Proceedings of the 18th World Congress
The International Federation of Automatic Control
Milano (Italy) August 28 - September 2, 2011

Change point detection of autoregressive process with unknown parameters ⋆

Sergey E. Vorobeychikov * **Yulya B. Burkatovskaya** **

* *Department of Applied Mathematics and Cybernetics, Tomsk State University, 36 Lenin Prospekt, 634050 Tomsk, Russia (e-mail: sev@mail.tsu.ru).*
** *Institute of Cybernetics, Tomsk Polytechnical University, 30 Lenin Prospekt, 634050 Tomsk, Russia (e-mail: tracey@tpu.ru).*

Abstract: The problem of detecting the parameters change point in the autoregressive process is considered. The values of the process parameters before and after the change point are supposed to be unknown. The procedure of change point detection based on the sequential estimations of unknown parameters is proposed, procedure characteristics are studied. Results of numerical simulations are presented.

Keywords: sequential algorithms, change point detection, weighted least square method, guaranteed estimation, autoregressive process.

1. INTRODUCTION AND PROBLEM STATEMENT

Evaluation of experimental data reveals that the mathematical models of a process at different intervals often significantly differ from each other. The problem of determining the change moment of the statistical characteristics of a random process is the classic one and is known as the change point detection problem. Various algorithms of the change point detection applicable for different levels of available information about the observed process and its parameters are known. For the processes with dependent values and, particularly, for the processes of the autoregressive type, several algorithms have been proposed (see Brodskiy B.B. et al. (1992), Basseville M. et al. (1985), Lai T.Z. (1995)), whose quality was studied by simulation. Theoretical investigation of algorithms properties for a fixed sample size is usually impossible; instead, the asymptotic properties for an unlimited increase of the sample size are studied. Recently, a method to detect the point of the parameter change was proposed (Vorobejchikov S. E., Ponomarjova J. S. (2002)) for a stochastic process which parameters are assumed to be known before the change point and unknown after it. In this study we propose a consistent procedure for detecting the change point from one set of unknown parameters to another set of unknown parameters. The procedure is based on a comparison of parameters estimators at different observation intervals.

We consider a scalar autoregressive process specified by the equation

$$x_{k+1} = A_k \lambda + B\xi_{k+1}, \tag{1}$$

where $\{\xi_k\}_{k\geq 0}$ is a sequence of independent identically distributed random variables with zero mean and unit variance. The density distribution function $f_\xi(x)$ of $\{\xi_k\}_{k\geq 0}$ is strictly positive for any x. The value $m > 1$ defines the order of the process, $A_k = [x_k, \ldots, x_{k-m+1}]$ is the $1 \times m$ matrix, $\lambda = [\lambda_1, \ldots, \lambda_m]$ is the parameter vector of dimension $m \times 1$, the noise variance B can be either known or unknown. The value of the parameter vector λ changes from μ_0 to μ_1 at the change point θ. Values of the parameters before and after θ are supposed to be unknown. The difference between μ_0 and μ_1 satisfies the condition

$$(\mu_0 - \mu_1)^T(\mu_0 - \mu_1) \geq \Delta, \tag{2}$$

where Δ is a known value defining the minimum difference between the parameters before and after the change point. The problem is to detect the change point θ from observations x_k.

2. SEQUENTIAL CHANGE POINT DETECTION

First we describe the general scheme of the change point detection procedure. Since the parameters both before and after the change point are unknown, it is logical to use estimators of the unknown parameters in the change point detection procedure. We use the sequential estimators proposed in Vorobejchikov S. E., Meder N. A. (2002) for a more general model with an arbitrary matrix A_k. The main advantage of the estimators is their preassigned mean square accuracy depending on the parameter of the estimation procedure.

At the first stage, we define intervals $[\tau_{i-1}+1, \tau_i]$, $i \geq 1$. The estimators λ_i^* of the parameters of process (1) are constructed on each interval. Then the estimators on intervals $[\tau_{i-l-1}+1, \tau_{i-l}]$ and $[\tau_{i-1}+1, \tau_i]$, where $l > 1$ is an integer, are compared. If the interval $[\tau_{i-1}+1, \tau_i]$ does not include the change point θ, then vector λ on this interval is constant. It can be equal to the initial value μ_0 or the final value μ_1. Thus, for certain i, if $\tau_{i-l} < \theta < \tau_{i-1}+1$, the difference between values of the parameters on intervals $[\tau_{i-l-1}+1, \tau_{i-l}]$ and $[\tau_{i-1}+1, \tau_i]$ is no less then Δ. This is the key property for the change point detection.

