A perturbation analysis of some Markov chains models with time-varying parameters

Lionel Truquet *†

Abstract

For some families of $V-$geometrically ergodic Markov kernels indexed by a parameter, we study the existence of a Taylor expansion of the invariant distribution in the space of signed measures. Our approach, which completes some previous results for the perturbation analysis of Markov chains, is motivated by a problem in statistics: a control of the bias for the nonparametric kernel estimation in some locally stationary Markov models. We illustrate our results with a nonlinear autoregressive process and a Galton-Watson process with immigration and time-varying parameters.

1 Introduction

The aim of this paper is to study some regularity properties of the local invariant measure in some locally stationary Markov models introduced in the literature. The notion of local stationarity has been introduced in Dahlhaus (1997) and offers a very interesting approach for the modelling of nonstationary time series for which the parameters are continuously changing with the time. The basic idea of local stationarity is to work with triangular arrays of random variables that can be locally approximated in some sense by stationary processes with the use of an infill asymptotic. In the literature, several stationary models have been extended to a locally stationary version, in particular some Markov models defined by autoregressive processes. See for instance Subba Rao (2006) Moulines et al. (2005) and Zhang and Wu (2012) for linear autoregressive processes, Dahlhaus and Rao (2006), Fryzlewicz et al. (2008) and Truquet (2016) for ARCH processes, Vogt (2012) and a recent contribution of Dahlhaus et al. (2017) for nonlinear autoregressive processes. In Truquet (2017), we introduced a notion of local stationarity for general Markov chains models, including most of the autoregressive processes introduced in the references given above but also finite-state Markov chains or integer-valued time series. To define these models, we use time-varying Markov kernels. Let $\{Q_u : u \in [0,1]\}$ be a family of ergodic Markov kernels on the same topological space $(E, \mathcal{E})$. We denote by π_u the invariant probability measure of Q_u. For an integer $n \geq 1$, we consider n random variables $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ taking values in G and such that

$$\mathbb{P}\left(X_{n,k} \in A | X_{n,k-1} = x\right) = Q_{k/n}(x, A), \quad (x, A) \in G \times \mathcal{B}(G), \quad 1 \leq k \leq n,$$

with the convention $X_{n,0} \sim \pi_0$.

In Truquet (2017), we extensively used some contraction properties and some Lipschitz continuity assumptions for the Markov kernels Q_u (or some of their iterations) to control the local approximation of $\pi_k^{(n)}$

*UMR 6625 CNRS IRMAR, University of Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France and
†ENSAI, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr*.

(the distribution of $X_{n,k}$) by π_u and in the same spirit the approximation of any finite dimensional distribution of the Markov chain with time-varying parameters. Various metric on the space of probability measures were discussed for controlling this approximation.

One of the important issues in the statistical inference of locally stationary processes is the curve estimation of some parameters of the kernels $\{Q_u : u \in [0,1]\}$ but also some parameters of the joint distribution, i.e $\int f d\pi_u$ for some functionals f. For instance, one can be interested in the estimation of the trend of a time series which is here the mean of the invariant probability. However, the nonparametric estimation of such functionals require to know their regularity, for instance the number of derivatives. For example, the local linear fit with a kernel density requires the existence of two derivatives for the function $u \mapsto \int f d\pi_u$. See Fan and Gijbels (1996) for an introduction to local polynomial modelling and Section 3.2 for the application to locally stationary Markov chains.

In the recent work of Dahlhaus et al. (2017), the authors study some autoregressive Markov processes with time-varying parameters. Taking benefit of the properties of some iterated random maps that define the time-varying process under consideration, they study some regularity properties of the stationary approximations through the notion of a derivative process. The regularity of some smooth functionals of the invariant measure is then deduced from this derivative process by using an adapted differential calculus. However, when the state space is discrete, the notion of derivative process is not relevant to evaluate such a regularity. This is in particular the case for some models we introduced in Truquet (2017).

In this paper, we study directly the regularity of the invariant measure $u \mapsto \pi_u$ when some regularity assumptions are available in some sense for $u \mapsto Q_u$. This approach has two benefits. First, it does not depend on the state space of the Markov process of interest and can be used for lots of locally stationary Markov processes. Secondly, with respect to Dahlhaus et al. (2017), one can get some regularity properties of some curves $u \mapsto \int f d\pi_u$ for non-smooth functionals f because our results are stated in total variation type norms. Of course, for the autoregressive processes studied in Dahlhaus (1997), the price to pay for this generality is to impose stronger regularity assumptions on the noise distribution.

In the literature of Markov chains, controlling the approximation of the invariant measure when the Markov kernel is perturbed has been extensively studied in the literature. A recent contribution to this theory is given in Rudolf and Schweizer (2017) with an application to stochastic algorithms. A few works also consider the problem of finding condition under which the invariant probability has some differentiability properties. See for instance Schweitzer (1968), Kartashov (1986) or Heidergott and Hordijk (2003). However most of the references devoted to this problem assume a continuity property of the Markov kernels with respect to the operator norm associated to some $V-$norms. As pointed out in Ferré et al. (2013), this assumption is too restrictive because it is not satisfied for the simple case of an autoregressive process of order 1. Let us precise this problem after introducing some notations. For a measurable function $V : E \to [1,\infty)$, we denote by $\mathcal{M}_V$ the set of signed measures μ on $(E, \mathcal{E})$ such that

$$\|\mu\|_V := |\mu| \cdot V = \int V d\mu < \infty.$$

We recall that $(\mathcal{M}_V, \|\cdot\|_V)$ is a Banach space. In this paper, we will mainly consider the Markov kernel Q_u as an operator T_u acting on $\mathcal{M}_V$, i.e $T_u \mu = \mu Q_u$ is the measure defined by

$$\mu Q_u(A) = \int \mu(dx) Q_u(x, A), \quad A \in \mathcal{E}.$$

Another convention in the literature is to consider Q_u as an operator acting on some functions spaces, i.e for a measurable function $g : G \to \mathbb{R}$ such that $|g|_V = \sup_{x \in E} \frac{|g(x)|}{V(x)} < \infty$, we set $Q_u g(x) = \int Q_u(x, dy) g(y)$. Then the operator T_u is often considered as the adjoint operator. Since we choose to work with measure spaces, we keep our convention. The corresponding $V-$operator norm of T_u is defined by

$$\|T_u\|_{V,V} = \sup_{\mu \in \mathcal{M}(V): \|\mu\|_V \leq 1} \|\mu Q_u\|_V = \sup_{|f|_V \leq 1} |Q_u f|_V .$$

However, this norm is often not appropriate. For instance, for a stationary AR(1) process $X_k = \alpha X_{k-1} + \xi_k$ with an integrable noise ξ_k having an absolutely continuous distribution with density f_ξ, Ferré et al. (2013) have shown that the corresponding Markov kernel $P_\alpha(x, dy) = f_\xi(y - \alpha x)dy$, is not continuous with respect to the lag coefficient α for this operator norm when $V(x) = 1 + |x|$. On the other hand, it is shown that a continuity property holds for the norm

$$\|P_\alpha\|_{V,1} = \sup_{|f|_1 \leq 1} |P_\alpha|_V = \sup_{\|\mu\|_V \leq 1} \|\mu P_\alpha\|_1,$$

where $\|\cdot\|_1$ is the total variation norm, corresponding to the case $V \equiv 1$. As pointed out in Ferré et al. (2013), this problem also occurs for the boundedness of the derivative operators because the successive derivatives of the conditional density involves a polynomial factor of increasing degree. For getting a Taylor expansion of the invariant probability, it is possible to use a general result given in Hervé and Pène (2010) (see Annex A of that paper) which is applied in Ferré et al. (2013) to the AR(1) process.

