Adaptive estimation for bifurcating Markov chains

Siméon Valère Bitseki Penda, Marc Hoffmann, Adélaïde Olivier

ADAPTIVE ESTIMATION FOR BIFURCATING MARKOV CHAINS

S. VALÈRE BITSEKI PENDA, MARC HOFFMANN AND ADÉLAÏDE OLIVIER

Abstract. In a first part, we prove Bernstein-type deviation inequalities for bifurcating Markov chains (BMC) under a geometric ergodicity assumption, completing former results of Guyon and Bitseki Penda, Djellout and Guillin. These preliminary results are the key ingredient to implement nonparametric wavelet thresholding estimation procedures: in a second part, we construct nonparametric estimators of the transition density of a BMC, of its mean transition density and of the corresponding invariant density, and show smoothness adaptation over various multivariate Besov classes under L^p-loss error, for $1 \leq p < \infty$. We prove that our estimators are (nearly) optimal in a minimax sense. As an application, we obtain new results for the estimation of the splitting size-dependent rate of growth-fragmentation models and we extend the statistical study of bifurcating autoregressive processes.

Mathematics Subject Classification (2010): 62G05, 62M05, 60J80, 60J20, 92D25,
Keywords: Bifurcating Markov chains, binary trees, deviations inequalities, nonparametric adaptive estimation, minimax rates of convergence, bifurcating autoregressive process, growth-fragmentation processes.

1. Introduction

1.1. **Bifurcating Markov chains.** Bifurcating Markov Chains (BMC) are Markov chains indexed by a tree (Athreya and Kang [1], Benjamini and Peres [6], Takacs [39]) that are particularly well adapted to model and understand dependent data mechanisms involved in cell division. To that end, bifurcating autoregressive models (a specific class of BMC, also considered in the paper) were first introduced by Cowan and Staudte [16]. More recently Guyon [28] systematically studied BMC in a general framework. In continuous time, BMC encode certain piecewise deterministic Markov processes on trees that serve as the stochastic realisation of growth-fragmentation models (see *e.g.* Doumic *et al.* [26], Robert *et al.* [38] for modelling cell division in *Escherichia coli* and the references therein).

For $m \geq 0$, let $\mathbb{G}_m = \{0,1\}^m$ (with $\mathbb{G}_0 = \{\emptyset\}$) and introduce the infinite genealogical tree

$$\mathbb{T} = \bigcup_{m=0}^{\infty} \mathbb{G}_m.$$

For $u \in \mathbb{G}_m$, set $|u| = m$ and define the concatenation $u0 = (u,0) \in \mathbb{G}_{m+1}$ and $u1 = (u,1) \in \mathbb{G}_{m+1}$. A bifurcating Markov chain is specified by **1)** a measurable state space $(\mathcal{S}, \mathfrak{S})$ with a Markov kernel (later called $\mathbb{T}$-transition) $\mathcal{P}$ from $(\mathcal{S}, \mathfrak{S})$ to $(\mathcal{S}\times\mathcal{S}, \mathfrak{S}\otimes\mathfrak{S})$ and **2)** a filtered probability space $\big(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m\geq 0}, \mathbb{P}\big)$. Following Guyon, we have the

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

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

for every $m \geq 0$ *and any family of (bounded) measurable functions* $(g_u)_{u\in\mathbb{G}_m}$*, where* $\mathcal{P}g(x) = \int_{\mathcal{S}\times\mathcal{S}} g(x,y,z)\mathcal{P}(x,dy\,dz)$ *denotes the action of* $\mathcal{P}$ *on* g*.*

The distribution of $(X_u)_{u\in\mathbb{T}}$ is thus entirely determined by $\mathcal{P}$ and an initial distribution for $X_\emptyset$. Informally, we may view $(X_u)_{u\in\mathbb{T}}$ as a population of individuals, cells or particles indexed by $\mathbb{T}$ and governed by the following dynamics: to each $u \in \mathbb{T}$ we associate a trait X_u (its size, lifetime, growth rate, DNA content and so on) with value in $\mathcal{S}$. At its time of death, the particle u gives rize to two children $u0$ and $u1$. Conditional on $X_u = x$, the trait $(X_{u0}, X_{u1}) \in \mathcal{S}\times\mathcal{S}$ of the offspring of u is distributed according to $\mathcal{P}(x, dy\,dz)$.

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. Assume we observe $\mathbb{X}^n = (X_u)_{u\in\mathbb{T}_n}$, *i.e.* we have $2^{n+1}-1$ random variables with value in $\mathcal{S}$. There are several objects of interest that we may try to infer from the data $\mathbb{X}^n$. Similarly to fragmentation processes (see *e.g.* Bertoin [9]) a key role for both asymptotic and non-asymptotic analysis of bifurcating Markov chains is played by the so-called *tagged-branch chain*, as shown by Guyon [28] and Bitseki Penda *et al.* [11]. The tagged-branch chain $(Y_m)_{m\geq 0}$ corresponds to a lineage picked at random in the population $(X_u)_{u\in\mathbb{T}}$: it is a Markov chain with value in $\mathcal{S}$ defined by $Y_0 = X_\emptyset$ and for $m \geq 1$,

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

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

$$\mathcal{Q} = (\mathcal{P}_0 + \mathcal{P}_1)/2,$$

obtained from the marginal transitions

$$\mathcal{P}_0(x,dy) = \int_{z\in\mathcal{S}} \mathcal{P}(x,dy\,dz) \ \text{ and } \ \mathcal{P}_1(x,dz) = \int_{y\in\mathcal{S}} \mathcal{P}(x,dy\,dz)$$

of $\mathcal{P}$. Guyon proves in [28] that if $(Y_m)_{m\geq 0}$ is ergodic with invariant measure ν, then the convergence

$$\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u) \to \int_{\mathcal{S}} g(x)\nu(dx) \tag{1}$$

holds almost-surely as $n \to \infty$ for appropriate test functions g. Moreover, we also have convergence results of the type

$$\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) \to \int_{\mathcal{S}} \mathcal{P}g(x)\nu(dx) \tag{2}$$

almost-surely as $n \to \infty$. These results are appended with central limit theorems (Theorem 19 of [28]) and Hoeffding-type deviations inequalities in a non-asymptotic setting (Theorem 2.11 and 2.12 of Bitseki Penda *et al.* [11]).

1.2. **Objectives.** The observation of $\mathbb{X}^n$ enables us to identify $\nu(dx)$ as $n \to \infty$ thanks to (1). Consequently, convergence (2) reveals $\mathcal{P}$ and therefore $\mathcal{Q}$ is identified as well, at least asymptotically. The purpose of the present paper is at least threefold:

1) Construct – under appropriate regularity conditions – estimators of $\nu, \mathcal{Q}$ and $\mathcal{P}$ and study their rates of convergence as $n \to \infty$ under various loss functions. When $\mathcal{S} \subseteq \mathbb{R}$ and when $\mathcal{P}$ is absolutely continuous w.r.t. the Lebesgue measure, we estimate the corresponding density functions under various smoothness class assumptions and build *smoothness adaptive* estimators, *i.e.* estimator that achieve an optimal rate of convergence without prior knowledge of the smoothness class.

2) Apply these constructions to investigate further specific classes of BMC. These include binary growth-fragmentation processes, where we subsequently estimate adaptively the splitting rate of a size-dependent model, thus extending previous results of Doumic *et al.* [26] and bifurcating autoregressive processes, where we complete previous studies of Bitseki Penda *et al.* [12] and Bitseki Penda and Olivier [13].

3) For the estimation of $\nu, \mathcal{Q}$ and $\mathcal{P}$ and the subsequent estimation results of **2)**, prove that our results are sharp in a minimax sense.

Our smoothness adaptive estimators are based on wavelet thresholding for density estimation (Donoho *et al.* [24] in the generalised framework of Kerkyacharian and Picard [32]). Implementing these techniques requires concentration properties of empirical wavelet coefficients. To that end, we prove new deviation inequalities for bifurcating Markov chains that we develop independently in a more general setting, when $\mathcal{S}$ is not necessarily restricted to $\mathbb{R}$. Note also that when $\mathcal{P}_0 = \mathcal{P}_1$, we have $\mathcal{Q} = \mathcal{P}_0 = \mathcal{P}_1$ as well and we retrieve the usual framework of nonparametric estimation of Markov chains when the observation is based on $(Y_i)_{1\leq i\leq n}$ solely. We are therefore in the line of combining and generalising the study of Clémençon [15] and Lacour [33, 34] that both consider adaptive estimation for Markov chains when $\mathcal{S} \subseteq \mathbb{R}$.

1.3. **Main results and organisation of the paper.** In Section 2, we generalise the Hoeffding-type deviations inequalities of Bitseki Penda *et al.* [11] for BMC to Bernstein-type inequalities: when $\mathcal{P}$ is uniformly geometrically ergodic (Assumption 3 below), we prove in Theorem 5 deviations of the form

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u, X_{u0}, X_{u1}) - \int \mathcal{P}g\, d\nu \geq \delta\Big) \leq \exp\Big(-\frac{\kappa|\mathbb{G}_n|\delta^2}{\Sigma_n(g) + |g|_\infty \delta}\Big)$$

and

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) - \int \mathcal{P}g\, d\nu \geq \delta\Big) \leq \exp\Big(-\frac{\tilde{\kappa} n^{-1}|\mathbb{T}_n|\delta^2}{\Sigma_n(g) + |g|_\infty \delta}\Big),$$

where $\kappa, \tilde{\kappa} > 0$ only depend on $\mathcal{P}$ and $\Sigma_n(g)$ is a variance term which depends on a combination of the L^p-norms of g for $p = 1, 2, \infty$ w.r.t. a common dominating measure for the family $\{\mathcal{Q}(x, dy), x \in \mathcal{S}\}$. The precise results are stated in Theorems 4 and 5.

Section 3 is devoted to the statistical estimation of $\nu, \mathcal{Q}$ and $\mathcal{P}$ when $\mathcal{S} \subseteq \mathbb{R}$ and the family $\{\mathcal{P}(x, dy\, dz), x \in \mathcal{S}\}$ is dominated by the Lebesgue measure on $\mathbb{R}^2$. In that setting, abusing notation slightly, we have $\nu(dx) = \nu(x)dx$, $\mathcal{Q}(x, dy) = \mathcal{Q}(x, y)dy$ and $\mathcal{P}(x, dy\, dz) = \mathcal{P}(x, y, z)dydz$ for some functions $x \rightsquigarrow \nu(x)$, $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$ and $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$ that we reconstruct nonparametrically. Our estimators are constructed in several steps:

i) We approximate the functions $\nu(x)$, $f_{\mathcal{Q}}(x,y) = \nu(x)\mathcal{Q}(x,y)$ and $f_{\mathcal{P}}(x,y,z) = \nu(x)\mathcal{P}(x,y,z)$ by atomic representations

$$\begin{aligned} \nu(x) &\approx \sum_{\lambda \in \mathcal{V}^1(\nu)} \langle \nu, \psi_\lambda^1 \rangle \psi_\lambda^1(x), \\ f_{\mathcal{Q}}(x,y) &\approx \sum_{\lambda \in \mathcal{V}^2(f_{\mathcal{Q}})} \langle f_{\mathcal{Q}}, \psi_\lambda^2 \rangle \psi_\lambda^2(x,y), \\ f_{\mathcal{P}}(x,y,z) &\approx \sum_{\lambda \in \mathcal{V}^3(f_{\mathcal{P}})} \langle f_{\mathcal{P}}, \psi_\lambda^3 \rangle \psi_\lambda^3(x,y,z), \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual L^2-inner product (over $\mathbb{R}^d$, for $d = 1,2,3$ respectively) and $\big(\psi_\lambda^d, \lambda \in \mathcal{V}^d(\cdot)\big)$ is a collection of functions (wavelets) in $L^2(\mathbb{R}^d)$ that are localised in time and frequency, indexed by a set $\mathcal{V}^d(\cdot)$ that depends on the signal itself[1].

ii) We estimate

$$\begin{aligned} \langle \nu, \psi_\lambda^1 \rangle \ &\text{by} \ |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \psi_\lambda^1(X_u), \\ \langle f_{\mathcal{Q}}, \psi_\lambda^2 \rangle \ &\text{by} \ |\mathbb{T}_n^\star|^{-1} \sum_{u \in \mathbb{T}_n^\star} \psi_\lambda^2(X_{u^-}, X_u), \\ \langle f_{\mathcal{P}}, \psi_\lambda^3 \rangle \ &\text{by} \ |\mathbb{T}_{n-1}|^{-1} \sum_{u \in \mathbb{T}_{n-1}} \psi_\lambda^3(X_u, X_{u0}, X_{u1}), \end{aligned}$$

where X_{u^-} denotes the trait of the parent of u and $\mathbb{T}_n^\star = \mathbb{T}_n \setminus \mathbb{G}_0$, and specify a selection rule for $\mathcal{V}^d(\cdot)$ (with the dependence in the unknown function somehow replaced by an estimator). The rule is dictated by hard thresholding over the estimation of the coefficients that are kept only if they exceed some noise level, tuned with $|\mathbb{T}_n|$ and prior knowledge on the unknown function, as follows by standard density estimation by wavelet thresholding (Donoho *et al.* [25], Kerkyacharian and Picard [32]).

iii) Denoting by $\widehat{\nu}_n(x)$, $\widehat{f}_n(x,y)$ and $\widehat{f}_n(x,y,z)$ the estimators of $\nu(x)$, $f_{\mathcal{Q}}(x,y)$ and $f_{\mathcal{P}}(x,y,z)$ respectively constructed in Step **ii)**, we finally take as estimators for $\mathcal{Q}(x,y)$ and $\mathcal{P}(x,y,z)$ the quotient estimators

$$\widehat{\mathcal{Q}}_n(x,y) = \frac{\widehat{f}_n(x,y)}{\widehat{\nu}_n(x)} \ \text{ and } \ \widehat{\mathcal{P}}_n(x,y,z) = \frac{\widehat{f}_n(x,y,z)}{\widehat{\nu}_n(x)}$$

provided $\widehat{\nu}_n(x)$ exceeds a minimal threshold.

Beyond the inherent technical difficulties of the approximation Steps i) and iii), the crucial novel part is the estimation Step ii), where Theorems 4 and 5 are used to estimate precisely the probability that the thresholding rule applied to the empirical wavelet coefficient is close in effect to thresholding the true coefficients.

When $\nu, \mathcal{Q}$ or $\mathcal{P}$ (identified with their densities w.r.t. appropriate dominating measures) belong to an isotropic Besov ball of smoothness s measured in L^π over a domain $\mathcal{D}^d$ in $\mathbb{R}^d$, with $s > d/\pi$ and $d = 1,2,3$ respectively, we prove in Theorems 8, 9 and 10 that if $\mathcal{Q}$ is uniformly geometrically

[1] The precise meaning of the symbol $\approx$ and the properties of the ψ_λ's are stated precisely in Section 3.1.

ergodic, then our estimators achieve the rate $|\mathbb{T}_n|^{-\alpha_d(s,p,\pi)}$ in $L^p(\mathcal{D})$-loss, up to additional $\log|\mathbb{T}_n|$ terms, where

$$\alpha_d(s,p,\pi) = \min\Big\{\frac{s}{2s+d}, \frac{s+d(1/p-1/\pi)}{2s+d(1-2/\pi)}\Big\}$$

is the usual exponent for the minimax rate of estimation of a d-variate function with order of smoothness s measured in L^π in L^p-loss error. This rate is nearly optimal in a minimax sense for $d=1$, as follows from particular case $\mathcal{Q}(x,dy)=\nu(dy)$ that boils down to density estimation with $|\mathbb{T}_n|$ data: the optimality is then a direct consequence of Theorem 2 in Donoho *et al.* [25]. As for the case $d=2$ and $d=3$, the structure of BMC comes into play and we need to prove a specific optimality result, stated in Theorems 9 and 10. We rely on classical lower bound techniques for density estimation and Markov chains (Hoffmann [31], Clémençon [15], Lacour [33, 34]).

We apply our generic results in Section 4 to two illustrative examples. We consider in Section 4.1 the growth-fragmentation model as studied in Doumic *et al.* [26], where we estimate the size-dependent splitting rate of the model as a function of the invariant measure of an associated BMC in Theorem 11. This enables us to extend the recent results of Doumic *et al.* in several directions: adaptive estimation, extension of the smoothness classes and the loss functions considered, and also a proof of a minimax lower bound. In Section 4.2, we show how bifurcating autoregressive models (BAR) as developped for instance in de Saporta *et al.* [8] and Bitseki Penda and Olivier [13] are embedded into our generic framework of estimation. A numerical illustration highlights the feasibility of our procedure in practice and is presented in Section 4.3. The proofs are postponed to Section 5.