⋆ Research was supported by RFBR 09-01-00172a Grant

2.1 Guaranteed parameter estimator

First we consider the case of the known variance B^2 in (1). Denote

$$C(N_1, N) = \sum_{k=N_1}^{N} v_k A_k^T A_k. \tag{3}$$

Non-negative weights $v_k = v_k(x)$ are found from the following condition:

$$\frac{\nu_{\min}(N_1, k)}{B^2} = \sum_{l=N_1+\sigma}^{k} v_l^2 A_l A_l^T, \tag{4}$$

where $\nu_{\min}(N_1, k)$ is the minimum eigenvalue of the matrix $C(N_1, k)$, σ is the minimum volume of observations for which the matrix $C(N_1, N_1+\sigma)$ is not degenerate. The weights on the interval $[N_1, N_1+\sigma]$ are defined as

$$v_k = \begin{cases} \dfrac{1}{B\sqrt{A_k A_k^T}}, & \text{if } A_{N_1}, \ldots, A_k \text{ are linearly independent;} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Choosing a certain positive parameter H, we find the random stopping time τ from the following condition:

$$\tau = \tau(H) = \inf\left(N > N_1 : \nu_{\min}(N_1, N) \geq H\right). \tag{6}$$

At the instant τ, the weight is found from the condition:

$$\frac{\nu_{\min}(N_1, \tau)}{B^2} \geq \sum_{l=N_1+\sigma}^{\tau} v_l^2 A_l A_l^T, \quad \nu_{\min}(N_1, \tau) = H. \tag{7}$$

The parameter estimator is defined in the form

$$\begin{aligned} \lambda^*(H) &= C^{-1}(N_1, \tau) \sum_{k=N_1}^{\tau} v_k A_k^T x_{k+1}; \\ C(N_1, \tau) &= \sum_{k=N_1}^{\tau} v_k A_k^T A_k. \end{aligned} \tag{8}$$

The properties of the estimator were established in Vorobejchikov S. E., Meder N. A. (2002)

Theorem 2.1. Let the parameter λ in (1) be constant and the weights v_k determined in (4–5) be such that

$$\sum_{k=0}^{\infty} v_k^2 A_k A_k^T = \infty \text{ a.s.} \tag{9}$$

Then the stopping time τ (6) is finite with probability one and the mean square accuracy of estimator (8) is bounded from above

$$\begin{aligned} E||\lambda^*(H) - \lambda||^2 &\leq \frac{P(H)}{H^2}, \\ P(H) &= H + m - 1. \end{aligned} \tag{10}$$

If the noise variance B^2 in (1) is unknown and the distribution of the noise ξ_k is known, then estimation procedure (4–8) is modified (Vorobejchikov S. E., Meder N. A. (2002)). The estimator is constructed in the form

$$\begin{aligned} \tilde{\lambda}^*(H) &= C^{-1}(N_1+n, \tau) \sum_{k=N_1+n}^{\tau} v_k A_k^T x_{k+1}; \\ C(N_1+n, \tau) &= \sum_{k=N_1+n}^{\tau} v_k A_k^T A_k. \end{aligned} \tag{11}$$

The weights v_k are found from the following condition:

$$\frac{\nu_{\min}(N_1+n, k)}{\Gamma(N_1, n)} = \sum_{l=N_1+n+\sigma}^{k} v_l^2 A_l A_l^T. \tag{12}$$

Here the multiplier $\Gamma(N_1, n)$ compensates for the influence of the unknown noise variance, n is an initial sample size for the estimation of the noise variance,

$$\begin{aligned} \Gamma(N_1, n) &= D(N_1, n) \sum_{l=N_1}^{N_1+n-1} x_l^2; \\ D(N_1, n) &= E\left(\sum_{l=N_1}^{N_1+n-1} \xi_l^2\right)^{-1}. \end{aligned} \tag{13}$$

Let the expectation $D(N_1, n)$ in (13) exists for given n. Then the weights on the interval $[N_1+n, N_1+n+\sigma]$ are

$$v_k = \begin{cases} \dfrac{1}{\sqrt{\Gamma(N_1, n) A_k A_k^T}}, & \text{if } A_{N_1}, \ldots, A_k \text{ are linearly independent;} \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Properties of the proposed estimator are the same as in the case of the known variance.

