Biometrics 60, 739–746
September 2004

Bayesian Multivariate Logistic Regression

Sean M. O'Brien* **and David B. Dunson**

Biostatistics Branch, MD A3-03, National Institute of Environmental Health Sciences,
P.O. Box 12233, Research Triangle Park, North Carolina 27709, U.S.A.
**email*: obrien4@niehs.nih.gov

Summary. Bayesian analyses of multivariate binary or categorical outcomes typically rely on probit or mixed effects logistic regression models that do not have a marginal logistic structure for the individual outcomes. In addition, difficulties arise when simple noninformative priors are chosen for the covariance parameters. Motivated by these problems, we propose a new type of multivariate logistic distribution that can be used to construct a likelihood for multivariate logistic regression analysis of binary and categorical data. The model for individual outcomes has a marginal logistic structure, simplifying interpretation. We follow a Bayesian approach to estimation and inference, developing an efficient data augmentation algorithm for posterior computation. The method is illustrated with application to a neurotoxicology study.

Key words: Block updating; Categorical data; Data augmentation; Latent variables; MCMC algorithm; Multiple binary outcomes; Proportional odds.

1. Introduction

In many application areas, such as epidemiologic and biomedical studies, logistic regression is the standard approach for the analysis of binary and ordered categorical outcome data. Regression coefficients from logistic models have simple interpretations in terms of odds ratios that are easily understood by subject-matter researchers. In many cases, outcome data are multivariate or correlated (e.g., due to repeated observations on the same study subject, or on subjects from the same family), and it is appealing to have a model that maintains a marginal logistic interpretation for the individual outcomes while appropriately accounting for the dependency structure.

To motivate the problem, we focus on data from a neurotoxicology study in which rat pups were exposed pre- and postnatally to the pesticide Methoxychlor (MXC) at doses of 0, 5, 50, or 150 mg/kg. From each of 39 litters (∼10/dose group), one male and one female pup were randomly selected for neuromotor testing. The activity level (0 = normal, 1 = elevated) was recorded for each pup at three ages (postnatal days 31, 47, 66). Scientific interest focuses on relating the probability of elevated activity to MXC dose, age, and gender. Investigators would like to base inferences on logistic regression models, which are easily interpretable for readers of biomedical journals.

Common frequentist approaches, which can be used for data of this type, include marginal logistic regression via generalized estimating equations (GEEs; Zeger and Liang, 1986; Prentice, 1988; Lipsitz, Laird, and Harrington, 1991; Carey and Zeger, 1993, among others) and mixed effects logistic regression (Stiratelli, Laird, and Ware, 1984). The GEE approach is often preferred due to simplicity of parameter interpretation and robustness to misspecification of the correlation structure. In contrast, the logistic structure is lost in integrating out the random effects from the mixed effects model. Although such models do have a cluster-specific logistic interpretation, this interpretation may be unfamiliar to subject-matter researchers. Although the GEE approach solves this problem, the justification relies on large sample arguments, and the performance in small to moderate samples is often unclear.

Motivated by the MXC application, in which the sample size is small, this article proposes an exact Bayesian analysis of a marginal logistic model. Although likelihood-based approaches to marginal logistic fitting have also been proposed (Fitzmaurice and Laird, 1993; Glonek and McCullagh, 1995), the complexity of model specification is a considerable practical limitation.

In analyzing multivariate categorical data, Bayesian approaches have several important advantages over quasi-likelihood and likelihood-based frequentist methods. First, by using Markov chain Monte Carlo (MCMC) algorithms to obtain estimates of exact posterior distributions, there is no need to rely on large sample justifications, except in terms of the number of MCMC iterations, which can be increased easily. In addition, Bayesian methods are more flexible in that outside substantive information and data can be incorporated through an informative prior distribution, when such information is available, or a noninformative prior can be chosen. This article is motivated by the need to develop Bayesian methods for multivariate logistic regression, which allow simple noninformative prior distributions (useful for reference analyses), result in straightforward and efficient computation, and are (conceptually) easy to generalize to more complicated data structures, involving mixed categorical and continuous outcomes, multilevel data, or informative censoring.

The key step is to choose an appropriate likelihood function. One possibility is to parameterize the likelihood in terms of marginal response probabilities and multiway association parameters (Fitzmaurice and Laird, 1993; Glonek and McCullagh, 1995). This approach is complicated and involves a large number of parameters unless simplifying assumptions are made. For example, in the MXC study, a total of $2^6 - 1 = 63$ parameters are required to specify the joint distribution of the six binary outcomes obtained from each sibling pair. Qu, Piedmonte, and Medendorp (1995) instead specify the likelihood in terms of latent normal response variables, which are related to predictors through a nonlinear random effects model. The link can be carefully chosen to produce a marginal logistic interpretation. Although this model is simple and parsimonious, it is not ideal for Bayesian inferences because the nonlinear random effects structure may lead to an improper posterior distribution or slow mixing when diffuse priors are chosen. Even with informative priors, efficient posterior computation is challenging, due to the lack of simple forms for the conditional posterior distributions.

Computational considerations have led to the wide use of probit models in Bayesian analyses of categorical data. Because probit models can be specified in terms of underlying normal variables following a linear regression model, posterior computation can proceed via a simple auxiliary variables Gibbs sampling algorithm (Albert and Chib, 1993; Chib and Greenberg, 1998; Chib, 2000, among others). This simple structure facilitates generalizations to more complicated data structures, such as clustered data with mixed categorical and continuous response variables (Dunson, Chen, and Harry, 2003). To generalize the underlying normal structure, Chen and Dey (2000) proposed a rich class of scale mixtures of multivariate normal underlying variables. An advantage of multivariate probit models is that the dependency structure can be described parsimoniously in terms of correlation coefficients of the underlying continuous variable distribution.

Although probit models are clearly appealing due to the relative ease in computation and modeling of the covariance structure, there are some problems with parameter interpretation. Specifically, since the standard normal distribution function does not have a closed form, the regression coefficients are not interpretable as changes in a simple function of the category probabilities. Instead, the coefficients have a more esoteric interpretation in terms of changes in the underlying normal mean. For this reason, logistic models may be preferred.

