EXPECTED ERROR RATE IN LINEAR DISCRIMINATION OF BALANCED SPATIAL GAUSSIAN TIME SERIES

M. Karaliutė[1], K. Dučinskas[1,2], L. Šaltytė-Vaisiauskė[2]
[1]*Institute of Data Science and Digital Technologies, Vilnius University*
[2] *Klaipėda University*
Vilnius, Klaipėda, LITHUANIA
e-mail: mkaraliute@gmail.com

Abstract

The problems of discriminant analysis of spatial-temporal correlated Gaussian data were intensively considered previously (see e.g. Saltyte-Benth and Ducinskas (2005)). However, theoretical results were derived under the assumption of statistical independence between observation to be classified and training sample. In the present paper, we avoid this tough restriction. The problem of supervised classifying of the spatial Gaussian time series (SGTS) observation into one of two populations, is specified by different regression mean models and by common covariance function, is considered. In the case of complete parametric certainty and with the fixed training sample locations, the formula of conditional Bayes error rate is derived. In the case of unknown regression parameters and temporal covariance matrix, their ML estimators are plugged into the Bayes discriminant function. The asymptotic approximation of expected error rate is derived. This result is multivariate generalization of previous ones.
Keywords: Gaussian random field, Bayes discriminant function, spatial correlation, conditional Bayes error rate, actual and expected error rate

1 Introduction

It is known that for completely specified populations an optimal classification rule in the sense of minimum misclassification probability is the Bayesian classification rule (BCR). In practice, however, some or all statistical parameters of populations are unknown. Training sample is used for the estimation of the parameters of both populations. Then the estimators of unknown parameters based on training sample are usually plugged in BCR. The expected error rate are usually considered as performance measure for the plug-in classification rule. To obtain closed-form expressions for the expected error rate are very cumbersome even for the simplest parametric structures of populations. This makes it difficult to build some qualitative conclusions. Therefore, asymptotic expansions of the expected error rate associated with plug-in BCR are especially important.

Many authors have investigated the performance of the plug-in version of the BCR when parameters are estimated from training samples with independent observations, or training samples where observations are temporally dependent (McLachlan(2004)). However, they did not analyze the error rate in classification of spatial-temporal data.

The main objective of this paper is to classify T observations of spatio-temporal GRF $\{Z(s,t) : s \in D \subset R^2, t \in [0,\infty)\}$ where s and t define spatial and temporal coordinates, respectively.

The model of observation $Z(s,t)$ in population Ω_l is

$$Z(s,t) = \mu_l(s,t) + \varepsilon(s,t),$$

where $\mu_l(s,t)$ - deterministic spatio-temporal trend, l - class number.

We modeled large-scale variation as the linear parametric trend

$\mu_l(s,t) = \beta_l' x(s)$

where $x(s) = (x_1(s), ..., x_q(s))'$ is vector of a spatial covariates and $\beta_l(t)$ is a q vector of parameters. The error term is generated by univariate zero - mean stationary GRF $\{\varepsilon(s,t) : s \in D \subset R^2, t \in [0,\infty)\}$, with covariance function defined by model for all $s, u \in D$

$$cov\{\varepsilon(s,t), \varepsilon(u,r)\} = C(s,u;t,r).$$

In this paper we restrict our attention to the separable case

$C(s,u;t,r) = R(s,u)\Sigma(t,r)$,

where $R(s,u)$ denotes spatial correlation between observations in locations s and u and $\Sigma(t,r)$ denotes temporal covariance between observations at moments t and r.

We consider isotropic spatial correlation belonging to **Mattern family (e.g. exponential model)**. Temporal dependence is described **by the AR(p) models**.

Let $S_n = \{s_i \in D; i = 1,...,n\}$ be a set of locations where training observation is taken. Call it the set of training locations (STL). So S_n is partitioned into the union of two disjoint subsets, i.e. $S_n = S^{(1)} \cup S^{(2)}$, where $S^{(l)}$ is the subset of S_n where observations of $Z(\cdot)$ from Ω_l are taken $l = 1,2$. Let $(S^{(l)}) = n_l, l = 1,2, n = n_1 + n_2$. The partition of STL denoted by $\xi = \{S^{(1)}, S^{(2)}\}$ will be called the spatial labels design (SLD) of training sample T.

Joint training sample M is stratified training sample, specified by $n \times T$ matrix $M = \binom{M_1}{M_2}$, where M_l is the $n_l \times T$ matrix of n_l observations of vectors $Z_i = (Z(s_i,1), ..., Z(s_i,T))'$ from $\Pi_l = \Omega_l^T$, where Ω_l^T denotes the T-fold direct product of from Ω_l, $l = 1,2$. Then for $l = 1,2$ $Z_i \sim N_T(B_l' x_i, \Sigma)$, where $x_i = x(s_i), i = 0,...,n$ and $B_l = (\beta_l(1), ..., \beta_l(T))$.

Let $B = \binom{B_1}{B_2}$ and $X_1 = (x_1, ..., x_{n_1})'$, $X_2 = (x_{n_1+1}, ..., x_n)'$ and $X = X_1 \oplus X_2$.

Consider the problem of classification of the vector of T observations of Z at location s_0 denoted by $Z_0 = (Z(s_0,1), ..., Z(s_0,T))'$ into one of two populations specified above with the given joint training sample M.

