Lietuvos matematikos rinkinys
Proc. of the Lithuanian Mathematical Society, Ser. A
Vol. **53**, 2012, 24–29

ISSN 0132-2818
www.mii.lt/LMR/

The influence of training sampling size on the expected error rate in spatial classification

Lina Dreižienė, Marta Karaliutė

Department of Statistics, Klaipėda University
H. Manto 84, LT-92294 Klaipėda
E-mail: l.dreiziene@gmail.com, destmkaraliute@gmail.com

Abstract. In this paper we use the pluged-in Bayes discriminant function (PBDF) for classification of spatial Gaussian data into one of two populations specified by different parametric mean models and common geometric anisotropic covariance function. The pluged-in Bayes discriminant function is constructed by using ML estimators of unknown mean and anisotropy ratio parameters. We focus on the asymptotic approximation of expected error rate (AER) and our aim is to investigate the effects of two different spatial sampling designs (based on increasing and fixed domain asymptotics) on AER.

Keywords: Bayes discriminant function, actual risk, expected error rate, Gaussian random field, increasing domain asymptotics, infill asymptotics.

Introduction

Bayesian discriminant function (BDF) is known as an optimal classification rule in the sense of minimum risk, then populations are completely specified and the loss function is known. If populations are not completely specified unknown parameters could be estimated from training sample and pluged-in BDF. The expected error rate (ER) is the performance measure of PBDF, but the expressions for the ER are very complicated even for the simplest forms of PBDF, therefore, asymptotic approximations of the ER are used.

The first investigation of PBDF quality for spatial classification was done by Switzer (1980) [7]. Later some extensions were done in [1, 8, 9]. However, correlations between observations to be classified and training sample were assumed to equal zero in all these publication. The first extension rejecting assumption about spatial independence was done by Dučinskas (2009) [4]. Here only the trend parameters and variance is assumed to be unknown. The extension of the latter approximation to the case of complete parametric uncertainty (all means and covariance function parameters are unknown) was implemented in Dučinskas and Dreižienė (2011) [5]. In all recently mentioned publications the pluged-in Bayes discriminant function (PBDF) was constructed using maximum likelihood estimators of unknown mean and covariance parameters.

The asymptotic behavior of spatial covariance parameter estimators can be different under the different asymptotic spatial frameworks. We consider two types of sampling framework in spatial statistics. One is the fixed-domain asymptotic framework or infill asymptotic framework. Here more and more observations might be sampled in the same finite domain (Cressie, 1993) [2]. The other type is called increasing domain asymptotic framework in which the minimum distance between sampling points is bounded away from zero and thus the spatial domain of observation is unbounded (Zhang, Zimmerman, 2005) [10]. This is the spatial analogue of the asymptotics observed in time series.

In this paper we seek to investigate the influence of training sample increase on AER using two different spatial sampling frameworks. We use the AER expression in the case of the unknown mean parameters and unknown anisotropy ratio. This AER expression in closed form was derived in [3].

1 Spatial classification problem

This paper deal with Gaussian random field (GRF) observations $\{Z(s)\colon s \in D \subset R^p\}$ and the main goal is to classify Z(s) into one of two populations Ω_j, $j = 1, 2$. Model of $Z(s)$ in the population Ω_j is

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

where $x(s)$ is a $q \times 1$ vector of non random regressors and β_j is a $q \times 1$ vector of parameters, $j = 1, 2$. $\varepsilon(s)$ is the error term with zero-mean and covariance function defined by model for all $s, u \in D$

$$\operatorname{cov}\big\{\varepsilon(s), \varepsilon(u)\big\} = C(s - u; \theta), \tag{2}$$

where $\theta \in \Theta$ is a $p \times 1$ parameter vector, Θ being an open subset of R^p.

Denote by $T = (Z(s_1), \dots, Z(s_n))'$ training sample and $S_n = \{s_i \in D;\ i = 1, \dots, n\}$ the set of locations where training sample T is taken and call it the set of training locations (STL).

We shall assume the deterministic spatial sampling design and all analyses are carried out conditionally on S_n.

S_n is partitioned into the 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 feature observations from Ω_j, $j = 1, 2$.

For given training sample T, consider the problem of classification of the $Z_0 = Z(s_0)$ into one of two populations when $x'(s_0)\beta_1 \neq x'(s_0)\beta_2, s_0 \in D$.

