Bayesian Multivariate Logistic Regression

Sean M. O'Brien and David B. Dunson

Biostatistics Branch
National Institute of Environmental Health Sciences
Research Triangle Park, NC

Goals

- Brief review of existing methods
- Illustrate some useful computational techniques
  - MCMC importance sampling (Hastings, 1970)
  - Data-augmentation Gibbs algorithm (Albert & Chib, 1993)
  - Methods for sampling from truncated multivariate normal (Devroye, 1989; Geweke, 1989)
  - Metropolis algorithms for sampling correlation matrices (Chib & Greenberg, 1998; Chen & Dey, 1998)

Introduction

- Correlated binary data arise in numerous application
  - Longitudinal studies
  - Cluster-randomized trials
  - Epidemiologic studies of twins
- Approaches to regression analysis of multivariate binary and ordinal categorical data
  - Generalized estimating equations (GEE)
  - Generalized linear mixed models (GLMMs)
- Logistic link is common in health sciences (odds ratios)
- Some approaches that work well for frequentist inference do not work as well in Bayesian context

Two main types of models

1. **Cluster-specific models**. Regression parameters have cluster-specific interpretation. For example,

$$\text{LogitPr}[y_{ij} = 1] = \mathbf{x}'_{ij}\beta + \mathbf{z}'_{ij}\mathbf{b}_i, \quad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

2. **Marginal models**. Regression parameters have population-average (marginal) interpretation (desirable for epidemiologic studies).

$$\text{LogitPr}[y_{ij} = 1] = \mathbf{x}'_{ij}\beta$$

Full likelihood not necessary for frequentist inference – can use GEE. Need a full likelihood for Bayesian inference.

Full Likelihood Approaches to Marginal Models

Multivariate logistic regression

Parameterization via cross-odds ratios (Glonek & McCullagh, 1995)

Let $\bar{y}_{ij} = 1 - y_{ij}$.

$$\text{Logit } \Pr(y_{ij} = 1) = \log \frac{E(y_{ij})}{E(\bar{y}_{ij})} = \mathbf{x}_{ij}\beta_j$$

$$\log \left\{ \frac{E(y_{ij}y_{ih})}{E(\bar{y}_{ij}y_{ih})} \Big/ \frac{E(y_{ij}\bar{y}_{ih})}{E(\bar{y}_{ij}\bar{y}_{ih})} \right\} = \mathbf{x}_{ijh}\theta_{jh}$$

$$\log \frac{\left\{ \frac{E(y_{ij}y_{ih}y_{ik})}{E(\bar{y}_{ij}y_{ih}y_{ik})} \Big/ \frac{E(y_{ij}\bar{y}_{ih}y_{ik})}{E(\bar{y}_{ij}\bar{y}_{ih}y_{ik})} \right\}}{\left\{ \frac{E(y_{ij}y_{ih}\bar{y}_{ik})}{E(\bar{y}_{ij}y_{ih}\bar{y}_{ik})} \Big/ \frac{E(y_{ij}\bar{y}_{ih}\bar{y}_{ik})}{E(\bar{y}_{ij}\bar{y}_{ih}\bar{y}_{ik})} \right\}} = \mathbf{x}_{ijhk}\boldsymbol{\gamma}_{jhk}$$

etc...

Typically impose additional restrictions to simplify model

Full Likelihood Approaches to Marginal Models

Multivariate Probit Models (Chib & Greenberg, 1998)

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mathbf{x}_{ij}\beta + e_{ij}$$

$$\mathbf{e}_i = (e_{i1}, \ldots, e_{ip})' \sim N(\mathbf{0}, \mathbf{R})$$

<u>Notation</u>

$\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ is a vector of binary outcomes

$\mathbf{x}_{ij}$ is a vector of predictors associated with y_{ij}

$\mathbf{R}$ is a correlation matrix for identifiability

β and $\mathbf{R}$ are parameters to be estimated

Multivariate Categorical Regression Methods

Generalized Estimating Equations (GEE)

- Cannot use GEE for Bayesian inference.

