Published in final edited form as:
Technometrics. 2010 August 1; 52(2): 340–348. doi:10.1198/TECH.2010.09055.

Likelihood Analysis of Multivariate Probit Models Using a Parameter Expanded MCEM Algorithm

Huiping Xu and
Department of Mathematics and Statistics, Mississippi State University, Mississippi State, MS 39762 (hxu@math.msstate.edu)

Bruce A. Craig
Department of Statistics Purdue University, West Lafayette, IN 47907 (bacraig@stat.purdue.edu)

Abstract

Multivariate binary data arise in a variety of settings. In this paper, we propose a practical and efficient computational framework for maximum likelihood estimation of multivariate probit regression models. This approach uses the Monte Carlo EM (MCEM) algorithm, with parameter expansion to complete the M-step, to avoid the direct evaluation of the intractable multivariate normal orthant probabilities. The parameter expansion not only enables a closed-form solution in the M-step but also improves efficiency. Using the simulation studies, we compare the performance of our approach with the MCEM algorithms developed by Chib and Greenberg (1998) and Song and Lee (2005), as well as the iterative approach proposed by Li and Schafer (2008). Our approach is further illustrated using a real-world example.

Keywords

Correlated binary data; Gibbs sampler; Monte Carlo EM algorithm; Multivariate probit; Parameter expansion

1 INTRODUCTION

Correlated binary data arise in a variety of applications related to biological, social, medical, and engineering research. In industrial quality control experiments, for example, several quality characteristics may be monitored simultaneously (Lu, 1998), or a single quality attribute may be monitored over time (Girard and Parent, 2001). Since the work of Ashford and Sowden (1970), the multivariate probit (MP) model has been a popular method for analyzing such data. The MP model is described in terms of a correlated multivariate normal variable that is linked to the observed multivariate discrete variable through a threshold specification.

Although the MP model can accommodate an arbitrarily complicated correlation structure, the computational burden associated with the evaluation of its likelihood function makes the general structure an uncommon choice. Instead special correlation structures have been proposed (see, e.g., Kolakowski and Bock, 1981; Ochi and Prentice, 1984). With these restrictive structures, the multivariate normal probabilities can be easily approximated using numerical methods. This simplifies the problem but at the expense of being able to examine the general correlation structure among variables.

Alternative approaches, including exploratory factor analysis models (Bock and Aitkin, 1981), have been proposed. These approaches are extendable to the general correlation case. In particular, the correlation matrix is assumed to have the form $\mathbf{\Lambda\Lambda}' + \mathbf{D}^2$, where $\mathbf{\Lambda}$ is a $p \times m$ factor loading matrix, $\mathbf{D}^2$ is a $p \times p$ diagonal matrix, and p and m are the number of dichotomous responses and the number of latent factors, respectively. This model has been extended by Bock and Gibbons (1996) to include fixed covariates and applied to studies on early childhood development of psychiatric disorders (Gibbons and Lavigne, 1998) and health service utilization and insurance coverage (Gibbons and Wilcox-Gok, 1998).

In these models, the normal orthant probabilities involved in the evaluation of the likelihood function were approximated using high dimensional Gaussian-Hermite quadrature with five or more quadrature points in each dimension. Meng and Schilling (1996) point out a reliability problem for the Gaussian-Hermite quadrature in approximating the high dimensional normal orthant integrals. Instead an MCEM algorithm was recommended for the maximum likelihood (ML) estimation.

This MCEM approach was applied by Song and Lee (2005) in their confirmatory factor analysis for MP models, which can also be extended to handle an arbitrary correlation structure. Specifically, the correlation matrix is assumed to have the form $\mathbf{\Gamma} + c\mathbf{I}$, where $\mathbf{I}$ is an identity matrix, c is a pre-assigned value, and $\mathbf{\Gamma}$ is a positive definite symmetric matrix. Although extendable to the general correlation case, the choice of c has to be fixed before model estimation, which is not realistic in practice. An inappropriate value of c can result in convergence to a non-optimal point.

Likelihood analysis of MP models with a general correlation structure was also considered by Chib and Greenberg (1998) using the MCEM algorithm. In the M-step, the correlation coefficients were updated with a Newton-Raphson type routine. Li and Schafer (2008) considered the MP model for longitudinal data, where the multivariate normal orthant probabilities were approximated using the Genz method (Genz, 1992, 1993). The parameter estimates were obtained by iteratively maximizing the log-likelihood with respect to one set of parameters (i.e., regression coefficients or correlation coefficients) with the other set of parameters fixed at their current values. While these approaches are general, the computational effort can be rather heavy since both approaches involve high-dimensional optimizations.

Inspired by the idea of parameter expansion and its use in the Bayesian analysis of correlated binary data (Lawrence, Bingham, Liu, and Nair, 2008), we propose a method for the maximum likelihood inference in the MP model with a general correlation structure using a similar expansion. This technique, in combination with the Monte Carlo technique, is used to overcome the previously mentioned computational difficulties of the other methods. The parameter expansion technique was originally proposed to accelerate the convergence rate of EM algorithms (Liu, Rubin, and Wu, 1998). This approach is used here to simplify the M-step of the MCEM algorithm, with the added benefit of faster convergence. Convergence of the parameter expanded MCEM (PX-MCEM) algorithm is addressed using the Genz method.

