KERNEL ESTIMATION OF THE TRANSITION DENSITY IN BIFURCATING MARKOV CHAINS

Siméon Valère Bitseki Penda

KERNEL ESTIMATION OF THE TRANSITION DENSITY IN BIFURCATING MARKOV CHAINS.

S. VALÈRE BITSEKI PENDA

Abstract. We study the kernel estimator of the transition density of bifurcating Markov chains. Under some ergodic and regularity properties, we prove that this estimator is consistent and asymptotically normal. Next, in the numerical studies, we propose two data-driven methods to choose the bandwidth parameters. These methods are based on the so-called two bandwidths approach.

Keywords: Kernel estimator, cross validation method, rule of thumb type method, bifurcating Markov chains, binary trees, asymptotic normality.

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

1. Introduction

This article is devoted to the study of the kernel estimators of the transition probability of bifurcating Markov chains. Before defining these estimators, let us first introduce useful definitions, notations and assumptions.

1.1. **Bifurcating Markov chains.** Let $d \geq 1$ be a natural integer. In order to simplify the notations in the sequel, we set $S = \mathbb{R}^d$ and we equip S with its Borel σ-algebra that we denote by $\mathscr{S}$. We denote by $\mathcal{B}(S)$ (resp. $\mathcal{B}_b(S)$, resp. $\mathcal{B}_+(S)$) the set of (resp. bounded, resp. non-negative) $\mathbb{R}$-valued measurable functions defined on S. For $f \in \mathcal{B}(S)$, we set $\| f \|_\infty = \sup\{|f(x)|, \, x \in S\}$. For a finite measure λ on $(S, \mathscr{S})$ and $f \in \mathcal{B}(S)$ we shall write $\langle \lambda, f\rangle$ for $\int f(x)\,\mathrm{d}\lambda(x)$ whenever this integral is well defined. We denote by $\mathcal{C}_b(S)$ (resp. $\mathcal{C}_+(S)$) the set of bounded (resp. non-negative) $\mathbb{R}$-valued continuous functions defined on S. For all natural integer $q \geq 1$, we equip S^q with $\mathscr{S}^{\otimes q} = \mathscr{S} \otimes \ldots \otimes \mathscr{S}$, the usual product σ-field on S^q.

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 $\mathcal{Q}f \in \mathcal{B}(S)$. For $n \in \mathbb{N}$, we denote by Q^n the n-th iterate of Q defined by $Q^0 = I_d$, 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. Recall that $\mathbb{N}$ is the set of non-negative integers and $\mathbb{N}^* = \mathbb{N}\setminus\{0\}$. 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 k-th generation, $\mathbb{T}_k$ to the tree up the k-th 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 (BMC) from Guyon [8].

Definition 1.1. *We say a stochastic process indexed by $\mathbb{T}$, $X = (X_i, i \in \mathbb{T})$, is a bifurcating Markov chain on a measurable space $(S, \mathscr{S})$ with initial probability distribution ν on $(S, \mathscr{S})$ and probability kernel $\mathcal{P}$ on $S \times \mathscr{S}^{\otimes 2}$, a BMC in short, if:*

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

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

We define three probability kernels P_0, P_1 and $\mathcal{Q}$ on $S \times \mathscr{S}$ by:

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

Notice that P_0 (resp. P_1) is the restriction of the first (resp. second) marginal of $\mathcal{P}$ to S. Following Guyon [8], 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 $x \in S$).

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:

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

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

For $i \in \mathbb{T}$ and $k \in \mathbb{N}$, we also define the σ-field:

$$\mathcal{H}_{i,k} = \sigma(X_u, u \in i\mathbb{T}_k)$$

We end this section with a useful notations. By convention, for $f, g \in \mathcal{B}(S)$, we define the function $f \otimes g$ by $(f \otimes g)(x, y) = f(x)g(y)$ for $x, y \in S$ and

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

Notice that $\mathcal{P}(g \otimes_{\text{sym}} \mathbf{1}) = \mathcal{Q}(g)$ for $g \in \mathcal{B}_+(S)$.

For all $u \in \mathbb{T}$, we denote by $X_u^\Delta = (X_u, X_{u0}, X_{u1})$ the mother-daughters triangle. For a finite subset $A \subset \mathbb{T}$, we define:

$$M_A(f) = \sum_{u\in A} f(X_u) \quad \text{if } f \in \mathcal{B}(S) \quad \text{and} \quad M_A(f) = \sum_{u\in A} f(X_u^\Delta) \quad \text{if } f \in \mathcal{B}(S^3).$$

In the sequel we will also use the following notation: let g and h be two functions which depend on one variable, x say; we denote by $g \oplus h$ the function of three variables, $xx_0x_1 := (x, x_0, x_1)$ say, defined by

$$(g \oplus h)(xx_0x_1) = g(x_0) + h(x_1).$$

1.2. **Assumptions on the law of the bifurcating Markov chains** $(X_i, i \in \mathbb{T})$**.** For a set $F \subset \mathcal{B}(S)$ of $\mathbb{R}$-valued functions, we write $F^2 = \{f^2; f \in F\}$, $F \otimes F = \{f_0 \otimes f_1; f_0, f_1 \in F\}$, and $P(F) = \{Pf; f \in F\}$ whenever a kernel P act on F. Following [8], we state a structural assumption on the set of functions we shall consider.

Assumption 1.2. *Let $F \subset \mathcal{B}(S)$ be a set of $\mathbb{R}$-valued functions such that:*

- (i) *F is a vector subspace which contains the constants;*
- (ii) *$F^2 \subset F$;*
- (iii) *$F \subset L^1(\nu)$;*
- (iv) *$F \otimes F \subset L^1(\mathcal{P}(x,\cdot))$ for all $x \in S$, and $\mathcal{P}(F \otimes F) \subset F$.*

The condition (iv) implies that $P_0(F) \subset F$,$P_1(F) \subset F$ as well as $\mathcal{Q}(F) \subset F$. Notice that if $f \in F$, then even if $|f|$ does not belong to F, using conditions (i) and (ii), we get, with $g = (1+f^2)/2$, that $|f| \leq g$ and $g \in F$. Typically, the set F can be the set $\mathcal{C}_b(S)$ of bounded real-valued functions, or the set of smooth real-valued functions such that all derivatives have at most polynomials growth.

Following [8], we also consider the following ergodic properties for $\mathcal{Q}$.

Assumption 1.3. *There exists a probability measure μ on $(S, \mathscr{S})$ such that $F \subset L^1(\mu)$ and for all $f \in F$, we have the point-wise convergence $\lim_{n\to\infty} \mathcal{Q}^n f = \langle \mu, f\rangle$ and there exists $g \in F$ with:*

$$|\mathcal{Q}^n(f)| \leq g \quad \text{for all } n \in \mathbb{N}. \tag{3}$$

Moreover, there exists a function $V : [1, +\infty) \mapsto (0, \infty)$ such that $V \in F$ and constants $\alpha \in (0, 1)$ and $M < \infty$ such that:

$$\sup_{|f|\leq V} |\mathcal{Q}^n f - \langle \mu, f\rangle| \leq M\alpha^n V \quad \text{for all } n \in \mathbb{N}. \tag{4}$$

Remark 1.4. In particular, (4) implies that for all $f \in \mathcal{B}_b(S)$, we have

$$|\mathcal{Q}^n f - \langle \mu, f\rangle| \leq M\|f\|_\infty\, \alpha^n\, V \quad \text{for all } n \in \mathbb{N}. \tag{5}$$

Next, we have the following assumption on the existence of the density of $\mathcal{P}$.

Assumption 1.5. *The transition kernel $\mathcal{P}$ has a density, still denoted by $\mathcal{P}$, with respect to the Lebesgue measure.*

Remark 1.6. Assumption 1.5 implies that the transition kernel $\mathcal{Q}$ has a density, still denoted by $\mathcal{Q}$, with respect to the Lebesgue measure. More precisely, we have $\mathcal{Q}(x, y) = 2^{-1} \int_S (\mathcal{P}(x, y, z) + \mathcal{P}(x, z, y))dz$. This implies in particular that the invariant probability μ has a density, still denoted by μ, with respect to the Lebesgue measure (for more details, we refer for e.g. to [7], chap 6).

Remark 1.7. Under Assumption 1.5, the probability measure $\mu^\triangle$ defined on S^3 by

$$\mu^\triangle(dxx_0x_1) = \mu(dx)P(x, dx_0, dx_1),$$

has density with respect to the Lebesgue measure, that we also denote by $\mu^\triangle$, given by $\mu^\triangle(xx_0x_1) = \mu(x)\mathcal{P}(x, x_0, x_1)$, for all $xx_0x_1 \in S^3$.

Assumption 1.8. *We assume that the following constant is finite:*

$$C_0 = \sup_{x,x_0,x_1\in S}(\mu(x) + \mathcal{Q}(x,x_0) + \mathcal{P}(x,x_0,x_1)).$$

Remark 1.9. We recall from [8, Theorem 11 and Corollary 15] that under Assumptions 1.2 and 1.3, we have for $f \in F$ the following convergence in L^2 (resp. a.s.):

$$\lim_{n\to\infty}|\mathbb{G}_n|^{-1}M_{\mathbb{G}_n}(f) = \langle\mu, f\rangle \quad \text{and} \quad \lim_{n\to\infty}|\mathbb{T}_n|^{-1}M_{\mathbb{T}_n}(f) = \langle\mu, f\rangle. \tag{6}$$

Now, the rest of the paper is organized follows. In Section 2, we define the estimators of the transition density $\mathcal{P}$ based on the observation of a subpopulation. We will see that these are quotient estimators. In Section 3, we study the consistency and the asymptotic normality of the numerators of the estimators of $\mathcal{P}$. Section 4 is dedicated to the study of consistency and asymptotic normality of the estimators of $\mathcal{P}$. In Section 5, we will illustrate the consistency of our estimators in a bifurcating Markov model called bifurcating autoregressive process (BAR, for short). In particular, we will develop two data-driven bandwidth selection methods: the least squares Cross-Validation in Section 5.1 and the rule of thumb type method in Section 5.2. Sections 6-8 are dedicated to the proofs of the main Theorems. In Section 9, we prove a useful inequality and in Section 10, we recall some useful results.

2. Kernel estimators of the transition density $\mathcal{P}$

Recall that $S = \mathbb{R}^d$. Our aim is to estimate the transition density $\mathcal{P}$ from the observation of the subpopulation $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$. For that purpose, assume we observe $\mathbb{X}^{\Delta n} = (X_u^{\Delta})_{u\in\mathbb{A}_n}$ *i.e.* we have $2^{n+2}-1$ (or 3×2^n) random variables with value in S. Let $K_0 : S \to \mathbb{R}$ and $K : S^3 \to \mathbb{R}$ be a functions such that $\int_S K_0(x)dx = 1$ and $K = K_0\otimes K_0\otimes K_0$. We also have $\int_{S^3} K(xx_0x_1)dxx_0x_1 = 1$. Let $(h_n, n\in\mathbb{N})$ be a sequence of positive numbers which converges to 0 as n goes to infinity. When there is no ambiguity, we write h for h_n. Let $\mathbb{A}_n \in \{\mathbb{T}_n, \mathbb{G}_n\}$. We define, for all $x \in S$:

$$\widehat{\mu}_{\mathbb{A}_n}(x) = \frac{1}{|\mathbb{A}_n|h^{d/2}}\sum_{u\in\mathbb{A}_n}K_{0h_n}(x - X_u), \tag{7}$$

where $K_{0h_n}(x-y) = h_n^{-d/2}K_0(h_n^{-1}(x-y))$ and for all $xx_0x_1 \in S^3$:

$$\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1) = \frac{1}{|\mathbb{A}_n|h^{3d/2}}\sum_{u\in\mathbb{A}_n}K_{h_n}(xx_0x_1 - X_u^{\Delta}) \quad \text{and} \quad \widehat{\mathcal{P}}_{\mathbb{A}_n}(xx_0x_1) = \frac{\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)}, \tag{8}$$

where

$$K_{h_n}(xx_0x_1 - yy_0y_1) = h_n^{-3d/2}\,K\left(h_n^{-1}(x-y), h_n^{-1}(x_0-y_0), h_n^{-1}(x_1-y_1)\right),$$

with the convention that $\widehat{\mathcal{P}}_{\mathbb{A}_n}(xx_0x_1) = 0$ if $\widehat{\mu}_{\mathbb{A}_n}(x) = 0$. However, we stress that if we assume that K_0 is strictly positive, then $\widehat{\mu}_{\mathbb{A}_n}(x) > 0$ for all $x \in S$.

From now on, we fix $xx_0x_1 \in S^3$, that is, we are interested in the estimation at the point xx_0x_1. We assume that $\mu(x) \neq 0$. We consider the function f_n defined by:

$$f_n(yy_0y_1) = K_h(xx_0x_1 - yy_0y_1). \tag{9}$$

If we want to be more rigorous, we must write f_{n,xx_0x_1} instead of f_n. But, we choose to write without the index xx_0x_1 in order to simplify the writing.

Remark 2.1. Note that asymptotic behavior (consistence and asymptotic normality) of $\widehat{\mu}_{\mathbb{A}_n}$ have been studied in [1].

Remark 2.2. We stress that the results of this paper can be straightforward extended the case where the bandwidth $\boldsymbol{h}$ is a vector of $\mathbb{R}^{3d}$, with possibly different coordinates. More precisely, one can take the bandwidth $\boldsymbol{h} = (h_i, 1 \leq i \leq 3d)$, where the h_i's may take different values. For our convenience, we choose to work with the case where all the coordinates are the same, that is $h_i = h$ for all $1 \leq i \leq 3d$.

3. Consistency and Asymptotic normality for $\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)$

First, we will study the consistency and the asymptotic normality of $\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)$. We set $\tilde{f}_n = f_n - \langle \mu, \mathcal{P} f_n \rangle$. We begin with the study asymptotic normality of $N_{n,\emptyset}(f_n) = |\mathbb{G}_n|^{-1/2} M_{\mathbb{A}_n}(\tilde{f}_n)$. This is motivated by the following decomposition:

$$\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1) - \mu^{\Delta}(xx_0x_1) = (|\mathbb{A}_n|\, |\mathbb{G}_n|^{-1/2}\, h^{3d/2})^{-1} N_{n,\emptyset}(f_n) + (h^{-3d/2} \langle \mu^{\Delta}, f_n \rangle - \mu^{\Delta}(xx_0x_1)). \tag{10}$$

We will need the following assumption on the bandwidth and on the kernel.

