SEGMENTATION AND CLASSIFICATION OF THE BLOOD FLOW SIGNALS

Abramovich M.S., Zhuk E.E.

Department of Mathematical Modeling and Data Analysis, Belarusian State University,4 Fr.Skariny av., 220050 Minsk, Belarus. E-mail: *abramovichms@bsu.by*

Abstract. The coronary ischemic disease is characterized by change of blood flow turbulence and, as a result, by appearance of high-frequency sounds that are caused by stenosis of arteries. The most informative part of this signal, which is used for diagnostics of the coronary ischemic disease, is the diastole. In order to extract the diastole from the entire signal, the moments, which correspond to the beginning and the end of a flap of the mitral valve damper, have to be determined. These moments can be detected as change-points of probabilistic characteristics of the signal, especially, as change-points of the spectral density of the signal. A training sample is formed from the extracted stationary fragments of the diastole for the healthy men and men with the coronary ischemic disease. The training sample is used for construction of a decision rule. In the paper the method for segmentation of the blood flow signal based on a spectral test for change-points detection is considered. The proposed decision rule is based on discriminant analysis of parameters of autoregressive models, which describe stationary fragments of the diastole.

Segmentation of the blood flow signals

Consider a mathematical model of a change-point in a blood flow signal. Let $X_1(t), t \in Z$ be a stationary signal with zero mean and the spectral density $S_1(\lambda), \lambda \in [-\pi, \pi]$, $X_2(t)$, $t \in Z$ be stationary signal independent from $X_1(t)$ with zero mean and the spectral density $S_2(\lambda)$ that is different from $S_1(\lambda)$: $S_1(\cdot) \neq S_2(\cdot)$. The registered signal $X = \{x_t, t = \overline{1,T}\}$ of length T has a change-point at an unknown time moment $t_0 \in \{\tau_-, \tau_- + 1, \ldots, \tau_+, T+1\}$ that means an abrupt change of a spectral density:

$$x_t = \begin{cases} X_1(t), t \in \{1, \ldots, t_0 - 1\}, \\ X_2(t), t \in \{t_0, t_0 + 1, \ldots, T\}, \end{cases} \tag{1}$$

where $1 < \tau_- < \tau_+ < T$ are limit values that are known a priory.

If $t_0 = T + 1$, then the signal $x_t = X_1(t)$ is homogeneous.

For $\tau \in \{\tau_-, \tau_- + 1, \ldots, \tau_+\}$ divide the observed signal into two fragments $X_1 = \{x_1, \ldots, x_{\tau-1}\}, X_2 = \{x_\tau, \ldots, x_T\}$ of lengths T_1 and T_2 correspondingly: $T_1 = \tau - 1$; $T_2 = T - \tau + 1, T_1 + T_2 = T$.

Under fixed τ define hypotheses:

$H_0 : t_0 = T + 1$, the signal is homogeneous;

$H_1 : t_0 = \tau$, the signal contains the change-point at the time moment τ.

Introduce a non-parametric estimator of the spectral density $\hat{S}_i(\lambda), \lambda \in [-\pi, \pi]$, $i = 1,2$ that is calculated from the fragment X_i [1]:

$$\hat{S}_i(\lambda) = \frac{2\pi}{T_i}\sum_{n=1}^{T_i-1} W^{(T_i)}\left(\lambda - \frac{2\pi n}{T_i}\right)\hat{I}^{(T_i)}\left(\frac{2\pi n}{T_i}\right), \tag{2}$$

$$W^{(T_i)}(\alpha) = K_{T_i}\sum_{j=-\infty}^{\infty} W\left(K_{T_i}(\alpha + 2\pi j)\right),$$

where K_{T_i} is an smoothing parameter, $\hat{I}^{(T_i)}(\lambda) = \frac{1}{2\pi T_i}\left|\hat{d}_i(\lambda)\right|^2$ is the periodogram, $\hat{d}_i(\lambda) = \sum_{t=T_{i-1}+1}^{T_{i-1}+T_i} x_t \exp\{i^*\lambda t\}$ is the Fourier transform of the fragment X_i, i^* is the imaginary unit, $W(\alpha), -\infty < \alpha < \infty$ is an weight function, $\int_{-\infty}^{\infty} W(\alpha)d\alpha = 1, W(-\alpha) = W(\alpha)$.