This article proposes a new likelihood for multivariate logistic regression defined in terms of underlying continuous variables. The dependency structure is modeled parsimoniously via parameters of a continuous multivariate logistic distribution. Because conventional multivariate logistic distributions are severely restricted in the amount of correlation that can be specified between two variables, we propose a new multivariate logistic density that has an unrestricted correlation structure. Posterior computation is very simple and is based on an extension of the Albert and Chib (1993) method used for probit models. Advantages of the proposed approach compared to Bayesian mixed effects logistic models include a marginal logistic interpretation for the regression parameters, a proper posterior distribution under a uniform improper prior, and potentially improved computational efficiency because random effects do not need to be computed.

Section 2 derives the proposed multivariate logistic density and discusses properties. Section 3 proposes applications to multivariate ordered categorical data. Section 4 investigates conditions for propriety of the posterior distribution. Section 5 outlines the algorithm for posterior computation. Section 6 applies the approach to the MXC example, and Section 7 discusses the results.

2. Multivariate Logistic Distributions

To illustrate model specification via underlying continuous variables, first consider the univariate logistic regression model:

$$\text{logit } \Pr(y_i = 1 \,|\, \mathbf{x}_i, \boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}, \tag{1}$$

where y_i is a 0/1 binary outcome, $\mathbf{x}_i$ is a $q \times 1$ vector of predictors, and $\boldsymbol{\beta}$ is a vector of unknown regression coefficients. This model is equivalent to letting $y_i = 1(z_i > 0)$, where $1(\cdot)$ is the indicator function and each $z_i \sim \mathcal{L}(\cdot \,|\, \mathbf{x}_i'\boldsymbol{\beta})$ is an independent logistically distributed random variable with location parameter $\mathbf{x}_i'\boldsymbol{\beta}$ and having density function

$$\mathcal{L}(z_i \,|\, \mathbf{x}_i'\boldsymbol{\beta}) = \frac{\exp\{-(z_i - \mathbf{x}_i'\boldsymbol{\beta})\}}{[1 + \exp\{-(z_i - \mathbf{x}_i'\boldsymbol{\beta})\}]^2}. \tag{2}$$

Using this latent variable specification, one can generalize the model to the multivariate case where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ is a vector of binary variables and $\mathbf{X}_i$ is a $p \times q$ matrix of predictors with rows $\mathbf{x}_{ij}$ by letting $y_{ij} = 1$ $(z_{ij} > 0)$ where $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})'$ follows a joint distribution having univariate logistic marginals $z_{ij} \sim \mathcal{L}(\cdot \,|\, \mathbf{x}_{ij}'\boldsymbol{\beta})$. The implementation of such a model is hampered by the lack of flexible multivariate distributions having logistic marginals (Gumbel, 1961; Malik and Abraham, 1973; Castillo, Sarabia, and Hadi, 1997). To accommodate a flexible correlation structure for our model, this article proposes a new multivariate logistic density.

In generalizing the logistic density to p dimensions, we are motivated to consider densities with a flexible correlation structure, easily interpretable parameters, and a form that results in simplified computation. A density that meets these criteria can be derived by transforming variables that follow a conventional multivariate distribution such as the multivariate normal or multivariate t distribution. To illustrate this approach, suppose that $e_1, \ldots, e_j$ are identically distributed continuous random variables arising from a continuous multivariate distribution with no location or scale parameters, and let $F(\cdot)$ denote the c.d.f. of the marginal univariate distribution of e_j. Since $F(e_j) \sim U(0, 1)$, it follows that $\log[F(e_j)/\{1 - F(e_j)\}]$ is logistically distributed with mean zero, and hence $z_j = \mu_j + [F(e_j)/\{1 - F(e_j)\}] \sim \mathcal{L}(\cdot \,|\, \mu_j)$. Because each z_j has a marginal logistic distribution, the joint distribution of $\mathbf{z} = (z_1, \ldots, z_p)'$ is a p-dimensional multivariate logistic distribution with location parameter $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$.

For notational simplicity, our multivariate logistic distribution is defined without scale parameters. It is parameterized in terms of a location vector $\boldsymbol{\mu}$, degrees of freedom ν, and a correlation matrix $\mathbf{R}$, which is required to be positive definite with 1's on the diagonal. The density function, denoted $\mathcal{L}_{p,\nu}(\cdot \,|\, \boldsymbol{\mu}, \mathbf{R})$, is implicitly defined as the density of

$\mathbf{z} = (z_1, \ldots, z_p)'$, where $z_j = \mu_j + \log[F_\nu(e_j)/\{1 - F_\nu(e_j)\}]$, $F_\nu(\cdot)$ denotes the c.d.f. of a standard univariate t distribution with ν degrees of freedom, and $\mathbf{e} = (e_1, \ldots, e_p)'$ follows a multivariate t distribution with ν degrees of freedom, mean $\mathbf{0}$, and scale matrix $\mathbf{R}$. The generic form of the conventional multivariate t density, including location and scale parameters, is

$$\mathcal{T}_{p,\nu}(\mathbf{t} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2)(\nu\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \right) \times \left\{ 1 + \frac{1}{\nu}(\mathbf{t} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{t} - \boldsymbol{\mu}) \right\}^{-(\nu+p)/2}.$$

The explicit form of our multivariate logistic density is

$$\mathcal{L}_{p,\nu}(\mathbf{z} \mid \boldsymbol{\mu}, \mathbf{R}) = \mathcal{T}_{p,\nu}(\{g_\nu(z_1 - \mu_1), \ldots, g_\nu(z_p - \mu_p)\}' \mid \mathbf{0}, \mathbf{R}) \times \prod_{j=1}^{p} \frac{\mathcal{L}(z_j \mid \mu_j)}{\mathcal{T}_{1,\nu}(g_\nu(z_j - \mu_j) \mid 0, 1)}, \tag{3}$$

where $g_\nu(x) = F_\nu^{-1}(e^x/(1 + e^x))$ and $\mathbf{R}$ is restricted to have 1's on the diagonal. Expression (3) is also the density of $\mathbf{z} = (z_1, \ldots, z_p)'$, where $z_j = \mu_j + \log[F(\phi^{-1}e_j)/(\{1 - F(\phi^{-1}e_j)\}]$, $\phi \sim \Gamma(\nu/2,\ \nu/2)$, and $\mathbf{e} = (e_1, \ldots, e_p)' \sim N_p(\cdot \mid \mathbf{0},\ \mathbf{R})$, with $N_p(\cdot \mid \boldsymbol{\mu},\ \mathbf{R})$ denoting a p-dimensional multivariate normal distribution.

