INFORMATICA, 2015, Vol. 26, No. 4, 557–568

DOI: http://dx.doi.org/10.15388/Informatica.2015.64

Actual Error Rates in Classification of the T-Distributed Random Field Observation Based on Plug-in Linear Discriminant Function

Kęstutis DUČINSKAS[1], Eglė ZIKARIENĖ[1,2]*

[1]*Department of Mathematics and Statistics, Klaipeda University*
H. Manto 84, 92294 Klaipeda, Lithuania
[2]*Recognition Processes Department, Institute of Mathematics and Informatics, Vilnius University*
Akademijos 4, 08663 Vilnius, Lithuania
e-mail: kestutis.ducinskas@ku.lt, egle.zikariene@gmail.com

Received: August 2014; accepted: March 2015

Abstract. In current paper a problem of classification of T-distributed random field observation into one of two populations specified by common scaling function is considered. The ML and LS estimators of the mean parameters are plugged into the linear discriminant function. The closed form expressions for the Bayes error rate and the actual error rate associated with the aforementioned discriminant functions are derived. This is the extension of one for the Gaussian case. The actual error rates are used to evaluate and compare the performance of the plug-in discriminant function by means of Monte Carlo study.

Key words: T-distributed random field, Bayes rule, spatial correlation, scaling function, actual error rate.

1. Introduction

Discriminant analysis (DA) (sometimes called supervised classification) traditionally assumes that observations to be classified are normally distributed, independent and, most often, identically distributed. Due to their mathematical tractability, Gaussian models have received the majority of the attention within the statistical modeling literature. However, according to numerous authors, not all data behave as a realization of a Gaussian distribution (see Roislien and Omre, 2011; Batsidis and Zografos, 2011). A well-known insufficiency of the normal distributions are their light tails. To tackle insufficiencies of the normal distributions, there has been an intense research in the use of non-normal distributions and there are several other parametric classes of multivariate distributions to choose from. The class of elliptically (spherically) contoured distributions is a particularly appealing family of multivariate symmetric distributions with simple density functions and possesses properties that provide a useful competitor of the multivariate normal model.

*Corresponding author.

This family of multivariate distributions, includes as the particular cases, the multivariate normal, multivariate T distribution, Pearson type II and VII, multivariate symmetric Kotz type distribution, scale mixtures of normal, etc. The multivariate T-distribution are suitable models to formulate and describe the random phenomenon involving high probability in the tails. To this respect T-distribution is a particularly useful model in, for instance, economics and actuarial sciences and many other disciplines (Sutradhar and Ali, 1986). Another motivation for considering of the multivariate T-distribution is its widely recognized capability to handle outliers more readily than multivariate Gaussian distribution. In observing spatial phenomena it is sometimes found that the T-distributed random field model is particularly useful whenever multiple, sparsely sampled realizations of the random field are available. The known estimation, discrimination and simulation methods for independent observations from multivariate T-distributions are reviewed by Nadarajah and Kotz (2008). Andrews *et al.* (2011) proposed classification technique of independent (or uncorrelated) observations based on mixtures of multivariate T-distributions. However, in practical situations with spatially distributed data the observations often are not independent. Data that are close together in space are likely to be correlated. Thus, is very important to include spatial dependencies in the prediction and classification problem. Kim and Mallick (2003) considered spatial prediction problems using the elliptical distribution. Batsidis and Zografos (2011) derived the asymptotic approximation of the distribution function for the probabilities of misclassification of elliptic random field observations. However their approach leads to the expression containing implicit function and they did not explore approximations and estimators of the expected error rate. We extend their work by considering empirical estimator of the expected error rate (ERR).

The current paper is concerned with Bayes rule (BR) that is an optimal classification rule in the sense of minimum overall misclassification probability in the case of completely specified populations. However, the complete statistical certainty of populations is usually not possible. Training sample is required for the estimation of the probabilistic characteristics of both populations. These estimators are plugged into the BR.

Many authors have investigated the performance of the plug-in version of the BR when the parameters are estimated from the training samples with independent observations, or training samples where observations are temporally dependent (see Okamoto, 1963; Lawoko and McLachlan, 1985; Shutoh, 2012). McLachlan (2004) have given a good review of the work done in this field. Switzer (1980) was the first to treat classification of spatial data, a work that was extended by Mardia (1984). However, neither of these authors analyzed the error rate of classification. Šaltytė and Dučinskas (2002) derived an asymptotic expansion of the expected error rate when classifying the observation of the univariate Gaussian random field into one of two classes with different regression mean models and common variance. This result was extended to multivariate spatial-temporal regression model in Šaltytė-Benth and Dučinskas (2005).