Generalized Linear Mixed Models (GLMMs)

- Posterior is improper when simple non-informative priors are chosen. (Natarajan and Kass, JASA, 2000)
- Regression parameters have subject-specific, not population averaged, interpretation.

Multivariate logistic regression

- Modeling dependency via multi-way odds ratios is unwieldy.

Multivariate probit regression

- Advantage: Simplified computation, modeling of dependency.

Objectives

- Propose new likelihood and computational algorithm for multivariate logistic regression
  - Model for individual outcomes is univariate logistic regression
  - Correlation structure is similar to probit models
- Advantages
  - Results can be summarized by odds ratios
  - Simple and flexible correlation structure
  - Computation is simple (like probit models)
  - Posterior is proper when non-informative priors are chosen

Model specification via underlying variables

Binary logistic regression

Univariate case

$$\text{Logit}\Pr[y_i = 1] = \mathbf{x}_i'\beta$$

$$\Updownarrow$$

Equivalent Model

$$y_i = 1(z_i > 0)$$

$$z_i \sim \text{Logistic}(\mathbf{x}_i'\beta, 1)$$

Logistic density

$$f(z;\, \mu) = \frac{\exp\big\{-(z-\mu)\big\}}{\big[1+\exp\big\{-(z-\mu)\big\}\big]^2}$$

Model specification via underlying variables

Multivariate generalization

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ denote vector of binary responses

Let $\mathbf{X}_i$ denote $(p \times q)$ matrix of predictors

$$y_{ij} = 1(z_{ij} > 0)$$

$$\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})' \sim \text{Multivariate Logistic}(\mathbf{X}_i\beta, \mathbf{R})$$

Choice of Multivariate Logistic Density

- There is a lack of flexible multivariate logistic distributions
- Need to define a new logistic density with a flexible correlation structure
- Approach: Transform variables that follow a standard multivariate distribution
- Let $\mathbf{t} = (t_1, \ldots, t_p)' \sim$ Multivariate $t_\nu(\mathbf{0}, \mathbf{R})$
- Let $z_j = \mu_j + \log\left(\frac{F(t_j)}{1-F(t_j)}\right)$, where $F(.)$ is CDF of t_j.
- Then $\mathbf{z} = (z_1, \ldots, z_p)'$ is Multivariate Logistic$(\boldsymbol{\mu}, \mathbf{R})$

Form of proposed multivariate logistic density

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

where the conventional multivariate t density is denoted by

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

$t_j = F^{-1}\left(e^{z_j}/(e^{z_j}+e^{\mu_j})\right)$ with $F^{-1}(.)$ denoting the inverse CDF of the $\mathcal{T}_{\tilde{\nu}}(0,1)$ density, $\mathbf{t} = (t_1,\ldots,t_p)'$, $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_p)'$, $\mathbf{R}$ is a correlation matrix (i.e., with 1's on the diagonal), and

Bayesian Implementation

Probability Model

$$y_{ij} = 1(z_{ij} > 0)$$

$$\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})' \sim \text{Multivariate Logistic}(\mathbf{X}_i\beta, \mathbf{R}_i)$$

$$\mathbf{R}_i = \mathbf{R}_i\,(\theta, \mathbf{X}_i)$$

Prior Specification

- Assume $\pi(\beta, \theta) = \pi(\beta)\pi(\theta)$
- Choose $\pi(\beta) \sim N(\beta_0, \mathbf{\Sigma}_\beta)$ or $\pi(\beta) \propto 1$
- Can use any prior for $\pi(\theta)$ (including uniform)

Posterior Computation

Use MCMC Importance Sampling (Hastings, 1970)

1. Use data-augmentation/Gibbs/Metropolis algorithm to sample from an approximation to the posterior, $\pi_{\text{approx}}(\theta|\text{data})$
2. Assign importance weights

$$\text{Let } \theta' \text{ denote sample from } \pi_{\text{approx}}(\theta|\text{data})$$

$$\text{Importance weight} = \frac{\pi_{\text{exact}}(\theta'|\text{data})}{\pi_{\text{approx}}(\theta'|\text{data})}$$

- Approximation is based on a (multivariate) t approximation to the (multivariate) logistic (see Albert & Chib, 1993).
- Nearly perfect approximation makes importance sampling highly efficient.