Then the model of M is

$$M = XB + E, \tag{1}$$

E is the $n \times T$ matrix of random errors that has matrix-variate normal distribution i.e.

$$E \sim N_{n \times p}(0, R \otimes \Sigma).$$

Here $R = (r_{ij}; i,j, = 1,...,n)$ denotes the spatial correlation matrix among observations in STL.

Denote by r_0 the vector of spatial correlations between Z_0 and observations in STL i.e $r_0 = (r_{01}, ..., r_{0n})$. Set

$$\alpha_0 = R^{-1} r_0, \ \rho = 1 - r_0' \alpha_0.$$

Notice that in population Ω_l, the conditional distribution of Z_0 given $M = m$ is Gaussian, i.e.

$$(Z_0 | M = m, \Omega_l) \sim N_T(\mu_{lm}^0, \Sigma_{0m}). \tag{2}$$

Then conditional squared Mahalanobis distance between populations for observation taken at location $s = s_0$ is

$$\Delta_0^2 = (\mu_{1m}^0 - \mu_{2m}^0)' \Sigma_{0m}^{-1} (\mu_{1m}^0 - \mu_{2m}^0) = \Delta^2 / \rho.$$

Let $H = (I_q, I_q)$ and $G = (I_q, -I_q)$, where I_q denotes the identity matrix of order q.

Under the assumption that the populations are completely specified and for known prior probabilities of populations $\pi_1(s)$ and $\pi_2(s)$ $(\pi_1(s)+\pi_2(s) = 1)$, the Bayes discriminant function (BDF) minimizing the probability of misclassification (PMC) is formed by log-ratio of conditional likelihood of distribution specified in (1)-(2), that is

$$W_m(Z_0) = \Big(Z_0 - \big(m - XB\big)'\alpha_0 - B'H'x_0/2\Big)' \Sigma^{-1} B'G'x_0/\rho + \gamma \tag{3}$$

where $\gamma = \ln(\pi_1(s_0)/\pi_2(s_0))$.

In this paper prior probabilities at location s_0 is assumed to be

$$\pi_1(s_0) = \sum_{i=1}^{n_1} \frac{1}{d(s_0, s_i)} \Big/ \sum_{i=1}^{n} \frac{1}{d(s_0, s_i)}, \pi_2(s_0) = 1 - \pi_1(s_0),$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance function between locations.

This discriminant function is optimal under the criterion of minimum of misclassification probability (see McLachlan, 2004).

The probability of misclassification for $W_T(Z_0)$ be called the Bayes error rate or optimal error rate. Denote it by P_n.

Lemma 1. *Bayes error rate for $W_m(Z_0)$ specified in (3) is*

$$P_n = \sum_{l=1}^{2} \pi_l \Phi(Q_l), \tag{4}$$

where $Q_l = -\Delta_0/2 + (-1)^l \gamma/\Delta_0$.

2 The error rates for plug-in BDF

When estimators of unknown parameters are plugged into BDF, the plug-in BDF is obtained. In this paper we assume that true values of parameters B and Σ are unknown. Let $\hat{B}$ and $\hat{\Sigma}$ be the estimators of B and Σ based on M.

The set of parameters that are to be estimated and the set of their estimators are denoted by $\Psi = \{B, \Sigma\}$ and $\hat{\Psi} = \{\hat{B}, \hat{\Sigma}\}$, respectively.

Then replacing Ψ by $\hat{\Psi}$ in (4) we get the plug-in BDF (PBDF)

$$W_M(Z_0; \hat{\Psi}) = \Big(Z_0 - \big(M - X\hat{B}\big)'\alpha_0 - \hat{B}'H'x_0/2\Big)'\hat{\Sigma}^{-1}\hat{B}'G'x_0/\rho + \gamma. \tag{5}$$

Definition 1. The actual error rate for BPDF $W_M(Z_0; \hat{\Psi})$ is defined as

$$P(\hat{\Psi}) = \sum_{l=1}^{2} \pi_l P((-1)^l W_M(Z_0; \hat{\Psi}) > 0|M). \tag{6}$$

Definition 2. The expectation of the actual error rate with respect to the distribution of M designated as $E_M\{P(\hat{\Psi})\}$, is called the expected error rate (EER).

So the EER for considered problem of Z_0 classification by BPDF is specified by $E_M\{P(\hat{\Psi})\}$.

Let $\Phi(x)$ be the standard normal distribution function.

Theorem 1. *Suppose that observation Z_0 to be classified by BPDF specified in (6), then the asymptotic approximation of EER based on second order Taylor expansion is*

$$AER = \sum_{l=1}^{2} \pi_l \Phi(-\Delta_0/2 + (-1)^l\gamma/\Delta_0) + \pi_1\varphi(-\Delta_0/2 - \gamma/\Delta_0) \times$$
$$\{\Lambda' R_B \Lambda \Delta_0/k + (T-1)x_0' G R_B G' x_0/(k\Delta_0) + (2\gamma^2/\Delta_0 + (T-1)\Delta_0)/(n-2q)\}/2. \tag{7}$$

where $\Lambda = X'\alpha_0 - (H'/2 + \gamma G'/\Delta_0^2)x_0$.

References

[1] McLachlan G.J. (2004). *Discriminant analysis and statistical pattern recognition.* Wiley, New York.

[2] Šaltytė-Benth J., Dučinskas K. (2005). Linear discriminant analysis of multivariate spatial-temporal regressions. *Scand. J. Statist.* . Vol. 32, pp. 281–294.