VELUT ARBOR ÆVO

Asymptotic behavior of the probabilities misclassification for discriminant functions with weighting

by

Tatijana Pavlenko
Department of Mathematics
Uppsala University

Technical Report No. 9701, January 15, (1997)

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

Asymptotic behavior of the probabilities misclassification for discriminant functions with weighting

By

TATIJANA PAVLENKO

Abstract: Discriminant analysis is considered when the dimension of observed vector and number of parameters are both large. Alternative to the standard approach is presented. This alternative is characterized by sequence of classification problems instead of one isolated problem and weighting discriminant function instead of ordinary classification rule. The asymptotic probabilities of misclassification are obtained via the weighting function when the distributions of populations are known exactly. The type of weighting function for which the probability of misclassification achives a minimum is defined.

Key words: Discriminant analysis, growing dimension asymptotic, weighting discriminant function, asymptotic probabilities of misclassification.

1. INTRODUCTION

In this study the term classification is understood as allocation of an object to one of two populations solely on the basis of the observed value of p-dimensional random vector accosiated with the object. Let populations are given by the density functions which depend on multivariate parameter. If the dimension of observed vector and the number of parameters are both large the problem of classification has to be formulated with taking into account the effect of growing dimension. To study such kind of problems Kolmogorov proposed an interesting approach when instead of one isolated problem of classification a sequence (with $n \to \infty$) of classification problems is considered. The dimension of observations and number of parameters grows indefinitely with the growth of problem's number n, and convergence of populations takes place. The crucial relation

$$\lim_{n\to\infty} \frac{p}{n} = c \tag{1.1,}$$

where $0 < c < \infty$ is a constant, is defined as the Kolmogorov asymptotic or, since p increases to infinity, the asymptotic of growing dimension.

Correspondence: Tatijana Pavlenko, Department of Mathematics, Box 480, S-751 06 Uppsala, Sweden

It turns out such statement of the problem results in elegant limit expressions for probabilities of misclassification. The first and indeed most important paper in this studies is the paper by Deev (1970). Here the explicit dependence of expected number of errors on number of estimated parameters is obtained assuming the observations have a multivariate normal densities. Later, the asymptotic behavior of probabilities misclassification has been investigated for different types of discriminant functions, see Meshalkin (1976), Serdobolskiy (1982), Friedman (1989).

Another approach to the classification problem with growing dimension has been presented by Girko (1987). It was suggested how to correct the traditional discrimination criteria in order to reduce the classification errors. In the papers by Girko, Pavlenko (1989) and Pavlenko (1989) such estimators were constructed for both linear and quadratic discriminant functions and for Mahalanobis distance under the assumption of a normal population's distribution. The classification problem with growing dimension when the variables are partitioned into an increasing number of blocks (non-empty subsets) with fixed dimension was treated in the paper by Pavlenko, von Rosen (1996). In this case the probabilities of misclassification are derived based on the asymptotic normality of the discriminant function. Asymptotic probabilities of misclassification are found via a certain distance between populations and the relation between the dimension of observed vector and size of training samples. The obtained formulae for misclassification can be applied to a reasonable wide class of population's distribution since no assumptions about normality of observations have been made. The conditions are required concern only uniform regularity of density functions and asymptotic convergence of populations. It is important to note that convergence of populations from Pavlenko, von Rosen (1996) leads to conclusion on equal input into block's distance made by each block of variables. In the present article we propose to examine the case when inputs of blocks are different. We shall consider the classification problem with weighting of the blocks of observed vectors depending on its input into the distance between populations. Concerning the vector (block) with given dimension we will assume the more differences of its probability distribution in populations Π_1 and Π_2, the more informativity of the vector. As the measure of difference between two probability distributions the information type of distance (Mahalanobis distanse, Kullback-Leibler distance, variation Kolmogorov distance, e.s.) is usually used. We take Kullback-Leibler distance to describe the informativity of independent sets of variables.In such a case, the more input of the block of variables into Kullback-Leibler distance, the block is more informative. The measure of informativity defines the weighting function which will be included in the discrimination rule. The question then arraises what is the effect such kind of weighting on classification. The main goal of this study is to analyse the dependence of misclassification probabilities upon the weighting function.

