

## An Edgeworth-type expansion for the distribution of a likelihood-based discriminant function

Emelyne Umunoza Gasana, Dietrich von Rosen & Martin Singull

**To cite this article:** Emelyne Umunoza Gasana, Dietrich von Rosen & Martin Singull (2023) An Edgeworth-type expansion for the distribution of a likelihood-based discriminant function, *Journal of Statistical Computation and Simulation*, 93:17, 3185-3202, DOI: 10.1080/00949655.2023.2219358

**To link to this article:** <https://doi.org/10.1080/00949655.2023.2219358>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 08 Jun 2023.



Submit your article to this journal [↗](#)



Article views: 276



View related articles [↗](#)



View Crossmark data [↗](#)

# An Edgeworth-type expansion for the distribution of a likelihood-based discriminant function

Emelyne Umunoza Gasana <sup>a,b</sup>, Dietrich von Rosen <sup>c</sup> and Martin Singull <sup>a</sup>

<sup>a</sup>Department of Mathematics, Linköping University, Linköping, Sweden; <sup>b</sup>Department of Mathematics, University of Rwanda, Butare, Rwanda; <sup>c</sup>Department of Energy and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden

## ABSTRACT

The exact distribution of a classification function is often complicated to allow for easy numerical calculations of misclassification errors. The use of expansions is one way of dealing with this difficulty. In this paper, approximate probabilities of misclassification of the maximum likelihood-based discriminant function are established via an Edgeworth-type expansion based on the standard normal distribution for discriminating between two multivariate normal populations.

## ARTICLE HISTORY

Received 8 November 2022  
Accepted 8 May 2023

## KEYWORDS

Classification rule;  
discriminant analysis;  
Edgeworth-type expansion;  
misclassification errors

## 1. Introduction


The concept of classification has interested several researchers already long time ago, such as [1–5]. It consists of developing a classification rule by examining the characteristics of individuals that best distinguish between the populations. There exists several approaches of deriving such a rule. For example a plug-in approach [6], a likelihood approach [7,8], as well as others.

Let  $\mathbf{y}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ ,  $i \in \{1, \dots, n_1\}$ , be a sample from population  $\pi_1$  collected in  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_{n_1})$  and  $\mathbf{z}_j \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ ,  $j \in \{1, \dots, n_2\}$ , be a sample from population  $\pi_2$  collected in  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_{n_2})$ . Let  $\bar{\mathbf{y}} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{y}_i$ , and  $\bar{\mathbf{z}} = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{z}_j$ . Assume that an observation  $\mathbf{x}$  is to be classified. Anderson's [6] classification rule that he labelled as the  $W$ -rule is based on the statistic

$$W = \mathbf{x}'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \bar{\mathbf{z}}) - \frac{1}{2}(\bar{\mathbf{y}} + \bar{\mathbf{z}})'\mathbf{S}^{-1}(\bar{\mathbf{y}} - \bar{\mathbf{z}}), \quad (1)$$

which is a linear function in  $\mathbf{x}$  with the rule that the new observation  $\mathbf{x}$  is classified to  $\pi_1$  if  $W \geq 0$  and to  $\pi_2$  if  $W < 0$ . The maximum likelihood-based classification rule; the  $Z$ -rule proposed by Kudo [7,9], as an alternative to the  $W$ -rule

$$Z = \frac{1}{2} \left[ \frac{n_2}{n_2 + 1} (\mathbf{x} - \bar{\mathbf{z}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{z}}) - \frac{n_1}{n_1 + 1} (\mathbf{x} - \bar{\mathbf{y}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{y}}) \right], \quad (2)$$

**CONTACT** Emelyne Umunoza Gasana  emelyne.umunoza.gasana@liu.se

which is a quadratic function in  $\mathbf{x}$ . The rule is to classify  $\mathbf{x}$  as coming from  $\pi_1$  if  $Z$  is positive and  $\pi_2$  if  $Z$  is negative. Alternatively, Gasana et al. [10] proposed two likelihood-based discriminant functions. If  $\Sigma$  is known

$$\tilde{D} = \frac{1}{2} \frac{n_2}{n_2 + 1} (\bar{\mathbf{z}} - \mathbf{x})^\top \Sigma^{-1} (\bar{\mathbf{z}} - \mathbf{x}) - \frac{1}{2} \frac{n_1}{n_1 + 1} (\bar{\mathbf{y}} - \mathbf{x})^\top \Sigma^{-1} (\bar{\mathbf{y}} - \mathbf{x}). \tag{3}$$

If  $\tilde{D} \geq 0$ ,  $\mathbf{x}$  is classified into  $\pi_1$  and if  $\tilde{D} < 0$  then  $\mathbf{x}$  is classified into  $\pi_2$ . If  $\Sigma$  is unknown

$$\hat{D} = \frac{n_2 m}{n_2 + 1} \frac{1}{2} (\bar{\mathbf{z}} - \mathbf{x})^\top \mathbf{S}_{yz}^{-1} (\bar{\mathbf{z}} - \mathbf{x}) - \frac{n_1 m}{n_1 + 1} \frac{1}{2} (\bar{\mathbf{y}} - \mathbf{x})^\top \mathbf{S}_{yz}^{-1} (\bar{\mathbf{y}} - \mathbf{x}), \tag{4}$$

where  $m = n_1 + n_2 - p - 3 > 0$ . The observation  $\mathbf{x}$  is classified into  $\pi_1$  if  $\hat{D} \geq 0$  and if  $\hat{D} < 0$  then  $\mathbf{x}$  is classified into  $\pi_2$ .

Whenever classifying an object, there is a chance that it might be wrongly classified. Therefore, in statistics, we are interested in the probability of such an error. Fujikoshi et al. [11] defines the probability of misclassification as a measure of the goodness of a classification rule. Generally, misclassification errors involve unknown parameters, thus they need to be estimated. Anderson [12], Sitgreaves [13,14] and Wald [15] mentioned that the distribution of the classification function is too complicated to be used numerically. Several authors, such as Anderson [6,12] and Okamoto [16] have investigated the distribution of the linear discriminant function. One way of handling this problem is to approximate the discriminant distribution function using expansions. Okamoto [16] and Memon and Okamoto [17] investigated asymptotic expansions for the discriminant function distribution up to terms of order  $n^{-2}$ ,  $n = n_1 + n_2 - 2$ , [12,18,19] extended the expansions to terms of  $O(n^{-3})$  the asymptotic expansions of misclassification errors for the distribution of the discriminant functions when  $n_1$  and  $n_2$  grow to infinity and the ratio  $n_1/n_2$  approaches a positive limit.

At early nineteenth century there was a general idea of approximating a frequency function using a series containing an uncomplicated density function, for instance Laplace [20] used normal density, Hermite polynomials and their expectations. On the basis of Laplace’s results, several other researchers came up with ideas of density approximations in terms of other quantities such as moments [21] and cumulants [22–26]. For details, see [27–29]. Popular density approximations are the Gram-Charlier expansion introduced but Thiele, Edgeworth expansion, the saddle point approximation and the Cornish-Fisher expansion. The Gram-Charlier and Edgeworth expansions are comparably similar but differ on how terms are organized. In addition, both expansions are disadvantaged by the fact that the approximations may not be densities. Gupta and Panchapakesan [30], reviewed Edgeworth expansions in statistics. This concept attracted several applications in many areas of statistics such as Ranga Rao [31] who obtained Edgeworth expansions and Davis [32] who introduced a multivariate Edgeworth expansion. In addition, Fujikoshi et al. [11] investigated the Edgeworth expansion and its validity. Furthermore, Kollo and von Rosen [33] described Edgeworth-type expansions where a complicated density function is described through a simpler one. In this paper, we will use an Edgeworth-type expansion to approximate the distribution of the likelihood-based discriminant function through a standard normal distribution and derive approximate probabilities of misclassification.

## 2. Proof preparation

Several researchers such as Anderson [6,12] and Okamoto [16], to mention a few, have investigated the distribution of the linear discriminant function. The exact distributions of the discriminant functions  $\tilde{D}$  and  $\hat{D}$ , given in (3) and (4), respectively, are very difficult to use for numerical calculations, as mentioned earlier. Therefore, it is practical to provide reasonable approximations of the distribution of these discriminant functions to get around the problem. With the help of the expansions misclassification probabilities can be obtained. However, it is important to note that an approximate density based on an Edgeworth expansion does not have to be a density.