The rest of the paper is organized as follows. Section 2 describes the structure of the general MP model. Implementation of the model using the PX-MCEM algorithm, along with the convergence of the algorithm and standard error computations, is presented in Section 3. Data from a real-world application are analyzed in Section 4, where a comparison of the performance of our approach to methods proposed by Chib and Greenberg (1998), Song and Lee (2005), and Li and Schafer (2008), is provided using simulation studies. This is followed by a brief discussion.

2 THE MULTIVARIATE PROBIT MODEL

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})'$ be a vector that denotes the binary responses of the i^{th} individual ($i = 1, 2, \ldots, N$). Let $\mathbf{z}_i$ denote a p-variate latent variable that is normally distributed with a mean vector $\boldsymbol{\beta}\mathbf{x}_i$ and variance-covariance matrix $\mathbf{\Sigma}$, where $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{i,q-1})'$ is a q-vector of covariates and $\boldsymbol{\beta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{q-1})$ is a $p \times q$ matrix of regression coefficients of $\mathbf{z}$ on $\mathbf{x}$. The observed binary vector $\mathbf{y}_i$ is associated with the underlying $\mathbf{z}_i$ in the following way:

$$y_{ij}=I(z_{ij}>0),\ j=1,2,\ldots,p,$$

where $I(\cdot)$ is an indicator function. This implies that the probability of the response $\mathbf{y}_i$, given the covariates $\mathbf{x}_i$ and the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, is

$$P(\mathbf{y}_i \mid \mathbf{x}_i, \beta, \boldsymbol{\Sigma})=\int_{B_{i1}} \cdots \int_{B_{ip}} \phi_p(\mathbf{t};\beta\mathbf{x}_i, \boldsymbol{\Sigma})d\mathbf{t},$$

where $\phi_p(\mathbf{t}; \boldsymbol{\beta}\mathbf{x}_i, \boldsymbol{\Sigma})$ is the density of a p-variate normal distribution with mean vector $\boldsymbol{\beta}\mathbf{x}_i$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The interval B_{ij} is $(-\infty, 0]$ if $y_{ij} = 0$ and $(0, \infty)$ if $y_{ij} = 1$.

It has been noted by Chib and Greenberg (1998) that the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are not identifiable according to the observed-data likelihood. For any diagonal matrix $\mathbf{D}$ with positive diagonal elements, it can be shown that

$$P(\mathbf{y}_i \mid \mathbf{x}_i, \beta, \boldsymbol{\Sigma})=P(\mathbf{y}_i \mid \mathbf{x}_i, \mathbf{D}\beta, \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}) \tag{1}$$

This implies that the variances in the matrix $\boldsymbol{\Sigma}$ cannot be estimated based on the likelihood function. For simplicity, we set them to be unity. Thus the variance-covariance matrix $\boldsymbol{\Sigma}$ is restricted to be a correlation matrix $\mathbf{R} = (\rho_{ij})$.

Augmenting the observed binary data $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$ with the latent variables $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N]$, the complete-data likelihood function can be written as

$$L_{\text{com}}(\theta \mid \mathbf{y}, \mathbf{z})=(2\pi)^{-\frac{Np}{2}}\ |\mathbf{R}|^{-\frac{N}{2}} \exp\left\{-\frac{1}{2}tr\left(\mathbf{R}^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\beta\mathbf{x}_i)(\mathbf{z}_i-\beta\mathbf{x}_i)'\right)\right\}\times\prod_{i=1}^{N}\prod_{j=1}^{p}I(z_{ij}\in B_{ij}), \tag{2}$$

where $\boldsymbol{\theta}$ denotes the model parameters $\boldsymbol{\beta}$ and ρ_{ij}'s. Integrating over $\mathbf{z}_i$'s in (2) yields the observed-data likelihood of the MP model,

$$L_{\text{obs}}(\theta \mid \mathbf{y})=\prod_{i=1}^{N}P(\mathbf{y}_i \mid \mathbf{x}_i, \beta, \mathbf{R}).$$

3 ML ESTIMATION VIA THE EM ALGORITHM

The EM algorithm is a powerful tool to turn to when incomplete data are involved (Dempster, Laird, and Rubin, 1977). For MP models, a challenge arises in that the variance-covariance matrix $\boldsymbol{\Sigma}$ is restricted to be a correlation matrix. The maximization step of the EM algorithm with respect to the correlation coefficients does not have a closed-form solution. Direct maximization of the conditional expectation of the complete-data likelihood function can be computationally intractable for high-dimensional problems.

