Annals of the Institute of Statistical Mathematics (2024) 76:511–534
https://doi.org/10.1007/s10463-024-00900-1

Using the growth curve model in classification of repeated measurements

Dietrich von Rosen[1,2] **· Martin Singull**[2]

Received: 8 March 2023 / Revised: 5 September 2023 / Accepted: 25 January 2024 /
Published online: 29 March 2024

Abstract
In this paper, discrimination between two populations following the growth curve model is considered. A likelihood-based classification procedure is established, in the sense that we compare the two likelihoods given that the new observation belongs to respective population. The possibility to classify the new observation as belonging to an unknown population is discussed, which is shown to be natural when considering growth curves. Several examples and simulations are given to emphasize this possibility.

Keywords Discriminant analysis · Growth curve model · Likelihood-based classification

1 Introduction

Pearson (1915) discussed the problem of sexing bones, where the analysis was based on the assumption that the population consisted of two overlapping populations. The Coefficient of Racial Likeness (CRL) was presented by Pearson and colleagues [e.g., see Pearson (1926)]. Based on measurements of bones, scalps, etc., the idea was to discriminate between races and an indicator was created, i.e., the CRL. Some years after the CRL first appeared, Mahalanobis (1936) introduced the idea of a distance measure between two populations. In an interesting contribution to Mahalanobis' work Dasgupta (1993) noted that Mahalanobis had been inspired by Seal (1911), who discussed group distances and variables as coordinates on a plane. At this point, Hotelling (1931) must also be mentioned. Hotelling developed a test for group differences based on correlated variables, and was inspired by Pearson and the CRL.

✉ Martin Singull
martin.singull@liu.se

[1] Department of Energy and Technology, Swedish University of Agricultural Science, Uppsala, Sweden

[2] Department of Mathematics, Linköping University, 581 83 Linköping, Sweden

The father of discriminant analysis Sir Ronald Fisher (1936), formulated and solved the mathematical problem of separating groups via a linear function of correlated variables. He also showed how to assign an observation into one of the groups. Fisher (1938) repeated many of the statements from the earlier article but also connected his work to Mahalanobis' and Hotelling's contributions. Moreover, Fisher (1938) discussed the direction of the observations from each group. In fact Barnard (1935), who communicated with Fisher, wanted to discriminate between groups when a linear regression appeared over time for the observed variables, which partly can be interpreted as a choice between directions. Barnard's approach can be said to be a univariate approach. Our research considers similar problems but it will take place within a likelihood ratio approach based on dependent variables. Wald (1944) had a clear derivation of a discrimination rule and discussed in detail distributional problems related to classification. In a very interesting paper, Rao (1948) clearly indicated that one purpose of discriminant analyses is to assign individuals to different populations with the possibility to assign the individual into an unknown population. Our article will show how this phenomenon in a specific case can be handled. Rao also connected discrimination to the Neyman–Pearson theory. An informative survey of discriminant analysis up to 1950 is presented by Hodges (1950), and 25 years later by Huberty (1975). The above preamble of discriminant analysis gives the background for this article. In particular it can be noted that the following had been discussed: (i) the separation of groups using linear regression (Barnard 1935; ii) the classification into an unknown population (Rao 1948). Rao (1973, Sect. 8e.2), there is a brief discussion of how to approach the problem with a third unknown population.

Multivariate analysis of variance (MANOVA) became included in text books in the 1950s and one-way MANOVA is of course linked to discrimination among groups. Moreover, the generalized multivariate analysis of variance (GMANOVA) started to be studied, which ended up in a general model by Potthoff and Roy (1964), i.e., the growth curve model was born. The model will constitute a basis in the proposed classification analysis. Estimation of the parameters of the growth curve model was considered by Khatri (1966), who derived the maximum likelihood estimators. However, there is a prehistory of the model where growth curves were estimated, e.g., see Rao (1958, 1959). Furthermore, in Rao (1958) the classification of growth curves was briefly mentioned. For references connected to the growth curve model see von Rosen (2018).

It is interesting to note that Burnaby (1966) discussed the classification of growth curves based on geometrical arguments including directions. However, Burnaby looked upon growth as a nuisance parameter when discussing classification, i.e., Burnaby wanted to exclude any growth effect when classifying observations. Rao had an influence on Burnaby's article and Rao (1966) provided several alternatives for how one can construct discriminant functions involving linear growth curves. In both Burnaby's and Rao's work, the dispersion is supposed to be known, or it can be estimated from "other" data sources. To some extent our contribution in this article is a direct follow-up of Burnaby's results, but now the likelihood is used, which indeed can be used to estimate the dispersion jointly with the construction of classification rule.

Kshirsagar and Davis (1983a, 1983b) [see also Albert and Kshirsagar (1993)] considered several populations and the construction of discriminant function for the populations via canonical correlations. The works are a continuation of ideas which originated in the 1930s [see Hotelling (1936), Bartlett (1938), Bartlett (1939), Bartlett (1951) and Kshirsagar (1962)]. Note that Kshirsagar and Davis (1983a, 1983b) and others do not consider the classification of objects. The aim was to find a discriminant function via classical testing of hypothesis. In this work, no testing of hypothesis is carried out. Instead the likelihood is studied where the observation which is to be classified is part of the likelihood. The criteria used in this article are expressed in the original matrices, which gives a clear interpretation of the results. Mentz and Kshirsagar (2004) studied classification via the growth curve model but their approach is not likelihood-based and differs from the one presented in the subsequent sections. Recently, Ngailo et al. (2021) considered Fisher's linear classification function for repeated measurements using the mean structure connected to the growth curve model, and derived misclassification errors. There is also a Bayesian literature connected to the growth curve model and Lee (1982) used the posterior distribution to classify linear growth curves. More references on Bayesian classification of growth curves can be found in Lee (1982).

For completeness, the growth curve model (Potthoff and Roy 1964) is presented:

$$\boldsymbol{X} = \boldsymbol{ABC} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}), \tag{1}$$

where $\boldsymbol{X} : p \times n$, $\boldsymbol{A} : p \times q$, $q < p$, $\boldsymbol{B} : q \times k$, $\boldsymbol{C} : k \times n$, $k < n$ and $\boldsymbol{E} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I})$. The unknown parameters are $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$. Moreover, $\boldsymbol{C}$ is the between-individuals design matrix which is the same design matrix as in linear models or MANOVA models, and $\boldsymbol{A}$ is the within individuals design matrix. The design matrix $\boldsymbol{C}$ will in our applications always be of full rank and later it will be indicated that without loss of generality it can be assumed that the rank of $\boldsymbol{A}$ equals q, i.e., $\boldsymbol{A}$ is of full rank.

Over the years, several specialized text books on classification have been written, e.g., McLachlan (2004), where a huge number of references on the subject can be found. Moreover, most books on multivariate analysis include classification and discriminant analysis (e.g., see Srivastava and Khatri 1979; Muirhead 1982; Anderson 2003). However, none of these books have information about classification of linear growth curves (McLachlan 2004, has a short Sect. 3.7.6).

This article considers the discrimination between two populations based on linear growth curves analysis, i.e., we have $k = 2$. Let the populations be denoted by Π_1 and Π_2 and then growth curve models are given for each population by

$$\Pi_1 : \boldsymbol{X}_1 = \boldsymbol{A}\boldsymbol{\beta}_1 \boldsymbol{1}'_{n_1} + \boldsymbol{E}, \qquad \boldsymbol{E} \sim N_{p,n_1}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_1}), \tag{2}$$

$$\Pi_2 : \boldsymbol{X}_2 = \boldsymbol{A}\boldsymbol{\beta}_2 \boldsymbol{1}'_{n_2} + \boldsymbol{E}, \qquad \boldsymbol{E} \sim N_{p,n_2}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2}), \tag{3}$$

where $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma}$ are unknown parameters, $\boldsymbol{1}_{n_i}$ is a vector of n_i ones, $'$ stands for the transpose, $N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ denotes the matrix normal distribution, where in the present work always $\boldsymbol{\Psi} = \boldsymbol{I}$ meaning that there are independent columns in $\boldsymbol{E}$, and $\boldsymbol{A}$: $p \times q$, $q < p$, is the within individuals design matrix which describes the growth structure.

The dispersion matrix $\boldsymbol{\Sigma}$ is supposed to be positive definite. The population $\boldsymbol{\Pi}_1$ consists of n_1 observations and $\boldsymbol{\Pi}_2$ consists of n_2 observations. The purpose with the analysis is to decide if a new observation $\boldsymbol{x}$ can be assigned to $\boldsymbol{\Pi}_1$, $\boldsymbol{\Pi}_2$ or to no one of these populations. A likelihood-based approach will be promoted which has not been taken place before when considering classification of linear growth curves.