The paper is organised as follows. In the part 2 the exact statement of classification problem with weighting in the growing dimension case is formulated with main assumptions about the structure of variables, regularity of distributions and convergence of populations.

In the part 3 the limit expression for the probabilities of misclassification via weigting function will be derived based on the asymptotic normality of discriminant function. Limit results are obtained provided the sequence of classification problems is considered. The type of weighting function for which the measure of misclassification attains a minimum is established.

2. Assumptions

Let $L(x,\theta)$ be a density (with respect to the σ-finite measure μ) of distribution of the p-dimensional random variable $x \in R^p$, depending on the l-dimensional parameter $\theta \in R^l$. Let θ^1 and θ^2 be parameter values corresponding to populations Π_1 and Π_2 respectively. The classification is performed according to the rule

$$d(x) > d_0,$$

where

$$d(x) = \ln \frac{L(x,\theta^1)}{L(x,\theta^2)}, \tag{2.1}$$

and d_0 is a constant. The misclassification probabilities are defined in the regular way

$$\alpha_1 = \int_{d(x)\leq d_0} L(x,\theta^1)\mu(\,dx),$$

$$\alpha_1 = \int_{d(x)> d_0} L(x,\theta^2)\mu(\,dx). \tag{2.2}$$

It is convenient to describe a sequence of classification problems in a form

$$\left\{p, l, x, \theta^1, \theta^2, L(x,\theta), d(x), \alpha_1, \alpha_2\right\}_n, \tag{2.3}$$

where $n = 1, 2, \ldots$ is the number of classification problem, p and l increase to infinity such a way that Kolmogorov's condition (1.1) is hold. Concerning the arguments in the sequence we shall consider the following series of assumptions.

2A. Structure of x and θ: We assume that the components of x as well as θ can be devided into k mutually independent non-empty subsets (blocks) $x = (x_1, \ldots, x_k)$, $\theta = (\theta_1, \ldots, \theta_k)$, where $x_i \in R^m$, is a random vector corresponding to the ith block, $\theta_i^\nu \in R^r$ is a vector of parameters corresponding to the ith block, and

$$\lim_{n\to\infty} \frac{m}{n} = 0,$$

$$\lim_{n\to\infty} \frac{r}{n} = 0, \tag{2.4}$$

$\nu = 1, 2$. Without loss of generality we will suppose that the dimension of x_i and θ^ν is the same for all $i = 1, \ldots, k$. Since the block dimension is bounded, and dimension of x and θ grows, with $n \to \infty$, k also grows indefinitely such a way that the limit

$$\lim_{n\to\infty} \frac{k}{n} = \rho < \infty \tag{2.5}$$

exist. With regard to block structure, (2.1) becomes

$$d(x) = \sum_{i=1}^{k} \ln \frac{L_i(x_i, \theta_i^1)}{L_i(x_i, \theta_i^2)}. \tag{2.6}$$

A special case is when each block consists of one observation, i.e. $m = 1$.

2B. FUNCTIONS $L(x, \theta^1)$ AND $L(x, \theta^2)$: We restrict the choice of functions $L(x, \theta^1)$ and $L(x, \theta^2)$ by the set of standard regularity conditions. Assume that the functions $L(x, \theta^\nu)$ are thrice differentiable in all components of $\theta^\nu, \nu = 1, 2$. All first-, second- and third- order derivatives of $\ln L_i(x_i, \theta_i^\nu)$ with respect to all components of θ_i^ν are integrable with respect to $L(x, \theta_\nu)\mu(\,dx)$. Let

$$l_i(x_i, \theta_i^\nu) = \ln L_i(x_i, \theta_i^\nu)$$

and we also suppose that

$$\left| \frac{\partial l_i(x_i, \theta_i^\nu)}{\partial \theta_{ij}^\nu} \right| < M_1(x_i),$$

for all $j = 1, ..., m$.