Density (3) has the correlation structure of a scale mixture of normals, with univariate logistic marginals. In the case where $p = 1$, it is clear that expression (3) reduces to the form in (2). For $p > 1$, the marginal univariate densities of z_j, for $j = 1, \ldots, p$, have univariate logistic, $\mathcal{L}(z_j \mid \mu_j)$, forms. By choosing nondiagonal $\mathbf{R}$, the density can account for general dependency structures.

A useful feature of the proposed multivariate logistic distribution is that the two densities $\mathcal{L}_\nu(\cdot \mid \boldsymbol{\mu}, \mathbf{R})$ and $\mathcal{T}_\nu(\cdot \mid \boldsymbol{\mu}, \sigma^2\mathbf{R})$ closely approximate one another when ν and σ^2 are chosen appropriately. This fact allows us to construct a highly efficient importance sampling algorithm that can be used to estimate the exact posterior distribution. As described subsequently, an approximation to the posterior distribution can be derived by substituting an appropriate multivariate t distribution for the multivariate logistic. A Gibbs sampler is used to generate samples from the approximate posterior, and we correct for the approximation by assigning importance weights to the samples when computing posterior summaries. The close correspondence between the logistic and t distribution in the univariate case was noted by Albert and Chib (1993). To make the approximation almost exact, we set $\sigma^2 = \tilde{\sigma}^2 \equiv \pi^2(\nu - 2)/3\nu$ (a value chosen to make the variances of the univariate t and logistic distributions equal) and set $\nu = \tilde{\nu} \equiv 7.3$ (a value chosen to minimize the integrated squared distance between the univariate t and univariate logistic densities). In actual data analysis, the performance of this approximation is reflected by the low coefficient of variation of sampling weights and the fact that weighting or not weighting the sample makes little difference in the results. In the sequel, we drop the parameter ν and use the notation $\mathcal{L}_p(\cdot \mid \boldsymbol{\mu}, \mathbf{R})$ to denote a p-dimensional multivariate logistic distribution with $\tilde{\nu} \equiv 7.3$ degrees of freedom. We use either $\mathcal{L}_1(\cdot \mid \mu, 1)$ or $\mathcal{L}(\cdot \mid \mu)$ to denote the univariate logistic density defined in (2).

3. Multiple Categorical Outcomes

When $\mathbf{y}_i$ consists of multiple binary outcomes, the likelihood under our model is

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{R}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{p} 1(z_{ij} > 0)^{y_{ij}} 1(z_{ij} \le 0)^{1-y_{ij}} \right\} \times \mathcal{L}_p(\mathbf{z}_i \mid \mathbf{X}_i\boldsymbol{\beta}, \mathbf{R})\, d\mathbf{z}_i. \tag{4}$$

Although there is no simple closed form expression for the correlation between y_{ij} and y_{ik}, following a similar strategy to that used in probit analysis, we can base inferences instead on ρ_{jk}, the element in the jth row and kth column of the correlation matrix $\mathbf{R}$. The elements of $\mathbf{R}$ measure dependencies among the elements of $\mathbf{y}_i$ through correlations in the underlying variables, $\mathbf{z}_i$.

Generalizing the likelihood to the case where the outcomes are polychotomous, so that $y_{ij} \in \{1, \ldots, d\}$ for $j = 1, \ldots, p$, we have

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i \mid \mathbf{X}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R}) = \int \cdots \int \left\{ \prod_{j=1}^{p} \sum_{k=1}^{d} 1(\alpha_{y_{ij}-1} < z_{ij} \le \alpha_{y_{ij}}) \right\} \times \mathcal{L}_p(\mathbf{z}_i \mid \mathbf{X}_i\boldsymbol{\beta}, \mathbf{R})\, d\mathbf{z}_i, \tag{5}$$

where $\boldsymbol{\alpha} = (\alpha_0,\ \alpha_1, \ldots, \alpha_{d-1},\ \alpha_d)'$ are threshold parameters, with

$$-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{d-1} < \alpha_d = \infty.$$

This likelihood implies the following cumulative logistic model for the jth ordinal outcome:

$$\text{logit}\ \Pr(Y_{ij} \le k \mid \mathbf{X}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_k - \mathbf{x}_{ij}'\boldsymbol{\beta}. \tag{6}$$

Thus, unlike commonly used logistic-normal random effects models for ordered categorical data, our model results in a marginal logistic regression model for each of the outcomes. This formulation implicitly assumes that the data consist of measurements of the same type of outcome for different subunits nested within subject i. By allowing the threshold parameters to vary with j, one can allow the data to consist of different types of categorical outcomes, potentially having different numbers of categories.

4. Prior Specification and Posterior Propriety

A Bayesian specification of the model is completed by assigning priors for the parameters. Although one could add flexibility to the model by placing a prior on the degrees of freedom ν, we treat ν as a fixed constant ($\equiv$7.3) rather than a parameter for reasons discussed above. For simplicity, we assume the prior density can be expressed as $\pi(\boldsymbol{\beta},\ \mathbf{R},\ \boldsymbol{\alpha}) = \pi(\boldsymbol{\beta})\pi(\mathbf{R})\pi(\boldsymbol{\alpha})$. For $\boldsymbol{\beta}$, we will either assign a normal prior density $\pi(\boldsymbol{\beta}) = \mathrm{N}_q(\boldsymbol{\beta};\ \boldsymbol{\beta}_0,\ \boldsymbol{\Sigma}_\beta)$, or else assign an improper uniform prior, $\pi(\boldsymbol{\beta}) \propto 1$. While any prior for $\mathbf{R}$ can be accommodated, a simple default prior for $\mathbf{R}$ is the uniform density with support on the space of correlation matrices. Finally, for the threshold parameters, we choose an improper uniform prior, $\pi(\boldsymbol{\alpha}) \propto 1(\alpha_1 < \alpha_2 < \cdots < \alpha_{d-1})$.