2 Likelihood-based classification

Let $\boldsymbol{X}_i$, $i \in \{1,2\}$, be the random matrices corresponding to the observations from $\boldsymbol{\Pi}_i$, $i \in \{1,2\}$ and $\boldsymbol{x}$ is the random vector which corresponds to the observation which is to be classified. If $\boldsymbol{x} \in \boldsymbol{\Pi}_1$, consider the following two models:

$$
\begin{aligned}
&(\boldsymbol{X}_1 : \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\beta}_1 \mathbf{1}'_{n_1+1} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_1+1}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_1+1}),\\
&\boldsymbol{X}_2 = \boldsymbol{A}\boldsymbol{\beta}_2 \mathbf{1}'_{n_2} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_2}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2}),
\end{aligned}
$$

where $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{x}$ are jointly independently distributed. These two equations form the likelihood function when $\boldsymbol{x} \in \boldsymbol{\Pi}_1$. However, if $\boldsymbol{x} \in \boldsymbol{\Pi}_2$ the model is specified as

$$
\begin{aligned}
&(\boldsymbol{X}_2 : \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{\beta}_2 \mathbf{1}'_{n_2+1} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_2+1}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2+1}),\\
&\boldsymbol{X}_1 = \boldsymbol{A}\boldsymbol{\beta}_1 \mathbf{1}'_{n_1} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_1}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_1}),
\end{aligned}
$$

which then will be used to create the likelihood function. Let

$$
\begin{aligned}
&\boldsymbol{X} = (\boldsymbol{X}_1 : \boldsymbol{x} : \boldsymbol{X}_2), \quad \boldsymbol{C}_1 = \begin{pmatrix} \mathbf{1}'_{n_1+1} & \mathbf{0}'_{n_2} \\ \mathbf{0}'_{n_1+1} & \mathbf{1}'_{n_2} \end{pmatrix}, \quad \boldsymbol{C}_2 = \begin{pmatrix} \mathbf{1}'_{n_1} & \mathbf{0}'_{n_2+1} \\ \mathbf{0}'_{n_1} & \mathbf{1}'_{n_2+1} \end{pmatrix},\\
&\boldsymbol{S}_i = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1}\boldsymbol{C}_i)\boldsymbol{X}', \quad i \in \{1,2\},
\end{aligned}
\tag{4}
$$

where $\mathbf{0}_{n_i}$ denotes a vector of n_i zeroes. If $\boldsymbol{x} \in \boldsymbol{\Pi}_i$, $i \in \{1,2\}$, the corresponding likelihood can be written

$$
L_i(\boldsymbol{B}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{1}{2}(n_1+n_2+1)p} |\boldsymbol{\Sigma}|^{-\frac{1}{2}(n_1+n_2+1)} e^{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}_i)(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}_i)'\}},
$$

where $\boldsymbol{B} = (\boldsymbol{\beta}_1 : \boldsymbol{\beta}_2)$, $q \times 2$. If $\boldsymbol{x} \in \boldsymbol{\Pi}_i$, $i \in \{1,2\}$, when taking into account all observations, the maximum likelihood estimators of $\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}_i$ and $\boldsymbol{\Sigma}$ equal, without assuming any full rank conditions on $\boldsymbol{A}$, (e.g., see von Rosen 2018, Theorem 3.1),

$$
\begin{aligned}
\boldsymbol{A}\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i =& \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}_i^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}_i^{-1}\boldsymbol{X}\boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1}\boldsymbol{C}_i,\\
(n+1)\widehat{\boldsymbol{\Sigma}}_i =& (\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i)(\boldsymbol{X} - \boldsymbol{A}\widehat{\boldsymbol{B}}_i\boldsymbol{C}_i)'\\
=& \boldsymbol{S}_i + (\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}_i^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}_i^{-1})\boldsymbol{X}\boldsymbol{C}'_i(\boldsymbol{C}_i\boldsymbol{C}'_i)^{-1}\boldsymbol{C}_i\boldsymbol{X}'\\
&\times (\boldsymbol{I} - \boldsymbol{S}_i^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}_i^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'),
\end{aligned}
$$

where "$^-$" denotes an arbitrary generalized inverse and $n = n_1 + n_2$. The estimators $\widehat{\boldsymbol{B}}_i$ and $\widehat{\boldsymbol{\Sigma}}_i$ denote estimators of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, respectively, when $\boldsymbol{x} \in \boldsymbol{\Pi}_i$, $i \in \{1,2\}$.

Given an inner product defined via $\boldsymbol{S}$ let $\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}$ be the projection on the column vector space of $\boldsymbol{A}$. Note the well known fact that the projector $\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}$ as well as the above-given maximum likelihood estimators do not depend on the choice of generalized inverses. Moreover, $\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}$ will be the same for all matrices generating the column space of $\boldsymbol{A}$. Thus, in principle, it can be assumed that $\boldsymbol{A}$ is of full rank. The likelihood ratio statistic for classifying $\boldsymbol{x}$ is equivalent to

$$\lambda = \frac{|\widehat{\boldsymbol{\Sigma}}_1|}{|\widehat{\boldsymbol{\Sigma}}_2|} = \frac{|\boldsymbol{S}_1 + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_1})\boldsymbol{X}\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-1}\boldsymbol{C}_1\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}'_{\boldsymbol{A},\boldsymbol{S}_1})|}{|\boldsymbol{S}_2 + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_2})\boldsymbol{X}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^{-1}\boldsymbol{C}_2\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}'_{\boldsymbol{A},\boldsymbol{S}_2})|}. \tag{5}$$

Some manipulations of (5) yield

$$\begin{aligned}\lambda =& \frac{|\boldsymbol{S}_1|}{|\boldsymbol{S}_2|}\frac{|\boldsymbol{I} + (\boldsymbol{S}_1^{-1} - \boldsymbol{S}_1^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_1})\boldsymbol{X}\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-1}\boldsymbol{C}_1\boldsymbol{X}')(\boldsymbol{I} - \boldsymbol{P}'_{\boldsymbol{A},\boldsymbol{S}_1})|}{|\boldsymbol{I} + (\boldsymbol{S}_2^{-1} - \boldsymbol{S}_2^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_2})\boldsymbol{X}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^{-1}\boldsymbol{C}_2\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}'_{\boldsymbol{A},\boldsymbol{S}_2})|} \\ =& \frac{|\boldsymbol{S}_1|}{|\boldsymbol{S}_2|}\frac{|\boldsymbol{I} + (\boldsymbol{S}_1^{-1} - \boldsymbol{S}_1^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_1})\boldsymbol{X}\boldsymbol{C}_1'(\boldsymbol{C}_1\boldsymbol{C}_1')^{-1}\boldsymbol{C}_1\boldsymbol{X}')|}{|\boldsymbol{I} + (\boldsymbol{S}_2^{-1} - \boldsymbol{S}_2^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_2})\boldsymbol{X}\boldsymbol{C}_2'(\boldsymbol{C}_2\boldsymbol{C}_2')^{-1}\boldsymbol{C}_2\boldsymbol{X}'|},\end{aligned} \tag{6}$$

since $\boldsymbol{P}'_{\boldsymbol{A},\boldsymbol{S}_i}(\boldsymbol{S}_i^{-1} - \boldsymbol{S}_i^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_i}) = \boldsymbol{0}$, $i \in \{1,2\}$, and for any matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ of proper sizes $|\boldsymbol{I} + \boldsymbol{U}\boldsymbol{V}| = |\boldsymbol{I} + \boldsymbol{V}\boldsymbol{U}|$.

Now $\boldsymbol{S}_i^{-1} - \boldsymbol{S}_i^{-1}\boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}_i} = \boldsymbol{A}^o(\boldsymbol{A}^{o'}\boldsymbol{S}_i\boldsymbol{A}^o)^{-1}\boldsymbol{A}^{o'}$, $i \in \{1,2\}$, where $\boldsymbol{A}^o$ is any matrix of full rank spanning the orthogonal complement to the column space of $\boldsymbol{A}$. Moreover, $\boldsymbol{S}_i + \boldsymbol{X}\boldsymbol{C}_i'(\boldsymbol{C}_i\boldsymbol{C}_i')^{-1}\boldsymbol{C}_i\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{X}'$, $i \in \{1,2\}$. Thus the λ-ratio in (6) can be written

$$\lambda = \frac{|\boldsymbol{S}_1||(\boldsymbol{A}^{o'}\boldsymbol{S}_1\boldsymbol{A}^o)^{-1}||\boldsymbol{A}^{o'}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}^o|}{|\boldsymbol{S}_2||(\boldsymbol{A}^{o'}\boldsymbol{S}_2\boldsymbol{A}^o)^{-1}||\boldsymbol{A}^{o'}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{A}^o|} = \frac{|\boldsymbol{S}_1||(\boldsymbol{A}^{o'}\boldsymbol{S}_1\boldsymbol{A}^o)^{-1}|}{|\boldsymbol{S}_2||(\boldsymbol{A}^{o'}\boldsymbol{S}_2\boldsymbol{A}^o)^{-1}|}. \tag{7}$$

In order to classify an observation $\boldsymbol{x}$ into Π_1, because $|\boldsymbol{\Sigma}|^{-1/2}$ appears in the normal density, the ratio λ should be small, i.e., $\lambda < 1$.

The next step will be to study the ratio $|\boldsymbol{S}_1|/|\boldsymbol{S}_2|$. Put the pooled sum of squares matrix based on the joint samples from Π_1 and Π_2, without taking into account that the new observation $\boldsymbol{x}$ belongs to one of these two populations, as

$$\boldsymbol{S} = \boldsymbol{X}_1\boldsymbol{X}_1' - n_1\bar{\boldsymbol{x}}_1\bar{\boldsymbol{x}}_1' + \boldsymbol{X}_2\boldsymbol{X}_2' - n_2\bar{\boldsymbol{x}}_2\bar{\boldsymbol{x}}_2', \tag{8}$$

where the averages are given by $\bar{\boldsymbol{x}}_i = n_i^{-1}\boldsymbol{X}_i\boldsymbol{1}_{n_i}$ for $i \in \{1,2\}$. The matrix $\boldsymbol{S}$ is Wishart distributed, with $n_1 + n_2 - 2$ degrees of freedom, which is denoted $\boldsymbol{S} \sim W_p(\boldsymbol{\Sigma}, n_1 + n_2 - 2)$.

The following lemma will be applied several times throughout the work.

