1

REPRESENTATION FORMULAE FOR PROBABILITIES OF CORRECT CLASSIFICATION

Wolf-Dieter Richter
University of Rostock, Germany

ABSTRACT Representation formulae for probabilities of correct classification are derived for a minimum-distance type decision rule. These formulae are essentially based upon the two-dimensional Gaussian distribution and allow new derivations of recent results in Krause and Richter(1999).

Keywords and phrases Generalized minimum-distance rule, repeated measurements, linear model approach, two-dimensional decision space, two-dimensional representation formulae, doubly noncentral F-distribution.

1.1 INTRODUCTION

Let an individual having a Gaussian distributed feature variable belong to one of two distinct populations. Assume that we are given measurements from a sample of individuals giving rise to independent and identically distributed Gaussian feature variables. Several methods of allocating the individual or the whole sample to one of the two populations have been studied in the literature. For an introduction to this area see, e.g., Anderson (1984) or McLachlan (1992). A certain subclass of classification rules is given by the so-called distance rules. Because of the great variety of distances existing in statistics, there are several approaches to distance based classification rules, as only to mention Cacoullos and Koutras (1985), Marco, Young and Turner (1987), Cacoullos (1992) and Cacoullos and Koutras (1996). When choosing a method for classifying an individual or a sample of individuals, one has to distinguish between the cases of known or unknown moments. Certain sample-distance based classification rules, however, work without assumptions concerning the second order moments and probabilities of correct classification can be described explicitly in terms of these moments. This advantage has been exploited to some extent recently in Krause and Richter(1999). The method developed there combines a geometric sample measure representation formula for the multivariate Gaussian measure with a certain non classic linear model approach due to Krause and Richter (1994). This linear model type approach will be modified in the present paper to derive new representation formulae for probabilities of correct classification which are based upon the two-dimensional Gaussian law. Further transformation of these formulae yields expressions in terms of the doubly noncentral F-distribution as has been derived recently in another way in Krause and Richter(1999).

Let n_1, n_2 and n_3 observations belong to the three populations Π_1, Π_2, and Π_3, respectively. We suppose that Π_1 and Π_2 are distinguishable with respect to their expectations and that population Π_3 can be understood as a copy of one of Π_1 or

Π_2. The overall sample vector

$$Y_{(n)} = \left(Y^{1T}_{(n_1)}, Y^{2T}_{(n_2)}, Y^{3T}_{(n_3)}\right)^T \qquad , \quad n = n_1 + n_2 + n_3,$$

consisting of the repeated measurement vectors $Y^i_{(n_i)} = (Y_{i1}, \ldots, Y_{in_i})^T$ from the three populations $\Pi_i, i = 1,2,3$ will be assumed to define a Gaussian statistical structure

$$S_n = (R^n, B^n, \{\Phi_{\mu,\Sigma}\,,\ \mu \in M, \Sigma \in \Theta\}).$$

Here, M denotes the range of $EY_{(n)}$ and will be called the model space, although it is not a linear vector space. It satisfies the representation

$$M = \{\mu \in R^n : \mu = \mu_1 1^{+00} + \mu_2 1^{0+0} + \mu_3 1^{00+},$$
$$\mu_3 \in \{\mu_1, \mu_2\}, (\mu_1, \mu_2) \in R^2, \mu_1 \neq \mu_2\},$$

where

$$1^{+00} = (1^T_{n_1} O^T_{n_2+n_3})^T \ , 1^{0+0} = (0^T_{n_1} 1^T_{n_2} 0^T_{n_3})^T, 1^{00+} = (0^T_{n_1+n_2} 1^T_{n_3})^T,$$
$$1_{n_i} = (1, \ldots, 1)^T \in R^{n_i}, 0_m = (0, \ldots, 0)^T \in R^m.$$

Further,

$$\Theta = \left\{ \Sigma = \begin{pmatrix} \sigma_1^2 I_{n_1} & & \\ & \sigma_2^2 I_{n_2} & \\ & & \sigma_3^2 I_{n_3} \end{pmatrix}, (\sigma_1^2, \sigma_2^2) \in R^+ \times R^+, \sigma_3^2 \in \{\sigma_1^2, \sigma_2^2\} \right\}$$

is a set of block diagonal matrices where I_{n_i} denotes a n_i x n_i unit matrix.
The problem of interest here is, on the basis of the overall sample vector $Y_{(n)}$, to decide between the hypotheses

$$H_{1/3} : \mu_3 = \mu_1 \ \text{ and } \ H_{2/3} : \mu_3 = \mu_2.$$

1.2 VECTOR ALGEBRAIC PRELIMINARIES

Put

$$1^{+0+} = 1^{+00} + 1^{00+}, 1^{0++} = 1^{0+0} + 1^{00+}, 1^{+++} = 1^{+00} + 1^{0+0} + 1^{00+}$$

and denote by

$$M_{1/3} = L(1^{+0+}, 1^{0+0}) \text{ and } M_{2/3} = L(1^{0++}, 1^{+00})$$

subspaces of the sample space spanned up by the vectors standing within the brackets. These spaces can be understood as hypotheses spaces or restricted model spaces under the hypotheses $H_{1/3}$ or $H_{2/3}$, respectively. Second basis representations for these spaces are

$$M_{1/3} = L(1^{+++}, 1^{0+0}), \quad M_{2/3} = L(1^{+++}, 1^{+00}).$$

Note that $M_{1/3}$ and $M_{2/3}$ are not orthogonal to each other,

$$M_{1/3} \cap M_{2/3} = L(1^{+++})$$

and

$$1^{0+0} \perp 1^{+00}.$$

While the dimensions of the hypotheses spaces satisfy the equations

$$\dim M_{i/3} = 2,\ i = 1, 2$$

the dimension of

$$\widetilde{M} = L(1^{+00}, 1^{0+0}, 1^{00+}),$$

i.e. the smallest subspace of the sample space containing both $M_{1/3}$ and $M_{2/3}$, equals three. A second basis representation for this so called extended model space is

