International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 3 (Special Issue), 2004, pp 209–219

Discrimination Between Two Generalized Multivariate Modified Bessel Populations

Lehana Thabane
Department of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton ON L8N 1G6, Canada
E-mail: thabanl@mcmaster.ca

Steve Drekic
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo ON N2L 3G1, Canada
E-mail: sdrekic@math.uwaterloo.ca

[Received March 8, 2004; Accepted May 20, 2004]

Abstract

Most of the classical literature on discriminant analysis assumes that the underlying populations are normally distributed. In this paper, we investigate discrimination problems under the assumption that the populations originate from a generalized multivariate modified Bessel distribution, which is a much more general model that includes both the multivariate normal and t distributions as special cases. We examine Fisher's linear discrimination criterion and the probabilities of misclassification under this model. It is shown that while the discriminant criterion remains robust, the misclassification probabilities depend heavily on the underlying model.

Keywords and Phrases: Dependent but uncorrelated samples; Discrimination criterion; Generalized multivariate modified Bessel distribution; Hotelling's T^2-statistic; Non-central F-Bessel distribution; Probability of misclassification.

AMS Classification: Primary 62H30 and Secondary 62F03.

1 Introduction

A discrimination problem deals with classifying an individual into one of several populations on the basis of a number of measurements taken on the individual. Fisher [9] was the first to introduce the terminology. Other descriptive terms used for this type of problem include *separation*, *classification*, or *allocation*. There is a vast amount of literature on discrimination problems using various criteria, namely:

- the use of linear discriminant functions (see Fisher [8, 9]);
- the use of information theoretic procedures (see Kullback [19]);
- the use of predictive or Bayesian techniques (see Geisser [12] and Logan and Gupta [22]);
- the use of the likelihood ratio test criterion (see Anderson [1], Choi [6], and Gupta [15, 16]);
- the use of linear programming approaches (see Freed and Glover [11], Lam and Choo [21], and Ragsdale and Stam [26]);
- the use of the k nearest-neighbour method, which is based on finding the k nearest neighbours of the observation to be classified in the sample, and then using a majority vote for classifying observations (see Fix and Hodges, Jr. [10]);
- the use of the CART (classification and regression trees) method, which starts with the entire sample space and constructs split binary tree structures of the sample space into subsets (see Breiman et al. [5]).

For a detailed treatment of the topic, we refer the reader to the following monographs: Hand [17], Lachenbruch [20], and Van Ryzin [32]. Other multivariate texts that cover the topic include Anderson [1], Johnson and Wichern [18], Morrison [24], and Muirhead [25].

At this stage, it is important to realize that much of the literature on classification problems assumes that the data are normally distributed. However, we remark that Sutradhar [28] examined the behaviour of Fisher's linear discriminant function assuming a multivariate t distribution for the data. In this paper, we focus on Fisher's linear discriminant criterion and study the problem under the assumption that the populations originate from a *generalized multivariate modified Bessel* (GMMB) distribution. The GMMB distribution provides a good model for practical situations where data may be dependent but uncorrelated, or the data distribution may possess heavier tails than the normal. Furthermore, because of its rich parametric structure, the results under the GMMB model are expected to provide a good general solution to the problem since the model includes as special cases several multivariate distributions of practical interest, including both the multivariate normal and t distributions.

The remainder of the paper is as follows. In the next section, we provide the basic probability model underlying a GMMB data distribution, as well as the joint probability density function (pdf) for the samples under consideration in this paper. In Section 3, we present Fisher's discriminant function for classifying an observation into one of two GMMB populations. We consider the discrimination criterion as well as the misclassification problem when parameters of the GMMB model are both known and unknown. We also provide some numerical results to assess the probabilities of misclassification and to make comparisons with the standard multivariate normal model. We conclude the paper with a numerical example.

