A METHOD OF DETECTING CHANGES IN THE BEHAVIOUR OF LOCALLY STATIONARY SEQUENCES

JIŘÍ MICHÁLEK

A method for the detection of abrupt changes in the course of a locally stationary sequence is proposed. The method is based on a suitable approximation of an observed sequence by autoregressive models that are compared by means of a similarity measure derived from the asymptotic I-divergence rate. The method is illustrated by several numerical results.

1. INTRODUCTION

The detection of changes in the behaviour of random sequences and processes is very intensively investigated in last years because of a direct use of suggested methods in practice. From the point of view of mathematical statistics the problem of detecting changes is a task belonging to the testing hypotheses region. We test a single hypothesis presenting a homogeneous course of an observed sequence without any change versus a composed alternative hypothesis not eliminating, in general a change at an arbitrary time instant during observation. The first kind error is the false alarm error and the second kind error is expressed via the average length between the occurring of a change and its recording. One can find in literature several survey papers dealing with the problem. One can recommend to the reader the papers due to Kligiene and Telksnys [5], Basseville and Benveniste [3], Willsky [11], Nikoforov [8]. The main difficulties occurring by solving this problem follow from the nonexistence of an optimal statistical test as shown by Deshayes and Picard in Basseville and Benveniste [3]. We can speak at most about the asymptotic optimality of proposed tests. From these reasons we can meet in literature many methods for detecting changes, which can be used under different circumstances. Most of these methods are based on the construction of a statistic, usually of a cumulative sum type, which under no change has vanishing mean value and after the occurring of a change its mean value changes as well. The usual approach in suggesting methods for detecting changes is the following. At first, the perfect knowledge of a mathematical model describing the situation before a change is supposed. Secondly, one supposes sometimes the perfect knowledge of the situation after a change, too.

Under these assumptions a method for detecting is derived. In practice, unknown parameters in the considered models are substituted by suitable estimates.

Some random nonstationary sequences can be approximated by locally stationary sequences. Such a random sequence can be split into mutually stochastically independent parts, which can be considered to be weakly stationary. The abrupt changes between adjoing stationary parts are to be detected. A change can occur both in mean values and in spectral densities. Although in practice one can meet a locally stationary sequence with a transitional period between two stationary segments our model can serve as a first approximation. Such a situation can happen during the monitoring of a signal coming from a measuring sensor bearing information about the behaviour of a technological process. A transitional nonstationary period can be caused by a transition from one regime to another one. We shall assume our observed sequence can be described by a locally stationary time series with unknown mean values, unknown spectral densities and unknown time instants of changes. Within every stationary segment the observed sequence is weakly stationary and regular, first we shall assume Gaussian observations, later this assumption will be dropped. The basic idea is to approximate the observed sequence by a suitable autoregressive models then to watch a similarity among these approximative models. The measure of similarity is given by the asymptotic I-divergence rate between two Gaussian autoregressive models. This measure of similarity serves as a testing statistic for detecting changes in the behaviour of a locally stationary sequence. Under the null hypothesis this statistic should be theoretically vanishing under a change it must be positive.

2. THEORETICAL BACKGROUND

Let P, Q be two Gaussian measures defined on the measurable space (R_∞, K_∞), where $R_\infty = \times_1^\infty R_1$, K_∞ is the Kolmogorov σ-algebra, by mean values m_P, m_Q and by spectral density functions $\varphi_P(\cdot)$, $\varphi_Q(\cdot)$. Let P_n, Q_n be the restriction of P, Q to the subspace (R_n, K_n). The problem of detecting changes in the behaviour of a random sequences is that of hypotheses testing.

The quality of a test is given by the first kind and second kind errors. Let us assume, for this moment only, we exactly know a change time instant and the mathematical models describing an observed sequence as well. Then, there exists the optimal test of the Neyman–Pearson type via the Radon–Nikodym derivative of the corresponding probability measure, let us say

$$\frac{\mathrm{d}P_n(\mathbf{x})}{\mathrm{d}Q_n(\mathbf{x})}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are observations at disposal given. The hypothesis "no change" is rejected if

$$\frac{\mathrm{d}P_n(\mathbf{x})}{\mathrm{d}Q_n(\mathbf{x})} > d_{\alpha,n}$$

where $d_{\alpha,n}$ is derived from the significance level equal to α. The test function

$$\varphi_n(\mathbf{x}) = 1 \iff \frac{\mathrm{d}P_n(\mathbf{x})}{\mathrm{d}Q_n(\mathbf{x})} > d_{\alpha,n}$$

$$\varphi_n(\mathbf{x}) = 0 \qquad \text{otherwise}$$

determines the second kind error $\beta_n(\alpha) = E_P\{1 - \varphi_n(\mathbf{x})\}$. In the i.i.d. case the behaviour of $\beta_n(\alpha)$ if $n \to \infty$ is given by the famous Stein theorem (cf. Bahadur [1]), generalizations to the case where an asymptotic I-divergence rate exists

