Journal of Applied Statistics
Vol. 38, No. 5, May 2011, 1021–1032

Taylor & Francis
Taylor & Francis Group

Classification with discrete and continuous variables via general mixed-data models

A.R. de Leon[a]*, A. Soo[b] and T. Williamson[b]

[a]*Department of Mathematics & Statistics, University of Calgary, Calgary, AB, Canada;* [b]*Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

(*Received 24 May 2009; final version received 16 February 2010*)

We study the problem of classifying an individual into one of several populations based on mixed nominal, continuous, and ordinal data. Specifically, we obtain a classification procedure as an extension to the so-called location linear discriminant function, by specifying a general mixed-data model for the joint distribution of the mixed discrete and continuous variables. We outline methods for estimating misclassification error rates. Results of simulations of the performance of proposed classification rules in various settings vis-à-vis a robust mixed-data discrimination method are reported as well. We give an example utilizing data on croup in children.

Keywords: error rate; general location model; grouped continuous model; maximum likelihood; measurement level; minimum distance probability; misclassification probability; plug-in estimates

1. Introduction

Many diagnostic studies in medicine entail the collection of data consisting of a mixture of discrete and continuous variables that are then used to classify patients into diagnostic or prognostic groups (e.g. healthy and sick, malignant and benign, bad and good prognosis). Consider, for example, a study involving child patients suffering from symptoms of croup (laryngotracheobronchitis), a common cause of upper airway obstruction in children and has an annual incidence of 6% in children under the age of 6 years in Alberta, Canada. Clinical practice guidelines for managing such patients presented at hospital emergency departments instruct health practitioners to treat them in the manner suggested by Bjornson *et al.* [1] and inform attending physicians' decisions on whether to admit them for hospitalization or treat them as outpatients. Such decisions are based on common characteristics used in the evaluation of croup severity which include continuous (e.g. temperature, respiratory rate) as well as nominal (e.g. presence/absence of certain symptoms) and ordinal measurements (e.g. croup score). Primary focus of the study is to evaluate the effectiveness

*Corresponding author. Email: adeleon@math.ucalgary.ca

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664761003758976
http://www.informaworld.com

of the guidelines in separating those croup patients in need of hospitalization from those who can be sent home.

An inefficient approach to handling such data is to convert one type of variable to another. If, for example, nominal variables can be subjected to some scoring scheme, then all the discrete variables can be treated as ordinal; alternatively, the ordinal scale may be treated as an interval, and the ordinal variables can then be considered continuous. Another option treats all discrete variables as nominal by categorizing ordinal variables through some grouping criteria. The general location model (GLOM) [18,21] and conditional grouped continuous model (CGCM) [20] have been used in practice, with the former treating ordinal variables as nominal (or continuous) and the latter treating nominal variables as ordinal. GLOMs have received much attention in the literature in the context of classification and discrimination [13]. Bar-Hen and Daudin [1] provide a survey of its adaptations in mixed-data discrimination via distances, and outline various approaches to variable selection. Mahat *et al.* [15] propose nonparametric smoothing of GLOM as a basis for discriminant analysis. Lee *et al.* [14] recently applied CGCM in discriminant analysis involving mixed continuous and ordered categorical data.

These options involve some element of subjectivity, with possible loss of information, and do not appear very satisfactory in general. CGCMs introduce considerable subjectivity in the adopted numerical scoring scheme and GLOMs either result in information loss when ordinal variables are categorized [2] or give rise to 'robustness' concerns when ordinal variables are treated as continuous [19]. Both models fail to make full and correct use of the information contained in the data, which may lead to potentially incorrect inferences. More importantly, they do not provide a mechanism for explicitly incorporating correlations between nominal and ordinal variables, and are thus unable to distinguish correlations between nominal and continuous variables from correlations between ordinal and continuous ones.

Another alternative is to use distance-based methods [6]. Unlike other model-based methods, this approach is able to handle different types of data including those with mixtures of discrete and continuous variables. Methods based on this approach rely on various distance functions between individual observations, and thus the choice of distance function is a crucial consideration. Recent references on this approach include Nuñez *et al.* [17] and Villaroya *et al.* [23]; see also Cuadras *et al.* [7].

The purpose of this paper is to develop classification procedures for multivariate data with mixtures of nominal, ordinal and continuous variables using so-called general mixed-data models (GMDMs) [9]. GMDMs are motivated by the need to account for various levels of measurement, and hence different types of information, in mixed data which many conventional approaches fail to incorporate in the analysis. GMDMs address the shortcomings of, and include as special cases, GLOM and CGCM. They can serve as platform for generalizing conventional multivariate methods to various mixed data settings [8].

The paper is organized as follows. Section 2 reviews GMDMs. Section 3 derives classification rules as generalizations of GLOM-based location linear discriminant functions (LLDF) [12]. Error rates for the proposed rules are investigated and methods for estimating them are outlined. The performance of the classification procedures in simulations is also reported in Section 3. The methodology is illustrated in Section 4 using mixed data from a study on croup among children. A brief summary concludes the paper in Section 5.

2. General mixed-data models

GMDMs are general models for the joint distribution of mixed discrete and continuous variables that extend and unify existing models into one general class flexible enough to accommodate various types of mixed data simultaneously. Let the mixed data consist of a vector $\mathbf{x} = (X_1, \ldots, X_S)^\top$ of binary variables with $\sum_s X_s = 1$, a vector $\mathbf{y} = (Y_1, \ldots, Y_C)^\top$ of continuous variables, and a

vector $\mathbf{z} = (Z_1, \ldots, Z_Q)^\top$ of ordinal variables with Z_q having $L_q + 1$ levels. The binary vector $\mathbf{x}$ represents nominal data from a contingency table with $S = \prod_d s_d$ nominal states (or cells) defined by each possible value of a vector $\mathbf{u} = (U_1, \ldots, U_D)^\top$ of nominal variables, each with s_d possible categories. In this case, we have $\mathbf{x} = \mathbf{x}_{(s)}$ (i.e. $X_s = 1$ and $X_{s'} = 0$ for all $s' \neq s$) if $\mathbf{u}$ falls in state $s = 1, \ldots, S$.

For the vector $\mathbf{z}$, an underlying continuous latent vector $\mathbf{y}^* = (Y_1^*, \ldots, Y_Q^*)^\top$ is assumed linked to $\mathbf{z}$ by the threshold relationship $Z_q = \ell_q \Leftrightarrow \alpha_q^{\ell_q - 1} < Y_q^* \leq \alpha_q^{\ell_q}$, where $\{\alpha_q^0 = -\infty, \alpha_q^1, \ldots, \alpha_q^{L_q}, \alpha_q^{L_q+1} = +\infty\}$ are unknown cutpoints, $\ell_q = 1, \ldots, L_q + 1$ are ordinal scores for Z_q, and $\text{var}(Y_1^*) = \cdots = \text{var}(Y_Q^*) = 1$.