$$\widetilde{M} = L(1^{+++}, 1^{+00}, 1^{0+0}).$$

The spaces $L(1^{+++})$ and $L(1^{+00}, 1^{0+0})$ are linearly independent but not orthogonal. A third basis representation for $\widetilde{M}$ is

$$\widetilde{M} = L(1^{+++}, 1^{-0+}, 1^{0-+}),$$

where

$$1^{-0+} = -\frac{1}{n_1}1^{+00} + \frac{1}{n_3}1^{00+},\ 1^{0-+} = -\frac{1}{n_2}1^{0+0} + \frac{1}{n_3}1^{00+}.$$

The spaces $L(1^{+++})$ and $L(1^{-0+}, 1^{0-+})$ are orthogonal but

$$(1^{-0+}, 1^{0-+}) = \frac{1}{n_3}. \tag{1.2.1}$$

Since

$$\Pi_{1^{-0+}}\mu = (\mu_3 - \mu_1)1^{-0+} \text{ and } \Pi_{1^{0-+}}\mu = (\mu_3 - \mu_2)1^{0-+},$$

the two-dimensional space

$$W = L(1^{-0+}, 1^{0-+})$$

will be called effect space or decision space. These notations correspond to the circumstances that changes of the differences $(\mu_3 - \mu_i), i = 1, 2$ are immediately reflected in the space W and decisions concerning the magnitude of these differences should be based upon considerations within this space. This can be taken as motivation to define the decision rules

$$d_c | R^n \longrightarrow \{1, 2\},\ c > 0$$

for deciding between the hypotheses $H_{1/3}$ and $H_{2/3}$ as

$$d_c(y_{(n)}) = 2 - I\{\|\Pi_{1^{-0+}}y_{(n)}\| < c\|\Pi_{1^{0-+}}y_{(n)}\|\} \tag{1.2.2}$$

for arbitrary $c > 0$. Here, $I(A)$ denotes the indicator of the random event A. Notice that

$$d_c(Y_{(n)}) = 1$$

holds iff

$$\|\Pi_{1^{-0+}}\Pi_W Y_{(n)}\| < c\|\Pi_{1^{0-+}}\Pi_W Y_{(n)}\|.$$

Recognize further that

$$d_1(Y_{(n)}) = 2 - I\{\|Y_{(n)} - \Pi_{M_{1/3}}Y_{(n)}\| < \|Y_{(n)} - \Pi_{M_{2/3}}Y_{(n)}\|\}.$$

Hence the geometrically motivated decision rule d_c will be called, throughout the present paper, a generalized minimum-distance classification rule.

Let us now consider the orthogonal projection of μ onto the effect space W, $\Pi_W\mu$. Note that $\mu \in M \subset \widetilde{M} = L(1^{+++}, W)$ and $W \perp 1^{+++}$. Hence, $\Pi_W\mu = \mu - \Pi_{1^{+++}}\mu$, i.e.

$$\Pi_W\mu = a(\mu_1, \mu_2, \mu_3)1^{+00} + b(\mu_1, \mu_2, \mu_3)1^{0+0} + c(\mu_1, \mu_2, \mu_3)1^{00+} \tag{1.2.3}$$

with
$n \cdot a(x,y,z) = n_2(x-y) + n_3(x-z), n \cdot b(x,y,z) = n_1(y-x) + n_3(y-z),$
$n \cdot c(x,y,z) = n_1(z-x) + n_2(z-y).$

It is easily seen from this $\widetilde{M}$–basis representation of $\Pi_W \mu$ that the coefficients must depend on each other. We shall therefore try to reduce the number of parameters included in the model. To this end we start with a first reparametrisation. This reparametrisation is based upon a certain partial orthogonalisation.

LEMMA 1.2.1

$$\Pi_W \mu = -n_1 a(\mu_1, \mu_2, \mu_3) 1^{-+0} - n_2 b(\mu_1, \mu_2, \mu_3) 1^{0-+},$$

$$\Pi_{1+++}\mu = \frac{n_1\mu_1 + n_2\mu_2 + n_3\mu_3}{n_1 + n_2 + n_3} 1^{+++}.$$

The proof of the second assertion is obvious. The first assertion follows immediately from the following lemma. Notice that, e.g., the dimension-depending new parameter

$$m(\mu_1, \mu_2, \mu_2) = \frac{1}{n} \sum_{i=1}^{3} \mu_i n_i$$

does not allow an immediate interpretation in the original problem.