2. Deviations inequalities for empirical means

In the sequel, we fix a (measurable) subset $\mathcal{D}\subseteq\mathcal{S}$ that will be later needed for statistical purposes. We need some regularity on the $\mathbb{T}$-transition $\mathcal{P}$ via its mean transition $\mathcal{Q}=\frac{1}{2}(\mathcal{P}_0+\mathcal{P}_1)$.

Assumption 2. *The family $\{\mathcal{Q}(x,dy), x\in\mathcal{S}\}$ is dominated by a common sigma-finite measure $\mathfrak{n}(dy)$. We have (abusing notation slightly)*

$$\mathcal{Q}(x,dy)=\mathcal{Q}(x,y)\mathfrak{n}(dy)\ \textit{ for every }\ x\in\mathcal{S},$$

for some $\mathcal{Q}:\mathcal{S}^2\to[0,\infty)$ such that

$$|\mathcal{Q}|_{\mathcal{D}}=\sup_{x\in\mathcal{S},y\in\mathcal{D}}\mathcal{Q}(x,y)<\infty.$$

An invariant probability measure for $\mathcal{Q}$ is a probability ν on $(\mathcal{S},\mathfrak{S})$ such that $\nu\mathcal{Q}=\nu$ where $\nu\mathcal{Q}(dy)=\int_{x\in\mathcal{S}}\nu(dx)\mathcal{Q}(x,dy)$. We set

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

for the r-th iteration of $\mathcal{Q}$. For a function $g:\mathcal{S}^d\to\mathbb{R}$ with $d=1,2,3$ and $1\leq p\leq\infty$, we denote by $|g|_p$ its L^p-norm w.r.t. the measure $\mathfrak{n}^{\otimes d}$, allowing for the value $|g|_p=\infty$ if $g\notin L^p(\mathfrak{n}^{\otimes d})$. The same notation applies to a function $g:\mathcal{D}^d\to\mathbb{R}$ tacitly considered as a function from $\mathcal{S}^d\to\mathbb{R}$ by setting $g(x)=0$ for $x\in\mathcal{S}\setminus\mathcal{D}$.

Assumption 3. *The mean transition $\mathcal{Q}$ admits a unique invariant probability measure ν and there exist $R>0$ and $0<\rho<1/2$ such that*

$$\big|\mathcal{Q}^m g(x)-\int_{\mathcal{S}}g\,d\nu\big|\leq R|g|_\infty\,\rho^m,\quad x\in\mathcal{S},\quad m\geq 0,$$

for every g integrable w.r.t. ν.

Assumption 3 is a uniform geometric ergodicity condition that can be verified in most applications using the theory of Meyn and Tweedie [36]. The ergodicity rate should be small enough ($\rho < 1/2$) and this point is crucial for the proofs. However this is sometimes delicate to check in applications and we refer to Hairer and Mattingly [29] for an explicit control of the ergodicity rate.

Our first result is a deviation inequality for empirical means over $\mathbb{G}_n$ or $\mathbb{T}_n$. We need some notation. Let

$$\begin{aligned}
\kappa_1 =& \kappa_1(\mathcal{Q},\mathcal{D}) = 32\max\big\{|\mathcal{Q}|_{\mathcal{D}}, 4|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2\big\},\\
\kappa_2 =& \kappa_2(\mathcal{Q}) = \tfrac{16}{3}\max\big\{1+R\rho, R(1+\rho)\big\},\\
\kappa_3 =& \kappa_3(\mathcal{Q},\mathcal{D}) = 96\max\big\{|\mathcal{Q}|_{\mathcal{D}}, 16|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2(1-2\rho)^{-2}\big\},\\
\kappa_4 =& \kappa_4(\mathcal{Q}) = \tfrac{16}{3}\max\big\{1+R\rho, R(1+\rho)(1-2\rho)^{-1}\big\},
\end{aligned}$$

where $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x\in\mathcal{S},y\in\mathcal{D}}\mathcal{Q}(x,y)$ is defined in Assumption 2. For $g:\mathcal{S}^d\to\mathbb{R}$, define $\Sigma_{1,1}(g) = |g|_2^2$ and for $n\geq 2$,

$$\Sigma_{1,n}(g) = |g|_2^2 + \min_{1\leq\ell\leq n-1}\big(|g|_1^2 2^{\ell} + |g|_\infty^2 2^{-\ell}\big). \tag{3}$$

Define also $\Sigma_{2,1}(g) = |\mathcal{P}g^2|_1$ and for $n\geq 2$,

$$\Sigma_{2,n}(g) = |\mathcal{P}g^2|_1 + \min_{1\leq\ell\leq n-1}\big(|\mathcal{P}g|_1^2 2^{\ell} + |\mathcal{P}g|_\infty^2 2^{-\ell}\big). \tag{4}$$

Theorem 4. *Work under Assumptions 2 and 3. Then, for every $n\geq 1$ and every $g:\mathcal{D}\subseteq\mathcal{S}\to\mathbb{R}$ integrable w.r.t. ν, the following inequalities hold true:*

(i) *For any $\delta>0$ such that $\delta\geq 4R|g|_\infty|\mathbb{G}_n|^{-1}$, we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u) - \int_{\mathcal{S}} g\,d\nu \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1\Sigma_{1,n}(g)+\kappa_2|g|_\infty\delta}\Big).$$

(ii) *For any $\delta>0$ such that $\delta\geq 4R(1-2\rho)^{-1}|g|_\infty|\mathbb{T}_n|^{-1}$, we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} g(X_u) - \int_{\mathcal{S}} g\,d\nu \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{T}_n|\delta^2}{\kappa_3\Sigma_{1,n}(g)+\kappa_4|g|_\infty\delta}\Big).$$

Theorem 5. *Work under Assumptions 2 and 3. Then, for every $n\geq 2$ and for every $g:\mathcal{D}^3\subseteq\mathcal{S}^3\to\mathbb{R}$ such that $\mathcal{P}g$ is well defined and integrable w.r.t. ν, the following inequalities hold true:*

(i) *For any $\delta>0$ such that $\delta\geq 4R|\mathcal{P}g|_\infty|\mathbb{G}_n|^{-1}$, we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(X_u,X_{u0},X_{u1}) - \int_{\mathcal{S}}\mathcal{P}g\,d\nu \geq \delta\Big) \leq \exp\Big(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1\Sigma_{2,n}(g)+\kappa_2|g|_\infty\delta}\Big).$$

(ii) *For any $\delta>0$ such that $\delta\geq 4(nR|\mathcal{P}g|_\infty+|g|_\infty)|\mathbb{T}_{n-1}|^{-1}$, we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u\in\mathbb{T}_{n-1}} g(X_u,X_{u0},X_{u1}) - \int_{\mathcal{S}}\mathcal{P}g\,d\nu \geq \delta\Big) \leq \exp\Big(\frac{-n^{-1}|\mathbb{T}_{n-1}|\delta^2}{\kappa_1\Sigma_{2,n-1}(g)+\kappa_2|g|_\infty\delta}\Big).$$

A few remarks are in order:

1) Theorem 4 (i) is a direct consequence of Theorem 5 (i) but Theorem 4 (ii) is not a corollary of Theorem 5 (ii): we note that a slow term or order $n^{-1} \approx (\log |\mathbb{T}_n|)^{-1}$ comes in Theorem 5 (ii).
2) Bitseki-Penda *et al.* in [11] study similar Hoeffding-type deviations inequalities for functionals of bifurcating Markov chains under ergodicity assumption and for uniformly bounded functions. In the present work and for statistical purposes, we need Bernstein-type deviations inequalities which require a specific treatment than cannot be obtained from a direct adaptation of [11]. In particular, we apply our results to multivariate wavelets test functions ψ_λ^d that are well localised but unbounded, and a fine control of the conditional variance $\Sigma_{i,n}(\psi_\lambda^d)$, $i = 1, 2$ is of crucial importance.
3) Assumption 3 about the uniform geometric ergodicity is quite strong, although satisfied in the two examples developed in Section 4 (at the cost however of assuming that the splitting rate of the growth-fragmentation model has bounded support in Section 4.1). Presumably, a way to relax this restriction would be to require a weaker geometric ergodicity condition of the form

$$\big|\mathcal{Q}^m g(x) - \int_{\mathcal{S}} g\, d\nu\big| \leq R|g|_\infty V(x)\, \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

for some Lyapunov function $V : \mathcal{S} \to [1, \infty)$. Analogous results could then be obtained via transportation information inequalities for bifurcating Markov chains with a similar approach as in Gao *et al.* [27], but this lies beyond the scope of the paper.

3. Statistical estimation

In this section, we take $(\mathcal{S}, \mathfrak{S}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$. As in the previous section, we fix a compact interval $\mathcal{D} \subseteq \mathcal{S}$. The following assumption will be needed here

Assumption 6. *The family $\{\mathcal{P}(x, dy\, dz), x \in \mathcal{S}\}$ is dominated w.r.t. the Lebesgue measure on $\big(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)\big)$. We have (abusing notation slightly)*

$$\mathcal{P}(x, dy\, dz) = \mathcal{P}(x, y, z) dy\, dz \;\; \textit{for every} \;\; x \in \mathcal{S}$$

for some $\mathcal{P} : \mathcal{S}^3 \to [0, \infty)$ such that

$$|\mathcal{P}|_{\mathcal{D},1} = \int_{\mathcal{S}^2} \sup_{x \in \mathcal{D}} \mathcal{P}(x, y, z) dy dz < \infty.$$

Under Assumptions 2, 3 and 6 with $\mathfrak{n}(dy) = dy$, we have (abusing notation slightly)

$$\mathcal{P}(x, dy\, dz) = \mathcal{P}(x, y, z) dy\, dz, \;\; \mathcal{Q}(x, dy) = \mathcal{Q}(x, y) dy \;\; \text{and} \;\; \nu(dx) = \nu(x) dx.$$

For some $n \geq 1$, we observe $\mathbb{X}_n = (X_u)_{u \in \mathbb{T}_n}$ and we aim at constructing nonparametric estimators of $x \rightsquigarrow \nu(x)$, $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$ and $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$ for $x, y, z \in \mathcal{D}$. To that end, we use regular wavelet bases adapted to the domain $\mathcal{D}^d$ for $d = 1, 2, 3$.

3.1. Atomic decompositions and wavelets.

Wavelet bases $(\psi_\lambda^d)_\lambda$ adapted to a domain $\mathcal{D}^d$ in $\mathbb{R}^d$, for $d = 1, 2, 3$ are documented in numerous textbooks, see *e.g.* Cohen [17]. The multi-index λ concatenates the spatial index and the resolution level $j = |\lambda|$. We set $\Lambda_j = \{\lambda,\; |\lambda| = j\}$ and $\Lambda = \cup_{j \geq -1} \Lambda_j$. Thus, for $g \in L^\pi(\mathcal{D}^d)$ for some $\pi \in (0, \infty]$, we have

$$g = \sum_{j \geq -1} \sum_{\lambda \in \Lambda_j} g_\lambda \psi_\lambda^d = \sum_{\lambda \in \Lambda} g_\lambda \psi_\lambda^d, \quad \text{with} \;\; g_\lambda = \langle g, \psi_\lambda^d \rangle,$$

where we have set $j = -1$ in order to incorporate the low frequency part of the decomposition and $\langle g, \psi_\lambda^d\rangle = \int g\psi_\lambda^d$ denotes the inner product in $L^2(\mathbb{R}^d)$. From now on, the basis $(\psi_\lambda^d)_\lambda$ is fixed. For $s > 0$ and $\pi \in (0,\infty]$, g belongs to $B^s_{\pi,\infty}(\mathcal{D})$ if the following norm is finite:

$$\|g\|_{B^s_{\pi,\infty}(\mathcal{D})} = \sup_{j\geq -1} 2^{j(s+d(1/2-1/\pi))}\big(\sum_{\lambda\in\Lambda_j} |\langle g, \psi_\lambda^d\rangle|^\pi\big)^{1/\pi} \tag{5}$$

with the usual modification if $\pi = \infty$. Precise connection between this definition of Besov norm and more standard ones can be found in [17]. Given a basis $(\psi_\lambda^d)_\lambda$, there exists $\sigma > 0$ such that for $\pi \geq 1$ and $s \leq \sigma$ the Besov space defined by (5) exactly matches the usual definition in terms of moduli of smoothness for g. The index σ can be taken arbitrarily large. The additional properties of the wavelet basis $(\psi_\lambda^d)_\lambda$ that we need are summarized in the next assumption.

Assumption 7. *For $p \geq 1$,*

$$\|\psi_\lambda^d\|_{L^p}^p \sim 2^{|\lambda|d(p/2-1)}, \tag{6}$$

for some $\sigma > 0$ and for all $s \leq \sigma$, $j_0 \geq 0$,

$$\|g - \sum_{j\leq j_0}\sum_{\lambda\in\Lambda_j} g_\lambda \psi_\lambda^d\|_{L^p} \lesssim 2^{-j_0 s}\|g\|_{B^s_{p,\infty}(\mathcal{D})}, \tag{7}$$

for any subset $\Lambda_0 \subset \Lambda$,

$$\int_{\mathcal{D}} \big(\sum_{\lambda\in\Lambda_0} |\psi_\lambda^d(x)|^2\big)^{p/2} dx \sim \sum_{\lambda\in\Lambda_0} \|\psi_\lambda^d\|_{L^p}^p. \tag{8}$$

If $p > 1$, for any sequence $(u_\lambda)_{\lambda\in\Lambda}$,

$$\big\|\big(\sum_{\lambda\in\Lambda} |u_\lambda\psi_\lambda^d|^2\big)^{1/2}\big\|_{L^p} \sim \|\sum_{\lambda\in\Lambda} u_\lambda\psi_\lambda^d\|_{L^p}. \tag{9}$$

The symbol $\sim$ means inequality in both ways, up to a constant depending on p and $\mathcal{D}$ only. The property (7) reflects that our definition (5) of Besov spaces matches the definition in term of linear approximation. Property (9) reflects an unconditional basis property, see Kerkyacharian and Picard [32], De Vore *et al.* [21] and (8) is referred to as a superconcentration inequality, or Temlyakov property [32]. The formulation of (8)-(9) in the context of statistical estimation is posterior to the original papers of Donoho and Johnstone [22, 23] and Donoho *et al.* [25, 24] and is due to Kerkyacharian and Picard [32]. The existence of compactly supported wavelet bases satisfying Assumption 7 is discussed in Meyer [35], see also Cohen [17].

3.2. **Estimation of the invariant density ν.** Recall that we estimate $x \rightsquigarrow \nu(x)$ for $x \in \mathcal{D}$, taken as a compact interval in $\mathcal{S} \subseteq \mathbb{R}$. We approximate the representation

$$\nu(x) = \sum_{\lambda\in\Lambda} \nu_\lambda\psi_\lambda^1(x),\ \ \nu_\lambda = \langle \nu, \psi_\lambda^1\rangle$$

by

$$\widehat{\nu}_n(x) = \sum_{|\lambda|\leq J} \widehat{\nu}_{\lambda,n}\psi_\lambda^1(x),$$

with

$$\widehat{\nu}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}\psi_\lambda^1(X_u)\Big),$$

and $\mathcal{T}_{\lambda,\eta}(x) = x\mathbf{1}_{|x|\geq\eta}$ denotes the standard threshold operator (with $\mathcal{T}_{\lambda,\eta}(x) = x$ for the low frequency part when $\lambda \in \Lambda_{-1}$). Thus $\widehat{\nu}_n$ is specified by the maximal resolution level J and the threshold η.

Theorem 8. *Work under Assumptions 2 and 3 with $\mathfrak{n}(dx) = dx$. Specify $\widehat{\nu}_n$ with*

$$J = \log_2 \frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|} \text{ and } \eta = c\sqrt{\log|\mathbb{T}_n|/|\mathbb{T}_n|}$$

for some $c > 0$. For every $\pi \in (0,\infty]$, $s > 1/\pi$ and $p \geq 1$, for large enough n and c, the following estimate holds

$$\Big(\mathbb{E}\big[\|\widehat{\nu}_n - \nu\|^p_{L^p(\mathcal{D})}\big]\Big)^{1/p} \lesssim \Big(\frac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_1(s,p,\pi)},$$

with $\alpha_1(s,p,\pi) = \min\big\{\frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi}\big\}$, up to a constant that depends on $s, p, \pi, \|\nu\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$, ρ, R and $|\mathcal{Q}|_{\mathcal{D}}$ and that is continuous in its arguments.

