Теор. ймовірност. та матем. статист.
Вип. 61, 1999, стор. 15–20

Teor. Imovirnost. ta Matem. Statyst.
No. 61, 1999, pp. 15–20

ASYMPTOTIC APPROXIMATION FOR THE EXPECTED RISK IN CLASSIFICATION OF DIFFERENT SPATIAL REGRESSIONS

UDC 519.21

K. DUCINSKAS AND J. SALTYTE

ABSTRACT. The paper deals with the problem of classifying an observation from spatial regression models with intrinsically stationary Gaussian residuals into one of two populations described by different mean and spatial correlation functions. Asymptotic approximation for the expected risk of a plug-in classification rule is obtained. Comparison of obtained asymptotic approximation with Monte Carlo simulations in one spatial case was done.

1. INTRODUCTION

Let $\{Z(s): s \in D \subset \mathbf{R}^2\}$ be an intrinsically stationary Gaussian random field having different mean and spatial covariance functions under populations Ω_1 and Ω_2. Assume that the model of $Z(s)$ in population Ω_l is

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

where $x'(s) = (x_1(s), \ldots, x_q(s))$ is a $q \times 1$ vector of nonrandom regressors and

$$\beta_l = (\beta_l^1, \ldots, \beta_l^q)' \in B, \qquad l = 1, 2,$$

are parameter vectors, B being an open subset of $\mathbf{R}^q$. Assume, that $\{\varepsilon_l(s): s \in D \subset \mathbf{R}^2\}$ is a zero-mean intrinsically stationary random Gaussian field with spatial covariance defined by a parametric model $\operatorname{cov}\{\varepsilon_l(t), \varepsilon_l(s)\} = \sigma(t - s; \theta_l)$ for all $t, s \in D$, where $\theta_l \in \Theta$ is a $p \times 1$ parameter vector, Θ being an open subset of $\mathbf{R}^p$, $l = 1, 2$. Intrinsically stationarity of $\{\varepsilon_l(s): s \in D \subset \mathbf{R}^2\}$ means that $\mathsf{E}\{(\varepsilon_l(t) - \varepsilon_l(s))^2\}$ depends only on $(t - s)$, for $l = 1, 2$ and $t, s \in D$. We restricted our attention to the homoscedastic models, i.e. $\sigma(0; \theta) = \sigma^2$, for any $\theta \in \Theta$. Assume that θ_l, $l = 1, 2$, are known and σ^2 unknown.

Then under Ω_l the mean function at location s is $\mu_{ls} = x'(s)\beta_l$ and the spatial covariance function is

$$\operatorname{cov}\{\varepsilon_l(t), \varepsilon_l(s)\} = \sigma^2 \rho(t - s; \theta_l),$$

where $\rho(t - s; \theta_l)$ is the spatial correlation function, $l = 1, 2$. If $p_l(z(s); \mu_{ls}, \sigma^2)$ denotes the probability density function (p.d.f.) of $Z(s)$ under Ω_l, then

$$p_l\left(z(s); \mu_{ls}, \sigma^2\right) = \exp\left\{-\left(z(s) - x'(s)\beta_l\right)^2 / 2\sigma^2\right\} / \left(\sqrt{2\pi}\sigma\right). \tag{1}$$

Let $Z(r)$ be an observation at $r \in D$ from one of the two populations Ω_1 and Ω_2. Under the assumption that the populations are completely specified and for known finite nonnegative losses $L(i, j)$, $i, j = 1, 2$, the Bayes classification rule (BCR) $d_B(\cdot)$ minimizing risk of classification of $z(r)$ the observed value of $Z(r)$ (see e.g., Hand 1997) is

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

1991 *AMS Mathematics Subject Classification.* Primary 62H30, 62H12; Secondary 62M30.

with $c_{lr} = \pi_{lr}(L(l, 3-l) - L(l, l))$, $l = 1, 2$, where π_{1r} and π_{2r} $(\pi_{1r} + \pi_{2r} = 1)$ are prior probabilities of the populations Ω_1 and Ω_2, respectively.

Denote by R_B^r the risk of BCR.

In practical applications the parameters β_1, β_2, and σ^2 are not known and must be estimated from training samples $T_l = \{Z_{l1}, \ldots, Z_{lN_l}\}$, where $Z_{lk} = Z(s_k^l)$ denotes the kth observation from Ω_l, $l = 1, 2$.