Inspired by the parameter expansion technique (Liu et al., 1998) and its use in the Bayesian analysis of multivariate probit models (Lawrence et al., 2008), we propose expanding the parameters in the following way:

$$\alpha = \mathbf{V}^{\frac{1}{2}}\beta,$$
$$\boldsymbol{\Sigma} = \mathbf{V}^{\frac{1}{2}}\mathbf{R}\mathbf{V}^{\frac{1}{2}},$$

where $\mathbf{V}$ is a $p \times p$ diagonal matrix with positive diagonal elements. The transformed matrix $\boldsymbol{\Sigma}$ becomes a general variance-covariance matrix. The parameter expansion preserves the observed-data likelihood according to (1) and leads to the following expanded complete-data likelihood:

$$L_{\text{x-com}}(\theta^* \mid \mathbf{y}, \mathbf{z})=(2\pi)^{-\frac{N_p}{2}}|\boldsymbol{\Sigma}|^{-\frac{N}{2}} \exp\left\{-\frac{1}{2}tr\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{N}(\mathbf{z}_i-\alpha\mathbf{x}_i)(\mathbf{z}_i-\alpha\mathbf{x}_i)'\right)\right\}\times\prod_{i=1}^{N}\prod_{j=1}^{p}I(z_{ij}\in B_{ij}).$$

Here $\boldsymbol{\theta}^*$ denotes the expanded parameters including $\boldsymbol{\alpha}$ and the distinct values in $\boldsymbol{\Sigma}$.

The E-step at iteration t of the EM algorithm with a current value of the parameter estimate $\boldsymbol{\theta}^{(t)}$ involves evaluating $Q(\boldsymbol{\theta}^* \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}) = E\{\log L_{\text{x-com}}(\boldsymbol{\theta}^* \mid \mathbf{y}, \mathbf{z}) \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\}$, where the expectation is taken with regard to the conditional distribution of $\mathbf{z}$ given the observed data $\mathbf{y}$. It can be shown that

$$\begin{aligned}Q(\theta^* \mid \mathbf{y}, \theta^{(t)})=&-\frac{N_p}{2}\log(2\pi)\\&-\frac{N}{2}\log|\boldsymbol{\Sigma}|\\&-\frac{1}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\left[E\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' \mid \mathbf{y}, \theta^{(t)}\right)-\alpha E\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{z}_i' \mid \mathbf{y}, \theta^{(t)}\right)-E\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i' \mid \mathbf{y}, \theta^{(t)}\right)\alpha'+\alpha\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right)\alpha'\right]\right\},\end{aligned}$$

This implies that the E-step involves only the computation of the sufficient statistics for the expanded complete data, $\left\{\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i', \sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right\}$. The details of this step are described in Section 3.1.

Implementing the M step is trivial since a closed-form solution to the maximization of the conditional expectation exists. Setting the derivative of $Q(\boldsymbol{\theta}^* \mid \mathbf{y}, \boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ equal to zero yields the following updates of the expanded parameters:

$$\begin{aligned}\alpha^{(t+1)}&=E\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{x}_i' \mid \mathbf{y}, \theta^{(t)}\right)\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right)^{-1},\\\boldsymbol{\Sigma}^{(t+1)}&=\frac{1}{N}\left\{E\left(\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i' \mid \mathbf{y}, \theta^{(t)}\right)-\left(\alpha^{(t+1)}\right)\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right)\left(\alpha^{(t+1)}\right)'\right\}.\end{aligned}$$

Let $\mathbf{D}$ be the diagonal matrix whose diagonal elements are the same as those of $\boldsymbol{\Sigma}^{(t+1)}$. The original parameters $\boldsymbol{\beta}$ and R can then be updated via parameter reduction:

$$\begin{aligned}\beta^{(t+1)}&=\mathbf{D}^{-\frac{1}{2}}\alpha^{(t+1)},\\\mathbf{R}^{(t+1)}&=\mathbf{D}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{(t+1)}\mathbf{D}^{-\frac{1}{2}}.\end{aligned}$$

3.1 Implementing the E-step via the Gibbs Sampler

To implement the E-step, we need to compute $\sum_{i=1}^{N} E\left(\mathbf{z}_i\mathbf{x}_i' \mid \mathbf{y}_i, \theta^{(t)}\right)$ and $\sum_{i=1}^{N} E\left(\mathbf{z}_i\mathbf{z}_i' \mid \mathbf{y}_i, \theta^{(t)}\right)$, where the expectations are with respect to the density function $f(\mathbf{z}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)})$ ($i = 1, 2, \ldots, N$), which is a p-variate normal density, $N_p(\boldsymbol{\beta}\mathbf{x}_i, \mathbf{R})$, truncated to the region specified by $B_i = B_{i1} \times B_{i2} \times \ldots \times B_{ip}$ (Chib and Greenberg, 1998). Computing these conditional expectations is equivalent to computing the first and second moments of a truncated multivariate normal distribution with a general correlation structure, which has been known to be a difficult task (Meng and Schilling, 1996). To ease the problem, we use the idea of the Monte Carlo EM algorithm to approximate these expectations via a Monte Carlo integration method (Wei and Tanner, 1990).