Two remarks are in order:

1) The upper-rate of convergence is the classical minimax rate in density estimation. We infer that our estimator is nearly optimal in a minimax sense as follows from Theorem 2 in Donoho *et al.* [25] applied to the class $\mathcal{Q}(x,y)dy = \nu(y)dy$, *i.e.* in the particular case when we have i.i.d. X_u's. We highlight the fact that n represents here the number of observed generations in the tree, which means that we observe $|\mathbb{T}_n| = 2^{n+1}-1$ traits.
2) The estimator $\widehat{\nu}_n$ is *smooth-adaptive* in the following sense: for every $s_0 > 0$, $0 < \rho_0 < 1/2$, $R_0 > 0$ and $\mathcal{Q}_0 > 0$, define the sets $\mathcal{A}(s_0) = \{(s,\pi), s \geq s_0, s_0 \geq 1/\pi\}$ and

$$\mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0) = \{\mathcal{Q} \text{ such that } \rho \leq \rho_0, R \leq R_0, |\mathcal{Q}|_{\mathcal{D}}, \leq \mathcal{Q}_0\},$$

where $\mathcal{Q}$ is taken among mean transitions for which Assumption 3 holds. Then, for every $C > 0$, there exists $c^\star = c^\star(\mathcal{D}, p, s_0, \rho_0, R_0, \mathcal{Q}_0, C)$ such that $\widehat{\nu}_n$ specified with $c^\star$ satisfies

$$\sup_n \sup_{(s,\pi)\in\mathcal{A}(s_0)} \sup_{\nu,\mathcal{Q}} \Big(\frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|}\Big)^{p\alpha_1(s,p,\pi)} \mathbb{E}\big[\|\widehat{\nu}_n - \nu\|^p_{L^p(\mathcal{D})}\big] < \infty$$

where the supremum is taken among $(\nu, \mathcal{Q})$ such that $\nu\mathcal{Q} = \nu$ with $\mathcal{Q} \in \mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0)$ and $\|\nu\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$ $\leq C$. In particular, $\widehat{\nu}_n$ achieves the (near) optimal rate of convergence over Besov balls simultaneously for all $(s,\pi) \in \mathcal{A}(s_0)$. Analogous smoothness adaptive results hold for Theorems 9, 10 and 11 below.

3.3. **Estimation of the density of the mean transition $\mathcal{Q}$.** In this section we estimate $(x,y) \rightsquigarrow \mathcal{Q}(x,y)$ for $(x,y) \in \mathcal{D}^2$ and $\mathcal{D}$ is a compact interval in $\mathcal{S} \subseteq \mathbb{R}$. In a first step, we estimate the density

$$f_{\mathcal{Q}}(x,y) = \nu(x)\mathcal{Q}(x,y)$$

of the distribution of (X_{u^-}, X_u) when $\mathcal{L}(X_\emptyset) = \nu$ (a restriction we do not need here) by

$$\widehat{f}_n(x,y) = \sum_{|\lambda|\leq J} \widehat{f}_{\lambda,n}\psi^2_\lambda(x,y),$$

with

$$\widehat{f}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}^\star_n|}\sum_{u\in\mathbb{T}^\star_n}\psi^2_\lambda(X_{u^-}, X_u)\Big),$$

and $\mathcal{T}_{\lambda,\eta}(\cdot)$ is the hard-threshold estimator defined in Section 3.2 and $\mathbb{T}_n^\star = \mathbb{T}_n \setminus \mathbb{G}_0$. We can now estimate the density $\mathcal{Q}(x,y)$ of the mean transition probability by

$$\widehat{\mathcal{Q}}_n(x,y) = \frac{\widehat{f}_n(x,y)}{\max\{\widehat{\nu}_n(x), \varpi\}} \tag{10}$$

for some threshold $\varpi > 0$. Thus the estimator $\widehat{\mathcal{Q}}_n$ is specified by J, η and ϖ. Define also

$$m(\nu) = \inf_x \nu(x) \tag{11}$$

where the infimum is taken among all x such that $(x,y) \in \mathcal{D}^2$ for some y.

Theorem 9. *Work under Assumptions 2 and 3 with $\mathfrak{n}(dx) = dx$. Specify $\widehat{\mathcal{Q}}_n$ with*

$$J = \tfrac{1}{2}\log_2 \frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{(\log|\mathbb{T}_n|)^2/|\mathbb{T}_n|}$$

for some $c > 0$ and $\varpi > 0$. For every $\pi \in [1,\infty]$, $s > 2/\pi$ and $p \geq 1$, for large enough n and c and small enough ϖ, the following estimate holds

$$\Big(\mathbb{E}\big[\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p\big]\Big)^{1/p} \lesssim \Big(\frac{(\log|\mathbb{T}_n|)^2}{|\mathbb{T}_n|}\Big)^{\alpha_2(s,p,\pi)}, \tag{12}$$

with $\alpha_2(s,p,\pi) = \min\big\{\frac{s}{2s+2}, \frac{s/2+1/p-1/\pi}{s+1-2/\pi}\big\}$, provided $m(\nu) \geq \varpi > 0$ and up to a constant that depends on $s, p, \pi, \|\mathcal{Q}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^2)}$, $m(\nu)$ and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define $\varepsilon_2 = s\pi - (p-\pi)$. We have

$$\inf_{\widehat{\mathcal{Q}}_n}\sup_{\mathcal{Q}} \Big(\mathbb{E}\big[\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p\big]\Big)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_2(s,p,\pi)} & \text{if } \varepsilon_2 > 0 \\ \Big(\dfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_2(s,p,\pi)} & \text{if } \varepsilon_2 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of $\mathcal{Q}$ based on $(X_u)_{u\in\mathbb{T}_n}$ and the supremum is taken among all $\mathcal{Q}$ such that $\|\mathcal{Q}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^2)} \leq C$ and $m(\nu) \geq C'$ for some $C, C' > 0$.

3.4. **Estimation of the density of the $\mathbb{T}$-transition $\mathcal{P}$.** In this section we estimate $(x,y,z) \rightsquigarrow \mathcal{P}(x,y,z)$ for $(x,y,z) \in \mathcal{D}^3$ and $\mathcal{D}$ is a compact interval in $\mathcal{S} \subseteq \mathbb{R}$. In a first step, we estimate the density

$$f_{\mathcal{P}}(x,y,z) = \nu(x)\mathcal{P}(x,y,z)$$

of the distribution of (X_u, X_{u0}, X_{u1}) (when $\mathcal{L}(X_\emptyset) = \nu$) by

$$\widehat{f}_n(x,y,z) = \sum_{|\lambda|\leq J} \widehat{f}_{\lambda,n}\psi^3_\lambda(x,y,z),$$

with

$$\widehat{f}_{\lambda,n} = \mathcal{T}_{\lambda,\eta}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u\in\mathbb{T}_{n-1}} \psi^3_\lambda(X_u, X_{u0}, X_{u1})\Big),$$

and $\mathcal{T}_{\lambda,\eta}(\cdot)$ is the hard-threshold estimator defined in Section 3.2. In the same way as in the previous section, we can next estimate the density $\mathcal{P}$ of the $\mathbb{T}$-transition by

$$\widehat{\mathcal{P}}_n(x,y,z) = \frac{\widehat{f}_n(x,y,z)}{\max\{\widehat{\nu}_n(x), \varpi\}} \tag{13}$$

for some threshold $\varpi > 0$. Thus the estimator $\widehat{\mathcal{P}}_n$ is specified by J, η and ϖ.

Theorem 10. *Work under Assumptions 2, 3 and 6. Specify $\widehat{\mathcal{P}}_n$ with*

$$J = \tfrac{1}{3}\log_2 \frac{|\mathbb{T}_n|}{\log|\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{(\log|\mathbb{T}_n|)^2/|\mathbb{T}_n|}$$

for some $c > 0$ and $\varpi > 0$. For every $\pi \in [1,\infty], s > 3/\pi$ and $p \geq 1$, for large enough n and c and small enough ϖ, the following estimate holds

$$\Big(\mathbb{E}\big[\|\widehat{\mathcal{P}}_n - \mathcal{P}\|^p_{L^p(\mathcal{D}^3)}\big]\Big)^{1/p} \lesssim \Big(\frac{(\log|\mathbb{T}_n|)^2}{|\mathbb{T}_n|}\Big)^{\alpha_3(s,p,\pi)}, \tag{14}$$

with $\alpha_3(s,p,\pi) = \min\big\{\frac{s}{2s+3}, \frac{s/3+1/p-1/\pi}{2s/3+1-2/\pi}\big\}$, provided $m(\nu) \geq \varpi > 0$ and up to a constant that depends on $s, p, \pi, \|\mathcal{P}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^3)}$ and $m(\nu)$ and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define $\varepsilon_3 = \frac{s\pi}{3} - \frac{p-\pi}{2}$. We have

$$\inf_{\widehat{\mathcal{P}}_n}\sup_{\mathcal{P}} \Big(\mathbb{E}\big[\|\widehat{\mathcal{P}}_n - \mathcal{P}\|^p_{L^p(\mathcal{D}^3)}\big]\Big)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_3(s,p,\pi)} & \text{if} \quad \varepsilon_3 > 0 \\ \Big(\dfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_3(s,p,\pi)} & \text{if} \quad \varepsilon_3 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of $\mathcal{P}$ based on $(X_u)_{u\in\mathbb{T}_n}$ and the supremum is taken among all $\mathcal{P}$ such that $\|\mathcal{P}\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^3)} \leq C$ and $m(\nu) \geq C'$ for some $C, C' > 0$.

4. Applications

4.1. **Estimation of the size-dependent splitting rate in a growth-fragmentation model.** Recently, Doumic *et al.* [26] have studied the problem of estimating nonparametrically the size-dependent splitting rate in growth-fragmentation models (see *e.g.* the textbook of Perthame [37]). Stochastically, these are piecewise deterministic Marvov processes on trees that model the evolution of a population of cells or bacteria: to each node (or cell) $u \in \mathbb{T}$, we associate as trait $X_u \in \mathcal{S} \subset (0,\infty)$ the size at birth of the cell u. The evolution mechanism is described as follows: each cell grows exponentially with a common rate $\tau > 0$. A cell of size x splits into two newborn cells of size $x/2$ each (thus $X_{u0} = X_{u1}$ here), with a size-dependent splitting rate $B(x)$ for some $B : \mathcal{S} \to [0,\infty)$. Two newborn cells start a new life independently of each other. If ζ_u denotes the lifetime of the cell u, we thus have

$$\mathbb{P}\big(\zeta_u \in [t, t+dt)\big|\zeta_u \geq t, X_u = x\big) = B\big(x\exp(\tau t)\big)dt \tag{15}$$

and

$$X_u = \tfrac{1}{2}X_{u^-}\exp(\tau\zeta_{u^-}) \tag{16}$$

so that (15) and (16) entirely determine the evolution of the population. We are interested in estimating $x \rightsquigarrow B(x)$ for $x \in \mathcal{D}$ where $\mathcal{D} \subset \mathcal{S}$ is a given compact interval. The process $(X_u)_{u\in\mathbb{T}}$ is a bifurcating Markov chain with state space $\mathcal{S}$ and $\mathbb{T}$-transition any version of

$$\mathcal{P}_B(x, dy\,dz) = \mathbb{P}\big(X_{u0} \in dy, X_{u1} \in dz\,|X_{u^-} = x\big).$$

Moreover, using (15) and (16), (see for instance the derivation of Equation (11) in [26]), it is not difficult to check that

$$\mathcal{P}_B(x, dy\,dz) = Q_B(x, dy) \otimes \delta_y(dz)$$

where δ_y denotes the Dirac mass at y and

$$Q_B(x, dy) = \frac{B(2y)}{\tau y}\exp\Big(-\int_{x/2}^{y}\frac{B(2z)}{\tau z}dz\Big)\mathbf{1}_{\{y\geq x/2\}}dy. \tag{17}$$

If we assume moreover that $x \rightsquigarrow B(x)$ is continuous, then we have Assumption 2 with $\mathcal{Q} = Q_B$ and $\mathfrak{n}(dx) = dx$.

Now, let $\mathcal{S}$ be a bounded and open interval in $(0, \infty)$ such that $\sup \mathcal{S} > 2 \inf \mathcal{S}$. Pick $r \in \mathcal{S}$, $0 < L < \tau \log 2$ and introduce the function class

$$\mathcal{C}(r, L) = \Big\{ B : \mathcal{S} \to [0, \infty), \int^{\sup \mathcal{S}} \frac{B(x)}{x} dx = \infty, \int_{\inf \mathcal{S}}^{r} \frac{B(x)}{x} dx \leq L \Big\}.$$

By Theorem 1.3 in Hairer and Mattingly [29] and the explicit representation (17) for Q_B, one can check that for every $B \in \mathcal{C}(r, L)$, we have Assumption 3 with $\mathcal{Q} = Q_B$. In particular, we comply with the stringent requirement $\rho = \rho_B \leq C(r, L)$ for some $C(r, L) < 1/2$, *i.e.* uniformly over $\mathcal{C}(r, L)$. Finally, we know by Proposition 2 in Doumic *et al.* [26] – see in particular Equation (24) – that

$$B(x) = \frac{\tau x}{2} \frac{\nu_B(x/2)}{\int_{x/2}^{y} \nu_B(z) dz},$$

where ν_B denotes the unique invariant probability of the transition $\mathcal{Q} = Q_B$. This yields a strategy for estimating $x \rightsquigarrow B(x)$ via an estimator of $x \rightsquigarrow \nu_B(x)$. For a given compact interval $\mathcal{D} \subset \mathcal{S}$, define

$$\widehat{B}_n(x) = \frac{\tau x}{2} \frac{\widehat{\nu}_n(x/2)}{\big(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbf{1}_{\{x/2 \leq X_u < x\}}\big) \vee \varpi}, \tag{18}$$

where $\widehat{\nu}_n$ is the wavelet thresholding estimator given in Section 3.2 specified by a maximal resolution level J and a threshold η and $\varpi > 0$. As a consequence of Theorem 8 we obtain the following

Theorem 11. *Specify $\widehat{B}_n$ with*

$$J = \tfrac{1}{2} \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c\sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$$

for some $c > 0$. For every $B \in \mathcal{C}(r, L)$, $s > 0, \pi \in (0, \infty]$ and $p \geq 1$, large enough n and c and small enough ϖ, the following estimate holds

$$\Big(\mathbb{E}\big[\|\widehat{B}_n - B\|_{L^p(\mathcal{D})}^p\big]\Big)^{1/p} \lesssim \Big(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_1(s,p,\pi)},$$

with $\alpha_1(s, p, \pi) = \min\big\{\frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi}\big\}$, up to a constant that depends on $s, p, \pi, \|B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$, r and L and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define $\varepsilon_1 = s\pi - \frac{1}{2}(p - \pi)$. We have

$$\inf_{\widehat{B}_n} \sup_{B} \Big(\mathbb{E}\big[\|\widehat{B}_n - B\|_{L^p(\mathcal{D})}^p\big]\Big)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_1(s,p,\pi)} & \text{if} \quad \varepsilon_1 > 0 \\ \Big(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_1(s,p,\pi)} & \text{if} \quad \varepsilon_1 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of B based on $(X_u)_{u \in \mathbb{T}_n}$ and the supremum is taken among all $B \in \mathcal{C}(r, L)$ such that $\|B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})} \leq C$.

Two remarks are in order:

1) We improve on the results of Doumic *et al.* [26] in two directions: we have smoothness-adaptation (in the sense described in Remark 2) after Theorem 8 in Section 3 for several loss functions over various Besov smoothness classes, while [26] constructs a non-adapative estimator

for Hölder smoothness in squared-error loss; moreover, we prove that the obtained rate is (nearly) optimal in a minimax sense.
2) We unfortunately need to work under the quite stringent restriction that $\mathcal{S}$ is bounded in order to obtain the uniform ergodicity Assumption 3, see Remark 3) after Theorem 5 in Section 2.

4.2. **Bifurcating autoregressive process.** Bifurcating autoregressive processes (BAR), first introduced by Cowan and Staudte [16], are yet another stochastic model for understanding cell division. The trait X_u may represent the growth rate of a bacteria $u \in \mathbb{T}$ in a population of *Escherichia Coli* but other choices are obviously possible. Contrary to the growth-fragmentation model of Section 4.1 the trait (X_{u0}, X_{u1}) of the two newborn cells differ and are linked through the autoregressive dynamics

$$\left\{\begin{array}{l} X_{u0} = f_0(X_u) + \sigma_0(X_u)\varepsilon_{u0}, \\[2ex] X_{u1} = f_1(X_u) + \sigma_1(X_u)\varepsilon_{u1}, \end{array}\right. \tag{19}$$

initiated with $X_\emptyset$ and where

$$f_0, f_1 : \mathbb{R} \to \mathbb{R} \;\text{ and }\; \sigma_0, \sigma_1 : \mathbb{R} \to (0, \infty)$$

are functions and $(\varepsilon_{u_0}, \varepsilon_{u_1})_{u\in\mathbb{T}}$ are i.i.d. noise variables with common density function $G : \mathbb{R}^2 \to [0, \infty)$ that specify the model.

