# Application of spatial auto-beta models in statistical classification

**Eglė Zikarienė**[a,b] , **Kęstutis Dučinskas**[a,b]

[a]*Institute of Data Science and Digital Technologies, Vilnius University*
Akademijos str. 4, LT-08412 Vilnius

[b]*Faculty of Marine Technology and Natural Sciences, Klaipeda University*
Bijūnų str. 17, LT-91225 Klaipėda

E-mail: egle.zikariene@ku.lt; kestutis.ducinskas@ku.lt

**Abstract.** In this paper, spatial data specified by auto-beta models is analysed by considering a supervised classification problem of classifying feature observation into one of two populations. Two classification rules based on conditional Bayes discriminant function (BDF) and linear discriminant function (LDF) are proposed. These classification rules are critically compared by the values of the actual error rates through the simulation study.

**Keywords:** Bayes discriminant function; linear discriminant function; actual error rate; supervised classification

**AMS Subject Classification:** 62H30, 62M30

## Introduction

An approach for spatial classification using Bayes rules was introduced by DuÄinskas [5]. This approach is based on conditional distributions of observations to be classified given training sample for continuous spatial index. Case with discrete spatial index for Gaussian Markov random fields is explored in [4, 6, 7]. General statistical analysis of spatial non-gaussian data associated with exponential family and based on generalized linear models has been analysed in [2, 13]. Spatial discrimination based on BDF for feature observations having elliptically contoured distributions is implemented in [1, 8].

In this paper we focus on auto-beta models introduced by Besag [2] for the case when the sufficient statistic as well as the canonical parameter are one-dimensional.

Moller [12] presented the simulation algorithms for several spatial one-parameter auto-models. Specific attention will be paid to the multi-parameter auto-models that are properly studied in [10, 9, 3]. We consider a particular case of spatial auto-beta models for solving classification problem of feature observation by using plug-in discriminant functions.

This paper is organized as follows: the problem description and the introduction of spatial auto-beta model are presented in the first section and discriminant functions and error rates are analyzed in the next section; in Section 3 numerical experiments are described and the conclusions are in the last section.

# 1  Discriminant functions based on spatial auto-beta model and corresponding error rates

In this paper we consider random fields $\{Z(s) : s \in D \subset R^2\}$ and $\{Y(s) : s \in D \subset R^2\}$ as the feature and class label, respectively. Assume that feature values belong to $(0, 1)$ and class label takes value of 1 or 2. Suppose that $S_n = \{s_i \in D, \ i = 1, \ldots, n\}$ – training locations (STL) where feature observations with known class label are taken and feature values and class labels are denoted by $Z(s_i) = Z_i$ and $Y(s_i) = Y_i$, respectively, here $i = 0, 1, \ldots, n$. A training sample is denoted by $T = (Z', Y')$ where $Y = (Y_1, \ldots, Y_n)'$, $Z = (Z_1, \ldots, Z_n)'$.

We focus on the spatial auto-beta models (SABE) and supervised classification problem with fixed STL, when feature observation $Z_0$, $T = (Z', Y')$ are given. Then conditional distribution for unlabeled observation $Z_0$ under SABE model is

$$Z_0 \big| T = t, Y_0 = l \sim Beta(\mu_{0l}, \phi_{0l}).$$

Denote the full conditional density function for the feature by

$$f(z|a, b) = z^{\mu_0^l \phi_0^l - 1}(1 - z)^{(1-\mu_0^l)\phi_0^l - 1} / Be(\mu_0^l \phi_0^l; (1 - \mu_0^l)\phi_0^l), \tag{1}$$

where $\mu_0^l = E(Z_0|T = t, Y_0 = l; \Psi) = \frac{1 + A_{01}^l}{2 + A_{01}^l + A_{02}^l}$, $\phi_0^l = 2 + A_{01}^l + A_{02}^l$, with parameters $A_{01}^l = \alpha_{01}^l - \sum_{j \neq 0}(\eta_{0j}\ln(1 - z_j))$, $A_{02}^l = \alpha_{02}^l - \sum_{j \neq 0}(\eta_{0j}\ln(z_j))$, $\alpha_0^l = x_0'\beta^l$ where $x_i - m$ vector of explanatory variables and $\beta_m^l$ unknown regression coefficients. $\Psi$ denotes the set of all model parameters. $Be(\mu_0^l \phi_0^l; (1 - \mu_0^l)\phi_0^l)$ – is Euler Beta function. Spatial auto-beta models have been recently studied by several authors [11].

