Mathematical Modelling and Analysis
Volume 7 Number 1, 2002, pages 169–176

COMPARISON OF TWO ESTIMATORS OF MEAN FUNCTION IN LDA OF SPATIALLY CORRELATED GAUSSIAN DATA

J. ŠALTYTĖ and K. DUČINSKAS

Klaipėda University, Department of System Research

H. Manto 84, Klaipėda LT-5808

E-mail: jsaltyte@gmf.ku.lt, duce@gmf.ku.lt

Received September 1 2001; revised November 5 2001

ABSTRACT

The problem of linear discriminant analysis of an observation of Gaussian random field into one of two populations is considered. In this paper we analyze the performance of the plug-in linear discriminant function, when unknown means are estimated from the training samples. The generalized least squares and the ordinary least squares estimators are used. Obtained asymptotic expansions for the expected error rate are compared numerically in the case of spherical models for population covariances.

1. INTRODUCTION

Let $\left\{Z(s) : s \in D \subset R^2\right\}$ be a univariate Gaussian random field having different means and factorized covariance matrices in populations Ω_1 and Ω_2. The model of $Z(s)$ in population Ω_l is

$$Z(s) = x_l^T(s)\,\beta_l + \varepsilon_l(s),$$

where $x_l^T(s) = (x_{l1}(s), \ldots, x_{lq}(s))$ is a $q \times 1$ vector of nonrandom regressors and $\beta_l = \left(\beta_l^1, \ldots, \beta_l^q\right)^T \in B$, $l = 1, 2$, are parameter vectors, B being an open subset of R^q. Assume, that $\left\{\varepsilon_l(s) : s \in D \subset R^2\right\}$ is a zero-mean intrinsically stationary random Gaussian field with spatial covariance defined by a parametric model $cov\left\{\varepsilon_l(t), \varepsilon_l(s)\right\} = \sigma(t - s; \theta_l)$ for all $t, s \in D$, where $\theta_l \in \Theta$ is a $p \times 1$ parameter vector, Θ being an open subset of R^p, $l = 1, 2$. We restricted our attention to the homoscedastic models, i.e. $\sigma(0; \theta) = \sigma^2$, for any $\theta \in \Theta$. Then in Ω_l the mean function at location s is $\mu_l(s) = x_l^T(s)\,\beta_l$ and

the spatial covariance function in $cov\{\varepsilon_l(t), \varepsilon_l(s)\} = \sigma^2 c(t-s;\theta_l)$, where $c(t-s;\theta_l)$ is the spatial correlation function, $l = 1,2$. It is assumed that the function $c(t-s;\theta_l)$ is positive definite (see [6]).

Assume that, for all $t, s \in D$, $t \neq s$,

$$cov\{\varepsilon_1(t), \varepsilon_2(s)\} = 0. \tag{1.1}$$

Consider the problem of classification of an observation $Z_m = (Z(r_1), \ldots, Z(r_m))^T$ with $r_i \in D_0 \subset D$, $i = 1, \ldots, m$ into one of two populations specified above. Instead of considering the classification of m observations, let us consider one of the observations, say $Z(r)$. Then the probability density function (p.d.f.) of $Z(r)$ in Ω_l, $l = 1,2$, is

$$p_l\left(z(r); \mu_l(r), \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(z(r) - x^T(r)\beta_l\right)^2\right\}.$$

Under the assumption, that the populations are completely specified and for known prior probabilities of populations $\pi_1(r), \pi_2(r)$ $(\pi_1(r) + \pi_2(r) = 1)$, the Bayesian classification rule (BCR) $d_B(\bullet)$ minimizing the probability of misclassification (PMC) is

$$d_B(z(r)) = \arg \max_{\{l=1,2\}} \pi_l(r)\, p_l(z(r)), \tag{1.2}$$

where $\pi_l(r)$ is a prior probability of $\Omega_l, l = 1,2$.

Denote by P_B^r the PMC for BCR usually called the Bayesian error rate.

In practical applications the p.d.f. are usually unknown and must be estimated. Very often unknown parameters are estimated from the training samples T_1 and T_2 taken separately from Ω_1 and Ω_2, respectively. When estimators of unknown parameters are used, the plug-in version of BCR is obtained.

The performance of the plug-in version of the BCR when parameters are estimated from training samples with independent observations was widely investigated (see e.g. [8]). However, it has been founded that the assumption of independence is frequently violated. [4] investigated the performance of sample linear discriminant function (LDF) when training samples follow a stationary autoregressive process.

