Journal of Theoretical Probability
https://doi.org/10.1007/s10959-022-01205-w

Central Limit Theorem for Kernel Estimator of Invariant Density in Bifurcating Markov Chains Models

S. Valère Bitseki Penda[1] · Jean-François Delmas[2]

Received: 8 February 2022 / Revised: 15 August 2022 / Accepted: 29 September 2022

Abstract
Bifurcating Markov chains are Markov chains indexed by a full binary tree representing the evolution of a trait along a population where each individual has two children. Motivated by the functional estimation of the density of the invariant probability measure which appears as the asymptotic distribution of the trait, we prove the consistency and the Gaussian fluctuations for a kernel estimator of this density based on late generations. In this setting, it is interesting to note that the distinction of the three regimes on the ergodic rate identified in a previous work (for fluctuations of average over large generations) disappears. This result is a first step to go beyond the threshold condition on the ergodic rate given in previous statistical papers on functional estimation.

Keywords Bifurcating Markov chains · Kernel estimator · Density estimation · Bifurcating autoregressive process · Binary trees · Fluctuations for trees indexed Markov chains

Mathematics Subject Classification (2020) 62G05 · 62G07 · 62G20 · 60J80 · 60F05

1 Introduction

Bifurcating Markov chains (BMCs) are a class of stochastic processes indexed by regular binary tree and which satisfy the branching Markov property (see below for a precise definition). This model represents the evolution of a trait along a popula-

✉ S. Valère Bitseki Penda
Simeon-Valere.Bitseki-Penda@u-bourgogne.fr

Jean-François Delmas
jean-francois.delmas@enpc.fr

1 Institut de Mathématiques de Bourgogne, UMR 5584, CNRS, Université de Bourgogne Franche-Comté, 21000 Dijon, France

2 CERMICS, École des Ponts, Champs-sur-Marne, France

Published online: 28 October 2022

tion where each individual has two children. The recent study of BMC models was motivated by the understanding of the cell division mechanism (where the trait of an individual is given by its growth rate). The first model of BMC, named "symmetric" bifurcating autoregressive process (BAR), see Sect. 3.2 for more details in a Gaussian framework, was introduced by Cowan and Staudte [6] in order to analyze cell lineage data. In [12], Guyon has studied more general asymmetric BMC to prove statistical evidence of aging in Escherichia Coli. We refer to [3] for more detailed references on this subject. Recently, several statistical works have been devoted to the estimation of cell division rates, see Doumic et al. [11], Bitseki et al. [4] and Hoffmann and Marguet [14]. Moreover, another studies, such as Doumic et al. [10], can be generalized using the BMC theory (we refer to the conclusion therein).

In this paper, our objective is to study the functional estimation of the density of the invariant probability measure μ associated with the BMC. For this purpose, we develop a kernel estimation in the $L^2(\mu)$ framework under reasonable hypothesis (which are in particular satisfied by the Gaussian symmetric BAR model from Sect. 3.2). This approach is in the spirit of the $L^2(\mu)$ approach developed [2]. In BMC model, the evolution of the trait along the genealogy of an individual taken at random is Markovian. Let us assume it is geometrically ergodic with rate $\alpha \in (0, 1)$, with μ is its invariant measure. In [2], three regimes where identified for the rate of convergence of averages over large generations according to the ergodic rate of convergence α with respect to the threshold $1/\sqrt{2}$. It is interesting, and surprising as well, to note that the distinction of those three regimes disappears for the rate of convergence when considering the kernel density estimation of the density of μ, see Theorem 3.5. (In [12], for different reasons, the distinction of the three regimes disappears also for additive functionals of the BMC with martingale increments.) However, let us mention that some further restriction on the admissible bandwidths of the kernel estimator is to be taken into account in the supercritical regime (*i.e.*, $\alpha > 1/\sqrt{2}$), to be precise see Condition (11) which is in force for Theorem 3.5. Furthermore, we get that estimations using different generations provide asymptotically independent fluctuations, see Remark 3.8 (see also the form of the asymptotic variance in Theorem 4.8 and Remark 4.9 in a more general framework); this phenomenon already appears in [7]. The convergence of the kernel estimator in Theorem 3.5 relies on different type of assumptions:

- Geometric ergodic rate $\alpha \in (0, 1)$ of convergence for the evolution of the trait along the genealogy of an individual taken at random, see Assumption 2.3.
- Regularity (density and integrability conditions) for the evolution kernel $\mathcal{P}$ and the initial distribution of the BMC, see Assumptions 3.1, and 3.2. The former is in the spirit of [2] (see Assumption 4.2 which is a consequence of Assumption 3.1).
- Regularity (isotropic Hölder regularity) of the density of μ with respect to the Lebesgue measure on $S = \mathbb{R}^d$, see Assumption 3.4 (i).
- Regularity of the kernel function K and on the bandwidth given in Assumption 3.3 and Assumption 3.4 (ii)–(iii).
- A condition on the bandwidth given in Equation (11) which add a further restriction only in the supercritical regime $\alpha > 1/\sqrt{2}$.

Eventually, we present some simulations on the kernel estimation of the density of μ. We note that in statistical studies which have been done in [4, 5, 11], the ergodic rate of convergence is assumed to be less than 1/2, which is strictly less than the threshold $1/\sqrt{2}$ for criticality. Moreover, in the latter works, the authors are interested in the non-asymptotic analysis of the estimators. Now, with the new perspective given by the present results, see in particular Remark 3.6, we think that the works in [4, 5, 11] can be extended to the case where the ergodic rate of convergence belongs to $(1/2, 1)$.

The paper is organized as follows. We introduce the BMC model in Sect. 2 as well as the L^2 ergodic assumption. We define the kernel estimator and state the main results on the estimation of the density of μ, see Theorem 3.5 (consistency and asymptotic normality), in Sect. 3.1. The proofs of those results rely on a general central limit theorem, see Theorem 4.8 in Sect. 4. In Sect. 3.2, we illustrate our results by studying the symmetric BAR, and we provide a numerical study in Sect. 3.3. Sections 4.2 and 5 are dedicated to the proofs of the main results.

2 Bifurcating Markov Chain (BMC)

We denote by $\mathbb{N}$ the set of nonnegative integers and $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. If $(E, \mathcal{E})$ is a measurable space, then $\mathcal{B}(E)$ (resp. $\mathcal{B}_b(E)$, resp. $\mathcal{B}_+(E)$) denotes the set of (resp. bounded, resp. nonnegative) $\mathbb{R}$-valued measurable functions defined on E. For $f \in \mathcal{B}(E)$, we set $\| f \|_\infty = \sup\{|f(x)|,\ x \in E\}$. For a finite measure λ on $(E, \mathcal{E})$ and $f \in \mathcal{B}(E)$ we shall write $\langle \lambda, f \rangle$ for $\int f(x)\, \mathrm{d}\lambda(x)$ whenever this integral is well defined, and $\| f \|_{L^2(\lambda)} = \langle \lambda, f^2 \rangle^{1/2}$. For $n \in \mathbb{N}^*$, the product space E^n is endowed with the product σ-field $\mathcal{E}^{\otimes n}$. If (E, d) is a metric space, then $\mathcal{E}$ will denote its Borel σ-field and the set $\mathcal{C}_b(E)$ (resp. $\mathcal{C}_+(E)$) denotes the set of bounded (resp. nonnegative) $\mathbb{R}$-valued continuous functions defined on E.

Let $(S, \mathscr{S})$ be a measurable space. Let Q be a probability kernel on $S \times \mathscr{S}$, that is: $Q(\cdot, A)$ is measurable for all $A \in \mathscr{S}$, and $Q(x, \cdot)$ is a probability measure on $(S, \mathscr{S})$ for all $x \in S$. For any $f \in \mathcal{B}_b(S)$, we set for $x \in S$:

$$(Qf)(x) = \int_S f(y)\ Q(x, \mathrm{d}y). \tag{1}$$

We define (Qf), or simply Qf, for $f \in \mathcal{B}(S)$ as soon as the integral (1) is well defined, and we have $Qf \in \mathcal{B}(S)$. For $n \in \mathbb{N}$, we denote by Q^n the nth iterate of Q defined by $Q^0 = I$, the identity map on $\mathcal{B}(S)$, and $Q^{n+1} f = Q^n(Qf)$ for $f \in \mathcal{B}_b(S)$.

Let P be a probability kernel on $S \times \mathscr{S}^{\otimes 2}$, that is: $P(\cdot, A)$ is measurable for all $A \in \mathscr{S}^{\otimes 2}$, and $P(x, \cdot)$ is a probability measure on $(S^2, \mathscr{S}^{\otimes 2})$ for all $x \in S$. For any $g \in \mathcal{B}_b(S^3)$ and $h \in \mathcal{B}_b(S^2)$, we set for $x \in S$:

$$(Pg)(x) = \int_{S^2} g(x, y, z)\ P(x, \mathrm{d}y, \mathrm{d}z) \quad \text{and} \quad (Ph)(x) = \int_{S^2} h(y, z)\ P(x, \mathrm{d}y, \mathrm{d}z). \tag{2}$$

We define (Pg) (resp. (Ph)), or simply Pg for $g \in \mathcal{B}(S^3)$ (resp. Ph for $h \in \mathcal{B}(S^2)$), as soon as the corresponding integral (2) is well defined, and we have that Pg and Ph belong to $\mathcal{B}(S)$.

We now introduce some notations related to the regular binary tree. We set $\mathbb{T}_0 = \mathbb{G}_0 = \{\emptyset\}$, $\mathbb{G}_k = \{0, 1\}^k$ and $\mathbb{T}_k = \bigcup_{0\leq r\leq k} \mathbb{G}_r$ for $k \in \mathbb{N}^*$, and $\mathbb{T} = \bigcup_{r\in\mathbb{N}} \mathbb{G}_r$. The set $\mathbb{G}_k$ corresponds to the kth generation, $\mathbb{T}_k$ to the tree up to the kth generation, and $\mathbb{T}$ the complete binary tree. For $i \in \mathbb{T}$, we denote by $|i|$ the generation of i ($|i| = k$ if and only if $i \in \mathbb{G}_k$) and $iA = \{ij;\ j \in A\}$ for $A \subset \mathbb{T}$, where ij is the concatenation of the two sequences $i, j \in \mathbb{T}$, with the convention that $\emptyset i = i\emptyset = i$.

We recall the definition of bifurcating Markov chain from [12].

Definition 2.1 We say a stochastic process indexed by $\mathbb{T}$, $X = (X_i, i \in \mathbb{T})$, is a bifurcating Markov chain (BMC) on a measurable space $(S, \mathscr{S})$ with initial probability distribution ν on $(S, \mathscr{S})$ and probability kernel $\mathcal{P}$ on $S \times \mathscr{S}^{\otimes 2}$ if:

- (Initial distribution.) The random variable $X_\emptyset$ is distributed as ν.
- (Branching Markov property.) For any sequence $(g_i, i \in \mathbb{T})$ of functions belonging to $\mathcal{B}_b(S^3)$, we have for all $k \geq 0$,

$$\mathbb{E}\Big[\prod_{i\in\mathbb{G}_k} g_i(X_i, X_{i0}, X_{i1})|\sigma(X_j;\ j \in \mathbb{T}_k)\Big] = \prod_{i\in\mathbb{G}_k} \mathcal{P}g_i(X_i).$$

Let $X = (X_i, i \in \mathbb{T})$ be a BMC on a measurable space $(S, \mathscr{S})$ with initial probability distribution ν and probability kernel $\mathcal{P}$. We define three probability kernels P_0, P_1 and $\mathcal{Q}$ on $S \times \mathscr{S}$ by:

$$P_0(x, A) = \mathcal{P}(x, A \times S), \quad P_1(x, A) = \mathcal{P}(x, S \times A) \quad \text{for}(x, A) \in S \times \mathscr{S}, \text{and}$$
$$\mathcal{Q} = \frac{1}{2}(P_0 + P_1).$$

Notice that P_0 (resp. P_1) is the restriction of the first (resp. second) marginal of $\mathcal{P}$ to S. Following [12], we introduce an auxiliary Markov chain $Y = (Y_n, n \in \mathbb{N})$ on $(S, \mathscr{S})$ with Y_0 distributed as $X_\emptyset$ and transition kernel $\mathcal{Q}$. The distribution of Y_n corresponds to the distribution of X_I, where I is chosen independently from X and uniformly at random in generation $\mathbb{G}_n$. We shall write $\mathbb{E}_x$ when $X_\emptyset = x$ (*i.e.*, the initial distribution ν is the Dirac mass at $x \in S$).

Remark 2.2 If the Markov chain Y is ergodic and if μ denotes its unique invariant probability measure, then Guyon proves in [12] that, when S is a metric space, for all $f \in \mathcal{C}_b(S)$,

$$|\mathbb{A}_n|^{-1} \sum_{u\in\mathbb{A}_n} f(X_u) \xrightarrow[n\to\infty]{} \langle \mu, f\rangle \quad \text{in probability,} \quad \text{where} \quad \mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}.$$

One can then see that the study of BMC is strongly related to the knowledge of μ. However, when it exists, the invariant probability μ is generally not known. The aim of this article is then to estimate μ and study, under appropriate hypotheses, the fluctuations of the estimators of μ.

We consider the following ergodic properties of $\mathcal{Q}$, which in particular implies that μ is indeed the unique invariant probability measure for $\mathcal{Q}$. We refer to [9] Section

22 for a detailed account on $L^2(\mu)$-ergodicity (and in particular Definition 22.2.2 on exponentially convergent Markov kernel).

Assumption 2.3 (*Geometric ergodicity*) The Markov kernel $\mathcal{Q}$ has an (unique) invariant probability measure μ, and $\mathcal{Q}$ is $L^2(\mu)$ exponentially convergent, that is there exists $\alpha \in (0, 1)$ and M finite such that for all $f \in L^2(\mu)$:

$$\| \mathcal{Q}^n f - \langle \mu, f \rangle \|_{L^2(\mu)} \leq M\alpha^n \| f \|_{L^2(\mu)} \quad \text{for all} n \in \mathbb{N}. \tag{3}$$

3 Main Result

3.1 Kernel Estimator of the Density μ

The purpose of this Section is to study asymptotic normality of kernel estimators for the density of the stationary measure of a BMC. Assume that $S = \mathbb{R}^d$, with $d \geq 1$, and that the invariant measure μ of the transition kernel $\mathcal{Q}$ exists is unique and has a density, still denoted by μ, with respect to the Lebesgue measure. Our aim is to estimate the density μ from the observation of the population over the nth generation $\mathbb{G}_n$ of over $\mathbb{T}_n$, that is up to generation n. For that purpose, assume we observe $\mathbb{X}^n = (X_u)_{u\in\mathbb{A}_n}$, where $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$, *i.e.,* we have $2^{n+1} - 1$ (or 2^n) random variables with value in S. We consider an integrable kernel function $K \in \mathcal{B}(S)$ such that $\int_S K(x)\,\mathrm{d}x = 1$ and a sequence of positive bandwidths $(h_n, n \in \mathbb{N})$ which converges to 0 as n goes to infinity. Then, we can define the estimation of the density of μ at $x \in S$ over individuals $\mathbb{A}_n \in \{\mathbb{T}_n, \mathbb{G}_n\}$ with kernel K and bandwidth $(h_n, n \in \mathbb{N})$ as:

$$\widehat{\mu}_{\mathbb{A}_n}(x) = |\mathbb{A}_n|^{-1} h_n^{-d/2} \sum_{u\in\mathbb{A}_n} K_{h_n}(x - X_u), \tag{4}$$

where for $h > 0$ the rescaled kernel function K_h is given for $y \in S$ by:

$$K_h(y) = h^{-d/2} K(h^{-1} y).$$

Those statistics are strongly inspired from [16, 18, 19]. For $h > 0$ and $u \in \mathbb{T}$, we set:

$$K_h \star \mu(x) = \mathbb{E}_\mu[K_h(x - X_u)] = \int_S K_h(x - y)\mu(y)\,\mathrm{d}y.$$

We have the following bias-variance type decomposition of the estimator $\widehat{\mu}_{\mathbb{A}_n}(x)$:

$$\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x) = B_{h_n}(x) + V_{\mathbb{A}_n,h_n}(x), \tag{5}$$

where for $h > 0$ and $\mathbb{A} \subset \mathbb{T}$ finite:

$$B_h(x) \; = h^{-d/2} K_h \star \mu(x) - \mu(x) \quad \text{and} \quad V_{\mathbb{A},h}(x)$$

$$= |\mathbb{A}|^{-1} h^{-d/2} \sum_{u \in \mathbb{A}} \Big(K_h(x - X_u) - K_h \star \mu(x) \Big).$$

Our aim is to study the convergence and the asymptotic normality of the estimator $\widehat{\mu}_{\mathbb{A}_n}(x)$ of $\mu(x)$. This relies on a series of assumption on the model, that is on $\mathcal{P}$, $\mathcal{Q}$ and μ, and on the kernel function K as well as the bandwidth $(h_n, n \in \mathbb{N})$.

