Central limit theorem for probabilities of correct classification

Wolf-Dieter Richter

Fachbereich Mathematik, Universität Rostock,D-18051 Rostock

Abstract: A minimum-distance classification rule studied in Krause and Richter (1994) for measurements from spherically contoured distributions will be used here for allocating measurements from unspecified distributions but having finite second order moments. The derivation of a representation formula for exact probabilities of correct classification similar to that derived in Richter (2000) for measurements from spherically contoured distributions will be the first step of proving a central limit theorem for probabilities of correct classification.

1 Introduction

A minimum Euclidean distance based classification rule for measurements from spherically contoured distributions was derived in Krause and Richter (1994) within a certain non classical linear model approach. This rule will be used here for allocating measurements from unspecified distributions but having finite second order moments. Probabilities of correct classification will be analysed in two steps. A representation formula for exact probabilities of correct classification similar to that derived in Richter (2000) for measurements from spherically contoured distributions will be checked in a first, vector algebraic, step. This representation formula will be the starting point for proving in a second, analytical, step a central limit theorem for probabilities of correct classification. To this end, the law of large numbers for the three dimensional vector of measurement means in the three populations and the corresponding central limit theorem will be combined with the help of the well known delta method.

2 Exact probabilities of correct classification

Let an investigator make n_i independent measurements from three populations $\Pi_i, i = 1, 2, 3$. Denote the over all sample size by $n, n = n_1 + n_2 + n_3$ and assume that the over all sample vector follows an unspecified distribution having finite variances σ_i^2 and expectations $\mu_i, i = 1, 2, 3$:

$$Y_{(n)} = \begin{pmatrix} Y^1_{(n_1)} \\ Y^2_{(n_2)} \\ Y^3_{(n_3)} \end{pmatrix} \sim \begin{pmatrix} V_1(\mu_1, \sigma_1^2)^{\times\, n_1} \\ V_2(\mu_2, \sigma_2^2)^{\times\, n_2} \\ V_3(\mu_3, \sigma_3^2)^{\times\, n_3} \end{pmatrix} .$$

We shall not demand that the distribution V_3 of the third population coincides with one of the distributions V_1, V_2 of the two other populations but that

$$\sigma_3^2 = \sigma_i^2 \text{ holds iff } \mu_3 = \mu_i,\ i \in \{1,2\}. \qquad (*)$$

The problem of interest here is to decide between two hypotheses:

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

where it is natural to assume that

$$\mu_1 \neq \mu_2.$$

The statistical structure under consideration is consequently

$$S_n = \left(R^n, \mathcal{B}^n, \{V_{(\mu,\Sigma)}, \mu \in \mathfrak{M}, \Sigma \in \Theta, (*)\}\right).$$

Here, $V = V_1^{\times n_1} \times V_2^{\times n_2} \times V_3^{\times n_3}$ is an unspecified but structured product measure, μ belongs to the model space $\mathfrak{M}$ which is not a linear space

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

$$1^{+00} = \begin{pmatrix} 1_{n_1} \\ 0_{n_2+n_3} \end{pmatrix}, 1^{0+0} = \begin{pmatrix} 0_{n_1} \\ 1_{n_2} \\ 0_{n_3} \end{pmatrix}, 1^{00+} = \begin{pmatrix} 0_{n_1+n_2} \\ 1_{n_3} \end{pmatrix},$$

$$1_{n_i} = (1, \dots, 1)^\top \in R^{n_i}, 0_{n_i} = (0, \dots 0)^\top \in R^{n_i}$$

and Σ belongs to the space of disturbing parameters

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

I_{n_i} being the $n_i \times n_i$ - unit matrix.

Remarks

- It is not assumed that $V_3 \in \{V_1, V_2\}$!
- Higher than second order moments are not assumed to exist!
- The case that the over all sample vector $Y_{(n)}$ follows an elliptically contoured distribution $V(\mu, \Sigma) = EC_n(\mu, \Sigma_i; g)$ with density generating function g has been dealt with in Krause and Richter (1994). Note that the function g makes it possible to reflect the influence of heavy or light distribution tails, respectively.

