

READER REACTION

Joint Regression Analysis for Discrete Longitudinal Data

L. Madsen* and Y. Fang**

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

*email: madsenl@onid.orst.edu

**email: fangya@science.oregonstate.edu

SUMMARY. We introduce an approximation to the Gaussian copula likelihood of Song, Li, and Yuan (2009, *Biometrics* **65**, 60–68) used to estimate regression parameters from correlated discrete or mixed bivariate or trivariate outcomes. Our approximation allows estimation of parameters from response vectors of length much larger than three, and is asymptotically equivalent to the Gaussian copula likelihood. We estimate regression parameters from the toenail infection data of De Backer et al. (1996, *British Journal of Dermatology* **134**, 16–17), which consist of binary response vectors of length seven or less from 294 subjects. Although maximizing the Gaussian copula likelihood yields estimators that are asymptotically more efficient than generalized estimating equation (GEE) estimators, our simulation study illustrates that for finite samples, GEE estimators can actually be as much as 20% more efficient.

KEY WORDS: Continuous extension; Correlated data; Dependent data; Discrete data; Gaussian copula.

1. Introduction

Song, Li, and Yuan (2009) presented a new joint regression analysis for multivariate outcomes based on Gaussian copulas and illustrate the technique in two examples. In the first example, one element of the bivariate response is continuous and the other is discrete. The response in the second example is a triple of dichotomous outcomes. For discrete multivariate outcomes the model is only tractable if the response vector is no more than four- or five-variate. In this article, we introduce an approximation to the model that allows larger-dimensional response vectors. We illustrate our model by analyzing the toenail infection data of De Backer et al. (1996) in which each subject was observed on up to seven occasions.

Song et al. (2009) compare the asymptotic efficiency of regression parameters estimators between their model and a standard generalized estimating equation (GEE) analysis. For this article, we conducted a simulation study to compare the observed efficiency between our estimator and the GEE estimator. The study shows that for finite samples, the GEE estimator is up to 20% more efficient than that obtained from the Gaussian copula model.

2. The Gaussian Copula with Continuous Extension

Song et al. (2009) arrive at the Gaussian copula as a convenient special case of vector generalized linear models (VGLMs). Here, we present a brief discussion of copula models and their utility in constructing multivariate distribution functions. For a more thorough introduction, see Nelsen (2006).

A *bivariate copula* is a joint distribution function $C(u, v)$ with uniform marginals. Sklar's theorem (Sklar, 1959; Nelsen, 2006, Section 2.3) establishes the correspondence between copulas and joint distribution functions. The theorem states that given any random variables Y_1 and Y_2 with marginal distributions F_1 and F_2 and joint distribution function H , there exists a copula C with $H(y_1, y_2) = C\{F_1(y_1), F_2(y_2)\}$ for any y_1, y_2 in \mathbb{R}^2 , and that conversely, given any copula C and univariate distribution functions F_1 and F_2 , $H(y_1, y_2) = C\{F_1(y_1), F_2(y_2)\}$ defines a joint distribution function with marginals F_1 and F_2 . If the F_i are continuous, C is unique.

The bivariate Gaussian copula model is $C(y_1, y_2; \alpha) = \Phi_\alpha[\Phi^{-1}\{F_1(y_1)\}, \Phi^{-1}\{F_2(y_2)\}]$, where Φ is the standard normal cumulative distribution function (cdf) and Φ_α is the bivariate normal cdf with correlation α . The parameter α is not the correlation between Y_1 and Y_2 but does determine the degree of dependence. Song (2007, p. 130) calls α the “normal scoring” between nonnormal Y_1 and Y_2 and discusses the connection between α and two other measures of monotone association, Spearman's ρ and Kendall's τ . Joe (1997, pp. 140–141) notes that $\alpha = 1$ corresponds to maximal positive dependence and $\alpha = -1$ corresponds to maximal negative dependence.

Nelsen (2006, pp. 116–117) catalogs a number of bivariate copula models but later notes that constructing multivariate copulas is “difficult.” The Gaussian copula is an exception, and has the additional advantage of allowing a flexible parametric dependence structure. Suppose Y_1, \dots, Y_n have marginal distributions F_i . Let Σ be a symmetric, non-negative definite matrix with diagonal entries equal to 1.