$$\left| \frac{\partial^2 l_i(x_i, \theta_i^\nu)}{\partial \theta_{ij_1}^\nu \theta_{ij_2}^\nu} \right| < M_2(x_i), \tag{2.7}$$

for all $j_1, j_2 = 1, ..., m$.

$$\left| \frac{\partial^3 l_i(x_i, \theta_i^\nu)}{\partial \theta_{ij_1}^\nu \theta_{ij_2}^\nu \theta_{ij_3}^\nu} \right| < M_3(x_i),$$

for all $j_1, j_2, j_3 = 1, ..., m$, where $M_r(x_i)$, $r = 1, 2, 3$, is independent of θ_i^ν, integrable and uniformly bounded over the k blocks. All integrals of $M_1^3(x_i), M_2^2(x_i), M_1(x_i), M_2(x_i)$ and $M_3(x_i)$ with respect to $L(x, \theta^\nu)\mu(\,dx)$ are uniformly bounded, $\nu = 1, 2$. The Fisher information matrices

$$I^\nu = I(\theta^\nu)_{ij} = \int \frac{\partial l(x, \theta^\nu)}{\partial \theta_i^\nu} \frac{\partial l(x, \theta^\nu)}{\partial \theta_j^\nu} L(x, \theta^\nu)\mu(\,dx) \tag{2.8}$$

are positive definite and their eigenvalues are bounded, $i, j = 1, ..., m$. Note that

$$\frac{\vec{\partial}}{\partial \theta^\nu} = (\frac{\partial}{\partial \theta_1^\nu}, ..., \frac{\partial}{\partial \theta_r^\nu}),$$

$$\nu = 1, 2.$$

By assumptions (2.4)-(2.5) the matrices are of block-diagonal form with blocks $I_i^\nu = I(\theta_i^\nu)$ of dimension $m \times m, \nu = 1, 2$.

2C. THE RATE OF CONVERGENCE OF POPULATIONS: We defined the information Kullback-Leibler distance between two populations as follows

$$J_n = \int \ln \frac{L(x, \theta^1)}{L(x, \theta^2)} [L(x, \theta^1) - L(x, \theta^2)]\mu(\,dx) = \sum_{i=1}^{k} J_{n_i} = \sum_{i=1}^{k} (J_{n_i}^1 - J_{n_i}^2), \tag{2.9}$$

where

$$J_{n_i}^{\nu} = \int \ln \frac{L_i(x_i,\theta_i^1)}{L_i(x_i,\theta_i^2)} L(x,\theta^{\nu})\mu(dx),$$

$\nu = 1,2, i = 1,...k$. Since $n \to \infty (k \to \infty)$ we have to suppose that

$$\lim_{n\to\infty} J_n = \lim_{n\to\infty} \sum_{i=1}^{k} (J_{n_i}^1 - J_{n_i}^2) < \infty. \tag{2.10}$$

The condition (2.9) is fairly natural if we study the case when the probabilities of misclassification do not tend to zero. In this case the distance between two populations must be bounded. Note that according to (2.9) $J_{n_i}^{\nu} = O(1/n)$ which is eqivalent to the condition

$$\mid \theta_{ij}^2 - \theta_{ij}^1 \mid = O(n^{-1/2}), \tag{2.11}$$

$\nu = 1,2, i = 1,...,k, j = 1,...,m$.

