Statistical estimation in a randomly structured branching population

Marc Hoffmann, Aline Marguet

STATISTICAL ESTIMATION IN A RANDOMLY STRUCTURED BRANCHING POPULATION

MARC HOFFMANN AND ALINE MARGUET

Abstract. We consider a binary branching process structured by a stochastic trait that evolves according to a diffusion process that triggers the branching events, in the spirit of Kimmel's model of cell division with parasite infection. Based on the observation of the trait at birth of the first n generations of the process, we construct nonparametric estimator of the transition of the associated bifurcating chain and study the parametric estimation of the branching rate. In the limit $n \to \infty$, we obtain asymptotic efficiency in the parametric case and minimax optimality in the nonparametric case.

Mathematics Subject Classification (2010): 62G05, 62M05, 60J80, 60J20, 92D25.
Keywords: Branching processes, bifurcating Markov chains, statistical estimation, geometric ergodicity, scalar diffusions.

1. Introduction

1.1. **Motivation.** The study of structured populations, with a strong input from evolutionary or cell division modelling in mathematical biology (see for instance the textbooks [27, 30] and the references therein) has driven the statistics of branching Markov processes over the last few years. Several models have been considered, with data processed either in discrete or continuous time. In this context, one typically addresses the inference of critical parameters like branching rates, modelled as functions of biological traits like age, size and so on. In many cases, this approach is linked to certain piecewise deterministic Markov models or bifurcating Markov chains (BMC) in discrete time. These models are well understood from a probabilist point of view (in discrete time Guyon [17], Bitseki-Penda *et al.* [6, 7], in continuous time Bansaye and Méléard [4], Bansaye *et al.* [3] or more recently Marguet [26] for a general approach). For the statistical estimation, we refer to [8, 14, 15, 21], and the references therein, see also Bitseki-Penda and Olivier [29], de Saporta *et al.* [12, 13], Azaïs *et al.* [1] or recently Bitseki-Penda and Roche [5]. In these models, the traits of a population between branching events like cell division evolve through time according to a dynamical system. The next logical step is to replace this deterministic evolution by a random flow, that allows one to account for traits that may have their own random evolution according to some exogeneous input. A paradigmatic example is Kimmel's model (see Kimmel [22] and Bansaye [2]) where the trait is given by a density of parasites within a cell that evolve according to a diffusion process. The statistical analysis of such models is the topic of the present paper.

We consider a population model with binary division triggered by a trait $x \in \mathcal{X}$ where $\mathcal{X} \subseteq \mathbb{R}$ is an open (possibly unbounded) interval. The trait $\phi_x(t)$ of each individual evolves according to

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x, \tag{1}$$

where $r, \sigma : \mathcal{X} \to \mathcal{X}$ are regular functions and $(W_t)_{t\geq 0}$ is a standard Brownian motion. Each individual with trait x dies according to a killing or rather division rate $x \mapsto B(x)$. (An individual

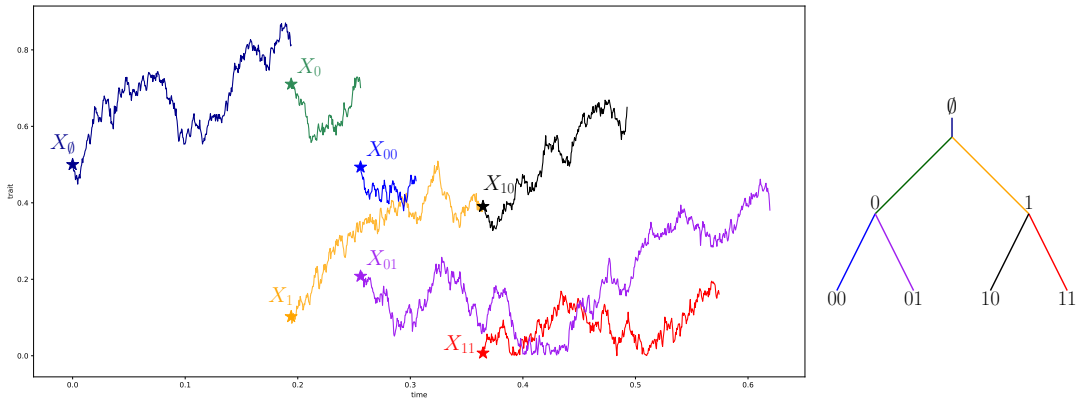

FIGURE 1. Example of a trajectory and its associated genealogy.

with trait $\phi_x(t)$ at time t dies with probability $B(\phi_x(t))dt$ during the interval $[t, t+dt]$.) At division, a particle is replaced by two new individuals with trait at birth given respectively by θx and $(1-\theta)x$ where θ is drawn according to $\kappa(x)dx$ for some symmetric probability density function $\kappa(x)$ on $[0,1]$. The model is described by the traits of the population, formally given as a Markov process

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

with values in $\bigcup_{k\geq 1} \mathcal{X}^k$, where the $X_i(t)$ denote the (ordered) traits of the living particles at time t. Its distribution is entirely determined by an initial condition at $t = 0$ and by the parameters (r, σ, B, κ).

1.2. **Statistical setting by reduction to a bifurcating Markov chain model.** We assume we have data at branching events (*i.e.* at cell division) and we wish to make inference on the parameters of the model. Using the Ulam-Harris-Neveu notation, for $m \geq 0$, let $\mathbb{G}_m = \{0,1\}^m$ (with $\mathbb{G}_0 = \{\emptyset\}$) and introduce the infinite genealogical tree

$$\mathbb{T} = \bigcup_{m\in\mathbb{N}} \mathbb{G}_m.$$

For $u \in \mathbb{G}_m$, set $|u| = m$ and define the concatenation $u0 = (u, 0) \in \mathbb{G}_{m+1}$ and $u1 = (u, 1) \in \mathbb{G}_{m+1}$. For $n \geq 0$, let $\mathbb{T}_n = \bigcup_{m=0}^n \mathbb{G}_m$ denote the genealogical tree up to the n-th generation. We denote by X_u the trait at birth of an individual $u \in \mathbb{T}$. From the branching events, we assume that we observe

$$\mathbb{X}^n = (X_u)_{u\in\mathbb{T}_n},$$

so that we have $2^{n+1} - 1$ random variables with value in $\mathcal{X}$. Asymptotics are taken as n grows to infinity. An example of trajectory is represented on Figure 1 with the associated genealogy.

There are several objects of interest that we may try to infer from the data $\mathbb{X}^n$. First, one may notice that the Markov structure of X in (2) turns $(X_u, u \in \mathbb{T})$ into a *bifurcating Markov chain* according to the terminology introduced by Delmas and Guyon [17]. A bifurcating Markov chain is specified by **1)** a measurable state space, here $\mathcal{X}$ (endowed with its Borel sigma-field) with a

Markov kernel $\mathcal{P}$ from $\mathcal{X}$ to $\mathcal{X}\times\mathcal{X}$ and **2)** a filtered probability space $\big(\Omega,\mathcal{F},(\mathcal{F}_m)_{m\geq 0},\mathbb{P}\big)$. Following Guyon, [17], Definition 2, we have the

Definition 1. *A bifurcating Markov chain (BMC) is a family $(X_u)_{u\in\mathbb{T}}$ of random variables with value in $\mathcal{X}$ such that X_u is $\mathcal{F}_{|u|}$-measurable for every $u\in\mathbb{T}$ and*

$$\mathbb{E}\big[\prod_{u\in\mathbb{G}_m}\psi_u(X_u,X_{u0},X_{u1})\,\big|\,\mathcal{F}_m\big]=\prod_{u\in\mathbb{G}_m}\mathcal{P}\psi_u(X_u) \tag{3}$$

for every $m\geq 0$ and any family of (bounded) measurable functions $(\psi_u)_{u\in\mathbb{G}_m}$, where $\mathcal{P}\psi(x)=\int_{\mathcal{X}\times\mathcal{X}}\psi(x,y_1,y_2)\mathcal{P}(x,dy_1dy_2)$ denotes the action of $\mathcal{P}$ on ψ.

The distribution of $(X_u)_{u\in\mathbb{T}}$ is thus entirely determined by $\mathcal{P}$ and an initial distribution for $X_\emptyset$. A key role for understanding the asymptotic behavior of the bifurcating Markov chain is the so-called *tagged-branch* chain, that consists in picking a lineage at random in the population $(X_u)_{u\in\mathbb{T}}$: it is a Markov chain with value in $\mathcal{X}$ defined by $Y_0=X_\emptyset$ and for $m\geq 1$:

$$Y_m=X_{\emptyset\epsilon_1\ldots\epsilon_m},$$

where $(\epsilon_m)_{m\geq 1}$ is a sequence of independent Bernoulli random variables with parameter $1/2$, independent of $(X_u)_{u\in\mathbb{T}}$, with transition

$$\mathcal{Q}=(\mathcal{P}_1+\mathcal{P}_2)/2 \tag{4}$$

obtained from the marginal transitions of $\mathcal{P}$:

$$\mathcal{P}_1(x,dy)=\int_{y_2\in\mathcal{X}}\mathcal{P}(x,dy\,dy_2)\ \text{ and }\ \mathcal{P}_2(x,dy)=\int_{y_1\in\mathcal{X}}\mathcal{P}(x,dy_1dy).$$

Guyon proves in [17] that if $(Y_m)_{m\geq 0}$ is ergodic with invariant measure $\nu(dx)$ on $\mathcal{X}$, then a convergence of the type

$$\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}\psi(X_u,X_{u0},X_{u1})\to\int_{\mathcal{X}}\mathcal{P}\psi(x)\nu(dx) \tag{5}$$

holds as $n\to\infty$ for appropriate test functions g, almost surely and appended with appropriate central limit theorems (Theorem 19 in [17]).

1.3. **Main results.** In this context, there are several quantities that can be inferred from the data $\mathbb{X}_n$ as n grows and that are important in order to understand the dynamics of $(X_u)_{u\in\mathbb{T}}$. Under suitable assumptions on the stochastic flow (1), the transition $\mathcal{Q}$ admits an invariant measure ν and we have fast convergence of the tagged-chain $(Y_m)_{m\geq 1}$ to equilibrium. This enables us to construct in a first part nonparametric estimators of ν and $\mathcal{Q}$ with an optimal rate of convergence. Notice that since the fragmentation kernel $\kappa(x)$ that distributes the trait of the parent between its two children is symmetric, we have $\mathcal{P}_1=\mathcal{P}_2=\mathcal{Q}$. This reveals the structure of the underlying BMC.

However, estimators of ν and $\mathcal{Q}$ do not give us any insight about the parameters (r,σ,B,κ) of the model. In a second part, we investigate the inference of the division rate $x\mapsto B(x)$ as a function of the trait $x\in\mathcal{X}$ when the other parameters r,σ and $\kappa(x)$ are known. This seemingly stringent assumption is necessary given the observation scheme $\mathbb{X}^n$. If extraneous data were available, estimators of the parameters r,σ and κ could be obtained in a relatively straightforward manner:

i) As soon as a discretisation of the values of the flow are available, standard techniques about inference in ergodic diffusions can be applied to recover $x\mapsto r(x)$ and $x\mapsto\sigma(x)^2$, see for instance [20, 23].

ii) The fact that an individual u distributed its traits to its offspring in a conservative way enables one to recover the fraction θ_u distributed among the children. Indeed the individual u born at b_u with lifespan $d_u - b_u$ has trait $\phi_{X_u}(d_u - b_u)$ at its time of death. It follows that its children have trait at birth given by

$$X_{u0} = \theta_u \phi_{X_u}(d_u - b_u), \quad X_{u1} = (1-\theta_u)\phi_{X_u}(d_u - b_u),$$

where the θ_u are drawn independently from the distribution $\kappa(x)dx$ and therefore, the relationship $\frac{X_{u0}}{X_{u1}} = \frac{\theta_u}{1-\theta_u}$ identifies θ_u. In turn, the estimation of $x \mapsto \kappa(x)$ reduces to a standard density estimation problem from data $(\theta_u)_{u\in\mathbb{T}_n}$, see for instance [19].

The identification and estimation of the branching rate $x \mapsto B(x)$ from data $\mathbb{X}^n$ is more delicate and is the topic of the second part of the paper. Under minimal regularity assumptions developed in Section 2 below, it is not difficult to obtain an explicit representation of the transition $\mathcal{Q}(x, dy) = \mathcal{Q}_B(x, dy) = q_B(x, y)dy$ that reads

$$q_B(x,y) = \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]dz, \tag{6}$$

where $L_t^y(\phi_x)$ denotes the local time at t in y of the semimartingale $(\phi_x(t))_{t\geq 0}$. Assuming (r, σ, κ) known (or identified by extraneous observation schemes) we study the estimation of $x \mapsto B(x)$ when B belongs to a parametric class of functions $\{B_\vartheta, \vartheta \in \Theta\}$ for some regular subset of the Euclidean space $\mathbb{R}^d$. Under a certain ordering property (Definition 13 in Section 3.2 below) that ensures identifiability of the model and suitable standard regularity properties, we realise a standard maximum likelihood proxy estimation of B thanks to (6) by maximising the contrast

$$\vartheta \mapsto \prod_{u\in\mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u), \quad \vartheta \in \Theta,$$

(with $\mathbb{T}_n^\star = \mathbb{T}_n \setminus \mathbb{G}_0$) and we prove that it achieves asymptotic efficiency and discuss its practical implementation. It is noteworthy that for the parametric estimation of B, there is no straightforward contrast minimisation procedure (at least we could not find any) whereas $q_B(x, y)$ is explicit. The fairly intricate dependence of B in the representation (6) makes however the whole scheme relatively delicate, both mathematically and numerically.

Clearly, other observation schemes are relevant in the context of cell division modelling. For instance, one could consider a (large) time $T > 0$ and observe the branching process X_t defined in (2) for every $t \in [0, T]$. This entails the possibility to extract the times (T_u) at which branching events occur, like *e.g.* in [21]. However, the continuous time setting is drastically different and introduce the additional difficulty of bias sampling, an issue we avoid in the present context. Alternatively, one could consider the augmented statistical experiment where one observes $(X_u, T_u)_{u\in\mathbb{T}_n}$, but the underlying mathematical structure is presumably not simpler. Our results show in particular that for the parametric estimation of the branching rate B, although the times at which branching event occur are statistically informative, their observation is not necessary to obtain optimal rates of convergence as soon as (r, σ, κ) are known.

1.4. **Organisation of the paper.** Section 2.1 is devoted to the construction of the stochastic model, our assumptions and the accompanying statistical experiments. In particular, we have a nice structure enough so that explicit representations of $\mathcal{P}$ and $\mathcal{Q}$ are available (Proposition 5). We give a first result on the geometric ergodicity of the model via an explicit Lyapunov function in Theorem 6 and derive in Theorem 7 a rate of convergence for the variance of empirical measures of the data $\mathbb{X}^n = (X_u)_{u\in\mathbb{T}_n}$ against test functions $\varphi(X_u)$ or $\psi(X_{u^-}, X_u)$ with a sharp control in

terms of adequate norms for φ, ψ that do not follow from the standard application of the geometric ergodicity of Theorem 6. This is crucial for the subsequent applications to the nonparametric estimation of $\mathcal{Q}$ and its invariant measure ν that are given in Theorem 10 of Section 3.1. Section 3.2 is devoted to the parametric estimation of the branching rate, where an asymptotically efficient result is proved for a maximum likelihood estimator in Theorem 20. It is based on a relatively sharp study of the transition $\mathcal{Q}$, thanks to local time properties of the stochastic flow that triggers the branching events. Section 4 is devoted to the numerical implementation of the parametric estimator of B. In particular, in order to avoid the computational cost of the explicit computation of $q_\vartheta(X_{u^-}, X_u)$, we take advantage of our preceding results and implement a nonparametric estimator on Monte-Carlo simulations instead, resulting in a feasible procedure for practical purposes. The proofs are postponed to Section 5 and an Appendix Section 6 contains useful auxiliary results.

2. A cell division model structured by a stochastic flow

2.1. Assumptions and well-posedness of the stochastic model.

Dynamics of the traits. Remember that $\mathcal{X} \subseteq \mathbb{R}$ is an open, possibly unbounded interval. The flow is specified by $r, \sigma : \mathcal{X} \to \mathcal{X}$ which are measurable and that satisfy the following assumption:

Assumption 2. *For some $r_1, \sigma_1, \sigma_2 > 0$, we have $|r(x)| \leq r_1(1 + |x|)$ and $\sigma_1 \leq \sigma(x) \leq \sigma_2$, for every $x \in \mathcal{X}$. Moreover, for some $r_2 > 0$, we have $\mathrm{sgn}(x)r(x) < 0$ for $|x| \geq r_2$ (with $\mathrm{sgn}(x) = \mathbf{1}_{\{x>0\}} - \mathbf{1}_{\{x\leq 0\}}$).*

Under Assumption 2, there is a unique strong solution to (1) (for instance [28], Theorem 5.2.1.). We denote by $(\Phi_x(t), t \geq 0)$ the unique solution to (1) with initial condition $x \in \mathcal{X}$. In particular, $(\Phi_x(t), t \geq 0)$ is a strong Markov process and is ergodic (cf. [23], Theorem 1.16.). Note that when $\mathcal{X}$ is bounded, the drift condition $\mathrm{sgn}(x)r(x) < 0$ for large enough x can be dropped.

Division events. An individual with trait x dies at an instantaneous rate $x \mapsto B(x)$, where $B : \mathcal{X} \to [0, \infty)$ satisfies the following condition:

Assumption 3. *The function $x \mapsto B(x)$ is continuous. Moreover, for some $b_1, b_2 > 0$ and $\gamma \geq 0$, we have $b_1 \leq B(x) \leq b_2 |x|^\gamma + b_1$ for every $x \in \mathcal{X}$.*

Under Assumptions 2 and 3, the process X in (2) is well defined and the size of the population does not explode in finite time almost-surely, see for instance Marguet [26]. Note that the lower bounds for σ and B are not needed for the well-posedness of X but rather for later statistical purposes.