A joint distribution function of Y_1, \dots, Y_n is $C(\mathbf{y}; \Sigma) = \Phi_{\Sigma}[\Phi^{-1}\{F_1(y_1)\}, \dots, \Phi^{-1}\{F_n(y_n)\}]$, where Φ_{Σ} is the multivariate normal cdf with covariance matrix Σ . If the Y_i have marginal density functions f_i , the joint density function is obtained by differentiating $C(\mathbf{y}; \Sigma)$ with respect to \mathbf{y} :

$$c(\mathbf{y}; \Sigma) = |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\Sigma^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \cdot \prod_{i=1}^n f_i(y_i), \quad (1)$$

where $\mathbf{z} = [\Phi^{-1}\{F_1(y_1)\}, \dots, \Phi^{-1}\{F_n(y_n)\}]'$ and \mathbf{I}_n denotes the $n \times n$ identity matrix.

If the Y_i are discrete, C is uniquely determined only on the range of $F_1 \times \dots \times F_n$, and the joint probability mass function of Y_1, \dots, Y_n is

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \sum_{j_1=1}^2 \dots \sum_{j_n=1}^2 (-1)^{j_1+\dots+j_n} \times \Phi_{\Sigma}\{\Phi^{-1}(u_{1j_1}), \dots, \Phi^{-1}(u_{nj_n})\}, \quad (2)$$

where $u_{i1} = F_i(y_i)$ and $u_{i2} = F_i(y_i-)$, the limit of F_i at y_i from the left (Song et al., 2009). We restrict our attention to integer-valued Y_i , so $F_i(y_i-) = F_i(y_i - 1)$.

Expression (2) contains 2^n terms, becoming unmanageable for n larger than four or five. Our model avoids the n -fold summation in (2) by using a continuous extension of the Y_i proposed by Denuit and Lambert (2005), which we now describe. With integer-valued Y_i , associate a continuous random variable $Y_i^* = Y_i - U_i$, where U_i is uniform on $(0, 1)$ and independent of Y_i and of U_j for $j \neq i$. This process “jitters” Y_i , filling in the gaps between the discrete elements of the support of F_i and producing continuous random variable Y_i^* with distribution function $F_i^*(y) = F_i([y]) + (y - [y])P(Y_i = [y + 1])$ and density function $f_i^*(y) = P(Y_i = [y + 1])$, where $[y]$ denotes the nearest integer less than or equal to y . Note that the parameters of F_i^* and f_i^* are exactly those of F_i , and that Y_i can be recovered from Y_i^* as $Y_i = [Y_i^* + 1]$, so no information is lost. Furthermore, Y_i^* and Y_j^* have the same dependence relationship as Y_i and Y_j (Denuit and Lambert, 2005).

Substituting f_i^* and F_i^* into (1) yields a joint probability density function for Y_1^*, \dots, Y_n^* . Averaging this density over the jitters $\mathbf{U} = (U_1, \dots, U_n)$ yields joint probability mass function for Y_1, \dots, Y_n :

$$h(\mathbf{y}; \Sigma) = E_{\mathbf{U}} \left[|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{Z}^*(\Sigma^{-1} - \mathbf{I}_n)\mathbf{Z}^*\right\} \prod_{i=1}^n f_i^*(y_i - U_i) \right], \quad (3)$$

where $\mathbf{Z}^* = [\Phi^{-1}\{F_1^*(y_1 - U_1)\}, \dots, \Phi^{-1}\{F_n^*(y_n - U_n)\}]'$. A proof that (3) is equal to the joint probability mass function (2) is given in the Web Appendix.

3. Maximum Likelihood Estimation

Given discrete multivariate observations $\mathbf{y}_i = [y_{i1}, \dots, y_{ic}]^T, i = 1 \dots n$, where n denotes the number of subjects and c denotes the number of observations per subject, suppose the marginal means $E(\mathbf{y}_{ij}) = \mu_{ij}$ depend on a vector of covariates \mathbf{X}_{ij} via a vector of parameters β and a link

function. Following Song et al. (2009), suppose the objective is to estimate β and the dispersion parameters ϕ of the marginal distributions of the y_{ij} , while accounting for the dependence between the observations on a single subject.

