Distance models for analysis of multivariate binary data

Worku, H.M.

Chapter 3

Properties of Ideal Point Classification Models for Bivariate Binary Data

Abstract

The Ideal Point Classification (IPC) model was originally proposed for analysing multinomial data in the presence of predictors. In this paper, we studied properties of the IPC model and extended it for analysing bivariate binary responses with a specific focus on three parameters: (1) the marginal probabilities; (2) the association structure between the two binary responses; and (3) the joint probabilities. We found that the IPC model with a specific class point configuration, represents either the marginal probabilities or the association structure. However, the IPC model is not able to represent both parameters at the same time. We then derived a new parameterization of the model, the Bivariate IPC (BIPC) model, which is able to represent both the marginal probabilities and the association structure. Like the standard IPC model, the results of the BIPC model can be displayed in a biplot, from which the effects of predictors on the binary responses and on their association can be read. We will illustrate our findings with a psychological example relating personality traits to depression and anxiety disorders.

This chapter was published as Worku, H. M. & De Rooij, M. (2017). Properties of Ideal Point Classification Models for Bivariate Binary Data. *Psychometrika, 82 (2), 308-328.* To address remarks of the PhD committee, this chapter is slightly modified.

3.1 Introduction

Multiple binary outcome data are often collected in epidemiology, psychology, medicine, and other life and behavioral sciences. For example, in the Netherlands Study of Depression and Anxiety (NESDA) data were collected on depression and anxiety disorders, and how these disorders are influenced by personality traits and background variables (Penninx et al., 2008; Spinhoven et al., 2009). In this paper, we focus on bivariate binary data in which two dichotomous response variables are observed for each subject in a study. Another example with bivariate binary data is the British coalminers study (Ashford et al., 1970), which investigated data on breathlessness ($1 =$ difficult; $0 =$ Normal) and wheeze ($1 =$ difficult; $0 =$ Normal) of coalminers in Britain, to study the impact of exposure on these respiratory indicators (Ashford et al., 1970; McCullagh & Nelder, 1989; Palmgren, 1989).

Let us denote the bivariate binary responses observed from the i-th subject by Y_{i1} and Y_{i2}. The p dimensional vector $\mathbf{x}_i$ represents the explanatory variables without including an intercept, where $i = 1, 2, \ldots, N$. The cross-classified binary responses are displayed in Table 3.1 in which the corresponding probabilities are also presented, i.e., the probabilities within the four cells represent the joint probabilities; and those at the margins represent the marginal probabilities. Empirical researchers working with bivariate binary data are often interested in the following parameters: (1) the marginal probabilities; (2) the association between the two binary responses; and (3) the joint (or multinomial) probabilities.

In marginal modelling, the main focus is on the analysis of the marginal probabilities separately in which the association structure between the binary responses could be a direct interest or treated as a nuisance parameter (Agresti, 2002, pp. 455; Molenberghs & Verbeke, 2005, pp. 55). In the margins of Table 3.1, the marginal probabilities are denoted by $\pi_{il\bullet} = \mathrm{Pr}(Y_{i1} = l)$ and $\pi_{i\bullet l} = \mathrm{Pr}(Y_{i2} = l)$, where $l = 0, 1$. Bahadur (1961) proposed a marginal model based on the full likelihood for analysing bivariate binary data. The

joint distribution was characterized by the two marginal distributions and the correlation between the two binary responses. Lipsitz, Laird and Harrington (1990) followed the idea of Bahadur (1961) and showed that other measures of association can also be used (e.g., the odds ratio or relative risk). For a 2×2 contingency table, the odds ratio is calculated as $\tau_i = (\pi_{i,11} \times \pi_{i,00})/(\pi_{i,10} \times \pi_{i,01})$ where $\pi_{i,11} = \mathsf{Pr}(Y_{i1} = 1, Y_{i2} = 1)$; $\pi_{i,00} = \mathsf{Pr}(Y_{i1} = 0, Y_{i2} = 0)$; $\pi_{i,10} = \mathsf{Pr}(Y_{i1} = 1, Y_{i2} = 0)$; and $\pi_{i,01} = \mathsf{Pr}(Y_{i1} = 0, Y_{i2} = 1)$.

Marginal model parameters can be fitted directly or by imposing restrictions on the joint distribution (Molenberghs & Verbeke, 2005, pp. 49). Aitchison and Silvey (1958, 1960) originally proposed constrains on parameters in maximum likelihood function. Their approach was later applied to categorical data by Lang and Agresti (1994), and other researchers (Lang, 1996; Bergsma, 1997; Bergsma & Rudas, 2002; Vermunt, Rodrigo, & Ato-Garcia, 2001). McCullagh and Nelder (1989) introduced a multivariate logistic transformation which can be used to relate the joint distribution to the marginal probabilities and the association structure. Their approach is widely used for marginal modelling of multivariate categorical responses (Glonek & McCullagh, 1995; Molenberghs & Lesaffre, 1994, 1999).

In recent years, the marginal modelling strategy has shifted from fitting and testing linear constraints on parameters to inequality constraints for addressing certain scientific questions (Colombi & Forcina, 2001; Bartolucci, Forcina, & Dardanoni, 2001; Bartolucci, Colombi, & Forcina, 2007). For ordinal responses, for example, it may be interesting to know whether the univariate distributions are stochastically ordered in some way, i.e., whether pairs of responses are positively correlated, or whether the degree of positive dependence changes with certain predictor variables (Colombi & Forcina, 2001).

The main drawback of a full likelihood-based marginal modelling approach is that it is computationally intensive and prone to model misspecification, especially when the number of response variables increases (Agresti, 2002, pp. 465; Molenberghs & Verbeke, 2005, pp. 151). Liang and Zeger (1986) proposed an extension of quasi-likelihood

Table 3.1: Cross-classification of bivariate binary data observed from i-th subject.

		Y_{i2}		
		1	0	
Y_{i1}	1	$\pi_{i,11}$	$\pi_{i,10}$	$\pi_{i1.}$
	0	$\pi_{i,01}$	$\pi_{i,00}$	$\pi_{i0.}$
		$\pi_{i.1}$	$\pi_{i.0}$	1.00

method, called Generalized Estimating Equations (GEE or GEE1), that does not require full specification of the response distribution. In GEE1 the association structure is treated as a nuisance parameter. Second-order GEE, called GEE2, (Liang et al., 1992) and Alternating Logistic Regression (ALR: Carey, Zeger, & Diggle, 1993) are commonly used for modelling both the marginal probabilities and the association structure.

The third parameter of interest are the joint probabilities. The joint probabilities as displayed in Table 3.1 (i.e., $\pi_{i,00}$; $\pi_{i,10}$; $\pi_{i,01}$; and $\pi_{i,11}$) correspond to a multinomial response variable, denoted by G_i, with four categories ($g = 4$). For simplicity, we use a single index to refer to the joint probabilities, i.e., $\pi_{ij} = \Pr(G_i = j)$. For example, the four cells in Table 3.1 can be represented as: $\pi_{i1} = \pi_{i,00}$; $\pi_{i2} = \pi_{i,10}$; $\pi_{i3} = \pi_{i,01}$; and $\pi_{i4} = \pi_{i,11}$. In the NESDA study, for example, a multinomial response variable can be defined from the two binary outcome variables. That is, $G_i = 1$ if the subject has no depression or anxiety; $G_i = 2$ if (s)he has an anxiety disorder, but no depression disorder; $G_i = 3$ if the subject has depression disorder, but no anxiety disorder; and $G_i = 4$ if there is co-morbidity. Statistical models such as the Multinomial Baseline-Category Logit (MBCL: Agresti, 2002, pp. 267) or Ideal Point Classification (IPC: De Rooij, 2009a), can be used to analyse multinomial response variables in the presence of predictors.

De Rooij (2009a) proposed the IPC model for analysing a multinomial response variable in the presence of predictors. The IPC model is a probabilistic multidimensional unfolding model and closely related to Ideal Point Discriminant Analysis (IPDA) as proposed by Takane, Bozdogan, and Shibayama (1987). Both IPDA and IPC models are classification methods based on multidimensional unfolding (MDU) (Heiser, 1981, 1987; De Leeuw,

2005). The objective of MDU is to find distances in Euclidean space between subjects and objects that approximate a set of proximities as good as possible. In IPC and IPDA models, the proximity is given by an indicator matrix that corresponds to the multinomial response.

