2019-07-11

Joint modeling of clustered binary data with crossed random effects via the Gaussian copula mixed model

Jaman, Ajmery

UNIVERSITY OF CALGARY

Joint modeling of clustered binary data with crossed random effects via the Gaussian copula mixed model

by

Ajmery Jaman

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA
JULY, 2019

Abstract

Models with crossed random effects are common in reader-based diagnostic studies, where the same group of readers evaluate patients for certain diseases; an example is diabetic retinopathy study in Alberta, Canada. Although generalized linear mixed models (GLMMs) are well developed for non-Gaussian responses (e.g., binary outcomes) with crossed random effects, evaluation of the marginal likelihood is still technically and computationally demanding and can become prohibitive in applications, since the data cannot be grouped into independent blocks. The available estimation methods are also not free from problems. A recent approach involves application of data cloning (DC) to obtain maximum likelihood (ML) estimates using a Bayesian framework. Their approach is proved to be superior over the other two alternatives they considered in terms of providing relatively unbiased and efficient parameter estimates. However, this approach is based on a multivariate latent Gaussian description of the multiple correlated binary outcomes. In this thesis, we relax this assumption by allowing for disparate non-Gaussian latent variables for the binary responses, and propose a joint modeling via the Gaussian copula mixed model (GCMM). We applied maximum pairwise likelihood (PL) estimation instead of doing full ML analysis to reduce computational complexities. We conducted simulation studies with a setting analogous to the diabetic retinopathy data to see the performance of PL estimators for GCMM with crossed random effects. Simulation results suggest that although the estimation of regression coefficients and correlation parameter exhibit no problem, a much bigger sample size is required for the other scale parameters to provide reasonably accurate approximate results. We also analyzed the retinopathy data with the proposed approach considering three different conditional margins.

Preface

I, hereby, declare that the work presented in this thesis is the outcome of the investigation performed by me under the supervision of Dr. Alexander de Leon, Associate Professor, Department of Mathematics and Statistics, University of Calgary. I also declare that no part of this thesis has been or is being submitted elsewhere for the award of any degree or diploma.

Acknowledgements

It gives me a great pleasure to express my sincere gratitude and deepest appreciation to my supervisor Dr. Alexander de Leon. In spite of his busy schedule, I was fortunate enough to have his exhaustive guidance for the current work to its competition. His effective supervision allow me to select a topic which has been an active field of research over the past few years, decide how far I would conclude; which made my work very flexible. His valuable suggestions, constant inspection helped me a great deal to complete this dissertation in time.

Besides my supervisor, I thank the members of the examination committee: Dr. Jingjing Wu, Dr. Bingrui (Cindy) Sun and Dr. Thuntida Ngamkham, for agreeing to read my thesis, for providing suggestions, and for making themselves available for my final oral examination.

I am also thankful to the Department of Mathematics and Statistics for their generous financial support through graduate assistantships.

Finally, I would like to thank my parents, and specially my husband for providing his support to me during the last two years of my masters.

To my unborn child...

Table of Contents

List of Tables

Chapter 1

Introduction

This thesis is motivated by the diabetic retinopathy study (De Leon et al., 2007; de Leon et al., 2009) introduced by Rudnisky et al. (2002), where patient- and reader-specific effects are crossed rather than nested. In an experiment, if every level of one effect co-occurs with every level of the other effect, then the two effects are considered crossed rather than nested. Nested effects are common, for example, in repeated measures data, where we make multiple observations on an individual, so that an individual's measurements are nested within the individual. Models with crossed random effects are common in psycholinguistic studies, where the experimenter presents multiple test items to multiple participants, and a particular sample of participants responds to the same test items. The same scenario is obtained in reader-based diagnostic studies, where the same group of readers evaluate patients for certain diseases. For several decades, it has been common practice to analyze such data by means of 2 analyses of variance (ANOVAs). In each ANOVA, one factor is considered as random and the solution is based on 2 F-statistics. The null hypothesis is rejected if both analyses showed significant F-values. However, these F-statistics are biased when the two factors are sampled randomly (Raaijmakers et al., 1999; Baayen et al., 2008). Clark (1973) discussed this issue and proposed a minimum-F statistic derived from separate F-statistics. However, this procedure can be too conservative (Raaijmakers et al., 1999). With the rela-

tively recent development of software for implementing mixed-effects models in applications, solving these types of problems became quite straightforward. Furthermore, the problem in the Gaussian setting becomes even more straightforward, as the responses marginally follows the multivariate Gaussian distribution. However, although generalized linear mixed models (GLMMs) are well developed for non-Gaussian responses (e.g., binary outcomes) with crossed random effects (see Section 1.1), evaluation of the marginal likelihood is still technically and computationally demanding and can become prohibitive in applications, since the data cannot be grouped into independent blocks. This has led various authors to propose a host of alternative estimation methods rather than carrying out full maximum likelihood (ML) analysis.

In the context of continuous outcomes, Verbeke et al. (2001) used conditional linear mixed models to estimate parameters as well as to calculate precision estimates in crossed random effects models. The major advantage of their approach is that, by appropriate conditioning, the original model maps into two hierarchical ones for which conventional and hence, computationally efficient and fast techniques can be used. Perhaps, the main weakness of the approach is that no cross-sectional effects can be estimated. Nevertheless, the method is applicable when within-cluster effects and variance components are of interest. Tibaldi et al. (2007) extended Verbeke's method (Verbeke et al., 2001) to binary data combining ideas from conditional logistic regression with composite likelihood estimation.

A well-known data set for correlated binary data with crossed random effects is the salamander mating data introduced by McCullagh and Nelder (1989). To obtain ML estimates, Chan and Kuk (1997) applied a Monte Carlo EM (expectation-maximization) algorithm, with the M-step greatly simplified under the assumption of a probit link and its E-step made feasible by Gibbs sampling. Booth and Hobert (1999) likewise applied a Monte Carlo EM algorithm, which used importance sampling to construct the Monte Carlo approximations at the E-step, with sample generation based on the exact distribution of the random effects given the data. However, as with other computationally intensive approaches, these

methods require significantly more computational time than conventional ones, and are likely to be beset with convergence problems. Jiang (1998) and Sutradhar and Rao (2003) proposed methods based on suitable estimating equations, which provide consistent estimates for the fixed parameters and variance components in GLMMs. However, these estimates generally suffer from some efficiency loss relative to ML estimates; in addition, these estimating equations are non-linear and generally yield multiple roots. Breslow and Clayton (1993) used penalized quasi-likelihood (PQL) approach, but their approach has been shown to yield inconsistent estimates with non-negligible bias, especially for binary data. Torabi (2012) adapted Lele *et al.*'s (Lele et al., 2007, 2010) data cloning (DC) approach — a method that provides ML estimates under a Bayesian framework — to a correlated two-factor model. Withanage et al. (2015) applied the DC approach to the diabetic retinopathy data considering a GLMM, more specifically, a multivariate probit model. Their approach proved to be superior over the other two alternatives (Laplace approximation and PQL) they considered in terms of providing relatively unbiased and efficient parameter estimates.

Bellio and Varin's (Bellio and Varin, 2005) pairwise likelihood approach provided another alternative approach to full ML estimation for the multivariate probit model with crossed random effects. Pairwise likelihood estimation has been shown to result in minimal efficiency loss (Renard et al., 2004). Wu and de Leon (2014) introduced the Gaussian copula mixed model (GCMM), a general mixed model for clustered data, where, in particular, the random effects are nested within clusters. In this thesis, we adapted the GCMM methodology to the setting of the diabetic retinopathy data and useed Bellio and Varin's (Bellio and Varin, 2005) pairwise likelihood approach instead of doing the full ML estimation.

1.1 Mixed models with crossed random effects

Conventional linear mixed models (LMMs) and generalized linear mixed models (GLMMs) for Gaussian and non-Gaussian responses can be viewed as special cases of GCMMs. We

discuss them in the sequel.

1.1.1 LMM

Let Y_{ijk} be the clustered continuous outcome for reader $k = 1, \ldots, K$ and eye $j = 1, 2$ in cluster (subject) $i = 1, \cdots, N$. We consider a model with two crossed random effects. The conventional LMM set-up assumes (see e.g. Verbeke et al. (2001))

$$Y_{ijk} = \mu_{ijk}(B_{1i}, B_{2k}) + \varepsilon_{ijk} = \mathbf{x}_{ijk}^{\top}\boldsymbol{\beta} + B_{1i} + B_{2k} + \varepsilon_{ijk}, \tag{1.1}$$

where $\mathbf{x}_{ijk}$ is known vector of covariates, $\boldsymbol{\beta}$ is the vector of regression coefficients, B_{1i} is the cluster-specific random effect, B_{2k} is the reader-specific random effect that occur in every cluster, and ε_{ijk} is residual error, usually assumed independent of B_{1i} and B_{2k}. Letting $\mathbf{B}_i = (B_{1i}, B_{21}, \ldots, B_{2K})^{\top}$ be a $(K+1)$-dimensional vector and $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijK})^{\top}$, the conventional LMM for Y_{ijk} assumes

$$\mathbf{B}_i \sim N_{K+1}\left(\mathbf{0}, \boldsymbol{\Sigma}^{(B)}\right), \tag{1.2}$$

$$\boldsymbol{\varepsilon}_{ij} \sim N_K\left(\mathbf{0}, \boldsymbol{\Sigma}^{(\varepsilon)}\right), \tag{1.3}$$

so that the respective conditional (on $\mathbf{B}_i$) and marginal distributions of $\mathbf{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijK})^{\top}$ are

$$\mathbf{Y}_{ij}|\mathbf{B}_i \sim N_K\left(\begin{pmatrix} \mu_{ij1}(B_{1i}, B_{21}) \\ \vdots \\ \mu_{ij2}(B_{1i}, B_{2K}) \end{pmatrix}, \boldsymbol{\Sigma}^{(\varepsilon)}\right), \tag{1.4}$$

$$\mathbf{Y}_{ij} \sim N_K\left(\begin{pmatrix} \mathbf{x}_{ij1}^{\top}\boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_{ijK}^{\top}\boldsymbol{\beta} \end{pmatrix}, \boldsymbol{\Sigma}^{(B)} + \boldsymbol{\Sigma}^{(\varepsilon)}\right). \tag{1.5}$$

The above model is called the Gaussian LMM (or normal-normal LMM). Note that Y_{ijk}'s are conditionally independent (given $\mathbf{B}_i$) if and only if ε_{ijk}'s are independent, i.e. the off-diagonal elements of $\boldsymbol{\Sigma}^{(\varepsilon)}$ are all zero.

1.1.2 GLMM

Now we consider Y_{ijk}'s are non-Gaussian responses. Unlike in (1.1), suppose Y_{ijk}'s are not suitable to be modeled via LMMs; for example, they may be positive continuous responses (e.g., time-to-event outcomes) or binary/categorical endpoints. As in Bellio and Varin (2005) we assume the following GLMM for Y_{ijk}:

$$E(Y_{ijk}|B_{1i}, B_{2k}) \;=\; \mu_{ijk}(B_{1i}, B_{2k}) \;=\; \hbar_k^{-1}(\mathbf{x}_{ijk}^\top \boldsymbol{\beta} + B_{1i} + B_{2k}), \tag{1.6}$$

where $\hbar_k(\cdot)$'s are suitable link functions (e.g., logit for binary, log-link for time-to-event), with $\mathbf{x}_{ijk}$ and $\boldsymbol{\beta}$ as defined in (1.1).

Conventional GLMMs conveniently assume that Y_{ijk}'s are conditionally independent, given $\mathbf{B}_i = (B_{1i}, B_{21}, \ldots, B_{2K})^\top$, in the absence of a viable and flexible (conditional) joint distribution for disparate non-Gaussian outcomes. An exception is correlated probit model for correlated binary outcomes (Gueorguieva and Agresti, 2001; Najita et al., 2009), where Gaussian latent variables are used to describe the binary data. The formulation is exactly the same as in the Gaussian LMM except that the Gaussian "responses" are latent. Such a latent formulation is statistically convenient since common binary regression models have equivalent formulations in terms of latent variables. For example, a Gaussian latent variable for a binary outcome results in a probit model for the latter; a logistic latent variable corresponds to a logistic model. However, due to the same limitations of Gaussian LMMs discussed previously, Gueorguieva and Agresti's (Gueorguieva and Agresti, 2001) approach is limited to a Gaussian latent model for the binary data.