LEMMA 1.2.2
The $\widetilde{M}$–vector

$$u = a1^{+00} + b1^{0+0} + c1^{00+} \tag{1.2.4}$$

belongs to the subspace W and allows the representation

$$u = \varphi 1^{-0+} + \psi 1^{0-+} \tag{1.2.5}$$

for some $(\varphi, \psi) \in R^2$ if and only if

$$an_1 + bn_2 + cn_3 = 0. \tag{1.2.6}$$

In this case,

$$\varphi = -n_1 a, \quad \psi = -n_2 b. \tag{1.2.7}$$

PROOF Replacing the two vectors in (1.2.5) by their definitions yields

$$u = -\frac{\varphi}{n_1} 1^{+00} - \frac{\psi}{n_2} 1^{0+0} + \frac{\varphi + \psi}{n_3} 1^{00+}.$$

Equating coefficients from the latter formula with corresponding coefficients from (1.2.4) gives (1.2.7) and

$$\frac{\varphi + \psi}{n_3} = -\frac{n_1 a + n_2 b}{n_3}.$$

The latter quantity coincides with the coefficient c if and only if condition (1.2.6) is fulfilled. ∎

Let us define by

$$\overline{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

the mean in the i-th population, $i = 1, 2, 3$ and by $m(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot})$ the overall mean. It follows then that

$$\Pi_{1+++} Y = m(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{+++}$$

and

$$\Pi_{\widetilde{M}} Y = \overline{Y}_{1\cdot} 1^{+00} + \overline{Y}_{2\cdot} 1^{0+0} + \overline{Y}_{3\cdot} 1^{00+}.$$

Hence, an $\widetilde{M}$–basis representation for $\Pi_W Y$, i.e. for $\Pi_{\widetilde{M}} Y - \Pi_{1+++} Y$ is

$$\Pi_W Y = a(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{+00} + b(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{0+0} + c(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{00+}.$$

The last three equations define $\widetilde{M}$–basis representations of the least squares estimations for the quantities $\Pi_{1+++}\mu$, $\Pi_{\widetilde{M}}\mu$ and $\Pi_W \mu$, respectively.

COROLLARY 1.2.3
Using the above defined functions a and b, a W-basis representation formula for $\Pi_W Y$ is given by

$$\Pi_W Y = -n_1 a(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{-0+} - n_2 b(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) 1^{-0+}.$$

Note that

$$W = L(b_1, b_2)$$

with

$$b_1 = 1^{-0+} \text{ and } b_2 = \frac{1}{n_1 + n_3} 1^{+0+} - \frac{1}{n_2} 1^{0+0} \tag{1.2.8}$$

defines an orthogonal basis representation for W where

$$b_2 = 1^{0-+} - \Pi_{1^{-0+}} 1^{0-+}.$$

From (1.2.1) and

$$\|1^{-0+}\|^2 = \frac{n_1 + n_3}{n_1 n_3} \tag{1.2.9}$$

we get

$$\Pi_{1^{-0+}} 1^{0-+} = \frac{n_1}{n_1 + n_3} 1^{-0+}.$$

Hence,

$$b_2 = -\frac{1}{n_2} 1^{0+0} + \frac{1}{n_3} 1^{00+} + \frac{1}{n_1 + n_3} 1^{+00} - \frac{n_1}{n_3(n_1 + n_3)} 1^{00+}.$$

This yields the second assertion in (1.2.8). The following lemma presents a reparametrisation which is based upon orthogonalisation.

LEMMA 1.2.4
The quantity $\Pi_W \mu$ can be written as

$$\Pi_W \mu = \frac{n_1 n_3}{n_1 + n_3} (\mu_3 - \mu_1) b_1 + d(\mu_1, \mu_2, \mu_3) b_2,$$

whereby the new parameter d satisfies the two representation formulae

$$n d(\mu_1, \mu_2, \mu_3) = n_1 n_2 (\mu_1 - \mu_2) + n_2 n_3 (\mu_3 - \mu_2), \tag{1.2.10}$$

and

$$n d(\mu_1, \mu_2, \mu_3) = n_2 (n_1 + n_3) \left(m^{(1/3)}(\mu_1, \mu_3) - \mu_2 \right) \tag{1.2.11}$$

with

$$m^{(1/3)}(x, y) = \frac{n_1 x + n_3 y}{n_1 + n_3}.$$

PROOF Making use of equations (1.2.3) and (1.2.7) above as well as (1.2.12) below one can see that

$$u = \Pi_W \mu = \Pi_W (\mu_1 1^{+00} + \mu_2 1^{0+0} + \mu_3 1^{00+})$$

allows the representation

$$u = \vartheta b_1 + \nu b_2$$

with

$$\vartheta = -n_1 a(\mu_1,\mu_2,\mu_3) - \frac{n_1 n_2}{n_1+n_3} b(\mu_1,\mu_2,\mu_3)$$

$$= \frac{n_1n_2(\mu_2-\mu_1)+n_1n_3(\mu_3-\mu_1)}{n} + \frac{n_1^2n_2(\mu_1-\mu_2)+n_1n_2n_3(\mu_3-\mu_2)}{n(n_1+n_3)}$$

and

$$\nu = \psi = -n_2 b(\mu_1,\mu_2,\mu_3) = \frac{n_2n_1(\mu_1-\mu_2)+n_2n_3(\mu_3-\mu_2)}{n}.$$

Hence

$$n(n_1+n_3)\vartheta = (n_1+n_3)[n_1n_2(\mu_2-\mu_1)+n_1n_3(\mu_3-\mu_1)]+n_1^2n_2(\mu_1-\mu_2)+n_1n_2n_3(\mu_3-\mu_2)$$

$$= n_1n_3n(\mu_3-\mu_1)$$

and

$$n\nu = n_2(n_1+n_3)\left[\frac{n_1\mu_1+n_3\mu_3}{n_1+n_3} - \mu_2\right],$$

which proves the assertions of the lemma. ▮

Observe that since the new parameters d and $m^{(1/3)}(\mu_1,\mu_3)$ depend on the sample sizes they do not allow immediate interpretations with respect to the original problem.

