

# 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, NC 27709

\**email*: obrien4@niehs.nih.gov

**SUMMARY.** Bayesian analyses of multivariate binary or categorical outcomes typically rely on probit or mixed effects logistic regression models which do not have a marginal logistic structure for the individual outcomes. In addition, difficulties arise when simple non-informative 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. Propriety results are provided under uniform improper priors, and the method is illustrated with application to a study of fetal growth restriction in twins.

**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 categorical outcome data. Publications in these areas report estimated odds ratios obtained from fitting of a logistic regression model as a standard component of the results. In many cases, data are multivariate or correlated (e.g., due to repeated observations on a study subject or for subjects within centers) and it is appealing to have a model that maintains a marginal logistic regression interpretation for the individual outcomes. Commonly used logistic random effects models (cf., Stiratelli, Laird and Ware, 1984) do not have this property, since the logistic structure is lost in integrating out the random effects (cf., Larsen et al, 2000). An alternative is to use a marginal analysis that avoids complete specification of the likelihood (cf., Zeger and Liang, 1986; Prentice, 1988; Lipsitz, Laird and Harrington, 1991; Carey and Zeger, 1993). In developing a likelihood for Bayesian inference, a useful approach is to specify the probability model for the categorical outcomes in terms of underlying continuous variables.

To illustrate this underlying variable specification, first consider the univariate logistic regression model:

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

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  $z_i$  is an underlying continuous variable that has a logistic density. Using this latent variable specification, one can generalize the model to the multivariate case, where  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ , by letting  $y_{ij} = 1(z_{ij} > 0)$  and choosing a multivariate logistic distribution for  $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})'$ .

Most likelihood-based methods avoid specifying a continuous multivariate logistic density, instead parameterizing correlations on the discrete scale (Fitzmaurice and Laird, 1993;

Glonek and McCullagh, 1995). One motivation for such an approach is the lack of flexible multivariate logistic distributions. (Refer to Gumbel, 1961; Malik and Abraham (1973); Castillo, Sarabia and Hadi, 1997). This article proposes a new multivariate logistic density, derived by transforming variables that follow a multivariate  $t$  distribution. The resulting logistic density is closely approximated by a multivariate  $t$  distribution, has an unrestricted correlation structure, and has properties that facilitate efficient computation.

Although the density can be used in constructing a likelihood for frequentist inference, this article focuses on Bayesian approaches. In Bayesian analyses of categorical data, probit models are widely used, motivated largely by the simplicity of posterior computation using the data augmentation algorithm of Albert and Chib (1993). In particular, after augmenting the data with normal underlying variables, computation can proceed using Gibbs samplers for normal linear models. This approach has been generalized for multivariate probit analysis (Chib and Greenberg, 1998), for random effects modeling of binary data (Chib, 2000), and for analysis of correlated ordinal (Chen and Dey, 2000), discrete-time survival (Albert and Chib, 2001), and mixed discrete and continuous (Dunson, 2000; Dunson, Chen and Harry, 2003) outcomes. Chen and Dey (2000) considered a rich class of scale mixtures of multivariate normal underlying variables.

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.

In generalizing the logistic regression model to account for multivariate categorical outcomes, normally distributed random effects are commonly incorporated (Diggle, Liang and Zeger, 1994; Stiratelli et al., 1984). Posterior computation can then proceed via a Gibbs

sampler that relies on adaptive rejection sampling (Gilks and Wild, 1992). A well known and vexing problem with this widely used approach is that the posterior density is improper when noninformative priors are chosen for the variance components and regression coefficients (Natarajan and Kass, 2000). This problem is not solved by using diffuse but proper priors, since the resulting posterior can be close to improper, causing extremely slow mixing of the Markov chain (Natarajan and McCulloch, 1998). A major advantage of the approach proposed in this article is that improper uniform priors can be used in applications where prior information is limited (or as a reference analysis) and the posterior will be proper under mild regularity conditions.

Using the proposed multivariate logistic density, we derive likelihoods for multiple binary and multiple ordinal outcomes. Then, augmenting the observed categorical outcomes with underlying variables, we develop a simple and efficient Markov chain Monte Carlo (MCMC) algorithm (Tierney, 1994) for posterior computation. In particular, we first use a Gibbs sampler to generate draws from an approximation to the posterior, derived using a scale mixture of normals to approximate the multivariate logistic density. Sampling from this approximation is highly efficient, since regression parameters can be updated in blocks from a multivariate normal density. Importance weights are then assigned in calculating summaries of the exact posterior density. Since the approximation is close to exact, the weights tend to be nearly constant.

