Exact probabilities of correct classifications for uncorrelated repeated measurements from elliptically contoured distributions

D. Krause, W.-D. Richter

Fachbereich Mathematik, Universität Rostock,
D-18051 Rostock, Germany
wolf-dieter.richter@mathematik.uni-rostock.de

Euclidean distance based classification rules are derived within a certain non classical linear model approach and applied to elliptically contoured samples having a density generating function g. Then a geometric measure theoretical method to evaluate exact probabilities of correct classification for multivariate uncorrelated feature vectors is developed. When doing this one has to measure suitably defined sets with certain standardized measures. The geometric key point is that the intersection percentage functions of the areas under investigation coincide with those of certain parabolic cylinder type sets. The intersection percentage functions of the latter sets can be described as threefold integrals. It turns out that these intersection percentage functions yield simultaneously geometric representation formulae for the doubly noncentral g-generalized F-distributions. Hence we get beyond new formulae for evaluating probabilities of correct classification new geometric representation formulae for the doubly noncentral g-generalized F-distributions. A numerical study concerning several aspects of evaluating both probabilities of correct classification and values of the doubly noncentral g-generalized F-distributions demonstrates advantageous computational properties of the present new approach. This impression will be supported by comparison with literature.

It is shown that probabilities of correct classification depend on the parameters of the underlying sample distribution through a certain well defined set of secondary parameters. If the underlying parameters are unknown, we propose to estimate probabilities of correct classification.

AMS 1991 subject classifications: 62H30, 62E15, 62-04, 65U05.

Key words and phrases: Classification probabilities, allocation, uncorrelated repeated measurements, elliptically contoured samples, exact distributions, doubly noncentral generalized F-distribution, intersection percentage function, geometric representation formulae.

W.-D. Richter
FB Mathematik
Universität Rostock
18051 Rostock
Germany

Tel: ++49 +381 4986663
Fax: ++49 +381 4986553
e-mail: wolf-dieter.richter@mathematik.uni-rostock.de

Manuscript sent in LaTeX

1 Introduction

Methods for deriving exact formulae for probabilities of correct classification are known in several situations when elements of a third population are to be allocated to one of two given populations. Exact results for the maximum likelihood rule in the normality case with unknown expectations, equal variances and sample size one in the third population can be found for one-dimensional measurements, e.g., in Hills (1966) and Schaafsma and van Vark (1977). For multivariate measurements Dorvlo (1993) shows that the distributions of the linear and quadratic discriminant functions in the cases of known or unknown expectations but known covariance matrices are the same as the distributions of certain linear combinations of independent noncentral chi-square distributed random variables. Furthermore John (1960), John (1961) and Moran (1975) showed for multivariate Gaussian measurements that probabilities of correct allocation when using the linear discriminant function with unknown expectations but known and equal covariance matrices can be expressed as values of the cumulative distribution function of the doubly noncentral F-distribution.

A certain non classical linear model approach for deriving classification rules including as a special case the maximum likelihood rule for one-dimensional measurements has been developed in Krause and Richter (1994a, 1994b) for the case that the continuous overall sample distribution is elliptically contoured with possibly unknown expectations and variances and for arbitrary sample sizes in all three populations. The subsamples corresponding to different populations as well as the measurements within a given population are assumed to be uncorrelated. Note, however, that uncorrelatedness does not imply here independence, unless for Gaussian joint distributions.

If one assumes, however, independent subsamples in the non Gaussian case then the model changes considerably. Such a model has been studied in Richter (2002), but from an asymptotic point of view.

The geometric idea in Krause and Richter (1994a,1994b) for deriving the classification rule is a nearest neighbor one and is connected therefore with the statistical idea of equal variances. But, because of the flexible choice of the cut-off point, the decision rule can also be applied in the case of unequal variances. It is shown furthermore how exact probabilities of correct classification can be derived by measuring suitably defined sets in certain subspaces of the sample space with measures, the location- and form parameters of which are carefully chosen. To this end representation formulae for probabilities of correct classification are derived from a general geometric measure representation formula for Gaussian and spherical distributions in Richter (1985) and in Richter (1991,95), respectively.

This approach will be developed in Section 2 of the present paper to derive rather explicit formulae which express precisely the dependence of probabilities of correct classifications on the parameters of the underlying model for one-dimensional measurements.

Interrelations between exact probabilities of correct classifications and suitably chosen values of a doubly noncentral g-generalized F-distribution are considered in Section 3

for multivariate measurements.

The main results of the present paper are based on the general approach from geometric measure theory mentioned above. Basic tools of investigation from that area like the intersection percentage function (i.p.f.) for certain Borel sets and a geometric measure representation formula for spherical measures with density generating function (d.g.f.) g as well as related representation formulae for noncentral g-generalized chi-square, Student- and Fisher-distributions are presented and illustrated by examples in the Appendix. The Appendix gives special insight into the similarities but also into the differences between the measure theoretical necessities for classification and other statistical decisions, the risks of which are described with chi-square, Student- or Fisher-distributions.

Numerical properties of the new formulae for one-dimensional measurements are discussed in Section 4 and compared with results from the literature, including such one concerning the doubly noncentral F-distribution as, e.g., Price (1964), Mudholkar et al. (1976) and Chou et al. (1985). The numerical results for multivariate measurements show that the probability of correct classification increases if the dimension of the feature vector increases in the considered case of uncorrelated feature vectors.

Note that our decision rules work without restricting assumptions concerning the variances and expectations. The formulae for evaluating probabilities of correct classification depend therefore strongly on several parameters which can either assumed to be known or to be unknown and to be estimated in the latter case. The resulting estimated probabilities of correct classifications are discussed in Section 5. Notice that we do not use so called plug-in rules and are not looking for their error rates.

2 Explicit formulae for probabilities of correct classifications

Let an experimenter make n_1, n_2 and n_3 one-dimensional measurements of individuals from certain categories or populations Π_1, Π_2 and Π_3, respectively. Populations Π_1 and Π_2 are assumed to be distinguishable in the sense that their expectations μ_1 and μ_2 are different and with respect to population Π_3 it is assumed that its distribution coincides with either that of Π_1 or that of Π_2. One could imagine Π_3 to be a copy of either Π_1 or Π_2.

Let the overall sample vector $Y_{(n)} = ({Y_{n_1}^1}^T, {Y_{n_2}^2}^T, {Y_{n_3}^3}^T)^T$, where $Y_{n_k}^k = (Y_{k1}, \ldots, Y_{kn_k})^T$, $k = 1, 2, 3$, satisfy the sample model equation

$$Y_{(n)} = \begin{pmatrix} 1_{n_1} & 0_{n_1} & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & 0_{n_2} \\ 0_{n_3} & 0_{n_3} & 1_{n_3} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \mathcal{E}_{(n)},$$

where $(\mu_1, \mu_2)^T \in \mathbb{R}^2 \backslash \{(\nu_1, \nu_2)^T : \nu_1 = \nu_2\}$, $\mu_3 \in \{\mu_1, \mu_2\}$, $n = n_1 + n_2 + n_3$, and where the random error vector $\mathcal{E}_{(n)}$ has a centered elliptically contoured distribution having a

density and finite second order moments. Adapting the notation in Fang et al. (1990), pp. 31 and 46, we have thus assumed that

$$\mathcal{E}_{(n)} \sim \mathrm{EC}_n^d(\mu, \Sigma_i^{\mathcal{E}_{(n)}}; g)$$

with expectation vector $\mu = 0 \in \mathbb{R}^n$, block diagonal form matrix

$$\Sigma_i^{\mathcal{E}_{(n)}} = \begin{pmatrix} \sigma_1^2 I_{n_1} & & \\ & \sigma_2^2 I_{n_2} & \\ & & \sigma_i^2 I_{n_3} \end{pmatrix}, i = \begin{cases} 1 & \text{if } \mu_3 = \mu_1 \\ 2 & \text{if } \mu_3 = \mu_2 \end{cases},$$

and density function

$$\begin{aligned} p_i\left(y_{(n)}; g\right) &= C(n,g) \left|\Sigma_i^{\mathcal{E}_{(n)}}\right|^{-1/2} g\left((y_{(n)} - \mathbb{E}_i Y_{(n)})^T (\Sigma_i^{\mathcal{E}_{(n)}})^{-1} (y_{(n)} - \mathbb{E}_i Y_{(n)})\right) \\ &= \frac{C(n,g)}{\sigma_1^{n_1} \sigma_2^{n_2} \sigma_i^{n_3}} g\left(\sum_{i=1}^{3} \sigma_i^{-2} \sum_{j=1}^{n_i} (y_{ij} - \mu_i)^2\right), y_{(n)} \in \mathbb{R}^n \end{aligned}$$

where $C(n,g) = \Gamma(n/2)/(2\pi^{n/2} \int_0^\infty r^{n-1} g(r^2) dr)$ denotes a normalizing constant and g is a d.g.f. satisfying the condition

$$0 < \int_0^\infty r^{n-1} g(r^2)\, dr < \infty. \tag{1}$$

Thereby,

$$1_{n_i} = (1, \ldots, 1)^T \in \mathbb{R}^{n_i},\ 0_{n_i} = (0, \ldots, 0)^T \in \mathbb{R}^{n_i}$$

and I_{n_i} denotes the $n_i \times n_i$ - unit matrix.
The question which the investigator wants to answer is which of the decisions

$$D_{1/3} :\ \mu_3 = \mu_1 \qquad \text{or} \qquad D_{2/3} :\ \mu_3 = \mu_2 \tag{2}$$

would be more reasonable. $D_{i/3}$ means to allocate or classify the individuals from population Π_3 having joint measurement vector $Y_{n_3}^3$ to population $\Pi_i, i \in \{1,2\}$. Put

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

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

and define two linear subspaces of $\mathbb{R}^n$ which can be interpreted as allocation or classification spaces reflecting $D_{1/3}$ and $D_{2/3}$ in the sample space $\mathbb{R}^n$, respectively:

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

The union of these spaces can be considered as the model space $\mathfrak{M}$, i.e. the set of all possible values of the expectation of $Y_{(n)}$:

$$\mathfrak{M} = \mathrm{range}(\mathbb{E}_1 Y_{(n)}) \cup \mathrm{range}(\mathbb{E}_2 Y_{(n)}) = \mathfrak{M}_{1/3} \cup \mathfrak{M}_{2/3}.$$

Here, $\mathbb{E}_i$ means expectation if $\mu_3 = \mu_i$ holds true. Note that $\mathfrak{M}$ is not a linear subspace of $\mathbb{R}^n$ but is a subset of the three-dimensional linear space

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

which is called the extended model space. Recognize furthermore that $\mathfrak{M}$ is not included in any two-dimensional subspace of $\mathbb{R}^n$ and that

$$\mathfrak{M}_{1/3} \cap \mathfrak{M}_{2/3} = \mathcal{L}(1_n).$$

The sample vector $Y_{(n)}$ leaves the model space only because of random deviations. We assume that the deviations from the two allocation spaces $\mathfrak{M}_{1/3}$ and $\mathfrak{M}_{2/3}$ have equal size what means equal variances in the populations Π_1 and Π_2. We use therefore the euclidean distance to describe what is meant with deviations from the allocation spaces. Nevertheless, we can apply our decision rule to the case of unequal variances, too. We shall study then the changes of the probabilities of correct classifications in the sense of a robustness-sensitivity study.

The experimenter could intend to use one of the following three allocation or classification rules.

Allocation rule d_1 This rule leads to rejection of $D_{1/3}$ for values of

$$\frac{\|Y_{(n)} - \Pi_{\mathfrak{M}_{1/3}} Y_{(n)}\|}{\|Y_{(n)} - \Pi_{\mathfrak{M}_{2/3}} Y_{(n)}\|} = \frac{\sum_{i=1,3}\sum_{k=1}^{n_i}(Y_{ik} - \overline{Y}_{..}^{(1,3)})^2 + \sum_{k=1}^{n_2}(Y_{2k} - \overline{Y}_{.}^{(2)})^2}{\sum_{i=2,3}\sum_{k=1}^{n_i}(Y_{ik} - \overline{Y}_{..}^{(2,3)})^2 + \sum_{k=1}^{n_1}(Y_{1k} - \overline{Y}_{.}^{(1)})^2}$$

larger than or equal to a certain positive real number c where Π means orthogonal projection, $\overline{Y}_{.}^{(i)} = \frac{1}{n_i}\sum_{k=1}^{n_i} Y_{ik}$ and $\overline{Y}_{..}^{(i,j)} = \frac{1}{n_+ n_j}(\sum_{k=1}^{n_i} Y_{ik} + \sum_{k=1}^{n_j} Y_{jk}), i \neq j$. It has been proved in Richter (1999) that d_1 is the likelihood ratio rule for the case that all parameters μ_1, μ_2, Σ_1 and Σ_2 are unknown but $\Sigma_1 = \Sigma_2$. Note that d_1 is called maximum likelihood rule if $c = 1$. In the case of a Gaussian error vector $\mathcal{E}_{(n)}$ and $n_3 = 1$ the maximum likelihood rule has been studied in Das Gupta (1965).

Allocation rule d_2 Classification rule d_2 rejects $D_{1/3}$ for values of

$$\frac{\text{dist}\left(\Pi_{\widetilde{\mathfrak{M}}} Y_{(n)}, \mathfrak{M}_{1/3}\right)}{\text{dist}\left(\Pi_{\widetilde{\mathfrak{M}}} Y_{(n)}, \mathfrak{M}_{2/3}\right)} = \frac{n_1(\overline{Y}_{.}^{(1)} - \overline{Y}_{..}^{(1,3)})^2 + n_3(\overline{Y}_{.}^{(3)} - \overline{Y}_{..}^{(1,3)})^2}{n_2(\overline{Y}_{.}^{(2)} - \overline{Y}_{..}^{(2,3)})^2 + n_3(\overline{Y}_{.}^{(3)} - \overline{Y}_{..}^{(2,3)})^2} \tag{3}$$

larger than or equal to c. Here,

$$\text{dist}\left(z, \mathfrak{M}_{i/3}\right) = \|z - \Pi_{\mathfrak{M}_{i/3}} z\|$$

means euclidean distance from the point $z \in \widetilde{\mathfrak{M}}$ to the space $\mathfrak{M}_{i/3}$.