The first extensions to the case when spatial correlations between Gaussian observations to be classified and observations in training sample are not assumed equal to zero is done in Dučinskas (2009) and Dučinskas and Stabingienė (2011). Dučinskas *et al.* (2015) also proposed a classification procedure for spatially correlated Gaussian observations for a case of more than two classes.

In the current paper a problem of classification of T-distributed random field observations into one of two populations specified by different parametric linear mean models and common known stationary dependence function is considered. The formula for Bayes error rate is derived. The maximum likelihood (ML) and ordinary least squares (OLS) estimators of the mean parameters obtained from training sample are plugged in the Bayes discriminant function corresponding to the BR. The closed form expressions of the actual error rates associated with the aforementioned plug-in linear discriminant functions are derived. These expressions are used in constructing the empirical estimator of the expected error rate by means of Monte Carlo simulations. This estimator is used for the evaluation and comparison of performances of the proposed classification procedures. This is the extension of the results previously obtained for the Gaussian case.

The outline of paper is as follows. In Section 2 the main concepts concerning multivariate T-distribution and T-distributed random field are presented. The exact formulas for the Bayes error rate and actual error rate are derived in Section 3. MC study of obtained analytical results and conclusions are presented in Section 4. Bulk density of the log observations from the wells in Gullfaks field in the North are modeled by a T-random field (see Roislien and Omre, 2011). These observations need to be classified to the different layer sections. This is an example of the real situation where obtained theoretical results could be effectively applied.

2. The Main Concepts and Definitions

The main objective of this paper is to classify the observations of T-distributed random field (TRF), that is a random field defined by the multivariate T-distribution.

Definition 1. A random vector $Z \in R^n$ is multivariate T-distributed, denoted by $Z \sim T_n(\mu, \Omega, m)$, with a mean vector $\mu \in R^n$, a positive definite $n \times n$ scaling matrix Ω and degrees of freedom $m > 0$ if its probability density function is $f(z) = \Gamma((m+n)/2)|\Omega|^{-1/2}[1 + (z-\mu)'\Omega^{-1}(z-\mu)/m]^{\frac{-(m+n)}{2}} / \left(\Gamma(m/2)(m\pi)^{n/2}\right)$ where $\Gamma(\bullet)$ is the gamma function.

This definition specifies a spherical-symmetric pdf centered at μ with Ω controlling scale and multivariate dependence, while m controls the tail behavior (Mardia *et al.*, 1979).

The moments of the multivariate T-distributed random vector are summarized below: $E(Z) = \mu, m \geqslant 2; cov(Z) = \Sigma = m\Omega/(m-2), m \geqslant 3$; while for m less than the specified values the moments are infinite. Next we define random field associated to the multivariate T-distribution (cf. Roislien and Omre, 2011).

Definition 2. A random field $\{Z(s) : s \in D \subset R^p\}$ is termed a T-distributed random field (TRF) if $Z = [Z(s_1), \ldots, Z(s_n)]' \sim T_n(\mu, \Omega, m)$ for all $n \in N_+$ and all configurations $(s_1, \ldots, s_n) \in D^n$.

Suppose that the model of observation $Z(s)$ in the population Π_l is

$$Z(s) = x'(s)\beta_l + \varepsilon(s), \tag{1}$$

where $x(s)$ is a $q \times 1$ vector of non random regressors and β_l is a $q \times 1$ vector of parameters, $l = 1, 2$. The error term is generated by zero-mean stationary TRF $\{\varepsilon(s) : s \in D\}$ with covariance function defined by model for all $s, u \in D$

$$cov\{\varepsilon(s), \varepsilon(u)\} = r(s-u)\sigma^2 = m\omega(s-u)/(m-2), \tag{2}$$

where $r(s-u)$ is a spatial correlation function, σ^2 is a variance, $\omega(s-u)$ is a scaling function, and $m \in R_+$ is degrees of freedom. So Ω is the $n \times n$ matrix with (i, j), element specified by $\omega(s_i - s_j)$, $i, j = 1, \ldots, n$.

Denote by $S_n = \{s_i \in D;\ i = 1, \ldots, n\}$ the set of locations (STL) where training sample $Z' = [Z(s_1), \ldots, Z(s_n)]$ is taken. It specifies the spatial sampling design or spatial framework for training sample (Shekhar *et al.*, 2002).

Assume that each training sample realization $Z = z$ and S_n are arranged in the following way. The first n_1 components are the observations of $Z(s)$ from Π_1 and remaining $n_2 = n - n_1$ components are the observations of $Z(s)$ from Π_2. So S_n is partitioned into a union of two disjoint subsets, i. e. $S_n = S^{(1)} \cup S^{(2)}$, where $S^{(j)}$ is the subset of S_n that contains n_j locations of the feature observations from Π_j, $j = 1, 2$. We shall assume that the deterministic spatial sampling design and all analyses are carried out conditional on the given STL.

