CLASSIFICATION OF A CORRELATED BINARY OBSERVATION

SANTOSH C. SUTRADHAR

Classification of a Correlated Binary Observation

by

©Santosh C. Sutradhar

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

May 1998

St. John's Newfoundland Canada

Abstract

A bivariate binary observation is traditionally classified into one of the two possible groups under the assumption that the cell counts follow a suitable multinomial distribution. But, in the traditional approach, the joint probability for each of these cell counts is unknown. Consequently it is not clear, how the traditional approach takes into account the correlation that may exist between two 2-dimensional binary observations. In this thesis, following Prentice [27] (Biometrics, 1988), we model the cell probabilities by a suitable bivariate binary distribution and examine the effect of this type of modelling in classifying a new correlated bivariate binary observation. The performance of the usual optimum classification procedure based on the proposed modelling of the cell probabilities are then compared with the model-free existing procedure. This is done through a simulation, by comparing the probabilities of misclassification for the two approaches, for various sample sizes and selected values of the marginal probabilities as well as correlation parameter between the two binary observations. We illustrate the use of the joint probability modelling in classification by analyzing a combined data set from two epidemiological surveys of 6-11 years old children conducted in Connecticut, the New Heaven Child Survey (NHCS) and the Eastern Connecticut Child Survey (ECCS).

Acknowledgments

I am grateful to my supervisor. Dr. B. C. Sutradhar for his encouragement to explore various kinds of research ideas that helped me to choose the present research topic. and also for his continuous guidance and helpful assistance in completing this thesis.

I am grateful to Drs Uditha Balasooria and John Hoenig, the members of my supervisor committee.

I sincerely acknowledge the financial support provided by the School of Graduate Studies and Department of Mathematics and Statistics in the form of Graduate Fellowships and Teaching Assistantships. Further I wish to thank Drs. Bruce Watson and Herbert Gaskil. the former and the present Department Heads. for providing me with a friendly atmosphere and the necessary facilities to complete the program.

Finally, I am grateful to my parents and brothers for their eternal love. emotional support. and encouragement during this program.

Last but not least. it is my great pleasure to thank my friends and well-wishers who directly or indirectly encouraged and helped me in the M. Sc. program and contributed to this dissertation.

Contents

List of Tables

Chapter 1

Introduction

1.1 Motivation of the Problem

Since R. A. Fisher's pioneering work (cf. Fisher [11]) in the thirties in the area of classification or discriminant analysis, there has been extensive work on this topic, mainly for variables of a continuous nature. This classification problem is quite important in practice. For example, in clinical studies, it may be very important to classify an incoming patient into a suspected disease group or into a non-disease group. Here, in this type of problem, it is customary to study the behavior of patients from both the disease and the non-disease groups and then base the classification of the new patient on the information available from these two groups. Similar problems frequently arise in other biomedical, social, natural, and physical sciences.

As mentioned earlier, most of the theory of discrimination and investigation of robustness properties for classification criterion are based on the normal

and other continuous distributions. In practice, there are many situations where the information may be binary or discrete. For example, consider a study in which information on rating of child's mental health status were collected from a parent and also from a child's teacher. This rating is a measure of emotional and "internalizing" disturbance, obtained by dichotomizing the corresponding scale score at the clinical "border-line" range. Here, it may be of interest from certain 'investigation' point of view to determine the sex of a child in question based on the information provided by both the teacher and the parent. It is clear that this is a classification problem for a bivariate binary observation.

The problem of discrimination with binary data is, however, not adequately addressed in the literature. There are some approaches suggested in the literature (cf. Seber [29] and the references there in) to deal with discrete data. Problems arise when the binary data are dependent. In our example, it is also reasonable to assume that the parents' and teachers' ratings are positively correlated as they are rating on the same child. Sometimes it is not easy to consider the pattern of dependency among the binary variables. This situation is noticed in the previous studies by Bahadur [5], Martin [24], Ott [26], Goldstein [16], Lachenbruch [22], McLachlan [25], and others. Consequently, in the existing literature, this type of correlated binary data have been classified based on a suitable multinomial distribution for the counts in each of the four cells, without modelling the probability structure in terms

of a correlation parameter. This observation motivated us to model the cell probabilities in the bivariate binary case by a suitable probability model taking the correlation into account in a natural way and to examine the effect of such modelling in classifying a bivariate binary observation into one of the two groups.

Furthermore, in practice, we may have correlated binary data with a set of covariates for each of the individuals in the study. In the thesis, we have also included this case and discussed the classification of a bivariate binary observation when covariates for individuals are available.

1.2 Objective of the Thesis

The main objectives of this thesis is to examine the effect of the modelling of the cell probabilities for a bivariate binary data set in classifying a new observation into one of the two groups. The specific plan of the thesis is as follows.

In chapter 2, we provide detail background of the classification problem for variables of a continuous and discrete nature.

Chapter 3 concentrates on the description of classification with correlated binary data by using an appropriate probability model. More specifically, in Section 3.1, we propose a joint probability model for correlated binary data as a function of marginal probabilities and the structural correlation param-

eter. Section 3.2 is dedicated to develop the optimum classification criterion based on the proposed approach where the cell probabilities of the four cells of the bivariate binary data set are modelled in terms of the marginal probabilities and the correlation parameter. Also in the same Section 3.2, we compare and contrast the existing classification procedure with the proposed procedure, where in the existing approaches, these do not have any specific probability structure for the cell counts under investigation. In Section 3.3, we have shown, using simulated correlated binary data, how our suggested probability modelling performs better as compared to the situation where the cell probabilities are not modelled by using any probability distribution. And finally in Section 3.4 we illustrate our method by a suitable example.

Chapter 4 is concerned with the classification of correlated binary data with covariates. The estimating equation for the regression parameters as well as the correlation parameter are computed based on the information of the covariates.

We conclude the thesis in chapter 5 with some remarks about the importance of modelling the cell probability in classifying a new bivariate binary observation into one of the two groups. In the same chapter, we have also discussed the possibilities of some future research in this area.

Chapter 2

Background of Classification Problems

The problem of classification arises when an investigator makes a number of measurements on an individual and wishes to classify the individual into one of several categories on the basis of these measurements. In brief. one may state the problem as follows: Given an individual with certain measurements; if several population exist from which this individual may have come. the question is, from which population did it arise?

2.1 Classification under Certain Continuous Distribution: Parametric Approach

There is a vast literature on discrimination for this case. In order to classify an observation into one of the populations, Fisher [11] suggested as a basis of classification decisions the use of a discriminant function linear in the components of the observations. Other bases for classification have included

likelihood ratio tests (cf. Anderson [4]), information theory (Kullback [21]), and Bayesian techniques (cf. Geisser [15]). In all cases, sampling theories have been considered under the assumption that the populations involved are multivariate normal. The problem of classification has also been studied for other continuous distributions. See, for example, Kariya [20], and Sutradhar [32] for discrimination analysis under general elliptical or t distribution set up.

2.2 Classification for Continuous or Discrete Data: Non-parametric Approach

In the continuous set up, there exist some other approaches where robust discrimination criteria are used to classify a new observation into one of the two or more groups. These classification procedures are not dependent on any particular distribution. For example, Chen and Muirhead [8] constructed a discriminant procedure by deriving robust discriminant functions using projection pursuit criteria. Projection pursuit, a computer-intensive methodology, was first successfully implemented on the computer by Friedman and Tukey [14], and thorough reviews have been given by Huber [18] and Jones and Sibson [19]. In order to evaluate the robustness and the performance of their discriminant rules under various distributional situations, Chen and Muirhead [8] (see also Chen [7]) did a Monte Carlo simulation based on the bivariate normal, Cauchy, log-normal, and contaminated normal distribu-

tions, which are continuous. There also exist similar but different (than the projection pursuit method) non-parametric classification approaches in the literature. For example, we refer to the references in McLachlan (chap 9) [25] for details. Among these approaches, the so-called kernel discriminant analysis is widely used in non-parametric classification analysis. The kernel density estimator, originally suggested by Fix and Hodges [13], can be used to estimate the density of both continuous and discrete feature data. The kernel method may be described in brief as follows:

Let $\mathbf{y}_{il}$ be the q-dimensional lth ($l = 1, \cdots, n_i$) observation in the ith group $G_i (i = 1, \cdots, g)$. For a continuous q-dimensional feature vector $\mathbf{Y}$, a non-parametric estimate, $\hat{f}_i^{(K)}(\mathbf{y})$, of the ith group density $f_i(\mathbf{y})$ provided by the kernel method is

$$\hat{f}_i^{(K)}(\mathbf{y}) = \left(\frac{1}{n_i}\right)\left(\frac{1}{h_i^q}\right)\sum_{l=1}^{n_i} K_q\left(\frac{\mathbf{y} - \mathbf{y}_{il}}{h_i}\right) \tag{2.2.1}$$

where K_q is a kernel function that integrates to one, and h_i is a smoothing parameter. The smoothing parameter h_i is known also as the bandwidth or window width which is a function of the ith group-sample size n_i. With most applications, the kernel K_q is fixed and the smoothing parameter h_i is specified as a function of the data. Usually, but not always, the kernel K_q is required to be nonnegative and symmetric, that is,

$$K_q(\mathbf{y}) \geq 0, \quad \text{and} \quad K_q(\mathbf{y}) = K_q(-\mathbf{y}) \quad \mathbf{y} \in \mathbb{R}^q$$

If the above condition holds, the kernel density estimate can be interpreted

as a mixture of n_i component densities in equal proportions.

By virtue of its definition, the kernel density approach to estimation is resistant to the effect of outliers. This is because $K_q[(\mathbf{y} - \mathbf{y}_{il})/h_i]$ must become small if $\mathbf{y}_{il}$ is far from $\mathbf{y}$. For computational aspects of kernel density estimation we refer to Silverman (section 3.5) [30]. Now the problem is to choose the kernel function in the definition (2.2.1) of the kernel density estimator. Epanechnikov [10] and Deheuvels [9] used an asymptotic argument to show that there is very little to choose between different kernel functions. Among the various kernels considered by Cacoullos [6] was the so-called product kernel,

$$K_q(\mathbf{y}) = \prod_{j=1}^{q} K_1(y_j),$$

where K_1 is a univariate probability density function. This yields

$$\hat{f}_i^{(K)}(\mathbf{y}) = \left(\frac{1}{n_i}\right)\left(\frac{1}{h_i^q}\right)\sum_{l=1}^{n_i}\prod_{j=1}^{q} K_1\left(\frac{\mathbf{y_j} - \mathbf{y}_{ilj}}{h_i}\right). \tag{2.2.2}$$

A common choice for the univariate kernel $K_1(\mathbf{y})$ is the univariate standard normal density function. With this choice, $\hat{f}_i^{(K)}(\mathbf{y})$ is estimated by a spherical normal kernel,

$$\hat{f}_i^{(K)}(\mathbf{y}) = \left(\frac{1}{n_i}\right)\sum_{l=1}^{n_i}\phi(\mathbf{y} - \mathbf{y}_{il}; \mathbf{0}, \mathbf{M}_i), \tag{2.2.3}$$

where $\mathbf{M}_i = h_i^2\mathbf{I}_q$, and $\phi(\mathbf{x};\mu,\Sigma)$ denotes the multivariate normal density with mean μ and covariance matrix Σ. The analogue of (2.2.1) for a discrete feature vector is considered in Section 2.2.1 in the context of a special multivariate binary data.

2.2.1 Kernel Discriminant Analysis: A Non-parametric Approach for Multivariate Binary Data

The classification problem based on binary data may arise in many biomedical situations. For an example of this type of problem we refer to Anderson et al. (1972) [3] where the condition keratoconjunctivities sicca, or dry eyes, is studied. The study refers to 10 symptoms (redness, itchiness, soreness or pain, burning, etc.) that are associated with this condition. Each symptom is either present or absent in each individual, and they are expected to be correlated to one another. For a given vector it is of interest to make a diagnosis (yes-no) for the disease. A training sample of 40 diseased patients and 37 non-diseased patients was available for use in diagnosis. Since the multivariate binary density is not known, one may give a special concentration towards kernel discriminant analysis in the context of this type of binary data.

Following Aitchison and Aitken [1] a binomial kernel may be used, whereby one estimates $\hat{f}_i(\mathbf{y})$ as

$$\tilde{f}_i^{(K)}(\mathbf{y}) = \frac{1}{n_i} \sum_{l=1}^{n_i} \mathbf{K_q}(\mathbf{y} : \mathbf{y}_{il}, h_i) \tag{2.2.4}$$

where

$$\mathbf{K_q}(\mathbf{y} : \mathbf{y}_{il}, h_i) = h_i^{q-d_{il}^2}(1-h_i)^{d_{il}^2}$$

with $\frac{1}{2} \leq h_i \leq 1$, and

$$d_{il}^2 = ||\mathbf{y} - \mathbf{y}_{il}||^2 = (\mathbf{y} - \mathbf{y}_{il})'(\mathbf{y} - \mathbf{y}_{il})$$

If we put $h_i = 1$, then $\tilde{f}_i^{(K)}(\mathbf{y})$ reduces to the multinomial estimate $n_i(\mathbf{y})/n_i$, where $n_i(\mathbf{y})$ is the number of sample points with $\mathbf{y} \equiv \mathbf{y}_{il}$ for all $l = 1, \cdots, n_i$. As h_i decreases from one, the smoothing of the multinomial estimates increases, so that at $h_i = 1/2$, it puts equal mass $1/2^q$ at possible realizations of $\mathbf{Y}$.

Once the density of the i-th group G_i $(i = 1, 2)$ is estimated by using the binomial kernel, the new observation $\mathbf{Y}$ may be classified to G_1 provided

$$\tilde{f}_1^{(K)}(\mathbf{y}) > \tilde{f}_2^{(K)}(\mathbf{y})$$

Ott and Kronmal [26] also introduced a non-parametric method of density estimation for multivariate binary data which is based on orthogonal expansion of the density in terms of a discrete Fourier series. Liang and Krishnaiah [23]

used the same approach, only with different coefficients. Both papers discuss the application of these procedures to the classification problem. Chen et al. [7] further extended his work, and Stoffer [31] expanded the discussion to binary time-series data.

We note here that this kind of classification problem for binary data has been studied in the literature using the semi-parametric approach. We now discuss this approach in the context of bivariate binary data in the following section.

2.3 Classification Rule for Bivariate Binary Data: Semi-parametric Approach

2.3.1 Basic Multinomial Approach

Suppose that $\mathbf{y}' = (y_1, y_2)$ is a vector of two binary variables, each taking the value 1 or 0 and it may arise from G_i for $i = 1, 2$. For $j = 1, 2$, now let $y_j = 1$ with probability p_{ij}, and $y_j = 0$ with probability $q_{ij} = 1 - p_{ij}$, if $\mathbf{y}$ comes from G_i $(i = 1, 2)$. Then $\mathbf{y}'$ can assume value of one of the following four multinomial cells

Cell	1	2	3	4
$\mathbf{y}$	(1,1)	(1,0)	(0,1)	(0,0)

Let $\theta_{(i)k}$ be the probability that the random vector $\mathbf{y}$ falls in the multinomial cell k ($k = 1, 2, 3, 4$) and π_i be the proportion of the i-th group in the whole population $\mathcal{P}$ of two groups G_1 and G_2. Now if an observation $\mathbf{y}$ falls into cell k, then the optimal classification rule is: Assign $\mathbf{y}$ to G_1 if

$$\frac{f_1(\mathbf{y})}{f_2(\mathbf{y})} = \frac{\theta_{(1)k}}{\theta_{(2)k}} > \frac{\pi_2}{\pi_1} \tag{2.3.5}$$

Note here that although we assume that $\theta_{(i)k}$ be the multinomial cell probability, no specific form of this probability is assumed here. Consequently, the approach considered here is a semi-parametric approach. Now the cell probabilities $\theta_{(i)k}$'s have to be estimated from the sample data. Without any loss of generality, let us assume that the y_j's are correlated. rather than independent. Also, suppose that we have a random sample of size n (fixed) from the population $\mathcal{P}$, of which n_i come from G_i, so that $n = n_1 + n_2$. Out of these n_i observations, let $n_{(i)k}$ fall into cell k so that $n_i = \sum_k n_{(i)k}$. Now, since

$$\begin{aligned} P[(\mathbf{y} \text{ in cell } k) \cap (\mathbf{y} \in G_i)] &= P[\mathbf{y} \text{ in cell } k | \mathbf{y} \in G_i] P[\mathbf{y} \in G_i] \\ &= \theta_{(i)k}\pi_i, \end{aligned}$$

the likelihood function based on the so-called mixture sampling approach (cf. Seber [29], Section 6.4.2) is given by

$$\begin{aligned} L(\theta_{(i)k}, \pi_i|\mathbf{y}) &= \prod_{i=1}^{2}\prod_{k=1}^{4} P[n_{(i)k} \text{ cell frequencies and } \mathbf{y} \in G_i] \\ &= \prod_{i=1}^{2}\prod_{k=1}^{4} (\theta_{(i)k}\pi_i)^{n_{(i)k}} \\ &= \prod_{i=1}^{2}[\pi_i^{n_i} \prod_{k=1}^{4} \theta_{(i)k}^{n_{(i)k}}] \end{aligned} \tag{2.3.6}$$

The maximum likelihood estimates of $\theta_(i)k$ and π_i are obtained by maximizing the likelihood with respect to $\theta_{(i)k}$, and π_i respectively. The estimates are:

$$\hat{\theta}_{(i)k} = \frac{n_{(i)k}}{n_i} \quad \text{and} \quad \hat{\pi}_i = \frac{n_i}{n}$$

Substituting these estimates in the optimum classification rule (2.3.5) gives the simple rule, for cell k, as follows: Assign $\mathbf{y}$ to G_1 if

$$\frac{n_{(1)k}}{n_1} > \frac{n_{(2)k}}{n_2} \quad k = 1, 2, 3, 4 \qquad \text{when} \quad \pi_1 = \pi_2 \tag{2.3.7}$$

which reduces to

$$n_{(1)k} > n_{(2)k} \quad k = 1, 2, 3, 4 \qquad \text{when} \quad \pi_1 \neq \pi_2 \tag{2.3.8}$$

2.3.2 Log-linear Representation in Basic Multinomial Approach

The cell probability involved in the multinomial model, for bivariate binary data, can be represented by a log-linear model for better understanding of the association between the two correlated binary variables. In this approach, though, the cell probability is not considered to have any parametric model, the log of any particular cell probability is expressed as a linear function of the main effects and interaction of the two variables. See equation (2.3.9) below for the specific relationship. But again, as there is no specific form for the cell probability, the approach is still considered as a semi-parametric approach.