The hypotheses $H_{1/3}$ and $H_{2/3}$ can be reflected in the hypotheses spaces

$$\mathfrak{M}_{1/3} = \mathcal{L}(1^{+0+}, 1^{0+0}) \text{ and } \mathfrak{M}_{2/3} = \mathcal{L}(1^{0++}, 1^{+00}),$$

respectively. Note that $\mathfrak{M}_{1/3}$ and $\mathfrak{M}_{2/3}$ are linear subspaces of R^n with

$$1^{+0+} = 1^{+00} + 1^{00+} \quad \text{and} \quad 1^{0++} = 1^{0+0} + 1^{00+}.$$

Recognize further that

$$\mathfrak{M}_{1/3} \cap \mathfrak{M}_{2/3} = \mathcal{L}(1^{+++}) \quad \text{with} \quad 1^{+++} = 1^{+0+} + 1^{0+0}$$

is a discrimination free space. The smallest linear subspace of R^n including the model space $\mathfrak{M}$ is the so called extended model space:

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

Another basis representation for this space is given by

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

where

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

Let us call

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

the action space or, alternatively, effect or decision space. The effects reflected therein are

$$\Pi_{1^{-0+}} \mu = \frac{n_1 n_3}{n_1 + n_3} (\mu_3 - \mu_1) 1^{-0+}$$

and

$$\Pi_{1^{0-+}} \mu = \frac{n_2 n_3}{n_2 + n_3} (\mu_3 - \mu_2) 1^{0-+}.$$

The least squares estimates for the effects $\mu_3 - \mu_1$ and $\mu_3 - \mu_2$ can be derived from the relations

$$\widehat{\Pi_{1^{-0+}} \mu} = \Pi_{1^{-0+}} Y_{(n)} = \frac{n_1 n_3}{n_1 + n_3} (\overline{Y}_{3\cdot} - \overline{Y}_{1\cdot}) 1^{-0+}$$

and

$$\widehat{\Pi_{1^{0-+}} \mu} = \Pi_{1^{0-+}} Y_{(n)} = \frac{n_2 n_3}{n_2 + n_3} (\overline{Y}_{3\cdot} - \overline{Y}_{2\cdot}) 1^{0-+} .$$

It follows that

$$\widehat{\mu_3 - \mu_i} = \overline{Y}_{3\cdot} - \overline{Y}_{i\cdot}, i = 1, 2.$$

A generalized minimum-distance classification rule $d_c | R^n \to \{1, 2\}$ studied in Krause and Richter (1994) is defined as

$$d_c(y_{(n)}) = 2 - I\{\widetilde{CC}_1(c)\}, c > 0,$$

where

$$\widetilde{CC}_1(c) = \left\{y_{(n)} \in R^n : \|\Pi_{1^{-0+}} y_{(n)}\| < c \|\Pi_{1^{0-+}} y_{(n)}\|\right\}$$

$$= \left\{ y_{(n)} \in R^n : |\overline{y}_{3\cdot} - \overline{y}_{1\cdot}| < c \sqrt{\frac{1 + n_3/n_1}{1 + n_3/n_2}} \cdot |\overline{y}_{3\cdot} - \overline{y}_{2\cdot}| \right\}.$$

If $Y_{(n)}$ is Gaussian distributed and $c = 1$ then this decision rule coincides with the maximum likelihood rule. The event of correct classification is $CC_i(c)$ if $H_{i/3}$ is true, $i \in \{1, 2\}$, where

$$CC_1(c) = Y_{(n)}^{-1}\left(\widetilde{CC}_1(c)\right), CC_2(c) = \overline{CC_1(c)}.$$

The probability of correct classification is under $H_{1/3}$ therefore

$$P\left(CC_1(c)\right) = P\left(\frac{\|\Pi_{1^{-0+}} \Pi_{\mathcal{M}} Y_{(n)}\|}{\|\Pi_{1^{0-+}} \Pi_{\mathcal{M}} Y_{(n)}\|} < c\right).$$

Note that

$$\Pi_{\widetilde{\mathcal{M}}} y_{(n)} = \overline{y}_{1\cdot} 1^{+00} + \overline{y}_{2\cdot} 1^{0+0} + \overline{y}_{3\cdot} 1^{00+}$$

and hence

$$\Pi_{\mathcal{M}} y_{(n)} = a\left(\overline{y}_{1\cdot}, \overline{y}_{2\cdot}, \overline{y}_{3\cdot}\right) 1^{-0+} + b\left(\overline{y}_{1\cdot}, \overline{y}_{2\cdot}, \overline{y}_{3\cdot}\right) 1^{0-+}$$

for suitably defined functions $a\left(\overline{y}_{1\cdot}, \overline{y}_{2\cdot}, \overline{y}_{3\cdot}\right)$ and $b\left(\overline{y}_{1\cdot}, \overline{y}_{2\cdot}, \overline{y}_{3\cdot}\right)$ given in Richter (2000). The last equation can be rewritten as