1.2 Diabetic retinopathy study

This is a reader-based diagnostic study in Alberta, Canada, in which at least two readers were used to diagnose the presence or absence of certain pathologies, e.g., clinically significant macular edema (CSME), microaneurysms, intra-retinal haemorrhage (IRH), hard exudates (HEX), that are indicative of retinal thickening among diabetic patients, who suffer from treatable diabetic retinopathy. In Canada, where a disproportionate share of diabetic patients are Aboriginal Canadians living in reserves in far-flung rural areas, sending retinal specialists on remote clinics can be costly and inefficient. Due to advances in digital imaging in recent years, a possible alternative is distance evaluation wherein patients undergo stereoscopic digital photography using a high-resolution digital camera. In this approach, digital images of patients' eyes are read by at least two specialists and patients are diagnosed as either positive (i.e., disease is present) or negative (i.e., disease is absent) for the pathologies. This cost-effective tele-ophthalmologic technique has the potential to increase rural accessibility to specialist eye care (Maberley et al., 2003), allowing for early detection and treatment of diabetic retinopathy. Only patients who need treatments would have to travel to a specialist; the transportation cost is thus also reduced. However, before wide implementation of any potential new diagnostic methodology, its accuracy must first be examined. The purpose of the study was thus to determine whether diabetic retinopathy can be identified with high-resolution stereoscopic digital photography and whether this identification correlates well with the accepted gold standard of clinical examination.

The data set-up for the case of $V \geq 1$ pathologies with $K \geq 1$ readers is presented in Table 1, where Y_{iLkv} and Y_{iRkv} represent the binary test results for the left and right eyes of patient $i = 1, \cdots, N$, respectively, as graded by reader $k = 1, \cdots, K$, for pathology $v = 1, \cdots, V$. For $K = V = 2$, the design can be considered as a full paired-patient-paired-reader design, whereby all digital images of a patient's left and right eyes undergo grading by every reader. Note the complex correlation structure in the data: in addition to the fellow-eye correlation induced by the binocular nature of the data, two other sources of correlation are present.

Since readers rely on the same image of the eye, their diagnoses are potentially correlated. Moreover, because the pathologies are all related to retinopathy, it is very likely that the presence or absence of one influences the presence or absence of another. Thus, diagnoses for pathologies are correlated as well.

Table 1.1: Data set-up for $V \geq 1$ pathologies and $K \geq 1$ readers.

Patient	Reader 1		$\cdots$	Reader K	
	Left eye	Right eye		Left eye	Right eye
1	Y_{1L11}	Y_{1R11}		Y_{1LK1}	Y_{1RK1}
	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	Y_{1L1V}	Y_{1R1V}		Y_{1LKV}	Y_{1RKV}
2	Y_{2L11}	Y_{2R11}		Y_{2LK1}	Y_{2RK1}
	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	Y_{2L1V}	Y_{2R1V}		Y_{2LKV}	Y_{2RKV}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
N	Y_{NL11}	Y_{NR11}		Y_{NLK1}	Y_{NRK1}
	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
	Y_{NL1V}	Y_{NR1V}		Y_{NLKV}	Y_{NRKV}

The accuracy of a medical test for diagnosing the presence or absence of a disease can be described by several measures, the most common of which are given by the test's sensitivity and specificity with respect to the true disease status as determined by a traditionally used and accepted test regarded as a 'gold standard'. Given binary variables Y and D denoting, respectively, the test's result and the disease status as determined by the gold standard, with 0 and 1 indicating negative and positive outcomes, respectively, the test's sensitivity and specificity are then given by $\text{Sen} = P(Y = 1|D = 1)$ and $\text{Spc} = P(Y = 0|D = 0)$, respectively. Other frequently used measures of diagnostic accuracy are the so-called post-test probabilities given by the test's positive predictive and negative predictive values. The former is defined as the probability $\text{PPV} = P(D = 1|Y = 1)$ of presence of disease given a positive test result while the latter is the probability $\text{NPV} = P(D = 0|Y = 0)$ of absence of disease given a negative test result. Positive and negative predictive values describe how

well a test predicts a patient's disease status, while sensitivity and specificity describe how well the test discriminates between positive disease status and negative disease status. Note that a diagnostic test's sensitivity and specificity are measures of the test's intrinsic accuracy and as such, unlike the predictive values, do not provide information on the accuracy of the diagnoses.

1.3 Objective of the thesis

For the diabetic retinopathy data, each patient yields two readings by a reader corresponding to the patient's left and right eyes, and consequently, patients are considered the clusters. A full likelihood analysis of the data was implemented in Withanage et al. (2015) using the data cloning (DC) approach. However, the analysis in Withanage et al. (2015) was based on a multivariate probit model based on a multivariate latent Gaussian description of the multiple correlated binary outcomes. In this thesis, our objective is to propose a joint modeling of clustered binary data with crossed random effects by relaxing this assumption by allowing for disparate non-Gaussian latent variables for the binary responses, and to analyze the diabetic retinopathy data applying GCMM. By accommodating a mixture of different Gaussian and non-Gaussian latent distributions (e.g., some Gaussian latent variables while others are logistic), a more flexible and realistic model can be constructed. A general approach entails assuming non-central t-latent distributions for binary responses, thus subsuming both probit and logit models while at the same time rendering the model robust to contamination and outliers.

Before analyzing the diabetic retinopathy data with GCMM, our objective is to see the performance of the estimators via a simulation study involving data with similar setup as the real data set. Instead of the full likelihood estimation, we adopted an alternative based on the idea of composite likelihood (Lindsay, 1988; Cox and Reid, 2004), specifically pairwise likelihood approach. The pairwise likelihood estimation resolves the computational complex-

ity involved in high-dimensional random effects models (Fieuws and Verbeke, 2006; Bellio and Varin, 2005; Renard et al., 2002). Estimation based on suitable composite likelihoods is generally consistent, and the efficiency loss with respect to ML estimation has been empirically shown to be insubstantial in many cases. In practice, efficiency losses are less important than the inability to directly and computationally efficiently fit the full multivariate model.

1.4 Outline

Chapter 2 discusses the GCMMs for clustered binary data with crossed random effects. Details of likelihood estimation procedure and inference for the model are also discussed in Chapter 2. Chapter 3 reports the results from simulation study on the finite-sample properties of the estimates. Results of our analysis of the diabetic retinopathy data for the pathology CSME are presented in Chapter 4. Chapter 5 concludes the thesis with a brief discussion.

Chapter 2

Gaussian copula mixed models (GCMMs) with crossed random effects

In this chapter, we briefly review how copula model works in Section 2.1, and discuss the Gaussian copula mixed model (GCMM) in Section 2.2. The likelihood estimation procedure and the variance calculation for the estimators are discussed in Section 2.3. Calculation of marginal associations is shown in Section 2.4.

2.1 Brief review of copulas

Copulas are not new in biomedical studies and recent references include Genest et al. (2013), de Leon and Wu (2011), Song et al. (2009), and Zimmer and Trivedi (2006), among many others.

Copula is a function that binds marginal distributions of a set of random variables (RVs) to form their multivariate distribution. To see this, consider P (possibly dependent) RVs $Y_1, \cdots, Y_P$, with corresponding CDFs $F_{Y_1}(\cdot), \cdots, F_{Y_P}(\cdot)$. It is well-known that the probability integral transformations (PITs) $U_1 = F_{Y_1}(Y_1), \cdots, U_P = F_{Y_P}(Y_P)$ are identically

distributed as *uniform*$[0,1]$, i.e., $U_j \sim$ *uniform*$[0,1]$, for all $j = 1, \cdots, P$. Hence, the joint CDF $F_{Y_1,\cdots,Y_P}(\cdot)$ of $Y_1, \cdots, Y_P$ can be written as

$$F_{Y_1,\cdots,Y_P}(y_1, \cdots, y_P) \;=\; P(U_1 \leq u_1, \cdots, U_P \leq u_P) \;=\; C(u_1, \cdots, u_P), \qquad (2.1)$$

where $u_1 = F_{Y_1}(y_1), \cdots, u_P = F_{Y_P}(y_P)$ are the realizations of the PITs $U_1, \cdots, U_P$, with $C(\cdot) \equiv F_{U_1,\cdots,U_P}(\cdot)$ the joint CDF of $U_1, \cdots, U_P$. The function $C(\cdot)$ is called a P-dimensional copula (or P-copula) and it is the unique copula associated with the CDF $F_{Y_1,\cdots,Y_P}(\cdot)$. Specifically, $C(\cdot) : [0,1]^P \to [0,1]$ and has the following properties:

1. $C(1, \cdots, 1, u_j, 1, \cdots, 1) = u_j$ for every $j \leq P$ and for all $u_j \in [0,1]$;

2. $C(u_1, \cdots, u_P) = 0$ if $u_j = 0$ for every $j \leq P$;

3. $C(u_1, \cdots, u_P)$ is P-increasing.

Hence, the joint CDF $F_{Y_1,\cdots,Y_P}(\cdot)$ of $Y_1, \cdots, Y_P$ is equivalent to the joint CDF $C(\cdot)$ of the marginally uniform (possibly dependent) RVs $U_1 = F_{Y_1}(Y_1), \cdots, U_P = F_{Y_P}(Y_P)$. The copula approach is thus a useful method for constructing a joint distribution when the marginal distributions are known or can be easily specified. It enables the construction of non-Gaussian multivariate/joint models suitable to applications involving non-Gaussian variables.

The corresponding joint density $f_{Y_1,\cdots,Y_P}(\cdot)$ of $Y_1, \cdots, Y_P$ can be obtained by taking the Pth mixed-partial derivative of (2.1). This yields

$$f_{Y_1,\cdots,Y_P}(y_1, \cdots, y_P) \;=\; \frac{\partial^P F_{Y_1,\cdots,Y_P}(y_1, \cdots, y_P)}{\partial y_1 \cdots \partial y_P} \;=\; c(u_1, \cdots, u_P) \prod_{j=1}^{P} f_{Y_j}(y_j), \qquad (2.2)$$

where $c(u_1, \cdots, u_P) = \partial^P C(u_1, \cdots, u_P)/\partial u_1 \cdots \partial u_P$ is the so-called copula density for copula $C(\cdot)$.

Sklar's Theorem (Sklar, 1959) states that the copula representation of $F_{Y_1,\cdots,Y_P}(\cdot)$ in (2.1) is unique if and only if all the margins are continuous (i.e., $Y_1, \cdots, Y_P$ are continuous RVs);

otherwise, for discrete margins, uniqueness property only holds in the product range of the margins (Song et al., 2007, see, e.g.,).

The suitability of copulas in applications depends on how well they are able to capture the unique dependencies in the data. A large number of copula families that can capture the different dependence features of the data have been studied in the literature. Examples include the product copula, Farlie-Gumbel-Morgenstern copula, the Frank copula, the Clayton copula, and the Student's t-copula, among others. Zimmer and Trivedi (2006) and Joe (1997) give comprehensive surveys of many of these families and their properties.