Let $\mathbf{Y}' = (Y_1, Y_2)$ be the 2×1 random vector of two correlated binary variables Y_1 are Y_2 with joint probability function given by

$$f(y_1, y_2) = \begin{cases} \pi_{11}, & y_1 = 1, \quad y_2 = 1, \\ \pi_{10}, & y_1 = 1, \quad y_2 = 0, \\ \pi_{01}, & y_1 = 0, \quad y_2 = 1, \\ \pi_{00}, & y_1 = 0, \quad y_2 = 0, \end{cases}$$

where $\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} = 1$. Note that in terms of the notation of the previous section, by omitting the suffix for group, we have

$$\pi_{11} = \theta_1, \quad \pi_{10} = \theta_2, \quad \pi_{01} = \theta_3, \quad \text{and} \quad \pi_{00} = \theta_4.$$

These probabilities may be represented in the form of the following (2×2)

table:

	y_2		
y_1	0	1	Total
0	π_{00}	π_{01}	$\pi_{00} + \pi_{01}$
1	π_{10}	π_{11}	$\pi_{10} + \pi_{11}$
Total	$\pi_{00} + \pi_{10}$	$\pi_{01} + \pi_{11}$	1

where the probability π_{ij} corresponds to the cell represented by i and j where these i and j are the possible values of y_1 and y_2.

Further, note that in general in log-linear models (see Agresti (1990)) for two dimensions, the log of the cell mean is expressed as a linear function of the parameters. Let m_{ij} be the mean of the (i,j)th cell. As m_{ij} $(= n\pi_{ij})$ is the constant multiple of the corresponding cell probability π_{ij}, one may like to express $\log \pi_{ij}$ instead of $\log m_{ij}$ as a linear function of the parameters. Let $\mu_{ij} = \log \pi_{ij}$ and

$$\mu_{i.} = \frac{\mu_{i0} + \mu_{i1}}{2}, \quad \mu_{.j} = \frac{\mu_{0j} + \mu_{1j}}{2}$$
$$\text{and} \quad \mu = \mu_{..} = \frac{\mu_{00} + \mu_{01} + \mu_{10} + \mu_{11}}{4}$$

Here μ denotes the overall mean of the $\{\log \pi_{ij}\}$. Then the log of π_{ij} may be expressed in the form of linear function given by

$$\log \pi_{ij} = \mu + \lambda_i^{y_1} + \lambda_j^{y_2} + \lambda_{ij}^{y_1 y_2}, \tag{2.3.9}$$

where

$$\begin{aligned} \lambda_i^{y_1} &= \mu_{i\cdot} - \mu \quad \text{is the } i\text{th row effect of } y_1, \\ \lambda_j^{y_2} &= \mu_{\cdot j} - \mu \quad \text{is the } j\text{th column effect of } y_2, \\ \text{and} \quad \lambda_{ij}^{y_1 y_2} &= \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu \quad \text{is the interaction between } y_1 \text{ and } y_2. \end{aligned}$$

The notation in (2.3.9) is similar to that for the usual two-way analysis of variance. The row and the column effects $\{\lambda_i^{y_1}\}$ and $\{\lambda_j^{y_2}\}$, respectively, are defined so that they are deviations about the mean and hence

$$\begin{aligned} \lambda_0^{y_1} + \lambda_1^{y_1} = 0 \quad &\Rightarrow \quad \lambda_1^{y_1} = -\lambda_0^{y_1} \\ \lambda_0^{y_2} + \lambda_1^{y_2} = 0 \quad &\Rightarrow \quad \lambda_1^{y_2} = -\lambda_0^{y_2} \end{aligned} \tag{2.3.10}$$

Thus there is one independent row effect parameter, say $u_1 = \lambda_0^{y_1}$ and one independent column effect parameter, say $u_2 = \lambda_0^{y_2}$. Also we have one independent association parameter, say $u_{12} = \lambda_{00}^{y_1 y_2}$ as

$$\begin{aligned} & \lambda_{00}^{y_1 y_2} + \lambda_{01}^{y_1 y_2} = 0, \quad \lambda_{10}^{y_1 y_2} + \lambda_{11}^{y_1 y_2} = 0 \\ \Rightarrow \quad & \lambda_{00}^{y_1 y_2} = \lambda_{11}^{y_1 y_2} = -\lambda_{01}^{y_1 y_2} = -\lambda_{10}^{y_1 y_2} = u_{12} \end{aligned} \tag{2.3.11}$$

Writing u for μ and using (2.3.10) and (2.3.11) one obtains from (2.3.9) that

$$\begin{aligned} \log \pi_{00} &= u - u_1 - u_2 + u_{12}, \\ \log \pi_{01} &= u - u_1 + u_2 - u_{12}, \\ \log \pi_{10} &= u + u_1 - u_2 - u_{12}, \\ \log \pi_{11} &= u + u_1 + u_2 + u_{12}. \end{aligned} \tag{2.3.12}$$

In the above approach, the model given by (2.3.9) yielding four equations (2.3.12), is known as the log-linear model for the multinomial cell probability.

Note that as mentioned before u_1 and u_2 in equation (2.3.12) are known as the main effects and u_{12} is known as the interaction effect and they can be expressed as

$$u_1 = \frac{1}{4}(-\log \pi_{00} - \log \pi_{01} + \log \pi_{10} + \log \pi_{11}) = \frac{1}{4} \log \frac{\pi_{11}\pi_{10}}{\pi_{01}\pi_{00}}$$

$$u_2 = \frac{1}{4}(-\log \pi_{00} + \log \pi_{01} - \log \pi_{10} + \log \pi_{11}) = \frac{1}{4} \log \frac{\pi_{11}\pi_{01}}{\pi_{10}\pi_{00}}$$

$$u_{12} = \frac{1}{4}(\log \pi_{00} - \log \pi_{01} - \log \pi_{10} + \log \pi_{11}) = \frac{1}{4} \log \frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}$$

It is clear that if the last odds ratio is unity then $u_{12} = 0$ indicating that y_1 and y_2 are independent.

Therefore the log linear representation helps to interpret the association between y_1 and y_2 without specific assumption about the joint cell probability of y_1 and y_2. This representation for the association, however, may not be meaningful, if the exact joint probability structure does not permit log-linear representation.

2.3.2.1 Classification Rule for Bivariate Binary Data

The log-linear representation in (2.3.12) can be rewritten, in general, for appropriate values of y_1 and y_2, as

$$\begin{aligned}
\log f(y_1, y_2) &= \log \pi_{y_1 y_2} \\
&= u - u_1(1-2y_1) - u_2(1-2y_2) + u_{12}(1-2y_1)(1-2y_2) \\
&= (u - u_1 - u_2 + u_{12}) + 2(u_1 - u_{12})y_1 + \\
&\quad +2(u_2 - u_{12})y_2 + 4u_{12}y_1y_2 \\
&= \beta_0' + \beta_1' y_1 + \beta_2' y_2 + \beta_{12}' y_1 y_2, \quad \text{say.}
\end{aligned} \tag{2.3.13}$$

Now suppose that $\mathbf{Y} \in G_i$, then one may write

$$\log f_i(y_1, y_2) = \beta_{(i)0}' + \beta_{(i)1}' y_1 + \beta_{(i)2}' y_2 + \beta_{(i)12}' y_1 y_2 \tag{2.3.14}$$

Then we have

$$\log \frac{f_1(y_1, y_2)}{f_2(y_1, y_2)} = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_{12} y_1 y_2 \tag{2.3.15}$$

where

$$\beta_0 = \beta_{(1)0}' - \beta_{(2)0}'; \quad \beta_1 = \beta_{(1)1}' - \beta_{(2)1}'$$

$$\beta_2 = \beta_{(1)2}' - \beta_{(2)2}'; \quad \text{and} \quad \beta_{12} = \beta_{(1)12}' - \beta_{(2)12}'$$

Therefore, according to the optimum classification rule (2.3.5), assign an individual with measurement $\mathbf{y}$ to G_1 if

$$\log \left[\frac{f_1(y_1, y_2)}{f_2(y_1, y_2)} \right] = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_{12} y_1 y_2 \geq 0. \tag{2.3.16}$$

In practice these 3 parameters are not known and must be estimated from the sample data. If we use the well-known conditional sampling and estimate the 3 parameters based on the posterior likelihood (see Seber [29], section 6.4.2) then we obtain the classification rule as

$$n_{(1)k} \geq n_{(2)k} \quad k = 1.2.3.4 \tag{2.3.17}$$

Now to verify the classification rule (2.3.17) for the unknown parameter case we rewrite the likelihood function in (2.3.6) as

$$\begin{aligned} L(\theta_{(i)k}, \pi_i|\mathbf{y}) &= \prod_{i=1}^{2}\prod_{k=1}^{4}(\pi_i\theta_{(i)k})^{n_{(i)k}} \\ &= \left[\prod_{i=1}^{2}\prod_{k=1}^{4}\left(\frac{\pi_i\theta_{(i)k}}{\theta_{(\cdot)k}}\right)^{n_{(i)k}}\right]\left[\prod_{k=1}^{4}\theta_{(\cdot)k}^{n_{(\cdot)k}}\right] \\ &= \left[\prod_{i=1}^{2}\prod_{k=1}^{4}\{q_i(k)\}^{n_{(i)k}}\right] L_0 \quad \text{say} \\ &= L_c L_0 \quad \text{say.} \end{aligned} \tag{2.3.18}$$

where $\theta_{(\cdot)k} = \pi_1\theta_{(1)k} + \pi_2\theta_{(2)k}$.

Note that quite often inference is made based on L_c rather than L_cL_0 (see Seber [29], section 6.4.2). We, in this section, follow this and observe that L_c can be explicitly written as

$$L_c = \prod_{i=1}^{2}\prod_{l=1}^{n_i} q_1^*(Y)^{z_l} q_2^*(Y)^{1-z_l}$$

where z_l is an indicator variable defined as

$$z_l = \begin{cases} 1 & \text{if } y_{(i)jl} \in G_i \\ 0 & \text{otherwise} \end{cases}$$

with $y_{(i)jl}$ as the l-th observation of the j-th binary variable in the i-th group G_i and for given vector $\mathbf{Y} = \mathbf{Y_0}$ the posterior distribution of G_1 is defined as

$$\begin{aligned} q_1^*(\mathbf{Y_0}) &= P(G_1|\mathbf{Y} = \mathbf{Y_0}) \\ &= \frac{P(\mathbf{Y} = \mathbf{Y_0}|G_1)P(G_1)}{P(\mathbf{Y} = \mathbf{Y_0}|G_1)P(G_1) + P(\mathbf{Y} = \mathbf{Y_0}|G_2)P(G_2)} \\ &= \frac{f_1(y_1, y_2)\pi_1}{f_1(y_1, y_2)\pi_1 + f_2(y_1, y_2)\pi_2} \\ &= \frac{\pi_1 e^{\mathbf{Y_0}'\beta}}{\pi_1 e^{\mathbf{Y_0}'\beta} + \pi_2} \quad \text{since by} \quad (2.3.15) \quad f_1(y_1, y_2) = f_2(y_1, y_2)e^{\mathbf{Y_0}'\beta} \\ &= \frac{e^{\mathbf{Y_0}'\beta}}{1 + e^{\mathbf{Y_0}'\beta}} \quad \text{for} \quad \pi_1 = \pi_2 \end{aligned}$$

with $Y_{0i}' = (1, y_{(i)1l}, y_{(i)2l}, y_{(i)1l}y_{(i)2l})$ and $\beta = (\beta_0, \beta_1, \beta_2, \beta_{12})'$ and

$$q_2^*(\mathbf{Y_0}) = 1 - q_1^*(\mathbf{Y_0})$$

Now to estimate the β parameters, we rewrite L_c, as

$$\begin{aligned} L_c(\beta|\mathbf{Y} = \mathbf{Y_0}) &= \prod_{i=1}^{2}\prod_{l=1}^{n_i}\left(\frac{e^{\mathbf{Y_0}'\beta}}{1 + e^{\mathbf{Y_0}'\beta}}\right)^{y_l}\left(\frac{1}{1 + e^{\mathbf{Y_0}'\beta}}\right)^{1-y_l} \\ &= \frac{\exp[\sum_{i=1}^{2}\sum_{l=1}^{n_i} y_l(\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12}y_{(i)1l}y_{(i)2l})]}{\prod_{i=1}^{2}\prod_{l=1}^{n_i}(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12}y_{(i)1l}y_{(i)2l}])} \end{aligned}$$

The log of this L_c is given by

$$
\begin{aligned}
l_c(\beta) &= \sum_{i=1}^{2}\sum_{l=1}^{n_i} y_l(\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}) \\
&\quad - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \log(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]) \\
&= \sum_{l=1}^{n_1} y_l(\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}) \\
&\quad - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \log(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]) \\
&= n_1\beta_0 + \beta_1 \sum_{l=1}^{n_1} y_{(i)1l} + \beta_2 \sum_{l=1}^{n_1} y_{(i)2l} + \beta_{12} \sum_{l=1}^{n_1} y_{(i)1l} y_{(i)2l} \\
&\quad - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \log(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}])
\end{aligned}
$$

Now the posterior likelihood estimates for β_0, β_1, β_2, and β_{12} may be obtained by solving the following likelihood estimating equations derived from the above log likelihood function. The likelihood estimating equations are:

$$
\begin{aligned}
\frac{\delta l_c(\beta)}{\delta \beta_0} &\approx n_1 - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \frac{\exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]}{(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}])} = 0 \\
\frac{\delta l_c(\beta)}{\delta \beta_1} &\approx \sum_{l=1}^{n_i} y_{(1)1l} - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \frac{y_{(1)1l} \exp[\beta_0 + \beta_1 y_{(i)1l} + t\beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]}{(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}])} = 0 \\
\frac{\delta l_c(\beta)}{\delta \beta_2} &\approx \sum_{l=1}^{n_i} y_{(1)2l} - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \frac{y_{(1)2l} \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]}{(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}])} = 0 \\
\frac{\delta l_c(\beta)}{\delta \beta_{12}} &\approx \sum_{l=l}^{n_i} y_{(1)1l} y_{(1)2l} \\
&\quad - \sum_{i=1}^{2}\sum_{l=1}^{n_i} \frac{y_{(1)1l} y_{(1)2l} \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}]}{(1 + \exp[\beta_0 + \beta_1 y_{(i)1l} + \beta_2 y_{(i)2l} + \beta_{12} y_{(i)1l} y_{(i)2l}])} = 0
\end{aligned}
$$

Let $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_{12}$ be the solutions of the above estimating equations for β_0, β_1, β_2, and β_{12}, respectively. Then, in terms of $n_{(i)k}$, the above four equations reduce to

$$\begin{aligned}
n_{(1)1}+n_{(1)2}+n_{(1)3}+n_{(1)4} &= \frac{(n_{(1)4}+n_{(2)4})e^{\hat{\beta}_0}}{1+e^{\hat{\beta}_0}}+\frac{(n_{(1)3}+n_{(2)3})e^{\hat{\beta}_0+\hat{\beta}_2}}{1+e^{\hat{\beta}_0+\hat{\beta}_2}}. \\
&\quad +\frac{(n_{(1)2}+n_{(2)2})e^{\hat{\beta}_0+\hat{\beta}_1}}{1+e^{\hat{\beta}_0+\hat{\beta}_1}}+\frac{(n_{(1)1}+n_{(2)1})e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}, \\
n_{(1)1}+n_{(1)2} &= \frac{(n_{(1)2}+n_{(2)2})e^{\hat{\beta}_0+\hat{\beta}_1}}{1+e^{\hat{\beta}_0+\hat{\beta}_1}}+\frac{(n_{(1)1}+n_{(2)1})e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}, \\
n_{(1)1}+n_{(1)3} &= \frac{(n_{(1)3}+n_{(2)3})e^{\hat{\beta}_0+\hat{\beta}_2}}{1+e^{\hat{\beta}_0+\hat{\beta}_2}}+\frac{(n_{(1)1}+n_{(2)1})e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}, \\
\text{and}\quad n_{(1)1} &= \frac{(n_{(1)1}+n_{(2)1})e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_{12}}}
\end{aligned}$$

respectively.