Denoting marginal/joint and conditional densities by $[\cdot]$ and $[\cdot|\cdot]$, respectively, GMDM models $[\mathbf{x}]$ as multinomial and $[\mathbf{y}, \mathbf{y}^*|\mathbf{x}]$ as multivariate normal whose mean depends on $\mathbf{x}$ but whose covariance matrix is constant across states, so that $[\mathbf{x}, \mathbf{y}, \mathbf{y}^*]$ is GLOM. de Leon and Carrière [9] show that $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ can be written as

$$[\mathbf{x} = \mathbf{x}_{(s)}, \mathbf{y}, \mathbf{z} = \ell] = \pi_s \phi_C(\mathbf{y} - \boldsymbol{\mu}_s | \boldsymbol{\Sigma}) \int_{\mathcal{S}(s,\mathbf{y},\ell)} \phi_Q(\mathbf{v}|\mathbf{R}) \mathrm{d}\mathbf{v}, \tag{1}$$

for $s = 1, \ldots, S$, with $\mathcal{S}(s, \mathbf{y}, \ell) = (\nu_{s1}^{\ell_1 - 1}, \nu_{s1}^{\ell_1}] \times \cdots \times (\nu_{sQ}^{\ell_Q - 1}, \nu_{sQ}^{\ell_Q}]$, where $\nu_{sq}^{\ell_q} = \gamma_q^{\ell_q} - \tau_{sq} - \boldsymbol{\beta}_q^\top \mathbf{y}$ and $\phi_K(\cdot|\mathbf{H})$ is a K-dimensional normal density with mean $\mathbf{0}$ and covariance matrix $\mathbf{H}$. We say $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are jointly distributed according to GMDM if their joint density $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ is given by Equation (1). The symmetric matrix $\mathbf{R}$ contains (conditional) polychoric correlations $r_{qq'}$ of $\mathbf{z}$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_S)^\top$ contains multinomial state probabilities ($\sum_s \pi_s = 1$), $\boldsymbol{\mu}_s$ is the mean vector of $\mathbf{y}$ for state s, $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{y}$, and $\ell = (\ell_1, \ldots, \ell_Q)^\top$ is the vector of ordinal scores. Note the following:

(i) standardized cutpoints $\gamma_q^{\ell_q}$ account for ordinal information in $\mathbf{z}$;
(ii) state effects τ_{sq} induce associations between $\mathbf{x}$ and $\mathbf{z}$;
(iii) regression effects $\boldsymbol{\beta}_q$ represent polyserial correlations between $\mathbf{y}$ and $\mathbf{z}$.

For model identifiability, state S is arbitrarily designated as the reference state. Parameters can then be represented by $\boldsymbol{\Theta}$, the stacked vector of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, where $\boldsymbol{\Theta}_1$ is the vector containing 'location' parameters $\boldsymbol{\pi}$, $\{\boldsymbol{\mu}_s : s = 1, \ldots, S\}$, $\{\gamma_q^{\ell_q} : q = 1, \ldots, Q; \ell_q = 1, \ldots, L_q\}$, and $\{\tau_{sq} : s = 1, \ldots, S - 1; q = 1, \ldots, Q\}$, and $\boldsymbol{\Theta}_2$ is the vector containing the rest of the parameters. A restricted GMDM may be defined by imposing restrictions on $\boldsymbol{\Theta}$ to reduce its dimension and streamline its structure [9]. Note that we put 'location' in quotes, as $\boldsymbol{\Theta}_1$ contains the state effects τ_{sq}, which measure associations between ordinal and nominal data.

Several models are obtained as special cases of GMDM. If $Q = 0$ (i.e. no ordinal variables), then GMDM specializes to GLOM. Similarly, GMDM reduces to CGCM when $S = 1$ (i.e. no nominal variables), in which $\mathbf{y}$ is multivariate normal and $\mathbf{z}$ depends on $\mathbf{y}$ via a multivariate probit model. Grouped continuous models [9] for ordinal data are obtained by taking $C = 0$ and $S = 1$.

The choice of $[\mathbf{y}, \mathbf{y}^*|\mathbf{x}]$ is completely arbitrary; however, modeling it by the multivariate normal distribution with constant covariance matrix across the states, as in GMDM, is convenient because of the normal distribution's nice marginal and conditional distributions. While normality may not hold in many cases, it can be easily checked in practice; in addition, transformations are readily available for normalizing non-normal data [11]. Non-normal latent distributions may also be considered, however, estimates tend to be robust with respect to the latent distribution even if the latter is skewed [22].

2.1 *Estimation*

For simplicity, consider two distinct populations $\mathbf{\Pi}^{(1)}$ and $\mathbf{\Pi}^{(2)}$ defined by GMDMs with respective parameters $\mathbf{\Theta}^{(1)}$ and $\mathbf{\Theta}^{(2)}$; extension of the methods below to the case of more than two populations is straightforward. Following de Leon and Carrière [8], we assume $\mathbf{\Theta}_1^{(1)} \neq \mathbf{\Theta}_1^{(2)}$ and $\mathbf{\Theta}_2^{(1)} = \mathbf{\Theta}_2^{(2)}$, that is, $\mathbf{\Pi}^{(1)}$ and $\mathbf{\Pi}^{(2)}$ differ only in their 'locations'. Since $\mathbf{\Theta}_2^{(1)} = \mathbf{\Theta}_2^{(2)}$ contains τ_{sq}, note that this is not exactly the mixed-data analog of complete homogeneity. It is possible to consider more flexible GMDMs in this case by allowing state effects τ_{sq} to vary across groups, a formulation analogous to the homogeneous case in normal data classification. However, relaxation of the assumption of equal state effects across groups, while rendering the methodology more appealing in practice, gives rise to identifiability issues [8,9].

