ELSEVIER

Stochastic Processes and their Applications 83 (1999) 79–102

stochastic processes and their applications

www.elsevier.com/locate/spa

Detection of multiple changes in a sequence of dependent variables

Marc Lavielle[a,b,*]

[a] Université René Descartes, Paris, France

[b] Université Paris-Sud, Paris, France

Received 19 November 1997; accepted 26 February 1999

Abstract

We present some results of convergence for a minimum contrast estimator in a problem of change-points estimation. Here, we consider that the changes affect the marginal distribution of a sequence of random variables. We only consider parametric models, but the results are obtained under very general conditions. We show that the estimated configuration of changes converges to the true configuration, and we show that the rate of convergence does not depend on the dependance structure of the process: we obtain the same rate for strongly mixing and strongly dependent processes. When the number of changes is unknown, it is estimated by minimizing a penalized contrast function. Some examples of application to real data are given. © 1999 Published by Elsevier Science B.V. All rights reserved.

MSC: 62F12; 62M99; 62C99

Keywords: Detection of change-points; Minimum contrast estimator; Penalized minimum contrast estimator; Strongly mixing processes; Strongly dependent processes

1. Introduction

The change-point problem is important in many applications, and has been well-studied for more than fourty years (see, for example, the books of Brodsky and Darkhovsky (1993), or Basseville and Nikiforov (1993), and the many references therein for a state-of-the-art). According to the method of data acquisition, there exist two different formulations of this problem. The a posteriori (or off-line) change-points problem arises when the series of observations is complete at the time to process it. In the sequential change-points problem, the detection is performed on-line.

* Université Paris-Sud, Equipe de Probabilité, Statistique et Modélisation, Bât 425, 91405 Orsay Cedex, France. Tel.: +33-1-6915-5743; fax: +33-16915-7234.

E-mail address: marc.lavielle@math.u-psud-fr (M. Lavielle)

PII: S0304-4149(99)00023-X

We only consider in this paper the a posteriori problem. Then, our problem consists in recovering the configuration of change points using the whole observed series. The detection of a unique change point has been widely studied by different authors in different contexts. Among many others, Picard (1985) has proposed some test statistics for detecting a change in the spectrum of a process or in the mean of an autoregressive process, the order of which is known. Bai (1994) has extended these results for detecting a change in the mean of a linear process. Kim (1994) has compared the likelihood ratio and the cumulative sum test for detecting a change point in a linear regression. The papers of Giraitis and Leipus (1992) and Giraitis et al. (1996) deal with nonparametric situations for detecting a change in the marginal distribution function and the spectral function.

In the case of multiple changes, the problem is much more intricate when the number of changes is unknown, and few papers are dedicated to this problem. Various authors consider the particular case of changes in a sequence of independent and univariate random variables. In particular, Yao estimates the number of jumps in the mean of an independent normal sequence via Schwarz' criterion. Lombard (1987) and Mia and Zhao (1988) propose some procedures based on a rank statistics to test for one or more change points. Schechtman and Wolfe (1985) present a sequential algorithm for estimating the number and the location of the change points. Some authors also considered the problem of dependent data: Epps (1988) proposes a chi-squared statistics for testing the stationarity of a Gaussian process. He confines attention to processes which changes abruptly at some known instants. Vostrikova (1981) studies an iterative method of detection, when the changes affect the mean of a process. Finally, Brodsky and Darkhovsky (1993) also propose an algorithm for estimating shifts in a sequence of mixing variables.

In a previous paper, Lavielle and Moulines (1999) study a penalized least-square estimate of an unknown number of shifts in a time series. We extend their results to the problem of detecting changes in the marginal distribution function of a sequence of dependent – including strongly dependent variables. We consider that this distribution depends on a parameter θ that changes abruptly at some unknown instants. We show with an example, that the method also applies to nonparametric distributions that are discretized.

When the number of change points is known, the configuration of change points is estimated by minimizing a contrast function. We obtain some asymptotical results when the length of each segment tends to infinity, at the same rate as the total number of observations n. It is shown, under very mild hypothesis, that, if the minimum contrast estimate of θ, computed in any segment of the true configuration, is consistent, then, the minimum contrast estimator of the configuration of change points $\hat{\boldsymbol{\tau}}_n$ converges to the true configuration $\boldsymbol{\tau}^\star$. The estimated parameter of vectors $\hat{\boldsymbol{\theta}}_n$ also converges to the true vector of parameters $\boldsymbol{\theta}^\star$. Furthermore, we precise the rate of convergence of the estimator $(\hat{\boldsymbol{\tau}}_n, \hat{\boldsymbol{\theta}}_n)$. In particular, we show that $\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^\star\|_\infty = \mathcal{O}_{\mathrm{P}}(n^{-1})$. An interesting result is that this rate of convergence does not depend on the covariance structure of the process: this rate holds for strongly mixing sequences and for strongly dependent sequences. On the other hand, the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ depends on the covariance structure of the process. Indeed, this rate is the same as in the absence of changes.

When the number of changes is unknown, it is estimated by minimizing a penalized contrast function. The penalization term has the form $\beta_n K$ where K is the number of segments, that is the number of parameters in the model. Then, this problem of change points detection can be seen as a problem of model selection via penalization (see Schwarz, 1978). This kind of method has been developed for estimating the order of an ARMA process (see Akaike, 1974; Hannan, 1980) or for estimating the order of a mixture (see Dacunha-Castelle and Gassiat (1997)). In a context of regression and density estimation, some precise risk bounds have been obtained by Barron et al. (1999), using theory of sieves. We show that the estimated number of change points converges to the true number if β_n goes to 0 at an appropriate rate. Indeed, if β_n decreases too quickly, the number of segments will be over-estimated. On the other hand, this number will be sub-estimated if β_n is too big. We must notice that Yao (1988) already proposed this kind of criterion with $\beta_n = \log n/n$ in the particular case of independent data. In the case of dependent data, the choice of β_n directly depends on the rate of convergence of $\hat{\theta}_n$, that is on the dependence structure of the data.

Some examples of application are finally proposed. First, we consider the problem of detecting changes in the mean and/or the variance of a process. We also consider the case of changes in a discrete distribution. We apply these methods to real data, for detecting change points in the CAC 40 index and the heart rate of a new-born baby. At this stage, it is important to underline that the results obtained in this paper are asymptotical. Of course, in front of a real data sequence, the penalization coefficient β_n takes a fixed value. The choice of this value is not justified by theoretical considerations, but by practical considerations: we choose β_n in order to obtain a resolution level, that is, a number of changes, that seems satisfactory (see Lavielle, 1997) for practical prescriptions for the choice of this tuning parameter). An automatic choice of β_n, for any value of n would require nonasymptotical results, but this is beyond the scope of this paper.

2. Detection of a known number of changes

2.1. Model and notations

Let Θ be a compact subset of $\mathbb{R}^d$ and $\vartheta^\star$ a function from $]0,1]$ to Θ. We consider a sequence of random variables $Y_1,\dots,Y_n$ that take its values in $\mathbb{R}^p$, and such that, for any $1\leqslant i\leqslant n$, the distribution of Y_i depends on the parameter $\vartheta^\star(i/n)$: there exists a function $F_i : \mathbb{R}^p \times \Theta \to [0,1]$ such that, for any $A \subset \mathbb{R}^p$ and any $1\leqslant i\leqslant n$,

$$P(Y_i \in A) = F_i\left(A; \vartheta^\star\left(\frac{i}{n}\right)\right).$$

We do not assume here that the F_i's are all identical. Indeed, consider the following model as an example, where the mean of the process varies:

$$Y_i = \vartheta^\star\left(\frac{i}{n}\right) + \varepsilon_i, \quad 1\leqslant i\leqslant n \tag{1}$$

where $(\varepsilon_i, 1 \leqslant i \leqslant n)$ is a sequence of centered random variables. Then, we can consider the problem of estimating the mean $\vartheta^\star$, without assuming that (ε_i) is identically distributed. This problem attracts considerable attention in different contexts. For example, Dahlhaus (1997) assumes that the function $\vartheta^\star$ is smooth, in order to guarantee that the process has locally a stationary behavior.

We consider here that this function is piecewise constant. Then, there exist $\tau_0^\star = 0 < \tau_1^\star < \cdots < \tau_{K-1}^\star < \tau_K^\star = 1$ and $\theta_1^\star, \ldots, \theta_K^\star \in \Theta \times \cdots \times \Theta$ such that, for all $x \in]0,1]$,

$$\vartheta^\star(x) = \sum_{j=1}^{K} \theta_j^\star \mathbb{1}_{\tau_{j-1}^\star < x \leqslant \tau_j^\star}. \tag{2}$$

This model means that $K-1$ changes affect the distribution of (Y_i) at some unknown instants $(t_j^\star, 1 \leqslant j \leqslant K-1)$ with $t_j^\star = [n\tau_j^\star]$: for any $A \subset \mathbb{R}^p$ and any $1 \leqslant i \leqslant n$,

$$P(Y_i \in A) = \sum_{j=1}^{K} F_i(A; \theta_j^\star) \mathbb{1}_{t_{j-1}^\star < i \leqslant t_j^\star}. \tag{3}$$

In this context, estimating the function $\vartheta^\star$ reduces to estimating the vector of parameters $\boldsymbol{\theta}^\star = (\theta_1^\star, \ldots, \theta_K^\star)$ and the configuration of normalized change points $\boldsymbol{\tau}^\star = (\tau_1^\star, \ldots, \tau_{K-1}^\star)$ from n observations $Y_1, Y_2, \ldots, Y_n$.

We use index j for the true configuration of change points $\boldsymbol{\tau}^\star = (\tau_j^\star, 1 \leqslant j \leqslant K-1)$, and for the true vector of parameters $\boldsymbol{\theta}^\star = (\theta_j^\star, 1 \leqslant j \leqslant K)$. We use index k for any other configuration. We denote by $\boldsymbol{Y}_j^\star$ the vector of observations that belong to segment j in the configuration $\boldsymbol{\tau}^\star$, and by $\boldsymbol{Y}_k$ the vector of observations that belong to segment k in the configuration $\boldsymbol{\tau} = (\tau_k, 1 \leqslant k \leqslant K-1)$:

$$\boldsymbol{Y}_j^\star = (Y_{t_{j-1}^\star+1}, \ldots, Y_{t_j^\star}),$$

$$\boldsymbol{Y}_k = (Y_{t_{k-1}+1}, \ldots, Y_{t_k}).$$

The lengths of $\boldsymbol{Y}_j^\star$ and $\boldsymbol{Y}_k$ are, respectively, $n_j^\star$ and n_k. Then, $n_j^\star/n = \tau_j^\star - \tau_{j-1}^\star$ is the proportion of observations that belong to segment j in $\boldsymbol{\tau}^\star$. The vector of observations that belong to segment j in the configuration $\boldsymbol{\tau}^\star$ and to segment k in the configuration $\boldsymbol{\tau}$ is

$$\boldsymbol{Y}_{kj} = (Y_i,\ i \in [t_{k-1}+1, t_k] \cap [t_{j-1}^\star + 1, t_j^\star]),$$

and the length of $\boldsymbol{Y}_{kj}$ is n_{kj}.

