ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 152, 2000

I. STATISTICAL MODELS

Mirosław Krzyśko, Waldemar Wołyński***

CLASSIFICATION INTO TWO POPULATIONS FOR TIME DEPENDENT OBSERVATIONS

Abstract. Optimal classification rules based on linear functions which maximize the area under the relative operating characteristic curve or which maximize the chosen probabilistic distance between two populations are studied here. We obtain an expression for the optimal linear discriminant function and show that the resulting procedure belongs to the Anderson-Bahadur admissible class. The asymptotic form of the discriminant function is also studied.

1. INTRODUCTION

A number of practical problems in the analysis of data reduce to classifying a realization of a stationary normal stochastic process as belonging to one or the other of two categories. Schumway (1982) has provided an extensive list of references and applications of discriminant analysis for time series. Applications listed there include discriminating between presumed earthquakes and underground nuclear explosions, the detection of a signal imbedded in a noise series, discriminating between different classes of brain wave recordings, and discriminating between various speakers or speech patterns on the basis of recorded speech data.

The admissible procedures for classification provided by the Neyman--Pearson theory as well as the Bayes' rule are based on the likelihood ratio. In the case of unequal covariance matrices this likelihood ratio depends on a quadratic function of observations. Unfortunately, the distribution of the quadratic discriminant function is very complicated. It involves the

* Adam Mickiewicz University, Faculty of Mathematics and Computer Science; Technical University of Zielona Góra.

** Adam Mickiewicz University, Faculty of Mathematics and Computer Science.

linear combination of non-central chi-square random variables so that computing error rates resulting from its use seems difficult. Hence, following Anderson and Bahadur (1962) we will consider a linear discriminant function.

2. THE RELATIVE OPERATING CHARACTERISTIC (ROC) CURVE AND ITS PROPERTIES

Let the T-dimensional time series $X = (X(0), X(1), \ldots, X(T-1))'$ be a realization of a stationary stochastic process from the population π_1 and let $Y = (Y(0), Y(1), \ldots, Y(T-1))'$ be an independent realization of a stationary stochastic process from the population π_2. Suppose that $X \sim N_T(\mu_1, \Sigma_1)$ and $Y \sim N_T(\mu_2, \Sigma_2)$, where $\mu_1 = (\mu_1(0), \mu_1(1), \ldots, \mu_1(T-1))'$ $\mu_2 = (\mu_2(0), \mu_2(1), \ldots, \mu_2(T-1))'$ and the covariance matrices $\Sigma_1 = (\sigma_1(|s-t|))$ and $\Sigma_2 = (\sigma_2(|s-t|))$, $s, t = 0, 1, \ldots, T-1$ are positive definite.

The parametres are assumed to be known. The phrase "X is stationary with mean μ and covariance matrix Σ" means that "$(X-\mu)$ is stationary with zero mean and covariance matrix Σ".

For each $a \in R^T$, $a \neq 0$, and each $c \in R$ let $R(a'X, c)$ denote the discriminant role that assigns the time series X to the population π_1 if $a'X \leqslant c$ and to the population π_2 if $a'X > c$.

We have

$$a'X \sim N(a'\mu_1, a'\Sigma_1 a), \quad a'Y \sim N(a'\mu_2, a'\Sigma_2 a).$$

Let

$$F_1(c) = P(a'X \leqslant c) = \Phi\left(\frac{c - a'\mu_1}{(a'\Sigma_1 a)^{1/2}}\right)$$

and

$$F_2(c) = P(a'Y \leqslant c) = \Phi\left(\frac{c - a'\mu_2}{(a'\Sigma_2 a)^{1/2}}\right),$$

where Φ is the distribution function of a $N(0, 1)$ randon variable.

Each discriminant rule is characterized in terms of the two probabilities of misclassification or in terms of the two conditional probabilities of the correct classification.

