 DOI: 10.1111/j.1541-0420.2008.01058.x

Joint Regression Analysis of Correlated Data Using Gaussian Copulas

Peter X.-K. Song,[1,*] Mingyao Li,[2] and Ying Yuan[3]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

[2]Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6021, U.S.A.

[3]Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, U.S.A.

[*]*email:* song@uwaterloo.ca

SUMMARY. This article concerns a new joint modeling approach for correlated data analysis. Utilizing Gaussian copulas, we present a unified and flexible machinery to integrate separate one-dimensional generalized linear models (GLMs) into a joint regression analysis of continuous, discrete, and mixed correlated outcomes. This essentially leads to a multivariate analogue of the univariate GLM theory and hence an efficiency gain in the estimation of regression coefficients. The availability of joint probability models enables us to develop a full maximum likelihood inference. Numerical illustrations are focused on regression models for discrete correlated data, including multidimensional logistic regression models and a joint model for mixed normal and binary outcomes. In the simulation studies, the proposed copula-based joint model is compared to the popular generalized estimating equations, which is a moment-based estimating equation method to join univariate GLMs. Two real-world data examples are used in the illustration.

KEY WORDS: Correlated data; Dispersion models; Gaussian copula; GEEs; Mixed outcomes.

1. Introduction

Multidimensional outcomes are frequently collected from biomedical studies. For example, a vector response variable might comprise multiple measurements from several response variables, such as blood pressure, heart rate, and some heart function indicators for a subject. Other examples of such data include clustered data collected from nuclear families of two parents and one affected child in genetic studies, and longitudinal data with a fixed number of repeated measurements collected, for example, from crossover trials (Jones and Kenward, 1989). Regression analysis is challenged by such data in studying the relationships between vector response variables and associated covariates. It is not uncommon that such data are analyzed via separate univariate regression models, one for each response component.

An estimating equations (EE)-based approach has been widely used in practice to join marginal models for correlated outcomes. The EE method enjoys the robustness against misspecification of fully parametric models, because typically it only requires the correct specification of the first two moments of data distributions. On the other hand, there are a few shortcomings associated with the EE method due to the lack of a fully parametric model, including (i) the loss of estimation efficiency, (ii) the lack of procedures for model assessment and selection, and (iii) the difficulty of incorporating vector outcomes of mixed types.

The burn injury data reported in Fan and Gijbels (1996) is an example of vector outcomes of mixed types. The data contain 981 cases of burn injuries, where two response variables, the disposition of death and the total burn area, are both related to a patient's age. To model the respective relationships, we let the severity of burn injury be $y_1 = \log(\text{burn area} + 1)$, which is a continuous response variable, and let the disposition y_2 be a binary response with 1 for death from burn injury and 0 for survival. To investigate how age (x) affects the severity of burn injury and the probability of death, one may propose two marginal mean models: $\mu_{i1} = \beta_{01} + \beta_{11}x_i$, and $\operatorname{logit}(\mu_{i2}) = \beta_{02} + \beta_{12}x_i$, where $\mu_{i1} = \mathrm{E}(y_{i1} \mid x_i)$ is the expected log-burn area, and $\mu_{i2} = P(y_{i2} = 1 \mid x_i)$ is the probability of death from burn injury for patient i, given the age of the patient. Note that the two regression models have different regression coefficients as well as different link functions (i.e., the identity and the logit). A simple analysis of the data would be just to fit the two marginal models separately, because this can be easily done using existing software, such as `SAS` or `R`. Obviously, this analysis is not efficient as it ignores the correlation between the two response variables. Moreover, as far as a joint regression analysis is concerned, a conditional interpretation for one response variable on others is often of interest. In this burn injury data analysis, we can obtain the conditional distribution of death given burn severity as a function of age, or vice versa (see

Figure 2 in Section 5). However, joining marginal models by a certain working correlation structure in the EE framework, unfortunately, does not lead to such a conditional distribution interpretation.

In this article, we attempt to develop a unified and flexible likelihood framework to join various marginal models and overcome the above shortcomings. For the ease of exposition, we consider the familiar generalized linear models (GLMs) as marginal models. To join marginal GLMs, we invoke Gaussian copulas (e.g., Joe, 1997; Song, 2000) as the link model, and the resulting joint regression model is referred to as the *vector GLM* (VGLM). Considering applications of the proposed method to analysis of biomedical data, we focus particularly on the development of the VGLM for correlated discrete outcomes and correlated mixed outcomes. Our simulation studies and data analyses will center on vector logistic models and vector models for mixed outcomes, with comparisons to the moment-based EE approach. In particular, we use two examples to show the gain of a joint analysis in improving estimation efficiency and the flexibility of handling mixed outcomes. The first example presents the detail concerning the analysis of the burn injury data discussed above. The other example illustrates the utility of a three-dimensional vector logistic model in the analysis of a randomized, placebo-controlled trial on multiple sclerosis patients to study the efficacy of fampridine, a compound to enhance nerve conduction (Davis, 2002). Although the illustration is oriented to the two limited scenarios, the proposed method is general enough to be applicable in a much broader range of problems that require a joint regression analysis.

The organization of the article is as follows. In Section 2, we first describe a general strategy of joining marginal models and then briefly review univariate GLM as well as Gaussian copulas. A general theory of the maximum likelihood inference is discussed in Section 3. To illustrate efficiency gain in parameter estimation, Section 4 focuses on three-dimensional VGLMs for discrete data, with comparisons of the respective asymptotic relative efficiencies (AREs) to the popular generalized estimating equations (GEEs). We present two data analyses in Section 5 and give some concluding remarks in Section 6. Some technical details can be found in the Web Supplementary Materials.

2. Framework

To develop a likelihood-based machinery to integrate univariate regression models into a joint analysis, a fully parametric multidimensional link model is inevitable. In this section, we first describe a general strategy of joining univariate regression models, and then give a brief review of Gaussian copula models.

