined in (5.5) and (5.9) respectively. Assume that the assumptions of Lemma B.1 hold, and there exists $\varepsilon > 0$ such that $\mathcal{K}(\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) \leq \varepsilon$. Then, there exists a sequence $(L_{d,n})_{n \in \mathbb{N}}$ of positive real numbers such that $\limsup_{n \rightarrow +\infty} L_{d,n} < +\infty$ and*

$$\left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - n_0 \right] \right\|_{L^\infty(\mathbb{R}_+)} \leq \alpha^{-1} \sqrt{\frac{2}{\pi}} \varepsilon + \alpha \sqrt{\frac{2}{\pi}} C_{\hat{n},d} + \frac{L_{d,N}}{N}. \tag{B.7}$$

Remark B.3. The minimum of the function $\alpha \in \mathbb{R}_+^* \mapsto \alpha^{-1} \sqrt{\frac{2}{\pi}} \varepsilon + \alpha \sqrt{\frac{2}{\pi}} C_{\hat{n},d}$ can be computed by analysing the sign of its derivative, and is in $\alpha^* = \left(\frac{\varepsilon}{C_{\hat{n},d}}\right)^{\frac{1}{2}}$. Then, in view of (B.7), the smoothing parameter providing the best bound on the error is α^* .

We now conclude this section by proving Lemma B.2 and Proposition 5.6.

B.2. Proof of Lemma B.2

First, decomposing $\left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - n_0 \right] \right\|_{L^\infty(\mathbb{R}_+)}$ into two terms thanks to the triangle inequality, and then using Theorem 2.7 to bound the second term, yields that there exists a sequence $(L_{d,n})_{n \in \mathbb{N}}$ of positive real numbers such that $\limsup_{n \rightarrow +\infty} L_{d,n} < +\infty$ and

$$\left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - n_0 \right] \right\|_{L^\infty(\mathbb{R}_+)} \leq \left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - \hat{n}_0^{(d)} \right] \right\|_{L^\infty(\mathbb{R}_+)} + \frac{L_{d,N}}{N}. \tag{B.8}$$

We thus only have to bound the term $\left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - \hat{n}_0^{(d)} \right] \right\|_{L^\infty(\mathbb{R}_+)}$, and (B.7) will be true. To do so, notice that by the change of variable $y = \log(x)$, the definition of $\bar{n}_0^{(1,\alpha)}$ and $\bar{n}_0^{(2k,\alpha)}$ given in (5.4)–(5.8), and the ones of $n_{\log}^{(d)}$ and $\bar{n}_{\log}^{(d)}$ given in (B.2) and (B.3), we have (recall that $\rho_\alpha = \frac{1}{\alpha} \rho(\frac{\cdot}{\alpha})$)

$$\begin{aligned} \left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - \hat{n}_0^{(d)} \right] \right\|_{L^\infty(\mathbb{R}_+)} &= \sup_{x \in \mathbb{R}_+} \left[x \left| \bar{n}_0^{(d,\alpha)}(x) - \hat{n}_0^{(d)}(x) \right| \right] \\ &= \sup_{y \in \mathbb{R}} \left[\exp(y) \left| \bar{n}_0^{(d,\alpha)}(\exp(y)) - \hat{n}_0^{(d)}(\exp(y)) \right| \right] \\ &= \left\| \rho_\alpha * \bar{n}_{\log}^{(d)} - n_{\log}^{(d)} \right\|_{L^\infty(\mathbb{R})}. \end{aligned}$$

Therefore, by first using Lemma B.1, and then the fact that $\mathcal{K}(\bar{n}_{\log}^{(d)} - n_{\log}^{(d)}) \leq \varepsilon$, we obtain

$$\left\| \text{Id} \left[\bar{n}_0^{(d,\alpha)} - \hat{n}_0^{(d)} \right] \right\|_{L^\infty(\mathbb{R}_+)} \leq \alpha^{-1} \sqrt{\frac{2}{\pi}} \varepsilon + \alpha \sqrt{\frac{2}{\pi}} \left\| \left(n_{\log}^{(d)} \right)' \right\|_{L^\infty(\mathbb{R})}.$$

We thus only have to plug the above in equation (B.8), and then using the following equality which comes from (B.2), the change of variable $x = \log(y)$, and (5.5)–(5.9), to conclude the proof of the lemma

$$\begin{aligned} \left\| \left(n_{\log}^{(d)} \right)' \right\|_{L^\infty(\mathbb{R})} &= \sup_{y \in \mathbb{R}} \left| \exp(2y) \left(\hat{n}_0^{(d)} \right)'(\exp(y)) + \exp(y) \hat{n}_0^{(d)}(\exp(y)) \right| \\ &= \sup_{x \in \mathbb{R}_+} \left| x^2 \left(\hat{n}_0^{(d)} \right)'(x) + x \hat{n}_0^{(d)}(x) \right| = C_{\hat{n},d}. \end{aligned}$$

