

A SIMPLE TEST FOR TRACKING THE REGULAR BEHAVIOUR OF A TECHNOLOGICAL PROCESS *

Jiří MICHÁLEK

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
182 08 Prague 8, Pod vodárenskou věží4, Czech Republic

Abstract. In the presented paper a method for treating a random signal bearing an information about the behaviour of a technological process is given. The main goal of the given method is to remove possible failures arising in analog sensors, which yield nonstationary behaviour of an observed signal. Then the smoothed signal is tested by a suitable test described in the paper for the regular or irregular behaviour of a technological process. One understands by the regular behaviour of a technological process that within prescribed bounds.

Key words: detection of changes, smoothing, test, confidence interval.

1. Introduction. The paper deals with the construction of a simple test for tracking the regular behaviour of a technological process, which is measured via an analog sensor. The question of a such test was provoked by practice and from the theoretical point of view this problem belongs to the detection of changes in random sequences. A signal coming from a sensor can be divided into two parts, namely

$$x(t) = m(t) + \xi(t),$$

* This paper is supported by the grant No. 201/93/0233 of the Grant Agency of the Czech Republic.

where $\{m(\cdot)\}$ is a technological process and $\{\xi(\cdot)\}$ is noise. Because of errors produced by a sensor or transmission channels this noise is nonstationary, in general. For simplicity, we shall assume $\{\xi(\cdot)\}$ being mutually independent. It is also convenient to understand the signal $m(\cdot)$ as a deterministic part of $\{x(\cdot)\}$. In the most of cases a mathematical model describing the evolution of $\{m(\cdot)\}$ need not be exactly known. One can suppose that $\{m(\cdot)\}$ is sufficient smooth, i.e., the existence of derivatives is quite natural. The technological process behaves regularly, roughly speaking where no abrupt changes are present. To specify this property it seems to be quite reasonably, e.g., to check the first and the second derivatives. If both the quantities are within limits prescribed in advance the signal $\{m(\cdot)\}$ behaves regularly. As we cannot observe the signal $\{m(\cdot)\}$ directly, the considered test for regularity must be carried out in two steps.

1. Smoothing $\{x(\cdot)\}$.
2. Testing regularity of smoothed $\{x(\cdot)\}$.

In order to remove possible nonstationarities in the noise $\{\xi(\cdot)\}$ it is necessary to treat data $\{x(\cdot)\}$ before testing them. Nonstationarities are understood as changes in the first and the second moments, i.e., an abrupt change in mean value and in dispersion, too. The detection of changes is based on a one step ahead predictor and on the comparison of residuals. After detecting a change in mean value or in dispersion the data $\{x(\cdot)\}$ are modified. The output from this procedure will be denoted by $\{\hat{x}(\cdot)\}$.

The second step is the test of regularity applied to $\{\hat{x}(\cdot)\}$. The information on the behaviour of the first and second derivatives is obtained in $\{\hat{x}(\cdot)\}$ via its first and second differences. It means a test statistic must be based on $\Delta_1 \hat{x}(t)$ and $\Delta_2 \hat{x}(t)$, where $\Delta_1 \hat{x}(t) = \hat{x}(t + \Delta) - \hat{x}(t)$, $\Delta_2 \hat{x}(t) = \hat{x}(t + 2\Delta) - 2\hat{x}(t + \Delta) + \hat{x}(t)$ where Δ is a sample period.

Here, an unknown parameter is of a vector type

$$\theta = \{ \dot{m}(t), \dot{m}(t+\Delta), \dots, \dot{m}(t+(n-1)\Delta), \\ \ddot{m}(t), \dots, \ddot{m}(t+(n-2)\Delta) \}$$

with dimensionality $2n-3$ if n is the number of observations. The hypothesis of regularity is valid if

$$|\dot{m}(t+j\Delta)| \leq \varepsilon_1$$

for each $j \in \{1, 2, \dots, n-1\}$ together with

$$|\ddot{m}(t+j\Delta)| \leq \varepsilon_2$$

for each $j \in \{1, 2, \dots, n-2\}$. The alternative hypothesis is complementary. The test is based on likelihood ratio under the Gaussianity of all random variables. The form of this test is not simple because of stochastic dependence among $\{\Delta_1 \hat{x}(t+j\Delta)\}$ and $\{\Delta_2 \hat{x}(t+k\Delta)\}$. In case we drop the fact of dependence among differences the test will become much easier. It results a sequence of simple inequalities

$$|\Delta_1 \hat{x}(t+js)| \leq \delta_1^*$$

for each $j \in \{1, 2, \dots, n-1\}$ and

$$|\Delta_2 \hat{x}(t+ks)| \leq \delta_2^*$$

for each $k \in \{1, 2, \dots, n-2\}$, where δ_1^* , δ_2^* are derived from significance level. If at least one of the previous inequalities is not valid regularity is rejected. As we work with random variables we can accept a wrong decision. From this reason in order to diminish the probability of a false alarm the first detection of irregular behaviour must be immediately confirmed in the next step. Otherwise we speak about latent irregularity.