Solving these equations for β's, we get

$$\begin{aligned}
\hat{\beta}_0 &= \log\left[\frac{n_{(1)4}}{n_{(2)4}}\right] \\
\hat{\beta}_1 &= \log\left[\frac{n_{(1)2}n_{(2)4}}{n_{(2)2}n_{(1)4}}\right] \\
\hat{\beta}_2 &= \log\left[\frac{n_{(1)3}n_{(2)4}}{n_{(2)3}n_{(1)4}}\right] \\
\hat{\beta}_{12} &= \log\left[\frac{n_{(1)1}n_{(1)4}}{n_{(1)2}n_{(1)3}}\Big/\frac{n_{(2)1}n_{(2)4}}{n_{(2)2}n_{(2)3}}\right] \qquad (2.3.19)
\end{aligned}$$

Consequently, by using the above $\hat{\beta}$'s in the classification criteria

$$\log\left[\frac{\hat{f}_1(y_1,y_2)}{\hat{f}_2(y_1,y_2)}\right]=\hat{\beta}_0+\hat{\beta}_1y_1+\hat{\beta}_2y_2+\hat{\beta}_{12}y_1y_2\geq 0,$$

we obtain

$$n_{(1)k} \geq n_{(2)k} \quad k = 1, 2, 3, 4$$

as the classification criteria for the unknown parameter to classify an observation (y_1, y_2) of cell k to G_1.

Note that the value of β_{12} computed by (2.3.19) helps one to understand the association between y_1 and y_2, provided the linearity assumption is valid. We further note here that for d (> 2) binary variables case, if higher order intersection are omitted from the log-linear representation, then the classification rule will be different than that found in the basic multinomial approach.

Remark that although, in general, the multinomial approach discussed in Section 2.3.1 and 2.3.2 is not parametric for correlated binary data, it is however parametric in the independent set-up as in the latter set-up, the joint probability directly depends on the marginal probabilities. We discuss this independent case in brief, as follows.

2.3.3 Independent Binary case: A Parametric Approach

In the independent set-up $\mathbf{y}' = (y_1, y_2)$ is a vector of independent binary variables, each taking the value 1 or 0. It then follows that for given $\mathbf{y} \in G_i$, the probability distribution of $\mathbf{y}$ is given by

Cell	1	2	3	4
$\mathbf{y}$	(1,1)	(1,0)	(0,1)	(0,0)
$f(\mathbf{y})$	$p_{i1}p_{i2}$	$p_{i1}q_{i2}$	$q_{i1}p_{i2}$	$q_{i1}q_{i2}$

allowing one to express the cell probability $\theta_{(i)k}$ as a function of p_{i1} and p_{i2} as

$$\theta_{(i)k} = f_i(\mathbf{y}) = p_{i1}^{y_1}(1-p_{i1})^{1-y_1}p_{i2}^{y_2}(1-p_{i2})^{1-y_2},$$

for y in cell k $(k = 1, 2, 3, 4)$.

Clearly as we can express each cell probability as a parametric function, it can be treated as parametric model for independent binary data and the classification criterion (2.3.5) can be simplified by replacing $f_i(\mathbf{y})$ with $\theta_{(i)k}$.

So far we have discussed in general the non-parametric or semi-parametric classification rule for binary data, though independent binary is a special case of the parametric approach. Note, however, that an allocation procedure can not be distribution-free in a literal sense (cf. T. W. Anderson [2]). For if it were, then its error rates would not depend on the group distributions of the feature vector and would be constant even when all the group distributions were identical (by a continuity argument). Therefore a parametric approach, if we know the model, is always a better approach.

It is clear that the classification criterion discussed above does not take into account the specific nature of the correlation coefficient. It is, however, known that correlated binary data can well be modelled as a function of the structural correlation parameter. Bahadur [5] suggested modelling binary data based on the adjustment factor on the correlation structure. Prentice [27] and Sutradhar and Das [33] have also analyzed correlated binary data. If the data really follows this distribution then naturally one would be able to do efficient analysis as compared to the ordinary (without considering correlation parameter) method. The purpose of the thesis is to examine the effect of the specific correlation structure over the classification when no distribution involving correlation parameter is considered.

Chapter 3

Classification of A Bivariate Binary Observation: A Model Based Approach

In the non-parametric approach, kernel methods are used to classify a multivariate binary observation into one of the two groups. In this approach, a kernel measures the distances of a given observation from sample observations of group 1 as well as of group 2 and classifies the given observation to a group based on the minimum distance. In the semi-parametric approach, however, this classification problem is formulated in a multinomial set-up. More specifically, for a d-dimensional binary data, it is assumed that an observation falls into one of the 2^d-cells with a certain multinomial probability which is unspecified in general. As discussed in the previous chapter in the context of bivariate binary data, the classification decision is made by comparing the corresponding cell probabilities of the two populations.

The non-parametric or semi-parametric approaches, mentioned above, are traditionally considered as suitable classification approach when one encounters difficulty in modelling the specific joint probability for the bivariate or multivariate binary observations. But, as bivariate binary data analysis is quite an important topic and as there exists suitable probability modelling for the case, in this thesis, we propose to classify a bivariate binary observation based on such probability modelling.

For bivariate binary probability modelling we refer, for example, to the probability model considered by Prentice [27], and Sutradhar and Das [33] and describe the modelling of bivariate binary case in section 3.1. This model will be exploited to classify a given bivariate binary observation in subsequent sections. The advantage of modelling the joint probability as compared to the semi-parametric approach, will be demonstrated through a simulation study in section 3.4. This will be done by comparing the misclassification probability of such model based classification criteria with that based on semi-parametric approach.

3.1 Joint Probability Model

Suppose that $y = (y_1, y_2)'$ is a pair of correlated binary variables each taking the values 1 or 0. Let $y_j = 1$ with probability p_{ij} and $y_j = 0$ with probability $q_{ij} = 1 - p_{ij}$, if y comes from group G_i $(i, j = 1, 2)$. Assume that y_1 and y_2 have a common correlation ϕ in both the groups G_1 and G_2. Following

Prentice [27] (1988, p1037: see also Sutradhar and Das [33]), one may then use the joint probability density of y_1 and y_2 for the ith $(i = 1, 2)$ group as

$$f(y_1, y_2|G_i) = p_{i1}^{y_1} q_{i1}^{1-y_1} p_{i2}^{y_2} q_{i2}^{1-y_2} \left[1 + \phi \frac{(y_1 - p_{i1})(y_2 - p_{i2})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \tag{3.1.1}$$

It is interesting to observe that this joint probability yields , as expected, the proper binary marginal densities for y_1 and y_2. Also the parameter ϕ is the proper correlation coefficient between y_1 and y_2, which is, however, restricted by

$$\max\left(-\left[\frac{p_{i1}p_{i2}}{q_{i1}q_{i2}}\right]^{\frac{1}{2}}, -\left[\frac{q_{i1}q_{i2}}{p_{i1}p_{i2}}\right]^{\frac{1}{2}}\right) < \phi < \min\left(\left[\frac{p_{i2}q_{i1}}{p_{i1}q_{i2}}\right]^{\frac{1}{2}}, \left[\frac{p_{i1}q_{i2}}{p_{i2}q_{i1}}\right]^{\frac{1}{2}}\right) \tag{3.1.2}$$

Note here that this restriction on ϕ, derived from the joint probability distribution (3.1.1), is necessary for (3.1.1) to be a proper joint density.

To verify the binary marginal density, we compute

$$\begin{aligned}
f(y_1|G_i) &= \sum_{y_2=0}^{1} f(y_1, y_2|G_i) \\
&= p_{i1}^{y_1} q_{i1}^{1-y_1} q_{i2} \left[1 + \phi \frac{(y_1 - p_{i1})(-p_{i2})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] + p_{i1}^{y_1} q_{i1}^{1-y_1} p_{i2} \left[1 + \phi \frac{(y_1 - p_{i1})q_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&= p_{i1}^{y_1} q_{i1}^{1-y_1} \left[(p_{i2} + q_{i2}) - \phi \frac{p_{i2}q_{i2}(y_1 - p_{i1})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}} + \phi \frac{p_{i2}q_{i2}(y_1 - p_{i1})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&= p_{i1}^{y_1} q_{i1}^{1-y_1},
\end{aligned}$$

which is the probability density of the binary variable y_1. Similarly we can show that

$$f(y_2|G_i) = p_{i2}^{y_2} q_{i2}^{1-y_2}$$

Thus y_1 and y_2 are marginally binary random variables with

$$E(y_j|G_i) = p_{ij} \quad i = 1,2; \quad j = 1,2$$

$$\text{and } Var(y_j|G_i) = p_{ij}q_{ij} \quad i = 1,2; \quad j = 1,2$$

To verify whether ϕ is the proper correlation coefficient, we compute the covariance between y_1 and y_2 as

$$\begin{aligned}
Cov[(y_1 - p_{i1})(y_2 - p_{i2})|G_i] &= E[(y_1 - p_{i1})(y_2 - p_{i2})|G_i] \\
&= p_{i1}p_{i2}q_{i1}q_{i2}\left[1 + \phi\frac{p_{i1}p_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&\quad -p_{i1}q_{i2}p_{i2}q_{i1}\left[1 + \phi\frac{p_{i1}q_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&\quad -q_{i1}p_{i2}p_{i1}q_{i2}\left[1 + \phi\frac{q_{i1}p_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&\quad +q_{i1}q_{i2}q_{i1}q_{i2}\left[1 + \phi\frac{q_{i1}q_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\
&= \frac{p_{i1}p_{i2}q_{i1}q_{i2}}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\phi[p_{i1}p_{i2} + p_{i1}q_{i2} + q_{i1}p_{i2} + q_{i1}q_{i2}] \\
&= \sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}\phi[p_{i1}(p_{i2} + q_{i2}) + q_{i1}(p_{i2} + q_{i2})] \\
&= \sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}\phi[(p_{i1} + q_{i1}) + (p_{i2} + q_{i2})] \\
&= \phi\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}},
\end{aligned}$$

yielding the correlation

$$\rho_{y_1y_2|G_i} = \frac{Cov[(y_1 - p_{i1})(y_2 - p_{i2})]}{\sqrt{Var(y_1|G_i)Var(y_2|G_i)}} = \phi,$$

between y_1 and y_2. This correlation parameter is usually referred to as the structural correlation parameter.

The bivariate binary density (3.1.1) has its natural generalization to the multivariate case (cf. Bahadur [5]) where as compared to (3.1.2), it becomes necessary to put severe restrictions on the higher order correlation and interaction. The analysis based on this type multinomial binary case is, however, beyond the scope of this thesis.

3.2 Classification Criterion

The random vector $Y' = (Y_1, Y_2)$ of two correlated binary variables Y_1 and Y_2 can take the four possible values $(1,1)$, $(1,0)$, $(0,1)$, and $(0,0)$. Therefore given $y \in G_i, y$ falls in the multinomial cell k with certain probability, say, $\theta_{(i)k}$ $(k = 1,2,3,4)$ which is determined from the joint probability function defined by (3.1.1) for the specific cell.

Thus if an observation y to be classified, belongs to cell k, then the optimum classification rule, due to Welch (1939), that minimizes the total probability of misclassification is the following: Assign y to G_1 if

$$\frac{f_1(y)}{f_2(y)} = \frac{\theta_{(1)k}}{\theta_{(2)k}} > \frac{\pi_2}{\pi_1} \tag{3.2.2}$$

and to G_2 otherwise, where π_1 is the proportion in G_1 and $\pi_2 (= 1 - \pi_1)$ is the remaining proportion in G_2 in a population $\mathcal{P}$ with only two groups. Since $\theta_{(i)k}$ for $i = 1,2$ is defined following (3.1.1), for $\pi_1 = \pi_2$, this rule (3.2.2)

classifies an observation with $y = (y_1, y_2)$ to G_1 if

$$\phi > \frac{(-1)^{y_1+y_2}}{w_1 - w_2}\left[p_{21}^{y_1}q_{21}^{1-y_1}p_{22}^{y_2}q_{22}^{1-y_2} - p_{11}^{y_1}q_{11}^{1-y_1}p_{12}^{y_2}q_{12}^{1-y_2}\right]. \tag{3.2.3}$$

where $w_i = \sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}$ for $i = 1, 2$.

In practice, the parameters $p_{11}, p_{12}, p_{21}, p_{22}$ and ϕ are usually not known and have to be estimated from the sample data. These parameters may however be estimated either by the traditional maximum likelihood estimation method or by using the well-known marginal estimating equation approach. The estimation of these parameters by these two methods is discussed in the following section.

3.2.1 Estimation of Parameters

Suppose that we have n_1 observations from group G_1 and n_2 observations from group G_2 so that in total we have $n = n_1 + n_2$ observations. Of these n_i observations that come from $G_i (i = 1, 2)$, let $n_{(i)k}$ fall into cell k, that is, $n_i = \sum_{k=1}^{4} n_{(i)k}$. Let y_{ijl} be the lth observation of the j-th variable in the i-th group G_i. Then the likelihood function, based on the mixture sampling approach, is given by

$$\begin{aligned} L(p_{i1}, p_{i2}, \phi) &= \prod_{i=1}^{2}\prod_{l=1}^{n_i}\left[\prod_{j=1}^{2}\pi_i p_{ij}^{y_{ijl}}q_{ij}^{1-y_{ijl}}\right]\left[1 + \phi\frac{(y_{i1l} - p_{i1})(y_{i2l} - p_{i2})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\ &= \frac{1}{2^n}\prod_{i=1}^{2}\prod_{l=1}^{n_i}\left[\prod_{j=1}^{2}p_{ij}^{y_{ijl}}q_{ij}^{1-y_{ijl}}\right]\left[1 + \phi\frac{(y_{i1l} - p_{i1})(y_{i2l} - p_{i2})}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\right] \\ &\quad \text{when} \quad \pi_1 = \pi_2 = \frac{1}{2} \end{aligned} \tag{3.2.4}$$

Equation (3.2.4) is an important special case (when $\pi_1 = \pi_2 = \frac{1}{2}$) which may be re-written as

$$L(p_{i1}, p_{i2}, o) = \frac{1}{2^n} \prod_{i=1}^{2} \prod_{k=1}^{4} \theta_{(i)k}^{n_{(i)k}} \tag{3.2.5}$$

where following (3.1.1), $\theta_{(i)k}$'s $(k = 1, 2, 3, 4)$ are given by

$$\begin{aligned}
\theta_{(i)1} &= p_{i1}p_{i2} + o\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} = p_{i1}p_{i2}\left[1 + o\sqrt{\frac{q_{i1}q_{i2}}{p_{i1}p_{i2}}}\right] \\
\theta_{(i)2} &= p_{i1}q_{i2} - o\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} = p_{i1}q_{i2}\left[1 - o\sqrt{\frac{q_{i1}p_{i2}}{p_{i1}q_{i2}}}\right] \\
\theta_{(i)3} &= q_{i1}p_{i2} - o\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} = q_{i1}p_{i2}\left[1 - o\sqrt{\frac{p_{i1}q_{i2}}{q_{i1}p_{i2}}}\right] \\
\theta_{(i)4} &= q_{i1}q_{i2} + o\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} = q_{i1}q_{i2}\left[1 + o\sqrt{\frac{p_{i1}p_{i2}}{q_{i1}q_{i2}}}\right]
\end{aligned}$$

Following (3.2.5) the appropriate likelihood estimating equations for p_{i1}, p_{i2} and o are given by

$$\begin{aligned}
\frac{\delta l}{\delta p_{i1}}|_{\mu=\hat{\mu}} &= \frac{n_{(i)1} + n_{(i)2}}{\hat{p}_{i1}} - \frac{n_{(i)3} + n_{(i)4}}{\hat{q}_{i1}} - \frac{\hat{o}}{2\hat{p}_{i1}\hat{q}_{i1}}\left[\frac{n_{(i)1}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{q}_{i2}} + \sqrt{\hat{p}_{i1}\hat{p}_{i2}}}\right. \\
&\quad \left. + \frac{n_{(i)2}\sqrt{\hat{q}_{i1}\hat{p}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{p}_{i2}} + \sqrt{\hat{p}_{i1}\hat{q}_{i2}}} - \frac{n_{(i)3}\sqrt{\hat{p}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{q}_{i2}} + \sqrt{\hat{q}_{i1}\hat{p}_{i2}}} - \frac{n_{(i)4}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{p}_{i2}} + \sqrt{\hat{q}_{i1}\hat{q}_{i2}}}\right] = 0.
\end{aligned}$$

$$\begin{aligned}
\frac{\delta l}{\delta p_{i2}}|_{\mu=\hat{\mu}} &= \frac{n_{(i)1} + n_{(i)3}}{\hat{p}_{i2}} - \frac{n_{(i)2} + n_{(i)4}}{\hat{q}_{i2}} - \frac{\hat{o}}{2\hat{p}_{i2}\hat{q}_{i2}}\left[\frac{n_{(i)1}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{q}_{i2}} + \sqrt{\hat{p}_{i1}\hat{p}_{i2}}}\right. \\
&\quad \left. - \frac{n_{(i)2}\sqrt{\hat{q}_{i1}\hat{p}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{p}_{i2}} + \sqrt{\hat{p}_{i1}\hat{q}_{i2}}} + \frac{n_{(i)3}\sqrt{\hat{p}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{q}_{i2}} + \sqrt{\hat{q}_{i1}\hat{p}_{i2}}} - \frac{n_{(i)4}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{p}_{i2}} + \sqrt{\hat{q}_{i1}\hat{q}_{i2}}}\right] = 0.
\end{aligned}$$

$$\begin{aligned}
\frac{\delta l}{\delta o}|_{\mu=\hat{\mu}} &= \sum_{i=1}^{2} \frac{n_{(i)1}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{q}_{i2}} + \sqrt{\hat{p}_{i1}\hat{p}_{i2}}} + \sum_{i=1}^{2} \frac{n_{(i)2}\sqrt{\hat{q}_{i1}\hat{p}_{i2}}}{\hat{o}\sqrt{\hat{q}_{i1}\hat{p}_{i2}} - \sqrt{\hat{p}_{i1}\hat{q}_{i2}}} \\
&\quad + \sum_{i=1}^{2} \frac{n_{(i)3}\sqrt{\hat{p}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{q}_{i2}} + \sqrt{\hat{q}_{i1}\hat{p}_{i2}}} + \sum_{i=1}^{2} \frac{n_{(i)4}\sqrt{\hat{q}_{i1}\hat{q}_{i2}}}{\hat{o}\sqrt{\hat{p}_{i1}\hat{p}_{i2}} + \sqrt{\hat{q}_{i1}\hat{q}_{i2}}} = 0.
\end{aligned}$$

respectively, where l denotes the log of the likelihood function given in (3.2.5) and $\mu = (p_{11}, p_{12}, p_{21}, p_{22}, \phi)'$. It is clear from the above derivatives that the likelihood estimates of p_{i1}, p_{i2} and ϕ do not have any closed form. To obtain these estimates, one needs to solve the above three equations by using a complicated iteration technique, which we do not pursue in the thesis. We rather estimate these parameters by using the well-known estimating equations approach which we discuss in the following section.