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 data on fetal growth restriction in twins, and Section 7 discusses the results.

## **2. Multivariate Logistic Distributions**

As noted in Section 1, the univariate logistic regression model (1) can be expressed as  $y_i = 1(z_i > 0)$ , where  $z_i \sim \mathcal{L}(\cdot | \mathbf{x}'_i \boldsymbol{\beta})$  are independent logistically distributed random variables, with

$$\mathcal{L}(z | \mu) = \frac{\exp\{- (z - \mu)\}}{[1 + \exp\{- (z - \mu)\}]^2} \quad (2)$$

denoting the logistic density with location parameter  $\mu$ . We can generalize this underlying variable specification to the case where  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$  is a vector of multiple binary outcomes by letting  $y_{ij} = 1(z_{ij} > 0)$ , where  $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})'$  follows a joint distribution having univariate logistic marginals  $z_{ij} \sim \mathcal{L}(\cdot | \mathbf{x}'_{ij} \boldsymbol{\beta})$ .

As noted previously (Albert and Chib, 1993), an excellent approximation to the univariate logistic density (2) can be obtained using the  $t$  distribution,

$$\mathcal{T}_\nu(z | \mu, \sigma^2) = \frac{\Gamma((\nu + 1)/2)}{\Gamma(\nu/2) \sqrt{\nu\pi}\sigma} \left\{ 1 + \frac{(z - \mu)^2}{\nu\sigma^2} \right\}^{-(\nu+1)/2}, \quad (3)$$

where  $\sigma$  is a scale parameter and  $\nu$  is the degrees of freedom. To closely approximate the univariate logistic distribution, we set  $\sigma^2 = \tilde{\sigma}^2 \equiv \pi^2(\nu - 2)/3\tilde{\nu}$  (a value chosen to make the variances equal) and  $\nu = \tilde{\nu} \equiv 7.3$  (a value chosen to minimize the integrated squared distance between the densities). The two densities are essentially indistinguishable when plotted together, and it may be verified analytically that the  $t$  density dominates the logistic in the tails. As we demonstrate subsequently, one can exploit this approximation to develop efficient algorithms for posterior computation.

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 is:

$$\mathcal{L}_p(\mathbf{z} | \boldsymbol{\mu}, \mathbf{R}) = \mathcal{T}_{p, \tilde{\nu}}(\mathbf{t} | \mathbf{0}, \mathbf{R}) \prod_{j=1}^p \frac{\mathcal{L}(z_j | \mu_j)}{\mathcal{T}_{\tilde{\nu}}(t_j | 0, 1)}, \quad (4)$$

where the conventional multivariate  $t$  density is denoted by

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

$t_j = F^{-1}(e^{z_j}/(e^{z_j} + e^{\mu_j}))$  with  $F^{-1}(\cdot)$  denoting the inverse CDF of the  $\mathcal{T}_{\tilde{\nu}}(0, 1)$  density,  $\mathbf{t} = (t_1, \dots, t_p)'$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ ,  $\mathbf{R}$  is a correlation matrix (i.e., with 1's on the diagonal), and  $\tilde{\nu}$  is the constant defined previously.  $\mathcal{L}_p(\mathbf{z} | \boldsymbol{\mu}, \mathbf{R})$  may be derived as the density of  $(z_1, \dots, z_p)'$  where  $z_j = \mu_j + \log[F(\phi^{-1}t_j)/\{1 - F(\phi^{-1}t_j)\}]$ ,  $\phi \sim \text{Gamma}(\tilde{\nu}/2, \tilde{\nu}/2)$ , and  $\mathbf{t} = (t_1, \dots, t_p)' \sim N_p(\cdot | \mathbf{0}, \mathbf{R})$ , with  $N_p(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$  denoting a  $p$  dimensional multivariate normal distribution.

Density (4) 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 (4) reduces to the form in (2). For  $p > 1$ , we verify in Appendix A that the marginal univariate densities of  $z_j$ , for  $j = 1, \dots, p$ , have univariate logistic,  $\mathcal{L}(z_j | \mu_j)$ , forms. In addition, for  $\mathbf{R} = \text{diag}(1, \dots, 1)$ ,  $\mathcal{L}_p(\boldsymbol{\mu}, \mathbf{R})$  reduces to a product of univariate logistic densities, and the elements of  $\mathbf{z}_i$  are uncorrelated. By choosing non-diagonal  $\mathbf{R}$ , the density can account for general dependency structures.

