

Stat 8931 Flu Homework

Charles J. Geyer

October 25, 2005

1 Introduction

1.1 Logit-Normal GLMM

In a Logit-Normal generalized linear mixed model (GLMM), the observed data is a vector y whose components are conditionally independent Bernoulli random variables given the missing data vector b , which is unconditionally jointly mean-zero multivariate normal. The model specification is completed by the specification of the *linear predictor*

$$\eta = X\beta + Zb \tag{1}$$

and the link function. In (1) X and Z are known matrices (the *model matrices* for fixed and random effects, respectively), β is a vector of unknown parameters (*fixed effects*), b is the vector of missing data (*random effects*), and the conditional expectation of y given η is $\text{logit}^{-1}(\eta)$.

The unknown parameters to be estimated are β and any unknown parameters determining the variance matrix of b . Usually this variance matrix has simple structure and involves only a few unknown parameters. Following Sung and Geyer (submitted), we simplify the structure of the model still further. The class of models they use is more easily described in R than in mathematical notation. The linear predictor has the form

$$\text{eta} = \mathbf{X} \%*\% \text{beta} + \mathbf{Z} \%*\% (\text{sigma}[\mathbf{i}] * \mathbf{b}) \tag{2}$$

where \mathbf{X} and \mathbf{Z} are the matrices X and Z in (1) and $\mathbf{X} \%*\% \text{beta}$ is the matrix multiplication $X\beta$ so the only way in which (1) differs from (2) other than notationally is that b in (1) is replaced by $(\text{sigma}[\mathbf{i}] * \mathbf{b})$ in (2), which, for readers not familiar with R, has the following interpretation: sigma is a vector of unknown parameters, \mathbf{i} is a vector of the same length as \mathbf{b} and having values that are possible indices for sigma , so $\text{sigma}[\mathbf{i}]$ is the

vector $(\sigma_{i_1}, \dots, \sigma_{i_m})$ in ordinary mathematical notation and $*$ in (2) denotes coordinatewise multiplication, so if z_{jk} are the components of the matrix Z the second term on the right hand side of (2) has j -th component

$$\sum_{k=1}^m z_{jk} \sigma_{i_k} b_k$$

written in conventional mathematical notation.

We also change assumptions; in (1) b is general multivariate normal, but in (2) \mathbf{b} is standard multivariate normal (mean vector zero, variance matrix the identity). Thus the only unknown parameters in our model are the vectors `beta` and `sigma`.

We also allow for independent and identically distributed (IID) data, in which case the data y is a matrix with IID columns, each column of y modeled as described above.

Summary First consider a single individual. The data on this individual consist of observed data y and missing data (random effects) b , both random vectors. Marginally, the random effects vector b has independent standard normal components. The conditional distribution of y given b has independent Bernoulli components. The success probabilities are given by

$$E(y_i|b) = \text{logit}^{-1}(\eta_i) = \frac{1}{1 + \exp(-\eta_i)}$$

where η_i are the components of the vector given by (2).

When we have multiple individuals, they are considered independent and identically distributed.

1.2 The Influenza Data

Coull and Agresti (2000) propose fitting Logit-Normal GLMM to data on influenza (flu).

1.2.1 Data

First we load the data, copied from Table 1 in Coull and Agresti (2000).

```
> data <- read.table("flu.txt", header = TRUE)
> names(data)
```

```
[1] "y1" "y2" "y3" "y4" "nobs"
```

```

> nobs <- data$nobs
> data$nobs <- NULL
> ypat <- as.matrix(data)
> ypat <- ypat[nobs > 0, ]
> nobs <- nobs[nobs > 0]
> foo <- cbind(ypat, nobs)
> dimnames(foo)[[1]] <- rep("", nrow(foo))

```

```

> foo

```

y1	y2	y3	y4	nobs
0	0	0	0	140
0	0	0	1	31
0	0	1	0	16
0	0	1	1	3
0	1	0	0	17
0	1	0	1	2
0	1	1	0	5
0	1	1	1	1
1	0	0	0	20
1	0	0	1	2
1	0	1	0	9
1	1	0	0	12
1	1	0	1	1
1	1	1	0	4

there are

```

> sum(nobs)

```

```

[1] 263

```

independent and identically distributed (IID) individuals, each with four Bernoulli observations (Y_1, \dots, Y_4) . The meaning is that the observations are IID random vectors $y = (y_1, y_2, y_3, y_4)$ and the i -th pattern of y values (the i -th row of `ypat`) is observed `nobs[i]` times.

