INFORMATICA, 1998, Vol. 9, No. 4, 401–414

Analysis of the Risk Regret for Classification of Gamma Populations

Kęstutis DUČINSKAS
Klaipėda University
H. Manto 84, 5808 Klaipėda, Lithuania
e-mail: duce@hgf.ku.lt

Received: April 1998

Abstract. The sample-based rule obtained from Bayes classification rule by replacing unknown parameters by ML estimates from stratified training sample is used for classification of random observations into one of two widely applicable Gamma distributions. The first order asymptotic expansions of the expected risk regret for different parametric structure cases are derived. These are used to evaluate performance of the proposed classification rule and to find the optimal training sample allocation minimizing the asymptotic expected risk regret.

Key words: Bayes classification rule, stratified training sample, Gamma distribution, actual risk, risk regret, asymptotic expected risk regret.

1. Introduction

Suppose that individuals come from one of two mutually exclusive and exhaustive populations Ω_1, Ω_2 with positive prior probabilities π_1, π_2, respectively, where $\sum_{i=1}^{2} \pi_i = 1$. Let $X \in \mathbf{X} \subset \mathbf{R}$ be a random feature variable which is measured on each individual. Assume that the distribution of X for the individual from Ω_i has the probability density function (p.d.f.) $p_i(x; \Theta_i)$ which belongs to the parametric family of regular densities $F_i = \{p_i(x; \Theta_i), \Theta_i \in \mathbf{K} \subset \mathbf{R}^m\}$, $(i = 1, 2)$.

Further, the dependence of any functions on any distribution parameters will be suppressed in the cases when functions are evaluated at the true values of these parameters denoted by asterisk $*$, e.g., $p_i(x; \Theta_i^*) = p_i(x)$. A decision is to be made as to which population an individual randomly chosen from $\Omega = \bigcup_{i=1}^{2} \Omega_i$, belongs on the basis of an observed value of X. Let $d(\cdot)$ denote a classification rule (CR) formed for this purpose, where $d(x) = i$ implies that an individual with feature vector $X = x$ is to be assigned to the population Ω_i $(i = 1, 2)$. In effect, CR divides the feature space $\mathbf{X}$ into L mutually exclusive and exhaustive assignment regions U_1, U_2, where if X falls in U_i, then the individual is allocated to Ω_i $(i = 1, 2)$. Let $C(i, j)$ denote the cost of allocation when an individual from Ω_i is allocated to Ω_j and let $C(i, j)$ always be finite, i.e., $\max_{i,j=1,2} C(i, j) = C_0 < \infty$.

When prior probabilities $\{\pi_i\}$ and densities $\{p_i(x)\}$ are known, the risk $R(d(\cdot))$ associated with rule $d(\cdot)$ can be expressed as

$$R\left(d(\cdot)\right) = \sum_{i=1}^{2} \pi_i \int_{\mathbf{X}} C\left(i, d(x)\right) p_i(x) dx. \tag{1}$$

Then Bayes classification rule (BCR) $d_B(x)$ minimising the risk $R(d(\cdot))$ is defined as

$$d_B(x) = \arg\max_{i=1,2} l_i p_i(x), \tag{2}$$

where

$$l_i = \pi_i \left(C(i, 3-i) - C(i,i)\right), \quad (i = 1, 2). \tag{3}$$

Therefore, Bayes risk R_B is

$$R_B = \sum_{i=1}^{2} \pi_i \int_{\mathbf{X}} C\left(i, d_B(x)\right) p_i(x) dx = \inf_{\{d(\cdot) \in D\}} R\left(d(\cdot)\right), \tag{4}$$

where D is the set of all CR $d(\cdot)$ defined before.

The risk becomes the probability of misclassification (PMC) when $C(i,j) = 1 - \delta_{ij}$, where δ_{ij} is Kronecker's delta.

In practical applications, the density functions $\{p_i(x)\}$ are seldom completely known. Often they are only known up to the parameters $\{\Theta_i\}$, i.e., we can only assert that $p_i(x)$ is an element of the parametric family of density functions F_i. Under such conditions, it is customary to estimate unknown parameters from given data.

Suppose that in order to estimate unknown parameters Θ_1, Θ_2 there are M individuals of known origin on which feature vector X has been recorded. That data is referred to in pattern recognition literature as training sample (TS). The only case of independent observations in TS will be considered in this paper. Suppose that TS realized under separate sampling (SS) design. This sample often is called stratified sample. Then the feature vectors are observed for a sample of M_i individuals taken separately from each population Ω_i $(i = 1, 2)$.

Suppose that there are m_1 elements of all $\{\Theta_i\}$ known a priori to be distinct and let θ_0 be the vector of m_0 elements known a priori to be equal, i.e., $\Theta_i = (\theta_0', \theta_i')' = (\theta_0^1, \ldots, \theta_0^{m_0},\ \theta_i^1, \ldots, \theta_i^{m_1})$, where $\theta_i^k \neq \theta_j^k$ for $i \neq j$, $(i, j = 1, 2;\ k = 1, \ldots, m_1)$, and $m_0 + m_1 = m$. The prime denotes vector transpose.

