Recursive Construction of Confidence Regions

Tomasz R. Bielecki [a], Igor Cialenco [a], Tao Chen [a]

First Circulated: May 18, 2016

Abstract: Assuming that one-step transition kernel of a discrete time, time-homogenous Markov chain model is parameterized by a parameter $\theta \in \mathbf{\Theta}$, we derive a recursive (in time) construction of confidence regions for the unknown parameter of interest, say θ^*. The key step in this construction is derivation of a recursive scheme for an appropriate point estimator of θ^*. To achieve this, we start by what we call the base recursive point estimator, using which we design a quasi-asymptotically linear recursive point estimator (a concept introduced in this paper). For the latter estimator we prove its weak consistency and asymptotic normality. The recursive construction of confidence regions is needed not only for the purpose of speeding up the computation of the successive confidence regions, but, primarily, for the ability to apply the dynamic programming principle in the context of robust adaptive stochastic control methodology.

Keywords: recursive confidence regions; stochastic approximation; recursive point estimators; statistical inference for Markov chains; ergodic processes; quasi-asymptotically linear estimator.

MSC2010: 62M05, 62F10, 62F12, 62F25, 60J05, 60J20.

1 Introduction

Suppose that a set of dynamic probabilistic models is selected and that it is parameterized in terms of a finite dimensional parameter θ taking values in the known parameter space $\mathbf{\Theta}$. We postulate that all these models are possible descriptions of some reality, which is of interest to us, and that only one of the models, say the one corresponding to $\theta^* \in \mathbf{\Theta}$, is the adequate, or true, description of this reality.

Motivated by discrete time robust stochastic control problems subject to model uncertainty (cf. [BCC+16]), we consider in the present paper discrete time, time-homogeneous Markov chain models only. Accordingly, it is the one-step transition kernel of the Markov chain model that is parameterized by θ. We postulate that the true parameter θ^* is not known, and we are interested in deriving a recursive (in time) construction of confidence regions for θ^*. Needless to say, we are seeking a recursive construction of confidence regions for θ^* that satisfy desired properties; in particular, some asymptotic properties, as the time series of observations increases. Robust stochastic control problems provide primary motivation for the present work, but, clearly, potential applications of the results presented here are far reaching.

[a] Department of Applied Mathematics, Illinois Institute of Technology
10 W 32nd Str, Building RE, Room 208, Chicago, IL 60616, USA
Emails: tbielecki@iit.edu (Bielecki), cialenco@iit.edu (Cialenco) and tchen29@hawk.iit.edu (Chen)
URLs: http://math.iit.edu/~bielecki (Bielecki) and http://math.iit.edu/~igor (Cialenco)

Of course, there is a vast literature devoted to recursive computation, also known as on-line computation, of point estimators. It is fair to say though, that, to the best of our knowledge, this work is the first to study a recursive construction of confidence regions. The geometric idea that underlies our recursive construction is motivated by recursive representation of confidence intervals for the mean of one dimensional Gaussian distribution with known variance, and by recursive representation of confidence ellipsoids for the mean and variance of one dimensional Gaussian distribution, where in both cases observations are generated by i.i.d. random variables. The recursive representation is straightforward in the former case, but it is not so any more in the latter one.

The recursive construction of confidence regions is needed not only for the purpose of speeding up the computation of the successive confidence regions, but, primarily, for the ability to apply the dynamic programming principle in the context of robust stochastic control methodology introduced in [BCC+16].

As it will be seen, one of the the key ingredients in our recursive construction of confidence regions is an appropriate recursive scheme for deriving a point estimator of θ^*.

The main contributions of this paper can be summarized as follows:

- We introduce the concept of *quasi-asymptotic linearity of a point estimator* of θ^*. This concept is related to the classic definition of asymptotic linearity of a point estimator, but it requires less stringent properties, which are satisfied by the recursive point estimation scheme that we develop in Section 4.

- Starting from what we call the base recursive point estimation scheme, we design a quasi-asymptotically linear recursive point estimation scheme, and we prove the weak consistency and asymptotic normality of the point estimator of generated by this scheme.

- We provide the relevant recursive construction of confidence regions for θ^*. We prove that these confidence regions are weakly consistent, that is, they converge in probability (in the Hausdorff metric) to the true parameter θ^*.

The paper is organized as follows. In Section 2 we introduce the Markov chain framework relevant for the present study, and we derive an important technical result (Proposition 2.2), which is crucial for recursive identification of the true Markov chain model.

Section 3 is devoted to the recursive construction of what we call the *base (recursive) point estimator* of θ^*. In our set-up, point-estimating of θ^* translates to finding solution to equation (2.10). This is an unknown equation. One of the most widely used iterative root finding procedures for unknown equations is the celebrated stochastic approximation method. Our base (recursive) point estimation scheme for θ^* is an adaptation of the stochastic approximation method. Also, here we prove the strong consistency of the base point estimator.

The key step to the desired recursive construction of confidence regions for θ^* is to establish the asymptotic normality of the underlying recursive point estimator. It turns out that the base (recursive) point estimator constructed in Section 3 may not be asymptotically normal. Therefore, in Section 4 we appropriately modify our base (recursive) point estimator, so to construct a quasi-asymptotically linear (recursive) point estimator, for which we prove weak consistency and asymptotic normality.

The main section of this paper is Section 5, which is devoted to recursive construction of confidence regions for θ^*, and to studying their asymptotic properties. In particular, we show that confidence regions derived from quasi-asymptotically linear (recursive) point estimators preserve a desired geometric structure. Such structure guarantees that we can represent the confidence regions in a recursive way in the sense that the region produced at step n is fully determined by the region produced at step $n-1$ and by the the newly arriving observation of the underlying reality.

Illustrating examples are provided in Section 6. Concluding remarks and open problems are gathered in Section 7, and the paper is completed with three technical Appendices.

2 Preliminaries

Let $(\Omega, \mathscr{F})$ be a measurable space, and $\mathbf{\Theta} \subset \mathbb{R}^d$ be a non-empty set, which will play the role of the parameter space throughout.[1] On the space $(\Omega, \mathscr{F})$ we consider a discrete time, real valued random process $Z = \{Z_n,\ n \geq 0\}$.[2] We postulate that this process is observed, and we denote by $\mathbb{F} = (\mathscr{F}_n, n \geq 0)$ its natural filtration. The (true) law of Z is unknown, and assumed to belong to a parameterized family of probability distributions on $(\Omega, \mathscr{F})$, say $\{\mathbb{P}_\theta, \theta \in \mathbf{\Theta}\}$. It will be convenient to consider $(\Omega, \mathscr{F})$ to be the canonical space for Z, and to consider Z to be the canonical process (see Appendix A for details). Consequently, the law of Z under $\mathbb{P}_\theta$ is the same as $\mathbb{P}_\theta$. The (true) law of Z will be denoted by $\mathbb{P}_{\theta^*}$; accordingly, $\theta^* \in \mathbf{\Theta}$ is the (unknown) true parameter.

The set of probabilistic models that we are concerned with is $\{(\Omega, \mathscr{F}, \mathbb{F}, Z, \mathbb{P}_\theta), \theta \in \mathbf{\Theta}\}$. The model uncertainty addressed in this paper occurs if $\mathbf{\Theta} \neq \{\theta^*\}$, which we assume to be the case. Our objective is to provide a recursive construction of confidence regions for θ^*, based on accurate observations of realizations of process Z through time, and satisfying desirable asymptotic properties.

In what follows, all equalities and inequalities between random variables will be understood in $\mathbb{P}_{\theta^*}$ almost surely sense. We denote by $\mathbb{E}_{\theta^*}$ the expectation operator corresponding to probability $\mathbb{P}_{\theta^*}$.

We impose the the following structural standing assumption.

Assumption M:

(i) Process Z is a time homogenous Markov process under any $\mathbb{P}_\theta,\ \theta \in \mathbf{\Theta}$.

(ii) Process Z is an ergodic Markov process under $\mathbb{P}_{\theta^*}$.[3]

(iii) The transition kernel of process Z under any $\mathbb{P}_\theta,\ \theta \in \mathbf{\Theta}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, that is, for any Borel subset of $\mathbb{R}$

$$\mathbb{P}_\theta(Z_1 \in A \mid Z_0 = x) = \int_A p_\theta(x, y) dy,$$

[1] In general, the parameter space may be infinite dimensional, consisting for example of dynamic factors, such as deterministic functions of time or hidden Markov chains. In this study, for simplicity, we chose the parameter space to be a subset of $\mathbb{R}^d$.

[2] The study presented in this paper extends to the case when process Z takes values in $\mathbb{R}^d$, for $d > 1$. We focus here the case of $d = 1$ for simplicity of presentation.

[3] See Appendix A for the definition of ergodicity that we postulate here.

for some positive and measurable function p_θ.[4]

For any $\theta \in \mathbf{\Theta}$ and $n \geq 1$, we define $\pi_n(\theta) := \log p_\theta(Z_{n-1}, Z_n)$.

Remark 2.1. In view of the Remark A.6, the process Z is a stationary process under $\mathbb{P}_{\theta^*}$. Consequently, under $\mathbb{P}_{\theta^*}$, for each $\theta \in \mathbf{\Theta}$ and for each $n \geq 0$, the law of $\pi_n(\theta)$ is the same as the law of $\pi_1(\theta)$.

We will need to impose several technical assumptions in what follows. We begin with the assumption

R0. For any $\theta \in \mathbf{\Theta}$, $\pi_1(\theta)$ is integrable under $\mathbb{P}_{\theta^*}$.

Then, we have the following result.

Proposition 2.2. *Assume that M and R0 hold. Then,*
(i) For any $\theta \in \Theta$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \pi_i(\theta) = \mathbb{E}_{\theta^*}[\pi_1(\theta)].$$

(ii) Moreover, for any $\theta \in \Theta$,

$$\mathbb{E}_{\theta^*}[\pi_1(\theta^*)] \geq \mathbb{E}_{\theta^*}[\pi_1(\theta)].$$

Proof. Fix $\theta \in \Theta$, since Z is ergodic under $\mathbb{P}_{\theta^*}$ and $\mathbb{E}_{\theta^*}[\pi_1(\theta)] < \infty$, then according to Proposition A.7 we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \pi_i(\theta) = \mathbb{E}_{\theta^*}[\pi_1(\theta)]$$

which proves (i).

Now we prove that (ii) holds. In fact, denote by f_{Z_1} the density function of Z_1 under $\mathbb{P}_{\theta^*}$, we have that

$$\begin{aligned}
&\mathbb{E}_{\theta^*}[\pi_1(\theta)] - \mathbb{E}_{\theta^*}[\pi_1(\theta^*)] \\
=&\mathbb{E}_{\theta^*}\left[\log \frac{p_\theta(Z_1, Z_2)}{p_{\theta^*}(Z_1, Z_2)}\right] = \int_{\mathbb{R}} \mathbb{E}_{\theta^*}\left[\log \frac{p_\theta(Z_1, Z_2)}{p_{\theta^*}(Z_1, Z_2)} \middle| Z_1 = z_1\right] f_{Z_1}(z_1)dz_1 \\
\leq& \int_{\mathbb{R}} \log \mathbb{E}_{\theta^*}\left[\frac{p_\theta(Z_1, Z_2)}{p_{\theta^*}(Z_1, Z_2)} \middle| Z_1 = z_1\right] f_{Z_1}(z_1)dz_1 \\
=& \int_{\mathbb{R}} \log \int_{\mathbb{R}} \frac{p_\theta(z_1, z_2)}{p_{\theta^*}(z_1, z_2)} p_{\theta^*}(z_1, z_2)dz_2 f_{Z_1}(z_1)dz_1 \\
=& \int_{\mathbb{R}} \log \int_{\mathbb{R}} p_\theta(z_1, z_2)dz_2 f_{Z_1}(z_1)dz_1 = 0,
\end{aligned}$$

where the inequality holds due to Jensen's inequality. □

[4]This postulate is made solely in order to streamline the presentation. In general, our methodology works for Markov processes for which the transition kernel is not absolutely continuous with respect to the Lebesgue.

