NONPARAMETRIC ESTIMATION OF THE DIVISION RATE OF AN AGE DEPENDENT BRANCHING PROCESS

Marc Hoffmann, Adélaïde Olivier

NONPARAMETRIC ESTIMATION OF THE DIVISION RATE OF AN AGE DEPENDENT BRANCHING PROCESS

MARC HOFFMANN AND ADÉLAÏDE OLIVIER

ABSTRACT. We study the nonparametric estimation of the branching rate $B(x)$ of a supercritical Bellman-Harris population: a particle with age x has a random lifetime governed by $B(x)$; at its death time, it gives rise to $k \geq 2$ offsprings with lifetime governed by the same division rate and so on. We observe continuously the process over a large time interval $[0,T]$; the data are stochastically dependent and one has to face simultaneously censoring, bias selection and non-ancillarity of the number of observations. In this setting, we construct a kernel-based estimator of $B(x)$ that achieves the rate of convergence $\exp(-\lambda_B \frac{\beta}{2\beta+1}T)$, where λ_B is the Malthus parameter and $\beta > 0$ is the smoothness of the function $B(x)$ in a viscinity of x. We prove that this rate is optimal in a minimax sense and we relate it explicitly to classical nonparametric models such as density estimation observed on an appropriate (parameter dependent) scale. We also shed some light on the fact that estimation with kernel estimators based on data alive at time T only is not sufficient to obtain optimal rates of convergence, a phenomenon which is specific to nonparametric estimation and that has been observed in other related growth-fragmentation models.

Mathematics Subject Classification (2010): 35A05, 35B40, 45C05, 45K05, 82D60, 92D25, 62G05, 62G20.

Keywords: Growth-fragmentation, cell division, nonparametric estimation, bias selection, minimax rates of convergence, Bellman-Harris processes.

1. Introduction

1.1. **Motivation.** Structured models have been paid particular attention over the last few years, both from a probabilistic and an applied analysis angle, in particular with a view toward a better understanding of population evolution in mathematical biology (see for instance the textbook by Perthame [19] and the references therein). In this context, a more specific focus and need for statistical methods has emerged recently (*e.g.* Doumic *et al.* [9, 8, 7] and the references therein) and this is the topic of the present paper. If x denotes a so-called structuring variable – for instance age, size, any measure of variability or DNA content of a cell or bacteria, and if $n(t,x)$ denotes the number or density of cells at time t of a population starting from a single ancestor at time $t=0$, a sound mathematical model can be obtained by specifying an evolution equation for $n(t,x)$.

Consider for instance the paradigmatic problem of age-dependent cell division, where the evolution of $n(t,x)$ is given by the simplest transport-fragmentation equation

$$(1) \qquad \begin{cases} \dfrac{\partial}{\partial t}n(t,x) + \dfrac{\partial}{\partial x}n(t,x) + B(x)n(t,x) = 0 \\[2ex] n(t,0) = m\int_0^\infty B(y)n(t,y)dy, \quad t>0, \quad n(0,x) = \delta_0, \end{cases}$$

where δ_0 denotes the Dirac mass at point 0. In this model, each cell dies according to a division rate $x \rightsquigarrow B(x)$ that depends on its age x only (a living cell of age x has probability $B(x)dx$ of dying in the interval $[x, x+dx]$) and it gives rise to $m \geq 2$ offsprings at its time of death. The parameters (m, B) specify the so-called age-dependent model.

In this seemingly simple context, we wish to draw statistical inference on the division rate function $x \rightsquigarrow B(x)$ and in m in the most rigorous way, when we observe the evolution of the population through time and when the shape of the function B can be arbitrary, to within a prescribed smoothness class, *i.e.* in a nonparametric setting. In order to do so, we transfer the deterministic description (1) into a probabilist model that consists of a system of (non-interacting) particles specified by a probability distribution p on the integers (the offspring distribution) and a probability density f on $[0, \infty)$. A particle has a random lifetime drawn according to $f(x)dx$; at the time of its death, it gives rise to k offsprings with probability p_k (with $p_0 = p_1 = 0$), each offspring having independent lifetimes distributed as $f(x)dx$, and so on. The resulting process is a classical supercritical Bellman-Harris, see for instance the textbooks of Harris [11] or Athreya and Ney [2]. It is described by a piecewise deterministic Markov process

$$X(t) = \big(X_1(t), X_2(t), \dots\big), \quad t \geq 0, \tag{2}$$

with values in $\bigcup_{k\geq 1}[0,\infty)^k$, where the $X_i(t)$'s denote the (ordered) ages of the living particles at time t. The formal link between $X(t)$ and $n(t,x)$ is obtained via $n(t,x) = \mathbb{E}\big[\sum_{i=1}^{\infty} \delta_{X_i(t)=x}\big]$ which has to be understood in a weak (measure) sense, *i.e.* the empirical measure (in expectation) of the particle system solves Equation (1).

The correspondence between (m, B) and (f, p) is given by

$$B(x) = \frac{f(x)}{1 - \int_0^x f(s)ds}, \ x \in [0, \infty), \ \text{ and } \ m = \sum_{k \geq 2} k p_k, \tag{3}$$

provided everything is well defined. Under fairly reasonable assumptions described below, it is one-to-one between B and f, but not between m and p. We are interested in the nonparametric estimation of $x \rightsquigarrow B(x)$, which is nothing but the hazard rate function of the lifetime density f of each particle, and also in the mean offspring m, the whole distribution p being considered as a nuisance parameter.

1.2. **Objectives and results.**

Observation schemes. We assume we observe the whole trajectory $(X(t), t \in [0,T])$, where $T > 0$ is a fixed (large) terminal time. Asymptotics are taken as $T \to \infty$. If we denote by $\mathcal{T}_T$ the population of individuals that are born before T and observed up to time T and if $(\zeta_u^T, u \in \mathcal{T}_T)$ denotes the values of the ages of the different individuals of $\mathcal{T}_T$ (at their time of death or at time T), we wish to draw inference on $B(x)$ based on

$$\big\{X(t), t \in [0,T]\big\} = \big\{\zeta_u^T, u \in \mathcal{T}_T\big\}.$$

Although the lifetimes of the individuals are independent (and identically distributed) with common density f, this is no longer the case for the population $(\zeta_u^T, u \in \mathcal{T}_T)$ considered as a whole: the tree structure plays a crucial role and we have to face several non-trivial difficulties:

1) *Bias selection:* particles with small lifetimes are more often observed than particles with large lifetimes since the observation of the process is stopped along all the branches at the fixed time T, as illustrated in Figure 1.
2) *Censoring:* if $\partial\mathcal{T}_T \subset \mathcal{T}_T$ denotes the population of individuals alive at time T (in red in Figure 1), they are censored in our observation scheme (we observe their lifetime only up to time T) but contribute to the whole estimation process at the same level as the population

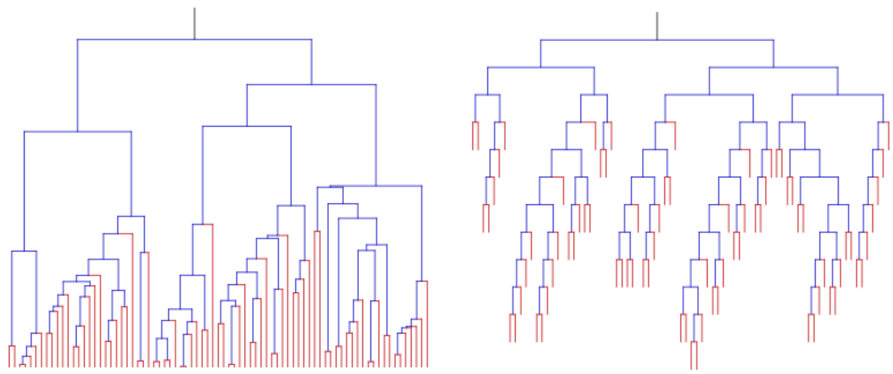

FIGURE 1. *The effect of bias selection. Simulation of a binary ($p_2 = 1$ so $m = 2$) age-dependent tree with B given in Section 4, up to time $T = 8$ ($|\mathcal{T}_T| = 145$). Left: the size of each segment represents the lifetime of an individual. Individuals alive at time T are represented in red. Right: genealogical representation of the same realisation of the tree. We see how*

$\mathring{\mathcal{T}}_T \subset \mathcal{T}_T$ of individuals born and dead before T: due to the supercriticality of the process ($m > 1$) we have $|\mathcal{T}_T| \approx |\mathring{\mathcal{T}}_T| \approx |\partial\mathcal{T}_T|$ as T grows to infinity, and this affects the statistical analysis, see Section 2.2 below.

3) *Non-ancillarity:* the number of observations $|\mathcal{T}_T|$ that govern the amount of statistical information is random and depends on B: we essentially have less observations if B is small (particles split at a slow rate) than if B is large (particles split at a fast rate). Moreover $|\mathcal{T}_T|$ is not ancillary (its distribution depends on the unknown parameter) so we cannot condition upon $|\mathcal{T}_T|$, according to Fisher's ancillary principle [16].

Main results. We first study in Section 2 the behaviour of empirical measures of the form

$$\mathcal{E}^T(\mathcal{V}, g) = |\mathcal{V}|^{-1} \sum_{u \in \mathcal{V}} g(\zeta_u^T), \quad \text{with } \mathcal{V} = \mathring{\mathcal{T}}_T \text{ or } \partial\mathcal{T}_T$$

for suitable test functions g. From the classical study of branching processes, it is known that $|\mathring{\mathcal{T}}_T| \approx |\partial\mathcal{T}_T| \approx e^{\lambda_B T}$, where $\lambda_B > 0$ is the Malthus parameter associated to the model (Harris [11] and (7) below). Both $\mathcal{E}^T(\mathring{\mathcal{T}}_T, g)$ and $\mathcal{E}^T(\partial\mathcal{T}_T, g)$ converge to their respective limits with rate $\exp(-\lambda_B T/2)$, with some uniformity in B and g as shown in Theorem 4 and 5 below. For the proof, we heavily rely on the recent studies of Cloez [5] and Bansaye *et al.* [3], two key references for this paper, adjusting the tools developed in [3] to the non-Markovian case: the essential ingredient is the use of many-to-one formulae that reduce the problem to studying the evolution of a particle picked at random along the genealogical tree (Propositions 11 and 12). The rate of convergence to equilibrium of this tagged particle, which governs the rates of convergence for statistical estimators, is obtained by a simple coupling argument (Proposition 13).

These preliminary results enable us to address the main issue of the paper: we construct in Section 3 a nonparametric estimator $\widehat{B}_T(x)$ of $B(x)$ that achieves the rate of convergence $\exp(-\lambda_B \frac{\beta}{2\beta+1} T)$ for pointwise error and uniformly over functions B with local smoothness of order $\beta > 0$ (Theorem 9). We show that this rate is optimal in a minimax sense in Theorem 10, thanks to statistical tools developed in Löcherbach [17]. We bypass the aformentioned bias selection difficulty 1) by weighting a kernel estimator by a de-biasing factor that depends on preliminary estimators of λ_B and m. These estimators (essentially) converge with rate $\exp(-\lambda_B T/2)$ as shown in Proposition 7. As for the censoring part 2), we base our nonparametric kernel estimator on $\mathcal{E}^T(\mathring{\mathcal{T}}_T, g)$ and *not on* $\mathcal{E}^T(\partial\mathcal{T}_T, g)$, since that latter quantity would lead to a subobtimal rate of convergence as discussed in Section 3.3. Finally, the non-ancillarity issue 3) is solved by specifying a random bandwidth for the kernel that also depends on the preliminary estimation of λ_B. This last point requires extra efforts in order to show a form of stability that is detailed in Proposition 18.

The statistical study of branching processes goes back to Athreya and Keiding [1] for deriving maximum likelihood theory in the case of a parametric (constant) division rate, relying on the fact that the number of living cells is then a Markov process, a property we lose here for a non-constant division rate $x \rightsquigarrow B(x)$. The textbook of Guttorp [10] gives an account of existing parametric methods in the 1990's. In the early 2000's the regularity in the sense of the LAN and LAMN property was established in the comprehensive study of Löcherbach [17, 18], see also Hyrien [14] for statistical computational methods and Johnson *et al.* [15] for Bayesian analysis, and Delmas and Marsalle [6] in discrete time. In nonparametric estimation, only few results exist; we mention the case when dynamics between jumps is driven by a diffusion in Höpfner *et al.* [13]. To the best or our knowledge, our study provides with the first fully nonparametric approch in continuous time in supercritical branching processes which are piecewise deterministic. Admittedly, the Bellman-Harris model is a toy model for the study of population dynamics, but we believe that the present contribution sheds some light in the intrinsinc difficulties that need to be solved in more elaborate models like cell equation for which only simplified statistical models have been considered so far (in discrete time or under additional deterministic or stochastic noise like in *e.g.* [9, 8, 7]).

Organisation of the paper. In Section 2, we define our rigorous statistical framework by means of continuous time rooted trees (Section 2.1) and study the convergence properties of the biased empirical measures $\mathcal{E}^T(\mathring{\mathcal{T}}_T, g)$ and $\mathcal{E}^T(\partial\mathcal{T}_T, g)$ in Section 2.3. We start by deriving heuristically the respective limits of the empirical measures in Section 2.2 (that can also be found in Cloez [5] and Bansaye *et al.* [3]) in order to shed some light on the specific methods of proof in the subsequent study of rate of convergence. We construct in Section 3 the estimators of m, λ_B and $B(x)$ and state our statistical results together with a discussion on the extensions and limitations of our findings. Section 4 tackles the problem of numerical implementation on simulated data, advocating for a reasonably use of our estimators in practice. Section 5 is devoted to the proofs. An appendix (Section 6) contains auxiliary useful results.

2. Rate of convergence for biased empirical measures

2.1. **Continuous time rooted trees.** It will prove more convenient to work with a representation of $(X(t))_{t\geq 0}$ in terms of a continuous time rooted tree. We need some notation and closely follow Bansaye *et al.* [3]. Let

$$\mathcal{U} = \bigcup_{k\geq 0} (\mathbb{N}^\star)^k$$

with $\mathbb{N}^\star = \{1, 2, \ldots\}$ and $(\mathbb{N}^\star)^0 = \{\varnothing\}$ denote the infinite genealogical tree. We use throughout the following standard notation: for $u = (u_1, u_2, \ldots, u_m)$ and $v = (v_1, \ldots, v_n)$ in $\mathcal{U}$, we write $uv = (u_1, \ldots, u_m, v_1, \ldots, v_n)$ for the concatenation, we identify $\varnothing u$, $u\varnothing$ and u, we write $u \preceq v$ if there exists w such that $uw = v$ and $u \prec v$ if $u \preceq v$ and $w \neq \varnothing$.

Given a family $(\nu_u, u \in \mathcal{U})$ of integers representing the number of offsprings of the individuals $u \in \mathcal{U}$, we construct an ordered rooted tree $\mathcal{T} \subset \mathcal{U}$ as follows:

i) $\varnothing \in \mathcal{T}$,
ii) If $v \in \mathcal{T}$, $u \preceq v$ implies $u \in \mathcal{T}$,
iii) For every $u \in \mathcal{T}$, we have $uj \in \mathcal{T}$ if and only if $1 \leq j \leq \nu_u$.

For a family $(\zeta_u, u \in \mathcal{U})$ of nonnegative numbers representing the lifetimes of the individuals $u \in \mathcal{U}$, we set

$$b_u = \sum_{v \prec u} \zeta_v \text{ and } d_u = b_u + \zeta_u \tag{4}$$

for the times of birth and death of the individual $u \in \mathcal{U}$. Let $\mathbb{U} = \mathcal{U} \times [0, \infty)$. A continuous time rooted tree is then a subset $\mathbb{T}$ of $\mathbb{U}$ such that

(i) $(\varnothing, 0) \in \mathbb{T}$,
(ii) The projection $\mathcal{T}$ of $\mathbb{T}$ on $\mathcal{U}$ is an ordered rooted tree,
(iii) There exists a family $(\zeta_u, u \in \mathcal{U})$ of nonnegative numbers such that $(u, s) \in \mathbb{T}$ if and only if $b_u \leq s < d_u$, where (b_u, d_u) are defined by (4).