LEMMA 1.2.5
The $\widetilde{M}$-vector u from (1.2.4) belongs to the subspace W and allows the representation

$$u = \vartheta b_1 + \nu b_2 \tag{1.2.12}$$

for certain values of $(\vartheta,\nu) \in R^2$ if and only if the condition (1.2.6) is satisfied. In this case we have, with (φ,ψ) from (1.2.7),

$$\vartheta = \varphi + \frac{n_1}{n_1+n_3}\psi, \;\; \nu = \psi.$$

PROOF Equating coefficients, (1.2.12) follows from

$$a1^{+00} + b1^{0+0} + c1^{00+} = u = \vartheta b_1 + \nu b_2$$

$$= \vartheta\left(-\frac{1}{n_1}1^{+00} + \frac{1}{n_3}1^{00+}\right) + \nu\left(\frac{1}{n_1+n_3}1^{+0+} - \frac{1}{n_2}1^{0+0}\right)$$

$$= \left(-\frac{\vartheta}{n_1} + \frac{\nu}{n_1+n_3}\right)1^{+00} - \frac{\nu}{n_2}1^{0+0} + \left(\frac{\vartheta}{n_3} + \frac{\nu}{n_1+n_3}\right)1^{00+}.$$

In the same way it follows that

$$c = \frac{\vartheta}{n_3} + \frac{\nu}{n_1+n_3} = -\frac{n_1}{n_3}a - \frac{n_1n_2}{n_3(n_1+n_3)}b - \frac{n_2}{n_1+n_3}b,$$

which is equivalent to (1.2.6). ▮

By definition of b_1,

$$\Pi_{b_1} Y = \frac{n_1 n_3}{n_1 + n_3} (\overline{Y}_{3\cdot} - \overline{Y}_{1\cdot}) b_1 .$$

Using the notation

$$\overline{Y_{\cdot\cdot}^{(1/3)}} = \frac{1}{n_1 + n_3} \sum_{i \in \{1,3\}} \sum_{j=1}^{n_i} Y_{ij} = m^{(1/3)}(\overline{Y}_{1\cdot}, \overline{Y}_{3\cdot})$$

for the pooled mean of the values from the union of the first and third subsamples $Y^1_{(n_1)}$ and $Y^3_{(n_3)}$, we get

$$\begin{aligned} \Pi_{b_2} Y &= d(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) b_2 \\ &= \frac{1}{n}(n_1 n_2(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}) + n_2 n_3(\overline{Y}_{3\cdot} - \overline{Y}_{2\cdot})) b_2 \\ &= \frac{n_2}{n}(n_1 + n_3) \left(\overline{Y_{\cdot\cdot}^{(1/3)}} - \overline{Y}_{2\cdot} \right) b_2 . \end{aligned}$$

Hence we arrive at least squares estimates for $\Pi_{b_1}\mu$ and $\Pi_{b_2}\mu$.
With

$$\|b_1\| = \sqrt{\frac{n_1 + n_3}{n_1 \cdot n_3}} \text{ and } \|b_2\| = \sqrt{\frac{n_1 + n_2 + n_3}{n_2(n_1 + n_3)}} \tag{1.2.13}$$

we get the following representation formula for the least squares estimate of $\Pi_W \mu$ with respect to the normalized orthogonal basis $\{B_1, B_2\}$ where $B_i = b_i / \|b_i\|$:

$$\Pi_W Y = \sqrt{\frac{n_1 n_3}{n_1 + n_3}} (\overline{Y}_{3\cdot} - \overline{Y}_{1\cdot}) B_1 + \sqrt{\frac{n_2(n_1 + n_3)}{n_1 + n_2 + n_3}} \left(\overline{Y_{\cdot\cdot}^{(1/3)}} - \overline{Y}_{2\cdot} \right) B_2 . \tag{1.2.14}$$

1.3 DISTRIBUTIONAL RESULTS

1.3.1 REPRESENTATION FORMULAE BASED UPON THE TWO-DIMENSIONAL GAUSSIAN LAW

The random variables $\overline{Y}_{3\cdot} - \overline{Y}_{1\cdot}$ and $\overline{Y_{\cdot\cdot}^{(1/3)}} - \overline{Y}_{2\cdot}$ play an essential role in the basic formula (1.2.14). They are coefficients of the projections of $Y_{(n)}$ onto the normalized orthogonal basis vectors B_1 and B_2 of the decision space W and are therefore uncorrelated. They are linear combinations of the components of the Gaussian random vector $Y_{(n)}$ and are therefore jointly Gaussian distributed and consequently independent from each other.
Let $\eta_1, \eta_2, \ldots$ denote independent standard Gaussian distributed random variables. The symbol $X_i \sim N_i(a_i, \lambda_i^2)$ will be used to indicate that X_i follows the same distribution law as $a_i + \lambda_i \eta_i$. Thus

$$\sqrt{\frac{n_1 n_3}{n_1 + n_3}} (\overline{Y}_{3\cdot} - \overline{Y}_{1\cdot}) \sim N_1 \left(\sqrt{\frac{n_1 n_3}{n_1 + n_3}} (\mu_3 - \mu_1), \frac{n_1 \sigma_3^2 + n_3 \sigma_1^2}{n_1 + n_3} \right) . \tag{1.3.1}$$

