KYBERNETIKA — VOLUME *28* (1992), NUMBER 6, PAGES 472 – 483

A METHOD OF TREND RECOGNITION IN TIME SERIES

Jiří Michálek and Jiří Skřivánek

The paper describes a method for detecting changes in the behaviour of a trend in time series. Simultaneously a method for distinguishing types of possible trends of time series is presented too.

1. INTRODUCTION

The problem of change detection in the behaviour of time series has been very intensively studied in last fifteen years. There is a lot of papers dealing with this problem but we refer only to some of them, which consider changes of mean value of time series. There is a publication [1], whose first chapter is devoted to the detection of jumps in the mean of a signal. The analysis of the most used methods is presented there. We can also mention a comparative study [2] of some sequential jump detection procedures. Usually the assumption about mutual stochastic independence among observations is accepted. A jump detection is equivalent to the acceptance of the hypothesis H_1 of a change against the hypothesis H_0 of no change during the observation. The detection of a change in the mean value can be understood as a hypothesis testing problem where one wishes to minimize the error of the first and second kind. Here, the first kind error is the false alarm and the detection delay presents the second kind error. If we knew the probability density function of observations the likelihood ratio test would be in a certain sense optimal as proved in [3] or [4]. But in the case of time series the situation is much more complicated because of a possible stochastic dependence. The first goal of the paper is to suggest a method detecting changes in a trend of time series and the other goal is to recognize the type of a detected change in the behaviour of a trend.

2. FORMULATING PROBLEM

In practice we can very often meet time series that can be described by an additive model consisting of two components $\{m_n\}$, $\{e_n\}$, i.e.

$$x_n = m_n + e_n$$

where $\{m_n\}$ presents a trend of time series and $\{e_n\}$ is an error.

We shall assume that the errors $\{e_n\}$ can be expressed by a colour noise with vanishing mean value and an unknown spectral density function. The trend $\{m_n\}$ can be understood either as the mean value of the sequence $\{x_n\}$ or as a random process usually stochastically independent on $\{e_n\}$. In the latter case one must find a suitable mathematical model describing the behavior of $\{m_n\}$ which could be a very difficult way, so we will assume $\{m_n\}$ can be considered as a deterministic trend. Since the trend $\{m_n\}$ can change in time the observed process $\{x_n\}$ is nonstationary in general. But within some time intervals when the trend assigns a constant value,

the sequence $\{x_n\}$ is weak stationary. In this sense we shall speak about local weak stationarity. Our task is obtain as much as possible information about the behaviour of $\{m_n\}$ via observed data $\{x_n\}$, i. e. from this point of view the problem of change detection belongs among typical statistical decision problems. As a basic state we mean the stationary situation when the trend $\{m_n\}$ shows a constant course. Any deviation from this state presents a change, which should be detected as quickly as possible. The next step is, very important in practice too, how to characterize the type of a detected change. It means we have a collection of typical courses occurring after changes and we must solve the decision problem of which typical course is the most similar to the course we have just observed. This decision problem we can be met very often in technical diagnostics. On the other hand, one must also determine the end of the change, which was detected before because we must be ready to accept next changes. For these purposes we need some state variables describing the course of the observed process after detecting a change. Such a state variable is defined on the trajectories of our process assigning its value in a state space. The state space is usually determined by experience and practice. Every situation after occurring any change may be understood as a state that is typical for some situations corresponding to the technical condition of a device. For the description of similarity between two states we need a similarity measure by means of which we could choose the nearest state with respect to the reality. In this way one can appreciate the course of the observed process with on-line regime.

