Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

CLASSIFYING SERIES REALISATIONS FROM ARMA (p, q) PROCESSES

M. KRZYŚKO, J. WACHOWIAK

This paper is concerned with discrimination among multivariate autoregressive-moving average (ARMA) processes with known parameters by the Bayes method. To calculate the value of the Bayesian risk we need to know the distribution of the discriminant function. Hence, the exact and asymptotic distributions of the discriminant function are investigated.

1. Classification procedure. Let π_i, $i=1, 2, \dots, k$, denote a class of the r-component vector ARMA (p_i, q_i) processes

$$(1)\qquad \sum_{u=0}^{p_i} A_i(u)\,[X_i(t-u)-\mu_i]=\sum_{v=0}^{q_i} B_i(v)\,e_i(t-v),\quad t=0,\ \pm 1,\ \pm 2, \dots,$$

where $X_i(t)$ are observable random vectors of finite size $r\geqq 1$, $\mu_i=E\{X_i(t)\}$, the $A_i(u)$ and $B_i(v)$ are $r\times r$ matrices of parameters, $A_i(0)=B_i(0)=I_r$, and the $e_i(t)$ are unobservable, independently normally distributed random vectors with $E\{e_i(t)\}=0$, $E\{e_i(t), e_i'(s)\}=\delta_{st}V$, where δ_{st} is Kronecker's delta function. We assume that V is positive definite. This model has r^2p_i parameters in the $A_i(u)$, r^2q_i in the $B_i(v)$, r parameters in μ_i and $\frac{1}{2}r(r+1)$ parameters in V, $i=1, 2, \dots, k$.

Let $\Gamma_i(n-m)=E\{(X_i(m)-\mu_i)(X_i(n)-\mu_i)'\}$ denote the covariance matrix of the process $\{X_i(t)\}$, $i=1, 2, \dots, k$. Assume that $\Theta_i=(A_i(1), \dots, A_i(p_i), B_i(1), \dots, B_i(q_i), \mu_i, V)$, $i=1, 2, \dots, k$, are known, which also implies that the covariance matrices $\Gamma_1(t)$, $\Gamma_2(t)$, $\dots, \Gamma_k(t)$ are known.

Let us assume that we have a realization $X=(X(1), X(2), \dots, X(T))$ of a process belonging to one of the classes $\pi_1, \pi_2, \dots, \pi_k$. We consider k alternative hypotheses $H_1, H_2, \dots, H_k$ forming a complete system of disjoint events, i. e. if $P_i>0$, $i=1, 2, \dots, k$, is the prior probability of the acceptability of hypothesis H_i, then $P_1+P_2+\dots+P_k=1$. The hypothesis H_i states that X is a realization of a process belonging to class π_i, $i=1, 2, \dots, k$. Our task is to verify the acceptability of one of these k alternative hypotheses. In the statistical literature (see, for example, Anderson, 1958, Chapt. 6) this problem is known as classification problem. We will make use of the Bayesian method of classification dividing $\mathcal{R}^N$ $(N=rT)$ into non-intersecting classification regions $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_k$ defined so as to minimize the Bayesian risk

$$R=1-\sum_{i=1}^{k} P_i \int_{\mathcal{R}_i} f(X/\Theta_i)\,dX,$$

where $f(X|\Theta_i)$ is the density function of the joint distribution of X in class π_i, $i=1, 2, \dots, k$. The classification region $\mathcal{R}_i$ minimizing the Bayesian risk is

$$\mathcal{R}_i=\{X\colon v_{ij}(X)\geq \ln (P_j/P_i),\quad j=1, 2, \dots, k,\quad j\neq i\},$$

where $v_{ij}(X)$ is a discriminant function of the following form $v_{ij}(X)=\ln(f(X|\Theta_i)/f(X|\Theta_j))$, $i, j,=1, 2, \ldots, k, j\neq i$. According to the Bayesian method, H_i is accepted if $X\in\mathcal{R}_i$, $i=1, 2, \ldots, k$. When the hypothesis H_i is true the probability α_i of rejecting it is equal to