2D. THE WEIGHTING OF THE BLOCKS IN THE DISCRIMINANT FUNCTION: Define the weighting function $\lambda_n = \lambda_n(\frac{nJ_{n_i}}{2})$ upon the input of ith block in the distance between populations, $i = 1,...,k$. Suppose that $\lambda_n(u)$ are nonnegative for all $u > 0$, uniformly bounded and

$$\lim_{n\to\infty} \lambda_n(u) = \lambda(u) < \infty. \tag{2.12}$$

The discriminant function with the weighting of the blocks by $\lambda_n(u)$ has the form

$$d_{\lambda_n}(x) = \sum_{i=1}^{k} \lambda_n(\frac{nJ_{n_i}}{2}) d_i(x_i). \tag{2.13}$$

Next in a sequence of classification problems (2.3) we will consider $d_{\lambda_n}(x)$ instead $d(x)$. We introduce the functions

$$H_n(u) = \frac{1}{k}\sum_{i=1}^{k} \chi(u - \frac{nJ_{n_i}}{2}),$$

where

$$\chi(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0. \end{cases}$$

Note that $H_n(u)$ is a function of bounded variation and continous from the left. Given some interval (a,b) and a function $f(u)$ we can form the sum

$$\Delta_n = \sum_{i=1}^{n} f(\tilde{u}_i)[H_n(u_i) - H_n(u_{i-1})], \tag{2.14}$$

for a division of (a,b) by points u_i such that $a < u_1 < ... < u_n < b$ and arbitrary $\tilde{u}_i \in (u_{i-1}, u_i)$. If the sum (2.14) tends to a limit

$$\Delta = \lim_{n\to\infty} \Delta_n$$

where $n \to \infty$ such that the length of each interval tends to zero we can defined Stieltjes integral of $f(u)$ with respect to $H(u)$ and denoted it by

$$\Delta = \int_a^b f(u)\, dH(u). \tag{2.15}$$

3. RESULTS

Under the conditions 1.-4. we establish the limit probabilities of misclassification depending on $m, \rho, d_0, H(u)$ and $\lambda(u)$. We start with the some important lemmas.

LEMMA 1. *Suppose that conditions (2.4)-(2.11) hold. Then for all* $i = 1, ..., k$

$$J_{n_i}^1 = -J_{n_i}^2 = \frac{1}{2}(\theta_i^2 - \theta_i^1)' I(\theta_i^1)(\theta_i^2 - \theta_i^1) + O(n^{-3/2}) \tag{3.1}$$

and therefore

$$J_{n_i} = 2J_{n_i}^1 + O(n^{-3/2}).$$

PROOF: Due to assumption 2. we can consider a Tailor series expansion of $l_i(x_i, \theta_i^2)$ from (2.8) around θ_i^1 up to the third order term. For the ith block we have

$$J_{n_i}^1 = \int \Big[(\theta_i^2 - \theta_i^1)' \frac{\partial}{\partial \theta_i} l(x_i, \theta_i^1) + \frac{1}{2}(\theta_i^2 - \theta_i^1)' (\frac{\partial}{\partial \theta_i} \times \frac{\partial'}{\partial \theta_i} l_i(x_i, \theta_i^1))(\theta_i^2 - \theta_i^1)$$

$$+ R_i(x_i, \theta_i^1, \theta_i^2)\Big] L(x, \theta^1) \mu(\, dx),$$

where

$$R_i(x_i, \theta_i^1, \theta_i^2) = \frac{1}{6}((\theta_i^2 - \theta_i^1), \frac{\partial}{\partial \theta_i})(\theta_i^2 - \theta_i^1)' (\frac{\partial}{\partial \theta_i} \times \frac{\partial'}{\partial \theta_i} l_i(x_i, \tilde{\theta}_i))(\theta_i^2 - \theta_i^1)$$

and $\tilde{\theta}_i = (\tilde{\theta_{i_1}}, ..., \tilde{\theta_{i_m}})$ represent an arbitrary value for each element in the interval $(\theta_{ij}^1, \theta_{ij}^2)$, $j = 1, ..., m$. For the first order term we obtain

$$\int \frac{\partial}{\partial \theta_i} L_i(x_i, \theta_i^1) \mu(\, dx) = 0. \tag{3.2}$$

