State-discretization of V -geometrically ergodic Markov chains and convergence to the stationary distribution

Loïc Hervé, James Ledoux

State-discretization of V-geometrically ergodic Markov chains and convergence to the stationary distribution

Loic HERVÉ and James LEDOUX *

September 26, 2019

Abstract

Let $(X_n)_{n\in\mathbb{N}}$ be a V-geometrically ergodic Markov chain on a measurable space $\mathbb{X}$ with invariant probability distribution π. In this paper, we propose a discretization scheme providing a computable sequence $(\widehat{\pi}_k)_{k\geq 1}$ of probability measures which approximates π as k growths to infinity. The probability measure $\widehat{\pi}_k$ is computed from the invariant probability distribution of a finite Markov chain. The convergence rate in total variation of $(\widehat{\pi}_k)_{k\geq 1}$ to π is given. As a result, the specific case of first order autoregressive processes with linear and non-linear errors is studied. Finally, illustrations of the procedure for such autoregressive processes are provided, in particular when no explicit formula for π is known.

AMS subject classification : 60J05; 60J22

Keywords : Markov chain, Rate of convergence, Autoregressive models.

1 Introduction

Let $(\mathbb{X}, d)$ denote a metric space equipped with its Borel σ-algebra $\mathcal{X}$. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with state space $(\mathbb{X}, \mathcal{X})$ and transition kernel P of the form

$$\forall x \in \mathbb{X}, \quad P(x, dy) = p(x, y)\, d\mu(y), \tag{1}$$

where $p : \mathbb{X}^2 \to [0, +\infty)$ is a measurable function and μ is a positive σ-additive measure on $(\mathbb{X}, \mathcal{X})$. Typically $\mathbb{X}$ is $\mathbb{R}^d$ and μ is the Lebesgue measure on $\mathbb{R}^d$. Moreover let $v : [0, +\infty) \to [1, +\infty)$ denote an unbounded increasing continuous function such that $v(0) = 1$, and let $V : \mathbb{X} \to [1, +\infty)$ be defined by

$$\forall x \in \mathbb{X}, \quad V(x) := v\big(d(x, x_0)\big), \tag{2}$$

where $x_0 \in \mathbb{X}$ is fixed. We assume that P admits an invariant probability measure π on $(\mathbb{X}, \mathcal{X})$. Since P is of the form (1), π is absolutely continuous with respect to μ, that is

*Univ Rennes, INSA Rennes, CNRS, IRMAR-UMR 6625, F-35000 Rennes, France. Loic.Herve@insa-rennes.fr, James.Ledoux@insa-rennes.fr

$d\pi(y) = \mathfrak{p}(y)\, d\mu(y)$ for some probability density function (pdf) $\mathfrak{p}$. Throughout the paper we assume that

$$\pi(V) := \int_{\mathbb{X}} V(y)\, \mathfrak{p}(y)\, d\mu(y) < \infty$$

and that P is V-geometrically ergodic, that is (e.g. see [MT93]): there exist $\rho \in (0,1)$ and a positive constant $C \equiv C(\rho)$ such that the following inequality holds for every measurable complex-valued function f on $\mathbb{X}$ satisfying $|f| \leq V$:

$$\forall n \geq 0, \quad \sup_{x \in \mathbb{X}} \frac{|(P^n f)(x) - \pi(f)|}{V(x)} \leq C\, \rho^n. \tag{3}$$

Mention that, for most of the classical V-geometrically ergodic Markov chains, the function V is of the form (2).

Even for simple models as first-order autoregressive models, the explicit computation of the stationary pdf $\mathfrak{p}$ is a difficult issue, and it is only possible for some specific examples. In this work, under suitable assumptions on the kernel $p(x,y)$, we propose a discretization procedure providing a computable sequence $(\widehat{\pi}_k)_{k\geq 1}$ of probability measures on $\mathbb{X}$ which approximates the stationary distribution π of P in total variation distance. Roughly speaking the probability measure $\widehat{\pi}_k$ on $\mathbb{X}$ is defined as follows. For every integer k, an explicit finite stochastic matrix B_k is derived from the Markov kernel P by discretization of the kernel $p(x,y)$. Then $\widehat{\pi}_k$ is defined as a natural extension of the left B_k-invariant probability vector. Then the above mentioned convergence of $(\widehat{\pi}_k)_{k\geq 1}$ to π in total variation distance is derived in Theorem 3.1 from the results of [HL14]. Moreover the absolutely continuous part $\mathfrak{p}_k$ of $\widehat{\pi}_k$ w.r.t. μ can be explicitly computed, and the sequence $(\mathfrak{p}_k)_{k\geq 1}$ is proved to converge to $\mathfrak{p}$ in the usual Lebesgue space $\mathbb{L}^1(\mathbb{X}, \mathcal{X}, \mu)$ (see Corollary 3.2). Applications to the first order (linear) autoregressive models AR(1) and to AR(1) processes with ARCH(1) errors are addressed in Sections 4. The computational issues to get $\mathfrak{p}_k$ are discussed in Section 5. Numerical illustrations are presented in Section 6.

The authors in [Hai98, AH00, ANR07] developed another method to approximate the stationary pdf $\mathfrak{p}$ of linear processes. Their approach consists in approximating the stationary pdf $\mathfrak{p}$ of an AR(1) process (i.e. $X_n = \varrho\, X_{n-1} + \vartheta_n$, see Subsection 4.1 for details) by the sequence $(h_n)_{n\in\mathbb{N}}$ of functions recursively defined by

$$h_0 := \nu \qquad \text{and} \qquad \forall n \geq 1, \quad h_n(x) := \int_{\mathbb{R}} \nu(x - \varrho u)\, h_{n-1}(u)\, du \tag{4}$$

where ν denotes the innovation pdf (i.e. the law of ϑ_1). Hainman in [Hai98] proved that $(h_n)_{n\in\mathbb{N}}$ uniformly converges to $\mathfrak{p}$ with geometric rate under strong assumptions on the support of ν. The authors in [AH00] proved that $(h_n)_{n\in\mathbb{N}}$ converges point-wise to $\mathfrak{p}$ under some mild assumptions on the Fourier transform of ν, and they established the uniform convergence with geometric rate in the case when ν is the exponential pdf. In [ANR07] the uniform convergence of $(h_n)_{n\in\mathbb{N}}$ to $\mathfrak{p}$, with geometric rate, is extended to general causal linear processes under mild assumptions on the noise process. Closely linked to these works, we also mention the paper [Log04] which studies the characteristic function of the stationary pdf $\mathfrak{p}$ for a threshold AR(1) model with noise process having Laplace distribution, as well as the paper [AR05] which investigates $\mathfrak{p}$ for absolute autoregressive associated with noise process having Gaussian, Cauchy or Laplace distribution (from [CT86] this issue may be reduced to the computation of the stationary pdf of an auxiliary AR(1) process).

Due to [ANR07], the approximation of $\mathfrak{p}$ by $(h_n)_{n\in\mathbb{N}}$ via Equation (4) is theoretically efficient for linear processes since the rate of convergence is geometric. However, except when the noise process has a special usual law, the exact calculation of the integral in (4) can not be carried out. Moreover, any numerical method recursively providing approximations of the integrals $h_1, \dots, h_p$ for some $p \geq 1$ induces some cumulative errors. For linear processes our method is thus an alternative way to approximate $\mathfrak{p}$: the rate of convergence in our work is not geometric (a priori), but for some $k \geq 1$ the approximation $\mathfrak{p}_k$ of $\mathfrak{p}$ as above described can be directly computed (without any recursive procedure). Section 6 provides numerical evidence for robustness of the method. Moreover our approach applies to any V-geometrical Markov chain (not only to linear processes) admitting a probability kernel $P(x, dy)$ of the form (1), provided that the kernel $p(\cdot, \cdot)$ has some suitable Lipschitz-regularity properties (see Assumption (18c)). For instance our method applies to autoregressive process with ARCH(1) errors (see Subsections 4.2 and 6.2.3).

The invariant pdf $\mathfrak{p}$ satisfies the functional equation $T\mathfrak{p} = \mathfrak{p}$, where T is the linear operator defined by $(Tf)(\cdot) = \int_{\mathbb{X}} p(y, \cdot)\, f(y)\, d\mu(y)$. However this operator T is not used in this work. Indeed that is not T, but P, which is approximated by a sequence of finite-rank operators $(\widehat{P}_k)_{k\geq 1}$. The reason for this is that P has good spectral properties on the usual weighted-supremum Banach space $\mathcal{B}_1$ associated with V due to the V-geometrical ergodicity assumption. Also note that the classical theory of perturbed operators does not apply here because the sequence $(\widehat{P}_k)_{k\geq 1}$ does not converge to P for the usual operator norm on $\mathcal{B}_1$ (in particular P is not a compact operator on $\mathcal{B}_1$). To get around this difficulty, we use the results of [HL14] based on the Keller-Liverani perturbation theorem [KL99]: this method requires an auxiliary weaker operator norm on $\mathcal{B}_1$ (see Lemma 3.4), as well as uniform (in k) drift inequalities for $\widehat{P}_k$ (see Lemma 3.3). In the context of perturbed V-geometrically ergodic Markov chains, the interest of using an auxiliary norm appears in [SS00] (see [Kel82] for similar issues in ergodic theory). For recent works related to this weak perturbation method in Markovian models, see [FHL13, RS18, Tru17] and the references therein.

2 Definition of the approximating probability measure $\widehat{\pi}_k$

Let $x_0 \in \mathbb{X}$ be fixed and, for every integer $k \geq 1$, let us consider any $\mathbb{X}_k \in \mathcal{X}$ such that

$$\big\{x \in \mathbb{X} \,:\, d(x, x_0) < k\big\} \;\subseteq\; \mathbb{X}_k \;\subseteq\; \big\{x \in \mathbb{X} \,:\, d(x, x_0) \leq k\big\}.$$

Let us introduce the following finite partitions of the sequence of spaces $(\mathbb{X}_k)_{k\geq 1}$.