Fragmentation of the trait at division. Finally, we make an additional set of assumptions on the fragmentation distribution $\kappa(z)dz$ that ensures in particular the non-degeneracy of the process.

Assumption 4. *We have*

$$\kappa(z) = \kappa(1 - z) \text{ for almost every } z \in [0, 1],$$
$$\mathrm{supp}(\kappa) \subset [\varepsilon, 1 - \varepsilon] \text{ for some } 0 < \varepsilon < 1/2,$$
$$\inf_{z \in [\varepsilon, 1-\varepsilon]} \kappa(z) \geq \delta.$$

The symmetry condition of Assumption 4 on the fragmentation distribution $\kappa(z)dz$ ensures that the law of the trait at birth is identical for both children. The other parts of Assumption 4 are slightly technical and may presumably be relaxed.

Representations of $\mathcal{P}$ *and* $\mathcal{Q}$. Under Assumptions 2, 3 and 4, we obtain closed-form formulae for the transition $\mathcal{P}$ defined via (3) and the mean or marginal transition $\mathcal{Q}$ of the BMC $(X_u)_{u\in\mathbb{T}}$, see (4) that also gives the transition probability of the discrete Markov chain with value in $\mathcal{X}$ corresponding to the trait at birth along an ancestral lineage. These representations are crucial for the subsequent analysis of the variance of the estimators of $\mathcal{P}$ and of the invariant measure ν.

Proposition 5. *Work under Assumptions 2, 3 and 4. For every* $x, y, y_1, y_2 \in \mathcal{X}$, *we have*

$$\mathcal{P}(x, dy_1 dy_2) = p(x, y_1, y_2)dy_1 dy_2 \quad \text{and} \quad \mathcal{Q}(x, dy) = q(x, y)dy,$$

with

$$p(x, y_1, y_2) = \frac{\kappa\big(y_1/(y_1+y_2)\big)}{y_1+y_2} B\,(y_1+y_2)\,\sigma(y_1+y_2)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y_1+y_2}(\phi_x)\Big] \tag{7}$$

and

$$q(x, y) = \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]dz, \tag{8}$$

where $L_t^y(\phi_x)$ *denotes the local time at* t *in* y *of the semimartingale* $(\phi_x(t))_{t\geq 0}$.

Note that by Assumption 4, we have $p(x, y_1, y_2) = p(x, y_2, y_1)$ for every $x, y_1, y_2 \in \mathcal{X}$ and also $\int_{\mathcal{X}} p(x, y_1, y)dy_1 = \int_{\mathcal{X}} p(x, y, y_2)dy_2 = q(x, y)$ for every $x, y \in \mathcal{X}$.

2.2. **Convergence of empirical measures.**

Convergence of $\mathcal{Q}$ *to equilibrium.* Assumptions 2,3 and 4 imply a drift condition for the Lyapunov function $V(x) = x^2$ on $\mathcal{X}$ and a minorisation condition over a small set so that in turn $\mathcal{Q}$ is geometrically ergodic.

Let $\mathbb{Q} = \mathbb{Q}(r_i, b_i, \sigma_i, z_i, \gamma, \varepsilon, \delta, i = 1, 2)$ be the class of all transitions $\mathcal{Q} = \mathcal{Q}(r, \sigma, B, \kappa)$ defined over $\mathcal{X}$ that satisfy Assumptions 2,3 and 4 with appropriate constants. An invariant probability measure for $\mathcal{Q}$ is a probability ν on $\mathcal{X}$ such that $\nu\mathcal{Q} = \nu$, where $\nu\mathcal{Q}(dy) = \int_{x\in\mathcal{X}} \nu(dx)\mathcal{Q}(x, dy)$. Define

$$\mathcal{Q}^r(x, dy) = \int Q(x, dz)Q^{r-1}(z, dy) \quad \text{with} \quad \mathcal{Q}^0(x, dy) = \delta_x(dy)$$

for the r-th iteration of $\mathcal{Q}$. For $\varphi : \mathcal{X} \to \mathbb{R}$, we set

$$|\varphi|_V = \sup_{x\in\mathcal{X}} \frac{|\varphi(x)|}{1+V(x)}$$

and write $\nu(\varphi) = \int_{\mathcal{X}} \varphi(x)\nu(dx)$ when no confusion is possible.

Theorem 6 (Convergence to equilibrium)**.** *Work under Assumptions 2, 3 and 4. Then any* $\mathcal{Q} \in \mathbb{Q}$ *admits an invariant probability distribution* ν. *Moreover, for* $V(x) = x^2$, *there exist* $C = C(\mathcal{Q}) > 0$ *and* $\rho = \rho(\mathcal{Q}) \in (0, 1)$ *such that for every* $m \geq 1$, *the bound*

$$\big|\mathcal{Q}^m\varphi - \nu(\varphi)\big|_V \leq C\rho^m\big|\varphi - \nu(\varphi)\big|_V$$

holds as soon as $|\varphi|_V < \infty$. *Moreover,* $\sup_{\mathcal{Q}\in\mathbb{Q}} C(\mathcal{Q}) < \infty$ *and* $\sup_{\mathcal{Q}\in\mathbb{Q}} \rho(\mathcal{Q}) < 1$.

In particular, we have $|\mathcal{Q}^m\varphi(x) - \nu(\varphi)| \leq C\rho^m(1+V(x))|\varphi - \nu(\varphi)|_\infty$ for every $x \in \mathcal{X}$.

Control of empirical variances. Theorem 6 is the key ingredient in order to control the rate of convergence of empirical means of the form

$$\mathcal{M}_n(\psi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \psi(X_{u^-}, X_u) \tag{9}$$

that converge to $\nu Q(\psi)$ for any reasonable test function $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. (If $\varphi : \mathcal{X} \to \mathbb{R}$ we set $\mathcal{M}_n(\varphi) = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \varphi(X_u) \to \nu(\varphi)$ and we have a formal correspondence between the two expressions by writing $\psi(x, y) = \varphi(y)$ as a function of the second variable.) In order to derive nonparametric estimators of ν and Q by means of kernel functions ψ that shall depend on n, we need sharp estimates in terms of ψ.

We shall restrict our study to transitions $Q \in \mathbb{Q}$ for which the geometric rate of convergence to equilibrium $\rho = \rho(Q)$ given in Theorem 6 satisfies $\rho(Q) \leq 1/2$. Let $\mathbb{Q}_{1/2} \subset \mathbb{Q}$ denote the set of such transitions.

We need some notation. We denote by $|\cdot|_1$ the usual L^1-norm w.r.t. the Lebesgue measure on $\mathcal{X} \times \mathcal{X}$. For a function $\psi = \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ we set $\psi^\star(x) = \sup_{y \in \mathcal{X}} |\psi(x, y)|$ and $\psi_\star(y) = \sup_{x \in \mathcal{X}} |\psi(x, y)|$ and define

$$|\psi|_{\wedge 1} = \int_{\mathcal{X} \times \mathcal{X}} |\psi(x, y)| dx dy \bigwedge \int_{\mathcal{X}} \sup_{x \in \mathcal{X}} |\psi(x, y)| dy.$$

Note in particular that when $\psi(x, y) = \varphi(y)$ is a function of y only, we may have that $|\psi_\star|_1 = \int_{\mathcal{X}} |\varphi(y)| dy$ is finite while ψ is not integrable on $\mathcal{X} \times \mathcal{X}$ as a function of two variables. For a positive measure ρ on $\mathcal{X}$, let also

$$|\psi|_\rho = \int_{\mathcal{X} \times \mathcal{X}} |\psi(x, y)| \rho(dx) dy + |\psi|_{\wedge 1}.$$

We write $\mathbb{P}_\mu$ for the law of $(X_u)_{u \in \mathbb{T}}$ with initial distribution μ for $X_\emptyset$. Remember that $V(x) = x^2$ from Theorem 6.

Theorem 7. *Work under Assumptions 2, 3 and 4. Let μ be a probability measure on $\mathcal{X}$ such that $\mu(V^2) < \infty$. Let $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ a bounded function such that $\psi_\star$ is compactly supported. The following estimate holds true:*

$$\mathbb{E}_\mu\big[\big(\mathcal{M}_n(\psi) - \nu(Q\psi)\big)^2\big] \lesssim |\mathbb{T}_n|^{-1} \big(|\psi^2|_{\mu+\nu} + |\psi^\star \psi|_\mu + \big(1 + \mu(V^2)^{1/2}\big) |\psi_\star|_1 |\psi|_\nu\big),$$

where the symbol $\lesssim$ means up to an explicitly computable constant that depends on Q and on $supp(\psi_\star)$ only. Moreover, the estimate is uniform in $Q \in \mathbb{Q}_{1/2}$.

Several remarks are in order: **1)** We have a sharp order in terms of the test functions ψ, that behave no worse than $\int_{\mathcal{X}^2} \psi^2$ under minimal regularity on ν which is satisfied, see Lemma 25 below (and of course μ, although this restriction could easily be relaxed). This behaviour is the one expected for instance in the IID case and is crucial for the subsequent statistical application of Theorem 10 where the functions ψ will be kernel depending on n. **2)** Theorem 7 can be derived with other more competitive tools in the uniform geometric ergodic case, see *e.g.* [8, 5], a case we do not recover here since $\mathcal{X}$ is not necessarily bounded. **3)** Theorem 7 has an analog in [15] for piecewise deterministic growth-fragmentation models, but our proof is somewhat simpler here and sharper (we do not pay the superfluous logarithmic term in [15]).

3. Statistical estimation

3.1. **Nonparametric estimation of $\mathcal{Q}$ and ν.** Under Assumptions 2, 3 and 4, any $\mathcal{Q}(x,dy) = q(x,y)dy$ admits an invariant probability measure $\nu(dx) = \nu(x)dx$, the regularity of $\nu(x)$ being inherited from that of $\mathcal{Q}$ via $\nu(x) = \int_{\mathcal{X}} q(z,x)\nu(dz)$.

Fix $(x_0,y_0) \in \mathcal{X}\times\mathcal{X}$. We are interested in constructing estimators of $q(x_0,y_0)$ and $\nu(x_0)$ from the observation $\mathbb{X}^n$ when both functions satisfy some Hölder regularity properties in the vicinity of (x_0,y_0). To that end, we need approximating kernels.

Definition 8. *A function $G : \mathcal{X} \to \mathbb{R}$ is a kernel of order k if it is compactly supported and satisfies $\int_{\mathcal{X}} x^\ell G(x)dx = \mathbf{1}_{\{\ell=0\}}$ for $\ell = 0,\dots,k$.*

The construction and numerical tractability of approximating kernels is documented in numerous textbooks, see for instance Tsybakov [33]. For bandwidth parameters $h, h_1, h_2 > 0$, we set

$$G_h(y) = h^{-1}G(h^{-1}y)$$

and

$$G^{\otimes 2}_{h_1,h_2}(x,y) = h_1^{-1}h_2^{-1}G(h_1^{-1}x)G(h_1^{-1}y)$$

and obtain approximations of $\nu(y_0)$ and $q(x_0,y_0)$ by setting

$$G_h \star \nu(y_0) = \int_{\mathcal{X}} G_h(y-y_0)\nu(y)dy$$

and

$$G_h \star q(x_0,y_0) = \int_{\mathcal{X}\times\mathcal{X}} G^{\otimes 2}_{h_1,h_2}(x-x_0,y-y_0)\nu(x)q(x,y)dxdy.$$

The convergence of $\mathcal{M}_n(\varphi)$ to $\nu(\varphi)$ suggests to pick $\varphi = G_h(\cdot - x_0)$ that converges to $G_h \star \nu(x_0)$ and use it as a proxy of $\nu(x_0)$. We obtain the estimator

$$\widehat{\nu}_n(x_0) = \mathcal{M}_n\big(G_h(\cdot - x_0)\big),$$

specified by the choice of $h > 0$ and the kernel G. Likewise, an estimator of $q(x_0,y_0)$ is obtained by considering the quotient estimator with numerator $\psi = G_{h_1,h_2}(\cdot - x_0, \cdot - y_0)$ that converges to $G_h \star \nu(x_0)q(x_0,y_0)$ and denominator $\widehat{\nu}_{n,h}(x_0)$ in order to balance the superfluous weight $\nu(x_0)$ in the numerator. We obtain the estimator

$$\widehat{q}_n(x_0,y_0) = \frac{\mathcal{M}_n\big(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\big)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi},$$

specified by the choice of $h, h_1, h_2 > 0$, a threshold $\varpi > 0$ and the kernel G. In order to quantify the kernel approximation, we introduce anisotropic Hölder classes. For $\alpha > 0$, we write $\alpha = \lfloor\alpha\rfloor + \{\alpha\}$ with $\lfloor\alpha\rfloor$ an integer and $0 < \{\alpha\} \le 1$.

Definition 9. *Let $\alpha, \beta > 0$ and $\mathcal{V}_{x_0}$ and $\mathcal{V}(y_0)$ be bounded neighbourhoods of x_0 and y_0.*

i) *The function $\varphi : \mathcal{V}_{x_0} \to \mathbb{R}$ belongs to the Hölder class $\mathcal{H}^\alpha(x_0)$ if*

$$|\varphi^{(\lfloor\alpha\rfloor)}(y) - \varphi^{(\lfloor\alpha\rfloor)}(x)| \le C|y-x|^{\{\alpha\}} \quad \text{for every } x,y \in \mathcal{V}_{x_0}. \tag{10}$$

ii) *The function $\psi : \mathcal{V}_{x_0}\times\mathcal{V}_{y_0} \to \mathbb{R}$ belongs to the anisotropic Hölder class $\mathcal{H}^{\alpha,\beta}(x_0,y_0)$ if*

$$x \mapsto \psi(x,y_0) \in \mathcal{H}^\alpha(x_0) \quad \text{and} \quad y \mapsto \psi(x_0,y) \in \mathcal{H}^\beta(y_0)$$

hold simultaneously.

We obtain a semi-norm on $\mathcal{H}^\alpha(x_0)$ by setting $|\varphi|_{\mathcal{H}^\alpha(x_0)} = \sup_{x\in\mathcal{V}_{x_0}} |\varphi(x)| + c_\alpha(\varphi)$, where $c_\alpha(\varphi)$ is the smallest constant for which (10) holds. Likewise, we equip $\mathcal{H}^{\alpha,\beta}(x_0,y_0)$ with the semi-norm $|\psi|_{\mathcal{H}^{\alpha,\beta}(x_0,y_0)} = |\psi(\cdot,y_0)|_{\mathcal{H}^\alpha(x_0)} + |\psi(x_0,\cdot)|_{\mathcal{H}^\beta(y_0)}$. The space $\mathcal{H}^{\alpha,\beta}(x_0,y_0)$ is appended with (semi) Hölder balls

$$\mathcal{H}^{\alpha,\beta}(x_0,y_0)(R) = \big\{\psi : \mathcal{X}\times\mathcal{X}\to\mathbb{R}, |\psi|_{\mathcal{H}^{\alpha,\beta}(x_0,y_0)} \le R\big\}, \quad R>0.$$

We are ready to state our convergence result over transitions $\mathcal{Q}$ that belong to

$$\mathbb{Q}^{\alpha,\beta}_{1/2}(R) = \mathbb{Q}_{1/2} \cap \mathcal{H}^{\alpha,\beta}(x_0,y_0)(R), \quad R>0,$$

with a slight abuse of notation.

Theorem 10. *Work under Assumptions 2,3, 4. Assume that the initial distribution μ is absolutely continuous w.r.t. the Lebesgue measure with a locally bounded density function and satisfies $\mu(V^2)<\infty$.*

Let $\alpha,\beta>0$. Specify $\widehat{\nu}_n(y_0)$ by a kernel of order $k>\max\{\alpha,\beta\}$ and $h = |\mathbb{T}_n|^{-1/(2\beta+1)}$ and $\widehat{q}_n(x_0,y_0)$ with the same kernel and $h_1 = |\mathbb{T}_n|^{-s(\alpha,\beta)/(\alpha\wedge\beta)(2s(\alpha,\beta)+1)}$, $h_2 = |\mathbb{T}_n|^{-s(\alpha,\beta)/\beta(2s(\alpha,\beta)+1)}$ and $\varpi_n\to 0$. Then, for every $R>0$

$$\sup_{\mathcal{Q}\in\mathbb{Q}^{\alpha,\beta}_{1/2}(R)} \big(\mathbb{E}\big[\big(\widehat{\nu}_n(y_0)-\nu(y_0)\big)^2\big]\big)^{1/2} \lesssim |\mathbb{T}_n|^{-\beta/(2\beta+1)}$$

and

$$\sup_{\mathcal{Q}\in\mathbb{Q}^{\alpha,\beta}_{1/2}(R)} \big(\mathbb{E}\big[\big(\widehat{q}_n(x_0,y_0)-q(x_0,y_0)\big)^2\big]\big)^{1/2} \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$$

hold true, where $s(\alpha,\beta)^{-1} = (\alpha\wedge\beta)^{-1}+\beta^{-1}$ is the effective anisotropic smoothness associated with (α,β).

Several remarks are in order: **1)** We obtain an optimal result in the minimax sense for estimating $\nu(y_0)$ and in the case $\beta\ge\alpha$ for estimating $q(x_0,y_0)$. This stems from the fact that the representation $\nu(x) = \int_{\mathcal{X}}\nu(y)q(y,x)dy$ henceforth $q\in\mathcal{H}^{\alpha,\beta}$ implies that $\nu\in\mathcal{H}^\beta$. In turn, the numerator of $\widehat{q}_n(x_0,y_0)$ is based on the estimation of the function $\nu(x)q(x,y)\in\mathcal{H}^{\alpha\wedge\beta,\beta}$. **2)** In the estimation of $q(x_0,y_0)$, we have a superfluous term ϖ_n^{-1} in the error that can be taken arbitrarily small, and that comes from the denominator of the estimator. It can be removed, however at a significant technical cost. Alternatively, one can get rid of it by weakening the error loss: it is not difficult to prove

$$\big(\mathbb{E}\big[\big(\widehat{q}_n(x_0,y_0)-q(x_0,y_0)\big)^p\big]\big)^{1/p} \lesssim |\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)} \quad \text{for every} \quad 0<p<2,$$

and the result of course also holds in probability. **3)** The assumption that μ is absolutely continuous can also be removed. **4)** Finally, a slightly annoying fact is that the estimators $\widehat{\nu}_n(x_0)$ and $\widehat{q}_n(x_0,y_0)$ require the knowledge of (α,β) to be tuned optimally, and this is not reasonable in practice. It is possible to tune our estimators in practice by cross-validation in the same spirit as in [21], but an adaptive estimation theory still needs to be established. This lies beyond the scope of the paper, and requires concentration inequalities, a result we do not have here, due to the fact that the model is not uniformly geometrically ergodic (otherwise, we could apply the same strategy as in [8, 5]).