The first and second order moments of this distribution will be denoted by ξ_1 and δ_1^2, respectively. In a similar way as above, one can show that the random variables $\overline{Y_{\cdot\cdot}^{(1/3)}}$ und $\overline{Y}_{2\cdot}$ are independent. Using the notations

$$\overline{Y}_{2\cdot} \sim N_3 \left(\mu_2, \frac{\sigma_2^2}{n_2} \right)$$

and

$$\overline{Y_{\cdot\cdot}^{(1/3)}} \sim N_4 \left(m^{(1/3)}(\mu_1, \mu_3), \frac{n_1 \sigma_1^2 + n_3 \sigma_2^3}{(n_1 + n_3)^2} \right) ,$$

we get

$$\sqrt{\frac{n_2(n_1+n_3)}{n_1+n_2+n_3}}(\overline{Y_{..}^{(1/3)}}-\overline{Y}_{2\cdot}) \sim N_2(\xi_2, \delta_2^2) \tag{1.3.2}$$

with

$$\xi_2 = \sqrt{\frac{n_2(n_1+n_3)}{n_1+n_2+n_3}}(m^{(1/3)}(\mu_1,\mu_3)-\mu_2) = \frac{n_1 n_2(\mu_1-\mu_2)+n_3 n_2(\mu_3-\mu_2)}{\sqrt{(n_1+n_2+n_3)n_2(n_1+n_3)}}$$

and

$$\delta_2^2 = \frac{(n_1+n_3)^2\sigma_2^2 + n_2(n_1\sigma_1^2+n_3\sigma_3^2)}{(n_1+n_2+n_3)(n_1+n_3)}.$$

Recall that

$$d_c(Y_{(n)}) = 1$$

holds if and only if

$$N_1 B_1 + N_2 B_2 \in \{z \in W : \|\Pi_{1^{-0+}} z\| < c\|\Pi_{1^{0-+}} z\|\}. \tag{1.3.3}$$

The following lemma concerns reducing dimension and norming.

LEMMA 1.3.1
The relation

$$N_1 B_1 + N_2 B_2 \in \{z \in W : \|\Pi_{1^{-0+}} z\| < c\,\|\Pi_{1^{0-+}} z\|\}$$

is true iff

$$\left(\frac{N_1}{\delta_1}, \frac{N_2}{\delta_2}\right)^T \in \left\{(t_1,t_2)^T \in R^2 : \frac{|t_1|}{|t_1+\kappa t_2|} < \zeta \cdot c\right\},$$

where

$$\kappa = \sqrt{\frac{n_3[(n_1+n_3)^2\sigma_2^2 + n_2(n_1\sigma_1^2+n_3\sigma_3^2)]}{n_1 n_2(n_1\sigma_3^2+n_3\sigma_1^2)}} \tag{1.3.4}$$

and

$$\zeta = \sqrt{\frac{n_1 n_2}{(n_1+n_3)(n_2+n_3)}} = \frac{1}{\sqrt{(1+n_3/n_1)(1+n_3/n_2)}}. \tag{1.3.5}$$

PROOF Put $A = \{z \in W : \|\Pi_{1^{-0+}} z\|/\|\Pi_{1^{0-+}} z\| < c\}$. Then

$$A = \left\{z_1 B_1 + z_2 B_2 : \frac{\|\Pi_{1^{-0+}}(z_1 B_1 + z_2 B_2)\|}{\|\Pi_{1^{0-+}}(z_1 B_1 + z_2 B_2)\|} < c,\ (z_1,z_2)^T \in R^2\right\}.$$

With

$$(B_1, 1^{0-+}) = \sqrt{\frac{n_1}{n_3(n_1+n_3)}}$$

and

$$(B_2, 1^{0-+}) = \sqrt{\frac{n_1+n_2+n_3}{n_2(n_1+n_3)}}$$

it follows that

$$A = \{z_1 B_1 + z_2 B_2 : |z_1|/|z_1 + \chi \cdot z_2| < \zeta \cdot c, (z_1,z_2)^T \in R^2\},$$

where

$$\chi = \sqrt{n_3(n_1+n_2+n_3)}/\sqrt{n_1 n_2}.$$

Hence,

$$N_1 B_1 + N_2 B_2 \in A$$

iff

$$(N_1,N_2)^T \in \left\{(z_1,z_2)^T \in R^2 : |z_1|/|z_1+\chi z_2| < \zeta \cdot c\right\}.$$

The assertion of the lemma follows now with $\kappa = (\delta_2/\delta_1)\chi$. ∎

Put

$$\nu_1 = \frac{\xi_1}{\delta_1} = \sqrt{\frac{n_1 n_3}{n_1\sigma_3^2 + n_3\sigma_1^2}}(\mu_3 - \mu_1) \tag{1.3.6}$$

and

$$\nu_2 = \frac{\xi_2}{\delta_2} = \frac{\sqrt{n_2}(n_1+n_3)[m^{(1/3)}(\mu_1,\mu_3) - \mu_2]}{\sqrt{(n_1+n_3)^2\sigma_2^2 + n_2(n_1\sigma_1^2 + n_3\sigma_3^2)}} \tag{1.3.7}$$

$$= \frac{n_1(\mu_1-\mu_2) + n_3(\mu_3-\mu_2)}{\sqrt{(n_1+n_3)^2\sigma_2^2 + n_2(n_1\sigma_1^2+n_3\sigma_3^2)}}.$$

If the hypothesis $H_{1/3}$ is true, then $m^{(1/3)}(\mu_1,\mu_3) = \mu_1$,

$$\nu_1 = 0 \tag{1.3.8}$$

and

$$\nu_2 = \sqrt{n_2}\frac{\mu_1-\mu_2}{\sigma_2}\sqrt{\frac{1+n_3/n_1}{1+n_3/n_1+(n_2\sigma_1^2)/(n_1\sigma_2^2)}} = \frac{\sqrt{n_1+n_3}(\mu_1-\mu_2)}{\sqrt{(n_1+n_3)\sigma_2^2+n_2\sigma_1^2}}. \tag{1.3.9}$$