De Rooij (2009a) showed that the IPC model in maximum dimensionality is equivalent to the MBCL model, i.e., if the dimensionality of the Euclidean space equals the number of categories of the response variable minus one. The MBCL is a natural extension of binary logistic regression to the case of nominal categorical variables. Both the IPC and the MBCL models use the joint probabilities to define their likelihood function. Unlike in the MBCL model, dimension reduction is possible in the IPC models. Thus, less model parameters are estimated in the reduced space. Furthermore, the results of the IPC model can be displayed using a biplot (Gower & Hand, 1996; Gower et al., 2011) which enhance interpretation of the model.

In this paper, our main aim is to study properties of the IPC model for bivariate binary data, specifically about the representation of the marginal probabilities and of the association structure. We will show that the IPC model either represents the marginal models or the association structure well. Next, we study a new parametrization of the IPC model, namely the Bivariate IPC (BIPC) model, in which both the marginal probabilities and the association structure are represented. This new model builds forward on the work of Bahadur (1961) and Lipsitz, Laird and Harrington (1990). Compared to this existing methodology for jointly modelling the marginal and association structure, our method has the advantage of dimension reduction and a graphical representation of the model using a biplot.

The paper is organized as follows. Section 2 presents the theoretical background. Section 3 studies properties of the IPC models both mathematically and with a simulation study. Section 4 proposes the BIPC model. Section 5 shows an example application and then we conclude in Section 6 with a discussion.

3.2 Background

3.2.1 The Ideal Point Classification Model

In the IPC model (De Rooij, 2009a) the conditional joint probabilities, i.e., $\pi_j(\mathbf{x}_i) = \Pr(G_i = j|\mathbf{x}_i)$, are modelled using a distance between two points in an Euclidean space of dimensionality M: one point representing subject i with coordinates $\boldsymbol{\eta}_i = [\eta_{i1}, \ldots, \eta_{iM}]^{\mathrm{T}}$, and the other representing class j with coordinates $\boldsymbol{\gamma}_j = [\gamma_{j1}, \ldots, \gamma_{jM}]^{\mathrm{T}}$. The smaller the relative distance between the two points, the larger the probability that the subject belongs to that class. The IPC model is defined as (De Rooij, 2009a),

$$\pi_j(\mathbf{x}_i) = \frac{\exp(-0.5 \times \delta_{ij})}{\sum_h \exp(-0.5 \times \delta_{ih})}, \tag{3.1}$$

where δ_{ij} is a squared Euclidean distance between the two points and is defined as

$$\delta_{ij} = \sum_{m=1}^{M} (\eta_{im} - \gamma_{jm})^2. \tag{3.2}$$

The coordinates of the subject points are assumed to be a linear combination of the predictor variables $\mathbf{x}_i$ and an intercept, i.e., $\boldsymbol{\eta}_i = \boldsymbol{\beta}_0 + \mathbf{x}_i\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a $(p \times M)$ matrix with regression weights and, $\boldsymbol{\beta}_0$ an M dimensional intercept. The parameters of this model are the regression weights and the class points.

Parameter estimates in the IPC model can be obtained by maximizing a multinomial log-likelihood function

$$\sum_{i=1}^{N} \left[\log \left(\prod_j \pi_j(\mathbf{x}_i)^{f_{ij}} \right) \right], \tag{3.3}$$

where $f_{ij} = 1$ if subject i is in category j, zero otherwise.

The IPC model has translation, rotational freedom, and multinomial indeterminacy (i.e., the class probability remains the same if a constant is added to each subject's squared

distance). The total number of restrictions needed is max$[M(M-1)/2, M(M+1)-(g-1)]$, and thus the total number of free parameters becomes npar $= (p+g)M-$max$[M(M-1)/2, M(M+1)-(g-1)]$ (De Rooij, 2009a). Depending on dimensionality of the fitted model, $\gamma-$parameters are set at fixed values to identify the model. For a multinomial response variable with $g = 4$ categories, for example, the maximum dimensionality of the IPC model is $M = 3(= g-1)$ and the total number of parameters in that case will be npar $= 3 \times (p+1)$ that corresponds to the regression parameters only since the class points can be set to fixed values that span the three-dimensional space. The class point coordinates can be specified, for example, as

$$\gamma = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.4}$$

The rows in (3.4) correspond to the response categories and the columns to the dimensions. In this case, the IPC model is equivalent to the MBCL model. The advantage of the IPC model over the MBCL model is that it provides the possibility of dimension reduction. For the multinomial response with $g = 4$, a 2-dimensional IPC model can be fitted with a total number of parameters npar $= 2 \times (p+1) + 3$, where the first part $(2 \times (p+1))$ represents the number of regression coefficients and the second part $(+3)$ the free class coordinates. From the eight class coordinates five need to be fixed for identification. This can be accomplished, for example, by defining

$$\gamma = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{bmatrix}, \tag{3.5}$$

where γ_{32}, γ_{41}, and γ_{42} are the free class coordinates, i.e., these can be estimated from the data.

3.2.2 The 2-step Approach of McCullagh and Nelder (1989)

We revisit a 2-step approach often used for constructing multivariate regression models using joint probabilities of multivariate (or bivariate) binary data, as proposed by McCullagh and Nelder (1989). We later apply this approach in the distance framework to study the properties of IPC models.

In the first step, a linear transformation is applied on the joint probabilities to obtain the marginal probabilities, i.e.,

$$\boldsymbol{\Lambda}_i = \mathbf{L}\boldsymbol{\pi}_i, \tag{3.6}$$

where $\mathbf{L}$ is a matrix of zeros and ones and $\boldsymbol{\pi}_i = [\pi_{i4}\ \pi_{i3}\ \pi_{i2}\ \pi_{i1}]^{\mathrm{T}}$. In the case of bivariate binary data, for example, the row margin is given by

$$\begin{aligned}
\boldsymbol{\Lambda}_{i1} &= \mathbf{L}_1\boldsymbol{\pi}_i \\
&= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \times [\pi_{i4}\ \pi_{i3}\ \pi_{i2}\ \pi_{i1}]^{\mathrm{T}} \\
&= \begin{bmatrix} \pi_{i4} + \pi_{i2} \\ \pi_{i3} + \pi_{i1} \end{bmatrix} = \begin{bmatrix} \pi_{i1\bullet} \\ \pi_{i0\bullet} \end{bmatrix}.
\end{aligned} \tag{3.7}$$

Similarly, the column margin is given by

$$\begin{aligned}\boldsymbol{\Lambda}_{i2} &= \mathbf{L}_2\boldsymbol{\pi}_i \\ &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \times \left[\pi_{i4}\ \pi_{i3}\ \pi_{i2}\ \pi_{i1}\right]^{\mathrm{T}} \\ &= \begin{bmatrix} \pi_{i4} + \pi_{i3} \\ \pi_{i2} + \pi_{i1} \end{bmatrix} = \begin{bmatrix} \pi_{i\bullet 1} \\ \pi_{i\bullet 0} \end{bmatrix}.\end{aligned} \tag{3.8}$$

In the second step, logarithmic contrasts of interest are formulated, i.e.,

$$\boldsymbol{\Psi}_i = \mathbf{C}^{\mathsf{T}} \log[\boldsymbol{\Lambda}_i], \tag{3.9}$$

for an appropriately chosen contrast matrix $\mathbf{C}^{\mathsf{T}}$. For the bivariate binary data, the contrast matrices can be chosen to be $\mathbf{C}^{\mathsf{T}} = \begin{bmatrix} 1 & -1 \end{bmatrix}$. Thus,

$$\begin{aligned}\psi_{i1} &= \begin{bmatrix} 1 & -1 \end{bmatrix} \log[\boldsymbol{\Lambda}_{i1}] \\ &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \log(\pi_{i1\bullet}) & \log(\pi_{i0\bullet}) \end{bmatrix}^{\mathsf{T}} \\ &= \log(\pi_{i1\bullet}) - \log(\pi_{i0\bullet}) \\ &= \log(\pi_{i1\bullet}/\pi_{i0\bullet}) \\ &= \mathsf{logit}(\pi_{i1\bullet}).\end{aligned} \tag{3.10}$$

Similarly, $\psi_{i2} = \log(\pi_{i\bullet 1}/\pi_{i\bullet 0}) = \mathsf{logit}(\pi_{i\bullet 1})$. In the presence of predictors these logits can be linked to the systematic part as used in Generalized Linear Models (Agresti, 2002); that is,

$$\begin{aligned}\mathsf{logit}(\pi_{i1\bullet}) &= \beta_{01} + \boldsymbol{\beta}_1^{\mathsf{T}}\mathbf{x}_i, \\ \mathsf{logit}(\pi_{i\bullet 1}) &= \beta_{02} + \boldsymbol{\beta}_2^{\mathsf{T}}\mathbf{x}_i.\end{aligned} \tag{3.11}$$