Lemma 1 *Let $\boldsymbol{S}_1, \boldsymbol{S}_2$ and $\boldsymbol{S}$ be defined in (4) and (8). Then*

$$\boldsymbol{S}_i = \boldsymbol{S} + \frac{n_i}{n_i + 1}(\bar{\boldsymbol{x}}_i - \boldsymbol{x})(\bar{\boldsymbol{x}}_i - \boldsymbol{x})', \quad i \in \{1,2\}.$$

Calculations to establish the lemma are the same as when establishing a likelihood-based classification rule in MANOVA Anderson (2003, Sect. 6.5.5). The given relation in Lemma 1 implies that

$$\frac{|\boldsymbol{S}_1|}{|\boldsymbol{S}_2|} = \frac{|\boldsymbol{S} + \frac{n_1}{n_1+1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})'|}{|\boldsymbol{S} + \frac{n_2}{n_2+1}(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})'|} = \frac{1 + \frac{n_1}{n_1+1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})}{1 + \frac{n_2}{n_2+1}(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})}. \tag{9}$$

According to (7) the ratio $|\boldsymbol{A}^{o'}\boldsymbol{S}_2\boldsymbol{A}^o|/|\boldsymbol{A}^{o'}\boldsymbol{S}_1\boldsymbol{A}^o|$ should also be considered. Lemma 1 implies

$$\boldsymbol{A}^{o'}\boldsymbol{S}_i\boldsymbol{A}^o = \boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^o + \frac{n_i}{n_i+1}\boldsymbol{A}^{o'}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{A}^o, \quad i \in \{1,2\}$$

and $\boldsymbol{A}^o(\boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^o)^{-1}\boldsymbol{A}^{o'} = \boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{S}^{-1}$. Thus

$$\begin{aligned}|\boldsymbol{A}^{o'}\boldsymbol{S}_i\boldsymbol{A}^o| = |\boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^o|\big(1 + \frac{n_i}{n_i+1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x}) \\ - \frac{n_i}{n_i+1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})\big), \quad i \in \{1,2\}.\end{aligned} \tag{10}$$

One should note that the factor $|\boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^o|$ cancels when taking the ratio $|\boldsymbol{A}^{o'}\boldsymbol{S}_2\boldsymbol{A}^o|/|\boldsymbol{A}^{o'}\boldsymbol{S}_1\boldsymbol{A}^o|$. Returning to the likelihood ratio in (7), the next theorem follows from (7)–(10).

Theorem 1 *Let the "likelihood ratio" λ be given by (7). Then*

$$\lambda = \frac{H_2}{H_1},$$

where

$$H_i = 1 - \frac{\frac{n_i}{n_i+1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})}{1 + \frac{n_i}{n_i+1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})}, \quad i \in \{1,2\}.$$

The classification rule states that $\boldsymbol{x}$ will be classified into Π_1 if $\lambda < 1$ and thus it is of interest to look closer at the inequality $H_2 < H_1$, which is equivalent to

$$\frac{L_1}{L_2} < \frac{1 + \frac{n_1}{n_1+1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{x})}{1 + \frac{n_2}{n_2+1}(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_2 - \boldsymbol{x})}, \tag{11}$$

where

$$L_i = \frac{n_i}{n_i+1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_i - \boldsymbol{x}), \quad i \in \{1,2\}. \tag{12}$$

The relation in (11) can be shown by noting that $H_2 < H_1$ is equivalent to

$$\frac{\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})}{1+\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})} > \frac{\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})}{1+\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})}.$$

From von Rosen (2018, Theorem 3.1) it follows that the maximum likelihood estimators of $\boldsymbol{\beta}_i$, $i \in \{1,2\}$, based on the samples from $\boldsymbol{\Pi}_i$, satisfy

$$\boldsymbol{A}\widehat{\boldsymbol{\beta}}_i = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}_i, \quad i \in \{1,2\}.$$

If it is of interest to estimate $\widehat{\boldsymbol{\beta}}_i$ then $\boldsymbol{A}$ has to be of full rank. The ratio L_1/L_2 can be written

$$\frac{L_1}{L_2} = \frac{n_1(n_2+1)}{n_2(n_1+1)} \frac{(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})}{(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})}. \tag{13}$$

Now the relations in (11) and (13) will be utilized to classify an observation where the mean follows the growth curve model.

Proposition 1 *All matrices used below have been introduced in Sect. 2. The n_1 observations in Population 1 ($\boldsymbol{\Pi}_1$) follow the model in (2), whereas the n_2 observations from Population 2 ($\boldsymbol{\Pi}_2$) follow the model in (3). Let $\boldsymbol{x}$ be a new observation which according to the following criteria is to be classified into $\boldsymbol{\Pi}_1$ or $\boldsymbol{\Pi}_2$ or it is classified to belong to an unknown population.*

The unknown observation $\boldsymbol{x}$ is classified into $\boldsymbol{\Pi}_1$ if the following two conditions are satisfied:

$$\begin{aligned}
&\text{(i)} \quad \tfrac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x}) < \tfrac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x}) \\
&\text{(ii)} \quad (\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x}) \\
&\qquad\qquad < c_0(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x}),
\end{aligned}$$

where

$$c_0 = \frac{n_2(n_1+1)}{n_1(n_2+1)} \frac{1+\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})}{1+\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})}.$$

Remark 1 If only one inequality, either (i) or (ii), is fulfilled the new observation $\boldsymbol{x}$ is classified as belonging to some unknown population. Furthermore, if none of the inequalities are fulfilled the new observation $\boldsymbol{x}$ is classified into $\boldsymbol{\Pi}_2$.

A plain likelihood ratio criterion would only have been based on (ii), which is natural because there is no contribution to the likelihood from observations from an unknown population. However, through the proposed method and the likelihood, more information about the observations can be extracted. Unfortunately, deriving misclassification errors will be difficult, but one can achieve results via the derivation of misclassification errors in both (i) and (ii) and the Bonferroni inequality.

In a way, the proposal is natural because the growth curve model is a bilinear model and in the data there are two types of observed variations (deviation from the model). First, it is the variation of the observations around the sample means [treated via (i)] and second, it is the variation of the sample means around the estimated model [treated via (ii)]. This has led to the proposed method presented in Proposition 1. In Fig. 1, several cases are illustrated where the observation which is to be classified does not belong to any of the two populations [see (b)–(d)].

Already Rao (1948) noted the importance of classifying into an unknown population but this case has received little attention over the years. If in Proposition 1 both (i) and (ii) suggest classifying the observation in, say, Population 1, this may still not be correct. Therefore, Proposition 1 has to be complemented in order to accurately classify the cases in Fig. 1 (of course still with some kind of probabilistic misclassification error). The reason for this is that $\widehat{\boldsymbol{\beta}}_i$ is of size q and some components in $\widehat{\boldsymbol{\beta}}_i$ can suggest classifying $\boldsymbol{x}$ into population $\boldsymbol{\Pi}_1$, whereas other components suggest classifying $\boldsymbol{x}$ into population $\boldsymbol{\Pi}_2$. In Fig. 1b this phenomenon is illustrated where the

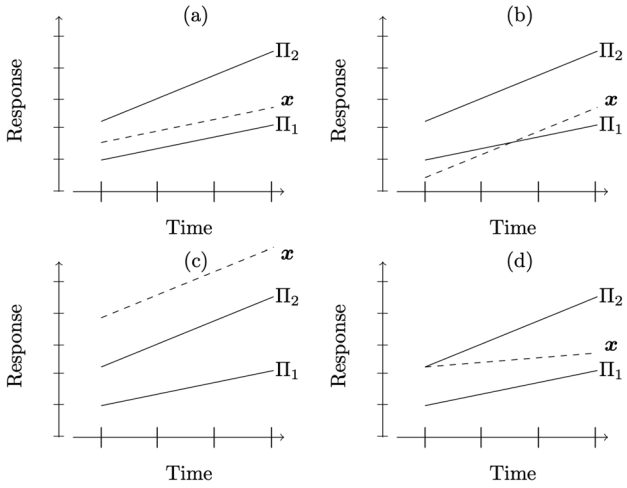

Fig. 1 In two populations, a response has been measured at four time points. In the present figures, $\boldsymbol{\Pi}_1$ represents Population 1, $\boldsymbol{\Pi}_2$ represents Population 2 and $\boldsymbol{x}$ (dashed line) stands for the new observation vector which is to be classified. In **a** it is shown that $\boldsymbol{x}$ belongs to $\boldsymbol{\Pi}_1$, whereas in **b** it is illustrated when $\boldsymbol{x}$ is close to the observations in $\boldsymbol{\Pi}_1$ but the growth follows the one of $\boldsymbol{\Pi}_2$. Moreover, it is shown in **c** when $\boldsymbol{x}$ follows the growth of $\boldsymbol{\Pi}_2$ but deviates from this population. In **d** it is indicated that $\boldsymbol{x}$ neither follows the growth of $\boldsymbol{\Pi}_1$ nor the growth of $\boldsymbol{\Pi}_2$. Thus only in **a** the new observation can be assigned to a population

intercept suggests classifying $\boldsymbol{x}$ into one population but the slope suggests classifying into the other population. Note that when $\widehat{\boldsymbol{\beta}}_i$ is discussed it is assumed that $\boldsymbol{A}$ is of full rank, otherwise estimable functions have to be considered.