We first state a series of assumption of the density of the kernel $\mathcal{P}$ and the initial distribution ν with respect to the invariant measure.

Assumption 3.1 (*Regularity of* $\mathcal{P}$ *and* ν_0) We assume that:

(i) There exists an invariant probability measure μ of $\mathcal{Q}$ and the transition kernel $\mathcal{P}$ has a density, denoted by p, with respect to the measure $\mu^{\otimes 2}$, that is, for all $x \in S$:

$$\mathcal{P}(x, \mathrm{d}y, \mathrm{d}z) = p(x, y, z)\, \mu(\mathrm{d}y)\mu(\mathrm{d}z).$$

(ii) The following function $\mathfrak{h}$ defined on S belongs to $L^2(\mu)$, where:

$$\mathfrak{h}(x) = \left(\int_S q(x, y)^2\, \mu(\mathrm{d}y) \right)^{1/2}, \tag{6}$$

with $q(x, y) = 2^{-1} \int_S (p(x, y, z) + p(x, z, y))\, \mu(\mathrm{d}z)$, the density of $\mathcal{Q}$ with respect to μ.

(iii) There exists $k_1 \geq 1$ such that $\mathfrak{h}_{k_1} \in L^6(\mu)$, where for $k \in \mathbb{N}^*$:

$$\mathfrak{h}_k = \mathcal{Q}^{k-1} \mathfrak{h}.$$

(iv) There exists $k_0 \in \mathbb{N}$, such that the probability measure $\nu \mathcal{Q}^{k_0}$ has a bounded density, say ν_0, with respect to μ:

$$\nu \mathcal{Q}^{k_0}(\mathrm{d}y) = \nu_0(y)\mu(dy) \quad \text{and} \quad \| \nu_0 \|_\infty < +\infty.$$

On the one hand, Conditions (i), (ii) and (iv) can be seen as standard L^2 condition for ergodic Markov chains. On the other hand, even in the simpler symmetric BAR model presented in Sect. 3.2, it may happen that $\mathfrak{h}$ has no finite higher moments (which are used in the proof of the asymptotic normality to check Lindeberg's condition using a fourth moment condition, see also Assumption 4.2). This motivated the introduction of Condition (iii).

Then, we consider the $\mathbb{R}^d$-valued case and assume further integrability condition on the density of $\mathcal{P}$ and $\mathcal{Q}$, and the existence of the density of μ with respect to the Lebesgue measure.

Assumption 3.2 (*Regularity of* μ *and integrability conditions*) Let $S = \mathbb{R}^d$ with $d \geq 1$. Assume that Assumption 3.1 (i) holds.

(i) The invariant measure μ of the transition kernel $\mathcal{Q}$ has a density, still denoted by μ, with respect to the Lebesgue measure.
(ii) The following constants are finite:

$$C_0 = \sup_{x,y\in\mathbb{R}^d} \big(\mu(x) + q(x,y)\mu(y)\big), \tag{7}$$

$$C_1 = \sup_{y,z\in\mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x)\mu(y)\mu(z)p(x,y,z), \tag{8}$$

$$C_2 = \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x) \sup_{z\in\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathrm{d}y\, \mu(y)\mathfrak{h}(y)\, \mu(z)\big(p(x,y,z) + p(x,z,y)\big) \right)^2 . \tag{9}$$

Following [17, Theorem 1A] (which we consider in dimension d, see Lemma 6.1 below), we shall consider the following assumptions. For $g \in \mathcal{B}(\mathbb{R}^d)$, we set $\| g \|_p = (\int_S |g(y)|^p \,\mathrm{d}y)^{1/p}$. Then, we consider condition of the kernel function.

Assumption 3.3 (*Regularity of the kernel function and the bandwidths*) Let $S = \mathbb{R}^d$ with $d \geq 1$.

(i) The kernel function $K \in \mathcal{B}(\mathbb{R}^d)$ satisfies:

$$\| K \|_\infty < +\infty,\ \| K \|_1 < +\infty,\ \| K \|_2 < +\infty,\ \int_{\mathbb{R}^d} K(x)\,\mathrm{d}x = 1 \quad \text{and} \quad \lim_{|x|\to+\infty} |x|K(x) = 0. \tag{10}$$

(ii) There exists $\gamma \in (0, 1/d)$ such that the bandwidths $(h_n, n \in \mathbb{N})$ are defined by $h_n = 2^{-n\gamma}$.

The following regularity assumptions on μ, the kernel function K and the bandwidth sequence $(h_n, n \in \mathbb{N})$ will be useful to control the bias term in (5). We follow Tsybakov [20], chapter 1. For $s \in \mathbb{R}_+$, let $\lfloor s \rfloor$ denote its integer part, that is the only integer $n \in \mathbb{N}$ such that $n \leq s < n+1$ and set $\{s\} = s - \lfloor s \rfloor$ its fractional part.

Assumption 3.4 (*Further regularity on the density μ, the kernel function and the bandwidths*) Suppose that there exists an invariant probability measure μ of $\mathcal{Q}$ and that Assumptions3.2 (i) and 3.3 hold. We assume there exists $s > 0$ such that the following holds:

(i) *The density μ belongs to the (isotropic) Hölder class of order $(s, \dots, s) \in \mathbb{R}^d$:* The density μ admits partial derivatives with respect to x_j, for all $j \in \{1, \dots d\}$, up to the order $\lfloor s \rfloor$ and there exists a finite constant $L > 0$ such that for all $x = (x_1, \dots, x_d), \in \mathbb{R}^d$, $t \in \mathbb{R}$ and $j \in \{1, \dots, d\}$:

$$\left| \frac{\partial^{\lfloor s \rfloor}\mu}{\partial x_j^{\lfloor s \rfloor}}(x_{-j}, t) - \frac{\partial^{\lfloor s \rfloor}\mu}{\partial x_j^{\lfloor s \rfloor}}(x) \right| \leq L|x_j - t|^{\{s\}},$$

where (x_{-j}, t) denotes the vector x where we have replaced the jth coordinate x_j by t, with the convention $\partial^0 \mu / \partial x_j^0 = \mu$.

(ii) *The kernel K is of order* $(\lfloor s \rfloor, \dots, \lfloor s \rfloor) \in \mathbb{N}^d$ We have $\int_{\mathbb{R}^d} |x|^s K(x)\, \mathrm{d}x < \infty$ and $\int_{\mathbb{R}} x_j^k K(x)\, \mathrm{d}x_j = 0$ for all $k \in \{1, \dots, \lfloor s \rfloor\}$ and $j \in \{1, \dots, d\}$.

(iii) *Bandwidth control* The bandwidths $(h_n, n \in \mathbb{N})$ satisfy $\lim_{n\to\infty} |\mathbb{G}_n|\, h_n^{2s+d} = 0$, that is $\gamma > 1/(2s + d)$.

Notice that Assumption 3.4-(i) implies that μ is at least Hölder continuous as $s > 0$.

The following theorem, which proof is given in Sect. 4.2, provides the consistency and the asymptotic normality of the estimator $\widehat{\mu}_{\mathbb{A}_n}(x)$ of $\mu(x)$, for x in the set of continuity of μ.

Theorem 3.5 (Convergence and asymptotic normality of the kernel density estimator) *Let $d \geq 1$. Let X be a $\mathbb{R}^d$-valued BMC with kernel $\mathcal{P}$ and initial distribution ν, K a kernel function and $(h_n, n \in \mathbb{N})$ a bandwidth sequence such that Assumptions 2.3 (on the geometric ergodicity), 3.1 (on the regularity of $\mathcal{P}$ and of ν), Assumptions 3.2 (on the density of μ and $\mathcal{P}$), Assumptions 3.3 (on the kernel function K and the bandwidths $(h_n, n \in \mathbb{N})$), and Assumptions 3.4 (on the density μ, K and $(h_n, n \in \mathbb{N})$) are in force.*

Furthermore, if the ergodic rate of convergence α (given in Assumption 2.3) is such that $\alpha > 1/\sqrt{2}$, then assume that the bandwidth rate γ (given in Assumption 3.3 (ii)) is such that:

$$2^{d\gamma} > 2\alpha^2. \tag{11}$$

Then, for x in the set of continuity of μ and $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$, we have the following convergence:

$$\widehat{\mu}_{\mathbb{A}_n}(x) \xrightarrow[n\to\infty]{\mathbb{P}} \mu(x) \quad \textit{in probability}, \tag{12}$$

$$|\mathbb{A}_n|^{1/2} h_n^{d/2} (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)) \xrightarrow[n\to\infty]{(\mathrm{d})} G \quad \textit{in distribution}, \tag{13}$$

where G is a centered Gaussian real-valued random variable with variance $\| K \|_2^2\, \mu(x)$.

Remark 3.6 The bandwidth must be a function of the geometric ergodic rate of convergence via the relation $2^{d\gamma} > 2\alpha^2$ given in Eq. (11). Notice this condition is automatically satisfied in the critical and subcritical case ($\alpha \leq 1/\sqrt{2}$) as $\gamma > 0$. In the supercritical case ($\alpha > 1/\sqrt{2}$), the geometric rate of convergence α could be interpreted as a regularity parameter for the bandwidth selection problems of the estimation of $\mu(x)$, just like the regularity of the unknown function μ. With this new perspective, we think that the results in [5] could be extended to $\alpha \in (1/2, 1)$ by studying an adaptive procedure with respect to the unknown geometric rate of convergence α.

Remark 3.7 We stress that the asymptotic variance is the same for $\mathbb{A}_n = \mathbb{G}_n$ and $\mathbb{A}_n = \mathbb{T}_n$. This is a consequence of the structure of the asymptotic variance σ^2 in (28) and (39), and the fact that $\lim_{n\to\infty} |\mathbb{T}_n|/|\mathbb{G}_n| = 2$.

Remark 3.8 One can prove that the estimators $|\mathbb{G}_{n-\ell}|^{1/2} h_{n-\ell}^{d/2}(\widehat{\mu}_{\mathbb{G}_{n-\ell}}(x) - \mu(x))$ are asymptotically independent for $\ell \in \{0, \dots, k\}$ for any $k \in \mathbb{N}$. This result relies on the additive structure of the asymptotic variance σ^2 in (28), see also Remark 4.9 or consider the functions $f_{\ell,n} = f_{\ell,n}^{\text{shift}}$ given by (37) in the proof of Theorem 3.5.

Now, for the applications (e.g., obtaining confidence interval for μ), it would be interesting in Theorem 3.5 to replace $\mu(x)$ in the expression of the asymptotic variance by an estimator. For that purpose, we consider $(\varpi_n, n \in \mathbb{N})$, a sequence of real numbers such that $\varpi_n \to 0$ as $n \to +\infty$. We consider the set $\mathbb{A}_n^*$ belonging to $\{\mathbb{G}_n, \mathbb{T}_n\}$. We will allow $\mathbb{A}_n$ and $\mathbb{A}_n^*$ to be identical. For two numbers a and b, we set $a \vee b = \max\{a, b\}$. Then the following result is a direct consequence of Theorem 3.5.

Corollary 3.9 *Under the hypotheses of Theorem 3.5, we have*

$$(\|K\|_2 \sqrt{\widehat{\mu}_{\mathbb{A}_n^*}(x)} \vee \varpi_n)^{-1} |\mathbb{A}_n|^{1/2} h_n^{d/2} (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)) \xrightarrow[n\to\infty]{\text{(d)}} Z \quad \textit{in distribution},$$

where Z is a centered Gaussian real-valued random variable with variance 1.

3.2 Application to the Study of Symmetric BAR

3.2.1 The Model

We consider a particular case from [6] of the real-valued bifurcating autoregressive process (BAR), see also [2, Section 4]. More precisely, let $a \in (-1, 1)$. We consider the process $X = (X_u, u \in \mathbb{T})$ on $S = \mathbb{R}$ where for all $u \in \mathbb{T}$:

$$\begin{cases} X_{u0} = aX_u + \varepsilon_{u0}, \\ X_{u1} = aX_u + \varepsilon_{u1}, \end{cases}$$

with $(\varepsilon_u, u \in \mathbb{T})$ an independent sequence of real Gaussian $\mathcal{N}(0, \sigma^2)$ random variables independent of $X_\emptyset$, with $\sigma > 0$. Then the process $X = (X_u, u \in \mathbb{T})$ is a BMC with transition probability $\mathcal{P}$ given by:

$$\mathcal{P}(x, \mathrm{d}y, \mathrm{d}z) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y - ax)^2 + (z - ax)^2}{2\sigma^2}\right) \mathrm{d}y\mathrm{d}z = \mathcal{Q}(x, \mathrm{d}y)\mathcal{Q}(x, \mathrm{d}z),$$

where the transition kernel $\mathcal{Q}$ of the auxiliary Markov chain is defined by:

$$\mathcal{Q}(x, \mathrm{d}y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - ax)^2}{2\sigma^2}\right) \mathrm{d}y.$$

We have $\mathcal{Q}f(x) = \mathbb{E}[f(ax + \sigma G)]$ and more generally:

$$\mathcal{Q}^n f(x) = \mathbb{E}\left[f\left(a^n x + \sqrt{1 - a^{2n}}\, \sigma_a G\right)\right], \tag{14}$$

where G is a standard $\mathcal{N}(0,1)$ Gaussian random variable and $\sigma_a = \sigma/\sqrt{1-a^2}$. The kernel $\mathcal{Q}$ admits a unique invariant probability measure μ, which is $\mathcal{N}(0,\sigma_a^2)$ and whose density, still denoted by μ, with respect to the Lebesgue measure is given by:

$$\mu(x) = \frac{\sqrt{1-a^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(1-a^2)x^2}{2\sigma^2}\right). \tag{15}$$

The density p (resp. q) of the kernel $\mathcal{P}$ (resp. $\mathcal{Q}$) with respect to $\mu^{\otimes 2}$ (resp. μ) is given by:

$$p(x,y,z) = q(x,y)q(x,z) \tag{16}$$

and

$$\begin{aligned} q(x,y) &= \frac{1}{\sqrt{1-a^2}} \exp\left(-\frac{(y-ax)^2}{2\sigma^2} + \frac{(1-a^2)y^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{1-a^2}}\, \mathrm{e}^{-(a^2y^2+a^2x^2-2axy)/2\sigma^2}. \end{aligned}$$

In particular, we have:

$$\mu(x)\, q(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-ay)^2}{2\sigma^2}\right).$$

3.2.2 Regularity of the Model, and Verification of the Assumptions

We first check that Assumption 2.3 on the geometric ergodicity holds. Since q is symmetric, the operator $\mathcal{Q}$ (in $L^2(\mu)$) is a symmetric integral Hilbert–Schmidt operator. Furthermore, its eigenvalues are given by $\sigma_p(\mathcal{Q}) = (a^n, n \in \mathbb{N})$, with their algebraic multiplicity being one. So Assumption 2.3 holds with $\alpha = |a|$ as $a \in (-1,1)$.

