



HAL
open science

Kullback-Leibler Approach to CUSUM Quickest Detection Rule for Markovian Time Series

Valerie Girardin, Victor Konev, Serguei Pergamenchtchikov

► **To cite this version:**

Valerie Girardin, Victor Konev, Serguei Pergamenchtchikov. Kullback-Leibler Approach to CUSUM Quickest Detection Rule for Markovian Time Series. *Sequential Analysis*, 2018, 37, pp.322 - 341. 10.1080/07474946.2018.1548846 . hal-02334914

HAL Id: hal-02334914

<https://hal.science/hal-02334914>

Submitted on 29 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Kullback-Leibler Approach to CUSUM Quickest Detection Rule for Markovian Time Series

Valérie Girardin

Laboratoire de Mathématiques N. Oresme, UMR6139 CNRS
Université de Caen Normandie, France

Victor Konev

Department of Applied Mathematics and Cybernetics,
Tomsk State University, Russian Federation

Serguei Pergamenchtchikov

Laboratoire de Mathématiques Raphaël Salem, CNRS, UMR 6085, Université de Rouen, France
and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance
of National Research Tomsk State University, Russian Federation

Abstract: Optimality properties of decision procedures are studied for the quickest detection of a change-point of parameters in autoregressive and other Markov type sequences. The limit of the normalized conditional log-likelihood ratios is shown to exist for Markov chains satisfying the ergodic theorem of information theory. Then closed-form expressions for this limit are derived by making use of the time average rate of Kullback-Leibler divergence. The good properties of the detection procedures based on a sequential analysis approach are proven to hold thanks to geometric ergodicity properties of the observation processes. In particular, the window-limited CUSUM rule is shown to be optimal for detecting the disruption point in autoregressive models. Sparre Andersen models are specifically studied.

Keywords: Change-point detection, CUSUM detection procedure, Ergodic Theorem of information theory, Kullback-Leibler divergence, Relative entropy, Sequential detection, Sparre Andersen models.

Subject Classifications: 62L10, 62L15, 60G40, 60J10, 62B10, 94A17..

1. INTRODUCTION

The problem of quick detection of an abrupt change in time series arises in many different areas, related to automatic control, segmentation of signals, biomedical signal processing, quality control engineering, finance, link failure detection in communication networks, intrusion detection in computer systems, and target detection in surveillance systems; see Page (1954), Basseville and Nikiforov (1993), Kent (2000), Tartakovsky et al. (2015) for details and further references. A challenging application area is intrusion detection in distributed computer networks; see Kent (2000) and Tartakovsky et al. (2004). Large scale

Address correspondence to V. Girardin, LMNO, Université de Caen Normandie, BP5186, 14032 Caen, France; E-mail: valerie.girardin@unicaen.fr

attacks, such as denial-of-service attacks, occur at unknown time points and need to be detected in the early stages by observing abrupt changes in the network traffic.

In change point analysis, a large variety of observation models is used. They include independent identically distributed (i.i.d.) sequences of random variables whose distributions change at the disruption time and also different kinds of processes with dependent variables; see Tartakovsky et al. (2015), Pergamenchtchikov and Tartakovsky (2016) for autoregressive processes, and Yakir (1994) for Markov chains of order 1 with finite state spaces, extended in Lai (1998) to general state spaces. Two aspects are essential: the decision rule must both provide a low false-alarm rate and ensure quick detection of the disruption after it occurs. However, in contrast to the analysis of detection procedures for i.i.d. models, most procedures for dependent models reduce to only one characteristic, the false alarm rate, while the delay time characteristic is studied only by means of numerical simulation for some specific models.

The asymptotic minimax theory for the i.i.d. case is developed in Lorden (1971), which both formulates the optimization problem as the minimization of the worst-case delay time subject to a lower bound on the mean time between false alarms, and proves the optimality of Page’s CUSUM algorithm. In recent years, efforts have been made to extend the theory of optimal detection beyond the i.i.d. setting. A general asymptotic minimax theory of change-point detection in stochastic systems is developed in Lai (1998), under an alternative constraint on false-alarm rate – precisely that the probability of a false alarm within a given period of length be uniformly small whenever the period starts. In particular, the asymptotic optimality of the procedures is proven under general conditions on the conditional log-likelihood ratio statistics (CLLRS).

Among possible conditions is the convergence of the normalized CLLRS to some finite positive quantity, say \mathbf{I} . In the i.i.d. case, this quantity is shown in Lorden (1971) – by standard arguments on error probabilities of sequential log-likelihood ratio tests, to amount to the Kullback-Leibler divergence (KLD) between the two involved pre and post-change distributions. A closed-form expression of \mathbf{I} is provided in Lai (1998) for Markov chains of order 1 in terms of the stationary distribution after the disruption and on the two involved transition densities. An expression in terms of the Fisher matrix of the model is also provided in Pergamenchtchikov and Tartakovsky (2016) for autoregressive models of order 1. None of them relate these expressions to the concept of entropy rate, classical in information theory.

The analysis of the properties of the detection procedures will here be essentially based on the Kullback-Leibler divergence (KLD) rates – see Girardin (2005) and Girardin and Limnios (2006), and the geometric ergodicity properties of stable autoregressive processes – see Galtchouk and Pergamenchtchikov (2014). Precisely, the present paper has two linked objectives. First, the problem of convergence of the normalized CLLRS is addressed. The existence of the limit \mathbf{I} is obtained through the ergodic theorem of information theory, also called asymptotic equipartition property (AEP); see Girardin (2005). Thus, \mathbf{I} is shown to be a KLD rate, with a closed-form expression for Markov chains of any order. Specific expressions are also derived for Gaussian autoregressive models. Second, in order to apply an optimal change-point detection rule to a specific model, some general conditions providing optimality have to be satisfied – all of them involving convergence of the normalized CLLRS to the KLD rate. For many models of interest in applications, the verification of these conditions becomes a stumbling block. Here we prove the optimality of the CUSUM and generalized CUSUM procedures for Markov chains and autoregressive models under simple conditions. To this end, a method is developed for checking their sufficiency in a general optimal detection framework, by using the closed-form expression obtained for the KLD rate.

This paper is organized as follows. In Section 2 we consider the problem of quick detection of a change-point in general models with dependent data. As in both Lai (1998) and Pergamenchtchikov and Tartakovsky (2016), we impose upper bounds on the uniform probability of the false alarms. We recall some general conditions on the CLLRS, required for the optimality of the CUSUM rule for the models treated thereafter. In Section 3, through an ergodic theory approach, we determine the limit \mathbf{I} of the normalized CLLRS under the form of a KLD rate, for Markov chains, and for autoregressive models; we also state conditions of convergence in terms of AEP. In Section 4, we treat the problem of change-point detection for the Sparre

Andersen model – widely used in insurance. In Section 5, we study the problem of the quickest detection in autoregressive models with Gaussian noise. Note that for sake of clarity and ease of reading, long proofs are postponed to the appendix.

2. OPTIMALITY PROPERTIES IN DETECTION PROCEDURES

In this section, we present the quickest change-point detection problems in the pointwise and minimax settings. The presentation follows the lines of the modified CUSUM procedures developed for general stochastic models in Lai (1998).

Let the time series $(X_k)_{k \geq 0}$ be specified by the conditional density function of X_k given $\mathbf{X}_0^{k-1} = (X_0, \dots, X_{k-1})$ for $k \geq 1$. The model with a disruption at time ν is given by the parametric family of conditional densities

$$f^{(\nu)}(x_k | \mathbf{x}_0^{k-1}) = f_0(x_k | \mathbf{x}_0^{k-1}) \mathbb{1}_{\{k \leq \nu\}} + f_1(x_k | \mathbf{x}_0^{k-1}) \mathbb{1}_{\{k > \nu\}}, \quad (2.1)$$

where $\mathbf{x}_0^{k-1} = (x_0, \dots, x_{k-1})$, together with the distribution of X_0 . The disruption time ν will be assumed to be a nonrandom unknown integer. We will denote by $\mathbb{P}^{(\nu)}$ the distribution of $(X_k)_{k \geq 0}$ defined by the two families of conditional densities f_0 and f_1 and the disruption time ν , and $\mathbb{E}^{(\nu)}$ will denote mean under $\mathbb{P}^{(\nu)}$.

The issue of sequential detection is to detect the change-point as soon as possible after it occurs. The considered procedures involve the system of CLLRS

$$Z_j = \log \frac{f_1(X_j | \mathbf{X}_0^{j-1})}{f_0(X_j | \mathbf{X}_0^{j-1})}, \quad j \geq 1. \quad (2.2)$$

In Lai (1998), the classical CUSUM procedure is changed into the so-called window-limited CUSUM rule. For each $0 < \alpha < 1$, let $\mathcal{M}_{\alpha, m}$ be the class of admissible stopping times τ such that $\sup_{k \geq 1} \mathbb{P}^{(0)}(k \leq \tau < k + m) \leq \alpha$.

Note that $\mathbb{P}^{(0)}$ stands for the distribution with density f_0 , that α is a preassigned upper bound for false alarm probability, and m a suitably chosen window size function of α such that

$$\lim_{\alpha \rightarrow 0} \frac{m_\alpha}{|\log \alpha|} = +\infty \quad \text{and} \quad \lim_{\alpha \rightarrow 0} \frac{\log m_\alpha}{|\log \alpha|} = 0.$$

The problem of optimal change-point detection thus becomes the determination of a procedure which minimizes in the class $\mathcal{M}_{\alpha, m}$ the positive part detection delay risk $\mathcal{R}_\nu(\tau) = \mathbb{E}^{(\nu)}(\tau - \nu)_+$, that is the mean time between the detection time τ and the true disruption time ν ; note that $(x)_+ = \max(0, x)$.