In the statement of the technical assumptions R1-R8 below we use the notations

$$\psi_n(\theta) = \nabla \pi_n(\theta), \quad \Psi_n(\theta) = \mathsf{H}\pi_n(\theta), \quad b_n(\theta) = \mathbb{E}_{\theta^*}[\psi_n(\theta)|\mathscr{F}_{n-1}], \tag{2.1}$$

where ∇ denotes the gradient vector and H denotes the Hessian matrix with respect to θ, respectively.

R1. For each $x, y \in \mathbb{R}$ the function $p_\cdot(x,y) : \mathbf{\Theta} \to \mathbb{R}_+$ is three times differentiable, and

$$\nabla \int_{\mathbb{R}} p_\theta(x,y)dy = \int_{\mathbb{R}} \nabla p_\theta(x,y)dy, \quad \mathsf{H}\int_{\mathbb{R}} p_\theta(x,y)dy = \int_{\mathbb{R}} \mathsf{H}p_\theta(x,y)dy. \tag{2.2}$$

R2. For any $\theta \in \mathbf{\Theta}$, $\psi_1(\theta)$ and $\Psi_1(\theta)$ are integrable under $\mathbb{P}_{\theta^*}$. The function $\mathbb{E}_{\theta^*}[\pi_1(\,\cdot\,)]$ is twice differentiable in θ, and

$$\nabla \mathbb{E}_{\theta^*}[\pi_1(\theta)] = \mathbb{E}_{\theta^*}[\psi_1(\theta)], \quad \mathsf{H}\mathbb{E}_{\theta^*}[\pi_1(\theta)] = \mathbb{E}_{\theta^*}[\Psi_1(\theta)].$$

R3. There exists a unique $\theta \in \mathbf{\Theta}$ such that

$$\mathbb{E}_{\theta^*}[\psi_1(\theta)] = 0.$$

R4. There exists a constant $c > 0$ such that, for any $n \geq 1$ and $\theta \in \mathbf{\Theta}$,

$$\mathbb{E}_{\theta^*}[\|\psi_n(\theta)\|^2 \mid \mathscr{F}_{n-1}] \leq c(1 + \|\theta - \theta^*\|^2). \tag{2.3}$$

R5. There exist some positive constants $K_i, i = 1, 2, 3$, such that for any $\theta, \theta_1, \theta_2 \in \mathbf{\Theta}$, and $n \geq 1$,[5]

$$(\theta - \theta^*)^T b_n(\theta) \leq -K_1\|\theta - \theta^*\|^2, \tag{2.4}$$

$$\|b_n(\theta_1) - b_n(\theta_2)\| \leq K_2\|\theta_1 - \theta_2\|, \tag{2.5}$$

$$\mathbb{E}_{\theta^*}[\|\Psi_n(\theta_1) - \Psi_n(\theta_2)\| \mid \mathscr{F}_{n-1}] \leq K_3\|\theta_1 - \theta_2\|. \tag{2.6}$$

R6. There exists a positive constant K_4, such that for any $\theta \in \mathbf{\Theta}$, and $n \geq 1$,

$$\mathbb{E}_{\theta^*}[\|\mathsf{H}\psi_n(\theta)\||\mathscr{F}_{n-1}] \leq K_4. \tag{2.7}$$

R7. For any $n \geq 1$,

$$\sup_{\theta \in \mathbf{\Theta}} \mathbb{E}_{\theta^*}\|\psi_n(\theta) - b_n(\theta)\|^2 < \infty. \tag{2.8}$$

R8. For each $\theta \in \mathbf{\Theta}$ the Fisher information matrix

$$I(\theta) := \mathbb{E}_\theta[\psi_1(\theta)\psi_1^T(\theta)]$$

exists and is positive definite. Moreover, $I(\theta)$ is continuous with respect to θ.

[5]Superscript T will denote the transpose.

R9.

$$\lim_{n\to\infty} \mathbb{E}_{\theta^*}\left[\sup_{0\le i\le n}\left|\frac{1}{\sqrt{n}}\psi_i(\theta^*)\right|\right] = 0. \tag{2.9}$$

Remark 2.3. (i) Note that in view of the Remark 2.1 properties assumed in R2, R3, and R8 imply that analogous properties hold with time n in place of time 1.
(ii) According to Proposition A.15, we have that if R4-R6 hold, then (2.3)-(2.7) are also satisfied for any $\mathscr{F}_{n-1}$-measurable random vector $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

As stated above, our aim is to provide a recursive construction of the confidence regions for θ^*. In the sequel, we will propose a method for achieving this goal that will be derived from a suitable recursive point estimator of θ^*. Note that due to Proposition 2.2 (ii) and Assumption R3, we have that θ^* is the unique solution of

$$\mathbb{E}_{\theta^*}[\psi_1(\theta)] = 0. \tag{2.10}$$

Therefore, point-estimating θ^* is equivalent to point-estimating the solution of the equation (2.10). Since θ^* is unknown, equation (2.10) is not really known to us. We will therefore apply an appropriate version of the so called *stochastic approximation* method, which is a recursive method used to point-estimate zeros of functions that can not be directly observed. This can be done in our set-up since, thanks to Proposition 2.2 (i), we are provided with a sequence of observed random variables $\frac{1}{n}\sum_{i=1}^n \psi_i(\theta)$ that $\mathbb{P}_{\theta^*}$ almost surely converges to $\mathbb{E}_{\theta^*}[\psi_1(\theta)]$ – a property, which will enable us to adopt the method of stochastic approximation. Accordingly, in the next two sections, we will introduce two recursive point estimators of θ^*, and we will derive properties of these estimators that are relevant for us.

3 $\sqrt{n}$-consistent base point estimator

In this section we consider a recursive point estimator $\tilde{\theta} = \{\tilde{\theta}_n,\ n\ge 1\}$ of θ^*, that will be defined in (3.1). Towards this end, we fix a positive constant β such that $\beta K_1 > \frac{1}{2}$, where K_1 was introduced in Assumption R6, and we define the process $\tilde{\theta}$ recursively as follows,

$$\tilde{\theta}_n = \tilde{\theta}_{n-1} + \frac{\beta}{n}\psi_n(\tilde{\theta}_{n-1}), \quad n\ge 1, \tag{3.1}$$

with the initial guess $\tilde{\theta}_0$ being an element in $\boldsymbol{\Theta}$, where ψ_n was defined in (2.1).

Given the definition of ψ_n, we see that $\tilde{\theta}_n$ is updated from $\tilde{\theta}_{n-1}$ based on new observation Z_n available at time n; of course, Z_{n-1} is used as well. We note that the recursion (3.1) is a version of the stochastic approximation method, which is meant to recursively approximate roots of the unknown equations, such as equation (2.10) (see e.g. [RM51], [KW52], [LS87], [KC78], [KY03]).

Remark 3.1. It is implicitly assumed in the recursion (3.1) that $\tilde{\theta}_n \in \boldsymbol{\Theta}$. One typical and easy way of making sure that this happens is to choose $\boldsymbol{\Theta}$ as the "largest possible set" that θ^* is an element of. So typically, one takes $\boldsymbol{\Theta} = \mathbb{R}^d$. However, this is not always possible, in which case one needs to implement a version of constrained stochastic approximation method (cf. e.g. [KC78] or [BK02]). We are not considering constrained stochastic approximation in this paper. This is planned for a future work.

As mentioned above, we are interested in the study of asymptotic properties of confidence regions that we will construct recursively in Section 5. These asymptotic properties crucially depend on the asymptotic properties of our recursive (point) estimators. One of such required properties is asymptotic normality. In this regard we stress that although the theory of asymptotic normality for stochastic approximation estimators is quite a mature field (see e.g. [Sac58], [Fab68], [LR79]), the existing results do not apply to $\tilde{\theta}$ as they require $\psi_n(\tilde{\theta}_{n-1}) - \mathbb{E}_{\theta^*}[\psi_n(\tilde{\theta}_{n-1})]$ to be a martingale difference, the property, which is not satisfied in our setup. Thus, we need to modify the base estimator $\tilde{\theta}$ to the effect of producing a recursive estimator that is asymptotically normal. In the next section we will construct such estimator, denoted there as $\hat{\theta}$, and we will study its asymptotic properties in the spirit of the method proposed by Fisher [Fis25]. Motivated by finding estimators that share the same asymptotic property as maximum likelihood estimators (MLEs), Fisher proposed in [Fis25] that if an estimator is $\sqrt{n}$-consistent (see below), then appropriate modification of the estimator has the same asymptotic normality as the MLE. This subject was further studied by LeCam in [LeC56] and [LeC60], where a more general class of observation than i.i.d. observations are considered.

Accordingly, we will show that $\tilde{\theta}$ is strongly consistent, and, moreover it maintains $\sqrt{n}$ convergence rate, i.e.

$$\mathbb{E}_{\theta^*}\|\tilde{\theta}_n - \theta^*\|^2 = O(n^{-1}). \tag{3.2}$$

An estimator that satisfies this equality is said to be $\sqrt{n}$-*consistent*.

We begin with the following proposition, which shows that the estimator $\tilde{\theta}$ is strongly consistent. For convenience, throughout, we will use the notation $\Delta_n := \tilde{\theta}_n - \theta^*$, $n \geq 1$.

Proposition 3.2. *Assume that* (2.3)*, and* (2.4) *are satisfied, then*

$$\lim_{n\to\infty} \tilde{\theta}_n = \theta^*, \quad \mathbb{P}_{\theta^*} - a.s.$$

Proof. Let us fix $n \geq 1$. Clearly, $\Delta_n = \Delta_{n-1} + \frac{\beta}{n}\psi_n(\theta^* + \Delta_{n-1})$, so that

$$\|\Delta_n\|^2 = \|\Delta_{n-1}\|^2 + \frac{2\beta}{n}\Delta_{n-1}^T\psi_n(\theta^* + \Delta_{n-1}) + \frac{\beta^2}{n^2}\|\psi_n(\theta^* + \Delta_{n-1})\|^2.$$

Taking conditional expectation on both sides leads to

$$\begin{aligned}
\mathbb{E}_{\theta^*}[\|\Delta_n\|^2|\mathscr{F}_{n-1}] =& \|\Delta_{n-1}\|^2 + \frac{2\beta}{n}\Delta_{n-1}^T b_n(\theta^* + \Delta_{n-1}) + \frac{\beta^2}{n^2}\mathbb{E}_{\theta^*}[\|\psi_n(\theta^* + \Delta_{n-1})\|^2|\mathscr{F}_{n-1}] \\
\leq& \|\Delta_{n-1}\|^2 + \frac{2\beta}{n}\Delta_{n-1}^T b_n(\theta^* + \Delta_{n-1}) + \frac{c\beta^2}{n^2}(1 + \|\Delta_{n-1}\|^2) \qquad (3.3) \\
\leq& \|\Delta_{n-1}\|^2 + \frac{c\beta^2}{n^2}(1 + \|\Delta_{n-1}\|^2), \qquad (3.4)
\end{aligned}$$

where the first inequality comes from (2.3) and the second is implied by (2.4). Let

$$Y_m := \|\Delta_m\|^2 \prod_{k=m+1}^{\infty}(1+\frac{c\beta^2}{k^2}) + \sum_{k=m+1}^{\infty}\frac{c\beta^2}{k^2}\prod_{j=k+1}^{\infty}(1+\frac{c\beta^2}{j^2}), \quad m \geq 0.$$

Then, (3.4) yields that

$$\mathbb{E}_{\theta^*}[Y_{m+1}|\mathscr{F}_m] \leq Y_m, \quad m \geq 0,$$

and therefore process Y is a supermartingale. Noting that Y is a positive process, and invoking the supermartingale convergence theorem, we conclude that hence the sequence $\{Y_m, m \geq 0\}$ converges $\mathbb{P}_{\theta^*}$ almost surely. This implies that the sequence $\{\|\Delta_m\|, m \geq 0\}$ converges, and we will show now that its limit is zero. According to (3.3), we have

$$\begin{aligned}\mathbb{E}_{\theta^*}\|\Delta_m\|^2 \leq & \mathbb{E}_{\theta^*}\|\Delta_{m-1}\|^2 + \frac{2\beta}{m}\mathbb{E}_{\theta^*}\left[\Delta_{m-1}^T b_m(\theta^* + \Delta_{m-1})\right] + \frac{c\beta^2}{m^2}\mathbb{E}_{\theta^*}\left[1 + \|\Delta_{m-1}\|^2\right] \\ \leq & \mathbb{E}_{\theta^*}\|\Delta_1\|^2 + \sum_{k=1}^{m}\frac{2\beta}{k}\mathbb{E}_{\theta^*}\left[\Delta_{k-1}^T b_k(\theta^* + \Delta_{k-1})\right] + \sum_{k=1}^{m}\frac{c\beta^2}{k^2}\mathbb{E}_{\theta^*}\left[1 + \|\Delta_{k-1}\|^2\right].\end{aligned}$$