Define a statistic for change-point detection based on difference between the spectral density estimators that are calculated from the fragments X_1, X_2:

$$\hat{\Delta}^2(\tau) = \sum_{s=1}^{m}\left(\hat{S}_1(\lambda_s) - \hat{S}_2(\lambda_s)\right)^2 \Big/ \sum_{s=1}^{m}\left(\hat{S}_1^2(\lambda_s) + \hat{S}_2^2(\lambda_s)\right), \tag{3}$$

where $\{\lambda_1,\ldots,\lambda_m\} \in [-\pi,\pi]$ is a fixed set of frequencies such that $\sum_{s=1}^{m}\left(\hat{S}_1(\lambda_s) - \hat{S}_2(\lambda_s)\right)^2 > 0$. The statistic (3) is similar to the statistic from [2] for bivariate time series.

It can be shown if the estimators of the spectral densities $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda)$ are consistent, then the statistic (3) converges on probability to a limiting functional that attains the maximal value when the change-point is present: $\tau = t_0$.

Introduce an auxiliary statistic:

$$\breve{\Delta}_1^2(\tau) = \sum_{s=1}^{m}\left(\hat{S}_1(\lambda_s) - \hat{S}_2(\lambda_s)\right)^2 \Big/ \sum_{s=1}^{m}\left(S_1^2(\lambda_s) + S_2^2(\lambda_s)\right).$$

Let $Q = \int_{-\infty}^{\infty} W^2(\alpha)d\alpha, c_i(u)$ be the covariance function of the fragment X_i.

Theorem 1 Let the following conditions hold $(i = 1,2)$:

C1) $\sum_{u=-\infty}^{\infty}|u||c_i(u)| < \infty$; C2) $\int_{-\infty}^{\infty}|\alpha||W(\alpha)|d\alpha < \infty$; C3) $T_i \to \infty, K_{T_i} \to \infty, \frac{K_{T_i}}{T_i} \to 0$; $m = O\left(T^{1-\gamma}\right), 0 < \gamma < 1, m \to \infty$; C4) Under asymptotics C3) lengths of the fragments are comparable values.

Then under true H_0 a statistic $F = \left(\sum_{s=1}^{m}\left(\hat{S}_1^2(\lambda_s) + \hat{S}_2^2(\lambda_s)\right)\hat{\Delta}_1^2(t_0) - A_m\right)\Big/B_m$

has asymptotically normal probability distribution $N_1(0,1)$, where

$$A_m = \sum_{s=1}^{m} a_s; B_m = \sqrt{\sum_{s=1}^{m} b_s},$$

$$a_s = 2\pi\left(\frac{K_{T_1}}{T_1}+\frac{K_{T_2}}{T_2}\right)(S_1(\lambda_s))^2 Q + o(1),\quad b_s = 8\pi^2\left(\frac{K_{T_1}}{T_1}+\frac{K_{T_2}}{T_2}\right)^2 (S_1(\lambda_s))^4 Q^2 + o(1).$$

On the base of Theorem 1 construct a test for change-point detection with significance level ε:

$$\text{decide}\begin{cases} H_0, \hat{\Delta}_1^2(\tau)\le\delta_\varepsilon, \\ H_{1\tau}, \hat{\Delta}_1^2(\tau)>\delta_\varepsilon, \end{cases} \tag{4}$$

$$\delta_\varepsilon = \pi\left(K_{T_1}/T_1 + K_{T_2}/T_2\right)\left(\sqrt{2}\Phi^{-1}(1-\varepsilon)+1\right)Q,$$

where $\Phi^{-1}(p)$ is the quantile of the standard normal probability distribution with the disribution function $\Phi(\cdot)$.

It can be proven that the statistic $\hat{\Delta}^2(\tau)$ along with the statistic $\hat{\Delta}_1^2(\tau)$ can be used as statistical measure of difference between the hypotheses H_0, $H_{1\tau}$.