Equation (3) can be used as a likelihood for the parameters of the model in the following way. Model the copula association matrix Σ_i for the i th subject as a function of a vector of parameters α . The method can accommodate different Σ_i , but for simplicity, assume $\Sigma_i = \Sigma$ is the same for all i , for example, Σ describes an exchangeable correlation structure. The expected likelihood is $L(\beta, \phi, \alpha; \mathbf{y}) = E_{\mathbf{U}} \prod_{i=1}^n [|\Sigma|^{-1/2} \exp\{-\frac{1}{2}\mathbf{Z}_i^*(\Sigma^{-1} - \mathbf{I}_c)\mathbf{Z}_i^*\} \prod_{j=1}^c P(Y_{ij} = y_{ij})]$, where $\mathbf{Z}_i^* = [\Phi^{-1}\{F_{i1}^*(y_{i1} - U_{i1})\}, \dots, \Phi^{-1}\{F_{ic}^*(y_{ic} - U_{ic})\}]'$. The expected likelihood can be approximated by averaging over a large number of jitters:

$$\begin{aligned} \tilde{L}(\beta, \alpha, \phi; \mathbf{y}) &= \frac{1}{m} \sum_{k=1}^m \left(\prod_{i=1}^n \left[|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}_{i,k}^*(\Sigma^{-1} - \mathbf{I}_c)\mathbf{z}_{i,k}^*\right\} \right. \right. \\ &\quad \left. \left. \times \prod_{j=1}^c P(Y_{ij} = y_{ij}) \right] \right), \quad (4) \end{aligned}$$

where $\mathbf{z}_{i,k}^* = [\Phi^{-1}\{F_{i1}^*(y_{i1} - u_{i1,k})\}, \dots, \Phi^{-1}\{F_{ic}^*(y_{ic} - u_{ic,k})\}]'$, and the $u_{i,j,k}$ are generated as independent uniform on $(0, 1)$ for $i = 1, \dots, n, j = 1, \dots, c$, and $k = 1, \dots, m$. Maximum likelihood (ML) estimates $\hat{\beta}, \hat{\alpha}$, and $\hat{\phi}$ can be found by maximizing the log of (4) with respect to parameter vector $\xi = [\beta, \alpha, \phi]$.

Under standard regularity conditions, as m and n go to infinity, the ML estimators will be consistent and asymptotically normal with asymptotic covariance matrix given by the Fisher information $-\{E(\mathbf{H})\}^{-1}$ where the ij th element of \mathbf{H} is $H_{ij} = \partial^2 \log(L) / (\partial \xi_i \partial \xi_j)$.

The variance of ML estimator $\hat{\xi}_i$ based on approximate likelihood (4) includes extra variance due to the jitters. When m is sufficiently large, this jittering variance will be small relative to the sampling variance, but “sufficiently large” depends on the model and data. For a particular data analysis, the choice of m can be justified by results obtained by repetition of the estimation procedure. For the analysis in Section 4, $m = 1000$ was chosen by inspecting the increased variance for the estimated parameter of interest over 500 repetitions of the algorithm. Details are given in Section 4.

4. Example

In this section, we analyze the toenail infection data of De Backer et al. (1996). Molenberghs and Verbeke (2005) use these data to illustrate GEE and generalized linear mixed model analyses. The data consist of up to seven binary observations on each of 294 subjects who had been randomly assigned to one of two treatment groups. The observations, taken at regularly scheduled time points, are coded as 1 if the subject’s infection was severe and 0 otherwise. The inferential goals are to determine if the two treatments differ and if the percentage of severe infections decreased over time. We restrict our attention to the 224 subjects observed at all

Table 1

Estimates and standard error from GEE and ML analyses of the toenail data. The adjusted SE accounts for the variance introduced by approximating the likelihood using $m = 1000$ jitters per observation in (4).