The aim of this paper is to study the behavior of an estimator of $(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star)$ when $n \to \infty$, when $(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star)$ is fixed. Then, the normalized lengths $(n_j^\star/n)$ of the segments in $\boldsymbol{\tau}^\star$, as well as the jumps $(\|\theta_j^\star - \theta_{j-1}^\star\|)$ are fixed, and bounded from below:

- There exists $0 < \Delta_\tau^\star < 1$ such that $\tau_j^\star - \tau_{j-1}^\star \geqslant \Delta_\tau^\star$, $1 \leqslant j \leqslant K$.
- There exists $\Delta_\theta^\star > 0$ such that $\|\theta_j^\star - \theta_{j-1}^\star\|_2 \geqslant \Delta_\theta^\star$, $2 \leqslant j \leqslant K$.

(Remark at this stage that we do not assume here that some minorants of $\Delta_\tau^\star$ and $\Delta_\theta^\star$ are known.)

2.2. Some preliminary hypothesis

Before estimating $(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star)$, it is clear that we must be able to estimate the parameter $\theta_j^\star$ in segment j of $\boldsymbol{\tau}^\star$, for any $1 \leqslant j \leqslant K$. Indeed, we shall assume that there exists a contrast function W_n such that the *minimum contrast estimator* $\hat{\theta}_j$ of $\theta_j^\star$, computed in segment j of $\boldsymbol{\tau}^\star$, and defined as the solution of the following minimisation problem:

$$W_n(\mathbf{Y}_j^\star, \hat{\theta}_j) \leqslant W_n(\mathbf{Y}_j^\star, \theta), \quad \forall \theta \in \Theta \tag{4}$$

converges in P-probability to $\theta_j^\star$ when $n \to \infty$. More precisely, we make the following hypothesis:

H1 (On the contrast function)

(i) There exist $\phi : \Theta \to \mathbb{R}$, $\psi : \Theta \to \mathbb{R}^m$ two twice continuously derivable functions, and $\xi : \mathbb{R}^p \to \mathbb{R}^m$ such that, for any $\theta \in \Theta$, the contrast function W_n can be written as:

$$W_n(Y_t \ldots Y_{t'}, \theta) = \frac{1}{n} \sum_{i=t}^{t'} (\phi(\theta) + \langle \psi(\theta), \xi(Y_i) \rangle), \quad 1 \leqslant t \leqslant t' \leqslant n. \tag{5}$$

(ii) There exists a function w of $\Theta \times \Theta$ in $\mathbb{R}$ such that, for any $1 \leqslant j \leqslant K$ and any $\theta \in \Theta$,

$$w(\theta_j^\star, \theta) = \phi(\theta) + \langle \psi(\theta), E\xi(Y_i) \rangle, \quad t_{j-1}^\star + 1 \leqslant i \leqslant t_j^\star. \tag{6}$$

and such that, for any $(\theta, \theta') \in \Theta \times \Theta$, $w(\theta, \theta) \leqslant w(\theta, \theta')$, with $w(\theta, \theta) = w(\theta, \theta')$ if and only if $\theta = \theta'$.

Furthermore, for any $1 \leqslant j \leqslant K$, there exist a neighborhood $\mathscr{V}(\theta_j^\star) \subset \Theta$ of $\theta_j^\star$ and a constant $B > 0$ such that

$$w(\theta_j^\star, \theta) - w(\theta_j^\star, \theta_j^\star) \geqslant B \|\theta_j^\star - \theta\|_2^2$$

for any $\theta \in \mathscr{V}(\theta_j^\star)$.

For any $(\theta, \theta') \in \Theta \times \Theta$, let

$$v(\theta, \theta') = w(\theta, \theta') - w(\theta, \theta). \tag{7}$$

Then, under H1, $v(\theta, \theta') \geqslant 0$ for any $(\theta, \theta') \in \Theta \times \Theta$ and $v(\theta, \theta') = 0$ if and only if $\theta = \theta'$.

Remark. We only consider here that the changes affect the marginal distribution of the Y_i's, then, we assume in H1(i) that the contrast function can be written as a sum (5). Let

$$g(y; \theta) = -\phi(\theta) - \langle \psi(\theta), \xi(y) \rangle. \tag{8}$$

In many applications, g is the log-likelihood of a probability distribution G_θ. As an example, we shall see in Section 4 that a Gaussian likelihood can be used as a contrast function for detecting changes in the mean and/or in the variance of a sequence of random (not necessarily Gaussian) variables. In such a case, $v(\theta, \theta')$ is the Kullback–Lieber distance beetween G_θ and $G_{\theta'}$, and H1(ii) is satisfied.

On the other hand, H1 is generally not satisfied when the changes affect the joint distribution of (Y_i). The contrast functions that can be used for detecting changes in

the spectrum of a random process, for example, are built up from the periodogram (see Lavielle, 1998), that is from some quadratic forms, and H1 does not hold anymore.

In order to control the fluctuations of the contrast process, we consider also the following condition:

H2 (on the fluctuations of the contrast process)

For any $\theta \in \Theta$, let

$$\eta_i(\theta) = \langle \psi(\theta), \xi(Y_i) - E\xi(Y_i) \rangle, \quad 1 \leqslant i \leqslant n. \tag{9}$$

Then, there exists $1 \leqslant h < 2$, such that,

$$E\left(\sum_{i=t}^{t+s} \eta_i(\theta)\right)^2 \leqslant C(\theta)s^h, \quad 1 \leqslant t \leqslant t+s \leqslant n. \tag{10}$$

Now, for any $\theta \in \Theta$, let

$$S_n(Y_t, \ldots, Y_{t+s}, \theta) = \sum_{i=t}^{t+s} \eta_i(\theta), \quad 1 \leqslant t \leqslant t+s \leqslant n. \tag{11}$$

Then, under H1, by using (5), (6), (9) and (11), $W_n(\boldsymbol{Y}_j^\star, \theta)$ can be written as follows:

$$W_n(\boldsymbol{Y}_j^\star, \theta) = (\tau_j^\star - \tau_{j-1}^\star) w(\theta_j^\star, \theta) + \frac{1}{n} S_n(\boldsymbol{Y}_j^\star, \theta). \tag{12}$$

Under H1 and H2, $n^{-1} S_n(\boldsymbol{Y}_j^\star, \theta)$ converges in probability to 0 at rate $n^{1-h/2}$, uniformly in θ. Then, $W_n(\boldsymbol{Y}_j^\star, \theta)$ converges in probability to $(\tau_j^\star - \tau_{j-1}^\star) w(\theta_j^\star, \theta)$, and $\hat{\theta}_j$ defined in (4) converges in probability to $\theta_j^\star$ at rate $n^{1-h/2}$.

2.3. Some remarks on condition H2

As we shall see in the next sections, the consistency of the estimator and the rate of convergence are obtained under H1 and H2, that is when the sequence $\eta_i(\theta)$ defined in Eq. (9) satisfies condition (10) for some $1 \leqslant h < 2$. In order to simplify the notations in this section, we shall omit θ and use the notation η_i instead of $\eta_i(\theta)$, since the results hold for any $\theta \in \Theta$. It is useful to remark that inequality (10) is satisfied for a wide family of processes $\eta = (\eta_i)$. In fact, if η is a second-order stationary process with zero-mean and autocovariance function γ_η, we have

$$E\left(\sum_{i=1}^{n} \eta_i\right)^2 = \sum_{k=-n+1}^{n-1} (n - |k|)\gamma_\eta(k) \leqslant 2n \sum_{k=0}^{n-1} |\gamma_\eta(k)|. \tag{13}$$

Thus, if $\gamma_\eta(k) = \mathcal{O}(n^{-a})$ for some $a > 0$, then η satisfies condition (10) for $h = \max(2-a, 1)$. For practical applications, it is necessary to obtain some conditions on the process Y, instead of η. We shall see some situations where the relation between the two processes is explicit.

(1) Y is a sequence of independent random variables. This is the simplest situation, and η is also a sequence of independent random variables such that (10) is satisfied for $h = 1$.

(2) Y is a second-order stationary ϕ-*mixing* process, see Doukhan (1994) for the definition. In this case, let $(\phi_Y(k))$ be the sequence of mixing coefficients of Y, such that

$$\phi_Y(k) = \mathcal{O}(k^{-a})$$

where $a > 0$. Hence, η is also a ϕ-*mixing* process such that

$$\phi_\eta(k) \leqslant \phi_Y(k), \quad \forall k \geqslant 0.$$

If $E\eta_0^2 \leqslant \infty$, and using the fact that, for any $k \geqslant 0$,

$$E(\eta_0\eta_k) \leqslant 2E(\eta_0^2)(\phi_\eta(k))^{1/2},$$

η satisfies (10) for $h = \min(2 - a/2, 1)$.

(3) Y is a second-order stationary α-*mixing* process, see Doukhan (1994). As before, let $(\alpha_Y(k))$ be the sequence of mixing coefficients of Y, such that

$$\alpha_Y(k) = \mathcal{O}(k^{-a}), \quad \forall k \geqslant 0,$$

where $a > 0$.

If there exists $\delta > 0$ and $A > 0$ such that $E|\eta_i|^{2(1+\delta)} \leqslant A$, for any $i \geqslant 0$, then η satisfies (10) for $h = \min(2 - a\delta/(1+\delta), 1)$.

We get this result by using the following inequality due to Davydov (1970):

$$E(\eta_0\eta_k) \leqslant 8(\alpha_\eta(k))^{1/p}\sqrt{E(|\eta_0|^q)E(|\eta_k|^q)}, \quad \frac{2}{q} + \frac{1}{p} = 1,$$

with $q = 2 + 2\delta$ and $p = 1 + 1/\delta$. In particular, if η is bounded, then $E(\eta_0\eta_k) \leqslant C\alpha_\eta(k)$.

(Here, we can remark that more precise inequalities have been obtained more recently in this context by Rio (1993).)

(4) Y is a strongly dependent stationary Gaussian process, see Taqqu (1977). Then, there exists $0 \leqslant a < 1$ such that

$$\mathrm{Cov}(Y_0, Y_k) = \mathcal{O}(k^{-a}), \quad \forall k \geqslant 0.$$

Because of the definition of η_i in (9), we can see η_i as a function of Y_i and consider the development of η_i on the basis of Hermite polynomials when $E\eta_0^2 < \infty$:

$$\eta_i = g(Y_i) - Eg(Y_i) = \sum_{j \geqslant m} c_j H_j(Y_i). \tag{14}$$

Here, m is the Hermite rank of η, that is the smallest value of j such that the coefficient c_j of H_j in (14) is different of zero. We shall see in the last section that $m = 1$ in the case of changes in the mean, and $m = 2$ in the case of changes in the variance of random variables.