We check Assumption 3.1 on the regularity of $\mathcal{P}$ and ν_0. Condition (i) therein holds thanks to (16). Recall $\mathfrak{h}$ defined in (6) and $\sigma_a = \sigma/\sqrt{1-a^2}$. It is not difficult to check that for $x \in \mathbb{R}$:

$$\mathfrak{h}(x) = (1-a^4)^{-1/4} \exp\left(\frac{a^2}{1+a^2}\, \frac{x^2}{2\sigma_a^2}\right), \tag{17}$$

and thus $\mathfrak{h} \in L^2(\mu)$ (that is $\int_{\mathbb{R}^2} q(x,y)^2\, \mu(x)\mu(y)\, \mathrm{d}x\mathrm{d}y < +\infty$). Thus Condition (ii) holds.

We now consider Condition (iii), that is $\mathfrak{h}_k = \mathcal{Q}^{k-1}\mathfrak{h}$ belongs to $L^6(\mu)$ for some $k \geq 1$. We deduce from (14) and (17) that there exists a finite constant C_k such that:

$$\mathfrak{h}_k(x) = \mathcal{Q}^{k-1}\mathfrak{h}(x) = C_k \exp\left(\frac{a^{2k}}{1+a^{2k}}\, \frac{x^2}{2\sigma_a^2}\right).$$

So we deduce that $\mathfrak{h}_k$ belongs to $L^6(\mu)$ if and only if $a^{2k} < 1/5$, which is satisfied for k large enough as $a \in (-1, 1)$. Thus, Condition (iii) holds.

Remark 3.10 As we shall see, Assumption 3.1 (iii) (the 6th moment of h_k being finite for some $k \in \mathbb{N}^*$) is used to check (25) and (26) from Assumption 4.2, see Sect. 4. So one could ask if those two inequalities could hold without Condition (iii). In fact, using elementary computations, it is possible to check the following. For $k_1 = 1$, (25) holds for $|a| < 3^{-1/4}$ and (26) also holds for $|a| \leq 0.724$ (but (26) fails for $|a| \geq 0.725$). (Notice that $2^{-1/2} < 0.724 < 3^{-1/4}$.) For $k_1 = 2$, (25) holds for $|a| < 3^{-1/6}$ and (26) also holds for $|a| \leq 0.794$ (but (26) fails for $|a| \geq 0.795$). So we see that checking (25) and (26) is rather tricky. This motivated the introduction of the stronger Condition (iii) from Assumption 3.1.

We now comment on Condition (iv) from Assumption 3.1. Notice that $\nu \mathcal{Q}^k$ is the probability distribution of $a^k X_\emptyset + \sigma_a \sqrt{1 - a^{2k}}\, G$, with G a $\mathcal{N}(0, 1)$ random variable independent of $X_\emptyset$. So Condition (iv) holds in particular if ν has compact support (with $k_0 = 1$) or if ν has a density with respect to the Lebesgue measure, which we still denote by ν, such that $\| \mathrm{d}\nu/\mathrm{d}\mu \|_\infty$ is finite (with $k_0 = 0$). Notice that if ν is the Gaussian probability distribution of $\mathcal{N}(m_0, \rho_0^2)$, then Condition (iv) holds if and only if $\rho_0 < \sigma_a$ and $m_0 \in \mathbb{R}$, or $\rho_0 = \sigma_a$ and $m_0 = 0$.

We now check Assumptions 3.2 on the regularity of μ and on the integrability conditions on the density of $\mathcal{P}$ and $\mathcal{Q}$. Condition (i) holds, see (15) for the density of μ with respect to the Lebesgue measure. We now check that Condition (ii) holds, that is the constants C_0, C_1 and C_2 defined in (7), (8) and (9) are finite. The fact that C_0 is finite is clear. Notice that:

$$\begin{aligned} C_1 &= \sup_{y,z \in \mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x)\mu(y)\mu(z) p(x, y, z) \\ &= \sup_{y,z \in \mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x)\mu(y)\mu(z) q(x, y) q(x, z) \leq C_0^2. \end{aligned}$$

We also have, using Jensen for the second inequality (and the probability measure $\mu(y)q(x, y)\,\mathrm{d}y$):

$$\begin{aligned} C_2 &= 4 \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x) \sup_{z \in \mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathrm{d}y\, \mu(y)\mathfrak{h}(y)\, \mu(z) q(x, y) q(x, z) \right)^2 \\ &\leq 4C_0^2 \int_{\mathbb{R}^d} \mathrm{d}x\, \mu(x) \left(\int_{\mathbb{R}^d} \mathrm{d}y\, \mu(y)\mathfrak{h}(y)\, q(x, y) \right)^2 \\ &\leq 4C_0^2 \, \| \mathfrak{h} \|_{L^2(\mu)}^2 . \end{aligned}$$

So, we get that the constants C_0, C_1 and C_2 are finite, and thus Condition (ii) holds.

Since the function μ given in (15) is of class $\mathcal{C}^\infty$ with all its derivative bounded, we get that the Hölder type Assumption 3.4 (i) holds (for any $s > 0$).

Many choices of the kernel function, K, and of the bandwidths parameter γ satisfy Assumptions 3.3 and 3.4 (ii) and (iii). Eventually, as $d = 1$ and $\alpha = |a|$, we get that Equation (11) becomes $2^\gamma > 2a^2$, which holds *a fortiori* if $2a^2 \leq 1$.

3.3 Numerical Studies for the Symmetric BAR Model

In order to illustrate the central limit theorem for the estimator of the invariant density μ, we simulate $n_0 = 500$ samples of a symmetric BAR $X = (X_u^{(a)}, u \in \mathbb{T}_n)$ with different values of the autoregressive coefficient $\alpha = a \in (-1, 1)$. For each sample, we compute the estimator $\widehat{\mu}_{\mathbb{A}_n}(x)$ given in (4) and its fluctuation given by

$$\zeta_n = |\mathbb{A}_n|^{1/2} h_n^{d/2} \left(\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)\right) \tag{18}$$

for $x \in \mathbb{R}$, the average over $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$, the Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-x^2/2} \tag{19}$$

and the bandwidth $h_n = 2^{-n\gamma}$ with $\gamma \in (0, 1)$. Next, in order to compare theoretical and empirical results, we plot in the same graphic, see Figs. 1 and 2:

- The histogram of ζ_n and the density of the centered Gaussian distribution with variance $\mu(x) \parallel K \parallel_2^2 = \mu(x)(2\sqrt{\pi})^{-1}$ (see Theorem 3.5).
- The empirical cumulative distribution of ζ_n and the cumulative distribution of the centered Gaussian distribution with variance $\mu(x) \parallel K \parallel_2^2 = \mu(x)(2\sqrt{\pi})^{-1}$.

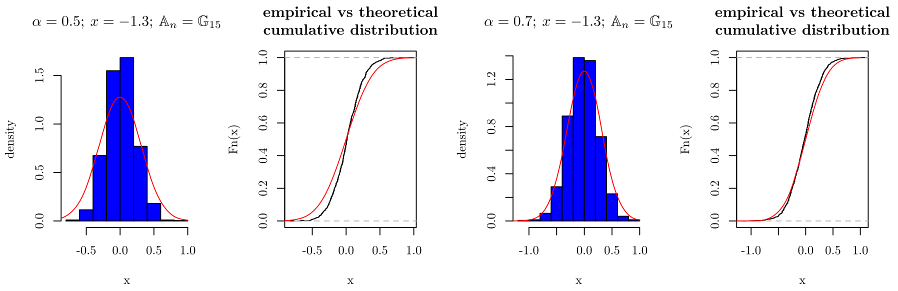
Since the Gaussian kernel is of order $s = 2$ and the dimension is $d = 1$, the bandwidth exponent γ must satisfy the condition $\gamma > 1/5$, so that Assumption 3.4-(iii) holds. Moreover, in the supercritical case, γ must satisfy the supplementary condition $2^\gamma > 2\alpha^2$, that is $\gamma > 1 + \log(\alpha^2)/\log(2)$, so that (11) holds. In Fig. 1, we take $\alpha = 0.5$ and $\alpha = 0.7$ (both of them corresponds to the subcritical case as $2\alpha^2 < 1$) and $\gamma = 1/5 + 10^{-3}$. The simulations agree with the results from Theorem 3.5. In Fig. 2, we take $\alpha = 0.9$ (supercritical case) and consider $\gamma = 0.696$ and $\gamma = 1/5 + 10^{-3}$. In the former case (11) is satisfied as $\gamma = 0.696 > 1 + \log((0.9)^2)/\log(2)$, and in the latter case (11) fails. As one can see in the graphics Fig. 2, the estimates agree with the theory in the former case ($\gamma = 0.696$), whereas they are poor in the latter case.

3.4 The Model of Asymmetric BAR

Now, we study an extension introduced in [12] of the symmetric BAR . We consider the real-valued Gaussian bifurcating autoregressive process (BAR) $X = (X_u, u \in \mathbb{T})$ where $X_\emptyset$ is arbitrary and for all $u \in \mathbb{T}$:

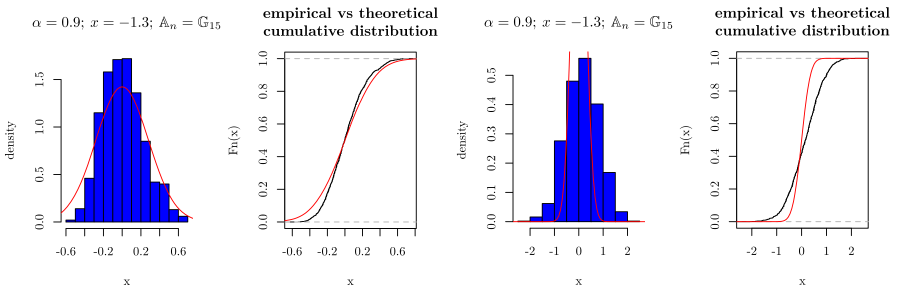
$$\begin{cases} X_{u0} = a_0 X_u + b_0 + \varepsilon_{u0} \\ X_{u1} = a_1 X_u + b_1 + \varepsilon_{u1}, \end{cases} \tag{20}$$

with $a_0, a_1 \in [-1, 1]$, $b_0, b_1 \in \mathbb{R}$ and $((\varepsilon_{u0}, \varepsilon_{u1}),\ u \in \mathbb{T})$ an independent sequence of bivariate Gaussian $\mathcal{N}(0, \Gamma)$ random vectors independent of $X_\emptyset$ with covariance



(a) $\alpha = 0.5$ **(b)** $\alpha = 0.7$

Fig. 1 Histogram and empirical cumulative distribution of ζ_n given in (18) with $x = -1.3$, $n = 15$, $\mathbb{A}_n = \mathbb{G}_n$ and $\gamma = 1/5 + 10^{-3}$. We consider the (subcritical) ergodic rate of convergence: $\alpha = 0.5$ and $\alpha = 0.7$ (Color figure online)



(a) $\gamma = 0.696$ **(b)** $\gamma = 1/5 + 10^{-3}$

Fig. 2 Histogram and empirical cumulative distribution of ζ_n given in (18) with $x = -1.3$, $n = 15$, $\mathbb{A}_n = \mathbb{G}_n$ and the ergodic rate $\alpha = 0.9$ (supercritical case). We consider the bandwidth exponent $\gamma = 0.696$ (which satisfies (11)) for the two left graphics and $\gamma = 1/5 + 10^{-3}$ (which does not satisfy (11)) for the two right (Color figure online)

matrix, with $\sigma > 0$ and $\rho \in \mathbb{R}$ such that $|\rho| < \sigma^2$:

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}.$$

Then the process $X = (X_u, u \in \mathbb{T})$ is a BMC with transition probability $\mathcal{P}$ given by:

$$\mathcal{P}(x, \mathrm{d}y, \mathrm{d}z) = \frac{1}{2\pi\sqrt{\sigma^4 - \rho^2}} \exp\left(-\frac{\sigma^2}{2(\sigma^4 - \rho^2)}\, g(x, y, z)\right) \mathrm{d}y\mathrm{d}z,$$

with

$$\begin{aligned} g(x, y, z) &= (y - a_0x - b_0)^2 - 2\rho\sigma^{-2}(y - a_0x - b_0)(z - a_1x - b_1) \\ &\quad + (z - a_1x - b_1)^2. \end{aligned}$$

The transition kernel $\mathcal{Q}$ of the auxiliary Markov chain is defined by:

$$\mathcal{Q}(x, \mathrm{d}y) = \frac{1}{2\sqrt{2\pi\sigma^2}} \left(\mathrm{e}^{-(y-a_0x-b_0)^2/2\sigma^2} + \mathrm{e}^{-(y-a_1x-b_1)^2/2\sigma^2} \right) \mathrm{d}y.$$

This process admits an invariant probability measure μ which is the law of the random variable Y_∞ defined by

$$Y_\infty := \sum_{k=1}^{\infty} A_1 A_2 \ldots A_{k-1} B_k, \tag{21}$$

where $A_k = a_{\zeta_k}$ and $B_k = b_{\zeta_k} + \varepsilon'_k$, with $(\varepsilon'_k, k \in \mathbb{N}^*)$ and $(\zeta_k, k \in \mathbb{N}^*)$ two independent sequences of i.i.d. random variables, independent of $X_\emptyset$. Each ε'_k is centered Gaussian random variable with variance σ^2 and each ζ_k has a Bernoulli law with parameter $1/2$, that is $\mathbb{P}(\zeta_k = 0) = \mathbb{P}(\zeta_k = 1) = 1/2$. Since the transition $\mathcal{Q}$ has a density with respect to the Lebesgue Measure, it follows that the probability measure μ also has a density, denoted by μ, with respect to the Lebesgue measure. To our best knowledge, the analytic expression of this density is unknown. Note, however, that an approximation of the density μ has been proposed in [8, 15]. We do not have any information on the ergodic convergence rate, except that it is bounded by $\sqrt{(a_0^2 + a_1^2)/2}$ (see, for example, the calculus in [12]). We will estimate the invariant density μ in a compact set $D \subset \mathbb{R}$. For that purpose, we use the estimator $\widehat{\mu}_{\mathbb{G}_n}(x)$, for all $x \in D$, given in (4), with the Gaussian kernel K defined in (19). Since the ergodicity rate is unknown, we have to develop a method based on data in order to select the bandwidth.

