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

LIONEL TRUQUET

CREST UMR 9194, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal – BP 37203, 35712 BRUZ cedex, France. E-mail: lionel.truquet@ensai.fr

We study some regularity properties in locally stationary Markov models which are fundamental for controlling the bias of nonparametric kernel estimators. In particular, we provide an alternative to the standard notion of derivative process developed in the literature and that can be used for studying a wide class of Markov processes. To this end, for some families of V-geometrically ergodic Markov kernels indexed by a real parameter u, we give conditions under which the invariant probability distribution is differentiable with respect to u, in the sense of signed measures. Our results also complete the existing literature for the perturbation analysis of Markov chains, in particular when exponential moments are not finite. Our conditions are checked on several original examples of locally stationary processes such as integer-valued autoregressive processes, categorical time series or threshold autoregressive processes.

Keywords: local stationarity; time-inhomogeneous Markov chains

1. Introduction

The notion of local stationarity has been introduced in Dahlhaus [2] and offers an interesting approach for the modeling of nonstationary time series for which the parameters are continuously changing with the time. In the literature, several stationary models have been extended to a locally stationary version, in particular Markov models defined by autoregressive processes. See, for instance, Subba Rao [33] Moulines et al. [27] and Zhang and Wu [40] for linear autoregressive processes, Dahlhaus and Subba Rao [5], Fryzlewicz et al. [16] and Truquet [34] for ARCH processes and a recent contribution of Dahlhaus et al. [4] for nonlinear autoregressive processes. In Truquet [35], a new notion of local stationarity was introduced 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 used time-varying Markov kernels. Let $\{Q_u : u \in [0, 1]\}$ be a family of Markov kernels on the same topological space $(E, \mathcal{E})$. We assume that for each $u \in [0, 1]$, Q_u has a unique invariant probability measure denoted by π_u. For an integer $n \geq 1$, we consider n random variables $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ such that

$$\mathbb{P}(X_{n,t} \in A | X_{n,t-1} = x) = Q_{t/n}(x, A), \quad (x, A) \in E \times \mathcal{E}, 1 \leq t \leq n, \tag{1}$$

with the convention $X_{n,0} \sim \pi_0$. Let us observe that $(X_{n,t})_{1 \leq t \leq n}$ is a time-inhomogeneous Markov chain as for the locally stationary autoregressive processes of order 1 introduced in the aforementioned references. Then formulation (1) is quite general for a locally stationary processes having Markov properties (application to p-order Markov process will be also discussed in Section 5,

but as in the homogeneous case, vectorization can be used to get a Markov chain of order 1). The main particularity of our approach, which is similar to that used in the literature of locally stationary processes, is the rescaling by the sample size n, taking $Q_{t/n}$ instead of Q_t for the transition kernel at time t. The aim of this non standard formulation is to overcome a main drawback of the standard large sample theory, from which it is mainly feasible to estimate parametric models, leading to very arbitrary statistical models for the time-varying Markov kernels Q_t. On the other hand, this rescaling allows to use a so-called infill asymptotic, from which local inference of some functional parameters defined on the compact unit interval $[0, 1]$ remains possible. We defer the reader to the monograph of Dahlhaus [3] for a thorough discussion of these asymptotic problems. One of the main issue for making this approach working is to show that the triangular array can be approximated marginally (in a sense to precise) by a stationary process with transition kernel Q_u when the ration t/n is close to a point $u \in [0, 1]$.

The approach used in Truquet [35] is based on Markov chains techniques. Let us introduce some notations. For two positive integers t, j such that $1 \le t \le n + 1 - j$, let $\pi_{t,j}^{(n)}$ be the probability distribution of the vector $(X_{n,t}, \ldots, X_{n,t+j-1})$ and $\pi_{u,j}$ the probability distribution of the vector $(X_1(u), \ldots, X_j(u))$, where $(X_t(u))_{t \in \mathbb{Z}}$ denotes a stationary Markov chain with transition kernel Q_u. Note that $\pi_{u,0} = \pi_u$. It is then possible to study the approximation of $\pi_{t,j}^{(n)}$ by $\pi_{u,j}$ using probability metrics. One of main idea of the paper is to use contraction/regularity properties for the Markov kernels Q_u which guarantee at the same time such approximation and the decay of some specific mixing coefficients. We will recall in Section 4, our approximation result for total variation type norms, from which a large class of locally stationary models can be studied. See also Section 4 in Truquet [35] for examples of such models and for results on their statistical inference.

One of the important issues in the statistical inference of locally stationary processes is the curve estimation of some parameters of the kernels $\{Q_u : u \in [0, 1]\}$. However, some parameters of the joint distributions and their regularity, e.g. $\int g \, d\pi_u$ for some measurable functionals $g : E \to \mathbb{R}$, have their own interest for two reasons.

1. First, one can be interested in estimating specific local parameters such as the trend of a time series (which is here the mean of the invariant probability measure) or the local covariance function $u \mapsto \mathrm{Cov}(X_0(u), X_1(u))$. Nonparametric estimation of such functionals typically require to know their regularity, for instance the number of derivatives. For example, estimating the expectation $m(u) := \int g \, d\pi_u = \mathbb{E}g(X_0(u))$ by a the local linear fit with a kernel density requires the existence of two derivatives for m. See, for instance, Fan and Gijbels [12] for an introduction to local polynomial modeling. We will discuss such a problem in Section 4.3. Typically, if m is k-times continuously differentiable, it can be estimated by a linear estimator $\hat{m}(u) = \sum_{t=1}^n w_{n,t}(u) g(X_{n,t})$ with some weights satisfying satisfying $\sum_{t=1}^n w_{n,t}(u) = 1$, $\sum_{t=1}^n w_{n,t}(u)(u - t/n)^\ell = 0$ for $1 \le \ell \le k - 1$ and we have

$$\mathbb{E}\hat{m}(u) - m(u) = \sum_{i=1}^n w_{n,t}(u)\left[\int g \, d\pi_t^{(n)} - \int g \, d\pi_{t/n}\right]$$

$$+ \sum_{i=1}^n w_{n,t}(u)(u - t/n)^{k-1} \frac{m^{(k-1)}(u + \tau_{n,t}(t/n - u)) - m^{(k-1)}(u)}{(k-1)!},$$

with $\tau_{n,t} \in [0, 1]$. The first term is of order $1/n$ while the second one is bounded by b^k (up to a constant) is the weights are such that $w_{n,t}(u) = 0$ when $|u - t/n| > b$ and $b \in (0, 1)$ is a bandwidth parameter. More regularity then entails a lower bias for the nonparametric estimator.

2. Moreover, as discussed in Truquet [35], Section 4.5, when $Q_u(x, dy) = Q_{\theta(u)}(x, dy)$ for a smooth function $\theta : [0, 1] \to \mathbb{R}^d$, getting a bias expression for the local likelihood estimator of θ requires existence of derivatives for an application of type $u \mapsto \int g \, d\pi_{2,u}$.

The results stated in Truquet [35] only guarantee Lipschitz continuity of such applications. See in particular Proposition 2 of that paper. One of the aim of the present paper is to complete such results by studying differentiability properties.

In the recent work of Dahlhaus et al. [4], the authors study some autoregressive Markov processes with time-varying parameters. These processes are defined by

$$X_{n,t} = F_{t/n}(X_{n,t-1}, \ldots, X_{n,t-p}, \varepsilon_t), \quad 1 \leq t \leq n.$$

Using contraction properties of the random maps $x \mapsto F_u(x, \varepsilon_1)$ in $\mathbb{L}^q$-norms, they study the local approximations of $X_{n,t}$ by a stationary process $(X_t(u))_{t \in \mathbb{Z}}$ where

$$X_t(u) = F_u\big(X_{t-1}(u), \ldots, X_{t-p}(u), \varepsilon_t\big), \quad t \in \mathbb{Z}.$$

Differentiability of some functionals of type $u \mapsto \mathbb{E}f(X_1(u), \ldots, X_j(u))$ for differentiable functions f are then studied through the notion of a derivative process $dX_t(u)/du$ which is an almost sure derivative of the application $u \mapsto X_t(u)$. See Proposition 3.8, Proposition 2.5 and Theorem 4.8 in Dahlhaus et al. [4]. The notion of derivative process is also fundamental for getting a bias expression of the local Maximum Likelihood Estimator. See Theorem 5.4 in Dahlhaus et al. [4] for details.

Note that here, the process $((X_t(u), \ldots, X_{t-p+1}(u)))_{t \in \mathbb{Z}}$ form a Markov chain with transition kernel $Q_{u,p}$ defined for $(x_1, \ldots, x_p) \in E^p$ and $(A_1, \ldots, A_p) \in \mathcal{E}^p$ by

$$Q_{u,p}\big((x_1, \ldots, x_p), A_1 \times \cdots \times A_p\big) = \mathbb{P}\big(F_u(x, \varepsilon_0) \in A_p\big) \prod_{i=2}^{p} \delta_{x_i}(A_{i-1}).$$

The previous functionals are then defined by some integrals of the invariant probability measure or more generally some integrals of other finite-dimensional distributions of the chain. Note also that any finite-dimensional distribution of a Markov chain still corresponds to the invariant probability measure of another Markov chain obtained from a vectorization of the initial stochastic process. Studying differentiability properties of an invariant probability measure depending on a parameter is then an important problem.

For the locally stationary models introduced in Truquet [35], the notion of derivative process is not relevant to evaluate such a regularity in particular when the state space is discrete. For instance, consider the following example of integer-valued time series, the INAR process with time-varying parameters. Its local stationary approximations are defined by

$$X_t(u) = \sum_{i=0}^{X_{t-1}(u)} \mathbb{1}_{\{U_{t,i} \leq \alpha(u)\}} + \varepsilon_t,$$

where $U_{t,i}$, ε_s, $(s, t, i) \in \mathbb{Z}^2 \times \mathbb{N}$ are i.i.d. with $U_{0,0}$ uniform over $[0, 1]$. A more general model will be discussed in Section 6.2. Here $Q_u(x, \cdot)$ is a convolution between the binomial distribution of parameters $(x, \alpha(u))$ and the distribution of ε_0. Of course, the usual derivative of $u \mapsto X_t(u)$ makes no sense here but when the functional parameter α is differentiable, so is $u \mapsto Q_u(x, y)$. We then expect that $u \mapsto \pi_u$ will be also differentiable. The same phenomenon already occurs for the simple case of a finite-state Markov chains with N states, that is, $\mathbb{P}(X_t(u) = y | X_{t-1}(u) = x) = Q_u(x, y)$, $x, y = 1, \ldots, N$. This example will be covered by the results given in Section 6.3.

Our aim with this paper is to study directly existence of derivatives for the applications $u \mapsto \pi_{u,j}$ under suitable regularity assumptions for $u \mapsto Q_u$. These derivatives will be understood in the sense of signed measures and using topologies defined by V-norms, where V denotes a drift function. See below for further details. Our approach will be applicable whatever the state space and can also be interesting for the autoregressive processes studied in Dahlhaus et al. [4]. We defer the reader to the Notes after Proposition 3 for a discussion of the differences between our results and that of Dahlhaus et al. [4] for a time-varying AR(1) process. We also stress that we study differentiability properties of any order whereas Dahlhaus et al. [4] only considered differentiability of order 1. The results given in this paper (see in particular Proposition 2 and Corollary 3) are then an interesting alternative to the existing notion of derivative process.

In what follows, to avoid confusions between the various Markov kernels and invariant probability measures that will appear in the paper and will depend on the targeted finite-dimensional distribution of the Markov chain, we adopt a generic notation and consider a topological space G endowed with its Borel σ-field $\mathcal{G}$ and a family of Markov kernels $\{P_u : u \in [0, 1]\}$ on $(G, \mathcal{G})$. We assume that for any $u \in [0, 1]$, P_u has a unique invariant probability measure denoted by μ_u. For the dynamic (1) and j a positive integer, G can then denote a cartesian product E^j, μ_u the probability measure $\pi_{u,j}$ and P_u the transition kernel $Q_{u,j}$ for which $\pi_{u,j}$ is an invariant probability measure (see (9) for a precise definition of $Q_{u,j}$).

