CHAPTER IX

LINEAR DISCRIMINANT ANALYSIS OF NORMAL POPULATIONS WITH COINCIDING COVARIANCE MATRICES

We consider the problem of the discrimination of a vector $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ from one of two populations $\mathfrak{S}_1$ and $\mathfrak{S}_2$ with common unknown covariance matrix $\Sigma = \text{cov}\,(\mathbf{x}, \mathbf{x})$.

The discrimination rule is $w(\mathbf{x}) \geq \theta$ against $w(\mathbf{x}) < \theta$, where $w(\mathbf{x})$ is a linear discriminant function, and θ is a threshold. Probabilities of errors of the discriminant analysis ('classification errors') are

$$\alpha_1 = \mathbf{P}(w(\mathbf{x}) < c|\mathfrak{S}_1) \quad \text{and} \quad \alpha_2 = \mathbf{P}(w(\mathbf{x}) \geq c|\mathfrak{S}_2). \tag{1}$$

for observations $\mathbf{x}$ from populations $\mathfrak{S}_1$ and $\mathfrak{S}_2$. If the populations are normal $\mathbf{N}(\vec{\mu}_\nu, \Sigma)$, $\nu = 1, 2$, with non-degenerate matrix Σ, then it is well known that, by the Neumann–Pearson lemma, the minimum of $(\alpha_1 + \alpha_2)/2$ is attained with $\theta = 0$ and $w(\mathbf{x})$ of the form

$$w(\mathbf{x}) = w^{\text{opt}}(\mathbf{x}) = \ln \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = (\vec{\mu}_1 - \vec{\mu}_2)^T \Sigma^{-1} (\mathbf{x} - (\vec{\mu}_1 + \vec{\mu}_2)/2),$$

where

$$f_\nu(\mathbf{x}) = (2\pi \ \det\, \Sigma)^{-1/2} \exp\, (-(\mathbf{x} - \vec{\mu}_\nu)^T \Sigma^{-1} (\mathbf{x} - \vec{\mu}_\nu)/2),$$

are normal distribution densities, $\nu = 1, 2$. The minimum of the half-sum $(\alpha_1 + \alpha_2)/2$ thus attained is $\Phi(-\sqrt{J}/2)$, where the quantity $J = (\vec{\mu}_1 - \vec{\mu}_2)^T \Sigma^{-1} (\vec{\mu}_1 - \vec{\mu}_2)$ is 'the square of the Mahalanobis distance'. Samples $\mathfrak{X}_1 = \{\mathbf{x}_m\}$, $m = 1, \dots, N_1$ and $\mathfrak{X}_2 = \{\mathbf{x}_m\}$, $m = N_1 + 1, \dots, N$, from $\mathfrak{S}_1$ and $\mathfrak{S}_2$ $(N = N_1 + N_2)$ are used to calculate sample means

$$\bar{\mathbf{x}}_1 = N_1^{-1} \sum_{m=1}^{N_1} \mathbf{x}_m, \qquad \bar{\mathbf{x}}_2 = N_2^{-1} \sum_{m=N_1+1}^{N} \mathbf{x}_m,$$

Typeset by $\mathcal{AMS}$-TeX

and 'pooled' sample covariance matrix

$$C = (N-2)^{-1}[\sum_{m=1}^{N_1}(\mathbf{x}_m - \bar{\mathbf{x}}_1)(\mathbf{x}_m - \bar{\mathbf{x}}_1)^T + \sum_{m=N_1+1}^{N}(\mathbf{x}_m - \bar{\mathbf{x}}_2)(\mathbf{x}_m - \bar{\mathbf{x}}_2)^T]. \quad (2)$$

Traditionally, the standard discrimination procedure is used with the consistent Wald discriminant function

$$w(\mathbf{x}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T C^{-1}(\mathbf{x} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2. \quad (3)$$

However, this procedure proves to be unstable and obviously not the best one even for low-dimensional problems (see Introduction). In this chapter, we consider a class of generalized stable discriminant functions using 'generalized ridge estimators' of the inverse covariance matrices and develop a limit theory with the purpose to search for improved and unimprovable discrimination procedures. This development is a continuation of researches under the Kolmogorov asymptotics described in Introduction.