We show that condition (9) holds true for process (1). Consider the case of the known noise variance. Without loss of generality, we can suppose that $B^2 = 1$. Series in (9) converges if and only if $\forall \varepsilon > 0$ (Shiryaev A. N. (1980))

$$P\left\{\sum_{k \geq 1} v_{N+k}^2 A_{N+k} A_{N+k}^T \geq \varepsilon\right\} \to 0, \quad N \to \infty. \tag{15}$$

Note that the matrix $C(N_1, N)$ is not diagonal. Indeed, the matrix $C(N_1, N_1) = v_{N_1} A_{N_1}^T A_{N_1}^T$ is not diagonal. If the matrix $C(N_1, N-1)$ is not diagonal, then the next matrix $C(N_1, N)$ is diagonal if and only if

$$\begin{aligned} \sum_{k=N_1}^{N} v_k x_k^2 = \sum_{k=N_1}^{N} v_k x_{k-1}^2 = \ldots = \sum_{k=N_1}^{N} v_k x_{k-m+1}^2; \\ \sum_{k=N_1}^{N} v_k x_{k-i} x_{k-j} = 0, \ \forall\, 0 < i < j < m. \end{aligned} \tag{16}$$

Thus, at the instant N we have $(m-1) + m(m-1)/2$ equation for x_N and v_N. Taking into account that x_N can take any value, we have that v_N satisfies equations (16) with zero probability for any $x_{N_1}, \ldots, x_{N-1}$.

Thus, $\nu_{\min}(N_1, N)$ is not unique eigenvalue of the matrix $C(N_1, N)$ for any N.

According to the definition of the minimum eigenvalue of matrix

$$\nu_{\min}(N_1, N) = \min_{x:||x||=1} (x, C(N_1, N)x),$$

where (x, y) is the scalar product of vectors x and y. Then using (3), we have

$$\begin{aligned} \nu_{\min}(N_1, N) &= \min_{y:||y||=1} (y, ((C(N_1, N-1) + v_N A_N^T A_N)y) \\ &= \min_{y:||y||=1} ((y, C(N_1, N-1)y) + v_N (A_N y)^2). \end{aligned}$$

Let y_N be the argument of the minimum in the last equation. According to (4), we have

$$(y_N, C(N_1, N-1)y_N) + v_{N+1}(A_N y_N)^2 \\ = \nu_{\min}(N_1, N-1) + v_N^2 A_N A_N^T.$$

So we have derived the quadratic equation for v_N with roots in the form

$$v_{1,2} = \frac{1}{2A_N A_N^T}\left[(A_N y_N)^2 \pm \left((A_N y_N)^4 \right.\right. \\ \left.\left. +4A_N A_N^T((y_N, C(N_1, N-1)y_N) - \nu_{\min}(N_1, N-1))\right)^{1/2}\right].$$

It is obvious that

$$(y_N, C(N_1, N-1)y_N) - \nu_{\min}(N_1, N-1) \geq 0.$$

Thus, the following two cases are possible.

Case 1. The equation has two zero roots: $v_1 = v_2 = 0$. This is possible if and only if y_N is the eigenvector of the matrix $C(N_1, N-1)$ corresponding to $\nu_{\min}(N_1, N-1)$ and $A_N y_N = 0$. However, at the instant N the first component of A_N, i.e. x_N, can take any value. Thus, the vector A_N is orthogonal to the given eigenvector of the matrix $C(N_1, N-1)$ with zero probability.

Case 2. The equation has one non-positive and one positive root. Taking the major root as v_N, we have

$$v_n^2 A_N A_N^T \geq \frac{(A_N y_N)^4}{2A_N A_N^T} \\ +(y_N, C(N_1, N-1)y_N) - \nu_{\min}(N_1, N-1). \tag{17}$$

The first term in (17) is equal to $A_N A_N^T \cos^4(\alpha_N)/2$, where α_N is the angle between A_N and y_N. Since $A_N A_N^T$ does not converge to zero, the first term converges to zero if and only if $\cos(\alpha_N) \to 0$ when $N \to \infty$. On the other hand, if the second term in (17) converges to zero then y_N converges to the eigenvector of the matrix $C(N_1, N-1)$ corresponding to $\nu_{\min}(N_1, N-1)$. If $v_n \to 0$, then the matrix $C(N_1, N)$ changes slightly with increasing N. Thus, the eigenvectors of the matrix change slightly too, and y_N converges to a certain vector y^*. Therefore, the right side of (17) converges to zero if the cosine of the angle between A_N and y^* converges to zero. However, because the first A_N component can take any value, this cosine can be sufficiently large with non-zero probability.

Thus, condition (15) does not hold true for AR process (1), and this implies (9).

An important property of any sequential estimation procedure is the mean number of the observation used for the parameter estimation. Theoretical investigation of this property for weighted methods is a serious problem, so numerical simulation was conducted. A typical example is presented below. We considered the AR(2) process specified by the equation

$$x_{k+1} = \lambda_1 x_k + \lambda_2 x_{k-1} + \xi_{k+1}.$$

The noise variance B^2 in (1) was supposed to be known, so estimator (8) with weights (4–5) was used. The table below presents the results of simulation of the proposed estimation procedure. Here H is a parameter of the procedure, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are estimators of the corresponding parameters, T is a mean length of the estimation interval, d^2 and D^2 are a theoretical and a sample standard deviation of the estimator from the parameter, respectively.