Note that d_1 and d_2 coincide if $c = 1$. Furthermore, d_1 can be looked as a sample space based classification rule while d_2 can be viewed as an extended model space based rule.

Allocation rule d_3 The decision space based rule d_3 is defined as leading to a rejection of $D_{1/3}$ for values of

$$\frac{\|\Pi_{1^{-0+}} Y_{(n)}\|}{\|\Pi_{1^{0-+}} Y_{(n)}\|} = \sqrt{\frac{n_1(n_2+n_3)}{n_2(n_1+n_3)}}\frac{|\overline{Y}_{.}^{(3)} - \overline{Y}_{.}^{(1)}|}{|\overline{Y}_{.}^{(3)} - \overline{Y}_{.}^{(2)}|} \tag{4}$$

larger than or equal to c, where

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

The vectors 1^{0-+} and 1^{-0+} build a basis of the two-dimensional decision space

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

Recognize that d_3 coincides with decision rule d_2 and that the least squares estimates $\overline{Y}_{\cdot}^{(3)} - \overline{Y}_{\cdot}^{(i)}$ for the effects $\mu_3 - \mu_i$, $i = 1, 2$ satify the equations

$$\Pi_{1^{-0+}} Y_{(n)} = 1^{-0+} \frac{n_1 n_3}{n_1 + n_3} (\overline{Y}_{\cdot}^{(3)} - \overline{Y}_{\cdot}^{(1)}),\ \Pi_{1^{0-+}} Y_{(n)} = 1^{0-+} \frac{n_2 n_3}{n_2 + n_3} (\overline{Y}_{\cdot}^{(3)} - \overline{Y}_{\cdot}^{(2)}).$$

Note further that the corresponding equations

$$\Pi_{1^{-0+}} \mathbb{E}_i Y_{(n)} = 1^{-0+} \frac{n_1 n_3}{n_1 + n_3} (\mu_i - \mu_1),\ \Pi_{1^{0-+}} \mathbb{E}_i Y_{(n)} = 1^{0-+} \frac{n_2 n_3}{n_2 + n_3} (\mu_i - \mu_2)$$

motivate the names effect or decision or action space because they reflect possible changes of the model parameters $\mu_i - \mu_j$, $j \in \{1, 2\}$, as well as deviations from the situations $\mu_i - \mu_j = 0$ and they allow a comparison of the quantities $\mu_3 - \mu_1$ and $\mu_3 - \mu_2$ which builds the background for the allocation rule d_3.

While the maximum likelihood rule occurs frequently in the literature, the equivalent representation d_3 suggests a new way for dealing with the risks of this decision. This way has been started in Krause and Richter (1994a) for $c = \sigma_1/\sigma_2$ and will be developed here for arbitrary positive cut-off point c. It leads, e.g., to the well interpretable fact that the probability of correct classification considered here depends on the four parameters

$$p_{1,1} = \frac{n_3}{\sigma_1^2}(\mu_1 - \mu_2)^2, \qquad p_{2,1} = \frac{\sigma_2^2}{\sigma_1^2}, \qquad p_{3,1} = \frac{n_3}{n_1}, \qquad p_{4,1} = \frac{n_3}{n_2} \tag{5}$$

if $\mu_3 = \mu_1$ is true and

$$p_{1,2} = \frac{n_3}{\sigma_2^2}(\mu_1 - \mu_2)^2, \qquad p_{2,2} = \frac{\sigma_1^2}{\sigma_2^2}, \qquad p_{3,2} = \frac{n_3}{n_2}, \qquad p_{4,2} = \frac{n_3}{n_1} \tag{6}$$

if $\mu_3 = \mu_2$ is true. In both cases, the probabilities of correct classification depend additionally on the cut-off point c and a certain function $\widetilde{g}$. Let

$$P_i(CC_i)(c) = P_i(CC_i)(c; \widetilde{g}) = P_i(CC_i)(c; p_{1,i}, p_{2,i}, p_{3,i}, p_{4,i}; \widetilde{g})$$

denote the probability of correct classification into the ith population, i.e. the probability of making decision $D_{i/3}$ if $D_{i/3}$ is actually correct. Concerning the function $\widetilde{g}$ in the definition of $P_i(CC_i)(c; \widetilde{g})$ recall that all two-dimensional marginal distributions of a continous elliptically contoured distribution follow a EC_2^d-distribution with a certain d.g.f. The latter one may be well defined up to a norming constant. The function $\widetilde{g}$

from $P_i(CC_i)(c;\widetilde{g})$ is such a d.g.f. of a two-dimensional marginal distribution which has a normalizing constant being equal to unity. We shall call therefore $\widetilde{g}$ a normalized d.g.f. or a density generator (d.g.). Finally, note that the parameters in (5) and (6) are defined in such a way, that we can derive below unified explicit formulae for $P_i(CC_i)(c)$, $i = 1, 2$.
We restrict our consideration for a moment to the Gaussian case $g = g^G$, where $g^G(r) = \mathrm{e}^{-r/2}$, $r > 0$. It has been shown in Ittrich et al. (2000) that the corresponding d.g. $\widetilde{g}$ in this two-dimensional case is $\widetilde{g} = \widetilde{g}^G$ with $\widetilde{g}^G(r) = (2\pi)^{-1}\mathrm{e}^{-r/2}$, $r > 0$.

Theorem 1 *If in the Gaussian case decision $D_{i/3}$ would be correct then the probability $P_i(CC_i)(c)$ of correct classification when using allocation rule d_3 allows the representation*

$$P_i(CC_i)(c) = 1 + \Big(\frac{1}{\pi}\arccos\frac{m_i^2-1}{\sqrt{h_i}} - 1\Big)\,\mathrm{e}^{-\|q_i\|^2/2} - \frac{1}{\pi}\int_{s_{min,i}}^{\|q_i\|} r\mathrm{e}^{-r^2/2}\arccos\frac{s_{min,i}}{r}dr + \frac{(-1)^{I_i}}{\pi}\int_{s_{max,i}}^{\|q_i\|} r\mathrm{e}^{-r^2/2}\arccos\frac{s_{max,i}}{r}\,dr \quad (7)$$

for $i \in \{1, 2\}$, where

$$m_1 = \frac{1}{c}\frac{\sqrt{1+p_{4,1}}}{\sqrt{1+p_{2,1}p_{4,1}}}, \quad m_2 = c\frac{\sqrt{1+p_{4,2}}}{\sqrt{1+p_{2,2}p_{4,2}}}, \quad h_i = (1+m_i^2)^2 - (2m_i\rho_i)^2,$$

$$\|q_i\| = \frac{a_i}{\sqrt{1-\rho_i^2}}, \qquad a_i = \frac{\sqrt{p_{1,i}}}{\sqrt{1+p_{2,i}p_{4,i}}}, \qquad \rho_i = \frac{1}{\sqrt{1+p_{3,i}}\sqrt{1+p_{2,i}p_{4,i}}},$$

$$s_{min,i} = \frac{a_i}{\sqrt{m_i^2+2m_i\rho_i+1}}, \qquad s_{max,i} = \frac{a_i}{\sqrt{m_i^2-2m_i\rho_i+1}},$$

$$I_1 = \begin{cases} 1 & \text{if } c^2 < (1+p_{3,1})(1+p_{4,1}) \\ 0 & \text{otherwise} \end{cases}, \qquad I_2 = \begin{cases} 1 & \text{if } 1/c^2 < (1+p_{3,2})(1+p_{4,2}) \\ 0 & \text{otherwise} \end{cases}.$$

Remark 1 The quantities I_i, m_i, $s_{min,i}$, $s_{max,i}$ and h_i occuring in Theorem 1 depend explicitely on the cut-off point c in the decision rule d_3 while the quantities a_i, ϱ_i and q_i do not depend on c.
Remark 2 Despite the fact that Theorem 1 offers two formulae for the probabilities of correct classification, for numerical computations in fact one has to implement only one formula and can use then the general relation

$$P_2(CC_2)(c) = P_1(CC_1)\Big(\frac{1}{c}; \frac{p_{1,1}}{p_{2,1}}, \frac{1}{p_{2,1}}, p_{4,1}, p_{3,1}; \widetilde{g}\Big).$$

Remark 3 One can read as usual $P(CC_i) := P_i(CC_i)(c)$ as "the probability of correct classification if there holds $\mu_3 = \mu_i$ ", $i \in \{1, 2\}$, although the symbols $P(CC_i)$ are formally not quite correct.

Remark 4 If $n_1 = n_2$, i.e. if the sample sizes are somehow balanced, then $p_{3,i} = p_{4,i}$. Hence, the number of parameters which influence the probabilities of correct classifications is reduced to three.

Before proving Theorem 1 let us continue now with the elliptically contoured case, i.e. with arbitrary d.g.f. g satisfying assumption (1). Recall that if, e.g., $g = g_P$ with $g_P(r) = (1 + r/m)^{-M}, r > 0, m > 0, M > 1/2$ is the Pearson type VII d.g.f. then the corresponding d.g. $\widetilde{g} = \widetilde{g}_P$ is

$$\widetilde{g}_P(r) = \frac{\Gamma(M)}{\pi m \Gamma(M - 1/2)} \left(1 + \frac{r}{m}\right)^{-M}, r > 0. \tag{8}$$

Theorem 2 *If decision $D_{i/3}$ would be correct then the probability of correct classification into the population Π_i allows the integral representation*

$$P_i(CC_i)(c) = \int_0^\infty r\, \widetilde{g}(r^2)\, \mathcal{F}_2(CC_i^{**}, r)\, dr / \int_0^\infty r\, \widetilde{g}(r^2)\, dr,\ i = 1, 2 \tag{9}$$

where

$$\mathcal{F}_2(CC_i^{**}, r) = \begin{cases} 1 & \text{if } r \le s_{min,i}, \\ 1 - \frac{1}{\pi} \arccos \frac{s_{min,i}}{r} & \text{if } s_{min,i} < r \le s_{max,i}, \\ 1 - \frac{1}{\pi} \arccos \frac{s_{min,i}}{r} + \frac{1}{\pi}(-1)^{I_i} \arccos \frac{s_{max,i}}{r} & \text{if } s_{max,i} < r \le \|q_i\|, \\ \frac{1}{\pi} \arccos \frac{m_i^2 - 1}{\sqrt{h_i}} & \text{if } \|q_i\| < r, \end{cases} \tag{10}$$

with m_i, ϱ_i, h_i, a_i, $\|q_i\|$, $s_{min,i}$, $s_{max,i}$, I_i being the same as in Theorem 1 and

$$\widetilde{g}(u) = \frac{\pi^{n/2-1}}{\Gamma(n/2 - 1)} \int_u^\infty (y - u)^{n/2-2} C(n, g) g(y)\, dy. \tag{11}$$

The following consideration will be based on a geometric measure representation, developed in Richter(1985, 1991, 1995a) and sketched in the Appendix. An important quantity from this representation is the intersection percentage function.

Definition 1 Denote by ω the uniform probability distribution on the unit sphere $S_n(1) = \{x \in \mathbb{R}^n : \|x\| = 1\}$ with respect to the euclidean norm in $\mathbb{R}^n$. Then

$$\mathcal{F}_n(A, r) = \omega\left(r^{-1}A \cap S_n(1)\right),\ r > 0 \tag{12}$$

is called *intersection percentage function (i.p.f.)* for the Borel set A.

Proof of Theorem 1 The n-dimensional problem of evaluating probabilities of correct classification has been reduced in Krause and Richter (1994a), formula (16), to a two-dimensional problem concerning a certain Gaussian random vector $Z_{(2)}$ and certain two-dimensional Borel sets CC_i^* such that: $P_i(CC_i) = P^{Z_{(2)}}(CC_i^*)$, $i = 1, 2$. The process of transforming the two-dimensional random vector $Z_{(2)}$ into a standardized

Gaussian vector was described, too. Simultaneous transformations of the areas CC_i^*, $i = 1, 2$, lead to well defined "double" cones, i.e. two-sided cones in the plane which were denoted by CC_i^{**}. Note that the cones

$$CC_1^{**} = \{ \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \mathbb{R}^2 : \frac{n_2+n_3}{n_2 n_3} \big(\frac{1}{2}\sqrt{2(1+\varrho_1)} w_1 + \frac{1}{2}\sqrt{2(1-\varrho_1)} w_2 \big)^2 \tag{13}$$

$$< \big(\frac{1}{n_2} + \frac{\sigma_1^2}{\sigma_2^2 n_3} \big) \big(\frac{1}{2}\sqrt{2(1+\varrho_1)} w_1 - \frac{1}{2}\sqrt{2(1-\varrho_1)} w_2 + \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_3}}} \big)^2 \},$$

$$CC_2^{**} = \{ \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \mathbb{R}^2 : \big(\frac{1}{n_1} + \frac{\sigma_2^2}{\sigma_1^2 n_3} \big) \big(\frac{1}{2}\sqrt{2(1+\varrho_2)} w_1 + \frac{1}{2}\sqrt{2(1-\varrho_2)} w_2 + \frac{\mu_2 - \mu_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_3}}} \big)^2$$

$$> \frac{n_1+n_3}{n_1 n_3} \big(\frac{1}{2}\sqrt{2(1+\varrho_2)} w_1 - \frac{1}{2}\sqrt{2(1-\varrho_2)} w_2 \big)^2 \} \tag{14}$$

depend on c through the ratio σ_1/σ_2. The resulting formulae
$P_i(CC_i) = \int_0^\infty r \exp\{-r^2/2\} \mathcal{F}_2(CC_i^{**}, r)\, dr$ and

$$\mathcal{F}_2(CC_i^{**}, r) = \begin{cases} 1 & \text{if } r \le s_{min,i} \\ 1 - \frac{1}{\pi} \arccos \frac{s_{min,i}}{r} & \text{if } s_{min,i} < r \le s_{max,i} \\ 1 - \frac{1}{\pi} \arccos \frac{s_{min,i}}{r} - \frac{1}{\pi} \arccos \frac{s_{max,i}}{r} & \text{if } (s_{max,i} < r \le \|q_i\|) \wedge \\ & \quad ((l_{1,i}, q_i)(l_{2,i}, q_i) > 0) \\ 1 - \frac{1}{\pi} \arccos \frac{s_{min,i}}{r} + \frac{1}{\pi} \arccos \frac{s_{max,i}}{r} & \text{if } (s_{max,i} < r \le \|q_i\|) \wedge \\ & \quad ((l_{1,i}, q_i)(l_{2,i}, q_i) \le 0) \\ \frac{1}{\pi} \arccos \frac{(l_{1,i} l_{2,i})}{\|l_{1,i}\| \|l_{2,i}\|} & \text{if } \|q_i\| < r \end{cases} \tag{15}$$