The approach used in this paper has an important connection with the literature of perturbation theory for Markov chains. A central problem in this field is to control an approximation of the invariant probability measure when the Markov kernel of the chain is perturbed. See, for instance, the recent contribution of Rudolf and Schweizer [31], motivated by an application to stochastic algorithms. Many contributions also provide some conditions under which the invariant probability has one or more derivatives with respect to an indexing parameter. See, for instance, Schweitzer [32], Kartashov [22], Pflug [28], Vázquez-Abad and Kushner [38] or Glynn and L'ecuyer [17]. For general state spaces, these contributions only focus on the existence of the first derivative. Higher-order differentiability is studied using operator techniques in Heidergott and Hordijk [19] or Heidergott et al. [20]. However, as we explain below, these results are restrictive for application to standard time series models. Let us first introduce some notations. For a measurable function $V : G \to [1, \infty)$, we denote by $\mathcal{M}_V(G)$ the set of signed measures μ on $(E, \mathcal{E})$ such that

$$\|\mu\|_V := \int V \, d|\mu| = \sup_{|f| \le V} \int f \, d\mu < \infty,$$

where $|\mu|$ denotes the absolute value of the signed measure μ. We recall that $(\mathcal{M}_V(G), \|\cdot\|_V)$ is a Banach space. In this paper, we will study differentiability of $u \mapsto \mu_u$, as an application from $[0, 1]$ to $\mathcal{M}_V(G)$. The function V will be mainly a drift function for the Markov chain, as in the

references mentioned above. We will consider the Markov kernel P_u as an operator T_u acting on $\mathcal{M}_V(E)$, i.e. $T_u \mu = \mu P_u$ is the measure defined by

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

For a measurable function $g : E \to \mathbb{R}$ such that $|g|_V = \sup_{x \in E} \frac{|g(x)|}{V(x)} < \infty$, we set $P_u g(x) = \int P_u(x, dy) g(y)$. The operator norm of the difference $T_u - T_v$ can be defined by the two following equivalent expressions

$$\|T_u - T_v\|_{V,V} := \sup_{\mu \in \mathcal{M}_V(E) : \|\mu\|_V \leq 1} \|\mu(P_u - P_v)\|_V = \sup_{|f|_V \leq 1} |P_u f - P_v f|_V.$$

Differentiability of the application $u \mapsto \mu_u$, considered as an application form $[0, 1]$ to $\mathcal{M}_V(G)$ could be obtained using the results of Heidergott and Hordijk [19] but it is necessary to assume continuity of the application $u \mapsto T_u$ for the previous operator norm. Such continuity assumption is also used in Kartashov [22]. In the literature of perturbation theory, exponential drift functions V are often used and such continuity property can be checked in many examples, such as for some queuing systems considered in Heidergott et al. [20]. However, exponential drift functions require exponential moments for the corresponding Markov chain. In time series analysis, existence of exponential moments is a serious restriction. On the other hand, for power drift functions (another classical choice in the literature of Markov chain), this continuity property often fails. For instance, let us consider the process $X_t(u) = u X_{t-1}(u) + \varepsilon_t$, $u \in (0, 1)$, where $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. integrable random variables having an absolutely continuous distribution with density f_ε. Ferré et al. [14] have shown that the corresponding Markov kernel $P_u(x, dy) = f_\varepsilon(y - ux) \, dy$ is not continuous with respect to u, when the classical drift function $V(x) = 1 + |x|$ is considered. Additional problems also occur in this example for the derivative operators, obtained by taking the successive derivatives of the conditional density, that is, $Q_u^{(\ell)}(x, dy) = (-1)^{(\ell)} x^\ell f_\varepsilon(y - ux) \, dy$, $\ell = 1, 2, \ldots$, which are not bounded operators for the operator norm $\| \cdot \|_{V,V}$. Boundedness of the derivative operators are required in Heidergott and Hordijk [19] or in Heidergott et al. [20] for studying the derivatives of $u \mapsto \pi_u$, as an application from $[0, 1]$ to $\mathcal{M}_V(G)$. Hence, the results of the two previous references cannot be applied here. For studying differentiability of the invariant probability measure, an alternative result can be found in Hervé and Pène [21] (see Appendix A of that paper). This result is applied in Ferré et al. [14] to the AR(1) process. However, it is formulated in a very abstract form, using operator theory and its application to on a general class of Markov chain models has not been discussed. In contrast, we provide an approach for studying derivatives of any finite-dimensional distribution for a wide class of Markov chains. This result has some similarities with that of Hervé and Pène [21] but our assumptions can be more easily checked and slightly better results can be obtained in the examples we will consider in Section 6. We defer the reader to the Notes (3.) after Theorem 1 and to the Notes (3.) after Proposition 3 for a discussion. Additionally, for a Markov chain and more generally a p-order Markov chain, we provide (see Proposition 2 and Corollary 3) easily verifiable conditions on the density of the transition kernels that guarantee differentiability properties for any finite-dimensional distribution of the process. To our knowledge, the existing literature on the perturbation theory of Markov chains does not contain such conditions in a this

general context. Our approach is particularly useful for models for which some power functions satisfy a drift condition. See Section 4.3 and Section 5 for details.

Though we motivate our results for bias reduction in nonparametric regression, we also point out that our results can be interesting for other problems.

1. First, choosing the bandwidth b by cross-validation for parameter estimation in some locally stationary time series models has been recently justified by Richter and Dahlhaus [30]. Their asymptotic results requires smoothness properties for some expectations of the stationary processes $(X_t(u))_{t \in \mathbb{Z}}$. We believe that our results could be used to adapt their approach for bandwidth selection for general models of the form (1), including integer-valued or categorical time series models that are not covered by the previous reference.

2. Our results are stated for locally stationary Markov chains, but one can get a straightforward extension to general parametric models of ergodic Markov processes, using partial derivatives in the multidimensional case. Such modifications will not change the core of our arguments and do not present additional difficulties, we then restrict our study to the case of a parameter $u \in [0, 1]$. In operations research, calibrating queuing systems or inventory levels often requires to optimize the expected value of a performance function $\theta \mapsto \mathbb{E}[g(X_0(\theta))]$ for a Markov chain $(X_t(\theta))_t$ for which the transition operator depends on a parameter $\theta \in \mathbb{R}^k$. Smoothness properties of such functions are then very important. See for instance Kushner and Yin [23], Section 2.4.3. Our results complement the existing results in perturbation theory of Markov chains, which are mainly stated when exponential stability is guaranteed. In particular, we do not exclude many simple Markov chains models such as the autoregressive process of order 1 whenever the noise has no exponential moments.

The paper is organized as follows. In Section 2, we give a general result, formulated using a pure operator-theoretic approach, for getting differentiability properties of an invariant probability measure depending on a parameter. In Section 3, we give a set of sufficient conditions, involving the transition densities of the Markov kernels. We also study differentiability of other finite-dimensional distributions of the Markov chain. Section 4 is devoted to the notion of local stationarity and the control of the bias in kernel smoothing. We also give simple sufficient conditions that ensure both local stationarity and differentiability properties. An extension of our results to p-order Markov processes is proposed in Section 5. Finally, we check our assumptions on several examples of locally stationary processes in Section 6. Some of these examples are new or are p-order extensions of existing Markov chain models. Proofs of some of our results can be found in the Supplementary Material Truquet [36].

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 σ-field $\mathcal{B}(G)$. For an integer $k \geq 1$, let $V_0, V_1, \ldots, V_k$ 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 $F_s = \mathcal{M}_{V_s}(G)$ and $\| \cdot \|_s = \| \cdot \|_{V_s}$ for $0 \leq s \leq k$. We remind

that $\{(F_\ell, \|\cdot\|_\ell) : 0 \le \ell \le k\}$ is a family of Banach spaces. Moreover, $0 \le \ell \le k-1$, 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 $0 \le i \le j \le k$ and a linear operator $T : (F_j, \|\cdot\|_j) \to (F_i, \|\cdot\|_i)$, we set $\|T\|_{j,i} = \sup_{\|\mu\|_j \le 1} \|T\mu\|_i$ and $\|T\|_{0,j,i} = \sup_{\|\mu\|_j \le 1, \mu \in F_{0,j}} \|T\mu\|_i$. Finally, for each $u \in [0, 1]$, we denote by T_u the linear operator acting on the space F_0 defined by $T_u\mu = \mu P_u$. For a positive integer m, T_u^m will denote the iteration of order m of the operator T_u.

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

$$\sup_{u \in [0,1]} \left\| T_u^{m_\ell} \right\|_{\ell,\ell} \le \kappa_\ell, \qquad \sup_{u \in [0,1]} \|T_u\|_{\ell,\ell} < \infty$$

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 any $1 \le \ell \le k$, there exists a continuous linear operator $T_u^{(\ell)} : (F_\ell, \|\cdot\|_\ell) \to (F_0, \|\cdot\|_0)$ such that for $0 \le s \le s + \ell \le k$, $T_u^{(\ell)} F_{s+\ell} \subset F_s$, $\sup_{u \in [0,1]} \|T_u^{(\ell)}\|_{s+\ell,s} < \infty$ and for $\mu \in F_{s+\ell}$, the function $u \mapsto T_u^{(\ell-1)}\mu$ is differentiable as a function from $[0, 1]$ to F_s with continuous derivative $u \mapsto T_u^{(\ell)}\mu$. We use the convention $T_u^{(0)} = T_u$.

Theorem 1. *Assume 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 $0 \le \ell \le k$. Moreover, the inverse of $I - T_u$ is given by $(I - T_u)^{-1} = \sum_{k \ge 0} T_u^k$.*
    • *We have $\max_{0 \le \ell \le k} \sup_{u \in [0,1]} \|(I - T_u)^{-1}\|_{\ell,\ell} < \infty$.*
    • *For $0 \le \ell \le k$ and $\mu \in F_{0,\ell}$, the application $u \mapsto (I - T_u)^{-1}\mu$ is continuous as an application from $[0, 1]$ to F_ℓ.*
    • *Moreover, for each $u \in [0, 1]$, we have for $0 \le \ell \le k-1$ and $\mu \in F_{\ell+1}$,*

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

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 application $\Gamma : [0, 1] \to F_k$ defined by $\Gamma(u) = \mu_u$, for $u \in [0, 1]$, is continuous. Moreover, there exist some functions $\Gamma^{(0)}, \ldots, \Gamma^{(k)}$ such that $\Gamma^{(0)} = \Gamma$ and*

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

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

Notes.

1. When $V_0 = V_1 = \cdots = V_k = V$, existence of the derivatives for the invariant probability measures is studied in Heidergott and Hordijk [19]. One can show that the condition C^k used for stating their result entails A2 because they use a continuity assumption of the derivative 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, the measure and the continuity assumption of the kernel for the V-operator norm entails the contraction A1 (for the contraction coefficient, see Section 3.2 below). We also deduce from our result the following Taylor-Lagrange formula that will be useful for controlling the bias of kernel estimators in Section 4.2. For $u \in [0, 1]$ and $h \in \mathbb{R}$ such that $u + h \in [0, 1]$, set $M = \sup_{v \in [0,1]} \|\Gamma^{(k)}(v)\|_0$. We then have

$$\left\| \Gamma(u + h) - \Gamma(u) - \sum_{\ell=1}^{k-1} \frac{\Gamma^{(\ell)}(u)}{\ell!} h^\ell \right\|_0 \leq \frac{M|h|^k}{k!}. \tag{2}$$

2. Let us discuss our assumptions. Assumption A1 guarantees the stability of the spaces $\mathcal{M}_{V_s}(G)$ by the application T_u (i.e., $\mu \in \mathcal{M}_{V_s}(G) \to \mu P_u \in \mathcal{M}_{V_s}(G)$). The contraction condition in the second part of this assumption guarantees some invertibility properties of the operator $I - T_u$ (see point 1 of Theorem 1) that are needed for getting an expression of the derivatives of $u \mapsto \mu_u$. Our assumptions involve some measure spaces with more and more moment restrictions $\mathcal{M}_{V_k}(G) \subset \cdots \subset \mathcal{M}_{V_0}(G)$. Assumption A2 allows the derivative operators of the Markov kernel to be only bounded for an operator norm involving a weaker final topology. This is particularly useful when the derivatives operators do not preserve a measure space of given regularity. For instance, for the AR(1) process $X_k(u) = a(u)X_{k-1}(u) + \varepsilon_k$ with a noise density f_ε, we have $T_u \mu(dy) = \int \mu(dx) f_\varepsilon(y - a(u)x) \, dy$ and a natural candidate for $T_u^{(\ell)}$ is

$$T_u^{(\ell)} \mu(dy) = a^{(\ell)}(u) \int \mu(dx)(-x)^\ell f_\varepsilon^{(\ell)}(y - a(u)x) \, dy.$$

Setting $V_s = 1 + |x|^s$, one can see that $|T_u^{(\ell)} \mu| \cdot V_s \leq C|\mu| \cdot V_{s+\ell}$ for a positive constant C. This means that μ has to have a moment of order $s + \ell$ for getting a finite upper bound in the previous inequality. This problem does not occur on this example when the V_s's are some exponential functions and the noise density and its derivatives have exponential moments. See in particular Proposition 7 given in the supplementary material. However, we do not want to use this restrictive moment condition.