2. Assumptions and model. Let us assume we receive a one-dimensional random signal $x(t)$, $t \in \{\Delta, 2\Delta, 3\Delta, \dots\}$ with a sample period Δ which comes from an analog sensor and bears information about the behaviour of a technological process. For simplicity we can also put $\Delta = 1$. We will assume that the signal $x(t)$ can be divided into two parts, i.e.,

$$x(t) = m(t) + \xi(t), \quad t = 1, 2, \dots, n,$$

where $m(t)$ is an unobserved technological process and $\xi(t)$ is a white noise (uncorrelated), which is not weakly stationary, in general. This nonstationary behaviour can be caused by a change in mean value or in the level of noise, i.e., in a change of second moments. Such a change can be expected as known on the basis of practice. If a sensor works well the level of noise is usually very small. One can also accept the Gaussianity of $\{\xi(t)\}$ hence $\{\xi(\cdot)\}$ is an independent white noise. Under the assumption that $\{m(\cdot)\}$ is also a random process we will assume mutual independence between $\{m(\cdot)\}$ and $\{\xi(\cdot)\}$. Further, as for the behaviour of the process $\{m(\cdot)\}$ we can assume its sufficient smoothness, i.e., the existence of the first two derivatives at least because it is obtained by sampling a physical variable (like pressure, temperature). We will demand by the original analog signal $\{m(\cdot)\}$ the existence of the first two derivatives, i.e., $\dot{m}(t)$, $\ddot{m}(t)$ exist. It is usually very difficult to construct a model describing the behaviour of $\{m(\cdot)\}$. In many cases a suitable mathematical model need not even exist because of nonstationary behaviour of $\{m(\cdot)\}$ in nonstandard situations. Such situations must be in practice detected because their occur means irregular behaviour of the technological process $\{m(\cdot)\}$. Under the regular behaviour of a technological process, roughly speaking one can imagine such a behaviour where the process is running within technological bounds prescribed in advance. A precise description will be given later.

3. Detecting changes. Before testing the regular behaviour of a technological process one must first detect abrupt changes in the behaviour of the observed process. These changes are caused by nonstationarities of a random noise, i.e., by changes in mean value or in dispersion of noise. Both these changes can occur simultaneously too. As the whole procedure must run in the on-line regime and hence is limited by time, the method of a sliding window was suggested. On the basis of practice and simulations the length of a sliding window is recommended to be 20–30 observations maximally. Within this sliding window we assume under the condition “no change occurred” that the observed random sequence can be understood to be stationary. Since it is very difficult to construct a stationary model describing precisely the behaviour of the observed random sequence within the sliding window an idea of “approximation” was used. We wish to find an approximating stationary sequence among autoregressive stationary sequences of a given order that could be the most similar to the actual sequence within the sliding window. The similarity is measured via asymptotic I -divergence rate (for definitions and basic properties see Vajda, 1989). When f_1, f_2 are the corresponding spectral densities on $(-\pi, \pi)$, then $\text{AIR}(f_1, f_2)$ can be evaluated as

$$\text{AIR}(f_1, f_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{(f_1 - f_2)^2}{f_1 f_2}.$$

If the spectral density is substituted by a suitable estimate and f_2 is running over the class of autoregressive stationary sequences of a prescribed order then the minimization of AIR gives the Yule–Walker estimates for the autoregressive model which is in this sense the most similar to the actual random sequence. The task of minimization can be solved effectively by the so called Levinson algorithm very suitable for computing; for details see Michálek, 1990. In practice it is sufficient to

use the first or second order of autoregressive models because longer models suffer from the inertia. When the most similar autoregressive model is chosen on its basis we evaluate a one step ahead prediction, let say $\tilde{x}(n+1)$, which is based on the information from the sliding window up to the time n . Then, this prediction is compared with the actual value $x(n+1)$ and the difference between them is evaluated, i.e.,

$$|x(n+1) - \tilde{x}(n+1)| \leq \varepsilon.$$

If the difference is smaller than a given ε in the absolute value then no change is detected. Otherwise, a latent failure is claimed. In order to diminish the probability of a false alarm, the detection of a change must be confirmed in the immediate future. In practice it means the failure must be detected at least in 3 of 5 observations.