$$\lim_{n \to \infty} \frac{1}{c(n)} I(P_n|Q_n) = \overline{I}(P|Q) = \lim_{n \to \infty} \frac{1}{c(n)} \ln \beta_n(\alpha)$$

were obtained by Vajda [9, 10]. In the stationary case we have $c(n) = n$. This fact says, roughly speaking, the larger $\overline{I}(P|Q)$ the better chance for distinguishing P from Q. The basic idea following from these conclusions: can we use a suitable estimate of $\overline{I}(P|Q)$ as a statistic for detecting changes in behaviour of a locally stationary sequence?

First, we need an expression for $\overline{I}(P|Q)$ in the Gaussian stationary case. The answer can be found, e. g. in the monograph Vajda [9].

Lemma 1. Let $\{P_n, Q_n\}_{n=1}^{\infty}$ be Gaussian stationary measures generated by random regular stationary sequences $\{x_j\}_{j=1}^{n}$, $\{y_j\}_{j=1}^{n}$, respectively. Then under the assumption of equivalence $P_n \sim Q_n$ for each $n \in \mathcal{N}$ there exists

$$\lim_{n \to \infty} \frac{1}{n} E_P \left\{ \ln \frac{\mathrm{d}P_n}{\mathrm{d}Q_n} \right\} = \overline{I}(P|Q) =$$

$$= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\varphi_P}{\varphi_Q} - \ln \frac{\varphi_P}{\varphi_Q} - 1 \right) (\lambda) \, \mathrm{d}\lambda + \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_Q^2}$$

where φ_P, φ_Q are the corresponding spectral densities, m_P, m_Q are mean values and σ_Q^2 is dispersion of the best one step ahead prediction error with respect to $\{Q_n\}$.

Proof. See [9], pp. 239–240. □

In case that the Gaussian measure Q is generated by an stationary autoregressive random sequence and the Gaussian measure P is given by a regular stationary random sequence the corresponding asymptotic I-divergence rate can be expressed more in detail. Then the explicit form is written in the following

Lemma 2. Let Q be a probability measure generated by an autoregressive stationary Gaussian sequence, let P be a probability measure generated by a linearly regular Gaussian stationary random sequence with a covariance function R and spectral density function φ_P. Then

$$\overline{I}(P|Q) = \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_Q^2} + \frac{1}{2} \left(\frac{\mathbf{a}^{\mathrm{T}} \mathbb{R} \, \mathbf{a}}{\sigma_Q^2} - \ln \frac{\sigma_P^2}{\sigma_Q^2} - 1 \right)$$

where $\mathbf{a}^{\mathrm{T}} = (1, a_1, \ldots, a_p)$ are the autoregressive coefficients due to Q, σ_Q^2 and σ_P^2 are mean square one step ahead prediction errors, $\mathbb{R}$ is the $(p+1) \times (p+1)$ Toeplitz matrix derived from the covariance function $R(\cdot)$.

P r o o f.   First, we shall calculate $\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} \, \mathrm{d}\lambda$. As Q is an autoregressive model then the corresponding spectral density $\varphi_Q(\cdot)$ is given by

$$\varphi_Q(\lambda) = \frac{1}{2\pi} \frac{\sigma_Q^2}{\left| \sum_{j=0}^{p} a_j \, e^{ij\lambda} \right|^2}$$

with $a_0 = 1$. Then

$$
\begin{aligned}
\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} \, \mathrm{d}\lambda &= \frac{2\pi}{4\pi} \int_{-\pi}^{\pi} \frac{\left| \sum_{j=0}^{p} a_j \, e^{ij\lambda} \right|^2}{\sigma_Q^2} \varphi_P(\lambda) \, \mathrm{d}\lambda \\
&= \frac{1}{2\sigma_Q^2} \int_{-\pi}^{\pi} \sum_{j=0}^{p} \sum_{k=0}^{p} a_j \, a_k \, e^{i(j-k)} \varphi_P(\lambda) \, \mathrm{d}\lambda = \\
&= \frac{1}{2\sigma_Q^2} \sum_{j=0}^{p} \sum_{k=0}^{p} a_j \, a_k \int_{-\pi}^{\pi} e^{i(j-k)} \varphi_P(\lambda) \, \mathrm{d}\lambda = \\
&= \frac{1}{2\sigma_Q^2} \sum_{j=0}^{p} \sum_{k=0}^{p} a_j \, a_k \, R(j-k) = \\
&= \frac{1}{2} \frac{\mathbf{a}^{\mathrm{T}} \mathbb{R} \, \mathbf{a}}{\sigma_Q^2}.
\end{aligned}
$$