The window-limited CUSUM procedure is defined by the stopping time

$$T_\alpha = \inf \left\{ n \geq 1 : \max_{(n-m_\alpha)_+ \leq k \leq n} \sum_{j=k}^n Z_j \geq \mathbf{c}_\alpha \right\}, \quad \text{where } \mathbf{c}_\alpha = \log \frac{2m_\alpha}{\alpha}. \quad (2.3)$$

The optimization problem is studied under the following conditions on the CLLRS defined in (2.2).

Condition 1. Some $\mathbf{I} \in \mathbb{R}_+^*$ exists such that, for any $\nu \geq 0$, the following almost sure (a.s.) convergence holds:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=\nu+1}^{\nu+n} Z_j = \mathbf{I}, \quad \mathbb{P}^{(\nu)} - a.s. \quad (2.4)$$

As established in Pergamenchtchikov and Tartakovsky (2016, Theorem 5), Condition 1 implies for the delay risk the sharp lower bound

$$\liminf_{\alpha \rightarrow 0} \frac{1}{|\ln \alpha|} \inf_{\tau \in \mathcal{M}_\alpha} \mathcal{R}_\nu(\tau) \geq \frac{1}{\mathbf{I}}. \quad (2.5)$$

In order to derive the upper bound for the detection procedure (2.3), a stronger condition is required.

Condition 2. Some $\mathbf{I} \in \mathbb{R}_+^*$ exists such that, for any $\delta > 0$,

$$\lim_{n \rightarrow \infty} \sup_{k \geq \nu} \text{essup} \mathbb{P}^{(\nu)} \left(\sum_{j=k+1}^{k+n} Z_j < (\mathbf{I} - \delta)n \mid \mathcal{F}_k \right) = 0. \quad (2.6)$$

where $\mathcal{F}_k = \sigma\{X_1, \dots, X_k\}$.

As proven in Lai (1998, Theorem 4 (ii)), Condition 2 provides the following sharp upper bound for the average time between the true disruption time ν and the detection time T_α ,

$$\limsup_{\alpha \rightarrow 0} \frac{1}{|\ln \alpha|} \sup_{\nu \geq 0} \mathcal{R}_\nu(T_\alpha) \leq \frac{1}{\mathbf{I}}. \quad (2.7)$$

3. KLD APPROACH TO THE LIMIT OF THE NORMALIZED CLRS

Here naturally arise questions about the assumptions to be fulfilled by a time series to ensure the validity of Conditions 1 and 2, and in which way the limit \mathbf{I} can be evaluated from observations. This section aims at providing conditions of existence and closed-form formulae for the case of Markov chains, and more particularly for autoregressive models. We will mainly use the geometric ergodicity properties of Markov chains together with concepts of entropy and information theory; see Gray (2011) and Girardin (2005). Even though autoregressive models constitute a subclass of Markov chains, specific methods and tools have to be developed in this case; see Meyn and Tweedie (1993) and Galtchouk and Pergamenchtchikov (2014).

Section 3.1 will develop an approach based on information theory for investigating the asymptotic convergence of the normalized CLLRS. Condition 1 will be shown to amount to a particular case of the classical AEP – also called Shannon-McMillan-Breiman theorem, or ergodic theorem of information theory; see Cover and Thomas (1991). Thus, for Markov chains of any fixed order, the limit \mathbf{I} will appear as the KLD rate between two random sequences with the conditional distributions before and after the disruption. In Section 3.2, the limit will be shown to take the form of an explicit function of the model parameters. In Section 3.3, for autoregressive models, assumptions on the existence of the limit \mathbf{I} will be specifically checked, yielding a closed-form formula for \mathbf{I} too. The change-point detection procedure will then be studied in full details, for Sparre Andersen models in Section 4 and autoregressive models in Section 5.

3.1. The Ergodic Theorem of Information Theory

Let us first recall some basics concerning entropy and divergence for random sequences. Entropy measures randomness or uncertainty in a random phenomenon. Its introduction into probability theory in Shannon (1948) gave rise to information theory, by measuring the information content of sources. The Shannon (or differential) entropy of a random vector \mathbf{X}_0^{n-1} is $\mathbb{S}(\mathbf{X}_0^{n-1}) = \mathbb{S}(f) = - \int_{\mathbb{R}^n} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}$, if its distribution is absolutely continuous with respect to the Lebesgue measure Λ , with density f . An extension of the definition leads to consider the relative entropy

$$\mathbb{S}_g(f) = - \int_{\mathbb{R}^n} \left[\log \frac{f(\mathbf{x})}{g(\mathbf{x})} \right] f(\mathbf{x}) d\mathbf{x},$$

if the support of f is included in the support of g , otherwise set as $+\infty$. This is the entropy of a measure with density f relative to a measure with density g – both absolutely continuous with respect to Λ . When both f and g are densities of probability measures on \mathbb{R}^n , the Kullback-Leibler divergence (KLD) of f with respect to g , also called neg-entropy and introduced in Kullback and Leibler (1951), is the opposite number of the entropy of f relative to g , precisely

$$\mathbb{K}(f|g) = \int_{\mathbb{R}^n} \left[\log \frac{f(\mathbf{x})}{g(\mathbf{x})} \right] f(\mathbf{x}) d\mathbf{x} = -\mathbb{S}_g(f).$$

Thus, the entropy of a random vector \mathbf{X}_0^{n-1} with density f relative to a measure with density g with respect to Λ on \mathbb{R}^n is defined as $\mathbb{S}_g(\mathbf{X}_0^{n-1}) = \mathbb{S}_g(f)$.

Similarly, the KLD of a random vector \mathbf{X}_0^{n-1} with density f with respect to a random vector \mathbf{Y}_0^{n-1} with density g on \mathbb{R}^n is defined as $\mathbb{K}(\mathbf{X}_0^{n-1}|\mathbf{Y}_0^{n-1}) = \mathbb{K}(f|g)$. The KLD measures dissimilarity between two random phenomena. Although it is not a true distance because of lack of symmetry and triangular inequality – see Cover and Thomas (1991), it constitutes a good tool for discriminating two distributions because

$$\mathbb{K}(f|g) \geq 0, \quad \text{with} \quad \mathbb{K}(f|g) = 0 \quad \text{if and only if} \quad f = g. \quad (3.1)$$

Let $f(\mathbf{x}, \mathbf{y})$ and $g(\mathbf{x}, \mathbf{y})$ denote densities (with respect to Λ) on \mathbb{R}^{m+n} , with conditional densities denoted by $f(\mathbf{y}|\mathbf{x})$ and $g(\mathbf{y}|\mathbf{x})$. The conditional KLD of f with respect to g is usually defined as the real number

$$\begin{aligned} \mathbb{K}(f(\mathbf{y}|\mathbf{x})|g(\mathbf{y}|\mathbf{x})) &= \int_{\mathbb{R}^{m+n}} f(\mathbf{x}, \mathbf{y}) \log \frac{f(\mathbf{y}|\mathbf{x})}{g(\mathbf{y}|\mathbf{x})} d\mathbf{x}d\mathbf{y} \\ &= \int_{\mathbb{R}^m} f(\mathbf{x}) \int_{\mathbb{R}^n} f(\mathbf{y}|\mathbf{x}) \log \frac{f(\mathbf{y}|\mathbf{x})}{g(\mathbf{y}|\mathbf{x})} d\mathbf{y}d\mathbf{x} = \mathbb{E}_f[\mathbb{K}(f, g; \mathbf{x})], \end{aligned} \quad (3.2)$$

where the function of \mathbf{x}

$$\mathbb{K}(f, g; \mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{y}|\mathbf{x}) \log \frac{f(\mathbf{y}|\mathbf{x})}{g(\mathbf{y}|\mathbf{x})} d\mathbf{y} \quad (3.3)$$

is called local divergence (in \mathbf{x}).

The marginal Shannon entropy of order n of a sequence $\mathbf{X} = (X_n)_{n \geq 0}$ taking values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined as the entropy of its n -dimensional marginal distribution, namely

$$\mathbb{S}(\mathbf{X}_0^{n-1}) = -\mathbb{E}_f[\log f(\mathbf{X}_0^{n-1})] = - \int_{\mathbb{R}^n} f(\mathbf{x}_0^{n-1}) \log f(\mathbf{x}_0^{n-1}) d\mu(\mathbf{x}_0^{n-1}),$$

where f is the density of the random vector \mathbf{X}_0^{n-1} with respect to the n -dimensional marginal of a reference measure μ on the infinite product space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Further, the Shannon entropy rate of a process \mathbf{X} is defined by $\mathbb{H}(\mathbf{X}) = \lim_n \frac{1}{n} \mathbb{S}(\mathbf{X}_0^{n-1})$ when the limit exists; see Cover and Thomas (1991, Chapter 9) and Gray (2011, Chapter 8). The convergence in probability of $-\frac{1}{n} \log f(\mathbf{X}_0^{n-1})$ to $\mathbb{H}(\mathbf{X})$ constitutes the weak AEP while the almost sure convergence is called the strong AEP – or ergodic theorem of information theory, or Shannon-MacMillan-Breiman theorem from the authors who originally proved it.

For any ergodic random sequence \mathbf{X} whose distribution $\mathbb{P}_{\mathbf{X}}$ is absolutely continuous with respect to Λ on $\mathbb{R}^{\mathbb{N}}$, with density f , the extension of AEP to relative entropy says that $-\frac{1}{n} \log \left[\frac{f}{g}(\mathbf{X}_0^{n-1}) \right]$ converges to $\mathbb{H}_g(\mathbf{X})$, where convergence holds in probability, in mean, or a.s. with respect to $\mathbb{P}_{\mathbf{X}}$, for any reference measure absolutely continuous with respect to Λ on $\mathbb{R}^{\mathbb{N}}$, with density g ; see Barron (1985) and Girardin (2005) and the references therein for details. Note that, for sake of simplification, f or g here denote as well the density of measures on $\mathbb{R}^{\mathbb{N}}$ as the density of all their marginals on \mathbb{R}^n for $n \in \mathbb{N}^*$.