If within the moving window no change is detected we can carry out the following smoothing procedure. If any change is detected then a loss of information is coming and then there are two possibilities. First one is to wait for a spontaneous recovery of a sensor, second one is to change completely a false sensor by human intervention.

4. Smoothing procedure. Since the process $\{m(\cdot)\}$ is not directly observed we must (in the case no change detected in the moving window) smooth the observed process $\{x(\cdot)\}$ to remove the additive noise $\{\xi(\cdot)\}$. Within the sliding window we will assume no detected change, i.e.,

$$E\{\xi(t)\} = 0, \quad D\{\xi(t)\} = \sigma^2$$

for each t from the sliding window. The level of noise dispersion σ^2 is not usually known. It is very difficult to look for a convenient mathematical model characterizing the technological process locally within every sliding window. As we assume

that $\{m(\cdot)\}$ and $\{\xi(\cdot)\}$ are mutually uncorrelated second order sequences we can consider the optimal filtering procedure with respect to minimization in the square mean. Using the basic relation

$$x(t) = m(t) + \xi(t),$$

one can in an easy way prove that the basic relation describing the optimal filtering procedure is given by

$$\hat{m}(t | t) = \left(1 - \frac{\sigma^2}{D \tilde{x}(t)}\right) x(t) + \frac{\sigma^2}{D \tilde{x}(t)} \hat{m}(t | t - 1),$$

where $\hat{m}(t | t)$ is the optimal filtered value of $m(t)$ at the time t under the history up to the time t , $\hat{m}(t | t - 1)$ is the optimal predicted value of $m(t)$ under the history up to the time $t - 1$, $\tilde{x}(t)$ is an innovation sequence and $D \tilde{x}(t)$ is its dispersion. In order to use this relation we must know $\hat{m}(t | t - 1)$. It is nothing else but to know the evaluation of $m(\cdot)$ what is the problem mentioned above. Further, we must also know the optimal value of the smoothing factor $\alpha_t = \frac{\sigma^2}{D \tilde{x}(t)}$ which can change in time. With respect to the length of a sliding window and real situation it is quite natural to consider the simplest case. Within a sliding window we shall assume that α_t is very slowly changing, i.e., $\alpha_t = \alpha$, where α must be between 0 and 1. As for the prediction $\hat{m}(t | t - 1)$ it is realistic to substitute its value by the latest filtered value, i.e.,

$$\hat{m}(t | t - 1) = \hat{m}(t - 1 | t - 1).$$

This situation occurs if we consider the following model for $\{m(\cdot)\}$

$$m(t + \Delta) = m(t) + \eta(t + \Delta),$$

where $\{m(\cdot)\}$ and $\{\eta(\cdot)\}$ are mutually uncorrelated and $E\{\eta(\cdot)\} = 0$ for every t . It is evident if the sample period Δ is sufficiently long such a model can work. In the case of a

shorter period Δ it is possible to assume a somewhat complicated model based on analogy with Taylor's expansion

$$m(t + \Delta) = m(t) + \dot{m}(t) + \eta(t + \Delta)$$

with

$$\dot{m}(t + \Delta) = \dot{m}(t) + \omega(t + \Delta).$$

In order to choose an optimal value for α within a sliding window we consider to minimize the errors of prediction one step ahead, i.e., α is chosen to minimize

$$\sum_{t \in W} (x(t) - \hat{m}_\alpha(t, t))^2.$$

As $x(t) = m(t) + \xi(t)$ then this criterion is de facto also the criterion of minimization of filtration errors. The basic relation for optimal square mean filtration can be simply rewritten as

$$\hat{m}(t | t) = m(t) + \xi(t) - \alpha \tilde{x}(t).$$

Denoting $e(t) = \xi(t) - \alpha \tilde{x}(t)$ then by the optimal choice for α we have $\{e(\cdot)\}$ is a white noise with

$$\mathbf{E}\{e(t)\} = 0, \quad D\{e(t)\} = (1 - \alpha)\sigma^2.$$

From this fact we have immediately $\hat{m}(t | t)$ is unbiased, i.e.,

$$E\{\hat{m}(t | t)\} = m(t).$$

On the other side the properties of $\{e(\cdot)\}$ can be used for checking the quality of smoothing because the smoothing procedure is optimal if and only if $\{e(t)\}$ is a white noise with moments described above.