In this paper, we will prove a general result for getting a Taylor expansion of the invariant probability. This result has some similarities with that of Hervé and Pène (2010) but it is easier to apply and can give slightly better results in the examples. Contrarily to Hervé and Pène (2010), we do not make use of the spectral theory of operators (except that this spectral theory is hidden in Proposition 2 which is a straightforward consequence of Theorem 1 in Ferré et al. (2013) and ensures simultaneous geometric ergodicity properties that can be useful to apply our results, see Section 2.2 for details). Our approach is more direct and exploits some weak continuity/differentiability assumptions. Note however that the primary goal of the method given in Hervé and Pène (2010) is to study the regularity of the resolvent operator $\alpha \mapsto (zI - P_\alpha)^{-1}$ whereas we focus here on some properties of the invariant measures. To see what advantages can offer some weak continuity assumptions, one can show that for the AR(1) process, the application $\alpha \mapsto \mu P_\alpha$ is continuous for any $\mu \in \mathcal{M}_V$. As a consequence, from the approach we used for proving Theorem 1 of the present paper, one can show that the invariant probability μ_α of P_α is continuous for the norm $\|\cdot\|_V$, even if the transition kernel is never continuous for the corresponding operator norm. This completes the result given in Ferré et al. (2013), where the continuity property of the invariant probabilities is obtained in a weaker sense.

Finally, let us mention that our approach is particularly useful for models for which some power functions are Lyapunov and the derivatives of the conditional moments are bounded by some power functions. Our results can be applied using some easily checked conditions on the conditional density of the Markov kernels. See Section 3.1 for details. Moreover, though our results are stated for locally stationary Markov chains, one can get a straightforward generalization to some parametric models of ergodic Markov processes, using partial derivatives in the multidimensional case.

The paper is organized as follows. In Section 2, we give a general result for getting the regularity of an invariant measure. In Section 3, we study some sufficient conditions that can be easily checked in the examples. We give an application to the study of the bias in Section 4 and we check our assumptions for two statistical models in Section 5.

2 Regularity of an invariant probability with respect to an indexing parameter

In this section, we consider a family $\{P_u : u \in [0,1]\}$ of Markov kernels on a topological space G endowed with its Borel $\sigma-$field $\mathcal{B}(G)$. For an integer $k \geq 1$, let $V_1, V_2, \ldots, V_k$ be $k+1$ measurable functions defined on G, taking values in $[1,+\infty)$ and such that $V_0 \leq V_1 \leq \cdots \leq V_k$. For simplicity of notations we set $F_s = \mathcal{M}_{V_s}$ and $\|\cdot\|_s = \|\cdot\|_{V_s}$ for $0 \leq s \leq k$. We remind that $\{(F_\ell, \|\cdot\|_\ell) : 0 \leq \ell \leq k\}$ is a family of Banach spaces. Moreover, $0 \leq \ell \leq k-1$, we have $F_{\ell+1} \subset F_\ell$ and the injection

$$i_\ell : (F_{\ell+1}, \|\cdot\|_{\ell+1}) \to (F_\ell, \|\cdot\|_\ell)$$

is continuous. For $j = 0, 1, \ldots, k$, we also denote by $F_{0,j}$ the set of measures $\mu \in F_j$ such that $\mu(G) = 0$. For $0 \leq i \leq j \leq k$ and a linear operator $T : \left(F_j, \|\cdot\|_j\right) \to (F_i, \|\cdot\|_i)$, we set $\|T\|_{j,i} = \sup_{\|x\|_j \leq 1} \|Tx\|_i$ and $\|T\|_{0,j,i} = \sup_{x \in F_{0,j}, \|x\|_j \leq 1} \|Tx\|_i$. Finally, for each $u \in [0,1]$, we denote by T_u the linear operator acting on the space F_0 and defined by $T_u\mu = \mu P_u$. Finally, for a positive integer m, T_u^m denotes the iteration of order m of the operator T_u.

[A1] We have $T_u F_\ell \subset F_\ell$ for $1 \leq \ell \leq k$. Moreover, for each $\ell = 0, 1, \ldots, k$, there exists an integer $m_\ell \geq 1$ and a real number $\kappa_\ell \in (0,1)$ such that,

$$\sup_{u \in [0,1]} \|T_u^{m_\ell}\|_{0,\ell,\ell} \leq \kappa_\ell, \quad \sup_{u \in [0,1]} \|T_u\|_{\ell,\ell} < \infty$$

and for each $\mu \in F_\ell$, the application $u \mapsto T_u\mu$ is continuous from $[0,1]$ to $(F_\ell, \|\cdot\|_\ell)$.

[A2] for $1 \leq \ell \leq k$, there exists a continuous operator $T_u^{(\ell)} : (F_\ell, \|\cdot\|_\ell) \to (F_0, \|\cdot\|_0)$ such that for $0 \leq s \leq s + \ell \leq k$, $T_u^{(\ell)} F_{s+\ell} \subset F_s$, $\sup_{u \in [0,1]} \left\|T_u^{(\ell)}\right\|_{s+\ell,s} < \infty$ and for $\mu \in F_{s+\ell}$,

$$\lim_{h \to 0} \left\|T_{u+h}^{(\ell)}\mu - T_u^{(\ell)}\mu\right\|_s = 0 \text{ and } \lim_{h \to 0} \left\|\frac{T_{u+h}^{(\ell-1)}\mu - T_u^{(\ell-1)}\mu}{h} - T_u^{(\ell)}\mu\right\|_s = 0,$$

with the convention $T_u^{(0)} = T_u$.

Theorem 1. *Assume the assumptions* **A1 − A2** *hold true. Then the following statements are true.*

1.
   - *For each $u \in [0,1]$, the operator $I - T_u$ defines an isomorphism on each space $(F_{0,\ell}, \|\cdot\|_\ell)$ for $0 \leq \ell \leq k$. Moreover the inverse of $I - T_u$ is given by $(I - T_u)^{-1} = \sum_{k \geq 0} T_u^k$.*
   - *We have $\max_{0 \leq \ell \leq k} \sup_{u \in [0,1]} \left\|(I - T_u)^{-1}\right\|_{0,\ell,\ell} < \infty$.*
   - *For $0 \leq \ell \leq k$ and $\mu \in F_{0,\ell}$, the application $u \mapsto (T - T_u)^{-1}\mu$ is continuous for the norm $\|\cdot\|_\ell$.*
   - *Moreover, for each $u \in [0,1]$, we have for $0 \leq \ell \leq k-1$ and all $\mu \in F_{0,\ell+1}$,*

$$\lim_{h \to 0} \left\|\frac{(I - T_{u+h})^{-1}\mu - (I - T_u)^{-1}\mu}{h} + (I - T_u)^{-1} T_u^{(1)} (I - T_u)^{-1}\mu\right\|_\ell = 0.$$

2. *For each $u \in [0,1]$, there exists a unique probability measure μ_u such that $T_u\mu_u = \mu_u$ (μ_u is an invariant probability for P_u). Moreover $\mu_u \in F_k$.*

3. *The function $f : [0,1] \to (F_k, \|\cdot\|_k)$ defined by $f(u) = \mu_u$, for $u \in [0,1]$, is continuous. Moreover there exist some functions $f^{(1)}, \ldots, f^{(k)}$ such that*

   - *for $1 \le \ell \le k$, the function $f^{(\ell)} : [0,1] \to (F_{k-\ell}, \|\cdot\|_{k-\ell})$ is continuous,*
   - *for $1 \le \ell \le k$ and $u \in [0,1]$, $\lim_{h\to 0} \left\| \frac{f^{(\ell-1)}(u+h) - f^{(\ell-1)}(u)}{h} - f^{(\ell)}(u) \right\|_{k-\ell} = 0$,*
   - *the derivatives of f are given recursively by*

$$f^{(\ell)}(u) = \sum_{s=1}^{\ell} \binom{\ell}{s} (I - T_u)^{-1}\, T_u^{(s)} f^{(\ell-s)}(u).$$

Note. When $V_0 = V_1 = \cdots = V_k = V$, existence of the derivatives for the invariant probabilities is studied in Heidergott and Hordijk (2003). Our results contain theirs. One can show that the condition C^k used for stating their result entails **A2** because they use a continuity assumption of the derivatives operators with respect to the $V-$operator norm. On the other hand, their geometric ergodicity result (see Result 2 in their paper) for each kernel P_u and the continuity assumption of the kernel for $V-$operator norm entails the contraction **A1** (for the contraction coefficient κ_ℓ, see Section 3.2 below).