**Definition 2.1:** The  $j$ th-degree Hermite polynomial,  $H_j(x, m, \sigma^2)$ , for the mean  $m$  and the variance  $\sigma^2 > 0$  is given by

$$H_j(x, m, \sigma^2) = \frac{(-1)^j}{f_x(x)} \frac{d^j f_x(x)}{dx^j}, \quad j \in \mathbb{N}_0, \tag{5}$$

where  $f_x(x)$  is the density function of the normal distribution  $\mathcal{N}(m, \sigma^2)$ , i.e.

$$f_x(x) = (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}}. \tag{6}$$

**Theorem 2.2:** Consider the case of a standard normal distribution,  $\mathcal{N}(0, 1)$ . The  $j^{\text{th}}$  univariate Hermite polynomials  $H_j(x, 1)$ ,  $j \in \{0, 1, 2, 3\}$ , equal

$$H_0(x, 1) = 1; \quad H_1(x, 1) = x; \quad H_2(x, 1) = x^2 - 1; \quad H_3(x, 1) = x^3 - 3x. \tag{7}$$

The Edgeworth-type expansion of a density presented by Kollo and von Rosen [33] is shown in the following theorem.

**Theorem 2.3:** Let  $d$  be a random variable with finite first three cumulants, then its density  $f_d(x)$  can be approximated through the density  $f_N(x)$  of the standard normal distribution,  $\mathcal{N}(0, 1)$ , by the Edgeworth-type expansion

$$f_d(x) \approx f_N(x) \left\{ 1 + E[d]H_1(x, 1) + \frac{1}{2} (\text{Var}[d] - 1 + (E[d])^2) H_2(x, 1) + \frac{1}{6} (c_3[d] + 3(\text{Var}[d] - 1)E[d] + (E[d])^3) H_3(x, 1) \right\}, \tag{8}$$

where  $c_3[\cdot]$  is the third cumulant,  $H_i(x, 1)$ ,  $i \in \{1, 2, 3\}$ , are obtained from Theorem 2.2 and  $f_N(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$ .

**Theorem 2.4:** Suppose the third cumulant is omitted and let  $d$  be a random variable with finite first three cumulants, then its density  $f_d(x)$  can be approximated through the density  $f_N(x)$  of the standard normal distribution,  $\mathcal{N}(0, 1)$ , by the Edgeworth-type expansion

$$f_d(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \{a_0 - a_1x + a_2x^2 + a_3x^3\}, \tag{9}$$

where

$$a_0 = 3 - \text{Var}[d] - (E[d])^2, \quad a_1 = 3\text{Var}[d]E[d] - 5E[d] + (E[d])^3,$$

$$a_2 = \text{Var}[d] - 1 + (E[d])^2, \quad a_3 = (\text{Var}[d] - 1)E[d] + \frac{1}{3}(E[d])^3. \tag{10}$$

**Proof:** Inserting the Hermite polynomials  $H_i(x, 1)$ ,  $i \in \{1, 2, 3\}$ , and the density of  $\mathcal{N}(0, 1)$  by their values, (8) can be rewritten as

$$\begin{aligned} f_d(x) &\approx f_N(x) \left\{ 1 - \frac{1}{2}(\text{Var}[d] - 1 + (E[d])^2) \right. \\ &\quad + \left( E[d] - \frac{1}{2}(3(\text{Var}[d] - 1)E[d] + (E[d])^3) \right) x \\ &\quad \left. + \frac{1}{2}(\text{Var}[d] - 1 + (E[d])^2)x^2 + \frac{1}{6}(3(\text{Var}[d] - 1)E[d] + (E[d])^3)x^3 \right\} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \left\{ 3 - \text{Var}[d] - (E[d])^2 - (3\text{Var}[d]E[d] - 5E[d] + (E[d])^3)x \right. \\ &\quad \left. + (\text{Var}[d] - 1 + (E[d])^2)x^2 + \left( (\text{Var}[d] - 1)E[d] + \frac{1}{3}(E[d])^3 \right) x^3 \right\}. \tag{11} \end{aligned}$$

■

Therefore,  $f_d(x)$  is approximated with a polynomial function in  $x$  and the coefficients are in terms of the first and second cumulants of  $d$ . Gasana et al. [10] found expressions of the first two cumulants of the discriminant function. An Edgeworth-type expansion, based on Theorem 2.4, of the discriminant function can be expressed up to three terms when the term including  $c_3[\cdot]$  is excluded.

**Theorem 2.5:** Let  $\tilde{D}$  be defined in (3). Then if  $\mathbf{x} \in \pi_1$ , the first two cumulants of  $\tilde{D}$  are given by

$$\begin{cases} E[\tilde{D}] = \frac{1}{2} \frac{n_2}{n_2 + 1} \Delta^2, \\ \text{Var}[\tilde{D}] = p(1 - q^2) + \frac{n_2}{n_2 + 1} \Delta^2, \end{cases}$$

where

$$q = \sqrt{\frac{n_1}{n_1 + 1}} \sqrt{\frac{n_2}{n_2 + 1}}, \tag{12}$$

and

$$\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{13}$$

is the Mahalanobis squared distance; if  $\mathbf{x} \in \pi_2$ ,  $\begin{cases} E[\tilde{D}] = -\frac{1}{2} \frac{n_1}{n_1 + 1} \Delta^2, \\ \text{Var}[\tilde{D}] = p(1 - q^2) + \frac{n_1}{n_1 + 1} \Delta^2. \end{cases}$

**Theorem 2.6:** Consider the discriminant function  $\hat{D}$  in (4). If  $\mathbf{x} \in \pi_1$ , the expected value and variance of the discriminant function are given by

$$\begin{cases} E[\hat{D}] = \frac{1}{2} \frac{n_2}{n_2 + 1} \Delta^2, \\ \text{Var}[\hat{D}] = (1 - q^2)m^2 c_0 p + m^2 \frac{n_2}{n_2 + 1} c_0 \Delta^2 + \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{2(m - 2)} (\Delta^2)^2, \end{cases}$$

where  $m = n_1 + n_2 - p - 3 > 2$ ,  $c_0 = \frac{m+p}{m(m-2)(m+1)}$  and  $\Delta^2$  is the Mahalanobis squared distance given in (13); and if  $\mathbf{x} \in \pi_2$

$$\begin{cases} E[\widehat{D}] = \frac{1}{2} \frac{n_2}{n_2 + 1} \Delta^2, \\ \text{Var}[\widehat{D}] = (1 - q^2)m^2 c_0 p + m^2 \frac{n_2}{n_2 + 1} c_0 \Delta^2 + \left(\frac{n_2}{n_2 + 1}\right)^2 \frac{1}{2(m - 2)} (\Delta^2)^2. \end{cases}$$

Proofs of Theorems 2.5 and 2.6 are given by Gasana et al. [10].

Since the discriminant function is one-dimensional, the Edgeworth-type expansions of the densities of  $\widehat{D}$  and  $\widetilde{D}$  through the density of  $\mathcal{N}(0, 1)$  in the next sections will take the form of (9), where  $a_i, i \in \{0, 1, 2, 3\}$ , will be in terms of the expectation and variance of the corresponding discriminant function, as in (10).

**Theorem 2.7:** Let  $\widetilde{D}$  be defined in (3), and  $\Delta^2$  in (13). For large  $n_1$  and  $n_2$ ,  $\widetilde{D}$  is normally distributed:

$$\begin{cases} \widetilde{D} \sim N\left(\frac{1}{2}\Delta^2, \Delta^2\right), & \text{if } \mathbf{x} \in \pi_1, \\ \widetilde{D} \sim N\left(-\frac{1}{2}\Delta^2, \Delta^2\right), & \text{if } \mathbf{x} \in \pi_2, \end{cases} \tag{14}$$

where  $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ .

**Proof:** As  $n_i \rightarrow \infty, i \in \{1, 2\}$ , we have

$$\frac{n_1}{n_1 + 1} \rightarrow 1, \quad \frac{n_2}{n_2 + 1} \rightarrow 1, \quad \bar{\mathbf{y}} \xrightarrow{p} \boldsymbol{\mu}_1, \quad \bar{\mathbf{z}} \xrightarrow{p} \boldsymbol{\mu}_2, \quad p(1 - q^2) \rightarrow 0. \tag{15}$$

Therefore, as  $n_i \rightarrow \infty, i \in \{1, 2\}$ ,

$$\begin{aligned} \widetilde{D} &\longrightarrow \frac{1}{2}(\boldsymbol{\mu}_2 - \mathbf{x})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \mathbf{x}) - \frac{1}{2}(\boldsymbol{\mu}_1 - \mathbf{x})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \mathbf{x}) \\ &= \mathbf{x}^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \end{aligned} \tag{16}$$

Then, for  $\mathbf{x} \in \pi_1$ ,

$$\begin{aligned} E[\widetilde{D}] &= E\left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right)\right] \\ &= \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \frac{1}{2}\Delta^2, \\ \text{Var}[\widetilde{D}] &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Delta^2. \end{aligned}$$

Similarly, for  $\mathbf{x} \in \pi_2$ ,

$$E[\widetilde{D}] = -\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

with the same variance as when  $\mathbf{x} \in \pi_1$ . ■

**Theorem 2.8:** Let  $\widehat{D}$  be given by (4) and  $\widetilde{D}$  by (3), with  $\bar{y}$  the mean of a sample of size  $n_1$  from  $\mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ ,  $\bar{z}$  the mean of a sample of size  $n_2$  from  $\mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ . The limiting distribution of both  $\widehat{D}$  and  $\widetilde{D}$  as  $n_1 \rightarrow \infty$  and  $n_2 \rightarrow \infty$  is  $\mathcal{N}(\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is distributed as  $\mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$  and  $\mathcal{N}(-\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is distributed as  $\mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ .

**Proof:** The limiting distribution of  $\widetilde{D}$  is already proven in Theorem 2.7. Moreover, as  $n_i \rightarrow \infty, i \in \{1, 2\}, m\mathbf{S}^{-1} \rightarrow \boldsymbol{\Sigma}^{-1}$ . Thus the proof for  $\widehat{D}$  follows from the proof of Theorem 2.7. ■

Hence it follows that the limiting distribution of both classification functions  $\widetilde{D}$  and  $\widehat{D}$  follow the distribution of a linear discriminant function. This result will be used for the standardization of the two classification rules in the next section.

### 3. Edgeworth-type expansion of the discriminant function

#### 3.1. Edgeworth-type expansion of $\widetilde{D}$

The basic properties of the discriminant function  $\widetilde{D}$ , defined in (3), such as its expectation and variance are given in Theorem 2.5. In the theorem below, an Edgeworth-type expansion of  $\widetilde{D}$  is presented in terms of its mean and variance via the standard normal distribution density.

**Theorem 3.1:** Let  $\widetilde{D}$  be the discriminant function defined in (3) and  $\Delta^2$  be as in (13). Let  $q$  be given by (12). Then the Edgeworth-type expansion,  $f_{\widetilde{D}}(x)$ , if  $\mathbf{x}$  is classified into  $\pi_1$ , equals

$$f_{\widetilde{D}}(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \{b_0 + b_1x + b_2x^2 + a_3x^3\}, \tag{17}$$

where,

$$\begin{aligned} b_0 &= 3 - p(1 - q^2) - \frac{n_2}{n_2 + 1} \Delta^2 - \frac{1}{4} \left(\frac{n_2}{n_2 + 1}\right)^2 (\Delta^2)^2, \\ b_1 &= \frac{1}{2} \left( (3p(1 - q^2) - 5) \frac{n_2}{n_2 + 1} \Delta^2 + 3 \left(\frac{n_2}{n_2 + 1}\right)^2 (\Delta^2)^2 + \frac{1}{4} \left(\frac{n_2}{n_2 + 1}\right)^3 (\Delta^2)^3 \right), \\ b_2 &= -1 + p(1 - q^2) + \frac{n_2}{n_2 + 1} \Delta^2 + \frac{1}{4} \left(\frac{n_2}{n_2 + 1}\right)^2 (\Delta^2)^2, \\ b_3 &= \frac{1}{2} \left( (p(1 - q^2) - 1) \frac{n_2}{n_2 + 1} \Delta^2 + \left(\frac{n_2}{n_2 + 1}\right)^2 (\Delta^2)^2 + \frac{1}{12} \left(\frac{n_2}{n_2 + 1}\right)^3 (\Delta^2)^3 \right). \end{aligned} \tag{18}$$

If  $\mathbf{x} \in \pi_2, f_{\widetilde{D}}(x)$  is the same as when  $\mathbf{x} \in \pi_1$ , except that  $n_1$  and  $n_2$  are interchanged and we have opposite signs on coefficients of  $x$  and  $x^3$ .

**Proof:** Consider the Edgeworth-type expansion of the form (9). For  $\widetilde{D}$  in (3) with the first moments defined from Theorem 2.5, suppose  $a_i, i \in \{0, 1, 2, 3\}$ , are given by equations (10).

Then, if  $\mathbf{x} \in \pi_1$  then  $b_i, i \in \{0, 1, 2, 3\}$ , are obtained by replacing the moments of  $\tilde{D}$  into  $a_i, i \in \{0, 1, 2, 3\}$ , which concludes the proof. ■

According to the results from Theorem 2.7, the background is that the limiting distribution of  $\tilde{D}$  is  $\mathcal{N}(\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is from  $\pi_1$  and is  $\mathcal{N}(-\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is from  $\pi_2$ . In the following theorem we evaluate the Edgeworth-type expansion of the standardized discriminant function with respect to its limiting distribution. This means that we are going to consider either  $y = \frac{\tilde{D}-\frac{1}{2}\Delta^2}{\Delta}$  or  $y = \frac{\tilde{D}+\frac{1}{2}\Delta^2}{\Delta}$ .

**Corollary 3.2:** *Let  $\tilde{D}$  be the discriminant function defined in (3) and  $\Delta^2$  is as in (13). Let  $y = \frac{\tilde{D}-\frac{1}{2}\Delta^2}{\Delta}$  and  $q$  be given by (12), then the Edgeworth-type expansion,  $f_y(x)$ , if  $\mathbf{x}$  is classified into  $\pi_1$ , equals*

$$f_y(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \{u_0 + u_1x + u_2x^2 + u_3x^3\}, \tag{19}$$

where

$$\begin{aligned} u_0 &= \frac{1}{\Delta^2}p(1 - q^2) + 3 - \frac{n_2}{n_2 + 1} - \frac{1}{2}\Delta^2 \left( \left( \frac{1}{2} \frac{n_2}{n_2 + 1} \right)^2 - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right) \\ u_1 &= \frac{3}{2} \frac{1}{\Delta} p(1 - q^2) \left( \frac{n_2}{n_2 + 1} - 1 \right) + \Delta \left( \frac{3}{2} \left( \frac{n_2}{n_2 + 1} \right)^2 - 4 \frac{n_2}{n_2 + 1} + \frac{5}{2} \right) \\ &\quad + \frac{1}{8} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 + 3 \left( \frac{n_2}{n_2 + 1} \right)^2 + 3 \frac{n_2}{n_2 + 1} - 1 \right), \\ u_2 &= \frac{1}{\Delta^2}p(1 - q^2) + 1 - \frac{n_2}{n_2 + 1} - \frac{1}{2}\Delta^2 \left( \frac{1}{2} \left( \frac{n_2}{n_2 + 1} \right)^2 - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right), \\ u_3 &= \frac{1}{2} \frac{1}{\Delta} p(1 - q^2) \left( \frac{n_2}{n_2 + 1} - 1 \right) + \Delta \left( \frac{1}{2} \left( \frac{n_2}{n_2 + 1} \right)^3 - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right) \\ &\quad + \frac{1}{24} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 - 3 \left( \frac{n_2}{n_2 + 1} \right)^2 + 3 \frac{n_2}{n_2 + 1} - 1 \right). \end{aligned} \tag{20}$$

When  $\mathbf{x} \in \pi_2$ , we have the same values as when  $\mathbf{x} \in \pi_1$ , except that  $n_1$  and  $n_2$  are interchanged and we have opposite signs on coefficients of  $x$  and  $x^3$ .

**Proof:** Given the standardized  $y = \frac{\tilde{D}-\frac{1}{2}\Delta^2}{\Delta}$  then, if  $\mathbf{x} \in \pi_1$ , its expectation and variance are  $E(y) = \frac{1}{\Delta} E[\tilde{D}] - \frac{1}{2}\Delta = \frac{1}{2\Delta} (n_2 - n_2 + 1 - 1)$  and  $Var[y] = \frac{1}{\Delta^2} Var[\tilde{D}] = \frac{1}{\Delta^2} p(1 - q^2) + \frac{n_2}{n_2 + 1}$ , where  $E[\tilde{D}]$  and  $Var[\tilde{D}]$  are given by Theorem 2.5. Now consider the Edgeworth-type expansion of the form (9). Therefore, substituting the mean and variance of  $y$  into  $a_i, i \in \{0, 1, 2, 3\}$ , given by (10) yields  $u_i, i \in \{0, 1, 2, 3\}$ , which establishes the proof. ■



### 3.2. Edgeworth-type expansion of $\widehat{D}$

The classification function  $\widehat{D}$  given in (4) follows a complicated distribution since it is a difference between two non-independent non-central F-distributions. In this section, an Edgeworth-type expansion of the density of  $\widehat{D}$  is provided in terms of its expected value and variance via a standard normal distribution density. Let  $f_{\widehat{D}}(x)$  be such an expansion. Then since  $\widehat{D}$  is one-dimensional,  $f_{\widehat{D}}$  will be a polynomial function of the form (9). In the following theorem the expansion is presented.

