



HAL
open science

Minimax and pointwise sequential changepoint detection and identification for general stochastic models

Serguei M Pergamenchikov, Alexander G Tartakovsky, Valentin S Spivak

► To cite this version:

Serguei M Pergamenchikov, Alexander G Tartakovsky, Valentin S Spivak. Minimax and pointwise sequential changepoint detection and identification for general stochastic models. 2021. hal-03318733

HAL Id: hal-03318733

<https://hal.science/hal-03318733>

Preprint submitted on 10 Aug 2021

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.

Minimax and pointwise sequential changepoint detection and identification for general stochastic models

Serguei M. Pergamenchtchikov^a, Alexander G. Tartakovsky^{b,*}, Valentin S. Spivak^c

^a*Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen Normandie, France and National Research Tomsk State University, International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk, Russia*

^b*Space Informatics Laboratory, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia and AGT StatConsult, Los Angeles, California, USA*

^c*Space Informatics Laboratory, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia*

Abstract

This paper considers the problem of joint change detection and identification assuming multiple composite post-change hypotheses. We propose a multihypothesis changepoint detection-identification procedure that controls the probabilities of false alarm and wrong identification. We show that the proposed procedure is asymptotically minimax and pointwise optimal, minimizing moments of the detection delay as probabilities of false alarm and wrong identification approach zero. The asymptotic optimality properties hold for general stochastic models with dependent observations. We illustrate general results for detection-identification of changes in multistream Markov ergodic processes. We consider several examples, including an application to rapid detection-identification of COVID-19 in Italy. Our proposed sequential algorithm allows much faster detection of COVID-19 than standard methods.

Keywords: Asymptotic optimality; changepoint detection; composite post-change hypotheses; detection and localization of epidemics; quickest change detection-identification.

2010 MSC: Primary 62L10; 62L15; Secondary 60G40; 60J05; 60J20.

1. Introduction

As discussed in [9, 15–17], in a variety of applications it is important not only to quickly detect abrupt changes but also to diagnose them (e.g., to determine which change in a set of possible changes has occurred). This problem of change detection and diagnosis applies, for example, to rapid detection and identification of intrusions in computer networks, object detection with various sensors, integrity monitoring of navigation systems, and early detection and localization of epidemics. Often called Change Detection and Isolation, the problem is a generalization of the quickest change detection problem to the case of multiple post-change hypotheses and can be formulated as joint change detection and identification. Nikiforov [9] first considered the change detection-isolation problem in a minimax setting for independent and identically distributed (i.i.d.) observations (in pre-change and post-change modes with different distributions) and simple post-change hypotheses. Several versions of the multihypothesis CUSUM-type and SR-type procedures, which have minimax optimality properties in the classes of rules with constraints imposed on the average run length to a false alarm and conditional probabilities of false isolation, are proposed by Nikiforov [10, 11] and Tartakovsky [14]. Dayanik et al. [2] proposed an asymptotically optimal Bayesian detection-isolation rule assuming that the prior distribution of the change point is geometric also in the i.i.d. case. However, in many practical applications, the i.i.d. assumption is too restrictive – the observations may be either non-identically distributed or dependent or both, i.e., non-i.i.d. Also, the post-change distribution is usually not completely known. Lai [7] provided a certain generalization for the non-i.i.d. case and composite hypotheses for a specific loss function. Recently, Tartakovsky [16] developed a general asymptotic multistream Bayesian theory of sequential change detection and identification for low rates of false alarms and misidentification, assuming (1) there are multiple data streams, (2) the change occurs in some

*Corresponding author. Email address: agt@phystech.edu

data stream(s) at an unknown random point in time, and (3) it is necessary to detect the change as soon as possible and identify which data streams are affected. However, a non-Bayesian multistream change detection-identification theory for non-i.i.d. data is still missing.

The primary goal of this paper is to provide a general non-Bayesian asymptotic multistream change detection-identification theory (minimax and pointwise) for non-i.i.d. data and composite post-change hypotheses. This theory generalizes changepoint detection theory (with no identification) developed by Pergamenchchikov and Tarkovsky [13]. In Section 2, we describe the general stochastic model and provide basic notation. In Section 3, we introduce main conditions. In Section 4, we introduce the change detection-identification rule. In Section 5, we derive the information lower bounds for moments of the detection delay in the class of changepoint detection-identification rules with constraints imposed on the probabilities of false alarm and wrong identification. In Section 6, we prove asymptotic optimality of the proposed detection-identification rule as the probabilities of false alarm and misidentification go to zero. We show that the lower bounds are attained for this procedure under very general conditions. In Section 7, we illustrate general results for detection-identification of changes in Markov ergodic processes. In Section 8, we consider two examples – detection-identification of changes in (1) the parameters of multivariate linear difference equations and (2) the correlation coefficients of multistream p -th order autoregressive models. In Section 9, we propose a specific model for epidemics and show that the proposed change detection-identification rule is asymptotically optimal. We also apply our rule for detection of COVID-19 in Italy and show that it allows for much earlier detection of COVID-19 than standard methods.

2. Basic notation

We consider the N independent streams of observations $(X_{1,l})_{l \geq 1}, \dots, (X_{N,l})_{l \geq 1}$. For any $1 \leq i \leq N$, $\nu \geq 0$ and θ_i from an open set $\Theta_i \subseteq \mathbb{R}^m$ we denote by $\mathbf{P}_{i,\nu,\theta_i}$ the distribution of the observations $(X_{i,l})_{l \geq 1}$ in \mathbb{R}^∞ . In the case when $\nu = \infty$, this distribution will be denoted by \mathbf{P}^* . We use the convention that $X_{i,\nu}$ is the last pre-change observation. Write $\mathbf{X}_i^n = (X_{i,1}, \dots, X_{i,n})$ for the concatenation of the first n observations in the i th data stream. Let now for any $1 \leq i \leq N$

$$\left(f_{i,j}^*(y_j|y_1, \dots, y_{j-1})\right)_{j \geq 1} \quad \text{and} \quad \left(f_{i,\theta_i,j}(y_j|y_1, \dots, y_{j-1})\right)_{j \geq 1} \quad (2.1)$$

be sequences of conditional densities of $X_{i,j}$ given \mathbf{X}_i^{j-1} with respect to some non-degenerate σ -finite measure. Note that for $1 \leq i \leq N$ the density \mathbf{q}_i of \mathbf{X}_i^n in \mathbb{R}^n has the following form

$$\mathbf{q}_{i,\nu,\theta_i}(y_1, \dots, y_n) = \begin{cases} \mathbf{q}_i^*(y_1, \dots, y_n) & \text{for } \nu \geq n; \\ \prod_{l=1}^{\nu} f_{i,l}^*(y_l|y_1, \dots, y_{l-1}) \prod_{l=\nu+1}^n f_{i,\theta_i,l}(y_l|y_1, \dots, y_{l-1}) & \text{for } \nu < n, \end{cases} \quad (2.2)$$

where $\mathbf{q}_i^*(y_1, \dots, y_n) = \prod_{l=1}^n f_{i,l}^*(y_l|y_1, \dots, y_{l-1})$.

Denote by γ a random variable with values in $\{1, \dots, N\}$ and assume that the change can occur only in the data stream $(X_{\gamma,l})_{l \geq 1}$ with probability $\nu_i = \mathbf{P}(\gamma = i)$. For $\nu \geq 1$ and $\theta = (\theta_1, \dots, \theta_N) \in \Theta = \Theta_1 \times \dots \times \Theta_N$ the joint density of the observations $\mathbf{X}_1^n, \dots, \mathbf{X}_N^n$ is given by

$$\mathbf{p}_{\nu,\theta}(y_{1,1}, \dots, y_{N,n}) = \sum_{i=1}^N \nu_i \mathbf{p}_{i,\nu,\theta}(y_{1,1}, \dots, y_{N,n}), \quad (2.3)$$

where

$$\mathbf{p}_{i,\nu,\theta_i}(y_{1,1}, \dots, y_{N,n}) = \mathbf{q}_{i,\nu,\theta_i}(y_{i,1}, \dots, y_{i,n}) \prod_{l \neq i} \mathbf{q}_l^*(y_{l,1}, \dots, y_{l,n}). \quad (2.4)$$

In the sequel we denote by \mathcal{M} the set of all Markov times with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$ where $\mathcal{F}_0 = \{\Omega, \emptyset\}$ and $\mathcal{F}_n = \sigma\{X_{i,j}, 1 \leq i \leq N, 1 \leq j \leq n\}$.

Note that when $n > k$ and $\vartheta \in \Theta_i$ the Radon-Nykodim density (likelihood ratio)

$$\mathbf{g}_{i,k,n}^* = \mathbf{g}_{i,k,n}^*(\vartheta) = \frac{d\mathbf{P}_{i,k,\vartheta}}{d\mathbf{P}^*} \Big|_{\mathcal{F}_n} = e^{Z_{i,n}^k(\vartheta)}, \quad (2.5)$$

where

$$Z_{i,n}^k(\vartheta) = \sum_{l=k+1}^n \log \frac{f_{i,\vartheta,l}(X_{i,l}|\mathbf{X}_i^{l-1})}{f_{i,l}^*(X_i|\mathbf{X}_i^{l-1})} \quad (2.6)$$

is the log-likelihood ratio, and for any $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$ with $i \neq j$ the Radom-Nikodym density

$$\mathbf{g}_{i,j,k,n} = \mathbf{g}_{i,j,k,n}(\theta_i, \theta_j) = \frac{d\mathbf{P}_{i,k,\theta_i}}{d\mathbf{P}_{j,k,\theta_j}} \Big|_{\mathcal{F}_n} = e^{Z_{i,n}^k(\theta_i) - Z_{j,n}^k(\theta_j)}. \quad (2.7)$$

A sequential change detection-identification procedure δ is a pair $\delta = (T, d)$, where T is a stopping time from \mathcal{M} , i.e., for any $1 \leq i \leq N$, $k \geq 0$ and $\vartheta \in \Theta_i$ the probability $\mathbf{P}_{i,k,\vartheta}(T < \infty) = 1$, and d is a decision rule, i.e., a random variable with the values in $\{1, \dots, N\}$ which is measurable with respect to the σ -field \mathcal{F}_T . We denote by \mathcal{S} the class of all sequential procedures. For $r \geq 1$ and $\theta_i \in \Theta_i$, define the risk for a sequential procedure $\delta = (T, d) \in \mathcal{S}$ associated with the conditional r -th moment of the detection delay

$$\mathcal{R}_{i,k,\theta_i}(\delta) = \mathbf{E}_{i,k,\theta_i} \left[(T - k)^r \mathbf{1}_{\{d=i\}} | T > k \right], \quad (2.8)$$

where $\mathbf{E}_{i,k,\theta_i}$ is the expectation with respect to the distribution $\mathbf{P}_{i,k,\theta_i}$ in \mathbb{R}^∞ .

Introduce the conditional probability of false alarm $\mathbf{P}^*(T < k + \mathbf{m}^*, d = i | T \geq k)$ on the event $\{d = i\}$ in the interval $[k, k + \mathbf{m}^*)$, i.e., the probability of raising the alarm with the decision $d = i$ that there is a change in the i th stream when there is no change. Also, introduce the misidentification probabilities $\mathbf{P}_{i,k,\theta_i}(d = j | T > k)$, $i \neq j$, $i, j = 1, \dots, N$.

For any $N \times N$ matrix $\beta = (\beta_{i,j})_{1 \leq i,j \leq N}$ with $0 < \beta_{i,j} < 1$, $\mathbf{m}^* \geq 1$ and $\mathbf{k}^* > \mathbf{m}^*$ we introduce the class of change detection-identification rules

$$\mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*) = \left\{ \delta \in \mathcal{S} : \sup_{1 \leq k \leq \mathbf{k}^* - \mathbf{m}^*} \max_{1 \leq i \leq N} \frac{\mathbf{P}^*(T < k + \mathbf{m}^*, d = i | T \geq k)}{\beta_{i,i}} \leq 1, \right. \\ \left. \max_{0 \leq k \leq \mathbf{k}^*} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} \max_{1 \leq j \neq i \leq N} \frac{\mathbf{P}_{i,k,\theta_i}(d = j | T > k)}{\beta_{i,j}} \leq 1 \right\}. \quad (2.9)$$

Our goal is to find a sequential procedure asymptotically optimal in two problems in the class of detection-identification rules $\mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$: the pointwise minimization

$$\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \mathcal{R}_{i,k,\theta_i}(\delta) \quad \text{for every } 1 \leq i \leq N, k \geq 0 \text{ and } \theta_i \in \Theta_i \quad (2.10)$$

and the minimax optimization

$$\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \sup_{k \geq 0} \mathcal{R}_{i,k,\theta_i}(\delta) \quad \text{for every } 1 \leq i \leq N \text{ and } \theta_i \in \Theta_i. \quad (2.11)$$

The parameters \mathbf{k}^* and \mathbf{m}^* will be specified later.

3. Main conditions

For a fixed $\theta_i \in \Theta_i$, we assume the following conditions for the log-likelihood ratio (LLR) processes $(Z_{i,n}^k(\theta_i))_{n \geq k+1}$ introduced in (2.6) for $1 \leq i \leq N$ and $\theta_i \in \Theta_i$.

(A₁) For any $1 \leq i, j \leq N$ there are $\Theta_i \times \Theta_j \rightarrow \mathbb{R}_+$ positive continuous functions $I_{i,j}$ with

$$0 < \inf_{(\theta_i, \theta_j) \in \Theta_i \times \Theta_j} I_{i,j}(\theta_i, \theta_j) \leq \sup_{(\theta_i, \theta_j) \in \Theta_i \times \Theta_j} I_{i,j}(\theta_i, \theta_j) < \infty \quad (3.1)$$

such that for any $k \geq 0$, $\varepsilon > 0$ and $\theta = (\theta_1, \dots, \theta_N) \in \Theta_1 \times \dots \times \Theta_N$

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \neq j \leq N} \mathbf{P}_{i,k,\theta_i} \left(Z_{i,k+n}^k(\theta_i) - Z_{j,k+n}^k(\theta_j) > (1 + \varepsilon) I_{i,j}(\theta_i, \theta_j) n \right) = 0 \quad (3.2)$$

and

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq N} \mathbb{P}_{i,k,\theta_i} \left(Z_{i,k+n}^k(\theta_i) > (1 + \varepsilon) \widetilde{I}_i(\theta_i)n \right) = 0, \quad (3.3)$$

where $\widetilde{I}_i(\theta_i) = I_{i,i}(\theta_i, \theta_i)$ for $\theta_i \in \Theta_i$.

In order to study asymptotic approximations to risks of the change detection-identification rule introduced below in Section 4 and for establishing its asymptotic optimality, we impose the following left-tail conditions:
(A₂(r)) For any $1 \leq i, j \leq N$ there are $\Theta_i \times \Theta_j \rightarrow \mathbb{R}_+$ positive continuous functions $I_{i,j}$ with the property (3.1) such that for every $\theta = (\theta_1, \dots, \theta_N) \in \Theta_1 \times \dots \times \Theta_N = \Theta$ and for any $0 < \varepsilon < 1$

$$\limsup_{\zeta \rightarrow 0} \max_{\theta \in \Theta} \max_{1 \leq i \leq N} \sum_{n=1}^{\infty} n^{r-1} \sup_{k \geq 0} \mathbb{P}_{i,k,\theta_i} \left(\inf_{|u-\theta_i| < \zeta} Z_{i,k+n}^k(u) < (1 - \varepsilon) \widetilde{I}_i(\theta_i)n \right) < \infty \quad (3.4)$$

and

$$\limsup_{\zeta \rightarrow 0} \max_{\theta \in \Theta} \max_{1 \leq i \neq j \leq N} \sum_{n=1}^{\infty} n^{r-1} \sup_{0 \leq k \leq \zeta n} \mathbb{P}_{i,k,\theta_i} \left(\inf_{|u-\theta_i| < \zeta} Z_{i,k+n}^k(u) - Z_{j,k+n}^* < (1 - \varepsilon) \widetilde{I}_i(\theta_i)n \right) < \infty, \quad (3.5)$$

where $Z_{j,n}^* = \max_{0 \leq l \leq n} \sup_{z \in \Theta_j} Z_{j,l+n}^l(z)$.

Remark 1. This is always true for i.i.d. data models with Kullback–Leibler informations given by

$$I_{i,j}(\theta_i, \theta_j) = \int \log \left(\frac{f_{i,\theta_i}(x)}{f_{j,\theta_j}(x)} \right) f_{i,\theta_i}(x) d\mu(x) \quad \text{and} \quad \widetilde{I}_i(\theta_i) = \int \log \left(\frac{f_{i,\theta_i}(x)}{f_i^*(x)} \right) f_{i,\theta_i}(x) d\mu(x)$$

for $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$.

4. Sequential change detection-identification procedure

First introduce weight distributions $(W_i(\vartheta))_{1 \leq i \leq N}$, which are probability measures on the sets Θ_i , i.e. $W_i(\Theta_i) = 1$ for any $1 \leq i \leq N$. In what follows, we assume that $W_i(\cdot)$ satisfy the following condition:

(C_W) For any $\varepsilon > 0$ and any $\vartheta \in \Theta_i$ the measure $W_i\{u \in \Theta_i : |u - \vartheta| < \varepsilon\} > 0$.