In this paper we shall consider the performance of the plug-in linear LDF when the unknown means are estimated from training sample following a Gaussian random field model described above assuming the spatial dependence parameters to be known. The maximum likelihood (ML) and ordinary least squares (OLS) procedures for the estimation of unknown means are used.

Suppose in region $D_1 \subset D$, $D_1 \cap D_0 = \emptyset$, we observe the training sample $T = \{T_1, T_2\}$ with $T_l = \{Z_{l1}, \ldots, Z_{lN_l}\}$, where $Z_{l\alpha} = Z(s_{l\alpha})$ denotes the α'th observation from Ω_l, $\alpha = 1, \ldots, N_l$, $l = 1,2$. Assume that D_1 is beyond the range (or the zone of influence) of D_0. Then $Z(r)$ is independent on T.

Let $\widehat{\mu}_l(r)$ be the estimator of $\mu_l(r)$, based on T. The plug-in rule $d_B(z(r); \widehat{\mu}_1(r), \widehat{\mu}_2(r))$ is obtained by replacing the parameters in (1.2) with their estimators.

Then the corresponding discriminant function W^r also known as the plug-in LDF is

$$W^r = \frac{1}{\sigma^2}\left(z(r) - \frac{1}{2}\left(\widehat{\mu}_1(r) + \widehat{\mu}_2(r)\right)\right)\left(\widehat{\mu}_1(r) - \widehat{\mu}_2(r)\right) + g(r), \tag{1.3}$$

where $g(r) = \ln \frac{\pi_1(r)}{\pi_2(r)}$.

DEFINITION 1.1. The actual error rate of $d_B(z(r); \widehat{\mu}_1(r), \widehat{\mu}_2(r))$ is defined as

$$P^r\left(\widehat{\mu}_1(r), \widehat{\mu}_2(r)\right) \triangleq \sum_{l=1}^{2} \pi_l(r) \tag{1.4}$$
$$\times \int \left(1 - \delta\left(l, d_B\left(z(r); \widehat{\mu}_1(r), \widehat{\mu}_2(r)\right)\right) p_l\left(z(r); \mu_l(r), \sigma^2\right)\right) dz(r).$$

In our case the actual error rate for $d_B(z(r); \widehat{\mu}_1(r), \widehat{\mu}_2(r))$ is defined as

$$P^r\left(\widehat{\mu}_1(r), \widehat{\mu}_2(r)\right) = \sum_{l=1}^{2} \pi_l(r) \tag{1.5}$$
$$\times \Phi\left((-1)^l \frac{\left(\mu_l(r) - \frac{1}{2}\left(\widehat{\mu}_1(r) + \widehat{\mu}_2(r)\right)\right)\left(\widehat{\mu}_1(r) - \widehat{\mu}_2(r)\right) + \sigma^2 g(r)}{\sigma\left|\widehat{\mu}_1(r) - \widehat{\mu}_2(r)\right|}\right)$$

where Φ is standard normal distribution function.

DEFINITION 1.2. The expectation of the actual error rate with respect to the distribution of T denoted as $E_T\left\{P^r\left(\widehat{\mu}_1(r), \widehat{\mu}_2(r)\right)\right\}$ is called the expected error rate (ER) for the $d_B(z(r); \widehat{\mu}_1(r), \widehat{\mu}_2(r))$ and expected error regret (EER) is defined by $EER = E_T\{P^r(\cdot,\cdot)\} - P_B$.

The goal of this paper is to find asymptotic expansions of ER associated with plug-in LDF for ML and OLS estimators. The case of normally distributed observations in training sample from the one of two populations with equal feature vector covariances was considered in [8]. The generalization for the case of arbitrary number of populations and regular class-conditional densities has been made in [2]. A similar problem of classifying the spatially distributed Gaussian observations with constant means is considered in [5].

In this paper we present the asymptotic expansion up to the order $O\left(N^{-2}\right)$, where $N = N_1 + N_2$, for the ER of classifying spatially distributed Gaussian observation to one of two populations with different means and common spatially factorized covariance. Terms of higher order are omitted from the asymptotic expansion since their contribution is in generally negligible [9]. The ML and OLS estimators of means are used in the plug-in version of the BCR. A set of calculations for a certain neighbourhood structure and spherical spatial correlation model is performed in order to estimate the plug-in BCR.