3. The idea of introducing nested spaces (such as $\mathcal{M}_{V_k}(G) \subset \cdots \subset \mathcal{M}_{V_0}(G)$ in our result) can also be found in Hervé and Pène [21] (see Annex A of that paper). In Proposition A of that paper, the authors study regularity properties of some resolvent operators depending on a parameter, also using an operator theoretic approach. An application of this result to study the regularity of the invariant probability measure of an AR(1) process with respect to its autoregressive coefficient is given in Ferré et al. [14], Proposition 1. However, application of such a result requires in our context to introduce additional operator norms for getting

continuity properties of applications $u \mapsto T_u^{(\ell)}$, as applications form $[0, 1]$ to some spaces of linear operators. See in particular the proof of Proposition 1 in Ferré et al. [14] and the transformation T_0 introduced in the proof of their Lemma 1. Here, in A2, we prefer to use pointwise continuity/differentiability assumptions for some applications $u \mapsto T_u^{(\ell)} \mu$ and that are sufficient for getting our result. We found our formulation easier to understand. We also defer the reader to the Notes (3.) after Proposition 3 for a comparison of our result with that of Ferré et al. [14] for an AR(1) process.

4. In Assumption A2, we assume that the operators $T_u^{(\ell)}$ satisfies some kind of weak continuity or weak differentiability with respect to u, in the sense that continuity and differentiability do not not hold for operator norms but simply for some applications $u \mapsto T_u^{(\ell)} \mu$. In the literature of perturbation of Markov chains, a notion of weak continuity or differentiability for measures depending on parameters can be found in Pflug [29] (see Section 3.2). Our condition is stronger since for an individual measure μ, the application $u \mapsto \mu P_u g$ is required to be continuous or differentiable but uniformly over a class of functions g. In contrast, Pflug [29] defined these notions for a fixed function g. But note that our final result entails existence of derivatives for the topology defined by some V-norms, which is stronger than getting derivatives for $u \mapsto \int g \, d\mu_u$ for a single function g.

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 Assumption A1, the series $\sum_{k \geq 0} T_u^k$, considered as an operator from $F_{0,\ell}$ to $F_{0,\ell}$ is normally convergent for the norm $\|\cdot\|_{0,\ell,\ell}$ and is the inverse of $I - T_u$. Then $I - T_u$ defines an isomorphism on the space $(F_{0,\ell}, \|\cdot\|_\ell)$.

   Using the expression $(I - T_u)^{-1} = \sum_{k \geq 0} T_u^k$, the second assertion is a consequence of Assumption A1.

   Next, we show that for $0 \leq \ell \leq k$ and $\mu \in F_{0,\ell}$, the application $u \mapsto (I - T_u)^{-1} \mu$ is continuous as an application from $[0, 1]$ to $F_{0,\ell}$. Considering all the operators as operators from $F_{0,\ell}$ to $F_{0,\ell}$, we use the decomposition

$$(I - T_{u+h})^{-1} - (I - T_u)^{-1} = (I - T_{u+h})^{-1}(T_{u+h} - T_u)(I - T_u)^{-1}. \tag{3}$$

   From the previous point, we have $\sup_{u \in [0,1]} \|(I - T_u)^{-1}\|_{0,\ell,\ell} < \infty$ and $(I - T_u)^{-1} \mu$ is an element of $F_{0,\ell}$. Moreover, if $v \in F_\ell$, Assumption A1 guarantees the continuity of the application $v \mapsto T_v v$ as an application from $[0, 1]$ to F_ℓ. Using (3), the continuity of the application $u \mapsto (I - T_u)^{-1} \mu$ follows.

   Finally, if $\mu \in F_{0,\ell+1}$, we show that the application $u \mapsto (I - T_u)^{-1} \mu$ is differentiable as an application from $[0, 1]$ to $F_{0,\ell}$. 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

$$h^{-1}\big[(I - T_{u+h})^{-1} x - (I - T_u)^{-1} x\big] = (I - T_{u+h})^{-1} z_{u,h}$$
$$= (I - T_{u+h})^{-1}(z_{u,h} - z_u) + (I - T_{u+h})^{-1} z_u.$$

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,

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

This shows that the application $u \mapsto (I - T_u)^{-1}\mu$ is differentiable, as an application from $[0,1]$ to $F_{0,\ell}$, 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. From Assumption A1 and the fixed point theorem, there exists a unique probability measure μ_u in F_k such that $\mu_u P_u = \mu_u$. But μ_u is in fact the single invariant probability measure for P_u. Indeed, since for any $x \in G$, we have $\delta_x \in F_{k,1}$, the fixed point theorem applied in $F_{k,1}$ entails that $\lim_{n \to \infty} \int g(y)P_u^n(x, dy) = \int g(y)\mu_u(dy)$ for all measurable function $g : G \to \mathbb{R}$ bounded by one. If $\overline{\mu}_u$ is an invariant probability measure, we get from the Lebesgue theorem,

$$\int g(y)\overline{\mu}_u(dy) = \int g(y)P_u^n(x, dy)\overline{\mu}(dx) \to_{n \to \infty} \int g(y)\mu_u(dy).$$

Necessarily, $\overline{\mu}_u = \mu_u$ and μ_u is then the unique invariant probability measure for P_u.

3. We first show that Γ is continuous. We have $\Gamma(u + h) - \Gamma(u) = (I - T_{u+h})^{-1}(T_{u+h} - T_u)\Gamma(u)$. From Assumption A1, we have $\lim_{h \to 0} \|T_{u+h}\Gamma(u) - T_u\Gamma(u)\|_k = 0$. Note that $(T_{u+h} - T_u)\Gamma(u)$ is an element of $F_{0,k}$. Using the second assertion of point 1. of the theorem, we get $\lim_{h \to 0}(\Gamma(u + h) - \Gamma(u)) = 0$ in $(F_{0,k}, \|\cdot\|_k)$.

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

(a) First, we assume that $\ell = 1$. Using the same decomposition as for proving continuity of Γ, we have

$$\frac{\Gamma(u + h) - \Gamma(u)}{h}$$

$$= (I - T_{u+h})^{-1}\frac{T_{u+h} - T_u}{h}\mu_u$$

$$= (I - T_{u+h})^{-1}\left[\frac{T_{u+h} - T_u}{h}\mu_u - T_u^{(1)}\mu_u\right] + (I - T_{u+h})^{-1}T_u^{(1)}\mu_u.$$

Here we consider the operators $T_{u+h} - T_u$ and $T_u^{(1)}$ as operator from F_k to $F_{0,k-1}$. The operators $(I - T_u)^{-1}$, $u \in [0,1]$, are considered as operators from $F_{0,k-1}$ to $F_{0,k-1}$. From 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

$$\lim_{h \to 0} \left\| \frac{\Gamma(u+h) - \Gamma(u)}{h} - \Gamma^{(1)}(u) \right\|_{k-1} = 0,$$

where $\Gamma^{(1)}(u) = (I - T_u)^{-1} T_u^{(1)} \mu_u$. It remains to prove the continuity of $\Gamma^{(1)}$ as an application from $[0, 1]$ to F_{k-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 Γ and of Assumption A2, using the decomposition

$$T_{u+h}^{(1)} \mu_{u+h} - T_u^{(1)} \mu_u = \left[T_{u+h}^{(1)} - T_u^{(1)} \right] \mu_u + T_{u+h}^{(1)} [\mu_{u+h} - \mu_u].$$

This shows the result for $\ell = 1$.

(b) Now let us assume that for $1 \le \ell \le k - 1$, Γ has ℓ derivatives such that for $1 \le s \le \ell$ and $u \in [0, 1]$, the function $\Gamma^{(s)} : [0, 1] \to F_{0,k-s}$ is continuous,

$$\lim_{h \to 0} \left\| \frac{\Gamma^{(s-1)}(u+h) - \Gamma^{(s-1)}(u)}{h} - \Gamma^{(s)}(u) \right\|_{k-s} = 0$$

and

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

For $1 \le s \le \ell$, we set $z_u = T_u^{(s)} \Gamma^{(\ell-s)}(u)$ and we consider $T_u^{(s)}$ as an operator from $F_{k-\ell+s}$ to $F_{k-\ell}$. We are going to show that the application $u \mapsto z_u$ from $[0, 1]$ to $F_{0,k-\ell}$ has a derivative. We have

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

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

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

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

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

Using Assumption A2, we have $\sup_{u \in [0,1]} \| T_u^{(s)} \|_{k-\ell+s-1, k-\ell-1} < \infty$. Then we get

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

Using again Assumption A2, we have

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

This shows that

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

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

Next, we compute the derivative of $u\mapsto y_u=(I-T_u)^{-1}z_u$, as an application from $[0,1]$ to $F_{0,k-\ell-1}$. 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 Assumption A2 and some previous results, we get

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

In what follows, 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 $(F_{k-\ell-1},\|\cdot\|_{k-\ell-1})$,

$$\lim_{h\to 0}\frac{\Gamma^{(\ell)}(u+h)-\Gamma^{(\ell)}(u)}{h}=\Gamma^{(\ell+1)}(u)\quad\text{where }\Gamma^{(\ell+1)}(u)=\sum_{s=1}^{\ell}\binom{\ell}{s}t_u^{(\ell,s)}.$$

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

Finally, using the induction hypothesis, the function $\Gamma^{(\ell+1-s)}$ is continuous as an application from $[0,1]$ to $F_{k-\ell+s-1}$, for each $1\le s\le \ell+1$. The proof of the continuity of $\Gamma^{(\ell+1)}$ is then similar to the proof of the continuity of $\Gamma^{(1)}$.

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

3. Sufficient conditions

Assumptions A1–A2 are expressed using an operator point of view and our aim is to provide sufficient conditions that can be checked on the conditional densities. In what follows, we 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 not depending on u. We also provide a natural expression for the derivative operators $T_u^{(\ell)}$. Assumption B1 given below is related to uniform ergodicity. Since there are several ways of checking this assumption, we discuss it in Section 3.2.

3.1. Another set of assumptions

Let k be a positive integer and $V_k \geq V_{k-1} \geq \cdots \geq V_0$ some measurable applications from G to $[1, \infty)$ such that the following conditions are satisfied.

B1 For $\ell = 0, 1, \ldots, k$, the family of Markov kernels $\{P_u : u \in [0, 1]\}$ is simultaneously V_ℓ-uniformly ergodic, that is, there exists $\kappa_\ell \in (0, 1)$ such that,

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

where the unique invariant probability measure μ_u of P_u satisfies $\mu_u V_k < \infty$.

B2 For all $(x, y) \in G^2$, the function $u \mapsto f(u, x, y)$ is k-times continuously differentiable 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 $0 \leq s \leq s + \ell \leq k$, $u \in [0, 1]$ and $x \in G$,

$$\sup_{u \in [0,1]} \int |\partial_1^{(\ell)} f(u, x, y)| V_s(y)\gamma(x, dy) \leq C V_{s+\ell}(x), \quad (4)$$

$$\lim_{h \to 0} \int |\partial_1^{(k-s)} f(u + h, x, y) - \partial_1^{(k-s)} f(u, x, y)| V_s(y)\gamma(x, dy) = 0. \quad (5)$$

Corollary 1. *The assumptions* B1–B3 *entail the assumptions* A1–A2. *Moreover the conclusions of Theorem* 1 *are valid for the derivative operators*

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

Proof of Corollary 1.