Then conditional BDF for SABE obtain the form

$$\begin{aligned} W(Z_0; \Psi) &= \ln\left(\pi_0^1 f_{01} / \pi_0^2 f_{02}\right) \\ &= \left(\alpha_{01}^1 - \alpha_{01}^2\right)\ln(Z_0) + \left(\alpha_{02}^1 - \alpha_{02}^2\right)\ln(1 - Z_0) + \gamma_0(\Psi), \end{aligned} \tag{2}$$

where $\gamma_0(\Psi) = ln(u)$, with $u = \frac{\pi_0^1 Be(A_{01}^2 + 1, A_{02}^2 + 1)}{\pi_0^2 Be(A_{01}^1 + 1, A_{02}^1 + 1)}$, $\pi_0^l$ – a prior probability in a population $\Omega_l$, $l = 1, 2$.

The prior probabilities depend on the location of focal observation and the number of neighbours: $\pi_0^l = \sum_{j \in NN_i^l} \frac{1}{d_{ij}} / \sum_{j \in NN_i} \frac{1}{d_{ij}}$ where $d_{ij}$ is the distance between sites $s_i$ and $s_j$, $i, j = 1, \ldots, n$, $NN_i = NN_i^1 + NN_i^2$, where $NN_i^l$ are sites belonging to the nearest neighbourhood set of $s_i$ in population $\Omega_l$, $l = 1, 2$.

So BDF allocates the observation in the following way: classify observation $Z_0$ given $Z = z$ to the population $\Omega_1$ if $W(Z_0; \Psi) \geqslant 0$ and to the population $\Omega_2$, otherwise.

We compare BDF with LDF for SABE in order to classify testing samples. In this work a modified LDF function is used where class conditional means and dispersions are used for the estimation. The modified LDF function:

$$L(Z_0; \Psi) = \left( \left( Z_0 - \frac{\mu^1 + \mu^2}{2} \right) \bigg/ \left( \pi_0^1 \sigma_1^2 + \pi_0^2 \sigma_2^2 \right) \right) (\mu^1 - \mu^2) + \gamma_0^*(\Psi), \qquad (3)$$

where $\gamma_0^*(\Psi) = \ln(\pi_0^1/\pi_0^2)$, $\pi_0^l$, – prior probability, $\mu^l$, $\sigma_l^2$ – conditional means and variance for the distribution $Beta(\mu_{0l}; \phi_{0l})$, $l = 1, 2$.

From the statistical decisions theory it is known that Bayes discriminant functions ensures the minimum of misclassification probability.

**Definition 1.** The Bayes error rate for the $W(Z_0; \Psi)$ specified in (2) is defined as

$$ER(\Psi) = \sum_{l=1}^{2} \pi_0^l P_l, \qquad (4)$$

where for $l = 1, 2$, $P_l = P_{lz}((-1)^l W(Z_0; \Psi) > 0) = \int H((-1)^l W(z_0; \Psi)) f_{0l}(z_0) dz_0$ with $H(\cdot)$ denoted the Heaviside step function: $H(\nu) := \mathbf{1}_{\nu > 0}$ and probability measure $P_{lz}$ is based on conditional Beta distribution with pdf $f_{0l}$ specified in (1).

The error rate for $L(Z_0; \Psi)$ specified in (3) which is denoted by $LER(\Psi)$ is defined in (4), when $W(Z_0; \Psi)$ is replaced by $L(Z_0; \Psi)$.

In practice parameter estimators are obtained by maximizing the pseudo-likelihood function, i.e.:

$$L_{PML}(\Psi) = \prod_{i=1}^{n_1} f_{i1} \prod_{i=n_1+1}^{n} f_{i2}, \qquad (5)$$
$$\widehat{\Psi} = \arg \max_{\Psi} \left\{ \log \left( L_{PML}(\Psi) \right) \right\}.$$

By replacing the parameters with their estimators in $W(Z_0; \Psi)$ and $L(Z_0; \Psi)$ we construct their plug-in versions denoted by $W(Z_0; \widehat{\Psi})$ and $L(Z_0; \widehat{\Psi})$.

**Definition 2.** The actual error rate for the $W(Z_0; \widehat{\Psi})$ is

$$AR(\widehat{\Psi}) = \sum_{l=1}^{2} \pi_0^l \widehat{P}_l, \qquad (6)$$

where $\widehat{P}_l = P_{lz}((-1)^l W(Z_0; \widehat{\Psi}) > 0)$, for $l = 1, 2$.

The actual error rate for the $L(Z_0; \widehat{\Psi})$ which is denoted by $LAR(\widehat{\Psi})$ is defined in (6), when $W(Z_0; \widehat{\Psi})$ is replaced by $L(Z_0; \widehat{\Psi})$.