To simulate samples efficiently from $f(\mathbf{z}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)})$, one can create a Gibbs sampler by cycling though the univariate conditional distributions, which are truncated normal variables (Horrace, 2005). Specifically, the conditional distribution $f(z_{ij} \mid y_{ij}, \boldsymbol{\theta}^{(t)})$ is a univariate normal distribution, $N(\nu, \sigma^2)$, truncated to B_{ij}, where

$$\sigma^2 = 1/(\mathbf{R}^{-1})_{jj}, \text{ and } \nu = \mu_{ij} - \sigma^2 (\mathbf{R}^{-1})_{j,-j}(z_{i,-j} - \mu_{i,-j}).$$

Here μ_{ij} is the jth element of the mean vector $\boldsymbol{\beta}\mathbf{x}_i$, $\mu_{i,-j}$ is the vector excluding the jth element of $\boldsymbol{\beta}\mathbf{x}_i$, $(\mathbf{R}^{-1})_{jj}$ is the jth diagonal element of $\mathbf{R}^{-1}$, and $(\mathbf{R}^{-1})_{j,-j}$ is the jth row of $\mathbf{R}^{-1}$ excluding the jth element. Samples from univariate truncated normal distributions are generated using the exponential accept-reject method (Robert, 1995) when the acceptance region is far away from the mean, and by a ratio of uniforms method otherwise (Kinderman and Monahan, 1977).

Once we have M draws of $\{\mathbf{z}_i^{(m)}, m = 1, 2, \ldots, M\}$ from $f(\mathbf{z}_i \mid \mathbf{y}_i, \boldsymbol{\theta}^{(t)})$ for each individual i, the conditional expectations of the sufficient statistics can be approximated as follows

$$\begin{aligned} \sum_{i=1}^{N} E\left(\mathbf{z}_i\mathbf{x}_i' \mid \mathbf{y}_i, \theta^{(t)}\right) &= \sum_{i=1}^{N} \left\{\frac{1}{M}\sum_{m=1}^{M} \mathbf{z}_i^{(m)}\right\}\mathbf{x}_i', \\ \sum_{i=1}^{N} E\left(\mathbf{z}_i\mathbf{z}_i' \mid \mathbf{y}_i, \theta^{(t)}\right) &= \sum_{i=1}^{N} \left\{\frac{1}{M}\sum_{m=1}^{M} \mathbf{z}_i^{(m)}\mathbf{z}_i^{(m)\prime}\right\}. \end{aligned}$$

It has been demonstrated by Wei and Tanner (1990) that the number of imputations M should be large in order to decrease the Monte Carlo error in the E-step, although it is inefficient to start with a large M when the parameter update is still far away from the true value. Because of this, it was suggested that M should be increased from one iteration to the next (Wei and Tanner, 1990). In situations where the imputation is expensive, the computational cost of this approach can be quite prohibitive.

Alternative algorithms that make more efficient use of the imputed missing values include the cumulative implementation of the MCEM algorithm (Kou, Liu, and Wu, 1998). It generates a small number of draws in each iteration of the EM algorithm and updates the parameter estimates using draws obtained from the current and the previous iterations within an adaptively growing window.

3.2 Convergence of PX-MCEM

A direct way of determining the convergence of the PX-MCEM algorithm is to monitor the plot of the parameter updates $\boldsymbol{\theta}^{(t)}$ against the iteration t. Due to the simulation variability introduced in the E-step, the parameter updates can still fluctuate after convergence. Wei and Tanner (1990) recommended that the algorithm be terminated when the iterates appear to fluctuate randomly. This may be impractical when the number of parameters is large. One alternative is to monitor the observed-data likelihood values, which involves computing the multivariate orthant probabilities, using the Monte Carlo procedure proposed by Genz (1992, 1993). This method allows efficient and accurate computation and can be used with a readily available routine, `pmvnorm`, in the `mvtnorm` package of the statistical programming language `R`. When the log-likelihood values are plotted against t, convergence can be claimed when the plot appears to fluctuate randomly.

3.3 Standard Error Calculations

In the EM literature several approaches have been proposed to obtain the asymptotic variance-covariance matrix of the ML estimates (see, e.g., Louis, 1982). We adopt an approach similar to what was proposed by Meilijson (1989). This approach requires the computation of the score vector, $\mathbf{s}_i$, based on the observed-data log-likelihood function. The score can be obtained through the expectation of $\mathbf{s}_i^*$, which is the score vector based on the complete data, conditional on the response $\mathbf{y}_i$. The asymptotic variance-covariance matrix is thus obtained by inverting the empirical Fisher information matrix,

$$\sum_{i=1}^{N}\mathbf{s}_i\mathbf{s}_i' = \sum_{i=1}^{N} E\left(\mathbf{s}_i^* \mid \mathbf{y}_i, \theta=\widehat{\theta}\right) E\left(\mathbf{s}_i^* \mid y_i, \theta=\widehat{\theta}\right)'.$$

Here the complete-data score vector $\mathbf{s}_i^*$ can be obtained as follows:

$$\begin{aligned}
\mathbf{s}_{i,\beta}^* &= \mathbf{R}^{-1}\left(\mathbf{z}_i\mathbf{x}_i' - \beta\mathbf{x}_i\mathbf{x}_i'\right),\\
\mathbf{s}_{i,\rho_{jk}}^* &= \tfrac{1}{2} tr\left\{\mathbf{R}^{-1}\mathbf{dR}_{jk}\mathbf{R}^{-1}\left[(\mathbf{z}_i - \beta\mathbf{x}_i)(\mathbf{z}_i - \beta\mathbf{x}_i)' - \mathbf{R}\right]\right\},
\end{aligned}$$

where $\mathbf{dR}_{jk}$ is a $p \times p$ matrix with the $(j, k)^{th}$ and $(k, j)^{th}$ elements being 1 and other elements being 0.

