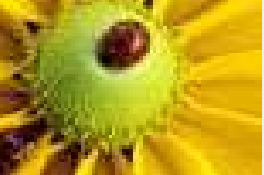


# Mixed Model Prediction: Part II

Jiming Jiang and Thuan Nguyen

University of California, Davis, USA

and Oregon Health & Science University, USA



# Generalized linear mixed models

- McCullagh & Nelder (1989) set up the foundation of generalized linear models (GLM)

Three basic elements for a linear model (LM):

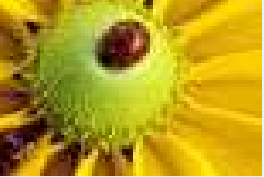
- (i) the observations are independent;
- (ii) the mean of the observation is a linear function of covariates, called *linear predictor*;
- (iii) the variance of the observation is a constant.

Generalized linear mixed

Examples

More about GLMM

Example: Salamander data



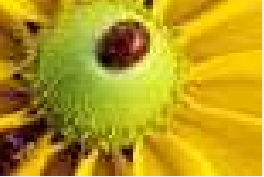
- Extending the three elements to GLM:

(i) stays the same;

(ii) the mean of the observation is associated with a linear predictor through a link function;

(iii) the variance of the observation is a function of the mean.

Examples of GLM: Logistic regression, log-linear model, etc.



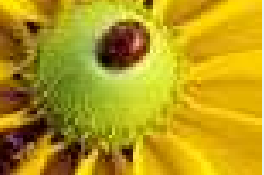
- GLMM may be viewed as an extension of GLM; alternatively, it can be viewed as an extension of LMM (covered in this course earlier).

(i) (conditional independence) given a vector  $\alpha$  of random effects, responses  $y_1, \dots, y_n$  are (conditionally) independent;

(ii) (conditional distribution) the conditional pdf (or pmf) of  $y_i$  given the random effects is a member of the exponential family, i.e.,

$$(1) \quad f_i(y_i | \alpha) = \exp \left\{ \frac{y_i \xi_i - b(\xi_i)}{a_i(\phi)} + c_i(y_i, \phi) \right\},$$

where  $b, a_i, c_i$  are known functions, and  $\phi$  is a dispersion parameter, which may or may not be known, and  $\xi_i$  is the natural parameter of the exponential family.

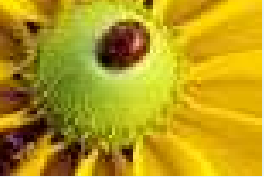


- Well-known cases where  $\phi$  is known: Binomial (including Bernoulli), Poisson; well-known cases where  $\phi$  is unknown: Normal, Gamma.

(iii) (conditional mean) by the properties of exponential family (e.g., McCullagh & Nelder 1989, sec. 2.2), the natural parameter,  $\xi_i$ , is associated with the conditional mean,  $\mu_i = E(y_i|\alpha)$ . Assume that  $\mu_i$  is, in turn, associated with a linear predictor through a link function,  $g$ :

$$(2) \quad g(\mu_i) = \eta_i = x_i' \beta + z_i' \alpha.$$

The right side of (2) is what we call the linear predictor, where  $x_i, z_i$  are known vectors,  $\beta$  is a vector of fixed effects, and  $\alpha$  is a vector of random effects.



- Note that the random effects are unobserved.

(iv) (distribution of random effects) assume further that

$$(3) \quad \alpha \sim N(0, G),$$

that is,  $\alpha$  is multivariate normal with mean vector 0 and covariance matrix  $G$ , which may depend on a vector  $\theta$  of variance components.

Assumptions (i)–(iv) define a GLMM.

Special cases of GLMM: LMM, mixed logistic model, Poisson mixed model, etc.



# Examples

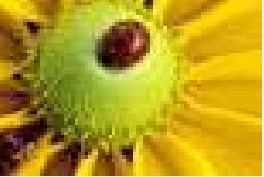
- 1. *Salamander mating data*: McCullagh & Nelder (1989, §14.5).

Two populations, Whiteside and Rough Butt, geographically isolated.

40 salamanders, 20 female and 20 male, 10 from each population; summer and fall, 1986.

A new group of 40 animals, with the same distribution, were added to the fall 86 experiment.

The responses are binary—1 if the mating between the  $i$ th female and  $j$ th male is “successful”, 0 otherwise, and are expected to be correlated.

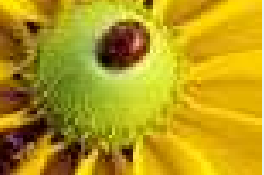


- Furthermore, the designs was partially crossed such that each female would encounter 6 males, and each male 6 females, during each of the mating experiments.