Definition (A). *Let $(\delta_k)_{k\geq 1}$ be a sequence of positive real numbers such that $\lim_k \delta_k = 0$. For every integer $k \geq 1$, we consider a finite family $\{\mathbb{X}_{j,k}\}_{j\in I_k}$ of disjoint measurable subsets of $\mathbb{X}_k$ such that*

$$\mathbb{X}_k = \bigsqcup_{j\in I_k} \mathbb{X}_{j,k} \qquad \textit{with} \;\; \forall j \in I_k, \quad \textit{diam}(\mathbb{X}_{j,k}) \leq \delta_k. \tag{5}$$

where $\textit{diam}(\mathbb{X}_{j,k}) := \sup\big\{d(x, x') \,:\, (x, x') \in \mathbb{X}_{j,k}\big\}$. The positive real number δ_k must be thought of as the mesh of the partition $\{\mathbb{X}_{j,k}\}_{j\in I_k}$.

Define

$$\forall k \geq 1,\ \forall (x,y) \in \mathbb{X}^2, \quad p_k(x,y) := 1_{\mathbb{X}_k}(y) \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x) \inf_{t \in \mathbb{X}_{i,k}} p(t,y),$$

Observe that $p_k \leq p$. Below $f : \mathbb{X} \to \mathbb{C}$ denotes any bounded measurable function on $\mathbb{X}$ where $\mathbb{C}$ denoted the set of complex numbers. We define the following non-negative kernel $\widehat{Q}_k$:

$$\begin{aligned} \forall x \in \mathbb{X}, \quad (\widehat{Q}_k f)(x) &:= \int_{\mathbb{X}} f(y)\, p_k(x,y)\, d\mu(y) \\ &= \sum_{i \in I_k} \Big(\int_{\mathbb{X}_k} f(y) \inf_{t \in \mathbb{X}_{i,k}} p(t,y)\, d\mu(y) \Big) 1_{\mathbb{X}_{i,k}}(x). \end{aligned} \tag{6}$$

Note that $\widehat{Q}_k f$ vanishes on $\mathbb{X} \setminus \mathbb{X}_k$. Let ψ_k be the non-negative function on $\mathbb{X}$ defined by

$$\psi_k := 1_{\mathbb{X}} - \widehat{Q}_k 1_{\mathbb{X}}.$$

We have $\psi_k \equiv 1$ on $\mathbb{X} \setminus \mathbb{X}_k$, and $0 \leq \psi_k \leq 1_{\mathbb{X}}$ since $0 \leq \widehat{Q}_k 1_{\mathbb{X}} \leq P 1_{\mathbb{X}} = 1_{\mathbb{X}}$. Next define the following kernel:

$$\forall x \in \mathbb{X}, \quad (\widehat{P}_k f)(x) := (\widehat{Q}_k f)(x) + f(x_0)\, \psi_k(x). \tag{7}$$

Then $\widehat{P}_k$ is a Markov kernel on $(\mathbb{X}, \mathcal{X})$, i.e. $\widehat{P}_k$ is non-negative ($f \geq 0 \Rightarrow \widehat{P}_k f \geq 0$) and $\widehat{P}_k 1_{\mathbb{X}} = 1_{\mathbb{X}}$.

Moreover we deduce from (7) and (6) that $\widehat{P}_k(f) \in \mathcal{F}_k$, where $\mathcal{F}_k$ is the finite-dimensional space spanned by the system of functions $\big\{1_{\mathbb{X}_{i,k}},\ i \in I_k\big\} \cup \{\psi_k\}$. Observe that $1_{\mathbb{X}} \in \mathcal{F}_k$ from $1_{\mathbb{X}} = \widehat{Q}_k 1_{\mathbb{X}} + \psi_k$ and (6). Now define

$$b_k := 1_{\mathbb{X}} - 1_{\mathbb{X}_k} = 1_{\mathbb{X} \setminus \mathbb{X}_k}.$$

Then $b_k \in \mathcal{F}_k$ since $1_{\mathbb{X}} \in \mathcal{F}_k$ and $b_k = 1_{\mathbb{X}} - \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}$. Thus another basis of $\mathcal{F}_k$ is given by

$$\mathcal{C}_k := \big\{1_{\mathbb{X}_{i,k}},\ i \in I_k\big\} \cup \{b_k\}. \tag{8}$$

Let $\{x_{i,k}\}_{i \in I_k}$ be such that $x_{i,k} \in \mathbb{X}_{i,k}$ and let $\overline{x}_k \in \mathbb{X} \setminus \mathbb{X}_k$. Then we have for every $g \in \mathcal{F}_k$:

$$g = \sum_{i \in I_k} g(x_{i,k})\, 1_{\mathbb{X}_{i,k}} + g(\overline{x}_k)\, b_k. \tag{9}$$

Now, from $\widehat{P}_k(\mathcal{F}_k) \subset \mathcal{F}_k$ we can define the linear map $P_k : \mathcal{F}_k \to \mathcal{F}_k$ as the restriction of $\widehat{P}_k$ to $\mathcal{F}_k$. Let $N_k := \dim \mathcal{F}_k = \mathrm{Card}\,(I_k) + 1$, and let B_k be the $N_k \times N_k-$matrix defined as the matrix of P_k with respect to the basis $\mathcal{C}_k$. Note that

$$P_k b_k = \widehat{P}_k b_k = \widehat{Q}_k b_k + b_k(x_0)\psi_k = 0, \tag{10}$$

and that for every $j \in I_k$

$$\begin{aligned} P_k 1_{\mathbb{X}_{j,k}} &= \widehat{P}_k 1_{\mathbb{X}_{j,k}} \\ &= \sum_{i \in I_k} (\widehat{P}_k 1_{\mathbb{X}_{j,k}})(x_{i,k})\, 1_{\mathbb{X}_{i,k}} + (\widehat{P}_k 1_{\mathbb{X}_{j,k}})(\overline{x}_k)\, b_k \qquad \text{(from (9))} \\ &= \sum_{i \in I_k} \big[(\widehat{Q}_k 1_{\mathbb{X}_{j,k}})(x_{i,k}) + 1_{\mathbb{X}_{j,k}}(x_0)\, \psi_k(x_{i,k})\big]\, 1_{\mathbb{X}_{i,k}} + \big[(\widehat{Q}_k 1_{\mathbb{X}_{j,k}})(\overline{x}_k) + 1_{\mathbb{X}_{j,k}}(x_0)\, \psi_k(\overline{x}_k)\big]\, b_k \\ &= \sum_{i \in I_k} \big[(\widehat{Q}_k 1_{\mathbb{X}_{j,k}})(x_{i,k}) + 1_{\mathbb{X}_{j,k}}(x_0)\, \psi_k(x_{i,k})\big]\, 1_{\mathbb{X}_{i,k}} + 1_{\mathbb{X}_{j,k}}(x_0)\, b_k. \end{aligned}$$

The previous equalities show that B_k is a non-negative matrix. Moreover Equality $P_k 1_{\mathbb{X}} = 1_{\mathbb{X}}$ reads as matrix equality $B_k \cdot \mathbf{1}_k = \mathbf{1}_k$ where $\mathbf{1}_k$ is the coordinate vector of $1_{\mathbb{X}}$ in the basis $\mathcal{C}_k$ and is given by $\mathbf{1}_k = (1, \dots, 1)^\top$. The symbol $\cdot^\top$ stands for the transpose operation. Thus B_k is a stochastic matrix. Accordingly there exists a non-zero row-vector $\pi_k \in [0, +\infty)^{N_k}$ such that

$$\pi_k \cdot B_k = \pi_k, \quad \text{and} \quad \pi_k \cdot \mathbf{1}_k = 1. \tag{11}$$

Note that the last component of π_k (i.e. the component associated with b_k) is zero since the last column of B_k is zero from (10). We denote by $\pi_{i,k}$ the component of π_k associated with the element $1_{\mathbb{X}_{i,k}}$ of the basis $\mathcal{C}_k$, so that the coordinate vector of π_k in $\mathcal{C}_k$ is $(\{\pi_{i,k}\}_{i \in I_k}, 0)$. For every $k \geq 1$ we set

$$\widehat{\pi}_k(f) := \pi_k \cdot F_k \tag{12}$$

where $F_k \equiv F_k(f)$ is the coordinate vector of $\widehat{P}_k f$ in the basis $\mathcal{C}_k$.

Proposition 2.1 *$\widehat{\pi}_k$ defines a $\widehat{P}_k$-invariant probability measure on $(\mathbb{X}, \mathcal{X})$. Moreover we have*

$$\widehat{\pi}_k(dy) = \mathfrak{p}_k(y)\, d\mu(y) + \left(1 - \int_{\mathbb{X}} \mathfrak{p}_k(y)\, d\mu(y)\right) \delta_{x_0}, \tag{13}$$

where δ_{x_0} is the Dirac distribution at x_0, and where $\mathfrak{p}_k$ is the non-negative function defined by

$$\forall y \in \mathbb{X}, \quad \mathfrak{p}_k(y) := 1_{\mathbb{X}_k}(y) \sum_{i \in I_k} \pi_{i,k} \inf_{t \in \mathbb{X}_{i,k}} p(t, y), \tag{14}$$

Note that Formula (14) involves the infimum of the function $t \mapsto p(t, y)$ on each subset $\mathbb{X}_{i,k}$. This is a technical choice to ensure that Lemma 3.3 holds true for $\widehat{P}_k$. Specifically, this a simple choice to simplify the convergence analysis in Section 3 of the approximation scheme.