2 Models for the data and samples

In the remainder of the paper, we adopt the usual convention that random variables will be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letter. Thus, let the joint pdf of the p-dimensional random variable $\mathbf{Y} = (Y_1, \ldots, Y_p)'$ be given by

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{|\Sigma|^{-\frac{1}{2}} \left(\frac{\lambda}{\psi}\right)^{\frac{p}{4}}}{(2\pi)^{\frac{p}{2}} K_\nu\left(\sqrt{\lambda\psi}\right)} \left\{1 + \frac{1}{\psi}(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}^{\frac{2\nu-p}{4}} \\ &\quad \times K_{\frac{2\nu-p}{2}}\left(\sqrt{\lambda\psi\left[1 + \frac{1}{\psi}(\mathbf{y}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]}\right) \quad \text{for} \quad -\infty < \mathbf{y} < \infty, \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}$ is a $p \times 1$ location vector, Σ is a $p \times p$ positive definite scale matrix, and $K_\nu(z)$ is the modified Bessel function of the third kind of order ν (see Gradshteyn and Ryzhik [14], p. 970). The domain of the shape parameters (ψ, λ, ν) is:

$$\begin{aligned} &\psi > 0,\ \lambda \geq 0 \quad \text{for} \quad \nu < 0, \\ &\psi > 0,\ \lambda > 0 \quad \text{for} \quad \nu = 0, \\ &\psi \geq 0,\ \lambda > 0 \quad \text{for} \quad \nu > 0. \end{aligned} \tag{2}$$

Equation (1) is the pdf of the GMMB distribution and some of its statistical properties were studied by Thabane and Haq [29]. A matrix-variate extension of (1) was also considered by Thabane and Haq [31]. Moreover, this pdf is: (i) a member of the elliptically symmetric class of distributions (see Ng et al. [7]), and (ii) a special case of the symmetric multivariate hyperbolic distributions of Barndorff-Nielsen [3]. We will denote $\mathbf{Y}$ having pdf of the form (1) by

$$\mathbf{Y} \sim GMMB_p(\boldsymbol{\mu}, \Sigma, \psi, \lambda, \nu).$$

In addition to the multivariate normal and t distributions, Thabane and Haq [29] have also indicated several other special cases of the GMMB distribution, such as the multivariate Bessel and modified Bessel distributions (see Bhattacharya and Saxena [2] and Fang et al. [7]), the Pearson Type VII distribution (see Fang et al. [7]), and the Rao-type t distribution (see Rao [27]).

The basic probability model (1) can be extended to model the situation where one utilizes two samples in order to assign a new observation into one of two groups. Analogous to the approach of Sutradhar [28], we can express the joint pdf of two uncorrelated samples chosen from two different p-dimensional GMMB data distributions as follows. Suppose we let $\mathbf{X} = (\mathbf{Y}_1, \mathbf{Y}_2)$ represent the two samples of size n_1 and n_2, respectively. Therefore, $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{in_i})$ is a $p \times n_i$ matrix for $i = 1, 2$, where $\mathbf{Y}_{ij}$ is a p-dimensional random variable for $j = 1, 2, \ldots, n_i$. Letting $N = n_1 + n_2$, we assume that the joint pdf of the two samples is given by

$$\begin{aligned} & f(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \\ &= \frac{|\Sigma|^{-\frac{N}{2}} \left(\frac{\lambda}{\psi}\right)^{\frac{Np}{4}}}{(2\pi)^{\frac{Np}{2}} K_\nu\left(\sqrt{\lambda\psi}\right)} \left\{1 + \frac{1}{\psi} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)\right\}^{\frac{2\nu - Np}{4}} \\ &\times K_{\frac{2\nu - Np}{2}} \left(\sqrt{\lambda\psi \left[1 + \frac{1}{\psi} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)\right]}\right). \end{aligned} \tag{3}$$

By applying successive integrations, one can readily verify from (3) that $\mathbf{Y}_{ij} \sim GMMB_p(\boldsymbol{\mu}_i, \Sigma, \psi, \lambda, \nu)$ for each $i = 1, 2$ and $j = 1, 2, \ldots, n_i$. Moreover, it follows from (3) that the elements of $\mathbf{X}$, the combined sample of size N, are pairwise uncorrelated, but not necessarily independent.