3 Comparison with other classification approaches

Over the years, many classification methods have been proposed. In this section four different methods, involving different classification functions, will be presented and compared with the approach in Proposition 1. These methods are all connected to the linear discriminant function. Following the derivation in Anderson (2003), with two normal populations and the same dispersion matrix, $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, say Π_1 and Π_2 and an unknown observation $\boldsymbol{x}$, the following classification function appears:

$$d = \boldsymbol{x}' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \tag{14}$$

If $d > 0$, $\boldsymbol{x}$ is classified to belong to Π_1 and otherwise $\boldsymbol{x}$ is classified into Π_2. Since $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are unknown these parameters have to be estimated via samples from Π_1 and Π_2. For example, if $\boldsymbol{X}_1 \in \Pi_1$, $\boldsymbol{X}_2 \in \Pi_2$,

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{\mu}_1 \mathbf{1}'_{n_1} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_1}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_1}), \\ \boldsymbol{X}_2 &= \boldsymbol{\mu}_2 \mathbf{1}'_{n_2} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_2}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2}), \end{aligned}$$

then maximum likelihood estimators are given by $\widehat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{x}}_1$, $\widehat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{x}}_2$ and ($n = n_1 + n_2$)

$$\begin{aligned} n\widehat{\boldsymbol{\Sigma}} &= \boldsymbol{S} = (\boldsymbol{X}_1 : \boldsymbol{X}_2)(\boldsymbol{I}_n - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C})(\boldsymbol{X}_1 : \boldsymbol{X}_2)', \\ \boldsymbol{C} &= \begin{pmatrix} \mathbf{1}'_{n_1} & \mathbf{0}'_{n_2} \\ \mathbf{0}'_{n_1} & \mathbf{1}'_{n_2} \end{pmatrix}, \end{aligned}$$

where $\boldsymbol{S}$ is the same sums of squares matrix as in (8). Thus, $\boldsymbol{x}$ is classified into Π_1 if $D > 0$, where

$$n^{-1}D = \boldsymbol{x}'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2). \tag{15}$$

Burnaby (1966) considered the following samples:

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{A}\boldsymbol{\beta}_1 \mathbf{1}'_{n_1} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_1+1}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_1}), \\ \boldsymbol{X}_2 &= \boldsymbol{A}\boldsymbol{\beta}_2 \mathbf{1}'_{n_2} + \boldsymbol{E}, \quad \boldsymbol{E} \sim N_{p,n_2}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2}), \end{aligned}$$

where $\boldsymbol{A}$ is the known within individuals design matrix modeling as before the responses' growth. Burnaby looked upon the growth as a nuisance factor, and to remove this factor, he introduced an orthogonal transformation (projection) perpendicular to the growth direction, i.e., $\boldsymbol{A}^{o'}\boldsymbol{X}_i$, which also implies that its dispersion equals $\boldsymbol{A}^{o'}\boldsymbol{\Sigma}\boldsymbol{A}^{o}$. Although Burnaby's idea was to estimate $\boldsymbol{\Sigma}$ via some other data set, it is known today that $n^{-1}\boldsymbol{S}$ (this is not the maximum likelihood estimator) can be used in the following classification function (Burnaby did not explicitly stated this result):

$$\begin{aligned} n^{-1}D_B = {} & \boldsymbol{x}'\boldsymbol{A}^{o}(\boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^{o})^{-1}\boldsymbol{A}^{o'}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \\ & - \frac{1}{2}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)'\boldsymbol{A}^{o}(\boldsymbol{A}^{o'}\boldsymbol{S}\boldsymbol{A}^{o})^{-1}\boldsymbol{A}^{o'}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2), \end{aligned} \tag{16}$$

where $\boldsymbol{x}$ is classified into Π_1 if $D_B > 0$.

Mentz and Kshirsagar (2004) also considered (14) where the following estimators were inserted ($\boldsymbol{A}$ was supposed to be of full rank which is not necessary):

$$\begin{aligned} \widehat{\boldsymbol{B}} &= (\widehat{\boldsymbol{\beta}}_1 \,:\, \widehat{\boldsymbol{\beta}}_2) = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}(\overline{\boldsymbol{x}}_1 \,:\, \overline{\boldsymbol{x}}_2), \\ n\widehat{\boldsymbol{\Sigma}}_{MK} &= ((\boldsymbol{X}_1 \,:\, \boldsymbol{X}_2) - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C})((\boldsymbol{X}_1 \,:\, \boldsymbol{X}_2) - \boldsymbol{A}\widehat{\boldsymbol{B}}\boldsymbol{C})'. \end{aligned}$$

The classification in (14) with these estimators equals

$$\begin{aligned} D_{MK} = {} & \boldsymbol{x}'\boldsymbol{A}(\boldsymbol{A}'\widehat{\boldsymbol{\Sigma}}_{MK}\boldsymbol{A})^{-1}\boldsymbol{A}'(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \\ & - \frac{1}{2}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)'\boldsymbol{A}(\boldsymbol{A}'\widehat{\boldsymbol{\Sigma}}_{MK}\boldsymbol{A})^{-1}\boldsymbol{A}'(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \end{aligned} \tag{17}$$

and if $D_{MK} > 0$ the observation $\boldsymbol{x}$ is classified into Π_1. A modification of this result is to use maximum likelihood estimators

$$\begin{aligned} \widehat{\boldsymbol{B}}_{ML} &= (\widehat{\boldsymbol{\beta}}_1 \,:\, \widehat{\boldsymbol{\beta}}_2) = (\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1 \,:\, \overline{\boldsymbol{x}}_2), \\ n\widehat{\boldsymbol{\Sigma}}_{ML} &= ((\boldsymbol{X}_1 \,:\, \boldsymbol{X}_2) - \boldsymbol{A}\widehat{\boldsymbol{B}}_{ML}\boldsymbol{C})((\boldsymbol{X}_1 \,:\, \boldsymbol{X}_2) - \boldsymbol{A}\widehat{\boldsymbol{B}}_{ML}\boldsymbol{C})'. \end{aligned}$$

Since $\boldsymbol{A}'\widehat{\boldsymbol{\Sigma}}_{ML}^{-1} = n\boldsymbol{A}'\boldsymbol{S}^{-1}$ the classification function equals

$$\begin{aligned} n^{-1}D_{ML} = {} & \boldsymbol{x}'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \\ & - \frac{1}{2}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2). \end{aligned} \tag{18}$$

If $D_{ML} > 0$ the observation $\boldsymbol{x}$ is classified into Π_1. It can be noted that $D = D_B + D_{ML}$, which is in agreement with intuition because with D_{ML} growth curves are studied, whereas Burnaby wanted to remove the growth effects when classifying observations.

In the next example a number of scenarios are discussed.

Example 1 In this example, two populations are considered and the aim is to illustrate the proposed method in Proposition 1, as well as to compare the results with

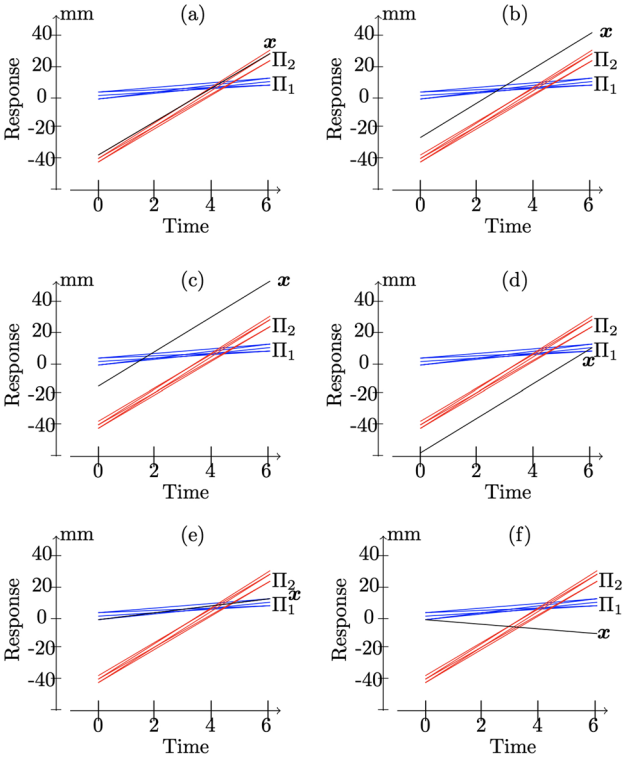

Fig. 2 In two populations a response deviating from baseline has been measured at four time points. In the present figures Π_1 (blue lines) represents Population 1, Π_2 (red lines) represents Population 2 and $\boldsymbol{x}$ (black line) stands for the new observation vector which is to be classified. In **a** it is shown that $\boldsymbol{x}$ belongs to Π_2, whereas in **b** it is illustrated when the response for $\boldsymbol{x}$ is shifted a bit from the observations in Π_2 and has come closer to the observations in Π_1, in comparison with **a**. In **c** the response for $\boldsymbol{x}$ has been shifted away from Π_2 more than in **b** whereas in **d** the response $\boldsymbol{x}$ has been shifted away from both the observations in Π_2 and Π_1. In **e** $\boldsymbol{x}$ belongs to Π_1 and in **f** $\boldsymbol{x}$ does not belong to any of the two given populations (colour figure online)

the earlier proposed methods presented above. Simulated data are used and the ideas are illustrated in Fig. 2. The following matrices have been used to generate data:

$$\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 4 & 6 \end{pmatrix}', \quad \boldsymbol{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = \begin{pmatrix} 1 & -40 \\ 1 & 10 \end{pmatrix},$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 10 & 5 & 2.50 & 1.25 \\ & 10 & 5 & 2.50 \\ & & 10 & 5 \\ & & & 10 \end{pmatrix}.$$