## 2 Numerical experiments

To evaluate proposed classification procedure a few different scenarios were chosen that differ in model shape defined by different parameter values. Based on the chosen parameter scenarios and using the first order neighbour scheme $S = [1, N] \times [1, N]$: each site $i \in S$ has four neighbours denoted as $\{i_e = i + (1,0), i_w = i - (1,0), i_n = i + (0,1), i_s = i - (0,1)\}$ with obvious neighbour adjustments near the boundary. Conditional natural parameter expressions are chosen for population:

$$A_{i1}^l = \beta_1^l x_1 + \beta_2^l x_2 - \big(\eta_1\big(\ln(1 - z_{je}) + \ln(1 - z_{jw})\big) + \eta_2\big(\ln(1 - z_{jn}) + \ln(1 - z_{js})\big)\big),$$
$$A_{i2}^l = \beta_3^l x_1 + \beta_4^l x_2 - \big(\eta_1\big(\ln(z_{je}) + \ln(z_{jw})\big) + \eta_2\big(\ln(z_{jn}) + \ln(z_{js})\big)\big).$$

In this case parameter vector $\Psi^l = (\beta_1^l, \beta_2^l, \beta_3^l, \beta_4^l, \eta_1, \eta_2)$ First, based on the chosen parameter scenarios, 100 replications of data were generated. Each simulation was divided into two sets: 80% training sample and 20% testing sample. In the learning stage, training sample is used to build the model and in the testing stage sample is used to compare classification rules. In the learning stage all feature values of the attributes and spatial dependency are used to build the model and in the testing step one value is hidden. In this stage model parameter vector is considered unknown and model parameters are evaluated using maximum pseudo-likelihood method described in (5). Therefore, simulations are conducted on the lattice size $n = 16 \times 16$. Two types of parametrical structures were chosen: when all parameters are fixed except the class 2 mean tendency parameter and when spatial dependency parameter that describes effects of the north-south neighbourhood points is changing. Chosen parameter values are presented in Table 1.

The calculations were performed using 3 kinds of prior probabilities: 1) when $\pi_0^1 = \pi_0^2 = 0.5$ and actual error rate is noted as $AR^1$; 2) using inverse distance function with all training sample points $AR^2$; 3) using inverse distance function for neighbour points of up to fourth order $AR^3$ The Actual error rate ratio for different priors is presented in Fig. 1.

In both cases when beta parameter changes and when eta parameters change the ratio $AR^1/AR^3$ and $AR^2/AR^3$ is greater than 1, i.e. the smallest AR estimates are obtained when priors are calculated using the third method. Actual error rate values were compared for different classification rules when prior probabilities are calculated by the third method described above.

Actual error rate ratio curves are presented in A part of Fig. 2. When beta values are chosen less than 1 the ratio is greater than 1 and LDF based classification rule

**Table 1.** Parametric structures.

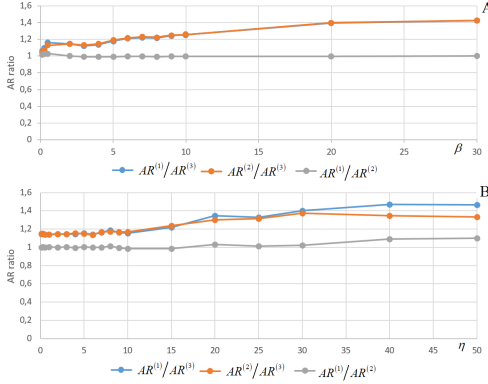| Parameter | Structure | Paramter values |
|---|---|---|
| $\beta^1$ | A, B | $\beta_1^1 = \beta_2^1 = \beta_3^1 = \beta_4^1$ |
| $\beta^2$ | A | $\beta_1^2 = \beta_2^2 = 1, \beta_3^2 = \beta_4^2 = \beta,$ $\beta = 0.125; 0.25; 0.5; 2; 3; 4; 5; 6; 7; 8; 9; 10; 20; 30; 40; 50.$ |
| | B | $\beta_1^2 = \beta_2^2 = 1, \beta_3^2 = \beta_4^2 = 2$ |
| $\eta$ | A | $\eta_1 = 1, \eta_2 = 2$ |
| | B | $\eta_1 = 1, \eta_2 = 0.125; 0.25; 0.5; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 20; 30; 40; 50$ |

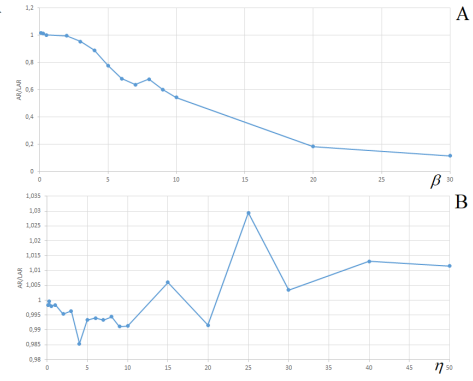**Fig. 1.** $AR(\widehat{\Psi})$ ratio curves with different prior probability.