### 3. Multiple Categorical Outcomes

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

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{R}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^p 1(z_{ij} > 0)^{y_{ij}} 1(z_{ij} \leq 0)^{1-y_{ij}} \right\} \mathcal{L}_p(\mathbf{z}_i | \boldsymbol{\mu}, \mathbf{R}) d\mathbf{z}_i. \quad (5)$$

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  $\rho_{jk}$ , the element in the  $j$ th row and  $k$ th 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  $y_{ij} \in \{1, \dots, d\}$  for  $j = 1, \dots, p$ , we have

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R}) = \int \cdots \int \left\{ \prod_{j=1}^p \sum_{k=1}^d 1(\alpha_{y_{ij}-1} < z_{ij} \leq \alpha_{y_{ij}}) \right\} \mathcal{L}_p(\mathbf{z}_i | \mathbf{X}_i, \boldsymbol{\beta}) d\mathbf{z}_i, \quad (6)$$

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

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

This likelihood implies the following cumulative logistic model for the  $j$ th ordinal outcome:

$$\text{logit Pr}(Y_{ij} \leq k | \mathbf{X}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_k - \mathbf{x}'_{ij}\boldsymbol{\beta}, \quad (7)$$

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. 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  $\pi(\boldsymbol{\beta}) = N_q(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ , 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,  $\pi(\boldsymbol{\alpha}) \propto 1(\alpha_1 < \alpha_2 < \dots < \alpha_{d-1})$ .

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 Kass, 2000; Natarajan and McCulloch, 1995, 1998; Natarajan, 2001; Sun, Tsutakawa and He, 2001). An attractive feature of the method we propose is that uniform improper priors result in proper posterior distributions under mild conditions, which we investigate in this section. In particular, we

derive practically useful sufficient conditions that are easy to verify in practice.

**Theorem 1** *Suppose (i) the data arise from model (5) and (ii) there exists a subset of the response data,  $\mathbf{y}^* = (y_{1j_1}, \dots, y_{n^*j_{n^*}})'$ ,  $n^* \leq n$ ,  $j_i \in \{1, \dots, p\}$ , comprising no more than a single outcome per subject, such that the function*

$$L(\boldsymbol{\beta}; \mathbf{y}^*) = \prod_{i=1}^{n^*} \left( \frac{e^{y_{ij_i} \mathbf{x}'_{ij_i} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{ij_i} \boldsymbol{\beta}}} \right) \quad (8)$$

*has a unique maximum. Under conditions (i) and (ii), the joint posterior distribution resulting from an improper prior of the form  $\pi(\boldsymbol{\beta}, \mathbf{R}) \propto \pi(\mathbf{R})$  is proper, even when  $\pi(\mathbf{R})$  is the uniform density on the space of correlation matrices.*

**Theorem 2** *Suppose (i) the data arise from model (6) and (ii) there exists a subset of the response data,  $\mathbf{y}^* = (y_{1j_1}, \dots, y_{n^*j_{n^*}})'$ ,  $n^* \leq n$ ,  $j_i \in \{1, \dots, p\}$ , comprising no more than a single outcome per subject, such that at least one of the functions*

$$L_1(\boldsymbol{\beta}; \mathbf{y}^*) = \prod_{i=1}^{n^*} \left( \frac{e^{1(y_{ij_i}=1) \mathbf{x}'_{ij_i} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{ij_i} \boldsymbol{\beta}}} \right) \quad (9)$$

$$L_d(\boldsymbol{\beta}; \mathbf{y}^*) = \prod_{i=1}^{n^*} \left( \frac{e^{1(y_{ij_i}=d) \mathbf{x}'_{ij_i} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{ij_i} \boldsymbol{\beta}}} \right) \quad (10)$$

*has a unique maximum. Under conditions (i) and (ii), the joint posterior distribution resulting from an improper prior of the form  $\pi(\boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\alpha}) \propto \pi(\mathbf{R})$  is proper.*

Expressions (8)–(10) are in the form of likelihoods for binary univariate logistic regression models. According to the Theorems, which are proven in Appendix B, the posterior density is proper if a unique finite MLE exists for at least one of these likelihoods. This is easy to verify, since programs like SAS PROC LOGISTIC automatically check for existence of the MLE. In most cases, propriety can be verified by simply choosing  $\mathbf{y}^*$  arbitrarily. For example, to verify propriety for the data set considered in Section 6, we constructed  $\mathbf{y}^*$  by setting  $n^* = n$  and  $j_i = 1$  for all  $i$ . If this approach fails, some guidance about the choice of  $\mathbf{y}^*$  can be gleaned from Silvapulle (1981) and Albert and Anderson (1984).

