Bernoulli **21**(3), 2015, 1760–1799
DOI: 10.3150/14-BEJ623

Statistical estimation of a growth-fragmentation model observed on a genealogical tree

MARIE DOUMIC[1], MARC HOFFMANN[2], NATHALIE KRELL[3] and LYDIA ROBERT[4,5]

[1]*INRIA Rocquencourt, projet BANG, Domaine de Voluceau, BP 105, 781153 Rocquencourt, France and Laboratoire J.L. Lions, CNRS-UMR 7598, 4 place Jussieu, 75005 Paris, France. E-mail: marie.doumic-jauffret@inria.fr*
[2]*Université Paris-Dauphine, CNRS-UMR 7534, Place du maréchal De Lattre de Tassigny, 75775 Paris Cedex 16, France. E-mail: hoffmann@ceremade.dauphine.fr*
[3]*Université de Rennes 1, CNRS-UMR 6625, Campus de Beaulieu, 35042 Rennes Cedex, France. E-mail: nathalie.krell@univ-rennes1.fr*
[4]*INRA, UMR 1319 Micalis, F-78350 Jouy-en-Josas, France*
[5]*AgroParisTech, UMR Micalis, F-78350 Jouy-en-Josas, France. E-mail: lydia.robert@upmc.fr*

We raise the issue of estimating the division rate for a growing and dividing population modelled by a piecewise deterministic Markov branching tree. Such models have broad applications, ranging from TCP/IP window size protocol to bacterial growth. Here, the individuals split into two offsprings at a division rate $B(x)$ that depends on their size x, whereas their size grow exponentially in time, at a rate that exhibits variability. The mean empirical measure of the model satisfies a growth-fragmentation type equation, and we bridge the deterministic and probabilistic viewpoints. We then construct a nonparametric estimator of the division rate $B(x)$ based on the observation of the population over different sampling schemes of size n on the genealogical tree. Our estimator nearly achieves the rate $n^{-s/(2s+1)}$ in squared-loss error asymptotically, generalizing and improving on the rate $n^{-s/(2s+3)}$ obtained in (*SIAM J. Numer. Anal.* **50** (2012) 925–950, *Inverse Problems* **25** (2009) 1–22) through indirect observation schemes. Our method is consistently tested numerically and implemented on *Escherichia coli* data, which demonstrates its major interest for practical applications.

Keywords: cell division equation; growth-fragmentation; Markov chain on a tree; nonparametric estimation

1. Introduction

1.1. Size-structured models and their inference

Growth-fragmentation models and structured population equations describe the temporal evolution of a population characterised by state variables such as age, size, growth, maturity, protein content and so on – see [26,32] and the references therein. This field continues to grow in interest as its applications appear to be wider and wider, ranging from the internet TCP/IP window size protocol [1] to neuronal activity [31], protein polymerization [17], cell division cycle [3], phase transitions [30], parasite proliferation [4] etc.

In order to quantitatively fit experimental observations and thus validate the relevance of the models, developing new and well-adapted statistical methods appears to be one of the major challenge for the coming years. A paradigmatic example, which can serve both as a toy model and a proof of concept for the methodology we develop here, is given by the growth-fragmentation or size-structured cell division equation [33]. When applied to the evolution of a bacterial population, it reads

$$\begin{cases} \partial_t n(t,x) + \tau\, \partial_x\big(xn(t,x)\big) + B(x)n(t,x) = 4B(2x)n(t,2x), \\ n(0,x) = n^{(0)}(x), \qquad x \geq 0, \end{cases} \tag{1}$$

and it quantifies the concentration $n(t,x)$ of individuals (cells) having size x (the state variable) at time t. A common stochastic mechanism for every single cell is attached to equation (1):

1. The size $x = x(t)$ of a cell at time t evolves exponentially according to the deterministic evolution $\mathrm{d}x(t) = \tau x(t)\,\mathrm{d}t$, where $\tau > 0$ is the growth rate of each cell, that quantifies their ability to ingest a common nutrient.
2. Each cell splits into two offsprings according to a division rate $B(x)$ that depends on its current size x.
3. At division, a cell of size x gives birth to two offsprings of size $x/2$ each, what is called *binary fission*.

Model (1) is thus entirely determined by the parameters $(\tau, B(x), x \in [0,\infty))$. Typically, the growth rate τ is assumed to be known or guessed [13], and thus inference about (1) mainly concerns the estimation of the division rate $B(x)$ that has to be taken from a nonparametric perspective.

By use of the general relative entropy principle, Michel, Mischler and Perthame showed that the approximation $n(t,x)\mathrm{e}^{-\lambda_0 t} \approx N(x)$ is valid [29], with $\lambda_0 > 0$, and where (λ_0, N) is the dominant eigenpair related to the corresponding eigenvalue problem, see [2,8,16,25,28,32]. The "stationary" density $N(x)$ of typical cells after some time has elapsed enables to recover $(B(x), x \in \mathcal{D})$ for a compact $\mathcal{D} \subset (0,\infty)$ by means of the regularisation of an inverse problem of ill-posedness degree 1. From a deterministic perspective, this is carried out in [14,15,34]. From a statistical inference perspective, if an n-sample of the distribution $N(x)$ is observed and if $B(x)$ has smoothness $s > 0$ in a Sobolev sense, it is proved in [12] that $B(x)$ can be recovered in squared-error loss over compact sets with a rate of convergence $n^{-s/(2s+3)}$. Both deterministic and stochastic methodology of [14] and [12] are motivated by experimental designs and data such as in [13,24]. However, they do not take into account the following two important aspects:

- Bacterial growth exhibits variations in the individual growth rate τ as demonstrated for instance in [36]. One would like to incorporate variability in the growth rate within the system at the level of a single cell. This requires to modify model (1).
- Recent evolution of experimental technology enables to track the whole genealogy of cell populations (along prescribed lines of descendants for instance), affording the observation of other state variables such as size at division, lifetime of a single individual and so on [38]. Making the best possible use of such measures is of great potential impact, and needs a complementary approach.

The availability of observation schemes at the level of cell individuals suggests an enhancement of the statistical inference of $(B(x), x \in \mathcal{D})$, enabling to improve on the rates of convergence obtained by indirect measurements such as in [12,14]. This is the purpose of the present paper. We focus on bacterial growth, for which we apply our method on experimental observations. This serves as a proof of concept for the relevance of our modelling and statistical methodology, which could adapt to other application fields and growth-fragmentation types.

1.2. Results of the paper

Statistical setting

Let

$$\mathcal{U} = \bigcup_{k=0}^{\infty} \{0,1\}^k$$

denote the binary genealogical tree (with $\{0,1\}^0 := \{\varnothing\}$). We identify each node $u \in \mathcal{U}$ with a cell that has a size at birth ξ_u and a lifetime ζ_u. In the paper, we consider the problem of estimating $(B(x), x \in [0,\infty))$ over compact sets of $(0,\infty)$. Our inference procedure is based on the observation of

$$\big((\xi_u, \zeta_u), u \in \mathcal{U}_n\big), \tag{2}$$

where $\mathcal{U}_n \subset \mathcal{U}$ denotes a connected subset of size n containing the root $u = \varnothing$. Asymptotics are taken as $n \to \infty$. Two important observation schemes are considered: the sparse tree case, when we follow the system along a given branch with n individuals, and the full tree case, where we follow the evolution of the whole binary tree up to the N_nth generation, with $N_n \approx \log_2 n$. In this setting, we are able to generalise model (1) and allow the growth rate τ to vary with each cell $u \in \mathcal{U}$. We assume that a given cell u has a random growth rate $\tau_u = v \in \mathcal{E} \subset (0,\infty)$ (later constrained to live on a compact set). Moreover, this value v is inherited from the growth rate v' of its parent according to a distribution $\rho(v', \mathrm{d}v)$. Since a cell splits into two offsprings of the same size, letting u^- denote the parent of u, we have the fundamental relationship

$$2\xi_u = \xi_{u^-} \exp(\tau_{u^-} \zeta_{u^-}) \tag{3}$$

that enables to recover the growth rate τ_u of each individual in $\mathcal{U}_n$ since $\mathcal{U}_n$ is connected by assumption, possibly leaving out the last generation of observed individuals, but this has asymptotically no effect on a large sample size approach.

Variability in growth rate

In the case where the growth rate can vary for each cell, the density $n(t,x)$ of cells of size x at time t does not follow equation (1) anymore and an extended framework needs to be considered. To that end, we structure the system with an additional variable $\tau_u = v$, which represents the growth rate and depends on each individual cell u. We construct in Section 2 a branching Markov chain $((\xi_u, \tau_u), u \in \mathcal{U})$ that incorporates variability for the growth rate in the mechanism

described in Section 1.1. Equivalently to the genealogical tree, the system may be described in continuous time by a piecewise deterministic Markov process

$$\big(X(t), V(t)\big) = \big(\big(X_1(t), V_1(t)\big), \big(X_2(t), V_2(t)\big), \ldots\big),$$

which models the process of sizes and growth rates of the living particles in the system at time t, with value in $\bigcup_{k=0}^{\infty} \mathcal{S}^k$, where $\mathcal{S} = [0, \infty) \times \mathcal{E}$ is the state space of size times growth rate. Stochastic systems of this kind that correspond to branching Markov chains are fairly well known, both from a theoretical angle and in applications; a selected list of contributions is [5,10, 27] and the references therein.

By fragmentation techniques inspired by Bertoin [7], see also Haas [21], we relate the process (X, V) to a growth-fragmentation equation as follows. Define

$$\big\langle n(t, \cdot), \varphi \big\rangle = \mathbb{E}\left[\sum_{i=1}^{\infty} \varphi\big(X_i(t), V_i(t)\big)\right]$$

as the expectation of the empirical measure of the process (X, V) over smooth test functions defined on $\mathcal{S}$. We prove in Theorem 1 that, under appropriate regularity conditions, the measure $n(t, \cdot)$ that we identify with the temporal evolution of the density $n(t, x, v)$ of cells having size x and growth rate v at time t is governed (in a weak sense[1]) by

$$\begin{cases} \partial_t n(t, x, v) + v\, \partial_x\big(x n(t, x, v)\big) + B(x) n(t, x, v) \\ \quad = 4B(2x) \displaystyle\int_{\mathcal{E}} \rho\big(v', v\big) n\big(t, 2x, \mathrm{d}v'\big), \\ n(0, x, v) = n^{(0)}(x, v), \qquad x \geq 0. \end{cases} \tag{4}$$

If we assume a constant growth rate $\tau > 0$, we then take $\rho(v', \mathrm{d}v) = \delta_\tau(\mathrm{d}v)$ (where δ denotes the Dirac mass) and we retrieve the standard growth-fragmentation equation (1). The proof of Theorem 1 is obtained via a so-called many-to-one formula, established in Proposition 3 in Section 5.1. Indeed, thanks to the branching property of the system, it is possible to relate the behaviour of additive functionals like the mean empirical measure to the behaviour of a so-called tagged cell (like a tagged fragment in fragmentation process), that consists in following the behaviour of a single line of descendants along a branch where each node is picked at random, according to a uniform distribution. This approach, inspired by fragmentation techniques, is quite specific to our model and enables to obtain a relatively direct proof of Theorem 4.

[1]For every $t \geq 0$, we actually have a Radon measure $n(t, \mathrm{d}x, \mathrm{d}v)$ on $\mathcal{S} = [0, \infty) \times \mathcal{E}$: If $\varphi(x, v)$ is a function defined on $\mathcal{S}$, we define $\langle n(t, \cdot), \varphi \rangle = \int_{\mathcal{S}} \varphi(x, v) n(t, \mathrm{d}x, \mathrm{d}v)$ whenever the integral is meaningful. Thus (4) has the following sense: for every sufficiently smooth test function φ with compact support in $\mathcal{E}$, we have

$$\begin{aligned} &\int_{\mathcal{S}} \partial_t n(t, \mathrm{d}x, \mathrm{d}v) \varphi(x, v) - v x n(t, \mathrm{d}x, \mathrm{d}v)\, \partial_x \varphi(x, v) + B(x) n(t, \mathrm{d}x, \mathrm{d}v) \varphi(x, v) \\ &\quad = 4 \int_{\mathcal{S}} \Big(B(2x) \int_{\mathcal{E}} \rho\big(v', \mathrm{d}v\big) n\big(t, 2\mathrm{d}x, \mathrm{d}v'\big) \varphi(x, v) \Big). \end{aligned}$$

Nonparametric estimation of the growth rate

In Section 3, we take over the problem of estimating $(B(x), x \in \mathcal{D})$ for some compact $\mathcal{D} \subset (0, \infty)$. We assume we have data of the form (2), and that the mean evolution of the system is governed by (4). The growth rate kernel ρ is unknown and treated as a nuisance parameter. A fundamental object is the transition kernel

$$\mathcal{P}_B\big(\mathbf{x}, \mathrm{d}\mathbf{x}'\big) = \mathbb{P}\big((\xi_u, \tau_u) \in \mathrm{d}\mathbf{x}' | (\xi_{u^-}, \tau_{u^-}) = \mathbf{x}\big)$$

of the size and growth rate distribution (ξ_u, τ_u) at the birth of a descendant $u \in \mathcal{U}$, given the size of birth and growth rate of its parent (ξ_{u^-}, τ_{u^-}). We define in Section 3.3 a class of division rates and growth rate kernels such that if (B, ρ) belongs to this class, then the transition $\mathcal{P}_B$ is geometrically ergodic and has a unique invariant measure $\nu_B(\mathrm{d}\mathbf{x}) = \nu_B(x, \mathrm{d}v)\,\mathrm{d}x$. From the invariant measure equation

$$\nu_B \mathcal{P}_B = \nu_B$$

we obtain in Proposition 2 the explicit representation

$$B(x) = \frac{x}{2} \frac{\nu_B(x/2)}{\mathbb{E}_{\nu_B}[(1/\tau_{u^-})\mathbf{1}_{\{\xi_{u^-} \le x, \xi_u \ge x/2\}}]}, \tag{5}$$

where $\nu_B(x) = \int_{\mathcal{E}} \nu_B(x, \mathrm{d}v)$ denotes the first marginal of the invariant distribution ν_B. A strategy for constructing and estimator B consists in replacing the right-hand size of (5) by its empirical counterpart, the numerator being estimated via a kernel estimator of the first maginal of $\nu_B(\mathrm{d}\mathbf{x})$. Under local Hölder smoothness assumption on B of order $s > 0$, we prove in Theorem 2 that for a suitable choice of bandwidth in the estimation of the invariant density, our estimator achieves the rate $n^{-s/(2s+1)}$ in squared-error loss over appropriate compact sets $\mathcal{D} \subset (0, \infty)$, up to an inessential logarithmic term when the full tree observation scheme is considered. We see in particular that we improve on the rate obtained in [12]. Our result quantifies the improvement obtained when estimating $B(x)$ from data $((\xi_u, \zeta_u), u \in \mathcal{U}_n)$, as opposed to overall measurements of the system after some time has elapsed as in [12]. We provide a quantitative argument based on the analysis of a PDE that explains the reduction of ill-posedness achieved by our method over [12] in Section 4.2.

In order to obtain the upper bound of Theorem 2, a major technical difficulty is that we need to establish uniform rates of convergence of the empirical counterparts to their limits in the numerator and denominator of (5) when the data are spread along a binary tree. This can be done via covariance inequalities that exploit the fact that the transition $\mathcal{P}_B$ is geometrically ergodic (Proposition 4) using standard Markov techniques, see [6,27]. The associated chain is however not reversible, and this yields an extraneous difficulty: the decay of the correlations between $\varphi(\xi_u, \tau_u)$ and $\varphi(\xi_v, \tau_v)$ for two nodes $u, v \in \mathcal{U}_n$ are expressed in terms of the sup-norm of φ, whenever $|\varphi(\mathbf{x})| \le \mathbb{V}(\mathbf{x})$ is dominated by a certain Lyapunov function $\mathbb{V}$ for the transition $\mathcal{P}_B$. However, the typical functions φ we use are kernels that depend on n and that are not uniformly bounded in sup-norm as $n \to \infty$. This partly explains the relative length of the technical Sections 5.5 and 5.6.

1.3. Organisation of the paper

In Section 2, we construct the model $((\xi_u, \tau_u), u \in \mathcal{U})$ of sizes and growth rates of the cells as a Markov chain along the genealogical tree. The discrete model can be embedded into a continuous time piecewise deterministic Markov process (X, V) of sizes and growth rates of the cells present at any time within the system. In Theorem 1, we explicit the relation between the mean empirical measure of (X, V) and the growth-fragmentation type equation 4. In Section 3, we explicitly construct an estimator $\widehat{B}_n$ of B by means of the representation given by (5) in Section 3.2. Two observation schemes are considered and discussed in Section 3.1, whether we consider data along a single branch (the sparse tree case) or along the whole genealogy (the full tree case). The specific assumptions and the class of admissible division rates B and growth rate kernels ρ are discussed in Section 3.3, and an upper bound for $\widehat{B}_n$ in squared-error loss is given in our main Theorem 2. Section 4 shows and discusses the numerical implementation of our method on simulated data. In particular, ignoring the variability in the reconstruction dramatically deterioriates the accuracy of estimation of B. We also explain from a deterministic point perspective the rate improvement of our method compared with [12] by means of a PDE analysis argument in Section 4.2. The parameters are inspired from real data experiments on *Escherichia coli* cell cultures. Section 5 is devoted to the proofs.

