KYBERNETIKA — VOLUME 29 (1993), NUMBER 2, PAGES 166–179

A METHOD OF DETECTING CHANGES IN THE BEHAVIOUR OF A RANDOM SEQUENCE BASED ON THE BAYES APPROACH

JIŘÍ MICHÁLEK AND JIŘÍ SKŘIVÁNEK

A method for the detection of changes in the behaviour of a locally stationary sequence is presented. A change can occur either as an abrupt jump in mean value or as a change in the spectral density function. The detection of changes is based on the Bayes approach. The suggested method is compared with a classical method based on maximal likelihood ratio and is presented by several numerical simulations.

1. INTRODUCTION

The problem of the detection of changes in the behaviour of time series has been very intensively studied during last fifteen years because this is a part of mathematical statistics having immediate application in practice, e. g. in technical diagnosis. The detection of changes in the behaviour of a time series belongs to the testing hypotheses region as we are testing the simple hypothesis "no change" versus the composed alternative hypothesis "a change occurred". With respect to the alternative hypothesis, which permits a possible change at every time instant in the future, the uniformly best test does not exist, in general that would be analogical to the classical Neyman–Pearson test. One can speak about an asymptotically best test as shown by Deshayes and Picard in Basseville and Benveniste [1]. This asymptotically optimal test is given by the generalized likelihood ratio but the precise knowledge of models before a change and after a change is assumed. Such a result is very interesting from the theoretical point of view, but in a practical case we usually don't know almost anything about an observed time series. One can find a series of different methods in the literature because of the nonexistence of an optimal test in a finite time period. We refer the reader for a better acquaintance to some survey paper dealing with the detection problem, e. g. Kligiene and Telksnys [3], Basseville and Benveniste [1], Nikoforov [6], Willsky [9]. A usual approach is based on the assumption that we know a mathematical model describing the situation before a change. Under the presence of a change in its behavior, e. g. in the change of its mean value in case when a change has occurred. In practice parameters defining a model before a change are supplied by suitable statistical estimates. The method

discussed in the paper is based on the Bayes approach for solving statistical decision problems. Firstly, the simplest case consisting of mutually independent observations before and also after a change is solved. Then, the proposed method is applied to the detection of changes in the behavior of a locally stationary sequence, which is a sequence consisting of mutually nonoverlapping time intervals within which such a sequence is weakly stationary. The time instants separating adjoing intervals are the instants of changes that have to be detected.

2. THEORETICAL BACKGROUND

First, we shall consider the case with mutually independent observations $x_1, \ldots, x_n$, where n represents time. We shall assume that the observations before a change are generated according to a probability density function $f(\cdot)$ and after a change according to $g(\cdot)$. The time instant of a change is not known and can occur quite arbitrarily. Under this situation we have at disposal the observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and we have to decide whether a change has occurred or not. If a change is detected we also have to estimate the time of change. Summarizing this situation we have n possibilities in the composed alternative hypothesis: when at the time j a change occurs the corresponding common probability density function for $\mathbf{x}$ has the form

$$f_j(\mathbf{x}) = \prod_{\ell=1}^{j-1} f(x_\ell) \prod_{\ell=j}^{n} g(x_\ell) \qquad j = 2, 3, \ldots, n$$

and

$$f_1(\mathbf{x}) = \prod_{\ell=1}^{n} g(x_\ell).$$

The hypothesis is created by the only probability density function

$$f_{n+1}(\mathbf{x}) = \prod_{\ell=1}^{n} f(x_\ell).$$

As we wish to follow the Bayes approach we need to determine a prior distribution function for a possible time of change. Let p_j, $j = 1, 2, \ldots, n+1$ present probability of occurring a change at the time j. The time $n+1$ means evidently no change has occurred during the time period $j = 1, 2, \ldots, n$. Surely, we must demand

$$\sum_{j=1}^{n+1} p_j = 1.$$

Let d_j be the decision that a change occurred at the time j. The decision d_{n+1} means, of course, no change was detected. Let $\ell(\cdot, \cdot)$ be the loss function that is given by

$$\ell(\theta_i, d_j) = 1 - \delta_{ij},$$

where $\theta_i \leftrightarrow f_i(\mathbf{x})$ and δ_{ij} is the Kronecker symbol. We look for a Bayes decision rule $\{\varphi(i|\mathbf{x})\}_{i=1}^{n+1}$, where $\varphi(i|\mathbf{x})$ is the conditional probability of the acceptation of d_i under the observation $\mathbf{x}$. The conditional risk function can be then expressed as

$$\begin{aligned} R(\theta_i, \varphi(\cdot|\mathbf{x})) &= \sum_{j=1}^{n+1} \ell(\theta_i, d_j)\, \mathsf{E}_{\theta_i}\{\varphi(j|\mathbf{x})\} = \\ &= 1 - \mathsf{E}_{\theta_i}\{\varphi(x|\mathbf{x})\}. \end{aligned}$$