Given independent samples from $\mathbf{\Pi}^{(1)}$ and $\mathbf{\Pi}^{(2)}$, the log-likelihood of the combined data is then given by

$$\ell(\mathbf{\Theta}^{(1)}, \mathbf{\Theta}^{(2)}) = \ell^{(1)}(\mathbf{\Theta}^{(1)}) + \ell^{(2)}(\mathbf{\Theta}^{(2)}), \tag{2}$$

which can be maximized simultaneously or separately for $\mathbf{\Theta}^{(1)}$ and $\mathbf{\Theta}^{(2)}$. The former – a more computationally intensive route – yields maximum likelihood estimates (MLEs) and results in a single estimate of $\mathbf{\Theta}_2^{(1)} = \mathbf{\Theta}_2^{(2)}$, while the latter – an application of inference functions or estimating equations [10] – provides two sets of estimates $\hat{\mathbf{\Theta}}_2^{(1)}$ and $\hat{\mathbf{\Theta}}_2^{(2)}$, which are then averaged. In either case, the estimates of the nominal and continuous data parameters are the usual MLEs for multinomial and multivariate normal samples given by $\bar{\mathbf{y}}_s^{(g)}$, the sample mean for state s in $\mathbf{\Pi}^{(g)}$, for $\boldsymbol{\mu}_s^{(g)}$; $\mathbf{S}_p$, the sample covariance matrix pooled across states and populations, for $\mathbf{\Sigma}$; and $\hat{\pi}_s^{(g)}$, the sample proportion for state s in $\mathbf{\Pi}^{(g)}$, for $\pi_s^{(g)}$. These work well when the sample sizes $N^{(1)}$ and $N^{(2)}$ are both appreciably larger than the total number SL of state-level combinations, with $L = \prod_q (L_q + 1)$. When this is not the case, a few states or levels may be collapsed to reduce the number of parameters. Alternatively, restrictions may be imposed on the model as in [9], and the original parameters are then expressed in terms of the restricted model parameters. Nonparametric smoothing methods as applied by Mahat *et al.* [15] to GLOM may also be adapted to GMDM.

Estimates obtained by separate maximization of $\ell^{(1)}(\mathbf{\Theta}^{(1)})$ and $\ell^{(2)}(\mathbf{\Theta}^{(2)})$ are, like MLEs, also consistent and have asymptotic normal distributions, albeit different from those of MLEs. Details, including numerical implementation of the methods, are found in [9].

3. Classification rules

Let $\alpha^{(g)}$ be the prior probability of $\mathbf{\Pi}^{(g)}$ and define $c(g|g')$ as the cost of misclassifying a member of $\mathbf{\Pi}^{(g')}$ as a member of $\mathbf{\Pi}^{(g)}$, $g \neq g'$. The optimum classification rule in the case of two populations [11, p. 581] assigns a mixed observation $\mathbf{w} \equiv \{\mathbf{x} = \mathbf{x}_{(s)}, \mathbf{y}, \mathbf{z} = \ell\}$ to $\mathbf{\Pi}^{(1)}$ if $[\mathbf{x}_{(s)}, \mathbf{y}, \ell]^{(1)}/[\mathbf{x}_{(s)}, \mathbf{y}, \ell]^{(2)} \geq (c(1|2)\alpha^{(2)})/(c(2|1)\alpha^{(1)})$, and to $\mathbf{\Pi}^{(2)}$ otherwise, where $[\cdot]^{(g)}$ is the density of $\mathbf{\Pi}^{(g)}$. Assuming, without loss of generality, that $\alpha^{(1)} = \alpha^{(2)} = 1/2$ and $c(1|2) = c(2|1)$, and using Equation (1), the optimum rule allocates $\mathbf{w}$ to $\mathbf{\Pi}^{(1)}$ if

$$\mathcal{R}_{s\ell} : (\boldsymbol{\mu}_s^{(1)} - \boldsymbol{\mu}_s^{(2)})^\top \mathbf{\Sigma}^{-1} \left(\mathbf{y} - \frac{\boldsymbol{\mu}_s^{(1)} + \boldsymbol{\mu}_s^{(2)}}{2} \right) + \log \left(\frac{p^{(1)}(\ell|\mathbf{x}_{(s)}, \mathbf{y})}{p^{(2)}(\ell|\mathbf{x}_{(s)}, \mathbf{y})} \right) \geq \log \left(\frac{\pi_s^{(2)}}{\pi_s^{(1)}} \right), \tag{3}$$

where $p^{(g)}(\ell|\mathbf{x}_{(s)}, \mathbf{y}) = \int_{\mathcal{S}^{(g)}(s,\mathbf{y},\ell)} \phi_Q(\mathbf{v}|\mathbf{R}) \mathrm{d}\mathbf{v}$, and $\mathcal{S}^{(g)}(s, \mathbf{y}, \ell)$ is as defined in Section 2, with $\nu_{sq}^{\ell_q,(g)} = \gamma_q^{\ell_q,(g)} - \tau_{sq}^{(g)} - \boldsymbol{\beta}_q^\top \mathbf{y}$. That is, we assign $\mathbf{w}$ to $\mathbf{\Pi}^{(1)}$ if $\mathbf{w} \in \mathcal{R}_{s\ell}$, and to $\mathbf{\Pi}^{(2)}$ if $\mathbf{w} \in \mathcal{R}_{s\ell}^c$. Note that Equation (3) can be expressed in terms of de Leon and Carrière's [9] generalized Mahalanobis distance.

Classification rule (3) provides discriminant functions for each of the SL state-level pairs. In the special case of only one ordinal variable with two levels (i.e. $Q = L = 1$), the conditional probabilities simplify to $p^{(g)}(1|\mathbf{x}_{(s)}, \mathbf{y}) = \Phi(\gamma^{(g)} - \tau_s^{(g)} - \boldsymbol{\beta}^\top \mathbf{y})$ and $p^{(g)}(2|\mathbf{x}_{(s)}, \mathbf{y}) = 1 - p^{(g)}(1|\mathbf{x}_{(s)}, \mathbf{y})$, where $\Phi(\cdot)$ is the cumulative standard normal distribution function. Observe that if there is no ordinal variable (i.e. $Q = 0$), Equation (3) reduces to Krzanowski's LLDF [12] based on GLOM. Moreover, if $S = 2$, then we get the double discriminant function of Chang and Afifi [4].