Put $T = \{T_1, T_2\}$ and $N = N_1 + N_2$. Let $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\sigma}^2$ be the estimators of β_1, β_2, and σ^2, respectively, based on T and let $\widehat{\mu}_l(r) = x'(r)\widehat{\beta}_l$. The plug-in rule $d_B(z(r), \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widehat{\sigma}^2)$ is obtained by replacing the parameters in (2) with their estimators, i.e.,

$$d_B\left(z(r), \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widehat{\sigma}^2\right) = \arg \max_{\{l=1,2\}} c_l p_l\left(z(r); \widehat{\mu}_{lr}, \widehat{\sigma}^2\right). \tag{3}$$

Definition 1. The actual risk for $d_B(z(r), \widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2)$ is defined as

$$R^r\left(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2\right) \triangleq \sum_{l=1}^{2} \pi_l \int \left(1 - \delta\left(l, d_B\left(z(r), \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widehat{\sigma}^2\right)\right) p_l\left(z(r); \mu_{lr}, \sigma^2\right)\right) dz(r). \tag{4}$$

Definition 2. The expectation of the actual risk with respect to the distribution of T denoted as

$$E_T\left\{R^r\left(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2\right)\right\}$$

is called the expected risk (ER) for the $d_B(z(r), \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widehat{\sigma}^2)$.

Asymptotic approximations and asymptotic expansions for ER in case of independent observations were considered by many authors (see e.g., Okamoto 1963, Ducinskas 1997). Mardia (1984) considered similar problem of classifying spatially distributed Gaussian observations with constant means. But he did not analyze ER or probabilities of misclassification. In presented paper we obtain asymptotic approximation for the ER of classifying spatially distributed Gaussian observation with different means depending on the locations. Maximum likelihood estimators (MLE) of means and bias adjusted MLE of variance were used in plug-in version of Byes classification rule. We also make a comparison for the accuracy of our asymptotic approximation with Monte Carlo simulations when training sample sizes are small.

2. Asymptotic approximation for ER

We shell restrict our attention to the case when the effect of cross-correlations between observations from different populations is negligible. In this paper we suppose, that if $Z(s)$ is from Ω_l and $Z(t)$ is from Ω_{3-l}, $l = 1, 2$, then

$$\operatorname{cov}\left(Z(s), Z(t)\right) = 0. \tag{5}$$

The expectation vector and the covariance matrix of $T_l^V = (Z_{l1}, \ldots, Z_{lN_l})'$ are

$$\mu_{lr}^+ = \mathbb{1}_{N_l} \cdot \mu_{lr}$$

and $\Sigma_l^+ = \sigma^2 \cdot C_l$, where $\mathbb{1}_{N_l}$ is the N_l-dimensional vector of ones and C_l is the spatial correlation matrix of order $N_l \times N_l$, whose (i, j)th element is $\rho(s_i^l - s_j^l)$, $i, j = 1, \ldots, N_l$, $l = 1, 2$.

Suppose, that C_l are known $(l = 1, 2)$. In this paper maximum likelihood estimators (MLE) of β_l, $l = 1, 2$, and σ^2 based on T_l are used. Let X^l be an $N_l \times q$ regressor matrix with ith column $(x_{1i}^l, \ldots, x_{N_l i}^l)'$, where

$$x_{ki}^l = x_i(s_k^l), \qquad i = 1, \ldots, q, \ k = 1, \ldots, N_l, \ l = 1, 2.$$

Lemma. *For $l = 1, 2$ MLE of β_1, β_2, and σ^2 based on T are*

$$\widehat{\beta}_l = \left(\left(X^l\right)' C_l^{-1} X^l \right)^{-1} \left(X^l\right)' C_l^{-1} T_l^V, \qquad l = 1, 2, \tag{6}$$

$$\widehat{\sigma}^2 = \left(\sum_{l=1}^{2} \left(T_l^V - X^l \widehat{\beta}_l \right)' C_l^{-1} \left(T_l^V - X^l \widehat{\beta}_l \right) \right) / N. \tag{7}$$