3.2.1.1 Marginal Estimating Equation (MEE) Approach

Since the marginal distributions of y_1 and y_2 are binary, in order to estimate p_{i1} and p_{i2}, we can use the marginal estimating equation, based on the sample from the i-th group G_i $(i = 1, 2)$, given by

$$n_i^{-1} \sum_{l=1}^{n_i} D_i' V_i^{-1} S_{il} = 0; \quad i = 1, 2 \tag{3.2.6}$$

where

$$S_{il} = [y_{i1l} - E(y_{i1l}),\ y_{i2l} - E(y_{i2l})]' = [y_{i1l} - p_{i1},\ y_{i2l} - p_{i2}]'.$$

for all $l = 1, 2, \cdots, n_i$ and V_i is the covariance matrix defined by

$$\begin{aligned} V_i &= \begin{pmatrix} Var(y_{i1l}) & Cov(y_{i1l}, y_{i2l}) \\ Cov(y_{i2l}, y_{i1l}) & Var(y_{i2l}) \end{pmatrix} \\ &= \begin{pmatrix} p_{i1}q_{i1} & \phi\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} \\ \phi\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}} & p_{i2}q_{i2} \end{pmatrix} \end{aligned}$$

and D_i is the matrix of first derivatives

$$D_i = \begin{pmatrix} \frac{\delta p_{i1}}{\delta p_{i1}} & \frac{\delta p_{i2}}{\delta p_{i1}} \\ \frac{\delta p_{i1}}{\delta p_{i2}} & \frac{\delta p_{i2}}{\delta p_{i2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Note that in the present case this V_i matrix does not depend on the individuals. This makes the estimation much easier as in such cases the MEE reduces to the moments equations. A remark is that this simplicity will not hold if the binary logistic regression case is considered where the covariates may be different for different individuals ($l = 1, 2, \cdots, n_i$, for $i = 1, 2$).

Turning back to the solutions for p_{i1} and p_{i2}, we obtain from (3.2.6) that

$$\hat{p}_{i1} = \frac{n_{(i)1} + n_{(i)2}}{n_i} \text{ and } \hat{p}_{i2} = \frac{n_{(i)1} + n_{(i)3}}{n_i} \text{ for } i = 1, 2, \qquad (3.2.7)$$

$$\text{with } \hat{q}_{i1} = 1 - \hat{p}_{i1} \text{ and } \hat{q}_{i2} = 1 - \hat{p}_{i2}.$$

Note that these estimates of the marginal probabilities are in fact the same as their maximum likelihood estimates based on the basic multinomial approach discussed in section (3.2.1). More specially, in the basic multinomial approach, the cell probabilities are estimated by using the maximum likelihood method which subsequently produces the marginal probability given by (3.2.7). As mentioned in the previous chapter, the cell probability in such cases, however, does not have a specific form based on the association parameter involved. The estimate of the association parameter under the present approach is given below, which in turn, will yield the cell probabilities corresponding to the four cells.

Temporarily we denote the correlation for the ith $(i = 1, 2)$ group by ϕ_i $(i = 1, 2)$ and estimate this correlation parameter ϕ_i assuming p_{ij} $(i, j = 1, 2)$ are known, by using the method of moments as

$$\begin{aligned}\hat{\phi}_i &= \frac{1}{n_i}\sum_{l=1}^{n_i}\frac{(y_{1il}-p_{i1})(y_{2il}-p_{i2})}{\sqrt{Var(y_{1il})Var(y_{2il})}} \\ &= \frac{1}{\sqrt{p_{i1}q_{i1}p_{i2}q_{i2}}}\left[\frac{n_{(i)1}}{n_i} - \frac{n_{(i)1}+n_{(i)3}}{n_i}p_{i1} - \frac{n_{(i)1}+n_{(i)2}}{n_i}p_{i2} + p_{i1}p_{i2}\right]\end{aligned}$$

Next, since it has been assumed that the two groups have common correlation ϕ we estimate this correlation coefficient by pooling the information collected from two samples, as

$$\begin{aligned}\hat{\phi} &= \frac{1}{n_1+n_2}\left[\sum_{l=1}^{n_1}\frac{(y_{11l}-p_{11})(y_{21l}-p_{12})}{\sqrt{Var(y_{11l})Var(y_{21l})}} + \sum_{l=1}^{n_2}\frac{(y_{12l}-p_{21})(y_{22l}-p_{22})}{\sqrt{Var(y_{12l})Var(y_{22l})}}\right] \\ &= \frac{1}{n_1+n_2}[n_1\hat{\phi}_1 + n_2\hat{\phi}_2]. \qquad (3.2.8)\end{aligned}$$

It is easy to see that for known p_{ij} $(i, j = 1, 2)$ this $\hat{\phi}$ is an unbiased estimate of ϕ as

$$\begin{aligned}E[\hat{\phi}_i] &= E\left[\frac{1}{n_i}\sum_{l=1}^{n_i}\frac{(y_{1il}-p_{i1})(y_{2il}-p_{i2})}{\sqrt{Var(y_{1il})Var(y_{2il})}}\right] \\ &= \frac{1}{n_i}\sum_{l=1}^{n_i}\frac{E(y_{1il}-p_{i1})(y_{2il}-p_{i2})}{\sqrt{Var(y_{1il})Var(y_{2il})}} \\ &= \phi_i,\end{aligned}$$

yielding

$$E[\hat{\phi}] = \phi \quad \text{by (3.2.8)} \quad \text{as} \quad \phi_1 = \phi_2 = \phi.$$

This moment estimate $\hat{\phi}$ is also consistent for ϕ. Following (3.1.2), this $\hat{\phi}$ should satisfy the restriction

$$\hat{\phi}_l < \hat{\phi} < \hat{\phi}_u$$

where

$$\hat{\phi}_l = \max\left(-\left[\frac{\hat{p}_{i1}\hat{p}_{i2}}{\hat{q}_{i1}\hat{q}_{i2}}\right]^{\frac{1}{2}}, \quad -\left[\frac{\hat{q}_{i1}\hat{q}_{i2}}{\hat{p}_{i1}\hat{p}_{i2}}\right]^{\frac{1}{2}}\right) \quad \text{and}$$

$$\hat{\phi}_u = \min\left(\left[\frac{\hat{p}_{i2}\hat{q}_{i1}}{\hat{p}_{i1}\hat{q}_{i2}}\right]^{\frac{1}{2}}, \quad \left[\frac{\hat{p}_{i1}\hat{q}_{i2}}{\hat{p}_{i2}\hat{q}_{i1}}\right]^{\frac{1}{2}}\right).$$

3.2.2 Performance of the Proposed Estimates: A Simulation Experiment

To examine the performance of the proposed estimates of the marginal probabilities p_{11}, p_{12}, p_{21}, and p_{22}, and the structural correlation parameter ϕ (discussed in the previous section), we conducted a simulation study as in the following.

Using the proposed density (3.1.1) for the i-th group G_i, we have the conditional distribution of y_{i2} given $y_{i1} = 0$ as

$$f(y_{i2}|y_{i2} = 0) = \frac{f(y_{i1}, y_{i2})}{f(y_{i1} = 0)} = p_{i2}\left[1 - \phi\sqrt{\frac{p_{i1}q_{i2}}{q_{i1}p_{i2}}}\right] \tag{3.2.9}$$

Similarly

$$f(y_{i2}|y_{i2} = 1) = \frac{f(y_{i1}, y_{i2})}{f(y_{i1} = 1)} = p_{i2}\left[1 + \phi\sqrt{\frac{q_{i1}q_{i2}}{p_{i1}p_{i2}}}\right] \tag{3.2.10}$$

In generating a correlated binary sample of size n_i from group G_i we use the following steps:

1. Generate binary y_{i1} with probability p_{i1}.

2. If $y_{i1} = 0$, then generate binary y_{i2} with probability given in (3.2.9) for a given ϕ and if $y_{i1} = 1$, then generate binary y_{i2} with probability given in (3.2.10) for the same ϕ.

3. Continue step 1-2 n_i times.

For various choices of p_{11}, p_{12} and p_{21}, p_{22}, as well as ϕ, we generate two bivariate samples of sizes n_1 and n_2 respectively. More specifically we have selected four different combinations of $(n_1, n_2) \equiv\{(25,20),(40,30),(50,40), (100,100)\}$ and three different choices of ϕ satisfying (3.1.2) under each of the three combinations of $(p_{11}, p_{12}) \equiv\{(0.10,0.10), (0.10,0.70), (0.50,0.30)\}$ and two different combinations of $(p_{21}, p_{22}) \equiv\{(0.50,0.70), (0.10,0.30)\}$.

We carry out 5000 simulations. Under each simulation, we estimate the parameters p_{ij} $(i, j = 1, 2)$ using the formula (3.2.7) and the structural correlation parameter ϕ by (3.2.8) and finally we compute the values for the parameters averaging the 5000 simulated estimates. The results are shown in Table 3.1- 3.6. Note that in each table we have also shown an effective number of simulation size, which we calculated based on the number of successful simulations depending on the computation of $\hat{\phi}$. To be more specific, the calculation for $\hat{\phi}$ fails if either $\hat{p}_{ij} = 0$ or $\hat{p}_{ij} = 1$. Any simulation yielding these estimated parameters is referred to as an unsuccessful simulation. The effective number of simulations is then the difference between the total

attempted simulations and the number of unsuccessful simulations. The estimates based on the effective simulation size are consequently reported in the second row for each ϕ. The results shown under 5000 simulations were computed by replacing $\hat{p}_{ij} = 0$ and $\hat{p}_{ij} = 1$ with $\hat{p}_{ij} = 0.02$ and $\hat{p}_{ij} = 0.98$, respectively.

It is clear from Table 3.1- 3.6 that as the sample size increases, the marginal probability estimates as well as ϕ estimate get very close to the true parameter values. More specifically, the large sample sizes yield significant gain in the estimation of ϕ. For example, when $n_1 = 25$, $n_2 = 20$ in Table 3.1 the absolute bias in estimating $\phi = 0.25$ is 0.0175 where for $n_1 = 100$, $n_2 = 100$ the absolute bias is 0.0006 which is very smaller. Also the standard errors estimates are found to be small and are not reported in the table. In the next section, these parameter estimates are used in the appropriate classification function in order to compare the misclassification rate of this procedure with that of the basic multinomial approach.

Table 3.1: Simulated Estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.10$, $p_{12} = 0.10$, $p_{21} = 0.50$, $p_{22} = 0.70$.

(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25,20)	-0.10	5000	0.1013	0.1011	0.4985	0.7002	-0.0903
		4303	0.1069	0.1070	0.5021	0.6945	-0.0908
	0.25	5000	0.1013	0.1016	0.4985	0.6996	0.2325
		4343	0.1094	0.1095	0.5034	0.6967	0.2530
	0.55	5000	0.1013	0.1020	0.4985	0.6984	0.5054
		4408	0.1099	0.1108	0.5037	0.7014	0.5435
(40,30)	-0.10	5000	0.1005	0.1002	0.4991	0.6985	-0.0927
		4849	0.1016	0.1011	0.5000	0.6976	-0.0929
	0.25	5000	0.1005	0.1011	0.4991	0.6979	0.2443
		4850	0.1021	0.1027	0.5003	0.6974	0.2490
	0.55	5000	0.1005	0.1008	0.4991	0.6997	0.5261
		4855	0.1024	0.1027	0.5006	0.7005	0.5352
(50,40)	-0.10	5000	0.1006	0.1007	0.4994	0.6985	-0.0931
		4957	0.1010	0.1009	0.4996	0.6983	-0.0931
	0.25	5000	0.1006	0.1010	0.4994	0.6984	0.2457
		4954	0.1011	0.1014	0.4998	0.6982	0.2470
	0.55	5000	0.1006	0.1011	0.4994	0.7000	0.5365
		4953	0.1011	0.1018	0.4998	0.7002	0.5393
(100,100)	-0.10	5000	0.1001	0.1001	0.4993	0.6987	-0.0957
		5000	0.1001	0.1001	0.4993	0.6987	-0.0957
	0.25	5000	0.1001	0.1004	0.4993	0.6989	0.2494
		5000	0.1001	0.1004	0.4993	0.6989	0.2494
	0.55	5000	0.1001	0.1001	0.4993	0.7000	0.5451
		5000	0.1001	0.1001	0.4993	0.7000	0.5451

Table 3.2: Simulated estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.10$, $p_{12} = 0.70$, $p_{21} = 0.50$, $p_{22} = 0.70$.

		Parameters	0.10	0.70	0.50	0.70	
(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25, 20)	-0.40	5000	0.1013	0.7015	0.4985	0.6993	-0.3796
		4635	0.1077	0.6966	0.5027	0.6974	-0.3831
	-0.25	5000	0.1013	0.7010	0.4985	0.6992	-0.2469
		4635	0.1077	0.6981	0.5027	0.6980	-0.2461
	0.10	5000	0.1013	0.6994	0.4985	0.6996	0.0797
		4635	0.1077	0.7005	0.5027	0.6999	0.0929
(40, 30)	-0.40	5000	0.1005	0.6999	0.4991	0.6991	-0.3861
		4920	0.1018	0.6990	0.5001	0.6989	-0.3871
	-0.25	5000	0.1005	0.6997	0.4991	0.6988	-0.2447
		4920	0.1018	0.6993	0.5001	0.6987	-0.2446
	0.10	5000	0.1005	0.6988	0.4991	0.6975	0.0920
		4920	0.1018	0.6993	0.5001	0.6977	0.0949
(50, 40)	-0.40	5000	0.1006	0.6989	0.4994	0.6986	-0.3893
		4976	0.1010	0.6987	0.4997	0.6986	-0.3895
	-0.25	5000	0.1006	0.6990	0.4994	0.6985	-0.2469
		4976	0.1010	0.6989	0.4997	0.6985	-0.2469
	0.10	5000	0.1006	0.6986	0.4994	0.6981	0.0958
		4976	0.1010	0.6988	0.4997	0.6982	0.0966
(100,100)	-0.40	5000	0.1001	0.6993	0.4993	0.6985	-0.3964
		5000	0.1001	0.6993	0.4993	0.6985	-0.3964
	-0.25	5000	0.1001	0.6992	0.4993	0.6987	-0.2489
		5000	0.1001	0.6992	0.4993	0.6987	-0.2489
	0.10	5000	0.1001	0.6983	0.4993	0.6983	0.0988
		5000	0.1001	0.6983	0.4993	0.6983	0.0988

Table 3.3: Simulated estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.50$, $p_{12} = 0.30$, $p_{21} = 0.50$, $p_{22} = 0.70$.

		Parameters	0.50	0.30	0.50	0.70	
(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25, 20)	-0.55	5000	0.4973	0.3021	0.4985	0.7012	-0.5259
		5000	0.4973	0.3021	0.4985	0.7012	-0.5259
	0.25	5000	0.4973	0.2999	0.4985	0.6996	0.2479
		5000	0.4973	0.2999	0.4985	0.6996	0.2479
	0.55	5000	0.4973	0.2986	0.4985	0.6984	0.5268
		5000	0.4973	0.2986	0.4985	0.6984	0.5268
(40, 30)	-0.55	5000	0.4984	0.3007	0.4991	0.7014	-0.5351
		5000	0.4984	0.3007	0.4991	0.7014	-0.5351
	0.25	5000	0.4984	0.3008	0.4991	0.6979	0.2497
		5000	0.4984	0.3008	0.4991	0.6979	0.2497
	0.55	5000	0.4984	0.2984	0.4991	0.6997	0.5351
		5000	0.4984	0.2984	0.4991	0.6997	0.5351
(50, 40)	-0.55	5000	0.4994	0.2995	0.4994	0.7012	-0.5386
		5000	0.4994	0.2995	0.4994	0.7012	-0.5386
	0.25	5000	0.4994	0.3012	0.4994	0.6984	0.2498
		5000	0.4994	0.3012	0.4994	0.6984	0.2498
	0.55	5000	0.4994	0.2991	0.4994	0.7000	0.5390
		5000	0.4994	0.2991	0.4994	0.7000	0.5390
(100,100)	-0.55	5000	0.4993	0.3000	0.4993	0.7018	-0.5456
		5000	0.4993	0.3000	0.4993	0.7018	-0.5456
	0.25	5000	0.4993	0.3013	0.4993	0.6989	0.2512
		5000	0.4993	0.3013	0.4993	0.6989	0.2512
	0.55	5000	0.4993	0.2982	0.4993	0.7000	0.5456
		5000	0.4993	0.2982	0.4993	0.7000	0.5456

Table 3.4: Simulated estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.10$, $p_{12} = 0.10$, $p_{21} = 0.10$, $p_{22} = 0.30$.