2. MAIN RESULTS

The expectation vector and the covariance matrix of the vectorized training sample T_l defined by $T_l^V = (Z_{l1}, \ldots, Z_{lN_l})^T$ are $\mu_l^V = (\mu_l(s_1), \ldots, \mu_l(s_{N_l}))^T$ and $\Sigma_l^V = \sigma^2 C_l$, respectively, where C_l is the spatial correlation matrix of order $N_l \times N_l$, whose $\alpha\beta$'th element is $c_{l;\alpha\beta} = c(s_\alpha - s_\beta)$, $\alpha, \beta = 1, \ldots, N_l$, $l = 1, 2$. Suppose σ^2 and C_l are known and $\widehat{\beta}_l^{\upsilon}$ is the estimator of β_l, based on T; here $\widehat{\mu}_l^{\upsilon}(s) = x_l^T(s)\,\widehat{\beta}_l^{\upsilon}$ and υ can take the value ML or OLS, $l = 1, 2$.

Put $C_l^{-1} = \left(c_l^{\alpha\beta}\right)$, $c_l^{\bullet\bullet} = \sum\limits_{\alpha,\beta=1}^{N_l} c_l^{\alpha\beta}$, $c_l^{\alpha\bullet} = \sum\limits_{\beta=1}^{N_l} c_l^{\alpha\beta}$, $l = 1, 2$. Let X_l be an $N_l \times q$ regressor matrix with j'th column $(x_{l,1j}, \ldots, x_{l,N_l j})^T$, where $x_{l,ij} = x_{l,j}(s_i)$, $j = 1, \ldots, q$, $i = 1, \ldots, N_l$, $l = 1, 2$.

Lemma 2.1. [1] *For $l = 1, 2$ the ML estimator of $\mu_l(s)$ based on T_l is*

$$\widehat{\mu}_l^{ML}(s) = x_l^T(s)\left(X_l^T C_l^{-1} X_l\right)^{-1} X_l^T C_l^{-1} T_l^V x_l(s). \tag{2.1}$$

It is known, that the OLS estimator of $\mu_l(s)$ based on T_l is

$$\widehat{\mu}_l^{OLS}(s) = x_l^T(s)\left(X_l^T X_l\right)^{-1} X_l^T T_l^V x_l(s). \tag{2.2}$$

It can be easily shown that $\widehat{\mu}_l^{\upsilon}(s)$ for finite N, $l = 1, 2$, $\upsilon = ML$ or OLS, have known exact distributions

$$\widehat{\mu}_l^{\upsilon}(s) \sim N\left(\mu_l(s), a_l^{\upsilon}\right), \tag{2.3}$$

where

$$a_l^{ML} = \sigma^2 x_l^T(s)\left(X_l^T C_l^{-1} X_l\right)^{-1} x_l(s) \tag{2.4}$$

and

$$a_l^{OLS} = \sigma^2 x_l^T(s)\left(X_l^T X_l\right)^{-1} X_l^T C_l X_l \left(X_l^T X_l\right)^{-1} x_l(s). \tag{2.5}$$

For simplicity we omit the superscript r in $P^r(\cdot,\cdot)$. Put $\Delta\widehat{\mu}_l^{\upsilon}(s) = \widehat{\mu}_l^{\upsilon}(s) - \mu_l(s)$, $l = 1,2$. Let $\varphi(\cdot)$ denotes the standard normal distribution density function. Denote by $P_l^{(1)} = \dfrac{\partial P(\cdot,\cdot)}{\partial \widehat{\mu}_l^{\upsilon}(s)}$ and $P_{k,l}^{(1)} = \dfrac{\partial^2 P(\cdot,\cdot)}{\partial \widehat{\mu}_k^{\upsilon}(s)\partial \widehat{\mu}_l^{\upsilon}(s)}$ the partial derivatives of $P\left(\widehat{\mu}_1^{\upsilon}(r), \widehat{\mu}_2^{\upsilon}(r)\right)$ up to the second order with respect to the corresponding parameters evaluated at $\widehat{\mu}_l^{\upsilon}(s) = \mu_l(s)$, $l = 1,2$.

Let $\lambda_{N_l}(C_l)$ be the largest eigenvalue of C_l and λ_1^l be the smallest eigenvalue of $X_l^T X_l$, $l = 1,2$.

Assumption 1. Assume, that $rank(X_l) = q$, for $l = 1,2$.