from the above mentioned paper allow more explicit descriptions of the probabilities of correct classification which will be deduced now. While doing this we drop from now on the assumption $c = \sigma_1/\sigma_2$ from the former paper. Here, $l_{1,i}$ and $l_{2,i}$ are directional vectors of the boundary lines $g_{j,i}$ of the double cone CC_i^{**} and the endpoint of the local vector q_i coincides with the vertex of the double cone:

$$g_{j,i} : \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = t_j l_{j,i} + q_i,\ t_j \in \mathbb{R},\ j \in \{1, 2\}, \tag{16}$$

with

$$l_{1,1} = \begin{pmatrix} \frac{1+m_1}{\sqrt{1+\varrho_1}} \\ \frac{1-m_1}{\sqrt{1-\varrho_1}} \end{pmatrix}, l_{2,1} = \begin{pmatrix} \frac{m_1-1}{\sqrt{1+\varrho_1}} \\ -\frac{m_1+1}{\sqrt{1-\varrho_1}} \end{pmatrix}, q_1 = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_2^2/n_2 + \sigma_1^2/n_3}} \frac{1}{\sqrt{2}} \begin{pmatrix} -\frac{1}{\sqrt{1+\varrho_1}} \\ \frac{1}{\sqrt{1-\varrho_1}} \end{pmatrix}, \tag{17}$$

$$l_{1,2} = \begin{pmatrix} \frac{m_2+1}{\sqrt{1+\varrho_2}} \\ -\frac{m_2-1}{\sqrt{1-\varrho_2}} \end{pmatrix}, l_{2,2} = \begin{pmatrix} \frac{m_2-1}{\sqrt{1+\varrho_2}} \\ -\frac{m_2+1}{\sqrt{1-\varrho_2}} \end{pmatrix}, q_2 = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_3}} \frac{1}{\sqrt{2}} \begin{pmatrix} -\frac{1}{\sqrt{1+\varrho_2}} \\ \frac{1}{\sqrt{1-\varrho_2}} \end{pmatrix}, \tag{18}$$

$$m_1 = \frac{1}{c} \frac{\sqrt{n_2+n_3}}{\sqrt{n_2 + \frac{\sigma_2^2}{\sigma_1^2} n_3}},\ m_2 = \frac{c\sqrt{n_1+n_3}}{\sqrt{n_1 + \frac{\sigma_1^2}{\sigma_2^2} n_3}},\ \varrho_i = \frac{\sigma_i^2 \sqrt{n_1 n_2}}{\sqrt{(n_3\sigma_1^2 + n_1\sigma_i^2)(n_3\sigma_2^2 + n_2\sigma_i^2)}},\ i = 1, 2.$$

The distances from the origin to the lines $g_{j,i}$ are

$$\mathrm{dist}(0, g_{1,i}) = \frac{a_i}{\sqrt{m_i^2 - 2m_i\varrho_i + 1}} = s_{min,i}, \mathrm{dist}(0, g_{2,i}) = \frac{a_i}{\sqrt{m_i^2 + 2m_i\varrho_i + 1}} = s_{max,i}.$$

Finally note that $\|q_i\| = a_i/\sqrt{1-\varrho_i^2}$ and and that the condition $(l_{1,i}, q_i)\,(l_{2,i}, q_i) > 0$ is equivalent to $c^2 < \left(1+\frac{n_3}{n_1}\right)\left(1+\frac{n_3}{n_2}\right)$ and $\frac{1}{c^2} < \left(1+\frac{n_3}{n_1}\right)\left(1+\frac{n_3}{n_2}\right)$ if $i=1$ and $i=2$, respectively. The proof is finished by introducing the parameters from (5) and (6). □

Proof of Theorem 2 The matrix

$$B^T := \begin{pmatrix} -\frac{1}{n_1} & \cdots & -\frac{1}{n_1} & 0 & \cdots & 0 & \frac{1}{n_3} & \cdots & \frac{1}{n_3} \\ 0 & \cdots & 0 & -\frac{1}{n_2} & \cdots & -\frac{1}{n_2} & \frac{1}{n_3} & \cdots & \frac{1}{n_3} \end{pmatrix}$$

transforms the overall sample vector $Y_{(n)}$ into the two dimensional vector $Z_{(2)} = B^T Y_{(n)}$. Our assumption $Y_{(n)} \sim \mathrm{EC}_n^d(\mathbb{E}_i Y_{(n)}, \Sigma_i^{Y_{(n)}}, g)$ corresponds to the assumption

$$Y_{(n)} \sim \mathrm{EC}_n^c(\mathbb{E}_i Y_{(n)}, \Sigma_i^{Y_{(n)}}, \phi) \tag{19}$$

for some characteristic generating function ϕ with the property $\phi(t^T t) = \varphi_{X_{(n)}}(t)$, $t \in \mathbb{R}^n$, where $\varphi_{X_{(n)}}$ denotes the characteristic function of $X_{(n)} := (\Sigma_i^{Y_{(n)}})^{-1/2}(Y_{(n)} - \mathbb{E}_i Y_{(n)})$. Using Theorem 2.16 in Fang et al. (1990) one gets

$$Z_{(2)} \sim \mathrm{EC}_2^c(\nu_i, \Gamma_i, \phi) \tag{20}$$

with the same characteristic generating function ϕ as above and with ν_i and Γ_i given in Krause and Richter (1994a). Now we want to determine the d.g.f.of the reduced statistic $Z_{(2)}$. Since a standardization doesn't change neither the characteristic generating function nor the d.g.f., $X_{(n)}$ follows the n-dimensional spherically symmetric distribution with characteristic and density generating functions ϕ and g, respectively. Let us consider now the two-dimensional random vector

$$X^{(1)} = \widetilde{B}^T X_{(n)} = \begin{pmatrix} 1 & 0 & 0_{n-2}^T \\ 0 & 1 & 0_{n-2}^T \end{pmatrix} X_{(n)}.$$

Due to Theorems 2.16 and 2.10 in Fang et al. (1990), its probability law is the two-dimensional spherically symmetric marginal law with characteristic and density generating functions ϕ and $g_{n,2}$, respectively, where according to formula (2.23) in Fang et al. (1990)

$$g_{n,2}(u) = \frac{\pi^{n/2-1}}{\Gamma(n/2-1)} \int_u^\infty (y-u)^{n/2-2} C(n,g) g(y)\, dy.$$

Note that $g_{n,2}$ has the normalization property $\int_{\mathbb{R}^2} g_{n,2}(\|u\|^2) du = 1$. As a consequence, $Y^{(1)} = \nu_i + \Gamma_i^{1/2} X^{(1)}$ satisfies both $Y^{(1)} \sim EC_2^c(\nu_i, \Gamma_i, \phi)$ and $Y^{(1)} \sim EC_2^d(\nu_i, \Gamma_i, g_{n,2})$. From this and (20) it follows

$$Z_{(2)} \sim \mathrm{EC}_2^d(\nu_i, \Gamma_i, g_{n,2}).$$

As in the proof of Theorem 1, it turns out that

$$P(CC_i) = P^{Z_{(2)}}(CC_i^*) = \mathrm{EC}_2^d(\nu_i, \Gamma_i, g_{n,2})(CC_i^*).$$

Since $W_{(2)} := \Gamma_i^{-1/2}(Z_{(2)} - \nu_i) \sim \mathrm{EC}_2^d(0_2, I_2, g_{n,2})$ it follows

$$P(CC_i) = \mathrm{EC}_2^d(0_2, I_2, g_{n,2})(CC_i^{**})$$

with the intersection-percentage function $\mathcal{F}_2(CC_i^{**}, r)$ being the same as in the Gaussian case. Now, we get the assertion of Theorem 2 from formula (32) below. □

Example 1 The normalizing constant of the n-dimensional Pearson type VII distribution is $C(n, g_P) = (\pi m)^{-n/2}\Gamma(M)/\Gamma(M - n/2)$. One can explicitely compute certain parts from the representation formula (9) and gets:

$$\begin{aligned} P_i(CC_i)(c) &= 1 + \big(\frac{1}{\pi}\arccos\frac{m_i^2 - 1}{\sqrt{h_i}} - 1\big)\big(1 + \frac{\|q_i\|^2}{m}\big)^{-M+\frac{n}{2}} \\ &\quad -\frac{2M-n}{\pi m}\int_{s_{min,i}}^{\|q_i\|} \arccos\frac{s_{min,i}}{r}\, r\,\big(1 + \frac{r^2}{m}\big)^{-M+\frac{n}{2}-1}\, dr \qquad (21) \\ &\quad +\frac{(-1)^{I_i}(2M-n)}{\pi m}\int_{s_{max,i}}^{\|q_i\|} \arccos\frac{s_{max,i}}{r}\, r\,\big(1 + \frac{r^2}{m}\big)^{-M+\frac{n}{2}-1}\, dr. \end{aligned}$$

3 Classification probabilities and doubly non central generalized Fisher distributions

Let us turn now to the case $p \geq 1$ of multivariate measurements. Let the sample vector

$$Y_{(np)} = \big(Y_{(n_1p)}^{1T}, Y_{(n_2p)}^{2T}, Y_{(n_3p)}^{3T}\big)^T$$

where

$$Y_{(n_ip)}^i = \big(Y_{i1}^T, \ldots, Y_{in_i}^T\big)^T,\ Y_{ij} = (Y_{1ij}, \ldots, Y_{pij})^T,\ j = 1, \ldots, n_i,\ i = 1, 2, 3,$$

satisfay the sample model equation

$$Y_{(np)} = \sum_{l=1}^{p}\sum_{i=1}^{3} 1_{li}\mu_{li} + \mathcal{E}_{(np)}.$$

It is essentially based upon the orthogonal vectors $1_{li} \in \mathbb{R}^{np}$ defined as

$$\begin{aligned} &1_{11} = \big(1, 0_{p-1}^T, \ldots, 1, 0_{p-1}^T, 0_{(n_2+n_3)p}^T\big)^T, \ldots, 1_{p1} = \big(0_{p-1}^T, 1, \ldots, 0_{p-1}^T, 1, 0_{(n_2+n_3)p}^T\big)^T, \\ &\vdots \\ &1_{13} = \big(0_{(n_1+n_2)p}^T, 1, 0_{p-1}^T, \ldots, 1, 0_{p-1}^T\big)^T, \ldots, 1_{p3} = \big(0_{(n_1+n_2)p}^T, 0_{p-1}^T, 1, \ldots, 0_{p-1}^T, 1\big)^T. \end{aligned}$$

The expectation vectors in the three populations are

$$\mu_i = (\mu_{1i}, \ldots, \mu_{pi})^T,\ i = 1, 2, 3, \text{ and } \mu_3 \in \{\mu_1, \mu_2\}$$

and the overall random error vector $\mathcal{E}_{(np)}$ is assumed to be distributed according to an elliptically contoured distribution $\Phi_{0_{np}, \Sigma_i^{\mathcal{E}_{(np)}}; g}$ with

$$\Sigma_i^{\mathcal{E}_{(np)}} = \begin{pmatrix} \mathrm{I}_{n_1} \otimes \Sigma_1 & 0_{pn_1, pn_2} & 0_{pn_1, pn_3} \\ 0_{pn_2, pn_1} & \mathrm{I}_{n_2} \otimes \Sigma_2 & 0_{pn_2, pn_3} \\ 0_{pn_3, pn_1} & 0_{pn_3, pn_2} & \mathrm{I}_{n_3} \otimes \Sigma_i \end{pmatrix},\ i \in \{1, 2\}.$$

Here, $0_{i,j}$ and "$\otimes$" denote the $i \times j$-zero matrix and the direct product, respectively. Put

$$1_l^{-0+} = -1_{l1}/n_1 + 1_{l3}/n_3,\ l = 1, \ldots, p.$$

Because of the projection property

$$\Pi_{1_l^{-0+}} \mathbb{E} Y_{(np)} = 1_l^{-0+} \frac{n_1 n_3}{n_1 + n_3} (\mu_{il} - \mu_{1l}),\ l = 1, \ldots, p$$

the vectors 1_l^{-0+} are suitable for describing the directions in the sample space $\mathbb{R}^{np}$ into which the quantities $\mu_{l3} - \mu_{l1}$ deviate from the zero vector if one of the partial assumptions $\mu_{l3} = \mu_{l1}, l = 1, \ldots, p$ would not be correct. The linear space $\mathcal{D}^{-0+}$ having the orthogonal basis $\{1_1^{-0+}, \ldots, 1_p^{-0+}\}$ can therefore be considered as the action space for the vector $\mu_3 - \mu_1$ or as the decision space corresponding to $D_{1/3}$. The action space $\mathcal{D}^{0-+}$ for the vector $\mu_3 - \mu_2$ which can also be interpreted as the decision space corresponding to $D_{2/3}$ is spanned up by the orthogonal vectors

$$1_l^{0-+} = -1_{l2}/n_2 + 1_{l3}/n_3,\ l = 1, \ldots, p.$$

Define the $2p$-dimensional decision space by

$$\mathcal{D} = \mathcal{L}\left(1_1^{-0+}, \ldots, 1_p^{-0+}, 1_1^{0-+}, \ldots, 1_p^{0-+}\right) = \mathcal{L}\left(\mathcal{D}^{-0+} \cup \mathcal{D}^{0-+}\right). \tag{22}$$