Assumption 3.1.
We assume that:

(i) $h_n = 2^{-n\gamma}$ *and* $2\alpha^2 < 2^{3d\gamma}$ *for some* $\gamma \in (0, 1/3d)$.
(ii) The kernel K_0 *(resp.* K_0^2*) is integrable and square integrable.*

Remark 3.2. Assumption 3.1, (i) implies in particular that

$$\lim_{n\to\infty} |\mathbb{G}_n| h_n^{3d} = \infty \quad \text{and} \quad \lim_{n\to\infty} (2\alpha^2)^n\, h_n^{3d} = 0. \tag{11}$$

Note that Assumption 3.1, (i) is automatically satisfied if $2\alpha^2 \leq 1$, regardless of the value of α. For $2\alpha^2 > 1$, this Assumption implies that the choice of the bandwidth is function of the ergodicity rate of the auxiliary Markov chain Y.

We have the following result.

Theorem 3.3. *Let X be a BMC with kernel $\mathcal{P}$ and initial distribution ν such that Assumptions 1.2, 1.3, 1.5, 1.8 and 3.1 hold. Then, we have the following convergence in distribution:*

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

where G is a centered Gaussian random variable with finite variance $\sigma^2 = 2\, \|K_0\|_2^6\, \mu^{\Delta}(x, x_0, x_1)$ if $\mathbb{A}_n = \mathbb{T}_n$ and $\sigma^2 = \|K_0\|_2^6\, \mu^{\Delta}(x, x_0, x_1)$ if $\mathbb{A}_n = \mathbb{G}_n$.

Proof. The proof of Theorem 3.3 is postponed to Section 6. □

Next, in order to study the asymptotic normality of $\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)$, we do the following additional hypothesis.

Assumption 3.4. *We assume that Assumption 3.1 holds and there exists $s > 0$ such that the following holds.*

(iv) ***The density μ^{Δ} (resp. μ) belongs to the (isotropic) Hölder class of order $(s, \ldots, s) \in \mathbb{R}^{3d}$ (resp. $(s, \ldots, s) \in \mathbb{R}^d$):*** *The density μ^{Δ} admits partial derivatives with respect to x_j, for all $j \in \{1, \ldots 3d\}$, up to the order $\lfloor s \rfloor$ and there exists a finite constant $L > 0$ such that for all $x = (x_1, \ldots, x_{3d}), \in \mathbb{R}^{3d}$, $t \in \mathbb{R}$ and $j \in \{1, \ldots, 3d\}$:*

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

where (x_{-j}, t) denotes the vector x where we have replaced the j^{th} coordinate x_j by t, with the convention $\partial^0 \mu^\Delta / \partial x_j^0 = \mu^\Delta$. The same thing for the density μ.

(v) ***The kernel K_0 is of order*** $(\lfloor s \rfloor, \ldots, \lfloor s \rfloor) \in \mathbb{N}^d$***:*** *We have $\int_{\mathbb{R}^d} |x|^s K_0(x)\, dx < \infty$ and $\int_{\mathbb{R}} x_j^k K_0(x)\, dx_j = 0$ for all $k \in \{1, \ldots, \lfloor s \rfloor\}$ and $j \in \{1, \ldots, d\}$.*

(vi) ***Bandwith control:*** *We have $\gamma > 1/(2s + 3d)$, that is $\lim_{n\to\infty} |\mathbb{G}_n| h_n^{2s+3d} = 0$.*

Notice that Assumption 3.4-(iv) implies that μ^Δ (resp. μ) is at least Hölder continuous as $s > 0$. We have the following result.

Theorem 3.5. *Let X be a BMC with kernel $\mathcal{P}$ and initial distribution ν. Under Assumptions of Theorem 3.3, we have for all (x, x_0, x_1) and $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$*

$$\widehat{\mu}^\Delta_{\mathbb{A}_n}(xx_0x_1) \xrightarrow[n\to\infty]{\mathbb{P}} \mu^\Delta(xx_0x_1) \quad \text{in probability.} \tag{12}$$

Moreover, under the additional Assumption 3.4, we have the following convergence in distribution:

$$|\mathbb{A}_n|^{1/2}\, h_n^{3d/2}\, (\widehat{\mu}^\Delta_{\mathbb{A}_n}(xx_0x_1) - \mu^\Delta(xx_0x_1)) \xrightarrow[n\to\infty]{(d)} G \quad \text{in distribution,}$$

where G is a centered Gaussian random variable with finite variance $\sigma^2 = \|K_0\|_2^6\, \mu^\Delta(x, x_0, x_1)$.

Proof. The proof is postponed to Section 7. □

4. Consistency and Asymptotic normality for $\widehat{\mathcal{P}}^\Delta_{\mathbb{A}_n}(xx_0x_1)$

We are now in position to state consistency and asymptotic normality of kernel estimator of the transition density $\mathcal{P}$. First, as a consequence of (8), (12) and (43) below, we have the following result.

Lemma 4.1. *Under the Assumptions of Theorem 3.3, we have for all (x, x_0, x_1) and $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$:*

$$\widehat{\mathcal{P}}_{\mathbb{A}_n}(x, x_0, x_1) \xrightarrow[n\to\infty]{\mathbb{P}} \mathcal{P}(x, x_0, x_1) \quad \text{in probability.}$$

Next, we have the following result.

Theorem 4.2. *Let X be a BMC with kernel $\mathcal{P}$ and initial distribution ν. Under the assumptions of Theorem 3.3 and the additional Assumption 3.4, we have,*

$$\sqrt{|\mathbb{A}_n| h_n^{3d}}(\widehat{\mathcal{P}}_{\mathbb{A}_n}(x, x_0, x_1) - \mathcal{P}(x, x_0, x_1)) \xrightarrow[n\to\infty]{(\mathrm{d})} G,$$

where G is a centered Gaussian real-valued random variable with mean 0 and variance

$$\sigma^2 \;=\; \|K_0\|_2^6\, \mathcal{P}(x, x_0, x_1)/\mu(x).$$

Proof. The proof is postponed to Section 8. □

5. Numerical studies

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 $u \in \mathbb{T}$:

$$\begin{cases} X_{u0} = a_0 X_u + b_0 + \varepsilon_{u0} \\ X_{u1} = a_1 X_u + b_1 + \varepsilon_{u1}, \end{cases} \tag{13}$$

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 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:

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

with

$$g(x,y,z) = (y - a_0 x - b_0)^2 - 2\rho\sigma^{-2}(y - a_0 x - b_0)(z - a_1 x - b_1) + (z - a_1 x - b_1)^2.$$

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

$$\mathcal{Q}(x, dy) = \frac{1}{2\sqrt{2\pi\sigma^2}} \left(\mathrm{e}^{-(y - a_0 x - b_0)^2/2\sigma^2} + \mathrm{e}^{-(y - a_1 x - b_1)^2/2\sigma^2}\right) dy.$$

We will estimate the transition density $\mathcal{P}$ in a compact set $D \subset \mathbb{R}^3$. For that purpose, we use the estimator $\widehat{\mathcal{P}}_{\mathbb{G}_n}(xx_0x_1)$, for all $xx_0x_1 \in D$, given in (8), with the Gaussian kernel K_0 defined by

$$K_0(x) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-x^2/2}. \tag{14}$$

Since the bandwidth is a function of the ergodicity rate which is unknown, we have to develop a method based on data in order to select it. To select the optimal bandwidth for $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ defined in (8), we will use the so-called "two bandwidths approach" (see for e.g. [6]). More precisely, since $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ is a quotient estimator, we select separately the bandwidths for the numerator (h_N, say) and the denominator (h_D, say). For that purpose, we propose two methods: the cross validation and the rule of thumb type method. The objective here is not to study nor to compare theoretically these two methods. This will be done in the future works. Our objective is only the see the numerical performances of each method. Our conclusion is that even if the rule of thumb developed in this paper give a crude approximation, it as more computational benefit with respect to the least squared cross validation.

5.1. **Bandwidth selection by least squares Cross-Validation method.** We choose the bandwidths which minimises the mean integrated squared errors (MISEs)

$$\mathbb{E}\left[\int_{\mathbb{R}^3} (\widehat{\mu}^{\Delta}_{\mathbb{G}_n} - \mu^{\Delta})^2(xx_0x_1)dxdx_0dx_1\right] \quad \text{and} \quad \mathbb{E}\left[\int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n} - \mu)^2(x)dx\right],$$

where $\widehat{\mu}^{\Delta}_{\mathbb{G}_n}$ and $\widehat{\mu}_{\mathbb{G}_n}$ are defined in (7) and (8). This is equivalent to minimise the functions J^{Δ} and J defined by

$$J^{\Delta}(h) = \mathbb{E}\left[\int_{\mathbb{R}^3} (\widehat{\mu}^{\Delta}_{\mathbb{G}_n})^2(xx_0x_1)dxdx_0dx_1\right] - 2\mathbb{E}\left[\int_{\mathbb{R}^3} (\widehat{\mu}^{\Delta}_{\mathbb{G}_n}\mu^{\Delta})(x,x_0,x_1)dxdx_0dx_1\right]$$

and

$$J(h) = \mathbb{E}\left[\int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n})^2(x)dx\right] - 2\mathbb{E}\left[\int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n}\mu)(x)dx\right].$$

The method to select the bandwidths is the following.

(1) We divide the sample $(X^{\Delta}_u)_{u\in\mathbb{G}_n}$ into K disjoints subsamples $\{(X^{\Delta}_u)_{u\in\mathbb{G}_n^{(k)}}, k \in \{1,\dots,K\}\}$, with $(\mathbb{G}_n^{(k)}, k \in \{1,\dots,K\})$ a partition of $\mathbb{G}_n$.
(2) For each subsample $(X_u)_{u\in\mathbb{G}_n^{(k)}}$:

(a) We set $\widehat{\mu}^{\Delta}_{\mathbb{G}_n[-k]}$ and $\widehat{\mu}_{\mathbb{G}_n[-k]}$ the estimators of μ^{Δ} and μ obtaining using the subsample $(X^{\Delta}_u)_{u\in\mathbb{G}_n} \setminus (X^{\Delta}_u)_{u\in\mathbb{G}_n^{(k)}}$, where for two sets A and B, $B\setminus A$ denotes the set of elements in B but not in A. More precisely,

$$\widehat{\mu}^{\Delta}_{\mathbb{G}_n[-k]}(xx_0x_1) = \frac{1}{|\mathbb{G}_n[-k]|h^3} \sum_{u\in\mathbb{G}_n[-k]} K_0\left(\frac{x-X_u}{h}\right) K_0\left(\frac{x_0-X_{u0}}{h}\right) K_0\left(\frac{x_1-X_{u1}}{h}\right)$$

and

$$\widehat{\mu}_{\mathbb{G}_n[-k]}(x) = \frac{1}{|\mathbb{G}_n[-k]|h} \sum_{u\in\mathbb{G}_n[-k]} K_0\left(\frac{x-X_u}{h}\right),$$

where we set $\mathbb{G}_n[-k] = \mathbb{G}_n \setminus \mathbb{G}_n^{(k)}$.

(b) We approximate J and J^{Δ} by

$$\widehat{J}^{(k)}(h) = \int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n[-k]})^2(x)dx - \frac{2}{|\mathbb{G}_n^{(k)}|} \sum_{u\in\mathbb{G}_n^{(k)}} \widehat{\mu}_{\mathbb{G}_n[-k]}(X_u) \quad \text{and}$$

$$\widehat{J}^{\Delta(k)}(h) = \int_{\mathbb{R}^3} (\widehat{\mu}^{\Delta}_{\mathbb{G}_n[-k]})^2(xx_0x_1)dxdx_0dx_1 - \frac{2}{|\mathbb{G}_n^{(k)}|} \sum_{u\in\mathbb{G}_n^{(k)}} \widehat{\mu}^{\Delta}_{\mathbb{G}_n[-k]}(X_u, X_{u0}, X_{u1}).$$

(3) Let $\mathcal{H} = \{h_1, \ldots, h_m\} \subset (0,1]$ be a bandwidth grid. Then, the selected bandwidths $\widehat{h}_N$ and $\widehat{h}_D$ for the numerator and the denominator of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ are given by:

$$\widehat{h}_N := \arg\min_{h\in\mathcal{H}} \frac{1}{K}\sum_{k=1}^{K} \widehat{J}^{\Delta(k)}(h) \quad \text{and} \quad \widehat{h}_D := \arg\min_{h\in\mathcal{H}} \frac{1}{K}\sum_{k=1}^{K} \widehat{J}^{(k)}(h).$$

Finally, the estimator used for numerical studies is $\widetilde{\mathcal{P}}_{\mathbb{G}_n}$ defined by:

$$\widetilde{\mathcal{P}}_{\mathbb{G}_n}(xx_0x_1) = \frac{\widetilde{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)}{\widetilde{\mu}_{\mathbb{A}_n}(x)},$$

with

$$\widetilde{\mu}_{\mathbb{G}_n}(x) = \frac{1}{|\mathbb{G}_n|\widehat{h}_D} \sum_{u\in\mathbb{G}_n} K_0\left(\frac{x-X_u}{\widehat{h}_D}\right) \quad \text{and}$$

$$\widetilde{\mu}^{\Delta}_{\mathbb{G}_n}(xx_0x_1) = \frac{1}{|\mathbb{G}_n|\widehat{h}_N^3} \sum_{u\in\mathbb{G}_n} K_0\left(\frac{x-X_u}{\widehat{h}_N}\right) K_0\left(\frac{x_0-X_{u0}}{\widehat{h}_N}\right) K_0\left(\frac{x_1-X_{u1}}{\widehat{h}_N}\right).$$

This method is known as the K-fold cross validation. One advantage of this method in the context of bifurcating Markov chains is that it is not requires the knowledge of the ergodicity rate. The main drawback being that it requires a lot of time for calculations.