The Gaussian copula family has been widely used in applications because of its convenient marginalization and conditionalization properties. In addition, the Gaussian copula can model positive and negative dependence in the data. The P-dimensional Gaussian copula is given by

$$C_{\Phi}(u_1, \cdots, u_P; \widetilde{\mathbf{R}}) \quad = \quad \Phi_P(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_P); \widetilde{\mathbf{R}}), \tag{2.3}$$

where $\Phi_P(\cdot; \widetilde{\mathbf{R}})$ is the P-dimensional standard Gaussian CDF (i.e., zero means and unit variances) with correlation matrix $\widetilde{\mathbf{R}}$. For continuous RVs $Y_1 \sim F_{Y_1}(\cdot), \cdots, Y_P \sim F_{Y_P}(\cdot)$ whose joint CDF $F_{Y_1,\cdots,Y_P}(\cdot)$ is determined by Gaussian copula (2.3), we have

$$F_{Y_1,\cdots,Y_P}(y_1, \cdots, y_P) \quad = \quad \Phi_P(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_P); \widetilde{\mathbf{R}}), \tag{2.4}$$

where $u_1, \cdots, u_P$ are respective realizations of the PITs $U_1 = F_{Y_1}(Y_1) \sim \mathit{uniform}[0,1], \cdots,$ $U_P = F_{Y_P}(Y_P) \sim \mathit{uniform}[0,1]$, with $\widetilde{\mathbf{R}}$ containing the normal correlations $\widetilde{\rho}_{jj'}$, which are the correlations between the so-called normal scores $\Phi^{-1}(U_j)$ and $\Phi^{-1}(U_{j'})$ given by

$$\widetilde{\rho}_{jj'} \quad = \quad \mathit{corr}(\Phi^{-1}(U_j), \Phi^{-1}(U_{j'})). \tag{2.5}$$

The corresponding density $f_{Y_1,\cdots,Y_P}(\cdot)$ obtained from (2.2) is given by

$$f_{Y_1,\cdots,Y_P}(y_1,\cdots,y_P) \;=\; \frac{\phi_P(\Phi^{-1}(u_1),\cdots,\Phi^{-1}(u_P);\widetilde{\mathbf{R}})}{\prod_{j=1}^{P}\phi(\Phi^{-1}(u_j))}\prod_{j=1}^{P} f_{Y_j}(y_j), \tag{2.6}$$

where $\phi_P(\cdot;\widetilde{\mathbf{R}})$ is the P-dimensional standard Gaussian density and $\phi(\cdot) \equiv \phi_1(\cdot)$ (i.e., the standard normal density).

Note that the correlation matrix $\widetilde{\mathbf{R}}$ is "margin-free" in the sense that $\Phi^{-1}(F_{Y_j}(Y_j)) \sim N(0,1)$, for any continuous margin $F_{Y_j}(\cdot)$, for all $j = 1,\cdots,P$. To see this, we have

$$P(\Phi^{-1}(U_j) \le y) \;=\; P(\Phi^{-1}(F_{Y_j}(Y_j)) \le y) \;=\; P(F_{Y_j}(Y_j) \le \Phi(y)) \;=\; \Phi(y),$$

for all real y, and for all j. Although the normal correlations $\widetilde{\rho}_{jj'}$ do not directly model the dependence among $Y_1,\cdots,Y_P$, they can be used to bound the correlations $\rho_{jj'} = corr(Y_j, Y_{j'})$, since $\rho_{jj'} \le |\widetilde{\rho}_{jj'}|$ (Klaassen and Wellner, 1997). Alternatively, a piecewise linear approximation may be used to recover $\rho_{jj'}$ from $\widetilde{\rho}_{jj'}$ (Kugiumtzis and Bora-Senta, 2010).

Since rank-based association measures, such as Kendall's tau, are invariant to monotonic transformations, copula models generally rely on them to evaluate the strength of dependence between variables. For example, the normal scores $\Phi^{-1}(U_j) = \Phi^{-1}(F_{Y_j}(Y_j))$ in (2.4) are monotonic transformations of the original variables Y_j, so that the Kendall's tau $\widetilde{\tau}_{jj'} = \tau(\Phi^{-1}(U_j), \Phi^{-1}(U_{j'}))$ between a pair of normal scores is the same as the corresponding Kendall's tau $\tau_{jj'} = \tau(Y_j, Y_{j'})$ between the original variables. Using the well-known relationship between $\widetilde{\tau}_{jj'}$ and $\widetilde{\rho}_{jj'}$, we get

$$\widetilde{\tau}_{jj'} \;=\; \frac{2}{\pi}\sin^{-1}(\widetilde{\rho}_{jj'}) \;=\; \tau_{jj'}. \tag{2.7}$$

2.2 GCMM with crossed random effects

Direct application of copula to binary data with the use of discrete margins — in contrast to the latent variable approach adopted in Withanage et al. (2015) and Wu and de Leon (2014), among others — is not new (Nikoloulopoulos and Karlis, 2008, 2009a,b; Song et al., 2009; Zimmer and Trivedi, 2006). As discussed in Wu and de Leon (2014), for example, Sklar's Thorem(Sklar, 1959) no longer holds when discrete margins are used in a copula, and as a consequence, the copula is unique only on the product range of the discrete margins, thus failing to completely uniquely determine the joint distribution of the discrete variables. However, the resulting copula model is still a proper and valid distribution, which explains the proliferation of copula models for discrete data. A recent reference is Genest et al. (2013), where meta-elliptical copulas are used to directly model correlated binary data.

The main issue about such copula models is more practical than theoretical, and concerns the interpretability of the dependence parameters. In particular, common rank-based association measures like Kendall's tau and Spearman's rho (Goodman and Kruskal, 1954), which are margin-free in traditional copula applications to continuous variables, may now depend on the margins (Nešlehová, 2007; Mesfioui and Tajar, 2005). In addition, their ranges may now be substantially restricted, so that re-scaled versions of them become necessary for proper interpretation (Genest et al., 2013). To avoid these complications, we adopt a latent variable formulation of the binary data as in Withanage et al. (2015) and Wu and de Leon (2014), and construct the copula model at the latent level. The joint model for the binary data is then constructed indirectly from the copula model. This is exactly the same approach developed in Withanage et al. (2014, 2015) except that we now accommodate a more flexible, possibly non-Gaussian, latent model; specifically, we adapt the GCMM to the binary data setting with crossed effects as in the diabetic retinopathy study. Dependence among the binary outcomes is measured by the normal correlations between the underlying latent variables; these normal correlations are akin to so-called tetrachoric correlations, which are quite commonplace as measures of association between discrete variables in psy-

chometrics. Note that the normal correlations in this case are margin-free, from which the tetrachoric correlations can be calculated via piecewise linear approximations (Kugiumtzis and Bora-Senta, 2010). We outline the methodology in what follows.

Let Y_{ijk} be the assessment of reader $k = 1, \cdots, K$, of eye $j = L, R$, of patient $i = 1, \cdots, N$. We assume that Y_{ijk} is observed by dichotomizing a continuous latent variable Y^*_{ijk}, where $Y_{ijk} = \mathrm{I}\{Y^*_{ijk} > 0\}$, for all i, j, k, where $\mathrm{I}\{\cdot\}$ is the indicator function. Suppose $Y^*_{ijk}|B_{1i}, B_{2k} \sim F_{Y^*_{ijk}|B_{1i},B_{2k}}(\cdot|\cdot)$, with conditional mean modeled as

$$\mu^*_{ijk}(B_{1i}, B_{2k}) \;=\; E(Y^*_{ijk}|B_{1i}, B_{2k}) \;=\; \mathbf{x}^\top_{ij}\boldsymbol{\beta} + B_{1i} + B_{2k}, \tag{2.8}$$

where $\mathbf{x}_{ij}$ is a known vector of covariates with corresponding vector $\boldsymbol{\beta}$ of unknown regression coefficients, with B_{1i} the subject-specific random effect indicating heterogeneity between subjects, and B_{2k} the random effect representing heterogeneity between readers. We assume that $B_{1i} \overset{iid}{\sim} f_{B_{1i}}(\cdot)$ independently of $B_{2k} \overset{iid}{\sim} f_{B_{2k}}(\cdot)$; further, given B_{1i}, a patient's left and right eyes are assumed independent, and in addition, we assume patients are independent, given $B_{21}, \cdots, B_{2K}$. The choice of the (conditional) margin $F_{Y^*_{ijk}|B_{1i},B_{2k}}(\cdot|\cdot)$ dictates the link function $\hbar_{ijk}(\cdot)$ such that $\mu_{ijk}(B_{1i}, B_{2k}) = E(Y_{ijk}|B_{1i}, B_{2k}) = P(Y_{ijk} = 1|B_{1i}, B_{2k}) = \hbar^{-1}_{ijk}(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + B_{1i} + B_{2k})$. For example, if $Y^*_{ijk}|B_{1i}, B_{2k} \sim N(\mu^*_{ijk}, 1)$, then we get

$$\mu_{ijk}(B_{1i}, B_{2k}) \;=\; P(Y^*_{ijk} > 0|B_{1i}, B_{2k}) \;=\; \Phi(\mu^*_{ijk}(B_{1i}, B_{2k})), \tag{2.9}$$

so that $\hbar_{ijk}(\cdot)$ is the probit link. For model identifiability, we assumed that $var(Y^*_{ijk}|B_{1i}, B_{2k}) = 1$ in (2.9). In general, if $var(Y^*_{ijk}|B_{1i}, B_{2k})$ depends on a scale parameter ξ for $F_{Y^*_{ijk}|B_{1i},B_{2k}}(\cdot|\cdot)$ (e.g., logistic latent distribution), then we assume that ξ is known; in any case, we assume that $var(Y^*_{ijk}|B_{1i}, B_{2k}) > 0$ is known if the scale parameter is the variance (Gueorguieva and Agresti, 2001). For a similar reason, we also assume a zero cutpoint for the threshold model linking Y_{ijk} to Y^*_{ijk}, so that $\boldsymbol{\beta}$ includes an intercept coefficient (Catalano and Ryan, 1992).

With conditional independence of Y_{ijk}, for all j, k, model (2.8) leads to an exchangeable

correlation structure for readers' assessments for the same eye or for fellow eyes of a patient. However, this is not true in many reader-based diagnostic studies, as assessments by different readers for the same eye are much more alike than their assessments for different eyes. Withanage et al. (2015) adopted a multivariate Gaussian latent distribution, so that each latent variable Y^*_{ijk} is Gaussian, for all j, k, and they included a third random effect to delineate these correlations. The additional computational burden a likelihood analysis of such model required was obviated by the use of DC, which involves neither numerical integration nor optimization of the likelihood function. In this paper, we alternatively adopt the GCMM, a copula-based approach (de Leon and Wu, 2011; Wu et al., 2013) to account for the different associations between readers' assessments for the same eye and for fellow eyes of a patient, without the addition of another random effect.

Let $\mathbf{Y}_{ij} = (Y_{ij1}, \cdots, Y_{ijK})^\top$, $\mathbf{Y}^*_{ij} = (Y^*_{ij1}, \cdots, Y^*_{ijK})^\top$, and $\mathbf{B}_2 = (B_{21}, \cdots, B_{2K})^\top$. Given the (conditional) margins $F_{Y^*_{ij1}|B_{1i},B_{21}}(\cdot|\cdot), \cdots, F_{Y^*_{ijK}|B_{1i},B_{2K}}(\cdot|\cdot)$, let the (conditional) cumulative distribution function (CDF) $F_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\cdot|\cdot)$ of $\mathbf{Y}^*_{ij}$, given B_{1i} and $\mathbf{B}_2$, be determined by a Gaussian copula as follows:

$$F_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\mathbf{y}^*_{ij}|b_{1i}, \mathbf{b}_2) \;=\; \Phi_K(\Phi^{-1}\{u^*_{ij1}(b_{1i}, b_{21})\}, \cdots, \Phi^{-1}\{u^*_{ijK}(b_{1i}, b_{2K})\}; \widetilde{\mathbf{R}}^*), \qquad (2.10)$$

where $u^*_{ijk}(b_{1i}, b_{2k}) = F_{Y^*_{ijk}|B_{1i},B_{2k}}(y^*_{ijk}|b_{1i}, b_{2k})$ is the realization of the latent (conditional) probability integral transform (PIT) $U^*_{ijk}(B_{1i}, B_{2k}) = F_{Y^*_{ijk}|B_{1i},B_{2k}}(Y^*_{ijk}|B_{1i}, B_{2k})$, and $\widetilde{\mathbf{R}}^*$ is the matrix of (conditional) normal correlations

$$\widetilde{\rho}^*_{kk'} \;=\; corr\{\Phi^{-1}(U^*_{ijk}(B_{1i}, B_{2k})), \Phi^{-1}(U^*_{ijk'}(B_{1i}, B_{2k'}))\} \qquad (2.11)$$

between the latent variables. These normal correlations measure the conditional association between assessments by different readers of the same eye j of patient i. In practice, readers' assessments are generally consistent due to their similar training. It is thus reasonable to assume an exchangeable structure for $\widetilde{\mathbf{R}}^*$, so that $\widetilde{\rho}^*_{kk'} = \widetilde{\rho}^*$, for all $k \neq k'$. Nevertheless,

if there are significant differences among the assessments by different readers, it may be worthwhile to assume an unstructured $\widetilde{\mathbf{R}}^*$.