The second step is to evaluate $\frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} \, \mathrm{d}\lambda$. Since we have regular spectral densities then

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} \, \mathrm{d}\lambda = \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \varphi_P(\lambda) \, \mathrm{d}\lambda - \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \varphi_Q(\lambda) \, \mathrm{d}\lambda.$$

At this moment we can use the formula for one step ahead prediction error in the quadratic mean sense, e. g. see [4] for detail

$$\sigma^2 = 2\pi \exp \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \varphi(\lambda) \, \mathrm{d}\lambda \right\}.$$

Hence, after elementary computations we find out that

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \ln \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} \, \mathrm{d}\lambda = \frac{1}{2} \ln \frac{\sigma_P^2}{\sigma_Q^2}.$$

Now, we are ready to state that

$$\overline{I}(P|Q) = \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_Q^2} + \frac{1}{2} \left(\frac{\mathbf{a}^{\mathrm{T}} \mathbb{R} \, \mathbf{a}}{\sigma_Q^2} - \ln \frac{\sigma_P^2}{\sigma_Q^2} - 1 \right).$$

$\square$

$\overline{I}(P|Q)$ possesses similar properties as statistical convex distances, e. g. the Kullback–Leibler information, Rényi's distances:

1. $\overline{I}(P|Q) \geq 0$
2. $\overline{I}(P|Q) = 0 \Longleftrightarrow P = Q$
3. $\overline{I}(P|Q)$ is not a metric in general because of non-valid triangular inequality.

$\overline{I}(P|Q)$ can be utilized for a construction of minimal distance estimates. This possibility is used in the proposed method for the construction of an autoregressive model of a given order that is in a certain sense the most similar to the observed sequence. If the measure P is derived from observations and Q belongs to the class of Gaussian autoregressive models of a given order $p \in \mathcal{N}$ then we wish to find an autoregressive model Q^* minimizing the distance $\overline{I}(P|Q)$. In case we choose the measure P derived from observations $\{x_1, x_2, \ldots, x_n\}$ via the sample spectral density function

$$\varphi_{\mathbf{x}}(\lambda) = \frac{1}{2\pi} \sum_{j=-p}^{p} e^{-i\lambda j} \hat{R}(j),$$

with

$$\hat{R}(j) = \frac{1}{n} \sum_{k=1}^{n-j} (x_k - \overline{x})(x_{k+j} - \overline{x}), \qquad j = 0, 1, \ldots, p, \ n \gg p$$

$$\hat{R}(-j) = \hat{R}(j), \qquad \overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

then as shown in Michálek [7] this minimization task is unambiguously solved. The solution is given by the Yule–Walker estimates of autoregressive coefficients $a_1, a_2, \ldots, a_p$. As for the order p this can be approximated by the Akaike criterion, e. g. A very progressive method suitable for computer implementation evaluating the Yule–Walker estimates is the so called Durbin–Levinson algorithm.

In case when the observed sequence $\{x_n\}$ is not Gaussian the approximative autoregressive model minimizing $\overline{I}(P|Q)$ can be considered as the closest projection of P into the class of Gaussian autoregressive models having the order p in the sense of the results presented in Künsch [6].

In the case of two Gaussian stationary autoregressive sequences the corresponding asymptotic I-divergence rate $\overline{I}(P|Q)$ can be expressed as

$$\overline{I}(P|Q) = \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_Q^2} + \frac{1}{2} \frac{\mathbf{a}^{\mathrm{T}} \mathbb{R}_P \mathbf{a} - \mathbf{b}^{\mathrm{T}} \mathbb{R}_P \mathbf{b}}{\sigma_Q^2} + \frac{1}{2} \left(\frac{\sigma_P^2}{\sigma_Q^2} - \ln \frac{\sigma_P^2}{\sigma_Q^2} - 1 \right) \quad .$$

where $Q \equiv (1, a_1, a_2, \ldots, a_p, \sigma_Q, m_Q)$, $P \equiv (1, b_1, \ldots, b_p, \sigma_P, m_P)$. This fact immediately follows from Lemma 2 and from the useful relation

$$\sigma_P^2 = \mathbf{b}^{\mathrm{T}} \mathbb{R}_P \mathbf{b},$$

which is a direct consequence of an autoregressive model. Sometimes one can use the symmetric version of $\overline{I}(P|Q)$ denoted as

$$\overline{J}(P, Q) = \frac{1}{2} \left(\overline{I}(P|Q) + \overline{I}(Q|P) \right).$$

For autoregressive models $\overline{J}(P, Q)$ has then a very simple form, namely

$$2 \cdot \overline{J}(P, Q) = \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_P^2} + \frac{1}{2} \frac{(m_P - m_Q)^2}{\sigma_Q^2} + \frac{1}{2} \frac{\mathbf{a}^T \mathbb{R}_P \mathbf{a}}{\sigma_Q^2} + \frac{1}{2} \frac{\mathbf{b}^T \mathbb{R}_Q \mathbf{b}}{\sigma_P^2} - 1.$$