5.2. **Gaussian symmetric BAR reference bandwidth selection.** In order to define a selection rule, we consider the special case of Gaussian BAR defined by (13) where $a_0 = a_1 := a$ and $\rho = 0$ as a reference model. It is well known (see [3]) that the densities of the transition kernel $\mathcal{Q}$ of the auxiliary Markov chain and the invariant probability μ associated to $\mathcal{Q}$ are given by:

$$\mathcal{Q}(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-ax)^2}{2\sigma^2}\right) \quad \text{and} \quad \mu(x) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left(-\frac{x^2}{2\sigma_a^2}\right),$$

where $\sigma_a = \sigma/\sqrt{1-a^2}$. The density of the transition probability $\mathcal{P}$ associated to this bifurcating Markov chain is defined by $\mathcal{P}(x,y,z) = \mathcal{Q}(x,y)\mathcal{Q}(x,z)$ and then, $\mu^{\Delta}(x,y,z) = \mu(x)\mathcal{Q}(x,y)\mathcal{Q}(x,z)$. We then have that the invariant densities μ and μ^{Δ} are square integrable and twice differentiable, with the second order derivative of μ and all the second order partial derivatives of μ^Δ bounded, continuous and square integrable. It is also well that the Markov chain with transition $\mathcal{Q}$ is geometrically ergodic and that the geometric ergodic rate of convergence is a (for more details, see for e.g Example 2.8 in [3]). In particular, following the proof of Proposition 28 in [8], one can prove that for all derivable function f such f and f' are bounded, we have

$$|\mathcal{Q}^n f(x) - \langle \mu, f\rangle| \leq \|f'\|_\infty (\sigma(1-a)^{-1} + |x|)a^n. \tag{15}$$

We assume that $\mathcal{L}(X_\emptyset) = \mu$, that is $X_\emptyset$ is distributed as μ, which implies that the process is stationary. We are now going to behave as if we did not know the invariant measures μ and μ^Δ and the transition probability $\mathcal{P}$. Recall the kernel density estimator of μ defined in (7) and the kernel K_0 defined in (14). Recall also the kernel estimator of the transition density $\mathcal{P}$ defined in (8). To ease notation, we write $\widehat{\mu}$ and $\widehat{\mu}^\Delta$ instead of $\widehat{\mu}_{\mathbb{G}_n}$ and $\widehat{\mu}^\Delta_{\mathbb{G}_n}$ respectively. We recall that our strategy is to select bandwidth for the numerator and the denominator in the estimation of $\mathcal{P}$. First, we treat the denominator $\widehat{\mu}$. The selection rule is based on the following asymptotic upper bound, known as asymptotic mean squared error:

$$\mathbb{E}\left[(\widehat{\mu}(x) - \mu(x))^2\right] \leq \frac{h^4}{2}\kappa_2^2 \mu''(x)^2 + \frac{4\|K_0\|_2^2}{|\mathbb{G}_n|h}\mu(x) + \frac{2\,C_{a,\sigma}}{a^2|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2a^2)^k + o(1). \tag{16}$$

where

$$C_{a,\sigma} = \frac{1}{\pi\sigma^2(1+a)(1-a)^2} \quad \text{and} \quad \kappa_2 = \int_{\mathbb{R}} y^2 K_0(y)dy = 1. \tag{17}$$

We postponed the proof of (16) in Section 9. Now, let p be a non negative probability density defined in $\mathbb{R}$ such that $\|p\|_\infty \leq 1$. Then, (16) implies that

$$\int_{\mathbb{R}} \mathbb{E}\left[(\widehat{\mu}(x) - \mu(x))^2\right] p(x)dx \leq \frac{h^4}{2}\kappa_2^2 \int_{\mathbb{R}} \mu''(x)^2 dx + \frac{4\|K_0\|_2^2}{|\mathbb{G}_n|h} + \frac{2\,C_{a,\sigma}}{a^2|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2a^2)^k + o(1). \tag{18}$$

The term in the left hand side of (18) is a modification of asymptotic mean integrated squared error that we call p-AMISE. We have introduced it because the last term in (16) does not depend on x. Finally, (18) suggests us to choose the bandwidth which minimises the function G defined by

$$G(h) = \frac{h^4}{2}\kappa_2^2 \int_{\mathbb{R}} \mu''(x)^2 dx + \frac{4\|K_0\|_2^2}{|\mathbb{G}_n|h} + M_{a,\sigma}\, a^{2n}\, \mathbf{1}_{\{2a^2>1\}},$$

where $M_{a,\sigma} = 2/(\pi\sigma^2 a^2(1+a)(1-a)^2(2a^2-1))$. Optimizing in h, we get that the optimal bandwidth h_D (for the denominator of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ defined in (8)) is given by

$$h_D = (c_1/4c_2)^{1/5}|\mathbb{G}_n|^{-1/5}\,\mathbf{1}_{\{2a^2\leq 2^{-1/5}\}} + (c_1/M_{a,\sigma})\,(2a^2)^{-n}\,\mathbf{1}_{\{2a^2>2^{-1/5}\}},$$

where

$$c_1 = 4\|K_0\|_2^2 = \frac{2}{\sqrt{\pi}} \quad \text{and} \quad c_2 = (1/2)\kappa_2^2\int_{\mathbb{R}}\mu''(x)^2 dx = \frac{3}{16\sqrt{\pi}}\,\sigma_a^{-5}.$$

Next, we treat the numerator $\widehat{\mu}^\Delta$ of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$. Recalling Remark 2.2, we consider the general case where for all $xx_0x_1 \in S^3$:

$$\widehat{\mu}^\Delta_{\mathbb{G}_n}(xx_0x_1) = \frac{1}{|\mathbb{G}_n|hh_0h_1}\sum_{u\in\mathbb{G}_n} K_0(h^{-1}(x-X_u))K_0(h_0^{-1}(x_0-X_{u0}))K_0(h_1^{-1}(x_1-X_{u1})).$$

Recall that for a vector $\boldsymbol{v}$, $\boldsymbol{v}^t$ denotes its transpose. As in (16), we have the following asymptotic upper bound:

$$\mathbb{E}\left[\left(\widehat{\mu}^{\Delta}(xx_0x_1) - \mu^{\Delta}(xx_0x_1)\right)^2\right] \leq \frac{1}{2}\kappa_2^2 \left(\boldsymbol{h}^t\, \boldsymbol{H}(xx_0x_1)\, \boldsymbol{h}\right)^2 + \frac{6\,\|K_0\|_2^6}{|\mathbb{G}_n|hh_0h_1}\mu^{\Delta}(xx_0x_1) + \frac{C_{a,\sigma}^{\Delta}\, \mathcal{P}(xx_0x_1)^2}{|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2\,a^2)^k + o(1), \tag{19}$$

where κ_2 is defined in (17), $\boldsymbol{h} = (h, h_0, h_1)^t$, and

$$\boldsymbol{H}(xx_0x_1) = \begin{pmatrix} \frac{\partial^2\mu^{\Delta}}{\partial x^2}(x, x_0, x_1) & 0 & 0 \\ 0 & \frac{\partial^2\mu^{\Delta}}{\partial x_0^2}(x, x_0, x_1) & 0 \\ 0 & 0 & \frac{\partial^2\mu^{\Delta}}{\partial x_1^2}(x, x_0, x_1) \end{pmatrix},$$

$$C_{a,\sigma}^{\Delta} = \frac{4}{e\,\pi\sigma^2a^2(1-a)^2(1+a)}.$$

We let the proof of (19) to the reader since it follows the same lines that of (16). Let p be a non negative probability density defined in $\mathbb{R}$ such that $\|p\|_\infty \leq 1$. Integrating (19) with respect to $p(x)dxdx_0dx_1$, we get

$$\iiint_{\mathbb{R}^3} \mathbb{E}\left[\left(\widehat{\mu}^{\Delta}(xx_0x_1) - \mu^{\Delta}(xx_0x_1)\right)^2\right] p(x)dxdx_0dx_1 \leq \frac{1}{2}\kappa_2^2 \iiint_{\mathbb{R}^3} \left(\boldsymbol{h}^t\, \boldsymbol{H}(xx_0x_1)\, \boldsymbol{h}\right)^2 dxdx_0dx_1 + \frac{6\,\|K_0\|_2^6}{|\mathbb{G}_n|hh_0h_1} + \frac{C_{a,\sigma}^{\Delta}}{4\pi\sigma^2|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2\,a^2)^k + o(1).$$

Now, the latter equation suggests us to choose the vector bandwidth $\boldsymbol{h}$ which minimises the function G^{Δ} defined by

$$G^{\Delta}(h, h_0, h_1) = \frac{1}{2}\kappa_2^2 \iiint_{\mathbb{R}^3} \left(\boldsymbol{h}^t\, \boldsymbol{H}(xx_0x_1)\, \boldsymbol{h}\right)^2 dxdx_0dx_1 + \frac{6\,\|K_0\|_2^6}{|\mathbb{G}_n|hh_0h_1} + M_{a,\sigma}^{\Delta}\, a^{2n}\mathbf{1}_{\{2a^2>1\}}.$$

where $M_{a,\sigma}^{\Delta} = (1/(4\,\pi\,\sigma^2(2a^2-1)))\,C_{a,\sigma}^{\Delta}$. Optimizing the function G^{Δ} in $\boldsymbol{h}$, we get that the optimal bandwidth $\boldsymbol{h}_N = (h_N, h_{0N}, h_{1N})$ is given by

$$\begin{aligned} h_N &= \left(c_2^{\Delta}/(4c^{\Delta})\right)^{1/7}|\mathbb{G}_n|^{-1/7}\,\mathbf{1}_{\{2a^2\leq 2^{-3/7}\}} + (c_2^{\Delta}/M_{a,\sigma}^{\Delta})^{1/3}\,(2a^2)^{-n/3}\,\mathbf{1}_{\{2a^2>2^{-3/7}\}}, \\ h_{0N} &= \left(c_2^{\Delta}/(4c_0^{\Delta})\right)^{1/7}|\mathbb{G}_n|^{-1/7}\,\mathbf{1}_{\{2a^2\leq 2^{-3/7}\}} + (c_2^{\Delta}/M_{a,\sigma}^{\Delta})^{1/3}\,(2a^2)^{-n/3}\,\mathbf{1}_{\{2a^2>2^{-3/7}\}}, \\ h_{1N} &= \left(c_2^{\Delta}/(4c_1^{\Delta})\right)^{1/7}|\mathbb{G}_n|^{-1/7}\,\mathbf{1}_{\{2a^2\leq 2^{-3/7}\}} + (c_2^{\Delta}/M_{a,\sigma}^{\Delta})^{1/3}\,(2a^2)^{-n/3}\,\mathbf{1}_{\{2a^2>2^{-3/7}\}}, \end{aligned}$$

where κ_2 is defined in (17), $c_2^{\Delta} = 6\,\|K_0\|_2^6 = 6/(8\pi\sqrt{\pi})$ and

$$\begin{aligned} c^{\Delta} &= \frac{1}{2}\,\kappa_2^2 \iiint_{\mathbb{R}^3}\left(\frac{\partial^2\mu^{\Delta}}{\partial x^2}(x, x_0, x_1)\right)^2 dxdx_0dx_1 = \frac{3(1+a^2)^2}{64\pi\sqrt{\pi}(1-a^2)^3}\sigma_a^{-7}, \\ c_0^{\Delta} &= \frac{1}{2}\,\kappa_2^2 \iiint_{\mathbb{R}^3}\left(\frac{\partial^2\mu^{\Delta}}{\partial x_0^2}(x, x_0, x_1)\right)^2 dxdx_0dx_1 = \frac{3}{64\pi\sqrt{\pi}(1-a^2)^3}\sigma_a^{-7}, \\ c_1^{\Delta} &= \frac{1}{2}\,\kappa_2^2 \iiint_{\mathbb{R}^3}\left(\frac{\partial^2\mu^{\Delta}}{\partial x_1^2}(x, x_0, x_1)\right)^2 dxdx_0dx_1 = \frac{3}{64\pi\sqrt{\pi}(1-a^2)^3}\sigma_a^{-7}. \end{aligned}$$

We have

$$\frac{c_1}{4c_2} = b_1\,\sigma_a^5, \quad \frac{c_1}{M_{a,\sigma}} = b_2\,\sigma_a^2, \quad \frac{c_2^\Delta}{4\,c^\Delta} \;=\; b_3\,\sigma_a^7, \quad \frac{c_2^\Delta}{4\,c_0^\Delta} = \frac{c_2^\Delta}{4\,c_1^\Delta} = b_4\,\sigma_a^7, \quad \frac{c_2^\Delta}{M_{a,\sigma}^\Delta} = b_5\;\sigma_a^4,$$

where

$$\begin{aligned} &b_1 = \frac{32}{3}, \quad b_2 = \sqrt{\pi}a^2(1-a^2)(1+a)(1-a)^2(2a^2-1), \quad b_3 = \frac{48(1-a^2)^3}{12(1+a^2)^2}, \\ &b_4 = 4(1-a^2)^3, \quad b_5 = \frac{3a^2(1-a^2)^3(1+a)(2a^2-1)}{4\sqrt{\pi}}. \end{aligned}$$