Thus, the expected value of $R\{\theta_i, \varphi(\cdot|\mathbf{x})\}$ with respect to a prior distribution $\{p_j\}_{j=1}^{n+1}$ is equal to

$$r\left(\{p_j\}_{j=1}^{n+1}, \varphi(\cdot|\mathbf{x})\right) = 1 - \sum_{j=1}^{n+1} p_j\, \mathsf{E}_{\theta_j}\{\varphi(j|\mathbf{x})\}. \tag{1}$$

The Bayesian decision rule must minimize (1). The answer is given by the following decision function $\Phi(\mathbf{x})$:

$$\Phi(\mathbf{x}) = \{\Phi(j|\mathbf{x}) : \; j = 1, 2, \ldots, n+1\}$$

where $\Phi(i|\mathbf{x}) = 0$ if and only if

$$p_i\, f_i(\mathbf{x}) < \max_{\substack{1 \le j \le n+1 \\ j \neq i}} \{p_j\, f_j(\mathbf{x})\}.$$

In other words, if

$$p_{i_0}\, f_{i_0}(\mathbf{x}) > p_j\, f_j(\mathbf{x})$$

for each $j = 1, 2, \ldots, n, n+1, \; j \neq i_0$, then

$$\Phi(i_0|\mathbf{x}) = 1$$

as we demand $\sum_{i=1}^{n+1} \Phi(i|\mathbf{x}) = 1$.

The proof of optimality is very easy and can be dropped out. Next, we shall show how this Bayesian decision rule can be used for the construction of a stopping rule. A change is detected at the time instant n if and only if

$$p_n\, f_n(\mathbf{x}) > \max_{\substack{1 \le j \le n+1 \\ j \neq n}} f_j(\mathbf{x}). \tag{2}$$

Let us suppose for simplicity $f_j(\mathbf{x}) > 0$ for every $\mathbf{x}$ and every j, then condition (2) can be rewritten as

$$1 > \frac{p_j}{p_n} \prod_{t=j}^{n-1} \frac{g(x_t)}{f(x_t)}$$

for each $j = 1, 2, \ldots, n-1$ and simultaneously

$$\frac{p_n\, g(x_n)}{p_{n+1}\, f(x_n)} > 1.$$

From these facts we immediately see that a change is not detected up to the time n including if and only if

$$\max_{1\le j\le n}\left\{\frac{p_j}{p_{n+1}}\prod_{\ell=j}^{n}\frac{g(x_\ell)}{f(x_\ell)}\right\}<1.$$

As long as we find such a time instant j_0 for which

$$\frac{p_{j_0}}{p_{n+1}}\prod_{\ell=j_0}^{n}\frac{g(x_\ell)}{f(x_\ell)}\ge 1$$

then a change in the behaviour of the observed sequence is detected. The decision rule derived from the Bayes approach is described detailly in the following

Lemma 1. Let $f(x)$, $g(x)>0$ for every real x, for simplicity. Then, we accept the decision d_{j_0} ($j_0=1,2,\dots n+1$) if and only if the following inequalities hold true:

for $j_0=1$

$$\frac{p_1}{p_j}\prod_{\ell=1}^{j-1}\frac{g(x_\ell)}{f(x_\ell)}>1 \qquad \text{for each } j=2,3,\dots,n,n+1$$

for $1<j_0<n+1$

$$\frac{p_{j_0}}{p_j}\prod_{\ell=j_0}^{j-1}\frac{g(x_\ell)}{f(x_\ell)}>1 \qquad \text{for each } j=j_0+1,\dots,n,n+1$$

and simultaneously

$$\frac{p_j}{p_{j_0}}\prod_{\ell=j}^{j_0-1}\frac{g(x_\ell)}{f(x_\ell)}>1 \qquad \text{for each } j=1,2,\dots,j_0-1$$

for $j_0=n+1$

$$\frac{p_j}{p_{n+1}}\prod_{\ell=j}^{n}\frac{g(x_\ell)}{f(x_\ell)}>1 \qquad \text{for each } j=1,2,\dots,n.$$

Proof. The proof of Lemma 1 is very simple and can be omitted because it immediately follows from the Bayesian decision rule function. □

Lemma 1 automatically answers the question about an estimate of a possible time of change. If a change is detected at the time j_0 then the instant j_0 is the very estimate of the time of change because it was chosen on the basis of the maximal aposteriori conditional probability $\Phi(j_0|\mathbf{x})$ given by the Bayesian rule.