$$\alpha_i=1-\int_{\mathcal{R}_i} f(X|\Theta_i)\,dX=1-P(v_{ij}(X)\geq\ln(P_j/P_i), \quad k, j=1, 2, \ldots, k.$$

$$j\neq i\,|\,\Theta=\Theta_i), \quad i=1, 2, \ldots, k,$$

and the minimum value of the Bayesian risk is

$$R=\sum_{i=1}^{k} P_i\alpha_i=1-\sum_{i=1}^{k} P_iP(v_{ij}(X)\geq\ln(P_j/P_i), \quad j=1, 2, \ldots, k, \quad j\neq i\,|\,\Theta=\Theta_i).$$

The probability α_i is very difficult to evaluate in this case because of the very complicated shapes of the classification regions. An alternative way of obtaining an upper bound for α_i is given by using the Bonferroni inequality which yields

$$\alpha_i\leq\sum_{\substack{j=1\\ j\neq 1}}^{k} P(v_{ij}(X)<\ln(P_j/P_i)), \quad i=1, 2, \ldots, k,$$

and

$$R\leq\sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq 1}}^{k} P_iP(v_{ij}(X)<\ln(P_j/P_i)).$$

If $k=2$, then equality holds in the above formulas.

Two equivalent forms of the density function $f(X|\Theta_i)$ are available.

1) By the normality of $e_i(t)$ we have

$$f(X|\Theta_i)=(2\pi)^{-N/2}|G_i|^{-1/2}\exp\{-\tfrac{1}{2}\operatorname{vec}'(X-m_i)G_i^{-1}\operatorname{vec}(X-m_i)\},$$

where $m_i=\mu_i 1'_N$, $1'_N=(1, 1, \ldots, 1)$, $i=1, 2, \ldots, k$, while vec A is a vector formed by stacking the columns of A, one on top of the other, in order from left to right. The matrix G_i is a block matrix of T^2 blocks, where block (m, n) contains the matrix $\Gamma_i(n-m)$, $i=1, 2, \ldots, k$. The matrix $\Gamma_i(n-m)$ can be expressed by the parameters of the equation (1).

We will need the following notation:

$$\varepsilon_i(t)=\sum_{v=0}^{q_i} B_i(v)\,e_i(t-v), \tag{2}$$

$$\Psi_i(t)=(\varepsilon_i'(t), 0, \ldots, 0), \quad Y_i(t)=((X_i(t)-\mu_i)', \ldots, (X_i(t-p_i+1)-\mu_i)')',$$

$$A_i=\begin{bmatrix} -A_i(1), & -A_i(2), \ldots, & -A_i(p_i) \\ & I_{r(p_i-1)}, & 0 \end{bmatrix}, \quad i=1, 2, \ldots, k.$$

In this notation the model (1) has the following form

$$Y_i(t)=A_iY_i(t-1)+\Psi_i(t), \quad i=1, 2, \ldots, k. \tag{3}$$

Note that $X_i(t)=\mu_i+W'Y_i(t)$ and $\varepsilon_i(t)=W'\Psi_i(t)$, where $W'=(I_r, 0, \ldots, 0)$ is a $r\times rp_i$ matrix, $i=1, 2, \ldots, k$, The process $Y_i(t)$ of the form (3) has the representation $Y_i(t)=\sum_{u=0}^{\infty}A_i^u\Psi_i(t-u)$ so that

$$X_i(t)=\mu_i+\sum_{u=0}^{\infty} W' A_i^u \Psi_i(t-u)$$

or $X_i(t)=\mu_i+\Sigma_{u=0}^{\infty} C_i(u)\varepsilon_i(t-u)$, where $C_i(u)=W' A_i^u W$, $i=1, 2, \ldots, k$.
Hence

$$\Gamma_i(n-m)=\sum_{u=0}^{\infty}\sum_{v=0}^{\infty} C_i(u)\, E[\varepsilon_i(m-u)\,\varepsilon_i'(n-v)]\, C_i'(v)$$

$$=\sum_{u=0}^{\infty}\sum_{r=0}^{q_i}\sum_{s=0}^{q_i} C_i(u)B_i(r)VB_i'(s)C_i'(n-m+r-s+u), \quad i=1, 2, \ldots, k.$$