Since for $a \in (0,1)$ the constants b_i, $i \in \{1,\dots,5\}$, are bounded, we can approximate h_D, h_N, h_{0N}, h_{1N} by:

$$\begin{aligned} &\widehat{h}_D = |\mathbb{G}_n|^{-1/5}\,\widehat{\sigma}_a\,\mathbf{1}_{\{2\,\widehat{a}^2 \leq 2^{-1/5}\}} \;+\; (2\,\widehat{a}^2)^{-n}\,\widehat{\sigma}_a\,\mathbf{1}_{\{2\widehat{a}^2 > 2^{-1/5}\}}, \\ &\widehat{h}_N = \widehat{h}_{0N} = \widehat{h}_{1N} = |\mathbb{G}_n|^{-1/7}\,\widehat{\sigma}_a\,\mathbf{1}_{\{2\,\widehat{a}^2 \leq 2^{-3/7}\}} \;+\; (2\,\widehat{a}^2)^{-n/3}\,\widehat{\sigma}_a\,\mathbf{1}_{\{2\widehat{a}^2 > 2^{-3/7}\}}, \end{aligned}$$

where $\widehat{\sigma}_a$ is the estimator of the standard deviation of the measure μ and $\widehat{a}$ is the estimator of the geometric ergodic rate. Note that in practice, the estimators $\widehat{h}_N$, $\widehat{h}_{0N}$ and $\widehat{h}_{1N}$ differ slightly. Indeed, for $\widehat{h}_D$ and $\widehat{h}_N$, $\widehat{\sigma}_a$ is computed using the sample $(X_u, u \in \mathbb{G}_n)$, for $\widehat{h}_{0N}$, $\widehat{\sigma}_a$ is computed using the sample $(X_{u0}, u \in \mathbb{G}_n)$ and for $\widehat{h}_{1N}$, $\widehat{\sigma}_a$ is computed using the sample $(X_{u1}, u \in \mathbb{G}_n)$. Recall that for $i \in \mathbb{T}$ and $A \subset \mathbb{T}$, $iA = \{ij, j \in A\}$, where ij is the concatenation of the two sequences $i, j \in \mathbb{T}$. For the geometric ergodic rate, we propose the following estimator, which is inspired from [9]:

$$\widehat{a} = \left(\frac{\sum_{u \in \mathbb{G}_{n-m+1}} \sum_{v \in u\mathbb{G}_{m-1}} (X_u - \overline{X})(X_v - \overline{X})}{\sum_{u \in \mathbb{G}_n} (X_u - \overline{X})^2} \right)^{1/m} \quad \text{with} \quad \overline{X} = \frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} X_u,$$

where m is a large enough natural integer such that $m = \mathcal{O}(n)$. The choice $m = \lfloor n/2 \rfloor + 1$ seems to be relevant. Finally, the estimator used for numerical studies is $\widetilde{\mathcal{P}}_{\mathbb{G}_n}$ defined by:

$$\widetilde{\mathcal{P}}_{\mathbb{G}_n}(xx_0x_1) = \frac{\widetilde{\mu}^\Delta_{\mathbb{A}_n}(xx_0x_1)}{\widetilde{\mu}_{\mathbb{A}_n}(x)},$$

with

$$\begin{aligned} &\widetilde{\mu}_{\mathbb{G}_n}(x) = \frac{1}{|\mathbb{G}_n|\widehat{h}_D} \sum_{u \in \mathbb{G}_n} K_0\left(\frac{x - X_u}{\widehat{h}_D}\right) \quad \text{and} \\ &\widetilde{\mu}^\Delta_{\mathbb{G}_n}(xx_0x_1) = \frac{1}{|\mathbb{G}_n|\widehat{h}_N^3} \sum_{u \in \mathbb{G}_n} K_0\left(\frac{x - X_u}{\widehat{h}_N}\right) K_0\left(\frac{x_0 - X_{u0}}{\widehat{h}_{0N}}\right) K_0\left(\frac{x_1 - X_{u1}}{\widehat{h}_{1N}}\right). \end{aligned}$$

This method is an adaptation of the rule of thumb developed by Silverman in [12]. The novelty here is that the ergodic rate of convergence is taken into account in the estimation procedure. In the context of BMC, The main advantage of this method is that it not requires a lot of time for calculations. However, this method is a crude approximation which works for approximately "Gaussian" bifurcating Markov chains.

5.3. **Numerical illustrations.** In order to validate our method, we consider two cases:

case 1: $(a_0, a_1, b_{0,1}, \sigma, \rho) = (0.7, 0.5, 0, 0, 1, 0)$;
case 2: $(a_0, a_1, b_{0,1}, \sigma, \rho) = (1.2, 0.7, 0, 0, 1, 0)$;

In case 2, we allow the dynamic of the new pole to be unstable, even if the entire dynamic of the system is stable. Following the terminology of Bitseki and Delmas in [2, 3], the case 2 corresponds to supercritical case.

As we can see, Figure 1-8, the two methods allow to recover the true function when the size of the data increases. Consequently, we conclude that our method is valid.

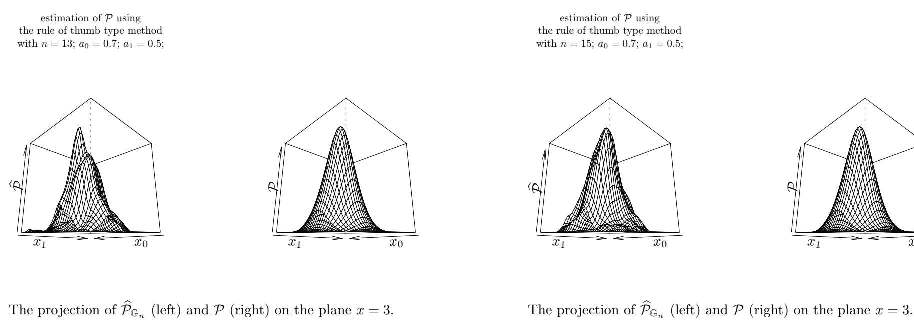


FIGURE 1

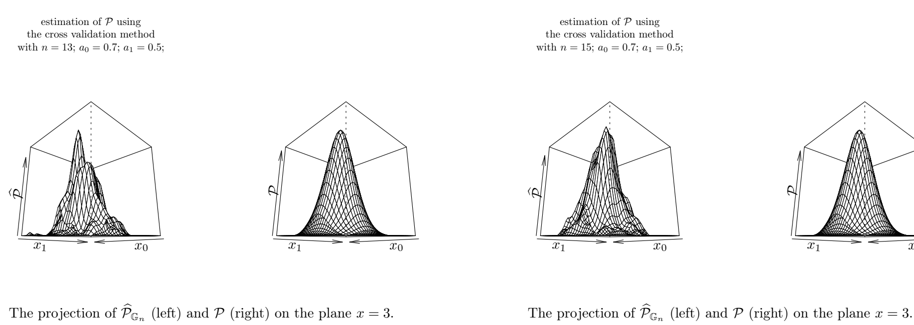


FIGURE 2

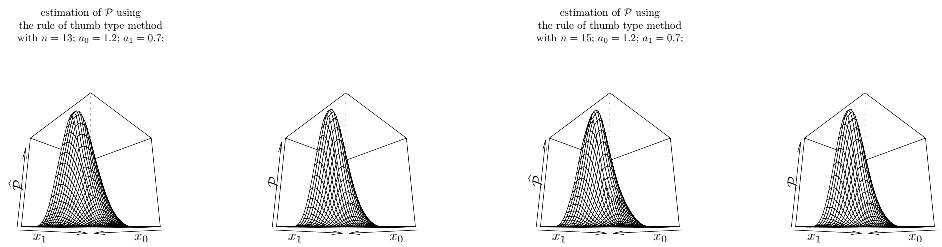
6. PROOF OF THEOREM 3.3

We begin the proof with $\mathbb{A}_n = \mathbb{T}_n$. Let $(p_n, n \in \mathbb{N})$ be a non-decreasing sequence of elements of $\mathbb{N}^*$ such that, for all $\lambda > 0$:

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

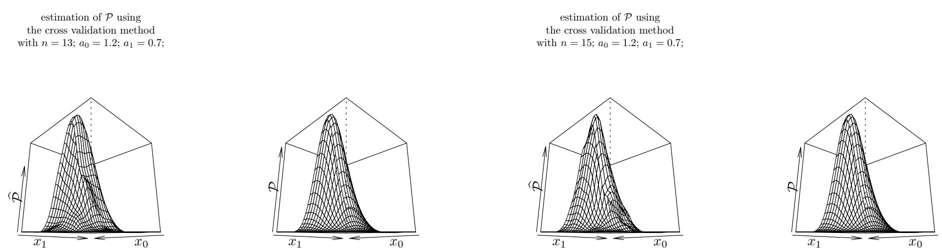
When there is no ambiguity, we write p for p_n. Recall the function f_n defined in (9). We have the following decomposition:

$$N_{n,\emptyset}(f_n) = R_0(n) + R_1(n) + \Delta_n(f_n), \tag{20}$$



The projection of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ (left) and $\mathcal{P}$ (right) on the plane $x = 3$. The projection of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ (left) and $\mathcal{P}$ (right) on the plane $x = 3$.

FIGURE 3



The projection of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ (left) and $\mathcal{P}$ (right) on the plane $x = 3$. The projection of $\widehat{\mathcal{P}}_{\mathbb{G}_n}$ (left) and $\mathcal{P}$ (right) on the plane $x = 3$.

FIGURE 4

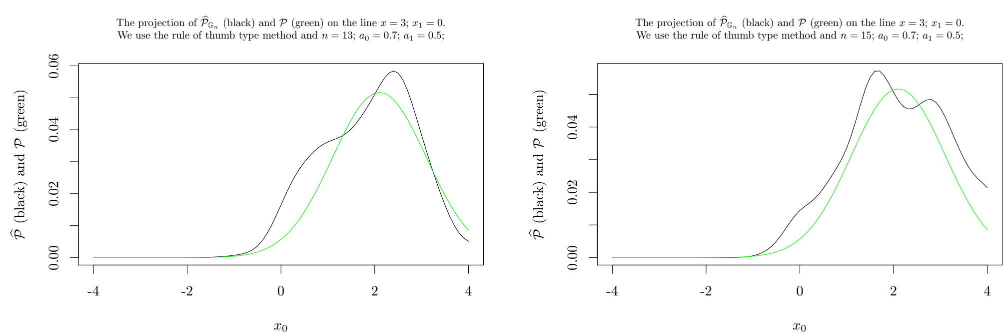


FIGURE 5

where:

$$R_0(n) = |\mathbb{G}_n|^{-1/2} \sum_{k=0}^{n-p-1} M_{\mathbb{G}_k}(\tilde{f}_n); \quad R_1(n) = \sum_{i \in \mathbb{G}_{n-p}} \mathbb{E}[N_{n,i}(f_n)|\mathcal{F}_i]; \quad \Delta_n(f_n) = \sum_{i \in \mathbb{G}_{n-p}} \Delta_{n,i}(f_n),$$

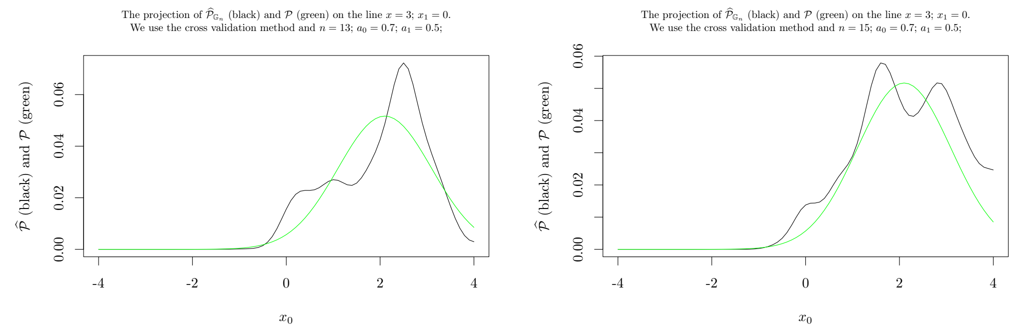


FIGURE 6

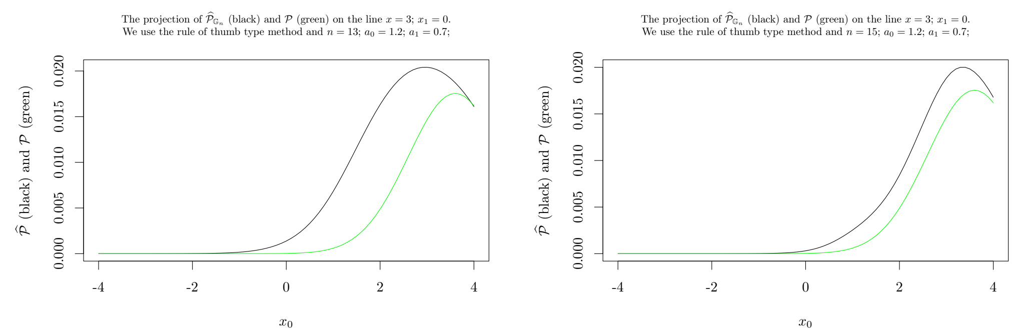


FIGURE 7

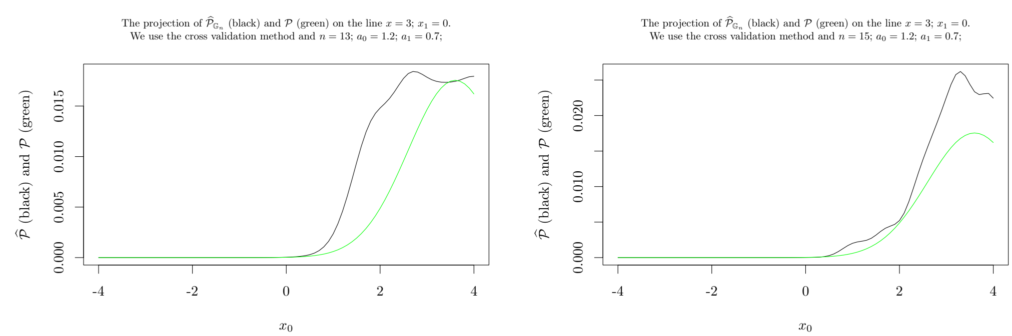


FIGURE 8

and for all $i \in \mathbb{G}_{n-p}$,

$$N_{n,i}(f_n) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{p} M_{i\mathbb{G}_{p-\ell}}(\tilde{f}_n) \quad \text{and} \quad \Delta_{n,i}(f_n) = N_{n,i}(f_n) - \mathbb{E}[N_{n,i}(f_n)|\mathcal{F}_i].$$

Note that using the branching Markov property, we have, for all $i \in \mathbb{G}_{n-p}$,

$$\mathbb{E}[M_{i\mathbb{G}_{p-\ell}}(\tilde{f}_n)|\mathcal{F}_i] = \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\mathcal{P}\tilde{f}_n)]. \tag{21}$$

We have the following convergence.

Lemma 6.1. *Under the assumptions of Theorem 3.3, we have that* $\lim_{n\to\infty} \mathbb{E}[R_0(n)^2] = 0$.