β	GEE		ML		Adj. SE
	$\hat{\beta}$	SE	$\hat{\beta}$	Unadj. SE	
Int.	-0.6472	0.2081	-0.6490	0.1380	0.1439
T	0.0019	0.3068	0.0748	0.1890	0.1992
t	-0.2166	0.0427	-0.2120	0.0304	0.0313
$T \times t$	-0.0486	0.0688	-0.0363	0.0438	0.0458

seven time points, though our approach accommodates differing cluster sizes.

Let Y_{ij} denote the j th observation on the i th subject. Assume $Y_{ij} \sim \text{Bernoulli}(\pi_{ij})$ and $\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$, where T_i indicates the treatment for subject i , and t_{ij} represents the j th time point for the i th subject (Molenberghs and Verbeke, 2005). Assuming an exchangeable correlation structure, we fit the GEE2 “alternating logistic regression” model of Molenberghs and Verbeke (2005) using SAS PROC GENMOD. Parameter estimates and standard errors are given in Table 1. These differ from Molenberghs and Verbeke (2005) because we use only the data from subjects with all seven observations.

Table 1 compares GEE estimates and ML estimates of $\beta_k, k = 0, \dots, 3$. ML estimates were obtained by maximizing the log of (4) with respect to $\xi = [\alpha, \beta_0, \beta_1, \beta_2, \beta_3]$, where α represents the off-diagonal elements of the exchangeable ML copula association matrix Σ .

We used $m = 1000$ jitters to approximate the likelihood. To verify that the chosen m is sufficiently large for the data and model, we calculated an adjusted standard error to account for the extra variance due to jittering. Table 1 gives two ML standard errors, unadjusted and adjusted. The unadjusted standard error was derived from a numerical approximation to the Hessian H for the single realization of (4) used to obtain the $\hat{\beta}_k$. The adjusted standard error was calculated as follows. The sampling variance of $\hat{\beta}_k$ can be decomposed as

$$\text{var}(\hat{\beta}_k) = E\{\text{var}(\hat{\beta}_k | \mathbf{U})\} + \text{var}\{E(\hat{\beta}_k | \mathbf{U})\}. \quad (5)$$

We estimate these two components by repeating the estimation procedure 500 times, producing 500 ML estimates of β_k and 500 variance estimates. Each ML estimate was obtained by maximizing the log of (4) for a different random \mathbf{U} . Each variance estimate was the appropriate diagonal element of the approximated inverse Hessian of the log of (4) given \mathbf{U} . The first component of (5) was estimated by averaging the 500 variance estimates, and the second component was estimated by the sample variance of the 500 ML estimates. The adjusted standard error is the sum of these two estimated components. Table 1 shows that adjusted standard errors are all less than 105% of the unadjusted standard errors, except for $SE(\hat{\beta}_1)$. Molenberghs and Verbeke (2005) point out that the randomization of treatments implies $\beta_1 = 0$ but include it in the model for generality. For the parameter of interest β_2 ,

the ratio of adjusted SE to unadjusted SE was less than 103%, so that inference is practically unchanged after accounting for the jitters. Increasing m will reduce the additional variance, but this reduction must be balanced with increased computational burden.

Matlab code for the ML analysis is available in the Web Appendix.

5. Simulations

Song et al. (2009) compared the efficiency of the ML (VGLM) estimators $\hat{\beta}_{ML}$ to the GEE estimators $\hat{\beta}_{GEE}$ by calculating the asymptotic relative efficiency for a hypothetical three-period crossover trial with binary response at each trial. Marginal means for the i th subject at the j th trial are modeled as

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + \beta_2(j - 1), \quad j = 1, 2, 3, \quad (6)$$

where the true value of β is $[\beta_0, \beta_1, \beta_2] = [0, 0.5, 0.5]$. Covariate x_{ij} indicates whether at time j subject i received a placebo ($x_{ij} = 0$) or an active drug ($x_{ij} = 1$), each of which was equally likely. The GEE working correlation structure was modeled as either exchangeable or independent, and the VGLM association was modeled as exchangeable. Because the response for each subject was trivariate, Song et al. (2009) were able to derive closed-form expressions for the asymptotic variances of $\hat{\beta}_{ML}$ and $\hat{\beta}_{GEE}$ as functions of association parameter α .

