Liet. matem. rink, **45***, spec. nr.*, 2005, 343–346

Discriminant analysis of Gaussian spatial data with exponential covariance structure

Kęstutis DUČINSKAS (KU)*
e-mail: duce@gmf.ku.lt

Abstract. This paper considers the discrimination of the observation of the stationary Gaussian random field belonging to one of two populations with different means and covariance functions. Assumming that covariance functions has exponential structure, the unknown means and covariance parameters are estimated by ML method. Approximation of the expected error rate associated with Bayes plug-in discriminant function is derived.

Keywords: spatial correlation function, exponential structure, actual error rate, expected error rate.

1. Introduction

Realizations of a random field $\{Z(s)\colon\ s \in D \subset R^d\}$ can be considered as spatial data, where s defines spatial coordinates.

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

$$Z(s) = \beta_l' x(s) + \varepsilon_l(s),$$

where $x(s) = (x_1(s), \ldots, x_q(s))'$ is a $q \times 1$ vector of nonrandom regressors and β_l is the vector of unknown parameters, $l = 1, 2$. Assume that $\{\varepsilon_l(s)\colon\ s \in D \subset R^d\}$ is a scalar zero-mean stationary spatial Gaussian random field with factorized stationary spatial covariance function

$$cov\big\{\varepsilon_l(s), \varepsilon_l(u)\big\} = c_l(s-u)\sigma^2,$$

where $c_l(s-u)$ is the positive definite spatial correlation function, $l = 1, 2$. Then, in Ω_l the mean function at location s is

$$\mu_l(s) = \beta_l' x(s), \quad l = 1, 2.$$

Consider the problem of classification of the observation $Z^0 = Z(s_0)$, with $s_0 \in D$ into one of two populations specified above. Under the assumption that the populations are completely specified and for known prior probabilities of populations π_1 and $\pi_2(\pi_1 + \pi_2 = 1)$, the Bayes classification rule (BCR) $d_B(\cdot)$ minimizing the probability of misclassification (PMC) is

$$d_B(z^0) = \arg \max_{\{l=1,2\}} \pi_l p_l(z^0), \tag{1}$$

*Darbą finansavo Lietuvos valstybinis mokslo ir studijų fondas (programos registr. Nr. T-05258).

where z^0 is the realization of Z^0 and

$$p_l(z^0) = \exp\Big(-\frac{1}{2}\,(z^0-\mu_l^0)'(z^0-\mu_l^0)/\sigma^2\Big)/\sqrt{2\pi\sigma^2}$$

is a probability density function (p.d.f.) of Z^0 in Ω_l, $l=1,2$. Here $\mu_l^0 = \beta_l' x^0$ with $x^0 = x(s_0)$, $l = 1,2$.

Denote by P_B the PMC of BCR, usually called Bayes error rate.

In practical applications the parameters of the p.d.f. are usually not known. Then the estimators of unknown parameters can be found from 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.

Suppose that we observe the training sample $T' = (T_1', T_2')$, where T_l is the $N_l \times 1$ vector of N_l observations of univariate $Z(s)$ from Ω_l, $l=1,2$. Then T is the $N \times 1$ vector, where $N = N_1 + N_2$.

Assume that Z^0 is independent on T and T_1 is independent on T_2.

Let $\hat\beta_1$, $\hat\beta_2$ and $\hat\sigma^2$ be the estimators of β_1, β_2 and σ^2, respectively, based on T, and let $\hat{\mu_l}(s) = \hat{\beta}_l' x(s)$, $l=1,2$. Put $\Psi = (\mu_1^0, \mu_2^0, \sigma^2)$ and $\hat\Psi = (\hat\mu_1^0, \hat\mu_2^0, \hat\sigma^2)$.

The plug-in rule $d_B(z^0; \hat\Psi)$ is obtained by replacing the parameters in (1) with their estimators. Then the corresponding sample LDF is defined as

$$W(z^0;\hat\Psi) = \Big(z^0 - \frac{1}{2}(\hat\mu_1^0 + \hat\mu_2^0)\Big)'(\hat\mu_1^0 - \hat\mu_2^0)/\hat\sigma^2 + \gamma,$$

where $\gamma = \ln\frac{\pi_1}{\pi_2}$.

DEFINITION 1. The actual error rate for $d_B(z^0; \hat\Psi)$ is defined as

$$P(\hat\Psi) = \sum_{l=1}^{2} \pi_l \int \Big(1 - \delta\big(l, d_B(z^0;\hat\Psi)\big) p_l(z^0;\Psi)\Big) dz^0,$$

where $\delta(\cdot,\cdot)$ is Kronecker delta.

In the considered case the actual error rate for $d_B(z^0; \hat\Psi)$ can be rewritten as

$$P(\hat\Psi) = \sum_{l=1}^{2} \pi_l^0 \Phi\Big((-1)^l \frac{(\mu_l^0 - \frac{1}{2}\,(\hat\mu_1^0 + \hat\mu_2^0))'(\hat\mu_1^0 - \hat\mu_2^0) + \gamma/\hat\sigma^2}{\sigma\sqrt{(\hat\mu_1^0 - \hat\mu_2^0)'(\hat\mu_1^0 - \hat\mu_2^0)}}\Big). \tag{2}$$

DEFINITION 2. The expectation of the actual error rate with respect to the distribution of T denoted as $E_T\{P(\hat\Psi)\}$ is called the expected error rate (EER) for the rule $d_B(z^0, \hat\Psi)$.