The same design used in all three experiments.

A mixed logistic model was proposed: Suppose that (i) given the female and male random effects  $u_i, v_j, (i, j) \in S$ , where  $S$  is a subset of  $\mathcal{I} = \{(i, j) : 1 \leq i \leq m_1, 1 \leq j \leq m_2\}$ , binary responses  $y_{ijk}, (i, j) \in S, k = 1, \dots, c_{ij}$  are conditionally independent;





- (ii) with  $p_{ijk} = P(y_{ijk} = 1|u, v)$ , we have  $\text{logit}(p_{ijk}) = x'_{ijk}\beta + u_i + v_j$ , where

$$\text{logit}(p) = \log \left( \frac{p}{1-p} \right),$$

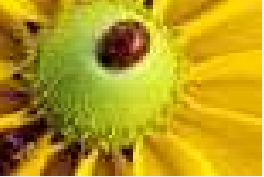
$x_{ijk}$  is a known vector of covariates,  $\beta$  is a unknown vector of parameters, and  $u_i, v_j$  denote the female and male random effects that are involved.

Specifically, one has

$$x'_{ij}\beta = \beta_0 + \beta_1 1_{\text{WSF}} + \beta_2 1_{\text{WSM}} + \beta_3 1_{\text{WSF}} \times 1_{\text{WSM}},$$

where  $1_{\text{WSF}}$  is the indicator of whiteside female, and  $1_{\text{WSM}}$  the indicator of whiteside male.

Note that the product,  $1_{\text{WSF}} \times 1_{\text{WSM}}$ , corresponds to the interaction between female and male.

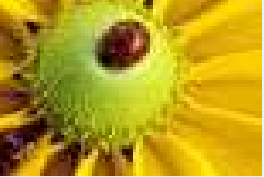


- (iii) the random effects  $u_i$ 's and  $v_j$ 's are independent with  $u_i \sim N(0, \sigma_1^2)$  and  $v_j \sim N(0, \sigma_2^2)$ , where  $\sigma_1^2, \sigma_2^2$  are unknown variances.

Assumptions (i)–(iii) define a mixed logistic model, which is a special case of GLMM.

Here, the exponential family is the Bernoulli distribution, and the link function is  $g(\mu) = \text{logit}(\mu)$  ( $0 < \mu < 1$  is a probability).

The salamander data is “infamous” due to the computational difficulty of computing the MLE (see below).

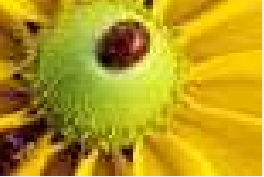


## ■ 2. *Epileptic seizure counts*

Poisson distribution is often used to model responses that are counts.

However, in many cases there is overdispersion (or underdispersion), so that the variance of the response does not follow that of a Poisson distribution.

Thall and Vail (1990) studied one of such cases, in which they presented data from a clinical trial involving 59 epileptics.



alized linear mixed

s

les

ce about GLMM

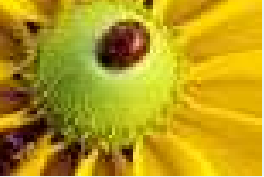
le: Salamander data

d

- These patients were randomized to a new drug (treatment) or a placebo (control).

The number of epileptic seizures was recorded for each patient during an eight-week period, namely, one seizure count during the two-week period before each of four clinic visits.

The responses are logitudinal; therefore, some correlations are expected among the responses, which may be the reason of the overdispersion that was observed.



alized linear mixed

les

ce about GLMM

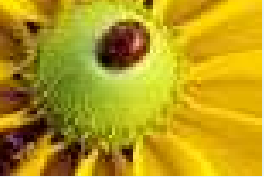
le: Salamander data

- Breslow and Clayton (1993) reanalyzed the data by proposing a Poisson log-linear mixed model.

It assumed that the seizure count  $y_{ij}$  for the  $i$ th patient on the  $j$ th visit ( $i = 1, \dots, 59, j = 1, \dots, 4$ ) was associated with an individual-specific random effect,  $\alpha_i = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \end{pmatrix}$ .