1. We first check A1. If P is a Markov kernel on $(G, \mathcal{B}(G))$, we define the following Dobrushin contraction coefficient

$$\Delta_V(P) := \sup_{\mu \in \mathcal{M}_V(G), \mu \neq 0, \mu(G) = 0} \frac{\|\mu P\|_V}{\|\mu\|_V} = \sup_{x, y \in G, x \neq y} \frac{\|\delta_x P - \delta_y P\|_V}{V(x) + V(y)}. \quad (6)$$

See, for instance, Douc et al. [9], Lemma 6.18 for the second expression. Note also that, with the notations of Section 2, we have if $T\mu = \mu P$, $\|T\|_{0,\ell,\ell} = \Delta_{V_\ell}(P)$.

First, note that from (4) applied with $\ell = 0$, we have $T_u F_s \subset F_s$ and $\sup_{u \in [0,1]} \|T_u\|_{s,s} < \infty$ for $s = 0, 1, \ldots, k$. Moreover, we have the bound

$$\|T_u^n\|_{0,\ell,\ell} = \Delta_{V_\ell}(P_u^n) \leq \sup_{x \in G} \frac{\|\delta_x P_u^n - \mu_u\|_\ell}{V_\ell(x)}. \quad (7)$$

This bound can be found for instance in Rudolf and Schweizer [31], Lemma 3.2. For completeness, we repeat the argument. We have,

$$\Delta_V\left(P_u^n\right) \leq \sup_{x \neq y} \frac{\|\delta_x P_u^n - \delta_y P_u^n\|_V}{V(x) + V(y)} \leq \sup_{x \neq y} \frac{\|\delta_x P_u^n - \mu_u\|_V + \|\delta_y P_u^n - \mu_u\|_V}{V(x) + V(y)},$$

which shows (7), using the inequality $(a + b)/(c + d) \leq \max\{a/c, b/d\}$ which is valid for all positive real numbers a, b, c, d,. This entails the existence of an integer $m_\ell \geq 1$ such that $\sup_{u \in [0,1]} \|T_u^{m_\ell}\|_{0,\ell,\ell} < 1$. It remains to show that if $\mu \in F_\ell$, $u \mapsto T_u \mu$ is continuous, as an application from $[0, 1]$ to F_ℓ. We have

$$\|T_{u+h}\mu - T_u\mu\|_\ell \leq \int\int |\mu|(dx)\gamma(x, dy)V_\ell(y)|f(u + h, x, y) - f(u, x, y)|.$$

We will use the Lebesgue theorem. Using the inequality $V_\ell \leq V_k$ and Assumption B3 (5) with $s = k$,

$$\lim_{h \to 0} c_h(u, x) := \int \gamma(x, dy)V_\ell(y)|f(u + h, x, y) - f(u, x, y)| = 0, \quad x \in G.$$

Moreover, from B3 (4) applied to the derivative of order 0, we have $c_h(u, x) \leq 2CV_\ell(x)$ and V_ℓ is $|\mu|$-integrable. The Lebesgue theorem then applies and gives $\lim_{h \to 0} T_{u+h}\mu = T_u\mu$ in F_ℓ and the last assertion in A1 follows.

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 from B3 (4),

$$\left\|T_u^{(\ell)}\mu\right\|_s \leq \int |\mu|(dx) \int \gamma(x, dy)|\partial_1^{(\ell)} f(u, x, y)|V_s(y)$$

$$\leq C \int |\mu|(dx)V_{s+\ell}(x) = C\|\mu\|_{s+\ell}.$$

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 $u \mapsto T_u^{(\ell)}\mu$, as an application from $[0, 1]$ to $F_{s+\ell}$. We have

$$\left\|T_{u+h}^{(\ell)}\mu - T_u^{(\ell)}\mu\right\|_s \leq \int |\mu|(dx) \int \gamma(x, dy)|\partial_1^{(\ell)} f(u + h, x, y) - \partial_1^{(\ell)} f(u, x, y)|V_s(y).$$

From the assertion (4) in B3 and the Lebesgue theorem, it is enough to prove that for all $x \in G$,

$$\lim_{h \to 0} \int \gamma(x, dy)|\partial_1^{(\ell)} f(u + h, x, y) - \partial_1^{(\ell)} f(u, x, y)|V_s(y) = 0.$$

We consider two cases. If $s + \ell = k$, this continuity is a direct consequence of the assertion (5) of Assumption B3. Now if $s + \ell + 1 \leq k$, the result follows from the assumption B3 (4)

and the bound

$$\int \gamma(x, dy) \left| \frac{\partial_1^{(\ell)} f(u+h, x, y) - \partial_1^{(\ell)} f(u, x, y)}{h} \right| V_s(y)$$

$$\leq \sup_{v \in [0,1]} \int \gamma(x, dy) \left| \partial_1^{(\ell+1)} f(v, x, y) \right| V_s(y).$$

Finally, we show the differentiability property of the operators. For $\mu \in F_{s+\ell}$, we have, using the mean value theorem,

$$\left\| \frac{T_{u+h}^{(\ell-1)} \mu - T_u^{(\ell-1)} \mu}{h} - T_u^{(\ell)} \mu \right\|_s$$

$$\leq \sup_{v \in [u, u+h]} \int |\mu|(dx) \int \gamma(x, dy) \left| \partial_1^{(\ell)} f(v, x, y) - \partial_1^{(\ell)} f(u, x, y) \right| V_s(y).$$

The result follows by using the same arguments as in the proof of the continuity of the application $u \mapsto T_u^{(\ell)} \mu$. This completes the proof of Corollary 1. $\square$

Note. When $\phi : G \to [1, \infty)$ is a measurable function such that for some $d \leq d_0$, $0 \leq \ell \leq k$ and $q_0, q_1, \ldots, q_k > 0$, $\int \gamma(x, dy) |\partial_1^{(\ell)} f(u, x, y)| \phi(y)^d \leq C \phi(x)^{d+q_\ell}$, Assumption B3 (4) is checked by setting $V_\ell(x) = \phi(x)^{d+q\ell}$ with $q = \max(q_1, q_2/2, \ldots, q_k/k)$ and assuming that $d + qk \leq d_0$.

3.2. Simultaneous uniform ergodicity

Assumption B1 is related to a simultaneous V-uniform ergodicity condition. Let us first give a precise definition of this notion.

Definition 1. We will say that a family of Markov kernel $\{P_u : u \in [0, 1]\}$ satisfies a simultaneous V-uniform ergodicity condition if there exists $C > 0$ and $\kappa \in (0, 1)$ such that for all $u \in [0, 1]$ and all $x \in G$,

$$\left\| \delta_x P_u^n - \mu_u \right\|_V \leq C V(x) \kappa^n.$$

This notion plays a central rule in our results and it is then important to provide sufficient conditions for B1. We also point out that this notion of simultaneous uniform ergodicity replaces stronger assumptions made in Heidergott and Hordijk [19]. These authors used pointwise uniform ergodicity and a continuity property for the application $u \mapsto P_u$, in the sense that

$$\lim_{h \to 0} \| P_{u+h} - P_u \|_{V,V} = 0.$$

See in particular Definition 3 and Condition 1–4 of that paper. We justify why the two previous conditions imply simultaneous uniform ergodicity in Proposition 8 given in the supplementary material.

For a single Markov kernel, V-uniform ergodicity is generally obtained under a drift condition and a small set condition. See Meyn and Tweedie [26], 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 (η, ν)-small set, where η a positive real number and ν a probability measure on $(G, \mathcal{B}(G))$ if $P(x, A) \geq \eta \nu(A)$, for all $A \in \mathcal{B}(G)$ and all $x \in C$.

We now present two approaches for getting simultaneous uniform ergodicity.

3.2.1. *Simultaneous V-uniform ergodicity via drift and small set conditions*

When simultaneous drift and small set conditions are satisfied, a result of Hairer and Mattingly [18] can be used to check simultaneous V-uniform ergodicity. For simplicity, we introduce the following condition.

Definition 2. For $\lambda \in (0, 1)$, $b, \eta, r > 0$ and ν a probability measure on $(G, \mathcal{B}(G))$, we will say that a Markov kernel P satisfies the condition $\mathcal{C}(V, \lambda, b, r, \eta, \nu)$ if

$$PV \leq \lambda V + b \quad \text{and} \quad \{x \in G : V(x) \leq r\} \quad \text{is a } (\eta, \nu) \text{ small set.} \tag{8}$$

The following result guarantees simultaneous uniform ergodicity.

Lemma 1. *Assume that*

- *there exists an integer $m \geq 1$ such that all the Markov kernels P_u^m, $u \in [0, 1]$, satisfy the condition $\mathcal{C}(V, \lambda, b, r, \eta, \nu)$ for $r > \frac{2b}{1-\lambda}$,*
- *there exists $K > 0$ such that $P_u V \leq KV$ for all $u \in [0, 1]$.*

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

Proof of Lemma 1. Using Theorem 1.3 in Hairer and Mattingly [18], our assumptions entail 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 $V_\delta = 1 + \delta V$ (see (6) for the definition of Δ_V). The result of Hairer and Mattingly [18] is in fact stated for a single Markov kernel but inspection its proof shows that the coefficients α and δ only depends on λ, b, r and η and the later constants are the same for the P_u's. Extension of this result to a family of Markov kernels $\{P_u : u \in [0, 1]\}$ is then immediate.

Next, using the equivalence of the norms $\|\cdot\|_V$ and $\|\cdot\|_{V_\delta}$, one can show as in Proposition 2 in Truquet [35] that there exists $C > 0$ and $\rho \in (0, 1)$ such that $\sup_{u \in [0,1]} \|\delta_x P_u^n - \pi_u\|_V \leq CV(x)\rho^j$. This completes the proof of the lemma. $\qquad \square$

The following result will be particularly important for checking simultaneous uniform ergodicity when the V_s's are a power of a given function ϕ, the most interesting case for our examples.

Proposition 1. *Let $\phi : G \to [1, \infty)$ be a measurable function and $V_s = \phi^{q_s}$ for $s = 0, \ldots, k$ and $0 < q_0 < \cdots < q_k$. One can then obtain simultaneous V_s-uniform ergodicity for $s = 0, 1, \ldots, k$ if*

1. *there exists a positive real number K such that for all $u \in [0, 1]$, $P_u V_k \le K V_k$,*
2. *there exist an integer $m \ge 1$, two real numbers $\lambda \in (0, 1)$, $b > 0$, a family of positive real numbers $\{\eta_r : r > 0\}$ and a family $\{\nu_r : r > 0\}$ of probability measures on G such that for all $r > 0$ and all $u \in [0, 1]$, the Markov kernel P_u^m satisfies Condition $\mathcal{C}(V_k, \lambda, b, r, \eta_r, \nu_r)$.*

Proof of Proposition 1. Let $s \in \{0, 1, \ldots, k - 1\}$. From the assumptions and Jensen's inequality, we have $P_u V_s \le K^{q_s/q_k} V_s$ for any $s = 0, \ldots, k$. Moreover for any $r > 0$, the family of Markov kernels $\{P_u^m : u \in [0, 1]\}$ satisfies Condition $\mathcal{C}(V_s, \lambda^{q_s}, b^{q_s}, r^{q_s/q_k}, \eta_r, \nu_r)$. The result is then a consequence of Lemma 1.   □

3.2.2. *Other approach*

Simultaneous uniform ergodicity can also be obtained from other conditions. For instance, if each kernel P_u is V-uniformly ergodic, then perturbation methods can be applied to get a local simultaneous V-uniform 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, but satisfies some weaker continuity properties, Ferré et al. [14] (see Theorem 1) obtained a nice result based on the Keller-Liverani perturbation theorem. However, construction of locally stationary Markov chain models considered in Truquet [35] is based on the simultaneous drift and small set conditions and we will not use this approach in the rest of the paper. The interested reader is deferred to Section 1 in the supplementary material for a discussion.