In this context, the following inequality holds (see Taqqu (1977), for example): there exists a constant D such that

$$E(\eta_0\eta_k) \leqslant D(\mathrm{Cov}(Y_0, Y_k))^m, \tag{15}$$

and η satisfies (10) for $h = \max(2 - am, 1)$.

Some results can also be obtained for non-Gaussian processes, using Appell polynomials, instead of Hermite polynomials, see Giraitis and Surgailis (1986), for more details.

2.4. A useful maximal inequality

The following lemma will be very useful in the following.

Lemma 2.1. *Let (η_i) be a sequence of random variables with zero mean. We assume that there exists $C < \infty$ and $1 < h < 2$ such that, for any $t \geqslant 0$ and any $s > 0$,*

$$E\left(\sum_{i=t+1}^{t+s} \eta_i\right)^2 \leqslant C s^h. \tag{16}$$

Then, there exists $A(h) > 0$ and $B(h) > 0$ such that, for any $\delta > 0$, for any $n > 0$ and any $m > 0$, the following maximal inequalities are satisfied:

$$P\left(\max_{1 \leqslant t \leqslant n} \left|\sum_{i=1}^{t} \eta_i\right| > \delta\right) \leqslant A(h)\frac{n^h}{\delta^2}, \tag{17}$$

$$P\left(\max_{t \geqslant m} \frac{1}{t}\left|\sum_{i=1}^{t} \eta_i\right| > \delta\right) \leqslant B(h)\frac{m^{h-2}}{\delta^2}. \tag{18}$$

Proof. The first inequality has been shown by Móricz et al. (1982). The same kind of proof by induction easily leads to the following generalisation of the Hájek and Rényi (1955) inequality: under the conditions of Lemma 2.1 and for any decreasing sequence $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_n > 0$,

$$P\left(\max_{1 \leqslant t \leqslant n} b_t \left|\sum_{i=1}^{t} \eta_i\right| > \delta\right) \leqslant A(h)\frac{n^{h-1}}{\delta^2}\sum_{i=1}^{n} b_i^2. \tag{19}$$

Then, we can show (18) by setting $b_t = 1/t$ and by splitting the set $\{t \geqslant m\}$ into the union of subsets of the form $\{2^p m \leqslant t < 2^{p+1} m\}$ for $p = 1, 2, \ldots$. We have

$$\begin{aligned} P\left(\max_{t \geqslant m} \frac{1}{t}\left|\sum_{i=1}^{t} \eta_i\right| > \delta\right) &\leqslant \sum_{p=0}^{\infty} P\left(\max_{2^p m \leqslant t < 2^{p+1} m} \frac{1}{t}\left|\sum_{i=1}^{t} \eta_i\right| > \delta\right) \\ &\leqslant \sum_{p=0}^{\infty} \frac{A(h)}{\delta^2}(2^p m)^{h-1} \sum_{i=2^p m}^{2^{p+1} m - 1} \frac{1}{i^2} \\ &\leqslant \frac{A(h)}{\delta^2}\sum_{p=0}^{\infty}(2^p m)^{h-2} \\ &\leqslant \frac{A(h)}{\delta^2}\frac{m^{h-2}}{1 - 2^{h-2}}. \qquad \square \end{aligned} \tag{20}$$

Remarks on Lemma 2.1. Inequalities (17) and (18) hold when (16) is satisfied for any $1 < h < 2$. Nevertheless, this result can be extended to $h = 1$, under some hypothesis on η. For example, these inequalities still hold with $h = 1$ if η is a sequence of independent variables, or if η is a linear process, the spectral density of which satisfies some conditions of regularity, see Bai (1994). The same result also holds for strongly

mixing processes when the sequence of strongly mixing coefficients goes to 0 quickly enough, see Oodaira and Yoshihara (1972). Under H1, let

$$\nabla S_n(Y_t,\dots,Y_{t'},\tilde{\theta}) = \sum_{i=t}^{t'} \nabla\eta_i(\theta) \tag{21}$$

where, using (9),

$$\nabla\eta_i(\theta) = \frac{\partial}{\partial\theta}\eta_i(\theta) = [\nabla\psi(\theta)](\xi(Y_i) - E\xi(Y_i)). \tag{22}$$

Then, under H2, using (11) and (21), a direct application of Lemma 2.1 leads to the following result: there exists $A_1 > 0$, $A_2 > 0$ and $A_3 > 0$, such that, for any $m > 0$,

$$P\left(\max_{1\leqslant t\leqslant t'\leqslant m}\sup_{\theta\in\Theta}|S_n(Y_{t+1}\dots Y_{t'},\theta)| > \delta\right) \leqslant \frac{A_1 m^h}{\delta^2}, \tag{23}$$

$$P\left(\max_{t\geqslant m}\sup_{\theta\in\Theta}\frac{|S_n(Y_1\dots Y_t,\theta)|}{t} > \delta\right) \leqslant \frac{A_2 m^{h-2}}{\delta^2}, \tag{24}$$

$$P\left(\max_{1\leqslant t\leqslant t'\leqslant m}\sup_{\theta\in\Theta}||\nabla S_n(Y_{t+1}\dots Y_{t'},\theta)||_2 > \delta\right) \leqslant \frac{A_3 m^h}{\delta^2}. \tag{25}$$

2.5. Definition of the estimator

Let $\mathscr{T}_K$ be the set of configurations and Θ_K the space of the parameters,

$$\Theta_K = \{\boldsymbol{\theta} = (\theta_1,\theta_2,\dots,\theta_K),\ \theta_k\in\Theta\},$$
$$\mathscr{T}_K = \{\boldsymbol{\tau} = (\tau_0,\tau_1,\dots,\tau_K),\ \tau_0 = 0 < \tau_1 < \cdots < \tau_{K-1} < \tau_K = 1\}.$$

We estimate $(\boldsymbol{\tau}^\star,\boldsymbol{\theta}^\star)$ by minimizing the function $J_n(\boldsymbol{\tau},\boldsymbol{\theta})$ in $\mathscr{T}_K\times\Theta_K$ defined by

$$J_n(\boldsymbol{\tau},\boldsymbol{\theta}) = \sum_{k=1}^{K} W_n(\boldsymbol{Y}_k,\theta_k), \tag{26}$$

where $W_n(\boldsymbol{Y}_k,\theta_k)$ is the contrast function computed over segment k of $\boldsymbol{\tau}$.

For a given configuration $\boldsymbol{\tau}$, $\hat{\theta}_k$ minimizes $W_n(\boldsymbol{Y}_k,\theta_k)$. Then, we can remark that, under H1 and H2 and when $\boldsymbol{\tau}=\boldsymbol{\tau}^\star$, the estimate of $\boldsymbol{\theta}^\star$ obtained by minimizing $J_n(\boldsymbol{\tau}^\star,\boldsymbol{\theta})$ converges to $\boldsymbol{\theta}^\star$.

For technical reasons, we will use the fact that, since $J_n(\boldsymbol{\tau}^\star,\boldsymbol{\theta}^\star)$ is constant, $(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)$ minimizes $U_n(\boldsymbol{\tau},\boldsymbol{\theta})$ defined by

$$U_n(\boldsymbol{\tau},\boldsymbol{\theta}) = J_n(\boldsymbol{\tau},\boldsymbol{\theta}) - J_n(\boldsymbol{\tau}^\star,\boldsymbol{\theta}^\star) = u(\boldsymbol{\tau},\boldsymbol{\theta}) + e_n(\boldsymbol{\tau},\boldsymbol{\theta}) \tag{27}$$

where, using (7) and (11),

$$u(\boldsymbol{\tau},\boldsymbol{\theta}) = EU_n(\boldsymbol{\tau},\boldsymbol{\theta}) = \sum_{k=1}^{K}\sum_{j=1}^{K}\frac{n_{kj}}{n}v(\theta_j^\star,\theta_k), \tag{28}$$

$$e_n(\boldsymbol{\tau},\boldsymbol{\theta}) = U_n(\boldsymbol{\tau},\boldsymbol{\theta}) - u(\boldsymbol{\tau},\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{j=1}^{K}(S_n(\boldsymbol{Y}_{kj},\theta_k) - S_n(\boldsymbol{Y}_{kj},\theta_j^\star)). \tag{29}$$

We shall establish now the consistency of this estimate, and compute its rate of convergence, under very general conditions.

2.6. Consistency of the estimator

We have the following result:

Theorem 2.2. *Let $\hat{\boldsymbol{\tau}}_n$ be the estimate of the change-points sequence and $\hat{\boldsymbol{\theta}}_n$ be the estimate of the parameters in the different segments, obtained as the solution of the following minimization problem*:

$$J_n(\hat{\boldsymbol{\tau}}_n, \hat{\boldsymbol{\theta}}_n) \leqslant J_n(\boldsymbol{\tau}, \boldsymbol{\theta}), \quad \forall (\boldsymbol{\tau}, \boldsymbol{\theta}) \in \mathscr{T}_K \times \Theta_K. \tag{30}$$

Then, under H1 *and* H2, $(\hat{\boldsymbol{\tau}}_n, \hat{\boldsymbol{\theta}}_n)$ *converges in P-probability to* $(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star)$.

Proof. First, we shall show the following lemma.

Lemma 2.3. *If condition* H1 *is satisfied, then there exists two constants* $C_{\theta^\star} > 0$ *and* $C_{\tau^\star} > 0$ *such that, for any* $(\boldsymbol{\tau}, \boldsymbol{\theta}) \in \mathscr{T}_K \times \Theta_K$,

$$u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \max(C_{\theta^\star} ||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty, C_{\tau^\star} ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V) \tag{31}$$

where $||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty = \max_j |\tau_j - \tau_j^\star|$ *and* $||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V = \max_j v(\theta_j^\star, \theta_j)$.