Let an allocation rule $d^{(p)} \,|\, \mathbb{R}^{np} \longrightarrow \{1, 2\}$ being defined as leading to decision $D_{2/3}$, i.e. as taking the value 2 if

$$\left\|\Pi_{\mathcal{D}^{-0+}} Y_{(np)}\right\| / \left\|\Pi_{\mathcal{D}^{0-+}} Y_{(np)}\right\| \geq c \tag{23}$$

for a suitably chosen cut-off point c. Note that if $p = 1$ then (23) is the same as (4), hence $d^{(1)}$ coincides with the above considered allocation rule d_3. Each of the sets

$$\left\{y \in \mathbb{R}^{np} : \frac{\left\|\Pi_{\mathcal{D}^{-0+}} y_{(np)}\right\|}{\left\|\Pi_{\mathcal{D}^{0-+}} y_{(np)}\right\|} < c\right\},\ \left\{y \in \mathbb{R}^{np} : \frac{\left\|\Pi_{\mathcal{D}^{-0+}} y_{(np)}\right\|}{\left\|\Pi_{\mathcal{D}^{0-+}} y_{(np)}\right\|} \geq c\right\}$$

is a cone in $\mathbb{R}^{np}$ with vertex in the origin. In the following it will be shown, that probabilities of correct classification can be equivalently determined by measuring cones in the $2p$-dimensional Euclidean space with suitably chosen non centered elliptically

contoured measures instead of measuring cones in the np-dimensional space. The decision is actually made in the $2p$-dimensional decision space $\mathcal{D}$, defined in (22). This circumstance is reflected by the fact that the statistic from the allocation rule can be reformulated in a data reduced form as

$$\frac{\left\|\Pi_{\mathcal{D}^{-0+}}Y_{(np)}\right\|^2}{\left\|\Pi_{\mathcal{D}^{0-+}}Y_{(np)}\right\|^2}=\frac{\left\|\sum_{l=1}^{p}1_l^{-0+}\frac{n_1n_3}{n_1+n_3}(\overline{Y_{l3\bullet}}-\overline{Y_{l1\bullet}})\right\|^2}{\left\|\sum_{l=1}^{p}1_l^{0-+}\frac{n_2n_3}{n_2+n_3}(\overline{Y_{l3\bullet}}-\overline{Y_{l2\bullet}})\right\|^2}=\frac{\frac{n_1}{n_1+n_3}\left\|(\overline{Y_{3\bullet}}-\overline{Y_{1\bullet}})\right\|^2}{\frac{n_2}{n_2+n_3}\left\|(\overline{Y_{3\bullet}}-\overline{Y_{2\bullet}})\right\|^2} \tag{24}$$

where

$$\overline{Y_{i\bullet}}:=\left(\overline{Y_{1i\bullet}},\dots\overline{Y_{pi\bullet}}\right)^T,\ \overline{Y_{li\bullet}}=\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{lij},\ i=1,2,3,\ l=1,\dots,p.$$

The reduced statistic

$$Z_{(2p)}:=\begin{pmatrix}\overline{Y_{3\bullet}}-\overline{Y_{1\bullet}}\\ \overline{Y_{3\bullet}}-\overline{Y_{2\bullet}}\end{pmatrix} \tag{25}$$

satisfies the equation $Z_{(2p)}=B^TY_{(np)}$ with

$$B^T:=\begin{pmatrix}-\frac{1}{n_1}I_p & \cdots & -\frac{1}{n_1}I_p & 0_{p,pn_2} & & & \frac{1}{n_3}I_p & \cdots & \frac{1}{n_3}I_p\\ 0_{p,pn_1} & & & -\frac{1}{n_2}I_p & \cdots & -\frac{1}{n_2}I_p & \frac{1}{n_3}I_p & \cdots & \frac{1}{n_3}I_p\end{pmatrix}\in\mathbb{R}^{2p\times np}.$$

Due to Theorems 2.16 and 2.10 in Fang et al. (1990),

$$Z_{(2p)}\overset{d}{\sim}\mathrm{EC}_{2p}^{d}\left(B^T(1_{l1}\mu_{l1}+1_{l2}\mu_{l2}+1_{l3}\mu_{li}),\ B^T\Sigma_i^{\mathcal{E}_{(np)}}B;g_{np,2p}\right), \tag{26}$$

where according to formula (2.23) of Fang et al. (1990), the d.g.f. is given by

$$g_{np,2p}(u)=\frac{\pi^{p(n/2-1)}}{\Gamma(p(n/2-1))}\int_u^\infty(y-u)^{p(n/2-1)-1}C(np,g)g(y)\,dy. \tag{27}$$

It follows from (24) and (25) that the areas which are to be measured with the $2p$-dimensional measures from (26) are

$$CC_1^*=\left\{z\in\mathbb{R}^{2p}:\frac{n_1}{n_1+n_3}\left\|(z_1,\dots,z_p)^T\right\|^2<c^2\frac{n_2}{n_2+n_3}\left\|(z_{p+1},\dots,z_{2p})^T\right\|^2\right\}$$

and $CC_2^*=\mathbb{R}^{2p}\setminus CC_1^*$ for $i=1$ and 2, respectively. These areas are cones in the $2p$-dimensional Euclidean space and can be rewritten in terms of quadratic forms as $CC_i^*=\left\{z\in\mathbb{R}^{2p}:z^TA_i^*z<0\right\}$ with indefinite form matrices

$$A_i^*=\begin{pmatrix}(-1)^{i+1}\frac{n_1}{n_1+n_3}I_p & 0_{p,p}\\ 0_{p,p} & (-1)^ic^2\frac{n_2}{n_2+n_3}I_p\end{pmatrix},\ i=1,2.$$

Lemma 1 *The probability of correct classification if $D_{i/3}$would be correct and when using the allocation rule $d^{(p)}$ allows in the case $\Sigma_1=\sigma_1^2I_p$, $\Sigma_2=\sigma_2^2I_p$ the representation:*

$$P_i\left(CC_i\right)(c)=\Phi_{0_{2p},I_{2p};g_{np,2p}}(CC_i^{**}(c)), \tag{28}$$

i = 1, 2, *where the sets*

$$CC_i^{**}(c) = CC_i^{**} = \left\{ w_{(2p)} \in \mathbb{R}^{2p} : (w_{(2p)} + \nu_i^{**})^T \Lambda_i \, (w_{(2p)} + \nu_i^{**}) < 0 \right\} \qquad (29)$$

are defined with the help of the form matrices

$$\Lambda_i = \begin{pmatrix} (1 - \frac{1}{m_i^2} + \frac{\sqrt{h_i}}{m_i^2}) \, I_p & 0_{p,p} \\ 0_{p,p} & (1 - \frac{1}{m_i^2} - \frac{\sqrt{h_i}}{m_i^2}) \, I_p \end{pmatrix}$$

and the shift vectors $\nu_i^{**} = \left(\frac{n_3}{2\sigma_i^2 N_i (1 + p_{2,i} p_{4,i})} \right)^{1/2}$

$$\times \begin{pmatrix} \left((-1)^i \frac{1}{\sqrt{1+\varrho_i}} (\sqrt{h_i} + m_i^2 \varrho_i - \varrho_i) + (-1)^{i+1} \sqrt{1+\varrho_i} (1 + m_i^2) \right) (\mu_2 - \mu_1) \\ \left(-\sqrt{1-\varrho_i} (1 + m_i^2) + (-\sqrt{h_i} - m_i^2 \varrho_i + \varrho_i) \frac{1}{\sqrt{1-\varrho_i}} \right) (\mu_2 - \mu_1) \end{pmatrix}.$$

Here, $N_i = (1 - \varrho_i^2)(1 + m_i^2)^2 + (\sqrt{h_i} + m_i^2 \varrho_i - \varrho_i)^2$, the quantities m_i, ϱ_i, h_i are the same as in Theorem 1 and the parameters $p_{2,i}, p_{3,i}$ and $p_{4,i}$ are defined in (5) and (6).

Let us recall that the quantities m_i and h_i depend on the cut-off point c.

Proof Starting from (26) we have $P_i(CC_i) = \Phi_{\nu_i^*, \Gamma_i^*; g_{np,2p}}(CC_i^*)$ with

$$\nu_i^* = B^T (1_{l1} \mu_{l1} + 1_{l2} \mu_{l2} + 1_{l3} \mu_{li}) = \begin{pmatrix} \mu_i - \mu_1 \\ \mu_i - \mu_2 \end{pmatrix}$$

and because of $\Sigma_1 = \sigma_1^2 I_p$, $\Sigma_2 = \sigma_2^2 I_p$,

$$\Gamma_i^* = B^T \Sigma_i^{\mathcal{E}_{(np)}} B = \begin{pmatrix} (\frac{\sigma_1^2}{n_1} + \frac{\sigma_i^2}{n_3}) I_p & \frac{\sigma_i^2}{n_3} I_p \\ \frac{\sigma_i^2}{n_3} I_p & (\frac{\sigma_2^2}{n_2} + \frac{\sigma_i^2}{n_3}) I_p \end{pmatrix}.$$

In the first step of the proof the form matrix of the reduced statistic $Z_{(2p)}$ will be transformed into the unit matrix. The matrix Γ_i^* has only two different eigenvalues. Let O_i be the orthogonal matrix which consists of the normalized eigenvectors of Γ_i^*. Put $M_i := O_i D_i^{1/2}$ where $D_i^{1/2}$ is a diagonal matrix consisting of the inverses of the square roots of the eigenvalues of Γ_i^*. We get then $\Gamma_i^* = M_i M_i^T$ and ${M_i}^{-1} \Gamma_i (M_i^T)^{-1} = I_{2p}$ with

$${M_i}^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{1+\varrho_i}} \sqrt{\frac{n_1 n_3}{n_3 \sigma_1^2 + n_1 \sigma_i^2}} \, I_p & \frac{1}{\sqrt{1+\varrho_i}} \sqrt{\frac{n_2 n_3}{n_3 \sigma_2^2 + n_2 \sigma_i^2}} \, I_p \\ \frac{1}{\sqrt{1-\varrho_i}} \sqrt{\frac{n_1 n_3}{n_3 \sigma_1^2 + n_1 \sigma_i^2}} \, I_p & -\frac{1}{\sqrt{1-\varrho_i}} \sqrt{\frac{n_2 n_3}{n_3 \sigma_2^2 + n_2 \sigma_i^2}} \, I_p \end{pmatrix}.$$

Consequently,

$$P_i(CC_i) = \Phi_{\nu_i, \Gamma_i^*; g_{np,2p}}\left(CC_i^*\right) = \Phi_{M_i^{-1} \nu_i^*, I_{(2p)}; g_{np,2p}}\left(CC_i^{\mathbb{T}_1}\right)$$

with $CC_i^{\mathbb{T}_1} = \left\{ u \in \mathbb{R}^{2p} : u^T A_i^{\mathbb{T}_1} u \; < 0 \right\}$, and where

$$A_i^{\mathbb{T}_1} := \begin{pmatrix} (1 + \varrho_i)(1 - \frac{1}{m_i^2}) \, I_p & (-1)^{i+1} \sqrt{1 - \varrho_i} (1 + \frac{1}{m_i^2}) \, I_p \\ (-1)^{i+1} \sqrt{1 - \varrho_i} (1 + \frac{1}{m_i^2}) \, I_p & (1 - \varrho_i)(1 - \frac{1}{m_i^2}) \, I_p \end{pmatrix}.$$

For proving this note that $M_1^T A_1 M_1 = \frac{1}{2}\frac{\sigma_1^2}{n_3} A_1^{\mathbb{T}_1}$, $M_2^T A_2 M_2 = \frac{1}{2}\frac{\sigma_2^2}{n_3} c^2 A_2^{\mathbb{T}_1}$ and therefore $u^T M_i^T A_i M_i\, u < 0$ iff $u^T A_i^{\mathbb{T}_1} u < 0$. In the second step of the proof the matrix $A_i^{\mathbb{T}_1}$ of the quadratic form will be transformed into a diagonal matrix. This will be done with an orthogonal matrix P_i the columns of which are the eigenvectors of $A_i^{\mathbb{T}_1}$:

$$P_i := \frac{1}{\sqrt{N_i}} \begin{pmatrix} \left(\sqrt{h_i} + m_i^2\varrho_i - \varrho_i\right) I_p & (-1)^{i+1}\sqrt{1-\varrho_i}\,(1+m_i^2)\, I_p \\ (-1)^{i+1}\sqrt{1-\varrho_i}\,(1+m_i^2)\, I_p & \left(-\sqrt{h_i} - m_i^2\varrho_i + \varrho_i\right) I_p \end{pmatrix}$$

with $N_i = (1-\varrho_i^2)(1+m_i^2)^2 + (\sqrt{h_i} + m_i^2\varrho_i - \varrho_i)^2$. Because of

$$\Lambda_i := P_i^T A_i^{\mathbb{T}_1} P_i = \begin{pmatrix} (1 - \frac{1}{m_i^2} + \frac{\sqrt{h_i}}{m_i^2})\, I_p & 0_{p,p} \\ 0_{p,p} & (1 - \frac{1}{m_i^2} - \frac{\sqrt{h_i}}{m_i^2})\, I_p \end{pmatrix} \text{ and } P_i I_{2p} P_i^T = I_{2p}.$$

it follows

$$\Phi_{M_i^{-1}\nu_i^*, I_{2p}; g_{np,2p}}\left(CC_i^{\mathbb{T}_i}\right) = \Phi_{\nu_i^{**}, I_{2p}; g_{np,2p}}\left(CC_i^{\mathbb{T}_2}\right)$$

with $CC_i^{\mathbb{T}_2} = \left\{v \in \mathbb{R}^{2p} : v^T \Lambda_i\, v < 0\right\}$ and $\nu_i^{**} := P_i\, M_i^{-1} \nu_i^* = \frac{\sqrt{n_3}}{\sqrt{2}\sigma_i\sqrt{1+p_{2,i}p_{4,i}}} \frac{1}{\sqrt{N_i}}$

$$\times \begin{pmatrix} \left[(-1)^i \frac{1}{\sqrt{1+\varrho_i}} \left(\sqrt{h_i} + m_i^2\varrho_i - \varrho_i\right)\right) + (-1)^{i+1}\sqrt{1+\varrho_i}(1+m_i^2)\right](\mu_2 - \mu_1) \\ \left[-\sqrt{1-\varrho_i}(1+m_i^2) + \frac{1}{\sqrt{1-\varrho_i}} \left(-\sqrt{h_i} - m_i^2\varrho_i + \varrho_i\right)\right](\mu_2 - \mu_1) \end{pmatrix}.$$

A third transformation with respect to the expectation vector results in

$$P_i(CC_i) = \Phi_{\nu_i^{**}, I_{2p}; g_{np,2p}}\left(CC_i^{\mathbb{T}_2}\right) = \Phi_{0_{2p}, I_{2p}; g_{np,2p}}\left(CC_i^{**}(c)\right).$$

□

Theorem 3 *The probability of correct classification satisfies the representation*

$$P_i(CC_i)(c) = F_{p,p,\delta_{1,i},\delta_{2,i};g}(t_i), i \in \{1,2\}$$

with

$$\begin{aligned} t_i &= \frac{1 - m_i^2 + \sqrt{h_i}}{-1 + m_i^2 + \sqrt{h_i}}, \\ \delta_{1,i}^2 &= \frac{a_i^2}{2N_i}\left(\frac{1}{\sqrt{1+\varrho_i}}(\sqrt{h_i} + m_i^2\varrho_i - \varrho_i) - \sqrt{1+\varrho_i}(1+m_i^2)\right)^2, \\ \delta_{2,i}^2 &= \frac{a_i^2}{2N_i}\left(\frac{1}{\sqrt{1-\varrho_i}}\,(\sqrt{h_i} + m_i^2\varrho_i - \varrho_i) + \sqrt{1-\varrho_i}\,(1+m_i^2)\right)^2, \end{aligned}$$

m_i, ϱ_i, h_i and a_i as in Theorem 1, N_i from Lemma 1 but the Parameters $p_{1,i}$ being defined now more general as

$$p_{1,i} = \frac{n_3}{\sigma_i^2} \sum_{l=1}^{p} (\mu_{1l} - \mu_{2l})^2,\ i \in \{1,2\}.$$

Note that the parameters $p_{1,i}$ coincide with those from (5) and (6) if $p = 1$.