Denote by α an $n = m_0 + 2m_1$-dimensional vector, which consists of θ_0 and (θ_1,θ_2), i.e.,

$$\alpha = (\theta_0', \theta_1', \theta_2')' = (\alpha^1, \ldots, \alpha^n). \tag{5}$$

Let $\mathbf{P} \subset \mathbf{R}^n$ be the set of all possible α, such that $\Theta_i \in \mathbf{K}$ $(i = 1, 2)$. Then suppose that

$$d(x, \alpha) = \arg\max_{i=1,2} l_i p_i(x, \Theta_i), \tag{6}$$

and

$$R_A(\alpha) = \sum_{i=1}^{2} \pi_i \int_{\mathbf{X}} C\left(i, d(x, \alpha)\right) p_i(x) dx. \tag{7}$$

The so-called estimative approach to the choice of sample-based classification rule $d_s(x)$ is used. The unknown parameters θ_0, θ_1, θ_2 are replaced by appropriate estimates $\widehat{\theta}_0$, $\widehat{\theta}_1$, $\widehat{\theta}_2$ obtained from the training data T in the BCR, i.e., $d_s(x) = d(x, \widehat{\alpha})$, where $\widehat{\alpha}' = (\widehat{\theta}'_0, \widehat{\theta}'_1, \widehat{\theta}'_2)'$. The case when $m_0 = 0$ means that all components of Θ_i are distinct for both populations.

The actual risk for the rule $d(x, \widehat{\alpha})$ is the risk of classifying a randomly selected individual with feature X and is designated by

$$R_A(\widehat{\alpha}) = \sum_{i=1}^{2} \pi_i \int_{\mathbf{X}} C\left(i, d(x, \widehat{\alpha})\right) p_i(x) dx. \tag{8}$$

For $C(i, j) = 1 - \delta_{ij}$, the actual risk becomes the actual error rate (AER), which is usually used for evaluation of performance of a sample-based rule.

It is obvious that $R_A(\alpha^*) = R_B$, where α^* is the true value of α.

DEFINITION 1. Risk regret (RR) for $d(x, \widehat{\alpha})$ is the difference between the actual risk $R_A(\widehat{\alpha})$ and Bayes risk R_B, and the expected regret risk (ERR) is the expectation of RR, i.e.,

$$ERR = E_T\left\{R_A(\widehat{\alpha})\right\} - R_B, \tag{9}$$

where $E_T\{R_A(\widehat{\alpha})\}$ denotes the expectation with respect to TS distribution.

It is obvious from (4), that RR is nonnegative random variable.

Unfortunately the exact distributions of RR usually are difficult to obtain. In those cases, large sample approximations to and asymptotic expansions for the distributions and expectations of RR are required. Only the situations when the feature variable X has the Gamma distribution for the individuals from Ω_i $(i = 1, 2)$ are considered.

The purpose of this paper is to find expansions of ERR when the maximum likelihood estimates (MLE) of unknown parameters of Gamma distributions for different parametric structure cases are used. These are used to evaluate the performance of sample-based CR and to find the optimal training sample allocation.

This is an extension of the result of Dučinskas (1995), who presented the asymptotic expansion of expected error regret in the situation when parameter vectors of classified distributions a priori had different all components. Kao *et al.* (1991) had also presented the asymptotic distribution of AER and asymptotic expansion for the expectation of AER. However, only the case of two normal populations with different means and common covariance was considered. Neil (1980) has found the general asymptotic distribution of AER for the classification into one of two populations. The probabilities of misclassification for exponential distributions (the special case of Gamma distributions) was obtained by Adegboye (1993).

The general asymptotic distribution of RR and asymptotic expansion of ERR in case of several populations and MLE of unknown parameters of classified distributions are derived in paper of Dučinskas (1997).

2. Notations and Auxiliary Results

Let ∇_α be the vector partial differential operator given by

$$\nabla_\alpha^t = \left(\frac{\partial}{\partial\alpha^1}, \ldots, \frac{\partial}{\partial\alpha^n}\right) \quad \text{and} \quad |\nabla_\alpha|^2 = \sum_{i=1}^{n} \left(\frac{\partial}{\partial\alpha^i}\right)^2$$

for any $\alpha = (\alpha^1, \ldots, \alpha^n) \in R^n$.

Similarly, ∇_α^2 denote the matrix second order differential operator

$$\nabla_\alpha^2 = \left\| \frac{\partial^2}{\partial\alpha^i \partial\alpha^j} \right\|_{i,j=1,2}.$$

Let

$$G_0(x) = l_1 p_1(x) - l_2 p_2(x)$$

and the real roots of the equation $G_0(x) = 0$ belonging to the support of $p_i(x)$ will be called threshold points and denoted by $x_{01}, \ldots, x_{0k}$.

Assume that $\mathbf{I}^i$ denotes the $m \times m$ Fisher information matrix for Θ_i, i.e.,

$$\mathbf{I}^i = E_i\left\{\nabla_{\Theta_i} \ln p_i(x) \nabla'_{\Theta_i} \ln p_i(x)\right\}, \tag{10}$$

where $E_i\{\cdot\}$ represents the expectation based on the distribution with the density function $p_i(x)$ $(i = 1, 2)$.