## 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,  $\mathbf{y}_1, \dots, \mathbf{y}_n$ , can be expressed as

$$\pi(\boldsymbol{\beta}, \mathbf{R}|\mathbf{y}) \propto \pi(\boldsymbol{\beta}, \mathbf{R})\pi(\mathbf{y}|\boldsymbol{\beta}, \mathbf{R}) \quad (11)$$

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 (5). Due to the complexity of the likelihood, efficient posterior computation is challenging. However, if one has a Markov Chain Monte Carlo (MCMC) algorithm for sampling efficiently from

$$\pi^*(\boldsymbol{\beta}, \mathbf{R}|\mathbf{y}) \propto \pi(\boldsymbol{\beta}, \mathbf{R})\pi^*(\mathbf{y}|\boldsymbol{\beta}, \mathbf{R}), \quad (12)$$

where  $\pi^*(\mathbf{y}|\boldsymbol{\beta}, \mathbf{R})$  denotes the likelihood under an alternative model which approximates (5), 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,

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\beta}, \mathbf{R}) \approx \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^p 1(z_{ij} > 0)^{y_{ij}} 1(z_{ij} \leq 0)^{1-y_{ij}} \right\} \mathcal{T}_{p, \tilde{\nu}}(\mathbf{z}_i | \mathbf{X}_i, \boldsymbol{\beta}, \tilde{\sigma}^2 \mathbf{R}) d\mathbf{z}_i. \quad (13)$$

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 (13) can be equivalently specified as follows:

$$\begin{aligned}
y_{ij} &= 1(z_{ij} > 0) \\
\mathbf{z}_i | \boldsymbol{\beta}, \mathbf{R}, \phi_i &\sim N_p(\mathbf{X}_i \boldsymbol{\beta}, \tilde{\sigma}^2 \phi_i^{-1} \mathbf{R}) \\
\phi_i | \boldsymbol{\beta}, \mathbf{R} &\sim \text{Gamma}\left(\frac{\tilde{\nu}}{2}, \frac{\tilde{\nu}}{2}\right).
\end{aligned} \tag{14}$$

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

$$\pi^*(\boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z} | \mathbf{y}) \propto \pi(\boldsymbol{\beta}) \pi(\mathbf{R}) \pi(\boldsymbol{\phi}) \pi^*(\mathbf{z} | \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}) \pi(\mathbf{y} | \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z}) \tag{15}$$

where  $\pi(\boldsymbol{\phi}) = \prod_{i=1}^n \text{Gamma}(\phi_i | \tilde{\nu}/2, \tilde{\nu}/2)$ ,  $\pi^*(\mathbf{z} | \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}) = \prod_{i=1}^n N_p(\mathbf{z}_i | \mathbf{X}_i \boldsymbol{\beta}, \tilde{\sigma}^2 \phi_i^{-1} \mathbf{R})$ ,  $\pi(\boldsymbol{\beta}) = N_q(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$  or  $\pi(\boldsymbol{\beta}) \propto 1$ ,  $\pi(\mathbf{y} | \boldsymbol{\beta}, \mathbf{R}, \boldsymbol{\phi}, \mathbf{z})$  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 to  $\boldsymbol{\beta}$  and setting  $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} = \mathbf{0}$ .

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  $t = 1, \dots, T$ ):

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

$$\mathbf{z}_i^{(t)} \sim N_p\left(\mathbf{X}_i \boldsymbol{\beta}^{(t-1)}, \tilde{\sigma}^2 / \phi_i^{(t-1)} \mathbf{R}^{(t-1)}\right),$$

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

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

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

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

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

$$\tilde{\boldsymbol{\mu}}_\beta = \tilde{\boldsymbol{\Sigma}}_\beta \left( \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\beta}_0 + \tilde{\sigma}^{-2} \sum_{i=1}^n \phi_i^{(t)} \mathbf{X}_i(\mathbf{R}^{(t-1)})^{-1} \mathbf{z}_i^{(t)} \right).$$