3.2. **Parametric estimation of the division rate.** In order to conduct inference on the division rate $x\mapsto B(x)$, we need more stringent assumptions on the model so that we can apply the results of Theorem 7. The main difficulty lies in the fact that we need to apply Theorem 7 to test functions of the form $\psi(x,y) = \log q(x,y)$ when applied to the loglikelihood of the data, and that these functions are possibly unbounded.

A stochastic trait model as a diffusion on a compact with reflection at the boundary. We circumvent this difficulty by assuming that the trait $\phi_x(t)$ of each individual evolves in a bounded interval with reflections at the boundary and with no loss of generality, we take $\mathcal{X} = [0, L]$ for some $L > 0$. The dynamics of the traits now follows

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t + d\ell_t, \tag{11}$$

where the solution $(\ell_t)_{t\geq 0}$ to $\ell_t = \int_0^t (\mathbf{1}_{\{\phi_x(s)=0\}} + \mathbf{1}_{\{\phi_x(s)=L\}})d\ell_s$ accounts for the reflection at the boundary and $(W_t)_{t\geq 0}$ is a standard Brownian motion. Under Assumption 2 (that reduces here to the boundedness of r, σ and the ellipticity of σ) there exists a unique strong solution to (11), see for instance Theorem 4.1. in [32].

A slight modification of Proposition (5) gives the following explicit formulae for the transitions $\mathcal{P}$ and $\mathcal{Q}$. Remember that by Assumption 4, we have $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$. Define

$$\mathcal{D} = \big\{0 < y_1 \leq \varepsilon L,\ \tfrac{\varepsilon}{1-\varepsilon} y_1 \leq y_2 \leq \tfrac{1-\varepsilon}{\varepsilon} y_1\big\} \cup \big\{\varepsilon L \leq y_1 \leq (1-\varepsilon)L,\ \tfrac{\varepsilon}{1-\varepsilon} y_1 \leq y_2 \leq \tfrac{L-y_1}{y_1}\big\}.$$

Then the explicit formula for $p(x, y_1, y_2)$ given in (7) remains unchanged provided $(x, y_1, y_2) \in \mathcal{X}\times\mathcal{D}$ and it vanishes outside of $\mathcal{X} \times \mathcal{D}$. For $q(x, y)$, the formula (8) now becomes

$$q(x,y) = \int_{y/L}^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\left[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\right] dz, \tag{12}$$

for $(x, y) \in \mathcal{X} \times \{0 \leq y \leq (1-\varepsilon)L\}$ and 0 otherwise.

Adapting the proof of Theorem 6 to the case of a diffusion living on a compact interval (formally replacing $[-w, w]$ by $[0, L]$ in the proof of Proposition 22 below) one easily checks that Theorem 6 remains valid in this setting (applying for instance Theorem 4.3.16 in [9]). In turn, Theorem 7 also holds true in the case of a reflected diffusion. For parametric estimation, the control on the variance of $\mathcal{M}_n(\psi)$ is less demanding and we will simply need the following

Corollary 11. *Work under Assumptions 2,3, 4 in the case of a reflected diffusion on $[0, L]$ for the evolution of the trait $(\phi_x(t),\ t \geq 0)$. Let $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Then, for any probability measure μ, we have*

$$\sup_{\mathcal{Q}\in\mathbb{Q}_{1/2}} \mathbb{E}_\mu\big[(\mathcal{M}_n(\psi) - \nu(Q\psi))^2\big] \lesssim |\mathbb{T}_n|^{-1} \sup_{x,y} \psi(x,y)^2.$$

Maximum likelihood estimation. From now on, we fix a triplet $(r_0, \sigma_0, \kappa_0)$ and we let the division rate $x \mapsto B(x)$ belong to a parametric class

$$\mathcal{B} = \big\{B : \mathcal{X} \to \mathbb{R}, B(x) = B_0(\vartheta, x), x \in \mathcal{X}, \vartheta \in \Theta\big\},$$

where $x \mapsto B_0(x, \vartheta)$ is known up to the parameter $\vartheta \in \Theta$, and $\Theta \subset \mathbb{R}^d$ for some $d \geq 1$ is a compact subset of the Euclidean space. In this setting, the model is entirely characterised by ϑ which is our parameter of interest. A first minimal stability requirement of the parametric model is the following

Assumption 12. *We have $\mathbb{Q}(\mathcal{B}) = \{\mathcal{Q} = \mathcal{Q}(r_0, \sigma_0, B, \kappa), B \in \mathcal{B}\} \subset \mathbb{Q}_{1/2}$.*

A second minimal requirement is the identifiability of the class $\mathcal{B}$, namely the fact that the map

$$B \mapsto \mathcal{Q}(r_0, \sigma_0, B, \kappa_0)$$

from $\mathcal{B}$ to $\mathbb{Q}$ is injective. This is satisfied in particular if $\mathcal{B}$ satisfies a certain orderliness property.

Definition 13. *A class $\mathcal{B}$ of functions from $\mathcal{X} \to [0,\infty)$ is orderly if $\varphi_1, \varphi_2 \in \mathcal{B}$ implies either $\varphi_1(x) \le \varphi_2(x)$ for every $x \in \mathcal{X}$ or $\varphi_2(x) \le \varphi_1(x)$ for every $x \in \mathcal{X}$.*

Proposition 14. *Let $\mathcal{B}$ be orderly in the sense of Definition 13 and $\mathcal{Q}(\mathcal{B}) \subset \mathbb{Q}$ for some $(r_0, \sigma_0, \kappa_0)$. Then $B \mapsto \mathcal{Q}(r_0, \sigma_0, B, \kappa_0)$ is injective.*

We further stress the dependence on ϑ by introducing a subscript in the notation whenever relevant. We formally obtain a statistical experiment

$$\mathcal{E}^n = \big\{\mathbb{P}^n_\vartheta, \vartheta \in \Theta\big\}$$

by letting $\mathbb{P}^n_\vartheta$ denote the law of $\mathbb{X}^n = (X_u, u \in \mathbb{T}_n)$ under $\mathcal{P}_\vartheta$ with initial condition $X_\emptyset$ distributed according to ν_ϑ on the product space $\mathcal{X}^{|\mathbb{T}_n|}$ endowed with its Borel sigma-field. The experiment $\mathcal{E}^n$ is dominated by the Lebesgue measure on $\mathcal{X}^{|\mathbb{T}_n|}$ and we obtain a likelihood function by setting

$$\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big) = \prod_{u \in \mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u), \tag{13}$$

where X_{u^-} denotes the trait of the parent of u. Taking any maximiser of (13) we obtain a maximum likelihood estimator

$$\widehat{\vartheta}_n \in \operatorname*{argmax}_{\vartheta \in \Theta} \mathcal{L}_n\big(\vartheta, \mathbb{X}^n\big)$$

provided a maximiser exists.

Convergence results and asymptotic efficiency. We first have an existence and consistency result of $\widehat{\vartheta}_n$ under the following non-degeneracy assumption that strengthens Assumption 3.

Assumption 15. *The function $B_0 : \Theta \times \mathcal{X} \to [0,\infty)$ is continuous and for some positive b_3, b_4, we have*

$$0 < b_3 \le \inf_{\vartheta,x} B_0(\vartheta, x) \le \sup_{\vartheta,x} B_0(\vartheta, x) \le b_4$$

Moreover, the class $\mathcal{B} = \big\{B_0(\vartheta, \cdot), \vartheta \in \Theta\big\}$ is orderly in the sense of Definition 13.

Theorem 16. *Work under Assumptions 2, 4, 12 and 15. Then, for every $\vartheta \in \Theta$, $\widehat{\vartheta}_n$ converges to ϑ in probability as $n \to \infty$.*

Our next result gives an explicit rate of convergence and asymptotic normality for $\widehat{\vartheta}_n$. We need further regularity assumptions.

Assumption 17. *The set Θ has non empty interior and, for every $x \in \mathcal{X}$ the map $\vartheta \mapsto B_0(\vartheta, x)$ is three times continuously differentiable. Moreover, for every $1 \le i, j, k \le d$:*

$$\sup_{\vartheta,x} |\partial_{\vartheta_i} B_0(\vartheta, x)| + \sup_{\vartheta,x} |\partial^2_{\vartheta_i \vartheta_j} B_0(\vartheta, x)| + \sup_{\vartheta,x} |\partial^3_{\vartheta_i \vartheta_j \vartheta_k} B_0(\vartheta, x)| < \infty.$$

Introduce the Fisher information $\Psi(\vartheta)$ at point $\vartheta \in \Theta$ as the $d \times d$-matrix with entries:

$$\Psi(\vartheta)_{i,j} = \nu_\vartheta \mathcal{Q}_\vartheta\Big(\frac{\partial_{\vartheta_i} q_\vartheta \, \partial_{\vartheta_j} q_\vartheta}{q_\vartheta^2}\Big) = \int_{\mathcal{X}\times\mathcal{X}} \frac{\partial_{\vartheta_i} q_\vartheta(x,y) \, \partial_{\vartheta_j} q_\vartheta(x,y)}{q_\vartheta(x,y)^2} \nu_\vartheta(x) q_\vartheta(x,y) dx dy,$$

for $1 \le i, j \le d$.

Assumption 18. *For every ϑ in the interior of Θ, the matrix $\Psi(\vartheta)$ is nonsingular.*

Although standard in regular parametric estimation, Assumption 18 is not obviously satisfied even if we have the explicit formula (12), for $q_\vartheta(x,y)$, due to its relatively intricate form. We can however easily show that it is satisfied in the special case of a trait evolving as a reflected diffusion with constant drift. More general parametrisations are presumably possible, adapting the proof delayed until Appendix 6.2.

Proposition 19. *Assume $d=1$, $B_0(\vartheta,x)=\vartheta$ for every $x\in\mathcal{X}$, with $\Theta=[\vartheta_1,\vartheta_2]\subset(0,\infty)$, $r(x)=r_1<0$ and $\sigma(x)=\sigma_0>0$ for every $x\in\mathcal{X}$. Let $\kappa(z)=(1-2\varepsilon)^{-1}$ for every $z\in[\varepsilon,1-\varepsilon]$. There exists an explicit open interval $\mathcal{I}\subset(0,1/2)$ such that Assumption 18 is satisfied as soon as $\varepsilon\in\mathcal{I}$.*

We are ready to state our final result on asymptotic normality of $\widehat{\vartheta}_n$.

Theorem 20. *Work under Assumptions 2, 4, 12, 15, 17 and 18. For every ϑ in the interior of Θ, we have*

$$\mathbb{T}_n^{1/2}(\widehat{\vartheta}_n-\vartheta)\to\mathcal{N}\big(0,\Psi(\vartheta)^{-1}\big)$$

in distribution as $n\to\infty$, where $\mathcal{N}(0,\Psi(\vartheta)^{-1})$ denote the d-dimensional Gaussian distribution with mean 0 and covariance the inverse of the Fisher matrix $\Psi(\vartheta)$.

Several remarks are in order: **1)** Although asymptotically optimal, the practical implementation of $\widehat{\vartheta}_n$ is a challenging question that we plan to address in a systematic way. **2)** As for classical estimation in diffusion processes (see *e.g.* [11, 16]), the assumptions of Theorem 20, especially Assumption 18 are standard. However, the fact that they hold true in the simple case of Proposition 19 and a glance at the proof is an indication that they are certainly true in wider generality.

4. Numerical implementation

We consider the implementation of the estimator $\widehat{\vartheta}_n$ in the case of a branching population structured by a trait drawn according to a Brownian motion reflected on $[0,1]$, namely

$$\phi_x(t)=x+W_t+\ell_t,\quad \ell_t=\int_0^t(\mathbf{1}_{\{\phi_x(s)=0\}}+\mathbf{1}_{\{\phi_x(s)=1\}})d\ell_s,\quad t\geq 0,$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion. We pick $\kappa(z)=(1-2\varepsilon)^{-1}\mathbf{1}_{[\varepsilon,1-\varepsilon]}(z)$ so that an individual with trait x at division splits into two individuals with traits Ux and $(1-U)x$ respectively, where U is uniformly distributed on $[\varepsilon,1-\varepsilon]$. We pick $\varepsilon=10^{-4}$.

4.1. Generation of simulated data.

Given a division rate B and an initial trait $x^{\mathrm{in}}\in\mathbb{R}$, we recursively construct a dataset of size $\mathbb{T}_n=2^{n+1}-1$. We pick T large enough so that each individual of the (N_n+1)-th generation is born before time T. For each individual u in the "current" population (composed of individuals from generations smaller than N_n) and born at time $t<T$,

i) we simulate a reflected diffusion using the Euler scheme for reflected stochastic differential equations of [24] with initial condition x^{in} and time step Δt until time T,
ii) we draw the lifetime of u by rejection sampling,
iii) we add to the population two new individuals with respective traits at birth given by Ux and $(1-U)x$ where U is a realisation of a uniform random variable on $[\varepsilon,1-\varepsilon]$ and x is the trait of u at division,
iv) we remove the individual u from the "current" population.

We can thus simulate a sample of size $2^{n+1}-1$ by considering all the couples composed of the traits at birth of an individual and its ancestor.

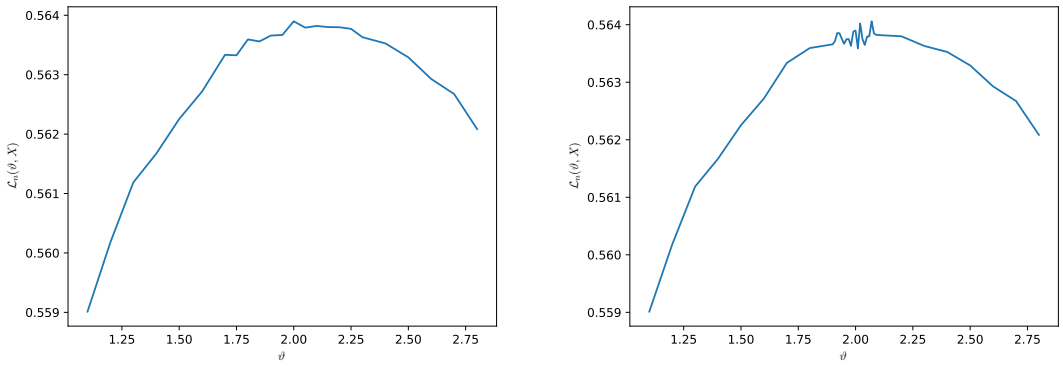

FIGURE 2. The contrast $\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big)$ for $n = 14$ and $B(x, \vartheta) \equiv 2$ with $\Delta\vartheta = 0.05$ (left) and $\Delta\vartheta = 0.01$ (right). We look at the results only for $\Delta\vartheta = 0.05$ because for smaller values of the increment, the noise due to computational errors is too important compared to the different values of the contrast.

4.2. **Implementation of the maximum likelihood.** For a given dataset $\mathbb{X}^n$, we approximate $\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big)$ using, for a given ϑ, the nonparametric estimator $\widehat{q}_n(X_{u^-}, X_u)$ introduced in Section 3.1.

More specifically, we implement $\widehat{q}_n(x_0, y_0) = \dfrac{\mathcal{M}_n\big(G^{\otimes 2}_{h_1,h_2}(\cdot - x_0, \cdot - y_0)\big)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi}$ for every (x_0, y_0) on a grid of mesh $n_1^{-1} \times n_2^{-1}$ of $[0,1] \times [0,1]$ with $n_1 = n_2 = 200$, $G(x) = (2\pi)^{-1/2} \exp(-x^2/2)$, $h = 2|\mathbb{T}_n|^{-1/3}$, $h_1 = h_2 = 0.1\sqrt{h}$, $\varpi = 10^{-6}$. We next use an interpolation scheme provided by the package *Interpolations* in Julia to compute the value of the transition at each $(X_{u^-}, X_u) \in \mathbb{X}^n$. For synthetic data, we pick $n = 19$, resulting in a tree of size $2^{20} - 1 = 1\,048\,575$ with initial value $x^{\text{in}} = 0.5$ and $\Delta t = 5\,10^{-4}$.

4.3. **Results.** We consider the following parametric classes $B_0(\vartheta, x) = \vartheta$ and $B_1(\vartheta, x) = 1 + \vartheta x$. We compute 300 Monte-Carlo samples of size $|\mathbb{T}_n| = 2^{15} - 1 = 32\,767$ for $\vartheta = \vartheta_1 = 2$ and $\vartheta = \vartheta_2 = 15$ in both cases. Therefore, we apply our results to four different cases. In each case, we approximate $q_\vartheta(X_u^-, X_u)$ for different values of $\vartheta \in \Theta = [\vartheta_{\min}, \vartheta_{\max}]$ and we compute the corresponding $\widehat{\vartheta}_n$. We progressively reduce the increment $\Delta\vartheta$ for the choice of ϑ until the contrast of likelihood starts to be noisy (see Figure 2), adapting at each level the choice for the upper and lower bounds of Θ. The results are displayed in Table 1. We recover the parameter in all four cases, with various accuracies. The most accurate value is obtained for B_1 with a small value of the parameter, *i.e.* $\vartheta = 2$. We did not reach the optimal accuracy $|T_n^\star|^{-1/2} \approx 6\,10^{-3}$. One could presumably obtain a better accuracy by choosing a finer discretisation of $[0,1] \times [0,1]$ for the computation of the $\widehat{q}_n$. But this choice leads to an important increase of the computational time. The results in the case of a linear division rate are less accurate. Those results could also probably be improved using a cross-validation procedure for the choice of the bandwidth parameters h, h_1, h_2.