The conditional expectations of the above complete-data score functions need to be evaluated only at the last iteration of the EM algorithm. Additional draws are generated to approximate the expectations by their Monte Carlo estimates. In our examples, 20,000 additional draws of the random samples are generated to compute the standard errors.

4 EXAMPLES

4.1 Simulation Study I

The goal of this simulation study is to compare the performance of the PX-MCEM algorithm with the MCEM approaches developed by Chib and Greenberg (1998) (CG-MCEM) and Song and Lee (2005) (SL-MCEM), and the iterative algorithm of Li and Schafer (2008) (LS). Multivariate probit models with $p = 3$ and $p = 6$ binary variables were considered. The latent variables $\mathbf{z}_i$'s were assumed to follow a multivariate normal distribution with mean vector $\boldsymbol{\beta}\mathbf{X}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_i$ and correlation matrix $\mathbf{R}$, where x_i's were drawn independently from a uniform distribution on (0, 1). For each of the binary variables, the true value of the intercept was taken to be −1 and the slope was taken to be 2. For the correlation matrix $\mathbf{R} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$, two choices of ρ were considered, a moderate correlation among the latent variables of $\rho = 0.5$ and a high correlation of $\rho = 0.9$. Here **I** is a $p \times p$ identity matrix and **J** is a $p \times p$ matrix of ones. For each of the four combinations of p and ρ, a sample of size 100 was generated. In order for the CG-MCEM and SL-MCEM approaches to be comparable to the PX-MCEM, we implemented all three methods using the cumulative Monte Carlo technique. A total of 5 draws was generated from each iteration and random draws from the last one third of the iterations were utilized in approximating the conditional expectations.

When fitting the model with an arbitrary correlation structure using the SL-MCEM algorithm, the correlation matrix $\mathbf{R}$ is expressed as $\mathbf{R} = \mathbf{\Gamma} + c\mathbf{I}$, where $c \in (0, 1)$ is a pre-assigned fixed value and $\mathbf{\Gamma}$ is a positive definite symmetric matrix with unknown off-diagonal elements and the diagonal elements fixed at $1 - c$. The value of c has to be very small for problems with high correlation. We chose $c = 0.2$ when $\rho = 0.5$ and $c = 0.02$ when $\rho = 0.9$.

Figure 1 compares the performance of these methods in terms of the proximity to the true MLE, which was obtained by directly maximizing the log-likelihood function. The three MCEM algorithms were run for 60 seconds of CPU time and the LS algorithm was run for 5 iterations. The log-likelihood values were plotted against the CPU time (in seconds). The starting values for these approaches were the same, with the regression coefficient matrix $\boldsymbol{\beta}$ being equal to the MLE based on the independent correlation structure and correlation matrix R being an identity matrix. The optimization routine `optim in R` was used to complete the M-step of the CG-MCEM approach and optimizations of the LS approach. The initial value for these optimizations is taken to be the value obtained in the previous iteration.

Compared to the CG-MCEM algorithm, the performance of the PX-MCEM algorithm is similar for the 3-variate model with a moderate correlation, with the PX-MCEM algorithm showing a slight advantage in terms of computational time. The efficiency of the PX-MCEM algorithm is more evident both in terms of the computational time and accuracy when the dimension or the correlation of the observed data is high. The computational time of each iteration for the PX-MCEM is slightly shorter than that for the CG-MCEM. This resulted in a 20% to 30% increase in the number of iterations within 60 seconds of CPU time. In addition, the PX-MCEM algorithm approaches the neighborhood of the MLE much more rapidly. For the 6-variate problems, at the 10th iteration, the sum of the squared deviations of the PX-MCEM estimate from the true MLE was 0.012 under $\rho = 0.5$ and 0.11 under $\rho = 0.9$, while the corresponding values were 0.22 and 1.35 for the CG-MCEM algorithm, respectively.

The SL-MCEM algorithm is competitive with the PX-MCEM in all four cases with the PX-MCEM algorithm showing a slight edge over the SL-MCEM when $p = 6$. The computational time for one iteration was comparable, hence the number of iterations completed in 60 seconds were about the same for both methods. However, a disadvantage of the SL-MCEM algorithm is that the fixed value c needs to be chosen before running the algorithm. In practice, this may not be realistic since the correlation is typically unknown before estimation.

Compared to the three MCEM algorithms, the LS method approaches the true MLE in only a few iterations. However, it is extremely time consuming to complete one iteration, especially in high-dimensional problems. For the 6-variate problem with $\rho = 0.9$, the sum of squared deviations of the LS approach was 3.08 after 5 iterations, which took about 2500 seconds to complete. On the other hand, the PX-MCEM algorithm had a comparable value of the sum of squared deviations at the 4th iteration, which only took 3.5 seconds. The sum of squared deviations for the SL-MCEM approach was comparable at the 6th iteration, which took approximately 4.7 seconds.