The above derivations (equation 3.6 - 3.11) can be summarized as follows.

$$\begin{aligned}\mathbf{C}^{\mathrm{T}}\log(\mathbf{L}_1\boldsymbol{\pi_i}) &= \beta_{01} + \boldsymbol{\beta}_1^{\mathsf{T}}\mathbf{x}_i,\\ \mathbf{C}^{\mathrm{T}}\log(\mathbf{L}_2\boldsymbol{\pi_i}) &= \beta_{02} + \boldsymbol{\beta}_2^{\mathsf{T}}\mathbf{x}_i.\end{aligned} \tag{3.12}$$

To obtain the association structure for bivariate binary data, the joint probabilities can also be transformed linearly. In this case $\mathbf{C}^{\mathsf{T}} = \begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}$ and $\mathbf{L} = \mathbf{I}$ such that,

$$\begin{aligned}\mathbf{C}^{\mathrm{T}}\log(\mathbf{L}\boldsymbol{\pi_i}) &= \begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}\log[\mathbf{I}\boldsymbol{\pi}_i]\\ &= \begin{bmatrix}1 & -1 & -1 & 1\end{bmatrix}\begin{bmatrix}\log(\pi_{i4}) & \log(\pi_{i3}) & \log(\pi_{i2}) & \log(\pi_{i1})\end{bmatrix}^{\mathsf{T}}\\ &= \log(\pi_{i4}) - \log(\pi_{i3}) - \log(\pi_{i2}) + \log(\pi_{i1})\\ &= \log\left[\frac{\pi_{i4}\times\pi_{i1}}{\pi_{i3}\times\pi_{i2}}\right]\\ &= \log(\tau_i).\end{aligned} \tag{3.13}$$

This odds ratio can be linked to predictors as

$$\log(\tau_i) = \beta_{03} + \boldsymbol{\beta}_3^{\mathsf{T}}\mathbf{x}_i. \tag{3.14}$$

3.3 Study-1: IPC Model as a Marginal Model

In this section, our aim is in how the IPC model represents both the marginal probabilities and the association structure for bivariate binary data. We use the 2-step approach of McCullagh and Nelder (1989) within the distance framework to transform the joint probabilities into the marginal probabilities and the association structure.

3.3.1 The 2-dimensional IPC Model

In this section, we show the representation of both the marginal probabilities and the association structure by a 2-dimensional IPC model. The class point matrix introduced in equation (3.5) will be used here with an additional restriction imposed on one of the free class points. That is, $\gamma_{32} = 1$ so that the first dimension pertains to a logistic regression of the first response and the second dimension to a logistic regression of the second response (i.e., no further scaling is required).

Representation of the Marginal Probabilities

Let us first show how the marginal probabilities of the two binary responses are represented by the 2-dimensional IPC model. The joint probability as defined by the IPC model in equation (3.1) will be used to define the marginal probabilities, that is,

$$\begin{aligned}
\log\left[\frac{\pi_{i1\bullet}}{\pi_{i0\bullet}}\right] &= \log\left[\frac{\pi_{i4}+\pi_{i2}}{\pi_{i3}+\pi_{i1}}\right] \\
&= \log\left[\frac{\dfrac{\exp(-0.5\delta_{i4})}{\sum_h \exp(-0.5\delta_{ih})}+\dfrac{\exp(-0.5\delta_{i2})}{\sum_h \exp(-0.5\delta_{ih})}}{\dfrac{\exp(-0.5\delta_{i3})}{\sum_h \exp(-0.5\delta_{ih})}+\dfrac{\exp(-0.5\delta_{i1})}{\sum_h \exp(-0.5\delta_{ih})}}\right] \\
&= \log\left[\frac{\exp(-0.5\delta_{i4})+\exp(-0.5\delta_{i2})}{\exp(-0.5\delta_{i3})+\exp(-0.5\delta_{i1})}\right].
\end{aligned} \tag{3.15}$$

Let us write out the Euclidean distances δ_{ij} as defined in equation (3.2). The marginal model (3.15) becomes,

$$\log\left[\frac{\pi_{i1\bullet}}{\pi_{i0\bullet}}\right] = \log\left[\frac{\exp\left[\gamma_{41}(\eta_{i1}-0.5\gamma_{41})\right]\times\exp\left[\gamma_{42}(\eta_{i2}-0.5\gamma_{42})\right]+\exp[\eta_{i1}-0.5]}{\exp(\eta_{i2}-0.5)+1}\right]. \tag{3.16}$$

In this paper, we find it convenient to re-parametrize γ_{41} and γ_{42} in terms of two other parameters, i.e., $\gamma_{41} = 1+\phi_1$ and $\gamma_{42} = 1+\phi_2$. The ϕ-parameters represent the deviation of the last category from (1, 1). By setting $\phi_1 = \phi_2 = 0$, the above result (16) can be simplified to:

$$\begin{aligned}
\text{logit}[\pi_{i1\bullet}] &= \log\left[\frac{[\exp(\eta_{i1}-0.5)\times\exp(\eta_{i2}-0.5)]+\exp(\eta_{i1}-0.5)}{\exp(\eta_{i2}-0.5)+1}\right] \\
&= \log\left[\frac{\exp(\eta_{i1}-0.5)\times[\exp(\eta_{i2}-0.5)+1]}{\exp(\eta_{i2}-0.5)+1}\right] \\
&= \eta_{i1}-0.5 \\
&= (\beta_{01}-0.5)+\boldsymbol{\beta}_1^{\mathsf{T}}\mathbf{x}_i \\
&= \beta_{01}^{*}+\boldsymbol{\beta}_1^{\mathsf{T}}\mathbf{x}_i.
\end{aligned} \tag{3.17}$$

Similarly,

$$\begin{aligned}
\log\left[\frac{\pi_{i\bullet 1}}{\pi_{i\bullet 0}}\right] &= \log\left[\frac{\pi_{i4}+\pi_{i3}}{\pi_{i2}+\pi_{i1}}\right] \\
&= \log\left[\frac{\dfrac{\exp(-0.5\delta_{i4})}{\sum_h\exp(-0.5\delta_{ih})}+\dfrac{\exp(-0.5\delta_{i3})}{\sum_h\exp(-0.5\delta_{ih})}}{\dfrac{\exp(-0.5\delta_{i2})}{\sum_h\exp(-0.5\delta_{ih})}+\dfrac{\exp(-0.5\delta_{i1})}{\sum_h\exp(-0.5\delta_{ih})}}\right] \\
&= \log\left[\frac{\exp[\gamma_{41}(\eta_{i1}-0.5\gamma_{41})]\times\exp[\gamma_{42}(\eta_{i2}-0.5\gamma_{42})]+\exp[\eta_{i2}-0.5]}{\exp[\eta_{i1}-0.5]+1}\right].
\end{aligned} \tag{3.18}$$

By setting $\phi_1 = \phi_2 = 0$ a straightforward marginal model is obtained, $\text{logit}[\pi_{i\bullet 1}] = (\beta_{02}-0.5)+\boldsymbol{\beta}_2^{\mathsf{T}}\mathbf{x}_i = \beta_{02}^{*}+\boldsymbol{\beta}_2^{\mathsf{T}}\mathbf{x}_i$; and, thus we call this the fixed class case. Without the constraints on the ϕ-parameters, the marginal models in (3.16) and (3.18) can not

be simplified further.