Section 2 introduced an approximation to the model of Song et al. (2009) that is feasible for larger cluster sizes. Because the asymptotic variances for this approximation are not available in closed form, we conducted a simulation study. We simulated longitudinal binary responses for $n = 50$ and $n = 250$ subjects, where each subject is observed on $c = 10$ occasions. We assumed model (6) but without the carry-over effect β_2 , that is, marginal means for the i th subject at the j th trial are $\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{ij}, j = 1, \dots, 10$, where the true value of β is $[\beta_0, \beta_1] = [0, 0.5]$. The correlation structure was simulated as exchangeable using four correlation values, $\rho \in \{0.01, 0.26, 0.63, 0.84\}$. These correlation values are the observed Monte Carlo correlations of the simulated data, not the association parameter of the Gaussian copula. For each of the eight combinations of ρ and n , 500 data sets were simulated. GEE estimates of the parameters were obtained using the procedure in Section 4 with exchangeable working correlation, and again with independent working correlation. ML estimates were obtained by maximizing the log of (4). The number of jitters for the likelihood approximation was set at $m = 1000$, because the results appeared nearly unchanged for $m = 5000$. Performing repeated estimations with each simulated data set and calculating an adjusted standard error, as in Section 4, was not computationally feasible, but informal exploration suggests that smaller m is sufficient when either the sample size is large or the dependence is weak, that is, when the information content of the sample is high.

Observed relative efficiencies $\text{RE}(\hat{\beta}_1) = \sum_{i=1}^{500} (\hat{\beta}_{1,ML} - \bar{\hat{\beta}}_{1,ML})^2 / [\sum_{i=1}^{500} (\hat{\beta}_{1,GEE} - \bar{\hat{\beta}}_{1,GEE})^2]^{-1}$ versus ρ are plotted in Figure 1. The simulations reveal an interesting phenomenon not apparent when only asymptotic relative efficiencies are considered. Song et al. (2009) observed that the asymptotic

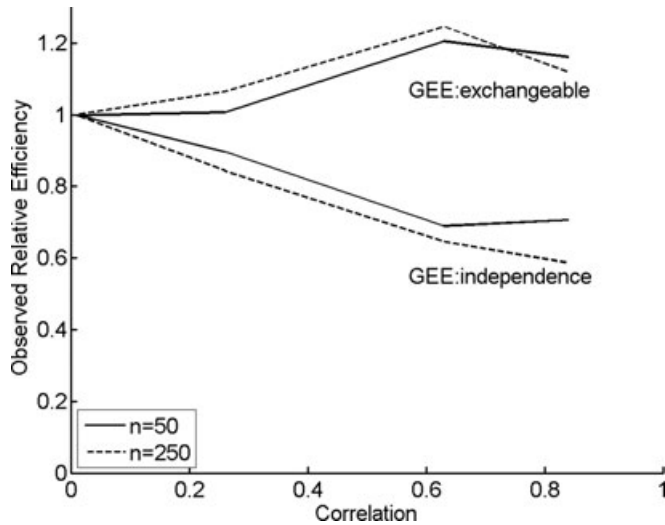


Figure 1. Observed efficiency of the ML estimator relative to the GEE estimators of β_1 , the coefficient of the treatment indicator in the simulated data, as a function of increasing correlation.

performance of the ML estimator is only slightly better than that of the GEE estimator and only at high correlations, when the GEE models the working correlation as exchangeable. The asymptotic performance of the ML is noticeably better when the GEE assumes an independent working correlation structure. Our simulations show that for moderate to high correlations, provided the correlation is not ignored, GEE estimators are actually *more* efficient than ML estimators. This pattern is apparent at sample sizes of both $n = 50$ and $n = 250$ subjects, though for $n = 250$, the observed relative efficiency drops as correlation increases from $\rho = 0.63$ to $\rho = 0.84$.

This apparent contradiction can be explained by the shrinking effective sample size as the degree of dependence increases. Although Song et al. (2009) showed that the ML estimator is asymptotically more efficient than the GEE estimator, in finite samples the amount of independent information decreases as dependence increases, widening the discrepancy between

observed and asymptotic quantities. The GEE estimator is apparently less sensitive to this loss of information, and performs better at high correlations.