Similarly, let $\mathbf{X} = (X_n)_{n \geq 0}$ and $\mathbf{Y} = (Y_n)_{n \geq 0}$ be two ergodic random sequences whose distributions are absolutely continuous with respect to Λ on $\mathbb{R}^{\mathbb{N}}$, with respective densities f and g . Averaging per time unit leads to the definition of the KLD rate

$$\mathbb{D}(\mathbf{X} | \mathbf{Y}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{K}(\mathbf{X}_0^{n-1} | \mathbf{Y}_0^{n-1}) = -\mathbb{H}_g(\mathbf{X}),$$

when the limit exists. Extended forms of AEP have been proven to hold for large classes of ergodic processes; see Girardin (2005) for a detailed review. In particular, the strong AEP was extended to Borel state spaces independently by Barron (1985) and Orey (1985). For full details on AEP and KLD rates in the special case of Markovian sequences, see Algoet and Cover (1988, Theorem 4) and Gray (2011, Corollary 8.4.1; in short, if both X and Y are ergodic Markov chains of finite order, with respective transition densities f and g , then their KLD rate is well-defined and the following a.s. convergence holds with respect to the distribution of \mathbf{X} :

$$\frac{1}{n} \log \left[\frac{f}{g}(\mathbf{X}_0^{n-1}) \right] \longrightarrow \mathbb{D}(\mathbf{X} | \mathbf{Y}). \quad (3.4)$$

3.2. Application to the CLLRS for Markov Chains

In Section 2, the quantity \mathbf{I} of interest in both Conditions 1 and 2 is the limit of the normalized CLLRS $\frac{1}{n} \sum_{j=\nu+1}^{\nu+n} Z_j$, with Z_j defined in (2.2). As remarked in Lai (1998), for the trivial case of i.i.d. observations, \mathbf{I} is the KLD $\mathbb{K}(f_1 | f_0)$ between the post-change and pre-change distributions $f_1(x)$ and $f_0(x)$ of X_n . For Markovian dependent observations, the next result states that \mathbf{I} is the KLD rate between two well-identified sequences.

Lemma 3.1. Let the time series $\mathbf{X} = (X_n)_{n \geq 0}$ be a model with a disruption at time ν , given by the parametric family of conditional densities (2.1). Let $\widehat{\mathbf{X}}$ be a time series with distribution given by the conditional densities $f_1(x_k | \mathbf{x}_0^{k-1})$. Let $\widetilde{\mathbf{X}}$ be another time series, with conditional densities $f_0(x_k | \mathbf{x}_0^{k-1})$. Suppose that both $\widehat{\mathbf{X}}$ and $\widetilde{\mathbf{X}}$ are ergodic homogeneous Markov chains of order d .

Then Condition 1 is fulfilled, with $\mathbf{I} = \mathbb{D}(\widehat{\mathbf{X}} | \widetilde{\mathbf{X}})$.

Proof. We can write

$$\sum_{j=\nu+1}^{\nu+n} Z_j = \log \left[\prod_{j=\nu+1}^{\nu+n} \frac{f_1(X_j | \mathbf{X}_0^{j-1})}{f_0(X_j | \mathbf{X}_0^{j-1})} \right] = \log \left[\frac{\prod_{j=1}^n f_1(X_{\nu+j} | \mathbf{X}_0^{\nu+j-1})}{\prod_{j=1}^n f_0(X_{\nu+j} | \mathbf{X}_0^{\nu+j-1})} \right].$$

For any density f defining a Markov chain of order d , we compute

$$\prod_{j=1}^n f(x_{\nu+j} | \mathbf{x}_0^{\nu+j-1}) = \prod_{j=1}^n f(x_{\nu+j} | \mathbf{x}_{\nu+j-d}^{\nu+j-1}) = \prod_{j=\nu+1}^n f(x_{\nu+j} | \mathbf{x}_{\nu+1}^{\nu+j-1}) \prod_{j=1}^d f(x_{\nu+j} | \mathbf{x}_{\nu+j-d}^{\nu+j-1}).$$

For any density f and time ν , we can also write

$$\begin{aligned} f(\mathbf{x}_\nu^{\nu+n}) &= f(x_{\nu+n} | \mathbf{x}_\nu^{\nu+n-1}) f(\mathbf{x}_\nu^{\nu+n-1}) = f(x_{\nu+n} | \mathbf{x}_\nu^{\nu+n-1}) f(x_{\nu+n-1} | \mathbf{x}_\nu^{\nu+n-2}) f(\mathbf{x}_\nu^{\nu+n-2}) \\ &= \dots = \prod_{j=1}^n f(x_{\nu+j} | \mathbf{x}_\nu^{\nu+j-1}) \cdot f(x_\nu). \end{aligned}$$

Therefore, for a Markov chain of order d , we get $\prod_{j=d+1}^n f(x_{\nu+j} | \mathbf{x}_{\nu+1}^{\nu+j-1}) = f(\mathbf{x}_{\nu+1}^{\nu+n}) / f(\mathbf{x}_{\nu+1}^{\nu+d})$. Specifically

$$\sum_{j=\nu+1}^{\nu+n} Z_j = \log \left[\frac{f_1(\mathbf{X}_{\nu+1}^{\nu+n})}{f_0(\mathbf{X}_{\nu+1}^{\nu+n})} \right] + h(\nu, d),$$

where $h(\nu, d)$ is a quantity depending on ν and d but not on n , and hence $h(\nu, d)/n$ converges to 0. Therefore, $\frac{1}{n} \sum_{j=\nu+1}^{\nu+n} Z_j$ converges to the same limit as $\frac{1}{n} \log [f_1(\mathbf{X}_{\nu+1}^{\nu+n})/f_0(\mathbf{X}_{\nu+1}^{\nu+n})]$.

By the strong AEP – see Gray (2011, Corollary 8.4.1) and (3.4), the latter converges to $\mathbf{I} = \mathbb{D}(\mathbf{X}|\tilde{\mathbf{X}})$, which proves the result. \square

The first explicit formula for entropy rates was already derived in Shannon (1948) for ergodic homogeneous Markov chains of order 1 with finite state spaces. Further, let \mathbf{X} be a Markov chain taking values in \mathbb{R} , with transition density $f(x|y)$ and stationary distribution with density $\pi(x)$. Then $-\frac{1}{n} \log f(\mathbf{X}_0^{n-1})$ converges almost surely to the entropy rate $\mathbb{H}(X) = -\int_{\mathbb{R}} \pi(y) \int_{\mathbb{R}} f(x|y) \log f(x|y) dx dy$; see for instance Girardin and Linnios (2006), where the extension to KLD is also proven to hold for semi-Markov processes. Considering straightforward extension of the above formula for Markov chains of order d together with Lemma 3.1 yields the following result for a Markov chain with a disruption.

Theorem 3.1. Let \mathbf{X} be an ergodic real Markov chain of order d with disruption time ν , with transition density

$$f^{(\nu)}(x_n|\mathbf{x}_0^{n-1}) = f_0(x_n|\mathbf{x}_{n-d}^{n-1})\mathbb{1}_{\{k \leq \nu\}} + f_1(x_n|\mathbf{x}_{n-d}^{n-1})\mathbb{1}_{\{k > \nu\}}, \quad n \geq d.$$

Let $\pi_1(x)$ denote the density of the stationary distribution for the Markov chain \hat{X} of order d , with transition density $f_1(x|\mathbf{y}_0^{d-1})$. Let similarly \tilde{X} be an ergodic homogeneous Markov chain of order d , with transition density $f_0(x|\mathbf{y}_0^{d-1})$. Then Condition 1 holds true, with

$$\mathbf{I} = \mathbb{D}(\hat{X} | \tilde{X}) = \int_{\mathbb{R}^d} \pi_1(\mathbf{y}_0^{d-1}) \left[\int_{\mathbb{R}} f_1(x|\mathbf{y}_0^{d-1}) \log \frac{f_1(x|\mathbf{y}_0^{d-1})}{f_0(x|\mathbf{y}_0^{d-1})} dx \right] dy_0 \dots dy_{d-1}. \quad (3.5)$$

Note that in Fuh (2003) a KLD number between Markov transition kernels is defined for proving optimality properties of CUSUM procedures in hidden Markov models of order 1. In particular, a formula similar to (3.5) is given for a Markov chain of order 1, but no link to the AEP and KLD rates is highlighted.

3.3. Autoregressive Models

For a scalar autoregressive sequence of order p with a disruption, we will provide an explicit expression for \mathbf{I} in terms of the Fisher matrix of the model. In this aim, we will consider the associated vector valued autoregressive sequence of dimension p , that is a Markov chain of order 1.

Precisely, let $X = (X_k)_{k \geq 1}$ be a random sequence governed by the autoregressive recursive equations of order p

$$X_k = \alpha_{1,k} X_{k-1} + \dots + \alpha_{p,k} X_{k-p} + \varepsilon_k \quad k \geq 1, \quad (3.6)$$

where $\alpha_{i,k} = a_i$ before the change-point time ν (for $k \leq \nu$), and $\alpha_{i,k} = \theta_i$ after ν , that is to say for $k > \nu$. Here $(\varepsilon_k)_{k \geq 1}$ is Gaussian white noise – an i.i.d. standard Gaussian sequence, and the initial vector (X_{1-p}, \dots, X_0) is independent of $(\varepsilon_k)_{k \geq 1}$.