The impropriety of the posterior distribution when uniform improper priors are chosen for parameters in mixed effects

binary regression models is a vexing problem that has prompted a considerable amount of research (Natarajan and McCulloch, 1995, 1998; Natarajan and Kass, 2000; Natarajan, 2001; Sun, Tsutakawa, and He, 2001). An attractive feature of the method proposed here is that uniform improper priors result in proper posterior distributions under conditions that are simple to verify in practice. A relatively simple technique for verifying propriety involves discarding all but a single outcome per subject leaving a reduced data set consisting of independent outcomes, say $\mathbf{y}_{\text{reduced}}$. If one can verify that $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R} | \mathbf{y}_{\text{reduced}})$ is proper, then clearly so is $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})$. Since the correlation parameter $\mathbf{R}$ does not appear in the reduced data likelihood, it follows that $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y}_{\text{reduced}}) = \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, | \,\mathbf{y}_{\text{reduced}})\pi(\mathbf{R})$, where $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, | \,\mathbf{y}_{\text{reduced}})$ is the posterior density obtained by fitting a univariate logistic regression model to $\mathbf{y}_{\text{reduced}}$ using the original q predictor variables and specifying an improper uniform prior distribution $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \propto 1(\alpha_1 < \alpha_2 < \cdots < \alpha_{d-1})$. Since the support of $\pi(\mathbf{R})$ is finite, it follows that $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y}_{\text{reduced}})$ is proper if and only if $\iint \pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \,|\, \mathbf{y}_{\text{reduced}})\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta} < \infty$. Sufficient conditions for the propriety of $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta} \,|\, \mathbf{y}_{\text{reduced}})$ under improper uniform priors are set forth in Chen and Shao (1999, 2000). In the special case of binary logistic regression, a necessary and sufficient condition for propriety is that the regression coefficients have a unique, finite MLE. This condition is easy to verify in practice, since programs like `SAS PROC LOGISTIC` automatically check for existence of the MLE.

5. Posterior Computation

Focusing on the multiple binary outcomes case, the joint posterior of $\boldsymbol{\beta}$ and $\mathbf{R}$, given a random sample of n observations, $\mathbf{y}_1, \ldots, \mathbf{y}_n$, can be expressed as

$$\pi(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y}) \propto \pi(\boldsymbol{\beta}, \mathbf{R})\pi(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R}), \tag{7}$$

where $\pi(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R}) = \prod_{i=1}^{n} \pi(\mathbf{y}_i \,|\, \boldsymbol{\beta}, \mathbf{R})$ is the likelihood function, $\pi(\boldsymbol{\beta}, \mathbf{R})$ is the prior density on the parameters, and $\pi(\mathbf{y}_i \,|\, \boldsymbol{\beta}, \mathbf{R})$ is the likelihood contribution of a single subject as defined in (4). Due to the complexity of the likelihood, efficient posterior computation is challenging. However, if one has an MCMC algorithm for sampling efficiently from

$$\pi^*(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y}) \propto \pi(\boldsymbol{\beta}, \mathbf{R})\pi^*(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R}), \tag{8}$$

where $\pi^*(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R})$ denotes the likelihood under an alternative model which approximates (4), then approximate inferences about $\pi(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})$ can be based on $\pi^*(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})$. Alternatively, with slightly more effort, the exact posterior $\pi(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})$ can be estimated to any desired level of precision by sampling from $\pi^*(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})$ and then assigning appropriate importance weights (Hastings, 1970, Section 2.5). When $\mathbf{y}_i$ consists of multiple binary outcomes, an excellent choice for $\pi^*(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R})$ is the t-link model,

$$\begin{aligned} &\Pr(\mathbf{Y}_i = \mathbf{y}_i \,|\, \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{R}) \\ &\quad \approx \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{p} 1(z_{ij} > 0)^{y_{ij}} 1(z_{ij} \le 0)^{1-y_{ij}} \right\} \\ &\qquad \times \mathcal{T}_{p,\tilde{\nu}}(\mathbf{z}_i \,|\, \mathbf{X}_i\boldsymbol{\beta}, \tilde{\sigma}^2\mathbf{R})\, d\mathbf{z}_i, \end{aligned} \tag{9}$$

where $\tilde{\nu}$ and $\tilde{\sigma}^2$ are constants defined in Section 2. Posterior computation under this approximation can be accomplished using an efficient MCMC algorithm, as described in the next section.

5.1 *MCMC Algorithm Using Approximation to Likelihood*

The proposed MCMC algorithm was motivated by the data augmentation algorithm of Albert and Chib (1993) and is also related to the algorithm of Chib and Greenberg (1998). First, note that the likelihood (9) can be equivalently specified as follows:

$$\begin{aligned} y_{ij} &= 1(z_{ij} > 0), \\ \mathbf{z}_i \,|\, \boldsymbol{\beta}, \mathbf{R}, \phi_i &\sim \mathrm{N}_p\big(\mathbf{X}_i\boldsymbol{\beta}, \tilde{\sigma}^2\phi_i^{-1}\mathbf{R}\big), \\ \phi_i \,|\, \boldsymbol{\beta}, \mathbf{R} &\sim \Gamma\left(\frac{\tilde{\nu}}{2}, \frac{\tilde{\nu}}{2}\right). \end{aligned} \tag{10}$$

Our algorithm for posterior computation focuses on the approximate joint posterior distribution of the parameters and latent variables, which can be expressed as follows:

$$\begin{aligned} &\pi^*(\boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z} \,|\, \mathbf{y}) \\ &\quad \propto \pi(\boldsymbol{\beta})\pi(\mathbf{R})\pi(\boldsymbol{\phi})\pi^*(\mathbf{z} \,|\, \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi})\pi(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z}), \end{aligned} \tag{11}$$

where $\pi(\boldsymbol{\phi}) = \prod_{i=1}^{n} \Gamma(\phi_i \,|\, \tilde{\nu}/2, \tilde{\nu}/2)$, $\pi^*(\mathbf{z} \,|\, \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}) = \prod_{i=1}^{n} \times \mathrm{N}_p(\mathbf{z}_i \,|\, \mathbf{X}_i\boldsymbol{\beta}, \tilde{\sigma}^2\phi_i^{-1}\mathbf{R})$, $\pi(\boldsymbol{\beta}) = \mathrm{N}_q(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})\pi(\mathbf{y} \,|\, \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z})$ is a (0/1) indicator function truncating $\mathbf{z}$ to the range implied by $\mathbf{y}$, and $\pi(\mathbf{R})$ is any distribution with support on the space of correlation matrices. Under the approximations, the full conditional distributions of $\boldsymbol{\beta}$, $\mathbf{z}$, and $\boldsymbol{\phi}$ have standard conjugate forms. Computation for the case where $\pi(\boldsymbol{\beta}) \propto 1$ can be handled by choosing the normal prior for $\boldsymbol{\beta}$ and setting $\sum_{\beta}^{-1} = \mathbf{0}$.