Using the matrices, 10 observations were generated from both populations which is in fact a small data set. From Population 1

$$\boldsymbol{X}_1 = \begin{pmatrix} 1.180192 & 3.330091 & 2.284808 & 0.5316934 & -7.324091 & 4.287237 \\ 0.963614 & 3.172805 & 1.613572 & 4.5971772 & -2.381949 & 2.014789 \\ 4.265160 & -1.586984 & 2.916197 & 1.8198309 & 4.239860 & 4.136507 \\ 1.148597 & 2.255889 & 11.616286 & 4.9593502 & 2.364634 & 10.201689 \\ -5.4962367 & 2.2856576 & 3.613856 & 4.134668 & & \\ 0.4055098 & 0.2124452 & 7.358171 & 4.918653 & & \\ 6.2437675 & 1.6049673 & 7.070861 & 5.996032 & & \\ 10.8431665 & 6.2032100 & 7.062818 & 2.647788 & & \end{pmatrix}$$

and from Population 2

$$\boldsymbol{X}_2 = \begin{pmatrix} -39.045120 & -40.929413 & -42.728002 & -31.026729 & -38.19440 & -40.21661 \\ -20.429371 & -20.086069 & -24.712534 & -18.288272 & -23.41076 & -21.42750 \\ -2.145189 & 0.276048 & -4.176433 & 6.028609 & -0.90917 & -1.54939 \\ 12.039531 & 20.310317 & 15.409145 & 25.366647 & 21.78737 & 22.63965 \\ -43.846127 & -41.193042 & -44.820454 & -40.031872 & & \\ -25.696265 & -19.383592 & -15.916768 & -19.625143 & & \\ -4.278519 & 5.790712 & 0.299045 & 6.308488 & & \\ 19.957922 & 24.849704 & 20.049941 & 20.884688 & & \end{pmatrix}.$$

Using the above data, the following estimators were obtained (note that the number of observations is small in this study)

$$\widehat{\boldsymbol{\beta}}_1 = \begin{pmatrix} 0.69 \\ 0.83 \end{pmatrix}, \quad \widehat{\boldsymbol{\beta}}_2 = \begin{pmatrix} -40.67 \\ 10.13 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 14.1 & 4.4 & 2.7 & 2.7 \\ & 7.7 & 4.0 & 1.5 \\ & & 10.9 & 5.9 \\ & & & 14.2 \end{pmatrix}.$$

Moreover it is noted that

$$\boldsymbol{S} = \begin{pmatrix} 279.9 & 90.6 & 52.3 & 53.4 \\ & 151.3 & 83.6 & 30.9 \\ & & 210.6 & 117.2 \\ & & & 283.6 \end{pmatrix}.$$

In Fig. 2, the data in $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are illustrated but for visibility reasons the real values are not presented (the variance indicated in the figures is too small). Corresponding to the six scenarios in Fig. 2, the observation which is to be classified equals

$$\text{Scenario 1, Fig. 2a} : \boldsymbol{x} = (-40, -20, 0, 20)'; \tag{19}$$

$$\text{Scenario 2, Fig. 2b} : \boldsymbol{x} = (-30, -10, 10, 30)'; \tag{20}$$

$$\text{Scenario 3, Fig. 2c} : \boldsymbol{x} = (-20, 0, 20, 40)'; \tag{21}$$

Table 1 The classification functions D, D_B, D_{MK}, D_{ML} and C_1/C_2, respectively, defined in (15)–(18), (25) and (26) are evaluated for each of the six classification scenarios presented in (19)–(24)

Scenario	D	D_B	D_{MK}	D_{ML}	C_1	C_2
Scenario 1, Fig. 2a	−0.21	0.0002	−85.3	−0.21	−7.78	−8.55
Scenario 2, Fig. 2b	−0.15	0.0002	−59.2	−0.15	−5.29	−3.05
Scenario 3, Fig. 2c	0.0014	0.0002	−2.4	0.001	0.05	−0.001
Scenario 4, Fig. 2d	−0.28	0.0002	−111.4	−0.28	−10.27	−6.25
Scenario 5, Fig. 2e	0.22	0.0002	87.2	0.22	8.01	0.98
Scenario 6, Fig. 2f	0.26	−0.00005	106.1	0.26	9.56	0.91

Note that negative values for D, D_B, D_{MK}, D_{ML} support classification into Population 2 (Π_2) whereas a positive value for C_1 and C_2 also supports a classification into Population 2

$$\text{Scenario 4, Fig. 2d} : \boldsymbol{x} = (-50, -30, -10, 10)'; \tag{22}$$

$$\text{Scenario 5, Fig. 2e} : \boldsymbol{x} = (1, 3, 5, 7)'; \tag{23}$$

$$\text{Scenario 6, Fig. 2f} : \boldsymbol{x} = (1, -4, -7, -10)'. \tag{24}$$

The classification criteria, which are used when classifying $\boldsymbol{x}$ will now be listed. From Proposition 1 two criteria, C_1 and C_2, are obtained as

$$C_1 = \frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x}) - \frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}), \tag{25}$$

$$\begin{aligned} C_2 = {} & c_0(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x}) \\ & - (\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x})'\boldsymbol{S}^{-1}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{P}_{\boldsymbol{A},\boldsymbol{S}}\boldsymbol{x}), \end{aligned} \tag{26}$$

where

$$c_0 = \frac{n_2(n_1+1)}{n_1(n_2+1)} \frac{1 + \frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})}{1 + \frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x})}.$$

If both C_1 and C_2 are positive, $\boldsymbol{x}$ is classified to Π_1 and if both C_1 and C_2 are negative $\boldsymbol{x}$ is classified to Π_2. Furthermore, if one is positive and the other negative, then the new observation $\boldsymbol{x}$ is classified to an unknown population. The six scenarios have been analyzed and in Table 1, and the results for D, D_B, D_{MK} and D_{ML} are reported together with C_1 and C_2.

It can once again be remarked that D_B differs from the other criteria, which is natural because D_B treats the growth effect as a nuisance parameter. Moreover, D_B is constant for Scenario 1–5 because $\boldsymbol{A}^{o'}\boldsymbol{x} = \boldsymbol{0}$. Except for Scenario 3, the same

conclusions about assigning $\boldsymbol{x}$ appear for D, D_{MK}, D_{ML} and C_1/C_2. In Scenario 3, C_1/C_2 classifies $\boldsymbol{x}$ into an unknown population, D and D_{ML} into Population 1, whereas D_{MK} assigns $\boldsymbol{x}$ into Population 2.

We continue with another example, which will be a small simulation study to compare and understand the performance for the different classifiers presented above.

Example 2 To compare the performance of the different classification methods, D, D_B, D_{MK}, D_{ML} and C_1/C_2, we find the outcomes from each method, using the same simulated data set for all of them. Let $q = 2$, $p = 5$ with $n_1 = n_2 = 25$, i.e., linear growth, five repeated measurements, at time points $t_i = i, i \in \{1, \dots, 5\}$, for two groups, each consisting of 25 independent observations. The considered mean parameter matrix equals

$$\boldsymbol{B} = \begin{pmatrix} 0 & -2 \\ 0.75 & 1.5 \end{pmatrix},$$

and the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{Q}'\boldsymbol{Q} + \sigma^2\boldsymbol{I}_p$, where $\boldsymbol{Q} = \left(q_{ij}\right)_{i,j\in\{1,\dots,p\}}$ with q_{ij} uniformly distributed in $[-2, 2]$ and $\sigma^2 = 1$.

For 1000 replications we note: (i) the decision for a new observation from Population 1, and (ii) the decision for new observations from three different "unknown" populations, which are characterizing the mean models

Case 1: between the two populations, i.e., the mean of $\boldsymbol{A}\boldsymbol{\beta}_1$ and $\boldsymbol{A}\boldsymbol{\beta}_2$;
Case 2: "far away" from the populations, i.e., far from $\boldsymbol{A}\boldsymbol{\beta}_1$ and $\boldsymbol{A}\boldsymbol{\beta}_2$;
Case 3: not in the linear subspace given by the columns of $\boldsymbol{A}$, e.g., a quadratic mean model.

At first, we consider the case when we simulate new observations from Population 1. In Fig. 3 (left) we can see one example of such a simulation and in Table 2 the proportions of simulations, when the new observation is classified to Population 1, Population 2 or as unknown. By comparing the proportions, one can conclude that all the rules perform similar. However, C_1/C_2, i.e., the proposed method in Proposition 1, has classified 6.0% of the simulations as unknown, which in this case is a

Table 2 The performance of the classification functions D, D_{MK}, D_{ML} and C_1/C_2, respectively, defined in (15), (17), (18), (25) and (26)

Population	D (%)	D_{ML} (%)	D_{MK} (%)	C_1/C_2 (%)
1	74.8	76.0	71.6	72.5
2	25.2	24.0	28.4	21.5
Unknown	–	–	–	6.0

The proportions given in the table, are classification proportions for new observations from Population 1, i.e., if new observations are classified into Population 1, Population 2 or classified as belonging to an unknown population

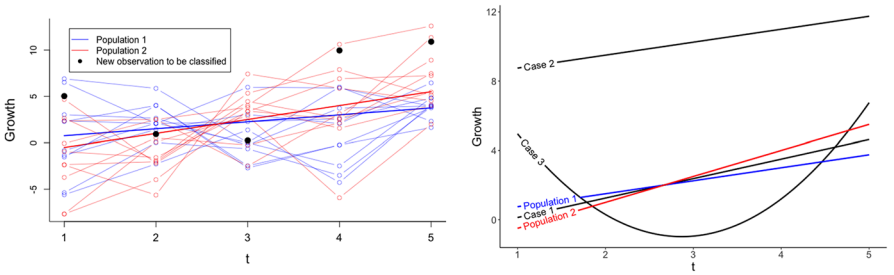

Fig. 3 *Left:* in Example 2 we simulated 25 profiles from each population. In this figure ten selected repeated measurement from each population is plotted together with the true growths (complete lines) and an observation which is to be classified. *Right:* the three different models, for which a new observation should be classified as unknown

wrong decision. The misclassification error for C_1/C_2 is the sum of the proportions for classifying into Population 2 and the proportion for classifying as unknown.