3.5 Bandwidth Selection by Cross Validation Method

To select optimal bandwidth, we choose the bandwidth h which minimizes the mean integrated squared error

$$\mathbb{E}\left[\int (\widehat{\mu}_{\mathbb{G}_n}(x) - \mu(x))^2 \mathrm{d}x\right] = \mathbb{E}\left[\int \widehat{\mu}^2_{\mathbb{G}_n}(x)\mathrm{d}x\right] - 2\mathbb{E}\left[\int \widehat{\mu}_{\mathbb{G}_n}(x)\mu(x)\mathrm{d}x\right] + \int \mu^2(x)\mathrm{d}x.$$

Since the last term of the previous equality does not depend on h, it suffices to choose the bandwidth h which minimizes the function J defined by

$$J(h) := \mathbb{E}\left[\int \widehat{\mu}^2_{\mathbb{G}_n}(x)\mathrm{d}x\right] - 2\mathbb{E}\left[\int \widehat{\mu}_{\mathbb{G}_n}(x)\mu(x)\mathrm{d}x\right].$$

Now, $J(h)$ can be approximated by

$$\widehat{J}(h) := \int \widehat{\mu}^2_{\mathbb{G}_n}(x)\mathrm{d}x - \frac{2}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} \widehat{\mu}_{\mathbb{G}_n, [-u]}(X_u),$$

where $\widehat{\mu}_{\mathbb{G}_n,[-u]}$ is the kernel density estimator of μ computed without the observation X_u. The motivation for considering the second approximation come from Remark 2.2. Let $\mathcal{H} := \{h_1, \ldots, h_m\}$ be a bandwidth grid. $\mathcal{H}$ is a subset of $(0, 1]$. Replacing J by $\widehat{J}$, we get that $\widehat{h}$ defined by

$$\widehat{h} := arg\,min_{h\in\mathcal{H}}\widehat{J}(h),$$

is an approximation of the optimal bandwidth for the estimation of the invariant density μ. This method is known in the literature as the leave one out cross-validation (see, for example, [20], Section 1.4). Now, in the numerical studies, instead of $\widehat{\mu}_{\mathbb{G}_n}$ defined in (4) , we use the estimator $\widetilde{\mu}_{\mathbb{G}_n}$ defined by

$$\widetilde{\mu}_{\mathbb{G}_n}(x) = |\mathbb{G}_n|^{-1}\widehat{h}^{-1/2}\sum_{u\in\mathbb{G}_n} K_{\widehat{h}}(x - X_u) \quad \forall x \in \mathbb{R}.$$

Remark 3.11 We do not study here the theoretical properties of $\widetilde{\mu}_{\mathbb{G}_n}$. We let this for future work.

3.6 Numerical Studies for the Asymmetric BAR Model

We consider two examples of the model (20):

(case 1) $(a_0, b_0, a_1, b_1) = (0.5, 0, 0.7, 0)$, $\sigma = 1$ and $\rho = 0$.
(case 2) $(a_0, b_0, a_1, b_1) = (1, 0, 0.5, 0)$, $\sigma = 1$ and $\rho = 0$.

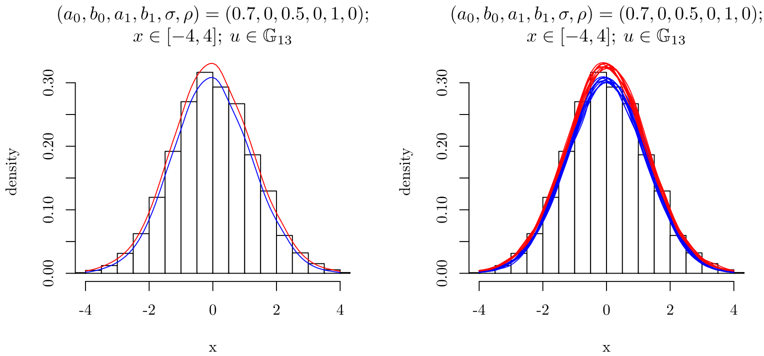
In the second example, we allow the evolution of the new pole to be instable. We have simulated the invariant distribution using the formula (21). Next, we have plotted in the same graph the histogram of data from the simulation of the invariant distribution and, using corollary 3.9, the 90% level confidence band for invariant density on the interval $D = [-4, 4]$. For the case $(a_0, b_0, a_1, b_1) = (0.5, 0, 0.7, 0)$, $\sigma = 1$ and $\rho = 0$, we plot in Fig. 3 the confidence band for one sample and ten samples of the process $(X_u, u \in \mathbb{G}_{13})$, simulated using (20). The bandwidths selected, using the previous cross validation, for the ten samples are

$$h_{CV} = (0.24, 0.21, 0.25, 0.24, 0.24, 0.23, 0.24, 0.25, 0.24, 0.26).$$

We do the same thing for the case $(a_0, b_0, a_1, b_1) = (1, 0, 0.5, 0)$, $\sigma = 1$ and $\rho = 0$. The bandwidths selected in this case are

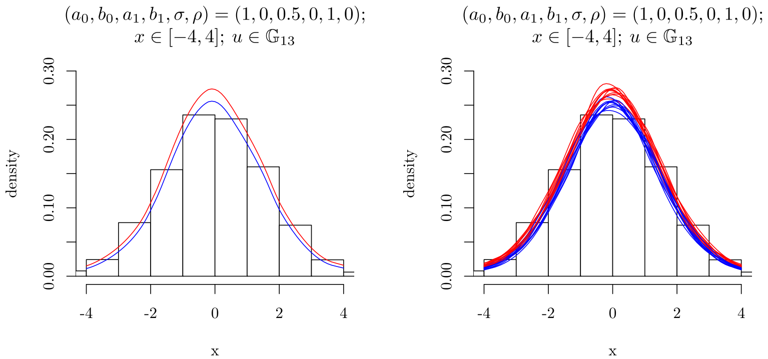
$$h_{CV} = (0.31, 0.32, 0.33, 0.27, 0.30, 0.32, 0.30, 0.32, 0.29, 0.30).$$

One can observe that the selected bandwidths in the second case are greater than those selected in the first case. This is due to the fact that in (case 2), we are certainly in the supercritical case. Indeed, is this case, the upper bound of the geometric ergodic rate of convergence is equal to $\sqrt{(1 + 0.5^2)/2} = 0.791$. This also certainly explains why the invariant distribution is more flattened in the second case, while it is more peaked in



(a) Confidence band for one sample **(b)** Confidence band for ten samples

Fig. 3 Histogram of the invariant distribution (simulated using the formula (21)) and the 90% level confidence band for invariant density for one sample (left) and ten samples (right). In red, the upper bound of the confidence bands and the lower bound in blue (Color figure online)



(a) Confidence band for one sample **(b)** Confidence band for ten samples

Fig. 4 Histogram of the invariant distribution (simulated using the formula (21)) and the 90% level confidence band for invariant density for one sample (left) and ten samples (right). In red, the upper bound of the confidence bands and the lower bound in blue (Color figure online)

the first case. (The upper bound of the geometric ergodic rate of convergence is equal to $\sqrt{(0.7^2 + 0.5^2)/2} = 0.61$.) In Fig. 4, one can observe that the upper and the lower bounds of the confidence bands overlap for different samples. The cross-validation method seems to be less efficient in this case, that is when the evolution of one of the poles is allowed to be instable. Choosing locally the bandwidth would certainly improve the construction of the confidence bands. This last question will be addressed in future work, in the same vein as the work done in [5] but considering this time the geometric ergodic rate of convergence as an unknown regularity parameter.

4 Seeing the Main Result in a More General Framework

The proof of Theorem 3.5 given in Sect. 4.2 relies on a general central limit result for additive functionals of BMC presented in the next section.

4.1 A General CLT for Additive Functionals of BMC

In the spirit of [2], we introduce the following series of assumptions in a general $L^2(\mu)$ framework, with increasing conditions as the geometric ergodic rate α exceed the critical threshold of $1/\sqrt{2}$. In fact, we believe that the general framework presented in this section may be used also for others nonparametric smoothing methods for BMC than the one presented in Sect. 3.1.

Let $X = (X_u, u \in \mathbb{T})$ be a BMC on $(S, \mathscr{S})$ with initial probability distribution ν, and probability kernel $\mathcal{P}$. Recall $\mathcal{Q}$ is the induced Markov kernel. We present some inequalities in the next remark.

Remark 4.1 By convention, for $f, g \in \mathcal{B}(S)$, we define the function $f \otimes g \in \mathcal{B}(S^2)$ by $(f \otimes g)(x, y) = f(x)g(y)$ for $x, y \in S$ and introduce the notations:

$$f \otimes_{\mathrm{sym}} g = \frac{1}{2}(f \otimes g + g \otimes f) \quad \text{and} \quad f\otimes^2 = f \otimes f.$$

Notice that $\mathcal{P}(g \otimes_{\mathrm{sym}} \mathbf{1}) = \mathcal{Q}(g)$ for $g \in \mathcal{B}_+(S)$. For $f \in \mathcal{B}_+(S)$, as $f \otimes f \le f^2 \otimes_{\mathrm{sym}} \mathbf{1}$, we get:

$$\mathcal{P}(f\otimes^2) = \mathcal{P}(f \otimes f) \le \mathcal{P}(f^2 \otimes_{\mathrm{sym}} \mathbf{1}) = \mathcal{Q}\left(f^2\right). \tag{22}$$

Assume μ is an invariant probability measure of $\mathcal{Q}$. By Cauchy–Schwarz we have for $f, g \in L^2(\mu)$:

$$|\mathcal{P}(f \otimes g)|^2 \le \mathcal{P}(f^2 \otimes 1)\, \mathcal{P}(1 \otimes g^2) \le 4\mathcal{Q}(f^2)\, \mathcal{Q}(g^2), \tag{23}$$

$$\langle \mu, \mathcal{P}(f \otimes g)\rangle \le 2 \, \| f \|_{L^2(\mu)} \; \| g \|_{L^2(\mu)} . \tag{24}$$

In the spirit of Assumption 2.4 and Remark 2.5 in [2], we consider the following hypothesis on asymptotic and non-asymptotic distribution of the process.

Assumption 4.2 ($L^2(\mu)$ *regularity for the probability kernel* $\mathcal{P}$ *and density of the initial distribution*)
There exists an invariant probability measure μ of $\mathcal{Q}$ and:

(i) There exists $k_1 \in \mathbb{N}$ and a finite constant M such that for all $f, g \in L^2(\mu)$:

$$\| \mathcal{P}(\mathcal{Q}^{k_1} f \otimes \mathcal{Q}^{k_1} g) \|_{L^2(\mu)} \le M \, \| f \|_{L^2(\mu)} \, \| g \|_{L^2(\mu)}, \tag{25}$$

and for all $h \in L^2(\mu)$, and all $m \in \{0, \dots, k_1\}$:

$$\| \mathcal{P}\left(\mathcal{Q}^m \mathcal{P}(\mathcal{Q}^{k_1} f \otimes_{\mathrm{sym}} \mathcal{Q}^{k_1} g) \otimes_{\mathrm{sym}} \mathcal{Q}^{k_1} h\right) \|_{L^2(\mu)}$$

$$\leq M \parallel f \parallel_{L^2(\mu)} \parallel g \parallel_{L^2(\mu)} \parallel h \parallel_{L^2(\mu)} . \tag{26}$$

(ii) item:k0 There exists $k_0 \in \mathbb{N}$, such that the probability measure $\nu \mathcal{Q}^{k_0}$ has a bounded density, say ν_0, with respect to μ:

$$\nu \mathcal{Q}^{k_0}(\mathrm{d}y) = \nu_0(y)\mu(dy) \quad \text{and} \quad \parallel \nu_0 \parallel_\infty < +\infty.$$

The next family of three assumptions are related to the sequence of functions which will be considered.

Assumption 4.3 (*Regularity of the approximation functions in the subcritical regime*) Let $(f_{\ell,n},\ n \geq \ell \geq 0)$ be a sequence of real-valued measurable functions defined on S such that:

(i) There exists $\rho \in (0, 1/2)$ such that $\sup_{n\geq\ell\geq0} 2^{-n\rho} \parallel f_{\ell,n} \parallel_\infty$ is finite.
(ii) The constants $\mathfrak{c}_2 = \sup_{n\geq\ell\geq0} \parallel f_{\ell,n} \parallel_{L^2(\mu)}$ and $\mathfrak{q}_2 = \sup_{n\geq\ell\geq0} \parallel \mathcal{Q}(f_{\ell,n}^2) \parallel_\infty^{1/2}$ are finite.
(iii) There exists a sequence $(\delta_{\ell,n}, n \geq \ell \geq 0)$ of positive numbers such that $\Delta = \sup_{n\geq\ell\geq0} \delta_{\ell,n}$ is finite, $\lim_{n\to\infty} \delta_{\ell,n} = 0$ for all $\ell \in \mathbb{N}$, and for all $n \geq \ell \geq 0$:

$$\langle \mu, |f_{\ell,n}|\rangle + |\langle \mu, \mathcal{P}(f_{\ell,n}\otimes^2)\rangle| \leq \delta_{\ell,n};$$

and for all $g \in \mathcal{B}_+(S)$:

$$\parallel \mathcal{P}(|f_{\ell,n}| \otimes_{\text{sym}} \mathcal{Q}g) \parallel_{L^2(\mu)} \leq \delta_{\ell,n} \parallel g \parallel_{L^2(\mu)} . \tag{27}$$

(iv) The following limit exists and is finite:

$$\sigma^2 = \lim_{n\to\infty} \sum_{\ell=0}^{n} 2^{-\ell} \parallel f_{\ell,n} \parallel_{L^2(\mu)}^2 < +\infty. \tag{28}$$

Remark 4.4 We stress that (i) and (ii) of Assumption 4.3 imply the existence of finite constant C such that for all $n \geq \ell \geq 0$:

$$\langle \mu, f_{\ell,n}^4\rangle \leq \|f_{\ell,n}\|_\infty^2 \langle \mu, f_{\ell,n}^2\rangle \leq C\, \mathfrak{c}_2^2\, 2^{2n\rho} \quad \text{and} \quad \langle \mu, f_{\ell,n}^6\rangle \leq C\, \mathfrak{c}_2^2\, 2^{4n\rho}.$$

We will use the following notations: for $n \in \mathbb{N}$, set $\mathfrak{f}_n = (f_{\ell,n}, \ell \in \mathbb{N})$ with the convention that $f_{\ell,n} = 0$ if $\ell > n$; and for $k \in \mathbb{N}^*$:

$$c_k(\mathfrak{f}_n) = \sup_{\ell\geq0} \parallel f_{\ell,n} \parallel_{L^k(\mu)} \quad \text{and} \quad q_k(\mathfrak{f}_n) = \sup_{\ell\geq0} \parallel \mathcal{Q}(f_{\ell,n}^k) \parallel_\infty^{1/k} . \tag{29}$$

In particular, we have $\mathfrak{c}_2 = \sup_{n\geq0} c_2(\mathfrak{f}_n)$ and $\mathfrak{q}_2 = \sup_{n\geq0} q_2(\mathfrak{f}_n)$.

For the critical case, $2\alpha^2 = 1$, we shall assume Assumption 4.3 as well as the following.