Now, we shall show an application of the above decision rule to the most important case. Let $f(\cdot)$ be the probability density function of the distribution $N(0,1)$ and let $g(\cdot)$ correspond to the distribution $N(\mu,\sigma^2)$. Then, we find out after simple computation that we accept the decision d_{j_0}, $j_0\in\{2,3,\dots,n-1,n\}$ if for each $j>j_0$

$$\sum_{\ell=j_0}^{j-1}\left(x_\ell^2-\frac{(x_\ell-\mu)^2}{\sigma^2}\right)>\ln\frac{p_j^2}{p_{j_0}^2}+(j-j_0)\ln\sigma^2$$

holds true and simultaneously for each $j < j_0$

$$\sum_{\ell=j}^{j_0-1}\left(x_\ell^2 - \frac{(x_\ell-\mu)^2}{\sigma^2}\right) < \ln\frac{p_{j_0}^2}{p_j^2} + (j-j_0)\ln\sigma^2.$$

The decision d_1 will be accepted when the following inequality will hold true for each $j > 1$, namely

$$\sum_{\ell=1}^{j-1}\left(x_\ell^2 - \frac{(x_\ell-\mu)^2}{\sigma^2}\right) > \ln\frac{p_j^2}{p_1^2} + (j-1)\ln\sigma^2.$$

The decision d_{n+1} will state, i.e. no change was detected if for each $j = 1, 2, \ldots, n$ the inequality

$$\sum_{\ell=j}^{n}\left(x_\ell^2 - \frac{(x_\ell-\mu)^2}{\sigma^2}\right) < \ln\frac{p_{n+1}^2}{p_j^2} + (n+1-j)\ln\sigma^2$$

is true. Although the above given decision procedure is based on the precise knowledge of the alternative hypothesis, i.e. the parameters μ, σ^2 must be known, in practice, where usually μ, σ^2 are unknown the suggested approach can be used. Very often, within the observation of changes in a technological process, we know some admissible changes, which can be reflected in changes of parameters μ, σ^2. These values can be then used in the construction of suitable decision rules. In this way, a collection of parallel decision rules with different values of μ, σ^2 can be used in practice.

Next, we shall compare the test given by Lemma 1 with the test derived from the likelihood ratio. The test based on the maximal likelihood ratio is given by the function $\lambda(\mathbf{x})$ defined on a sample space of $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\mathrm{H}} f(\mathbf{x},\theta)}{\sup_{\theta\in\mathrm{H}\cup\mathrm{A}} f(\mathbf{x},\theta)},$$

where θ is a parameter, H is a hypothesis and A is an alternative hypothesis. Surely, $0 \le \lambda(\mathbf{x}) \le 1$. The critical region of the maximal likelihood ratio is given by the test function $\Phi(\mathbf{x})$ where

$$\begin{aligned}\Phi(\mathbf{x}) &= 1 \quad \text{iff} \quad \lambda(\mathbf{x}) < d_\alpha\\ \Phi(\mathbf{x}) &= 0 \quad \text{iff} \quad \lambda(\mathbf{x}) > d_\alpha\\ \Phi(\mathbf{x}) &= \mu \quad \text{iff} \quad \lambda(\mathbf{x}) = d_\alpha,\end{aligned}$$

the constants μ, d_α are determined so that

$$\mathsf{E}_\theta\{\Phi(\mathbf{x})\} \le \alpha$$

for each $\theta \in \mathrm{H}$. In our case we have $\mathrm{H} = \{n+1\}$ and $\mathrm{A} = \{1, 2, \ldots, n\}$ as possible time instants of the change. For simplicity, we shall assume stochastic independence

among observations. Then, with a probability density function $f(\cdot)$ before a change and with $g(\cdot)$ after a change the ratio, $\lambda(\mathbf{x})$ can be expressed as

$$\lambda(\mathbf{x}) = \frac{f_n(\mathbf{x})}{\max\limits_{j \le j \le n+1} \{f_j(\mathbf{x})\}} = \frac{1}{\max\limits_{1 \le j \le n} \left\{ 1, \prod\limits_{\ell=j}^{n} \frac{g(x_\ell)}{f(x_\ell)} \right\}} \tag{3}$$