Now, for some fixed $0 < \varrho < 1$ we set

$$\pi_k = \pi_k(\varrho) = \varrho (1 - \varrho)^k, \quad k = 0, 1, 2, \dots \quad (4.1)$$

and using this distribution we set

$$\mathbf{L}_{i,n} = \sum_{k=0}^{n-1} \pi_k \int_{\Theta_i} \mathbf{g}_{i,k,n}^*(\vartheta) W_i(d\vartheta) \quad \text{and} \quad \widehat{\mathbf{L}}_{i,n} = \sum_{k=0}^{n-1} \pi_k \sup_{\theta \in \Theta_i} \mathbf{g}_{i,k,n}^*(\theta). \quad (4.2)$$

Using these statistics we define the following random $N \times N$ matrix \mathbf{U}_n as

$$\langle \mathbf{U}_n \rangle_{i,j} = \frac{\mathbf{L}_{i,n}}{\widehat{\mathbf{L}}_{j,n}} \quad \text{if } i \neq j \quad \text{and} \quad \langle \mathbf{U}_n \rangle_{i,i} = \frac{\mathbf{L}_{i,n}}{\sum_{l \geq n} \pi_l}, \quad (4.3)$$

where $\langle \mathbf{U} \rangle_{i,j}$ is the (i, j) th element of the matrix \mathbf{U} . Finally, using this matrix we set

$$T_{i,A}^* = \inf \left\{ n \geq 1 : \min_{1 \leq j \leq N} \frac{\langle \mathbf{U}_n \rangle_{i,j}}{A_{i,j}} \geq 1 \right\}, \quad (4.4)$$

where $A = (A_{i,j})_{1 \leq i, j \leq N}$ is a $N \times N$ matrix with positive elements which will be specified later. In the definitions of stopping times we set $\inf\{\emptyset\} = +\infty$. The sequential change detection-identification procedure $\delta_A^* = (T_A^*, d_A^*)$ that will be studied in this paper has the form

$$T_A^* = \min_{1 \leq i \leq N} T_{i,A}^* \quad \text{and} \quad d_A^* = i \quad \text{if} \quad T_{i,A}^* = T_A^*. \quad (4.5)$$

If there are several numbers i for which $T_{i,A}^* = T_A^*$ we can take arbitrary. Note that, as we will see later in Proposition 2, the condition $(\mathbf{A}_2^*(r))$ implies that T_A^* is a $\mathbb{P}_{i,k,\theta_i}$ -proper stopping time, that is, for any $1 \leq i \leq N, k \geq 1$ and $\theta_i \in \Theta_i$

$$\mathbb{P}_{i,k,\theta_i}(T_A^* < \infty) = 1. \quad (4.6)$$

Now, for any sequential procedure $\delta = (T, d) \in \mathcal{S}$ we set

$$\text{PFA}_i(\delta) = \sum_{k=0}^{\infty} \pi_k \mathbb{P}^*(T \leq k, d = i) \quad (4.7)$$

and

$$\text{PMI}_{i,j}(\delta) = \sum_{k=0}^{\infty} \pi_k \sup_{\theta_i \in \Theta_i} \mathbb{P}_{i,k,\theta_i}(T > k, d = j). \quad (4.8)$$

For some $N \times N$ matrix $\alpha = (\alpha_{i,j})$ with $0 < \alpha_{i,j} < 1$ and some fixed $0 < \varrho < 1$, define the following *Bayesian class*:

$$\Delta(\alpha, \varrho) = \left\{ \delta \in \mathcal{S} : \max \left(\max_{1 \leq i \leq N} \frac{\text{PFA}_i(\delta)}{\alpha_{i,i}}, \max_{1 \leq i, j \leq N, i \neq j} \frac{\text{PMI}_{i,j}(\delta)}{\alpha_{i,j}} \right) \leq 1 \right\}. \quad (4.9)$$

Next, for any arbitrary fixed matrix $\beta = (\beta_{i,j})_{1 \leq i, j \leq N}$ and $0 < \varrho < 1$ introduce two matrices $\alpha_1 = (\alpha_{i,j}^{(1)})_{1 \leq i, j \leq N}$ and $\alpha_2 = (\alpha_{i,j}^{(2)})_{1 \leq i, j \leq N}$ as

$$\alpha_{i,j}^{(1)} = \frac{\beta_{i,i}(1-\varrho)^{\mathbf{k}^*}}{1 + \text{tr} \beta} \mathbf{1}_{\{i=j\}} + \frac{\beta_{i,j}\varrho(1-\varrho)^{\mathbf{k}^*}}{1 + \text{tr} \beta} \mathbf{1}_{\{i \neq j\}} \quad (4.10)$$

and

$$\alpha_{i,j}^{(2)} = (\beta_{i,i} + (1-\varrho)^{\mathbf{m}^*+1}) \mathbf{1}_{\{i=j\}} + (\beta_{i,j} + (1-\varrho)^{\mathbf{k}^*+1}) \mathbf{1}_{\{i \neq j\}}. \quad (4.11)$$

The following proposition compares classes (2.9) and (4.9).

Proposition 1. *For any matrix $\beta = (\beta_{i,j})_{1 \leq i, j \leq N}$, $1 \leq \mathbf{m}^* < \mathbf{k}^*$ and $0 < \varrho < 1$ the following inclusions hold*

$$\Delta(\alpha_1, \varrho) \subset \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*) \subset \Delta(\alpha_2, \varrho). \quad (4.12)$$

Proof. First note that if $\delta = (T, d) \in \Delta(\alpha_1, \varrho)$, then for any $1 \leq i \leq N$ and $k \geq 1$

$$\alpha_{i,i}^{(1)} \geq \sum_{l=0}^{\infty} \pi_l \mathbb{P}^*(T \leq l, d = i) \geq \mathbb{P}^*(T \leq k, d = i) \sum_{l=k}^{\infty} \pi_l = (1-\varrho)^k \mathbb{P}^*(T \leq k, d = i),$$

i.e., for $k \geq 1$

$$\mathbb{P}^*(T \leq k, d = i) \leq (1-\varrho)^{-k} \alpha_{i,i}^{(1)} \quad \text{and} \quad \mathbb{P}^*(T \leq k) \leq \sum_{j=1}^N \mathbb{P}^*(T \leq k, d = j) \leq (1-\varrho)^{-k} \text{tr} \alpha_1. \quad (4.13)$$

Therefore, for any $1 \leq i \leq N$ and $1 \leq k \leq \mathbf{k}^* - \mathbf{m}^*$

$$\begin{aligned} \mathbb{P}^*(T < k + \mathbf{m}^*, d = i | T \geq k) &= \frac{\mathbb{P}^*(k \leq T < k + \mathbf{m}^*, d = i)}{1 - \mathbb{P}^*(T < k)} \\ &\leq \frac{\mathbb{P}^*(T \leq \mathbf{k}^*, d = i)}{1 - \mathbb{P}^*(T \leq \mathbf{k}^*)} \leq \frac{(1-\varrho)^{-\mathbf{k}^*} \alpha_{i,i}^{(1)}}{1 - (1-\varrho)^{-\mathbf{k}^*} \text{tr} \alpha_1} := \beta_{i,i}. \end{aligned}$$

Moreover, for $i \neq j$ and any $k \geq 1$

$$\alpha_{i,j}^{(1)} \geq \pi_k \sup_{\theta_i \in \Theta_i} \mathbb{P}_{i,k,\theta_i}(T > k, d = j) = \varrho(1-\varrho)^k \sup_{\theta_i \in \Theta_i} \mathbb{P}_{i,k,\theta_i}(T > k, d = j),$$

i.e., in view of (4.13), for $1 \leq k \leq \mathbf{k}^*$ and $\theta_i \in \Theta_i$

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i}(d = j | T > k) &= \frac{\mathbf{P}_{i,k,\theta_i}(T > k, d = j)}{\mathbf{P}_{i,k,\theta_i}(T > k)} = \frac{\mathbf{P}_{i,k,\theta_i}(T > k, d = j)}{1 - \mathbf{P}^*(T > k)} \\ &\leq \frac{\alpha_{i,j}^{(1)} \varrho^{-1} (1 - \varrho)^{-k}}{1 - \mathbf{P}^*(T \leq k)} \leq \frac{\alpha_{i,j}^{(1)} \varrho^{-1} (1 - \varrho)^{-\mathbf{k}^*}}{1 - (1 - \varrho)^{-\mathbf{k}^*} \text{tr } \alpha_1} = \beta_{i,j}. \end{aligned}$$

This implies, that $\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$, i.e., we get the first inclusion in (4.12).

Let now $\delta = (T, d) \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$, i.e., for any $1 \leq k \leq \mathbf{k}^* - \mathbf{m}^*$ and $1 \leq i \leq N$

$$\beta_{i,i} \geq \mathbf{P}^*(T < k + \mathbf{m}^*, d = i | T \geq k) = \frac{\mathbf{P}^*(k \leq T < k + \mathbf{m}^*, d = i)}{\mathbf{P}^*(T \geq k)} \geq \mathbf{P}^*(k \leq T < k + \mathbf{m}^*, d = i)$$

and, in particular,

$$\beta_{i,i} \geq \mathbf{P}^*(T < 1 + \mathbf{m}^*, d = i).$$

Therefore,

$$\begin{aligned} \sum_{k \geq 0} \pi_k \mathbf{P}^*(T \leq k, d = i) &\leq \mathbf{P}^*(T < 1 + \mathbf{m}^*, d = i) + \sum_{k \geq \mathbf{m}^* + 1} \pi_k \\ &\leq \beta_{i,i} + (1 - \varrho)^{\mathbf{m}^* + 1} = \alpha_{i,i}^{(2)}. \end{aligned}$$

Furthermore, for any $i \neq j$, $\theta_i \in \Theta_i$ and $0 \leq k \leq \mathbf{k}^*$

$$\beta_{i,j} \geq \mathbf{P}_{i,k,\theta_i}(d = j | T > k) \geq \mathbf{P}_{i,k,\theta_i}(k < T < \infty, d = j),$$

i.e.,

$$\sum_{k \geq 0} \pi_k \mathbf{P}_{i,k,\theta_i}(T > k, d = j) \leq \beta_{i,j} + \sum_{k > \mathbf{k}^*} \pi_k = \beta_{i,j} + (1 - \varrho)^{\mathbf{k}^* + 1} = \alpha_{i,j}.$$

Thus, we obtain the last inclusion (4.12). Hence Proposition 1. \square

The first question we ask is how to select the thresholds in the procedure (4.5) to imbed it into class $\Delta(\alpha, \rho)$. To study this question we need the following probability measures on $\mathbb{N} \times \mathbb{R}^\infty$ which for any $1 \leq i \leq N$ are defined as

$$\mathbf{Q}_i(J \times A) = \sum_{k \in J} \pi_k \int_{\Theta_i} \mathbf{P}_{i,k,\vartheta}(A) W_i(d\vartheta), \quad J \subseteq \mathbb{N} \quad \text{and} \quad A \in \mathcal{B}(\mathbb{R}^\infty), \quad (4.14)$$

where $\mathbb{N} = \{0, 1, \dots\}$ and $\mathcal{B}(\mathbb{R}^\infty)$ is the cylinder field in \mathbb{R}^∞ . In the sequel we denote by $\mathbf{E}^{\mathbf{Q}_i}$ the expectation over the probability measure \mathbf{Q}_i . One can check directly that

$$\mathbf{Q}_i(v \geq n | \mathcal{F}_n^*) = \frac{1}{1 + \langle \mathbf{U}_n \rangle_{i,i}}, \quad (4.15)$$

where the matrix \mathbf{U}_n is defined in (4.3).

Lemma 1. For all $1 \leq i \leq N$ the probabilities (4.7) satisfy the inequalities

$$\text{PFA}_i(\delta_A^*) \leq \frac{1}{1 + A_{i,i}}. \quad (4.16)$$

Proof. Note that

$$\text{PFA}_i(\delta_A^*) = \sum_{k=0}^{\infty} \pi_k \mathbf{P}^*(T_A^* \leq k, d = i) = \sum_{k=0}^{\infty} \pi_k \int_{\Theta_i} \mathbf{P}_{i,k,\vartheta}(T_{i,A}^* \leq k, d = i) W_i(d\vartheta)$$

$$\leq \sum_{k=0}^{\infty} \pi_k \int_{\Theta_i} \mathbf{P}_{i,k,\theta} (T_{i,A}^* \leq k) W_i(d\theta) = \mathbf{Q}_i (v \geq T_{i,A}^*) = \mathbf{E}^{\mathbf{Q}_i} \left[\mathbf{Q}_i (v \geq T_{i,A}^* | \mathcal{F}_{T_{i,A}^*}) \right].$$

Therefore, using (4.15) and (4.4) we obtain that

$$\text{PFA}_i(\delta_A^*) \leq \mathbf{E}^{\mathbf{Q}_i} \left[\frac{1}{1 + \langle \mathbf{U}_{T_{i,A}^*} \rangle_{i,i}} \right] \leq \frac{1}{1 + A_{i,i}},$$

which completes the proof. \square

Lemma 2. For any $1 \leq i, j \leq N$, $i \neq j$ the PMI probabilities of the procedure (4.5) satisfy the inequalities

$$\text{PMI}_{i,j}(\delta_A^*) \leq \frac{1}{A_{j,i}}. \quad (4.17)$$

Proof. First note that for the rule (4.4) we obtain that for any $j \neq i$ and $\theta_i \in \Theta_i$

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i} (T_A^* > k, d = j) &\leq \mathbf{P}_{i,k,\theta_i} (k < T_{j,A}^* < \infty) \\ &\leq \frac{1}{A_{j,i}} \mathbf{E}_{i,k,\theta_i} \left[\langle \mathbf{U}_{T_{j,A}^*} \rangle_{j,i} \mathbf{1}_{\{k < T_{j,A}^* < \infty\}} \right] \\ &= \frac{1}{A_{j,i}} \mathbf{E}^* \left[\langle \mathbf{U}_{T_{j,A}^*} \rangle_{j,i} \mathbf{g}_{i,k,T_{j,A}^*}(\theta_i) \mathbf{1}_{\{k < T_{j,A}^* < \infty\}} \right]. \end{aligned}$$

Therefore, in view of the definition in (4.2), we get

$$\begin{aligned} \sum_{k \geq 0} \pi_k \mathbf{P}_{i,k,\theta_i} (T_A^* > k, d = j) &\leq \frac{1}{A_{j,i}} \mathbf{E}^* \left[\langle \mathbf{U}_{T_{j,A}^*} \rangle_{j,i} \sum_{k=0}^{T_{j,A}^*-1} \pi_k \mathbf{g}_{i,k,T_{j,A}^*}(\theta_i) \mathbf{1}_{\{T_{j,A}^* < \infty\}} \right] \\ &\leq \frac{1}{A_{j,i}} \mathbf{E}^* \left[\langle \mathbf{U}_{T_{j,A}^*} \rangle_{j,i} \widehat{\mathbf{L}}_{i,T_{j,A}^*} \mathbf{1}_{\{T_{j,A}^* < \infty\}} \right] \\ &= \frac{1}{A_{j,i}} \mathbf{E}^* \left[\mathbf{L}_{j,T_{j,A}^*} \mathbf{1}_{\{T_{j,A}^* < \infty\}} \right]. \end{aligned}$$

Moreover, note that

$$\begin{aligned} \mathbf{E}^* \left[\mathbf{L}_{j,T_{j,A}^*} \mathbf{1}_{\{T_{j,A}^* < \infty\}} \right] &= \sum_{k \geq 0} \pi_k \int_{\Theta_i} \mathbf{E}^* \left[\mathbf{g}_{j,k,T_{j,A}^*}(\theta_i) \mathbf{1}_{\{k < T_{j,A}^* < \infty\}} \right] W_i(d\theta_i) \\ &= \sum_{k \geq 0} \pi_k \int_{\Theta_i} \mathbf{P}_{i,k,\theta_i} (k < T_{j,A}^* < \infty) W_i(d\theta_i) \leq 1, \end{aligned}$$

which implies upper bound (4.17) \square

Now, if we take in (4.4)

$$A_{i,j} = \left(\frac{1}{\alpha_{i,i}} - 1 \right) \mathbf{1}_{\{i=j\}} + \frac{1}{\alpha_{j,i}} \mathbf{1}_{\{i \neq j\}} \quad (4.18)$$

then using the property (4.6) and the upper bounds (4.16) and (4.17) we obtain that under condition $(\mathbf{A}_2^*(r))$ the sequential procedure (4.5) belongs to class $\Delta(\varrho, \alpha)$ for any $0 < \varrho < 1$ and $\alpha = (\alpha_{i,j})_{1 \leq i, j \leq N}$ with $0 < \alpha_{i,j} < 1$. Therefore, if we take

$$A_{i,j} = \left(\frac{1 + \text{tr} \beta}{\beta_{i,i}(1 - \varrho)^{\mathbf{k}^*}} - 1 \right) \mathbf{1}_{\{i=j\}} + \frac{1 + \text{tr} \beta}{\beta_{j,i} \varrho (1 - \varrho)^{\mathbf{k}^*}} \mathbf{1}_{\{i \neq j\}}, \quad (4.19)$$

we obtain that for any $0 < \varrho < 1$ the sequential procedure (4.5) belongs to class $\mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$.

5. Information lower bounds

5.1. Bayesian setting

For any matrix $\alpha = (\alpha_{i,j})_{1 \leq i,j \leq N}$ and any parameter value $\theta_i \in \Theta_i$ define

$$\mathbf{b}_{i,\alpha}(\theta_i) = \max_{1 \leq j \leq N} \frac{|\log \alpha_{ji}|}{\iota_{i,j}(\theta_i)} \quad \text{and} \quad \iota_{i,j}(\theta_i) = \widetilde{I}_i(\theta_i) \mathbf{1}_{\{j=i\}} + \widehat{I}_{i,j}(\theta_i) \mathbf{1}_{\{j \neq i\}}, \quad (5.1)$$

where the function $\widetilde{I}_i(\cdot)$ is defined in (3.3) and $\widehat{I}_{i,j}(\theta_i) = \inf_{\vartheta \in \Theta_j} I_{i,j}(\theta_i, \vartheta)$ for $\theta_i \in \Theta_i$. In what follows we will always suppose without special emphasis that

$$\inf_{\theta_j \in \Theta_j} I_{i,j}(\theta_i, \theta_j) > 0 \quad \text{for all } \theta_i \in \Theta_i, j \neq i \text{ and } i = 1, \dots, N$$

(see condition (3.1)).

Write $\alpha_{max} = \max_{1 \leq i,j \leq N} \alpha_{i,j}$ and $\Delta_\alpha = \Delta(\alpha, \varrho_\alpha)$ (in case where $\varrho = \varrho_\alpha$ depends on α). The following theorem establishes information lower bounds in the Bayesian problem. These bounds will be used to obtain asymptotic lower bounds for $\mathcal{R}_{i,k,\theta_i}(\delta)$ in class $\mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$ (see Theorem 2) and to prove asymptotic optimality of the proposed detection-identification procedure in this class.