MCMC Importance Sampling (Hastings, 1970)

- A method to calculate population means, moments, percentiles, and other expectations of the form

$$E = \int_{-\infty}^{\infty} g(x)\pi(x)dx$$

- Suppose we have an MCMC algorithm that $\longrightarrow \pi^*(x) \approx \pi(x)$.
- Draw sample $\left\{x^{(1)}, \ldots, x^{(T)}\right\}$ from an MCMC that $\longrightarrow \pi^*(x)$
- Define importance weight $w^{(t)} = \pi\left(x^{(t)}\right)/\pi^*\left(x^{(t)}\right)$
- Can prove that as $T \to \infty$

$$\hat{E} = \frac{\sum_{t=1}^{T} w^{(t)} g(x^{(t)})}{\sum_{t=1}^{T} w^{(t)}} \longrightarrow \int_{-\infty}^{\infty} g(x)\pi(x)dx$$

Computation for Multivariate Logistic Regression

True Model – Logistic	Approximation – t-link
$y_{ij} = 1(z_{ij} > 0)$	$y_{ij} = 1(z^*_{ij} > 0)$
$z_{ij} = \mathbf{x}_{ij}\beta + \log\left(\frac{F(t_{ij})}{1-F(t_{ij})}\right)$	$z^*_{ij} = \mathbf{x}_{ij}\beta + \sigma t_{ij}$
$\mathbf{t}_i \sim \mathrm{N}\big(\mathbf{0}, \phi_i^{-1}\mathbf{R}\big)$	$\mathbf{t}_i \sim \mathrm{N}\big(\mathbf{0}, \phi_i^{-1}\mathbf{R}\big)$
$\phi_i \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$	$\phi_i \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$

- When ν and σ^2 are appropriately chosen, these two models yield virtually identical inferences about β and $\mathbf{R}$.
- We sample from posterior under multivariate t-link model.
  - Use data-augmentation approach (Albert & Chib, 1993)
  - Gibbs steps to update β, $\mathbf{t}_i$'s, ϕ_i's.
  - Metropolis step to update $\mathbf{R}$.

Full conditionals for t-link model

Likelihood: $y_{ij} = 1(z_{ij} > 0), \quad \phi_i \overset{\text{i.i.d}}{\sim} \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$

$$\mathbf{z}_i = \{z_{i1}, \ldots, z_{ip}\}' \sim N\left(\mathbf{X}_i\beta, \frac{\sigma^2}{\phi_i}\mathbf{R}\right)$$

Prior: $\beta \sim N(\beta_0, \mathbf{\Sigma}_\beta), \quad \mathbf{R} \sim \pi[\mathbf{R}]$

Full conditionals:

1. $\beta|\mathbf{z}, \mathbf{R}, \mathbf{y}, \boldsymbol{\phi} \sim N\left(\mathbf{AB}, \mathbf{A}\right)$

   $\mathbf{A} = \left[\mathbf{\Sigma}_\beta^{-1} + \sigma^{-2}\sum_{i=1}^n \phi_i \mathbf{X}_i'\mathbf{R}^{-1}\mathbf{X}_i\right]^{-1}$

   $\mathbf{B} = \left[\mathbf{\Sigma}_\beta^{-1}\beta_0 + \sigma^{-2}\sum_{i=1}^n \phi_i \mathbf{X}_i'\mathbf{R}^{-1}\mathbf{z}_i\right]$

2. $\mathbf{z}_i|\beta, \mathbf{\Sigma}, \mathbf{y}, \mathbf{z}_{(-i)}, \boldsymbol{\phi} \sim TN_{\Omega_y}(\mathbf{X}_i\beta, \frac{\sigma^2}{\phi_i}\mathbf{R})$

3. $\phi_i|\beta, \mathbf{\Sigma}, \mathbf{y}, \mathbf{z}, \boldsymbol{\phi}_{(-i)} \sim \text{Gamma}\left(\frac{\nu+p}{2}, \frac{\nu+\sigma^{-2}(\mathbf{z}_i-\mathbf{X}_i\beta)'\mathbf{R}^{-1}(\mathbf{z}_i-\mathbf{X}_i\beta)}{2}\right)$