The model of training sample has the following form $T = X\beta + E$, where $\beta = (\beta_1', \beta_2')'$ is $2q \times 1$ vector of regression parameters, X is $n \times 2q$ design matrix of training sample T. E is the $n \times 1$ vector of random errors that has multivariate Gaussian distribution $N_n(0, C(\theta))$.

Denote by c_0 the covariance between Z_0 and T. Let t denote the realization of T.

Since Z_0 follows model specified in (1), the conditional distribution of Z_0 given $T = t$, Ω_j is Gaussian with mean

$$\mu_{lt}^0 = E(Z_0 | T = t; \Omega_j) = x_0'\beta_j + \alpha_0'(t - X\beta), \quad j = 1, 2 \tag{3}$$

and variance

$$\sigma_0^2(\theta) = \operatorname{var}(Z_0 | T = t; \Omega_j) = C(0) - c_0' C^{-1} c_0, \tag{4}$$

where $x_0' = x'(s_0)$, $\alpha_0' = c_0' C^{-1}$.

Under the assumption of complete parametric certainty of populations and for known finite nonnegative losses $\{L(i,j),\ i,j=1,2\}$, the BDF has the following form

$$W_t(Z_0,\Psi) = \Big(Z_0 - \frac{1}{2}\big(\mu^0_{1t} + \mu^0_{2t}\big)\Big)\big(\mu^0_{1t} - \mu^0_{2t}\big)/\sigma_0^2 + \gamma, \tag{5}$$

$\Psi = (\beta', \theta')'$, $\gamma = \ln(\pi_1^*/\pi_2^*)$, $\pi_j^* = \pi_j(L(j,3-j) - L(j,j))$, $j = 1,2$, where π_1, π_2 ($\pi_1 + \pi_2 = 1$) are prior probabilities of the populations Ω_1 and Ω_2, respectively.

Denote by $\hat\beta$, $\hat\theta$ the estimators of corresponding parameters. Replacing parameters with their estimates in BDF (5) we form the PBDF

$$W_t(Z_0;\hat\Psi) = \Big(Z_0 - \hat\alpha_0'(t - X\hat\beta) - \frac{1}{2}x_0'H\hat\beta\Big)\big(x_0'G\hat\beta\big)/\hat\sigma_0^2 + \gamma, \tag{6}$$

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

We will use the maximum likelihood (ML) estimators of parameters based on the training sample. The asymptotic properties of ML estimators established by Mardia and Marshall (1984) [6] under increasing domain asymptotic framework and subject to some regularity conditions are essentially exploited. Hence, the ML estimator $\hat\Psi$ is weakly consistent and asymptotically Gaussian [5].

Denote $\Lambda' = \alpha_0'X - x_0'(H/2 + \gamma G/\Delta_0^2)$, $K_\beta = \Lambda' J_\beta^{-1}\Lambda$, $J_\beta = X'C^{-1}X$. Here Δ_0^2 is the squared Mahalanobis distance between conditional distributions of Z_0 given $T = t$.

Under assumptions (A1) and (A2) in Dučinskas, Dreižienė (2011) [5] approximation of ER in the case of estimated unknown mean parameters and estimated unknown covariance parameters is

$$AER = R(\Psi) + \pi_1^*\varphi(-\Delta_0/2 - \gamma/\Delta_0)\Delta_0(K_\beta + K_\theta)/2\sigma_0^2, \tag{7}$$

where

$$K_\theta = \operatorname{tr}\big(CBJ_\theta^{-1}B'\big) + \gamma^2\big(\big(\hat\sigma_0^2\big)^{(1)}_\theta\big)'J_\theta^{-1}\big(\hat\sigma_0^2\big)^{(1)}_\theta/\Delta_0^2\sigma_0^2, \tag{8}$$

$R(\Psi)$ is the risk of BDF, (i,j)-th element of J_θ is $\operatorname{tr}(C^{-1}C_iC^{-1}C_j)/2$. $B = \partial\hat\alpha_0/\partial\hat\theta$ is the $n \times k$ matrix of partial derivatives evaluated at point $\hat\theta = \theta$. $\varphi(\cdot)$ denotes the standard normal distribution density function and $(\hat\sigma_0^2)^{(1)}_\theta$ is first order partial derivatives of $\hat\sigma_0^2$ (4) evaluated at point $\hat\theta = \theta$. More details can be found in [5].