Assumption 4.5 (*Regularity of the approximation functions in the critical regime*)
Keeping the same notations as in Assumption 4.3, we further assume that:

(v)
$$\lim_{n\to\infty} n \sum_{\ell=0}^{n} 2^{-\ell/2}\, \delta_{\ell,n} = 0. \tag{30}$$

(vi) For all $n \geq \ell \geq 0$:
$$\|\mathcal{Q}(|f_{\ell,n}|)\|_\infty \leq \delta_{\ell,n}. \tag{31}$$

For the supercritical case, $2\alpha^2 > 1$, we shall assume Assumptions 4.3, 4.5 as well as the following.

Assumption 4.6 (*Regularity of the approximation functions in the supercritical regime*)
Keeping the same notations as in Assumption 4.5, we further assume that Assumption 2.3 holds with $2\alpha^2 > 1$ and that:

$$\sup_{0\leq\ell\leq n} (2\alpha^2)^{n-\ell}\delta_{\ell,n}^2 < +\infty \quad \text{and, for all} \ell \in \mathbb{N}, \quad \lim_{n\to\infty} (2\alpha^2)^{n-\ell}\delta_{\ell,n}^2 = 0. \tag{32}$$

Notice that condition (32) implies (30) as well as $\Delta < +\infty$ and $\lim_{n\to\infty} \delta_{\ell,n} = 0$ for all $\ell \in \mathbb{N}$ (see Assumption 4.3 (iii)) when $2\alpha^2 > 1$.

Following [2], for a finite set $\mathbb{A} \subset \mathbb{T}$ and a function $f \in \mathcal{B}(S)$, we set:

$$M_{\mathbb{A}}(f) = \sum_{i\in\mathbb{A}} f(X_i). \tag{33}$$

We shall be interested in the cases $\mathbb{A} = \mathbb{G}_n$ (the nth generation) and $\mathbb{A} = \mathbb{T}_n$ (the tree up to the nth generation). We shall assume that μ is an invariant probability measure of $\mathcal{Q}$. In view of Remark 2.2, one is interested in the fluctuations of $|\mathbb{G}_n|^{-1} M_{\mathbb{G}_n}(f)$ around $\langle \mu, f\rangle$. So, we will use frequently the following notation:

$$\tilde{f} = f - \langle \mu, f\rangle \quad \text{for} f \in L^1(\mu). \tag{34}$$

Let $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$ be a sequence of elements of $L^1(\mu)$. We set for $n \in \mathbb{N}$:

$$N_{n,\emptyset}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{n} M_{\mathbb{G}_{n-\ell}}(\tilde{f}_\ell). \tag{35}$$

The notation $N_{n,\emptyset}$ means that we consider the average from the root $\emptyset$ up to the nth generation.

Remark 4.7 The following two simple cases are frequently used in the literature. Let $f \in L^1(\mu)$ and consider the sequence $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$. If $f_0 = f$ and $f_\ell = 0$ for $\ell \in \mathbb{N}^*$, then we get:

$$N_{n,\emptyset}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_n}(\tilde{f}).$$

If $f_\ell = f$ for $\ell \in \mathbb{N}$, then we get, as $|\mathbb{T}_n| = 2^{n+1} - 1$ and $|\mathbb{G}_n| = 2^n$:

$$N_{n,\emptyset}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} M_{\mathbb{T}_n}(\tilde{f}) = \sqrt{2 - 2^{-n}}\ |\mathbb{T}_n|^{-1/2} M_{\mathbb{T}_n}(\tilde{f}).$$

Thus, we will easily deduce the fluctuations of $M_{\mathbb{T}_n}(f)$ and $M_{\mathbb{G}_n}(f)$ from the asymptotics of $N_{n,\emptyset}(\mathfrak{f})$.

The main result of this section is motivated by the decomposition given in (5). It will allow us to treat the variance term of kernel estimators defined in (4). The proof is given in Sect. 5 for the subcritical case ($\alpha \in (0, 1/\sqrt{2})$), with α the rate defined in Assumption 2.3; it follows closely the approach given in [2]. For the critical case ($\alpha = 1/\sqrt{2}$) and the supercritical case ($\alpha \in (1/\sqrt{2}, 1)$, the proof is an adaptation of the subcritical case and it is therefore omitted; the interested reader can find the details in [1]. Recall $N_{n,\emptyset}(\mathfrak{f})$ defined in (35).

Theorem 4.8 *Let X be a BMC with kernel $\mathcal{P}$ and initial distribution ν, such that Assumption 2.3 (on the geometric ergodic rate $\alpha \in (0, 1)$), Assumption 4.2 (on the regularity of $\mathcal{P}$ and of ν) and Assumption 4.3 (on the approximation functions $(f_{\ell,n}, n \geq \ell \geq 0)$) are in force.*

Furthermore, if $\alpha = 1/\sqrt{2}$ then assume that Assumption 4.5 holds; and if $\alpha > 1/\sqrt{2}$ then assume that Assumption 4.5 and Assumption 4.6 hold. Then, we have the following convergence in distribution:

$$N_{n,\emptyset}(\mathfrak{f}_n) \xrightarrow[n\to\infty]{(d)} G,$$

where $\mathfrak{f}_n = (f_{\ell,n},\ \ell \in \mathbb{N})$ and the convention that $f_{\ell,n} = 0$ for $\ell > n$, and with G a centered Gaussian random variable with finite variance σ^2 defined in (28).

Remark 4.9 Assume $\sigma_\ell^2 = \lim_{n\to\infty} \| f_{\ell,n} \|^2_{L^2(\mu)}$ exists for all $\ell \in \mathbb{N}$; so that σ^2 defined in (28) is also equal to $\sum_{\ell\in\mathbb{N}} 2^{-\ell}\sigma_\ell^2$. According to the additive form of the variance σ^2, we deduce that for fixed $k \in \mathbb{N}$, the random variables $\left(|\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_{n-\ell}}(\tilde{f}_{\ell,n}),\ \ell \in \{0, \dots, k\}\right)$ converges in distribution, as n goes to infinity toward $(G_\ell,\ \ell \in \{0, \dots, k\})$ which are independent real-valued Gaussian centered random variables with variance $\mathrm{Var}(G_\ell) = 2^{-\ell}\sigma_\ell^2$.

4.2 Proof of Theorem 3.5

We suppose that $S = \mathbb{R}^d$, with $d \geq 1$, and that Assumptions 3.1, 3.2 hold. Let K be a kernel function satisfying Assumption 3.3 (i) and bandwidths $(h_n, n \in \mathbb{N})$ satisfying Assumption 3.3 (ii). For $x \in \mathbb{R}^d$, we define the sequences of functions $(f_\ell^x,\ \ell \in \mathbb{N})$ given by:

$$f_\ell^x(y) = K_{h_\ell}(x - y) = h_\ell^{-d/2} K\left(\frac{x - y}{h_\ell}\right) \quad \text{for} y \in \mathbb{R}^d.$$

We consider the sequences of functions:

$$\mathfrak{f}^{\text{shift}} = (f_{\ell,n}^{\text{shift}},\ n \geq \ell \geq 0), \quad \mathfrak{f}^{\text{id}} = (f_{\ell,n}^{\text{id}},\ n \geq \ell \geq 0) \quad \text{and} \quad \mathfrak{f}^{0} = (f_{\ell,n}^{0},\ n \geq \ell \geq 0), \tag{36}$$

defined by:

$$f_{\ell,n}^{\text{shift}} = f_{n-\ell}^{x}, \quad f_{\ell,n}^{\text{id}} = f_{n}^{x} \quad \text{and} \quad f_{\ell,n}^{0} = f_{n}^{x}\mathbf{1}_{\{\ell=0\}}. \tag{37}$$

Let x be in the set of continuity of μ. Thanks to Bochner's recall results in Lemma 6.1 of Appendix, we have:

$$\lim_{\ell\to\infty} \| f_{\ell}^{x} \|_{L^2(\mu)}^2 = \lim_{\ell\to\infty} \langle \mu, (f_{\ell}^{x})^2 \rangle = \mu(x) \| K \|_2^2 . \tag{38}$$

The proof of the following lemma is not difficult and left to the reader (see, for example, [2] for more details).

Lemma 4.10 *Under the assumption of Theorem 3.5 and when considering any of the sequence $\mathfrak{f}^{shift}$, $\mathfrak{f}^{id}$ or $\mathfrak{f}^{0}$, the Assumptions 4.2, 4.3, 4.5 and 4.6 hold with σ^2 defined by (28), respectively, given by:*

$$(\sigma^{shift})^2 = 2\mu(x) \| K \|_2^2, \quad (\sigma^{id})^2 = 2\mu(x) \| K \|_2^2 \quad \textit{and} \quad (\sigma^{0})^2 = \mu(x) \| K \|_2^2 . \tag{39}$$

The subcritical case and $\mathbb{A}_n = \mathbb{T}_n$. We have the following decomposition:

$$\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x) = \frac{\sqrt{|\mathbb{G}_n|}}{|\mathbb{T}_n| h_n^{d/2}} N_{n,\emptyset}(\mathfrak{f}_n) + B_{h_n}(x), \tag{40}$$

where $\mathfrak{f}_n = (f_{\ell,n},\ \ell \in \mathbb{N})$ with the functions $f_{\ell,n} = f_{\ell,n}^{\text{id}}$ defined in (37) for $n \geq \ell \geq 0$ and $f_{\ell,n} = 0$ otherwise; $N_{n,\emptyset}$ is defined in (35) with $\mathfrak{f}$ replaced by $\mathfrak{f}_n$; and the bias term:

$$B_{h_n}(x) = \frac{1}{|\mathbb{T}_n| h_n^{d/2}} \sum_{\ell=0}^{n} 2^{n-\ell} \langle \mu, f_{\ell,n} \rangle - \mu(x) = \langle \mu, h_n^{-d} K(h_n^{-1}(x - \cdot)) \rangle - \mu(x).$$

Since $\lim_{n\to\infty} |\mathbb{G}_n| h_n^d = \infty$ as $\gamma < 1$, we get that $\lim_{n\to\infty} |\mathbb{G}_n|^{1/2} / |\mathbb{T}_n| h_n^{d/2} = 0$. Thus, as a direct consequence of Theorem 4.8, we get the following convergence in probability:

$$\lim_{n\to\infty} \frac{\sqrt{|\mathbb{G}_n|}}{|\mathbb{T}_n| h_n^{d/2}} N_{n,\emptyset}(\mathfrak{f}_n) = 0.$$

Next, it follows from Lemma 6.1 that $\lim_{n\to\infty} B_{h_n}(x) = 0$. This gives Equation (12) on the consistency of the estimator.

Using the value of $\sigma = \sigma^{\text{id}}$ in (39), thanks to Theorem 4.8 and the decomposition (40), we see that to get the asymptotic normality of the estimator (13) it suffices to prove that:

$$\lim_{n\to\infty} |\mathbb{T}_n|^{1/2} h_n^{d/2} B_{h_n}(x) = 0. \tag{41}$$

Using that:

$$\mu(x - h_n y) - \mu(x) = \sum_{j=1}^{d} (\mu(x_1 - h_n y_1, \ldots, x_j - h_n y_j, x_{j+1}, \ldots, x_d)$$
$$-\mu(x_1 - h_n y_1, \ldots, x_{j-1} - h_n y_{j-1}, x_j, x_{j+1}, \ldots, x_d)),$$

the Taylor expansion and Assumption 3.4, we get that, for some finite constant $C > 0$:

$$\begin{aligned}
|\mathbb{T}_n|^{1/2} h_n^{d/2} B_{h_n}(x) &= \sqrt{|\mathbb{T}_n| h_n^d} \left| \int_{\mathbb{R}^d} h_n^{-d} K(h_n^{-1}(x-y))\mu(y)\mathrm{d}y - \mu(x) \right| \\
&= \sqrt{|\mathbb{T}_n| h_n^d} \left| \int_{\mathbb{R}^d} K(y)(\mu(x - h_n y) - \mu(x))\, \mathrm{d}y \right| \\
&\le C\sqrt{|\mathbb{T}_n| h_n^d} \sum_{j=1}^{d} \int_{\mathbb{R}^d} K(y) \frac{(h_n |y_j|)^s}{\lfloor s \rfloor !} \mathrm{d}y \\
&\le C\sqrt{|\mathbb{T}_n| h_n^{2s+d}}.
\end{aligned}$$

Then Eq. (41) follows, since $\lim_{n\to\infty} |\mathbb{T}_n| s_n^{2s+d} = 0$. This ends the proof for $\mathbb{A}_n = \mathbb{T}_n$.
The subcritical case and $\mathbb{A}_n = \mathbb{G}_n$. The proof is similar, using instead the functions $f_{\ell,n} = f_{\ell,n}^0$ defined in (37).
The critical and supercritical cases. The proof follows the same lines, using Theorem 4.8 in the critical and supercritical cases and the decomposition (40).

5 Proof of Theorem 4.8 in the Subcritical Case ($2\alpha^2 < 1$)

Recall the definition of $M_{\mathbb{A}}$ given in (33) and of $\tilde{f} = f - \langle \mu, f \rangle$ in (34). We follow the approach of [2]. In order to study the asymptotics of $M_{\mathbb{G}_{n-\ell}}(\tilde{f})$ as n goes to infinity and ℓ is fixed, it is convenient to consider the contribution of the descendants of the individual $i \in \mathbb{T}_{n-\ell}$ for $n \ge \ell \ge 0$:

$$N_{n,i}^{\ell}(f) = |\mathbb{G}_n|^{-1/2} M_{i\mathbb{G}_{n-|i|-\ell}}(\tilde{f}), \tag{42}$$

where $i\mathbb{G}_{n-|i|-\ell} = \{ij,\ j \in \mathbb{G}_{n-|i|-\ell}\} \subset \mathbb{G}_{n-\ell}$. For all $k \in \mathbb{N}$ such that $n \ge k + \ell$, we have:

$$M_{\mathbb{G}_{n-\ell}}(\tilde{f}) = \sqrt{|\mathbb{G}_n|} \sum_{i \in \mathbb{G}_k} N_{n,i}^{\ell}(f) = \sqrt{|\mathbb{G}_n|}\, N_{n,\emptyset}^{\ell}(f).$$

Let $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$ be a sequence of elements of $L^1(\mu)$. We set for $n \in \mathbb{N}$ and $i \in \mathbb{T}_n$:

$$N_{n,i}(\mathfrak{f}) := \sum_{\ell=0}^{n-|i|} N_{n,i}^{\ell}(f_\ell) := |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{n-|i|} M_{i\mathbb{G}_{n-|i|-\ell}}(\tilde{f}_\ell). \tag{43}$$

We deduce that $\sum_{i \in \mathbb{G}_k} N_{n,i}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{n-k} M_{\mathbb{G}_{n-\ell}}(\tilde{f}_\ell)$. For $k = 0$, we recover Eq. (35).

We consider the notations of Theorem 4.8. Recall that $\mathfrak{f}_n = (f_{\ell,n}, \ \ell \in \mathbb{N})$ with the convention that $f_{\ell,n} = 0$ for $\ell > n$. In the following proofs, we will denote by C any unimportant finite constant which may vary from line to line (in particular C does not depend on n nor on $\mathfrak{f}_n$).