Proof Let $\mathfrak{M} := \mathcal{L}(e_1, \ldots, e_p)$ and $\mathfrak{M}^\perp := \mathcal{L}(e_{p+1}, \ldots, e_{2p})$ be orthogonal subspaces of $\mathbb{R}^{2p}$ where e_j is the j'th unit vector of $\mathbb{R}^{2p}$, $j = 1, \ldots, 2p$. Let further $\nu_i^{(1)} := \Pi_{\mathfrak{M}} \nu_i^{**}$ and $\nu_i^{(2)} := \Pi_{\mathfrak{M}^\perp} \nu_i^{**}, i = 1, 2$. Then

$$CC_i^{**}(c) = CC_i^{**} = \big\{ w_{(2p)} \in \mathbb{R}^{2p} : \\ \big(1 - \frac{1}{m_i^2} + \frac{\sqrt{h_i}}{m_i^2}\big)\|w^{(1)} + \nu_i^{(1)}\|^2 + \big(1 - \frac{1}{m_i^2} - \frac{\sqrt{h_i}}{m_i^2}\big)\|w^{(2)} + \nu_i^{(2)}\|^2 < 0 \big\}.$$

One can show that $1 - \frac{1}{m_i^2} + \frac{\sqrt{h_i}}{m_i^2} > 0$, consequently

$$CC_i^{**} = \big\{ w_{(2p)} \in \mathbb{R}^{2p} : \frac{\|w^{(1)} + \nu_i^{(1)}\|^2}{\|w^{(2)} + \nu_i^{(2)}\|^2} < \frac{\sqrt{h_i} - m_i^2 + 1}{\sqrt{h_i} + m_i^2 - 1} \big\}.$$

From this representation of CC_i^{**} it follows that the sets CC_i^{**} from (29) belong to the class of Borel sets $\mathcal{A}_{p,p,\delta_{1,i}\delta_{2,i},t_i}$ defined in the Appendix. Notice that the functions $t_i \,|\, \mathbb{R}^+ \to \mathbb{R}^+$ arising there are chosen here as the constants t_i, i.e. $t_i(r) = t_i, r > 0, i = 1, 2$. With the notion $C_{n,n-m,\delta_1,\delta_2}(x)$ for a certain cone as in the Appendix it follows

$$CC_i^{**} \cap S_{2p}(r) = C_{p,p,\delta_{1,i},\delta_{2,i}}(t_i(r)) \cap S_{2p}(r),$$

where

$$\delta_{1,i}^2 = \|\nu^{(1)}\|^2 = \frac{n_3}{2\sigma_i^2 N_i (1 + p_{2,i} p_{4,i})} \\ \times \left(\frac{1}{\sqrt{1 + \varrho_i}} (\sqrt{h_i} + m_i^2 \varrho_i - \varrho_i) - \sqrt{1 + \varrho_i}(1 + m_i^2) \right)^2 \sum_{l=1}^{p} (\mu_{l2} - \mu_{l1})^2,$$

$$\delta_{2,i}^2 = \|\nu^{(2)}\|^2 = \frac{n_3}{2\sigma_i^2 N_i (1 + p_{2,i} p_{4,i})} \\ \times \left(-\sqrt{1 - \varrho_i}(1 + m_i^2) + \frac{1}{\sqrt{1 - \varrho_i}} (-\sqrt{h_i} - m_i^2 \varrho_i + \varrho_i) \right)^2 \sum_{l=1}^{p} (\mu_{l2} - \mu_{l1})^2.$$

Hence, $CC_i^{**} \in \mathcal{A}_{p,p,\delta_{1,i},\delta_{2,i},t_i}$. Recall that the functions $t_i(r)$ actually do not depend on r. Due to Remark 5 in the Appendix, the probabilities of correct classification can be written in terms of the c.d.f. of the doubly noncentral g-generalized F-distribution:

$$P_i(CC_i) = F_{p,p,\delta_{1,i},\delta_{2,i};g}(t_i),$$

with $\delta_{1,i}, \delta_{2,i}$, t_i as given above. □

4 Numerical results

The representation formulae (7) and (9) with (15) are the basis for efficient procedures to evaluate the probabilities of correct classifications in various situations. Single integrals are to be calculated. This was done by numerical integration according to a

combination of Simpson's rule and the extended trapezoidal rule, see, e.g., in Press et al. (1989), implemented as a Turbo-Pascal program. In this section, several tables of probabilities of correct classification for various parameter configurations are presented.

First of all, Table 1 will be given for a comparison of our one-dimensional results with related results from the literature. We apply formula (7) for probabilities of correct classification into the first population in the special case of $c = 1$, equal variances $\sigma_1^2 = \sigma_2^2 =: \sigma^2$ and only one individual in the third population, $n_3 = 1$, i.e.

$$p_{1,1} = \frac{(\mu_1 - \mu_2)^2}{\sigma^2},\ p_{2,1} = 1,\ p_{3,1} = \frac{1}{n_1},\ p_{4,1} = \frac{1}{n_2}.$$

Table 1 Comparison with literature: Gaussian sample distributions.

Δ	n_1	n_2	Theorem 1	Sch./Vark	Dejew	Mem./Oka.
1	3	3	0.638943	0.638943	0.678324	0.623803
0.1	5	5	0.502388	0.502388	0.519012	1.083490
0.3	5	5	0.520740	0.520740	0.556864	0.709856
0.5	5	5	0.553771	0.553771	0.594206	0.656126
1	5	5	0.662364	0.662364	0.683266	0.665046
3	5	5	0.923668	0.923668	0.929415	0.874694
5	5	5	0.991429	0.991429	1.102311	1.072366
10	5	5	0.999999	0.999999	-0.577656	1.000932
1	5	10	0.677435	0.677435	0.689032	0.678786
1	10	5	0.668284	0.668284	0.681573	0.672178
0.1	10	10	0.503443	0.503443	0.513459	0.662283
1	10	10	0.682462	0.682462	0.687254	0.682865
3	10	10	0.928383	0.928383	0.934892	0.883520
1	20	20	0.689003	0.689003	0.689348	0.688235
1	50	50	0.690588	0.690588	0.690638	0.690420
0.1	100	100	0.510352	0.510352	0.519889	0.521717
0.3	100	100	0.557455	0.557455	0.559472	0.560179
0.5	100	100	0.598425	0.598425	0.598471	0.598861
1	100	100	0.691024	0.691024	0.691074	0.690982
5	100	100	0.993680	0.993680	1.142290	0.988134
10	100	100	1.000000	1.000000	-0.029009	0.999991

The exact formula of Schaafsma and van Vark (1977) is based on assumptions of normality, equality of the variances and sample size one in the third population and is given in terms of the Mahalanobis distance $\Delta = \sigma^{-1}|\mu_1 - \mu_2|$ and the sample sizes n_1 and n_2. The values in column 5 coming from application of the formula of Schaafsma and van Vark (1977) coincide with our values in column 4 in all cases. Note that there exist approximation formulae for the special case under consideration here but with the extension to higher dimensional measurements.

In column 6, we tabled values derived from Dejew's approximation formulae including second order terms given in Ahrens and Läuter (1981). Values derived from Memon

and Okamoto's approximation formulae as given in Siotani (1982) are tabled in column 7. These approximation formulae are in some situations not very precisely, especially in cases of probabilities near an half and near one, respectively.

On the basis of Theorem 1, we are in a position to compute probabilities of correct classification in more general cases of unequal variances and more than one individual in the third population, see Tables 2 and 3.

Table 2 Unequal variances and repeated measurements, the case $\mu_3 = \mu_1$.

c	n_1	n_2	n_3	μ_1	μ_2	σ_1^2	σ_2^2	$p_{1,1}$	$p_{2,1}$	$p_{3,1}$	$p_{4,1}$	$P_1(CC_1)(c)$
1	4	4	4	0.5	0	1	1	1	1	1	1	0.582475
$\frac{1}{2}$	4	4	4	0.5	0	1	1	1	1	1	1	0.345909
2	4	4	4	0.5	0	1	1	1	1	1	1	0.774654
1	4	4	4	0.25	0	1	1	0.25	1	1	1	0.522347
1	4	4	4	1	0	1	1	4	1	1	1	0.746820
1	4	4	4	1	0	4	2	1	0.5	1	1	0.561523
1	4	4	4	0.5	0	1	2	1	2	1	1	0.620593

Table 3 The case $\mu_3 = \mu_2$: $P_2(CC_2)(c; p_{1,2}, p_{2,2}, p_{3,2}, p_{4,2}; g^G)$
$= P_1(CC_1)(1/c; p_{1,1}/p_{2,1}, 1/p_{2,1}, p_{4,1}, p_{3,1}; g^G)$.

c	n_1	n_2	n_3	μ_1	μ_2	σ_1^2	σ_2^2	$p_{1,2}$	$p_{2,2}$	$p_{3,2}$	$p_{4,2}$	$P_2(CC_2)(c)$
1	4	4	4	0.5	0	1	1	1	1	1	1	0.582475
$\frac{1}{2}$	4	4	4	0.5	0	1	1	1	1	1	1	0.774654
2	4	4	4	0.5	0	1	1	1	1	1	1	0.345909
1	4	4	4	0.25	0	1	1	0.25	1	1	1	0.522347
1	4	4	4	1	0	1	1	4	1	1	1	0.746820
1	4	4	4	1	0	4	2	2	2	1	1	0.664547
1	4	4	4	0.5	0	1	2	$\frac{1}{2}$	0.5	1	1	0.507431

Using Theorem 2 we can also exactly evaluate probabilities of correct classification for elliptically contoured error variables. In Table 4 such values are given for a Pearson Typ VII distribution with parameters N and m. This $g_{n,2}$-marginal case is described in Example 1 and is a distribution with heavier tails then that of the normal distribution. With $N = (m+n)/2$ one gets the multivariate t-distribution which is considered here. Note that the probabilities of correct classification depend on c, $p_{1,i}, \ldots, p_{4,i}$, m, N and via $g_{n,2}$ on n.

Table 4 Multivariate-t sample distribution.

c	$p_{1,i}$	$p_{2,i}$	$p_{3,i}$	$p_{4,i}$	m	N	$P_1\left(CC_1\right)(c; g_{12,2}^{PVII})$	$P_2(CC_2)\left(c; g_{12,2}^{PVII}\right)$
1	1	1	1	1	1	6.5	0.570024	0.570024
1	1	1	1	1	2	7	0.575506	0.575506
$\frac{1}{2}$	1	1	1	1	1	6.5	0.337745	0.767720
$\frac{1}{2}$	1	1	1	1	2	7	0.341436	0.770805
2	1	1	1	1	1	6.5	0.767720	0.337745
2	1	1	1	1	2	7	0.770805	0.341436
1	1	2	1	1	1	6.5	0.615027	0.615027
1	1	2	1	1	2	7	0.617577	0.617577

The importance of the case $n_3 > 1$ was inter alia mentioned in Schaafsma and van Vark (1977) and in McLachlan (1992). In actual practice the researcher will usually have to classify more individuals, for instance if one wants to assign all the skeletal remains found in a particular specified burial site to one of two prehistoric populations.

Under the assumptions under which Table 5 has been generated, it can be seen that the increase of the number of individuals in the third population yields a greater effect onto the probabilities of correct classification into the first population than the increase of the number of individuals in the first population. The effect of the increase of n_2, however, is again more significant.

Table 5 Most effective sample size increase when $\mu_3 = \mu_1$, $c = 1$, $\mu_1 = 1$, $\mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$.

$n_2 = 3$, $n_3 = 3$	
n_1	$P_1(CC_1)(1)$
1	0.695219
2	0.699711
3	0.702820
4	0.705093
5	0.706825
10	0.711578
50	0.717412
100	0.718330

$n_1 = 3$, $n_3 = 3$	
n_2	$P_1(CC_1)(1)$
1	0.611121
2	0.668726
3	0.702820
4	0.724939
5	0.740276
10	0.776183
50	0.809411
100	0.813715

$n_1 = 3$, $n_2 = 3$	
n_3	$P_1(CC_1)(1)$
1	0.638943
2	0.679132
3	0.702820
4	0.718766
5	0.730307
10	0.759937
50	0.793355
100	0.798400

The free choice of c ($c \in \mathbb{R}^+$) in the decision function allows the statistician in practice to determine c on such a way, that certain demands on the probabilities are fulfilled.