The advantage of the PX-MCEM algorithm over the LS approach is even stronger when the sample size is large. The LS approach uses a stepwise ascent method that iteratively maximizes the log-likelihood function with respect to one parameter at a time, with others fixed at their current value. With large sample sizes, the evaluation of the log-likelihood function involves a large number of calculations of the multivariate normal probabilities, which can be quite time consuming.

These methods were also compared via this study framework using a binary predictor variable x, which was drawn from a Bernoulli distribution with probability 0.5. Again, the PX-MCEM algorithm approached the neighborhood of the true MLE much more rapidly and achieved a greater accuracy than other methods within a fixed amount of time. For such grouped data, the computational burden of evaluating the likelihood is much less since the multivariate normal probabilities only need to be evaluated for groups of individuals. Hence the relative computational time for the LS approach was reduced, but it was still much longer than that of the MCEM algorithms.

4.2 Simulation Study II

The goal of this simulation study is to further examine the performance of the SL-MCEM algorithm when an inappropriate c value is used. A data set with $p = 6$ and $\rho = 0.9$ was generated according to the same simulation scheme as in Simulation Study I. Three values of c were considered: a sensible value $c = 0.05$, an overly large value $c = 0.2$, and an overly small value $c = 0.001$. Both PX-MCEM and SL-MCEM algorithms were run for 120 seconds in CPU time. A total of 50 draws was generated from each iteration and random draws from the last one-third of the iterations were utilized in approximating the conditional expectations.

Since $1 - c$ serves as an upper bound on the correlations, an inappropriate choice of c (typically one too large) would prevent the algorithm from exploring the proper correlation space. This can lead to convergence to a non-optimal point, which is seen from Figure 2 when $c = 0.2$. On the other hand, a small value of c does not necessarily work without any problems. The latent variable $\mathbf{z}_i$ of the Song and Lee (2005)'s model is specified as follows:

$$\mathbf{z}_i=\beta\mathbf{x}_i+\omega_i+\boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\omega}_i \sim N(\mathbf{0}, \boldsymbol{\Gamma})$, $\varepsilon_i \sim N(\mathbf{0}, c\mathbf{I})$, and $\boldsymbol{\omega}_i$ and ε_i are independent. In the Monte Carlo E-step, a Gibbs sampler was used to simulate random draws from $f(\mathbf{z}_i, \boldsymbol{\omega}_i \mid \mathbf{y}_i)$, which is the joint density function of $\mathbf{z}_i$ and $\boldsymbol{\omega}_i$ given the observed data. When the value of c is small, the variances of the conditional distributions $f(\mathbf{z}_i \mid \boldsymbol{\omega}_i, \mathbf{y}_i)$ and $f(\boldsymbol{\omega}_i \mid \mathbf{z}_i, \mathbf{y}_i)$ in the Gibbs sampler would be much smaller than the variances of their corresponding marginals $f(\mathbf{z}_i \mid \mathbf{y}_i)$ and $f(\boldsymbol{\omega}_i \mid \mathbf{y}_i)$. This implies that the sampler would not mix very well and a large number of random draws would be needed in order to estimate the quantities of the E-step. This can be seen from Figure 2 when $c = 0.001$. Although for many problems with low to moderate correlations, it is not necessary to choose a value of c as low as 0.001, our simulation study showed that the SL-MCEM algorithm becomes less efficient as c decreases. Marked improvement on the convergence of the SL-MCEM approach when $c = 0.001$ was noted when 500 instead of 50 random draws were generated within each iteration.

The performance of the SL-MCEM algorithm is similar to that of the PX-MCEM algorithm when $c = 0.05$, although the rate of convergence is still slightly slower. Further examination of the log-likelihood values indicates that the speed of convergence of the SL-MCEM algorithm slows down greatly as it nears the true MLE.

Similar results were found when other values of p and ρ were examined. When an inappropriate value of c is chosen, the SL-MCEM algorithm either fails to converge to the optimal point, or requires a large number of Monte Carlo draws to converge. This suggests that multiple values of c should be used for the SL-MCEM algorithm in order to estimate the general correlation structure. This, however, reduces the efficiency of the SL-MCEM algorithm.

4.3 Simulation Study III

The objective of this simulation study is to examine the performance of the large-sample standard errors for the PX-MCEM algorithm. Five hundred samples of sizes $N = 100$ and $N = 500$, each with $p = 4$, were generated with the true parameter values shown in Table 1. The latent $\mathbf{z}_i$'s were assumed to follow a multivariate normal distribution with mean vector $\boldsymbol{\beta}\mathbf{X}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_i$, where the covariate x_i's were drawn independently from a uniform distribution on (0, 1).

The results of the simulation are summarized in Table 1 using the mean and the standard deviation of the MLEs of 500 samples. The 95% confidence interval was constructed based on the large-sample standard errors and the coverage levels are also presented in the table. For both sample sizes, the parameter estimates agree closely to the true values. The coverage levels are very close to the nominal 95% except for the two high correlations between items 1 and 2, and items 3 and 4 when the sample size is small ($N = 100$). Further examination of the distributions of the MLEs for these two correlations revealed that the distributions of the MLEs are highly skewed. This implies that a bootstrap approach may be used to produce more appropriate confidence intervals in this case.