Proof. The log-likelihood of T_l is

$$\ln L_l = \text{const} - \frac{1}{2} (N_l \ln \sigma + \ln |C|) - \left(T_l^V - X^l \beta^l \right)' C_l^{-1} \left(T_l^V - X^l \beta^l \right) / \left(2\sigma^2 \right), \quad l = 1, 2.$$

Solving the equations

$$\frac{\partial \ln L_l}{\partial \beta_l} = 0, \qquad l = 1, 2,$$

and

$$\sum_{l=1}^{2} \frac{\ln L_l}{\partial \sigma^2} = 0,$$

we complete the proof of Lemma. □

Since MLE $\widehat{\sigma}^2$ is biased, i.e., $\mathsf{E}\{\widehat{\sigma}^2\} = \sigma^2 (N - 2q)/N$, we will use bias adjusted estimator

$$\widetilde{\sigma}^2 = \widehat{\sigma}^2 \frac{N}{N - 2q}. \tag{8}$$

Mardia and Marshall (1984) have given sufficient conditions in the considered cases for the consistency and asymptotic normality of exact ML estimators of β and σ^2. They also discussed numerical methods for estimating the parameters of spatial covariance functions.

The sufficient conditions of Mardia and Marshall in considered case are satisfied by the following assumptions.

Assumption 1 (continuity assumption). Assume that $\rho(t - s; \theta)$ is twice differentiable with respect to θ at all points of Θ for all $t, s \in D$, and that it is positive-definite in the sense that for every finite subset $D_n = \{t_1, \ldots, t_n\}$ of D the correlation matrix

$$C = \{\rho(t_i - t_j; \theta)\}$$

is positive definite.

Assumption 2 (growth in information). Assume that $\lim (X_n' C^{-1} X_n)^{-1} = 0$ as $n \to \infty$, where X_n is an $n \times q$ regressor matrix with jth column

$$x_j = \left(x_j(t_1), \ldots, x_j(t_n) \right)'.$$

Put $\gamma_r = \ln(c_{1r}/c_{2r})$, $\Delta\widehat{\mu}_{lr} = \widehat{\mu}_{lr} - \mu_{lr}$, and $\Delta\widetilde{\sigma}^2 = \widetilde{\sigma}^2 - \sigma^2$. Let $\Phi(\cdot)$ and $\varphi(\cdot)$ denote standard normal distribution and density functions, respectively.

The plug-in discriminant function can be written in the form

$$d_B\left(z(r); \widehat{\mu}_1, \widehat{\mu}_2, \widetilde{\sigma}^2\right) = \left(z(r) - \frac{1}{2} \left(\widehat{\mu}_{1r} + \widehat{\mu}_{2r} \right) \right) \frac{\left(\widehat{\mu}_{1r} - \widehat{\mu}_{2r} \right)}{\widetilde{\sigma}^2} + \gamma_r.$$

Then the actual risk for $d_B(x_r, \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ (see McLaclan (1974)) is

$$R^r\left(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2\right) = \pi_{1r}\Phi\left(-\frac{\left(\mu_{1r} - \frac{1}{2}\left(\widehat{\mu}_{1r} + \widehat{\mu}_{2r}\right)\right)\left(\widehat{\mu}_{1r} - \widehat{\mu}_{2r}\right)/\widetilde{\sigma}^2 + \gamma_r}{\sqrt{\left(\widehat{\mu}_{1r} - \widehat{\mu}_{2r}\right)^2 \sigma^2/\left(\widetilde{\sigma}^2\right)^2}}\right)$$
$$+ \pi_{2r}\Phi\left(\frac{\left(\mu_{2r} - \frac{1}{2}\left(\widehat{\mu}_{1r} + \widehat{\mu}_{2r}\right)\right)\left(\widehat{\mu}_{1r} - \widehat{\mu}_{2r}\right)/\widetilde{\sigma}^2 + \gamma_r}{\sqrt{\left(\widehat{\mu}_{1r} - \widehat{\mu}_{2r}\right)^2 \sigma^2/\left(\widetilde{\sigma}^2\right)^2}}\right), \tag{9}$$

where superscript in R^r indicates the location $r \in D$ of the observation being classified.