Problem Setting

We consider a sequence of discrimination problems

$$\mathfrak{P} = \{(\mathfrak{S}_1, \mathfrak{S}_2, \vec{\mu}_1, \vec{\mu}_2, \Sigma, N_1, N_2, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, C, w(\mathbf{x}), \alpha_1, \alpha_2)_n\}, \quad (4)$$

$n = 1, 2, \ldots$ (we do not write out the subscripts n for the arguments of $\mathfrak{P}$). The observation vectors $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are taken from one of two populations $\mathfrak{S}_1$ and $\mathfrak{S}_2$; the population means are $\vec{\mu}_1 = \mathbf{E}_1\mathbf{x}$ and $\vec{\mu}_2 = \mathbf{E}_2\mathbf{x}_2$, where (and in the following) $\mathbf{E}_1$ and $\mathbf{E}_2$ are expectation operators for $\mathbf{x}$ in $\mathfrak{S}_1$ and $\mathfrak{S}_2$, respectively. It is assumed that both population have a common covariance matrix

$$\Sigma = \mathbf{E}_1\ (\mathbf{x} - \vec{\mu}_1)(\mathbf{x} - \vec{\mu}_1)^T = \mathbf{E}_2\ (\mathbf{x} - \vec{\mu}_2)(\mathbf{x} - \vec{\mu}_2)^T.$$

The discriminant function $w(\mathbf{x})$ is a function of sample means and sample covariance matrix. The discrimination error probabilities

$$\alpha_1 = \mathbf{P}(w(\mathbf{x}) < c|\mathbf{x} \text{ in } \mathfrak{S}_1), \quad \alpha_2 = \mathbf{P}(w(\mathbf{x}) \geq c|\mathbf{x} \text{ in } \mathfrak{S}_2)$$

are functions of samples.

We assume that populations are normal. The results can be generalized to a wide class of populations by using the 'Normal Evaluation Principle' offered in Chapter 4.

We restrict $\mathfrak{P}$ with following conditions.

A. The populations $\mathfrak{S}_\nu$ are normal $\mathbf{N}(\vec{\mu}_\nu, \Sigma)$, $\nu = 1, 2$.

B. All eigenvalues of Σ are located on the segment $[c_1, c_2]$, where $c_1 > 0$ and c_2 do not depend on n.

C. In $\mathfrak{P}$ the limits exist $y_\nu = \lim\limits_{n\to\infty} n/N_\nu \geq 0$, $\nu = 1, 2$, such that $y = \lim\limits_{n\to\infty} n/(N_1 + N_2) = y_1 y_2/(y_1 + y_2)$, where $y < 1$.

We introduce the empiric distribution functions

$$F_{on}(u) = n^{-1} \sum_{i=1}^{n} \text{ind } (\lambda_i \leq u),$$

of eigenvalues λ_i of Σ, $i = 1, \ldots, n$, and

$$B_n(u) = \sum_{i=1}^{n} \mu_i^2/\lambda_i \text{ ind } (\lambda_i \leq u),$$

where μ_i are components of $\vec{\mu} = \vec{\mu}_1 - \vec{\mu}_2$ in a system of coordinates, where Σ is diagonal.

D. For $u > 0$, $F_{0n}(u) \to F_0(u)$ and $B_n(u) \to B(u)$ almost everywhere.

Under conditions A–D, the limit exists

$$J = B(c_2) = \lim_{n\to\infty} (\vec{\mu}_1 - \vec{\mu}_2)^T \Sigma^{-1} (\vec{\mu}_1 - \vec{\mu}_2)$$

and $(\vec{\mu}_1 - \vec{\mu}_2)^2 \leq c_2 B(c_2)$.

E. The discriminant function is of the form

$$w(\mathbf{x}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \Gamma(C) \big(\mathbf{x} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2\big), \tag{5}$$

where the matrix $\Gamma(C)$ depends on C so that $\Gamma(C)$ is diagonalized together with C and has eigenvalues $\Gamma(\lambda)$ for the eigenvalues λ of C; it is defined by the scalar function $\Gamma : \mathbb{R}^1 \to \mathbb{R}^1$

$$\Gamma = \Gamma(u) = \int_{t\geq 0} (1 + ut)^{-1} d\eta(t),$$

where $\eta(t)$ is a function of finite variation on $[0, \infty)$ not depending on n.

To be more concise, we denote
$\vec{\mu} = \vec{\mu}_1 - \vec{\mu}_2$, $\bar{\mathbf{x}} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$, and $N_0 = N_1 N_2/(N_1 + N_2)$.