Using the corresponding spectral density functions φ_P, φ_Q $\overline{J}(P, Q)$ can be expressed in a very clear form as

$$\overline{J}(P, Q) = \frac{1}{8\pi} \int_{-\pi}^{\pi} \frac{(\varphi_P - \varphi_Q)^2}{\varphi_P \varphi_Q} (\lambda) \, \mathrm{d}\lambda$$

(if for simplicity $m_P = m_Q = 0$). This last expression for $\overline{J}(P, Q)$ follows directly from **Lemma 1** because

$$\overline{I}(P|Q) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} - \ln \frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} - 1 \right) \mathrm{d}\lambda,$$

$$\overline{I}(Q|P) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\varphi_Q(\lambda)}{\varphi_P(\lambda)} - \ln \frac{\varphi_Q(\lambda)}{\varphi_P(\lambda)} - 1 \right) \mathrm{d}\lambda,$$

hence

$$\frac{1}{2} \left[\overline{I}(P|Q) + \overline{I}(Q|P) \right] =$$

$$= \frac{1}{8\pi} \int_{-\pi}^{\pi} \left(\frac{\varphi_P(\lambda)}{\varphi_Q(\lambda)} + \frac{\varphi_Q(\lambda)}{\varphi_P(\lambda)} - 2 \right) \mathrm{d}\lambda =$$

$$= \frac{1}{8\pi} \int_{-\pi}^{\pi} \frac{(\varphi_P(\lambda) - \varphi_Q(\lambda))^2}{\varphi_Q(\lambda) \varphi_Q(\lambda)} \, \mathrm{d}\lambda.$$

3. DESCRIPTION OF THE METHOD

Let us assume the observations $\{x_j\}_{j=1}^n$ form a locally stationary sequence with unknown mean values and spectral densities. The method for detecting changes is based on the principle of two sliding windows. The first one serves as a reference window identifying sequences before a change the other one is a testing window moving ahead along observations detecting a possible change. The reference window contains a certain number of observations, let us say m, which must be sacrificed for starting up the procedure. The first testing window begins at the position $m + 1$. The testing window can be relatively short (in practice we can consider from 20 up to 100 observations). Let L be the length of a testing window.

Under the hypothesis "no change" the sequence $\{x_j\}_{j=1}^n$ forms a regular weakly stationary sequence as assumed. Within each window the observed sequence will be approximated by a suitable autoregressive model of the order p, where p is chosen in advance. In this way we obtained the estimates $\hat{\mathbf{a}} = (1, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}_Q^2)$ and $\mathbb{R}_Q$ from the reference window, $\hat{\mathbf{b}} = (1, \hat{b}_1, \hat{b}_2, \ldots, \hat{b}_p, \hat{\sigma}_P^2)$ and $\mathbb{R}_P$ from the testing

window. Using these quantities we can estimate $\overline{I}(\hat{P}|\hat{Q})$ expressing the similarity between both the windows. This estimate has the following form

$$\overline{I}(\hat{P}|\hat{Q}) = \frac{1}{2}\frac{(\hat{m}_P - \hat{m}_Q)^2}{\hat{\sigma}_Q^2} + \frac{1}{2}\frac{\hat{\mathbf{a}}^{\mathrm{T}}\hat{\mathbb{R}}_P\hat{\mathbf{a}} - \hat{\mathbf{b}}^{\mathrm{T}}\hat{\mathbb{R}}_P\hat{\mathbf{b}}}{\hat{\sigma}_Q^2} + \frac{1}{2}\left(\frac{\hat{\sigma}_P^2}{\hat{\sigma}_Q^2} - \ln\frac{\hat{\sigma}_P^2}{\hat{\sigma}_Q^2} - 1\right)$$

where $\hat{m}_Q$ is the arithmetic mean in the reference window and $\hat{m}_P$ is the arithmetic mean in the testing window. Under the validity of the hypothesis "no change" the considered statistics $\overline{I}(\hat{P}|\hat{Q})$ forms a sequence with values close to zero. If a change occurs this constant course must be violated since new observations after the change are starting to fill up the testing window. Using the statistic $\overline{I}(\hat{P}|\hat{Q})$ the problem of the detection of changes in spectral characteristics is transformed into the problem of the detection of changes in constant mean value. When no change is detected the testing window is innovated by the latest observation. We see this method cannot distinguish two changes that are too close each other.