**Theorem 3.3:** *Let  $\widehat{D}$  be the discriminant function defined in (4) and  $\Delta^2$  is as in (13). Then, for  $m = n_1 + n_2 - p - 3$ ,  $q$  in (12) and  $c_0 = \frac{m+p}{m(m-2)(m+1)}$ , the Edgeworth-type expansion,  $f_{\widehat{D}}(x)$  when the observation  $\mathbf{x}$  is classified in population  $\pi_1$  equals*

$$f_{\widehat{D}}(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \{v_0 + v_1x + v_2x^2 + v_3x^3\}, \tag{21}$$

where

$$\begin{aligned} v_0 &= 3 - (1 - q^2)m^2c_0p - m^2 \frac{n_2}{n_2 + 1} c_0\Delta^2 + \left(\frac{n_2}{n_2 + 1}\right)^2 \frac{m}{4(m - 2)} (\Delta^2)^2, \\ v_1 &= \frac{1}{2} \left\{ (3(1 - q^2)m^2c_0p - 5) \frac{n_2}{n_2 + 1} \Delta^2 + 3m^2 \left(\frac{n_2}{n_2 + 1}\right)^2 c_0(\Delta^2)^2 \right. \\ &\quad \left. + \left(\frac{n_2}{n_2 + 1}\right)^3 \frac{m + 4}{4(m - 2)} (\Delta^2)^3 \right\}, \\ v_2 &= (1 - q^2)m^2c_0p - 1 + m^2 \frac{n_2}{n_2 + 1} c_0\Delta^2 - \left(\frac{n_2}{n_2 + 1}\right)^2 \frac{m}{4(m - 2)} (\Delta^2)^2, \\ v_3 &= \frac{1}{2} \left[ ((1 - q^2)m^2c_0p - 1) \frac{n_2}{n_2 + 1} \Delta^2 + m^2 \left(\frac{n_2}{n_2 + 1}\right)^2 c_0(\Delta^2)^2 \right. \\ &\quad \left. + \left(\frac{n_2}{n_2 + 1}\right)^3 \frac{m + 4}{12(m - 2)} (\Delta^2)^3 \right]. \end{aligned} \tag{22}$$

Moreover, when  $\mathbf{x} \in \pi_2$ , the coefficients  $v_i$ ,  $i \in \{0, 1, 2, 3\}$  in  $f_{\widehat{D}}(x)$  are similar to the above except that we have opposite signs on  $v_1$  and  $v_3$  and that  $n_1$  and  $n_2$  are interchanged.

**Proof:** If  $\mathbf{x} \in \pi_1$ , for  $m = n_1 + n_2 - p - 3$ , consider the form (9) of the Edgeworth-type expansion, where  $\widehat{D}$  is defined as (4) with expectation and variance given by Theorem 2.6. Substituting the cumulants into  $a_i$ ,  $i \in \{0, 1, 2, 3\}$ , given by (10), we get the results. ■

Moreover, from Theorem 2.8, the limiting distribution of  $\widehat{D}$  is  $\mathcal{N}(\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is said to come from  $\pi_1$  and is  $\mathcal{N}(-\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  is said to come from  $\pi_2$ . The following theorem depicts the standardized Edgeworth-type expansion of  $\widehat{D}$ .

**Theorem 3.4:** *Let  $\widehat{D}$  be the discriminant function defined in (4) and  $\Delta^2$  is as in (13) and let  $z = \frac{\widehat{D} - \frac{1}{2}\Delta^2}{\Delta}$ . Then, for  $m = n_1 + n_2 - p - 3$ ,  $q$  in (12) and  $c_0 = \frac{m+p}{m(m-2)(m+1)}$ , the*

Edgeworth-type expansion,  $f_z(x)$  when the observation  $\mathbf{x}$  is classified in population  $\pi_1$  equals

$$f_{\widehat{D}}(x) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \frac{1}{2} \{w_0 + w_1x + w_2x^2 + w_3x^3\}, \tag{23}$$

where

$$\begin{aligned} w_0 &= -\frac{1}{\Delta^2} (1 - q^2) m^2 c_0 p + 3 - m^2 c_0 \frac{n_2}{n_2 + 1} - \frac{1}{2} \Delta^2 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{m}{2(m - 2)} - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right), \\ w_1 &= \frac{1}{\Delta} \frac{3}{2} (1 - q^2) m^2 c_0 p \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{2} \Delta \left( 3m^2 c_0 \frac{n_2}{n_2 + 1} - 5 \right) \left( \frac{n_2}{n_2 + 1} - 1 \right) \\ &\quad + \frac{1}{4} \Delta^3 \left( 3 \left( \frac{n_2}{n_2 + 1} \right)^3 \frac{1}{m - 2} - 3 \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{m - 2} + \frac{n_2}{n_2 + 1} - 1 \right), \\ w_2 &= \frac{1}{\Delta^2} (1 - q^2) m^2 c_0 p - 1 + m^2 c_0 \frac{n_2}{n_2 + 1} + \frac{1}{2} \Delta^2 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{m}{2(m - 2)} - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right), \\ w_3 &= \frac{1}{\Delta} \frac{1}{2} (1 - q^2) m^2 c_0 p \left( \frac{n_2}{n_2 + 1} - 1 \right) + \Delta \frac{1}{2} \left( m^2 c_0 \frac{n_2}{n_2 + 1} - \frac{4}{3} \right) \left( \frac{n_2}{n_2 + 1} - 1 \right) \\ &\quad + \frac{1}{4} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{m - 2} \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{3} \frac{n_2}{n_2 + 1} - \frac{1}{3} \right). \end{aligned} \tag{24}$$

**Proof:** Consider  $E[\widehat{D}]$  and  $Var[\widehat{D}]$  given by Theorem 2.6. If  $\mathbf{x} \in \pi_1$  then,

$$E[z] = \frac{1}{\Delta} E[\widehat{D}] - \frac{1}{2} \Delta = \frac{1}{2} \Delta \left( \frac{n_2}{n_2 + 1} - 1 \right)$$

and

$$\begin{aligned} Var[z] &= \frac{1}{\Delta^2} Var[\widehat{D}] \\ &= \frac{1}{\Delta^2} (1 - q^2) m^2 c_0 p + m^2 c_0 \frac{n_2}{n_2 + 1} + \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{2(m - 2)} \Delta^2. \end{aligned}$$

Now consider the Edgeworth-type expansion of the form (9). Hence, if  $\mathbf{x} \in \pi_1$ , replacing the expected value and variance of  $z$  in  $a_i, i \in \{0, 1, 2, 3\}$  given by (10) by their values yields

$$\begin{aligned} w_0 &= 3 - Var[z] - (E[z])^2 \\ &= -\frac{1}{\Delta^2} (1 - q^2) m^2 c_0 p + 3 - m^2 c_0 \frac{n_2}{n_2 + 1} - \frac{1}{2} \Delta^2 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{m}{2(m - 2)} - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right), \\ w_1 &= 3Var[z]E[z] - 5E[z] + (E[z])^2 \\ &= \frac{1}{\Delta} \frac{3}{2} (1 - q^2) m^2 c_0 p \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{2} \Delta \left( 3m^2 c_0 \frac{n_2}{n_2 + 1} - 5 \right) \left( \frac{n_2}{n_2 + 1} - 1 \right) \\ &\quad + \frac{1}{4} \Delta^3 \left( 3 \left( \frac{n_2}{n_2 + 1} \right)^3 \frac{1}{m - 2} - 3 \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{m - 2} + \frac{n_2}{n_2 + 1} - 1 \right), \end{aligned}$$

$$\begin{aligned}
 w_2 &= 2 - a_0 \\
 &= \frac{1}{\Delta^2} (1 - q^2) m^2 c_0 p - 1 + m^2 c_0 \frac{n_2}{n_2 + 1} + \frac{1}{2} \Delta^2 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{m}{2(m - 2)} - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right), \\
 w_3 &= \frac{2}{3} E[z] + \frac{1}{3} a_1 \\
 &= \frac{1}{\Delta} \frac{1}{2} (1 - q^2) m^2 c_0 p \left( \frac{n_2}{n_2 + 1} - 1 \right) + \Delta \frac{1}{2} \left( m^2 c_0 \frac{n_2}{n_2 + 1} - \frac{4}{3} \right) \left( \frac{n_2}{n_2 + 1} - 1 \right) \\
 &\quad + \frac{1}{4} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^2 \frac{1}{m - 2} \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{3} \frac{n_2}{n_2 + 1} - \frac{1}{3} \right). \tag{25}
 \end{aligned}$$