The (random) probabilities of errors (1) are

$$\alpha_1 = \Phi\left(-\frac{\mathbf{E}_1 w(\mathbf{x}) - \theta}{\sqrt{\text{var } w(\mathbf{x})}}\right), \quad \alpha_2 = \Phi\left(-\frac{\theta - \mathbf{E}_2 w(\mathbf{x})}{\sqrt{\text{var } w(\mathbf{x})}}\right),$$

where the conditional expectation operators $\mathbf{E}_1$ and $\mathbf{E}_2$ and conditional variance var $(w(\mathbf{x}))$ (identical in the both populations) are calculated for chosen samples. The half-sum $(\alpha_1 + \alpha_2)/2$ is a minimum for

$$\theta = \theta_n^{\text{opt}} = 1/2\ (\mathbf{E}_1 w(\mathbf{x}) - \mathbf{E}_2 w_2(\mathbf{x}))$$

and is equal to

$$\alpha = \frac{\alpha_1 + \alpha_2}{2}\bigg|_{\theta = \theta_n^{\text{opt}}} = \Phi\left(-\frac{\vec{\mu}^T \Gamma \bar{\mathbf{x}}}{2\sqrt{\bar{\mathbf{x}}^T \Gamma \Sigma \Gamma \bar{\mathbf{x}}}}\right), \tag{6}$$

where the matrix $\Gamma = \Gamma(C)$.

Expectation and Variance of Generalized Discriminant Functions

We study the resolvent $H = H(z) = (I - zC)^{-1}$ of the matrices C and matrices $\Gamma = \Gamma(C) = \int (I + tC)^{-1} d\eta(t)$. First, we need to estimate the variance of expressions quadratic with respect to H.

Remark 1. (a corollary of Theorem 2.2). Under assumptions A–D as $n \to \infty$, for z outside any ε-neighbourhood of the half-axis $z > 0$ uniformly,

$$\mathbf{E}\ n^{-1}\text{tr } \Sigma H(z) = \begin{cases} \dfrac{h(z) - 1}{z s(z)} & \text{if } z \neq 0, \\ \Lambda_1 = \lim\limits_{n \to \infty} n^{-1}\text{tr } \Sigma & \text{if } z = 0. \end{cases}$$

Remark 2. Under assumptions A–D for z outside of the half-axis $z > 0$, as $n \to \infty$, we have var $(n^{-1}\text{tr } H(z)) \to 0$,
and var $(n^{-1}\text{tr } \Sigma H(z)) \to 0$.

These relations are a corollary of Lemma 2.3.

Denote $N_0 = N_1 N_2/(N_1 + N_2)$.

LEMMA 9.1. *Under assumptions A–D for any $z, z' \le 0$ and for any non-random vector* $\mathbf{e}$ *of unit length, we have in* $\mathfrak{P}$

$$\mathrm{var}\ (\mathbf{e}^T H(z)\Sigma H(z')\mathbf{e}) = O(N^{-1}),$$
$$\mathrm{var}\ (\bar{\mathbf{x}}^T H(z)\Sigma H(z')\bar{\mathbf{x}}) = O(N_0^{-1}).$$

Proof. Denote $H = H(z)$, $H' = H(z')$. To estimate the variances we use the method of the alternating elimination of sample vectors. Since the distributions are normal the matrix C can be transformed to the form of the Gram matrix $S = N^{-1}\sum \mathbf{x}_m\mathbf{x}_m^T$ with $m = 1, \ldots, N-1$. Let us eliminate the vector $\mathbf{x}_1$. Excluding $\mathbf{x}_1$, we denote $S^1 = S - N^{-1}\mathbf{x}_1\mathbf{x}_1^T$, $H^1 = (I - zS^1)^{-1}$. The identity holds $H = H^1 + zN^{-1}H^1\mathbf{x}_1\mathbf{x}_1^T H$. Denote $\Omega = H(z)\Sigma H(z')$, $\Omega^1 = H^1(z)\Sigma H^1(z')$. We have

$$\begin{aligned} f \stackrel{\mathrm{def}}{=} & \mathbf{e}^T H(z)\Sigma H(z')\mathbf{e} \\ = & \mathbf{e}^T\Omega^1\mathbf{e} + zN^{-1}\mathbf{e}^T H^1(z)\mathbf{x}_1\mathbf{x}_1^T\Omega\mathbf{e} \\ & + z'N^{-1}\mathbf{e}^T\Omega\mathbf{x}_1\mathbf{x}_1^T H^1(z')\mathbf{e} + zz'N^{-2}q\mathbf{e}^T H^1(z)\mathbf{x}_1\ \mathbf{x}_1^T H(z')^1\mathbf{e}, \end{aligned}$$

where $q = \mathbf{x}_1^T\Omega\mathbf{x}_1$. Here the first summand does not depend on $\mathbf{x}_1$. By Lemma 2.2, var f is not greater that $N\mathbf{E}\ (|\xi_2|^2 + |\xi_3|^2 + |\xi_4|^2)$, where ξ_2, ξ_3, ξ_4 are the remaining three summands of f. It suffices to prove that the expectation of the square of each of these is $O(N^{-2})$. As in Chapter 2 denote $v_1(z) = \mathbf{e}^T H^1(z)\mathbf{x}_1$. Using the Schwarz inequality we find that

$$\mathbf{E}\ \xi_1^2 \le |z|^2 N^{-2}\mathbf{E}\ v_1^2\mathbf{E}\ \mathbf{e}^T\Omega^1\Sigma\Omega^1\mathbf{e} \le N^{-2}|z|^2\|\Sigma\|^2 = O(N^{-2}).$$