1.2.2 Model

We fit the model the random effects having variance matrix

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_1 & \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_2 & \rho_2 & 1 \end{bmatrix} \quad (3)$$

which is given by equation (7) in Coull and Agresti (2000). In order to use our restricted model random vector having variance matrix (3) in the form $Z\Delta b$ where Z is a fixed known matrix that does not contain parameters, Δ is a diagonal positive semi-definite matrix, the diagonal being $\mathbf{sigma}[i]$ in the notation of the preceding section, and b is a standard normal random vector (having IID standard normal components). We try

$$Z = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

Then

$$\begin{aligned} \text{var}(Z\Delta b) &= Z\Delta^2 Z^T \\ &= \begin{pmatrix} \delta_1^2 + \delta_2^2 + \delta_3^2 & \delta_1^2 + \delta_2^2 & \delta_1^2 + \delta_2^2 & \delta_1^2 - \delta_2^2 \\ \delta_1^2 + \delta_2^2 & \delta_1^2 + \delta_2^2 + \delta_4^2 & \delta_1^2 + \delta_2^2 & \delta_1^2 - \delta_2^2 \\ \delta_1^2 + \delta_2^2 & \delta_1^2 + \delta_2^2 & \delta_1^2 + \delta_2^2 + \delta_5^2 & \delta_1^2 - \delta_2^2 \\ \delta_1^2 - \delta_2^2 & \delta_1^2 - \delta_2^2 & \delta_1^2 - \delta_2^2 & \delta_1^2 + \delta_2^2 + \delta_6^2 \end{pmatrix} \end{aligned} \quad (5)$$

and we see that (3) and (5) do indeed have the same form if we impose the constraints

$$\delta_3 = \delta_4 = \delta_5 = \delta_6 \quad (6)$$

with one caveat. The representation (3) is positive semi-definite if

$$-\frac{1}{2} \leq \rho_1 \leq 1 \quad (7a)$$

$$\rho_2^2 \leq \frac{1 + 2\rho_1}{3} \quad (7b)$$

whereas the representation (5) is necessarily positive semi-definite, but as the δ_i range over all non-negative numbers the correlations

$$\rho_1 = \frac{\delta_1^2 + \delta_2^2}{\delta_1^2 + \delta_2^2 + \delta_3^2}$$

$$\rho_2 = \frac{\delta_1^2 - \delta_2^2}{\delta_1^2 + \delta_2^2 + \delta_3^2}$$

only fill out the set

$$0 \leq \rho_1 \leq 1 \tag{8a}$$

$$-\rho_1 \leq \rho_2 \leq \rho_1 \tag{8b}$$

Hence our $Z\Delta b$ model is a restriction of the full model specified by (3).

Nevertheless, if the MLE is in the interior of the parameter space for our $Z\Delta b$ model, then it is the same as the MLE for the full model.

To finish the model specification we must specify the design matrices

```
> x <- diag(4)
> z <- rbind(c(1, 1, 1, 0, 0, 0), c(1, 1, 0, 1, 0, 0), c(1, 1,
+ 0, 0, 1, 0), c(1, -1, 0, 0, 0, 1))
> idx <- c(1, 2, 3, 3, 3, 3)
```

the last vector being used to impose the constraint (6). Now `sigma[idx]` is the diagonal of Δ . (We switched from `i` to `idx` because it is too hard to remember not to use the variable name `i` for anything else).