2. A Markov model on a tree

2.1. The genealogical construction

Recall that $\mathcal{U} := \bigcup_{n=0}^{\infty}\{0, 1\}^n$ (with $\{0, 1\}^0 := \{\varnothing\}$) denotes the infinite binary genealogical tree. Each node $u \in \mathcal{U}$ is identified with a cell of the population and has a mark

$$(\xi_u, b_u, \zeta_u, \tau_u),$$

where ξ_u is the size at birth, τ_u the growth rate, b_u the birthtime and ζ_u the lifetime of u. The evolution $(\xi_t^u, t \in [b_u, b_u + \zeta_u))$ of the size of u during its lifetime is governed by

$$\xi_t^u = \xi_u \exp\big(\tau_u(t - b_u)\big) \qquad \text{for } t \in [b_u, b_u + \zeta_u). \tag{6}$$

Each cell splits into two offsprings of the same size according to a division rate $B(x)$ for $x \in (0, \infty)$. Equivalently

$$\mathbb{P}\big(\zeta_u \in [t, t + \mathrm{d}t] | \zeta_u \geq t, \xi_u = x, \tau_u = v\big) = B\big(x \exp(vt)\big)\,\mathrm{d}t. \tag{7}$$

At division, a cell splits into two offsprings of the same size. If u^- denotes the parent of u, we thus have

$$2\xi_u = \xi_{u^-} \exp(\tau_{u^-}\zeta_{u^-}). \tag{8}$$

Finally, the growth rate τ_u of u is inherited from its parent τ_{u^-} according to a Markov kernel

$$\rho\big(v, \mathrm{d}v'\big) = \mathbb{P}\big(\tau_u \in \mathrm{d}v' | \tau_{u^-} = v\big), \tag{9}$$

where $v > 0$ and $\rho(v, \mathrm{d}v')$ is a probability measure on $(0, \infty)$ for each $v > 0$. Equations (6), (7), (8) and (9) completely determine the dynamics of the model $((\xi_u, \tau_u), u \in \mathcal{U})$, as a Markov chain on a tree, given an additional initial condition $(\xi_\varnothing, \tau_\varnothing)$ on the root. The chain is embedded into a piecewise deterministic continuous Markov process thanks to (6) by setting

$$\big(\xi_t^u, \tau_t^u\big) = \big(\xi_u \exp\big(\tau_u(t - b_u)\big), \tau_u\big) \qquad \text{for } t \in [b_u, b_u + \zeta_u)$$

and $(0, 0)$ otherwise. Define

$$\big(X(t), V(t)\big) = \big(\big(X_1(t), V_1(t)\big), \big(X_2(t), V_2(t)\big), \ldots\big)$$

as the process of sizes and growth rates of the living particles in the system at time t. We have an identity between point measures

$$\sum_{i=1}^{\infty} \mathbf{1}_{\{X_i(t)>0\}} \delta_{(X_i(t), V_i(t))} = \sum_{u \in \mathcal{U}} \mathbf{1}_{\{b_u \le t < b_u + \zeta_u\}} \delta_{(\xi_t^u, \tau_t^u)}, \tag{10}$$

where δ denotes the Dirac mass. In the sequel, the following basic assumption is in force.

Assumption 1 (Basic assumption on B and ρ). *The division rate $x \rightsquigarrow B(x)$ is continuous. We have $B(0) = 0$ and $\int^{\infty} x^{-1} B(x)\, \mathrm{d}x = \infty$. The Markov kernel $\rho(v, \mathrm{d}v')$ is defined on a compact set $\mathcal{E} \subset (0, \infty)$.*

Proposition 1. *Work under Assumption 1. The law of*

$$\big(\big(X(t), V(t)\big), t \ge 0\big) \quad \text{or} \quad \big((\xi_u, \tau_u), u \in \mathcal{U}\big) \quad \text{or} \quad \big(\big(\xi_t^u, \zeta_t^u\big), t \ge 0, u \in \mathcal{U}\big)$$

is well-defined on an appropriate probability space with almost-surely no accumulation of jumps.

If μ is a probability measure on the state space $\mathcal{S} = [0, \infty) \times \mathcal{E}$, we shall denote indifferently by $\mathbb{P}_\mu$ the law of any of the three processes above where the root $(\xi_\varnothing, \tau_\varnothing)$ has distribution μ. The construction is classical (see for instance [7] and the references therein) and is outlined in Appendix A.1.

2.2. Behaviour of the mean empirical measure

Denote by $\mathcal{C}_0^1(\mathcal{S})$ the set of real-valued test functions with compact support in the interior of $\mathcal{S}$.

Theorem 1 (Behaviour of the mean empirical measure). *Work under Assumption 1. Let μ be a probability distribution on $\mathcal{S}$. Define the distribution $n(t, \mathrm{d}x, \mathrm{d}v)$ by*

$$\big\langle n(t, \cdot), \varphi \big\rangle = \mathbb{E}_\mu\left[\sum_{i=1}^{\infty} \varphi\big(X_i(t), V_i(t)\big)\right] \qquad \textit{for every } \varphi \in \mathcal{C}_0^1(\mathcal{S}).$$

Then $n(t,\cdot)$ solves (in a weak sense)

$$\begin{cases} \partial_t n(t,x,v) + v\,\partial_x\big(xn(t,x,v)\big) + B(x)n(t,x,v) \\ \quad = 4B(2x)\displaystyle\int_{\mathcal{E}} \rho\big(v',v\big)n\big(t,2x,\mathrm{d}v'\big), \\ n(0,x,v) = n^{(0)}(x,v), \qquad x \geq 0, \end{cases}$$

with initial condition $n^{(0)}(\mathrm{d}x,\mathrm{d}v) = \mu(\mathrm{d}x,\mathrm{d}v)$.

Theorem 1 somehow legitimates our methodology: by enabling each cell to have its own growth rate and by building-up new statistical estimators in this context, we still have a translation in terms of the approach in [14]. In particular, we will be able to compare our estimation results with [12]. Our proof is based on fragmentation techniques, inspired by Bertoin [7] and Haas [21]. Alternative approaches to the same kind of questions include the probabilistic studies of Chauvin *et al.* [9], Bansaye *et al.* [5] or Harris and Roberts [22] and references therein.

3. Statistical estimation of the division rate

3.1. Two observation schemes

Let $\mathcal{U}_n \subset \mathcal{U}$ denote a subset of size n of connected nodes: if u belongs to $\mathcal{U}_n$, so does its parent u^-. We look for a nonparametric estimator of the division rate

$$y \rightsquigarrow \widehat{B}_n(y) = \widehat{B}_n\big(y, (\xi_u,\tau_u), u \in \mathcal{U}_n\big) \qquad \text{for } y \in (0,\infty).$$

Statistical inference is based on the observation scheme

$$\big((\xi_u,\tau_u), u \in \mathcal{U}_n\big)$$

and asymptotic study is undertaken as the population size of the sample $n \to \infty$. We are interested in two specific observation schemes.

The full tree case. We observe every pair (ξ_u,τ_u) over the first N_n generations of the tree:

$$\mathcal{U}_n = \big\{u \in \mathcal{U}, |u| \leq N_n\big\}$$

with the notation $|u| = n$ if $u = (u_0,u_1,\ldots,u_n) \in \mathcal{U}$, and N_n is chosen such that 2^{N_n} has order n.

The sparse tree case. We follow the first n offsprings of a single cell, along a fixed line of descendants. This means that for some $u \in \mathcal{U}$ with $|u| = n$, we observe every size ξ_u and growth rate τ_u of each node (u_0), (u_0,u_1), (u_0,u_1,u_2) and so on up to a final node $u = (u_0,u_1,\ldots,u_n)$.

Remark 1. For every $n \geq 1$, we tacitly assume that there exists a (random) time $T_n < \infty$ almost surely, such that for $t \geq T_n$, the observation scheme $\mathcal{U}_n$ is well-defined. This is a consequence of the behaviour of B near infinity that we impose later on in (17) below.

3.2. Estimation of the division rate

Identification of the division rate

We denote by $\mathbf{x} = (x, v)$ an element of the state space $\mathcal{S} = [0, \infty) \times \mathcal{E}$. Introduce the transition kernel

$$\mathcal{P}_B\big(\mathbf{x}, \mathrm{d}\mathbf{x}'\big) = \mathbb{P}\big((\xi_u, \tau_u) \in \mathrm{d}\mathbf{x}' | (\xi_{u^-}, \tau_{u^-}) = \mathbf{x}\big)$$

of the size and growth rate distribution (ξ_u, τ_u) at the birth of a descendant $u \in \mathcal{U}$, given the size at birth and growth rate of its parent (ξ_{u^-}, τ_{u^-}). From (7), we infer that $\mathbb{P}(\zeta_{u^-} \in \mathrm{d}t | \xi_{u^-} = x, \tau_{u^-} = v)$ is equal to

$$B\big(x\exp(vt)\big)\exp\bigg(-\int_0^t B\big(x\exp(vs)\big)\,\mathrm{d}s\bigg)\,\mathrm{d}t.$$

Using formula (8), by a simple change of variables

$$\mathbb{P}\big(\xi_u \in \mathrm{d}x' | \xi_{u^-} = x, \tau_{u^-} = v\big) = \frac{B(2x')}{vx'}\mathbf{1}_{\{x' \geq x/2\}}\exp\bigg(-\int_{x/2}^{x'}\frac{B(2s)}{vs}\,\mathrm{d}s\bigg)\,\mathrm{d}x'.$$

Incorporating (9), we obtain an explicit formula for

$$\mathcal{P}_B\big(\mathbf{x}, \mathrm{d}\mathbf{x}'\big) = \mathcal{P}_B\big((x, v), x', \mathrm{d}v'\big)\,\mathrm{d}x',$$

with

$$\mathcal{P}_B\big((x, v), x', \mathrm{d}v'\big) = \frac{B(2x')}{vx'}\mathbf{1}_{\{x' \geq x/2\}}\exp\bigg(-\int_{x/2}^{x'}\frac{B(2s)}{vs}\,\mathrm{d}s\bigg)\rho\big(v, \mathrm{d}v'\big). \tag{11}$$

Assume further that $\mathcal{P}_B$ admits an invariant probability measure $\nu_B(\mathrm{d}\mathbf{x})$, that is, a solution to

$$\nu_B\mathcal{P}_B = \nu_B, \tag{12}$$

where

$$\mu\mathcal{P}_B(\mathrm{d}\mathbf{y}) = \int_{\mathcal{S}}\mu(\mathrm{d}\mathbf{x})\mathcal{P}_B(\mathbf{x}, \mathrm{d}\mathbf{y})$$

denotes the left action of positive measures $\mu(\mathrm{d}\mathbf{x})$ on $\mathcal{S}$ for the transition $\mathcal{P}_B$.

Proposition 2. *Work under Assumption 1. If $\mathcal{P}_B$ admits an invariant probability measure ν_B of the form $\nu_B(\mathrm{d}\mathbf{x}) = \nu_B(x, \mathrm{d}v)\,\mathrm{d}x$, then we have*

$$\nu_B(y) = \frac{B(2y)}{y}\mathbb{E}_{\nu_B}\bigg[\frac{1}{\tau_{u^-}}\mathbf{1}_{\{\xi_{u^-} \leq 2y, \xi_u \geq y\}}\bigg] \tag{13}$$

where $\mathbb{E}_{\nu_B}[\cdot]$ denotes expectation when the initial condition $(\xi_\varnothing, \tau_\varnothing)$ has distribution ν_B and where we have set $\nu_B(y) = \int_{\mathcal{E}}\nu_B(y, \mathrm{d}v')$ in (13) for the marginal density of the invariant probability measure ν_B with respect to y.

We exhibit below a class of division rates B and growth rate kernels ρ that guarantees the existence of such an invariant probability measure.

Construction of a nonparametric estimator

Inverting (13) and applying an appropriate change of variables, we obtain

$$B(y) = \frac{y}{2} \frac{\nu_B(y/2)}{\mathbb{E}_{\nu_B}[(1/\tau_{u^-})\mathbf{1}_{\{\xi_{u^-} \le y, \xi_u \ge y/2\}}]}, \tag{14}$$

provided the denominator is positive. This formula has no easy interpretation: it is obtained by some clever manipulation of the equation $\nu_B = \mathcal{P}_B \nu_B$. A tentative interpretation in the simplest case with no variability (so that $\tau_u = \tau$ for some fixed $\tau > 0$ and for every $u \in \mathcal{U}$ is proposed in Section 4.2). Representation (14) also suggests an estimation procedure, replacing the marginal density $\nu_B(y/2)$ and the expectation in the denominator by their empirical counterparts. To that end, pick a kernel function

$$K : [0, \infty) \to \mathbb{R}, \qquad \int_{[0,\infty)} K(y)\, \mathrm{d}y = 1,$$

and set $K_h(y) = h^{-1} K(h^{-1} y)$ for $y \in [0, \infty)$ and $h > 0$. Our estimator is defined as

$$\widehat{B}_n(y) = \frac{y}{2} \frac{n^{-1} \sum_{u \in \mathcal{U}_n} K_h(\xi_u - y/2)}{n^{-1} \sum_{u \in \mathcal{U}_n} (1/\tau_{u^-}) \mathbf{1}_{\{\xi_{u^-} \le y, \xi_u \ge y/2\}} \vee \varpi}, \tag{15}$$

where $\varpi > 0$ is a threshold that ensures that the estimator is well defined in all cases and $x \vee y = \max\{x, y\}$. Thus $(\widehat{B}_n(y),\, y \in \mathcal{D})$ is specified by the choice of the kernel K, the bandwidth $h > 0$ and the threshold $\varpi > 0$.

Assumption 2. *The function K is bounded with compact support, and for some integer $n_0 \ge 1$, we have $\int_{[0,\infty)} x^k K(x)\, \mathrm{d}x = \mathbf{1}_{\{k=0\}}$ for $0 \le k \le n_0$.*

3.3. Error estimates

We assess the quality of $\widehat{B}_n$ in squared-loss error over compact intervals $\mathcal{D}$. We need to specify local smoothness properties of B over $\mathcal{D}$, together with general properties that ensure that the empirical measurements in (15) converge with an appropriate speed of convergence. This amounts to impose an appropriate behaviour of B near the origin and infinity.

Model constraints

For $\lambda > 0$ such that $2^\lambda > \sup \mathcal{E} / \inf \mathcal{E} > 0$ and a vector of positive constants $\mathfrak{c} = (r, m, \ell, L)$, introduce the class $\mathcal{F}^\lambda(\mathfrak{c})$ of continuous functions $B : [0, \infty) \to [0, \infty)$ such that

$$\int_0^{r/2} x^{-1} B(2x)\, \mathrm{d}x \le L, \qquad \int_{r/2}^{r} x^{-1} B(2x)\, \mathrm{d}x \ge \ell \tag{16}$$

and

$$B(x) \geq mx^{\lambda} \qquad \text{for } x \geq r. \tag{17}$$

Remark 2. Similar conditions on the behaviour of B can also be found in [16], in a deterministic setting.

Remark 3. Assumption 1 is satisfied as soon as $B \in \mathcal{F}^{\lambda}(\mathfrak{c})$ (and ρ is defined on a compact $\mathcal{E} \subset (0, \infty)$ of course).

Let $\rho_{\min}, \rho_{\max}$ be two positive finite measures on $\mathcal{E}$ such that $\rho_{\max} - \rho_{\min}$ is a positive measure and $\rho_{\min}(\mathcal{E}) > 0$. We define $\mathcal{M}(\rho_{\min}, \rho_{\max})$ as the class of Markov transitions $\rho(v, \mathrm{d}v')$ on $\mathcal{E}$ such that

$$\rho_{\min}(A) \leq \rho(v, A) \leq \rho_{\max}(A), \qquad A \subset \mathcal{E}, v \in \mathcal{E}. \tag{18}$$

Remark 4. Control (18) ensures the geometric ergodicity of the process of variability in the growth rate.

Let us be given in the sequel a vector of positive constants $\mathfrak{c} = (r, m, \ell, L)$ and $0 < e_{\min} \leq e_{\max}$ such that $\mathcal{E} \subset [e_{\min}, e_{\max}]$. We introduce the Lyapunov function

$$\mathbb{V}(x, v) = \mathbb{V}(x) = \exp\left(\frac{m}{e_{\min}\lambda} x^{\lambda}\right) \qquad \text{for } (x, v) \in \mathcal{S}. \tag{19}$$

The function $\mathbb{V}$ controls the rate of the geometric ergodicity of the chain with transition $\mathcal{P}_B$ and will appear in the proof of Proposition 4 below. Define

$$\delta = \delta(\mathfrak{c}) := \frac{1}{1 - 2^{-\lambda}} \exp\left(-\left(1 - 2^{-\lambda}\right)\frac{m}{e_{\max}\lambda} r^{\lambda}\right) \rho_{\max}(\mathcal{E}).$$

Assumption 3 (The sparse tree case). *Let $\lambda > 0$. We have $\delta(\mathfrak{c}) < 1$.*

In the case of the full tree observation scheme, we will need more stringent (and technical) conditions on $\mathfrak{c}$. Let $\gamma_{B,\mathbb{V}}$ denote the spectral radius of the operator $\mathcal{P}_B - 1 \otimes \nu_B$ acting on the Banach space of functions $g : \mathcal{S} \to \mathbb{R}$ such that

$$\sup\left\{\left|g(\mathbf{x})\right| / \mathbb{V}(\mathbf{x}), \mathbf{x} \in \mathcal{S}\right\} < \infty,$$

where $\mathbb{V}$ is defined in (19) above.