Proof. For any $1 \leqslant k \leqslant K-1$, let

$$f_k(\theta^\star, \alpha) = \inf_{\theta \in \Theta} (\alpha v(\theta_{k+1}^\star, \theta) + (1-\alpha) v(\theta_k^\star, \theta)).$$

Then, we have $f_k(\theta^\star, 0) = f_k(\theta^\star, 1) = 0$. Furthermore, $f_k(\theta^\star, \cdot)$ is a concave function (as the inferior hull of a family of linear functions). Then, let

$$A_k(\theta^\star) = 2 f_k(\theta^\star, \tfrac{1}{2}).$$

Thus,

$$f_k(\theta^\star, \alpha) \geqslant \alpha A_k(\theta^\star), \quad \forall\ 0 \leqslant \alpha \leqslant \tfrac{1}{2}. \tag{32}$$

We can see that $A_k(\theta^\star) > 0$ if $\theta_k^\star \neq \theta_{k+1}^\star$, and we set

$$\underline{A}(\theta^\star) = \min_{1 \leqslant k \leqslant K-1} A_k(\theta^\star)$$

with $\underline{A}(\theta^\star) > 0$. Let us consider now a configuration $\boldsymbol{\tau} \in \mathscr{T}_K$ such that $||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty \leqslant \Delta_\tau^\star / 4$. A change point τ_k can be on the left, or on the right of the original change-point $\tau_k^\star$:

- For any k such that $\tau_{k-1} \leqslant \tau_k^\star \leqslant \tau_k$ and using (28), we have that

$$u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \frac{n_{k,k+1}}{n} v(\theta_{k+1}^\star, \theta_k) + \frac{n_{kk}}{n} v(\theta_k^\star, \theta_k). \tag{33}$$

Let

$$\alpha_{k,k+1} = \frac{n_{k,k+1}}{n_{k,k+1} + n_{kk}}.$$

By hypothesis, $\alpha_{k,k+1} \leqslant 1/2$ and

$$\begin{aligned} u(\boldsymbol{\tau}, \boldsymbol{\theta}) &\geqslant \frac{n_{k,k+1} + n_{kk}}{n} (\alpha_{k,k+1} v(\theta_{k+1}^\star, \theta_k) + (1 - \alpha_{k,k+1}) v(\theta_k^\star, \theta_k)) \\ &\geqslant (\tau_k - \tau_k^\star) \underline{A}(\theta^\star). \end{aligned} \tag{34}$$

- For any k such that $\tau_k \leqslant \tau_k^\star \leqslant \tau_{k+1}$, we set in a same way

$$\alpha_{k,k-1} = \frac{n_{k,k-1}}{n_{k,k-1} + n_{kk}}.$$

As before, and using the fact that $\alpha_{k,k-1} \leqslant 1/2$,

$$u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant (\tau_k^\star - \tau_k)\underline{A}(\theta^\star). \tag{35}$$

Thus, if $||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty \leqslant \Delta_\tau^\star/4$, $u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \underline{A}(\theta^\star)||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty$.

On the other hand, from (28), $u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \max_k v(\theta_k^\star, \theta_k) n_{kk}/n$. Now, if $||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty \leqslant \Delta_\tau^\star/4$, $n_{kk}/n \geqslant \Delta_\tau^\star/2$ for any k, and then, $u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \Delta_\tau^\star ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V/2$.

If now $||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty > \Delta_\tau^\star/4$, there clearly exists a pair (k, j) such that $n_{kj} \geqslant n\Delta_\tau^\star/4$ and $n_{k,j+1} \geqslant n\Delta_\tau^\star/4$. Let

$$\alpha_{k,j+1} = \frac{n_{k,j+1}}{n_{k,j+1} + n_{kj}}.$$

For any $\boldsymbol{\theta} \in \Theta_K$, we have

$$\begin{aligned}
u(\boldsymbol{\tau}, \boldsymbol{\theta}) &\geqslant \frac{n_{k,j+1} + n_{kj}}{n} (\alpha_{k,j+1} v(\theta_{j+1}^\star, \theta_k) + (1 - \alpha_{k,j+1}) v(\theta_j^\star, \theta_k)) \\
&\geqslant \frac{n_{k,j+1} + n_{kj}}{n} \min(\alpha_{k,j+1}, 1 - \alpha_{k,j+1}) \underline{A}(\theta^\star) \\
&\geqslant \frac{n_{k,j+1} + n_{k,j+1}}{n} \min\left(\frac{n_{k,j+1}}{n}, \frac{n_{k,j+1}}{n}\right) \underline{A}(\theta^\star) \\
&\geqslant \frac{\Delta_\tau^{\star 2}}{8} \underline{A}(\theta^\star).
\end{aligned} \tag{36}$$

Finally, using the fact that for $0 \leqslant a, b \leqslant 1$, $\min(a, b) \geqslant ab$, we obtain that

$$\begin{aligned}
u(\boldsymbol{\tau}, \boldsymbol{\theta}) &\geqslant \underline{A}(\theta^\star) \min\left(\frac{\Delta_\tau^\star}{2}, ||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty\right) \\
&\geqslant \frac{\Delta_\tau^\star \underline{A}(\theta^\star)}{2} ||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
u(\boldsymbol{\tau}, \boldsymbol{\theta}) &\geqslant \frac{\Delta_\tau^\star}{2} \min\left(\frac{\Delta_\tau^\star}{4}, ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V\right) \\
&\geqslant \frac{\Delta_\tau^{\star 2}}{8} ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V.
\end{aligned} \tag{38}$$

Thus, setting $C_{\theta^\star} = \Delta_\tau^{\star 2} \underline{A}(\theta^\star)/8$ and $C_{\tau^\star} = \Delta_\tau^{\star 2}/8$,

$$u(\boldsymbol{\tau}, \boldsymbol{\theta}) \geqslant \max(C_{\theta^\star} ||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty, C_{\tau^\star} ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V).$$

This achieves the proof of Lemma 2.3. □

Proof of Theorem 2.2 (*conclusion*). For any $\delta > 0$, let us define

$$\mathscr{T}_{K,\delta} = \{\boldsymbol{\tau} \in \mathscr{T}_K;\ ||\boldsymbol{\tau} - \boldsymbol{\tau}^\star||_\infty > \delta\}, \tag{39}$$

$$\Theta_{K,\delta} = \{\boldsymbol{\theta} \in \Theta_K;\ ||\boldsymbol{\theta} - \boldsymbol{\theta}^\star||_V > \delta\}. \tag{40}$$

Thus, we have

$$\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K} u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant C_{\theta^\star}\delta,$$

$$\inf_{(\tau,\theta)\in\mathcal{T}_K\times\Theta_{K,\delta}} u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant C_{\tau^\star}\delta$$

and then,

$$\begin{aligned} P(\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty>\delta) &\leqslant P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K} U_n(\boldsymbol{\tau},\boldsymbol{\theta})<0\right)\\ &\leqslant P\left(\sup_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K}|e_n(\boldsymbol{\tau},\boldsymbol{\theta})|>\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K} u(\boldsymbol{\tau},\boldsymbol{\theta})\right). \end{aligned} \tag{41}$$

Using Lemma 2.3 and (29), we have that

$$\begin{aligned} P(\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty>\delta) &\leqslant 2P\left(\max_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K}\sum_{k=1}^{K}|n^{-1}S_n(\boldsymbol{Y}_k,\theta_k)|>\frac{C_{\theta^\star}\delta}{2}\right)\\ &\leqslant 2KP\left(\sup_{1\leqslant t\leqslant t'\leqslant n,\theta\in\Theta}|n^{-1}S_n(Y_t,Y_{t+1},\ldots,Y_{t'},\theta)|>\frac{C_{\theta^\star}\delta}{2K}\right). \end{aligned} \tag{42}$$

In the same way, we show that

$$P(\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^\star\|_V>\delta)\leqslant 2KP\left(\max_{1\leqslant t\leqslant t'\leqslant n,\,\theta\in\Theta}|n^{-1}S_n(Y_t,Y_{t+1},\ldots,Y_{t'},\theta)|>\frac{C_{\tau^\star}\delta}{2K}\right). \tag{43}$$

From (23), the right-hand terms of (42) and(43) converge to 0 and $(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)$ converges to $(\boldsymbol{\tau}^\star,\boldsymbol{\theta}^\star)$ in probability. □

2.7. The rate of convergence

We have the following result concerning the rate of convergence of the sequence of estimates $\{(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)\}$:

Theorem 2.4. *Let*

$$\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty=\max_{1\leqslant j\leqslant K}|\hat{\tau}_j-\tau_j^\star| \quad \textit{and} \quad \|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^\star\|_\infty=\max_{1\leqslant j\leqslant K}\|\hat{\theta}_j-\theta_j^\star\|_2.$$

Assume that conditions H1 *and* H2 *are satisfied. Then, the sequences* $(n\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty)$ *and* $(\sqrt{n^{2-h}}\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^\star\|_\infty)$ *are uniformly tight in P-probability*:

(i) $$\lim_{\delta\to\infty}\lim_{n\to\infty} P(n\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty\geqslant\delta)=0,$$

(ii) $$\lim_{\eta\to\infty}\lim_{n\to\infty} P(\sqrt{n^{2-h}}\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^\star\|_\infty\geqslant\eta)=0.$$

This result means that the rate of convergence of $\hat{\boldsymbol{\tau}}_n$ does not depend on the covariance structure of the sequence (Y_i). For strongly mixing sequences, as well as for strongly

dependent sequences, $\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty=\mathcal{O}_{\mathrm{P}}(n^{-1})$. On the other hand, the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ directly depends on the covariance structure of (Y_i). Indeed, this rate is the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ in the absence of changes.

Proof of Theorem 2.4. We shall show first that, under H1 and H2,

$$\lim_{\delta\to\infty}\lim_{n\to\infty} P(n^{2-h}\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|_\infty \geqslant \delta)=0. \tag{44}$$

For any $\delta>0$ and for $1\leqslant h<2$, let

$$\mathcal{T}_{K,\delta n^{h-2}}=\{\boldsymbol{\tau}\in\mathcal{T}_K;\ \|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty>\delta n^{h-2}\}.$$

Then, for any $\delta>0$,

$$P(n^{2-h}\|\hat{\boldsymbol{\tau}}_n-\boldsymbol{\tau}^\star\|>\delta)\leqslant P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta n^{h-2}}\times\Theta_K} U_n(\boldsymbol{\tau},\boldsymbol{\theta})<0\right). \tag{45}$$

In (29), $e_n(\boldsymbol{\tau},\boldsymbol{\theta})$ is decomposed into a sum. Since the consistency of $\hat{\boldsymbol{\tau}}_n$ is established, $\hat{\tau}_k$ converges to $\tau_k^\star$ for any k. Thus, only the pairs (k,j), such that $|k-j|\leqslant 1$, are present in the sum. Thus, using (27) and(29), expression of U_n reduces to

$$U_n(\boldsymbol{\tau},\boldsymbol{\theta})=u(\boldsymbol{\tau},\boldsymbol{\theta})+\frac{1}{n}\sum_{k=1}^{K}\sum_{j=k-1}^{k+1}(S_n(\boldsymbol{Y}_{kj},\theta_k)-S_n(\boldsymbol{Y}_{kj},\theta_j^\star)). \tag{46}$$

Thus, using (45) and (46), Eq. (44) will be a direct consequence of the two following equalities:

$$\lim_{\delta\to\infty}\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta n^{h-2}}\times\Theta_K}\left(\frac{S_n(\boldsymbol{Y}_{kk},\theta_k)-S_n(\boldsymbol{Y}_{kk},\theta_k^\star)}{n}+Cu(\boldsymbol{\tau},\boldsymbol{\theta})\right)<0\right)=0, \tag{47}$$

$$\lim_{\delta\to\infty}\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta n^{h-2}}\times\Theta_K}\frac{S_n(\boldsymbol{Y}_{k,k+1},\theta_k)}{n}+Cu(\boldsymbol{\tau},\boldsymbol{\theta})<0\right)=0, \tag{48}$$

for any $C>0$.