Joint training sample Z is specified by $Z' = (Z_1', Z_2')$, where training sample Z_l is the $n_l \times 1$vector of n_l observations of $Z(s)$ from Π_l, $l = 1, 2$. Then Z is the $n \times 1$ vector specified by the model

$$Z = X\beta + E, \tag{3}$$

where X is the $n \times 2q$ design matrix, $\beta' = (\beta_1', \beta_2')$ and E is the $n \times 1$ vector of the random errors that has the multivariate T-distribution $T_n(0, \Omega, m)$.

The design matrix X in Eq. (3) is specified by $X = X_1 \oplus X_2$ where the symbol $\oplus$ denotes the direct sum of matrices and X_l is the $n_l \times q$ matrix of regressors for $Z_l, l = 1, 2$.

Consider the problem of classification of the single observation of TRF at the location s_0 denoted by Z_0 into one of two populations specified above with the given training sample Z.

Denote by r_0 the vector of spatial correlations between Z_0 and Z and let $R = \Sigma/\sigma_2 = m\Omega/(\sigma_2(m-2))$ denote a matrix of spatial correlation of Z. The conditional distribution of Z_0 given $Z = z$ in population Π_l is $T_1(\mu_{lz}, \omega_{0z}, m+n)$ (see Roislien and Omre, 2011) with the mean function which is linear in the training sample observations:

$$\mu_{lz}^0 = E(Z_0|Z = z; \Pi_l) = x_0'\beta_l + \alpha'(z - X\beta), \quad l = 1, 2 \tag{4}$$

and the scaling parameter

$$\omega_{0z} = (m+n-2)Var(Z_0|Z = z; \Pi_l)/(m+n) = \sigma^2\rho_0\xi(z)(m-2)/m,$$

where $\alpha = R^{-1}r_0$, $\rho_0 = 1 - r_0'\alpha$ and

$$\xi(Z) = \big[1 + (z - X\beta)'\Omega^{-1}(z - X\beta)\big/m\big]\big/(1 + n/m).$$

3. Error Rates of Classification

Under the assumption that the populations are completely specified and for the known prior probabilities of the populations π_1 and π_2 ($\pi_1 + \pi_2 = 1$), the BR is based on the log ratio of the conditional densities described above.

Set $\mu_l^0 = x_0'\beta_l$, $l = 1, 2$ and assume $\pi_1 = 0,5$ under insignificant loss of generality. Then the BR is associated with the linear discriminant function (LDF)

$$L_z(Z_0) = \left(Z_0 - \frac{1}{2}\big(\mu_{1z}^0 + \mu_{2z}^0\big) - \alpha'(z - X\beta)\right)\big(\mu_{1z}^0 - \mu_{2z}^0\big). \tag{5}$$

Put $S_n(\bullet)$ and $t_n(\bullet)$ as cdf and pdf of $T_1(0, 1, n)$ and let $\Delta\mu^0 = \mu_1^0 - \mu_2^0 > 0$ and $\Delta_0 = \Delta\mu^0/(\sigma\sqrt{\rho_0})$.

Lemma 1. *The probability of misclassification based on LDF is*

$$P_B(z) = S_{m+n}\big(-(\Delta_0/2)\sqrt{\big(m\big/\big((m-2)\xi(z)\big)\big)}\big). \tag{6}$$

Proof. The probability of misclassification for $L_z(Z_0)$ is defined as $P_B(z) = \sum_{l=1}^{2}\pi_l P_{0l}$, where, for $l = 1, 2$, $P_{0l} = P_{0z}((-1)^l L_z(Z_0) > 0|\Pi_l)$ is the conditional probability that $L_z(Z_0)$ specified in Eq. (5) misclassifies Z_0, when it comes from Ω_l.

It is obvious that $(L_z(Z_0)|\Pi_l) \sim T_1(E_l, \omega_L, m + n)$, where

$$E_l = E\big(L_z(Z_0)\big|\Pi_l\big) = (-1)^{l+1}\big(\Delta\mu^0\big)^2/2,$$

$\omega_L = (m + n - 2)Var(L_z(Z_0)|\Pi_l)/(m + n) = \sigma_2(\Delta\mu^0)^2\xi(z)\rho_0(m - 2)/m$.