Theorem 2 Let the conditions C1-C4 of Theorem 1 hold. Then under the true alternative hypothesis H_{1t_0} the statistic $\hat{\Delta}_1^2(t_0)$ has the asymptotically normal probability distribution:

$$L\left\{\frac{\hat{\Delta}_1^2(t_0)-a(T_1,T_2)}{\sqrt{B(T_1,T_2)}}\right\}\xrightarrow{D} N_1(0,1),\quad a(T_1,T_2)=\frac{\sum_{s=1}^{m}(S_1(\lambda_s)-S_2(\lambda_s))^2}{\sum_{s=1}^{m}\left(S_1^2(\lambda_s)+S_2^2(\lambda_s)\right)},$$

$$B(T_1,T_2)=4Q\left(\sum_{s=1}^{m}(S_1(\lambda_s)-S_2(\lambda_s))\right)^2\left(\frac{K_{T_1}S_1^2(\lambda_s)}{T_1}+\frac{K_{T_2}S_2^2(\lambda_s)}{T_2}\right)\Bigg/\left(\sum_{s=1}^{m}\left(S_1^2(\lambda_s)+S_2^2(\lambda_s)\right)\right)^2.$$

Corollary The following approximation of the power of the test (4) takes place:

$$W = P_{H_{1t_0}}\left\{\hat{\Delta}^2(t_0)>\delta_\varepsilon\right\}\approx 1-\Phi\left(\frac{\delta_\varepsilon - a(T_1,T_2)}{\sqrt{B(T_1,T_2)}}\right).$$

An algorithm for estimation of a time moment of the change-point is based on the extremum property of the statistic (3) and the test (4).

At the time moment $\tau = l+1, K, T-l+1$ the decision on change-point presence is made on the base of a sample $x_{\tau-l},\dots,x_{\tau+l-1}$ ("moving window") for fixed $N=2l$ (l is a parameter).

The algorithm for a change-point detection in the signal consists in successive displacement of the moving window with a fixed step along the signal, calculation of the statistic $\hat{\Delta}^2(\tau)$, finding of its local maximum values, and comparison of the found maximum values with threshold value. An estimator of the change-point time moment τ is defined as:

$$\hat{t} = \begin{cases} \arg\max \hat{\Delta}^2(\tau), \max\left\{\hat{\Delta}^2(\tau)\mid \tau\in\{\tau_-,\dots,\tau_+\}\right\}>\delta_\varepsilon, \\ T+1, \mathit{otherwise}. \end{cases}$$

Statistical classification of the diastole fragments by the autoregressive model

The diastole fragment $\tilde{X}^n = \{X_1,...,X_n\}$ of size n consists of the observations from $L \geq 2$ classes $\{\Omega_1,...,\Omega_L\}$. The observation X_t belongs to the class with a random index $d_t^0 \in S$, $S = \{1,2,...,L\}$ $(t = \overline{1,n})$.

Under the fixed class index $d_t^0 = i$ ($i \in S$) the observation $X_t = (x_{t1},...,x_{t,T_t})' \in R^{T_t}$ is the realization of the length T_t of the autoregressive time series of the order $p \geq 1$ (AR(p)):

$$x_l^i + \theta_{i1}^0 x_{l-1}^i + ... + \theta_{ip}^0 x_{l-p}^i = \xi_l^i, l \in Z = \{0, \pm 1, \pm 2, ...\}, \tag{5}$$

where $\theta_i^0 = (\theta_{i1}^0,...,\theta_{ip}^0)' \in R^p$ is the vector of the autoregressive coefficients for the i-th class; $\{\xi_l^i\}_{l=-\infty}^{+\infty}$ are jointly independent normal (Gaussian) random variables:

$$E\{\xi_l^i\} = 0, \ D\{\xi_l^i\} = E\left\{(\xi_l^i)^2\right\} = \sigma^2 < +\infty, \ l \in Z, \ i \in S. \tag{6}$$

Also the classes $\{\Omega_i\}_{i \in S}$ are characterized by the prior probabilities:

$$\pi_i^0 = P\{d_t^0 = i\} > 0, \ i \in S \ (\pi_1^0 + ... + \pi_L^0 = 1). \tag{7}$$

Thus, the classes $\{\Omega_i\}_{i \in S}$ are determined by the characteristics $\{\pi_i^0, \theta_i^0\}_{i \in S}$ and the variance σ^2 [3,4].