The probability of misclassifying an observation when it comes from the first population is

$$P(\pi_2|\pi_1) = P(a'X > c) = 1 - P(a'X \leqslant c) = 1 - F_1(c)$$

and the probability of misclassifying an observation when it comes from the second population is

$$P(\pi_1|\pi_2) = P(a'Y \leqslant c) = F_2(c).$$

The corresponding conditional probabilities of the correct classification are equal to

$$P(\pi_1|\pi_1) = P(a'X \leqslant c) = F_1(c)$$

and

$$P(\pi_2|\pi_2) = P(a'Y > c) = 1 - F_2(c).$$

The probability $P(\pi_1|\pi_1)$ is called the specificity of the discriminant rule and the probability $P(\pi_2|\pi_2)$ is called the sensitivity of the discriminant rule.

In the parametric representation, the curve of the form

$$x = F_1(c),\ y = 1 - F_2(c),\ -\infty \leqslant c \leqslant \infty$$

is called the Relative Operating Characteristic (ROC) curve of the class rules $R(a'X,\ \cdot)$. A plot of the ROC curve is given in Fig. 1.

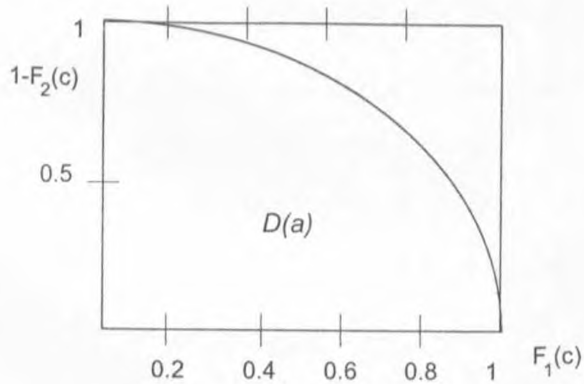

Fig. 1. A plot of the ROC curve

The area $D(a)$ under the ROC curve is the index which evaluates the accuracy of a class of discriminant rules $R(a'X,\ c)$. A large area indicates that the linear combination $a'X$ discriminates well between the two populations being compared.

The area $D(a)$ under the ROC curve has a simple probabilistic interpretation.

Theorem 1. The area $D(a)$ under the ROC curve is equal to the probability that the random variable $a'Y$ is stochastically larger than the random variable $a'X$.

$$D(a) = P(a'Y > a'X) = \Phi\left(\frac{a'\delta}{(a'\Sigma_1 a + a'\Sigma_2 a)^{1/2}}\right) \tag{1}$$

where

$$\delta = \mu_2 - \mu_1.$$

A proof of the theorem is in Krzyśko (1998) and it will be published elsewhere.

3. THE CLASSIFICATION RULE

A comparison of the area under the different ROC curves may be used to determine which linear combination $a'X$ is best.

Hence, we want to find the linear discriminant function for which area under the corresponding ROC curve is maximized. This maximal area is the ROC criterion, measuring how well the vector of characteristics distinguishes between the two populations.

Theorem 2. The vector a which maximizes the area $D(a)$ given by (1) has the form

$$a = (\Sigma_1 + \Sigma_2)^{-1}\delta \tag{2}$$

For a of the form (2) we have

$$D(a) = \Phi((\delta'(\Sigma_1 + \Sigma_2)^{-1}\delta)^{1/2}) \tag{3}$$

A proof of the theorem is in Krzyśko [1998] and it will be published elsewhere.

Remark 1. Since $\delta'(\Sigma_1 + \Sigma_2)^{-1}\delta \geqslant 0$ and $\Phi((\delta'(\Sigma_1 + \Sigma_2)^{-1}\delta)^{1/2}) \geqslant 1/2$, we have

$$\frac{1}{2} \leqslant D(a) \leqslant 1.$$

The area $D(a)$ close to 1 indicates that the T characteristics distinguishes well between the populations π_1 and π_2, and $D(a)$ close to 1/2 indicates that these two populations are not well separated.

4. THE SECOND CLASSIFICATION RULE

Now we would like to pass on to the second method of finding the optimal linear classification rule. We would like to find the linear classification procedure which maximizes the function

$$\rho(g_1, g_2),$$

where g_1 is the density function of the random variable $a'X$, g_2 is the density function of the random variable $a'Y$ and ρ is the Matusita (1956) distance, or the Morisita (1959) distance or the Kullback distance (Kullback, Leibler, 1951).