Setting $\Phi_k = (X_k, \dots, X_{k-p+1})'$, $a = (a_1, \dots, a_p)'$ and $\theta = (\theta_1, \dots, \theta_p)'$, the model (3.6) becomes

$$X_k = \Phi'_{k-1} (a \mathbb{1}_{\{k \leq \nu\}} + \theta \mathbb{1}_{\{k > \nu\}}) + \varepsilon_k, \quad k \geq 1. \quad (3.7)$$

In this case the conditional densities (2.1) take the form

$$f_0(x_k|\mathbf{x}_1^{k-1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_k - \Phi'_{k-1} a)^2}{2}} \quad \text{and} \quad f_1(x_k|\mathbf{x}_1^{k-1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_k - \Phi'_{k-1} \theta)^2}{2}}.$$

Since the CLLRS (2.2) can be written

$$Z_k = X_k \Phi'_{k-1} (\theta - a) - \frac{1}{2} [(\Phi'_{k-1} \theta)^2 - (\Phi'_{k-1} a)^2], \quad (3.8)$$

the ergodic properties of $\Phi = (\Phi_k)_{k \geq \nu}$ will induce a closed-form formula for \mathbf{I} , involving the Fisher matrix of the model. Indeed, the vector valued process Φ is a Markov chain of order 1, and is governed after the change-point by the multivariate autoregressive equation of order 1,

$$\Phi_k = A\Phi_{k-1} + \xi_k \quad k \geq \nu + 1, \quad (3.9)$$

where $\xi_k = (\varepsilon_k, 0, \dots, 0)' \in \mathbb{R}^p$ and

$$A = A(\theta) = \begin{pmatrix} \theta_1 & \dots & \cdot & \theta_p \\ 1 & 0 & \dots & 0 \\ \cdot & \dots & \cdot & \cdot \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

Applying (3.9) repeatedly yields $\Phi_k = A^{k-\nu}\Phi_\nu + \sum_{j=\nu+1}^k A^{k-j}\xi_j$ for $k \geq \nu + 1$, and hence, for any $\mathbf{v} \in \mathbb{R}^p$,

$$\mathbb{E} (\Phi_k \Phi_k' | \Phi_\nu = \mathbf{v}) = A^{k-\nu} \mathbf{v} \mathbf{v}' (A')^{k-\nu} + \sum_{l=0}^{k-\nu-1} A^l B (A')^l, \quad (3.10)$$

where the prime denotes transpose, and $B = (b_{ij})_{1 \leq i, j \leq p}$ is the matrix with $b_{ij} = 1$ if $(i, j) = (1, 1)$ and 0 otherwise.

In the sequel we will denote by Θ_{stb} the set of all $\theta \in \mathbb{R}^p$ for which all eigenvalues of A are less than one in modulus, and say accordingly that Φ – or X – is a stable sequence. If $\theta \in \Theta_{\text{stb}}$, then Φ is ergodic, the Fisher matrix $F(\theta) = F$ of the model (3.9) is well-defined, and, for any $\mathbf{v} \in \mathbb{R}^p$,

$$\lim_{k \rightarrow \infty} \mathbb{E} (\Phi_k \Phi_k' | \Phi_\nu = \mathbf{v}) = \mathbb{E} \Phi_\infty \Phi_\infty' = \sum_{l \geq 0} A^l B (A')^l = F, \quad (3.11)$$

where $\Phi_\infty = \sum_{l \geq 1} A^{l-1} \xi_l \sim \mathcal{N}_p(0, F(\theta))$ is the ergodic distribution of Φ_k ; see for example Basseville and Nikiforovbook (1993). Moreover the following strong law of large numbers, proven in the appendix, holds true.

Theorem 3.2. If $\theta \in \Theta_{\text{stb}}$ for the model (3.6), then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=\nu+1}^{\nu+n} \Phi_j \Phi_j' = F, \quad \mathbb{P}_\nu - \text{a.s.}$$

Condition 1 follows, with explicit limit.

Theorem 3.3. Let \mathbf{X} be governed by the model (3.6). Then Condition 1 holds true for all $\theta \in \Theta_{\text{stb}}$ and any $\nu \geq 0$. Moreover, the limit in (2.4) is

$$\mathbf{I} = \mathbb{K}(f_\theta(x|\mathbf{v})|f_a(x|\mathbf{v})) = \frac{1}{2} (\theta - a)' F(\theta) (\theta - a). \quad (3.12)$$

with the KLD defined in (3.2) for the conditional densities defined in (3.8).

Proof. Equation (3.9) implies that

$$\sum_{j=\nu+1}^{\nu+n} Z_j(\theta) = (\theta - \alpha)' M_n + \frac{1}{2} (\theta - a)' \sum_{j=\nu}^{\nu+n-1} \Phi_j \Phi_j' (\theta - a), \quad (3.13)$$

where $M_n = \sum_{j=\nu+1}^{\nu+n} \Phi_{j-1} \varepsilon_j$. The Chebyshev inequality easily induces that for any $\delta > 0$

$$\sum_{n \geq 1} \mathbb{P}(|M_n| > \delta n) \leq \delta^{-4} \sup_{k \geq 1} \frac{\mathbb{E} M_k^4}{k^2} \sum_{n \geq 1} \frac{1}{n^2}.$$

Since the process $(M_n)_{n \geq 1}$ is a martingale, using successively the Burkholder-Gundy inequality and the Cauchy-Schwarz inequality proves that some constant $\mathbf{C}^* > 0$ exists such that

$$\mathbb{E}(M_n^4) \leq \mathbf{C}^* \mathbb{E} \left(\sum_{j=\nu+1}^{\nu+n} |\Phi_{j-1}|^2 \varepsilon_j^2 \right)^2 \leq \mathbf{C}^* n \sum_{j=\nu+1}^{\nu+n} \mathbb{E}(|\Phi_{j-1}|^4 \varepsilon_j^4) = \mathbf{C}^* \mathbb{E} \left(\varepsilon_1^4 n \sum_{j=\nu+1}^{\nu+n} \mathbb{E} |\Phi_{j-1}|^4 \right).$$

From (3.11), it is easy to deduce that $\sup_{k \geq \nu+1} \mathbb{E} |\Phi_k|^4$ is finite, where $|x|$ denotes the Euclidean norm of the vector $x \in \mathbb{R}^p$. Therefore,

$$\lim_{n \rightarrow \infty} \frac{|\Phi_n|^2}{n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{n} M_n = 0, \quad \mathbb{P}_\nu - \text{a.s.}, \quad (3.14)$$

where the second limit follows from $\sup_{n \geq 1} (\mathbb{E} M_n^4)/n^2 < \infty$, thanks to the Borel-Cantelli lemma.

Equation (2.4) follows directly from Theorem 3.2, and the convergence is proven. Finally, (3.12) is given in Example 6.5 of Pergamenchtchikov and Tartakovsky (2016), with the link to KLD detailed for an autoregressive model of order 1 in Example 6.2. \square

4. SPARRE ANDERSEN MODELS

In this section we study the quickest detection problem for a specific case of the discrete time Sparre Andersen models that is widely used in insurance; see for example Asmussen and Albrecher (2010, p. 427) and Back et al. (2003). We will show that, after the disruption, this observation process is a geometric ergodic homogeneous Markov chain under some technical assumptions. We will explicitly compute its stationary distribution, and finally obtain a closed form expression for the limit of the normalized CLLRS through specific means.

Let $(J_k)_{k \geq 0}$ be an autoregressive process of order 1 with disruption time ν , governed by the recursive equation

$$J_k = (a_0 \mathbf{1}_{\{k \leq \nu\}} + a_1 \mathbf{1}_{\{k > \nu\}}) J_{k-1} + \varepsilon_k, \quad k \geq 1, \quad (4.1)$$

where the real parameters $a_0 \neq a_1$ are known and the initial variable J_0 is independent of the Gaussian white noise $(\varepsilon_k)_{k \geq 1}$. Let $N_k = \sum_{m \geq 1} \mathbf{1}_{\{S_m \leq k\}}$, where $N_0 = 0$ and the S_m are sums of m i.i.d. exponential random variables with parameter $\lambda > 0$. The renewal process $(N_k)_{k \geq 0}$ and the sequence $(J_k)_{k \geq 0}$ are independent. The process is assumed to be observed only through $\tilde{J}_k = J_{N_k}$ for $k \geq 0$.

The problem is to construct a detection rule for ν on the basis of the observations

$$X_k = \tilde{J}_{\sigma_k}, \quad k \geq 0, \quad (4.2)$$

where $\sigma_0 = 0$ and $\sigma_k = \inf\{l \geq \sigma_{k-1} + 1 : N_l - N_{\sigma_{k-1}} > 0\}$ for $k \geq 1$. Note that $(\sigma_k)_{k \geq 0}$ is an increasing sequence of stopping times with respect to the filtration $(\sigma(N_k, k \leq n))_{n \geq 0}$. The shift from the sequence (\tilde{J}_k) to the subsequence (\tilde{J}_{σ_k}) is necessary because, in the case of equal values of N_k , the corresponding values of \tilde{J}_k coincide, and hence carry no information on the disruption; therefore such \tilde{J}_k must be excluded from the detection rule.

For this construction, we need to study the observation process (4.2) after ν , that is $(X_k)_{k > \nu}$. Applying (4.1) repeatedly for $k > \nu$, we get $J_{m+k} = a_1^k J_m + \sum_{j=1}^k a_1^{k-j} \varepsilon_{j+m}$, and hence $(X_k)_{k > \nu}$ is governed by the recursive equation

$$X_k = \theta_k X_{k-1} + \eta_k, \quad k \geq \nu + 1, \quad (4.3)$$

where $\eta_k = \sum_{j=1}^{s_k} a_1^{s_k-j} \varepsilon_{j+\sigma_{k-1}}$, $s_k = \sigma_k - \sigma_{k-1}$ and $\theta_k = a_1^{s_k}$. The associated conditional densities are

$$f(x, y; a_i) = f_i(y|x) = \sum_{k \geq 1} \gamma_k \phi_{i,k}(y, x), \quad i = 0, 1, \quad (4.4)$$

where $\gamma_k = (1 - e^{-\lambda})e^{-\lambda(k-1)}$ and

$$\phi_{i,k}(y, x) = \frac{1}{\varsigma_{i,k} \sqrt{2\pi}} e^{-\frac{(y-a_i^k x)^2}{2\varsigma_{i,k}^2}} \quad \text{with} \quad \varsigma_{i,k}^2 = \frac{1 - a_i^{2k}}{1 - a_i^2}. \quad (4.5)$$

Now let us compute the transition probability function for the random sequence governed by (4.3).