1.2.3 Putative MLE

Coull and Agresti (2000) give the following MLE for this model.

```
> beta.putative <- c(-4, -4.4, -4.7, -4.5)
> sigmasq.putative <- 4.05^2
> rho1.putative <- 0.43
> rho2.putative <- (-0.25)
> delta3.putative <- sqrt(sigmasq.putative * (1 - rho1.putative))
> delta2.putative <- sqrt(sigmasq.putative * (rho1.putative - rho2.putative)/2)
> delta1.putative <- sqrt(sigmasq.putative * (rho1.putative + rho2.putative)/2)
> delta.putative <- c(delta1.putative, delta2.putative, delta3.putative)
> beta.putative

[1] -4.0 -4.4 -4.7 -4.5

> delta.putative

[1] 1.215000 2.361536 3.057683
```

2 The Assigned Homework

Check the MLE found by Coull and Agresti (2000) (given in Section 1.2.3 above) using the `metrop` function in the MCMC package to calculate the score. More precisely if $f_\theta(b, y)$ is the complete data joint density (you have to work this out for yourself, although it is described implicitly in the first section), do the following.

1. Simulate using the `metrop` function the conditional distribution of b given y under the parameter value θ (which denotes a vector of seven parameters, four betas and three deltas).
2. If b_1, b_2, \dots are the resulting Markov chain, calculate

$$\frac{1}{N} \sum_{i=1}^N \nabla \log f_\theta(b_i, y) \quad (9)$$

where ∇ denotes differentiation w. r. t. θ (and hence is a seven-dimensional vector of partial derivatives).

3. Calculate MCSE (using the method of batch means or any other valid method) for each component of (9).
4. You should get zero (to within MCSE) for (9), assuming Coull and Agresti (2000) are correct, which all the calculations I have done seem to support. As statisticians we know that you can never “accept” a null hypothesis, you can only “fail to reject” it (in the words of one intro textbook I have used). So we can’t be sure from the MCMC that this is the true MLE, but we can check it to arbitrary accuracy. Get your MCSE for the components of (9) to less than 0.25 (**note:** this is increased from the paper handed out in class).

Hints

For each individual, the observed data is a Bernoulli vector of length four (a row of `ypat`) and the missing data is continuous random vector of length four (marginally normal, but not conditionally given the observed data). Although there are `sum(nobs) = 263` individuals, all individuals with the same observed data pattern (belonging to the same row of `ypat`) make the same contribution to the log likelihood. Therefore we need only simulate as many missing data patterns as there are observed data patterns, which is `nrow(ypat) = 14`.

In order to make this clear, write n_j for `nobs[j]`, and write y_j for the data pattern `ypat[j,]` (note that y_j is a four-vector), and write b_{ij} for the missing data for an individual with data pattern y_j simulated in the Markov chain. Here i indicates an iteration of the Markov chain, j a data pattern, and b_{ij} is a six-vector (so that Zb_{ij} makes sense).

Then we can rewrite (9) as

$$\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^J n_j \nabla \log f_{\theta}(b_{ij}, y_j) \quad (10)$$

where now f_{θ} denotes the complete data density for one individual with observed data y_j and missing data b_{ij} .

Thus we see that we can have the dimension of the state of the Markov chain be `nrow(ypat) × ncol(z) = 84`.

An alternative method would be to run `nrow(ypat) = 14` independent Markov chains, each with state of dimension `ncol(z) = 6`. We still denote the realizations b_{ij} but now for each fixed j b_{1j}, b_{2j}, \dots is a Markov chain (all by itself).

Then we can calculate the contribution to the log likelihood

$$\frac{1}{N} \sum_{i=1}^N \nabla \log f_{\theta}(b_{ij}, y_j) \quad (11)$$

from this Markov chain and calculate standard errors for that too. Of course, (11) is not supposed to be zero. Only when the terms of (11) are multiplied by n_j and added to give (10) do we get zero (theoretically).

References

- Sung, Y. J. and Geyer, C. J. (submitted). Monte Carlo likelihood inference for missing data models. <http://www.stat.umn.edu/geyer/bernor/ms.pdf>.
- Coull, B. A. and Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, **56**, 73–80.