If the hypothesis $H_{1/3}$ is true, then

$$\kappa = \sqrt{\frac{n_3}{n_1} + \frac{(n_1+n_3)n_3\sigma_2^2}{n_1 n_2\sigma_1^2}}, \tag{1.3.10}$$

but if $H_{2/3}$ is true, then

$$\kappa = \sqrt{\frac{n_3\sigma_2^2(n_2n_3 + (n_1+n_3)^2) + n_1n_2n_3\sigma_1^2}{n_1n_2(n_1\sigma_2^2 + n_3\sigma_1^2)}}. \tag{1.3.11}$$

If $H_{1/3}$ is true, then the random event of correct classification

$$CC_1(c) = \{d_c(Y_{(n)}) = 1\},$$

in view of (1.3.3) and Lemma 1.3.1, obtains the representation

$$CC_1(c) = \left\{\frac{|N_1(0,1)|}{|N_1(0,1)+\kappa N_2(\nu_2,1)|} < \zeta\cdot c\right\} \tag{1.3.12}$$

with independent random variables N_1 and N_2 , ν_2 from (1.3.9), κ from (1.3.10) and ζ from (1.3.5). This proves the following theorem.

THEOREM 1.3.2
If the hypothesis $H_{1/3}$ is true, then $P_1(CC_1(c))$, the probability of correct classification into the population Π_1, satisfies the representation formula

$$P_1(CC_1(c)) = \Phi_{(0,\nu_2)^T,\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}}(CC_1^*(c)) \tag{1.3.13}$$

for all $c > 0$, with

$$CC_1^*(c) = \left\{(t_1,t_2)^T \in R^2 : \frac{|t_1|}{|t_1+\kappa t_2|} < \zeta\cdot c\right\}$$

and ν_2, κ and ζ as in (1.3.9), (1.3.10) and (1.3.5),respectively.

If the hypothesis $H_{2/3}$ is true, then the random event of correct classification into the population Π_2,

$$CC_2(c) = \{d_c(Y_{(m)}) = 2\},$$

satisfies for all $c > 0$ the representation

$$CC_2(c) = \left\{ \frac{|N_1(\nu_1, 1)|}{|N_1(\nu_1, 1) + \kappa N_2(\nu_2, 1)|} > \zeta \cdot c \right\}, \tag{1.3.14}$$

where κ and ζ are to be chosen according to (1.3.11) and (1.3.5), respectively,

$$\nu_1 = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_2^2/n_3 + \sigma_1^2/n_1}} = \frac{\sqrt{n_1 n_3}(\mu_2 - \mu_1)}{\sqrt{n_1\sigma_2^2 + n_3\sigma_1^2}} \tag{1.3.15}$$

and, since

$$m^{1/3}(\mu_1, \mu_2) - \mu_2 = \frac{\mu_1 - \mu_2}{1 + n_3/n_1},$$

ν_2 is defined as

$$\nu_2 = \sqrt{n_2}\frac{\mu_1 - \mu_2}{\sigma_2} \Big/ \sqrt{\left(1 + \frac{n_3}{n_1}\right)^2 + \frac{n_2}{n_1}\frac{\sigma_1^2}{\sigma_2^2} + \frac{n_2 n_3}{n_1^2}} \tag{1.3.16}$$

$$= \frac{n_1(\mu_1 - \mu_2)}{\sqrt{\sigma_2^2(n_2 n_3 + (n_1 + n_3)^2) + \sigma_1^2 n_1 n_2}}.$$

The following theorem is an immediate consequence.

THEOREM 1.3.3
If $H_{2/3}$ is true then the probability of correct classification into the population Π_2, $P_2(CC_2(c))$, satisfies the representation formula

$$P_2(CC_2(c)) = \Phi_{(\nu_1,\nu_2)^T, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}(CC_2^*(c)) \tag{1.3.17}$$

for all $c > 0$, with

$$CC_2^*(c) = \left\{ (t_1, t_2)^T \in R^2 : \frac{|t_1|}{|t_1 + \kappa t_2|} > \zeta \cdot c \right\}$$

and ν_1, ν_2, κ, ζ as in (1.3.15), (1.3.16), (1.3.11) and (1.3.5), respectively.

The representation formulae for $P_1(CC_1(c))$ in Theorem 1.3.2 and for $P_2(CC_2(c))$ in Theorem 1.3.3 do not only reflect different quantitative situations but they are also of different qualitative nature. The most obvious difference between $\nu_1 = 0$ in (1.3.8) and $\nu_1 \neq 0$ in (1.3.15) can be easily detected. The following theorem presents a second representation formula for $P_2(CC_2(c))$ which corresponds in quality to that for $P_1(CC_1(c))$ in Theorem 1.3.2. Its proof repeats that of Theorem1.3.2 and will be suppressed therefore here.

THEOREM 1.3.4
If $H_{2/3}$ is true then the probability of correct classification into the population Π_2 satisfies for all $c > 0$ the representation

$$P_2(CC_2(c)) = \Phi_{(0,\tilde{\nu}_2)^T, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}(\widetilde{CC}_2(c)), \tag{1.3.18}$$

where

$$\widetilde{CC}_2(c) = \left\{ (t_1, t_2)^T \in R^2 : \frac{|t_1|}{|t_1 + \tilde{\kappa} t_2|} < \frac{\zeta}{c} \right\}$$

with ζ as in (1.3.5),

$$\tilde{\kappa} = \sqrt{\frac{n_3}{n_2} + \frac{n_3(n_2+n_3)}{n_1 n_2} \frac{\sigma_1^2}{\sigma_2^2}} \tag{1.3.19}$$

and

$$\tilde{\nu}_2 = \frac{\sqrt{n_2+n_3}(\mu_2 - \mu_1)}{\sqrt{(n_2+n_3)\sigma_1^2 + n_1\sigma_2^2}}. \tag{1.3.20}$$

The next theorem follows by analogy.