Given the random effects, the seizure counts  $y_{ij}$  are independent  $\sim \text{Poisson}(\mu_{ij})$ , where

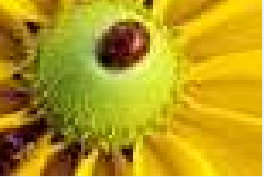
$$\log(\mu_{ij}) = x'_{ij}\beta + \alpha_{1i} + \alpha_{2i}(\text{Visit}_j/10) + \epsilon_{ij}.$$



- Here  $x_{ij}$  is a vector of covariates including indicators of the treatment, visit, the logarithm of  $1/4$  times the number of baseline seizures (Base), the logarithm of age (Age) and some interactions;  $Visit_j$  is the visit code that equals  $-3$ ,  $-1$ ,  $1$ , and  $3$ , respectively, for  $j = 1, \dots, 4$ ,

and  $\epsilon_{ij}$  is an additional random error that represents the overdispersion not captured by the individual-specific random effects.

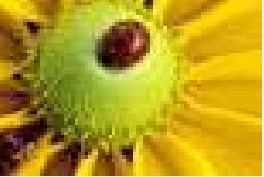
Finally, the random effects  $\alpha_i$  are independent bivariate normal with means zero, unknown variances and correlation coefficient, which are part of the variance components.



- The data associated with this example is very typical in medical studies. It is called longitudinal data because responses from the same subject (e.g., patient) are collected over time; and multiple subjects are involved.

As a result, the responses are not independent. It is important to capture the correlations among the responses; otherwise, the inference results may be inaccurate, or even invalid.

The proposed model is a special case of GLMM, called Poisson mixed model: The exponential family is Poisson distribution, with the link function  $g(\mu) = \log(\mu)$ .

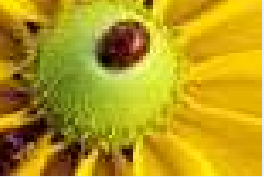


### ■ 3. *Small area estimation of Mammography Rates*

The Behavioral Risk Factor Surveillance System (BRFSS) is a Center for Disease Control and Prevention coordinated, state-based random-digit-dialing telephone survey.

A data set of particular interest involved the use of mammography among women aged 40 or older, from 1993 to 1995, and for areas from three federal regional offices: Boston (including Maine, Vermont, Massachusetts, Connecticut, Rhode Island, and New Hampshire), New York (including New York and New Jersey) and Philadelphia (including Pennsylvania, Delaware, Washington DC, Maryland, Virginia, and West Virginia).





- Overall, there were 118 health service areas (HSAs) in the region, which were considered as small areas.

Initial analysis of the data suggested that mammography rates gradually increase from age groups 40–44 to 50–54, and then decrease.

To catch this curvature phenomena, Jiang et al. (2001) proposed a mixed logistic model for the proportion of women having had mammography.

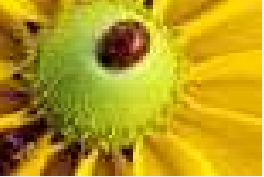
Under this model, there is a random effect corresponding to each HSA.

alized linear mixed

les

ce about GLMM

le: Salamander data



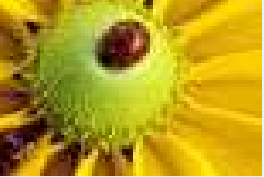
- Given a HSA, the proportion of women having had mammography,  $p$ , satisfies

$$\text{logit}(p) = \beta_0 + \beta_1 * \text{age} + \beta_2 * \text{age}^2 + \beta_3 * \text{Race} + \beta_4 * \text{Edu} + \text{HSA effect},$$

where Age is grouped as 40–44, 45–49, ..., 75–79, and 80 and over; Race as white and others; and Edu as the percentage of people in the HSA aged 25 or older with at least a high school education.

Finally, the HSA effects are assumed to be independent normal with mean zero and unknown variance.

The proposed model is, again, a case of mixed logistic model.



# Inference about GLMM

- Difficulty: The likelihood function under a GLMM typically involves integrals, possibly high dimensional, with no analytic expression.

For example, under the mixed logistic model for the salamander data, the likelihood function involves an 80-dimensional integral with no analytic form.

Due to such difficulties, some alternative methods of inference had been proposed, before the latest computational advances that made computation of the MLE under GLMM possible.

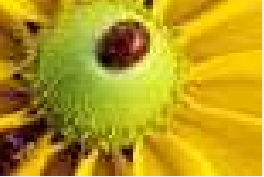
For the most part, the methods include 1. approximate inference; 2. generalized estimating equations; 3. Monte-Carlo EM; 4. data cloning; and 5. Bayesian inference.

alized linear mixed

les

ce about GLMM

le: Salamander data



## ■ 1. Approximate inference

PQL (penalized quasi-likelihood; Breslow & Clayton 1993).

Idea: Approximate the likelihood function via Laplace approximation.

Advantage: Computationally attractive (available in major software, such as SAS and R).