3 Discrimination between two populations

3.1 Known parameters

In what follows, let π_1 and π_2 represent the two populations $GMMB_p(\boldsymbol{\mu}_1, \Sigma, \psi, \lambda, \nu)$ and $GMMB_p(\boldsymbol{\mu}_2, \Sigma, \psi, \lambda, \nu)$, respectively, where the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and Σ are assumed to be known. We also assume that an observation $\mathbf{y}$ is equally likely to have been drawn from either π_1 or π_2. Therefore, according to the classical rule for discrimination (see Anderson [1], pp. 199-202), we would assign $\mathbf{y}$ to π_1 if

$$p(\mathbf{y}|\boldsymbol{\mu}_1, \Sigma) \geq p(\mathbf{y}|\boldsymbol{\mu}_2, \Sigma). \tag{4}$$

However, Thabane and Haq [29] have shown that

$$p(\mathbf{y}|\boldsymbol{\mu}_i, \Sigma) = \int_0^\infty g(\mathbf{y}|\boldsymbol{\mu}_i, \Sigma, \tau) h(\tau) d\tau, \tag{5}$$

where

$$g(\mathbf{y}|\boldsymbol{\mu}_i, \Sigma, \tau) = \frac{|\Sigma|^{-\frac{1}{2}}}{(2\pi\tau)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2\tau}(\mathbf{y}-\boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}_i)\right\} \quad \text{for} \quad -\infty < \mathbf{y} < \infty \quad (6)$$

and

$$h(\tau) = \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{\nu}{2}}}{2K_\nu\left(\sqrt{\lambda\psi}\right)} \tau^{\nu-1} \exp\left\{-\frac{1}{2}\left(\frac{\psi}{\tau} + \lambda\tau\right)\right\} \quad \text{for} \quad \tau > 0. \quad (7)$$

We remark that (6) is the joint pdf of an $N_p(\boldsymbol{\mu}_i, \tau\Sigma)$ random variable, while (7) is the pdf of the generalized inverse Gaussian distribution (see Barndorff-Nielsen [3] and Barndorff-Nielsen et al. [4]). The domain of the parameters (ψ, λ, ν) in (7) is also given by (2). Let T denote the random variable having pdf (7).

Equation (5) implies that, conditional on $T = \tau > 0$, the observation $\mathbf{y}$ either comes from an $N_p(\boldsymbol{\mu}_1, \tau\Sigma)$ or an $N_p(\boldsymbol{\mu}_2, \tau\Sigma)$ population. According to the discrimination criterion for two known multivariate normal populations (see Anderson [1], pp. 204-205), (4) is equivalent to

$$\frac{1}{\tau}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}\mathbf{y} - \frac{1}{2\tau}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq 0,$$

from which it immediately follows that we assign $\mathbf{y}$ to π_1 if

$$w(\mathbf{y}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}\mathbf{y} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq 0. \quad (8)$$

We note that (8) is Fisher's linear discriminant rule, which therefore implies that the classical discrimination criterion remains robust under the GMMB model. This is consistent with Sutradhar's [28] findings in the multivariate t model. This result is not too surprising however, since, as remarked by Giri [13], most classical procedures studied under the normal model remain robust under the elliptical class of distributions. Nonetheless, while the discriminant rule remains robust under the t model, Sutradhar [28] did indicate that the probabilities of misclassification depended heavily on the degrees of freedom of the model. In the following subsection, we explore the behaviour of the rule through misclassification probabilities and compare the results with those obtained under the standard normal model.