In the second order terms the Fisher's information matrix for the ith block appears, for which

$$I^i(\theta_i^1) = \int \frac{\partial}{\partial \theta_i} \times \frac{\partial}{\partial \theta_i} l_i(x_i, \theta_i^1) L(x, \theta^1) \mu(\, dx). \tag{3.3}$$

For the third order term we have

$$\int R_i(x_i, \theta_i^1, \theta_i^2) L(x, \theta_1) \mu(\, dx) < O(n^{-3/2}) \int M_3(x) L(x, \theta^1) \mu(\, dx) = O(n^{-3/2}),$$

according to (2.6) and (2.8). Therefore

$$J_{n_i} = \frac{1}{2}(\theta_i^2 - \theta_i^1)' I^i(\theta_i^1)(\theta_i^2 - \theta_i^1) + O(n^{-3/2}).$$

Next we consider the Taylor series expansion for $l_i(x_i, \theta_i^2)$ and $L_i(x_i, \theta_i^2)$ from (2.10) around θ_i^1 up to the third order terms. By the conditions (2.6), (2.7) and (2.8) we obtain

$$J_{n_i}^2 = -\int \frac{\partial'}{\partial\theta_i} l_i(x_i,\theta_i^1) L_i(x_i,\theta_i^1)\mu(\,dx)$$

$$+\frac{1}{2}(\theta_i^2-\theta_i^1)' \int \frac{\partial}{\partial\theta_i} \times \frac{\partial'}{\partial\theta_i} l_i(x_i,\theta_i^1) L_i(x_i,\theta_i^1)\mu(\,dx)(\theta_i^2-\theta_i^1)$$

$$-(\theta_i^2-\theta_i^1)' \int \frac{\partial}{\partial\theta_i} l_i(x_i,\theta_i^1) \times \frac{\partial'}{\partial\theta_i} L_i(x_i,\theta_i^1)\mu(\,dx)(\theta_i^2-\theta_i^1) + O(n^{-3/2}) \tag{3.4}$$

$$= \frac{1}{2}(\theta_i^2-\theta_i^1) I^i(\theta_i^1)(\theta_i^2-\theta_i^1)$$

$$-(\theta_i^2-\theta_i^1)' \int \frac{\partial}{\partial\theta_i} l_i(x_i,\theta_i^1) \times \frac{\partial'}{\partial\theta_i} l_i(x_i,\theta_i^1) L_i(x_i,\theta_i^1)\mu(\,dx)(\theta_i^2-\theta_i^1) + O(n^{-3/2})$$

$$= -\frac{1}{2}(\theta_i^2-\theta_i^1)' I^i(\theta_i^1)(\theta_i^2-\theta_i^1) + O(n^{-3/2}).$$

It follows that $J_{n_i} = 2J_{n_i}^1 + O(n^{-3/2})$. Lemma 1 is proved.

LEMMA 2. *Let the conditions (2.6)-(2.8) hold. Then for all* i

$$J_{n_i} = \frac{2}{n}(\beta_i)^2 + O(n^{-3/2}) \tag{3.5}$$

where

$$\beta_i = \frac{1}{2}\sqrt{n}\big[I^i(\theta_i^1)\big]^{1/2}(\theta_i^2-\theta_i^1), i = 1, ..., k.$$

Proof is immediately clear from the expression (3.4). We are now in a position to introduce our main results. In the next lemma, limit expressions for first three moments of the weighting discriminant function will be obtained via the weighting function, dimension of blocks and constant ρ. Let E denote the expectation with respect to x.