2.1 *Joining Marginal Models*

GLMs (Nelder and Wedderburn, 1972) have been playing a central role in the regression analysis of one-dimensional nonnormal data. In a GLM, a univariate response y is assumed to follow an exponential dispersion family distribution (Jørgensen, 1987, 1997), denoted by $ED(\mu, \varphi)$, with mean μ and dispersion parameter φ. The density function of $ED(\mu, \varphi)$ is given by

$$g(y; \mu, \varphi) = c(y; \varphi) \exp[\{\theta y - \kappa(\theta)\}/\varphi], \quad y \in \mathcal{R}, \quad \theta \in \Theta, \quad (1)$$

where $\kappa(\cdot)$ is the cumulant generating function, Θ is an open interval, and φ varies in a subset of $(0, \infty)$. It is known that the mean and variance are, respectively, $\mu = \mathrm{E}(y) = \tau(\theta)$ and $\mathrm{var}(y) = \varphi\, v(\mu)$, where $v(\cdot)$ is the unit variance function, with $\tau(\cdot) = \dot{\kappa}(\cdot)$ and $v(\cdot) = \dot{\tau}\{\tau^{-1}(\cdot)\}$ being the respective first derivatives of $\kappa(\cdot)$ and $\tau(\cdot)$.

A GLM also postulates that the mean μ is related to p covariates $\mathbf{x} = (x_1, \ldots, x_p)^T$ by a known link function h,

$$h(\mu) = \eta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \quad (2)$$

where $\boldsymbol{\beta} = (\beta_0,\ \beta_1, \ldots, \beta_p)^T$ is a vector of regression coefficients. Statistical inference for $\boldsymbol{\beta}$ is one of the main tasks in the theory of GLMs.

To jointly analyze vector data by the GLM approach, multidimensional GLMs, or VGLMs, specify the conditional distribution of a vector response $\mathbf{y}$ given $\mathbf{x}$ as follows:

$$f(\mathbf{y} \mid \mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\varphi}, \Gamma) = \delta(\mathbf{y}, \eta_1, \ldots, \eta_m; \boldsymbol{\varphi}, \Gamma), \quad (3)$$

with, in general, the regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_m^T)^T$ and the linear predictors $\eta_j = \eta_j(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_j, j = 1, \ldots, m$. The parametric link model $\delta(\cdot; \boldsymbol{\varphi}, \Gamma)$ is parameterized by the vector of dispersion parameters $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_m)^T$ and the association matrix Γ. Here $\Gamma = (\gamma_{ij})$ characterizes the association among the components of $\mathbf{y}$. The regression coefficients $\boldsymbol{\beta}_j$'s are allowed to be different.

Examples of the VGLM (3) with the common regression parameter include the log-linear representation (Bishop, Fienberg, and Holland, 1975) or the Bahadur (1961) representation for correlated binary responses (see Zhao and Prentice, 1990; Fitzmaurice et al., 1993), and generalized linear mixed models (see Breslow and Clayton, 1993; McCulloch and Searle, 2001; Diggle et al., 2002). Examples of models with different $\boldsymbol{\beta}_j$'s include the bivariate logit model (see McCullagh and Nelder, 1989, section 6.5.6) and the bivariate probit model (Ashford and Sowden, 1970) for correlated binary responses, among others.

In this article, we consider a new class of parametric link models $\delta(\cdot)$ via the multivariate distributions generated by Gaussian copulas (Song, 2000). The resulting framework provides a unified and flexible approach to a joint regression analysis of correlated continuous, discrete, or mixed outcomes. Advantages of the copula link model and the resulting VGLMs include (i) the association coefficients in the copula model are not constrained by marginal means, (ii) the VGLMs are reproducible or marginally closed, and (iii) the regression coefficients have marginal interpretation.

2.2 *Multivariate ED Family Distributions*

We now give a brief review of the Gaussian copula and the resulting multivariate ED (MED) distributions (Song, 2000), which pertain to the specification of the link model $\delta(\cdot)$.

For component $j, j = 1, \ldots, m$, denote the marginal cumulative distribution function (CDF) of $\mathrm{ED}(\mu_j,\ \varphi_j)$ by $G_j(y_j; \mu_j, \varphi_j)$ or simply $G_j(y_j)$. Following Sklar (1959), a joint CDF with m ED margins can be constructed by the Gaussian copula in the form

$$F(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\varphi}, \Gamma) = C\left\{G_1(y_1; \mu_1, \varphi_1), \ldots, G_m(y_m; \mu_m, \varphi_m) \mid \Gamma\right\}, \quad (4)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T$ is the vector of m means, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_m)^T$ is the vector of m dispersion parameters, and $C(\cdot)$ is the m-variate Gaussian copula with the CDF given by

$$C(\mathbf{u} \,|\, \Gamma) = \Phi_m\big\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m) \,\big|\, \Gamma\big\},$$
$$\mathbf{u} = (u_1, \ldots, u_m)^T \in (0,1)^m. \tag{5}$$

Here Φ_m (or ϕ_m) and Φ (or ϕ) are the respective CDFs (or densities) of m-variate normal $N_m(0,\,\Gamma)$ with a correlation matrix Γ and the standard univariate normal $N(0,1)$ marginals. Note that all marginal parameters are brought into the F, and the parameters for association are inherited from the correlation matrix Γ of the multivariate normal. It is known that the Gaussian copula in equation (4) is a joint CDF of m uniform random variables on (0, 1) with association matrix $\Gamma = (\gamma_{jj'})_{m\times m}$ in which the diagonals $\gamma_{jj} = 1$ and off-diagonals $|\gamma_{jj'}| < 1$. Like the multivariate Gaussian distribution, an MED distribution is fully parameterized by the three sets of parameters, $\boldsymbol{\mu}$, $\boldsymbol{\varphi}$, and Γ, and it can accommodate both positive and negative associations.

Clearly, the multivariate normal distribution is a special case of the MED when all margins are univariate Gaussian. In this case $\boldsymbol{\mu}$ is the vector of mean parameters, $\boldsymbol{\varphi}$ is the vector of variance parameters, and Γ is the Pearson correlation matrix. With non-Gaussian margins, the (i,j)th element of Γ becomes a pairwise nonlinear association, namely, the van der Waerden coefficient (Klaassen and Wellner, 1997), defined by

$$\gamma_{ij} = \operatorname{corr}\left[\Phi^{-1}\{G_i(y_i)\}, \Phi^{-1}\{G_j(y_j)\}\right]. \tag{6}$$

When both marginal CDFs $G_t(\cdot)$, $t = i,\ j$ are continuous, γ_{ij} represents the linear correlation of two normal scores $\Phi^{-1}\{G_t(y_t)\}$, $t = i,\ j$. When y_i and y_j are discrete, the equation (6) still holds, but the interpretation would be different with different data types. For example, when y_t, $t = i,\ j$ are both binary, the resulting bivariate binary model will have the same joint probability mass function as that induced from the threshold latent variable model via dichotomization (Song, 2007, p. 133). This implies that the association parameter γ_{ij} can be interpreted as the *tetrachoric correlation* (Harris, 1988). See Song (2000) for more details in other distribution cases such as Poisson distribution.