$$\Pi_{\mathcal{M}} y_{(n)} = \eta_1(y_{(n)}) B_1 + \eta_2(y_{(n)}) B_2$$

or

$$\Pi_{\mathcal{M}} y_{(n)} = \sqrt{\frac{n_1 n_3}{n_1 + n_3}} (\overline{y}_{3\cdot} - \overline{y}_{1\cdot}) B_1 + \sqrt{\frac{n_2(n_1 + n_3)}{n}} \left(\overline{y}^{(1/3)} - \overline{y}_{2\cdot}\right) B_2,$$

where

$$B_1 = \frac{1^{-0+}}{\|1^{-0+}\|} \quad \text{and} \quad B_2 = \frac{1^{0-+} - \Pi_{1^{-0+}} 1^{0-+}}{\|1^{0-+} - \Pi_{1^{-0+}} 1^{0-+}\|}$$

are orthogonal unit vectors and

$$\overline{y}^{(1/3)} = \frac{n_1 \overline{y}_{1\cdot} + n_3 \overline{y}_{3\cdot}}{n_1 + n_3}.$$

Notice that the corresponding representation formula for $\Pi_{\mathcal{M}}\mu$,

$$\Pi_{\mathcal{M}}\mu = \sqrt{\frac{n_1 n_3}{n_1 + n_3}}(\mu_3 - \mu_1)B_1 + \sqrt{\frac{n_2(n_1 + n_3)}{n}} \left(\overline{\mu}^{(1/3)} - \mu_2\right) B_2,$$

reflects a certain reparametrisation based upon orthogonalisation and that

$$\Pi_{\mathcal{M}}\mu|_{H_{1/3}} = \sqrt{\frac{n_2(n_1 + n_3)}{n}}(\mu_1 - \mu_2)B_2.$$

The new parameter

$$\overline{\mu}^{(1/3)} = \frac{n_1 \mu_1 + n_3 \mu_3}{n_1 + n_3}$$

depends on the sample sizes n_1 and n_3 and can therefore not easily be interpreted within the original problem. Let

$$\Pi_{\mathcal{M}} Y_{(n)} = H_1 B_1 + H_2 B_2.$$

As a result from the above vector algebraic consideration we got the following dimension reduction formula.

Theorem 1 With

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

and

$$\zeta = 1/\sqrt{(1 + n_3/n_1)(1 + n_3/n_2)}$$

it holds

$$P\left(CC_1(c)\right) = P^{H_1 B_1 + H_2 B_2}\left(\{\eta_1 B_1 + \eta_2 B_2 \in R^n : |\eta| < |\eta_1 + \chi \eta_2| c\zeta\}\right)$$

$$= P^{(H_1, H_2)^\top}\left(\left\{\begin{pmatrix}\eta_1 \\ \eta_2\end{pmatrix} \in R^2 n : |\eta_1| < |\eta_1 + \chi \eta_2| c\zeta\right\}\right).$$

3 Central limit theorem

In this section we exploit the well known delta method. Notice that

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \sqrt{n_1} f_{n_1}(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) \qquad (**)$$

where

$$f_{n_1}(x,y,z) = \left(\frac{z-x}{\sqrt{n_1}\|1^{-0+}\|}, \frac{1}{\sqrt{n_1}\|1^{0-+} - \Pi_{1^{-0+}} 1^{0-+}\|} \left[\frac{n_1 x + n_3 z}{n_1 + n_3} - y \right] \right)^\top .$$

In what follows we shall assume that the condition of asymptotically comperable sample sizes will be fulfilled:

$$\frac{n_i(n_1)}{n_1} \to \lambda_i, \; n_1 \to \infty, \; i \in \{2,3\}. \qquad (***)$$

If condition (***) is satisfied then

$$f_{n_1}(x,y,z) \to f(x,y,z) = \begin{pmatrix} f_1(x,y,z) \\ f_2(x,y,z) \end{pmatrix}$$

where

$$f_1(x,y,z) = \frac{z-x}{\sqrt{1+1/\lambda_3}}, f_2(x,y,z) = \sqrt{\frac{\lambda_2(1+\lambda_3)}{1+\lambda_2+\lambda_3}} \left(\frac{x+\lambda_3 z}{1+\lambda_3} - y \right).$$

The law of large numbers ensures that under (***) a.s.

$$\begin{pmatrix} \overline{Y}_{1\cdot} \\ \overline{Y}_{2\cdot} \\ \overline{Y}_{3\cdot} \end{pmatrix} \longrightarrow \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, n_1 \to \infty.$$