3. THEORETICAL BACKGROUND

The basic assumption is that the colored noise is a weakly stationary regular random sequence with vanishing mean value. A trend $\{m_n\}$ is added to this noise, whose basic state is a constant level. Until any change occurs the result of our observation is a weakly stationary sequence with a certain mean value. Since we do not know the parameters of the process $\{x_n\}$ it is necessary to obtain some estimates characterizing the course of $\{x_n\}$. At this situation we use the following idea to substitute the completely unknown sequence $\{x_n\}$ we have observed by a simpler sequence, which would be similar to the observed sequence as much as possible. After constructing such a suitable approximation we shall make our decisions on the basis of the approximating sequence. It seems to be quite reasonable to choose the family of Gaussian autoregressive sequences. We suggest to measure the similarity between the observed sequence and an autoregressive approximation by means of the asymptotic I-divergence rate having the form

$$\bar{R}_1\left(\hat{\varphi}_x, \varepsilon_a\right) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\hat{\varphi}_x}{\varphi_a} - \ln \frac{\hat{\varphi}_x}{\varphi_a} - 1 \right) \mathrm{d}\lambda$$

where $\hat{\varphi}_x$ is the spectral density function derived from observation and φ_a is the spectral density function of an autoregressive model with a suitable order. For more information we refer to the monograph [5]. The problem how to choose a suitable order of an autoregressive sequence is not solved in this paper but one can say the more abrupt changes the shorter order is sufficient for a good detection. Let us assume we have chosen the order p for an autoregressive model

$$y_{n+1} + \sum_{k=1}^{p} a_k \, y_{n-k+1} = \sigma_p \, \xi_{n+1}$$

where $\{\xi_n\}$ is a standard Gaussian white noise. Minimizing $\bar{R}_1\left(\hat{\varphi}_x, \hat{\varphi}_a\right)$ with

$$\hat{\varphi}_a(\lambda) = \frac{1}{2\pi} \frac{\sigma_p^2}{\left| \sum_{j=1}^{p} a_j e^{i\lambda j} \right|^2}$$

and

$$\hat{\varphi}_x(\lambda) = \frac{1}{2\pi} \sum_{j=-p}^{p} e^{-ij\lambda} \hat{R}_j,$$

where for $j = 0, 1, \ldots, p$

$$\hat{R}_j = \frac{1}{N} \sum_{k=1}^{N-j} (x_k - \bar{x})\,(x_{k+j} - \bar{x})\,, \qquad \hat{R}_j = \hat{R}_{-j}$$

and

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k,$$

we will reach the system of Yule–Walker equations for $(a_1, a_2, \ldots, a_p, \sigma_p)$ as proved in [6]. It means the most similar autoregressive model is given by the Yule–Walker estimates. The system of the Yule–Walker equations can be very quickly solved by means of the Levinson algorithm. The number N presents the length of a sliding window moving over the observations.

At this moment we have at our disposal the estimates $(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}_p)$ obtained by the procedure described above and we can work with the approximating model

$$y_{n+1} + \sum_{j=1}^{p} \hat{a}_p\, y_{n+1-j} = \hat{\sigma}_p\, \xi_{n+1}.$$

Knowing the values $x_n, x_{n-1}, \ldots, x_{n-p}$ we can, using this model, construct a "best" predictor of x_{n+1} under the history of the length p. This predictor has the form

$$\hat{x}_{n+1} = -\sum_{j=1}^{p} \hat{a}_j x_{n-j} + \left(1 + \sum_{j=1}^{p} \hat{a}_j\right) \bar{x}.$$

There is an opportunity to use this predictor for the detection of changes in the behaviour of $\{m_n\}$. Coming out from the assumption we are in the basic state, i. e. the weak stationarity of $\{x_n\}$, we can expect under the maintenance of stationarity the residuum $x_{n+1} - \hat{x}_{n+1}$ should be relatively small. In such a case with high probability x_{n+1} should fall into a small neighbourhood of $\hat{x}_{n+1}$, which can be determined as an interval of reliability by means of the estimate $\hat{\sigma}_p^2$. The correctness of this approach is based on the result given in [7] that showed that increasing the order p of the approximating autoregressive model we can reach a very good similarity between the approximation and reality because