B	ϑ	Mean	Std Dev.	$\vartheta_{\min}$	$\vartheta_{\max}$	$\Delta\vartheta$
B_0	2	1.9908	0.0845	1.8	2.2	0.05
	15	15.029	0.4763	14	16	0.25
B_1	2	2.1392	0.3485	1.375	3	0.125
	15	15.023	0.6425	13.75	16.5	0.25

TABLE 1. Results for B in $\mathcal{B}_1$ and $\mathcal{B}_2$. For each parametric class of functions $\mathcal{B}_1$ and $\mathcal{B}_2$, we display the results for $\vartheta = 2$ and $\vartheta = 15$. The third and fourth columns corresponds respectively to the mean value and the standard deviation of $\widehat{\vartheta}_n$, computed with 300 different data sets of size $2^{n+1} - 1 = 32767$. The fifth and sixth columns correspond to a 95% confidence interval. The last column corresponds to the value of the step for the discretisation of Θ, which limits the accuracy of the result.

5. Proofs

5.1. **Proof of Proposition 5.** We first prove (8). By the occupation times formula, a change of variable and Fubini's theorem, for any bounded $\varphi : \mathcal{X} \to \mathbb{R}$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}
\mathcal{Q}\varphi(x) &= \mathbb{E}\big[\varphi(X_u)\big|X_{u^-} = x\big] \\
&= \mathbb{E}\Big[\int_0^1 \kappa(z) \int_0^\infty \varphi\big(z\phi_x(t)\big) B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} dt dz\Big] \\
&= \mathbb{E}\Big[\int_0^1 \frac{\kappa(z)}{z} \int_{\mathbb{R}} \varphi(y) B(y/z) \int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} \frac{dL_t^{y/z}(\phi_x)}{\sigma(y/z)^2} dy dz\Big] \\
&= \int_{\mathbb{R}} \varphi(y) \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x) dz\Big] dy,
\end{aligned}$$

where $(L_t^y(\phi_x))_{t\geq 0}$ is the local time of ϕ_x at $y \in \mathcal{X}$. Since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, the above integrals are well defined and (8) is established. We turn to (7). By (3), for any bounded $\psi : \mathcal{X}^3 \to \mathbb{R}$ and $x \in \mathcal{X}$, we have

$$\begin{aligned}
\mathcal{P}\psi(x) &= \mathbb{E}\Big[\int_0^1 \kappa(z) \int_0^\infty \psi\big(x, z\phi_x(t), (1-z)\phi_x(t)\big) B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} dt dz\Big] \\
&= \int_{\mathbb{R}} \int_0^1 \psi\big(x, y_1, \tfrac{1-z}{z} y_1\big) B\big(y_1/z\big)\sigma(y_1/z)^{-2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y_1/z}(\phi_x)\Big] \frac{\kappa(z)}{z} dz dy_1 \\
&= \int_{\mathcal{D}} \psi\big(x, y_1, y_2\big) \frac{\kappa\big(y_1/(y_1+y_2)\big)}{y_1+y_2} B\left(y_1+y_2\right) \sigma(y_1+y_2)^{-2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y_1+y_2}(\phi_x)\Big]
\end{aligned}$$

where we set $y_2 = \frac{1-z}{z} y_1$ in order to obtain the last line. The integral is taken over the domain

$$\mathcal{D} = \big\{(y_1, y_2) \in \mathbb{R}^2,\ \tfrac{\varepsilon}{1-\varepsilon} y_1 \leq y_2 \leq \tfrac{1-\varepsilon}{\varepsilon} y_1\big\} \subset \mathrm{supp}\big((y_1, y_2) \mapsto \kappa(z_1 + z_2)\big)$$

therefore the above integral is well defined and the representation (7) is proved.

5.2. **Proof of Theorem 6.** The proof goes along a classical path: we establish a drift and a minorisation condition in Proposition 21 and 22 below, and then apply for instance Theorem 1.2. in [18], see also the references therein.

Proposition 21 (Drift condition). *Let $V(x) = x^2$. Work under Assumptions 2,3 and 4. There exist explicitly computable $0 < v_1 = v_1(\varepsilon) < 1$ and $v_2 = v_2(\varepsilon, r_1, r_2, \sigma_1, \sigma_2, b_1) > 0$ such that*

$$\mathcal{Q}V(x) \leq v_1 V(x) + v_2.$$

Proposition 22 (Minorisation condition). *Work under Assumption 2, 3 and 4. For large enough $w > 0$, there exists $\alpha \in (0,1)$ and a probability measure μ on $\mathcal{X}$ such that*

$$\inf_{\{x, |x| \leq w\}} \mathcal{Q}(x, \mathcal{A}) \geq \alpha \mu(\mathcal{A})$$

for every Borel set $\mathcal{A} \subset \mathcal{X}$.

Proof of Proposition 21. Fix $x \in \mathcal{X}$ and let $m(\kappa) = \int_0^1 z^2 \kappa(z) dz$. By Itô formula, we obtain the decomposition

$$\mathcal{Q}V(x) = \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\Big[z^2 \phi_x(t)^2 B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} \Big] dt dz = m(\kappa)(I + II + III + IV),$$

where

$$\begin{aligned}
I &= x^2 \int_0^\infty \mathbb{E}\Big[B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} \Big] dt, \\
II &= 2 \int_0^\infty \mathbb{E}\Big[\int_0^t \phi_x(u) r(\phi_x(u)) du \, B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} \Big] dt, \\
III &= \int_0^\infty \mathbb{E}\Big[\int_0^t \sigma\big(\phi_x(u)\big)^2 du \, B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} \Big] dt, \\
IV &= 2 \int_0^\infty \mathbb{E}\Big[\int_0^t \phi_x(u) \sigma\big(\phi_x(u)\big) dW_u \, B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} \Big] dt.
\end{aligned}$$

First, note that $\int_0^\infty B(\phi_x(t)) e^{-\int_0^t B(\phi_x(s))ds} dt = 1$, therefore $I = x^2$ by Fubini's theorem. We turn to II. By Fubini's theorem again:

$$\begin{aligned}
II &= 2\,\mathbb{E}\Big[\int_0^\infty \phi_x(u) r\big(\phi_x(u)\big) \Big(\int_u^\infty B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} dt \Big) du \Big] \\
&= 2\,\mathbb{E}\Big[\int_0^\infty \phi_x(u) r\big(\phi_x(u)\big) e^{-\int_0^u B(\phi_x(s))ds} du \Big] \\
&= 2 \int_{\mathbb{R}} z r(z) \sigma(z)^{-2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^z(\phi_x) \Big] dz.
\end{aligned}$$

by the occupation times formula. By Assumption 2 we have $zr(z) < 0$ for $|z| \geq r_2$, therefore:

$$\begin{aligned}
II &\leq 2 \int_{-r_2}^{r_2} z r(z) \sigma(z)^{-2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^z(\phi_x) \Big] dz \\
&\leq 2 r_1 r_2 (1 + r_1) \sigma_1^{-2} \int_{-r_2}^{r_2} \mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^z(\phi_x) \Big] dz \\
&\leq 2 r_1 r_2 (1 + r_1) \sigma_1^{-2} \int_0^\infty e^{-b_1 t} \mathbb{P}\big(-t_2 \leq \phi_x(t) \leq r_2 \big) dt \leq 2 r_1 r_2 (1 + r_1) \sigma_1^{-2} b_1^{-1}.
\end{aligned}$$

using successively Assumption 2, 3 and the occupation times formula. For the term III, we have

$$\begin{aligned} III &= \int_0^\infty \mathbb{E}\big[\sigma\big(\phi_x(u)\big)^2 \int_u^\infty B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds}\Big] du \\ &= \int_0^\infty \mathbb{E}\left[\sigma(\phi_x(u))^2 e^{-\int_0^u B(\phi_x(s))ds}\right] du \end{aligned}$$

by Fubini again and this last quantity is less than $\sigma_2^2 \int_0^\infty e^{-b_1 t} dt = \sigma_2^2 b_1^{-1}$ by Assumption 2 and 3. Similarly for the term IV, we have

$$\begin{aligned} IV &= 2\,\mathbb{E}\Big[\int_0^\infty \sigma\big(\phi_x(u)\big) \int_u^\infty B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds} dt dW_u\Big] \\ &= 2\,\mathbb{E}\Big[\int_0^\infty \sigma\big(\phi_x(u)\big) e^{-\int_0^u B(\phi_x(s))ds} dW_u\Big] \end{aligned}$$

and this last quantity vanishes. Putting the estimates for I, II, III and IV together, we conclude

$$\mathcal{Q}V(x) \le m(\kappa)x^2 + m(\kappa)(2r_1r_2(1+r_1)\sigma_1^{-2} + \sigma_2^2)b_1^{-1}.$$

Since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, we have $m(\kappa) \le (1-\varepsilon)^2 < 1$ and this completes the proof with $v_1 = m(\kappa)$ and $v_2 = m(\kappa)(2r_1r_2(1+r_1)\sigma_1^{-2} + \sigma_2^2)b_1^{-1}$.

□

Proof of Proposition 22. Step 1). Let $x \in [-w, w]$ and $\mathcal{A} \subset \mathcal{X}$ be a Borel set. Applying Assumption 4, introducing the event $\mathcal{W}(\phi_x(t)) = \{2w \le \phi_x(t) \le 2w\sqrt{(1-\varepsilon)/\varepsilon}\}$, applying Fubini's theorem and a change of variable, we successively obtain

$$\begin{aligned} \mathcal{Q}(x,\mathcal{A}) &= \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\{z\phi_x(t)\in\mathcal{A}\}} B\big(\phi_x(t)\big) e^{-\int_0^t B(\phi_x(s))ds}\Big] dt dz \\ &\ge b_1 \int_0^1 \kappa(z) \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))} \mathbf{1}_{\{z\phi_x(t)\in\mathcal{A}\}} e^{-\int_0^t B\left(\phi_x(s)\right)ds}\Big] dt dz \\ &\ge b_1 \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))} e^{-\int_0^t B(\phi_x(s))ds} \int_0^{\phi_x(t)} \mathbf{1}_{\{y\in\mathcal{A}\}} \kappa\big(y/\phi_x(t)\big)\phi_x(t)^{-1} dy\Big] dt \\ &= \int_{\mathbb{R}} \mathbf{1}_{\{y\in A\}} f(x,y) dy, \end{aligned}$$

with $f(x,y) = b_1 \int_0^\infty \mathbb{E}\big[\mathbf{1}_{\mathcal{W}(\phi_x(t))} \mathbf{1}_{\{y\le\phi_x(t)\}} e^{-\int_0^t B(\phi_x(s))ds} \kappa\big(y/\phi_x(t)\big)\phi_x(t)^{-1}\big] dt$.

Step 2). We now prove that f is bounded below independently of x. By Assumption 4 and the definition of $\mathcal{W}(\phi_x(t))$, we have

$$\begin{aligned} f(x,y) &\ge \delta b_1 \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t))} \mathbf{1}_{\{\varepsilon\phi_x(t)\le y\le(1-\varepsilon)\phi_x(t)\}} e^{-\int_0^t B(\phi_x(s))ds} \phi_x(t)^{-1}\Big] dt \\ &\ge \frac{\delta b_1}{2w\sqrt{(1-\varepsilon)/\varepsilon}} \mathbf{1}_{\left\{2w\sqrt{\varepsilon(1-\varepsilon)}\le y\le 2w(1-\varepsilon)\right\}} h(x,y), \end{aligned}$$

with $h(x,y) = \int_0^\infty \mathbb{E}\big[\mathbf{1}_{\mathcal{W}(\phi_x(t))} e^{-\int_0^t B(\phi_x(s))ds}\big] dt$. Let $\Delta = (1+\sqrt{(1-\varepsilon)/\varepsilon})w$ denote the mid-point of the interval $[2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]$. Let also $T_y^x = \inf\{t > 0, \phi_x(t) \ge y\}$ denote the exit time of the

interval $(\inf \mathcal{X}, y)$ by $(\phi_x(t))_{t\geq 0}$. It follows that

$$\begin{aligned} h(x,y) &\geq \mathbb{E}\Big[\int_{T_\Delta^x}^{\infty} \mathbf{1}_{\mathcal{W}(\phi_x(t))} e^{-\int_0^t B(\phi_x(s))ds} dt\Big] \\ &= \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))} e^{-\int_0^{t+T_\Delta^x} B(\phi_x(s))ds}\Big] dt \\ &= \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))} e^{-\int_0^{T_\Delta^x} B(\phi_x(s))ds} e^{-\int_0^t B(\phi_x(s+T_\Delta^x))ds}\Big] dt \\ &\geq \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_x(t+T_\Delta^x))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-\int_0^t B(\phi_x(s+T_\Delta^x))ds}\Big] dt, \end{aligned}$$

by Assumption 3 and because $\phi_x(s) \leq \Delta$ for every $s \leq T_\Delta^x$. Applying the strong Markov property, we further obtain

$$h(x,y) \geq \int_0^\infty \mathbb{E}\Big[\mathbf{1}_{\mathcal{W}(\phi_\Delta(t))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-\int_0^t B(\phi_\Delta(s))ds}\Big] dt \tag{14}$$

since $\phi_x(T_\Delta^x) = \Delta$ for $x \leq w < \Delta$. Introduce next $T_{\mathcal{W}}^\Delta = \inf\{t \geq 0,\ \phi_\Delta(t) \notin [2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]\}$, *i.e.* the exit time of $[2w, 2w\sqrt{(1-\varepsilon)/\varepsilon}]$ by $(\phi_\Delta(t))_{t\geq 0}$. By (14) and Assumption 3 again, it follows that

$$\begin{aligned} h(x,y) &\geq \mathbb{E}\Big[\int_0^{T_{\mathcal{W}}^\Delta} \mathbf{1}_{\mathcal{W}(\phi_\Delta(t))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-\int_0^t (b_2|\phi_x(s)|^\gamma+b_1)ds} dt\Big] \\ &\geq \mathbb{E}\left[\int_0^{T_{\mathcal{W}}^\Delta} \mathbf{1}_{\mathcal{W}(\phi_\Delta(t))} e^{-T_\Delta^x(b_2\Delta^\gamma+b_1)} e^{-v_3 t} dt\right], \end{aligned}$$

using that $\phi_x(t) \leq 2w\sqrt{(1-\varepsilon)/\varepsilon}$ for $t \leq T_{\mathcal{W}}^\Delta$ and where $v_3 = b_2(2w\sqrt{(1-\varepsilon)/\varepsilon})^\gamma + b_1$. Since $\Delta > -w$, the event $\{T_\Delta^x \leq T_\Delta^{-w}\}$ holds almost-surely for every $x \in [-w, w]$ and therefore

$$h(x,y) \geq \mathbb{E}\Big[\int_0^{T_{\mathcal{W}}^\Delta} e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)} e^{-tv_3} dt\Big] \geq v_3^{-1}\mathbb{E}\big[e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)}\big]\mathbb{E}\big[1 - e^{-T_{\mathcal{W}}^\Delta v_3}\big]. \tag{15}$$

by the independence of $T_{\mathcal{W}}^\Delta$ and T_Δ^{-w}. Furthermore, for every $a, x \in \mathcal{X}$ with $x < a$, we have

$$\mathbb{P}\left(T_a^x < \infty\right) = \lim_{u\to-\infty} \frac{s(u) - s(x)}{s(u) - s(a)},$$

where $s(x) = \int_{\inf\mathcal{X}}^x \exp(-2\int_{\inf\mathcal{X}}^y r(z)\sigma(z)^{-2}dz)dy$, is the scale function associated to $(\phi_x(t))_{t\geq 0}$. By the classical Feller classification of scalar diffusions (see *e.g.* Revuz and Yor [31]), we have the equivalence $\mathbb{P}\left(T_a^x < \infty\right) = 0$ if only if $\int_{\inf\mathcal{X}} s(x)dx = 0$ but that latter property contradicts Assumption 2. Therefore, there exist $w_1, \delta_1 > 0$ such that $\mathbb{P}(T_\Delta^{-w} \leq w_1) > \delta_1$. It follows that

$$\mathbb{E}\big[e^{-T_\Delta^{-w}(b_2\Delta^\gamma+b_1)}\big] \geq \mathbb{E}\big[e^{-w_1(b_2\Delta^\gamma+b_1)}\mathbf{1}_{\{T_\Delta^{-w}\leq w_1\}}\big] \geq \delta_1 e^{-w_1(b_2\Delta^\gamma+b_1)} \tag{16}$$

and since $T_{\mathcal{W}}^\Delta > 0$ almost surely, there exists $\delta_2 > 0$, independent of x, such that

$$\mathbb{E}\big[1 - e^{-T_{\mathcal{W}}^\Delta v_3}\big] > \delta_2. \tag{17}$$

Back to (15), putting together (16) and (17), we obtain

$$h(x,y) \geq v_3^{-1}\delta_1 e^{-w_1(b_2\Delta^\gamma+b_1)}\delta_2$$

and eventually

$$f(x,y) \geq \frac{\delta b_1 \delta_1 \delta_2}{2w\sqrt{(1-\varepsilon)/\varepsilon} v_3} \mathbf{1}_{\left\{2w\sqrt{\varepsilon(1-\varepsilon)} \leq y \leq 2w(1-\varepsilon)\right\}} e^{-w_1(b_2\Delta^\gamma + b_1)}.$$

Step 3). Define the probability measure $\mu(dy) = f(y)dy$ on $\mathcal{X}$ by

$$f(y) = \frac{1}{2w\big(1-\varepsilon-\sqrt{\varepsilon(1-\varepsilon)}\big)} \mathbf{1}_{\left\{2w\sqrt{\varepsilon(1-\varepsilon)} \leq y \leq 2w(1-\varepsilon)\right\}},$$

and let $\alpha = (1-\varepsilon-\sqrt{\varepsilon(1-\varepsilon)}) \frac{b_1\delta\delta_1\delta_2}{\sqrt{(1-\varepsilon)/\varepsilon} v_3} e^{-w_1(b_2\Delta^\gamma + b_1)}$. We may assume that $0 < \alpha < 1$ (the lower bound remains valid if we replace δ by $\delta' < \delta$ for instance) and we thus have established

$$\mathcal{Q}(x, \mathcal{A}) \geq \alpha \mu(\mathcal{A}),$$

for an arbitrary Borel set $\mathcal{A} \subset \mathcal{X}$. The proof of Proposition 22 is complete. □

5.3. **Proof of Theorem 7.**

Preparations. We first state a useful estimate on the local time of $L_t^y(\phi_x)$ as $t \to \infty$. Its proof is delayed until Appendix 6.1.