5. Testing regularity. First, we must define precisely what is the regular behaviour of a technological process. It

means there are bounds δ_1, δ_2 derived from the properties of an observed technological process such that

$$\begin{aligned} |\dot{m}(t)| &< \delta_1, \\ |\ddot{m}(t)| &< \delta_2. \end{aligned}$$

There is no problem to show that after smoothing it holds

$$\begin{aligned} \mathbf{E}\{\Delta_1 \hat{m}(t | t)\} &= \Delta_1 m(t) = \dot{m}\Delta + o(\Delta); \\ \mathbf{E}\{\Delta_2 \hat{m}(t | t)\} &= \Delta_2 m(t) = \ddot{m}\Delta^2 + o(\Delta^2). \end{aligned}$$

Here, Δ_i is the operator of the i -th difference. When we neglect the remainders in Taylor's expansions then we can understand the problem of regular behaviour as a decision problem of testing hypothesis where the number of observations is given by the length of a sliding window, an unknown parameter θ is of a vector type

$$\theta = \{\dot{m}(t_1), \dot{m}(t_1 + \Delta), \dots, \dot{m}(t_1 + (n-1)\Delta), \\ \ddot{m}(t_1), \ddot{m}(t_1 + \Delta), \dots, \ddot{m}(t_1 + (n-2)\Delta)\}$$

with the dimension $2n-3$; n is the length of a sliding window. The hypothesis H is formed by the part of the $(2n-3)$ -dimensional Euclidean space.

$$H = \prod_{i=1}^{n-1} \langle -\delta_1, \delta_1 \rangle \times \prod_{i=1}^{n-2} \langle -\delta_2, \delta_2 \rangle,$$

where δ_1, δ_2 are given as technological limits. The alternative hypothesis is complementary, i.e.,

$$A = \mathbf{E}_{2n-3} - H.$$

The observations we have at disposal are the smoothed values

$$\Delta_1 \hat{m}(t | t), \quad \Delta_2 \hat{m}(t | t),$$

where t is running over a sliding window. If we accept the Gaussianity of $\{\hat{m}(t | t)\}$ then the problem of testing the regularity of $\{m(\cdot)\}$ becomes that of the behaviour of mean value of a multidimensional Gaussian random variable, which is formed by dependent coordinates, in general. One can easily show that the covariance structure is given by the following relations

$$\begin{aligned}
 & \mathbb{E}\{\Delta_1 \hat{m}(t | t) \Delta_1 \hat{m}(t + j\Delta | t + j\Delta)\} \\
 &= \begin{cases} (1 - \alpha) \sigma^2, & j = 0, \\ -(1 - \alpha) \sigma^2, & j = \pm 1, \\ 0, & \text{otherwise;} \end{cases} \\
 & \mathbb{E}\{\Delta_2 \hat{m}(t | t) \Delta_2 \hat{m}(t + j\Delta | t + j\Delta)\} \\
 &= \begin{cases} -3(1 - \alpha) \sigma^2, & j = 0, \\ (1 - \alpha) \sigma^2, & j = 1, \\ -(1 - \alpha) \sigma^2, & j = -1, -2, \\ 0, & \text{otherwise;} \end{cases} \\
 & \mathbb{E}\{\Delta_2 \hat{m}(t | t) \Delta_1 \hat{m}(t + j\Delta | t + j\Delta)\} \\
 &= \begin{cases} -3(1 - \alpha) \sigma^2, & j = 0, \\ -(1 - \alpha) \sigma^2, & j = 1, 2, \\ (1 - \alpha) \sigma^2, & j = 0, \\ 0, & \text{otherwise;} \end{cases} \\
 & \mathbb{E}\{\Delta_2 \hat{m}(t | t) \Delta_2 \hat{m}(t + j\Delta | t + j\Delta)\} \\
 &= \begin{cases} 6(1 - \alpha) \sigma^2, & j = 0, \\ -4(1 - \alpha) \sigma^2, & j = \pm 1, \\ (1 - \alpha) \sigma^2, & j = \pm 2, \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}$$

From this fact one can immediately see that the corresponding covariance matrix \mathbf{C} can be expressed as

$$\mathbf{C} = (1 - \alpha) \sigma^2 \mathbf{K}_{2n-3},$$

where \mathbf{K}_{2n-3} is a fixed matrix independent on α and σ^2 , it is dependent on the number of observations only.