Secondly, we simulate new observations from the three different cases, discussed above and which can be seen in Fig. 3 (right). For those cases we would like our classifier to identify the new observations as belonging to an unknown population. In Table 3 we can see that this happens, but it seems not to be good enough and there is more information to gain from the data as we will see in Sect. 4. Furthermore, one can note that in Case 1, the classifications are roughly 50/50, which seems to be reasonable since the mean model for the new observation is just between the two populations. Similar result for Case 2, since the mean model is "far away" from the populations. In both cases the proposed rule with C_1/C_2 gives a proportion of

Table 3 The performance of the classification functions D, D_{MK}, D_{ML} and C_1/C_2, respectively, defined in (15), (17), (18), (25) and (26)

Population	D (%)	D_{MK} (%)	D_{ML} (%)	C_1/C_2 (%)
Case 1				
1	51.4	51.1	50.8	45.4
2	48.6	48.9	49.2	42.8
Unknown	–	–	–	11.8
Case 2				
1	41.2	41.5	37.7	35.2
2	58.8	58.5	62.3	52.4
Unknown	–	–	–	12.4
Case 3				
1	93.5	94.3	70.1	91.6
2	6.5	5.7	29.9	3.9
Unknown	–	–	–	4.5

The proportions given in the table, are classification proportions for new observations from the three cases seen in Fig. 3 and described in Example 2, i.e., if the observations are classified into Population 1, Population 2 or as coming from an unknown population

simulations classified to an unknown population, roughly 12%, but lower (4.5%) when we consider the quadratic mean model.

4 More criteria for classifying into an unknown population

Already in the end of Sect. 2 it was noted, since the size of the mean parameters $\boldsymbol{\beta}_i$, $i \in \{1,2\}$, is q that many kind of differences can appear. In this section we propose some new ideas how to control for different effects in $\boldsymbol{\beta}_i$, $i \in \{1,2\}$.

It is informative to calculate the expectation of the quadratic forms in (13) (given $\boldsymbol{S}$). The results presented in the next lemma will be used in what follows. The proof is based on knowledge about moments for quadratic forms in normally distributed variables and the first moment for the inverse Wishart distribution. Note, for $i \in \{1,2\}$,

$$\boldsymbol{y}_i = \sqrt{\frac{n_i}{n_i+1}}(\overline{\boldsymbol{x}}_i - \boldsymbol{x}) \sim N_p\left(\sqrt{\frac{n_i}{n_i+1}}(\boldsymbol{A}\boldsymbol{\beta}_i - E[\boldsymbol{x}]), \boldsymbol{\Sigma}\right), \tag{27}$$

and

$$\boldsymbol{S} \sim W_p(\boldsymbol{\Sigma}, n-2), \tag{28}$$

which are independent. Moreover, since $E[\boldsymbol{x}]$ equals either $\boldsymbol{A}\boldsymbol{\beta}_i$ for $i \in \{1,2\}$, $E[\boldsymbol{x}]$ belongs to the column space of $\boldsymbol{A}$.

Lemma 2 *For $\boldsymbol{y}_i$, $\boldsymbol{x}$ and $\boldsymbol{S}$ in (27) and (28), then*

(i) for $\boldsymbol{Q}$: $p \times p$, and $\boldsymbol{A}$: $p \times q$, $q < p$,

$$E[\boldsymbol{y}_i'\boldsymbol{Q}\boldsymbol{y}_i] = \mathrm{tr}\{\boldsymbol{Q}\boldsymbol{\Sigma}\} + \frac{n_i}{n_i+1}(\boldsymbol{A}\boldsymbol{\beta}_i - E[\boldsymbol{x}])'\boldsymbol{Q}(\boldsymbol{A}\boldsymbol{\beta}_i - E[\boldsymbol{x}]);$$

(ii) for $\boldsymbol{A}$: $p \times q$ of rank q and $n > p-3$

$$E[\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}] = \frac{1}{n-p+q-3}\left(\frac{q}{n-p-3}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\right).$$

Using a vector space decomposition, the next lemma can be proved (it is a special case of Theorem 1.2.25 in Kollo and von Rosen (2005)).

In order to handle the situation with many parameters, the following lemma will be utilized.

Lemma 3 *Let $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)$ be of rank q and $\boldsymbol{A}_{(k)}$ equals $\boldsymbol{A}$ without the column $\boldsymbol{a}_k$. Let $\boldsymbol{T}_k = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_k^o(\boldsymbol{H}_k^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_k^o)^{-1}\boldsymbol{H}_k^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}$, where $\boldsymbol{H}_k = \boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}_{(k)}$. Then, for $k \in \{1, \ldots, q\}$,*

$$\boldsymbol{T}_k = \boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1} - \boldsymbol{S}^{-1}\boldsymbol{A}_{(k)}(\boldsymbol{A}'_{(k)}\boldsymbol{S}^{-1}\boldsymbol{A}_{(k)})^{-1}\boldsymbol{A}'_{(k)}\boldsymbol{S}^{-1}.$$

Later it will be used that $\boldsymbol{A}'_{(k)}\boldsymbol{T}_k = \boldsymbol{0}$. The expectation $E[\boldsymbol{T}_k]$ follows from Lemma 2 (ii).

Suppose the size of $\boldsymbol{\beta}_i$, $i \in \{1, 2\}$, equals q, i.e., there are q different parameters which should be considered. The next proposed approach is not based on a factorization of the likelihood. Instead, if Proposition 1 suggests classifying into, say, Π_1 a strategy will be proposed which can support the outcome of the proposition.

In (12) the matrix $\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}$ appears and according to Lemma 3 it can be split into two terms of which one equals ($k \in \{1, \ldots, q\}$)

$$\boldsymbol{T}_k = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_k^o(\boldsymbol{H}_k^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_k^o)^{-1}\boldsymbol{H}_k^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1},$$

where $\boldsymbol{H}_k = \boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}_{(k)}$ and $\boldsymbol{A}_{(k)}$ is the matrix $\boldsymbol{A}$ without the kth element, as in Lemma 3. This can be done in q different ways, i.e., one split for each parameter. Note that $\boldsymbol{T}_k$ works like a filter and selects $\boldsymbol{a}_k$ when premultiplying by $\boldsymbol{A}' = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q)'$. From (12) it follows that $\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_i - \boldsymbol{x})$ is of interest, i.e., it is natural to calculate

$$\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}) - \frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_2 - \boldsymbol{x}).$$

Therefore, the expectation of $\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_i - \boldsymbol{x})$, given $\boldsymbol{S}$, is calculated. Given $\boldsymbol{S}$ the matrix $\boldsymbol{T}_k$ is constant and using Lemma 2(i), as well as independence between $\bar{\boldsymbol{x}}_1 - \boldsymbol{x}$ and $\boldsymbol{S}$,

$$\begin{aligned} E\left[\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_i - \boldsymbol{x})|\boldsymbol{S}\right] &= \operatorname{tr}\{\boldsymbol{T}_k\boldsymbol{\Sigma}\} \\ &\quad + \frac{n_i}{n_i+1}(\boldsymbol{A}\boldsymbol{\beta}_i - E[\boldsymbol{x}])'\boldsymbol{T}_k(\boldsymbol{A}\boldsymbol{\beta}_i - E[\boldsymbol{x}]). \end{aligned} \tag{29}$$

Depending on which population $\bar{\boldsymbol{x}}_i$ and $\boldsymbol{x}$ belong to the expression in (29) is of the following form:

– if $\bar{\boldsymbol{x}}_i$ and $\boldsymbol{x}$ belong to the same population

$$E\left[\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_i - \boldsymbol{x})|\boldsymbol{S}\right] = \operatorname{tr}\{\boldsymbol{T}_k\boldsymbol{\Sigma}\}, \quad i \in \{1, 2\};$$

– if $\bar{\boldsymbol{x}}_i \in \Pi_1$ and $\boldsymbol{x} \in \Pi_2$

$$\begin{aligned} &E\left[\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})|\boldsymbol{S}\right] \\ &\quad = \operatorname{tr}\{\boldsymbol{T}_k\boldsymbol{\Sigma}\} + \frac{n_1}{n_1+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'\boldsymbol{A}'\boldsymbol{T}_k\boldsymbol{A}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2); \end{aligned}$$

– if $\bar{\boldsymbol{x}}_i \in \Pi_2$ and $\boldsymbol{x} \in \Pi_1$

$$E\left[\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_2-\boldsymbol{x})|\boldsymbol{S}\right]$$
$$=\mathrm{tr}\{\boldsymbol{T}_k\boldsymbol{\Sigma}\}+\frac{n_2}{n_2+1}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'\boldsymbol{A}'\boldsymbol{T}_k\boldsymbol{A}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2).$$

Based on Lemmas 2 and 3, it is possible to find the expectation of $\boldsymbol{T}_k$ in the above formulas but we gain nothing by doing so. Instead it is more fruitful to see that

$$(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'\boldsymbol{A}'\boldsymbol{T}_k\boldsymbol{A}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)=(\beta_{2k}-\beta_{1k})^2\boldsymbol{a}_k'\boldsymbol{T}_k\boldsymbol{a}_k,$$

where β_{1k} and β_{2k} are the kth element in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively.

Based on the above moment calculations, it makes sense to set up, additionally to the conditions in Propositions 1, q new conditions to evaluate how an observation $\boldsymbol{x}$ should be classified.