LEMMA 3. *For* $\lambda_n = \lambda_n(\frac{nJ_{n_i}}{2}), i = 1, ..., k$ *the limits*

$$E\big[d_\lambda(x)\big] = \lim_{n\to\infty} \int d_{\lambda_n}(x) L(x,\theta^1)\mu(\,dx) = \rho \int u\lambda(u)\,dH(u), \tag{3.6}$$

$$E\big[d_\lambda(x)\big]^2 = \lim_{n\to\infty} \int \big[d_{\lambda_n}(x)\big]^2 L(x,\theta^1)\mu(\,dx) = 2\rho \int (u+m)\lambda^2(u)\,dH(u), \tag{3.7}$$

exist and

$$E\big[d_{\lambda_n}(x)\big]^3 = O(n^{-3/2}). \tag{3.8}$$

PROOF: Due to assumption 2B the functions $l_i(x_i, \theta_i^2)$ for all $i = 1, ..., k$ can be expanded in the Taylor series around θ_i^1 up to the third order term. By using the proof of Lemma 1 we obtain

$$d_{\lambda_n}(x) = \sum_{i=1}^{k} \lambda_n\big(\frac{nJ_{n_i}}{2}\big)\big[\frac{\partial}{\partial\theta_i} l(x_i,\theta_i^1)(\theta_i^2-\theta_i^1)$$

$$+\frac{1}{2}(\theta_i^2-\theta_i^1)'(\frac{\partial}{\partial\theta_i}\times\frac{\partial'}{\partial\theta_i}l_i(x_i,\theta_i^1))(\theta_i^2-\theta_i^1)+O(n^{-3/2})].$$

Next according to the Lemma 2 and conditions (2.12) we have

$$\lim_{n\to\infty} E[d_{\lambda_n}(x)]$$

$$=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{k}\lambda_n(\frac{nJ_{n_i}}{2})[\frac{nJ_{n_i}}{2}+O(n^{-3/2})]$$

$$=\lim_{n\to\infty}\frac{k}{n}\sum_{i=1}^{k}\lambda_n(\frac{nJ_{n_i}}{2})\frac{nJ_{n_i}}{2}[H_n(u_i)-H_n(u_{i-1})]$$

$$=\rho\int u\lambda(u)\,dH(u),$$

where $u_i, i=1,...,k$ are the points of a division such that $0<u_1<u_2...<u_k<\infty$ and

$$\frac{nJ_{n_i}}{2}\in(u_{i-1},u_i).$$

In order to prove (3.7) we need the Taylor series expansion of $l_i(x_i,\theta_i^2)$ up to the second order term. Then

$$[\ln\frac{L_i(x_i,\theta_i^1)}{L_i(x_i,\theta_i^2)}]^2=[(\theta_i^2-\theta_i^1)'\frac{\partial}{\partial\theta_i}l_i(x_i,\tilde{\theta}_i)]^2, \tag{3.9}$$

where $\tilde{\theta}_i=(\tilde{\theta}_{i_1},...,\tilde{\theta}_{i_m}), \tilde{\theta}_{i_j}\in(\theta_{i_j}{}^1,\theta_{i_j}{}^2)$, for all $j=1,...,m, i=1,...,k$. Note that

$$l_i(x_i,\tilde{\theta}_i)=l_i(x_i,\theta_i^1)+(\tilde{\theta}_i-\theta_i^1)'\frac{\partial}{\partial\theta_i}l_i(x_i,\tilde{\theta_i^1}), \tag{3.10}$$

where $\tilde{\theta_i^1}$ is defined the same way as $\tilde{\theta}_i$. Then by integraiting $[d_i(x_i)]^2$ with respect to $L(x,\theta^1)$ and using the properties of mixed derivatives of $l(x_i,\theta_i)$ from 2. we obtain

$$(\theta_i^2-\theta_i^1)'I^i(\theta_i^1)(\theta_i^2-\theta_i^1)+2(\theta_i^2-\theta_i^1)'(\theta_i^2-\theta_i^1)+O(n^{-2}),$$