Proof of Theorem 1

1. - First, one can note that $(F_{0,\ell}, \|\cdot\|_\ell)$ is a closed vector subspace of $(F_\ell, \|\cdot\|_\ell)$ and then a Banach space. Moreover, from the assumption **A1**, the series $\sum_{k\ge 0} T_u^k$ is normally convergent for the norm $\|\cdot\|_{\ell,\ell}$ and is the inverse of $I - T_u$. Then $I - T_u$ defines an isomorphism on the space $(F_{0,\ell}, \|\cdot\|_\ell)$.
   - From the assumption **A1**, the second assertion is straightforward.
   - Next, we show that for $0 \le \ell \le k$ and $\mu \in F_{0,\ell}$, the application $u \mapsto (I - T_u)^{-1}\mu$ taking values in $(F_{0,\ell}, \|\cdot\|_\ell)$ is continuous. We use the decomposition

$$(I - T_{u+h})^{-1} - (I - T_u)^{-1} = -(I - T_{u+h})^{-1}(T_{u+h} - T_u)(I - T_u)^{-1}.$$

   From the previous point, we have $\sup_{u\in[0,1]} \|(I - T_u)^{-1}\|_{0,\ell,\ell} < \infty$ and $v \mapsto T_v(I - T_u)^{-1}\mu$ is continuous. Then the continuity of the application $u \mapsto (I - T_u)^{-1}\mu$ follows.
   - Finally, we show that the application $u \mapsto (I - T_u)^{-1}\mu$ is differentiable. Setting $z_{u,h} = h^{-1}(T_{u+h} - T_u)(I - T_u)^{-1}\mu$, we deduce from assumption **A2** that $\lim_{h\to 0} z_{u,h} = z_u = T_u^{(1)}(I - T_u)^{-1}\mu$ in $(F_{0,\ell}, \|\cdot\|_\ell)$. We use the decomposition

$$\begin{aligned} h^{-1}\left[(I - T_{u+h})^{-1}\mu - (I - T_u)^{-1}\mu\right] &= -(I - T_{u+h})^{-1} z_{u,h} \\ &= -(I - T_{u+h})^{-1}(z_{u,h} - z_u) - (I - T_{u+h})^{-1} z_u. \end{aligned}$$

   From the previous point, we have $\lim_{h\to 0}(I - T_{u+h})^{-1} z_u = (I - T_u)^{-1} z_u$ in $(F_{0,\ell}, \|\cdot\|_\ell)$. Moreover,

$$\left\|(I - T_{u+h})^{-1}(z_{u,h} - z_u)\right\|_\ell \le \sup_{u\in[0,1]} \left\|(I - T_u)^{-1}\right\|_{0,\ell,\ell} \|z_{u,h} - z_u\|_\ell \overset{h\to 0}{\to} 0.$$

   This shows that the application $u \mapsto (I - T_u)^{-1}\mu$ is differentiable with derivative $u \mapsto -(I - T_u)^{-1} T_u^{(1)} (I - T_u)^{-1}\mu$.

2. The space $F_{k,1} = \{\mu \in F_k : \mu \text{ is a probability measure }\}$ endowed with the norm $\|\cdot\|_k$ is a complete metric space. Then the result follows from the fixed point theorem.

3. We first show that f is continuous. We have $f(u+h) - f(u) = (I - T_{u+h})^{-1}(T_{u+h} - T_u)f(u)$. From the assumption **A1**, we have $\lim_{h\to 0} \|T_{u+h}f(u) - T_u f(u)\|_k = 0$. Then using the second assertion of point 1. of the theorem, we get $\lim_{h\to 0}(f(u+h) - f(u)) = 0$.

   Next, we prove the existence of the derivative and derive its properties by induction on ℓ with $1 \leq \ell \leq k$.

   - First, we assume that $\ell = 1$. Since for each $u \in [0,1]$, μ_u is a fixed point of T_u, we have
   $$\frac{\mu_{u+h} - \mu_u}{h} = \frac{T_{u+h} - T_u}{h} \cdot \mu_u + T_{u+h} \cdot \left(\frac{\mu_{u+h} - \mu_u}{h}\right).$$
   Then we get
   $$\begin{aligned} & \frac{\mu_{u+h} - \mu_u}{h} \\ = \; & (I - T_{u+h})^{-1} \, \frac{T_{u+h} - T_u}{h} \mu_u \\ = \; & (I - T_{u+h})^{-1} \left[\frac{T_{u+h} - T_u}{h} \mu_u - T_u^{(1)} \mu_u\right] + \left[(I - T_{u+h})^{-1} - (I - T_u)^{-1}\right] T_u^{(1)} \mu_u. \end{aligned}$$
   From the assumption **A2**, we have
   $$\lim_{h\to 0} \left\| \frac{T_{u+h}\mu_u - T_u \mu_u}{h} - T_u^{(1)} \mu_u \right\|_{k-1} = 0.$$
   From the second and the third assertions of the point 1., we get
   $$\lim_{h\to 0} \left\| \frac{\mu_{u+h} - \mu_u}{h} - f^{(1)}(u) \right\|_{k-1} = 0,$$
   where $f^{(1)}(u) = (I - T_u)^{-1} T_u^{(1)} \mu_u$. It remains to prove the continuity of $f^{(1)}$. As previously, it is sufficient to show that
   $$\lim_{h\to 0} \left\| T_{u+h}^{(1)} \mu_{u+h} - T_u^{(1)} \mu_u \right\|_{k-1} = 0.$$
   But this is a consequence of the continuity of f and of the assumption **A2**, using the decomposition
   $$T_{u+h}^{(1)} \mu_{u+h} - T_u^{(1)} \mu_u = \left[T_{u+h}^{(1)} - T_u^{(1)}\right] \mu_u + T_{u+h}^{(1)} \left[\mu_{u+h} - \mu_u\right].$$
   This shows the result for $\ell = 1$.

4. Now let us assume that for $1 \leq \ell \leq k-1$, f has ℓ derivatives such that for $1 \leq s \leq \ell$ and $u \in [0,1]$, the function $f^{(s)} : [0,1] \to (F_{k-s}, \|\cdot\|_{k-s})$ is continuous,
   $$\lim_{h\to 0} \left\| \frac{f^{(s-1)}(u+h) - f^{(s-1)}(u)}{h} - f^{(s)}(u) \right\|_{k-s} = 0$$
   and
   $$f^{(\ell)}(u) = \sum_{s=1}^{\ell} \binom{\ell}{s} (I - T_u)^{-1} \, T_u^{(s)} f^{(\ell-s)}(u).$$

- For $1 \leq s \leq \ell$, we set $z_u = T_u^{(s)} f^{(\ell-s)}(u)$ and first we are going to show that the application $u \mapsto z_u$ has a derivative. We have

$$\frac{z_{u+h} - z_u}{h} = \frac{T_{u+h}^{(s)} - T_u^{(s)}}{h} f^{(\ell-s)}(u) + T_{u+h}^{(s)} \frac{f^{(\ell-s)}(u+h) - f^{(\ell-s)}(u)}{h}.$$

Since $f^{(\ell-s)}(u) \in F_{k-\ell+s}$, we have from assumption **A2**,

$$\lim_{h\to 0} \left\| \frac{T_{u+h}^{(s)} - T_u^{(s)}}{h} f^{(\ell-s)}(u) - T^{(s+1)} f^{(\ell-s)}(u) \right\|_{k-\ell-1} = 0.$$

Next we set $w_{u,h} = \frac{f^{(\ell-s)}(u+h) - f^{(\ell-s)}(u)}{h}$. By the induction hypothesis, we have

$$\lim_{h\to 0} \left\| w_{u,h} - f^{(\ell-s+1)}(u) \right\|_{k-\ell+s-1} = 0.$$

Using the assumption **A2**, we have $\sup_{u\in[0,1]} \left\| T_{u+h}^{(s)} \right\|_{k-\ell+s-1,k-\ell-1} < \infty$. Then we get

$$\lim_{h\to 0} \left\| T_{u+h}^{(s)} \left(w_{u,h} - f^{(\ell-s+1)}(u) \right) \right\|_{k-\ell-1} = 0.$$

Using again the assumption **A2**, we have

$$\lim_{h\to 0} \left\| T_{u+h}^{(s)} f^{(\ell-s+1)}(u) - T_u^{(s)} f^{(\ell-s+1)}(u) \right\|_{k-\ell-1} = 0.$$

This shows that

$$\lim_{h\to 0} \left\| \frac{z_{u+h} - z_u}{h} - T_u^{(s+1)} f^{(\ell-s)}(u) - T_u^{(s)} f^{(\ell-s+1)}(u) \right\|_{k-\ell-1} = 0.$$

In the sequel we set $z_u^{(1)} = T_u^{(s+1)} f^{(\ell-s)}(u) + T_u^{(s)} f^{(\ell-s+1)}(u)$.