3.3. Regularity of higher-order finite dimensional distributions

We now study existence of some derivatives for $u \mapsto \pi_{u,j}$ ($j \ge 2$) in model (1). We remind that $\pi_{u,j}(d\mathbf{x}) = \pi_u(dx_1) Q_u(x_1, dx_2) \cdots Q_u(x_{j-1}, dx_j)$. For $x_1, \ldots, x_j \in E$ and $0 \le \ell \le k$, we set $V_{\ell,j}(x_1, \ldots, x_j) = \sum_{i=1}^{j} V_\ell(x_i)$. For an integer $j \ge 1$, we denote by $\mathcal{M}_V(E^j)$ the space of signed measures on E^j such that

$$\|\mu\|_V := \sup\left\{ \int f \, d\mu : \left| f(x_1, \ldots, x_j) \right| \le V(x_1) + \cdots + V(x_j) \right\}.$$

Finally, let $M_\ell(x_1) = \sup_{u \in [0,1]} \int |\partial_1^{(\ell)} f(u, x_1, y_1)| \gamma(x_1, dy_1)$.

The following additional assumption will be needed.

B4 There exists $C > 0$ such that for $0 \le s \le s + \ell \le k$ and all $x_1, x_2 \in E$, we have

$$V_s(x_1) M_\ell(x_2) \le C\left(V_{s+\ell}(x_1) + V_{s+\ell}(x_2) \right).$$

Constant C can be the same as in assumption B2, this is why we use the same notation. The following result is a consequence of Corollary 1.

Corollary 2. *Let $\{Q_u : u \in [0, 1]\}$ be a family of Markov kernels on E satisfying the assumptions B1–B4. Then the application $u \mapsto \pi_{u,j}$ from $[0, 1]$ to $\mathcal{M}_{V_0}(E^j)$, is k-times continuously differentiable.*

Note. Assumption B4 will be satisfied if there exists a function $\phi : E \rightarrow [1, \infty)$ such that

$$\int \phi(y_j)^d |\partial_1^{(\ell)} f(u, x_j, y_j)| \gamma(x_j, dy_j) \leq C\phi(x_j)^{d+r\ell}$$

for $0 \leq d \leq d_0$. Indeed in this case, one can take (up to a constant) $V_\ell(x_j) = \phi(x_j)^{d+r\ell}$ with $r = \max(r_1, r_2/2, \ldots, r_k/k)$ and k such that $d + rk \leq d_0$.

Proof of Corollary 2. Here we set for $\mathbf{x} \in E^j$ and $A \in \mathcal{E}^{\otimes j}$,

$$Q_{u,j}(\mathbf{x}, A) = \int_A f(u, x_j, y_j)\gamma_j(\mathbf{x}, d\mathbf{y}) \quad \text{with } \gamma_j(\mathbf{x}, d\mathbf{y}) = \gamma(x_j, dy_j) \prod_{i=1}^{j-1} \delta_{x_{i+1}}(dy_i). \qquad (9)$$

Let us first check that $Q_{u,j}$ satisfies assumption B1. Let $1 \leq s \leq k$. For an integer $h \geq j$ and a measurable function $g : E^j \rightarrow \mathbb{R}$ such that $|g| \leq V_{s,j}$, we have

$$|Q_{u,j}^j g(\mathbf{x})| \leq \int |g(y_1, \ldots, y_j)| Q_u(x_j, dy_1) Q_u(y_1, dy_2) \cdots Q_u(y_{j-1}, dy_j)$$

$$\leq \sum_{i=1}^{j} Q_u^i V_s(x_j) \leq C_j V_s(x_j),$$

with $C_j = \sum_{i=1}^{j} C^i$ and C defined in (4). We then get

$$\sup_{|g| \leq V_{s,j}} |Q_{u,j}^h g(\mathbf{x}) - \pi_{u,j}g| \leq \sup_{|\overline{g}| \leq C_j V_s} |Q_u^{h-j}\overline{g}(x_j) - \pi_u\overline{g}| \leq C_j \sup_{|\overline{g}| \leq V_s} |Q_u^{h-j}\overline{g}(x_j) - \pi_u\overline{g}|.$$

From the simultaneous V_s-uniform ergodicity property for $\{Q_u : u \in [0, 1]\}$, the previous bounds entail automatically 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 (5). Let us check assumption B3 (4). We have

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

Using assumption B4, we have $V_s(x_i) M_\ell(x_j) \leq C(V_{s+\ell}(x_i) + V_{s+\ell}(x_j))$ and B3 (4) is also satisfied for the family $\{Q_{u,j} : u \in [0, 1]\}$. This completes the proof. $\qquad \square$

4. Locally stationary Markov chains

In this section, we consider a topological space E endowed with its Borel σ-field $\mathcal{B}(E)$ and a triangular array of Markov chains $\{X_{n,t} : 1 \leq t \leq n, n \geq 1\}$ defined by (1).

4.1. Some results about locally stationary Markov chains

We first recall some results obtained in Truquet [35]. 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 $1 \le i, j \le m$.

L1 There exist a measurable function $V : E \to [1, \infty)$, an integer $m \ge 1$, some positive real numbers ϵ, K, λ, b, r, η with $\lambda < 1$, $r > 2b/(1 - \lambda)$ 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 Condition $\mathcal{C}(V, \lambda, b, r, \eta, \nu)$. Moreover, there exists $K > 0$ such that $Q_u V \le K V$ for all $u \in [0, 1]$.

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

L3 For all $(u, v) \in [0, 1]^2$, we have

$$\|\delta_x Q_u - \delta_x Q_v\|_1 \le L(x)|u - v|, \quad \text{with} \sup_{\substack{u \in [0,1] \\ 1 \le \ell' \le \ell}} \mathbb{E}\big[L\big(X_\ell(u)\big) V\big(X_{\ell'}(u)\big)\big] < \infty.$$

Here, $(X_t(u))_{t \in \mathbb{Z}}$ denotes a stationary time-homogeneous Markov chain with transition kernel Q_u.

Under the conditions L1–L3, it is shown in Truquet [35] (see Theorem 3) that for all integer $j \ge 1$, the distribution $\pi_{t,j}^{(n)}$ of $(X_{n,t}, \ldots, X_{n,t+j-1})$ satisfies

$$\big\|\pi_{t,j}^{(n)} - \pi_{u,j}\big\|_V \le C_j \left[\left|u - \frac{t}{n}\right| + \frac{1}{n}\right], \tag{10}$$

where $C_j > 0$ does not depend on u, n, t and $V_j(x_1, \ldots, x_j) = V(x_1) + \cdots + V(x_j)$. Note that under Assumption L1 entails simultaneous V-uniform ergodicity for the family $\{Q_u : u \in [0, 1]\}$ (see Section 3.2.1). Condition L1 is useful to guarantee some β-mixing properties for the triangular array. See Proposition 3 in Truquet [35] for details. Note also that condition L2 is always satisfied for $(V, V') = (V_0, V_1)$ if assumption B3 (4) holds true. In Truquet [35], Proposition 2 and its proof, it has been shown that Assumptions L1–L3 entail, for each $u \in [0, 1]$, geometric ergodicity of a Markov chain with transition kernel Q_u. Moreover, the finite dimensional distribution $\pi_{u,j}$ are shown to be Lipschitz with respect to u, when the space of signed measure on E^j is endowed with the V-norm. However higher-order regularity (such as differentiability) has not been studied and this is precisely the aim of this paper.

For more clarity, we introduce the following terminology.

Definition 3. A triangular array of Markov chains $\{X_{n,t} : 1 \le t \le n, n \ge 1\}$ associated to a family of Markov kernel $\{Q_u : u \in [0, 1]\}$ will be said V-locally stationary if (10) is satisfied.

4.2. Simple sufficient conditions

In order to check more easily our assumptions for specific examples, we give below a set of conditions that guarantee, for the same topology, local stationarity as well as differentiability of

the applications $u \mapsto \pi_{u,j}$ for $j \geq 1$. In particular, the following set of assumptions will imply at the same time L1–L3 and B1–B3. Proposition 2 given below is then important for practical applications of our results to locally stationary Markov models. We only consider the case of power functions, that is, for each integer s, V_s is a power of a measurable function $\phi : E \to [1, \infty)$. This is the most interesting case in practice.

SC1 There exist an integer $m \geq 1$, some positive real numbers d_0, ϵ, K, λ, b with $\lambda < 1$, $d_0 \geq 1$, a family of positive real number $\{\eta_r : r > 0\}$ and a family $\{\nu_r : r > 0\}$ of probability measures on E such that for all $r > 0$ and 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 Condition $\mathcal{C}(\phi^{d_0}, \lambda, b, r, \eta_r, \nu_r)$. Moreover, there exists $K > 0$ such that $Q_u V \leq KV$ for all $u \in [0, 1]$.

SC2 There exists an integer $k \geq 1$ such that for all $(x, y) \in E^2$, the function $u \mapsto f(u, x, y)$ is k-times continuously differentiable.

SC3 There exist some real numbers $d_1 > 0$ and $q \geq 0$ such that $d_1 + kq \leq d_0$ and for all $1 \leq \ell \leq k$ and $d \leq d_1 + (k - \ell)q$,

$$\int \phi^d(y) \big|\partial_1^{(\ell)} f(u, x, y)\big| \gamma(x, dy) \leq C\phi^{d + q\ell}(x).$$

Moreover, for $s = 0, \ldots, k$,

$$\lim_{h \to 0} \int \phi^{d_1 + qs}(y) \big|\partial_1^{(k-s)} f(u + h, x, y) - \partial_1^{(k-s)} f(u, x, y)\big| \gamma(x, dy) = 0.$$

Proposition 2. *Assume that* SC1–SC3 *hold true. Set* $V_0 = \phi^{d_1}$. *The triangular array of Markov chain* $\{X_{n,t} : 1 \leq t \leq n, n \geq 1\}$ *is* V_0-*locally stationary. Moreover, for any integer* $j \geq 1$, *the application* $u \mapsto \pi_{u,j}$, *from* $[0, 1]$ *to* $\mathcal{M}_{V_0}(E^j)$, *is* k-*times continuously differentiable.*

Proof of Proposition 2. For $s = 0, \ldots, k$, we set $V_s = \phi^{d_1 + qs}$. Note that from SC1, Assumption L1 is automatically satisfied for each function V_s, $s = 0, \ldots, k$. This is a consequence of Proposition 1.

Moreover, from SC3 (set $d = d_1$ and $\ell = 1$), Assumption L2 holds true for $V = V_0$ and $V' = V_1$.

Next, we check L3. Using SC3 with $d = 0$ and $\ell = 1$, we see that one can choose $L = C\phi^q$. Setting $V = V_0$, we know from L1–L2 that $\sup_{u \in [0,1]} \int \phi^{d_1 + q} d\pi_u < \infty$. See Truquet [35], Proposition 2. This shows that the integrability condition in L3 is satisfied. The proof of local stationary then follows.

We next check B1–B4. B1 follows from L1 which holds true for all the functions V_s, $s = 0, \ldots, k$. See the discussion of Section 3.2 for details. Finally, B2–B4 follow directly from SC2–SC3. See also the Note after Corollary 2 for checking B4. Differentiability of the marginal distributions then follows from Corollary 2. The proof is now complete. $\qquad \square$

4.3. Application to bias control in nonparametric estimation

In this section, we discuss why differentiability properties of the application $u \mapsto \pi_{u,j}$ are fundamental for controlling the bias in nonparametric estimation of some parameter curves. Let

$\{X_{n,t} : 1 \le t \le n, n \ge 1\}$ be a triangular array of V-locally stationary Markov chains. For a given integer $1 \le j \le n$ and $1 \le t \le n - j + 1$, set $Z_{n,t} = (X_{n,t}, \ldots, X_{n,t+j-1})$. We also assume that the application $u \mapsto \pi_{u,j}$ is k-times continuously differentiable, as an application from $[0, 1]$ to $\mathcal{M}_V(E^j)$. If Assumptions SC1–SC3 are satisfied, Proposition 2, given in the previous section, guarantees V_0-local stationarity and also that $u \mapsto \pi_{u,j}$ is k-times continuously differentiable, as an application from $[0, 1]$ to $\mathcal{M}_{V_0}(E^j)$.