Remark 5.1 Recall k_0 given in Assumption 4.2 (iii). Recall that from Assumption 4.3 (ii), the sequence $\mathfrak{f}_n$ is bounded in $L^2(\mu)$. We have

$$N_{n,\emptyset}(\mathfrak{f}_n) = N_{n,\emptyset}^{[k_0]}(\mathfrak{f}_n) \ + \ |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{k_0-1} M_{\mathbb{G}_\ell}(\tilde{f}_{n-\ell,n}), \tag{44}$$

where we set:

$$N_{n,\emptyset}^{[k_0]}(\mathfrak{f}_n) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{n-k_0} M_{\mathbb{G}_{n-\ell}}(\tilde{f}_{\ell,n}). \tag{45}$$

Using the Cauchy–Schwarz inequality, we get

$$|\mathbb{G}_n|^{-1/2} \Big| \sum_{\ell=0}^{k_0-1} M_{\mathbb{G}_\ell}(\tilde{f}_{n-\ell,n}) \Big| \leq C c_2(\mathfrak{f}) |\mathbb{G}_n|^{-1/2} \ + \ |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{k_0-1} M_{\mathbb{G}_\ell}(|f_{n-\ell,n}|). \tag{46}$$

Since the sequence $\mathfrak{f}_n$ is bounded in $L^2(\mu)$ and since k_0 is finite, we have, for all $\ell \in \{0, \dots, k_0 - 1\}$, $\lim_{n\to\infty} |\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_\ell}(|f_{n-\ell,n}|) = 0$ a.s. and then that (used (46))

$$\lim_{n\to\infty} |\mathbb{G}_n|^{-1/2} \Big| \sum_{\ell=0}^{k_0-1} M_{\mathbb{G}_\ell}(\tilde{f}_{n-\ell}) \Big| = 0 \quad \text{a.s.}$$

Therefore, from (44), the study of $N_{n,\emptyset}(\mathfrak{f}_n)$ is reduced to that of $N_{n,\emptyset}^{[k_0]}(\mathfrak{f}_n)$ defined in (45).

Let $(p_n, n \in \mathbb{N})$ be a non-decreasing sequence of elements of $\mathbb{N}^*$ such that, for all $\lambda > 0$:

$$p_n < n, \quad \lim_{n\to\infty} p_n/n = 1 \quad \text{and} \quad \lim_{n\to\infty} n - p_n - \lambda \log(n) = +\infty. \tag{47}$$

In the rest of the paper, when there is no ambiguity, we write p for p_n.

Let $i, j \in \mathbb{T}$. We write $i \preccurlyeq j$ if $j \in i\mathbb{T}$. We denote by $i \wedge j$ the most recent common ancestor of i and j, which is defined as the only $u \in \mathbb{T}$ such that if $v \in \mathbb{T}$ and $v \preccurlyeq i$, $v \preccurlyeq j$ then $v \preccurlyeq u$. We also define the lexicographic order $i \leq j$ if either $i \preccurlyeq j$ or $v0 \preccurlyeq i$ and $v1 \preccurlyeq j$ for $v = i \wedge j$. Let $X = (X_i, i \in \mathbb{T})$ be a BMC with kernel $\mathcal{P}$ and initial measure ν. For $i \in \mathbb{T}$, we define the σ-field:

$$\mathcal{F}_i = \{X_u; u \in \mathbb{T} \text{ such that } \quad u \leq i\}.$$

By construction, the σ-fields $(\mathcal{F}_i;\ i \in \mathbb{T})$ are nested as $\mathcal{F}_i \subset \mathcal{F}_j$ for $i \leq j$.

Recalling $N_{n,i}(\mathfrak{f})$ defined in (43), we define for $n \in \mathbb{N}$, $i \in \mathbb{G}_{n-p_n}$ and $\mathfrak{f}_n$ the martingale increments:

$$\Delta_{n,i}(\mathfrak{f}_n) = N_{n,i}(\mathfrak{f}_n) - \mathbb{E}\left[N_{n,i}(\mathfrak{f}_n)|\, \mathcal{F}_i\right] \quad \text{and} \quad \Delta_n(\mathfrak{f}_n) = \sum_{i \in \mathbb{G}_{n-p_n}} \Delta_{n,i}(\mathfrak{f}_n). \quad (48)$$

Thanks to (43), we have:

$$\sum_{i \in \mathbb{G}_{n-p_n}} N_{n,i}(\mathfrak{f}_n) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{p_n} M_{\mathbb{G}_{n-\ell}}(\tilde{f}_{\ell,n}) = |\mathbb{G}_n|^{-1/2} \sum_{k=n-p_n}^{n} M_{\mathbb{G}_k}(\tilde{f}_{n-k,n}).$$

Using the branching Markov property, and (43), we get for $i \in \mathbb{G}_{n-p_n}$:

$$\mathbb{E}\left[N_{n,i}(\mathfrak{f}_n)|\, \mathcal{F}_i\right] = \mathbb{E}\left[N_{n,i}(\mathfrak{f}_n)|\, X_i\right] = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{p_n} \mathbb{E}_{X_i}\left[M_{\mathbb{G}_{p_n-\ell}}(\tilde{f}_{\ell,n})\right].$$

Assume that n is large enough so that $n - p_n - 1 \geq k_0$. We have:

$$N_{n,\emptyset}^{[k_0]}(\mathfrak{f}) = \Delta_n(\mathfrak{f}) + R_0^{k_0}(n) + R_1(n),$$

where $N_{n,\emptyset}^{[k_0]}(\mathfrak{f})$ is defined in (45), $\Delta_n(\mathfrak{f})$ is defined in (48) and:

$$R_0^{k_0}(n) = |\mathbb{G}_n|^{-1/2} \sum_{k=k_0}^{n-p_n-1} M_{\mathbb{G}_k}(\tilde{f}_{n-k}) \quad \text{and} \quad R_1(n) = \sum_{i \in \mathbb{G}_{n-p_n}} \mathbb{E}\left[N_{n,i}(\mathfrak{f}_n)|\, \mathcal{F}_i\right].$$

We have the following result:

Lemma 5.2 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we have that*

$$\lim_{n\to\infty} \mathbb{E}\left[\left(N_{n,\emptyset}^{[k_0]}(\mathfrak{f}_n) - \Delta_n(\mathfrak{f}_n)\right)^2\right] = 0.$$

Proof We deduce from Remark 5.5 in [2] that $\mathbb{E}\left[\left(N_{n,\emptyset}^{[k_0]}(\mathfrak{f}_n) - \Delta_n(\mathfrak{f}_n)\right)^2\right] \leq a_{0,n}\, \mathfrak{c}_2^2$ for a sequence $(a_{0,n}, n \in \mathbb{N})$ which converges to 0 and does not depend on the sequences $\mathfrak{f}_n$. □

We consider the bracket of the martingale $\Delta_n(\mathfrak{f}_n)$ given by $V(n) = \sum_{i\in\mathbb{G}_{n-p_n}} \mathbb{E}\left[\Delta_{n,i}(\mathfrak{f}_n)^2|\mathcal{F}_i\right]$. Using (43) and (48), we write:

$$V(n) = |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p_n}} \mathbb{E}_{X_i}\left[\left(\sum_{\ell=0}^{p_n} M_{\mathbb{G}_{p_n-\ell}}(\tilde{f}_{\ell,n})\right)^2\right] - R_2(n) = V_1(n) + 2V_2(n) - R_2(n), \tag{49}$$

with:

$$\begin{aligned} V_1(n) &= |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p_n}} \sum_{\ell=0}^{p_n} \mathbb{E}_{X_i}\left[M_{\mathbb{G}_{p_n-\ell}}(\tilde{f}_{\ell,n})^2\right], \\ V_2(n) &= |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p_n}} \sum_{0\leq\ell<k\leq p_n} \mathbb{E}_{X_i}\left[M_{\mathbb{G}_{p_n-\ell}}(\tilde{f}_{\ell,n}) M_{\mathbb{G}_{p_n-k}}(\tilde{f}_{k,n})\right], \\ R_2(n) &= \sum_{i\in\mathbb{G}_{n-p_n}} \mathbb{E}\left[N_{n,i}(\mathfrak{f}_n)|X_i\right]^2, \end{aligned}$$

where $N_{n,i}(\mathfrak{f})$ is defined in (43).

Lemma 5.3 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we have that $R_2(n)$ converges in probability toward 0.*

Proof We deduce from Remark 5.7 in [2] that $\mathbb{E}[|R_2(n)|] \leq C\mathfrak{c}_2^2 a_n$ for a sequence $(a_n, n \in \mathbb{N})$ which converges to 0 and does not depend on the sequence $\mathfrak{f}_n$. □

Lemma 5.4 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we have that $V_2(n)$ converges in probability toward 0.*

Proof First, we have the following preliminary results. Let $f \in L^2(\mu)$ and recall that $\tilde{f} = f - \langle \mu, f\rangle$. We deduce from $\langle \mu, f\rangle = \langle \mu, \mathcal{Q}f\rangle \leq \| \mathcal{Q}f \|_\infty \leq \| \mathcal{Q}(f^2) \|_\infty^{1/2}$ that:

$$\| \mathcal{Q}\tilde{f} \|_\infty \leq 2 \| \mathcal{Q}(f^2) \|_\infty^{1/2} \quad \text{and} \quad \| \mathcal{Q}(\tilde{f}^2) \|_\infty \leq 4 \| \mathcal{Q}(f^2) \|_\infty. \tag{50}$$

Note that thanks to Assumption 4.3 we have, for all $k, \ell, r \in \mathbb{N}$, and $j > 0$:

$$\lim_{n\to\infty} |\langle \mu, \tilde{f}_{k,n} \mathcal{Q}^j \tilde{f}_{\ell,n}\rangle| = 0 \quad \text{and} \quad \lim_{n\to\infty} |\langle \mu, \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^j \tilde{f}_{\ell,n}\right)\rangle| = 0. \tag{51}$$

Indeed, we have thanks to Assumption 4.3 (iii):

$$|\langle \mu, \tilde{f}_{k,n} \mathcal{Q}^j \tilde{f}_{\ell,n} \rangle| \leq \| \mathcal{Q} \tilde{f}_{\ell,n} \|_\infty \langle \mu, |\tilde{f}_{k,n}| \rangle \leq 4 \| \mathcal{Q} f_{\ell,n}^2 \|_\infty^{1/2} \langle \mu, |f_{k,n}| \rangle \leq 4 \mathfrak{q}_2 \, \delta_{k,n}.$$

We also have thanks to Assumption 4.3 (iii), for $g = \mathcal{Q}^{j-1} |\tilde{f}_{\ell,n}|$ and $r = 0$:

$$\begin{aligned} |\langle \mu, \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^j \tilde{f}_{\ell,n}\right)\rangle| &\leq \langle \mu, \mathcal{P}\left(|\tilde{f}_{k,n}| \otimes_{\text{sym}} \mathcal{Q} g\right)\rangle \\ &\leq \langle \mu, \mathcal{P}\left(\mathbf{1} \otimes_{\text{sym}} \mathcal{Q} g\right)\rangle \langle \mu, |f_{k,n}| \rangle \\ &\quad + \| \mathcal{P}(|f_{k,n}| \otimes_{\text{sym}} \mathcal{Q} g) \|_{L^2(\mu)} \\ &\leq 2 \| g \|_{L^2(\mu)} \, \delta_{k,n} \\ &\leq 2 \mathfrak{c}_2 \, \delta_{k,n}, \end{aligned}$$

and for $r \geq 1$ using (50) and that $\langle \mu, \mathcal{P}(\mathbf{1} \otimes_{\text{sym}} h) \rangle = \langle \mu, h \rangle$:

$$|\langle \mu, \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^j \tilde{f}_{\ell,n}\right)\rangle| \leq \langle \mu, \mathcal{P}\left(\mathbf{1} \otimes_{\text{sym}} \mathcal{Q} g\right)\rangle \; \| \mathcal{Q}^r \tilde{f}_{k,n} \|_\infty \leq 2 \, \mathfrak{q}_2 \, \delta_{\ell,n}.$$

Then use that for all $k \in \mathbb{N}$ fixed, we have $\lim_{n \to \infty} \delta_{k,n} = 0$ to conclude that (51) holds.

Using (76), we get:

$$V_2(n) = V_5(n) + V_6(n), \tag{52}$$

with

$$\begin{aligned} V_5(n) &= |\mathbb{G}_n|^{-1} \sum_{i \in \mathbb{G}_{n-p}} \sum_{0 \leq \ell < k \leq p} 2^{p-\ell} \mathcal{Q}^{p-k} \left(\tilde{f}_{k,n} \mathcal{Q}^{k-\ell} \tilde{f}_{\ell,n}\right)(X_i), \\ V_6(n) &= |\mathbb{G}_n|^{-1} \sum_{i \in \mathbb{G}_{n-p}} \sum_{0 \leq \ell < k < p} \sum_{r=0}^{p-k-1} 2^{p-\ell+r} \, \mathcal{Q}^{p-1-(r+k)} \\ &\quad \left(\mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}\right)\right)(X_i). \end{aligned}$$

First, we consider the term $V_6(n)$. We have:

$$V_6(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{6,n}),$$

with

$$\begin{aligned} H_{6,n} &= \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} h_{k,\ell,r}^{(n)} \, \mathbf{1}_{\{r+k<p\}} \quad \text{and} \quad h_{k,\ell,r}^{(n)} = 2^{r-\ell} \, \mathcal{Q}^{p-1-(r+k)} \\ &\quad \left(\mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}\right)\right). \end{aligned}$$

Define

$$H_6^{[n]}(\mathfrak{f}_n) = \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} h_{k,\ell,r} \, \mathbf{1}_{\{r+k<p\}}, \tag{53}$$

with $h_{k,\ell,r} = 2^{r-\ell} \langle \mu, \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}\right)\rangle = \langle \mu, h_{k,\ell,r}^{(n)} \rangle$.