The spectral decomposition of a matrix A_i has the form $A_i=P_i\Lambda_i P_1^{-1}$, where $\Lambda_i=\text{diag}(\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{i,rp_i})$, $i=1, 2, \ldots, k$. Hence

$$\Gamma_i(n-m)=W' P_i\Big(\sum_{j=0}^{q_i} H_i^*(j, n-m)\Big)P_i' W, \tag{4}$$

where

$$H_i^*(j, n-m)=(h_{uv}^*(i, j, n-m))=\sum_{u=0}^{\infty}\Lambda_i^u H_i(j)\,\Lambda_i^{n-m+j+u},$$

$$H_i(j)=(h_{uv}(i, j))=\sum_{s=0}^{q_i} P_i^{-1} WB_i(j)VB_i'(s)W'(P_i')^{-1}\Lambda_i^s,$$

$$h_{uv}^*(i, j, n-m)=\frac{h_{uv}(i, j)\lambda_{iv}^{n-m+j}}{1-\lambda_{iu}\lambda_{iv}}.$$

The discriminant function $v_{ij}(X)$ takes the following form

$$v_{ij}(X)=-\frac{1}{2}[\text{vec}'(X-m_i)G_i^{-1}\text{vec}(X-m_i) \tag{5}$$

$$-\text{vec}'(X-m_j)G_j^{-1}\text{vec}(X-m_j)-\ln(|G_j|/|G_i|)],$$

where the matrix G_i contains the matrices $\Gamma_i(n-m)$ given by (4), $i, j=1, 2, \ldots, k$, $j\neq i$.

2) The density function $f(X|\Theta_i)$ may be written as follows: $f(X|\Theta_i)=f_1(X_{p_i}|\Theta_i)\cdot f_2(X(p_i+1), \ldots, X(T)|X_{p_i})$, where $X_{p_i}=(X(1), X(2), \ldots, X(p_i))$, $i=1, 2, \ldots, k$.
We have

$$f_1(X_{p_i}|\Theta_i)=(2\pi)^{-rp_i/2}|G_{p_i}|^{-1/2}\exp\{-\frac{1}{2}w_{p_i}^2\},$$

where $w_{p_i}^2=\text{vec}'(X_{p_i}-m_{p_i})G_{p_i}^{-1}\text{vec}(X_{p_i}-m_{p_i})$, $m_{p_i}=\mu_i 1_{p_i}'$, $1_{p_i}'=(1, 1, \ldots 1)$, $i=1, 2, \ldots, k$
The matrix G_{p_i} is a block matrix of p_i^2 blocks, where block (m, n) contains the matrix $\Gamma_i(n-m)$, $i=1, 2, \ldots, k$.

The equation (2) for $t=p_i+1, p_i+2, \ldots, T$ we can express in the following form $\varepsilon_i=B_i e_i$, where $\varepsilon_i'=(\varepsilon_i'(p_i+1), \varepsilon_i'(p_i+2), \ldots, \varepsilon_i'(T))'$, $e_i'=(e_i'(p_i+1-q_i), e_i'(p_i+2-q_i), \ldots, e_i'(T))'$,

$$B_i=\begin{bmatrix} & B_i(q_i), \ldots, B_i(1), I, 0, \ldots, 0 \\ 0, & B_i(q_i), \ldots, B_i(1), I, \quad \ldots, 0 \\ \ldots & \ldots \\ 0, \ldots, 0, & B_i(q_i), \ldots, B_i(1), I \end{bmatrix}, \quad i=1, 2, \ldots, k.$$

It is known that $e_i \sim N(0, I_{T-p_i+q_i}\otimes V)$, $i=1, 2, \ldots, k$.