According to the values $\{\theta_i^0\}_{i \in S}$ we suppose:

$$z: \quad z^p + \sum_{j=1}^{p} \theta_{ij}^0 z^{p-j} = 0, \ |z| < 1, \ i \in S. \tag{8}$$

The true classification vector $D^0 = (d_1^0,...,d_n^0)' \in S^n$ is known. The discriminant analysis problem is to consruct $\hat{d}_{n+1}, \hat{d}_{n+2} \in S$ for the true unknown class indices o f the new-registered observations X_{n+1}, X_{n+2} on the sample $\tilde{X}^n = \{X_1,...,X_n\}$ of size n.

Transform the source sample $\tilde{X}^n = \{X_1,...,X_n\}$ to the sample $\tilde{Y}^n = \{Y_1,...,Y_n\}$, where $Y_t \in R^p$ $(t = \overline{1,n})$ is the ML-estimator for the p-vector of the autoregressive coefficients $\theta_{d_t^0}^o \in R^p$ constructed on the observation $X_t \in R^{T_t}$:

$$\{Y_t, \hat{\sigma}_t\} = \arg\max_{\{\bar{\theta},\sigma\}} \ln p(X_t; \bar{\theta}, \sigma). \tag{9}$$

Under the fixed class index $d_t^0 = i$ ($i \in S$):

$$p(X_t; \theta_i^0, \sigma) = n_p(X_t^p \mid O_p, R_p(\theta_i^0, \sigma))(2\pi)^{-(T_t - p)/2} \sigma^{-(T_t - p)} \times \tag{10}$$

$$\times \exp\left(-\frac{1}{2\sigma^2} \sum_{l=p+1}^{T_t} \left(x_{tl} + \theta_{i1}^0 x_{t,l-1} + ... + \theta_{ip}^0 x_{t,l-p}\right)^2\right),$$

where O_p is zero p-vector;

$$X_t = (x_{t1},...,x_{t,T_t})' = ((X_t^p)', x_{t,p+1},...,x_{t,T_t})' \in R^{T_t}; \ X_t^p = (x_{t1},...,x_{tp})' \in R^p;$$

$$n_p(y \mid \mu,\Sigma) = (2\pi)^{-p/2}(\det(\Sigma))^{-1/2}\exp(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu))$$

is the probability density function of the p-dimensional ($y \in R^p$) normal probability distribution $N_p(\mu,\Sigma)$ with the mean vector $\mu \in R^p$ and the covariance ($p \times p$)-matrix Σ ($\det(\Sigma) \neq 0$);

$$R_p(\theta_i^0,\sigma) = \left(\rho_{|k-l|}(\theta_i^0,\sigma)\right)_{k,l=1}^p = E\{X_t^p(X_t^p)' \mid d^0 = i\} \quad (11)$$

is the nonsingular covariance ($p \times p$)-matrix formed the autocovariances: $\rho_k(\theta_i^0,\sigma) = E\{x_{tj}x_{t,j+k} \mid d^0 = i\}$, $k = 0,1,2,...$, determined by the Yule-Walker equations:

$$\sum_{j=1}^{p} \theta_{ij}^0 \rho_j(\theta_i^0,\sigma) + \rho_0(\theta_i^0,\sigma) = \sigma^2; \quad (12)$$

$$\sum_{j=1}^{p} \theta_{ij}^0 \rho_{|k-j|}(\theta_i^0,\sigma) + \rho_k(\theta_i^0,\sigma) = 0, \quad k = 1,2,...$$

Under increasing observation lengths:

$$T_t \to +\infty, \quad t = \overline{1,n}, \quad (13)$$

the ML-estimators $\{Y_t\}_{t=1}^n$ from (9) can be approximately replaced by the LS-estimators:

$$Y_t = -\left(\sum_{l=p+1}^{T_t} X_{tl}^p (X_{tl}^p)'\right)^{-1} \sum_{l=p+1}^{T_t} x_{tl} X_{tl}^p, \quad (14)$$