The Matusita distance is defined as follows.

Let P_1, P_2 be distributions defined on the p-dimensional Euclidean space R^p and denote by f_1, f_2 their densities with respect to the Lebesgue measure in R^p. The Matusita distance has the following form

$$\rho_1(f_1, f_2) = -\ln \int_{R^p} \sqrt{f_1(x)f_2(x)dx}.$$

If f_1 and f_2 are square integrable with respect to the Lebesgue measure in R^p then the Morisita distance has the form

$$\rho_2(f_1, f_2) = -\ln \frac{2\Delta(f_1, f_2)}{\Delta(f_1)+\Delta(f_2)},$$

$$\Delta(f_1, f_2) = \int_{R^p} f_1(x)f_2(x)dx,$$

$$\Delta(f_i) = \Delta(f_i, f_i), \; i = 1, 2.$$

It is obvious that $\rho_2(f_1, f_2)$ is closely related to the usual distance

$$\left(\int_{R^p} (f_1(x)-f_2(x))^2 dx\right)^{\frac{1}{2}}.$$

The Kullback distance has the form:

$$\rho_3(f_1, f_2) = \int_{R^p} [f_1(x)-f_2(x)] \ln \frac{f_1(x)}{f_2(x)} \, dx.$$

If g_1 is the probability density of $a'X$ and g_2 is the probability density of $a'Y$ then one can easily show that

$$\rho_1(g_1, g_2) = \frac{1}{4}\frac{(a'\delta)}{a'(\Sigma_1+\Sigma_2)a} + \frac{1}{2}\ln\left[\frac{1}{2}a'(\Sigma_1+\Sigma_2)a\right] - \frac{1}{4}\ln(a'(\Sigma_1 a)(a'\Sigma_2 a),$$

$$\rho_2(g_1, g_2) = \frac{1}{2}\frac{(a'\delta)^2}{a'(\Sigma_1+\Sigma_2)a} + \frac{1}{2}\ln\left[\frac{1}{2}a'(\Sigma_1+\Sigma_2)a\right] + \ln\left[(a'(\Sigma_1 a)^{\frac{1}{2}} + (a'\Sigma_2 a)^{\frac{1}{2}}\right]$$

$$-\frac{1}{2}\ln[(a'(\Sigma_1 a) + (a'\Sigma_2 a)] - \ln(2\sqrt{2}),$$

$$\rho_3(g_1, g_2) = \frac{1}{2}\left[(a'\delta)^2\left(\frac{1}{a'\Sigma_1 a} + \frac{1}{a'\Sigma_2 a}\right) + a'(\Sigma_1 - \Sigma_2)a\left(\frac{1}{a'\Sigma_2 a} - \frac{1}{a'\Sigma_1 a}\right)\right].$$

These three distances are invariant under scalar multiplication of a.

Theorem 3. (K r z y ś k o, W o ł y ń s k i, 1997). The vector a which maximizes these three distances has the form

$$a = (\Sigma_1 + \theta\Sigma_2)^{-1}\delta \tag{4}$$

where

$$\theta = \frac{t_{i2}(a)}{t_{i1}(a)}, \; i = 1, 2, 3 \tag{5}$$

are such that the matrices $\Sigma_1 + \theta\Sigma_2$ are nonsingular and

$$t_{11}(a) = A_1 + \frac{1}{a'\Sigma_1 a}, \; t_{12}(a) = A_1 + \frac{1}{a'\Sigma_2 a},$$

$$A_1 = \frac{(a'\delta)^2}{[a'(\Sigma_1+\Sigma_2)a]^2} - \frac{2}{a'(\Sigma_1+\Sigma_2)a},$$

$$t_{21}(a) = A_2 - B_2(a'\Sigma_1 a)^{-\frac{1}{2}} + (a'\Sigma_1 a)^{-1}, \; t_{22}(a) = A_2 - B_2(a'\Sigma_2 a)^{-\frac{1}{2}} + (a'\Sigma_2 a)^{-1},$$