In the following, we give density functions of MEDs with different marginal distributions. When all m margins are continuous, the joint density of an MED in equation (5) is given by

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\varphi}, \Gamma) = c\,\{G_1(y_1), \ldots, G_m(y_m) \,|\, \Gamma\} \prod_{i=1}^{m} g_i(y_i; \mu_i, \varphi_i), \tag{7}$$

where $c(\cdot)$ is the density of the copula $C(\cdot)$ in (4) given by

$$c(\mathbf{u} \,|\, \Gamma) = |\Gamma|^{-1/2} \exp\left\{\frac{1}{2}\mathbf{q}^T\left(I_m - \Gamma^{-1}\right)\mathbf{q}\right\},$$

with $\mathbf{q} = (q_1, \ldots, q_m)^T$ being a vector of normal scores $q_i = \Phi^{-1}(u_i)$, $i = 1, \ldots, m$, and I_m being the m-dimensional identity matrix. Obviously, $\Gamma = I_m$ implies the independence of the m components, similar to the multivariate normal.

Consequently, when the function δ required in the VGLM (3) is chosen to be the density f specified by equation (7), the VGLM yields a large class of multidimensional regression models for various correlated continuous data, including the vector normal linear model, the vector gamma GLM model, the vector inverse Gaussian GLM model, and the vector compound Poisson GLM model.

When all m margins are discrete, the joint probability function of a discrete MED distribution takes the form

$$\begin{aligned} f(\mathbf{y}) &= \mathrm{P}(Y_1 = y_1, \ldots, Y_m = y_m) \\ &= \sum_{j_1=1}^{2} \cdots \sum_{j_m=1}^{2} (-1)^{j_1+\cdots+j_m} C(u_{1,j_1}, \ldots, u_{m,j_m} \,|\, \Gamma), \end{aligned} \tag{8}$$

where $u_{j,1} = G_j(y_j-)$ and $u_{j,2} = G_j(y_j)$. Here $G_j(y_j-)$ is the left-hand limit of G_j at y_j.

Likewise, a large class of multidimensional regression models for correlated discrete data is specified under a unified framework, by taking this probability mass function f in equation (8) as the δ for the VGLM in equation (3). In this article, a special vector GLM from this class, the vector logistic model for correlated dichotomous data, will be studied in detail in Section 4.

When the m margins appear to be mixed outcomes, say, the first m_1 margins being continuous and the remaining $m_2 = m - m_1$ margins being discrete, the joint density function is given as follows. Let $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)^T$, with $\mathbf{u}_1 = (u_1, \ldots, u_{m_1})^T$ and $\mathbf{u}_2 = (u_{m_1+1}, \ldots, u_m)^T$. The same partition and notation are applied for vectors $\mathbf{x}$ and $\mathbf{q}$. Let

$$\begin{aligned} &C_1^{m_1}(\mathbf{u}_1, \mathbf{u}_2 \,|\, \Gamma) \\ &= \frac{\partial^{m_1}}{\partial u_1 \cdots \partial u_{m_1}} C(u_1, \ldots, u_m \,|\, \Gamma) \\ &= (2\pi)^{-\frac{m_2}{2}} |\Gamma|^{-\frac{1}{2}} \int_{-\infty}^{\Phi^{-1}(u_{m_1+1})} \\ &\quad \cdots \int_{-\infty}^{\Phi^{-1}(u_m)} \exp\left\{\frac{1}{2}\left(\mathbf{q}_1^T, \mathbf{x}_2^T\right)\Gamma^{-1}\left(\mathbf{q}_1^T, \mathbf{x}_2^T\right)^T - \frac{1}{2}\mathbf{q}_1^T\mathbf{q}_1\right\} d\mathbf{x}_2. \end{aligned}$$

Then, the joint density is given by

$$\begin{aligned} f(\mathbf{y}) = \prod_{j=1}^{m_1} g_j(y_j) \times \sum_{j_{m_1+1}=1}^{2} \cdots \sum_{j_m=1}^{2} (-1)^{j_{m_1+1}+\cdots+j_m} C_1^{m_1} \\ \times \left(G_1(y_1), \ldots, G_{m_1}(y_{m_1}), u_{m_1+1, j_{m_1+1}}, \ldots, u_{m,j_m} \,|\, \Gamma\right), \end{aligned} \tag{9}$$

where u_{t,j_t}'s are the same as defined in equation (8). Section 5 presents an application of this joint modeling in the analysis of correlated mixed outcomes from a study of burn injuries.

3. Simultaneous Maximum Likelihood Inference

Given covariates $(X_1, \ldots, X_n)$, suppose the responses $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ follow an m-variate MED distribution,

$$\mathbf{y}_i \,|\, X_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im}) \sim \mathrm{MED}_m(\boldsymbol{\mu}_i, \boldsymbol{\varphi}_i, \Gamma), \quad i = 1, \ldots, n,$$

where response vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^T$ has mean $\boldsymbol{\mu}_i = (\mu_{i1}(\mathbf{x}_{i1}), \ldots, \mu_{im}(\mathbf{x}_{im}))^T$ and dispersion $\boldsymbol{\varphi}_i = (\varphi_{i1}, \ldots, \varphi_{im})^T$. Here $X_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im})$ is a $p \times m$ matrix of covariates. Moreover, the mean μ_{ij} follows a marginal GLM, $h_j(\mu_{ij}) = \eta_j(\mathbf{x}_{ij})$ with $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta}_j$ and link function $h_j, j = 1, \ldots, m$. When modeling the dispersion parameter is also of interest, similar to Smyth (1989), one may specify an additional GLM to characterize the varying dispersion as a function of covariates. The primary objective is to establish simultaneous maximum likelihood inferences for all the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\varphi}, \Gamma)$.

In many cases, the general model above may become more specific. For example, typically in longitudinal or clustered data analysis, a VGLM (3) takes a common regression parameter vector $\boldsymbol{\beta}$ with a common link function. In addition, the association matrix Γ may be further parameterized by a parameter vector $\boldsymbol{\alpha}$, denoted by $\Gamma(\boldsymbol{\alpha})$, such as exchangeable, AR(1) or 1-dependence. In this case, we have $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\varphi}, \boldsymbol{\alpha})$.