- Next we compute the derivative of $u \mapsto y_u = (I - T_u)^{-1} z_u$. We have

$$\frac{y_{u+h} - y_u}{h} = \frac{(I - T_{u+h})^{-1} - (I - T_u)^{-1}}{h} z_u + (I - T_{u+h})^{-1} \left(\frac{z_{u+h} - z_u}{h} - z_u^{(1)} \right) + (I - T_{u+h})^{-1} z_u^{(1)}.$$

Using the assumption **A2** and the previous point, we get

$$\lim_{h\to} \left\| \frac{y_{u+h} - y_u}{h} - (I - T_u)^{-1} T_u^{(1)} (I - T_u)^{-1} z_u - (I - T_u)^{-1} z_u^{(1)} \right\|_{k-\ell-1} = 0.$$

In the sequel, we set

$$t^{(\ell,s)}(u) = (I - T_u)^{-1} T_u^{(1)} (I - T_u)^{-1} z_u + (I - T_u)^{-1} z_u^{(1)}.$$

- Finally we get in $(F_{k-\ell-1}, \|\cdot\|_{k-\ell-1})$,

$$\lim_{h\to 0} \frac{f^{(\ell)}(u+h) - f^{(\ell)}(u)}{h} = f^{(\ell+1)}(u),$$

where

$$f^{(\ell+1)}(u) = \sum_{s=1}^{\ell} \binom{\ell}{s} t_u^{(\ell,s)}.$$

The expression for $f^{(\ell+1)}(u)$ given in the statement of the theorem follows from straightforward computations.

- Finally, using the induction assumption, the function $f^{(\ell+1-s)}$ is continuous for the norm $\|\cdot\|_{k-\ell+s-1}$ for each $1 \leq s \leq \ell+1$. Then the proof of the continuity of $f^{(\ell+1)}$ is similar to the proof of the continuity of $f^{(1)}$.

The the properties of the successive derivatives $f^{(1)}, \ldots, f^{(k)}$ follow by induction and the proof of Theorem 1 is now complete.□

3 Sufficient conditions

We now provide some sufficient conditions for **A1 – A2**. The most important situation in our examples concerns the case $V_\ell = V^{r_\ell}$ where $r_k = 1 \geq r_{k-1} \geq \cdots \geq r_0 \geq 0$ and V is a drift function which is a power of a norm. We will also assume that the kernel P_u is defined by

$$P_u(x, A) = \int_A f(u, x, y)\gamma(x, dy), \quad A \in \mathcal{B}(G),$$

where $f : [0,1] \times G^2 \to \mathbb{R}_+$ is a measurable function and γ is a kernel.

3.1 A sufficient set of conditions

We will use the following assumptions.

B1 For $\ell = 0, 1, \ldots, k$, the family of Markov kernels $\{P_u : u \in [0,1]\}$ is uniformly $V_\ell-$geometrically ergodic, i.e there exists $\kappa_\ell \in (0,1)$ such that for all $x \in G$,

$$\sup_{u \in [0,1]} \frac{\|\delta_x P_u^n - \mu_u\|_\ell}{V_\ell(x)} = O\left(\kappa_\ell^n\right),$$

where the invariant measure μ_u of P_u satisfies $\mu_u V < \infty$.

B2 For all $(x, y) \in G^2$, the function $u \mapsto f(u, x, y)$ is of class C^k and for $1 \leq \ell \leq k$, we denote by $\partial_1^{(\ell)} f$ its partial derivative of order ℓ.

B3 There exist $C > 0$ such that for integers $0 \leq s \leq s + \ell \leq k$ and $x \in G$,

$$\sup_{u \in [0,1]} \int \left|\partial_1^{(\ell)} f(u, x, y)\right| V_s(y)\mu_0(x, dy) \leq C V_{s+\ell}(x) \tag{1}$$

and for each $u \in [0,1]$,

$$\lim_{h \to 0} \int \left|\partial_1^{(k-s)} f(u+h, x, y) - \partial_1^{(k-s)} f(u, x, y)\right| V_s(y)\gamma(x, dy) = 0. \tag{2}$$

Corollary 1. *Assume that $k \geq 1$. The assumptions* **B1-B3** *entail the assumptions* **A1-A2** *and Theorem 2 applies for the derivatives operators*

$$T_u^{(\ell)}\mu = \int \mu(dx)\partial_1^{(\ell)} f(u, x, y)\gamma(x, dy), \quad 1 \leq \ell \leq k, \quad \mu \in \mathcal{M}_V.$$

Proof of Corollary 1

1. First, note that from (1) applied with $\ell = 0$ we get $\sup_{u\in[0,1]} \|T_u\|_{s,s} < \infty$ for $s = 0, 1, \ldots, k$. Then assumption **A1** follows from assumption **B1**. Indeed, we have (see for instance Rudolf and Schweizer (2017) Lemma 3.2)
$$\|T_u^n\|_{0,\ell,\ell} = \Delta_{V_\ell}(P_u^n) \leq \frac{\|\delta_x P_u^n - \pi_u\|_\ell}{V_\ell(x)}.$$
See (3) for the definition of Δ_{V_ℓ}. This entails the existence of an integer $m_\ell \geq 1$ such that $\sup_{u\in[0,1]} \|T_u^{m_\ell}\|_{0,\ell,\ell} < 1$.

2. Next, we check the assumption **A2**. We first notice that for $0 \leq s \leq s+\ell \leq k$ and $\mu \in F_{s+\ell}$ we have
$$\begin{aligned}\|T_u^{(\ell)}\mu\|_s &\leq \int |\mu|(dx) \int \gamma(x,dy) \left|\partial_1^{(\ell)} f(u,x,y)\right| V_s(y) \\ &\leq C \int |\mu|(dx) V_{s+\ell}(x) = C\|\mu\|_{s+\ell}.\end{aligned}$$
This shows that $T_u^{(\ell)} F_{s+\ell} \subset F_s$ and $\sup_{u\in[0,1]} \|T_u^{(\ell)}\|_{s+\ell,s} \leq C$. Next, for $\mu \in F_s$, we show the continuity of the application $u \mapsto T_u^{(\ell)}\mu$ taking values in $F_{s+\ell}$. We have
$$\|T_{u+h}^{(\ell)}\mu - T_u^{(\ell)}\mu\|_s \leq \int |\mu|(dx) \int \mu_0(x,dy) \left|\partial_1^{(\ell)} f(u+h,x,y) - \partial_1^{(\ell)} f(u,x,y)\right| V_s(y).$$
From **B2** (1) and the Lebesgue theorem, it is enough to prove that for all $x \in G$,
$$\lim_{h\to 0} \int \gamma(x,dy) \left|\partial_1^{(\ell)} f(u+h,x,y) - \partial_1^{(\ell)} f(u,x,y)\right| V_s(y) = 0.$$
We consider two cases.