Representation of the Association

The odds ratio is defined in terms of the joint probabilities as shown in (3.13). Let us rewrite the probabilities in terms of the IPC model as in equation (3.1); that is,

$$\begin{aligned}
\log(\tau_i) &= \log\left[\frac{\pi_{i4}\times\pi_{i1}}{\pi_{i2}\times\pi_{i3}}\right] \\
&= \log\left[\frac{\frac{\exp(-0.5\delta_{i4})}{\sum_h \exp(-0.5\delta_{ih})}\times\frac{\exp(-0.5\delta_{i1})}{\sum_h \exp(-0.5\delta_{ih})}}{\frac{\exp(-0.5\delta_{i2})}{\sum_h \exp(-0.5\delta_{ih})}\times\frac{\exp(-0.5\delta_{i3})}{\sum_h \exp(-0.5\delta_{ih})}}\right] \\
&= \log\left[\frac{\exp(-0.5\delta_{i4})\times\exp(-0.5\delta_{i1})}{\exp(-0.5\delta_{i2})\times\exp(-0.5\delta_{i3})}\right] \\
&= 0.5\times[\delta_{i2}+\delta_{i3}-\delta_{i4}-\delta_{i1}]. \qquad (3.19)
\end{aligned}$$

This result implies that the differences between pairs of squared Euclidean distances correspond to the log-odds ratio. The distances can be written out and the association model becomes,

$$\log(\tau_i) = \phi_1\times(\eta_{i1}-1)+\phi_2\times(\eta_{i2}-1)-0.5*(\phi_1^2+\phi_2^2). \qquad (3.20)$$

In the case of $\phi_1=\phi_2=0$, $\log(\tau_i)=0$ which is equal to $\tau_i=1$. An odds ratio of unity indicates no association between the two binary responses, i.e., independence.

3.3.2 The 3-dimensional IPC Model

We now show the representation of the marginal probabilities and the association structure in a 3-dimensional IPC model. The class point introduced in equation (3.4) will be used

in the next derivations of the 3-dimensional IPC model.

Representation of the Marginal Probabilities

We follow the same derivation as before, but now the joint probabilities are defined in the 3-dimensional Euclidean space. For the marginal probabilities, we have

$$\begin{aligned}
\log\left[\frac{\pi_{i1\bullet}}{\pi_{i0\bullet}}\right] &= \log\left[\frac{\pi_{i4}+\pi_{i2}}{\pi_{i3}+\pi_{i1}}\right] \\
&= \log\left[\frac{\frac{\exp(-0.5\times\delta_{i4})}{\sum_h \exp(-0.5\times\delta_{ih})}+\frac{\exp(-0.5\times\delta_{i2})}{\sum_h \exp(-0.5\times\delta_{ih})}}{\frac{\exp(-0.5\times\delta_{i3})}{\sum_h \exp(-0.5\times\delta_{ih})}+\frac{\exp(-0.5\times\delta_{i1})}{\sum_h \exp(-0.5\times\delta_{ih})}}\right] \\
&= \log\left[\frac{\exp\left[\eta_{i1}+\eta_{i2}+\eta_{i3}-(3/2)\right]+\exp\left[\eta_{i1}-0.5\right]}{\exp\left[\eta_{i2}-0.5\right]+1}\right]. \quad (3.21)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\log\left[\frac{\pi_{i\bullet 1}}{\pi_{i\bullet 0}}\right] &= \log\left[\frac{\pi_{i4}+\pi_{i3}}{\pi_{i2}+\pi_{i1}}\right] \\
&= \log\left[\frac{\frac{\exp(-0.5\times\delta_{i4})}{\sum_h \exp(-0.5\times\delta_{ih})}+\frac{\exp(-0.5\times\delta_{i3})}{\sum_h \exp(-0.5\times\delta_{ih})}}{\frac{\exp(-0.5\times\delta_{i2})}{\sum_h \exp(-0.5\times\delta_{ih})}+\frac{\exp(-0.5\times\delta_{i1})}{\sum_h \exp(-0.5\times\delta_{ih})}}\right] \\
&= \log\left[\frac{\exp\left[\eta_{i1}+\eta_{i2}+\eta_{i3}-(3/2)\right]+\exp\left[\eta_{i2}-0.5\right]}{\exp\left[\eta_{i1}-0.5\right]+1}\right]. \quad (3.22)
\end{aligned}$$

It is not possible to simplify the above formulas further because of the parameters η_{i3}. Compared to the 2-dimensional IPC model with *fixed* class point, the marginal models

are not clearly represented in the 3-dimensional IPC models.

Representation of the Association

Using the formula derived in equation (3.19), but with the distances defined in three dimensions, the association model becomes

$$\begin{aligned}
\log[\tau_i] &= 0.5 \times [\delta_{i2} + \delta_{i3} - \delta_{i4} - \delta_{i1}] \\
&= 0.5 \times \left\{ \left[\sum_{m=1}^{3} (\eta_{im}^2 - 2\eta_{i1} + 1) \right] + \left[\sum_{m=1}^{3} (\eta_{im}^2 - 2\eta_{i2} + 1) \right] \right. \\
&- \left. \left[\sum_{m=1}^{3} (\eta_{im}^2 - 2\eta_{i1} - 2\eta_{i2} - 2\eta_{i3} + 3) \right] - \left[\sum_{m=1}^{3} \eta_{im}^2 \right] \right\} \\
&= \eta_{i3} - 0.5. \qquad (3.23)
\end{aligned}$$

This result proves that the 3-dimensional IPC model represents the association structure where the third dimension uniquely pertains to the association model.

3.3.3 Discussion

We studied both 2- and 3-dimensional IPC models in terms of marginal probabilities and association structure of bivariate binary data in the presence of predictors. We showed that both models with a specific class point specification are able to recover either the marginal probabilities or the association structure. That is, the 2-dimensional IPC model with *fixed* class point, i.e., $\phi_1 = \phi_2 = 0$, is equivalent to the marginal model with an *independence* association structure. In the case of a 3-dimensional model, the association structure is represented by the third dimension.

Based on the results of Section 3.1.1 and 3.1.2, we showed that a 2-dimensional IPC model with *fixed* class points, i.e., $\gamma_{41} = \gamma_{42} = 1$, represents a marginal model with an *independence* association structure. Each of the dimensions in the IPC model is related to one of the two binary responses. As shown in equation (3.20), the 2-dimensional IPC

model with free ϕ-parameters represents the association structure by a mixture of the marginal parameters and the ϕ-parameters.

According to the analytical results shown in equations (3.16) and (3.18), the marginal models can not be further simplified unless $\phi_1 = \phi_2 = 0$. When $\phi_1 \neq 0$ and $\phi_2 \neq 0$, neither the marginal model nor the association structure is well represented. At this stage, however, we do not know whether the IPC model is capable of recovering the models for the marginal probabilities and the association structure; therefore, we conducted a simulation study.

3.3.4 Simulation Study

We were able to show mathematically the performance of both the 2-dimensional IPC model with *fixed* class point, denoted by `IPC(2D-FIXED)`, and the 3-dimensional IPC model, denoted by `IPC(3D)`, in representing the marginal probabilities and the association structure for bivariate binary data. The analytical derivation under the 2-dimensional IPC model with *free* class points, denoted by `IPC(2D-FREE)`, however, was cumbersome. We conducted a simulation study to fully understand to what degree the `IPC(2D-FREE)` model recovers the marginal models and/ or the association model.

Data-generating Model

Bivariate binary data were generated from a Bivariate Logistic Regression model (Palmgren, 1989). The data generating model for the marginal probabilities is defined as follows,

$$\begin{aligned} \text{logit}[\pi_{i1\bullet}] &= \beta_{01} + \beta_{11}X_{1i} + \beta_{21}X_{2i} + \beta_{31}X_{3i} + \beta_{41}X_{4i} + \beta_{51}X_{5i}, \\ \text{logit}[\pi_{i\bullet 1}] &= \beta_{02} + \beta_{12}X_{1i} + \beta_{22}X_{2i} + \beta_{32}X_{3i} + \beta_{42}X_{4i} + \beta_{52}X_{5i}. \end{aligned} \tag{3.24}$$

We set $(\beta_{01}, \beta_{02}) = (-2.20, -1.50)$; $(\beta_{11}, \beta_{12}) = (0.00, -0.25)$; $(\beta_{21}, \beta_{22}) = (0.20, 0.00)$; $(\beta_{31}, \beta_{32}) = (-0.15, -0.15)$; $(\beta_{41}, \beta_{42}) = (1.05, 1.15)$; and $(\beta_{51}, \beta_{52}) = (-0.45, -0.15)$.

To generate data we need a representation of the association structure, i.e., $\log[\tau_i] = \beta_{03} + \beta_{13}X_{1i} + \beta_{23}X_{2i} + \beta_{33}X_{3i} + \beta_{43}X_{4i} + \beta_{53}X_{5i}$. In the 2-dimensional IPC model, the association structure is defined in terms of the other parameters as shown in (3.20). That is, $\beta_{03}^{*} = \phi_1 \times \beta_{01} + \phi_2 \times \beta_{02} - 0.5 \times \phi_1^2 - 0.5 \times \phi_2^2 - \phi_1 - \phi_2$ and $\beta_{k3}^{*} = \phi_1 \times \beta_{k1} + \phi_2 \times \beta_{k2}$, where $k = 1, 2, \ldots, 5$. Therefore, the data generating model for the association is $\log[\tau_i] = \beta_{03}^{*} + \beta_{13}^{*}X_{1i} + \beta_{23}^{*}X_{2i} + \beta_{33}^{*}X_{3i} + \beta_{43}^{*}X_{4i} + \beta_{53}^{*}X_{5i}$. We set $\phi_1 = -0.20$ and $\phi_2 = -0.45$; thus, the association parameters become $\beta_{03}^{*} = 1.65$ and $\boldsymbol{\beta}_{k3}^{*} = (0.10, -0.05, 0.10, -0.70, 0.15)$.