The solution of the question testing mean value of multi-dimensional Gaussian variable by an unknown covariance matrix is given by the Hotelling T^2 -statistic and this solution can be found almost in every textbook of mathematical statistics. But, the corresponding confidence intervals are given by ellipsoids what is unsuitable for a practical situation where we wish to decide about every coordinate in mean value individually. From this reason we should have a rectangular confidence interval. This is evidently satisfied in the case of independent observations but it is not our case. We can find an answer for our problem in the paper [Šidák]. From this article we can use the following

Theorem (Šidák, 1967). *Let $\mathcal{X} = (\mathcal{X}_\infty, \mathcal{X}_\epsilon, \dots, \mathcal{X}_\parallel)$ be the vector of random variables having k -dimensional normal distribution with zero means, arbitrary variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ and arbitrary correlation matrix $\mathbf{R} = \{\rho_{ij}\}$. Then, for any positive numbers c_1, c_2, \dots, c_k*

$$\mathbf{P} \left\{ \bigcap_{i=1}^k |X_i| \leq c_i \right\} \geq \prod_{i=1}^k \mathbf{P} \{|X_i| \geq c_i\}.$$

If we wish to construct a rectangular confidence interval with $k = 2n - 3$ and with the significance level $1 - p$ we determine $c_1, c_2, \dots, c_{2n-3}$ such that the inequality

$$\prod_{i=1}^{2n-3} \mathbf{P} \{|X_i| \leq c_i\} = 1 - p$$

holds. Since we do not know a precise value of dispersion σ^2 , we can use for our situation a modification of Theorem 1 also presented in Šidák, 1967 (see Theorem 2, p. 629). Using these results we can construct a rectangular confidence interval such

that

$$\begin{aligned} \Delta_1 \hat{m}(t + j\Delta | t + j\Delta) - c_\alpha s &\leq \dot{m}(t + j\Delta) \\ &\leq \Delta_1 \hat{m}(t + j\Delta | t + j\Delta) + c_\alpha s, \\ \Delta_2 \hat{m}(t + j\Delta | t + j\Delta) - c_\alpha s &\leq \ddot{m}(t + j\Delta) \\ &\leq \Delta_2 \hat{m}(t + j\Delta | t + j\Delta) + c_\alpha s, \end{aligned}$$

where c_α is the quantile of Student distribution and s is a suitable estimate of dispersion σ^2 in a sliding window. In other words speaking if we find a rectangular confidence interval with the significance level $1 - p$ then this interval is also a confidence interval for arbitrary dependent variables with not less significance level. In practice it means we can decide about every coordinate $\dot{m}(t + j\Delta)$, $\ddot{m}(t + j\Delta)$ individually. The hypothesis of regularity is not rejected if for each j from a sliding window

$$\begin{aligned} |\Delta_1 \hat{m}(t + j\Delta | t + j\Delta)| &\leq \delta_1^*, \\ |\Delta_2 \hat{m}(t + j\Delta | t + j\Delta)| &\leq \delta_2^*, \end{aligned}$$

where δ_1^* , δ_2^* are properly chosen.

6. Conclusion. The presented method for treating a technological signal is simple. But it is limited by the sample period Δ . In practice we can approximately consider $\Delta = 1$ sec to have a sufficient time interval for carrying out one step of the proposal procedure. This method was tested both by simulations and in practice, too. Its results are quite satisfactory.

REFERENCES

- Michálek, J. (1990). Yule-Walker estimates and asymptotic I -divergence rate. *Problems of Control and Information Theory*, 19(5-6), 387-398.

- Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Statist. Assoc.*, **62**, 626–633.
- Vajda, I. (1989). *Theory of Statistical Inference and Information*. Kluwer Academic Publishers Dordrecht, Boston, London.

Received November 1993

J. Michálek graduated from the Faculty of Mathematics and Physics of the Charles University of Prague in 1966. In 1968 he joined the Department of Information Theory in the Institute of Information Theory and Automation of Czechoslovak Academy of Sciences where he has been working up to nowadays. He received the Degree of Candidate of Mathematical Sciences in 1975. His professional interests are concentrated on statistical problems in random processes.