$X_{tl}^p = (x_{t,l-1},...,x_{t,l-p})' \in R^p$, $t = \overline{1,n}$,

or by the Yule-Walker estimators determined by the equations (12):

$$Y_t = -\left(\hat{R}_t^p\right)^{-1}\hat{r}_t^p, \quad \hat{R}_t^p = (\hat{\rho}_{|k-l|}^t)_{k,l=1}^p, \quad \hat{r}_t^p = (\hat{\rho}_1^t,...,\hat{\rho}_p^t)', \quad t = \overline{1,n}, \quad (15)$$

where $\hat{\rho}_j^p = \sum_{s=1}^{T_t-j} x_{ts}x_{t,s+j}/(T_t - j)$ is the estimator of the autocovariance $\rho_j(\theta_{d_t^0}^0,\sigma)$ ($j = 0,1,...,p$) on the observation-realization X_t of length T_t.

Theorem 3 Let under the model (5)–(7) the stationarity condition (8) be satisfied. Then under the fixed $D^0 = (d_1^0,...,d_n^0)' \in S^n$ the ML-estimators $Y_1,...,Y_n$ from (9) are consistent estimators of the corresponding autoregressive parameters ($t = \overline{1,n}$):

$$Y_t \xrightarrow{P} \theta_{d_t^0}^0, \quad T_t \to +\infty, \quad (16)$$

and have the asymptotically normal distributions:

$$\sqrt{T_t}(Y_t - \theta_{d_t^0}^0) \sim N_p(O_p, \sigma^2 R_p^{-1}(\theta_{d_t^0}^0,\sigma)), \quad T_t \to +\infty, \quad (17)$$

where the covariance matrices $\{R_p(\theta_i^0,\sigma)\}_{i\in S}$ are from (11), (12).

To classify the sample $\tilde{Y}^n = \{Y_1,...,Y_n\}$ the following discriminant analysis procedure can be used [4].

1. On the transformed sample $\tilde{Y}^n = \{Y_1, ..., Y_n\}$ and the true classification vector $D^0 = (d_1^0, ..., d_n^0)' \in S^n$ the unknown class «centres» $\{\theta_i^0\}_{i \in S}$ are estimated:

$$\hat{\theta}_i = (\sum_{t=1}^{n} \delta_{d_t^0, i})^{-1} \sum_{t=1}^{n} \delta_{d_t^0, i} Y_t, \quad i \in S, \tag{18}$$

where $\delta_{j,i} = \{1, if\ j = i; 0, if\ j \neq i\}$ is the Kronecker symbol.

2. The new observation Y (constructed on $X \in \{X_{n+1}, X_{n+2}, ...\}$) is classified:

$$\hat{d} = d(Y; \hat{\theta}) = \arg\min_{i \in S} \left| Y - \hat{\theta}_i \right|. \tag{19}$$

Now investigate the efficiency of the procedure (18), (19).

This procedure is based on the decision rule (DR)

$$d = d(Y; \theta^0) = \arg\min_{i \in S} \left| Y - \theta_i^0 \right|, \quad Y \in R^p, \tag{20}$$

and on the "plug-in" DR $\hat{d}_t = d(Y_t; \hat{\theta}) = \arg\min_{i \in S} \left| Y_t - \hat{\theta}_i \right|$, $t = 1, n, n+1, n+2, ...$, obtained from the DR (20) by substituting the statistical estimator $\hat{\theta} = (\hat{\theta}_1', ..., \hat{\theta}_L')'$ instead of the unknown vector $\theta^o \in R^{Lp}$ of the true autoregressive parameters $\{\theta_i^0\}_{i \in S}$.

Let us evaluate the risk (the classification error probability) of the DR (20):

$$r_T = P\{d(Y; \theta^0) \neq d^0\}, \tag{21}$$

under the conditions when length T of the source observation $X \in R^T$ corresponded the ML-estimator $Y \in R^p$ is large ($T \to +\infty$, the asymptotics (13)).