Four of the predictors were generated from the standard normal distribution, $X_{qi} \sim N(0, 1)$ where $q = 2, \ldots, 5$, and one from a binomial distribution, i.e., $X_{1i} \sim \text{BIN}(0.67)$. The VGAM package in the R software was used for generating the bivariate binary data (Yee, 2010).

Design and Analysis

A sample size of $N = 500$ was used in the simulations and each simulation was replicated $R = 1000$ times to obtain the sampling distributions of model parameters.

The performance of the proposed methods was evaluated by Bias (B), Root Mean Squared Error (RMSE), and Coverage. The bias of a parameter is defined as the difference between true value and the average of estimated values, i.e., $\text{B}(\hat{\beta}) = \bar{\hat{\beta}} - \beta$, with

$$\bar{\hat{\beta}} = \sum_{r=1}^{1000} \hat{\beta}_r / 1000,$$

and $\hat{\beta}_r$ is the estimate obtained from $r-$th replication. The RMSE is defined as

$$\text{RMSE} = \sqrt{\sum_{r=1}^{1000} \Big[(\hat{\beta}_r - \beta)^2 / 1000\Big]}.$$

Finally, the coverage is defined as the proportion of times the $100(1-\alpha)\%$ confidence

interval (CI) includes the true β value, where α corresponds to the nominal level of significance. The CI is defined as $[\hat{\beta}_r \pm Z_{1-\alpha/2}\widehat{\mathsf{SE}}(\hat{\beta}_r)]$ in which SE stands for the standard error of a parameter.

Simulation Study Results

The simulation results of the 2- and 3-dimensional IPC models are summarized in Table 3.2. The results for `IPC(2D-FIXED)` are given in columns 4-6, for `IPC(3D)` in columns 7-9, and for `IPC(2D-FREE)` in the last three columns. Because we showed analytically that the marginal models are represented well by the 2-dimensional fixed IPC model, and the association structure is represented well by the 3-dimensional IPC model, we focus here on the contrast of the 2-dimensional free model with the other two.

Compared to the `IPC(2D-FIXED)` results, marginal parameters under the `IPC(2D-FREE)` model were more biased. Specifically, two of the effects (i.e., X_2 and X_4) including the intercept, were poorly estimated. More specifically, $\mathsf{B}(\beta_{21}) = 0.037$ is about nine times bigger compared to the `IPC(2D-FIXED)` result, $\mathsf{B}(\beta_{22}) = -0.016$, $\mathsf{B}(\beta_{41}) = 0.106$, and $\mathsf{B}(\beta_{42}) = 0.050$ which all are about three times bigger than those obtained from the `IPC(2D-FIXED)`. All the RMSE results for the `IPC(2D-FREE)` model were higher than those obtained from the `IPC(2D-FIXED)` model. The coverage of the marginal parameters by the `IPC(2D-FREE)` model, compared to the former results, seems promising. However, both the intercepts and some of the effects were not covered well (i.e., β_{01}: 85.2%; β_{02}: 91.0%; β_{21}: 92.5%; β_{41}: $92.6\%\%$; β_{52}: 91.9%). Unlike the marginal parameters, the association parameters were fairly well estimated by the `IPC(2D-FREE)`. This is evident if we compare the results of the association parameters under the `IPC(2D-FREE)` and the `IPC(3D)` models.

Table 3.2: Summarized results of the simulation study for studying the performance of the IPC model for analysing bivariate binary data. `IPC(2D-FIXED)` corresponds to the 2-dimensional IPC model with *fixed* class points, i.e., $\phi_1 = \phi_2 = 0$; `IPC(3D)` to the 3-dimensional IPC model; and `IPC(2D-FREE)` to the 2-dimensional IPC model with *free* class points.

			IPC (2D-FIXED)			IPC (3D)			IPC (2D-FREE)*		
Effect	Parameter	True	Bias	RMSE	Coverage	Bias	RMSE	Coverage	Bias	RMSE	Coverage
Intercept	β_{01}	-2.20	-0.083	0.337	96.3	-0.487	0.652	88.3	-0.461	0.617	85.2
	β_{02}	-1.50	-0.044	0.260	94.8	-0.236	0.368	91.5	-0.236	0.362	91.0
	β_{03}	1.65	—	—	—	-0.045	0.786	94.4	0.020	0.586	94.8
X_1	β_{11}	0.00	0.018	0.373	94.4	0.079	0.486	95.9	0.040	0.456	94.9
	β_{12}	-0.25	-0.008	0.287	96.0	-0.024	0.335	95.7	-0.020	0.323	95.2
	β_{13}	0.10	—	—	—	-0.076	0.717	96.2	-0.031	0.411	98.9
X_2	β_{21}	0.20	0.004	0.174	93.0	0.031	0.228	94.2	0.037	0.215	92.5
	β_{22}	0.00	-0.006	0.144	93.9	-0.026	0.163	95.9	-0.016	0.158	94.8
	β_{23}	-0.05	—	—	—	0.001	0.372	93.0	-0.035	0.238	95.8
X_3	β_{31}	-0.15	-0.009	0.167	95.2	-0.011	0.206	95.5	-0.015	0.195	95.5
	β_{32}	-0.15	-0.005	0.136	96.3	-0.004	0.160	96.1	-0.009	0.151	95.9
	β_{33}	0.10	—	—	—	-0.027	0.341	96.6	-0.007	0.172	98.9
X_4	β_{41}	1.05	0.033	0.198	94.7	0.065	0.255	95.1	0.106	0.270	92.6
	β_{42}	1.15	0.019	0.178	94.6	0.034	0.207	94.2	0.050	0.201	95.4
	β_{43}	-0.70	—	—	—	0.083	0.430	93.0	-0.023	0.308	95.6
X_5	β_{51}	-0.45	-0.001	0.163	96.4	-0.032	0.212	96.0	-0.040	0.205	95.9
	β_{52}	-0.15	-0.004	0.149	93.5	0.033	0.175	92.9	0.012	0.173	91.9
	β_{53}	0.15	—	—	—	-0.034	0.352	95.8	0.035	0.240	96.9

$\beta^*_{03} = \phi_1 \times \beta_{01} + \phi_2 \times \beta_{02} - 0.5 \times \phi_1^2 - 0.5 \times \phi_2^2 - \phi_1 - \phi_2$; $\quad \beta^*_{k3} = \phi_1 \times \beta_{k1} + \phi_2 \times \beta_{k2}$, where $k = 1, 2, \ldots, 5$.

3.3.5 Summary of Study-1

De Rooij (2009a) studied IPC model for categorical data and showed its equivalence to logistic regression models. It was shown that the MBCL model is equivalent to the IPC model in maximum dimensionality. These models represent the joint probabilities.

In this Section we studied properties of the IPC model and extended it for analysing bivariate binary data, focusing on the marginal probabilities and the association structure. We showed their connection both mathematically and using a simulation study. We found that a 2-dimensional IPC model with *fixed* class point (i.e., $\phi_1 = \phi_2 = 0$) represents the marginal models with an *independence* association structure. We also found that a 3-dimensional IPC model with a specific class point configuration represents the association model in the third dimension.

We also studied the performance of a 2-dimensional IPC model with *free* class point. Since its analytical part was cumbersome, we conducted a simulation study to see if it can recover both the marginal models and the association model. This model represents the association model well, but the marginal models were misspecified. Therefore, we conclude that a given IPC model can recover either the marginal models or the association model of bivariate binary data, but not both of them at the same time.

3.4 Study-2: The Bivariate IPC Model

In the first study, we investigated properties of the standard IPC models for the representation of both the marginal probabilities and the association structure. It was concluded that a given IPC model is not able to represent both types of the models at the same time. In this section, we re-parametrize the IPC model in order to provide a better representation of both the marginal probabilities and the association structure.