4.4 Drivers' Perceptions of Headlight Glare

To further illustrate the performance of the PX-MCEM algorithm relative to the other MCEM algorithms, we analyzed a real-world data set obtained from the Bureau of Transportation Statistics (BTS). The Omnibus Survey is a national probability sample conducted monthly by the BTS to monitor the public's satisfaction with various transportation issues. Here, we focus on drivers' perceptions of headlight glare. Three questions pertinent to glare, specifically glare from oncoming and following vehicles at night and from daytime running lights, were included in the survey during the first six months of 2002.

Perceptions of glare were expressed on a 5-point scale, "not noticeable" "barely noticeable", "noticeable but acceptable", "disturbing", or "caused a crash or near miss". These ratings were dichotomized at the neutral category to create binary responses indicating whether traffic glare was "disturbing" or "not disturbing". To investigate whether difference with respect to glare perceptions exists due to demographic factors, age and gender were included in the analysis. Potential influence of the number of dark hours on drivers' perceptions was investigated by including the month of the interview.

The starting value of the parameters was chosen to be the MLE based on independent correlations. A total of 5 draws was generated from each iteration and random draws from the last one third of the iterations were utilized in approximating the conditional expectations. The algorithms were stopped when the maximum change in parameter estimates was less than 0.001 for 5 consecutive iterations. After examining the estimates of the correlation coefficients from the other two MCEM approaches, the value of c was chosen to be 0.2 for the SL-MCEM algorithm.

The trajectory of the log-likelihood values across iterations (Figure 3) shows that the log-likelihood values stabilized after 100 seconds in CPU time (approximately 20 iterations). While all three algorithms converged to the optimal point, the PX-MCEM algorithm shows a slight advantage of efficiency compared to the others. Table 2 shows the maximum likelihood estimates of the regression coefficients and correlations and the asymptotic standard errors obtained from the PX-MCEM algorithm. Compared to males, female drivers were significantly more concerned about the headlight glare from both oncoming and following traffic. Also, drivers older than 35 years had more concerns about the oncoming traffic glare. The analysis showed that for both oncoming and following glare, as the number of dark hours decreases from January to June, drivers became less concerned. However, it did not have a significant effect on the glare from the daytime running lights. Although perceptions on the traffic glares are positively correlated, the highest correlation occurred for concerns with oncoming and following traffic at night.

5 DISCUSSION

Correlated binary data are common in many areas including engineering and medical, social, and biological sciences. In this article, we develop an efficient approach to the ML estimation of multivariate probit models for analyzing these data. The proposed methodology can handle variables that have arbitrarily complicated correlation structures. By using a parameter expanded MCEM method, we not only avoid the direct evaluation of the likelihood values, which involves computing multivariate normal orthant probabilities, but also improve the efficiency of the algorithm. Another advantage of the PX-MCEM algorithm is its simplicity compared to the CG-MCEM algorithm in that the PX-MCEM algorithm has an analytically tractable M-step and hence does not require numerical optimization techniques. Although the SL-MCEM algorithm also has an analytic M-step, it is sensitive to the choice of c. The simulation studies show that, compared to the CG-MCEM, SL-MCEM, and LS approaches, the PX-MCEM algorithm approaches the neighborhood of the MLE rapidly and provides higher accuracy per unit CPU time. The large-sample standard errors are reliable measures of the uncertainty in the regression coefficients. Our approach may be readily extended to likelihood analysis of multivariate probit models for ordinal data.

Acknowledgments

The authors thank Professor Chuanhai Liu for his insightful discussion and comments as well as many helpful suggestions for writing the paper. The authors also thank the editor, the associate editor, and two referees for their constructive comments that substantially improved the article. This research was supported by the National Institute of Health grant R01AG-022090 to Dr. Laura Sands and the National Cancer Institute grant P01 CA48112 to Dr. Bruce Craig.

References

Ashford J, Sowden R. Multi-variate probit analysis. Biometrics 1970;26:535–546. [PubMed: 5480663]

Bock D, Gibbons R. High-dimensional multivariate probit analysis. Biometrics 1996;52:1183–1194. [PubMed: 8962449]

Bock R, Aitkin M. Marginal maximum likelihood estimation of item parameters: application to an EM algorithm. Psychometrika 1981;46:443–445.

Chib S, Greenberg E. Analysis of multivariate probit models. Biometrika 1998;85:347–361.

Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 1977;39:1–38.

Genz A. Numerical computation of multivariate normal probabilities. Journal of Computational and Graphical Statistics 1992;1:141–149.

Genz A. Comparison of methods for the computation of multivariate normal probabilities. Computing Sciences and Statistics 1993;25:400–405.

Gibbons R, Lavigne J. Emergence of childhood psychiatric disorders: A multivariate probit analysis. Statistics in Medicine 1998;17:2487–2499. [PubMed: 9819840]

Gibbons R, Wilcox-Gok V. Health service utilization and insurance coverage: A multivariate probit analysis. Journal of the American Statistical Association 1998;93:63–72.

Girard P, Parent E. Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. Technometrics 2001;43:180–191.

Horrace W. Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis 2005;94:209–221.

Kinderman A, Monahan J. Computer generation of random variables using the ratio of uniform deviates. ACM transactions on Mathematical Software 1977;3:257–260.