Let the log-likelihood function of the given model be

$$\ell(\boldsymbol{\theta}; Y, X) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}; \mathbf{y}_i, X_i). \tag{10}$$

Then, the MLE of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \ell(\boldsymbol{\theta}; Y, X).$$

To find the MLE $\hat{\boldsymbol{\theta}}$ numerically, we implement a Gauss–Newton-type algorithm that allows us to search for the MLE with no need of second derivatives of the log-likelihood function (Ruppert, 2005). As seen in Section 1 of Web Supplementary Materials, explicitly deriving the second-order derivatives of the log likelihood may appear to be very difficult. This issue could be more troublesome in the case of discrete data.

It follows from the standard ML theory that under certain regularity conditions, the MLE $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normal. When the analytic expressions of second-order derivatives of the log likelihood are unavailable, following Ruppert (2005), we estimate the observed Fisher Information using the following sandwich form:

$$\hat{\mathcal{I}} = \mathbf{A}_n^{-1}(\hat{\boldsymbol{\theta}})\mathbf{B}_n(\hat{\boldsymbol{\theta}})\mathbf{A}_n^{-1}(\hat{\boldsymbol{\theta}}), \tag{11}$$

where $\mathbf{A}_n(\boldsymbol{\theta})$ is the numerical Hessian matrix and $\mathbf{B}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_i(\boldsymbol{\theta}; \mathbf{y}_i, X_i)\dot{\ell}_i(\boldsymbol{\theta}; \mathbf{y}_i, X_i)^T$.

The key step in the above Gauss–Newton algorithm adopted in the search for the MLE is to take step-halving that guarantees likelihood increases progressively over iterations. Precisely, the $(k+1)$th iteration updates the parameter $\boldsymbol{\theta}$ by

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} + \epsilon\left\{\mathbf{B}_n(\boldsymbol{\theta}^{k})\right\}^{-1}\dot{\ell}(\boldsymbol{\theta}^{k}),$$

where ϵ is the step-halving term that is chosen as follows: starting at 1, it halves each time until $\ell(\boldsymbol{\theta}^{k+1}) > \ell(\boldsymbol{\theta}^{k})$ holds in one iteration. Finally, the algorithm stops when the increase in the likelihood is no longer possible or the difference between two consecutive updates is smaller than a prespecified precision level.

Among several optimization algorithms that we have tried, such as Newton–Raphson, downhill simplex, and a quasi-Newton based on numerical derivatives, the Gauss–Newton-type algorithm appears to provide the best trade-off between computational efficiency and analytic complexity, as well as the best numerical stability.

To make the presentation of this article less technical, a Web supplementary document has been created to list all general results regarding the log-likelihood functions and their scores corresponding to various VGLMs for continuous, discrete, and mixed data types.

4. VGLMs for Trivariate Discrete Data

The moment-based EE method, such as GEEs, has been widely used in practice to conduct a joint analysis of correlated data that follow marginally the same type of univariate distribution. This section concerns trivariate VGLMs for discrete data. Then we compare the ARE with the popular GEE method. Note that in the case of correlated binary data, the corresponding GEEs estimation is coincident with the MLE derived from the log-linear model representation (Diggle et al., 2002, section 8.2). Under the two different link models for correlated binary data, it is of interest to examine which method, the Gaussian copula or the log-linear model representation, leads to better estimation efficiency, if the marginal logistic models are specified the same.

4.1 *Trivariate VGLMs*

For simplicity, we consider the exchangeable correlation structure, namely, all off-diagonal elements of the association matrix Γ equal to a constant α, $|\alpha| < 1$. Refer to the supplementary document for other types of association structures.

The trivariate probability mass function f is obtained from equation (8) as

$$\begin{aligned} f(\mathbf{y}_i; \boldsymbol{\theta}) &= P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, Y_{i3} = y_{i3}) \\ &= \sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\sum_{j_3=1}^{2} (-1)^{j_1+j_2+j_3} C(u_{i,1,j_1}, u_{i,2,j_2}, u_{i,3,j_3} \,|\, \alpha), \end{aligned}$$

where $C(u_{i,1,j_1}, u_{i,2,j_2}, u_{i,3,j_3} \,|\, \alpha) = \Phi_3\{\Phi^{-1}(u_{i,1,j_1}), \Phi^{-1}(u_{i,2,j_2}), \Phi^{-1}(u_{i,3,j_3}) \,|\, \alpha\}$ with $u_{i,j,1} = G_i(y_{ij})$ and $u_{i,j,2} = G_i(y_{ij}-)$, $j = 1, 2, 3$. $\Phi_m(\cdot \,|\, \alpha)$ (or $\phi_m(\cdot \,|\, \alpha)$) denotes the CDF (or the density) of m-dimensional normal distribution with the standard normal margins and the exchangeable correlation coefficient α. The parameter vector is then $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$ as all dispersion parameters $\varphi_{ij} = 1$.

Let $\mathbf{u}_{i,j_1,j_2,j_3} = (u_{i,1,j_1}, u_{i,2,j_2}, u_{i,3,j_3})$, and let $\dot{f}_k(\cdot)$ be the first-order derivative of density f with respect to θ_k. Then, the scores are

$$\begin{aligned} \dot{\ell}_{\theta_k}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \dot{f}_k(\mathbf{y}_i; \boldsymbol{\theta})/f(\mathbf{y}_i; \boldsymbol{\theta}) \\ &= \sum_{i=1}^{n}\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\sum_{j_3=1}^{2} \left\{(-1)^{j_1+j_2+j_3}\dot{C}_{\theta_k}(\mathbf{u}_{i,j_1,j_2,j_3} \,|\, \alpha)\right\} \\ &\quad /f(\mathbf{y}_i; \boldsymbol{\theta}). \end{aligned}$$

Moreover, by the chain rule, the scores with respect to β_j are given by

$$\frac{\partial C(\mathbf{u}_{i,j_1,j_2,j_3} \,|\, \alpha)}{\partial \beta_j} = \sum_{t=1}^{3} \frac{\partial C(\mathbf{u}_{i,j_1,j_2,j_3} \,|\, \alpha)}{\partial u_{i,t,j_t}} \frac{\partial u_{i,t,j_t}}{\partial \beta_j},$$

where the first factor on the right-hand side takes the following forms:

$$\frac{\partial C(\mathbf{u}_{i,j_1,j_2,j_3}\,|\,\alpha)}{\partial u_{i,1,j_1}} = \Phi_2\left\{\Delta_\alpha(u_{i,2,j_2},u_{i,1,j_1}),\Delta_\alpha(u_{i,3,j_3},u_{i,1,j_1})\,|\,\gamma\right\},$$