3.2 Investigating misclassification probabilities

Let p_{ij} be the probability of classifying an observation $\mathbf{y}$ into π_j when it really belongs π_i for $i, j = 1, 2$. Therefore, it immediately follows that the misclassification probabilities are given by

$$p_{12} = Pr\{w(\mathbf{Y}) \leq 0 | \mathbf{Y} \in \pi_1\}$$

and

$$p_{21} = Pr\{w(\mathbf{Y}) \geq 0 | \mathbf{Y} \in \pi_2\},$$

where $w(\mathbf{Y})$ is defined by (8). In order to calculate p_{12} and p_{21}, we clearly require the distribution of the random variable $w(\mathbf{Y})$. Thabane and Haq [29] showed that if

$$\mathbf{Y} \sim GMMB_p\left(\boldsymbol{\mu}, \Sigma, \psi, \lambda, \nu\right),$$

then for a $p \times q$ matrix A of rank q $(q < p)$ and a constant vector $\mathbf{b}$ of dimension q, the linear combination

$$A^{'}\mathbf{Y} + \mathbf{b} \sim GMMB_q\left(A^{'}\boldsymbol{\mu} + \mathbf{b}, A^{'}\Sigma A, \psi, \lambda, \nu\right).$$

Applying this result with $A^{'} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{'}\Sigma^{-1}$ and $\mathbf{b} = -\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{'}\Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$, it can be easily shown that for $\mathbf{Y} \in \pi_1$,

$$w(\mathbf{Y}) \sim GMMB_1\left(\frac{\Delta^2}{2}, \Delta^2, \psi, \lambda, \nu\right)$$

where

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

is the coefficient of separation between the two populations (known as Mahalanobis' squared distance). In a similar fashion, one can also show that if $\mathbf{Y} \in \pi_2$, then

$$w(\mathbf{Y}) \sim GMMB_1\left(-\frac{\Delta^2}{2}, \Delta^2, \psi, \lambda, \nu\right).$$

Therefore, by straightforward algebra, it readily follows that

$$p_{12} = p_{21} = \int_{-\infty}^{-\Delta/2} \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{1}{4}}}{\sqrt{2\pi}K_{\nu}\left(\sqrt{\lambda\psi}\right)} \left\{1 + \frac{z^2}{\psi}\right\}^{\frac{2\nu-1}{4}} K_{\frac{2\nu-1}{2}}\left(\sqrt{\lambda\psi\left[1 + \frac{z^2}{\psi}\right]}\right) dz. \qquad (9)$$

Table 1 displays the probabilities of misclassification (to 4 decimal places) for selected values of Δ^2 under several different variants of the GMMB model, namely:

1. $\psi = r, \lambda = 0, \nu = -r/2,\ r \to \infty$ (*Normal distribution*)
2. $\psi = 2, \lambda = 0, \nu = -2$ (*Rao-type t distribution*)
3. $\psi = 5, \lambda = 0, \nu = -3.5$ (*Rao-type t distribution*)
4. $\psi = 1/2, \lambda = 0, \nu = -1.25$ (*Pearson Type VII distribution*)
5. $\psi = 0, \lambda = 1, \nu = 0.5$ (*Bessel distribution*)

6. $\psi = 0, \lambda = 4, \nu = 2$ (*Bessel distribution*)

7. $\psi = 0.5, \lambda = 2, \nu = 0.5$ (*Modified Bessel distribution*)

There are two important observations one can make concerning the results in Table 1. First of all, the calculation of misclassification probabilities under the normal model generally leads to overestimation if the populations are not well separated (i.e. $\Delta^2 \leq 9$). Secondly, if the populations are well separated (i.e. $\Delta^2 > 9$), the normality assumption leads to underestimation of the misclassification probabilities. This is not surprising, however, given the heavier-tailed behaviour of the GMMB models under consideration in Table 1. Moreover, it is worth noting that these observations are consistent with Sutradhar's findings under the multivariate t model (see [28], p. 831).