The third summand can be estimated likewise. Now we notice that $q/N \le \|\Sigma\|\sqrt{|\psi_1(z)|}\sqrt{|\psi_1(z')|}$ where $\psi_1(z) = \mathbf{x}_1^T H(z)\mathbf{x}_1/N$. But $|z\psi(z)| \le \alpha$ (see Chapter 2) and it follows that

$$\mathbf{E}\ |\xi_4|^2 \le |zz'|^2 N^{-2}\mathbf{E}\ |v_1(z)v_1(z')|^2 = O(N^{-2}).$$

We conclude that the first statement of our lemma holds.

To prove the second statement we notice that for normal distributions sample means do not depend on the matrix C. Therefore

$$\mathrm{var}\ \left(\bar{\mathbf{x}}^T\Omega\bar{\mathbf{x}}\right) = \mathrm{var}\ \left(\bar{\mathbf{x}}^T A\bar{\mathbf{x}}\right) + \mathbf{E}\ |\bar{\mathbf{x}}|^4\mathrm{var}\ (\mathbf{e}^T\Omega\mathbf{e}),$$

where $A = \mathbf{E}\,\Omega$, the vectors $\mathbf{e} = \bar{\mathbf{x}}/|\bar{\mathbf{x}}|$, and the last variance is conditional under fixed $\bar{\mathbf{x}}$. Here, in the first summand, var $(\bar{\mathbf{x}}^T\Omega\bar{\mathbf{x}}) = 2N_0^{-2}$ tr $A\Sigma A\Sigma$. It follows that the first variance is $O(N_0^{-1})$. In the second variance we use the first statement of our lemma. We find that the second summand is of the same order of magnitude. The lemma is proved. □

LEMMA 9.2. *Suppose conditions $A--E$ are satisfied. Then*

$$\vec{\mu}^T\Gamma\bar{\mathbf{x}} = \vec{\mu}^T\Gamma\vec{\mu} + \xi_1,$$
$$\bar{\mathbf{x}}^T\Gamma\Sigma\Gamma\bar{\mathbf{x}} = \vec{\mu}^T\Gamma\Sigma\Gamma\vec{\mu} + N_0^{-1}\mathrm{tr}\,(\Sigma\Gamma\Sigma\Gamma) + \xi_2,$$
$$\theta_n^{\mathrm{opt}} = 1/2\ (N_2^{-1} - N_1^{-1})\ \mathrm{tr}\ (\Sigma\Gamma) + \zeta\,,$$

where $\Gamma = \Gamma(C)$, *and random* ξ_ν *and* ζ *are such that* $\mathbf{E}\ \xi_\nu^2 = O(N_0^{-1})$, $\nu = 1, 2$, *and* $\mathbf{E}\ \zeta^2 = O(N_0^{-1})$ *as* $n \to \infty$.

Proof. The random value $\vec{\mu}^T\Gamma\bar{\mathbf{x}} = \vec{\mu}^T\Gamma\vec{\mu} + \xi_1$, where $\xi_1 = \mathbf{v}^T\overset{o}{\bar{\mathbf{x}}}$, $\overset{o}{\bar{\mathbf{x}}} = \bar{\mathbf{x}} - \vec{\mu}$ and $\mathbf{v} = \Gamma\vec{\mu}$. For normal distributions, the random values $\bar{\mathbf{x}}$ and $\mathbf{v}$ are independent. We have

$$\mathbf{E}\ \xi_1^2 \le 2V(\eta)\mathbf{E}\ \mathbf{v}^T\Sigma\mathbf{v}/N_0 \le \mathrm{const}\ \vec{\mu}^2||\Sigma||/N_0\,,$$

where $V(\eta)$ is the variation of the function $\eta(\cdot)$. The spectral norm $\|\Sigma\| \le c_2$. Thus $\mathbf{E}\ \xi_1^2 = O(N_0^{-1})$.