We now work on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this setting, we have the following

Definition 1 (The Bellman-Harris model). *A random continuous time rooted tree is a Bellman-Harris model with offspring distribution $p = (p_k)_{k \geq 1}$ and division rate $B : [0, \infty) \to [0, \infty)$ if*

(i) *The family of the number of offsprings $(\nu_u, u \in \mathcal{U})$ are independent random variables with common distribution p.*
(ii) *The family of lifetimes $(\zeta_u, u \in \mathcal{U})$ are independent random variables such that*

$$\mathbb{P}\big(\zeta_u \in [x, x + dx) \,\big|\, \zeta_u \geq x\big) = B(x)dx \tag{5}$$

with

$$\int^{\infty} B(x)dx = \infty,$$

(iii) *The families of random variables $(\nu_u, u \in \mathcal{U})$ and $(\zeta_u, u \in \mathcal{U})$ are independent.*

Going back to the process $(X(t))_{t \geq 0}$ defined in (2),we have an identity between point measures on $(0, \infty)$ that reads

$$\sum_{i \geq 1} \mathbf{1}_{\{X_i(t) > 0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{T}} \mathbf{1}_{\{t \in [b_u, d_u)\}} \delta_{t - b_u}.$$

The following assumption will be in force in the paper:

Assumption 2. *The offspring distribution $p = (p_k)_{k \geq 0}$ satisfies*

$$p_0 = p_1 = 0, \;\; 2 \leq m = \sum_{k \geq 2} k p_k < \infty, \quad \sum_{k \geq 2} k^2 p_k < \infty \;\; \textit{and} \;\; \bar{m} = \sum_{i \neq j} \sum_{k \geq i \vee j} p_k < \infty.$$

2.2. **First estimates.** In order to extract information about $x \rightsquigarrow B(x)$, we consider the empirical distribution function over the lifetimes indexed by some $\mathcal{V} \subset \mathcal{T}_T$ for a test function g, that is

$$\mathcal{E}^T(\mathcal{V}, g) = |\mathcal{V}|^{-1} \sum_{u \in \mathcal{V}} g(\zeta_u^T),$$

and expect a law of large number as $T \to \infty$. Without much of a surprise, it turns out that depending whether $\zeta_u^T = \zeta_u$ or not, *i.e.* if the data are still alive at time T, therefore censored or not, we have a different limit. More precisely, define

$$\mathring{\mathcal{T}}_T = \{u \in \mathcal{T}, b_u < T \text{ and } d_u \leq T\} \text{ and } \partial\mathcal{T}_T = \{u \in \mathcal{T}, b_u \leq T < d_u\},$$

i.e. the set of particles that are born and that die before T, and the set of particles alive at time T, so that

$$\mathcal{T}_T = \mathring{\mathcal{T}}_T \cup \partial \mathcal{T}_T.$$

Information from $\mathcal{E}^T(\partial \mathcal{T}_T, g)$. Heuristically, we postulate for large T the approximation

$$\mathcal{E}^T(\partial \mathcal{T}_T, g) \sim \frac{1}{\mathbb{E}[|\partial \mathcal{T}_T|]} \mathbb{E}\Big[\sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T)\Big].$$

Then, a classical result based on renewal theory gives the estimate

$$\mathbb{E}\big[|\partial \mathcal{T}_T|\big] \sim \kappa_B e^{\lambda_B T}, \tag{6}$$

where $\lambda_B > 0$ is the Malthus parameter of the model, defined as the unique solution to

$$\int_0^\infty B(x) e^{-\lambda_B x - \int_0^x B(u)du} dx = \frac{1}{m}, \tag{7}$$

and $\kappa_B > 0$ is an explicitly computable constant (that also depends on m, see [11] and also Lemma 14 below). As for the numerator call χ_t, the size of a particle at time t along a branch of the tree picked at random. The process $(\chi_t)_{t \geq 0}$ is Markov process with values in $[0, \infty)$ with infinitesimal generator

$$\mathcal{A}_B g(x) = g'(x) + B(x)\big(g(0) - g(x)\big) \tag{8}$$

densely defined on bounded continuous functions. Assume temporarily (for simplicity) that each cell $u \in \mathcal{U}$ has exactly m offsprings at each division. It is then relatively straightforward to obtain the identity

$$\mathbb{E}\Big[\sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T)\Big] = \mathbb{E}\big[m^{N_T} g(\chi_T)\big], \tag{9}$$

where $N_t = \sum_{s \leq t} \mathbf{1}_{\{\chi_s - \chi_{s_-} > 0\}}$ is the counting process associated to $(\chi_t, t \geq 0)$, see Proposition 11 below for a general setting. Putting together (6) and (9), we thus expect

$$\mathcal{E}^T(\partial \mathcal{T}_T, g) \sim \kappa_B^{-1} e^{-\lambda_B T} \mathbb{E}\big[m^{N_T} g(\chi_T)\big],$$

and we anticipate that the term $e^{-\lambda_B T}$ should somehow be compensated by the term m^{N_T} within the expectation. To that end, following Cloez [5] (and also in Bansaye *et al.* [3] when B is constant) one introduces an auxiliary "biased" Markov process $(\widetilde{\chi}_t)_{t \geq 0}$, with generator $\mathcal{A}_{H_B}$ for a biasing function $H_B(x)$ characterised by

$$f_{H_B}(x) = m e^{-\lambda_B x} f_B(x),\ x \geq 0, \tag{10}$$

and where

$$f_B(x) = B(x) \exp\big(-\int_0^x B(y)dy\big)$$

denotes the density associated to the division rate B, as follows from (3) or (5) . This choice (and this choice only, see Proposition 11 below) enables us to obtain

$$e^{-\lambda_B T} \mathbb{E}\big[m^{N_T} g(\chi_T)\big] = m^{-1} \mathbb{E}\Big[g(\widetilde{\chi}_T) B(\widetilde{\chi}_T)^{-1} H_B(\widetilde{\chi}_T)\Big] \tag{11}$$

with $\widetilde{\chi}_0 = 0$ under $\mathbb{P}$. Moreover $(\widetilde{\chi}_t)_{t\geq 0}$ is geometrically ergodic, with invariant probability $c_B \exp(-\int_0^x H_B(y)dy)dx$. We further anticipate

$$\mathbb{E}\Big[g(\widetilde{\chi}_T)B(\widetilde{\chi}_T)^{-1}H_B(\widetilde{\chi}_T)\Big] \sim c_B \int_0^\infty g(x)B(x)^{-1}H_B(x)e^{-\int_0^x H_B(y)dy}dx$$
$$= mc_B \int_0^\infty g(x)e^{-\lambda_B x}B(x)^{-1}f_B(x)dx$$

assuming everthing is well-defined, since $H_B(x)\exp(-\int_0^x H_B(y)dy) = f_{H_B}(x) = me^{-\lambda_B}f_B(x)$ by (10). Finally, we have $\kappa_B^{-1}c_B = \lambda_B \frac{m}{m-1}$ by Lemma 14 below which enables us to conclude

$$\mathcal{E}^T(\partial\mathcal{T}_T, g) \sim \partial\mathcal{E}(g) := \lambda_B \frac{m}{m-1}\int_0^\infty g(x)e^{-\lambda_B x}e^{-\int_0^x B(y)dy}dx. \tag{12}$$

Unfortunately, the statistical information extracted from $\mathcal{E}^T(\partial\mathcal{T}_T, g)$ does not enable us to obtain optimal rates of convergence, since the form of $\partial\mathcal{E}(g)$ involves an antiderivative of B leading to so-called ill-posedness. This is discussed at length in Section 3.3 below. We thus investigate in a second step the statistical information we can get from $\mathring{\mathcal{T}}_T$.

Information from $\mathcal{E}(\mathring{\mathcal{T}}_T, g)$. The situation is a bit different if we allow for data in $\mathring{\mathcal{T}}_T$. Note first that $\zeta_u^T = \zeta_u$ on the event $u \in \mathring{\mathcal{T}}_T$. We also have in that case a many-to-one formula that now reads

$$\mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u^T)\Big] = \mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u)\Big] = m^{-1}\int_0^T e^{\lambda_B s}\mathbb{E}\big[g(\widetilde{\chi}_s)H_B(\widetilde{\chi}_s)\big]ds, \tag{13}$$

where $(\widetilde{\chi}_t)_{t\geq 0}$ is the auxiliary one-dimensional auxiliary Markov process with generator $\mathcal{A}_{H_B}$, see (8), where H_B is characterised by (10) above. Assuming again ergodicity, we approximate the right-hand side of (13) and obtain

$$\mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u)\Big] \sim c_B m^{-1}\frac{e^{\lambda_B T}}{\lambda_B}\int_0^\infty g(x)H_B(x)e^{-\int_0^x H_B(u)du}dx$$
$$= c_B\frac{e^{\lambda_B T}}{\lambda_B}\int_0^\infty g(x)e^{-\lambda_B x}f_B(x)dx.$$

since $H_B(x)\exp(-\int_0^x H_B(y)dy) = f_{H_B}(x) = me^{-\lambda_B x}f_B(x)$ by (10) We again have an approximation of the type (6) with another constant κ_B', see Lemma 15 and we eventually expect

$$\mathcal{E}^T(\mathring{\mathcal{T}}_T, g) \sim \mathring{\mathcal{E}}(g) = \frac{c_B}{\lambda_B\kappa_B'}\int_0^\infty g(x)e^{-\lambda_B x}f_B(x)dx = m\int_0^\infty g(x)e^{-\lambda_B x}f_B(x)dx$$

as $T\to\infty$, where the last equality stems from the identity $c_B = \lambda_B\kappa_B' m$ that can be readily derived by picking $g = 1$ and using (10) together with the fact that f_{H_B} is a density function.

2.3. **Convergence results for biased empirical measures.** The parameter m (the mean of the offspring distrubution) is fixed once for all. For $b, C > 0$, we introduce the sets

$$\mathcal{L}_C = \Big\{g : [0,\infty) \to \mathbb{R},\ \sup_x |g(x)| \leq C\Big\},$$
$$\mathbb{B}_{b,C} = \Big\{B : [0,\infty) \to [0,\infty), \forall x \geq 0 :\ b \leq B(x) \leq b\max\{C, 1\}\Big\},$$
$$\mathcal{B}_b = \Big\{B \in \mathbb{B}_{b,m/(m-1)},\ B \text{ differentiable and } \forall x \geq 0 : B'(x) \leq B(x)^2\Big\},$$

so that the following inclusions hold

$$\mathcal{B}_b \subset \mathbb{B}_{b,m/(m-1)} \subset \mathcal{L}_{bm/(m-1)}.$$

Definition 3. *For a family $\Upsilon_T = \Upsilon_T(\gamma), t \geq 0$ of real-valued random variables, depending on some parameter γ , with $\gamma \in \mathcal{G}$ (or with distribution depending on γ), we say that Υ_T is $\mathcal{G}$-tight for the parameter γ if*

$$\limsup_{T\to\infty} \sup_{\gamma\in\mathcal{G}} \mathbb{P}\big(|\Upsilon_T(\gamma)| \geq K\big) \to 0 \;\; as \;\; K \to \infty.$$

Theorem 4 (Rate of convergence for particles living at time T)**.** *Work under Assumption 2. For every $b, C > 0$,*

$$e^{\lambda_B T/2}\big(\mathcal{E}^T\big(\partial\mathcal{T}_T, g\big) - \partial\mathcal{E}(g)\big)$$

is $\big(\mathcal{B}_b, \mathcal{L}_C\big)$-tight for the parameter (B, g).

Theorem 5 (Rate of convergence for particles dying before T)**.** *In the same setting as Theorem 4,*

$$e^{\lambda_B T/2}\big(\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, g\big) - \mathring{\mathcal{E}}(g)\big)$$

is $\big(\mathcal{B}_b, \mathcal{L}_C\big)$-tight for the parameter (B, g).

Some comments are in order:

About the rate of convergence. First, since $|\mathring{\mathcal{T}}_T|$ and $|\partial\mathcal{T}_T|$ are of the order $e^{\lambda_B T}$, we obtain a standard rate of convergence. Also, we could replace $e^{\lambda_B T/2}$ by the random normalisations $|\mathring{\mathcal{T}}_T|^{1/2}$ and $|\partial\mathcal{T}_T|^{1/2}$ in Theorems 4 and 5 respectively, as stems from the well known estimates

$$e^{-\lambda_B T}|\mathring{\mathcal{T}}_T| \to Z_B \;\text{ and }\; e^{-\lambda_B T}|\partial\mathcal{T}_T| \to \widetilde{Z}_B \;\text{ in }\; L^2(\mathbb{P}) \;\text{ as }\; T \to \infty,$$

with $\mathbb{P}(Z_B > 0) = \mathbb{P}(\widetilde{Z}_B > 0) = 1$ together with $\mathrm{Var}(Z_B) > 0$ and $\mathrm{Var}(\widetilde{Z}_B) > 0$, see Lemmas 14 and 15 below.

About the tightness. What we need in order to handle the random normalisation is actually the convergence of $e^{\lambda_B T}|\mathring{\mathcal{T}}_T|^{-1}\mathbf{1}_{\{|\mathring{\mathcal{T}}_T|\neq 0\}}$ and $e^{\lambda_B T}|\partial\mathcal{T}_T|^{-1}$. This convergence still holds in probability but not necessarily in $L^2(\mathbb{P})$, so we only have tightness in Theorems 4 and 5. However, if we replace $\mathcal{E}^T(\mathring{\mathcal{T}}_T, g)$ by

$$\frac{1}{\mathbb{E}[|\mathring{\mathcal{T}}_T|]} \sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u^T),$$

then we have a bound in $L^2(\mathbb{P})$ together with a control on g, see Proposition 16 below. Such a finer control is mandatory for the subsequent statistical analysis, since we need to pick a function g that depends on T and that mimicks the behaviour of the Dirac mass δ_x, see Section 3 below.

About the class $\mathcal{B}_b$. We are able to have a uniform result for smooth division rates $x \rightsquigarrow B(x)$ that are bounded below and above over the range $[b, bm/(m-1)]$ for every $b > 0$. The maximal range is obtained for $m = 2$ and gives $[b, 2b]$. This seemingly frustrating result stems from our method of proof of Theorems 4 and 5. In order to control the decorrelation between $g(\zeta_u)$ and $g(\zeta_v)$ for $u, v \in \mathcal{U}$, transferring the estimates via many-to-one formulae like in (9) and (11), we need to control the rate of convergence to equilibrium of the Markov process $(\widetilde{\chi}_t)_{t\geq 0}$ that appears in (11): we have, for every $x \in (0,\infty)$,

$$\Big|P_{H_B}^t g(x) - \int_0^\infty g(y)\mu_B(y)dy\Big| \leq 4\sup_y |g(y)| \exp\big(-\rho_B t\big),$$

where $(P^t_{H_B})_{t\geq 0}$ is the semigroup associated to $(\widetilde{\chi}_t)_{t\geq 0}$ and μ_B its invariant probability, see Proposition 13 below, where $\rho_B = \inf H_B(x)$. The class $\mathcal{B}_b \subset \mathbb{B}_{b,m/(m-1)}$ is then constructed in order to guarantee $\lambda_B \leq \rho_B$ for $B \in \mathcal{B}_b$, an essential condition in order to obtain Proposition 16 below, a key result for the subsequent statistical results. A careful glance at the proof of Theorems 4 and 5 show that we still have a rate without this assumption. For $B \in \mathbb{B}_{b,C}$, define

$$v_T(B) = \begin{cases} e^{-\min\{\lambda_B, 2\rho_B\}T/2} & \text{if} \quad \lambda_B \neq 2\rho_B \\ T^{1/2}e^{-\lambda_B T/2} & \text{otherwise.} \end{cases}$$

Corollary 6. *Work under Assumption 2. For every $b, C, C' > 0$, we have that*

$$v_T(B)^{-1}\big(\mathcal{E}^T\big(\partial\mathcal{T}_T, g\big) - \partial\mathcal{E}(g)\big)$$

and

$$v_T(B)^{-1}\big(\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, g\big) - \mathring{\mathcal{E}}(g)\big)$$

are $\big(\mathbb{B}_{b,C}, \mathcal{L}_{C'}\big)$-tight for the parameter (B, g).

The restriction $B \in \mathbb{B}_{b,C}$ could presumably also be relaxed further, but at the cost of loosing the uniformity in B.

3. Statistical estimation

3.1. **Construction of an estimation procedure.**

Estimation of m and λ_B. To a particle sitting at node $u \in \mathring{\mathcal{T}}_T$, we associate its number of offsprings ν_u. Note that the knowledge of $\mathcal{T}_T$ enables us to reconstruct ν_u for every $u \in \mathring{\mathcal{T}}_T$. This enables us to define an estimator for m by setting

$$\widehat{m}_T = |\mathring{\mathcal{T}}_T|^{-1} \sum_{u \in \mathring{\mathcal{T}}_T} \nu_u \tag{14}$$

on the set $|\mathring{\mathcal{T}}_T| \neq 0$ and 2 otherwise. In order to estimate λ_B, we first observe that for $\mathrm{Id}(x) = x$, we can write

$$\begin{aligned}\mathring{\mathcal{E}}(\mathrm{Id}) &= m\int_0^\infty x\big(B(x)+\lambda_B\big)e^{-\int_0^x (B(y)+\lambda_B)dy}dx - m\lambda_B \int_0^\infty xe^{-\lambda_B x}e^{-\int_0^x B(y)dy}dx \\ &= m\int_0^\infty e^{-\int_0^x (B(y)+\lambda_B)dy}dx - m\lambda_B \tfrac{m-1}{m\lambda_B}\partial\mathcal{E}(\mathrm{Id}) = m\tfrac{m-1}{m\lambda_B} - (m-1)\partial\mathcal{E}(\mathrm{Id}),\end{aligned}$$

integrating by part to obtain the last equality. So we obtain the following representation

$$\lambda_B = \Big(\tfrac{1}{m-1}\mathring{\mathcal{E}}(\mathrm{Id}) + \partial\mathcal{E}(\mathrm{Id})\Big)^{-1}$$

and this yields the estimator

$$\widehat{\lambda}_T = \Big(\tfrac{1}{\widehat{m}_T - 1}|\mathring{\mathcal{T}}_T|^{-1}\sum_{u \in \mathring{\mathcal{T}}_T}\zeta_u + |\partial\mathcal{T}_T|^{-1}\sum_{u \in \partial\mathcal{T}_T}\zeta_u^T\Big)^{-1}. \tag{15}$$

The following convergence result for $\widehat{\lambda}_T$ is then a consequence of Theorems 4 and 5.

Proposition 7. *In the same setting as Theorem 4,*

$$e^{\lambda_B T/2}\big(\widehat{m}_T - m\big) \;\; \text{and} \;\; T^{-1/2}e^{\lambda_B T/2}\big(\widehat{\lambda}_T - \lambda_B\big)$$

are $\mathcal{B}_b$-tight for the parameter B.