To compare variance estimators for the two procedures, we calculated the coverage rate of nominal 95% confidence intervals of β_1 . The ML variance estimator was the appropriate diagonal entry of a numerical approximation of the observed inverse information matrix. The GEE variance estimator was the robust sandwich variance estimator as calculated by the `gee` function of **R** (Carey, Lumley, and Brian, 2007; **R** Development Core Team, 2008). Confidence intervals were calculated as $\hat{\beta}_1 \pm 1.96 \cdot SE(\hat{\beta}_1)$. Table 2 lists the results. All coverage rates were close to 95% except for ML intervals with high dependence, where the coverage was about 89%.

To assess if the confidence coverage was high due to positive bias in variance estimators, we calculated the percentage of intervals containing 0 (recall the true parameter is $\beta_1 = 0.5$). These rates are shown in Table 3. For GEE assuming independence, and $n = 50$, well over 90% of intervals contained zero, whereas for $n = 250$ and GEE (independence), only one out of 500 intervals contained zero. GEE (exchangeable) and ML intervals for $n = 250$ all excluded zero. For $n = 50$, and $\rho = 0.01$, about 20% of both GEE (exchangeable) and ML intervals contained zero. For higher correlations, the rates dropped, but substantially fewer GEE intervals contained zero.

6. Conclusion

This article gives an approximation to the Gaussian copula VGLM of Song et al. (2009), allowing ML estimation of model parameters for larger clusters. The approximation enjoys the same advantages as the VGLM, such as the ability to model multivariate data where dependent observations have marginal distributions from different families, as well as the strengths of a ML approach. However, for parameter estimation from longitudinal data, our simulations suggest that the GEE estimator can be more efficient in finite samples, and that its robust variance estimator gives good confidence coverage without inordinate positive bias.

The Gaussian copula model has shown more promise in the spatial setting. Spatial data sets typically consist of over 100 observations, all of which are assumed dependent, so imple-

Table 2
Nominal 95% confidence coverage from simulation study

	$\rho = 0.01$		$\rho = 0.26$		$\rho = 0.63$		$\rho = 0.84$	
	$n = 50$	$n = 250$	$n = 50$	$n = 250$	$n = 50$	$n = 250$	$n = 50$	$n = 250$
ML	0.948	0.949	0.946	0.951	0.982	0.968	0.886	0.888
GEE exch.	0.933	0.947	0.934	0.940	0.952	0.962	0.932	0.940
GEE indep.	1.000	0.945	1.000	0.938	1.000	0.964	1.000	0.952

Table 3
Rates of nominal 95% confidence intervals containing 0

	$\rho = 0.01$		$\rho = 0.26$		$\rho = 0.63$		$\rho = 0.84$	
	$n = 50$	$n = 250$	$n = 50$	$n = 250$	$n = 50$	$n = 250$	$n = 50$	$n = 250$
ML	0.197	0.000	0.198	0.000	0.078	0.000	0.060	0.000
GEE exch.	0.197	0.000	0.140	0.000	0.010	0.000	0.008	0.000
GEE indep.	0.960	0.000	0.978	0.000	0.970	0.000	0.940	0.002

mentation of the copula likelihood (2) is not feasible. Our approximated likelihood gives a model for discrete spatial data that is appealingly similar to a traditional geostatistical model and, unlike a spatial GEE approach, is able to model high dependence among counts (Madsen, 2009). Simulation results in Madsen (2009) show that ML estimators are more efficient than GEE estimators for data with strong spatial dependence, contradicting the results of the simulation study in Section 5. This suggests that the superior asymptotic efficiency of ML over GEE estimators may manifest for larger clusters than $c = 10$ considered here.