$$A_2 = \frac{(a'\delta)^2}{[a'(\Sigma_1+\Sigma_2)a]^2} - [a'(\Sigma_1+\Sigma_2)a]^{-1}, \; B_2 = \left((a'\Sigma_1 a)^{\frac{1}{2}} + (a'\Sigma_2 a)^{\frac{1}{2}}\right)^{-1},$$

$$t_{31}(a) = A_3 - (a'\Sigma_2 a)^{-1}, \; t_{32}(a) = B_3 - (a'\Sigma_1 a)^{-1},$$

$$A_3 = \frac{(a'\delta)^2 + a'\Sigma_2 a}{(a'\Sigma_1 a)^2}, \quad B_3 = \frac{(a\delta)^2 + a'\Sigma_1 a}{(a'\Sigma_2 a)^2}.$$

It is clear that the equation (4) is an implicit equation in a. Hence an iterative procedure must be employed to solve for a.

Since Σ_1 and Σ_2 are positive definite matrices by assumption, there always exists a non-singular matrix P such that $\Sigma_1 = P'\ P$, $\Sigma_2 = P'\Lambda P$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ and $\lambda_l' s\,(l = 1, \ldots, p)$ are the characteristic roots of $\Sigma_2\Sigma_1^{-1}$.

Then

$$a'\Sigma_1 a = \beta'\beta, \quad a'\Sigma_2 a = \beta'\Lambda\beta, \quad a'\delta = \beta'\eta,$$

where

$$\beta = Pa \quad \text{and} \quad \delta = P'\eta.$$

Now, β can be written as

$$\beta = Pa = P(\Sigma_1 + \theta\Sigma_2)^{-1}\delta = P(p'P + \theta P'\Lambda P)^{-1}\delta = (1 + \theta\Lambda)^{-1}\eta.$$

Thus equation (5) reduces to

$$\theta = \psi_i(\theta), \quad i = 1,\ 2,\ 3 \tag{6}$$

where

$$\psi_1(\theta) = \frac{\beta'\beta}{\beta'\Lambda\beta} + (1 - \theta)A_1(\beta'\beta),$$

$$A_1 = \frac{(\beta'\eta)^2}{[\beta'(I + \Lambda)\beta]^2} - \frac{2}{\beta'(I + \Lambda)\beta},$$

$$\psi_2(\theta) = \left[\frac{B_2}{A_2}(\beta'\beta)^{-\frac{1}{2}} - (A_2\beta'\beta)^{-1}\right]\theta + 1 - \frac{B_2}{A_2}(\beta'\Lambda\beta)^{-\frac{1}{2}} + [A_2(\beta'\Lambda\beta)]^{-1},$$

$$A_2 = \frac{(\beta'\eta)^2}{[\beta'(I + \Lambda)\beta]^2} - (\beta'(I + \Lambda)\beta)^{-1}, \quad B_2 = \left((\beta'\beta)^{\frac{1}{2}} + (\beta'\Lambda\beta)^{\frac{1}{2}}\right)^{-1},$$

$$\psi_3 = (\theta) = A_3(\beta'\Lambda\beta)\theta - B_3(\beta'\Lambda\beta) + (\beta'\Lambda\beta)(\beta'\beta)^{-1},$$

$$A_3 = \frac{(\beta'\eta)^2 + \beta'\Lambda\beta}{(\beta'\beta)^2}, \quad B_3 = \frac{(\beta'\eta)^2 + \beta'\beta}{(\beta'\Lambda\beta)^2}.$$

The convergence of the iteration process follows from the following theorem.

Theorem 4. (Vilenkin 1979, p. 69). Let the function $\theta = \psi(\theta)$ be the mapping of the interval $[a, b]$ into itself and suppose in this interval the inequality $|\psi'(\theta)| < q$, where $q < 1$, holds. Then for any point θ_0 of the interval $[a, b]$ the sequence of points $\theta_0, \theta_1, ..., \theta_n$, where $\theta_{n+1} = \psi(\theta_n)$, converges to the root of the equation $\theta = \psi(\theta)$.