It is obvious that matrix $\mathbf{I}^i$ can be expressed as a block matrix

$$\mathbf{I}^i = \begin{pmatrix} \mathbf{I}_0^i & \mathbf{I}_{i0} \\ \mathbf{I}_{0i} & \mathbf{I}_i \end{pmatrix}, \tag{11}$$

where

$$\mathbf{I}_i = E_i\left\{\nabla_{\theta_i} \ln p_i(x) \nabla'_{\theta_i} \ln p_i(x)\right\}, \ \mathbf{I}_0^i = E_i\left\{\nabla_{\theta_0} \ln p_i(x) \nabla'_{\theta_0} \ln p_i(x)\right\}, \tag{12}$$

$$\mathbf{I}_{i0} = \mathbf{I}'_{0i} = E_i\left\{\nabla_{\theta_i} \ln p_i(x) \nabla'_{\theta_0} \ln p_i(x)\right\}, \quad (i = 1, 2). \tag{13}$$

Denote convergence in law by $\xrightarrow{\mathcal{L}}$.

Let the training sample realized under SS scheme is

$$T = (T'_1, T'_2), \tag{14}$$

where

$$T'_i = \left(X'_{i1}, \dots, X'_{iM_i}\right),$$

X_{ij} is the j-th observation from Ω_i, $i = 1, 2$ and $M = M_1 + M_2$.

Suppose that the regularity assumptions S for Lemma 3 of Dučinskas (1997) hold. Then the MLE $\widehat{\alpha}$ from T is a consistent estimate and as $M_i \to \infty$, $M_i/M \to r_i > 0$ $(i = 1, 2)$ satisfies

$$\sqrt{M}(\widehat{\alpha} - \alpha^*) \xrightarrow{\mathcal{L}} N_n(0, J_0^{-1}), \tag{15}$$

where

$$J_0(\alpha) = \begin{pmatrix} \sum_{i=1}^{2} r_i \mathbf{I}_0^i & r_1 \mathbf{I}_{01} & r_1 \mathbf{I}_{02} \\ & r_1 \mathbf{I}_1 & 0 \\ & & r_2 \mathbf{I}_2 \end{pmatrix}. \tag{16}$$

Let the random variable X_i has p.d.f. $p_i(x)$ $(i = 1, 2)$ and

$$V_j = \sum_{i=1}^{2} l_r (-1)^r \left(\nabla_{\theta_0} p_r(x_{0j}) - \mathbf{I}_{0r} \mathbf{I}_r^{-1} \nabla_{\theta_r} p_r(x_{0j})\right). \tag{17}$$

Theorem 1. *Let the regularity assumption S holds and let $R_A(\alpha)$ be twice continuously differentiable as a function of α in some neighborhood U_{α^*} and let $F(x_1, x_2)$ be a real valued function defined on R^2 that satisfies*

$$M(R_A - R_B) < F(X_1, X_2), \tag{18}$$

where $E\{F(X_1, X_2)\} < H$, $0 < H < \infty$.

Then the first order asymptotic expansion of the EER is

$$EER = \beta/M + \sum_{i=1}^{2} \rho_i/M_i + o(M^{-1}), \tag{19}$$

where

$$\beta = \sum_{j=1}^{k} V_j' \Lambda V_j \left| \nabla_x G_0(x_{0j}) \right|^{-1} / 2, \tag{20}$$

$$\rho_i = \sum_{j=1}^{k} l_i^2 \nabla_{\theta_i}' p_i(x_{0j}) \mathbf{I}_i^{-1} \nabla_{\theta_i} p_i(x_{0j}) \left| \nabla_x G_0(x_{0j}) \right|^{-1} / 2, \tag{21}$$

$$\Lambda = \left(\sum_{i=1}^{2} r_i (\mathbf{I}_0^i - \mathbf{I}_{0i} \mathbf{I}_i^{-1} \mathbf{I}_{i0}) \right)^{-1}. \tag{22}$$

Proof. The assertion of the stated theorem directly follows from one of Theorem 5 of Dučinskas (1997), after collection the terms at M^{-1} for two populations case.

REMARK 1. Let $m_0 = 0$, i.e., all true values of components of unknown distribution parameters are distinct. Then $\beta = 0$ in the first order asymptotic expansion of ERR defined in (19).

REMARK 2. If $m_0 \neq 0$, but θ_0^* are known, i.e., θ_0 is a nuisance parameter, then also $\beta = 0$ in (19).

Define the asymptotic expected risk regret as

$$AERR = \beta / M + \sum_{i=1}^{2} \rho_i / M_i. \tag{23}$$

The training sample allocation problem is viewed as follows. For a fixed value of M, let $W_i = M_i/M$ denote the proportions of observations taken from Ω_i. The design problem is to choose a value W_i^* for W_i that minimizes the $AERR$ defined in (23).