Theorem 1. *Assume that the right-tail probability convergence condition (\mathbf{A}_1) holds and in (4.1) the parameter of the geometric prior distribution ϱ is a function of α , i.e., $\varrho = \varrho_\alpha$, such that*

$$\lim_{\alpha_{max} \rightarrow 0} \left(\varrho_\alpha + \frac{|\log \varrho_\alpha|}{|\log \alpha_{max}|} \right) = 0. \quad (5.2)$$

Then, for any $r \geq 1$, $k \geq 0$, $\theta_i \in \Theta_i$, and $1 \leq i \leq N$ the following asymptotic lower bounds hold:

$$\liminf_{\alpha_{max} \rightarrow 0} \frac{\inf_{\delta \in \Delta_\alpha} \mathbb{E}_{i,k,\theta_i} [(T-k)_+^r \mathbf{1}_{\{d=i\}}]}{|\mathbf{b}_{i,\alpha}(\theta_i)|^r} \geq 1. \quad (5.3)$$

Proof. To prove this theorem it suffices to show that for any $j \neq i$ and $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$ ($j = 1, \dots, N$)

$$\inf_{\delta \in \Delta_\alpha} \mathbb{E}_{i,k,\theta_i} [(T-k)_+^r \mathbf{1}_{\{d=i\}}] \geq (1 + \psi_{\alpha,i,j}(\theta_i, \theta_j)) \frac{|\log \alpha_{ji}|^r}{I_{i,j}^r(\theta_i, \theta_j)}, \quad (5.4)$$

and for any $j = i$ and $\theta_i \in \Theta_i$

$$\inf_{\delta \in \Delta_\alpha} \mathbb{E}_{i,k,\theta_i} [(T-k)_+^r \mathbf{1}_{\{d=i\}}] \geq (1 + \psi_{\alpha,i,i}(\theta_i, \theta_i)) \frac{|\log \alpha_{ii}|^r}{\widetilde{I}_i^r(\theta_i)}, \quad (5.5)$$

where the term $\psi_{\alpha,i,j}(\theta_i, \theta_j)$ is such that

$$\lim_{\alpha_{max} \rightarrow 0} |\psi_{\alpha,i,j}(\theta_i, \theta_j)| = 0 \quad \text{for any } (\theta_i, \theta_j) \in \Theta_i \times \Theta_j \text{ and } 1 \leq j \leq N.$$

To prove (5.4) note that condition (3.2) implies that for any $\varepsilon > 0$, $k \geq 0$ and $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$ with $j \neq i$

$$\mathbb{P}_{i,k,\theta_i} \left\{ Z_{i,k+M}^k(\theta_i) - Z_{j,k+M}^k(\theta_j) \geq (1 + \varepsilon) I_{i,j} M \right\} \xrightarrow{M \rightarrow \infty} 0, \quad (5.6)$$

where $I_{i,j} = I_{i,j}(\theta_i, \theta_j)$. Define for $(\theta_i, \theta_j) \in \Theta_i \times \Theta_j$

$$\mathbf{D}_{i,j,k}(\delta) = \mathbb{P}_{i,k,\theta_i} (k < T \leq k + M_{i,j}, d = i) \quad \text{and} \quad M_{i,j} = M_{i,j}(\theta_i, \theta_j) = (1 - \varepsilon) \frac{|\log \alpha_{ji}|}{I_{i,j}}.$$

We now show that for any $k \geq 0$, $0 < \varepsilon < 1$ and $1 \leq j \neq i \leq N$

$$\lim_{\alpha_{max} \rightarrow 0} \sup_{\delta \in \Delta_\alpha} \mathbf{D}_{i,j,k}(\delta) = 0. \quad (5.7)$$

Using definition (2.7) we can obtain that for $\mathbf{m} = k + M_{i,j}$

$$\begin{aligned} \mathbf{D}_{i,j,k}(\delta) &= \mathbb{E}_{j,k,\theta_j} \left[\mathbf{g}_{i,j,k,\mathbf{m}} \mathbf{1}_{\{k \leq T \leq \mathbf{m}, d=i\}} \right] \leq e^{(1+\varepsilon)I_{i,j}M_{i,j}} \mathbb{P}_{j,k,\theta_j}(k < T \leq \mathbf{m}, d=i) \\ &\quad + \mathbb{P}_{i,k,\theta_i} \left(Z_{i,\mathbf{m}}^k(\theta_i) - Z_{j,\mathbf{m}}^k(\theta_j) \geq (1+\varepsilon)I_{i,j}M_{i,j} \right). \end{aligned} \quad (5.8)$$

Using the definition of $\text{PMI}_{ji}(\delta)$ in (4.8) along with the fact that $\text{PMI}_{ji}(\delta) \leq \alpha_{ji}$ for any $\delta \in \Delta(\alpha, \varrho_\alpha)$ we get

$$\begin{aligned} \alpha_{ji} &\geq \sum_{l \geq 0} \pi_l \sup_{\theta_j \in \Theta_j} \mathbb{P}_{j,l,\theta_j}(T > l, d=i) \geq \pi_k \mathbb{P}_{j,l,\theta_j}(T > l, d=i) \geq \pi_k \mathbb{P}_{j,k,\theta_j}(k < T \leq \mathbf{m}, d=i) \\ &= \varrho_\alpha (1 - \varrho_\alpha)^k \mathbb{P}_{j,k,\theta_j}(k < T \leq \mathbf{m}, d=i) \quad \text{for any } \theta_j \in \Theta_j \text{ and any } k \geq 0, \end{aligned}$$

so that

$$\sup_{\delta \in \Delta_\alpha} \mathbb{P}_{j,k,\theta_j}(k < T \leq \mathbf{m}, d=i) \leq \varrho_\alpha^{-1} (1 - \varrho_\alpha)^{-k} \alpha_{ji} = e^{-|\log \alpha_{ji}| - \log \varrho_\alpha + k\varpi_\alpha},$$

where in view of (5.2) the term $\varpi_\alpha = -\log(1 - \varrho_\alpha) \rightarrow 0$ as $\alpha_{max} \rightarrow 0$. So the first term on the right-hand side of the inequality (5.8) can be estimated as

$$\exp\left\{(1+\varepsilon)I_{i,j}M_{i,j} - \log \varrho_\alpha + k\varpi_\alpha + \log \alpha_{ji}\right\} \leq \exp\left\{-\varepsilon^2 |\log \alpha_{ji}| - \log \varrho_\alpha + k\varpi_\alpha\right\} \leq \exp\left\{-\varepsilon^2 |\log \alpha_{max}| - \log \varrho_\alpha + k\varpi_\alpha\right\}$$

and by condition (5.2) it goes to zero as $\alpha_{max} \rightarrow 0$. Therefore, (5.8) and (5.6) imply (5.7) for any $j \neq i$.

Let now $i = j$. Using the definition (2.5) we can rewrite the inequality (5.8) as

$$\begin{aligned} \mathbf{D}_{i,i,k}(\delta) &= \mathbb{E}^* \left[\mathbf{g}_{i,i,k,\mathbf{m}}^* \mathbf{1}_{\{k \leq T \leq \mathbf{m}, d=i\}} \right] \leq e^{(1+\varepsilon)\widetilde{I}_i(\theta_i)M_{ii}} \mathbb{P}^*(k < T \leq \mathbf{m}, d=i) \\ &\quad + \mathbb{P}_{i,k,\theta_i} \left(Z_{i,\mathbf{m}}^k(\theta_i) \geq (1+\varepsilon)\widetilde{I}_i(\theta_i)M_{ii} \right), \end{aligned} \quad (5.9)$$

where $\widetilde{I}_i(\theta_i) = I_{i,i}(\theta_i, \theta_i)$ and where by condition (3.3)

$$\lim_{M_{ii} \rightarrow \infty} \mathbb{P}_{i,k,\theta_i} \left(Z_{i,k+M_{ii}}^k(\theta_i) > (1+\varepsilon)\widetilde{I}_i(\theta_i)M_{ii} \right) = 0 \quad \text{for any } \theta_i \in \Theta_i. \quad (5.10)$$

Now, the definition of class $\Delta_\alpha = \Delta(\alpha, \varrho_\alpha)$ in (4.9) implies that for any $\delta \in \Delta_\alpha$, any $k \geq 1$ and all $i = 1, \dots, N$

$$\alpha_{ii} \geq \sum_{\ell \geq 1} \pi_\ell \mathbb{P}^*(T \leq \ell, d=i) \geq \mathbb{P}^*(T \leq k, d=i) \sum_{\ell=1}^k \varrho_\alpha (1 - \varrho_\alpha)^{\ell-1} = (1 - \varrho_\alpha)^k \mathbb{P}^*(T \leq k, d=i),$$

which yields

$$\sup_{\delta \in \Delta_\alpha} \mathbb{P}^*(T \leq k, d=i) \leq \alpha_{ii} (1 - \varrho_\alpha)^{-k} = e^{-|\log \alpha_{ii}| + k\varpi_\alpha}. \quad (5.11)$$

Therefore, the first term on the right side of the inequality (5.9) may be estimated as

$$e^{(1+\varepsilon)\widetilde{I}_i(\theta_i)M_{ii} - |\log \alpha_{ii}| + \varpi_\alpha k + \varpi_\alpha M_{ii}} \leq e^{-\varepsilon^2 |\log \alpha_{ii}| + \varpi_\alpha k + \varpi_\alpha M_{ii}}$$

and it goes to zero for any fixed $0 \leq k < \infty$ as $\alpha_{max} \rightarrow 0$, which along with (5.10) implies (5.7) for $i = j$.

To obtain lower bounds (5.4) and (5.5) note that for any $1 \leq j \leq N$

$$\mathbb{E}_{i,k,\theta_i} \left[(T-k)_+^r \mathbf{1}_{\{d=i\}} \right] \geq \mathbb{E}_{i,k,\theta_i} \left[(T-k)_+^r \mathbf{1}_{\{T > k+M_{i,j}, d=i\}} \right] \geq M_{i,j}^r \mathbb{P}_{i,k,\theta_i}(T > k+M_{i,j}, d=i)$$

$$= M_{i,j}^r \left(\mathbf{P}_{i,k,\theta_i}(T > k, d = i) - \mathbf{D}_{i,j,k}(\delta) \right). \quad (5.12)$$

Using the upper bound (5.11), we get

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i}(T > k, d = i) &= \mathbf{P}_{i,k,\theta_i}(d = i) - \mathbf{P}_{i,k,\theta_i}(T \leq k, d = i) \\ &= \mathbf{P}_{i,k,\theta_i}(d = i) - \mathbf{P}^*(T \leq k, d = i) \\ &\geq 1 - \sum_{l=1, l \neq i}^N \mathbf{P}_{i,k,\theta_i}(d = l) - \alpha_{i,i}(1 - \varrho_\alpha)^{-k}. \end{aligned}$$

Next, it follows from (4.8) and (4.9) that for $l \neq i$

$$\sup_{\theta_i \in \Theta_i} \mathbf{P}_{i,k,\theta_i}(T > k, d = l) \leq \frac{\alpha_{i,l}}{\varrho_\alpha} (1 - \varrho_\alpha)^{-k} = e^{\log \alpha_{i,l} - \log \varrho_\alpha + k \varpi_\alpha},$$

i.e., for any $l \neq i$ and $\theta_i \in \Theta_i$

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i}(d = l) &= \mathbf{P}^*(T \leq k, d = l) + \mathbf{P}_{i,k,\theta_i}(T > k, d = l) \\ &\leq \alpha_{i,l}(1 - \varrho)^{-k} + \frac{\alpha_{i,l}}{\varrho_\alpha} (1 - \varrho_\alpha)^{-k} = e^{\log \alpha_{i,l} + k \varpi_\alpha} + e^{\log \alpha_{i,l} - \log \varrho_\alpha + k \varpi_\alpha}. \end{aligned}$$

Thus, in view of (5.2) for any $\theta_i \in \Theta_i$

$$\mathbf{P}_{i,k,\theta_i}(T > k, d = i) \rightarrow 1 \quad \text{as} \quad \alpha_{max} \rightarrow 0$$

and using (5.12) and (5.7), we finally obtain the asymptotic inequality

$$\mathbb{E}_{i,k,\theta_i} \left[(T - k)_+^r \mathbf{1}_{\{d=i\}} \right] \geq \left[(1 - \varepsilon) \frac{|\log \alpha_{j,i}|}{I_{i,j}(\theta_i, \theta_j)} \right]^r (1 + o(1)), \quad \alpha_{max} \rightarrow 0$$

(where $o(1) \rightarrow 0$), which holds for an arbitrary $\varepsilon \in (0, 1)$, so letting $\varepsilon \rightarrow 0$ implies lower bounds (5.4) (for $j \neq i$) and (5.5) (for $j = i$). The proof is complete. \square

5.2. The local constraints setting

To find asymptotic lower bounds for the problems (2.10) and (2.11) in addition to condition (\mathbf{A}_1) we impose the following condition.

(H₁) The parameters ϱ , \mathbf{m}^* and \mathbf{k}^* in (4.11) are functions of β , i.e. $\varrho = \varrho_\beta$, $\mathbf{m}^* = \mathbf{m}_\beta^*$ and $\mathbf{k}^* = \mathbf{k}_\beta^*$, such that

$$\lim_{\beta_{max} \rightarrow 0} \varrho_\beta = 0, \quad \lim_{\beta_{max} \rightarrow 0} \frac{|\log \varrho_\beta|}{|\log \alpha_{max}^{(2)}|} = 0 \quad \text{and} \quad \lim_{\beta_{max} \rightarrow 0} \max_{1 \leq i, j \leq N} \left| \frac{|\log \alpha_{i,j}^{(2)}|}{|\log \beta_{i,j}|} - 1 \right| = 1, \quad (5.13)$$

where $\alpha_{max}^{(2)} = \max_{1 \leq i, j \leq N} \alpha_{i,j}^{(2)}$ and $\beta_{max} = \max_{1 \leq i, j \leq N} \beta_{i,j}$.

For example, we can take

$$\mathbf{m}_\beta^* = \lceil |\log \beta_{min}| / \varrho_\beta \rceil, \quad \mathbf{k}_\beta^* = \check{\mathbf{k}} \mathbf{m}_\beta^* \quad \text{and} \quad \varrho_\beta = \frac{1}{1 + |\log \beta_{max}|}, \quad (5.14)$$

where $[x]$ is the integer part of the x , $\beta_{min} = \min_{1 \leq i, j \leq N} \beta_{i,j}$ and $\check{\mathbf{k}} > 1$ is some fixed number.

The following theorem establishes asymptotic lower bounds in class of detection-identification procedures $\mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)$.

Theorem 2. Assume that conditions (\mathbf{A}_1) and (\mathbf{H}_1) hold. Then, for any $r \geq 1$, $k \geq 0$, $1 \leq i \leq N$ and $\theta_i \in \Theta_i$,

$$\liminf_{\beta_{max} \rightarrow 0} \frac{\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \mathcal{R}_{i,k,\theta_i}(\delta)}{\mathbf{b}_{i,\beta}^r(\theta_i)} \geq 1, \quad (5.15)$$

where the denominator $\mathbf{b}_{i,\beta}(\theta_i)$ is defined in (5.1) by replacing the matrix α with β .

Proof. First of all, note that the last condition in (5.13) implies that for any $1 \leq i \leq N$ and $\theta_i \in \Theta_i$

$$\lim_{\beta_{\max} \rightarrow 0} \frac{\mathbf{b}_{i,\alpha_2}(\theta_i)}{\mathbf{b}_{i,\beta}(\theta_i)} = 1.$$

It is clear that the last inclusion in (4.12) implies that for all $k \geq 0$ and sufficiently small $\beta > 0$

$$\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \mathcal{R}_{i,k,\theta_i}(\delta) \geq \inf_{\delta \in \Delta(\alpha_2, \varrho_\beta)} \mathcal{R}_{i,k,\theta_i}(\delta) \geq \inf_{\delta \in \Delta(\alpha_2, \varrho_\beta)} \mathbf{E}_{i,k,\theta_i} \left[(T-k)_+^r \mathbf{1}_{\{d=i\}} \right].$$

Now, the lower bounds (5.15) directly follow from the lower bounds (5.3) and condition (\mathbf{H}_1) . \square

6. Upper bounds and asymptotic optimality

We begin with studying the sequential procedure (4.5) for large threshold values $A_{i,j}$. For any matrix $A = (A_{i,j})_{1 \leq i,j \leq N}$ with $A_{i,j} > 1$ and any $\vartheta \in \Theta_i$ define

$$\mathbf{B}_{i,A}(\vartheta) = \max_{1 \leq j \leq N} \frac{\log A_{i,j}}{\iota_{i,j}(\vartheta)}, \quad (6.1)$$

where the ‘‘information’’ functions $\iota_{i,j}(\cdot)$ are defined in (5.1). We need the following condition:

(\mathbf{H}_2) The matrix $A = (A_{i,j})_{1 \leq i,j \leq N}$ is such that

$$\lim_{A_{\min} \rightarrow \infty} \max_{1 \leq i \leq N} \left| \frac{\max_{1 \leq j \leq N} \log A_{i,j}}{\log A_{i,i}} - 1 \right| = 0.$$

Proposition 2. *If conditions (\mathbf{A}_2) and (\mathbf{H}_2) hold true, then for any $1 \leq i \leq N$ and any compact set $\mathbf{K} \subset \Theta$ the sequential procedure (4.5), in which $|\log \varrho| = o(\log A_{\min})$ as $A_{\min} \rightarrow \infty$, admits the following upper bound*

$$\limsup_{A_{\min} \rightarrow \infty} \max_{1 \leq i \leq N} \sup_{\theta_i \in \mathbf{K}} \sup_{1 \leq k \leq \mathbf{k}_*} \frac{\mathcal{R}_{i,k,\theta_i}(\delta_A^*)}{\mathbf{B}_{i,A}^r(\theta_i)} \leq 1, \quad (6.2)$$

where $\theta = (\theta_1, \dots, \theta_N)$, $A_{\min} = \min_{1 \leq i,j \leq N} A_{i,j}$ and \mathbf{k}_* is such that $\mathbf{k}_* = o(\log A_{\min})$ as $A_{\min} \rightarrow \infty$.

Proof. First, note that in view of (4.17) δ_A^* belongs to class $\Delta(\varrho, \alpha)$ with

$$\alpha_{i,j} = \frac{1}{1 + A_{i,i}} \mathbf{1}_{\{i=j\}} + \frac{1}{1 + A_{j,j}} \mathbf{1}_{\{i \neq j\}}.$$

Therefore, using the upper bound (4.13), we obtain that uniformly over $1 \leq k \leq \mathbf{k}_*$

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i}(T_A^* > k) &= \mathbf{P}^*(T_A^* > k) \geq 1 - (1 - \varrho)^{-k} \sum_{j=1}^N \frac{1}{1 + A_{j,j}} \\ &\geq 1 - (1 - \varrho)^{-k} \sum_{j=1}^N \frac{1}{1 + A_{j,j}} \rightarrow 1, \quad A_{\min} \rightarrow \infty. \end{aligned}$$

Therefore, to obtain the inequality (6.2) it suffices to show that

$$\limsup_{A_{\min} \rightarrow \infty} \max_{1 \leq i \leq N} \sup_{\theta_i \in \mathbf{K}} \sup_{1 \leq k \leq \mathbf{k}_*} \frac{\mathbf{E}_{i,k,\theta_i} [(T_A^* - k)_+^r] \mathbf{1}_{\{d_A^*=i\}}}{\mathbf{B}_{i,A}^r(\theta_i)} \leq 1. \quad (6.3)$$

Note also that by condition (A_2) for arbitrary $0 < \varepsilon < 1$ we can chose such $0 < \zeta < 1$ for which $\{\vartheta : |\vartheta - \theta_i| < \zeta\} \subset \Theta_i$ for all $1 \leq i \leq N$,

$$\mathbf{U}_1^*(\zeta) = \max_{1 \leq i \leq N} \sup_{\theta_i \in \mathbf{K}} \sum_{n=1}^{\infty} n^{r-1} \sup_{k \geq 0} \mathbf{P}_{i,k,\theta_i} \left(\inf_{|u-\theta_i| < \zeta} Z_{i,k+n}^k(u) < (1-\varepsilon)\widetilde{I}_i(\theta_i)n \right) < \infty$$

and

$$\mathbf{U}_2^*(\zeta) = \max_{i \neq j} \sup_{\theta_i \in \mathbf{K}} \sum_{n=1}^{\infty} n^{r-1} \max_{0 \leq k \leq \zeta n} \mathbf{P}_{i,k,\theta_i} \left(\inf_{|u-\theta_i| < \zeta} Z_{i,k+n}^k(u) - Z_{j,k+n}^* < (1-\varepsilon)\widetilde{I}_i(\theta_i)n \right) < \infty.$$

Now, for an arbitrary $\vartheta \in \Theta_i$, we set

$$\mathbf{n}_i^* = \mathbf{n}_{i,A}^*(\vartheta) = \left\lceil \frac{1+\varepsilon}{1-\varepsilon} \varkappa_{i,A} \mathbf{B}_{i,A}(\vartheta) \right\rceil + 1 \quad \text{and} \quad \varkappa_{i,A} = \frac{\max_{1 \leq j \leq N} \log A_{i,j}}{\log A_{i,i}}. \quad (6.4)$$