Reconstruction formula for $B(x)$. An *estimator* $\widehat{B}_T : [0,\infty) \to \mathbb{R}$ of B is a random function

$$\widehat{B}_T(x) = \widehat{B}_T\big(x, (X(t))_{t\in[0,T]}\big), \ \ x \in [0,\infty)$$

that is also measurable as a function of x. By (3), we have

$$B(x) = \frac{f_B(x)}{1 - \int_0^x f_B(y)dy}$$

and from the definition $\mathring{\mathcal{E}}(g) = m\int_0^\infty g(x)e^{-\lambda_B x} f_B(x)dx$ we obtain the formal reconstruction formula

$$B(x) = \frac{\mathring{\mathcal{E}}\big(m^{-1}e^{\lambda_B\cdot}\delta_x(\cdot)\big)}{1 - \mathring{\mathcal{E}}\big(m^{-1}e^{\lambda_B\cdot}\mathbf{1}_{\{\cdot\leq x\}}\big)} \tag{16}$$

where $\delta_x(\cdot)$ denotes the Dirac function at x. Therefore, substituting m and λ_B by the estimators defined in (14) and (15) and taking g as a weak approximation of δ_x, we obtain a strategy for estimating $B(x)$ replacing $\mathring{\mathcal{E}}(\cdot)$ by its empirical version $\mathcal{E}^T(\mathring{\mathcal{T}}_T, \cdot)$.

Construction of a kernel estimator and function spaces. Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel function. For $h > 0$, set $K_h(x) = h^{-1}K(h^{-1}x)$. In view of (16), we define the estimator

$$\widehat{B}_T(x) = \frac{\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, \widehat{m}_T^{-1}e^{\widehat{\lambda}_T\cdot}K_h(x-\cdot)\big)}{1 - \mathcal{E}^T\big(\mathring{\mathcal{T}}_T, \widehat{m}_T^{-1}e^{\widehat{\lambda}_T\cdot}\mathbf{1}_{\{\cdot\leq x\}}\big)}$$

on the set $\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, \widehat{m}_T^{-1}e^{\widehat{\lambda}_T\cdot}\mathbf{1}_{\{\cdot\leq x\}}\big) \neq 1$ and 0 otherwise. Thus $\widehat{B}_T(x)$ is specified by the choice of the kernel K and the bandwidth $h > 0$.

We need the following property on K:

Assumption 8. *The kernel $K : \mathbb{R} \to \mathbb{R}$ is differentiable with compact support and for some integer $n_0 \geq 1$, we have $\int_{-\infty}^{\infty} x^k K(x)dx = \mathbf{1}_{\{k=0\}}$ for $k = 1, \ldots, n_0$.*

Assumption 8 will enable us to have nice approximation results over smooth functions B, described in the following way: for a compact interval $\mathcal{D} \subset (0,\infty)$ and $\beta > 0$, with $\beta = \lfloor\beta\rfloor + \{\beta\}$, $0 < \{\beta\} \leq 1$ and $\lfloor\beta\rfloor$ an integer, let $\mathcal{H}^\beta_{\mathcal{D}}$ denote the Hölder space of functions $g : \mathcal{D} \to \mathbb{R}$ possessing a derivative of order $\lfloor\beta\rfloor$ that satisfies

$$|g^{\lfloor\beta\rfloor}(y) - g^{\lfloor\beta\rfloor}(x)| \leq c(g)|x-y|^{\{s\}}. \tag{17}$$

The minimal constant $c(g)$ such that (17) holds defines a semi-norm $|g|_{\mathcal{H}^s_{\mathcal{D}}}$. We equip the space $\mathcal{H}^s_{\mathcal{D}}$ with the norm $\|g\|_{\mathcal{H}^\beta_{\mathcal{D}}} = \sup_x |g(x)| + |g|_{\mathcal{H}^\beta_{\mathcal{D}}}$ and the balls

$$\mathcal{H}^\beta_{\mathcal{D}}(L) = \{g : \mathcal{D} \to \mathbb{R}, \ \|g\|_{\mathcal{H}^\beta_{\mathcal{D}}} \leq L\}, \ L > 0.$$

3.2. **Convergence results for $\widehat{B}_T(x)$.** We are ready to give our main result, which controls the rate of convergence of $\widehat{B}_T(x)$ for x restricted to a compact interval $\mathcal{D}$, uniformly over Hölder balls intersected with sets of the form $\mathcal{B}_b$.

Theorem 9 (Upper rate of convergence)**.** *Specify $\widehat{B}_T$ with a kernel satisfying Assumption 8 for some $n_0 > 1$ and*

$$h = \widehat{h}_T = \exp\big(-\tfrac{1}{2\beta+1}\widehat{\lambda}_T\big) \tag{18}$$

for some $\beta \in (1, n_0)$. For every $b > 0, L > 0$, every compact interval $\mathcal{D}$ in $(0, \infty)$ (with non-empty interior) and every $x \in \mathcal{D}$,

$$e^{\lambda_B \frac{\beta}{2\beta+1} T}\big(\widehat{B}_T(x) - B(x)\big)$$

is $\mathcal{B}_b \cap \mathcal{H}^\beta_{\mathcal{D}}(L)$-tight for the parameter B.

This rate is indeed optimal in the following minimax sense:

Theorem 10 (Lower rate of convergence)**.** *Let $\mathcal{D}$ be a compact interval in $(0, \infty)$. For every $x \in \mathcal{D}$ and every positive b, β, L, there exists $C > 0$ such that*

$$\liminf_{T \to \infty} \inf_{\widehat{B}_T} \sup_{B} \mathbb{P}\big(e^{\lambda_B \frac{\beta}{2\beta+1} T}\big|\widehat{B}_T(x) - B(x)\big| \geq C\big) > 0,$$

where the supremum is taken among all $B \in \mathcal{B}_b \cap \mathcal{H}^\beta_{\mathcal{D}}(L)$ and the infimum is taken among all estimators.

3.3. **Discussion of the results.**

Rates of convergence. The "parametric case" for a constant division rate $B(x) = b$ with $b > 0$ has a statistical simpler structure, but also a nice probabilistic feature since the process $t \rightsquigarrow |\partial\mathcal{T}_t|$, *i.e.* the number of cells alive at time t is Markov. In that setting, explicit (asymptotic) information bounds are available (Athreya and Keiding [1]). In particular, the model is regular with asymptotic Fisher information of order $e^{\lambda_B T}$, thus the best-achievable (normalised) rate of convergence is $e^{-\lambda_B T/2}$. This is consistent with the minimax rate $\exp(-\lambda_B \frac{\beta}{2\beta+1} T)$ that we have here, for the class $\mathcal{H}^\beta_{\mathcal{D}}(L)$, and we retrieve the parametric rate by formally setting $\beta = \infty$ in the previous formula.

However, this rate is strongly parameter dependent in the sense that it also depends on B via λ_B. This dependence is severe, since it appears at the same level as the smoothness exponent $\beta/(2\beta + 1)$ in the rate exponent $\frac{\beta}{2\beta+1}\lambda_B$. For instance, in the simplest case of a constant function $B(x) = b$ for every $x \geq 0$, we have $\lambda_B = (m-1)b$, and we see that B (b here) plays at the same level as $\beta/(2\beta+1)$. This also has a non-trivial technical cost in establishing rates of convergence for the estimator $\widehat{B}_T(x)$: in order to minimise the bias-variance tradeoff, the (log)-bandwidth has to be chosen as $-\lambda_B \frac{1}{2\beta+1} T\big(1+o(1)\big)$ exactly, and this is achieved by the plug-in rule $-\widehat{\lambda}_T \frac{1}{2\beta+1} T$ thanks to Proposition 18. We then have to carefully check that our estimator is not too sensitive to this further approximation, and this requires the analysis of the smoothness of the process $h \rightsquigarrow \widehat{B}_{T,h}(x)$ where h is the bandwidth of $\widehat{B}_T(x)$, as shown in Proposition 18.

Other loss functions. If $\mathcal{K} \subset \mathring{\mathcal{D}}$ is a closed interval ($\mathring{\mathcal{D}}$ denotes the interior of $\mathcal{D}$), then Theorem 9 also holds uniformly in $x \in \mathcal{K}$. So we also have that

$$e^{2\lambda_B \frac{\beta}{2\beta+1} T} \int_{\mathcal{K}} \big(\widehat{B}_T(x) - B(x)\big)^2 dx$$

is $\mathcal{B}_b \cap \mathcal{H}^\beta_{\mathcal{D}}(L)$-tight for the parameter B. For integrated squared error-loss, we could weaken the smoothness constraint $B \in \mathcal{H}^\beta_{\mathcal{D}}(L)$ to Sobolev smoothness (see *e.g.* [22]) when the smoothness is measured in L^2-norm. An extension of Theorem 10 can be obtained likewise.

Smoothness adaptation. Our estimator $\widehat{B}_T(x)$ *is not* β-adaptive, in the sense that the choice of the optimal (log) bandwidth $-\widehat{\lambda}_T \frac{1}{2\beta+1} T$ still depends on β, which is unknown in principle. In the numerical implementation Section 4 below, we address this issue from a practical point of view. However, a theoretical result is still needed. The classical analysis of adaptive (or other) kernel methods à la Lepski for instance shows that this boils down to proving concentration inequalities of the type

$$\mathbb{P}\big(\big|\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, g_h\big) - \mathring{\mathcal{E}}(g_h)\big| \geq e^{\lambda_B T/2} c(q,T)\big) \leq e^{-q\lambda_B T}, \;\; q > 0, \tag{19}$$

where, for $0 < h^{-1} \leq e^{\lambda_B T}$, the test function g_h has the form $g_h(y) = h^{-1/2} g\big(h^{-1}(x-y)\big)$ with $x \in \mathcal{D}$ and $g \in \mathcal{L}_C$. The threshold $c(q,T)$ should be of order $q\lambda_B T$ and would inflate the risk by a slow term (of order T). By a suitable choice of q, it would then be possible to obtain adaptation for β in compact intervals. Concentration inequalities like (19) have been explored in [4] in discrete time. To the best of our knowledge, such inequalities are not yet available in continuous time and lie beyond the scope of the paper.

Information from $\mathring{\mathcal{T}}_T$ *versus* $\partial\mathcal{T}_T$*.* Having

$$\partial\mathcal{E}(g) = \lambda_B \frac{m}{m-1} \int_0^\infty g(x) e^{-\lambda_B x} \exp\big(-\int_0^x B(y)dy\big)dx$$

and ignoring the fact that the constants m and λ_B are unknown (or rather knowing that they can be estimated at the superoptimal rate $e^{\lambda_B T/2}$), we can anticipate that by picking a suitable test function g mimicking a delta function $g(x) \approx \delta_x$, the information about $B(x)$ can only be inferred through $\exp(-\int_0^x B(y)dy)$, which imposes to further take a derivative hence some ill-posedness.

We can briefly make all these arguments more precise: we assume that we have estimators of $\widehat{m}_T$ of m and $\widehat{\lambda}_T$ of λ_B (using the ones defined in (14) and (15) or by any other means) that converge with rate $e^{\lambda_B T/2}$ as in Proposition 7. Consider the quantity

$$\widehat{f}_{h,T}(x) = -\mathcal{E}^T\Big(\partial\mathcal{T}_T, \frac{\widehat{m}_T - 1}{\widehat{\lambda}_T \widehat{m}_T}\big(K_h\big)'(x-\cdot)\Big)$$

for a kernel satisfying Assumption 8. By Theorem 4 and integrating by part, we readily see that

$$\widehat{f}_{h,T} \to -\partial\mathcal{E}\Big(\frac{m-1}{\lambda_B m}\big(K_h\big)'(x-\cdot)\Big) = \int_0^\infty K_h(x-y) f_{B+\lambda_B}(y)dy \tag{20}$$

in probability as $T \to \infty$, where $f_{B+\lambda_B}$ is the density associate to the division rate $x \rightsquigarrow B(x) + \lambda_B$. On the one hand, it is not difficult to show that Proposition 16 (used in the proof of Theorem 9 below) is valid when substituting $\mathring{\mathcal{T}}_T$ by $\partial\mathcal{T}_T$, so we can anticipate that the rate of convergence in (20) is of order $h^{-3/2}e^{\lambda_B T/2}$ due to the order of the functional $\Phi(g)$ that appears in Proposition 16 and which plays the role of a variance, when applied to the test function $g = \big(K_h\big)'(x-\cdot)$. On the other hand, the limit $\int_0^\infty K_h(x-y) f_{B+\lambda_B}(y)dy$ approximates $f_{B+\lambda_B}(x)$ with an error of order h^β if $B \in \mathcal{H}^\beta_{\mathcal{D}}$. Balancing the two error terms in h, we see that we can estimate $f_{B+\lambda_B}(x)$ with an error of (presumably optimal) order $\exp(-\lambda_B \frac{\beta}{2\beta+3} T)$. Due to the fact that the denominator in representation (3) can be estimated with parametric error rate $\exp(-\lambda_B T/2)$, we end up with the rate of estimation $\exp(-\lambda_B \frac{\beta}{2\beta+3} T)$ for $B(x)$ as well, and that can be related to an ill-posed problem of order 1 (see for instance [22]).

This phenomenon, namely the structure of an ill-posed problem of order 1 in restriction to data alive at time T, has already been observed in other settings: for the estimation of a size-division rate from living cells at a given large time in Doumic *et al.* [9, 8] or for the estimation of the dislocation measure for a homogeneous fragmentation in Hoffmann and Krell [12]. There is however no proof so far of the optimality of these rates, *i.e.* in our case an analogue of Theorem 10 with rate $\exp(-\lambda_B \frac{\beta}{2\beta+3}T)$, with the infimum being taken among all estimators based on data $\{\zeta_u^T, u \in \partial\mathcal{T}_T\}$ solely. Note also that this phenomenon does not appear in parametric estimation, since the number of data in $\mathring{\mathcal{T}}_T$ and $\partial\mathcal{T}_T$ are of the same order of magnitude (or put differently, the rates in Theorems 4 and 5 are the same and govern the rate of estimation of a one dimensional parameter).

4. Numerical implementation

We assume that each cell $u \in \mathcal{U}$ has exactly two offsprings at each division ($p_2 = 1$). This can model the evolution of a population of cells reproducing by binary division, as described deterministically by (1). We pick a trial division rate B defined analytically by

$$B(x) = \begin{cases} \frac{1}{3}x^3 - \frac{7}{8}x^2 + \frac{5}{8}x + \frac{4}{10} & \text{if} \quad 0 \leq x \leq \frac{3}{2} \\ \frac{119}{160} - \frac{1}{4}\exp\big(-(x-\frac{3}{2})\big) & \text{if} \quad x > \frac{3}{2} \end{cases}$$

and represented in Figure 2 (bold red line). We have $B \in \mathcal{B}_b$ for $b = 0.4$. Given $T > 0$ we simulate the life length of the rooted cell $\zeta_\emptyset$ with probability density f_B and set $d_\emptyset = \zeta_\emptyset$. For $u \in \mathcal{U}$ such that $d_u > T$, we do not simulate the life lengths of its descendants since they are not in the observation scheme $\mathring{\mathcal{T}}_T \cup \partial\mathcal{T}_T$. For $u \in \mathcal{U}$ such that $d_u \leq T$ we simulate ζ_{u0} and ζ_{u1} independently with probability density f_B; we set $d_{u0} := d_u + \zeta_{u0}$ and $d_{u1} := d_u + \zeta_{u1}$. We generate $M = 100$ trees up to time $T = 23$. Figure 1 represents a typical observation scheme with continuous or discrete representation. The (random) number of observations fluctuates a lot as shown in Table 1 where some elementary statistics are given (note that for a binary tree, we always have the identity $|\partial\mathcal{T}_T| = |\mathring{\mathcal{T}}_T| + 1$).

Min.	**1st Qu.**	**Med.**	**Mean**	**3rd Qu.**	**Max.**	**Std.**
3 726	43 930	96 480	115 800	144 100	561 200	102 408

Table 1. *Fluctuations of the number of observations $|\mathring{\mathcal{T}}_T|$ for $M = 100$ Monte-Carlo trees observed up to time $T = 23$.*

We take a Gaussian kernel $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and the bandwidth $\widehat{h}_T$ is chosen here according to the rule-of-thumb $1.06\hat{\sigma}|\mathring{\mathcal{T}}_T|^{-1/5}$ where $\hat{\sigma}$ is the empirical standard deviation of $(\zeta_u, u \in \mathring{\mathcal{T}}_T)$. We also implemented standard cross-validation with less success. We evaluate $\widehat{B}_T$ on a regular grid of $\mathcal{D} = [0.25, 0.5]$ with mesh $\Delta x = 0.01$. For each sample we compute the empirical error

$$e_i = \frac{\|\widehat{B}_T^{(i)} - B\|_{\Delta x}}{\|B\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where $\|\cdot\|_{\Delta x}$ denotes the discrete norm over the numerical sampling. Table 2 displays the mean-empirical error $\overline{e} = M^{-1}\sum_{i=1}^M e_i$ together with the empirical standard deviation $\big(M^{-1}\sum_{i=1}^M (e_i - \overline{e})^2\big)^{1/2}$. The comparison of f_B, the density of interest, and f_{H_B}, the biased density, on Figure 2, highlights the bias selection since f_{H_B} gives more weight to small lifetimes than f_B. The error deteriorates as x grows since the biased density f_{H_B} (bold blue line - we approximate the Malthus

$\boldsymbol{T}$	13	15	17	19	21	23
Mean $\lvert\mathring{\boldsymbol{\mathcal{T}}}_{\boldsymbol{T}}\rvert$	652	1 847	5 202	14 634	41 151	115 760
$\overline{\boldsymbol{e}}$	0.1624	0.1046	0.0735	0.0448	0.0307	0.0178
Std. dev.	0.1052	0.0764	0.0599	0.0260	0.0197	0.0092

TABLE 2. *Mean empirical relative error $\overline{e}$ and its standard deviation, with respect to T, for the division rate B reconstructed over the interval $\mathcal{D} = [0.25, 2.5]$ by the estimator $\widehat{B}_T$.*

parameter using (7) and we find $\lambda_B \approx 0.5173$) decreases, see Figure 2. Close to 0, our estimator exhibits a large bias, and this is presumably a boundary effect. The larger T, the better the reconstruction at a visual level, as shown on Figure 2 where 95%-level confidence bands are built so that for each point x, the lower and upper bounds include 95% of the estimators $(\widehat{B}_T^{(i)}(x), i = 1 \dots M)$. The error is close to $\exp(-2\lambda_B T/5)$ as expected: indeed, for a kernel of order n_0, the bias term in density estimation is of order $h^{\beta \wedge (n_0+1)}$. For the smooth B we have here, we rather expect for the rate $\exp(-\lambda_B \frac{(n_0+1)}{2(n_0+1)+1} T) = \exp(-2\lambda_B T/5)$ for the Gaussian kernel with $n_0 = 1$ that we use here, and this is consistent with what we observe in Figure 3.