Let $g : E^j \to \mathbb{R}$ be a measurable function such that $|g|_V < \infty$. We want to estimate the quantity $\psi_g(u) = \int g \, d\pi_{u,j}$ using local polynomials. We precise that the approach used here is very classical in nonparametric estimation and, except for the local approximation, is identical to that used for i.i.d. data. See Tsybakov [37], Section 1.8, for a general approach for studying of the bias of local polynomial estimators. 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 $\hat{\psi}_g(u)$ of $\psi_g(u)$ is given by the first component of the vector

$$\hat{\mathcal{H}}_g(u) := \left(\hat{\psi}_g(u), b\hat{\psi}'_g(u), \ldots, b^{k-1}\hat{\psi}_g^{(k-1)}(u) \right)'$$

$$= \arg \min_{\alpha_0, \ldots, \alpha_{k-1} \in \mathbb{R}} \sum_{t=1}^{n} K_b\left(u - \frac{t}{n} \right) \left[g(Z_{n,t}) - \sum_{i=0}^{k-1} \alpha_i \frac{(t/n - u)^i}{b^i i!} \right]^2.$$

For $1 \le t \le n$, we set $v_t(u) = (1, \frac{t/n - u}{b}, \ldots, \frac{(t/n-u)^{k-1}}{b^{k-1}(k-1)!})'$ and

$$D(u) = \sum_{t=1}^{n-j+1} \frac{K_b(t/n - u)}{n - j + 1} v_t(u) v_t(u)', \qquad \hat{N}_g(u) = \sum_{t=1}^{n-j+1} \frac{K_b(t/n - u)}{n - j + 1} v_t(u) f(Z_{n,t}).$$

From (10), we have $\max_{1 \le t \le n-j+1} \sup_{|g|_V \le 1} |\mathbb{E}g(Z_{n,t}) - \psi_g(t/n)| = O(1/n)$. Next, setting $\mathcal{H}_g(u) = (\psi_g(u), b\psi'_g(u), \ldots, b^{k-1}\psi_g^{(k-1)}(u))'$ and using the differentiability properties of ϕ, one can apply the bound (2). There exists $C > 0$ such that for all $n \ge 1$, $1 \le t \le n - j + 1$ and $u \in [0, 1]$,

$$\sup_{|g|_V \le 1} \left| \psi_g(t/n) - \mathcal{H}_g(u)' v_t(u) \right| \le C(u - t/n)^k.$$

We deduce that $\sup_{u \in [0,1]} \sup_{|g|_V \le 1} |\mathbb{E}\hat{N}_g(u) - D(u)\mathcal{H}_g(u)| = O(b^k + \frac{1}{n})$. The rest of the proof consists in bounding the matrix $D(u)^{-1}$ using very classical arguments available in the literature. Using our assumptions on the kernel and on the design $X_i = i/n$, the assumptions LP(1)-LP(3) of Tsybakov [37] are satisfied and Lemma 1.5 and Lemma 1.7 in Tsybakov [37] guaranty that $\max_{u \in [0,1]} \|D(u)^{-1}\| = O(1)$. Then we get

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

In conclusion, up to a term of order $1/n$ which is negligible and can be interpreted as a deviation term with respect to stationarity, the bias is of order b^k when ψ_g is k-times continuously differentiable. We then recover a classical property of local polynomial estimators.

Notes.

1. We will not discuss the variance of the estimator $\hat{\mathcal{H}}_g(u)$. As shown in Truquet [35], Proposition 3, Assumption SC1 ensures geometric β-mixing properties for the triangular array of Markov chains $\{X_{n,t} : 1 \leq t \leq n, n \geq 1\}$. Using standard arguments, one can then show that such variance is of order $1/nb$, as usual for nonparametric curve kernel estimators. Since this problem is not the scope of this paper, we omit the details.
2. Differentiability of $u \mapsto \pi_{u,2}$ is also important for deriving an expression of the bias for the local maximum likelihood estimator of some parameter curves. We defer the reader to Section 4.5 in Truquet [35] for a discussion of this problem.

5. Extension to *p*-order Markov chains

Let us now give an extension of our results to p-order Markov processes. We choose here to present a version which can be applied directly to the examples of the last section of the paper. We consider a family $\{R_u : u \in [0, 1]\}$ of probability kernel from $(E^p, \mathcal{E}^{\otimes p})$ to $(E, \mathcal{E})$. We assume that for $u \in [0, 1]$, $R_u(\mathbf{x}, A) = \int f(u, \mathbf{x}, y)\gamma(dy)$, for a measurable function $f : [0, 1] \times E^{p+1} \to \mathbb{R}$ and a measure γ on E. We also consider a triangular array $\{Y_{n,t} : 1 \leq t \leq n, n \geq 1\}$ of p-order Markov processes such that

$$\mathbb{P}(Y_{n,t} \in A | Y_{n,t-1}, \ldots, Y_{n,t-p}) = R_{t/n}(Y_{n,t-p}, \ldots, Y_{n,t-1}, A), \quad A \in \mathcal{B}(E), 1 \leq t \leq n.$$

For simplicity, we also define a sequence $(Y_{n,t})_{t \leq 0}$ which is a time-homogeneous Markov process with transition kernel R_0. Note that setting $X_{n,t} = (Y_{n,t-p+1}, \ldots, Y_{n,t})$, one can define a triangular array $\{X_{n,t} : 1 \leq t \leq n, n \geq 1\}$ of Markov chains. To this end, let Q_u be the Markov kernel on E^p defined by

$$Q_u(\mathbf{x}, d\mathbf{y}) = R_u(\mathbf{x}, dy_p) \prod_{i=1}^{p-1} \delta_{x_{i+1}}(dy_i).$$

We then have $\mathbb{P}(X_{n,t} \in A | X_{n,t-1}) = Q_{t/n}(X_{n,t-1}, A)$, $A \in \mathcal{E}^{\otimes p}$, $1 \leq t \leq n$. One can then use the results available for locally stationary Markov chains to define and study some locally stationary Markov processes of order $p \geq 2$. For convenience, we give below a set of assumptions on the family of kernels $\{R_u : u \in [0, 1]\}$ which ensure local stationary and differentiability properties of $u \mapsto \pi_{u,j}$ for Markov chains with local transition kernels $\{Q_u : u \in [0, 1]\}$. These properties will be derived using Proposition 1.

Now let $\phi : E \to [1, \infty)$ a measurable function satisfying the following properties.

SCp1 There exists a real number $d_0 \geq 1$ and some positive real numbers $\alpha_{1,u}, \ldots, \alpha_{p,u}, \alpha_0$ such that $\sup_{u \in [0,1]} \sum_{i=1}^p \alpha_{i,u} < 1$ and

$$R_u \phi^{d_0}(\mathbf{x}) \leq \sum_{i=1}^p \alpha_{i,u} \phi^{d_0}(x_i) + \alpha_0, \quad \mathbf{x} \in E^p, u \in [0, 1].$$

Moreover, for each $r > 0$, there exist a positive real number η_r and a probability measure ν_r on E such that,

$$R_u(\mathbf{x}, A) \geq \eta_r \nu_r(A), \quad A \in \mathcal{E}, \ \max_{1 \leq i \leq p} \phi(x_i)^{d_0} \leq r.$$

SCp2 There exists an integer $k \geq 1$ such that for all $(\mathbf{x}, y) \in E^p \times E$, the function $u \mapsto f(u, \mathbf{x}, y)$ is k-times continuously differentiable.

SCp3 There exist some real numbers $d_1 > 0$ and $q \geq 0$ such that $d_1 + kq \leq d_0$ and for all $1 \leq \ell \leq k$ and $d \leq d_1 + (k - \ell)q$,

$$\int \phi^d(y) \big| \partial_1^{(\ell)} f(u, \mathbf{x}, y) \big| \gamma(dy) \leq C \sum_{i=1}^p \phi^{d + q\ell}(x_i).$$

Moreover, $\lim_{h \to 0} \int \phi^{d_1 + qs}(y) |\partial_1^{(k-s)} f(u + h, \mathbf{x}, y) - \partial_1^{(k-s)} f(u, \mathbf{x}, y)| \gamma(dy) = 0$.

Corollary 3. *Assume that assumptions* SCp1–SCp3 *hold true. The triangular array of Markov chains* $\{X_{n,k} : 1 \leq k \leq n, n \geq 1\}$ *is* V_0-*locally stationary, with* $V_0(x_1, \dots, x_p) = \sum_{i=1}^p \phi^{d_1}(x_i)$. *Moreover, for each integer* $j \geq 1$, *the finite dimensional distribution* $u \mapsto \pi_{u,j}$ *of the Markov chains with transition* Q_u *it* k-*times continuously differentiable, as an application from* $[0, 1]$ *to* $\mathcal{M}_{V_0}(E^{pj})$.

6. Examples

In this section, we consider several examples of locally stationary Markov processes satisfying our assumptions and for which some parameter curves $u \mapsto \int g \, d\pi_{u,j}$ $(j \geq 1)$ can be estimated with local polynomials as explained in Section 4.3. We precise that our goal is not to estimate some parameter curves for the Markov kernel $Q_u = Q_{\theta(u)}$. However, as explained in Section 4.3, the results stated below are essential for getting an expression of the bias for minimum contrast estimators of $\theta(\cdot)$. With respect to the examples discussed in Truquet [35], Section 6.3 provide a new example of locally stationary processes whereas Section 6.1 and Section 6.2 give extensions to the order p of some existing models. A comparison of our results with that of Dahlhaus et al. [4] is given in Section 6.1.

6.1. Nonlinear autoregressive process

We consider the following real-valued autoregressive process

$$X_{n,t} = m(t/n, X_{n,t-1}, \dots, X_{n,t-p}) + \sigma(t/n)\varepsilon_t, \quad 1 \leq t \leq n,$$

where $m : [0, 1] \times \mathbb{R}^p \to \mathbb{R}$ and $\sigma : [0, 1] \to \mathbb{R}_+$ are two measurable functions and $(\varepsilon_t)_{t \in \mathbb{Z}}$ is a sequence of i.i.d. random variables. In what follows, we set $E = \mathbb{R}$ and for $\mathbf{y} \in \mathbb{R}^p$, $|\mathbf{y}| = \sum_{i=1}^p |y_i|$. We will use the following assumptions.

E21 The function $u \mapsto \sigma(u)$ is k-times continuously differentiable. Moreover, $\sigma_- := \inf_{u \in [0,1]} \sigma(u) > 0$ and $\max_{0 \le \ell \le k} \sup_{u \in [0,1]} |\sigma^{(\ell)}(u)| < \infty$.

E22 For all $\mathbf{y} \in \mathbb{R}^p$, the function $u \mapsto m(u, \mathbf{y})$ is k-times continuously differentiable. Moreover, there exists a family of nonnegative real numbers $\{\beta_{i,u} : 1 \le i \le p, u \in [0,1]\}$ such that $\sup_{u \in [0,1]} \sum_{i=1}^p \beta_{i,u} < 1$ and four positive real numbers β_0, q', C_1, C_2 such that for all $(u, \mathbf{y}) \in [0,1] \times \mathbb{R}^p$,

$$\left| m(u, \mathbf{y}) \right| \le \sum_{i=1}^p \beta_{i,u} |y_i| + \beta_0, \quad \max_{1 \le \ell \le k} \sup_{u \in [0,1]} \left| \partial_1^{(\ell)} m(u, \mathbf{y}) \right| \le C_1 |\mathbf{y}|^{q'} + C_2.$$

E23 The noise ε_1 has a moment of order d_0 such that $d_0 - q'k > 0$ and has a density f_ε, k-times continuously differentiable, positive everywhere and such that

$$\int |y|^{d_0 + (1 - q')s} \left| f_\varepsilon^{(s)}(y) \right| dy < \infty, \quad s = 0, \dots, k.$$

Setting $R_u(\mathbf{x}, dy) = \frac{1}{\sigma(u)} f_\varepsilon\left(\frac{y - m(u, \mathbf{x})}{\sigma(u)}\right) dy$, the family $\{Y_{n,k} : 1 \le k \le n, n \ge 1\}$ is a triangular array of time-inhomogeneous p-order Markov processes associated to the transition kernels R_u, $u \in [0,1]$.