Theorem 4.1. The sequence $(X_k)_{k \geq \nu}$ is a homogeneous Markov chain of order 1, with transition function given for any Borel set $B \subset \mathbb{R}$ by

$$P(x, B) = \int_B f_1(y|x) dy, \quad (4.6)$$

where f_1 is given by (4.4) for $i = 1$.

Proof. First, let us check that (s_k) is an i.i.d. sequence. We compute

$$\begin{aligned} \mathbb{P}(s_1 = m_1, \dots, s_j = m_j) &= \mathbb{E} [\mathbb{P}(s_1 = m_1, \dots, s_j = m_j | \mathcal{F}_{\sigma_{j-1}})] \\ &= \mathbb{E} [\mathbf{1}_{(s_1=m_1)} \cdots \mathbf{1}_{(s_{j-1}=m_{j-1})}] \mathbb{P}(s_j = m_j | \mathcal{F}_{\sigma_{j-1}}), \end{aligned}$$

where $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and $\mathcal{F}_k = \sigma\{X_1, \dots, X_k\}$ for $k \geq 1$. It follows from the definition of σ_j that

$$\begin{aligned} \mathbb{P}(s_j = m_j | \mathcal{F}_{\sigma_{j-1}}) &= \mathbb{P}(N_{\sigma_{j-1}} = \dots = N_{\sigma_{j-1}+m_j-1}, N_{\sigma_{j-1}+m_j} > N_{\sigma_{j-1}} | \mathcal{F}_{\sigma_{j-1}}) \\ &= (1 - e^{-\lambda})e^{-\lambda(m_j-1)}. \end{aligned}$$

Thus, X is a homogeneous Markov chain with transition probability function given by

$$\begin{aligned} \mathbb{P}(x, B) &= \mathbb{P}(a_1^{s_1} x + \sum_{j=1}^{t_1} a_1^{s_1-j} \varepsilon_{j+1}) \\ &= \sum_{k \geq 1} \mathbb{P}(a_1^k x + \sum_{j=1}^k a_1^{k-j} \varepsilon_{j+1} \in B) \mathbb{P}(s_1 = k) = \sum_{k \geq 1} \gamma_k \int_B \phi_{1,k}(y, x) dy, \end{aligned}$$

and the desired result follows. \square

Lemma 4.1. The Markov chain X is geometrically ergodic if and only if $|a_1| < 1$.

Proof. Let us check the assumptions of Theorem A.1 from the appendix. Let V be defined by

$$V(x) = \mathbf{c}_*(1 + x^2), \quad (4.7)$$

for some parameter $\mathbf{c}_* \geq 1$. Then

$$\mathbb{E}_x[V(X_1)] = \mathbf{c}_* [1 + \mathbb{E}_x(X_1^2)] = \mathbf{c}_* (1 + \mathbb{E}_x[(\theta_1 x + \eta_1)^2]) = \mathbf{c}_* [1 + x^2 \mathbb{E}(\theta_1^2) + \mathbb{E}(\eta_1^2)].$$

It is easy to check that $\mathbb{E}(\eta_1^2) = (1 - a^2 e^{-\lambda})^{-1}$ and

$$\mathbb{E}(\theta_1^2) = \sum_{m \geq 1} a_1^{2m} \mathbb{P}(s_1 = m) = \frac{(1 - e^{-\lambda}) a_1^2}{1 - a_1^2 e^{-\lambda}} := \tilde{a}_1 < 1.$$

Therefore, for any $\tilde{a}_1 < \rho < 1$,

$$\mathbb{E}_x[V(X_1)] = \mathbf{c}_* \left(\frac{a_1^2}{e^\lambda - a_1^2} + \tilde{a}_1 x^2 \right) \leq \tilde{\mathbf{c}}_* + \tilde{a}_1 V(x) \leq \rho V(x), \quad x \in C,$$

where $\tilde{\mathbf{c}}_* = \mathbf{c}_*(e^\lambda - a_1^2)^{-1}a_1^2$ and

$$C = \left\{ x \in \mathbb{R} : |x| > \sqrt{\frac{\tilde{\mathbf{c}}_*}{\mathbf{c}_*(\rho - \tilde{a}_1)} - 1} \right\}.$$

Consequently, both Conditions \mathbf{D}_1 and \mathbf{D}_2 of Theorem A.1 hold for the set C . Applying this theorem yields that X is geometrically ergodic. \square

Lemma 4.2. If $|a_1| < 1$, then the stationary distribution of the Markov chain $(X_k)_{k>\nu}$ is centered Gaussian with variance $(1 - a_1^2)^{-1}$.

Proof. Let $\mathcal{S} = \sigma\{N_t, t \geq 0\}$ be the σ -algebra generated by the renewal counting process $(N_t)_{t \geq 0}$. We get by induction from (4.3) that

$$X_n = \left(\prod_{j=\nu+1}^n \theta_j \right) X_\nu + \sum_{k=\nu+1}^n \eta_k \prod_{j=k+1}^n \theta_j, \quad n > \nu, \quad (4.8)$$

where an empty product is set to 1 by convention, and η_k is \mathcal{S} -conditionally centered Gaussian distributed with variance

$$\Delta_k = \sum_{j=1}^{s_k} a_1^{2(s_k-j)} = \frac{1 - a_1^{2s_k}}{1 - a_1^2}.$$

Since $(\eta_n)_{n>\nu}$ is a sequence of \mathcal{S} -conditionally independent random variables, the distribution of X_n for $n > \nu$ is \mathcal{S} -conditionally Gaussian distributed, with mean

$$\mathbf{m}_n = \mathbb{E}(X_n | \mathcal{S}) = \prod_{j=\nu+1}^n \theta_j \mathbb{E}(X_\nu | \mathcal{S}) = a_1^{2(\sigma_n - \sigma_\nu)} \mathbb{E}(X_\nu | \mathcal{S}),$$

and variance

$$\mathbf{d}_n = \mathbb{E} [X_n - \mathbb{E}(X_n | \mathcal{S})^2 | \mathcal{S}] = \sum_{k=\nu+1}^n \Delta_k \prod_{j=k+1}^n \theta_j = \frac{1 - a_1^{2\sigma_n - \sigma_\nu}}{1 - a_1^2}. \quad (4.9)$$

Moreover \mathbf{m}_n converges a.s. to 0 and \mathbf{d}_n to $(1 - a_1^2)^{-1}$ when n tends to infinity. Thus, both the conditional distribution of X_n and its ordinary one are asymptotically centred Gaussian distributed with variance $(1 - a_1^2)^{-1}$. Therefore the candidate for the stationary distribution of the process is the probability measure defined by

$$\pi(B) = \sqrt{\frac{1 - a_1^2}{2\pi}} \int_A e^{-(1-a_1^2)u^2/2} du, \quad B \in \mathcal{B}(\mathbb{R}). \quad (4.10)$$

Finally, it is straightforward to show that $\pi(B) = \int_{\mathbb{R}} P(x, B)\pi(dx)$ for any Borel set $B \in \mathcal{B}(\mathbb{R})$, where $P(x, B)$ is given in (4.6). \square

Thanks to (4.6), the conditional densities in (2.1) take the form (4.4), or $f_i(x_j | \mathbf{x}_0^{j-1}) = f(x_{j-1}, x_j; a_i)$, for $i = 0, 1$, and the CLLRS in (2.2) takes the form

$$Z_j = g(X_{j-1}, X_j), \quad \text{where } g(x, y) = \log \frac{f(x, y; a_1)}{f(x, y; a_0)}. \quad (4.11)$$

The following properties of the conditional KLD are necessary for obtaining a closed-form formula for \mathbf{I} . The proof is postponed to the appendix.

Lemma 4.3. Let $f_1(y|x)$ and $f_0(y|x)$ be given by (4.4).

1. If $a_1 \neq a_0$, then the local divergence $\mathbb{K}(f_1, f_0; x)$ defined by (3.3) is positive for all $x \in \mathbb{R}$.
2. Let V be defined by (4.7). If $\mathbf{c}_* \geq 6(1-a_1^2)^{-1} + 2 \log \sqrt{2\pi} - 2 \log(1-e^{-\lambda})$, then $\mathbb{K}(f_1, f_0; x) \leq V(x)$ for all $x \in \mathbb{R}$.

It follows from Lemma 4.3 that the local divergence satisfies

$$\mathbb{K}(f_1, f_0; x) \leq 6(1-a_1^2)^{-1} + 2 \log \sqrt{2\pi} - 2 \log(1-e^{-\lambda}).$$

Condition 1 can then be deduced from Theorem 3.1 and a closed-form expression for the limit derives from (3.5). Still, the following given proof is specific to the case under study.

Theorem 4.2. Let \mathbf{X} be a Sparre-Anderson model governed by (4.3). Then for any $\nu \geq 0$, Condition 1 is satisfied, with

$$\mathbf{I} = \int_{\mathbb{R}} \pi(x) \mathbb{K}(f_1(y|x) | f_0(y|x)) dx = \sqrt{\frac{1-a_1^2}{2\pi}} \int_{\mathbb{R}} \mathbb{K}(f_1, f_0; x) e^{-(1-a_1^2)x^2/2} dx, \quad \mathbb{P}^{(\nu)} - \text{a.s.}$$

where the conditional densities are given in (4.4).

Proof. We can write

$$\sum_{j=\nu+1}^{\nu+n} Z_j = \sum_{j=\nu+1}^{\nu+n} U_j + \sum_{j=\nu+1}^{\nu+n} \tilde{g}(X_{j-1}) = A_n + B_n,$$

where $U_j = g(X_{j-1}, X_j) - \tilde{g}(X_{j-1})$ and $\tilde{g}(x) = E_x[g(x, X_1)] = \int_{\mathbb{R}} g(x, y) f_1(y|x) dy$. Let us show that, $\mathbb{P}^{(\nu)}$ -a.s., B_n/n converges to $\mathbb{K}(f_1|f_0)$ and A_n/n converges to 0 as n tends to infinity.