$$\frac{\partial C(\mathbf{u}_{i,j_1,j_2,j_3}\,|\,\alpha)}{\partial u_{i,2,j_2}} = \Phi_2\left\{\Delta_\alpha(u_{i,1,j_1},u_{i,2,j_2}),\Delta_\alpha(u_{i,3,j_3},u_{i,2,j_2})\,|\,\gamma\right\},$$

$$\frac{\partial C(\mathbf{u}_{i,j_1,j_2,j_3}\,|\,\alpha)}{\partial u_{i,3,j_3}} = \Phi_2\left\{\Delta_\alpha(u_{i,1,j_1},u_{i,3,j_3}),\Delta_\alpha(u_{i,2,j_2},u_{i,3,j_3})\,|\,\gamma\right\},$$

with $\Delta_\alpha(u_{i,t,j_t},u_{i,s,j_s}) = \frac{\Phi^{-1}(u_{i,t,j_t})-\alpha\Phi^{-1}(u_{i,s,j_s})}{\sqrt{1-\alpha^2}}$ and $\gamma = \frac{\alpha}{1+\alpha}$. On the other hand, the second factor is given by

$$\frac{\partial u_{i,t,j_t}}{\partial\beta_j} = \frac{\partial u_{i,t,j_t}}{\partial\mu_{it}}\frac{x_{itj}}{\dot{h}_t(\mu_{it})}.$$

Note that derivatives $\partial u_{i,j,j_t}/\partial\mu_{it}$ can have closed form expressions when certain marginal distributions are assumed. For example, the Bernoulli margin for binary data leads to

$$\frac{\partial u_{i,t,1}}{\partial\mu_{it}} = -1[y_{it}=0], \frac{\partial u_{i,t,2}}{\partial\mu_{it}} = -1[y_{it}=1],$$

where $1[A]$ denotes the indicator function on set A, whereas the Poisson margin for count data gives

$$\frac{\partial u_{i,t,1}}{\partial\mu_{it}} = G_{it}(y_{it}-1) - G_{it}(y_{it}),$$

$$\frac{\partial u_{i,t,2}}{\partial\mu_{it}} = G_{it}(y_{it}-2) - G_{it}(y_{it}-1),$$

where $G_{it}(\cdot)$ is the Poisson CDF with mean μ_{it}.

Similarly, for the association parameter α,

$$\frac{C(\mathbf{u}_{i,j_1,j_2,j_3}\,|\,\alpha)}{\partial\alpha} = \int_{-\infty}^{\Phi^{-1}(u_{i,1,j_1})}\int_{-\infty}^{\Phi^{-1}(u_{i,2,j_2})}\int_{-\infty}^{\Phi^{-1}(u_{i,3,j_3})} \times\frac{\partial}{\partial\alpha}\left\{\ln\phi_3(z_1,z_2,z_3\,|\,\alpha)\right\} \times\phi_3(z_1,z_2,z_3\,|\,\alpha)\,dz_1\,dz_2\,dz_3, \qquad (12)$$

with

$$\frac{\partial}{\partial\alpha}\left\{\ln\phi_3(z_1,z_2,z_3\,|\,\alpha)\right\} = -\frac{1}{2(1-\alpha)}\left[\frac{z_1^2+z_2^2+z_3^2}{1-\alpha} - \frac{1+2\alpha^2}{(1-\alpha)(1+2\alpha)^2}(z_1+z_2+z_3)^2 - \frac{6\alpha}{1+2\alpha}\right].$$

The integral in (12) will be evaluated using the Gaussian–Hermite quadrature method (Lange, 1998, chapter 16).

4.2 *Comparison of Asymptotic Efficiency*

Now we present a comparison of the ARE between the Gaussian copula link model and the moment-based GEEs link method for trivariate binary data. We focus only on the regression parameters $\boldsymbol{\beta}$, because the association parameter α is usually treated as a nuisance parameter in the GEEs and is not really comparable between the two classes of link models.

According to Diggle et al. (2002), the sandwich estimator of the asymptotic covariance matrix of the GEE estimator $\hat{\boldsymbol{\beta}}_{gee}$ is $\text{var}_{gee} = \mathbf{I}_0^{-1}(\boldsymbol{\theta})\mathbf{I}_1(\boldsymbol{\theta})\mathbf{I}_0^{-1}(\boldsymbol{\theta})$, where

$$\mathbf{I}_0(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial\boldsymbol{\mu}_i^T}{\partial\boldsymbol{\beta}}\mathbf{V}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}}, \quad \text{and}$$

$$\mathbf{I}_1(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial\boldsymbol{\mu}_i^T}{\partial\boldsymbol{\beta}}\mathbf{V}_i^{-1}\text{cov}(\mathbf{y}_i)\mathbf{V}_i^{-1}\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\beta}}$$

with $\text{cov}(\mathbf{y}_i)$ being the covariance matrix of $\mathbf{y}_i$, usually estimated by $(\mathbf{y}_i-\boldsymbol{\mu}_i)(\mathbf{y}_i-\boldsymbol{\mu}_i)^T$. The working covariance matrix $\mathbf{V}_i = \varphi\mathbf{A}_i^{\frac{1}{2}}R(\alpha)\mathbf{A}_i^{\frac{1}{2}}$, where $\mathbf{A}_i$ is an $m\times m$ diagonal matrix $\text{diag}\{v(\mu_{ij}),\ j=1,\dots,m\}$, and $R(\alpha)$ is a working correlation matrix. The variance function is $v(\mu)=\mu(1-\mu)$ for the binomial distribution. Under the exchangeable correlation structure, the correlation parameter α can be estimated by

$$\hat{\alpha} = \frac{1}{3n}\sum_{i=1}^n(e_{i1}e_{i2}+e_{i1}e_{i3}+e_{i2}e_{i3}), \qquad (13)$$

where $e_{ij} = (y_{ij}-\mu_{ij})/\sqrt{v(\mu_{ij})}$ are the Pearson residuals. Let var_{vglm} be the asymptotic covariance of the ML estimator $\boldsymbol{\beta}_{vglm}$ from the VGLM. The ARE takes the form

$$\text{ARE}(\boldsymbol{\beta}) = \text{diag}\{\text{var}_{vglm}\}[\text{diag}\{\text{var}_{gee}\}]^{-1}. \qquad (14)$$

The comparison is based on a hypothetical clinical trial in which a binary response is repeatedly measured over three time periods. Following Fitzmaurice et al. (1993), we assume at each trial period that placebo ($x_t=0$) or an active drug ($x_t=1$) is randomly assigned among the subjects, and all the eight possible covariate configurations have equal probability of occurrence. A logistic model for the marginal expectation is specified as

$$\text{logit}(\mu_{it}) = \beta_0+\beta_1x_{it}+\beta_2(t-1), \quad t=1,2,3, \qquad (15)$$

where $\beta_0=0$, and $\beta_1=\beta_2=0.5$. The exchangeable correlation structure arises from the design of this three-period crossover trial. Hence it seems reasonable to assume that the logistic VGLM model describes adequately the mechanism of data generation from this trial. Note that in such a setting, we are able to analytically derive the closed form expressions of var_{gee} and var_{vglm}, so the calculation of the ARE does not require simulation data. Related details concerning the derivations of these closed form expressions are listed in Section 2 of the Web Supplementary Materials.