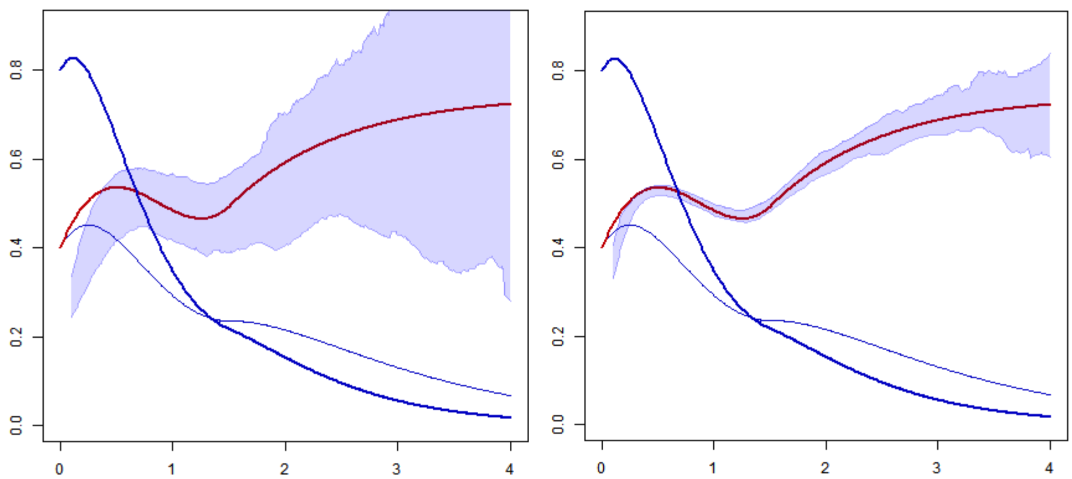

FIGURE 2. *Reconstruction of B over $\mathcal{D} = [0.1, 4]$ with 95%-level confidence bands constructed over $M = 100$ Monte-Carlo trees. In bold red line: $x \rightsquigarrow B(x)$; in bold blue line: f_{H_B}; in blue line: f_B. Left: $T = 15$. Right: $T = 23$.*

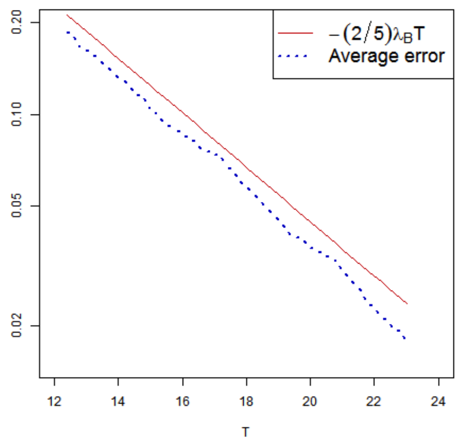

Figure 3. *The log-average relative empirical error over $M = 100$ Monte-Carlo trees vs. T (i.e. the log-rate) for $x \rightsquigarrow B(x)$ reconstructed over $\mathcal{D} = [0.25, 2.5]$ with $x \rightsquigarrow \widehat{B}_T(x)$ (dashed blue line) compared to the expected log-rate (solid red line).*

5. Proofs

For a locally integrable $B : [0,\infty) \to [0,\infty)$ such that $\int^\infty B(y)dy = \infty$, set

$$f_B(x) = B(x)e^{-\int_0^x B(y)dy}, \quad x \geq 0.$$

Define also $H_B(x)$ via

$$f_{H_B}(x) = me^{-\lambda_B x} f_B(x).$$

5.1. **Preliminaries.**

Many-to-one formulae. For $u \in \mathcal{U}$, we write ζ_u^t for the age of the cell u at time $t \in I_u = [b_u, d_u)$, *i.e.* $\zeta_u^t = (t - b_u)\mathbf{1}_{\{t \in I_u\}}$. We extend ζ_u^t over $[0, b_u)$ by setting $\zeta_u^t = \zeta_{u(t)}^t$, where $u(t)$ is the ancestor of u living at time t, defined by $u(t) = v$ if $v \preceq u$ and $(v,t) \in \mathbb{T}$. For $t \geq d_u$ we set $\zeta_u^t = \zeta_u$. Note that $\zeta_u^T = \zeta_u$ on the event $u \in \mathring{\mathcal{T}}_T$.

Let $(\chi_t)_{t\geq 0}$ and $(\widetilde{\chi}_t)_{t \geq 0}$ denote the one-dimensional Markov processes with infinitesimal generators (densely defined on continuous functions) $\mathcal{A}_B$ and $\mathcal{A}_{H_B}$ respectively, where

$$\mathcal{A}_B g(x) = g'(x) + B(x)\big(g(0) - g(x)\big)$$

and such that $\mathbb{P}(\chi_0 = 0) = \mathbb{P}(\widetilde{\chi}_0 = 0) = 1$.

Proposition 11 (Many-to-one formulae)**.** *For any $g \in \mathcal{L}_C$, we have*

$$\mathbb{E}\Big[\sum_{u \in \partial\mathcal{T}_T} g(\zeta_u^T)\Big] = \frac{e^{\lambda_B T}}{m}\mathbb{E}\Big[g(\widetilde{\chi}_T)B(\widetilde{\chi}_T)^{-1}H_B(\widetilde{\chi}_T)\Big], \tag{21}$$

and

$$\mathbb{E}\Big[\sum_{u \in \mathring{\mathcal{T}}_T} g(\zeta_u^T)\Big] = \mathbb{E}\Big[\sum_{u \in \mathring{\mathcal{T}}_T} g(\zeta_u)\Big] = \frac{1}{m}\int_0^T e^{\lambda_B s}\mathbb{E}\Big[g(\widetilde{\chi}_s)H_B(\widetilde{\chi}_s)\Big]ds. \tag{22}$$

In order to compute rates of convergence, we will also need many-to-one formulae over pairs of individuals. We can pick two individuals in the same lineage or over forks, *i.e.* over pairs of individuals that are not in the same lineage. If $u, v \in \mathcal{U}$, $u \wedge v$ denote their most recent common ancestor. Define

$$\mathcal{FU} = \{(u,v) \in \mathcal{U}^2, |u \wedge v| < |u| \wedge |v|\} \text{ and } \mathcal{FT} = \mathcal{FU} \cap \mathcal{T}^2.$$

Introduce also $\bar{m} = \sum_{i \neq j} \sum_{k \geq i \vee j} p_k$ which is finite by Assumption 2. Finally, denote by $(P^t_{H_B})_{t \geq 0}$ the Markov semigroup associated to $\mathcal{A}_{H_B}$.

Proposition 12 (Many-to-one formulae over pairs)**.** *For any $g \in \mathcal{L}_C$, we have*

$$\mathbb{E}\Big[\sum_{\substack{u,v \in \partial \mathcal{T}_T,\\ u \neq v}} g(\zeta_u^T) g(\zeta_v^T)\Big] = \frac{\bar{m}}{m^3} \int_0^T e^{\lambda_B s}\Big(e^{\lambda_B (T-s)} P_{H_B}^{T-s}\big(g \frac{H_B}{B}\big)(0)\Big)^2 P^s_{H_B} H_B(0) ds, \tag{23}$$

$$\mathbb{E}\Big[\sum_{(u,v) \in \mathcal{FT} \cap \mathring{\mathcal{T}}_T^2} g(\zeta_u) g(\zeta_v)\Big] = \frac{\bar{m}}{m^3} \int_0^T e^{\lambda_B s}\Big(\int_0^{T-s} e^{\lambda_B t} P^t_{H_B}(g H_B)(0) dt\Big)^2 P^s_{H_B} H_B(0) ds, \tag{24}$$

and

$$\mathbb{E}\big[\sum_{\substack{u,v \in \mathring{\mathcal{T}}_T,\\ u \prec v}} g(\zeta_u) g(\zeta_v)\big] = \int_0^T e^{\lambda_B s}\Big(\int_0^{T-s} e^{\lambda_B t} P^t_{H_B}\big(g H_B\big)(0) dt\Big) P^s_{H_B}(g H_B)(0) ds. \tag{25}$$

The identity (23) is a particular case of Lemma 3.9 of Cloez [5]. In order to obtain identity (24), we closely follow the method of Bansaye *et al.* [3]. Although the setting in [3] is much more general than ours, it formally only applies for exponential renewal times (corresponding to constant functions B) so we need to slightly accomodate their proof. The same ideas enable us to prove (25). This is set out in details in the appendix.

Geometric ergodicity of the auxiliary Markov process. Define the probability measure

$$\mu_B(x)dx = c_B \exp(-\int_0^x H_B(y) dy) dx \text{ for } x \geq 0.$$

We have the fast convergence of $P^T_{H_B}$ toward μ_B as $T \to \infty$. More precisely,

Proposition 13. *Let $\rho_B = \inf_x H_B(x) \geq 0$.*

(i) *For any $B \in \mathbb{B}_{b, m/(m-1)}$, $g \in \mathcal{L}_C$, $t \geq 0$ and $x \in (0, \infty)$, we have*

$$\Big| P^t_{H_B} g(x) - \int_0^\infty g(y) \mu_B(y) dy \Big| \leq 4 \sup_y |g(y)| \exp\big(-\rho_B t\big).$$

(ii) *For every $b > 0$ and every $B \in \mathcal{B}_b$, we have $\lambda_B \leq \rho_B$.*

Proof. First, one readily checks that $\int_0^\infty \mathcal{A}_{H_B} f(x) \mu_B(x) dx = 0$ for any continuous f, and since moreover $P^t_{H_B}$ is strongly Feller, it admits $\mu_B(x)dx$ as an invariant probability. In order to prove (i), it is sufficient to show

$$\|Q_B^{x,t} - \mu_B\|_{TV} \leq \exp(-\rho_B t)$$

where $Q_B^{x,t}$ denotes the law of of the Markov process with infinitesimal generator $\mathcal{A}_{H_B}$ started from x at time $t = 0$ and $\|\cdot\|_{TV}$ is the total variation norm between probability measures. Let

$N(ds\,dt)$ be a Poisson random measure with intensity $ds \otimes dt$ on $[0,\infty) \times [0,\infty)$. Define on the same probability space two random processes $(Y_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ such that

$$Y_t = x + t - \int_0^t \int_0^\infty Y_{s_-} \mathbf{1}_{\{z \leq H_B(Y_{s^-})\}} N(dz\,ds), \quad t \geq 0,$$

$$Z_t = Z_0 + t - \int_0^t \int_0^\infty Z_{s_-} \mathbf{1}_{\{z \leq H_B(Z_{s^-})\}} N(dz\,ds), \quad t \geq 0,$$

where Z_0 is a random variable with distribution μ_B. We have that both $(Y_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ are Markov processes driven by the same Poisson random measure with generator $\mathcal{A}_{H_B}$. Moreover, if N has a jump in $[0,t) \times [0, \inf_x H_B(x)]$, then Y_t and Z_t both necessarily start from 0 after this jump and coincide further on. It follows that

$$\mathbb{P}(Y_t \neq Z_t) \leq \mathbb{P}\Big(\int_0^t \int_0^{\inf_x H_B(x)} N(ds\,dt) = 0\Big) = \exp(-\inf_x H_B(x)t) = \exp(-\rho_B t).$$

Observing that Y_t and Z_t have distribution $Q_B^{x,t}$ and μ_B respectively, we conclude thanks to the fact that $\|Q_B^{x,t} - \mu_B\|_{TV} \leq 2\mathbb{P}(Y_t \neq Z_t)$.

We now turn to (ii). By representation (3), we readily obtain

$$\begin{aligned} H_B(x) &= \frac{me^{-\lambda_B x} f_B(x)}{1 - m\int_0^x e^{-\lambda_B y} f_B(y)dy} \\ &= \frac{me^{-\lambda_B x} B(x) e^{-\int_0^x B(y)dy}}{1 - m\int_0^x e^{-\lambda_B y} B(y) e^{-\int_0^y B(u)du} dy}. \end{aligned}$$

Set

$$G_B(x) = me^{-\lambda_B x} B(x) e^{-\int_0^x B(y)dy} - \lambda_B\big(1 - m\int_0^x e^{-\lambda_B y} B(y) e^{-\int_0^y B(u)du} dy\big).$$

Statement (ii) is equivalent to proving that $\inf_{x\geq 0} G_B(x) \geq 0$. We first claim that

$$B(x) \leq \widetilde{B}(x) \ \text{ for every } \ x \in (0,\infty) \ \text{ implies } \ \lambda_B \leq \lambda_{\widetilde{B}}.$$

Indeed, in that case, one can construct on the same probability space two random variables τ_B and $\tau_{\widetilde{B}}$ such that $\tau_B \geq \tau_{\widetilde{B}}$. It follows that $\phi_B(\lambda) = \mathbb{E}[e^{-\lambda\tau_B}] \leq \phi_{\widetilde{B}}(\lambda) = \mathbb{E}[e^{-\lambda\tau_{\widetilde{B}}}]$ for every $\lambda \geq 0$. Also, ϕ_B and $\phi_{\widetilde{B}}$ are both non-increasing, vanishing at infinity, and $\phi_B(0) = \phi_{\widetilde{B}}(0) = 1 > \frac{1}{m}$. Consequently, the values λ_B and $\lambda_{\widetilde{B}}$ such that $\phi_B(\lambda_B) = \phi_{\widetilde{B}}(\lambda_{\widetilde{B}}) = \frac{1}{m}$ necessarily satisfy $\lambda_B \leq \lambda_{\widetilde{B}}$ hence the claim. Now, for constant functions $B(x) = \alpha$, we clearly have $\lambda_B = (m-1)\alpha$ and this enables us to infer

$$\lambda_B \leq (m-1)\sup_x B(x).$$

Remember now that $B \in \mathcal{B}_b$ implies $b \leq B(x) \leq \frac{m}{m-1}b$ for every $x \geq 0$. Therefore

$$\lambda_B \leq (m-1)\frac{m}{m-1}b = mb \leq mB(0) \tag{26}$$

and $G_B(0) = mB(0) - \lambda_B \geq 0$ follows. Moreover, one readily checks that

$$G_B'(x) = me^{-\lambda_B x} e^{-\int_0^x B(y)dy}\big(B'(x) - B(x)^2\big) \leq 0$$

since $B'(x) - B(x)^2 \leq 0$ as soon as $B \in \mathcal{B}_b$. So G_B is non-increasing, $G_B(0) \geq 0$ and its infimum is thus attained for $x \to \infty$. Since $G_B(\infty) = 0$, we conclude $\inf_{x\geq 0} G_B(x) \geq 0$.

□

5.2. **Proof of Theorems 4 and 5.** In order to ease notation, when no confusion is possible, we abbreviate $\mathcal{B}_b$ by $\mathcal{B}$, $\mathcal{L}_C$ by $\mathcal{L}$ and so on.

Proof of Theorem 4. Writing

$$e^{\lambda_B T/2}\big(\mathcal{E}(\partial\mathcal{T}_T,g)-\partial\mathcal{E}(g)\big)=\frac{e^{\lambda_B T}}{\big|\partial\mathcal{T}_T\big|}\,e^{-\lambda_B T/2}\sum_{u\in\partial\mathcal{T}_T}\big(g(\zeta_u^T)-\partial\mathcal{E}(g)\big),$$

Theorem 4 is then a consequence of the following two facts: first we claim that

$$e^{\lambda_B T}\big|\partial\mathcal{T}_T\big|^{-1}\to W_B\ \text{ in probability as }\ T\to\infty, \tag{27}$$

uniformly in $B\in\mathbb{B}$, where the random variable W_B satisfies $\mathbb{P}(W_B>0)=1$, and second, for $B\in\mathcal{B}$ and $g\in\mathcal{L}$, we claim that the following estimate holds:

$$\mathbb{E}\Big[\Big(\sum_{u\in\partial\mathcal{T}_T}\big(g(\zeta_u^T)-\partial\mathcal{E}(g)\big)\Big)^2\Big]\lesssim e^{\lambda_B T}, \tag{28}$$

where $\lesssim$ means up to a constant (possibly varying from line to line) that only depends on $\mathcal{B}$ (or $\mathbb{B}$) and $\mathcal{L}$.

Step 1. The convergence (27) is a consequence of the following lemma:

Lemma 14. *For every $B\in\mathbb{B}$, there exists $\widetilde{W}_B$ with $\mathbb{P}(\widetilde{W}_B>0)=1$ such that*

$$\mathbb{E}\Big[\Big(\frac{|\partial\mathcal{T}_T|}{\mathbb{E}\big[|\partial\mathcal{T}_T|\big]}-\widetilde{W}_B\Big)^2\Big]\to 0\ \text{ as }\ T\to\infty, \tag{29}$$

uniformly in $B\in\mathbb{B}$ and

$$\kappa_B^{-1}e^{\lambda_B T}\mathbb{E}\big[|\partial\mathcal{T}_T|\big]\to 1\ \text{ as }\ T\to\infty, \tag{30}$$

uniformly in $B\in\mathbb{B}$, where $\kappa_B^{-1}=\lambda_B\frac{m}{m-1}\int_0^\infty\exp(-\int_0^x H_B(y)dy)dx$.

