INFORMATICA, 1992, Vol.3, No.1, 80–87

A PRACTICAL METHOD FOR SEGMENTATION AND ESTIMATION OF MODEL PARAMETERS OF THE PROCESSES WITH FREQUENTLY AND INSTANTLY CHANGING PROPERTIES

Egidijus OSTAŠEVIČIUS

Institute of Mathematics and Informatics
Lithuanian Academy of Sciences
2600 Vilnius, Akademijos St.4, Lithuania

Abstract. A practical method for segmentation and estimation of model parameters of processes is proposed in this paper. A pseudo-stationary random process with instantly changing properties is divided into stationary segments. Every segment is described by an autoregressive model. A maximum likehood method is used for segmentation of the random process and estimation of unknown model parameters. An example with simulated data is presented.

Key words: random process, segmentation, maximum likelihood estimation, pseudo-stationary time series, autoregressive model.

1. Introduction. Many practical time series exhibit non-stationary characteristics. One of the categories of such non-stationarity is a pseudo-stationary time series which may be considered as the "sewing" together of a number of stationary time series (see Fig. 1(a)).

We can treat such time series as random processes with instantly and frequently changing properties. There are instant and noticeable changes of the average in the realization of the random process and they separate individual segments. We may assume first that the random process in a segment is stationary between changes of the average . Then, the components of the random process in different segments are independent though they may be dependent within the segment. Every segment as part of the

stationary random process may be modelled by an autoregressive (AR) model and belong to a certain class. The segments with similar properties may belong to the same class. A nonsupervised control process performs switching among classes. It may be a Markov chain which characterizes the frequency of switching among classes.

A maximum likelihood method is used here for segmentation of random processes with instantly changing properties and estimation of unknown autoregressive model parameters for each class. Besides, estimation of the matrix of state transition probabilities for the Markov chain is proposed. The results of segmentation and estimation of the model parameters, using simulated time series, are presented.

2. Statement of the problem. Let $\{X_t\}$, $t = 1,\ldots,N$ be a discrete random process which characterizes the features of the observed object. At any moment of time t an object may be in one of L states, i.e., a component of the process X_t may belong to one of L classes. We shall introduce a segment of the process $\{X_t\}$ as a sequence of the process components X_t belonging to the same class. We shall denote a sequence of the states of process components X_t $(t = 1,\ldots,N)$ by $\{s_t : s_t = l,\ l = 1,\ldots,L\}$. The state sequence $\{s_t\}$ is a Markov chain with the matrix of state transition probabilities $Q = \{q_{ij} : q_{ij} = P[s_t = j \mid s_{t-1} = i],\ i,j = 1,\ldots,L\}$.

The random process $\{X_t\}$ is modelled by an autoregressive model with instantly changing parameters depending on the state s_t:

$$X_t = X_t(s_t) = \mu(s_t) - \sum_{i=1}^{p(s_t)} a_i(s_t)\left[X_{t-i}(s_t) - \mu(s_t)\right] + b(s_t)V_t, \quad (1)$$

where the model parameters depend upon the state of the object. Each state corresponds to the vector of model parameters $A_l = (\mu_l, a_1^{(l)},\ldots,a_{p_l}^{(l)}, p_l, b_l)$, $(l = 1,\ldots,L)$. $\{V_t\}$ are the normal independent random variables with zero mean and the variance equal to 1. Besides we assume the components of $\{X_t\}$ in different segments to be independent.

The problem lies in estimating the state sequence $\{s_t\}$, i.e., carrying out the segmentation of the random process according to the classes, unknown parameters of autoregressive models A_l for each class l $(l = 1,\ldots,L)$ and the matrix Q of state transition probablities using the realization $\{x_t\}$ of the random process $\{X_t\}$ $(t = 1,\ldots,N)$. The number of classes L is assumed to be known.

3. Estimation of the state sequence $\{s_t\}$. We solve this problem temporarily assuming that $p(s_t) = 0$, i.e., all components of the process $\{X_t\}$ are independent. According to (1) and since V_t has a normal distributation, the probability density for X_t is a mixture of normals with weight coefficients $\pi_l (l = 1,\ldots,L)$:

$$h(X_t \mid \theta) = \sum_{l=1}^{L} \pi_l f(X_t \mid \mu_l, \sigma_l^2), \tag{2}$$

where L is the number of components of normal probability densities in the mixture which is the same as the number of classes; $\sum_{l=1}^{L} \pi_l = 1$, (μ_l, σ_l^2) are the parameters of the normal density function.