Our algorithm alternates between sampling from the full conditionals of $\boldsymbol{\beta}, \mathbf{z}$, and $\boldsymbol{\phi}$, and updating $\mathbf{R}$ using a Metropolis step. In particular, after choosing initial values for $\boldsymbol{\beta}$, $\mathbf{R}$, and $\boldsymbol{\phi}$, the algorithm repeats the following steps (for $t = 1, \ldots, T$):

1. For $i = 1, \ldots, n$, sample $\mathbf{z}_i^{(t)}$ from the full conditional distribution of $\mathbf{z}_i$

$$\mathbf{z}_i^{(t)} \sim \mathrm{N}_p\big(\mathbf{X}_i\boldsymbol{\beta}^{(t-1)}, \tilde{\sigma}^2/\phi_i^{(t-1)}\mathbf{R}^{(t-1)}\big),$$

with z_{ij} truncated above (below) zero if $y_{ij} = 1$ ($y_{ij} = 0$).

2. For $i = 1, \ldots, n$, sample $\phi_i^{(t)}$ from the full conditional distribution of ϕ_i

$$\phi_i^{(t)} \sim \Gamma\left(\frac{\tilde{\nu}+p}{2}, \frac{\tilde{\nu} + \tilde{\sigma}^{-2}\left(\mathbf{z}_i^{(t)} - \mathbf{X}_i\boldsymbol{\beta}^{(t-1)}\right)'(\mathbf{R}^{(t-1)})^{-1}\left(\mathbf{z}_i^{(t)}, - \mathbf{X}_i\boldsymbol{\beta}^{(t-1)}\right)}{2}\right).$$

3. Sample $\boldsymbol{\beta}^{(t)}$ from its full conditional distribution: $\boldsymbol{\beta}^{(t)} \sim \mathrm{N}_q(\widetilde{\boldsymbol{\mu}}_{\beta}, \widetilde{\boldsymbol{\Sigma}}_{\beta})$, where

$$\widetilde{\boldsymbol{\Sigma}}_{\beta} = \left(\boldsymbol{\Sigma}_{\beta}^{-1} + \tilde{\sigma}^{-2}\sum_{i=1}^{n} \phi_i^{(t)}\mathbf{X}_i'(\mathbf{R}^{(t-1)})^{-1}\mathbf{X}_i\right)^{-1},$$

$$\widetilde{\boldsymbol{\mu}}_{\beta} = \widetilde{\boldsymbol{\Sigma}}_{\beta}\left(\boldsymbol{\Sigma}_{\beta}^{-1}\boldsymbol{\beta}_0 + \tilde{\sigma}^{-2}\sum_{i=1}^{n} \phi_i^{(t)}\mathbf{X}_i'(\mathbf{R}^{(t-1)})^{-1}\mathbf{z}_i^{(t)}\right).$$

4. Sample a candidate value for the $p^* = p(p-1)/2$ unique elements of $\mathbf{R}$:

$$\text{unique } \widetilde{\mathbf{R}} \sim \mathrm{N}_{p^*}(\text{unique } \mathbf{R}^{(t-1)}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega}$ is chosen by experimentation to yield a desirable acceptance probability. If $\widetilde{\mathbf{R}}$ is positive definite, set

$\mathbf{R}^{(t)} = \widetilde{\mathbf{R}}$ with probability

$$\min\left\{1, \frac{\pi(\widetilde{\mathbf{R}})\prod_{i=1}^{n} \mathrm{N}_p\left(\mathbf{z}_i^{(t)} \,\middle|\, \mathbf{X}_i\boldsymbol{\beta}^{(t)}, \tilde{\sigma}^2/\phi_i^{(t)}\widetilde{\mathbf{R}}\right)}{\pi(\mathbf{R}^{(t-1)})\prod_{i=1}^{n} \mathrm{N}_p\left(\mathbf{z}_i^{(t)} \,\middle|\, \mathbf{X}_i\boldsymbol{\beta}^{(t)}, \tilde{\sigma}^2/\phi_i^{(t)}\mathbf{R}^{(t-1)}\right)}\right\},$$

and set $\mathbf{R}^{(t)} = \mathbf{R}^{(t-1)}$ otherwise. If $\widetilde{\mathbf{R}}$ is not positive definite, then set $\mathbf{R}^{(t)} = \mathbf{R}^{(t-1)}$.

In this algorithm, the elements of $\mathbf{R}$ are updated jointly in a single random walk Metropolis step. An alternate strategy is to update them one at a time using multiple Metropolis steps, one for each element of $\mathbf{R}$.

5.2 *Using Importance Weights for Exact Inference*

Posterior expectations of functionals of the parameters,

$$\mathrm{E}\{h(\boldsymbol{\beta}, \mathbf{R}) \,|\, \mathbf{y}\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\boldsymbol{\beta}, \mathbf{R})\pi(\boldsymbol{\beta}, \mathbf{R} \,|\, \mathbf{y})\, d\boldsymbol{\beta}\, d\mathbf{R}$$

can be estimated consistently by a weighted sample average of the form

$$\widehat{\mathrm{E}}\{h(\boldsymbol{\beta}, \mathbf{R}) \,|\, \mathbf{y}\} = \frac{\sum_{t=B+1}^{T} w^{(t)} h\left(\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)}\right)}{\sum_{t=B+1}^{T} w^{(t)}}, \tag{12}$$

where B is the burn-in interval of the MCMC algorithm, and $w^{(t)}$ denotes the sampling importance weight at iteration t (Hastings, 1970). Because the stationary distribution of the Markov chain is $\pi^*(\boldsymbol{\beta}, \mathbf{R}, \mathbf{z} \,|\, \mathbf{y})$, not $\pi(\boldsymbol{\beta}, \mathbf{R}, \mathbf{z} \,|\, \mathbf{y})$, the appropriate sampling weight is

$$w^{(t)} \propto \pi\left(\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)}, \mathbf{z}^{(t)} \,\middle|\, \mathbf{y}\right) \Big/ \pi^*\left(\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)}, \mathbf{z}^{(t)} \,\middle|\, \mathbf{y}\right).$$