Proposition 2 *Let the samples and the observation which is to be classified be as in Proposition 1.*

(i) Suppose that the conditions in Proposition 1(i) and (ii) are satisfied. If for $k\in\{1,\ldots,q\}$

$$\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_1-\boldsymbol{x})<\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_2-\boldsymbol{x}),$$

then $\boldsymbol{x}$ is classified as belonging to Π_1.

(ii) Suppose that the conditions in Proposition 1(i) and (ii) are not satisfied. If for $k\in\{1,\ldots,q\}$

$$\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_1-\boldsymbol{x})>\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{T}_k(\bar{\boldsymbol{x}}_2-\boldsymbol{x}),$$

then $\boldsymbol{x}$ is classified as belonging to Π_2.

(iii) If one of the above conditions in (i) or (ii) is not satisfied then $\boldsymbol{x}$ is classified as belonging to some unknown population.

Proposition 2 is based on $\boldsymbol{T}_k$ and from Lemma 3 it follows that also

$$\boldsymbol{S}^{-1}\boldsymbol{A}_{(k)}(\boldsymbol{A}_{(k)}'\boldsymbol{S}^{-1}\boldsymbol{A}_{(k)})^{-1}\boldsymbol{A}_{(k)}'\boldsymbol{S}^{-1} \tag{30}$$

can be of interest. This matrix can be split into a sum of rank one matrices which will lead to the possibility to investigate a number of new criteria.

In this article, however, only the case $q=2$ is considered. If $q=2$ and $k=1$ the expression in (30) equals

$$\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}. \tag{31}$$

The expectation of the quadratic form, given $\boldsymbol{S}$, equals

$$\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_i-\boldsymbol{x}), \tag{32}$$

which follows from (29) by replacing $\boldsymbol{T}_k$ with the expression in (31). Thus, if $\boldsymbol{\beta}_i=(\beta_{1i},\beta_{2i})'$, $i\in\{1,2\}$, using the independence between $(\bar{\boldsymbol{x}}_i-\boldsymbol{x})$ and $\boldsymbol{S}$,

$$\begin{aligned}E\left[\frac{n_i}{n_i+1}(\bar{\boldsymbol{x}}_i-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_i-\boldsymbol{x})|\boldsymbol{S}\right]\\ =\operatorname{tr}\{\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\}\\ +\frac{n_i}{n_i+1}(\boldsymbol{A}\boldsymbol{\beta}_i-E[\boldsymbol{x}])'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}(\boldsymbol{A}\boldsymbol{\beta}_i-E[\boldsymbol{x}]).\end{aligned} \tag{33}$$

Then, if $\boldsymbol{x}\in\Pi_j$, for $j\in\{1,2\}$, the last term in (33) equals

$$\begin{aligned}&\frac{n_i}{n_i+1}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_j)'\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{A}(\boldsymbol{\beta}_i-\boldsymbol{\beta}_j)\\ &\quad=\left(\beta_{1i}-\beta_{1j}+(\beta_{2i}-\beta_{2j})\frac{\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2}{\boldsymbol{a}_1'\boldsymbol{S}^{-1}\boldsymbol{a}_2}\right)^2\frac{(\boldsymbol{a}_1'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^2}{\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2}.\end{aligned}$$

An unconditional expectation could have been obtained by applying Lemma 2 (ii) but this is not essential for the presentation.

As above, we use criteria C_1 and C_2 given in (25) and (26), respectively, to classify the new observation $\boldsymbol{x}$. Moreover, if the criteria indicate that $\boldsymbol{x}$ belongs to either Π_1 or Π_2 it is suggested to continue with evaluating a number of other criteria, denoted C_3 to C_6. From Proposition 2(i)

$$C_3=\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{T}_1(\bar{\boldsymbol{x}}_2-\boldsymbol{x})-\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{T}_1(\bar{\boldsymbol{x}}_1-\boldsymbol{x}), \tag{34}$$

where $\boldsymbol{T}_1=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_1^o(\boldsymbol{H}_1^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_1^o)^{-1}\boldsymbol{H}_1^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}$, and $\boldsymbol{H}_1=\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{a}_2$. If $C_3>0$, $\boldsymbol{x}$ is classified to come from Π_1, otherwise Π_2. Moreover, also from Proposition 2(i),

$$C_5=\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{T}_2(\bar{\boldsymbol{x}}_2-\boldsymbol{x})-\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{T}_2(\bar{\boldsymbol{x}}_1-\boldsymbol{x}), \tag{35}$$

where $\boldsymbol{T}_2=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_2^o(\boldsymbol{H}_2^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{H}_2^o)^{-1}\boldsymbol{H}_2^{o'}\boldsymbol{A}'\boldsymbol{S}^{-1}$, and $\boldsymbol{H}_2=\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{a}_1$. If $C_5>0$, $\boldsymbol{x}$ is classified as coming from Π_1, otherwise Π_2. Finally there are the criteria which are "complementary" to C_3 and C_5, see (32):

$$\begin{aligned}C_4=&\frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2-\boldsymbol{x})\\ &-\frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_2(\boldsymbol{a}_2'\boldsymbol{S}^{-1}\boldsymbol{a}_2)^{-1}\boldsymbol{a}_2'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1-\boldsymbol{x}),\end{aligned} \tag{36}$$

$$C_6 = \frac{n_2}{n_2+1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_1(\boldsymbol{a}_1'\boldsymbol{S}^{-1}\boldsymbol{a}_1)^{-1}\boldsymbol{a}_1'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_2 - \boldsymbol{x}) - \frac{n_1}{n_1+1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x})'\boldsymbol{S}^{-1}\boldsymbol{a}_1(\boldsymbol{a}_1'\boldsymbol{S}^{-1}\boldsymbol{a}_1)^{-1}\boldsymbol{a}_1'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{x}). \tag{37}$$

If $C_4 > 0$ then $\boldsymbol{x}$ is classified as belonging to Π_1 and the same rule applies to C_6. Thus, a proposed decision rule equals: if C_1, C_2 and C_3 to C_6 have the same sign, i.e., if all criteria are positive, classify $\boldsymbol{x}$ into Π_1, if all are negative classify $\boldsymbol{x}$ into Π_2, and otherwise classify $\boldsymbol{x}$ into an unknown population.

Example 3 Consider again the two populations and the six scenarios in Example 1. According to Table 4, it follows that according to the classification philosophy in this article, i.e., applying all six criteria $(C_1, ..., C_6)$, only $\boldsymbol{x}$ in Scenarios 1 and 5, will be classified into one of the two given populations, i.e., in Scenario 1 $\boldsymbol{x}$ will be assigned to Π_2 and in Scenario 5, to Π_1. Moreover, it is only in Scenario 3 where C_1 and C_2 do not agree about the classification. In Scenarios 2, 4 and 6, it is C_3 to C_6 which hint that $\boldsymbol{x}$ should be assigned to an unknown population.

Example 4 With the same simulations as in Example 2 we can study the classifications using all six criteria $(C_1, ..., C_6)$. Hence, criteria $C_3, ..., C_6$ will further help us to classify to an unknown population. Table 5 is the same as Table 2, but we have added the last column with the classification proportions using all six criteria $(C_1, ..., C_6)$. When using $(C_1, ..., C_6)$ one can note that a much smaller proportion of simulations are misclassified to Population 2. The proportion of simulations classified to an unknown population are larger than only using C_1 and C_2.

Table 6 is the same as Table 3, but we have again added the last column with the classification rule using all six criteria $(C_1, ..., C_6)$. In all three cases the use of all six criteria are very promising, since the proportion of classifying the simulations as belonging to an unknown population has increased significantly. For Case 2, it is as high as 99.1.

Table 4 The classification functions C_1 to C_6, respectively, defined in (25)–(37), are evaluated for each of the six classification scenarios presented in (19)–(24)

Scenario	C_1	C_2	C_3	C_4	C_5	C_6	Classify to
Scenario 1, Fig. 2a	−7.78	−8.55	−7.13	−0.66	−6.08	−1.71	Π_2
Scenario 2, Fig. 2b	−5.29	−3.05	−3.57	−1.74	−6.08	0.78	Unknown
Scenario 3, Fig. 2c	0.05	−0.001	3.20	−3.16	−4.74	4.78	Unknown
Scenario 4, Fig. 2d	−10.27	−6.25	−10.69	0.42	−6.08	−4.19	Unknown
Scenario 5, Fig. 2e	8.01	0.98	7.48	0.53	6.02	1.98	Π_1
Scenario 6, Fig. 2f	9.56	0.91	7.20	2.36	9.81	−0.25	Unknown

A negative value means that the criteria supports a classification in Population 2 (Π_2)

Table 5 The performance of the classification functions D, D_{MK}, D_{ML} and $(C_1, ..., C_6)$, defined in (15), (17), (18) and (25)–(37), is presented for the data in Example 2

Population	D (%)	D_{MK} (%)	D_{ML} (%)	C_1/C_2 (%)	$(C_1, ..., C_6)$ (%)
1	74.8	76.0	71.6	72.5	27.2
2	25.2	24.0	28.4	21.5	5.5
Unknown	–	–	–	6.0	67.3

The proportions given in the table, are classification proportions for new observations from Population 1, i.e., if it is classified into Population 1, Population 2 or to an unknown population

5 Comparisons of classification approaches using a data set

In this section we will consider a real data set comprising repeated measurements and compare the performance of the five different classification procedures which were discussed in the previous section. In the supplementary material, one can find studies of five more real data sets, with similar results and conclusions.

We will do it by a leave-one-out cross-validation procedure, i.e., removing one vector of repeated measurements at time, which then should be classified. The data set consists of data from two populations with n_1 and n_2 independent observations. Hence, we will classify $n = n_1 + n_2$ observations and find the proportions of correctly, misclassified and unknown classified measurements, respectively. A quadratic growth will be assumed.