Lemma 14 is well known, and follows from classical renewal arguments, see Chapter 6 in the book of Harris [11]. Only the uniformity in $B\in\mathbb{B}$ requires an extra argument, but with a uniform version of the key renewal theorem of [21], it readily follows from the proof of Harris, so we omit it. Note that (29) and (30) entail the convergence $e^{\lambda_B T}|\partial\mathcal{T}_T|^{-1}\to\kappa_B\widetilde{W}_B^{-1}=W_B$ in probability as $T\to\infty$ uniformly in $B\in\mathbb{B}$, and this entails (27).

Step 2. We now turn to the proof of (28). Without loss of generality, we may (and will) assume that $\partial\mathcal{E}(g)=0$. We have

$$\mathbb{E}\big[\big(\sum_{u\in\partial\mathcal{T}_T}g(\zeta_u^T)\big)^2\big]=\mathbb{E}\big[\sum_{u\in\partial\mathcal{T}_T}g(\zeta_u^T)^2\big]+\mathbb{E}\big[\sum_{\substack{u,v\in\partial\mathcal{T}_T,\\ u\neq v}}g(\zeta_u^T)g(\zeta_v^T)\big]=I+II,$$

say. By (21) in Proposition 11, we write

$$\begin{aligned}I&=\frac{e^{\lambda_B T}}{m}\mathbb{E}\Big[g(\widetilde{\chi}_T)^2B(\widetilde{\chi}_T)^{-1}H_B(\widetilde{\chi}_T)\Big]\\&\leq\frac{e^{\lambda_B T}}{m}\int_0^\infty g(x)^2\tfrac{H_B(x)}{B(x)}\mu_B(x)dx+\tfrac{e^{\lambda_B T}}{m}\Big|P_{H_B}^T\big(g^2\tfrac{H_B}{B}\big)(0)-\int_0^\infty g(x)^2\tfrac{H_B(x)}{B(x)}\mu_B(x)dx\Big|.\end{aligned}$$

Since $g\in\mathcal{L}$ and $B\in\mathbb{B}$, we successively have

$$m^{-1}\int_0^\infty g(x)^2\tfrac{H_B(x)}{B(x)}\mu_B(x)dx\lesssim 1\ \text{ and }\ g(x)^2\tfrac{H_B(x)}{B(x)}\lesssim 1,$$

so applying (i) of Proposition 13 we derive

$$\Big|P_{H_B}^T\big(g^2\tfrac{H_B}{B}\big)(0)-\int_0^\infty g(x)^2\tfrac{H_B(x)}{B(x)}\mu_B(x)dx\Big|\lesssim 1,$$

and we conclude that $I\lesssim e^{\lambda_B T}$. In order to bound II, we use identity (23) of Proposition 12 and obtain

$$II=\frac{\bar m e^{2\lambda_B T}}{m^3}\int_0^T e^{-\lambda_B s}\Big(P_{H_B}^{T-s}\big(g\tfrac{H_B}{B}\big)(0)\Big)^2 P_{H_B}^s H_B(0)ds.$$

We have $P_{H_B}^s H_B(0)\lesssim 1$ since $B\in\mathbb{B}$ and also $|g(x)|\frac{H_B(x)}{B(x)}\lesssim 1$ since $g\in\mathcal{L}$ so applying Proposition 13 to the test function $g(x)\frac{H_B(x)}{B(x)}$ which has vanishing integral under μ_B, we obtain

$$\big|P_{H_B}^{T-s}\big(g\tfrac{H_B}{B}\big)(0)\big|\lesssim e^{-\rho_B(T-s)}$$

hence

$$|II|\lesssim e^{2\lambda_B T}\int_0^T e^{-2\rho_B(T-s)}e^{-\lambda_B s}ds\leq \lambda_B^{-1}e^{\lambda_B T}$$

using (ii) of Proposition 13. We conclude $|II|\lesssim e^{\lambda_B T}$ since $\inf_{B\in\mathcal{B}}\lambda_B>0$. Indeed, reasoning as in the proof of Proposition 13 (ii), we readily obtain the estimate $(m-1)b\leq\lambda_B$ for $B\in\mathcal{B}$. $\square$

Proof of Theorem 5. The proof goes along the same line but is slightly more intricate. First, we implicitly work on the event $\{|\mathring{\mathcal{T}}_T|\geq 1\}$ which has probability that goes to 1 as $T\to\infty$, uniformly in $B\in\mathbb{B}$. Next, we again write

$$e^{\lambda_B T/2}\big(\mathcal{E}(\mathring{\mathcal{T}}_T,g)-\mathring{\mathcal{E}}(g)\big)=e^{\lambda_B T}\frac{e^{-\lambda_B T/2}}{|\mathring{\mathcal{T}}_T|}\sum_{u\in\mathring{\mathcal{T}}_T}\big(g(\zeta_u^T)-\mathring{\mathcal{E}}(g)\big),$$

and we claim that

$$e^{\lambda_B T}\big|\mathring{\mathcal{T}}_T\big|^{-1}\to W_B'>0\ \text{ in probability as }\ T\to\infty,\tag{31}$$

uniformly in $B\in\mathbb{B}$, where W_B' satisfies $\mathbb{P}(W_B'>0)=1$ and that the following estimate holds:

$$\mathbb{E}\Big[\Big(\sum_{u\in\mathring{\mathcal{T}}_T}\big(g(\zeta_u^T)-\mathring{\mathcal{E}}(g)\big)\Big)^2\Big]\lesssim e^{\lambda_B T},\tag{32}$$

uniformly in $B\in\mathcal{B}$ and $g\in\mathcal{L}$. In the same way as in the proof of Theorem 4, (31) is a consequence of the following classical result, which can be obtained in the same way as for Lemma 14 and proof of which we omit.

Lemma 15. *For every $B\in\mathbb{B}$, there exists $\widetilde{W}_B'>0$ with $\mathbb{P}(\widetilde{W}_B'>0)=1$ such that*

$$\mathbb{E}\Big[\Big(\frac{|\mathring{\mathcal{T}}_T|}{\mathbb{E}\big[|\mathring{\mathcal{T}}_T|\big]}-\widetilde{W}_B'\Big)^2\Big]\to 0\ \text{ as }\ T\to\infty,$$

uniformly in $B\in\mathbb{B}$ and

$$(\kappa_B')^{-1}e^{\lambda_B T}\mathbb{E}\big[|\mathring{\mathcal{T}}_T|\big]\to 1\ \text{ as }\ T\to\infty,$$

uniformly in $B\in\mathbb{B}$, where $(\kappa_B')^{-1}=\lambda_B m\int_0^\infty\exp(-\int_0^x H_B(y)dy)dx$.

It remains to prove (32). We again assume without loss of generality that $\mathring{\mathcal{E}}(g) = 0$ and we plan to use the following decomposition:

$$\mathbb{E}\big[\big(\sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u)\big)^2\big] = I + II + III,$$

with

$$I = \mathbb{E}\big[\sum_{u\in\mathring{\mathcal{T}}_T} g(\zeta_u)^2\big],$$

$$II = \mathbb{E}\big[\sum_{(u,v)\in\mathcal{F}\mathcal{T}\cap\mathring{\mathcal{T}}_T^2} g(\zeta_u)g(\zeta_v)\big]$$

and

$$III = 2\mathbb{E}\big[\sum_{\substack{u,v\in\mathring{\mathcal{T}}_T,\\ u\prec v}} g(\zeta_u)g(\zeta_v)\big].$$

Step 1. By (22) of Proposition 11, we have

$$I = \frac{1}{m}\int_0^T e^{\lambda_B s}\mathbb{E}\big[g(\widetilde{\chi}_s)^2 H_B(\widetilde{\chi}_s)\big]ds,$$

In the same way as for the term I in the proof of Theorem 4, we readily check that $g \in \mathcal{L}$ and $B \in \mathbb{B}$ guarantee that $\mathbb{E}\big[g(\widetilde{\chi}_s)^2 H_B(\widetilde{\chi}_s)\big] \lesssim 1$ uniformly in $s \in [0, T]$, so $I \lesssim e^{\lambda_B T}$.

Step 2. By (24) of Proposition 12, we obtain

$$II = \frac{\bar{m}}{m^3}\int_0^T e^{\lambda_B s}\Big(\int_0^{T-s} e^{\lambda_B t}P^t_{H_B}(gH_B)(0)dt\Big)^2 P^s_{H_B}(H_B)(0)ds.$$

We work as for the term II in the proof of Theorem 4: we successively have $P^s_{H_B}(H_B)(0) \lesssim 1$ and $\big|P^t_{H_B}(gH_B)(0)\big| \lesssim \exp(-\rho_B t)$ by applying Proposition 13 and the fact that gH_B has vanishing integral under μ_B. Hence

$$|II| \lesssim \int_0^T e^{\lambda_B s}\Big(\int_0^{T-s} e^{(\lambda_B-\rho_B)t}dt\Big)^2 ds \lesssim e^{\lambda_B T}$$

proceeding as before, and we readily check that the estimates are uniform in $B \in \mathcal{B}$ and $g \in \mathcal{L}$.

Step 3. Using (25) of Proposition 12,

$$|III| \leq \int_0^T e^{\lambda_B s}\Big|\int_0^{T-s} e^{\lambda_B t}P^t_{H_B}\big(gH_B\big)(0)dt\Big|P^s_{H_B}(|g|H_B)(0)ds.$$

In the same way as for the term II, we have $|P^t_{H_B}(gH_B)(0)| \lesssim \exp(-\rho_B t)$ and therefore

$$\Big|\int_0^{T-s} e^{\lambda_B t}P^t_{H_B}\big(gH_B\big)(0)dt\Big| \lesssim T - s$$

as before, the uniformity in $B \in \mathcal{B}$ being granted by Proposition 13 (ii). Plugging this estimate, we derive

$$|III| \lesssim \int_0^T e^{\lambda_B s}(T-s)P^s_{H_B}(|g|H_B)(0)ds \lesssim e^{\lambda_B T}$$

thanks to $|P^s_{H_B}(|g|H_B)(0)| \lesssim 1$. $\square$

5.3. **Proof of Proposition 7.** Conditional on $\mathring{\mathcal{T}}_T$, the random variables $(\nu_u, u \in \mathring{\mathcal{T}}_T)$ are independent, with common distribution p_k. It follows that

$$\mathbb{E}\big[(\widehat{m}_T - m)^2 \,|\mathring{\mathcal{T}}_T\big] \leq |\mathring{\mathcal{T}}_T|^{-1} \sum_k k^2 p_k.$$

Since $e^{\lambda_B T}|\mathring{\mathcal{T}}_T|^{-1}$ is $\mathbb{B}$-tight thanks to Lemma 15, we obtain the result for $e^{\lambda_B T/2}(\widehat{m}_T - m)$. The $\mathbb{B}$-tightness of $T^{-1/2}e^{\lambda_B T/2}(\widehat{\lambda}_T - \lambda_B)$ is a consequence of Theorem 4 and 5, together with the convergence of the preliminary estimators $\widehat{m}_T$. For $M > 0$,

$$\begin{aligned}\mathcal{E}^T(\mathrm{Id}, \partial\mathcal{T}_T) - \partial\mathcal{E}(\mathrm{Id}) &= \big(\mathcal{E}^T(\min\{\mathrm{Id}, M\}, \partial\mathcal{T}_T) - \partial\mathcal{E}(\min\{\mathrm{Id}, M\})\big) \\ &\quad + \Big(\mathcal{E}^T\big((\cdot - M)\mathbf{1}_{\{\cdot > M\}}, \partial\mathcal{T}_T\big) - \partial\mathcal{E}\big((\cdot - M)\mathbf{1}_{\{\cdot > M\}}\big)\Big) = I + II\end{aligned}$$

say. We choose $M = M_T = 2T$ and we apply Theorem 4 for the test functions $g_T(x) = \min\{x, M_T\}/M_T$ which are uniformly bounded in T to get the $\mathbb{B}$-tightness of $T^{-1/2}e^{\lambda_B T/2}I$. Since $\zeta_u \leq T$ when $u \in \partial T_T$, we also have $|II| = \partial\mathcal{E}\big((\cdot - 2T)\mathbf{1}_{\{\cdot > 2T\}}\big)$ and we deduce that $e^{\lambda_B T/2}II$ is $\mathbb{B}$-tight. We study in the same way $\mathcal{E}^T(\mathrm{Id}, \mathring{\mathcal{T}}_T)$ to conclude.

5.4. **Proof of Theorem 9.** The proof of Theorem 9 goes along the classical line of a bias-variance analysis in nonparametrics (see for instance the classical textbook [22]). However, we have two kind of extra difficulties: first we have to get rid of the random bandwidth $\widehat{h}_T = \exp(-\frac{1}{2\beta+1}\widehat{\lambda}_T)$ defined in (18) (actually this is the most delicate part of the proof) and second, we have to get rid of the preliminary estimators $\widehat{m}_T$ and $\widehat{\lambda}_T$.

The point $x \in (0, \infty)$ where we estimate $B(x)$ is fixed throughout, and further omitted in the notation. We first need a slight extension of Theorem 5 – actually of the estimate (32) – in order to accomodate test functions $g = g_T$ such that $g_T \to \delta_x$ weakly as $T \to \infty$. To that end, define, for $C > 0$

$$\mathcal{C}_C = \big\{g : \mathbb{R} \to \mathbb{R},\ \mathrm{supp}(g) \subset [0, C] \text{ and } \sup_y |g(y)| \leq C\big\}.$$

For a function $g : [0, \infty) \to \mathbb{R}$ let

$$|g|_1 = \int_0^\infty |g(y)|dy,\ \ |g|_2^2 = \int_0^\infty g(y)^2 dy \ \text{ and } \ |g|_\infty = \sup_y |g(y)|$$

denote the usual L^p-norm over $[0, \infty)$ for $p = 1, 2, \infty$ and define also

$$\Phi(g) = |g|_2^2 + \inf_{1 \leq \omega \leq e^{\lambda_{\mathcal{B}} T}} \big(|g|_1^2\omega + |g|_\infty^2\omega^{-1}\big) + |g|_1|g|_\infty, \tag{33}$$

where $\lambda_{\mathcal{B}} = \inf_{B \in \mathcal{B}} \lambda_B \geq (m-1)b$ (recall the argument which leads to (26) in the proof of (ii) of Proposition 13).

Proposition 16. *In the same setting as Theorem 5, we have, for any $g \in \mathcal{C}_C$,*

$$\mathbb{E}\Big[\Big(\sum_{u \in \mathring{\mathcal{T}}_T} \big(g(\zeta_u^T) - \mathring{\mathcal{E}}(g)\big)\Big)^2\Big] \lesssim e^{\lambda_B T}\Phi(g) + |g|_\infty^2, \tag{34}$$

where the symbol $\lesssim$ means here uniformly in $B \in \mathcal{B}$ and independently of g.

Let us briefly comment Proposition 16. If $g \in \mathcal{C}_C$, consider the function $g_{h_T}(y) = h_T^{-1}g\big(h_T^{-1}(x - y)\big)$ that mimicks the Dirac function δ_x for $h_T \to 0$. It is noteworthy that in the left-hand side of (34), $g_{h_T}(\zeta_u^T)^2$ is of order h_T^{-2} since $h_T^{1/p}g_{h_T}$ is of order 1, $h^{-1/2}$, h^{-1} in L^p for respectively

$p = 1, 2, \infty$ if we pick $\omega = h_T^{-1}$. We can thus expect the right-hand side to be of order $e^{\lambda_B T} h_T^{-1}$ if h_T is not too small and gain a crucial factor h_T thanks to averaging over $\mathring{\mathcal{T}}_T$.

Proof. We carefully revisit the estimate (32) in the proof of Theorem 5 keeping up with the same notation and assuming with no loss of generality that $\mathring{\mathcal{E}}(g) = 0$.

Step 1. For the term I, we insert $\int_0^\infty g(y)^2 H_B(y)\mu_B(y)dy = mc_B \int_0^\infty g(y)^2 e^{-\lambda_B y} f_B(y)dy$ to obtain $I = IV + V$, where

$$IV \lesssim e^{\lambda_B T} \int_0^\infty g(y)^2 e^{-\lambda_B y} f_B(y)dy$$

and

$$|V| \leq \frac{1}{m}\int_0^T e^{\lambda_B s}\Big|P^s_{H_B}\big(g^2 H_B\big)(0) - \int_0^\infty g(y)^2 H_B(y)\mu_B(y)dy\Big|ds.$$

Clearly, $|IV| \lesssim e^{\lambda_B T}|g|_2^2$. Also, by Proposition 13 (i) we further infer

$$|V| \lesssim |g|_\infty^2 \int_0^T e^{\lambda_B s} e^{-\rho_B s} ds \lesssim |g|_\infty^2$$

where we used (ii) of Proposition 13 for the last estimate.

Step 2. For the term II, using $P^s_{H_B}(H_B)(0) \lesssim 1$ we now obtain

$$II \lesssim e^{\lambda_B T}\int_0^T e^{-\lambda_B s}\Big(\int_0^s e^{\lambda_B t} P^t_{H_B}(gH_B)(0)dt\Big)^2 ds.$$

A new difficulty appears here, since the crude bound

$$|P^t_{H_B}(gH_B)(0)| \lesssim |g|_\infty \exp(-\rho_B t) \tag{35}$$

given by Proposition 13 (i) does not yield to the correct order for small value of t in the above integral. We need the following refinement (for small values of t), based on a renewal argument and proved in the Appendix:

Lemma 17. *For every $t \geq 0$ and $g \in \mathcal{C}_C$, we have*

$$\big|P^t_{H_B}\big(gH_B\big)(0)\big| \lesssim |g(t)|e^{-\lambda_B t} + |g|_1$$

uniformly in $B \in \mathbb{B}$.