Figure 1 displays the ARE for the estimator of treatment effect β_1 as a function of the association parameter $\alpha\in[0,1)$, with the exchangeable structure for the VGLM, where $\alpha=0$ corresponds to the case of independence. We invoked a Monte Carlo method to convert the value of α in the VGLM into the value of correlation in the GEE via equation (13). For the GEEs, both the exchangeable and independence working covariance matrices are considered in the comparison, respectively. When the exchangeable covariance matrix was used,

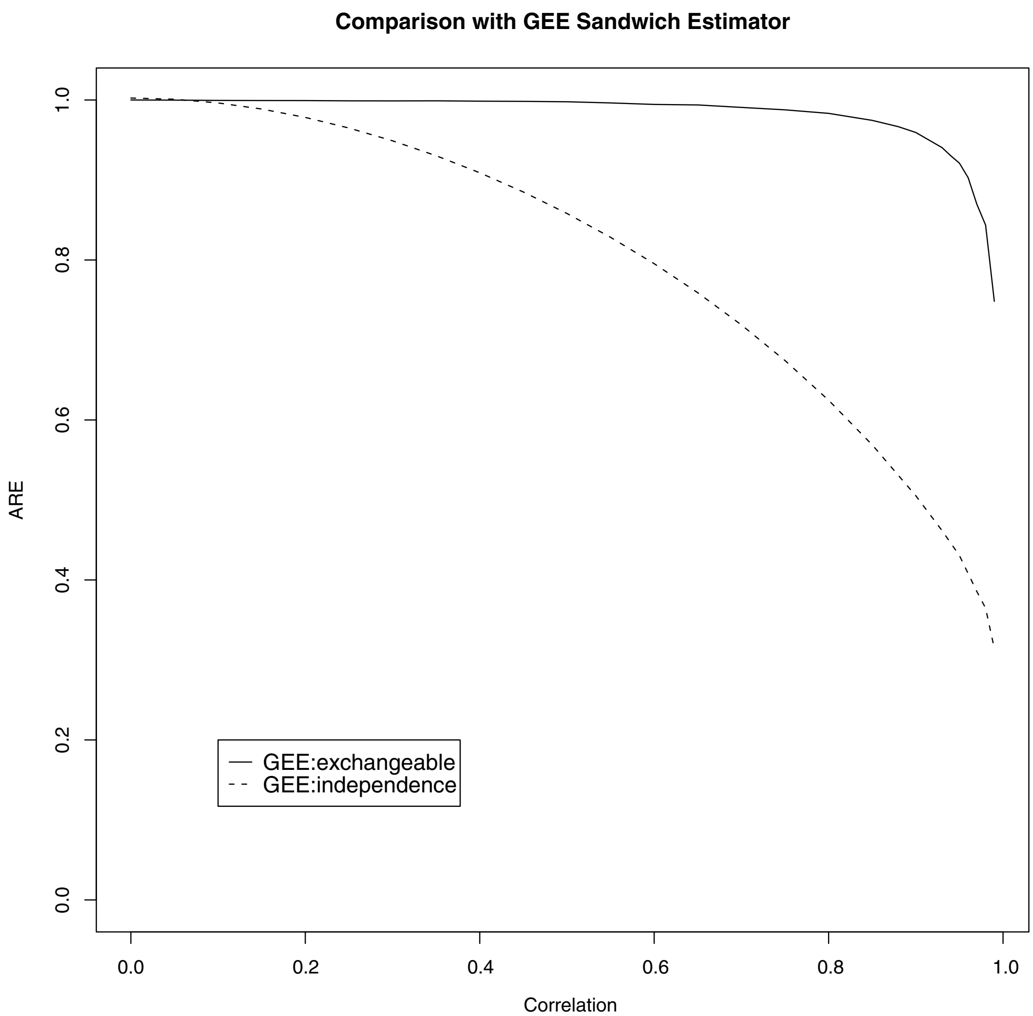

Figure 1. Asymptotic efficiencies of the VGLM estimator of the slope parameter β_1 relative to the estimators, respectively, from the GEE under the trivariate logistic model.

i.e., both the first and second moments of the outcome variable are correctly specified, the GEE is highly efficient, so we observe that the VGLM performs only marginally better than the GEE, and the efficacy gain, although small, occurs only at high correlations. This is in agreement with the results of Liang and Zeger (1986). However, when the independence working covariance matrix is employed in the GEE, which is somewhat similar to running univariate logistic regression separately across time periods, the estimator from the VGLM is clearly more efficient than the corresponding GEE estimator.

5. Data Examples

We present two data examples to further illustrate the proposed VGLMs. One is to join two GLMs for mixed outcomes of binary and normal responses and the other is to join three cross-sectional logit regression models for a three-visit longitudinal trial data.

5.1 *Joint Analysis of Burn Injury Data*

To demonstrate the flexibility of the proposed VGLM, we now apply a VGLM to analyze the burn injury data introduced in Section 1. There are two response variables, of which the severity of burn injury by y_1 is continuous and the disposition of death y_2 is binary. It is of interest to investigate how age (x) affects the severity of burn injury and the probability of death. Therefore, we proposed two marginal mean models: one is a normal linear model for the expected log-burn area, $\mu_{i1} = \beta_{01} + \beta_{11}x_i$, and the other is a logistic model for the probability of death from burn injury, $\text{logit}(\mu_{i2}) = \beta_{02} + \beta_{12}x_i$.