From the central limit theorem and assumption (***) it follows that

$$\mathcal{L}\left(\sqrt{n_1} \left[\begin{pmatrix} \overline{Y}_{1\cdot} \\ \overline{Y}_{2\cdot} \\ \overline{Y}_{3\cdot} \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \right] \right) \Longrightarrow \Phi_{O_3,\Lambda} \quad \text{as} \quad n_1 \to \infty$$

where

$$\Lambda = \begin{pmatrix} \sigma_1^2 & & \\ & \sigma_2^2/\lambda_2 & \\ & & \sigma_3^2/\lambda_3 \end{pmatrix}.$$

Recall that, in general, the delta method applies with a fixed function, e.g. f. We must switch therefore in relation (**) from the sequence of functions f_{n_1} to the fixed function f. The following lemma proves then what we need:

$$\left(f_{n_1}(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) - f(\mu_1, \mu_2, \mu_3)\right)\sqrt{n_1} = o_P(1) \text{ as } n_1 \to \infty.$$

Lemma If (x, y, z) belongs to a sufficiently small neighbourhood of (μ_1, μ_2, μ_3) then

$$(f_{n_1}(x, y, z) - f(x, y, z))\sqrt{n_1} = o(1) \text{ as } n_1 \to \infty.$$

Theorem 2 If condition (***) is satisfied then

$$\mathcal{L}\left(\left(f_{n_1}(\overline{Y}_{1\cdot}, \overline{Y}_{2\cdot}, \overline{Y}_{3\cdot}) - f(\mu_1, \mu_2, \mu_3)\right)\sqrt{n_1}\right) \Longrightarrow \Phi_{O_2,\, \mathcal{F}\Lambda\mathcal{F}^\top},$$

where

$$\mathcal{F} = \frac{1}{\lambda_3 + 1}\begin{pmatrix} \lambda_3\sigma_1^2 + \sigma_3^2 & \dfrac{\sqrt{\lambda_2\lambda_3}[\sigma_3^2 - \sigma_1^2]}{\sqrt{1 + \lambda_2 + \lambda_3}} \\ & \dfrac{\lambda_2\sigma_1^2 + (1 + \lambda_3)^2\sigma_2^2 + \lambda_2\lambda_3\sigma_3^2}{1 + \lambda_2 + \lambda_3} \end{pmatrix}.$$

Approximation Formula 1 If sample sizes are sufficiently large then

$$P\left((H_1, H_2) \in A\right) \approx \Phi_{\sqrt{n_1}f(\mu_1,\mu_2,\mu_3),\, \mathcal{F}\Lambda\mathcal{F}^\top}(A).$$

Recognize that if $H_{1/3}$ is true then

$$f_1(\mu_1, \mu_2, \mu_3) = 0 \text{ and } f_2(\mu_1, \mu_2, \mu_3) = \sqrt{\frac{\lambda_2(1 + \lambda_3)}{1 + \lambda_2 + \lambda_3}}(\mu_1 - \mu_2)$$

as well as

$$\mathcal{F} = \mathcal{F}_0 = \begin{pmatrix} \sigma_1^2 & 0 \\ & \dfrac{\lambda_2\sigma_1^2 + (1 + \lambda_3)\sigma_2^2}{1 + \lambda_2 + \lambda_3} \end{pmatrix}.$$

Remark Notice that the parameters χ and ζ in the definition of the set ${CC_1}^0(c)$ depend on the sample size n_1. Hence, the central limit theorem and Approximation Formula 1 do not immediately apply for this set. However, the dependence of χ and ζ on n_1 is not very strong as can be seen from the following asymptotic relations which hold for $n_1 \to \infty$:

$$\zeta = \zeta(n_1) = \sqrt{\lambda_2(1 + \lambda_3)^{-1}(\lambda_2 + \lambda_3)^{-1}} + 0\left(\frac{1}{\sqrt{n_1}}\right),$$

$$\chi = \chi(n) = \sqrt{\lambda_3(1 + \lambda_2 + \lambda_3)/\lambda_2} + 0\left(\frac{1}{\sqrt{n_1}}\right).$$

As a consequence, we can derive the Approximation Formula 2 which is, from an applied point of view, the main result of the present paper.

Approximation Formula 2 If $H_{1/3}$ is true and the sample sizes are sufficiently large then

$$P\left(CC_1(c)\right) \approx \Phi_{\left(\begin{array}{c}0\\ \sqrt{n_1}f_2\end{array}\right),\, \mathcal{F}_0 \Lambda \mathcal{F}_0^\top} \left(CC_1^\circ(c)\right)$$

where

$$CC_1^\circ(c) = \left\{ \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \in R^2 : |\eta_1| < |\eta_1 + \chi\eta_2| c\zeta \right\}.$$

References

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.

RICHTER, W.-D. (2000): Representation Formulae for Probabilities of Correct Classification. Submitted.