Let $v \in [0, T]$ be arbitrary. For $0 \leq s \leq v$, applying Lemma 17, we obtain

$$\mathcal{I}_s = \Big(\int_0^s e^{\lambda_B t}|P^t_{H_B}(gH_B)(0)|dt\Big)^2 \lesssim \Big(\int_0^s |g(t)|dt + |g|_1\int_0^s e^{\lambda_B t}dt\Big)^2 \lesssim |g|_1^2 e^{2\lambda_B s}.$$

For $s \geq v$, using that $e^{-\rho_B t} \leq e^{-\lambda_B t}$ in (35) for $B \in \mathcal{B}$, we have

$$\mathcal{I}_s \lesssim \mathcal{I}_v + |g|_\infty^2 (s-v)^2 \mathbf{1}_{\{s \geq v\}}.$$

On the one hand, we deduce

$$\int_0^v e^{-\lambda_B s}\mathcal{I}_s ds \lesssim |g|_1^2 e^{\lambda_B v},$$

and on the other hand $\int_v^T e^{-\lambda_B s}\mathcal{I}_s ds$ is less than

$$\mathcal{I}_v \int_v^T e^{-\lambda_B s}ds + |g|_\infty^2 \int_v^T e^{-\lambda_B s}(s-v)^2 ds \lesssim |g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{-\lambda_B v}$$

whence

$$|II| \lesssim e^{\lambda_B T}\big(|g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{-\lambda_B v}\big)$$

follows.

Step 3. Finally going back to Step 3 in the proof of Theorem 5 we readily obtain

$$\begin{aligned}|III| &\lesssim \int_0^T e^{\lambda_B s} P_{H_B}^s\big(|g|H_B\big)(0) \int_0^{T-s} e^{\lambda_B t}\big|P_{H_B}^t\big(gH_B\big)(0)\big|dtds \\ &\lesssim \int_0^T e^{\lambda_B s}(|g(s)|e^{-\lambda_B s} + |g|_1)|g|_\infty \int_0^{T-s} e^{\lambda_B t} e^{-\rho_B t} dtds\end{aligned}$$

by applying Lemma 17 for the term involving $P_{H_B}^s$ and the estimate (35) for the term involving $P_{H_B}^t$, therefore $|III| \lesssim e^{\lambda_B T}|g|_1|g|_\infty$. □

Proposition 16 enables us to obtain the next result, that is the key ingredient to get rid of the random bandwidth $\widehat{h}_T$, thanks to the fact that it is concentrated around its estimated value $h_T(\beta)$.

Denote by $\mathcal{C}_C^1$ (later abbreviated by $\mathcal{C}^1$) the subset of $\mathcal{C}_C$ of functions that are moreover differentiable, with derivative uniformly bounded by C. For $a, b \geq 0$ we set $[a \pm b] = [\max\{0, a-b\}, a+b]$.

Proposition 18. *Assume that $\beta > 1$. For every $\kappa > 0$,*

$$e^{\lambda_B T/2} \sup_{h \in [h_T(\beta) \pm \kappa T^{3/2} e^{-\lambda_B T/2}]} \big|\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, h^{1/2} f g_h\big) - \mathring{\mathcal{E}}(h^{1/2} f g_h)\big|$$

is $(\mathcal{B}, \mathcal{L}, \mathcal{C}^1)$-tight, for the parameter (B, f, g), and where $h_T(\beta) = e^{-\frac{1}{2\beta+1}\lambda_B T}$.

Proof. Step 1. Define $\overline{fg_h} = fg_h - \mathring{\mathcal{E}}(fg_h)$. Writing

$$e^{\lambda_B T/2}\Big(\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, h^{1/2} f g_h\big) - \mathring{\mathcal{E}}(h^{1/2} f g_h)\Big) = \frac{e^{\lambda_B T}}{|\mathring{\mathcal{T}}_T|} e^{-\lambda_B T/2} \sum_{u \in \mathring{\mathcal{T}}_T} h^{1/2}\overline{fg_h}(\zeta_u),$$

we see as in the proof of Theorem 5 that thanks to Lemma 15, it is enough to prove the $\mathcal{B}$-tightness of

$$\sup_{h \in [h_T(\beta) \pm \kappa T^{3/2} e^{-\lambda_B T/2}]} |V_h^T| = \sup_{s \in [0,1]} |V_{h_s}^T|,$$

where

$$V_h^T = e^{-\lambda_B T/2} \sum_{u \in \mathring{\mathcal{T}}_T} h^{1/2}\overline{fg_h}(\zeta_u), \;\; \text{with} \;\; h_s = (h_T(\beta) - \kappa T^{3/2} e^{-\lambda_B T/2}) + 2s\kappa T^{3/2} e^{-\lambda_B T/2}, \;\; s \in [0,1].$$

Step 2. We claim that

$$\mathbb{E}\big[\big(V_{h_t}^T - V_{h_s}^T\big)^2\big] \leq C'(t-s)^2 \;\; \text{for} \;\; s, t \in [0,1], \tag{36}$$

for some constant $C' > 0$ that does not depend on T or $B \in \mathcal{B}$. Then, by Kolmogorov continuity criterion, this implies in particular that

$$\limsup_{T \to \infty} \sup_{B \in \mathcal{B}} \mathbb{E}\big[\sup_{s \in [0,1]} |V_{h_s}^T|\big] < \infty$$

hence the result (see for instance [20] to track the constant and obtain a uniform version of the continuity criterion). We have

$$V_{h_t}^T - V_{h_s}^T = e^{-\lambda_B T/2} \sum_{u \in \mathring{\mathcal{T}}_T} \Big(\Delta_{s,t}(h^{1/2} f g_h)(\zeta_u) - \mathring{\mathcal{E}}\big(\Delta_{s,t}(h^{1/2} f g_h)\big)\Big)$$

where $\Delta_{s,t}(h^{1/2}fg_h)(y) = h_t^{1/2}f(y)g_{h_t}(y) - h_s^{1/2}f(y)g_{h_s}(y)$. By Proposition 16, we derive

$$\mathbb{E}\big[\big(V_{h_t}^T - V_{h_s}^T\big)^2\big] \lesssim \Phi\big(\Delta_{s,t}\big(h^{1/2}fg_h\big)\big) + e^{-\lambda_B T}|\Delta_{s,t}\big(h^{1/2}fg_h\big)|_\infty^2. \tag{37}$$

and the remainder of the proof amounts to check that each term in the definition of Φ in (33) with test function $\Delta_{s,t}(h^{1/2}fg_h)$ together with the second term in the right-hand side of (37) have all the right order.

Step 3. For every y, we have

$$\Delta_{s,t}\big(h^{1/2}fg_h\big)(y) = (h_t - h_s)\partial_h\big(h^{1/2}f(y)g_h(y)\big)_{|h=h^\star(y)}$$

for some $h^\star(y) \in [h_t, h_s]$. Observe now that since $g \in \mathcal{C}^1$ and $f \in \mathcal{L}$, we have

$$\partial_h\big(h^{1/2}fg_h(y)\big) = -\frac{1}{2}h^{-3/2}f(y)g\big(h^{-1}(x-y)\big) - h^{-5/2}(x-y)f(y)g'\big(h^{-1}(x-y)\big)$$

therefore, for small enough h (which is always the case for T large enough, uniformly in $B \in \mathcal{B}$), we obtain the estimate

$$|\partial_h\big(h^{1/2}fg_h(y)\big)| \lesssim h^{-3/2}\mathbf{1}_{[0,C]}\big(h^{-1}(x-y)\big).$$

Assume with no loss of generality that $s \leq t$ so that $h_s \leq h(y)^\star \leq h_t$. It follows that

$$\begin{aligned}\big|\Delta_{s,t}\big(h^{1/2}fg_h\big)(y)\big| &\lesssim (h_t-h_s)h^\star(y)^{-3/2}\mathbf{1}_{[0,C]}\big(h^\star(y)^{-1}(x-y)\big)\\ &\leq (h_t-h_s)h_s^{-3/2}\mathbf{1}_{[0,C]}\big(h_t^{-1}(x-y)\big).\end{aligned}$$

Using that $h_t - h_s = 2(t-s)\kappa T^{3/2}e^{-\lambda_B T/2}$, we successively obtain the following estimates, for every $u > 0$:

$$\begin{aligned}
&e^{-\lambda_B T}\big|\Delta_{s,t}(h^{1/2}fg_h)\big|_\infty^2 \lesssim (h_t-h_s)^2e^{-\lambda_B T}h_s^{-3} \lesssim (t-s)^2T^3e^{\lambda_B(\frac{3}{2\beta+1}-2)T},\\
&\big|\Delta_{s,t}(h^{1/2}fg_h)\big|_2^2 \lesssim (h_t-h_s)^2h_s^{-3}h_t \lesssim (t-s)^2T^3e^{\lambda_B(\frac{2}{2\beta+1}-1)T},\\
&\big|\Delta_{s,t}(h^{1/2}fg_h)\big|_1^2\omega \lesssim (h_t-h_s)^2h_s^{-3}h_t^2\omega \lesssim (t-s)^2T^3\omega e^{\lambda_B(\frac{1}{2\beta+1}-1)T},\\
&\big|\Delta_{s,t}(h^{1/2}fg_h)\big|_\infty^2\omega^{-1} \lesssim (h_t-h_s)^2h_s^{-3}\omega^{-1} \lesssim (t-s)^2T^3\omega^{-1}e^{\lambda_B(\frac{3}{2\beta+1}-1)T},\\
&\big|\Delta_{s,t}(h^{1/2}fg_h)\big|_1\big|\Delta_{s,t}\big(h^{1/2}fg_h\big)\big|_\infty \lesssim (h_t-h_s)^2h_s^{-3/2}h_th_s^{-3/2} \lesssim (t-s)^2T^3e^{\lambda_B(\frac{2}{2\beta+1}-1)T}.
\end{aligned}$$

For the choice $\omega = h_T(\beta)^{-1}$, the above terms have maximal order $(t-s)^2T^3e^{\lambda_B(\frac{3}{2\beta+1}-1)T}$ and are thus all bounded by a constant times $(t-s)^2$, uniformly in T as soon as $\beta > 1$, a condition we have by assumption. So (36) is established and Proposition 18 is proved. $\square$

We now get rid of the preliminary estimators $\widehat{m}_T$ and $\widehat{\lambda}_T$:

Lemma 19. *Assume that $\beta > 1$. Let either $G_T(y) = g_{\widehat{h}_T}(y)$ with $g \in \mathcal{C}^1$ or $G_T(y) = \mathbf{1}_{\{y \leq x\}}$ for $y \in [0,\infty)$. Then*

$$e^{\lambda_B\frac{\beta}{2\beta+1}T}\big(\mathcal{E}^T(\mathring{\mathcal{T}}_T, \widehat{m}_T^{-1}e^{\widehat{\lambda}_T\cdot}G_T) - \mathcal{E}^T\big(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B\cdot}G_T)\big)$$

is $\mathcal{B}$-tight for the parameter B.

Proof. Define

$$\gamma_T(u) = e^{\lambda_B\frac{\beta}{2\beta+1}T}\big(\widehat{m}_T^{-1}e^{\widehat{\lambda}_T\zeta_u} - m^{-1}e^{\lambda_B\zeta_u}\big)G_T(\zeta_u).$$

Lemma 19 amounts to show that $|\mathring{\mathcal{T}}_T|^{-1}\sum_{u\in\mathring{\mathcal{T}}_T}\gamma_T(u)$ is $\mathcal{B}$-tight. Set $h_T(\beta) = \exp(-\lambda_B\frac{1}{2\beta+1}T)$ and note that

$$e^{\lambda_B\frac{\beta}{2\beta+1}T} = e^{\lambda_B T/2}h_T(\beta)^{1/2}.$$

We first treat the case $G_T(y) = g_{\widehat{h}_T}(y)$.

Step 1. By Proposition 7, we have

$$\widehat{\lambda}_T = \lambda_B + T^{1/2}e^{-\lambda_B T/2}r_T \text{ and } \widehat{m}_T^{-1} = m^{-1} + e^{-\lambda_B T/2}r'_T,$$

where both r_T and r'_T are $\mathcal{B}$-tight. We then have the decomposition $\gamma_T(u) = I + II$, with

$$I = T^{1/2}h_T(\beta)^{1/2}\widehat{m}_T^{-1}r_T\zeta_u e^{\vartheta_T\zeta_u}g_{\widehat{h}_T}(\zeta_u),$$

where $\vartheta_T \in [\min\{\lambda_B, \widehat{\lambda}_T\}, \max\{\lambda_B, \widehat{\lambda}_T\}]$, and

$$II = h_T(\beta)^{1/2}e^{\lambda_B\zeta_u}r'_T g_{\widehat{h}_T}(\zeta_u).$$

Since $g \in \mathcal{C}^1 \subset \mathcal{C}$ and $\widehat{m}_T^{-1}$, ϑ_T and $\widehat{h}_T$ are $\mathcal{B}$-tight, we can write

$$|I| \leq T^{1/2}h_T(\beta)^{1/2}\widehat{m}_T^{-1}r_T(C\widehat{h}_T + x)e^{\vartheta_T(C\widehat{h}_T+x)}|g_{\widehat{h}_T}(\zeta_u)| \leq T^{1/2}h_T(\beta)^{1/2}|g_{\widehat{h}_T}(\zeta_u)|\widetilde{r}_T$$

and

$$|II| \leq h_T(\beta)^{1/2}e^{\lambda_B(C\widehat{h}_T+x)}r'_T|g_{\widehat{h}_T}(\zeta_u)| \leq h_T(\beta)^{1/2}|g_{\widehat{h}_T}(\zeta_u)|\widetilde{r}'_T$$

where $\widetilde{r}_T$ and $\widetilde{r}'_T$ are tight uniformly in $B \in \mathcal{B}$.

Step 2. We are left to proving the tightness of $T^{1/2}h_T(\beta)^{1/2}|g_{\widehat{h}_T}(\zeta_u)|$ when averaging over $\mathring{\mathcal{T}}_T$ that it to say the tightness of $T^{1/2}h_T(\beta)^{1/2}\mathcal{E}^T(\mathring{\mathcal{T}}_T, |g_{\widehat{h}_T}|)$. We plan to use Proposition 18. For $\varepsilon > 0$, on the event

$$\mathcal{A}_{T,\varepsilon} = \big\{|\widehat{h}_T - h_T(\beta)| \leq \varepsilon\big\},$$

we have

$$T^{1/2}h_T(\beta)^{1/2}\mathcal{E}^T(\mathring{\mathcal{T}}_T, |g_{\widehat{h}_T}|) \leq III + IV,$$

with

$$III = T^{1/2}h_T(\beta)^{1/2}\sup_{h\in[h_T(\beta)\pm\varepsilon]}\mathring{\mathcal{E}}(|g_h|)$$

and

$$IV = T^{1/2}h_T(\beta)^{1/2}(h_T(\beta)-\varepsilon)^{-1/2}\sup_{h\in[h_T(\beta)\pm\varepsilon]}\big|\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, h^{1/2}|g_h|\big) - \mathring{\mathcal{E}}(h^{1/2}|g_h|)\big|.$$

Concerning the main term III, we write

$$\begin{aligned}\mathring{\mathcal{E}}(|g_h|) &= m\int_0^\infty h^{-1}|g\big(h^{-1}(x-y)\big)|e^{-\lambda_B y}f_B(y)dy\\ &\leq m\sup_y\big(e^{-\lambda_B y}f_B(y)\big)\int_0^\infty |g(y)|dy \lesssim 1\end{aligned}$$

since $B \in \mathcal{B}$, so we have a bound that does not depend on h and we readily conclude $III \lesssim 1$ on $\mathcal{A}_{T,\varepsilon}$. For the remainder term IV, we pick $\varepsilon = \varepsilon_T = \kappa T^{3/2}e^{-\lambda_B T/2}$ for an arbitrary $\kappa > 0$ and we can apply Proposition 18 since $\beta > 1$. For this choice, we also have $\varepsilon_T \ll h_T(\beta)$ so we obtain the $\mathcal{B}$-tightness of IV (that actually goes to 0 at a fast rate) on $\mathcal{A}_{T,\varepsilon_T}$.

Step 3. It remains to control the probability of $\mathcal{A}_{T,\varepsilon_T}$. By Proposition 7, we have $\widehat{\lambda}_T = \lambda_B + T^{1/2}e^{-\lambda_B T/2}r_T$, where r_T is $\mathcal{B}$-tight. It follows that

$$\mathbb{P}(\mathcal{A}_{T,\varepsilon_T}^c) = \mathbb{P}\big(T^{-3/2}e^{\lambda_B T/2}\big|e^{-\widehat{\lambda}_T\frac{1}{2\beta+1}T} - e^{-\lambda_B\frac{1}{2\beta+1}T}\big| \geq \kappa\big) = \mathbb{P}\big(\big|\tfrac{1}{2\beta+1}r_T e^{-\vartheta_T T}\big| \geq \kappa\big)$$

where $\vartheta_T \in [\min\{\lambda_B, \widehat{\lambda}_T\}, \max\{\lambda_B, \widehat{\lambda}_T\}]$ and r_T is tight, so this term can be made arbitrarily small by taking κ large enough.