The corresponding model for $\mathbf{Y}_{ij}$ is then

$$P(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | B_{1i} = b_{1i}, \mathbf{B}_2 = \mathbf{b}_2) \quad = \quad \int_{A_{ij1} \times \cdots \times A_{ijK}} f_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\mathbf{y}^*_{ij}|b_{1i}, \mathbf{b}_2) d\mathbf{y}^*_{ij}, \qquad (2.12)$$

where $f_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\cdot|\cdot)$ is the (conditional) density of $\mathbf{Y}^*_{ij}$ corresponding to CDF (2.10), and the intervals A_{ijk} are either $(-\infty, 0]$ or $(0, +\infty)$ according to whether $y_{ijk} = 0$ or 1. Note that (2.12) involves the computation of multivariate Gaussian orthant probabilities, for which a general expression can be found in Song et al. (2009) and Genest et al. (2013).

Model (2.12) is a GCMM, and as such inherits all the nice properties of GCMMs. For example, the latent distributions can be flexibly chosen and need not come from the same parametric family. It is possible to consider Gaussian latent variables for some binary outcomes, logistic for others, and t for still others. This contrasts with Gueorguieva and Agresti (2001) correlated probit model based on a multivariate Gaussian latent distribution. Moreover, our use of the continuous latent vector $\mathbf{Y}^*_{ij}$ in (2.12) to describe the binary vector $\mathbf{Y}_{ij}$ and to build the joint model for $\mathbf{Y}_{ij}$ from the GCMM for $\mathbf{Y}^*_{ij}$ allows us to sidestep the complications of using discrete margins in the Gaussian copula. The CDF (2.10) is uniquely determined by the Gaussian copula; hence, the joint model in (2.12) is likewise unique. The normal correlation matrix $\widetilde{\mathbf{R}}^*$ is also margin-free, so that the dependence model for $\mathbf{Y}_{ij}$, as captured by $\widetilde{\mathbf{R}}^*$, is independent of its marginal specification.

A potential drawback is that the dependence in $\mathbf{Y}_{ij}$ is measured by the dependence in $\mathbf{Y}^*_{ij}$ (i.e., association among $Y_{ij1}, \cdots, Y_{ijK}$ is gauged at the latent level). However, we view this as an advantage rather than a disadvantage. For one, the use of latent-level correlations, such as polychoric correlations between discrete variables (or tetrachoric correlations in the case of binary variables) is standard practice in many disciplines. For another, such correlations are not artificially constrained by the marginal probabilities of the discrete vari-

ables (unlike Pearson's correlations) and their number does not increase with the number of levels/categories of ordinal/categorical variables (unlike odds ratios).

2.3 Likelihood estimation

Let $\mathbf{\Theta}$ be the vector containing all parameters of (2.12). The conditional likelihood contribution $L_{ij}(\mathbf{\Theta}|b_{1i}, \mathbf{b}_2)$ of eye j of patient i is

$$L_{ij}(\mathbf{\Theta}|b_{1i}, \mathbf{b}_2) \quad = \quad P(\mathbf{Y}_{ij} = \mathbf{y}_{ij}|B_{1i} = b_{1i}, \mathbf{B}_2 = \mathbf{b}_2).$$

The conditional independence of $\mathbf{Y}^*_{iL}$ and $\mathbf{Y}^*_{iR}$, given B_{1i}, for all i, and of $\mathbf{Y}^*_{ij}$ and $\mathbf{Y}^*_{i'j'}$, given $\mathbf{B}_2$, for all $i \neq i'$ and $j \neq j'$, gives the marginal likelihood function $L(\mathbf{\Theta})$ as

$$L(\mathbf{\Theta}) = \int_{\mathcal{R}^{K+N}} \prod_{i,j} L_{ij}(\mathbf{\Theta}|b_{1i}, \mathbf{b}_2) f_{B_{1i}}(b_{1i}) f_{B_{21}}(b_{21}) \cdots f_{B_{2K}}(b_{2K}) db_{1i} db_{21} \cdots db_{2K}, \qquad (2.13)$$

For convenience, let $B_{1i} \overset{iid}{\sim} N(0, \sigma_1^2)$ be independent of $B_{2k} \overset{iid}{\sim} N(0, \sigma_2^2)$, so that $\mathbf{\Theta} = (\boldsymbol{\beta}^\top, \sigma_1, \sigma_2, \widetilde{\rho}^*)^\top$; note that non-Gaussian choices for the random effects distributions are also possible; for example, Lin et al. (2010) used a bridge-distributed random effect to facilitate marginal interpretability in logistic regressions. Marginal likelihood function $L(\mathbf{\Theta})$ now becomes

$$L(\mathbf{\Theta}) = \frac{1}{\sigma_1^N \sigma_2^K} \int_{\mathcal{R}^{K+N}} \prod_{i,j} L_{ij}(\mathbf{\Theta}|b_{1i}, \mathbf{b}_2) \phi\left(\frac{b_{1i}}{\sigma_1}\right) \phi\left(\frac{b_{21}}{\sigma_2}\right) \cdots \phi\left(\frac{b_{2K}}{\sigma_2}\right) db_{1i} db_{21} \cdots db_{2K}, \qquad (2.14)$$

where $\phi(\cdot)$ is the standard normal density. The integration involved in (2.14) suffers from the curse of dimensionality, as the number of integrals in (2.14) increases with the number of patients and readers. This necessitates the use of some numerical, stochastic or analytical approximation. Gaussian-Hermite quadrature (Evans and Swartz, 2000; Lesaffre and

Spiessens, 2001) is the preferred method in GLMMs when a few random factors are involved. The likelihood function (2.13) involves $N+K$ integrals and evaluation via Gausssian-Hermite quadrature is not feasible. In addition, no simple and effective error bounds are available for high-dimensional integrals approximated by quadrature methods (Renard et al., 2004).

Alternatively, Monte Carlo-based methods, such as the Monte Carlo EM (McCulloch, 1997; Booth and Hobert, 1999) and Monte Carlo Newton-Raphson (McCulloch, 1994, 1997) algorithms, are widely used when the dimension of the random effects is high. However, these methods are generally quite computationally intensive.

Rather than undertake a full likelihood analysis in this case, we adopt instead the pairwise likelihood (PL) approach. Bellio and Varin (2005) showed that the inferential and computational gain provided by the PL approach is remarkable in models for binary data with crossed random effects.

2.3.1 Pairwise likelihood estimation

The PL function $PL(\boldsymbol{\Theta})$ is obtained by the product of the bivariate probabilities for all possible pairs. However, we choose to include only those pairs that share at least one common random effect, as in Bellio and Varin (2005). Hence, we get

$$\begin{aligned} PL(\boldsymbol{\Theta}) &= \left(\prod_{\substack{i \\ j,\, j'}} \prod_{k<k'} P(Y_{ijk} = y_{ijk}, Y_{ij'k'} = y_{ij'k'}) \right) \left(\prod_{\substack{i<i' \\ j,\, j'}} \prod_{k} P(Y_{ijk} = y_{ijk}, Y_{i'j'k} = y_{i'j'k}) \right) \\ &\quad \times \left(\prod_{\substack{i \\ j \neq j'}} \prod_{k} P(Y_{ijk} = y_{ijk}, Y_{ij'k} = y_{ij'k}) \right), \end{aligned} \tag{2.15}$$

with

$$P(Y_{ijk} = y_{ijk}, Y_{ij'k'} = y_{ij'k'}) \quad = \quad \begin{cases} P(Y^*_{ijk} > 0, Y^*_{ij'k'} > 0) & \text{, if } y_{ijk} = y_{ij'k'} = 1 \\ P(Y^*_{ijk} > 0, Y^*_{ij'k'} \leq 0) & \text{, if } y_{ijk} = 1, y_{ij'k'} = 0 \\ P(Y^*_{ijk} \leq 0, Y^*_{ij'k'} > 0) & \text{, if } y_{ijk} = 0, y_{ij'k'} = 1 \\ P(Y^*_{ijk} \leq 0, Y^*_{ij'k'} \leq 0) & \text{, if } y_{ijk} = y_{ij'k'} = 0 \end{cases} .$$

From (2.15), the number of random effects included in a pair varies from two to three — refer to (2.16) and (2.19) to see why — and thus, Gaussian-Hermite quadrature may be used. The different pairs in (2.15) reflect the following association types:

1. association between Y_{ijk} and $Y_{ijk'}$, the assessments by different readers of a patient's eye;
2. association between Y_{iLk} and $Y_{iRk'}$, the assessments by different readers, one reading the left and the other the right;
3. association between Y_{iLk} and Y_{iRk}, the assessments by the same reader of a patient's fellow eyes; and
4. association between Y_{ijk} and $Y_{i'j'k}$, the assessments by the same reader of matching or non-matching eyes of different patients.

Each of these pairwise contributions can then be written in terms of univariate or bivariate probabilities as follows:

$$
\begin{aligned}
P(Y_{ijk} = y_{ijk}, Y_{ijk'} = y_{ijk'}) &= \frac{1}{\sigma_1\sigma_2^2}\int_{\mathcal{R}^3} \Phi_2(\Phi^{-1}\{u_{ijk}(b_{1i}, b_{2k})\}, \Phi^{-1}\{u_{ijk'}(b_{1i}, b_{2k'})\}; \widetilde{\rho}^*) \\
&\quad \times \phi\left(\frac{b_{1i}}{\sigma_1}\right)\phi\left(\frac{b_{2k}}{\sigma_2}\right)\phi\left(\frac{b_{2k'}}{\sigma_2}\right) db_{1i}db_{2k}db_{2k'}, \qquad (2.16) \\
P(Y_{iLk} = y_{iLk}, Y_{iRk'} = y_{iRk'}) &= \frac{1}{\sigma_1\sigma_2^2}\int_{\mathcal{R}^3} P(Y_{iLk} = y_{iLk}|b_{1i}, b_{2k})P(Y_{iRk'} = y_{iRk'}|b_{1i}, b_{2k'}) \\
&\quad \times \phi\left(\frac{b_{1i}}{\sigma_1}\right)\phi\left(\frac{b_{2k}}{\sigma_2}\right)\phi\left(\frac{b_{2k'}}{\sigma_2}\right) db_{1i}db_{2k}db_{2k'}, \qquad (2.17) \\
P(Y_{iLk} = y_{iLk}, Y_{iRk} = y_{iRk}) &= \frac{1}{\sigma_1\sigma_2}\int_{\mathcal{R}^2} P(Y_{iLk} = y_{iLk}|b_{1i}, b_{2k})P(Y_{iRk} = y_{iRk}|b_{1i}, b_{2k}) \\
&\quad \times \phi\left(\frac{b_{1i}}{\sigma_1}\right)\phi\left(\frac{b_{2k}}{\sigma_2}\right) db_{1i}db_{2k}, \qquad (2.18) \\
P(Y_{ijk} = y_{ijk}, Y_{i'j'k} = y_{i'j'k}) &= \frac{1}{\sigma_1^2\sigma_2}\int_{\mathcal{R}^3} P(Y_{ijk} = y_{ijk}|b_{1i}, b_{2k})P(Y_{i'j'k} = y_{i'j'k}|b_{1i'}, b_{2k}) \\
&\quad \times \phi\left(\frac{b_{1i}}{\sigma_1}\right)\phi\left(\frac{b_{1i'}}{\sigma_1}\right)\phi\left(\frac{b_{2k}}{\sigma_2}\right) db_{1i}db_{1i'}db_{2k}, \qquad (2.19)
\end{aligned}
$$

for $j, j' = L, R$. The PL estimate $\widehat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta} = (\boldsymbol{\beta}^\top, \sigma_1, \sigma_2, \widetilde{\rho}^*)^\top$ is then obtained by solving the pairwise score equations $\mathbf{U}_{p\ell}(\boldsymbol{\Theta}) = \partial \log PL/\partial\boldsymbol{\Theta} = \mathbf{0}$. These estimates we refer to as pairwise-maximum-likelihood estimates (PMLEs) as they are obtained my maximizing the pairwise likelihood function.