The process $(X_u)_{u\in\mathbb{T}}$ is a bifurcating Markov chain with state space $\mathcal{S} = \mathbb{R}$ and $\mathbb{T}$-transition

$$\mathcal{P}(x, dy\, dz) = G\Big(\sigma_0(x)^{-1}\big(y - f_0(x)\big), \sigma_1(x)^{-1}\big(z - f_1(x)\big)\Big) dy\, dz. \tag{20}$$

This model can be seen as an adaptation of nonlinear autoregressive model when the data have a binary tree structure. The original BAR process in [16] is defined for linear link functions f_0 and f_1 with $f_0 = f_1$ Several extensions have been studied from a parametric point of view, see *e.g.* Basawa and Huggins [2, 3] and Basawa and Zhou [4, 5]. More recently, de Saporta *et al.* [8, 19] introduces asymmetry and take into account missing data while Blandin [14], Bercu and Blandin [7], and de Saporta *et al.* [20] study an extension with random coefficients. Bitseki-Penda and Djellout [10] prove deviation inequalities and moderate deviations for estimators of parameters in linear BAR processes. From a nonparametric point of view, we mention the applications of [12] (Section 4) where deviations inequalities are derived for the Nadaraya-Watson type estimators of f_0 and f_1 with constant and known functions σ_0 and σ_1). A detailed nonparametric study of these estimators is carried out in Bitseki Penda and Olivier [13].

We focus here on the nonparametric estimation of the characteristics of the tagged-branch chain ν and $\mathcal{Q}$ and on the $\mathbb{T}$-transition $\mathcal{P}$, based on the observation of $(X_u)_{u\in\mathbb{T}_n}$ for some $n \geq 1$. Such an approach can be helpful for the subsequent study of goodness-of-fit tests for instance, when one needs to assess whether the data $(X_u)_{u\in\mathbb{T}}$ are generated by a model of the form (19) or not.

We set $G_0(x) = \int_{\mathcal{S}} G(x, y)dy$ and $G_1(y) = \int_{\mathcal{S}} G(x, y)dx$ for the marginals of G, and define, for any $M > 0$,

$$\delta(M) = \min\big\{ \inf_{|x|\leq M} G_0(x), \inf_{|x|\leq M} G_1(x)\big\}.$$

Assumption 12. *For some $\ell > 0$ and $\underline{\sigma} > 0$, we have*

$$\max\big\{ \sup_x |f_0(x)|, \sup_x |f_1(x)|\big\} \leq \ell < \infty$$

and

$$\min\big\{\inf_x \sigma_0(x), \inf_x \sigma_1(x)\big\} \geq \underline{\sigma} > 0.$$

Moreover, G_0 and G_1 are bounded and there exists $\mu > 0$ and $M > \ell/\underline{\sigma}$ such that $\delta\big((\mu+\ell)/\underline{\sigma}\big) > 0$ and $2(M\underline{\sigma} - \ell)\delta(M) > 1/2$.

Using that G_0 and G_1 are bounded, and (20), we readily check that Assumption 6 is satisfied. We also have Assumption 2 with $\mathfrak{n}(dx) = dx$ and

$$\mathcal{Q}(x,y) = \tfrac{1}{2}\Big(G_0\big(y - f_0(x)\big) + G_1\big(y - f_1(x)\big)\Big),$$

Assumption 12 implies Assumption 3 as well, as follows from an straightfroward adaptation of Lemma 25 in Bitseki Penda and Olivier [13]. Denoting by ν the invariant probability of $\mathcal{Q}$ we also have $m(\nu) > 0$ with $m(\nu)$ defined by (11), for every $\mathcal{D} \subset [-\mu, \mu]$, see the proof of Lemma 24 in [13]. As a consequence, the results stated in Theorems 8, 9 and 10 of Section 3 carry over to the setting of BAR processes satisfying Assumption 12. We thus readily obtain smoothness-adaptive estimators estimators for $\nu, \mathcal{Q}$ and $\mathcal{P}$ in this context and these results are new.

4.3. **Numerical illustration.** We focus on the growth-fragmentation model and reconstruct its size-dependent splitting rate. We consider a perturbation of the baseline splitting rate $\widetilde{B}(x) = x/(5-x)$ over the range $x \in \mathcal{S} = (0,5)$ of the form

$$B(x) = \widetilde{B}(x) + \mathfrak{c}\, T\big(2^j(x - \tfrac{7}{2})\big)$$

with $(\mathfrak{c}, j) = (3,1)$ or $(\mathfrak{c}, j) = (9,4)$, and where $T(x) = (1+x)\mathbf{1}_{\{-1 \leq x < 0\}} + (1-x)\mathbf{1}_{\{0 \leq x \leq 1\}}$ is a tent shaped function. Thus the trial splitting rate with parameter $(\mathfrak{c}, j) = (9,4)$ is more localized around $7/2$ and higher than the one associated with parameter $(\mathfrak{c}, j) = (3,1)$. One can easily check that both $\widetilde{B}$ and B belong to the class $\mathcal{C}(r, L)$ for an appropriate choice of (r, L). For a given B, we simulate $M = 100$ Monte Carlo trees up to the generation $n = 15$. To do so, we draw the size at birth of the initial cell $X_\emptyset$ uniformly in the interval $[1.25, 2.25]$, we fix the growth rate $\tau = 2$ and given a size at birth $X_u = x$, we pick X_{u0} according to the density $y \rightsquigarrow Q_B(x,y)$ defined by (17) using a rejection sampling algorithm (with proposition density $y \rightsquigarrow Q_{\widetilde{B}}(x,y)$) and set $X_{u1} = X_{u0}$. Figure 1 illustrates quantitatively how fast the decorrelation on the tree occurs.

Computational aspects of statistical estimation using wavelets can be found in Härdle *et al.*, Chapter 12 of [30]. We implement the estimator $\widehat{B}_n$ defined by (18) using the Matlab wavelet toolbox. We take a wavelet filter corresponding to compactly supported Daubechies wavelets of order 8. As specified in Theorem 11, the maximal resolution level J is chosen as $\frac{1}{2}\log_2(|\mathbb{T}_n|/\log|\mathbb{T}_n|)$ and we threshold the coefficients $\widehat{\nu}_{\lambda,n}$ which are too small by hard thresholding. We choose the threshold proportional to $\sqrt{\log|\mathbb{T}_n|/|\mathbb{T}_n|}$ (and we calibrate the constant to 10 or 15 for respectively the two trial splitting rates, mainly by visual inspection). We evaluate $\widehat{B}_n$ on a regular grid of $\mathcal{D} = [1.5, 4.8]$ with mesh $\Delta x = (|\mathbb{T}_n|)^{-1/2}$. For each sample we compute the empirical error

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

where $\|\cdot\|_{\Delta x}$ denotes the discrete L^2-norm over the numerical sampling and sum up the results through the mean-empirical error $\bar{e} = M^{-1}\sum_{i=1}^M e_i$, together with the empirical standard deviation $\big(M^{-1}\sum_{i=1}^M (e_i - \bar{e})^2\big)^{1/2}$.

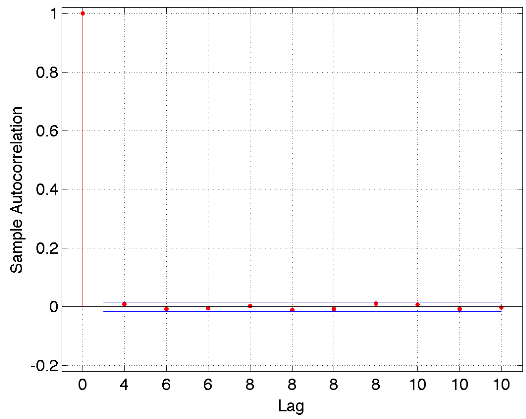

FIGURE 1. *Sample autocorrelation of ordered* $(X_{u0}, u \in \mathbb{G}_{n-1})$ *for* $n = 15$. *Note: due to the binary tree structure the lags are* $\{4, 6, 6, \dots\}$. *As expected, we observe a fast decorrelation.*

Table 1 displays the numerical results we obtained, also giving the compression rate (columns %) defined as the number of wavelet coefficients put to zero divided by the total number of coefficients. We choose an oracle error as benchmark: the oracle estimator is computed by picking the best resolution level J^* with no coefficient thresholded. We also display the results when constructing $\widehat{B}_n$ with $\mathbb{G}_n$ (instead of $\mathbb{T}_n$), in which case an analog of Theorem 11 holds. For the large spike, the thresholding estimator behaves quite well compared to the oracle for a large spike and achieves the same performance for a high spike.

		$n = 12$				**$n = 15$**			
		Oracle		**Threshold est.**		**Oracle**		**Threshold est.**	
		Mean (sd.)	J^*	**Mean (sd.)**	%	**Mean (sd.)**	J^*	**Mean (sd.)**	%
Large spike	$\mathbb{T}_n$	0.0677 (0.0159)	5	0.1020 (0.0196)	96.6	0.0324 (0.0055)	6	0.0502 (0.0055)	97.1
	$\mathbb{G}_n$	0.0933 (0.0202)	5	0.1454 (0.0267)	97.9	0.0453 (0.0081)	6	0.0728 (0.0097)	96.7
High spike	$\mathbb{T}_n$	0.1343 (0.0180)	7	0.1281 (0.0163)	97.4	0.0586 (0.0059)	8	0.0596 (0.0060)	97.7
	$\mathbb{G}_n$	0.1556 (0.0222)	7	0.1676 (0.0228)	97.7	0.0787 (0.0079)	8	0.0847 (0.0087)	97.9

TABLE 1. Mean empirical relative error $\bar{e}$ and its standard deviation, with respect to n, for the trial splitting rate B specified by $(\mathfrak{c}, j) = (3, 1)$ (large spike) or $(\mathfrak{c}, j) = (4, 9)$ (high spike) reconstructed over the interval $\mathcal{D} = [1.5, 4.8]$ by the estimator $\widehat{B}_n$. Note: for $n = 15$, $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$ and $\frac{1}{2}|\mathbb{G}_n| = 16\ 384$; for $n = 12$, $\frac{1}{2}|\mathbb{T}_n| = 4\ 095$ and $\frac{1}{2}|\mathbb{G}_n| = 2\ 048$.

Figure 2 and Figure 3 show the reconstruction of the size-dependent splitting rate B and the invariant measure ν_B in the two cases (large or high spike) for one typical sample of size $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$. In both cases, the spike is well reconstructed and so are the discontinuities in the derivative of B. As expected, the spike being localized around $\frac{7}{2}$ for B, we detect it around $\frac{7}{4}$ for the invariant measure of the sizes at birth ν_B. The large spike concentrates approximately 50% of the mass of ν_B whereas the large only concentrates 20% of the mass of ν_B.

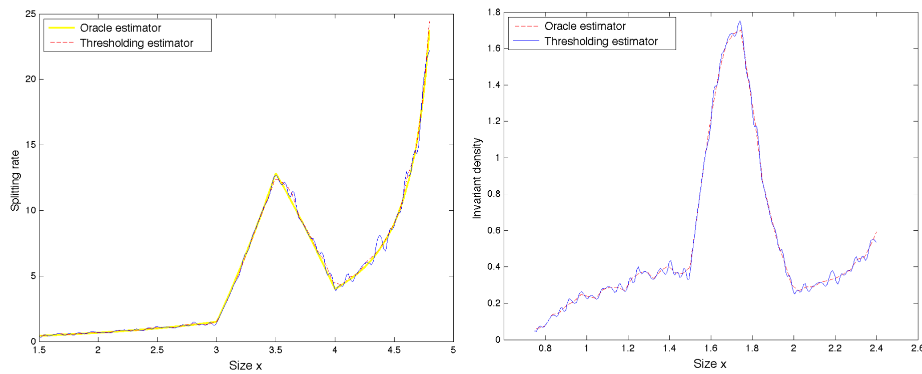

FIGURE 2. *Large spike: reconstruction of the trial splitting rate B specified by $(\mathfrak{c}, j) = (3, 1)$ over $\mathcal{D} = [1.5, 4.8]$ and reconstruction of ν_B over $\mathcal{D}/2$ based on one sample $(X_u, u \in \mathbb{T}_n)$ for $n = 15$ (i.e. $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$).*

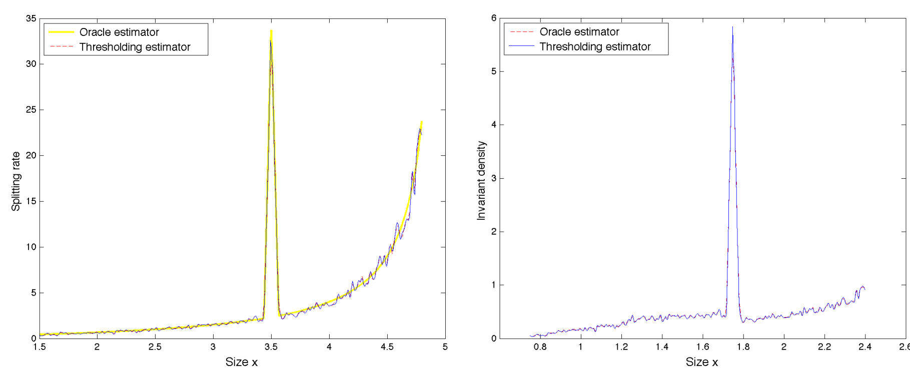

FIGURE 3. *High spike : reconstruction of the trial splitting rate B specified by $(\mathfrak{c}, j) = (9, 4)$ over $\mathcal{D} = [1.5, 4.8]$ and reconstruction of ν_B over $\mathcal{D}/2$ based on one sample $(X_u, u \in \mathbb{T}_n)$ for $n = 15$ (i.e. $\frac{1}{2}|\mathbb{T}_n| = 32\ 767$).*

5. PROOFS

5.1. **Proof of Theorem 4 (i).** Let $g : \mathcal{S} \to \mathbb{R}$ such that $|g|_1 < \infty$. Set $\nu(g) = \int_{\mathcal{S}} g(x)\nu(dx)$ and $\widetilde{g} = g - \nu(g)$. Let $n \geq 2$. By the usual Chernoff bound argument, for every $\lambda > 0$, we have

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u) \geq \delta\Big) \leq \exp\big(-\lambda|\mathbb{G}_n|\delta\big)\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u)\big)\Big]. \tag{21}$$

Step 1. We have

$$\begin{aligned}\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-1}\Big] &= \mathbb{E}\Big[\prod_{u\in\mathbb{G}_{n-1}}\exp\Big(\lambda\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\\ &= \prod_{u\in\mathbb{G}_{n-1}}\mathbb{E}\Big[\exp\Big(\lambda\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\end{aligned}$$

thanks to the conditional independence of the $(X_{u0},X_{u1})_{u\in\mathbb{G}_{n-1}}$ given $\mathcal{F}_{n-1}$, as follows from Definition 1. We rewrite this last term as

$$\prod_{u\in\mathbb{G}_{n-1}}\mathbb{E}\Big[\exp\Big(\lambda\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big)\Big)\big|\mathcal{F}_{n-1}\Big]\exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big),$$

inserting the $\mathcal{F}_{n-1}$-measurable random variable $2\mathcal{Q}\widetilde{g}(X_u)$ for $u\in\mathbb{G}_{n-1}$. Moreover, the bifurcating struture of $(X_u)_{u\in\mathbb{T}}$ implies

$$\mathbb{E}\big[\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big|\mathcal{F}_{n-1}\big]=0,\quad u\in\mathbb{G}_{n-1}, \tag{22}$$

since $\mathcal{Q}=\frac{1}{2}(\mathcal{P}_0+\mathcal{P}_1)$. We will also need the following bound, proof of which is delayed until Appendix

Lemma 13. *Work under Assumptions 2 and 3. For all $r=0,\dots,n-1$ and $u\in\mathbb{G}_{n-r-1}$, we have*

$$\big|2^r\big(\mathcal{Q}^r\widetilde{g}(X_{u0})+\mathcal{Q}^r\widetilde{g}(X_{u1})-2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big)\big|\leq c_1|g|_\infty$$

and

$$\mathbb{E}\Big[\Big(2^r\big(\mathcal{Q}^r\widetilde{g}(X_{u0})+\mathcal{Q}^r\widetilde{g}(X_{u1})-2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big)\Big)^2\Big|\mathcal{F}_{n-r-1}\Big]\leq c_2\sigma_r^2(g),$$

with $c_1=4\max\big\{1+R\rho,R(1+\rho)\big\}$, $c_2=4\max\{|\mathcal{Q}|_{\mathcal{D}},4|\mathcal{Q}|_{\mathcal{D}}^2,4R^2(1+\rho)^2\}$ and

$$\sigma_r^2(g)=\left\{\begin{array}{ll}|g|_2^2 & r=0,\\ \min\big\{|g|_1^2 2^{2r},|g|_\infty^2(2\rho)^{2r}\big\} & r=1,\dots,n-1.\end{array}\right. \tag{23}$$

(Recall that $|\mathcal{Q}|_{\mathcal{D}}=\sup_{x\in\mathcal{S},y\in\mathcal{D}}\mathcal{Q}(x,y)$ and R,ρ are defined via Assumption 3.)