Table 1: Probabilities of Misclassification

	(ψ, λ, ν)						
Δ^2	Normal	(2,0,-2)	(5,0,-3.5)	(0.5,0,-1.25)	(0,1,0.5)	(0,4,2)	(0.5,2,0.5)
0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
1	0.3085	0.2593	0.2864	0.1797	0.2049	0.2759	0.2681
4	0.1587	0.1151	0.1377	0.0645	0.1045	0.1353	0.1283
9	0.0668	0.0506	0.0596	0.0287	0.0561	0.0622	0.0598
16	0.0228	0.0237	0.0249	0.0152	0.0309	0.0275	0.0277
25	0.0062	0.0121	0.0106	0.0090	0.0173	0.0118	0.0129
36	0.0014	0.0066	0.0047	0.0059	0.0098	0.0050	0.0060
49	0.0002	0.0039	0.0022	0.0040	0.0056	0.0021	0.0028

3.3 Unknown Parameters

This is perhaps the most realistic case that commonly arises in many situations. Let $\mathbf{Y}_1 = (\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1})$ and $\mathbf{Y}_2 = (\mathbf{Y}_{21}, \ldots, \mathbf{Y}_{2n_2})$ represent two samples of sizes n_1 and n_2, respectively, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have joint pdf given by (3). Therefore, $\mathbf{Y}_{ij} \sim GMMB_p(\boldsymbol{\mu}_i, \Sigma, \psi, \lambda, \nu)$ for each $i = 1, 2$ and $j = 1, 2, \ldots, n_i$. In what follows, we assume that the location vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, as well as the common scale matrix Σ, are unknown. However, we assume that the shape parameters ψ, λ, and ν are known. The situation where all (or even a subset) of these shape parameters are unknown is a much more complicated problem, and beyond the scope of this paper.

In order to discriminate between the two GMMB populations in a meaningful fashion, we should first assess whether they are significantly separated by testing the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against the alternative hypothesis $H_a : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. To aid in this regard, we define the following random variables:

$$\bar{\mathbf{Y}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{Y}_{ij}, \quad i = 1, 2,$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i\right) \left(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i\right)', \quad i = 1, 2,$$

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2},$$
$$T_d^2 = \frac{n_1 n_2}{n_1+n_2}\left(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2\right)' S_p^{-1}\left(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2\right),$$
$$F_d = \frac{n_1+n_2-p-1}{(n_1+n_2-2)p} T_d^2.$$

We remark that S_p is the pooled estimate of the unknown scale matrix Σ based on the two samples, and T_d^2 is the classical Hotelling's T^2-statistic.

The above hypothesis testing problem was considered recently by Thabane and Drekic [30], who showed that the test statistic F_d has a *non-central F-Bessel distribution* with parameters $(p, n_1+n_2-p-1, \delta_d^2, \psi, \lambda, \nu)$ where $\delta_d^2 = n_1 n_2 \Delta^2/(n_1+n_2)$ (we refer the reader to Thabane and Drekic [30], pp. 368-369, for the complete details of this result). We also remark that under the null hypothesis H_0, $\delta_d^2 = 0$ and the distribution of F_d simplifies to the central F distribution with p and n_1+n_2-p-1 degrees of freedom, a result that is consistent with the null distribution obtained under the multivariate normal model (see Muirhead [25], p. 216) and the multivariate t model (see Sutradhar [28]).

If the populations are found to be well separated after applying the above testing procedure, then a suitable discrimination criterion can be formulated to classify an observation $\mathbf{y}$ as originating from either a $GMMB_p(\boldsymbol{\mu}_1, \Sigma, \psi, \lambda, \nu)$ population or a $GMMB_p(\boldsymbol{\mu}_2, \Sigma, \psi, \lambda, \nu)$ population. In particular, one natural approach is to use the sample discriminant function between the two populations given by

$$d(\mathbf{y}) = \left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)' s_p^{-1}\mathbf{y} - \frac{1}{2}\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)' s_p^{-1}\left(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2\right), \tag{10}$$