Note that in view of the properties (3.1) and the fact that $\mathbf{k}_* = o(\log A_{\min})$ we can conclude that for a sufficiently large A_{\min} we have $\mathbf{k}_* \leq \zeta \mathbf{n}_i^*$. Moreover,

$$\mathbb{E}_{i,k,\theta_i} \left[(T_A^* - k)_+ \mathbf{1}_{\{d_A^* = i\}} \right] \leq \sum_{n \geq 0} n^{r-1} \mathbf{P}_{i,k,\theta_i} (T_{i,A}^* > k+n) \leq 1 + (\mathbf{n}_i^*)^r + \sum_{n > \mathbf{n}_i^*} n^{r-1} \mathbf{P}_{i,k,\theta_i} (T_{i,A}^* > k+n). \quad (6.5)$$

Now, the definition (4.4) implies that

$$\mathbf{P}_{i,k,\theta_i} (T_{i,A}^* > k+n) = \mathbf{P}_{i,k,\theta_i} \left(\max_{1 \leq l \leq n+k} \min_{1 \leq j \leq N} \frac{\langle \mathbf{U}_l \rangle_{i,j}}{A_{i,j}} < 1 \right) \leq \sum_{j=1}^N \mathbf{P}_{i,k,\theta_i} (\langle \mathbf{U}_{k+n} \rangle_{i,j} < A_{i,j}),$$

i.e.

$$\mathbf{P}_{i,k,\theta_i} (T_{i,A}^* > k+n) \leq \sum_{j=1}^N \mathbf{P}_{i,k,\theta_i} (\log \langle \mathbf{U}_{k+n} \rangle_{i,j} < \log A_{i,j}). \quad (6.6)$$

Using (4.3) we obtain that for $i \neq j$

$$\mathbf{P}_{i,k,\theta_i} (\log \langle \mathbf{U}_{k+n} \rangle_{i,j} < \log A_{i,j}) = \mathbf{P}_{i,k,\theta_i} (\log \mathbf{L}_{i,k+n} - \log \widehat{\mathbf{L}}_{j,k+n} < \log A_{i,j}) \quad (6.7)$$

and for $i = j$

$$\mathbf{P}_{i,k,\theta_i} (\log \langle \mathbf{U}_{k+n} \rangle_{i,i} < \log A_{i,i}) = \mathbf{P}_{i,k,\theta_i} (\log \mathbf{L}_{i,k+n} - (k+n) \log(1-\varrho) < \log A_{i,i}). \quad (6.8)$$

Note that

$$\log \mathbf{L}_{i,k+n} \geq \log \pi_k \int_{\{|v-\theta_i| < \zeta\}} \mathbf{g}_{i,k,k+n}^*(v) W_i(dv) \geq \inf_{\{|v-\theta_i| < \zeta\}} Z_{i,k+n}^k(v) + \mathbf{I}_i(\varrho),$$

where $\mathbf{I}_i(\varrho) = -\log \varrho - \mathbf{k}_* \log(1-\varrho) - \log W_i(\zeta \in \Theta_i : |\zeta - \theta_i| < \zeta)$. Since $|\log \varrho| = o(\log A_{\min})$ and $\mathbf{k}_* = o(\log A_{\min})$, we get

$$\lim_{A_{\min} \rightarrow \infty} \frac{\max_{1 \leq i \leq N} \mathbf{I}_i(\varrho)}{\log A_{\min}} = 0.$$

Obviously, $\log \widehat{\mathbf{L}}_{j,k+n} \leq Z_{j,k+n}^*$. Moreover, taking into account that for any $1 \leq j \leq N$

$$\mathbf{n}_i^* \geq \frac{(1+\varepsilon) \log A_{i,j}}{(1-\varepsilon)\widetilde{I}_i(\theta_i)}$$

we can obtain that for any $j \neq i$, $\theta_i \in \Theta_i$, $n > \mathbf{n}_i^* > \mathbf{k}_*/\zeta$ and $0 \leq k \leq \mathbf{k}_*$ and for sufficiently large A_{\min} for which $\max_{1 \leq i \leq N} \mathbf{I}_i(\varrho) \leq \varepsilon \log A_{\min}$

$$\mathbf{P}_{i,k,\theta_i} (\log \langle \mathbf{U}_{k+n} \rangle_{i,j} < \log A_{i,j}) \leq \mathbf{P}_{i,k,\theta_i} \left(\inf_{\{|v-\theta_i| < \zeta\}} Z_{i,k+n}^k(v) - Z_{j,k+n}^* < \log A_{i,j} + \mathbf{I}_i(\varrho) \right)$$

$$\leq \max_{0 \leq l \leq \zeta n} \mathbf{P}_{i,l,\theta_i} \left(\inf_{|v-\theta_i| < \zeta} Z_{i,l+n}^l(v) - Z_{j,l+n}^* < (1-\varepsilon) \widetilde{I}_i(\theta_i) n \right)$$

and

$$\begin{aligned} \mathbf{P}_{i,k,\theta_i} \left(\log \langle \mathbf{U}_{k+n} \rangle_{i,i} < \log A_{i,i} \right) &\leq \mathbf{P}_{i,k,\theta_i} \left(\inf_{\{|v-\theta_i| < \zeta\}} Z_{i,k+n}^k(v) < \log A_{i,i} + \mathbf{I}_i(\varrho) \right) \\ &\leq \max_{0 \leq l \leq \zeta n} \mathbf{P}_{i,l,\theta_i} \left(\inf_{|v-\theta_i| < \zeta} Z_{i,l+n}^l(v) < (1-\varepsilon) \widetilde{I}_i(\theta_i) n \right). \end{aligned}$$

Therefore, from (6.5) we get

$$\begin{aligned} \mathbf{E}_{i,k,\theta_i} \left[(T_A^* - k)_+^r \mathbf{1}_{\{d_A^* = i\}} \right] &\leq 1 + (\mathbf{n}_i^*)^r + \mathbf{U}_1^*(\zeta) + \mathbf{U}_2^*(\zeta) \\ &\leq 1 + \left(\frac{1+\varepsilon}{1-\varepsilon} \right)^r \mathcal{Z}_{i,A}^r \mathbf{B}_{i,A}^r(\theta_i) + \mathbf{U}_1^*(\zeta) + \mathbf{U}_2^*(\zeta), \end{aligned}$$

and using the condition (\mathbf{H}_2) , we get

$$\limsup_{A_{\min} \rightarrow \infty} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta} \sup_{1 \leq k \leq \mathbf{k}_*} \frac{\mathbf{E}_{i,k,\theta_i} \left[(T_A^* - k)_+^r \mathbf{1}_{\{d_A^* = i\}} \right]}{\mathbf{B}_{i,A}^r(\theta_i)} \leq \left(\frac{1+\varepsilon}{1-\varepsilon} \right)^r.$$

Since ε can be arbitrarily small, taking the limit as $\varepsilon \rightarrow 0$, we obtain the bound (6.3), which completes the proof of Proposition 2. \square

Remark 2. If both left-tail and right-tail conditions (\mathbf{A}_1) and (\mathbf{A}_2) hold along with conditions (\mathbf{H}_1) and (\mathbf{H}_2) , then inverting the equality (4.19) and using Theorem 2 (with β replaced with A^{-1}) and Proposition 2 simultaneously it can be shown that the following asymptotic equalities for the moments of delay of the procedure δ_A hold for any fixed k , $\theta_i \in \Theta_i$ and all $i = 1, \dots, N$:

$$\mathcal{R}_{i,k,\theta_i}(\delta_A) = \max_{1 \leq j \leq N} \frac{\log A_{i,j}}{I_{i,j}(\theta_i)} (1 + o(1)) = \max \left\{ \frac{\log A_{i,i}}{I_i(\theta_i)}, \max_{1 \leq j \neq i \leq N} \frac{\log A_{i,j}}{I_{i,j}(\theta_i)} \right\} (1 + o(1)) \quad \text{as } A_{\min} \rightarrow \infty.$$

To obtain the optimal detection rate we need to impose the following condition:

(\mathbf{H}_3) Parameters ϱ^{opt} and \mathbf{k}^* are functions of β , i.e. $\varrho^{opt} = \varrho_\beta^{opt}$, $\mathbf{k}^* = \mathbf{k}_\beta^*$ and $\mathbf{m}^* = \mathbf{m}_\beta^*$, such that

$$\lim_{\beta_{\max} \rightarrow 0} \frac{|\log \varrho^{opt}|}{|\log \beta_{\max}|} = 0 \quad \text{and} \quad \lim_{\beta_{\max} \rightarrow 0} \max_{1 \leq i,j \leq N} \left| \frac{|\log \alpha_{i,j}^{(1)}|}{|\log \beta_{i,j}|} - 1 \right| = 0, \quad (6.9)$$

where

$$\alpha_{i,j}^{(1)} = \beta_{i,i} (1 - \varrho^{opt})^{\mathbf{k}^*} \mathbf{1}_{\{i=j\}} + \beta_{i,j} \varrho^{opt} (1 - \varrho^{opt})^{\mathbf{k}^*} \mathbf{1}_{\{i \neq j\}}.$$

For example, for some $\check{\mathbf{k}} > 1$ we can take

$$\mathbf{m}_\beta^* = \left[\frac{|\log \beta_{\min}|}{\varrho_\beta} \right], \quad \mathbf{k}_\beta^* = \check{\mathbf{k}} \mathbf{m}_\beta^*, \quad \varrho_\beta^{opt} = \frac{|\log \beta_{\max}| \varrho_\beta}{|\log \beta_{\min}| (1 + |\log \varrho_\beta|)} \quad \text{and} \quad \varrho_\beta = \frac{1}{1 + |\log \beta_{\max}|}. \quad (6.10)$$

Then under the conditions

$$\lim_{\beta_{\max} \rightarrow 0} \frac{\log |\log \beta_{\min}|}{|\log \beta_{\max}|} = 0 \quad \text{and} \quad \lim_{\beta_{\max} \rightarrow 0} \max_{1 \leq i \leq N} \left| \frac{\max_{1 \leq j \leq N} |\log \beta_{j,i}|}{|\log \beta_{i,i}|} - 1 \right| = 0, \quad (6.11)$$

we obtain that the conditions $(\mathbf{H}_1) - (\mathbf{H}_3)$ hold.

Denote by $\delta_\beta^{opt} = (T^{opt}, d^{opt})$ the procedure (4.5) with $\varrho = \varrho^{opt}$ and

$$A_{i,j} = A_{i,j}^{opt}(\beta) = \left(\frac{1 + \text{tr} \beta}{\beta_{i,i}(1 - \varrho^{opt})^{\mathbf{k}^*}} - 1 \right) \mathbf{1}_{\{i=j\}} + \frac{1 + \text{tr} \beta}{\beta_{j,i}(\varrho^{opt})(1 - \varrho^{opt})^{\mathbf{k}^*}} \mathbf{1}_{\{i \neq j\}}, \quad (6.12)$$

that is,

$$T_\beta^{opt} = \min_{1 \leq i \leq N} T_{i\beta}^{opt} \quad \text{and} \quad d_\beta^{opt} = i \quad \text{if} \quad T_{i\beta}^{opt} = T_\beta^{opt}, \quad (6.13)$$

where

$$T_{i\beta}^{opt} = \inf \left\{ n \geq 1 : \min_{1 \leq j \leq N} \frac{\langle \mathbf{U}_n \rangle_{i,j}}{A_{i,j}^{opt}(\beta)} \geq 1 \right\}. \quad (6.14)$$

The following theorem deduces the pointwise and minimax optimality properties of the procedure δ_β^{opt} .

Theorem 3. *Assume that conditions (\mathbf{A}_1) – (\mathbf{A}_2) and (\mathbf{H}_1) – (\mathbf{H}_3) hold true. Then the procedure δ_β^{opt} is optimal in the pointwise sense, i.e., for any $\theta_i \in \Theta_i$, $1 \leq i \leq N$, and for every fixed $k \geq 0$*

$$\lim_{\beta_{max} \rightarrow 0} \frac{\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \mathcal{R}_{i,k,\theta_i}(\delta)}{\mathcal{R}_{i,k,\theta_i}(\delta_\beta^{opt})} = 1 \quad \text{and} \quad \lim_{\beta_{max} \rightarrow 0} \frac{\mathcal{R}_{i,k,\theta_i}(\delta_\beta^{opt})}{\mathbf{b}_{i\beta}^r(\theta_i)} = 1. \quad (6.15)$$

Also, for any $\mathbf{k}_* = o(|\log \beta_{max}|)$ as $\beta_{max} \rightarrow 0$ the procedure δ_β^{opt} is optimal in the minimax sense, i.e., for any $\theta_i \in \Theta_i$ and $1 \leq i \leq N$,

$$\lim_{\beta_{max} \rightarrow 0} \frac{\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \max_{1 \leq k \leq \mathbf{k}_*} \mathcal{R}_{i,k,\theta_i}(\delta)}{\max_{1 \leq k \leq \mathbf{k}_*} \mathcal{R}_{i,k,\theta_i}(\delta_\beta^{opt})} = 1 \quad \text{and} \quad \lim_{\beta_{max} \rightarrow 0} \frac{\max_{1 \leq k \leq \mathbf{k}_*} \mathcal{R}_{i,k,\theta_i}(\delta_\beta^{opt})}{\mathbf{b}_{i\beta}^r(\theta_i)} = 1. \quad (6.16)$$

Proof. By condition (\mathbf{H}_3)

$$\lim_{\beta_{max} \rightarrow 0} \max_{1 \leq i, j \leq N} \left| \frac{|\log A_{i,j}^{opt}(\beta)|}{|\log \beta_{j,i}|} - 1 \right| = 0,$$

so that using the asymptotic upper bound (6.2) in Proposition 2 we obtain the asymptotic upper bound

$$\limsup_{\beta_{max} \rightarrow 0} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} \sup_{1 \leq k \leq \mathbf{k}_*} \frac{\mathcal{R}_{i,k,\theta_i}(\delta_\beta^{opt})}{\mathbf{B}_{i\beta}^r(\theta_i)} \leq 1. \quad (6.17)$$

Comparing this bound with the lower bound (5.3) in Theorem 1 yields (6.15) and (6.16). \square

The next theorem also shows that the procedure $\delta_\beta^{opt} = (T^{opt}, d^{opt})$ is “robust” in the following sense

$$\mathcal{R}_\beta^*(\delta) = \sup_{\theta \in \Theta} \max_{1 \leq i \leq N} \max_{1 \leq k \leq \mathbf{k}_*} \frac{\mathcal{R}_{i,k,\theta_i}(\delta)}{\mathbf{b}_{i\beta}^r(\theta_i)}. \quad (6.18)$$

Theorem 4. *Suppose that conditions (\mathbf{A}_1) – (\mathbf{A}_2) and (\mathbf{H}_1) – (\mathbf{H}_3) hold and $\mathbf{k}_* = o(|\log \beta_{max}|)$ as $\beta_{max} \rightarrow 0$. Then*

$$\lim_{\beta_{max} \rightarrow 0} \frac{\inf_{\delta \in \mathcal{H}(\beta, \mathbf{k}^*, \mathbf{m}^*)} \mathcal{R}_\beta^*(\delta)}{\mathcal{R}_\beta^*(\delta_\beta^{opt})} = 1. \quad (6.19)$$

The proof is similar to the proof of Theorem 2 in [13] for the single-stream detection problem and is omitted.

7. Detection-identification of changes in homogeneous Markov models

Let the observations $(X_{i,n})_{n \geq 1}$ be time homogeneous Markov processes with values in a measurable space $(\mathcal{X}_i, \mathcal{B}_i)$ defined by a family of the transition probabilities $(P_i^{\theta_i}(x, A))_{\theta_i \in \Theta_i}$ for some fixed parameter set $\Theta_i \subseteq \mathbb{R}^p$. In the sequel we denote by $E_{i,x}^{\theta_i}(\cdot)$ the expectation with respect to this probability. Moreover, we assume that for any $1 \leq i \leq N$ the observations $(X_{i,n})_{n \geq 1}$ are Markov processes, such that $(X_{i,n})_{1 \leq n \leq \nu}$ is a homogeneous process with the transition (from x to y) density $f_i^*(y|x)$ and in the case when $\nu = +\infty$ this process is ergodic with the ergodic distribution $\lambda_{*,i}$. We denote by \mathbf{P}^* the distribution of the observations $(X_{i,n})_{1 \leq i \leq N, n \geq 1}$ of this process when $\nu = \infty$. The expectation with respect to this distribution will be denoted by $E^*(\cdot)$. In addition, we assume that for any $1 \leq i \leq N$ the process $(X_{i,n})_{n > \nu}$ is homogeneous positive ergodic with the transition density $f_{i,\theta_i}(y|x)$ and the ergodic (stationary) distribution $\lambda_{\theta_i,i}$ ($\theta_i \in \Theta_i$). The densities $f_i^*(y|x)$ and $(f_{i,\theta_i}(y|x))_{\theta_i \in \Theta_i}$ are calculated with respect to a sigma-finite positive measure μ_i on \mathcal{B}_i . In this case, we can represent the LLR process $Z_{i,n}^k(u)$ defined in (2.6) as

$$Z_{i,n}^k(u) = \sum_{j=k+1}^n g_i(u, X_{i,j}, X_{i,j-1}), \quad g_i(u, y, x) = \log \frac{f_{i,u}(y|x)}{f_i^*(y|x)}. \quad (7.1)$$

We also assume that densities $f_{i,u}(y|x)$ are continuously differentiable with respect to u in a compact set $\bar{\Theta}_i \subseteq \Theta_i$. Now we set

$$h(x, y) = \max_{1 \leq i \leq N} \max_{u \in \bar{\Theta}_i} \max_{1 \leq j \leq p} |\partial g_i(u, y, x) / \partial u_j| \quad (7.2)$$

and

$$\bar{h}_i(\theta_i, y) = \int_{\mathcal{X}_i} h(y, x) f_{i,\theta_i}(y|x) \mu_i(dy), \quad \theta_i \in \Theta_i.$$

For some $\mathbf{q} > 0$ define

$$g_{\mathbf{q}}^*(x) = \sup_{n \geq 1} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} E_{i,x}^{\theta_i} |g_i(\theta_i, X_{i,n}, X_{i,n-1})|^{\mathbf{q}}$$

and

$$h_{\mathbf{q}}^*(x) = \sup_{n \geq 1} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} E_{i,x}^{\theta_i} |h(X_{i,n}, X_{i,n-1})|^{\mathbf{q}}. \quad (7.3)$$

Also, define

$$J_i(\theta_i, x) = \int_{\mathcal{X}_i} g_i(\theta_i, y, x) f_{i,\theta_i}(y|x) \mu_i(dy) \quad \text{and} \quad J_i^*(\theta_i, x) = \int_{\mathcal{X}_i} g_i(\theta_i, y, x) f_i^*(y|x) \mu_i(dy). \quad (7.4)$$

Obviously, $J_i(\theta_i, x) \geq 0$ and $J_i^*(\theta_i, x) \leq 0$. Write

$$\bar{J}_i(\theta_i) = \int_{\mathcal{X}_i} J_i(\theta_i, x) \lambda_{\theta_i,i}(dx) \quad \text{and} \quad \bar{J}_i^*(\theta_i) = \int_{\mathcal{X}_i} J_i^*(\theta_i, x) \lambda_{*,i}(dx). \quad (7.5)$$

Introduce the following conditions.

(C₁) For any $1 \leq i \leq N$ there exist sets $C_i \in \mathcal{B}_i$ with $\mu_i(C_i) < \infty$ such that

(C1.1) $f_* = \min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \inf_{x, y \in C_i} f_{i,\theta_i}(y|x) > 0$.