**Fig. 2.** $AR(\widehat{\Psi})$ and $LAR(\widehat{\Psi})$ ratio curve with respect to $\beta$ and $\eta$.

is performing better. When $\beta > 1$ the ratio decreases and BDF based classification rule gains advantage. In B part when eta is chosen less than 10 the ratio is less than 1 and BDF based classification rule is performing better. When eta value is chosen 25 or greater LDF based classification rule gains advantage.

## 3   Discussion and conclusions

In this paper we proposed two classification rules for non-gaussian spatial data based on auto-beta models in the frameworks of Bayesian and linear discriminant functions. Simulation data study was conducted to estimate and empirically compare the BDF classifier with LDF classifier for various parametric structures and prior class probability models. Numerical analysis showed that:

1. While considering the situations with different prior probabilities better results are achieved by including fourth order neighbours in calculating prior probabilities in cases when prior probabilities are equal and when prior probabilities include all training points.

2. In situation when all parameters are fixed except for the second class mean tendency parameter the results show that class separability increases when mean tendency parameters increase. When $\beta < 1$ LDF has advantage and when $\beta > 1$ DBF performs better.

3. In situation when all parameters are fixed except for the spatial dependency parameter that is common for both classes the results yield that class separability decreases when spatial dependency parameter increases, and Actual error rate approximates 1.

The results of performed calculations in all examples give us the strong argument to encourage the users to consider non-gaussian spatial data models directly ignoring various data normalization procedures.

# References

[1] A. Batsidis, K. Zografos. Errors of misclassification in discrimination of dimensional coherent elliptic random field observations. *Stat. Neerl.*, **65**(4):446–461, 2011.

[2] J. Besag. Spatial interactions and the statistical analysis of lattice systems. *J. R. Stat. Soc. B*, **148**(4):1–36, 1974.

[3] E.D. Cepeda, A.V. Nunez. Spatial double generalized beta regression models extensions and application to study quality of education in Colombia. *J. Educat. Behav. Stat.*, **38**:604–628, 2013.

[4] L. Dreiziene, K. Ducinskas, L. Saltyte-Vaisiauske. Statistical classification of multivariate conditionally autoregressive Gaussian random field observations. *Spat. Stat.*, **28**:216–225, 2018.

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

[6] K. Ducinskas, L. Dreiziene. Risks of classification of the Gaussian Markov random field observations. *J. Classif.*, **35**:422–436, 2018.

[7] K. Dučinskas, L. Dreižienė, E. Zikarienė. Multiclass classification of the scalar Gaussian random field observation with known spatial correlation function. *Stat. Prob. Lett.*, **98**:107–114, 2015. https://doi.org/10.1016/j.spl.2014.12.008.

[8] K. Dučinskas, E. Zikarienė. Actual error rates in classification of the t-distributed random field observation based on plug-in linear discriminant function. *Informatica*, **26**(4):557–568, 2015.

[9] C. Hardouin, Y. Jian-Feng. Multi-parameter auto-models and their application. *Biometrika*, **95**(2):335–349, 2008.

[10] M.S. Kaiser, N. Cressie, J. Lee. Spatial mixture models based on exponential family conditional distributions. *Stat. Sinica*, **12**:449–474, 2002.

[11] B.M. Lagos-Álvarez, R. Fustos-Toribio, J. Figueroa-Zúniga. Geostatistical mixed beta regression: a Bayesian approach. *Stoch. Environ. Res. Risk Assess.*, **31**:571–584, 2017.

[12] J. Moller. Perfect simulation of conditionally specified models. *J. R. Stat. Soc. Ser. B*, **61**(1):251–264, 1999.

[13] H. Zhang. On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, **58**(1):129–136, 2002.

REZIUMĖ

**Erdvinių auto-beta modelių taikymas statistiniame klasifikavime**

*E. Zikarienė, K. Dučinskas*

Straipsnyje pristatomos naujos statistinio klasifikavimo taisyklės erdviams auto-beta modeliams. Jos paremtos sąlygine Bajeso ir tiesine diskriminantinėmis funkcijomis. Sprendžiamas uždavinys, kai erdvės taškas klasifikuojamas į 1 iš 2 populiacijų, su žinoma požymio reikšme ir mokymo aibe. Populiacijos apibrėžiamos regresoriais, bendraisiais ir klasių parametrais. Visi skaičiavimai atlikti simuliuotiems duomenims, su keletu skirtingų modelio parametrų rinkinių. Siūlomos klasifikavimo taisyklės palyginamos skaičiuojant tikrąją klasifikavimo klaidą, su skirtingais apriorinių tikimybių vertinimais.

*Raktiniai žodžiai*: Bajeso diskriminantinė funkcija; tiesinė diskriminantinė funkcija; tikroji klasifikavimo klaida; prižiūrimas klasifikavimas