4. $\mathbf{R}$ use Metropolis step

Metropolis step for R

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

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

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

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

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

Weights for Importance Sampling

$$\text{weight} \propto \pi_{\text{true}}\left(\beta, \mathbf{R}, \mathbf{z} | \mathbf{y}\right) \Big/ \pi_{\text{approx}}\left(\beta, \mathbf{R}, \mathbf{z} | \mathbf{y}\right).$$

An equivalent computational formula is

$$\begin{aligned} \text{weight} &= \frac{\pi_{\text{logistic}}\left(\mathbf{z} | \beta, \mathbf{R}\right)}{\pi^*_{t\text{-link}}\left(\mathbf{z} | \beta, \mathbf{R}\right)} \\ &= \prod_{i=1}^{n} \left(\frac{\mathcal{T}_{p,\tilde{\nu}}(\mathbf{t}_i |\, \mathbf{0}, \mathbf{R})}{\mathcal{T}_{p,\tilde{\nu}}(\mathbf{z}_i |\, x_{ij}\beta, \tilde{\sigma}^2 \mathbf{R})} \right) \prod_{j=1}^{p} \left(\frac{\mathcal{L}(z_{ij} |\, x_{ij}\beta)}{\mathcal{T}_{1,\tilde{\nu}}(t_{ij} |\, 0)} \right), \end{aligned}$$

where $\mathbf{t}_i = (t_{i1}, \ldots, t_{ip})'$ is defined by $t_j = F^{-1}(e^{z_j}/(e^{z_j} + e^{\mu_j}))$ with $F^{-1}(.)$ denoting the inverse CDF of the $\mathcal{T}_\nu(0, 1)$ density.

Extension to ordered categorical data

Cumulative logits model

Univariate case

$$\text{Logit}\Pr[y_i \leq k] = \alpha_k - \mathbf{x}_i'\beta$$

$$\Updownarrow$$

$$y_i = \begin{cases} 1 & \text{if } z_i \in (-\infty, \alpha_1) \\ 2 & \text{if } z_i \in (\alpha_1, \alpha_2) \\ \vdots & \vdots \\ k & \text{if } z_i \in (\alpha_{k-1}, \infty) \end{cases}, \; z_i \sim \text{Logistic}(\mathbf{x}_i'\beta, 1)$$

If $\pi(\alpha) \propto 1$, then full conditional of α_j is uniform.

Full conditional of z_i is truncated to fall in $(\alpha_{y_i-1}, \alpha_{y_i})$

Approaches to sampling from truncated normal

1. Inverse CDF method
   - Draw $u \sim \text{Uniform}\left(\Phi(\frac{a-\mu}{\sigma}), \Phi(\frac{b-\mu}{\sigma})\right)$
   - Set $z = \mu + \sigma\Phi^{-1}(u)$
   - Computing $\Phi^{-1}(.)$ is slow
   - Splus crashes when $(a-\mu)/\sigma > 8$
2. Importance sampling with exponential density for (a, ∞)
   - Draw $E_i \overset{\text{ind}}{\sim} \text{Exponential}(1), i = 1, 2$
   - Repeat until $E_1^2 \leq 2\sigma^{-2}(a-\mu)^2 E_2$
   - Set $x = a + \frac{E_1}{\frac{a-\mu}{\sigma}}$
3. Geweke (1989) proposed a mixed-rejection algorithm that chooses between i) normal rejection sampling, ii) uniform rejection sampling, iii) exponential rejection sampling.