Assumption 4 (The full tree case). *We have $\delta(\mathfrak{c}) < \frac{1}{2}$ and moreover*

$$\sup_{B \in \mathcal{F}^{\lambda}(\mathfrak{c})} \gamma_{B,\mathbb{V}} < \frac{1}{2}. \tag{20}$$

Remark 5. It is possible to obtain bounds on $\mathfrak{c}$ so that (20) holds, by using explicit (yet intricate) bounds on $\gamma_{B,\mathbb{V}}$ following Fort *et al.* [18] or Douc *et al.* [11], see also Baxendale [6].

Rate of convergence

We are ready to state our main result. For $s > 0$, with $s = \lfloor s \rfloor + \{s\}$, $0 < \{s\} \leq 1$ and $\lfloor s \rfloor$ an integer, introduce the Hölder space $\mathcal{H}^s(\mathcal{D})$ of functions $f : \mathcal{D} \to \mathbb{R}$ possessing a derivative of order $\lfloor s \rfloor$ that satisfies

$$\left| f^{\lfloor s \rfloor}(y) - f^{\lfloor s \rfloor}(x) \right| \leq c(f)|x - y|^{\{s\}}. \tag{21}$$

The minimal constant $c(f)$ such that (21) holds defines a semi-norm $|f|_{\mathcal{H}^s(\mathcal{D})}$. We equip the space $\mathcal{H}^s(\mathcal{D})$ with the norm

$$\|f\|_{\mathcal{H}^s(\mathcal{D})} = \|f\|_{L^\infty(\mathcal{D})} + |f|_{\mathcal{H}^s(\mathcal{D})}$$

and the Hölder balls

$$\mathcal{H}^s(\mathcal{D}, M) = \left\{ B, \|B\|_{\mathcal{H}^s(\mathcal{D})} \leq M \right\}, \qquad M > 0.$$

Theorem 2. *Work under Assumption 3 in the sparse tree case and Assumption 4 in the full tree case. Specify $\widehat{B}_n$ with a kernel K satisfying Assumption 2 for some $n_0 > 0$ and*

$$h_n = c_0 n^{-1/(2s+1)}, \qquad \varpi_n \to 0.$$

For every $M > 0$, there exist $c_0 = c_0(\mathfrak{c}, M)$ and $d(\mathfrak{c}) \geq 0$ such that for every $0 < s < n_0$ and every compact interval $\mathcal{D} \subset (d(\mathfrak{c}), \infty)$ such that $\inf \mathcal{D} \geq r/2$, we have

$$\sup_{\rho, B} \mathbb{E}_\mu \left[\|\widehat{B}_n - B\|^2_{L^2(\mathcal{D})} \right]^{1/2} \lesssim \varpi_n^{-1} (\log n)^{1/2} n^{-s/(2s+1)},$$

where the supremum is taken over

$$\rho \in \mathcal{M}(\rho_{\min}, \rho_{\max}) \quad \text{and} \quad B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(\mathcal{D}, M),$$

and $\mathbb{E}_\mu[\cdot]$ denotes expectation with respect to any initial distribution $\mu(\mathrm{d}\mathbf{x})$ for $(\xi_\varnothing, \tau_\varnothing)$ on $\mathcal{S}$ such that $\int_{\mathcal{S}} \mathbb{V}(\mathbf{x})^2 \mu(\mathrm{d}\mathbf{x}) < \infty$.

Several remarks are in order: (1) Since ϖ_n is arbitrary, we obtain the classical rate $n^{-s/(2s+1)}$ (up to a log term) which is optimal in a minimax sense for density estimation. It is presumably optimal in our context, using for instance classical techniques for nonparametric estimation lower bounds on functions of transition densities of Markov chains, see, for instance, [19]. (2) The extra logarithmic term is due to technical reasons: we need it in order to control the decay of correlations of the observations over the full tree structure. (3) The knowledge of the smoothness s that is needed for the construction of $\widehat{B}_n$ is not realistic in practice. An adaptive estimator could be obtained by using a data-driven bandwidth in the estimation of the invariant density $\nu_B(y/2)$ in (15). The Goldenschluger–Lepski bandwidth selection method [20], see also [12] would presumably yield adaptation, but checking the assumptions still requires a proof in our setting. We implement data-driven bandwidth in the numerical Section 4 below.

4. Numerical implementation

4.1. Protocol and results

Generating simulated data

Given a division rate $B(x)$, a growth rate kernel ρ, an initial distribution $\mu(\mathrm{d}\mathbf{x})$ for the node $(\xi_\varnothing, \tau_\varnothing)$ (as in Theorem 2) and a dataset size $n = 2^{N_n}$, we simulate the full tree and the sparse tree schemes recursively:

1. Given (ξ_{u^-}, τ_{u^-}), we select at random its lifetime ζ_{u^-} (by a rejection sampling algorithm) with probability density
$$t \rightsquigarrow B\big(\xi_{u^-}\exp(\tau_{u^-}t)\big)\exp\bigg(-\int_0^t B\big(\xi_{u^-}\exp(\tau_{u^-}s)\big)\,\mathrm{d}s\bigg),$$
following the computations of Section 3.2.
2. We derive the size at birth ξ_u for the two offsprings (with $u = (u^-, 0)$ and $(u^-, 1)$ with obvious notation) by formula (8).
3. We simulate at random the growth rates τ_u (by the rejection sampling algorithm) according to the distribution $\rho(\tau_{u^-}, \mathrm{d}v)$.
4. For the sparse tree case, we select only one offspring (either $(u^-, 0)$ of $(u^-, 1)$), whereas we keep both for the full tree case.

In order to stay in line with previous simulations of [12] we pick $B(x) = x^2$. We fix $\mu(\mathrm{d}\mathbf{x})$ as the uniform distribution over $[1/3, 3] \times \mathcal{E}$, with $\mathcal{E} = [0.2, 3]$. As for the growth rate kernel, we implement
$$\rho\big(v, \mathrm{d}v'\big) = g\big(v' - v\big)\,\mathrm{d}v',$$
where g is a uniform distribution over $[1-\alpha, 1+\alpha]$ for some $\alpha > 0$, and dilated by a scaling factor so that $(\int (v'-v)^2\rho(v, \mathrm{d}v'))^{1/2} = 1/2$. We also condition the values of τ_u to stay in $\mathcal{E}$ (by rejection sampling).

Implementing $\widehat{B}_n$

We implement $\widehat{B}_n$ using formula (15). We pick a standard Gaussian kernel $K(x) = (2\pi)^{-1/2} \times \exp(-x^2/2)$, for which $n_0 = 1$ in Assumption 2; henceforth we expect a rate of convergence of order $n^{-1/3}$ at best. We evaluate $\widehat{B}_n$ on a regular grid $x_1 = \Delta x, \ldots x_m = m\Delta x$ with $\Delta x = n^{-1/2}$ and $x_m = 5$. Thus x_m is large enough so that $v_B(x/2)$ becomes negligible for $x \geq x_m$ and Δx is smaller than $n^{-1/3}$ to avoid numerical discrepancies. For tractability purposes, we wish to avoid the use of any relationship between the nodes $u \in \mathcal{U}_n$. Indeed, whereas it is quite easy to label u^- and u in the sparse tree case, it is a bit more difficult to track the parent of each individual in the full tree case if we do not want to double the memory. As a consequence, we simply reformulate (15) into
$$\widehat{B}_n(y) = \frac{y}{2}\frac{n^{-1}\sum_{u\in\mathcal{U}_n} K_h(\xi_u - y/2)}{n^{-1}\sum_{u\in\mathcal{U}_n}(1/\tau_u)\mathbf{1}_{\{\xi_u \leq y \leq \xi_u \mathrm{e}^{\tau_u\zeta_u}\}} \vee \varpi}. \tag{22}$$

We take $h_n = n^{-1/3}$ for the bandwidth according to Theorem 2 to serve as a proof of concept. Data-driven choices could of course be made, such as the Goldenschluger and Lepski's method [12,20], and improve the already fairly good results shown in Figure 2. Finally, we also test whether taking into account variability in the growth rate improves significantly or not the estimate of B, replacing τ_u by its mean value $n^{-1}\sum_{u\in\mathcal{U}_n}\tau_u$ everywhere in formula (22), thus ignoring growth variability in that case.

Numerical results

We display our numerical results as specified above in Figures 1, 2 and 3. Figure 1 displays the reconstruction of B on the full tree scheme for a simulated sample of size $n = 2^{17}$. At a visual level, we see that the estimation deteriorates dramatically when the variability is ignored in the region where ν_B is small, while our estimator (22) still shows good performances.

In Figure 2, we plot on a log–log scale the empirical mean error of our estimation procedure for both full tree and sparse tree schemes. The numerical results agree with the theory. The empirical error is computed as follows: we compute

$$e_i = \frac{\|\widehat{B} - B\|_{\Delta x,m}}{\|B\|_{\Delta x,m,\varpi}}, \qquad i = 1,\dots,M, \tag{23}$$

where $\|\cdot\|_{\Delta x,m,\varpi}$ denotes the discrete norm over the numerical sampling described above, conditioned on the fact that the denominator in (22) is larger than $\varpi/\log(n)$. We end up with a mean-empirical error $\overline{e} = M^{-1}\sum_{i=1}^{M} e_i$. The number of Monte-Carlo samples is chosen as $M = 100$. In Figure 3, we explore further the degradation of the estimation process on the region where ν_B

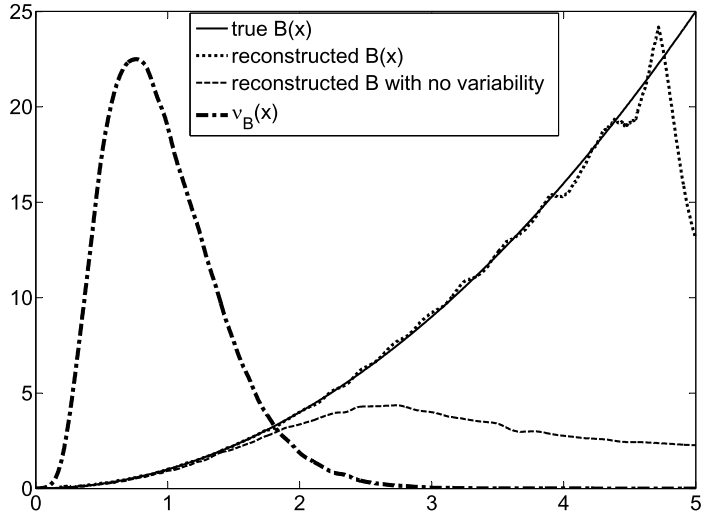

Figure 1. Reconstruction for $n = 2^{17}$. When the variability in the growth rate is ignored, the estimate reveals unsatisfactory. The parameter values are the reference ones.

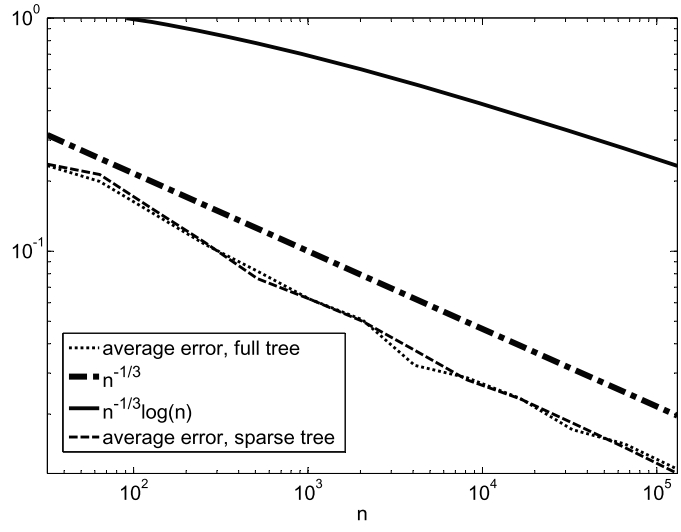

Figure 2. Error vs. n for the full tree and the sparse tree case on a log–log scale. The error actually proves better than the upper rate of convergence announced in Theorem 2, and ϖ may be taken smaller than $\log(n)$. Estimates are comparable for both schemes. The parameter values are the reference ones.

is small, plotting 95% confidence intervals of the empirical distribution of the estimates, based on $M = 100$ Monte-Carlo samples. Finally, Table 1 displays the relative error for the reconstruction of B according to (23). The standard deviation is computed as $(M^{-1}\sum_{i=1}^{M}(e_i - \overline{e})^2)^{1/2}$. We also carried out control experiments for other choices of variability kernel $\rho(v, \mathrm{d}v')$ for the growth rate. These include $\rho(v, \mathrm{d}v') = g(v')\,\mathrm{d}v'$, so that the variability of an individual is not inherited

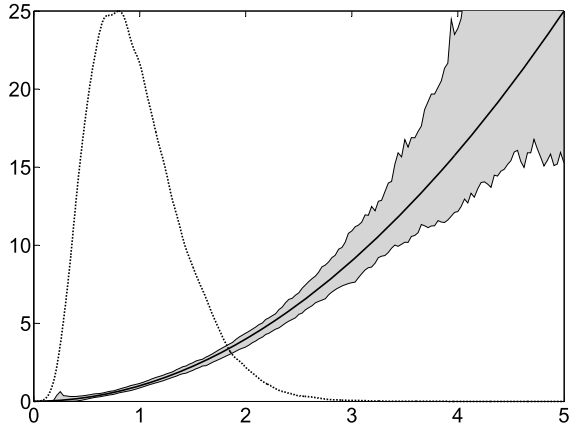

Figure 3. Reconstruction for $n = 2^{10}$, error band for 95%, full tree case, over $M = 100$ simulations, with $\varpi = 1/n$ in order to emphasise that the larger x, the smaller v_B and the larger the error estimate.

Table 1. Relative error $\overline{e}$ for B and its standard deviation, with respect to n (on a log scale). The error is computed using (23) with $\varpi = 1/\log(n)$

$\log_2(n)$	5	6	7	8	9	10
$\overline{e}$	0.2927	0.1904	0.1460	0.1024	0.0835	0.0614
std. dev.	0.1775	0.0893	0.0627	0.0417	0.0364	0.0241

from its parent, a Gaussian density for g with the same prescription for the mean and the variance as in the uniform case, conditioned to live on $[e_{\min}, e_{\max}]$. We also tested the absence of variability, with $\rho(v, \mathrm{d}v') = \delta_\tau(\mathrm{d}v')$, with $\tau = 1$. None of these control experiments show any significant difference from the case displayed in Figures 1, 2 and 3.

Analysis on E. coli *data*

Finally, we analyse a dataset obtained through microscopic time-lapse imaging of single bacterial cells growing in rich medium, by Wang *et al.* [38] and by Stewart *et al.* [35]. Thanks to a microfluidic set-up, the experimental conditions are well controlled and stable, so that the cells are in a steady state of growth (so-called balanced growth). The observation scheme corresponds to the sparse tree case for the data from Wang *et al.* [38]: at each generation, only one offspring is followed. On the contrary, data corresponds to the full tree case for the data by Stewart *et al.*, where the cells grow in a culture. The growth and division of the cells is followed by microscopy, and image analysis allows to determine the time evolution of the size of each cell, from birth to division. We picked up the quantities of interest for our implementation: for each cell, its size at birth, growth rate and lifetime. We consider that cells divide equally into two daughter cells, neglecting the small differences of size at birth between daughter cells. Each cell grows exponentially fast, but growth rates exhibit variability.

Our data is formed by the concatenation of several lineages, each of them composed with a line of offsprings coming from a first single cell picked at random in a culture. Some of the first and last generations were not considered in order to avoid any experimental disturbance linked either to non stationary conditions or to aging of the cells.

We proceed as in the above protocol. Figure 4 shows the reconstructed B and ν_B for a sample of $n = 2335$ cells for the sparse tree data, $n = 748$ for the full tree data. Though much more precise and reliable, thanks both to the experimental device and the reconstruction method, our results are qualitatively in accordance with previous indirect reconstructions carried out in [13] on old datasets published in [24] back in 1969.