Although the proposed method need not assume almost anything about the observed sequence except its local stationarity there exists a weak point, namely the setup of a limit whose crossing means the detection of a change. From this theoretical point of view the calculation of such a limit is a very complicated problem because of a strong dependence of J-divergence rates on moving windows. Further, such a limit would be also strongly dependent on parameters defining the observed sequence before a change and after it. But, the parameters after a change are not kn\, wn to the observer a priori. It means if we had any theoretical bound for detecting changes, from the practical point of view such a limit would be without any applicability as long as we did not know the parameters after the change. Such a situation is, unfortunately very rare in practice. There is in fact the only possibility how to overcome this difficulty. We must have a possibility to recognize the behaviour of J-divergence rates before a change. In other words speaking, we must sacrifice a certain part of observations for determining a reliable upper bound for a J-divergence rate. This approach is systematically used in our simulations described in the last part of the paper. It is evident that the setup of an upper bound for J-divergence rates is very closely connected both with the probability of a false alarm and with the delayed time after detecting a change. From this reason we considered the different length of the testing window to show its influence upon the delayed time.

The method with the fixed reference window and with moving detection window is suitable in such a situation when we can expect one change only. In case we can watch a series of changes the proposed method must be rather modified, namely both the windows must be moving simultaneously, i.e. the reference window must immediately follow the detection window. Their length need not be the same, in general. On the basis of practical results it is more convenient by this modified method to calculate

$$\ln(|\overline{J}(\hat{P}, \hat{Q})| + 1)$$

instead of $\overline{J}(\hat{P}, \hat{Q})$ to give chance for detecting changes with relatively smaller differences in $\overline{J}(\hat{P}, \hat{Q})$.

4. NUMERICAL RESULTS AND CONCLUSION

The proposed method is illustrated in 6 examples of locally stationary sequences. The Models A, B, C, D concern autoregressive sequences before the change and after it, too. The Model E deals with a very simple ARMA(1,1) sequence. The last Model F considers a sequence of several changes and shows the application of the modified method described just above. All the cases A, B, C, D and E are organized identically. All the sequences contain 1 000 observations, 500 before the change and 500 after the change, i. e. the change occurs at the position 501. When the reference window has the length m, and the detecting window L then the change in the figures is depicted at the position $501 - m - L$. For better orientation the change in the figures is represented by a vertical line.

It means the first observation after the change is taken into the detecting window having the length L at the position $501 - m - L$. All the figures concerning the cases A, B, C, D and E present the corresponding J-divergence rate

$$\overline{J}(\hat{P}, \hat{Q}) = \frac{1}{2} \frac{\mathbf{a}^\mathrm{T} \mathbb{R}_{\hat{P}} \, \mathbf{a}}{\sigma_{\hat{Q}}^2} + \frac{1}{2} \frac{\mathbf{b}^\mathrm{T} \mathbb{R}_{\hat{Q}} \, \mathbf{b}}{\sigma_{\hat{P}}^2} - 1$$

between the moving detection window and fixed reference window at the beginning of observations. If $J(\hat{P}, \hat{Q})$ is depicted in the figure at the nth position it means this value presents the J-divergence rate between the detecting window $\{x_{m+n}, \dots, x_{n+L+m}\}$ and the reference window $\{x_1, x_2, \dots, x_m\}$ when the whole experiments contains observations $x_1, x_2, \dots, x_{1000}$. In order to show how the proposed method works also in the case of nongaussian variables all the sequences were generated by white noises from the uniform distribution over the interval $\langle -\frac{1}{2}, \frac{1}{2} \rangle$.

Model A: before the change: $x_{n+1} = 1,67\, x_n - 1,01\, x_{n-1} + 0,2\, x_{n-2} + e_{n+1}$
after the change: $x_{n+1} = 0,85\, x_n - 0,25\, x_{n-1} + 0,06\, x_{n-2} + e_{n+1}$

Model B: before the change: $x_{n+1} = 0,6\, x_n + e_{n+1}$
after the change: $x_{n+1} = 0,1\, x_n + 2\, e_{n+1}$

Model C: before the change: $x_{n+1} = 0,3\, x_n + 0,5\, x_{n-1} + 4\, e_{n+1}$
after the change: $x_{n+1} = 0,3\, x_n + 0,5\, x_{n-1} + 0,25\, e_{n+1}$

Model D: before the change: $x_{n+1} = 0,85\, x_n - 0,25\, x_{n-1} + 0,06\, x_{n-2} + 3\, e_{n+1}$
after the change: $x_{n+1} = 1,33\, x_n - 0,45\, x_{n-1} - 0,04\, x_{n-2} + 0,5\, e_{n+1}$

Model E: before the change: $x_{n+1} = -0,6\, x_n + e_{n+1} - 0,6\, e_n$
after the change: $x_{n+1} = -0,1\, x_n + e_{n+1} + 0.5\, e_n.$

The last model F was constructed from the previous models A, B, C, D, E and can be schematically expressed as

$$F = A + B + C + D + E,$$

i.e. $5\,000$ observations together with 9 changes at the positions $501 + k \cdot 500$, $k = 0, 1, \ldots, 8$.