We establish (47) first. Recall that $\boldsymbol{Y}_{kk}$ represents the observations that belong to segment k in both configurations, $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^\star$: $\boldsymbol{Y}_{kk}=(Y_t,\max(t_{k-1}^\star,t_{k-1})<t\leqslant\min(t_k^\star,t_k))$. Since the consistency of the estimate is established, $\hat{n}_k=\hat{t}_k-\hat{t}_{k-1}$ converges to $n_k^\star=t_k^\star-t_{k-1}^\star$, and $\hat{\theta}_j$ converges to $\theta_j^\star$. Thus, we only consider the configurations $(\boldsymbol{\tau},\boldsymbol{\theta})$ where $n_{kk}>n_k^\star/2$, and where $\theta_k\in\mathcal{V}(\theta_k^\star)$. Thus, using H1 (ii) and (28),

$$u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant\frac{n_{kk}}{n}v(\theta_k^\star,\theta_k)\geqslant\frac{n_{kk}}{n}B\|\theta_k-\theta_k^\star\|_2^2$$

and, from Lemma 2.3,

$$\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta n^{h-2}}\times\Theta_K}u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant\max\left(C_{\theta^\star}\delta n^{h-2},\frac{n_k^\star}{2n}B\|\theta_k-\theta_k^\star\|_2^2\right). \tag{49}$$

Under H1, there exists $\tilde{\theta}\in\Theta$ such that

$$S_n(\boldsymbol{Y}_{kk},\theta_k)-S_n(\boldsymbol{Y}_{kk},\theta_k^\star)=\langle\nabla S_n(\boldsymbol{Y}_{kk},\tilde{\theta}),\theta_k-\theta_k^\star\rangle \tag{50}$$

where $\nabla S_n(\boldsymbol{Y}_{kk},\tilde{\theta})$ was defined in (21). On the other hand, using the fact that $\max(b/a, ca) \geqslant \sqrt{bc}$ for $a,b,c>0$, we have

$$P\left(\sup_{(\tau,\theta)\in\mathscr{T}_{K,\delta n^{h-2}}\times\Theta_K}\left(\frac{1}{n}\sup_{\tilde{\theta}\in\Theta}|\langle\nabla S_n(\boldsymbol{Y}_{kk},\tilde{\theta}),\theta_k-\theta_k^\star\rangle| - Cu(\boldsymbol{\tau},\boldsymbol{\theta})\right)\geqslant 0\right)$$

$$\leqslant P\left(\max_{t_{k-1}^\star<t_{k-1}<t_k\leqslant t_k^\star}\sup_{\tilde{\theta}\in\Theta}||\nabla S_n(Y_{t_{k-1}+1},\ldots,Y_{t_k},\tilde{\theta})||\geqslant\sqrt{C_{\theta^\star}BC\left(\frac{n_k^\star}{2n}\right)}\sqrt{\delta n^h}\right)$$

$$\leqslant\frac{2A_1}{C_{\theta^\star}BC\delta}\left(\frac{n_k^\star}{n}\right)^{h-1} \tag{51}$$

by using (25). This term converges to 0 when $\delta\to\infty$ since $n_k^\star/n$ is a constant.

We show now (48). Here, $\tau_k>\tau_k^\star$ and $\boldsymbol{Y}_{k,k+1}$ represents the observations that belong to segment k in $\boldsymbol{\tau}$ and to segment $k+1$ in $\boldsymbol{\tau}^\star$: $\boldsymbol{Y}_{k,k+1}=(Y_t,\ t_k^\star<t\leqslant t_k)$. (The case $\tau_k<\tau_k^\star$ would be treated in a same way: $\boldsymbol{Y}_{k+1,k}$ represents the observations that belong to segment $k+1$ in $\boldsymbol{\tau}$ and segment k in $\boldsymbol{\tau}^\star$.)

If $\tau_k-\tau_k^\star\leqslant\delta n^{h-2}$, that is, $t_k-t_k^\star\leqslant\delta n^{h-1}$, then

$$P\left(\min_{0\leqslant t_k-t_k^\star\leqslant\delta n^{h-1}}\inf_{\theta\in\Theta_K}\left(\frac{S_n(\boldsymbol{Y}_{k,k+1},\theta_k)}{n}+Cu(\boldsymbol{\tau},\boldsymbol{\theta})\right)<0\right)$$

$$\leqslant P\left(\max_{0\leqslant t_k-t_k^\star\leqslant\delta n^{h-1}}\sup_{\theta_k\in\Theta}||S_n(Y_{t_k^\star+1},\ldots,Y_{t_k},\theta_k)||>C_{\theta^\star}C\delta n^{h-1}\right)$$

$$\leqslant\frac{A_1}{C_{\theta^\star}^2C^2\delta^{2-h}}n^{(h-1)(h-2)} \tag{52}$$

by using (23). On the other hand, if $\tau_k-\tau_k^\star\geqslant\delta n^{h-2}$, we can remark, from Lemma 2.3, that $u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant C_{\theta^\star}(\tau_k-\tau_k^\star)$. Then we have

$$P\left(\min_{t_k-t_k^\star\geqslant\delta n^{h-1}}\inf_{\theta\in\Theta_K}\left(\frac{S_n(\boldsymbol{Y}_{k,k+1},\theta_k)}{n}+Cu(\boldsymbol{\tau},\boldsymbol{\theta})\right)<0\right)$$

$$\leqslant P\left(\min_{t_k-t_k^\star\geqslant\delta n^{h-1}}\inf_{\theta_k\in\Theta}\frac{t_k-t_k^\star}{n}\left(\frac{S(\boldsymbol{Y}_{k,k+1},\theta_k)}{t_k-t_k^\star}+C_{\theta^\star}C\right)<0\right)$$

$$\leqslant P\left(\max_{t_k-t_k^\star\geqslant\delta n^{h-1}}\sup_{\theta_k\in\Theta}\frac{||S_n(Y_{t_k^\star+1},\ldots,Y_{t_k},\theta)||}{t_k-t_k^\star}>C_{\theta^\star}C\right)$$

$$\leqslant\frac{A_2}{C_{\theta^\star}^2C^2\delta^{2-h}}n^{(h-1)(h-2)} \tag{53}$$

by using (24). Eq. (48) is satisfied since the right-hand terms of (53) and (52) converge to 0 when $\delta\to\infty$, since $1\leqslant h<2$.

That achieves the proof of (44). The point (ii) of Theorem 2.4 can be shown in a same way. Indeed, for any $\eta>0$, let $\delta=B\eta^2$, where B has been defined in condition H1, and let

$$\Theta_{K,\delta n^{h-2}}=\{\boldsymbol{\theta}\in\Theta_K;\ ||\boldsymbol{\theta}-\boldsymbol{\theta}^\star||_V>\delta n^{h-2}\}.$$

Then,

$$P(\sqrt{n^{2-h}}\|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_{\infty}>\eta)\leqslant P(n^{2-h}\|\boldsymbol{\theta}-\boldsymbol{\theta}^{\star}\|_{V}>B\eta^{2})$$

$$\leqslant P\left(\inf_{(\tau,\theta)\in\mathscr{T}_K\times\Theta_{K,\delta n^{h-2}}} U_n(\boldsymbol{\tau},\boldsymbol{\theta})<0\right). \tag{54}$$

Following the proof of (44), we show that the right-hand term of (54) goes to 0 when $\eta\to\infty$ and $n\to\infty$.

We shall now be able to improve the rate of convergence of $\hat{\boldsymbol{\tau}}_n$. For a given configuration $\boldsymbol{\tau}\in\mathscr{T}_K$, let $\boldsymbol{\theta}(\boldsymbol{\tau})=(\theta_k(\boldsymbol{\tau}),\ 1\leqslant k\leqslant K)$ be the value of $\boldsymbol{\theta}$ that minimizes $U_n(\boldsymbol{\tau},\boldsymbol{\theta})$. Thus, using this notation, $\hat{\boldsymbol{\theta}}_n=\boldsymbol{\theta}(\hat{\boldsymbol{\tau}}_n)$.

Following again the proof of (44), we must show that

$$\lim_{\delta\to\infty}\lim_{n\to\infty}P\left(\inf_{\tau\in\mathscr{T}_{K,\delta n^{-1}}}\left(\frac{S_n(\boldsymbol{Y}_{kk},\theta_k(\boldsymbol{\tau}))-S_n(\boldsymbol{Y}_{kk},\theta_k(\boldsymbol{\tau}^{\star}))}{n}\right.\right.$$

$$\left.\left.+C(u(\boldsymbol{\tau},\boldsymbol{\theta}(\boldsymbol{\tau}))-u(\boldsymbol{\tau}^{\star},\boldsymbol{\theta}(\boldsymbol{\tau}^{\star})))\right)<0\right)=0, \tag{55}$$

$$\lim_{\delta\to\infty}\lim_{n\to\infty}P\left(\inf_{\tau\in\mathscr{T}_{K,\delta n^{-1}}}\frac{S_n(\boldsymbol{Y}_{k,k+1},\theta_k(\boldsymbol{\tau}))}{n}+C(u(\boldsymbol{\tau},\boldsymbol{\theta}(\boldsymbol{\tau}))-u(\boldsymbol{\tau}^{\star},\boldsymbol{\theta}(\boldsymbol{\tau}^{\star})))<0\right)=0 \tag{56}$$

for any $C>0$. We obtain (56) exactly as we have shown (48). Indeed, (48) remains true if we set $h=1$ in (52) and (53).

We must now show (55). To do that, assume that $\hat{\boldsymbol{\tau}}_n$ is a configuration of change points where $\tau_k\geqslant\tau_k^{\star}$ for any k. From (44) and from Theorem 2.4, $n_{k,k+1}=n(\hat{\tau}_k-\tau_k^{\star})=\mathcal{O}_{\mathrm{p}}(n^{h-1})$ and $\theta_k(\hat{\boldsymbol{\tau}}_n)-\theta_k^{\star}=\mathcal{O}_{\mathrm{p}}(n^{h/2-1})$. Then, using the definition of u and S_n, we can show that

$$u(\boldsymbol{\tau},\boldsymbol{\theta}(\boldsymbol{\tau}))-u(\boldsymbol{\tau}^{\star},\boldsymbol{\theta}(\boldsymbol{\tau}^{\star}))=\sum_{k=1}^{K-1}\frac{n_{k,k+1}}{n}v(\theta_k^{\star},\theta_{k+1}^{\star})(1+\mathcal{O}_{\mathrm{p}}(1)) \tag{57}$$

and

$$\frac{S_n(\boldsymbol{Y}_{kk},\theta_k(\hat{\boldsymbol{\tau}}_n))-S_n(\boldsymbol{Y}_{kk},\theta_k(\boldsymbol{\tau}^{\star}))}{n}=\frac{n_{k,k+1}}{n}\frac{S_n(\boldsymbol{Y}_{kk},\theta_k^{\star})}{n}(1+\mathcal{O}_{\mathrm{p}}(1)). \tag{58}$$

We use the fact that $v(\theta_k^{\star},\theta_{k+1}^{\star})$ is bounded from below, and that $n^{-1}S_n(\boldsymbol{Y}_{kk},\theta_k^{\star})$ goes to 0, uniformly in $\boldsymbol{\tau}$ to show (55). □

3. Detection of an unknown number of changes

Now, we assume that the number of segments K is unknown. Nevertheless, we assume that that this number is upper bounded by a known $\bar{K}$.