Suppressing the subject index, from equation (9) we can write the joint density of $\mathbf{y} = (y_1, y_2)$ as follows:

$$f(y_1, y_2) = \begin{cases} \phi(y_1; \mu_1, \varphi_1)\{1 - C_1^*(\mu_2, z_1 \,|\, \alpha)\}, & \text{if } y_2 = 0, \\ \phi(y_1; \mu_1, \varphi_1) C_1^*(\mu_2, z_1 \,|\, \alpha), & \text{if } y_2 = 1, \end{cases} \tag{16}$$

where $\phi(\cdot;\, \mu_1,\, \varphi_1)$ is the density of $N(\mu_1, \varphi_1)$, $z_1 = (y_1 - \mu_1)/\sqrt{\varphi_1}$, and $C_1^*(a, b \,|\, \alpha) = \Phi(\frac{\Phi^{-1}(a) - \alpha b}{\sqrt{1-\alpha^2}})$. An advantage of this joint copula modeling is that it avoids the artificial bimodal mixture of two normal margins, which is usually incurred by the conditional modeling approach such as Fitzmaurice and Laird's (1995).

For the burn injury data $\{\mathbf{y}_i, (\mathbf{x}_{i1}, \mathbf{x}_{i2})\}$, $i = 1, \ldots, n$, the log likelihood for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \varphi_1, \alpha)$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{i\in I_0} \ln\left[\phi(y_{i1};\mu_{i1},\varphi_1)\left\{1 - C_1^*(\mu_{i2}, z_{i1}\,|\,\alpha)\right\}\right] \\ &\quad + \sum_{i\in \bar{I}_0} \ln\left[\phi(y_{i1};\mu_{i1},\varphi_1)C_1^*(\mu_{i2}, z_{i1}\,|\,\alpha)\right] \\ &= \sum_{i=1}^{n} \ln\phi(y_{i1};\mu_{i1},\varphi_1) + \sum_{i\in I_0}\ln\left\{1 - C_1^*(\mu_{i2}, z_{i1}\,|\,\alpha)\right\} \\ &\quad + \sum_{i\in \bar{I}_0} \ln C_1^*(\mu_{i2}, z_{i1}\,|\,\alpha),\end{aligned}$$

where $I_0 = \{i : y_{i2} = 0\}$ and $\bar{I}_0 = \{i : y_{i2} = 1\}$ are subsets of indices for survived and dead subjects, respectively.

Both joint model and individual univariate models were applied to fit the data, and the results are summarized in Table 1. The estimated association parameter α by the VGLM was 0.80, which indicates a strong association between these two responses. Overall, the point estimates obtained by the VGLM and the separate univariate models are very similar to each other. However, the VGLM appears to be much more efficient than the separate univariate analysis. More specifically, the effect of age on the burn severity is found to be statistically significant by the VGLM but marginally significant by the univariate linear regression model. So, ignoring the strong association between the response variables will greatly reduce the power of the statistical inference. As discussed previously, one important advantage of our approach over the EE-based approach is that we can obtain the conditional distribution of death given burn severity based on the VGLM, as shown in Figure 2. When the logarithm of burn area is below 7, the probability of death is weakly associated with age; however, when the logarithm of the burn area is between 7 and 10, the older age has a significantly higher probability of death than the younger age given the same burn area. In conclusion, the joint modeling approach in this example provides substantial improvements in efficiency over the univariate analysis and hence is preferred.

5.2 *Joint Analysis of Longitudinal Trial Data*

Data arising from longitudinal clinical trials can be analyzed by the proposed VGLMs. This example concerns a data set collected from 80 subjects with multiple sclerosis in a randomized, placebo-controlled trial studying the efficacy of fampridine, a compound to enhance nerve conduction (Davis, 2002).

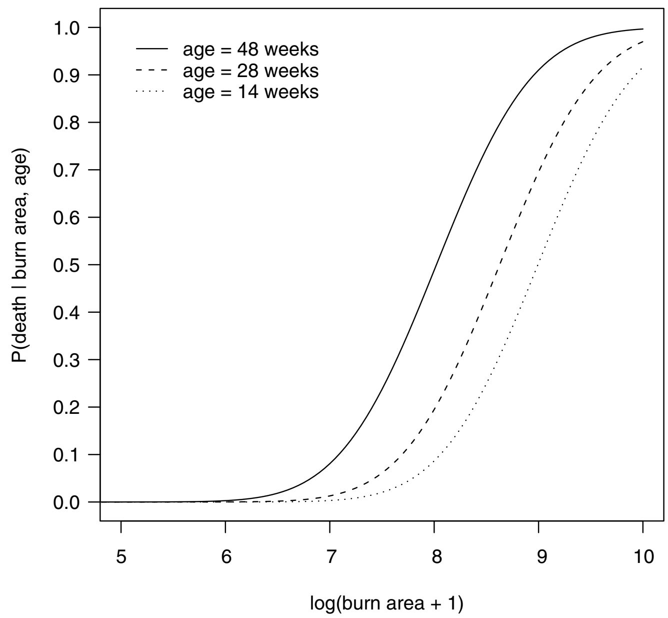

Figure 2. Conditional distribution of death given the burn area under various ages.

Prior to the initiation of treatment, the time that each patient was required to walk a specified distance was recorded as the baseline reference. During the treatment period, the same process was repeated for each patient at weeks 2, 4, and 6, and the outcome variable was the change from baseline in ambulation time. For individual i at visit j, let y_{ij} be the response variable with 1 indicating improvement from baseline, and 0 otherwise. The marginal expectations were specified as follows:

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{ij2}, \quad i = 1, \ldots, 80, \quad j = 1, 2, 3,$$

where $\mu_{ij} = P(y_{ij} = 1)$, and two covariates, treatment (x_1) and visit (x_2), are defined as $x_{i1} = 1$ for fampridine and 0 for placebo, and $x_{ij2} = j$ for visit j.

We applied the proposed VGLM to fit the data. The results are reported in Table 2. The estimate of the association parameter by the VGLM was 0.71 under the exchangeable correlation matrix. Clearly, our analysis suggested that fampridine was effective in improving the multiple sclerosis patients' nerve conduction. That is, the odds of improvement for the patients taking fampridine was 2.66 times as high as

Table 1

The estimates and standard errors obtained from the analysis of the burn injury data, where both joint model and separate univariate models are applied

		VGLM			Univariate models		
Model	β	$\hat{\beta}$	S.E.	Z	$\hat{\beta}$	S.E.	Z
Linear	Intercept	6.6980	0.0479	139.73	6.7118	0.0690	97.24
(log(burn area+1))	Age	0.0039	0.0012	3.16	0.0035	0.0018	1.97
Logit	Intercept	−4.0521	0.1658	−24.44	−3.6891	0.2342	−17.78
(death)	Age	0.0527	0.0028	19.13	0.0509	0.0046	11.07

Table 2
Estimated regression coefficients, standard errors, and Z-statistics from the VGLM fit

	VGLM				GEE			
Variable	$\hat{\beta}$	S.E.	Z	p-value	$\hat{\beta}$	S.E.	Z	p-value
Intercept	0.4456	0.3619	1.23	0.2183	0.4601	0.4021	1.14	0.2525
Treatment (x_1)	0.9803	0.4376	2.24	0.0251	0.9451	0.4447	2.13	0.0336
Visit (x_2)	−0.0250	0.1232	−0.20	0.8390	−0.0347	0.1430	−0.24	0.8084

the patients taking placebo. However, the odds of improvement were not related to a specific visit. Given the level of the association around 0.7, according to the simulation study in Section 4.2, it is not surprising to see that the analysis based on the GEE yielded similar results. In this particular example, VGLM did not show an appreciable efficacy gain compared with the GEE. This may be due to the fact that the exchangeable correlation matrix well approximates the true underlying covariance structure.