In practice, the parameters $\boldsymbol{\Theta}^{(1)}$ and $\boldsymbol{\Theta}^{(2)}$ are usually unknown and need to be estimated. Given independent samples of sizes $N^{(1)}$ and $N^{(2)}$ from $\boldsymbol{\Pi}^{(1)}$ and $\boldsymbol{\Pi}^{(2)}$, respectively, we estimate $\mathcal{R}_{s\ell}$ by simply replacing the parameters with their corresponding estimates as follows:

$$\hat{\mathcal{R}}_{s\ell} : (\bar{\mathbf{y}}_s^{(1)} - \bar{\mathbf{y}}_s^{(2)})^\top \mathbf{S}_p^{-1} \left(\mathbf{y} - \frac{\bar{\mathbf{y}}_s^{(1)} + \bar{\mathbf{y}}_s^{(2)}}{2} \right) + \log \left(\frac{\hat{p}^{(1)}(\ell|\mathbf{x}_{(s)}, \mathbf{y})}{\hat{p}^{(2)}(\ell|\mathbf{x}_{(s)}, \mathbf{y})} \right) \geq \log \left(\frac{\hat{\pi}_s^{(2)}}{\hat{\pi}_s^{(1)}} \right), \quad (4)$$

where $\hat{p}^{(g)}(\ell|\mathbf{x}_{(s)}, \mathbf{y}) = \int_{\hat{\mathcal{S}}^{(g)}(s,\mathbf{y},\ell)} \phi_Q(\mathbf{v}|\hat{\mathbf{R}})\mathrm{d}\mathbf{v}$. Sample classification rule (4) then allocates an observation $\mathbf{w}$ to $\boldsymbol{\Pi}^{(1)}$ if $\mathbf{w} \in \hat{\mathcal{R}}_{s\ell}$, and to $\boldsymbol{\Pi}^{(2)}$ if $\mathbf{w} \in \hat{\mathcal{R}}_{s\ell}^c$. While we used the estimates described in Section 2.1 in Equation (4), any set of consistent estimates can in fact be used to estimate the classification rules.

The above approach easily extends to the case of several groups. Given populations $\boldsymbol{\Pi}^{(1)}, \ldots, \boldsymbol{\Pi}^{(G)}$, $G \geq 2$, defined by GMDMs such that $\boldsymbol{\Theta}_1^{(g)} \neq \boldsymbol{\Theta}_1^{(g)}$ $\forall g' \neq g$ and $\boldsymbol{\Theta}_2^{(1)} = \cdots = \boldsymbol{\Theta}_2^{(G)}$, and assuming equal misclassification costs, the optimum rule [11, p. 611] classifies a mixed observation $\mathbf{w} \equiv \{\mathbf{x} = \mathbf{x}_{(s)}, \mathbf{y}, \mathbf{z} = \ell\}$ as belonging to $\boldsymbol{\Pi}^{(g^*)}$ if $\delta_{s\ell}^{(g^*)}(\mathbf{w}) = \max_g \delta_{s\ell}^{(g)}(\mathbf{w})$, where

$$\delta_{s\ell}^{(g)}(\mathbf{w}) = (\boldsymbol{\mu}_s^{(g)})^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} (\boldsymbol{\mu}_s^{(g)})^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_s^{(g)} + \log \pi_s^{(g)} + \log p^{(g)}(\ell|\mathbf{w}) + \log \alpha^{(g)},$$

for $g = 1, \ldots, G$. Observe that the linear discriminant scores in normal data classification are a special case of $\delta_{s\ell}^{(1)}, \ldots, \delta_{s\ell}^{(G)}$. The latter reduces to the former for $S = 1$ and $Q = 0$. In practice, estimates $\hat{\delta}_{s\ell}^{(1)}(\mathbf{w}), \ldots, \hat{\delta}_{s\ell}^{(G)}(\mathbf{w})$ are obtained by plug-in method using any set of consistent estimates (e.g. MLEs) of the parameters.

3.1 *Misclassification probabilities*

For convenience, let $G = 2$. With $\alpha^{(1)} = \alpha^{(2)} = 1/2$ and $c(1|2) = c(2|1)$, misclassification probabilities $P(g|g') = P(\mathbf{w}$ is allocated to $\boldsymbol{\Pi}^{(g)}|\mathbf{w} \in \boldsymbol{\Pi}^{(g')})$, $g \neq g'$, may be derived from Equations (1) and (4) as follows:

$$P(g|g') = \sum_{s,\ell} \pi_s^{(g')} \left(\int_{\hat{\mathcal{R}}_{s\ell}^c} I(g' = 1) + \int_{\hat{\mathcal{R}}_{s\ell}} I(g' = 2) \right) p^{(g')}(\ell|\mathbf{x}_{(s)}, \mathbf{y}) \phi_C(\mathbf{y} - \boldsymbol{\mu}_s^{(g')}|\boldsymbol{\Sigma})\mathrm{d}\mathbf{y}, \quad (5)$$

where $I(\cdot)$ is the indicator function. In the special case of $Q = L = 1$, Equation (5) simplifies to

$$P(g|g') = \sum_s \pi_s^{(g')} \left\{ \left(\int_{\hat{\mathcal{R}}_{s1}^c} I(g' = 1) - \int_{\hat{\mathcal{R}}_{s2}} I(g' = 2) \right) \Phi(\gamma^{(g')} - \tau_s^{(g')} - \boldsymbol{\beta}^\top \mathbf{y}) + \int_{\hat{\mathcal{R}}_{s2}} I(g' = 2) \right\} \phi_C(\mathbf{y} - \boldsymbol{\mu}_s^{(g')}|\boldsymbol{\Sigma})\mathrm{d}\mathbf{y}. \quad (6)$$

The actual error rate is given by AER $= (P(1|2) + P(2|1))/2$. To obtain an estimate of its mean μ_{AER}, AER may be evaluated by using consistent estimates of $\boldsymbol{\Theta}$ and then evaluating the

integrals in Equations (5) and (6) by Monte Carlo methods. For example, to obtain a Monte Carlo approximation of

$$\hat{I}_s = \int_{\hat{\mathcal{R}}_{s1}} \Phi(\hat{\gamma}^{(g')} - \hat{\tau}_s^{(g')} - \hat{\boldsymbol{\beta}}^\top \mathbf{y}) \phi_C(\mathbf{y} - \bar{\mathbf{y}}_s^{(g')} | \mathbf{S}_p) \mathrm{d}\mathbf{y}$$

in Equation (6), we have $\hat{I}_s \approx \sum_{\forall \mathbf{y}_m^\dagger \in \hat{\mathcal{R}}_{s1}} \Phi(\hat{\gamma}^{(g')} - \hat{\tau}_s^{(g')} - \hat{\boldsymbol{\beta}}^\top \mathbf{y}_m^\dagger)/M$, where $\mathbf{y}_1^\dagger, \ldots, \mathbf{y}_M^\dagger$ are a Monte Carlo sample of size M (usually, 5000) generated from the multivariate normal distribution with mean $\bar{\mathbf{y}}_s^{(g')}$ and covariance matrix $\mathbf{S}_p$. We call $\hat{\mu}_{\mathrm{AER}}^{\mathrm{M}}$ obtained in this way the plug-in Monte Carlo estimate.