Hence, we get

$$\sum_{k=1}^{m}\frac{2\beta}{k}\mathbb{E}_{\theta^*}\left|\Delta_{k-1}^T b_k(\theta^* + \Delta_{k-1})\right| \leq \mathbb{E}_{\theta^*}\|\Delta_1\|^2 - \mathbb{E}_{\theta^*}\|\Delta_m\|^2 + \sum_{k=1}^{m}\frac{c\beta^2}{k^2}\mathbb{E}_{\theta^*}\left[1 + \|\Delta_{k-1}\|^2\right].$$

Since

$$\lim_{m\to\infty}\|\Delta_m\|^2 = \lim_{m\to\infty} Y_m < \infty,$$

and

$$\lim_{m\to\infty}\sum_{k=1}^{m}\frac{c\beta^2}{k^2}\mathbb{E}_{\theta^*}\left[1 + \|\Delta_{k-1}\|^2\right] \leq \sum_{k=1}^{\infty}\frac{c\beta\mathbb{E}_{\theta^*}[1+Y_1]}{k^2} < \infty,$$

then, the series

$$\sum_{k=1}^{m}\frac{1}{k}\mathbb{E}_{\theta^*}\left|\Delta_{k-1}^T b_k(\theta^* + \Delta_{k-1})\right|, \quad m \geq 1,$$

converges $\mathbb{P}_{\theta^*}$ almost surely, and thus

$$\lim_{k\to\infty}\mathbb{E}_{\theta^*}\left|\Delta_{k-1}^T b_k(\theta^* + \Delta_{k-1})\right| = 0.$$

This implies that there exists a subsequence $\Delta_{m_k-1}^T b_{m_k}(\theta^* + \Delta_{m_k-1})$ which converges $\mathbb{P}_{\theta^*}$ almost surely to zero, as $k \to \infty$. According to (2.4), we also have that

$$\|\Delta_{m_k-1}\|^2 \leq \frac{1}{K_1}\|\Delta_{m_k-1}^T b_{m_k}(\theta^* + \Delta_{m_k-1})\|.$$

Therefore, $\lim_{k\to\infty}\Delta_{m_k-1} = 0$, $\mathbb{P}_{\theta^*}$ almost surely, and this concludes the proof. □

Proposition 3.3. *Assume that* (2.3), (2.4), (2.5) *and* (2.8) *hold. Then,*

$$\mathbb{E}_{\theta^*}\|\tilde{\theta}_n - \theta^*\|^2 = O(n^{-1}).$$

Proof. Putting $V_n(\tilde\theta_{n-1}) := \psi_n(\tilde\theta_{n-1}) - b_n(\tilde\theta_{n-1})$, from (3.1) we immediately have that

$$\Delta_n = \Delta_{n-1} + \frac{\beta}{n} b_n(\tilde\theta_{n-1}) + \frac{\beta}{n} V_n(\tilde\theta_{n-1}),$$

that yields

$$\mathbb{E}_{\theta^*}\|\Delta_n\|^2 = \mathbb{E}_{\theta^*}\|\Delta_{n-1} + \frac{\beta}{n} b_n(\tilde\theta_{n-1})\|^2 + \frac{\beta^2}{n^2}\mathbb{E}_{\theta^*}\|V_n(\tilde\theta_{n-1})\|^2.$$

From here, applying consequently (2.8), (2.5), (2.4), and note that $b_n(\theta^*) = 0$, we get

$$\begin{aligned}
\mathbb{E}_{\theta^*}\|\Delta_n\|^2 &= \mathbb{E}_{\theta^*}\left\|\Delta_{n-1} + \frac{\beta}{n} b_n(\tilde\theta_{n-1})\right\|^2 + O(n^{-2}) \\
&\le \mathbb{E}_{\theta^*}\left[\|\Delta_{n-1}\|^2 + \frac{\beta^2 K_2^2}{n^2}\|\Delta_{n-1}\|^2 + \frac{2\beta}{n}\Delta_{n-1}^T b_n(\tilde\theta_{n-1})\right] + O(n^{-2}) \\
&\le \left(1 + \frac{\beta^2 K_2^2}{n^2} - \frac{2\beta K_1}{n}\right)\mathbb{E}_{\theta^*}\|\Delta_{n-1}\|^2 + D_1 n^{-2}.
\end{aligned}$$

Clearly, for any $\varepsilon > 0$, and for large enough n, we get

$$\mathbb{E}_{\theta^*}\|\Delta_n\|^2 \le (1 - (2K_1\beta - \varepsilon)n^{-1})\mathbb{E}_{\theta^*}\|\Delta_{n-1}\|^2 + D_1 n^{-2}. \tag{3.5}$$

For ease of writing, we put $p := 2K_1\beta - \varepsilon$ and $c_n := \mathbb{E}_{\theta^*}\|\Delta_n\|^2$. Take ϵ sufficiently small, so that $p > 1$, and then chose an integer $N > p$. Then, for $n > N$ we have by (3.5) that

$$\begin{aligned}
c_n &\le c_N \prod_{j=N+1}^{n}(1 - \frac{p}{j}) + D_1 \sum_{j=N+1}^{n}\frac{1}{j^2}\prod_{k=j+1}^{n}(1 - \frac{p}{k}) \\
&\le c_N \prod_{j=N+1}^{n}(1 - \frac{p}{j}) + D_1 \sum_{j=N+1}^{n}\frac{1}{j^2}.
\end{aligned}$$

Using the fact that $\sum_{j=m}^n 1/j^2 \sim 1/n$ and $\prod_{j=m}^n (1 - p/j) \sim 1/n^p$, for any fixed $m, p \ge 1$, we immediately get that $c_n \le O(n^{-1})$. This concludes the proof. □

4 Quasi-asymptotically linear estimator

In this section we define a new estimator denoted as $\{\hat\theta_n, n \ge 1\}$ and given recursively by

$$\begin{aligned}
\hat\theta_n &= -I^{-1}(\tilde\theta_n) I_n \tilde\theta_n + I^{-1}(\tilde\theta_n)\Gamma_n, \\
\Gamma_n &= \frac{n-1}{n}\Gamma_{n-1} + \frac{1}{n}(\mathrm{Id} + \beta I_n)\psi_n(\tilde\theta_{n-1}), \\
I_n &= \frac{n-1}{n} I_{n-1} + \frac{1}{n}\Psi_n(\tilde\theta_{n-1}), \quad n \ge 1, \\
\Gamma_0 &= 0, \quad I_0 = 0,
\end{aligned} \tag{4.1}$$

where Id is the unit matrix. Since $\tilde\theta_n$, I_n, and Γ_n are updated from time $n-1$ based on the new observation Z_n available at time n, then the estimator $\hat\theta$ indeed is recursive. This estimator will be used in Section 6 for recursive construction of confidence regions for θ^*.

Remark 4.1. (i) In Section 7 we propose an alternative algorithm to (4.1) that uses I_n^+, the Moore–Penrose pseudoinverse matrix of I_n, instead of $I^{-1}(\tilde{\theta}_n)$. However, the convergence of this alternative algorithm remains an open problem.
(ii) In the argument below we will use the following representations of Γ_n and I_n,

$$\Gamma_n = \sum_{j=1}^{n} (\mathrm{Id} + \beta I_j)\, \psi_j(\tilde{\theta}_{j-1}), \quad I_n = \frac{1}{n} \sum_{i=1}^{n} \Psi_i(\tilde{\theta}_{i-1}).$$

Next, we will show that $\hat{\theta}$ is weakly consistent and asymptotically normal. We will derive asymptotic normality of $\hat{\theta}$ from the property of quasi-asymptotic linearity, which is related to the property of asymptotic linearity (cf. [Shi84]), and which is defined as follows:

Definition 4.2. An estimator $\{\bar{\theta}_n, n \geq 1\}$ of θ^* is called a *quasi-asymptotically linear estimator* if there exist a $\mathbb{P}_{\theta^*}$-convergent, adapted matrix valued process G, and adapted vector valued processes ϑ and ε, such that

$$\bar{\theta}_n - \vartheta_n = \frac{G_n}{n} \sum_{i=1}^{n} \psi_i(\theta^*) + \varepsilon_n, \ n \geq 1, \quad \vartheta_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} \theta^*, \quad \sqrt{n}\varepsilon_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} 0.$$

Our definition of quasi-asymptotically linear estimator is motivated by the classic concept of asymptotically linear estimator (see e.g. [Sha10]): $\check{\theta}$ is called (locally) asymptotically linear if there exists a matrix process $\{\check{G}_n, n \geq 1\}$ such that

$$\check{\theta}_n - \theta^* = \check{G}_n \sum_{i=1}^{n} \psi_i(\theta^*) + \varepsilon_n,$$

where $\check{G}_n^{-1/2} \varepsilon_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} 0$. Asymptotic linearity is frequently used in the proof of asymptotic normality of estimators. However, in general, asymptotic linearity can not be reconciled with the full recursiveness of the estimator. The latter property is the key property involved in construction of recursive confidence regions. Moreover, the property of asymptotic linearity requires that the matrices $\check{G}_n$ are invertible, which is a very stringent requirement, not easily fulfilled. These are the reasons why we propose the concept of quasi-asymptotic linearity since, it can be reconciled with recursiveness and does not require that matrices G_n are invertible. As it will be shown below, the fully recursive estimator $\hat{\theta}$ is quasi-asymptotically linear.

In what follows, we will make use of the following representation for $\hat{\theta}$

$$\hat{\theta}_n = -I^{-1}(\tilde{\theta}_n) I_n \tilde{\theta}_n + \frac{1}{n} I^{-1}(\tilde{\theta}_n) \sum_{j=1}^{n} (\mathrm{Id} + \beta I_j)\, \psi_j(\tilde{\theta}_{j-1}). \tag{4.2}$$

Theorem 4.3. *Assume that R1–R8 hold, then the estimator $\hat{\theta}$ is $\mathbb{P}_{\theta^*}$–weakly consistent.*[6] *Moreover, $\hat{\theta}$ is quasi-asymptotically linear estimator for θ^*.*

[6]That is, $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} \theta^*$.