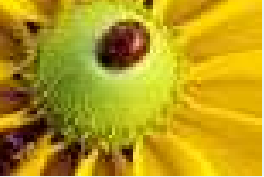
Disadvantage: The estimators are biased and, in fact, inconsistent.

alized linear mixed

les

ce about GLMM

le: Salamander data



## ■ 2. GEE (generalized estimating equations)

Method of simulated moments (Jiang 1998).

Robust estimation (Jiang & Zhang 2001).

Advantages: Computationally attractive; estimators are consistent.

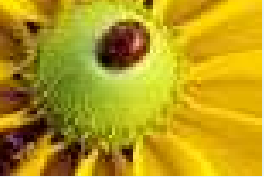
Disadvantage: The estimators may not be as efficient as the MLE.

alized linear mixed

les

ce about GLMM

le: Salamander data

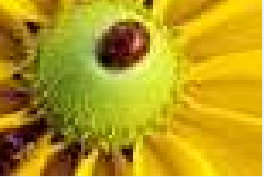


### ■ 3. Monte-Carlo EM

If the likelihood function only involves low-dimensional (e.g., one-dimensional) integrals, numerical integration can be used to compute the MLE.

But the method does not work if the likelihood involves high-dimensional integrals, such as for the salamander data.

Alternatively, one may consider using the E-M algorithm (Dempster *et al.* 1977).



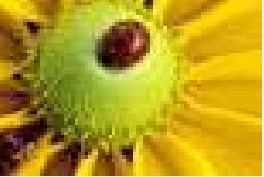
- Idea: “complete data”  $(y, \alpha)$ , but  $\alpha$  is not observed.

For example,  $\alpha$  is the vector of random effects.

E-step: Compute the conditional expectation of the “log-likelihood” based on the complete data, given the observed data,  $y$ .

The conditional expectations are evaluated at the current estimated parameters.

M-step: Maximize the conditional expectation to find updates of the parameter estimates.



alized linear mixed

S

les

ce about GLMM

le: Salamander data

d

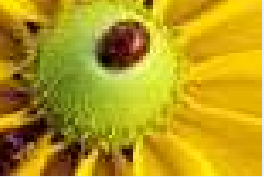
- For GLMMs, the E-step is difficult, as it involves integrals with no analytic expressions.

Therefore, Monte-Carlo methods are used to evaluate the conditional expectations in the E-step.

Due to the complexity of GLMM, advanced Monte-Carlo methods may be needed.

In particular, Markov-chain Monte-Carlo (MCMC) methods have been used, such as the Gibbs sampler (Geman & Geman 1984) and Metropolis-Hastings algorithm (Metropolis *et al.* 1953, Hastings 1970). See, for example, Jiang (2007), sec. 4.1, for details.





## ■ 4. Data cloning

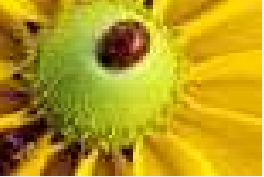
In a latest development, Lele *et al.* (2007, 2010) proposed an interesting idea called *data cloning*.

It uses the Bayesian computational approach, namely, MCMC, for frequentist purposes.

Let  $\pi$  denote the prior pdf of  $\psi$ . Then, one has the posterior,

$$\pi(\psi|y) = \frac{p(y|\psi)\pi(\psi)}{p(y)},$$

where  $p(y)$  is the integral of the numerator with respect to  $\psi$ , which does not depend on  $\psi$ .

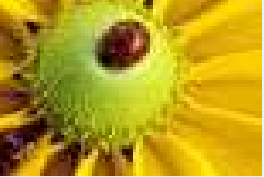


- There are computational tools using the MCMC method for posterior simulation that generate random variables from the posterior without having to compute the posterior.

Thus, we can assume that one can generate random variables from the posterior, whatever it is.

Here comes the key idea: Suppose the observations  $y$  were repeated independently from  $M$  different individuals such all of these result in exactly the same data,  $y$ , denoted by

$$y^{[M]} = \underbrace{(y, \dots, y)}_{M \text{ of them}}.$$

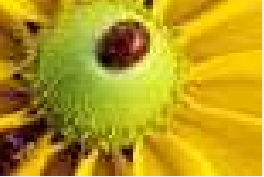


- Then, the posterior based on  $y^{[M]}$  is

$$(4) \quad \pi_M(\psi|y^{[M]}) = \frac{\{p(y|\psi)\}^M \pi(\psi)}{\int \{p(y|\psi)\}^M \pi(\psi) d\psi}.$$

It can be shown that, as  $M$  increases, the right side of (4) converges to a multivariate normal distribution whose mean vector is equal to the MLE,  $\hat{\psi}$ , and whose covariance matrix is approximately equal to  $M^{-1} I^{-1}(\hat{\psi})$ , where  $I(\psi)$  is the Fisher information matrix.