(C1.2) For any $1 \leq i \leq N$ there exists $\mathcal{X}_i \rightarrow [1, \infty)$ Lyapunov's function \mathbf{V}_i such that

- $\mathbf{V}_i(x) \geq J_i(\theta_i, x)$ and $\mathbf{V}_i(x) \geq \bar{h}_i(\theta_i, x)$ for any $\theta_i \in \Theta_i$ and $x \in \mathcal{X}_i$.
- $\max_{1 \leq i \leq N} \sup_{x \in C_i} \mathbf{V}_i(x) < \infty$.
- There exist $0 < \rho < 1$ and $\mathbf{m}^* > 0$ such that for all $1 \leq i \leq N$, $x \in \mathcal{X}_i$ and $\theta_i \in \Theta_i$,

$$E_{i,x}^{\theta_i}[\mathbf{V}_i(X_{i,1})] \leq (1 - \rho)\mathbf{V}_i(x) + \mathbf{m}^* \mathbf{1}_{C_i}(x). \quad (7.6)$$

(C₂(q)) There exists $q > 2$ such that for any $1 \leq i \leq N$

$$\sup_{k \geq 1} \mathbb{E}^* [g_q^*(X_{i,k})] < \infty, \quad \sup_{k \geq 1} \mathbb{E}^* [h_q^*(X_{i,k})] < \infty \quad \text{and} \quad \sup_{k \geq 1} \mathbb{E}^* [v_q^*(X_{i,k})] < \infty,$$

where the functions $g_q^*(x)$ and $h_q^*(x)$ are given in (7.3) and

$$v_q^*(x) = \sup_{1 \leq i \leq N, n \geq 0} \sup_{\theta_i \in \Theta_i} \mathbb{E}_{i,x}^{\theta_i} [\mathbf{V}_i(X_{i,n})]^q. \quad (7.7)$$

(C₃(q)) The function $g_i(u, y, x)$ can be represented as

$$g_i(u, y, x) - \bar{J}_i^*(u) = \sum_{l=1}^m \alpha_{j,l}(u) \check{g}_{i,l}(y, x) \quad (7.8)$$

with $\alpha_{j,l}(u)$ and $\check{g}_{i,l}(y, x)$ such that for any $1 \leq i \leq N$ and $1 \leq l \leq m$

$$\sup_{u \in \Theta_i} |\alpha_{j,l}(u)| < \infty \quad \text{and} \quad \sup_{n \geq 1} n^{-q/2} \mathbb{E}^* \left| \sum_{j=1}^n \check{g}_{i,l}(X_{i,j}, X_{i,j-1}) \right|^q < \infty.$$

Theorem 5. Assume that conditions (C₁) – (C₃(q)) hold true and the functions $\bar{J}_i(\theta_i)$ and $-\bar{J}_i^*(\theta_i)$ defined in (7.4) are continuous and positive for $\theta_i \in \Theta_i$. Then conditions (A₁) and (A₂(r)) are satisfied for any $0 < r < q/2$ with

$$I_{i,j}(\theta_i, \theta_j) = \bar{J}_i(\theta_i) \mathbf{1}_{\{i=j\}} + [\bar{J}_i(\theta_i) - \bar{J}_j^*(\theta_j)] \mathbf{1}_{\{i \neq j\}}. \quad (7.9)$$

Proof. Note first that conditions (3.3) and (3.4) follow from Theorem 8 in [13] that uses the uniform geometric ergodicity property and concentration inequalities methods developed in [3, 4]. To prove condition (3.5) we observe that condition (C₃) and the Chebyshev inequality imply that for any $a > 0$

$$\sup_{n \geq 1} n^{q/2} \max_{1 \leq i \leq N} \max_{1 \leq k \leq n} \mathbb{P}^* \left(\sup_{\theta_i \in \Theta_i} |\Delta_{i,n}^k(\theta_i)| > an \right) < \infty, \quad (7.10)$$

where

$$\Delta_{i,n}^k(\theta_i) = Z_{i,n}^k - \bar{J}_i^*(\theta_i)(n-k) = \sum_{j=k+1}^n (g_i(\theta_i, X_{i,j}, X_{i,j-1}) - \bar{J}_i^*(\theta_i)).$$

Taking into account that $\bar{J}_j^*(\theta_j) \leq 0$, we obtain that for any $\varepsilon > 0$ and $1 \leq j \leq N$

$$\mathbb{P}^* (Z_{j,n}^* > \varepsilon n) \leq \mathbb{P}^* \left(\max_{1 \leq k \leq n} \sup_{\theta_j \in \Theta_j} |\Delta_{j,n}^k(\theta_j)| > \varepsilon n \right) \leq n \max_{1 \leq k \leq n} \mathbb{P}^* \left(\sup_{\theta_j \in \Theta_j} |\Delta_{j,n}^k(\theta_j)| > \varepsilon n \right).$$

Therefore, (7.10) implies the condition (3.5) for $r < q/2$. \square

Remark 3. The function $\bar{J}_i(\cdot)$ is called the Kullback-Leibler divergence for the Markov processes (see, e.g., [5]).

Note that condition (C1.1) does not always hold for the process $(X_{i,n})_{n \geq 1}$ directly. For example, this condition does not hold for the practically important autoregression process of the order more than one. For this reason, we need to weaken this requirement. Similarly to [12] we assume that there exists $p \geq 2$ for which the process $(\bar{X}_{i,n}^t)_{n \geq \bar{v}_i}$ for $\bar{v}_i = \nu/p - \iota$ and $0 \leq \iota \leq p - 1$ defined as $\bar{X}_{i,n}^t = X_{i, np + \iota}$ satisfies the following conditions:

(C'₁) There exist sets $C_i \in \mathcal{B}_i$ with $\mu_i(C_i) < \infty$ such that

$$(C'1.1) \quad \min_{1 \leq i \leq N} \inf_{1 \leq \iota \leq p} \inf_{\theta_i \in \Theta_i} \inf_{x, y \in C} \bar{f}_{i, \theta_i}^t(y|x) > 0, \quad \text{where } \bar{f}_{i, \theta_i}^t(y|x) \text{ is the transition density for the process } (\bar{X}_{i,n}^t)_{n \geq \bar{v}_i}.$$

(C'1.2) For any $1 \leq i \leq N$ there exists $\mathcal{X}_i \rightarrow [1, \infty)$ Lyapunov's function \mathbf{V}_i such that

$$\max_{1 \leq i \leq N} \max_{1 \leq j \leq p} \sup_{\theta_i \in \Theta_i} \sup_{x \in \mathcal{X}_i} \frac{\mathbb{E}_{i,x}^{\theta_i} [\mathbf{V}_i(X_{i,j})]}{\mathbf{V}_i(x)} < \infty \quad \text{and} \quad \sup_{\theta_i \in \Theta_i} \lambda_{\theta_i,i}(\mathbf{V}_i) < \infty.$$

- $\mathbf{V}_i(x) \geq J_i(\theta_i, x)$ and $\mathbf{V}_i(x) \geq \bar{h}_i(\theta_i, x)$ for $\theta_i \in \Theta_i$ and $x \in \mathcal{X}_i$ and $\max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} \sup_{x \in C_i} \mathbf{V}_i(x) < \infty$.
- For any $1 \leq i \leq N$, there exist $0 < \rho < 1$ and $\mathbf{m}^* > 0$ such that for all $x \in \mathcal{X}_i$, $\theta_i \in \Theta_i$, and $0 \leq \iota \leq p-1$

$$\mathbb{E}_{i,x}^{\theta_i} [\mathbf{V}_i(\bar{X}_{i,1}^\iota)] \leq (1-\rho)\mathbf{V}_i(x) + \mathbf{m}^* \mathbf{1}_{C_i}(x). \quad (7.11)$$

Similarly to Theorem 5 we can prove, using Theorem 9 in [13], the following result.

Theorem 6. Assume that conditions (C₁) and (C₂(q)) hold, the processes $(\bar{X}_{i,n}^\iota)_{n \geq \bar{\nu}}$ satisfy the condition (C₃(q)), and the functions $\bar{J}_i(\theta_i)$ and $-\bar{J}_i^*(\theta_i)$ defined in (7.4) are continuous and positive for $\theta_i \in \Theta_i$. Then conditions (A₁) and (A₂(r)) hold for any $0 < r < \mathbf{q}/2$ with the functions $I_{i,j}$ defined in (7.9).

To check the condition C₃(q) we need to obtain concentration inequality for the homogenous Markov process $(X_{i,n})_{n \geq 1}$ with the transition density $f_i^*(y|x)$. The following condition is sufficient for this purpose:
(C₃^{*}(q)) For any $1 \leq i \leq N$ there exist sets $C_i \in \mathcal{B}_i$ with $\mu_i(C_i) < \infty$ such that

1. $\min_{1 \leq i \leq N} \inf_{x,y \in C_i} f_i^*(y|x) > 0$.
2. For any $1 \leq i \leq N$ there exists $\mathcal{X}_i \rightarrow [1, \infty)$ Lyapunov's function \mathbf{V}_i such that

- $\max_{1 \leq i \leq N} \sup_{x \in C_i} \mathbf{V}_i(x) < \infty$.
- There exist $0 < \rho < 1$ and $\mathbf{m}^* > 0$ such that for all $1 \leq i \leq N$, $x \in \mathcal{X}_i$ and $\theta_i \in \Theta_i$,

$$\mathbb{E}_x^*[\mathbf{V}_i(X_{i,1})] \leq (1-\rho)\mathbf{V}_i(x) + \mathbf{m}^* \mathbf{1}_{C_i}(x). \quad (7.12)$$

- There exists $\mathbf{q} > 0$ such that for any $x_i \in \mathcal{X}_i$

$$\max_{1 \leq i \leq N} \sup_{j \geq 1} \mathbb{E}_{x_i}^* [\mathbf{V}_i^{\mathbf{q}}(X_{i,j})] < \infty. \quad (7.13)$$

3. The functions $\check{g}_{i,l}(y, x)$ in (7.8) are such that for any $1 \leq i \leq N$ and $x \in \mathcal{X}_i$

$$\max_{1 \leq l \leq m} \left| \int_{\mathcal{X}_i} \check{g}_{i,l}(y, x) f_i^*(y|x) \mu_i(dy) \right| \leq \mathbf{V}_i(x) \quad \text{and} \quad \sup_{j \geq 1} \mathbb{E}_x^* |\check{g}_{i,l}(X_{i,j}, X_{i,j-1})|^q < \infty.$$

Proposition 1 from [12] provides the following result.

Proposition 3. The condition (C₃^{*}(q)) implies the condition (C₃(q)).

8. Examples

8.1. Example 1: Change in the parameters of multivariate linear difference equations

Consider the multivariate models in \mathbb{R}^p given by

$$X_{i,n} = \left(\Gamma_{i,n}^* \mathbb{1}_{\{n \leq \nu\}} + \Gamma_{i,n} \mathbb{1}_{\{n > \nu\}} \right) X_{i,n-1} + w_{i,n}, \quad (8.1)$$

where $\Gamma_{i,n}^*$ and $\Gamma_{i,n}$ are $p \times p$ random matrixes and $(w_{i,n})_{n \geq 1}$ is an i.i.d. sequence of Gaussian random vectors $\mathcal{N}(0, Q_i^*)$ in \mathbb{R}^p with the positive definite $p \times p$ matrix Q_i^* . Assume also that

$$\Gamma_{i,n}^* = \theta_i^* + B_{i,n} \quad \text{and} \quad \Gamma_{i,n} = \theta_i + B_{i,n} \quad (8.2)$$

and $(B_{i,n})_{n \geq 1}$ are i.i.d. Gaussian random matrices $\mathcal{N}(0, Q_i)$, where the $p^2 \times p^2$ matrix $Q_i = \mathbf{E}[B_{i,1} \otimes B_{i,1}]$ is positive defined. Assume, in addition, that all eigenvalues of the matrix

$$\mathbf{E}[\Gamma_{i,1}^* \otimes \Gamma_{i,1}^*] = \theta_i^* \otimes \theta_i^* + Q_i$$

are less than one in module. Define

$$\Theta_i^{st} = \{\theta_i \in \mathbb{R}^{p^2} : \max_{1 \leq j \leq p_i^4} \mathbf{e}_j(\theta_i \otimes \theta_i + Q_i) < 1\} \setminus \{\theta_i^*\}, \quad (8.3)$$

where $\mathbf{e}_j(\Gamma)$ is the j th eigenvalue of matrix Γ , and assume further that in (8.2) the matrices $\theta_i \in \Theta_i^{st}$. In this case, the processes $(X_{i,n})_{n \geq 1}$ (for $\nu = \infty$) and $(X_{i,n})_{n > \nu}$ (for $\nu < \infty$) are ergodic with the ergodic distributions given by the vectors [[6]]

$$\varsigma_i^* = \sum_{l \geq 1} \prod_{j=1}^{l-1} \Gamma_{i,j}^* w_{i,l} \quad \text{and} \quad \varsigma_{i,\theta} = \sum_{l \geq 1} \prod_{j=1}^{l-1} \Gamma_{i,j} w_{i,l}$$

i.e., the corresponding invariant measures $\lambda_{*,i}$ and $\lambda_{\theta,i}$ on \mathbb{R}^p are defined as

$$\lambda_{*,i}(dx) = \mathbf{P}(\varsigma_i^* \in dx) \quad \text{and} \quad \lambda_{\theta,i}(dx) = \mathbf{P}(\varsigma_{i,\theta} \in dx)$$

Note that in this case the Markov processes $(X_{i,n})_{n \leq \nu}$ and $(X_{i,n})_{n > \nu}$ have the following transition densities in \mathbb{R}^p , $\theta_i \in \Theta_i$

$$f_i^*(y|x) = \frac{\exp\left\{-\frac{|\eta_i^*(y,x)|^2}{2}\right\}}{(2\pi)^{p/2} \sqrt{\det(G_i(x))}} \quad \text{and} \quad f_{i,\theta_i}(y|x) = \frac{\exp\left\{-\frac{|\eta_{\theta_i,i}(y,x)|^2}{2}\right\}}{(2\pi)^{p/2} \sqrt{\det(G_i(x))}}, \quad (8.4)$$

where $\eta_i^*(y, x) = G_i^{-1/2}(x)(y - \theta_i^* x)$, $\eta_{\theta_i,i}(y, x) = G_i^{-1/2}(x)(y - \theta_i x)$, $\theta_i \in \Theta_i$ and

$$G_i(x) = \mathbf{E}[B_{i,1} x x' B_{i,1}'] + Q_i^* = Q_i \text{Vect}(x x') + Q_i^*.$$

Therefore, in this case,

$$g_i(\theta_i, y, x) = \log \frac{f_{i,\theta_i}(y|x)}{f_i^*(y|x)} = y' G_i^{-1}(x)(\theta_i - \theta_i^*) x + \frac{x' (\theta_i^*)' G_i^{-1}(x) \theta_i^* x - x' \theta_i' G_i^{-1}(x) \theta_i x}{2}.$$

Now we set

$$\mathbf{L}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \int_{\Theta_i} e^{\sum_{j=k+1}^n g_i(\theta_i, X_{i,j}, X_{i,j-1})} W_i(d\theta_i) \quad (8.5)$$

and

$$\widehat{\mathbf{L}}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \sup_{\theta \in \Theta_i} e^{\sum_{j=k+1}^n g_i(\theta_i, X_{i,j}, X_{i,j-1})}, \quad (8.6)$$

where ϱ^{opt} is defined in (6.10). The random $N \times N$ matrix (4.3) has the following form

$$\langle \mathbf{U}_n \rangle_{i,j} = \frac{\mathbf{L}_{i,n}}{\widehat{\mathbf{L}}_{j,n}} \quad \text{if } i \neq j \quad \text{and} \quad \langle \mathbf{U}_n \rangle_{i,i} = \frac{\mathbf{L}_{i,n}}{(1 - \varrho^{opt})^n}, \quad (8.7)$$

and the corresponding change detection-identification procedure $\delta_\beta^{opt} = (T_\beta^{opt}, d_\beta^{opt})$ is defined by (6.13)-(6.14) with the threshold matrix $A = A_\beta^{opt}$ given by (6.12).

As shown in [13], conditions (C_1) and $(C_2(r))$ hold for any $r > 0$. Moreover, one can calculate directly that

$$\bar{J}_i(\theta_i) = \frac{1}{2} \mathbf{E} \left[\varsigma'_{i,\theta_i} (\theta_i - \theta_i^*)' G_i^{-1}(\varsigma_{i,\theta_i}) (\theta_i - \theta_i^*) \varsigma_{i,\theta_i} \right]$$

and

$$\bar{J}_i^*(\theta_i) = -\frac{1}{2} \mathbb{E} \left[(\zeta_i^*)' (\theta_i - \theta_i^*)' G_i^{-1} (\zeta_i^*) (\theta_i - \theta_i^*) \zeta_i^* \right].$$

To check the condition (\mathbf{C}_3) denote $\bar{\theta}_i = \theta_i^* - u$. It can be easily shown that for $1 \leq i \leq N$

$$\begin{aligned} g_i(u, y, x) - \bar{J}_i^*(u) &= - \sum_{s_1, s_2=1}^p \langle \bar{\theta}_i \rangle_{s_1, s_2} \sum_{k=1}^p D_{s_1, s_2, k}^{(i,1)}(x) \langle \eta_i^*(y, x) \rangle_k \\ &\quad - \frac{1}{2} \sum_{s_1, s_2, s_3, s_4=1}^p \langle \bar{\theta}_i \rangle_{s_1, s_2} \langle \bar{\theta}_i \rangle_{s_3, s_4} D_{s_1, s_2, s_3, s_4}^{(i,2)}(x), \end{aligned}$$

where $D_{s_1, s_2, k}^{(i,1)}(x) = \langle G_i^{-1/2}(x) \rangle_{s_1, k} \langle x \rangle_{s_2}$, $D_{s_1, s_2, s_3, s_4}^{(i,2)}(x) = \mathbf{b}_{s_1, s_2, s_3, s_4}(x) - \mathbb{E}^*[\mathbf{b}_{s_1, s_2, s_3, s_4}(\zeta_i^*)]$ and

$$\mathbf{b}_{s_1, s_2, s_3, s_4}(x) = \langle x \rangle_{s_2} \langle x \rangle_{s_4} \langle G_i^{-1}(x) \rangle_{s_1, s_3}.$$

Note that for any $z \in \mathbb{R}^p$ and $|z| = 1$

$$z' G_i(x) z = |x|^2 (\text{Vect}(\bar{x}z'))' Q_i \text{Vect}(\bar{x}z') + z' Q_i^* z,$$

where $\bar{x} = x/|x|$. Taking into account that the matrices Q_i are positive definite, we obtain that for some $\mathbf{c}_* > 0$

$$\max_{1 \leq i \leq N} |G_i^{-1}(x)| \leq \frac{\mathbf{c}_*}{1 + |x|^2}. \quad (8.8)$$

Therefore, the functions $D_{s_1, s_2, k}^{(i,1)}(x)$ and $\mathbf{b}_{s_1, s_2, s_3, s_4}(x)$ are bounded. Moreover, as shown in [13] (Example 1), the Lyapunov function for the process (8.1) (with $\nu = +\infty$) has the form

$$\mathbf{V}(x) = \mathbf{v}^* \left(1 + (x' T x)^\delta \right)$$

for some constant $\mathbf{v}^* \geq 1$, a fixed matrix T and any $\delta > 0$. Since in this case $\max_{1 \leq i \leq N} \sup_{j \geq 1} \mathbb{E}^*[X_{i,j}^2] < \infty$, we obtain that condition $(\mathbf{C}_3^*(\mathbf{q}))$ holds true for any $\mathbf{q} > 0$. Now, taking into account that under \mathbf{P}^* the random vectors $(\eta_i^*(X_{i,j}, X_{i,j-1}))_{1 \leq i \leq N, j \geq 1}$ are i.i.d. $(0, \mathcal{I}_p)$ Gaussian (\mathcal{I}_p is the unity matrix in \mathbb{R}^p), we obtain the condition $(\mathbf{C}_3(q))$ for any $q > 0$ using Proposition 3. Therefore, Theorems 3 and 4 imply that the sequential procedure δ_β^{opt} defined in (6.13)-(6.14) is asymptotically optimal and robust in the pointwise and minimax senses for any compact sets $\Theta_i \subset \Theta_i^{st}$ and for any $r > 0$.