Kolakowski D, Bock R. A multivariate generalization of probit models. Biometrics 1981;37:541–551.

Kou, S.; Liu, C.; Wu, Y. Technical report, Department of Statistics. University of Michigan; 1998. Cumulative implementation of Monte Carlo EM.

Lawrence E, Bingham D, Liu C, Nair V. Bayesian inference for multivariate ordinal data using parameter expansion. Technometrics 2008;50:182–191.

Li Y, Schafer D. Likelihood analysis of the multivariate ordinal probit regression model for repeated ordinal responses. Computational Statistics and Data Analysis 2008;52:3474–3492.

Liu C, Rubin D, Wu Y. Parameter expansion to accelerate EM: the PX-EM algorithm. Biometrika 1998;85:755–770.

Louis T. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B 1982;44:226–233.

Lu X. Control charts for multivariate attribute processes. International Journal of Production Research 1998;36:3477–3489.

Meilijson I. A fast improvement to the EM algorithm on its own terms. Journal of the Royal Statistical Society, Series B 1989;51:127–138.

Meng X, Schilling S. Fitting full-information item factor models and an empirical investigation of bridge sampling. Journal of the American Statistical Association 1996;91:1254–1267.

Ochi Y, Prentice R. Likelihood inference in a correlated probit regression model. Biometrika 1984;71:531–543.

Robert C. Simulation of truncated normal variables. Statistics and Computing 1995;5:121–125.

Song X, Lee S. A multivariate probit latent variable model for analyzing dichotomous responses. Statistica Sinica 2005;15:645–664.

Wei G, Tanner M. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association 1990;85:699–704.

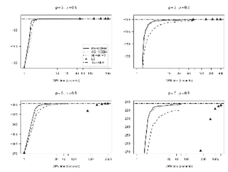

Figure 1.

Values of log-likelihood by Genz method against CPU time for the simulation study I

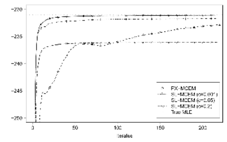

Figure 2.
Values of log-likelihood across iterations for the SL-MCEM with difference values of *c* in comparison to the PX-MCEM for the simulation study II

NIH-PA Author Manuscript
NIH-PA Author Manuscript

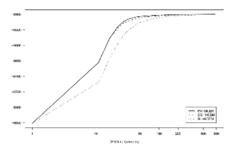

Figure 3.
Values of log-likelihood against CPU time for Drivers' Perceptions data

Table 1

Results of simulation study III

			N = 100			N = 500		
		True	**Mean**	**SD**	**Coverage**	**Mean**	**SD**	**Coverage**
β_0	1	−1.0	−1.042	0.279	97.4%	−0.999	0.117	96.4%
	2	−1.0	−1.042	0.275	98.0%	−1.006	0.115	96.6%
	3	−1.0	−1.034	0.297	96.0%	−1.002	0.115	98.0%
	4	−1.0	−1.016	0.277	97.8%	−0.998	0.119	97.4%
β_1	1	2.0	2.081	0.491	97.8%	1.998	0.211	96.8%
	2	2.0	2.081	0.489	96.8%	2.013	0.208	96.6%
	3	2.0	2.064	0.516	96.4%	2.005	0.204	97.0%
	4	2.0	2.025	0.493	96.6%	1.994	0.205	96.2%
R	(1,2)	0.8	0.794	0.086	93.8%	0.800	0.038	95.6%
	(1,3)	0.2	0.196	0.170	95.0%	0.199	0.074	96.2%
	(1,4)	0.2	0.200	0.166	96.4%	0.200	0.074	96.0%
	(2,3)	0.2	0.182	0.171	96.2%	0.192	0.073	95.4%
	(2,4)	0.2	0.196	0.169	96.6%	0.197	0.075	95.8%
	(3,4)	0.8	0.808	0.086	91.6%	0.802	0.036	93.4%

Table 2

ML estimates and asymptotic standard errors of the Drivers' Perceptions data

	Nighttime Oncoming lights	Nighttime Following lights	Daytime running lights
Regression Coefficients			
Intercept	−0.577	−0.589	−1.963
	(0.054)	(0.054)	(0.118)
Age: 35–54 years	0.096	0.050	0.076
	(0.044)	(0.044)	(0.104)
Age: 55 years or above	0.098	−0.024	0.121
	(0.050)	(0.050)	(0.115)
Gender: female	0.153	0.218	−0.181
	(0.037)	(0.037)	(0.082)
Month: February	0.039	−0.097	0.117
	(0.060)	(0.061)	(0.123)
Month: March	0.013	−0.017	−0.104
	(0.061)	(0.061)	(0.138)
Month: April	−0.145	−0.144	−0.091
	(0.062)	(0.062)	(0.136)
Month: May	−0.093	−0.103	−0.158
	(0.062)	(0.063)	(0.150)
Month: June	−0.135	−0.152	−0.105
	(0.061)	(0.061)	(0.135)
Correlations			
Nighttime following lights	0.561	-	-
	(0.018)	-	-
Daytime running lights	0.241	0.264	-
	(0.050)	(0.051)	-