One advantage of the VGLM approach is the availability of likelihood, which enables us to conduct statistical inference and model selection easily. To illustrate, we output the $-2\times$log likelihood in the analysis, namely 213.7212, 213.7894, and 218.7406 for the model with both covariates, the model with only visit covariate, and the model with only treatment covariate, respectively. Then, the standard likelihood ratio test leads to p-values of 0.0251 for treatment and of 0.794 for visit under the chi-square distribution with one degree of freedom. These p-values are comparable to those reported in Table 2, which were found via the traditional z-test statistics. In a similar fashion, one can calculate the Akaike information criterion (AIC) or Bayesian information criterion (BIC) to assess the goodness of fit for model selection.

6. Concluding Remarks

This article presents a class of multidimensional GLMs that can accommodate a variety of correlated discrete, continuous, and mixed outcomes. We developed a simultaneous maximum likelihood inference that was implemented by a Gauss–Newton-type algorithm. Our focus in this article is on the joint analysis of correlated discrete and mixed outcomes through the Gaussian copula link model, because this approach has not been studied thoroughly in the literature. An advantage of the proposed theory is that all different types of data can be treated under one unified MLE framework. Another advantage is that the proposed joint regression analysis leads to a more powerful inference than separate univariate regression analysis.

Because few probability model-based methods are available for practitioners to analyze correlated discrete and mixed data, the models proposed in the present article provide a useful arsenal to conduct maximum likelihood statistical inference in a joint model that integrates univariate GLMs via the Gaussian copula. With available likelihoods, it is straightforward to define likelihood ratio statistics in the context of hypothesis testing. In contrast, for the GEE approach testing hypotheses based on likelihood ratios for discrete data becomes a difficult issue because of the lack of a likelihood function. In addition, model selection along the lines of the AIC or BIC can be readily established in the VGLM framework.

The proposed VGLMs for correlated discrete data need to evaluate multivariate normal CDFs. `R` software provides a package `mvtnorm` based on Genz's algorithm (Genz, 1992) that can compute the normal CDF of 100 dimensions or less. Although deriving joint probability mass functions may become analytically tedious for high dimensions, one may invoke software such as `MAPLE` to obtain derivatives of the log likelihood. Nevertheless, for discrete outcomes with moderate to large dimensions, one may overcome the tediousness by invoking the method of composite likelihood (Lindsay, 1988; Varin and Vidoni, 2005), which comprise valid low-dimensional likelihood objects. Such a dimensional reduction procedure may be initiated from, say, pairwise analysis that infers all pairwise associations, which then gives useful clues to break down the likelihood function according to the sparsity of the association matrix. Recently, a new idea, called the continued extension argument proposed by Denuit and Lambert (2005), shed light on overcoming difficulties in the application of copulas to discrete marginals. This continued extension idea is to transform a discrete response variable into a continuous one by adding a uniformly distributed continuous random variable. As a result, estimation and inference can be made in the framework of copula models with continuous marginals, where numerical calculations appear much simpler. Both directions are worth serious further exploration.

As always in the application of parametric models for data analysis, model assumption diagnostics are necessary before the results are employed to draw final conclusions. Checking assumptions on the marginal model specifications can be done similarly as in the classical GLM theory. Other copulas than the Gaussian copula may also be possible to establish similar frameworks to that presented in this article. Consequently, there is an issue of copula selection, which needs some future research.

7. Supplementary Materials

The Web Supplementary Materials referenced in Sections 3 and 4.2 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The first author's research was supported by a Canadian NSERC Discovery Grant. The authors are grateful to the associate editor and an anonymous referee for their constructive comments.

References

Ashford, J. R. and Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics* **26,** 535–546.

Bahadur, R. R. (1961). A representation of the joint distribution of responses on n dichotomous items. In *Studies in Item Analysis and Prediction, Stanford Mathematical Studies in the Social Sciences VI*, H. Solomon (ed), 158–168. Stanford, CA: Stanford University Press.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

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

Davis, C. S. (2002). *Methods for the Analysis of Repeated Measurements*. New York: Springer.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93,** 40–57.

Diggle, P. J., Liang, K.-Y., Heggerty, P., and Zeger, S. L. (2002). *The Analysis of Longitudinal Data*, 2nd edition. Oxford: Oxford University Press.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

Fitzmaurice, G. M. and Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* **90,** 845–852.

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses. *Statistical Science* **8,** 284–309.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1,** 141–149.

Harris, B. (1988). Tetrachoric correlation coefficient. In *Encyclopedia of Statistical Sciences,* Volume 9, L. Kotz and N. L. Johnson (eds), 223–225. New York: Wiley.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.

Jones, B. and Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials*. London: Chapman and Hall.

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B* **49,** 127–162.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman and Hall.

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

Lang, K. (1998). *Numerical Analysis for Statisticians*. New York: Springer.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lindsay, B. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, N. U. Prabhu (ed), 221–239. Providence, RI: American Mathematical Society.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. New York: Wiley.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135,** 370–384.

Ruppert, D. (2005). Discussion of "Maximization by parts in likelihood inference" by Song, Fan and Kalbfleisch. *Journal of the American Statistical Association* **100,** 1161–1163.

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics* **27,** 305–320.

Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. New York: Springer.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de L'Institut de Statistiques de l'Universite de Paris* **8,** 229–231.

Smyth, G. K. (1989). Generalised linear models with varying dispersion. *Journal of the Royal Statistical Society, Series B* **51,** 47–60.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92,** 519–528.

Zhao, P. L. and Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic model. *Biometrika* **77,** 642–648.

Received February 2007. *Revised January* 2008.
Accepted February 2008.