We propose to estimate the configuration of changes $\boldsymbol{\tau}$, the vector of parameters $\boldsymbol{\theta}$ and the number of segments K, by minimizing a penalized contrast function $\tilde{J}_n(\boldsymbol{\tau},\boldsymbol{\theta},K)$

defined by

$$\tilde{J}_n(\boldsymbol{\tau},\boldsymbol{\theta},K)=\sum_{k=1}^{K} W_n(\boldsymbol{Y}_k,\theta_k)+\beta_n K \tag{59}$$

for any $K\in\{1,2,\dots,\bar{K}\}$, and any $(\boldsymbol{\tau},\boldsymbol{\theta})\in\mathscr{T}_K\times\Theta_K$.

The sequence $\{\beta_n\}$ is positive and tends to 0 when n tends to infinity. This method is classical for many problems of model selection (see Akaike, 1974; Schwarz, 1978; Hannan, 1980; Yao, 1988; Barron et al., 1999; Dacunha-Castelle and Gassiat, 1997). The parameter β_n is a trade-off between the fit with the observations and the size of the model, that is, the number of segments that cannot be too big. The particular choice $\beta_n=\log n/n$ corresponds to the so-called Schwarz criterion, already proposed by Yao (1988) for estimating an unknown number of jumps in an independent random sequence. In a more general framework, we must adjust the rate of convergence of $\{\beta_n\}$ to 0 according to the rate of convergence of the estimate $(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)$. Indeed, when the number of changes is unknown, we have the following result.

Theorem 3.1. *Let $\{\beta_n\}$ be a positive sequence of real numbers such that*

$$\beta_n \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad n^{2-h}\beta_n \underset{n\to\infty}{\longrightarrow} \infty. \tag{60}$$

Then, under H1 *and* H2, *the minimum penalized contrast estimator $(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n,\hat{K}_n)$, obtained as the solution of the following minimization problem*:

$$\tilde{J}_n(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n,\hat{K}_n)\leqslant\tilde{J}_n(\boldsymbol{\tau},\boldsymbol{\theta},K),\quad \forall(\boldsymbol{\tau},\boldsymbol{\theta},K)\in\mathscr{T}_K\times\Theta_K\times\{1,2,\dots,\bar{K}\}. \tag{61}$$

converges in P-probability to $(\boldsymbol{\tau}^\star,\boldsymbol{\theta}^\star,K^\star)$.

Proof. For any $n>0$, we have

$$U_n(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)+\beta_n\hat{K}_n\leqslant\beta_n K^\star, \tag{62}$$

that is,

$$\hat{K}_n\leqslant-\frac{U_n(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)}{\beta_n}+K^\star. \tag{63}$$

We define now $||\boldsymbol{\tau}-\boldsymbol{\tau}^\star||_\infty$ by

$$||\boldsymbol{\tau}-\boldsymbol{\tau}^\star||_\infty=\max_j\min_k|\tau_k-\tau_j^\star|$$

and Lemma 2.3 still applies: $u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant C_{\theta^\star}||\boldsymbol{\tau}-\boldsymbol{\tau}^\star||_\infty$. Then, for any configuration $\boldsymbol{\tau}$ of K segments, with $K<K^\star$, $||\boldsymbol{\tau}-\boldsymbol{\tau}^\star||_\infty\geqslant\Delta_\tau^\star/2$ and

$$u(\boldsymbol{\tau},\boldsymbol{\theta})\geqslant\frac{C_{\theta^\star}}{2}\Delta_\tau^\star,\quad\forall\theta\in\Theta. \tag{64}$$

Then, if $\hat{K}_n<K^\star$,

$$\beta_n K^\star\geqslant u(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)+e_n(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)+\beta_n\hat{K}_n \tag{65}$$

$$\geqslant\frac{C_{\theta^\star}\Delta_\tau^\star}{2}+e_n(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n)+\beta_n\hat{K}_n \tag{66}$$

and

$$P(\hat{K}_n < K^\star) \leqslant P\left(\max_{1\leqslant K\leqslant K^\star-1}\ \sup_{(\tau,\theta)\in\mathcal{T}_K\times\Theta_K} |e_n(\boldsymbol{\tau},\boldsymbol{\theta})| \geqslant \frac{C_{\theta^\star}\Delta_\tau^\star}{2} - \beta_n K^\star\right). \tag{67}$$

We use the fact that β_n converges to 0 and that $e_n(\boldsymbol{\tau},\boldsymbol{\theta})$ converges to 0 uniformly to conclude that $P(\hat{K}_n < K^\star) \to 0$ when $n \to \infty$. On the other hand, we have

$$\begin{aligned}P(\hat{K}_n > K^\star) &\leqslant P\left(\max_{K^\star+1\leqslant K\leqslant \bar{K}}\ \sup_{(\tau,\theta)\in\mathcal{T}_K\times\Theta_K} -\frac{U_n(\boldsymbol{\tau},\boldsymbol{\theta})}{\beta_n} \geqslant 1\right)\\ &\leqslant P\left(\min_{K^\star+1\leqslant K\leqslant \bar{K}}\ \inf_{(\tau,\theta)\in\mathcal{T}_K\times\Theta_K} u(\boldsymbol{\tau},\boldsymbol{\theta}) + \beta_n + e_n(\boldsymbol{\tau},\boldsymbol{\theta}) \leqslant 0\right).\end{aligned} \tag{68}$$

Thus, we can adapt the proof of Theorem 2.4 by changing $u(\boldsymbol{\tau},\boldsymbol{\theta})$ into $u(\boldsymbol{\tau},\boldsymbol{\theta})+\beta_n$:

Let $\delta_n = \beta_n n^{2-h}$. Under the hypotheses of Theorem 3.1, $\delta_n \to \infty$ when $n \to \infty$. Thus, by using Theorem 2.4, we have that for any $K^\star + 1 \leqslant K \leqslant \bar{K}$,

$$\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K,\beta_n}\times\Theta_K} u(\boldsymbol{\tau},\boldsymbol{\theta}) + e_n(\boldsymbol{\tau},\boldsymbol{\theta}) \leqslant 0\right) = 0,$$

$$\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\mathcal{T}_{K}\times\Theta_{K,\beta_n}} u(\boldsymbol{\tau},\boldsymbol{\theta}) + e_n(\boldsymbol{\tau},\boldsymbol{\theta}) \leqslant 0\right) = 0$$

where $\mathcal{T}_{K,\beta_n}$ and Θ_{K,β_n} have been defined in (39) and (40). Let

$$\begin{aligned}\tilde{\mathcal{T}}_{K,\beta_n} &= \mathcal{T}_K \setminus \mathcal{T}_{K,\beta_n} = \{\boldsymbol{\tau}\in\mathcal{T}_K;\ \|\boldsymbol{\tau}-\boldsymbol{\tau}^\star\|_\infty \leqslant \beta_n\},\\ \tilde{\Theta}_{K,\beta_n} &= \Theta_K \setminus \Theta_{K,\beta_n} = \{\boldsymbol{\theta}\in\Theta_K;\ \|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_V \leqslant \beta_n\}.\end{aligned}$$

We have to show that, for any $K^\star + 1 \leqslant K \leqslant \bar{K}$, for any $\delta > 0$,

$$\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\tilde{\mathcal{T}}_{K,\beta_n}\times\tilde{\Theta}_{K,\beta_n}} \beta_n + e_n(\boldsymbol{\tau},\boldsymbol{\theta}) \leqslant 0\right) = 0. \tag{69}$$

Thus, using the decomposition (29) of $e_n(\boldsymbol{\tau},\boldsymbol{\theta})$ into a sum, it is enough to show that, for any $K^\star + 1 \leqslant K \leqslant \bar{K}$, for any pair (k,j), and for any $C > 0$,

$$\lim_{n\to\infty} P\left(\inf_{(\tau,\theta)\in\tilde{\mathcal{T}}_{K,\beta_n}\times\tilde{\Theta}_{K,\beta_n}} \left(\frac{S_n(\boldsymbol{Y}_{kj},\theta_k)}{n} - \frac{S_n(\boldsymbol{Y}_{kj},\theta_j^\star)}{n} + C\beta_n\right) \leqslant 0\right) = 0. \tag{70}$$

Consider first that k and j are such that $t_{j-1}^\star \leqslant t_{k-1} < t_k \leqslant t_j^\star$: segment k of $\boldsymbol{\tau}$ is included in segment j of $\boldsymbol{\tau}^\star$. Since $\boldsymbol{\theta}\in\tilde{\Theta}_{K,\beta_n}$, $\|\theta_k - \theta_j^\star\| \leqslant \sqrt{\beta_n/B}$ and then,

$$\begin{aligned}&P\left(\min_{t_{j-1}^\star \leqslant t_{k-1} < t_k \leqslant t_j^\star}\ \inf_{\theta\in\tilde{\Theta}_{K,\beta_n}} \left(\frac{S_n(\boldsymbol{Y}_{kj},\theta_k)}{n} - \frac{S_n(\boldsymbol{Y}_{kj},\theta_j^\star)}{n} + C\beta_n\right) \leqslant 0\right)\\ &\quad\leqslant P\left(\max_{t_{j-1}^\star \leqslant t_{k-1} < t_k \leqslant t_j^\star}\ \sup_{\tilde{\theta}\in\Theta} \|\Delta S_n(\boldsymbol{Y}_{kj},\tilde{\theta})\| \geqslant \frac{\sqrt{B}Cn}{\sqrt{\beta_n}}\right)\\ &\quad\leqslant \frac{8A_1}{BC^2}\left(\frac{n^{h-2}}{\beta_n}\right).\end{aligned} \tag{71}$$

Consider now that k and j are such that $t_{k-1} < t_j^\star < t_k$: a part of segment k of $\boldsymbol{\tau}$ belongs to segment j of $\boldsymbol{\tau}^\star$ and another part belongs to segment $j+1$. Since $\boldsymbol{\tau} \in \tilde{\mathcal{T}}_{K,\beta_n}$, $t_k - t_j^\star \leqslant \delta n^{h-1}$, and then,

$$
\begin{aligned}
&P\left(\min_{t_j^\star \leqslant t_k \leqslant t_j^\star + n\beta_n} \inf_{\theta \in \tilde{\Theta}_{K,\beta_n}} \left(\frac{S_n(\boldsymbol{Y}_{kj}, \theta_k)}{n} + C\beta_n\right) \leqslant 0\right) \\
&\quad \leqslant P\left(\max_{s \leqslant n\beta_n} \sup_{\theta \in \Theta} |S_n(Y_{t_j^\star+1} \ldots Y_{t_j^\star+s}, \theta)| > Cn\beta_n\right) \\
&\quad \leqslant \frac{A_1}{C^2}(n\beta_n)^{h-2}.
\end{aligned} \tag{72}
$$

When the conditions of Theorem 3.1 are satisfied, the right terms of (72) and (73) converge to 0 when $n \to \infty$ since $n\beta_n \to \infty$ when $n \to \infty$.