Thus, for large  $M$ , one can approximate the MLE by the sample mean vector of, say,  $\psi^{(1)}, \dots, \psi^{(B)}$ , generated from the posterior (4) (using MCMC).



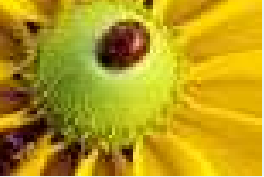
■ That is,

$$(5) \quad \hat{\psi} \approx \bar{\psi} = \frac{1}{B} \sum_{b=1}^B \psi^{(b)}.$$

Furthermore,  $I^{-1}(\hat{\psi})$  can be approximated by  $M$  times the sample covariance matrix of  $\psi^{(1)}, \dots, \psi^{(B)}$ . That is,

$$(6) \quad I^{-1}(\hat{\psi}) \approx \frac{M}{B-1} \sum_{b=1}^B \{\psi^{(b)} - \bar{\psi}\} \{\psi^{(b)} - \bar{\psi}\}',$$

The right side of (6) can be used to obtain the standard errors of  $\hat{\psi}$ , the MLE.



## ■ 5. Bayesian Inference

A hierarchical GLMM:  $y|\beta, \alpha, H \sim \text{Exponential Family};$

$\alpha \sim N(0, G); G$  is part of  $H$ ;

$(\beta, H) \sim \pi, \text{ a prior.}$

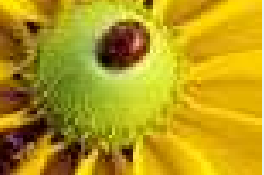
Note: Without the prior, the model is the same as GLMM.

alized linear mixed

bles

ce about GLMM

le: Salamander data



- The posterior for  $(\beta, H)$  is

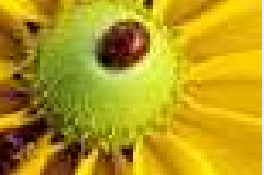
$$p(\beta, H|y) = \frac{\int f(y|\beta, \alpha, H)f(\alpha|G)\pi(\beta, H)d\alpha}{\int \int f(y|\beta, \alpha, H)f(\alpha|G)\pi(\beta, H)d\alpha d\beta dH}.$$

The posterior for  $\alpha$  is

$$p(\alpha|y) = \frac{\int f(y|\beta, \alpha, H)f(\alpha|G)\pi(\beta, H)d\beta dH}{\int \int f(y|\beta, \alpha, H)f(\alpha|G)\pi(\beta, H)d\alpha d\beta dH}.$$

Here,  $f(y|\beta, \alpha, H) = \prod_{i=1}^n f(y_i|\beta, \alpha, \phi)$ , with  $H = (G, \phi)$ , and  $f(\alpha|G)$  is the pdf of  $N(0, G)$ .

In most cases, the posterior cannot be evaluated analytically; MCMC methods need to be used to draw samples from the posterior.



# Example: Salamander data revisited

- Torabi (2012) used data cloning to carry out a maximum likelihood analysis under the mixed logistic model proposed earlier, using Gibbs sampler for the posterior sampling. The MLE and their standard errors were obtained as follows:

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
.96 (.39)	-2.84 (.61)	-.66 (.36)	3.56 (.64)
$\sigma_u^2$	$\sigma_v^2$		
1.34 (.62)	1.03 (.53)		

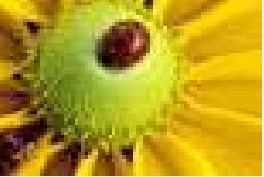
It appears that WS is a worse population for mating regardless of the gender. Nevertheless, the interaction between female and male is found highly significant, which makes sense. The female random effects are found significant at 5% level; the male random effects are barely insignificant at 5% level.

alized linear mixed

les

ce about GLMM

le: Salamander data



alized linear mixed

les

ce about GLMM

le: Salamander data

- Karim & Zeger (1992) used Gibbs sampler to evaluate the posteriors (e.g., Jiang 2007, p. 179).  
The authors reported posterior medians as point estimates of the GLMM parameters, and the difference between the 95th and 5th percentiles of the posterior divided by  $2 \times 1.645 = 3.29$  as the standard errors.

The results are given below

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
1.03 (.43)	-3.01 (.60)	-.69 (.50)	3.74 (.68)
$\sigma_u^2$	$\sigma_v^2$		
1.50 (NA)	1.36 (NA)		

The interpretations are similar.