Proof. Recall that b_k is defined by $b_k = 1_{\mathbb{X}} - \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}$. From $\psi_k := 1_{\mathbb{X}} - \widehat{Q}_k 1_{\mathbb{X}}$ it follows that $\psi_k = b_k + \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}} - \widehat{Q}_k 1_{\mathbb{X}}$. Define

$$m_{i,k}(f) := \int_{\mathbb{X}_k} f(y) \inf_{t \in \mathbb{X}_{i,k}} p(t, y)\, d\mu(y) \tag{15}$$

and observe that $\widehat{Q}_k f = \sum_{i \in I_k} m_{i,k}(f)\, 1_{\mathbb{X}_{i,k}}$. Then we deduce from (6) and (7) that

$$\begin{aligned} \widehat{P}_k f := (\widehat{Q}_k f) + f(x_0)\, \psi_k &= \sum_{i \in I_k} m_{i,k}(f)\, 1_{\mathbb{X}_{i,k}} + f(x_0)\big(b_k + \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}} - \widehat{Q}_k 1_{\mathbb{X}}\big) \\ &= \sum_{i \in I_k} \big[m_{i,k}(f) + f(x_0) - f(x_0)\, m_{i,k}(1_{\mathbb{X}})\big]\, 1_{\mathbb{X}_{i,k}} + f(x_0) b_k, \end{aligned}$$

so that (12) and $\sum_{i \in I_k} \pi_{i,k} = 1$ give

$$\begin{aligned} \widehat{\pi}_k(f) &:= \sum_{i \in I_k} \pi_{i,k} \left[m_{i,k}(f) + f(x_0) - f(x_0)\, m_{i,k}(1_{\mathbb{X}})\right] \\ &= \sum_{i \in I_k} \pi_{i,k}\, m_{i,k}(f) + f(x_0)\left(1 - \sum_{i \in I_k} \pi_{i,k}\, m_{i,k}(1_{\mathbb{X}})\right). \end{aligned} \tag{16}$$

This proves Formula (13). Now we prove that $\widehat{\pi}_k$ defines a $\widehat{P}_k$-invariant probability measure on $(\mathbb{X}, \mathcal{X})$. Note that

$$\forall i \in I_k, \quad m_{i,k}(1_{\mathbb{X}}) \leq \int_{\mathbb{X}} p(x_{i,k}, y)\, d\mu(y) = (P1_{\mathbb{X}})(x_{i,k}) = 1,$$

thus

$$\int_{\mathbb{X}} \mathfrak{p}_k(y)\, d\mu(y) = \sum_{i \in I_k} \pi_{i,k}\, m_{i,k}(1_{\mathbb{X}}) \leq 1.$$

It follows from this remark and from (16) that $\widehat{\pi}_k$ is a probability measure on $\mathbb{X}$. Finally $B_k \cdot F_k$ is the coordinate vector of $\widehat{P}_k^2 f$ in $\mathcal{C}_k$ since $\widehat{P}_k f \in \mathcal{F}_k$ and F_k is the coordinate vector of $\widehat{P}_k f$ in $\mathcal{C}_k$. Consequently we deduce from (12) and (11) that

$$\widehat{\pi}_k(\widehat{P}_k f) := \pi_k \cdot B_k \cdot F_k = \pi_k \cdot F_k = \widehat{\pi}_k(f).$$

Thus $\widehat{\pi}_k$ is $\widehat{P}_k$-invariant. □

3 Convergence of $(\widehat{\pi}_k)_{k \geq 1}$ to π in total variation distance

The metric space $\mathbb{X}$ is equipped with a sequence of partitions satisfying Definition **(A)**. The Markov kernel P on $\mathbb{X}$ is assumed to be of the form (1). Let $\theta \in (0, 1]$. For $k \geq 1$, $i \in I_k$, and $y \in \mathbb{X}_k$, we denote by $L_{i,k,\theta}(y)$ the following quantity in $[0, +\infty]$

$$L_{i,k,\theta}(y) := \sup \left\{ \frac{|p(x,y) - p(x',y)|}{d(x,x')^\theta},\ (x,x') \in \mathbb{X}_{i,k} \times \mathbb{X}_{i,k},\ x \neq x' \right\}. \tag{17}$$

Finally we assume that P satisfies the following assumptions

$$\exists\, \delta \in (0,1),\ \exists M \in (0, +\infty), \quad PV \leq \delta V + M \tag{18a}$$

$$\alpha_k := \sup_{u \in \mathbb{X}_k} \frac{P\big(u, \mathbb{X} \setminus \mathbb{X}_k\big)}{V(u)} \longrightarrow 0 \quad \text{when } k \to +\infty \tag{18b}$$

$$\exists\, \theta \in (0,1],\ \forall k \geq 1,\ \ell_{k,\theta} := \max_{i \in I_k} \int_{\mathbb{X}_k} L_{i,k,\theta}(y)\, d\mu(y) < \infty, \quad \text{and } \lim_{k \to +\infty} \ell_{k,\theta}\, \delta_k^\theta = 0. \tag{18c}$$

Actually (18a) is a drift type inequality (see [MT93]) which comes from the V-geometric ergodicity assumption (3). Technical conditions (18b) and (18c) are used to control the weak convergence of $(\widehat{P}_k)_{k \geq 1}$ to P (see Lemma 3.4). In the first order autoregressive models of Section 4, condition (18b) reduces to a polynomial moment condition on the noise (see (25) for instance), and Condition (18c) reduces to the control of the derivative of the noise (see (26) for instance).

Theorem 3.1 *Let $(\delta_k)_{k \geq 1}$ be a sequence of positive real numbers from Definition* **(A)**. *Assume that P is a V-geometrically ergodic Markov kernel of the form (1), and finally that Assumptions (18a)-(18c) hold. Then the probability measures $\widehat{\pi}_k$ on $\mathbb{X}$ given in (13) are such that $\|\pi - \widehat{\pi}_k\|_{TV} \to 0$ when $k \to +\infty$, more precisely:*

$$\|\pi - \widehat{\pi}_k\|_{TV} = O\big(|\ln \tau_k|\, \tau_k\big) \quad \textit{with} \quad \tau_k = 2 \max\left(\frac{1}{v(k)}\ ,\ \alpha_k + \ell_{k,\theta}\, \delta_k^\theta \right). \tag{19}$$

Let $(\mathcal{B}_0, \|\cdot\|_0)$ denote the Banach space of bounded measurable $\mathbb{C}$-valued functions on $\mathbb{X}$ equipped with the norm $\|f\|_0 := \sup_{x\in\mathbb{X}} |f(x)|$. Then (19) means that

$$\forall k \geq 1,\ \forall f \in \mathcal{B}_0, \quad \big|\pi(f) - \widehat{\pi}_k(f)\big| \leq \gamma_k \, \|f\|_0$$

with $\gamma_k = \mathrm{O}\big(|\ln \tau_k|\, \tau_k\big)$. Recall that $\pi(dy) = \mathfrak{p}(y) d\mu(y)$. Assume that $\mu(\{x_0\}) = 0$. Then, using (13), the previous inequalities applied to $f := 1_{\{x_0\}}$ imply

$$0 \leq 1 - \int_{\mathbb{X}} \mathfrak{p}_k(y)\, d\mu(y) \leq \gamma_k.$$

Hence

$$\forall k \geq 1,\ \forall f \in \mathcal{B}_0, \quad \Big| \int_{\mathbb{X}} f(y)\, \mathfrak{p}(y)\, d\mu(y) - \int_{\mathbb{X}} f(y)\, \mathfrak{p}_k(y)\, d\mu(y)\Big| \leq 2\, \gamma_k \, \|f\|_0,$$

from which we deduce the following corollary.

Corollary 3.2 *Assume that the assumptions of Theorem 3.1 hold and that $\mu(\{x_0\}) = 0$. Then the sequence $(\mathfrak{p}_k)_{k\geq 1}$ given in (14) converges to $\mathfrak{p}$ in the usual Lebesgue space $\mathbb{L}^1(\mathbb{X}, \mathcal{X}, \mu)$, more precisely*

$$\int_{\mathbb{X}} \big|\mathfrak{p}(y) - \mathfrak{p}_k(y)\big|\, d\mu(y) = O\big(|\ln \tau_k|\, \tau_k\big). \tag{20}$$

Proof of Theorem 3.1. We apply [HL14, Prop. 2.1(b)] based on the Keller-Liverani perturbation theorem [KL99]. Define $(\mathcal{B}_1, \|\cdot\|_1)$ as the weighted-supremum Banach space

$$\mathcal{B}_1 := \big\{\ f : \mathbb{X} \to \mathbb{C},\ \text{measurable}\ : \|f\|_1 := \sup_{x\in\mathbb{X}} |f(x)| V(x)^{-1} < \infty\ \big\}.$$

Note that Inequality (3) writes as follows

$$\forall n \geq 0,\ \forall f \in \mathcal{B}_1, \quad \|P^n f - \pi(f)\, 1_{\mathbb{X}}\|_1 \leq C\, \rho^n\, \|f\|_1.$$

Since $p_k(x,y) \leq p(x,y)$, $\widehat{P}_k$ continuously acts on both $\mathcal{B}_0$ and $\mathcal{B}_1$. In fact $\widehat{P}_k$ is finite-rank, more precisely

$$\widehat{P}_k(\mathcal{B}_1) \subset \mathcal{F}_k$$

with $\mathcal{F}_k$ given in Section 2 (see (8)). Note that $\widehat{\pi}_k$ clearly defines a non-negative bounded linear form on $\mathcal{B}_1$. Then, according to [HL14, Prop. 2.1(b)], Property (19) follows from the next Lemmas 3.3 and 3.4. □

Lemma 3.3 *We have*

$$\forall k \geq 1, \quad \widehat{P}_k V \leq \delta V + L \qquad \textit{with}\ \ L := M + 1\ \textit{and}\ M\ \textit{given in (18a).}$$

Proof. If $x \in \mathbb{X} \setminus \mathbb{X}_k$, then $(\widehat{P}_k V)(x) = V(x_0)\, \psi_k(x) \leq 1$. If $x \in \mathbb{X}_k$, then we obtain (see (15)):

$$\begin{aligned}
(\widehat{P}_k V)(x) &= \sum_{i\in I_k} m_{i,k}(V)\, 1_{\mathbb{X}_{i,k}}(x) + V(x_0)\, \psi_k(x) \\
&\leq \sum_{i\in I_k} \Big(\int_{\mathbb{X}} V(y)\, p(x,y)\, d\mu(y) \Big) 1_{\mathbb{X}_{i,k}}(x) + 1 \\
&\leq (PV)(x) + 1.
\end{aligned}$$

The desired inequality follows from (18a). □

Lemma 3.4 *For every $k \geq 1$ we have:* $\displaystyle\sup_{f \in \mathcal{B}_0,\, \|f\|_0 \leq 1} \|\widehat{P}_k f - Pf\|_1 \;\leq\; \tau_k.$

Proof. Let $f \in \mathcal{B}_0$, $\|f\|_0 \leq 1$. If $x \in \mathbb{X} \setminus \mathbb{X}_k$, it follows from $(\widehat{P}_k f)(x) = f(x_0)\, \psi_k(x)$ that

$$\frac{\big|(\widehat{P}_k f)(x) - (Pf)(x)\big|}{V(x)} \leq \frac{\psi_k(x) + (P|f|)(x)}{V(x)} \leq \frac{2}{V(x)} \leq \frac{2}{v(k)}. \tag{21}$$