In view of (22) and Lemma 13 for $r=0$, we plan to use the bound

$$\mathbb{E}\big[\exp(\lambda Z)\big]\leq\exp\Big(\frac{\lambda^2\sigma^2}{2(1-\lambda M/3)}\Big) \tag{24}$$

valid for any $\lambda\in(0,3/M)$, any random variable Z such that $|Z|\leq M$, $\mathbb{E}[Z]=0$ and $\mathbb{E}[Z^2]\leq\sigma^2$. Thus, for any $\lambda\in\big(0,3/c_1|g|_\infty\big)$ and any $u\in\mathbb{G}_{n-1}$, with $Z=\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)$, we obtain

$$\mathbb{E}\Big[\exp\Big(\lambda\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\leq\exp\Big(\frac{\lambda^2c_2\sigma_0^2(g)}{2(1-\lambda c_1|g|_\infty/3)}\Big).$$

It follows that

$$\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-1}\Big]\leq\exp\Big(\frac{\lambda^2c_2\sigma_0^2(g)|\mathbb{G}_{n-1}|}{2(1-\lambda c_1|g|_\infty/3)}\Big)\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big). \tag{25}$$

Step 2. We iterate the procedure in Step 1. Conditioning with respect to $\mathcal{F}_{n-2}$, we need to control

$$\mathbb{E}\Big[\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-2}\Big],$$

and more generally, for $1 \leq r \leq n-1$:

$$\begin{aligned}
&\mathbb{E}\Big[\prod_{u \in \mathbb{G}_{n-r}} \exp\big(\lambda 2^r \mathcal{Q}^r \widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-r-1}\Big] \\
&= \prod_{u \in \mathbb{G}_{n-r-1}} \mathbb{E}\Big[\exp\Big(\lambda 2^r\big(\mathcal{Q}^r \widetilde{g}(X_{u0}) + \mathcal{Q}^r \widetilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big)\Big)\Big|\mathcal{F}_{n-r-1}\Big] \\
&\quad \times \exp\big(\lambda 2^{r+1}\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big),
\end{aligned}$$

the last equality being obtained thanks to the conditional independence of the $(X_{u0}, X_{u1})_{u \in \mathbb{G}_{n-r-1}}$ given $\mathcal{F}_{n-r-1}$. We plan to use (24) again: for $u \in \mathbb{G}_{n-r-1}$, we have

$$\mathbb{E}\big[2^r\big(\mathcal{Q}^r \widetilde{g}(X_{u0}) + \mathcal{Q}^r \widetilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big)\big|\mathcal{F}_{n-r-1}\big] = 0$$

and the conditional variance given $\mathcal{F}_{n-r-1}$ can be controlled using Lemma 13. Using recursively (24), for $r = 1, \ldots, n-1$,

$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_{n-1}} \exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_0\Big] \leq \prod_{r=1}^{n-1}\exp\Big(\frac{\lambda^2 c_2 \sigma_r^2(g)|\mathbb{G}_{n-r-1}|}{2\big(1-\lambda c_1|g|_\infty/3\big)}\Big)\exp\big(\lambda 2^n \mathcal{Q}^n \widetilde{g}(X_\emptyset)\big)$$

for $\lambda \in \big(0, 3/c_1|g|_\infty\big)$. By Assumption 3,

$$\exp\big(\lambda 2^n \mathcal{Q}^n \widetilde{g}(X_\emptyset)\big) \leq \exp(\lambda 2^n R(2|g|_\infty)\rho^n) \leq \exp(\lambda 2R|g|_\infty)$$

since $\rho < 1/2$. In conclusion

$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_{n-1}} \exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big)\Big] \leq \exp\Big(\frac{\lambda^2 c_2 \sum_{r=1}^{n-1}\sigma_r^2(g)|\mathbb{G}_{n-r-1}|}{2\big(1-\lambda c_1|g|_\infty/3\big)}\Big)\exp(\lambda 2R|g|_\infty).$$

Step 3. Let $1 \leq \ell \leq n-1$. By definition of $\sigma_r^2(g)$ – recall (23) – and using the fact that $(2\rho)^{2r} \leq 1$, since moreover $|\mathbb{G}_{n-r-1}| = 2^{n-r-1}$, we successively obtain

$$\begin{aligned}
\sum_{r=1}^{n-1}\sigma_r^2(g)2^{n-r-1} &\leq 2^{n-1}\big(|g|_1^2\sum_{r=1}^{\ell}2^r + |g|_\infty^2\sum_{r=\ell+1}^{n-1}2^{-r}(2\rho)^{2r}\big) \\
&\leq 2^n\big(|g|_1^2 2^\ell + |g|_\infty^2 2^{-\ell}\big) \\
&\leq |\mathbb{G}_n|\phi_n(g)
\end{aligned}$$

for an appropriate choice of ℓ, with $\phi_n(g) = \min_{1 \leq \ell \leq n-1}\big(|g|_1^2 2^\ell + |g|_\infty^2 2^{-\ell}\big)$. It follows that

$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_{n-1}} \exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big)\Big] \leq \exp\Big(\frac{\lambda^2 c_2|\mathbb{G}_n|\phi_n(g)}{2\big(1-\lambda c_1|g|_\infty/3\big)} + \lambda 2R|g|_\infty\Big). \tag{26}$$

Step 4. Putting together the estimates (25) and (26) and coming back to (21), we obtain

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u \in \mathbb{G}_n}\widetilde{g}(X_u) \geq \delta\Big) \leq \exp\Big(-\lambda|\mathbb{G}_n|\delta + \frac{\lambda^2 c_2|\mathbb{G}_n|\Sigma_{1,n}(g)}{2\big(1-\lambda c_1|g|_\infty/3\big)} + \lambda 2R|g|_\infty\Big)$$

with $\Sigma_{1,n}(g) = |g|_2^2 + \phi_n(g)$ for $n \geq 2$ and $\Sigma_{1,1}(g) = \sigma_0^2(g) = |g|_2^2$. Since δ is such that $2R|g|_\infty \leq |\mathbb{G}_n|\delta/2$, we obtain

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u \in \mathbb{G}_n}\widetilde{g}(X_u) \geq \delta\Big) \leq \exp\Big(-\lambda|\mathbb{G}_n|\frac{\delta}{2} + \frac{\lambda^2 c_2|\mathbb{G}_n|\Sigma_{1,n}(g)}{2\big(1-\lambda c_1|g|_\infty/3\big)}\Big).$$

The admissible choice $\lambda = \delta/\big(\frac{2}{3}\delta c_1|g|_\infty + 2c_2\Sigma_{1,n}(g)\big)$ yields the result.

5.2. **Proof of Theorem 4 (ii).** *Step 1.* Similarly to (21), we plan to use

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}\widetilde{g}(X_u)\geq\delta\Big)\leq \exp\big(-\lambda|\mathbb{T}_n|\delta\big)\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{T}_n}\widetilde{g}(X_u)\big)\Big] \tag{27}$$

for a specific choice of $\lambda > 0$. We first need to control

$$\mathbb{E}\big[\exp\big(\lambda\sum_{u\in\mathbb{T}_n}\widetilde{g}(X_u)\big)\big|\mathcal{F}_{n-1}\big] = \prod_{u\in\mathbb{T}_{n-1}}\exp\big(\lambda\widetilde{g}(X_u)\big)\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-1}\Big].$$

Using (25) to control $\mathbb{E}\big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(X_u)\big)\big|\mathcal{F}_{n-1}\big]$, we obtain

$$\begin{aligned}\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{T}_n}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-1}\Big]&\\ \leq \exp\Big(\frac{\lambda^2c_2\sigma_0^2(g)|\mathbb{G}_{n-1}|}{2(1-\lambda c_1|g|_\infty/3)}\Big)&\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}\widetilde{g}(X_u)\big)\prod_{u\in\mathbb{T}_{n-1}}\exp\big(\lambda\widetilde{g}(X_u)\big).\end{aligned}$$

Step 2. We iterate the procedure. At the second step, conditioning w.r.t. $\mathcal{F}_{n-2}$, we need to control

$$\mathbb{E}\Big[\prod_{u\in\mathbb{T}_{n-2}}\exp\big(\lambda\widetilde{g}(X_u)\big)\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda\widetilde{g}(X_u)+2\lambda\mathcal{Q}\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-2}\Big]$$

and more generally, at the $(r+1)$-th step (for $1\leq r\leq n-1$), we need to control

$$\begin{aligned}\mathbb{E}\Big[\prod_{u\in\mathbb{T}_{n-r-1}}&\exp\big(\lambda\widetilde{g}(X_u)\big)\prod_{u\in\mathbb{G}_{n-r}}\exp\big(\lambda\sum_{m=0}^{r}2^m\mathcal{Q}^m\widetilde{g}(X_u)\big)\Big|\mathcal{F}_{n-r-1}\Big]\\ &=\prod_{u\in\mathbb{T}_{n-r-2}}\exp\big(\lambda\widetilde{g}(X_u)\big)\prod_{u\in\mathbb{G}_{n-r-1}}\exp\big(\lambda\sum_{m=0}^{r+1}2^m\mathcal{Q}^m\widetilde{g}(X_u)\big)\\ &\qquad\qquad\times\mathbb{E}\big[\exp\big(\lambda\,\Upsilon_r(X_u,X_{u0},X_{u1})\big)\big|\mathcal{F}_{n-r-1}\big],\end{aligned}$$

where we set

$$\Upsilon_r(X_u,X_{u0},X_{u1})=\sum_{m=0}^{r}2^m\big(\mathcal{Q}^m\widetilde{g}(X_{u0})+\mathcal{Q}^m\widetilde{g}(X_{u1})-2\mathcal{Q}^{m+1}\widetilde{g}(X_u)\big).$$

This representation successively follows from the $\mathcal{F}_{n-r-1}$-measurability of the random variable $\prod_{u\in\mathbb{T}_{n-r-1}}\exp\big(\lambda\widetilde{g}(X_u)\big)$, the identity

$$\prod_{u\in\mathbb{G}_{n-r}}\exp\big(F(X_u)\big)=\prod_{u\in\mathbb{G}_{n-r-1}}\exp\big(F(X_{u0})+F(X_{u1})\big),$$

the independence of $(X_{u0},X_{u1})_{u\in\mathbb{G}_{n-r-1}}$ conditional on $\mathcal{F}_{n-r-1}$ and finally the introduction of the term $2\sum_{m=0}^{r}2^m\mathcal{Q}^{m+1}\widetilde{g}(X_u)$.

We have, for $u\in\mathbb{G}_{n-r-1}$

$$\mathbb{E}\big[\Upsilon_r(X_u,X_{u0},X_{u1})\big|\mathcal{F}_{n-r-1}\big]=0,$$

and we prove in Appendix the following bound

Lemma 14. *For any $r = 1, \dots, n-1$, $u \in \mathbb{G}_{n-r-1}$, we have*

$$|\Upsilon_r(X_u, X_{u0}, X_{u1})| \leq c_3 |g|_\infty$$

and

$$\mathbb{E}\big[\Upsilon_r(X_u, X_{u0}, X_{u1})^2 \big| \mathcal{F}_{n-r-1}\big] \leq c_4 \sigma_r^2(g) < \infty$$

where $c_3 = 4R(1+\rho)(1-2\rho)^{-1}$, $c_4 = 12 \max\big\{|\mathcal{Q}|_{\mathcal{D}}, 16|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2(1-2\rho)^{-2}\big\}$ and

$$\sigma_r^2(g) = |g|_2^2 + \min_{\ell \geq 1} \big(|g|_1^2 2^{2(\ell \wedge r)} + |g|_\infty^2 (2\rho)^{2\ell} \mathbf{1}_{\{r > \ell\}}\big). \tag{28}$$

(Recall that $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x,y)$ and R, ρ are defined via Assumption 3.)

In the same way as for Step 2 in the proof of Theorem 4 (i), we apply recursively (24) for $r = 1, \dots, n-1$ to obtain

$$\mathbb{E}\big[\exp\big(\lambda \sum_{u \in \mathbb{T}_n} \widetilde{g}(X_u)\big) \big| \mathcal{F}_0\big] \leq \prod_{r=0}^{n-1} \exp\Big(\frac{c_4 \lambda^2 \sigma_r^2(g) |\mathbb{G}_{n-r-1}|}{2(1 - c_3' \lambda |g|_\infty / 3)}\Big) \exp\big(\lambda \sum_{m=0}^{n} 2^m \mathcal{Q}^m \widetilde{g}(X_\emptyset)\big),$$

if $\lambda \in \big(0, 3/c_3'|g|_\infty\big)$ with $c_3' = \max\{c_1, c_3\} = 4\max\{1 + R\rho, R(1+\rho)(1-2\rho)^{-1}\}$ and $\sigma_0^2(g) = |g|_2^2$ in order to include Step 1 (we use $c_4 \geq c_2$ as well). Now, by Assumption 3, this last term can be bounded by

$$\exp\big(\lambda \sum_{m=0}^{n} 2^m (R|\widetilde{g}|_\infty \rho^m)\big) \leq \exp\big(\lambda 2R(1-2\rho)^{-1}|g|_\infty\big)$$

since $\rho < 1/2$. Since $|\mathbb{G}_{n-r-1}| = 2^{n-r-1}$, by definition of $\sigma_r^2(g)$ – recall (28) – for any $1 \leq \ell \leq n-1$ and using moreover that $(2\rho)^\ell \leq 1$, we obtain

$$\begin{aligned}
&\sum_{r=0}^{n-1} \sigma_r^2(g) |\mathbb{G}_{n-r-1}| \\
&\leq 2^{n-1}\Big(|g|_2^2 \sum_{r=0}^{n-1} 2^{-r} + |g|_1^2 \Big(\sum_{r=1}^{\ell} 2^{2r} 2^{-r} + \sum_{r=\ell+1}^{n-1} 2^{2\ell} 2^{-r}\Big) + |g|_\infty^2 \sum_{r=\ell+1}^{n-1} 2^{-r}\Big) \\
&\leq |\mathbb{T}_n| \Sigma_{1,n}(g),
\end{aligned}$$

where $\Sigma_{1,n}(g)$ is defined in (3). Thus

$$\mathbb{E}\Big[\exp\big(\lambda \sum_{u \in \mathbb{T}_n} \widetilde{g}(X_u)\big)\Big] \leq \exp\Big(\frac{c_4 \lambda^2 |\mathbb{T}_n| \Sigma_{1,n}(g)}{2\big(1 - c_3' \lambda |g|_\infty / 3\big)} + \lambda 2R(1-2\rho)^{-1}|g|_\infty\Big).$$

Step 3. Coming back to (27), for $\delta > 0$ such that $2R(1-2\rho)^{-1}|g|_\infty \leq |\mathbb{T}_n|\delta/2$, we obtain

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \widetilde{g}(X_u) \geq \delta\Big) \leq \exp\Big(-\lambda |\mathbb{T}_n| \frac{\delta}{2} + \frac{c_4 \lambda^2 |\mathbb{T}_n| \Sigma_{1,n}(g)}{2\big(1 - c_3' \lambda |g|_\infty / 3\big)}\Big).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

5.3. **Proof of Theorem 5 (i).** The strategy of proof is similar as for Theorem 4. Let $g : \mathcal{S}^3 \to \mathbb{R}$ such that $|g|_1 < \infty$ and set $\widetilde{g} = g - \nu(\mathcal{P}g)$. Let $n \geq 2$ (if $n = 1$, set $\Sigma_{2,1}(g) = |\mathcal{Q}(\mathcal{P}g)|_\infty$). Introduce the notation $\Delta_u = (X_u, X_{u0}, X_{u1})$ for simplicity. For every $\lambda > 0$, the usual Chernoff bound reads

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n}\widetilde{g}(\Delta_u) \geq \delta\Big) \leq \exp(-\lambda|\mathbb{G}_n|\delta)\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(\Delta_u)\big)\Big]. \tag{29}$$

Step 1. We first need to control

$$\begin{aligned}\mathbb{E}\big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(\Delta_u)\big)\big|\mathcal{F}_{n-1}\big] &= \mathbb{E}\Big[\prod_{u\in\mathbb{G}_{n-1}}\exp\Big(\lambda\big(\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\\ &= \prod_{u\in\mathbb{G}_{n-1}}\mathbb{E}\Big[\exp\Big(\lambda\big(\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\end{aligned}$$

using the conditional independence of the $(\Delta_{u0}, \Delta_{u1})$ for $u \in \mathbb{G}_{n-1}$ given $\mathcal{F}_{n-1}$. Inserting the term $2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)$, this last quantity ia also equal to

$$\prod_{u\in\mathbb{G}_{n-1}}\mathbb{E}\Big[\exp\Big(\lambda\big(\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})-2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big)\Big)\Big|\mathcal{F}_{n-1}\Big]\exp\big(\lambda 2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big).$$

For $u \in \mathbb{G}_{n-1}$ we successively have

$$\mathbb{E}\big[\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})-2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big|\mathcal{F}_{n-1}\big]=0,$$

$$|\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})-2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)| \leq 4(1+R\rho)|g|_\infty$$

and

$$\mathbb{E}\big[\big(\widetilde{g}(\Delta_{u0})+\widetilde{g}(\Delta_{u1})-2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big)^2\big|\mathcal{F}_{n-1}\big] \leq 4|\mathcal{Q}|_\mathcal{D}|\mathcal{P}g^2|_1,$$

with $|\mathcal{Q}|_\mathcal{D} = \sup_{x\in\mathcal{S},y\in\mathcal{D}}\mathcal{Q}(x,y)$ and R, ρ defined via Assumption 3. The first equality is obtained by conditioning first on $\mathcal{F}_n$ then on $\mathcal{F}_{n-1}$. The last two estimates are obtained in the same line as the proof of Lemma 13 for $r = 0$, using in particular $\mathcal{Q}(\mathcal{P}g^2)(x) = \int_\mathcal{S}\mathcal{P}g^2(y)\mathcal{Q}(x,y)\mathfrak{n}(dy) \leq |\mathcal{Q}|_\mathcal{D}|\mathcal{P}g^2|_1$ since $\mathcal{P}g^2$ vanishes outside $\mathcal{D}$.