8.2. Example 2: Change in the correlation coefficients of autoregressive models

Consider the problem of detecting the change of the correlation coefficient in the p th order AR process which in the i th stream satisfies the recursion

$$X_{i,n} = \mathbf{a}_{i,1}^{(n)} X_{i,n-1} + \dots + \mathbf{a}_{i,p}^{(n)} X_{i,n-p} + w_{i,n}, \quad \text{for } n \geq 1, \quad (8.9)$$

where $\mathbf{a}_{i,l}^{(n)} = \theta_l^* \mathbb{1}_{\{n \leq \nu\}} + \theta_{i,l} \mathbb{1}_{\{n > \nu\}}$ and $(w_{i,n})_{n \geq 1}$ are i.i.d. Gaussian random variables with $\mathbb{E}[w_{i,1}] = 0$, $\mathbb{E}[w_{i,1}^2] = 1$. In the sequel, we use the notation $\theta_i^* = (\theta_{i,1}^*, \dots, \theta_{i,p}^*)'$ and $\theta_i = (\theta_{i,1}, \dots, \theta_{i,p})'$. Hereafter \prime denotes the transposition operation. The corresponding conditional densities $X_{i,n} | X_{i,n-1}, \dots, X_{i,n-p}$ for $\nu \leq n$ and $\nu > n$, $\theta_i \in \Theta_i$, are

$$f_i^*(y|x) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{(\eta_i^*(y, x))^2}{2} \right\} \quad \text{and} \quad f_{i,\theta_i}(y|x) = \frac{1}{(2\pi)^{p/2}} \exp \left\{ -\frac{(\eta_{\theta_i}(y, x))^2}{2} \right\}, \quad (8.10)$$

where $\eta_i^*(y, x) = y - (\theta_i^*)' x$ and $\eta_{\theta_i}(y, x) = y - (\theta_i)' x$. Therefore, for any $\theta_i \in \mathbb{R}^p$, $y \in \mathbb{R}$ and $x = (x_1, \dots, x_p)' \in \mathbb{R}^p$

$$g_i(\theta_i, y, x) = \log \frac{f_{i,\theta_i}(y|x)}{f_i^*(y|x)} = y_1 (\theta_i - \theta_i^*)' x + \frac{((\theta_i^*)' x)^2 - (\theta_i' x)^2}{2}. \quad (8.11)$$

The process (8.9) is not Markov, but the p -dimensional processes

$$\Phi_{i,n} = (X_{i,n}, \dots, X_{i,n-p+1})' \in \mathbb{R}^p, \quad i = 1, \dots, N \quad (8.12)$$

are Markov. Now, for any $\vartheta = (\vartheta_1, \dots, \vartheta_p) \in \mathbb{R}^p$ we define

$$\Lambda(\vartheta) = \begin{pmatrix} \vartheta_1 & \vartheta_2 & \dots & \vartheta_p \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Using this matrix it is easy to show that the processes $(\Phi_{i,n})_{n \leq \nu}$ and $(\Phi_{i,n})_{n > \nu+p}$ satisfy the following stochastic linear equations:

$$\Phi_{i,n} = \Lambda_{*,i} \Phi_{i,n-1} + \widetilde{w}_{i,n} \quad \text{for } n \leq \nu \quad \text{and} \quad \Phi_{i,n} = \Lambda_i \Phi_{i,n-1} + \widetilde{w}_{i,n} \quad \text{for } n > \nu, \quad (8.13)$$

where $\Lambda_{*,i} = \Lambda(\theta_i^*)$, $\Lambda_i = \Lambda(\theta_i)$ and $\widetilde{w}_{i,n} = (w_{i,n}, 0, \dots, 0)' \in \mathbb{R}^p$. Obviously,

$$\mathbb{E}[\widetilde{w}_n \widetilde{w}_n'] = B = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}.$$

Assume that all eigenvalues of the matrices $\Lambda_{*,i} = \Lambda_i(\theta_i^*)$ in modules are less than 1 and that θ_i belongs to the set

$$\Theta_i^{st} = \{\vartheta \in \mathbb{R}^p : \max_{1 \leq j \leq p} |\mathbf{e}_j(\Lambda_i(\vartheta))| < 1\} \setminus \{\theta_i^*\}, \quad (8.14)$$

where $\mathbf{e}_j(\Lambda)$ is the i th eigenvalue of the matrix Λ . In this case, the processes (8.13) have the ergodic distributions defined by the random vectors

$$s_i^* = \sum_{l \geq 1} (\Lambda_{*,i})^{l-1} \widetilde{w}_{i,l} \quad \text{and} \quad s_{i,\vartheta} = \sum_{l \geq 1} (\Lambda_i)^{l-1} \widetilde{w}_{i,l}$$

which are $(0, F^*)$ and $(0, F_i(\vartheta))$ Gaussian vectors in \mathbb{R}^p , where

$$F^* = \sum_{n \geq 0} (\Lambda_{*,i})^n B (\Lambda_{*,i}')^n \quad \text{and} \quad F_i(\vartheta) = \sum_{n \geq 0} \Lambda_i^n B (\Lambda_i')^n.$$

Now we set

$$\mathbf{L}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \int_{\Theta_i} e^{\sum_{j=k+1}^n g_i(\vartheta, X_{i,j}, \Phi_{i,j-1})} W_i(d\vartheta) \quad (8.15)$$

and

$$\widehat{\mathbf{L}}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \sup_{\vartheta \in \Theta_i} e^{\sum_{j=k+1}^n g_i(\vartheta, X_{i,j}, \Phi_{i,j-1})}, \quad (8.16)$$

where ϱ^{opt} is defined in (6.10). The random $N \times N$ matrix (4.3) has the following form

$$\langle \mathbf{U}_n \rangle_{i,j} = \frac{\mathbf{L}_{i,n}}{\widehat{\mathbf{L}}_{j,n}} \quad \text{if } i \neq j \quad \text{and} \quad \langle \mathbf{U}_n \rangle_{i,i} = \frac{\mathbf{L}_{i,n}}{(1 - \varrho^{opt})^n}. \quad (8.17)$$

and the corresponding change detection-identification procedure $\delta_\beta^{opt} = (T_\beta^{opt}, d_\beta^{opt})$ is defined in (6.13)-(6.14) with the threshold matrix $A = A_\beta^{opt}$ given by (6.12).

As shown in [13], conditions (C_1^r) and $(C_2(r))$ hold for any $r > 0$ and any compact sets $\Theta_i \subset \Theta_i^{st}$ for the function $I_{i,j}$ defined in (7.9) with

$$\bar{J}_i(\vartheta) = \frac{1}{2} (\vartheta - \theta_i^*)' F(\vartheta) (\vartheta - \theta_i^*) \quad \text{and} \quad \bar{J}_i(\vartheta) = -\frac{1}{2} (\vartheta - \theta_i^*)' F^* (\vartheta - \theta_i^*). \quad (8.18)$$

It should be noted that in the scalar case, i.e., when $p_1 = \dots = p_N = 1$,

$$\bar{J}_i(\vartheta) = \frac{(\vartheta - \theta_i^*)^2}{2(1 - \vartheta^2)} \quad \text{and} \quad \bar{J}_i(\vartheta) = -\frac{(\vartheta - \theta_i^*)^2}{2(1 - (\vartheta_i^*)^2)}.$$

Write $\bar{\theta}_i = \theta_i^* - u$. To check the condition (C_3) direct calculations show that for $1 \leq i \leq N$

$$g_i(u, y, x) - \bar{J}_i(u) = -\bar{\theta}_i' \eta_i(y, x) - \frac{1}{2} \bar{\theta}_i' (xx' - \mathbb{E}[S_{i,1}^* (S_{i,1}^*)']) \bar{\theta}_i.$$

Therefore, taking into account that in this case $\max_{1 \leq i \leq N} \sup_{j \geq 1} \mathbb{E}^* |X_{i,j}|^q < \infty$ for any $q > 0$ we obtain that condition $(C_3^*(\mathbf{q}))$ holds for any $\mathbf{q} > 0$. Now, taking into account that under the probability measure \mathbb{P}^* the random variables $(\eta_i^*(X_{i,j}, \Phi_{i,j-1}))_{1 \leq i \leq N, j \geq 1}$ are i.i.d. $\mathcal{N}(0, 1)$ we obtain the condition $(C_3(q))$ for any $q > 0$ using Proposition 3. Therefore, Theorems 3 and 4 imply that the sequential procedure δ_β^{opt} is asymptotically optimal and robust in the pointwise and minimax senses for any compact sets $\Theta_i \subset \Theta_i^{st}$ and for any $r > 0$.

9. Application to epidemics detection and localization

9.1. Near optimality

We begin with considering the epidemiological statistical models proposed in [1]. Assume that for any $1 \leq i \leq N$ the observations $(X_{i,n})_{1 \leq n \leq \nu}$ and $(X_{i,n})_{n > \nu}$ are homogenous Markov processes with the values in the finite space (\mathcal{X}, μ) , $\mathcal{X}_i = \{0, \dots, D\}$ and $\mu\{0\} = \dots = \mu\{D\} = 1$. In this model, the conditional $X_{i,n} | X_{i,n-1}$ densities for $n \leq \nu$ and for $n > \nu$ are defined respectively as

$$f_i^*(y|x) = \binom{x}{y} (p_i^*)^{x-y} (1 - p_i^*)^y \quad \text{and} \quad f_{i,\theta_i}(y|x) = \binom{x}{y} \theta_i^{x-y} (1 - \theta_i)^y, \quad (9.1)$$

where $0 < p_i^* < 1$ and $\theta_i \in \Theta_i \subset [0, 1]$. The probabilities p_i^* are non-epidemic (normal) infection rates and the Θ_i are the sets of epidemic values of the infection parameters θ_i . In this case, the functions g_i defined in (7.1) for any $0 < \theta_i < 1$, $x, y \in \mathcal{X}$ have the following forms

$$g_i(\theta_i, y, x) = \log \frac{f_{i,\theta_i}(y|x)}{f_i^*(y|x)} = (x - y) \log \frac{\theta_i}{p_i^*} + y \log \frac{1 - \theta_i}{1 - p_i^*}, \quad i = 1, \dots, N. \quad (9.2)$$

So the functions (7.4) are

$$J_i(\theta_i, x) = \sum_{y=0}^x \left((x - y) \log \frac{\theta_i}{p_i^*} + y \log \frac{1 - \theta_i}{1 - p_i^*} \right) \binom{x}{y} \theta_i^{x-y} (1 - \theta_i)^y$$

and

$$J_i^*(\theta_i, x) = \sum_{y=0}^x \left((x - y) \log \frac{\theta_i}{p_i^*} + y \log \frac{1 - \theta_i}{1 - p_i^*} \right) \binom{x}{y} (p_i^*)^{x-y} (1 - p_i^*)^y.$$

One can check directly that the set $C = \{0\}$ is an accessible atom for the Markov chains $(X_{i,n})_{n > \nu}$. Obviously, if $X_{i,n-1} = 0$ then $X_{i,n} = 0$ almost surely. Define the Markov time

$$\tau = \inf\{n \geq 1 : X_{i,n} = 0\}.$$

If $X_{i,0} = 0$ then $\tau = 1$, i.e., $\mathbb{E}_{i,x=0}^{\theta_i} [\tau] = 1$. Therefore, for any $0 < \theta_i < 1$ the chain is ergodic with the ergodic distribution $\lambda^{\theta_i}(\Gamma) = \mathbf{1}_{\{0 \in \Gamma\}}$ for any $\Gamma \subseteq \mathcal{X}$ (point measure). See, e.g., Theorems 10.2.1 and 10.2.2 in [8]. In this case, for any compact sets $\Theta_i \subset \Theta$

$$\min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \inf_{x, y \in C} f_{i,\theta_i}(y|x) = 1.$$

Let us select

$$\mathbf{V}(x) = \mathbf{V}_* e^{\gamma x}, \quad (9.3)$$

where $\gamma > 0$ and $\mathbf{V}_* \geq 1$. For any $x \geq 1$ we have

$$\begin{aligned} \frac{\mathbb{E}_x^{\theta_i}[\mathbf{V}(X_1)]}{\mathbf{V}(x)} &= \sum_{y=0}^x \frac{\mathbf{V}(y)}{\mathbf{V}(x)} \binom{x}{y} \theta_i^{x-y} (1-\theta_i)^y \\ &= (1-\theta_i)^x + \sum_{y=0}^{x-1} \frac{\mathbf{V}(y)}{\mathbf{V}(x)} \binom{x}{y} \theta_i^{x-y} (1-\theta_i)^y \\ &\leq (1-\theta_i)^x + e^{-\gamma} \sum_{y=0}^{x-1} \binom{x}{y} \theta_i^{x-y} (1-\theta_i)^y = (1-e^{-\gamma})(1-\theta_i)^x + e^{-\gamma} \\ &\leq (1-e^{-\gamma})(1-\theta_i) + e^{-\gamma}. \end{aligned}$$

So, if we take $\gamma = \log 2$ in (9.3), we obtain that for $1 \leq i \leq N$ and $\theta_i \in \Theta_i$

$$\frac{\mathbb{E}_x^{\theta_i}[\mathbf{V}(X_{i,1})]}{\mathbf{V}(x)} \leq \frac{2-\theta_i}{2} \leq 1-\rho \quad \text{with} \quad \rho = \min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \frac{\theta_i}{2}.$$

By Theorem A1 in [13], the Markov chain $(X_{i,n})_{n \geq 1}$ is uniformly geometric ergodic and for some positive constants κ^* and R^*

$$\sup_{n \geq 0} e^{\kappa^* n} \max_{1 \leq i \leq N} \sup_{x \in X_i} \sup_{\theta_i \in \Theta_i} \sup_{0 \leq g \leq \mathbf{V}} \frac{1}{\mathbf{V}(x)} |\mathbb{E}_{i,x}^{\theta_i} [g(X_{i,n})] - g(0)| \leq R^*.$$

Therefore, Theorem 5 implies conditions (\mathbf{A}_1) and $(\mathbf{A}_2(r))$ with $\bar{J}_i(\theta_i) = J_i(\theta_i, 0) = 0$ and $\bar{J}_j^*(\theta_j) = J_j^*(\theta_j, 0) = 0$ for all $r > 0$. This means that we cannot use the procedures (4.4) for this problem directly. However, in practice the values of the observations $X_{i,n}$ are sufficiently large, i.e., $D \rightarrow \infty$, and usually the number of the infected populations is not too large, i.e., $X_{i,n} \geq \epsilon D$ for some $0 < \epsilon < 1$. So it is more natural to modify the initial model and study the limiting model when D is sufficiently large. Note that in this case observations in the binomial models (9.1) can be represented as

$$X_{i,n} = (1-\vartheta)X_{i,n-1} + \sum_{j=1}^{X_{i,n-1}} (\eta_{n,j} - 1 + \vartheta),$$

where $(\eta_{n,j})_{j \geq 1}$ is i.i.d. sequence of Bernoulli random variables with $\mathbf{P}(\eta_{n,j} = 1) = 1 - \vartheta$ and independent from $X_{i,n-1}$ and where $\vartheta = \theta_i$ and p_i^* in the post-change and pre-change modes, respectively. Using the Gaussian approximation for the last sum

$$\frac{1}{\sqrt{X_{i,n-1}}} \sum_{j=1}^{X_{i,n-1}} (\eta_{n,j} - 1 + \vartheta) \sim \mathcal{N}(0, \sigma_\vartheta^2), \quad \sigma_\vartheta^2 = \vartheta(1-\vartheta),$$

we obtain the following model

$$X_{i,n} = (1-\vartheta)X_{i,n-1} + \sigma_\vartheta \sqrt{|X_{i,n-1}|} \xi_{i,n},$$

where $(\xi_{i,n})_{n \geq 1}$ is the sequence of i.i.d. normal $\mathcal{N}(0, 1)$ random variables.

Thus, in place of the original Bernoulli model we will use the following model: the observations $X_{i,k}$ before change are defined as

$$X_{i,n} = (1-p_i^*)X_{i,n-1} + \sigma_i^* \sqrt{|X_{i,n-1}|} \xi_{i,n}, \quad \sigma_i^* = \sqrt{p_i^*(1-p_i^*)}, \quad (9.4)$$

and after change as

$$X_{i,n} = (1-\theta_i)X_{i,n-1} + \sigma_{\theta_i} \sqrt{|X_{i,n-1}|} \xi_{i,n}, \quad (9.5)$$

where $\theta_i \in \Theta_i \subset [0, 1]$ and $(\xi_{i,n})_{n \geq 1}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables. In this case, the spaces $(\mathcal{X}_i, \mathcal{B}_i, \mu_i)$ are: $\mathcal{X} = \mathbb{R}_* = \mathbb{R} \setminus \{0\}$, $\mathcal{B}_i = \mathcal{B}(\mathbb{R}_*)$ is the Borel field and $\mu_i = \mu$ is the Lebesgue measure on $\mathcal{B}(\mathbb{R}_*)$. Obviously,

$$f_i^*(y|x) = \frac{1}{\sigma_i^* \sqrt{2\pi|x|}} \exp\left\{-\frac{[y - (1 - p^*)x]^2}{2(\sigma_i^*)^2|x|}\right\} \quad \text{and} \quad f_{\theta_i}(y|x) = \frac{1}{\sigma_{\theta_i} \sqrt{2\pi|x|}} \exp\left\{-\frac{[y - (1 - \theta_i)x]^2}{2\sigma_{\theta_i}^2|x|}\right\}. \quad (9.6)$$

Using definitions (7.2) and (7.4) we obtain that for $x \in \mathbb{R}_*$

$$g_i(\theta_i, y, x) = \log \frac{\sigma_i^*}{\sigma_{\theta_i}} + \frac{(\eta_i^*(y, x))^2}{2} - \frac{(\eta_{\theta_i, i}(y, x))^2}{2}, \quad (9.7)$$

where

$$\eta_i^*(y, x) = \frac{y - (1 - p^*)x}{\sigma_i^* \sqrt{|x|}} \quad \text{and} \quad \eta_{\theta_i, i}(y, x) = \frac{y - (1 - \theta_i)x}{\sigma_{\theta_i} \sqrt{|x|}}.$$

In this case,

$$\mathbf{L}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \int_{\Theta_i} e^{\sum_{j=k+1}^n g_i(\theta_i, X_{i,j}, X_{i,j-1})} W_i(d\theta_i) \quad (9.8)$$

and

$$\widehat{\mathbf{L}}_{i,n} = \varrho^{opt} \sum_{k=0}^{n-1} (1 - \varrho^{opt})^k \sup_{\theta \in \Theta_i} e^{\sum_{j=k+1}^n g_i(\theta, X_{i,j}, X_{i,j-1})}, \quad (9.9)$$

where ϱ^{opt} is defined in (6.10). The elements of the random $N \times N$ matrix (4.3) have the following form

$$\langle \mathbf{U}_n \rangle_{i,j} = \frac{\mathbf{L}_{i,n}}{\mathbf{L}_{j,n}} \quad \text{if } i \neq j \quad \text{and} \quad \langle \mathbf{U}_n \rangle_{i,i} = \frac{\mathbf{L}_{i,n}}{(1 - \varrho^{opt})^n}, \quad (9.10)$$

and the corresponding change detection-identification procedure $\delta_\beta^{opt} = (T_\beta^{opt}, d_\beta^{opt})$ is defined in (6.13)-(6.14) with the threshold matrix $A = A_\beta^{opt}$ given by (6.12).