and to classify $\mathbf{y}$ as a $GMMB_p(\boldsymbol{\mu}_1, \Sigma, \psi, \lambda, \nu)$ observation if $d(\mathbf{y}) \geq 0$. As noted by Sutradhar [28], however, the distribution of (10) is quite complicated to deal with. Nonetheless, one could estimate the probabilities of misclassification using

$$\hat{p}_{12} = \hat{p}_{21} = \int_{-\infty}^{-\hat{\Delta}/2} \frac{\left(\frac{\lambda}{\psi}\right)^{\frac{1}{4}}}{\sqrt{2\pi} K_\nu\left(\sqrt{\lambda\psi}\right)} \left\{1 + \frac{z^2}{\psi}\right\}^{\frac{2\nu-1}{4}} K_{\frac{2\nu-1}{2}}\left(\sqrt{\lambda\psi\left[1+\frac{z^2}{\psi}\right]}\right) dz, \tag{11}$$

where $\hat{\Delta}^2 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' s_p^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$.

3.4 A Numerical Example

To illustrate the procedure outlined in the previous subsection, we consider the flea beetle data analyzed by Lubischew [23] and subsequently by Sutradhar [28]. In particular, Sutradhar considered the two variables

$$Y_1 \equiv \text{the length of elytra (measured in 0.01 mm)}$$

and

$$Y_2 \equiv \text{the length of the second antennal joint (measured in microns)}$$

of two species of the flea beetle, namely $\pi_1 \equiv$ Haltica oleracea and $\pi_2 \equiv$ Haltica carduorum. We refer the reader to the data in Table 2 of Lubischew [23], p. 461, which details the $n_1 = 19$ and $n_2 = 20$ measurements sampled under π_1 and π_2, respectively.

Sutradhar [28] proceeded to provide evidence which indicated that these data were: (i) non-normal, and (ii) dependent but uncorrelated. Consequently, Sutradhar proposed that the 19 observations under π_1 and 20 observations under π_2 were realizations from a multivariate t distribution, which is contained as a special case of the more general GMMB model given by (3). Moreover, Sutradhar applied the same hypothesis testing procedure described above and found that π_1 and π_2 are well separated. We omit the details here, but refer the reader to Sutradhar [28], p. 833, for complete details.

For comparative purposes, we consider some other GMMB models as possible alternatives to Sutradhar's multivariate t model. In particular, Table 2 displays the estimated probabilities of misclassification (computed to 4 decimal places via (11)) for several different choices of (ψ, λ, ν) of the GMMB model (see Section 3.2 for a further description of these models). In calculating the probabilities, we used the raw data in Lubischew [23], Table 2, to re-calculate $\hat{\Delta}^2$, as there appears to be a typographical error with the value of $\hat{\Delta}^2$ reported in Sutradhar [28], p. 834. Specifically, we calculated $\hat{\Delta}^2 = 3.3696$. It is worth remarking that for $\hat{\Delta}^2 = 3.3696$, we found that $\hat{p}_{12} = \hat{p}_{21} = 0.1794$. Once again, these findings are consistent with those of Sutradhar [28], who suggested that the use of normal-based misclassification probabilities generally leads to overestimation.

Table 2: Estimated Probabilities of Misclassification

(ψ, λ, ν)	$(0.5, 0, -1.25)$	$(0, 1, 0.5)$	$(0.5, 2, 0.5)$	$(0, 4, 2)$
Probability	0.0752	0.1162	0.1453	0.1530

Acknowledgements:

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

[1] T. W. Anderson, "An Introduction to Multivariate Statistical Analysis", 2nd Edition, Wiley, New York, 1984.

[2] S. K. Bhattacharya, A. K. Saxena, Multivariate modified Bessel distribution, *Estadistica* **41** (1989) 1-11.

[3] O. Barndorff-Nielsen, Hyperbolic distributions and distributions on hyperbolae, *Scandinavian Journal of Statistics* **5** (1978) 151-157.