The W_1^* could be expressed explicitly (see Theorem 3 in Dučinskas (1995))

$$W_1^* = 1 \Big/ \left(1 + \sqrt{\rho_2 / \rho_1} \right). \tag{24}$$

3. Asymptotic Expansions of *ERR* for Gamma Populations

Suppose that distribution of X for the individual from Ω_i is Gamma with p.d.f.

$$p_i(x; \lambda_i, \eta_i) = \lambda_i^{\eta_i} x^{\eta_i - 1} e^{-\lambda_i x} \big/ \Gamma(\eta_i), \quad x > 0,\ \lambda_i > 0,\ \eta_i > 0\ (i = 1, 2). \tag{25}$$

Clearly, λ_i is a scale parameter and η_i is a shape parameter $(i = 1, 2)$.

The Gamma distributions are used by many authors in various models of reliability theory and in queing system analysis. This model is often suggested as a lifetime distribution. Special cases of Gamma distribution are Chi Square, Erlang and Exponential distributions.

Three parametric structure cases will be considered in this paper.

$$\textit{Case A.}\quad \lambda_1^* > \lambda_2^*,\ \eta_1^* = \eta_2^* = \eta^*,\ (\eta \ln(\lambda_2^*/\lambda_1^*) - \gamma) \big/ (\lambda_2^* - \lambda_1^*) > 0. \tag{26}$$

$$\textit{Case B.}\quad \lambda_1^* = \lambda_2^* = \lambda^*,\ \eta_1^* > \eta_2^*,$$
$$\exp\{-\gamma + (\eta_2^* - \eta_1^*)\ln\lambda + \ln(\Gamma(\eta_2^*)/\Gamma(\eta_1^*))\} \Big/ (\eta_1^* - \eta_2^*) > 0. \tag{27}$$

$$\textit{Case C.}\quad \lambda_1^* > \lambda_2^*,\ \eta_1^* > \eta_2^*. \tag{28}$$

Further superscripts A, B and C will be used to identify the considered cases.

Threshold points and assignment regions for these parametric structure cases will be shown in Figs. 1–3, assuming $\pi_1 = \pi_2 = 0.5$ and $C(i,j) = 1 - \delta_{ij}$ for simplicity.

Let $\Psi(\eta) = d\ln\Gamma(\eta)/d\eta$ and $\Psi'(\eta) = d^2\ln\Gamma(\eta)/d\eta^2$ are "digamma" and "trigamma" functions, respectively. Values of $\Psi(\eta)$ and $\Psi'(\eta)$ are tabulated in Abramowitz *et al.* (1964) for $1 \leqslant \eta \leqslant 2$. For $\eta > 2$ and $\eta < 1$, these functions can be evaluated from tabulated values and the recurrence relations

$$\Psi(\eta+1) = \Psi(\eta) + 1/\eta, \quad \Psi'(\eta+1) = \Psi'(\eta) - 1/\eta^2.$$

Unfortunately, the MLE of parameters of Gamma distribution usually can not be found in explicit form (see, e.g., Stacy and Mihram (1965)). So, further methods for solving of the MLE equations for three considered cases of parametric structure are proposed. Denote $a = \Psi'(\eta^*)$ and $a_i = \Psi'(\eta_i^*)$, $(i = 1,2)$.

The ML equations in the case A are

$$\frac{\partial \ln L}{\partial \lambda_i} = \sum_{i=1}^{M_i} \left(\widehat{\eta}/\widehat{\lambda}_i - x_{ij}\right) = 0 \quad (i = 1,2), \tag{29}$$

$$\frac{\partial \ln L}{\partial \eta} = \sum_{i=1}^{2} M_i \left(\ln\widehat{\lambda}_i + \ln\overline{x}_{iG}\right) - M\Psi(\widehat{\eta}) = 0 \quad (i = 1,2), \tag{30}$$

where L is the likelihood function for training sample T and $\overline{x}_{iG} = \prod_{j=1}^{M_i} x_{ij}^{1/M_i}$ is the geometric mean for sample T_i $(i = 1,2)$. Combining the ML equations (29), (30) gives

$$ln(\widehat{\eta}) - \Psi(\widehat{\eta}) = g, \tag{31}$$

where $g = \sum_{i=1}^{2} W_i \ln(\overline{x}_i/\overline{x}_{iG})$.

The ML estimates $\widehat{\eta}$ can be obtained from (31) by inverse interpolation for a given value of g. Alternatively, the following highly accurate approximation (see Greenwood and Durand (1960)) can be used to obtain $\widehat{\eta}$ directly:

$$\widehat{\eta} = (0.5001 + 0.1649g - 0.0544g^2)/g, \quad 0 < g \leqslant 0.577, \tag{32}$$