   - If $s+\ell = k$, this continuity is a direct consequence of the assumption **B2** (2).
   - We next assume that $s+\ell+1 \leq k$. We have
   $$\int \gamma(x,dy) \left|\partial_1^{(\ell)} f(u+h,x,y) - \partial_1^{(\ell)} f(u,x,y)\right| V_s(y) \leq h \sup_{v\in[0,1]} \int \gamma(x,dy) \left|\partial_1^{(\ell+1)} f(v,x,y)\right| V_s(y).$$
   Then the result follows from the assumption **B2** (1).

   Finally, we show the differentiability property of the operators. For $\mu \in F_{s+\ell}$, we have
$$\begin{aligned}&\left\|\frac{T_{u+h}^{(\ell-1)}\mu - T_u^{(\ell-1)}\mu}{h} - T_u^{(\ell)}\mu\right\|_s \\ \leq\ & \int_{[u,u+h]} dv |\mu|(dx) \int \gamma(x,dy) \left|\partial_1^{(\ell)} f(v,x,y) - \partial_1^{(\ell)} f(u,x,y)\right| V_s(y).\end{aligned}$$
The result follows by using the same arguments than for the proof of the continuity of the application $u \mapsto T_u^{(\ell)}\mu$. This completes the proof of Corollary 1.□

Note. When G is a normed space and if for some $p \leq p_0$, $0 \leq \ell \leq k$ and $r_0, r_1, \ldots, r_k > 0$,

$$\int \gamma(x, dy) \left|\partial_1^{(\ell)} f(u, x, y)\right| (1 + \|y\|)^p \leq C\,(1 + \|x\|)^{p + r_\ell},$$

then assumption **B3** (1) is checked by setting $V_\ell(x) = (1 + |x|)^{p + r\ell}$ with $r = \max(r_1, r_2/2, \ldots, r_k/k)$ and assuming that $p + rk \leq p_0$.

3.2 Sufficient conditions for B1

Assumption **B1** is related to some simultaneous $V-$uniform ergodicity condition. $V-$uniform ergodicity is generally obtained under a drift condition and a small set assumption. See Meyn and Tweedie (2009), Chapter 16 for details. Let us first recall the definition of a small set. For a Markov kernel P on $(G, \mathcal{B}(G))$, a set $C \in \mathcal{B}(G)$ is called a $(\eta, \nu)-$small set, where η a positive real number and ν a probability measure on $(G, \mathcal{B}(G))$ if

$$P(x, A) \geq \eta \nu(A), \text{ for all } A \in \mathcal{B}(G) \text{ and all } x \in C.$$

In Truquet (2017), we extensively used a result of Hairer and Mattingly (2011) which we recall below. For simplicity we introduce the following condition. For $\lambda \in (0,1)$, $b, \eta > 0$, $R > \frac{2b}{1-\lambda}$ and ν a probability measure on $(G, \mathcal{B}(G))$, we will say that a Markov kernel P satisfies the condition $C\,(V, \lambda, b, R, \eta, \nu)$ if

$$PV \leq \lambda V + b \quad \text{and} \quad \{x \in G : V(x) \leq R\} \text{ is a } (\eta, \nu) \text{ small set.}$$

If P is a Markov kernel on $(G, \mathcal{B}(G))$, we also define the following Dobrushin coefficient

$$\Delta_V(P) = \sup_{\mu \in \mathcal{M}_V, \mu \neq 0, \mu(G) = 0} \frac{\|\mu P\|_V}{\|\mu\|_V} = \sup_{x, y \in G, x \neq y} \frac{\|\delta_x P - \delta_y P\|_V}{V(x) + V(y)}. \tag{3}$$

Note also that, with the notations of Section 2, we have if $T\mu = \mu P$, $\|T\|_{0,\ell,\ell} = \Delta_{V_\ell}(P)$.

Under such a condition, Theorem 1.3 in Hairer and Mattingly (2011) guarantees the existence of $\alpha \in (0,1)$ and $\delta > 0$, not depending on $u \in [0,1]$ such that $\Delta_{V_\delta}(P_u^m) \leq \alpha$ with $V_\delta = 1 + \delta V$. Then, using the equivalence of the norms $\|\cdot\|_V$ and $\|\cdot\|_{V_\delta}$, one can show as in Proposition 6 in Truquet (2017) that there exists $C > 0$ and $\rho \in (0,1)$ such that

$$\sup_{u \in [0,1]} \|\delta_x Q_u^n - \pi_u\|_V \leq C V(x) \rho^n.$$

Then the family of Markov kernels $\{P_u : u \in [0,1]\}$ is simultaneously $V-$geometrically ergodic.

When the level sets of V are bounded and all bounded sets are small, the following proposition is interesting for many examples. Its proof is straightforward, since a drift condition on V entails a drift condition on V^r for all $r \in (0,1)$.

Proposition 1. *Assume that there exist a positive real number K such that $P_u V \leq KV$ and an integer $m \geq 1$ such that the family of Markov kernels $\{P_u^m : u \in [0,1]\}$ satisfies the a condition $C\,(V, \lambda, b, R, \eta, \nu_u)$ whatever the value of $R > 0$ (the other constants possibly depending on R). Then this family is simultaneously V^r-uniform geometrically ergodic for all $r \in (0,1]$.*

In Truquet (2017), we used the type of condition given above for our locally stationary models. However, if the problem is just to derive the regularity of the application $u \mapsto \pi_u$, the assumption **B1** can be guaranteed by another set of conditions. For instance, if each kernel P_u is $V-$ geometrically ergodic, then perturbation methods can be applied to get a local simultaneous $V-$geometric ergodicity property which can easily be extended to the interval $[0,1]$ by compactness. When the Markov kernel is not continuous with respect to the operator norm, Ferré et al. (2013) (see Theorem 1) obtained a nice result based on the Keller-Liverani perturbation theorem. The following proposition is an easy consequence of their result.

Proposition 2. *Let $\{P_u : u \in [0,1]\}$ be a family of Markov kernels on a measurable space $(G,\mathcal{G})$ and $V : G \to [1,\infty)$ a measurable function satisfying the three following conditions.*

1. *For each $u \in [0,1]$, the Markov kernel P_u admits a unique invariant measure μ_u such that $\mu_u \cdot V < \infty$ and there exists $\kappa_u \in (0,1)$ such that $\|P_u^n - \mu_u \mathbb{1}_G\|_{V,V} = O(\kappa_u^n)$.*
2. *There exist an integer $m \geq 1$, a real number $\lambda \in (0,1)$ and two positive real numbers K and L such that for all $u \in [0,1]$,*
$$P_u V \leq KV, \quad P_u^m V \leq \lambda V + L.$$
3. *The application $u \to P_u$ is continuous for the norm $\|\cdot\|_{V,1}$.*

Then, there exists $\kappa \in (0,1)$ such that

$$\sup_{u\in[0,1]} \|P_u^n - \mu_u \mathbb{1}_G\|_{V,V} = O(\kappa^n).$$

Moreover $\sup_{u\in[0,1]} \Delta_V(P_u^n) = O(\kappa^n)$.

Note For the AR(1) process $X_n(u) = \alpha(u)X_{n-1}(u) + \xi_n$, with $u \mapsto \alpha(u) \in (-1,1)$ continuous and ξ_1 has absolutely continuous error distribution with a density denoted by f_ξ having a moment of order $a > 0$, it is well known that $P_u(x,dy) = f_\xi(y - \alpha(u)x)\,dy$ is V_a-geometrically ergodic with $V_a(x) = (1+|x|)^a$. See Guibourg et al. (2011), Section 4 for a discussion of the geometric ergodicity of some classical autoregressive processes. Moreover, the continuity of $u \mapsto P_u$ holds for the norm $\|\cdot\|_{V_a,1}$ as shown in Ferré et al. (2013), Example 1 (the result is shown for the case $a = 1$ but extension to the case $a > 0$ is straightforward). Then Proposition 2 applies to this example. This approach is more interesting here because it is not necessary to assume another property for the density f_ξ which is often require to check the small set condition in $C(V,\lambda,b,R,\eta,\nu_u)$, e.g f_ξ is lower-bounded by a positive constant on each compact set.