Lemma 23. *Work under Assumption 2. For every compact $\mathcal{K} \subset \mathcal{X}$ and for every $t \geq 0$, we have*

$$\sup_{x \in \mathcal{X}, y \in \mathcal{K}} \mathbb{E}\left[L_t^y(\phi_x)\right] \lesssim 1 + t^{3/2},$$

up to a constant that only depends on r_1, r_2 and σ_2. In particular, for every $c > 0$, the function

$$y \mapsto \int_0^\infty e^{-ct} \sup_{x \in \mathcal{X}} \mathbb{E}\left[L_t^y(\phi_x)\right] dt$$

is well-defined and locally bounded, uniformly over $\mathbb{Q}$.

Lemma 23 enables us to obtain estimates on the action of $\mathcal{P}$ and $\mathcal{Q}$ on functions $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with nice behaviours that will prove essential for obtaining Theorem 7. We set

$$\mathcal{Q}\psi(x) = \int_{\mathcal{X}} \psi(x,y) q(x,y) dy, \quad \mathcal{P}(\psi \otimes \psi)(x) = \int_{\mathcal{X}\times\mathcal{X}} \psi(x,y_1)\psi(x,y_2) p(x,y_1,y_2) dy_1 dy_2,$$

where $p(x,y_1,y_2)$ and $q(x,y)$ are given in Proposition 5.

Lemma 24. *Work under Assumptions 2, 3 and 4. Let $\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be bounded and such that $\psi_\star$ has compact support. There exists a constant $c_{\mathrm{supp}(\psi_\star)}$ depending on $\mathrm{supp}(\psi_\star)$ (and $\mathcal{Q}$) such that*

$$|\mathcal{Q}\psi(x)| \leq c_{\mathrm{supp}(\psi_\star)} \int_{\mathcal{X}} |\psi(x,y)| dy \tag{18}$$

and

$$|\mathcal{P}(\psi \otimes \psi)(x)| \leq c_{\mathrm{supp}(\psi_\star)} \psi^\star(x) \int_{\mathcal{X}} |\psi(x,y)| dy.$$

Note in particular that (18) implies in turn the estimate

$$|\mathcal{Q}^2\psi(x)| \leq c^2_{\mathrm{supp}(\psi_\star)} |\psi|_1 \wedge c_{\mathrm{supp}(\psi_\star)} \Big| \int_{\mathcal{X}} \psi_\star(y) dy \Big| \lesssim |\psi|_{\wedge 1}.$$

Proof. By Assumption 3, we have

$$\begin{aligned}|\mathcal{Q}\psi(x)| &= \Big|\int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\big[\psi(x,z\phi_x(t))B\big(\phi_x(t)\big)e^{-\int_0^t B(\phi_x(s))ds}\big]dtdz\Big| \\ &\le \int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\big[\big|\psi(x,z\phi_x(t))\big|\big(b_2\,|\phi_x(t)|^\gamma + b_1\big)e^{-b_1 t}\big]dtdz \\ &\le \int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\Big[\big|\psi(x,z\phi_x(t))\big|\big(b_2\,|\phi_x(t)|^\gamma + b_1\big)\int_t^\infty b_1 e^{-b_1 s}ds\Big]dtdz.\end{aligned}$$

Next, by Fubini's theorem and the occupation times formula, we derive

$$\begin{aligned}|\mathcal{Q}\psi(x)| &\le \int_0^1 \kappa(z)\int_0^\infty b_1 e^{-b_1 s}\int_0^s \mathbb{E}\big[|\psi(x,z\phi_x(t))|(b_2|\phi_x(t)|^\gamma + b_1)\big]dtdsdz \\ &\le \int_0^1 \kappa(z)\int_0^\infty b_1 e^{-b_1 s}\int_{\mathbb{R}} |\psi(x,zy)|(b_2|y|^\gamma + b_1)\sigma(y)^{-2}\mathbb{E}[L_s^y(\phi_x)]dydsdz \\ &= \int_0^1 \kappa(z)z^{-1}\int_0^\infty b_1 e^{-b_1 s}\int_{\mathbb{R}} |\psi(x,y)|(b_2|y/z|^\gamma + b_1)\sigma(y/z)^{-2}\mathbb{E}[L_s^{y/z}(\phi_x)]dydsdz \\ &\le c_{\mathrm{supp}(\psi_\star)}\int_{\mathbb{R}} |\psi(x,y)|dy,\end{aligned}$$

and (18) is proved with

$$c_{\mathrm{supp}(\psi_\star)} = \sup_{y\in\mathrm{supp}(\psi_\star),x\in\mathcal{X},z\in[\varepsilon,1-\varepsilon]} b_1(b_2|y/z|^\gamma + b_1)\sigma(y/z)^{-2}z^{-1}\int_0^\infty e^{-b_1 s}\mathbb{E}[L_s^{y/z}(\phi_x)]ds.$$

This last quantity is finite by Lemma 23. For the second estimate, we have

$$\begin{aligned}|\mathcal{P}(\psi\otimes\psi)(x)| &\le \int_0^1 \kappa(z)\int_0^\infty \mathbb{E}\big[\big|\psi\big(x,z\phi_x(t)\big)\psi\big(x,(1-z)\phi_x(t)\big)\big|B(\phi_x(t))e^{-\int_0^t B(\phi_x(s))ds}\big]dtdz \\ &\le \big|\mathcal{Q}\psi(x)\big|\sup_y\big|\psi(x,y)\big|\end{aligned}$$

and we conclude by applying (18). □

Completion of proof of Theorem 7. Without loss of generality, we may assume that $\nu(\mathcal{Q}\psi) = 0$, the general case being obtained by considering the function $\psi(x,y) - \nu(\mathcal{Q}\psi)$. Of course, the compact support property is lost by adding a constant and one has to be careful when revisiting the estimates of *Step 1)* to *Step 4)* below. They exhibit additional error terms that all have the right order using Lemma 24 and the fact that $\mathcal{P}\mathbf{1} = \mathcal{Q}\mathbf{1} = \mathbf{1}$.

Step 1). We start with a standard preliminary decomposition, see for instance [6, 7], expanding the sum in $u \in \mathbb{T}_n^\star$. We have

$$\begin{aligned}\mathbb{E}[\mathcal{M}_n(\psi)^2] &= |\mathbb{T}_n^\star|^{-2}\mathbb{E}_\mu\Big[\Big(\sum_{m=1}^n\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\Big)^2\Big] \\ &\le |\mathbb{T}_n^\star|^{-2}\Big(\sum_{m=1}^n\Big(\mathbb{E}_\mu\Big[\Big(\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\Big)^2\Big]\Big)^{1/2}\Big)^2\end{aligned}$$

by triangle inequality. Thus Theorem 7 amounts to control

$$\mathbb{E}_\mu\big[\big(\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\big)^2\big]=I_m+II_m,$$

with

$$I_m=\mathbb{E}_\mu\big[\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)^2\big],$$
$$II_m=\mathbb{E}_\mu\big[\sum_{u,v\in\mathbb{G}_m,u\neq v}\psi(X_{u^-},X_u)\psi(X_{v^-},X_v)\big].$$

Step 2). The control of the term I_m is straightforward: by Lemma 24 we have

$$I_1=2\mu(\mathcal{Q}\psi^2)\le 2c_{\mathrm{supp}(\psi_\star)}\int_{\mathcal{X}\times\mathcal{X}}\psi(x,y)^2\mu(dx)dy\ \text{ for }\ m=1,$$
$$I_m=2^m\mu(\mathcal{Q}^m\psi^2)\le 2^m(c^2_{\mathrm{supp}(\psi_\star)}|\psi^2|_1\wedge c_{\mathrm{supp}(\psi_\star)}|\psi^2_\star|_1)\ \text{ for }\ m\ge 2,$$

therefore $I_m\lesssim 2^m|\psi^2|_\mu$ holds for every $m\ge 1$. In the case $\nu(Q\psi)\neq 0$, we replace $|\psi^2|_\mu$ by $|\psi^2|_{\mu+\nu}$.

Step 3). We further decompose the main term $II_m=III_m+IV_m$, having

$$III_m=\mathbb{E}_\mu\big[\sum_{w\in\mathbb{G}_{m-1}}\psi(X_w,X_{w0})\psi(X_w,X_{w1})\big],$$
$$IV_m=\mathbb{E}_\mu\big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v}\sum_{i,j=0,1}\psi(X_u,X_{ui})\psi(X_v,X_{vj})\big].$$

The control of III_m is straightforward:

$$III_m=\mathbb{E}_\mu\big[\sum_{w\in\mathbb{G}_{m-1}}\mathcal{P}(\psi\otimes\psi)(X_w)\big]=2^{m-1}\mu\big(\mathcal{Q}^{m-1}\mathcal{P}(\psi\otimes\psi)\big).$$

In the same way as for the term I_m, by Lemma 24, one readily checks that $|III_m|\lesssim 2^{m-1}|\psi^\star\psi|_\mu$.

Step 4). We now turn to the main term IV_m. Writing here $u\wedge v$ for the most common recent ancestor of u and v, conditioning w.r.t. $\mathcal{F}_{|u\wedge v|+1}$ and using the conditional independence of (X_u,X_{ui}) and (X_v,X_{vj}) given $\mathcal{F}_{|u\wedge v|+1}$ thanks to the BMC property (3), we successively obtain

$$\begin{aligned}
IV_m&=\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v}\sum_{i,j=0,1}\mathbb{E}_\mu\left[\psi(X_u,X_{ui})\psi(X_v,X_{vj})\middle|\mathcal{F}_{|u\wedge v|+1}\right]\Big]\\
&=\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v}\sum_{i,j=0,1}\mathbb{E}_\mu\left[\psi(X_u,X_{ui})\middle|\mathcal{F}_{|u\wedge v|+1}\right]\mathbb{E}_\mu\left[\psi(X_v,X_{vj})\middle|\mathcal{F}_{|u\wedge v|+1}\right]\Big]\\
&=\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v}\sum_{i,j=0,1}\mathbb{E}_\mu\left[\mathcal{Q}\psi(X_u)\middle|\mathcal{F}_{|u\wedge v|+1}\right]\mathbb{E}_\mu\left[\mathcal{Q}\psi(X_v)\middle|\mathcal{F}_{|u\wedge v|+1}\right]\Big]\\
&=4\,\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v}\mathcal{Q}^{|u|-|u\wedge v|}\psi(X_{u^*})\mathcal{Q}^{|u|-|u\wedge v|}\psi(X_{v^*})\Big],
\end{aligned}$$

where u^* (respectively v^*) is the descendant of $u \wedge v$ which is an ancestor of u (respectively v). Conditioning further w.r.t. $\mathcal{F}_{|u\wedge v|}$ we obtain

$$\begin{aligned} IV_m &= 4\,\mathbb{E}_\mu\Big[\sum_{u,v\in\mathbb{G}_{m-1},u\neq v} \mathcal{P}\big(\mathcal{Q}^{|u|-|u\wedge v|}\psi\otimes\mathcal{Q}^{|u|-|u\wedge v|}\psi\big)(X_{u\wedge v})\Big] \\ &= 4\sum_{u,v\in\mathbb{G}_{m-1},u\neq v} \mu\big(\mathcal{Q}^{|u\wedge v|}\mathcal{P}(\mathcal{Q}^{|u|-|u\wedge v|}\psi\otimes\mathcal{Q}^{|u|-|u\wedge v|}\psi)\big) \\ &= 4\sum_{l=1}^{m-1}\sum_{u,v\in\mathbb{G}_{m-1},|u\wedge v|=m-l-1} \mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi\otimes\mathcal{Q}^l\psi)\big), \end{aligned}$$

obtaining the last term by rearranging the sum $u,v\in\mathbb{G}_{m-1}$ that expands over indices $|u|-|u\wedge v| = m-1-|u\wedge v|$ that vary from 1 to $m-1$. Notice also that for $l=1,\dots,m-1$,

$$\big|\{u\neq v\in\mathbb{G}_{m-1}, |u\wedge v| = m-l-1\}\big| = |\mathbb{G}_{m-l-1}|\times 2^l\times 2^{l-1},$$

where $|\mathbb{G}_{m-l-1}|$ is the number of common ancestors, 2^l is the number of choices for u (the first descendant in generation $m-1$ of the ancestor from generation $m-l-1$) and 2^{l-1} is the number choices of v (the second descendant, satisfying $u\wedge v\in\mathbb{G}_{m-1}$). We finally obtain

$$IV_m = 4\sum_{l=1}^{m-1} 2^{m+l-2}\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi\otimes\mathcal{Q}^l\psi)\big).$$

By Lemma 24 and Theorem 6 one easily obtains the following estimates:

$$|\mu\left(\mathcal{Q}^{m-2}\mathcal{P}\left(\mathcal{Q}\psi\otimes\mathcal{Q}\psi\right)\right)| \lesssim |\psi_\star|_1|\psi|_{\wedge 1} \quad \text{for} \quad l=1,$$

and for $l\geq 2$:

$$|\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}(\mathcal{Q}^l\psi\otimes\mathcal{Q}^l\psi)\big)| \lesssim |\psi|^2_{\wedge 1} \wedge \rho^{2(l-1)}|\psi_\star|^2_1\,\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}((1+V)\otimes(1+V))\big).$$

In the case $\nu(Q\psi)\neq 0$, we replace $|\psi|_{\wedge 1}$ by $|\psi|_\nu$. We claim that

$$\mu\big(\mathcal{Q}^{m-l-1}\mathcal{P}((1+V)\otimes(1+V))\big) \lesssim 1+\mu(V^2) \tag{19}$$

and postpone the proof of (19) to Step 6 below. It follows that for any $\ell\geq 1$:

$$\begin{aligned} |IV_m| &\lesssim 2^m|\psi_\star|_1|\psi|_{\wedge 1} + 2^m\sum_{l=1}^{\infty} 2^\ell\big(|\psi|^2_{\wedge 1}\wedge\rho^{2(l-1)}|\psi_\star|^2_1\big(1+\mu(V^2)\big)\big) \\ &\leq 2^m|\psi_\star|_1|\psi|_{\wedge 1} + 2^m\Big(|\psi|^2_{\wedge 1}\sum_{l=1}^{\ell}2^l + |\psi_\star|^2_1\big(1+\mu(V^2)\big)\sum_{l=\ell+1}^{\infty}2^{-l}(2\rho)^{2l}\Big) \\ &\leq 2^m|\psi_\star|_1|\psi|_{\wedge 1} + 2^m\inf_{\ell\geq 1}\Big(|\psi|^2_{\wedge 1}2^\ell + |\psi_\star|^2_1\big(1+\mu(V^2)\big)2^{-\ell}\Big) \\ &\lesssim 2^m\big(1+\mu(V)^{1/2}\big)|\psi_\star|_1|\psi|_{\wedge 1}. \end{aligned}$$

Step 5). Putting together the estimates obtained for I_m in Step 2, III_m in Step 3 and IV_m in Step 4, and recalling $II_m = III_m + IV_m$ we eventually derive:

$$\mathbb{E}_\mu\big[\big(\sum_{u\in\mathbb{G}_m}\psi(X_{u^-},X_u)\big)^2\big] \lesssim 2^m\big(|\psi^2|_\mu + |\psi^\star\psi|_\mu + \big(1+\mu(V)^{1/2}\big)|\psi_\star|_1|\psi|_{\wedge 1}\big).$$

In the case $\nu(Q\psi)\neq 0$, we replace $|\psi^2|_\mu$ by $|\psi^2|_{\mu+\nu}$ and $|\psi|_{\wedge 1}$ by $|\psi|_\nu$ as follows from Step 2 and 4. Taking square root, summing in $1\leq m\leq n$, taking square again and normalising by $|\mathbb{T}^\star_n|^{-2}$ we

obtain Theorem 7.

Step 6). It remains to establish (19). We only sketch the argument which is similar to the proof of Proposition 21. First, one easily obtains

$$\mathcal{P}((1+V)\otimes(1+V)) \lesssim 1+\mathcal{Q}V(x)+QV^2(x),$$

and it follows that

$$\begin{aligned}\mu(\mathcal{Q}^{m-l-1}\mathcal{P}\big((1+V)\otimes(1+V)\big)) &\lesssim 1+\mu(\mathcal{Q}^{m-l}V)+\mu(\mathcal{Q}^{m-l}V^2)\\ &\lesssim 1+\mu(V)+\mu(\mathcal{Q}^{m-l}V^2)\end{aligned}$$

by Proposition 21. Applying Itô's formula and using Assumptions 2 and 3 on can check that

$$\mathcal{Q}V^2(x) \lesssim 1+V(x)^2+\mathcal{Q}V(x) \lesssim 1+V(x)^2$$

by Proposition 21 again. We obtain (19) by integrating w.r.t. μ.

5.4. **Proof of Theorem 10.**

Preparations. We first establish local estimates on the invariant density ν.

Lemma 25. *Work under Assumptions 2, 3 and 4. Let $\mathcal{Q}\in\mathbb{Q}$ and let ν be the associated invariant density of Theorem 6. Let $x_0\in\mathcal{X}$. There exist positive constants $c_i=c_i(x_0,\mathcal{Q})$ and a bounded neighbourhood $\mathcal{V}_{x_0}$ with non-empty interior such that*

$$0<c_1\le \inf_{x\in\mathcal{V}_{x_0}}\nu(x)\le \sup_{x\in\mathcal{V}_{x_0}}\nu(x)\le c_2.$$

Moreover, $0<\inf_{\mathcal{Q}\in\mathbb{Q}}c_1(x_0,\mathcal{Q})\le\sup_{\mathcal{Q}\in\mathbb{Q}}c_2(x_0,\mathcal{Q})<\infty$.