$$\lim_{p \to \infty} \bar{R}_1\,(\hat{\varphi}_x,\, \hat{\varphi}_a(p)) = 0,$$

if we knew, of course precisely the spectral measure $\hat{\varphi}(\cdot)$ corresponding to the stationary state. Here, $\hat{\varphi}_x$ is derived from the observations as mentioned above. As long as the difference between $x_{n+1} - \hat{x}_{n+1}$ is larger then a change in the behaviour of $\{m_n\}$ is detected. In this way all the information on the behaviour of the time series is expressed via the sequence of residua. It is evident as long as a change is sufficiently expressive then the residuum is with high probability expressive too since the predictor is somewhat delayed behind the reality. This fact can be utilized for the construction of an adaptive prediction in time series, which is switched on immediately if a change in the trend was detected. This adaptive prediction consists in the simple fact that we add to the predicted value $\hat{x}_{n+2}$ the previous residuum $x_{n+1} - \hat{x}_{n+1}$. But, we will solve the problem how to use the behaviour of residua for the characterization of the course of the observed time series. The course of the

sequence will be characterized locally which means we will treat the last residua only. How a long sequence of residua one must use we will not discuss here, but as the experience says on the basis of simulations one can already use 15 – 20 samples of residua. Everything is depending on the types of changes and on the sampling period. As long as a change is very slow it may happen such a situation will not be detected at all.

In practice we can use the following state variable $\{S_n\}$ describing the course of time series. This variable is derived of residua and assigns three values only: $0, +1, -1$ according to the magnitude and sign of the residuum. If

$$|x_{n+1} - \hat{x}_{n+1}| < \delta,$$

where $\delta > 0$ is the threshold determined by $\hat{\sigma}_p$, then the state variable S_{n+1} equals 0. As long as

$$x_{n+1} - \hat{x}_{n+1} > \delta,$$

then we put $S_{n+1} = +1$. On the other hand, if

$$x_{n+1} - \hat{x}_{n+1} < -\delta$$

S_{n+1} will be equal to the value –1. In this way the course of $\{x_n\}$ is characterized by the sequence of signs from the alphabet $\{-1, 0, +1\}$. It could be possible to enlarge the alphabet to 5 or 7 signs according to the magnitudes of residua so that one can distinguish the intensity of changes too.

At this moment we will assume the state variable $\{S_n\}$ forms a Markov chain acquiring three states: -1, 0, $+1$. Further, we will assume that this chain is locally stationary, i. e. relatively short periods of residua can be understood as a realization of a stationary chain. For the evaluation of the time series course we will compute the relative frequency

$$\hat{p}_{ij} = \frac{n_{ij}}{n-1}$$

corresponding to the number of the immediate transitions from the state i into the state j within the window containing the last values of the state variable $\{S_n\}$. Here, the number n presents the length of this moving window. We obtained in such a way a matrix

$$\hat{\mathbf{P}} = \{\hat{p}_{ij}\}_{i,\, j\in\{-1,0,1\}},$$

an estimate of the probability distribution for the random variable $\{S_n(\cdot),\, S_{n+1}(\cdot)\}$

$$P\{\omega :\ S_n(\omega) = i,\ S_{n+1}(\omega) = j\}.$$

Now, there is the problem how to characterize the similarity between two courses. Let some typical examples based on practice and experience be chosen. These typical courses are interesting for us from some reasons; e. g. from the point of view of technical diagnostics. Every chosen course is characterized by a matrix

$$\mathbf{Q}_k = \left\{q_{ij}^{(k)}\right\}_{i,\, j\in\{-1,0,+1\}}$$

$k = 1, 2, \ldots, M$. Now, the task is to choose a course described by the matrix $\mathbf{Q}_k,\ k = 1, 2, \ldots, M$ in order to be the most similar to the observed course given by the matrix $\hat{\mathbf{P}}$. As a similarity measure we will use the I-divergence between the probability distributions $\mathbf{Q}_k$ and $\hat{\mathbf{P}}$, i. e.