Hence $\varepsilon_i \sim N(0, D_i)$, where $D_i = B_i(I_{T-p_i+q_i} \otimes V)B_i'$, $i=1, 2, \ldots, k$. In order to find the form of the function f_2, we may use our knowledge of the distribution of the vector ε_i, $i=1, 2, \ldots, k$. Regarding $\Sigma_{u=0}^{p_i} A_i(u)[X_i(t-u)-\mu_i] = \varepsilon_i(t)$ as a transformation $\varepsilon(t)$ to $X(t)$ for $t=p_i+1, \ldots, T$, with the Jacobian of the transformation being 1, we obtain

$$f_2(X(p_i+1), \ldots, X(T) \mid X_{p_i}, \Theta_i) = (2\pi)^{-\frac{r(T-p_i)}{2}} |D_i|^{-\frac{1}{2}} \exp\{-\frac{1}{2} y_i' D_i^{-1} y_i\},$$

where $y_i' = (y_i'(p_i+1), y_i'(p_i+2), \ldots, y_i'(T))$, $y_i(t) = \Sigma_{u=0}^{p_i} A_i(u)[X(t-u)-\mu_i]$, $i=1,2,\cdots,k$. Hence

$$f(X \mid \Theta_i) = (2\pi)^{-\frac{rT}{2}} |D_i|^{-\frac{1}{2}} |G_{p_i}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}[y_i' D_i^{-1} y_i + w_{p_i}^2]\}, \quad i=1,2, \ldots, k.$$

The discriminant function $v_{ij}(X)$ takes the following form

(6) $$2v_{ij}(X) = y_j' D_j^{-1} y_j - y_i' D_i^{-1} y_i + w_{p_j}^2 - w_{p_i}^2 + 2A_{ij},$$

where $A_{ij} = \frac{1}{2} \ln(|G_{p_j}|/|G_{p_i}|) + \frac{1}{2} \ln(|D_j|/|D_i|)$, $i, j=1, 2, \ldots, k$, $j \neq i$.
To calculate the minimum value of the Bayesian risk we need to know the distribution of $v_{ij}(X)$, $i, j=1, 2, \ldots, k$, $j \neq i$.

2. Distribution of the discriminant function. First of all we take into consideration the discriminant function of the form (5). Since G_i^{-1} is a symmetric p. d. matrix, there exists a nonsingular matrix A_i such that $G_i^{-1} = A_i' A_i$, $i=1, 2, \ldots, k$. For $n=1, \ldots, N$ let $\lambda_{ij,n}$ be the eigenvalue and $P_{ij,n}$ be the corresponding eigenvector of $R_{ij} = (A_i^{-1})' G_j^{-1} A_i^{-1}$, and let $P_{ij} = (P_{ij,1}, \ldots, P_{ij,N})$, where $i, j=1, 2, \ldots, k$, $j \neq i$. We will need the following notation

$$b_{ij,n} = \frac{1}{2}(\lambda_{ij,n} - 1), \quad \rho_{ij} = 2(A_i^{-1})' G_j^{-1} \operatorname{vec}(m_i - m_j),$$

$$\gamma_{ij} = (\gamma_{ij,1}, \ldots, \gamma_{ij,N}) = P_{ij}^{-1} \rho_{ij}, \quad \delta_{ij,n}^2 = \gamma_{ij,n}^2 / 16\, b_{ij,n}^2,$$

$$k_{ij} = \frac{1}{2}[\operatorname{vec}'(m_i - m_j) G_j^{-1} \operatorname{vec}(m_i - m_j) + \ln(|G_j|/|G_i|)],$$

$$a_{ij} = k_{ij} - \sum_{n=1}^{N} \gamma_{i,jn}^2 / 16\, b_{ij,n}, \quad i, j=1, 2, \ldots, k, \ j \neq i.$$