We finally show that $P(\hat{K}_n \neq K^\star)$ converges to 0. □

4. Some examples of application

4.1. Detection of changes in the mean of a sequence of random variables

We consider the following model:

$$Y_i = \mu(i) + \varepsilon_i, \quad i = 0, 1, 2, \ldots, n \tag{73}$$

where μ is piecewise constant, and where (ε_i) is a second-order stationary process with zero mean and variance $\sigma^2(i)$. The function $\sigma(i)$ is assumed to be constant here and the changes affect the function μ, that is the mean of Y.

Here, for $1 \leqslant k \leqslant K$, $\mu_k = \mu(i)$ for any $t_{k-1} + 1 \leqslant i \leqslant t_k$, and the vector of parameter to be estimated is $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K)$.

We estimate $\theta_k = \mu_k$ in segment k by minimizing $W_n(\boldsymbol{Y}_k, \mu_k)$ defined by

$$W_n(\boldsymbol{Y}_k, \mu_k) = \frac{1}{n} \sum_{i=t_{k-1}+1}^{t_k} (Y_i - \mu_k)^2. \tag{74}$$

(Of course, $\hat{\mu}_k$ is the empirical mean of Y computed in segment k: $\hat{\mu}_k = \bar{\boldsymbol{Y}}_k$.) We estimate $(\boldsymbol{\tau}, \boldsymbol{\theta})$ by minimizing the function $J_n(\boldsymbol{\tau}, \boldsymbol{\theta})$ defined by

$$J_n(\boldsymbol{\tau}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^{K} ||\boldsymbol{Y}_k - \mu_k||^2, \tag{75}$$

that is, by minimizing the function $U_n(\boldsymbol{\tau}, \boldsymbol{\theta})$ defined in (27) by

$$
\begin{aligned}
U_n(\boldsymbol{\tau}, \boldsymbol{\theta}) &= J_n(\boldsymbol{\tau}, \boldsymbol{\theta}) - J_n(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star) \\
&= \sum_{k=1}^{K} \sum_{j=1}^{K} \frac{n_{kj}}{n} (\mu_j^\star - \mu_k)^2 - 2 \sum_{k=1}^{K} \sum_{j=1}^{K} \frac{n_{kj}}{n} \overline{\varepsilon_{kj}} (\mu_k - \mu_j^\star)
\end{aligned} \tag{76}
$$

where $\overline{\varepsilon_{kj}}$ is the empirical mean of ε_{kj}, that is, the ε_t that belong to segment j in configuration $\boldsymbol{\tau}^\star$ and to segment k in configuration $\boldsymbol{\tau}$.

Here, condition H1 is satisfied, $v(\theta_1, \theta_2) = (\theta_1 - \theta_2)^2$ and $\xi(Y_i) = Y_i$.

Then, using the definition of $\eta_i(\theta)$ given in (9), we have, for any $1 \leqslant i \leqslant n$, and any $\theta = \mu$,

$$\eta_i(\theta) = -2\mu\varepsilon_i, \tag{77}$$

$$\nabla\eta_i(\theta) = -2\varepsilon_i. \tag{78}$$

Condition H2 is satisfied when (ε_i) satisfies (16) for some $1 \leqslant h < 2$. This example was studied in detail by Lavielle and Moulines (1999).

4.2. Detection of changes in the mean and the variance of a sequence of random variables

Here, we want to detect simultaneously changes in the mean and the variance of (Y_i) defined in (73). That means that the changes affect both function μ and σ. Let $\theta_k = (\mu_k, \sigma_k^2)$ for any k, where μ_k and σ_k^2 are the mean and the variance of Y_i in segment k. The set Θ has the form $\Theta = [A, B] \times [C, D]$ where $C > 0$. We estimate θ_k in segment k by minimizing $W_n(\boldsymbol{Y}_k, \theta_k)$ defined by

$$W_n(\boldsymbol{Y}_k, \theta_k) = \frac{||\boldsymbol{Y}_k - \mu_k||^2}{n\sigma_k^2} + \frac{n_k}{n}\log\sigma_k^2. \tag{79}$$

We estimate $(\boldsymbol{\tau}^\star, \boldsymbol{\theta}^\star)$ by minimizing the function $J_n(\boldsymbol{\tau}, \boldsymbol{\theta})$ defined by

$$J_n(\boldsymbol{\tau}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\left(\frac{||\boldsymbol{Y}_k - \mu_k||^2}{\sigma_k^2} + n_k\log\sigma_k^2\right) \tag{80}$$

that is, $(\hat{\boldsymbol{\tau}}_n, \hat{\boldsymbol{\theta}}_n)$ minimizes

$$U_n(\boldsymbol{\tau}, \boldsymbol{\theta}) = \sum_{k=1}^{K}\sum_{j=1}^{K}\frac{n_{kj}}{n}v(\theta_j^\star, \theta_k) + \frac{1}{n}\left(\sum_{k=1}^{K}\sum_{i=t_{k-1}}^{t_k}\eta_i(\theta_k) - \sum_{j=1}^{K}\sum_{i=t_{j-1}^\star}^{t_j^\star}\eta_i(\theta_j^\star)\right) \tag{81}$$

where

$$v(\theta, \theta') = \frac{(\mu - \mu')^2}{\sigma'^2} + \log\left(\frac{\sigma'^2}{\sigma^2}\right) + \frac{\sigma^2}{\sigma'^2} - 1 \tag{82}$$

is the Kullback–Lieber distance beetween two Gaussian distributions. Here, for any $1 \leqslant i \leqslant n$ and any $\theta = (\mu, \sigma^2)$,

$$\eta_i(\theta) = -2\frac{\mu}{\mathrm{Var}\, Y_i}\varepsilon_i + \frac{1}{\sigma^2}\zeta_i \tag{83}$$

where

$$\varepsilon_i = Y_i - EY_i,$$

$$\zeta_i = (Y_i - EY_i)^2 - \mathrm{Var}\, Y_i.$$

Theorems 2.2, 2.4 and 3.1 apply if both ε and ζ satisfy (16) for $1 \leqslant h < 2$. For example, if ε is a stationary Gaussian process with autocovariance function γ_ε,

$$E\left(\sum_{i=1}^{t}\varepsilon_i\right)^2 = \sum_{l=-t+1}^{t-1}(t-|l|)\gamma_\varepsilon(l),$$

$$E\left(\sum_{i=1}^{t}\zeta_i\right)^2 = \sum_{l=-t+1}^{t-1}(t-|l|)\,\gamma_\varepsilon^2(l).$$

Thus, it is enough to check that ε satisfies (16).

In this particular case, (84) gives a decomposition of $\eta_i(\theta)$ in the Hermite polynomials. We see that the rank of Hermite of $\eta_i(\theta)$ is $m=1$. Hence, we can conclude that, if $\gamma_\varepsilon(t)=\mathcal{O}(t^{-a})$ for some $a>0$, the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ is $n^{\min(a/2,1/2-\delta)}$, for any $\delta>0$.

Remark. Assume that we are looking for changes only in the variance of the process, and not in the mean, $\boldsymbol{\theta}=(\sigma_k^2)$. Assuming that $EY_i=\mu$ for any $i\geqslant 1$, we estimate $(\boldsymbol{\tau},\boldsymbol{\theta})$ by minimizing the function $J_n(\boldsymbol{\tau},\boldsymbol{\theta})$ defined by

$$J_n(\boldsymbol{\tau},\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\left(\frac{||\boldsymbol{Y}_k-\mu||^2}{\sigma_k^2} + n_k\log\sigma_k^2\right). \tag{84}$$

In this case, for any $\theta=\sigma^2$,

$$\eta_i(\theta) = \frac{1}{\sigma^2}((Y_i-EY_i)^2 - \operatorname{Var} Y_i) \tag{85}$$

and the Hermite rank of η is 2. Then, if $\gamma_\varepsilon(t)=\mathcal{O}(t^{-a})$ for some $a>0$, the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ is $n^{\min(a,1/2-\delta)}$, for any $\delta>0$.

We apply this method for detecting changes in the distribution of financial assests. The series of the CAC 40 index is displayed in Fig. 1. We estimate the configuration of change points by minimizing the penalized contrast function proposed in (59) with the function J_n proposed in (80):

$$(\hat{\boldsymbol{\tau}}_n,\hat{\boldsymbol{\theta}}_n,\hat{K}_n) = \arg\min_{1\leqslant K\leqslant\bar{K}}\inf_{(\tau,\theta)\in\mathcal{T}_{K,\delta}\times\Theta_K}\left\{\frac{1}{n}\sum_{k=1}^{K}\left(\frac{||\boldsymbol{Y}_k-\mu_k||^2}{\sigma_k^2}+n_k\log\sigma_k^2\right)+\ \beta_n K\right\}. \tag{86}$$

The estimated configuration displayed in Fig. 1 was obtained with $\beta_n=20$. Of course, with a finite number of observations, we must adjust the parameter β_n in order to obtain a good resolution level in the segmentation.

We can see that there are no sudden changes in the mean which remains very close to zero. However, we detect three regimes in the variances of returns, that is, three different regimes in the pattern of the market volatility: the second interval is very short and corresponds to high volatility in the market place, the first and the fourth interval may refer to a regime of low volatility, while the third interval corresponds to a period of stable volatility.

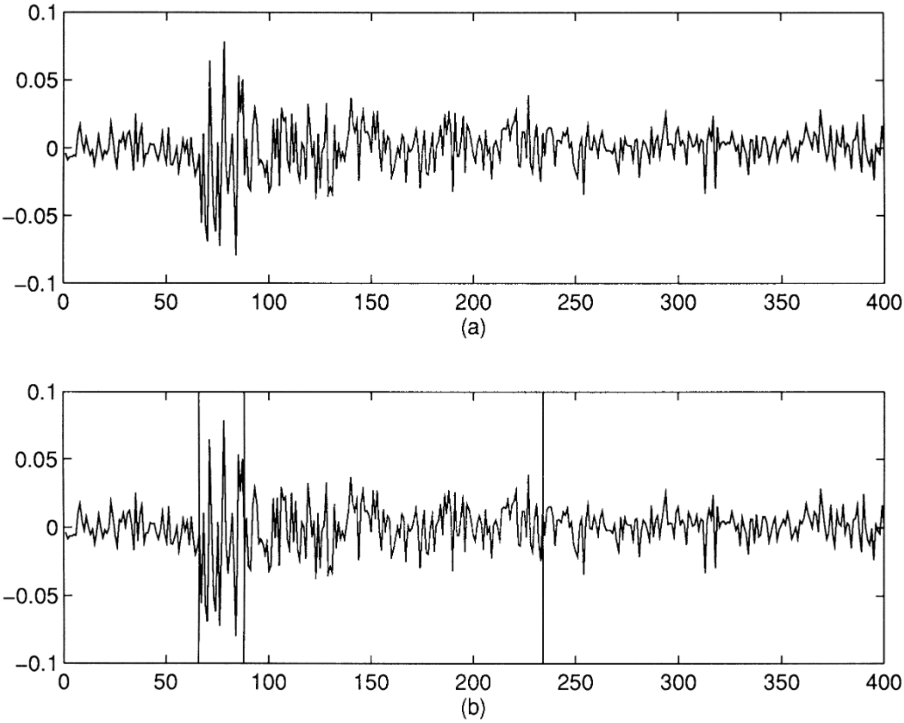

Fig. 1. Detection of changes in the CAC 40 index: (a) the observed series, (b) the estimated configuration of changes.