Now, we have

$$\bar{\mathbf{x}}^T\Gamma\Sigma\Gamma\bar{\mathbf{x}} = \vec{\mu}^T\Gamma\Sigma\Gamma\vec{\mu} + 2\vec{\mu}^T\Gamma\Sigma\Gamma\overset{o}{\bar{\mathbf{x}}} + \overset{o}{\bar{\mathbf{x}}}\Gamma\Sigma\Gamma\overset{o}{\bar{\mathbf{x}}}\,.$$

Here the second summand equals $\xi_3 = 2\mathbf{v}^T\overset{o}{\bar{\mathbf{x}}}$, where $\mathbf{v} = 2\Gamma\Sigma\Gamma\vec{\mu}$. The vector $\bar{\mathbf{x}}$ does not depend on $\mathbf{v}$ and is distributed normally as $\mathbf{N}(0, N_0^{-1}\Sigma)$. We obtain $\mathbf{E}\ \xi_3^2 = O(N_0^{-1})$. For the third summand, $\mathbf{E}\ (\overset{o}{\bar{\mathbf{x}}}^T\Gamma\Sigma\Gamma\overset{o}{\bar{\mathbf{x}}}) = N_0^{-1}\mathrm{tr}\ (\Sigma\Gamma\Sigma\Gamma)$. The variance of the third summand vanishes as $n \to \infty$ by Lemma 9.1. We obtain statement 2 of our lemma.

Further, consider the random value

$$\begin{aligned}\theta_n^{\mathrm{opt}} &= 1/2\ \bar{\mathbf{x}}^T\Gamma(\overset{o}{\bar{\mathbf{x}}}_1 + \overset{o}{\bar{\mathbf{x}}}_2)\\ &= 1/2\ \vec{\mu}^T\Gamma(\overset{o}{\bar{\mathbf{x}}}_1 + \overset{o}{\bar{\mathbf{x}}}_2) + 1/2\ \overset{o}{\bar{\mathbf{x}}}{}_2^T\Gamma\bar{\mathbf{x}}_2 - 1/2\ \overset{o}{\bar{\mathbf{x}}}{}_1^T\Gamma\bar{\mathbf{x}}_1\,.\end{aligned}$$

For fixed C the conditional expectation

$$\mathbf{E}\ (\theta_n^{\text{opt}}|C) = 1/2\ (N_2^{-1} - N_1^{-1})\ \text{tr}\ (\Sigma\Gamma).$$

It suffices to show that the variance of each of the last three summands in the expression for θ_n^{opt} is $O(N_0^{-1})$. For normal distributions,

$$\mathbf{E}\ (\vec{\mu}^T\Gamma\overset{o}{\mathbf{x}}_\nu)^2 = \mathbf{E}\ \vec{\mu}^T\Gamma\Sigma\Gamma\vec{\mu}/N_\nu = O(N_\nu^{-1}), \quad \nu = 1, 2.$$

The variance of last two terms can be estimated quite analogously to the second statement of Lemma 9.1. The proof of Lemma 9.2 is complete. □

Now let us study the resolvent of the matrix C. From the additivity property of the Wishart distribution it follows that C also is a Wishart matrix, and its probability density is

$$f_W = \text{const}\ |\Sigma|^{(M-1)/2}|C|^{(M-n-2)/2}\exp\ \left(-\frac{M}{2}\ \text{tr}\ \Sigma^{-1}C\right), \quad (7)$$

where $M = N_1 + N_2 - 1$. Define the square

$$\mathfrak{K}_T = \{z, z' : -T \le z \le 0,\ -T \le z' \le 0\}.$$

LEMMA 9.3. *Under assumptions A–E for any $T > 0$ as $n \to \infty$,*

$$\mathbf{E}\ H(z)\Sigma H(z') = \begin{cases} \dfrac{\mathbf{E}\ H(z) - \mathbf{E}\ H(z')}{s(z)s(z')(z-z')} + K_n^{(1)} & \text{if } z \ne 0, \\ \dfrac{1}{s^2(z)}\mathbf{E}\ CH^2(z) + K_n^{(2)} & \text{if } z = 0. \end{cases} \quad (8)$$

where the spectral norms of matrices $K_n^{(1)}$ and $K_2^{(2)}$ are $O(M^{-1})$ uniformly in $(z, z') \in \mathfrak{K}_T$.

Proof. We use the method of the reduction of volume integrals with respect to the Wishart distribution to the surface integrals. Denote $\nabla_{kl} = \varepsilon_{kl}\partial/\partial C_{kl}$, where $\varepsilon_{kk} = 1$, and $\varepsilon_{kl} = 1/2$ if $k \ne l$, $k, l = 1, \dots, n$, and C_{kl} is an entry of the matrix C. Let the matrix operator ∇ denote the symmetric matrix of derivatives $\{\nabla_{ij}\}$

described in Introduction. To be more concise, we denote $H = H(z)$, $H' = H(z')$ and let subscripts for H and H' define their matrix elements. We note that the volume integral