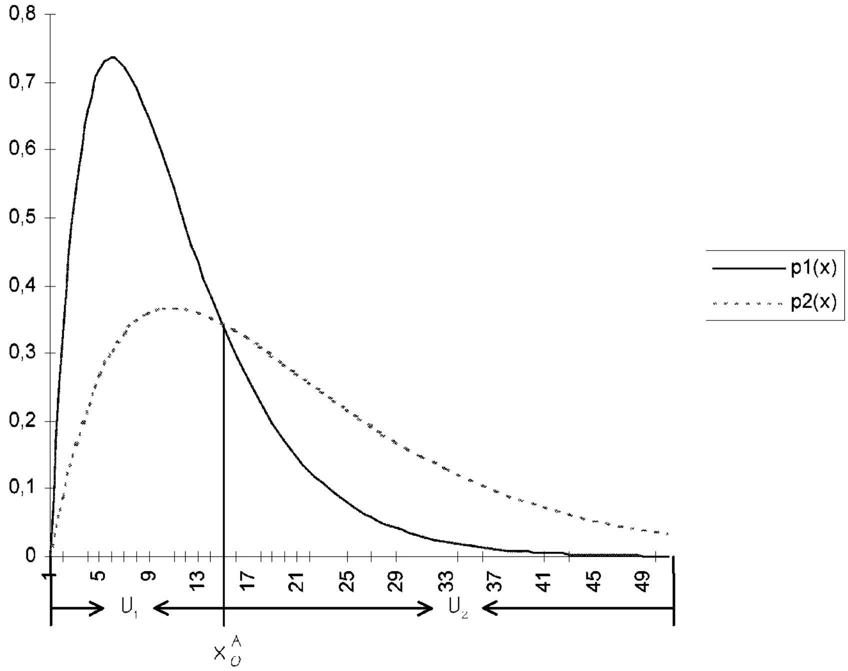

Fig. 1. Threshold point and assignment regions for the Case A ($\lambda_1^* = 2, \lambda_2^* = 1, \eta^* = 2$).

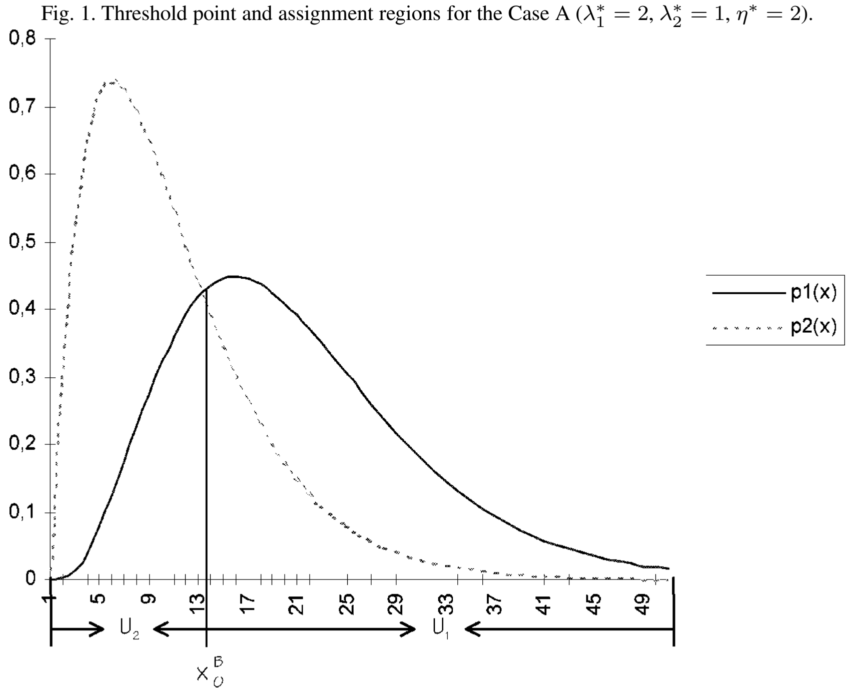

Fig. 2. Threshold point and assignment regions for the Case B ($\lambda^* = 2, \eta_1^* = 4, \eta_2^* = 2$).

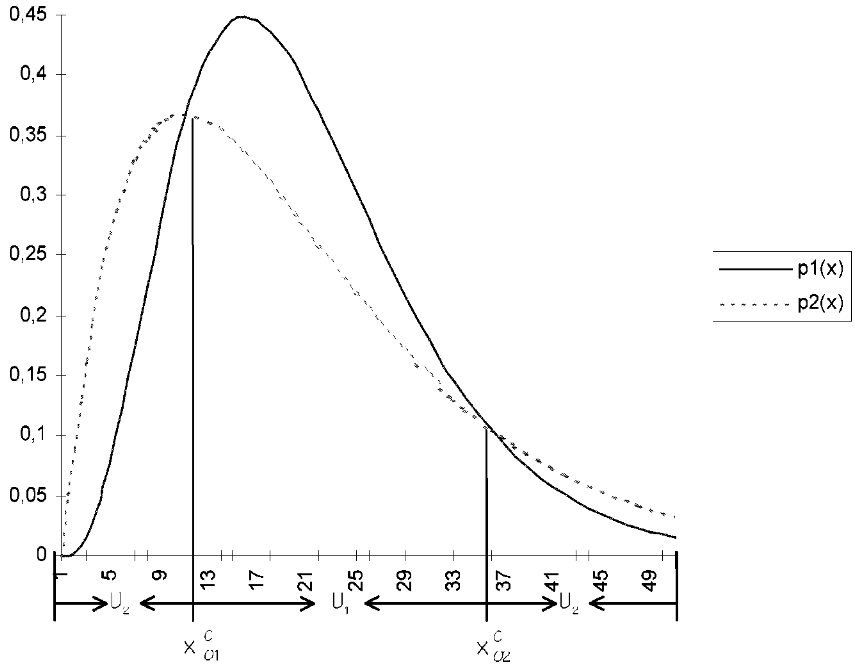

Fig. 3. Threshold point and assignment regions for the Case C ($\lambda_1^* = 2$, $\lambda_2^* = 1$, $\eta_1^* = 4$, $\eta_2^* = 2$).

$$\widehat{\eta} = (8.899 + 9.060g + 0.9775g^2)/\left((17.80 + 11.97g + g^2)g\right), \quad 0.577 \leqslant g \leqslant 17. \tag{33}$$

The error of either formula is less than 0.01%. The ML estimates $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ follow from (29).