if we suppose again $f_{n+1}(\mathbf{x}) > 0$. From (3) we immediately see that

$$\lambda(\mathbf{x}) = 1 \qquad \text{iff} \qquad f_{n+1}(\mathbf{x}) > \max_{1 \le j \le n} \{f_j(\mathbf{x})\}$$

and otherwise

$$\lambda(\mathbf{x}) = \prod_{\ell=j_0}^{n} \frac{f(x_\ell)}{g(x_\ell)} \qquad \text{iff} \qquad f_{j_0}(\mathbf{x}) > \max_{\substack{j \in \{1,2,\ldots,n,n+1\} \\ j \ne j_0}} \{f_j(\mathbf{x})\} .$$

Here, we see a close connection with the Bayesian test that considers the uniform prior distribution $\{p_j\}_{j=1}^{n+1}$. The first kind error of this maximal likelihood ratio test equals

$$P_{\mathrm{H}} \{\mathbf{x} : \lambda(\mathbf{x}) < d_\alpha\} \le P_{\mathrm{H}} \{\mathbf{x} : \lambda(\mathbf{x}) < 1\}$$

because $d_\alpha \le 1$. But, $P_{\mathrm{H}} \{\mathbf{x} : \lambda(\mathbf{x}) < 1\}$ expresses the probability of the first kind error for the Bayesian test having the uniform prior distribution. This fact says that the Bayesian test has larger probability of "false alarms" than the test based on the maximal likelihood ratio. On the other hand, the Bayesian test possesses the advantage of the explicit form over the maximal likelihood ratio test where the problem is to evaluate the values μ, d_α. Further, the probability of the second kind errors behaves quite on the contrary. Let $j \in \mathrm{A}$ (alternative hypothesis), then the probability of second kind error equals

$$\beta_j(\alpha) = P_j \{\mathbf{x} : \lambda(\mathbf{x}) > d_\alpha\} \ge P_j \{\mathbf{x} : \lambda(\mathbf{x}) = 1\} .$$

The right hand side is nothing else but the probability of second kind error of the Bayesian test. From these reasons both the tests are not comparable. If there exists $\alpha \in (0, 1)$ such that

$$\alpha = P_{\mathrm{H}} \{\mathbf{x} : \lambda(\mathbf{x}) = 1\} ,$$

then both the tests would be identical.

3. APPLICATION TO LOCALLY STATIONARY SEQUENCES

A locally stationary sequences is a random sequence that consists of mutually independent weakly stationary parts; i.e. there are time intervals nonoverlapping each other within which the sequence is weakly stationary. We don't know the mean values and spectral density functions and we don't know either, of course, the beginnings and the ends of individual stationary parts. The problem is how to detect possible changes in the mean values or in spectral densities when a realization of a

locally stationary sequence passes from one stationary part to another one. In practice such a sequence we can meet seldomly because of the existence of a temporary section between two adjoing stationary parts. But, like the first approximation to the reality such a statistical model is admissible. As we know nothing about stationary sections the basic idea is to approximate the observed sequence by a suitable stationary sequence, which would be the most "similar" to realizations. Let us assume each stationary section is regular and Gaussian, later the assumption on Gaussiality will be dropped out. A suitable approximation will be the class of autoregressive sequences. At this moment we must solve the problem of the order for autoregressive model and then the problem of estimates of unknown parameters describing an autoregressive model. The order p can be estimated by the classical Akaike criterion and can be adapted with the running time. Let us imagine, at the time instant n we have the observations $x_1, x_2, \ldots, x_n$ belonging to a stationary section. The construction of a suitable autoregressive model will be given in the following steps. First, we must estimate the order p, e. g. as explained earlier. Secondly, we must estimate unknown mean value. Here, the arithmetic mean value seems to be admissible because of strong consistence in every case. In the third step, we need estimates of unknown autoregressive coefficients $a_1, a_2, \ldots, a_p$, σ forming the autoregressive model

$$(x_{\ell+1} - \mu) + \sum_{k=1}^{p} a_k (x_{\ell+1-k} - \mu) = \sigma \ell_{\ell+1}$$

with a standard Gaussian white noise $\{\ell_\ell\}$ and mean value μ. On the basis of observations $x_1, x_2, \ldots, x_n$ we shall construct estimates of unknown covariance function R_k, $k \in \{-p, -p+1, \ldots, -1, 0, 1, \ldots, p\}$. For our purpose it is convenient to choose the estimates of the form