The reconstruction of the division rate is prominent here since it appears to be the last component needed for a full calibration of the model. Thus, our method provides biologists with a complete understanding of the size dependence of the biological system. Phenotypic variability between genetically identical cells has recently received growing attention with the recognition that it can be genetically controlled and subject to selection pressures [23]. Our mathematical framework allows the incorporation of this variability at the level of individual growth rates. It should allow the study of the impact of variability on the population fitness and should be of particular importance to describe the growth of populations of cells exhibiting high variability

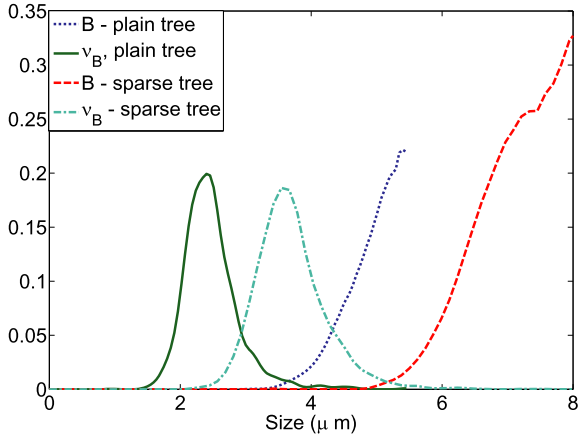

Figure 4. Estimation of B (dotted line and dashed line resp.) and ν_B (solid line and dash-dotted line resp.) on experimental data of *E. coli* dividing cells, for resp. a sparse tree and a full tree experiment. $n = 2335$ for the sparse tree, $n = 748$ for the full tree. The experimental conditions being different (temperature and nutrient), these results are not supposed to be identical, yet a generic pattern appears for B, that could serve as a basis for future biological studies.

of growth rates. Several examples of high variability have been described, both in genetically engineered or natural bacterial populations [36,37].

4.2. Link with the deterministic viewpoint

Considering the reconstruction formula (15), let us give here some insight from a deterministic analysis perspective. For the sake of clarity, let us focus on the simpler case when there is no variability, so that for all $u \in \mathcal{U}_n$ we have $\tau_u = \tau > 0$ a fixed constant. Formula (15) comes from (14), which in the case $\tau_u = \tau$ simplifies further into

$$B(y) = \frac{\tau y}{2} \frac{\nu_B(y/2)}{\mathbb{E}_{\nu_B}[\mathbf{1}_{\{\xi_{u^-} \le y, \xi_u \ge y/2\}}]} = \frac{\tau y}{2} \frac{\nu_B(y/2)}{\int_{y/2}^{y} \nu_B(z)\,\mathrm{d}z}. \tag{24}$$

We also notice that, in this particular case, we do not need to measure the lifetime of each cell in order to implement (24). Define $N(y) = \frac{1}{2}\frac{\nu_B(y/2)}{B(y)}$, or equivalently $\nu_B(x) = 2B(2x)N(2x)$. Differentiating (24), we obtain

$$\partial_x(\tau x N) = 2B(2x)N(2x) - B(x)N(x)$$

which corresponds to the stationary state linked to the equation

$$\begin{cases} \partial_t n(t,x) + \tau\, \partial_x\big(x n(t,x)\big) = 2B(2x)n(t,2x) - B(x)n(t,x), \\ n(0,x) = n^{(0)}(x), \qquad x \ge 0. \end{cases} \tag{25}$$

Equation (25) exactly corresponds to the behaviour of the tagged cell of Section 5.1 below, in a (weak) sense:

$$n(t, \mathrm{d}x) = \mathbb{P}\big(\chi(t) \in \mathrm{d}x\big),$$

where $\chi(t)$ denotes the size at time t along a branch picked at random, see Section 5.1. Existence and uniqueness of an invariant measure ν_B has an analogy to the existence of a steady state solution for the PDE (25), and the convergence of the empirical measure to the invariant rejoins the stability of the steady state [31]. The equality $\nu_B(x) = 2B(2x)N(2x)$ may be interpreted as follows: $N(x)$ is the steady solution of equation (25), and represents the probability density of a cell population dividing at a rate B and growing at a rate $x\tau$, but when only one offspring remains alive at each division so that the total quantity of cells remains constant. The fraction of dividing cells is represented by the term $B(x)N(x)$ in the equation, with distribution given by $\frac{1}{2}\nu_B(x/2)$, whereas the fraction of newborn cells is $2B(2x)N(2x)$. Equation (24) can be written in terms of BN as

$$B(y) = \frac{\tau y BN(y)}{\int_y^{2y} B(z)N(z)\,\mathrm{d}z}. \tag{26}$$

This also highlights why we obtain a rate of convergence of order $n^{-s/(2s+1)}$ rather than the rate $n^{-s/(2s+3)}$ obtained with indirect measurements as in [12]. In that latter case, we observe a n-sample with distribution N. As shown in [12], one differentiation is necessary to estimate B therefore we have a degree of ill-posedness of order 1. In the setting of the present paper, we rather observe a sample with distribution BN, and B can be recovered directly from (26) and we have here a degree of ill-posedness of order 0.

5. Proofs

We first prove Theorem 1 in Sections 5.1 and 5.2. The strategy consists in obtaining a so-called many-to-one formula (Proposition 3) that enables to relate additive functionals of the whole Markov tree to a special one-dimensional process that consists of following at random a branch on the tree. It suffices to check in Section 5.2 that this randomly tagged branch integrated against appropriate test functions satisfies the desired transport-fragmentation equation. Section 5.3 studies at length the Markov transition $\mathcal{P}_B$. We first prove the key representation formula for B obtained in Proposition 2. We then quantify the geometric ergodicity of the model by a standard Lyapunov technique (Proposition 4). In Section 5.5, we subsequently apply the geometric ergodicity of the transition $\mathcal{P}_B$ by establishing covariance inequalities on a tree in Propositions 5 and 6; these are the crucial tools to later control the convergence rate of the estimator. We need in particular to study the covariance of delta-like functions with supremum norm increasing to infinity with our asymptotic, and this explains the relative technical length of our estimates. This enables to further control in Section 5.6 a rate of convergence for the empirical measure in Propositions 7 and 8. The fact that we work on a tree with an non-reversible Markov transition and delta-like test functions is an extra technical difficulty. Finally, we can prove Theorem 2 in Section 5.7 for the rate of convergence of our estimator with a classical trade-off technique between a bias and a variance term, thanks to the tools developed in the preceding sections and in particular in Section 5.4 where some useful estimates for the invariant measure are established.

The notation $\lesssim$ means inequality up to a constant that does not depend on n. We set $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously. A mapping $f : \mathcal{E} \to \mathbb{R}$ or $g : [0, \infty) \to \mathbb{R}$ is implicitly identified as a function on $\mathcal{S}$ via $f(x, v) = f(x)$ and $g(x, v) = g(v)$.

5.1. A many-to-one formula via a tagged cell

For $u \in \mathcal{U}$, we set $m^i u$ for the ith parent along the genealogy of u. Define

$$\overline{\tau_t^u} = \sum_{i=1}^{|u|} \tau_{m^i u} \zeta_{m^i u} + \tau_t^u (t - b_u) \qquad \text{for } t \in [b_u, b_u + \zeta_u)$$

and 0 otherwise for the cumulated growth rate along its ancestors up to time t. In the same spirit as tagged fragments in fragmentation processes (see the book by Bertoin [7] for instance) we pick a branch at random along the genealogical tree at random: for every $k \geq 1$, if ϑ_k denotes the node of the tagged cell at the kth generation, we have

$$\mathbb{P}(\vartheta_k = u) = 2^{-k} \qquad \text{for every } u \in \mathcal{U} \text{ such that } |u| = k,$$

and 0 otherwise. For $t \geq 0$, the relationship

$$b_{\vartheta_{C_t}} \leq t < b_{\vartheta_{C_t}} + \zeta_{\vartheta_{C_t}}$$

uniquely defines a counting process $(C_t, t \geq 0)$ with $C_0 = 0$. The process C_t enables in turn to define a tagged process of size, growth rate and cumulated growth rate via

$$\big(\chi(t), \mathcal{V}(t), \overline{\mathcal{V}}(t)\big) = \big(\xi_t^{\vartheta_{C_t}}, \tau_t^{\vartheta_{C_t}}, \overline{\tau_t^{\vartheta_{C_t}}}\big) \qquad \text{for } t \in [b_{\vartheta_{C_t}}, b_{\vartheta_{C_t}} + \zeta_{\vartheta_{C_t}})$$

and 0 otherwise. We have the representation

$$\chi(t) = \frac{x \mathrm{e}^{\overline{\mathcal{V}}(t)}}{2^{C_t}} \tag{27}$$

and since $\mathcal{V}(t) \in [e_{\min}, e_{\max}]$, we note that

$$e_{\min} t \leq \overline{\mathcal{V}}(t) \leq e_{\max} t. \tag{28}$$

The behaviour of $(\chi(t), \mathcal{V}(t), \overline{\mathcal{V}}(t))$ can be related to certain functionals of the whole particle system via a so-called many-to-one formula. This is the key tool to obtain Theorem 1.

Proposition 3 (A many-to-one formula). *Work under Assumption 1. For $x \in (0, \infty)$, let $\mathbb{P}_x$ be defined as in Lemma 1. For every $t \geq 0$, we have*

$$\mathbb{E}_x\big[\phi\big(\chi(t), \mathcal{V}(t), \overline{\mathcal{V}}(t)\big)\big] = \mathbb{E}_x\Bigg[\sum_{u \in \mathcal{U}} \xi_t^u \frac{\mathrm{e}^{-\overline{\tau_t^u}}}{x} \phi\big(\xi_t^u, \tau_t^u, \overline{\tau_t^u}\big)\Bigg]$$

for every $\phi : \mathcal{S} \times [0, \infty) \to [0, \infty)$.

Proof. For $v \in \mathcal{U}$, set $I_v = [b_v, b_v + \zeta_v)$. By representation (27), we have

$$\mathbb{E}_x\big[\phi\big(\chi(t), \mathcal{V}(t), \overline{\mathcal{V}}(t)\big)\big] = \mathbb{E}_x\left[\phi\left(\frac{x\mathrm{e}^{\overline{\mathcal{V}}(t)}}{2^{C_t}}, \mathcal{V}(t), \overline{\mathcal{V}}(t)\right)\right]$$
$$= \mathbb{E}_x\left[\sum_{v\in\mathcal{U}} \phi\left(\frac{x\mathrm{e}^{\overline{\tau_t^v}}}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)\mathbf{1}_{\{t\in I_v, v=\vartheta_{C_t}\}}\right].$$

Introduce the discrete filtration $\mathcal{H}_n$ generated by (ξ_u, ζ_u, τ_u) for every u such that $|u| \le n$. Conditioning with respect to $\mathcal{H}_{|v|}$ and noting that on $\{t \in I_v\}$, we have

$$\mathbb{P}(\vartheta_{C_t} = v|\mathcal{H}_{|v|}) = \frac{1}{2^{|v|}} = \frac{\xi_v \mathrm{e}^{-\overline{\tau_{b_v}^v}}}{x},$$

we derive

$$\mathbb{E}_x\left[\sum_{v\in\mathcal{U}} \phi\left(\frac{x\mathrm{e}^{\overline{\tau_t^v}}}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)\mathbf{1}_{\{t\in I_v, v=\vartheta_{C_t}\}}\right] = \mathbb{E}_x\left[\sum_{v\in\mathcal{U}} \xi_v \frac{\mathrm{e}^{-\overline{\tau_{b_v}^v}}}{x}\phi\left(\frac{x\mathrm{e}^{\overline{\tau_t^v}}}{2^{|v|}}, \tau_t^v, \overline{\tau_t^v}\right)\mathbf{1}_{\{t\in I_v\}}\right]$$
$$= \mathbb{E}_x\left[\sum_{u\in\mathcal{U}} \xi_t^u \frac{\mathrm{e}^{-\overline{\tau_t^u}}}{x}\phi\big(\xi_t^u, \tau_t^u, \overline{\tau_t^u}\big)\right].$$

□

5.2. Proof of Theorem 1

We fix $x \in (0, \infty)$ and prove the result for an initial measure μ_x as in Proposition 3. Let $\varphi \in \mathcal{C}_0^1(\mathcal{S})$ be nonnegative. By (10), we have

$$\big\langle n(t,\cdot), \varphi\big\rangle = \mathbb{E}_x\left[\sum_{i=1}^{\infty} \varphi\big(X_i(t), Z_i(t)\big)\right] = \mathbb{E}_x\left[\sum_{u\in\mathcal{U}} \varphi\big(\xi_t^u, \tau_t^u\big)\right]$$

and applying Proposition 3, we derive

$$\big\langle n(t,\cdot), \varphi\big\rangle = x\mathbb{E}_x\left[\varphi\big(\chi(t), \mathcal{V}(t)\big)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right]. \tag{29}$$

For $h > 0$, introduce the difference operator

$$\Delta_h f(t) = h^{-1}\big(f(t+h) - f(t)\big).$$

We plan to study the convergence of $\Delta_h\langle n(t,\cdot), \varphi\rangle$ as $h \to 0$ using representation (29) in restriction to the events $\{C_{t+h} - C_t = i\}$, for $i = 0, 1$ and $\{C_{t+h} - C_t \ge 2\}$. Denote by $\mathcal{F}_t$ the filtration generated by the tagged cell $(\chi(s), \mathcal{V}(s), s \le t)$. The following standard estimate proved in Appendix A.2 will be later useful.

Lemma 1. *Assume that B is continuous. Let $x \in (0,\infty)$ and let μ_x be a probability measure on $\mathcal{S}$ such that $\mu_x(\{x\}\times\mathcal{E}) = 1$. Abbreviate $\mathbb{P}_{\mu_x}$ by $\mathbb{P}_x$. For small $h > 0$, we have*

$$\mathbb{P}_x(C_{t+h} - C_t = 1|\mathcal{F}_t) = B\big(\chi(t)\big)h + h\varepsilon(h),$$

with the property $|\varepsilon(h)| \le \epsilon(h) \to 0$ as $h \to 0$, for some deterministic $\epsilon(h)$, and

$$\mathbb{P}_x(C_{t+h} - C_t \ge 2) \lesssim h^2.$$

Since $\varphi \in \mathcal{C}_0^1(\mathcal{S})$, there exists $c(\varphi) > 0$ such that $\varphi(y,v) = 0$ if $y \ge c(\varphi)$. By (28), we infer

$$\left|\varphi\big(\chi(t),\mathcal{V}(t)\big)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right| \le \sup_{y,v}\varphi(y,v)\frac{\exp(e_{\max}t)}{c(\varphi)}. \tag{30}$$

By Lemma 1 and (30), we derive

$$\mathbb{E}_x\left[\Delta_h\left(\varphi\big(\chi(t),\mathcal{V}(t)\big)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right)\mathbf{1}_{\{C_{t+h}-C_t\ge 2\}}\right] \lesssim h. \tag{31}$$

On the event $\{C_{t+h} - C_t = 0\}$, the process $\mathcal{V}(s)$ is constant for $s \in [t, t+h)$ and so is $\frac{\mathrm{e}^{\overline{\mathcal{V}}(s)}}{\chi(s)}$ thanks to (27). It follows that

$$\Delta_h\left(\varphi\big(\chi(t),\mathcal{V}(t)\big)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right) = \Delta_h\varphi\big(\chi(t),\mathcal{V}(s)\big)\big|_{s=t}\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}$$

on $\{C_{t+h} - C_t = 0\}$ and also

$$\left|\Delta_h\varphi\big(\chi(t),\mathcal{V}(s)\big)\big|_{s=t}\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right| \le \sup_{y,v}\big|\,\partial_y\varphi(y,v)\big|xe_{\max}\frac{\exp(2e_{\max}t)}{c(\varphi)}$$

on $\{C_{t+h} - C_t = 0\}$ likewise. Since $\mathbb{P}_x(C_{t+h} - C_t = 0) \to 1$ as $h \to 0$, by dominated convergence

$$\begin{aligned} &x\mathbb{E}_x\left[\Delta_h\left(\varphi\big(\chi(t),\mathcal{V}(t)\big)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right)\mathbf{1}_{\{C_{t+h}-C_t=0\}}\right] \\ &\quad\to x\mathbb{E}_x\big[\partial_1\varphi\big(\chi(t),\mathcal{V}(t)\big)\mathcal{V}(t)\mathrm{e}^{\overline{\mathcal{V}}(t)}\big] \qquad \text{as } h\to 0. \end{aligned} \tag{32}$$

By Proposition 3 again, this last quantity is equal to $\langle n(t,\mathrm{d}x,\mathrm{d}v), xv\,\partial_x\varphi\rangle$. On $\{C_{t+h} - C_t = 1\}$, we successively have

$$\begin{aligned} \chi(t+h) &= \tfrac{1}{2}\chi(t) + \varepsilon_1(h), \\ \varphi\big(\chi(t+h),\mathcal{V}(t+h)\big) &= \varphi\big(\chi(t)/2,\mathcal{V}(t+h)\big) + \varepsilon_2(h) \end{aligned}$$

and

$$\exp\bigl(\overline{\mathcal{V}}(t+h)\bigr) = \exp\bigl(\overline{\mathcal{V}}(t)\bigr) + \varepsilon_3(h)$$

with the property $|\varepsilon_i(h)| \le \epsilon_1(h) \to 0$ as $h \to 0$, where $\epsilon_1(h)$ is deterministic, thanks to (27) and (28). Moreover,

$$\mathcal{V}(t+h) = \tau_{\vartheta_{C_t+1}} \qquad \text{on } \{C_{t+h} - C_t = 1\}.$$