$$I\left(\hat{\mathbf{P}}|\mathbf{Q}_k\right) = \sum_{i,\, j\in\{-1,0,+1\}} \hat{p}_{ij} \ln \frac{\hat{p}_{ij}}{q_{ij}^{(k)}}.$$

The solution of our classification problem is given by minimizing

$$\min_{k=1,2,\ldots,M} I\left(\hat{\mathbf{P}}|\mathbf{Q}_k\right).$$

Using this method one can meet the situation $q_{ij}^{(k)} = 0$ for some pair i, j. In order to overcome these difficulties it seems to be reasonable to put $q_{ij}^{(k)} = \varepsilon > 0$ in such a case where ε is a small positive member chosen below $\min_{i,j}\left\{q_{ij}^{(k)} > 0\right\}$.

The suggested approach can be made more severe by setting up a threshold that must not be overcome by the minimum value of I-divergence. In this way we want to minimize the error choosing the closest course to the observed one, but this closest course will not be similar to the reality as we would wish.

There is of course a possibility to use another measure of similarity. A very frequently used measure is the χ^2-distance having the famous form

$$\chi^2\left(\mathbf{P}|\mathbf{Q}_k\right) = \sum_{i,j\in\{-1,0,1\}} \frac{\left(p_{ij} - q_{ij}^{(k)}\right)^2}{q_{ij}^{(k)}}.$$

Since we assumed the stationarity of the Markov property the next measure derived from stationarity is possible too. The measure in question is so called asymptotic I-divergence rate equal to

$$\mathrm{AIR}\left(\hat{\mathbf{P}}|\mathbf{Q}_k\right) = \sum_{i,j\in\{-1,0,1\}} p_{ij} \ln \frac{p_{i|j}}{q_{i|j}^{(k)}}$$

where $p_{i|j} = P\{S_{n+1} = i \,|\, S_n = j\}$. This measure of similarity is introduced by the following Lemma.

Lemma. Let $\{\xi_n\}$, $\{\eta_n\}$ be two stationary Markov chains with a finite number of states $\{1, 2, \ldots, N\}$. Let $\mathbf{P}$, $\mathbf{Q}$ be their transition probability matrices, $\mathbf{p} = (p_1, p_2, \ldots, p_N)$, $\mathbf{q} = (q_1, q_2, \ldots, q_N)$ their initial probability distributions; $p_{ij} = 0$ if and only if $q_{ij} = 0$. Then there exists the following limit

$$\lim_{n\to\infty} \mathsf{E}_{P_n}\left\{P_n\left(\xi_1, \xi_2, \ldots, \xi_n\right) \ln \frac{P_n(\xi_1, \xi_2, \ldots, \xi_n)}{Q_n(\eta_1, \eta_2, \ldots, \eta_n)}\right\} =$$

$$= \sum_{i,j=1}^{N} p_{ij} \ln \frac{p_{i|j}}{q_{i|j}}$$

where $\mathbf{P} = \left\{p_{i|j}\right\}_{i,j=1}^{N}$, $\mathbf{Q} = \left\{Q_{i|j}\right\}_{i,j=1}^{N}$, $p_{ij} = p_j\, p_{i|j}$, $q_{ij} = q_j\, q_{i|j}$,

$$\begin{aligned} P_n\left(\xi_1, \xi_2, \ldots, \xi_n\right) &= P\left\{\xi_1 = i_1,\, \xi_2 = i_2, \ldots, \xi_n = i_n\right\}, \\ Q_n\left(\eta_1, \eta_2, \ldots, \eta_n\right) &= P\left\{\eta_1 = i_1,\, \eta_2 = i_2, \ldots, \eta_n = i_n\right\} \end{aligned}$$

with $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, N\}$.