Finally, thanks to (24) with $Z = \widetilde{g}(\Delta_{u0}) + \widetilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)$, we infer

$$\mathbb{E}\big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(\Delta_u)\big)\big|\mathcal{F}_{n-1}\big] \leq \exp\Big(\frac{\lambda^2 4|\mathcal{Q}|_\mathcal{D}|\mathcal{P}g^2|_1}{2(1-\lambda 4(1+R\rho)|g|_\infty/3)}\Big)\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big) \tag{30}$$

for $\lambda \in \big(0, 3/(4(1+R\rho)|g|_\infty)\big)$.

Step 2. We wish to control $\mathbb{E}\big[\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big)\big]$. We are back to Step 2 and Step 3 of the proof of Theorem 4 (i), replacing $\widetilde{g}$ by $\mathcal{P}\widetilde{g}$, which satisfies $\nu(\mathcal{P}\widetilde{g}) = 0$. Equation (26) entails

$$\mathbb{E}\Big[\prod_{u\in\mathbb{G}_{n-1}}\exp\big(\lambda 2\mathcal{Q}(\mathcal{P}\widetilde{g})(X_u)\big)\Big] \leq \exp\Big(\frac{\lambda^2 c_2|\mathbb{G}_n|\phi_n(\mathcal{P}g)}{2\big(1-\lambda c_1|\mathcal{P}g|_\infty/3\big)} + \lambda 2R|\mathcal{P}g|_\infty\Big) \tag{31}$$

with $\phi_n(\mathcal{P}g) = \min_{1\leq\ell\leq n-1}\big(|\mathcal{P}g|_1^2 2^\ell + |\mathcal{P}g|_\infty^2 2^{-\ell}\big)$ and $c_1 = 4\max\big\{1+R\rho, R(1+\rho)\big\}$, $c_2 = 4\max\{|\mathcal{Q}|_\mathcal{D}, 4|\mathcal{Q}|_\mathcal{D}^2, 4R^2(1+\rho)^2\}$.

Step 3. Putting together (30) and (31), we obtain

$$\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{G}_n}\widetilde{g}(\Delta_u)\big)\Big] \leq \exp\Big(\frac{\lambda^2 c_2|\mathbb{G}_n|\Sigma_{2,n}(g)}{2(1-\lambda c_1|g|_\infty/3)} + \lambda 2R|\mathcal{P}g|_\infty\Big) \tag{32}$$

with $\Sigma_{2,n}(g) = |\mathcal{P}g^2|_1 + \phi_n(\mathcal{P}g)$ and using moreover $|g|_\infty \geq |\mathcal{P}g|_\infty$ and $c_1 \geq 4(1+R\rho)$. Back to (29), since $2R|\mathcal{P}g|_\infty \leq |\mathbb{G}_n|\delta/2$ we finally infer

$$\mathbb{P}\Big(\frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} g(\Delta_u) - \nu(\mathcal{P}g) \geq \delta\Big) \leq \exp\Big(-\lambda|\mathbb{G}_n|\frac{\delta}{2} + \frac{\lambda^2 c_2 |\mathbb{G}_n|\Sigma_{2,n}(g)}{2(1-\lambda c_1 |g|_\infty/3)}\Big).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

5.4. **Proof of Theorem 5 (ii).** In the same way as before, for every $\lambda > 0$,

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u\in\mathbb{T}_{n-1}} \widetilde{g}(\Delta_u) \geq \delta\Big) \leq e^{-\lambda|\mathbb{T}_{n-1}|\delta}\,\mathbb{E}\Big[\exp\big(\lambda\sum_{u\in\mathbb{T}_{n-1}}\widetilde{g}(\Delta_u)\big)\Big]. \tag{33}$$

Introduce $\Sigma'_{2,0}(g) = |\mathcal{P}g^2|_1$ and

$$\Sigma'_{2,n}(g) = |\mathcal{P}g^2|_1 + \inf_{\ell\geq 1}\big(|\mathcal{P}g|_1^2 2^{\ell\wedge(n-1)} + |\mathcal{P}g|_\infty^2 2^{-\ell}\mathbf{1}_{\{\ell<n-1\}}\big),\quad \text{for } n \geq 1.$$

It is not difficult to check that (32) is still valid when replacing $\Sigma_{2,n}$ by $\Sigma'_{2,n}$. We plan to successively expand the sum over the whole tree $\mathbb{T}_{n-1}$ into sums over each generation $\mathbb{G}_m$ for $m = 0,\ldots,n-1$, apply Hölder inequality, apply inequality (32) repeatedly (with $\Sigma'_{2,m}$) together with the bound

$$\sum_{m=0}^{n-1}|\mathbb{G}_m|\Sigma'_{2,m}(g) \leq |\mathbb{T}_{n-1}|\Sigma_{2,n-1}(g).$$

We thus obtain

$$\begin{aligned}
\mathbb{E}\big[\exp\big(\lambda\sum_{u\in\mathbb{T}_{n-1}}\widetilde{g}(\Delta_u)\big)\big] &= \mathbb{E}\big[\prod_{m=0}^{n-1}\exp\big(\lambda\sum_{u\in\mathbb{G}_m}\widetilde{g}(\Delta_u)\big)\big] \\
&\leq \Big(\mathbb{E}\big[\exp\big(n\lambda\widetilde{g}(\Delta_\emptyset)\big)\big]\prod_{m=1}^{n-1}\mathbb{E}\big[\exp\big(n\lambda\sum_{u\in\mathbb{G}_m}\widetilde{g}(\Delta_u)\big)\big]\Big)^{1/n} \\
&\leq \Big(\exp\big(n\lambda 2|g|_\infty\big)\prod_{m=1}^{n-1}\exp\big(\frac{(n\lambda)^2 c_2|\mathbb{G}_m|\Sigma'_{2,m}(g)}{2(1-(n\lambda)c_1|g|_\infty/3)} + (n\lambda)2R|\mathcal{P}g|_\infty\big)\Big)^{1/n} \\
&\leq \exp\Big(\frac{\lambda^2 c_2 n|\mathbb{T}_{n-1}|\Sigma_{2,n-1}(g)}{2(1-c_1(n\lambda)|g|_\infty/3)} + 2\lambda(nR|\mathcal{P}g|_\infty + |g|_\infty)\Big).
\end{aligned}$$

Coming back to (33) and using $2(nR|\mathcal{P}g|_\infty + |g|_\infty) \leq |\mathbb{T}_{n-1}|\delta/2$, we obtain

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_{n-1}|}\sum_{u\in\mathbb{T}_{n-1}}\widetilde{g}(\Delta_u) \geq \delta\Big) \leq \exp\Big(-\lambda|\mathbb{T}_{n-1}|\frac{\delta}{2} + \frac{\lambda^2 c_2 n|\mathbb{T}_{n-1}|\Sigma_{2,n-1}(g)}{2(1-(n\lambda)c_1|g|_\infty/3)}\Big).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

5.5. **Proof of Theorem 8.** Put $c(n) = (\log|\mathbb{T}_n|/|\mathbb{T}_n|)^{1/2}$ and note that the maximal resolution $J = J_n$ is such that $2^{J_n} \sim c(n)^{-2}$. Theorem 8 is a consequence of the general theory of wavelet threshold estimators, see Kerkyacharian and Picard [32]. We first claim that the following moment bounds and moderate deviation inequalities hold: for every $p \geq 1$,

$$\mathbb{E}\big[|\widehat{\nu}_{\lambda,n} - \nu_\lambda|^p\big] \lesssim c(n)^p \;\text{ for every }\; |\lambda| \leq J_n \tag{34}$$

and

$$\mathbb{P}\big(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq p\varkappa c(n)\big) \leq c(n)^{2p} \;\text{ for every }\; |\lambda| \leq J_n \tag{35}$$

provided $\varkappa > 0$ is large enough, see Condition (37) below. In turn, we have Conditions (5.1) and (5.2) of Theorem 5.1 of [32] with $\Lambda_n = J_n$ (with the notation of [32]). By Corollary 5.1 and Theorem 6.1 of [32] we obtain Theorem 8.

It remains to prove (34) and (35). We plan to apply Theorem 4 (ii) with $g = \psi_\lambda$ and $\delta = \delta_n = p\varkappa c(n)$. First, we have $|\psi_\lambda^1|_p \leq C_p 2^{|\lambda|(1/2-1/p)}$ for $p = 1, 2, \infty$ by (6), so one readily checks that for

$$\varkappa \geq \tfrac{4}{p} R(1-2\rho)^{-1} C_\infty (\log |\mathbb{T}_n|)^{-1},$$

the condition $\delta_n \geq 4R(1-2\rho)^{-1}|\psi_\lambda^1|_\infty |\mathbb{T}_n|^{-1}$ is satisfied, and this is always true for large enough n. Furthermore, since $2^{|\lambda|} \leq 2^{J_n} \leq c(n)^{-2}$ it is not difficult to check that

$$\Sigma_{1,n}(\psi_\lambda^1) = |\psi_\lambda^1|_2^2 + \min_{1 \leq \ell \leq n-1} \big(|\psi_\lambda^1|_1^2 2^\ell + |\psi_\lambda^1|_\infty^2 2^{-\ell} \big) \leq C \tag{36}$$

for some $C > 0$ and thus $\kappa_3 \Sigma_{1,n}(\psi_\lambda) \leq \kappa_3 C = C'$ say. Also $\kappa_4 |\psi_\lambda^1|_\infty \delta_n \leq \kappa_4 C_\infty 2^{|\lambda|/2} c(n) p\varkappa \leq C'' p\varkappa$, where $C'' > 0$ does not depend on n since $2^{|\lambda|/2} \leq c(n)^{-1}$. Theorem 4 (ii) yields

$$\mathbb{P}\big(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq p\varkappa c(n)\big) \leq 2\exp\Big(-\frac{|\mathbb{T}_n| p^2 \varkappa^2 c(n)^2}{C' + C'' p\varkappa}\Big) \leq c(n)^{2p}$$

for $\varkappa$ such that

$$\varkappa \geq \tfrac{1}{2} C'' + \sqrt{(C'')^2 + \tfrac{4}{p} C'} \tag{37}$$

and large enough n. Thus (35) is proved. Straightforward computations show that (34) follows using $\mathbb{E}\big[|\widehat{\nu}_{\lambda,n} - \nu_\lambda|^p\big] = \int_0^\infty p u^{p-1} \mathbb{P}\big(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq u\big) du$ and (35) again. The proof of Theorem 8 is complete.

5.6. **Preparation for the proof of Theorem 9.** For $h : \mathcal{S}^2 \to \mathbb{R}$, define $|h|_{\infty,1} = \displaystyle\int_{\mathcal{S}} \sup_{x \in \mathcal{S}} |h(x,y)| dy$. For $n \geq 2$, set also

$$\Sigma_{3,n}(h) = |h|_2^2 + \min_{1 \leq \ell \leq n-1} \big(|h|_1^2 2^\ell + |h|_{\infty,1}^2 2^{-\ell} \big). \tag{38}$$

Recall that under Assumption 3 with $\mathfrak{n}(dx) = dx$, we set $f_{\mathcal{Q}}(x,y) = \nu(x)\mathcal{Q}(x,y)$. Before proving Theorem 9, we first need the following preliminary estimate

Lemma 15. *Work under Assumption 2 with $\mathfrak{n}(dx) = dx$ and Assumption 3. Let $h : \mathcal{D}^2 \to \mathbb{R}$ be such that $|hf_{\mathcal{Q}}|_1 < \infty$. For every $n \geq 1$ and for any $\delta \geq 4|h|_\infty (Rn+1)|\mathbb{T}_n^\star|^{-1}$, we have*

$$\mathbb{P}\Big(\frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} h(X_{u^-}, X_u) - \langle h, f_{\mathcal{Q}} \rangle \geq \delta\Big) \leq \exp\Big(\frac{-n^{-1}|\mathbb{T}_n^\star| \delta^2}{\kappa_5 \Sigma_{3,n}(h) + \kappa_2 |h|_\infty \delta}\Big)$$

where $\mathbb{T}_n^\star = \mathbb{T}_n \setminus \{\emptyset\}$ and $\kappa_5 = \max\{|\mathcal{Q}|_{\mathcal{D}}, |\mathcal{Q}|_{\mathcal{D}}^2\} \kappa_1(\mathcal{Q}, \mathcal{D})$.