Let us check conditions $(\mathbf{C}_1) - (\mathbf{C}_3)$. To this end, first note that

$$J_i(\theta_i, x) = \frac{1}{2} \left(\log \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} - 1 + \frac{\theta_i(1 - \theta_i)}{p_i^*(1 - p_i^*)} + \frac{(\theta_i - p_i^*)^2}{p_i^*(1 - p_i^*)} |x| \right) \quad (9.11)$$

and

$$J_i^*(\theta_i, x) = \frac{1}{2} \left(\log \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} + 1 - \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} - \frac{(\theta_i - p_i^*)^2}{\theta_i(1 - \theta_i)} |x| \right). \quad (9.12)$$

Taking into account that the function $z - \log z - 1 > 0$ for all $z > 0$ and $z \neq 1$ we obtain that $\inf_x J_i(\theta_i, x) > 0$ and $\sup_x J_i(\theta_i, x) < 0$ for all $\theta_i \neq p_i^*$. Recall that p_i^* and θ_i are the infection rates, where p^* is normal non-epidemic value and θ_i is epidemic value. So, if \check{p}_i is the epidemic threshold for the i th stream, then $\Theta_i \subset (\check{p}_i, 1)$ and for all $p_i^* < \check{p}_i$

$$\min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \inf_x J_i(\theta_i, x) > 0. \quad (9.13)$$

Now we need to check the conditions of Theorem 5. First, note that the definition (9.6) yields that for any $\mathbf{B} > 0$

$$\min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \inf_{|y| \leq \mathbf{B}} \inf_{|x| \leq \mathbf{B}} f_{\theta_i}(y|x) > 0.$$

From (9.11) it is easy to deduce that for any compact sets $\Theta_i \subset (\check{p}_i, 1)$

$$\mathbf{g}^* = \max_{1 \leq i \leq N} \max_{x \in \mathbb{R}} \sup_{\theta_i \in \Theta_i} \frac{J_i(\theta_i, x)}{1 + |x|} < \infty. \quad (9.14)$$

Moreover, note that

$$\frac{\partial g_i(\vartheta, y, x)}{\partial \vartheta} = \frac{2\vartheta - 1}{\vartheta(1 - \vartheta)} - \frac{(y - (1 - \vartheta)x)x}{\vartheta(1 - \vartheta)|x|} + \frac{(y - (1 - \vartheta)x)^2}{2\vartheta^2(1 - \vartheta)^2|x|}$$

and, therefore, using the definition (7.2) we obtain

$$h(x, y) = \max_{1 \leq i \leq N} \max_{u \in \Theta_i} |\partial g_i(u, y, x) / \partial u| \leq \check{h} \left(1 + |x| + |y| + \frac{y^2}{|x|} \right), \quad (9.15)$$

where $\check{h} = \max_{1 \leq i \leq N} \max_{\vartheta \in \Theta_i} \vartheta^{-2}(1 - \vartheta)^{-2}$. Therefore, taking into account, that in this case $f_{i,\vartheta}(y|x) = f_\vartheta(y|x)$ for all $1 \leq i \leq N$, we get for any $\vartheta \in \Theta_i$

$$\begin{aligned} \bar{h}_i(\vartheta, y) &= \bar{h}(\vartheta, y) = \int_{\mathbb{R}} h(y, x) f_\vartheta(y|x) dy \leq \check{h} \left(1 + |x| + \int_{\mathbb{R}} |y| f_\vartheta(y|x) dy + \frac{1}{|x|} \int_{\mathbb{R}} y^2 f_\vartheta(y|x) dy \right) \\ &\leq \check{h} \left(1 + |x| + (1 - \vartheta)|x| + \sigma_\vartheta \sqrt{|x|} + (1 - \vartheta)^2|x| + \sigma_\vartheta^2|x| \right) \leq 4\check{h}(1 + |x|). \end{aligned}$$

To check the conditions (C1.2) we set

$$\mathbf{V}(x) = \mathbf{V}^*(1 + |x|) \quad \text{and} \quad \mathbf{V}^* = \mathbf{g}^* + 4\check{h}.$$

For the model (9.5) we have

$$\mathbf{E}_{i,x}^{\theta_i} [\mathbf{V}(X_{i,1})] \leq \mathbf{V}^* \left(1 + (1 - \theta_i)|x| + \sigma_{\theta_i} \sqrt{|x|} \mathbf{E}|\xi_1| \right),$$

i.e.,

$$\limsup_{\mathbf{B} \rightarrow \infty} \sup_{|x| > \mathbf{B}} \max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} \frac{\mathbf{E}_{i,x}^{\theta_i} [\mathbf{V}(X_{i,1})]}{\mathbf{V}(x)} \leq 1 - \min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \theta_i.$$

Therefore, there exists $\mathbf{B} > 0$ such that for all $|x| > \mathbf{B}$

$$\max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} \mathbf{E}_{i,x}^{\theta_i} [\mathbf{V}(X_{i,1})] \leq (1 - \rho)\mathbf{V}(x) \quad \text{and} \quad \rho = \frac{1}{2} \min_{1 \leq i \leq N} \inf_{\theta_i \in \Theta_i} \theta_i.$$

Obviously, this inequality implies condition (C₁) with $C = \{x \in \mathbb{R} : |x| \leq \mathbf{B}\}$. Using Theorem 15.01 in [8] and Proposition A.1 in the Appendix, it is easy to deduce that the processes (9.4) and (9.5) are stationary with the ergodic distributions defined by the random variables ζ_i^* and $\zeta_{\theta_i,i}$ such that for any $\mathbf{q} > 0$

$$\mathbf{E}|\zeta_i^*|^{\mathbf{q}} < \infty \quad \text{and} \quad \mathbf{E}|\zeta_{\theta_i,i}|^{\mathbf{q}} < \infty. \quad (9.16)$$

Hence, from (9.11) and (9.12) we get

$$\bar{J}_i(\theta_i) = \frac{1}{2} \left(\log \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} - 1 + \frac{\theta_i(1 - \theta_i)}{p_i^*(1 - p_i^*)} + \frac{(\theta_i - p_i^*)^2}{p_i^*(1 - p_i^*)} \mathbf{E}|\zeta_{\theta_i,i}| \right) \quad (9.17)$$

and

$$\bar{J}_i^*(\theta_i) = \frac{1}{2} \left(\log \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} + 1 - \frac{p_i^*(1 - p_i^*)}{\theta_i(1 - \theta_i)} - \frac{(\theta_i - p_i^*)^2}{\theta_i(1 - \theta_i)} \mathbf{E}|\zeta_i^*| \right). \quad (9.18)$$

As far as the condition (C₂) is concerned it follows from (9.7) and (9.15) that there exists a constant $\mathbf{c}_* > 0$ such that for all $x, y \in \mathbb{R}_*$

$$\max_{1 \leq i \leq N} \sup_{\theta_i \in \Theta_i} (|g_i(\theta_i, y, x)| + h_i(x, y)) \leq \mathbf{c}_* \left(1 + |x| + |y| + \frac{y^2}{|x|} \right).$$

Note that for any $m \geq 1$ there exists a constant $C_m > 0$ for which for any $1 \leq i \leq N$ and $\theta_i \in \Theta_i$

$$\mathbf{E}_{i,\theta_i} [X_{i,n}^{2m} | X_{i,n-1}] \leq C_m (X_{i,n-1}^{2m} + |X_{i,n-1}|^m).$$

Hence, for any $m \geq 1$ there exists a constant $C_m > 0$ such that

$$\mathbb{E}_{i,x}^{\theta_i} [g_i^{2m}(\theta_i, X_{i,n}, X_{i,n-1})] \leq C_m \left(1 + \mathbb{E}_{i,x}^{\theta_i} [X_{i,n-1}^{2m}] + \mathbb{E}_{i,x}^{\theta_i} [|X_{i,n-1}|^m]\right) \leq C_m \left(1 + \mathbb{E}_{i,x}^{\theta_i} [X_{i,n-1}^{2m}]\right)$$

and

$$\mathbb{E}_{i,x}^{\theta_i} [h_i^{2m}(X_{i,n}, X_{i,n-1})] \leq C_m \left(1 + \mathbb{E}_{i,x}^{\theta_i} [X_{i,n-1}^{2m}]\right).$$

To check the condition (\mathbf{C}_3) we can obtain directly from (9.18) that for $1 \leq i \leq N$

$$\begin{aligned} g_i(u, y, x) - \bar{J}_i^*(u) &= \frac{1}{2} \left(1 - \frac{p_i^*(1-p_i^*)}{u(1-u)}\right) ((\eta_i^*(y, x))^2 - 1) - \frac{(u-p_i^*)^2}{2u(1-u)} (|x| - \mathbb{E}[|\zeta_i^*|]) \\ &\quad - \sqrt{\frac{p_i^*(1-p_i^*)}{u(1-u)}} (u-p_i^*) x \eta_i^*(y, x). \end{aligned}$$

By Proposition A.1 (see the appendix) $\max_{1 \leq i \leq N} \sup_{j \geq 1} \mathbb{E}^* |X_{i,j}|^q < \infty$ for any $q > 0$. Also, under \mathbf{P}^* the random variables $(\eta_i^*(X_{i,j}, X_{i,j-1}))_{1 \leq i \leq N, j \geq 1}$ are i.i.d. $\mathcal{N}(0, 1)$, so that condition $(\mathbf{C}_3^*(\mathbf{q}))$ holds for any $q > 0$, which implies condition $(\mathbf{C}_3(q))$ for any $q > 0$ (see Proposition 3). Thus, it follows from Theorems 3 and 4 that the sequential detection-identification procedure $\delta_{\beta}^{opt} = (T_{\beta}^{opt}, d_{\beta}^{opt})$ defined in (6.13)-(6.14) is asymptotically optimal (as $\beta_{max} \rightarrow 0$) and robust in the pointwise and minimax senses for any $r > 0$.

9.2. Monte Carlo

To get operating characteristics of the proposed detection-identification algorithm not only in the asymptotic case but also for reasonable probabilities of false alarm and misidentification, we perform Monte Carlo (MC) simulations for the modified Bernoulli model (9.4), (9.5) with $X_{i,n} = Y_{i,n}/V_i$. The values of $Y_{i,n}$ correspond to the number of susceptible at the n -th point in time for the i th population ($n \geq 0, i = 1, \dots, N$) and the values of V_i to the number of susceptible at the initial moment, i.e. $Y_{i,0} = V_i$. In simulations, we set the initial value $Y_{i,0} = V_i = 0.5(i+1) \cdot 10^4$ for $1 \leq i \leq N$. Without loss of generality we assume that the change occurs in the N th stream. Then (9.4) for $1 \leq i \leq N-1$ reduces to

$$X_{i,n} = (1-p_i^*)X_{i,n-1} + \sigma_i^* \sqrt{|X_{i,n-1}|} \xi_{i,n}, \quad \sigma_i^* = \sqrt{\frac{p_i^*(1-p_i^*)}{V_i}}, \quad X_{i,0} = 1,$$

and (9.5) to

$$X_{N,n} = (1-\vartheta)X_{N,n-1} + \sigma_{\vartheta} \sqrt{|X_{N,n-1}|} \xi_{N,n}, \quad \sigma_{\vartheta} = \sqrt{\frac{\vartheta(1-\vartheta)}{V_N}}, \quad X_{N,0} = 1,$$

where $\vartheta = \vartheta_n = p_N^* + (\bar{p} - p_N^*) \mathbf{1}_{\{n > \nu\}}$.

In each MC run m , using formulas (9.7)–(9.10) and (6.12), we get a pair $\delta_A^{*,m} = (T_A^{*,m}, d_A^{*,m})$ — the stopping time and the number of the stream where the change is detected ($m = 1, \dots, M$, M is the total number of MC runs).

The theoretic estimate of the expected detection delay for $i = N$ is given by the second asymptotic formula in (6.15) with $r = 1$, i.e.,

$$\mathcal{R}_{N,\nu,\vartheta_N} \approx \max_{1 \leq j \leq N} \frac{|\log \beta_{j,N}|}{\iota_{N,j}(\vartheta_N)} \quad \text{for } \vartheta_N = \bar{p}.$$

Since calculation of $\iota_{N,j}(\vartheta_N)$ analytically is difficult we evaluate it using MC simulations. To this end, we first estimate the conditional informations

$$J_i(\vartheta, x) = \int_{\mathbb{R}} g_i(\vartheta, y, x) f_{\vartheta}(y|x) dy \quad \text{and} \quad J_i^*(\vartheta, x) = \int_{\mathbb{R}} g_i^*(\vartheta, y, x) f_i^*(y|x) dy$$

and then we calculate the Kullback-Leibler divergences by MC as

$$\bar{I}_N(\vartheta) = \frac{1}{K} \sum_{n=\nu+1}^{\nu+K} J_N(\vartheta, X_{N,n}) \quad \text{and} \quad \bar{I}_i^*(\vartheta) = \frac{1}{K} \sum_{n=1}^K J_i^*(\vartheta, X_{i,n}) \quad \text{for } 1 \leq i \leq N-1. \quad (9.19)$$

By the law of large numbers for Markov chains the MC estimates $\bar{I}_i(\vartheta)$ and $\bar{I}_i^*(\vartheta)$ converge to the true values $\bar{J}_i(\vartheta)$ and $\bar{J}_i^*(\vartheta)$ defined in (7.5). Then (5.1) and (7.9) reduce to

$$\iota_{N,j}(\theta_N) = \bar{I}_N(\theta_N)\mathbf{1}_{\{j=N\}} + (\bar{I}_N(\theta_N) - \max_{\vartheta \in \Theta_j} \bar{I}_j^*(\vartheta))\mathbf{1}_{\{j \neq N\}}. \quad (9.20)$$

The MC estimate of the expected detection delay (in the N th stream) is calculated from the formula:

$$\widehat{\mathcal{R}} = \frac{\sum_{m=1}^M (T_A^{*,m} - \nu)\mathbf{1}_{\{T_A^{*,m} > \nu\}}\mathbf{1}_{\{d_A^{*,m} = N\}}}{\sum_{m=1}^M \mathbf{1}_{\{T_A^{*,m} > \nu\}}}.$$

In particular, for $\nu = 0$, which is used in simulations, it reduces to

$$\widehat{\mathcal{R}} = \frac{1}{M} \sum_{m=1}^M T_A^{*,m} \mathbf{1}_{\{d_A^{*,m} = N\}}.$$

The MC estimate of the false alarm probability ($\nu = +\infty$) is:

$$\widehat{\mathcal{P}}_N = \max_{1 \leq \ell \leq \mathbf{k}^* - \mathbf{m}^*} \frac{\sum_{m=1}^M \mathbf{1}_{\{\ell \leq T_A^{*,m} < \ell + \mathbf{m}^*\}} \mathbf{1}_{\{d_A^{*,m} = N\}}}{\sum_{m=1}^M \mathbf{1}_{\{T_A^{*,m} \geq \ell\}}},$$

and the MC estimates of the miss identification probabilities are:

$$\check{\mathcal{P}}_{j,N} = \max_{\nu < \ell \leq \nu + \mathbf{k}^*} \frac{\sum_{m=1}^M \mathbf{1}_{\{T_A^{*,m} > \ell\}} \mathbf{1}_{\{d_A^{*,m} = j\}}}{\sum_{m=1}^M \mathbf{1}_{\{T_A^{*,m} > \ell\}}} \quad \text{for } 1 \leq j \leq N - 1.$$

In simulations, we assume that the number of streams $N = 5$; the parameters of the observed process are $p_i^* = 1/(100 + i)$, $1/(50 + i)$; $q = \bar{p}/p_N^* = 1.1, 1.15, 1.2$; for calculation of thresholds $A_{i,j}$ we use (6.10) and (6.12) with $\beta_{i,j} = \frac{\varepsilon}{i+j}$, $\varepsilon = 0.3, 0.1, 0.01$ and $\check{\mathbf{k}} = 2, 1.55, 1.23$. We also assume that the change occurs from the very beginning, i.e., at the time $\nu = 0$, in which case $\vartheta = \bar{p}$.

The results are shown in Table 1 and Table 2. It is seen that the detection-identification algorithm has good performance. Even for small false alarm and miss identification probabilities the average detection delay is small. Therefore, we recommend using this algorithm in practice for the detection and localization of epidemics. Also, the asymptotic approximations for the average detection delay are quite accurate and, therefore, can be used for the evaluation of the performance of the detection-identification procedure in practice.

Table 1: Operating characteristics of the detection-identification procedure for $p_i^* = \frac{1}{100+i}$ (MC simulations with 10^5 runs).

ε	$\check{\mathbf{k}}$	q	$\check{\mathcal{P}}_{1,N}$	$\check{\mathcal{P}}_{2,N}$	$\check{\mathcal{P}}_{3,N}$	$\check{\mathcal{P}}_{4,N}$	$\widehat{\mathcal{P}}_N$	$\widehat{\mathcal{R}}$	$\mathcal{R}_{N,0,\theta_N}$
0.3	2	1.1	0.0024	0.0027	0.0011	0.0007	0.00088	6.46	5.17
0.3	2	1.15	0.0018	0.0041	0.0020	0.0007	0.00154	3.32	2.95
0.3	2	1.2	0.0036	0.0091	0.0044	0.0021	0.00459	2.02	2.03
0.1	1.55	1.1	0.0009	0.0014	0.0008	0.0003	0.00028	7.52	6.95
0.1	1.55	1.15	0.0004	0.0023	0.0013	0.001	0.0007	3.75	3.96
0.1	1.55	1.2	0.0014	0.0056	0.0023	0.0013	0.0023	2.26	2.72
0.01	1.23	1.1	0.00016	0.00062	0.00014	0.00011	$< 10^{-5}$	9.96	10.50
0.01	1.23	1.15	0.0001	0.0006	0.0004	0.0002	0.00014	4.78	5.99
0.01	1.23	1.2	0.0002	0.0019	0.0006	0.0005	0.0008	2.77	4.12

Table 2: Operating characteristics of the detection-identification procedure for $p_i^* = \frac{1}{50+i}$ (MC simulations with 10^5 runs).