An equivalent computational formula is

$$w^{(t)} = \frac{\pi\left(\mathbf{z}^{(t)} \,\middle|\, \boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)}\right)}{\pi^*\left(\mathbf{z}^{(t)} \,\middle|\, \boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)}\right)} = \prod_{i=1}^{n}\left(\frac{\mathcal{T}_{p,\tilde{\nu}}\left(\mathbf{e}_i \,\middle|\, \mathbf{0}, \mathbf{R}^{(t)}\right)}{\mathcal{T}_{p,\tilde{\nu}}\left(\mathbf{z}_i \,\middle|\, x_{ij}\boldsymbol{\beta}^{(t)}, \tilde{\sigma}^2\mathbf{R}^{(t)}\right)}\right)\prod_{j=1}^{p}\left(\frac{\mathcal{L}\left(z_{ij} \,\middle|\, x_{ij}\boldsymbol{\beta}^{(t)}\right)}{\mathcal{T}_{1,\tilde{\nu}}(e_{ij} \,|\, 0, 1)}\right), \tag{13}$$

where $e_{ij} = g_{\tilde{\nu}}(z_{ij} - x_{ij}\boldsymbol{\beta}^{(t)}) = F_{\tilde{\nu}}^{-1}(e^{z_{ij} - x_{ij}\boldsymbol{\beta}^{(t)}}/\{1 + e^{z_{ij} - x_{ij}\boldsymbol{\beta}^{(t)}}\})$ is the transformation defined in expression (3) and $\mathbf{e}_i = (e_{i1}, \ldots, e_{ip})'$. Posterior means, moments, and percentiles can all be computed using expression (12) with an appropriate choice of $h(\cdot)$.

5.3 *Modifications for Ordered Categorical Response Variables*

Suppose $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$, $i = 1, \ldots, n$, are vectors of ordered categorical variables arising from the cumulative logistic model (5), where $y_{ij} \in \{1, \ldots, d\}$. The MCMC sampling approach described in Section 4.1 can be generalized to this situation by incorporating a Step 5 for updating the threshold parameters, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)'$, by sampling from their uniform full conditional distributions, and by appropriately modifying the conditional distribution in Step 1. In particular, in Step 1, $\mathbf{z}_i^{(t)}$ is drawn from the same normal distribution, but $z_{ij}^{(t)}$ is truncated to fall between $\alpha_{y_{ij}-1}$ and $\alpha_{y_{ij}}$. Steps 2–4 proceed without modification. To update $\boldsymbol{\alpha}$, in Step 5:

5*. For $k = 1, \ldots, d-1$ sample α_k from a uniform distribution on the interval $[\max\{z_{ij}^{(t)} : y_{ij} = k\},\ \min\{z_{ij}^{(t)} : y_{ij} = k+1\}]$.

The importance sampling weights are as shown in expression (13).

6. Application

In this section, we return to the neurotoxicology study of MXC (Chapin et al., 1997) that was briefly described in Section 1. Pregnant dams were exposed to MXC at 0 mg/kg, 5 mg/kg ($n = 11$), ($n = 10$), 50 mg/kg ($n = 10$), or 150 mg/kg ($n = 8$) daily for the week before and the week after birth. The pups were then dosed directly at the same dose as the mother from postnatal day 7 to postnatal day 21. One male and one female pup from each litter were tested at each of three ages (day 31, day 47, and day 66) using a battery of neurobehavioral items measured on different scales. Due to missing data at the highest dose level, three of the litters are incomplete and yield data only for a single pup.

For ease of exposition, we have restricted attention to a single binary outcome variable, high motor activity, coded as 1 if the pup exhibited an elevated amount of motor activity compared to the control mean, and 0 otherwise. Thus, for litters with complete data, $\mathbf{y}_i$ is a 6×1 vector obtained by concatenating three binary observations from each of two pups in the ith litter. The marginal probability of elevated activity is modeled as $\operatorname{logit}\Pr(y_{ij} = 1 \,|\, \mathbf{x}_{ij}, \boldsymbol{\beta}) = \mathbf{x}_{ij}'\boldsymbol{\beta}$, where $\mathbf{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 1(\text{test time} = 47) + \beta_2 1(\text{test time} = 66) + \beta_3 1(\text{dose} = 5\text{ mg/kg}) + \beta_4 1(\text{dose} = 50\text{ mg/kg}) + \beta_5 1(\text{dose} = 150\text{ mg/kg}) + \beta_6 1(\text{male pup})$.

Our analysis accounts for two sources of correlation between outcome variables: (i) dependence between different pups in the same litter and (ii) dependence between repeated observations on the same pup nested within a litter. To obtain a minimally structured correlation matrix, we assume that pups within a litter are exchangeable. Hence, the 6×6 matrix $\mathbf{R}$ can be patterned as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_2 \\ \mathbf{R}_2 & \mathbf{R}_1 \end{pmatrix}, \tag{14}$$

where $\mathbf{R}_1$ is a symmetric 3×3 correlation matrix and $\mathbf{R}_2$ is also symmetric. The ijth element of $\mathbf{R}_1$ describes the correlation between the ith and jth longitudinal observation on the same pup. The ijth element of $\mathbf{R}_2$ describes the correlation between the ith and jth observations for two different pups in the same litter. Data from incomplete litters contribute information about $\mathbf{R}_1$ but not $\mathbf{R}_2$.