First, applying the strong law of large numbers – see Meyn and Tweedie (1993, Theorem 17.1.7) – to the series $B_n = \sum_{j=\nu+1}^{\nu+n} \tilde{g}(X_{j-1})$, for $n \geq 1$, yields

$$\lim_{n \rightarrow \infty} \frac{1}{n} B_n = \int_{\mathbb{R}} \tilde{g}(y) d\pi(y) = \mathbb{K}(f_1|f_0) \quad \mathbb{P}^{(\nu)} - \text{a.s.}$$

Second, $(U_j)_{j>\nu}$ is a martingale-difference with respect to the natural filtration $\mathcal{F}_j = \sigma(\mathbf{X}_0^j)$ for $j > \nu$. Therefore A_n/n will converge to 0 if

$$\sum_{j>\nu} \frac{1}{j^2} \mathbb{E}(U_j^2 | \mathcal{F}_{j-1}) < \infty \quad \mathbb{P}^{(\nu)} - \text{a.s.} \quad (4.12)$$

Thanks to the bound (A.5), we get

$$\begin{aligned} \mathbb{E}(U_j^2 | \mathcal{F}_{j-1}) &\leq \mathbb{E}[g^2(X_{j-1}, X_j) | \mathcal{F}_{j-1}] \leq \mathbb{E}[(\mathbf{u}^* + X_{j-1}^2 + X_j^2)^2 | \mathcal{F}_{j-1}] \\ &\leq 3[(\mathbf{u}^*)^2 + X_{j-1}^4 + \mathbb{E}(X_j^4 | \mathcal{F}_{j-1})]. \end{aligned}$$

Thanks to the latter inequality, it is sufficient for proving (4.12) to show that

$$\sup_{n>\nu} \mathbb{E}(X_n^4 | \mathcal{F}_\nu) < \infty. \quad (4.13)$$

First (4.8) gives $X_n^4 \leq 8X_\nu^4 \prod_{j=\nu+1}^n \theta_j^4 + 8\tilde{X}_n^4 \leq 8X_\nu^4 + 8Y_n^4$, where $Y_n = \sum_{k=\nu+1}^n \eta_k \prod_{j=k+1}^n \theta_j$. So, proving (4.13) amounts to proving that $\sup_{n>\nu} \mathbb{E}(Y_n^4 | \mathcal{F}_\nu) < \infty$.

For $n > \nu$, applying the Hölder inequality yields

$$Y_n^4 \leq \left[\sum_{k=\nu+1}^n \left(|\eta_k| \prod_{j=k+1}^n |\theta_j|^{1/2} \right) \left(\prod_{j=k+1}^n |\theta_j|^{1/2} \right) \right]^4 \leq c_n \sum_{k=\nu+1}^n \eta_k^4 \prod_{j=k+1}^n \theta_j^2,$$

where $c_n = \left(\sum_{k=\nu+1}^n \prod_{j=k+1}^n |\theta_j|^{2/3} \right)^3$. Therefore, thanks to (4.9),

$$\mathbb{E} (Y_n^4 | \mathcal{S}) \leq 3c_n \sum_{k=\nu+1}^n \Delta_k^2 \prod_{j=k+1}^n \theta_j^2 \leq \frac{3c_n}{1-a_1^2} \sum_{k=1}^n \Delta_k \prod_{j=k+1}^n \theta_j^2 \leq \frac{3c_n}{(1-a_1^2)^2}.$$

Since $\sigma_j - \sigma_{j-1} \geq 1$ in (4.2), we get $c_n \leq \left(\sum_{k \geq 1} |a_1|^{2k/3} \right)^3 < \infty$, and (4.13) is proven. Finally, the series (4.12) is convergent, and hence A_n/n converges \mathbb{P}_ν -a.s. to 0. \square

5. OPTIMALITY PROPERTIES FOR AUTOREGRESSIVE MODELS

We will here prove the optimality of the window-limited CUSUM procedure (2.3) for the autoregressive model (3.6) for which both pre-change and post-change parameters are known. We will apply the detecting procedures proposed in Lai (1998) and shortly described in Section 2. In this aim, we will develop analytical tools leading to the conditions that provide the lower and upper bounds for the detection delay.

First the detection delay has an asymptotic upper bound, because Condition 2 is satisfied, as stated in the following proposition, whose proof is given in the appendix.

Theorem 5.1. If the autoregressive model (3.6) is stable – that is for $\theta \in \Theta_{\text{stab}}$, then Condition 2 is satisfied.

In Lai (1998, Theorem 4 (ii)), the upper bound (2.7) is proven to hold true provided that the system of statistics (3.8) satisfies (2.6). Therefore, the lower bound (2.5) together with the upper bound (2.7) imply the minimax property of the CUSUM procedure, as stated below.

Theorem 5.2. If the autoregressive model (3.6) is stable, then the detection delay risk of the window-limited procedure (2.3) based on the statistics (3.8) is upper bounded as in (2.7), with \mathbf{I} given by (3.12), and the procedure is asymptotically minimax with

$$\lim_{\alpha \rightarrow 0} \frac{\inf_{\tau \in \mathcal{M}_\alpha} \sup_{\nu \geq 0} \mathcal{R}_\nu(\tau)}{\sup_{\nu \geq 0} \mathcal{R}_\nu(T_\alpha)} = 1.$$

Moreover, this procedure is pointwise optimal. In mathematical words, for any $\nu \geq 0$,

$$\lim_{\alpha \rightarrow 0} \frac{\inf_{\tau \in \mathcal{M}_\alpha} \mathcal{R}_\nu(\tau)}{\mathcal{R}_\nu(T_\alpha)} = 1.$$

ACKNOWLEDGEMENTS

This work is partly supported by the French Federation Normandie Mathématiques, by the RSF grant 14-49-00079 (National Russian Research University "MPEI" 14 Krasnokazarmennaya, 111250 Moscow, Russia), and by the RFBR Grant 16-01-00121 A.

REFERENCES

- Algoet, P. and Cover, T. (1988). A Sandwich Proof of the Shannon-McMillan- Breiman Theorem, *Annals of Probability* 16: 899–909.
- Asmussen, S. and Albrecher, H. (2010). *Ruin Probabilities*, 2d edition, Singapore: World Scientific.
- Back, K., Bielecki, T.R., Hipp, C., Peng, Sh. and Schachermayer, W. (2003). *Stochastic Methods in Finance, C.I.M.E.-E.M.S. Summer School*, July 6–12, Bressanone/Brixen, Italy.
- Barron, A. (1985). The strong Ergodic Theorem for Densities: Generalized Shannon-McMillan-Breiman Theorem, *Annals of Probability*, 13: 1292–1303.
- Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Applications*, Englewood Cliffs: Prentice Hall.
- Cover, L. and Thomas, J. (1991). *Elements of Information Theory*, New York: Wiley series in telecommunications.
- Feigin, P.D. and Tweedie, R. (1985). Random Coefficient Autoregressive Processes: a Markov chain Analysis of Stationarity and Finiteness of Moments, *Journal of Time Series Analysis* 6: 1-14.
- Fuh, C.-D. (2003). SPRT and CUSUM in hidden Markov models, *Annals of Statistics* 31: 942-977.
- Galtchouk, L. and Pergamenchtchikov, S. (2014). Geometric Ergodicity for Classes of Homogeneous Markov Chains, *Stochastic Processes and Applications* 124, 3362–3391.
- Girardin, V. (2005). On the Different Extensions of the Ergodic Theorem of Information Theory, *Recent Advances in Applied Probability* R. Baeza-ates, J. Glaz, H. Gzyl, J. Hüsler and J.L. Palacios, eds., pp. 163 – 179, San Francisco: Springer-Verlag.
- Girardin, V. and Limnios, N. (2006). Entropy for Semi-Markov Processes with Borel State Spaces: Asymptotic Equipartition Properties and Invariance Principles, *Bernoulli* 12: 1–19.
- Gray, R. M. (2011). *Entropy and Information Theory*, 2d edition, New York: Springer.
- Kent, S. (2000). On the Trial of Intrusions into Information Systems, *IEEE Spectrum*, 37: 52–56.
- Konev, V. V. and Pergamenchtchikov, S. M. (1981). The Sequential Plans of Parameters Identification in Dynamic Systems, *Automation and Remote Control* 7: 84–92.
- Konev, V. V. and Pergamenchtchikov, S. M. (1990). Truncated Sequential Estimation of the Parameters in a Random Regression, *Sequential Analysis* 9: 19–41.
- Kullback S. and Leibler R. A. (1951). On Information and Sufficiency, *Annals of Mathematical Statistics* 22: 49–86.
- Lai, T.L. (1998) Informations Bounds and Quick Detection of Parameters Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917–2929
- Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897-1908.
- Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability*, New York: Springer Verlag.
- Orey, S. (1985) On the Shannon-Perez-Moy Theorem, *Contemporary Mathematics* 41: 319–27.
- Page, E.S. (1954). Continuous Inspection Schemes, *Biometrika* 41: 100–115.
- Pergamenchtchikov, S. M. and Tartakovsky, A.G. (2016). Asymptotically Optimal Pointwise and Minimax quickest Change-Point Detection for Dependent Data, *Statistical inference for stochastic processes*, DOI: 10.1007/s11203-016-9149-x
- Shannon, C. (1948). A Mathematical Theory of Communication, *Bell Systems Technical Journal* 27: 379–423, 623-656.
- Tartakovsky, A.G. and Veeravalli, V.V. (2004). Change-Point Detection in Multichannel and Distributed Systems with Applications, *Applications of Sequential Methodologies* (N. Mukhopadhyay, S. Datta and S. Chattopadhyay, eds.), pp. 331–363, New York: Marcel Dekker, Inc..

Tartakovsky, A., Nikiforov, I. and Basseville, M. (2015). *Sequential Analysis: Hypothesis Testing and Change-point Detection*, Boca Raton: Chapman and Hall/CRC.