ε	$\check{\mathbf{k}}$	q	$\check{P}_{1,N}$	$\check{P}_{2,N}$	$\check{P}_{3,N}$	$\check{P}_{4,N}$	\widehat{P}_N	$\widehat{\mathcal{R}}$	$\mathcal{R}_{N,0,\theta_N}$
0.3	2	1.1	0.00076	0.00062	0.00033	0.00016	0.0004	3.92	3.89
0.3	2	1.15	0.00078	0.00162	0.0004	0.0003	0.0007	2.03	2.29
0.3	2	1.2	0.0073	0.0078	0.0019	0.0009	0.0045	1.24	1.65
0.1	1.55	1.1	0.0002	0.00022	0.00004	0.00006	0.00014	4.50	5.23
0.1	1.55	1.15	0.00018	0.00082	0.00026	0.00016	0.00044	2.26	3.08
0.1	1.55	1.2	0.00342	0.00447	0.00142	0.00049	0.0017	1.33	2.21
0.01	1.23	1.1	0.00004	0.00004	0.00002	0.00002	0.00002	5.74	7.91
0.01	1.23	1.15	0.00002	0.00022	0.00008	0.00001	0.00008	2.78	4.65
0.01	1.23	1.2	0.00084	0.00145	0.00056	0.00019	0.0003	1.57	3.35

9.3. Detection of COVID-19 in Italy

In Subsection 9.1, we applied the proposed sequential detection-identification algorithm to epidemic models and showed it to be asymptotically optimal when the probabilities of wrong identification and false alarm are small. In this subsection, we demonstrate that the proposed detection-identification procedure can be effectively applied for the localization of COVID-19, i.e., for the detection of the epidemic anomalies and identification of the affected region. Consider the case of Italy.

Let $H_{i,n}$ be the number of hospitalized people at the n -th moment for the i th region. (Since the shortage of hospital beds presented a major challenge in Italy during the first wave of COVID-19, we focus on hospitalizations. However, this model also applies to other kinds of observations, e.g., number of infected people, number of visits to the doctor [1].) Then, $Y_{i,n} = V_i - H_{i,n}$, where V_i is the total number of hospital beds, i.e., $Y_{i,n}$ is potentially free beds for new hospitalizations at the n -th moment for the i th region. Then the observation, as in (9.2), will be $X_{i,n} = Y_{i,n}/V_i$.

We use the data provided by *Sito del Dipartimento della Protezione Civile - Emergenza Coronavirus: la risposta nazionale* (the Italian Department of Civil Protection). This data includes information on hospitalizations by region each day. We consider five Italian regions: Sicily, Lazio, Tuscany, Venice, and Lombardy. We use the proposed detection-identification algorithm to detect the presence of COVID-19 in a given region. Fig. 1 shows raw observations for five different regions, detection and identification of a region with a COVID outbreak in Italy by the proposed algorithm (blue vertical line), and the official introduction of a regional quarantine (red vertical line) in Lombardy.

It is known that Lombardy became the epicenter of the spread of COVID not only in Italy but throughout Europe. According to Fig. 1, the proposed algorithm detected COVID in Italy 9 days prior to the imposition of quarantine protocols in Lombardy (February 28, 2020 vs. March 8, 2020). The proposed detection-identification algorithm could therefore be a useful tool for researchers and public health authorities in detecting and localizing epidemics.

10. Conclusion

1. In this paper, we ignore the possible indifference zone Θ_{ind} of parameter values where the probabilities of false alarms and misidentification are too close to be reasonably distinguishable. In the indifference zone, the constraints on the erroneous decisions are not imposed, but still, the expected detection delays (or more generally moments of delay) have to be minimized for all possible parameter values, including those in the indifference zone. The modification of the proposed procedure to take into account an indifference zone, if needed, is straightforward. For the sake of brevity, the details are omitted.

2. As in the recent paper by Tartakovsky [16], we focus on the multistream changepoint model (2.1)–(2.2). It is worth noting that the same results hold in the single-stream detection-isolation problem when the observations $\{X_n\}_{n \geq 1}$ represent either a scalar process or a vector process but all components of this process change at time ν . Specifically, in change detection and isolation, the post-change hypothesis H_{ν,i,θ_i} , $\theta_i \in \Theta_i$ corresponding to the i th type of change usually involves unknown parameters θ_i and, therefore, is composite. Under the hypothesis H_{ν,i,θ_i} , $\theta_i \in \Theta_i$

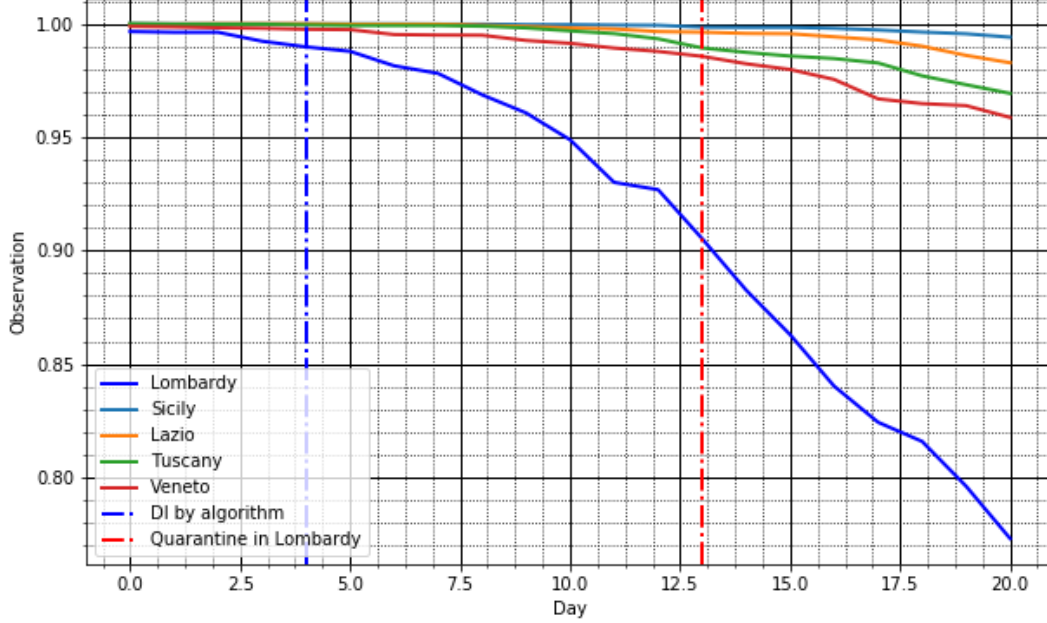


Fig. 1: Detection and identification of the region with an outbreak of the COVID-19 epidemic in Italy: the proposed algorithm vs. the imposition of quarantine protocols in Lombardy. The time of detection and identification by our algorithm is shown by blue vertical line and the time of the official imposition of quarantine protocols by red vertical line.

the post-change conditional density function is $f_{i,\theta_i,n}(X_n|\mathbf{X}_1^{n-1})$, $n > \nu$, while the pre-change density is $f_n^*(X_n|\mathbf{X}_1^{n-1})$, $n \leq \nu$, where $\mathbf{X}^l = (X_1, \dots, X_l)$. Hence, introducing parametric families of densities $\{f_{i,\theta_i,n}(X_n|\mathbf{X}_1^{n-1}), \theta_i \in \Theta_i\}$ and for $i = 1, \dots, N$ and $\Theta_i \subset \Theta$ considering the model¹

$$p(\mathbf{X}^n | H_{\nu,i,\theta_i}) = \begin{cases} \mathbf{q}^*(\mathbf{X}^n) & \text{for } \nu \geq n; \\ \mathbf{q}^*(\mathbf{X}^\nu) \prod_{l=\nu+1}^n f_{i,\theta_i,l}(X_l|\mathbf{X}^{l-1}) & \text{for } \nu < n \end{cases}, \quad (10.1)$$

where $p(\mathbf{X}^n | H_{\nu,i,\theta_i})$ stands for the joint density of the first n observations \mathbf{X}^n conditioned on the hypothesis H_{ν,i,θ_i} and $\mathbf{q}^*(\mathbf{X}^s) = \prod_{l=1}^s f_l^*(X_l|\mathbf{X}^{l-1})$ for $s \geq 1$, we arrive at the single-stream model that has all features of the previous multistream model (2.1)–(2.2). In fact, setting $X_n = (X_{1,n}, \dots, X_{N,n})$, where the components of this vector are mutually independent and assuming that the change may occur only in a single component, we obtain

$$p(\mathbf{X}^n | H_{\nu,i,\theta_i}) = \begin{cases} \prod_{i=1}^N \mathbf{q}_i^*(\mathbf{X}_i^n) & \text{for } \nu \geq n; \\ \mathbf{q}_i^*(\mathbf{X}_i^\nu) \prod_{l=\nu+1}^n f_{i,\theta_i,l}(X_{i,l}|\mathbf{X}_i^{l-1}) \prod_{1 \leq j \neq i \leq N} \prod_{l=1}^n f_j^*(X_{j,l}|\mathbf{X}_j^{l-1}) & \text{for } \nu < n \end{cases},$$

where $\mathbf{X}_i^n = (X_{1,i}, \dots, X_{i,n})$, $\mathbf{X}^n = (\mathbf{X}_1^n, \dots, \mathbf{X}_N^n)$, and $\mathbf{q}_i^*(\mathbf{X}_i^s) = \prod_{l=1}^s f_{i,l}^*(X_{i,l}|\mathbf{X}_i^{l-1})$. Obviously, this joint density is the same as the one in (2.4), so that in the case of mutually independent streams the multistream model defined in (2.1)–(2.2) is a particular case of the model (10.1).

3. All previous results can be generalized for the case when the change points are different for different streams, i.e., when $\nu = \nu_i$.

4. For independent observations as well as for a variety of Markov and certain hidden Markov models (see, e.g., Subsections 8.1 and (8.2) and Section 9), the decision statistics $\langle \mathbf{U}_n \rangle_{i,j}$ defined in (4.3) can be computed relatively

¹Often, $\theta_i \equiv \theta$ does not depend on i in practice.

easily, in which case implementation of the proposed change detection-identification procedure is not an issue. In general, however, the computational complexity of rule δ_A may be high. To avoid computational difficulties rule δ_A can be modified into a window-limited version where the summation in the decision statistics over potential change points ν is restricted to the sliding window of specific fixed size $\ell = \ell_\beta$, which is a function of the error probabilities constraints β . Following guidelines of [16] (Ch 3, Sec 3.10) where asymptotic optimality of mixture window-limited rules was established in the single-stream case, it can be shown that the window-limited version of the multihypothesis detection-identification procedure also has asymptotic optimality properties as long as the size of the window ℓ_β goes to infinity as $\beta_{max} \rightarrow 0$ at such a rate that $\ell_\beta/|\log \beta_{max}| \rightarrow \infty$.

Acknowledgements

The work of S.M. Pergamenchikov was partially supported by the RSF grant 20-61-47043 (National Research Tomsk State University). The work of A.G. Tartakovsky was supported in part by the Russian Science Foundation Grant 18-19-00452 at the Moscow Institute of Physics and Technology (Space Informatics Laboratory).

Appendix A. Moment properties of the epidemic models

Proposition A.1. *For any integer $m \geq 1$ and for any compact set $\Theta_i \subseteq \Theta$ the process $X_{i,n}$ defined in (9.4) and (9.5) has the following moment properties*

$$\sup_{x \in \mathbb{R}} \sup_{n \geq 0} \max_{1 \leq i \leq N} \frac{\mathbb{E}_{i,x}^* (X_{i,n}^*)^{2m}}{1 + x^{2m}} < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}} \sup_{n \geq 0} \max_{1 \leq i \leq N} \sup_{\vartheta \in \Theta_i} \frac{\mathbb{E}_{i,x}^\vartheta (X_{i,n})^{2m}}{1 + x^{2m}} < \infty. \quad (\text{A.1})$$

Proof. We prove only the second inequality in (A.1) since the proof of the first one is essentially similar. To this end, we first show that for any $x \in \mathbb{R}$

$$\sup_{n \geq 0} \max_{1 \leq i \leq N} \sup_{\vartheta \in \Theta_i} \mathbb{E}_{i,x}^\vartheta [X_{i,n}^2] \leq x^2 + 1. \quad (\text{A.2})$$

It is easily seen that for the model (9.5) we have

$$\mathbb{E}_{i,x}^\vartheta [X_{i,n}^2] \leq (1 - \vartheta)^2 \mathbb{E}_{i,x}^\vartheta [X_{i,n-1}^2] + \sigma_\vartheta^2 \mathbb{E}_{i,x}^\vartheta |X_{i,n-1}|.$$

For the sake of brevity write $y_n = \mathbb{E}_{i,x}^\vartheta [X_{i,n}^2]$. Taking into account that $x \leq 1 + x^2$, we obtain

$$y_n \leq (1 - \vartheta)^2 y_{n-1} + \sigma_\vartheta^2 \sqrt{y_{n-1}} \leq [(1 - \vartheta)^2 + \sigma_\vartheta^2] y_{n-1} + \sigma_\vartheta^2 = (1 - \vartheta) y_{n-1} + \sigma_\vartheta^2.$$

Let now $v_n = y_n - (1 - \vartheta)^2 y_{n-1}$. Clearly $v_n \leq \sigma_\vartheta^2$ and

$$y_n = x^2 (1 - \vartheta)^n + \sum_{j=1}^n \vartheta^{n-j} v_j \leq x^2 + \sigma_\vartheta^2 \sum_{j \geq 0} (1 - \vartheta)^j = x^2 + \frac{\sigma_\vartheta^2}{\vartheta} = x^2 + 1 - \vartheta.$$

This implies inequality (A.2) and, therefore, the second inequality in (A.1) for $m = 1$.

For an arbitrary $m \geq 1$ this inequality can be proved by induction as follows. Assume that the second inequality in (A.1) is true for $m - 1$ and $m \geq 2$, i.e., there exists a constant $C_m \geq 1$ such that for any $x \in \mathbb{R}$, $k \geq 1$, $1 \leq i \leq N$ and $\vartheta \in \Theta_i$

$$\mathbb{E}_{i,x}^\vartheta [(X_{i,k})^{2(m-1)}] \leq C_m (1 + x^{2(m-1)}). \quad (\text{A.3})$$

To show that it holds for m , using the initial condition $X_{i,0} = x$, we represent the process (9.5) as

$$X_{i,k} = (1 - \vartheta)^k x + \sigma_\vartheta \sum_{j=1}^k (1 - \vartheta)^{k-j} \sqrt{|X_{i,j-1}|} \xi_{i,j}.$$

By the Hölder inequality,

$$\begin{aligned} \mathbb{E}_{i,x}^\vartheta [X_{i,k}^{2m}] &\leq 2^{2m-1} \left(x^{2m} + \frac{\sigma_\vartheta^{2m}}{\vartheta^{2m-1}} \sum_{j=1}^k (1-\vartheta)^{k-j} \mathbb{E}_{i,x}^\vartheta [|X_{i,k}|^m \xi_{i,j}^{2m}] \right) \\ &\leq 2^{2m-1} \left(x^{2m} + \frac{(2m-1)!! \sigma_\vartheta^{2m}}{\vartheta^{2m}} \sup_{k \geq 0} \mathbb{E}_{i,x}^\vartheta |X_{i,k}|^m \right). \end{aligned}$$

Now, using the induction assumption (A.3) and that $|x+y|^\alpha \leq |x|^\alpha + |y|^\alpha$ for $0 < \alpha \leq 1$, we obtain

$$\mathbb{E}_{i,x}^\vartheta |X_{i,k}|^m \leq \left(\mathbb{E}_{i,x}^\vartheta |X_{i,k}|^{2(m-1)} \right)^{m/(2m-2)} \leq C_m^{m/(2m-2)} \left(1 + x^{2(m-1)} \right)^{m/(2m-2)} \leq C_m \left(1 + x^{2m} \right).$$

This implies the second inequality in (A.1), completing the proof. \square

References

- [1] BARON, M., CHOUDHARY K. AND YU, X. (2013) Change-Point Detection in Binomial Thinning Processes, with Applications in Epidemiology. *Sequential Analysis: Design Methods and Applications*, **32**, 350–367.
- [2] DAYANIK, S., POWELL, W. B. AND YAMAZAKI, K. (2013) Asymptotically optimal Bayesian sequential change detection and identification rules. *Annals of Operations Research*. **208** 337–370.
- [3] GALTHOUK, L.I. AND PERGAMENSHCHIKOV, S.M. (2013) Uniform concentration inequality for ergodic diffusion processes observed at discrete times. *Stochastic Processes Applic.* **123** 91–109.
- [4] GALTHOUK, L.I. AND PERGAMENSHCHIKOV, S.M. (2014) Geometric ergodicity for classes of homogeneous Markov chains. *Stochastic Processes Applic.* **124** 3362–3391.
- [5] GIRARDIN, V., KONEV, V.V. AND PERGAMENSHCHIKOV, S. M. (2018) Kullback-Leibler Approach to CUSUM Quickest Detection Rule for Markovian Time Series. *Sequential Analysis*. **37** 322–341.
- [6] KLÜPPELBERG, C. AND PERGAMENSHCHIKOV, S.M. (2004) The tail of the stationary distribution of a random coefficient AR(q) process with applications to an ARCH(q) process. *Ann. Appl. Probab.* **14** 971–1005.
- [7] LAI, T.L. (2000) Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Trans. Inform. Theory*. **46** 595–608.
- [8] MEYN, S. AND TWEEDIE, R. (1993) *Markov Chains and Stochastic Stability*. Springer Verlag, Berlin, New York.
- [9] NIKIFOROV, I.V. (1995) A generalized change detection problem. *IEEE Trans. Inform. Theory*. **41** 171–187.
- [10] NIKIFOROV, I.V. (2000) A simple recursive algorithm for diagnosis of abrupt changes in random signals. *IEEE Trans. Inform. Theory*. **46** 2740–2746.
- [11] NIKIFOROV, I.V. (2003) A lower bound for the detection/isolation delay in a class of sequential tests. *IEEE Trans. Inform. Theory*. **49** 3037–3046.
- [12] PERGAMENCHTCHIKOV, S.M. AND TARTAKOVSKY, A.G. (2018) Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data. *Stat Inference Stoch Process* **21** 217–259
- [13] PERGAMENCHTCHIKOV, S.M. AND TARTAKOVSKY, A.G. (2019) Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis. *Journal of Multivariate Analysis*, **174**.
- [14] TARTAKOVSKY, A.G. (2008) Multidecision quickest change-point detection: Previous achievements and open problems. *Sequential Analysis*. **27** 201–231
- [15] TARTAKOVSKY, A. (2020) *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*. Chapman & Hall/CRC Press (Monographs on Statistics and Applied Probability), Boca Raton, London, New York.
- [16] TARTAKOVSKY, A.G. (2021) An asymptotic theory of joint sequential changepoint detection and identification for general stochastic models. *IEEE Trans. Inform. Theory* **67** 4768–4783.
- [17] TARTAKOVSKY, A., NIKIFOROV, I., AND BASSEVILLE, M. (2014) *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman & Hall/CRC Press (Monographs on Statistics and Applied Probability), Boca Raton, London, New York.