P r o o f. Let us put $\ln \frac{0}{0} = 0$ to overcome problems with division by 0. Thanks to the Markov property we can write

$$\begin{aligned} P_n\left(\xi_1, \xi_2, \ldots, \xi_n\right) &= p_{i_1}\, p_{i_2|i_1}\, p_{i_3|i_2} \cdots p_{i_n|i_{n-1}} \\ Q_n\left(\eta_1, \eta_2, \ldots, \eta_n\right) &= q_{i_1}\, q_{i_2|i_1}\, q_{i_3|i_2} \cdots q_{i_n|i_{n-1}}. \end{aligned}$$

Then, the I-divergence $I\left(P_n|Q_n\right)$ equals

$$\mathsf{E}_{P_n}\left\{P_n\left(\xi_1, \xi_2, \ldots, \xi_n\right) \ln \frac{P_n(\xi_1, \xi_2, \ldots, \xi_n)}{Q_n(\eta_1, \eta_2, \ldots, \eta_n)}\right\} =$$

$$= \sum_{(i_1, i_2, \ldots, i_n)} p_{i_1} p_{i_2|i_1} \cdots p_{i_n|i_{n-1}} \ln \frac{p_{i_1} p_{i_2|i_1} \cdots p_{i_n|i_{n-1}}}{q_{i_1} q_{i_2|i_1} \cdots q_{i_n|i_{n-1}}}.$$

Since the considered Markov chains are stationary $I(P_n|Q_n)$ can be expressed as follows:

$$I(P_n|Q_n) = \sum_{(i_1,i_2,\ldots,i_n)} p_{i_1} \prod_{(i,j)} p_{i|j}^{n_{ij}} \ln \left\{ \frac{p_{i_1} \prod_{(i,j)} p_{i|j}^{n_{ij}}}{q_{i_1}} \prod_{(i,j)} q_{i|j}^{n_{ij}} \right\}$$

where n_{ij} is the number of the immediate transitions $j \to i$ within the time period of the length n; evidently $\sum_{(i,j)} n_{ij} = n - 1$.

In this way we obtained the following form

$$P(P_n|Q_n) = \sum_{(i_1,i_2,\ldots,i_n)} P_n(\xi_1, \xi_2, \ldots, \xi_n) \left\{ \sum_{i,j=1}^{N} n_{ij} \ln \frac{p_{i|j}}{q_{i|j}} + \ln \frac{p_{i_1}}{q_{i_1}} \right\}.$$

Then

$$\frac{1}{n} I(P_n|Q_n) = \sum_{(i_1,i_2,\ldots,i_n)} \left\{ \sum_{i,j=1}^{N} \frac{n_{ij}}{n} \ln \frac{p_{i|j}}{q_{i|j}} + \frac{1}{n} \ln \frac{p_{i_1}}{q_{i_1}} \right\} P_n\{\xi_1, \xi_2, \ldots, \xi_n\}.$$

Now, thanks to the law of large numbers and to the finite number of states one can easily prove that

$$\lim_{n\to\infty} \frac{1}{n} I(P_n|Q_n) = \sum_{i,j=1}^{N} p_{ij} \ln \frac{p_{i|j}}{q_{i|j}}$$

because $\frac{n_{ij}}{n} \longrightarrow_{n\to\infty} p_j\, p_{i|j}$ in probability. □

4. APPLICATION IN PRACTICE

For illustration of the given method we will show a few examples and will give a detailed discussion of application. In practice, one must very often know whether the course of an observed process is regular or irregular. Among regular courses a stationary course without any changes naturally belongs. A regular change of course is the behaviour of an observed process corresponding to a controlled process. Such a typical regular change is a monotone change expressing a transition from one level to another one. The recognition of those changes and their intensity too is very important for technical diagnostics. A monotonic change belonging to the class of regular courses is given in Figure 1. An irregular course with very rapid changes in the behaviour of a technological process caused by a not working feed-back is presented in Figure 2. Figure 3 describes an

Fig. 1. Monotonic change.

Fig. 2. Irregular behaviour.

Fig. 3. Admissible behaviour.

admissible course of a technological process. The course of a technological process is denoted by a full line. The sequence of small squares presents the course of the

one step predictor following possible changes. The difference between the measured value and its prediction is the residuum. All the courses were obtained from practice.