Trick for verifying propriety

- Let $\mathbf{y}$ denote the $(n \times p)$ matrix of outcomes and let $\mathbf{y}_j^*$ denote the data in the jth column of $\mathbf{y}$. In other words, $\mathbf{y}_j^* = \{y_{1j}, \ldots, y_{nj}\}'$.
- Let $\pi[\beta|\mathbf{y}_j^*]$ denote the posterior distribution of β given $\mathbf{y}_j^*$ obtained by fitting a univariate logistic regression model with $\pi[\beta] \propto 1$.
- Theorem: If at least one $\pi[\beta|\mathbf{y}_j^*]$ is proper then $\pi[\beta, \mathbf{R}|\mathbf{y}]$ is proper.
  Proof:

$$\begin{aligned} \pi[\beta, R|\mathbf{y}] &\propto \int\int \Pr(\mathbf{y}|\beta, \mathbf{R}) d\beta d\mathbf{R} \\ &= \int\int \Pr(\mathbf{y}_j^*|\beta, \mathbf{R}) \times \Pr(\mathbf{y}|\beta, \mathbf{R}, \mathbf{y}_j^*) d\beta d\mathbf{R} \\ &\leq \int \Pr(\mathbf{y}_j^*|\beta, \mathbf{R}) d\beta \int d\mathbf{R} = \int \pi[\beta|\mathbf{y}_j^*] d\beta \end{aligned}$$

- Note: $\pi[\beta|\mathbf{y}_j^*]$ is proper if MLE exists. Programs like SAS PROC LOGISTIC automatically check for existence of the MLE.

Example

Data: All twin pregnancies ($n = 584$) enrolled in the Collaborative Perinatal Project from 1959 to 1965

Outcome: Small for gestational age (SGA) birth.

Covariates: Gender, maternal age, years of cigarette smoking, weight gain during pregnancy, gestational age at delivery, and variables relating to obstetric history.

Previously analyzed by: Ananth and Preisser analyzed data via maximum likelihood using a different bivariate logistic model. Used odds ratios to model within-twins association

Goal: (i) To assess efficiency of importance sampling. (ii) To assess whether two different models yield similar results.

Example - Model and Prior Specification

Marginal probability model:

(Same as Ananth & Preisser)

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

$$\mathbf{x}'\beta = \text{ to be defined}$$

Correlation model:

Let ρ_i denote single free correlation parameter in $\mathbf{R}_i$.

$$\rho_i = \begin{cases} \theta_1 & \text{if subject } i \text{ is primiparous} \\ \theta_2 & \text{if subject } i \text{ is multiparous} \end{cases}$$

Prior:

$$\pi(\beta, \theta_1, \theta_2) \propto 1$$

Table 1. Bivariate logistic regression analysis of SGA in twins

		A & P, 1999[†]	Posterior Summary	
Covariate		MLE (SE)	Mean (SD)	OR (95% CI)
Intercept	β_0	3.10 (1.62)	2.97 (1.63)	
Female infant	β_1	0.36 (0.17)	0.35 (0.17)	1.43 (1.01–2.01)
Pregnancy history	β_2	-0.36 (0.26)	-0.39 (0.26)	0.68 (0.41–1.13)
	β_3	-0.92 (0.38)	-0.97 (0.38)	0.38 (0.18–0.79)
	β_4	0.44 (0.33)	0.42 (0.32)	1.53 (0.81–2.87)
	β_5	0.38 (0.46)	0.45 (0.44)	1.57 (0.66–3.68)
log(age)	β_6	-1.03 (0.37)	-1.00 (0.47)	0.37 (0.15–0.91)
log(wt gain + 6)	β_7	-0.47 (0.17)	-0.46 (0.17)	0.63 (0.45–0.88)
log(yrs smoking + 1)	β_8	0.26 (0.10)	0.25 (0.09)	1.28 (1.07–1.55)
(Gest age − 37)	β_9	0.21 (0.04)	0.21 (0.04)	
$(\text{Gest age} - 37)^2$	β_{10}	0.02 (0.01)	0.02 (0.01)	
Correlation(Primiparous)	θ_1*	–	0.16 (0.16)	
Correlation (Multiparous)	θ_2*	–	0.35 (0.07)	

$* \Pr[\theta_2 > \theta_1 | \text{data}] = 86\%$

Conclusions

- Computational algorithm is easy to program and efficient
- Posterior is proper under mild conditions
- Uses underlying normal framework, similar to probit models
- Has marginal logistic interpretation for individual outcomes
- Generalizations are straightforward.
  - Multivariate polychotomous outcomes
  - Mixed discrete and continuous outcomes