In Table 6 we determined c in such a way that we got equal probabilities of correct classification

$$P_1(CC_1)(c) = P_2(CC_2)(c) =: P(CC)(c).$$

To this end we used the bisection method (Press et al. (1989)) as a root finding algorithm.

Table 6 Risk equalizing cut-off points.

$\frac{n_3(\mu_1-\mu_2)^2}{\sigma_1^2}$	$\frac{\sigma_2^2}{\sigma_1^2}$	$\frac{n_3}{n_1}$	$\frac{n_3}{n_2}$	c	$P(CC)(c)$
0.25	1	0.5	0.5	1	0.533877
1	1	0.5	0.5	1	0.617859
1	1	1	1	1	0.582475
4	1	1	1	1	0.746820
1	2	0.5	0.5	0.883739	0.593300
1	0.5	0.5	0.5	1.223383	0.578692
1	1	0.5	1	1.036860	0.598719
1	1	1	0.5	0.964450	0.598719

In Table 7 we will consider additionally costs of misclassification. Let $C(i|j) > 0$ be the costs of misclassification which arise, if one classifies an object from Π_j as coming from Π_i. The statistician could search for such a cut-off point c that

$$\frac{C(2|1)}{C(1|2)} = \frac{P_2(MC_2)(c)}{P_1(MC_1)(c)}$$

holds for the probabilities of misclassification $P_i(MC_i)(c) = 1 - P_i(CC_i)(c)$, $i = 1, 2$.

Table 7 Cost driven cut-off points.

$\frac{n_3(\mu_1-\mu_2)^2}{\sigma_1^2}$	$\frac{\sigma_2^2}{\sigma_1^2}$	$\frac{n_3}{n_1}$	$\frac{n_3}{n_2}$	$\frac{C(2\|1)}{C(1\|2)}$	c	$1 - P_1(CC_1)(c)$	$1 - P_2(CC_2)(c)$
1	1	1	1	3	2.047427	0.220413	0.661240
1	1	1	1	1	1	0.417525	0.417525
1	1	1	1	0.5	0.644801	0.570775	0.285387
0.25	1	1	1	3	2.151153	0.242446	0.727339
0.25	1	1	1	0.5	0.625045	0.641410	0.320705
1	0.5	1	1	3	2.078769	0.230112	0.690335
1	0.5	1	1	0.5	0.641438	0.598374	0.299187

Note that it was assumed in the former discussion for determining c that all parameters of the classification problem are known. Let us assume now that the expectations μ_1 and μ_2 are not known, but all other parameters are known. It can be recommended then to choose c as

$$c = \Big(\frac{1 + \frac{n_3}{n_2}}{1 + \frac{n_3}{n_1}} \frac{1 + \frac{\sigma_1^2}{\sigma_2^2}\frac{n_3}{n_2}}{1 + \frac{\sigma_2^2}{\sigma_1^2}\frac{n_3}{n_1}}\Big)^{1/4}. \tag{30}$$

It follows from the explicit formulae in Theorems 1 and 2 that the probabilities of correct classification $P_i(CC_i)(c)$ are nearly the same for $i = 1$ and $i = 2$ if c is chosen as in (30). Note that the quantities $\|q_i\|$, a_i and ϱ_i occurring in Theorems 1 and 2 do not depend on the cut-off point c and that m_1 and m_2 depend explicitly on c. Starting therefore from the equation $m_1 = m_2$ gives similar values for $s_{min,1}$ and $s_{min,2}$ as well as for $s_{max,1}$ and $s_{max,2}$. Table 7a gives an impression of how the choice of c influences

the probabilities of correct classification. This table includes four rows for each of the examples (i) up to (iv). The first row corresponds each time to the case $c = 1$, the second to the case $c = \sigma_1/\sigma_2$, the third deals with c from (30) and the forth with the risk equalizing c for equal costs of misclassification.

Table 7a Several choices of the cut-off point.

Ex.	n_1	n_2	n_3	μ_2	σ_1^2	σ_2^2	c	$P_1(CC_1)$	$P_2(CC_2)$	$P_1 + P_2$
(i)	4	4	4	1	1	2	1	0.736220	0.648230	1.384451
							0.707107	0.636353	0.743300	1.379653
							0.840896	0.689517	0.699714	1.389232
							0.856268	0.694712	0.694712	1.389424
(ii)	4	4	4	1	1	5	1	0.750662	0.525090	1.275752
							0.447214	0.514512	0.764610	1.279121
							0.668740	0.643891	0.662209	1.306100
							0.689554	0.652981	0.652981	1.305962
(iii)	3	5	7	0.5	1	2	1	0.641850	0.492028	1.133878
							0.707107	0.532184	0.609111	1.141296
							0.681732	0.520103	0.620756	1.140859
							0.795160	0.570549	0.570549	1.141097
(iv)	50	50	50	0.1	1	2	1	0.596201	0.477100	1.073301
							0.707107	0.364174	0.714035	1.078210
							0.840896	0.478646	0.604921	1.083567
							0.845594	0.540085	0.540085	1.080170

Table 7a shows that among the considered four cases formula (30) yields relatively good results and can therefore be recommended to be used.

Due to Example 5 it is also possible to evaluate probabilities of correct classification as values of the c.d.f. of doubly noncentral g-generalized F-distributions with (1,1) degrees of freedom. For special parameters of the underlying classification problem the n.c.p.'s and arguments of this distribution are evaluated from Example 5 and given in Table 8 only with 6 digits. Note that we internally used all available digits to get exact as possible probabilities of correct classification. This circumstance is indicated by the symbol "$\approx$" in columns 6, 7 and 8. Then the exact values of the g-generalized F-distribution are determined and tabled for the Gaussian and the multivariate t density generating functions, respectively. For the first case let $F_{g^G} := F_{1,1,\delta_{1,1}^2,\delta_{2,1}^2;g^G}(t_1)$ and for the latter case let $F_{g_{12,2}^{PVII}} := F_{1,1,\delta_{1,1}^2,\delta_{2,1}^2;g_{12,2}^{PVII}}(t_1)$ and note that the parameters are choosen as $m = 1$ and $N = 6.5$.

Table 8 Doubly noncentral g-generalized F-distributions.

c	$p_{1,1}$	$p_{2,1}$	$p_{3,1}$	$p_{4,1}$	$\delta^2_{1,1} \approx$	$\delta^2_{2,1} \approx$	$t_1 \approx$	F_{g^G}	$F_{g^{PVII}_{12,2}}$
1	1	1	1	1	0.044658	0.622008	1.000000	0.582475	0.570024
$\frac{1}{2}$	1	1	1	1	0.006006	0.660660	0.208715	0.345909	0.337145
2	1	1	1	1	0.115115	0.551551	4.791288	0.774654	0.767720
1	1	2	1	1	0.025426	0.374574	1.558258	0.620593	0.615027
1	1	1	2	1	0.026139	0.573861	1.000000	0.579059	0.567869

Note that the numerical results coincide with the results determined with the explicit representations from Theorems 1 and 2.

For the case of the Gaussian density generating function we can compare our numerical results with available special results from the literature.

Price (1964) gives explicit formulae for the c.d.f. of the usual doubly noncentral F-distribution, where the numbers of degrees of freedom are either both even or both odd, respectively. By using formula (4.8) in Price (1964) we received the numerical values given in column 5 of Table 9. Note that the formulae of Price (1964) were used in Moran (1975) to evaluate error rates.

The approximate normality of the cube root of the noncentral chi-square distribution and an Edgeworth-series expansion are used in Mudholkar et al. (1976) to derive an approximation requiring only normal approximation. Formulae (3.2) and (3.3) yield the values of the culumn 6 in Table 9.

Representations of the c.d.f. of the doubly noncentral F- distribution are presented in Chou et al. (1985) in terms of the c.d.f. of the noncentral chi-square distribution. Specializing formula (1) in Chou et al. (1985) to the case of (1,1) d.f. and using a representation of the c.d.f. of the noncentral chi-square distribution with one degree of freedom:

$$CQ(1, \delta^2)(c^2) = \Phi(c - \delta) + \Phi(c + \delta) - 1$$

cited therein and the p.d.f. of the non central chi-square distribution we get column 7.

Table 9 Classical doubly non central F-distributions. Comparison with the literature.

$\delta^2_{1,1}$	$\delta^2_{2,1}$	t_1	Theorem 5	Price	Mudh.	Chou
0.044658	0.622008	1.000000	0.582475	0.582475	0.624917	0.582437
0.006006	0.660660	0.208715	0.345911	0.345911	0.342071	0.345894
0.115115	0.551551	4.791288	0.774654	0.774654	0.882719	0.774570
0.025426	0.374574	1.558258	0.620593	0.620593	0.669155	0.620539
0.026139	0.573861	1.000000	0.579059	0.579059	0.622287	0.579020

Let us remark, that the values of Price's formula coincide with our values. The results of Mudholkar et al. are not satisfactory for the case of (1,1) d.f. and the values of Chou et al. coincide with our values at least to three digits.

To get a more detailed impression of the accuracy of our formula in comparison with that of Price's formula Table 10 is given.

Table 10 Many digits comparison with Price's formula.

row	$\delta_{1,1}^2$	$\delta_{2,1}^2$	t_1
(1)	0.0446581987385205	0.622008467928146	1.000000000000000
(2)	0.00600649797934476	0.660660168687322	0.208712152522080
(3)	0.115115443097341	0.551551223569326	4.79128784747792
(4)	0.0254256878112061	0.374574312188794	1.55825756949558
(5)	0.0261387212474169	0.573861278752583	1.000000000000000

row	Example 5	Price
(1)	0.582475444211846	0.582475444228883
(2)	0.345909195073269	0.345909195089144
(3)	0.774653756627723	0.774653756607881
(4)	0.620593176025199	0.620593175997704
(5)	0.579059127459032	0.579059127517264

The values of the c.d.f. of the doubly non central F-distribution are given there with 15 digits. One numerical integration over a domain from zero to infinity is necessary for our formula and two numerical integrations, each over a finite domain, are necessary for the formula of Price. Note that we took the value 12 as the upper integration limit for our numerical integration because of the fast decreasing density generating function in the integrand. The values of columns "Example 5" and "Price" in Table 10 are received by numerical integration which is performed by Simpson's rule with a large number of steps. Note that for the considered cases the numerical results of both formulae coincide (after rounding) in 10 digits. Finally note that we don't know competing results from the literature for the g-generalized case dealt with here.

Due to Theorem 3 we can evaluate probabilities of correct classification also for a p-dimensional feature vector as values of the doubly non central g-generalized F-distribution with (p, p) degrees of freedom. For the Gaussian density generating function we get the usual doubly non central F-distribution. The latter can be written as a linear combination of independent non central chi-square variables. Such a distribution can be evaluated by using formula (3.2) in Imhof (1961). Specifying it to the situation of Example 5 gives

$$F_{p,p,\delta_{1,i}^2,\delta_{2,i}^2;g^G}(t_i) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u)}{u\gamma(u)}du \tag{31}$$

with

$$\theta(u) = \frac{1}{2}\sum_{k=1}^{2}\big(p\arctan(\lambda_k u) + \delta_{k,i}^2\lambda_k u(1+\lambda_k^2u^2)^{-1}\big),$$

$$\gamma(u) = \prod_{k=1}^{2}(1+\lambda_k^2u^2)^{p/4}\exp\Big\{\frac{1}{2}\sum_{k=1}^{2}(\delta_{k,i}\lambda_k u)^2(1+\lambda_k^2u^2)^{-1}\Big\}$$

and $\lambda_1 = 1$, $\lambda_2 = -t_i$.

Table 11 illustrates that increasing dimensions p of the feature vector yield increasing probabilities of correct classification when the distance of each coordinate of the expectation vector from zero is constant. Note that we assumed uncorrelated features.

Table 11 Uncorrelated features: values $P(CC)$ for increasing dimensions, $c = 1$, $n_1 = n_2 = n_3 = 4$, $\mu_{1l} = 0$, $\sigma_1^2 = \sigma_2^2 = 1$. Application of formula (31).

Feature distance	Dimension p						
μ_{2l}	1	2	3	5	10	50	100
0.10	0.5037	0.5057	0.5073	0.5097	0.5141	0.5320	0.5453
0.25	0.5223	0.5346	0.5437	0.5577	0.5827	0.6817	0.7483
0.50	0.5824	0.6230	0.6517	0.6943	0.7646	0.9468	0.9888
0.75	0.6634	0.7307	0.7748	0.8346	0.9147	0.9989	1.0000
1.00	0.7468	0.8271	0.8744	0.9292	0.9805	1.0000	1.0000

Note that formula (31) is not applicable in the case of a non-Gaussian density generator. We are interested therefore in getting numerical results from the geometric measure representation formula using the threefold integral (33). This intersection percentage function was determined by threefold calling a Simpson quadrature formula. For comparison Table 11a contains again results for the Gaussian case.

Table 11a Increasing dimensions in the situation of Table 11. Application of formula (33).

Feature distance	Dimension p						
μ_{2l}	1	2	3	5	10	50	100
0.25	0.5223	0.5340	0.5432	0.5577	0.5827	0.6794	0.7483

Note that the numerical evaluations based upon formula (33) are relatively time consuming and do not yield as accurate results as when using Imhof's formula (31). However, on principle, we can evaluate with one and the same intersection percentage function which we used for establishing Table 11a the probabilities of correct classification for arbitrary continuous spherically symmetric sampling error distributions. The case of a multivariate t-distribution is dealt with in Table 11b.

Table 11b Uncorrelated features: values $P(CC)$ for increasing dimensions, density generator of the multivariate t-distribution with $m = 2$, $c = 1$, $n_1 = n_2 = n_3 = 4$, $\mu_{1l} = 0$, $\sigma_1^2 = \sigma_2^2 = 1$.

Feature distance	Dimension p				
μ_{2l}	1	2	3	5	10
0.25	0.5218	0.5325	0.5406	0.5527	0.5739

The next study concerns fixed Mahalanobis distance $\Delta^2 = 1$ as it was considered, e.g., in Ahrens and Läuter (1981). Note that in the present situation the parameters

$p_{1,i} = \frac{n_3}{\sigma_i^2} \sum_{l=1}^{p} (\mu_{1l} - \mu_{2l})^2$, $i \in \{1, 2\}$ coincide with Δ^2 if $n_3 = 1$, $\sigma_1^2 = \sigma_2^2$ and that fixing Δ^2 results in decreasing feature distance for increasing dimensions. We have chosen the distances $\mu_{2l} = 1/\sqrt{p}$, $l = 1, \ldots, p$, for all features in Table 12.