Theorem 1. *The distribution of the discriminant function $v_{ij}(X)$ is*

(7) $$P(v_{ij}(X) < y \mid \Theta = \Theta_i) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin(s_{ij}(u, y))}{u t_{ij}(u)} du,$$

where

$$s_{ij}(u, y) = \frac{1}{2} \sum_{n=1}^{N} (\tan^{-1}(b_{ij,n} u) + \delta_{ij,n}^2 b_{ij,n} u (1 + b_{ij,n}^2 u^2)^{-1}) - \frac{1}{2}(y - a_{ij}) u,$$

$$t_{ij}(u) = \sum_{n=1}^{N} (1 + b_{ij,n}^2 u^2)^{\frac{1}{4}} \exp\{\frac{1}{2} \sum_{n=1}^{N} (\delta_{ij,n} b_{ij,n} u)^2 / (1 + b_{ij,n}^2 u^2)\}, \quad i, j=1, 2, \ldots, k, \ j \neq i.$$

Remark 1. The proof of this theorem is based on Imhof's (1961) result concerning the distribution of quadratic forms in normal variables.

Remark 2. Since the distribution function of $v_{ij}(X)$ is very complicated we will find its asymptotic distribution in the case of the ARMA$(p_i, 0)$ processes, $i, j = 1, 2, \ldots, k$, $j \neq i$.

3. Asymptotic distribution of the discriminant function in the case of the ARMA $(p_i, 0)$ processes. Now, we take into consideration the discriminant function of the form (6). In the special case $q_i = 0$, the discriminant function (6) has the following form

$$2v_{ij}(X) = \sum_{t=p_j+1}^{T} z_j^2(t) - \sum_{t=p_i+1}^{T} z_i^2(t) + w_{p_j}^2 - w_{p_i}^2 + 2A_{ij},$$

where $z_i^2(t) = y_i'(t) V^{-1} y_i(t)$, $i = 1, 2, \ldots, k$.

We will need the following notation

$$Z_{ij} = z_j^2(t) - z_i^2(t), \quad n = T - p, \quad p = \max_{1 \le i \le k} p_i,$$

$$S_{ij}(n) = \sum_{t=1}^{n} Z_{ij}(t), \quad m_{ij} = E\{Z_{ij}(t) \mid \Theta = \Theta_i\},$$

$$\sigma_{ij}^2(n) = E\{(S_{ij}(n) - nm_{ij})^2 \mid \Theta = \Theta_i\}, \quad i, j = 1, 2, \ldots, k, \; j \neq i.$$

Using a central limit theorem for dependent random variables due to Ibragimov (1975) one may show the following result (see Krzyśko (1983)).

Theorem 2. *For all pairs (i, j), $i \neq j$, as $T \to \infty$*

$$(v_{ij}(X) - A_{ij} - \frac{1}{2}(T-p)m_{ij}) / \frac{1}{2}\sigma_{ij}(T-p)$$

is asymptotically, normally distributed with zero mean and unit variance.

4. Deviation of the distribution of the discriminant function $v_{ij}(X)$ from the asymptotic normal distribution. In the case ARMA $(p_i, 0)$ processes Theorem 2 shows that the limit distribution of $v_{ij}(X)$ is normal. Now, we will estimate the precision of this approximation according to the length T of the time series.

The expected value and the variance of $v_{ij}(X)$, $i, j = 1, 2, \ldots, k$, $j \neq i$, are

$$m_{ij}(N) = E\{v_{ij}(X) \mid \Theta = \Theta_i\} = k_{ij} + \sum_{n=1}^{N} b_{ij,n},$$

$$\sigma_{ij}^2(N) = \mathrm{Var}\{v_{ij}(X) \mid \Theta = \Theta_i\} = 2\sum_{n=1}^{N} b_{ij,n}^2 + \frac{1}{4}\sum_{n=1}^{n} \gamma_{ij,n}^2,$$

respectively.