Let

$$a_l^r = \sigma^2 x'(r)\left(\left(X^l\right)' C_l^{-1} X^l\right)^{-1} x(r)$$

and

$$\Delta^2(r) = \frac{(\mu_{1r} - \mu_{2r})^2}{\sigma^2} = \frac{\left(x'(r)\left(\beta^1 - \beta^2\right)\right)^2}{\sigma^2}$$

for any $r \in D$ and $l = 1, 2$.

Let $R_l^{(1)}$ be the first-order derivatives of $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ by $\widehat{\mu}_{lr}$ evaluated at μ_{lr} and $R_{l,k}^{(2)}$ denotes the second-order derivatives of $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ by $\widehat{\mu}_{lr}$ and $\widehat{\mu}_{kr}$ evaluated at μ_{lr} and μ_{kr}, respectively, $(l, k = 1, 2)$. Similarly, $R_{\sigma^2}^{(k)}$, $k = 1, 2$, means the k-th order derivative of $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ with respect to $\widetilde{\sigma}^2$ evaluated at $\widetilde{\sigma}^2 = \sigma^2$.

Theorem. *Suppose Assumptions 1 and 2 hold for training samples T_1 and T_2. Then the asymptotic approximation of expected risk for the $d_B(z(r), \widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ is*

$$E_T\left\{R^r\left(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2\right)\right\} \simeq \sum_{l=1}^{2}\left(c_{lr}\Phi\left(-\frac{\Delta(r)}{2} + (-1)^l\frac{\gamma_r}{\Delta(r)}\right) + \pi_{lr}L(l,l)\right)$$
$$+ c_{1r}\varphi\left(-\frac{\Delta(r)}{2} - \frac{\gamma_r}{\Delta(r)}\right)$$
$$\times \sum_{l=1}^{2} a_l^r\left(-\frac{\Delta(r)}{2} + (-1)^l\frac{\gamma_r}{\Delta(r)}\right)^2/2\Delta(r)$$
$$+ \frac{\gamma_r^2}{\Delta(r)}c_{1r}\varphi\left(-\frac{\Delta(r)}{2} - \frac{\gamma_r}{\Delta(r)}\right)/(N - 2q). \tag{10}$$

Proof. Since $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ is invariant under linear transformations of data we use the convenient canonical form of $\sigma^2 = 1$ and $\mu_{1r} = \Delta(r)$, $\mu_{2r} = 0$ (see Dunn (1971)). Expand $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ in Taylor series about the point $\widehat{\mu}_{1r} = \Delta(r)$, $\widehat{\mu}_{2r} = 0$, and $\widetilde{\sigma}^2 = 1$. Taking the expectation with respect to the distribution of T and dropping the third order terms we have

$$E_T\left(R^r\left(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2\right)\right) \simeq R_B^r + \sum_{l=1}^{2} R_l^{(1)} E_T\{\Delta\widehat{\mu}_{lr}\} + R_{\sigma^2}^{(1)} E_T\left\{\Delta\widetilde{\sigma}^2\right\}$$
$$+ \frac{1}{2}\sum_{l,k=1}^{2}\left(R_{kl}^{(2)} E_T\{\Delta\widehat{\mu}_{lr}\Delta\widehat{\mu}_{kr}\} + R_{\sigma^2}^{(2)} E_T\left\{\left(\Delta\widetilde{\sigma}^2\right)^2\right\}\right). \tag{11}$$

Since $R^r(\widehat{\mu}_{1r}, \widehat{\mu}_{2r}, \widetilde{\sigma}^2)$ is minimized at $\widehat{\mu}_{lr} = \mu_{lr}$ and $\widetilde{\sigma}^2 = \sigma^2$, then

$$R_l^{(1)} = 0, \qquad l = 1, 2, \tag{12}$$

and $R_{\sigma^2}^{(1)} = 0$.