Proof. We plan to apply Theorem 5 (ii) to $g(x, x_0, x_1) = \frac{1}{2}\big(h(x,x_0) + h(x,x_1)\big)$. Since $\mathcal{Q} = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$ we readily have $\mathcal{P}g(x) = \int_{\mathcal{D}} h(x,y)\mathcal{Q}(x,y)dy$. Moreover, in that case,

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} g(X_u, X_{u0}, X_{u1}) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} h(X_{u^-}, X_u)$$

and $\int_{\mathcal{S}} \mathcal{P}g(x)\nu(x)dx = \int_{\mathcal{S} \times \mathcal{D}} h(x,y)\mathcal{Q}(x,y)\nu(x)dxdy = \langle h, f_{\mathcal{Q}} \rangle$. We then simply need to estimate $\Sigma_{2,n}(g)$ defined by (4). It is not difficult to check that the following estimates hold

$$|\mathcal{P}g|_1^2 \leq |\mathcal{Q}|_{\mathcal{D}}^2 |h|_1^2, \;\; |\mathcal{P}g|_\infty^2 \leq |\mathcal{Q}|_{\mathcal{D}}^2 |h|_{\infty,1}^2 \;\text{ and }\; |\mathcal{P}g^2|_1 \leq |\mathcal{Q}|_{\mathcal{D}} |h|_2^2$$

since $(\mathcal{P}g^2)(x) \leq \int_{\mathcal{D}} h(x,y)^2 \mathcal{Q}(x,y)dy$. Thus $\Sigma_{2,n}(g) \leq \max\{|\mathcal{Q}|_{\mathcal{D}}, |\mathcal{Q}|_{\mathcal{D}}^2\}\Sigma_{3,n}(h)$ and the result follows. ☐

5.7. **Proof of Theorem 9, upper bound.** *Step 1.* We proceed as for Theorem 8. Putting $c(n) = (n \log |\mathbb{T}_n^\star|/|\mathbb{T}_n^\star|)^{1/2}$ and noting that the maximal resolution $J = J_n$ is such that $2^{dJ_n} \sim c(n)^{-2}$ with $d = 2$, we only have to prove that for every $p \geq 1$,

$$\mathbb{E}\big[|\widehat{f}_{\lambda,n} - f_\lambda|^p\big] \lesssim c(n)^p \ \text{ for every } \ |\lambda| \leq J_n \tag{39}$$

and

$$\mathbb{P}\big(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p\varkappa c(n)\big) \leq c(n)^{2p} \ \text{ for every } \ |\lambda| \leq J_n. \tag{40}$$

We plan to apply Lemma 15 with $h(x,y) = \psi_\lambda^d(x,y) = \psi_\lambda^2(x,y)$ and $\delta = \delta_n = p\varkappa c(n)$. With the notation used in the proof of Theorem 8 one readily checks that for

$$\varkappa \geq \tfrac{4}{p}(1-2\rho)^{-1}C_\infty(Rn+1)(\log|\mathbb{T}_n^\star|)^{-1}$$

the condition $\delta_n \geq 4|\psi_\lambda^2|_\infty(Rn+1)|\mathbb{T}_n^\star|^{-1}$ is satisfied, and this is always true for large enough n and

$$\varkappa \geq \tfrac{4}{p}(1-2\rho)^{-1}C_\infty(2R+1). \tag{41}$$

Furthermore, since $|\psi_\lambda^d|_p \leq C_p 2^{d|\lambda|(1/2-1/p)}$ for $p = 1, 2, \infty$ and $2^{d|\lambda|} \leq 2^{dJ_n} \leq c(n)^{-2}$ we can easily check

$$\Sigma_{3,n}(\psi_\lambda^d) = |\psi_\lambda^d|_2^2 + \min_{1\leq \ell \leq n-1}\big(|\psi_\lambda^d|_1^2 2^\ell + |\psi_\lambda^d|_{\infty,1}^2 2^{-\ell}\big) \leq C$$

for some $C > 0$, and thus $\kappa_5\Sigma_{3,n}(g) \leq \kappa_5 C = C'$ say. Also, $\kappa_2|\psi_\lambda^d|_\infty\delta_n \leq \kappa_2 C_\infty 2^{d|\lambda|/2}c(n)p\varkappa \leq C''p\varkappa$, where C'' does not depend on n. Applying Lemma 15, we derive

$$\mathbb{P}\big(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p\varkappa c(n)\big) \leq 2\exp\Big(-\frac{n^{-1}|\mathbb{T}_{n-1}|p^2\varkappa^2 c(n)^2}{C' + C''p\varkappa}\Big) \leq c(n)^{2p}$$

as soon as $\varkappa$ satisfies (41) and (37) (with appropriate changes for C' and C''). Thus (40) is proved and (39) follows likewise. By [32] (Corollary 5.1 and Theorem 6.1), we obtain

$$\mathbb{E}\Big(\big[\|\widehat{f}_n - f_{\mathcal{Q}}\|_{L^p(\mathcal{D}^2)}^p\big]\Big)^{1/p} \lesssim \Big(\frac{n\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_2(s,p,\pi)} \tag{42}$$

as soon as $\|f_{\mathcal{Q}}\|_{B^s_{\pi,\infty}(\mathcal{D}^2)}$ is finite, as follows from $f_{\mathcal{Q}}(x,y) = \mathcal{Q}(x,y)\nu(x)$ and the fact that $\|\nu\|_{B^s_{\pi,\infty}(\mathcal{D})}$ is finite too. The last statement can be readily seen from the representation $\nu(x) = \int_{\mathcal{S}} \nu(y)\mathcal{Q}(y,x)dy$ and the definition of Besov spaces in terms of moduli of continuity, see *e.g.* Meyer [35] or Härdle *et al.* [30], using moreover that $\pi \geq 1$.

Step 2. Since $\mathcal{Q}(x,y) = f_{\mathcal{Q}}(x,y)/\nu(x)$ and $\widehat{\mathcal{Q}}_n(x,y) = \widehat{f}_n(x,y)/\max\{\widehat{\nu}_n(x), \varpi\}$, we readily have

$$|\widehat{\mathcal{Q}}_n(x,y) - \mathcal{Q}(x,y)|^p \lesssim \tfrac{1}{\varpi^p}\big(|\widehat{f}_n(x,y) - f_{\mathcal{Q}}(x,y)|^p + \tfrac{|f_{\mathcal{Q}}|_\infty^p}{m(\nu)^p}|\max\{\widehat{\nu}_n(x),\varpi\} - \nu(x)|^p\big),$$

where the supremum for $f_{\mathcal{Q}}$ can be restricted over $\mathcal{D}^2$. Since $m(\nu) \geq \varpi$, we have $|\max\{\widehat{\nu}_n(x),\varpi\} - \nu(x)| \leq |\widehat{\nu}_n(x) - \nu(x)|$ for $x \in \mathcal{D}$, therefore

$$\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p \lesssim \tfrac{1}{\varpi^p}\big(\|\widehat{f}_n - f_{\mathcal{Q}}\|_{L^p(\mathcal{D}^2)}^p + \tfrac{|f_{\mathcal{Q}}|_\infty^p}{m(\nu)^p}\|\nu - \nu_n\|_{L^p(\mathcal{D})}^p\big)$$

holds as well. We conclude by applying successively the estimate (42) and Theorem 8.

5.8. **Proof of Theorem 9, lower bound.** We only give a brief sketch: the proof follows classical lower bounds techniques, bounding appropriate statistical distances along hypercubes, see [25, 30] and more specifically [15, 31, 34] for specific techniques involving Markov chains. We separate the so-called *dense* and *sparse* case.

The dense case $\varepsilon_2 > 0$. Let $\psi_\lambda : \mathcal{D}^2 \to \mathbb{R}$ a family of (compactly supported) wavelets adapted to the domain $\mathcal{D}$ and satisfying Assumption 7. For j such that $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+1)}$, consider the family

$$\mathcal{Q}_{\epsilon,j}(x,y) = |\mathcal{D}^2|^{-1}\mathbf{1}_{\mathcal{D}^2}(x,y) + \gamma|\mathbb{T}_n|^{-1/2}\sum_{\lambda\in\Lambda_j}\epsilon_\lambda\psi_\lambda^2(x,y)$$

where $\epsilon \in \{-1,1\}^{\Lambda_j}$ and $\gamma > 0$ is a tuning parameter (independent of n). Since $|\psi_\lambda^2|_\infty \leq C_\infty 2^{|\lambda|} = C_\infty 2^j$ and since the number of overlapping terms in the sum is bounded (by some fixed integer N), we have

$$\gamma|\mathbb{T}_n|^{-1/2}\Big|\sum_{\lambda\in\Lambda_j}\epsilon_\lambda\psi_\lambda^2(x,y)\Big| \leq \gamma|\mathbb{T}_n|^{-1/2}NC_\infty 2^j \lesssim \gamma.$$

This term can be made smaller than $|\mathcal{D}^2|^{-1}$ by picking γ sufficiently small. Hence $\mathcal{Q}_{\epsilon,j}(x,y) \geq 0$ and since $\int\psi_\lambda = 0$, the family $\mathcal{Q}_{\epsilon,j}(x,y)$ are all admissible *mean* transitions with common invariant measure $\nu(dx) = \mathbf{1}_{\mathcal{D}}(x)dx$ and belong to a common ball in $\mathcal{B}^s_{\pi,\infty}(\mathcal{D}^2)$. For $\lambda \in \Lambda_j$, define $T_\lambda : \{-1,1\}^{\Lambda_j} \to \{-1,1\}^{|\Lambda_j|}$ by $T_\lambda(\epsilon_\lambda) = -\epsilon_\lambda$ and $T_\lambda(\epsilon_\mu) = \epsilon_\mu$ if $\mu \neq \lambda$. The lower bound in the dense case is then a consequence of the following inequality

$$\limsup_n \max_{\epsilon\in\{-1,1\}^{\Lambda_j},\lambda\in\Lambda_j} \|\mathbb{P}^n_{\epsilon,j} - \mathbb{P}^n_{T_\lambda(\epsilon),j}\|_{TV} < 1, \tag{43}$$

where $\mathbb{P}^n_{\epsilon,j}$ is the law of $(X_u)_{u\in\mathbb{T}_n}$ specified by the $\mathbb{T}$-transition $\mathcal{P}_{\epsilon,j} = \mathcal{Q}_{\epsilon,j}\otimes\mathcal{Q}_{\epsilon,j}$ and the initial condition $\mathcal{L}(X_\emptyset) = \nu$.

We briefly show how to obtain (43). By Pinsker's inequality, it is sufficient to prove that $\mathbb{E}^n_{\epsilon,j}\big[\log\frac{d\mathbb{P}^n_{\epsilon,j}}{d\mathbb{P}^n_{T_\lambda(\epsilon),j}}\big]$ can be made arbitrarily small uniformly in n (but fixed). We have

$$\begin{aligned}
\mathbb{E}^n_{\epsilon,j}\Big[-\log\frac{d\mathbb{P}^n_{T_\lambda(\epsilon),j}}{d\mathbb{P}^n_{\epsilon,j}}\Big] &= -\sum_{u\in\mathbb{T}_n}\mathbb{E}^n_{\epsilon,j}\Big[\log\frac{\mathcal{P}_{T_\lambda(\epsilon),j}(X_u,X_{u0},X_{u1})}{\mathcal{P}_{\epsilon,j}(X_u,X_{u0},X_{u1})}\Big] \\
&= -\sum_{u\in\mathbb{T}^\star_{n+1}}\mathbb{E}^n_{\epsilon,j}\Big[\log\frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(X_{u^-},X_u)}{\mathcal{Q}_{\epsilon,j}(X_{u^-},X_u)}\Big] \\
&= -|\mathbb{T}^\star_{n+1}|\int_{\mathcal{D}^2}\log\Big(\frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x,y)}{\mathcal{Q}_{\epsilon,j}(x,y)}\Big)\mathcal{Q}_{\epsilon,j}(x,y)\nu(dx)dy \\
&\leq \ \ |\mathbb{T}^\star_{n+1}|\int_{\mathcal{D}^2}\Big(\frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x,y)}{\mathcal{Q}_{\epsilon,j}(x,y)}-1\Big)^2\mathcal{Q}_{\epsilon,j}(x,y)\nu(dx)dy
\end{aligned}$$

using $-\log(1+z) \leq z^2 - z$ valid for $z \geq -1/2$ and the fact that $\nu(dx)$ is an invariant measure for both $\mathcal{Q}_{T_\lambda(\epsilon),j}$ and $\mathcal{Q}_{\epsilon,j}$. Noting that

$$\mathcal{Q}_{T_\lambda(\epsilon),j}(x,y) = \mathcal{Q}_{\epsilon,j}(x,y) - 2\gamma|\mathbb{T}_n|^{-1/2}\epsilon_\lambda\psi_\lambda^2(x,y),$$

we derive

$$\Big|\frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x,y)}{\mathcal{Q}_{\epsilon,j}(x,y)} - 1\Big| \leq \frac{2\gamma|\mathbb{T}_n|^{-1/2}C_\infty 2^j}{1-\gamma|\mathbb{T}_n|^{-1/2}NC_\infty 2^j} \lesssim \gamma|\mathbb{T}_n|^{-1/2}$$

hence the squared term within the integral is of order $\gamma^2|\mathbb{T}_n|^{-1}$ so that, by picking γ sufficiently small, our claim about $\mathbb{E}^n_{\epsilon,j}\big[\log\frac{d\mathbb{P}^n_{\epsilon,j}}{d\mathbb{P}^n_{T_\lambda(\epsilon),j}}\big]$ is proved and (43) follows.

The sparse case $\epsilon_2 \leq 0$. We now consider the family

$$\mathcal{Q}_{\lambda,j}(x,y) = |\mathcal{D}^2|^{-1}\mathbf{1}_{\mathcal{D}^2}(x,y) + \gamma\big(\tfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2}\epsilon_\lambda\psi^2_\lambda(x,y)$$

with $\epsilon_\lambda \in \{-1,+1\}$ and $\lambda \in \Lambda_j$, with j such that $\big(\frac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2} \lesssim 2^{-j(s+1-2/\pi)}$. The lower bound then follows from the representation

$$\log\frac{d\mathbb{P}^n_{\lambda,j}}{d\mathbb{P}^n_\nu} = \mathcal{U}^n_\lambda - \omega_\lambda\log 2^j$$

where $\mathbb{P}^n_{\lambda,j}$ and $\mathbb{P}^n_\nu$ denote the law of $(X_u)_{u\in\mathbb{T}_n}$ specified by the $\mathbb{T}$-transitions $\mathcal{Q}_{\lambda,j}\otimes\mathcal{Q}_{\lambda,j}$ and $\nu\otimes\nu$ respectively (and the initial condition $\mathcal{L}(X_\emptyset)=\nu$); the ω's are such that $\sup_n\max_{\lambda\in\Lambda_j}\omega_\lambda < 1$, and $\mathcal{U}^n_\lambda$ are random variables such that $\mathbb{P}^n_{\lambda,j}\big(\mathcal{U}^n_\lambda \geq -C_1\big) \geq C_2 > 0$ for some $C_1, C_2 > 0$. We omit the details, see *e.g.* [15, 31, 34].

5.9. **Proof of Theorem 10.**

Proof of Theorem 10, upper bound. We closely follow Theorem 9 with $c(n) = (n\log|\mathbb{T}_{n-1}|/|\mathbb{T}_{n-1}|)^{1/2}$ and $J = J_n$ such that $2^{dJ_n} \sim c(n)^{-2}$ with $d=3$ now. With $\delta = \delta_n = p\varkappa c(n)$, for $\varkappa \geq \frac{4}{p}(1-2\rho)^{-1}C_\infty(2R+1)$, we have $\delta_n \geq 4|\psi^3_\lambda|_\infty(Rn+1)|\mathbb{T}^\star_n|^{-1}$.

Furthermore, since $|\psi^d_\lambda|_p \leq C_p 2^{d|\lambda|(1/2-1/p)}$ for $p=1,2,\infty$ and $2^{d|\lambda|} \leq 2^{dJ_n} \leq c(n)^{-2}$ it is not difficult to check that

$$\Sigma_{2,n}(\psi_\lambda) \leq \max\big\{|\mathcal{P}|_{\mathcal{D},1}|\mathcal{Q}|_{\mathcal{D}}, |\mathcal{P}|^2_{\mathcal{D},1}\big\}\Sigma_{1,n}(\psi_\lambda) \leq C$$

thanks to Assumption 6 and (36), and thus $\kappa_1\Sigma_{2,n}(g) \leq \kappa_1 C = C'$. We also have $\kappa_2|\psi^d_\lambda|_\infty\delta_n \leq \kappa_2 C_\infty 2^{|\lambda|d/2}c(n)p\varkappa \leq C''p\varkappa$, where C'' does not depend on n. Noting that $f_\lambda = \langle f_{\mathcal{P}}, \psi^d_\lambda\rangle = \int\mathcal{P}\psi^d_\lambda d\nu$, we apply Theorem 5 (ii) to $g=\psi_\lambda$ and derive

$$\mathbb{P}\big(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p\varkappa c(n)\big) \leq 2\exp\Big(-\frac{n^{-1}|\mathbb{T}_{n-1}|p^2\varkappa^2c(n)^2}{C'+C''p\varkappa}\Big) \leq c(n)^{2p}$$

for every $|\lambda| \leq J_n$ as soon as κ is large enough and the estimate

$$\mathbb{E}\Big(\big[\|\widehat{f}_n - f_{\mathcal{P}}\|^p_{L^p(\mathcal{D}^3)}\big]\Big)^{1/p} \lesssim \Big(\frac{n\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\Big)^{\alpha_3(s,p,\pi)}$$

follows thanks to the theory of [32]. The end of the proof follows Step 2 of the proof of Theorem 9 line by line, substituting $f_{\mathcal{Q}}$ by $f_{\mathcal{P}}$. □

Proof of Theorem 10, lower bound. This is a slight modification of the proof of Theorem 9, lower bound. For the dense case $\epsilon_3 > 0$, we consider an hypercube of the form

$$\mathcal{P}_{\epsilon,j}(x,y,z) = |\mathcal{D}^3|^{-1}\mathbf{1}_{\mathcal{D}^3}(x,y,z) + \gamma|\mathbb{T}_n|^{-1/2}\sum_{\lambda\in\Lambda_j}\epsilon_\lambda\psi^3_\lambda(x,y,z)$$

where $\epsilon \in \{-1,1\}^{\Lambda_j}$ with j such that $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+3/2)}$ and $\gamma > 0$ a tuning parameter, while for the sparse case $\epsilon_3 \leq 0$, we consider the family

$$\mathcal{P}_{\lambda,j}(x,y,z) = |\mathcal{D}^3|^{-1}\mathbf{1}_{\mathcal{D}^3}(x,y,z) + \gamma\big(\tfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2}\epsilon_\lambda\psi^3_\lambda(x,y,z)$$

with $\epsilon_\lambda \in \{-1,+1\}$, $\lambda \in \Lambda_j$, and j such that $\big(\frac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2} \lesssim 2^{-j(s+3(1/2-1/\pi))}$. The proof then goes along a classical line. □