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

$$\text{unique } \tilde{\mathbf{R}} \sim N_{p^*}(\text{unique } \mathbf{R}^{(t-1)}, \boldsymbol{\Omega}),$$

where  $\boldsymbol{\Omega}$  is chosen by experimentation to yield a desirable acceptance probability. If

$\tilde{\mathbf{R}}$  is positive definite, set  $\mathbf{R}^{(t)} = \tilde{\mathbf{R}}$  with probability

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

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

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,  $g(\boldsymbol{\beta}, \mathbf{R})$ ,

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

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

$$\hat{E}\{g(\boldsymbol{\beta}, \mathbf{R}) | \mathbf{y}\} = \frac{\sum_{t=B+1}^T w^{(t)} g(\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)})}{\sum_{t=B+1}^T w^{(t)}}, \quad (16)$$

where  $B$  is the burn-in interval of the MCMC algorithm, and  $w^{(t)}$  denotes the sampling importance weight at iteration  $t$  (Hastings, 1970). Since 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

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

An equivalent computational formula is

$$w^{(t)} = \frac{\pi(\mathbf{z}^{(t)}|\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)})}{\pi^*(\mathbf{z}^{(t)}|\boldsymbol{\beta}^{(t)}, \mathbf{R}^{(t)})} = \prod_{i=1}^n \left( \frac{\mathcal{I}_{p,\tilde{\nu}}(\mathbf{t}_i|\mathbf{0}, \mathbf{R}^{(t)})}{\mathcal{I}_{p,\tilde{\nu}}(\mathbf{z}_i|x_{ij}\boldsymbol{\beta}^{(t)}, \tilde{\sigma}^2\mathbf{R}^{(t)})} \right) \prod_{j=1}^p \left( \frac{\mathcal{L}(z_{ij}|x_{ij}\boldsymbol{\beta}^{(t)})}{\mathcal{I}_{1,\tilde{\nu}}(t_{ij}|0)} \right), \quad (17)$$

where  $\mathbf{t}_i = (t_{i1}, \dots, t_{ip})'$  is defined as in expression (4). Posterior means, moments, and percentiles can all be computed using expression (16) with an appropriate choice of  $g(\cdot)$ .

### 5.3 Modifications for Ordered Categorical Response Variables

Suppose  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ ,  $i = 1, \dots, n$ , are vectors of ordered categorical variables arising from the cumulative logistic model (6), where  $y_{ij} \in \{1, \dots, d\}$ . 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,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)'$ , 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, 3, and 4 proceed without modification. To update  $\boldsymbol{\alpha}$ , in Step 5:

- 5\*. For  $k = 2, \dots, d$  sample  $\alpha_k$  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 (17).

## 6. Perinatal epidemiology application

We illustrate the methodology through an analysis of risk factors for fetal growth restriction in twin pregnancies. The data consist of all twin pregnancies ( $n = 584$ ) enrolled in the Collaborative Perinatal Project from 1959 to 1965, as analyzed by Ananth and Preisser (1999). The outcome of interest is small for gestational age (SGA) birth, defined as a birthweight below the 10th percentile for a given gestational age in a reference population. The available covariates include infant gender, maternal age, years of cigarette smoking, weight gain during

pregnancy, gestational age at delivery, and variables relating to obstetric history. Ananth and Preisser analyzed the same data via maximum likelihood using a marginal logistic model for the probability of low birth weight and a pairwise log-odds ratio model for the correlation between twin outcomes. A goal of our analysis is to assess whether our Bayesian approach, which has a fundamentally different model for the correlation structure than proposed by Ananth and Preisser (and others), yields similar estimates for the regression parameters and inferences on the correlation structure when non-informative priors are chosen.

### 6.1 Model specification

Following Ananth and Preisser, the marginal probability of SGA for the  $j$ th infant of the  $i$ th pregnancy is modeled as

$$\text{logit Pr}(y_{ij} = 1 \mid \mathbf{x}_{ij}, \boldsymbol{\beta}) = \mathbf{x}'_{ij}\boldsymbol{\beta},$$

where

$$\begin{aligned} \mathbf{x}'\boldsymbol{\beta} = & \beta_0 + \beta_1 1(\text{female infant}) + \beta_2 1(\text{multiparous, PFD absent, PPT absent}) \\ & + \beta_3 1(\text{PFD present, PPT absent}) + \beta_4 1(\text{PFD absent, PPT present}) \\ & + \beta_5 1(\text{PFD present, PPT present}) + \beta_6 \log(\text{maternal age}) \\ & + \beta_7 \log(\text{weight gain} + 6) + \beta_8 \log(\text{years smoked} + 1) \\ & + \beta_9 (\text{gestational age} - 37) + \beta_{10} (\text{gestational age} - 37)^2 \end{aligned}$$