We set $A_{6,n}(\mathfrak{f}_n) = H_{6,n} - H_6^{[n]}(\mathfrak{f}_n) = \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} (h_{k,\ell,r}^{(n)} - h_{k,\ell,r}) \, \mathbf{1}_{\{r+k<p\}}$, so that from the definition of $V_6(n)$, we get that:

$$V_6(n) - H_6^{[n]}(\mathfrak{f}_n) = |\mathbb{G}_{n-p}|^{-1} \, M_{\mathbb{G}_{n-p}}(A_{6,n}(\mathfrak{f}_n)).$$

We now study the second moment of $|\mathbb{G}_{n-p}|^{-1} \, M_{\mathbb{G}_{n-p}}(A_{6,n}(\mathfrak{f}_n))$. Using (77), we get for $n - p \geq k_0$:

$$|\mathbb{G}_{n-p}|^{-2} \, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{6,n}(\mathfrak{f}_n))^2\right] \leq C \, |\mathbb{G}_{n-p}|^{-1} \sum_{j=0}^{n-p} 2^j \, \| \mathcal{Q}^j (A_{6,n}(\mathfrak{f}_n)) \|_{L^2(\mu)}^2 .$$

We deduce that

$$\begin{aligned}
\| \mathcal{Q}^j (A_{6,n}(\mathfrak{f}_n)) \|_{L^2(\mu)} &\leq \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} \| \mathcal{Q}^j h_{k,\ell,r}^{(n)} - h_{k,\ell,r} \|_{L^2(\mu)} \mathbf{1}_{\{r+k<p\}} \\
&\leq C \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} 2^{r-\ell} \, \alpha^{p-1-(r+k)+j} \\
&\quad \| \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}\right) \|_{L^2(\mu)} \mathbf{1}_{\{r+k<p\}} \\
&\leq C \mathfrak{c}_2^2 \, \alpha^j \sum_{\substack{0 \leq \ell < k \\ r \geq k_1}} 2^{r-\ell} \, \alpha^{p-(r+k)} \alpha^{k-\ell+2r} \, \mathbf{1}_{\{r+k<p\}} \qquad (54) \\
&\quad + C\alpha^j \sum_{\substack{0 \leq \ell < k \\ 0 \leq r \leq k_1 - 1}} 2^{-\ell} \, \alpha^{p-k} \qquad (55) \\
&\quad \| \mathcal{P}\left(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}\right) \|_{L^2(\mu)} \mathbf{1}_{\{r+k<p\}}, \qquad (56)
\end{aligned}$$

where we used the triangular inequality for the first inequality; (3) for the second; (25) for $r \geq k_1$ and (3) again for the third. The term (55) can be bounded from above using (50) and

$$\begin{aligned}
\| \mathcal{P}(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_{\ell,n}) \|_{L^2(\mu)} &\leq \| \mathcal{Q} \tilde{f}_{\ell,n} \|_\infty \, \| \mathcal{P}(\mathcal{Q}^r \tilde{f}_{k,n} \otimes_{\text{sym}} \mathbf{1}) \|_{L^2(\mu)} \\
&\leq 2\mathfrak{q}_2 \, \mathfrak{c}_2 \quad \text{as} k > \ell,
\end{aligned}$$

and thus (54) and (55) imply that

$$\begin{aligned}\| \mathcal{Q}^j (A_{6,n}(\mathfrak{f}_n)) \|_{L^2(\mu)} &\leq C \mathfrak{c}_2 (\mathfrak{c}_2 + \mathfrak{q}_2)\, \alpha^j \sum_{\substack{0\leq \ell < k\\ r\geq 0}} 2^{r-\ell}\, \alpha^{p-(r+k)} \alpha^{k-\ell+2r}\, \mathbf{1}_{\{r+k<p\}} \\ &\leq C \mathfrak{c}_2 (\mathfrak{c}_2 + \mathfrak{q}_2)\, \alpha^j, \end{aligned} \tag{57}$$

where we used that $\sum_{0\leq \ell<k,\, r\geq 0} 2^{r-\ell} \alpha^{k-\ell+2r}$ is finite for the last inequality. As $\sum_{j=0}^{\infty} (2\alpha^2)^j$ is finite, we deduce that:

$$\begin{aligned}\mathbb{E}\left[\left(V_6(n) - H_6^{[n]}(\mathfrak{f}_n) \right)^2 \right] &= |\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{6,n}(\mathfrak{f}_n))^2 \right] \\ &\leq C \mathfrak{c}_2^2 (\mathfrak{c}_2 + \mathfrak{q}_2)^2\, 2^{-(n-p)}. \end{aligned} \tag{58}$$

We now consider the term $V_5(n)$ defined just after (52):

$$V_5(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{5,n}),$$

with

$$H_{5,n} = \sum_{0\leq \ell<k} h_{k,\ell}^{(n)}\, \mathbf{1}_{\{k\leq p\}} \quad \text{and} \quad h_{k,\ell}^{(n)} = 2^{-\ell}\, \mathcal{Q}^{p-k} \left(\tilde{f}_{k,n} \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n} \right).$$

We consider the constant

$$H_5^{[n]}(\mathfrak{f}_n) = \sum_{0\leq \ell<k} h_{k,\ell} \mathbf{1}_{\{k\leq p\}} \quad \text{with} \quad h_{k,\ell} = 2^{-\ell} \langle \mu,\, \tilde{f}_{k,n} \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n} \rangle. \tag{59}$$

We set $A_{5,n}(\mathfrak{f}_n) = H_{5,n} - H_5^{[n]}(\mathfrak{f}_n) = \sum_{0\leq \ell<k} (h_{k,\ell}^{(n)} - h_{k,\ell})\, \mathbf{1}_{\{k\leq p\}}$, so that from the definition of $V_5(n)$, we get that:

$$V_5(n) - H_5^{[n]}(\mathfrak{f}_n) = |\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(A_{5,n}(\mathfrak{f}_n)).$$

We now study the second moment of $|\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(A_{5,n}(\mathfrak{f}_n))$. Using (77), we get for $n-p\geq k_0$:

$$|\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{5,n}(\mathfrak{f}_n))^2 \right] \leq C\, |\mathbb{G}_{n-p}|^{-1} \sum_{j=0}^{n-p} 2^j\, \| \mathcal{Q}^j (A_{5,n}(\mathfrak{f}_n)) \|_{L^2(\mu)}^2.$$

We also have that:

$$\| \mathcal{Q}^j (A_{5,n}(\mathfrak{f}_n)) \|_{L^2(\mu)} \leq \sum_{0\leq \ell<k} \| \mathcal{Q}^j h_{k,\ell}^{(n)} - h_{k,\ell} \|_{L^2(\mu)}\, \mathbf{1}_{\{k\leq p\}}$$

$$\leq C \sum_{0\leq \ell<k} 2^{-\ell}\, \alpha^{p-k+j} \,\|\, \tilde{f}_{k,n}\, \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n} \,\|_{L^2(\mu)}\, \mathbf{1}_{\{k\leq p\}}, \tag{60}$$

where we used the triangular inequality for the first inequality and (3) for the last. The term (60) can be bounded from above using $\|\, \tilde{f}_{k,n}\, \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n} \,\|_{L^2(\mu)} \leq \|\, \tilde{f}_{k,n} \,\|_{L^2(\mu)}\, \|\, \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n} \,\|_\infty \leq \mathfrak{c}_2\, \mathfrak{q}_2$ as $k > \ell$. This implies that

$$\|\, \mathcal{Q}^j\, (A_{5,n}(\mathfrak{f}_n)) \,\|_{L^2(\mu)} \leq C\, \mathfrak{c}_2 \mathfrak{q}_2\, \alpha^j.$$

As $\sum_{j=0}^\infty (2\alpha^2)^j$ is finite, we deduce that:

$$\mathbb{E}\left[\left(V_5(n) - H_5^{[n]}(\mathfrak{f}_n)\right)^2\right] = |\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{5,n}(\mathfrak{f}_n))^2\right] \leq C\, \mathfrak{c}_2 \mathfrak{q}_2\, 2^{-(n-p)}. \tag{61}$$

We deduce from (58) and (61), as $V_2(n) = V_5(n) + V_6(n)$ (see (52)), that:

$$\mathbb{E}\left[\left(V_2(n) - H_2^{[n]}(\mathfrak{f}_n)\right)^2\right] \leq C\, \left(\mathfrak{c}_2^4 + \mathfrak{c}_2^2\, \mathfrak{q}_2^2\right)\, 2^{-(n-p)}, \quad \text{with} \quad H_2^{[n]}(\mathfrak{f}_n) = H_6^{[n]}(\mathfrak{f}_n) + H_5^{[n]}(\mathfrak{f}_n). \tag{62}$$

Since according to (ii) in Assumption 4.3 $\mathfrak{c}_2$ and $\mathfrak{q}_2$ are finite, we deduce that $\lim_{n\to\infty} V_2(n) - H_2^{[n]}(\mathfrak{f}_n) = 0$ in probability.

We now check that $\lim_{n\to\infty} H_2^{[n]}(\mathfrak{f}_n) = 0$. Using (53) and (59), we get that:

$$|H_2^{[n]}(\mathfrak{f}_n)| \;\leq \sum_{k>\ell\geq 0} 2^{-\ell} |\langle \mu,\, \tilde{f}_{k,n}\, \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n}\rangle| + \sum_{\substack{k>\ell\geq 0\\ r\geq 0}} 2^{r-\ell} |\langle \mu, \mathcal{P}\left(\mathcal{Q}^r\, \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r}\, \tilde{f}_{\ell,n}\right)\rangle|.$$

Recall the definition of Δ in Assumption 4.3 (iii). Thanks to (3) and (24) we have:

$$|\langle \mu,\, \tilde{f}_{k,n}\, \mathcal{Q}^{k-\ell}\, \tilde{f}_{\ell,n}\rangle| \leq \mathfrak{c}_2^2\, \alpha^{k-\ell},$$
$$|\langle \mu, \mathcal{P}\left(\mathcal{Q}^r\, \tilde{f}_{k,n} \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r}\, \tilde{f}_{\ell,n}\right)\rangle| \leq C\, \mathfrak{c}_2^2\, \alpha^{k-\ell+2r}. \tag{63}$$

Since $\sum_{0\leq\ell<k} 2^{-\ell} \alpha^{k-\ell} + \sum_{\substack{0\leq\ell<k\\ r\geq 0}} 2^{r-\ell}\alpha^{k-\ell+2r}$ is finite, we deduce from (53), (59), (51) and dominated convergence that $\lim_{n\to\infty} H_2^{[n]}(\mathfrak{f}_n) = 0$. This implies that $\lim_{n\to\infty} V_2(n) = 0$ in probability. □

Recall $V_1(n)$ defined after (49). We have the following result.

Lemma 5.5 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we have that $V_1(n)$ converges in probability toward σ^2 defined by (28).*

Proof Using (75), we get:

$$V_1(n) = V_3(n) + V_4(n), \tag{64}$$

with

$$V_3(n) = |\mathbb{G}_n|^{-1} \sum_{i \in \mathbb{G}_{n-p}} \sum_{\ell=0}^{p} 2^{p-\ell} \, \mathcal{Q}^{p-\ell}(\tilde{f}_{\ell,n}^2)(X_i),$$

$$V_4(n) = |\mathbb{G}_n|^{-1} \sum_{i \in \mathbb{G}_{n-p}} \sum_{\ell=0}^{p-1} \sum_{k=0}^{p-\ell-1} 2^{p-\ell+k} \, \mathcal{Q}^{p-1-(\ell+k)} \left(\mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right)\right)(X_i).$$

We first consider the term $V_4(n)$. We have:

$$V_4(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{4,n}),$$

with:

$$H_{4,n} = \sum_{\ell \geq 0,\, k \geq 0} h_{\ell,k}^{(n)} \, \mathbf{1}_{\{\ell+k<p\}} \quad \text{and} \quad h_{\ell,k}^{(n)} = 2^{k-\ell} \, \mathcal{Q}^{p-1-(\ell+k)} \left(\mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right)\right).$$

Define the constant

$$H_4^{[n]}(\mathfrak{f}_n) = \sum_{\ell \geq 0,\, k \geq 0} h_{\ell,k} \, \mathbf{1}_{\{\ell+k<p\}} \quad \text{with} \quad h_{\ell,k} = 2^{k-\ell} \, \langle \mu, \mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right) \rangle. \tag{65}$$

We set $A_{4,n}(\mathfrak{f}_n) = H_{4,n} - H_4^{[n]}(\mathfrak{f}_n) = \sum_{\ell \geq 0,\, k \geq 0} (h_{\ell,k}^{(n)} - h_{\ell,k}) \, \mathbf{1}_{\{\ell+k<p\}}$, so that from the definition of $V_4(n)$, we get that:

$$V_4(n) - H_4^{[n]}(\mathfrak{f}_n) = |\mathbb{G}_{n-p}|^{-1} \, M_{\mathbb{G}_{n-p}}(A_{4,n}(\mathfrak{f}_n)).$$

We now study the second moment of $|\mathbb{G}_{n-p}|^{-1} \, M_{\mathbb{G}_{n-p}}(A_{4,n}(\mathfrak{f}_n))$. Using (77), we get for $n - p \geq k_0$:

$$|\mathbb{G}_{n-p}|^{-2} \, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{4,n}(\mathfrak{f}_n))^2\right] \leq C \, |\mathbb{G}_{n-p}|^{-1} \sum_{j=0}^{n-p} 2^j \, \| \mathcal{Q}^j(A_{4,n}(\mathfrak{f}_n)) \|_{L^2(\mu)}^2 .$$

Using (22) and (50), we obtain that for all $0 \leq k < k_1$, $\| \mathcal{P}(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2) \|_{L^2(\mu)} \leq \| \mathcal{Q} \tilde{f}_{\ell,n}^2 \|_{L^2(\mu)} \leq 4\mathfrak{q}_2^2$. We deduce that:

$$\| \mathcal{Q}^j(A_{4,n}(\mathfrak{f}_n)) \|_{L^2(\mu)} \leq \sum_{\ell \geq 0,\, k \geq 0} \| \mathcal{Q}^j h_{\ell,k}^{(n)} - h_{\ell,k} \|_{L^2(\mu)} \, \mathbf{1}_{\{\ell+k<p\}}$$

$$\le C \sum_{\ell\ge 0,\, k\ge 0} 2^{k-\ell}\, \alpha^{p-1-(\ell+k)+j}\, \| \mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right) \|_{L^2(\mu)} \mathbf{1}_{\{\ell+k<p\}}$$
$$\le C\, \mathfrak{c}_2^2\, \alpha^j \sum_{\ell\ge 0,\, k\ge k_1} 2^{k-\ell}\, \alpha^{p-(\ell+k)} \alpha^{2k}\, \mathbf{1}_{\{\ell+k<p\}}$$
$$+ C\, \alpha^j \sum_{\substack{\ell\ge 0\\ 0\le k<k_1}} 2^{k-\ell}\, \alpha^{p-(\ell+k)}\, \| \mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right) \|_{L^2(\mu)} \mathbf{1}_{\{\ell<p\}}$$
$$\le C\, (\mathfrak{c}_2^2 + 4\mathfrak{q}_2^2)\, \alpha^j,$$

where we used the triangular inequality for the first inequality;(3) for the second; (25) for $k \ge k_1$ and (3) again for the third; (22) and (50) for the last. As $\sum_{j=0}^{\infty} (2\alpha^2)^j$ is finite, we deduce that:

$$\mathbb{E}\left[\left(V_4(n) - H_4^{[n]}(\mathfrak{f}_n)\right)^2\right] = |\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{4,n}(\mathfrak{f}_n))^2\right] \le C\, (\mathfrak{c}_2^2 + \mathfrak{q}_2^2)\, 2^{-(n-p)}. \tag{66}$$

We now consider the term $V_3(n)$ defined just after (64):

$$V_3(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{3,n}),$$

with

$$H_{3,n} = \sum_{\ell\ge 0} h_\ell^{(n)}\, \mathbf{1}_{\{\ell\le p\}} \quad \text{and} \quad h_\ell^{(n)} = 2^{-\ell}\, \mathcal{Q}^{p-\ell}\left(\tilde{f}_{\ell,n}^2\right).$$

We consider the constant

$$H_3^{[n]}(\mathfrak{f}_n) = \sum_{\ell\ge 0} h_\ell\, \mathbf{1}_{\{\ell\le p\}} \quad \text{with} \quad h_\ell = 2^{-\ell} \langle \mu, \tilde{f}_{\ell,n}^2 \rangle = \langle \mu, h_\ell^{(n)} \rangle. \tag{67}$$

We set $A_{3,n}(\mathfrak{f}_n) = H_{3,n} - H_3^{[n]}(\mathfrak{f}_n) = \sum_{\ell\ge 0} (h_\ell^{(n)} - h_\ell)\, \mathbf{1}_{\{\ell\le p\}}$, so that from the definition of $V_3(n)$, we get that:

$$V_3(n) - H_3^{[n]}(\mathfrak{f}_n) = |\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(A_{3,n}(\mathfrak{f}_n)). \tag{68}$$

We now study the second moment of $|\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(A_{3,n}(\mathfrak{f}_n))$. Using (77), we get for $n - p \ge k_0$:

$$|\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{3,n}(\mathfrak{f}_n))^2\right] \le C\, |\mathbb{G}_{n-p}|^{-1} \sum_{j=0}^{n-p} 2^j\, \| \mathcal{Q}^j (A_{3,n}(\mathfrak{f}_n)) \|_{L^2(\mu)}^2. \tag{69}$$