$$\int_{\det\, C>0} \sum \nabla_{kl}(H'_{ij}\Sigma_{jk}H_{lm}f_W)dC = 0$$

for each i and m, since it can be reduced to the integral over the hypersurface $\det C = 0$, where the integrand vanishes if $N > n+2$ (here the summation is carried out over all repeated indexes). Differentiating by the rules described in Introduction and using the relation $C^{-1}H = C^{-1} + zH$, we transform this expression to

$$\int_{\det\, C>0} [z'M^{-1}H\Sigma H'H + z'H'HM^{-1}\mathrm{tr}\,(\Sigma H') \\ + zH'\Sigma HM^{-1}\mathrm{tr}\,H + zM^{-1}H'\Sigma H^2 + LM^{-1}H'\Sigma C^{-1} \\ + LM^{-1}zH'\Sigma H - (1 - M^{-1})H'H]\; f_W dC = 0,$$

where $L = M - n - 2$. The first and the fourth summands present matrices whose norms are $O(N^{-1})$ as $n \to \infty$ uniformly in $\mathfrak{K}_T$. In view of Remarks 1 and 2 we have

$$zM^{-1}\mathrm{tr}\,(\Sigma H) \xrightarrow{2} y(h(z) - 1)/s(z), \qquad n^{-1}\mathrm{tr}\,H \xrightarrow{2} h(z).$$

Singling out the leading term we obtain

$$\int_{\det\, C>0} [-H'H/s(z') + LM^{-1}H'\Sigma C^{-1} + zs(z)H'\Sigma H] f_W dC = \Xi_n\,,$$

where the norm of the matrix Ξ_n is $O(N^{-1})$. Setting $z = 0$ we have

$$\int_{\det\, C>0} [-H'/s(z') + LM^{-1}H'\Sigma C^{-1}] f_W dC = \Xi_n.$$

Thus we can write

$$z\mathbf{E}\; H'\Sigma H = \frac{\mathbf{E}\; H'(H - I)}{s(z)s(z')} + \Xi_n.$$

Substituting $HH' = (zH - z'H')/(z - z')$ (if $z \neq z'$), we obtain (8). Lemma 9.3 is proved. □

From Remark 1, we obtain

$$b_n(z) = \vec{\mu}^T H(z)\vec{\mu} \xrightarrow{2} b(z) \stackrel{\text{def}}{=} \int (1 - zs(z)u)^{-1} u dB(u).$$

Define

$$k(z) = \begin{cases} b(z) + (y_1 + y_2)(h(z) - 1)/(zs(z)) & \text{if} \quad z \neq 0, \\ b(0) + (y_1 + y_2)(h(z) - 1)\Lambda_1 & \text{if} \quad z = 0. \end{cases} \tag{9}$$

LEMMA 9.4. *Under assumptions A–E for any $T > 0$ as $n \to \infty$,*

$$\mathbf{E}\ \bar{\mathbf{x}}^T H(z)\Sigma H(z')\bar{\mathbf{x}} = \frac{k(z) - k(z')}{s(z)s(z')(z - z')} + o_n\,, \tag{10}$$

where the expression in the right hand side is extended by continuity to $z = z'$ and $o_n = O(N^{-1})$ uniformly in $\mathfrak{K}_T$.

Proof. Denote $\delta\bar{\mathbf{x}} = \bar{\mathbf{x}} - \vec{\mu}$. The random value $\delta\bar{\mathbf{x}}$ does not depend on C, $\mathbf{E}\ \delta\bar{\mathbf{x}} = 0$, and $\mathbf{E}\ \delta\bar{\mathbf{x}}\ \delta\bar{\mathbf{x}}^T = N_0^{-1}\Sigma$. We find that

$$\mathbf{E}\ \bar{\mathbf{x}}^T \Omega \mathbf{x} = \mathbf{E}\ \vec{\mu}^T \Omega \vec{\mu} + \mathbf{E}\ N_0^{-1} \mathrm{tr}\ (\Sigma\Omega),$$

where $\Omega = H(z)\Sigma H(z')$. Let us replace $\mathbf{E}\ \Omega$ using Lemma 9.3. The right hand side of (10) equals

$$\frac{1}{s(z)s(z')} \left(\frac{b_n(z) - b_n(z')}{z - z'} + N_0^{-1} \mathrm{tr}\ \frac{\Sigma H(z) - \Sigma H(z')}{z - z'} \right) + o_n, \tag{11}$$

where $b_n(z) = \vec{\mu}^T H(z)\vec{\mu}$, the expressions are extended by continuity to $z = z'$, and the estimate $o_n = O(N_0^{-1})$ is uniform in $\mathfrak{K}_T$. In view of Remarks 1 and 2 the limit transition as $n \to \infty$ gives the right hand side of (10). The proof of Lemma 9.4 is complete. □