Asymptotic approximations and asymptotic expansions for EER in the case of independent observations were considered by many authors (see, e.g., Dučinskas [1]). Mardia [2] considered a similar problem of classifying spatially distributed Gaussian observations with constant means. But he did not analyze the EER. In the present paper the asymptotic approximation for the EER of the classifying spatially distributed Gaussian observation with different means depending on regressors and parametric covariance function having exponential structure with common variance is derived. Maximum likelihood estimators (MLE) of all parameters were used in the plug-in version of the Bayes classification rule.

2. Asymptotic approximation for EER

Denote by C_l the $N_l \times N_l$ spatial correlation matrix of training sample T_l for $l = 1, 2$ and by X_l the $N_l \times q$ matrix of regressors or design matrix.

Set $C = C_1 \oplus C_2$, where $\oplus$ denotes direct sum of matrices.

Assume that the mathematical model of T is

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

where $X = X_1 \oplus X_2$, $\beta = \beta_1 \oplus \beta_2$ and $E \sim N_N(0, \sigma^2 C)$.

Suggest that $c_l(h)$ belongs to the parametric exponential family defined by $c_l(h) = \exp(-|h|/\theta_l)$, $l = 1, 2$.

The situation when true parameters $\{\theta_l\}$ are known was considered by J. Šaltytė and K. Dučinskas [4].

In this paper vector of parameters $\delta = (\sigma^2, \beta_1, \beta_2, \theta_1, \theta_2)$ are assumed unknown. Maximum likelihood method studied by Mardia and Marshal [3] was used for parameter estimations.

Let $\hat{\delta}$ be the ML estimator of δ by training sample T.

LEMMA. *Suppose that regularity assumptions of Mardia, Marshall* [3] *for model* (3) *and sampling design for T hold. Then as $N \to \infty$*

$$J(\hat{\delta} - \delta) \xrightarrow{D} N_{2q+3}(0, I),$$

where

$$J = J_\beta \oplus J_V.$$

Proof. Lemma directly follows from the Theorem 1 [3] applied for model (3).

Partial derivatives of matrix C_l by θ_l was defined by D_l.

Let

$$J_\beta = X'C^{-1}X/\sigma^2, \quad J_V = \frac{1}{2}(t_{ij}), \quad i, j = 1, 2, 3,$$

where J_V is symmetric 3×3 matrix with elements

$$t_{11} = n/\sigma^4, \quad t_{23} = 0,$$
$$t_{1i} = tr(C_{i-1}^{-1} D_{i-1})/(\sigma^2 \theta_{i-1}^2), \quad t_{ii} = tr(C_{i-1}^{-1} D_{i-1})^2/\theta_i^4, \quad i = 2, 3.$$

THEOREM. *Under assumptions of Lemma the asymptotic approximation of the EER is*

$$E\big(P(\hat{\Psi})\big) \simeq P_B + \pi_1\varphi\Big(-\frac{\Delta}{2} - \frac{\gamma}{\Delta}\Big)\Big(\sum_{l=1}^{2} a_l + \gamma^2 b\Big)/2\Delta,$$

where

$$a_l = \sigma^2 x'(s_0)(X_l' C_l^{-1} X_l)^{-1} x(s_0)\Big(-\frac{\Delta}{2} + (-2)^l \frac{\gamma}{\Delta}\Big)^2,$$

$$b = 2\sigma^4\Big(N - \sum_{l=1}^{2}\big(tr(C_l^{-1}D_l)\big)^2/tr(C_l^{-1}D_l)^2\Big).$$

Proof. Lemma imlies that covariance of $\hat{\delta}$ is asymptotically equivalent to J^{-1}, as $N \to \infty$ i.e.

$$cov(\hat{\delta} - \delta) \simeq J^{-1}. \tag{4}$$

Taylor series expansion of $P(\hat{\Psi})$ given by formula (2) up to the second order derivatives about true values of parameters are used. Taking the expectation of it by sampling distribution and using (4) the proof is completed.

The accuracy of the proposed approximation depends on the expectation of the remainder term of the Taylor Series expansion of $(P(\hat{\Psi}))$. Under given assumptions, it is of order $o((X'C^{-1}X)^{-1})$.

References

1. K. Dučinskas, An asymptotic analysis of the regret risk in discriminant analysis under various training schemes, *Lith. Math. J.*, **37**(4), 337–351 (1997).
2. K.V. Mardia, Spatial discrimination and classification maps, *Commun. Statist. theory Meth.*, **13**(18), 2181–2197 (1984).
3. K.V. Mardia, R.J. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, **61**(1), 131–135 (1984).
4. J. Šaltytė, K. Dučinskas, Comparison of ML and OLS estimators in discriminant analysis of spatially correlated observations, *Informatica*, **13**(2), 227–238 (2002).

REZIUMĖ

K. Dučinskas. Diskriminantinė erdvinių Gauso duomenų su eksponentine kovariacija analizė

Straipsnyje nagrinėjamas Gauso lauko realizacijų klasifikavimo uždavinys, kai klasės skiriasi regresiniais vidurkiais ir kovariacinėmis funkcijomis. Kovariacinės funkcijos turi parametrinę eksponentinę struktūrą. Pateikta laukiamos klaidos tikimybės aproksimacija atvejui, kai visi nežinomi klasių parametrai vertinami maksimalaus tikėtinumo metodu.