Proof. We have

$$\begin{aligned}\mathbb{E}[R_0(n)^2] &= |\mathbb{G}_n|^{-1}\, \mathbb{E}[(\sum_{k=0}^{n-p-1} \sum_{u\in\mathbb{G}_k} \tilde{f}_n(X_u^{\Delta})^2] \\ &\leq |\mathbb{G}_n|^{-1}\, (\sum_{k=0}^{n-p-1} \mathbb{E}[(\sum_{u\in\mathbb{G}_k} \tilde{f}_n(X_u^{\Delta})^2]^{1/2})^2,\end{aligned} \tag{22}$$

where we used the Minkowski inequality for the first inequality. By developing the term in the expectation, we get

$$\begin{aligned}\mathbb{E}[(\sum_{u\in\mathbb{G}_k} \tilde{f}_n(X_u^{\Delta})^2] &= \mathbb{E}[\sum_{u\neq v\in\mathbb{G}_k} \mathbb{E}[\tilde{f}_n(X_u^{\Delta})\tilde{f}_n(X_v^{\Delta})|X_u, X_v]] + \mathbb{E}[\sum_{u\in\mathbb{G}_k} \mathbb{E}[(\tilde{f}_n)^2(X_u^{\Delta})|X_u]] \\ &= \mathbb{E}[\sum_{u\neq v\in\mathbb{G}_k} \mathcal{P}\tilde{f}_n(X_u)\mathcal{P}\tilde{f}_n(X_v)] + \mathbb{E}[\sum_{u\in\mathbb{G}_k} \mathcal{P}((\tilde{f}_n)^2)(X_u)] \\ &= \mathbb{E}[(\sum_{u\in\mathbb{G}_k} \mathcal{P}\tilde{f}_n(X_u^{\Delta}))^2] + \mathbb{E}[\sum_{u\in\mathbb{G}_k} (\mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2)(X_u)],\end{aligned}$$

where we used the branching Markov property for the second inequality and the fact that $\mathcal{P}((\tilde{f}_n)^2) - (\mathcal{P}\tilde{f}_n)^2 = \mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2$ for the third equality. Using (22) and using the inequalities $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $(a+b)^2 \leq 2a^2 + 2b^2$, we get

$$\begin{aligned}\mathbb{E}[R_0(n)^2] &\leq |\mathbb{G}_n|^{-1}\, (\sum_{k=0}^{n-p-1} (\mathbb{E}[(\sum_{u\in\mathbb{G}_k} \mathcal{P}\tilde{f}_n(X_u^{\Delta}))^2]^{1/2} + \mathbb{E}[\sum_{u\in\mathbb{G}_k} (\mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2)(X_u^{\Delta})]^{1/2}))^2 \\ &\leq 2|\mathbb{G}_n|^{-1}((\sum_{k=0}^{n-p-1} \mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}(\tilde{f}_n))^2]^{1/2})^2 + (\sum_{k=0}^{n-p-1} \mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2)]^{1/2})^2).\end{aligned}$$

Note that from Lemma 10.1, we have $\|\mathcal{P}(f_n)\|_\infty \leq C h_n^{d/2}$ and $\|\mathcal{P}(f_n^2)\|_\infty \leq C$. Recall $\mathcal{P}(\tilde{f}_n) = \mathcal{P}(f_n) - \langle \mu, \mathcal{P}(f_n)\rangle$. Then, using (53), (5) and Lemma 10.1, we get

$$\mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}\tilde{f}_n)^2] \leq C\, h_n^d \quad \text{if } k \in \{0, 1\}$$

and for all $k \geq 2$,

$$\mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}\tilde{f}_n)^2] \leq \begin{cases} C\, h_n^{2d}\, 2^k & \text{if } 2\alpha^2 \leq 1 \\ C\, 2^k\, (h_n^{2d} \,+\, (2\alpha^2)^k\, h_n^{3d}) & \text{if } 2\alpha^2 > 1. \end{cases}$$

It follows for the two last inequalities that

$$|\mathbb{G}_n|^{-1}\, (\sum_{k=0}^{n-p-1} \mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}(\tilde{f}_n))^2]^{1/2})^2 \leq C\, 2^{-n}\, h_n^{3d} \,+\, C\, 2^{-p}\, h_n^{2d} \,+\, 2^{-p}\, (2\alpha^2)^{n-p}\, h_n^{3d}\, \mathbf{1}_{\{2\alpha^2>1\}}.$$

Using (11), Assumption 3.1 and since $\lim_{n\to\infty} p_n = \infty$, it follows that

$$\lim_{n\to\infty} |\mathbb{G}_n|^{-1}\Big(\sum_{k=0}^{n-p-1} \mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}(\tilde{f}_n))^2]^{1/2}\Big)^2 = 0.$$

Next, using Lemma 10.1, we get $\mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2)] \leq C\, 2^k$. This implies that

$$\lim_{n\to\infty} |\mathbb{G}_n|^{-1}\Big(\sum_{k=0}^{n-p-1} \mathbb{E}[M_{\mathbb{G}_k}(\mathcal{P}((f_n)^2) - (\mathcal{P}f_n)^2)]^{1/2}\Big)^2 \leq C \lim_{n\to\infty} 2^{-p} = 0$$

and this ends the proof. □

Next, we have the following convergence.

Lemma 6.2. *Under the assumptions of Theorem 3.3, we have that* $\lim_{n\to\infty} \mathbb{E}[R_1(n)^2] = 0$.

Proof. Using (21), we get

$$R_1(n) = \sum_{k=0}^{p} R_1(k,n),$$

with

$$R_1(k,n) = |\mathbb{G}_n|^{-1/2}\, |\mathbb{G}_{p-k}|\, M_{\mathbb{G}_{n-p}}(\mathcal{Q}^{p-k}\mathcal{P}\tilde{f}_n).$$

It follows that

$$\mathbb{E}\left[R_1(n)^2\right]^{1/2} \leq \sum_{k=0}^{p} \left(\mathbb{E}\left[R_1(k,n)^2\right]\right)^{1/2}. \tag{23}$$

Following the proof of Lemma 4.2 in [3] and using (5) and Lemma 10.1, we find that

$$\mathbb{E}\left[R_1(k,n)^2\right] \leq C\, 2^{-p}\, h_n^{2d}\, \mathbf{1}_{\{k=p\}} + \begin{cases} C\, h_n^{3d}\, 2^{-p}\, (2\alpha)^{2(p-k)} & \text{if } 2\alpha^2 < 1 \\ C\,(n-p)\, h_n^{3d}\, 2^{-k} & \text{if } 2\alpha^2 = 1 \\ C\, 2^{-p}\, (2\alpha^2)^{n-p}\, h_n^{3d} (2\alpha)^{2(p-k)} & \text{if } 2\alpha^2 > 1. \end{cases}$$

From (23), this implies that

$$\mathbb{E}\left[R_1(n)^2\right]^{1/2} \leq C\, 2^{-p/2}\, h_n^{d} + \begin{cases} C\, h_n^{3d/2} \sum_{k=0}^{p} 2^{-k/2}\, (2\alpha^2)^{(p-k)/2} & \text{if } 2\alpha^2 < 1 \\ C\,(n-p)^{1/2}\, h_n^{3d/2} & \text{if } 2\alpha^2 = 1 \\ C\,(2\alpha^2)^{n/2}\, h_n^{3d/2} & \text{if } 2\alpha^2 > 1. \end{cases}$$

From the latter inequality and using Assumption 3.1, we deduce that $\lim_{n\to\infty} \mathbb{E}[R_1(n)^2] = 0$. □

We now study the bracket

$$V(n) = \sum_{i\in\mathbb{G}_{n-p}} \mathbb{E}[\Delta_{n,i}(f_n)^2 | \mathcal{F}_i].$$

Note that for $i \in \mathbb{G}_{n-p}$, we have

$$\mathbb{E}[\Delta_{n,i}(f_n)^2|\mathcal{F}_i] = |\mathbb{G}_n|^{-1}\mathbb{E}[(\sum_{k=0}^{p} M_{i\mathbb{G}_{p-k}}(\tilde{f}_n))^2|\mathcal{F}_i] - |\mathbb{G}_n|^{-1}(\mathbb{E}[\sum_{k=0}^{p} M_{i\mathbb{G}_{p-k}}(\tilde{f}_n)|\mathcal{F}_i])^2.$$

Using the branching Markov chain property, this implies that

$$V(n) = V_1(n) + V_2(n) - R_2(n), \tag{24}$$

with

$$
\begin{aligned}
&V_1(n) = |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{k=0}^{p} \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}(\tilde{f}_n)^2],\\
(25)\qquad &V_2(n) = 2|\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{0\leq k<\ell\leq p} \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\tilde{f}_n)M_{\mathbb{G}_{p-k}}(\tilde{f}_n)],\\
&R_2(n) = |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} (\sum_{k=0}^{p} 2^{p-k}\mathcal{Q}^{p-k}\mathcal{P}\tilde{f}_n(X_u))^2.
\end{aligned}
$$

We have the following result.

Lemma 6.3. *Under the assumptions of Theorem 3.3, we have the following convergence:*

$$\lim_{n\to\infty} \mathbb{E}[R_2(n)] = 0.$$

Proof. We have using (52), (5) and Lemma 10.1:

$$
\begin{aligned}
\mathbb{E}\left[R_2(n)\right] &= |\mathbb{G}_n|^{-1}\,|\mathbb{G}_{n-p}|\,\langle \nu, \mathcal{Q}^{n-p}\left(\Big(\sum_{k=0}^{p}|\mathbb{G}_{p-k}|\,\mathcal{Q}^{p-k}\mathcal{P}\tilde{f}_n\Big)^2\right)\rangle\\
&\leq C\,2^{-p}\langle \nu, \mathcal{Q}^{n-p}((\mathcal{P}\tilde{f}_n)^2)\rangle + C\,2^{-p}\langle \nu, \mathcal{Q}^{n-p}((\sum_{k=0}^{p-1} 2^{p-k}\mathcal{Q}^{p-k-1}(\mathcal{Q}\mathcal{P}\tilde{f}_n))^2)\rangle\\
&\leq\; C\,2^{-p}\,h_n^{2d} \;+\; C\,2^{-p}\,h_n^{3d}\,a_n,
\end{aligned}
$$

where the sequence $(a_n, n\geq 1)$ is defined by

$$
a_n = \begin{cases} 1 & \text{if } 2\alpha < 1\\ p^2 & \text{if } 2\alpha = 1\\ (2\alpha)^{2p} & \text{if } 2\alpha > 1. \end{cases}
$$

Using Assumption 3.1, and in particular Remark 3.2, it follows that $\lim_{n\to\infty}\mathbb{E}[R_2(n)] = 0$. □

Next, we have the following result.

Lemma 6.4. *Under the assumptions of Theorem 3.3, we have the following convergence:*

$$\lim_{n\to\infty} \mathbb{E}[V_2(n)^2] = 0.$$

Proof. Let $0 \leq k < \ell \leq p$ and $i \in \mathbb{G}_{n-p}$. Conditioning two times, first by $\mathcal{H}_{i,p-k}$ and next by $\mathcal{H}_{i,p-\ell+1}$, and using the branching Markov property, we get

$$\mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\tilde{f}_n)M_{\mathbb{G}_{p-k}}(\tilde{f}_n)] = 2^{\ell-k-1}\mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\tilde{f}_n)M_{\mathbb{G}_{p-\ell}}(g_{k,\ell,n})],$$

where we set $g_{k,\ell,n} = \mathcal{Q}^{\ell-k-1}\mathcal{P}\tilde{f}_n \oplus \mathcal{Q}^{\ell-k-1}\mathcal{P}\tilde{f}_n$. Next, conditioning by $\mathcal{H}_{i,p-\ell}$ and using the branching Markov property, we get

$$
\begin{aligned}
\mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\tilde{f}_n)M_{\mathbb{G}_{p-\ell}}(g_{k,\ell,n})] &= \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\mathcal{P}(\tilde{f}_n g_{k,\ell,n}) - \mathcal{P}\tilde{f}_n\mathcal{P}g_{k,\ell,n})]\\
&\quad + \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-\ell}}(\mathcal{P}\tilde{f}_n)M_{\mathbb{G}_{p-\ell}}(\mathcal{P}g_{k,\ell,n})].
\end{aligned}
$$

From the foregoing and using (52), (54) and (25), it follows that:

$$V_2(n) = V_5(n) + V_6(n),$$

where
$$V_5(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{5,n}) \quad \text{and} \quad V_6(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{6,n}),$$
with
$$H_{5,n} = \sum_{0\leq k<\ell} 2^{-k} \mathcal{Q}^{p-\ell}(\mathcal{P}(\tilde{f}_n\, g_{k,\ell,n}))\mathbf{1}_{\{\ell\leq p\}} \quad \text{and}$$
$$H_{6,n} = \sum_{\substack{0\leq k<\ell\\ r\geq 0}} 2^{-k+r} \mathcal{Q}^{p-\ell-r-1}\mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^r\mathcal{P} g_{k,\ell,n})\mathbf{1}_{\{r+\ell<p\}}.$$
First, we treat the term $V_6(n)$. Note that we have
$$\mathcal{Q}^r\mathcal{P} g_{k,\ell,n} = 2\, \mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n.$$
We set
$$h_{k,\ell,r}^{(n)} = 2^{r-k+1} \mathcal{Q}^{p-1-(r+\ell)}\mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n) \quad \text{and}$$
$$h_{k,\ell,r} = 2^{r-k+1}\langle \mu, \mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n)\rangle.$$
We consider the following sums:
$$H_6^{[n]} = \sum_{\substack{0\leq k<\ell\\ r\geq 0}} h_{k,\ell,r}\mathbf{1}_{\{r+\ell<p\}} \quad \text{and} \quad A_{6,n} = H_{6,n} - H_6^{[n]} = \sum_{\substack{0\leq k<\ell\\ r\geq 0}} (h_{k,\ell,r}^{(n)} - h_{k,\ell,r})\mathbf{1}_{\{r+\ell<p\}}.$$
Using Lemma 10.1, we have for all $0 \leq k < \ell$ and $r \geq 0$:
$$\begin{aligned} |\mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n)| &\leq \|\mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n)\|_\infty \, \|\mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathbf{1})\|_\infty \\ &\leq \|\mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n)\|_\infty \, \|\mathcal{Q}^{r+1}\mathcal{P}\tilde{f}_n\|_\infty \leq C\, h_n^{3d}. \end{aligned} \tag{26}$$
Moreover, using (5) and Lemma 10.1, we have for all $0 \leq k < \ell$ and $r \geq 1$:
$$\begin{aligned} |\mathcal{P}(\mathcal{Q}^r\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^{r+\ell-k}\mathcal{P}\tilde{f}_n)| &\leq C\, \alpha^{2r+\ell-k}\, \|\mathcal{Q}(\mathcal{P}f_n)\|_\infty^2 \, \mathcal{P}(V \otimes_{\text{sym}} V) \\ &\leq C\, \alpha^{2r+\ell-k}\, h_n^{3d}\, \mathcal{P}(V \otimes_{\text{sym}} V). \end{aligned} \tag{27}$$
Distinguishing the cases $r = 0$ and $r \geq 1$ and using (26), (27), (iv) of Assumption 1.2 and (3), we get, for some $g_1, g \in F$,
$$\begin{aligned} |H_{6,n} - H_6^{[n]}| &= \sum_{0\leq k<\ell} |h_{k,\ell,0}^{(n)} - h_{k,\ell,0}|\, \mathbf{1}_{\{\ell<p\}} + \sum_{\substack{0\leq k<\ell\\ r\geq 1}} |h_{k,\ell,r}^{(n)} - h_{k,\ell,r}|\, \mathbf{1}_{\{r+\ell<p\}} \\ &\leq C \sum_{0\leq k<\ell} 2^{-k}\alpha^{p-\ell-1}\, \|\mathcal{P}(\mathcal{P}\tilde{f}_n \otimes_{\text{sym}} \mathcal{Q}^{\ell-k}\mathcal{P}\tilde{f}_n)\|_\infty\, V \\ &\quad + C\, h_n^{3d} \sum_{\substack{0\leq k<\ell\\ r\geq 1}} 2^{r-k}\, \alpha^{2r+\ell-k}\, \mathcal{Q}^{p-\ell-r-1}\mathcal{P}(V \otimes_{\text{sym}} V)\, \mathbf{1}_{\{r+\ell<p\}} \\ &\leq C\, h_n^{3d}\, (V + (\sum_{\substack{0\leq k<\ell\\ r\geq 1}} 2^{r-k}\, \alpha^{2r+\ell-k}\, \mathbf{1}_{\{r+\ell<p\}})\, g_1) \leq C\, h_n^{3d}\, a_n\, g, \end{aligned} \tag{28}$$
where
$$a_n = \begin{cases} 1 & \text{if } 2\alpha^2 < 1 \\ p & \text{if } 2\alpha = 1 \\ (2\alpha^2)^p & \text{if } 2\alpha^2 > 1. \end{cases}$$

Using (28), we find that

$$|V_6(n) - H_6^{[n]}| \leq |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(|H_6(n) - H_6^{[n]}|) \leq C\, a_n\, h_n^{3d}\, |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(g).$$

Using (6), (11) and that $g \in L^1(\mu)$, we get

$$\lim_{n\to+\infty} |V_6(n) - H_6^{[n]}| = 0 \quad \text{a.s. and in } L^2.$$

Next, as for (28), using (26), (27) and that $F \subset L^1(\mu)$, we find that $|H_6^{[n]}| \leq C\, a_n\, h_n^{3d}$. Using (11), we get $\lim_{n\to+\infty} H_6^{[n]} = 0$. Now, since we can write $V_6(n) = (V_6(n) - H_6^{[n]}) + H_6^{[n]}$, we conclude that $\lim_{n\to\infty} \mathbb{E}[V_6(n)^2] = 0$.

Next, we treat the term $V_5(n)$. We have $V_5(n) = (V_5(n) - H_5^{[n]}) + H_5^{[n]}$, where

$$H_5^{[n]} = \sum_{0\leq k<\ell} 2^{-k} \langle \mu, \mathcal{P}(\tilde{f}_n\, g_{k,\ell,n})\rangle \mathbf{1}_{\{\ell\leq p\}}.$$

Using Lemma 10.1, we get, for all $0 \leq k < \ell$,

$$|\mathcal{P}(\tilde{f}_n g_{k,\ell,n}) - \langle \mu, \mathcal{P}(\tilde{f}_n g_{k,\ell,n})\rangle| \leq\ C\, \|\mathcal{Q}^{\ell-k-1}\mathcal{P}\tilde{f}_n\|_\infty\, \|\mathcal{P}\tilde{f}_n\|_\infty \ \leq\ C\, h_n^d \mathbf{1}_{\{k=\ell-1\}} + C\, h_n^{2d}\, \mathbf{1}_{\{k\leq\ell-2\}}.$$

Using the latter inequality and distinguishing the cases $\ell = p$ and $\ell \leq p-1$, we find that

$$|V_5(n) - H_5^{[n]}| \ \leq C\, (2^{-p} h_n^d \ +\ h_n^{2d} \ +\ p\, h_n^{2d}).$$

This implies that $\lim_{n\to+\infty} |V_5(n) - H_5^{[n]}| = 0$ a.s. and in L^2.

Next, using Lemma 10.1, we have

$$|H_5^{[(n)]}| \leq \sum_{0\leq k<\ell} 2^{-k} |\langle \mu, \mathcal{P}(\tilde{f}_n\, g_{k,\ell,n})\rangle| \mathbf{1}_{\{\ell\leq p\}} \leq C \sum_{\ell>0} (2^{-\ell+1}\, h_n^{2d} + \sum_{k=0}^{\ell-2} 2^{-k}\, \alpha^{\ell-k}\, h_n^{3d/2}) \leq C\, h_n^{3d/2}.$$

This implies that $\lim_{n\to\infty} H_5^{[(n)]} = 0$ and then that $\lim_{n\to\infty} V_5(n) = 0$ a.s. and in L^2.

Finally, since $V_2(n) = V_5(n) + V_6(n)$, it follows from the foregoing that $\lim_{n\to\infty} \mathbb{E}[V_2(n)^2] = 0$ and this ends the proof. □

Now we treat the term $V_1(n)$. Recall

$$V_1(n) = |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{k=0}^{p} \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}(\tilde{f}_n)^2].$$

We have the following convergence.

Lemma 6.5. *Under the assumptions of Theorem 3.3, we have the following convergence:*

$$\lim_{n\to\infty} V_1(n) = 2\, \|K_0\|_2^6\, \mu^\Delta(x, x_0, x_1) \quad \textit{in probability.}$$

Proof. Let $k \in \{0, \ldots, p\}$ and $i \in \mathbb{G}_{n-p}$. Conditioning by $\mathcal{H}_{i,p-k}$ and using the branching Markov property, we get

$$\mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}(\tilde{f}_n)^2] \ =\ \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}(\mathcal{P}(\tilde{f}_n^2) - (\mathcal{P}\tilde{f}_n)^2)] \ +\ \mathbb{E}_{X_i}[(M_{\mathbb{G}_{p-k}}(\mathcal{P}\tilde{f}_n))^2].$$

Using the latter inequality and the fact that $\mathcal{P}(\tilde{f}_n^2) - (\mathcal{P}\tilde{f}_n)^2 = \mathcal{P}(f_n^2) - (\mathcal{P}f_n)^2$, we get

$$V_1(n) = V_3(n) + V_4(n) - V_7(n),$$

where

$$\begin{aligned}
V_3(n) &= |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{k=0}^{p} \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}(\mathcal{P}(f_n^2))];\\
V_7(n) &= |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{k=0}^{p} \mathbb{E}_{X_i}[M_{\mathbb{G}_{p-k}}((\mathcal{P}f_n)^2)];\\
V_4(n) &= |\mathbb{G}_n|^{-1} \sum_{i\in\mathbb{G}_{n-p}} \sum_{k=0}^{p} \mathbb{E}_{X_i}[(M_{\mathbb{G}_{p-k}}(\mathcal{P}\tilde{f}_n))^2].
\end{aligned}$$

First we treat $V_7(n)$. We set

$$H_{7,n} = \sum_{k=0}^{p} 2^{-k}\, \mathcal{Q}^{p-k}((\mathcal{P}f_n)^2).$$

Using (52), we have $V_7(n) = |\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(H_{7,n})$. Using Lemma 10.1 and distinguishing the cases $k = p$ and $k \leq p-1$, we get $|V_7(n)| \leq C\,(2^{-p}\, h_n^d + h_n^{2d})$. It then follows that $\lim_{n\to\infty} V_7(n) = 0$ in probability.

Next, we treat the term $V_3(n)$. We set $A_{3,n} = H_{3,n} - H_3^{[n]}$, with:

$$H_{3,n} = \sum_{k=0}^{p} 2^{-k}\, \mathcal{Q}^{p-k}(\mathcal{P}(f_n^2)) \quad \text{and} \quad H_3^{[n]} = \sum_{k=0}^{p} 2^{-k}\, \langle \mu, \mathcal{P}(f_n^2)\rangle = 2(1-2^{-p-1})\langle \mu, \mathcal{P}(f_n^2)\rangle. \tag{29}$$

We set $g_n = \mathcal{P}(f_n^2) - \langle \mu, \mathcal{P}(f_n^2)\rangle$. Using (5) and Lemma 10.1, we have

$$\begin{aligned}
|V_3(n) - H_3^{[n]}| &\leq |\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(2^{-p}\,|g_n|) + M_{\mathbb{G}_{n-p}}(\sum_{k=0}^{p-1} 2^{-k}|\mathcal{Q}^{p-k-1}(\mathcal{Q}(g_n))|)\\
&\leq C\,2^{-p} \,+\, C\,|\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(\sum_{k=0}^{p-1} 2^{-k}\, \alpha^{p-k}\, \|\mathcal{Q}(\mathcal{P}(f_n^2))\|_\infty\, V)\\
&\leq C\,2^{-p} \,+\, C\,a_n\,|\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(V), \qquad (30)
\end{aligned}$$

where

$$a_n = \begin{cases} 2^{-p} & \text{if } 2\alpha < 1\\ p\,2^{-p} & \text{if } 2\alpha = 1\\ \alpha^p & \text{if } 2\alpha > 1. \end{cases}$$

Using (6) and the fact that $\lim_{n\to+\infty} a_n = 0$, we find that

$$\lim_{n\to+\infty} C\,2^{-p} \,+\, C\,a_n\,|\mathbb{G}_{n-p}|^{-1} M_{\mathbb{G}_{n-p}}(V) = 0 \quad \text{a.s. and in } L^2.$$

From (30), this implies that

$$\lim_{n\to+\infty} |V_3(n) - H_3^{[n]}| = 0 \quad \text{in probability.}$$

Using Lemma 10.2, we get

$$\lim_{n\to\infty} H_3^{[n]} = \lim_{n\to\infty} 2\,(1-2^{-p-1})\,\langle \mu, \mathcal{P}(f_n^2) = 2\,\|K_0\|_2^6\, \mu^{\triangle}(x, x_0, x_1).$$

From the foregoing, we conclude that $\lim_{n\to\infty} V_3(n) = 2\,\|K_0\|_2^6\, \mu^{\triangle}(x, x_0, x_1)$ in probability.

Finally, we treat the term $V_4(n)$. Using (53), we have

$$V_4(n) \;=\; V_8(n) \;+\; V_9(n),$$

where

$$V_8(n) \;=\; |\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(H_{8,n}) \quad \text{and} \quad V_9(n) \;=\; |\mathbb{G}_{n-p}|^{-1}\, M_{\mathbb{G}_{n-p}}(H_{9,n}),$$

with

$$H_{8,n} = \sum_{k=0}^{p} 2^{-k} \mathcal{Q}^{p-k}((\mathcal{P}\tilde{f}_n)^2) \quad \text{and} \quad H_{9,n} = \sum_{k\geq 0, \ell \geq 0} 2^{-k+\ell} \mathcal{Q}^{p-k-\ell-1} \mathcal{P}(\mathcal{Q}^\ell \mathcal{P}\tilde{f}_n \otimes \mathcal{Q}^\ell \mathcal{P}\tilde{f}_n) \mathbf{1}_{\{k+\ell<p\}}.$$

Writing

$$H_8^{[n]} = \sum_{k=0}^{p} 2^{-k} \langle \mu, (\mathcal{P}\tilde{f}_n)^2 \rangle \quad \text{and} \quad H_9^{[n]} = \sum_{k\geq 0, \ell\geq 0} 2^{-k+\ell} \langle \mu, \mathcal{P}(\mathcal{Q}^\ell \mathcal{P}\tilde{f}_n \otimes \mathcal{Q}^\ell \mathcal{P}\tilde{f}_n)\rangle \mathbf{1}_{\{k+\ell<p\}},$$

we prove, as previously, that

$$\lim_{n\to\infty} |V_8(n) - H_8^{[n]}| = \lim_{n\to\infty} |V_9(n) - H_9^{[n]}| = 0 \quad \text{a.s. and in } L^2; \quad \lim_{n\to\infty} H_8^{[n]} = \lim_{n\to\infty} H_9^{[n]} = 0.$$

As a result, we find that

$$\lim_{n\to\infty} |V_8(n)| = \lim_{n\to\infty} |V_9(n)| = 0 \quad \text{in probability.}$$

Since $V_4(n) = V_8(n) + V_9(n)$, we conclude that $\lim_{n\to\infty} V_4(n) = 0$ in probability. Finally, since $V_1(n) = V_3(n) + V_4(n) - V_7(n)$, the result of the Lemma follows from the foregoing. $\square$

As a consequence of (24), Lemmas 6.3, 6.4 and 6.5, we have the following result.