Proof of Proposition 2 Let $u \in [0,1]$. Using Theorem 1 of Ferré et al. (2013), one can find an open interval $I_u \ni u$ of $[0,1]$, two real numbers $C_u > 0$ and $\kappa_u \in [0,1]$ such that $\sup_{v\in I_u} \|P_v^n - \mu_v \mathbb{1}_G\|_{V,V} \leq C_u \kappa_u^n$. From a compactness argument, $[0,1]$ can be covered by a finite number of such intervals $I_{u_1}, \ldots, I_{u_p}$. Then the simultaneous geometrical ergodicity condition follows by setting $\kappa = \max_{1\leq i\leq p} \kappa_{u_i}$ and defining the constant $C = \max_{1\leq i\leq p} C_{u_i}$. Moreover, using Lemma 3.2 in Rudolf and Schweizer (2017), we have $\Delta_V(P_u^n) \leq \|P_u^n - \mu_u \mathbb{1}_G\|_{V,V}$ which gives the second conclusion of the proposition.□

4 Locally stationary Markov chains

In this section, we consider a topological space E endowed with its Borel $\sigma-$field $\mathcal{B}(E)$ and a triangular array of Markov chains $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ such that for all $(x,A) \in E \times \mathcal{B}(E)$ and $1 \leq k \leq n$,

$$\mathbb{P}(X_{n,k} \in A | X_{n,k-1} = x) = Q_{k/n}(x,A), \quad X_{n,0} \sim \pi_0.$$

We remind that for $u \in [0,1]$, π_u denotes the invariant probability of Q_u.

4.1 Some results about locally stationary Markov chains

We first recall some results obtained in Truquet (2017), Section 4. For simplicity, we introduce the two following conditions. For $\epsilon > 0$, we denote $I_m(\epsilon)$ the subsets of $[0,1]^m$ such that $(u_1,\ldots,u_m) \in I_m(\epsilon)$ if and only if $|u_i - u_j| < \epsilon$ for $1 \leq i,j \leq m$.

L1 There exist an integer $m \geq 1$, some positive real numbers $\epsilon, K, \lambda, b, R, \eta$ with $\lambda < 1$ and a probability measure ν such that for all $(u_1,u_2,\ldots,u_m) \in I_m(\epsilon)$, the kernel $Q_{u_1}Q_{u_2}\cdots Q_{u_m}$ satisfies the condition $C(V,\lambda,b,R,\eta,\nu)$ and $Q_{u_1}V \leq KV$.

L2 There exists a measurable function $\widetilde{V} : E \to [1,\infty)$ such that $\sup_{u\in[0,1]} \pi_u \widetilde{V} < \infty$ and for all $x \in E$, $\|\delta_x Q_u - \delta_x Q_v\|_V \leq \widetilde{V}(x)|u-v|$.

Under the conditions **L1-L2**, it is shown in Truquet (2017) (see Theorem 3) that the distribution $\pi_k^{(n)}$ of $X_{n,k}$ satisfies

$$\|\pi_k^{(n)} - \pi_u\|_V \leq C\left[\left|u - \frac{k}{n}\right| + \frac{1}{n}\right], \tag{4}$$

where $C > 0$ does not depend on u,n,k. One can also get the same type of results for other joint distributions, i.e the probability distribution of $\left(X_{n,k}, X_{n,k+1},\ldots,X_{n,k+j-1}\right)$ for $j \geq 2$, by adding an additional moment condition. Note that under the two conditions **L1-L2**, assumption **B1** is automatically satisfied, see Section 3.2 for details. With respect to the assumption used in Proposition 1, i.e. the products Q_u^m satisfies a simultaneous small set and drift condition, the condition **L1** is more restrictive but is useful to guarantee some $\beta-$mixing properties for the triangular array. Note also that condition **L2** is always satisfied if assumption **B3** (1) holds true, setting $V = V_0$ and $\widetilde{V} = V_1$.

In the sequel, we will say that the triangular array of Markov chains $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ satisfying (4) is $V-$locally stationary.

4.2 Application to bias control in nonparametric estimation

In this subsection, we assume that $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ is a triangular array of $V-$locally stationary Markov chains. Assume that the application $\phi : [0,1] \to \mathcal{M}_V$ defined by $\phi(u) = \pi_u$ is of class C^k. Let $f : E \to \mathbb{R}$ be a measurable function such that $|f|_V < \infty$. A functional $\psi_f(u) = \int f d\pi_u$ of the local invariant measure can be estimated by local polynomials. To this end, let K be a continuous probability density, bounded and supported on $[-1,1]$ and $b \in (0,1)$ a bandwidth parameter such that $b = b_n \to 0$ and $nb \to \infty$. We set $K_b = \frac{1}{b}K(\cdot/b)$. An estimator $\hat{g}(u)$ of $g(u)$ is given by the first component of the vector

$$\hat{\mathcal{H}}_f(u) := \left(\hat{\psi}_f(u), b\hat{\psi}'_f(u),\ldots,b^{k-1}\hat{\psi}_f^{(k-1)}(u)\right)' = \arg\min_{\alpha_0,\ldots,\alpha_{k-1}\in\mathbb{R}} \sum_{t=1}^n K_b\left(u - \frac{t}{n}\right)\left[f(X_{n,t}) - \sum_{i=0}^{k-1}\alpha_i \frac{(t/n-u)^i}{b^i i!}\right]^2.$$

For $1 \leq t \leq n$, we set

$$v_t(u) = \left(1, \frac{t/n-u}{b},\ldots,\frac{(t/n-u)^{k-1}}{b^{k-1}(k-1)!}\right)'$$

and

$$D(u) = \frac{1}{n}\sum_{t=1}^n K_b(t/n-u)\, v_t(u)v_t(u)', \quad \hat{N}_f(u) = \frac{1}{n}\sum_{t=1}^n K_b(t/n-u)\, v_t(u) f(X_{n,t}).$$

From (4), we have

$$\max_{1\le t\le n}\sup_{|f|_V\le 1}\left|\mathbb{E}f\left(X_{n,t}\right)-\psi_f(t/n)\right|=O\left(1/n\right).$$

Next, setting $\mathcal{H}_f(u)=\left(\psi_f(u),b\psi_f'(u),\ldots,b^{k-1}\psi_f^{(k-1)}(u)\right)'$ and using the differentiability properties of ϕ, there exists $C>0$ such that for all $n\ge 1$ and $1\le t\le n$,

$$\sup_{u\in[0,1]}\sup_{|f|_V\le 1}\left|\psi_f(t/n)-\mathcal{H}_f(u)'v_t(u)\right|\le C(u-t/n)^k.$$

We deduce that

$$\sup_{u\in[0,1]}\sup_{|f|_V\le 1}\left|\mathbb{E}\hat{N}_f(u)-D(u)\mathcal{H}_f(u)\right|=O\left(b^k+\frac{1}{n}\right).$$

Moreover, it is well-known that $\max_{u\in[0,1]}\|D(u)^{-1}\|=O(1)$. See for instance Tsybakov (2009), Lemma 1.5. Then we get

$$\sup_{u\in[0,1]}\sup_{|f|_V\le 1}\left|\mathbb{E}\hat{\mathcal{H}}_f(u)-\mathcal{H}_f(u)\right|=O\left(b^k+\frac{1}{n}\right).$$

In conclusion, up to a term of order $1/n$ which is negligible can be interpreted as a deviation term with respect to stationarity, the bias is of order b^k when ψ_f is C^k. We then recover a classical property of local polynomials. Note that the variance of the estimator $\hat{\mathcal{H}}_f(u)$ is as usual of order $1/nb$ under some mixing properties for the triangular array.