2.3.2 Variance of the PMLEs

The standard errors (SEs) of the PMLEs can be obtained from the inverse of the Godambe information $\mathbf{G}(\boldsymbol{\Theta}) = \mathbf{H}(\boldsymbol{\Theta})\mathbf{J}(\boldsymbol{\Theta})^{-1}\mathbf{H}(\boldsymbol{\Theta})$, where $\mathbf{H}(\boldsymbol{\Theta}) = E[-\partial\mathbf{U}_{p\ell}(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}]$ and $\mathbf{J}(\boldsymbol{\Theta}) = E[\mathbf{U}_{p\ell}(\boldsymbol{\Theta})\mathbf{U}_{p\ell}(\boldsymbol{\Theta})^\top]$. The matrix $\mathbf{H}(\boldsymbol{\Theta})$ can be estimated by

$$
\widehat{\mathbf{H}(\boldsymbol{\Theta})} = \mathbf{H}(\widehat{\boldsymbol{\Theta}}) = -\frac{\partial}{\partial\boldsymbol{\Theta}}\mathbf{U}_{p\ell}(\boldsymbol{\Theta})\bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}}. \qquad (2.20)
$$

However, the estimate of $\mathbf{J}(\boldsymbol{\Theta})$ is possible only if independent or pseudo-independent replicates of the data are available. Similarly, the jackknife estimate (Zhao and Joe, 2005; Lipsitz et al., 1994) of $\mathbf{J}(\boldsymbol{\Theta})$ is possible provided that the data may be decomposed into pseudo-independent subunits. Both methods fail for models with crossed random effects because of the non-decaying correlation pattern, as in model (2.8). Following Bellio and Varin (2005), we thus estimate $\mathbf{J}(\boldsymbol{\Theta})$ by pure Monte Carlo computation. Specifically, given $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma}_1, \widehat{\sigma}_2, \widehat{\widetilde{\rho}}^*)^\top$, Monte Carlo samples were generated based on the following algorithm:

1. Generate B_{2k} from a $N(0, \widehat{\sigma}_2^2)$, for $k = 1, \ldots, K$.

2. Generate B_{1i} from a $N(0, \widehat{\sigma}_1^2)$.

3. Jointly generate $y^*_{ij1}, \ldots, y^*_{ijK}$ from $F_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(.|.)$ with the conditional mean $\widehat{\mu}^*_{ijk}(B_{1i}, B_{2k}) = \mathbf{x}_{ij}^\top \widehat{\boldsymbol{\beta}} + B_{1i} + B_{2k}$.

4. Dichotomize y^*_{ijk} as $y_{ijk} = \mathrm{I}\{y^*_{ijk} > 0\}$ to obtain $y^{(b)}_{ijk}$.

5. Repeat Steps 2 to 4 for N times.

After completing the above algorithm we have the bth data set. Generate B such data sets, where B is the number of Monte Carlo samples, repeating the above algorithm B times and the Monte Carlo estimate of $\mathbf{J}(\boldsymbol{\Theta})$ is given by

$$\widehat{\mathbf{J}(\boldsymbol{\Theta})} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{U}^{(b)}_{p\ell}(\widehat{\boldsymbol{\Theta}}) \mathbf{U}^{(b)\top}_{p\ell}(\widehat{\boldsymbol{\Theta}}), \tag{2.21}$$

where $\mathbf{U}^{(b)}_{p\ell}(\widehat{\boldsymbol{\Theta}})$ is the score vector evaluated at $\widehat{\boldsymbol{\Theta}}$ for the bth Monte Carlo sample. For the simulation study in Chapter 3 and data analysis in Chapter 4, we have used the `fdHess` function of `nlme` package in `R` to calculate these score vectors.

2.4 Marginal associations

The marginal associations between the binary responses are measured via the marginal correlations at the latent level. The marginal correlations can be calculated from the conditional correlations using the total covariance formula. Since $E[Y^*_{ijk}|b_{1i}, b_{2k}] = \mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}$, $E[Y^*_{ijk'}|b_{1i}, b_{2k'}] = \mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k'}$, and $var(Y^*_{ijk}|b_{1i}, b_{2k}) = var(Y^*_{ijk'}|b_{1i}, b_{2k'}) = 1$, using the total covariance formula we have

$$
\begin{aligned}
cov(Y^*_{ijk}, Y^*_{ijk'}) &= E[cov(Y^*_{ijk}, Y^*_{ijk'}|b_{1i}, b_{2k}, b_{2k'})] + cov(E[Y^*_{ijk}|b_{1i}, b_{2k}], E[Y^*_{ijk'}|b_{1i}, b_{2k'}]) \\
&= E[corr(Y^*_{ijk}, Y^*_{ijk'}|b_{1i}, b_{2k}, b_{2k'})\sqrt{var(Y^*_{ijk}|b_{1i}, b_{2k})\, var(Y^*_{ijk'}|b_{1i}, b_{2k'})}] \\
&\quad + cov(E[Y^*_{ijk}|b_{1i}, b_{2k}], E[Y^*_{ijk'}|b_{1i}, b_{2k'}]) \\
&= E[\rho\sqrt{1 \times 1}] + cov(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}, \mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k'}) \\
&= \rho + \sigma_1^2, \\
var(Y^*_{ijk}) &= var(E[Y^*_{ijk}|b_{1i}, b_{2k}]) + E[var(Y^*_{ijk}|b_{1i}, b_{2k})] \\
&= var(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}) + E[1] \\
&= \sigma_1^2 + \sigma_2^2 + 1
\end{aligned}
$$

and similarly, $var(Y^*_{ijk'}) = \sigma_1^2 + \sigma_2^2 + 1$. So the marginal correlation between Y^*_{ijk} and $Y^*_{ijk'}$ becomes

$$
\begin{aligned}
corr(Y^*_{ijk}, Y^*_{ijk'}) &= \frac{cov(Y^*_{ijk}, Y^*_{ijk'})}{\sqrt{var(Y^*_{ijk})\, var(Y^*_{ijk'})}} \\
&= \frac{\rho + \sigma_1^2}{1 + \sigma_1^2 + \sigma_2^2}.
\end{aligned}
$$

Using the similar approach we have

$$cov(Y^*_{ijk}, Y^*_{ij'k}) = 0 + cov(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}, \mathbf{x}^\top_{ij'}\boldsymbol{\beta} + b_{1i} + b_{2k}) = \sigma_1^2 + \sigma_2^2,$$
$$cov(Y^*_{ijk}, Y^*_{ij'k'}) = 0 + cov(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}, \mathbf{x}^\top_{ij'}\boldsymbol{\beta} + b_{1i} + b_{2k'}) = \sigma_1^2,$$
$$cov(Y^*_{ijk}, Y^*_{i'j'k}) = 0 + cov(\mathbf{x}^\top_{ij}\boldsymbol{\beta} + b_{1i} + b_{2k}, \mathbf{x}^\top_{i'j'}\boldsymbol{\beta} + b_{1i'} + b_{2k}) = \sigma_2^2,$$

leading to the following marginal correlations

$$corr(Y^*_{ijk}, Y^*_{ij'k}) = \frac{\sigma_1^2 + \sigma_2^2}{1 + \sigma_1^2 + \sigma_2^2},$$
$$corr(Y^*_{ijk}, Y^*_{ij'k'}) = \frac{\sigma_1^2}{1 + \sigma_1^2 + \sigma_2^2},$$
$$corr(Y^*_{ijk}, Y^*_{i'j'k}) = \frac{\sigma_2^2}{1 + \sigma_1^2 + \sigma_2^2}.$$

Note that in the above calculation we have used the fact that the pairs $\{Y^*_{ijk}, Y^*_{ij'k}\}$, $\{Y^*_{ijk}, Y^*_{ij'k'}\}$ and $\{Y^*_{ijk}, Y^*_{i'j'k}\}$ are independent conditional on the random effects. Also note that the correlation $corr(Y^*_{ijk}, Y^*_{ijk'})$ depends on the (conditional) tetrachoric correlation

$$\rho = corr(Y^*_{ij1}, Y^*_{ij2} | b_{1i}, b_{2k}, b_{2k'}),$$

which can be obtained from $\widetilde{\rho}^*$ by piecewise linear approximation (Kugiumtzis and Bora-Senta, 2010).

Chapter 3

Simulation Study

In this chapter, a simulation study is carried out to study the finite-sample behaviors of the estimates obtained via pairwise likelihood method for the GCMM fitted to clustered binary data with crossed random effects. The data were generated using the package `copula` in `R` and the models were fitted using self written `R` codes that utilize the `optim` function for pairwise likelihood estimation. The simulations design and the results are discussed in the following sections.

3.1 Simulation design

A design analogous to the diabetic retinopathy data was adopted for the simulation study. In particular, we considered one pathology ($V = 1$) and two readers ($K = 2$). We assumed that the assessment Y_{ijk} on this single pathology by reader k for eye j of patient i, has an underlying continuous latent variable Y^*_{ijk}, such that $Y_{ijk} = \mathrm{I}\{Y^*_{ijk} > 0\}$ and $Y^*_{ijk}|B_{1i}, B_{2k} \sim logistic(\mu^*_{ijk}(B_{1i}, B_{2k}), 1)$ (i.e., the logistic distribution with unit scale). So the conditional density $f_{Y^*_{ijk}|B_{1i},B_{2k}}(\cdot|\cdot)$ of Y^*_{ijk}, given B_{1i} and B_{2k}, which is the conditional margin to be

considered for the GCMM in (2.12), is

$$f_{Y^*_{ijk}|B_{1i},B_{2k}}(y^*_{ijk}|b_{1i},b_{2k}) = \frac{e^{-\{y^*_{ijk}-\mu^*_{ijk}(b_{1i},b_{2k})\}}}{\left[1+e^{-\{y^*_{ijk}-\mu^*_{ijk}(b_{1i},b_{2k})\}}\right]^2},$$

where $i = 1, \ldots, N$, $j = L, R$, $k = 1, \ldots, K$. The conditional mean using which data were simulated is

$$\mu^*_{ijk}(B_{1i}, B_{2k}) = \beta_0 + \beta_1\, D_{ij} + B_{1i} + B_{2k}, \tag{3.1}$$

where D_{ij} is the true disease status for eye j of patient i, $B_{1i} \overset{iid}{\sim} N(0, \sigma_1^2)$, and $B_{2k} \overset{iid}{\sim} N(0, \sigma_2^2)$. The (conditional) cumulative distribution function of the bivariate Gaussian copula distribution for $\mathbf{Y}^*_{ij} = (Y^*_{ij1}, Y^*_{ij2})^\top$, given B_{1i} and $\mathbf{B}_2 = (B_{21}, B_{22})^\top$, is

$$F_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\mathbf{y}^*_{ij}|b_{1i},\mathbf{b}_2) \;=\; \Phi_2(\Phi^{-1}\{u^*_{ij1}(b_{1i},b_{21})\}, \Phi^{-1}\{u^*_{ij2}(b_{1i},b_{22})\}; \widetilde{\rho}^*),$$

where $u^*_{ijk}(b_{1i}, b_{2k}) = F_{Y^*_{ijk}|B_{1i},B_{2k}}(y^*_{ijk}|b_{1i}, b_{2k})$ and

$$\widetilde{\rho}^* = corr\{\Phi^{-1}(u^*_{ij1}(b_{1i},b_{21})), \Phi^{-1}(u^*_{ij1}(b_{1i},b_{22}))\}.$$

For simulations we generated data using the following two different true parameter settings:

Scenario I : $\beta_0 = -3$, $\beta_1 = 5$, $\sigma_1 = 1$, $\sigma_2 = 0.1$ and $\widetilde{\rho}^* = 0.6$

Scenario II : $\beta_0 = -3$, $\beta_1 = 5$, $\sigma_1 = 1$, $\sigma_2 = 0.2$ and $\widetilde{\rho}^* = 0.4$.

We kept the value of the variance component σ_2 smaller than the component σ_1, because it is expected that the variation of the assessments on same eye by different readers should not be big as they receive similar training. After specifying the parameters $\mathbf{\Theta} = (\beta_0, \beta_1, \sigma_1, \sigma_2, \widetilde{\rho}^*)^\top$ the clustered binary responses were generated using the following algorithm:

1. Generate B_{2k} independently from $N(0, \sigma_2^2)$, for $k = 1, 2$.

2. Independently generate the true disease statuses D_{iL} and D_{iR}, for patient i from the Bernoulli distribution with success probability 0.5.