and the indicator variables PFD and PPT represent prior fetal death and prior preterm delivery, respectively. The joint probabilities of SGA for a twin pair are given by the distribution function of the bivariate logistic distribution (4). Following Ananth and Preisser, we allow the magnitude of the correlation of SGA for twins of the same pregnancy to differ depending on the mother's parity status. To accommodate this extension, the single free correlation parameter in  $\mathbf{R}$ , denoted  $\rho$ , is replaced by  $\rho_i$  where  $\rho_i = r_1$  if subject  $i$  is primiparous and  $\rho_i = r_2$  if subject  $i$  is multiparous.

## 6.2 Posterior computation

Posterior distributions of the parameters were estimated using the Markov chain Monte Carlo algorithm described in Section 4. A non-informative prior distribution for  $(\boldsymbol{\beta}, r_1, r_2)$  was specified by setting  $\pi(\boldsymbol{\beta}) = 1$  and choosing  $\pi(r_1, r_2)$  to be the bivariate uniform distribution on  $(-1, 1) \times (-1, 1)$ . The parameters  $r_1$  and  $r_2$  were updated in separate random walk Metropolis steps, each using a univariate normal proposal density centered at the current value. The variance of the proposal density was chosen to yield an acceptance probability of approximately 45%.

To assess convergence, we ran 3 separate chains with diffuse starting values for  $(r_1, r_2)$  equal to  $(-0.5, -0.5)$ ,  $(-0.5, 0.5)$ , and  $(0.5, 0.5)$  and with all of the regression coefficients, except the intercept, initialized to zero. Two chains were run for 350,000 iterations, and the last chain was run for 1,000,000 iterations. The first 5,000 iterations were discarded as a burn-in. There was no evidence of lack of convergence based on an examination of trace plots and the convergence diagnostic procedure of Gelman and Rubin (1992). Posterior summaries for all three chains were virtually identical. In addition, autocorrelation tended to be low, suggesting high sampling efficiency. Lag-10 autocorrelations for the regression parameters,  $\boldsymbol{\beta}$ , ranged from .002 to 0.03. Autocorrelations for  $r_1$  and  $r_2$  were somewhat higher, as these were updated in random walk steps, but the lag-20 autocorrelations were still less than 0.05. The sampling weights were relatively constant (coefficient of variation = 0.83) and centered close to one (mean = 1.4, median = 1.1) suggesting that the multivariate  $t$  distribution closely approximates the multivariate logistic.

[Table 1 about here.]

## 6.3 Results

Table 1 presents posterior summaries of the regression coefficients, as well as maximum likelihood estimates from the bivariate logistic model of Ananth and Preisser. Female gender,

cigarette smoking, low weight gain, and prior preterm delivery are all associated with an increased risk of SGA. As expected, outcomes for twins are highly correlated. There is some evidence that the level of correlation is higher among multiparous women than among primiparous women, with  $\hat{r}_1 = 0.16$  (95% CI=[-0.16,0.45]) for primiparous women,  $\hat{r}_2 = 0.35$  (95% CI=[0.21,0.48]) for multiparous women, and  $\Pr(r_2 > r_1 | \mathbf{y}) = 86\%$ .

Ananth and Preisser (1999) reached an identical conclusion, though their correlation estimates are based on odds ratios, and so are not directly comparable. Although the two models for the correlation structure are very different, we get nearly identical estimates for the regression coefficients. This is not unexpected, since both models have the same marginal logistic structure, while accommodating within-twin dependency. For this data set, when non-informative priors are chosen, Bayesian- and MLE-based procedures yield identical inferences. However, for studies with smaller sample sizes and prior information, the Bayesian and MLE-based results may differ substantially.

## 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 generalization 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 (mixed discrete and continuous outcomes, multi-level 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.

## REFERENCES

- Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1-10.
- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.

- Albert, J. H. and Chib, S. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics* **57**, 829-836.
- Ananth, C. V. and Preisser, J. S. (1999). Bivariate logistic regression: Modelling the association of small for gestational age births in twin gestations. *Statistics in Medicine* **18**, 2011-2023.
- 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.
- 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. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347-361.
- 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.