From the properties of the multivariate T-distribution it follows that

$$\big(L_z(Z_0) - E_l\big)\big/\sqrt{\omega_L} \sim T_1(0, 1, m + n)$$

in population Π_l, $l = 1, 2$. Therefore we easily complete the proof of lemma. □

Derived probability of misclassification is usually called the Bayes error rate.

In practical applications not all statistical parameters of populations are known. Then the estimators of unknown parameters can be found from the joint training sample. When the estimators of unknown parameters are plugged into the LDF specified in Eq. (6), the plug-in LDF (PLDF) is obtained. In this paper we assume that the true values of the parameters β are unknown. Let $\hat{\beta}$ be an estimator of β based on Z.

Then the PLDF is obtained from the LDF by replacing β with $\hat{\beta}$ in Eq. (5)

$$L_z(Z_0;\hat{\beta}) = \left(Z_0 - \alpha'(z - X\hat{\beta}) - \frac{1}{2}x_0'H\hat{\beta}\right)\left(x_0'G\hat{\beta}\right), \tag{7}$$

with $H = (I_q, I_q)$ and $G = (I_q, -I_q)$, where I_q denotes the identity matrix of order q. Performance of the plug-in discriminant functions is usually evaluated by an actual error rate (AER) (Le Roux *et al.*, 1997).

Put $b = \alpha' X - x_0'H/2$, $a_l = x_0'\beta_l - \alpha' X\beta$, $l = 1, 2$.

Lemma 2. *The actual error rate for PLDF specified in Eq.* (7) *is*

$$P_z(\hat{\beta}) = \sum_{l=1}^{2} S_{m+n}(\hat{Q}_l)/2, \tag{8}$$

where $\hat{Q}_l = (-1)^l \frac{(a_l + b\hat{\beta})sgn(x_0'G\hat{\beta})}{\sigma\sqrt{\rho_0\xi(z)}}\sqrt{(m+n)/(m+n-2)}$.

Proof. AER for PLDF $L_z(Z_0;\hat{\beta})$ is defined as $P_z(\hat{\beta}) = \sum_{l=1}^{2}\pi_l\hat{P_{0l}}$ where, for $l = 1, 2$, $\hat{P_{0l}} = P_{0z}((-1)^l L_z(Z_0;\hat{\beta}) > 0 \big| \Pi_l)$ is the conditional probability that $L_z(Z_0;\hat{\beta})$ misclassifies Z_0 when it comes from Ω_l (conditional probability is based on conditional distribution of Z_0 with mean μ_{lz}^0 Eq. (4) and variance σ_{0z}^2).

It is obvious that $(L_z(Z_0;\hat{\beta}\big|\Pi_l)) \sim T_1(\hat{E}_l, \hat{\omega}_L, m+n)$, where

$$\hat{E}_l = E\left(L_z(Z_0;\hat{\beta})\big|\Pi_l\right) = (a_l + b\hat{\beta})\left(\Delta\hat{\mu}^0\right),$$

$\hat{\omega}_L = (m+n-2)Var(L_z(Z_0;\hat{\beta})|\Pi_l)/(m+n) = \sigma^2(\Delta\hat{\mu}^0)^2\xi(z)\rho_0(m-2)/m$.

From the properties of the multivariate T-distribution it follows that

$$\left(L_z(Z_0;\hat{\beta}) - \hat{E}_l\right)\Big/\sqrt{\hat{\omega}_L} \sim T_1(0, 1, m+n)$$

in the population Π_l, $l = 1, 2$. Therefore we easily complete the proof of lemma. □

Definition 3. The expected error rate is obtained by averaging the actual error rate with respect to the distribution of the training sample and is defined as $EER = E_z(P_z(\hat{\beta}))$.

The actual error rate is useful in providing a guide to the performance of the plug-in classification rule when it is actually formed from training sample. It depends on observed values of training observations as well as their locations. The EER is the performance measure of PLDF before a training sample is observed and it depends on the set of training locations and the location of the observation to be classified s_0. It plays a similar role as the mean squared prediction error (MSPE) plays in the evaluating the performance of the plug-in kriging predictor (Diggle *et al.*, 2002). MSPE and its estimators are used for the spatial sampling design criterion for prediction (Zhu and Zhang, 2006; Zimmerman,

2006). These facts strengthen the motivation for the deriving of the estimators of the EER associated with the PLDF.