3. Independently generate $B_{1i} \sim N(0, \sigma_1^2)$.

4. Use package `copula` (Yan, 2007) to jointly generate y^*_{ij1} and y^*_{ij2} using the logistic margins $logistic(\beta_0 + \beta_1 D_{ij} + B_{1i} + B_{2k}, 1)$ for $k = 1, 2$.

5. Dichotomize y^*_{ijk} as $y_{ijk} = \mathrm{I}\{y^*_{ijk} > 0\}$ to obtain y_{ijk}, $k = 1, 2$.

6. Repeat Steps 2 to 5 for N times.

The univariate and bivariate probabilities for constructing the pairwise likelihood in (2.15) are computed as follows:

$$
\begin{aligned}
P(Y_{ijk} = 0 | b_{1i}, b_{2k}) &= F_L(-\mu^*_{ijk}(b_{1i}, b_{2k})), \\
P(Y_{ijk} = 1 | b_{1i}, b_{2k}) &= F_L(\mu^*_{ijk}(b_{1i}, b_{2k})), \\
P(Y_{ij1} = Y_{ij2} = 0 | b_{1i}, b_{21}, b_{22}) &= \Phi_2 \left(\begin{array}{c} \Phi^{-1}\{F_L(-\mu^*_{ij1}(b_{1i}, b_{21}))\}, \\ \Phi^{-1}\{F_L(-\mu^*_{ij2}(b_{1i}, b_{22}))\} \end{array} ; \widetilde{\rho}^* \right), \\
P(Y_{ij1} = 1, Y_{ij2} = 0 | b_{1i}, b_{21}, b_{22}) &= F_L(-\mu^*_{ij2}(b_{1i}, b_{22})) \\
&\quad - \Phi_2 \left(\begin{array}{c} \Phi^{-1}\{F_L(-\mu^*_{ij1}(b_{1i}, b_{21}))\}, \\ \Phi^{-1}\{F_L(-\mu_{ij2}(b_{1i}, b_{22}))\} \end{array} ; \widetilde{\rho}^* \right), \\
P(Y_{ij1} = 0, Y_{ij2} = 1 | b_{1i}, b_{21}, b_{22}) &= F_L(-\mu^*_{ij1}(b_{1i}, b_{21})) \\
&\quad - \Phi_2 \left(\begin{array}{c} \Phi^{-1}\{F_L(-\mu^*_{ij1}(b_{1i}, b_{21}))\}, \\ \Phi^{-1}\{F_L(-\mu^*_{ij2}(b_{1i}, b_{22}))\} \end{array} ; \widetilde{\rho}^* \right), \\
P(Y_{ij1} = Y_{ij2} = 1 | b_{1i}, b_{21}, b_{22}) &= 1 + \Phi_2 \left(\begin{array}{c} \Phi^{-1}\{F_L(-\mu^*_{ij1}(b_{1i}, b_{21}))\}, \\ \Phi^{-1}\{F_L(-\mu^*_{ij2}(b_{1i}, b_{22}))\} \end{array} ; \widetilde{\rho}^* \right) \\
&\quad - F_L(-\mu^*_{ij1}(b_{1i}, b_{21})) - F_L(-\mu^*_{ij2}(b_{1i}, b_{22})),
\end{aligned}
$$

where $F_L(.)$ is the CDF of the *logistic*-distribution. The above probabilities were computed using the functions `pmvnorm`, `qnorm`, and `plogis` in `R`. The maximum pairwise likelihood estimate (MPLE) for $\boldsymbol{\Theta}$ was obtained using the `optim` function in `R`.

Sensitivities, specificities, and their standard errors were calculated for each simulation runs. Expressions for sensitivities and specificities were obtained analogously as in Withanage et al. (2015) as follows:

$$\begin{aligned}
\text{Sen} &= \frac{1}{\widehat{\sigma}_1\widehat{\sigma}_2}\int_{\mathcal{R}^2} F_L(\widehat{\beta}_0+\widehat{\beta}_1+b_{1i}+b_{2k})\phi\left(\frac{b_{1i}}{\widehat{\sigma}_1}\right)\phi\left(\frac{b_{2k}}{\widehat{\sigma}_2}\right)db_{1i}db_{2k},\\
\text{Spc} &= 1-\frac{1}{\widehat{\sigma}_1\widehat{\sigma}_2}\int_{\mathcal{R}^2} F_L(\widehat{\beta}_0+b_{1i}+b_{2k})\phi\left(\frac{b_{1i}}{\widehat{\sigma}_1}\right)\phi\left(\frac{b_{2k}}{\widehat{\sigma}_2}\right)db_{1i}db_{2k},
\end{aligned} \tag{3.2}$$

where $F_L(.)$ is the CDF of logistic distribution with unit scale. The variances for the estimates of these accuracy measures were obtained using the delta method as follows:

$$\begin{aligned}
var(\widehat{\text{Sen}}) &= \Big(\frac{\partial}{\partial\boldsymbol{\Theta}}\text{Sen}\Big)\mathbf{G}(\boldsymbol{\Theta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\Theta}}\text{Sen}\Big)^{\top}\Big|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}},\\
var(\widehat{\text{Spc}}) &= \Big(\frac{\partial}{\partial\boldsymbol{\Theta}}\text{Spc}\Big)\mathbf{G}(\boldsymbol{\Theta})^{-1}\Big(\frac{\partial}{\partial\boldsymbol{\Theta}}\text{Spc}\Big)^{\top}\Big|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}},
\end{aligned}$$

where $\mathbf{G}(\widehat{\boldsymbol{\Theta}})$ is the estimated Godambe information matrix. Standard errors were calculated by taking square root of these variances. The derivatives in the above variances were calculated using the `fdHess` function of `nlme` package in `R`.

3.2 Simulation results

We conducted simulation study to see the performance of the parameter estimates obtained via pairwise likelihood estimation under Gaussian copula mixed model approach for clustered binary data with crossed random effects. For each of the parameter settings (Scenario I and II), as defined in Section 3.1, simulation studies were conducted for two different sample sizes, $N = 100$ and 500. The results are presented in Tables 3.1, 3.2, 3.3 and 3.4. Tables

3.1 and 3.2 display the results for Scenario I with sample sizes 100 and 500, respectively. Whereas Tables 3.3 and 3.4 display the results for Scenario II for the two sample sizes. The performance of the estimates were evaluated based on the following:

$$\text{Relative bias} = 100 \times \frac{\text{Average estimate} - \text{True parameter value}}{\text{True parameter value}},$$

$$\text{Relative efficiency} = \frac{\text{Average SE}}{\text{Empirical SD}},$$

where SE and SD are standard error and standard deviation, respectively, and the coverage probability (with true level 95%). The coverage probability was calculated as the proportion of times out of 100 repeated samples, the 95% confidence intervals contained the true parameter value. Sensitivities and specificities (averaged over the repeated samples), their relative biases, average standard errors, empirical standard deviations and relative efficiencies are also presented in the tables.

In Table 3.1, we see that the relative biases are very small for the location parameters (β_0 and β_1), as well as for the normal correlation ($\widetilde{\rho}^*$); ranging from one to five percent. But the variance components, specially the component (σ_2) associated with the reader-specific random effects exhibits a high downward relative bias. Same behavior is observed in case of relative efficiency and empirical coverage probability. Relative efficiencies are close to one for all the parameters except for σ_2. The coverage probabilities are close to the true level 0.95 for β_0, β_1 and $\widetilde{\rho}^*$, but a bit far from the true level for σ_1. For σ_2 the coverage probability is far less than the true level, which is definitely the result of high bias in estimation. Since sensitivity and specificity involve the variance components, the biases and inefficiencies in those estimates caused a high bias in sensitivity estimate. The numbers for relative efficiencies are also very big for the sensitivity and specificity estimates.

However, the interesting thing was observed in simulation results when we increased the sample size from 100 to 500 keeping the parameter values same (results are shown in Table 3.2). The absolute relative bias decreased from 14.8% to 3.4% for variance component σ_1,

while for σ_2 a huge improvement was observed; absolute bias decreased from 52.7% to 12%. We can see similar improvement in coverage probabilities for these variance components. The coverage probability increased from 0.79 to 0.89 for σ_1 and from 0.48 to 0.92 for σ_2 (a notable improvement indeed!). Improvements in the estimation of the variance components have largely improved the estimation of sensitivity and specificity as reflected by the relative biases, which are close to zero, and the relative efficiencies, which are close to one.

Table 3.1: Estimates from pairwise likelihood estimation under Gaussian copula mixed model for Scenario I, with 100 repeated samples, $K = 2$ readers, $V = 1$ pathology, $N = 100$ patients.

Parameter	Average estimate	Relative bias (%)	Average SE	Empirical SD	Relative efficiency	Coverage probability
$\beta_0 = -3$	–3.0718	2.3947	0.5357	0.6411	0.8356	0.9100
$\beta_1 = 5$	5.0505	1.0100	0.7953	0.8742	0.9098	0.8700
$\sigma_1 = 1$	0.8517	–14.8317	0.5479	0.6042	0.9067	0.7900
$\sigma_2 = 0.1$	0.0473	–52.6556	0.0441	0.0567	0.7772	0.4800
$\widetilde{\rho}^* = 0.6$	0.5731	–4.4880	0.1596	0.1747	0.9132	0.9100
Sen = 0.8442	0.6483	–23.2068	153.8060	0.3305	465.3523	
Spc = 0.9305	0.9463	1.7007	15.1639	0.0337	449.7245	

Table 3.2: Estimates from pairwise likelihood estimation under Gaussian copula mixed model for Scenario I, with 100 repeated samples, $K = 2$ readers, $V = 1$ pathology, $N = 500$ patients.

Parameter	Average estimate	Relative bias (%)	Average SE	Empirical SD	Relative efficiency	Coverage probability
$\beta_0 = -3$	–3.0781	2.6033	0.3365	0.2805	1.1996	0.9000
$\beta_1 = 5$	5.1246	2.4924	0.5072	0.3776	1.3433	0.9000
$\sigma_1 = 1$	1.0340	3.4027	0.3559	0.3139	1.1340	0.8900
$\sigma_2 = 0.1$	0.1120	12.0393	0.0668	0.0437	1.5294	0.9200
$\widetilde{\rho}^* = 0.6$	0.5820	–3.0063	0.1017	0.0896	1.1350	0.8800
Sen = 0.8442	0.8447	0.0562	0.0168	0.0165	1.0219	
Spc = 0.9305	0.9309	0.0528	0.0107	0.0112	0.9550	

For the second setting of parameters, where we slightly increased the value of σ_2 and decreased the value of $\widetilde{\rho}^*$ than those in setting one, we observed better relative efficiencies for

all parameter estimates for both sample sizes (see Tables 3.3 and 3.4). The estimates of regression coefficients exhibit negligible biases even for sample size 100 like previous scenario. As before the scale parameter estimates exhibit high bias for smaller sample size and the bias decreases with increase in sample size. With increasing sample size efficiencies get closer to one, also the coverage fractions for the variance components, specially for σ_1, get closer to the nominal level 0.95.

Table 3.3: Estimates from pairwise likelihood estimation under Gaussian copula mixed model for Scenario II, with 100 repeated samples, $K = 2$ readers, $V = 1$ pathology, $N = 100$ patients.

Parameter	Average estimate	Relative bias (%)	Average SE	Empirical SD	Relative efficiency	Coverage probability
$\beta_0 = -3$	–3.0596	1.9870	0.5825	0.5872	0.9919	0.9300
$\beta_1 = 5$	5.0789	1.5785	0.8793	0.8346	1.0536	0.9300
$\sigma_1 = 1$	0.8891	–11.0888	0.6428	0.5518	1.1649	0.8300
$\sigma_2 = 0.2$	0.0814	–59.3012	0.1241	0.0961	1.2923	0.5100
$\widetilde{\rho}^* = 0.4$	0.3661	–8.4709	0.2303	0.2047	1.1248	0.9400
Sen = 0.8433	0.7014	–16.8200	0.0259	0.2811	0.0920	
Spc = 0.9298	0.9426	1.3836	0.0175	0.0306	0.5716	

Table 3.4: Estimates from pairwise likelihood estimation under Gaussian copula mixed model for Scenario II, with 100 repeated samples, $K = 2$ readers, $V = 1$ pathology, $N = 500$ patients.