Lemma 6.6. *Under the assumptions of Theorem 3.3, we have the following convergence:*

$$\lim_{n\to\infty} V(n) = 2\, \|K_0\|_2^6\, \mu^\Delta(x, x_0, x_1) \quad \textit{in probability.}$$

We now check the Lindeberg condition using a fourth moment condition. We set:

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

Lemma 6.7. *Under the assumptions of Theorem 3.3, we have that* $\lim_{n\to\infty} R_3(n) = 0$

Proof. We have

$$\begin{aligned} R_3(n) &\leq 16\,(p+1)^3 |\mathbb{G}_n|^{-2} \sum_{i\in\mathbb{G}_{n-p}} \sum_{\ell=0}^{p} \mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(\tilde{f}_n))^4] \\ &\leq 128\,(p+1)^3\, |\mathbb{G}_n|^{-2} \sum_{i\in\mathbb{G}_{n-p}} \sum_{\ell=0}^{p} \mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(f_n - \mathcal{P}f_n))^4] \\ &\qquad + \, 128\,(p+1)^3\, |\mathbb{G}_n|^{-2} \sum_{i\in\mathbb{G}_{n-p}} \sum_{\ell=0}^{p} \mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(\mathcal{P}\tilde{f}_n))^4], \end{aligned} \tag{31}$$

where we used that $(\sum_{k=0}^r a_k)^4 \leq (r+1)^3 \sum_{k=0}^r a_k^4$ for the two inequalities (resp. with $r = 1$ and $r = p$), Jensen inequality for the first inequality and the decomposition $f_n = (f_n - \mathcal{P}f_n) + \mathcal{P}f_n$

for the last inequality. For the second term of the right hand side of (31), we follow the proof of Lemma 5.6 in [4] and Lemma 4.7 in [3] to find that

$$n^3\,|\mathbb{G}_n|^{-2}\sum_{i\in\mathbb{G}_{n-p}}\sum_{\ell=0}^{p}\mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(\mathcal{P}\tilde{f}_n))^4]\leq C\,n^5(2^{-n+p}h_n\mathbf{1}_{\{2\alpha^2\leq 1\}}+2^{-n+p}\,(2\alpha^2)^{2p}h_n^{6d}\mathbf{1}_{\{2\alpha^2>1\}}),$$

and using (11), this implies that

$$\lim_{n\to+\infty}n^3\,|\mathbb{G}_n|^{-2}\sum_{i\in\mathbb{G}_{n-p}}\sum_{\ell=0}^{p}\mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(\mathcal{P}\tilde{f}_n))^4]=0. \tag{32}$$

We are now going to treat the first term of (31). Since $\mathcal{P}(f_n-\mathcal{P}(f_n))=0$, we have, (see Remark 2.3 in [5] for more details),

$$\mathbb{E}_x[(M_{\mathbb{G}_{p-\ell}}(f_n-\mathcal{P}f_n))^4]\leq\; g_{n,\ell}(x)\;+6\,h_{n,\ell}(x), \tag{33}$$

with:

$$g_{n,\ell}(x)=\mathbb{E}_x[M_{\mathbb{G}_{p-\ell}}(\mathcal{P}((f_n-\mathcal{P}f_n)^4))]\quad\text{and}\quad h_{n,\ell}(x)=\mathbb{E}_x[(M_{\mathbb{G}_{p-\ell}}(\mathcal{P}((f_n-\mathcal{P}f_n)^2)))^2].$$

We set

$$R_{3,1}(n)=(p+1)^3\,2^{-2n}\sum_{i\in\mathbb{G}_{n-p}}\sum_{\ell=0}^{p}\mathbb{E}[(M_{i\mathbb{G}_{p-\ell}}(f_n-\mathcal{P}f_n))^4].$$

Using the branching Markov property and (33) for the first inequality and (52) for equality, we get

$$\begin{aligned}R_{3,1}(n)&\leq C\,n^3\,2^{-2n}\sum_{\ell=0}^{p}\mathbb{E}[M_{\mathbb{G}_{n-p}}(g_{n,\ell})]+C\,n^3\,2^{-2n}\sum_{\ell=0}^{p}\mathbb{E}[M_{\mathbb{G}_{n-p}}(h_{n,\ell})]\\
&=\;C\,n^3\,2^{-n-p}\sum_{\ell=0}^{p}\langle\nu,\mathcal{Q}^{n-p}g_{n,\ell}\rangle\;+\;C\,n^3\,2^{-n-p}\sum_{\ell=0}^{p}\langle\nu,\mathcal{Q}^{n-p}h_{n,\ell}\rangle.\end{aligned} \tag{34}$$

Using Lemma 10.1, we get

$$\langle\nu,\mathcal{Q}^{n-p}g_{n,\ell}\rangle\;=\;2^{p-\ell}\,\langle\nu,\mathcal{Q}^{n-\ell}(\mathcal{P}((f_n-\mathcal{P}f_n)^4))\rangle\;\leq\;2^{p-\ell}\,\|\mathcal{Q}\mathcal{P}((f_n-\mathcal{P}f_n)^4)\|_\infty\leq\;C\,h_n^{-3d}\,2^{p-\ell}.$$

The latter inequality implies that

$$n^3\,2^{-n-p}\sum_{\ell=0}^{p}\langle\nu,\mathcal{Q}^{n-p}g_{n,\ell}\rangle\leq C\,n^3\,(2^nh_n^{3d})^{-1}. \tag{35}$$

Using (53), the fact that $\mathcal{P}((f_n-\mathcal{P}f_n)^2)\leq\mathcal{P}(f_n^2)$ for the first inequality and Lemma 10.1 for the second inequality, we get

$$\begin{aligned}\langle\nu,\mathcal{Q}^{n-p}h_{n,\ell}\rangle&\leq 2^{p-\ell}\,\langle\nu,\mathcal{Q}^{n-\ell}(\mathcal{P}(f_n^2))^2\rangle+\sum_{k=0}^{p-\ell-1}2^{p-\ell+k}\langle\nu,\mathcal{Q}^{n-\ell-k-1}\mathcal{P}(\mathcal{Q}^k\mathcal{P}(f_n^2\otimes^2))\rangle\\
&\leq\;C\,2^{p-\ell}\;+\;C\,2^{2(p-\ell)}.\end{aligned}$$

The latter inequality implies that

$$n^3\,2^{-n-p}\sum_{\ell=0}^{p}\langle\nu,\mathcal{Q}^{n-p}h_{n,\ell}\rangle\;\leq\;C\,n^3(2^{-n}\;+\;2^{-n+p}). \tag{36}$$

From (34), (35) and (36), we conclude that $\lim_{n\to\infty}R_{3,1}(n)=0$. Finally, from (31) and (32), this proves that $\lim_{n\to\infty}R_3(n)=0$. □

We can now use Theorem 3.2 and Corollary 3.1, p. 58, and the remark p. 59 from [10] to deduce from Lemmas 6.6 and 6.7 that $\Delta_n(f_n)$ converges in distribution towards a Gaussian real-valued random variable with deterministic variance σ^2. Using (20) and Lemmas 6.1 and 6.2, we then deduce Theorem 3.3 for $A_n = \mathbb{T}_n$.

For $A_n = \mathbb{G}_n$, we have

$$N_{n,\emptyset}(f_n) = R_1(n) + \Delta_n(f_n),$$

where

$$R_1(n) = \sum_{i \in \mathbb{G}_{n-p}} \mathbb{E}[N_{n,i}(f_n)|\mathcal{F}_i] \quad \text{and} \quad \Delta_n(f_n) = \sum_{i \in \mathbb{G}_{n-p}} \Delta_{n,i}(f_n),$$

and for all $i \in \mathbb{G}_{n-p}$,

$$N_{n,i}(f_n) = M_{i\mathbb{G}_p}(\tilde{f}_n) \quad \text{and} \quad \Delta_{n,i}(f_n) = N_{n,i}(f_n) - \mathbb{E}[N_{n,i}(f_n)|\mathcal{F}_i].$$

Following exactly the proof of Lemma 6.2, 6.6 and 6.7, we get the result for this case. We note that for $\mathbb{A}_n = \mathbb{G}_n$, the factor 2 is missing in the asymptotic variance. This come from the fact that here, $H_3^{[n]}$ defined in (29) is simply equal to $\langle \mu, \mathcal{P}(f_n^2) \rangle$.

7. Proof of Theorem 3.5

We begin the proof with $\mathbb{A}_n = \mathbb{T}_n$. From (10), we have

$$|\mathbb{T}_n|^{1/2}\, h^{3d/2}\, (\mu^{\Delta}_{\mathbb{T}_n}(xx_0x_1) - \mu^{\Delta}(xx_0x_1)) = (|\mathbb{G}_n|/|\mathbb{T}_n|)^{1/2}\, N_{n,\emptyset}(f_n) \,+\, B_n(xx_0x_1),$$

where the bias term $B_n(xx_0x_1)$ is defined by

$$B_n(xx_0x_1) = |\mathbb{T}_n|^{1/2}\, h_n^{3d/2}\, (h^{-3/2}\, \langle \mu^{\Delta}, f_n \rangle - \mu^{\Delta}(xx_0x_1)).$$

Since $\lim_{n\to\infty}(|\mathbb{G}_n|/|\mathbb{T}_n|)^{1/2} = 1/\sqrt{2}$, from Theorem 3.3, it suffices, to obtain the result of Theorem 3.5, to prove that $\lim_{n\to\infty} B_n(xx_0x_1) = 0$. Using the Taylor expansion and Assumption 3.4, one can prove that (see [1] for more details)

$$B_n(xx_0x_1) \,\leq\, C\, \sqrt{|\mathbb{T}_n| h_n^{2s+3d}}.$$

Since $\lim_{n\to\infty} |\mathbb{T}_n| h_n^{2s+3d} = 0$, we conclude that $\lim_{n\to\infty} B_n(xx_0x_1) = 0$ and this ends the proof for $\mathbb{A}_n = \mathbb{T}_n$.

For $\mathbb{A}_n = \mathbb{G}_n$ the proof follows exactly the same lines.

8. Proof of Theorem 4.2

First of all, we have the following decomposition:

$$\begin{aligned}\sqrt{|\mathbb{A}_n| h_n^{3d}}(\widehat{\mathcal{P}}_{\mathbb{A}_n}(xx_0x_1) - \mathcal{P}(xx_0x_1)) = (|\mathbb{A}_n| h_n^{3d})^{1/2}\, &(\frac{\widehat{\mu}^{\Delta}_{\mathbb{A}_n}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)} - \frac{\mu^{\Delta}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)}) \\ &- \frac{\mu^{\Delta}(xx_0x_1)}{\mu(x)\widehat{\mu}_{\mathbb{A}_n}(x)}\, (|\mathbb{A}_n| h_n^{3d})^{1/2}\, (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)).\end{aligned}$$

Then, the proof of Theorem 4.2 is a direct consequence of the previous decomposition and Lemmas 8.1 and 8.2 below.

Lemma 8.1. *Under Assumptions of Theorem 4.2, we have*

$$\lim_{n\to\infty} \frac{\mu^{\Delta}(xx_0x_1)}{\mu(x)\widehat{\mu}_{\mathbb{A}_n}(x)}\, (|\mathbb{A}_n| h_n^{3d})^{1/2}\, (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)) = 0 \quad \textit{in probability.}$$

Proof. We consider the function g_n defined on S by: $g_n(y) = h_n^{-d/2} K_0(h_n^{-1}(x-y))$ for all $y \in S$. We begin the proof with $\mathbb{A}_n = \mathbb{T}_n$. We set $\tilde{g}_n = g_n - \langle \mu, g_n \rangle$. We have the following decomposition:

$$\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x) = |\mathbb{T}_n|^{-1} h_n^{-d/2} (\sum_{\ell=0}^{2} M_{\mathbb{G}_\ell}(\tilde{g}_n) + \sum_{\ell=3}^{n} M_{\mathbb{G}_\ell}(\tilde{g}_n)) + \langle \mu, h_n^{-d/2} g_n \rangle - \mu(x). \tag{37}$$

Using the fact that K_0 is bounded, integration by parts and Assumption 1.8, we have the following upper bounds:

$$\|g_n\|_\infty \leq \|K_0\|_\infty \, h_n^{-d/2}; \quad \|\mathcal{Q} g_n\|_\infty + |\langle \mu, g_n \rangle| \leq 2\, C_0 \, \|K_0\|_1 \, h_n^{d/2}. \tag{38}$$

Using (38), we find that

$$|\mathbb{T}_n|^{-1} h_n^{-d/2} |\sum_{\ell=0}^{2} M_{\mathbb{G}_\ell}(\tilde{g}_n)| \leq C\, (|\mathbb{T}_n| \, h_n^d)^{-1}. \tag{39}$$

Next, from Minkowski's inequality, we have

$$\mathbb{E}[(\sum_{\ell=3}^{n} M_{\mathbb{G}_\ell})^2] \leq (\sum_{\ell=3}^{n} (\mathbb{E}[(M_{\mathbb{G}_\ell}(\tilde{g}_n))^2])^{1/2})^2.$$

Using (53), (5) and (38), we get:

$$\begin{aligned} \mathbb{E}[(M_{\mathbb{G}_\ell}(\tilde{g}_n))^2] &\leq C\, 2^\ell \, \langle \nu, \mathcal{Q}^\ell(\tilde{g}_n^2) \rangle \; + \sum_{r=0}^{\ell-1} 2^{\ell+r} \langle \nu, \mathcal{Q}^{\ell-r-1}(\mathcal{P}(|\mathcal{Q}^r \tilde{g}_n| \otimes^2)) \rangle \\ &\leq \; C\, 2^\ell \; + \; C\, h_n^d \sum_{r=0}^{\ell-1} 2^\ell \, (2\alpha^2)^r \; \leq \; C\, 2^\ell \, \mathbf{1}_{\{2\alpha^2 \leq 1\}} \; + \; C\, 2^\ell (1 + (2\alpha^2)^\ell \, h_n^d) \mathbf{1}_{\{2\alpha^2 > 1\}}. \end{aligned}$$

The latter inequality implies that

$$\mathbb{E}[(|\mathbb{T}_n|^{-1} h_n^{-d/2} \sum_{\ell=3}^{n} M_{\mathbb{G}_\ell}(\tilde{g}_n))^2] \leq C\, (|\mathbb{T}_n| h_n^{3d})^{-1} \, (h_n^{2d} + (2\alpha^2)^n h_n^{3d} \, \mathbf{1}_{\{2\alpha^2 > 1\}}). \tag{40}$$

Using (39), (40) and (11), we deduce that

$$\lim_{n \to +\infty} \mathbb{E}[(|\mathbb{T}_n|^{-1} h_n^{-d/2} \sum_{\ell=0}^{n} M_{\mathbb{G}_\ell}(\tilde{g}_n))^2] = 0. \tag{41}$$