		Parameters	0.10	0.10	0.10	0.30	
(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25, 20)	-0.10	5000	0.1013	0.1011	0.1029	0.3005	-0.0840
		4069	0.1102	0.1070	0.1138	0.3044	-0.0884
	0.30	5000	0.1013	0.1015	0.1029	0.2993	0.2538
		4121	0.1126	0.1105	0.1161	0.3085	0.2967
	0.50	5000	0.1013	0.1019	0.1029	0.2994	0.4213
		4170	0.1131	0.1121	0.1164	0.3140	0.4890
(40, 30)	-0.10	5000	0.1005	0.1002	0.1017	0.3021	-0.0872
		4709	0.1034	0.1010	0.1053	0.3022	-0.0884
	0.30	5000	0.1005	0.1014	0.1017	0.3013	0.2804
		4706	0.1041	0.1036	0.1061	0.3037	0.2931
	0.50	5000	0.1005	0.1009	0.1017	0.3015	0.4543
		4723	0.1041	0.1037	0.1060	0.3058	0.4754
(50, 40)	-0.10	5000	0.1006	0.1007	0.1007	0.3019	-0.0886
		4915	0.1016	0.1008	0.1018	0.3017	-0.0889
	0.30	5000	0.1006	0.1009	0.1007	0.3009	0.2864
		4912	0.1018	0.1016	0.1020	0.3015	0.2902
	0.50	5000	0.1006	0.1012	0.1007	0.3012	0.4685
		4905	0.1018	0.1023	0.1020	0.3025	0.4758
(100,100)	-0.10	5000	0.1001	0.1001	0.1001	0.3017	-0.0940
		5000	0.1001	0.1001	0.1001	0.3017	-0.0940
	0.30	5000	0.1001	0.1003	0.1001	0.3007	0.2956
		5000	0.1001	0.1003	0.1001	0.3007	0.2956
	0.50	5000	0.1001	0.1000	0.1001	0.3009	0.4823
		4999	0.1001	0.1000	0.1001	0.3009	0.4824

Table 3.5: Simulated estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.10$. $p_{12} = 0.70$. $p_{21} = 0.10$. $p_{22} = 0.30$.

		Parameters	0.10	0.70	0.10	0.30	
(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25. 20)	-0.20	5000	0.1013	0.7008	0.1029	0.3002	-0.1861
		4384	0.1110	0.6964	0.1144	0.2966	-0.1890
	0.10	5000	0.1013	0.6994	0.1029	0.2998	0.0716
		4386	0.1110	0.7003	0.1145	0.3024	0.0955
	0.20	5000	0.1013	0.6997	0.1029	0.2995	0.1538
		4385	0.1110	0.7023	0.1145	0.3043	0.1863
(40. 30)	-0.20	5000	0.1005	0.6998	0.1017	0.3017	-0.1852
		4780	0.1036	0.6990	0.1055	0.3001	-0.1878
	0.10	5000	0.1005	0.6988	0.1017	0.3017	0.0896
		4780	0.1036	0.6996	0.1055	0.3022	0.0968
	0.20	5000	0.1005	0.6988	0.1017	0.3014	0.1767
		4780	0.1036	0.7001	0.1055	0.3027	0.1863
(50. 40)	-0.20	5000	0.1006	0.6990	0.1007	0.3016	-0.1879
		4934	0.1016	0.6989	0.1018	0.3010	-0.1888
	0.10	5000	0.1006	0.6986	0.1007	0.3012	0.0936
		4934	0.1016	0.6991	0.1018	0.3012	0.0956
	0.20	5000	0.1006	0.6988	0.1007	0.3011	0.1842
		4934	0.1016	0.6994	0.1018	0.3014	0.1872
(100.100)	-0.20	5000	0.1001	0.6992	0.1001	0.3015	-0.1924
		5000	0.1001	0.6992	0.1001	0.3015	-0.1924
	0.10	5000	0.1001	0.6983	0.1001	0.3015	0.0980
		5000	0.1001	0.6983	0.1001	0.3015	0.0980
	0.20	5000	0.1001	0.6984	0.1001	0.3008	0.1924
		5000	0.1001	0.6984	0.1001	0.3008	0.1924

Table 3.6: Simulated estimates of the marginal probabilities and correlation parameter for $p_{11} = 0.50$, $p_{12} = 0.30$, $p_{21} = 0.10$, $p_{22} = 0.30$.

		Parameters	0.50	0.30	0.10	0.30	
(n_1, n_2)	ϕ	Simulation #	$\hat{p}_{11}$	$\hat{p}_{12}$	$\hat{p}_{21}$	$\hat{p}_{22}$	$\hat{\phi}$
(25, 20)	-0.20	5000	0.4973	0.3000	0.1029	0.3002	-0.1845
		4384	0.5031	0.2998	0.1144	0.2966	-0.1889
	0.30	5000	0.4973	0.2997	0.1029	0.2993	0.2748
		4384	0.5031	0.3023	0.1145	0.3062	0.2945
	0.50	5000	0.4973	0.2979	0.1029	0.2994	0.4446
		4387	0.5031	0.3007	0.1145	0.3105	0.4722
(40, 30)	-0.20	5000	0.4984	0.3016	0.1017	0.3017	-0.1887
		4780	0.5005	0.3012	0.1055	0.3001	-0.1902
	0.30	5000	0.4984	0.3008	0.1017	0.3013	0.2881
		4780	0.5005	0.3012	0.1055	0.3033	0.2953
	0.50	5000	0.4984	0.2983	0.1017	0.3015	0.4654
		4780	0.5005	0.2991	0.1055	0.3050	0.4753
(50, 40)	-0.20	5000	0.4994	0.3019	0.1007	0.3016	-0.1895
		4934	0.5002	0.3016	0.1018	0.3010	-0.1900
	0.30	5000	0.4994	0.3013	0.1007	0.3009	0.2921
		4934	0.5002	0.3013	0.1018	0.3013	0.2941
	0.50	5000	0.4994	0.2991	0.1007	0.3012	0.4731
		4934	0.5002	0.2997	0.1018	0.3021	0.4764
(100,100)	-0.20	5000	0.4993	0.3014	0.1001	0.3015	-0.1941
		5000	0.4993	0.3014	0.1001	0.3015	-0.1941
	0.30	5000	0.4993	0.3015	0.1001	0.3007	0.2982
		5000	0.4993	0.3015	0.1001	0.3007	0.2982
	0.50	5000	0.4993	0.2981	0.1001	0.3009	0.4837
		5000	0.4993	0.2981	0.1001	0.3009	0.4837

3.3 Performance of The Classification Rules: A Simulation Study:

As mentioned in chapter 2, in classification context, the basic multinomial approach uses the so-called optimum classification criteria which appears to be a function of only the cell counts irrespective of the model for the data. The purposes of the simulation study is to generate data, based on the model 3.1.1 and compare the performance of the basic multinomial approach (BMA) with that of the model based approach (MBA), where in the latter case, the classification criteria depends on the estimates of the parameters of the model including the structural correlation parameter.

To be more specific, for the cases with unknown parameters, the MBA classification rule (3.2.3) is given as: classify a given observation $y = (y_1, y_2)$ into the group G_1 if

$$\hat{\phi} > \frac{(-1)^{y_1+y_2}}{\hat{w}_1 - \hat{w}_2} \left[\hat{p}_{21}^{y_1}\hat{q}_{21}^{1-y_1}\hat{p}_{22}^{y_2}\hat{q}_{22}^{1-y_2} - \hat{p}_{11}^{y_1}\hat{q}_{11}^{1-y_1}\hat{p}_{12}^{y_2}\hat{q}_{12}^{1-y_2}\right]. \tag{3.3.11}$$

Or equivalently, if an observation y belongs to cell k $(k = 1, 2, 3, 4)$, then the optimum classification rule is to assign y to G_1 if

$$\hat{\theta}_{(1)k} > \hat{\theta}_{(2)k} \quad k = 1, 2, 3, 4 \tag{3.3.12}$$

where $\hat{\theta}_{(i)k}$ is the estimated cell probability under model (3.1.1) for $i = 1, 2$. For example for $k = 1$, we have

$$\hat{\theta}_{(i)1} = \hat{p}_{i1}\hat{p}_{i2} + \hat{\phi}\sqrt{\hat{p}_{i1}\hat{q}_{i1}\hat{p}_{i2}\hat{q}_{i2}},$$

for the ith ($i = 1, 2$) group under the model (3.1.1).

In BMA, the classification rule is to assign y to G_1 if

$$\tilde{\theta}_{(1)k} > \tilde{\theta}_{(2)k} \quad k = 1.2.3.4. \tag{3.3.13}$$

$\tilde{\theta}_{(i)k}$ being the maximum likelihood estimate of the multinomial cell probability $\theta_{(i)k}$, where unlike in the proposed approach, $\theta_{(i)k}$ does not have any specific structure, mainly in terms of marginal probabilities of y_1 and y_2 and their correlation.

It is now clear that to examine the performance of the classification rules (3.3.11) and (3.3.13) in classifying an individual with two correlated binary measurements y_1 and y_2 into one of the two groups, one needs to derive the distributions of these classification functions, which is extremely complicated. Consequently, we have chosen to examine their performances empirically as follows.

For a given set of p_{11}, p_{12}, p_{21}, p_{22}, and ϕ, we first compute all the cell probabilities under each group using model (3.1.1) and compare the respective cell probabilities of the two groups to determine the classification criterion to classify a new observation belonging to that cell, into any of the two groups G_1 and G_2. For example, for a particular choice of parameters, say, $\phi = 0.2$, and $p_{11} = 0.5, p_{12} = 0.7$, and $p_{21} = 0.3, p_{22} = 0.3$, we obtain the cell

probabilities for the first group as

$$\theta_{(1)1} = 0.13, \quad \theta_{(1)2} = 0.17, \quad \theta_{(1)3} = 0.17, \quad \theta_{(1)4} = 0.53$$

and the cell probabilities for the second group as

$$\theta_{(2)1} = 0.39, \quad \theta_{(2)2} = 0.11, \quad \theta_{(2)3} = 0.31, \quad \theta_{(2)4} = 0.19,$$

yielding

$$\theta_{(1)1} \leq \theta_{(2)1}, \quad \theta_{(1)2} \geq \theta_{(2)2}, \quad \theta_{(1)3} \leq \theta_{(2)3}, \quad \theta_{(1)4} \geq \theta_{(2)4}.$$

Now, according to the classification rule (3.2.3), any new observation that belongs to the cell (1,1) is to be assigned into group G_2, as $\theta_{(1)1} \leq \theta_{(2)1}$. Similarly as $\theta_{(1)2} \geq \theta_{(2)2}$, the same classification rule (3.2.3) leads to classify any new observation that belongs to the cell (1,0) into group G_1. The classification of the observation belonging to the other two cells may be similarly interpreted.

Next to compare the performance of the proposed Model-Based Approach (MBA) as compared to the Basic Multinomial Approach (BMA) we may generate two bivariate correlated binary samples of sizes n_1 and n_2 following the proposed model (3.1.1) and then compare the performance of the estimated classification rules under both the approaches in classifying the selected observation into the correct group. To be more specific, suppose that we generate two samples of sizes n_1 and n_2 based on the above selection of the parameters ($p_{11} = 0.5$, $p_{12} = 0.7$, and $p_{21} = 0.3$, $p_{22} = 0.3$). We then

estimate $\theta_{(1)k}$ and $\theta_{(2)k}$ for $k = 1, 2, 3, 4$ by MBA and BMA and examine, for example, whether the conditions $\hat{\theta}_{(1)1} \leq \hat{\theta}_{(2)1}$ and $\tilde{\theta}_{(1)1} \leq \tilde{\theta}_{(2)1}$ for cell 1 are satisfied. If any of the two methods fails to satisfy this condition, then the individual in question (with value (1,1)) will not be classified into G_2 leading to misclassification due to the method of estimation. Now, if this behavior of classification is repeatedly tested for say R times, the proportion of unsuccessful cases will lead to the probability of misclassification due to that particular method. Note that although the data is generated according to the bivariate binary distribution in (3.1.1) with structural correlation parameter ϕ, the classification rule based on BMA does not require the estimation of the ϕ parameter whereas the the classification rule based on MBA does require the estimation of this ϕ parameter.

As mentioned above, to check the classification performance of the two approaches, we carry out a Monte-Carlo experiment based on $R = 5000$ simulations. In each of the simulations, we generate two samples of correlated bivariate binary observations of sizes $(n_1, n_2) \equiv \{(100, 100), (200, 200)\}$ and 3 to 5 different choices of ϕ depending on the restriction in (3.1.2) under each of the several combinations of (p_{11}, p_{12}) and two different combinations of $(p_{11}, p_{12}) \equiv \{(0.50, 0.70), (0.10, 0.30)\}$. Under each simulation, we estimate the parameters p_{11}, p_{12}, p_{21}, p_{22}, and ϕ and hence $\theta_{(i)k}$ $(i = 1, 2;$ $k = 1, 2, 3, 4)$ based on MBA and we estimate $\theta_{(i)k}$ $(i = 1, 2;$ $k = 1, 2, 3, 4)$ by maximum likelihood method based on BMA. Next we compute the sim-

ulated probability of misclassification for each method on the basis of the estimated classification criteria in terms of these estimated cell probabilities $\hat{\theta}_{(i)k}$ $(i = 1, 2;\ k = 1, 2, 3, 4)$ and $\tilde{\theta}_{(i)k}$ $(i = 1, 2;\ k = 1, 2, 3, 4)$. The results are shown in Table 3.7-3.33 below. In all the Tables 3.7-3.33, columns 4 to 7 contain the number of misclassification cases for a new individual belonging to cells 1, 2, 3, and 4 respectively by both methods. The number in the 8th column represents the total of columns from 4 to 7 on that row and the last column exhibits the total probability of misclassification obtained by dividing the figure in the 8th column by 5000, the total number of simulations.

It is clear from these tables that in almost all cases the MBA is found to be better than the BMA in terms of probability of misclassification. Here one method is considered to be superior to the other when the probability of misclassification (PM) due to this particular method is less than that of the other method. For some specific combinations (see Table 3.11, Table 3.12, Table 3.16, Table 3.17, Table 3.21), the MBA is substantially better than BMA as the PM is considerably higher for the latter method. For example, when $n_1 = n_2 = 100$ and $\phi = 0.40$ in Table 3.12, the probability of misclassification based on MBA and BMA are 0.0498 and 0.1236 respectively indicating that MBA is far superior to the BMA in classifying a new observation to the correct group. Note that as the sample size increases, the probability of misclassification generally decreases for both the methods. But the probability of misclassification still remains higher for BMA as compared

to MBA. For example, when $n_1 = n_2 = 200$ and $\phi = 0.40$ it is clear from the same Table 3.12 that the probability of misclassification is 4.78% higher for BMA as compared to MBA.

Remark that when the cell probability of bivariate binary observations under any group is close to zero, or when the relative difference between the two corresponding cell probabilities of the two groups are negligible, the probability of misclassification is generally higher under both approaches. This is obvious as the performance of any classification rule depends on the fact about whether the two groups, into which an observation is to be classified, are well-separated. In these types of unusual situations, even the PM based on MBA can be worse as compared to that of BMA.

Table 3.7: Probability of misclassification (PM) for $p_{11} = 0.10$, $p_{12} = 0.10$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
0.10	100	MBA	0	485	0	0	485	0.0970
		BMA	0	614	0	0	614	0.1228
	200	MBA	0	184	0	0	184	0.0368
		BMA	0	308	0	0	308	0.0616
0.20	100	MBA	0	855	0	0	855	0.1710
		BMA	0	903	0	0	903	0.1806
	200	MBA	0	480	0	0	480	0.0960
		BMA	0	622	0	0	622	0.1244
0.30	100	MBA	0	1411	0	0	1411	0.2822
		BMA	0	1318	0	0	1318	0.2636
	200	MBA	0	1088	0	0	1088	0.2176
		BMA	0	1119	0	0	1119	0.2238

Table 3.8: Probability of misclassification (PM) for $p_{11} = 0.10$, $p_{12} = 0.30$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.10	100	MBA	0	30	531	0	561	0.1122
		BMA	0	109	575	0	684	0.1368
-0.10	200	MBA	0	1	179	0	180	0.0360
		BMA	0	12	216	0	228	0.0456
0.10	100	MBA	0	74	809	0	883	0.1766
		BMA	0	166	850	0	1016	0.2032
0.10	200	MBA	0	1	379	0	380	0.0760
		BMA	0	31	453	0	484	0.0968
0.20	100	MBA	0	101	886	0	987	0.1974
		BMA	0	72	946	0	1018	0.2036
0.20	200	MBA	0	10	469	0	479	0.0958
		BMA	0	9	575	0	584	0.1168
0.30	100	MBA	0	150	1016	0	1166	0.2332
		BMA	0	64	1067	0	1131	0.2262
0.30	200	MBA	0	19	580	0	599	0.1198
		BMA	0	8	713	0	721	0.1442
0.40	100	MBA	0	194	1164	0	1358	0.2716
		BMA	0	30	1254	0	1284	0.2568
0.40	200	MBA	0	30	745	0	775	0.1550
		BMA	0	0	889	0	889	0.1778

Table 3.9: Probability of misclassification (PM) for $p_{11} = 0.10$, $p_{12} = 0.50$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	0	283	0	283	0.0566
		BMA	0	23	305	0	328	0.0656
-0.20	200	MBA	0	0	66	0	66	0.0132
		BMA	0	0	87	0	87	0.0174
-0.10	100	MBA	0	1	212	0	213	0.0426
		BMA	0	45	217	0	262	0.0524
-0.10	200	MBA	0	0	44	0	44	0.0088
		BMA	0	2	55	0	57	0.0114
0.10	100	MBA	0	5	293	0	298	0.0596
		BMA	0	1	365	0	366	0.0732
0.10	200	MBA	0	0	62	0	62	0.0124
		BMA	0	0	98	0	98	0.0196
0.20	100	MBA	0	4	217	0	221	0.0442
		BMA	0	0	314	0	314	0.0628
0.20	200	MBA	0	0	32	0	32	0.0064
		BMA	0	0	76	0	76	0.0152