Parameter	Average estimate	Relative bias (%)	Average SE	Empirical SD	Relative efficiency	Coverage probability
$\beta_0 = -3$	–3.0303	1.0095	0.2539	0.2556	0.9933	0.8700
$\beta_1 = 5$	5.0733	1.4660	0.3543	0.3563	0.9944	0.8700
$\sigma_1 = 1$	1.0369	3.6853	0.2352	0.2630	0.8942	0.9200
$\sigma_2 = 0.2$	0.1307	–34.6501	0.0729	0.0690	1.0566	0.7400
$\widetilde{\rho}^* = 0.4$	0.3690	–7.7519	0.0974	0.0996	0.9774	0.9400
Sen = 0.8433	0.8438	0.0618	0.0178	0.0204	0.8672	
Spc = 0.9298	0.9282	–0.1724	0.0110	0.0141	0.7771	

3.3 Summary

In this chapter, we carried out simulation studies to investigate the finite-sample properties of PMLES obtained from GCMM approach to clustered binary data with crossed random effects. Simulation results reveal that the estimation of the variance components is really problematic when we have smaller sample size (like 100), although the estimation of the regression coefficients and the normal correlation is okay. The estimates of regression coefficients and the normal correlation exhibit negligible biases, efficiencies close to one and coverage fractions close to the nominal level (0.95) even for smaller sample size. The estimates of variance components (σ_1 and σ_2), specially the component σ_2 associated with the reader-specific random effects, exhibit high biases, efficiencies far from one and coverage fractions far from 0.95 for smaller sample size. But with increase in the sample size from 100 to 500, the bias decreases and the efficiency gets closer to one. Also, the coverage fraction tends to be closer to the nominal level with larger sample size.

Chapter 4

Analysis of diabetic retinopathy data

In this chapter, we first describe the diabetic retinopathy data in Section 4.1, and then discuss the models in Section 4.2, which we have considered for analyzing the data. Section 4.3 presents the analysis results followed by a brief summary in the end of this chapter.

4.1 Diabetic retinopathy data

The diabetic retinopathy study was conducted in Edmonton, Canada, between February 1, 2000, and June 1, 2000. All new diabetic patients referred to a comprehensive retina practice were eligible for inclusion in the study irrespective of the reason for the referral. Patients underwent clinical examination (i.e., 'gold standard') of the retina after pupillary dilatation using 1 drop of diophenyl T in each eye. At this time, the presence or absence of CSME, microaneurysms, intraretinal hemorrhage, hard exudate, and other disease of note, were recorded as present or absent (i.e., 'true disease status'). Patients then underwent stereoscopic digital fundus photography by a trained ophthalmic photographer using a high-resolution digital camera. Digital photographs were taken on the same day as the clinical examination. Two retinal specialists (i.e., 'readers') enrolled all study patients, and a minimum of 2 months between clinical examination and photographic grading was allowed to minimize reader recall. The readers were masked to the clinical grading of each eye. Digital

photographs of the left eye were reviewed in random order, with a minimum of 2 months before review of right eyes. Images were viewed on a computer monitor through liquid crystal diode shutter goggles using a three-dimensional viewing software. The number of images viewed per eye was at the discretion of the reader, with the sharpest images used for the final grading. The reader had the option of zooming into view the image at the maximum pixel resolution. During the enrollment period, there were 139 new patients with diabetes mellitus. Twenty three patients (32 eyes) were not eligible for enrollment: 6 patients (12 eyes) were physically unable to sit at the fundus camera because of fatigue, prior stroke, incontinence, illness, or 9 physical size, 10 patients (16 eyes) had media opacities preventing adequate clinical evaluation, 6 patients (3 eyes) were unwilling to be photographed on the same day as the clinical examination, and 1 patient (1 eye) had retinal disease preventing differentiation of diabetic retinopathy. A total of 116 patients were examined clinically and received same-day high resolution stereoscopic digital fundus photography, of which 11 patients (19 eyes) were excluded after enrollment because the digital image files were lost, and the photos could therefore not be graded, and 5 patients who had data on only 1 eye (3 right eyes and 2 left eyes) were eventually dropped from the analysis. In total, 200 eyes of 100 patients were included in the final analysis. Power calculations were performed to verify that the sample size was adequate to evaluate the extent of agreement between the diagnostic methods. Finally, there were N = 94 patients with complete data. More information about the study can be found in Rudnisky et al. (2002).

4.2 Fitted models

In this section, we illustrate our methodology on the diabetic retinopathy data for pathology CSME. Let Y_{ijk} be the assessment on eye j of patient i by reader k, where $i = 1, \ldots, N = 94$, $j = L, R$, and $k = 1, 2$. The GCMM we considered for modeling the binary 2-dimensional

response vector $\mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2})^\top$ is

$$P(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | B_{1i} = b_{1i}, \mathbf{B}_2 = \mathbf{b}_2) \;=\; \int_{A_{ij1} \times A_{ij2}} f_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\mathbf{y}^*_{ij}|b_{1i}, \mathbf{b}_2) d\mathbf{y}^*_{ij}, \tag{4.1}$$

where $\mathbf{Y}^*_{ij} = (Y^*_{ij1}, Y^*_{ij2})^\top$ is the underlying latent response vector, $\mathbf{b}_2 = (b_{21}, b_{22})^\top$, $B_{1i} \overset{iid}{\sim} N(0, \sigma_1^2)$, $B_{2k} \overset{iid}{\sim} N(0, \sigma_2^2)$, the intervals A_{ijk} are either $(-\infty, 0]$ or $(0, +\infty)$ according as whether $y_{ijk} = 0$ or 1, and $f_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\cdot|\cdot)$ is the (conditional) density of $\mathbf{Y}^*_{ij}$ corresponding to CDF

$$F_{\mathbf{Y}^*_{ij}|B_{1i},\mathbf{B}_2}(\mathbf{y}^*_{ij}|b_{1i}, \mathbf{b}_2) \;=\; \Phi_2(\Phi^{-1}\{u^*_{ij1}(b_{1i}, b_{21})\}, \Phi^{-1}\{u^*_{ij2}(b_{1i}, b_{22})\}; \widetilde{\rho}^*),$$

where $\widetilde{\rho}^*$ is the normal correlation and $u^*_{ijk}(b_{1i}, b_{2k}) = F_{Y^*_{ijk}|B_{1i},B_{2k}}(y^*_{ijk}|b_{1i}, b_{2k})$ is the realization of the latent (conditional) probability integral transform (PIT) $U^*_{ijk}(B_{1i}, B_{2k}) = F_{Y^*_{ijk}|B_{1i},B_{2k}}(Y^*_{ijk}|B_{1i}, B_{2k})$. For the analysis of diabetic retinopathy data, we considered GCMM with three different conditional margins; t-t, logistic-logistic, and Gaussian-Gaussian margins. Following are the conditional densities used as margins in the GCMM (4.1):

$$t\text{-density} : f_{Y^*_{ijk}|B_{1i},B_{2k}}(y_{ijk}|b_{1i}, b_{2k}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu}\{y^*_{ijk} - \mu^*_{ijk}(b_{1i}, b_{2k})\}^2\right)^{-(\nu+1)/2},$$

$$\text{logistic-density} : f_{Y^*_{ijk}|B_{1i},B_{2k}}(y_{ijk}|b_{1i}, b_{2k}) = \frac{e^{-\left(y^*_{ijk} - \mu^*_{ijk}(b_{1i}, b_{2k})\right)}}{\left[1 + e^{-\left(y^*_{ijk} - \mu^*_{ijk}(b_{1i}, b_{2k})\right)}\right]^2},$$

$$\text{Gaussian-density} : f_{Y^*_{ijk}|B_{1i},B_{2k}}(y_{ijk}|b_{1i}, b_{2k}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y^*_{ijk} - \mu^*_{ijk}(b_{1i}, b_{2k})\right)^2},$$

where ν is the parameter (degrees of freedom) of the t distribution, and the scale parameters of the logistic and Gaussian distributions are considered as unity. The conditional mean is expressed as

$$\mu^*_{ijk}(b_{1i}, b_{2k}) = \beta_0 + \beta_1 D_{ij} + b_{1i} + b_{2k},$$

where D_{ij} is the true disease status for eye j of patient i.

The use of t-latent distributions is equivalent to the so-called robit regression, a generalization of and more robust alternative to both logistic and probit regression (de Leon and Wu, 2011). To estimate the degrees of freedom ν, we used the method of profile likelihood (Song et al., 2007) adopted in de Leon and Wu (2011) and Wu and de Leon (2014). The likelihood attained its maximum at $\nu = \widehat{\nu} = 2.5$, so that the PL estimate $\widehat{\boldsymbol{\Theta}}$ was obtained at $\widehat{\nu} = 2.5$. For all three margins SEs of $\widehat{\boldsymbol{\Theta}}$ were computed based on Monte Carlo simulation as outlined in Section 2.3.2. For simulations we used package `copula` to jointly generate y^*_{ij1} and y^*_{ij2} with location parameter $\mu^*_{ijk}(B_{1i}, B_{2k})$ under the three different margins. Since `copula` only supports central t_ν-distributions (i.e., with zero means), we first generated $y^\dagger_{ij1}$ and $y^\dagger_{ij2}$ jointly using the central t_ν-margins and used the transformations $y^*_{ij1} = y^\dagger_{ij1} + \widehat{\beta}_0 + \widehat{\beta}_1 D_{ij} + B_{1i} + B_{21}$ and $y^*_{ij2} = y^\dagger_{ij2} + \widehat{\beta}_0 + \widehat{\beta}_1 D_{ij} + B_{1i} + B_{22}$, for $j = L, R$. According to the properties of t-distribution (Kotz and Nadarajah, 2004, p. 15) this is equivalent to jointly generating y^*_{ij1} and y^*_{ij2} from a bivariate Gaussian copula with normal correlation $\widehat{\widetilde{\rho}}^*$, and respective margins $t_\nu(\widehat{\beta}_0 + \widehat{\beta}_1 D_{ij} + B_{1i} + B_{21}, 1)$ and $t_\nu(\widehat{\beta}_0 + \widehat{\beta}_1 D_{ij} + B_{1i} + B_{22}, 1)$.

4.3 Analysis results

Table 4.1 compares the PL estimates for CSME for the GCMM under three different margins we considered. The estimates of regression coefficients and the variance components, and their SEs are similar under t-t and logistic-logistic margins, but for Gaussian margins these estimates are quite different. The SEs obtained under Gaussian margins are notably smaller than those obtained under the other two margins. This suggests that more careful investigation is needed before concluding which margin better fits the diabetic retinopathy data for pathology CSME. However, the estimates of normal correlation $\widetilde{\rho}^*$ and its SEs are similar under all three margins. Interestingly, estimates of sensitivities and specificities for CSME under all three margins are almost same. Furthermore, compared with the results in Withanage et al. (2015) obtained under DC, the estimated sensitivities and specificities are

slightly higher in the present case.

Table 4.1: PL estimates of the GCMM, with their SEs, for pathology CSME along with the corresponding estimates of sensitivities and specificities under different margins.

Parameter	$t-t$		logistic-logistic		Gaussian-Gaussian	
	Est	SE	Est	SE	Est	SE
β_0	–2.7919	0.6692	–3.1712	0.5677	-1.7598	0.2775
β_1	4.5071	1.0717	5.1824	0.8595	2.8979	0.4564
σ_1	1.2163	0.5622	1.3001	0.5254	0.6929	0.3202
σ_2	0.1497	0.3692	0.1382	0.3837	0.0573	0.3762
$\widetilde{\rho}^*$	0.5554	0.1998	0.5697	0.1832	0.5716	0.1670
ρ	0.4381	—	0.5511	—	0.5572	—
Sen	0.8249	0.0484	0.8249	0.0458	0.8250	0.0492
Spc	0.9258	0.0220	0.9258	0.0225	0.9257	0.0218

The piecewise linear approximation of Kugiumtzis and Bora-Senta (2010) was used to estimate the (conditional) tetrachoric correlation $\rho = corr(Y^*_{ij1}, Y^*_{ij2}|b_{1i}, b_{21}, b_{22})$ between assessments by the two readers of a patient's eye. The marginal correlations assessing different associations between the readers' assessments are shown in Table 4.2. As we expected, the highest correlation is reported for those assessments by different readers of the same eye of a patient. Comparing the correlations with those reported in Withanage et al. (2015), the ones based on the GCMM with t-t margins are much lower than those from the DC method. One reason might be the underlying latent distribution adopted, which is a $t_{2.5}$-distribution. As such, $var(Y_{ijk}|B_{1i}, B_{2k}) = 5 > 1$ is considerably higher than the unit variance of the standard Gaussian latent distribution in Withanage et al. (2015).