Next, using Taylor expansion and Assumption 3.4, we get (see [1] for more details)

$$|\langle \mu, h_n^{-d/2} g_n \rangle - \mu(x)| \leq C\, h_n^s. \tag{42}$$

From (37), (41) and (42), we deduce that

$$\lim_{n \to \infty} \widehat{\mu}_{\mathbb{T}_n}(x) = \mu(x) \quad \text{in probability.} \tag{43}$$

We further deduce that

$$\mathbb{E}[|\mathbb{T}_n| h_n^{3d} (\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x))^2] \; \leq \; C\, (h_n^{2d} \; + \; (2\alpha^2)^n h_n^{3d} \mathbf{1}_{\{2\alpha^2 > 1\}} \; + \; |\mathbb{T}_n| h_n^{3d+2s}).$$

Using (11), the latter inequality implies that

$$\lim_{n \to \infty} (|\mathbb{T}_n| h_n^{3d})^{1/2} \, (\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x)) = 0 \quad \text{in probability.} \tag{44}$$

From (43), (44) and using Slutsky's Lemma, we get

$$\lim_{n \to \infty} \frac{\mu^\Delta(x x_0 x_1)}{\mu(x) \widehat{\mu}_{\mathbb{T}_n}(x)} \, (|\mathbb{T}_n| h_n^{3d})^{1/2} \, (\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x)) = 0 \quad \text{in probability.}$$

For $\mathbb{A}_n = \mathbb{G}_n$, we follows exactly the same lines and this ends the proof. $\square$

Lemma 8.2. *Under Assumptions of Theorem 4.2, we have*

$$(|\mathbb{A}_n| h_n^{3d})^{1/2} \left(\frac{\widehat{\mu}^{\triangle}_{\mathbb{A}_n}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)} - \frac{\mu^{\triangle}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)}\right) \xrightarrow[n\to\infty]{\text{(d)}} G,$$

where G is a centered Gaussian real-valued random variable with mean 0 and variance

$$\sigma^2 \;=\; \|K_0\|_2^6\, \mathcal{P}(x, x_0, x_1)/\mu(x).$$

Proof. This is a direct consequence of Theorem 3.5, (43) and Slutsky's Lemma. $\square$

9. Proof of (16)

We set $f_{0h}(y) = h^{-1}K_0(h^{-1}(x-y))$ and recall $\langle \mu, f_{0h}\rangle = \int_{\mathbb{R}} f_{0h}(y)\mu(y)dy$. Using the decomposition

$$\widehat{\mu}(x) - \mu(x) = \frac{1}{|\mathbb{G}_n|}\sum_{u\in\mathbb{G}_n} \tilde{f}_{0h}(X_u) \,+ \langle \mu, f_{0h}\rangle \,-\, \mu(x),$$

we obtain the following biais-variance type decomposition.

$$\mathbb{E}\left[(\widehat{\mu}(x) - \mu(x))^2\right] \leq 2(|\mathbb{G}_n|)^2 \mathbb{E}\big[(\sum_{u\in\mathbb{G}_n} \tilde{f}_{0h}(X_u))^2\big] + 2\big(\langle \mu, f_{0h}\rangle \,-\, \mu(x)\big)^2. \tag{45}$$

Using (53) and the fact that the process $\mathcal{L}(X_\emptyset) = \mu$ (which implies that $\mu\mathcal{Q} = \mu$), we get

$$|\mathbb{G}_n|^{-2}\, \mathbb{E}\big[(\sum_{u\in\mathbb{G}_n} \tilde{f}_{0h}(X_u))^2\big] \;\leq\; \frac{2\,\langle \mu, \tilde{f}_{0h}^2\rangle}{|\mathbb{G}_n|} \;+\; \frac{1}{|\mathbb{G}_n|}\sum_{k=1}^{n-1} 2^k \langle \mu, (\mathcal{Q}^{k-1}(\mathcal{Q}\tilde{f}_{0h}))^2\rangle. \tag{46}$$

We now plan to use (15) with $f = \mathcal{Q}f_{0h}$. For all $y \in \mathbb{R}$, we get, after the change of variable $t = h^{-1}(x-z)$ and the use of the first-order Taylor's expansion,

$$\begin{aligned}
(\mathcal{Q}f_{0h})'(y) &= \frac{1}{h}\int_{\mathbb{R}} K_0(h^{-1}(x-z))\frac{\partial \mathcal{Q}}{\partial y}(y,z)dz \\
&= \int_{\mathbb{R}} K_0(z)\frac{\partial \mathcal{Q}}{\partial y}(y, x \,-\, hz)\,dz \\
&= \frac{\partial \mathcal{Q}}{\partial y}(y,x)\int_{\mathbb{R}} K_0(z)\,dz + o(1) = \frac{\partial \mathcal{Q}}{\partial y}(y,x) \,+\, o(1).
\end{aligned}$$

We then have that

$$\|(\mathcal{Q}f_{0h})'\|_\infty = \sup_{y\in\mathbb{R}}\left\{\frac{\partial \mathcal{Q}}{\partial y}(\cdot, x)\right\} \,+\, o(1) \;=\; \frac{1}{\sqrt{2\pi}\sigma^2}\,\mathrm{e}^{-1/2} + o(1).$$

Using the latter equality and (15), we get, for all $k \geq 1$,

$$|\mathcal{Q}^{k-1}(\mathcal{Q}\tilde{f}_{0h})|(y) \leq \frac{1}{\sqrt{2\pi}\sigma^2\,\mathrm{e}^{1/2}(1-a)}(\sigma(1+a)^{-1} + |y|)a^{k-1} + o(1). \tag{47}$$

Recall μ is the Gaussian law $\mathcal{N}(0, \sigma^2(1-a^2))$. Using (47) and $(\sigma(1+a)^{-1} + |y|)^2 \leq 2\,\sigma^2(1+a)^{-2} + 2\,y^2$, we get

$$\langle \mu, (\mathcal{Q}^{k-1}(\mathcal{Q}\tilde{f}_{0h}))^2\rangle \leq \frac{1}{e\pi\sigma^2}\Big(\frac{1}{(1-a)^2} + \frac{1}{1-a^2}\Big)a^{2(k-1)} \,+\, o(1) \,\leq\, \frac{C_{a,\sigma}}{a^2}\,a^{2k} \,+\, o(1). \tag{48}$$

Now, (48) and (46) implies that

$$(|\mathbb{G}_n|)^2\mathbb{E}\big[(\sum_{u\in\mathbb{G}_n}\tilde{f}_{0h}(X_u))^2\big] \;\leq\; \frac{2\,\langle\mu,\tilde{f}_{0h}^2\rangle}{|\mathbb{G}_n|} \;+\; \frac{M_{a,\sigma}}{a^2|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2a^2)^k \;+\; \mathcal{O}(1).$$

Putting the latter inequality into (45), we obtain

$$\mathbb{E}\left[(\widehat{\mu}(x)-\mu(x))^2\right] \leq \frac{2M_{a,\sigma}}{a^2|\mathbb{G}_n|}\sum_{k=1}^{n-1}(2a^2)^k \;+\; \frac{4\,\langle\mu,\tilde{f}_{0h}^2\rangle}{|\mathbb{G}_n|} \;+\; 2\left(\langle\mu,f_{0h}\rangle \;-\; \mu(x)\right)^2 \;+\; \mathcal{O}(1). \tag{49}$$

Finally, it is very standard to get asymptotic equivalence of the second and the third term of the right hand side of (49) (see for e.g. [12], Section 3.3.1 for more details). This ends the proof of (16).

10. Appendix

First, we give some useful upper bounds. We recall that $S=\mathbb{R}^d$. Recall f_n defined in (9).

Lemma 10.1. *Under Assumption* (1.8)*, we have:*

$$\begin{aligned}
&\|\mathcal{P}|f_n|\|_\infty \leq \|K_0\|_1^2\,\|K_0\|_\infty\,\|\mathcal{P}\|_\infty\,h_n^{d/2};\\
&\|\mathcal{Q}\mathcal{P}|f_n|\|_\infty \leq \|\mathcal{P}\|_\infty\,\|\mathcal{Q}\|_\infty\,\|K_0\|_1^3\,h_n^{3d/2};\\
&|\langle\mu,\mathcal{P}f_n\rangle| \leq \|\mathcal{P}\|_\infty\,\|\mu\|_\infty\,\|K_0\|_1^3\,h_n^{3d/2}\\
&\langle\mu,\mathcal{P}(f_n^2)\rangle \leq \|\mu\|_\infty\,\|\mathcal{P}\|_\infty\,\|K_0\|_2^6\\
&\|\mathcal{Q}\mathcal{P}f_n^2\|_\infty \leq \|K_0\|_2^6\,\|\mathcal{P}\|_\infty\,\|\mathcal{Q}\|_\infty;\\
&\|\mathcal{P}(\mathcal{P}f_n\otimes\mathcal{P}f_n)\|_\infty \leq \|\mathcal{P}\|_\infty^3\,\|K_0\|_1^6\,h_n^{3d};\\
&\|\mathcal{Q}((\mathcal{P}f_n)^2)\|_\infty \leq \|K_0\|_2^2\,\|K_0\|_1^4\,\|\mathcal{Q}\|_\infty\,\|\mathcal{P}\|_\infty^2\,h^{2d};\\
&\|\mathcal{P}((\mathcal{P}f_n)^2\otimes(\mathcal{P}f_n)^2)\|_\infty \leq \|K_0\|_2^4\,\|K_0\|_1^8\,\|\mathcal{P}\|_\infty^5\,h^{4d};\\
&\|\mathcal{Q}(\mathcal{P}(f_n^4))\|_\infty \leq \|\mathcal{P}\|_\infty\,\|\mathcal{Q}\|_\infty\,\|K_0\|_4^{12}\,h_n^{-3d}.
\end{aligned}$$

Proof. Using a change of variables, we have, for all $y\in S$:

$$\mathcal{P}|f_n|(y) = h^{d/2}\,|K_0(h^{-1}(x-y))|\int_{S^2}|K_0|(y_0)\,|K_0|(y_1)\mathcal{P}(y,x_0-h\,y_0,x_1-h\,y_1)dy_0\,dy_1. \tag{50}$$

This implies that

$$\|\mathcal{P}f_n\|_\infty \leq \|K_0\|_1^2\,\|K\|_\infty\,\|\mathcal{P}\|_\infty\,h_n^{d/2}.$$

From (50) and using again a change of variables, we have, for all $t\in S$:

$$\mathcal{Q}\mathcal{P}|f_n|(t) = h_n^{3/2}\int_{S^3}|K_0|(y)\,|K_0|(y_0)\,|K_0|(y_1)\,\mathcal{P}(x-h\,y,x_0-h\,y_0,x_1-h\,y_1)\,\mathcal{Q}(t,x-h\,y)dy\,dy_0\,dy_1.$$

This implies that

$$\|\mathcal{Q}\mathcal{P}|f_n|\|_\infty \leq \|\mathcal{P}\|_\infty\,\|\mathcal{Q}\|_\infty\,\|K_0\|_1^3\,h_n^{3d/2}. \tag{51}$$

As for (51), we have

$$|\langle\mu,\mathcal{P}f_n\rangle| \leq \|\mathcal{P}\|_\infty\,\|\mu\|_\infty\,\|K_0\|_1^3\,h_n^{3d/2}.$$

Now, following the same ideas, we easily get the others upper bounds. □

We recall the following result due to Bochner (see [11, Theorem 1A] which can be easily extended to any dimension $d\geq 1$).

Lemma 10.2. *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* $\int_{\mathbb{R}^d} |g(x)|dx < +\infty$*. Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a measurable function such that* $\| f \|_\infty < +\infty$*,* $\int_{\mathbb{R}^d} |f(y)|\, dy < +\infty$ *and* $\lim_{|x|\to+\infty} |x|f(x) = 0$*. Define*

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

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

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

In this section, we recall useful results on BMC from Bitseki-Delmas [1].

Lemma 10.3. *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:*

$$\mathbb{E}_x\left[M_{\mathbb{G}_n}(f)\right] = |\mathbb{G}_n|\, \mathcal{Q}^n f(x) = 2^n\, \mathcal{Q}^n f(x), \tag{52}$$

$$\mathbb{E}_x\left[M_{\mathbb{G}_n}(f)^2\right] = 2^n\, \mathcal{Q}^n(f^2)(x) + \sum_{k=0}^{n-1} 2^{n+k}\, \mathcal{Q}^{n-k-1}\left(\mathcal{P}\left(\mathcal{Q}^k f \otimes \mathcal{Q}^k f\right)\right)(x), \tag{53}$$

$$\begin{aligned}\mathbb{E}_x\left[M_{\mathbb{G}_n}(f)M_{\mathbb{G}_m}(g)\right] = {} & 2^n \mathcal{Q}^m \left(g\mathcal{Q}^{n-m} f\right)(x) \\ & + \sum_{k=0}^{m-1} 2^{n+k}\, \mathcal{Q}^{m-k-1}\left(\mathcal{P}\left(\mathcal{Q}^k g \otimes_{\mathrm{sym}} \mathcal{Q}^{n-m+k} f\right)\right)(x).\end{aligned} \tag{54}$$

References

[1] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for bifurcating markov chains, 2020.

[2] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for bifurcating Markov chains under L^2 ergodic conditions. *Advances in Applied Probability*, pages 1–33, 2022.

[3] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for bifurcating markov chains under pointwise ergodic conditions. *The Annals of Applied Probability*, 32(5):3817–3849, 2022.

[4] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for kernel estimator of invariant density in bifurcating markov chains models. *Journal of Theoretical Probability*, pages 1–35, 2022.

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

[6] F. Comte and N. Marie. On a nadaraya-watson estimator with two bandwidths. *Electronic Journal of Statistics*, 15(1):2566–2607, 2021.

[7] M. Duflo. *Random iterative models*, volume 34. Springer Science & Business Media, 2013.

[8] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.*, 17(5-6):1538–1569, 2007.

[9] B. M. Gyori and D. Paulin. Hypothesis testing for markov chain monte carlo. *Statistics and Computing*, 26(6):1281–1292, 2016.

[10] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

[11] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[12] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26. CRC Press, 1986.

S. Valère Bitseki Penda, IMB, CNRS-UMR 5584, Université Bourgogne Franche-Comté, 9 avenue Alain Savary, 21078 Dijon Cedex, France.

E-mail address: `simeon-valere.bitseki-penda@u-bourgogne.fr`