A noninformative prior distribution was specified for the parameters by setting $\pi(\boldsymbol{\beta}, \mathbf{R}) \propto 1$. Posterior distributions of the parameters were estimated using the MCMC algorithm described in Section 4. To accommodate incomplete litters, we modified Step 1 of the Gibbs sampling algorithm as follows. The latent variable $\mathbf{z}_i$ is drawn from the multivariate normal

Table 1
Multivariate logistic regression analysis of activity in pups exposed to MXC

Covariate		Mean (95% CI†)	OR (95% CI†)	$\Pr(\beta_j < 0)$
Intercept	β_0	0.36 (−0.41, 1.13)		
Day 47	β_1	−1.05 (−1.8, −0.33)	0.35 (0.16, 0.72)	0.998
Day 66	β_2	−0.42 (−1.19, 0.33)	0.65 (0.3, 1.4)	0.862
Dose = 5 mg/kg	β_3	0.66 (−0.18, 1.53)	1.94 (0.83, 4.6)	0.062
Dose = 50 mg/kg	β_4	1.94 (1.03, 2.92)	6.94 (2.79, 18.57)	<0.001
Dose = 150 mg/kg	β_5	1.05 (0.05, 2.1)	2.87 (1.05, 8.19)	0.020
Male	β_6	−1.06 (−1.68, −0.46)	0.35 (0.19, 0.63)	>0.999

†CI = Credible interval.

distribution shown in Step 1, but the truncation bounds are only applied to z_{ij} if y_{ij} is observed. The Markov chain was initialized with all of the regression coefficients, except the intercept, equal to zero. The first 10,000 observations were discarded as burn-in and an additional 1 million iterations were used to compute posterior summaries.

Autocorrelation for the parameters tended to be low, suggesting high sampling efficiency. Lag-10 autocorrelations for the regression parameters, $\boldsymbol{\beta}$, ranged from 0.006 to 0.035. Autocorrelations for the correlation parameters were somewhat higher, as these were updated in random walk steps, but the lag-50 autocorrelations were still less than 0.30. The sampling weights were relatively constant (coefficient of variation = 0.39) and centered close to one (mean = 1.02, median = 1.02) suggesting that the multivariate t distribution closely approximates the multivariate logistic. There was no evidence of lack of convergence based on standard diagnostic tests and examination of trace plots.

Table 1 presents posterior summaries of the regression coefficients. The summaries indicate clear differences in activity level with respect to gender (males are less active than females) and age (activity was highest at day 31 and lowest at 47). There are also strong dose-related differences in activity level, with exposed pups having an increased activity level relative to the unexposed controls. In particular, activity level increases with dose in the first three treatment groups, with a downturn at the 150 mg/kg dose level. At low dose levels, MXC appears to cause the animals to become agitated, leading to abnormally high activity. As dose increases, the animals may become sick, leading to a downturn in the activity level in the high dose group. An advantage of the Bayesian outlook compared to frequentist analysis is that the significance of the downturn can be assessed in terms of posterior probability. Since the posterior probability of a downturn at the highest dose level is 0.95 (data not shown), there is strong posterior belief that the downturn is real. An advantage of our marginal logistic analysis, compared to a mixed effects regression analysis, is that we can summarize the results in terms of marginal odds ratios, as shown in Table 1.

Table 2 presents posterior summaries of the correlation matrices of the three longitudinal measurements made on the same pup and for two different pups in the same litter. The analysis reveals relatively low correlations, with all 95% credible intervals containing zero. An advantage of our method compared to other likelihood-based multivariate logistic models is that we are able to present a very simple description of the association between dependent observations. Model specification via multiway odds ratios would be considerably more complicated, with as many as 57 parameters appearing in the association model.

To assess the sensitivity of our results to the choice of prior distribution, the analysis was repeated by choosing an entirely unstructured correlation matrix and placing a uniform prior distribution on the unstructured correlation model parameters. Unlike the previous analysis, which assumes that the correlation structure is identical for male and female rats, this prior specification allows the correlation parameters to differ by sex. Despite the added flexibility of the correlation model, the resulting posterior distribution does not reveal any clear gender differences in the correlation parameters, with all of the 95% credible intervals including 0.0. Inferences regarding the regression coefficients are also insensitive to the choice of correlation structure, with all of the posterior means differing by less than 0.05 (posterior means: $\beta_0 = -0.34$, $\beta_1 = 1.03$, $\beta_2 = 0.41$, $\beta_3 = -0.66$, $\beta_4 = -1.90$, $\beta_5 = -1.03$, $\beta_6 = 1.05$).

Table 2
Posterior summary of correlation matrices for measurements on days 31, 47, *and* 66

	Posterior mean (95% CI)					
	Same pup			Different pup		
	Day 31	Day 47	Day 66	Day 31	Day 47	Day 66
Day 31	1.00 (1.00, 1.00)	0.08 (−0.12, 0.28)	−0.03 (−0.23,0.17)	0.07 (−0.21,0.34)	−0.01 (−0.21,0.20)	0.03 (−0.18, 0.23)
Day 47		1.00 (1.00, 1.00)	0.16 (−0.05, 0.36)		0.08 (−0.22, 0.36)	0.05 (−0.17, 0.26)
Day 66			1.00 (1.00, 1.00)			0.16 (−0.13, 0.44)

To further investigate the effect of the model structure and prior distribution, the analysis was also repeated by setting $\mathbf{R}_2 = \gamma \mathbf{R}_1$ in equation (14), and using the following two prior specifications: $\pi(\boldsymbol{\beta}, \gamma, \mathbf{R}_1) \propto 1$ and $\pi(\boldsymbol{\beta}, \gamma, \mathbf{R}_1) \propto |\mathbf{R}_1|^{-(3+1)/2}$. Regression estimates for both analyses were similar to Table 1, with posterior means of regression coefficients differing by less than 0.05. For both analyses, the posterior mean of γ was 0.11, and the 95% credible interval included zero, indicating that the correlation between observations on two different pups within the same litter is substantially less than the correlation between observations made on the same pup. Finally, the analysis was repeated by choosing a multivariate normal prior for the regression coefficients with $\beta_j \overset{\text{ind}}{\sim} N(0, 4)$ for $j = 0, \ldots, p$. Results of this analysis were virtually identical to Tables 1 and 2.

To compare our results to frequentist quasi-likelihood estimation, we also fit an identical marginal logistic model using the GEEs methodology as implemented in `SAS PROC GENMOD`. Point estimates for the regression coefficients varied somewhat depending on the choice of working correlation matrix but were similar to the Bayesian posterior means presented in Table 1, especially when an independence working correlation matrix was specified ($\hat{\beta}_0 = -0.36, \hat{\beta}_1 = 1.02, \hat{\beta}_2 = 0.39, \hat{\beta}_3 = -0.62, \hat{\beta}_4 = -1.82, \hat{\beta}_5 = -0.98, \hat{\beta}_6 = 1.01$). Within-rat correlation estimates ranged from -0.12 to 0.55 when the working correlation structure was specified by treating individual rats as an unstructured subcluster nested within litters. These correlation estimates are not directly comparable to the results presented in Table 2, since the Bayesian estimates pertain to unobserved continuous variables as opposed to observed binary variables. Unlike the Bayesian approach, which produces exact credible intervals as a by-product of the MCMC computations, `SAS PROC GENMOD` provides only point estimates for the parameters of the working correlation matrix. In addition, since the sample sizes are small in the MXC application, our exact Bayesian approach seems more defensible than methods that rely on asymptotic arguments for inference.

