DRAFT: Comparison of Fitting Methods of Generalized Linear Mixed Models

Jake Rich

Coming to Theaters Sometime Spring 2018

Contents

Running Notes

1.

1 Introduction

Talk about modeling in Ecology/Evolutionary Biology to motivate GLMMs - keep in mind writing for statistics audience about applications in EE.

Zuur's Owl example to start?

1.1 Background

To motivate why there are a variety of fitting methods for GLMMs, we will first look at a very specific case of a GLMM, an ordinary linear regression model. Suppose we have the following linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times p$ design matrix which is assumed to be full column rank and have fixed values, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors. In this model, the $\boldsymbol{X}\boldsymbol{\beta}$ portion is the deterministic portion of the model while the $\boldsymbol{\varepsilon}$ portion is the probabilistic part of the model.

1. $\boldsymbol{\varepsilon} \overset{iid}{\sim} N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

2. $\boldsymbol{X}\boldsymbol{\beta}$ is linear in the β's.

The parameters in $\boldsymbol{\beta}$ are typically estimated using least squares estimation. Least squares are obtained through minimizing the sum of squared residuals (Faraway text). Provided error distribution and linearity assumptions hold, the OLS estimates will be BLUE, the best linear unbiased estimator, for the β's. The $\boldsymbol{\beta}$ effects and are specific to the sample and are not random variables. Importantly, the ordinary least squares estimator, $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$, is also the maximum likelihood (ML) estimator since it was specifed that the errors are assumed to be normally distributed.

In many applications, these assumptions are reasonably satisfied and OLS estimation can be used without issue. However, there are also many applications where one or more of those necessary assumptions are violated, which has lead to the development of other models. As of right now, the most versatile linear model available is the generalized linear mixed model (GLMM), which uses theory from linear mixed models which allow dependent errors (LMM) and generalized linear models (GLM) which allow for non-normally

3

distributed errors. Building up to GLMMs, we will take a quick look at these two other classes of models and we will start with linear mixed models.

1.1.1 Mixed Models

The assumption of independence of errors is a strong assumption that is quickly violated in many study designs with repeated measurements, clustered observations, or within spatial elements.

Suppose we wish to compare across a number of fifth grade classes in a school. It's reasonable to assume that students in one classroom with one teacher will be similar to some degree since they're all being taught by the same teacher, and the math scores might differ between classes. The independence assumption then would be violated as the responses (student scores) would be correlated at the classroom level due to many students being taught by the same teacher. Violating the independence assumption means that ordinary linear models could be used, but the estimates will no longer be BLUE.

Mixed models are an alternative to the independence violation as they account for correlation in between observations. In the testing example, students in each classroom would be correlated with one another. Having correlated observations leads to two major differences in the model. The first is that we would now have a second set of model components known as random effects, which changes the model to,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{Z}$ is the $n \times q$ random effects design matrix and $\boldsymbol{b}$ is a $q \times n$ vector of random effects. The second is that the errors are now allowed to be correlated in some way, whether that is by group membership, spatial, temporal, or spatio-temporal correlation of observations. The allowing of corrlated responses arises in the distribution of the errors, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \Sigma)$, where Σ is the error variance-covariance matrix.

Random effects are unobservable random variables with known distributions, typically $N(\boldsymbol{0}, D)$ where D is the unknown random effects variance-covariance matrix, that we want to estimate. For our math testing example, each class room would have a random effect in the $\boldsymbol{b}$ vector. Random effects differ from fixed effects in that the random effects are assumed to be a random sample of all possible levels and values of the variables with the intent to make inferences about the populations which the variables represent (506 Notes, mention fixed effects?). Random effects can take two forms. Random intercepts allow for different group means to vary around the overall mean while random slopes allow for a covariate to have a different slope for each group. For the test scores, random effects would allow for each class room to have it's own baseline mean by using random intercepts.

In addition to the random effects, the variance components of the random effects are also of interest. With the variance components, we can determine how our groups vary compared to the individuals and see what varies more. Additionally, we can use the estimated variance components to calculate the intraclass correlation, a measure of the correlation between subjects within the levels of the mixed model. In order to estimate the multiple variance components in addition to the slope parameters, we need to change the estimation method from ordinary least squares.

When using OLS while assuming errors are normally distributed, it was mentioned that the OLS estimator is not only BLUE, but it is the maximum likelihood estimate. With mixed models, we also could use ML estimation for the model parameters. Because of the additional components that need to be estimated though, we actually will end up with biased estimates of $\boldsymbol{\beta}$ since those are treated as fixed quantities. Instead of ML estimation, we need to use restricted maximum likelihood (REML) estimation. With REML, we are modifying the likelihood such that the residuals $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$ conditional on the random effect

1.1.2 Generalized Linear Models

As we saw in the owlet calls example, there are times when we would like to model a process but the errors are known to follow a distribution other than a normal distribution. In that case we knew that the number of calls per nest was likely to be distributed Poisson. While we might be able in certain cases to approximate a Poisson distribution with a normal distribution, we generally need to modify our linear model framework to accept other response error distributions. In these non-normal cases, we can turn to the generalized linear model.

Generalized linear models allow for the modeling of responses from any distribution from the exponential family of distributions. The most common distributions used in GLMs include the binomial, Poisson, negative binomial, and gamma distributions.

1.2 Generalized Linear Mixed Models

2 Methods of Fitting GLMMs

The theory and implementation of generalized linear mixed models is a fairly new method compared to others like ANOVA or ordinary linear regression, with the most development in computational methods for GLMMs occurring in the past twenty-five years. The first wide spread introduction to these models came from Mccullagh and Nelder when they introduced penalized quasi-likelihood for estimation of the

mean parameters of GLMMs (1989). Since then, a range of methods from both frequentist and Bayesian perspectives have been developed for inference on GLMMs

2.1 Frequentist Approaches

Most common frequentist statistical approaches rely on the likelihood (usually the log-likelihood for simplicity) for use in estimation. As has been pointed out before, the use of the full likelihood function with GLMMs is No frequentist approaches rely on the true likelihood due to the issues integration introduced with GLMMs - all are approximations.

Several methods exist and are categorized in to three categories based on the most commonly used methods in statistical practice.

2.1.1 Penalized Quasi-Likelihood

The first common method of estimating parameters for a GLMM is penalized quasi-likelihood, which works through approximating the likelihood integrand or the data (both are equivalent conceptually). Penalized quasi-likelihood was introduced by Breslow and Clayton

Essentially, the GLMM to be fit is approximated by a linear mixed model, and then the approximate mixed model is estimated.

Uses a first-order Taylor series approximation to linearize the link function, effectively linearizing the data and allowing for LMM methods to be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

2.1.2 Laplace Approximation

2.1.3 Adaptive Gauss-Hermite Quadrature

GHQ is at its heart a generalization of Laplace approximation, and when the number of points is $n = 1$, GHQ is Laplace approximation.

2.2 Bayesian Approaches

Bayesian methods give an altogether different f

2.2.1   Metropolis-Hastings/MCMC Methods

2.3   Monte Carlo EM-Algorithm Method

3   Simulation Study of Performance of Fitting Methods

3.1   Data Generation

3.2   Model Fitting

3.3   Model Evaluation

4   Discussion

5   References

6   R Code Appendix