$$\begin{aligned} \hat{R}_k &= \frac{1}{n} \sum_{i=1}^{n-k} (x_{i+k} - \overline{x})\,(x_i - \overline{x}) \qquad k = 0, 1, 2, \ldots, p \\ \hat{R}_k &= \overline{\hat{R}_{-k}} \qquad k = -p, -p+1, \ldots, -1 \end{aligned}$$

and $\overline{x}$ is the arithmetic mean. This form of estimates $\hat{R}_k$ is chosen quite deliberately because then the Toeplitz matrix $\hat{\mathbb{R}} = \left\{\hat{R}_{j-k}\right\}_{j,k=0}^{p}$ is positive definite with probability one. We shall look for an autoregressive model having the same mean value as $\overline{x}$ and the same covariance function as $\hat{R}_j$, $j = \{-p, -p+1, \ldots, 0; \ldots, p\}$. Michálek in [5] showed that this problem can be unambiguously solved among all the autoregressive model of the order p and further, this solution minimizes the so called asymptotic I-divergence rate. The asymptotic I-divergence rate is derived from I-divergence between two stationary Gaussian measure and can serve as a similarity measure among Gaussian stationary measures, in more detail see the monograph by Vajda [8]. For a better orientation we present here the explicit form of the asymptotic I-divergence rate.

Lemma 2. Let $\{P_n, Q_n\}_{n=1}^{\infty}$ be Gaussian stationary measures generated by regular stationary sequences $\{x_j\}_{j=1}^{n}$, $\{y_j\}_{j=1}^{n}$ respectively. Then under the assumption

$P_n \sim Q_n$ for each $n \in \mathcal{N}$ there exists the limit

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}_{P_n} \left\{ \ln \frac{dP_n}{dQ_n} \right\} = \mathrm{AIR}(P|Q) =$$

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

where m_P, m_Q are mean values, φ_P, φ_Q spectral spectral densities and σ_Q^2 is the dispersion of the best one step ahead prediction for the sequence $\{y_j\}_{j=1}^{\infty}$.

Proof. See Vajda [8], pp. 343–344. □

The quantity $\overline{I}(P|Q)$ possesses some properties reminding a metric except the triangular inequality, in general. But, in every case $\overline{I}(P|Q) \geq 0$, $\overline{I}(P|Q) = 0$ iff $P = Q$. Sometimes one can use a symmetric version

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

When the measure P is determined in a some suitable sense by observations $x_1, \ldots, x_n$, let us denote it as $\hat{P}$, and the measure Q is determined by the autoregressive model with coefficients $(1, a_1, a_2, \ldots, a_p, \sigma_p)$, denoted as Q_a, the quantity $\overline{I}(\hat{P}|Q_a)$ can be used for the construction of so called minimal distance estimate $(1, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}_p)$, i.e.

$$\operatorname*{argmin}_{Q_a \in \mathcal{A}_p} \overline{I}(\hat{P}|Q_a) = (1, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p, \hat{\sigma}_p) \qquad (*)$$

where $\mathcal{A}_p$ is the set of all the autoregressive models of the order p.
Choosing $\hat{P} \sim (\overline{x}, \hat{\varphi}(\cdot))$, $\overline{x}$ is the arithmetic mean of $x_1, x_2, \ldots, x_n$ and

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

is a sample spectral density function, the estimates $(*)$ can be given in the explicit form determined by the Yule–Walker equations, i.e.

$$\sum_{j=0}^{p} \hat{a}_j \hat{R}_{k-j} = 0 \qquad k = 1, 2, \ldots, p,$$

$$\sum_{j=0}^{p} \hat{a}_j \hat{R}_j = \hat{\sigma}_p^2$$

(see [5]). In practice, these equations can be evaluated very effectively by the so called Levinson algorithm, see e.g. [7]. In this way we described the procedure of constructing the most similar autoregressive model to the observed sequence. In the case that the observations $x_1, x_2, \ldots, x_n$ are not Gaussian we can consider the obtained estimates $(1, \hat{a}_1, \ldots, \hat{a}_p, \hat{\sigma}_p)$ as a projection of the measure $\hat{P}$ into the

family of Gaussian autoregressive models $\mathcal{A}_p$. This possibility is based on the results due to [4] where it is proved it for Gauss–Markov random fields, which contain autoregressive models as a special case. Using the approximating autoregressive model we can construct a predictor in the form

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

Roughly speaking, if the residua $\varepsilon_\ell = x_\ell - \hat{x}_\ell$ behave "reasonably", i.e. with mean zero and are mutually independent, then there is a hope to assume that stationarity is not violated. By means of the residua $\{\varepsilon_\ell\}$ the detection of changes for a locally stationary sequence can be transformed into the case of the detection of changes in mean value and dispersion for mutually independent observation as described in the previous part.