Roughly speaking this theorem says that the process of successive approximations enables us to find those roots θ of the equation $\theta = \psi(\theta)$ for which the inequality $|\psi'(\theta)| < 1$ is satisfied.

In our case one can easily check on which interval or intervals of the real line the condition $|\psi'(\theta)| < 1$ is satisfied.

5. ADMISSIBILITY OF PROCEDURES

Each classification procedure is characterized in terms of the two probabilities of misclassification. The probability of misclassifying an observation when it comes from the first population is

$$P(\pi_2|\pi_1) = 1 - F_1(c) = 1 - \Phi\left(\frac{c - a'\mu_1}{(a'\Sigma_1 a)^{\frac{1}{2}}}\right)$$

and the probability of misclassifying an observation when it comes from the second population is

$$P(\pi_1|\pi_2) = F_2(c) = \Phi\left(\frac{c - a'\mu_{21}}{(a'\Sigma_2 a)^{\frac{1}{2}}}\right).$$

It is desired to make these probabilities small. One classification procedure is better than another if each probability of misclassification of the former is not greater than the corresponding one of the latter and at least one is less. A procedure is admissible if there is no other procedure which is better.

The following theorem is true.

Theorem 5. (Krzyśko, Wołyński, 1997). The linear classification procedure defined by (4) and

$$c = a'\mu_1 + a'\Sigma_1 a = a'\mu_2 - \theta(a'\Sigma_2 a)$$

for any θ such that $\Sigma_1 + \theta\Sigma_2$ is positive definite is admissible within the class of linear procedures.

This result follows from the Anderson-Bahadur's theorem on admissible class of linear procedures (Anderson, Bahadur, 1962).

Remark 2. The linear discriminant procedure for which the area under the corresponding ROC curve is maximized is admissible. In our case $\theta = 1$ and the matrix $\Sigma_1 + \Sigma_2$ is positive definite.

Remark 3. If $\Sigma_1 = \Sigma_2 = \Sigma$, then for Matusita, Morisita and Kullback distances $\theta = 1$ and

$$a'X - c = \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}\left[X - \frac{1}{2}(\mu_1 - \mu_2)\right].$$

Hence all these distances and the ROC curve give the same well-known Fisher linear discriminant function.

Remark 4. The two probabilities of misclassification resulting from the use of the linear admissible classification procedures have the following form

$$P(\pi_2|\pi_1) = 1 - \Phi(\sqrt{a'\Sigma_1 a}),\ P(\pi_1|\pi_2) = 1 - \Phi(\theta\sqrt{a'\Sigma_2 a}).$$

Remark 5. We have

$$a = (\Sigma_1 + \theta\Sigma_2)^{-1}\delta$$

or

$$(\Sigma_1 + \theta\Sigma_2)a = \delta,$$

or

$$a'(\Sigma_1 + \theta\Sigma_2)a = a'\delta$$

If $\Sigma_1 + \theta\Sigma_2 > 0$ then the discriminant procedure is admissible. The matrix $\Sigma_1 + \theta\Sigma_2$ is positive definite if $a'(\Sigma_1 + \theta\Sigma_2)a > 0$ for all $a \neq 0$. Hence $a'\delta \neq 0$ for all $a \neq 0$ or $a'\mu_1 \neq a'\mu_2$ or $\mu_1 \neq \mu_2$. This means that every admissible linear discriminant rule makes some use of the fact that $\mu_1 \neq \mu_2$.

6. THE ASYMPTOTIC FORMS OF THE LINEAR DISCRIMINANT FUNCTIONS

We now consider a spectral approximation to the linear discriminant functions under the following assumptions:

1. In the population π_1 the stationary process $Z(t)$ has covariance function

$$\sigma_j(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda k} h_j(\lambda)(d\lambda)$$

with $h_j(\lambda)(j = 1, 2)$ assumed to be continuous, positive $[-\pi, \pi]$ absolutely integrable spectral densities.

We note that for every admissible linear classification procedure the matrix $\Sigma_1 + \theta\Sigma_2$ is positive definite. For stationary process, this implies that the spectral density

$$h_\theta(\lambda) = h_1(\lambda) + \theta h_2(\lambda)$$

is strictly positive for $\lambda \in [-\pi, \pi]$.