4.3 Estimation of other joint distributions

One can be also be interested in estimating a functional $u\mapsto\int gd\pi_{u,j}$ for $j\ge 2$, $g:E^j\to\mathbb{R}$ a measurable function and $\pi_{u,j}(d\mathbf{x})=\pi_u(dx_1)Q_u(x_1,dx_2)\cdots Q_u(x_{j-1},dx_j)$, for instance the local covariance $u\mapsto\mathrm{Cov}\left(X_0(u),X_1(u)\right)$ where $(X_k(u))_k$ is a Markov chain with kernel Q_u. Here we set for $\mathbf{x}\in E^j$,

$$Q_{u,j}g(\mathbf{x})=\int g(\mathbf{y})f\left(u,x_j,y_j\right)\gamma_j(\mathbf{x},d\mathbf{y}),$$

with $\gamma_j(\mathbf{x},d\mathbf{y})=\gamma(x_j,dy_j)\prod_{i=1}^{j-1}\delta_{x_{i+1}}(dy_i)$. We also $V_s^{(j)}(\mathbf{x})=\sum_{i=1}^j V_s(x_i)$.

- Let us check that $Q_{u,j}$ satisfies the assumption **B1** when Q_u satisfies itself this assumption as well as assumption **B2** (1) with $\ell=0$. Let $1\le s\le k$. For an integer $h\ge j$ and a measurable function $g:E^j\to\mathbb{R}$ such that $|g|\le V_s^{(j)}$, we have

$$\begin{aligned}\left|Q_{u,j}^j g(\mathbf{x})\right| &= \int\left|g(y_1,\ldots,y_j)\right|Q_u(x_j,dy_1)Q_u(y_1,dy_2)\cdots Q_u(y_{j-1},dy_j)\\ &\le \sum_{i=1}^j Q_u^i V_s(x_j)\\ &\le C_jV_s(x_j),\end{aligned}$$

  with $C_j=\sum_{i=1}^j C^i$ and C defined in (1).. Then we get

$$\sup_{|g|\le V_s^{(j)}}\left|Q_{u,j}^h g(x)-\pi_{u,j}g\right|\le\sup_{|f|\le C_jV_s}\left|Q_u^{h-j}f(x_j)-\pi_uf\right|\le C_j\sup_{|f|\le V_s}\left|Q_u^{h-j}f(x_j)-\pi_uf\right|.$$

  This bounds entails **B1**.

- Now assume that the family $\{Q_u : u \in [0,1]\}$ satisfies the assumptions **B2-B3**. Then the family $\{Q_{u,j} : u \in [0,1]\}$ automatically satisfies the assumption **B2** and **B3** (2). Checking assumption **B3** (1) requires an additional assumption. Indeed, we have

$$\int V_s^{(j)}(\mathbf{y}) \left|\partial_1^{(\ell)} f(u, x_j, y_j)\right| \gamma_j(\mathbf{x}, d\mathbf{y}) \leq C \left[V_{s+\ell}(x_j) + \sum_{i=2}^{j} V_s(x_i) M_\ell(x_j) \right],$$

where $M_\ell(x_j) = \sup_{u\in[0,1]} \int \left|\partial_1^{(\ell)} f(u, x_j, y_j)\right| \gamma(x_j, dy_j)$. Hence condition $V_s(x_i)M_\ell(x_j) \leq C\left(V_{s+\ell}(x_i) + V_{s+\ell}(x_j)\right)$ for a constant $C > 0$ is required for applications of our results. However, this condition will be satisfied for a normed space $(G, \|\cdot\|)$ if

$$\int \left(1 + \|y_j\|\right)^p \left|\partial_1^{(\ell)} f(u, x_j, y_j)\right| \gamma(x_j, dy_j) \leq C\left(1 + \|x\|\right)^{p+r_\ell}$$

for $0 \leq p \leq p_0$. Indeed in this case, one can take (up to a constant) $V_\ell(x_j) = \left(1 + \|x_j\|\right)^{p+r\ell}$ with $r = \max\left(r_1, r_2/2, \ldots, r_k/k\right)$ and k such that $p + rk \leq p_0$.

5 Examples

In this section, we consider too examples with unbounded state spaces and for which our results offer a complement with respect to that of Dahlhaus et al. (2017). In the two examples given below, an extension to higher order Markov chain is possible with more tedious arguments by using a vectorization. For simplicity, we only consider the case of a Markov chain of order 1.

5.1 Autoregressive processes

We consider the following autoregressive process

$$X_{n,k} = m\left(k/n, X_{n,k-1}\right) + \sigma(k/n)\varepsilon_k, \quad 1 \leq k \leq n,$$

where $E = \mathbb{R}$, $m : [0,1] \times E \to E$ and $\sigma : [0,1] \to \mathbb{R}_+$ are two measurable functions and $(\varepsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d random variables. We will use the following assumptions.

E1 The function σ is of class C^k. Moreover $\min_{u\in[0,1]} \sigma(u) > 0$.

E2 For all $x \in \mathbb{R}$, the function $u \mapsto m(u, x)$ is of class C^k. Moreover there exists $\alpha \in (0, 1)$ and four positive real numbers β, δ, C_1, C_2 such that for all $x \in \mathbb{R}$,

$$\sup_{u\in[0,1]} |m(u, x)| \leq \alpha|x| + \beta,$$

$$\max_{1\leq\ell\leq k} \sup_{u\in[0,1]} \left|\partial_1^{(\ell)} m(u, x)\right| \leq C_1|x|^\delta + C_2.$$

E3 The noise ε_1 has a density f_ε of class C^2, positive everywhere and there exists $p > 0$ such that

$$\int |y|^{p+k\delta+(1-\delta)s} \left|f_\epsilon^{(s)}(y)\right| < \infty, \quad s = 0, 1, \ldots, k.$$

Proposition 3. *Under the assumptions* **E1-E3***, the triangular array of Markov chains* $\{X_{n,k} : 1 \le k \le n, n \ge 1\}$ *is locally stationary. Moreover the unique invariant probability* π_u *associated to the kernel*

$$Q_u(x, dy) = \frac{1}{\sigma(u)} f_\varepsilon\left(\frac{y - m(u, x)}{\sigma(u)}\right) dy$$

is of class C^k *in the space* $\mathcal{M}_{V_0}$ *where* $V_0 = 1 + |x|^p$.

Proof of Proposition 3 We use Proposition 1 with $m = 1$ to check **B1**. We set $V_\ell(x) = (1 + |x|)^{p+\ell\delta}$ for $0 \le \ell \le k$. The drift condition follows as for the standard AR(1) process, using **E2**. The small set condition easily follows from the fact that f_ε is lower-bounded on any compact set. Assumption **B3** can be checked after some computations. One can use the following representation of the derivatives of $f(u, x, y) = \frac{1}{\sigma(u)} f_\varepsilon\left(\frac{y-m(u,x)}{\sigma(u)}\right)$:

$$\partial_1^{(\ell)} f(u, x, y) = \sum_{s=0}^{\ell} f_\varepsilon^{(s)}\left(b_{x,y}(u)/\sigma(u)\right) \mathcal{P}_{u,s}\left(b_{x,y}(u), b_{x,y}^{(1)}(u), \ldots, b_{x,y}^{(\ell-s+1)}(u)\right),$$

with $b_{x,y} = y - m(u, x)$ and $\mathcal{P}_{u,s}$ is a polynomial of degree s with coefficients of type $h(u)$ for continuous functions h. Details are omitted.□

5.2 Galton-Watson process with immigration

For $u \in [0, 1]$, let p_u and q_u be two probability distributions supported on the nonnegative integers and for $x \in \mathbb{Z}_+$, $Q_u(x, \cdot)$ will denote the probability distribution given by the convolution product $p_u^{*x} * q_u$ with $p_u^{*x} = p_u^{*(x-1)} * p_u$ if $x \ge 1$, $p_u^{*1} = p_u$ and the convention $p_u^{*0} = \delta_0$. Then Q_u is the transition matrix of a Galton-Watson process with immigration. Such a process is also used in time series analysis of discrete data. When $p_u = p$ is a Bernoulli distribution of parameter $\alpha \in (0, 1)$, the stationary process $X_k = \alpha \circ X_{k-1} + \varepsilon_k$ is called an INAR process and is studied in Al Osh and Alzaid (1987). Here $\alpha \circ x$ denotes a random variable following a binomial distribution of parameters (x, α).

We will use the following assumptions.

G1 We have $\alpha := \sup_{u\in[0,1]} \sum_{x\ge0} x p_u(x) < 1$ and there exists an integer x_0 such that $\beta := \inf_{u\in[0,1]} q_u(x_0) > 0$.