During the construction of a program for a computer some modifications of the described method were done. With regard to practice one of the important information is to know whether an observed error signal represented in a moving window as a state variable series $\{S_n\}$ is increasing or decreasing. The probabilistic approach given above cannot distinguish these two dissimilar cases. For a better data processing the original values in $\{S_n\}$, i. e. $0, +1, -1$ were substituted by $0, 2, 1$, so that 2 represents the increase and 1 the decrease of the error signal. One can easily imagine $\{S_n\}$, e. g. 0, 0, 2, 2, 2, 0, 1, 2, 1, 1, 1, 0, 0, $\ldots, 0$ describing an "increase" and $\{S'_m\}$ as
0, 0, 1, 1, 1, 0, 2, 1, 2, 2, 2, 0, 0, $\ldots, 0$ describing a "decrease", while their transition matrices $\hat{\mathbf{P}}_n$ and $\hat{\mathbf{P}}_m$ are identical. For this reason a switch before the proper classification procedure was installed, dividing all the error signals into two basic groups according to the beginning of a respective state variable series.

Further, in a moving window with n data we can observe $(n-1)$ transitions only, i. e. for the last value the transition is unknown as far as in the next step when the $(n+1)$-st value becomes the n-th one. So, the last value has a nonequivalent position with respect to other data in a moving window. It shows to be convenient to know an auxiliary $(n+1)$-st value determining the last transition from n to $(n+1)$. The following example demonstrates the usefulness of the knowledge of the $(n+1)$-st value in an $\{S_n\}$-sequence. Let us assume the 10-data moving window: 0, 0, 0, 2, 2, 2, 1, 1, 2, 1. In the next step of observation the error signal – after shifting one step from the right to the left – changes into 0, 0, 2, 2, 2, 1, 1, 2, 1, 0. Although dealing evidently with the same error signal, the $\{S_n\}$-sequence before shifting has no transition from 1 to 0, and thus it has a different qualitative transition matrix as the shifted sequence. Comparing these two sequences, we meet the problem described earlier: how to treat the term $q_{ij}^{(k)} = 0$; always the I-divergence value will be greater than in the case when the transition $(1 \rightarrow 0)$ is known, the similarity measure is then distorted.

This just described problem can be generalized: how to avoid such cases where $\mathbf{P}$ and $\mathbf{Q}$ contain $p_{ij} = 0$ or $q_{ij} = 0$ at different positions? By practical examples it can be shown that the I-divergence value is mostly influenced by these undefined terms. The only way ensuring that these undefined terms cannot appear at all is the use of $\mathbf{P}$ and $\mathbf{Q}$ with all the same types of transitions. In our described 9-elements transition matrix there exist about 70 combinations of different transition types for an increasing error signal and the same number for a decreasing signal, which one can meet in practice. Other examples are almost impossible. Fortunately, we can deal with only one of them, e. g. with the increasing signal and thanks to the installed switch before the proper classification procedure, every decreasing signal can be transformed on the basis of mirror symmetry into a similar increasing one, as it is shown in the first example (1 is transformed into 2, 2 into 1 and 0 remains without any change). The detailed analysis of all the possible combinations of different transitions types leads to the conclusion that for practice about 20 of them can be utilized only.

The classification procedure after the basic determination of an error signal (increasing/decreasing) at the beginning starts with the error type identification selecting one type from 20 possibilities mentioned above. Each of these 20 error type classes can contain several representants $\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_r$ that are successively compared with the observed error signal and the respective values of similarity measure are calculated. The optimal similarity measure is then chosen to be minimum of them. The corresponding I-divergence value is finally compared with a prescribed threshold that must not be overcome for the similarity to be considered as satisfactory.

Several typical courses were tested by the described method to achieve a sufficient number of combinations in error type classes, e. g. step, S-curve, jump, oscillations and some others. Signals that were obtained as actual regular processes are characterized with several numbers of error types only.