Next assume that $x \in \mathbb{X}_k$. Then we obtain from the definition of $\widehat{Q}_k$ that

$$\begin{aligned}
\frac{\big|(\widehat{Q}_k f)(x) - (Pf)(x)\big|}{V(x)} &\leq \int_{\mathbb{X}} \frac{\big|p_k(x,y) - p(x,y)\big|}{V(x)}\, d\mu(y) \\
&\leq \underbrace{\int_{\mathbb{X}\setminus\mathbb{X}_k} \frac{\big|p_k(x,y) - p(x,y)\big|}{V(x)}\, d\mu(y)}_{:=\alpha_k(x)} + \underbrace{\int_{\mathbb{X}_k} \frac{\big|p_k(x,y) - p(x,y)\big|}{V(x)}\, d\mu(y)}_{:=\beta_k(x)}
\end{aligned}$$

Since $p_k(x,y) = 0$ when $y \in \mathbb{X} \setminus \mathbb{X}_k$, we obtain that

$$\alpha_k(x) = \int_{\mathbb{X}\setminus\mathbb{X}_k} \frac{p(x,y)}{V(x)}\, d\mu(y) = \frac{P\big(x, \mathbb{X} \setminus \mathbb{X}_k\big)}{V(x)} \leq \alpha_k$$

from the definition (18b) of α_k. Now, since $V \geq 1$, it follows from Conditions (5) and (18c) that

$$\begin{aligned}
\beta_k(x) &= \int_{\mathbb{X}_k} \Big| \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x) \inf_{t \in \mathbb{X}_{i,k}} p(t,y) - \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x)\, p(x,y) \Big|\, d\mu(y) \\
&\leq \int_{\mathbb{X}_k} \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x) \big|p(x,y) - \inf_{t \in \mathbb{X}_{i,k}} p(t,y)\big|\, d\mu(y) \\
&\leq \int_{\mathbb{X}_k} \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x) \sup_{u \in \mathbb{X}_{i,k}} \big|p(x,y) - p(u,y)\big|\, d\mu(y) \\
&\leq \delta_k^\theta \int_{\mathbb{X}_k} \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x)\, L_{i,k,\theta}(y)\, d\mu(y) \\
&\leq \delta_k^\theta \sum_{i \in I_k} 1_{\mathbb{X}_{i,k}}(x) \int_{\mathbb{X}_k} L_{i,k,\theta}(y)\, d\mu(y) \\
&\leq \delta_k^\theta\, \ell_{k,\theta}.
\end{aligned}$$

We have proved that, for every $f \in \mathcal{B}_0$ such that $\|f\|_0 \leq 1$ and for every $x \in \mathbb{X}_k$, we have

$$\frac{\big|(\widehat{Q}_k f)(x) - (Pf)(x)\big|}{V(x)} \leq \alpha_k + \ell_{k,\theta}\, \delta_k^\theta. \tag{22}$$

Moreover we deduce from the definition of ψ_k and from (22) that

$$0 \leq \frac{\psi_k(x)}{V(x)} = \frac{1 - (\widehat{Q}_k 1_{\mathbb{X}})(x)}{V(x)} = \frac{(P1_{\mathbb{X}})(x) - (\widehat{Q}_k 1_{\mathbb{X}})(x)}{V(x)} \leq \alpha_k + \ell_{k,\theta}\, \delta_k^\theta. \tag{23}$$

It follows from Inequalities (22) and (23) that, for every $f \in \mathcal{B}_0$ such that $\|f\|_0 \leq 1$ and for every $x \in \mathbb{X}_k$, we have:

$$\begin{aligned}\frac{\big|(\widehat{P}_k f)(x) - (Pf)(x)\big|}{V(x)} &= \frac{\big|(\widehat{Q}_k f)(x) + f(x_0)\psi_k(x) - (Pf)(x)\big|}{V(x)} \\ &\leq \frac{\psi_k(x)}{V(x)} + \frac{\big|(\widehat{Q}_k f)(x) - (Pf)(x)\big|}{V(x)} \\ &\leq 2\big(\alpha_k + \ell_{k,\theta}\,\delta_k^\theta\big).\end{aligned}$$

This inequality and (21) provide the conclusion of Lemma 3.4. □

Remark 3.5 *The inequality (b) of [HL14, Prop. 2.1] provides explicit bounds in (19) and (20) in terms of the constants δ, L in (18a) and the constants C and ρ in (3). Unfortunately, finding explicit constants $\rho \in (0,1)$ and $C > 0$ in (3) is a difficult issue, even for simple models as AR(1). Such constants can be obtained in our context by applying the procedure of [HL14, Th. 4.1], but the resulting constant C is too large to be numerically interesting. An alternative way is to use, for k larger and larger, the bound provided by Inequality (a) of [HL14, Prop. 2.1(a)], which is only based on the spectral properties of the finite stochastic matrix B_k. But again the resulting constants are too large. In fact the numerical applications presented in Section 6 show that the convergence in (19) and (20) is much better than what is provided by using the constants derived from [HL14, Prop. 2.1].*

4 Applications to first order autoregressive processes

4.1 The standard AR(1) process

Let $(X_n)_{n\in\mathbb{N}}$ be a standard first order linear autoregressive process, that is

$$\forall n \geq 1, \quad X_n = \varrho\, X_{n-1} + \vartheta_n, \tag{24}$$

where X_0 is a real-valued random variable (r.v.) and $(\vartheta_n)_{n\in\mathbb{N}}$ is a sequence of real-valued independent and identically distributed (i.i.d.) random variables, also assumed to be independent from X_0. We suppose that $|\varrho| < 1$, that ϑ_1 has a pdf ν, called the innovation density function, with respect to the Lebesgue measure $d\mu(y) := dy$ on $\mathbb{R}$. Assume that the three following conditions are satisfied:

(a) ϑ_1 has a moment of order m for some $m \in [1, +\infty)$, namely

$$\exists\, m \in [1,+\infty), \quad \eta_m := \int_{\mathbb{R}} |x|^m\, \nu(x)\, dx < \infty; \tag{25}$$

(b) ν is continuously differentiable on $\mathbb{R}$ and its derivative ν' is assumed to be right differentiable on $\mathbb{R}$;

(c) finally

$$I' := \int_{\mathbb{R}} |\nu'(y)|\, dy < \infty \quad \text{and} \quad M'' := \sup_{t\in\mathbb{R}} |\nu_r''(t)| < \infty \tag{26}$$

where $\nu_r''(t)$ denotes the right derivative of ν' at t.

Let $\mathbb{X} := \mathbb{R}$ be equipped with its usual distance $d(x,x') := |x-x'|$ and with its Borel σ-algebra $\mathcal{X}$. Recall that $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with transition kernel P defined by

$$\forall A \in \mathcal{X}, \quad P(x,A) := \int_{\mathbb{R}} 1_A(y)\, p(x,y)\, dy \qquad \text{with} \;\; p(x,y) := \nu(y-\varrho x). \tag{27}$$

It is well-known from [MT93] that $(X_n)_{n\in\mathbb{N}}$ admits a unique stationary probability measure π on $\mathbb{R}$, and that π is absolutely continuous with respect to the Lebesgue measure, with density function $\mathfrak{p}$ such that $\int_{\mathbb{R}} |y|^m \mathfrak{p}(y) dy < \infty$ and satisfying

$$\forall x \in \mathbb{R}, \quad \mathfrak{p}(x) = \int_{\mathbb{R}} \nu(x-\varrho u)\, \mathfrak{p}(u)\, du.$$

For $x \in \mathbb{R}$, we define $V(x) = \lfloor 1 + |x|^m \rfloor$ where m is the positive real number given in (25) and where $\lfloor \cdot \rfloor$ denotes the integer part function on $\mathbb{R}$. According to (25) and $|\varrho| < 1$, P is V-geometrically ergodic (see [MT93]). In the sequel we fix any $\delta \in (|\varrho|^m, 1)$. It can be easily deduced from (25) that there exists $M \equiv M(\delta)$ such that

$$PV \le \delta V + M. \tag{28}$$

For every $k \ge 1$, we choose $\delta_k > 0$ such that $\delta_k = \mathrm{O}(1/k)$ and (for the sake of simplicity) such that $q_k := 2k/\delta_k \in \mathbb{N}$. Set $\mathbb{X}_k = [-k, k[$, and consider the following partition of $\mathbb{X}_k$:

$$\mathbb{X}_k := \bigsqcup_{i=0}^{q_k-1} \mathbb{X}_{i,k} \qquad \text{with} \;\; \mathbb{X}_{i,k} := \big[x_{i,k}, x_{i+1,k}\big[, \quad x_{i,k} = -k + i\,\delta_k. \tag{29}$$

The associated discretized Markov kernels $\widehat{P}_k$ and the probability measures $\widehat{\pi}_k$ on $\mathbb{R}$ are defined by (7) and (12) respectively. The associated function $\mathfrak{p}_k$ is given in (14).