Another method for determining misclassification error rates is Lachenbruch's hold-out procedure [11, p. 599]. It works well for moderate-sized samples and provides a nearly unbiased estimate of μ_{AER}. Although it is relatively easy to implement using *R*, for example, this approach is more computationally intensive than the plug-in Monte Carlo method since the classification rule (5) needs to be re-estimated at each hold-out step. We call $\hat{\mu}_{\mathrm{AER}}^{\mathrm{H}}$ obtained in this way the hold-out estimate.

3.2 *Simulations*

In the simulations, $G = 2$ and GMDMs with $C = L = Q = 1$ and $S = 2$ are considered, which correspond to the case of one continuous variable, one nominal vector with two states, and one ordinal variable with two levels. Suppressing superscripts for convenience, the parameters are then $\boldsymbol{\Theta}_1 = (\pi, \mu_1, \mu_2, \gamma, \tau)^\top$ and $\boldsymbol{\Theta}_2 = (\sigma^2, \rho, \beta)^\top$, where μ_s is the sth state mean of continuous variable Y, γ the (standardized) cutpoint α for latent variable Y^* underlying ordinal variable Z, and τ the effect of State 1 on Z relative to that of State 2. Note that $\gamma = \alpha/\sqrt{1-\rho^2} - (\mu_2^*/\sqrt{1-\rho^2} - \beta\mu_2)$, $\beta = \rho/(\sigma\sqrt{1-\rho^2})$, and $\tau = \xi^*/\sqrt{1-\rho^2} - \beta\xi$, where $\xi = \mu_1 - \mu_2$, $\xi^* = \mu_1^* - \mu_2^*$, $\mu_s^* = E(Y^*|\mathbf{x} = \mathbf{x}_{(s)})$, and ρ is the correlation between Y and Y^*. Note also that $Z = 2$ if $Y^* > \alpha$ and $Z = 1$ if $Y^* \leq \alpha$. Similar to de Leon and Carrière [8], the following four cases are considered:

(I) there is difference between populations only with respect to nominal vector;
(II) there is difference between populations only with respect to continuous variable;
(III) there is difference between populations only with respect to ordinal variable;
(IV) populations are different with respect to all variables.

To evaluate the performance of the classification rule, independent samples of sizes $(N^{(1)}, N^{(2)}) = (100, 50)$, $(100, 100)$, and $(250, 200)$ were generated from GMDMs for $\boldsymbol{\Pi}^{(1)}$ and $\boldsymbol{\Pi}^{(2)}$ with $\boldsymbol{\Theta}_2^{(1)} = \boldsymbol{\Theta}_2^{(2)} = (\sigma^2, \rho, \beta)^\top = (1, 0.5, 0.58)^\top$ and $\boldsymbol{\Theta}_1^{(g)} = (\pi^{(g)}, \mu_1^{(g)}, \mu_2^{(g)}, \gamma^{(g)}, \tau^{(g)})^\top$, $g = 1, 2$, given by (I) $(0.3, 0, 2, 0, -1.15)^\top$ for $\boldsymbol{\Pi}^{(1)}$ and $(0.8, 0, 2, 0, -1.15)^\top$ for $\boldsymbol{\Pi}^{(2)}$, (II) $(0.5, 3, -1, 1.73, 2.31)^\top$ for $\boldsymbol{\Pi}^{(1)}$ and $(0.5, 4, 0, 1.15, 2.31)^\top$ for $\boldsymbol{\Pi}^{(2)}$, (III) $(0.5, 0, 0.5, 0.87, -0.29)^\top$ for $\boldsymbol{\Pi}^{(1)}$ and $(0.5, 0, 0.5, -2.02, -2.02)^\top$ for $\boldsymbol{\Pi}^{(2)}$, and (IV) $(0.3, 2.5, 1.1, 0.52, 0.81)^\top$ for $\boldsymbol{\Pi}^{(1)}$ and $(0.8, 0, -1.8, 1.27, 0.12)^\top$ for $\boldsymbol{\Pi}^{(2)}$. Parameter estimates were obtained by maximum likelihood.

Table 1 displays hold-out and plug-in Monte Carlo estimates of the true mean AER μ_{AER}. Estimates were obtained from 500 simulated samples, with plug-in Monte Carlo estimates evaluated using 5000 Monte Carlo samples. True values for μ_{AER}, obtained by Monte Carlo approximation, are 0.25 for Case (I), 0.3076 for Case (II), 0.254 for Case (III), and 0.0669 for Case (IV). The simulation results suggest that the estimates $\hat{\mu}_{\mathrm{AER}}^{\mathrm{M}}$ and $\hat{\mu}_{\mathrm{AER}}^{\mathrm{H}}$ both perform well. While the means of the estimates indicate some bias, particularly $\hat{\mu}_{\mathrm{AER}}^{\mathrm{H}}$ – with slight improvements as the sample sizes increase – the estimates are able to capture the true error rate for the classification rule.

Table 1. Simulation study on misclassification error rates of GMDM-based classification rule (4) with $C = L = Q = 1$, $S = 2$ versus IS-MDP method.

Case	Sample sizes		Mean of estimates		Mean number of hold-out misclassifications	
	$N^{(1)}$	$N^{(2)}$	$\hat{\mu}^{\mathrm{M}}_{\mathrm{AER}}$	$\hat{\mu}^{\mathrm{H}}_{\mathrm{AER}}$	GMDM	IS-MDP
I	100	50	0.2485	0.2862	42.93	40.85
	100	100	0.2497	0.2559	51.02	51.18
	250	200	0.2507	0.2579	116.06	116.65
II	100	50	0.3094	0.3076	46.14	46.2
	100	100	0.306	0.3126	62.52	66.94
	250	200	0.3094	0.3065	137.93	143.42
III	100	50	0.2449	0.2574	38.61	36.84
	100	100	0.2462	0.2544	50.88	49.96
	250	200	0.2449	0.2496	112.32	111.78
IV	100	50	0.071	0.0741	11.12	14.69
	100	100	0.0681	0.0732	14.64	21.18
	250	200	0.0659	0.0658	29.61	42.03

Notes: Parameter configurations for Cases (I)–(IV) are given in Section 3.2. For the plug-in Monte Carlo and hold-out estimates, the true mean AER μ_{AER} is given by 0.25 for Case (I), 0.3076 for Case (II), 0.254 for Case (III), and 0.0669 for Case (IV). Hold-out error estimates were based on 500 simulation repeats and plug-in Monte Carlo estimates were based on 5000 Monte Carlo samples.