Proof. First, we show the generalized asymptotic linearity of $\hat{\theta}$. Due to Taylor's expansion, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta^*) - \frac{1}{n}\sum_{i=1}^{n}\psi_i(\tilde{\theta}_{i-1}) = -\frac{1}{n}\sum_{i=1}^{n}\Psi_i(\tilde{\theta}_{i-1})\Delta_{i-1} + \frac{1}{n}\sum_{i=1}^{n}\Delta_{i-1}^T \mathsf{H}\psi_i(\eta_{i-1})\Delta_{i-1}$$
$$=: A_n + B_n, \tag{4.3}$$

where $\eta_{i-1}, 1 \leq i \leq n$, is in a neighborhood of θ^* such that $\|\eta_{i-1} - \theta^*\| \leq \|\tilde{\theta}_{i-1} - \theta^*\|$. Note that

$$\begin{aligned} A_n =& -\frac{1}{n}\sum_{i=1}^{n}\Psi_i(\tilde{\theta}_{i-1})\Big(\Delta_n - \sum_{j=i}^{n}\frac{\beta}{j}\psi_j(\tilde{\theta}_{j-1})\Big) \\ =& -I_n\Delta_n + \frac{\beta}{n}\sum_{i=1}^{n} I_i\psi_i(\tilde{\theta}_{i-1}), \end{aligned}$$

and by (4.3), we get

$$I_n\Delta_n = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Id} + \beta I_i)\psi_i(\tilde{\theta}_{i-1}) - \frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta^*) + B_n.$$

Therefore, using the representation (4.2), we immediately have

$$\hat{\theta}_n + I^{-1}(\tilde{\theta}_n)I_n\theta^* = \frac{I^{-1}(\tilde{\theta}_n)}{n}\sum_{i=1}^{n}\psi_i(\theta^*) - I^{-1}(\tilde{\theta}_n)B_n. \tag{4.4}$$

Next we will show that

$$\mathbb{P}_{\theta^*}\text{-}\lim_{n\to\infty} I_n = -I(\theta^*). \tag{4.5}$$

First, by (2.6), we deduce that

$$\mathbb{E}_{\theta^*}\Big[\frac{1}{n}\sum_{i=1}^{n}\|\Psi_i(\tilde{\theta}_{i-1}) - \Psi_i(\theta^*)\|\Big] \leq \frac{K_3}{n}\sum_{i=1}^{n}\mathbb{E}_{\theta^*}\|\Delta_{i-1}\|.$$

Due to Proposition 3.3, $\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_{\theta^*}\|\Delta_{i-1}\| \leq \frac{1}{n}\sum_{j=1}^{n} j^{-1/2} = O(n^{-1/2})$. Hence,

$$\frac{1}{n}\sum_{i=1}^{n}\|\Psi_i(\tilde{\theta}_{i-1}) - \Psi_i(\theta^*)\| \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} 0. \tag{4.6}$$

Therefore,

$$\mathbb{P}_{\theta^*}\text{-}\lim_{n\to\infty} I_n = \mathbb{P}_{\theta^*} - \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Psi_i(\tilde{\theta}_{i-1}) = \mathbb{P}_{\theta^*} - \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Psi_i(\theta^*). \tag{4.7}$$

Next, observe that in view of Proposition A.7 we get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Psi_i(\theta^*)=\mathbb{E}_{\theta^*}[\Psi_1(\theta^*)]=\mathbb{E}_{\theta^*}[\mathsf{H}\pi_1(\theta^*)]=\mathbb{E}_{\theta^*}[\mathsf{H}\log p_{\theta^*}(Z_0,Z_1)].$$

Invoking the usual chain rule we obtain that

$$\mathsf{H}\log p_{\theta^*}(Z_0,Z_1)=\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}-\frac{\nabla p_{\theta^*}(Z_0,Z_1)\nabla p_{\theta^*}(Z_0,Z_1)^T}{p^2_{\theta^*}(Z_0,Z_1)}=\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}-\psi_1(\theta^*)\psi_1^T(\theta^*),$$

so that

$$\mathbb{E}_{\theta^*}[\mathsf{H}\log p_{\theta^*}(Z_0,Z_1)]=\mathbb{E}_{\theta^*}[\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}]-I(\theta^*).$$

We will now show that $\mathbb{E}_{\theta^*}\left[\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}\right]=0$. In fact, denote by f_{Z_0} the density function of Z_0 under $\mathbb{P}_{\theta^*}$ and in view of (2.2), we have

$$\begin{aligned}
\mathbb{E}_{\theta^*}\left[\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}\right]=&\mathbb{E}_{\theta^*}\left[\mathbb{E}_{\theta^*}\left[\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}\middle|Z_0\right]\right]\\
=&\int_{\mathbb{R}}\mathbb{E}_{\theta^*}\left[\frac{\mathsf{H}p_{\theta^*}(Z_0,Z_1)}{p_{\theta^*}(Z_0,Z_1)}\middle|Z_0=z_0\right]f_{Z_0}(z_0)dz_0\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{\mathsf{H}p_{\theta^*}(z_0,z_1)}{p_{\theta^*}(z_0,z_1)}p_{\theta^*}(z_0,z_1)dz_1f_{Z_0}(z_0)dz_0\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}}\mathsf{H}p_{\theta^*}(z_0,z_1)dz_1f_{Z_0}(z_0)dz_0\\
=&\int_{\mathbb{R}}\mathsf{H}\int_{\mathbb{R}}p_{\theta^*}(z_0,z_1)dz_1f_{Z_0}(z_0)dz_0\\
=&\int_{\mathbb{R}}(\mathsf{H}1)f_{Z_0}(z_0)dz_0=0.
\end{aligned}$$

Recalling (4.7) we conclude that (4.5) is satisfied.

By Assumption R8 and strong consistency of $\tilde{\theta}$ we obtain that

$$\lim_{n\to\infty}I^{-1}(\tilde{\theta}_n)=I^{-1}(\theta^*)\quad\mathbb{P}_{\theta^*}-a.s.,\tag{4.8}$$

which, combined with (4.5) implies that

$$-I^{-1}(\tilde{\theta}_n)I_n\theta^*\xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}}\theta^*.\tag{4.9}$$

Next, we will show that

$$\sqrt{n}B_n\xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}}0.\tag{4.10}$$

Indeed, by (2.7), $\sqrt{n}\mathbb{E}_{\theta^*}\|B_n\|\leq\frac{K_4}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}_{\theta^*}\|\Delta_{i-1}\|^2$, and consequently, in view of Proposition 3.3,

$$\lim_{n\to\infty}\sqrt{n}\mathbb{E}_{\theta^*}\|B_n\|\leq\lim_{n\to\infty}\frac{K_4}{\sqrt{n}}\log n=0,$$

which implies (4.10).

Now, taking $\vartheta_n = -I^{-1}(\tilde{\theta}_n)I_n\theta^*$, $G_n = I^{-1}(\tilde{\theta}_n)$ and $\varepsilon_n = I^{-1}(\tilde{\theta}_n)B_n$, we deduce quasi-asymptotic linearity of $\hat{\theta}$ from (4.4), (4.8), (4.9) and (4.10).

Finally, we will show the weak consistency of $\hat{\theta}$. By ergodicity of Z, in view of Proposition A.7, and using the fact that θ^* is a (unique) solution of (2.10), we have that

$$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\theta^*) = \mathbb{E}_{\theta^*}[\psi_1(\theta^*)] = 0, \quad \mathbb{P}_{\theta^*} - \text{a.s.}$$

Thus, $\lim_{n\to\infty}\frac{I^{-1}(\tilde{\theta}_n)}{n}\sum_{i=1}^{n}\psi_i(\theta^*) = 0$ $\mathbb{P}_{\theta^*}$ almost surely. This, combined with (4.4), (4.9) and (4.10) implies that $\hat{\theta}_n \xrightarrow{\mathbb{P}_{\theta^*}} \theta^*$, as $n\to\infty$. The proof is complete. □

The next result, which will be used in analysis of asymptotic properties of the recursive confidence region for θ^* in Section 6, is an application of Theorem 4.3.

Proposition 4.4. *Assume that R1–R9 are satisfied. Then, there exists an adapted process ϑ such that*

$$\vartheta_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} \theta^*, \tag{4.11}$$

and

$$\sqrt{n}(\hat{\theta}_n - \vartheta_n) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, I^{-1}(\theta^*)). \tag{4.12}$$

Proof. Let $\vartheta_n = -I^{-1}(\tilde{\theta}_n)I_n\theta^*$, $G_n = I^{-1}(\tilde{\theta}_n)$ and $I^{-1}(\tilde{\theta}_n)B_n = \varepsilon_n$. Then, property (4.11) follows from (4.9).

In order to prove (4.12), we note that according to Theorem 4.3 we have

$$\hat{\theta}_n - \vartheta_n = \frac{G_n}{n}\sum_{i=1}^{n}\psi_i(\theta^*) + \varepsilon_n, \quad \sqrt{n}\varepsilon_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta^*}} 0.$$

Next, Proposition A.14 implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i(\theta^*) \xrightarrow[n\to\infty]{d} N(0, I(\theta^*)).$$

Consequently, since by (4.8) $G_n \xrightarrow{\mathbb{P}_{\theta^*}} I^{-1}(\theta^*)$, using Slutsky's theorem we get

$$\frac{G_n}{\sqrt{n}}\sum_{i=1}^{n}\psi_i(\theta^*) \xrightarrow[n\to\infty]{d} N(0, I^{-1}(\theta^*)).$$

The proof is complete. □

We end this section with the following technical result, which will be used in our construction of confidence region in Section 6. Towards this end, for any $\theta \in \boldsymbol{\Theta}$ and

$n \geq 1$, we define[7]

$$U_n(\theta) := n(\hat{\theta}_n - \theta)^T I(\tilde{\theta}_n)(\hat{\theta}_n - \theta) \tag{4.13}$$
$$= n \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_n^{ij} (\hat{\theta}_n^i - \theta^i)(\hat{\theta}_n^j - \theta^i),$$

where $(\sigma_n^{ij})_{i,j=1,\dots,d} = I(\tilde{\theta}_n)$, and, as usual, we denote by χ_d^2 a random variable that has the chi-squared distribution with d degrees of freedom.

Corollary 4.5. *With $\vartheta_n = -I^{-1}(\tilde{\theta}_n) I_n \theta^*$, we have that*

$$U_n(\vartheta_n) \xrightarrow[n\to\infty]{d} \chi_d^2.$$

Proof. From Assumption R8, strong consistency of $\tilde{\theta}$ and Proposition 4.4, and employing the Slutsky's theorem again, we get that

$$\sqrt{nI(\tilde{\theta}_n)}(\hat{\theta}_n - \vartheta_n) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \mathrm{Id}).$$

Therefore,

$$U_n(\vartheta_n) = n(\hat{\theta}_n - \vartheta_n)^T I(\tilde{\theta}_n)(\hat{\theta}_n - \vartheta_n) \xrightarrow{d} \xi^T \xi,$$

where $\xi \sim \mathcal{N}(0, \mathrm{Id})$. The proof is thus complete since $\xi^T \xi \stackrel{d}{=} \chi_d^2$.

□

5 Recursive construction of confidence regions

This section is devoted to the construction of the recursive confidence region based on quasi-asymptotically linear estimator $\hat{\theta}$ developed in Section 4. We start with introducing the definition of the approximated confidence region.

Definition 5.1. Let $V_n : \mathbb{R}^{n+1} \to 2^{\boldsymbol{\Theta}}$ be a set valued function such that $V_n(z)$ is a connected set[8] for any $z \in \mathbb{R}^{n+1}$. The set $V_n(Z_0^n)$, with $Z_0^n := (Z_0, \dots, Z_n)$, is called an *approximated confidence region* for θ^*, at significant level $\alpha \in (0,1)$, if there exists a weakly consistent estimator ϑ of θ^*, such that

$$\lim_{n\to\infty} \mathbb{P}_{\theta^*}(\vartheta_n \in V_n(Z_0^n)) = 1 - \alpha.$$

Such approximated confidence region can be constructed, as next result shows, by using the asymptotic results obtained in Section 4. Recall the notation $U_n(\theta) = n(\hat{\theta}_n - \theta)^T I(\tilde{\theta}_n)(\hat{\theta}_n - \theta)$, for $\theta \in \boldsymbol{\Theta}$, $n \geq 1$.

[7] We use superscripts here to denote components of vectors and matrices.

[8] A connected set is a set that cannot be represented as the union of two or more disjoint nonempty open subsets.

Proposition 5.2. *Fix a confidence level α, and let $\kappa \in \mathbb{R}$ be such that $\mathbb{P}_{\theta^*}(\chi_d^2 < \kappa) = 1-\alpha$. Then, the set*

$$\mathcal{T}_n := \{\theta \in \mathbf{\Theta} : U_n(\theta) < \kappa\}$$

is an approximated confidence region for θ^*.

Proof. As in Section 4, we take $\vartheta_n = -I^{-1}(\hat{\theta}_n)I_n\theta^*$, which in view of Proposition 4.4 is a weakly consistent estimator of θ^*. Note that $U_n(\,\cdot\,)$ is a continuous function, and thus $\mathcal{T}_n$ is a connected set, for any $n \geq 1$. By Corollary 4.5, $U_n(\vartheta_n) \xrightarrow{d} \chi_d^2$, and since $\mathbb{P}_{\theta^*}(\vartheta_n \in \mathcal{T}_n) = \mathbb{P}_{\theta^*}(U_n(\vartheta_n) < \kappa)$, we immediately have that $\lim_{n\to\infty} \mathbb{P}_{\theta^*}(\vartheta_n \in \mathcal{T}_n) = 1-\alpha$. This concludes the proof. □

Next, we will show that the approximated confidence region $\mathcal{T}_n$ can be computed in a recursive way, by taking into account its geometric structure. By the definition, the set $\mathcal{T}_n$ is the interior of a d-dimensional ellipsoid, and hence cT_n is uniquely determined by its extreme $2d$ points. Thus, it is enough to establish a recursive formula for computing the extreme points. Let us denote by

$$(\theta_{n,k}^1, \ldots, \theta_{n,k}^d), \quad k = 1, \ldots, 2d,$$

the coordinates of these extreme points; that is $\theta_{n,k}^i$, denotes the ith coordinate of the kth extreme point of ellipsoid $\mathcal{T}_n$.