[4] O. Barndorff-Nielsen, P. Blaesild, C. Halgreen, First hitting time models for the generalized inverse Gaussian distribution, *Stochastic Processes and Their Applications* **7** (1978) 49-54.

[5] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone. "Classification and Regression Trees", Wadsworth International Group, California, 1984.

[6] S. C. Choi, Classification of multiple observed data, *Biometrische Zeitschrift* **14** (1972) 8-11.

[7] K. Fang, S. Kotz, K. Ng, "Symmetric Multivariate and Related Distributions", Chapman & Hall, New York, 1990.

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

[9] R. A. Fisher, The staistical utilization of multiple measurements, *Annals of Eugenics* **8** (1938) 376-386.

[10] E. Fix, J. L. Hodges, Jr., Discriminatory analysis, nonparametric discrimination: Consistency properties, Technical Report 4, Project Number 21-49-004, 1951, pp. 261-279, USAF School of Aviation Medicine, Randolph Field, Texas.

[11] N. Freed, F. Glover, A linear programming approach to the discriminant problem, *Decision Sciences* **12** (1981) 68-74.

[12] S. Geisser, Posterior odds for multivariate normal classification, *Journal of the Royal Statistical Society: Series B* **26** (1964) 69-76.

[13] N. C. Giri, "Multivariate Statistical Analysis", Marcel Dekker, New York, 1996.

[14] I. S. Gradshteyn, I. W. Ryzhik, "Tables of Integrals, Series and Products", Academic Press, New York, 1980.

[15] A. K. Gupta, On a multivariate classification model, in "Multivariate Statistical Analysis" (R. P. Gupta, Editor), North Holland, 83-93.

[16] A. K. Gupta, On a classification rule for multiple measurements, *Computers and Mathematics with Applications* **12** (1986) 301-308.

[17] D. J. Hand, "Discrimination and Classification", Wiley, New York, 1981.

[18] R. A. Johnson, D. W. Wichern, "Applied Multivariate Statistical Analysis", Prentice Hall, New York, 1992.

[19] S. Kullback, "Information Theory and Statistics", Wiley, New York, 1959.

[20] P. A. Lachenbruch, "Discriminant Analysis", Hafner Press, New York, 1975.

[21] K. F. Lam, E. U. Choo, A linear goal programming model for classification with non-monotonic attributes, *Journal of Computers and Operations Research* **20** (1993) 403-408.

[22] T. P. Logan, A. K. Gupta, Bayesian discrimination using multiple observations, *Communications in Statistics - Theory and Methods* **22** (1993) 1735-1754.

[23] A. A. Lubischew, On the use of discriminant functions in taxonomy, *Biometrics* **18** (1962) 455-477.

[24] D. F. Morrison, "Multivariate Statistical Methods", McGraw-Hill, New York, 1967.

[25] R. J. Muirhead, "Aspects of Multivariate Statistical Theory", Wiley, New York, 1982.

[26] C. T. Ragsdale, A. Stam, Mathematical programming formulations for the discriminant problem: An old dog does new tricks, *Decision Sciences* **22** (1991) 296-306.

[27] C. R. Rao, "Linear Statistical Inference and Its Applications" Wiley, New York, 1973.

[28] B. C. Sutradhar, Discrimination of observations into one of two t populations, *Biometrics* **46** (1990) 827-835.

[29] L. Thabane, M. S. Haq, The generalized multivariate modified Bessel distribution and its Bayesian applications, *Journal of Statistical Sciences* **11** (2003) 255-267.

[30] L. Thabane, S. Drekic, Hypothesis testing for the generalized multivariate modified Bessel model, *Journal of Multivariate Analysis* **86** (2003) 360-374.

[31] L. Thabane, M. S. Haq, On the generalized matrix-variate hyperbolic distribution and its application to the Bayesian analysis of the normal linear model, to appear in *Statistics*.

[32] J. Van Ryzin, "Classification and Clustering", Academic Press, New York, 1977.

BLANK PAGE