Bahadur (1961) proposed a full likelihood-based marginal model for bivariate binary data by characterizing the multinomial probabilities in terms of both the marginal prob-

abilities and the correlation coefficient between the two responses (Y_{i1} and Y_{i2}). Lipsitz, Laird and Harrington (1990) followed the Bahadur (1961) approach and showed that other measures of association, such as the odds ratio and the relative risk, can also be used.

In this second study, our aim is to adopt the Lipsitz, Laird and Harrington (1990) approach into the IPC model framework for better representation of the required statistical models. As shown in equation (3.3), parameter estimation under the IPC model is based on the multinomial likelihood function. To avoid confusion with the former IPC model presented in Section 2.1, we refer the Bahadur-based IPC model as the Bivariate IPC (BIPC) model.

In the BIPC model framework, the Euclidean distance defined in equation (3.2) will be used only to define the joint probabilities which are related to the association structure. For defining the marginal models, we use another Euclidean distance definition emphasizing the marginal models. That is,

$$\begin{aligned} \pi_{i1\bullet} &= \frac{\exp(-0.5\delta_{i1\bullet})}{\exp(-0.5\delta_{i0\bullet}) + \exp(-0.5\delta_{i1\bullet})}; \\ \pi_{i\bullet 1} &= \frac{\exp(-0.5\delta_{i\bullet 1})}{\exp(-0.5\delta_{i\bullet 0}) + \exp(-0.5\delta_{i\bullet 0})}, \end{aligned} \tag{3.25}$$

where $\delta_{il\bullet} = \sum_{m=1}^{2} (\eta_{im} - \gamma_{l\bullet m})^2$ and $\delta_{i\bullet l} = \sum_{m=1}^{2} (\eta_{im} - \gamma_{\bullet lm})^2$, $l = 0, 1$. As shown in Appendix B, the class points of the BIPC model are defined as

$$\boldsymbol{\gamma_1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

and

$$\boldsymbol{\gamma_2} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

where $\boldsymbol{\gamma_1}$ is the class point matrix that corresponds to the first response variable and $\boldsymbol{\gamma_2}$

to the second response variable.

The first step according to Bahadur (1961) is to rewrite the the association structure between the two binary responses, i.e., the odds ratio in our case, using the marginal probabilities. That is,

$$\tau_i = \frac{\pi_{i4} \times \pi_{i1}}{\pi_{i2} \times \pi_{i3}} = \frac{\pi_{i4} \times (1 - \pi_{i1\bullet} - \pi_{i\bullet 1} + \pi_{i4})}{(\pi_{i1\bullet} - \pi_{i4}) \times (\pi_{i\bullet 1} - \pi_{i4})}. \tag{3.26}$$

We showed in (3.19) that given the IPC model, the odds ratio can be defined in terms of Euclidean distances, i.e., $\tau_i = \exp[0.5 \times (\delta_{i2} + \delta_{i3} - \delta_{i1} - \delta_{i4})]$. We will use this representation as the defining characteristics of the association in the BIPC model. With free class points, i.e., $\phi_1 \neq 0$ and $\phi_1 \neq 0$, the odds ratio becomes,

$$\tau_i = \exp[\phi_1 \times (\eta_{i1} - 0.5\phi_1 - 1) + \phi_2 \times (\eta_{i2} - 0.5\phi_2 - 1)]. \tag{3.27}$$

We can then replace τ_i in (3.26) by (3.27), and solve the quadratic equation to get solutions for π_{i4} (Mardia, 1967). The valid solution for π_{i4} is,

$$\pi_{i4} = \begin{cases} \dfrac{w_i - \{w_i^2 - 4\exp(a_i)[\exp(a_i) - 1]\pi_{i1\bullet}\pi_{i\bullet 1}\}^{1/2}}{2[\exp(a_i) - 1]} & \text{if } a_i \neq 0 \\ \pi_{i1\bullet} \times \pi_{i\bullet 1} & \text{if } a_i = 0, \end{cases} \tag{3.28}$$

where $w_i = 1 - [1 - \exp(a_i)][\pi_{i1\bullet} + \pi_{i\bullet 1}]$ and $a_i = \phi_1 \times (\eta_{i1} - 0.5\phi_1 - 1) + \phi_2 \times (\eta_{i2} - 0.5\phi_2 - 1)$.

The final step is to rewrite the joint probabilities in the multinomial likelihood in terms of the marginal probabilities and the association structure, i.e., $\pi_{i2} = \pi_{i1\bullet} - \pi_{i4}$

and $\pi_{i3} = \pi_{i\bullet 1} - \pi_{i4}$ in which π_{i4} will be replaced by (3.28). That is,

$$\boldsymbol{\pi}_i^* = \begin{bmatrix} \pi_{i4} \\ \pi_{i\bullet 1} - \pi_{i4} \\ \pi_{i1\bullet} - \pi_{i4} \end{bmatrix}. \tag{3.29}$$

This modified likelihood will be used for estimating the parameters of the BIPC model.

3.4.1 Simulation Study Results

The simulation results of the BIPC model are summarized in Table 3.3. We compare these results against those in Table 3.2, particularly the results from `IPC(2D-FIXED)` and `IPC(3D)` models.

The bias and RMSE results for the marginal parameters under the BIPC model are very close to those under the `IPC(2D-FIXED)` model, which proves that the BIPC model represents the marginal models well. Almost all the coverages of the marginal parameters were satisfactory, except two of the effects, one for β_{22} equal to 92.8% and for β_{52} equal to 92.2%. Their coverage by the `IPC(2D-FIXED)` model was 93.9% and 93.5%, respectively.

Compared to the results presented in Table 3.2 for `IPC(3D)`, the BIPC model produced smaller bias, except for two of the effects, i.e., $\text{B}(\beta_{43}) = -0.176$ and $\text{B}(\beta_{53}) = 0.087$. However, all the RMSEs under the BIPC model were smaller than those obtained from the IPC models. Almost all the parameters were covered well by the BIPC model, with a coverage above 95.0%. Compared to the IPC models, the BIPC model estimates are generally less biased, more accurate, and well covered parameters for both the marginal models and the association model.

We conclude that the BIPC model represents not only the marginal models, but also the association model for the analysis of bivariate binary data in the presence of predictors.

Table 3.3: Summarized results of the simulation study for studying the performance of the BIPC model for analysing bivariate binary data.

Effect	Parameter	True	Bias	RMSE	Coverage
Intercept	β_{01}	-2.20	-0.074	0.333	95.5
	β_{02}	-1.50	-0.048	0.262	94.9
	β^*_{03}	1.65	0.109	0.602	94.5
X_1	β_{11}	0.00	0.021	0.365	94.3
	β_{12}	-0.25	0.001	0.288	95.4
	β^*_{13}	0.10	-0.022	0.369	98.7
X_2	β_{21}	0.20	0.005	0.171	94.2
	β_{22}	0.00	-0.010	0.144	92.8
	β^*_{23}	-0.05	-0.044	0.230	95.1
X_3	β_{31}	-0.15	-0.007	0.162	95.2
	β_{32}	-0.15	0.004	0.137	96.1
	β^*_{33}	0.10	0.013	0.158	99.3
X_4	β_{41}	1.05	0.022	0.195	94.5
	β_{42}	1.15	0.009	0.179	93.1
	β^*_{43}	-0.70	-0.176	0.382	95.7
X_5	β_{51}	-0.45	0.007	0.162	96.5
	β_{52}	-0.15	0.001	0.149	92.2
	β^*_{53}	0.15	0.087	0.249	96.8

$\beta^*_{03} = \phi_1 \times \beta_{01} + \phi_2 \times \beta_{02} - 0.5 \times \phi_1^2 - 0.5 \times \phi_2^2 - \phi_1 - \phi_2$; $\quad \beta^*_{k3} = \phi_1 \times \beta_{k1} + \phi_2 \times \beta_{k2}$, where $k = 1, 2, \ldots, 5$.

3.5 Application

The NESDA data introduced earlier (Penninx et al., 2008), were analysed using the proposed distance models. The sample comprised of $N = 2,938$ subjects aged 18 to 65 years (Mean=42; S.D.=13.1). About 66.5% were female and the average number of years of education attained was 12.2 with S.D. $= 3.3$. The responses of interest were diagnoses of dysthmia (DYST: 1 if diseased; 0, otherwise) and generalized anxiety disorder (GAD: 1 if diseased; 0, otherwise). About 10.2% and 15.3% of the subjects in the study developed DYST and GAD, respectively.