2. The covariance sequence $(\sigma_j(t))$ satisfies

$$\sum_{t=-\infty}^{\infty} |t|^{1+\beta}|\sigma_j(t)| < \infty$$

for $j = 1, 2$ and for some β, $0 < \beta < 1$.

3. The sequence of mean differences

$$\delta(t) = \mu_2(t) - \mu_1(t)$$

satisfies

(i)

$$\sup_t |\delta(t)| < \infty$$

and

(ii)

$$\rho_T(\tau) = T^{-1} \sum_{t=0}^{T-1-|\tau|} \delta(t + |\tau|)\delta(t)$$

has a limit given by

$$\rho(\tau) = \lim_{T\to\infty} \rho_T(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda\tau}\, dM(\lambda),$$

where $M(\lambda)$ is a monotone nondecreasing function uniquely defined by the conditions $M(-\pi) = 0$ and continuity from the right.

Under the assumptions stated above we have the following theorem.

Theorem 6. (Krzyśko, Wołyński, 1997). If $h_\theta(\lambda) = h_1(\lambda) + \theta h_2(\lambda) > 0$ for $\lambda \in [-\pi, \pi]$, then

$$\lim_{T\to\infty} T^{-1}\rho_1(g_1, g_2) = \frac{1}{4}G(\theta),$$

$$\lim_{T\to\infty} T^{-1}\rho_2(g_1, g_2) = \frac{1}{2}G(\theta),$$

$$\lim_{T\to\infty} T^{-1}\rho_3(g_1, g_2) = \frac{1}{2}H(\theta),$$

where

$$G(\theta) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{dM(\lambda)}{h_\theta(\lambda)}\right)^2 \Bigg/ \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_1(\lambda)+h_2(\lambda)}{h_\theta^2(\lambda)}\,dM(\lambda)\right) \tag{7}$$

$$H(\theta) = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{dM(\lambda)}{h_\theta(\lambda)}\right)^2 \left[\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_1(\lambda)}{h_\theta^2(\lambda)}\,dM(\lambda)\right)^{-1} + \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_2(\lambda)}{h_\theta^2(\lambda)}\,dM(\lambda)\right)^{-1}\right] \tag{8}$$

The optimal vector a in the sense of maximizing the Matusita distance asymptotically or Morisita distance asymptotically has the form (4) where θ is the value for which the function $G(\theta)$ has a global maximum.

The optimal vector a in the sense of maximizing the Kullback distance asymptotically has the form (4) where θ is the value for which the function $H(\theta)$ has a global maximum.

The following theorem characterizes the value of θ for which the functions $G(\theta)$ and $H(\theta)$ have a global maximum.

Theorem 7. (Krzyśko, Wołyński, 1997). The functions $G(\theta)$ and $H(\theta)$ defined in (7) and (8) have a global maximum at $\theta = 1$.

From the Theorem 7, the asymptotically optimal vector a is given by

$$a_\infty = (\Sigma_1 + \Sigma_2)^{-1}\delta.$$

The vector a which maximizes the area under the ROC curve has the same form.

7. AN ILLUSTRATIVE EXAMPLE

Let $X(t) - \mu(t) = Z(t)$ where $E(X(t)) = \mu(t)$ and let $\{Z(t), t \geqslant 0\}$ be a stationary normal process with $E(Z(t)) = 0$ which satisfies the assumption 1) Take $E_{\pi_1}(X(t)) = \cos(\pi/2)t$ and $E_{\pi_2}(X(t)) = 0$. Then $\delta(t) = \cos(\pi/2)t$ and

$$\lim_{T\to\infty} T^{-1} \sum_{t=0}^{T-1-|\tau|} \delta(1+|\tau|)\delta(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda\tau} dM(\lambda),$$

where $M(\lambda)$ is a step function having jumps at $\pm(\pi/2)$ of height $\pi/2$. Hence the assumption 3) is satisfied.