Proposition 4.1 *Let δ_k be such that $\delta_k > 0$ and $\delta_k = O(1/k)$. Assume that the innovation density function $\nu(\cdot)$ satisfies Conditions (25) and (26). Then*

$$\|\pi - \widehat{\pi}_k\|_{TV} = O\big(|\ln \tau_k|\, \tau_k\big), \quad \|\mathfrak{p} - \mathfrak{p}_k\|_{\mathbb{L}^1(\mathbb{R})} = O\big(|\ln \tau_k|\, \tau_k\big) \quad \textit{with } \tau_k = \frac{1}{k^m} + \delta_k. \tag{30}$$

Proof. Proposition 4.1 follows from Theorem 3.1 and Corollary 3.2, provided that Assumptions (18b) and (18c) are satisfied (all the others assumptions of Theorem 3.1 have been already checked above). First the real number α_k in (18b) satisfies

$$\alpha_k = \int_{|y|>k(1-|\varrho|)} \nu(y)\, dy \;\le \frac{\eta_m}{(1-|\varrho|)^m\, k^m}$$

from Markov's inequality. Thus (18b) holds. Second, we obtain for every $(x,x') \in \mathbb{X}_{i,k}$

$$\begin{aligned}
|p(x,y) - p(x',y)| &= \big|\nu(y-\varrho x) - \nu(y-\varrho x')\big| \\
&\le |\varrho|\,|x-x'|\,\big|\nu'(y-\varrho\, c)\big| \quad \text{for some} \;\; c \equiv c_{x,x',y} \in \mathbb{X}_{i,k} \\
&\le |\varrho|\,|x-x'|\,\big(\big|\nu'(y-\varrho\, c) - \nu'(y-\varrho\, x_{i,k})\big| + \big|\nu'(y-\varrho\, x_{i,k})\big|\big) \\
&\le |\varrho|\,|x-x'|\,\big(|\varrho|\, M''\, \delta_k + \big|\nu'(y-\varrho\, x_{i,k})\big|\big)
\end{aligned}$$

so that

$$\begin{aligned} L_{i,k,1}(y) &:= \sup\left\{ \frac{|p(x,y)-p(x',y)|}{|x-x'|},\ (x,x') \in \mathbb{X}_{i,k} \times \mathbb{X}_{i,k},\ x \neq x' \right\} \\ &\leq |\varrho| \left(|\varrho|\, M''\, \delta_k + \left|\nu'(y - \varrho\, x_{i,k})\right| \right). \end{aligned}$$

Using the notations of (26), we obtain that

$$\ell_{k,1} := \max_{i \in I_k} \int_{-k}^{k} L_{i,k,\theta}(y)\, dy \leq 2\, |\varrho|^2\, M'' k\, \delta_k + |\varrho|\, I'. \tag{31}$$

Recall that $\delta_k = \mathrm{O}(1/k)$ by hypothesis, so that $\sup_{k \geq 1} \ell_{k,1} < \infty$. Hence (18c) holds. □

Remark 4.2 *Alternative assumptions (instead of (26)) on the innovation density function ν are possible. For instance, in place of (26), we may suppose that the derivative ν' of ν exists and that $D := \sup_{x \in \mathbb{R}} |\nu'(x)| < \infty$. Then $L_{i,k,1}(\cdot) \leq D$, so that $\ell_{k,1} = O(k)$. Thus, provided that $\delta_k = o(1/k)$, the statements of Proposition 4.1 is replaced with the following ones: $\|\pi - \widehat{\pi}_k\|_{TV}$ and $\|\mathfrak{p} - \mathfrak{p}_k\|_{\mathbb{L}^1(\mathbb{R})}$ are both $O\big(|\ln \tau_k|\, \tau_k\big)$ with $\tau_k = 1/k^m + k\, \delta_k$ and both $\|\pi - \widehat{\pi}_k\|_{TV}$ and $\|\mathfrak{p} - \mathfrak{p}_k\|_{\mathbb{L}^1(\mathbb{R})}$ converge to 0 when $k \to +\infty$.*

Remark 4.3 *In [DDGMR00] a similar state-discretization procedure is proposed to estimate the spectrum of the Markov kernel P given in (27). Because the authors of [DDGMR00] use the standard perturbation theory, they have to assume that the innovation density function is compactly supported in some interval $[a,b]$ (the action of P is then considered on the usual Lebesgue space $\mathbb{L}^2([a,b])$. The use of the Keller-Liverani perturbation theorem in our work (see the proof of Theorem 3.1) allows us to consider innovation density functions with unbounded support.*

Remark 4.4 *If $(X_n)_{n \in \mathbb{N}}$ is a first order autoregressive model given by (24), then for any $\ell \geq 1$ the sequence $(X_n^{(\ell)})_{n \geq 0}$ defined by $X_n^{(\ell)} := X_{\ell n}$ satisfies the following linear recursion*

$$\forall n \geq 1, \quad X_n^{(\ell)} = \varrho^{\ell}\, X_{n-1}^{(\ell)} + \vartheta_n^{(\ell)} \quad \text{with } \vartheta_n^{(\ell)} = \sum_{k\, =\, \ell(n-1)+1}^{\ell\, n} \varrho^{\ell n - k}\, \vartheta_k. \tag{32}$$

The sequence $(\vartheta_n^{(\ell)})_{n \geq 0}$ is i.i.d., and $(X_n^{(\ell)})_{n \geq 0}$ is a first order autoregressive model having the same stationary density function as $(X_n)_{n \in \mathbb{N}}$. The transition kernel of $(X_n^{(\ell)})_{n \geq 0}$ is P^{ℓ}, which is of the form (1) too. More precisely, for every $x \in \mathbb{R}$ we have $P^{\ell}(x, dy) = p_{\ell}(x,y)\, dy$ with $p_{\ell}(x,y) := \nu_{\ell}(y - \varrho^{\ell} x)$, where ν_{ℓ} is the pdf of $\vartheta_1^{(\ell)}$, that is $\nu_{\ell} := \mu_{\ell} \star \cdots \star \mu_1$, where μ_k denotes the pdf of the r.v. $\varrho^{\ell-k}\, \vartheta_k$ for $k = 1, \ldots, \ell$, and where the symbol "$\star$" stands for the standard convolution product. This fact may be relevant since ν_{ℓ} is more and more regular as ℓ increases, so that ν_{ℓ} may satisfy the regularity condition required in (26) for ℓ large enough. In this case the stationary density function of $(X_n)_{n \in \mathbb{N}}$ can be approximated by applying Proposition 4.1 to P^{ℓ} (thus with ν_{ℓ} in place of ν). For instance, if the innovation law is the uniform distribution on $[0,1]$, then Proposition 4.1 applies to the Markov kernel P^3 since the associated innovation density function (i.e. the law of $\vartheta_1^{(3)} := \varrho^2\, \vartheta_1 + \varrho\, \vartheta_2 + \vartheta_3$) is continuously differentiable on $\mathbb{R}$ and satisfies (26).

4.2 The AR(1) process with ARCH(1) errors

The following example is derived from [BK01]. Let $\mathbb{X} := \mathbb{R}$ be equipped with its usual distance $d(x, x') := |x - x'|$ and with its Borel σ-algebra $\mathcal{X}$. Let $\alpha \in \mathbb{R}$ and let $\beta, \lambda > 0$. We consider the autoregressive process $(X_n)_{n\in\mathbb{N}}$ with ARCH(1) errors, defined by

$$\forall n \geq 1, \quad X_n = \alpha\, X_{n-1} + \big(\beta + \lambda X_{n-1}^2\big)^{1/2}\, \vartheta_n, \tag{33}$$

where X_0 is a real-valued r.v. and $(\vartheta_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. real-valued random variables which are independent from X_0. We suppose that ϑ_1 has a pdf ν with respect to the Lebesgue measure $d\mu(y) := dy$ on $\mathbb{R}$, that ν is a bounded continuously differentiable and symmetric function with full support $\mathbb{R}$, that its derivatives ν' satisfies $|\nu'(x)| = \mathrm{O}_{\pm\infty}(1/|x|)$, and finally that ϑ_1 has a second-order moment. Then $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with transition kernel P defined by $P(x, A) := \int_{\mathbb{R}} 1_A(y)\, p(x, y)\, dy \quad (A \in \mathcal{X})$ with

$$p(x, y) := \big(\beta + \lambda x^2\big)^{-1/2}\, \nu\left(\frac{y - \alpha x}{\big(\beta + \lambda x^2\big)^{1/2}}\right). \tag{34}$$

As in Section 4, for every $k \geq 1$ we consider $q_k := 2k/\delta_k$, $\mathbb{X}_k = [-k, k[$, and $\mathbb{X}_{i,k}$ as in (29). Moreover assume that

$$\mathbb{E}\big[\, \ln\big|\alpha + \sqrt{\lambda}\, \vartheta_1\big|\, \big] < 0. \tag{35}$$

Then there exists $\kappa > 0$ such that, for every $u \in]0, \kappa[$, we have $\mathbb{E}[|\alpha + \sqrt{\lambda}\, \vartheta_1|^u] < 1$, see [BK01, Prop. 2]. Let $\eta \in]0, \min(\kappa, 2)[$.

Proposition 4.5 *Under the previous assumptions and notations, the following estimates hold true:*

$$\|\pi - \widehat{\pi}_k\|_{TV} = O\big(|\ln \tau_k|\, \tau_k\big), \quad \|\mathfrak{p} - \mathfrak{p}_k\|_{\mathbb{L}^1(\mathbb{R})} = O\big(|\ln \tau_k|\, \tau_k\big) \quad \textit{with } \tau_k = \frac{1}{k^{\eta/2}} + k\, \delta_k. \tag{36}$$

Thus $\|\pi - \widehat{\pi}_k\|_{TV}$ and $\|\mathfrak{p} - \mathfrak{p}_k\|_{\mathbb{L}^1(\mathbb{R})}$ converge to 0 when $k \to +\infty$ provided that $\delta_k = o(1/k)$.

Proof. For $x \in \mathbb{R}$, we define $V(x) = 1 + |x|^\eta$. The V-geometrical ergodicity of P, together with Condition (18a), are proved in [BK01, Th. 1]. To study (18b), we assume that $\alpha > 0$ (similar arguments hold if $\alpha < 0$). Note that

$$\int_k^{+\infty} p(x, y)\, dy = \int_{\phi_k(x)} \nu(t)\, dt \qquad \text{with} \quad \phi_k(x) := \frac{k - \alpha x}{\big(\beta + \lambda x^2\big)^{1/2}}.$$

Let $x \in [-k, k]$. If $\sqrt{k} \leq |x| \leq k$, then

$$\frac{1}{1 + |x|^\eta} \int_k^{+\infty} p(x, y)\, dy \leq \frac{1}{1 + k^{\eta/2}}.$$

If $|x| \leq \sqrt{k}$, then $\frac{1}{1+|x|^\eta} \leq 1$ and

$$\int_{\phi_k(x)} \nu(t)\, dt \leq \int_{\phi_k(\sqrt{k})} \nu(t)\, dt \;\; = \mathrm{O}\big(1/k\big)$$

from $\phi_k(\sqrt{k}) \leq \phi_k(x)$, $\phi_k(\sqrt{k}) \sim_{+\infty} (k/\lambda)^{1/2}$, and from Markov's inequality (since by hypothesis ν has a second-order moment). Since $\eta < 2$, we have proved that

$$\sup_{|x|\leq k} \int_k^{+\infty} p(x,y)\, dy = \mathrm{O}\left(\frac{1}{k^{\eta/2}}\right).$$