2 Numerical example

Numerical example is considered to investigate the influence of training sample increase on AER using different spatial frameworks. Assume that D is a regular 2-dimensional lattice with unit spacing. Consider the case $s_0 = (4,4)$ and eight fixed STL $S_{m,n}$, $m = 1,2$, where 1 denotes infill asymptotic sampling framework, 2 denotes increasing domain asymptotic sampling framework and n represents the size of training sample, $n = 8$, 16, 32, 98. For example $S_{m,8}$ contains 8 neighbors of s_0, $S_{m,16}$ contains 16 neighbors of s_0 and so on. $S_{m,n}$ is partitioned into a union of two disjoint subsets, i.e. $S_{m,n} = S^{(1)} \cup S^{(2)}$, where $S^{(j)}$ is the subset of $S_{m,n}$ that contains n_j locations of feature observations from Ω_j, $j = 1,2$ and $n_1 = n_2$.

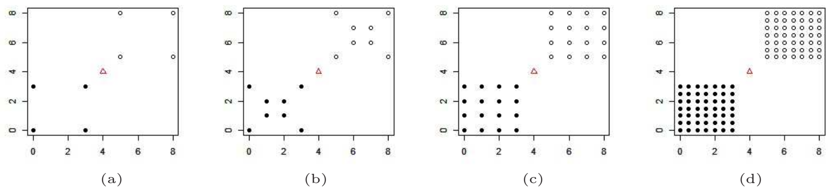

Fig. 1. Infill asymptotic sampling framework with different training sample sizes (a) $n = 8$, (b) $n = 16$, (c) $n = 32$, (d) $n = 98$; the symbols $\bullet$, $\circ$, and the triangle $\triangle$ represent $S^{(1)}$, $S^{(2)}$, and s_0, respectively.

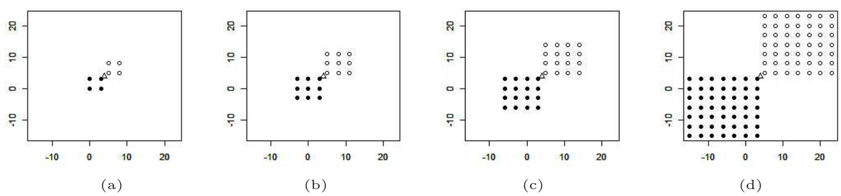

Fig. 2. Increasing domain asymptotic sampling frameworks with different training sample sizes (a) $n = 8$, (b) $n = 16$, (c) $n = 32$, (d) $n = 98$; the symbols $\bullet$, $\circ$, and the triangle $\triangle$ represent $S^{(1)}$, $S^{(2)}$, and s_0, respectively.

Figure 1 represents STL for infill asymptotic sampling framework ($S_{1,n}$), here extra locations are taken from between observed locations. The STL for increasing domain asymptotic sampling framework ($S_{2,n}$) are shown in Fig. 2. In this case number extra locations are taken by increasing the domain of observations.

With an insignificant loss of generality the case with $\pi_j = 0.5$ and $L(i, j) = 1 - \delta_{ij}$, $i, j = 1, 2$ is considered. Observations are assumed to arise from stationary GRF with different constant mean and common nugget-less covariance function given by $C(h) = \sigma^2 r(h)$, where σ^2 is variance (sill) and $r(h) = \exp\{-\sqrt{h_x^2 + \lambda^2 h_y^2}/\alpha\}$ is the exponential geometric anisotropic correlation function with anisotropy ratio λ and anisotropy angle $\varphi = \pi/2$. Here α denotes the range parameter.

We consider the case with unknown mean and anisotropy ratio parameters.

Figure 3 shows the values of AER using infill asymptotic sampling frameworks and increasing domain asymptotic sampling frameworks. AER are calculated assuming Mahalanobis distance between marginal distributions $\Delta = (\mu_1 - \mu_2)/\sigma = 1$, α=0.6, $\sigma^2 = 1$. The results show that less values of AER are obtained using increasing domain sampling framework. AER values are decreasing while training sample size increases for both sampling frameworks and for both isotropic and anisotropic cases ($\lambda = 1$ and $\lambda = 2$).