Recall $c_k(\mathfrak{f}_n)$ and $q_k(\mathfrak{f}_n)$ defined in (29). We have that

$$\begin{aligned}\| \mathcal{Q}^j(A_{3,n}(\mathfrak{f}_n)) \|_{L^2(\mu)} &\leq \sum_{\ell\geq 0} \| \mathcal{Q}^j h_\ell^{(n)} - h_\ell \|_{L^2(\mu)} \mathbf{1}_{\{\ell\leq p\}} \\ &\leq C \sum_{\ell\geq 0} 2^{-\ell} \| \mathcal{Q}^{j+p-\ell} \tilde{g} \|_{L^2(\mu)} \mathbf{1}_{\{\ell\leq p\}} \quad \text{with} \quad g = \tilde{f}_{\ell,n}^2 \\ &= 2^{-p} \| \tilde{g} \|_{L^2(\mu)} \mathbf{1}_{\{j=0\}} + \sum_{\ell=0}^{p} 2^{-\ell} \| \mathcal{Q}^{j+p-\ell-1} \mathcal{Q}\tilde{g} \|_{L^2(\mu)} \mathbf{1}_{\{j+p-\ell>0\}} \\ &\leq C\, c_4^2(\mathfrak{f}_n)\, 2^{-p} \mathbf{1}_{\{j=0\}} + C \sum_{\ell\geq 0} 2^{-\ell} \alpha^{j+p-\ell} \| \mathcal{Q}\tilde{g} \|_{L^2(\mu)} \\ &\leq C\, c_4^2(\mathfrak{f}_n)\, 2^{-p} \mathbf{1}_{\{j=0\}} + C\, q_2^2(\mathfrak{f}_n)\, \alpha^j, \end{aligned} \tag{70}$$

where we used the triangular inequality for the first inequality; (3) for the third and (50) for the last inequality. As $\sum_{j=0}^{\infty}(2\alpha^2)^j$ is finite, we deduce that:

$$\begin{aligned}\mathbb{E}\left[\left(V_3(n) - H_3^{[n]}(\mathfrak{f}_n)\right)^2\right] &= |\mathbb{G}_{n-p}|^{-2}\, \mathbb{E}\left[M_{\mathbb{G}_{n-p}}(A_{3,n}(\mathfrak{f}_n))^2\right] \\ &\leq C\, c_4^4(\mathfrak{f}_n)\, 2^{-n} + C\, q_2^4(\mathfrak{f}_n)\, 2^{-(n-p)}. \end{aligned} \tag{71}$$

As $V_1 = V_4 + V_3$, we deduce from (66) and (71) that:

$$\mathbb{E}\left[\left(V_1(n) - H_1^{[n]}(\mathfrak{f}_n)\right)^2\right] \leq C\, \left((c_2^4(\mathfrak{f}_n) + q_2^4(\mathfrak{f}_n))\, 2^{-(n-p)} + c_4^4(\mathfrak{f}_n)\, 2^{-n}\right),$$

with $H_1^{[n]}(\mathfrak{f}_n) = H_3^{[n]}(\mathfrak{f}_n) + H_4^{[n]}(\mathfrak{f}_n)$. Since $c_4^4(\mathfrak{f}_n) \leq c_2^2(\mathfrak{f}_n)\, c_\infty^2(\mathfrak{f}_n) \leq C_\rho\, c_2^2(\mathfrak{f}_n)\, 2^{2n\rho}$ with $\rho \in (0, 1/2)$ and some finite constant C_ρ according to (i) in Assumption 4.3, and since $\lim_{n\to\infty} p/n = 1$ so that $2^{-n(1-2\rho)} \leq 2^{-(n-p)}$ (at least for n large enough), we deduce from (ii) in Assumption 4.3 that:

$$\mathbb{E}\left[\left(V_1(n) - H_1^{[n]}(\mathfrak{f}_n)\right)^2\right] \leq C\, \left(\mathfrak{c}_2^4 + \mathfrak{q}_2^4 + C_\rho \mathfrak{c}_2^2\right)\, 2^{-(n-p)} \tag{72}$$

and thus $\lim_{n\to\infty} V_1(n) - H_1^{[n]}(\mathfrak{f}_n) = 0$ in probability.

We check that $\lim_{n\to\infty} H_1^{[n]}(\mathfrak{f}_n) = \sigma^2$. Recall (see (67) and (65)) that:

$$\begin{aligned} H_3^{[n]}(\mathfrak{f}_n) &= \sum_{\ell\geq 0} 2^{-\ell} \langle \mu, \tilde{f}_{\ell,n}^2 \rangle\, \mathbf{1}_{\{\ell\leq p\}} \quad \text{and} \quad |H_4^{[n]}(\mathfrak{f}_n)| \\ &\leq \sum_{\ell\geq 0,\, k\geq 0} 2^{k-\ell}\, |\langle \mu, \mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right)\rangle|. \end{aligned}$$

Thanks to (22) and (3), we have:

$$|\langle \mu, \mathcal{P}\left(\mathcal{Q}^k \tilde{f}_{\ell,n} \otimes^2\right)\rangle| \leq \| \mathcal{Q}^k \tilde{f}_{\ell,n} \|_{L^2(\mu)}^2 \leq C\, \alpha^{2k}\, \| f_{\ell,n} \|_{L^2(\mu)}^2 \leq C\, \alpha^{2k}\, \mathfrak{c}_2^2.$$

Using Assumption 4.3 (iii), we get that

$$|\langle \mu, \mathcal{P}(\tilde{f}_{\ell,n} \otimes^2)\rangle| \leq |\langle \mu, \mathcal{P}(f_{\ell,n} \otimes^2)\rangle| + \langle \mu, f_{\ell,n}\rangle^2 \leq (1+\Delta)\delta_{\ell,n}. \tag{73}$$

We deduce from (51) (for $k \geq 1$) and the previous upper-bound (for $k = 0$) and dominated convergence that $\lim_{n\to\infty} H_4^{[n]}(\mathfrak{f}_n) = 0$.

We now prove that $\lim_{n\to\infty} H_3^{[n]}(\mathfrak{f}_n) = \sigma^2$. We define $\sigma_n^2 = \sum_{\ell=0}^n 2^{-\ell} \parallel f_{\ell,n} \parallel_{L^2(\mu)}^2$, so that by Assumption 4.3 (iv), $\lim_{n\to\infty} \sigma_n^2 = \sigma^2$. We have:

$$|H_3^{[n]}(\mathfrak{f}_n) - \sigma_n^2| \leq \sum_{\ell=p+1}^{n} 2^{-\ell}\langle \mu, f_{\ell,n}^2\rangle + \sum_{\ell=0}^{p} 2^{-\ell}\langle \mu, f_{\ell,n}\rangle^2 \leq \mathfrak{c}_2^2 2^{-p} + \Delta \sum_{\ell=0}^{p} 2^{-\ell}\delta_{\ell,n}.$$

Then use dominated convergence to deduce that $\lim_{n\to\infty} |H_3^{[n]}(\mathfrak{f}_n) - \sigma_n^2| = 0$. This implies that $\lim_{n\to\infty} V_1(n) = \sigma^2$ in probability. □

Using (49), we have the following result as a direct consequence of Lemmas 5.3, 5.4 and 5.5. Recall $V(n)$ defined in (49).

Lemma 5.6 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we have that $V(n)$ converges in probability toward σ^2 defined by (28).*

We now check the Lindeberg's condition using a fourth moment condition. Recalling $\Delta_{n,i}(\mathfrak{f}_n)$ defined in (48), we set

$$R_3(n) = \sum_{i\in\mathbb{G}_{n-p_n}} \mathbb{E}\left[\Delta_{n,i}(\mathfrak{f}_n)^4\right]. \tag{74}$$

Lemma 5.7 *Under the assumptions of Theorem 4.8 ($2\alpha^2 < 1$), we get $\lim_{n\to\infty} R_3(n) = 0$.*

The proof of Lemma 5.7 is omitted since it uses the same arguments as in Lemmas 5.4, 5.5 and the same sort of techniques as were used in the proof of Lemma 5.11 in [2].

We can now use Theorem 3.2 and Corollary 3.1, p. 58, and the Remark p. 59 from [13] to deduce from Lemmas 5.6 and 5.7 that $\Delta_n(\mathfrak{f}_n)$ converges in distribution toward a Gaussian real-valued random variable with deterministic variance σ^2 defined by (28). Using Remark 5.1 and Lemma 5.2, we then deduce Theorem 4.8.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10959-022-01205-w.

Acknowledgements We warmly thank Adélaïde Olivier for numerous discussions in the preliminary phases of this project.

Data Availability All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

In this section, we recall the following result due to Bochner (see [17, Theorem 1A] which can be easily extended to any dimension $d \geq 1$).

Lemma 6.1 *Let $(h_n, n \in \mathbb{N})$ be a sequence of positive numbers converging to 0 as n goes to infinity. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that $\int_{\mathbb{R}^d} |g(x)|\mathrm{d}x < +\infty$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that $\| f \|_\infty < +\infty$, $\int_{\mathbb{R}^d} |f(y)|\,\mathrm{d}y < +\infty$ and $\lim_{|x|\to+\infty} |x|f(x) = 0$. Define*

$$g_n(x) = h_n^{-d} \int_{\mathbb{R}^d} f(h_n^{-1}(x-y))g(y)\mathrm{d}y.$$

Then, we have at every point x of continuity of g,

$$\lim_{n\to+\infty} g_n(x) = g(x) \int_{\mathbb{R}} f(y)\mathrm{d}y.$$

We also recall useful results on BMC which are recalled in [2].

Lemma 6.2 *Let $f, g \in \mathcal{B}(S)$, $x \in S$ and $n \geq m \geq 0$. Assuming that all the quantities below are well defined, we have:*

$$\begin{aligned} \mathbb{E}_x\left[M_{\mathbb{G}_n}(f)\right] &= |\mathbb{G}_n|\, \mathcal{Q}^n f(x) = 2^n\, \mathcal{Q}^n f(x), \\ \mathbb{E}_x\left[M_{\mathbb{G}_n}(f)^2\right] &= 2^n\, \mathcal{Q}^n(f^2)(x) + \sum_{k=0}^{n-1} 2^{n+k}\, \mathcal{Q}^{n-k-1}\left(\mathcal{P}\left(\mathcal{Q}^k f \otimes \mathcal{Q}^k f\right)\right)(x), \end{aligned} \tag{75}$$

$$\begin{aligned} \mathbb{E}_x\left[M_{\mathbb{G}_n}(f)M_{\mathbb{G}_m}(g)\right] &= 2^n \mathcal{Q}^m\left(g\mathcal{Q}^{n-m} f\right)(x) \\ &\quad + \sum_{k=0}^{m-1} 2^{n+k}\, \mathcal{Q}^{m-k-1}\left(\mathcal{P}\left(\mathcal{Q}^k g \otimes_{\mathrm{sym}} \mathcal{Q}^{n-m+k} f\right)\right)(x). \end{aligned} \tag{76}$$

Lemma 6.3 *Let X be a BMC with kernel $\mathcal{P}$ and initial distribution ν such that (iii) from Assumption 4.2 (with $k_0 \in \mathbb{N}$) is in force. There exists a finite constant C, such that for all $f \in \mathcal{B}_+(S)$ all $n \geq k_0$, we have:*

$$\begin{aligned} |\mathbb{G}_n|^{-1}\mathbb{E}[M_{\mathbb{G}_n}(f)] \leq C \| f \|_{L^1(\mu)} \quad \text{and} \quad |\mathbb{G}_n|^{-1}\mathbb{E}\left[M_{\mathbb{G}_n}(f)^2\right] \\ \leq C \sum_{k=0}^{n} 2^k \| \mathcal{Q}^k f \|_{L^2(\mu)}^2. \end{aligned} \tag{77}$$

References

1. Bitseki Penda, S.V., Delmas, J.F.: Central Limit Theorem for Bifurcating Markov Chains. hal-03047744 (2020)
2. Bitseki Penda, S.V., Delmas, J.F.: Central limit theorem for bifurcating Markov chains under L^2 ergodic conditions. Adv. Appl. Probab. (2022). https://doi.org/10.1017/apr.2022.3
3. Bitseki Penda, S.V., Delmas, J.F.: Central limit theorem for bifurcating Markov chains under pointwise ergodic conditions. Ann. Appl. Probab. **32**(5), 3817–3849 (2022). https://doi.org/10.1214/21-AAP1774
4. Bitseki Penda, S.V., Hoffmann, M., Olivier, A.: Adaptive estimation for bifurcating Markov chains. Bernoulli **23**(4B), 3598–3637 (2017). https://doi.org/10.3150/16-BEJ859
5. Bitseki Penda, S.V., Roche, A.: Local bandwidth selection for kernel density estimation in a bifurcating Markov chain model. J. Nonparametr. Stat. **32**(3), 535–562 (2020)
6. Cowan, R., Staudte, R.: The bifurcating autoregression model in cell lineage studies. Biometrics **42**(4), 769–783 (1986). https://doi.org/10.2307/2530692
7. Delmas, J.F., Marsalle, L.: Detection of cellular aging in a Galton–Watson process. Stoch. Process. Appl. **120**(12), 2495–2519 (2010). https://doi.org/10.1016/j.spa.2010.07.002
8. Devroye, L., Neininger, R.: Density approximation and exact simulation of random variables that are solutions of fixed-point equations. Adv. Appl. Probab. **34**(2), 441–468 (2002)
9. Douc, R., Moulines, E., Priouret, P., Soulier, P.: Markov Chains. Springer Series in Operations Research and Financial Engineering. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97704-1
10. Doumic, M., Escobedo, M., Tournus, M.: Estimating the division rate and kernel in the fragmentation equation. Ann. Inst. H. Poincaré Anal. Non Linéaire **35**(7), 1847–1884 (2018). https://doi.org/10.1016/j.anihpc.2018.03.004
11. Doumic, M., Hoffmann, M., Krell, N., Robert, L.: Statistical estimation of a growth-fragmentation model observed on a genealogical tree. Bernoulli **21**(3), 1760–1799 (2015). https://doi.org/10.3150/14-BEJ623
12. Guyon, J.: Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. Ann. Appl. Probab. **17**(5–6), 1538–1569 (2007). https://doi.org/10.1214/105051607000000195
13. Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. Probability and Mathematical Statistics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York (1980)
14. Hoffmann, M., Marguet, A.: Statistical estimation in a randomly structured branching population. Stoch. Process. Appl. **129**(12), 5236–5277 (2019). https://doi.org/10.1016/j.spa.2019.02.015
15. Knape, M., Neininger, R.: Approximating perpetuities. Methodol. Comput. Appl. Probab. **10**(4), 507–529 (2008)
16. Masry, E.: Recursive probability density estimation for weakly dependent stationary processes. IEEE Trans. Inf. Theory **32**(2), 254–267 (1986)
17. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**(3), 1065–1076 (1962)
18. Roussas, G.G.: Nonparametric estimation in Markov processes. Ann. Inst. Stat. Math. **21**(1), 73–87 (1969)
19. Roussas, G.G.: Estimation of transition distribution function and its quantiles in Markov processes: Strong consistency and asymptotic normality. In: Nonparametric Functional Estimation and Related Topics, pp. 443–462. Springer (1991)
20. Tsybakov, A.B.: Introduction to Nonparametric Estimation. Springer (2008)