The same conclusion can be similarly obtained for the term $\sup_{|x|\leq k} \int_{-\infty}^{-k} p(x,y)\, dy$. Consequently α_k in (18b) satisfies: $\alpha_k = \mathrm{O}\big(k^{-\eta/2}\big)$. Next, to obtain (18c) set $M := \sup_{x\in\mathbb{R}} \nu(x)$, $M' := \sup_{x\in\mathbb{R}} \nu'(x)$, and $C := \sup_{x\in\mathbb{R}} |x\, \nu'(x)|$. An easy computation gives

$$\left|\frac{\partial p}{\partial x}(x,y)\right| \;\leq\; \frac{M\lambda|x|}{(\beta+\lambda x^2)^{3/2}} + \frac{M'|\alpha|}{\beta+\lambda x^2} + \frac{C\lambda|x|}{(\beta+\lambda x^2)^{3/2}}.$$

Thus $D := \sup_{(x,y)\in\mathbb{R}^2} |\frac{\partial p}{\partial x}(x,y)| < \infty$, so that the function $L_{i,k,\theta}$ defined in (17) satisfies (with $\theta = 1$): $\forall y \in \mathbb{R}$, $L_{i,k,1}(y) \leq D$. Therefore the real numbers $\ell_{k,\theta}$ in (18c) are such that $\ell_{k,1} \leq 2Dk$. The above inequalities and Theorem 3.1 provide the desired statement in Proposition 4.5. □

5 A generic algorithm to get $\mathfrak{p}_k(y)$

Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition kernel $p(\cdot,\cdot)$. In this section, we propose a generic algorithm to get the material provided by Section 2. Specifically, the focus is on the non-negative function $\mathfrak{p}_k$ (14) which allows us to obtain the approximating invariant probability given by Proposition 2.1. According to Section 2, the following algorithm can be proposed.

1. Fix the positive integer k such that $\mathbb{X}_k := [-k, k[$ and choose the integers k^- et k^+ such that $[k^-, k^+[\subset X_k$ (you can take $k^- = -k, k^+ = k$).
2. Choose a mesh δ_k of the partition of $[k^-, k^+[$ such that the number of intervals of the subdivision is $q_{\max} := (k^+ - k^-)/\delta_k \in \mathbb{N}^*$,

   Let us introduce the $(q_{\max}+1)$ points of the subdivision $\big\{x_{i,k} := k^- + i\delta_k, i = 0, \ldots, q_{\max}\big\}$ and consider the finite partition $\{\mathbb{X}_{i,k} = [x_{i,k}, x_{i+1,k}[, i = 0, \ldots, q_{\max}-1\}$ of $[k^-, k^+[$
3. Introduce
$$p_{i,k}(y) := \inf_{t\in\mathbb{X}_{i,k}} p(t,y).$$
4. Choose $j_0 \in \{1, \ldots, q_{\max}-1\}$, then for $j = j_0$ compute:
$$B_k(i,j_0) \;:=\; \int_{x_{j_0,k}}^{x_{j_0+1,k}} p_{i,k}(y)\, dy + 1 - \int_{k^-}^{k^+} p_{i,k}(y)\, dy \qquad \text{for } i = 0, \ldots, q_{\max}-1$$
$$B_k(q_{\max}, j_0) \;:=\; 1$$
   Compute for $j = 0, \ldots, q_{\max}-1$, $j \neq j_0$,
$$B_k(i,j) \;:=\; \int_{x_{j,k}}^{x_{j+1,k}} p_{i,k}(y)\, dy \qquad \text{for } i = 0, \ldots, q_{\max}-1$$
$$B_k(q_{\max}, j) \;:=\; 0 \qquad \text{pour } i = q_{\max}$$
   Set $B(i,j) = 0$ for $j = q_{\max}$ et $i = 1, \ldots, q_{\max}$.

5. Compute the B_k-invariant probability vector π_k of B_k: it has the form $\pi_k = (\{\pi_{i,k}\}_{0\leq i<q_{\max}}, 0)$
6. Finally, the non-negative function $\mathfrak{p}_k(\cdot)$ is defined by (see (14)):

$$\forall y \in \mathbb{R}, \quad \mathfrak{p}_k(y) := 1_{[k^-,k^+[}(y) \sum_{i=0}^{q_{\max}-1} \pi_{i,k}\, p_{i,k}(y).$$

The third step of the algorithm involves the computation of an extreme value of the function $t \mapsto p(t,y)$ on a small interval (length δ_k). Such a numerical minimization may be computationally expensive. But it can be checked that, for AR(1) models in Subsections 6.1, 6.2, the function $t \mapsto p(t,y)$ has no local minima so that the minimum may be setted to $\min(p(x_{i,k},y), p(x_{i+1,k},y))$. The case of the AR(1) with ARCH(1) errors may produce local minima for some parameter (β,α,λ). But it can be expected that any approximation of $p_{i,k}(y)$ in Step 3. does not provide large numerical errors from the fact that it is made on a very small interval of length $\delta_k << 1$.

Such an algorithm has been implemented using MATLAB software to obtain the numerical results of Section 6.

Remark 5.1 (Multivariate Markov models) *A natural issue is the generalization of the material of Sections 2 and 3 to multivariate Markov models. A general discussion is beyond the scope of this paper. We only mention that technical Conditions (18a,18b,18c) have natural counterparts for multivariate autoregressive models (e.g. see [MT93]). Thus it can seen from this section that the main difficulties in a multidimensional framework are computational issues due to computation of extreme values and integrals.*

6 Numerical examples

6.1 Application to the Gaussian AR(1)

The benchmark model is the Gaussian linear model where the random variables ϑ_n in (24) have Gaussian distribution $\mathcal{N}(0,\sigma^2)$. In such a context, it is well-known that the invariant probability π of the Markov chain $(X_n)_{n\in\mathbb{N}}$ specified by (24) is $\mathcal{N}(0,\sigma^2/(1-\varrho^2))$. Therefore the pdf's ν and $\mathfrak{p}$ are

$$\nu(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right) \qquad \mathfrak{p}(y) = \frac{\sqrt{1-\varrho^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(1-\varrho^2)y^2}{2\sigma^2}\right). \tag{37}$$

Using the algorithm in Section 5, we obtain the following numerical results. For the sake of simplicity, set $\sigma^2 := 1$. The support of the approximation is $\mathbb{X}_k := [-k,k[$ for specific value of the positive integer k, and δ_k is the mesh of the partition of $\mathbb{X}_k$ used for the computation. The supremum norm of the error vector $v_k := \big(\mathfrak{p}_k(x_{i,k}) - \mathfrak{p}(x_{i,k})\big)_{i=0,\ldots,q_k}$ between $\mathfrak{p}_k$ and $\mathfrak{p}$ on the grid of points $\{x_{i,k}, i = 0,\ldots,q_k\}$ given by the partition of $\mathbb{X}_k$ (see (29)) is denoted by $\|v_k\|_\infty$ and reported in Table 1. The Riemann sum estimation $\|v_k\|_{1,R} := \delta_k \|v_k\|_1$ of $\|\mathfrak{p}_k - \mathfrak{p}\|_{\mathbb{L}^1}$ is provided. These errors are computed using a decreasing sequence of meshes δ_k and a support $\mathbb{X}_k$ selected according to the comments of Remark 6.1. As it can be seen, the

quality of the approximation is quite satisfactory. Figure 1 gives the graphs of the two pdf $\mathfrak{p}_k$ and ν. Note that the exact invariant pdf is not reported in Figure 1 since the estimated and exact graphs cannot be distinguished at this scale. From Table 1, whatever the value of ϱ, the errors norms $\|v_k\|_\infty$ or $\|v_k\|_{1,R}$ scale linearly with the mesh δ_k.

	ϱ								
	0.5			**0.7**			**0.9**		
k	8			14			40		
δ_k	0.05	0.02	0.005	0.05	0.02	0.005	0.05	0.02	0.005
$\|v_k\|_{1,R}$	0.01	0.004	0.001	0.0151	0.0061	0.0015	0.0540	0.025	0.0058
$\|v_k\|_\infty$	0.0025	0.001	2.45×10^{-04}	0.0035	0.0014	3.45×10^{-4}	0.0099	0.0041	0.0011

Table 1: Numerical results for the Gaussian linear model

Remark 6.1 *The algorithm is sensitive to the support $\mathbb{X}_k$ of the approximate function $\mathfrak{p}_k$. Indeed, if the value of k in $\mathbb{X}_k$ is too small with respect to the support of the target pdf $\mathfrak{p}$ of π, then the approximate function $\mathfrak{p}_k$ may appear to be far from the target $\mathfrak{p}$.*

6.2 Applications to AR(1) where the invariant pdf $\mathfrak{p}$ is unknown

When the target pdf $\mathfrak{p}$ is unknown, the set $\mathbb{X}_k$ (the support of $\mathfrak{p}_k$, see (29)) can be chosen as follows. If the innovation pdf ν has a support contained in $[-a, a]$ and if $X_0 = 0$, then $\mathbb{P}\{X_n \in [-s_n, s_n]\} = 1$ with $s_n := a\sum_{k=0}^{n-1} \varrho^k$, so that the pdf $\mathfrak{p}$ has support $[-s, s]$ with $s := a/(1-\varrho)$. If ν is not compactly supported, then the previous remark may be applied with a such that $\nu(x)$ is meaningless for $|x| > a$. Obviously this remark may be easily adapted when the exact or approximated support of ν is contained in $[0, a]$. In the previous Gaussian case, although this question is less relevant since the target pdf $\mathfrak{p}$ is known, we take $a = 4$ and $s = 4/(1-\varrho)$ in Figure 1.