First, note that the matrix $I(\tilde{\theta}_n)$ is positive definite, and hence it admits the Cholesky decomposition:

$$I(\tilde{\theta}_n) = L_n L_n^T = \begin{bmatrix} l_n^{11} & 0 & \cdots & 0 \\ l_n^{21} & l_n^{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ l_n^{d1} & l_n^{d2} & \cdots & l_n^{dd} \end{bmatrix} \begin{bmatrix} l_n^{11} & l_n^{21} & \cdots & l_n^{d1} \\ 0 & l_n^{22} & \cdots & l_n^{d2} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & l_n^{dd} \end{bmatrix},$$

where l_n^{ij} $i,j = 1, \ldots, d$, are given by

$$l_n^{ii} = \sqrt{\sigma_n^{ii} - \sum_{k=1}^{i-1} (l_n^{ik})^2},$$

$$l_n^{ij} = \frac{1}{l_n^{ii}}\Big(\sigma_n^{ij} - \sum_{k=1}^{j-1} l_n^{ik} l_n^{jk}\Big).$$

Thus, we have that $U_n(\theta) = n(u_{n,1}^2(\theta) + u_{n,2}^2(\theta) + \cdots + u_{n,d}^2(\theta))$, where

$$u_{n,i}(\theta) = \sum_{j=i}^{d} l_n^{ji}(\hat{\theta}_n^j - \theta^j), \quad i = 1, \ldots, d,$$

and thus $\mathcal{T}_n = \{\theta : \sum_{j=1}^d (u_{n,j}(\theta))^2 < \frac{\kappa}{n}\}$.

By making the coordinate transformation $\theta \mapsto \rho$ given by $\rho = L_n^T(\hat{\theta}_n - \theta)$, the set $\mathcal{T}_n$ in the new system of coordinates can be written as $\mathcal{T}_n = \{\rho : \sum_{i=1}^d (\rho^i)^2 < \frac{\kappa}{n}\}$. Hence, $\mathcal{T}_n$, in the new system of coordinates, is determined by the following $2d$ extreme points of the ellipsoid:

$$(\rho_1^1, \ldots, \rho_1^d) = (\sqrt{\frac{\kappa}{n}}, 0, \ldots, 0),$$
$$(\rho_2^1, \ldots, \rho_2^d) = (-\sqrt{\frac{\kappa}{n}}, 0, \ldots, 0),$$
$$\ldots$$
$$(\rho_{2d-1}^1, \ldots, \rho_{2d-1}^d) = (0, \ldots, 0, \sqrt{\frac{\kappa}{n}}),$$
$$(\rho_{2d}^1, \ldots, \rho_{2d}^d) = (0, \ldots, 0, -\sqrt{\frac{\kappa}{n}}).$$

Then, in the original system of coordinates, the extreme points (written as vectors) are given by

$$\begin{aligned}(\theta_{n,2j-1}^1, \ldots, \theta_{n,2j-1}^d)^T &= \hat{\theta}_n - \sqrt{\frac{\kappa}{n}}(L_n^T)^{-1}e_j,\\ (\theta_{n,2j}^1, \ldots, \theta_{n,2j}^d)^T &= \hat{\theta}_n + \sqrt{\frac{\kappa}{n}}(L_n^T)^{-1}e_j,\end{aligned} \qquad j = 1, \ldots, d, \tag{5.1}$$

where $\{e_j\}$, $j = 1, \ldots, d$, is the standard basis in $\mathbb{R}^d$.

Finally, taking into account the recursive constructions (3.1), (4.1), and the representation (5.1), we have the following recursive scheme for computing the approximate confidence region.

Recursive construction of the confidence region

Initial Step: $\Gamma_0 = 0, \quad I_0 = 0, \quad \tilde{\theta}_0 \in \mathbf{\Theta}.$

n^{th} Step:

Input: $\tilde{\theta}_{n-1}, I_{n-1}, \Gamma_{n-1}, Z_{n-1}, Z_n.$

Output:
$$\tilde{\theta}_n = \tilde{\theta}_{n-1} + \frac{\beta}{n}\psi_n(\tilde{\theta}_{n-1}),$$
$$I_n = \frac{n-1}{n}I_{n-1} + \frac{1}{n}\Psi_n(\tilde{\theta}_{n-1}),$$
$$\Gamma_n = \frac{n-1}{n}\Gamma_{n-1} + \frac{1}{n}(\mathrm{Id} + \beta I_n)\psi_n(\tilde{\theta}_{n-1}),$$
$$(\theta_{n,2j}^1, \ldots, \theta_{n,2j}^d)^T = -I^{-1}(\tilde{\theta}_n)I_n\tilde{\theta}_n + I^{-1}(\tilde{\theta}_n)\Gamma_n + \sqrt{\frac{\kappa}{n}}(I_n^{-1/2})^T e_j,$$
$$(\theta_{n,2j-1}^1, \ldots, \theta_{n,2j-1}^d)^T = -I^{-1}(\tilde{\theta}_n)I_n\tilde{\theta}_n + I^{-1}(\tilde{\theta}_n)\Gamma_n - \sqrt{\frac{\kappa}{n}}(I_n^{-1/2})^T e_j,$$
$$j = 1, \ldots, d.$$

From here, we also conclude that there exists a function τ, independent of n, such that

$$\mathcal{T}_n = \tau(\mathcal{T}_{n-1}, Z_n). \tag{5.2}$$

The above recursive relationship goes to heart of application of recursive confidence regions in the robust adaptive control theory originated in [BCC+16], since it makes it possible to take the full advantage of the dynamic programming principle in the context of such control problems.

We conclude this section by proving that the confidence region converges to the singleton θ^*. Equivalently, it is enough to prove that the extreme points converge to the true parameter θ^*.

Proposition 5.3. *For any* $k \in \{1, \dots, 2d\}$*, we have that*

$$\mathbb{P}_{\theta^*}\text{-}\lim_{n\to\infty} \theta_{n,k} = \theta^*.$$

Proof. By Assumption R8 and Theorem 3.2 (strong consistency of $\tilde{\theta}$), we have that $L_n \xrightarrow[n\to\infty]{\text{a.s.}} I^{1/2}(\theta^*)$, and consequently, we also have that

$$\sqrt{\frac{\kappa}{n}} e_j^T L_n^{-1} \xrightarrow[n\to\infty]{\text{a.s}} 0. \tag{5.3}$$

Of course, the last convergence holds true in the weak sense too. Passing to the limit in (5.1), in $\mathbb{P}_{\theta^*}$ probability sense, and using (5.3) and weak consistency of $\hat{\theta}$ (Theorem 4.3), we finish the proof.

□

6 Examples

In this section we will present three illustrative examples of the recursive construction of confidence regions developed above. We start with our main example, Example 6.1, of a Markov chain with Gaussian transitional densities where both the conditional mean and conditional standard deviation are the parameters of interest. Example 6.2 is dedicated to the case of i.i.d. Gaussian observations, which is a particular case of the first example.

Generally speaking, the simple case of i.i.d. observations for which the MLE exists and asymptotic normality holds true, one can recursively represent the sequence of confidence intervals constructed in the usual (off-line) way, and the theory developed in this paper is not really needed. The idea is illustrated in Example 6.3 by considering again the same experiment as in Example 6.2. In fact, as mentioned above, this idea served as the starting point for the general methodology presented in the paper.

Example 6.1. Let us consider a Markov process $\{Z_n\}$ with a Gaussian transition density function

$$p_\theta(x, y) = \frac{1}{\sqrt{1-\rho^2}\sqrt{2\pi}\sigma} e^{-\frac{(y-\rho x-(1-\rho)\mu)^2}{2\sigma^2(1-\rho^2)}}, \quad n \geq 1,$$

and such that $Z_0 \sim \mathcal{N}(\mu, \sigma^2)$.

We assume that the correlation parameter $\rho \in (-1, 1)$ is known, and the unknown parameter is $\theta = (\mu, \sigma) \in \Theta$, where $\Theta = [a_1, a_2] \times [b_1, b_2]$, and $a_1 \leq a_2,\ b_1 \leq b_2$ are some fixed real numbers with $b_1 > 0$.

In the Appendix A.3 we show that the process Z satisfies the Assumption M, and the conditions R0-A9.

Thus, all the results derived in the previous sections hold true. Moreover, for a given confidence level α, we have the following explicit formulas for the nth step of the recurrent construction of the confidence regions:

$$\begin{aligned}
\tilde{\mu}_n &= \tilde{\mu}_{n-1} + \frac{\beta(Z_n - \rho Z_{n-1} - (1-\rho)\tilde{\mu}_{n-1})}{n\tilde{\sigma}^2_{n-1}(1+\rho)}, \\
\tilde{\sigma}^2_n &= \tilde{\sigma}^2_{n-1} - \frac{\beta}{n\tilde{\sigma}_{n-1}} + \frac{\beta(Z_n - \rho Z_{n-1} - (1-\rho)\tilde{\mu}_{n-1})^2}{n(1-\rho^2)\tilde{\sigma}^3_{n-1}}), \\
I_n &= \frac{n-1}{n} I_{n-1} + \frac{1}{n}\begin{bmatrix} -\frac{1-\rho}{(1+\rho)\tilde{\sigma}^2_{n-1}} & -\frac{2(Z_n-\rho Z_{n-1}-(1-\rho)\tilde{\mu}_{n-1})}{(1+\rho)\tilde{\sigma}^3_{n-1}} \\ -\frac{2(Z_n-\rho Z_{n-1}-(1-\rho)\tilde{\mu}_{n-1})}{(1+\rho)\tilde{\sigma}^3_{n-1}} & \frac{1}{\tilde{\sigma}^2_{n-1}} - \frac{3(Z_n-\rho Z_{n-1}-(1-\rho)\tilde{\mu}_{n-1})^2}{(1-\rho^2)\tilde{\sigma}^4_{n-1}} \end{bmatrix}, \\
\Gamma_n &= \frac{n-1}{n}\Gamma_{n-1} + \frac{1}{n}(\mathrm{Id} + \beta I_n)\begin{bmatrix} \tilde{\mu}_{n-1} \\ \tilde{\sigma}^2_{n-1} \end{bmatrix},
\end{aligned}$$

and, for $j \in \{1,2,3,4\}$,

$$\begin{bmatrix} \mu_{n,j} \\ \sigma^2_{n,j} \end{bmatrix} = -\begin{bmatrix} \frac{(1+\rho)\tilde{\sigma}^2_n}{1-\rho} & 0 \\ 0 & \frac{\tilde{\sigma}^2_n}{2} \end{bmatrix} I_n \begin{bmatrix} \tilde{\mu}_n \\ \tilde{\sigma}^2_n \end{bmatrix} + \begin{bmatrix} \frac{(1+\rho)\tilde{\sigma}^2_n}{1-\rho} & 0 \\ 0 & \frac{\tilde{\sigma}^2_n}{2} \end{bmatrix} \Gamma_n + \varpi_j \frac{\kappa}{n}\begin{bmatrix} \sqrt{\frac{1+\rho}{1-\rho}}\tilde{\sigma}_n & 0 \\ 0 & \frac{\tilde{\sigma}_n}{\sqrt{2}} \end{bmatrix} u_j,$$

where $\varpi_1 = \varpi_3 = -1$, $\varpi_2 = \varpi_4 = 1$, $u_1 = u_2 = e_1$, $u_3 = u_4 = e_2$, β is a constant such that $\beta > \frac{b_2^3}{4b_1}$, $\beta > \frac{(1+\rho)b_2^3}{2(1-\rho)b_1}$, and $\mathbb{P}_{\theta^*}(\chi^2_2 < \kappa) = 1-\alpha$.