THEOREM 1.3.5
If $H_{1/3}$ is true then it holds for all $c > 0$

$$P_1(CC_1(c)) = \Phi_{(\tilde{\nu}_1, \tilde{\nu}_2)^T, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} (\widetilde{CC_1(c)}), \tag{1.3.21}$$

where

$$\widetilde{CC}_1(c) = \left\{ (t_1, t_2)^T \in R^2 : \frac{|t_1|}{|t_1 + \tilde{\kappa} t_2|} > \frac{\zeta}{c} \right\}$$

with ζ as in (1.3.5),

$$\tilde{\kappa} = \sqrt{\frac{n_3\sigma_1^2(n_1n_3 + (n_2+n_3)^2) + n_1n_2n_3\sigma_2^2}{n_1n_2(n_1\sigma_1^2 + n_3\sigma_2^2)}} \tag{1.3.22}$$

and

$$\tilde{\nu}_1 = \frac{\sqrt{n_2n_3}(\mu_1 - \mu_2)}{\sqrt{n_2\sigma_1^2 + n_3\sigma_2^2}} \tag{1.3.23}$$

as well as

$$\tilde{\nu}_2 = \frac{n_2(\mu_2 - \mu_1)}{\sqrt{\sigma_1^2(n_1n_3 + (n_2+n_3)^2) + n_1n_2\sigma_2^2}}. \tag{1.3.24}$$

1.3.2 Representation formulae based upon the doubly non central F–distribution

It was shown in John (1961) and Moran (1975) that the probabilities of correct classification can be expressed in terms of the doubly noncentral F–distribution if expectations are unknown but covariance matrices are known and equal and the linear discriminant function is used for classifying an individual into one of the populations Π_1 and Π_2. It has been recently proved, in Krause and Richter(1999), that the probabilities of correct classification can be also expressed in terms of the doubly noncentral F–distribution if both expectations and covariance matrices are unknown but a certain generalized minimum-distance rule is used for making the decision. Here, a result will be derived which is equivalent in content to the latter one but different from it in form. The method of proving this result developed here differs from that in Krause and Richter(ibid) in using basically a two-dimensional representation formula from the preceding section whereas the proof of the corresponding result in Krause and Richter(ibid) starts from a sample space measure representation formula.
The aim of what follows is to determine the Gaussian measure of $CC_1^*(c)$ in accordance with Theorem 1.3.2. To this end we describe the boundary of the set of

points satisfying the inequality

$$\frac{|t_1|}{|t_1+\kappa t_2|}<\zeta c \tag{1.3.25}$$

with the help of the straight lines

$$g_j : t_2 = -\frac{1}{\kappa}\left(1+\frac{(-1)^j}{\zeta c}\right)t_1, j=1,2. \tag{1.3.26}$$

It turns out that the set of solutions of (1.3.25) includes the t_2–axis in its inner part. The straight line g_2 belongs for all values of ζc to the union of the second and third quadrants in a cartesian coordinate system, i.e. it belongs to the union of the sets $\{t_1<0, t_2>0\}$ and $\{t_1>0, t_2>0\}$.
The straight line g_1 belongs to the same set if $\zeta c>1$ but to the union of the first and fourth quadrants if $0<\zeta c<1$. The straight lines g_1 and g_2 intersect within the set (1.3.25) under an angle α satisfying

$$\alpha=\pi-\arctan\left(-\frac{1}{\kappa}\left(1+\frac{1}{\zeta c}\right)\right)+(-1)^{I\{\zeta c<1\}}\cdot\arctan\left(-\frac{1}{\kappa}\left(1-\frac{1}{\zeta c}\right)\right). \tag{1.3.27}$$

Notice that the set of points $(t_1,t_2)^T$ corresponding to (1.3.25) represents a cone. That is why there exists a vector $(t_{10},t_{20})^T\in R^2$ and a positive real number $d=d(\zeta c)$ such that a vector $(t_1,t_2)^T\in R^2$ satisfies condition (1.3.25) if and only if it satisfies the condition

$$\frac{\|(t_1,t_2)^T-\Pi_{(t_{10},t_{20})^T}(t_1,t_2)^T\|^2}{\|\Pi_{(t_{10},t_{20})^T}(t_1,t_2)^T\|^2}<d^2. \tag{1.3.28}$$

The latter condition is equivalent to

$$\frac{(t_1t_{20}-t_2t_{10})^2}{(t_1t_{10}+t_2t_{20})^2}<d^2$$

or

$$(q-t_2/t_1)^2<d^2(1+qt_2/t_1)^2 \tag{1.3.29}$$

where

$$q=t_{20}/t_{10}.$$

Let us determine now a solution (q, d). Recall that the boundary of (1.3.25) can be described by the equations (1.3.26). The temporary assumption that we have in (1.3.25) and (1.3.29) equalities instead of inequalities leads us to the equation systems

$$\frac{q-d}{dq+1}=-\frac{1}{\kappa}+\frac{1}{\kappa\zeta c} \quad , \quad \frac{q+d}{1-dq}=-\frac{1}{\kappa}-\frac{1}{\kappa\zeta c} \tag{1.3.30}$$

and

$$\frac{q-d}{dq+1}=-\frac{1}{\kappa}-\frac{1}{\kappa\zeta c} \quad , \quad \frac{q+d}{1-dq}=-\frac{1}{\kappa}+\frac{1}{\kappa\zeta c}. \tag{1.3.31}$$