since $|\tilde{\theta}_{i_j}-\theta_{i_j}{}^1|=O(n^{-1/2})$ for all $j=1,...,m$. According to the Lemma 2 and conditions (2.12) we hawe

$$\lim_{n\to\infty}E[d_\lambda(x)]^2$$

$$=\lim_{n\to\infty}\frac{2}{n}\sum_{i=1}^{k}\lambda_n^2(\frac{nJ_{n_i}}{2})[\frac{nJ_{n_i}}{2}+m]$$

$$=2\lim_{n\to\infty}\frac{k}{n}\sum_{i=1}^{k}\lambda_n^2(\frac{nJ_{n_i}}{2})[\frac{nJ_{n_i}}{2}+m][H_n(u_i)-H_n(u_{i-1})]$$

$$=2\rho\int(u+m)\lambda^2(u)\,dH(u).$$

Finally for (3.8) we apply the Taylor series expansion of $d_i(x_i)$ up to the first order term and use conditions (2.6). Hence the proof of Lemma 3. is complete.

Having Lemma 2 and the normal convergence of sum of independent random variables we can prove the folloving theorem.

THEOREM 1. *If the classification is performed by the rule $d_\lambda(x) > d_0$ with the weighting functions $\lambda_n = \lambda_n(\frac{nJ_{n_i}}{2}), i = 1, ..., k$ in the above-given formulation of the problem, then the quantities α_1 and α_2 have the limits*

$$\lim_{n\to\infty} \alpha_1 = \Phi(-\frac{E + d_0}{2\sqrt{D}}), \tag{3.11}$$

and

$$\lim_{n\to\infty} \alpha_2 = \Phi(-\frac{E - d_0}{2\sqrt{D}}), \tag{3.12}$$

where

$$\Phi(z) = \frac{1}{2\pi}\int_{-\infty}^{z} \exp(-y^2/2)\, dy.$$

PROOF: Observe that $d_\lambda(x)$ is the random value and define the distribution functions of $d_\lambda(x)$ as $F_\nu(d_\lambda)$ for x from the populations Π_1 and Π_2. Since the density functions corresponding to Π_ν are $L(x, \theta^\nu), \nu = 1, 2$ we have

$$\alpha_1 = \int_{d_{\lambda_n}(x) \leq d_0} L(x, \theta^1)\mu(\, dx) = P(d_\lambda(x) < d_0 | x \in \Pi_1) = F_1(d_0),$$

$$\alpha_2 = \int_{d_{\lambda_n}(x) > d_0} L(x, \theta^2)\mu(\, dx) = 1 - P(d_\lambda(x) < d_0 | x \in \Pi_2) = 1 - F_2(d_0).$$

But according to (2.12) $d_\lambda(x)$ can be considered as a sum of growing number of independent random values with the same distribution. As it is clear from the proof of Lemma 3. the conditions of applicability of the centeral limit theorem are fulfilled. From the asymptotic normality of $d_\lambda(x)$ and Lemma 3 we conclude that the probabilities of misclassification are established by (3.11) and (3.12). It has to be noted that the asymptotic probabilities of misclassification are functionals depending on $\lambda(u)$. We investigate the influence of weighting of the blocks input on the discrimination. Suppose

$$A = \lim_{n\to\infty} \frac{1}{2}(\alpha_1 + \alpha_2).$$

The minimum A is achived for $\alpha_1 = \alpha_2$ with $d_0 = 0$ and hence

$$A = \Phi(-\frac{E}{2\sqrt{D}}). \tag{3.13}$$

We will vary the function $\lambda_n(u)$ in (3.6), (3.7) and (3.13) for constant m, ρ and $H(u)$. Let Λ be the set of all functions satisfying the assumptions (2.12).