$\lambda_1 = 0,5,\ \lambda_2 = 0,1$

$\hat{\lambda}_1$	$\hat{\lambda}_2$	H	T	d^2	D^2
0,4979	0,1083	50	169	0,0204	0,0198
0,4968	0,0936	100	340	0,0101	0,0093
0,5050	0,1030	150	516	0,0067	0,0051
0,4963	0,1049	200	658	0,0050	0,0048
0,4997	0,1005	250	826	0,0040	0,0040

One can see that the mean length of the estimation interval increases linearly with H. Besides, the sample standard deviation does not exceed the theoretical one. At the same time, the difference between them is not significant, so the bound (10) is rather tight.

2.2 Change point detection procedure

Consider now the change point detection problem for process (1). We construct a series of sequential estimation plans

$$(\tau_i, \lambda_i^*) = (\tau_i(H), \lambda_i^*(H)),\ i \geq 1,$$

where $\{\tau_i\}$, $i \geq 0$ is the increasing sequence of the stopping instances ($\tau_0 = -1$), and λ_i^* is the guaranteed parameter estimator on the interval $[\tau_{i-1}+1, \tau_i]$. The following condition holds true for the estimator

$$E\,||\lambda_i^*(H) - \lambda||^2 \leq \frac{P(H)}{H^2}. \tag{18}$$

Then we choose an integer $l > 1$. We associate the statistic J_i with the $i-$th interval $[\tau_{i-1}+1, \tau_i]$ for all $i > l$

$$J_i = \left(\lambda_i^* - \lambda_{i-l}^*\right)^T \left(\lambda_i^* - \lambda_{i-l}^*\right). \tag{19}$$

This statistic is the squared deviation of the estimators with numbers i and $i-l$.

Denote the deviation of the estimator λ_i^* from the true value of the parameter λ as ζ_i. Let the parameter value remains unchanged until the instant τ_i, i.e., $\theta > \tau_i$. In this case, $\lambda_i = \mu_0 + \zeta_i$, $\lambda_{i-l} = \mu_0 + \zeta_{i-l}$, and statistic (19) can be written in the form

$$J_i = ||\zeta_i - \zeta_{i-l}||^2.$$

Let the change of the parameters takes place on the interval $[\tau_{i-l}, \tau_{i-1}]$ i.e. $\tau_{i-l} < \theta \leq \tau_{i-1}$. In this case, $\lambda_i = \mu_1 + \zeta_i$, $\lambda_{i-l} = \mu_0 + \zeta_{i-l}$, and statistic (19) is

$$J_i = ||\mu_1 - \mu_0 + \zeta_i - \zeta_{i-l}||^2.$$

Thus, the change of the expectation of the statistic J_i allows us to construct the following change point detection algorithm. The J_i values are compared with a certain threshold δ. The change point is considered to be detected, when the value of the statistic exceeds δ.

The probabilities of a false alarm and a delay in the change point detection in any observation cycle are important characteristics of any change point detection procedure. Due to the application of the guaranteed parameter estimators in the statistics, we can bound these probabilities from above.

Theorem 2.2. Let $0 < \delta < \Delta$. Then the probability of false alarm P_0 and the probability of delay P_1 in any observation cycle $[\tau_{i-1}+1, \tau_i]$ are bounded from above

$$P_0 \le 4\frac{P(H)}{\delta H^2}, \quad P_1 \le 4\frac{P(H)}{(\sqrt{\Delta}-\sqrt{\delta})^2 H^2}. \tag{20}$$

Proof. First, we consider the false alarm probability, i.e., the probability that the statistic J_i exceeds the threshold before the change point. Using the norm properties and the Chebyshev inequality, we have

$$\begin{aligned} P_0 = \mathcal{P}\{\, J_i > \delta |\, \tau_i < \theta\} &= \mathcal{P}\left\{ ||\zeta_i - \zeta_{i-l}||^2 > \delta \right\} \\ &\le \frac{2E\left(||\zeta_i||^2 + ||\zeta_{i-l}||^2\right)}{\delta}. \end{aligned}$$

This and (18) imply the first inequality from (20).