Similar calculations if  $\mathbf{x} \in \pi_2$  concludes the proof. ■

### 4. Misclassification errors

It is not enough to make a decision based on a classification rule. In addition to a classification rule, it is important to study probabilities of misclassifications in order to know if a classification approach is reliable. The probability of misclassification is a measure of the goodness of the proposed classification rule as stated by Fujikoshi et al. [11]. The classification of the new observation  $\mathbf{x}$  depends on whether the discriminant function is greater than zero or not. Denote by  $P(\mathbf{x} \rightarrow \pi_i)$  the probability that an observation  $\mathbf{x}$  is classified into population  $\pi_i$ . Let  $i, j \in \{1, 2\}$ , then  $P(\mathbf{x} \rightarrow \pi_i | \mathbf{x} \in \pi_i)$  is the probability of correctly classifying the observation  $\mathbf{x}$  into population  $\pi_i$  and  $P(\mathbf{x} \rightarrow \pi_j | \mathbf{x} \in \pi_i)$  is the probability that the observation  $\mathbf{x}$  comes from population  $\pi_i$  but is classified as coming from  $\pi_j$ . In this section, we investigate misclassification errors of the quadratic classifiers  $\tilde{D}$  in (3) and  $\hat{D}$  given in (4). The probabilities of misclassifications for the classification functions  $\tilde{D}$  and  $\hat{D}$  can be estimated by integrating the Edgeworth-type expansions of the discriminant functions through the density of the standard normal distribution. As we observed before, the exact distributions of  $\tilde{D}$  and  $\hat{D}$  are complicated to calculate. Therefore, one would like to approximate the misclassification errors to know how good these classification rules are.

According to Theorem 2.8 we assume the asymptotic distribution of  $\hat{D}$  and  $\tilde{D}$  to be  $\mathcal{N}(\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  comes from  $\pi_2$  and to be  $\mathcal{N}(-\frac{1}{2}\Delta^2, \Delta^2)$  if  $\mathbf{x}$  comes from  $\pi_2$ .

**Lemma 4.1:** *Let  $a_i, i \in \{0, 1, 2, 3\}$  be given by (10) and  $\Delta$  in (13). Let  $\Phi(\cdot)$  and  $\phi(\cdot)$  be the cumulative distribution function and density of  $\mathcal{N}(0, 1)$ , respectively. For the discriminant function  $D$ ,*

$$\begin{aligned}
 &P \left\{ \frac{D - \frac{1}{2}\Delta^2}{\Delta} \leq -\frac{1}{2}\Delta | \mathbf{x} \in \pi_1 \right\} \\
 &\approx \Phi \left( -\frac{1}{2}\Delta \right) + \phi \left( \frac{1}{2}\Delta \right) \left( \frac{1}{2}a_1 + \frac{1}{4}\Delta a_2 - \left( 1 + \frac{1}{8}\Delta^2 \right) a_3 \right), \tag{26}
 \end{aligned}$$

and  $P\left\{ \frac{-D + \frac{1}{2}\Delta^2}{\Delta} \leq -\frac{1}{2}\Delta | \mathbf{x} \in \pi_2 \right\}$  is the same but with opposite signs on coefficients of  $a_1$  and  $a_3$ , where  $D$  can either be  $\tilde{D}$  defined in (3) or  $\hat{D}$  defined in (4).

**Remark 4.1:** Note that the values of  $a_1$ ,  $a_2$  and  $a_3$  depend on whether it is  $\widehat{D}$  or  $\widetilde{D}$  that is approximated.

**Proof:** Consider the Edgeworth-type expansion of the form (9) and put  $t = \frac{D - \frac{1}{2}\Delta^2}{\Delta}$ . Then, using integration by parts

$$\begin{aligned} P(\mathbf{x} \rightarrow \pi_2 | \mathbf{x} \in \pi_1) &\approx P\left\{ \frac{D - \frac{1}{2}\Delta^2}{\Delta} \leq -\frac{1}{2}\Delta | \mathbf{x} \in \pi_1 \right\} = \int_{-\infty}^{-\frac{1}{2}\Delta} f_t(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{1}{2}\Delta} e^{-\frac{1}{2}x^2} \frac{1}{2} \{a_0 - a_1x + a_2x^2 + a_3x^3\} dx \\ &= \Phi\left(-\frac{1}{2}\Delta\right) \frac{1}{2}(a_0 + a_2) + \phi\left(-\frac{1}{2}\Delta\right) \left(\frac{1}{2}a_1 + \frac{1}{4}\Delta a_2 - \left(1 + \frac{1}{8}\Delta^2\right)a_3\right) \end{aligned}$$

and the proof follows since

$$\frac{1}{2}(a_0 + a_2) = \frac{1}{2} (3 - \text{Var}(t) - (E(t))^2 + \text{Var}(t) - 1 + (E(t))^2) = 1.$$

■

The classification rule using  $\widetilde{D}$  is to assign  $\mathbf{x}$  to  $\pi_1$  if  $\widetilde{D} \geq 0$  and to  $\pi_2$  if  $\widetilde{D} < 0$ . The probabilities of misclassification are given by Theorem 4.2 below.

**Theorem 4.2:** Consider the classification function,  $\widetilde{D}$ , given by (3). Let  $\Delta^2$  be given by (13) and let  $\Phi(\cdot)$  and  $\phi(\cdot)$  be the cumulative distribution function and density of  $\mathcal{N}(0, 1)$ , respectively. Then, with  $q$  given by (12),

$$\begin{aligned} P(\mathbf{x} \rightarrow \pi_2 | \mathbf{x} \in \pi_1) &\approx \Phi\left(-\frac{1}{2}\Delta\right) + \frac{1}{2}\phi\left(\frac{1}{2}\Delta\right) \left\{ \frac{1}{2} \frac{1}{\Delta} \frac{n_2}{n_2 + 1} p(1 - q^2) + \frac{3}{2} \left(\frac{n_2}{n_2 + 1}\right)^2 - 4 \frac{n_2}{n_2 + 1} + \frac{5}{2} \right. \\ &\quad - \Delta \left( \left(\frac{n_2}{n_2 + 1}\right)^3 - \frac{3}{2} \frac{n_2}{n_2 + 1} - \frac{1}{16} p(1 - q^2) \left(\frac{n_2}{n_2 + 1} - 1\right) + \frac{1}{2} \right) \\ &\quad + \frac{1}{2} \Delta^3 \left( \frac{1}{3} \left(\frac{n_2}{n_2 + 1}\right)^3 + \left(\frac{n_2}{n_2 + 1}\right)^2 - \frac{1}{2} \frac{n_2}{n_2 + 1} - \frac{3}{4} \right) \\ &\quad \left. + \frac{1}{64} \Delta^5 \left( \left(\frac{n_2}{n_2 + 1}\right)^3 - 3 \left(\frac{n_2}{n_2 + 1}\right)^2 + 3 \frac{n_2}{n_2 + 1} - 1 \right) \right\}, \tag{27} \end{aligned}$$

$$P(\mathbf{x} \rightarrow \pi_1 | \mathbf{x} \in \pi_2)$$

$$\approx \Phi\left(-\frac{1}{2}\Delta\right) + \frac{1}{4}\phi\left(\frac{1}{2}\Delta\right) \left\{ \frac{1}{2} \frac{1}{\Delta} \frac{n_1}{n_1 + 1} p(1 - q^2) + \frac{3}{2} \left(\frac{n_1}{n_1 + 1}\right)^2 - 4 \frac{n_1}{n_1 + 1} + \frac{5}{2} \right.$$

$$\begin{aligned}
 & - \Delta \left( \left( \frac{n_1}{n_1 + 1} \right)^3 - \frac{3}{2} \frac{n_1}{n_1 + 1} - \frac{1}{16} p(1 - q^2) \left( \frac{n_1}{n_1 + 1} - 1 \right) + \frac{1}{2} \right) \\
 & + \frac{1}{2} \Delta^3 \left( \frac{1}{3} \left( \frac{n_1}{n_1 + 1} \right)^3 + \left( \frac{n_1}{n_1 + 1} \right)^2 - \frac{1}{2} \frac{n_1}{n_1 + 1} - \frac{3}{4} \right) \\
 & + \frac{1}{64} \Delta^5 \left( \left( \frac{n_1}{n_2 + 1} \right)^3 - 3 \left( \frac{n_1}{n_1 + 1} \right)^2 + 3 \frac{n_1}{n_1 + 1} - 1 \right) \Big\}, \tag{28}
 \end{aligned}$$

The difference between the two expressions in (27) and (28) is that  $n_1$  and  $n_2$  are interchanged.