□

B.3. Proof of Proposition 5.6

Our objective here is to apply Lemma B.2. To do so, we begin by computing the probability that $\mathcal{K}(\bar{n}_{\log}^{(2k)} - n_{\log}^{(2k)}) \leq \varepsilon_p$, where $\varepsilon_p = \left(\frac{\log\left(\frac{2}{p}\right)}{2n_s}\right)^{\frac{1}{4k}}$. We denote for all $y \in \mathbb{R}$

$$R(y) := \left[1 - \frac{1}{n_s} \sum_{i=1}^{n_s} 1_{\left\{\log\left(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i}\right) \leq y\right\}}\right]^{\frac{1}{2k}} - \int_y^{+\infty} n_{\log}^{(2k)}(s) \, ds, \quad (\text{B.9})$$

and for all $t \geq 0$: $N_{\partial}(t) = \int_t^{\infty} n_{\partial}^{(2k)}(s) \, ds$. In view of (B.2), the change of variable $s' = \exp(s)$, and the fact that $\hat{n}_0^{(2k)}$ is the derivative of $x \mapsto -(N_{\partial})^{\frac{1}{2k}}\left(\frac{2x}{b\tilde{m}_1}\right)$ (see (2.10)), we have for all $y \in \mathbb{R}$

$$\int_y^{+\infty} n_{\log}^{(2k)}(s) \, ds = \int_{\exp(y)}^{+\infty} \hat{n}_0^{(2k)}(s') \, ds' = N_{\partial}\left(\frac{2e^y}{b\tilde{m}_1}\right)^{\frac{1}{2k}}. \quad (\text{B.10})$$

Our aim is to use the above equality to bound the probability that $\sup_{y \in \mathbb{R}} |R(y)| > \varepsilon_p$. To do so, we first use that $|a^{\frac{1}{2k}} - c^{\frac{1}{2k}}| \leq |a - c|^{\frac{1}{2k}}$ for all $a, c \geq 0$ to bound from above $|R|$. This inequality is a consequence of the reverse triangle inequality that holds for the quasi-distance $D(a, c) := |a - c|^{\frac{1}{2k}}$, where $(a, c) \in \mathbb{R}^2$. Then, we apply (B.10) to simplify the expression in the measure. Finally, we use the Dvoretzky–Kiefer–Wolfowitz–Massart inequality [14, 25], in view of the fact that $x \mapsto 1 - N_{\partial}\left(\frac{2e^x}{b\tilde{m}_1}\right)$ is the cumulative distribution function of the random variables $(\log(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i}))_{1 \leq i \leq n_s}$ (as $1 - N_{\partial}$ is the one of $(T_{2k,i})_{1 \leq i \leq n_s}$). We obtain

$$\begin{aligned} \mathbb{P}\left[\sup_{y \in \mathbb{R}} |R(y)| > \varepsilon_p\right] &\leq \mathbb{P}\left[\sup_{y \in \mathbb{R}} \left|1 - \frac{1}{n_s} \sum_{i=1}^{n_s} 1_{\left\{\log\left(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i}\right) \leq y\right\}} - \left[\int_y^{+\infty} n_{\log}^{(2k)}(s) \, ds\right]^{2k}\right|^{\frac{1}{2k}} > \varepsilon_p\right] \\ &= \mathbb{P}\left[\sup_{y \in \mathbb{R}} \left| \left[1 - N_{\partial}\left(\frac{2e^y}{b\tilde{m}_1}\right)\right] - \frac{1}{n_s} \sum_{i=1}^{n_s} 1_{\left\{\log\left(\frac{\tilde{b}\tilde{m}_1}{2} T_{2k,i}\right) \leq y\right\}} \right| > (\varepsilon_p)^{2k}\right] \\ &\leq 2e^{-2n_s(\varepsilon_p)^{4k}}. \end{aligned} \quad (\text{B.11})$$

In addition, in view of (B.9), (B.3), and the reasoning that allowed us to obtain Proposition 5.5, we have that $R' = n_{\log}^{(2k)} - \bar{n}_{\log}^{(2k)}$. This yields, in view of the fact that $R \in L^{\infty}(\mathbb{R})$ and the infimum in (B.1), that it holds $\mathcal{K}(n_{\log}^{(2k)} - \bar{n}_{\log}^{(2k)}) \leq \|R\|_{L^{\infty}(\mathbb{R})}$. Then, by combining this inequality with (B.11), we obtain that

$$\mathbb{P}\left[\mathcal{K}\left(n_{\log}^{(2k)} - \bar{n}_{\log}^{(2k)}\right) \leq \varepsilon_p\right] \geq \mathbb{P}\left[\sup_{y \in \mathbb{R}} |R(y)| \leq \varepsilon_p\right] \geq 1 - 2e^{-2n_s(\varepsilon_p)^{4k}} = 1 - p.$$

We thus only have to combine the above with Lemma B.2 for $\varepsilon = \varepsilon_p$ and $\alpha = \alpha_p$ to conclude that (5.10) is true with probability at least $1 - p$ (we notice that α_p is of the form given in Rem. B.3, so is optimal). \square