G2 For each integer $x \ge 0$, the application $u \mapsto p_u(x)$ and $u \mapsto q_u(x)$ are of class C^k. Moreover, there exists a positive integer d such that for $s = 0, 1, \ldots, k$,

$$\lim_{M\to\infty} \sup_{u\in[0,1]} \sum_{x\ge M} x^{d+k-s} \left[|p_u^{(s)}(x)| + |q_u^{(s)}(x)|\right] = 0.$$

Proposition 4. *Assume that the assumptions* **G1-G3** *hold true and set* $V_0(x) = 1 + x^d$ *for a nonnegative integer* x*. Then* $u \mapsto \pi_u$ *is of class* C^k *for the norm* $\|\cdot\|_0$.

Note. Assumption **G2** is satisfied for Bernoulli, Poisson or negative binomial distributions provided the real-valued parameter of these distributions is a C^k function taking values in the usual intervals $(0, 1)$ (for the Bernoulli or negative binomial distribution) or $(0, \infty)$ (for the Poisson distribution).

Proof of Proposition 4 Once again, we use Proposition 1 to check **B1**.

- The small set condition is satisfied for any finite set $C = \{0, 1, \ldots, c\}$. Indeed, from assumption **G1**, we have $p_u(0) \geq 1 - \alpha$ and for $x \in C$,

$$Q_u(x, x_0) \geq (1 - \alpha)^c \beta.$$

Then the small set condition is satisfied for the probability measure $\nu = \delta_{x_0}$ and $\eta = (1 - \alpha)^c \beta$.

- Now, the drift condition for the function $V_k(x) = (1 + x)^{d+k}$ can be checked as in Truquet (2017) (the proof is done for the Bernoulli distribution but it can be easily extended) by using the assumption **G1** and the assumption **G2** with $s = 0$.

- Next we check the assumption **B3** (1). In the sequel we set for an integer $M \geq 0$,

$$\mathcal{D}_M := \sup_{u \in [0,1]} \max_{s=0,1,\ldots,k} \sum_{x \geq M} (1 + x)^{d+k-s} \left[|p_u^{(s)}(x)| + |q_u^{(s)}(x)| \right].$$

For a nonnegative integer x, the conditional density $f(u, x, \cdot)$ with respect to the counting measure on $\mathbb{Z}_+$ is given by the convolution product $p_u^{*x} * q_u$. Then we have for $1 \leq \ell \leq k$,

$$\partial_1^{(\ell)} f(u, x, y) = \sum_{j_1 + \cdots + j_{x+1} = y} \sum_{\ell_1 + \ell_2 + \cdots + \ell_{x+1} = \ell} \frac{\ell!}{\prod_{i=1}^{x+1} \ell_i!} p_u^{(\ell_1)}(j_1) \cdots p_u^{(\ell_x)}(j_x) q_u^{(\ell_{x+1})}(j_{x+1}).$$

If $0 \leq s \leq s + \ell \leq k$, we have, using the convexity of the application $y \mapsto V_s(y) := (1 + y)^{d+s}$,

$$\sum_{y \geq 0} V_s(y) \left| \partial_1^{(\ell)} f(u, x, y) \right| \sum_{\ell_1 + \ell_2 + \cdots + \ell_{x+1} = \ell} \frac{\ell!}{\prod_{i=1}^{x+1} \ell_i!} (x + 1)^{d+s} \mathcal{D}_0^{\ell+1} = V_{s+\ell}(x) \mathcal{D}_0^{\ell+1}.$$

In the previous inequality, we have used the following property: if $\ell_1 + \ell_2 + \cdots + \ell_{x+1} = \ell$ then at most ℓ of these integers are positive and then for $i = 1, 2, \ldots, x + 1$,

$$\sum_{j_1, \ldots, j_{x+1} \geq 0} V_s(j_i) \left| p_u^{(\ell_1)}(j_1) \cdots p_u^{(\ell_x)}(j_x) q_u^{(\ell_{x+1})}(j_{x+1}) \right| \leq \mathcal{D}_0^{\ell+1}.$$

- Finally, we check the assumption **B3** (2). Let x be a nonnegative integer and s an integer such that $0 \leq s \leq k$. It is easily seen that for $M > 0$,

$$\lim_{h \to 0} \sum_{y \geq 0} \left| \partial_1^{(k-s)} f(u + h, x, y) - \partial_1^{(k-s)} f(u, x, y) \right| V_s(y) = 0.$$

Then (2) will follow if we show that

$$\lim_{M \to \infty} \sup_{u \in [0,1]} \sum_{y \geq M} \left| \partial_1^{(k-s)} f(u, x, y) \right| V_s(y) = 0. \tag{5}$$

But as for the proof of (1), we have

$$\sum_{y \geq M} V_s(y) \left| \partial_1^{(k-s)} f(u, x, y) \right| \leq \sum_{\ell_1 + \ell_2 + \cdots + \ell_{x+1} = k-s} \frac{(k - s)!}{\prod_{i=1}^{x+1} \ell_i!} (x + 1)^{d+s} \mathcal{D}_{\frac{M}{x+1}}^{k-s+1} = V_k(x) \mathcal{D}_{\frac{M}{x+1}}^{k-s+1}.$$

From the assumption **G2**, we get (5) and then (2).□

Acknowledgements. The authors would like to thank Loïc Hervé and James Ledoux for some clarifications about the perturbation theory of Markov operators.

References

M. Al Osh and A. Alzaid. Firs-order integer-valued autoregressive process. *J. Time Series Anal.*, 8:261–275, 1987.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1997.

R. Dahlhaus and S.S. Rao. Statistical inference for time-varying arch processes. *Ann. Statist.*, 34:1075–1114, 2006.

R. Dahlhaus, S. Richter, and W.B. Wu. Towards a general theory for non-linear locally stationary processes. *arXiv:1704.02860*, 2017.

J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Springer-Science, 1996.

D. Ferré, L. Hervé, and J. Ledoux. Regular perturbation of $v-$geometrically ergodic markov chains. *J. Appl. Prob.*, 50:184–194, 2013.

P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. Normalized least-squares estimation in time-varying arch models. *Ann. Statist.*, 36:742–786, 2008.

D. Guibourg, L. Hervé, and J. Ledoux. Quasi-compactness of markov kernels on weighted-supremum spaces and geometrical ergodicity. *hal-00632580v5*, 2011.

M. Hairer and J.C. Mattingly. Yet another look at harris' ergodic theorem for markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications IV*, volume 63, pages 109–117. Birkhäuser,/Springer Basel AG, Basel, 2011.

B. Heidergott and A. Hordijk. Taylor series expansions for stationary markov chains. *Adv. in App. Probab.*, 35:1046–1070, 2003.

L. Hervé and F. Pène. The nagaev-guivarc'h method via the keller-liverani theorem. *Bull. Soc. Math. France*, 138:415–489, 2010.

N. V. Kartashov. Inequalities in theorems of ergodicity and stability for markov chains with common phase space. i. *Theory Probab. Appl.*, 30:247–259, 1986.

S. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability 2nd*. Cambridge University Press New York, 2009.

E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for time varying autoregressive processes. *The Annals of Statistics*, 33:2610–2654, 2005.

D. Rudolf and N. Schweizer. Perturbation theory for markov chains via wassertein distance. *Bernoulli*, 2017.

P.J. Schweitzer. Perturbation theory and finite markov chains. *J. Appl. Prob.*, 5:401–413, 1968.

S. Subba Rao. On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in App. Probab.*, 38:1155–1172, 2006.

L. Truquet. Parameter stability and semiparametric inference in time-varying arch models. *Forthcoming in J. R. Stat. Soc. Ser. B*, 2016.

L. Truquet. Local stationarity and time-inhomogeneous markov chains. *ArXiv*, 2017.

A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics, 2009.

M. Vogt. Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40:2601–2633, 2012.

T. Zhang and W.B. Wu. Inference of time-varying regression models. *Ann. Statist.*, 40:1376–1402, 2012.