Proposition 3. *Under the assumptions* E21–E24, *the conclusions of Corollary 3 hold true with* $q = q', d_1 = d_0 - q'k$ *and* $\phi(y) = 1 + |y|, y \in E$.

Example. Consider the case for $p = 1$ with $m(u, x) = \sum_{i=1}^I (a_i(u)x + b_i(u)) \mathbb{1}_{x \in R_i}, \{R_1, \dots, R_I\}$ a partition of $\mathbb{R}$ and a_i, b_i are functions k-times continuously differentiable with $\max_{1 \le i \le I} \max_{u \in [0,1]} |a_i(u)| < 1$. This corresponds to a threshold model with non time-varying regions for the different regimes. If E21 holds true, E22 follows with $q' = 1$. If Assumption E23 is also valid for some $q_0 > 1$, Proposition 3 applies. This example is a generalization of the SETAR model discussed in Truquet [35] (see Example 3 in Section 4.4).

Notes.

1. Let us compare our result with that of Dahlhaus et al. [4] who studied nonlinear autoregressive processes. For simplicity, we restrict the study to $p = 1$. Suppose that for some $d_0 \ge 1$, we have $\mathbb{E}|\varepsilon_1|^{d_0} < \infty$ and there exist $c > 0$ and $\beta \in (0,1)$ such that

$$\sup_{u \in [0,1]} \left| m(u, x) - m(u, x') \right| \le \beta |x - x'|,$$

$$\max_{i=1,2} \sup_{u \in [0,1]} \left| \partial_i m(u, x) - \partial_i m(u, x') \right| \le C |x - x'|.$$

Theorem 4.8 and Proposition 3.8 in Dahlhaus et al. [4] show that the function $u \mapsto \int g \, d\pi_{u,j}$ is continuously differentiable whenever the function $g : E^j \to \mathbb{R}$ is continuously differentiable and satisfies for some $C > 0$,

$$\left| g(z) - g(z') \right| \le C \left(1 + |z|^{d_0 - 1} + |z'|^{d_0 - 1} \right) |z - z'|.$$

These authors also prove that there exists some positive constants C_1 and C_2 such that for $t \in \mathbb{Z}$,

$$\mathbb{E}^{1/d_0}\big|X_{n,t} - X_t(u)\big|^{d_0} \leq C_1\big[|u - t/n| + 1/n\big] \quad \text{with } X_t(u) = m\big(u, X_t(u)\big) + \sigma(u)\varepsilon_t$$

and $|\int g \, d\pi_{t,j}^{(n)} - \int g \, d\pi_{u,j}| \leq C_2[|u - t/n| + 1/n]$.

In contrast, when $d_0 > 1$, $\int |y|^{d_0}[f_\varepsilon(y) + |f'_\varepsilon(y)|]\,dy < \infty$ and there exist $\beta \in (0,1)$, $\beta', C > 0$ such that

$$\sup_{u \in [0,1]} \big|m(u,x)\big| \leq \beta|x| + \beta', \qquad \sup_{u \in [0,1]} \big|\partial_1 m(u,x)\big| \leq C\big(1 + |x|\big),$$

Proposition 3 guarantees that $u \mapsto \int g\,d\pi_{u,j}$ is continuously differentiable, provided that $|g(z)| \leq C(1 + |z|)$ for some constant $C > 0$.

Contrarily to [4], differentiability and even continuity of the application $x \mapsto m(u,x)$ is not required for applying our results. Moreover, one can consider very irregular functions g (see, e.g., the indicator of any Borel set). On the other hand, we impose much more regularity assumptions on the noise distribution (existence of a smooth density and a moment condition for its derivative). Our approach is then more interesting for non smooth regression functions.

To illustrate the benefit of our results for a model discussed in Dahlhaus et al. [4], we consider the threshold model with $m(u,x) = a_1(u)\max(x,0) + a_2(u)\max(-x,0)$ with a_1, a_2 continuously differentiable and $\max_{u \in [0,1]} |a_i(u)| < 1$, $i = 1, 2$. In this case, Lemma 4.5 and Theorem 4.8 in Dahlhaus et al. [4] guaranty Lipschitz continuity of $u \mapsto \int g\,d\pi_{u,j}$ but not its differentiability because the application $x \mapsto m(u,x)$, which is required to be continuously differentiable, is not differentiable at point $x = 0$. In contrast, we can prove differentiability $u \mapsto \int g\,d\pi_{u,j}$ for a different class of functions g. More general threshold models with a discontinuous regression function can also be considered with our approach, some of them are given after the statement of Proposition 3.

2. Exponential stability can be used for such models provided that $f_\varepsilon^{(s)}$ has some exponential moments for $s = 0, \ldots, k$. In this case, one can take $V_s(y) = \exp(\kappa|y|)$ for all s. A precise result is given in Proposition 7 given in the supplementary material. In this case, the approach of Heidergott and Hordijk [19] can also be used for studying existence of derivatives and our general result, which also covers this case, is not useful (except that we provide a criterion for p-order Markov chain, which is new). These exponential moments induce a serious restriction on the noise distribution because fatter tails distributions such as Student distributions are excluded. However, the local stationarity property of this model, resulting from Proposition 7, is a new result.

3. Our result can be also applied to the AR(1) process $X_t = \alpha X_{t-1} + \varepsilon_t$ for getting derivatives of the applications $\alpha \mapsto \pi_\alpha$, as in Ferré et al. [14]. The index u is replaced with α and the interval $[0,1]$ with $I = [-1+\epsilon, 1-\epsilon]$ for some $\epsilon \in (0,1)$. Let $Q_\alpha(x, dy) = f_\varepsilon(y - \alpha x)\,dy$. In this case, one can take $q' = 1$, $\phi(x) = 1 + |x|$ and if $k < d_0 < k + 1$, $d_1 = d_0 - k$. Under some assumptions that guaranty E23, Ferré et al. [14] showed in their Proposition 1 that $\alpha \mapsto \pi_\alpha$, considered as an application from I to $\phi^\beta(\mathbb{R})$, is k-times continuously

differentiable, provided that $0 < \beta < d_1$. See their condition on β given after the statement of their Lemma 1. One can then see that our result is stronger. We claim that the slight difference between the two results is explained by the additional topologies used in their Lemma 1 for studying continuity of the application

$$\alpha \mapsto Q_\alpha^{(\ell)}(x, dy) = (-1)^\ell x^\ell f_\varepsilon^{(\ell)}(y - \alpha x) \, dy.$$

Let us enlighten why by supposing that $k = 1$. From Theorem 1, we have, using our notations $T_\alpha^{(\ell)} \mu = \mu Q_\alpha^{(\ell)}$,

$$\pi_\alpha^{(1)} = (I - T_\alpha)^{-1} T_\alpha^{(1)} \pi_\alpha.$$

Denoting by $\mathcal{L}(\phi^{d_1}, \phi^{d_1'})$ the set of bounded linear operators from $\mathcal{M}_{\phi^{d_1}}(\mathbb{R})$ to $\mathcal{M}_{\phi^{d_1'}}(\mathbb{R})$, the application $\alpha \mapsto T_\alpha^{(1)}$, as an application from I to $\mathcal{L}(\phi^{d_1}, \phi^{d_1'})$ is only continuous when $d_1' < d_1$. This shows that one can only get continuity $\alpha \mapsto \pi_\alpha^{(1)}$ for $\| \cdot \|_{\phi^{d_1'}}$ if we use operator norms. On the other hand, if $\mu \in \mathcal{M}_{\phi^{d_1}}(\mathbb{R})$, one can show that the application $\alpha \mapsto T_\alpha^{(1)} \mu$, as an application from I to $\mathcal{M}_{\phi^{d_1}}(\mathbb{R})$ is continuous. As shown in Theorem 1, this weaker continuity condition is sufficient for getting continuity of $\alpha \mapsto \pi_\alpha^{(1)}$, as an application from I to $\mathcal{M}_{\phi^{d_1}}(\mathbb{R})$.

6.2. Integer-valued time series

For $u \in [0, 1]$ and $1 \leq i \leq p$, let $\zeta_{i,u}$ and ξ_u be some probability distributions supported on the nonnegative integers and for $\mathbf{x} \in \mathbb{Z}_+^p$, $R_u(\mathbf{x}, \cdot)$ will denote the probability distribution given by the convolution product $\zeta_{1,u}^{*x_1} * \zeta_{2,u}^{*x_2} * \cdots * \zeta_{p,u}^{*x_p} * \xi_u$ with $\zeta_{i,u}^{*x} = \zeta_{i,u}^{*(x-1)} * \zeta_{i,u}$ if $x \geq 1$, $\zeta_{i,u}^{*1} = \zeta_{i,u}$ and the convention $\zeta_{i,u}^{*0} = \delta_0$.

Let us comment this Markov structure. When $p = 1$, R_u is the transition matrix of a Galton-Watson process with immigration. Such Markov processes are also used in time series analysis of discrete data. For instance, if $\zeta_{i,u}$ denotes the Bernoulli distribution of parameter $\alpha_{i,u}$, such Markov processes are called INAR processes and were studied by McKenzie [25], Al-Osh and Alzaid [1] and Jin-Guan and Yuan [10] among others. Note that in this case, we have the autoregressive representation $X_k = \sum_{i=1}^p \alpha_{i,u} \circ X_{k-i} + \varepsilon_k$, where $\alpha \circ x$ denotes a random variable following a binomial distribution of parameters (x, α) and independent from ε_k, an integer-valued random variable with probability distribution q_u. When $\zeta_{i,u}$ denotes the Poisson distribution of parameter $\alpha_{i,u}$ and ξ_u denotes the Poisson distribution of parameter $\alpha_{0,u}$, then $R_u(\mathbf{x}, \cdot)$ is the Poisson distribution of parameter $\alpha_{0,u} + \sum_{i=1}^p \alpha_{i,u}$ and the Markov process coincides with the INARCH process studied in Ferland et al. [13]. The distributions $\zeta_{i,u}$ and ξ_u can also have a general form as in the generalized INAR processes studied by Latour [24] and are not required to have exponential moments. For instance the log-logistic distribution ζ with parameters $\alpha, \beta > 0$ and defined by $\zeta(x) = (\beta/\alpha)(x/\alpha)^\beta (1 + (x/\alpha)^\beta)^{-2}$ for $x \in \mathbb{Z}_+$, has only a finite moment of order $k < \beta$. When $p = 1$, conditions ensuring local stationarity for the INARCH and INAR processes are discussed in Truquet [35]. Here, we propose an extension to the case $p \geq 1$, with general probability distributions $\zeta_{i,u}$ and ξ_u. We will use the following assumptions.

E31 We have $\alpha := \sup_{u \in [0,1]} \sum_{i=1}^p \sum_{x \geq 0} x \zeta_{i,u}(x) < 1$ and there exists an integer x_0 such that $\beta := \inf_{u \in [0,1]} \xi_u(x_0) > 0$.

E32 For each integer $x \geq 0$, the applications $u \mapsto \zeta_{i,u}(x)$ and $u \mapsto \xi_u(x)$ are of class $\mathcal{C}^k$. Moreover, there exists a positive integer d_1 such that for $s = 0, 1, \ldots, k$,

$$\lim_{M \to \infty} \sup_{u \in [0,1]} \sum_{i=1}^p \sum_{x \geq M} x^{d_1 + k - s} \big[\big| \zeta_{i,u}^{(s)}(x) \big| + \big| \xi_u^{(s)}(x) \big| \big] = 0.$$

Proposition 4. *Assume that the assumptions E31–E32 hold true and set $\phi(x) = 1 + x$ for $x \in \mathbb{N}$ and $d_0 = d_1 + k$. Then the conclusions of Corollary 3 hold true.*

Note. Assumption E32 is satisfied for Bernoulli, Poisson or negative binomial distributions provided the real-valued parameter of these distributions is a k-times continuously differentiable function taking values in the usual intervals $(0, 1)$ (for the Bernoulli or negative binomial distribution) or $(0, \infty)$ (for the Poisson distribution).