Then we consider delay probability, i.e., the probability that the statistic J_i does not exceed the threshold after the change point

$$\begin{aligned} P_1 &= \mathcal{P}\{\, J_i < \delta |\, \tau_{i-l} < \theta < \tau_{i-1}\} \\ &= \mathcal{P}\left\{ ||\mu_1 - \mu_0 + \zeta_i - \zeta_{i-l}||^2 < \delta \right\} \\ &= \mathcal{P}\left\{ ||\mu_1 - \mu_0 + \zeta_i - \zeta_{i-l}|| < \sqrt{\delta} \right\}. \end{aligned}$$

Taking into account that $||\mu_0 - \mu_1||^2 > \Delta$ and using the norm properties and the Chebyshev inequality, one has

$$\begin{aligned} P_1 &\le \mathcal{P}\left\{ ||\mu_1 - \mu_0|| - ||\zeta_i - \zeta_{i-l}|| < \sqrt{\delta} \right\} \\ &\le \mathcal{P}\left\{ \sqrt{\Delta} - ||\zeta_i - \zeta_{i-l}|| < \sqrt{\delta} \right\} \\ &= \mathcal{P}\left\{ ||\zeta_i - \zeta_{i-l}|| > \sqrt{\Delta} - \sqrt{\delta} \right\} \le \frac{2E\left(||\zeta_i||^2 + ||\zeta_{i-l}||^2\right)}{(\sqrt{\Delta}-\sqrt{\delta})^2}. \end{aligned}$$

This and (18) imply the second inequality from (20).

This theorem implies the following constraint for the parameters of the procedure

$$\frac{P(H)}{H^2} \le \frac{\min\left\{ \delta, (\sqrt{\Delta}-\sqrt{\delta})^2 \right\}}{4}.$$

The minimum value of H is obtained when $\delta = \Delta/4$. In this case the upper bounds of the probabilities of false alarm and delay are equal.

2.3 Asymptotic properties of the statistics

The following theorem gives us asymptotic properties of the proposed procedure for $H \to \infty$ for a stable AR process. To simplify the investigation of the properties of the procedure we assume additionally that the weights v_k satisfy the following condition:

$$v_k^2 A_k A_k^T \le \gamma(H)H, \tag{21}$$

where $\gamma(H)H \to \infty$ as $H \to \infty$. It is rather weak constraint, because for a stable process the probability of this event tends to unity with increasing H.

Theorem 2.3. Let AR process (1) be stable. If the conditions of Theorem 1 and (21) hold true, and $E\xi_k^4 < \infty$, then for a sufficiently large H

$$\mathcal{P}\left\{ ||\lambda^* - \lambda||^2 > x \right\} \le 1 - \left(2\Phi\left(\sqrt{\frac{xH^2}{P(H)}} \right) - 1 \right)^m, \tag{22}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Proof. We consider estimator (8). Using the norm properties, one has

$$\begin{aligned} ||\lambda^* - \lambda||^2 &= \left\| BC^{-1}(N_1,\tau) \sum_{k=N_1}^{\tau} v_k A_k^T \xi_{k+1} \right\|^2 \\ &\le \left\| BC^{-1}(N_1,\tau) \right\|^2 \left\| \sum_{k=N_1}^{\tau} v_k A_k^T \xi_{k+1} \right\|^2. \end{aligned}$$

Using (6) and (7), one has

$$||\lambda^* - \lambda||^2 \le \frac{B^2}{H^2} \left\| \sum_{k=N_1}^{\tau} v_k A_k^T \xi_{k+1} \right\|^2. \tag{23}$$

Denote $Z = [z_1, \dots, z_m]$ and consider a linear combination of the components of the vector from the last equation

$$X_\tau = \frac{B}{\sqrt{P(H)}} \sum_{k=N_1}^{\tau} v_k Z A_k^T \xi_{k+1}.$$

Further we find the limit distribution of X_τ along the lines of the proof of the martingale central limit theorem (see Shiryaev A. N. (1980)). We find the characteristic function of X_τ. Denote

$$\begin{aligned} \varepsilon_k = \varepsilon_k(H) &= \frac{B}{\sqrt{P(H)}} v_k Z A_k^T \xi_{k+1} \chi_{[\tau \ge k]}, \\ X_n &= \sum_{k=N_1}^{n} \varepsilon_k. \end{aligned} \tag{24}$$

Then choose the sequence $\gamma = \gamma(H) \downarrow 0$, $\gamma(H)H \uparrow \infty$ as $H \to \infty$. We can consider additional constraint (21) as an inequality for v_k, i.e., if (21) does not hold true for v_k, specified by (4–7), then we choose it in the form

$$v_k = \sqrt{\gamma H/(A_k A_k^T)}.$$

Note that inequality (7) holds true under this condition.

It is evident that under the assumptions of Theorem 1 as $n \to \infty$

$$|X_\tau - X_n| \to^{\mathcal{P}} 0.$$

Thus, in order to find the characteristic function of X_τ, one needs to find the limit of the characteristic function of X_n. Denote

$$\mathcal{E}^n(\eta) = \prod_{k=N_1}^{n} E\left(e^{i\eta\varepsilon_k} \middle| \mathcal{F}_k \right),$$

Lemma 2.1. (Shiryaev A. N. (1980)) If (for given η) $|\mathcal{E}^n(\eta)| \ge c(\eta) > 0$, $n > 1$, then convergence in probability $\mathcal{E}^n(\eta) \to E\left(e^{i\eta X}\right)$ is sufficient for convergence $E\left(e^{i\eta X_n}\right) \to E\left(e^{i\eta X}\right)$.