The probability density function for the process $\{X_t\}$ is

$$h(X_1,\ldots,X_N) = \prod_{T=1}^{N} \sum_{l=1}^{L} \pi_l f(X_t \mid \mu_l, \sigma_l^2), \tag{3}$$

where

$$f(X_t \mid \mu_l, \sigma_l^2) = \frac{1}{\sqrt{\pi}\sigma_l} \exp\left\{\frac{(x_t - \mu_l)^2}{2\sigma_l^2}\right\}. \tag{4}$$

The parameters π_l, μ_l, σ_l^2 $(l = 1,\ldots,L)$ will be estimated by the maximum likelihood method using the logarithm of the likelihood function:

$$L(\theta) = \sum_{t=1}^{N} \log \sum_{l=1}^{L} \pi_l f_l(x_t \mid \mu_l, \sigma_l^2), \tag{5}$$

where $\bar{\theta} = (\pi_1,\ldots,\pi_L,\mu_1,\ldots,\mu_L,\sigma_1^2,\ldots,\sigma_L^2)$ – is the vector of parameters of a mixture of normal distribution densities.

We shall use the EM recurrent procedure (Redner, 1984) which maximizes (5).

Step 1. Computation of the a posteriori probability g_{lt} for the component X_t of the random process realization:

$$\hat{g}_{lt} = \frac{\hat{\pi}_l f(x_t \mid \hat{\mu}_l, \hat{\sigma}_l^2)}{\sum_{n=1}^{L} \hat{\pi}_n f(x_t \mid \hat{\mu}_n, \hat{\sigma}_n^2)}. \tag{6}$$

Step 2. Precise estimates of parameters π_l, μ_l, σ_l^2:

$$\hat{\pi}_l = \frac{1}{N} \sum_{t=1}^{N} \hat{g}_{lt}, \tag{7}$$

$$\hat{\mu}_l = \frac{\sum_{t=1}^{N} \hat{g}_{lt} x_t}{\sum_{t=1}^{N} \hat{g}_{lt}}, \tag{8}$$

$$\hat{\sigma}_l^2 = \frac{\sum_{t=1}^{N} \hat{g}_{lt}(x_t) - \hat{\mu}_l)^2}{\sum_{t=1}^{N} \hat{g}_{lt}} \qquad (l = 1, \ldots, L). \tag{9}$$

Step 3. Computation of $L(\hat{\theta}^{(i)})$ by (5), where i is the number of iteration.

Step 4. If

$$|L(\hat{\theta}^{i-1}) - L(\hat{\theta}^{(i)}| < r \tag{10}$$

is true (where r is some threshold), then the recurrent procedure is completed, otherwise, we repeat step 1.

Initial meanings of the parameters $(\pi_l^0, \mu_l^0, \sigma_l^{20})$ $(l = 1, \ldots, L)$ may be computed by the algorithm (Ostaševičius, 1986). The threshold r is determined experimentally.

The estimate of the state sequence $\{\hat{s}_t\}$, using the sequence of a posteriori probabilities $\{\hat{g}_{lt}\}$ $(l = 1, \ldots, L,\ t = 1, \ldots, N)$, is

$$\hat{s}_t = \arg \max_{1 \leqslant l \leqslant L} \{\hat{g}_{lt}\} \qquad (t = 1, \ldots, N). \tag{11}$$

4. Segmentation. Let us denote a segment by w_{ln}, where l is the number of the class $(l = 1, \ldots, k)$, n is the number of the segment in the class. The number N_l of segments for the class l $(l = 1, \ldots, L)$ is being defined during the formation of segments.