It follows that

$$\begin{aligned}
&\mathbb{E}_x\left[\varphi\bigl(\chi(t+h), \mathcal{V}(t+h)\bigr)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t+h)}}{\chi(t+h)}\mathbf{1}_{\{C_{t+h}-C_t=1\}}\right] \\
&\quad= \mathbb{E}_x\left[\varphi\bigl(\chi(t)/2, \tau_{\vartheta_{C_t+1}}\bigr)\frac{2\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\mathbf{1}_{\{C_{t+h}-C_t=1\}}\right] + \epsilon_2(h) \\
&\quad= \mathbb{E}_x\left[\varphi\bigl(\chi(t)/2, \tau_{\vartheta_{C_t+1}}\bigr)\frac{2\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\mathbf{1}_{\{C_{t+h}-C_t\ge 1\}}\right] + \epsilon_3(h),
\end{aligned}$$

where $\epsilon_2(h), \epsilon_3(h) \to 0$ as $h \to 0$, and where we used the second part of Lemma 1 in order to obtain the last equality. Conditioning with respect to $\mathcal{F}_t \vee \tau_{\vartheta_{C_t+1}}$ and using that $\{C_{t+h} - C_t \ge 1\}$ and $\tau_{\vartheta_{C_t+1}}$ are independent, applying the first part of Lemma 1, this last term is equal to

$$\begin{aligned}
&\mathbb{E}_x\left[\varphi\bigl(\chi(t)/2, \tau_{\vartheta_{C_t+1}}\bigr)\frac{2\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}B\bigl(\chi(t)\bigr)h\right] + \epsilon_4(h) \\
&\quad= \mathbb{E}_x\left[\int_{\mathcal{E}} \varphi\bigl(\chi(t)/2, v'\bigr)\rho\bigl(\mathcal{V}(t), \mathrm{d}v'\bigr)\frac{2\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}B\bigl(\chi(t)\bigr)h\right] + \epsilon_4(h),
\end{aligned}$$

where $\epsilon_4(h) \to 0$ as $h \to 0$. Finally, using Lemma 1 again, we derive

$$\begin{aligned}
&\mathbb{E}_x\left[\Delta_h\left(\varphi\bigl(\chi(t), \mathcal{V}(t)\bigr)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}\right)\mathbf{1}_{\{C_{t+h}-C_t=1\}}\right] \\
&\quad\to \mathbb{E}_x\left[\left(\int_{\mathcal{E}} 2\varphi\bigl(\chi(t)/2, v'\bigr)\rho\bigl(\mathcal{V}(t), \mathrm{d}v'\bigr) - \varphi\bigl(\chi(t), \mathcal{V}(t)\bigr)\right)\frac{\mathrm{e}^{\overline{\mathcal{V}}(t)}}{\chi(t)}B\bigl(\chi(t)\bigr)\right]
\end{aligned} \tag{33}$$

as $h \to 0$. By Proposition 3, this last quantity is equal to

$$\left\langle n(t, \mathrm{d}x, \mathrm{d}v), \left(\int_{\mathcal{E}} 2\varphi\bigl(x/2, v'\bigr)\rho\bigl(v, \mathrm{d}v'\bigr) - \varphi(x, v)\right)B(x)\right\rangle$$

which, in turn, is equal to

$$\left\langle n(t, 2\mathrm{d}x, \mathrm{d}v), \int_{\mathcal{E}} 4\varphi\bigl(x, v'\bigr)\rho\bigl(v, \mathrm{d}v'\bigr)B(2x)\right\rangle - \bigl\langle n(t, \mathrm{d}x, \mathrm{d}v), \varphi(x, v)B(x)\bigr\rangle$$

by a simple change of variables. Putting together the estimates (31), (32) and (33), we conclude

$$\begin{aligned}&\partial_t\big\langle n(t,\mathrm{d}x,\mathrm{d}v),\varphi\big\rangle-\big\langle n(t,\mathrm{d}x,\mathrm{d}v),xv\,\partial_x\varphi\big\rangle+\big\langle n(t,\mathrm{d}x,\mathrm{d}v)B(x),\varphi\big\rangle\\&\quad=\left\langle n(t,2\mathrm{d}x,\mathrm{d}v),\int_{\mathcal{E}}4\varphi\big(x,v'\big)\rho\big(v,\mathrm{d}v'\big)B(2x)\right\rangle,\end{aligned}$$

which is the dual formulation of (4). The proof is complete.

5.3. Geometric ergodicity of the discrete model

We keep up with the notations of Sections 2 and 3. We first prove Proposition 2.

Proof of Proposition 2. The fact that $\nu_B(\mathrm{d}\mathbf{x})=\nu_B(x,\mathrm{d}v)\,\mathrm{d}x$ readily follows from the representation $\mathcal{P}_B(\mathbf{x},\mathrm{d}\mathbf{x}')=\mathcal{P}_B((x,v),x',\mathrm{d}v')\,\mathrm{d}x'$ together with the invariant measure equation (12). It follows that for every $y\in(0,\infty)$,

$$\begin{aligned}\nu_B\big(y,\mathrm{d}v'\big)&=\int_{\mathcal{S}}\nu_B(x,\mathrm{d}v)\,\mathrm{d}x\mathcal{P}_B\big((x,v),y,\mathrm{d}v'\big)\\&=\frac{B(2y)}{y}\int_{\mathcal{E}}\int_0^{2y}\nu_B(x,\mathrm{d}v)\exp\left(-\int_{x/2}^{y}\frac{B(2s)}{vs}\,\mathrm{d}s\right)\frac{\rho(v,\mathrm{d}v')}{v}\,\mathrm{d}x.\end{aligned}$$

By Assumption 1, we have $\int_{x/2}^{\infty}\frac{B(2s)}{s}\,\mathrm{d}s=\infty$ hence

$$\exp\left(-\int_{x/2}^{y}\frac{B(2s)}{vs}\,\mathrm{d}s\right)=\int_y^{\infty}\frac{B(2s)}{vs}\exp\left(-\int_{x/2}^{s}\frac{B(2s')}{vs'}\,\mathrm{d}s'\right)\mathrm{d}s,$$

therefore $\nu_B(y,\mathrm{d}v')$ is equal to

$$\begin{aligned}&\frac{B(2y)}{y}\int_{\mathcal{E}}\int_0^{2y}\nu_B(x,\mathrm{d}v)\,\mathrm{d}x\int_y^{\infty}\frac{B(2s)}{vs}\exp\left(-\int_{x/2}^{s}\frac{B(2s')}{vs'}\,\mathrm{d}s'\right)\mathrm{d}s\frac{\rho(v,\mathrm{d}v')}{v}\\&=\frac{B(2y)}{y}\int_{\mathcal{S}}\int_{[0,\infty)}\mathbf{1}_{\{x\le 2y,s\ge y\}}v^{-1}\nu_B(x,\mathrm{d}v)\,\mathrm{d}x\mathcal{P}_B\big((x,v),s,\mathrm{d}v'\big)\,\mathrm{d}s.\end{aligned}$$

Integrating with respect to $\mathrm{d}v'$, we obtain the result. □

Geometric ergodicity

We extend $\mathcal{P}_B$ as an operator acting on functions $f:\mathcal{S}\to[0,\infty)$ via

$$\mathcal{P}_Bf(\mathbf{x})=\int_{\mathcal{S}}f(\mathbf{y})\mathcal{P}_B(\mathbf{x},\mathrm{d}\mathbf{y}).$$

If $k\ge 1$ is an integer, define $\mathcal{P}_B^k=\mathcal{P}_B^{k-1}\circ\mathcal{P}_B$.

Proposition 4. *Let* $\mathfrak{c}$ *satisfy Assumption 3. Then, for every* $B \in \mathcal{F}^\lambda(\mathfrak{c})$ *and* $\rho \in \mathcal{M}(\rho_{\min})$, *there exists a unique invariant probability measure of the form* $\nu_B(\mathrm{d}\mathbf{x}) = \nu_B(x, \mathrm{d}v)\,\mathrm{d}x$ *on* $\mathcal{S}$. *Moreover, there exist* $0 < \gamma < 1$, *a function* $\mathbb{V} : \mathcal{S} \to [1, \infty)$ *and a constant* R *such that*

$$\sup_{B \in \mathcal{F}^\lambda(\mathfrak{c}), \rho \in \mathcal{M}(\rho_{\min})} \sup_{|g| \le V} \left| \mathcal{P}_B^k g(\mathbf{x}) - \int_{\mathcal{S}} g(\mathbf{z}) \nu_B(\mathrm{d}\mathbf{z}) \right| \le R\mathbb{V}(\mathbf{x})\gamma^k \tag{34}$$

for every $\mathbf{x} \in \mathcal{S}$, $k \ge 0$, *and where the supremum is taken over all functions* $g : \mathcal{S} \to \mathbb{R}$ *satisfying* $|g(\mathbf{x})| \le \mathbb{V}(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{S}$. *Moreover, under Assumption 4, we can take* $\gamma < \frac{1}{2}$. *Finally, the function* $\mathbb{V}$ *is* ν_B*-integrable for every* $B \in \mathcal{F}^\lambda(\mathfrak{c})$ *and* (34) *is well defined.*

We will show in the proof that the function $\mathbb{V}$ defined in (19) satisfies the properties announced in Proposition 4.

Proof of Proposition 4. We follow the classical line of establishing successively a condition of minorisation, strong aperiodicity and drift for the transition operator $\mathcal{P}_B$ (see, for instance, [6,18,27]. We keep in with the notation of Baxendale [6]). Recall that $0 < e_{\min} \le e_{\max}$ is such that $\mathcal{E} \subset [e_{\min}, e_{\max}]$.

Minorisation condition. Let $B \in \mathcal{F}^\lambda(\mathfrak{c})$. Define

$$\varphi_B(y) = \frac{B(2y)}{e_{\max} y} \exp\left(- \int_0^y \frac{B(2s)}{e_{\min} s}\,\mathrm{d}s \right). \tag{35}$$

Set $\mathcal{C} = (0, r) \times \mathcal{E}$, where r is specified by $\mathfrak{c}$. For any measurable $\mathcal{X} \times A \subset \mathcal{S}$ and $(x, v) \in \mathcal{C}$, we have

$$\begin{aligned} \mathcal{P}_B\big((x, v), \mathcal{X} \times A\big) &= \int_A \rho\big(v, \mathrm{d}v'\big) \int_{\mathcal{X} \cap [x/2, \infty]} \frac{B(2y)}{vy} \exp\left(- \int_{x/2}^y \frac{B(2s)}{vs}\,\mathrm{d}s \right) \mathrm{d}y \\ &\ge \rho_{\min}(A) \int_{\mathcal{X} \cap [r/2, \infty]} \varphi_B(y)\,\mathrm{d}y. \end{aligned}$$

Define

$$\Gamma_B(\mathrm{d}y, \mathrm{d}v) = c_B^{-1} \mathbf{1}_{[r/2, \infty)}(y) \varphi_B(y)\,\mathrm{d}y \frac{\rho_{\min}(\mathrm{d}v)}{\rho_{\min}(\mathcal{E})},$$

where

$$c_B = \rho_{\min}(\mathcal{E}) \int_{r/2}^\infty \varphi_B(y)\,\mathrm{d}y \ge \frac{e_{\min} \rho_{\min}(\mathcal{E})}{e_{\max}} \exp\left(- \frac{L}{e_{\min}} \right) =: \widetilde{\beta} > 0$$

by (16) since $B \in \mathcal{F}^\lambda(\mathfrak{c})$. We have thus exhibited a small set $\mathcal{C}$, a probability measure Γ_B and a constant $\widetilde{\beta} > 0$ so that the minorisation condition

$$\mathcal{P}_B\big((x, v), \mathcal{X} \times A\big) \ge \widetilde{\beta} \Gamma_B(\mathcal{X} \times A) \tag{36}$$

holds for every $(x, v) \in \mathcal{C}$ and $\mathcal{X} \times A \subset \mathcal{S}$, uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$.

Strong aperiodicity condition. We have

$$\begin{aligned}\widetilde{\beta}\Gamma_B(\mathcal{C}) &= \widetilde{\beta}c_B^{-1}\int_{\mathcal{E}}\rho_{\min}(\mathrm{d}v)\int_{r/2}^{r}\frac{B(2y)}{e_{\max}y}\exp\biggl(-\int_{x/2}^{y}\frac{B(2s)}{e_{\min}s}\,\mathrm{d}s\biggr)\,\mathrm{d}y\\ &\geq \widetilde{\beta}c_B^{-1}\int_{r/2}^{r}\varphi_B(y)\,\mathrm{d}y\\ &\geq \widetilde{\beta}\biggl(1-\exp\biggl(-\int_{r/2}^{r}\frac{B(2y)}{e_{\min}y}\,\mathrm{d}y\biggr)\biggr)\\ &\geq \widetilde{\beta}\biggl(1-\exp\biggl(-\frac{\ell}{e_{\min}}\biggr)\biggr)=:\beta>0,\end{aligned}\tag{37}$$

where we applied (16) for the last inequality.

Drift condition. Let $B\in\mathcal{F}^{\lambda}(\mathfrak{c})$. Let $\mathbb{V}:\mathcal{S}\to[1,\infty)$ be continuously differentiable and such that for every $v\in\mathcal{E}$,

$$\lim_{y\to\infty}\mathbb{V}(y,v)\exp\biggl(-2^{\lambda}\frac{m}{v\lambda}y^{\lambda}\biggr)=0.\tag{38}$$

For $x\geq r$, by (17) and integrating by part with the boundary condition (38), we have, for every $v\in\mathcal{E}$,

$$\begin{aligned}\mathcal{P}_B\mathbb{V}(x,v) &= \int_{\mathcal{E}}\rho\bigl(v,\mathrm{d}v'\bigr)\int_{x/2}^{\infty}\mathbb{V}\bigl(y,v'\bigr)\frac{B(2y)}{vy}\exp\biggl(-\int_{x/2}^{y}\frac{B(2s)}{vs}\,\mathrm{d}s\biggr)\,\mathrm{d}y\\ &\leq \int_{\mathcal{E}}\rho\bigl(v,\mathrm{d}v'\bigr)\int_{x/2}^{\infty}\partial_y\mathbb{V}\bigl(y,v'\bigr)\exp\biggl(-\frac{m2^{\lambda}}{v}\int_{x/2}^{y}s^{\lambda-1}\,\mathrm{d}s\biggr)\,\mathrm{d}y\\ &\leq \exp\biggl(\frac{m}{v\lambda}x^{\lambda}\biggr)\int_{\mathcal{E}}\rho\bigl(v,\mathrm{d}v'\bigr)\int_{(m2^{\lambda})/(v\lambda)(x/2)^{\lambda}}^{\infty}\mathbb{V}\biggl(\biggl(y\frac{v\lambda}{m2^{\lambda}}\biggr)^{1/\lambda},v'\biggr)\mathrm{e}^{-y}\,\mathrm{d}y.\end{aligned}$$

Pick $\mathbb{V}(x,v)=\mathbb{V}(x)=\exp(\frac{m}{e_{\min}\lambda}x^{\lambda})$ defined in (19) and note that (38) is satisfied for an appropriate choice of $e_{\min}$ and $e_{\max}$ since $2^{\lambda}>\sup\mathcal{E}/\inf\mathcal{E}$. With this choice, we further infer

$$\begin{aligned}\mathcal{P}_B\mathbb{V}(x,v) &\leq \mathbb{V}(x,v)\int_{\mathcal{E}}\rho\bigl(v,\mathrm{d}v'\bigr)\int_{(m2^{\lambda})/(v\lambda)(x/2)^{\lambda}}^{\infty}\exp\bigl(-\bigl(1-2^{-\lambda}\bigr)y\bigr)\,\mathrm{d}y\\ &\leq \mathbb{V}(x,v)\frac{1}{1-2^{-\lambda}}\exp\biggl(-\bigl(1-2^{-\lambda}\bigr)\frac{m}{v\lambda}r^{\lambda}\biggr)\rho_{\max}(\mathcal{E})\end{aligned}$$

since $x\geq r$. Recall that

$$\delta(\mathfrak{c})=\frac{1}{1-2^{-\lambda}}\exp\biggl(-\bigl(1-2^{-\lambda}\bigr)\frac{m}{e_{\max}\lambda}r^{\lambda}\biggr)\rho_{\max}(\mathcal{E}).$$

We obtain, for $x\geq r$ and $v\in\mathcal{E}$

$$\mathcal{P}_B\mathbb{V}(x,v)\leq\delta(\mathfrak{c})\mathbb{V}(x,v)\tag{39}$$

and we have $\delta(\mathfrak{c}) < 1$ by Assumption 3. We next need to control $\mathcal{P}_B\mathbb{V}$ outside $x \in [r, \infty)$, that is on the small set $\mathcal{C}$. For every $(x, v) \in \mathcal{C}$, we have

$$\begin{aligned}\mathcal{P}_B\mathbb{V}(x,v) &\le \int_{\mathcal{E}} \rho\big(v, \mathrm{d}v'\big)\bigg(\int_{x/2}^{r/2} \mathbb{V}\big(y, v'\big)\frac{B(2y)}{vy}\,\mathrm{d}y \\ &\qquad + \int_{r/2}^{\infty} \mathbb{V}\big(y, v'\big)\frac{B(2y)}{vy}\exp\bigg(-\int_{r/2}^{y}\frac{B(2s)}{vs}\,\mathrm{d}s\bigg)\,\mathrm{d}y\bigg) \qquad (40)\\ &\le \rho_{\max}(\mathcal{E})\Big(e_{\min}^{-1}\sup_{y\in[0,r]}\mathbb{V}(y)L + \delta(\mathfrak{c})\mathbb{V}(r/2)\Big) =: Q < \infty,\end{aligned}$$

where we used (16) and the fact that $B \in \mathcal{F}^{\lambda}(\mathfrak{c})$. Combining together (39) and (40), we conclude

$$\mathcal{P}_B\mathbb{V}(\mathbf{x}) \le \delta(\mathfrak{c})\mathbb{V}(\mathbf{x})\mathbf{1}_{\{\mathbf{x}\notin\mathcal{C}\}} + Q\mathbf{1}_{\{\mathbf{x}\in\mathcal{C}\}}. \qquad (41)$$

Completion of proof of Proposition 4. The minorisation condition (36) together with the strong aperiodicity condition (37) and the drift condition (41) imply inequality (34) by Theorem 1.1 in Baxendale [6], with R and γ that explicitly depend on $\delta(\mathfrak{c})$, β, $\widetilde{\beta}$, V and Q. By construction, this bound is uniform in $B \in \mathcal{F}^{\lambda}(\mathfrak{c})$ and $\rho \in \mathcal{M}(\rho_{\min})$. More specifically, we have

$$\gamma < \min\big\{\max\big\{\delta(c), \gamma_{\mathbb{V},B}\big\}, 1\big\}$$

therefore under Assumption 3 we have $\gamma < 1$ and under Assumption 4, we obtain the improvement $\gamma < \frac{1}{2}$. □

5.4. Further estimates on the invariant probability

Lemma 2. *For any* $\mathfrak{c}$ *such that Assumption 3 is satisfied and any compact interval* $\mathcal{D} \subset (0, \infty)$, *we have*

$$\sup_{B\in\mathcal{F}^{\lambda}(\mathfrak{c})\cap\mathcal{H}^{s}(\mathcal{D},M)}\ \sup_{x\in 2^{-1}\mathcal{D}} \nu_B(x) < \infty,$$

with $\nu_B(x) = \int_{\mathcal{E}} \nu_B(x, \mathrm{d}v)$.