Assumption 2. Suppose, that $\lambda(C_l) < \kappa_l$, $0 < \kappa_l < \infty$, for $l = 1,2$.

Assumption 3. Suppose, that $\lambda_1^l \longrightarrow \infty$, as $N_l \longrightarrow \infty$, for $l = 1,2$.

Theorem 2.1. *Suppose, that assumptions* $1-3$ *hold for training samples* T_1, T_2. *Then the asymptotic expansion of the ER for the* $d_B\left(z(r); \widehat{\mu}_1^{\upsilon}(r), \widehat{\mu}_2^{\upsilon}(r)\right)$, *where* υ *can take the value ML or OLS, is*

$$E_T\left\{P\left(\widehat{\mu}_1^{\upsilon}(r), \widehat{\mu}_2^{\upsilon}(r)\right)\right\} = \sum_{l=1}^{2} \pi_l(r)\,\Phi\left(-\frac{\Delta(r)}{2} + (-1)^l \frac{g(r)}{\Delta(r)}\right) + \frac{1}{2\Delta(r)}$$

$$\times\, \pi_l(r)\varphi\left(-\frac{\Delta(r)}{2} - \frac{g(r)}{\Delta(r)}\right) \sum_{l=1}^{2} a_l^{\upsilon}\left(-\frac{\Delta(r)}{2} + (-1)^l \frac{g(r)}{\Delta(r)}\right)^2 + O\left(N^{-2}\right).$$

Proof. Without loss of generality we use the convenient canonical form of $\sigma^2 = 1$, $\mu_1(r) = \frac{\Delta(r)}{2}$, $\mu_2(r) = -\frac{\Delta(r)}{2}$ (see, [7]). By a Taylor expansion of the $P\left(\widehat{\mu}_1(r), \widehat{\mu}_2(r)\right)$, for $\upsilon = ML$ or OLS, about the true values of parameters we have for $P \equiv P\left(\widehat{\mu}_1^{\upsilon}(r), \widehat{\mu}_2^{\upsilon}(r)\right)$:

$$P = P_B + \sum_{l=1}^{2} P_l^{(1)} \Delta\widehat{\mu}_l^{\upsilon}(r) + \frac{1}{2}\sum_{k,l=1}^{2} P_{k,l}^{(2)} \Delta\widehat{\mu}_k^{\upsilon}(r)\,\Delta\widehat{\mu}_l^{\upsilon}(r) + O_3, \tag{2.6}$$

where $P_B = \sum_{l=1}^{2} \pi_l(r)\,\Phi\left(-\frac{\Delta(r)}{2} + (-1)^l \frac{g(r)}{\Delta(r)}\right)$ and O_3 is the third and higher order terms of $\Delta\widehat{\mu}_l^{\upsilon}(r)$ and their products. Since $P\left(\widehat{\mu}_1^{\upsilon}(r), \widehat{\mu}_2^{\upsilon}(r)\right)$ is minimised at $\mu_l(r) = (-1)^{l+1}\frac{\Delta(r)}{2}$, $l = 1,2$, then $P_l^{(1)} = 0$.

Using (2.1) – (2.5), for $l = 1,2$, under the independence of estimators $\widehat{\mu}_l^{\upsilon}(r)$ we have $E\left\{\widehat{\mu}_l^{\upsilon}(r)\right\} = 0$ and

$$E\left\{\left(\widehat{\mu}_l^{ML}(r)\right)^2\right\} = x_l^T(r)\left(X_l^T C_l^{-1} X_l\right) x_l(r)\,, \tag{2.7}$$

$$E\left\{\left(\widehat{\mu}_l^{OLS}(r)\right)^2\right\} = x_l^T(r)\left(X_l^T X_l\right)^{-1} X_l^T C_l X_l \left(X_l^T X_l\right)^{-1} x_l(r)\,, \tag{2.8}$$

$$E\left\{\widehat{\mu}_1^{\upsilon}(r)\,\widehat{\mu}_1^{\upsilon}(r)\right\} = 0\,. \tag{2.9}$$