The ML equations in the Case B are

$$W_1\widehat{\eta}_1 + W_2\widehat{\eta}_2 = \widehat{\lambda}\overline{x}, \tag{34}$$

and

$$\ln\widehat{\lambda} + \ln\overline{x}_{iG} = \Psi(\widehat{\eta}_i), \quad (i = 1, 2), \tag{35}$$

where $\overline{x} = \sum_{i=1}^{2} W_i\overline{x}_i$ is the total mean of T.

These nonlinear equations could be solved numerically by root isolation procedure described by Wingo (1987).

In the Case C, the ML equations can be written as

$$\widehat{\lambda}_i = \widehat{\eta}_i/\overline{x}_i, \tag{36}$$

$$\ln\widehat{\eta}_i - \Psi(\widehat{\eta}_i) = g_i, \tag{37}$$

where $g_i = \ln(\overline{x}_i/\overline{x}_{iG})$ $(i = 1, 2)$.

For solving of (37) the approximations given in (32), (33) are used substituting g by g_1 and g_2. Further the ML estimates $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ are obtained from (36).

Let $D = \sum_{i=1}^{2} r_i(a_i\eta_i^* - 1)/a_i$.

Theorem 2. *Suppose that the ML estimates of parameters of the classified Gamma distributions are used in the sample-based rule. Then the coefficients of the first order expansions of ERR in parametric structure Cases A and B are*

$$\beta^A = l_1 p_1(x_0^A)\gamma^2 / \left((a\eta^* - 1)(\lambda_1^* - \lambda_2^*)\eta^*\right), \tag{38}$$

$$\rho_i^A = l_1 p_1(x_0^A)(\eta^* - x_0^A\lambda_i^*)^2 / \left((\lambda_1^* - \lambda_2^*)\eta^*\right), \tag{39}$$

and

$$\begin{aligned}\beta^B = l_1 p_1(x_0^B)x_0^B \Bigg(\eta_1^* - \eta_2^* + \ln(\lambda^* x)\left(\frac{1}{a_1} - \frac{1}{a_2}\right) \\ + \Psi(\eta_2^*)/a_2 - \Psi(\eta_1^*)/a_1\Bigg)^2 \Big/ (\eta_1^* - \eta_2^*)D,\end{aligned} \tag{40}$$

$$\rho_i^B = l_1 p_1(x_0^B)x_0^B(\eta_i^* - x_0^B\lambda^*)^2 \Big/ \left((\eta_1^* - \eta_2^*)a_i(\lambda^*)^2\right), \tag{41}$$

where

$$x_0^A = \left(\eta \ln(\lambda_2^*/\lambda_1^*) - \gamma\right) / (\lambda_1^* - \lambda_2^*), \tag{42}$$

$$x_0^B = \exp\left\{-\gamma + (\eta_2^* - \eta_1^*)\ln\lambda + \ln\left(\Gamma(\eta_2^*)/\Gamma(\eta_1^*)\right)\right\} \Big/ (\eta_1^* - \eta_2^*), \tag{43}$$

are the threshold points for the Cases A and B, respectively.

Proof. In the Case A, the common shape parameter η corresponds to θ_0 and the scale parameters λ_1 and λ_2 correspond to θ_1 and θ_2, respectively. Then

$$\nabla_{\theta_i} p_i(x) = \partial p_i(x)/\partial\lambda_i = p_i(x)(\eta^*/\lambda_i^* - x),$$

$$\nabla_{\theta_0} p_i(x) = \partial p_i(x)/\partial\eta = p_i(x)\left(\ln(\lambda_i^* x) - \Psi(\eta^*)\right),$$

$$\mathbf{I}_i = \eta^*/(\lambda_i)^2, \quad \mathbf{I}_{0i} = -1/\lambda_i^*, \quad \mathbf{I}_0^i = \Psi'(\eta),$$

$$\nabla_x G_0(x) = \frac{\eta^* - 1}{x} G_0(x) - \left(l_1 p_1(x)\lambda_1^* - l_2 p_2(x)\lambda_2^*\right).$$

Let x_0^A be a positive root of $G(x_0^A) = 0$.

Evaluating these expressions in the threshold point $x = x_0^A$ and using the assertion of the Theorem 1 the proof was completed.

In the Case B, the common scale parameter corresponds to θ_0 and the shape parameters η_1 and η_2 correspond to θ_1 and θ_2, respectively.