One of the objectives of NESDA is to measure the effect of personality traits on the

risk of developing mental disorders (Spinhoven et al., 2009). We considered the Big-Five personality variables, i.e., Neuroticism (N), Extraversion (E), Openness to experience (O), Agreeableness (A), and Conscientiousness (C). We also took into account the background variables, i.e., age (AGE), years of educations attained (EDU), and gender (GEN: 1=female; 0=male). Both the personality traits and the background variables will be treated as predictors.

In the final fitted (B)IPC models, all background variables and two of the personality traits such as neuroticism and extraversion, are retained since the other traits (such as O, A and C) are not statistically significant on both dimensions.

Table 3.4: Parameter estimates with corresponding standard errors (between the parenthesis) obtained from the IPC and BIPC models fitted on the NESDA data. IPC(2D-IND) corresponds to the 2-dimensional IPC model with *fixed* class coordinates; IPC(2D-FREE) to the 2-dimensional IPC model with *free* class coordinates; and IPC(3D) to the 3-dimensional IPC model.

		Models			
Effect	Parameter	IPC(2D-FIXED)	IPC(2D-FREE)†	IPC(3D)	BIPC†
		Dysthymia			
Intercept	β_{01}	$-2.20(0.131)$	$-2.57(0.148)$	$-2.55(0.167)$	$-2.21(0.131)$
Gender	β_{11}	$-0.18(0.140)$	$-0.21(0.143)$	$-0.25(0.180)$	$-0.17(0.139)$
Age	β_{21}	$0.20(0.072)$*	$0.20(0.073)$*	$0.18(0.093)$*	$0.20(0.072)$*
Education	β_{31}	$-0.15(0.066)$*	$-0.17(0.067)$*	$-0.18(0.085)$*	$-0.15(0.065)$*
Neuroticism	β_{41}	$1.03(0.102)$*	$1.14(0.127)$*	$1.13(0.133)$*	$1.03(0.102)$*
Extraversion	β_{51}	$-0.46(0.085)$*	$-0.47(0.087)$*	$-0.47(0.11)$*	$-0.45(0.085)$*
		Generalized Anxiety Disorder			
Intercept	β_{02}	$-1.51(0.105)$	$-1.69(0.118)$	$-1.69(0.118)$	$-1.51(0.103)$
Gender	β_{12}	$-0.26(0.119)$*	$-0.31(0.136)$*	$-0.31(0.137)$*	$-0.26(0.117)$*
Age	β_{22}	$0.06(0.060)$	$0.03(0.068)$	$0.02(0.069)$	$0.05(0.059)$
Education	β_{32}	$-0.13(0.056)$*	$-0.14(0.064)$*	$-0.14(0.065)$*	$-0.12(0.055)$*
Neuroticism	β_{42}	$1.16(0.086)$*	$1.22(0.098)$*	$1.22(0.098)$*	$1.14(0.085)$*
Extraversion	β_{52}	$-0.15(0.070)$*	$-0.10(0.080)$	$-0.10(0.081)$	$-0.14(0.070)$*
		Association			
Intercept	β_{03}	—	$1.75(0.199)$	$2.19(0.274)$	$1.69(0.207)$
Gender	β_{13}	—	$0.23(0.116)$*	$0.30(0.281)$	$0.16(0.081)$*
Age	β_{23}	—	$-0.02(0.055)$	$0.01(0.145)$	$-0.06(0.043)$
Education	β_{33}	—	$0.10(0.051)$*	$0.14(0.133)$	$0.09(0.034)$*
Neuroticism	β_{43}	—	$-0.92(0.187)$*	$-0.89(0.211)$*	$-0.73(0.170)$*
Extraversion	β_{53}	—	$0.08(0.072)$	$0.07(0.169)$	$0.16(0.067)$*

† $\beta_{03} = \phi_1 \times \beta_{01} + \phi_2 \times \beta_{02} - 0.5 \times \phi_1^2 - 0.5 \times \phi_2^2 - \phi_1 - \phi_2$; $\beta_{k3} = \phi_1 \times \beta_{k1} + \phi_2 \times \beta_{k2}$, where $k = 1, 2, \ldots, 5$.

* statistically significant, i.e., $p < 0.05$.

3.5.1 The IPC Models

The results of 2- and 3-dimensional IPC models fitted on the NESDA data are shown in Table 3.4.

The 2-dimensional IPC Model

The 2-dimensional IPC model with *fixed* class points, which is a marginal model with an *independence* association structure, is presented in the third column of Table 3.4 and has a fit statistic of BIC $= 3,784.1$ with twelve parameters.

We found a strong positive effect of neuroticism on risk of developing both mental disorders, i.e., $\hat{\beta}_{41} = 1.03$ with DYST; and $\hat{\beta}_{42} = 1.16$ with GAD. This implies that on average neurotic (i.e., emotionally unstable) people have a higher chance of developing the mental disorders. The other personality trait with stronger effect was extraversion with a moderate negative effect, i.e., $\hat{\beta}_{51} = -0.46$ with DYST; and $\hat{\beta}_{52} = -0.15$ with GAD. Being an introvert (i.e., having lower social engagement) seems to increase the chance of developing the mental disorders.

Among the background variables, education was the only predictor with statistically significant association with both disorders, i.e., $\hat{\beta}_{31} = -0.15$ with DYST; and $\hat{\beta}_{32} = -0.13$ with GAD. That is, less educated people had a higher chance of developing the disorders. The other vulnerable groups were males (i.e., $\hat{\beta}_{12} = -0.26$ with GAD) and elders ($\hat{\beta}_{21} = 0.20$ with DYST).

The fourth column shows the results of the 2-dimensional IPC model with *free* class points; its fit statistics was BIC $= 3,723.6$ with fourteen parameters. The additional two parameters are due to the estimated class points, i.e., $\hat{\phi}_1 = -0.01$ and $\hat{\phi}_2 = -0.74$. The association parameters presented in the last row block of Table 3.4 under `IPC(2D-FREE)`, are not free parameters because they are estimated using the other parameters including the class coordinates as shown in equation (3.20). Gender, education, and neuroticism

had significant effect on the log-odds ratio, i.e., $\hat{\beta}_{13} = 0.23$, $\hat{\beta}_{33} = 0.10$ and $\hat{\beta}_{43} = -0.92$, respectively. Neuroticism had a negative strong effect on the log-odds ratio, which implies that the association between the two disorders became weaker when the level of neuroticism for a given person increased; and the rate of change was about 0.92 for a unit change in neuroticism. In the case of education, the direction was positive which implies that the association between the disorders became stronger when a person became more educated and the rate of change was about 0.10 for a unit change in education.

The results of `IPC(2D-FIXED)` and `IPC(2D-FREE)` models are not comparable as shown mathematically in Section 3.1. This is also evident if we compare the effect of extraversion under these models, i.e., $\hat{\beta}_{52} = -0.15$ under the `IPC(2D-FIXED)` model which is statistically significant, but it became insignificant under the `IPC(2D-FREE)` model, i.e., $\hat{\beta}_{52} = -0.10$.

The 3-dimensional IPC Model

The results of the 3-dimensional IPC model are presented in the fifth column that corresponds to `IPC(3D)` and its fit statistic was BIC $= 3,755.4$ with eighteen parameters. The first two row blocks of parameters under the `IPC(3D)` model have the same interpretation as the other models for the joint probabilities. Thus, we focus on the additional parameters that are displayed in the last row block, which corresponds to the association model as shown in equation (3.23).

It is important to note that these parameters are not comparable to those under the 2-dimensional IPC model, because the latter are specified in a lower-dimensional space and thus are restricted, while the former handles the association structure using separate parameters on third dimension. Only neuroticism had a significant effect on the log-odds ratio, i.e., $\hat{\beta}_{43} = -0.89$. This implies that the association between the two disorders became weaker when the level of neuroticism for a given person increased. The rate of change was about 0.89 for a unit change in neuroticism.

3.5.2 The BIPC Model

The last column of Table 3.4 shows the results from the BIPC model which had a fit statistic BIC $= 3,735.6$ with fourteen parameters. The first two row blocks display the marginal parameters. These results are equivalent to the `IPC(2D-FIXED)`, and thus they both have the same interpretation.

The last row block shows the parameters of the association model that are obtained using the other parameters and the estimated class points, i.e., $\hat{\phi}_1 = -0.21$ and $\hat{\phi}_2 = -0.46$. Except age, all the predictors were statistically significant in the association model. The effect of extraversion was $\hat{\beta}_{53} = 0.16$, which implies that the association between the two disorders became stronger when the level of extraversion increased. In the case of neuroticism, the effect was negative, $\hat{\beta}_{43} = -0.73$. Thus the more neurotic a person was the weaker the association between the disorders.