In this paper we propose the empirical estimator of the EER obtained by the rule based on the proposed PLDF. Maximum likelihood (ML) and ordinary least squares (LS) estimators of β are denoted by $\hat{\beta}^{ML}$ and $\hat{\beta}^{LS}$ respectively. Properties of ML estimators for parameters of stationary GRF are explored by numerous authors (see e.g. Mardia and Marshall, 1984; Sakalauskas, 2013). The following steps are performed to construct this empirical estimator of EER:

1. Simulate M training sample Z realizations according to the model specified in Eqs. (1)–(3).
2. For each simulated realization of $Z = z_{(l)}$, $l = \overline{1, M}$ compute the appropriate estimates of parameter β and denote them by $\hat{\beta}^{\kappa}_{(l)}$, where κ denotes the abbreviation of the estimator type, i.e. it takes the values ML or LS.
3. By using the derived formula for AER equation (8) compute the empirical estimator of the EER

$$\widehat{ER}_{\kappa} = \sum_{l=1}^{M} P_z\left(\hat{\beta}^{\kappa}_{(l)}\right)/M. \tag{9}$$

Denote by $\widehat{ER}_{ML}$ and $\widehat{ER}_{LS}$ the empirical estimators of the EER given in Eq. (9) with the implemented ML and LS parameter estimators, respectively.

Set $\overline{P}_B = \sum_{l=1}^{M} P_B(z_{(l)})/M$.

In the paper we consider the maximum likelihood and the ordinary least squares estimators and denote them by $\hat{\beta}^{ML}$ and $\hat{\beta}^{LS}$, respectively.

It is easy to show (see Sutradhar and Ali, 1986), that the ML estimator of β from the training sample Z is given by $\hat{\beta}^{ML} = \left(X' R^{-1} X\right)^{-1} X' R^{-1} Z$.

From the properties of the multivariate T-distribution it follows that

$$\hat{\beta}^{ML} \sim T_{2q}\left(\beta, \sigma^2 R_{ML}(m-2)/m, m\right)$$

and

$$\hat{\beta}^{LS} = \left(X'X\right)^{-1} X'Z \sim T_{2q}\left(\beta, \sigma^2 R_{LS}(m-2)/m, m\right),$$

where $R_{ML} = \left(X' R^{-1} X\right)^{-1}$ and $R_{LS} = (X'X)^{-1} X' R X (X'X)^{-1}$.

4. Example and Discussions

To investigate the influence of the parameter estimation methods to the proposed empirical estimators of the EER in the finite (even small) training sample case a numerical example is considered.

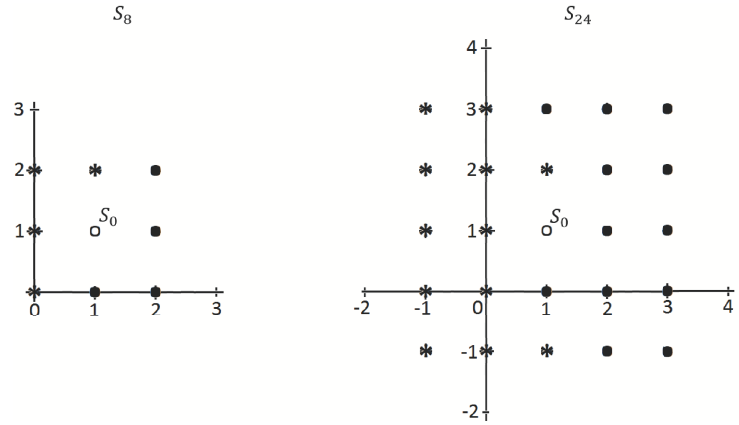

Fig. 1. S_8, S_{24} with $S^{(1)}$ and $S^{(2)}$ points marked as asterisks and dots.

In this example, the observations are assumed to arise from the stationary TRF with a constant mean in each population and a covariance function given by $C(h) = \sigma^2 r(h)$, where σ^2 is a known variance (sill) and $r(h)$ is the spatial correlation function specified by $r(h) = \exp\{-h/\varphi\}$ is considered. Here h is the distance between the locations, and denotes the range parameter (see Cressie, 1993).

Assume that D is a regular two-dimensional lattice with unit spacing. Consider the case $s_0 = (1, 1)$ and two fixed STL for S_8 and S_{24} respectively specified by

$$S_8 = \left\{S^{(1)} = \left\{(0,0), (0,1), (0,2), (1,2)\right\} S^{(2)} = \left\{(2,2), (2,1), (2,0), (1,0)\right\}\right\}$$

and

$$\begin{aligned} S_{24} = \big\{ & S^{(1)} = \big\{(0,0), (0,1), (0,2), (1,2), (1,-1), (0,-1), (-1,-1), (-1,0), \\ & (-1,1), (-1,2), (-1,3), (0,3)\big\}, \\ & S^{(2)} = \big\{(2,2), (2,1), (2,0), (1,0), (1,3), (2,3), (3,3), (3,2), \\ & (3,1), (3,0), (3,-1), (2,-1)\big\}\big\} \end{aligned}$$

STL distributions are shown in Fig. 1.