Table 1 shows the ratio of AER calculated using an increasing domain asymptotic sampling framework (AER_{inc}) to infill asymptotic sampling framework (AER_{inf}). It is obvious that $\mathrm{AER}_{Inc}/\mathrm{AER}_{Inf}$ increases while α is increasing for all training sample sizes. This leads to the conclusion that for greater α values and greater training sample size the infill asymptotic sampling framework gives lower values of AER in comparison with the increasing domain asymptotic sampling framework.

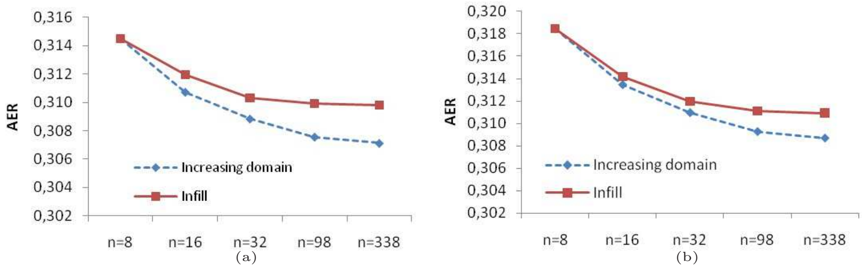

Fig. 3. AER values using different asymptotic frameworks: (a) isotropic case ($\lambda = 1$), (b) anisotropic case ($\lambda = 2$).

Table 1. $\mathrm{AER}_{inc}/\mathrm{AER}_{inf}$ with different α values and fixed $\Delta = 1$ and $\lambda = 1$.

α	$N = 16$	$N = 32$	$N = 98$	$N = 338$
0.8	0.9954	0.9947	0.9925	0.9920
1.2	0.9960	0.9962	0.9955	0.9960
1.6	0.9972	0.9983	0.9988	1.0001
2.0	0.9981	1.0000	1.0015	1.0034
2.4	0.9987	1.0014	1.0037	1.0061
2.8	0.9992	1.0025	1.0055	1.0084
3.2	0.9996	1.0035	1.0070	1.0104

References

[1] A. Batsidis and K. Zografos. Discrimination of observations into one of two elliptic populations based on monotone training samples. *Metrika*, **64**:221–241, 2006.

[2] N. Cressie. *Statistics for Spatial Data.* Wiley & Sons, New York, 1993.

[3] L. Dreižienė. Linear discriminant analysis of spatial Gaussian data with estimated anisotropy ratio. *Liet. mat. rink.: LMD darbai*, **52**:315–320, 2011.

[4] K. Dučinskas. Approximation of the expected error rate in classification of the Gaussian random field observations. *Stat. Prob. Lett.*, **79**:138–144, 2009.

[5] K. Dučinskas and L. Dreižienė. Supervised classification of the scalar gaussian random field observations under a deterministic spatial sampling design. *Austrian Journal of Statistics*, **40**(1,2):25–36, 2011.

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

[7] P. Switzer. Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Math. Geol.*, **12**:367–376, 1980.

[8] J. Šaltytė and K. Dučinskas. Comparison of ML and OLS estimators in discriminant analysis of spatially correlated observations. *Informatica*, **13**(2):227–238, 2002.

[9] J. Šaltytė and K. Dučinskas. Linear discriminant analysis of multivariate spatial-temporal regressions. *Scand. J. Statist.*, **32**:281–294, 2005.

[10] H. Zhang and D.L. Zimmerman. Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrica*, **92**:921–936, 2005.

REZIUMĖ

Skirtingų erdvinių asimptotikų įtaka klasifikavimo klaidai

L. Dreižienė, M. Karaliutė

Straipsnyje tiriama mokymo imčių didumo įtaka klasifikavimo klaidai, kai naudojami du skirtingi erdvinio ėmimo tipai, pagrįsti didėjančios ir fiksuotos erdvės asimptotikomis. Analizuojamas Gauso erdvinių duomenų klasifikavimo į vieną iš dviejų populiacijų, kurių vidurkio modeliai skirtingi, o geometriškai anizotropinė kovariacinė funkcija tokia pat, uždavinys. Naudojama tiesioginio pakeitimo Bajeso diskriminantinė funkcija (PBDF), kuri gaunama vietoj nežinomų parametrų tiesioginio pakeitimo būdu įstačius jų įvertinius.

Raktiniai žodžiai: Bajeso diskriminantinė funkcija, klasifikavimo rizika, anizotropijos koeficientas, atsitiktinis Gauso laukas, didėjančios srities asimptotika, baigtinės srities asimptotika.