**Proof:** Consider the misclassification probability in Lemma 4.1 with the constants  $a_i, i \in \{1, 2, 3\}$ , in (20) and the cumulants of  $\tilde{D}$  from Theorem 2.5. If  $\mathbf{x} \in \pi_1$

$$\begin{aligned}
 \frac{1}{2} a_1 + \frac{1}{4} \Delta a_2 - (1 + \frac{1}{8} \Delta^2) a_3 &= \frac{3}{4} \frac{1}{\Delta} p(1 - q^2) \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{2} \left( \frac{3}{2} \left( \frac{n_2}{n_2 + 1} \right)^2 - 4 \frac{n_2}{n_2 + 1} + \frac{5}{2} \right) \\
 & + \frac{1}{16} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 + 3 \left( \frac{n_2}{n_2 + 1} \right)^2 - 1 \right) + \frac{1}{4} \frac{1}{\Delta} p(1 - q^2) \\
 & - \left( \frac{n_2}{n_2 + 1} - 1 \right) \frac{1}{4} \Delta - \frac{1}{8} \Delta^3 \left( \frac{1}{2} \left( \frac{n_2}{n_2 + 1} \right)^2 - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right) \\
 & - \frac{1}{2} \frac{1}{\Delta} p(1 - q^2) \left( \frac{n_2}{n_2 + 1} - 1 \right) - \Delta \left( \frac{1}{2} \left( \frac{n_2}{n_2 + 1} \right)^3 - \frac{n_2}{n_2 + 1} + \frac{1}{2} \right) \\
 & - \frac{1}{24} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 - 3 \left( \frac{n_2}{n_2 + 1} \right)^2 + 3 \frac{n_2}{n_2 + 1} - 1 \right) \\
 & + \frac{1}{16} \Delta p(1 - q^2) \left( \frac{n_2}{n_2 + 1} - 1 \right) + \frac{1}{8} \Delta^3 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 + 3 \frac{n_2}{n_2 + 1} - 1 \right) \\
 & + \frac{1}{128} \Delta^5 \left( \left( \frac{n_2}{n_2 + 1} \right)^3 - 3 \left( \frac{n_2}{n_2 + 1} \right)^2 + 3 \frac{n_2}{n_2 + 1} - 1 \right) \\
 & = (p(1 - q^2) - 1) \Delta + (p(1 - q^2) - 3) \frac{n_2}{n_2 + 1} \Delta^2 + \frac{n_2}{n_2 + 1} \Delta^3 \\
 & + \left( \frac{n_2}{n_2 + 1} - \frac{1}{2} p(1 - q^2) \right) \frac{n_2}{n_2 + 1} \Delta^4 + \frac{1}{4} \left( \frac{n_2}{n_2 + 1} \right)^2 \Delta^5 \\
 & - \frac{1}{2} \left( \frac{1}{6} \frac{n_2}{n_2 + 1} + 1 \right) \left( \frac{n_2}{n_2 + 1} \right)^2 \Delta^6 - \frac{1}{24} \left( \frac{n_2}{n_2 + 1} \right)^3 \Delta^8.
 \end{aligned}$$

Hence the expression in (27) follows. Furthermore, similar calculations when  $\mathbf{x} \in \pi_2$  conclude the proof. ■

The classification rule using  $\widehat{D}$  is to classify  $\mathbf{x}$  to  $\pi_1$  if  $\widehat{D} \geq 0$  and to  $\pi_2$  if  $\widehat{D} < 0$ . The theorem below expresses the misclassification errors for this rule.

**Theorem 4.3:** Consider the classification function,  $\widehat{D}$ , given by (4) and  $\Delta^2$  given by (13). Then, for  $m = n_1 + n_2 - p - 3$ ,  $q$  in (12) and  $c_0 = \frac{m+p}{m(m-2)(m+1)}$ , let  $\Phi(\cdot)$  and  $\phi(\cdot)$  be the cumulative distribution function and density of  $\mathcal{N}(0, 1)$ , respectively. Then,

$$\begin{aligned}
 P(\mathbf{x} \rightarrow \pi_2 | \mathbf{x} \in \pi_1) &\approx \Phi\left(-\frac{1}{2}\Delta\right) + \frac{1}{2}\phi\left(\frac{1}{2}\Delta\right) \left\{ \frac{1}{\Delta} \frac{1}{4}(1-q^2)m^2c_0p \frac{n_2}{n_2+1} \right. \\
 &+ \frac{1}{2}\Delta \left[ \frac{1}{2}m^2c_0 \left(\frac{n_2}{n_2+1}\right)^2 - \frac{1}{8}(1-q^2)m^2c_0p \left(\frac{n_2}{n_2+1} - 1\right) \right. \\
 &- \left. \frac{3}{4} \frac{n_2}{n_2+1} + \frac{1}{2} \right] + \frac{1}{2}\Delta^3 \left[ \frac{1}{24} \left(\frac{n_2}{n_2+1}\right)^3 \frac{12+m^2(m-2)}{m^2(m-2)} \right. \\
 &- \left. \frac{1}{16} \left(\frac{n_2}{n_2+1}\right)^2 \frac{4+m^2(m-2)}{m^2(m-2)} - \frac{1}{8}m^2c_0 \frac{n_2}{n_2+1} \left(\frac{n_2}{n_2+1} - 1\right) - \frac{1}{24} \right] \\
 &- \left. \frac{1}{192}\Delta^5 \left[ \left(\frac{n_2}{n_2+1}\right)^3 \frac{12+m^2(m-2)}{m^2(m-2)} \right. \right. \\
 &- \left. \left. 3 \left(\frac{n_2}{n_2+1}\right)^2 \frac{4+m^2(m-2)}{m^2(m-2)} + 3 \frac{n_2}{n_2+1} - 1 \right] \right\}, \\
 P(\mathbf{x} \rightarrow \pi_1 | \mathbf{x} \in \pi_2) &\approx \Phi\left(-\frac{1}{2}\Delta\right) + \frac{1}{2}\phi\left(\frac{1}{2}\Delta\right) \left\{ \frac{1}{\Delta} \frac{1}{4}(1-q^2)m^2c_0p \frac{n_1}{n_1+1} \right. \\
 &+ \frac{1}{2}\Delta \left[ \frac{1}{2}m^2c_0 \left(\frac{n_1}{n_1+1}\right)^2 - \frac{1}{8}(1-q^2)m^2c_0p \left(\frac{n_1}{n_1+1} - 1\right) \right. \\
 &- \left. \frac{3}{4} \frac{n_1}{n_1+1} + \frac{1}{2} \right] + \frac{1}{2}\Delta^3 \left[ \frac{1}{24} \left(\frac{n_1}{n_1+1}\right)^3 \frac{12+m^2(m-2)}{m^2(m-2)} \right. \\
 &- \left. \frac{1}{16} \left(\frac{n_1}{n_1+1}\right)^2 \frac{4+m^2(m-2)}{m^2(m-2)} - \frac{1}{8}m^2c_0 \frac{n_1}{n_1+1} \left(\frac{n_1}{n_1+1} - 1\right) - \frac{1}{24} \right] \\
 &- \left. \frac{1}{192}\Delta^5 \left[ \left(\frac{n_1}{n_1+1}\right)^3 \frac{12+m^2(m-2)}{m^2(m-2)} \right. \right. \\
 &- \left. \left. 3 \left(\frac{n_1}{n_1+1}\right)^2 \frac{4+m^2(m-2)}{m^2(m-2)} + 3 \frac{n_1}{n_1+1} - 1 \right] \right\}. \tag{29}
 \end{aligned}$$

**Proof:** Consider  $a_i, i \in \{1, 2, 3\}$ , given by (24). If  $\mathbf{x} \in \pi_2$ , then from Lemma 4.1 we get

$$\begin{aligned}
 \frac{1}{2}a_1 + \frac{1}{4}\Delta a_2 - (1 + \frac{1}{8}\Delta^2)a_3 &= \frac{3}{4}\frac{1}{\Delta}(1-q^2)m^2c_0p \left(\frac{n_2}{n_2+1} - 1\right) + \frac{1}{4}\Delta \left(\frac{n_2}{n_2+1} - 1\right) \left(3m^2c_0 \frac{n_2}{n_2+1} - 5\right) \\
 &+ \frac{1}{16}\Delta^3 \left(\left(\frac{n_2}{n_2+1}\right)^3 \frac{12+m^2(m-2)}{m^2(m-2)} - 3\Delta^3 \left(\frac{n_2}{n_2+1}\right)^3 \frac{4+m^2(m-2)}{m^2(m-2)}\right)
 \end{aligned}$$