5.10. **Proof of Theorem 11.**

Proof of Theorem 11, upper bound. Set $\widehat{v}_n(x) = \frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}\mathbf{1}_{\{x/2\leq X_u\leq x\}}$ and $v_\nu(x) = \int_{x/2}^x \nu_B(y)dy$. By Propositions 2 and 4 in Doumic *et al.* [26], one can easily check that $\sup_{x\in\mathcal{D}}\nu_B(x) < \infty$ and $\inf_{x\in\mathcal{D}} v_\nu(x) > 0$ with some uniformity in B by Lemma 2 and 3 in [26]. For $x \in \mathcal{D}$, we have

$$\begin{aligned}\big|\widehat{B}_n(x) - B(x)\big|^p &\lesssim \tfrac{1}{\varpi^p}\big|\widehat{\nu}_n(x) - \nu_B(x)\big|^p + \frac{\sup_{x\in\mathcal{D}}\nu_B(x)^p}{\inf_{x\in\mathcal{D}} v_\nu(x)^p}\big|\max\{\widehat{v}_n(x),\varpi\} - v_\nu(x)\big|^p \\ &\lesssim \big|\widehat{\nu}_n(x) - \nu_B(x)\big|^p + \big|\widehat{v}_n(x) - v_\nu(x)\big|^p.\end{aligned}$$

By Theorem 4 (ii) with $g = \mathbf{1}_{\{x/2\leq\cdot\leq x\}}$, one readily checks

$$\mathbb{E}\big[|\widehat{v}_n(x) - v_\nu(x)|^p\big] = \int_0^\infty pu^{p-1}\mathbb{P}\big(|\widehat{v}_n(x) - v_\nu(x)| \geq u\big)du \lesssim |\mathbb{T}_n|^{-p/2}$$

and this term is negligible. Finally, it suffices to note that $\|\nu_B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$ is finite as soon as $\|B\|_{\mathcal{B}^s_{\pi,\infty}(\mathcal{D})}$ is finite. This follows from

$$\nu_B(x) = \int_{\mathcal{S}}\nu_B(y)Q_B(y,x)dy = \frac{B(2x)}{\tau x}\int_0^{2x}\nu_B(y)\exp\Big(-\int_{y/2}^x\frac{B(2z)}{\tau z}dz\Big)dy.$$

We conclude by applying Theorem 8. □

Proof of Theorem 11, lower bound. This is again a slight modification of the proof of Theorem 9, lower bound. For the dense case $\epsilon_1 > 0$, we consider an hypercube of the form

$$B_{\epsilon,j}(x) = B_0(x) + \gamma|\mathbb{T}_n|^{-1/2}\sum_{\lambda_j}\epsilon_k\psi^1_\lambda(x)$$

where $\epsilon \in \{-1,1\}^{\Lambda_j}$ with j such that $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+1/2)}$ and $\gamma > 0$ a tuning parameter. By picking B_0 and γ in an appropriate way, we have that B_0 and $B_{\varepsilon,j}$ belong to a common ball in $\mathcal{B}^s_{\pi,\infty}(\mathcal{D})$ and also belong to $\mathcal{C}(r,L)$. The associated $\mathbb{T}$-transition $\mathcal{P}_{B_{\epsilon,j}}$ defined in (17) admits as mean transition

$$\mathcal{Q}_{B_{\epsilon,j}}(x,dy) = \frac{B_{\epsilon,j}(2y)}{\tau y}\exp\Big(-\int_{x/2}^y\frac{B_{\epsilon,j}(2z)}{\tau z}dz\Big)\mathbf{1}_{\{y\geq x/2\}}dy$$

which has a unique invariant measure $\nu_{B_{\epsilon,j}}$. Establishing (43) is similar to the proof of Theorem 9, lower bound, using the explicit representation for $\mathcal{Q}_{B_{\epsilon,j}}$ with a slight modification due to the fact that the invariant measures $\nu_{B_{\epsilon,j}}$ and $\nu_{B_{T_\lambda(\epsilon),j}}$ do not necessarily coincide. We omit the details.

For the sparse case $\epsilon_1 \leq 0$, we consider the family

$$B_{\lambda,j}(x) = B_0(x) + \gamma\big(\tfrac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2}\epsilon_\lambda\psi^1_\lambda(x)$$

with $\epsilon_\lambda \in \{-1,+1\}$, $\lambda \in \Lambda_j$, with j such that $\big(\frac{\log|\mathbb{T}_n|}{|\mathbb{T}_n|}\big)^{1/2} \lesssim 2^{-j(s+1/2-1/\pi)}$. The proof is then similar. □

6. Appendix

6.1. **Proof of Lemma 13.** *The case* $r=0$. By Assumption 3,

$$\big|\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big| \leq 2\big(|\widetilde{g}|_\infty + R|\widetilde{g}|_\infty \rho\big) \leq 4(1+R\rho)|g|_\infty.$$

This proves the first estimate in the case $r=0$. For $u\in\mathbb{G}_{n-1}$,

$$\begin{aligned}
&\mathbb{E}\big[\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big)^2|\mathcal{F}_{n-1}\big]\\
&\quad= \mathbb{E}\big[\big(g(X_{u0})+g(X_{u1})-2\mathcal{Q}g(X_u)\big)^2|\mathcal{F}_{n-1}\big]\\
&\quad\leq \mathbb{E}\big[\big(g(X_{u0})+g(X_{u1})\big)^2|\mathcal{F}_{n-1}\big] \leq 2\big(\mathcal{P}_0 g^2(X_u)+\mathcal{P}_1 g^2(X_u)\big) = 4\mathcal{Q}g^2(X_u)
\end{aligned}$$

and for $x\in\mathcal{S}$, by Assumption 2,

$$\mathcal{Q}g^2(x)=\int_{\mathcal{S}} g(y)^2\mathcal{Q}(x,y)\mathfrak{n}(dy) \leq |\mathcal{Q}|_{\mathcal{D}}|g|_2^2$$

since g vanishes outside $\mathcal{D}$. Thus

$$\mathbb{E}\big[\big(\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-2\mathcal{Q}\widetilde{g}(X_u)\big)^2|\mathcal{F}_{n-1}\big] \leq 4|\mathcal{Q}|_{\mathcal{D}}|g|_2^2 \tag{44}$$

hence the result for $r=0$.

The case $r\geq 1$. On the one hand, by Assumption 3,

$$\begin{aligned}
\big|2^r\big(\mathcal{Q}^r\widetilde{g}(X_{u0})+\mathcal{Q}^r\widetilde{g}(X_{u1})-2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big)\big| &\leq 2^r\big(2R|\widetilde{g}|_\infty(\rho^r+\rho^{r+1})\big)\\
&\leq 4R(1+\rho)|g|_\infty(2\rho)^r.
\end{aligned} \tag{45}$$

On the other hand, since

$$|\mathcal{Q}g(x)| \leq \int_{\mathcal{S}} |g(y)|\mathcal{Q}(x,y)\mathfrak{n}(dy) \leq |\mathcal{Q}|_{\mathcal{D}}|g|_1,$$

we also have

$$\begin{aligned}
2^r\big|\mathcal{Q}^r\widetilde{g}(X_{u0})+\mathcal{Q}^r\widetilde{g}(X_{u1})-2\mathcal{Q}^{r+1}\widetilde{g}(X_u)\big| &= 2^r\big|\mathcal{Q}^r g(X_{u0})+\mathcal{Q}^r g(X_{u1})-2\mathcal{Q}^{r+1}g(X_u)\big|\\
&\leq 2^r 4|\mathcal{Q}|_{\mathcal{D}}|g|_1.
\end{aligned} \tag{46}$$

Putting together these two estimates yields the result for the case $r\geq 1$.

6.2. **Proof of Lemma 14.** By Assumption 3,

$$|\Upsilon_r(X_u,X_{u0},X_{u1})| \leq 2\sum_{m=0}^{r} 2^m R|\widetilde{g}|_\infty \rho^m(1+\rho) \leq 4R|g|_\infty(1+\rho)(1-2\rho)^{-1}$$

since $\rho<1/2$. This proves the first bound. For the second bound we balance the estimates (45) and (46) obtained in the proof of Lemma 13. Let $\ell\geq 1$. For $u\in\mathbb{G}_{n-r-1}$, we have

$$|\Upsilon_r(X_u,X_{u0},X_{u1})| \leq I+II+III,$$

with

$$\begin{aligned}
I &= \big|\widetilde{g}(X_{u0})+\widetilde{g}(X_{u1})-\mathcal{Q}\widetilde{g}(X_u)\big|,\\
II &= \sum_{m=1}^{\ell\wedge r} 2^m\big|\mathcal{Q}^m\widetilde{g}(X_{u0})+\mathcal{Q}^m\widetilde{g}(X_{u1})-2\mathcal{Q}^{m+1}\widetilde{g}(X_u)\big|,\\
III &= \sum_{m=\ell\wedge r+1}^{r} 2^m\big|\mathcal{Q}^m\widetilde{g}(X_{u0})+\mathcal{Q}^m\widetilde{g}(X_{u1})-2\mathcal{Q}^{m+1}\widetilde{g}(X_u)\big|,
\end{aligned}$$

with $III = 0$ if $\ell > r$. For $u \in \mathbb{G}_{n-r-1}$, by (44), we successively have

$$\mathbb{E}[I^2 | \mathcal{F}_{n-r-1}] \leq 4|\mathcal{Q}|_{\mathcal{D}} |g|_2^2,$$

$$II \leq 4|\mathcal{Q}|_{\mathcal{D}} |g|_1 \sum_{m=1}^{\ell \wedge r} 2^m \leq 8|\mathcal{Q}|_{\mathcal{D}} |g|_1 2^{\ell \wedge r}$$

by (46), while for $\ell \leq r$,

$$III \leq 4R(1+\rho)|g|_\infty \sum_{m=\ell+1}^{r} (2\rho)^m \leq 4R(1+\rho)(1-2\rho)^{-1}|g|_\infty (2\rho)^{\ell+1}$$

by (45). The result follows.

Acknowledgements. We are grateful to A. Guillin for helpful discussions. The research of S.V. B.P. is supported by the Hadamard Mathematics Labex of the Fondation Mathématique Jacques Hadamard. The research of M.H. is partly supported by the Agence Nationale de la Recherche, (Blanc SIMI 1 2011 project CALIBRATION).

References

[1] K. B. Athreya and H.-J. Kang. *Some limit theorems for positive recurrent branching Markov chains: I.* Advances in Applied Probability, 30 (1998), 693–710.
[2] I. V. Basawa and R. M. Huggins. *Extensions of the bifurcating autoregressive model for cell lineage studies.* Journal of Applied Probability, 36 (1999), 1225–1233.
[3] I. V. Basawa and R. M. Huggins. *Inference for the extended bifurcating autoregressive model for cell lineage studies.* Australian & New Zealand Journal of Statistics, 42 (2000), 423–432.
[4] I. V. Basawa and J. Zhou. *Non-Gaussian bifurcating models and quasi-likelihood estimation.* Journal of Applied Probability, 41 (2004), 55-64.
[5] I. V. Basawa and J. Zhou. *Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors.* Journal of Time Series Analysis, 26 (2005), 825–842.
[6] I. Benjamini, Y. Peres. *Markov chains indexed by trees.* The Annals of Probability (1994), 219–243.
[7] B. Bercu, V. Blandin. *A Rademacher-Menchov approach for randon coefficient bifurcating autoregressive processes.* Stochastic Processes an their Applications, 125 (2015), 1218-1243.
[8] B. Bercu, B. de Saporta and A. Gégout-Petit. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach.* Electronic Journal of Probability, 87 (2009), 2492–2526.
[9] J. Bertoin *Random fragmentation and coagulation processes.* Cambridge University Press, 2006.
[10] S. V. Bitseki Penda, H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models.* Annales de l'Institut Henri Poincaré, 50 (2014), 806–844.
[11] S. V. Bitseki Penda, H. Djellout and A. Guillin. *Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application.* The Annals of Applied Probability, 24 (2014), 235–291.
[12] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains. arXiv:1501.06693.*
[13] S. V. Bitseki Penda, A. Olivier. *Nonparametric estimation of the autoregressive functions in bifurcating autoregressive models. arXiv:1506.01842.*
[14] V. Blandin. *Asymptotic results for random coefficient bifurcating autoregressive processes.* Statistics 48 (2014), 1202–1232.
[15] S. Clémençon. *Adaptive Estimation of the Transition Density of a Regular Markov Chain by Wavelet Methods.* Mathematical Methods of Statistics, 9 (2000), 323–357.
[16] R. Cowan and R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies.* Biometrics, 42 (1986), 769–783.
[17] A. Cohen. *Wavelets in Numerical Analysis.* In: Ciarlet, P.G., Lions, J.L. (eds.) Handbook of Numerical Analysis, vol. VII. Elsevier, Amsterdam, 2000.
[18] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data.* Electronic Journal of Statistics, 5 (2011), 1313–1353.

[19] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Asymmetry tests for Bifurcating Auto-Regressive Processes with missing data.* Statistics & Probability Letters, 82 (2012), 1439–1444.
[20] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Random coefficients bifurcating autoregressive processes.* ESAIM PS, in press. *arXiv:1205.3658.*
[21] R. DeVore, S.V. Konyagin, V.N. Temlyakov. *Hyperbolic Wavelet Approximation. Constructive Approximation.* Springer, New York, 1998.
[22] D.L. Donoho, I.M. Johnstone. *Ideal spatial adaptation via wavelet shrinkage.* Biometrika, 81 (1994), 425–455.
[23] D.L. Donoho, I.M. Johnstone. *Adapting to unknown smoothness via wavelet shrinkage.* Journal of the American Statistical Association, 90 (1995), 1200–1224.
[24] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard. *Wavelet shrinkage: Asymptopia?* Journal of the Royal Statistical Society, 57 (1995), 301–369.
[25] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard. *Density estimation by wavelet thresholding.* The Annals of Statistics, 54 (1996), 508–539.
[26] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree.* Bernoulli, 21 (2015), 1760–1799.
[27] F. Gao, A. Guillin and L. Wu. *Bernstein-type concentration inequalities for symmetric Markov processes.* Theory of Probability and its Applications, 58 (2014), 358–382.
[28] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging.* The Annals of Applied Probability, 17 (2007), 1538–1569.
[29] M. Hairer and J. Mattingly. *Yet Another Look at Harris' Ergodic Theorem for Markov Chains.* Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability, 63 (2011), 109–117.
[30] W. Härdle, G. Kerkyacharian, D. Picard and A. Tsybakov. *Wavelets, Approximation and Statistical Applications.* Lecture Notes in Statistics, Springer, 1998.
[31] M. Hoffmann. *Adaptive estimation in diffusion processes.* Stochastic Processes and their Applications, 79 (1999), 135-163.
[32] G. Kerkyacharian and D. Picard. *Thresholding algorithms, maxisets and well-concentrated bases.* Test, 9 (2000), 283-344.
[33] C. Lacour. *Adaptive estimation of the transition density of a Markov chain* Annales de l'Institut Henri Poincaré, Probabillités et Statistique, 43 (2007), 571–597.
[34] C. Lacour. *Nonparametric estimation of the stationary density and the transition density of a Markov chain* Stochastic Process and their Applications 118 (2008), 232–260.
[35] Y. Meyer. *Ondelettes et Opérateurs*, vol. 1. Hermann, Paris, 1990.
[36] S. Meyn and R. Tweedie. *Markov chains and stochastic stability.* Springer Berlin Heidelberg, 1993.
[37] B. Perthame. *Transport Equations in Biology.* Frontiers in Mathematics. Basel: Birkhäuser, 2007.
[38] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert and M. Doumic. *Division control in Escherichia coli is based on a size-sensing rather than a timing mechanism.* BMC Biology, 02/2014 12(1):17.
[39] C. Takacs. *Strong law of large numbers for branching Markov chains.* Markov Processes and Related Fields, 8 (2001), 107–116.

S. Valère Bitseki Penda, IMB, CNRS-UMR 5584, Université de Bourgogne, 9 avenue Alain Savary, 21078 Dijon Cedex, France.
E-mail address: `simeon-valere.bitseki-penda@u-bourgogne.fr`

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

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