Check the lemma conditions

$$\begin{aligned} |\mathcal{E}^n(\eta)| &= \prod_{k=N_1}^{n} \left| E\left[e^{i\eta\varepsilon_k} \middle| \mathcal{F}_k \right] \right| \\ &= \prod_{k=N_1}^{n} \left| 1 + E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k \middle| \mathcal{F}_k \right] \right|. \end{aligned}$$

Using the inequality $\left|e^{i\eta x} - 1 - i\eta x\right| \le (\eta x)^2/2$, we have

$$|\mathcal{E}^n(\eta)| \ge \prod_{k=N_1}^{n} \left(1 - E\left[\left|e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\right|\middle| \mathcal{F}_k\right]\right)$$
$$\ge \prod_{k=N_1}^{n} \left(1 - \frac{1}{2}E\left[(\eta\varepsilon_k)^2\middle| \mathcal{F}_k\right]\right)$$
$$= \prod_{k=N_1}^{n} \left(1 - \frac{(\eta B v_k Z A_k^T)^2 \chi_{[\tau\ge k]}}{2P(H)} E\xi_{k+1}^2\right)$$
$$= \exp\left\{\sum_{k=N_1}^{n} \ln\left(1 - \frac{(\eta B v_k Z A_k^T)^2 \chi_{[\tau\ge k]}}{2P(H)}\right)\right\}.$$

Constraint (21) implies $(v_k Z A_k^T)^2/P(H) \to 0$ as $H \to \infty$. Using the inequality $\ln(1-x) \ge -2x$, where $x \in (0, 1/2]$, for any $H \ge H_0(\eta)$, one has

$$|\mathcal{E}^n(\eta)| \ge \exp\left\{-\sum_{k=N_1}^{\min(n,\tau)} \frac{(\eta B v_k Z A_k^T)^2}{P(H)}\right\}$$
$$\ge \exp\left\{-\frac{\eta^2 B^2}{P(H)} \sum_{k=N_1}^{\tau} \left(v_k Z A_k^T\right)^2\right\}.$$

Taking into account (7), we have

$$|\mathcal{E}^n(\eta)| \ge \exp\left\{-\frac{\eta^2 B^2}{P(H)} \frac{P(H)}{B^2}\right\} = e^{-\eta^2}.$$

The lemma conditions hold true.

Further we investigate an asymptotic behavior of $\mathcal{E}^n(\eta)$. Write this value in the form

$$\mathcal{E}^n(\eta) = \exp\left\{\sum_{k=N_1}^{n} E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]\right\} \times$$
$$\times \exp\left\{-\sum_{k=N_1}^{n} E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]\right\} \times \tag{25}$$
$$\times \prod_{k=N_1}^{n} \left(1 + E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]\right).$$

Then we show that the product of the last two factors tends to 1. Denote $\alpha_k = E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]$. Using the inequality $|e^x - 1| \le e^{|x|}|x|$, we have

$$\left|\prod_{k=N_1}^{n} (1+\alpha_k)\, e^{-\alpha_k} - 1\right|$$
$$= \left|\exp\left\{\ln \prod_{k=N_1}^{\tau} (1+\alpha_k)\, e^{-\alpha_k}\right\} - 1\right|$$
$$\le \exp\left\{\left|\ln \prod_{k=N_1}^{n} (1+\alpha_k)\, e^{-\alpha_k}\right|\right\} \left|\ln \prod_{k=N_1}^{n} (1+\alpha_k)\, e^{-\alpha_k}\right|.$$

Using the inequalities $|\ln(1+x) - x| \le 2|x|^2$ for $|x| < 1/2$ and $\left|e^{i\eta x} - 1 - i\eta x\right| \le (\eta x)^2/2$, as $H > H_0(\eta)$, we have

$$\left|\ln \prod_{k=N_1}^{n} (1+\alpha_k)\, e^{-\alpha_k}\right| \le \sum_{k=N_1}^{n} |\ln(1+\alpha_k) - \alpha_k|$$
$$\le 2\sum_{k=N_1}^{n} |\alpha_k|^2 = 2\sum_{k=N_1}^{n} \left(E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]\right)^2$$
$$\le \frac{\eta^4 B^4}{P(H)^2} \sum_{k=N_1}^{\tau} (v_k Z A_k^T)^4.$$

Using (21) and (7) one has

$$\left|\ln \prod_{k=N_1}^{n} (1+\alpha_k)\, e^{-\alpha_k}\right|$$
$$\le \frac{\eta^4 B^4 \gamma H ||Z||^2}{P(H)^2} \sum_{k=N_1}^{\tau} (v_k Z A_k^T)^2 \le \eta^4 ||Z||^4 \gamma \to 0.$$

Thus the product of the last two multipliers in (25) tends to 1 in probability as $n \to \infty$, $H \to \infty$.