4. DESCRIPTION OF METHOD

Let us observe a locally stationary sequence without knowing any next information. The method of detection is based on the principle of two windows moving over observations. A detecting window is shorter (about 20 samples) and contains the latest observations. The other window, used for identification contains all the remaining observations. This situation is typical before any change. Under the validity of hypothesis "no change" all the observation inside the identificating window forms a section of a weakly stationary sequence and by use of the previous approximating procedure we can find the most suitable autoregressive model, which order p is determined in every step, e. g. by the Akaike criterion as mentioned above. The estimates $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_p$, $\hat{\sigma}_p$ are evaluated by the Levinson algorithm, which is very fast and stable. Using the one step ahead predictor $\hat{x}_{n+1}$ we obtain the corresponding residuum $\varepsilon_{n+1} = x_{n+1} - \hat{x}_{n+1}$ for each x_{n+1} from a detecting window. When no change occurs inside the detecting window the probability distribution function of $\{\varepsilon_{n+1}/\hat{\sigma}_p\}$ must be very close to $N(0,1)$. Now, if we choose an alternative hypothesis, let's say $N(\mu, \sigma^2)$, we can detect a possible change using the Bayesian approach explained earlier. As for the choice of the parameters μ, σ^2 this problem must be solved individually for every situation. For the better ability of detecting changes one could consider a collection of parameters μ_i, σ_i^2, $i = 1, 2, \ldots, M$, and test the simple hypothesis "no change" simultaneously versus the alternatives $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, M$. In this way, different values for μ can express the different intensity of abrupt jumps in mean value. It would be possible, of course, to consider the shortest detecting window containing the last observation only. In this situation we would make decision based on one residuum only. Such an approach is more or less identical with method proposed by in [2], but this method is critized in Basseville and Benveniste [1] because its reaction on a change is very slow and the proposed test is almost "blind". From this reason one can recommend a longer detecting window, cca 20 – 30 samples. As long as no change is detected, the detecting window is shifted forward without changing its length, i.e. the newest observation is added and the oldest one is dropped out and added to the identificating window. In

this way the windows are moving over observations. As long as a change is detected, the procedure stops and everything starts from the beginning. As we see changes coming very quickly one after the other cannot be well detected in this way. But, this is a disadvantage of every method for detection based on moving windows. We need, in every case a sufficiently long time interval for the good identification of a stationary section before a change. In practice, it seems to be reasonable to confirm every detection of change by detections in next detecting windows. This measure decreases the number of false alarms. On practical experience, the rule 3 is recommended, i.e., a change is detected if this change is confirmed at least by 3 detections coming one after the other.

5. RESULTS OF SIMULATIONS

The proposed method was verified by simulated data, which consisted of autoregressive models. Our method was compared with the classical method based on the χ^2-statistics, i.e. the sum of squared rezidua belonging to a detecting window must be distributed according to χ^2-distribution under the validity of the hypothesis "no change". For our purposes we constructed realizations of locally weakly stationary sequences with two sections having the length 500 observations, i.e. the change was starting with the sample 501 in every case. We compared autoregressive sequences from the order $p = 0$ up to $p = 3$. The detection of change was based on the rule "3" to decrease the number of false alarms. The obtained realizations are depicted in Figures 1, 2, 3, 4, 5. We shall present all simulated models:

		1. section	2. section
$p = 0$	Series A + B	$N(0,1)$ – white noise	$N(1,1)$ – white noise
	Series A + C	$N(0,1)$ – white noise	$N(0,2)$ – white noise
$p = 1$	Series 1 a 2	$x_{n+1} + 0,6\,x_n = e_{n+1}$	$x_{n+1} + 0,1\,x_n = 2e_{n+1}$
$p = 2$	Series 5 a 6	$x_{n+1} - 0,3\,x_n - 0,3\,x_{n-1} = e_{n+1}$	$x_{n+1} - 0,5\,x_n - 0,5\,x_{n-1} = 2e_{n+1}$
$p = 3$	Series 9 a 10	$x_{n+1} - 0,85\,x_n + 0,25\,x_{n-1} - -0,06\,x_{n-2} = e_{n+1}$	$x_{n+1} - 1,33\,x_n + 0,45\,x_{n-1} + +0,04\,x_{n-2} = 2e_{n+1}$

During simulations the parameters standing in the Bayesian decision rule were changed to find out their influence on the behaviour of the test. We also carried out experiments with the length of the detecting window. The obtained results are summarized in the following Table 1.