The data set consists of five repeated measurements, at start (baseline) and after 6, 12, 20 and 24 months, of the serum cholesterol (Srivastava 2002, pp. 379–381). The patients constitute of two groups, the first group ($n_1 = 31$) received an active drug and

Table 6 The performance of the classification functions D, D_{MK}, D_{ML}, and $(C_1, ..., C_6)$, defined in (15), (17), (18) and (25)–(37), is presented for the data in Example 2

Population	D (%)	D_{MK} (%)	D_{ML} (%)	C_1/C_2 (%)	$(C_1, ..., C_6)$ (%)
Case 1					
1	51.4	51.1	50.8	45.4	17.9
2	48.6	48.9	49.2	42.8	13.7
Unknown	–	–	–	11.8	68.4
Case 2					
1	41.2	41.5	37.7	35.2	0.9
2	58.8	58.5	62.3	52.4	0.0
Unknown	–	–	–	12.4	99.1
Case 3					
1	93.5	94.3	70.1	91.6	36.7
2	6.5	5.7	29.9	3.9	0.7
Unknown	–	–	–	4.5	62.6

The proportions given in the table, are classification proportions for new observations from the three cases seen in Fig. 3, i.e., if it is classified into Population 1, Population 2 or to an unknown population

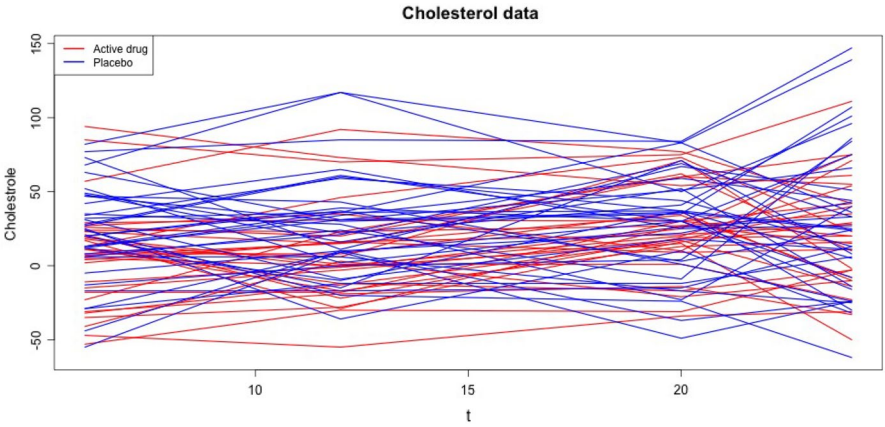

Fig. 4 Plot of cholesterol data. Four repeated measurements at 6, 12, 20 and 24 months for the two groups, an active drug intake ($n_1 = 31$) and placebo ($n_2 = 36$) are presented

Table 7 The performance of the classification functions D, D_{MK}, D_{ML}, $(C_1, ..., C_6)$, respectively, defined in (15), (17), (18), and (25)–(37), i.e., the proportions of the patients based on the cholesterol measurements that are correctly, misclassified or classified to come from an unknown population, respectively

Population	D (%)	D_{MK} (%)	D_{ML} (%)	C_1/C_2 (%)	$(C_1, ..., C_6)$ (%)
Correctly	64	63	60	52	19
Misclassified	36	37	40	30	7
Unknown	–	–	–	18	73

the second group ($n_2 = 36$) received placebo. To include the baseline effect, a new data set was created by individually subtracting baseline values from the original observations. This data set is exposed in Fig. 4, and in Table 7 we can find the proportions of correctly, misclassified and unknown unclassified patients, respectively.

As one can see in Table 7, the performance of the classification functions D, D_{MK} and D_{ML} are fairly similar and the number of correctly classified observations are as low as 60%. The data set is not designed for classification and the profiles for each group are fairly similar, which can be depicted in Fig. 4, hence the poor performance. Furthermore, when considering the two classifiers proposed in this paper, respectively, based on C_1/C_2 and $(C_1, ..., C_6)$, several of the measurements are classified as unknown, which seems to be reasonable due to similar growths.

6 Concluding remarks

When classifying growth curves in multivariate linear models following a normal distribution, it is of interest and possible to include different aspects in the analysis. Hence, the mean structure for the specified populations can be on different levels, but they can

also follow different profiles, e.g., slopes. In this paper, we have developed a likelihood-based classification procedure for linear growth curves, i.e., a new observation of p repeated measurements is assigned to one of two specified populations. A key feature of the proposed classification procedure is that it can suggest that the new observation does not belong to any of the two populations, i.e., it should be classified to an unknown population. This can be the case, for example, when the new observation has the same level as one of the population, but different profiles or vice versa.

Given a new classification procedure, it is of interest to find out how well it will classify a new observation. However, in general, it can be difficult to find the exact expression of the misclassification probabilities. Even in the simplest case for Fisher's linear classification function with estimated parameters, the misclassification probabilities can only be found approximately. In this paper we have not discussed the misclassification probabilities for the proposed likelihood-based procedure, but the ideas can serve for future research on this problem.

We conclude that the proposed classifiers based on C_1/C_2 and $(C_1, ..., C_6)$, have significant smaller number of misclassified observations compared to the other methods. Although, they deliver a smaller number of correctly classified objects. Relatively to the other classification procedures the approach of this paper is more "robust".

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-024-00900-1.

Acknowledgements The authors would like to thank the anonymous reviewers for valuable and helpful suggestions which have significantly improved the paper.

References

Albert, J. M., Kshirsagar, A. M. (1993). The reduced-rank growth curve model for discriminant analysis of longitudinal data. *Australian Journal of Statistics, 35*, 345–357.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis. Wiley series in probability and statistics* (3rd ed.). Hoboken: Wiley.

Barnard, M. M. (1935). The secular variations of skull characters in four series of Egyptian skulls. *Annals of Eugenics, 6*, 352–371.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Mathematical Proceedings of the Cambridge Philosophical Society, 34*, 33–40.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society, 35*, 180–185.

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Annals of Eugenics, 16*, 199–214.

Burnaby, T. P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics, 22*, 96–110.

Dasgupta, S. (1993). The evolution of the D^2 statistic of Mahalanobis. *Sankhyā, 55*, 442–459.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*, 179–188.

Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, *8*, 376–386. (Revised 1950, in Contributions to Mathematical Statistics, Wiley, New York).

Hodges, J. L. (1950). Discriminatory analysis. I. Survey of discriminatory analysis (Report No. 1, Project No. 21-49-004). USAF School of Aviation Medicine, October 1950. (NTIS No. PB 102-690).

Hotelling, H. (1931). The generalization of Student's ratio. *The Annals of Mathematical Statistics, 2*, 360–378.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika, 28*, 321–377.

Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research, 45*, 543–598.
Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics, 18*, 75–86.
Kollo, T., and von Rosen, D. (2005). *Advanced multivariate statistics with matrices*. Dordrecht: Springer.
Kshirsagar, A. M. (1962). A note on the direction and collinearity factors in canonical analysis. *Biometrika, 9*, 255–259.
Kshirsagar, A. M., Davis, T. M. (1983a). Direction and collinearity factors of Wilks' Λ associated with the Growth Curve model. *Australian Journal of Statistics, 25*, 467–481.
Kshirsagar, A. M., Davis, T. M. (1983b). Discriminant functions from the dummy variable space for growth curve models. *Communications in Statistics-Theory and Methods, 12*, 427–440.
Lee, J. C. (1982). Classification of Growth Curves. In *Classification, pattern recognition and reduction of dimensionality, Handbook of statist.* (Vol. 2, pp. 121–137). Amsterdam: North-Holland.
Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India, 2*, 49–55.
McLachlan, G. J. (2004). *Discriminant analysis and statistical pattern recognition*. Hoboken: Wiley.
Mentz, G. B., Kshirsagar, A. M. (2004). Classification using growth curves. *Communications in Statistics-Theory and Methods, 33*, 2487–2502.
Muirhead, R. J. (1982). *Aspects of multivariate statistical theory. Wiley series in probability and mathematical statistics*. New York: Wiley.
Ngailo, E. K., von Rosen, D., Singull, M. (2021). Asymptotic approximation of misclassification probabilities in linear discriminant analysis with repeated measurements. *Acta et Commentationes Universitatis Tartuensis de Mathematica, 25*, 67–85.
Pearson, K. (1915). On the problem of sexing osteometric material. *Biometrika, 10*, 479–487.
Pearson, K. (1926). On the coefficient of racial likeness. *Biometrika, 18*, 105–117.
Potthoff, R. F., Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*, 313–326.
Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B, 10*, 159–203.
Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14*, 1–17.
Rao, C. R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika, 46*, 49–58.
Rao, C. R. (1966). Discriminant function between composite hypotheses and related problems. *Biometrika, 53*, 339–345.
Rao, C. R. (1973). *Linear statistical inference and its applications. Wiley series in probability and mathematical statistics* (2nd ed.). New York: Wiley.
Seal, B. (1911). Meaning of race, tribe, nation. In G. Spiller (Ed.) *Inter-racial problems* (pp. 1–12). London: Universal Races Congress. https://archive.org/details/papersoninterrac00univiala/page/n51/mode/2up?q=Seal.
Srivastava, M. S. (2002). *Method of multivariate statistics*. New York: Wiley-Interscience.
Srivastava, M. S., Khatri, C. G. (1979). *An introduction to multivariate statistics*. New York-Oxford: North-Holland.
von Rosen, D. (2018). *Bilinear regression analysis: An introduction. Lecture notes in statistics* (Vol. 220). New York: Springer.
Wald, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. *The Annals of Mathematical Statistics, 15*, 145–162.