Remark 3. If z and z' are outside of any ε-neighbourhood of the half-axis $z > 0$ and assumptions A–E hold, then as $n \to \infty$ uniformly,

$$\vec{\mu}^T H(z)\Sigma H(z')\vec{\mu} \xrightarrow{2} \frac{1}{s(z)s(z')} \frac{b(z) - b(z')}{z - z'},$$

$$\bar{\mathbf{x}}^T H(z)\Sigma H(z')\bar{\mathbf{x}} \xrightarrow{2} \frac{1}{s(z)s(z')} \frac{k(z) - k(z')}{z - z'}, \tag{12}$$

where the right hand sides are extended by continuity to $z = z'$.

Note that

$$\mathbf{E}_1\ (w(\mathbf{x}) - \theta_n^{\text{opt}}) = \mathbf{E}_2\ (w(\mathbf{x}) + \theta_n^{\text{opt}}) = \vec{\mu}^T \Gamma(C)\vec{\mu}\,.$$

THEOREM 9.1. *Under assumptions A–E, there exist the limits*

$$\operatorname*{plim}_{n\to\infty} \theta_n^{\rm opt} = \theta^{\rm opt},$$
$$\operatorname*{plim}_{n\to\infty} \vec{\mu}^T\Gamma(C)\bar{\mathbf{x}} = 2G\,, \quad \operatorname*{plim}_{n\to\infty} \bar{\mathbf{x}}^T\Gamma(C)\Sigma\Gamma(C)\bar{\mathbf{x}} = D,$$

where

$$\theta^{\rm opt} = 1/2\ (y_2 - y_1) \int \frac{1-h(-t)}{ts(-t)} d\eta(t)\,,$$
$$G = 1/2 \int b(-t) d\eta(t)\,,$$
$$D = \iint \frac{k(-t)-k(-t')}{s(-t)s(-t')(t'-t)} d\eta(t) d\eta(t')\,, \tag{13}$$

and the last integrand is extended by continuity to $t = t'$ and to $t = 0$.

Proof. The first statement immediately follows from (12). To obtain the second statement, it suffices to prove the convergence

$$\vec{\mu}^T\Gamma(C)\vec{\mu} = \int \vec{\mu}^T(I+tC)^{-1}\vec{\mu}\ d\eta(t) = \int b_n(-t)d\eta(t)\,.$$

Here $b_n(-t) \xrightarrow{2} b(-t)$ uniformly in $t \le T$. Since $\eta(t)$ is of bounded variation, the contribution of large t can be made arbitrarily small. Hence $\vec{\mu}^T\Gamma(C)\vec{\mu} \xrightarrow{2} \int b(-t)d\eta(t)$. The second statement of our theorem follows. Our theorem is proved. □

Example 1. Let us choose a special form of $\eta(t)$: $\eta(t') = 0$ for $t' \le t$, and $\eta(t') = t$ for $t' \ge t$. Then $\Gamma(C) = t(I+tC)^{-1}$ is a ridge-estimator of the matrix Σ^{-1}. In this case

$$\theta^{\rm opt} = 1/2\ (y_2 - y_1)\ (1 - h(-t))/s(-t),$$

$$G = 1/2\ t\ b(-t),$$

$$D = -t^2[s(-t)^{-2}\frac{dk(-t)}{dt}].$$

Limit Probabilities of the Discrimination Errors

THEOREM 9.2. *Suppose assumptions A–E hold, $D > 0$, the discrimination function* (5) *is used, and the threshold θ is chosen for each n so that it minimizes $(\alpha_1 + \alpha_2)/2$. Then*

$$\operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\sqrt{J^{\text{eff}}}/2),$$

where $J^{\text{eff}} = 4G^2/D$ with G and D defined by (13).

Proof. The minimum of $(\alpha_1+\alpha_2)/2$ is attained for $\theta = \theta_n^{\text{opt}}$, where θ_n^{opt} is defined by (6). The statement of Theorem 9.2 immediately follows from Theorem 9.1. □

Example 2. Let $\Sigma = \sigma^2 I$, $\sigma > 0$ for all $n = 1, 2, \ldots,$ and let $\eta(t) = 1$ for all $t > 0$. Then $\Gamma(C) = I$, $G = J/2$, $D = J + y_1 + y_2$, where $J = B(c_2)$ and $\theta^{\text{opt}} = (y_2 - y_1)/2$ in the agreement with limit formulas for block-dependent components of $\mathbf{x}$ (see Introduction).