6.2.1 Exponential innovation distribution

In this part, the innovation distribution ν is set to the exponential one with parameter 1. Recall that the pdf ν must satisfy the regularity conditions of Proposition 4.1. Therefore, as discussed in Remark 4.4, the pdf ν_3 is used as input in the algorithm instead of ν:

$$\nu_{3,\varrho}(x) = \frac{1}{(1-\varrho)^2}\left(e^{-x} - e^{-x/\varrho} + \frac{\varrho\left(e^{-x/\varrho^2} - e^{-x}\right)}{1+\varrho}\right) 1_{[0,+\infty[}(x)$$

and the dynamics is given by (32) with $\ell := 3$. The support of $\nu_{3,\varrho}$ may be truncated to $[0, a_{3,\varrho}]$ with $a_{3,0.5} = 11, a_{3,0.7} = 12, a_{3,0.9} = 14$, so that the support of $\mathfrak{p}$ may be truncated to $[0, s_\varrho]$ with $s_\varrho = \lfloor a_{3,\varrho}/(1-\varrho^3)\rfloor + 1$, that is $s_{0.5} = 13, s_{0.7} = 19, s_{0.9} = 52$. Thus we use the interval $[0, 13], [0, 19], [0, 52]$ as $\mathbb{X}_k$ for $\varrho := 0.5, 0.7, 0.9$ (apply the above remark with $\nu_{3,\varrho}$ and ϱ^3 in place of ν and ϱ). In Figure 2 are reported the graphs of the estimated $\mathfrak{p}_k$ of the (unknown) invariant pdf $\mathfrak{p}$ for $\varrho = 0.5, 0.7, 0.9$.

The invariant probability distribution π with pdf $\mathfrak{p}$ satisfies $\pi P = \pi$, that is: $\int_{\mathbb{R}} \mathfrak{p}(x)\, p(x, \cdot)dx = \mathfrak{p}(\cdot)$. Such a relation can be checked on the grid of points $\{x_{i,k}, i = 0, \ldots, q_k\}$ given by the

partition of $\mathbb{X}_k$ (see (29)):

$$\forall x_{i,k}, \quad \int_{\mathbb{R}} \mathfrak{p}(x)\, p(x, x_{i,k}) dx = \mathfrak{p}(x_{i,k}).$$

The integral on the left hand side can be estimated using the Riemann sum denoted by $\widetilde{\mathfrak{p}}_k(x_{i,k}) := \sum_j \mathfrak{p}_k(x_{j,k}) P(x_{j,k}, x_{i,k}) \delta_k$. Therefore, in order to get some confidence into the estimated invariant pdf $\mathfrak{p}_k$, the uniform norm of the following vector $w_k := \big(\mathfrak{p}_k(x_{i,k}) - \widetilde{\mathfrak{p}}_k(x_{i,k})\big)_{i=0,\dots,q_k}$ is reported in Table 2. As it can be seen, the results are satisfactory.

Remark 6.2 *The case $\varrho = 0.9$ (even $\varrho = 0.8$) shows that the graphs of $\nu_{3,\varrho}$ and $\mathfrak{p}$ (given by the approximation $\mathfrak{p}_k$) are very far. Consequently, in this case, the method of [Hai98, AH00, ANR07] requires to compute h_N via (4) for some quite large integer N. Since the use of (4) is recursive, the successive approximations $h_1, \dots, h_N$ should involve large cumulative errors. A similar comment holds true in the forthcoming case of the uniform innovation distribution. As mentioned in Introduction, our method does not contain this drawback since it is not based on a recursive algorithm.*

6.2.2 Uniform innovation distribution

Here, the innovation distribution ν is set to the uniform one on $[0, 1]$. As discussed in Remark 4.4, the pdf ν_3 is used as input in the algorithm instead of ν (see [ANR07, p 281-282] for an explicit formula). Here the dynamics is given by (32) with $\ell := 3$. The graphs of $\nu_{3,\varrho}$ with $\varrho = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ are reported in Figures 3, 4 (blue curves). The support of $\nu_{3,\varrho}$ is $[0, 1 + \varrho + \varrho^2]$ so that the support of the target pdf $\mathfrak{p}_\varrho$ is included into $[0, s_\varrho]$ with $s_\varrho := \lfloor (1+\varrho+\varrho^2)/(1-\varrho^3) \rfloor + 1$. Thus we use the intervals $[0, 2], [0, 2], [0, 3], [0, 4], [0, 5], [0, 10]$ as set $\mathbb{X}_k$ for $\varrho := 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. In Figure 3, we report the graphs of the approximated function $\mathfrak{p}_k$ of the (unknown) invariant pdf $\mathfrak{p}$ and the pdf $\nu_{3,\varrho}$ for $\varrho = 0.4, 0.5, 0.6$. The graphs for $\varrho = 0.7, 0.8, 0.9$ are reported in Figure 4. As in the exponential case, the expected invariance of the estimated pdf $\mathfrak{p}_k$ is evaluated by the uniform norm of the following vector $w_k := \big(\mathfrak{p}_k(x_{i,k}) - \widetilde{\mathfrak{p}}_k(x_{i,k})\big)_{i=0,\dots,q_k}$ (see Table 2). The results are still satisfactory.

	Expo		
ϱ	0.5	0.7	0.9
$\|w_k\|_\infty$	8.73×10^{-4}	9.47×10^{-4}	0.0013

	Unif		
ϱ	0.7	0.8	0.9
$\|w_k\|_\infty$	0.0045	0.0050	0.0051

Table 2: AR(1): checking P-invariance of the estimated pdf $\mathfrak{p}_k$ with $\delta_k := 0.02$

6.2.3 AR(1) with ARCH(1) errors

In this part, we apply our generic algorithm to the autoregressive model with ARCH(1) errors and transition kernel defined in (34). The innovation distribution is the standard Gaussian one, that is $\vartheta_n \sim \mathcal{N}(0, 1)$. The estimated invariant pdf $\mathfrak{p}_{15}$ with support $\mathbb{X}_{15} = [-15, 15]$ and the Gaussian pdf are reported in Figure 5 when $(\beta, \alpha, \lambda) = (1, 0.7, 0.2)$. As for AR(1) models, the invariance property of the estimated pdf $\mathfrak{p}_{15}$ is evaluated by $w_{15} := \|\big(\mathfrak{p}_{15}(x_{i,15}) - \widetilde{\mathfrak{p}}_{15}(x_{i,15})\big)_{i=0,\dots,q_{15}}\|_\infty = 0.0223$.

References

[AH00] Jiří Anděl and Karel Hrach. On calculation of stationary density of autoregressive processes. *Kybernetika (Prague)*, 36(3):311–319, 2000.

[ANR07] J. Anděl, I. Netuka, and P. Ranocha. Methods for calculating stationary distribution in linear models of time series. *Statistics*, 41(4):279–287, 2007.

[AR05] J. Anděl and P. Ranocha. Stationary distribution of absolute autoregression. *Kybernetika (Prague)*, 41(6):735–742, 2005.

[BK01] M. Borkovec and C. Klüppelberg. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.*, 11(4):1220–1241, 2001.

[CT86] K. S. Chan and H. Tong. A note on certain integral equations associated with nonlinear time series analysis. *Probab. Theory Relat. Fields*, 73(1):153–158, 1986.

[DDGMR00] J. A. De Doná, G. C. Goodwin, R. H. Middleton, and I. Raeburn. Convergence of eigenvalues in state-discretization of linear stochastic systems. *SIAM J. Matrix Anal. Appl.*, 21(4):1102–1111, 2000.

[FHL13] D. Ferré, L. Hervé, and J. Ledoux. Regular perturbation of V-geometrically ergodic Markov chains. *J. Appl. Probab.*, 50:184–194, 2013.

[Hai98] G. Haiman. Upper and lower bounds for the tail of the invariant distribution of some AR(1) processes. In *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)*, pages 723–730. North-Holland, Amsterdam, 1998.

[HL14] L. Hervé and J. Ledoux. Approximating Markov chains and V-geometric ergodicity via weak perturbation theory. *Stochastic Processes and their Applications*, 124:613–638, 2014.

[Kel82] G. Keller. Stochastic stability in some chaotic dynamical systems. *Monatsh. Math.*, 94(4):313–333, 1982.

[KL99] G. Keller and C. Liverani. Stability of the spectrum for transfer operators. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze Sér. 4*, 28:141–152, 1999.

[Log04] W. Loges. The stationary marginal distribution of a threshold AR(1) process. *J. Time Ser. Anal.*, 25(1):103–125, 2004.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag London Ltd., London, 1993.

[RS18] D. Rudolf and N. Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018.

[SS00] T. Shardlow and A. M. Stuart. A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.*, 37:1120–1137, 2000.

[Tru17] L. Truquet. A perturbation analysis of some Markov chains models with time-varying parameters. *ArXiv e-prints*, June 2017.

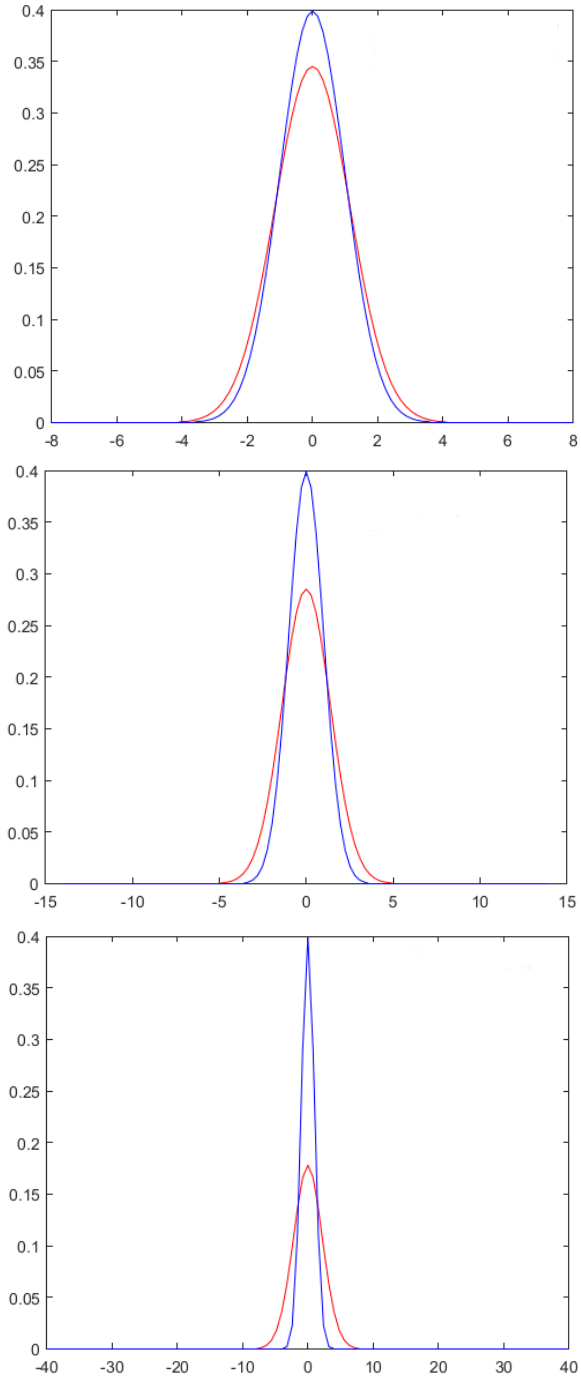

Figure 1: (Gauss) For $\varrho = 0.5, 0.7, 0.9$ and $\delta_k = 0.02$: graphs of the invariant pdf $\mathfrak{p}_8, \mathfrak{p}_{14}, \mathfrak{p}_{40}$ (red) and the innovation pdf ν (blue)