Table 3.10: Probability of misclassification (PM) for $p_{11} = 0.10$, $p_{12} = 0.90$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.30	100	MBA	0	0	0	1207	1207	0.2414
		BMA	0	0	0	973	973	0.1946
-0.30	200	MBA	0	0	0	799	799	0.1598
		BMA	0	0	0	699	699	0.1398
-0.20	100	MBA	0	0	0	657	657	0.1314
		BMA	0	0	0	560	560	0.1120
-0.20	200	MBA	0	0	0	249	249	0.0498
		BMA	0	0	0	267	267	0.0534
-0.10	100	MBA	0	0	0	286	286	0.0572
		BMA	0	0	0	317	317	0.0634
-0.10	200	MBA	0	0	0	70	70	0.0140
		BMA	0	0	0	97	97	0.0194

Table 3.11: Probability of misclassification (PM) for $p_{11} = 0.30$, $p_{12} = 0.30$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.30	100	MBA	0	561	21	0	582	0.1164
		BMA	0	729	77	0	806	0.1612
-0.30	200	MBA	0	194	0	0	194	0.0388
		BMA	0	342	3	0	345	0.0690
-0.20	100	MBA	0	515	36	0	551	0.1102
		BMA	0	718	100	0	818	0.1636
-0.20	200	MBA	0	163	1	0	164	0.0328
		BMA	0	365	5	0	370	0.0740
0.20	100	MBA	0	262	32	0	294	0.0588
		BMA	0	457	104	0	561	0.1122
0.20	200	MBA	0	50	1	0	51	0.0102
		BMA	0	167	10	0	177	0.0354
0.40	100	MBA	0	91	8	0	99	0.0198
		BMA	0	58	68	0	126	0.0252
0.40	200	MBA	0	4	0	0	4	0.0008
		BMA	0	6	3	0	9	0.0018
0.50	100	MBA	0	10	0	0	10	0.0020
		BMA	0	29	1	0	30	0.0060
0.50	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000

Table 3.12: Probability of misclassification (PM) for $p_{11} = 0.30$, $p_{12} = 0.70$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.50	100	MBA	0	67	3	0	70	0.0140
		BMA	0	112	6	0	118	0.0236
-0.50	200	MBA	0	0	0	0	0	0.0000
		BMA	0	10	0	0	10	0.0020
-0.40	100	MBA	0	155	9	85	249	0.0498
		BMA	51	565	2	0	618	0.1236
-0.40	200	MBA	0	13	0	6	19	0.0038
		BMA	5	253	0	0	258	0.0516
-0.10	100	MBA	0	0	0	0	0	0.0000
		BMA	0	12	0	0	12	0.0024
-0.10	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.20	100	MBA	0	0	0	25	25	0.0050
		BMA	0	0	0	60	60	0.0120
0.20	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	8	8	0.0016
0.30	100	MBA	0	0	0	97	97	0.0194
		BMA	1	0	0	108	109	0.0218
0.30	200	MBA	0	0	0	10	10	0.0020
		BMA	0	0	0	28	28	0.0056

Table 3.13: Probability of misclassification (PM) for $p_{11} = 0.30$, $p_{12} = 0.90$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.40	100	MBA	489	0	0	26	515	0.1030
		BMA	594	0	0	41	635	0.1270
-0.40	200	MBA	155	0	0	1	156	0.0312
		BMA	271	0	0	3	274	0.0548
-0.30	100	MBA	325	0	0	8	333	0.0666
		BMA	397	0	0	41	438	0.0876
-0.30	200	MBA	79	0	0	0	79	0.0158
		BMA	144	0	0	1	145	0.0290
-0.20	100	MBA	213	0	0	3	216	0.0432
		BMA	281	0	0	35	316	0.0632
-0.20	200	MBA	40	0	0	0	40	0.0080
		BMA	78	0	0	4	82	0.0164
-0.10	100	MBA	138	0	0	3	141	0.0282
		BMA	208	0	0	36	244	0.0488
-0.10	200	MBA	17	0	0	0	17	0.0034
		BMA	42	0	0	2	44	0.0088
0.10	100	MBA	46	0	0	1	47	0.0094
		BMA	72	0	0	17	89	0.0178
0.10	200	MBA	3	0	0	0	3	0.0006
		BMA	7	0	0	0	7	0.0014

Table 3.14: Probability of misclassification (PM) for $p_{11} = 0.50$, $p_{12} = 0.50$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.50	100	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
-0.50	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
-0.40	100	MBA	6	6	6	6	24	0.0048
		BMA	190	190	0	0	380	0.0760
-0.40	200	MBA	0	0	0	0	0	0.0000
		BMA	36	36	0	0	72	0.0144
0.10	100	MBA	10	10	10	10	40	0.0080
		BMA	0	0	282	282	564	0.1128
0.10	200	MBA	1	1	1	1	4	0.0008
		BMA	0	0	74	74	148	0.0296
0.40	100	MBA	3	3	3	3	12	0.0024
		BMA	0	0	187	187	374	0.0748
0.40	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	39	39	78	0.0156
0.50	100	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.50	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000

Table 3.15: Probability of misclassification (PM) for $p_{11} = 0.70$, $p_{12} = 0.30$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.50	100	MBA	15	0	0	119	134	0.0268
		BMA	27	0	1	85	113	0.0226
-0.50	200	MBA	0	0	0	12	12	0.0024
		BMA	2	0	0	10	12	0.0024
-0.40	100	MBA	11	0	0	200	211	0.0422
		BMA	69	0	0	229	298	0.0596
-0.40	200	MBA	0	0	0	24	24	0.0048
		BMA	5	0	0	51	56	0.0112
-0.10	100	MBA	121	0	0	640	761	0.1522
		BMA	121	0	0	614	735	0.1470
-0.10	200	MBA	11	0	0	258	269	0.0538
		BMA	16	0	0	270	286	0.0572
0.20	100	MBA	141	0	0	962	1103	0.2206
		BMA	138	0	0	949	1087	0.2174
0.20	200	MBA	15	0	0	545	560	0.1120
		BMA	15	0	0	556	571	0.1142
0.30	100	MBA	144	0	0	1054	1198	0.2396
		BMA	123	0	0	1054	1177	0.2354
0.30	200	MBA	19	0	0	646	665	0.1330
		BMA						

Table 3.16: Probability of misclassification (PM) for $p_{11} = 0.70$, $p_{12} = 0.70$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.30	100	MBA	10	664	112	12	798	0.1596
		BMA	70	749	173	145	1137	0.2274
-0.30	200	MBA	1	286	13	1	301	0.0602
		BMA	4	404	15	26	449	0.0898
-0.20	100	MBA	18	468	95	24	605	0.1210
		BMA	90	656	157	276	1179	0.2358
-0.20	200	MBA	2	148	10	1	161	0.0322
		BMA	9	337	19	72	437	0.0874
0.20	100	MBA	140	42	17	257	456	0.0912
		BMA	183	268	99	563	1113	0.2226
0.20	200	MBA	17	2	2	33	54	0.0108
		BMA	27	75	8	243	353	0.0706
0.40	100	MBA	252	59	11	488	810	0.1620
		BMA	232	101	56	629	1018	0.2036
0.40	200	MBA	29	4	0	158	191	0.0382
		BMA	33	11	6	287	337	0.0674
0.50	100	MBA	301	85	18	637	1041	0.2082
		BMA	265	47	56	680	1048	0.2096
0.50	200	MBA	54	6	0	265	325	0.0650
		BMA	46	3	3	323	375	0.0750

Table 3.17: Probability of misclassification for $p_{11} = 0.70$, $p_{12} = 0.90$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.10	100	MBA	0	10	584	0	594	0.1188
		BMA	0	85	601	3	689	0.1378
-0.10	200	MBA	0	0	217	0	217	0.0434
		BMA	0	6	271	0	277	0.0554
0.10	100	MBA	0	17	765	0	782	0.1564
		BMA	1	101	828	13	943	0.1886
0.10	200	MBA	0	0	364	0	364	0.0728
		BMA	0	12	444	0	456	0.0912
0.20	100	MBA	1	8	905	0	914	0.1828
		BMA	2	121	934	13	1070	0.2140
0.20	200	MBA	0	0	454	0	454	0.0908
		BMA	0	31	544	0	575	0.1150
0.30	100	MBA	5	22	1017	0	1044	0.2088
		BMA	1	119	1077	14	1211	0.2422
0.30	200	MBA	0	0	564	0	564	0.1128
		BMA	0	19	688	0	707	0.1414
0.40	100	MBA	7	68	1171	1	1247	0.2494
		BMA	4	100	1230	14	1348	0.2696
0.40	200	MBA	1	5	746	0	752	0.1504
		BMA	1	15	867	0	883	0.1766

Table 3.18: Probability of misclassification for $p_{11} = 0.90$, $p_{12} = 0.30$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.40	100	MBA	1180	0	0	262	1442	0.2884
		BMA	1287	0	0	165	1452	0.2904
-0.40	200	MBA	718	0	0	62	780	0.1560
		BMA	862	0	0	31	893	0.1786
-0.30	100	MBA	993	0	0	218	1211	0.2422
		BMA	1119	0	0	191	1310	0.2620
-0.30	200	MBA	563	0	0	37	600	0.1200
		BMA	706	0	0	45	751	0.1502
-0.20	100	MBA	862	0	0	189	1051	0.2102
		BMA	991	0	0	189	1180	0.2360
-0.20	200	MBA	448	0	0	26	474	0.0948
		BMA	553	0	0	44	597	0.1194
-0.10	100	MBA	869	0	0	147	1016	0.2032
		BMA	862	0	0	188	1050	0.2100
-0.10	200	MBA	412	0	0	24	436	0.0872
		BMA	453	0	0	33	486	0.0972
0.10	100	MBA	627	0	0	100	727	0.1454
		BMA	576	0	0	146	722	0.1444
0.10	200	MBA	226	0	0	7	233	0.0466
		BMA	241	0	0	26	267	0.0534

Table 3.19: Probability of misclassification for $p_{11} = 0.90$, $p_{12} = 0.50$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	235	0	0	10	245	0.0490
		BMA	323	1	0	25	349	0.0698
-0.20	200	MBA	40	0	0	0	40	0.0080
		BMA	90	0	0	1	91	0.0182
-0.10	100	MBA	306	0	0	9	315	0.0630
		BMA	384	0	0	33	417	0.0834
-0.10	200	MBA	69	0	0	0	69	0.0138
		BMA	118	0	0	1	119	0.0238
0.10	100	MBA	354	0	0	5	359	0.0718
		BMA	361	0	0	51	412	0.0824
0.10	200	MBA	79	0	0	0	79	0.0158
		BMA	103	0	0	3	106	0.0212
0.20	100	MBA	475	0	0	8	483	0.0966
		BMA	473	0	0	49	522	0.1044
0.20	200	MBA	142	0	0	0	142	0.0284
		BMA	161	0	0	4	165	0.0330

Table 3.20: Probability of misclassification for $p_{11} = 0.90$, $p_{12} = 0.90$, $p_{21} = 0.50$, $p_{22} = 0.70$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
0.10	100	MBA	0	263	0	0	263	0.0526
		BMA	0	376	0	0	376	0.0752
0.10	200	MBA	0	42	0	0	42	0.0084
		BMA	0	147	0	0	147	0.0294
0.20	100	MBA	0	573	0	0	573	0.1146
		BMA	0	650	0	0	650	0.1300
0.20	200	MBA	0	208	0	0	208	0.0416
		BMA	0	340	0	0	340	0.0680
0.30	100	MBA	0	1220	0	0	1220	0.2440
		BMA	0	1079	0	0	1079	0.2158
0.30	200	MBA	0	797	0	0	797	0.1594
		BMA	0	827	0	0	827	0.1654

Table 3.21: Probability of misclassification for $p_{11} = 0.10$, $p_{12} = 0.10$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
0.20	100	MBA	0	0	0	0	0	0.0000
		BMA	600	600	0	0	1200	0.2400
0.20	200	MBA	0	0	0	0	0	0.0000
		BMA	300	300	0	0	600	0.1200
0.40	100	MBA	0	0	0	0	0	0.0000
		BMA	203	203	0	0	406	0.0812
0.40	200	MBA	0	0	0	0	0	0.0000
		BMA	45	45	0	0	90	0.0180
0.50	100	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.50	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000

Table 3.22: Probability of misclassification for $p_{11} = 0.10$, $p_{12} = 0.70$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	59	59	59	59	236	0.0472
		BMA	0	0	86	86	172	0.0344
-0.20	200	MBA	2	2	2	2	8	0.0016
		BMA	0	0	10	10	20	0.0040
-0.10	100	MBA	46	46	46	46	184	0.0368
		BMA	0	0	80	80	160	0.0320
-0.10	200	MBA	2	2	2	2	8	0.0016
		BMA	0	0	8	8	16	0.0032
0.10	100	MBA	0	0	0	0	0	0.0000
		BMA	942	942	0	0	1884	0.3768
0.10	200	MBA	0	0	0	0	0	0.0000
		BMA	641	641	0	0	1282	0.2564
0.20	100	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.20	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.30	100	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
0.30	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000

Table 3.23: Probability of Misclassification for $p_{11} = 0.10$, $p_{12} = 0.90$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	0	0	0	0	0.0000
		BMA	1	1	0	0	2	0.0004
-0.20	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
-0.10	100	MBA	0	0	0	0	0	0.0000
		BMA	44	44	0	0	88	0.0176
-0.10	200	MBA	0	0	0	0	0	0.0000
		BMA	3	3	0	0	6	0.0012
0.10	100	MBA	0	0	0	0	0	0.0000
		BMA	212	212	0	0	424	0.0848
0.10	200	MBA	0	0	0	0	0	0.0000
		BMA	50	50	0	0	100	0.0200

Table 3.24: Probability of Misclassification for $p_{11} = 0.30$, $p_{12} = 0.30$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	0	194	0	194	0.0388
		BMA	0	0	187	0	187	0.0374
-0.20	200	MBA	0	0	37	0	37	0.0074
		BMA	0	0	42	0	42	0.0084
-0.10	100	MBA	0	0	25	0	25	0.0050
		BMA	0	0	59	0	59	0.0118
-0.10	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	2	0	2	0.0004
0.20	100	MBA	0	0	53	3	56	0.0112
		BMA	136	6	0	0	142	0.0284
0.20	200	MBA	0	0	3	0	3	0.0006
		BMA	25	0	0	0	25	0.0050
0.40	100	MBA	0	0	6	35	41	0.0082
		BMA	20	0	0	9	29	0.0058
0.40	200	MBA	0	0	0	2	2	0.0004
		BMA	1	0	0	0	1	0.0002
0.50	100	MBA	0	0	0	70	70	0.0140
		BMA	15	0	0	25	40	0.0080
0.50	200	MBA	0	0	0	2	2	0.0004
		BMA	1	0	0	1	2	0.0004

Table 3.25: Probability of Misclassification for $p_{11} = 0.30$, $p_{12} = 0.50$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	3	410	0	413	0.0826
		BMA	0	9	564	0	573	0.1146
-0.20	200	MBA	0	0	133	0	133	0.0266
		BMA	0	1	227	0	228	0.0456
-0.10	100	MBA	0	5	488	0	493	0.0986
		BMA	0	17	657	0	674	0.1348
-0.10	200	MBA	0	0	158	0	158	0.0316
		BMA	0	1	287	0	288	0.0576
0.20	100	MBA	0	7	366	0	373	0.0746
		BMA	0	12	218	0	230	0.0460
0.20	200	MBA	0	0	104	0	104	0.0208
		BMA	0	1	62	0	63	0.0126
0.40	100	MBA	0	27	557	0	584	0.1168
		BMA	0	0	425	0	425	0.0850
0.40	200	MBA	0	1	215	0	216	0.0432
		BMA	0	0	182	0	182	0.0364
0.50	100	MBA	0	31	690	0	721	0.1442
		BMA	0	0	632	0	632	0.1264
0.50	200	MBA	0	0	317	0	317	0.0634
		BMA	0	0	300	0	300	0.0600

Table 3.26: Probability of Misclassification for $p_{11} = 0.30$, $p_{12} = 0.70$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

φ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	570	1	0	571	0.1142
		BMA	0	744	3	0	747	0.1494
-0.20	200	MBA	0	224	0	0	224	0.0448
		BMA	0	442	0	0	442	0.0884
-0.10	100	MBA	0	775	5	0	780	0.1560
		BMA	0	1006	10	0	1016	0.2032
-0.10	200	MBA	0	403	0	0	403	0.0806
		BMA	0	723	0	0	723	0.1446
0.20	100	MBA	0	1966	9	0	1975	0.3950
		BMA	0	1705	20	0	1725	0.3450
0.20	200	MBA	0	1826	0	0	1826	0.3652
		BMA	0	1636	1	0	1637	0.3274

Table 3.27: Probability of Misclassification for $p_{11} = 0.30$, $p_{12} = 0.90$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

o	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	280	0	0	280	0.0560
		BMA	0	459	0	0	459	0.0918
-0.20	200	MBA	0	54	0	0	54	0.0108
		BMA	0	189	0	0	189	0.0378
-0.10	100	MBA	0	208	0	0	208	0.0416
		BMA	0	512	0	0	512	0.1024
-0.10	200	MBA	0	23	0	0	23	0.0046
		BMA	0	221	0	0	221	0.0442
0.10	100	MBA	0	94	0	0	94	0.0188
		BMA	0	205	0	0	205	0.0410
0.10	200	MBA	0	8	0	0	8	0.0016
		BMA	0	44	0	0	44	0.0088