Proof. Since $B \in \mathcal{F}^{\lambda}(\mathfrak{c})$, ν_B is well-defined and satisfies

$$\nu_B(x, \mathrm{d}v) = \frac{B(2x)}{x}\int_{\mathcal{E}}\int_0^{2x} \nu_B\big(y, \mathrm{d}v'\big)\,\mathrm{d}y\exp\bigg(-\int_{y/2}^{x}\frac{B(2s)}{v's}\,\mathrm{d}s\bigg)\frac{\rho(v', \mathrm{d}v)}{v'}.$$

Hence, $\nu_B(x, \mathrm{d}v) \le B(2x)(e_{\min}x)^{-1}\rho_{\max}(\mathrm{d}v)$ and we also have $\nu_B(x) \le B(2x)(e_{\min}x)^{-1} \times \rho_{\max}(\mathcal{E})$. Since $B \in \mathcal{H}^{s}(\mathcal{D}, M)$ implies $\sup_{x\in 2^{-1}\mathcal{D}} B(2x) = \|B\|_{L^{\infty}(\mathcal{D})} \le M$, the conclusion follows. □

Lemma 3. *For any* $\mathfrak{c}$ *such that Assumption 3 is satisfied, there exists a constant* $d(\mathfrak{c}) \geq 0$ *such that for any compact interval* $\mathcal{D} \subset (d(\mathfrak{c}), \infty)$, *we have*

$$\inf_{B \in \mathcal{F}^{\lambda}(\mathfrak{c})} \inf_{x \in \mathcal{D}} \varphi_B(x)^{-1} \nu_B(x) > 0,$$

where $\varphi_B(x)$ *is defined in* (35).

Proof. Let $g : [0, \infty) \to [0, \infty)$ satisfy $g(x) \leq \mathbb{V}(x) = \exp(\frac{m}{e_{\min}\lambda} x^{\lambda})$ for every $x \in [0, \infty)$. By Proposition 4, we have

$$\sup_{B \in \mathcal{F}^{\nu}_{\lambda}(\mathfrak{c})} \int_{[0,\infty)} g(x) \nu_B(x) \, \mathrm{d}x < \infty, \tag{42}$$

as a consequence of (34) with $n = 1$ together with the property that $\sup_{B \in \mathcal{F}^{\nu}_{\lambda}(\mathfrak{c})} \mathcal{P}_B \mathbb{V}(\mathbf{x}) < \infty$ for every $\mathbf{x} \in \mathcal{S}$, as follows from (41) in the proof of Proposition 4. Next, for every $x \in (0, \infty)$, we have

$$\int_{2x}^{\infty} \nu_B(y) \, \mathrm{d}y \leq \exp\left(-\frac{m}{e_{\min}\lambda} (2x)^{\lambda} \right) \int_{[0,\infty)} \mathbb{V}(y) \nu_B(y) \, \mathrm{d}y$$

and this bound is uniform in $B \in \mathcal{F}^{\lambda}(\mathfrak{c})$ by (42). Therefore, for every $x \in (0, \infty)$, we have

$$\sup_{B \in \mathcal{F}^{\nu}_{\lambda}(\mathfrak{c})} \int_{2x}^{\infty} \nu_B(y) \, \mathrm{d}y \leq c(\mathfrak{c}) \exp\left(-\frac{m}{e_{\min}\lambda} (2x)^{\lambda} \right) \tag{43}$$

for some $c(\mathfrak{c}) > 0$. Let

$$d(\mathfrak{c}) > \left(\frac{e_{\min}\lambda 2^{-\lambda}}{m} \log c(\mathfrak{c}) \right)^{1/\lambda} \mathbf{1}_{\{c(\mathfrak{c}) \geq 1\}}. \tag{44}$$

By definition of ν_B, for every $x \in (0, \infty)$, we now have

$$\begin{aligned}
\nu_B(x, \mathrm{d}v) &= \frac{B(2x)}{x} \int_{\mathcal{E}} \int_0^{2x} \nu_B\big(y, \mathrm{d}v'\big) \exp\left(-\int_{y/2}^{x} \frac{B(2s)}{v's} \, \mathrm{d}s \right) \mathrm{d}y \frac{\rho(v', \mathrm{d}v)}{v'} \\
&\geq \frac{B(2x)}{e_{\max}x} \exp\left(-\int_0^{x} \frac{B(2s)}{e_{\min}s} \, \mathrm{d}s \right) \int_0^{2x} \nu_B(y) \, \mathrm{d}y \rho_{\min}(\mathrm{d}v) \\
&\geq \frac{B(2x)}{e_{\max}x} \exp\left(-\int_0^{x} \frac{B(2s)}{e_{\min}s} \, \mathrm{d}s \right) \left(1 - c(\mathfrak{c}) \exp\left(-\frac{m}{e_{\min}\lambda} (2x)^{\lambda} \right) \right) \rho_{\min}(\mathrm{d}v),
\end{aligned}$$

where we used (43) for the last inequality. By (44), for $x \geq d(\mathfrak{c})$ we have

$$\left(1 - c(\mathfrak{c}) \exp\left(-\frac{m}{e_{\min}\lambda} (2x)^{\lambda} \right) \right) > 0$$

and the conclusion readily follows by integration. □

5.5. Covariance inequalities

If $u, w \in \mathcal{U}$, we define $a(u, w)$ as the node of the most recent common ancestor between u and w. Introduce the distance

$$\mathbb{D}(u, w) = |u| + |w| - 2\big|a(u, w)\big|.$$

Proposition 5. *Work under Assumption 3. Let μ be a probability distribution on $\mathcal{S}$ such that $\int_{\mathcal{S}} \mathbb{V}(\mathbf{x})^2 \mu(\mathrm{d}\mathbf{x}) < \infty$. Let $G : \mathcal{S} \to \mathbb{R}$ and $H : [0, \infty) \to \mathbb{R}$ be two bounded functions. Define*

$$Z(\xi_{u^-}, \tau_{u^-}, \xi_u) = G(\xi_{u^-}, \tau_{u^-})H(\xi_u) - \mathbb{E}_{\nu_B}\big[G(\xi_{u^-}, \tau_{u^-})H(\xi_u)\big].$$

For any $u, w \in \mathcal{U}$ with $|u|, |w| \geq 1$, we have

$$\big|\mathbb{E}_\mu\big[Z(\xi_{u^-}, \tau_{u^-}, \xi_u)Z(\xi_{w^-}, \tau_{w^-}, \xi_w)\big]\big| \lesssim \gamma^{\mathbb{D}(u,w)} \tag{45}$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$, with γ and ν_B defined in (34) of Proposition 4.

Proof. In view of (45), with no loss of generality, we may (and will) assume that for every $(x, v) \in \mathcal{S}$

$$\big|G(x, v)\big| \leq \mathbb{V}(x) \quad \text{and} \quad \big|H(x)\big| \leq \mathbb{V}(x). \tag{46}$$

Applying repeatedly the Markov property along the branch that joins the nodes $a^-(u, w) := a(u^-, w^-)$ and w, we have

$$\begin{aligned}
&\mathbb{E}_\mu\big[G(\xi_{u^-}, \tau_{u^-})H(\xi_u)|\xi_{a^-(u,w)}, \tau_{a^-(u,w)}\big] \\
&\quad = \mathcal{P}_B^{|u^-|-|a^-(u,w)|}(G\mathcal{P}_B H)(\xi_{a^-(u,w)}, \tau_{a^-(u,w)}) \\
&\quad = \mathcal{P}_B^{|u|-|a(u,w)|}(G\mathcal{P}_B H)(\xi_{a^-(u,w)}, \tau_{a^-(u,w)})
\end{aligned}$$

with an analogous formula for $G(\xi_{w^-}, \tau_{w^-})H(\xi_w)$. Conditioning with respect to $\xi_{a^-(u,w)}$, $\tau_{a^-(u,w)}$, it follows that

$$\begin{aligned}
&\mathbb{E}_\mu\big[Z(\xi_{u^-}, \tau_{u^-}, \xi_u)Z(\xi_{w^-}, \tau_{w^-}, \xi_w)\big] \\
&\quad = \mathbb{E}_\mu\Big[\big(\mathcal{P}_B^{|u|-|a(u,w)|}(G\mathcal{P}_B H)(\xi_{a^-(u,w)}, \tau_{a^-(u,w)}) - \mathbb{E}_{\nu_B}\big[G\mathcal{P}_B H(\xi_\varnothing, \tau_\varnothing)\big]\big) \\
&\qquad \times \big(\mathcal{P}_B^{|w|-|a(u,w)|}(G\mathcal{P}_B H)(\xi_{a^-(u,w)}, \tau_{a^-(u,w)}) - \mathbb{E}_{\nu_B}\big[G\mathcal{P}_B H(\xi_\varnothing, \tau_\varnothing)\big]\big)\Big].
\end{aligned}$$

Applying Proposition 4 thanks to Assumption 3 and (46), we further infer

$$\begin{aligned}
\mathbb{E}_\mu\big[Z(\xi_{u^-}, \tau_{u^-}, \xi_u)Z(\xi_{w^-}, \tau_{w^-}, \xi_w)\big] &\leq R^2 \sup_x H(x)^2 \mathbb{E}_\mu\big[\mathbb{V}(\xi_{a^-(u,w)})^2\big]\gamma^{\mathbb{D}(u,w)} \\
&\lesssim \int_{\mathcal{S}} \mathcal{P}_B^{|a^-(u,w)|}\big(\mathbb{V}^2\big)(\mathbf{x})\mu(\mathrm{d}\mathbf{x})\gamma^{\mathbb{D}(u,w)}.
\end{aligned}$$

We leave to the reader the straightfoward task to check that the choice of $\mathbb{V}$ in (19) implies that $\mathbb{V}^2$ satisfies (41). It follows that Proposition 4 applies, replacing $\mathbb{V}$ by $\mathbb{V}^2$ in (34). In particular,

$$\sup_{B\in\mathcal{F}^\lambda(\mathfrak{c})} \mathcal{P}_B^{|a^-(u,w)|}\big(\mathbb{V}^2\big)(\mathbf{x}) \lesssim 1+\mathbb{V}(\mathbf{x})^2. \tag{47}$$

Since $\mathbb{V}^2$ is μ-integrable by assumption, inequality (45) follows. □

Proposition 6. *Work under Assumption 3. Let μ be a probability on $\mathcal{S}$ such that $\int_{\mathcal{S}} \mathbb{V}(\mathbf{x})^2\mu(\mathrm{d}\mathbf{x}) < \infty$. Let x_0 be in the interior of $\frac{1}{2}\mathcal{D}$. Let $H:\mathbb{R}\to\mathbb{R}$ be bounded with compact support. Set*

$$\widetilde{H}\left(\frac{\xi_u - x_0}{h}\right) = H\left(\frac{\xi_u - x_0}{h}\right) - \mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_\varnothing - x_0}{h}\right)\right].$$

For any $u, w \in \mathcal{U}$ with $|u|, |w| \geq 1$, we have

$$\left|\mathbb{E}_\mu\left[\widetilde{H}\left(\frac{\xi_u - x_0}{h}\right)\widetilde{H}\left(\frac{\xi_w - x_0}{h}\right)\right]\right| \lesssim \gamma^{\mathbb{D}(u,w)} \wedge h\gamma^{\mathbb{D}(u,a(u,w))\vee\mathbb{D}(w,a(u,w))} \tag{48}$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(\mathcal{D}, M)$ for sufficiently small $h > 0$.

Proof. The first part of the estimate in the right-hand side of (48) is obtained by letting $G = 1$ in (45). We turn to the second part. Repeating the same argument as for (45) and conditioning with respect to $\xi_{a(u,w)}$, we obtain

$$\begin{aligned}
&\mathbb{E}_\mu\left[\widetilde{H}\left(\frac{\xi_u - x_0}{h}\right)\widetilde{H}\left(\frac{\xi_w - x_0}{h}\right)\right] \\
&\quad= \mathbb{E}_\mu\left[\left(\mathcal{P}_B^{|u|-|a(u,w)|}H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right) - \mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_\varnothing - x_0}{h}\right)\right]\right)\right. \\
&\qquad\left.\times\left(\mathcal{P}_B^{|w|-|a(u,w)|}H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right) - \mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_\varnothing - x_0}{h}\right)\right]\right)\right].
\end{aligned} \tag{49}$$

Assume with no loss of generality that $|u| \leq |w|$ (otherwise, the same subsequent arguments apply exchanging the roles of u and w). On the one hand, applying (34) of Proposition 4, we have

$$\left|\mathcal{P}_B^{|w|-|a(u,w)|}H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right) - \mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_\varnothing - x_0}{h}\right)\right]\right| \leq R\mathbb{V}(\xi_{a(u,w)})\gamma^{|w|-|a(u,w)|}. \tag{50}$$

On the other hand, identifying H as a function defined on $\mathcal{S}$, for every $(x, v) \in \mathcal{S}$, we have

$$\begin{aligned}
&\left|\mathcal{P}_B H\left(\frac{x - x_0}{h}\right)\right| \\
&\quad= \left|\int_{x/2}^\infty H\big(h^{-1}(y - x_0)\big)\frac{B(2y)}{vy}\exp\left(-\int_{x/2}^y \frac{B(2s)}{vs}\,\mathrm{d}s\right)\mathrm{d}y\right|
\end{aligned} \tag{51}$$

$$\leq \int_{[0,\infty)} \big|H\big(h^{-1}(y-x_0)\big)\big| \frac{B(2y)}{e_{\min} y}\,\mathrm{d}y$$
$$\leq e_{\min}^{-1} \sup_{y\in\{x_0+h\,\mathrm{supp}(H)\}} \frac{B(2y)}{y} h \int_{[0,\infty)} \big|H(x)\big|\,\mathrm{d}x \lesssim h.$$

Indeed, since x_0 is in the interior of $\frac{1}{2}\mathcal{D}$ we have $\{x_0 + h\,\mathrm{supp}(H)\} \subset \frac{1}{2}\mathcal{D}$ for small enough h hence $\sup_{y\in\{x_0+h\,\mathrm{supp}(H)\}} B(2y) \leq M$. Now, since $\mathcal{P}_B$ is a positive operator and $\mathcal{P}_B\mathbf{1} = \mathbf{1}$, we derive

$$\left|\mathcal{P}_B^{|u|-|a(u,w)|} H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right)\right| \lesssim h \tag{52}$$

as soon as $|u| - |a(u,w)| \geq 1$, uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(\mathcal{D}, M)$. If $|u| = |a(u,w)|$, since $\int_{\mathcal{E}} \nu_B(\mathrm{d}x, \mathrm{d}v) = \nu_B(x)\,\mathrm{d}x$, we obtain in the same way

$$\left|\mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right)\right]\right| \leq \int_{[0,\infty)} \left|H\left(\frac{x-x_0}{h}\right)\right| \nu_B(x)\,\mathrm{d}x \lesssim h \tag{53}$$

using Lemma 2. We have $|\mathbb{E}_{\nu_B}[H(\frac{\xi_u - x_0}{h})]| \lesssim h$ likewise. Putting together (52) and (53), we derive

$$\left|\mathcal{P}_B^{|u|-|a(u,w)|} H\left(\frac{\xi_{a(u,w)} - x_0}{h}\right) - \mathbb{E}_{\nu_B}\left[H\left(\frac{\xi_u - x_0}{h}\right)\right]\right| \lesssim h$$

and this estimate is uniform in $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(\mathcal{D}, M)$. In view of (49) and (50), we obtain

$$\mathbb{E}_\mu\left[\widetilde{H}\left(\frac{\xi_u - x_0}{h}\right)\widetilde{H}\left(\frac{\xi_w - x_0}{h}\right)\right] \lesssim h\gamma^{|w|-|a(u,w)|}\mathbb{E}_\mu\big[\mathbb{V}(\xi_{a(u,w)})\big].$$

We conclude in the same way as in Proposition 5. □

5.6. Rate of convergence for the empirical measure

For every $y \in (0,\infty)$ and $u \in \mathcal{U}$ with $|u| \geq 1$, define

$$D(y) = \mathbb{E}_{\nu_B}\left[\frac{1}{\tau_{u^-}}\mathbf{1}_{\{\xi_{u^-}\leq 2y, \xi_u \geq y\}}\right], \tag{54}$$

$$D_n(y) = n^{-1}\sum_{u\in\mathcal{U}_n} \frac{1}{\tau_{u^-}}\mathbf{1}_{\{\xi_{u^-}\leq 2y, \xi_u \geq y\}} \tag{55}$$

and

$$D_n(y)_\varpi = D_n(y) \vee \varpi.$$

Proposition 7. *Work under Assumption 3 in the sparse tree case and Assumption 4 in the full tree case. Let μ be a probability on $\mathcal{S}$ such that $\int_{\mathcal{S}} \mathbb{V}(\mathbf{x})^2 \mu(\mathrm{d}\mathbf{x}) < \infty$. If $1 \geq \varpi = \varpi_n \to 0$ as $n \to \infty$, we have*

$$\sup_{y \in \mathcal{D}} \mathbb{E}_\mu\big[\big(D_n(y)_{\varpi_n} - D(y)\big)^2\big] \lesssim n^{-1} \tag{56}$$

uniformy in $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(2^{-1}\mathcal{D}, M)$ and $\rho \in \mathcal{M}(\rho_{\min}, \rho_{\max})$.