4.4 Summary

In this chapter, we analyzed the diabetic retinopathy data, which arose from a reader-based diagnostic study in Alberata, Canada, where two retinal specialists (readers) diagnosed the presence or absence of particular pathologies in the fellow eyes of several diabetic patients.

Table 4.2: Estimated marginal tetrachoric correlations under different margins for the diabetic retinopathy data for pathology CSME. No SE is given for $corr(Y_{ij1}, Y_{ij2})$, as it involves ρ, for which no SE was available.

Correlation	$t-t$		logistic-logistic		Gaussian-Gaussian	
	Est	SE	Est	SE	Est	SE
$corr(Y^*_{iLk}, Y^*_{iRk})$	0.2310	0.1608	0.6309	0.1863	0.3259	0.2043
$corr(Y^*_{iL1}, Y^*_{iR2})$	0.2275	0.1630	0.6239	0.1918	0.3237	0.2018
$corr(Y^*_{iLk}, Y^*_{i'Rk})$	0.0034	0.0170	0.0070	0.0391	0.0022	0.0290
$corr(Y^*_{ij1}, Y^*_{ij2})$	0.5646	—	0.8273	—	0.7090	—

We considered the diagnosis of pathology CSME (clinically significant macular edema) for the analysis purpose. Since there are four observations from the same patients, we have binary data with clustered nature. Also the readers have received similar training, so their assessments of the same eye of a patient are likely to be correlated. Since the goal was to estimate the sensitivity and specificity with respect to true disease status and there are two random sources of variation in the data (patients and readers), it is reasonable to fit a mixed effects model to this data. We applied the Gaussian copula mixed model (GCMM), as considered in the simulation study, where the normal correlation of the bivariate Gaussian copula distribution accounts for the between reader correlation for the same eye of a patient. We fitted the GCMM to this data using three different conditional margins. The pairwise likelihood estimation was used to obtain the estimates of regression coefficients, the variance components associated with the random effects and the conditional correlation. The estimates for sensitivity and specificity were almost same under all three margins, although there was slight variation in the estimates and their SEs for the rest of the parameters. The estimated sensitivity and specificity were 82.5% and 92.5% (irrespective of the margins), respectively, which are slightly higher than those obtained for the same data under multivariate probit model with full likelihood estimation using data cloning approach in Withanage et al. (2015).

Chapter 5

Discussion

Motivated by the data from a diabetic retinopathy study, where several readers (retinal specialists or opthalmologists) assessed the presence or absence of certain pathologies in a diabetic patient's left and right eyes, we proposed a regression approach for correlated binocular binary diagnostic data based on a GCMM. There is a complex correlation structure present in the data: in addition to the fellow-eye correlation induced by the binocular nature of the data, two other sources of correlation are present. Since readers rely on the same image of the eye, their diagnoses are potentially correlated. Moreover, because the pathologies are all related to retinopathy, it is very likely that the presence or absence of one influences the presence or absence of another. Thus, diagnoses for pathologies are correlated as well. Since we considered a single pathology CSME for our analysis, cross-correlation between pathologies was not a problem for our case. The correlations between the assessments on fellow eyes of a patient by the same or different readers, were handled by adding two random effects in the model. The correlation between the assessments on the same eye by different readers was introduced in the model through the normal correlation of a bivariate Gaussian copula distribution (since there were two readers only). A pairwise maximum likelihood estimation was used instead of full ML estimation to reduce the computational complexity duo to presence of a large number of integrals in the likelihood function. Before analyzing

the data under GCMM with PL estimation, we conducted a simulation study to see the performance of PMLEs using a simulation setup analogous to the diabetic retinopathy data.

The simulation studies were carried out for two different parameter settings each with two different sample sizes ($N = 100, 500$). Although the estimation of the regression coefficients and the normal correlation did not seem to be problematic, the estimation of the variance components was greatly affected by the smaller sample size. The estimates of regression coefficients and the normal correlation exhibit negligible biases, efficiencies close to one and coverage fractions close to the true nominal level (0.95); irrespective of different parameter settings and even for the smaller sample size. The estimates of variance components (σ_1 and σ_2), specially the component (σ_2) associated with the reader-specific random effects, exhibit high biases, efficiencies far from one and coverage fractions far from 0.95 for the smaller sample size. But for these estimates, the bias decreases and the efficiency gets closer to one with increase in sample size. Also, the coverage fractions tend to be closer to the nominal level with larger sample size. However, $N = 500$ seems to be (still!) not sufficiently large for asymptotic normality to kick in for the estimate of σ_2 as reflected by the performance measures.In most settings, $N = 500$ may be large enough for MLEs to exhibit reasonably good approximate normality, but: (1) we used PMLE, not MLE, and the PMLE is less efficient than the MLEs (so that a sample size larger than N is necessary to attain the same efficiency of the MLE at sample size N), and (2) we had crossed random effects, indicating that that the data were not really independent, unlike in other mixed models with nested random effects, and this again has implications on efficiency. Overall, our simulation results so far suggest that it will take a much bigger sample size for the normal approximation (to the exact (sampling) distribution of PMLEs for the scale parameters) to provide reasonably accurate approximate results.

We analyzed the diabetic retinopathy data with GCMM under three different conditional margins. The estimates of sensitivities and specificities were almost same under all three margins, although the model parameter estimates and SEs varied slightly. The estimated

marginal associations were found as what we had expected. The assessments on same eyes of a patients by different readers had much larger correlation than the other pairs of assessments. Also, the between subject marginal correlations were close to zero, which is common in almost all study. Since the sample size for this data was 94 and as we noted in the simulation study that small sample size may lead to really bad PMLEs for scale parameters, one should be careful while analyzing such small data set with GCMM under PL estimation.

Due to time constraint we conducted the simulation study with only 100 independent repeated samples and a small number of Monte Carlo samples (only 20 per replication!) for variance estimation of the PMLEs. This small numbers may also have effect on calculation of relative biases, efficiencies and coverage fractions. Further simulation studies can be done increasing these numbers, say for example, 500 repeated samples with 1000 Monte Carlo samples per replication, keeping rest of the simulation setting same to see if the performance measures improve. Also, as PMLEs have shown to perform better with increasing sample size, further detail investigation is needed for the finite sample properties of PMLEs with a bigger sample size than 500.

Bibliography

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Bellio, R. and Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, 5(3):217–227.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):265–285.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, 87(419):651–658.

Chan, J. S. and Kuk, A. Y. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics*, pages 86–97.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4):335–359.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.

De Leon, A., Guo, M., Rudnisky, C., and Singh, G. (2007). A likelihood approach to estimating sensitivity and specificity for binocular data: Application in ophthalmology. *Statistics in Medicine*, 26(17):3300–3314.

de Leon, A. R., Soo, A., Bonzo, D. C., and Rudnisky, C. J. (2009). Joint estimation of diagnostic accuracy measures for paired organs–application in ophthalmology. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 51(5):837–850.

de Leon, A. R. and Wu, B. (2011). Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, 30(2):175–185.

Evans, M. and Swartz, T. (2000). *Approximating integrals via Monte Carlo and deterministic methods*, volume 20. OUP Oxford.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431.

Genest, C., Nikoloulopoulos, A. K., Rivest, L.-P., Fortin, M., et al. (2013). Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Brazilian Journal of Probability and Statistics*, 27(3):265–284.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. volume 49, pages 732–764.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, 96(455):1102–1112.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association*, 93(442):720–729.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC.

Klaassen, C. A. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 3(1):55–77.

Kotz, S. and Nadarajah, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press.

Kugiumtzis, D. and Bora-Senta, E. (2010). Normal correlation coefficient of non-normal variables using piece-wise linear approximation. *Computational Statistics*, 25(4):645–662.

Lele, S. R., Dennis, B., and Lutscher, F. (2007). Data cloning: easy maximum likelihood estimation for complex ecological models using bayesian markov chain monte carlo methods. *Ecology letters*, 10(7):551–563.

Lele, S. R., Nadeem, K., and Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105(492):1617–1625.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):325–335.

Lin, L., Bandyopadhyay, D., Lipsitz, S. R., and Sinha, D. (2010). Association models for clustered data with binary and continuous responses. *Biometrics*, 66(1):287–293.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–239.

Lipsitz, S. R., Dear, K. B., and Zhao, L. (1994). Jackknife estimators of variance for parameter estimates from estimating equations with applications to clustered survival data. *Biometrics*, pages 842–846.

Maberley, D., Walker, H., Koushik, A., and Cruess, A. (2003). Screening for diabetic retinopathy in james bay, ontario: a cost-effectiveness analysis. *Cmaj*, 168(2):160–164.

McCullagh, P. and Nelder, J. (1989). Generalized linear models., 2nd edn.(chapman and hall: London).

McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89(425):330–335.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170.

Mesfioui, M. and Tajar, A. (2005). On the properties of some nonparametric concordance measures in the discrete case. *Nonparametric Statistics*, 17(5):541–554.

Najita, J. S., Li, Y., and Catalano, P. J. (2009). A novel application of a bivariate regression model for binary and continuous outcomes to studies of fetal toxicity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(4):555–573.

Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, 98(3):544–567.

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine*, 27(30):6393–6406.

Nikoloulopoulos, A. K. and Karlis, D. (2009a). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, 139(11):3878–3890.

Nikoloulopoulos, A. K. and Karlis, D. (2009b). Modeling multivariate count data using copulas. *Communications in Statistics-Simulation and Computation*, 39(1):172–187.

Raaijmakers, J. G., Schrijnemakers, J. M., and Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and language*, 41(3):416–426.

Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002). Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 44(8):921–935.

Renard, D., Molenberghs, G., and Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis*, 44(4):649–667.

Rudnisky, C. J., Hinz, B. J., Tennant, M. T., de Leon, A. R., and Greve, M. D. (2002). High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema. *Ophthalmology*, 109(2):267–274.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1):60–68.

Song, P. X.-K., Zhang, P., and Qu, A. (2007). Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions. *Statistica Sinica*, pages 929–943.

Sutradhar, B. C. and Rao, R. P. (2003). On quasi-likelihood inference in generalized linear mixed models with two components of dispersion. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 415–435.

Tibaldi, F. S., Verbeke, G., Molenberghs, G., Renard, D., Van den Noortgate, W., and De Boeck, P. (2007). Conditional mixed models with crossed random effects. *British Journal of Mathematical and Statistical Psychology*, 60(2):351–365.

Torabi, M. (2012). Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning. *Computational Statistics & Data Analysis*, 56(12):4259–4265.

Verbeke, G., Spiessens, B., and Lesaffre, E. (2001). Conditional linear mixed models. *The American Statistician*, 55(1):25–34.

Withanage, N., de Leon, A. R., and Rudnisky, C. J. (2014). Joint estimation of disease-specific sensitivities and specificities in reader-based multi-disease diagnostic studies of paired organs. *Journal of Applied Statistics*, 41(10):2282–2297.

Withanage, N., de Leon, A. R., and Rudnisky, C. J. (2015). Joint estimation of multiple disease-specific sensitivities and specificities via crossed random effects models for correlated reader-based diagnostic data: Application of data cloning. *Statistics in Medicine*, 34(29):3916–3928.

Wu, B. and de Leon, A. R. (2014). Gaussian copula mixed models for clustered mixed outcomes, with application in developmental toxicology. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(1):39–56.

Wu, B., de Leon, A. R., and Withanage, N. (2013). Joint analysis of mixed discrete and continuous outcomes via copula models. *Analysis of Mixed Data: Methods and Applications*, pages 139–156.

Yan, J. (2007). Enjoy the joy of copulas: with a package copula. *Journal of Statistical Software*, 21(4):1 – 21.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.

Zimmer, D. M. and Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.