Figure 4 shows three courses of signals that were considered being representation types. Firstly, the type 0 presents a relatively quiet course (output window contains zeros only); secondly, the type B is a sequence of dumped oscillations (output window contains 0, 1 and 2) and the third course, type A, is an increase (output window contains 0 and 2 only).

Fig. 4. Reference signal.

Figure 5 shows three representants of oscillations (possible subtypes) that have a different number of transitions but with the same kind of them. Similarly, Figure 6 shows 3 subtypes of increases. All the subtypes were chosen according to the representative types but with different amplitudes. In Figure 7 a real signal coming to the input of the discrimination procedure is displayed. The suggested procedure works in two levels. At the first level, the corresponding type of a signal according to the kind of transitions determined within a sliding window is found. In this step we choose a family of all courses having the same types of transitions. The second step is already carried out within this family. We prepared some typical courses (subtypes) in advance, which are interesting for us for some reasons. Here, see Fig.

5 and Fig. 6, where some typical

Fig. 5. Oscillations – reference subtypes.

Fig. 6. Increase – reference subtypes.

subtypes were chosen in advance. We evaluate the value $I(P, Q_i)$ for every pair P, Q_i where the matrix P is derived from a real signal and the matrices $Q_i,\ i = 1, 2, 3$ were chosen in advance as typical. The matrices $Q_i,\ i = 1, 2, 3$ represent the chosen subtypes

Fig. 7. Real signal.

in every family. After finding the minimum value, let's say $I(P, Q_1)$ the real signal is declared to be the most similar to the reference signal No 1. The following table brings the values of $I(P, Q)$ for four chosen windows covering some parts of a real signal.

window	$I_{pq}(1)$	$I_{pq}(2)$	$I_{pq}(3)$	minimal value	most similar subtype
32 – 51	0,065	0,347	0,641	0,063	A1
61 – 80	0,137	0,129	0,066	0,066	B3
133 – 152	0,013	0,047	0,183	0,013	A1
169 – 188	0,117	0,019	0,069	0,019	B2

$$I_{pq}(i) = I(P, Q_i), \qquad i = 1, 2, 3.$$

In the window 32–51 the type A was detected and among the chosen subtypes A_1, A_2, A_3 the subtype A_1 is the most similar to the real signal. Similarly, in the window 61–80 we first detected the type B and then among the subtypes B_1, B_2, B_3 the subtype B_3 was chosen according to the minimal value of I-divergence. In the same way the signal in the window 133–152 in Figure 7 is the most similar to the subtype A_1. The last window 169–188 is the most similar to the subtype B_2.

(Received October 4, 1991.)

REFERENCES

[1] M. Basseville and A. Benveniste (eds.): Detection of Abrupt Changes in Signals and Dynamical Systems. (Lecture Notes in Control and Information Sciences 77.) Springer–Verlag, Berlin – Heidelberg – New York 1986.

[2] M. Basseville and A. Benveniste: Design and comparative study of some sequential jump detection algorithms for digital signals. IEEE Trans. Acoust. Speach. Sig. Process *A.S.S.P.–31* (1983), 3, 521–534.

[3] G. Lorden: Procedures for reacting to a change in distribution. Ann. Math. Statist. *42* (1971), 6, 1897–1908.

[4] A. N. Shiriyaev: On optimum methods in quickest detection problems. Theory Probab. Appl. *8* (1963), 1, 22–46.

[5] I. Vajda: Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht 1989.

[6] J. Michálek: Yule–Walker estimates and asymptotic I-divergence rate. Problems Control Inform. Theory *19* (1990), 5-6, 387–398.

[7] J. Michálek: Levinson's algorithm and I-divergence. In: Proc. 10-th Prague Conf. on Inform. Theory, Stat. Dec. Funct. Random Proc., vol. B, Academia, Prague 1992, pp. 239–250.

RNDr. Jiří Michálek, CSc., Ing. Jiří Skřivánek, Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation – Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.