We first need the following estimate.

Lemma 4. *Work under Assumption 3. Let $d(\mathfrak{c})$ be defined as in Lemma 3. For every compact interval $\mathcal{D} \subset (d(\mathfrak{c}), \infty)$ such that $\inf \mathcal{D} \leq r/2$, we have*

$$\inf_{B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(2^{-1}\mathcal{D}, M)} \inf_{y \in \mathcal{D}} D(y) > 0.$$

Proof. By (13) and the definition of φ_B in (35), we readily have

$$D(y) = \frac{1}{e_{\max}} \varphi_B(y)^{-1} \nu_B(y) \exp\bigg(- \int_0^y \frac{B(2s)}{e_{\min} s} \, \mathrm{d}s \bigg).$$

Since $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(2^{-1}\mathcal{D}, M)$, by applying (16) and (17) successively, we obtain

$$\begin{aligned} \int_0^{\sup \mathcal{D}} \frac{B(2s)}{e_{\min} s} \, \mathrm{d}s &\leq e_{\min}^{-1} L + \int_{r/2}^{\sup \mathcal{D}} \frac{B(2s)}{e_{\min} s} \, \mathrm{d}s \\ &\leq e_{\min}^{-1} \bigg(L + M \log \frac{\sup \mathcal{D}}{r/2} \bigg) < \infty, \end{aligned}$$

where we used that $\inf \mathcal{D} \leq r/2$. It follows that

$$\inf_{y \in \mathcal{D}} \exp\bigg(- \int_0^y \frac{B(2s)}{e_{\min} s} \, \mathrm{d}s \bigg) \geq \exp\bigg(- e_{\min}^{-1} \bigg(L + M \log \frac{\sup \mathcal{D}}{r/2} \bigg) \bigg) > 0$$

and Lemma 4 follows by applying Lemma 3. □

Proof of Proposition 7. Since $D_n(y)$ is bounded, we have

$$\big(D_n(y)_{\varpi_n} - D(y)\big)^2 \lesssim \big(D_n(y) - D(y)\big)^2 + \mathbf{1}_{\{D_n(y) < \varpi_n\}}. \tag{57}$$

Next, take n sufficiently large, so that

$$0 < \varpi_n \leq q = \frac{1}{2} \inf_{B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(2^{-1}\mathcal{D}, M)} \inf_{y \in \mathcal{D}} D(y)$$

a choice which is possible thanks to Lemma 4. Since

$$\big\{ D_n(y) < \varpi_n \big\} \subset \big\{ D_n(y) - D(y) < -q \big\},$$

integrating (57), we have that $\mathbb{E}_\mu[(D_n(y)_{\varpi_n} - D(y))^2]$ is less than a constant times

$$\mathbb{E}_\mu\big[\big(D_n(y) - D(y)\big)^2\big] + \mathbb{P}_\mu\big(\big|D_n(y) - D(y)\big| \geq q\big),$$

which in turn is less than a constant times $\mathbb{E}_\mu[(D_n(y) - D(y))^2]$. Set $G(x, v) = \frac{1}{v}\mathbf{1}_{\{x \leq 2y\}}$ and $H(x) = \mathbf{1}_{\{x \geq y\}}$ and note that G and H are bounded on $\mathcal{S}$ (and also uniformly in $y \in \mathcal{D}$). It follows that

$$D_n(y) - D(y) = n^{-1} \sum_{u \in \mathcal{U}_n} \big(G(\xi_{u^-}, \tau_{u^-})H(\xi_u) - \mathbb{E}_{\nu_B}\big[G(\xi_{u^-}, \tau_{u^-})H(\xi_u)\big]\big).$$

We then apply (45) of Proposition 5 to infer, with the same notation that

$$\begin{aligned}
&\mathbb{E}_\mu\big[\big(D_n(y) - D(y)\big)^2\big] \\
&\quad = n^{-2} \sum_{u,w \in \mathcal{U}_n} \mathbb{E}_\mu\big[Z(\xi_{u^-}, \tau_{u^-}, \xi_u)Z(\xi_{w^-}, \tau_{w^-}, \xi_w)\big] \\
&\quad \lesssim n^{-2} \sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)}
\end{aligned}$$

uniformly in $y \in \mathcal{D}$ and $B \in \mathcal{F}^\lambda(\mathfrak{c})$. We further separate the sparse and full tree cases.

The sparse tree case. We have $\sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} = \sum_{1 \leq |u|, |w| \leq n} \gamma^{||u|-|w||}$ and by Proposition 4, this last quantity is of order n.

The full tree case. We have $n \sim 2^{N_n}$, where N_n is the number of generations used to expand $\mathcal{U}_n$. We evaluate

$$\mathcal{N}_k = \sum_{|u|=k} \sum_{w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} \qquad \text{for } k = 0, \ldots, N. \tag{58}$$

For $k = 0$, we have

$$\mathcal{N}_0 = 1 + 2\gamma + 4\gamma^2 + \cdots + 2^{N_n}\gamma^{N_n} = \frac{1 - (2\gamma)^{N_n+1}}{1 - 2\gamma} =: \phi_\gamma(N_n).$$

Under Assumption 4, by Proposition 4, we have $\gamma < \frac{1}{2}$ therefore $\phi_\gamma(N_n)$ is bounded as $n \to \infty$. For $k = 1$, if we start with the node $u = (\varnothing, 0)$, then the contribution of its descendants in (58) is given by $\phi_\gamma(N_n - 1)$, to which we must add γ for its ancestor corresponding to the node $u = \varnothing$ and also $\gamma\phi_\gamma(N_n)$ for the contribution of the second lineage of the node $u = \varnothing$. Finally, we must repeat the argument for the node $u = (\varnothing, 1)$. We obtain

$$\mathcal{N}_1 = 2\big(\phi_\gamma(N_n - 1) + \gamma + \gamma^2\phi_\gamma(N_n - 1)\big).$$

More generally, proceeding in the same manner, we derive

$$\begin{aligned}
\mathcal{N}_k = 2^k\big(&\phi_\gamma(N_n - k) + \big(\gamma + \gamma^2\phi_\gamma(N_n - k)\big) + \cdots \\
&+ \gamma^i + \gamma^{i+1}\phi_\gamma\big(N_n - k + (i-1)\big) + \cdots + \big(\gamma^k + \gamma^{k+1}\phi_\gamma(N_n - 1)\big)\big)
\end{aligned} \tag{59}$$

for $k = 1, \dots, N_n$, and this last quantity is of order 2^k. It follows that

$$\sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} = \sum_{k=0}^{N_n} \mathcal{N}_k \lesssim \sum_{k=1}^{N_n} 2^k \lesssim 2^{N_n} \lesssim n$$

and the conclusion follows likewise.

Putting together the sparse and full tree case, we obtain the proposition. □

Let $\widehat{\nu}_n(\mathrm{d}y) = n^{-1} \sum_{u \in \mathcal{U}_n} \delta_{\xi_u}(\mathrm{d}y)$ denote the empirical measure of the observation $(\xi_u, u \in \mathcal{U}_n)$.

Proposition 8. *Work under Assumption 3 in the sparse tree case and Assumption 4 in the full tree case. Let μ be a probability on $\mathcal{S}$ such that $\int_{\mathcal{S}} \mathbb{V}(\mathbf{x})^2 \mu(\mathrm{d}\mathbf{x}) < \infty$. We have*

$$\sup_{y \in \mathcal{D}} \mathbb{E}_\mu\big[\big(K_{h_n} \star \widehat{\nu}_n(y) - K_{h_n} \star \nu_B(y)\big)^2\big] \lesssim |\log h_n| (n h_n)^{-1} \tag{60}$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$.

Proof. We have, with the notation of Proposition 6

$$\begin{aligned}
&\mathbb{E}_\mu\big[\big(K_{h_n} \star \widehat{\nu}_n(y) - K_{h_n} \star \nu_B(y)\big)^2\big] \\
&= (n h_n)^{-2} \mathbb{E}_\mu\left[\left(\sum_{u \in \mathcal{U}_n} K\left(\frac{\xi_u - y}{h_n}\right) - \mathbb{E}_{\nu_B}\left[K\left(\frac{\xi_u - y}{h_n}\right)\right]\right)^2\right] \\
&= (n h_n)^{-2} \sum_{u,w \in \mathcal{U}_n} \mathbb{E}_\mu\left[\widetilde{K}\left(\frac{\xi_u - y}{h_n}\right)\widetilde{K}\left(\frac{\xi_w - y}{h_n}\right)\right] \\
&\lesssim (n h_n)^{-2} \sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} \wedge h_n \gamma^{\mathbb{D}(u,a(u,w)) \vee \mathbb{D}(w,a(u,w))}
\end{aligned} \tag{61}$$

by applying (48) of Proposition 6. It remains to estimate (61).

The sparse tree case. We have $a(u, w) = u$ if $|u| \le |w|$ and $a(u, w) = w$ otherwise. It follows that

$$\mathbb{E}_\mu\big[\big(K_{h_n} \star \widehat{\nu}_n(y) - K_{h_n} \star \nu_B(y)\big)^2\big] \lesssim n^{-2} h_n^{-1} \sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)},$$

and since $\sum_{u,w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} = \sum_{1 \le |u|,|w| \le n} \gamma^{||u|-|w||}$ is of order n as soon as $\gamma < 1$, we obtain the result.

The full tree case. The computations are a bit more involved. Let us evaluate

$$\sum_{|u|=k} \sum_{w \in \mathcal{U}_n} \gamma^{\mathbb{D}(u,w)} \wedge h_n \gamma^{\mathbb{D}(u,a(u,w)) \vee \mathbb{D}(w,a(u,w))}.$$

We may repeat the argument displayed in (59) in order to evaluate the contribution of the term involving $\gamma^{\mathbb{D}(u,a(u,w))}$. However, in the estimate $\mathcal{N}_k$, each term $\gamma^i + \gamma^{i+1}\phi_\gamma(N_n - k + (i-1))$ in formula (59) may be replaced by $h_n(\gamma^i + \gamma\phi_\gamma(N_n - k + (i-1)))$ up to constants. This corresponds to the correction given by $h_n\gamma^{\mathbb{D}(u,a(u,w))\vee\mathbb{D}(w,a(u,w))}$. As a consequence, we obtain

$$\begin{aligned}
&\sum_{|u|=k}\sum_{w\in\mathcal{U}_n}\gamma^{\mathbb{D}(u,w)}\wedge h_n\gamma^{\mathbb{D}(u,a(u,w))\vee\mathbb{D}(w,a(u,w))}\\
&\quad\lesssim 2^k\sum_{i=1}^{k}h_n\big(\gamma^i+\gamma\phi_\gamma(N_n-k+i-1)\big)\wedge\gamma^i\big(1+\gamma\phi_\gamma(N_n-k+i-1)\big)\\
&\quad\lesssim 2^k\sum_{i=1}^{k}h_n\wedge\gamma^i.
\end{aligned}$$

Define $k_n^\star = \lfloor\frac{|\log h_n|}{|\log\gamma|}\rfloor$. We readily derive

$$2^k\sum_{i=1}^{k}h_n\wedge\gamma^i = 2^k\left(\sum_{i=1}^{k_n^\star}h_n+\sum_{i=k_n^\star+1}^{k}\gamma^i\right)\lesssim 2^k h_n|\log h_n|,$$

ignoring the second term if $k_n^\star+1\geq k$. Going back to (61), it follows that

$$\begin{aligned}
&(nh_n)^{-2}\sum_{u,w\in\mathcal{U}_n}\gamma^{\mathbb{D}(u,w)}\wedge h_n\gamma^{\mathbb{D}(u,a(u,w))\vee\mathbb{D}(w,a(u,w))}\\
&\quad=(nh_n)^{-2}\sum_{k=0}^{N_n}\sum_{|u|=k}\sum_{v\in\mathcal{U}_n}\gamma^{\mathbb{D}(u,w)}\wedge h_n\gamma^{\mathbb{D}(u,a(u,w))\vee\mathbb{D}(w,a(u,w))}\\
&\quad\lesssim(nh_n)^{-2}\sum_{k=0}^{N_n}2^k h_n|\log h_n|\lesssim|\log h_n|(nh_n)^{-1}
\end{aligned}$$

and the conclusion follows in the full case.

Putting together the sparse and full tree cases, we obtain the proposition. □

5.7. Proof of Theorem 2

From

$$\widehat{B}_n(2y) = y\frac{n^{-1}\sum_{u\in\mathcal{U}_n}K_{h_n}(\xi_u-y)}{n^{-1}\sum_{u\in\mathcal{U}_n}(1/\tau_{u^-})\mathbf{1}_{\{\xi_{u^-}\leq 2y,\xi_u\geq y\}}\vee\varpi_n}$$

and

$$B(2y) = y\frac{\nu_B(y)}{\mathbb{E}_{\nu_B}[(1/\tau_{u^-})\mathbf{1}_{\{\xi_{u^-}\le 2y, \xi_u \ge y\}}]},$$

we plan to use the following decomposition

$$\widehat{B}_n(2y) - B(2y) = y(I + II + III),$$

with

$$I = \frac{K_{h_n} \star \nu_B(y) - \nu_B(y)}{D(y)},$$
$$II = \frac{K_{h_n} \star \widehat{\nu}_n(y) - K_{h_n} \star \nu_B(y)}{D_n(y)_{\varpi_n}},$$
$$III = \frac{K_{h_n} \star \nu_B(y)}{D_n(y)_{\varpi_n} D(y)}\big(D(y) - D_n(y)_{\varpi_n}\big),$$

where $D(y)$ and $D_n(y)_\varpi$ are defined in (54) and (55) respectively. It follows that

$$\|\widehat{B}_n - B\|^2_{L^2(\mathcal{D})} = 2\int_{(1/2)\mathcal{D}} \big(\widehat{B}_n(2y) - B(2y)\big)^2 \,\mathrm{d}y \lesssim IV + V + VI,$$

with

$$IV = \int_{(1/2)\mathcal{D}} \big(K_{h_n} \star \nu_B(y) - \nu_B(y)\big)^2 \frac{y^2}{D(y)^2}\,\mathrm{d}y,$$
$$V = \int_{(1/2)\mathcal{D}} \big(K_{h_n} \star \widehat{\nu}_n(y) - K_{h_n} \star \nu_B(y)\big)^2 D_n(y)_\varpi^{-2} y^2\,\mathrm{d}y,$$
$$VI = \int_{(1/2)\mathcal{D}} \big(D_n(y)_\varpi - D(y)\big)^2 \big(K_{h_n} \star \nu_B(y)\big)^2 \big(D_n(y)_\varpi D(y)\big)^{-2} y^2\,\mathrm{d}y.$$

The term IV. We get rid of the term $\frac{y^2}{D(y)^2}$ by Lemma 4 and the fact that $\mathcal{D}$ is bounded. By Assumption 2 and classical kernel approximation, we have for every $0 < s \le n_0$

$$IV \lesssim \|K_{h_n} \star \nu_B - \nu_B\|^2_{L^2(2^{-1}\mathcal{D})} \lesssim |\nu_B|^2_{\mathcal{H}^s(2^{-1}\mathcal{D})} h_n^{2s}. \tag{62}$$

Lemma 5. *Let $\mathcal{D} \subset (0,\infty)$ be a compact interval. Let $B \in \mathcal{F}^\lambda(\mathfrak{c})$ for some $\mathfrak{c}$ satisfying Assumption 3. We have*

$$\|\nu_B\|_{\mathcal{H}^s(2^{-1}\mathcal{D})} \le \psi\big(e_{\min}, e_{\max}, \mathcal{D}, \|B\|_{\mathcal{H}^s(\mathcal{D})}\big)$$

for some continuous function ψ.