Consider the first multiplier

$$\exp\left\{\sum_{k=N_1}^{n} E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k\middle| \mathcal{F}_k\right]\right\} =$$
$$= \exp\left\{-\frac{1}{2}\sum_{k=N_1}^{n} E\left[(\eta\varepsilon_k)^2\middle| \mathcal{F}_k\right]\right\} \times \tag{26}$$
$$\times \exp\left\{\sum_{k=N_1}^{n} E\left[e^{i\eta\varepsilon_k} - 1 - i\eta\varepsilon_k + \frac{(\eta\varepsilon_k)^2}{2}\middle| \mathcal{F}_k\right]\right\}.$$

Using the inequality $\left|e^{i\eta x} - 1 - i\eta x + (\eta x)^2/2\right| \le |\eta x|^3/6$ and (7), one has

$$\left|\sum_{k=N_1}^{n} E\left[e^{i\eta\varepsilon_t} - 1 - i\eta\varepsilon_k + \frac{(\eta\varepsilon_k)^2}{2}\middle| \mathcal{F}_k\right]\right|$$
$$\le \frac{1}{6P(H)^{3/2}} \sum_{k=N_1}^{n} E\left[\left|\eta v_k Z A_k^T \xi_{k+1}\right|^3 \chi_{[\tau\ge k]}\middle| \mathcal{F}_k\right]$$
$$= \frac{B^3|\eta|^3 \mathcal{M}|\xi_{k+1}|^3}{6P(H)^{3/2}} \sum_{k=N_1}^{\tau} \left|v_k Z A_k^T\right|^3 \chi_{[\tau\ge k]}.$$

According to (21) and (7), the last expression tends to 0. So the second multiplier in (26) tends to 1 as $H \to \infty$. Consider the first multiplier

$$\exp\left\{-\frac{1}{2}\sum_{k=N_1}^{n} E\left[(\eta\varepsilon_k)^2\middle| \mathcal{F}_k\right]\right\}$$
$$= \exp\left\{-\frac{B^2\eta^2}{2P(H)} \sum_{k=N_1}^{\min(n,\tau)} (v_k Z A_k^T)^2\right\} = \exp\left\{-\frac{\eta^2}{2}\langle X_n\rangle\right\}.$$

Note that according to (7), $\langle X_n\rangle$ is a bounded submartingale. Thus, the limit $\langle X_\infty\rangle = \lim_{n\to\infty}\langle X_n\rangle$ exists almost surely, and $\langle X_\infty\rangle \le ||Z||^2$. On the other hand, $\langle X_n\rangle \to \langle X_\tau\rangle$ as $n \to \infty$. So the distribution X_τ is asymptotically normal. Thus, the random vector

$$Y = \frac{B}{\sqrt{P(H)}} \sum_{k=N_1}^{\tau} v_k A_k^T \xi_{k+1}$$

is asymptotically normal with the parameters $(0, \Sigma)$.

Estimate the probability (22). Denote $y = [y_1, \dots, y_m]$ and $C = xH^2/P(H)$. Using (23) one has

$$\mathcal{P}\left\{||\lambda^* - \lambda||^2 > x\right\} \le \mathcal{P}\left\{\frac{P(H)}{H^2}||Y||^2 > x\right\}$$
$$\le \frac{1}{\sqrt{(2\pi)^m|\Sigma|}} \int\limits_{yy^T > C} \exp\left\{-\frac{1}{2}y\Sigma^{-1}y^T\right\} dy$$
$$= \frac{1}{(2\pi)^{m/2}} \int\limits_{t\Sigma t^T > C} \exp\left\{-\frac{1}{2}tt^T\right\} dt$$
$$\le \frac{1}{(2\pi)^{m/2}} \int\limits_{\mathrm{tr}(\Sigma)\max\limits_{1\le i\le m} t_i^2 > C} \exp\left\{-\frac{1}{2}tt^T\right\} dt.$$

Using (7), it is easy to show that

$$\mathrm{tr}(\Sigma) = \frac{B^2}{P(H)} \sum_{k=N_1}^{\tau} v_k^2 A_k A_k^T \le 1.$$

Thus, one has

$$\mathcal{P}\left\{||\lambda^* - \lambda||^2 > x\right\}$$
$$\le 1 - \frac{1}{(2\pi)^{m/2}} \int\limits_{\max\limits_{1\le i\le m} t_i^2 \le C} \exp\left\{-\frac{1}{2}tt^T\right\} dt$$
$$= 1 - \left(2\Phi\left(\sqrt{C}\right) - 1\right)^m,$$

and (22) holds true.