For comparison, we considered the minimum distance probability (MDP) method for mixed discrete and continuous data recently developed by Nuñez *et al.* [17]. The method is a robust discrimination algorithm based on a distance function. Its robustness owes to the fact that, while still model-based, it requires only marginals of the data and incorporates correlations by using a score transformation. In the simulations, we used the so-called individual-score (IS) distance, a Mahalanobis distance between scores, since this emerged as the best choice from simulations [17]. To obtain estimates of misclassification rates, a two-step hold-out procedure is adopted [17]. A FORTRAN program called MDP was used to carry out the simulations reported in Table 1. The results in Table 1 demonstrate that the GMDM-based classification rule is relatively effective in separating the two mixed-variate populations under various settings, producing generally fewer misclassifications than the MDP method. Its ability to correctly classify individuals also improves when differences between populations exist for both discrete and continuous data.

To supplement these results, we ran additional simulations with $G = 2$ and GMDMs with $C = Q = 1$ and $L = S = 2$. These correspond to the same setting as the previous one, except that the ordinal variable now has three levels. The same parameters as before need to be estimated with the addition of one extra cutpoint; suppressing superscripts, we have $\boldsymbol{\Theta}_1 = (\pi, \mu_1, \mu_2, \gamma_1, \gamma_2, \tau)^\top$ and $\boldsymbol{\Theta}_2 = (\sigma^2, \rho, \beta)^\top$. The following four scenarios are studied, with $\boldsymbol{\Theta}_2^{(1)} = \boldsymbol{\Theta}_2^{(2)} = (1, 0.5, 0.5774)^\top$ and common (unstandardized) cutpoints $\alpha_1 = 1$ and $\alpha_2 = 2$:

(A) $\boldsymbol{\Theta}_1^{(1)} = (0.4, 2.5, 1.1, 0.5196, 1.6743, 0.8083)^\top$ and $\boldsymbol{\Theta}_1^{(2)} = (0.6, 1.5, 0, 1.1547, 2.3094, 0.866)^\top$;
(B) $\boldsymbol{\Theta}_1^{(1)} = (0.3, 2.5, 1, 0.5774, 1.7321, 0.866)^\top$ and $\boldsymbol{\Theta}_1^{(2)} = (0.8, 0, 2, 0, 1.1547, -1.1547)^\top$;
(C) $\boldsymbol{\Theta}_1^{(1)} = (0.3, 3.5, 2, 0, 1.1547, 0.866)^\top$ and $\boldsymbol{\Theta}_1^{(2)} = (0.8, 0, 1, -1.7321, 1.7321, -0.5774)^\top$;
(D) $\boldsymbol{\Theta}_1^{(1)} = (0.3, 3.5, 2, 0, 2.3094, 0.866)^\top$ and $\boldsymbol{\Theta}_1^{(2)} = (0.8, -2, 1, -2.8868, -1.7321, -1.7321)^\top$.

Cases (A)–(D) correspond to increasing separation between the two populations. Equal sample sizes $N^{(1)} = N^{(2)} = 100$ were generated and then were used in the classification using rule (4)

Table 2. Simulation study on misclassification error rates of GMDM-based classification rule (4) with $C = Q = 1$, $L = S = 2$ versus IS-MDP method.

	Mean of $\hat{\mu}_{\text{AER}}$		Mean number of hold-out misclassifications	
Case	GMDM	IS-MDP	GMDM	IS-MDP
A	0.2862	0.3107	57.24	62.14
B	0.1349	0.3303	26.98	66.06
C	0.101	0.1757	20.2	35.14
D	0.0397	0.1032	7.95	20.64

Notes: Parameter configurations for Cases (A)–(D) are given in Section 3.2. Error rate estimates $\hat{\mu}_{\text{AER}}$ were obtained by hold-out for both GMDM-based and IS-MDP methods. Hold-out error estimates were based on 500 simulation repeats.

and the IS-MDP method. Table 2 shows the results, displaying average error rates and average number of hold-out misclassifications for both GMDM-based rule (4) and the IS-MDP method. With the addition of another ordinal level, it appears that rule (4) outperforms the MDP method to a considerable degree, with the latter's error rates ranging between two to three times those of the former. This suggests that using GMDM-based methods for mixed data with non-binary multi-level ordinal variables results in significant gains in classificatory performance.

While the MDP method generally performed well compared with the GMDM-based rule for the cases with only binary ordinal variables (or so-called 'low-information' variables), exhibiting the method's claimed robustness [17] and flexibility in handling different mixed-data populations, this was not the case when polytomous ordinal variables are involved. The superiority of the GMDM-based method was apparent in mixed-data with 'high-information' ordinal variables. With its closed-form classification rules that are easily applied in practice, the GMDM-based method becomes specially useful in clinical settings, where classifications need to be determined by health practitioners, as illustrated by the example in the next section.

4. Application to croup data

For the croup data, we considered characteristics of mixed types associated with patients that fell into one of the two groups: those discharged home ('outpatients') and those admitted to the hospital ('inpatients'). Data in our analysis were based on $N^{(1)} = 46$ inpatients and $N^{(2)} = 152$ outpatients on the following variables: TRT, a nominal variable with three states representing the treatment received by patients (1, 2, and 3, corresponding respectively, to 'received both nebulized racemic epinephrine and corticosteroids,' 'received only nebulized racemic epinephrine,' and 'received only corticosteroids or neither'); HRT, a continuous variable representing patient's heart rate; TEMP, a continuous variable representing patient's temperature; RESP, a continuous variable measuring patient's respiratory rate; WT, a continuous variable for patient's weight; and CRP, an ordinal variable with two levels representing patient's croup score (1 if low, 2 if high).

GMDM is appropriate in this case, as it allows us to correctly model CRP as an ordinal outcome, and not a nominal one. In the terminology of Cox and Wermuth [5, p. 3], intermediate continuous variables HRT, TEMP, RESP, and WT are conditioned on explanatory nominal variable TRT. In addition, because CRP is the variable of primary interest, GMDM conditions CRP on HRT, TEMP, RESP, WT, and TRT. Thus, GMDM appropriately treats the variables according to their natural hierarchy, with CRP as the ultimate response. Moreover, GMDM, unlike GLOM and CGCM, enables the explicit accounting of the ordinal information in CRP and its associations with the nominal variable TRT, and the continuous variables TEMP, HRT, RESP, and WT, resulting in a

meaningful delineation of the correlations between CRP and TRT and between CRP and HRT, for example. This becomes impossible when CRP is treated as another nominal variable as in GLOM, for example, as this embeds its relationships with TRT within those with the continuous variables.