Example 6.2. Let Z_n, $n \geq 0$, be a sequence of i.i.d. Gaussian random variables with an unknown mean μ and unknown standard deviation σ. Clearly, this important case is a particular case of Example 6.1, with $\rho = 0$, and the same recursive formulas for confidence regions by taking $\rho = 0$ in the above formulas.

Example 6.3. We take the same setup as in the previous example - i.i.d Gaussian random variables with unknown mean and standard deviation. We will use the fact that in this case, the MLE estimators for μ and σ^2 are computed explicitly and given by

$$\hat{\mu}_n = \frac{1}{n+1}\sum_{i=0}^n Z_i, \quad \hat{\sigma}^2_n = \frac{1}{n+1}\sum_{i=0}^n (Z_i - \hat{\mu}_n)^2, \quad n \geq 1,$$

It is well known that $(\hat{\mu}, \hat{\sigma}^2)$ are asymptotically normal, namely

$$\sqrt{n}(\hat{\mu}_n - \mu^*, \hat{\sigma}^2_n - (\sigma^*)^2) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, I^{-1}),$$

where

$$I = \begin{bmatrix} (\sigma^*)^2 & 0 \\ 0 & 2(\sigma^*)^4 \end{bmatrix}.$$

First, note that $(\hat{\mu}_n, \hat{\sigma}^2_n)$ satisfies the following recursion:

$$\begin{aligned}
\hat{\mu}_n &= \frac{n}{n+1}\hat{\mu}_{n-1} + \frac{1}{n+1}Z_n, \\
\hat{\sigma}^2_n &= \frac{n}{n+1}\hat{\sigma}^2_{n-1} + \frac{n}{(n+1)^2}(\hat{\mu}_n - Z_n)^2, \quad n \geq 1.
\end{aligned} \tag{6.1}$$

Second, due to asymptotic normality, we also have that, $U_n \xrightarrow[n\to\infty]{d} \chi^2_2$, where $U_n := \frac{n}{\hat{\sigma}^2_n}(\hat{\mu}_n - \mu^*)^2 + \frac{n}{2\hat{\sigma}^4_n}(\hat{\sigma}^2_n - (\sigma^*)^2)^2$. Now, for a given confidence level α, we let $\kappa \in \mathbb{R}$ be such that $\mathbb{P}_{\theta^*}(\chi^2_2 < \kappa) = 1 - \alpha$, and then, the confidence region for (μ, σ^2) is given by

$$\mathcal{T}_n := \left\{ (\mu, \sigma^2) \in \mathbb{R}^2 : \frac{n}{\hat{\sigma}^2_n}(\hat{\mu}_n - \mu)^2 + \frac{n}{2\hat{\sigma}^4_n}(\hat{\sigma}^2_n - \sigma^2)^2 < \kappa \right\}.$$

Similar to the previous cases, we note that $\mathcal{T}_n$ is the interior of an ellipse (in $\mathbb{R}^2$), that is uniquely determined by its extreme points

$$\begin{aligned}
(\mu_{n,1}, \sigma^2_{n,1}) &= \left(\hat{\mu}_n + \sqrt{\frac{\kappa}{n}}\hat{\sigma}_n, \hat{\sigma}^2_n\right), & (\mu_{n,2}, \sigma^2_{n,2}) &= \left(\hat{\mu}_n - \sqrt{\frac{\kappa}{n}}\hat{\sigma}_n, \hat{\sigma}^2_n\right), \\
(\mu_{n,3}, \sigma^2_{n,3}) &= \left(\hat{\mu}_n, \left(1 + \sqrt{\frac{2\kappa}{n}}\right)\hat{\sigma}^2_n\right), & (\mu_{n,4}, \sigma^2_{n,4}) &= \left(\hat{\mu}_n, \left(1 - \sqrt{\frac{2\kappa}{n}}\right)\hat{\sigma}^2_n\right).
\end{aligned}$$

Therefore, taking into account (6.1), we have a recursive formula for computing these extreme points, and thus the desired recursive construction of the confidence regions $\mathcal{T}_n$.

7 Concluding remarks and open problems

In this paper we initiated the theory of recursive confidence regions. In part, this theory hinges on the theory of recursive identification for stochastic dynamical systems, such as a Markov chain, which is the main model studied here. Although the results in the existing literature on statistical inference for Markov processes are quite general, not much work has been done on the recursive identification methods for Markov processes. Our results provide a useful contribution in this regard, but, they are subject to assumption of ergodicity imposed on our Markov chain. We leave the study of more general cases to the future work.

Also, we leave for the future work the study of recursive confidence regions generated via constrained recursive point estimation algorithms.

A Appendix

A.1 Ergodic Theory for Markov Chains

In this section, we will briefly discuss the theory of ergodicity for (time homogeneous) Markov processes in discrete time. Note that for fixed transition kernel Q and initial distribution μ, all the corresponding Markov processes have the same law. With this in mind, we will present results regarding ergodicity of Markov processes associated to the canonical construction from Q and μ. We start with recalling the notion for ergodicity of general dynamical systems.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $T : \Omega \to \Omega$ be a measure preserving map, i.e. a map such that $\mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ for every $A \in \mathscr{F}$. Then, the corresponding dynamical system is defined as the quadruple $(\Omega, \mathscr{F}, \mathbb{P}, T)$. Define $\mathcal{G} := \{A \in \mathscr{F} : T^{-1}(A) = A\}$, and note that $\Omega, \emptyset \in \mathcal{G}$. Then, we have the following

Definition A.1. A dynamical system $(\Omega, \mathscr{F}, \mathbb{P}, T)$ is said to be ergodic if for any $A \in \mathcal{G}$ we have $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

One important result in the theory of dynamical system is the celebrated Birkhoff's Ergodic Theorem (See e.g. [Bir31], [vN32b], [vN32a]).

Theorem A.2 (Birkhoff's Ergodic Theorem)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P}, T)$ be an ergodic dynamical system. If $f \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Then,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n\omega) = \mathbb{E}_{\mathbb{P}}[f] \quad \mathbb{P} - a.s.$$

We now proceed by introducing the canonical construction of time homogeneous Markov chains. Let $(\mathcal{X}, \mathfrak{X})$ be a measurable space. Also, let $Q : \mathcal{X} \times \mathfrak{X} \to [0, 1]$ be a transition kernal and π be a probability measure on $(\mathcal{X}, \mathfrak{X})$ such that $\pi(A) = \int_{\mathcal{X}} Q(x, A)\pi(dx)$, for any $A \in \mathfrak{X}$. Such measure π is called an invariant probability measure of Q. For every $n \geq 0$, we define a probability measure $\mathbb{P}_\pi^{Q,n}$ on $(\mathcal{X}^{n+1}, \mathfrak{X}^{n+1})$, where $\mathfrak{X}^{n+1}$ is the product σ-algebra on $\mathcal{X}^{n+1}$, by

$$\mathbb{P}_\pi^{Q,n}(A_0 \times \ldots \times A_n) = \int_{A_0} \cdots \int_{A_n} Q(x_{n-1}, dx_n) \cdots Q(x_0, dx_1)\pi(dx_0),$$

for any $A_0, \ldots, A_n \in \mathfrak{X}$. The sequence of measures $\{\mathbb{P}_\pi^{Q,n}\}_{n>0}$ is consistent. That is,

$$\mathbb{P}_\pi^{Q,n}(A_0 \times A_1 \ldots \times A_n) = \mathbb{P}_\pi^{Q,n+m}(A_0 \times A_1 \times \ldots \times A_n \times \mathcal{X}^m),$$

holds true for any integer $m > 0$, and $A_0, \ldots, A_n \in \mathfrak{X}$. Therefore, by Kolmogorov's extension theorem, such family of measures extends to a unique measure $\mathbb{P}_\pi^Q$ on $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}})$, such that

$$\mathbb{P}_\pi^Q(A_0 \times A_1 \ldots \times A_n \times \mathcal{X}^\infty) = \mathbb{P}_\pi^{Q,n}(A_0 \times A_1 \ldots \times A_n), \quad A_0, \ldots, A_n \in \mathfrak{X}. \tag{A.1}$$

With a slight abuse of notation, we denote by T the (one step) shift map on $\mathcal{X}^{\mathbb{N}}$

$$(T(\omega))_k = \omega_{k+1}, \quad \omega \in \mathcal{X}^{\mathbb{N}}.$$

Due to the construction of $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q)$ and the fact that π is an invariant measure, then it can be verified that T is measure preserving, and therefore $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q, T)$ is a dynamical system.

Next, define a process X on $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q)$ by

$$X(\omega) = \omega, \quad \omega \in \mathcal{X}^{\mathbb{N}},$$

so that, in particular, $X_n(\omega) = \omega(n)$ for any integer $n \geq 0$. A process defined in this way is called a *canonical process* on $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q)$.

We now state and prove the following result,

Lemma A.3. *A canonical process X on $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q)$ is a time homogenous Markov chain with transition kernel Q, and thus it is called the canonical Markov chain on $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_\pi^Q)$. Moreover, the initial distribution of X coincides with π, so that*

$$\mathbb{P}_\pi^Q(X_0 \in A) = \pi(A).$$

Consequently, process X is a stationary process, that is, for any $n \geq 1$, the law of $(X_j, X_{j+1}, \ldots, X_{j+n})$ under $\mathbb{P}_\pi^Q$ is independent of j, $j \geq 0$.

Proof. For any $n > 1$, denote by $dx_{0:n} := dx_n \times \cdots \times dx_0$. According to (A.1) and the definition of $\mathbb{P}_\pi^{Q,n}$, we obtain that

$$\mathbb{P}_\pi^Q(dx_{0:n}) = \mathbb{P}_\pi^{Q,n}(dx_{0:n}) = Q(x_{n-1}, dx_n) \cdots Q(x_0, dx_1)\pi(dx_0).$$

Next, for any $A_0, \ldots, A_n \in \mathfrak{X}$, we get that

$$\begin{aligned} \mathbb{P}_\pi^Q(A_n \times \cdots \times A_0) &= \mathbb{E}_\pi^Q[\mathbb{1}_{A_n \times \cdots \times A_0}] \\ &= \int_{A_{n-1} \times \cdots \times A_0} \mathbb{E}_\pi^Q[\mathbb{1}_{A_n} \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0] \mathbb{P}_\pi^Q(dx_{0:n-1}). \end{aligned} \tag{A.2}$$

On the other hand, we also have that

$$\begin{aligned} \mathbb{P}_\pi^Q(A_n \times \cdots A_0) = \mathbb{P}_\pi^{Q,n}(A_n \times \cdots A_0) &= \int_{A_0} \cdots \int_{A_n} Q(x_{n-1}, dx_n) \cdots Q(x_0, dx_1)\pi(dx_0) \\ &= \int_{A_{n-1} \times \cdots \times A_0} \int_{A_n} Q(x_{n-1}, dx_n) \mathbb{P}_\pi^Q(dx_{0:n-1}). \end{aligned} \tag{A.3}$$

(A.2) and (A.3) yield that

$$\begin{aligned} \mathbb{P}_\pi^Q(X_n \in A_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= \mathbb{E}_\pi^Q[\mathbb{1}_{A_n} \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0] \\ &= \int_{A_n} Q(x_{n-1}, dx_n) = \mathbb{P}_\pi^Q(X_n \in A_n \mid X_{n-1} = x_{n-1}). \end{aligned}$$

Therefore, we conclude that X is a Markov chain.

Now we prove the initial distribution of X is π. By definition of X we have

$$\mathbb{P}_\pi^Q(X_0 \in A) = \mathbb{P}_\pi^Q(\omega(0) \in A) = \mathbb{P}_\pi^Q(A \times \mathcal{X}^\infty).$$

Then, according to (A.1), it is true that

$$\mathbb{P}_\pi^Q(A \times \mathcal{X}^\infty) = \mathbb{P}_\pi^{Q,0}(A) = \pi(A).$$

Therefore,

$$\mathbb{P}_\pi^Q(X_0 \in A) = \pi(A),$$

and π is the initial distribution of X.