The values of empirical estimators of EER calculated by Eq. (9) with $M = 1000$ for S_8 and S_{24} for various Δ and φ are presented in Table 1.

Analyzing the figures in Table 1 we see that both estimators of the EER monotonically decreases when Δ and φ decreases. For all cases $\widehat{ER}_{LS} \geqslant \widehat{ER}_{ML}$ for S_8 and S_{24}. For all cases both $\widehat{ER}_{LS}$ and $\widehat{ER}_{ML}$ are larger than $\overline{P}_B$.

So from Table 1 we can conclude that the ML case have an advantage against the LS case by the sense of minimal value of the empirical estimator of the EER. The visual comparison of these two cases of parameter estimators is also done by plotting the values of index $\eta = \widehat{ER}_{ML}/\widehat{ER}_{LS}$. The dependence of values of this index on the parameter φ is shown for $\Delta = 0.5; 1.5$ (Fig. 2).

Table 1
Values of $\overline{P}_B$, $\widehat{ER}_{ML}$, $\widehat{ER}_{LS}$ for S_8 and S_{24}, for various Δ and φ.

φ/Δ	Error type	0.5		0.7		0.9		1.1		1.3		1.5	
		S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}	S_8	S_{24}
0.1	$\overline{P}_B$	0.3467	0.3467	0.2931	0.2930	0.2457	0.2456	0.2049	0.2048	0.1705	0.1704	0.1418	0.1417
	$\widehat{ER}_{ML}$	0.3539	0.3501	0.3025	0.2971	0.2565	0.2501	0.2164	0.2095	0.1820	0.1750	0.1530	0.1461
	$\widehat{ER}_{LS}$	0.3539	0.3501	0.3025	0.2971	0.2565	0.2501	0.2164	0.2095	0.1820	0.1750	0.1530	0.1461
0.5	$\overline{P}_B$	0.3421	0.3426	0.2872	0.2879	0.2391	0.2399	0.1980	0.1989	0.1636	0.1645	0.1351	0.1360
	$\widehat{ER}_{ML}$	0.3508	0.3470	0.2985	0.2932	0.2520	0.2457	0.2115	0.2048	0.1770	0.1702	0.1480	0.1414
	$\widehat{ER}_{LS}$	0.3510	0.3472	0.2988	0.2935	0.2523	0.2460	0.2118	0.2051	0.1773	0.1705	0.1483	0.1417
0.9	$\overline{P}_B$	0.3213	0.3210	0.2617	0.2615	0.2114	0.2111	0.1700	0.1698	0.1366	0.1366	0.1101	0.1101
	$\widehat{ER}_{ML}$	0.3327	0.3276	0.2762	0.2694	0.2273	0.2196	0.1862	0.1782	0.1522	0.1444	0.1246	0.1173
	$\widehat{ER}_{LS}$	0.3338	0.3298	0.2775	0.2719	0.2288	0.2223	0.1876	0.1809	0.1536	0.1470	0.1259	0.1197
1.3	$\overline{P}_B$	0.2989	0.2992	0.2353	0.2357	0.1837	0.1842	0.1432	0.1436	0.1119	0.1124	0.0880	0.0885
	$\widehat{ER}_{ML}$	0.3139	0.3093	0.2535	0.2473	0.2030	0.1960	0.1619	0.1549	0.1292	0.1226	0.1035	0.0975
	$\widehat{ER}_{LS}$	0.3167	0.3160	0.2567	0.2546	0.2063	0.2034	0.1652	0.1619	0.1324	0.1290	0.1064	0.1032
1.7	$\overline{P}_B$	0.2789	0.2792	0.2127	0.2130	0.1612	0.1614	0.1223	0.1224	0.0935	0.0934	0.0722	0.0720
	$\widehat{ER}_{ML}$	0.2983	0.2915	0.2348	0.2265	0.1833	0.1747	0.1429	0.1347	0.1118	0.1043	0.0880	0.0814
	$\widehat{ER}_{LS}$	0.3032	0.3000	0.2401	0.2362	0.1886	0.1845	0.1479	0.1439	0.1162	0.1125	0.0918	0.0885
2.1	$\overline{P}_B$	0.2592	0.2598	0.1910	0.1916	0.1400	0.1407	0.1032	0.1037	0.0768	0.0773	0.0580	0.0584
	$\widehat{ER}_{ML}$	0.2822	0.2719	0.2161	0.2047	0.1643	0.1531	0.1249	0.1147	0.0955	0.0866	0.0737	0.0662
	$\widehat{ER}_{LS}$	0.2878	0.2831	0.2223	0.2174	0.1704	0.1657	0.1305	0.1263	0.1005	0.0968	0.0780	0.0749
2.5	$\overline{P}_B$	0.2447	0.2456	0.1759	0.1770	0.1261	0.1273	0.0911	0.0923	0.0668	0.0679	0.0498	0.0508
	$\widehat{ER}_{ML}$	0.2670	0.2585	0.2001	0.1903	0.1490	0.1396	0.1113	0.1029	0.0837	0.0767	0.0637	0.0579
	$\widehat{ER}_{LS}$	0.2724	0.2740	0.2061	0.2069	0.1550	0.1553	0.1167	0.1168	0.0884	0.0884	0.0677	0.0676
2.9	$\overline{P}_B$	0.2311	0.2280	0.1621	0.1588	0.1137	0.1106	0.0808	0.0780	0.0584	0.0560	0.0431	0.0411
	$\widehat{ER}_{ML}$	0.2549	0.2382	0.1867	0.1697	0.1362	0.1207	0.0999	0.0866	0.0742	0.0630	0.0560	0.0467
	$\widehat{ER}_{LS}$	0.2619	0.2566	0.1944	0.1885	0.1436	0.1376	0.1065	0.1008	0.0798	0.0745	0.0606	0.0558
3.3	$\overline{P}_B$	0.2188	0.2169	0.1499	0.1482	0.1032	0.1017	0.0722	0.0709	0.0516	0.0506	0.0377	0.0369
	$\widehat{ER}_{ML}$	0.2460	0.2315	0.1771	0.1623	0.1273	0.1139	0.0923	0.0809	0.0679	0.0584	0.0508	0.0430
	$\widehat{ER}_{LS}$	0.2536	0.2478	0.1850	0.1800	0.1345	0.1304	0.0984	0.0951	0.0728	0.0703	0.0547	0.0527