Joint normality checks using chi-square probability plots [11, p. 182] were constructed (not shown) and suggested log-transformed variables lTEMP = log(TEMP), lRESP = log(RESP), and lWT = log(WT), in addition to HRT. To supplement these, individual normal *Q–Q* plots for these variables are shown in Figure 1 for each of the three TRT states for outpatients; normal *Q–Q* plots were similarly constructed for inpatients but are not shown here. The plots generally appear to suggest that normality is a reasonable assumption.

Six discriminant functions from Equation (4) corresponding to state-level pairs for TRT and CRP were estimated by plug-in using MLEs. For example, (TRT = 1,CRP = 1) yields the following discriminant function

$$\hat{\delta}_{11} = 62.12 + 0.02\text{HRT} + 0.75\text{lTEMP} - 18.52\text{lRESP} + 0.18\text{lWT} + \log\left(\frac{\Phi(1.51 - 0.01\text{HRT} - 1.44\text{lTEMP} + 2.7\text{lRESP} - 1.11\text{lWT})}{\Phi(-86.93 - 0.03\text{HRT} - 0.37\text{lTEMP} + 26.27\text{lRESP} - 0.74\text{lWT})}\right),$$

and a patient in State 1 with low croup score is admitted for hospitalization if $\hat{\delta}_{11} \geq 0$ and discharged, otherwise. Similar allocation rules were obtained from discriminant functions $\hat{\delta}_{12}, \ldots, \hat{\delta}_{32}$ for (TRT, CRP) = (1, 2), . . . , (3, 2). These functions are simple and easy to use in actual clinical situations where a health professional needs to make a determination of whether to admit or discharge a patient.

Classifications using the GMDM-based rule and IS-MDP are shown in Table 3, assuming equal prior probabilities for the two patient groups and equal costs of misclassification. The

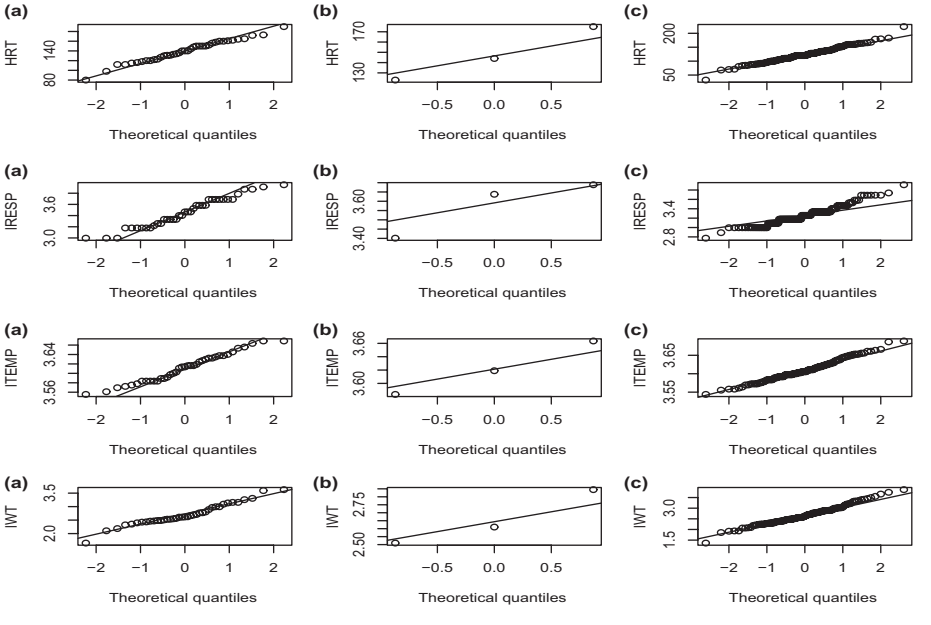

Figure 1. *Q–Q* normal plots of HRT, lTEMP, lRESP, and lWT by TRT for outpatients: (a) corresponds to 'State 1: received both', (b) to 'State 2: received only racemic', and (c) to 'State 3: received only corticosteroids or neither'.

Table 3. Number of hold-out misclassifications of GMDM-based rule versus IS-MDP for croup data.

Group	Sample Size	Method	
		GMDM	IS-MDP
Inpatients	46	20 (43.5)	29 (63)
Outpatients	152	56 (36.8)	39 (25.6)
Total	198	76 (38.4)	68 (34.3)

Notes: Classification for both methods was done with equal prior probabilities for the two groups and equal costs of misclassification. Percentages of misclassified patients are shown in parentheses.

results are very similar for both methods, with overall misclassification error rates in the mid-30%. This is to be expected since the data contain only 'low-information' binary nominal and ordinal variables. The plug-in Monte Carlo estimate of μ_{AER} is $\hat{\mu}^{\text{M}}_{\text{AER}} = 0.38$. Note that we could minimize the number of misclassifications by adopting the trivial rule, which always classifies a patient as outpatient. In this case, the total number of misclassified patients would be 46, the sample size from the inpatient group. This is generally the result we obtain by incorporating prior probabilities equal to the observed sample proportions when the sample sizes are quite different, with one being considerably larger than the other. For example, rule (4) in this case yields 9 out of 152 misclassified patients for the inpatient group and 43 out of 46 for the outpatient group.

5. Discussion

In this paper, we developed a classification methodology based on GMDM, a general model for mixed discrete and continuous variables which incorporates associations and correlations between different variable types and accounts for their measurement levels. The methodology extends GLOM-based LLDF to data with mixtures of nominal, ordinal, and continuous variables. Simulations showed that the methodology is effective in correctly classifying individuals, showing relatively superior performance compared with robust MDP method, especially in mixed-data involving 'high-information' multi-level ordinal variables. Error rates can be estimated either by plug-in Monte Carlo or hold-out method. Both approaches yield nearly unbiased estimates for large samples. We presented an application of the methodology to data from a study on croup in children. Rules based on GMDM and MDP performed almost identically. Our results indicate satisfactory performance of GMDM-based methods. They provide reasonable alternatives to current mixed-data methods and have the advantage of yielding closed-form classification rules that are very useful in clinical settings.