Proof. Define

$$\Lambda_B(x,y)=\int_{\mathcal{E}}\frac{\nu_B(y,\mathrm{d}v')}{v'}\exp\Big(-\int_{y/2}^{x}\frac{B(2s)}{v's}\,\mathrm{d}s\Big).$$

If $B\in\mathcal{H}^s(\mathcal{D})$, then $x\rightsquigarrow\Lambda_B(x,y)\in\mathcal{H}^s(2^{-1}\mathcal{D})$ for every $y\in[0,\infty)$, and we have

$$\big\|\Lambda_B(\cdot,y)\big\|_{\mathcal{H}^s(2^{-1}\mathcal{D})}\le\psi_1\big(y,\|B\|_{\mathcal{H}^s(\mathcal{D})},e_{\min},e_{\max}\big)$$

for some continuous function ψ_1. The result is then a consequence of the representation $\nu_B(x)=\frac{B(2x)}{x}\int_0^{2x}\Lambda_B(x,y)\,\mathrm{d}y$. □

Going back to (62) we infer from Lemma 5 that $|\nu_B|_{\mathcal{H}^s(2^{-1}\mathcal{D})}$ is bounded above by a constant that depends on $e_{\min}$, $e_{\max}$, $\mathcal{D}$ and $\|B\|_{\mathcal{H}^s(\mathcal{D})}$ only. It follows that

$$IV\lesssim h_n^{2s}\tag{63}$$

uniformly in $B\in\mathcal{H}^s(\mathcal{D},M)$.

The term V. We have

$$\mathbb{E}_\mu[V]\le\varpi_n^{-2}|\mathcal{D}|\sup_{y\in2^{-1}\mathcal{D}}y^2\mathbb{E}_\mu\big[\big(K_{h_n}\star\widehat{\nu}_n(y)-K_{h_n}\star\nu_B(y)\big)^2\big].$$

By (60) of Proposition 8, we derive

$$\mathbb{E}_\mu[V]\lesssim\varpi_n^{-2}|\log h_n|(nh_n)^{-1}\tag{64}$$

uniformly in $B\in\mathcal{F}^\lambda(\mathfrak{c})$.

The term VI. First, by Lemma 4, the estimate

$$\inf_{B\in\mathcal{F}^\lambda(\mathfrak{c})}\inf_{y\in2^{-1}\mathcal{D}}D_n(y)_\varpi D(y)\gtrsim\varpi_n$$

holds. Next,

$$\begin{aligned}\sup_{y\in2^{-1}\mathcal{D}}\big|K_{h_n}\star\nu_B(y)\big|&=\sup_{y\in2^{-1}\mathcal{D}}\Big|\int_{[0,\infty)}K_{h_n}(z-y)\nu_B(z)\,\mathrm{d}z\Big|\\&\le\sup_{y\in2^{-1}\mathcal{D}_{h_n}}\nu_B(y)\|K\|_{L^1([0,\infty))},\end{aligned}\tag{65}$$

where $2^{-1}\mathcal{D}_{h_n}=\{y+z,\,y\in2^{-1}\mathcal{D},\,z\in\mathrm{supp}(K_{h_n})\}\subset\widetilde{\mathcal{D}}$, for some compact interval $\widetilde{\mathcal{D}}$ since K has compact support by Assumption 2. By Lemma 2, we infer that (65) holds uniformly in $B\in\mathcal{F}^\lambda(\mathfrak{c})$. We derive

$$\mathbb{E}_\mu[VI]\lesssim\varpi_n^{-2}\sup_{y\in2^{-1}\mathcal{D}}\mathbb{E}_\mu\big[\big(D_n(y)_{\varpi_n}-D(y)\big)^2\big].$$

Applying (56) of Proposition 7, we conclude

$$\mathbb{E}_\mu[VI] \lesssim \varpi_n^{-2} n^{-1} \tag{66}$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c})$.

Completion of proof of Theorem 2. We put together the three estimates (63), (64) and (66). We obtain

$$\mathbb{E}_\mu\big[\|\widehat{B}_n - B\|^2_{L^2(\mathcal{D})}\big] \lesssim h_n^{2s} + \varpi_n^{-2} |\log h_n| (nh_n)^{-1} + \varpi_n^{-2} n^{-1}$$

uniformly in $B \in \mathcal{F}^\lambda(\mathfrak{c}) \cap \mathcal{H}^s(\mathcal{D}, M)$. The choice $h_n \sim n^{-1/(2s+1)}$ and the fact that ϖ_n^{-2} grows logarithmically in n yields the rate $n^{-s/(2s+1)}$ up to log terms and the inessential supplementary multiplicative error factor ϖ_n^{-1}. The proof is complete.

Appendix

A.1. Construction of the discrete model

Fix an initial condition $\mathbf{x} = (x, v) \in \mathcal{S}$. On a rich enough probability space, we consider a Markov chain on the binary tree $(\tau_u, u \in \mathcal{U})$ with transition $\rho(v, \mathrm{d}v')$ and initial condition v: if $u = (u_1, \ldots, u_k) \in \mathcal{U}$, we write $ui = (u_1, \ldots, u_k, i)$, $i = 0, 1$ for the two offsprings of u; we set $\tau_\varnothing = v$ and

$$\tau_{u0} \sim \rho\big(\tau_u, \mathrm{d}v'\big) \quad \text{and} \quad \tau_{u1} \sim \rho\big(\tau_u, \mathrm{d}v'\big)$$

so that conditional on τ_u, the two random variables τ_{u0} and τ_{u1} are independent. We also pick a sequence of independent standard exponential random variables $(\mathbf{e}_u, u \in \mathcal{U})$, independent of $(\tau_u, u \in \mathcal{U})$. The model $((\xi_u, \tau_u), u \in \mathcal{U})$ is then constructed recursively. We set

$$\xi_\varnothing = x, \qquad b_\varnothing = 0, \qquad \tau_\varnothing = v \quad \text{and} \quad \zeta_\varnothing = F^{-1}_{x,v}(\mathbf{e}_\varnothing),$$

where $F_{x,v}(t) = \int_0^t B(x \exp(vs))\,\mathrm{d}s$. For $u \in \mathcal{U}$ and $i = 0, 1$, we put

$$\xi_{u0} = \xi_{u1} = \mathrm{e}^{\tau_u \zeta_u} \frac{\xi_u}{2}, \qquad b_{u0} = b_{u1} = b_u + \zeta_u, \qquad \zeta_{ui} = F^{-1}_{\xi_{ui}, \tau_{ui}}(\mathbf{e}_{ui}).$$

To each node $u \in \mathcal{U}$, we then associate the mark $(\xi_i, b_u, \zeta_u, \tau_u)$ of the size, date of birth, lifetime and growth rate respectively of the individual labeled by u. One easily checks that Assumption 1 guarantees that the model is well defined. In particular, since B is locally bounded, we see that there is no accumulation of jumps almost-surely.

A.2. Proof of Lemma 1

Note first that

$$\{C_{t+h} - C_t \geq 1\} = \{t < b_{\vartheta_{C_t}} + \zeta_{\vartheta_{C_t}} \leq t + h\}.$$

Since moreover $\xi_{\vartheta_{C_t}} = x\exp(\overline{\mathcal{V}}(b_{\vartheta_{C_t}}))2^{-C_t}$, it follows by (7) that

$$\begin{aligned}&\mathbb{P}(C_{t+h} - C_t \geq 1|\mathcal{F}_t)\\&= \int_{t-b_{\vartheta_{C_t}}}^{t+h-b_{\vartheta_{C_t}}} B\left(\frac{x\mathrm{e}^{\overline{\mathcal{V}}(b_{\vartheta_{C_t}})+s\mathcal{V}(s)}}{2^{C_t}}\right)\exp\left(-\int_0^s B\left(\frac{x\mathrm{e}^{\overline{\mathcal{V}}(b_{\vartheta_{C_t}})+s'\mathcal{V}(s')}}{2^{C_t}}\right)\mathrm{d}s'\right)\mathrm{d}s.\end{aligned}$$

Introduce the quantity $B(x\mathrm{e}^{\overline{\mathcal{V}}(b_{\vartheta_{C_t}})+\mathcal{V}(t)(t-b_{\vartheta_{C_t}})}2^{-C_t})$ within the integral. Noting that $\overline{\mathcal{V}}(b_{\vartheta_{C_t}}) + \mathcal{V}(t)(t - b_{\vartheta_{C_t}}) = \overline{\mathcal{V}}(t)$ we obtain the first part of the lemma thanks to the representation (27) and the uniform continuity of B over compact sets. For the second part, introduce the $(\mathcal{F}_t)$-stopping time

$$\Upsilon_t = \inf\{s > t, C_s - C_t \geq 1\}$$

and note that $\{C_{t+h} - C_t \geq 1\} = \{\Upsilon_t \leq t + h\} \in \mathcal{F}_{\Upsilon_t}$. Writing

$$\{C_{t+h} - C_t \geq 2\} = \{\Upsilon_t < t + h, \Upsilon_{\Upsilon_t} \leq t + h\}$$

and conditioning with respect to $\mathcal{F}_{\Upsilon_t}$, we first have

$$\begin{aligned}&\mathbb{P}(C_{t+h} - C_t \geq 2)\\&= \mathbb{E}\left[\int_t^{t+h-\Upsilon_t} B\left(\frac{x\mathrm{e}^{\overline{\mathcal{V}}(b_{\vartheta_{C_t}})+s\mathcal{V}(s)}}{2^{C_t}}\right)\mathrm{e}^{-\int_0^s B(x\mathrm{e}^{\overline{\mathcal{V}}(b_{\vartheta_{C_t}})+s'\mathcal{V}(s')}/2^{C_t})\,\mathrm{d}s'}\,\mathrm{d}s\mathbf{1}_{\{\Upsilon_t<t+h\}}\right]\\&\leq h\sup_{y\leq x\exp(2e_{\max}t)} B(y)\mathbb{P}(\Upsilon_t < t + h).\end{aligned}$$

In the same way, $\mathbb{P}(\Upsilon_t < t + h) \lesssim h$ and the conclusion follows.

Acknowledgements

The research of M. Doumic is partly supported by the Agence Nationale de la Recherche, Grant No. ANR-09-BLAN-0218 TOPPAZ. The research of M. Hoffmann is partly supported by the Agence Nationale de la Recherche (Blanc SIMI 1 2011 project CALIBRATION). We are grateful to V. Bansaye, A. Dalalyan, A. Marguet, E. Moulines, A. Olivier, B. Perthame, V. Rivoirard and V. C. Tran for helpful discussion and comments.

References

[1] Baccelli, F., McDonald, D.R. and Reynier, J. (2002). A mean-field model for multiple TCP connections through a buffer implementing RED. *Performance Evaluation* **49** 77–97.

[2] Balagué, D., Canizo, J. and Gabriel, P. (2013). Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates. *Kinetic and Related Models* **6** 219–243.

[3] Banks, H.T., Sutton, K.L., Thompson, W.C., Bocharov, G., Roosec, D., Schenkeld, T. and Meyerhanse, A. (2011). Estimation of cell proliferation dynamics using CFSE data. *Bull. Math. Biol.* **73** 116–150.

[4] Bansaye, V. (2008). Proliferating parasites in dividing cells: Kimmel's branching model revisited. *Ann. Appl. Probab.* **18** 967–996. MR2418235

[5] Bansaye, V., Delmas, J.-F., Marsalle, L. and Tran, V.C. (2011). Limit theorems for Markov processes indexed by continuous time Galton–Watson trees. *Ann. Appl. Probab.* **21** 2263–2314. MR2895416

[6] Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[7] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge: Cambridge Univ. Press. MR2253162

[8] Cáceres, M.J., Cañizo, J.A. and Mischler, S. (2011). Rate of convergence to an asymptotic profile for the self-similar fragmentation and growth-fragmentation equations. *J. Math. Pures Appl.* (9) **96** 334–362. MR2832638

[9] Chauvin, B., Rouault, A. and Wakolbinger, A. (1991). Growing conditioned trees. *Stochastic Process. Appl.* **39** 117–130. MR1135089

[10] Cloez, B. (2011). Limit theorems for some branching measure-valued processes. Available at arXiv:1106.0660v2.

[11] Douc, R., Moulines, E. and Rosenthal, J.S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.* **14** 1643–1665. MR2099647

[12] Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2012). Nonparametric estimation of the division rate of a size-structured population. *SIAM J. Numer. Anal.* **50** 925–950.

[13] Doumic, M., Maia, P. and Zubelli, J.P. (2010). On the calibration of a size-structured population model from experimental data. *Acta Biotheor.* **58** 405–413.

[14] Doumic, M., Perthame, B. and Zubelli, J.P. (2009). Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems* **25** 1–22. MR2482159

[15] Doumic, M. and Tine, L.M. (2012). Estimating the division rate for the growth-fragmentation equation. *J. Math. Biol.* **67** 69–103.

[16] Doumic, M. and Gabriel, P. (2010). Eigenelements of a general aggregation–fragmentation model. *Math. Models Methods Appl. Sci.* **20** 757–783. MR2652618

[17] Engler, H., Prüss, J. and Webb, G.F. (2006). Analysis of a model for the dynamics of prions. II. *J. Math. Anal. Appl.* **324** 98–117. MR2262459

[18] Fort, G., Moulines, E. and Priouret, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.* **39** 3262–3289. MR3012408

[19] Gobet, E., Hoffmann, M. and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509

[20] Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214

[21] Haas, B. (2003). Loss of mass in deterministic and random fragmentations. *Stochastic Process. Appl.* **106** 245–277. MR1989629

[22] Harris, S.C. and Roberts, M.I. (2012). The many-to-few lemma and multiple spines. Available at arXiv:1106.4761v3.

[23] Kaern, M., Elston, T.C., Blake, W.J. and Collins, J.J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6** 451–464.

[24] Kubitschek, H.E. (1969). Growth during the bacterial cell cycle: Analysis of cell size distribution. *Biophys. J.* **9** 792–809.

[25] Laurençot, P. and Perthame, B. (2009). Exponential decay for the growth-fragmentation/cell-division equation. *Commun. Math. Sci.* **7** 503–510. MR2536450

[26] Metz, J.A.J. and Diekmann, O., eds. (1986). *The Dynamics of Physiologically Structured Populations*. *Lecture Notes in Biomathematics* **68**. Berlin: Springer. Papers from the colloquium held in Amsterdam, 1983. MR0860959
[27] Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*. Berlin: Springer.
[28] Michel, P. (2006). Existence of a solution to the cell division eigenproblem. *Math. Models Methods Appl. Sci*. **16** 1125–1153. MR2250122
[29] Michel, P., Mischler, S. and Perthame, B. (2005). General relative entropy inequality: An illustration on growth models. *J. Math. Pures Appl*. (9) **84** 1235–1260. MR2162224
[30] Niethammer, B. and Pego, R.L. (1999). Non-self-similar behavior in the LSW theory of Ostwald ripening. *J. Stat. Phys*. **95** 867–902. MR1712441
[31] Pakdaman, K., Perthame, B. and Salort, D. (2012). Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation. Available at http://hal.upmc.fr/hal-00794841.
[32] Perthame, B. (2007). *Transport Equations in Biology*. *Frontiers in Mathematics*. Basel: Birkhäuser.
[33] Perthame, B. and Ryzhik, L. (2005). Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations* **210** 155–177. MR2114128
[34] Perthame, B. and Zubelli, J.P. (2007). On the inverse problem for a size-structured population model. *Inverse Problems* **23** 1037–1052. MR2329931
[35] Stewart, E.J., Madden, R., Paul, G. and Taddei, F. (2005). Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Comput. Biol*. **3** e45.
[36] Sturm, A., Heinemann, M., Arnoldini, M., Benecke, A., Ackermann, M., Benz, M., Dormann, J. and Hardt, W.-D. (2011). The cost of virulence: Retarded growth of *Salmonella typhimurium* cells expressing type III secretion system 1. *PLoS Pathog*. **7** e1002143.
[37] Tan, C., Marguet, P. and You, L. (2009). Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol*. **5** 842–848.
[38] Wang, P., Robert, L., Pelletier, J., Dang, W.L., Taddei, F., Wright, A. and Jun, S. (2010). Robust growth of *Escherichia coli*. *Curr. Biol*. **20** 1099–1103.

Received September 2013 and revised March 2014