Table 3.28: Probability of Misclassification for $p_{11} = 0.50$, $p_{12} = 0.10$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	205	0	0	1	206	0.0412
		BMA	0	0	0	3	3	0.0006
-0.20	200	MBA	35	0	0	0	35	0.0070
		BMA	0	0	0	0	0	0.0000
-0.10	100	MBA	359	0	0	1	360	0.0720
		BMA	475	0	0	3	478	0.0956
-0.10	200	MBA	98	0	0	0	98	0.0196
		BMA	250	0	0	0	250	0.0500
0.10	100	MBA	682	0	0	5	687	0.1374
		BMA	927	0	0	8	935	0.1870
0.10	200	MBA	264	0	0	0	264	0.0528
		BMA	615	0	0	0	615	0.1230
0.20	100	MBA	798	0	0	8	806	0.1612
		BMA	1144	0	0	6	1150	0.2300
0.20	200	MBA	388	0	0	0	388	0.0776
		BMA	863	0	0	0	863	0.1726
0.30	100	MBA	884	0	0	10	894	0.1788
		BMA	1178	0	0	13	1191	0.2382
0.30	200	MBA	488	0	0	0	488	0.0976
		BMA	920	0	0	1	921	0.1842

Table 3.29: Probability of Misclassification for $p_{11} = 0.50$, $p_{12} = 0.50$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.10	100	MBA	0	0	2136	0	2136	0.4272
		BMA	0	0	2090	0	2090	0.4180
-0.10	200	MBA	0	0	2069	0	2069	0.4138
		BMA	0	0	2075	0	2075	0.4150
0.20	100	MBA	0	0	833	0	833	0.1666
		BMA	0	0	648	0	648	0.1296
0.20	200	MBA	0	0	430	0	430	0.0860
		BMA	0	0	297	0	297	0.0594
0.40	100	MBA	0	0	228	0	228	0.0456
		BMA	0	0	116	0	116	0.0232
0.40	200	MBA	0	0	36	0	36	0.0072
		BMA	0	0	10	0	10	0.0020
0.50	100	MBA	0	0	63	0	63	0.0126
		BMA	0	0	18	0	18	0.0036
0.50	200	MBA	0	0	4	0	4	0.0008
		BMA	0	0	3	0	3	0.0006

Table 3.30: Probability of Misclassification for $p_{11} = 0.50$, $p_{12} = 0.70$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	29	411	0	440	0.0880
		BMA	0	61	461	0	522	0.1044
-0.20	200	MBA	0	1	107	0	108	0.0216
		BMA	0	5	137	0	142	0.0284
-0.10	100	MBA	0	30	531	0	561	0.1122
		BMA	0	109	575	0	684	0.1368
-0.10	200	MBA	0	1	179	0	180	0.0360
		BMA	0	12	216	0	228	0.0456
0.20	100	MBA	0	101	886	0	987	0.1974
		BMA	0	72	946	0	1018	0.2036
0.20	200	MBA	0	10	469	0	479	0.0958
		BMA	0	9	575	0	584	0.1168
0.30	100	MBA	0	150	1016	0	1166	0.2332
		BMA	0	64	1067	0	1131	0.2262
0.30	200	MBA	0	19	580	0	599	0.1198
		BMA	0	8	713	0	721	0.1442
0.40	100	MBA	0	194	1164	0	1358	0.2716
		BMA	0	30	1254	0	1284	0.2568
0.40	200	MBA	0	30	745	0	775	0.1550
		BMA	0	0	889	0	889	0.1778

Table 3.31: Probability of Misclassification for $p_{11} = 0.50$, $p_{12} = 0.90$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	1450	6	0	1456	0.2912
		BMA	0	1393	6	0	1399	0.2798
-0.20	200	MBA	0	1074	1	0	1075	0.2150
		BMA	0	1160	0	0	1160	0.2320
-0.10	100	MBA	0	1304	7	0	1311	0.2622
		BMA	0	1383	8	0	1391	0.2782
-0.10	200	MBA	0	907	1	0	908	0.1816
		BMA	0	1096	1	0	1097	0.2194
0.10	100	MBA	0	946	12	0	958	0.1916
		BMA	0	949	26	0	975	0.1950
0.10	200	MBA	0	524	0	0	524	0.1048
		BMA	0	682	0	0	682	0.1364
0.20	*100*	MBA	*0*	*748*	*12*	*0*	*760*	*0.1520*
		BMA	0	557	24	0	581	0.1162
0.20	200	MBA	0	320	1	0	321	0.0642
		BMA	0	270	2	0	272	0.0544
0.30	100	MBA	0	559	15	0	574	0.1148
		BMA	0	147	35	0	182	0.0364
0.30	200	MBA	0	174	0	0	174	0.0348
		BMA	0	41	2	0	43	0.0086

Table 3.32: Probability of Misclassification for $p_{11} = 0.70$, $p_{12} = 0.10$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	31	0	0	0	31	0.0062
		BMA	23	0	0	0	23	0.0046
-0.20	200	MBA	0	0	0	0	0	0.0000
		BMA	0	0	0	0	0	0.0000
-0.10	100	MBA	65	0	0	0	65	0.0130
		BMA	200	0	0	0	200	0.0400
-0.10	200	MBA	3	0	0	0	3	0.0006
		BMA	42	0	0	0	42	0.0084
0.10	100	MBA	283	0	0	0	283	0.0566
		BMA	521	0	0	0	521	0.1042
0.10	200	MBA	63	0	0	0	63	0.0126
		BMA	225	0	0	0	225	0.0450

Table 3.33: Probability of Misclassification for $p_{11} = 0.70$, $p_{12} = 0.70$, $p_{21} = 0.10$, $p_{22} = 0.30$ based on MBA and BMA

ϕ	n	Model	Cell-1	Cell-2	Cell-3	Cell-4	Total	PM
-0.20	100	MBA	0	17	1234	0	1251	0.2502
		BMA	0	20	1140	0	1160	0.2320
-0.20	200	MBA	0	0	817	0	817	0.1634
		BMA	0	0	821	0	821	0.1642
-0.10	100	MBA	0	11	1051	0	1062	0.2124
		BMA	0	14	981	0	995	0.1990
-0.10	200	MBA	0	0	660	0	660	0.1320
		BMA	0	0	649	0	649	0.1298
0.20	100	MBA	0	5	474	0	479	0.0958
		BMA	0	5	519	0	524	0.1048
0.20	200	MBA	0	0	169	0	169	0.0338
		BMA	0	0	215	0	215	0.0430
0.40	100	MBA	0	1	156	0	157	0.0314
		BMA	0	1	120	0	121	0.0242
0.40	200	MBA	0	0	23	0	23	0.0046
		BMA	0	0	24	0	24	0.0048
0.50	100	MBA	0	0	47	0	47	0.0094
		BMA	0	0	63	0	63	0.0126
0.50	200	MBA	0	0	2	0	2	0.0004
		BMA	0	0	4	0	4	0.0008

3.4 An Illustration: Connecticut Child Survey data (CCSD)

In this section we illustrate the methods described in the previous section using an epidemiologic survey data on the school children of ages 6 to 11. This particular data set was collected in Connecticut through two epidemiologic surveys namely, the New Haven Child Survey (NHCS) and the Eastern Connecticut Child Survey (ECCS). For original sources of the data, we refer to Zahner et al [17] and Fitmaurice et al [12]. In both surveys emotional and behavioral information on each child was obtained from a parent or primary care-giver, and also from the child's teacher. By design, there was no overlap of children within families or within teachers. The child's emotional and behavioral problems were assessed using a standardized scale completed by both the parents and teachers. Altogether 2,501 children of both sexes participated in the survey. In addition to their emotional status (determined by their parent or teacher), a covariate measuring the parental dissatisfaction with family life was also recorded. In this illustration however, we ignore the covariate for simplicity. Our main objective is to see the difference between the behavior of male and female children. Considering the sex as a covariate and parental dissatisfaction as a second covariate, Fitmaurice et al [12] mainly studied the effect of these covariates on the emotional pattern of the child.

Of the 2,501 children about which information was collected in the two stud-

ies, we have considered only 1,428 with complete information. For each of these 1,428 children, both parents and teachers provided information on emotional status measured by the scale mentioned above. Note that information on each child given by either the parent or teacher was in a dichotomized form obtained from the corresponding scale score at the "clinical-borderline" range. More specifically, if either parent or teacher rated the child as emotionally disturbed, then this status was symbolized as '1', a binary outcome. Otherwise, the status was indicated by '0'. Further note that the binary information (0 or 1) referred by the teacher would be positively correlated with the binary information (0 or 1) rated by the parent as they are rating on same child. This correlation would be denoted by ϕ_1 for male children and ϕ_2 for female children.

3.4.1 Classifying Parent-Teacher Information into Male or Female Group

In order to illustrate our methodologies developed in the previous section, we now formulate the above CCSD problem as follows. Suppose that the ratings of both the teacher and parent is available as $(y_1,\ y_2) \equiv \{(1,\ 1),\ (1,\ 0),\ (0,\ 1),\ (0,\ 0)\}$. The question, based on some sample information, is whether it is possible to recognize a new bivariate information (y_{10}, y_{20}), say, arises from a male or from a female child? To answer this, we first exhibit the sample information collected by the CCSD as below.

Table 3.34: Cross-classification of Parent and Teacher Ratings of Male and Female.

Male Children

Teacher	Parent		Total
	0	1	
0	440	116	556
1	96	50	146
Total	536	166	702

Female Children

Teacher	Parent		Total
	0	1	
0	516	86	602
1	91	33	124
Total	607	119	726

Next based on both BMA and MBA as discussed in the previous section we estimate their respective parameter estimation performance as follows.

Estimation of Parameter by BMA

In order to develop the classification we have to estimate the unknown parameter in both the groups (1=Male and 2=Female). For group i, if $n_{(i)k}$ is the cell count for the kth ($k = 1, \cdots, 4$) cell out of n_i observations, then the maximum likelihood estimates of kth cell probability in ith ($i = 1,\ 2$) group is given by

$$\tilde{\theta}_{(i)k} = \frac{n_{(i)k}}{n_i} \text{ for } k = 1, \cdots, 4$$

By using the data from the Table 3.34, we obtain the likelihood estimates as follows.

For Male Group:

$$\tilde{\theta}_{(1)1} = 0.0712 \quad \tilde{\theta}_{(1)2} = 0.1368$$

$$\tilde{\theta}_{(1)3} = 0.1652 \quad \tilde{\theta}_{(1)4} = 0.6268$$

For Female Group:

$$\tilde{\theta}_{(2)1} = 0.0455 \quad \tilde{\theta}_{(2)2} = 0.1253$$

$$\tilde{\theta}_{(2)3} = 0.1185 \quad \tilde{\theta}_{(2)4} = 0.7107$$

Estimation of Parameters by MBA

For Male Group:

$$\hat{p}_{11} = \frac{n_{(1)1} + n_{(1)2}}{n_1} = 0.2080 \quad \hat{p}_{12} = \frac{n_{(1)1} + n_{(1)3}}{n_1} = 0.2365 \quad \text{and}$$

$$\hat{\phi}_1 = \frac{1}{\sqrt{\hat{p}_{11}\hat{q}_{11}\hat{p}_{12}\hat{q}_{12}}}\left(\frac{n_{(1)1}}{n_1} - \hat{p}_{11}\hat{p}_{12}\right) = 0.1276$$

For Female Group:

$$\hat{p}_{21} = \frac{n_{(2)1} + n_{(2)2}}{n_2} = 0.1708 \quad \hat{p}_{22} = \frac{n_{(2)1} + n_{(2)3}}{n_2} = 0.1639 \quad \text{and}$$

$$\hat{\phi}_2 = \frac{1}{\sqrt{\hat{p}_{21}\hat{q}_{21}\hat{p}_{22}\hat{q}_{22}}}\left(\frac{n_{(2)1}}{n_2} - \hat{p}_{21}\hat{p}_{22}\right) = 0.1257$$

Thus we have the pooled estimate of the correlation coefficient as

$$\hat{\phi} = \frac{n_1\hat{\phi}_1 + n_2\hat{\phi}_2}{n_1 + n_2} = 0.1266$$

Therefore. substituting these estimates in the corresponding formula. we can estimate the cell probability for each group for the proposed model as given below:

For Male Group:

$$\hat{\theta}_{(1)1} = \hat{p}_{11}\hat{p}_{12} + \hat{o}\sqrt{\hat{p}_{11}\hat{q}_{11}\hat{p}_{12}\hat{q}_{12}} = 0.0710.$$

$$\hat{\theta}_{(1)2} = \hat{p}_{11}\hat{q}_{12} - \hat{o}\sqrt{\hat{p}_{11}\hat{q}_{11}\hat{p}_{12}\hat{q}_{12}} = 0.1370.$$

$$\hat{\theta}_{(1)3} = \hat{q}_{11}\hat{p}_{12} - \hat{o}\sqrt{\hat{p}_{11}\hat{q}_{11}\hat{p}_{12}\hat{q}_{12}} = 0.1655.$$

$$\hat{\theta}_{(1)4} = \hat{q}_{11}\hat{q}_{12} + \hat{o}\sqrt{\hat{p}_{11}\hat{q}_{11}\hat{p}_{12}\hat{q}_{12}} = 0.6265.$$

For Female Group:

$$\hat{\theta}_{(2)1} = \hat{p}_{21}\hat{p}_{22} + \hat{o}\sqrt{\hat{p}_{21}\hat{q}_{21}\hat{p}_{22}\hat{q}_{22}} = 0.0456.$$

$$\hat{\theta}_{(2)2} = \hat{p}_{21}\hat{q}_{22} - \hat{o}\sqrt{\hat{p}_{21}\hat{q}_{21}\hat{p}_{22}\hat{q}_{22}} = 0.1252.$$

$$\hat{\theta}_{(2)3} = \hat{q}_{21}\hat{p}_{22} - \hat{o}\sqrt{\hat{p}_{21}\hat{q}_{21}\hat{p}_{22}\hat{q}_{22}} = 0.1183.$$

$$\hat{\theta}_{(2)4} = \hat{q}_{21}\hat{q}_{22} + \hat{o}\sqrt{\hat{p}_{21}\hat{q}_{21}\hat{p}_{22}\hat{q}_{22}} = 0.7109.$$

Classification Criterion:

Now following the classification rule (3.3.12), it is readily seen that any new observation belonging to cell

(1, 1) will be classified to the Male group as $\hat{\theta}_{11} \geq \hat{\theta}_{21}$.

Similarly, any new observation that belongs to cell

(1.0) will be classified to the Male group as $\hat{\theta}_{12} \geq \hat{\theta}_{22}$,

(0,1) will be classified to the Male group as $\hat{\theta}_{13} \geq \hat{\theta}_{23}$, and

(0,0) will be classified to the Female group as $\hat{\theta}_{14} \leq \hat{\theta}_{24}$.

3.4.2 A CCSD Based Simulation Study to Examine the Performance of BMA and MBA for Classification

As the sample sizes $n_1 = 702$ and $n_2 = 726$ are sufficiently large the parameter estimates can be considered to be very close to the population values. In order to compute the empirical probability of misclassification we considered two hypothetical groups G_1 and G_2. Under G_1 let $p_{11} = 0.2080$ and $p_{12} = 0.2365$ and correlation coefficient $o = 0.1266$. Similarly, under G_2 let $p_{21} = 0.1708$ and $p_{22} = 0.1639$ and same correlation coefficient $o = 0.1266$. These parameters are exactly the same as the corresponding estimates obtained for the CCSD data. The steps involved in the simulations are as follows:

1 Using the same procedure as discussed in section 3.2.2, a bivariate binary random sample of size $n_1 = 702$ is generated from group G_1 (Male) with marginal probabilities $p_{11} = 0.2080$ and $p_{12} = 0.2365$ and correlation parameter $\phi = 0.1266$ and on the basis of this generated sample

the quaitities $\hat{p}_{11}$, $\hat{p}_{12}$, and $\hat{\phi}_1$ are calculated by using the formulae given in section (3.2.1.1).

2 Similarly a bivariate binary random sample of size $n_1 = 726$ is generated from group G_2 (Female) with marginal probabilities $p_{11} = 0.1708$ and $p_{12} = 0.1369$ and the same correlation parameter $\phi = 0.1266$ and the quantities $\hat{p}_{21}$, $\hat{p}_{22}$, and ϕ_2 are calculated.

3 Using $\hat{\phi}_1$ and $\hat{\phi}_2$, we calculate the pooled estimate $\hat{\phi}$ by (3.2.8). And finally we calculate the cell probabilities $\hat{\theta}_{(i)k}$ based on MBA and $\tilde{\theta}_{(i)k}$ based on BMA. Comparing these estimated cell probabilities for the two groups (Male and Female) we check whether there is misclassification in each of the two approaches and record it.

4 Continue step 1-3 5000 times and then we compute the probability of misclassification for both the approaches MBA and BMA and the results are shown below.

For MBA

Probability of misclassification = 0.0214.

For BMA

Probability of misclassification = 0.0574.

The above results indicate that the probability of misclassification based on BMA is 3.6% higher than that of MBA. This clearly demonstrates the advantage of modelling the correlated binary data using the joint probability function given in (3.1.1), in classifying a new bivariate binary observation into one of the two groups.

Chapter 4

Classification of A Correlated Binary Observation With Covariates: A Model Based Approach

Let $Y_{il} = [y_{il1}, y_{il2}]'$ be the 2×1 vector of two correlated binary variables for the lth ($l = 1, \cdots, n_i$) subject in the ith group G_i. Also let $X_{il} = (x_{il1}, \cdots, x_{ilu}, \cdots, x_{ilp})'$ be the corresponding $p \times 1$ vector of covariates. A layout for data of this type, for the ith group G_i, is given below.

Table 4.1: Data for Correlated Binary Model with p covariates.