Theorem 3. *The following inequality holds*

$$\sup_{y} \left| P\{v_{ij}(X) < y \mid \Theta = \Theta_i\} - \Phi\left(\frac{y - m_{ij}(N)}{\sigma_{ij}(N)}\right) \right| \le c(\Theta_i, \Theta_j, N),$$

where

$$c(\Theta_i, \Theta_j, N) = \frac{8.1 \max_{1 \le n \le N} |\lambda_{ij,n} - 1|}{\sqrt{2\pi}\, 6_{ij}^3(N)} \{1.8\,\sigma_{ij}^2(N) - \frac{1}{6}\sum_{n=0}^{N} (\lambda_{ij,n} - 1)^2\}, \tag{8}$$

$$i, j = 1, 2, \ldots, k, \; j \neq i.$$

Remark 3. The proof of this theorem is based on the Berry-Esséen type inequality given by Zolotarev (1967).

Remark 4. Theorem 3 with $p=1$ (i. e. $N=T$) reduces to the result due to Misiukas (1978). One should observe that the two constants which appear in the expression $c(\Theta_i, \Theta_j, T)$ of the Misiukas theorem are miscalculated. Instead 8.455 and 1.68 should be 8.1 and 1.8, respectively.

5. Numerical example. Let us consider the three classes of the two-dimensiona second-order autoregressive series with the parameters

$$\Theta_1=\left(\begin{bmatrix}-1.0 & -0.3\\ 3.3 & 1.0\end{bmatrix}, \begin{bmatrix}-0.02 & 0.0\\ 0.0 & -0.02\end{bmatrix}, \begin{bmatrix}0.0\\ 0.0\end{bmatrix}, \begin{bmatrix}1.0 & 0.5\\ 0.5 & 1.25\end{bmatrix}\right),$$

$$\Theta_2=\left(\begin{bmatrix}-1.0 & -1.0\\ 0.24 & 0.0\end{bmatrix}, \begin{bmatrix}0.0 & -0.05\\ 0.048 & 0.11\end{bmatrix}, \begin{bmatrix}20.0\\ 30.0\end{bmatrix}, \begin{bmatrix}1.0 & 0.5\\ 0.5 & 1.25\end{bmatrix}\right),$$

and

$$\Theta_3=\left(\begin{bmatrix}-1.0 & -2.0\\ 0.3 & 0.6\end{bmatrix}, \begin{bmatrix}-0.04 & 0.0\\ 0.035 & 0.03\end{bmatrix}, \begin{bmatrix}-20.0\\ 60.0\end{bmatrix}, \begin{bmatrix}1.0 & 0.5\\ 0.5 & 1.25\end{bmatrix}\right).$$

The values of $c(\Theta_i, \Theta_j, T)$ given by (8) according to the length T of the time series are given in Table 1.

From Table 1 we see that $c(\Theta_i, \Theta_j, T)\to 0$ if $T\to\infty$.

Table 1

The values of $c(\theta_i, \theta_j, T)$

T	$c(\theta_1, \theta_2, T)$	$c(\theta_1, \theta_3, T)$	$c(\theta_2, \theta_3, T)$
10	0.184	0.107	0.073
20	0.129	0.075	0.056
30	0.105	0.061	0.047
40	0.090	0.052	0.041
50	0.081	0.047	0.037

REFERENCES

1. T. W. Anderson. An Introduction to Multivariate Statistical Analysis. New York, 1958.
2. E. J. Hannan. Multiple Time Series. New York, 1970.
3. I. A. Ibragimov. A note on the central limit theorem for dependent random variables. *Teor. Verojatonost. i Primenen.*, 20, 1975, 134-139.
4. J. P. Imhof. Computing the distribution of quadratic forms in normalivariables. Biometrika, 48 1961, 419-426.
5. M. Krzyśko. Asymptotic distribution of the discriminant function. *Statist. Prob. Letters*, 1, 1983, 243-250.
6. R. Misiukas. Estimate of the probability of misclassification of autoregressive sequences. In: Statistical Problems of Control 30 (Institute of Mathematics and Cybernetics of the Academy of Siences of the Lithuanian SSR, Vilnius), 29-52.
7. V. M. Zolotarev. A sharpening of the inequality of Berry—Esseen. *Z. Wahrsch. verw. Geb.*, 8, 1967, 332-342.

Mickiewicz University,
Institute of Mathematics
60-769 Poznan, Poland

Receied 4. 10. 1985