Example 3. Let $\eta(t') = \text{ind}\ (t' \le t)$, where $t \to \infty$. This corresponds to a transition to the standard discriminant function. Let us tend t to infinity in the limit expressions (13). As $t \to \infty$ we find that $h(-t) \to 0$, $s(-t) \to 1 - y$, and $tb(-t) \to J/(1 - y)$, where $J = \lim_{n\to\infty} \vec{\mu}^T \Sigma^{-1} \vec{\mu}$. From (13) we obtain

$$G \to J/2\ (1 - y)^{-1}, \quad D \to (J + y_1 + y_2)(1 - y)^{-3},$$
$$\text{and} \quad J^{\text{eff}} \to J^2(1 - y)/(J + y_1 + y_2),$$

in the agreement with Deev's formula (see Introduction).

Example 4. Let $\eta(t') = \text{ind}\ (t' \le t)$, and consider a special case when $\mu_i^2/\lambda_i = J/n$ for all components μ_i of the vector $\vec{\mu}$ in a system of coordinates where Σ is diagonal, where λ_i are the corresponding eigenvalues of Σ, $i = 1, \ldots, n$ (the case of 'equal contributions to the Mahalanobis distance'). Suppose that the limit spectrum of sample covariance matrices is given by the 'ρ-model' that was described in Chapter 2. Then the dependence of J^{eff} on t can be found in an explicit form and is as follows:

$$J^{\text{eff}} = J^{\text{eff}}(t) = J^2(J - y + 2yh - (\rho + y)h^2)/(J + y_1 + y_2),$$

where $h = h(-t)$ is

$$h = 2/[1 + \rho + \kappa(1 - y)t + \sqrt{(1 + \rho + \kappa(1 - y)t)^2 + 4(\kappa yt - \rho)},$$

where $\kappa = \sigma^2(1 - \rho)^2$ and σ and $\rho < 1$ are parameters of the model. We note that $h(-t)$ is a monotone function of t, and the maximum of J^{eff} is attained for $h = h^{\mathrm{opt}} = y(\rho + y)^{-1},\ t = t^{\mathrm{opt}} = \rho\kappa^{-1}y^{-1}$ if $\rho > 0,\ \sigma > 0$, and $y > 0$. The maximum value is

$$J^{\mathrm{eff}}(t^{\mathrm{opt}}) = J^2\,[1 - \rho y/(\rho + y)]\,/(J + y_1 + y_2). \tag{14}$$

The limit error probability $\Phi(-\sqrt{J^{\mathrm{eff}}(t^{\mathrm{opt}})}/2)$ provides an extension of Deev's formula for a regularized and optimized discriminant function. The ratio $(\rho + y(1 - \rho))\,/\,\big(\rho + y(1 - \rho) - y^2\big)$ characterizes the gain in the limit provided by the improved discrimination procedure. In the case when $\rho = 0$, the value $J^{\mathrm{eff}}(t^{\mathrm{opt}}) = J^2/(J + y_1 + y_2)$; that corresponds to the discrimination in the system of coordinates where $\Sigma = I$ (for independent variables, although this fact is unknown to the observer). In the case, when $y = 1$ we have $J^{\mathrm{eff}} = J^2/[(1 + \rho)(J + y_1 + y_2)]$ in contrast with the standard discrimination procedure for which $J^{\mathrm{eff}} = 0$.

We notice that the both functions G^2 and D are quadratic in $\eta(t)$ and an obvious maximization of G^2/D is possible .

Theorem 9.3. *Suppose that a function $\eta(t) = \eta^{\mathrm{opt}}(t)$ of finite variation exists such that*

$$\int \frac{k(-t) - k(-t')}{s(-t)s(-t')(t' - t)}\ d\eta^{\mathrm{opt}}(t') = b(-t), \quad t \geq 0 \tag{15}$$

Then for any function of finite variation $\eta(t)$ on $[0, \infty)$, we have $J^{\mathrm{eff}}(\eta) \leq J^{\mathrm{eff}}(\eta^{\mathrm{opt}})$.

This assertion immediately follows from Theorems 9.1 and 9.2.

Note that the solution of (15) may not exist. If it exists, serious difficulties are still to be expected when we try to replace $k(-t)$ and $s(-t)$ by their natural estimators. Relation (15) is the Fredholm integral equation of the first kind and its solution may be ill-conditioned. For applications some more detailed theoretical investigations are necessary.

However, the recent researches show that these difficulties can be overcome. In [41] the extremum problem was reformulated and partly solved. Under some additional assumptions an unimprovable in the limit solution was found explicitly. In [63] and [64] the asymptotic expressions for this solution were used to construct a practical improved discrimination procedure ('Elda', see [64]). This procedure was examined numerically by academic tests and successfully applied to real problems.