Table 12 Increasing dimensions: the case of fixed Mahalanobis distance and decreasing feature distances.
$c = 1$, $n_1 = n_2 = 4$, $n_3 = 1$, $\mu_{1l} = 0$, $l = 1, \ldots, p$, $\sigma_1^2 = \sigma_2^2 = 1$.

p	1	2	3	5	10	100
$P_i(CC_i)$	0.6528	0.6297	0.6154	0.5975	0.5749	0.5262

Furthermore we studied how the coordinates of the expectation vector must be chosen to achieve a constant probability of correct classification $P_1(CC_1) = P_2(CC_2) = P(CC)$ for increasing dimension of the feature vector. The results of this study are given in Tables 13 and 14.

Table 13 Increasing dimensions: the case of a fixed probability of correct classification $P(CC) = 0.7468$.
$c = 1$, $n_1 = n_2 = n_3 = 4$, $\mu_{1l} = 0$, $l = 1, \ldots, p$, $\sigma_1^2 = \sigma_2^2 = 1$.

p	1	2	3	5	10	100
μ_{2l}	1	0.7468	0.6908	0.5879	0.4768	0.24905

Note that Table 14 gives an answer to the question of which order should be the rate of convergence of $|\mu_2 - \mu_1| := f(n)$ towards zero as n with $n/3 = n_1 = n_2 = n_3$ tends to infinity if one wants ensure that the probability of correct classification $P(CC)$ does not change as $n \to \infty$. It turns out from Table 14 that $f(n) = 1/\sqrt{n}$ which corresponds to the often considered so called $n^{-1/2}$-local alternatives from test theory.

Table 14 Increasing sample sizes: the case of a fixed $P_1(CC_1) = 0.582475$.
$c = 1$, $n_1 = n_2 = n_3 = 4$, $\mu_{1l} = 0$, $l = 1, \ldots, p$, $\sigma_1^2 = \sigma_2^2 = 1$.

n	4	10^2	10^4	10^6
μ_{2l}	0.5	0.1	0.01	0.001

5 Estimating probabilities of correct classification

Let us restrict the considerations in this section to the Gaussian case, i.e. put $g = g^G$. The probabilities of correct classification depend on the possibly unknown parameters μ_1, μ_2, μ_3, σ_1^2, σ_2^2, σ_3^2. If $D_{1/3}$ would be correct then we can estimate the parameters μ_2 and σ_2^2 as well as the parameters μ_1 and σ_1^2 by the single sample based estimators $\overline{y_{2\bullet}}$, s_2^2 and the pooled sample estimators $\overline{y_\bullet^{(1/3)}}$ and ${s_{n_1+n_3}^{(1/3)}}^2$, respectively, where

$${s_{n_1+n_3}^{(1/3)}}^2 = \frac{1}{n_1 + n_3 - 1} \left[\sum_{j=1}^{n_1} \left(y_i^{(1)} - \overline{y_\bullet^{(1/3)}} \right)^2 + \sum_{j=1}^{n_3} \left(y_i^{(3)} - \overline{y_\bullet^{(1/3)}} \right)^2 \right].$$

Based upon

$$\widehat{p_{1,1}} = \frac{n_3}{s^{(1/3)\ 2}_{n_1+n_3}}\left(\overline{y_\bullet^{(1/3)}} - \overline{y_{2\bullet}}\right)^2 \text{ and } \widehat{p_{2,1}} = \frac{s_2^2}{s^{(1/3)\ 2}_{n_1+n_3}}$$

we can estimate the probabilities of correct classification into the first population by

$$\widehat{P_1(CC_1)} = P_1(CC_1)(c, \widehat{p_{1,1}}, \widehat{p_{2,1}}, p_{3,1}, p_{4,1}; g_G).$$

A related simulation study is organized as follows: For fixed n_1, n_2 and n_3 each Gaussian distributed subsample vector of dimension n_i with expectation $\mu_i 1_{n_i}$ and covariance matrix $\sigma_i^2 I_{n_i}$ was simulated N=500 times, $i = 1, 2, 3$. For every repetition, the single and the pooled sample estimators $\widehat{p_{1,1}}^{(j)}$ and $\widehat{p_{2,1}}^{(j)}$ were computed, and based upon them the estimators

$$\widehat{p_1(CC_1)}^{(j)} = P_1(CC_1)(\widehat{p_{1,1}}^{(j)}, \widehat{p_{2,1}}^{(j)}, p_{3,1}, p_{4,1}; g_G),\ j = 1, \ldots, 500$$

of the probability of correct classification $P_1(CC_1)$ were evaluated, too. The arithmetic mean $\overline{\widehat{p}} = \overline{\widehat{p_1}}^{(\bullet)}$ of these probability estimators was tabled in column 3 of Table 15 and their empirical variance $s^2_{\hat{P}}$ in column 4. Furthermore the relative errors

$$r^{(j)} := \frac{|\widehat{p_1(CC_1)}^{(j)} - P_1(CC_1)|}{P_1(CC_1)},\ j = 1, \ldots, 500$$

were computed and their arithmetic mean $\overline{r}$ and empirical variance s_R^2 were tabled in columns 5 and 6, respectively. It turns out from Table 15 that increasing probabilities of correct classification correspond to decreasing relative estimation errors.

Table 15 Increasing balanced sample sizes $n_1 = n_2 = n_3$: the case of a fixed feature difference. $c = 1$, $\mu_2 = 0$, $\mu_1 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$.

n_1	$P(CC)(1)$	$\overline{\widehat{p}}$	$s^2_{\hat{P}}$	$\overline{r}$	s_R^2
3	0.702820	0.740398	0.025	0.1938	0.025
4	0.746820	0.742211	0.023	0.1703	0.013
5	0.782989	0.765938	0.022	0.1591	0.011
10	0.891467	0.860205	0.015	0.1021	0.010
20	0.965326	0.943363	0.004	0.0437	0.003
30	0.987274	0.973138	0.002	0.0214	0.001
100	0.999978	0.999892	0.000	0.0001	0.000

The probability of correct classification is fixed in Table 16 at a preassumed level and the parameters $p_{1,i}$ up to $p_{4,i}$ are fixed as well for $i = 1, 2$. To achieve this, the difference $|\mu_1 - \mu_2|$ of the expectations in Π_1 and Π_2 becomes smaller when the overall sample size n increases. Simulation studies like in the case of Table 15 were made for producing the results in Table 16 which reflect increasing estimation accuracy for increasing sample sizes.

Table 16 Increasing balanced sample sizes $n_1 = n_2 = n_3$: the case of a fixed probability of correct classification $P(CC) = 0.582475$.
$c = 1$, $\mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$.

n_1	μ_1	$\overline{\widehat{p}}$	$s^2_{\widehat{p}}$	$\overline{r}$	s^2_R
4	0.5	0.648	0.022	0.2248	0.028
100	0.1	0.635	0.017	0.1835	0.023
10^4	0.01	0.635	0.016	0.1813	0.023
10^6	0.001	0.595	0.010	0.1305	0.011

The next simulation study in Table 17 was done for the case that the sample distribution is the multivariate-t distribution.

Table 17 Constant probability of correct classification $P_1(CC_1) = 0.570024$.
$c = 1$, $\mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$.
$g = g^P$ with $m = 1$, $N = (1 + n)/2$.

n_1	n_2	n_3	μ_1	$\overline{\widehat{p}}$	$s^2_{\widehat{p}}$	$\overline{r}$	s^2_R
4	4	4	0.5	0.605290	0.012	0.1649	0.012
100	100	100	0.1	0.595109	0.007	0.1224	0.008
10^4	10^4	10^4	0.01	0.598885	0.007	0.1251	0.009
10^6	10^6	10^6	0.001	0.594235	0.007	0.1232	0.009

Another question arises if the parameters μ_1, μ_2, σ_1^2 and σ_2^2 are not known and one wants to determine a "good" cut-off point c.
First, we want to estimate c from (30) by plugging in the estimates s_1^2 and s_2^2 for the variances. The resulting estimator $\hat{c}$ was evaluated for 500 randomly choosen normal samples, and then the arithmetic mean $\overline{\hat{c}}$ was evaluated. The results are given in Table 17a for the examples (i) to (iv) of Table 7a.

Table 17a Estimating c from (30).

Ex.	$\overline{\hat{c}}$	$\overline{\hat{p_1}}$	$\overline{\hat{p_2}}$	$\overline{\hat{p_1}} + \overline{\hat{p_2}}$	$s^2_{\hat{c}}$	$s^2_{\hat{p_1}}$	$s^2_{\hat{p_2}}$	$\overline{r}$	s^2_R
(i)	0.904836	0.685638	0.681861	1.367499	0.088	0.009	0.008	0.2679	0.059
(ii)	0.714178	0.636020	0.649391	1.285411	0.071	0.011	0.011	0.2944	0.076
(iii)	0.737871	0.514831	0.613284	1.128116	0.108	0.018	0.016	0.3552	0.113
(iv)	0.847527	0.539907	0.539957	1.079864	0.003	0.001	0.001	0.0567	0.002

Table 17a contains furthermore the arithmetic means $\overline{\hat{p_1}}$ and $\overline{\hat{p_2}}$ of the probabilities of correct classifications into the two populations, the empirical variances of $\hat{c}$, $\hat{p_1}$, $\hat{p_2}$, the arithmetic mean of the relative errors $r^{(j)} := |\hat{c} - c|/c$, $j = 1, \ldots, 500$, and the empirical variance of the relative error. One can see in Table 17a that for the examples (i) to (iv) the sums of the probabilities of correct classifications are smaller in tendency for unknown variances than for known variances.

Analogous considerations will be made now for the risk-equalizing cut-off-point c. In this case it is additionally necessary to estimate the expectations μ_1 and μ_2. The results are given in Table 17b.

Table 17b Estimating risk-equalizing c.

Ex.	$\overline{\hat{c}}$	$\overline{\hat{p_1}}$	$\overline{\hat{p_2}}$	$\overline{\hat{p_1}} + \overline{\hat{p_2}}$	$s^2_{\hat{c}}$	$s^2_{\hat{p_1}}$	$s^2_{\hat{p_2}}$	$\overline{r}$	s^2_R
(i)	0.891769	0.742211	0.742211	1.484422	0.087	0.023	0.023	0.2621	0.052
(ii)	0.739055	0.739093	0.739093	1.478186	0.072	0.020	0.020	0.2775	0.080
(iii)	0.801861	0.712718	0.712718	1.425436	0.091	0.023	0.023	0.2892	0.060
(iv)	0.853702	0.646150	0.646150	1.292299	0.004	0.016	0.016	0.0594	0.002

Note that one could argue from the latter four examples that the received probabilities of correct classifications are greater in the case of unknown than in the case of known parameters. However, we counted how often the probabilities of correct classification were greater in the case of unknown parameters than that in the case of known parameters. The hypothesis that the probability for this event is equal 1/2 was not rejected by a respective significance test at the level $\alpha = 0.001$.

Acknowledgement The authors are grateful to H. H. Bock, the referees and the Editor-in-Charge for making valuable remarks to former versions of this paper.

Appendix
Geometry behind noncentral generalized chi-square, Student- and Fisher distributions

A unified geometric approach to several statistical distributions for elliptically contoured populations is based upon a geometric representation formula for Gaussian and spherical measures in Richter (1985) and in Richter (1991,95), respectively. Let $Y_{(n)}$ be a n-dimensional spherically symmetric distributed random vector having Lebesgue density with a d.g.f. g. The spherical measure $\Phi(\,\cdot\,; g) = \mathrm{EC}_n^d(0_n, I_n, g)(\cdot)$ allows the representation

$$\Phi(A; g) = \int_0^\infty \mathcal{F}_n(A, r)\, r^{n-1} g(r^2)\, dr \Big/ \int_0^\infty r^{n-1} g(r^2)\, dr,\ A \in \mathfrak{B}^n \tag{32}$$

where g is assumed to fulfill assumption (1), $\mathcal{F}_n(A, r)$ is the i.p.f., see Definition 1. If $\Phi_{\mu,\Sigma;g}$ denotes an elliptically contoured probability distribution with expectation $\mu \in \mathbb{R}^n$, form matrix $\Sigma \in \mathbb{R}^n \times \mathbb{R}^n$ and d.g.f. g then $\Phi_{0,I_n;g}(A)$ coincides with $\Phi(A; g)$. It's well known that different statistics generate in a canonical way different types of sets in the sample space $\mathbb{R}^n$.

Example 2 Let $A(x) = \left\{y_{(n)} \in \mathbb{R}^n :\ \|y_{(n)} + \mu\|^2 < x^2\right\}$, $x > 0$ be a family of balls in $\mathbb{R}^n$ and $\delta^2 = \|\mu\|^2$, then

$$\Phi(A(x); g) = CQ(n, \delta^2; g)(x^2)$$

is the c.d.f. of the noncentral g-generalized chi-square distribution with n d.f. and n.c.p. δ^2. This approach has been dealt with in Ittrich et al. (2000) as well as in Richter and Schumacher (2000).

Example 3 Assume that $\mathcal{N}_1$ and $\mathcal{N}_2 = \mathcal{L}(\mu)$ with $\mu \in S_n(1)$ are linear subspaces of $\mathbb{R}^n$ having dimensions $n-2$ and 1, respectively, and being orthogonally to each other. Let

$$A(x) = \left\{ y_{(n)} \in \mathbb{R}^n : \frac{\|\Pi_{\mathcal{N}_2} y_{(n)} + \delta\mu\| \operatorname{sign}(y_{(n)} + \delta\mu, \mu)}{\|\Pi_{\mathcal{N}_1} y_{(n)}\|/\sqrt{n-2}} > x \right\}$$

denote a class of cones from $\mathfrak{B}^n$. Then $\Phi(A(x); g) = t^g_{n-2}(\delta)$, $x \in \mathbb{R}$, has been called in Richter (1994) noncentral g-generalized Student distribution with $n-2$ d.f. and n.c.p. $\delta \in \mathbb{R}$. It is shown in Richter (1994) which spaces $\mathcal{N}_1$ and $\mathcal{N}_2$ play a role when evaluating certain probabilities of correct selection.