Some of the considered sequences were borrowed from the paper [2].

We studied the role of the varying length of the n.oving windows by all the models A, B, C, D, E. Similarly, we studied the influence of the varying order p of an autoregressive approximation. The order p was chosen up to the value 4. The influence of these parameter on the delayed time after the detection is summarized in the following table.

Table of delay

Model A

$L^{\,P}$	1	2	3	4
100	82	x	x	x
150	75	100	140	147
200	90	108	109	149

Model B

$L^{\,P}$	1	2	3	4
25	5	11	·	·
50	12	5	·	·
100	10	17	16	18

Model C

$L^{\,P}$	1	2	3	4
25	23	·	·	·
50	45	48	·	·
100	30	89	92	·

Model D

$L^{\,P}$	1	2	3	4
50	39	·	·	·
100	59	79	68	·
150	69	71	72	89

Model E

$L^{\,P}$	1	2	3	4
50	x	31	18	·
100	x	35	21	·
150	81	19	10	·

The mark x means the method is failing, i.e. there was no possibility to detect unambiguously the change from the behaviour of $\overline{J}(\hat{P}, \hat{Q})$, the mark · means this case was not carried out.

In the figure F concerning the model F $\overline{J}(\hat{P}, \hat{Q})$ is pictured with the choice $L = 150$, $m = 150$ and $p = 2$. It means, in the figure F the changes are at the points $201 + k \cdot 500$, $' = 0, 1, \ldots, 8$. At the first sight one can see that the modified method detected each change. This approach can be used in practice for recognizing homogeneous parts in locally stationary sequences.

A.· follows from the above results there is probably no general receipt how to determine the parameters L, m and p of the method. It is evident the more complicated sequence the longer detecting and reference windows must be used. In the case of autoregressive sequences it is sufficient to use the simplest approximation models with $p = 1$ or $p = 2$. But, it is not true in the case of ARMA sequences as follows from the model E. Here, the autoregressive approximation must be longer otherwise the method is failing. All these particular results can be summarized into a recommendation for practical use. If we have enough time to handle with data quite arbitrarily we have to try finding "optimal" values for L, m and p by experiments. In the "on-line" situation there is the only possibility to watch several parallel moving windows with different lengths simultaneously. Such an approach diminishes the probability of a false alarm and gets shorter delay time after a change.
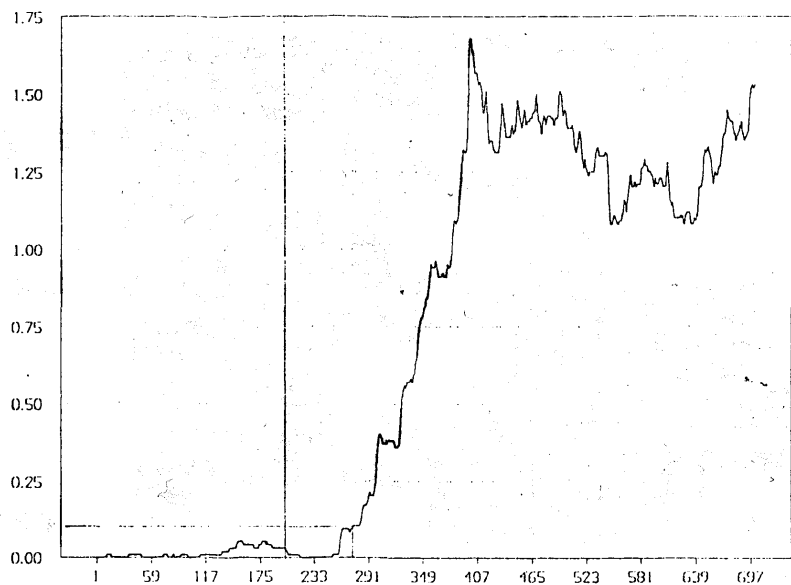
Fig. A. $L = 150$, $m = 150, p = 1$, change $= 201$, delay $= 75$, upper bound $= 0, 1$.
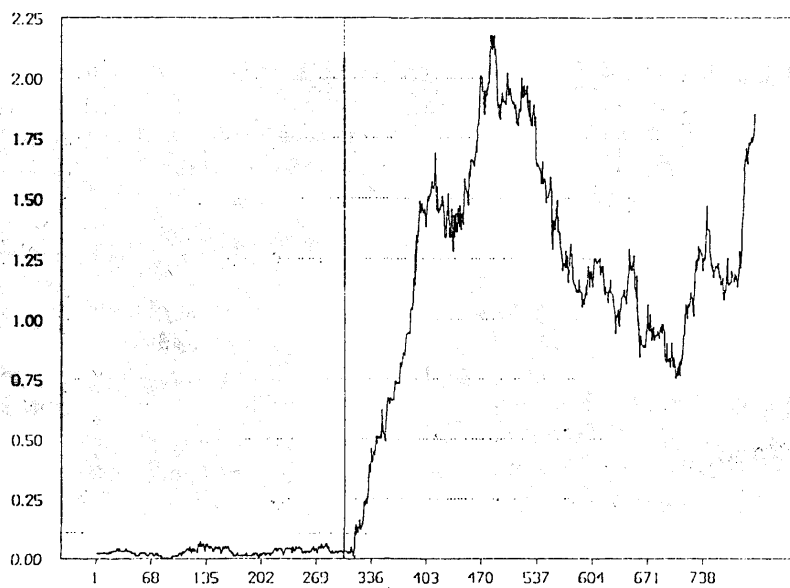