Let $\{\varepsilon(t)\}$ be a normal process with $E(\varepsilon(t))=0$ and

$$\mathrm{Cov}(\varepsilon(t),\ \varepsilon(t+k)) = \begin{cases} 0, & k=1,\ 2,\ \ldots, \\ 1, & k=0. \end{cases}$$

Let now $\{Z(t),\ t \geqslant 0\}$ be the second order autoregressive AR(2) process i.e. $Z(t)$ satisfies

$$Z(t) = \beta_1 Z(t-1) + \beta_2 Z(t-2) + \varepsilon(t).$$

The AR(2) process is always invertible. The stationary condition of the AR(2) process is given by the following inequalities

$$\beta_2 + \beta_1 < 1,$$

$$\beta_2 - \beta_1 < 1,$$

$$-1 < \beta_2 < 1.$$

The autocovariances of the AR(2) process are

$$\sigma(k) = \begin{cases} \dfrac{1-\beta_2}{1+\beta_2}\,\dfrac{1}{(1-\beta_2)^2-\beta_1^2}, & k=0, \\ \beta_1\sigma(k-1)+\beta_2\sigma(k-2), & k \geqslant 1. \end{cases}$$

Then it is easy to check that $(\sigma(k))_{k=0}^{\infty}$ satisfies the assumption 2).

Let

$$\pi_1 : Z(t) = 0.6Z(t-1) + 0.3Z(t-2) + \varepsilon(t),$$

$$\pi_2 : Z(t) = 0.8Z(t-1) + 0.3Z(t-2) + \varepsilon(t).$$

The solution of the implicit equation (6) for the Morisita distance and the value of the area under the ROC curve for these processes and $T = 10, 11, \ldots, 25$, are given in Tab. 1.

Table 1

The solutions θ of the equation (6) and the values of the area under the ROC curve for Morisita distance and AR(2) processes

T	θ	$D(a)$
10	0.786808	0.965627
11	0.778806	0.973469
12	0.817571	0.977832
13	0.808964	0.982712
14	0.839575	0.985532
15	0.844328	0.988685
16	0.857253	0.990494
17	0.852721	0.992534
18	0.871248	0.993716
19	0.867817	0.995051
20	0.882811	0.995826
21	0.879934	0.996705
22	0.892438	0.997217
23	0.890074	0.997799
24	0.900613	0.998139
25	0.898603	0.998525

It is clear from Tab. 1 that approximating the solution of (6) by $\theta = 1$, becomes increasingly accurate as T becomes larger. The Morisita distance gives the best results. From this Table we see also that if T is increasing then $D(a) \to 1$.

REFERENCES

Anderson T. W., Bahadur R. R. (1962), *Classification into Two Multivariate Normal Distributions with Different Covariance Matrices*, Ann. Math. Statist., **33**, 420–431.

Fisher R. A. (1936), *The Use of Multiple Measurements in Taxonomic Problems*, Ann. Eugen., **7**, 179–188.

Krzyśko M. (1998), *Linear Discriminant Functions which Maximize the Area under the ROC Curve*, (unpublished manuscript).

Krzyśko M., Wołyński W. (1997), *Linear Discriminant Functions for Stationary Time Series*, "Biometrical Journal", **39**, 955–973.

Kullback S., Leibler A. (1951), *On Information and Sufficiency*, Ann. Math. Statist., **22**, 79–86.

Matusita K. (1956), *Decision Rule, Based on Distance, for the Classification Problem*, Ann. Inst. Statist. Math., **8**, 67–77.

Morisita M. (1959), *Measuring of Interspecific Association and Similarity between Communities*, Mem. Fac. Sci. Kyushu Univ., Ser. **E**, 65–80.

Schumway R. H. (1982), *Discriminant Analysis for Time Series*, [in:] *Handbook of Statistics*, Vol. 2, P. R. Krishnaiah and L. N. Kanal (eds), North-Holland, Amsterdam, 1–46.

Van Belle G., Ahmad I. (1974), *Measuring Affinity of Distributions*, [in:] *Reliability and Biometry. Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling (eds), SIAM, Philadelphia, 651–668.

Vilenkin N. Ya. (1979), *Method of Successive Approximations*, Mir Publishers, Moscow.