Theorem 4 Under the conditions of Theorem 3 the risk (21) of the DR (20) satisfies the relation:

$$r_T / r_T^* \to 1, \quad T \to +\infty; \tag{22}$$

$$r_T^* = 1 - \sum_{i \in S} \pi_i^0 \int_{R^p} \prod_{\substack{j \in S \\ j \neq i}} U\left((\theta_i^0 - \theta_j^0)' z + \sqrt{T} \frac{\left| \theta_i^0 - \theta_j^0 \right|^2}{2\sigma} \right) n_p(z \mid O_p, R_p^{-1}(\theta_i^0, \sigma)) dz,$$

where $U(\omega) = \{1$, if $\omega \geq 0$; 0, if $\omega < 0\}$ is the unit function.

For the case of two classes ($L = 2$):

$$r_T^* = \pi_1^0 \Phi\left(-\sqrt{T} \frac{\left| \theta_1^0 - \theta_2^0 \right|^2}{2\sigma \Delta(\theta_1^0, \theta_2^0)} \right) + \pi_2^0 \Phi\left(-\sqrt{T} \frac{\left| \theta_1^0 - \theta_2^0 \right|^2}{2\sigma \Delta(\theta_2^0, \theta_1^0)} \right), \tag{23}$$

where $\Delta(\theta_i^0, \theta_j^0) = \sqrt{(\theta_i^0 - \theta_j^0)' R_p^{-1}(\theta_i^0, \sigma)(\theta_i^0 - \theta_j^0)}$; $\Phi(\cdot)$ is the distribution function of the standard Gaussian law $N_1(0,1)$.

In practice at large lengths of observations ($T \to +\infty$) the proved relations (22), (23) allows approximately evaluate the risk (21): $r_T \approx r_T^*$.

Experimental results

The numerical analysis of efficiency of the proposed change-points detection algorithm was performed by statistical modeling.

Acoustic signals of the blood flow were simulated in the following way: duration of the signal is 1000 milliseconds, the systole lasts from 0 millisecond to 300 milliseconds, the flap of the mitral valve damper lasts from 301 milliseconds to 350 milliseconds, the diastole lasts from 351 milliseconds to 1000 milliseconds. All fragments of the signal are described by AR(2) model with different parameters. The accuracy of change-point detection at a specified time moment in the blood flow signal for healthy men and men with the coronary ischemic disease is presented in Table 1.

Time, milliseconds	Accuracy of a change-point detection at the moment t, %					
	t=275 (no change-points)	t=300 (there is a change-point)	t=325 (no change-points)	t=350 (there is a change-point)	t=375 (no change-points)	t=400 (no change-points)
Healthy men	100%	95%	95%	95%	97%	100%
Men with the coronary ischemic disease	100%	97%	97%	94%	95%	100%

Table 1 Accuracy of a change-point detection

Thus, in order to form a training sample of the blood flow signals for healthy men and men with the coronary ischemic disease one needs to extract diastole fragment from the signal from t=400 milliseconds, which presents stationary part of the diastole.

The training samples were formed by 500 stationary fragments of diastole of healthy men and men with the coronary ischemic disease. To each sample the classification procedure (19), (20) based on the different types of estimators (the ML-estimators (9), the LS-estimators (14) and the Yule-Walker estimators (15)) is used. Estimated frequencies of error decision (FED) and the risk of the classification procedure are presented in Table 2.

T	FED; Estimators			Risk, r_T^*
	ML	LS	Yule-Walker	
75	0.186	0.188	0.232	0.159
100	0.112	0.112	0.153	0.125
125	0.090	0.090	0.126	0.099

Table 2 Frequency of error decision (FED)

These investigations were supported by ISTC (Project B-705).

References

[1] Brillinger D. R. Time Series. Data analysis and theory. Holt, Rinehart and Winston, Inc. New York. 1975.

[2] Kharin Yu.S., Abramovich M.S. On detection of spectral change-point in bivariate time series // Optoelectronics, Instrumentation and Data Processing. 1999. No. 2.

[3] Zhuk E.E., Kharin Yu.S. Robustness in Cluster Analysis of Multivariate Observations. BSU, Minsk, 1998. (in Russian).

[4] Zhuk E.E. Cluster Analysis of the Realizations of Autoregression Time Series // Automation and Remote Control. 2003. Vol. 64. No.1.