Since $E\left\{\left(\widehat{\mu}_l^{\upsilon}(r)\right)^3\right\}=0$, $E\left\{\left(\widehat{\mu}_l^{ML}(r)\right)^4\right\}=3\left(x_l^T(r)\left(X_l^TC_l^{-1}X_l\right)x_l(r)\right)^2$,

$$E\left\{\left(\widehat{\mu}_l^{OLS}(r)\right)^4\right\}=3\left(x_l^T(r)\left(X_l^TX_l\right)^{-1}X_l^TC_lX_l\left(X_l^TX_l\right)^{-1}x_l(r)\right)^2,$$

under the assumptions 1–3, we have $x_l^T(r)\left(X_l^TC_l^{-1}X_l\right)x_l(r)=O\left(\frac{1}{\lambda_1^l}\right)$, $x_l^T(r)\left(X_l^TX_l\right)^{-1}X_l^TC_lX_l\left(X_l^TX_l\right)^{-1}x_l(r)=O\left(\frac{1}{N_l}\right)$, as $N_l\to\infty$, $l=1,2$, $\upsilon=ML$ or OLS. Note, that

$$P_{l,l}^{(2)}=\frac{\pi_1(r)}{\Delta(r)}\varphi\left(-\frac{\Delta(r)}{2}-\frac{g(r)}{\Delta(r)}\right)\left(-\frac{\Delta(r)}{2}+(-1)^{l+1}\frac{g(r)}{\Delta(r)}\right)^2.$$

By substituting estimators (2.1) and (2.2) into (2.6), taking the expectation of the right side of (2.6) and using (2.7) – (2.9), we complete the proof of the theorem. ■

The asymptotic EER for ML and OLS estimators are

$$AEER_{ML}=\frac{1}{2\Delta(r)}\pi_1(r)\,\varphi\left(-\frac{\Delta(r)}{2}-\frac{g(r)}{\Delta(r)}\right)\times$$
$$\times\sum_{l=1}^{2}a_l^{ML}\left(-\frac{\Delta(r)}{2}+(-1)^l\frac{g(r)}{\Delta(r)}\right)^2,$$
$$AEER_{OLS}=\frac{1}{2\Delta(r)}\pi_1(r)\,\varphi\left(-\frac{\Delta(r)}{2}-\frac{g(r)}{\Delta(r)}\right)\times$$
$$\times\sum_{l=1}^{2}a_l^{OLS}\left(-\frac{\Delta(r)}{2}+(-1)^l\frac{g(r)}{\Delta(r)}\right)^2.$$

These quantities are used for the evaluation of the performance of the LDF. The comparison of these two asymptotic regrets is given in the example below.

3. EXAMPLE

Here we compare the asymptotic expected error regrets when the ML and OLS estimators of unknown large-scale-variation (mean) parameters are used. The results of this comparison are presented in Tab. 1.

As an example consider the integer regular 2-dimensional lattice. There are six observations in the first training sample and nine in the second (Fig.1).

Assume, that the correlation functions are the same for both populations. Consider the spherical correlation function

$$c_s(|h|,\theta)=\begin{cases}\frac{\theta_1}{\theta_0+\theta_1}\left(1-\frac{3}{2}\frac{|h|}{\theta_2}+\frac{1}{2}\frac{|h|^3}{\theta_2^3}\right), & 0\le|h|\le\theta_2,\\ 1, & |h|=0,\\ 0, & |h|>\theta_2,\end{cases}$$

Table 1.
Comparison of the asymptotic expected error regrets (for $\pi_1 = 0.4$)

Δ	$AEER_{ML}$	$AEER_{OLS}$	$\frac{AEER_{ML}}{AEER_{OLS}}$
0.6	0.1176	0.1305	0.9011
1.0	0.0150	0.0184	0.8143
1.4	0.0171	0.0538	0.3179
1.8	0.0203	0.0868	0.2336
2.2	0.0216	0.1042	0.2071
2.6	0.0209	0.1072	0.1952
3.0	0.0188	0.0994	0.1888
3.4	0.0157	0.0851	0.1849
3.8	0.0124	0.0683	0.1823
4.2	0.0093	0.0516	0.1805
4.6	0.0066	0.0370	0.1791
5.0	0.0045	0.0251	0.1782

for nonnegative θ_0, θ_1, θ_2. The nugget effect is θ_0 and the sill is $\theta_0 + \theta_1$. For this model, observations more than θ_2 units apart are uncorrelated, so the range is θ_2.