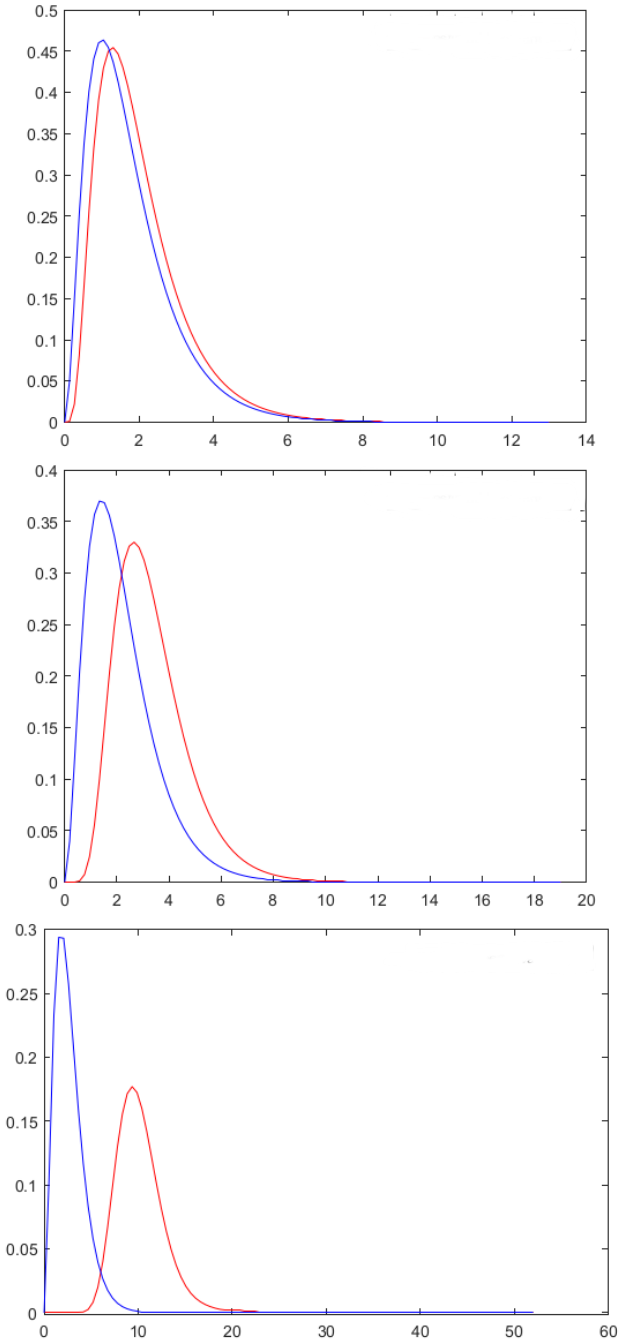

Figure 2: (Expo) For $\varrho = 0.5, 0.7, 0.9$ with $\delta_k = 0.02$: graphs of the estimated invariant pdf $\mathfrak{p}_k$ (red) and the pdf $\nu_{3,\varrho}$ (blue)

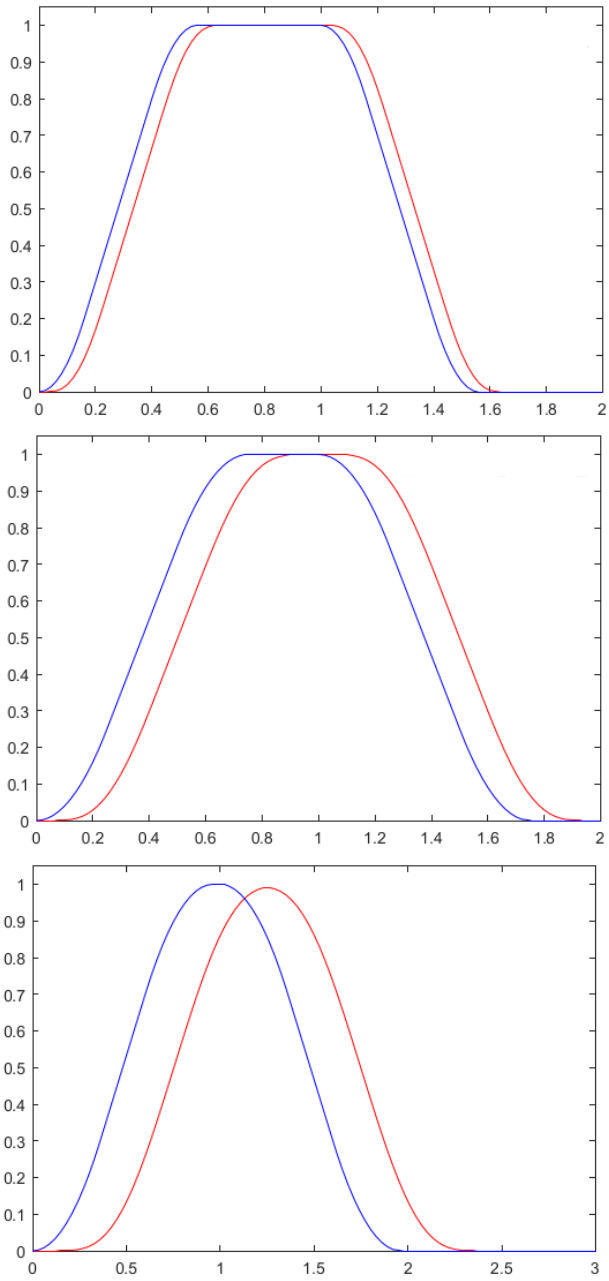

Figure 3: (Unif) For $\varrho = 0.4, 0.5, 0.6$ with $\delta_k = 0.02$: graphs of the estimated pdf $\mathfrak{p}_k$ (red) and $\nu_{3,\varrho}$ (blue)

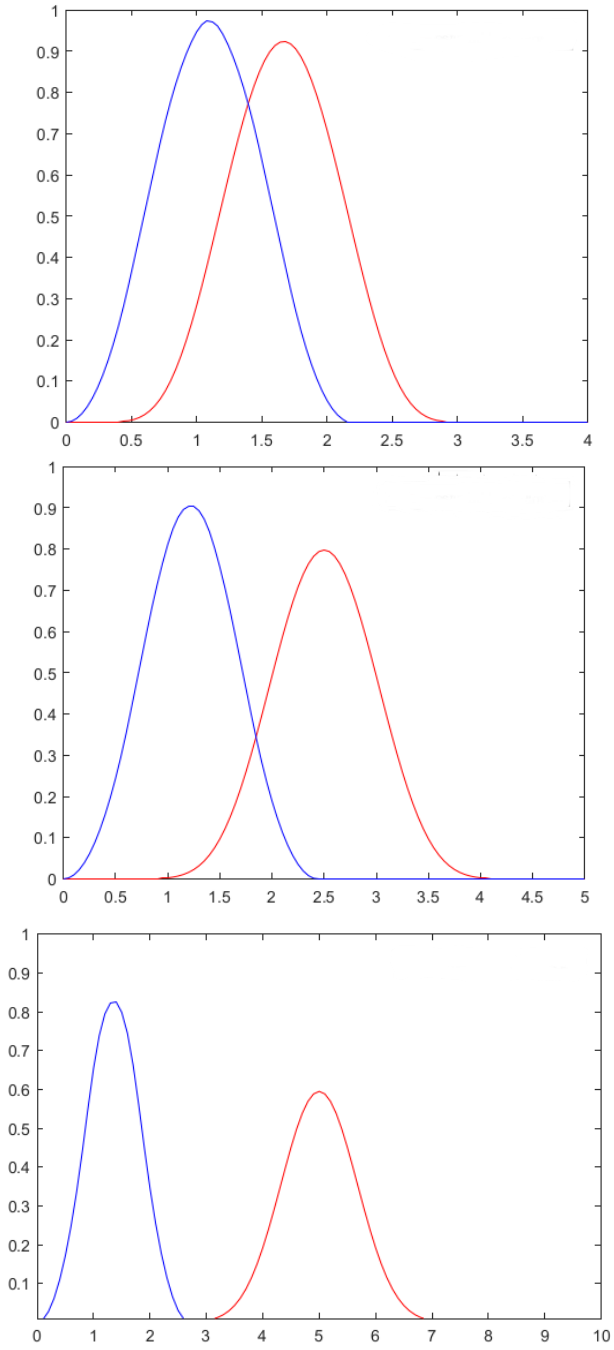

Figure 4: (Unif) For $\varrho = 0.7, 0.8, 0.9$ with $\delta_k = 0.02$: graphs of the estimated pdf $\mathfrak{p}_k$ (red) and $\nu_{3,\varrho}$ (blue)

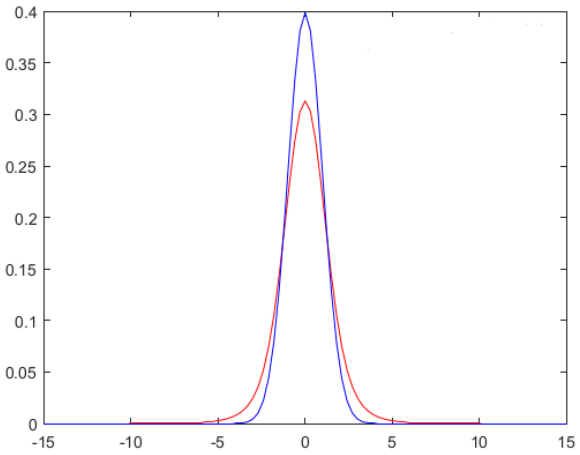

Figure 5: (ARCH) For $\delta_{15} = 0.02$: graphs of the estimated pdf $\mathfrak{p}_{15}$ (red) and the innovation pdf (blue)