6.3. Markov chain in a Markovian random environment

We consider a state space $E = E_1 \times E_2$ with E_1 a finite set and E_2 an arbitrary metric space. Let $\{P(u, \cdot, \cdot; z) : u \in [0,1], z \in E_2\}$ is a family of stochastic matrices on E_1 and $\{\overline{Q}_u : u \in [0,1]\}$ a family of Markov kernels on E_2. We assume that for all $u \in [0,1]$, $\overline{Q}_u(x_2, dy_2) = \overline{f}(u, x_2, y_2) \overline{\gamma}(x_2, dy_2)$ for a measurable function $\overline{f} : [0,1] \times E_2^2 \to \mathbb{R}_+$ and a measure kernel $\overline{\gamma}$ on E_2. We consider the family of Markov kernels $\{Q_u : u \in [0,1]\}$ such that

$$Q_u\big((y_1, z_1), (dy_2, dz_2)\big) = P(u, y_1, y_2; z_2) \overline{Q}_u(z_1, dz_2), \quad u \in [0,1]. \tag{11}$$

Setting $f(u, (y_1, z_1), (y_2, z_2)) = P(u, y_1, y_2; z_2) \overline{f}(u, z_1, z_2)$ we have

$$Q_u\big((y_1, z_1), (dy_2, dz_2)\big) = f\big(u, (y_1, z_1), (y_2, z_2)\big) c(dy_2) \overline{\gamma}(z_1, dz_2),$$

where c denotes the counting measure on E_1. We then set $\gamma((y_1, z_1), (dy_2, dz_2)) = c(dy_2) \overline{\gamma}(z_1, dz_2)$.

For $u \in [0,1]$, $P(u, \cdot, \cdot; z)$ is the transition matrix of a process in a Markovian random environment. The kernels Q_u can also be seen as a transition operator for a categorical time series with exogenous covariates. Indeed, if $\{X_{n,t} = (Y_{n,t}, Z_{n,t}) : 1 \leq t \leq n, n \geq 1\}$ is a triangular array associated to the family $\{Q_u : u \in [0,1]\}$, we have

$$\mathbb{P}\big(Y_{n,t} = y' | Y_{n,t-1}, Z_{n,1}, \ldots, Z_{n,n}\big) = P\big(t/n, y, y'; Z_{n,t}\big), \quad 1 \leq t \leq n, n \geq 1.$$

In the time-homogeneous case, Fokianos and Truquet [15] recently studied Markov chains models with exogenous covariates of a general form and discussed their link with Markov chains in a random environment for studying ergodicity properties. We provide here a locally stationary analogue but with a restriction on the covariate process which is given by a locally stationary

Markov chain. An important important example of such models is the autoregressive logistic model with $E_1 = \{0, 1\}$, $E_2 = \mathbb{R}^g$ and

$$P(u, y, 1, z) = \frac{\exp(a_0(u) + a_1(u)y + z'\beta(u))}{1 + \exp(a_0(u) + a_1(u)y + z'\beta(u))}$$

for some continuous functions $a_0, a_1 : [0, 1] \to \mathbb{R}$ and $\beta : [0, 1] \to \mathbb{R}^g$.

We will use the following set of assumptions.

E41 For all $(y_1, y_2, z_2) \in E_1^2 \times E_2$, the functions $u \mapsto P_u(y_1, y_2; z_2)$ is k-times continuously differentiable and positive.

E42 The family of Markov kernels $\{\overline{Q}_u : u \in [0, 1]\}$ satisfies Assumptions SC1–SC3. We denote the different constants involved in the assumptions by an overline.

E43 For $\ell = 0, \ldots, k$, we have

$$\sup_{u \in [0, 1]} \max_{y, y' \in E_1} \left| \partial_1^{(\ell)} P(u, y, y'; z) \right| \leq C\overline{\phi}(z)^{\overline{q}\ell}.$$

Proposition 5. *Under the assumptions* E41–E43, *the conclusions of Proposition* 1 *are valid for* $(d_0, d_1, q) = (\overline{d}_0, \overline{d}_1, \overline{q})$ *and* $\phi(y, z) = \overline{\phi}(z)$.

Note. Discrete-valued time series have been extensively studied. See for instance Weiß [39] for a recent overview of the literature on this topic. Let us mention two additional classes of time series models. One important class concerns the models of Durbin and Koopman [11] which are defined by a conditional distribution depending on a latent process parameter. See equations (3, 4) in Durbin and Koopman [11] for a precise definition. Let us mention that for categorical time series, (11) contains a time-varying version of such models as a special case. To this end, it is necessary to assume that $P(u, y_1, y_2; z_2)$ does not depend on y_1 and $\overline{Q}_u$ is the transition of an AR(1) process, $Z_t(u) = a(u)Z_{t-1}(u) + \sigma(u)\varepsilon_t$. To check E42, the noise ε_t is required to have a sufficiently smooth density.

The second class concern observation-driven models for time series of counts as in Davis et al. [6] or Davis and Liu [7]. In this class of model, the conditional distribution of the process $(Y_t)_{t \in \mathbb{Z}}$ at time t depends on a random parameter, say λ_t, which is a function of the lag values $\lambda_{t-j}, Y_{t-j}, 1 \leq j \leq p$. While the bivariate process $((\lambda_t, Y_t))_{t \in \mathbb{Z}}$ has Markovian properties, the usual irreducibility condition does not hold in general and the small set assumption used in the present paper is not valid. Different techniques are then employed for studying such models, for instance the contraction in average method used by Davis and Liu [7] or other methods for non irreducible Markov chains, as in Douc et al. [8]. It is then not possible to use our results for studying time-varying versions of such processes.

Acknowledgments

Supplementary Material

Supplement to "A perturbation analysis of Markov chains models with time-varying parameters" (DOI: 10.3150/20-BEJ1210SUPP; .pdf). We provide additional proofs that are not given in the paper.

References

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal*. **8** 261–275. MR0903755 https://doi.org/10.1111/j.1467-9892.1987.tb00438.x

[2] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist*. **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[3] Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics* **30** 351–413. Elsevier.

[4] Dahlhaus, R., Richter, S. and Wu, W.B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

[5] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist*. **34** 1075–1114. MR2278352 https://doi.org/10.1214/009053606000000227

[6] Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B. (2003). Observation-driven models for Poisson counts. *Biometrika* **90** 777–790. MR2024757 https://doi.org/10.1093/biomet/90.4.777

[7] Davis, R.A. and Liu, H. (2016). Theory and inference for a class of nonlinear models with application to time series of counts. *Statist. Sinica* **26** 1673–1707. MR3586234

[8] Douc, R., Doukhan, P. and Moulines, E. (2013). Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator. *Stochastic Process. Appl*. **123** 2620–2647. MR3054539 https://doi.org/10.1016/j.spa.2013.04.010

[9] Douc, R., Moulines, E. and Stoffer, D.S. (2014). *Nonlinear Time Series*. *Chapman & Hall/CRC Texts in Statistical Science Series*. Boca Raton, FL: CRC Press/CRC. Theory, methods, and applications with R examples. MR3289095

[10] Du, J.G. and Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. *J. Time Ser. Anal*. **12** 129–142. MR1108796 https://doi.org/10.1111/j.1467-9892.1991.tb00073.x

[11] Durbin, J. and Koopman, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **62** 3–56. MR1745604 https://doi.org/10.1111/1467-9868.00218

[12] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. London: CRC Press. MR1383587

[13] Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal*. **27** 923–942. MR2328548 https://doi.org/10.1111/j.1467-9892.2006.00496.x

[14] Ferré, D., Hervé, L. and Ledoux, J. (2013). Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab*. **50** 184–194. MR3076780 https://doi.org/10.1239/jap/1363784432

[15] Fokianos, K. and Truquet, L. (2019). On categorical time series models with covariates. *Stochastic Process. Appl*. **129** 3446–3462. MR3985569 https://doi.org/10.1016/j.spa.2018.09.012

[16] Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist*. **36** 742–786. MR2396814 https://doi.org/10.1214/07-AOS510

[17] Glynn, P.W. and L'Ecuyer, P. (1995). Likelihood ratio gradient estimation for stochastic recursions. *Adv. in Appl. Probab*. **27** 1019–1053. MR1358905 https://doi.org/10.2307/1427933

[18] Hairer, M. and Mattingly, J.C. (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis*, *Random Fields and Applications VI*. *Progress in Probability* **63** 109–117. Basel: Birkhäuser/Springer Basel AG. MR2857021 https://doi.org/10.1007/978-3-0348-0021-1_7

[19] Heidergott, B. and Hordijk, A. (2003). Taylor series expansions for stationary Markov chains. *Adv. in Appl*. *Probab*. **35** 1046–1070. MR2014269 https://doi.org/10.1239/aap/1067436334

[20] Heidergott, B., Hordijk, A. and Weisshaupt, H. (2006). Measure-valued differentiation for stationary Markov chains. *Math*. *Oper*. *Res*. **31** 154–172. MR2205525 https://doi.org/10.1287/moor.1050.0171

[21] Hervé, L. and Pène, F. (2010). The Nagaev-Guivarc'h method via the Keller-Liverani theorem. *Bull*. *Soc*. *Math*. *France* **138** 415–489. MR2729019 https://doi.org/10.24033/bsmf.2594

[22] Kartashov, N.V. (1985). Inequalities in stability and ergodicity theorems for Markov chains with a general phase space. I. *Theory Probab. Appl*. **30** 247–259.

[23] Kushner, H.J. and Yin, G.G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **35**. New York: Springer. Stochastic Modelling and Applied Probability. MR1993642

[24] Latour, A. (1997). The multivariate GINAR(p) process. *Adv. in Appl*. *Probab*. **29** 228–248. MR1432938 https://doi.org/10.2307/1427868

[25] McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Adv. in Appl. Probab.* **18** 679–705. MR0857325 https://doi.org/10.2307/1427183

[26] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. With a prologue by Peter W. Glynn. MR2509253 https://doi.org/10.1017/CBO9780511626630

[27] Moulines, E., Priouret, P. and Roueff, F. (2005). On recursive estimation for time varying autoregressive processes. *Ann*. *Statist*. **33** 2610–2654. MR2253097 https://doi.org/10.1214/009053605000000624

[28] Pflug, G.Ch. (1992). Gradient estimates for the performance of Markov chains and discrete event processes. *Ann*. *Oper*. *Res*. **39** 173–194. MR1204975 https://doi.org/10.1007/BF02060941

[29] Pflug, G.Ch. (1996). *Optimization of Stochastic Models*: *The interface between simulation and optimization*. *The Kluwer International Series in Engineering and Computer Science* **373**. Boston, MA: Kluwer Academic. MR1492446 https://doi.org/10.1007/978-1-4613-1449-3

[30] Richter, S. and Dahlhaus, R. (2019). Cross validation for locally stationary processes. *Ann*. *Statist*. **47** 2145–2173. MR3953447 https://doi.org/10.1214/18-AOS1743

[31] Rudolf, D. and Schweizer, N. (2018). Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli* **24** 2610–2639. MR3779696 https://doi.org/10.3150/17-BEJ938

[32] Schweitzer, P.J. (1968). Perturbation theory and finite Markov chains. *J. Appl. Probability* **5** 401–413. MR0234527 https://doi.org/10.2307/3212261

[33] Subba Rao, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. MR2285698 https://doi.org/10.1239/aap/1165414596

[34] Truquet, L. (2017). Parameter stability and semiparametric inference in time varying auto-regressive conditional heteroscedasticity models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1391–1414. MR3731668 https://doi.org/10.1111/rssb.12221

[35] Truquet, L. (2019). Local stationarity and time-inhomogeneous Markov chains. *Ann*. *Statist*. **47** 2023–2050. MR3953443 https://doi.org/10.1214/18-AOS1739

[36] Truquet, L. (2020). Supplement to "A perturbation analysis of Markov chains models with time-varying parameters." https://doi.org/10.3150/20-BEJ1210SUPP

[37] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. New York: Springer. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[38] Vázquez-Abad, F.J. and Kushner, H.J. (1992). Estimation of the derivative of a stationary measure with respect to a control parameter. *J. Appl. Probab.* **29** 343–352. MR1165219 https://doi.org/10.2307/3214571

[39] Weiß, C.H. (2018). *An Introduction to Discrete-valued Time Series*. New York: Wiley.

[40] Zhang, T. and Wu, W.B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029 https://doi.org/10.1214/12-AOS1010