Along the lines of the proof, one can show that

$$\mathcal{P}\left\{||\zeta_i - \zeta_{i-l}||^2 > x\right\} \le 1 - \left(2\Phi\left(\sqrt{\frac{xH^2}{2P(H)}}\right) - 1\right)^m.$$

Thus, the probabilities of false alarm and delay can be estimated as

$$P_0 \le 1 - \left(2\Phi\left(\frac{\delta H}{\sqrt{2(H+m-1)}}\right) - 1\right)^m;$$
$$P_1 \le 1 - \left(2\Phi\left(\frac{\left(\sqrt{\Delta} - \sqrt{\delta}\right)H}{\sqrt{2(H+m-1)}}\right) - 1\right)^m. \tag{27}$$

We can use these estimators instead of (20) for a sufficiently large H.

3. NUMERICAL SIMULATION

We considered the AR(2) process. The change point $\theta = 20000$. For every H, 1000 replications of the experiment were performed. The noise variance B^2 in (1) was supposed to be known, so estimator (8) with weights (4–5) was used. The table below presents the results of simulation of the proposed estimation procedure. Here μ_0 and μ_1 are values of the vector $[\lambda_1, \lambda_2]$ before and after change point, respectively; H and δ are parameters of the procedure; T_1 is a mean delay in the change point detection; T_0 is a mean interval between false alarms; $\hat{P}_0$ and $\hat{P}_1$ are the sample probabilities of a false alarm and a delay, respectively; P is their theoretical upper bound (27) (we choose $\delta = \Delta/4$, so the both probabilities are bounded from above by the same value).

$\mu_0 = [0,5;0,1]$, $\mu_1 = [0,1;0,3]$, $\delta = 0,05$

H	T_1	T_0	$\ln(T_0)$	$\hat{P}_1$	$\hat{P}_0$	P
100	284	4467	8,41	0,028	0,074	0,217
120	348	10152	9,22	0,015	0,039	0,162
140	405	15384	9,64	0,000	0,030	0,121
160	469	45454	10,72	0,000	0,011	0,090
180	541	68965	11,14	0,000	0,008	0,067
200	563	117647	11,68	0,000	0,005	0,051

The mean delay and the logarithm of the mean time between the false alarms increase linearly with H. The procedures having such property are referred as optimal (see Lorden G. (1971), Brodskiy B.B. et al. (1992)). The sample error probabilities are less then their theoretical upper bounds, so the asymptotic estimation of these probabilities can be used for the choice of the parameters.

Results of numerical simulation prove that the suggested procedure can be used for a change point detection of recurrent processes.

4. CONCLUSIONS

The change point detection procedure for the autoregressive process with unknown parameters before and after the change point has been constructed. The procedure is based on the weighted least square method. The guaranteed sequential estimators of unknown parameters are used. The choice of weights and stopping rule guarantees the prescribed accuracy of the estimation and, hence, the prescribed probabilities of delay and false alarm in every observation interval. Results of numerical simulation prove the efficiency of the suggested procedure.

REFERENCES

Brodskiy B.B., Darhovskiy B.S. (1992). Nonparametric methods in change-point problems. *Kluwer Academic, Boston.*, 1992.

Basseville M., Willsky A.S., Benveniste A. et al.(1985. Detection of Abrupt Changes in Signals and Dynamical Systems. *Lecture Notes in Control and Information Sciences, LNCIS-77, Springer-Verlag, Berlin, DE.*, 1985.

Lai T.Z.(1995). Sequential change-point detection in quality control and dynamical systems. *Journ. Res. Statist. Soc. B.*, 1995. V. 57. No. 4. P. 613–658.

Vorobejchikov S. E., Ponomarjova J. S. (2002). On change point detection of stochastic processes. *Optoelectronics, Instrumentation and Data Processing*, 2002. V. 38. No. 5. P. 68–78.

Vorobejchikov S. E., Meder N. A. (2002). On guaranteed estimation of the parameter of random processes by the weighted least squares method. *Preprints of the 15th Triennial World Congress of the International Federation of Automatic Control, Barcelona*, 21-26 July 2002, No. 1200.

Shiryaev A. N. (1980). Probability. *Moscow: Nauka, Chief Editorial Staff of Physical-Mathematical Literature*, 1980.

Lorden G. (1971). Procedures for reacting to a change in distribution. *Annals. Math. Statist.*, 1971. V. 42. 1897–1908.