An important issue about mixed-data models like GMDM is the sample size requirements necessary to be able to estimate the model parameters well. Nakanishi [16] recently suggested that roughly $10 \times 2^S \times C$ mixed observations on C continuous and S binary variables are necessary for hypothesis tests involving GLOM, which has $L - Q$ fewer parameters than GMDM. For the simplest GMDM with $C = L = Q = 1$, our experience indicates that reasonably good estimates can be obtained with samples of sizes as small as 50, which agrees with Nakanishi's rule. We thus anticipate similar sample size requirements for GMDM estimation; however, this requires an extensive investigation encompassing various scenarios and parameter configurations. Related to this is the comparative performance of the two estimation methods discussed in Section 2, especifically as they relate to classification. We hope to more fully address this issue in a future work.

Another issue relevant to practitioners concerns the relative performance of classification rules compared with other simpler and/or more complex competitors. One way the classification rules in Section 3 may be simplified is to assume independence of certain variables; for example, we can take $[\mathbf{x}, \mathbf{y}, \mathbf{z}] = [\mathbf{x}][\mathbf{y}][\mathbf{z}]$, which is equivalent to assuming $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_S$, $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_Q = \mathbf{0}$, and $\tau_{sq} = 0\ \forall s, q$ [9]. Classification rules in this case are akin to diagonal linear discriminant analysis, which has been shown to work remarkably well, and in some cases, even better than other more sophisticated methods, for high-dimensional data with small sample sizes. Merits of such a method include its simplicity and the ease of its implementation, and while ignoring correlations may be problematic, it has been demonstrated to be as good as those methods that incorporate them. In the context of mixed-data classification, it is of interest to explore if such parsimonious simplifications, which greatly alleviate estimation, can perform as well as the rules in Section 3.

Finally, GMDM-based classification rules rely on a number of assumptions that may prove stringent in practice. Our simulations only confirmed their good performance in settings where such assumptions hold. While we expect them to be robust to certain violations (e.g. nonnormality), similar to LDA, a thorough study is required to confirm this. We plan to report the results of such a study in a separate work. Of particular interest is the relative robustness of rules based on certain independence assumptions to violations of such assumptions – leading to greatly simplified rules – especially in problems with sample sizes that are not large relative to the number of parameters to be estimated.

Acknowledgements

This research was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. A.S. was an NSERC summer research student when this research was completed. T.W. was supported by a Studentship Award from the Alberta Heritage Foundation for Medical Research. The authors are grateful to Prof. J. Oller, Universitat de Barcelona, for providing the MDP program, and to G. Duggan and J. Owoc for computational help. They are thankful to several anonymous reviewers for insightful comments leading to an improved article. This work was partially carried out while A.R.L. was enjoying the hospitality of the School of Statistics, University of the Philippines.

References

[1] A. Bar-Hen and J.-J. Daudin, *Discriminant analysis based on continuous and discrete variables*, in *Statistical Methods for Biostatistics and Related Fields*, W. Härdle, Y. Mori, and P. Vieu, eds., Springer, Berlin, 2007, pp. 3–28.

[2] Y.M. Bishop, S.E. Fienberg, and P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, MA, 1975.

[3] C.L. Bjornson, T.P. Klassen, J. Williamson, R. Brant, C. Mitton, A. Plint, B. Bulloch, L. Evered, and D.W. Johnson, *A randomized trial of a single dose of oral Dexamethasone for mild croup*, New Engl. J. Med. 351 (2004), pp. 1306–1313.

[4] P.C. Chang and A.A. Afifi, *Classifcation based on dichotomous and continuous variables*, J. Am. Stat. Assoc. 69 (1974), pp. 336–339.

[5] D.R. Cox and N. Wermuth, *Multivariate Dependencies: Models, Analysis and Interpretation*, Chapman & Hall, Boca Raton, 1996.

[6] C.M. Cuadras, *Some examples of distance based discrimination*, Biometrical Lett. 29 (1992), pp. 3–20.

[7] C.M. Cuadras, J. Fortiana, and F. Oliva, *The proximity of an individual to a population with applications in discriminant analysis*, J. Classif. 14 (1997), pp. 117–136.

[8] A.R. de Leon and K.C. Carrière, *A generalized Mahalanobis distance for mixed data*, J. Multivariate Anal. 92 (2005), pp. 174–185.

[9] A.R. de Leon and K.C. Carrière, *General mixed-data model: extension of general location and grouped continuous models*, Can. J. Stat. 35 (2007), pp. 533–548.

[10] V.P. Godambe, *Estimating Functions*, Oxford University Press, Oxford, 1991.

[11] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., Prentice Hall, New York, 2007.

[12] W.J. Krzanowski, *Discrimination and classification using both binary and continuous variables*, J. Am. Stat. Assoc. 70 (1975), pp. 782–790.

[13] W.J. Krzanowski, *The location model for mixtures of categorical and continuous variables*, J. Classif. 10 (1993), pp. 25–49.
[14] S.-Y. Lee, X.-Y. Song, and B. Lu, *Discriminant analysis using mixed continuous, dichotomous, and ordered categorical variables*, Multivariate Behav. Res. 42 (2007), pp. 631–645.
[15] N.I. Mahat, W.J. Krzanowski, and A. Hernandez, *Variable selection in discriminant analysis based on the location model for mixed variables*, Adv. Data Anal. Classif. 1 (2007), pp. 105–122.
[16] H. Nakanishi, *Tests of hypotheses for the distance between populations on the mixture of categorical and continuous variables*, J. Jpn Soc. Comput. Stat. 16 (2003), pp. 53–62.
[17] M. Nuñez, A. Villaroya, and J.M. Oller, *Minimum distance probability discriminant analysis for mixed variables*, Biometrics 59 (2003), pp. 248–253.
[18] I. Olkin and R.F. Tate, *Multivariate correlation models with mixed discrete and continuous variables*, Ann. Math. Stat. 32 (1961), pp. 448–465.
[19] U. Olsson, F. Drasgow, and N.J. Dorans, *On the robustness of factor analysis against crude classification of the observations*, Multivariate Behav. Res. 14 (1979), pp. 485–500.
[20] W.-Y. Poon and S.-Y. Lee, *Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients*, Psychometrika 52 (1987), pp. 409–430 (correction in 53, p. 301).
[21] J.L. Schafer, *Analysis of Incomplete Multivariate Data*, Chapman & Hall, New York, 1997.
[22] M. Tan, Y. Qu, and J.S. Rao, *Robustness of the latent variable model for correlated binary data*, Biometrics 55 (1999), pp. 258–263.
[23] A. Villaroya, M. Riós, and J.M. Oller, *Discriminant analysis algorithm based on a distance function and on a Bayesian decision*, Biometrics 51 (1995), pp. 908–919.