Proof. Let $\mathcal{V}_{x_0}=[a,b]\subset\mathcal{X}$ be a bounded neighbourhood of x_0 and

$$\mathcal{V}_{x_0}^{\varepsilon}=\big[a/(1-\varepsilon)\wedge a/\varepsilon,\ b/\varepsilon\vee b/(1-\varepsilon)\big].$$

By Proposition 5, using Assumptions 2, 3 and 4, for every $x\in\mathcal{V}_{x_0}$, we successively obtain

$$\begin{aligned}\nu(x)&=\int_{\mathcal{X}}\nu(y)q(y,x)dy\\ &=\int_{\mathcal{X}}\nu(y)\int_0^1\frac{\kappa(z)}{z}B(x/z)\sigma(x/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_y(s))ds}dL_t^{x/z}(\phi_y)\Big]dzdy\\ &\le(b_2|x|^\gamma+b_1)\varepsilon^{-(1+\gamma)}\sigma_1^{-2}\int_{\mathcal{X}}\nu(y)\sup_{x\in\mathcal{V}_{x_0}^{\varepsilon}}\mathbb{E}\Big[\int_0^\infty e^{-b_1t}dL_t^x(\phi_y)\Big]dy\\ &=(b_2|x|^\gamma+b_1)\varepsilon^{-(1+\gamma)}\sigma_1^{-2}b_1\int_{\mathcal{X}}\nu(y)\int_0^\infty e^{-b_1t}\sup_{x\in\mathcal{V}_{x_0}^{\varepsilon}}\mathbb{E}\big[L_t^x(\phi_y)\big]dtdy.\end{aligned}$$

By Lemma 23, we have $\sup_{x\in\varepsilon\mathcal{V}_{x_0}}\mathbb{E}\big[L_t^x(\phi_y)\big]\lesssim 1+t^{3/2}$ uniformly over $\mathbb{Q}$ and the first part of the lemma easily follows. For the second part of the lemma, we have

$$\begin{aligned}\nu(x)&\ge b_1\sigma_2^{-2}\int_{\mathcal{X}}\nu(y)\inf_{x\in\mathcal{V}_{x_0}^{\varepsilon}}\mathbb{E}\Big[\int_0^\infty e^{-b_2\int_0^t|\phi_y(s)|^\gamma ds-b_1t}dL_t^x(\phi_y)\Big]dy\\ &\ge b_1\sigma_2^{-2}\int_{[-N,N]\cap\mathcal{X}}\nu(y)\inf_{x\in\mathcal{V}_{x_0}^{\varepsilon}}\mathbb{E}\Big[\int_0^T e^{-(b_2M^\gamma+b_1)t}dL_t^x(\phi_y)\mathbf{1}_{\{\sup_{t\le T}|\phi_y(t)|\le M\}}\Big]dy\\ &\ge b_1\sigma_2^{-2}e^{-(b_2M^\gamma+b_1)T}\int_{[-N,N]\cap\mathcal{X}}\nu(y)\inf_{x\in\mathcal{V}_{x_0}^{\varepsilon}}\mathbb{E}\big[L_T^x(\phi_y)\mathbf{1}_{\{\sup_{t\le T}|\phi_y(t)|\le M\}}\big]dy\end{aligned}$$

for arbitrary constants $M, N, T > 0$. Since $\mathbb{E}\big[L_T^x(\phi_y)\mathbf{1}_{\{\sup_{t\le T}|\phi_y(t)|\le M\}}\big] \uparrow \mathbb{E}\big[L_T^x(\phi_y)\big]$ uniformly in $(x,y) \in \mathcal{V}_{x_0} \times [-N,N]$ as M grows, pick M large enough so that for every $y \in [-N,N] \cap \mathcal{X}$, we have

$$\inf_{x\in\mathcal{V}_{x_0}^\varepsilon} \mathbb{E}\big[L_T^x(\phi_y)\mathbf{1}_{\{\sup_{t\le T}|\phi_y(t)|\le M\}}\big] \ge \tfrac{1}{2} \inf_{x\in\mathcal{V}_{x_0}^\varepsilon} \mathbb{E}\big[L_T^x(\phi_y)\big].$$

Next, we use the fact that Assumption 2 implies that the law of the random variable $\phi_y(t)$ admits a density $\rho_t(y,x)$ w.r.t. the Lebesgue measure and that this density is bounded away from zero on compact sets in (y,x), see for instance [11, 16]. In turn $\mathbb{E}\big[L_T^x(\phi_y)\big] = \int_0^T \rho_t(y,x)dt \ge \tau_T > 0$ for some τ_T depending also on M and N and we infer

$$\nu(x) \ge \frac{\tau_T}{2} b_1 \sigma_2^{-2} e^{-(b_2 M^\gamma + b_1)T} \int_{[-N,N]\cap\mathcal{X}} \nu(y)dy$$

and we obtain the result by taking N sufficiently large. The proof is complete. □

Completion of proof of Theorem 10. Step 1). Write $\widehat{\nu}_n(y_0) - \nu(y_0) = I + II$, with

$$I = \frac{1}{|\mathbb{T}_n|} \sum_{u\in\mathbb{T}_n} G_h(X_u - y_0) - \nu(\mathcal{Q}G_h(\cdot - y_0)) \quad \text{and} \quad II = G_h \star \nu(y_0) - \nu(y_0).$$

We plan to apply Theorem 7 to I with $\psi(x,y) = \varphi(y) = G_h(y - y_0)$. By Lemma 25, ν is locally bounded and we easily check that

$$|\psi^2|_{\mu+\nu} \lesssim \int_{\mathcal{X}} |G_h(y-y_0)^2|dy = h^{-1}\int_{\mathcal{X}} G(y)^2 dy \lesssim h^{-1},$$

$$|\psi^\star\psi|_\mu \lesssim \sup_y |G_h(y)| \int_{\mathcal{X}} |G_h(y-y_0)|dy \lesssim h^{-1},$$

and

$$|\psi_\star|_1 |\psi|_{\wedge 1} \lesssim \Big(\int_{\mathcal{X}} |G_h(y-y_0)|dy\Big)^2 \lesssim 1.$$

Therefore, by Theorem 7, we have $\mathbb{E}_\mu[I^2] \lesssim |\mathbb{T}_n|^{-1}h^{-1}$ and this term is of order $|\mathbb{T}_n|^{-2\beta/(2\beta+1)}$ from the choice of h. For the term II, Lemma 25 and the representation $\nu(x) = \int_{\mathcal{X}} \nu(y)q(y,x)dy$ show that $\nu \in \mathcal{H}^\beta(y_0)$ as soon as $q \in \mathcal{H}^{\alpha,\beta}(x_0,y_0)$. Then, by classical kernel approximation (see *e.g.* the textbook by Tsybakov [33]) we have that $II^2 \lesssim h^{2\beta}$ since the order k of the kernel G satisfies $k > \beta$, and thus II^2 the same order as I^2 from the choice of h.

Step 2). For the estimation of $q(x_0,y_0)$, write

$$\widehat{q}_n(x_0,y_0) - q(x_0,y_0) = I + II,$$

with

$$I = \frac{\mathcal{M}_n\big(G_{h_1,h_2}^{\otimes 2}(\cdot - x_0, \cdot - y_0)\big) - \nu(x_0)q(x_0,y_0)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n},$$

and

$$II = \frac{q(x_0,y_0)\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n\big)}{\mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n}.$$

We have $\big|I\big| \le III + IV$, with

$$III = \varpi_n^{-1}\big|\mathcal{M}_n\big(G_{h_1,h_2}^{\otimes 2}(\cdot - x_0, \cdot - y_0)\big) - G_{h_1,h_2}^{\otimes 2} \star \nu(x_0)q(x_0,y_0)\big|$$

and

$$IV = \varpi_n^{-1}\big|G_{h_1,h_2}^{\otimes 2} \star \nu(x_0)q(x_0,y_0) - \nu(x_0)q(x_0,y_0)\big|.$$

We plan to apply Theorem 7 to bound III with $\psi(x,y) = G_{h_1,h_2}(x-x_0, y-y_0)$. Using Lemma 25 and the fact that μ is absolutely continuous, we have $|\psi|_{\mu+\nu} \lesssim |\psi|_1$. It readily follows that

$$|\psi^2|_{\mu+\nu} \lesssim |G_{h_1}(\cdot - x_0)^2|_1 |G_{h_2}(\cdot - y_0)^2|_1 \lesssim h_1^{-1}h_2^{-1},$$

$$\begin{aligned}|\psi^\star\psi|_\mu &\lesssim |G_{h_1}(\cdot - x_0)\sup_y |G_{h_2}(y-y_0)|G_{h_1}(\cdot - x_0)G_{h_2}(\cdot - y_0)|_1 \\ &= |G_{h_1}(\cdot - x_0)^2|_1 \sup_y |G_{h_2}(y-y_0)||G_{h_2}(\cdot - y_0)|_1 \lesssim h_1^{-1}h_2^{-1},\end{aligned}$$

and

$$|\psi_\star|_1|\psi|_1 \lesssim \sup_x |G_{h_1}(x-x_0)||G_{h_2}(\cdot - y_0)|_1^2 |G_{h_1}(\cdot - x_0)|_1 \lesssim h_1^{-1}.$$

We conclude

$$\mathbb{E}_\mu\big[III^2\big] \lesssim \varpi_n^{-2}|\mathbb{T}_n|^{-1}h_1^{-1}h_2^{-1},$$

and this term has order $\varpi_n^{-2}|\mathbb{T}_n|^{-2s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$ from the choice of h_1 and h_2. By kernel approximation and the fact that G has order $k > \max\{\alpha,\beta\}$, noting that $(x,y) \mapsto \mu(x)q(x,y) \in \mathcal{H}^{\alpha\wedge\beta,\beta}$, we have

$$|IV| \lesssim h_1^{\alpha\wedge\beta} + h_2^\beta \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-s(\alpha,\beta)/(2s(\alpha,\beta)+1)}$$

from the choice of h_1, h_2.

We turn to the term II. We plan to use

$$\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n\big)^2 \lesssim \big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big)\big)^2 + \mathbf{1}_{\{\mathcal{M}_n(G_h(\cdot - x_0)) < \varpi_n\}}.$$

Pick n large enough so that $0 < \varpi_n \le \tau(x_0) = \frac{1}{2}\inf_{\mathcal{Q}\in\mathbb{Q}, x\in\mathcal{V}_{x_0}} \nu(x)$, a choice which is possible by Lemma 25. Since $\{\mathcal{M}_n(G_h(\cdot - x_0)) < \varpi_n\} \subset \{\mathcal{M}_n(G_h(\cdot - x_0)) - \nu(x_0) < -\tau(x_0)\}$, we further infer

$$\begin{aligned}&\mathbb{E}_\mu\big[\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big) \vee \varpi_n\big)^2\big] \\ &\le \mathbb{E}_\mu\big[\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big)\big)^2\big] + \mathbb{P}_\mu\big(\big|\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big)\big| \ge \tau(x_0)\big) \\ &\lesssim \mathbb{E}_\mu\big[\big(\nu(x_0) - \mathcal{M}_n\big(G_h(\cdot - x_0)\big)\big)^2\big].\end{aligned}$$

Applying *Step 1)* of the proof, we derive

$$\mathbb{E}_\mu\big[II^2\big] \lesssim \varpi_n^{-2}|\mathbb{T}_n|^{-2\beta/(2\beta+1)}$$

and this term has negligible order. The proof of Theorem 10 is complete.

5.5. **Proof of Proposition 14.** Let $s(x) = \int_0^x \exp\Big(-2\int_0^y \frac{r(z)}{\sigma(z)^2}dz\Big)$ and $m(x) = \frac{2}{\sigma(x)^2 s'(x)}$. Consider the infinitesimal generator $\mathcal{L}$ associated to the diffusion process (11), written in its divergence form

$$\mathcal{L}f(x) = \frac{1}{m(x)}\frac{d}{dx}\Big(\frac{1}{s(x)}\frac{d}{dx}f(x)\Big), \quad f \in \mathcal{D}(\mathcal{L}),$$

with domain $\mathcal{D}(\mathcal{L})$ densely defined on twice continuously differentiable functions f satisfying the boundary condition $f'(0) = f'(L) = 0$. By Itô formula and Fubini's theorem, for $f \in \mathcal{D}(\mathcal{L})$, we have

$$\begin{aligned}\int_{\mathcal{X}} f(y)q(x,y)dy &= -\int_0^1 \kappa(z)\mathbb{E}\Big[\int_0^\infty f\left(z\phi_x(t)\right)\frac{d}{dt}e^{-\int_0^t B(\phi_x(s))ds}dt\Big]dz \\ &= \int_0^1 \kappa(z)f_z(x)dz + \int_0^1 \kappa(z)\mathbb{E}\Big[\int_0^\infty \mathcal{L}f_z\left(\phi_x(t)\right)e^{-\int_0^t B(\phi_x(s))ds}dt\Big]dz,\end{aligned}$$

where we set $f_z(x) = f(zx)$ for $z \in [\varepsilon, 1-\varepsilon]$ since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$ by Assumption 4. Pick $f(x) = \int_0^x \exp\big(2\varepsilon^{-1}\int_0^y e^{2\int_0^{u\varepsilon^{-1}}\frac{|r(v)|}{\sigma(v)^2}dv}s(u)^{-1}du\big)dy$, and note that

$$f''(zx) = \frac{2}{\varepsilon s(zx)}\exp\Big(2\int_0^{zx\varepsilon^{-1}}\frac{|r(v)|}{\sigma(v)^2}dv\Big)f'(zx).$$

It follows that for $z \geq \varepsilon$ and every $x \in \mathcal{X}$, we have

$$\begin{aligned}\mathcal{L}f_z(x) &= \frac{z}{m(x)}\frac{zs(x)f''(zx) - s'(x)f'(zx)}{s^2(x)} \\ &= \frac{zf'(zx)}{m(x)s(x)^2}\Big(\frac{2zs(x)}{\varepsilon s(zx)}e^{2\int_0^{zx\varepsilon^{-1}}\frac{|r(v)|}{\sigma(v)^2}dv} - e^{-2\int_0^x\frac{r(v)}{\sigma(v)^2}dv}\Big) > 0.\end{aligned}$$

Now let $B_1, B_2 : \mathcal{X} \to [0,\infty)$ be two functions in an orderly class $\mathcal{B}$ according to Definition 13 and write q_{B_1} and q_{B_2} for the associated transition densities. With no loss of generality, we may (and will) assume that $B_1(x) \leq B_2(x)$ for every $x \in \mathcal{X}$. Assume that $q_{B_1} = q_{B_2}$. Since $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$, we have

$$\begin{aligned}&\int_{\mathcal{X}} f(y)\big(q_{B_1}(x,y) - q_{B_2}(x,y)\big)dy \\ &= \int_\varepsilon^{1-\varepsilon}\kappa(z)\mathbb{E}\Big[\int_0^\infty \mathcal{L}f_z(\phi_x(t))\big(e^{-\int_0^t B_1(\phi_x(s))ds} - e^{-\int_0^t B_2(\phi_x(s))ds}\big)\Big]dtdz = 0.\end{aligned}$$

Our choice of f and the property $B_1 \leq B_2$ implies that the integrand is non-negative. It follows that

$$\kappa(z)\big(e^{-\int_0^t B_1(\phi_x(s))ds} - e^{-\int_0^t B_2(\phi_x(s))ds}\big) = 0$$

$dzdt \otimes \mathbb{P}$-a.s. Picking z such that $\kappa(z) > 0$, we obtain $\int_0^t B_1(\phi_x(s))ds = \int_0^t B_2(\phi_x(s))ds$ $\mathbb{P}$-a.s. for every $t \geq 0$ by continuity of the integrand in t. By the occupation times formula, it follows that $\int_X \big(B_1(y) - B_2(y)\big)L_t^y(\phi_x)dy = 0$, almost-surely, and by the ordering property, $B_1(y) = B_2(y)$ for every y such that $L_t^y(\phi_x) > 0$, *i.e.* for $y \in [\inf_{0\leq s\leq t}\phi_x(s), \sup_{0\leq s\leq t}\phi_x(s)] \to \mathcal{X}$ as $t \to \infty$. The proof of Proposition 14 is complete.

5.6. **Proof of Theorem 16.**

Preparation for the proof. We first establish uniform bounds for $q_\vartheta(x,y)$. Remember that in the reflected case, we have $\mathcal{X} = [0, L]$ and $\mathrm{supp}(\kappa) \subset [\varepsilon, 1-\varepsilon]$ under Assumption 4.

Lemma 26. *Work under Assumptions 2, 4 and 15. For sufficiently small $\eta > 0$, we have:*

$$0 < \inf_{x\in\mathcal{X}, y\in\mathcal{X}_\eta, \vartheta\in\Theta} q_\vartheta(x,y) \leq \sup_{x,y\in\mathcal{X},\vartheta\in\Theta} q_\vartheta(x,y) < \infty,$$

where $\mathcal{X}_\eta = [0, (1-\varepsilon)L - \eta]$.