4.3. Detection of changes in a discrete distribution

We consider here the case where Y_i takes its values in a finite set, $\mathscr{A}=\{c_1,c_2,\dots,c_M\}$. For any $t^\star_{k-1}+1\leqslant i\leqslant t^\star_k$, let $p_{km}=P(Y_i=c_m)$.

The changes affect here the vector of probabilities $\theta_k=(p_{km},\ 1\leqslant m\leqslant M)$, with $\sum_{m=1}^M p_{km}=1$. We assume that there exists $0<a<b<1$ such that $a<p_{km}<b$ for any $1\leqslant k\leqslant K$ and any $1\leqslant m\leqslant M$. Thus,

$$\Theta=\left\{\theta=(p_1,\dots,p_M),\ a<p_m<b,\ \sum_{m=1}^M p_m=1\right\}.$$

We estimate θ_k in segment k by minimizing $W_n(\boldsymbol{Y}_k,\theta)$ in Θ defined by

$$W_n(\boldsymbol{Y}_k,\theta)=-\frac{1}{n}\sum_{m=1}^M n_{km}\log p_{km} \tag{87}$$

where n_{km} is the number of observations in segment k that take the value c_m. Here,

$$v(\theta_k,\theta_j)=\sum_{m=1}^M p_{jm}\log\left(\frac{p_{jm}}{p_{km}}\right)$$

is the Kullback–Lieber distance beetween two discrete distributions, and condition H1 is satisfied. On the other hand, for any $\theta \in \Theta$ and any $1 \leqslant i \leqslant n$, we have

$$\eta_i(\theta) = \sum_{m=1}^{M} \log p_m(\mathbb{1}_{Y_i=c_m} - P(Y_i = c_m)) \tag{88}$$

and condition H2 is satisfied. Here, inequality (16) is satisfied with a value of h that depends on the autocovariance structure of $(\eta_i(\theta))$, that is on the strong mixing coefficients of (Y_i). In fact, there exists a constant $C > 0$ such that, for any $\theta \in \Theta$,

$$\begin{aligned} E\eta_t(\theta)\eta_{t+s}(\theta) = {} & \sum_{m=1}^{M}\sum_{\ell=1}^{M} \log p_m \log p_\ell (P(Y_t = c_m, Y_{t+s} = c_\ell) \\ & -P(Y_t = c_m)P(Y_{t+s} = c_\ell)) \\ \leqslant {} & C\alpha(s) \end{aligned} \tag{89}$$

where $(\alpha(s))$ is the sequence of strong mixing coefficients of (Y_i). Then, if $\alpha(s) = \mathcal{O}(s^{-a})$, the rate of convergence of $\hat{\boldsymbol{\theta}}_n$ is $n^{\min(a/2, 1/2-\delta)}$, for any $\delta > 0$.

We present an application to real data to illustrate the method. Fig. 2 represents the heart rate of a new-born baby. It can be very useful to identify automatically heavy and light sleep periods from this series.

In this example, the observed process $Z = (Z_i)$ is not discrete. Nevertheless, we define a discrete process Y as follows:

$$Y_i = m \quad \text{if } x_{m-1} < Z_t \leqslant x_m$$

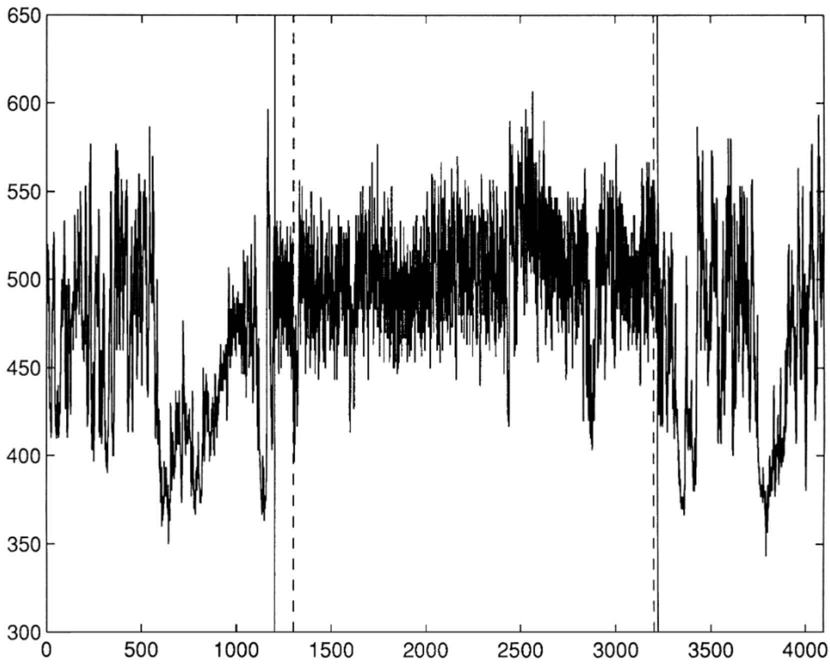

Fig. 2. Detection of changes in the heart-rate of a new-born baby: (- - -) the configuration of changes obtained with external measurements, (—) the estimated configuration of changes.

where $(x_m,\ 0 \leqslant m \leqslant M)$ is a sequence of real numbers such that $x_0 < x_1 < \cdots < x_M$. For this application, we used a equally spaced sequence (x_m) with $M = 20$. Assuming that the changes that affect the distribution of Z affect the distribution of Y, we shall recover the configuration of changes as follows:

$$(\hat{\tau}_n, \hat{\theta}_n, \hat{K}_n) = \arg \min_{1 \leqslant K \leqslant \bar{K}} \inf_{(\tau,\theta) \in \mathcal{T}_{K,\delta} \times \Theta_K} \left\{ -\frac{1}{n} \sum_{k=1}^{K} \sum_{m=1}^{M} n_{km} \log p_{km} + \beta_n K \right\}. \tag{90}$$

External measurements (such as that of the eye-lids' movements) allow us to know that the heavy sleep period is approximatively between data 1300 and data 3200. We can see in Fig. 2 that the changes detected by the algorithm with $\beta_n = 300$ agree with the exact instants of change.

Of course, looking at Fig. 2, one is tempted to introduce an additional change at about $t = 500$. Indeed, this change is well detected with a smaller value of β_n, but it is very difficult to decide, with only one trajectory, if there is really a change which affects the marginal distribution of the data, or if this "jump" is due to the dependence structure of the data. Nevertheless, this example is very interesting, since the two significative changes are well recovered without any assumptions on the dependence structure of the data.

Acknowledgements

The author wishes to thank Paul Doukham for many helpful discussions which helped to improve the manuscript.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control AC-19, 716–723.

Bai, J., 1994. Least Square estimation of a shift in linear processes. J. Time Ser. Anal. 15 (5), 453–472.

Barron, A., Birgé, L., Massart, P., 1999. Risk bounds for model selection via penalization. Probab. Theory Related Fields 113 (3), 301–413.

Basseville, M., Nikiforov, N., 1993. The Detection of Abrupt Changes – Theory and Applications. Information and System Sciences Series, Prentice-Hall, Englewood Cliffs, NJ.

Brodsky, B., Darkhovsky, B., 1993. Nonparametric Methods in Change-Point Problems. Kluwer Academic Publishers, the Netherlands.

Dacunha-Castelle, D., Gassiat, E., 1997. Estimation of the order of a mixture. Bernoulli 3, 279–299.

Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Ann. Statist. 25, 1–37.

Davydov, Y., 1970. The invariance principle for stationary processes. Theory Probab. Appl. 15, 487–498.

Doukhan, P., 1994. Mixing, Properties and Examples. Lecture Notes in Statistics, vol. 85. Springer, Berlin.

Epps, T., 1988. Testing that a Gaussian process is stationary. Ann. Statist. 16, 1667–1683.

Giraitis, L., Leipus, R., 1992. Testing and estimating in the change-point problem of the spectral function. Lithuanian Math. J. 32, 20–38.

Giraitis, L., Leipus, R., Surgailis, D., 1996. The change-point problem for dependent observations. J. Statist. Plan. Inference 53, 297–310.

Giraitis, L., Surgailis, D., 1986. Multivariate Appell polynomial and the central limit theorem. In: Eberlein, E., Taqqu, M. (Eds.), Dependence in Probability and Statistics. Birkhauser, Boston, pp. 27–51.

Hájek, J., Rényi, A., 1955. Generalization of an inequality of Kolmogorov. Acta. Math. Acad. Sci. 6, 281–283.

Hannan, J., 1980. The estimation of the order of an ARMA process. Ann. Statist. 8, 1071–1081.

Kim, H., 1994. Likelihood ratio and cumulative sum tests for a change-point in linear regression. J. Multivariate Anal. 51, 54–70.
Lavielle, M., 1998. Optimal segmentation of random processes. IEEE Trans. Signal Process. 46 (5) 1365–1373.
Lavielle, M., Moulines, E., 1999. Least Squares estimation of an unknown number of shifts in a time series. J. Time-Ser. Anal., in press.
Lombard, F., 1987. Rank tests for change point problems. Biometrika 74, 615–624.
Mia, B., Zhao, L., 1988. Detection of change points using rank methods. Comm. Statist. – Theory Methods 17, 3207–3217.
Móricz, F., Serfling, R., Stout, W., 1982. Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. Ann. Probab. 10 (4), 1032–1040.
Oodaira, H., Yoshihara, K., 1972. Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. Kodai Math. Sem. Rep. 24, 259–269.
Picard, D., 1985. Testing and estimating change points in time series. J. Appl. Probab. 17, 841–867.
Rio, E., 1993. Covariance inequalities for strongly mixing processes. Ann. I.H.P. 29, 587–597.
Schechtman, E., Wolfe, D., 1985. Multiple change points problem – nonparametric procedures for estimation of the points of change. Comm. Statist. – Simul. Comput. 14, 615–631.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Taqqu, M., 1977. Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit long range dependence. Z. Wahrsch. verw. Geb. 40, 203–238.
Vostrikova, L., 1981. Detecting "disorder" in multidimensional random processes. Soviet. Math. Dokl. 24, 55–59.
Yao, Y., 1988. Estimating the number of change-points via Schwarz criterion. Statist. Probab. Lett. 6, 181–189.