7. Supplementary Materials

A Web Appendix containing the proof of the equivalence of equations (2) and (3), as well as `Matlab` code and data for the analysis in Section 4, are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

REFERENCES

- Carey, V. J., Lumley, T., and Brian, R. (2007). *gee: Generalized Estimation Equation Solver*. R package version 4.13–13. <http://cran.r-project.org>
- De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: Terbafine 250 mg/day vs. itraconazole 200 mg/day—a double-blind comparative trial. *British Journal of Dermatology* **134**, 16–17.
- Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93**, 40–57.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall/CRC.
- Madsen, L. (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. *Journal of Agricultural, Biological, and Environmental Statistics* **14**, 375–391.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edition. New York: Springer.
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* **8**, 229–231.
- Song, P. X.-K. (2007). *Correlated Data Analysis*. New York: Springer.
- Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* **65**, 60–68.

Received July 2009. Revised February 2010.

Accepted March 2010.

The authors replied as follows:

This short note discusses the efficiency and computational issues of the maximum likelihood estimator for vector generalized linear models proposed in Song, Li, and Yuan (2009, *Biometrics* **65**, 60–68). A simulation study is further performed to evaluate the relative efficiency of the proposed vector generalized linear model with the generalized estimating equations (GEE).

1. Introduction

A major contribution of Madsen and Fang (2010) is the application of Denuit and Lambert's (2005) idea of “continuous fill-in” (CFI), which essentially fills simulated continuous values in a gap between two adjacent integer values given by a discrete random variable. As a result, discrete-valued observations are then converted into continuous ones. An advantage of this approach is that the likelihood function can be derived under the copula model with continuous margins, which is known to be analytically more convenient and computationally simpler. In particular, we are pleased to learn this new approach as an alternative to Song et al.'s (2009) direct maximum likelihood estimation (MLE) for correlated discrete data.

Although we appreciate very much the novelty of Denuit and Lambert's CFI idea, we do see some potential confusion and dilemmas of this approach, which could complicate related data analysis and interpretations. In our views, Denuit and Lambert's CFI may be regarded as an approach of data augmentation, which is widely applied in various statistical analyses. What appears special in Madsen and Fang's paper is that their data augmentation method leads to a different likelihood function from that of the original data, and thus additional effort is needed to integrate out those filled values. As shown in their paper, such effort of integration turns out to be a highly nontrivial and computationally tedious operation. Several approximations have been taken in the implementation of this method. We found that the simulation method proposed in their paper is similar to the method of importance sampling (Robert and Casella, 1999, Section 3.3). It is known that the Monte Carlo expectation–maximization (EM) algorithm is another popular method to handle augmented data. Our experience suggests that the Monte Carlo EM works better in most cases than the importance sampling due to the use of posterior distribution. Thus, we wonder if the authors ever considered the Monte Carlo EM algorithm in their implementation of the CFI idea.

The computational burden in the direct MLE considered in Song et al. (2009) lies in the choice of precisely evaluating the cumulative multivariate normal distribution. There exist other choices of carrying out such evaluation, e.g., Monte Carlo simulation as considered in Madsen and Fang's paper. Thus, when the Monte Carlo method is adopted to modify Song et al.'s direct MLE, the resulting method would handle correlated discrete data of large dimension. As a matter of fact, a similar idea has been explored by Pitt, Chan, and Kohn (2006). Also, as suggested in Song et al. (2009), composite likelihood approach (Varin, 2008) is a promising alternative to extend the direct MLE of Song et al. (2009) to deal with correlated discrete data of arbitrary dimensions.

In addition, we are concerned with the quality control of numerical implementation in Madsen and Fang's approach, because most of the key terms need to be evaluated via intensive simulation. In their paper, the likelihood function is evaluated by simulation, and the standard errors are calculated via bootstrap. It is not clear to us if the simulation error has been carefully assessed before final conclusions are delivered.

Table 1
Observed relative efficiency of direct MLE versus GEE based on 500 simulated data sets

Parameter	Correlation parameter				
	0.02	0.25	0.50	0.75	0.98
<i>n</i> = 50					
β_0	0.933	0.906	0.942	0.755	0.511
β_1	0.971	0.896	0.913	0.737	0.526
β_2	0.935	0.929	0.944	0.798	0.593
<i>n</i> = 80					
β_0	0.905	0.932	0.936	0.711	0.558
β_1	0.958	0.884	0.923	0.725	0.533
β_2	0.906	0.932	0.943	0.761	0.633
<i>n</i> = 250					
β_0	0.938	0.912	0.921	0.751	0.715
β_1	0.932	0.904	0.910	0.757	0.705
β_2	0.935	0.911	0.923	0.787	0.828