The solution of (1.3.30) is given by

$$d=\frac{1}{2}\left(\kappa\zeta c+\frac{\zeta c}{\kappa}-\frac{1}{\kappa\zeta c}\right)+\frac{1}{2}\sqrt{\left(\kappa\zeta c+\frac{\zeta c}{\kappa}-\frac{1}{\kappa\zeta c}\right)^2+4}, \tag{1.3.32}$$

$$q=\left[d-\frac{1}{\kappa}+\frac{1}{\kappa\zeta c}\right]/\left[1+\frac{d}{\kappa}-\frac{d}{\kappa\zeta c}\right]. \tag{1.3.33}$$

The solution of (1.3.31) is given by

$$d = -\frac{1}{2}\left(\kappa\zeta c + \frac{\zeta c}{\kappa} - \frac{1}{\kappa\zeta c}\right) + \frac{1}{2}\sqrt{\left(\kappa\zeta c + \frac{\zeta c}{\kappa} - \frac{1}{\kappa\zeta c}\right)^2 + 4}, \tag{1.3.34}$$

$$q = \left[d - \frac{1}{\kappa} - \frac{1}{\kappa\zeta c}\right] / \left[1 + \frac{d}{\kappa} + \frac{d}{\kappa\zeta c}\right]. \tag{1.3.35}$$

Consequently, the representation (1.3.25) for the cone under consideration is equivalent to the representation (1.3.29) if the quantities d and q are chosen there either according to (1.3.32) and (1.3.33) or according to (1.3.34) and (1.3.35), respectively. Hence, the following lemma has been proved.

LEMMA 1.3.6

The probability of correct classification into Π_1 satisfies the representation $P_1(CC_1(c)) =$

$$\Phi_{(0,v_2)^T, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}\left(\left\{(t_1,t_2)^T \in R^2 : \frac{\|(t_1,t_2)^T - \Pi_{(t_{10},t_{20})^T}(t_1,t_2)^T\|^2}{\|\Pi_{(t_{10},t_{20})^T}(t_1,t_2)^T\|^2} < d^2\right\}\right).$$

Here, v_2 is chosen as in (1.3.9), $(t_{10}, t_{20})^T \in R^2$ is an arbitrary vector satisfying $t_{20}/t_{10} = q$ and d and q are to be chosen according to either (1.3.32) and (1.3.33) or (1.3.34) and (1.3.35).

It follows from the definition of the doubly noncentral F–distribution with (1,1) degrees of freedom and from the invariance of the two-dimensional standard Gaussian measure with respect to orthogonal transformations that $P_1(CC_1(c))$ can be expressed as a suitable value of the cumulative distribution function $F_{1,1,\Delta_1^2,\Delta_2^2}$. The noncentrality parameters of this distribution are

$$\Delta_2^2 = \left\|\Pi_{(t_{10},t_{20})^T}(0,\nu_2)^T\right\|^2 = \frac{t_{20}^2\nu_2^2}{t_{10}^2 + t_{20}^2}$$

and

$$\Delta_1^2 = \left\|(0,\nu_2)^T - \Pi_{(t_{10},t_{20})^T}(0,\nu_2)^T\right\|^2 = \frac{t_{10}^2\nu_2^2}{t_{10}^2 + t_{20}^2}.$$

As a result, the following theorem has been proved.

THEOREM 1.3.7

If $H_{1/3}$ is true, then

$$P_1(CC_1(c)) = F_{1,1,\Delta_1^2,\Delta_2^2}(d^2)$$

with ν_2, d, q as in Lemma 1.3.6 and

$$\Delta_1^2 = \frac{{\nu_2}^2}{1+q^2},\ \Delta_2^2 = \frac{\nu_2^2}{1+1/q^2}.$$

Notice that a corresponding result in Krause and Richter (1999) is different in form and it is not obvious how to transform one of the results into the other in a direct way.

References

1. Anderson, T. W. (1984). *Multivariate Statistical Analysis.* J.Wiley and Sons, New York.
2. Cacoullos, T. (1992). Two LDF Characterizations of the Normal as a Spherical Distribution. *J.Multivar.Anal.* **40**, 205-212.
3. Cacoullos, T. and Koutras, M. (1985). Minimum distance discrimination for spherical distributions. In: Matusita, K., ed., *Statistical Theory and Data Analysis*, 91-102, North Holland.
4. Cacoullos, T. and Koutras, M. (1996). On the Performance of Minimum-Distance Classification Rules for Kotz-type Elliptical Distributions. In: Johnson, N.L. and Balakrishnan, N., eds., *Advances in the theory and Practice of Statistics: A Volume in Honor of Samuel Kotz* , 209-224, Wiley, New York.
5. Hills, M. (1966). Allocation rules and their error rates. *Journal of the Royal Statistical Society*, Series B, **28**, 1-31.
6. John, S. (1961). Errors in discrimination.*Annals Math. Statistics*, **32**, 1125-1144.
7. Krause, D. and Richter, W.-D. (1994). Geometric Approach to Evaluating Probabilities of Correct Classification into two Gaussian or Spherical Categories. In: Bock, H.-H., Lenski, W. and Richter, M.,eds., *Information Systems and Data Analysis*, 242-250. Springer-Verlag, Berlin.
8. Krause, D. and Richter,W.-D.(1999). Exact Probabilities of Correct Classifications for Repeated Measurements from Elliptically Contoured Distributions. Submitted.
9. McLachlan, G. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* Wiley and Sons, New York.
10. Moran, M.A. (1975). On the expectation of errors of allocation associated with a linear discrimination function. *Biometrika*, **62**, 1, 141-148.
11. Schaafsma, W., van Vark, G. (1977). Classification and discrimination problems with applications, Part I. *Statist. Neerlandica*, **31**, 25-45.