Assume that $n-1$ segments are formed for the class l. The components $x_i, x_{i+1}, \ldots, x_j$ will belong to the segment w_{ln} of class l if it is true

$$s_{i-1} \neq l \quad \text{and} \quad s_i = l \quad \text{and} \quad s_{i+1} = l \quad \text{and} \quad \ldots$$
$$\ldots \quad \text{and} \quad s_{j-1} = l \quad \text{and} \quad s_j = l \quad \text{and} \quad s_{j+1} \neq l. \tag{12}$$

All the segments $w_{l1}, w_{l2}, \ldots, w_{lN_l}$ for each class l will be formed in the same way.

5. Estimation of parameters of the autoregressive model for the class. Let $w_{l1}, w_{l2}, \ldots, w_{lN_l}$ be the segments which belong to the same class l. The autoregressive model parameters for this class will be estimated by solving the system of linear equations with respect to the parameters $a_1^{(l)}, \ldots, a_{p_l}^{(l)}$ when model order p_l is fixed. We define some indicator function

$$J(t,l) = \begin{cases} 1, & \text{if } s_{t-p_l} = l \text{ and} \\ & s_{t-p_l+1} = l \text{ and } s_{t-1} = l \text{ and } s_t = l, \\ 0, & \text{otherwise } (l = 1, \ldots, L;\ t = 1, \ldots, N). \end{cases} \tag{13}$$

The system of linear equations is

$$C^{(l)} a^{(l)} = c^{(l)}, \tag{14}$$

where $C^{(l)}$ is the matrix and $a^{(l)}$, $c^{(l)}$ are column-vectors with the elements:

$$c_{ij}^{(l)} = \sum_{t=1}^{N} J(t,l)(x_{t-i} - \hat{\mu}_l)(x_{t-j} - \hat{\mu}_l), \tag{15}$$

$$\hat{a}_i^{(l)},$$

$$c_j^{(l)} = -c_{0j}^{(l)}; \tag{16}$$

$(i = 1, \ldots, p_l;\ j = 0, 1, \ldots, p_l)$
and the estimates of parameters $\hat{b}_l$ are

$$\hat{b}_l = \left\{ \frac{1}{M_l} \sum_{t=1}^{N} J(t,l) \left[x_t - \hat{\mu}_l + \sum_{i=1}^{p_l} a_i^{(l)} (x_{t-i} - \hat{\mu}_l) \right]^2 \right\}^{1/2}, \tag{17}$$

where $M_l = \sum_{t=1}^{N} J(t,l)$.

Model order p_l may be estimated by what is commonly known as information criterion measures. A number of information criteria have been proposed, e.g., Hannan-Quinn (Hannan, 1979) that we use.

Let $p_{\max}$ be the maximum order of the model. Model order p_l will be determined as

$$\hat{p}_l = \arg\min \left\{ M_l \log \left[c_0^{(l)} - \sum_{i=1}^{p} a_i^{(l)} c_i^{(l)} \right] + \gamma \log\left(\log(M_l)\right) \right\}, \tag{18}$$

where $\gamma > 2$.

6. Estimation of the matrix Q of state transition probabilities. We compute the matrix $\nu = [\nu_{ij},\ i,j = 1,\ldots,L]$ of state transition frequencies using the estimated state sequence $\{\hat{s}_t\}$ $(t = 1,\ldots,N)$.

Let us introduce indicator functions such as

$$I(i,i,t) = \begin{cases} 0, & \text{if} \quad s_{t-1} \neq s_t, \\ 1, & \text{if} \quad s_{t-1} = i \quad \text{and} \quad s_t = i; \end{cases} \tag{19}$$

$(i = 1,\ldots,L;\ t = 2,\ldots,N)$

$$J(i,j,t) = \begin{cases} 0, & \text{if} \quad s_{t-1} = s_t, \\ 1, & \text{if} \quad s_{t-1} = i \quad \text{and} \quad s_t = j \quad \text{and} \quad i \neq j; \end{cases} \tag{20}$$

$(i,j = 1,\ldots,L;\ t = 2,\ldots,N)$.