$$\begin{aligned}
 & +3\frac{n_2}{n_2+1} - 1) + \frac{1}{4\Delta}(1-q^2)m^2c_0p - \frac{1}{4}\Delta\left(1-m^2c_0\frac{n_2}{n_2+1}\right) \\
 & + \frac{1}{8}\Delta^3\left(\left(\frac{n_2}{n_2+1}\right)^2\frac{4+m^2(m-2)}{m^2(m-2)} - \frac{n_2}{n_2+1} + 1\right) \\
 & - \frac{1}{2\Delta}(1-q^2)m^2c_0p\left(\frac{n_2}{n_2+1} - 1\right) - \frac{1}{2}\Delta\left(\frac{n_2}{n_2+1} - 1\right)\left(m^2c_0\frac{n_2}{n_2+1} - 1\right) \\
 & - \frac{1}{8}\Delta^3\left(\frac{1}{3}\left(\frac{n_2}{n_2+1}\right)^3\frac{12+m^2(m-2)}{m^2(m-2)} + \left(\frac{n_2}{n_2+1}\right)^2\frac{4+m^2(m-2)}{m^2(m-2)}\right. \\
 & \left. - \frac{n_2}{n_2+1} + \frac{1}{3}\right) - \frac{1}{16}\Delta(1-q^2)m^2c_0p\left(\frac{n_2}{n_2+1} - 1\right) \\
 & - \frac{1}{16}\Delta^3\left(\frac{n_2}{n_2+1} - 1\right)\left(m^2c_0\frac{n_2}{n_2+1} - 1\right) + \frac{1}{192}\Delta^5\left(\frac{1}{3}\left(\frac{n_2}{n_2+1}\right)^3\right. \\
 & \left.\times\frac{12+m^2(m-2)}{m^2(m-2)} - 3\left(\frac{n_2}{n_2+1}\right)^2\frac{4+m^2(m-2)}{m^2(m-2)} + 3\frac{n_2}{n_2+1} - 1\right) \\
 & = \frac{1}{4}\frac{1}{\Delta}(1-q^2)m^2c_0p\frac{n_2}{n_2+1} + \frac{1}{2}\Delta\left[\frac{1}{2}m^2c_0\left(\frac{n_2}{n_2+1}\right)^2\right. \\
 & \left.- \frac{1}{8}(1-q^2)m^2c_0p\left(\frac{n_2}{n_2+1} - 1\right) - \frac{3}{4}\frac{n_2}{n_2+1} + \frac{1}{2}\right] \\
 & + \frac{1}{2}\Delta^3\left[\frac{1}{24}\left(\frac{n_2}{n_2+1}\right)^3\frac{12+m^2(m-2)}{m^2(m-2)} - \frac{1}{16}\left(\frac{n_2}{n_2+1}\right)^2\right. \\
 & \left.\times\frac{4+m^2(m-2)}{m^2(m-2)} - \frac{1}{8}m^2c_0\frac{n_2}{n_2+1}\left(\frac{n_2}{n_2+1} - 1\right) - \frac{1}{24}\right] \\
 & - \frac{1}{192}\Delta^5\left[\left(\frac{n_2}{n_2+1}\right)^3\frac{12+m^2(m-2)}{m^2(m-2)}\right. \\
 & \left.- 3\left(\frac{n_2}{n_2+1}\right)^2\frac{4+m^2(m-2)}{m^2(m-2)} + 3\frac{n_2}{n_2+1} - 1\right].
 \end{aligned}$$

Therefore, substituting the above expression into (26) the proof follows. To prove a similar result when  $x \in \pi_2$  we can use the same approach as when  $x \in \pi_1$  and the result is established. ■

### 5. Conclusion

We have observed that it is difficult to evaluate the exact distributions of  $\tilde{D}$  and  $\hat{D}$ . Therefore, we handled this problem by performing Edgeworth-type expansions in order to further obtain misclassification probabilities. Note that an Edgeworth-type expansion does not have to be a density. In our case, the integration of both Edgeworth-type expansions of

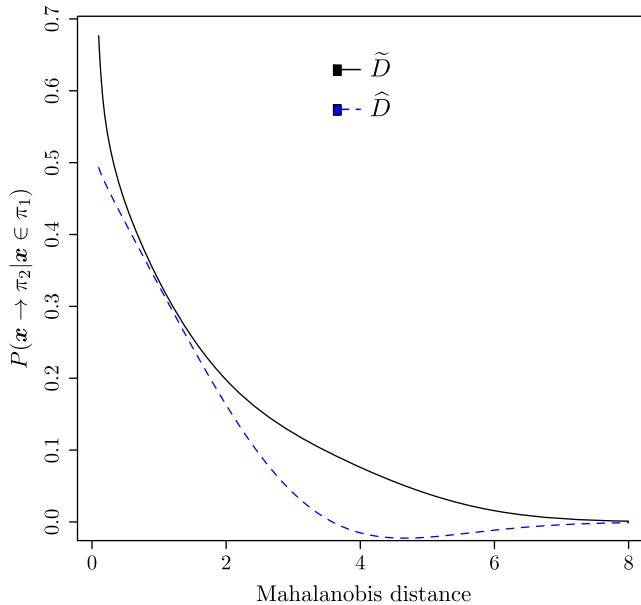
the discriminant functions over the entire space equals one, that is,

$$\int_{-\infty}^{+\infty} f_{\tilde{D}}(x) dx = \int_{-\infty}^{+\infty} f_{\hat{D}}(x) dx = 1.$$

However, we have not shown nonnegativity of the expansions. The Edgeworth-type expansion was performed in terms of cumulants of the discriminant functions  $\hat{D}$  and  $\tilde{D}$  through a standard normal distribution up to three terms. The expansion is a polynomial function in  $\mathbf{x}$ .

In addition, in order to estimate misclassification probabilities for the classification rules whether  $\tilde{D} > 0$  or  $\tilde{D} < 0$  and whether  $\hat{D} > 0$  or  $\hat{D} < 0$ , we used the Edgeworth-type expansions  $f_{\tilde{D}}(x)$  and  $f_{\hat{D}}(x)$  of standardized variables with their limiting distribution given in Theorem 2.8. Misclassification errors from Theorem 4.2 and Theorem 4.3 are functions of the Mahalanobis distance. Figure 1 shows the change in misclassification probabilities using  $\tilde{D}$  and  $\hat{D}$  with respect to the Mahalanobis distance ( $\Delta$ ), where  $P(\mathbf{x} \rightarrow \pi_2 | \mathbf{x} \in \pi_1)$  is the probability that an observation  $\mathbf{x}$  is wrongly classified into population  $\pi_2$ . The curves for the discrimination rules with  $\tilde{D}$  and  $\hat{D}$  are more or less similar.

Now, let us discuss our results in comparison to well-known results. Consider the discriminant functions  $W$  given by (1) and  $Z$  in (2). Like the discriminant function  $\hat{D}$ , they both involve a known covariance matrix. However, the  $W$ -rule is a linear function whereas the  $Z$ -rule is a quadratic function, similarly to  $\tilde{D}$  and  $\hat{D}$  discriminant functions. Note that when  $n_1 = n_2 \equiv n$ ,  $\hat{D} = -\frac{mn}{n+1} W = mZ$ , where  $m = 2n - p - 3$ . The following two theorems by Anderson [12] illustrate misclassification errors with respect to  $W$  and  $Z$  classification functions, respectively.



**Figure 1.** Misclassification errors for  $\tilde{D}$  in (3) and  $\hat{D}$  in (4) rules when  $n_1 = n_2 = 100$  and  $p = 10$ .



**Theorem 5.1:** For large  $n_1$  and  $n_2$ , and  $n = n_1 + n_2 - 2$  and  $u = -\frac{1}{2}\Delta$ , where  $\Delta$  is the Mahalanobis distance,

$$\begin{aligned}
 P\left\{\frac{W - \frac{1}{2}\Delta^2}{\Delta} \leq u \mid \mathbf{x} \in \pi_1\right\} &= \Phi(u) - \phi(u) \left\{ \frac{1}{2n_1\Delta^2} [u^3 + (p-3)u - p\Delta] \right. \\
 &\quad + \frac{1}{2n_2\Delta^2} [u^3 + 2\Delta u^2 + (p-3 + \Delta^2)u + (p-2)\Delta] \\
 &\quad \left. + \frac{1}{4n} [4u^3 + 4\Delta u^2 + (6p-6 + \Delta^2)u + 2(p-1)\Delta] \right\} + O(n^{-2})
 \end{aligned}
 \tag{30}$$

and  $P\{-\left(\frac{W + \frac{1}{2}\Delta^2}{\Delta}\right) \leq u \mid \mathbf{x} \in \pi_2\}$  is similar to (30) (only  $n_1$  and  $n_2$  have to be interchanged).