- Diggle, P. J., Liang, K-Y., and Zeger, S. L. (1994). *Longitudinal Data Analysis*. New York: Oxford University Press.
- Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62**, 355-366.
- Dunson, D. B., Chen, Z. and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics*, in press.
- Fitzmaurice G. M. and Laird N. M. (1993). A likelihood based method for analysing longitudinal binary responses. *Biometrika* **80**, 141-151.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457-472.
- Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.
- 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.
- Larsen, K., Petersen, J. H., Budtz-Jorgensen, E., and Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics* **56**, 909-914.

- 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 Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* **95**, 227-237.
- 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.
- Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033-1048.
- Silvapulle, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response models. *Journal of the Royal Statistical Society, Series B* **43**, 310-313.

- 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.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**, 1701-1728.
- Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.

## APPENDIX A

*Proof that marginal univariate density is logistic.*

If  $g(z_j; \mu_j) = F_\nu^{-1}(e^{z_j} / (e^{z_j} + e^{\mu_j}))$ , with  $F_\nu^{-1}(\cdot)$  denoting the inverse CDF of the  $\mathcal{T}_{1,\nu}(0, 1)$  density, then the Jacobian of the transformation  $\mathbf{t}_j = \mathbf{g}(\mathbf{z}_j; \boldsymbol{\mu}_j) = (t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_p)'$ , for  $t_h = g(z_h, \mu_h)$ , is equal to

$$|J_j| = \prod_{h \neq j}^p \frac{dz_h}{dt_h} = \prod_{h \neq j}^p \frac{\mathcal{T}_{1,\nu}(t_h)}{\mathcal{L}(z_h | \mu_h)}.$$

It follows that

$$\begin{aligned} & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{L}(\mathbf{z} | \boldsymbol{\mu}, \mathbf{R}) \prod_{h \neq j}^p dz_h \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{T}_{p,\nu}(\mathbf{t} | \mathbf{0}, \mathbf{R}) \left\{ \prod_{h=1}^p \frac{\mathcal{L}(z_h | \mu_h)}{\mathcal{T}_{1,\nu}(t_h | 0, 1)} \right\} \left\{ \prod_{h \neq j}^p \frac{\mathcal{T}_{1,\nu}(z_h | \mu_h)}{\mathcal{L}(t_h | 0, 1)} \right\} \prod_{h \neq j}^p dt_h \\ &= \frac{\mathcal{L}(z_j | \mu_j)}{\mathcal{T}(t_j | 0, j)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{T}_{p,\nu}(\mathbf{t} | \mathbf{0}, \mathbf{R}) \prod_{h \neq j}^p dt_h \\ &= \mathcal{L}(z_j | \mu_j). \end{aligned}$$

## APPENDIX B

### *Proof of Theorems*

**Proof of Theorem 1.** The posterior density has the form

$$\pi(\boldsymbol{\beta}, \mathbf{R}|\mathbf{y}) \propto \pi(\boldsymbol{\beta})\pi(\mathbf{R}) \prod_{i=1}^n \Pr(\mathbf{Y}_i = \mathbf{y}_i|\boldsymbol{\beta}, \mathbf{R})$$

where  $\Pr(\mathbf{Y}_i = \mathbf{y}_i|\boldsymbol{\beta}, \mathbf{R})$  is defined in expression (5). Let  $\mathbf{y}^* = (y_{1j_1}, \dots, y_{n^*j_{n^*}})'$  denote a subset of the data consisting of the  $j_i$ th response for subject  $i$ ,  $j_i \in \{1, \dots, p\}$ , for  $i = 1, \dots, n^*$ ,  $n^* \leq n$ . The likelihood for  $\mathbf{y}^*$  is as follows:

$$L(\boldsymbol{\beta}|\mathbf{y}^*) = \prod_{i=1}^{n^*} \Pr(Y_{ij_i} = y_{ij_i}|\boldsymbol{\beta}, \mathbf{R}) = \prod_{i=1}^{n^*} \left( \frac{e^{y_{ij_i} \mathbf{x}'_{ij_i} \boldsymbol{\beta}}}{1 + e^{\mathbf{x}'_{ij_i} \boldsymbol{\beta}}} \right)$$

which does not depend on  $\mathbf{R}$ . If  $\mathbf{y}^*$  satisfies condition (ii), then  $L(\boldsymbol{\beta}|\mathbf{y}^*)$  is bounded and it follows from Theorems 2.1 and 3.1 in Chen and Shao (2000), that  $M = \int L(\boldsymbol{\beta}|\mathbf{y}^*) d\boldsymbol{\beta} < \infty$ .