Assume, that there is no nugget effect, i.e. $\theta_0 = 0$. It is obvious from the Fig. 1 that the appropriate range is $\theta_2 = 4.5$. [3] suggests represent mean as a polynomial function of coordinates of a specified order, that is $\mu_l(s) = A_l^T(s)\lambda_l$, where $A_l(s)$ is a vector of location co-ordinates of the point s and λ_l is a vector of trend surface parameters so that

$$A_l^T(s) = \left(a_{s1}^l, a_{s2}^l, \left(a_{s1}^l\right)^2, \left(a_{s2}^l\right)^2, a_{s1}^l a_{s2}^l, \ldots, \left(a_{s1}^l\right)^p \left(a_{s2}^l\right)^q\right), \tag{3.1}$$

where $a_{s1}^l a_{s2}^l$ defines the location of point s, and for $l = 1, 2$,
$\lambda_l = (\lambda_{10}, \lambda_{01}, \lambda_{20}, \lambda_{02}, \lambda_{11}, \ldots, \lambda_{pq})^T$. Let A_l be an $N_l \times k$ matrix with i'th column being defined in (3.1), $i = 1, \ldots, N_l$, $l = 1, 2$. This is so-called trend surface model. The sum $p + q = k$ represents the order of the trend surface: zero-order ($p + q = 0$) which is the same as the constant mean case; linear or first-order ($p + q = 1$) which generates a sloping plane surface; quadratic or second-order ($p + q = 2$); cubic or the third-order ($p + q = 3$), and so on. Successively higher orders generate surfaces of increasing complexity.

It is easy to see, that the trend surface model could be considered as a special case of regression model, where X_l is replaced with A_l, $l = 1, 2$. Here we use the trend surface model for the case $k = 1$, as an example.

As it was expected, the $AEER_{ML}$ and $AEER_{OLS}$ are decreasing when the distance increases. It is seen from the table that the expansion when the ML estimator is used is smaller than that obtained by using the OLS estimator. This difference is insignificant for close populations, however the ML estimator would be especially appropriate for the estimation of parameters, when populations are more separated.

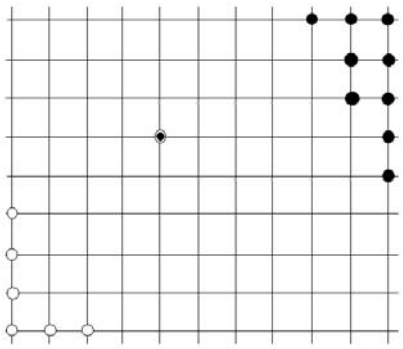

Figure 1. The positions of points in training samples

REFERENCES

[1] J. Šaltytė. *The asymptotic expansion of the expected risk for the LDA of spatially correlated Gaussian observations.* PhD thesis. Klaipėda university, 2001.

[2] K. Dučinskas. An asymptotic analysis of the regret risk in discriminant analysis under various training schemes. *Lith. Math. J.*, **37**(4), 337 – 351, 1997.

[3] R.P. Haining. *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge University Press, 1990.

[4] C.R.O. Lawoko and G.J. McLachlan. Discrimination with autocorrelated observations. *Pattern Recognition*, **18**(2), 145 – 149, 1985.

[5] K.V. Mardia. Spatial discrimination and classification maps. commun. *Statist.-Theor. Meth.*, **13**(18), 2181 – 2197, 1984.

[6] K.V. Mardia and R.J. Marshall. Maximum likelihood estimation of models for residual covariance and spatial regression. *Biometrika*, **71**, 135 – 146, 1984.

[7] G.J. McLachlan. *Discriminant Analysis and Statistical pattern recognition.* Wiley& Sons, 1992.

[8] M. Okamoto. An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, **34**, 1286 – 1301, 1963.

[9] M.J. Scherwish. Asymptotic expansions for correct classification rates in discriminant analysis. *The Annals of Statistics*, **9**(5), 1002 – 1009, 1981.

Dviejų vidurkio įvertinių palyginimas tiesinėje erdvėje koreliuotų Gauso stebėjimų diskriminantinėje analizėje

J. Šaltytė, K. Dučinskas

Straipsnyje sprendžiamas atsitiktinio Gauso lauko stebėjimų tiesinės diskriminantinės analizės uždavinys dviejų klasių atveju. Gauti pirmos eilė asimptotiniai tikėtinos klasifikavimo klaidos skleidiniai atvejui, kai į Bajeso klasifikavimo taisyklę įstatome maksimalaus tikėtinumo bei empirinį vidurkių įverčius. Atliktas skaitinis asimptotinių klasifikavimo klaidų palyginimas tam tikrai kaimynystės schemai bei sferinei koreliacijų funkcijai.