2. Simulation Study

We are particularly concerned with the authors’ conclusion “Our simulations show that... GEE estimators are actually more efficient than ML estimator,” which contradicts the theoretical asymptotic results that ML estimators are more efficient than the GEE estimators (Song et al., 2009). Thus, we conducted a simulation study to assess the efficiency of MLE estimator in comparison to GEE estimator under both small and large sample sizes. Because the authors’ simulation cannot be reproduced, due to lack of information on the simulation model, and because it is rare to use a 10-period crossover trial design in practice, we have chosen to conduct a set of simple yet plausible simulations. Specifically, under the setting of two-arm randomized trials, we simulated longitudinal binary responses for $n = 50, 80,$ and 250 subjects. We assumed that the subjects were equally randomized into two arms and each subject was measured repeatedly three times. We generated data from the Gaussian copula model (involving a total of eight cases), in which the marginal means for the i th subject at the j th measurement time were given by

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij},$$

where x_{1i} is a binary treatment indicator defined as $x_{1i} = 1$ for treatment and 0 for control, and x_{2ij} is the measurement time with $x_{2ij} = j, j = 1, 2, 3, 4$. We set $(\beta_0, \beta_1, \beta_2) = (0.5, 0.9, -0.4)$ and considered five values for the correlation parameter of the Gaussian copula, $\rho = 0.02, 0.25, 0.50, 0.75,$ and 0.98 . We simulated 500 data sets and reported the observed relative efficiencies $RE = \sum_{i=1}^{500} (\hat{\beta}_{k, MLE} - \tilde{\beta}_{k, MLE})^2 / \sum_{i=1}^{500} (\hat{\beta}_{k, GEE} - \tilde{\beta}_{k, GEE})^2$ in Table 1. In both estimation methods, we used exchangeable working correlation matrix. Note that Song et al. (2009) calculated the RE directly using available analytic formulas. Here the simulation-based RE would provide confirmative evidence for or against that obtained from the formulas.

As shown in Table 1, our simulation result confirms that the direct MLE is clearly more efficient than the GEE under finite samples, and such efficiency gain increases when the

within-subject correlation increases. For example, even under the small sample size of $n = 50$, when estimating the treatment effect parameter β_1 , the RE of the direct MLE against the GEE are 0.971, 0.913, and 0.526, respectively, with the correlation parameter equal to 0.02, 0.5, and 0.98. This result is fully consistent with the theoretical asymptotic RE that derived in Song et al. (2009) (see Equation 14 and Figure 1), and also is in full agreement with the results of Liang and Zeger (1986).

3. Conclusion

We found the development given in Madsen and Fang’s paper interesting. It provides a potential approach to address the computational challenge in the direct MLE. However, in terms of the relative efficiency of the GEE estimator and the MLE estimator in the Gaussian copula joint model, for theoretical reasons, and based on our simulation study, we reach a different conclusion, which is certainly worth more exploration.

The authors thank the editor for the helpful comments and the invitation for this rejoinder.

REFERENCES

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93**, 40–57.
 Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
 Madsen, L. and Fang, Y. (2010). Joint regression analysis for discrete longitudinal data. *Biometrics* xx–xx.
 Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* **93**, 537–554.
 Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.
 Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* **65**, 60–68.
 Varin, C. (2008). On composite marginal likelihood. *Advances in Statistical Analysis* **92**, 1–28.

Peter X.-K. Song
 Department of Biostatistics,
 University of Michigan,
 Ann Arbor, Michigan 48109-2029, U.S.A.
 email: pxsong@umich.edu

Mingyao Li
 Department of Biostatistics and Epidemiology,
 University of Pennsylvania,
 Philadelphia, Pennsylvania, 19104-6021, U.S.A.
 email: mingyao@mail.med.upenn.edu

and

Ying Yuan
 Department of Biostatistics,
 The University of Texas MD Anderson Cancer Center,
 Houston, Texas, 77030, U.S.A.
 email: yyuan@mdanderson.org