Then

$$\nabla_{\theta_0} p_i(x) = \partial p_i(x)/\partial\lambda = p_i(x)(\eta_i^*/\lambda^* - x),$$

$$\nabla_{\theta_i} p_i(x) = \partial p_i(x)/\partial\eta_i = p_i(x)\left(\ln(\lambda_i^* x) - \Psi(\eta_i^*)\right),$$

$$\mathbf{I}_i = \Psi'(\eta_i^*), \quad \mathbf{I}_{0i} = -1/\lambda^*, \quad \mathbf{I}_0^i = \eta_i^*/(\lambda^*)^2,$$

$$\partial G_0/\partial x = l_1 p_1(x)\left(\frac{\eta_1 - 1}{x} - \lambda_1\right) - \pi_2 p_2(x)\left(\frac{\eta_2 - 1}{x} - \lambda_2\right).$$

Substituting corresponding terms in (20), (21) the formulae (40), (41) were obtained.

REMARK 3. If $\gamma = 0$, then from (38) follows that $\beta^A = 0$. Thus the first order asymptotic expansion of the EER is the same as in the situation when η^* is known.

The threshold points x_0^A, coefficients β^A, $\{\rho_i^A\}$ and W_1^* for $C(i,j) = 1-\delta_{ij}$, $\lambda_2^* = 1$, $\eta^* = 2$ are given in Table 1. The threshold points x_0^B, coefficients β^B, $\{\rho_i^B\}$ and W_1^* for $C(i,j) = 1 - \delta_{ij}$, $\lambda^* = 2$, $\eta_2^* = 1$ are given in Table 2.

Theorem 3. *Let the assumptions of Theorem* 2 *hold. Then the coefficients of the first order expansion of ERR in parametric structure Case* C *are equal*

$$\beta^C = 0, \tag{44}$$

$$\rho_i^C = l_1 \sum_{j=1}^{k} p_1(x_{0j}^C)\Bigg(\left(\ln(\lambda_i^* x_{0j}^C) - \Psi(\eta_i^*)\right)^2 \eta_i^* + 2(\eta_i^* - \lambda_i^* x_{0j}^C)\left(\ln(\lambda_i^* x_{0j}^C) - \Psi(\eta_i^*)\right) + (\eta_i^* - \lambda_i^* x_{0j}^C)^2 a_i\Bigg) \Bigg/ \left((a_i\eta_i^* - 1)\left|\frac{\eta_1^* - \eta_2^*}{x_{0j}^C} + \lambda_2^* - \lambda_1^*\right|\right), \tag{45}$$

where $\{x_{0j}^C\}$ are the positive roots of the equation

$$x(\lambda_2^* - \lambda_1^*) + \ln x(\eta_1^* - \eta_2^*) + \gamma + \eta_1^* \ln\lambda_1^* - \eta_2^* \ln\lambda_2^* + \ln\left(\Gamma(\eta_2^*)/\Gamma(\eta_1^*)\right) = 0. \tag{46}$$

The proof of the stated theorem is directly analogous to one of Theorem 2 of this paper.

REMARK 4. If equation (46) has no positive roots, then $\rho_i^C = 0$. This is because the support of Gamma p.d.f. is the set of all positive real numbers.

The results of Theorems 2 and 3 can be used in evaluation of performance for the sample-based CR with plugged ML estimates of the shape and scale parameters

Table 1
The values of x_0^A, β^A, $\{\rho_i^A\}$ and W_1^* for $\lambda_2^* = 1$, $\eta^* = 2$

λ_1^*	x_0^A	β^A	ρ_1^A	ρ_2^A	W_1^*
	$\pi_1 = 0.5$				
2	1.38629	0.00000	0.05172	0.03263	0.44268
3	1.09861	0.00000	0.07687	0.03719	0.41022
4	0.92420	0.00000	0.08800	0.03537	0.38800
5	0.80472	0.00000	0.09211	0.03214	0.37135
6	0.71670	0.00000	0.09260	0.02882	0.35810
7	0.64864	0.00000	0.09118	0.02580	0.34723
8	0.59413	0.00000	0.08878	0.02315	0.33803
9	0.54931	0.00000	0.08588	0.02086	0.33014
10	0.51169	0.00000	0.08278	0.01887	0.32316
11	0.47958	0.00000	0.07962	0.01716	0.31705
12	0.45180	0.00000	0.07651	0.01567	0.31156
13	0.42749	0.00000	0.07350	0.01436	0.30652
14	0.40601	0.00000	0.07061	0.01322	0.30202
15	0.38686	0.00000	0.06786	0.01221	0.29784
16	0.36968	0.00000	0.06525	0.01132	0.29404
17	0.35415	0.00000	0.06277	0.01052	0.29047
18	0.34004	0.00000	0.06044	0.00981	0.28718
19	0.32716	0.00000	0.05823	0.00917	0.28410
20	0.31534	0.00000	0.05615	0.00859	0.28116
	$\pi_2 = 0.8$				
2	2.77259	0.03861	0.21779	0.01034	0.17891
3	1.79176	0.03327	0.17011	0.00065	0.05822
4	1.38629	0.02574	0.14519	0.00435	0.14755
5	1.15129	0.02028	0.12844	0.00656	0.18434
6	0.99396	0.01639	0.11560	0.00745	0.20246
7	0.87969	0.01356	0.10516	0.00764	0.21231
8	0.79217	0.01142	0.09641	0.00748	0.21786
9	0.72259	0.00977	0.08893	0.00716	0.22103
10	0.66572	0.00847	0.08245	0.00677	0.22273
11	0.61821	0.00742	0.07677	0.00636	0.22350
12	0.57783	0.00657	0.07175	0.00596	0.22373
13	0.54302	0.00586	0.06729	0.00558	0.22358
14	0.51265	0.00526	0.06330	0.00522	0.22310
15	0.48589	0.00476	0.05971	0.00489	0.22250
16	0.46210	0.00432	0.05646	0.00459	0.22187
17	0.44080	0.00395	0.05350	0.00431	0.22108
18	0.42159	0.00363	0.05081	0.00405	0.22017
19	0.40418	0.00334	0.04835	0.00382	0.21941
20	0.38830	0.00309	0.04608	0.00360	0.21845