The theorem follows since

$$\begin{aligned} \iint \pi(\boldsymbol{\beta}, \mathbf{R}|\mathbf{y}) d\boldsymbol{\beta} d\mathbf{R} &\propto \iint \left( \prod_{i=1}^n \Pr(\mathbf{Y}_i = \mathbf{y}_i|\boldsymbol{\beta}, \mathbf{R}) \right) \pi(\boldsymbol{\beta})\pi(\mathbf{R}) d\boldsymbol{\beta} d\mathbf{R} \\ &= \iint \prod_{i=1}^n \Pr \left( \bigcap_{j=1}^p Y_{ij} = y_{ij} \middle| \boldsymbol{\beta}, \mathbf{R} \right) \pi(\mathbf{R}) d\boldsymbol{\beta} d\mathbf{R} \\ &\leq \iint \left( \prod_{i=1}^{n^*} \Pr(Y_{ij_i} = y_{ij_i}|\boldsymbol{\beta}, \mathbf{R}) \right) \pi(\mathbf{R}) d\boldsymbol{\beta} d\mathbf{R} \\ &= \int L(\boldsymbol{\beta}|\mathbf{y}^*) d\boldsymbol{\beta} \int \pi(\mathbf{R}) d\mathbf{R} \\ &= M < \infty. \end{aligned}$$

**Proof of Theorem 2.** Denote the likelihood for  $\mathbf{y}^*$  as follows:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{y}^*) = \prod_{i=1}^{n^*} \Pr(Y_{ij_i} = y_{ij_i}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R}),$$

where  $\Pr(\mathbf{Y}_i = \mathbf{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{R})$  is defined in expression (6). For  $r \in \{1, d\}$ , let  $z_{ij_i}^{(r)} = 1 - 2 \times 1(y_{ij_i} = r)$ , and define  $\mathbf{X}_r^*$  as the  $n^* \times q$  matrix with rows  $z_{ij_i}^{(r)} \mathbf{x}_{ij_i}$ . If condition (ii) of

Theorem 2 holds, then by Theorem 3.1 of Chen and Shao (2000), there exists a positive vector  $\mathbf{a} = (a_1, \dots, a_{n^*})' \in R^n$ , i.e. each component  $a_i > 0$ , such that  $\mathbf{a}'\mathbf{X}_1^* = \mathbf{0}$  or  $\mathbf{a}'\mathbf{X}_d^* = \mathbf{0}$ . It follows from Proposition (3.1) of Chen and Shao (1999), that  $M = \iint L(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{y}^*)d\boldsymbol{\alpha}d\boldsymbol{\beta} < \infty$ . The theorem follows using an argument similar to the proof of Theorem 1.

Table 1: Bivariate logistic regression analysis of small for gestational age births in twins

Covariate		AP <sup>†</sup>		
		MLE (SE)	Mean (SD)	Posterior Summary OR (95% CI)
Intercept	$\beta_0$	3.10 (1.62)	2.97 (1.63)	
Female infant	$\beta_1$	0.36 (0.17)	0.35 (0.17)	1.43 (1.01–2.01)
Multiparity				
PFD absent, PPT absent	$\beta_2$	-0.36 (0.26)	-0.39 (0.26)	0.68 (0.41–1.13)
PFD present, PPT absent	$\beta_3$	-0.92 (0.38)	-0.97 (0.38)	0.38 (0.18–0.79)
PFD absent, PPT present	$\beta_4$	0.44 (0.33)	0.42 (0.32)	1.53 (0.81–2.87)
PFD present, PPT present	$\beta_5$	0.38 (0.46)	0.45 (0.44)	1.57 (0.66–3.68)
log(maternal age)	$\beta_6$	-1.03 (0.37)	-1.00 (0.47)	0.37 (0.15–0.91)
log(weight gain + 6)	$\beta_7$	-0.47 (0.17)	-0.46 (0.17)	0.63 (0.45–0.88)
log(years smoked + 1)	$\beta_8$	0.26 (0.10)	0.25 (0.09)	1.28 (1.07–1.55)
(Gestational age – 37)	$\beta_9$	0.21 (0.04)	0.21 (0.04)	
(Gestational age – 37) <sup>2</sup>	$\beta_{10}$	0.02 (0.01)	0.02 (0.01)	

<sup>†</sup> taken from Ananth and Preisser (AP; 1999)