It is easy to see (Fig. 2) that $\eta \leqslant 1$ and the values of this index decreases when φ increases. The same situation is valid for all levels of the population separation specified by the values of Δ.

5. Conclusions

In this paper, the closed form expression for the Bayes error rate equation (6) and the actual error rate equation (8) in classification of TRF observation based on the linear discriminant function and the plug-in linear discriminant function are derived. Based on these formulas, the comparison of the two PLDF based ML and LS estimators of the mean parameters is done by the simulated values of the empirical estimators of the EER. The simulation experiment shows that the advantage of the PLDF based on the ML estimator against the one based on the LS estimator. This advantage is greater for the cases with stronger spatial dependence between observations (i. e. larger values of φ). This conclusion is valid

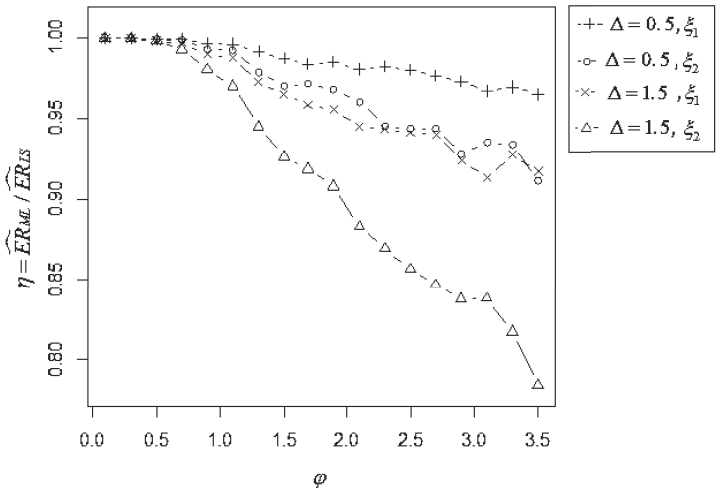

Fig. 2. Values of η for $\Delta = 0, 5; 1, 5$.

to different levels of separation between populations. Hence the results of this paper give us strong arguments to conclude that often untractable ML estimators of spatial mean parameters should be used in highly correlated spatial data modeled by TRF, and these estimators could be replaced by the simpler LS estimators for weakly correlated spatial data without significant loss of the PLDF performance.

References

Andrews, J.L., McNicholas, P.D., Subedi, S. (2011). Model-based classification via mixtures of multivariate t-distributions. *Computational Statistics and Data Analysis*, 55(1), 520–529.

Batsidis, A., Zografos, K. (2011). Errors of misclassification in discrimination of dimensional coherent elliptic random field observations. *Statistica Neerlandica*, 65, 446–461.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Diggle, P.J., Ribeiro, P.J., Christensen, O.F. (2002). An introduction to model-based geostatistics. *Lecture Notes in Statistics*, 173, 43–86.