Table 2
The values of x_0^B, β^B, $\{\rho_i^B\}$ and W_1^* for $\lambda_2^* = 1, \eta^* = 2$

η_1^*	x_0^B	β^B	ρ_1^B	ρ_2^B	W_1^*
			$\pi_1 = 0.2$		
2	2.00000	0.491125721972	0.090877575641	0.080168927531	0.484
3	0.70711	0.000071605740	0.054731161047	0.000896540668	0.113
4	0.43679	0.062239130662	0.023257603988	0.000006561420	0.016
5	0.31947	0.013758025648	0.002515806805	0.000002320187	0.029
6	0.25325	0.001083157093	0.000141107831	0.000000125527	0.028
7	0.21042	0.000046986756	0.000005008052	0.000000003622	0.026
8	0.18033	0.000001326421	0.000000123978	0.000000000070	0.023
9	0.15796	0.000000026637	0.000000002273	0.000000000001	0.020
			$\pi_1 = 0.8$		
2	0.12500	0.306642814028	0.046227155626	0.003328974161	0.211
3	0.17678	0.198095844344	0.027516323925	0.000394189966	0.106
4	0.17334	0.046735602810	0.005336322964	0.000029445362	0.069
5	0.15974	0.005042310896	0.000498583963	0.000001418139	0.050
6	0.14545	0.000303233532	0.000027147656	0.000000046164	0.039
7	0.13256	0.000011561928	0.000000965771	0.000000001073	0.032
8	0.12135	0.000000304299	0.000000024199	0.000000000018	0.027
9	0.11169	0.000000005864	0.000000000450	0.000000000001	0.023

of Gamma distributions. The presented first order expansions of the ERR enable researchers to find the optimal stratified training sample allocation, i.e., W_1^*, when the total training sample size M is fixed. The calculations given in Tables 1 and 2 show that equal training sample sizes for both populations often are not optimal, even when the prior probabilities of populations are equal, i.e., $\pi_1 = \pi_2$.

References

Abramowitz, M., and I. Stegun (1965). *Handbook of Mathematical Functions*. New York: Dover.

Adegboye, O.S. (1993). The optimal classification rule for exponential populations. *Austral. J. Statist.*, **35**(2), 185–194.

Dučinskas, K. (1995). Optimal training sample allocation and asymptotic expansions for error rates in discriminant analysis. *Acta Appl. Math.*, **38**, 3–11.

Dučinskas, K. (1997). An asymptotic analysis of the regret risk in discriminant analysis under various training schemes. *Lith. Math. J.*, **37**(4), 337–351.

Greenwood, J.A., and D. Durand (1960). Aids for fitting the Gamma distribution by maximum likelihood. *Technometrics*, **2**.

Kao, T.C., and G.P. McCabe (1991). Optimal allocation for normal discrimination and logistic regression under stratified sampling. *J. Amer. Statist. Assoc.*, **88**, 432–436.

O'Neill, T.J. (1980). The general distribution of the error rate of a classification procedure with application to logistic regression discriminant. *J. Amer. Statist. Assoc.*, **75**, 154–160.

Wingo, D.R. (1987). Computing maximum-likelihood parameter estimates of the generalized Gamma distribution by numerical root isolation. *IEEE Transactions on reliability*, **R-36**(5), 586–590.

K. Dučinskas graduated from the Vilnius University in 1976 in applied mathematics, where he received Doctor Degree in 1983. He is a head of System Research Depertment and associate professor of Klaipėda University. Present research interests include discriminant analysis and statistical pattern recognition.

Gama skirstinių klasifikavimo rizikos analizė

Kęstutis DUČINSKAS

Straipsnyje nagrinėjamas stebėjimų, pasiskirsčiusių pagal Gama dėsnį, klasifikavimo uždavinys. Pateikti pirmos eilės asimptotiniai skleidiniai laukiamam klasifikavimo rizikos padidėjimui, kai nežinomos parametrų reikšmės vertinamos iš mokymo imties, panaudojant maksimalaus tikėtinumo metodą. Gautos formulės gali būti naudojamos vertinant klasifikavimo taisyklės "kokybę", nustatant optimalius mokymo imčių dydžius.