The results of the BIPC model can also be displayed using a biplot (Gower & Hand, 1996; Gower et al., 2011). Figure 1 displays the biplot for the final BIPC model in which only the predictors having significant effect on both dimensions are considered. The labels of the predictors are placed at the positive side of the variable axis. On the variable axes markers are placed that represent $\mu_X \pm t\sigma_X$, where μ_X is the mean of X, σ_X is the standard deviation and $t = 0, 1, 2, 3$. From the biplot it is evident that neuroticism had a strong association with both mental disorders because its variable axis is long. The second influential predictor was extraversion pointing to the reverse direction compared to neuroticism.

The axes of the biplot corresponds to the marginal models, i.e., the horizontal axis corresponds to DYST and the vertical axis to GAD. The angle between a variable axis and each axis of the biplot, can be used to evaluate the strength of their association, i.e., the smaller the angle the stronger the association between them. For example, the angle between extraversion and DYST is smaller compared to the angle between extraversion

and GAD, which indicates that the association between extraversion and dysthmia is stronger. This result is in line with the estimates shown in the last column of Table 3.4 under extraversion, i.e., $\hat{\beta}_{51} = -0.45$ with DYST and $\hat{\beta}_{52} = -0.14$ with GAD.

The effect of predictors on the association model can also be read from the biplot. We showed mathematically in Section 3.1 that the `IPC(2D-FIXED)` is a marginal model with an *independence* association structure. This would correspond to the spatial solution in the biplot if the last category was positioned at $(\gamma_{41}, \gamma_{42}) = (1, 1)$. In the biplot displayed in Figure 1, however, the last category was positioned at (0.79, 0.54) because $\hat{\phi}_1 = -0.21$ and $\hat{\phi}_2 = -0.46$. With every unit increase of neuroticism the log odds ratio of dysthimia and GAD changes by $\beta_{43} = \phi_1\beta_{41} + \phi_2\beta_{42}$. Both β_{41} and β_{42} were positive while ϕ_1 and ϕ_2 were negative. Therefore, with an increase of neuroticism the log odds ratio goes down. Along similar lines, we can derive that the log odds ratio increases with an increase of extraversion. These derivations show explicitly that the marginal model and the association structure are intuitively coupled, i.e., the same regression coefficients are used and only the ϕ-parameters can be used to adjust sign and strength. The adjustment by ϕ_1 and ϕ_2 is the same for every predictor variable.

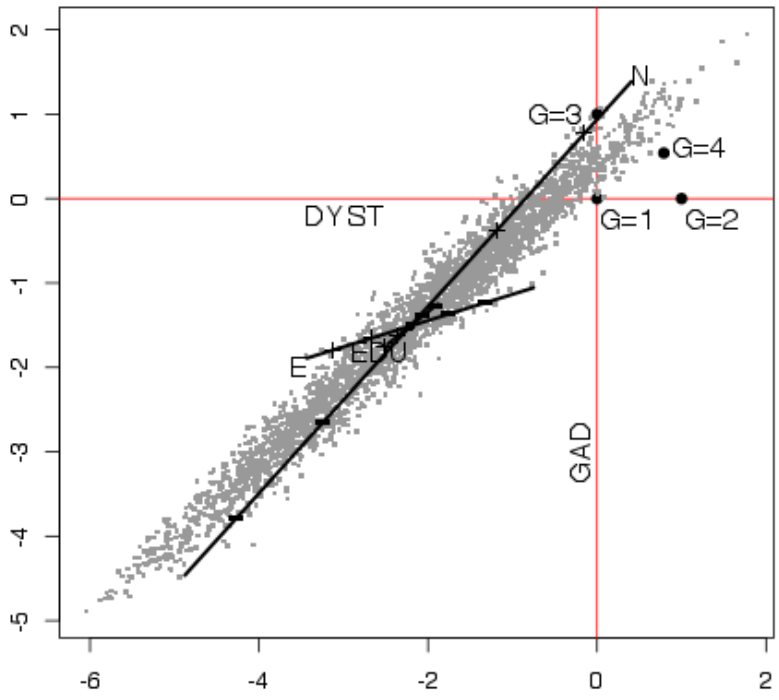

Figure 3.1: Biplot of the final BIPC model fitted on the NESDA data. The predictors neuroticism, represented by `N`; extraversion, by `E`; and education, by `EDU`. The bivariate binary responses are dysthmia, represented by `DYST`; and generalized anxiety disorder, by `GAD`. The class coordinates that correspond to the multinomial response variable, denoted by `G`, are also displayed.

3.6 Conclusion and Discussion

In this paper, we studied properties of the IPC model and extended it for analysing bivariate binary data in the presence of predictors, focusing on the marginal probabilities and the association structure. Researchers often model the marginal probability of an outcome variable without the influence of the other outcome variable. Such models are

referred as marginal models since the effect of the other outcome variable is marginalized. In addition to the marginal models, investigators are sometimes interested in modelling the association structure between the binary responses. It is expected that the two binary responses are correlated as they are measured on the same subject.

We found the following three results about the IPC model for analysing bivariate binary data. The 2-dimensional IPC model with *fixed* class point(`IPC(2D-FIXED)`) represents the marginal models with an *independence* association structure between the binary responses. Each dimension under the `IPC(2D-FIXED)` model pertains to one of the binary response variables. This result does agree with the finding by Liang and Zeger (1986) in which they showed that fitting a separate logistic regression model for each binary response variable gives consistent parameter estimates but biased standard errors. In the IPC model, however, the standard errors are not biased because estimation of model parameters are based on a multinomial likelihood function.

The 3-dimensional IPC model (`IPC(3D)`) represents the association structure in the third dimension. This model, however, misspecifies the models for the marginal probabilities. The compromise between the former two IPC models is a 2-dimensional IPC model with *free* class points (`IPC(2D-FREE)`). We showed, using simulation studies, that this latter model represents the association model as a form of restricted model. Like the `IPC(3D)` model, the `IPC(2D-FREE)` model misspecified the models for the marginal probabilities. Therefore, we conclude that the IPC model represents either the models for the marginal probabilities or the model for the association structure, but not both of them at the same time.

We therefore considered a possible extensions of the IPC model for representing both the marginal models and the association model at the same time. We modified the multinomial likelihood function following Bahadur (1961) and Lipsitz, Laird and Harrington (1990). The extended IPC model is called the Bivariate IPC (BIPC) model. Using simulation studies we showed that the BIPC model represented both the models for the

marginal probabilities and the model for the association structure well.

Unlike existing marginal models for bivariate binary data, the results of the BIPC model can be displayed graphically in a biplot which enhances the interpretation of the model. The axes in the biplot correspond to marginal models of the bivariate binary data, i.e., the horizontal axis corresponds to the first response variable and the vertical axis to the second response variable. The angle between the variable axis and each axis of the biplot is used to explain the strength of their association. In the same biplot, one can also read the relationship between a predictor variable and association structure (i.e., odds ratio). Therefore, we use both the ϕ-parameters and the marginal parameters to explain the direction and strength of their relationship. If both ϕ-parameters are found to be positive, it is an indication of a strong positive relationship between a predictor variable and the association structure. Similarly, an inverse relationship is characterized by the presence of negative estimates for both ϕ-parameters.

In this paper our focus was on application of the (B)IPC model for analysing bivariate binary data. Marginal modelling of multivariate polytomous type of responses has been an interest in social and other empirical sciences (Bergsma, 1997; Bergsma, Croon, & Hagenaars, 2009; Molenberghs & Lesaffre, 1994, 1999). The BIPC model can easily be extended for analysing bivariate polytomous responses by modifying the class coordinates to accomodate the additional response categories. At this stage, it is, however, not straight forward to extend the BIPC model for analysing multivariate binary responses. This is due to the fact that both the pairwise and higher-order association structure parameters must be specified in the likelihood function. With three binary responses (i.e., Y_1, Y_2, and Y_3), for example, three pairwise associations and a three-way association parameters must be specified which makes the computation cumbersome. If the interest is only on the pairwise association, the BIPC model for bivariate binary data can be extended by modifying the class point matrix.

We made the data and source codes (R / SAS) used in the simulation studies and in

the application available on the online repository system GitHub. The following link can be used to get access to the files: https://github.com/workuhm1/BIPCM.