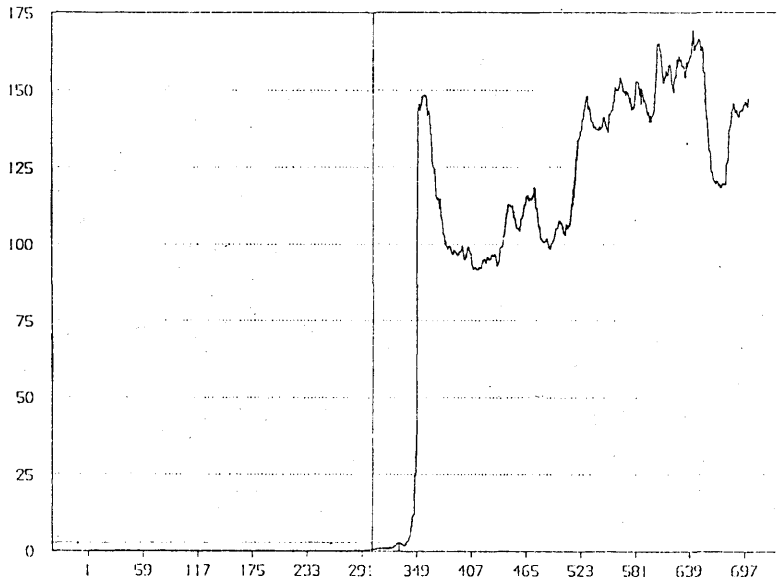**Fig. B.** $L = 100$, $m = 100, p = 2$, change $= 301$, delay $= 17$, upper bound $= 0, 1$.

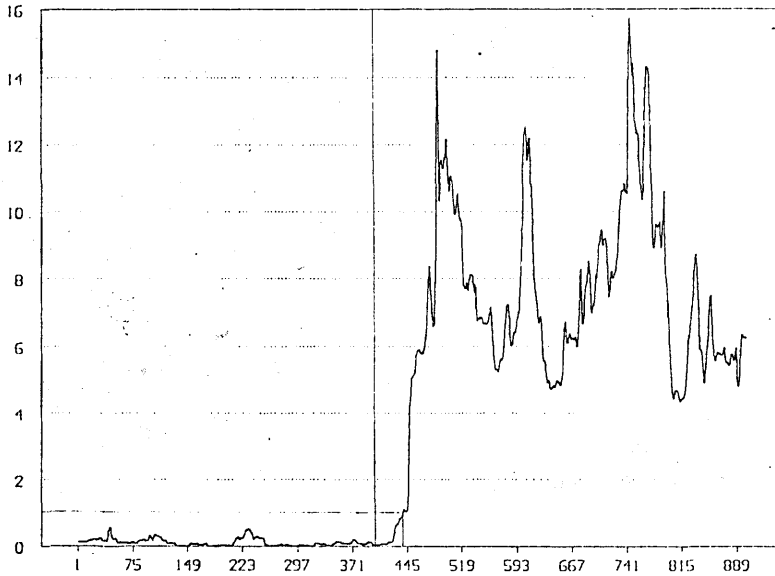**Fig. C.** $L = 100,\ m = 100, p = 3$, change $= 301$, delay $= 30$, upper bound $= 2, 5$.