Example 4 Let $\mathfrak{M} \subset \mathbb{R}^n$ be a m-dimensional linear subspace of $\mathbb{R}^n$ and $\mathfrak{M}^\perp$ the orthogonal complement of $\mathfrak{M}$, put $y^{(1)} = \Pi_{\mathfrak{M}} y_{(n)}$ and $y^{(2)} = \Pi_{\mathfrak{M}^\perp} y_{(n)}$, and consider the family of cones

$$A(x) = \left\{ y_{(n)} \in \mathbb{R}^n : \frac{\|y^{(1)}\|^2}{\|y^{(2)}\|^2} < \frac{m}{n-m} x \right\}, \; x \in \mathbb{R}.$$

Then, according to Richter (1991, 1995),

$$\Phi_{\mu, I_n, g}\left(A(x)\right), \; x \in \mathbb{R} = F_{m,n-m,\delta_1^2,\delta_2^2;g}(x)$$

is the c.d.f. of the *doubly noncentral g-generalized Fisher distribution* with m and $n-m$ degrees of freedom (d.f.), respectively, and noncentrality parameters (n.c.p.) $\delta_1^2 = \|\Pi_{\mathfrak{M}}\mu\|^2$ and $\delta_2^2 = \|\Pi_{\mathfrak{M}^\perp}\mu\|^2$. The special cases $\delta_1^2 = 0$ or $\delta_2^2 = 0$ correspond to noncentral g-generalized Fisher distributions of second and first kind, respectively. With $\mu^{(1)} = \Pi_{\mathfrak{M}}\mu$ and $\mu^{(2)} = \Pi_{\mathfrak{M}^\perp}\mu$ and $\delta_1 = \|\mu^{(1)}\|$, $\delta_2 = \|\mu^{(2)}\|$ it follows $\Phi_{\mu, I_n; g}\left(A(x)\right) = \Phi_{0, I_n; g}\left(C_{m,n-m,\delta_1,\delta_2}(x)\right)$ where

$$C_{m,n-m,\delta_1,\delta_2}(x) = \left\{ y \in \mathbb{R}^n : \frac{\left\|y^{(1)} + \mu^{(1)}\right\|^2}{\left\|y^{(2)} + \mu^{(2)}\right\|^2} < \frac{m}{n-m}\, x \right\}.$$

Let $\left(Y^{(1)}, Y^{(2)}\right) \sim \Phi_{0, I_n; g^G}$. The distribution of $Q = \|Y^{(1)} + \mu^{(1)}\|^2 / \|Y^{(2)} + \mu^{(2)}\|^2$ depends on the vectors $\mu^{(1)}$ and $\mu^{(2)}$ only through $\|\mu^{(1)}\|^2 = \delta_1^2$ and $\|\mu^{(2)}\|^2 = \delta_2^2$ because the noncentral chisquare distributions of $\|Y^{(1)} + \mu^{(1)}\|^2$ and $\|Y^{(2)} + \mu^{(2)}\|^2$ depend only on δ_1^2 or δ_2^2, respectively. The following definition is thus motivated.

Definition 2 A Borel set $A \subset \mathbb{R}^n$ will be said to belong to the *class* $\mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$ if there exists a function $t \,|\, \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ such that

$$A \cap S_n(r) = C_{m,n-m,\delta_1,\delta_2}(t(r)) \cap S_n(r), \; r > 0.$$

Remark 5 Clearly, $C_{m,n-m,\delta_1,\delta_2}(\nu) \in \mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$ with $t(r) \equiv \nu$ for all $r > 0$ and

$$F_{m,n-m,\delta_1^2,\delta_2^2;g}(x) = \Phi_{0, I_n; g}\left(C_{m,n-m,\delta_1,\delta_2}\left(x\right)\right).$$

We will give an additional description for the sets A from $\mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$. To this end let

$$Z_{m,n-m,\delta_1,\delta_2,r}(t(r)) = \left\{x \in \mathbb{R}^n : x_{m+1} > b_1\left(x_1^2 + \cdots + x_m^2\right) + b_2 x_1 + b_3\right\}$$

denote a certain parabolic cylinder type set from $\mathbb{R}^n$ where

$$b_1 = \frac{n-m}{2t(r)m\delta_2} + \frac{1}{2\delta_2},\ b_2 = \frac{(n-m)\delta_1}{t(r)m\delta_2},\ b_3 = -\frac{r^2+\delta_2^2}{2\delta_2} + \frac{(n-m)\delta_1^2}{2t(r)m\delta_2}$$

and where $x_1, \ldots, x_m$ and $x_{m+1}, \ldots, x_n$ denote the coordinates of x with respect to orthonormal bases in $\mathfrak{M}$ and $\mathfrak{M}^\perp$, respectively. The sets A from $\mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$ satisfy the equations

$$A \cap S_n(r) = Z_{m,n-m,\delta_1,\delta_2,r}(t(r)) \cap S_n(r),\ r > 0.$$

It follows that the i.p.f. for $C_{m,n-m,\delta_1,\delta_2}(t(r))$ coincides with that for the cylinder type set $Z_{m,n-m,\delta_1,\delta_2,r}\,(t(r))$, i.e.

$$\mathcal{F}_n\left(C_{m,n-m,\delta_1,\delta_2}\left(t(r)\right), r\right) = \mathcal{F}_n\left(Z_{m,n-m,\delta_1,\delta_2,r}\left(t(r)\right), r\right),\ r > 0.$$

Theorem 4 *The i.p.f. for a set A from the class $\mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$ with $m \geq 2$, $n-m \geq 2$ satisfies the representation formula*

$$\begin{aligned}
&\omega_n \mathcal{F}(A,r) = \omega_{m-1}\omega_{n-m-1} \\
&\times \int_0^\pi \int_0^\pi \int_0^{\pi/2} (\sin\phi_1)^{m-2} (\sin\phi_m)^{m-1} (\cos\phi_m)^{n-m-1} (\sin\phi_{m+1})^{n-m-2} \qquad (33) \\
&I\left\{b_1 r^2 (\sin\phi_m)^2 + b_2 r \sin\phi_m \cos\phi_1 - r\cos\phi_m \cos\phi_{m+1} + b_3 < 0\right\} d\phi_{m+1} d\phi_m d\phi_1,
\end{aligned}$$

were $\omega_k = 2\pi^{k/2}/\Gamma(k)$ denotes the surface area of $S_k(1)$.

The proof follows the general line in Richter (1991) and will therefore be omitted here.

Simulating the intersection percentage function

The numerical evaluation of the i.p.f. $\mathcal{F}_n(A,\cdot)$ for a given Borel set A is relatively timeconsuming, in general. So it is of some interest to simulate values of this function. To this end, one has to generate first N uniformly distributed random vectors $(x_1, \ldots, x_n)$ on the n-dimensional sphere with radius r. Second, one has to check, whether $(x_1, \ldots, x_n)$ belongs to $A \cap S_n(r)$ or not. This means in the case of Theorem 4, i.e. if $A \in \mathcal{A}_{m,n-m,\delta_1,\delta_2,t}$, that one has to check whether the condition

$$\frac{(x_1+\delta_1)^2 + x_2^2 + \cdots + x_m^2}{(x_{m+1}+\delta_2)^2 + x_{m+2}^2 + \cdots + x_n^2} < \frac{m}{n-m} t(r)$$

is fulfilled or not. The number of cases when the condition is fulfilled divided by N provides a simulated value for $\mathcal{F}_n(A,r) = \mathcal{F}_n(C_{m,n-m,\delta_1,\delta_2}\,(t(r)),r)$. For simulating the probabilities

$$\Phi(A;g) = \int_0^\infty \mathcal{F}_n(A,r) r^{n-1} g(r^2)\,dr \Big/ \int_0^\infty r^{n-1} g(r^2)\,dr$$

one needs the values of $\mathcal{F}_n(A,r)$ for r from zero up to r_{max}, where r_{max} depends on the tails of the elliptically contoured distribution, determined by g. Writing the values of $\mathcal{F}_n(A,r)$ for certain values of the parameters into files gives the possibility, to exploit the geometric representation formula with only one file for all admissible d.g.f. g. Here lies a main advantage of the geometric method. Instead of simulating new in each case random vectors x with certain elliptically contoured distributions one has only one times to simulate $\mathcal{F}_n(A,r)$ and can then approximate with this estimator the spherical measures $\Phi(A;g)$ for all d.g.f. g satisfying assumption (1).

Example 5 Two-dimensional geometric consideration shows that the sets CC_i^{**} from (13) and (14) belong to the classes $\mathcal{A}_{1,1,\delta_{1,i},\delta_{2,i},t_i}$, $i=1,2$, where the functions $t_i = t_i(r) = \frac{\sqrt{h_i}-m_i^2+1}{\sqrt{h_i}+m_i^2-1}$ actually do not depend on r and where the noncentrality parameters are

$$\begin{aligned}
\delta_{1,i}^2 &= \frac{a_i^2}{\sqrt{h_i}}\frac{m_i^2-\sqrt{h_i}+1}{m_i^2+\sqrt{h_i}-1}, \\
\delta_{2,i}^2 &= \frac{a_i^2}{\sqrt{h_i}}\frac{m_i^2\left(\sqrt{h_i}+m_i^2+1\right)+\varrho_i^2\left(-\sqrt{h_i}-3m_i^2+1\right)}{\left(1-\varrho_i^2\right)\left(\sqrt{h_i}+m_i^2-1\right)}
\end{aligned}$$

with m_i, ϱ_i, h_i and a_i being the same as in Theorem 1. Consequently,

$$P_i(CC_i)(c;g) = \Phi_{0_2,I_2;g_{n,2}}(CC_i^{**}) = F_{1,1,\delta_{1,i}^2,\delta_{2,i}^2;g_{n,2}}(t_i).$$

This result was the starting point for our general consideration in Section 3.

References

Ahrens, H., Läuter, J. (1981): *Mehrdimensionale Varianzanalyse*. Akademie-Verlag, Berlin.

Chou, Y.-M., Arthur, K., Rosenstein, R., Owen, D. (1985): New representations for the doubly noncentral F-distribution and derived distributions. *Commun. Statist.-Theor. Meth.*, 14, 3, 527–534.

Das Gupta, S. (1965): Optimum classification rules for classification into two multivariate normal populations. *Ann. Math. Statist.*, 36, 1174–1184.

Dorvlo, A. S. S. (1993): A unified method for computing probabilities of misclassification when the covariance matrices are known. *Commun. Statist.-Simula.*, 22, 3, 679–687.

Fang, K.-T., Kotz, S., Ng, K. (1990): *Symmetric multivariate and related distributions*. Chapman and Hall, London, New York.

Hills, M. (1966): Allocation rules and their error rates. *Journal of the Royal Statistical Society, Series B*, 28, 1–31.

Imhof, J. P. (1961): Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 3/4, 419–426.

Ittrich, C., Krause, D., Richter, W.-D. (2000): Probabilities and large quantiles of noncentral generalized chi-square distributions. *Statistics*, 34, 53–101.

John, S. (1960): On some classification statistics. *Sankhya*, 22, 309–317.

John, S. (1961): Errors in discrimination. *Annals Math. Statistics*, 32, 1125–1144.

Krause, D., Richter, W.-D. (1994a): Geometric approach to evaluating probabilities of correct classification into two Gaussian or spherical categories. In: Bock, H.-H., Lenski, W., Richter, M., eds., *Information Systems and Data Analysis*, pages 242–250. Springer-Verlag, Berlin.

Krause, D., Richter, W.-D. (1994b): Probabilities of correct classification. In: Herrendörfer, G., Miescke, K. J., eds., *Selection Procedures II*, pages 26–31. Schriftenreihe des FBN, Dummerstorf.

McLachlan, G. (1992): *Discriminant Analysis and Statistical Pattern Recognition*. Wiley & Sons, New York.

Moran, M. A. (1975): On the expectation of errors of allocation associated with a linear discriminant function. *Biometrika*, 62, 1, 141–148.

Mudholkar, G., Chaubey, Y., Lin, C.-C. (1976): Approximations for the doubly noncentral F-distribution. *Commun. Statist.-Theor. Meth.*, A5, 1, 49–63.

Press, W., Flannery, B., Teukolsky, S., Vetterling, W. (1989): *Numerical Recipes in Pascal*. Cambridge University Press, Cambridge.

Price, R. (1964): Some non-central F-distributions expressed in closed form. *Biometrika*, 51, 1/2, 107–122.

Richter, W.-D. (1985): Laplace–gauss integrals, gaussian measure asymptotic behavior and probabilities of moderate deviations. *Zeitschrift für Analysis und ihre Anwendungen*, 4, 3, 257–267.

Richter, W.-D. (1991): Eine geometrische Methode in der Stochastik. *Rostocker Mathematisches Kolloquium*, 44, 63–72.

Richter, W.-D. (1994): Geometric approach to evaluating probabilities of correct selection for spherically distributed samples. In: Herrendörfer, G., Miescke, K. J., eds., *Selection Procedures II*. Schriftenreihe des FBN, Dummerstorf.

Richter, W.-D. (1999): Classification for repeated measurements in Gaussian and related populations. In: Gaul, W., Locarek-Junge, H., eds., *Classification in the Information Age*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 241–251. Springer, Berlin.

Richter, W.-D. (2002): Central limit theorem for probabilities of correct classification. In: Gaul, W., Ritter, G., eds., *Classification, Automation, and New Media*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 123–130. Springer, Berlin.

Richter, W.-D., Schumacher, J. (2000): Asymptotic expansions for large deviation probabilities of noncentral generalized chi-square distributions. *Journal of Multivariate Analysis*, 75, 184–218.

Schaafsma, W., van Vark, G. (1977): Classification and discrimination problems with applications, Part I. *Statist. Neerlandica*, 31, 25–45.

Siotani, M. (1982): Large sample approximations and asymptotic expansions of classification statistics. In: Krishnaiah, P. R., Kanal, L., eds., *Handbook of Statistics (Vol. 2)*, pages 61–100. North-Holland, Amsterdam.