Yakir, B. (1994). Optimal Detection of a Change in Distribution when Observations Form a Markov Chain with a Finite State Space, *Change-Point problems*, E. Carlstein, H. Muller and D. Siegmund, eds., pp. 346–358, Hayward: Institute of Mathematical Statistics.

APPENDIX

A.1. Geometric Ergodicity for Homogeneous Markov Processes

We will follow the so-called Meyn–Tweedie approach; some definitions from Meyn and Tweedie (1993) and Galtchouk and Pergamenchtchikov (2014) are necessary.

For a homogeneous Markov chain $(X_n)_{n \geq 0}$ with measurable state space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, the transition probability of the chain is defined by $P(x, A) = \mathbb{P}(X_1 \in A | X_0 = x)$ for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$. The n -step transition probability is $P^n(x, A) = \mathbb{P}(X_n \in A | X_0 = x)$. A measure π on $\mathcal{B}(\mathcal{X})$ is said to be invariant – or stationary, or ergodic – for the chain if $\pi(A) = \int_{\mathcal{X}} P(x, A) \pi(dx)$ for all $A \in \mathcal{B}(\mathcal{X})$. If an invariant positive measure π with $\mu(\mathcal{X}) = 1$ exists, then the chain is said to be positive.

The following conditions are minorization and drift conditions:

D₁. There exist $\delta > 0$, a set $C \in \mathcal{B}(\mathcal{X})$, and a probability measure Q on $\mathcal{B}(\mathcal{X})$ with $Q(C) = 1$, such that, for any $A \in \mathcal{B}(\mathcal{X})$ for which $Q(A) > 0$, $\inf_{x \in C} P(x, A) > \delta Q(A)$.

D₂. There exist a function $\mathbf{V} : \mathcal{X} \rightarrow [1, \infty)$, constants $0 < \rho < 1$ and $d \geq 1$, and a set $C \in \mathcal{B}(\mathcal{X})$ such that $W = \sup_{x \in C} |\mathbf{V}(x)| < \infty$ and, for all $x \in \mathcal{X}$, $\mathbb{E}_x[\mathbf{V}(X_1)] \leq (1 - \rho)\mathbf{V}(x) + d\mathbb{1}_C(x)$.

If **D₂** is satisfied, \mathbf{V} is called the Lyapunov function of the chain. Note that, obviously, Condition **D₁** implies that $\eta = \inf_{x \in C} P(x, C) - \delta > 0$.

The following theorem is used in the proof of Lemma 4.1 in Section 4; see Feigin and Tweedie (1985), and also Galtchouk and Pergamenchtchikov (2014).

Theorem A.1. Let $(X_n)_{n \geq 0}$ be a homogeneous Markov chain satisfying conditions **D₁** and **D₂** with the same set $C \in \mathcal{B}(\mathcal{X})$. Then $(X_n)_{n \geq 0}$ is a positive geometric ergodic process, i.e.,

$$\sup_{n \geq 0} e^{\kappa^* n} \sup_{x \in \mathcal{X}} \sup_{0 \leq g \leq \mathbf{V}} \frac{1}{\mathbf{V}(x)} |\mathbb{E}_x g(X_n) - \pi(g)| \leq R^*.$$

Note that the positive constants κ^* and R^* are determined in Galtchouk and Pergamenchtchikov (2014).

A.2. Proof of Theorem 3.2

Setting $\Psi_n = \sum_{j=\nu+1}^{\nu+n} \Phi_j \Phi_j'$, we obtain from (3.9) that

$$\Psi_n = A \Psi_n A' + A \left(\Phi_\nu \Phi_\nu' - \Phi_{\nu+n} \Phi_{\nu+n}' \right) A' + A \widetilde{M}_n + \widetilde{M}_n' A' + \sum_{j=\nu+1}^{\nu+n} \xi_j \xi_j',$$

where $\widetilde{M}_n = \sum_{j=\nu+1}^{\nu+n} \Phi_{j-1} \xi_j'$ is the $p \times p$ matrix whose first column is the vector M_n defined in (3.13) and all other columns are null. So, if we set

$$D_n = A \left(\Phi_\nu \Phi_\nu' - \Phi_{\nu+n} \Phi_{\nu+n}' \right) A' + A \widetilde{M}_n + \widetilde{M}_n' A' + \sum_{j=\nu+1}^{\nu+n} \xi_j \xi_j',$$

we obtain for the matrix Ψ_n the linear equation

$$\Psi_n - A\Psi_n A' = D_n. \quad (\text{A.1})$$

Thanks to both (3.14) and the strong law of large numbers, we get that $\frac{1}{n} D_n$ converges \mathbb{P}_ν -a.s. to the matrix B in (3.10).

A vector operation $\mathbf{vec}(\cdot)$ is defined by setting $\mathbf{vec}(V) = (v_{1,1}, \dots, v_{p,1}, \dots, v_{1,p}, \dots, v_{p,p})' \in \mathbb{R}^{p^2}$, for any $p \times p$ matrix V . Thus we can write (A.1) as

$$(I - A \otimes A)\mathbf{vec}(\Psi_n) = \mathbf{vec}(D_n), \quad (\text{A.2})$$

where $U \otimes V = (u_{i,j}v_{k,l})_{1 \leq i,j,k,l \leq p}$ is the $p^2 \times p^2$ matrix called the Kronecker product of the two $p \times p$ matrices $U = (u_{i,j})_{1 \leq i,j \leq p}$ and $V = (v_{i,j})_{1 \leq i,j \leq p}$; see, for example, Feigin and Tweedie (1985, Section 3.3)

One can check directly that the eigenvalues of the matrix $A \otimes A$ are $(\lambda_i \lambda_j)_{1 \leq i,j \leq p}$, where $(\lambda_i)_{1 \leq i \leq p}$ are the eigenvalues of A . If $|\lambda_i| < 1$ for all i , then $|\lambda_i \lambda_j| < 1$ for all (i, j) too. Therefore, (A.2) has a unique solution which can be written

$$\mathbf{vec}\left(\frac{1}{n}\Psi_n\right) = (I - A \otimes A)^{-1} \mathbf{vec}\left(\frac{1}{n}D_n\right). \quad (\text{A.3})$$

Obviously, the Fisher matrix F defined in (3.11) satisfies the linear equation $F - AFA' = B$, that can be rewritten as $\mathbf{vec}(F) = (I - A \otimes A)^{-1} \mathbf{vec}(B)$. Letting n tend to infinity in (A.3) yields the desired convergence. \square

A.3. Proof of Lemma 4.3

1. Thanks to (3.1), it is sufficient to show that, for all $x \in \mathbb{R}$, some set $\Gamma(x)$ with positive Lebesgue measure exists such that $f_1(y|x) \neq f_0(y|x)$ in (4.4) for all $y \in \Gamma(x)$. Let us suppose, without loss of generality, that $a_1 > a_0$ and show that

$$\phi_{1,k}(y, x) > \phi_{0,k}(y, x), \quad k \geq 1. \quad (\text{A.4})$$

According to (4.5), the above inequality amounts to $y^2 A_{1,k} + y A_{2,k} + A_{3,k} > 0$, where

$$A_{1,k} = \frac{1}{2} \left(\frac{1}{\varsigma_{0,k}^2} - \frac{1}{\varsigma_{1,k}^2} \right), \quad A_{2,k} = x \left(\frac{a_0^k}{\varsigma_{0,k}^2} - \frac{a_1^k}{\sigma_{1,k}^2} \right), \quad \text{and} \quad A_{3,k} = \frac{x^2}{2} \left(\frac{a_0^k}{\varsigma_{0,k}^2} - \frac{a_1^k}{\varsigma_{1,k}^2} \right) - \log \frac{\varsigma_{1,k}^2}{\varsigma_{0,k}^2}.$$

Since $A_{1,k} > 0$ for all k , (A.4) will follow for example if $y > y_k''(x)$, where

$$y_k''(x) = \frac{\sqrt{|A_{2,k}^2 - 4A_{1,k}A_{3,k}|} - A_{2,k}}{2A_{1,k}}.$$

Note that, if the discriminant $A_{2,k}^2 - 4A_{1,k}A_{3,k}$ is negative, then the inequality (A.4) holds for any $y \in \mathbb{R}$. Taking into account that $\lim_{k \rightarrow +\infty} A_{1,k} = (a_1^2 - a_0^2)/2$, $\lim_{k \rightarrow \infty} A_{2,k} = 0$ and $\lim_{k \rightarrow \infty} A_{3,k} = -\log(1 - a_0^2)/(1 - a_1^2)$, we get

$$\lim_{k \rightarrow \infty} y_k''(x) = 2 \frac{\sqrt{2(a_1^2 - a_0^2) \log(1 - a_0^2)/(1 - a_1^2)}}{a_1^2 - a_0^2}.$$

Therefore, for any $x \in \mathbb{R}$, $z^*(x) = \sup_{k \geq 1} y_k''(x) < \infty$, and the interval $\Gamma(x) = (z^*(x), +\infty)$ is a suitable choice. On this set, (A.4) holds, and hence, $f_1(y|x) > f_0(y|x)$, as required. Point 1. follows.