We finish the proof by showing the stationarity of X. That is to prove for any fixed $n \geq 1$, the probability $\mathbb{P}_\pi^Q(X_j \in A_0, \dots, X_{n+j} \in A_n)$ is independent of $j \geq 0$. Since π is invariant measure Q, then it is clear that $\mathbb{P}_\pi^Q(X_j \in A_0) = \pi(A_0)$. Next, we have

$$\mathbb{P}_\pi^Q(X_j \in A_0, \dots, X_{n+j} \in A_n) = \mathbb{P}_\pi^Q(\mathcal{X}^j \times A_0 \times \cdots \times A_n \times \mathcal{X}^\infty),$$

where the right hand side is equal to $\mathbb{P}_\pi^{Q,n+j}(\mathcal{X}^j \times A_0 \times \cdots \times A_n)$ by (A.1). Finally, according to the definition of $\mathbb{P}_\pi^{Q,n+j}$, we have

$$\begin{aligned}
\mathbb{P}_\pi^{Q,n+j}(\mathcal{X}^j \times A_0 \times \ldots \times A_n) &= \int_{\mathcal{X}} \int_{A_0} \cdots \int_{A_n} Q(x_{n+j-1}, dx_{n+j}) \\
&\qquad \cdots Q(x_{n-1}, dx_n) \cdots Q(x_0, dx_1)\pi(dx_0), \\
&= \int_{A_0} \cdots \int_{A_n} Q(x_{n+j-1}, dx_{n+j}) \cdots Q(x_j, dx_{j+1})\pi(dx_j) \\
&= \mathbb{P}_\pi^{Q,n}(A_0 \times \ldots \times A_n) = \mathbb{P}_\pi^Q(A_0 \times \ldots \times A_n \times \mathcal{X}^\infty) \\
&= \mathbb{P}_\pi^Q(X_0 \in A_0, \dots, X_n \in A_n).
\end{aligned}$$

We now conclude that X is a stationary process. □

Remark A.4. If a transition kernel Q admits an invariant measure π, then it is customary to say that π is an invariant measure for any Markov chain corresponding to Q. In particular, π is the invariant measure for the canonical Markov chain X on $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q)$.

We proceed by defining the notion of ergodicity for a canonical Markov chain X.

Definition A.5. The canonical Markov chain X on $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q)$ is said to be ergodic if $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q, T)$ is an ergodic dynamical system.

Remark A.6. Note that since an ergodic Markov chain X is, in particular, a canonical Markov chain on $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q)$, then it is a stationary process.

Through the rest of this section X denote the canonical Markov chain defined on $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q)$. The following technical result is on of the keys technical results used in this paper. In its formulation we denote by $\mathbb{E}_\pi^Q$ the expectation under measure $\mathbb{P}_\pi^Q$.

Proposition A.7. *Let X be ergodic. Then for any g such that $\mathbb{E}_\pi^Q[g(X_0, \dots, X_n)] < \infty$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} g(X_i, \dots, X_{i+n}) = \mathbb{E}_\pi^Q[g(X_0, \dots, X_n)] \quad \mathbb{P}_\pi^Q - a.s.$$

Proof. By definition, we have that $(\mathcal{X}^\mathbb{N}, \mathfrak{X}^\mathbb{N}, \mathbb{P}_\pi^Q, T)$ is an ergodic dynamical system.

For fixed $n > 0$, take $f : \mathcal{X}^\mathbb{N} \to \mathbb{R}$ defined as $f(\omega) := g(\omega(0), \dots, \omega(n))$ for any $\omega \in \mathcal{X}^\mathbb{N}$. Note that

$$\omega(j) = X_j(\omega), \quad j \geq 0,$$

and

$$T^i(\omega)(j) = X_{i+j}(\omega), \quad i, j \geq 0.$$

Then, according to Birkhoff's ergodic theorem, we get that for almost every $\omega \in \mathcal{X}^{\mathbb{N}}$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} g(X_i(\omega), \ldots, X_{i+n}(\omega)) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i(\omega)) = \mathbb{E}_{\pi}^{Q}[f] = \mathbb{E}_{\pi}^{Q}[g(X_0, \ldots, X_n)].$$

□

We finish this section with providing a brief discussion regarding sufficient conditions for the Markov chain X to be ergodic. Towards this end, first note that, in general, a transition kernel Q possesses more than one invariant measures, and we quote the following structural result regarding the set of invariant measures of Q,

Proposition A.8. *[Var01] Let $Q : \mathcal{X} \times \mathfrak{X} \to [0, 1]$ be a (one step) transition kernel. Then, the set Π_Q of all invariant probability measures of Q is convex. Also, given a measure $\pi \in \Pi$, the system $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_{\pi}^{Q}, T)$ is ergodic if and only if π is an extremal point of Π. Furthermore, any two ergodic measures are either identical or mutually singular.*

Proposition A.8 implies

Corollary A.9. *If a transition kernel Q has a unique invariant probability measure π, then the system $(\mathcal{X}^{\mathbb{N}}, \mathfrak{X}^{\mathbb{N}}, \mathbb{P}_{\pi}^{Q}, T)$ is ergodic.*

One powerful tool for checking the uniqueness of invariant probability measure is the notion of positive Harris chain. There are several equivalent definitions of positive Harris Markov chain, and we will use the one from [HLL00].

Definition A.10. The Markov chain X with transition kernel Q is called a positive Harris chain if

(a) there exists a σ-finite measure μ on $\mathfrak{X}$ such that for any $x_0 \in \mathcal{X}$, and $B \in \mathfrak{X}$ with $\mu(B) > 0$

$$\mathbb{P}(X_n \in B \text{ for some } n < \infty | X_0 = x_0) = 1,$$

(b) there exists an invariant probability measure for Q.

Proposition A.11. *If X is a positive Harris chain, then X is ergodic.*

Proof. It is well known (cf. e.g. [MT93]) that a positive Harris chain admits a unique invariant measure. Thus, the result follows form Corollary (A.9).

□

A.2 CLT for Multivariate Martingales

In this section, for a matrix A with real valued entries we denote by $|A|$ the sum of the absolute values of its entries.

In [CP05] Proposition 3.1, the authors gave the following version of the central limit theorem for discrete time multivariate martingales.

Proposition A.12. *On a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ let $D = \{D_{n,j}, 0 \leq j \leq k_n, n \geq 1\}$ be a triangular array of d-dimensional real random vectors, such that, for each n, the finite sequence $\{D_{n,j}, \mathbb{1} \leq j \leq k_n\}$ is a martingale difference process with respect to some filtration $\{\mathscr{F}_{n,j}, j \geq 0\}$. Set*

$$D_n^* = \sup_{1 \leq j \leq k_n} |D_{n,j}|, \quad U_n = \sum_{j=1}^{k_n} D_{n,j} D_{n,j}^T.$$

Also denote by $\mathscr{U}$ the σ-algebra generated by $\bigcup_j \mathscr{H}_j$ where $\mathscr{H}_j := \liminf_n \mathscr{F}_{n,j}$. Suppose that D_n^* converges in L^1 to zero and that U_n converges in probability to a $\mathscr{U}$ measurable d-dimensional, positive semi-definite matrix U. Then, the random vector $\sum_{j=1}^{k_n} D_{n,j}$ converges $\mathscr{U}$-stably to the Gaussian kernel $\mathcal{N}(0, U)$.

Remark A.13. $\mathscr{U}$-stable convergence implies convergence in distribution; it is enough to take the entire Ω in the definition of $\mathscr{U}$-stable convergence. See for example [AE78] or [HL15].

We will apply the above proposition to the process $\{\psi_n(\theta^*), n \geq 0\}$ such that Assumption M, R8 and R9 are satisfied. To this end, let us define the triangular array $\{D_{n,j}, 1 \leq j \leq n, n \geq 1\}$ as

$$D_{n,j} = \frac{1}{\sqrt{n}} \psi_j(\theta^*),$$

and let us take $\mathscr{F}_{n,j} = \mathscr{F}_j$.

First, note that $\mathbb{E}_{\theta^*}[\psi_j(\theta^*)|\mathscr{F}_{j-1}] = 0$, so that for any $n \geq 1$, $\{D_{n,j}, 1 \leq j \leq n\}$ is a martingale difference process with respect to $\{\mathscr{F}_j, 0 \leq j \leq n\}$. Next, R9 implies that $D_n^* := \sup_{1 \leq j \leq n} \frac{1}{\sqrt{n}} |\psi_j(\theta^*)|$ converges in L^1 to 0. Finally, stationarity, R8 and ergodicity guarantee that

$$U_n := \frac{1}{n} \sum_{j=1}^{n} \psi_j(\theta^*) \psi_j^T(\theta^*) \to \mathbb{E}_{\theta^*}[\psi_1(\theta^*) \psi_1^T(\theta^*)] \quad \mathbb{P}_{\theta^*} - a.s.$$

The limit $I(\theta^*) = \mathbb{E}_{\theta^*}[\psi_1(\theta^*) \psi_1^T(\theta^*)]$ is positive semi-definite, and it is deterministic, so that it is measurable with respect to any σ-algebra. Therefore, applying Proposition A.12 and Remark A.13 we obtain

Proposition A.14. *Assume that Assumption M, R8, and R9 are satisfied. Then,*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi_j(\theta^*) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, I(\theta^*)).$$

A.3 Technical Supplement

Assumptions R4–R6 are stated for any deterministic vector $\theta \in \boldsymbol{\Theta}$. In this section, we show that if (2.3)-(2.7) hold for $\theta \in \boldsymbol{\Theta}$, then for any random vectors $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ that are $\mathscr{F}_{n-1}$ measurable and take values in $\boldsymbol{\Theta}$, analogous inequalities are true.

Proposition A.15. *Assume that R4-R6 are satisfied. Then, for any fixed $n \geq 1$ and for any random vectors $\boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ that are $\mathscr{F}_{n-1}$ measurable and take values in $\boldsymbol{\Theta}$, we have*

$$\mathbb{E}_{\theta^*}[\|\psi_n(\boldsymbol{\theta})\|^2|\mathscr{F}_{n-1}] \leq c(1 + \|\boldsymbol{\theta} - \theta^*\|^2), \tag{A.4}$$

$$(\boldsymbol{\theta} - \theta^*)^T b_n(\boldsymbol{\theta}) \leq -K_1\|\boldsymbol{\theta} - \theta^*\|^2, \tag{A.5}$$

$$\|b_n(\boldsymbol{\theta})\| \leq K_2\|\boldsymbol{\theta} - \theta^*\|, \tag{A.6}$$

$$\mathbb{E}_{\theta^*}[\|\Psi_n(\boldsymbol{\theta}_1) - \Psi_n(\boldsymbol{\theta}_2)\||\mathscr{F}_{n-1}] \leq K_3\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|, \tag{A.7}$$

$$\mathbb{E}_{\theta^*}[\|\mathsf{H}\psi_n(\boldsymbol{\theta})\||\mathscr{F}_{n-1}] \leq K_4. \tag{A.8}$$

Proof. We will only show that (A.5) is true. The validity of the remaining inequalities can be proved similarly. Also, without loss of generality, we assume that $d = 1$.

From (2.4), we have for any $\theta \in \boldsymbol{\Theta}$, $(\theta - \theta^*)\mathbb{E}_{\theta^*}[\psi_n(\theta) \mid \mathscr{F}_{n-1}] \leq K_1|\theta - \theta^*|$. If $\boldsymbol{\theta}$ is a simple random variable, i.e. there exists a partition $\{A_m, 1 \leq m \leq M\}$ of Ω, where M is a fixed integer, such that $A_m \in \mathcal{F}_{n-1}$, $1 \leq m \leq M$, and $\boldsymbol{\theta} = \sum_{m=1}^M c_m \mathbb{1}_{A_m}$, where $c_m \in \boldsymbol{\Theta}$. Then, we have that

$$\begin{aligned}
(\boldsymbol{\theta} - \theta^*)\, b_n(\boldsymbol{\theta}) &= (\sum_{m=1}^M c_m \mathbb{1}_{A_m} - \theta^*)\mathbb{E}_{\theta^*}[\psi_n(\boldsymbol{\theta}) \mid \mathscr{F}_{n-1}] \\
&= \sum_{m=1}^M \mathbb{1}_{A_m}(c_m - \theta^*)\mathbb{E}_{\theta^*}[\mathbb{1}_{A_m}\psi_n(\boldsymbol{\theta}) \mid \mathscr{F}_{n-1}] \\
&= \sum_{m=1}^M \mathbb{1}_{A_m}(c_m - \theta^*)\mathbb{E}_{\theta^*}[\mathbb{1}_{A_m}\psi_n(c_m) \mid \mathscr{F}_{n-1}] \\
&= \sum_{m=1}^M \mathbb{1}_{A_m}(c_m - \theta^*)\mathbb{E}_{\theta^*}[\psi_n(c_m) \mid \mathscr{F}_{n-1}] \\
&\leq -\sum_{m=1}^M \mathbb{1}_{A_m} K_1|c_m - \theta^*|^2 = -\sum_{m=1}^M K_1|\boldsymbol{\theta} - \theta^*|^2.
\end{aligned}$$

From here, using the usual limiting argument we conclude that (A.5) holds true for any $\mathscr{F}_{n-1}$ measurable random variable $\boldsymbol{\theta}$. □

In the rest of this section we will verify that the Assumption M and the properties R0–R9 are satisfies in Example 6.1.