Observation	1	2	1	2	$\cdots$	u	$\cdots$	p
1	y_{i11}	y_{i12}	x_{i11}	x_{i12}	$\cdots$	x_{i1u}	$\cdots$	x_{i1p}
2	y_{i21}	y_{i22}	x_{i21}	x_{i22}	$\cdots$	x_{i2u}	$\cdots$	x_{i2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
l	y_{il1}	y_{il2}	x_{il1}	x_{il2}	$\cdots$	x_{ilu}	$\cdots$	x_{ilp}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
n_i	y_{in_i1}	y_{in_i2}	x_{in_i1}	x_{in_i2}	$\cdots$	x_{in_iu}	$\cdots$	x_{in_ip}

4.1 Covariates Based Joint Probability Model

Recall that for the ith group G_i the bivariate binary variables Y_{il1} and Y_{il2} are jointly distributed as

$$f(y_{il1}, y_{il2}|G_i) = p_{il1}^{y_{il1}} q_{il1}^{1-y_{il1}} p_{il2}^{y_{il2}} q_{il2}^{1-y_{il2}} \left[1 + \phi \frac{(y_{il1} - p_{il1})(y_{il2} - p_{il2})}{\sqrt{p_{il1}q_{il1}p_{il2}q_{il2}}}\right] \quad (4.1.1)$$

as in (3.1.1), where p_{il1} be the probability for $y_{il1} = 1$ and p_{il2} be the probability for $y_{il2} = 1$. In the present case, however, the marginal probabilities p_{il1} and p_{il2} will be modelled as functions of covariates x_{il}. More specifically we use a binary logistic function to model these probabilities as follows:

$$p_{ilj} = h(X_{il}\beta_{ij}) = \frac{e^{X_{il}'\beta_{ij}}}{1 + e^{X_{il}'\beta_{ij}}},$$

for $i = 1, 2$; $l = 1, \cdots, n_i$; and $j = 1, 2$. Here for $j = 1, 2$, $\beta_{ij} = [\beta_{ij1}, \cdots, \beta_{ijp}]'$ is the p-dimensional vector of regression parameters.

4.1.1 Estimation of Parameters: An Estimating Equation (EE) Approach

For the ith ($i = 1, 2$) group G_i and for known ϕ, we first estimate the $\beta_{ij} = [\beta_{ij1}, \cdots, \beta_{ijp}]'$ for $j = 1, 2$, by using the estimating equation approach discussed by Prentice and Zhao [28], and other authors. More specifically $\beta_i = (\beta_{i1}', \beta_{i2}')'$ is the root of the quasi-likelihood estimating equations

$$n_i^{-1} \sum_{l=1}^{n_i} D_{il}' V_{il}^{-1} S_{il} = 0 \quad i = 1, 2 \quad (4.1.2)$$

where $S_{il} = Y_{il} - E(Y_{il}) = [y_{il1} - p_{il1},\ y_{il2} - p_{il2}]'$, V_{il} is the variance-covariance matrix defined as

$$\begin{aligned} V_{il} &= \begin{pmatrix} Var(y_{il1}) & Cov(y_{il1}, y_{il2}) \\ Cov(y_{il2}, y_{il1}) & Var(y_{il2}) \end{pmatrix} \\ &= \begin{pmatrix} p_{il1}q_{il1} & \phi\sqrt{p_{il1}q_{il1}p_{il2}q_{il2}} \\ \phi\sqrt{p_{il1}q_{il1}p_{il2}q_{il2}} & p_{il2}q_{il2} \end{pmatrix}. \end{aligned}$$

$$\begin{aligned} D_{il} &= \begin{pmatrix} \frac{\delta p_{il1}}{\delta \beta_{i11}} & \cdots & \frac{\delta p_{il1}}{\delta \beta_{i1p}} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{\delta p_{il2}}{\delta \beta_{i21}} & \cdots & \frac{\delta p_{il2}}{\delta \beta_{i2p}} \end{pmatrix} \\ &= \begin{pmatrix} d_{il1}^{11} & \cdots & d_{ilp}^{11} & d_{il1}^{12} & \cdots & d_{ilp}^{12} \\ d_{il1}^{21} & \cdots & d_{ilp}^{21} & d_{il1}^{22} & \cdots & d_{ilp}^{22} \end{pmatrix}, \end{aligned}$$

with

$$d_{ilu}^{lm} = \begin{cases} \frac{\delta p_{ilw}}{\delta \beta_{imu}} = \frac{x_{ilu}e^{X_{il}'\beta_{im}}}{(1+e^{X_{il}'\beta_{im}})^2}, & \text{for } w = m; \quad w = 1,2;\ m = 1,2. \\ 0 & \text{for } w \neq m \end{cases}.$$

The solution for β_{ij} for $j = 1, 2$ may be obtained from (4.1.2) by using the well-known Newton-Rapson method. This is, however, equivalent to use the iterative equation given by

$$\hat{\beta}_i(t+1) = \hat{\beta}_i(t) + \Big(\sum_{l=1}^{n_i} D_{il}'\hat{V}_{il}^{-1}D_{il}\Big)_t^{-1}\Big(\sum_{l=1}^{n_i} D_{il}'\hat{V}_{il}^{-1}S_{il}\Big)_t \qquad (4.1.3)$$

where $\hat{\beta}_i(t)$ is the tth iteration value of $\hat{\beta}_i$ and the expression $(\cdot)_t$ denotes that the quantities within the brackets are evaluated at $\hat{\beta}_i(t)$.

Now at the $(t+1)$-th iteration the probabilities p_{il1} and p_{il2} are estimated as

$$\hat{p}_{il1}(t+1) = \frac{e^{X'_{il}\hat{\beta}_{i1}(t+1)}}{1+e^{X'_{il}\hat{\beta}_{i1}(t+1)}} \quad \text{for } l = 1, \cdots, n_i,$$

and

$$\hat{p}_{il2}(t+1) = \frac{e^{X'_{il}\hat{\beta}_{i2}(t+1)}}{1+e^{X'_{il}\hat{\beta}_{i2}(t+1)}} \quad \text{for } l = 1, \cdots, n_i$$

respectively.

For convenience, let ϕ computed from the ith group be denoted by ϕ_i. Now to estimate ϕ_i, that is to obtain $\phi_i(t+1)$ from $\hat{\beta}_i(t+1)$ and hence from $\hat{p}_{ilj}(t+1)$, we use the method of moment and compute

$$\hat{\phi}_i(t+1) = \sum_{l=1}^{n_i} \frac{(y_{il1} - \hat{p}_{il1}(t+1))(y_{il2} - \hat{p}_{il2}(t+1))}{\sqrt{\hat{p}_{il1}(t+1)\hat{q}_{il1}(t+1)\hat{p}_{il2}(t+1)\hat{q}_{il2}(t+1)}}. \tag{4.1.4}$$

This new value of $\hat{\phi}_i(t+1)$ is then used in (4.1.3) to obtain $\hat{\beta}_i(t+2)$ which in turn produces $\hat{\phi}_i(t+2)$ by (4.1.4). This cycle of iteration continues until convergence. Let the final estimates be $\hat{\beta}_i$ and $\hat{\phi}_i$. Next as we have assumed common correlation in (4.1.1), we estimate this common correlation parameter ϕ by pooling ϕ_1 and ϕ_2 as

$$\hat{\phi} = \frac{n_1\hat{\phi}_1 + n_2\hat{\phi}_2}{n_1+n_2}$$

We remark here that one could also estimate the ϕ_i $(i = 1, 2)$ and eventually the ϕ parameter by the EE approach. This approach, however, will require the computations for the third and fourth order moments of the joint binary probability distribution, which appears to be complicated. Alternatively, as the probability model of the correlated binary variables is known by (4.1.1), one may also use the likelihood method to estimate ϕ_i. But the computations for this type of likelihood estimates also appear to be complicated.

4.2 Covariates Based Classification Criterion

In classifying a new correlated binary observation with covariates into one of the two groups G_1 and G_2, we use the same classification rule (3.2.3). That is, classify a new $y_0 = (y_{10},\ y_{20})$ into G_1 if

$$\phi > \frac{(-1)^{y_{10}+y_{20}}}{w_1 - w_2}\left[p_{2l_01}^{y_{10}} q_{2l_01}^{1-y_{10}} p_{2l_02}^{y_{20}} q_{2l_02}^{1-y_{20}} - p_{1l_01}^{y_{10}} q_{1l_01}^{1-y_{10}} p_{1l_02}^{y_{20}} q_{1l_02}^{1-y_{20}}\right], \qquad (4.2.5)$$

where X_{l_0} is the covariate associated with the new observation y_0 with

$$p_{il_0k} = \frac{e^{X_{l_0}'\beta_{ij}}}{1 + e^{X_{l_0}'\beta_{ij}}}, \quad i = 1, 2;\ \ j = 1, 2,$$

and

$$w_i = \sqrt{p_{il_01} q_{il_01} p_{il_02} q_{il_02}} \quad (i = 1, 2)$$

When the parameters are unknown, we use the classification rule (4.2.5) after replacing the β and ϕ parameters with their estimates. The classification rule

(4.2.5) then reduces to:

$$\hat{\phi} > \frac{(-1)^{y_{10}+y_{20}}}{\hat{w}_1 - \hat{w}_2} \left[\hat{p}_{2l_01}^{y_{10}} \hat{q}_{2l_01}^{1-y_{10}} \hat{p}_{2l_02}^{y_{20}} \hat{q}_{2l_02}^{1-y_{20}} - \hat{p}_{1l_01}^{y_{10}} \hat{q}_{1l_01}^{1-y_{10}} \hat{p}_{1l_02}^{y_{20}} \hat{q}_{1l_02}^{1-y_{20}} \right], \qquad (4.2.6)$$

to classify y_0 in G_1.

Note that for the two cases whether covariates are associated with responses or not, the classification rule appears quite similar. The difference lies only in the estimation of the parameters β's and ϕ. In the first case, when there is no covariate, the estimating equation (3.2.6) for p_{i1}, and p_{i2}, $i = 1, 2$ does not involve the variance covariance matrix as it is constant for both groups G_1 and G_2, whereas in the second case the variance covariance matrix plays an important role as shown in (4.1.2). As there is no extra burden in computing the classification rule in the second case (as compared to the first one without considering any covariates) except this difference in estimation, we do not pursue further simulation for the classification problem with covariates.

Chapter 5

Concluding Remarks

Classification of a multi-dimensional observation into one of two groups is an important practical problem. For the cases, when a multi-dimensional observation follows a Gaussian or a continuous distribution, there exits numerous studies (cf. McLachlan [25] and Seber [29]) for this type of classification problem. But in certain specific situations, for example, in biomedical applications the multi-dimensional observation may be discrete or, more specifically, it may follow the multivariate binary distribution. In a further specialized but important situation, one may even deal with the classification of a bivariate binary observation. For example, we refer to the CCSD data discussed in chapter 3.

Unlike in the continuous case, this type of classification problems for bivariate binary data are dealt with by using a suitable distribution-free approach or a certain semi-parameter approach such as BMA discussed in the thesis. In the BMA, the joint probabilities under two groups are estimated by using

the multinomial based maximum likelihood estimation technique. This is mainly done as there is no joint probability model known (or available) for the bivariate binary observation that may belong to any of the four cells. It is, therefore, not clear how the correlation between the two correlated binary data is taken into account in such a basic multinomial approach. Alternatively, in higher dimensional cases, a certain log-linear approach is used to interpret the association of the variables. But as shown in chapter 2, under some conditions, this approach is basically the same as the BMA in classifying a bivariate binary observation.

As argued in the thesis, we have chosen to model the joint probability of a correlated bivariate binary variables by following the idea of Prentice [27] (see also Sutradhar and Das [33]). This modelling takes the correlation between the two binary variables into account in a natural way. It is shown in this thesis that for the large sample case, the parameter of the model including the correlation parameter (ϕ) may be estimated with sufficient accuracy. We have then used these estimates to estimate the joint probability under each group and used the optimum classification rule to classify a new observation, based on the magnitude of the estimated probabilities in the corresponding cells of the two groups. We have conducted a Monte Carlo experiment with 5,000 simulations to examine the performance of this new modelling in classifying a correlated bivariate binary observation as compared to the BMA, where no model is known or available. It has been shown that in general, the

modelling of the cell probability has a significant effect in classifying such an observation into one of the two groups. More specifically it was found that the probability of misclassification is less if classification is based on the proper modelling.

We remark, however, that although there is an immediate generalization of the bivariate binary probability distribution to the multivariate binary distribution, this type of generalization puts severe restrictions on the possible values of the correlation parameters. Therefore, it may be better to search for a new approach to model such higher dimensional binary distributions. Alternatively, we may examine the performance of the distribution-free approach or kernel approach to classify such observations. This is, however, beyond the scope of the present thesis.

Bibliography

[1] J. Aitchison and C. G. G. Aitken. *Multivariate binary discrimination by the kernel method.*. Biometrika **63** (1976). 413–420.

[2] J. Anderson. *Some nonparametric multivariate procedures based on statistically equivalent blocks.*. Multivariate Analysis (1966). 5–27.

[3] J. A. Anderson. K. Williamson J. Whaley. and W. W. Buchanan. *A statistical aid to the diagnosis of keratoconjuctivitis sicca.*. Quarterly Journal of Medicine **162** (1972). 175–189.

[4] T. W. Anderson. *An introduction to multivariate statistical analysis.* John Wiley and Sons. New York. 1984.

[5] R. R. Bahadur. *A representation of the joint distribution of response to n dichotomous item.*. Studies in Item Analysis and Prediction (1961). 158–168.

[6] T. Cacoullos. *Estimation of a multivariate density.* Ann. Inst. Statist. Math. **18** (1966). 179–189.

[7] Z. Y. Chen. *Robust linear discriminant procedures using projection pursuit methods.*. PhD. Dessertation. Department of Statistics. University of Michigan. Ann Arbor. 1989.

[8] Z. Y. Chen and R. J. Muirhead. *A comparison of robust linear discriminant procedures using projection pursuit methods.*. Multivariate Analysis and its Aplication. IMS Lecurer Notes - Monograph Series **24** (1994), 163–176.

[9] P. Deheuvels. *Estimation non parametrique de la densite par histogrammes generalises (ii)*, Publications de l'institute de Statistique de l'Universite de Paris **XXII** (1977), 1–23.

[10] V. A. Epanechnikov. *Non-parametric estimation of a multivariate probability density*, Theory of Probability and Application (1969), 153–158.

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

[12] G. M. Fitmaurice, N. M. Laird, and G. E. P. Zahner. *Multivariate logistic models for incomplete binary responses.* Journal of Amerian Statistical Association **91** (1996), no. 433. 99–108.

[13] E. Fix and J. L. Hodges, *Discriminatory analysis-nonparametric discrimination. consistency properties.*. Report No. 4 (Randolph Field. Texas), U. S. Air Force School of Aviation Medicine.. 1977. pp. 261–279.

[14] J. H. Friedman and J. W. Tukey, *A projection pursuit algorithm for explanatory data analysis.*, IEEE Trans. Comput. **C-23** (1974), 881–890.

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

[16] M. Goldstain and W. R. Dillon, *Discrete multivariate analysis.* John Wiley and Sons, New York, 1978.

[17] J. H. Jacobs J. H., D. H. Freeman, G. E. P. Zahner, and K. F. Trainor. *Rural-urban child psychopathology in a northeastern us state: 1986-1989.* Journal of the American Academy of Child and Adolescent Psychiatry **32** (1993), 378–387.

[18] P. J. Huber, *Projection pursuit (with discussion).*, Annuals of Statistics. **13** (1985), 435–475.

[19] M. C. Jones and Tukey R. Sibson, *What is projection pursuit? (with discussion)*, Journal of the Royal Statistical Society, Series A **150** (1987), 1–36.

[20] T. Kariya, *Robustness of multivariate tests*, Annuals of Statistics. **9** (1981), 1267–1275.

[21] S. Kullback, *Information theory and statistics*, John Wiley and Sons, New York, 1959.

[22] P. A. Lachenbruch and M. Goldstein, *Discriminant analysis*, Biometrics **35** (1979), 69–85.

[23] W. Q. Liang and P. R. Krishnaiah, *Nonparametric iterative estimation of multivariate binary density*, Journal of Multivatiate Analysis **16** (1985), 162–172.

[24] D. C. Martin and R. A. Bradly, *Probability models, estimation, and classification for multivariate dichotomous populations*, Biometrics **28** (1972), 203–222.

[25] Geoffrey J. McLachlan, *Discriminant analysis and statistical pattern recognition*, John Wiley and Sons, New York-Chichester-Brisbane-Toronto-Singapore, 1992.

[26] J. Ott and R. A. Kronmal, *Some classification procedures for binary data using orthogonal functions*, Journal of Amerian Statistical Association **71** (1976), 391–399.

[27] R. L. Prentice, *Correlated binary regression with covariates specific to each binary observation*, Biometrics **44** (1988), 1033–1048.

[28] R. L. Prentice and L. P. Zao, *Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses*, Biometrics **47** (1991), 825–839.

[29] G. A. F. Seber, *Multivariate observations*, John Wiley and Sons, New York-Chichester-Brisbane-Toronto-Singapore, 1984.

[30] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall., London, 1986.

[31] D. S. Stoffer, *Walsh-fourier analysis and its statistical applications (with discussion)*., Journal of Amerian Statistical Association **86** (1991), 461–485.

[32] B. C. Sutradhar, *Discrimination of observations into one of two t population*, Biometrics **46** (1990), 827–835.

[33] B. C. Sutradhar and K. Das, *Generalized linear models for beta correlated binary longitudinal data*, Cummunication in Statistics, Theory and Method **26** (1997), no. 3, 617–635.