Table 1.[1]

				Bayes test		χ^2-tests		
	L	μ	σ^2	false alarms	delay	false alarms	5 % delay	2 %
Series A + B	20	0,00	4,00	0	5	0	24	∞
	20	0,00	2,00	1	5	0	24	∞
	20	0,00	3,00	1	5	0	24	∞
	20	0,00	5,00	0	5	0	24	∞
	20	0,00	9,00	0	∞	0	24	∞
	20	0,00	4,00	0	5	0	24	∞

[1] ∞ ... change was not detected

				Bayes test		χ^2-tests		
	L	μ	σ^2	false alarms	delay	false alarms	5 % delay	2 %
Series A + C	20	0,00	4,00	0	5	0	24	∞
	20	0,10	5,00	0	5	0	24	∞
	20	0,10	4,00	0	5	0	24	∞
	25	0,10	4,00	0	5	0	24	∞
	15	0,10	5,00	0	5	0	∞	∞
	20	0,00	4,00	0	5	0	24	∞
Series 1 a 2	20	1,00	9,00	2	4	1	6	7
	20	1,00	16,00	0	6	1	6	7
	20	1,00	25,00	0	6	1	6	7
	20	1,00	36,00	0	6	1	6	7
	20	1,00	49,00	0	6	1	6	7
	20	1,00	100,00	0	6	1	6	7
	20	1,00	400,00	0	6	1	6	7
	20	1,00	900,00	0	6	1	6	7
	20	1,00	36,00	0	4	0	5	5
	20	1,00	16,00	0	4	0	5	5
	20	1,00	9,00	0	4	0	5	5
	20	1,00	4,00	1	4	0	5	5
	50	1,00	3,00	2	0	3	∞	∞
	30	1,00	3,00	2	0	2	∞	∞
	20	1,00	3,00	3	0	1	7	8
	15	1,00	3,00	3	0	0	7	8
	10	1,00	3,00	2	0	1	9	10
	5	1,00	3,00	3	0	1	14	14
	20	1,00	10000	0	11	1	6	7
Series 5 a 6	30	1,00	9,00	0	∞	1	∞	∞
	20	1,00	0,04	2	∞	1	∞	∞
	20	1,00	0,01	1	27	1	∞	∞
	20	0,50	0,04	1	6	1	∞	∞
	20	1,00	0,04	0	35	1	∞	∞
	20	1,00	0,01	0	6	1	∞	∞
	20	1,00	36,00	0	∞	1	∞	∞
	20	1,00	9,00	0	∞	1	∞	∞
	20	1,00	0,04	0	1	0	13	14
	20	0,50	0,04	2	6	0	13	14
	20	1,50	0,04	0	1	0	13	14
	20	1,50	1,00	2	0	0	13	14
	20	2,00	4,00	0	0	0	13	14
	20	0,50	1,00	6	0	0	13	14
	20	1,00	90,00	0	16	1	13	14
	20	1,00	2,00	3	0	0	∞	∞
	20	1,00	4,00	2	0	1	∞	∞
	20	1,00	9,00	1	0	∞	∞	∞
Series 9 a 10	30	1,00	9,00	0	0	0	1	1
	20	1,00	90,00	0	0	0	1	1
	20	1,00	9,00	0	0	0	1	1
	20	1,00	25,00	0	0	0	1	1

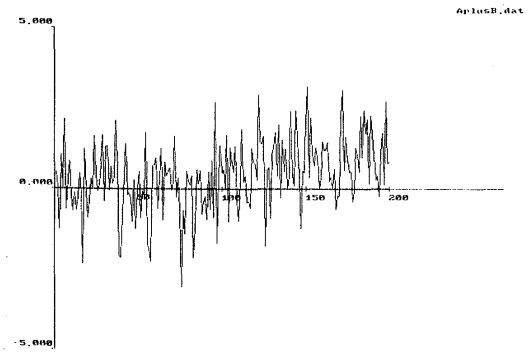

Fig. 1.