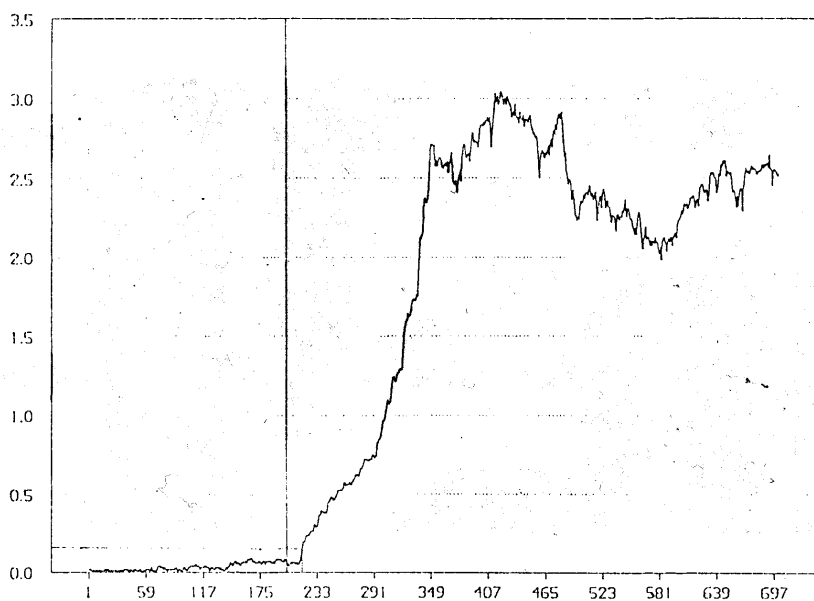**Fig. D.** $L = 50,\ m = 50, p = 1$, change $= 401$, delay $= 31$, upper bound $= 1$.

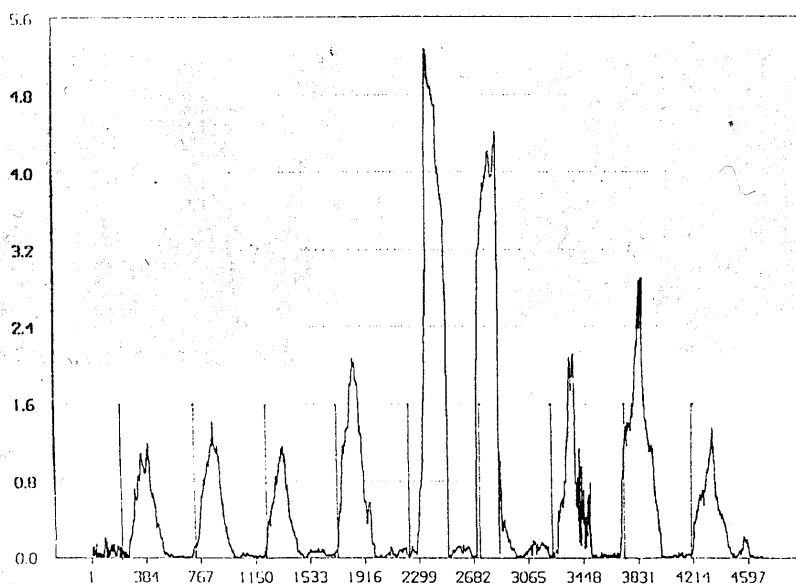**Fig. E.** $L = 150$, $m = 150$, $p = 2$, change $= 201$, delay $= 19$, upper bound $= 0, 15$.

Fig. F. $L = 150$, $m = 150$, $p = 2$, change $= 201 + k \cdot 500$, $k = 0, 1, 2, \ldots, 8$.

ACKNOWLEDGEMENT

(Received October 26, 1993.)

REFERENCES

[1] R. R. Bahadur: Some Limits Theorem in Statistics: Regional Conference Series in Appl. Math. SIAM Pubs., Philadelphia 1971.
[2] M. Basseville and A. Benveniste: Sequential detection of abrupt changes in spectral characteristics of digital signals. IEEE Trans. Inform. Theory *IT-29* (1983), 5, 709–724.
[3] M. Basseville and A. Benveniste (eds.): Detection of Abrupt Changes in Signals and Dynamic Systems (Lecture Notes in Control and Inform. Sci. 77). Springer, Berlin 1986.
[4] V. Grenander and M. Rossenblatt: Statistical Analysis of Stationary Time Series. J. Wiley, New York 1957.
[5] N. Kligiene and L. A. Telksnys: Methods of detecting instants of change of random process properties. Automat. Remote Control *44* (1983), 1241–1283.
[6] H. Künsch: Thermodynamics and statistical analysis of Gaussian random fields. Z. Wahrschein. verw. Gebiete *58* (1981), 407–421.
[7] J. Michálek: Yule–Walker estimates and asymptotic *I*-divergence rate. Problems Control Inform. Theory *19* (1990), 5–6, 387–398.
[8] I. V. Nikoforov: Sequential Detection of Abrupt Changes in Time Series Properties. Nauka, Moscow 1983 (in Russian).
[9] I. Vajda: Theory of Statistical Inference and Information. Kluwer, Dordrecht, Boston 1989.
[10] I. Vajda: Distances and discrimination rates of stochastic processes. Stochastic Process. Appl. *3* (1990), 47–57.
[11] A. S. Willsky A survey of design methods for failure detection in dynamic systems. Automatica *12* (1976), 601–611.

RNDr. Jiří Michálek, CSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.