2. Since $\varsigma_{i,1} = 1$ and $\varsigma_{i,k}^2 \geq 1$ for all $k \geq 2$, we get for all x and y ,

$$\frac{\gamma_1}{\sqrt{2\pi}} e^{-(x^2+y^2)} \leq f_i(y|x) \leq \frac{1}{\sqrt{2\pi}},$$

or $|\log f_i(y|x)| \leq \log \sqrt{2\pi} - \log(1 - e^{-\lambda}) + x^2 + y^2$. Therefore,

$$|g(x, y)| \leq |\log f_0(y|x)| + |\log f_1(y|x)| \leq \mathbf{u}^* + 2(x^2 + y^2), \quad (\text{A.5})$$

where $\mathbf{u}^* = 2 \log \sqrt{2\pi} - 2 \log(1 - e^{-\lambda})$ and g is defined in (4.11). Further,

$$\mathbb{K}(f_1|f_0; x) = \int_{\mathbb{R}} f_1(y|x)g(x, y)dy \leq \mathbf{u}^* + 2x^2 + 2 \int_{\mathbb{R}} y^2 f_1(y|x) dy. \quad (\text{A.6})$$

Elementary algebra gives

$$\int_{\mathbb{R}} y^2 f_1(y|x) dy = \sum_{k \geq 1} \frac{\gamma_k}{\varsigma_{1,k} \sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-\frac{(y-a_1^k x)^2}{2\varsigma_{1,k}^2}} dy = \sum_{k \geq 1} (\varsigma_{1,k}^2 + a_1^{2k} x^2) \gamma_k \leq (1 - a_1^2)^{-1} + x^2.$$

The above equality jointly with (A.6) yield the desired result, precisely

$$\mathbb{K}(f_1|f_0; x) \leq \mathbf{u}^* + \frac{2}{1 - a_1^2} + 4x^2 \leq V(x),$$

provided that $\mathbf{c}_* \geq \mathbf{u}^* + 6(1 - a_1^2)^{-1}$. \square

A.4. Proof of Theorem 5.1

First note that, thanks to the time homogeneity of the process $(X_j)_{j > \nu}$, for any $k \geq \nu$,

$$\text{essup } \mathbb{P}^{(\nu)} \left(\sum_{j=k+1}^{k+n} Z_j < (\mathbf{I} - \delta)n \mid \mathcal{F}_k \right) = \sup_{\mathbf{v} \in \mathbb{R}^p} \mathbb{P}^{(0)} \left(\sum_{j=1}^n Z_j < (\mathbf{I} - \delta)n \mid \Phi_0 = \mathbf{v} \right). \quad (\text{A.7})$$

where $\mathbb{P}^{(0)}$ denotes the probability measure when the disruption occurs at time 0.

In order to simplify the analysis of the right hand side conditional probability in (A.7), let us introduce, for any initial value $\mathbf{v} \in \mathbb{R}^p$, the autoregressive process

$$\tilde{\Phi}_k = \Phi_k - A^k \mathbf{v} = \sum_{l=1}^k A^{k-l} B \xi_l.$$

Note that $(\tilde{\Phi}_k)_{k \geq 1}$ satisfies (3.9) with initial value zero. Substituting X_k from (3.7) into (3.8), we get

$$Z_k = \Phi'_{k-1} \bar{\theta} \varepsilon_k + \frac{1}{2} \Phi'_{k-1} \bar{\theta} \bar{\theta}' \Phi_{k-1}, \quad (\text{A.8})$$

where $\bar{\theta} = \theta - a$. For the process $(\tilde{Z}_k)_{k \geq 1}$ with initial value zero associated to $(Z_k)_{k \geq 1}$, we have $\tilde{Z}_k = \tilde{\Phi}'_{k-1} \bar{\theta} \varepsilon_k + \frac{1}{2} \tilde{\Phi}'_{k-1} \bar{\theta} \bar{\theta}' \tilde{\Phi}_{k-1}$. Substituting $\tilde{\Phi}_{k-1}$ from (??) into (A.8) yields

$$Z_k = \tilde{Z}_k + \frac{\mathbf{v}'(A')^{k-1} \bar{\theta} \bar{\theta}' A^{k-1} \mathbf{v}}{2} + u_k \bar{\theta}' A^{k-1} \mathbf{v},$$

where $u_k = \bar{\theta}' \tilde{\Phi}_{k-1} + \varepsilon_k$. It follows that the sums $S_n = \sum_{j=1}^n Z_j$ and $\tilde{S}_n = \sum_{j=1}^n \tilde{Z}_j$ are related by $S_n = \tilde{S}_n + g(\mathbf{v})$ where

$$g(\mathbf{v}) = \frac{1}{2} \mathbf{v}' V_n \mathbf{v} + U_n' \mathbf{v}, \quad (\text{A.9})$$

with $V_n = \sum_{k=1}^n (A')^{k-1} \bar{\theta} \bar{\theta}' A^{k-1}$ and $U_n = \sum_{k=1}^n u_k (A')^{k-1} \bar{\theta}$. Moreover, the function g is bounded as stated in Lemma A.1 below.

Using Theorem 3.3 and this bound will allow us to prove that for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{v} \in \mathbb{R}^p} \mathbb{P}^{(0)} \left(\sum_{j=1}^n Z_j < (\mathbf{I} - \delta)n \mid \Phi_0 = \mathbf{v} \right) = 0.$$

Indeed, thanks to Lemma A.1 below,

$$\begin{aligned} \mathbb{P}^{(0)} (S_n < (\mathbf{I} - \delta)n \mid \Phi_0 = \mathbf{v}) &= \mathbb{P}^{(0)} \left(\frac{\tilde{S}_n}{n} + \frac{g(\mathbf{v})}{n} < \mathbf{I} - \delta \right) \\ &\leq \mathbb{P}^{(0)} \left(\frac{\tilde{S}_n}{n} - \mathbf{I} < -\delta + \frac{|\zeta_*|}{n} \right) \leq \mathbb{P}^{(0)} \left(\left| \frac{\tilde{S}_n}{n} - \mathbf{I} \right| > \delta/4 \right) + \mathbb{P}^{(0)} \left(\frac{|\zeta_*|}{n} \geq \delta/4 \right). \end{aligned}$$

Thus

$$\mathbb{P}^{(0)} \left(\sum_{j=1}^n Z_j < (\mathbf{I} - \delta)n \mid \mathbb{P}^{(0)} = \mathbf{v} \right) \leq \mathbb{P}^{(0)} \left(\left| \frac{\tilde{S}_n}{n} - \mathbf{I} \right| > \delta/4 \right) + \mathbb{P}^{(0)} \left(\frac{|\zeta_*|}{n} \geq \delta/4 \right).$$

Letting n tend to infinity and applying Theorem 3.3 yield (2.6). Thus Theorem 5.1 is proven. \square

Lemma A.1. The function g in (A.9) is lower bounded by some integrable random variable ζ_* , that is $\inf_{\mathbf{v} \in \mathbb{R}^p} g(\mathbf{v}) \geq \zeta_*$.

Proof. Since the matrix V_n is symmetric and nonnegative definite, an orthogonal matrix Q_n exists such that $Q_n V_n Q_n' = \text{diag}(\mathbf{v}_{1,n}, \dots, \mathbf{v}_{p,n})$, where $\mathbf{v}_{1,n} \geq \dots \geq \mathbf{v}_{p,n} \geq 0$ are the eigenvalues of V_n .

A non-degenerate transformation is defined onto \mathbb{R}^p by setting $t = (t_1, \dots, t_p)' = Q_n' \mathbf{v}$. Therefore, the function $g(\cdot)$ can be viewed as a function of t , precisely

$$g(\mathbf{v}) = \tilde{g}(t) = \sum_{j=1}^p \left(\frac{\mathbf{v}_{j,n} t_j^2}{2} + \beta_{j,n} t_j \right) \quad (\text{A.10})$$

with $\beta_{j,n} = \langle Q_n' U_n \rangle_j$, where $\langle x \rangle_j$ denotes the j -th component of the vector x . For any j , we can write $\mathbf{v}_{j,n} = \tilde{e}_j' V_n \tilde{e}_j$ with $\tilde{e}_j = Q_n e_j$, where $e_j = (0, \dots, 1, \dots, 0)' \in \mathbb{R}^p$. Note that $\beta_{j,n} = U_n' \tilde{e}_j$.

If $\mathbf{v}' V_n \mathbf{v} = \sum_{k=1}^n (\bar{\theta}' A^{k-1} \mathbf{v})^2 = 0$, (where $\bar{\theta} = \theta - a$ as in (A.8)), then $\bar{\theta}' A^{k-1} \mathbf{v} = 0$ for all $1 \leq k \leq n$, and hence $U_n' \mathbf{v} = \sum_{k=1}^n u_k \bar{\theta}' A^{k-1} \mathbf{v} = 0$, and finally $g(\mathbf{v})$ is null. Therefore, if $\mathbf{v}_{j,n} = 0$ for some $1 \leq j \leq p$, then $\beta_{j,n} = 0$.

Thus the function $\tilde{g}(\cdot)$ in (A.10) can be rewritten as

$$\tilde{g}(t) = \sum_{j=1}^p \left(\frac{\mathbf{v}_{j,n} t_j^2}{2} + \beta_{j,n} t_j \right) \mathbf{1}_{\{\mathbf{v}_{j,n} > 0\}}.$$

It is easy to check that $\tilde{g}(\cdot)$ is lower bounded, that is

$$\tilde{g}(t) \geq - \sum_{j=1}^p \frac{\beta_{j,n}^2}{\mathbf{v}_{j,n}} \mathbf{1}_{\{\mathbf{v}_{j,n} > 0\}}. \quad (\text{A.11})$$

Further, since $\theta \in \Theta_{\text{stb}}$, the limit of V_n exists when n tends to infinity. Precisely, V_n converges to $\sum_{l \geq 0} (A')^l \bar{\theta} \bar{\theta}' A^l$, and, for each $1 \leq j \leq p$, the sequence $(\mathbf{v}_{j,n})_{n \geq 1}$ is increasing with a finite limit.

Finally,

$$\mathbb{E} \left(\sup_{n \geq 1} |U_n| \right) \leq |\bar{\theta}| \sum_{k \geq 1} \mathbb{E} \left(|u_k| |(A')^{k-1}| \right) \leq |\bar{\theta}| \sup_{k \geq 1} \mathbb{E} (|u_k|) \sum_{k \geq 1} |(A')^{k-1}| < \infty.$$

Together with (A.11), this yields the desired result with

$$\zeta_* = - \inf_{n \geq 1} \sum_{j=1}^p \frac{\beta_{j,n}^2}{\mathbf{v}_{j,n}} \mathbf{1}_{\{\mathbf{v}_{j,n} > 0\}}.$$

□