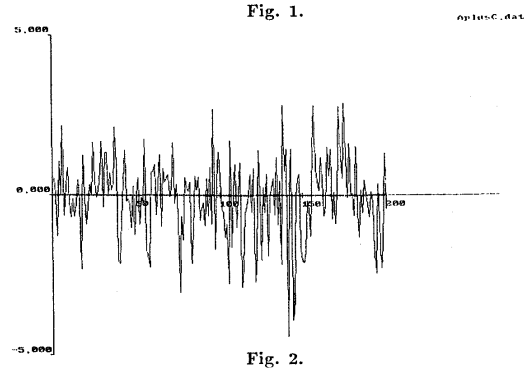

Fig. 2.

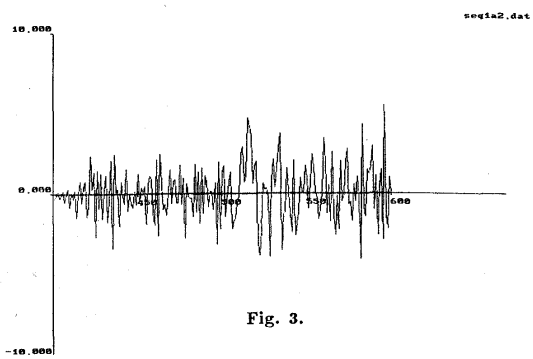

Fig. 3.

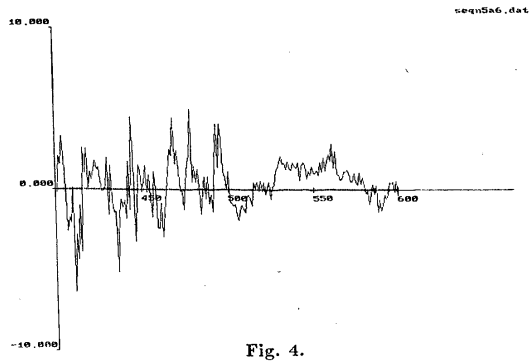

Fig. 4.

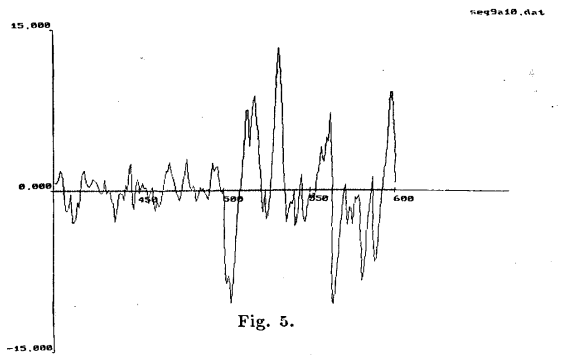

Fig. 5.

6. CONCLUSIONS

The results obtained by simulations show that the Bayesian test is in general better than the test based on χ^2-statistic. Although the Bayesian test has a larger number of false alarms, but the difference is not so big, its ability for detecting changes is much more better than the χ^2-test, which is in some situations actually "blind". The conclusions following from the simulations can be summarized in the rule the greater σ the smaller number of false alarms. We can recommend quite unambiguously for practice: in case we don't know anything about the "intensity" of a possible change there is necessary to use a set of parallel Bayesian tests with different μ's and σ's.

(Received June 11, 1992.)

REFERENCES

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

[2] D. H. Crowell, R. J. Jones and L. E. Kapuniai: A method for detecting change in a time series applied to newborn EEG. Electroencephalography and Clinical Neurophysiology *27* (1969), 434–440.

[3] N. Kligiene and L. A. Telksnys: Methods of detecting instants of change of random process properties. Autom. Remote Control *44* (1983), 1241–1283.

[4] H. Künsch: Thermodynamics and statistical analysis of Gaussian random fields. Z. Wahrsch. verw. Gebiete *58* (1981), 407–421.

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

[6] I. V. Nikoforov: Sequential Detection of Abrupt Changes in Time Series Properties (in Russian). Nauka, Moscow 1983.

[7] O. Šefl and I. Vajda: Spectrum oriented source coding theory. Kybernetika *23* (1987), 458–475.

[8] I. Vajda: Theory of Statistical Inference and Information. Kluwer, Dordrecht – Boston 1990.

[9] A. J. Willsky: A survey of design methods for failure detection in dynamic systems. Automatica *12* (1976), 601–611.

RNDr. Jiří Michálek, CSc., Ing. Jiří Skřivánek, Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 18208 Praha 8. Czech Republic.