Dučinskas, K. (2009). Approximation of the expected error rate in classification of the Gaussian random field observations. *Statistics and Probability Letters*, 79, 138–144.

Dučinskas, K., Stabingienė, L. (2011). Expected Bayes error rate in supervised classification of spatial Gaussian data. *Informatica*, 22(3), 371–381.

Dučinskas, K., Dreižienė, L., Zikarienė E. (2015). Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. *Statistics and Probability Letters*, 98, 107–114.

Kim, H.M., Mallick, B.K. (2003). A note on Bayesian spatial prediction using the elliptical distribution. *Statistics and Probability Letters*, 64, 271–276.

Lawoko, C.R.O., McLachlan, G.J. (1985). Discrimination with Autocorrelated Observations. *Pattern Recognition*, 18(2), 145–149.

Le Roux, N.J., Steel, S.J., Louw, N. (1997). Variable selection and error rate estimation in discriminant analysis. *Journal of Statistical Computation and Simulation*, 59, 195–219.

Mardia, K.V. (1984). Spatial discrimination and classification maps. *Communications in Statistics – Theory and Methods*, 13(18), 2181–2197.

Mardia, K.V., Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135–146.

Mardia, K.V., Kent, J.T., Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

McLachlan, G.J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York.

Nadarajah, S., Kotz, S. (2008). Estimation methods for the multivariate T-distribution. *Acta Applicandae Mathematicae*, 102, 99–118.

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *The Annals of Mathematical Statistics*, 34, 1268–1301.

Roislien, J., Omre, H. (2006). T-distributed random fields: a parametric model for heavy-tailed well-log data. *Mathematical Geology*, 38(7), 821–849.

Sakalauskas, L. (2013). Locally Homogeneous and Isotropic Gaussian Fields in Kriging. *Informatica*, 24(2), 253–274.

Shekhar, S., Schrater, P.R., Vatsavai, R.R., Wu, W., Chawla, S. (2002). Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. *IEEE Transactions of Multimedia*, 4, 174–188.

Shutoh, N. (2012). An asymptotic expansion for the distribution of the linear discriminant function based on monotone missing data. *Journal of Statistical Planning and Inference*, 82(2), 241–259.

Sutradhar, B.C., Ali, M.M. (1986). Estimation of the parameters of a regression model with a multivariate t error variable. *Communications in Statistics, Theory and Methods*, 15, 429–450.

Switzer, P. (1980). Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Mathematical Geology*, 12(4), 367–376.

Šaltytė, J., Dučinskas, K. (2002). Comparison of ML and OLS estimators in discriminant analysis of spatially correlated observations. *Informatica*, 13(2), 297–238.

Šaltytė-Benth, J., Dučinskas, K. (2005). Linear discriminant analysis of multivariate spatial-temporal regressions. *Scandinavian Journal of Statistics, Theory and Applications*, 32, 281–294.

Zhu, Z., Zhang, H. (2006). Spatial sampling design under infill asymptotic framework. *Environmetrics*, 17, 323–337.

Zimmerman, D.L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics*, 17, 635–652.

K. Dučinskas graduated from Vilnius University in 1976 in applied mathematics, where he received his doctor degree in 1983. He is a head of Mathematics and Statistics Department and a professor of Klaipeda University. Present research interests include discriminant analysis of spatially correlated data, statistical pattern recognition, geostatistics and Markov random fields.

E. Zikarienė graduated from Klaipeda University in 2009 in mathematics and is currently studying for a doctor degree in Vilnius University. She is an assistant in Klaipeda University. Present research interests include statistical pattern recognition, discriminant analysis of spatially correlated data, Bayes estimation and inference, spatial data mining.

Tikrosios T-atsitiktinio lauko stebinių klasifikavimo klaidų tikimybės panaudojant iterptąją tiesinę diskriminantinę funkciją

Kęstutis DUČINSKAS, Eglė ZIKARIENĖ

Šiame straipsnyje nagrinėjamas T-atsitiktinio lauko stebinių klasifikavimo į dvi populiacijas, besiskiriančias vidurkio funkcijomis, uždavinys. Klasifikavimo procedūra pagrįsta įterptąja tiesine diskriminantine funkcija, naudojančia vidurkių parametrų maksimalaus tikėtinumo ir mažiausių kvadratų įvertinius. Išvedamos orginalios Bajeso klaidos tikimybės ir tikrosios klasifikavimo klaidos tikimybės formulės. Pastarosios panaudojamos skaitiniame siūlomų klasifikavimo procedūrų efektyvumo vertinime ir palyginime.