7. Discussion

This article has proposed an approach for Bayesian multivariate logistic regression based on a new multivariate logistic density, which is structured to facilitate posterior computation. In particular, a modification of the data augmentation algorithm of Albert and Chib (1993) can be used to obtain samples from an approximation to the posterior, and these samples are then assigned importance weights in constructing summaries of the exact posterior. This algorithm is easy to program and has high efficiency in cases we have considered.

Mixed effects logistic regression is currently one of the most commonly used approaches for the analysis of multivariate categorical data. Due to difficulties in routinely implementing maximum likelihood-based methods and in generalizing these methods to more complex settings (variable selection, mixed discrete and continuous outcomes, multilevel data structures, joint modeling with survival times, etc.), Bayesian approaches have become increasingly popular. This popularity is largely due to the availability of the WinBUGS software (Lunn, Thomas, and Spiegelhalter, 2000), which can implement Gibbs sampling for a wide variety of Bayesian models.

A major drawback of the Bayesian approach to mixed effects logistic regression is poor performance when diffuse or improper priors are chosen. These problems arise due to impropriety of the posterior when uniform improper priors are chosen, and to near impropriety when diffuse but proper priors are chosen. In addition to having a marginal logistic interpretation, which logistic-normal mixed effects models lack, a major advantage of our approach is that the posterior is proper under mild regularity conditions. In addition, as we have demonstrated in the application, our algorithm for posterior computation is efficient even when uniform improper priors are chosen.

Currently, probit models are very widely used in complex applications involving categorical or discrete-time survival data, due to the ease of modeling and computation using the underlying normal framework. Our approach uses the appealing features of the underlying normal framework to develop methods for analysis of logistic models. Logistic models are much easier to interpret than probit models, particularly for biomedical researchers who routinely use logistic regression in analyses of univariate outcomes. As it is straightforward to generalize our approach to essentially any data structure in which a probit model has been used, the proposed methodology should prove widely useful.

Résumé

Les analyses bayésiennes de résultats multivariés binaires ou qualitatifs reposent sur des modèles de régression probit ou logistique à effets mixtes, n'ayant pas de structure marginale logistique pour les résultats individuels. De plus on se heurte à des difficultés lorsqu'on choisit des a-priori non informatifs simples pour les paramètres de covariance. Ces questions nous ont conduit à proposer un nouveau type de distribution logistique multivariée pouvant être utilisée à la construction d'une vraisemblance pour l'analyse de données binaires et qualitatives par régression logistique multivariée. Le modèle possède une structure marginale logistique pour les résultats individuels, simplifiant l'interprétation. Nous suivons une approche bayésienne pour l'estimation et l'inférence, en développant un algorithme efficient d'accroissement de données pour le calcul des a-posteriori. La méthode est illustrée par une application à une étude de neurotoxicologie.

References

Carey, V. and Zeger, S. L. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80,** 517–526.

Castillo, E., Sarabia, J. M., and Hadi, A. S. (1997). Fitting continuous bivariate distributions to data. *The Statistician* **46,** 355–369.

Chapin, R. E., Harris, M. W., Davis, B. J., et al. (1997). The effects of perinatal/juvenile methoxychlor exposure on adult rat nervous, immune, and reproductive system function. *Fundamental and Applied Toxicology* **40,** 138–157.

Chen, M.-H. and Dey, D. K. (2000). Bayesian analysis for correlated ordinal data models. In *Generalized Linear Models: A Bayesian Perspective*, D. K. Dey, S. K. Gosh, and B. K. Mallick (eds). New York: Marcel Dekker.

Chen, M.-H. and Shao, Q. M. (1999). Existence of Bayesian estimates for the polychotomous quantal response models. *Annals of the Institute of Statistics and Mathematics* **51,** 637–656.

Chen, M.-H. and Shao, Q. M. (2000). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society* **129,** 293–302.

Chib, S. (2000). Bayesian methods for correlated binary data. In *Generalized Linear Models: A Bayesian Perspective*, D. K. Dey, S. K. Ghosh, and B. K. Mallick (eds). New York: Marcel Dekker.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85,** 347–361.

Dunson, D. B., Chen, Z., and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59,** 521–530.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood based method for analysing longitudinal binary responses. *Biometrika* **80,** 141–151.

Glonek, G. F. V. and McCullagh, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* **57,** 533–546.

Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association* **56,** 335–349.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57,** 97–109.

Lipsitz, S., Laird, N., and Harrington D. (1991). Generalized estimating equations for correlated binary data: Using odds ratios as a measure of association. *Biometrika* **78,** 153–160.

Lunn, D. J., Thomas, A., and Spiegelhalter, D. (2000). WinBUGS—A Bayesian modeling framework: Concepts, structure and extensibility. *Statistics and Computing* **10,** 325–337.

Malik, H. J. and Abraham, B. (1973). Multivariate logistic distributions. *The Annals of Statistics* **1,** 588–590.

Natarajan, R. (2001). On the propriety of a modified Jeffreys's prior for variance components in binary random effects models. *Statistics & Probability Letters* **51,** 409–414.

Natarajan, R. and McCulloch, C. E. (1995). A note on the existence of the posterior distribution for a class of mixed models for binomial responses. *Biometrika* **82,** 639–643.

Natarajan, R. and McCulloch, C. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* **7,** 267–277.

Natarajan, R. and Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95,** 227–237.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.

Qu, Y., Piedmonte, M. R., and Medendorp, S. V. (1995). Latent variable models for clustered ordinal data. *Biometrics* **51,** 268–275.

Stiratelli, R., Laird, N., and Ware, J. (1984). Random effects models for serial observations with binary responses. *Biometrics* **40,** 961–970.

Sun, D., Tsutakawa, R. K., and He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica* **11,** 77–95.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.

Received March 2003. *Revised February* 2004.
Accepted March 2004.