THEOREM 2. *In the case $\lambda_n = \lambda_n(\frac{nJ_{n_i}}{2})$ for all i when the function $\lambda(u)$ runs through the elements of Λ, such that E and D are not equal zero, A attains a minimum for $\lambda(u)$ in the form*

$$\lambda_0 = \frac{u}{m + u},$$

and minimal value of A is

$$A_0 = \Phi(-\frac{1}{2}\sqrt{2\rho\int u\lambda_0(u)\,dH(u)}).$$

PROOF: One can see that A is an functional on λ and minimal value of A is achived under the condition that the quantity $E/\sqrt{D}$ is maximal with respect to λ. The necessary condition for the functional to have an extremal value is given by

$$\frac{d}{d\lambda}(\frac{E}{\sqrt{D}}) = 0.$$

We now define the value of λ which is the solution of this equation under the boundary conditions

$$\lambda_0(0) = 0,$$

$$\lambda_0(\infty) = 1. \tag{3.14}$$

By the differentiating $E/\sqrt{D}$ with respect to λ we obtain

$$\frac{\int u\,dH(u)}{\int (u+m)\lambda(u)\,dH(u)} = \frac{\int u\lambda(u)\,dH(u)}{\int (u+m)\lambda^2(u)\,dH(u)}. \tag{3.15}$$

It can be easily verified that the solution of (3.15) which satisfies to the conditions (3.14) is

$$\lambda_0(u) = \frac{u}{u+m}.$$

It follows that

$$\frac{E}{\sqrt{D}} = \frac{\rho\int u\lambda_0(u)\,dH(u)}{\sqrt{2\rho\int (u+m)\lambda_0^2(u)\,dH(u)}} = \sqrt{2\rho\int u\lambda_0(u)\,dH(u)},$$

and

$$A_0 = \min_{\lambda(u)\in\Lambda} A = \Phi(-\frac{1}{2}\sqrt{2\rho\int u\lambda_0(u)\,dH(u)}).$$

Theorem 2 is proved.

ACKNOWLEDGEMENTS

I am grateful to Dr. D. von Rosen for useful discussions. I wish to thank Prof. M. Srivastava for giving me an opportunity to complete this study in the Department of Statistics at the University of Toronto, and for his valuable comments and suggestions.

References

DEEV, A.D. (1970). Rrepresentation of statistics of the discriminant analysis and asymptotical expansions when space dimension is comparable with sample size. *Dokl. Akad. nauk USSR* **195**, 1547-1550.

FRIEDMAN, J.H. (1989). Regularized discriminant analysis. *J. of Amer. Statist. Assoc.* **405**, 165-175.

GIRKO, V.L. (1987). An introduction to the general statistical analysis. *Theory of probability and its applications.* **4**, 741-749.

GIRKO, V.L. (1990). *Theory of Random Determinants.* Kluwer, London.

GIRKO, V.L. & PAVLENKO, T.V. (1989). G-estimator of quadratic discriminant function. *Ukrainskiy matematicheskiy jurnal.* **41**, 1700-1705.

MCLACHLAN, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* Wiley, New York.

MESHALKIN, L.D. (1976). Assignment of numerical values to qualitstive criteria. *Statistical control problems* **14**, 49-55, (in Russian).

PAVLENKO T.V. (1989). Classification of observations in the case when the space dimension is compared with the volume of samples. In *Proc. of the IV All-Union Conf.*, Mathematical methods of images testing, **1** 20-21, Riga.

PAVLENKO, T.V. & VON ROSEN, D. (1996). Estimating the discriminant function when the number of variables is large.*Report 6, Uppsala University, Dep. of Mathematics.*

SERDOBOLSKIY, V.I. (1982). Minimum probability of misclassification in discriminant analysis. *Dokl. Akad. Nauk USSR* **9**, 1066-1070.

SRIVASTAVA, M.S. & KHATRI, C.G. (1979). *An Introduction to Multivariate Statistics.* North Holland, New York.

Uppsala University
Department of Mathematics
Box 480
S-751 06 Uppsala
Sweden
e-mail: pavlenko@snobben.teknikum.uu.se