It is clear that the Markov chain $\{Z_n, n \geq 0\}$, as defined in Example 6.1, satisfies (i) and (iii) in Assumption M. Next we will show that Z is a positive Harris chain (see Definition A.10). For any Borel set $B \in \mathcal{B}(\mathbb{R})$ with strictly positive Lebesgue measure, and any $z_0 \in \mathbb{R}$, we have that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}_{\theta^*}(Z_n \notin B, \dots, Z_1 \notin B \mid Z_0 = z_0) \\
&\qquad = \lim_{n\to\infty} \mathbb{P}_{\theta^*}(Z_n \notin B \mid Z_{n-1} \notin B) \cdots \mathbb{P}_{\theta^*}(Z_2 \notin B \mid Z_1 \notin B)\mathbb{P}_{\theta^*}(Z_1 \notin B \mid Z_0 = z_0) \\
&\qquad = \lim_{n\to\infty} \mathbb{P}_{\theta^*}(Z_2 \notin B \mid Z_1 \notin B)^{n-1}\mathbb{P}_{\theta^*}(Z_1 \notin B \mid Z_0 = z_0) = 0,
\end{aligned}$$

and thus Z satisfies Definition A.10.(a). Also, since the density (with respect to the Lebesgue measure) of Z_1 is

$$f_{Z_1,\theta^*}(z_1) = \int_{\mathbb{R}} p_{\theta^*}(z_0, z_1) f_{Z_0,\theta^*}(z_0) dz_0 = \frac{1}{\sqrt{2\pi}\sigma^*} e^{-\frac{(z_1-\mu^*)^2}{2(\sigma^*)^2}},$$

then $Z_1 \sim \mathcal{N}(\mu^*, (\sigma^*)^2)$, and consequently, we get that $Z_n \sim \mathcal{N}(\mu^*, (\sigma^*)^2)$ for any $n \geq 0$. This implies that $\mathcal{N}(\mu^*, (\sigma^*)^2)$ is an invariant distribution for Z. Thus, Z is a positive Harris chain, and respectively, by Proposition A.11, Z is an ergodic process.

As far as propreties R0–R9, we fist note that

$$\begin{aligned}
\psi_n(\theta) &= \nabla \log p_\theta(Z_{n-1}, Z_n) \\
&= \Big(\frac{Z_n - \rho Z_{n-1} - (1-\rho)\mu}{\sigma^2(1+\rho)}, -\frac{1}{\sigma} + \frac{(Z_n - \rho Z_{n-1} - (1-\rho)\mu)^2}{(1-\rho^2)\sigma^3} \Big)^T, \\
b_n(\theta) &= \mathbb{E}_{\theta^*}[\psi_n(\theta) | \mathscr{F}_{n-1}] \\
&= \Big(-\frac{(1-\rho)(\mu-\mu^*)}{\sigma^2(1+\rho)}, \frac{\sigma^{*,2} - \sigma^2}{\sigma^3} + \frac{(1-\rho)(\mu-\mu^*)^2}{(1+\rho)\sigma^3} \Big)^T, \\
\Psi_n(\theta) &= \begin{bmatrix} -\frac{1-\rho}{(1+\rho)\sigma^2} & -\frac{2(Z_n - \rho Z_{n-1} - (1-\rho)\mu)}{(1+\rho)\sigma^3} \\ -\frac{2(Z_n - \rho Z_{n-1} - (1-\rho)\mu)}{(1+\rho)\sigma^3} & \frac{1}{\sigma^2} - \frac{3(Z_n - \rho Z_{n-1} - (1-\rho)\mu)^2}{(1-\rho^2)\sigma^4} \end{bmatrix}.
\end{aligned}$$

We denote by $Y_n := Z_n - \rho Z_{n-1} - (1-\rho)\mu$, and we immediately deduce that that

$$\begin{aligned}
\mathbb{E}_{\theta^*}[Y_n \mid \mathscr{F}_{n-1}] &= (1-\rho)(\mu^* - \mu), \\
\mathbb{E}_{\theta^*}[Y_n^2 \mid \mathscr{F}_{n-1}] &= (1-\rho)^2(\mu-\mu^*)^2 + (\sigma^*)^2(1-\rho^2), \\
\mathbb{E}_{\theta^*}[Y_n^4 \mid \mathscr{F}_{n-1}] &= (1-\rho)^4(\mu^*-\mu)^4 + 6(1+\rho)(1-\rho)^3(\mu^*-\mu)^2(\sigma^*)^2 \\
&\quad + 3(\sigma^*)^4(1-\rho^2)^2.
\end{aligned} \tag{A.9}$$

From here, and using the fact that $\mathbf{\Theta}$ is bounded, it is straightforward, but tedious,[9] to show that R4, R5, R6, and R7 are satisfied. Also, it is clear note that R0 is true, and using (A.9) by direct computations we get that R1 and R2 are satisfied.

Since

$$\mathbb{E}_{\theta^*}[\psi_1(\theta)] = \left(\frac{(1-\rho)(\mu^*-\mu)}{\sigma^2(1+\rho)}, \frac{(\sigma^*)^2 - \sigma^2}{\sigma^3} + \frac{(1-\rho)(\mu-\mu^*)^2}{(1+\rho)\sigma^3} \right),$$

then R3 is clearly satisfied.

[9]The interested reader can contact the authors for details.

Again by direct evaluations, we have that

$$I(\theta) = \mathbb{E}_\theta[\psi_1(\theta)\psi_1(\theta)^T] = \begin{bmatrix} \frac{1-\rho}{(1+\rho)\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix},$$

which is positive definite matrix, and thus R8 is satisfied.

Finally, we will verify R9. By Jensen's inequality and Cauchy-Schwartz inequality, we have that

$$\begin{aligned}
\exp\left(\mathbb{E}_{\theta^*} \sup_{0\leq i\leq n} |\psi_i(\theta^*)|\right) &\leq \mathbb{E}_{\theta^*} \exp\left(\sup_{0\leq i\leq n} |\psi_i(\theta^*)|\right) = \mathbb{E}_{\theta^*}\left[\sup_{0\leq i\leq n} \exp|\psi_i(\theta^*)|\right] \\
&\leq \sum_{i=1}^n \mathbb{E}_{\theta^*} \exp|\psi_i(\theta^*)| \leq \sum_{i=1}^n \mathbb{E}_{\theta^*} \exp(\frac{|Y_i|}{\sigma^2(1+\rho)} + \frac{1}{\sigma} + \frac{Y_n^2}{(1-\rho)^2\sigma^3}) \\
&\leq \sum_{i=1}^n \left(\mathbb{E}_{\theta^*} \exp(\frac{2|Y_i|}{\sigma^2(1+\rho)})\right)^{\frac{1}{2}} \left(\mathbb{E}_{\theta^*} \exp(\frac{2}{\sigma} + \frac{2Y_i^2}{(1-\rho)^2\sigma^3})\right)^{\frac{1}{2}}.
\end{aligned}$$

Note that for $Y_i, i = 0, \ldots, n$ is normally distributed, and therefore, there exist two constants C_1 and C_2, that depend on θ^* such that

$$\mathbb{E}_{\theta^*} \exp\left(\frac{2|Y_i|}{\sigma^2(1+\rho)}\right) = C_1, \quad \mathbb{E}_{\theta^*} \exp\left(\frac{2}{\sigma} + \frac{2Y_i^2}{(1-\rho)^2\sigma^3}\right) = C_2.$$

Hence, we have that

$$\mathbb{E}_{\theta^*} \sup_{0\leq i\leq n} |\psi_i(\theta^*)| \leq \log n + \frac{1}{2}\log C_1C_2,$$

and, thus R9 is satisfied:

$$\lim_{n\to\infty} \mathbb{E}_{\theta^*}\left[\sup_{0\leq i\leq n}\left|\frac{1}{\sqrt{n}}\psi_i(\theta^*)\right|\right] \leq \lim_{n\to\infty}\left(\frac{\log n}{\sqrt{n}} + \frac{\log C_1C_2}{2\sqrt{n}}\right) = 0.$$

Acknowledgments

Part of the research was performed while Igor Cialenco was visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation.

References

[AE78] D. J. Aldous and G. K. Eagleson. On Mixing and Stability of Limit Theorems. *The Annals of Probability*, 6(2):325–331, April 1978.

[BCC+16] T.R. Bielecki, I. Cialenco, T. Chen, A. Cousin, and M. Jeanblanc. Adaptive Robust Hedging Under Model Uncertainty. In preparation, 2016.

[Bir31] G. Birkhoff. Proof of the Ergodic Theorem. *Proc. Natl. Acad. Sci. USA*, 17(12):656–660, 1931.

[BK02] R. Buche and H. J. Kushner. Rate of convergence for constrained stochastic approximation algorithms. *SIAM J. Control Optim.*, 40(4):1011–1041, 2001/02.

[CP05] I. Crimaldi and L. Pratelli. Convergence Result for Multivariate Martingales. *Stoch. Proc. and their Applic.*, 115:571–577, 2005.

[Fab68] V. Fabian. On Asymptotic Normality in stochastic approximation. *Ann. Math. Statist.*, 39(4):1327–1332, 1968.

[Fis25] R. A. Fisher. Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, 22:700–725, 1925.

[HL15] E. Häusler and H. Luschgy. *Stable Convergence and Stable Limit Theorems.* Springer, 2015.

[HLL00] O. Hernández-Lerma and J. Lasserre. Further criteria for positive Harris recurrence of Markov chains. *Proceedings of the American Mathematical Society*, 129(5):1521–1524, 2000.

[KC78] H. Kushner and D. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, volume 26 of *Applied Mathematical Sciences.* Springer-Verlag New York, 1978.

[KW52] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[KY03] H. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, volume 35 of *Stochastic Modelling and Applied Probability.* Springer-Verlag New York, 2003.

[LeC56] L. LeCam. On the asymptotic theory of estimation and testing hypothesis. *Proccedings in the Third Berkely Symposium on Mathematical Statist. and Probab.*, 1:129–156, 1956.

[LeC60] L. LeCam. Locally asymptotic normal families of distributions. *Univ. California Publ. in Statist.*, 3(2):37–98, 1960.

[LR79] T. L. Lai and H. Robbins. Adaptive design and stochastic approximation. *Ann. Statist.*, 7(6):1196–1221, 1979.

[LS87] L. Ljung and Söderström. *Theory and Practice of Recursive Identification*, volume 4 of *The MIT Press Series in Signal Processing, Optimization, and Control.* MIT Press, 1987.

[MT93] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge Mathematical Library. Cambridge University Press, 1993.

[RM51] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[Sac58] J. Sacks. Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.*, 29(2):373–405, 1958.

[Sha10] T. Sharia. Recursive parameter estimation: asymptotic expansion. *Annals of the Institute of Statistical Mathematics*, 62(2):343–362, 2010.

[Shi84] A. Shiryayev. *Probability*, volume 95 of *Graduate Texts in Mathematics.* Springer New York, 1984.

[Var01] S. R. S. Varadhan. *Probability Theory*, volume 7 of *Courant Lecture Notes.* American Mathematical Society, 2001.

[vN32a] J. von Neumann. Physical applications of the ergodic hypothesis. *Proc. Natl. Acad. Sci. USA*, 18(3):263–266, 1932.

[vN32b] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA*, 18(1):70–82, 1932.