Proof. The proof is close to that of Lemma 25. Let $x \in \mathcal{X}$ and $y \in \mathcal{X}_\eta$. We have

$$\begin{aligned}\inf_{\vartheta\in\Theta} q_\vartheta(x,y) &\geq b_3\sigma_2^{-2}\int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon}\frac{\kappa(z)}{z}\mathbb{E}\big[\int_0^\infty e^{-b_4 t}dL_t^{y/z}(\phi_x)\big]dz\\ &\geq b_3\sigma_2^{-2}b_4\int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon}\frac{\kappa(z)}{z}\int_0^\infty e^{-b_4 t}\mathbb{E}\big[L_t^{y/z}(\phi_x)\big]dtdz\\ &\geq (1-\varepsilon)^{-1}b_3\sigma_2^{-2}b_4\frac{\eta}{L}\int_0^\infty e^{-b_4 t}\inf_{x,y\in\mathcal{X}}\mathbb{E}\big[L_t^y(\phi_x)\big]dt\end{aligned}$$

According to [10], Section 5, proof of Lemma 5.37, the law of $\phi_x(t)$ is absolutely continuous with density with $y \mapsto \rho_t(x,y)$ that can be taken continuous and that satisfies $\inf_{x,y\in\mathcal{X}}\rho_t(x,y) > 0$ for every $t > 0$. Therefore

$$\inf_{x,y\in\mathcal{X}}\mathbb{E}\big[L_t^y(\phi_x)\big] = \inf_{x,y\in\mathcal{X}}\int_0^t \rho_s(x,y)ds \geq \tau_t > 0$$

and the result follows. The upper bound readily follows from

$$\begin{aligned}\sup_{\vartheta\in\Theta} q_\vartheta(x,y) &\leq b_4\sigma_1^{-2}\int_\varepsilon^{1-\varepsilon}\frac{\kappa(z)}{z}dz\sup_{x,y\in\mathcal{X}}\mathbb{E}\big[\int_0^\infty e^{-b_3 t}dL_t^y(\phi_x)\big]\\ &\leq \varepsilon^{-1}b_4\sigma_1^{-2}\sup_{x,y\in\mathcal{X}}\mathbb{E}\big[\int_0^\infty e^{-b_3 t}dL_t^y(\phi_x)\big]\\ &= \varepsilon^{-1}b_4\sigma_1^{-2}b_3\sup_{x,y\in\mathcal{X}}\int_0^\infty e^{-b_3 t}\mathbb{E}\big[L_t^y(\phi_x)\big]dt\end{aligned}$$

which is finite by Lemma 23. □

Completion of proof of Theorem 16. This proof is classical (see for instance van der Vaart [34] Theorem 5.14). We nevertheless give it for self-containedness. For $a \in \Theta$, let

$$\mathcal{M}(a,\vartheta) = \int_\mathcal{X}\nu_\vartheta(dx)\int_\mathcal{X}\log q_a(x,y)q_\vartheta(x,y)dy.$$

First, $a \mapsto \mathcal{M}(a,\vartheta)$ has a unique maximum at $a = \vartheta$, as stems from the inequality $\log(x) \leq 2(\sqrt{x}-1)$ for $x \geq 0$. Indeed

$$\begin{aligned}\mathcal{M}(a,\vartheta) - \mathcal{M}(\vartheta,\vartheta) &= \int_\mathcal{X}\nu_\vartheta(dx)\int_\mathcal{X}\log\frac{q_a(x,y)}{q_\vartheta(x,y)}q_\vartheta(x,y)dy\\ &\leq \int_\mathcal{X}\nu_\vartheta(dx)\Big(\int_\mathcal{X} 2\sqrt{q_a(x,y)}\sqrt{q_\vartheta(x,y)}dy - 2\Big)\\ &\leq -\int_\mathcal{X}\nu_\vartheta(dx)\int_\mathcal{X}\Big(\sqrt{q_a(x,y)}-\sqrt{q_\vartheta(x,y)}\Big)^2 dy \leq 0.\end{aligned}$$

Next, writing $m_\mathcal{U}(x,y) = \sup_{a\in\mathcal{U}}\log q_a(x,y)$, we prove that for every $a \neq \vartheta \in \Theta$, there exists a neighbourhood $\mathcal{U}_a$ of a such that:

$$\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_a}\big) < \nu_\vartheta\big(\mathcal{Q}_\vartheta\log q_\vartheta\big) = \mathcal{M}(\vartheta,\vartheta). \tag{20}$$

Pick a decreasing sequence of open balls $(\mathcal{U}_\ell(a))_{\ell\geq 1}$ around a with vanishing diameters. For every $x, y \in \mathcal{X}$ we have $m_{\mathcal{U}_\ell(a)}(x,y) \downarrow \log q_a(x,y)$ by continuity of $a \mapsto \log q_a(x,y)$ thanks to the

continuity of B_0 according to Assumption 15. By Lemma 26, we also have $\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}}\big) < \infty$ for any $\mathcal{U} \subset \Theta$ therefore

$$\nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_l(a)}\big) \downarrow \nu_\vartheta\big(\mathcal{Q}_\vartheta \log q_a\big) = \mathcal{M}(a,\vartheta) \leq \mathcal{M}(\vartheta,\vartheta)$$

by monotone convergence with equality only if $a = \vartheta$, and this proves the existence of $\mathcal{U}_a$ such that (20) holds. We are now ready to prove the consistency result. For $\eta' > 0$, the compact ball

$$\mathcal{C}_{\eta'}(\vartheta) = \{a \in \Theta, |a - \vartheta| \geq \eta'\}$$

can be covered by finitely many open neighbourhoods $\mathcal{U}_{a_1}, \dots \mathcal{U}_{a_p}$ with $a_i \in \mathcal{C}_{\eta'}(\vartheta)$ and such that (20) holds for every $\mathcal{U}_{a_i}$. For $\eta > 0$, let

$$m_{\mathcal{U}}^{(\eta)}(x,y) = \sup_{a\in\mathcal{U}} \log q_a(x,y) \mathbf{1}_{\{q_a(x,y)\geq\eta\}}.$$

Abbreviating $\mathcal{L}_n(a,(X_u)_{u\in\mathbb{T}_n})$ by $\mathcal{L}_n(a)$, it follows that

$$\begin{aligned} |\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \log \mathcal{L}_n(a) &\leq \max_{1\leq i\leq p} |\mathbb{T}_n^\star|^{-1} \sum_{u\in\mathbb{T}_n^\star} m_{\mathcal{U}_{a_i}}^{(\eta)}(X_{u^-}, X_u) \\ &\to \max_{1\leq i\leq p} \nu_\vartheta\big(\mathcal{Q}_\vartheta m_{\mathcal{U}_{a_i}}\big) < \mathcal{M}(\vartheta,\vartheta) \end{aligned} \tag{21}$$

in probability as $n \to \infty$ and letting $\eta \to 0$, as stems from Corollary 11 and the fact that $\sup_{x,y\in\mathcal{X}_\eta} m_{\mathcal{U}_{a_i}}(x,y) < \infty$ by Lemma 26. Finally, if $\widehat{\vartheta}_n \in \mathcal{C}_{\eta'}(\vartheta)$, then, by definition of $\widehat{\vartheta}_n$, we have

$$|\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \log \mathcal{L}_n(a) \geq |\mathbb{T}_n^\star|^{-1} \log \mathcal{L}_n(\widehat{\vartheta}_n) \geq |\mathbb{T}_n^\star|^{-1} \log \mathcal{L}_n(\vartheta) > \mathcal{M}(\vartheta,\vartheta) - \epsilon_n,$$

where $\epsilon_n \to 0$ in probability, as follows from Corollary 11. We conclude the proof by noticing that

$$\big\{\widehat{\vartheta}_n \in \mathcal{C}_{\eta'}(\vartheta)\big\} \subset \big\{|\mathbb{T}_n^\star|^{-1} \sup_{a\in\mathcal{C}_{\eta'}(\vartheta)} \mathcal{L}_n(a) \geq \mathcal{M}(\vartheta,\vartheta) - \epsilon_n\big\}$$

and the fact that the probability of this last event converges to 0 by (21) as $n \to \infty$.

5.7. **Proof of Theorem 20.**

Preparation for the proof. We start by proving some useful estimates on the gradient and Hessian of $\log q_\vartheta$. Let

$$\Gamma_\vartheta = \nabla_\vartheta \log q_\vartheta = (\partial_{\vartheta_1} \log q_\vartheta, \dots, \partial_{\vartheta_d} \log q_\vartheta), \quad \Gamma_{\vartheta,i} = \partial_{\vartheta_i} \log q_\vartheta, \quad 1 \leq i \leq d.$$

Lemma 27. *Work under Assumptions 2, 4, 15 and 17. For every $1 \leq i,j \leq d$ and $\eta > 0$, we have*

$$\sup_{x\in\mathcal{X},y\in\mathcal{X}_\eta,\vartheta\in\Theta} |\Gamma_{\vartheta,i}(x,y)| < \infty, \sup_{x\in\mathcal{X},y\in\mathcal{X}_\eta,\vartheta\in\Theta} |\partial_\vartheta \Gamma_\vartheta(x,y)_{i,j}| < \infty, \sup_{x\in\mathcal{X},y\in\mathcal{X}_\eta,\vartheta\in\Theta} \big\|\partial_\vartheta^2 \Gamma_\vartheta(x,y)\big\| < \infty$$

where $\|\cdot\|$ corresponds to the operator norm for the Hessian $\partial_\vartheta^2 \Gamma_\vartheta(x,y)$.

Proof. According to Lemma 26, since

$$\Gamma_\vartheta(x,y) = \frac{\partial_\vartheta q_\vartheta(x,y)}{q_\vartheta(x,y)}$$

componentwise, it suffices to show $|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim 1$ in order to establish the first bound. We have:

$$\partial_{\vartheta_i} q_\vartheta(x,y) = \int_{\varepsilon \vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z} \partial_{\vartheta_i} B_0(\vartheta, y/z)\sigma(y/z)^{-2}\,\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]dz$$
$$- \int_{\varepsilon \vee yL^{-1}}^{1-\varepsilon} \frac{\kappa(z)}{z} B_0(\vartheta, y/z)\sigma(y/z)^{-2}\,\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} \int_0^t \partial_{\vartheta_i} B_0(\vartheta,\phi_x(s))ds\, dL_t^{y/z}(\phi_x)\Big].$$

By Assumption 2, 15 and 17 we readily obtain

$$|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim \int_\varepsilon^{1-\varepsilon} \mathbb{E}\Big[\int_0^\infty (1+t) e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]\kappa(z)dz$$

By Assumption 17 and integration by part, we have

$$\mathbb{E}\Big[\int_0^\infty (1+t)e^{-\int_0^t B_0(\vartheta,\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big] \le \mathbb{E}\Big[\int_0^\infty (1+t)e^{-b_3 t} dL_t^{y/z}(\phi_x)\Big]$$
$$= \int_0^\infty (1 - b_3(1+t))e^{-b_3 t}\mathbb{E}[L_t^{y/z}(\phi_x)]dt.$$

This last term is bounded by Lemma 23 and $|\partial_{\vartheta_i} q_\vartheta(x,y)| \lesssim 1$ follows. We turn to the second bound: clearly, for $1 \le i,j \le d$

$$\partial_\vartheta \Gamma_\vartheta(x,y)_{i,j} = \frac{\partial^2_{\vartheta_i\vartheta_j} q_\vartheta(x,y) q_\vartheta(x,y) - \partial_{\vartheta_i} q_\vartheta(x,y)\partial_{\vartheta_j} q_\vartheta(x,y)}{q_\vartheta(x,y)^2}$$

and thanks to Lemma 26 and the first bound, we only need to show $|\partial^2_{\vartheta_i\vartheta_j} q_\vartheta(x,y)| \lesssim 1$ in order to obtain the second bound. Define $\omega_t(y,z,\vartheta) = \mathbf{1}_{\{\varepsilon\vee yL^{-1}, 1-\varepsilon\}}(z)\frac{\kappa(z)}{z\sigma(y/z)^2}\exp(-\int_0^t B_0(\vartheta,\phi_x(s))ds)$. We have

$$\partial^2_{\vartheta_i\theta_j} q_\vartheta(x,y) = \int_0^1 \partial^2_{\vartheta_i\vartheta_j} B_0(\vartheta, y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta) dL_t^{y/z}(\phi_x)\Big]dz$$
$$+ \int_0^1 B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial_{\vartheta_i} B_0(\vartheta,\phi_x(s))ds \int_0^t \partial_{\vartheta_j} B_0(\vartheta,\phi_x(s))ds\, dL_t^{y/z}(\phi_x)\Big]dz$$
$$- \sum_{(\ell,\ell')=(i,j),(j,i)} \int_0^1 \partial_{\theta_\ell} B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial_{\theta_{\ell'}} B_0(\vartheta,\phi_x(s))ds dL_t^{y/z}(\phi_x)\Big]dz$$
$$+ \int_0^1 B_0(\vartheta,y/z)\mathbb{E}\Big[\int_0^\infty \omega_t(y,z,\vartheta)\int_0^t \partial^2_{\vartheta_i\vartheta_j} B_0(\vartheta,\phi_x(s))ds dL_t^{y/z}(\phi_x)\Big]dz$$

and we proceed in the same way as for the first estimate, using repeatedly Assumption 2, 15 and 17. The proof of the third bound is analogous. □

Completion of proof of Theorem 20. This proof is classical (see for instance van der Vaart [34] Theorem 5.41). We nevertheless give it for self-containedness. By definition of $\widehat{\vartheta}_n$ and a Taylor expansion around ϑ, we have

$$0 = \sum_{u\in\mathbb{T}_n^\star} \Gamma_{\widehat{\vartheta}_n}(X_{u^-}, X_u)$$
$$= \sum_{u\in\mathbb{T}_n^\star} \Big(\Gamma_\vartheta(X_{u^-},X_u) + \partial_\vartheta\Gamma_\vartheta(X_{u^-},X_u)(\widehat{\vartheta}_n - \vartheta) + (\widehat{\vartheta}_n - \vartheta)^T \partial^2_\vartheta \Gamma_{\widetilde{\vartheta}_n}(X_{u^-},X_u)(\widehat{\vartheta}_n - \vartheta)\Big),$$

for some $\widetilde{\vartheta}_n$ on the segment line between ϑ and $\widehat{\vartheta}_n$. Rearranging the sum and introducing the normalisation $|\mathbb{T}_n^\star|^{1/2}$, we derive

$$\begin{aligned}(22)\qquad &\big(|\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}\partial_\vartheta\Gamma_\vartheta(X_{u^-},X_u)+|\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}(\widehat{\vartheta}_n-\vartheta)^T\partial^2_\vartheta\Gamma_{\widetilde{\vartheta}_n}(X_{u^-},X_u)\big)|\mathbb{T}_n^\star|^{1/2}(\widehat{\vartheta}_n-\vartheta)\\ &=-|\mathbb{T}_n^\star|^{-1/2}\sum_{u\in\mathbb{T}_n^\star}\Gamma_\vartheta(X_{u^-},X_u).\end{aligned}$$

We plan to apply the central limit theorem for bifurcating Markov chain proved by Guyon, see [17] Corollary 24 on the right-hand side. By Lemma 26 and 27 we have that $\mathcal{Q}_\vartheta(\Gamma_{\vartheta,i}\Gamma_{\vartheta,j})$ and $\mathcal{Q}_\vartheta(\Gamma_{\vartheta,i}\Gamma_{\vartheta,j}\Gamma_{\vartheta,k}\Gamma_{\vartheta,l})$ are bounded functions on $\mathcal{X}$ for all $1\le i,j,k,l\le d$. Moreover, we have $\nu_\vartheta(\mathcal{Q}_\vartheta\Gamma_{\vartheta,i})=0$. Therefore

$$(23)\qquad |\mathbb{T}_n^\star|^{-1/2}\sum_{u\in\mathbb{T}_n^\star}\Gamma_\vartheta(X_{u^-},X_u)\to\mathcal{N}\big(0,\Psi(\vartheta)\big)$$

in distribution as $n\to\infty$, where $\Psi(\vartheta)$ is the Fisher information matrix defined after Assumption 17. Next, since $\partial_\vartheta\Gamma_\vartheta$ is bounded by Lemma 27, we have

$$(24)\qquad |\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}\partial_\vartheta\Gamma_\vartheta(X_{u^-},X_u)\to\Psi(\vartheta)$$

in probability as $n\to\infty$. Moreover, by Lemma 27, we have: $\sup_{x,y\in\mathcal{X},\vartheta\in\Theta}\big\|\partial^2_\vartheta\Gamma_\vartheta(x,y)\big\|<\infty$ and since $\widehat{\vartheta}_n-\vartheta$ converges to 0 by Theorem 16, it follows that

$$(25)\qquad |\mathbb{T}_n^\star|^{-1}\sum_{u\in\mathbb{T}_n^\star}(\widehat{\vartheta}_n-\vartheta)^T\partial^2_\theta\Gamma_{\widetilde{\vartheta}_n}(X_{u^-},X_u)\to 0$$

in probability as $n\to\infty$ tends to infinity. Combining (23), (24) and (25) in (22) we finally obtain

$$\Psi(\vartheta)|\mathbb{T}_n^\star|^{1/2}(\widehat{\vartheta}_n-\vartheta)\to\mathcal{N}\big(0,\Psi(\vartheta)\big)$$

in distribution as $n\to\infty$. We conclude thanks to the invertibility of $\Psi(\vartheta)$ granted by Assumption 18.