The case $G_T(y) = \mathbf{1}_{\{y \leq x\}}$ is obtained in the same way and is actually much simpler, since there is no factor $\widehat{h}_T^{-1}$ in the Step 2 which is therefore straightforward and there is also no need for a Step 3. We omit the details. $\square$

Proof of Theorem 9. We are ready to prove the main result of the paper. The key ingredient is Proposition 18.

Step 1. In view of Lemma 19 with test function $g = K$, it is now sufficient to prove Theorem 9 replacing $\widehat{B}_T(x)$ by $\widetilde{B}_T(x)$, with

$$\widetilde{B}_T(x) = \frac{\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot} K_{\widehat{h}_T}(x-\cdot)\big)}{1-\mathcal{E}^T(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot}\mathbf{1}_{\{\cdot \leq x\}})}.$$

Since $(x,y) \rightsquigarrow x/(1-y)$ is Lipschitz continuous on compact sets that are bounded away from $\{y=1\}$, this simply amounts to show the $\mathcal{B}$-tightness of

$$e^{\lambda_B \frac{\beta}{2\beta+1}T}\big(\mathcal{E}^T(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot}\mathbf{1}_{\{\cdot \leq x\}}) - \mathring{\mathcal{E}}(m^{-1}e^{\lambda_B \cdot}\mathbf{1}_{\{\cdot \leq x\}}\big) \tag{38}$$

and

$$e^{\lambda_B \frac{\beta}{2\beta+1}T}\big(\mathcal{E}^T\big(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot}K_{\widehat{h}_T}(x-\cdot)\big) - f_B(x)\big). \tag{39}$$

We readily obtain the $\mathcal{B}$-tightness of (38) by applying Theorem 5 with test function $g(y) = m^{-1}e^{\lambda_B y}\mathbf{1}_{\{y \leq x\}}$ (we even have the convergence to 0).

Step 2. We turn to the main term (39). First, for $h>0$, introduce the notation

$$K_h f_B(x) = \mathring{\mathcal{E}}(m^{-1}e^{\lambda_B \cdot}K_h) = \int_0^\infty K_h(x-y)f_B(y)dy.$$

For $\kappa > 0$ and $\varepsilon_T = \kappa T e^{-\lambda_B T/2}$, on the event $\mathcal{A}_{T,\varepsilon_T} = \big\{|\widehat{h}_T - h_T(\beta)| \leq \varepsilon_T\big\}$, by introducing the approximation term $K_h f_B(x)$, we obtain a bias-variance bound that reads

$$\big|\mathcal{E}^T(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot}K_{\widehat{h}_T}) - f_B(x)\big| \leq I + II,$$

with

$$I = \sup_{h \in [h_T(\beta) \pm \varepsilon_T]} \big|K_h f_B(x) - f_B(x)\big|$$

and

$$II = \sup_{h \in [h_T(\beta) \pm \varepsilon_T]} \big|\mathcal{E}^T(\mathring{\mathcal{T}}_T, m^{-1}e^{\lambda_B \cdot}K_h) - \mathring{\mathcal{E}}(m^{-1}e^{\lambda_B \cdot}K_h)\big|.$$

The term I is treated by the following classical argument in nonparametric estimation: since $B \in \mathcal{H}^\beta_{\mathcal{D}}(L)$ we also have $f_B \in \mathcal{H}^\beta_{\mathcal{D}}(L')$ for another constant L' that only depends on $\mathcal{D}, L$ and β. Write $\beta = n + \{\beta\}$ with n a non-negative integer. By a Taylor expansion up to order n (recall that the number n_0 of vanishing moments of K in Assumption 8 satisfies $n_0 > \beta$), we obtain

$$I \lesssim \sup_{h \in [h_T(\beta) \pm \varepsilon_T]} h^\beta = (e^{-\lambda_B \frac{1}{2\beta+1}T} + \varepsilon_T)^\beta,$$

see for instance, Proposition 1.2 in Tsybakov [22]. The choice of ε_T shows that this term has the right order.

Step 3. We further bound the term II on $\mathcal{A}_{T,\varepsilon_T}$ as follows:

$$|II| \leq (h_T(\beta) - \varepsilon_T)^{-1/2} \sup_{h \in [h_T(\beta) \pm \varepsilon_T]} \big|\mathcal{E}^T(\mathring{\mathcal{T}}_T, h^{1/2} m^{-1} e^{\lambda_B \cdot} K_h) - \mathring{\mathcal{E}}(h^{1/2} m^{-1} e^{\lambda_B \cdot} K_h)\big|.$$

By assumption, we have $\beta > 1$, so by Proposition 18 applied to $f(y) = m^{-1}e^{\lambda_B y}\mathbf{1}_{\{y \leq x+C\}} \in \mathcal{L}_{C+x}$ and $g = K \in \mathcal{C}_{C+x}$ and using $\varepsilon_T \ll h_T(\beta)$, we conclude that $e^{\lambda_B T/2} h_T(\beta)^{1/2}|II|$ is $\mathcal{B}$-tight and this term also has the right-order.

Step 4. It remains to control the probability of $\mathcal{A}_{T,\varepsilon_T}$. This is done exactly in the same way as for Step 4 in the proof of Lemma 19. □

5.5. **Proof of Theorem 10.** We will prove actually a slightly stronger result, by restricting the supremum in B over a neighbourhood of an arbitrary function B_0, provided B_0 is slightly smoother in $\mathcal{H}^\beta_\mathcal{D}$ norm and not identically equal to the maximal element of $\mathcal{B}_b$.

Remember that the evolution of the Bellman-Harris model can be described by a piecewise deterministic Markov process

$$X(t) = \big(X_1(t), X_2(t), \dots\big), \quad t \geq 0$$

with values in $\mathcal{S} = \bigcup_{k \geq 1} [0,\infty)^k$ and where the $X_i(t)$ denote the (ordered) ages of the living particles at time t. Following Löcherbach [17], we set $\mathbb{D}([0,\infty),\mathcal{S})$ for the Skorokhod space of càdlàg functions $\varphi : [0,\infty) \to \mathcal{S}$ and introduce the subset $\Omega \subset \mathbb{D}([0,\infty),\mathcal{S})$ of functions φ such that:

(i) There is an increasing sequence of jump times $T_0 = 0 < T_1 < T_2 < \cdots$ such that the restriction $\varphi_{\big|[T_k,T_{k+1})}$ is continuous with values in $[0,\infty)^{l_{k,\varphi}}$ for some $l_{k,\varphi} \geq 0$ and every $k \geq 0$.

(ii) We have $\ell\big(\varphi(T_k)\big) \neq \ell\big(\varphi(T_{k+1})\big)$ for every $k \geq 0$, where we set $\ell(x) = \sum_{k \geq 0} k \mathbf{1}_{\{x \in [0,\infty)^k\}}$ for $x \in \mathcal{S}$.

We endow Ω with its Borel sigma-field $\mathcal{F}$, its canonical process $X_t(\varphi) = (\varphi_1(t), \varphi_2(t), \dots)$ and its canonical filtration $(\mathcal{F}_t)_{t \geq 0}$ (modified in order to be right-continuous). By Proposition 3.3 of Löcherbach [17], there is a unique probability measure $\mathbb{P}_B$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$ such that X is strongly Markov under $\mathbb{P}_B$ with $\mathbb{P}_B(X(0) = 0) = 1$ (*i.e.* we start with one common ancestor with age 0 at time 0) and such that the random continuous time rooted tree associated to X via

$$\sum_{i \geq 1} \mathbf{1}_{\{X_i(t) > 0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{T}} \mathbf{1}_{\{t \in [b_u, d_u)\}} \delta_{t - b_u}$$

is a Harris-Bellman process according to Definition 1. The strategy for proving the lower bound is a classical two point information inequality: we nevertheless need to be careful since the target lower bound rate $e^{-\lambda_B \frac{\beta}{2\beta+1} T}$ is parameter dependent in a non-trivial way.

Step 1. Let $\delta > 0$. Fix $B_0 \in \mathcal{B}_b \cap \mathcal{H}^\beta_\mathcal{D}(L - \delta)$ and $x \in \mathcal{D}$. Then, for large enough T, setting $h_T(B) = e^{-\lambda_B \frac{1}{2\beta+1} T}$, we construct a perturbation B_T of B_0 defined by

$$B_T(y) = B_0(y) + a h_T(B_0)^{\beta+1} K_{h_T(B_0)}\big(y - x\big), \quad y \in [0,\infty),$$

for some nonnegative smooth kernel K with compact support, such that $K(0) = 1$, and for some $a = a_{\delta,K} > 0$ chosen in such a way that $B_T \in \mathcal{B}_b \cap \mathcal{H}^\beta_\mathcal{D}(L)$ for every $T \geq 0$. Such a choice is always possible (if $B_0 \neq mb/(m-1)$ identically in a neighbourhood of x, which we may an will assume from now on) thanks to the assumption $\|B_0\|_{\mathcal{H}^\beta_\mathcal{D}} \leq L-\delta$; it suffices then to impose $\|ah_T^{\beta+1}K_{h_T}(\cdot - x)\|_{\mathcal{H}^\beta_\mathcal{D}} \leq \delta$ which is easily obtained by picking $a_{\delta,K}$ sufficiently small.

Also, by construction, we have $B_0(y) \leq B_T(y)$ for every $y \geq 0$ hence $\lambda_{B_0} \leq \lambda_{B_T}$, compare the proof of Proposition 13 (ii) and at $y = x$, the lower estimate $|B_0(x) - B_T(x)| = a_{\delta,K}h_T^\beta(B_0)$ holds, and this quantity is of order $e^{-\lambda_{B_0}\frac{\beta}{2\beta+1}T}$.

Step 2. Abusing notation slightly, we further write $\mathbb{P}_B$ for $\mathbb{P}_{B|\mathcal{F}_T}$, *i.e.* the measure in restriction to the sigma field generated by the observation $(X(t))_{0\leq t\leq T}$. Since $B_0, B_T \in \mathcal{B}_b \cap \mathcal{H}^\beta_\mathcal{D}(L)$, for an arbitrary estimator $\widehat{B}_T(x)$ and a constant $C > 0$ the maximal risk is bounded below by

$$\begin{aligned}
&\max_{B\in\{B_0,B_T\}} \mathbb{P}_B\big(e^{\lambda_B\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B(x)| \geq C\big)\\
\geq& \tfrac{1}{2}\Big(\mathbb{P}_{B_0}\big(e^{\lambda_{B_0}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B_0(x)| \geq C\big) + \mathbb{P}_{B_T}\big(e^{\lambda_{B_T}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B_T(x)| \geq C\big)\Big)\\
\geq& \tfrac{1}{2}\,\mathbb{E}_{B_0}\Big[\mathbf{1}_{\big\{e^{\lambda_{B_0}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x)-B_0(x)|\geq C\big\}} + \mathbf{1}_{\big\{e^{\lambda_{B_T}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x)-B_T(x)|\geq C\big\}}\Big] - \tfrac{1}{2}\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}.
\end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned}
&e^{\lambda_{B_0}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B_0(x)| + e^{\lambda_{B_T}\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B_T(x)|\\
\geq& e^{\min\{\lambda_{B_0},\lambda_{B_T}\}\frac{\beta}{2\beta+1}T}|B_0(x) - B_T(x)| \geq a_{K,\delta}
\end{aligned}$$

by Step 1, so if we now take $C < a_{K,\delta}/2$, one of the two indicators within the expectation above must be equal to one with full $\mathbb{P}_{B_0}$-probability. In that case

$$\max_{B\in\{B_0,B_T\}} \mathbb{P}_B\big(e^{\lambda_B\frac{\beta}{2\beta+1}T}|\widehat{B}_T(x) - B(x)| \geq C\big) \geq \tfrac{1}{2}(1 - \|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV})$$

and Theorem 10 is thus proved if $\limsup_{T\to\infty}\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV} < 1$.

Step 3. By Pinsker's inequality, we have $\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV} \leq \frac{\sqrt{2}}{2}\Big(\mathbb{E}_{B_0}\Big[\log\frac{d\mathbb{P}_{B_0}}{d\mathbb{P}_{B_T}}\Big]\Big)^{1/2}$. By Theorem 3.5 in [17], the measures $\mathbb{P}_{B_0}$ and $\mathbb{P}_{B_T}$ are equivalent on $\mathcal{F}_T$ and we have

$$\log\Big(\frac{d\mathbb{P}_{B_T}}{d\mathbb{P}_{B_0}}\Big) = \sum_{u\in\mathring{\mathcal{T}}_T}\log\Big(\frac{B_T}{B_0}(\zeta_u)\Big) - \int_0^T\sum_{u\in\partial\mathcal{T}_s}(B_T - B_0)(\zeta_u^s)\,ds,$$

where ζ_u^t denotes the age of the cell u at time $t \in I_u = [b_u, d_u)$. Using $-\log(1+x) \leq x^2 - x$ if $x \geq -1/2$ and setting $\varepsilon_T(y) = a_{K,\delta}h_T(B_0)^{\beta+1}K_{h_T(B_0)}\big(y - x\big)$, we further infer

$$\begin{aligned}
\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}^2 &\leq \frac{1}{2}\Big(\mathbb{E}_{B_0}\Big[\sum_{u\in\mathring{\mathcal{T}}_T}\frac{\varepsilon_T^2}{B_0^2}(\zeta_u)\Big] - \mathbb{E}_{B_0}\Big[\sum_{u\in\mathring{\mathcal{T}}_T}\frac{\varepsilon_T}{B_0}(\zeta_u)\Big] + \int_0^T\mathbb{E}_{B_0}\Big[\sum_{u\in\partial\mathcal{T}_s}\varepsilon_T(\zeta_u^s)\Big]ds\Big)\\
&= \frac{1}{2m}\int_0^T e^{\lambda_{B_0}s}\,\mathbb{E}_{B_0}\Big[\frac{\varepsilon_T^2}{B_0^2}(\widetilde{\chi}_s)H_{B_0}(\widetilde{\chi}_s)\Big]ds
\end{aligned}$$

by (21) and (22) in Proposition 11 and the fact that the last two terms cancel. We now use the same kind of estimates as in the proof of Proposition 16, Step 1 with test function $g = \varepsilon_T/B_0$ to finally get

$$\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}^2 \lesssim e^{\lambda_{B_0}T}\big|B_0^{-1}\varepsilon_T\big|_2^2 + \big|B_0^{-1}\varepsilon_T\big|_\infty^2 \lesssim a_{K,\delta}^2$$

and this term can be made arbitrarily small by picking $a_{K,\delta}$ small enough.

6. Appendix

6.1. **Proof of Proposition 11.** We start with a continuous time rooted tree which is a Bellman Harris process in the sense of Definition 1, so we have random variables $(\zeta_u, \nu_u, u \in \mathcal{U})$ satisfying properties (i), (ii) and (iii) of the definition. For $u \in \mathcal{U}$, and $t \geq 0$, let $\Lambda_t^u = \sum_{v \prec u(t)} \log(\nu_v), t \geq 0$ denote the process that encodes the birth times and the numbers of offsprings of the ancestors of u. Let $\vartheta = (\vartheta_k)_{k \geq 0}$ with $\vartheta_k \in \mathcal{U}$ be such that $|\vartheta_k| = k$ for $k \geq 1$ (with $\vartheta_0 = \varnothing$) and $\vartheta_k \preceq \vartheta_l$ for $k \leq l$. We associate to ϑ a counting process $(N_t)_{t \geq 0}$ via the relationship

$$b_{\vartheta_{N_t}} \leq t < d_{\vartheta_{N_t}}, \quad t \geq 0.$$

This enables us to further obtain a "tagged process of age" such that $\chi_t = \zeta^t_{\vartheta_{N_t}}$ for $t \in I_{\vartheta_{N_t}}$ and also a process $(\Lambda_t)_{t \geq 0}$ that encodes the genealogy of the tagged branch

$$\Lambda_t = \sum_{k=0}^{N_t} \log(\nu_{\vartheta_k}), \quad t \geq 0.$$

Step 1. Let us pick ϑ at random along the genealogical tree $\mathcal{T}$. This means that if $\mathcal{H}_n$ denotes the sigma-field generated by $(\zeta_u, \nu_u, u \in \mathcal{T}, |u| \leq n)$, then on the event $\{t \in I_u\}$ (*i.e.* the particle u is living at time t), we have (or rather, we set)

$$\mathbb{P}\big(\vartheta_{N_t} = u \big| \mathcal{H}_{|u|}\big) = \prod_{v \preceq u} \frac{1}{\nu_v} = e^{-\Lambda_t^u}.$$

It is not difficult to see that $(\chi_t)_{t \geq 0}$ is a Markov process with generator $\mathcal{A}_B$. By definition of $(\chi_t)_{t \geq 0}$ and $(\Lambda_t)_{t \geq 0}$, it follows that $\mathbb{E}[e^{\Lambda_T} g(\chi_T)]$ can be rewritten as

$$\sum_{u \in \mathcal{U}} \mathbb{E}[e^{\Lambda_T} g(\chi_T) \mathbf{1}_{\{T \in I_u, u = \vartheta_{N_T}\}}] = \sum_{u \in \mathcal{U}} \mathbb{E}[e^{\Lambda_T^u} g(\zeta_u^T) \mathbf{1}_{\{T \in I_u, u = \vartheta_{N_T}\}}] = \sum_{u \in \mathcal{U}} \mathbb{E}[g(\zeta_u^T) \mathbf{1}_{\{T \in I_u\}}],$$

where the last equality is obtained by conditioning with respect to $\mathcal{H}_{|u|}$.