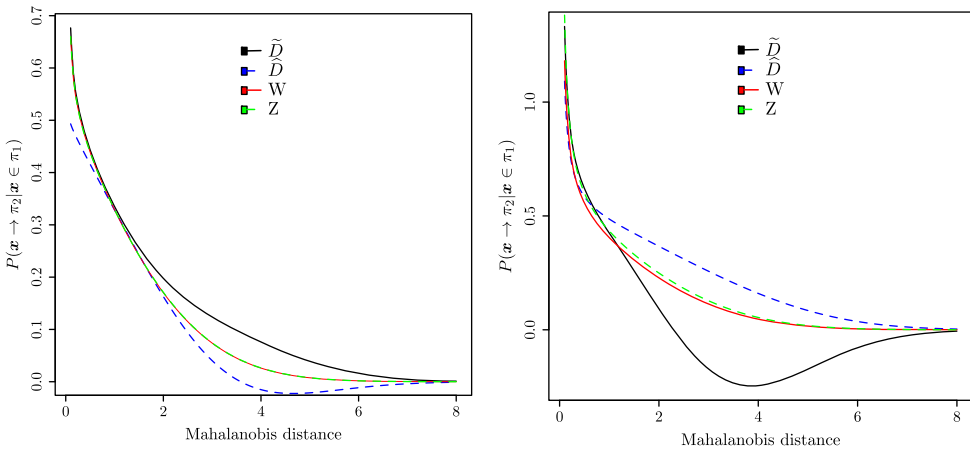
**Theorem 5.2:** For large  $n_1$  and  $n_2$ ,  $n_1/n_2$  and  $n = n_1 + n_2 - 2$  and  $u = -\frac{1}{2}\Delta$ ,

$$\begin{aligned}
 P\left\{\frac{Z - \frac{1}{2}\Delta^2}{\Delta} \leq u \mid \mathbf{x} \in \pi_1\right\} &= \Phi(u) - \phi(u) \left\{ \frac{1}{2n_1\Delta^2} [u^3 + \Delta u^2 + (p-3)u - \Delta] \right. \\
 &\quad + \frac{1}{2n_2\Delta^2} [u^3 + \Delta u^2 + (p-3 + \Delta^2)u - \Delta^3 - \Delta] \\
 &\quad \left. + \frac{1}{4n} [4u^3 + 4\Delta u^2 + (6p-6 + \Delta^2)u + 2(p-1)\Delta] \right\} + O(n^{-2}),
 \end{aligned}
 \tag{31}$$

and  $P\{-\left(\frac{Z + \frac{1}{2}\Delta^2}{\Delta}\right) \leq u \mid \mathbf{x} \in \pi_2\}$  equals (31) with  $n_1$  and  $n_2$  interchanged.

Proofs of Theorem 5.1 and Theorem 5.2 are provided by Anderson [12].

Figure 2(a,b) display the change in misclassification probabilities using the rules  $\tilde{D}$ ,  $\hat{D}$ ,  $W$  and  $Z$  with respect to  $\Delta$ . We can see from both figures that the rules  $\hat{D}$  and  $\tilde{D}$  are comparably



**Figure 2.** Misclassification errors for  $\tilde{D}$ ,  $\hat{D}$ ,  $W$  and  $Z$  rules. (a)  $n_1 = n_2 = 100$  and  $p = 10$  and (b)  $n_1 = 5$ ,  $n_2 = 4$  and  $p = 3$ .

similar to  $W$  and  $Z$  rules. However, Figure 1 also shows that the rule with  $\widehat{D}$  may have negative misclassification error. Moreover, for small  $n$ , ( $n < 10$ ), see Figure 2(b), misclassification errors using  $\widetilde{D}$  can be negative. This is due to the fact that the Egdeworth-type expansion is not always a density. Hence, misclassification errors are not really probabilities and might yield irrelevant values. However, a negative value can be interpreted that the misclassification error is small.

We consider classification only if the misclassification error is less than 50%. Therefore, it is not meaningful to classify with  $\widetilde{D}$  when  $\Delta < 0.4$ . On the other hand, the misclassification error using  $\widehat{D}$  is always less than 50%.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Emelyne Umunoza Gasana  <http://orcid.org/0000-0002-5559-4120>

Dietrich von Rosen  <http://orcid.org/0000-0002-3135-4325>

Martin Singull  <http://orcid.org/0000-0001-9896-4438>

## References

- [1] Barnard MM. The secular variations of skull characters in four series of Egyptian skulls. *Ann Eugen.* 1935;6:352–371.
- [2] Fisher RA. The use of multiple measurements in 10 taxonomic problems. *Ann Eugen.* 1936;7:179–188.
- [3] Fisher RA. The statistical utilization of multiple measurements. *Ann Eugen.* 1938;8:376–386.
- [4] Mahalanobis PC. Analysis of race-mixture in Bengal. *J Proc Asiat Soc Bengal New Ser.* 1925;23:301–333.
- [5] Pearson K. On the problem of sexing osteometric material. *Biometrika.* 1915;10(4):479–487.
- [6] Anderson TW. Classification by multivariate analysis. *Psychometrika.* 1951;16:31–50.
- [7] Kudo A. The classificatory problem viewed as a two-decision problem. *Mem Fac Sci Kyushu Univ Ser A Math.* 1959;13:96–125.
- [8] Srivastava MS, Khatri C. An introduction to multivariate statistics. New York: Elsevier North-Holland; 1979.
- [9] Kudo A. The classificatory problem viewed as a two-decision problem-II. *Mem Fac Sci Kyushu Univ Ser A Math.* 1960;14:63–83.
- [10] Gasana EU, von Rosen D, Singull M. Moments of the likelihood-based discriminant function. *Commun Statist-Theory Methods.* 2022;1–13.
- [11] Fujikoshi Y, Ulyanov VV, Shimizu R. Multivariate statistics: high-dimensional and large-sample approximations. Vol. 760. Hoboken: John Wiley & Sons, Inc.; 2011.
- [12] Anderson TW. An introduction to multivariate statistical analysis. 3rd ed. Hoboken: John Wiley & Sons, Inc.; 2003.
- [13] Sitgreaves R. On the distribution of two random matrices used in classification procedures. *Ann Math Stat.* 1952;23:263–270.
- [14] Sitgreaves R. Some results on the distribution of the  $W$ -classification statistic. Chapter 19 in Solomon H, editor. *Studies in Item Analysis and Prediction.* 1961. p. 241–261.
- [15] Wald A. On a statistical problem arising in the classification of an individual into one of two groups. *Ann Math Stat.* 1944;15:145–162.
- [16] Okamoto M. An asymptotic expansion for the distribution of the linear discriminant function. *Ann Math Stat.* 1963;34:1286–1301.

- [17] Memon AZ, Okamoto M. Asymptotic expansion of the distribution of the  $Z$  statistic in discriminant analysis. *J Multivar Anal.* **1971**;1(3):294–307.
- [18] Siotani M. Asymptotic expansions for error rates and comparison of the  $W$ -procedure and the  $Z$ -procedure in discriminant analysis. Amsterdam: North-Holland; **1977**.
- [19] Siotani M, Wang R-H. Further Expansion Formulae for Error Rates and Comparison of the  $W$ -and  $Z$ -procedures in Discriminant Analysis. Manhattan, Kansas: Department of Statistics, Kansas State University; 1975. (Technical Report No. 33).
- [20] Laplace P-S. Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mémoires de l'Académie Royale des Sciences de Paris*, 1811. p. 279–347.
- [21] Gram JP. Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode [PhD thesis]. København: Andr. Fred. Høst & Son; 1879.
- [22] Cornish EA, Fisher RA. Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique*, 1938. p. 307–320.
- [23] Edgeworth F. The law of error. *Trans Cambridge Philoso Soc.* **1905**;20:36–66. 113–141.
- [24] Edgeworth F. On the representation of statistical frequency by a series. *J R Stat Soc.* **1907**;70(1):102–106.
- [25] Thiele TN. Om en tilnærmelsesformel. *J Math.* **1873**;3:22–31.
- [26] Thiele TN. Forlaesinger over almindelig iagttagelseslaere: sandsynlighedsregning og mindste kvadraters metode. København: Reitzel; **1889**.
- [27] Hald A. The early history of the cumulants and the gram-charlier series. *Int Stat Rev.* **2000**;68(2):137–153.
- [28] Hald A, Steffensen J. On the history of series expansions of frequency functions and sampling distributions, 1873–1944. Copenhagen: Det Kongelige Danske Videnskabernes Selskab; **2002**.
- [29] Lauritzen SL, Thiele TN, Thiele J, et al. Thiele: pioneer in statistics. New York: Oxford University Press Inc.; **2002**.
- [30] Gupta SS, Panchapakesan S. Edgeworth expansions in statistics: a brief review. Purdue university Lafayette in department of statistics; 1982. (Technical report).
- [31] Ranga Rao R. Some problems in probability theory [PhD thesis]. Calcutta University; 1960.
- [32] Davis A. Statistical distributions in univariate and multivariate edgeworth populations. *Biometrika.* **1976**;63:661–670.
- [33] Kollo T, von Rosen D. Advanced multivariate statistics with matrices. Vol. 579. Dordrecht: Springer Science & Business Media; **2005**.