6. Appendix

6.1. **Proof of Lemma 23.** *Step 1).* Fix $\delta>0$ and let $\mathcal{K}_\delta=\{y\in\mathcal{X},\inf_{z\in\mathcal{K}}|y-z|\le\delta\}$ denote the δ-enlargement of $\mathcal{K}$. For $x\in\mathcal{X}$, let

$$\tau_x=\inf\{t\ge 0,\phi_x(t)\in\mathcal{K}_\delta\},\quad \inf\emptyset=\infty,$$

and

$$\phi_x^{\mathcal{K}_\delta}(t)=\begin{cases}\phi_{\sup\mathcal{K}_\delta}\big((t-\tau_x)_+\big) & \text{if}\quad x>\sup\mathcal{K}_\delta\\ \phi_x(t) & \text{if}\quad x\in\mathcal{K}_\delta\\ \phi_{\inf\mathcal{K}_\delta}\big((t-\tau_x)_+\big) & \text{if}\quad x<\inf\mathcal{K}_\delta.\end{cases}$$

For every $y\in\mathcal{K}$, we have $L_t^y(\phi_x)=L_t^y(\phi_x^{\mathcal{K}_\delta})$, and by Itô-Tanaka's formula, it follows that

$$L_t^y(\phi_x)=L_t^y(\phi_x^{\mathcal{K}_\delta})=\big|\phi_x^{\mathcal{K}_\delta}(t)-y\big|-|\phi_x^{\mathcal{K}_\delta}(0)-y|-\int_0^t\operatorname{sgn}\big(\phi_x^{\mathcal{K}_\delta}(s)-y\big)d\phi_x^{\mathcal{K}_\delta}(s).$$

Assume first that $x > \sup \mathcal{K}_\delta$. Observing that $L_t^y(\phi_x) = 0$ on $\{\tau_x \geq t\}$, and that $d\phi_x^{\mathcal{K}_\delta}(s)$ vanishes on $[0, \tau_x)$ on $\{\tau_x < t\}$, we readily have

$$L_t^y(\phi_x) = \big|\phi_x^{\mathcal{K}_\delta}(t) - y\big| - |\sup \mathcal{K}_\delta - y| - \int_{\tau_x \wedge t}^t \operatorname{sgn}\big(\phi_x^{\mathcal{K}_\delta}(s) - y\big) d\phi_x^{\mathcal{K}_\delta}(s)$$

$$= \big|\phi_{\sup \mathcal{K}_\delta}\big((t-\tau_x)_+\big) - y\big| - |\sup \mathcal{K}_\delta - y| - \int_0^{(t-\tau_x)_+} \operatorname{sgn}\big(\phi_{\sup \mathcal{K}_\delta}(s) - y\big) d\phi_{\sup \mathcal{K}_\delta}(s). \tag{26}$$

We plan to bound each term separately.

Step 2). By Itô's formula, $(\phi_{\sup \mathcal{K}_\delta}(t) - y)^2 = (\sup \mathcal{K}_\delta - y)^2 + I + II$, with

$$I = \int_0^t \big(2(\phi_{\sup \mathcal{K}_\delta}(s) - y) r(\phi_{\sup \mathcal{K}_\delta}(s)) + \sigma(\phi_{\sup \mathcal{K}_\delta}(s))^2\big) ds,$$
$$II = 2\int_0^t (\phi_{\sup \mathcal{K}_\delta}(s) - y)\sigma(\phi_{\sup \mathcal{K}_\delta}(s)) dW_s.$$

First,

$$I \leq 2\sigma_1^{-2} \int_{\mathbb{R}} (z-y) r(z) L_t^z(\phi_{\sup \mathcal{K}_\delta}) dz + t\sigma_2^2$$

by the occupation times formula and Assumption 2. Introduce $|y|_{r_2} = |y| \vee r_2$, where r_2 is defined in Assumption 2. Since $z - y > 0$ and $r(z) < 0$ for $z > |y|_{r_2}$, we have $\int_{|y|_{r_2}}^\infty (z-y) r(z) L_t^z(\phi_x) dz < 0$. Similarly $\int_{-\infty}^{-|y|_{r_2}} (z-y) r(z) L_t^z(\phi_x) dz < 0$. It follows that

$$\begin{aligned}
\int_{\mathbb{R}} (z-y) r(z) L_t^z(\phi_x) dz &\leq \int_{-|y|_{r_2}}^{|y|_{r_2}} (z-y) r(z) L_t^z(\phi_x) dz \\
&\leq r_1 \int_{-|y|_{r_2}}^{|y|_{r_2}} |z-y|(1+|z|) L_t^z(\phi_x) dz \\
&\leq r_1 (|y|_{r_2} - y)(1 + |y|_{r_2}) \int_{\mathbb{R}} L_t^z(\phi_x) dz \\
&\leq r_1 (|y|_{r_2} - y)(1 + |y|_{r_2}) t,
\end{aligned}$$

therefore

$$I \leq 2\sigma_1^{-2} r_1 (|y|_{r_2} - y)(1 + |y|_{r_2}) t + \sigma_2^2 t = t\kappa(y)$$

say. Since $\mathbb{E}[II] = 0$, we derive by Cauchy-Schwarz's inequality

$$\mathbb{E}\big[\big|\phi_{\sup \mathcal{K}_\delta}(t) - y\big|\big] \leq \sqrt{(\sup \mathcal{K}_\delta - y)^2 + t\kappa(y)}. \tag{27}$$

Step 3). We are ready to control each term of (26). We have

$$\begin{aligned}
&\mathbb{E}\big[\big|\phi_{\sup\mathcal{K}_\delta}\big((t-\tau_x)_+\big)-y\big|\big]\\
\le&|\sup\mathcal{K}_\delta-y|+\mathbb{E}\Big[\int_0^{(t-\tau_x)_+}\big|r\big(\phi_{\sup\mathcal{K}_\delta}(s)\big)\big|ds\Big]+\mathbb{E}\Big[\Big|\int_0^{(t-\tau_x)_+}\sigma\big(\phi_{\sup\mathcal{K}_\delta}(s)\big)dW_s\Big|\Big]\\
\le&|\sup\mathcal{K}_\delta-y|+r_1\mathbb{E}\Big[\int_0^t(1+|\phi_{\sup\mathcal{K}_\delta}(s)|)ds\Big]+\mathbb{E}\Big[\sup_{u\le t}\Big(\int_0^u\sigma\big(\phi_{\sup\mathcal{K}_\delta}(s)\big)dW_s\Big)^2\Big]^{1/2}\\
\le&|\sup\mathcal{K}_\delta-y|+r_1t+r_1\mathbb{E}\Big[\int_0^t|\phi_{\sup\mathcal{K}_\delta}(s)|ds\Big]+\sqrt{2}\sigma_2t\\
\le&|\sup\mathcal{K}_\delta-y|+r_1t+r_1\int_0^t\sqrt{(\sup\mathcal{K}_\delta)^2+s\kappa(0)}ds+\sqrt{2}\sigma_2t\\
\lesssim&1+t^{3/2},
\end{aligned}$$

where we successively applied Assumption 2, Doob's inequality and (27). In the same way

$$\begin{aligned}
&\Big|-\int_0^{(t-\tau_x)_+}\operatorname{sgn}\big(\phi_{\sup\mathcal{K}_\delta}(s)-y\big)d\phi_{\sup\mathcal{K}_\delta}(s)\Big|\\
\le&r_1\int_0^t(1+|\phi_{\sup\mathcal{K}_\delta}(s)|)ds+\sup_{u\le t}\Big|\int_0^u\operatorname{sgn}(\phi_{\sup\mathcal{K}_\delta}(s)-y)\sigma\big(\phi_{\sup\mathcal{K}_\delta}(s)\big)dW_s\Big|.
\end{aligned}$$

Taking expectation and using the foregoing arguments, this last quantity is also of order $1+t^{3/2}$ and Lemma 23 is proved for $x>\sup\mathcal{K}_\delta$.

Step 4). If $x<\inf\mathcal{K}_\delta$, we apply the same arguments, replacing $|\sup\mathcal{K}_\delta|$ by $|\inf\mathcal{K}_\delta|$ with obvious changes. Likewise if $x\in\mathcal{K}_\delta$ we may replace $|\sup\mathcal{K}_\delta|$ by $\max\{|\sup\mathcal{K}_\delta|,|\inf\mathcal{K}_\delta|\}$.

6.2. **Proof of Proposition 19.** Remember that

$$\Psi(\vartheta)=\nu_\vartheta\Big(\mathcal{Q}_\vartheta\Big(\frac{\partial_\vartheta q_\vartheta}{q_\vartheta}\Big)^2\Big)=\int_{\mathcal{X}}\nu_\vartheta(dx)\int_{\mathcal{X}}\frac{\big(\partial_\vartheta q_\vartheta(x,y)\big)^2}{q_\vartheta(x,y)}dy.$$

If $\mathcal{A}\subset\mathcal{X}$ is a Borel set with $\mathrm{Leb}(\mathcal{A})>0$, we have

$$\nu_\vartheta(\mathcal{A})=\int_{\mathcal{X}\times\mathcal{X}}\mathbf{1}_{\mathcal{A}}(y)q_\vartheta(x,y)\nu_\vartheta(dx)dy\ge\inf_{x,y}q_\vartheta(x,y)\,\mathrm{Leb}(\mathcal{A})>0$$

since $\inf_{x,y\in\mathcal{X}}q_\vartheta(x,y)>0$ by Lemma 26. By continuity of $y\mapsto\partial_\vartheta q_\vartheta(x,y)$ on $[0,L]$, it suffices then to show the existence $x,y\in\mathcal{X}$ such that $\partial_\vartheta q_\vartheta(x,y)>0$. For $x,y\in\mathcal{X}$, we have

$$\begin{aligned}
\partial_\vartheta q_\vartheta(x,y)&=\int_{\varepsilon\vee yL^{-1}}^{1-\varepsilon}\frac{\kappa(z)}{z}\sigma^{-2}\mathbb{E}\Big[\int_0^\infty(1-\vartheta t)e^{-\vartheta t}dL_t^{y/z}(\phi_x)\Big]dz\\
&=\int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}}\kappa(y/u)\sigma^{-2}\mathbb{E}\Big[\int_0^\infty(1-\vartheta t)e^{-\vartheta t}dL_t^u(\phi_x)\Big]\frac{du}{u}\\
&=\mathbb{E}\Big[\int_0^\infty\kappa\big(y/\phi_x(t)\big)(1-\vartheta t)e^{-\vartheta t}\mathbf{1}_{\{y(1-\varepsilon)^{-1}\le\phi_x(t)\le y\varepsilon^{-1}\}}\frac{dt}{\phi_x(t)}\Big]\\
&=\frac{1}{1-2\varepsilon}\int_0^\infty(1-\vartheta t)e^{-\vartheta t}\mathbb{E}\Big[\mathbf{1}_{\{y(1-\varepsilon)^{-1}\le\phi_x(t)\le y\varepsilon^{-1}\}}\frac{1}{\phi_x(t)}\Big]dt
\end{aligned}$$

by the change of variable $u = yz^{-1}$, the occupation times formula, and the specific form of κ. For $t \geq 0$, define

$$A_t(x,y) = \mathbb{E}\Big[\mathbf{1}_{\{y(1-\varepsilon)^{-1} \leq \phi_x(t) \leq y\varepsilon^{-1}\}} \frac{1}{\phi_x(t)}\Big] = \int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} \rho_t(x,z)\frac{dz}{z},$$

for which a closed-form formula is known, see for instance [25], Section 4.1, given by

$$\rho_t(x,z) = \frac{2r_1 e^{2r_1 x}}{e^{2r_1 L}-1} + \frac{2}{L}e^{r_1(z-x)}\sum_{n=1}^{\infty}\frac{e^{-a(n)t/2}}{a(n)}g(n,x)g(n,z),$$

with

$$g(n,x) = \frac{\pi n}{L}\cos\Big(x\frac{\pi n}{L}\Big) + r_1 \sin\Big(x\frac{\pi n}{L}\Big), \quad \text{and} \quad a(n) = r_1^2 + \pi^2 n^2/L^2.$$

It follows that

$$A_t(x,y) = \frac{2r_1 e^{2r_1 x}}{e^{2r_1 L}-1}\log\left(\frac{1-\varepsilon}{\varepsilon}\right) + \frac{2}{L}e^{-r_1 x}\sum_{n=1}^{\infty}\frac{e^{-a(n)t/2}}{a(n)}g(n,x)\mathcal{I}(n,y)$$

with $\mathcal{I}(n,y) = \int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} e^{r_1 z}g(n,z)\frac{dz}{z}$, and therefore

$$\begin{aligned}
\partial_\vartheta q_\vartheta(x,y) &= \frac{1}{1-2\varepsilon}\int_0^\infty (1-\vartheta t)e^{-\vartheta t}A_t(x,y)dt \\
&= \frac{1}{1-2\varepsilon}\frac{2}{L}e^{-r_1 x}\sum_{n=1}^{\infty}\Big(\int_0^\infty (1-\vartheta t)e^{-\vartheta t}\frac{e^{-a(n)t/2}}{a(n)}dt\Big)g(n,x)\mathcal{I}(n,y) \\
&= \frac{1}{(1-2\varepsilon)L}\int_{y(1-\varepsilon)^{-1}}^{y\varepsilon^{-1}} e^{-r_1(x-z)}\sum_{n=1}^{\infty}\frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2}\frac{dz}{z}.
\end{aligned}$$

Let $x \in [0,L]$ be such that $g(n,x) \neq 0$ for every $n \geq 1$. Since $x \mapsto g(n,x)$ is continuous on $[0,L]$, there exists $0 < \varepsilon_n < \frac{1}{2}$ such that $g(n,x)g(n,z) > 0$ for all $z \in \mathcal{J}(\varepsilon_n, x) = [2\varepsilon_n x, 2(1-\varepsilon_n)x]$. Let $N > 0$ be such that for all $z \in \mathcal{J}(\varepsilon_1, x)$:

$$|R_N(x,z)| = \Big|\sum_{n=N+1}^{\infty}\frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2}\Big| < \frac{g(1,x)g(1,z)}{(\vartheta + a(1)/2)^2},$$

which exists because by normal convergence of the above series. Then, for every $z \in \mathcal{J}(\max\{\varepsilon_n, 1 \leq n \leq N\}, x)$ we have

$$\sum_{n=1}^{\infty}\frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2} = \sum_{n=1}^{N}\frac{g(n,x)g(n,z)}{(\vartheta + a(n)/2)^2} + R_N(x,z) > \frac{g(1,x)g(1,z)}{(\vartheta + a(n)/2)^2} - |R_N(x,z)| > 0.$$

Finally, for $\varepsilon > \max\{\varepsilon_n, 1 \leq n \leq N\}$, picking $y = 2\varepsilon(1-\varepsilon)x$ yields $[y(1-\varepsilon)^{-1}, y\varepsilon^{-1}] = \mathcal{J}(\varepsilon, x) \subset \mathcal{J}(\max\{\varepsilon_n, 1 \leq n \leq N\}, x)$ so that $\partial_\vartheta q_\theta(x,y) > 0$.

Acknowledgement. We are grateful to V. Bansaye for helpful discussion and comments. A.M. acknowledge partial support by the Chaire Modélisation Mathématique et Biodiversité of Veolia Environnement - École Polytechnique - Museum National Histoire Naturelle - F.X.

References

[1] Romain Azaïs, François Dufour, and Anne Gégout-Petit. Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes. *Scand. J. Stat.*, 41(4):950–969, 2014.

[2] Vincent Bansaye. Proliferating parasites in dividing cells: Kimmel's branching model revisited. *Ann. Appl. Probab.*, 18(3):967–996, 2008.

[3] Vincent Bansaye, Jean-François Delmas, Laurence Marsalle, and Viet Chi Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *Ann. Appl. Probab.*, 21(6):2263–2314, 2011.

[4] Vincent Bansaye and Sylvie Méléard. *Stochastic models for structured populations*, volume 1 of *Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems*. Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH, 2015. Scaling limits and long time behavior.

[5] S. Valère Biteski-Penda and Roche Angelina. Local bandwidth selection for kernel density estimation in bifurcating markov chain model. *arXiv:1706.07034*, 2017.

[6] S. Valère Bitseki Penda, Hacène Djellout, and Arnaud Guillin. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Ann. Appl. Probab.*, 24(1):235–291, 2014.

[7] S. Valère Bitseki Penda, Mikael Escobar-Bach, and Arnaud Guillin. Transportation and concentration inequalities for bifurcating Markov chains. *Bernoulli*, 23(4B):3213–3242, 2017.

[8] S. Valère Bitseki Penda, Marc Hoffmann, and Adélaïde Olivier. Adaptive estimation for bifurcating Markov chains. *Bernoulli*, 23(4B):3598–3637, 2017.

[9] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, Inc., 2005.

[10] P. Cattiaux. Stochastic calculus and degenerate boundary value problems. *Ann. Inst. Fourier (Grenoble)*, 42(3):541–624, 1992.

[11] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19(4):263–284, 1986.

[12] Benoîte de Saporta, Anne Gégout-Petit, and Laurence Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *Statist. Probab. Lett.*, 82(7):1439–1444, 2012.

[13] Benoîte de Saporta, Anne Gégout-Petit, and Laurence Marsalle. Random coefficients bifurcating autoregressive processes. *ESAIM Probab. Stat.*, 18:365–399, 2014.

[14] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM J. Numer. Anal.*, 50(2):925–950, 2012.

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

[16] Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1):119–151, 1993.

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

[18] M. Hairer and J. C Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117, 2011.

[19] V. H. Hoang. Estimating the division kernel of a size-structured population. *arXiv:1509.02872*, 2015.

[20] Marc Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79(1):135–163, 1999.

[21] Marc Hoffmann and Adélaïde Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Process. Appl.*, 126(5):1433–1471, 2016.

[22] Marek Kimmel. Quasistationarity in a branching model of division-within-division. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 157–164. Springer, New York, 1997.

[23] Y. A Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Science & Business Media, 2013.

[24] D. Lépingle. Euler scheme for reflected stochastic differential equations. *Mathematics and Computers in Simulation*, 38(1-3):119–126, 1995. Probabilités numériques (Paris, 1992).

[25] V. Linetsky. On the transition densities for reflected diffusions. *Advances in Applied Probability*, 37(2):435–460, 2005.

[26] A. Marguet. Uniform sampling in a structured branching population. *arXiv:1609.05678*, 2016.

[27] S. Méléard. *Modèles aléatoires en écologie et évolution. (French) [Random models in ecology and evolution]*. Mathématiques et Applications (Berlin) [Mathematics and Applications]. Springer-Verlag, Berlin, 2016.

[28] Bernt Øksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2003.

[29] S. Valère Bitseki Penda and Adélaïde Olivier. Autoregressive functions estimation in nonlinear bifurcating autoregressive models. *Stat. Inference Stoch. Process.*, 20(2):179–210, 2017.
[30] Benoît Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.
[31] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
[32] H. Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Stochastic Processes: Selected Papers of Hiroshi Tanaka*, 9:157, 1979.
[33] Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.
[34] A. W. van der Vaart. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

Marc Hoffmann, Université Paris-Dauphine, PSL Research University, CNRS, CEREMADE, 75016 Paris, France.
E-mail address: hoffmann@ceremade.dauphine.fr

Aline Marguet, CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France.
E-mail address: aline.marguet@cmap.polytechnique.fr