Step 2. For $j \geq 1$, let $\tau_j = \inf\{t \geq 0, N_t \geq j\} - \inf\{t \geq 0, N_t \geq j-1\}$ denote the durations between the jumps of $(\chi_t)_{t \geq 0}$, so that

$$e^{\Lambda_T} g(\chi_T) = \sum_{k=0}^{\infty} e^{\sum_{j=1}^k \log(\nu_{\vartheta_j})} g(T - \sum_{j=1}^k \tau_j) \mathbf{1}_{\left\{ \sum_{j=1}^k \tau_j \leq T < \sum_{j=1}^{k+1} \tau_j \right\}}.$$

By properties (i)-(iii) of Definition 1, the τ_i are independent with common distribution $f_B(x)dx$, and independent of the ν_{ϑ_k} that are independent with common distribution $(p_k)_{k \geq 1}$. We thus infer that $\mathbb{E}[e^{\Lambda_T} g(\chi_T)]$ is equal to

$$\sum_{k=0}^{\infty} \sum_{h_j \geq 1, j \leq k} e^{\sum_{j=1}^k \log(h_j)} \prod_{j=1}^k p_{h_j} \int_{[0,\infty)^{k+1}} g(T - \sum_{j=1}^k t_j) \mathbf{1}_{\{\sum_{j=1}^k t_j \leq T < \sum_{j=1}^{k+1} t_j\}} \prod_{j=1}^{k+1} f_B(t_j) dt_1 \dots dt_{k+1}.$$

We set $\overline{F}_B(x) = 1-\int_0^x f_B(y)dy$ and $q_k = m^{-1}kp_k$, so that $(q_k)_{k\geq 1}$ defines a probability distribution. Using $f_{H_B}(x) = me^{-\lambda_B x}f_B(x)$, we can rewrite the preceding formula so that

$$e^{-\lambda_B T}\mathbb{E}\big[e^{\Lambda_T}g(\chi_T)\big] = \sum_{k=0}^{\infty}\sum_{h_j\geq 1, j\leq k}\prod_{j=1}^{k} q_{h_j}\int_{[0,\infty)^k} g(T-\sum_{j=1}^{k}t_j)\mathbf{1}_{\{T-\sum_{j=1}^k t_j\geq 0\}}e^{-\lambda_B(T-\sum_{j=1}^k t_j)}$$
$$\times\, \overline{F}_B(T-\sum_{j=1}^{k}t_j)\prod_{j=1}^{k}f_{H_B}(t_j)dt_1\dots dt_k.$$

Step 3. Putting $W_B(x) = me^{-\lambda_B x}\overline{F}_B(x)/\overline{F}_{H_B}(x)$, we finally obtain the representation

$$e^{-\lambda_B T}\mathbb{E}\big[e^{\Lambda_T}g(\chi_T)\big] = \frac{1}{m}\mathbb{E}\Big[g(\widetilde{\chi}_T)W_B(\widetilde{\chi}_T)\Big],$$

where $(\widetilde{\chi}_t)_{t\geq 0}$ is a Markov process with generator $\mathcal{A}_{H_B}$ that can be constructed in the same way as $(\chi_t)_{t\geq 0}$, substituting f_B by f_{H_B}. Straightfoward computations give $W_B(x) = \frac{H_B(x)}{B(x)}$. Putting together all the three steps, we have proved

$$\sum_{u\in\mathcal{U}}\mathbb{E}\big[g(\zeta_u^T)\mathbf{1}_{\{T\in I_u\}}\big] = \mathbb{E}\big[e^{\Lambda_T}g(\chi_T)\big] = \frac{e^{\lambda_B T}}{m}\mathbb{E}\Big[g(\widetilde{\chi}_T)\frac{H_B(\widetilde{\chi}_T)}{B(\widetilde{\chi}_T)}\Big].$$

Noticing that $\sum_{u\in\mathcal{U}}\mathbb{E}[g(\zeta_u^T)\mathbf{1}_{\{T\in I_u\}}]$ is nothing but $\mathbb{E}\big[\sum_{u\in\partial\mathcal{T}_T}g(\zeta_u^T)\big]$ establishes (21).

Step 4. By definition of the set $\mathring{\mathcal{T}}_T$,

$$\mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T}g(\zeta_u)\Big] = \sum_{u\in\mathcal{U}}\mathbb{E}\Big[g(\zeta_u)\mathbf{1}_{\{b_u+\zeta_u\leq T\}}\mathbf{1}_{\{u\in\mathcal{T}\}}\Big].$$

We denote by $\mathcal{F}_t$ the sigma-field generated by $(\zeta_u^s, u\in\partial\mathcal{T}_s, s\leq t)$ and we note that $d_u\mathbf{1}_{\{u\in\mathcal{T}\}}$ is a stopping time for the filtration $(\mathcal{F}_t)_{t\geq 0}$. Conditioning w.r.t $\mathcal{F}_{b_u}$, using that the ζ_u are independent of $\mathcal{F}_{b_u}$, we successively obtain

$$\begin{aligned}
\mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T}g(\zeta_u)\Big] &= \sum_{u\in\mathcal{U}}\mathbb{E}\Big[\mathbf{1}_{\{u\in\mathcal{T}\}}\int_0^\infty g(x)\mathbf{1}_{\{b_u+x\leq T\}}B(x)e^{-\int_0^x B(y)dy}dx\Big] \\
&= \sum_{u\in\mathcal{U}}\mathbb{E}\Big[\mathbf{1}_{\{u\in\mathcal{T}\}}\int_0^\infty\Big(\int_0^y g(x)B(x)\mathbf{1}_{\{b_u+x\leq T\}}dx\Big)B(y)e^{-\int_0^y B(z)dz}dy\Big] \\
&= \sum_{u\in\mathcal{U}}\mathbb{E}\Big[\mathbf{1}_{\{u\in\mathcal{T}\}}\int_0^{\zeta_u} g(x)B(x)\mathbf{1}_{\{b_u+x\leq T\}}dx\Big] \\
&= \sum_{u\in\mathcal{U}}\mathbb{E}\Big[\mathbf{1}_{\{u\in\mathcal{T}\}}\int_{b_u}^{d_u} g(\zeta_u^s)B(\zeta_u^s)\mathbf{1}_{\{s\leq T\}}ds\Big].
\end{aligned}$$

using that $\zeta_u^s = s-b_u$ for $s\in I_u$ in order to obtain the last equality. Finally, observing that $\{s\in I_u\} = \{u\in\partial\mathcal{T}_s\}$, we finally infer

$$\mathbb{E}\Big[\sum_{u\in\mathring{\mathcal{T}}_T}g(\zeta_u)\Big] = \int_0^\infty\mathbb{E}\Big[\sum_{u\in\partial\mathcal{T}_s}g(\zeta_u^s)B(\zeta_u^s)\Big]\mathbf{1}_{\{s\leq T\}}ds.$$

Using (21) completes the proof of (22).

6.2. **Proof of (24) of Proposition 12.** Whenever $(u,v)\in\mathcal{FU}$ there exist $w,\tilde u$ and $\tilde v\in\mathcal U$ together with integers $i\neq j$, such that $u=wi\tilde u$ and $v=wj\tilde v$. Conditioning w.r.t $\mathcal F_{d_w}$, using the branching property between descendents of w and the strong Markov property at time d_w, we have

$$\begin{aligned}\mathbb E\Big[\sum_{(u,v)\in\mathcal{FT}\cap\mathring{\mathcal T}_T^2} g(\zeta_u)g(\zeta_v)\Big] &= \sum_{(u,v)\in\mathcal{FU}}\mathbb E\Big[g(\zeta_u)\mathbf 1_{\{d_u<T\}}\mathbf 1_{\{u\in\mathcal T\}}g(\zeta_v)\mathbf 1_{\{d_v<T\}}\mathbf 1_{\{v\in\mathcal T\}}\Big]\\ &=\sum_{w\in\mathcal U}\sum_{i\neq j}\mathbb E\Big[\mathbb E\big[\sum_{\tilde u\in\mathcal U}g(\zeta_{wi\tilde u})\mathbf 1_{\{d_{wi\tilde u}<T\}}\mathbf 1_{\{wi\tilde u\in\mathcal T\}}\big|\mathcal F_{d_w}\big]\\ &\qquad\times\mathbb E\big[\sum_{\tilde v\in\mathcal U}g(\zeta_{wj\tilde v})\mathbf 1_{\{d_{wj\tilde v}<T\}}\mathbf 1_{\{wj\tilde v\in\mathcal T\}}\big|\mathcal F_{d_w}\big]\Big]\\ &=\sum_{w\in\mathcal U}\sum_{i\neq j}\mathbb E\Big[\mathbf 1_{\{wi\in\mathcal T,wj\in\mathcal T\}}\big(\mathbb E\big[\sum_{u\in\mathcal T}g(\zeta_u)\mathbf 1_{\{d_u<T-t\}}\big]_{|t=d_w}\big)^2\mathbf 1_{\{d_w<T\}}\Big].\end{aligned}$$

Notice that $\{wi\in\mathcal T,wj\in\mathcal T\}=\{w\in\mathcal T\}\cap\{i\leq\nu_w,j\leq\nu_w\}$, and ν_w is independent of d_w and has distribution $(p_k)_{k\geq1}$. We conclude by using (22) of Proposition 11 (slightly generalized for test functions that depend on d_u and ζ_u). Let us now turn to (25). For $u,v\in\mathcal T$ with $u\prec v$, we have $uiw=v$ for some $w\in\mathcal T$ and some integer i. It follows that

$$\begin{aligned}\mathbb E\big[\sum_{\substack{u,v\in\mathring{\mathcal T}_T,\\ u\prec v}}g(\zeta_u)g(\zeta_v)\big] &=\sum_{u\in\mathcal U}\sum_i\mathbb E\Big[g(\zeta_u)\mathbf 1_{\{d_u<T\}}\mathbf 1_{\{u\in\mathcal T\}}\mathbb E\big[\sum_{w\in\mathcal U}g(\zeta_{uiw})\mathbf 1_{\{d_{uiw}<T\}}\mathbf 1_{\{uiw\in\mathcal T\}}\,\big|\mathcal F^u_{d_u}\big]\Big]\\ &=\sum_{u\in\mathcal U}\sum_i\mathbb E\Big[g(\zeta_u)\mathbf 1_{\{0\leq d_u<T\}}\mathbf 1_{\{ui\in\mathcal T\}}\mathbb E\big[\sum_{w\in\mathcal T}g(\zeta_w)\mathbf 1_{\{d_w<T-s\}}\big]_{|s=d_u}\mathbf 1_{\{d_u<T\}}\Big]\end{aligned}$$

conditioning w.r.t. $\mathcal F^u_{d_u}$ on $\{d_u<T\}$ and applying the branching property. Next, we have

$$\mathbb E\big[\sum_{w\in\mathcal T}g(\zeta_w)\mathbf 1_{\{d_w<T-s\}}\big]=\mathbb E\big[\sum_{w\in\mathring{\mathcal T}_{T-s}}g(\zeta_w)\big]=\frac1m\int_0^{T-s}e^{\lambda_Bt}P^t_{H_B}\big(gH_B\big)(0)dt$$

by (22) of Proposition 11. Since $\{ui\in\mathcal T\}=\{i\leq\nu_u\}$, and ν_u is independent of ζ_u and d_u and has distribution with expectation m, we obtain

$$\mathbb E\big[\sum_{\substack{u,v\in\mathring{\mathcal T}_T,\\ u\prec v}}g(\zeta_u)g(\zeta_v)\big]=\mathbb E\Big[\sum_{u\in\mathring{\mathcal T}_T}g(\zeta_u)\int_0^{T-d_u}e^{\lambda_Bt}P^t_{H_B}\big(gH_B\big)(0)dt\Big]$$

and we conclude by using once more (22) of Proposition 11 (slightly generalized for test functions that depend on d_u and ζ_u).

6.3. **Proof of Lemma 17.** Let τ denote the first jump time of the process $\widetilde\chi_t$. Conditioning with respect to the event $\{\tau>t\}$ and applying the strong Markov property, we have

$$P^t_{H_B}\big(gH_B\big)(0)=g(t)H_B(t)\mathbb P(\tau>t)+\int_0^tP^{t-u}_{H_B}\big(gH_B\big)(0)f_{H_B}(u)du.$$

The function $t\rightsquigarrow u(t)=P^t_{H_B}\big(gH_B\big)(0)$ satisfies a renewal equation of the form $u=u_0+u\star f_{H_B}$, with locally bounded initial condition $u_0=gH_B\mathbb P(\tau>\cdot)$ and renewal distribution $f_{H_B}(y)dy$. Its solution is thus given by

$$P^t_{H_B}\big(gH_B\big)(0)=g(t)H_B(t)\mathbb P(\tau>t)+\int_0^tg(t-s)H_B(t-s)\mathbb P(\tau>t-s)d\mathbb E[\widetilde N_s],$$

where $\widetilde{N}_t = \sum_{s\leq t} \mathbf{1}_{\{\widetilde{\chi}_s - \widetilde{\chi}_{s-} > 0\}}$ is the counting process associated to $(\widetilde{\chi}_t)_{t\geq 0}$. By construction, we have $\mathbb{E}[\widetilde{N}_t] = \mathbb{E}\big[\int_0^t H_B(\widetilde{\chi}_s)ds\big]$ and $\mathbb{P}(\tau > t) = \int_t^\infty f_{H_B}(y)dy = m\int_t^\infty e^{-\lambda_B y} f_B(y)dy \leq me^{-\lambda_B t}$, therefore

$$|P^t_{H_B}\big(gH_B\big)(0)| \leq |g(t)|e^{-\lambda_B t} m|H_B|_\infty + |H_B|^2_\infty \int_0^t |g(u)|du$$

and we obtain the desired estimate thanks to the fact that H_B is uniformly bounded over $\mathbb{B}$.

Acknowledgements. We are grateful to V. Bansaye. M. Doumic for helpful discussion and comments. The illuminating coupling argument for proving Proposition 13 was indicated to us by N. Fournier. Part of this work was completed while M.H. was visiting Humboldt-Universität zu Berlin. The research of M.H. is partly supported by the Agence Nationale de la Recherche, (Blanc SIMI 1 2011 project CALIBRATION).

References

[1] K. B. Athreya and N. Keiding. *Estimation theory for continuous-time branching processes.* Sankhya: The Indian Journal of Statistics, Series A 39 (1977), 101–123.
[2] K. B. Athreya and P. Ney. *Branching processes.* Springer-Verlag, New-York, 1972.
[3] V. Bansaye, J.-F. Delmas, L. Marsalle and V. C. Tran. *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees.* The Annals of Applied Probability, 21 (2011) 2263–2314.
[4] S. V. Bisteki-Penda, H. Djellout and A. Guillin. *Deviations inequalities, moderate deviations and some limits theorems for bifurcating Markov chains with application.* Annals of Applied Probability 24 (2014) 235-291.
[5] B. Cloez. *Limit theorems for some branching measure-valued processes*, hal-00598030 (2011).
[6] J.-F. Delmas and L. Marsalle. *Detection of cellular aging in a Galton–Watson process.* Stochastic Processes and their Applications, 120 (2010) 2495–2519.
[7] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree.* Bernoulli, to appear.
[8] M. Doumic, M. Hoffmann, P. Reynaud-Bouret and V. Rivoirard. *Nonparametric estimation of the division rate of a size-structured population.* SIAM Journal on Numerical Analysis, 50 (2012) 925–950.
[9] M. Doumic, B. Perthame and J. P. Zubelli. *Numerical solution of an inverse problem in size-structured population dynamics.* Inverse Problems, 25 (2009) 25pp.
[10] P. Guttorp. *Statistical Inference for Branching Processes.* Wiley, 1991.
[11] T. Harris. *The theory of branching processes.* Springer-Verlag, New-York, 1963.
[12] M. Hoffmann and N. Krell. *Statistical analysis of self-similar fragmentation chains.* Bernoulli, 17 (2011) 395–423.
[13] R. Höpfner, M. Hoffmann and E. Löcherbach. *Nonparametric estimation of the death rate in branching diffusions.* Scandinavian Journal of Statistics, 29 (2002) 665-690.
[14] O. Hyrien. *Pseudo-likelihood estimation for discretely observed multitype Bellman-Harris branching processes.* Journal of Statistical Planning and Inference, 137 (2007) 1375 – 1388.
[15] R. Johnson, V. Susarla and J. van Ryzin. *Bayesian nonparametric estimation for age-dependent branching processes.* Stochastic Processes and their Applications, 9 (1979) 307 – 318.
[16] J. Kiefer *Conditional inference.* In: Encyclopedia of Statistical Science, Volume 2 (1972) 103–109, John Wiley, New York.
[17] E. Löcherbach. *Likelihood ratio processes for Markovian particle systems with killing and jumps.* Statistical inference for stochastic processes, 5 (2002a) 153–177.
[18] E. Löcherbach. *LAN and LAMN for systems of interacting diffusions with branching and immigration.* Annales of the Institute Henri Poincaré, 38 (2002b) 59–90.
[19] B. Perthame. *Transport equations arising in biology.* Birckhäuser Frontiers in mathematics edition, 2007.
[20] R. L. Schilling. *Sobolev embeddings for stochastic processes.* Expositiones Mathematicae, 18 (2000) 239–242.
[21] B. Tsirelson *From uniform renewal theorem to uniform large and moderate deviations for renewal-reward processes.* Electronic Communication in Probability, 18 (2013) 1–13.
[22] A. Tsybakov. *Introduction to nonparametric estimation.* Springer series in statistics, Springer-Verlag, New-York, 2009.

Marc Hoffmann, CEREMADE, CNRS-UMR 7534, Université Paris Dauphine, Place du maréchal De Lattre de Tassigny 75775 Paris Cedex 16, France.
E-mail address: hoffmann@ceremade.dauphine.fr

Adélaïde Olivier, CEREMADE, CNRS-UMR 7534 and CREST-ENSAE, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France.
E-mail address: adelaide.olivier@ensae.fr