The elements of the state transition frequency matrix are computed by

$$\nu_{ii} = \frac{1}{N-1} \sum_{t=2}^{N} I(i,i,t), \tag{21}$$

$$\nu_{ij} = \frac{1}{N-1} \sum_{t=2}^{N} J(i,j,t). \tag{22}$$

If $N \to \infty$ for the Markov chain, then the elements ν_{ij} of the state transition frequency matrix ν converge to the elements q_{ij} of the state transition probability matrix Q. For large N

$$\hat{q}_{ij} \simeq \nu_{ij} \qquad (i,j = 1,\ldots,L). \tag{23}$$

7. Example. Let as consider an example. A realization of a random process was generated using the model parameters from Table 1. The number of class was $L = 2$ and the length of realization was $N = 3000$. Fig. 1 shows a fragment of the generated realization (a) and the result of segmentation (b). Estimated parameters are presented in Table 2. The number of iterations was 15.

Table 1. Model parameters

l	μ_l	b_l	p_l	$a_1^{(l)}$	$a_2^{(l)}$	Q
1	-3.000	1.000	2	-0.750	0.500	$\begin{pmatrix} 0.950 & 0.050 \\ 0.050 & 0.950 \end{pmatrix}$
2	3.000	1.000	2	0.250	-0.500	

Table 2. Estimated parameters

l	$\widehat{\mu}_l$	$\widehat{b}_l$	$\widehat{p}_l$	$\widehat{a}_1^{(l)}$	$\widehat{a}_2^{(l)}$	$\widehat{Q}$
1	-3.082	1.246	2	-0.719	0.474	$\begin{pmatrix} 0.934 & 0.066 \\ 0.059 & 0.941 \end{pmatrix}$
2	3.02	1.313	2	0.237	-0.504	

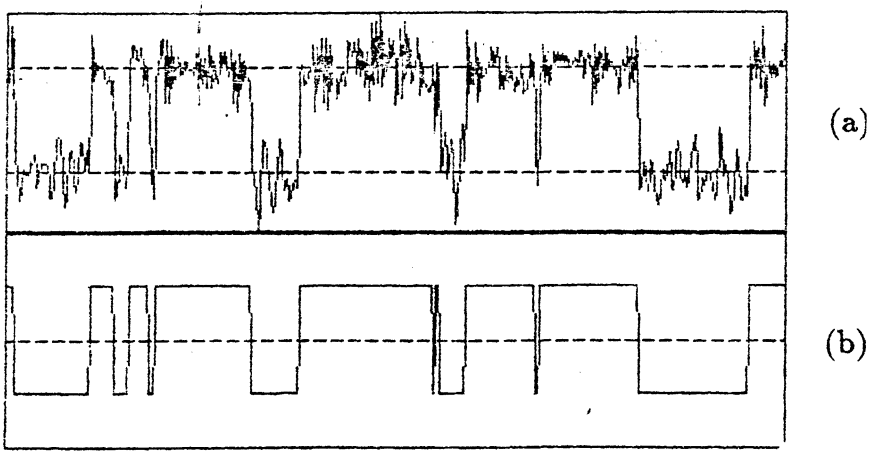

Fig. 1. Generated realization $\{x_t\}$ of the random process (a) and estimated state sequence $\{s_t\}$ (b).

8. Conclusions. We propose a practical method for analyzing the pseudo-stationary random processes in this paper. This method may be applied especially to random processes with an exactly and instantly changing average which separates individual segments. First of all this method includes the segmentation of a pseudo-stationary random process into stationary segments. Then, the parameter estimation technique for the linear model is used. The estimation of the state transition probability matrix is presented as well. By applying the method to the simulated random process, it is observed that the method works well. Thus, the method can be recommended for solving practical problems.

REFERENCES

Hannan, E.J., and B.G.Quinn (1979). The determination of the order of autoregression. *Journal of Royal Statistical Society,* Ser. B, **41**(2).

Ostaševičius, E. (1986). An algorithm for the estimation of unknown parameters of random processes with the distribution of a mixture of normals. In *Statisticheskie Problemy Upravleniya,* Vol.77. Vilnius. pp. 26–34. (in Russian).

Redner, L., and J.Walkner (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review,* **26**(2).

Received October 1991

E. Ostaševičius is a researcher of the Recognition Processes Department at the Institute of Mathematics and Informatics, Lithuanian Academy of Sciences. He received the degree of Candidate of Technical Sciences from the Kaunas Polytechnic Institute, Kaunas, Lithuania in 1989. His main research interests are: analysis of pseudo-stationary random processes, application of computers.