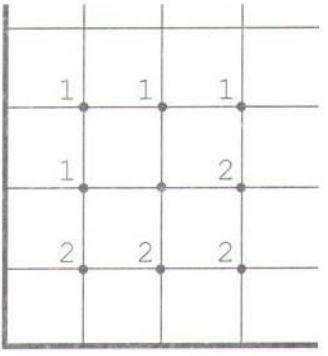

FIGURE 1. Second-order neighborhood scheme

Using Lemma, (5) and (8) we get, that

$$E_T\left\{(\Delta\widehat{\mu}_{lr})^2\right\} = a_l^r, \qquad l = 1, 2, \tag{13}$$

$$E_T\{\Delta\widehat{\mu}_{1r}\Delta\widehat{\mu}_{2r}\} = 0, \tag{14}$$

$$E_T\left\{\left(\Delta\widetilde{\sigma}^2\right)^2\right\} = \frac{2}{(N-2q)}. \tag{16}$$

Note that

$$R_l^{(2)} = c_{1r}\varphi\left(-\frac{\Delta(r)}{2} - \frac{\gamma_r}{\Delta(r)}\right)\left(-\frac{\Delta(r)}{2} + (-1)^l\frac{\gamma_r}{\Delta(r)}\right)^2/\Delta(r), \tag{17}$$

$$R_{\sigma^2}^{(2)} = c_{1r}\gamma_r^2\varphi\left(-\frac{\Delta(r)}{2} - \frac{\gamma_r}{\Delta(r)}\right)/\Delta(r). \tag{18}$$

Then putting (12)–(18) into (11) we complete the proof of stated theorem. □

3. EXAMPLE

Here we make the comparison of our approximation of ER (denoted by R_A) with Monte Carlo simulations (denoted by R_{MC}) for one special case.

As an example we consider the integer regular 2-dimensional lattice and use the second-order neighborhood scheme for training sample. Suppose, that there are 4 spatially symmetric observations in training sample for each class (see Figure 1).

Suppose, that regressor for both populations is of the form

$$x'(s) = \frac{1}{(|s|^2 + 2.5)}.$$

Considered spatial correlation function for both populations is $\rho(s) = \exp(-2\,|s|^2)$.

Let $\pi_1 = 0.2$ and loss function is zero-one function.

In Table 1 the values of asymptotic approximation of ER and Monte Carlo simulation values obtained by taking 100 replications at each location are presented. Column with ratio R_A/R_{MC} allow us to estimate the accuracy of proposed approximation. We can conclude that this approximation is sometimes appropriate even for small training sample sizes.

BIBLIOGRAPHY

1. K. Dučinskas, *An asymptotic analysis of the regret risk in discriminant analysis under various training schemes*, Lith. Math. J. **37** (1997), № 4, 337–351.

TABLE 1. Comparison of approximation with simulation

Δ	R_A	R_{MC}	R_A/R_{MC}
0.8	0.831118	0.272062	3.054879
1.0	0.523008	0.261229	2.002106
1.2	0.377941	0.255751	1.477773
1.4	0.296024	0.233617	1.267134
1.6	0.242301	0.212316	1.141228
1.8	0.202832	0.186737	1.086192
2.0	0.171508	0.173034	0.991182
2.2	0.145441	0.166917	0.871334
2.4	0.123171	0.146899	0.838466
2.6	0.103905	0.126339	0.822421
2.8	0.087171	0.121036	0.720208
3.0	0.072655	0.111749	0.650162

2. O. J. Dunn, *Some expected values for probabilities of correct classification in discriminant analysis*, Technometrics **13** (1971), 345–353.
3. D. J. Hand, *Construction and assessment of classification rules*, John Wiley & Sons, New York, 1997.
4. K. V. Mardia, *Spatial discrimination and classification maps*, Commun. Statist. - Theory Meth. **13** (1984), № 18, 2181–2197.
5. K. V. Mardia and R. J. Marshall, *Maximum likelihood estimation of models for residual covariance in spatial regression*, Biometrika **71** (1984), 135–146.
6. G. J. McLaclan, *The asymptotic distributions of the conditional error rate and risk in discriminant analysis*, Biometrika **61** (1974), № 1, 131–135.
7. M. Okamoto, *An asymptotic expansion for the distribution of the linear discriminant function*, Ann. Math. Statist. **34** (1963), 1286–1301.

KLAIPEDA UNIVERSITY, DEPARTMENT OF SYSTEM RESEARCH, H. MANTO 84, KLAIPEDA LT-5808
E-mail address: duce@gmf.ku.lt

KLAIPEDA UNIVERSITY, DEPARTMENT OF SYSTEM RESEARCH, H. MANTO 84, KLAIPEDA LT-5808
E-mail address: jsaltyte@gmf.ku.lt

Received 10/01/2000